

**UNIVERSITE TOULOUSE III – PAUL SABATIER**  
**U.F.R Sciences de la Vie et de la Terre**

**THÈSE**

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE TOULOUSE III**

***Discipline : Innovation pharmacologique***

présentée et soutenue

par

Pascal, GP MARTIN

le 29 novembre 2007

***ETUDE DU POTENTIEL DE REGULATION GENIQUE EXERCÉ PAR  
L'ISOFORME ALPHA DU RECEPTEUR ACTIVE PAR LES  
PROLIFERATEURS DE PEROXYSOMES (PPAR $\alpha$ )***

---

***Directeur de thèse :***  
Dr. Thierry PINEAU

---

**JURY**

Président :	<i>Pr Dominique Langin</i>
Rapporteur :	<i>Pr Béatrice Desvergne</i>
Rapporteur :	<i>Dr Paul Grimaldi</i>
Examinateur :	<i>Pr Robert Barouki</i>
Examinateur :	<i>Pr Stéphane Robin</i>
Examinateur :	<i>Dr Thierry Pineau</i>



**AUTEUR :** Pascal, Guy, Pierre MARTIN

**TITRE :** Etude du potentiel de régulation génique exercé par l'isoforme alpha du récepteur activé par les proliférateurs de peroxyosomes (PPAR $\alpha$ )

**DIRECTEUR DE THESE :** Dr Thierry PINEAU

**LIEU ET DATE DE SOUTENANCE :** Toulouse, le 29 novembre 2007

---

### **RESUME en français**

L'isoforme  $\alpha$  du récepteur activé par les proliférateurs de peroxyosomes (PPAR $\alpha$ ) est un récepteur nucléaire (Chapitre I). Suite à son activation par diverses molécules, PPAR $\alpha$  (Chapitre II) régule l'expression de nombreux gènes impliqués notamment dans le métabolisme des lipides. Grâce au développement d'une puce à ADN (Chapitre III) dédiée à PPAR $\alpha$  et à son partenaire de dimérisation le récepteur X aux rétinoïdes (RXR), nous avons étudié les régulations géniques provoquées par PPAR $\alpha$  lors de son activation par des stimuli i) pharmacologique (activateurs pharmacologiques de PPAR $\alpha$  et RXR, Chapitre IV), ii) nutritionnel (acides gras alimentaires, Chapitre V) et iii) physiopathologique (jeûne, Chapitre VI). Ces travaux montrent que PPAR $\alpha$  est très vraisemblablement l'objet d'une activation physiologique par les acides gras alimentaires ou issus de la lipolyse du tissu adipeux et que ses voies de signalisation subissent l'influence croisée de celles de son partenaire de signalisation RXR.

---

### **MOTS-CLES**

Récepteurs nucléaires, transcriptomique, macroarray, puce à ADN, proliférateur de peroxyosome, PPAR, fibrate, réxinoïde, acide gras, jeûne, foie

---

### **DISCIPLINE ADMINISTRATIVE :** Innovations pharmacologiques

---

**INTITULE ET ADRESSE DU LABORATOIRE :** Laboratoire de Pharmacologie et Toxicologie (UR66), INRA BP3, 180 chemin de Tournefeuille, 31931 Toulouse Cedex 9, FRANCE



## **Remerciements**

**A mes parents,**

**A ma famille,**

**A mes amis,**

**A ma femme Michela et à mon fils Lucas,**

Je tiens à remercier chaleureusement :

Monsieur le **Docteur Thierry Pineau**, Directeur de Recherche au laboratoire de Pharmacologie et Toxicologie de l'INRA, qui m'a accueilli dans son équipe comme doctorant puis comme ingénieur, m'a toujours accompagné, soutenu et fait confiance. Au delà de toute l'expérience que tu m'as permis d'acquérir, c'est le cadeau de ton amitié qui compte le plus pour moi. Les mots me manquent pour t'exprimer ma profonde reconnaissance, mon respect et mon amitié.

Madame le **Professeur Béatrice Desvergne**, Professeur au Centre de Génomique Intégrative de l'Université de Lausanne, pour m'avoir fait l'honneur d'accepter d'évaluer ce travail de doctorat en tant que rapporteur. Votre parcours, vos travaux et votre ouverture d'esprit constituent pour moi des références et des exemples à suivre.

Monsieur le **Docteur Paul Grimaldi**, Directeur de Recherche à l'Université de Nice, pour m'avoir également fait l'honneur d'accepter d'évaluer, en tant que rapporteur, ce volumineux travail de thèse. Votre passion et votre enthousiasme pour la science et la recherche ainsi que la perspective de pouvoir peut-être un jour travailler en collaboration avec vous sont pour moi des moteurs essentiels dans mon début de carrière.

Monsieur le **Professeur Dominique Langin**, Directeur de l'unité INSERM U586 de recherche sur les obésités pour avoir accepté de présider le jury d'évaluation de ce travail de

doctorat. Ta bienveillance et tes compétences scientifiques sont également des références pour moi. Avec l'espoir que nos travaux nous conduisent dans le futur à collaborer de manière plus étroite, je souhaite t'exprimer ici toute ma reconnaissance.

Monsieur le **Professeur Robert Barouki**, Directeur de l'UMR 747 INSERM-Université Paris Descartes de Pharmacologie, Toxicologie et Signalisation Cellulaire pour avoir accepté de participer à l'évaluation de ce travail de doctorat. Je souhaite ici vous exprimer toute ma gratitude pour votre bienveillance et votre regard toujours constructif sur nos travaux.

Monsieur le **Professeur Stéphane Robin**, Professeur à l'UMR 518 INRA-AgroParisTech de Mathématiques et Informatique Appliquées pour m'avoir fait l'honneur d'accepter d'évaluer ce travail de doctorat en biologie au sein duquel sont discrètement disséminées quelques notions de statistique. Nos trop rares échanges ont toujours été pour moi source de réflexion et de motivation supplémentaire et je souhaite vous exprimer ici toute ma reconnaissance pour votre bienveillance à mon égard.

J'adresse également mes plus sincères remerciements,

A mes **parents Denis et Dominique** qui m'ont toujours soutenus et encouragés dans mes études et dans ma vie professionnelle et m'ont constamment apporté leur amour.

A mon frère **Patrick** et à ses trois princesses : **Marianne, Isabelle et Adèle** pour avoir toujours été à mes côtés dans les moments importants.

A ma femme **Michela** pour ton amour et toute la logistique que tu as assuré au quotidien lorsque la rédaction occupait mes journées. Je n'aurais pas pu réaliser ce travail sans ton aide et ton soutien.

A mon fils **Lucas** dont les sourires me comblent de bonheur chaque jour.

A mes **collègues** du Laboratoire de Pharmacologie et Toxicologie de l'INRA de Toulouse et en particulier aux membres de l'équipe de Pharmacologie Moléculaire : **Frédéric**,

**Arnaud, Laila, Colette, Gérard, Arlette, Cécile et Gilberte** pour m'avoir accompagné dans la réalisation de ces travaux et surtout pour participer à créer dans l'équipe et l'unité une ambiance aussi agréable. Mes pensées vont également à **Jocelyne Laumailler** et à **Nicole Chevreau** qui nous ont quittés trop tôt.

Aux **stagiaires**, doctorants ou masters, qui sont passés dans l'équipe et ont grandement contribués à ces travaux, en particulier à **Alain** Roulet pour m'avoir accueilli dans l'équipe, à **Romain** Barnouin pour tes travaux sur le jeûne et l'ambiance inoubliable que tu as su créer lors de ton passage au labo, à **Marie** De Tayrac, pour tes travaux sur les phtalates et sur l'analyse des données ainsi que pour nos discussions enrichissantes sur l'interface entre disciplines, à **Alexandre** Eveillard pour tes travaux sur les plastifiants et l'évolution de la puce INRArray, pour nos échanges footbalistiques, scientifiques ou de jeunes papas (c'est quoi ta marque préférée de biberons?) et pour ton amitié, à **Alice** Vigneron pour la mise en place du script R de traitement des données et pour nos discussions quasi-philosophiques de temps en temps, et bien sûr à tous les autres pour leurs contributions.

Aux nombreux **collaborateurs** qui nous ont fait confiance et nous ont apporté de nouveaux points de vue enrichissants sur nos travaux. Je pense en particulier à Maurice Bugaut qui nous a quitté bien trop tôt, Stéphane Savary, Jean-Marc Pascussi, Isabelle Niot, Mauro Dacasto, Sandrine Lagarrigue, Alain Paris, Yves Lecarpentier, David Causier, Christelle Robert-Granié, Magali San Cristobal et Xavier Collet. Je pense aussi à tous les autres que je ne cite pas nommément ici mais à qui je souhaite exprimer ma profonde reconnaissance.

Aux membres de l'**IFC** : **Alex, JP, Cédric, Alex, Vince, Ben, Sylvain, Bruno, Nono, et Manu**, pour votre amitié et les moments indispensables passés à se défouler ensemble autour d'un ballon rond.

A **Grégory** Chevillard, **Stéphane** Fourcade, **Cathy** et **Laurent** Corcos et tout le LBMC de Dijon pour nos collaborations fructueuses mais surtout pour les bons moments passés ensemble et l'amitié que nous avons créée.

Au trio gagnant du LSP : **Alain** Baccini, **Philippe** Besse et **Sébastien** Déjean pour m'avoir accepté au sein de vos discussions sur l'analyse des données, pour nos collaborations fructueuses et perennes, pour m'avoir accompagné vers une meilleure connaissance de votre domaine d'expertise et pour votre amitié.

Au laboratoire de Biochimie de l'ENSA de Rennes pour la collaboration fructueuse et enrichissante sur les aspects "lipides" de ce travail. Plus particulièrement à **Hervé Guillou** pour ton amitié constante et pour nos discussions interminables, qu'elles soient scientifiques ou plus personnelles, qui m'apportent toujours un regain de motivation, d'enthousiasme et de réflexion.

A mes **amis**, en particulier les cassins de Bourges, les toulousains Nono, Jerem', Sly et Delphine, les parisiens Aram et Arash, Charly et Man et Gaia ainsi qu'à mes amis de Rennes, Jules, Tanguy, Pierre-Yves, Max, TyBen, Mumu et tous les autres. Votre amitié est un moteur essentiel de ma vie.

Enfin à tous ceux que j'ai involontairement oubliés. Qu'ils trouvent ici la marque de ma reconnaissance et de ma sympathie.

# Table des matières

<i>Liste des Illustrations</i>	p5
<i>Liste des abréviations</i>	p10
<i>Liste des valorisations scientifiques</i>	p11
• <i>Articles en premier auteur</i>	p11
• <i>Articles dans le domaine de la statistique</i>	p11
• <i>Articles en tant que co-auteur</i>	p11
• <i>Communications orales (congrès et séminaires)</i>	p13
• <i>Communications par poster</i>	p16
• <i>Formations (Intervenant)</i>	p18
<b>INTRODUCTION</b>	<b>p19</b>
<b><u>Chapitre I : Les récepteurs nucléaires</u></b>	<b>p21</b>
<b>I.1. Nomenclature</b>	p21
<b>I.2. Structure canonique des récepteurs nucléaires</b>	p24
I.2.1. Les domaines A et B	p24
I.2.2. Le domaine C de fixation à l'ADN	p25
I.2.3. Le domaine charnière D	p33
I.2.4. Le domaine E de fixation du ligand	p35
I.2.5 Le domaine F	p36
<b>I.3. Eléments de réponse et dimérisation des récepteurs nucléaires</b>	p43
I.3.1. Eléments de réponse et classe des récepteurs nucléaires	p43
I.3.2. Dimérisation des récepteurs nucléaires	p49
I.3.3. Les tétramères de RXR	p50
<b>I.4. Ligands des récepteurs nucléaires</b>	p51
I.4.1. Ligands endogènes des récepteurs nucléaires	p54
I.4.2. Méthodes d'analyse des interactions récepteur-ligand	p56
I.4.3. Origines et concentrations physiologiques des ligands endogènes des récepteurs nucléaires	p58
I.4.4. Ligands synthétiques ou naturels des récepteurs nucléaires	p71
I.4.5. Mécanismes et conséquences de la fixation des ligands	p78
<b>I.5. Corégulateurs et régulation de la transcription</b>	p84
I.5.1. Facteurs généraux de la transcription	p85

I.5.2. Corégulateurs et remodelage de la chromatine	p87
I.5.3. Corégulateurs et modifications covalentes des histones	p89
I.5.4. Complexes Médiateur	p97
I.5.5. Autres corégulateurs	p99
I.5.6. Dynamique de fonctionnement des corégulateurs	p103
<b>I.6. Modèles généraux du fonctionnement des voies de signalisation des récepteurs nucléaires</b>	p103
 <b><u>Chapitre II : Le récepteur PPAR<math>\alpha</math></u></b>	<b>p107</b>
<b>II.1. Clonage, localisation chromosomique et expression</b>	p107
I.1.1. Clonage des récepteurs PPAR	p107
I.1.2. Localisation chromosomique et organisation des gènes PPAR	p109
I.1.3. Expression et régulation des PPARs	p110
<b>II.2. Ligands des PPARs</b>	p114
II.2.1. Ligands endogènes potentiels	p114
II.2.2. Ligands synthétiques	p121
II.2.3. Autres situations d'activation de PPAR $\alpha$	p126
<b>II.3. Fonctionnement de la voie de signalisation PPAR<math>\alpha</math></b>	p128
II.3.1. Interaction avec les protéines chaperones	p128
II.3.2. Domaines A/B et phosphorylation de PPAR $\alpha$	p128
II.3.3. Domaine C et éléments de réponse de PPAR $\alpha$	p132
II.3.4. Domaines D et E et corégulateurs de PPAR $\alpha$	p133
II.3.5. Dynamique de fonctionnement de la voie PPAR $\alpha$	p139
<b>II.4. Régulation du métabolisme par PPAR<math>\alpha</math></b>	p141
II.4.1. Régulation du métabolisme des lipoprotéines par PPAR $\alpha$	p141
II.4.2. Régulation du transport et du trafic intracellulaire des acides gras par PPAR $\alpha$	p151
II.4.3. Régulation du métabolisme intracellulaire des lipides par PPAR $\alpha$	p160
II.4.4. Régulation du métabolisme des glucides et des acides aminés par PPAR $\alpha$	p186
<b>II.5. Autres fonctions de PPAR<math>\alpha</math></b>	p191
II.5.1. Régulation du métabolisme des xénobiotiques par PPAR $\alpha$	p191

II.5.2. Rôles de PPAR $\alpha$ dans l'inflammation au niveau vasculaire	p197
II.5.3. Rôles de PPAR $\alpha$ dans l'inflammation au niveau hépatique	p202
<b><u>Chapitre III : Méthodes d'analyse de l'expression des gènes</u></b>	<b>p207</b>
<b>III.1. Techniques de mesure de l'expression génique à « bas débit »</b>	p209
III.1.1. Le northern blot	p209
III.1.2. La protection aux RNases	p213
III.1.3. La PCR en temps réel	p215
<b>III.2. Techniques de mesure de l'expression génique à « haut-débit »</b>	p237
III.2.1. Les méthodes basées sur le séquençage d'étiquettes (ou tags)	p239
III.2.2. Les méthodes basées sur l'amplification de sous-populations d'ARNm	p244
III.2.3. Les puces à ADN	p249
<b>III.3. Développement de la puce INRAarray</b>	p261
III.3.1. Justification des choix technologiques	p261
III.3.2. Etapes de mise au point de la puce INRAarray	p263
III.3.3. Clonage des ADNc	p271
III.3.4. Générations de la puce INRAarray	p275
<b>III.4. Analyse statistique des données issues des puces à ADN</b>	p280
III.4.1. Qualité des données, filtres et prétraitements	p280
III.4.2. Normalisation des données	p284
III.4.3. Comparaison de classes	p289
III.4.4. Découverte de classes	p299
III.4.5. Prédiction de classes	p305
III.4.6. Aide à l'interprétation / Aspects bioinformatiques	p308
III.4.7. Analyse des puces INRAarray 01	p310
<b>III.5. Mise en œuvre de diverses méthodes sur un même jeu de données</b>	p313

**Article n°1 :** Alain Baccini, Philippe Besse, Sébastien Déjean, Pascal G. P. Martin, Christelle Robert-Granié, Magali San Cristobal. Stratégies pour l'analyse statistique de données transcriptomiques. *Journal de la Société Française de Statistique*. 2005 ; 146(1-2): 5-44.

<b><u>Chapitre IV : Etude d'agonistes synthétiques de RXR</u></b>	<b>p315</b>
<b>IV.1. Introduction</b>	p315
<b>IV.2. Présentation des résultats publiés</b>	p316
<b>Article n°2 :</b> Pascal G. P. Martin, Frédéric Lasserre, Cécile Calléja, Armelle van Es, Alain Roulet, Didier Concordet, Michela Cantiello, Romain Barnouin, Béatrice Gauthier, Thierry Pineau. Transcriptional modulations by RXR agonists are only partially subordinated to PPAR $\alpha$ signalling and attest additional, organ-specific, molecular cross-talks. <i>Gene Expression</i> . 2005 ; 12: 177-192.	
<b>IV.3. Conclusions / perspectives</b>	p317
<b><u>Chapitre V : Etude des effets de lipides alimentaires sur la voie PPAR<math>\alpha</math></u></b>	<b>p319</b>
<b>V.1. Introduction</b>	p319
<b>V.2. Présentation des résultats publiés</b>	p320
<b>Article n°3 :</b> Pascal G. P. Martin, Hervé Guillou, Frédéric Lasserre, Sébastien Déjean, Annaïg Lan, Jean-Marc Pascussi, Magali SanCristobal, Philippe Legrand, Philipe Besse, Thierry Pineau. Novel aspects of PPAR $\alpha$ -mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study. <i>Hepatology</i> . 2007 ; 45(3): 767-777	
<b>V.3. Résultats complémentaires</b>	p321
<b>V.4. Conclusions / perspectives</b>	p328
<b><u>Chapitre VI : Etude cinétique des effets du jeûne sur l'expression génique</u></b>	<b>p331</b>
<b>VI.1. Introduction</b>	p331
<b>VI.2. Matériels et méthodes</b>	p336
<b>VI.3. Résultats</b>	p339
<b>VI.4. Discussion</b>	p374
<b>CONCLUSION</b>	<b>p382</b>
<b>BIBLIOGRAPHIE</b>	<b>p385</b>
<b><u>ANNEXE 1 : Protocole d'utilisation des puces INRAarray 01</u></b>	<b>p437</b>
<b><u>ANNEXE 2 : Liste des clones présents sur la puce INRAarray 01.4</u></b>	<b>p449</b>
<b><u>ANNEXE 3 : Article soumis sur l'analyse canonique régularisée</u></b>	<b>p455</b>
<b><u>ANNEXE 4 : Article publié sur l'analyse des données de cinétique</u></b>	<b>p457</b>

## Liste des illustrations

<b>Chapitre I : Les récepteurs nucléaires</b>	<b>p21</b>
<u>Table 1</u> : Liste des 49 récepteurs nucléaires murins.....	p22
<u>Figure 1</u> : Organisation des récepteurs nucléaires en six domaines fonctionnels indépendants.....	p23
<u>Figure 2</u> : Mécanismes de fixation des récepteurs nucléaires à l'ADN .....	p26
<u>Figure 3</u> : Eléments de réponse des récepteurs nucléaires .....	p26
<u>Figure 4</u> : Alignement des domaines de fixation à l'ADN des récepteurs nucléaires murins.....	p27
<u>Figure 5</u> : Représentation schématique du domaine de fixation à l'ADN de RXR ..	p28
<u>Figure 6</u> : Structure tridimensionnelle du domaine de fixation à l'ADN de RXR....	p28
<u>Figure 7</u> : Exemples de structures de domaines de fixation à l'ADN de différents récepteurs nucléaires .....	p30
<u>Table 2</u> : Structures élucidées des DBD des récepteurs nucléaires.....	p32
<u>Figure 8</u> : Domaine de fixation du ligand de RXR $\alpha$ .....	p34
<u>Table 3</u> : Structures élucidées des LBD des récepteurs nucléaires .....	p37-43
<u>Figure 9</u> : Les quatre classes de récepteurs nucléaires .....	p44
<u>Table 4</u> : Exemples d'éléments de réponse naturels des récepteurs RAR .....	p47
<u>Figure 10</u> : Interface de dimérisation du LBD des récepteurs nucléaires .....	p48
<u>Table 5</u> : Principaux ligands endogènes des récepteurs nucléaires .....	p52-54
<u>Table 6</u> : Exemples de ligands non endogènes des récepteurs nucléaires.....	p67-71
<u>Figure 11</u> : Changements conformationnels du LBD de RXR $\alpha$ engendrés par la fixation d'un ligand .....	p77
<u>Figure 12</u> : Comparaison des sites de fixation des coactivateurs et corépresseurs sur le LBD de PPAR $\alpha$ .....	p77
<u>Figure 13</u> : Représentation schématique du fonctionnement des corégulateurs agissant par remodelage de la chromatine .....	p86
<u>Figure 14</u> : Représentation schématique du fonctionnement des coactivateurs provoquant des modifications covalentes des histones .....	p88
<u>Figure 15</u> : Domaines fonctionnels et structuraux des coactivateurs SRC/p160 .....	p90
<u>Figure 16</u> : Représentation schématique du fonctionnement des complexes Médiateur .....	p96

<u>Figure 17</u> : Schéma de fonctionnement des voies de signalisation des RNs de classe I .....	p102
<u>Figure 18</u> : Schéma de fonctionnement des voies de signalisation des RNs de classe II.....	p104
<b><u>Chapitre II : Le récepteur PPAR<math>\alpha</math></u></b> .....	<b>p107</b>
<u>Figure 19</u> : Alignement des séquences protéiques des PPAR murins .....	p108
<u>Figure 20</u> : La $\beta$ -oxydation peroxyxsomale des acides gras à longue chaîne .....	p116
<u>Figure 21</u> : Effets du fénofibrate chez les souris TTD sur l'expression hépatique d'un panel de gènes .....	p130
<u>Figure 22</u> : Fonctionnement général de la voie PPAR $\alpha$ .....	p140
<u>Figure 23</u> : Schéma général du transport du cholestérol .....	p142
<u>Table 7</u> : Composition biochimique et propriétés des principales lipoprotéines humaines.....	p143
<u>Table 8</u> : Panel d'éléments de réponse fonctionnels des récepteurs PPAR $\alpha$ et Rev-erb $\alpha$ .....	p145-146
<u>Figure 24</u> : Principaux effets de PPAR $\alpha$ sur le métabolisme des lipoprotéines .....	p150
<u>Table 9</u> : Exemples d'éléments de réponse naturels du récepteur LXR.....	p152
<u>Figure 25</u> : Entrée dans la mitochondrie et activation des acides gras à longue chaîne.....	p156
<u>Figure 26</u> : Principaux effets de PPAR $\alpha$ sur le captage et le trafic intracellulaire des acides gras .....	p158
<u>Figure 27</u> : La $\beta$ -oxydation mitochondriale des acides gras .....	p162
<u>Figure 28</u> : Biosynthèse des corps cétoniques (cétogenèse) .....	p164
<u>Figure 29</u> : Chaîne respiratoire mitochondriale et protéines de découplage .....	p168
<u>Figure 30</u> : Principales étapes de la biosynthèse des acides biliaires.....	p172
<u>Table 10</u> : Exemples d'éléments de réponse naturels de FXR .....	p174
<u>Figure 31</u> : Biosynthèse des acides gras et origine de l'acétyl-CoA cytosolique .....	p178
<u>Figure 32</u> : Biosynthèse des acides gras longs polyinsaturés.....	p180
<u>Figure 33</u> : Glycolyse et néoglucogenèse.....	p184
<u>Figure 34</u> : Principaux effets de PPAR $\alpha$ sur l'inflammation .....	p196

<b>Chapitre III : Méthodes d'analyse de l'expression des gènes .....</b>	<b>p207</b>
<u>Figure 35</u> : Principe général du northern blot .....	p208
<u>Figure 36</u> : Principe général du dosage par protection aux RNases .....	p212
<u>Figure 37</u> : Principe général de la PCR .....	p216
<u>Figure 38</u> : La courbe de PCR .....	p217
<u>Figure 39</u> : SYBR Green et sondes TaqMan.....	p218
<u>Figure 40</u> : Courbes de dissociation en SYBR Green .....	p220
<u>Figure 41</u> : Répliqats techniques en PCR en temps réel.....	p222
<u>Figure 42</u> : Définition des cycles seuils (C <sub>t</sub> ) .....	p224
<u>Figure 43</u> : Utilisation d'une courbe standard en PCR en temps réel .....	p226
<u>Figure 44</u> : Effets de l'efficacité sur la courbe de PCR.....	p230
<u>Figure 45</u> : Estimation de l'efficacité à partir d'une courbe de PCR .....	p232
<u>Figure 46</u> : Principe général de la méthode SAGE .....	p238
<u>Figure 47</u> : Principe général du Differential Display .....	p242
<u>Figure 48</u> : Principe général des puces à ADN (simple couleur) .....	p250
<u>Figure 49</u> : Principe des puces à ADN double couleur .....	p252
<u>Figure 50</u> : Mise au point de la quantité de matériel spotté sur les puces INRArray .....	p264
<u>Figure 51</u> : Exemple d'hybridation dun ARN en quadruplicat sur les puces INRArray 01 .....	p266
<u>Figure 52</u> : Comparaison de résultats issus d'hybridations distinctes sur les puces INRArray 01 .....	p268
<u>Table 11</u> : Rapports d'expression théorique et observés de 8 spikes de levure .....	p271
<u>Figure 53</u> : Principales étapes du clonage des ADNc déposés sur les puces INRArray 01 .....	p272
<u>Figure 54</u> : Principales fonctions biologiques explorées par l'INRArray 01.4 .....	p276
<u>Figure 55</u> : Risques et taux d'erreurs pour les procédures de tests simples et multiples .....	p292
<u>Figure 56</u> : Principe général des analyses statistiques exploratoires multidimensionnelles .....	p298

**Chapitre IV : Etude d'agonistes synthétiques de RXR** **p315**

**Chapitre V : Etude des effets de lipides alimentaires sur la voie PPAR $\alpha$**  p319

- Figure 57 : Effets du TCPOBOP sur quelques gènes cibles hépatiques ..... p322  
Figure 58 : Induction de Cyp2b10 par les activateurs de CAR..... p323  
Figure 59 : Fixation de CAR sur le promoteur de Cyp2b10 ..... p324  
Figure 60 : ACP des profils d'expression intestinaux dans l'étude de nutrition ..... p326

**Chapitre VI : Etude cinétique des effets du jeûne sur l'expression génique** p331

- Figure 61 : Le cycle des Cori ..... p330  
Figure 62 : Le cycle de l'urée ..... p332  
Figure 63 : Evolution pondérale des souris et de leurs organes au cours du jeûne... p338  
Figure 64 : Dosages de glucose, d'urée et de  $\beta$ -hydroxybutyrate circulants ..... p340  
Figure 65 : Evolution des teneurs en acides gras du TABe et du foie au cours du jeûne ..... p342  
Figure 66 : ACP de la composition en acides gras du TABe ..... p344  
Figure 67 : Profils des proportions d'acides palmitique et stéarique dans le TABe.. p346  
Figure 68 : Profils des proportions d'acides gras monoinsaturés dans le TABe ..... p347  
Figure 69 : Profils des proportions d'acides gras polyinsaturés dans le TABe ..... p348  
Figure 70 : ACP de la composition hépatique en acides gras ..... p350  
Figure 71 : Profils des proportions d'acides gras saturés dans le foie ..... p352  
Figure 72 : Profils des proportions d'acides gras monoinsaturés dans le foie..... p353  
Figure 73 : Profils des proportions d'acides gras polyinsaturés dans le foie..... p354  
Figure 74 : Classification des gènes et des échantillons sur la base des profils d'expression ..... p356  
Figure 75 : Profils d'expression hépatique de 4 apolipoprotéines ..... p358  
Figure 76 : Profils d'expression hépatique de l'ASAT, de la PEPCK, de la L-CPTI et de Tp $\beta$ ..... p360  
Figure 77 : Profils d'expression hépatique de gènes impliqués dans l'oxydation peroxysonmale ..... p362  
Figure 78 : Profils d'expression hépatique de gènes impliqués dans l'oxydation mitochondriale..... p364  
Figure 79 : Profils d'expression hépatique de gènes impliqués dans l'oxydation des AGPI ..... p366  
Figure 80 : Profils d'expression hépatique des cytochromes P450 4a..... p367

Figure 81 : Profils d'expression hépatique de L-FABP, catalase, ALDH1 et LEF1 . p368

Figure 82 : Profils d'expression hépatique de gènes impliqués dans la biosynthèse

des acides gras ..... p370

Figure 83 : Profils d'expression hépatique de gènes de la biosynthèse

du cholestérol ..... p372

## Liste des abréviations

Seules les abréviations les plus utilisées dans ce manuscrit sont présentées ici. Les autres sont précisées dans le texte au fur et à mesure de leur utilisation.

<b>ADN :</b>	Acide Désoxyribonucléique
<b>ADNc :</b>	Acide Désoxyribonucléique complémentaire
<b>AG :</b>	Acide gras
<b>AGPI :</b>	Acide gras polyinsaturé
<b>Apo :</b>	Apolipoprotéine
<b>ARNm :</b>	Acide Ribonucléique messager
<b>C/EBP :</b>	CAAT/enhancer binding protein
<b>CYP :</b>	Cytochrome P450
<b>FAD :</b>	Flavine adénine dinucléotide
<b>FADH:</b>	forme réduite du FAD
<b>FADH<sub>2</sub>:</b>	forme réduite du FAD
<b>HDL :</b>	High density lipoprotein
<b>IDL :</b>	Intermediate density lipoprotein
<b>LDL :</b>	Low density lipoprotein
<b>Lp-B :</b>	Lipoprotéines à apoB (VLDL, IDL, LDL et chylomicrons)
<b>NAD<sup>+</sup> :</b>	Nicotinamide adénine dinucléotide (forme oxydée)
<b>NADH :</b>	Nicotinamide adénine dinucléotide (forme réduite)
<b>NF-κB :</b>	Nuclear Factor-kappa B
<b>pb :</b>	paire de bases (unité de mesure de l'ADN qui se décline en Kb, Mb, etc.)
<b>PP :</b>	Proliférateur de peroxysomes
<b>PPAR :</b>	Peroxisome proliferator-activated receptor
<b>RN :</b>	Récepteur nucléaire
<b>RXR:</b>	Retinoid X Receptor
<b>SDS:</b>	Sodium dodécyl sulfate
<b>SREBP:</b>	Sterol regulatory element-binding protein
<b>SSC:</b>	Tampon salin sodium citrate
<b>STAT :</b>	Signal transducer and activator of transcription
<b>TNF-α:</b>	Tumor necrosis factor α
<b>VLDL :</b>	Very low density lipoprotein

## **Liste des valorisations scientifiques**

### **Articles en premier auteur**

1. Martin PGP, Lasserre F, Calléja C, van Es A, Roulet A, Concorde D, Cantiello M, Barnouin R, Gauthier B, Pineau T. Transcriptional modulations by RXR agonists are only partially subordinated to PPAR $\alpha$  signalling and attest additional, organ-specific, molecular cross-talks. *Gene Expression* 2005 ; 12: 177-192
2. Martin PGP, Guillou H, Lasserre F, Déjean S, Lan A, Pascussi JM, SanCristobal M, Legrand P, Besse P, Pineau T. Novel aspects of PPAR $\alpha$ -mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study. *Hepatology* 2007 ; 45(3): 767-777

### **Articles dans le domaine de la statistique**

1. Baccini A, Besse P, Déjean S, Martin PGP, Robert-Granié C, San Cristobal M. Stratégies pour l'analyse statistique de données transcriptomiques. *J. Soc. Fr. Statistique* 2005 ; 146(1-2):5-44
2. Déjean S, Martin PGP, Baccini A, Besse P., Clustering Time-Series Gene Expression Data Using Smoothing Spline Derivatives. *EURASIP Journal on Bioinformatics and Systems Biology*, vol. 2007, Article ID 70561, 10 pages, 2007. doi:10.1155/2007/70561

### **Articles en tant que co-auteur**

1. Lecarpentier Y, Krokidis X, Martin P, Pineau T, Hébert J-L, Quillard J, Cortes-Morichetti M, Coirault C. Increased entropy production in diaphragm muscle of PPAR alpha knockout mice. *J. Theor Biol* 2007 ; sous presse
2. Loiseau N, Debrauwer L, Sambou T, Bouhet S, Miller JD, Martin PG, Viadère JL, Pinton P, Puel O, Pineau T, Tulliez J, Galtier P, Oswald IP. Fumonisins B(1) exposure and its selective effect on porcine jejunal segment: Sphingolipids, glycolipids and trans-epithelial passage disturbance. *Biochem Pharmacol* 2007 ; 74(1):144-52.
3. Petit V, Arnould L, Martin P, Monnot MC, Pineau T, Besnard P, Niot I. Chronic high-fat diet affects intestinal fat absorption and postprandial triglyceride levels in the mouse. *J Lipid Res* 2007 ; 48(2):278-87.

4. Bassissi MF, Alvinerie M, Martin PG, Perret B, Lespine A Influence of dyslipidemia on moxidectin distribution in plasma lipoproteins and on its pharmacokinetics. *Pharm Res* 2006 ; 23(11):2672-80.
5. Goncaille C, Depreter M, Fourcade S, Lecca MR, Leclercq S, Martin P, Pineau T, Cadepond F, El Etr M, Bertrand N, Beley A, Duclos S, de Craemer D, Roels F, Savary S, Bugaut M. Phenyl butyrate up-regulates the Adrenoleucodystrophy-related gene as a non-classical peroxisome proliferator. *Journal of Cell Biology* 2005 ; 169:93-104.
6. Pascussi JM, Robert A, Nguyen M, Walrant-Debray O, Garabedian M, Martin P, Pineau T, Saric J, Navarro F, Maurel P, Vilarem MJ. Possible involvement of pregnane X receptor-enhanced CYP24 expression in drug-induced osteomalacia. *Journal of Clinical Investigation* 2005 ; 115:177-186.
7. Le Josic-Corcos C, Duclos S, Ramirez L, Zaghini I, Chevillard G, Martin P, Pineau T, Bournot P. Effects of Peroxisome proliferator-activated receptor  $\alpha$  (PPAR $\alpha$ ) activation on pathways contributing to cholesterol homeostasis in rat hepatocytes. *Biochim. Biophys. Acta.(Mol. Cell. Biol. L.)* 2004 ; 1683:49-58.
8. Chevillard G, Clémencet M-C, Etienne P, Martin P, Pineau T, Latruffe N, Nicolas-Frances V. Molecular cloning, gene structure and expression profile of the two mouse peroxisomal 3-ketoacyl-CoA thiolase genes. *BMC Biochem.* 2004 ; 5:3 (Epub.).
9. Bugaut M, Fourcade S, Gondcaille C, Gueugnon F, Depreter M, Roels F, Netik A, Berger J, Martin P, Pineau T, Cadepond F, El Etr M, Savary S. Pharmacological induction of redundant genes for a therapy of X-ALD: phenylbutyrate and other compounds. *Adv Exp Med Biol.* 2003 ; 544:281-91.
10. Chevillard G, Clemencet MC, Etienne P, Martin P, Pineau T, Latruffe N, Nicolas-Frances V. Tissue-specific expression of two peroxisomal 3-ketoacyl-CoA thiolase genes in wild and PPAR alpha-null mice and induction by fenofibrate. *Adv Exp Med Biol.* 2003 ; 544:55-6.
11. Guillou H, Martin P, Jan S, D'Andrea S, Roulet A, Catheline D, Rioux V, Pineau T, Legrand P. Comparative effect of fenofibrate on hepatic desaturases in wild-type and peroxisome proliferator-activated receptor alpha-deficient mice. *Lipids.* 2002 ; 37(10):981-9.
12. Lambert G., Amar M. J.A., Martin P., Fruchart-Najib J., Föger B., Shamburek R.D., Brewer H.B. Jr. and Santamarina-Fojo S. Hepatic lipase deficiency decreases the selective uptake of HDL-cholesteryl esters in vivo. *J. Lipid Res.* 2000 ; 41:667-672

## Communications orales (congrès et séminaires)

[S] : communication orale sélectionné par les organisateurs du congrès

[I] : communication orale sur invitation

Le nom de la personne qui a réalisé la communication orale est souligné

1. Eveillard A, Martin P, Lasserre F, Lakhal L, Guillou H, Roulet A, De Tayrac M, Calléja C, Déjean S, Claus S, Canlet C, Paris A, Pineau T. Contribution of Genomics in Toxicology. International Symposium on Animal Genomics for Animal Health - Paris, France, 23-25 october 2007 [I]
2. Pineau T, Lasserre F, Eveillard A, Mselli-Lakhal L, Martin PGP. Endocrine disruptive compounds (EDCs) : mechanisms and effects in domestic animals. 10th International Congress of the European Association for Veterinary Pharmacology and Toxicology EAVPT - Torino, Italia, 17-22 September 2006 [I]
3. Pineau T, Lasserre F, Eveillard A, Mselli-Lakhal L, Martin PGP. Comprehensive investigation of the transcriptome. 10th International Congress of the European Association for Veterinary Pharmacology and Toxicology EAVPT - Torino, Italia, 17-22 September 2006 [I]
4. Cantiello M, Carletti M, Dacasto M, Capolongo F, Gardini G, Martin PGP, Pineau T, Nebbia C. Catalytic, immunochemical and molecular characterization of the xenobiotic-metabolising enzyme induction by phenobarbital in the bovine liver. 10th International Congress of the European Association for Veterinary Pharmacology and Toxicology EAVPT - Torino, Italia, 17-22 September 2006 [S]
5. Robert-Granié C, Baccini A, Besse P, Déjean S, Ferré PJ, Liaubet L, Martin PGP, San Cristobal M. Kinetics analysis of microarray data using semiparametric mixed models. 8th World Congress on Genetics Applied to Livestock Production. Belo-Horizonte, Brazil, August 13-18, 2006 [S]
6. Robert-Granié C, San Cristobal M, Martin PGP. Kinetics analysis of microarray data using semiparametric mixed models. The XXIIIth International Biometric Conference. Montréal, Québec, Canada, July 16-21, 2006 [S]
7. Eveillard A, de Tayrac M, Claus S, Canlet C, Gotardi G, Lasserre F, Martin P, Paris A, Pineau T. Assessing the low dose-impact of a plasticizer (DEHP) by means of coordinated gene profiling/NMR investigations. SAFE Consortium International Congress, Budapest, Hungary, 11-14 June 2006 [S]
8. Baccini A, Besse P, Déjean S, Martin P. Analysis of time series gene expression data. 4<sup>th</sup> workshop on statistical method for post-genomic data, Toulouse, 30-31 mars 2006 [S]

9. Martin PGP, Guillou H, Lasserre F, Déjean S, Legrand P, Besse P, Pineau T. Empreintes transcriptionnelles de lipides alimentaires sur modèles souris. Analyse intégrée de signatures variées. GENOTOUL, Toulouse, 18 novembre 2005 [I]
10. Cantiello M, Giantin M, Nebbia C, Bollo E, Lasserre F, Martin P, Pineau T, Dacasto M, Sequencing and cloning of some relevant drug metabolism genes in cattle: the “Programma Galileo” 2003 summary report. Galileo Project – 2005 Meeting, Siena, Italy, 12-13 September 2005 [S]
11. Baccini A, Besse P, Déjean S, Martin PGP, Robert-Granié C, San Christobal M, Etude de données cinétiques issues de biopuces, XXXVIIIèmes Journées de Statistiques de la Société Française de Statistiques, Pau - France, 6-10 juin 2005 [S]
12. Martin PGP, Pineau T. La méthodologie QSAR en Toxicologie prédictive. Congrès de l’Association pour la recherche en Toxicologie (ARET), Paris, 19-20 Mai 2005 [I]
13. Martin PGP, Barnouin R, Lasserre F, Guillou H, Cantiello M, Déjean S, Besse P, Pineau T. A kinetic view of coordinate modulations of gene expression and metabolism in wild-type and PPAR $\alpha$ -/- mice during fasting. Cycles de conférences de l'IFR31, Toulouse, 10 mai 2005. [I]
14. Martin PGP, Lasserre F, Guillou H, Barnouin R, Dejean S, Gotardi G, Legrand P, Baccini A, Besse P, Paris A. & Pineau T. Influence des acides gras alimentaires sur les voies de signalisation des récepteurs nucléaires : une étude comparée de signatures transcriptomique et lipidomique. Journées Scientifiques du GIS « PA-3S », Boussens, 9-10 Mai 2005 [I]
15. Martin PGP, Lasserre F, Guillou H, Barnouin R, Dejean S, Gotardi G, Legrand P, Baccini A, Besse P, Paris A. & Pineau T. PPAR $\alpha$  remplit-il véritablement la fonction de senseur nutritionnel vis-à-vis des acides gras alimentaires ? Une étude comparée de signatures transcriptomique et lipidomique. Journée GERLI de lipidomique, Toulouse, 11 Avril 2005 [I]
16. Barnouin R, Lasserre F, Cantiello M, Guillou H, Pineau T, Martin PGP. A kinetic view of coordinate modulations of gene expression and metabolism in wild-type and PPAR $\alpha$ -/- mice during fasting. Gene signature symposia, Paris, France, April 5<sup>th</sup> 2005 [S]
17. Martin PGP, Guillou H, Lasserre F, Déjean S, Legrand P, Besse P, Pineau T. Caractérisation des effets hépatiques induits par les lipides alimentaires par analyse de leurs signatures biologiques chez la souris. Journée d’animation scientifique de l’IFR30, Toulouse, 14 Décembre 2004 [S]
18. Bietrix F, Yan D, Nauze M, Rolland C, Schaak S, Martin PGP, Pineau T, Barbaras R, Perret B, Tercet F, Collet X. Etude de l’absorption intestinale de lipides dans un modèle murin surexprimant au niveau intestinal le Scavenger Receptor Class B type I : SR-BI. Journée d’animation scientifique de l’IFR30, Toulouse, 14 Décembre 2004 [S]
19. Martin PGP, Lasserre F, Guillou H, Barnouin R, Dejean S, Gotardi G, Legrand P, Baccini A, Besse P, Paris A. & Pineau T. Is PPAR $\alpha$  a true nutritional sensor for dietary fatty acids: a

coordinate transcriptomic and lipidomic study. Journée d'animation scientifique de l'IFR31, Toulouse, 19 Novembre 2004 [I]

20. Martin PGP, Lasserre F, Guillou H, Barnouin R, Déjean S, Gotardi G, Legrand P, Baccini A, Besse P, Paris A, Pineau T. Biological effects of nutrients : what interdisciplinary strategies to grasp and analyze complex signatures. Journée d'animation du Département d'Alimentation Humaine de l'INRA, Paris, France, 15 Juin 2004 [I]
21. Martin PGP, Guillou H, Lasserre F, Paris A, Déjean S, Legrand P, Besse P, Pineau T. Characterization of hepatic effects induced by food lipids from the analysis of their biological signatures in mice. Journée d'animation du Département d'Alimentation Humaine de l'INRA, Paris, France, 15 Juin 2004 [I]
22. Barnouin R, Lasserre F, Cantiello M, Guillou H, Pineau T, Martin PGP. A kinetic view of coordinate modulations of gene expression and metabolism in wild type and PPAR $\alpha$ -/- mice during fasting. Conference of the Paul Hamel Institute / New developements in the metabolic syndrome, Monaco, June 7-8 2004 [S]
23. Martin PGP, Lasserre F, Guillou H, Barnouin R, Dejean S, Gotardi G, Legrand P, Baccini A, Besse P, Paris A. & Pineau T. Comment s'approprier les technologies de la biologie à grande échelle (transcriptomique, métabolomique, ...)? Quels enjeux ? Quels défis à relever ? Journées d'animation scientifique du groupe de recherche thématique « Enzymes du Métabolisme des Xénobiotiques » de la SFBBM (Société Française de Biochimie et de Biologie Moléculaire), Dijon, France, 27-28 Mai 2004 [I]
24. Martin P, Ouamrane L, Roulet A, Lasserre F, Calléja C, Van Es A, Larrieu G, Dumas M-E, Gottardi G, Paris A, Gautier B, Pineau T. Multiorgan gene expression profiling in murine models: new prospects in assessing benefits and risks. 8<sup>th</sup> European ISSX (International Society for the Study of Xenobiotics) Meeting, Dijon, France, April 27-May 1 2003 [I]
25. Martin P, Ouamrane L, Lasserre F, Calléja C, Van Es A, Roulet A, Gauthier B, Pineau T. Differential effects of fenofibrate and rexinoids on an array of class II nuclear receptor-target genes. Second International Symposium on PPARs : From Basic Science to Clinical Applications, Florence, Italy, 19-22 March 2003 [S]
26. Guillou H, Martin P, Lan A, Jan S, Rioux V, Ouamrane L, Pineau T, Legrand P. Impact des acides gras alimentaires sur le développement de la stéatose hépatique chez la souris déficiente pour l'isoforme a du récepteur activé par les proliférateurs de peroxysomes (PPAR $\alpha$ ). 4èmes Journées francophones de nutrition, Dijon, France, 27-29 novembre 2002 [S]. (*Abstract publié dans Nutrition Clinique et Métabolisme, Nov 2002 Vol 16- Suppl.1, p. 1-52.*)
27. Chevillard G, Clémencet M-C, Etienne P, Martin P, Pineau T, Latruffe N, Nicolas-Francès V. Tissue-specific expression of two peroxisomal 3-ketoacyl-CoA Thiolase genes in wild-type and PPAR $\alpha$ -null mice and induction by fenofibrate. International Symposium on Peroxisomal

Disorders and Regulation of Genes, Ghent, Belgium, September 25-28 2002 [S]. (*Proceedings publiés dans "Peroxisomal Disorders and Regulation of Genes"*, Kluwer academic/Plenum Publishers 2003).

28. Bugaut M, Fourcade S, Gondcaille C, Gueugnon F, Depreter M, Roels F, Netik A, Berger J, Martin P, Pineau T, CadepondF, El Etr M, Savary S. Pharmacological induction of redundant genes for a therapy of X-ALD: phenylbutyrate and other compounds. International Symposium on Peroxisomal Disorders and Regulation of Genes, Ghent, Belgium, September 25-28 2002 [S]. (*Proceedings publiés dans "Peroxisomal Disorders and Regulation of Genes"*, Kluwer academic/Plenum Publishers 2003)
29. Martin P, Guillou H, Lasserre F, Calléja C, Van Es A, Roulet A, Ouamrane L, Jan S, Lan A, Legrand P, Pineau T. Modulation de la voie de signalisation du récepteur PPAR $\alpha$  par la fraction lipidique alimentaire. Profils d'expression de gènes-cibles de récepteurs nucléaires de classe II. 29ème Forum des Jeunes Chercheurs, Dijon, France, 1-5 juillet 2002 [S]
30. Martin P, Ouamrane L, Guillou H, Roulet A, Lasserre F, Dumas M, Calléja C, Pineau T. Etudes génétique, expressionnelle et fonctionnelle du récepteur nucléaire activé par les proliférateurs de peroxyosomes (PPAR $\alpha$ ) chez la souris. 29ème Forum des Jeunes Chercheurs, Dijon, France, 1-5 juillet 2002 [I]
31. Martin P, Guillou H, Lasserre F, Calléja C, Van Es A, Roulet A, Ouamrane L, Jan S, Legrand P, Pineau T. The PPAR $\alpha$  signaling pathway and lipid nutrients: Expression profiling of class II nuclear receptor target genes. 93<sup>rd</sup> American Oil Chemists Society Annual Meeting & Expo, Montréal, Québec/Canada, May 5-8 2002 [I].
32. Guillou H, Martin P, Lasserre F, Roulet A, Jan S, Legrand P, Roche H, Pineau T. Nutrient regulation of gene expression : a case for dietary fatty acids and the PPARs. Joint Meeting of the Nutrition Society, the Association Française de Nutrition and the Société de Nutrition et de Diététique de Langue Française, London, England, Feb 13-14 2002 [I].

## Communications par poster

Le nom de la (ou des) personne(s) ayant réalisé(es) et présenté(es) le poster est souligné

1. González I, Déjean S, Martin P, Gonçalvez O, Baccini A, Besse P. Mise en relation du niveau d'expression d'ABC transporteurs humains avec l'efficacité de molécules candidates dans des lignées cellulaires cancéreuses. 3ème journée du Cancéropôle Grand Sud Ouest, Bordeaux, France, 2007

2. Radulescu O, Siegel A, Le Borgne M, Veber P, Pécou E, Lagarrigue S, Martin P. Qualitative, integrated modeling of lipid metabolism in hepatocytes. 7th International Conference on Systems Biology - Yokohama, Japan, 9-13 October 2006
3. Cantiello M, Giantin M, Nebbia C, Bollo E, Lasserre F, Martin P, Pineau T, Dacasto M, Sequencing and cloning of some relevant drug metabolism genes in cattle: the “Programma Galileo” 2003 summary report. Galileo Project – 2005 Meeting, Siena, Italy, 12-13 September 2005
4. Barnouin R, Lasserre F, Cantiello M, Guillou H, Pineau T, Martin PGP. A kinetic view of coordinate modulations of gene expression and metabolism in wild-type and PPAR $\alpha$ -/- mice during fasting. 3<sup>rd</sup> International Symposium on PPARs: Efficacy and Safety, Monte Carlo, Monaco, March 19-23 2005
5. Martin P, Guillou H, Lasserre F, Calléja C, Jan S, Paris A, Legrand P, Pineau T. Effects of dietary fatty acids on the expression of an array of class II nuclear receptor target genes. 8th European ISSX (International Society for the Study of Xenobiotics) Meeting, Dijon, France, 27 avril-1<sup>er</sup> mai 2003.
6. Guillou H, Martin P, Jan S, d'Andrea S, Roulet A, Catheline D, Rioux V, Pineau T, Legrand P. Regulation of hepatic desaturases by PPAR $\alpha$ . 8th European ISSX Meeting, Dijon, France, April 27-May 1 2003
7. Guillou H, Martin P, Lan A, Jan S, Rioux V, Ouamrane L, Pineau T, Legrand P. Dietary fatty acids influence the hepatic steatosis in PPARalpha-deficient mice. Second International Symposium on PPARs : From Basic Science to Clinical Applications, Florence, Italy, March 19-22 2003
8. Martin P, Guillou H, Lasserre F, Calléja C, Dumas M, Paris A, Legrand P, Pineau T. Modulations de la voie de signalisation du récepteur PPAR $\alpha$  par la fraction lipidique alimentaire. Profils d'expression de gènes cibles des récepteurs nucléaires de classe II. 4èmes Journées francophones de nutrition, Dijon, France, 27-29 novembre 2002. (*Abstract publié dans Nutrition Clinique et Métabolisme, Nov 2002 Vol 16- Suppl.1, p. 1-52.*)
9. Fourcade S, Gondcaille C, Savary S, Albet S, Martin P, Pineau T, Gueugnon F, Bellenger J, Bugaut M. The ABCD2 gene is induced by phenylbutyrate independently of PPAR $\alpha$ . International Symposium on Peroxisomal Disorders and Regulation of Genes, Ghent, Belgium, September 25-28 2002.
10. Guillou H, Martin P, Jan S, d'Andréa S, Roulet A, Catheline D, Rioux V, Pineau T, Legrand P. Fenofibrate up-regulates hepatic  $\Delta$ 6-desaturase in mice through a PPAR $\alpha$ -dependent pathway. 6<sup>ème</sup> Congrès Annuel de la Société Française de Pharmacologie, Rennes, France, 8-10 avril 2002

11. Chevillard G, Clémencet M-C, Etienne P, Martin P, Pineau T, Latruffe N, Nicolas-Francès V  
Effet d'un proliférateur de peroxysomes, le fénofibrate, sur le niveau d'expression du gène de la 3-oxoacyl-CoA thiolase peroxysomiale chez la souris sauvage et chez la souris invalidée pour le gène PPAR $\alpha$ . 29ème Forum des Jeunes Chercheurs, Dijon, France, 1-5 juillet 2002.
12. Fourcade S, Gondcaille C, Savary S, Albet S, Martin P, Pineau T, Bellenger J, Bugaut M. Le phénylbutyrate (PBA) induit le gène ABCD2 (ou ALDR –Adrenoleukodystrophy-related-) indépendamment de PPAR $\alpha$ . 29ème Forum des Jeunes Chercheurs, Dijon, France, 1-5 juillet 2002.

## **Formations (intervenant) :**

- Formation des personnels de l'UR66 et de l'UMR1089 de l'INRA à la PCR en temps réel. P. Martin, O. Puel, F. Guerraud. Toulouse, 09 septembre 2004.
- Atelier « Analyse des données issues des technologies à haut-débit », M2R Innovation pharmacologique. P. Martin, S. Déjean. 04 février 2005, 16 décembre 2005, 15 décembre 2006, 14 décembre 2007
- Cours « Analyse statistique des expériences sur le transcriptome ». Cours de 2h aux étudiants du M2P Bioingénierie. P. Martin. 08 novembre 2005, 14 novembre 2006, 07 novembre 2007
- Cours « Microarrays / biochips technology: a “new” molecular biology tool to assess the control of gene expression ». Cours de 2h aux étudiants de dernière année et aux doctorants de l'Université Vétérinaire. P. Martin. Padoue (Italie), novembre 2004.
- Cours « Studying Nuclear Receptors Signaling Pathways : advanced technologies to measure gene expression ». Cours de 3h aux étudiants de dernière année et aux doctorants de l'Université Vétérinaire. P. Martin. Padoue (Italie), 21 décembre 2006.
- Cours « Advanced methodologies to assess gene expression ». Cours de 3h aux étudiants de dernière année et aux doctorants de l'Université Vétérinaire. P. Martin. Turin (Italie), 08 décembre 2005 puis novembre 2006.
- Module au choix « Analyse statistique des données biologiques à haut-débit » de l'ENSA Rennes. P. Martin, D. Causeur, S. Lagarrigue. Rennes, 13-23 juin 2006
- Formation permanente INRA « Analyse statistique des données X-omiques ». P. Martin, D. Causeur, S. Lagarrigue. Rennes, 19-23 février 2007, 21-25 janvier 2008
- Cours européen « Genomics on obesity ». Intervention de 2h « Statistics and bioinformatics applied to -omics technologies ». P. Martin. Toulouse, 09 juin 200

## Introduction

Ce travail de doctorat s'inscrit dans l'objectif global du laboratoire de pharmacologie et toxicologie et du pôle de toxicologie alimentaire de l'INRA de Toulouse qui consiste à évaluer **l'impact de xénobiotiques sur l'organisme** pour éclairer en particulier les évaluations qui guident souvent les gestionnaires du risque et l'autorité administrative en matière de toxicologie et de pharmacologie. Le laboratoire s'intéresse depuis de nombreuses années aux enzymes de biotransformation impliquées dans le métabolisme des substances exogènes et notamment des médicaments et de toxiques divers. La création de l'équipe de pharmacologie moléculaire a permis de développer un axe de recherche dédié aux voies de signalisation impliquées dans la régulation de l'expression des enzymes du métabolisme des xénobiotiques. Cette équipe a donc initié des travaux sur les **récepteurs nucléaires** régulant la transcription de certains de ces gènes mais également d'autres gènes jouant des rôles importants dans de nombreux processus biologiques : le développement, le métabolisme énergétique, la prolifération et la différentiation cellulaire, les balances ioniques, l'inflammation, la reproduction, etc. Dans le cadre de ce doctorat, nous avons cherché à caractériser le potentiel de régulation génique exercé par l'isoforme  $\alpha$  du récepteur activé par les proliférateurs de peroxysomes (**PPAR $\alpha$**  pour peroxisome proliferator-activated receptor). Pour cela, deux approches ont été combinées : 1) le développement d'une puce à ADN dédiée à l'étude du récepteur PPAR $\alpha$ , dénommée INRArray 01 et destinée à assurer une évaluation assez large des modulations d'expression provoquées par ce récepteur et 2) l'étude à l'aide de cette puce à ADN et d'outils complémentaires, des modulations d'expression génique provoquées par le récepteur PPAR $\alpha$  activé par des signaux très différents, de type pharmacologique, nutritionnel et physiopathologique.

Ce travail a été conduit dans un contexte particulier : celui d'un doctorat réalisé en situation professionnelle puisqu'en qualité d'ingénieur agronome j'ai été recruté à l'INRA comme ingénieur de recherche (IR2) en 2004, dans l'équipe où ce doctorat était initié. Dès lors, ce débouché professionnel rendant moins prégnante la conclusion rapide du doctorat, j'ai pu satisfaire en parallèle des contraintes de service en développement de méthodes et d'outils adjacents au sujet initial de doctorat (certains figurant dans ce mémoire), et une expérimentation plus détaillée et plus diversifiée que nous ne l'avions initialement prévu (enrichissement progressif en études et valorisations). Il en résulte une durée atypique de doctorat, mais également une densité accrue de valorisations sous formes d'articles, de

conférences, de communications et de participations à des enseignements. Au moment venu de la synthèse de ces travaux et de la rédaction d'un manuscrit, j'ai souhaité de plus consacrer tout le temps et l'attention nécessaires à réaliser un bilan relativement complet de mes connaissances de notre domaine de recherche et des technologies que nous utilisons. J'ai volontairement écrit les trois premiers chapitres de cette thèse dans l'objectif d'un usage comme document de référence initial pour les stagiaires et les nouveaux arrivants dans notre équipe (un collègue CR2 arrive en 2008). Etant de plus en plus impliqué dans diverses formations, j'ai également souhaité que ce manuscrit puisse me servir de base pour la rédaction de documents à visée pédagogique ou de transfert. Cela m'a naturellement conduit à rédiger un document dense, comportant parfois des informations contextuelles qui ne sont pas déterminantes dans le cadre de l'expérimentation conduite mais qui revêt une valeur intrinsèque de synthèse et dont la lecture, agrémentée de nombreuses figures originales, peut être réalisée presque indépendamment chapitre par chapitre.

Dans ce manuscrit, je présente d'abord une synthèse de la bibliographie sur les récepteurs nucléaires (**Chapitre I**) puis me concentre sur le principal récepteur étudié dans le cadre de ce doctorat : PPAR $\alpha$  (**Chapitre II**). Je dresse ensuite un inventaire des méthodes d'étude des régulations géniques à l'échelle d'un gène et à l'échelle du transcriptome entier et justifie les choix technologiques réalisés dans le cadre de ce doctorat (**Chapitre III**). Dans ce troisième chapitre, je présente également un éventail de méthodes d'analyse des données issues de la biologie à haut-débit et situe le contexte de mes travaux dans ce domaine. Enfin, je présente successivement les trois principales études réalisées à l'aide des outils de transcriptomique développés dans le cadre de ce doctorat. Elles ont permis de préciser la nature des impacts transcriptionnels de trois situations d'activations de PPAR $\alpha$  spécialement choisies pour explorer des signalisations de trois natures :

1. Pharmacologique avec une étude sur trois tissus des impacts d'activateurs de PPAR $\alpha$  et de RXR utilisés en thérapeutique humaine (**Chapitre IV**)
2. Nutritionnelle avec une étude de l'activation du récepteur PPAR $\alpha$  par des acides gras polyinsaturés  $\omega 6$  et  $\omega 3$  alimentaires (**Chapitre V**)
3. Physiopathologique avec une étude du jeûne réalisée en cinétique (**Chapitre VI**).

# Chapitre I : Les récepteurs nucléaires

## I.1 NOMENCLATURE

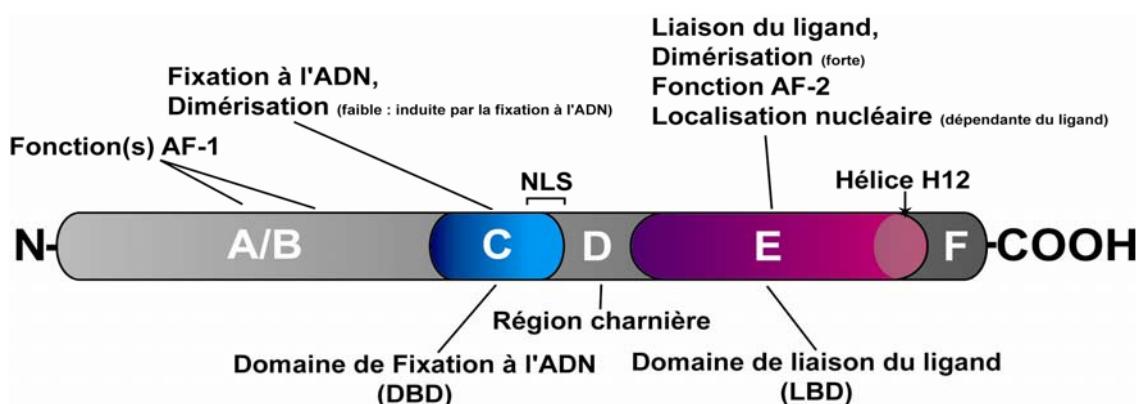
Les récepteurs nucléaires (RNs) sont des facteurs de transcription dont l'activité sur les promoteurs de leurs gènes cibles est dépendante de la fixation de ligands. Les ligands connus de récepteurs nucléaires sont en règle générale des molécules de petite taille et de nature lipophile. Un certain nombre de récepteurs nucléaires n'a pas de ligand connu. Ils sont alors qualifiés par certains auteurs de récepteurs orphelins (*orphan nuclear receptors*). J'ai fait le choix d'éviter cette nomenclature dans le cadre de ce mémoire car elle est évidemment dépendante de l'acquisition de nouvelles connaissances. Ainsi, certains auteurs qualifient toujours "d'orphelins" des récepteurs nucléaires pour lesquels des ligands, y compris endogènes, ont été identifiés (cas fréquent pour les PPAR). D'autres auteurs introduisent dans ce cas la notion de "récepteurs orphelins adoptés". Notons que des données structurales suggèrent que certains récepteurs nucléaires pourraient effectivement ne pas autoriser la fixation de ligands en raison d'une poche de fixation du ligand absente ou de taille très réduite (voir par exemple Greschik et al., 2002; Kallen et al., 2004b; Wang et al., 2003). Cependant, des réarrangements conformationnels, parfois importants, autorisent la fixation de ligands synthétiques sur certains de ces récepteurs (Greschik et al., 2004; Wang et al., 2006a) et la recherche de ligands de taille réduite se poursuit.

C'est en se basant sur l'alignement et l'évolution des séquences nucléotidiques et protéiques de certains domaines fonctionnels (domaine de fixation à l'ADN : DBD et de fixation des ligands : LBD) des récepteurs nucléaires de nombreuses espèces qu'une nomenclature des récepteurs nucléaires a été mise en place en 1999 (Committee, 1999). L'alignement des séquences des LBD et DBD des récepteurs de diverses espèces a conduit à la construction d'un dendrogramme ou arbre de classification. Les auteurs définissent familles et groupes en identifiant les branches les plus internes qui s'avèrent robustes au sens d'une méthode de ré-échantillonnage (bootstrap, seuil fixé à 90%). Ce système identifie 6 sous-familles dans la super-famille des récepteurs nucléaires, numérotées de 1 à 6 de manière arbitraire (c'est-à-dire sans lien évolutif hiérarchique). La sous-famille 0, additionnelle, regroupe les récepteurs nucléaires "atypiques" qui ne contiennent pas l'un des deux domaines DBD ou LBD. Des groupes sont ensuite définis à l'intérieur des sous-familles par une nouvelle identification des branches les plus internes présentant au moins 90% de robustesse.

Sous-famille	Groupe	Nom officiel	Nom alternatif	Nom complet (anglais)		Nom complet (français)		Ref GenBank	GenID	Ref protéine (Genbank)
1	A	NTR1A1	TTRB, THRB, T3Rβ[1], Thrb1, Thrb2, Rvr, c-erbA-α	Thyroid Hormone Receptor alpha		Recepteur à l'hormone thyroïdienne alpha		NM_178060	21833	NP_835161
1	A	NTR1A2	TTRB, THRB, T3Rβ[1], Thrb1, Thrb2, c-erbA-β	Thyroid Hormone Receptor beta		Recepteur à l'hormone thyroïdienne beta		NM_009380	21834	NP_033406
1	B	NTR1B1	RAR <sub>α</sub> , RAR, RAR <sub>α</sub> beta, Hap	Retinoid acid receptor alpha		Recepteur à la facioïde rétinotique alpha		NM_005024	19401	NP_033050
1	B	NTR1B2	RAR <sub>β</sub> , RAR <sub>β</sub> beta, Hap	Retinoid acid receptor beta		Recepteur à la facioïde rétinotique beta		NM_011243	19472	NP_035373
1	B	NTR1B3	RAR <sub>γ</sub> , RAR <sub>γ</sub> beta, RAR <sub>γ</sub> gamma	Retinoid acid receptor gamma		Recepteur à la facioïde rétinotique gamma		NM_011244	19411	NP_035374
1	C	NTR1C1	PPAR <sub>α</sub> , PPAR <sub>α</sub> beta, PPAR <sub>α</sub> gamma	Peroxisome proliferator-activated receptor alpha		Recepteur activé par les proliférateurs de peroxysomes, isoforme α		NM_011144	19013	NP_035274
1	C	NTR1C2	PPAR <sub>β</sub> , PPAR <sub>β</sub> delta, PPAR <sub>β</sub> gamma	Peroxisome proliferator-activated receptor beta (ou delta)		Recepteur activé par les proliférateurs de peroxysomes, isoforme β/δ		NM_011145	19015	NP_035275
1	C	NTR1C3	PPAR <sub>γ</sub> , PPAR <sub>γ</sub> alpha, PPAR <sub>γ</sub> gamma, PPAR-gamma2	Peroxisome proliferator-activated receptor gamma		Recepteur activé par les proliférateurs de peroxysomes, isoforme γ		NM_011146	19016	NP_035276
1	D	NTR1D1	Rev-erbα, Rev-erbbeta, Rev-erbalpha, EAR1, EAR1A	Rev-erbα alpha		Rev-erbα alpha		NM_145442	217166	NP_683409
1	D	NTR1D2	Rev-erb <sub>β</sub> , Rev-erbbeta, Rev-erb <sub>α</sub> , Rev-erb <sub>α</sub> beta, EAR1b, BD73, HZ72	Rev-erbα alpha-related receptor		Recepteur relié à Rev-erbα alpha		NM_011584	353187	NP_035714
1	F	NTR1F1	ROR <sub>α</sub> , ROR <sub>α</sub> beta, ROR <sub>α</sub> gamma	RAR-related orphan receptor alpha		Recepteur orphelin relié à RAR, isoforme alpha		NM_036364	19883	NP_038867
1	F	NTR1F2	ROR <sub>β</sub> , ROR <sub>β</sub> beta, ROR <sub>β</sub> gamma	RAR-related orphan receptor beta		Recepteur orphelin relié à RAR, isoforme beta		NM_146095	217998	NP_666207
1	F	NTR1F3	ROR <sub>γ</sub> , ROR <sub>γ</sub> beta, ROR <sub>γ</sub> gamma	RAR-related orphan receptor gamma		Recepteur orphelin relié à RAR, isoforme gamma		NM_011281	19885	NP_035411
1	H	NTR1H2	LXR <sub>β</sub> , LXReba, LXR <sub>α</sub> , LXR <sub>α</sub> beta, NER1, OR-1, Uni2, RIP15	Liver X receptor beta		Recepteur X aux oxystéroïds, isoforme beta		NM_008473	22260	NP_033499
1	H	NTR1H3	LXR <sub>α</sub> , LXReba, LXReba, LXReba, LXR <sub>α</sub> beta, Uni1	Liver X receptor alpha		Recepteur X aux oxystéroïds, isoforme alpha		NM_013839	22259	NP_038867
1	H	NTR1H4	FXR <sub>α</sub> , FXR <sub>β</sub> , FXR <sub>γ</sub> , HNR1, RIP14, RXRIP14	Farnesoid X receptor alpha		Recepteur X aux farnésoides, isoforme alpha		NM_009108	20186	NP_033134
1	H	NTR1H5	FXR <sub>β</sub> , FXR <sub>β</sub>	Farnesoid X receptor beta		Recepteur X aux farnésoides, isoforme beta		NM_199658	361463	NP_941060
1	I	NTR1I1	VDR	Vitamin D receptor		Recepteur à la vitamine D		NM_008504	22337	NP_033530
1	I	NTR1I2	PXR, SKR, BXR, ONR1, PXR1, PXR1, PXR <sub>2</sub>	Pregnane X receptor		Recepteur X aux pregnanées		NM_010936	18171	NP_033499
1	I	NTR1I3	CAR, CAR <sub>β</sub> , CAR <sub>β</sub> , CAR <sub>α</sub> , CAR1, CAR2, MB67, Cere2	Constitutive androstane receptor		Recepteur constitutif aux androstanes		NM_008083	12255	NP_033333
2	A	NTR2A1	HNF4α, LNF4, LNF4-alpha, LNF4, HNF4, MODY1, TCF14	Hepatic nuclear factor 4 alpha		Facteur nucléaire hépatique 4, isoforme alpha		NM_008261	15378	NP_032287
2	A	NTR2A2	HNF4γ, HNF4α, HNF4gamma	Hepatic nuclear factor 4 gamma		Facteur nucléaire hépatique 4, isoforme gamma		NM_013920	30942	NP_038448
2	B	NTR2B1	RXR <sub>α</sub> , RXR <sub>α</sub> beta, RXR <sub>α</sub> gamma	Retinoid X receptor alpha		Recepteur X aux rétinoides, isoforme alpha		NM_011303	20182	NP_035435
2	B	NTR2B2	RXR <sub>β</sub> , RXR <sub>β</sub> beta, RXR <sub>β</sub> gamma	Retinoid X receptor beta		Recepteur X aux rétinoides, isoforme beta		NM_008107	20183	NP_033133
2	B	NTR2B3	RXR <sub>γ</sub> , RXR <sub>γ</sub> beta, RXR <sub>γ</sub> gamma	Retinoid X receptor gamma		Recepteur X aux rétinoides, isoforme gamma		NM_011620	21907	NP_035759
2	C	NTR2C1	TR2, EENR, TR2-11	Testicular receptor 2		Recepteur testiculaire 2		NM_152229	21907	NP_038415
2	C	NTR2C2	TR4, TAK-1		Homologue de tailless		NM_011630	22026	NP_035760	
2	E	NTR2E1	TLL, TLX, FRC, MTLL, MTLL, fierce					NM_152229	21907	NP_038415
2	E	NTR2E3	PNR, RNR, rd7					NM_013708	23958	NP_038736
2	F	NTR2F1	COUP-TF1, COUP-TF1, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2	Photoreceptor-specific nuclear receptor		Recepteur nucléaire spécifique des cellules photoréceptrices		NM_010151	13815	NP_034281
2	F	NTR2F2	COUP-TF1, COUP-TF1, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2, COUP-TF2	chicken ovalbumin upstream promoter-transcription factor 2		Facteur de transcription 1 du promoteur amont de l'ovalbumine de poulet		NM_009637	11819	NP_033827
2	F	NTR2F6	EAR2, EAR-2, Et1β2, COUP-TF3	Et1β-related gene-2		Gène 2 relié à Et1β		NM_010150	13864	NP_034280
3	A	NTR3A1	ER <sub>α</sub> , ER <sub>α</sub> beta, ESR, ESR <sub>1</sub> , ESTR, ESTR <sub>α</sub> , ER <sub>β</sub>	Estrogen receptor alpha		Recepteur aux oestrogénies, isoforme alpha		NM_007956	13982	NP_031982
3	A	NTR3A2	ER <sub>β</sub> , Er, Er <sub>β</sub> beta, ESR <sub>2</sub> , ESTR <sub>β</sub> , ER <sub>β</sub>	Estrogen receptor beta		Recepteur aux oestrogénies, isoforme beta		NM_007953	13983	NP_987590
3	B	NTR3B1	ER <sub>β</sub> alpha, ER <sub>β</sub> beta, ER <sub>β</sub> beta, ESTR <sub>β</sub>	Estrogen-related receptor alpha		Recepteur relié aux oestrogénies, isoforme alpha		NM_007959	26379	NP_031979
3	B	NTR3B2	ER <sub>β</sub> beta, ER <sub>β</sub> beta, ER <sub>β</sub> beta, ESTR <sub>β</sub>	Estrogen-related receptor beta		Recepteur relié aux oestrogénies, isoforme beta		NM_011934	26380	NP_036064
3	B	NTR3B3	ER <sub>β</sub> gamma, ER <sub>β</sub> gamma, ER <sub>β</sub> gamma	Estrogen-related receptor gamma		Recepteur relié aux oestrogénies, isoforme gamma		NM_011935	26381	NP_036065
3	C	NTR3C1	GR, GR <sub>1</sub> , GR <sub>1</sub> -L <sub>1</sub>	Glucocorticoid receptor		Recepteur aux glucocorticoïdes		NM_008173	14819	NP_036065
3	C	NTR3C2	IMR, IMCR, MLR	Mineralocorticoid receptor		Recepteur aux minéralocorticoïdiques		XM_366093	110784	XP_366093
3	C	NTR3C3	PR, PRA, PR <sub>B</sub> , PGR	Progesterone receptor		Recepteur à la progéstérone		NM_008829	18667	NP_032355
3	C	NTR3C4	AR, TMF	Androgen receptor		Recepteur aux androgènes		NM_013476	11835	NP_038504
4	A	NTR4A1	NGF <sub>β</sub> , NGF <sub>β</sub> , Nutt7, Nutr, Ntr, TR3, Grp, Hb1, NP10, TIS1, GFRP <sub>1</sub> , Htr-1	Nerve growth factor induced protein I-B		Protéine I-B induite par facteur de croissance des nerfs		NM_010444	15370	NP_034574
4	A	NTR4A2	Nutr, NOT, HZF-3, RNR-1, TINOR, TINUR	Nutr-related protein 1		Protéine I-2 reliée à Nur		NM_013613	18227	NP_038641
4	A	NTR4A3	Nor1, NOR-1, CHN, TEC, CSMF, MINOR	Nor1, NOR-1	Recepteur orphelin neuronal 1		NM_015743	18124	NP_056558	
5	A	NTR5A1	SF1, SF1, ADDBP, ELP-3, Flz1, Flz2-F1	Steriologenic factor 1		Facteur stéroïdogénique 1		XM_620639	26423	NP_030674
5	A	NTR5A2	LRIH1, LRIH1, FTF, AFE17A, FFLR, PHR	Liver receptor homolog 1		Homologue 1 du récepteur thérapeutique		NM_030672	109861	NP_032355
6	A	NTR6A1	GONF, RTR, NCNF	Germ cell nuclear factor		Facteur nucléaire des cellules germinales		NM_010264	14536	NP_034394
6	B	NTRB1	DAX-1, AHD, AHCH	Dax-1, Adrenohypophysis congenita		Gène 1 de la région DSS-AHC du chromosome X		NM_007430	11614	NP_031456
0	B	NTRB2	SHP-1, SHP-1, SHP-1	Small heterodimer partner		Petit partenaire dihétérodimérisation		NM_011850	23957	NP_055380

Table 1 : liste des 49 récepteurs nucléaires muriens

Le nom d'un récepteur nucléaire est, dans ce système, NR<sub>x</sub>y<sub>z</sub>, où x est le numéro de la sous-famille, y la lettre (en capitale) correspondant au groupe et z le chiffre permettant d'identifier le récepteur dans ce groupe. NR1A1 est ainsi l'isoforme  $\alpha$  du récepteur à l'hormone thyroïdienne (TR $\alpha$ ) qui est le premier membre du groupe A de la sous-famille 1. Cette nomenclature sera peu utilisée dans ce mémoire car, si elle a le mérite de fournir des noms officiels pour tous les membres de la superfamille et de permettre d'adoindre de nouveaux membres, peu de gens l'utilisent de manière systématique, sans doute parce que les noms des récepteurs sont alors peu évocateurs de leurs fonctions, de leurs ligands, du cadre de leur découverte ou d'autres informations généralement associées aux noms initialement utilisés. La [Table 1](#) présente les 49 récepteurs nucléaires connus chez la souris. Elle fournit le nom officiel et une liste de noms alternatifs qui sont retrouvés dans la littérature ou les bases de données et permettra au lecteur de toujours retrouver le lien avec la nomenclature officielle. A la même date, on connaît 48 récepteurs nucléaires chez l'homme (Duarte et al., 2002; Maglich et al., 2001; Ruau et al., 2004). Le récepteur NR1H5 (FXR $\beta$ ) n'a pas été identifié chez l'homme et semble être spécifique de la souris. Le rat disposerait de 47 récepteurs nucléaires (Bertrand et al., 2004) mais il est possible que l'avancement du séquençage de son génome en révèle d'autres. La drosophile (*Drosophila melanogaster*) dispose quant à elle de 21 récepteurs nucléaires (Adams et al., 2000; Bertrand et al., 2004) dont le très étudié récepteur à l'ecdysone. Enfin, le nématode modèle (*Caenorhabditis elegans*) possède plus de 270 récepteurs nucléaires (Maglich et al., 2001; Sluder et al., 1999). L'origine de ce nombre pléthorique viendrait de multiples événements de duplication.



[Figure 1](#) : Organisation des récepteurs nucléaires en six domaines fonctionnels indépendants

Les légendes du bas indiquent les noms de certaines régions (DBD : DNA-Binding Domain, LBD : Ligand-Binding Domain). Les légendes du haut fournissent des informations sur la fonction de certaines zones (AF : Activation Function, NLS : Nuclear Localization Signal)

## I.2. STRUCTURE CANONIQUE DES RECEPTEURS NUCLEAIRES

Les récepteurs nucléaires présentent une structure commune. Ils sont organisés en domaines fonctionnels indépendants. La plupart des récepteurs possèdent 6 domaines nommés domaine A à domaine F ([Figure 1](#)). Les domaines A et B sont en général étudiés ensemble et leur borne commune n'est pas toujours clairement définie.

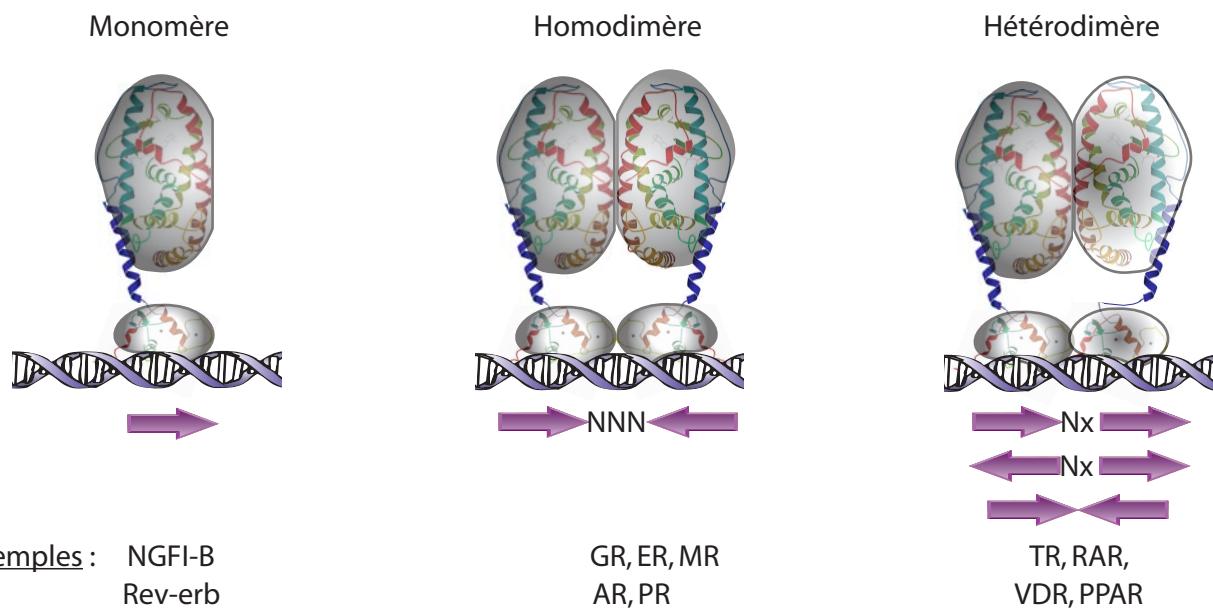
### I.2.1. Les domaines A et B

La région N-terminale (**domaines A/B**) est la zone la plus variable à la fois en taille et en séquence parmi les récepteurs nucléaires. Ainsi, sa taille s'étend de 23 (VDR) jusqu'à plus de 600 acides aminés (MR). De plus, plusieurs récepteurs présentent des isoformes qui diffèrent uniquement par leur domaine A/B du fait d'épissages alternatifs et/ou de l'utilisation de promoteurs alternatifs (Chen et al., 2005; Wilson and McPhaul, 1994; Zelent et al., 1991). Cette région présente au moins une zone de transactivation (activation de la transcription) constitutivement active dénommée AF-1 (Activation Function 1). La fonction AF-1 est capable d'activer la transcription de manière autonome et indépendante de la présence de ligand dans des cellules de mammifère (Bocquel et al., 1989) ou de levure (Metzger et al., 1995) d'une manière dépendante du récepteur (Bocquel et al., 1989), du type cellulaire (Tora et al., 1989) et de l'élément de réponse étudiés (Lees et al., 1989). La structure tridimensionnelle des domaines A/B n'est pas connue. Des études de résonance magnétique nucléaire (RMN) et de dichroïsme circulaire ont montré que la région AF-1 de GR (Dahlman-Wright et al., 1995) et les régions N-terminales des récepteurs aux estrogènes ER $\alpha$  et ER $\beta$  (Warnmark et al., 2001) n'étaient pas structurées en solution aqueuse. Des expériences de protéolyse partielle sur la région N-terminale de PR indiquent également une absence de structure bien déterminée (Bain et al., 2000). En revanche, en présence du domaine C de fixation à l'ADN (DBD), la région N-terminale présente une plus faible sensibilité à la protéolyse (Bain et al., 2000), suggérant qu'elle possède une structure secondaire lorsque le récepteur est intact. Toujours en présence du DBD, la partie N-terminale du récepteur GR présente les caractéristiques d'un enrichissement en hélices  $\alpha$  qui semble accrue et complétée par une structure tertiaire suite à la fixation du récepteur à l'ADN (Kumar et al., 1999). En présence de trifluoroéthanol, un agent stabilisant des structures secondaires, ou de TMAO (triméthylamine N-oxyde), un osmolyte (agent promoteur et stabilisant des structures secondaires), les régions AF-1 de plusieurs récepteurs montrent des caractéristiques d'hélices

$\alpha$  (pour une synthèse, voir Kumar and Thompson, 2003). La formation d'hélices  $\alpha$  potentielles dans le domaine AF-1 de ER $\alpha$  semble également influer sur la capacité de transactivation du récepteur (Metivier et al., 2000). De plus, des interactions avec d'autres facteurs de transcription comme TBP (TATA box binding protein) (Warnmark et al., 2001) ou TFIIF (transcription factor II F) (Reid et al., 2002) pourraient favoriser la formation de structures secondaires. En lien avec ces aspects structuraux, des travaux récents ont démontré que les régions AF-1 de plusieurs récepteurs nucléaires étaient capables d'interagir avec des cofacteurs (Benecke et al., 2000, et références incluses) comme SRC-1, TIF2, CBP ou Ada2. Comme nous le verrons plus loin, ces cofacteurs agissent comme "ponts" entre les récepteurs et la machinerie transcriptionnelle et/ou induisent des remodelages locaux de la chromatine influant sur la transcription. Certaines interactions AF1-cofacteur semblent régulées par des phosphorylations (Hammer et al., 1999; Tremblay et al., 1999), constituant un lien potentiel avec d'autres voies de signalisation. Enfin, plusieurs études suggèrent que la fonction AF-1 interagirait avec d'autres parties du récepteur, en particulier la partie C-terminale contenant la ou les fonctions AF-2, et fonctionnerait en synergie avec ces autres domaines, notamment *via* le recrutement coopératif de cofacteurs (Benecke et al., 2000; Bommer et al., 2002). L'ensemble de ces résultats tend à montrer que les domaines A/B sont intrinsèquement peu structurés mais que l'acquisition d'une structure secondaire fonctionnelle par la partie AF-1 joue un rôle dans l'activité du récepteur et pourrait être induite ou sélectionnée sous l'effet d'interactions avec d'autres domaines du récepteur, de la fixation d'autres protéines et/ou de la fixation du récepteur à l'ADN. La fonctionnalité des domaines A/B et plus particulièrement de la fonction AF-1 semble donc démontrée dans un certain nombre de cas précis mais aucun schéma général ne semble aujourd'hui établi. Enfin, on peut penser que l'acquisition de données cristallographiques sur des récepteurs entiers, si elle s'avère possible, permettrait de mieux comprendre le rôle et le fonctionnement de ces domaines A/B.

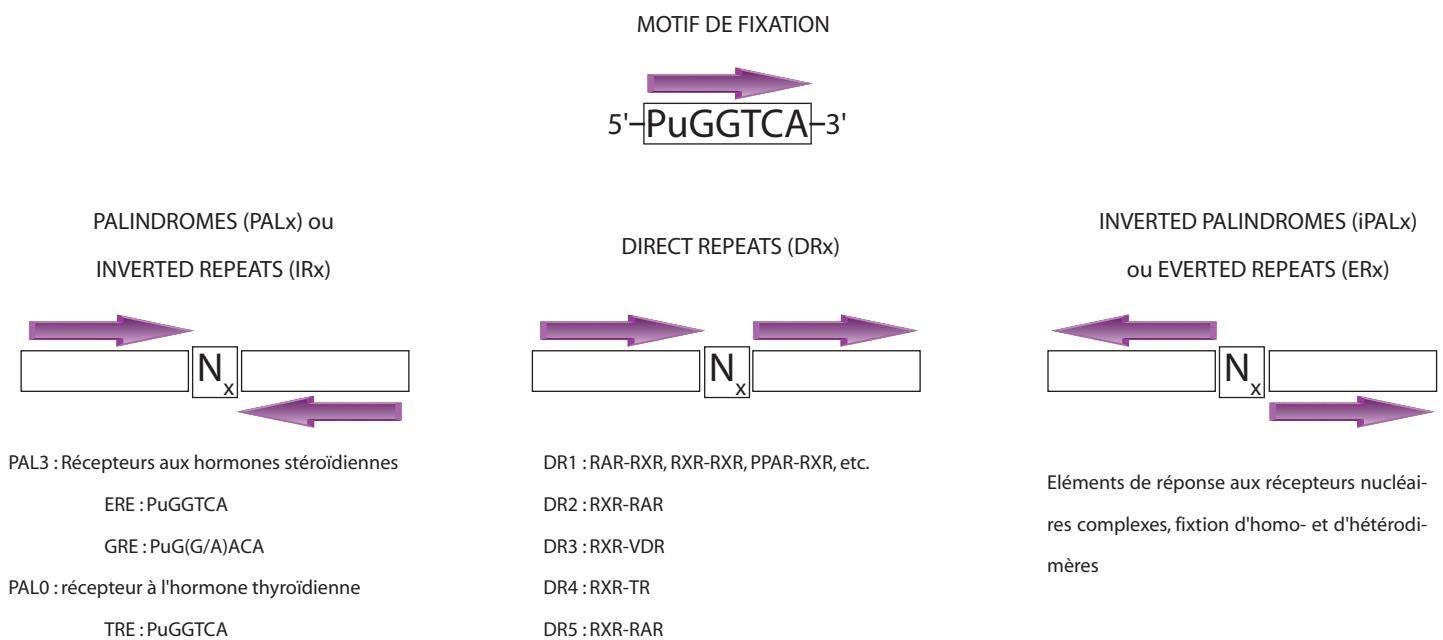
### I.2.2. Le domaine C de fixation à l'ADN

Le **domaine C** constitue le domaine de fixation à l'ADN (**DBD** pour DNA Binding Domain). Les récepteurs nucléaires se fixent à l'ADN sous forme de monomères, d'homodimères ou d'hétérodimères ([Figure 2](#)) (pour une synthèse, voir Glass, 1994). Comme nous le verrons, de nombreux récepteurs forment des hétérodimères avec les récepteurs X aux rétinoïdes (RXR). Ce sont ces partenaires des RXR qui ont été le principal objet des études menées dans le cadre de ce doctorat. D'autres interactions hétérodimériques ont cependant été rapportées dont les rôles physiologiques ne seraient pas négligeables (Glass, 1994). La



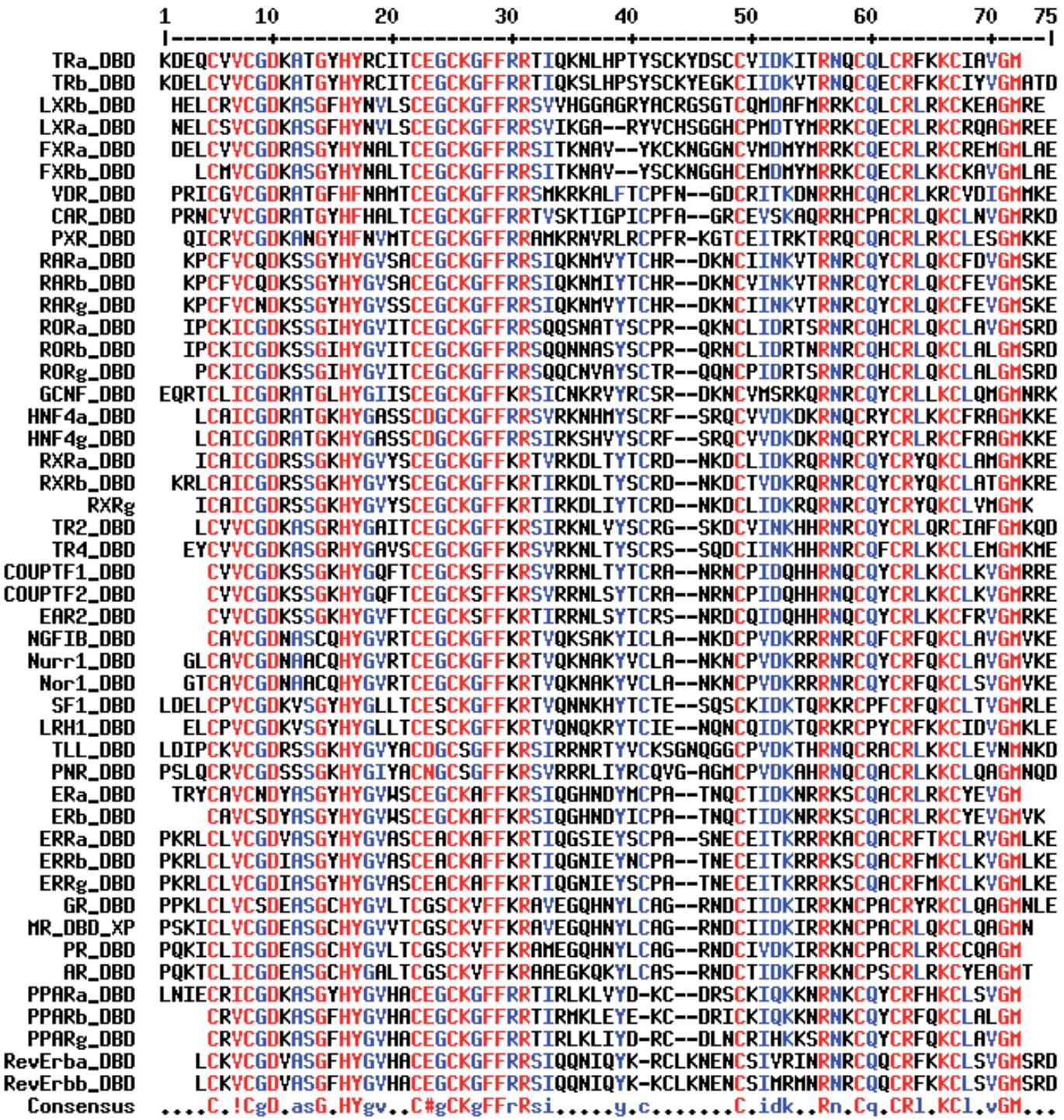
**Figure 2 : Mécanismes de fixation à l'ADN des récepteurs nucléaires**

Cette figure est inspirée de CK Glass, 1994. Le domaine de fixation du ligand (LBD) et le domaine de fixation à l'ADN (DBD) du récepteur RXR (NR2B1) ont été artificiellement fusionnés pour représenter les récepteurs nucléaires sur leurs éléments de réponse (source des structures 3D : Protein Data Bank, [www.rcsb.org](http://www.rcsb.org))



**Figure 3 : Eléments de réponse des récepteurs nucléaires**

Cette figure est inspirée de Laudet & Gronenmeyer, 2002. La fixation de monomères fait intervenir des bases supplémentaires en 5' du motif de fixation (ex : NGFI-B a pour élément de réponse 5'-AAPuGGTCA-3'). La duplication du motif de fixation génère des palindromes (PAL ou IR), des répétitions directes (DR) et des palindromes inversés (iPAL ou ER) où les deux motifs sont séparés par x paires de bases. La plupart des éléments de réponse identifiés s'écartent substantiellement d'un motif canonique et des mutations sont courantes, au moins dans l'un des deux motifs constituant l'élément de réponse. En général, les palindromes ne fixent que des homodimères alors que les DR fixent des homo- ou des hétérodimères (la polarité de fixation est précisée pour certains dimères fixant des DR : à gauche en 5' et à droite en 3'). Le récepteur situé en 3' réalise des contacts avec le petit sillon de l'ADN au niveau des bases séparant les deux motifs. Il existe des éléments de réponse constitués d'arrangements complexes du motif de fixation canonique. ERE : Estrogen Response Element, GRE : glucocorticoid RE, TRE : Thyroid RE.



**Figure 4 : Alignement des domaines de fixation à l'ADN des récepteurs nucléaires murins**

Les 47 séquences protéiques ont été obtenues à partir des bases de données GenBank et NuReBase (<http://www.ens-lyon.fr/LBMC/laudet/nurebase/nurebase.html>). L'alignement a été réalisé sous Multalin (<http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html>) avec les paramètres par défaut. Cet alignement illustre bien la forte conservation de séquence des DBD, notamment au niveau des deux doigts à atome de zinc et en particulier des quatre cystéines impliquées, dans chacune de ces structures, dans la chélation de l'atome de zinc. Les récepteurs SHP et Dax-1 ne possédant pas de domaine de fixation à l'ADN, ils n'apparaissent pas dans cet alignement.

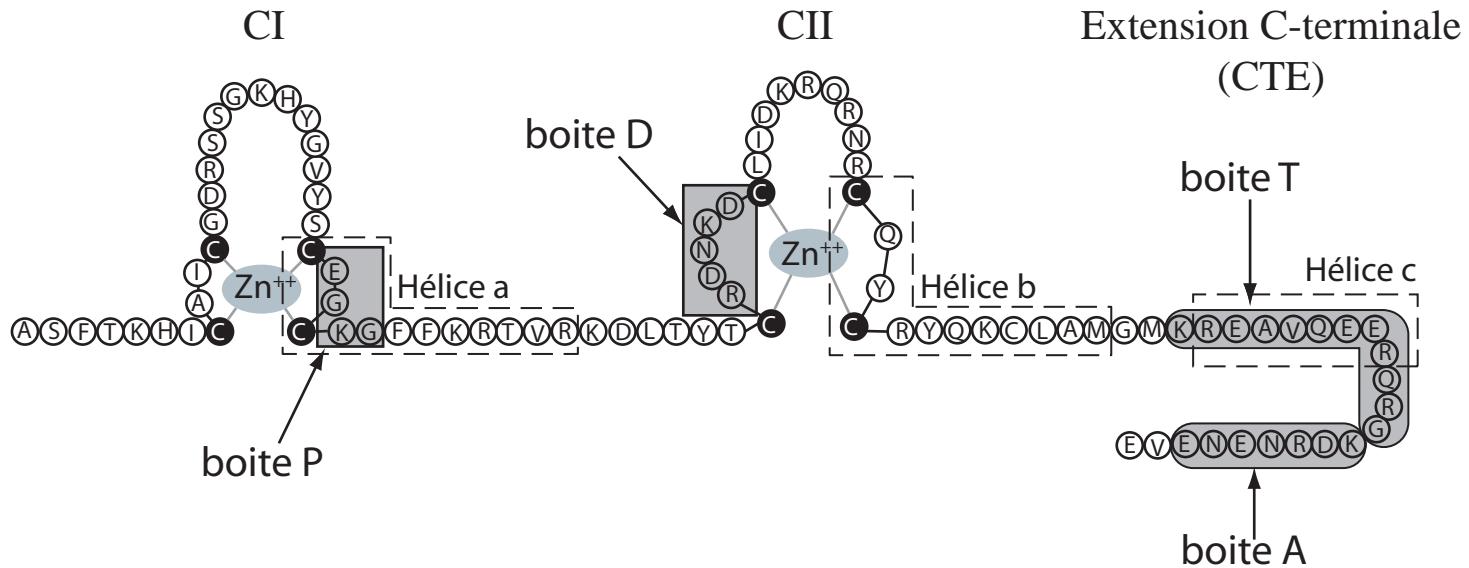


Figure 5 : Représentation schématique du domaine de fixation à l'ADN (DBD) de RXR.  
CI : doigt à atome de zinc N-terminal. C II : doigt à atome de zinc C-terminal. CTE : C-terminal extension

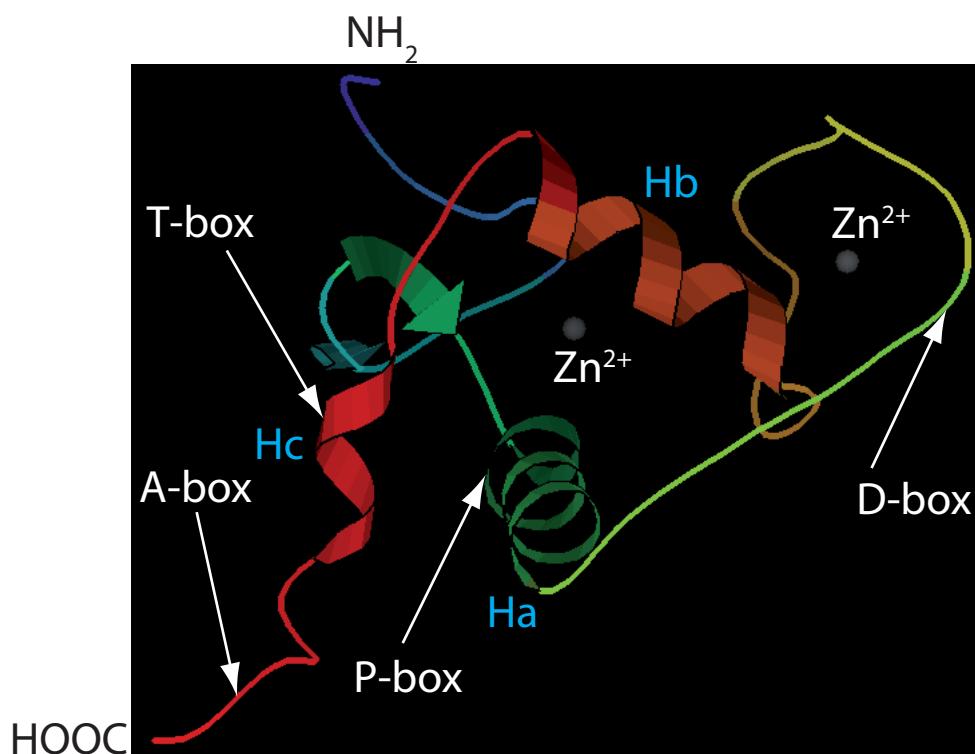
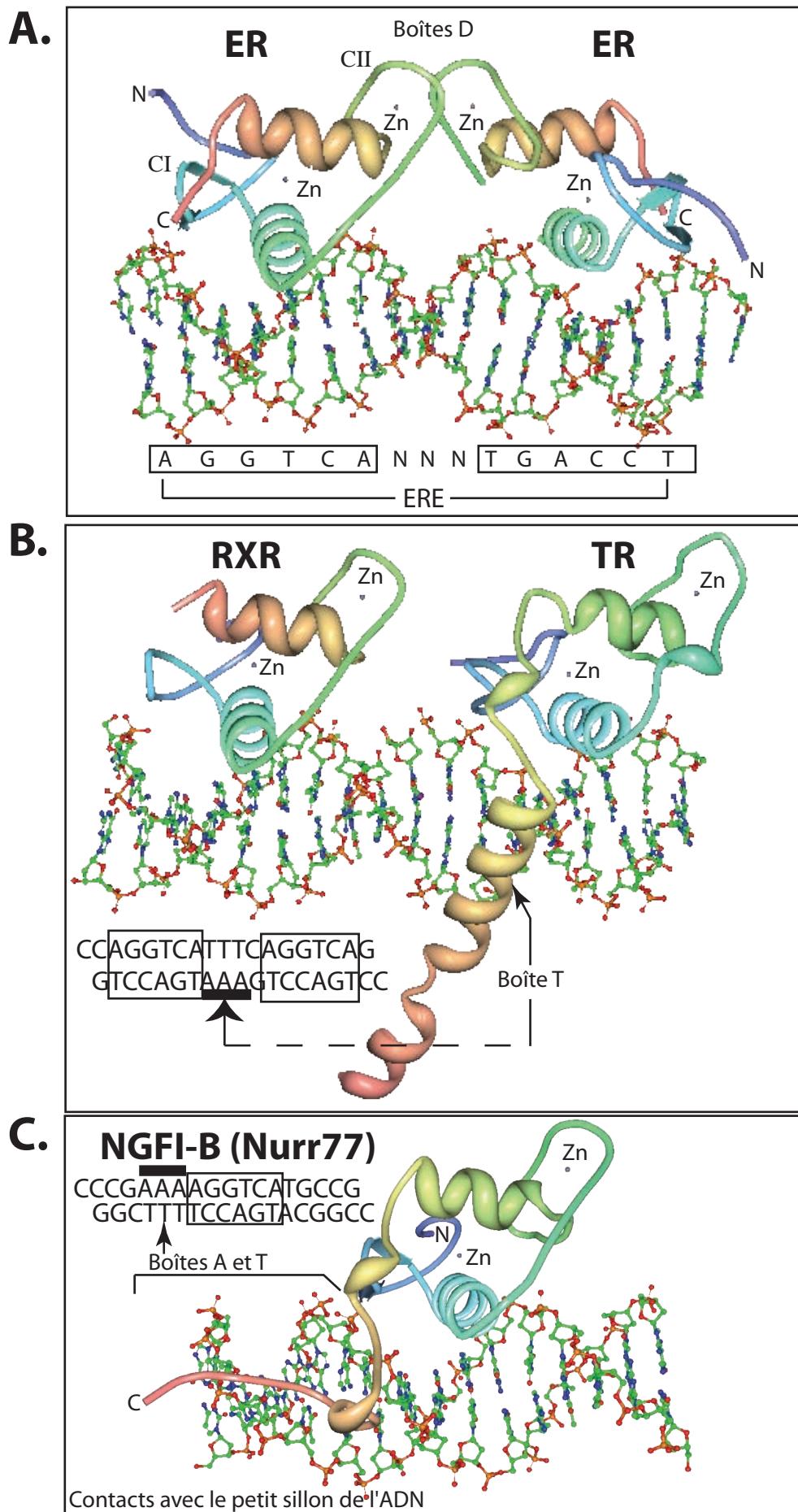


Figure 6 : Structure tridimensionnelle du domaine de fixation à l'ADN (DBD) de RXR.  
Les principaux éléments présentés en Figure 4 sont replacés sur la structure tridimensionnelle. Ha, Hb et Hc : Hélices a, b et c. P-, D-, T- et A-box : boîtes P, D, T et A. Source : Protein Data Bank, structure 1RXR, Holmbeck SM et al. J. Mol. Biol (1998)

séquence cible minimale reconnue par les récepteurs nucléaires est un hexamère que l'on qualifiera dans la suite de ce mémoire de **motif de fixation**. Certaines bases azotées de ce motif semblent communes aux motifs de fixation de presque tous les récepteurs. D'autres déterminent la spécificité de fixation de certains récepteurs uniquement. La séquence consensus de cet hexamère est 5'-PuGGTCA où Pu est une base purique (A ou G). Cependant, mis à part ER, les récepteurs aux hormones stéroïdiennes (GR, MR, PR et AR) reconnaissent préférentiellement le motif AGAACCA. Les séquences AGGTCA et AGTTCA sont généralement considérées comme des motifs de fixation "standards" des autres récepteurs. Les récepteurs se fixant à l'ADN sous la forme de monomères, comme NGFI-B par exemple, reconnaissent un unique motif de fixation précédé en 5' par une séquence riche en A/T. Dans le cas des dimères, le site de fixation, également appelé **élément de réponse**, est constitué par deux motifs de fixation (parfois appelés hémi-sites) configurés en répétitions directes (DR pour Direct Repeat), palindromiques (PAL ou IR pour Inverted Repeat) ou palindromiques inversées (iPAL ou ER pour Everted Repeat) et séparés par 0 à 5 bases azotées ([Figure 3](#)). Ainsi, trois caractéristiques des éléments de réponse déterminent la spécificité de fixation des dimères de récepteurs nucléaires : 1) la séquence du ou des motif(s) de fixation, 2) la configuration des motifs l'un par rapport à l'autre (DR, ER ou IR) et 3) l'écartement entre les deux motifs. C'est le domaine de fixation à l'ADN (DBD) qui confère aux récepteurs nucléaires leur spécificité de fixation aux éléments de réponses. Il participe également, avec d'autres interfaces d'interaction protéine-protéine au mode de fonctionnement des récepteurs en monomère, homo- ou hétérodimères. Le DBD est le domaine le plus conservé des récepteurs nucléaires ([Figure 4](#)). Il a notamment été utilisé pour identifier de nouveaux récepteurs dans le cadre de criblages de banques d'ADNc ou d'ADN génomique (voir par exemple la découverte de RAR $\alpha$  ou bien celle des PPARs Dreyer et al., 1992; Issemann and Green, 1990; Petkovich et al., 1987) et nous avons vu qu'il a été utilisé pour définir la nomenclature des RNs (Nuclear Receptor Nomenclature Committee, 1999). Le DBD est principalement constitué de deux modules en doigts à atome de zinc (CI et CII) couvrant 66-70 acides aminés et une extension C-terminale couvrant environ 25 acides aminés ([Figure 5](#)). Au sein de chacun de ces motifs, quatre cystéines chélatent un ion Zn<sup>2+</sup>. Les parties C-terminales des deux doigts à atome de zinc sont structurés en hélice  $\alpha$  (Hélices a et b, [Figure 6](#)). Plusieurs éléments de séquence du DBD ont été définis et caractérisés (boîtes P, D, T et A). La **boîte P** (Proximale) contribue à la spécificité de reconnaissance de l'élément de réponse (Freedman, 1992). En effet, l'hélice a, contenant la boîte P établie de multiples contacts de type polaires avec l'ADN lors de la fixation du DBD (voir différents exemples en



**Figure 7 : Exemples de structures de domaines de fixation à l'ADN (DBD) de différents RNs**

A. Homodimère de DBD du récepteur ER sur un Estrogen Response Element (ERE). B. Hétérodimère des DBD 5'-RXR-TR-3' sur un DR4. La boîte T détermine l'espacement minimum (4bp) entre RXR et TR par encombrement stérique. C. Monomère du DBD de NGFI-B. Les boîtes T et A interagissent avec les bases en amont du motif de fixation.

**Figure 7).** La **boîte D** (Distale) constitue une interface de dimérisation. Elle participe également dans certains cas à la sélection d'un élément de réponse spécifique comme c'est le cas pour la différenciation entre un IR3, reconnu par ER, et un IR0, reconnu par TR (Umesono and Evans, 1989). Deux autres boîtes situées dans l'extension C-terminale : les boîtes A et T, ont été décrites dans les DBD de récepteurs nucléaires pouvant fonctionner en hétérodimères. La **boîte A** établi des contacts avec le petit sillon de l'ADN déterminant ainsi : 1) des spécificités de bases jouxtant un motif de fixation, comme dans le cas de NGFI-B fixé sous la forme d'un monomère sur un NBRE (NGFI-B Response Element) : 5'-(A/T)AAAGGTCA (Meinke and Sigler, 1999; Wilson et al., 1993) ou bien 2) le nombre (et la nature dans une certaine mesure) de bases séparant deux motifs de fixation comme c'est le cas pour le dimère 5'-RXR-TR fixé sur un DR4 où la boîte A de TR détermine une distance minimale de 4 bases entre les deux motifs par un phénomène d'encombrement stérique (Figure 7 et Rastinejad et al., 1995; Zechel et al., 1994; Zelent et al., 1991). Enfin, la **boîte T** intervient comme surface d'homo- ou d'hétérodimérisation dans le cas des dimères de RXR (Lee et al., 1993; Zechel et al., 1994). La structure tridimensionnelle du DBD de plusieurs récepteurs nucléaires a été résolue, soit en solution par RMN (éventuellement complexé à l'ADN), soit sous la forme de cristaux, en complexe avec l'ADN sous la forme de monomères ou de dimères par diffraction aux rayons X. La [Table 2](#) présente les différentes structures de DBD de récepteurs nucléaires disponibles (seuls les récepteurs présentés en [Table 1](#) sont considérés) et la [Figure 7](#) illustre trois exemples de complexe DBD de récepteur(s)-ADN (Références PDB [Fig. 7A](#) : 1HCQ, [7B](#) : 2NLL et [7C](#) : 1CIT). L'étude de ces structures a permis de mieux comprendre les déterminants de la fixation à l'ADN et de la reconnaissance des éléments de réponse situés dans les promoteurs des gènes cibles. La structure de la zone comportant les deux doigts de zinc est particulièrement conservée entre les différents récepteurs nucléaires qui ont été étudiés. La [Figure 6](#) présente la structure tridimensionnelle du DBD de RXR obtenue par RMN (Holmbeck et al., 1998). La forme globulaire et la disposition tridimensionnelle des deux doigts de zinc et des hélices  $\alpha$  a et b est globalement conservée dans tous les DBD présentés en [Table 2](#). En particulier, les hélices a et b sont perpendiculaires et les résidus de l'hélice a, en particulier ceux de la boîte P, établissent des contacts étroits avec le motif de fixation (voir les exemples en [Fig. 7](#)). La structure de l'extension C-terminale des DBD de différents récepteurs nucléaires qui contient les boîtes T et A, est en revanche moins conservée ([Fig. 7](#)). En fonction des récepteurs, cette extension C-terminale joue des rôles divers et souvent multiples : contacts avec l'ADN augmentant l'affinité pour des séquences d'ADN spécifiques (TR) ou déterminant des spécificités de bases

qui jouxtent ou séparent les motifs de fixation (Rev-Erb), encombrement stérique déterminant le nombre de bases séparant deux motifs de fixation (TR, VDR) ou encore interface de dimérisation. Il est intéressant de noter que, pour plusieurs récepteurs nucléaires, les structures obtenues par RMN (récepteur en solution) indiquent que l'extension C-terminale (voir Fig. 5) est peu structurée en l'absence d'ADN, alors qu'elle apparaît structurée dans les complexes récepteur(s)-ADN obtenus par cristallographie. Ceci semble indiquer que l'ADN joue un rôle dans la formation et/ou la stabilisation de sous-structures des récepteurs nucléaires qui elles-mêmes déterminent les séquences qui jouxtent ou séparent les motifs de fixation. L'ADN participe ainsi à un accroissement coopératif de l'affinité spécifique et jouerait donc le rôle d'un effecteur allostérique positif de sa propre reconnaissance (Meinke and Sigler, 1999).

Récepteur	Ref PDB	Element de réponse	Références
<b>Structures RMN</b>			
GR	2GDA	-	(Baumann et al., 1993)
GR	1GDC	-	(Baumann et al., 1993)
GR	1RGD	-	(Hard et al., 1990; van Tilborg et al., 1995)
ER	1HCP	-	(Schwabe et al., 1993b)
RXR	1RXR	-	(Holmbeck et al., 1998)
RAR $\beta$	1HRA	-	(Katahira et al., 1992)
SF-1	2FF0	atypique (promoteur de l'inhibine $\alpha$ )	(Little et al., 2006)
ERR2	1LO1	typique (TCAAGGTCA)	(Gearhart et al., 2003)
<b>Homodimères</b>			
GR	1GLU, 1R4O, 1R4R	GRE	(Luisi et al., 1991)
GR mutant	1LAT	atypique	(Gewirth and Sigler, 1995)
ER	1HCQ	ERE	(Schwabe et al., 1993a)
Rev-erb	1A6Y, 1GA5, 1HLZ	DR2	(Sierk et al., 2001; Zhao et al., 1998)
RXR	1BY4	DR1	(Zhao et al., 2000)
AR	1R4I	ADR3 (Androgen DR3)	(Shaffer et al., 2004)
VDR	1KB2 1KB4 1KB6	SPP (mouse osteopontin VDRE) DR3 OC (rat osteocalcin VDRE)	(Shaffer and Gewirth, 2002)
<b>Hétérodimères</b>			
5'-RXR-TR-3'	2NLL	DR4	(Rastinejad et al., 1995)
5'-RAR-RXR-3'	1DSZ	DR1	(Rastinejad et al., 2000)
5'-VDR-RXR-3' (polarité non physiologique)	1YNW	DR3	(Shaffer and Gewirth, 2004)
<b>Monomères</b>			
NGFI-B	1CIT	NBRE	(Meinke and Sigler, 1999)
LRH-1	2A66	hCYP7A1 promoter (TCAAGGCCA)	(Solomon et al., 2005)

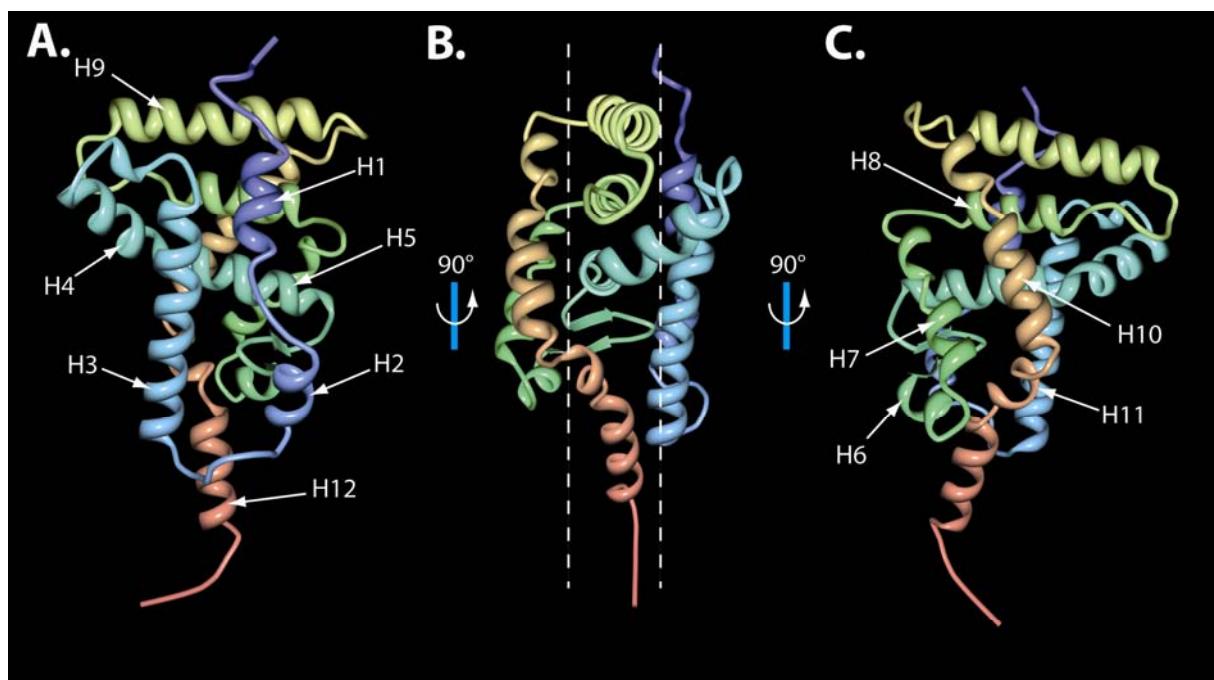
**Table 2 :** structures élucidées des domaines de fixation à l'ADN (DBD) des récepteurs nucléaires.

La structure de l'hétérodimère VDR-RXR déterminée par Schaffer et al., correspond à une forme mutée de VDR. L'observation d'un dimère dont la polarité est non physiologique (le dimère RXR-VDR fonctionne physiologiquement avec RXR localisé en 5') est liée à des contraintes dues à la cristallisation. PDB : Protein Data Bank ([www.pdb.org](http://www.pdb.org)).

### I.2.3. Le domaine charnière D

Le **domaine D** est une région moins conservée qui agit comme une charnière flexible entre les domaines C (DBD) et E (LBD). Il est donc souvent dénommé **domaine charnière** (hinge region en anglais). Le domaine D comprend, pour certains récepteurs, une partie de l'extension C-terminale du domaine C, en particulier les boîtes T et A. Cette région permettrait aux domaines C et E d'adopter plusieurs conformations différentes tout en évitant les gênes stériques. Le domaine D contient généralement un signal de localisation nucléaire (NLS pour Nuclear Localization Signal) qui peut parfois s'étendre sur le domaine C (pour revue voir Gronemeyer and Laudet, 1995). Il semble exister trois catégories différentes de récepteurs nucléaires en fonction de leurs localisations subcellulaires : 1) les récepteurs exclusivement cytoplasmiques comme GR (Htun et al., 1996) ou AR (Georget et al., 1998), 2) les récepteurs à la fois cytoplasmiques et nucléaires comme MR (Fejes-Toth et al., 1998) ou TR $\beta$  (Zhu et al., 1998) et enfin 3) les récepteurs exclusivement (ou très majoritairement) nucléaires comme PR (Guiochon-Mantel et al., 1989; Lim et al., 1999a), VDR (Michigami et al., 1999) ou RXR (Yasmin et al., 2005). Cependant, dans plusieurs cas, ces localisations subcellulaires restent controversées et il n'est pas rare d'obtenir des observations contradictoires entre des lignées cellulaires et des cellules *in vivo* (Kawana et al., 2003), voire entre différentes lignées cellulaires (Aschrafi et al., 2006). Il faut noter que les récepteurs situés dans le cytoplasme en l'absence de ligand (forme apo-récepteur) subissent une translocation dans le noyau suite à la fixation du ligand (Fejes-Toth et al., 1998; Georget et al., 1998; Htun et al., 1996; Kawana et al., 2003; Lim et al., 1999a; Yasmin et al., 2005; Zhu et al., 1998). La localisation subcellulaire des RNs est vraisemblablement le résultat d'un équilibre dynamique de transport bidirectionnel entre le cytoplasme et le noyau (navette nucléo-cytoplasmique) comme cela a été montré pour PR (Guiochon-Mantel et al., 1991) ou GR (Madan and DeFranco, 1993). De courtes séquences riches en acides aminés basiques, appelées signaux de localisation nucléaire (NLS) permettent la translocation nucléaire des récepteurs cytoplasmiques. D'autre part, notons que le domaine C (DBD) fonctionne chez de nombreux récepteurs comme signal d'export nucléaire (NES pour Nuclear Export Signal Black et al., 2001). Un ou plusieurs NLS peuvent être présents dans les récepteurs nucléaires. Ils sont en général situés dans les domaines C, D et E. Les NLS situés dans les régions C et D sont le plus souvent actifs constitutivement, en particulier lorsque le récepteur est fortement exprimé. En revanche, les NLS situés dans le domaine de fixation du ligand (région E) ne sont actifs qu'en présence d'un ligand. Leur séquence et leur localisation précise dans le domaine E

reste néanmoins largement inconnue. Il est possible que la fixation du ligand entraîne une exposition du NLS permettant l'interaction avec des facteurs d'import nucléaire qui facilitent la translocation à travers les pores nucléaires (Nakielny and Dreyfuss, 1999; Pemberton and Paschal, 2005). Ce modèle reste néanmoins débattu car des travaux récents ont montré que des importines peuvent interagir avec le NLS situé dans le domaine E de GR, y compris en l'absence de ligand (Freedman and Yamamoto, 2004). Un modèle plus complexe, prenant en compte les protéines chaperones et leurs interactions avec le cytosquelette est donc pressenti (Pemberton and Paschal, 2005). La faible conservation du domaine D entre les différents RNs suggère que cette région pourrait avoir des rôles divers et spécifiques de certains RNs. Ainsi, le domaine D est nécessaire à la fixation de la T3 (3,5,3'-tri-iodo-L-thyronine) et à l'hétérodimérisation avec RXR pour les récepteurs TR (Lin et al., 1991; Miyamoto et al., 2001) alors qu'il semble nécessaire à la reconnaissance spécifique de l'ADN et à l'activation de la transcription par VDR (Miyamoto et al., 2001; Shaffer et al., 2005). Enfin, le domaine D serait également, au moins pour certains récepteurs, une région d'interaction avec des cofacteurs comme c'est le cas pour TR interagissant avec les corépresseurs N-CoR ou SMRT (Chen and Evans, 1995; Horlein et al., 1995).



**Figure 8 : Domaine de fixation du ligand de RXR $\alpha$**  (Bourguet et al., 1995).  
Identifiant PDB : 1LBD. Les différentes hélices sont présentées sur les vues A et C. La vue B illustre la configuration générale en "sandwich d'hélices  $\alpha$ ".

#### I.2.4. Le domaine E de fixation du ligand

Le **domaine E** est le domaine de fixation du ligand (**LBD** pour Ligand Binding Domain). Sa séquence est moins conservée que celle du DBD mais sa structure secondaire, en 12 hélices  $\alpha$  apparaît plus conservée que sa séquence primaire (Wurtz et al., 1996). Le domaine E est responsable de plusieurs fonctions, principalement régulées par la fixation d'un ligand. En particulier, il porte la fonction de transactivation dépendante du ligand (AF-2 pour Activation Function 2), une interface de dimérisation, un autre signal de localisation nucléaire (NLS) et souvent des fonctions de répression. La structure tridimensionnelle du LBD de la plupart des récepteurs nucléaires a été élucidée en présence (holo-LBD) ou en l'absence (apo-LBD) d'un ligand. En fonction des récepteurs, des monomères, des dimères (homo- ou hétérodimères) voire des tétramères dans le cas de RXR (Gampe et al., 2000b) ont été cristallisés et étudiés par diffraction aux rayons X. La [Table 3](#) comprend l'ensemble des structures de LBD de récepteurs nucléaires présentes dans la base de données PDB fin août 2006. La structure tridimensionnelle du LBD des RNs est étonnamment similaire d'un récepteur à l'autre. Elle se présente sous la forme d'un "sandwich d'hélices  $\alpha$ " illustré en [Figure 8](#) par le récepteur RXR $\alpha$  (Bourguet et al., 1995). On distingue, dans le LBD de RXR, 12 hélices  $\alpha$  notées H1 à H12 et deux feuillets  $\beta$  situés entre H5 et H6 (Egea et al., 2002). Les hélices H4, H5, H8, H9, H11 et H12 sont "prises en sandwich" entre H1, H2 et H3 d'une part et H6, H7 et H10 d'autre part (les pointillés sur la [Figure 8B](#) délimitent la "tranche" centrale de ce sandwich). Toutes les structures de LBD de RNs élucidées jusqu'à aujourd'hui présentent une configuration similaire. On y distingue parfois des hélices  $\alpha$  ou des feuillets  $\beta$  supplémentaires ou absents. La numérotation des hélices  $\alpha$  supplémentaires reste néanmoins basée sur 12 hélices (voir par exemple l'hélice H2\* située entre les hélices H2 et H3 de PPAR $\gamma$  et ROR $\alpha$  Kallen et al., 2002; Nolte et al., 1998). Dans la structure du LBD de RXR ([Fig. 8](#)), on note une moindre densité d'hélices  $\alpha$  dans la partie basse du récepteur. Les hélices H3, H5, H7 et H11 ainsi que les deux feuillets  $\beta$  définissent une poche hydrophobe qui permet la fixation des ligands : on parle de **poche de fixation du ligand** (LBP pour Ligand Binding Pocket en anglais). Comme nous le verrons plus loin, la fixation du ligand engendre des changements conformationnels dans le LBD qui conduisent à l'apparition d'un sillon à la surface du LBD permettant l'interaction des cofacteurs qui joueront un rôle de médiateurs entre le récepteur et la machinerie transcriptionnelle et affecteront parfois la structure locale de la chromatine. L'étude combinée des structures de LBD en l'absence ou en présence de

ligands et/ou de cofacteurs a permis de mieux comprendre 1) les déterminants de la fixation des ligands 2) les modifications de conformations engendrées par les ligands de type agoniste ou antagoniste et 3) le processus de recrutement des cofacteurs. Signalons également que l'observation et l'étude de ligands dans des LBD cristallisés sans ajout de ligands par le manipulateur a permis pour certains récepteurs comme HNF4 $\alpha$ , HNF4 $\gamma$ , ROR $\alpha$  ou encore RXR $\alpha$  de découvrir de nouveaux ligands endogènes potentiels . (Dhe-Paganon et al., 2002; Kallen et al., 2004a; Kallen et al., 2002; Wisely et al., 2002; Xu et al., 2004). Enfin, l'étude de dimères de LBD a permis de mieux comprendre les déterminants de l'homo- ou de l'hétérodimérisation des récepteurs nucléaires. La fixation des ligands et les changements conformationnels engendrés par cette fixation, le recrutement des cofacteurs ainsi que les phénomènes de dimérisation sont développés dans les paragraphes suivants.

### I.2.5. Le domaine F

Certains récepteurs disposent également d'un **domaine F** dont la séquence est extrêmement variable et dont ni la structure ni la (ou les) fonction(s) n'ont été élucidées.

**Table 3 :** Structures élucidées des LBD des RNs (début de la table)

Récepteur	Ligand	Identifiants PDB	Remarque	Références
<b>Monomères</b>				
RAR $\gamma$	all- <i>trans</i> RA	2LBD	Agoniste	(Renaud et al., 1995)
	9-cis RA, BMS961	3LBD, 4LBD	Agonistes	(Klaholz et al., 1998)
	BMS270394, BMS270395	1EXA, 1EXX	Agonistes	(Klaholz et al., 2000a)
	BMS184394, CD564, BMS181156	1FCX, 1FCY, 1FCZ	Agonistes spécifiques d'isoformes	(Klaholz et al., 2000b)
	SR11254	1FD0	Agoniste	(Klaholz and Moras, 2002)
RAR $\beta$	TTNBP	1XAP	Agoniste	(Germain et al., 2004)
RXR $\alpha$	9-cis RA	1FBY	Agoniste	(Egea et al., 2000)
TR $\alpha$	T3, Dimit	-	Agoniste	(Wagner et al., 1995)
	T3	2H77, 2H79	Agoniste	(Nascimento et al., 2006)
	agoniste synthétique	1NAV	Agoniste (forte affinité pour TR $\beta$ )	(Ye et al., 2003)
TR $\beta$	T3	1BSX	Complexe avec agoniste et boîte RN de GRIP1	(Daramont et al., 1998)
	T3, T4	1XZX, 1Y0X	Agonistes	(Sandler et al., 2004)
	T3	2H6W	Agoniste	(Nascimento et al., 2006)
	agonistes synthétiques	1NAX, 1R6G, 1Q4X, 1N46	Agonistes (forte affinité pour TR $\beta$ )	(Borngraeber et al., 2003; Dow, 2003 #94; Hangeland et al., 2004; Ye et al., 2003)
	3,3',5-triodo L-acetic acid (TRIAC)	1NUO, 1NQ0, 1NQ1, 1NQ2	mutants responsables de résistances à l'hormone thyroïdienne	(Huber et al., 2003a; Huber et al., 2003b)
PPAR $\alpha$	AZ242	1I7G	Agoniste	(Cronet et al., 2001)
	GW409544	1K7L	Complexe avec agoniste et boîte RN de SRC-1	(Xu et al., 2001a)
	GW6471	1KKQ	Antagoniste et complexe avec peptide SMRT	(Xu et al., 2002a)
PPAR $\delta$	Vaccenic acid, GW2433, EPA	1GWX, 2GWX, 2BAW 3GWX	Agonistes	(Fyffe et al., 2006b; Xu et al., 1999)
	GW2331	1Y0S	Agoniste	(Takada et al., 2000)
	Vaccenic acid	2AWH, 2B50	Ligand	(Fyffe et al., 2006a)
PPAR $\gamma$	Apo	3PRG	-	(Uppenberg et al., 1998)
	AZ242	1I7I	Agoniste	(Cronet et al., 2001)
	2-BABA	1WMO	Complexe avec agoniste et boîte RN de GRIP1	(Ostberg et al., 2004)
	pyrazol-5-ylbenzenesulfonamides	2G0H, 2G0G	Agonistes spécifiques de PPAR $\gamma$	(Lu et al., 2006a)

**Table 3 :** Structures élucidées des LBD des RNs (suite)

PPAR $\gamma$	Acides Indol-1-yl acétique	2F4B	Agoniste	(Mahindroo et al., 2006)
	Rosiglitazone	1ZGY	Complexe avec agoniste et la boîte RN-2 de SHP (NR5A1)	(Li et al., 2005b)
	PA-082	2FVJ	Complexe avec agoniste spécifique de PPAR $\gamma$ et boîte RN de SRC1	(Burgermeister et al., 2006)
VDR	Vitamine D <sub>3</sub>	1DB1	Agoniste	(Rochel et al., 2000)
	Analogues 20-epi de la vitamine D <sub>3</sub> (MC1288 et KH1060)	1IE8, 1IE9	Agonistes	(Tocchini-Valentini et al., 2001)
	Calcipotriol, seocalcitol	1S19, 1S0Z	Agonistes	(Tocchini-Valentini et al., 2004)
	Analogues de la vitamine D <sub>3</sub>	1RJK, 1RK3, 1RKG, 1RKH	Complexes avec agonistes et boîte RN-2 de DRIP205	(Vanhooke et al., 2004)
	Analogue 14-epi de la vitamine D <sub>3</sub> (TX522)	1TXI	Agoniste	(Eelen et al., 2005)
LXR $\beta$	T-0901317	1UPV, 1UPW	Agoniste	(Hoerer et al., 2003)
FXR	Fexaramine	1OSH	Agoniste	(Downes et al., 2003)
	acide 6 $\alpha$ -ethyl-chenodeoxycholique, acide 3-deoxy-chenodeoxycholique	1OSV, 1OT7	Complexes avec agonistes et boîte RN-3 de GRIP1	(Mi et al., 2003)
PXR	Apo	1ILG	-	(Watkins et al., 2001)
	SR12813	1ILH	Agoniste	(Watkins et al., 2001)
	Hyperforin	1M13	Agoniste	(Watkins et al., 2003b)
	Rifampicine	1SKX	Agoniste	(Chrencik et al., 2005)
CAR	Androstenol	1XNX	Agoniste inverse	(Shan et al., 2004)
Nurr1	Apo	1OVL	-	(Wang et al., 2003)
NGFI-B	Apo	1YJE	-	(Flaig et al., 2005)
ROR $\alpha$	Cholesterol	1N83	Agoniste endogène supposé	(Kallen et al., 2002)
	Cholesterol sulfate	1S0X	Agoniste endogène supposé	(Kallen et al., 2004a)
ROR $\beta$	Acide stéarique	1K4W	Complexe avec ligand et boîte RN2 de SRC1	(Stehlin et al., 2001)
	Acide all- <i>trans</i> rétinoïque, ALRT1550	1N4H, 1NQ7	Complexes avec antagoniste endogène ou synthétique et boîte RN2 de SRC1	(Stehlin-Gaon et al., 2003)
LRH-1	Apo	1PK5	-	(Sablin et al., 2003)
	Phospholipides	1YUC	Complexe avec ligands endogènes supposés et boîte RN1 de SHP	(Ortlund et al., 2005)
	Phospholipides	1YOK	Complexe avec ligands endogènes supposés et boîte RN de TIF2	(Krylova et al., 2005)

**Table 3 :** Structures élucidées des LBD des RNs (suite)

LRH-1	Phospholipides	1ZDU	Complexe avec ligands endogènes supposés et boîte RN3 de TIF2	(Wang et al., 2005a)
	Apo	1ZH7	Complexe avec boîte RN1 de SHP	(Li et al., 2005b)
SF-1	Phospholipides	1YP0	Complexe avec ligands endogènes potentiels et boîte RN1 de SHP	(Li et al., 2005a)
	Phospholipides	1YMT	Complexe avec ligands endogènes supposés et boîte RN1 de SHP	(Krylova et al., 2005)
	Phospholipides	1YOW	Complexe avec ligands endogènes supposés et boîte RN de TIF2	(Krylova et al., 2005)
	Phospholipides	1ZDT	Complexe avec ligands endogènes supposés et boîte RN3 de TIF2	(Wang et al., 2005a)
ER $\alpha$	17 $\beta$ -estradiol	1QKT	Triple mutant (C381S / C417S / C530S), agoniste	(Gangloff et al., 2001)
	ICI 164,384	1HJ1	Antagoniste	(Pike et al., 2001)
	Tetrahydroisoquinoline	1UOM	Agoniste	(Renaud et al., 2003)
	Genistein	1X7R	Complexe avec phytoestrogène et boîte RN de SRC3	(Manas et al., 2004b)
	Dérivés de chromanes	1YIM, 1YIN	Antagonistes	(Tan et al., 2005)
	Dihydrobenzoxathiins	1SJ0	Agoniste sélectif de ER $\alpha$	(Kim et al., 2004)
	LY2066948-HCl	2AYR	Agoniste	(Hummel et al., 2005)
	dérivés de dihydrobenzoxathiins	1XP1, 1XP6, 1XP9, 1XPC	Antagonistes et agonistes partiels	(Blizzard et al., 2005)
ER $\beta$	5,11- <i>cis</i> -diethyl-5,6,11,12-tetrahydrochrysène-2,8-diol (THC)	1L2J	Antagoniste	(Shiau et al., 2002)
	Modulateur sélectif de type triazine	1NDE	agoniste sélectif de ER $\beta$	(Henke et al., 2002)
ERR $\alpha$	Apo	1XB7	Complexe avec motif L3 de PGC-1 $\alpha$	(Kallen et al., 2004b)
GR	Dexaméthasone	1P93	Complexe avec agonsite et boîte RN-3 de TIF2	(Kauppi et al., 2003)
	RU-486	1NHZ	Antagoniste	(Kauppi et al., 2003)
PR	Progesterone	1A28	Agoniste	(Williams and Sigler, 1998)
	Mometasone, norethindrone	1SR7, 1SQN	Agonistes	(Madauss et al., 2004)
	Metribolone (R1881)	1E3K	Agoniste	(Matias et al., 2000)

**Table 3 :** Structures élucidées des LBD des RNs (suite)

PR	Tanaproget	1ZUC	Agoniste non stéroïdien	(Zhang et al., 2005b)
MR	Aldostéron, progestérone, deoxycorticostérone	2AA2, 2AA5, 2AA7	Mutant C808S; agonistes (agoniste faible pour la progestérone)	(Bledsoe et al., 2005)
	Progestérone, cortisone, spironolactone	2AA6, 2AAX, 2AB2	Mutant C808S/S810L; agonistes (pour le mutant S810L uniquement)	(Bledsoe et al., 2005)
	Corticostérone	2A3I	Mutant C808S en complexe avec agoniste et boîte RN-4 de SRC1	(Li et al., 2005c)
	Deoxycorticostérone, progestérone	1Y9R, 1YA3	Mutant C910A/S810L; agonistes (pour le mutant uniquement)	(Fagart et al., 2005)
	Deoxycorticostérone	2ABI	Mutant C910A; agoniste	Huyet J et al., non publié
AR	Metribolone (R1881)	1E3G	Agoniste	(Matias et al., 2000)
	Dihydrotestostérone (DHT)	1I38, 1I37	Présence (1I38) ou absence (1I37) de la mutation T877A	(Sack et al., 2001)
	DHT	1T63, 1T5Z, 1T65, 1XJ7	Complexe avec agoniste et peptides coactivateurs (SRC2-3, ARA70, SRC2-2 ou SRC3-2)	(Estebanez-Perpina et al., 2005)
	Testostérone, DHT, tetrahydrogestrinone	2AM9, 2AMA, 2AMB	Agonistes naturels et synthétique	(Pereira de Jesus-Tran et al., 2006)
	R1881	1XOW, 2AO6, 1XQ3	Complexe avec agoniste seul (1XQ3) ou boîte RN3 de TIF2 (2AO6) ou motif FXXLF du domaine A/B de AR (1XOW)	(He et al., 2004)
<b>Homodimères</b>				
PPAR $\gamma$	Apo	1PRG	-	(Nolte et al., 1998)
	Rosiglitazone	2PRG	Complexe avec agoniste et boîtes RN de SRC1	(Nolte et al., 1998)
	GW0072	4PRG	Agoniste partiel	(Oberfield et al., 1999)
	Dérivé d'acide tricyclique-alpha-alkyloxyphényl-propionique	1KNU	Agoniste sélectif pour PPAR $\alpha$ et $\gamma$	(Sauerberg et al., 2002)
	Ragaglitazar	1NYX	Agoniste sélectif pour PPAR $\alpha$ et $\gamma$	(Ebdrup et al., 2003)
	Dérivé d'acide alpha-aryloxyphénylacétique	1ZEO	Agoniste sélectif pour PPAR $\alpha$ et partiel pour PPAR $\gamma$	(Shi et al., 2005)

**Table 3 :** Structures élucidées des LBD des RNs (suite)

Homodimères (suite)				
RXR $\alpha$	Apo	1LBD	-	(Bourguet et al., 1995)
	BMS649, acide docosahexaenoïque (DHA)	1MVC, 1MZN, 1MV9	Complexes avec agonistes et boîte RN2 de GRIP1	(Egea et al., 2002)
RXR $\beta$	LG100268	1H9U	Agoniste	(Love et al., 2002)
LXR $\beta$	24(S),25-epoxycholesterol	1P8D	Complexe avec agoniste et boîte RN de SRC1	(Williams et al., 2003)
	GW3965, T-0901317	1PQ6, 1PQ9, 1PQC	Agonistes	(Farnegardh et al., 2003)
PXR	SR12813	1NRL	Complexe avec agoniste et boîte RN-2 de SRC1	(Watkins et al., 2003a)
HNF4 $\alpha$	Acides gras saturés et monoinsaturés (C14-C18)	1M7W	Ligands endogènes supposés	(Dhe-Paganon et al., 2002)
	Acides gras saturés et monoinsaturés	1PZL	Ligands endogènes supposés	(Duda et al., 2004)
HNF4 $\gamma$	Acides gras saturés et monoinsaturés (C14-C18)	1LV2	Ligands endogènes supposés	(Wisely et al., 2002)
ER $\alpha$	17 $\beta$ -oestradiol, raloxifène	1ERE, 1ERR	Agoniste, antagoniste partiel	(Brzozowski et al., 1997)
	Diethylstilbestrol (DES)	3ERTD	Complexe avec agoniste et boîte RN-2 de GRIP1	(Shiau et al., 1998)
	4-hydroxytamoxifen (OHT)	3ERT	Antagoniste partiel	(Shiau et al., 1998)
	17 $\beta$ -estradiol	1A52	Agoniste	(Tanenbaum et al., 1998)
	17 $\beta$ -estradiol	1G50	Agoniste	(Eiler et al., 2001)
	THC	1L2I	Complexe avec agoniste et boîte RN-2 de GRIP1	(Shiau et al., 2002)
	Noyau aroylbenzothiophène du raloxifène (RALcore)	1GWQ	Complexe avec agoniste et boîte RN-2 de TIF2	(Warnmark et al., 2002)
	17 $\beta$ -estradiol	1GWR	Complexe avec agoniste et boîte RN-3 de TIF2	(Warnmark et al., 2002)
	WAY-244	1X7E	Complexe avec agoniste sélectif de ER $\beta$ et boîte RN de SRC3	(Manas et al., 2004a)
	OHT	2BJ4	Complexe avec antagoniste et peptide corépresseur	(Kong et al., 2005)
	Dérivés de chromanes	1YIM, 1YIN	Antagonistes	(Tan et al., 2005)
	GW5638	1R5K	Antagoniste	(Wu et al., 2005b)
	Dérivés de tetrahydroisoquinoline	1XQC	Antagoniste sélectif de ER $\alpha$	(Renaud et al., 2005)
	OBCP-1M, -2M et -3M	2B1V, 2FAI, 1ZKY	complexes avec agonistes sélectifs de ER $\beta$ et boîte RN-2 de GRIP1	(Hsieh et al., 2006)

**Table 3 :** Structures élucidées des LBD des RNs (suite)

Homodimères (suite)				
ER $\alpha$	17 $\beta$ -estradiol	1PCG	Complexe avec agoniste et un peptide synthétique à action antagoniste	(Leduc et al., 2003)
ER $\beta$	Raloxifène, genistein	1QKN, 1QKM	Antagoniste, agoniste partiel	(Pike et al., 1999)
	WAY-697, WAY-338, WAY-797, WAY-397, WAY-244, ERB-041, CL-272, WAY-202196, 1-chloro-6-(4-hydroxy-phenyl)-naphthalen-2-ol	1X76, 1U3R, 1U3S, 1U9E, 1X78, 1X7B, 1U3Q, 1YYE, 1YY4	Complexes avec ligands sélectifs de ER $\beta$ et boîte RN de SRC1	(Malamas et al., 2004; Manas et al., 2004a; Mewshaw et al., 2005)
	Genistein	1X7J	Complexe avec phytoestrogène et boîte RN de SRC1	(Manas et al., 2004b)
	2-arylindene-1-one	1ZAF	Complexe avec ligand et peptide SRC1	(McDevitt et al., 2005)
	OHT	2FSZ	Antagoniste	(Wang et al., 2006c)
ERR $\gamma$	Apo	1KV6, 1TFC	Complexe avec boîte RN-2 de SRC1	(Greschik et al., 2004; Greschik et al., 2002)
	DES, OHT	1S9P, 1S9Q, 1VJB	Antagonistes	(Greschik et al., 2004)
	GSK5182	2EWP	Antagoniste	(Chao et al., 2006)
GR	Dexaméthasone	1M2Z	Complexe avec agoniste et boîte RN-3 de TIF2	(Bledsoe et al., 2002)
Hétérodimères				
RAR $\alpha$ /RXR $\alpha$	BMS614 / acide oléique	1DKF	Antagoniste / agoniste partiel	(Bourguet et al., 2000b)
RAR $\beta$ /RXR $\alpha$	acide 9- <i>cis</i> rétinoïque	1DXD	Complexe avec agoniste et boîte RN2 de TRAP220	(Pogenberg et al., 2005)
PPAR $\gamma$ /RXR $\alpha$	Rosiglitazone / acide 9- <i>cis</i> rétinoïque	1FM6	Complexe avec agonistes et boîte RN de SRC1	(Gampe et al., 2000a)
PPAR $\gamma$ /RXR $\alpha$	GI262570 (farglitazar) / acide 9- <i>cis</i> rétinoïque	1FM9	Complexe avec agonistes et boîte RN de SRC1	(Gampe et al., 2000a)
PPAR $\gamma$ /RXR $\alpha$	GI262570 / acide tetrahydrobenzofuranyl propenoïque	1RDT	Complexe avec agonistes et boîte RN de SRC1	(Haffner et al., 2004)
PPAR $\gamma$ /RXR $\alpha$	GW409544 / acide 9- <i>cis</i> rétinoïque	1K74	Complexe avec agonistes et boîte RN de SRC1	(Xu et al., 2001a)
LXR $\alpha$ /RXR $\beta$	T-0901317 / acide méthoprène	1UHL	Complexe avec agonistes et peptide GRIP1	(Svensson et al., 2003)
LXR $\alpha$ /RXR $\alpha$	GSK3987 / acide rétinoïque	2ACL	Agonistes	(Jaye et al., 2005)

**Table 3 :** Structures élucidées des LBD des RNs (suite)

Hétérodimères (suite)				
hCAR/hRXR $\alpha$	CITCO / acides gras mono- et polyinsaturés (C16, C18)	1XVP	Complexe avec agonistes et boîte RN-2 de SRC1	(Xu et al., 2004)
hCAR/hRXR $\alpha$	5 $\beta$ -pregnane-3,20-dione / acides gras mono- et polyinsaturés (C16, C18)	1XV9	Complexe avec agonistes et boîte RN-2 de SRC1	(Xu et al., 2004)
mCAR/RXR $\alpha$	TCPOBOP / acide 9- <i>cis</i> rétinoïque	1XLS	Complexe avec agonistes et boîte RN-3 de TIF2	(Suino et al., 2004)
Tetramère				
RXR $\alpha$	Apo	1G1U	-	(Gampe et al., 2000b)
	<i>Trans</i> -isomère de l'acide rétinoïque	1G5Y	Ligand inactivant	(Gampe et al., 2000b)

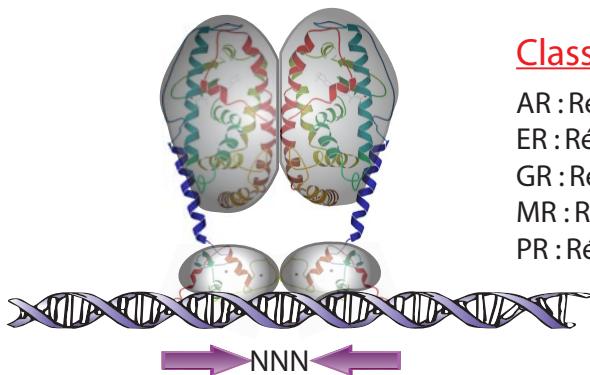
**Table 3 :** Structures élucidées des domaines de fixation des ligands (LBD) des récepteurs nucléaires.

Pour plusieurs récepteurs, l'introduction de mutations s'est avérée nécessaire à l'obtention de cristaux. Ces mutations sont soit précisées dans cette table soit dans les entrées PDB correspondantes. PDB : Protein Data Bank ([www.pdb.org](http://www.pdb.org)). Apo : absence de ligand. Boîte RN : boîte Récepteur Nucléaire (motif LXXLL, voir paragraphe sur les co-facteurs). Les abréviations des cofacteurs sont fournies dans le paragraphe correspondant. Date de mise à jour de la table : 28/08/2006.

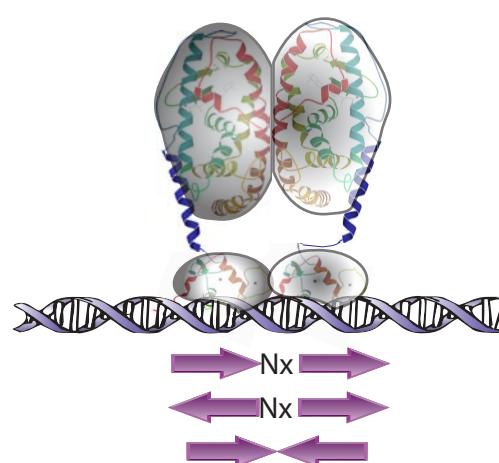
### I.3. ELEMENTS DE REONSE ET DIMERISATION DES RECEPTEURS NUCLEAIRES

#### I.3.1 Eléments de réponse et classe des récepteurs nucléaires

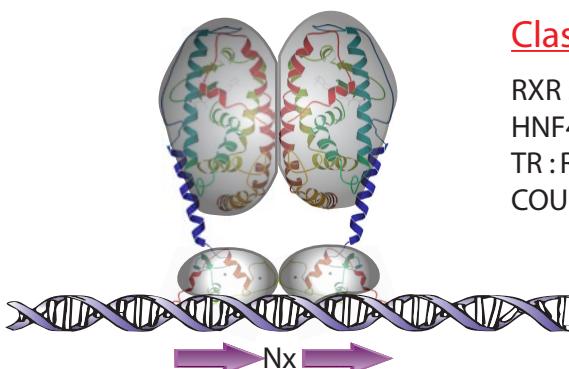
Comme nous l'avons vu précédemment, les récepteurs nucléaires reconnaissent des séquences spécifiques (**éléments de réponse**) situées dans les régions régulatrices de leurs gènes cibles. Ces régions régulatrices se localisent au niveau du promoteur ou bien dans des "enhancers" situés plus en amont (jusqu'à plusieurs kilobases) du site d'initiation de la transcription. Ces séquences sont constituées d'un arrangement particulier de **motifs de fixation** (Fig. 3). Ces motifs de fixation sont du type AGGTCA ou AGTTCA pour la plupart des récepteurs et de type AGAACCA pour les récepteurs aux hormones stéroïdiennes (sauf pour ER). Certains récepteurs reconnaissent également des bases qui flanquent ou séparent les motifs de fixation *via* leur DBD (cas de NGFI-B Fig. 7C). L'arrangement des motifs de fixation reflète le mode de fonctionnement en monomère, en homodimère ou en hétérodimère (Fig. 2). La séquence, la disposition et l'écartement des motifs sont les éléments clefs déterminant la fixation des récepteurs sur un élément de réponse.

Classe I

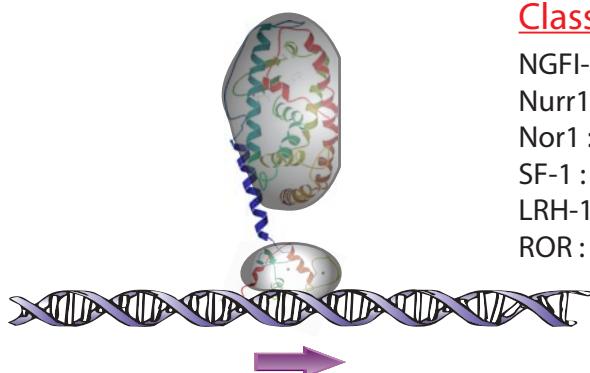
AR : Récepteur aux androgènes	IR3
ER : Récepteurs aux oestrogènes	IR3
GR : Récepteur aux glucocorticoïdes	IR3
MR : Récepteur aux minéralocorticoïdes	IR3
PR : Récepteur à la progestérone	IR3

Classe II

RAR : Récepteur à l'acide rétinoïque	DR2, DR5, IR0, ER8
PPAR : Réc. aux proliférateurs de peroxyosomes	DR1, DR2
LXR : Récepteur X du foie	DR4
PXR : Récepteur X aux pregnanes	DR3, DR4, ER6
VDR : Récepteur à la vitamine D <sub>3</sub>	DR3, ER9
TR : Récepteurs à l'hormone thyroïdienne	DR4, IR0, ER6-8
CAR : Récepteur constitutif aux androstanes	DR4, DR5, ER6
FXR : Récepteur X aux farnésoides	DR5, IR1

Classe III

RXR : Récepteur X aux rétinoïdes	DR1, IR
HNF4 : Facteur nucléaire hépatique 4	DR1, DR2
TR : Récepteurs testiculaires	DR1, DR3-5
COUP-TF : Récepteurs COUP-TF	DR1, DR6, DR4,...

Classe IV

NGFI-B : Nerve growth factor induced protein I-B	hémisite, IR, DR5
Nurr1 : Protéine 1 reliée à Nur	hémisite, IR, DR5
Nor1 : Récepteur orphelin neuronal 1	hémisite, IR
SF-1 : Facteur stéroidogénique 1	hémisite
LRH-1 : Homologue 1 du récepteur du foie	hémisite
ROR : récepteurs reliés à RAR	hémisite, DR2

**Figure 9 : Les quatre classes de récepteurs nucléaires**

Adapté et complété à partir de Mangelsdorf DJ &amp; Evans RM, 1995, Cell, 83:841-50.

En fonction de leurs propriétés de dimérisation et des éléments de réponse auxquels ils sont capables de se lier, quatre classes de récepteurs nucléaires ont été définies (Mangelsdorf and Evans, 1995). Ces quatre classes sont illustrées en [Figure 9](#) et détaillées ci-dessous :

- La **classe I** regroupe les récepteurs aux hormones stéroïdiennes AR, ER, GR, MR et PR. Ces récepteurs fonctionnent en homodimères et reconnaissent des éléments de réponse de type palindrome où les motifs de fixation sont séparés par 3 paires de base (IR3, [Fig. 7A](#)). La séquence consensus des motifs de fixation est AGAACCA, sauf pour ER qui reconnaît préférentiellement la séquence AGGTCA. La discrimination entre les éléments de réponse de ER et de GR est donc due aux deux bases centrales du motif de fixation qui réalisent des interactions spécifiques avec les boîtes P des DBD de ces deux RNs (Luisi et al., 1991; Schwabe et al., 1993b).
- La **classe II** regroupe les partenaires d'hétérodimérisation du récepteur RXR. Ils se fixent de manière générale sur des éléments de réponse organisés en répétition directe (DRx où x précise le nombre de bases séparant les deux motifs) bien que certains récepteurs reconnaissent également des éléments de réponse constitués de motifs symétriques (IR ou ER).
- La **classe III** regroupe les récepteurs non stéroïdiens fonctionnant en homodimères et reconnaissant généralement des éléments de réponse de type répétition directe (DRx) comme RXR et HNF4.
- La **classe IV** est constituée par les récepteurs se fixant sous la forme de monomère et reconnaissant, le plus souvent, des motifs de fixation qui s'étendent au-delà de la séquence consensus.

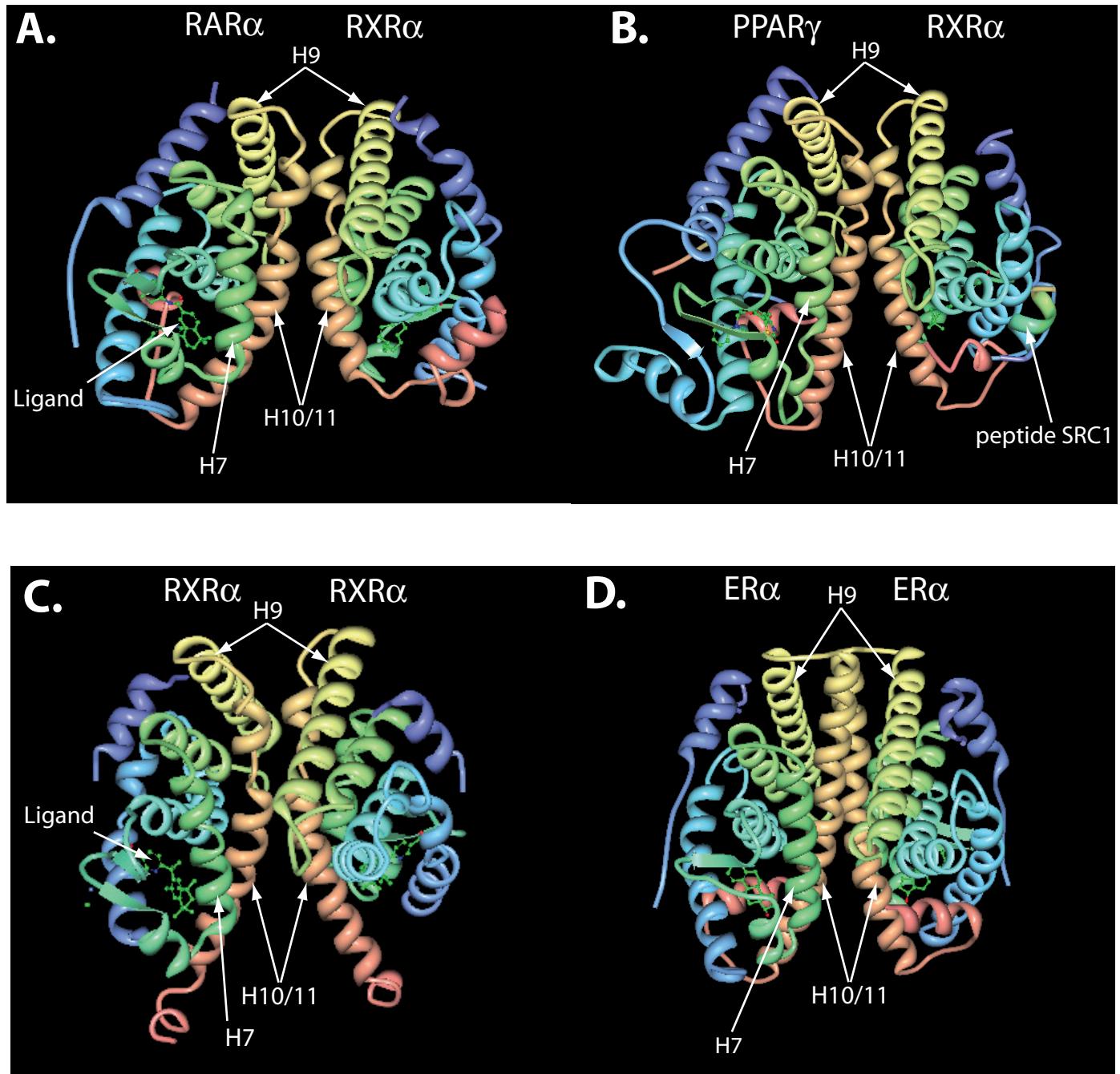
La [Figure 9](#) (adaptée de Mangelsdorf and Evans, 1995) présente les principaux récepteurs de ces quatre classes ainsi que leurs préférences en termes d'éléments de réponse. Cette classification reste naturellement imparfaite et certains récepteurs pourraient appartenir à différentes classes. Par exemple les récepteurs Rev-erb et ROR se fixent à l'ADN sous la forme de monomères ou sous la forme d'homodimères (DR2) et pourraient donc entrer dans les classes III ou IV (Adelman et al., 1996; Harding et al., 1997; Harding and Lazar, 1993; Harding and Lazar, 1995). Enfin, les éléments de réponse de nombreux gènes présentent des motifs qui divergent de manière notable de la séquence consensus. Une dizaine d'éléments de réponse naturels des récepteurs RAR est présentée dans la [Table 4](#). Ils ont été identifiés et caractérisés en particulier grâce aux techniques de gène-rapporteur et de retard sur gel. L'exemple des récepteurs RAR ([Table 4](#)) illustre bien la diversité des éléments de réponse reconnus par un récepteur donné. Dans le Chapitre II, des éléments de réponse naturels pour

d'autres récepteurs de classe II sont présentés, permettant ainsi d'observer la conservation plus ou moins importante des règles canoniques que nous avons évoquées précédemment. Un grand nombre d'éléments de réponse naturels ou synthétiques a également été présenté par H. Gronemeyer & V. Laudet (Gronemeyer and Laudet, 1995).

Gène cible	Gene ID	Elément de réponse	Remarques	Référence
Mouse RAR $\beta$ 2	218772	CGGGAAAG <u>GGTCA</u> CCGAA <u>AGTTC</u> ACTGCATA <u>T</u> -53 -37	RAR $\alpha$ , $\beta$ ou $\gamma$ /RXR $\alpha$ sur un DR5 ( $\beta$ RARE)	(de The et al., 1990; Hoffmann et al., 1990; Sucov et al., 1990; Vivanco Ruiz et al., 1991)
Rat ABCC2 (MRP2)	25303	<u>GAGTTC</u> ACAGAGAT <u>GGTAA</u> ATACC <u>C</u> -422 -398	RAR $\alpha$ /RXR $\alpha$ sur un DR5	(Denson et al., 2000)
Mouse Cyp26	13082	<u>CAAATTAGTTCA</u> OCCAA <u>AGTTC</u> ATCT <u>C</u> -95 -120	RAR $\gamma$ /RXR $\alpha$ sur un DR5	(Loudig et al., 2000)
Mouse ABCA1	11303	CGAGGCGAG <u>GGTCA</u> TACTAT <u>GGTCA</u> AAGGCCAGT -55 -70	RAR $\gamma$ /RXR $\alpha$ sur un DR4 reconnu par LXR/RXR	(Costet et al., 2003)
Mouse CRBPI	19659	TTAGTAG <u>GGTCAAA</u> <u>AGGTCA</u> GACAC -1011 -998	RAR $\alpha$ , $\beta$ ou $\gamma$ /RXR $\alpha$ sur un DR2	(Smith et al., 1991)
Mouse CRABPII	12904	RARE1 : <u>TOCCAGTTC</u> AC <u>AGTC</u> ACGGCT -163 -140 RARE2 : GAGGCTAG <u>AGGGCA</u> <u>GAGGTCA</u> CAGC -633 -658	RAR $\alpha$ , $\beta$ ou $\gamma$ /RXR $\alpha$ , $\beta$ ou $\gamma$ sur DR2 ou DR1	(Durand et al., 1992)
Mouse MMP11	17385	<b>3'-DR2 :</b> AGCTTCTGA <u>AGGTCA</u> <u>GGTCA</u> AATCCA -2141 -2128 <b>DR1 :</b> AGCTTCTGA <u>GGTCA</u> CATT <u>TC</u> ATAGTA -367 -355	RAR $\alpha$ , $\beta$ ou $\gamma$ /RXR $\alpha$ , $\beta$ ou $\gamma$ sur un DR1 et deux DR2 (un seul présent ici)	(Ludwig et al., 2000)
Rat Ntcp (Slc10a1)	24777	<u>T</u> GGGGCAT <u>A</u> GGTT <u>T</u> TGG -56 -37	RAR $\alpha$ /RXR $\alpha$ sur un DR2	(Denson et al., 2000)
Human MCAD	34	<u>GGGT</u> <u>TGACT</u> TTC <u>CTCCGGTAA</u> <u>AGGTCA</u> AGG -341 -308	RAR $\alpha$ ou $\beta$ /RXR $\alpha$ sur un ER8. L'élément en vert intervient dans la fixation d'HNF4	(Carter et al., 1994; Raisher et al., 1992)
Mouse TR2 (NR2C1)	22025	<u>GTGGCTCA</u> <u>CGAACCTC</u> GT -32 -15	RAR $\alpha$ /RXR $\beta$ sur un IRO	(Lee and Wei, 1999; Lee and Wei, 2000)
Mouse Laminin B1	16777	<u>GAGGTCA</u> GCT <u>AGGTAA</u> (N <sub>13</sub> ) <u>GGTCA</u> AC -432 -469	RAR $\alpha$ , $\beta$ ou $\gamma$ /RXR $\alpha$ sur un élément complexe	(Vasios et al., 1991)

**Table 4 :** Exemples d'éléments de réponse naturels des récepteurs RAR.

En fonction des séquences, les numérotations prennent pour origine soit le site d'initiation de la transcription, soit le site d'initiation de la traduction. Dans la plupart des cas, ces numérotations correspondent à celles utilisées par les auteurs cités en référence.



**Figure 10 : Interface de dimérisation du LBD des récepteurs nucléaires.**

Les hélices  $\alpha$  intervenant dans les interfaces de dimérisation sont indiquées par des flèches. Les structures de deux hétérodimères (A : RAR/RXR et B : PPAR/RXR) et de deux homodimères (C : RXR/RXR et D : ER/ER) sont présentées. Un ligand est présent au sein de chaque LBD. Un peptide SRC1 est également visible dans la structure B. PPAR $\gamma$ /RXR $\alpha$ .

Identifiants PDB (référence) : A. 1DKF (Bourguet *et al.*, 2000), B. 1FM6 (Gampe *et al.*, 2000), C. 1H9U (Love *et al.*, 2002), D. 1ERE (Brzozowski *et al.*, 1997).

### I.3.2. Dimérisation des récepteurs nucléaires

Pour les récepteurs fonctionnant en dimère, on distingue deux interfaces de dimérisation. La première est qualifiée d'interface faible. Elle est constituée par la boîte D du domaine C (DBD) et n'est observée qu'en présence d'ADN. La deuxième interface de dimérisation des RN est située dans le domaine E de fixation du ligand. Il s'agit là d'une interface forte de dimérisation et des dimères de LBD sont observés en solution. Les premières cristallisations d'hétérodimères de LBD (Bourguet et al., 2000b; Gampe et al., 2000a) ont permis de comparer les interfaces de dimérisation des homo- et des hétérodimères. De manière générale, l'arrangement des deux LBD dans les hétérodimères ressemble fortement à celui observé dans les homodimères (Bourguet et al., 1995; Brzozowski et al., 1997; Nolte et al., 1998). La [Figure 10](#) présente les structures 3D de deux hétérodimères (RAR/RXR et PPAR/RXR) et de deux homodimères (RXR/RXR et ER/ER). Les interfaces de dimérisation impliquent des résidus des hélices H7, H9, H10 et H11 et des boucles L8-9 et L9-10 (Bourguet et al., 2000b). Les hélices H10 des deux récepteurs sont parallèles à l'axe de symétrie reliant les deux récepteurs engagés dans le dimère ([Figure 10 A-D](#)). Au sein des hétérodimères, les deux protomères ne contribuent pas de manière équivalente à l'interface de dimérisation. Par exemple, dans le dimère RAR/RXR ([Fig. 10A](#)) RAR fournit significativement plus de groupements polaires que RXR, alors ce dernier fournit à l'interface la majorité des groupements atomiques chargés positivement. De plus, les deux récepteurs présentent des différences en terme d'aire de contact impliquée dans l'interface. L'hélice H7 de RXR contribue quatre fois plus en terme de surface à l'interface de dimérisation que l'hélice H7 de RAR. En revanche, la contribution de la boucle L8-9 de RAR est trois fois plus importante que la boucle correspondante de RXR. Les parties du LBD de RXR engagées dans l'interface d'hétérodimérisation ([Fig. 10 A-B](#)) sont à peu près les mêmes que celles engagées dans l'interface d'homodimérisation ([Fig. 10C](#)). Dans les deux types de dimères, les hélices H9 et H10 constituent le cœur de l'interface de dimérisation et contribue à plus de 75% au total de cette surface. Cependant, quelques différences existent dans l'implication relative de certains éléments structuraux de RXR aux interfaces des hétérodimères ou de l'homodimère. Ces différences sont liées à la nature du protomère interagissant avec RXR et s'observent même si les résidus de RXR impliqués dans les surfaces d'homodimérisation et d'hétérodimérisation sont exactement les mêmes. Ces différences pourraient engendrer des différences de stabilité des différents dimères et expliquer notamment le fait qu'un mélange de RAR et de RXR conduise à la formation quasi-exclusive d'hétérodimères RAR-RXR. Les éléments

structuraux de ER $\alpha$  impliqués dans son interface d'homodimérisation sont les mêmes que ceux décrits précédemment pour RXR et RAR (Fig. 10D Brzozowski et al., 1997). Cependant, la longueur de certains de ces éléments et la réalisation de contacts additionnels conduit à enfouir une surface plus importante de ER $\alpha$  dans l'interface de dimérisation (environ 1700 Å<sup>2</sup>). L'interface plus réduite de RXR dans les homodimères ou les hétérodimères (environ 1000 Å<sup>2</sup>) suggère un lien plus faible avec le partenaire de dimérisation et pourrait être lié à la capacité de RXR à former des dimères avec de nombreux récepteurs (classe II). La stabilité modérée des dimères engageant RXR pourraient ainsi favoriser le passage entre différents partenaires.

Le fait que les LBDs cristallisés de différents RNs présentent un repliement identique et que les interfaces de dimérisation des différents dimères cristallisés soient également très comparables suggèrent que tous les RNs peuvent établir des surfaces de dimérisation similaires. Ce serait donc des différences en termes d'acides aminés engagés dans ces interfaces qui détermineraient les caractéristiques de la dimérisation.

### I.3.3. Les tétramères de RXR

Enfin, signalons que les travaux de Kersten *et al.* (Kersten et al., 1995a) ont mis en évidence que le récepteur RXR était capable, *in vitro* et en l'absence de ligand, de s'auto-associer sous la forme de tétramères avec une forte affinité (Kd ~3-5 nM pour l'association dimère-dimère). L'ajout du ligand activateur 9cRA (acide 9-cis rétinoïque) conduisant à une rapide dissociation de ces tétramères en dimères et monomères (Kersten et al., 1995b; Kersten et al., 1995c), il a été proposé que les tétramères de RXR étaient transcriptionnellement inactifs et que leur dissociation sous l'effet d'un ligand constituait la première étape vers l'activation de RXR (Kersten et al., 1997). Cette proposition est en accord avec l'activité transcriptionnelle de mutants de RXR présentant des capacités différentes à former des tétramères (Kersten et al., 1998). Ainsi, le mutant RXR $\alpha$  (R321A) forme des tétramères incapables de se dissocier suite à la fixation de 9cRA et présente conjointement un défaut d'induction de son activité transcriptionnelle par le 9cRA (Kersten et al., 1998). A l'opposé, le mutant RXR $\alpha$  (F318A) est incapable de former des tétramères stables et présente une activité transcriptionnelle fortement accrue, y compris en l'absence de ligand (Kersten et al., 1998). Les tétramères de RXR présentent également un rôle fonctionnel dans la régulation du gène CRBPII (Cellular Retinol Binding Protein II Mangelsdorf et al., 1991). Ils régulent

efficacement la transcription de CRBPII via plusieurs éléments de réponse qui sont faiblement reconnus par les dimères de RXR. De plus, la formation coopérative de tétramères de RXR sur le promoteur de CRBPII est nécessaire à l'activation de la transcription de ce gène par le 9cRA (Chen and Privalsky, 1995). Des structures tridimensionnelles de tétramères de RXR en l'absence de ligand ou en présence d'un isomère du 9cRA agissant comme ligand inactif de RXR ont été obtenues (Table 3, Gampe et al., 2000b). Le tétramère se présente sous la forme de deux homodimères associés *via* une interface de tétramérisation. Cette interface implique principalement les hélices H3 et H11 ainsi que l'hélice H12 portant la fonction AF-2. L'interface d'homodimérisation observée au sein de chacun des dimères d'un tétramère ressemble fortement à celle observée dans les structures de dimères de RXR précédemment élucidées (Bourguet et al., 1995). Dans le tétramère, l'hélice H12 prend une conformation inactive très particulière qui permet de comprendre l'absence d'activité transcriptionnelle du tétramère. L'hélice AF-2 de chaque monomère interagit avec le site de fixation des coactivateurs du monomère adjacent situé dans le dimère symétrique. Ainsi, les sites de fixation des coactivateurs des quatre RXR engagés dans le tétramère sont occupés par une hélice AF-2 et ne peuvent donc jouer leur rôle d'interaction avec les cofacteurs nécessaire à l'activité transcriptionnelle des récepteurs (voir paragraphe I.5). Plusieurs caractéristiques de RXR impliquées dans la formation de tétramères semblent conservées chez d'autres récepteurs nucléaires de la sous-famille 2 présentant la capacité de s'homodimériser tels HNF4 ou COUP-TF (Gampe et al., 2000b). Il est donc possible que d'autres récepteurs nucléaires soient également capables de former des tétramères jouant un rôle dans leur activité transcriptionnelle. Cette hypothèse semble renforcée par divers travaux qui suggèrent également la formation de tétramères chez d'autres récepteurs nucléaires (Figueira et al., 2006; Garlatti et al., 1994).

#### I.4. LIGANDS DES RECEPTEURS NUCLEAIRES

Les caractéristiques communes des ligands des RNs sont leur taille relativement réduite et leur caractère hydrophobe. Contrairement aux ligands de nombreux récepteurs membranaires, les ligands des RNs doivent entrer dans la cellule, et parfois dans le noyau pour atteindre leur récepteur cible. Les passages au travers de la membrane plasmique et de l'enveloppe nucléaire sont ainsi facilités par ces deux caractéristiques (petite taille et forte hydrophobicité). En dehors de ces caractéristiques communes, les ligands des RNs présentent des structures très diverses, en particulier les ligands synthétiques. Ces ligands peuvent être

des hormones (hormone thyroïdienne, hormones stéroïdiennes: œstrogènes, progestérone, androgènes, glucocorticoïdes ou minéralocorticoïdes), des vitamines (vitamine D<sub>3</sub>, métabolites de la vitamine A), des lipides (acides gras, cholestérol, phospholipides) ou leurs dérivés (prostaglandines, leucotriènes, acides biliaires) ou encore des xénobiotiques (médicaments, pesticides, contaminants alimentaires ou environnementaux).

**Table 5 : Principaux ligands endogènes des récepteurs nucléaires (début)**

RN	Ligand	Structure	Affinité (Kd)	Références
TR $\alpha$ , $\beta$	Triiodothyronine (T3)		TR $\alpha$ ~0,2-0,3 nM TR $\beta$ ~2 nM	(Nakai et al., 1988; Sap et al., 1986; Thompson et al., 1987; Weinberger et al., 1986)
	Thyroxine (T4)		TR $\alpha$ ~1-5 nM TR $\beta$ ~50 nM	(Nakai et al., 1988; Sap et al., 1986; Thompson et al., 1987; Weinberger et al., 1986)
	TRIAC (acide 3,3',5-triiodothyroacétique)		~1-9 nM	(Evans and Braverman, 1986; Thompson et al., 1987; Weinberger et al., 1986)
RAR $\alpha$ , $\beta$ , $\gamma$	Acide all-trans rétinoïque		RAR ~0,2-0,7 nM ROR $\beta$ ~280 nM	(Allenby et al., 1993; Keidel et al., 1992; Stehlin-Gaon et al., 2003)
RXR $\alpha$ , $\beta$ , $\gamma$	Acide 9-cis rétinoïque		RAR ~0,2-0,7 nM RXR ~10-20 nM	(Allenby et al., 1993)
PPAR $\alpha$	Acide oléique		~2-6 nM	(Gottlicher et al., 1992; Hostetler et al., 2005; Lin et al., 1999; Murakami et al., 1999)
	Acide linoléique		~1-5 nM	(Forman et al., 1997; Gottlicher et al., 1992; Hostetler et al., 2005; Kliewer et al., 1997; Krey et al., 1997; Lin et al., 1999; Murakami et al., 1999)
	Acide $\alpha$ -linolénique		~8 nM	(Forman et al., 1997; Gottlicher et al., 1992; Kliewer et al., 1997; Krey et al., 1997; Lin et al., 1999; Murakami et al., 1999)
	Acide arachidonique		~17-20 nM	(Forman et al., 1997; Gottlicher et al., 1992; Hostetler et al., 2006; Hostetler et al., 2005; Kliewer et al., 1997; Krey et al., 1997; Lin et al., 1999; Murakami et al., 1999)

**Table 5 : Principaux ligands endogènes des récepteurs nucléaires (suite)**

<b>PPAR<math>\alpha</math></b>	Acide 8(S)-hydroxyeicosatétraenoïque (8S-HETE)		EC <sub>50</sub> / IC <sub>50</sub> ~0,1-1 $\mu$ M	(Forman et al., 1997; Kliewer et al., 1997; Krey et al., 1997; Murakami et al., 1999)
<b>LXR<math>\alpha, \beta</math></b>	24(S),25-époxy-cholestérol		~200 nM	(Janowski et al., 1999)
	24(S)-hydroxy-cholestérol (Cérébrostérol)		~100 nM	(Janowski et al., 1999)
	24-déshydro-cholestérol (Desmostérol)		~1 $\mu$ M	(Yang et al., 2006a)
<b>FXR</b>	Acide chénodesoxy-cholique		~3,4-8 $\mu$ M	(Fujino et al., 2003; Makishima et al., 1999; Parks et al., 1999; Wang et al., 1999)
<b>VDR</b>	1 $\alpha$ ,25-dihydroxy-vitamine D <sub>3</sub>		~0,01-0,05 nM	(Baker et al., 1988; Craig and Kumar, 1996; Nakajima et al., 1993; Ross et al., 1991)
<b>PXR</b>	5 $\beta$ -pregnane-3,20-dione		EC <sub>50</sub> ~1-5 $\mu$ M	(Jones et al., 2000; Kliewer et al., 1998; Lehmann et al., 1998; Moore et al., 2000b)
	5 $\beta$ -cholestane-3 $\alpha$ ,7 $\alpha$ ,12 $\alpha$ -triol		mPXR : EC <sub>50</sub> ~2,5-5 $\mu$ M	(Dussault et al., 2003; Goodwin et al., 2003)
<b>CAR</b>	Androsténol (5 $\alpha$ -androst-16-en-3 $\alpha$ -ol)		mCAR : EC <sub>50</sub> ~0,8 $\mu$ M	(Forman et al., 1998; Shan et al., 2004)
	Androstanol (5 $\alpha$ -androstan-3 $\alpha$ -ol)		mCAR : EC <sub>50</sub> ~0,8 $\mu$ M	(Forman et al., 1998; Moore et al., 2000b)
<b>ER<math>\alpha, \beta</math></b>	17 $\beta$ -estradiol		ER $\alpha$ ~0,1 nM ER $\beta$ ~0,4 nM	(Fitzpatrick et al., 1999; Green et al., 1986; Greene et al., 1986; Kuiper et al., 1996)
<b>GR</b>	Cortisol		hGR ~10-20 nM rGR ~40-80 nM hMR ~0,5-1,5 nM	(Giannopoulos and Keichline, 1981; Hellal-Levy et al., 1999; Lind et al., 2000; Lu et al., 2006b; Rupprecht et al., 1993)

**Table 5 :** Principaux ligands endogènes des récepteurs nucléaires (suite)

<b>GR</b>	Corticostéron		rGR~10-20 nM hGR~15-60 nM rMR~0,5 nM	(Giannopoulos and Keichline, 1981; Lu et al., 2006b; Sheppard and Funder, 1987; Yoneda et al., 1995)
<b>MR</b>	Aldostérone		hMR ~0,5-1,5 nM rMR ~0,5 nM hGR ~15-50 nM	(Arriza et al., 1987; Hellal-Levy et al., 1999; Sheppard and Funder, 1987)
<b>PR</b>	Progesterone		~1-1,5 nM	(Birnbaumer et al., 1983; Eul et al., 1989; Gronemeyer et al., 1987)
<b>AR</b>	5α-dihydro-testostérone		~0,2-1 nM	(Chang et al., 1988a; Chang et al., 1988b; Lubahn et al., 1988; Tilley et al., 1989; Trapman et al., 1988)

**Table 5 :** Principaux ligands endogènes des récepteurs nucléaires

Sauf indication contraire, l'affinité est exprimée en termes de constante de dissociation (Kd). Pour une quantité fixe de récepteur, le Kd correspond à la concentration de ligand donnant une occupation de 50% des sites de fixation disponibles. Pour certains récepteurs (PXR, CAR) l'affinité des ligands endogènes n'a pas été mesurée bien que la fixation au récepteur ait été démontrée. Pour ces récepteurs, cette table présente la puissance des molécules via l'EC<sub>50</sub> qui est la concentration nécessaire à l'obtention d'un effet biologique (en général l'induction d'un gène rapporteur contrôlé par un élément de réponse spécifique ou bien le recrutement d'un cofacteur ou encore la fixation sur un élément de réponse) représentant 50% de l'effet maximal observable. Tous les ligands présentés, à l'exception des ligands de CAR agissent comme agonistes. L'androsténol et l'androstanol agissent comme agonistes inverses du récepteur CAR. Les préfixes h, m et r devant les noms des récepteurs se rapportent à l'espèce étudiée (homme, souris et rat respectivement).

#### I.4.1. Ligands endogènes des récepteurs nucléaires

La [Table 5](#) présente les principaux ligands endogènes connus de plusieurs récepteurs nucléaires. Tous les ligands évoqués dans la littérature ne sont naturellement pas présentés dans cette table. En particulier, les ligands jouant un rôle plus structural que fonctionnel ne sont pas présentés. Ainsi, des études structurales ont mis en évidence que les LBD de HNF4α (Dhe-Paganon et al., 2002) et de HNF-4γ (Wisely et al., 2002) étaient occupés par des acides gras saturés ou monoinsaturés (en position Δ9) comportant 16 à 18 carbones et provenant des bactéries ayant servi à surexprimer les récepteurs. Cependant, ces acides gras n'ont pu être

déplacés de leur site de fixation par des acides gras marqués (Dhe-Paganon et al., 2002, Wisely, 2002 #170), suggérant qu'ils ne se comportent pas comme des ligands échangeables, mais plutôt comme des cofacteurs structuraux des HNF-4. De plus, en présence de ces acides gras, l'hélice H12 d'HNF-4 $\alpha$ , se présente sous deux conformations différentes (Dhe-Paganon et al., 2002) correspondant à une forme active (i.e. capable de fixer des coactivateurs, voir paragraphes suivants) et à une forme inactive du LBD (i.e. incapable de fixer un coactivateur). Ces données suggèrent que c'est la fixation des coactivateurs qui pourrait jouer le rôle de signal activateur pour HNF-4, le ligand facilitant mais ne dictant pas une conformation active du récepteur. La structure du complexe ternaire HNF-4 $\alpha$ /acide gras/peptide coactivateur (boîte RN de SRC-1) élucidée ultérieurement (Duda et al., 2004) semble confirmer cette hypothèse car l'hélice H12 y adopte uniquement la conformation active, i.e. repliée contre le LBD et permettant les interactions avec le peptide coactivateur (voir paragraphes suivants). Les récepteurs HNF-4 nouvellement traduits adopteraient donc spontanément une conformation leur permettant d'être activés après avoir capturé un acide gras endogène approprié au sein des cellules. La fixation de coactivateurs stabiliseraient définitivement une conformation pleinement active du récepteur. L'activité, apparemment constitutive d'HNF-4, serait donc vraisemblablement régulée par son niveau d'expression, par des modifications post-traductionnelles (Viollet et al., 1997) et par la formation de complexes avec des cofacteurs (Lee et al., 2000b; Yoon et al., 2001). De manière similaire, plusieurs études structurales des LBD des récepteurs LRH-1 et SF-1 ont mis en évidence la présence inattendue de phospholipides au sein de la poche de fixation du ligand (Krylova et al., 2005; Li et al., 2005a; Ortlund et al., 2005; Wang et al., 2005a). Contrairement à son homologue humain, le LBD du récepteur LRH-1 murin ne présente pas de phospholipide lié à sa poche de fixation du ligand (Sablin et al., 2003). Ceci semble dû à des mutations précises au sein de la forme murine du LBD qui modifieraient la forme de sa poche de fixation du ligand (Wang et al., 2005a), sans pour autant occuper tout le volume de cette dernière qui reste accessible au solvant (Sablin et al., 2003). Ces différentes études ont montré que la fixation de phospholipides ou de phosphatidyl inositols jouait effectivement un rôle dans la fonction des récepteurs de la famille NR5A, et notamment le recrutement de coactivateurs (Krylova et al., 2005; Li et al., 2005a; Ortlund et al., 2005; Wang et al., 2005a). Cependant, il reste encore à déterminer quels phospholipides se fixent sur ces récepteurs et si leur rôle est effectivement celui de ligands échangeables ou bien s'ils jouent un rôle plus structural comme cela semble être le cas pour les acides gras et HNF-4. Récemment, des ligands synthétiques de haute affinité, agonistes pour LRH-1 et SF-1 ont été développés (Whitby et al., 2006). Ils devraient

permettre de mieux délimiter les fonctions et de préciser la capacité de modulation de l'activité des récepteurs NR5A, les faisant peut-être entrer dans la catégorie des cibles thérapeutiques potentielles. Enfin, la structure du LBD du récepteur ROR $\alpha$  a révélé la présence de cholestérol dans la poche de fixation du ligand (Kallen et al., 2002). Le cholestérol sulfate a ultérieurement été identifié comme un meilleur ligand de ROR $\alpha$  (Bitsch et al., 2003; Kallen et al., 2004a). Cependant, le manque de données fonctionnelles ne permet pas actuellement de définir de manière indubitable le ligand de ce récepteur. Notons finalement que des études structurales ont également identifié des récepteurs ne possédant pas de poche de fixation du ligand tels ERR $\gamma$  (Greschik et al., 2002) ou Nurr1 (Wang et al., 2003), définissant ainsi un mode de fonctionnement des récepteurs nucléaires potentiellement indépendant des ligands.

Comme l'illustre la [Table 5](#), on connaît relativement mal à l'heure actuelle les ligands endogènes des RNs. Certes, les études sur les récepteurs aux hormones stéroïdiennes (ER, GR, PR, MR et AR) ou encore sur les récepteurs à l'hormone thyroïdienne ou à la vitamine D3 ont permis d'accumuler de nombreuses connaissances sur ces récepteurs, leurs ligands et les interactions récepteur-ligand. Cependant, pour la plupart des RNs, ou bien on ne connaît pas de ligand endogène, ou bien les ligands endogènes identifiés ne permettent pas de rendre compte de manière exhaustive des activités et des modulations d'activité du récepteur. A partir du bilan des ligands endogènes des RNs présenté en [Table 5](#), deux principales observations peuvent être faites : 1) on distingue deux types de ligands endogènes, ceux présentant une forte affinité pour leur récepteur ( $K_d$  de l'ordre du nM) et ceux présentant une affinité plus faible ( $K_d$  de l'ordre du  $\mu\text{M}$ ) et 2) de nombreux ligands endogènes présentés dans cette table proviennent du métabolisme du cholestérol.

En ce qui concerne l'affinité des ligands endogènes pour leur récepteur deux points principaux méritent d'être développés : 1) les méthodes de mesure des affinités récepteur-ligand et 2) les concentrations physiologiques des ligands endogènes des RNs.

#### I.4.2 méthodes d'analyse des interactions récepteur-ligand

En premier lieu, diverses méthodes de dosage de l'affinité ont été utilisées dans la littérature. Les méthodes classiques de liaison (ou binding) ne peuvent être utilisées que lorsque soit le ligand à étudier (méthode de saturation) soit un ligand de forte affinité (méthode de déplacement) est disponible sous forme radiomarquée. La pureté de la préparation de récepteurs et en particulier la présence d'autres sites de fixation potentiels du

ligand sont alors critiques dans l'évaluation de la constante de dissociation (voir par exemple Sheppard and Funder, 1987). Notons que ces méthodes fonctionnent également mal dans les cas de faibles affinités car la liaison récepteur-ligand risque d'être rompue lors de la séparation des ligands non fixés (Dr Marc Poirot, Institut Claudius Regaud, Toulouse, communication personnelle). Des méthodes plus indirectes ont également vu le jour plus récemment. En particulier, la méthode CARLA (coactivator-dependent receptor ligand assay), développée par l'équipe du Pr W. Wahli (Krey et al., 1997) a permis de cibler de manière efficace et peu coûteuse un grand nombre de ligands potentiels des récepteurs PPAR. Cette méthode constitue une alternative aux méthodes classiques de liaison ligand-récepteur et utilise la capacité du récepteur (ou de son domaine de fixation du ligand) ayant fixé son ligand à interagir avec un coactivateur. Plusieurs développements technologiques ultérieurs de cette méthode ont permis de la rendre encore plus apte à la réalisation de criblages à haut-débit. Les technologies de TR-FRET (Time-resolved fluorescence energy transfer), de TRF (Time-resolved fluorescence), d'ALPHAScreen (amplified luminescent proximity homogeneous assay) ou encore de résonance plasmonique de surface ont ainsi été utilisées pour mesurer les interactions récepteur (ou LBD)-peptide coactivateur suite à l'addition de divers ligands (exemples avec le récepteur FXR: Fujino et al., 2003; Glickman et al., 2002; Parks et al., 1999). Ces méthodes indirectes ont été utilisées pour l'identification et l'étude des ligands de nombreux récepteurs nucléaires de classe II (en particulier PPAR, FXR, LXR ou PXR) alors que des études de liaison directe ont été utilisées pour évaluer les affinités des hormones stéroïdiennes aux récepteurs de classe I. A ma connaissance, aucune étude n'a évalué de manière systématique le lien entre les affinités mesurées par ces deux types de méthode, rendant difficile la comparaison des résultats obtenus *via* ces deux familles de techniques. Signalons toutefois que l'utilisation d'autres méthodes directes de liaison, non radioactives mais basée sur les phénomènes de quenching de fluorescence (méthode de liaison d'un ligand fluorescent, méthode de liaison d'un ligand non fluorescent basé sur le quenching de l'émission des tyrosines ou méthode de déplacement d'un ligand fluorescent) et utilisées en particulier pour les récepteurs PPAR $\alpha$  (Hostetler et al., 2006; Hostetler et al., 2005; Lin et al., 1999) et HNF-4 $\alpha$  (Petrescu et al., 2002) ont permis d'estimer des constantes de dissociation de l'ordre du nM alors que les techniques utilisées précédemment fournissaient des affinités de l'ordre du  $\mu$ M. En conclusion, si les affinités des récepteurs de classe I pour leurs ligands respectifs semblent faire l'objet d'un assez large consensus, des travaux complémentaires semblent en revanche nécessaires pour les membres de la superfamille clonés plus récemment.

### I.4.3. Origines et concentrations physiologiques des ligands endogènes des récepteurs nucléaires

Un deuxième paramètre à prendre en considération pour comparer les affinités des ligands endogènes des récepteurs nucléaires est la concentration physiologique de ces ligands. Ces concentrations physiologiques peuvent naturellement être considérées à plusieurs niveaux d'observation (concentrations circulantes, concentrations tissulaires, concentrations dans certains sous-compartiments cellulaires comme le noyau) et peuvent varier en fonction de nombreux paramètres (rythme circadien, cycle menstruel chez les femelles, alimentation et environnement en général, stress, etc.). Les taux circulants sont naturellement les plus facilement accessibles et donc les mieux connus mais ne reflètent qu'imparfaitement la biodisponibilité du ligand pour son récepteur dans son compartiment biologique spécifique.

De manière générale, les hormones stéroïdiennes circulent à des taux plasmatiques faibles (de l'ordre du ng/mL, sources principales pour les concentrations chez l'Homme : [www.doctissimo.fr](http://www.doctissimo.fr) et [www.medisite.fr](http://www.medisite.fr)). La sécrétion de testostérone, de progestérone et d'estradiol est régulée *via* la sécrétion de GnRH (Gonadotropin-releasing hormone) par des neurones de l'hypothalamus qui induit la sécrétion de LH (Luteinizing hormone) par l'hypophyse antérieure (ainsi que de FSH –follicle stimulating hormone-, une autre hormone gonadotrope secrétée par l'hypophyse antérieure). La testostérone, la progestérone et l'estradiol exercent un rétrocontrôle négatif sur cet axe hypothalamo-hypophysaire conduisant à une sécrétion pulsatile de LH. La prolactine est la troisième hormone gonadotrope secrétée par l'hypophyse antérieure sous l'effet positif du PRF (Prolactin-releasing factor) et de la TRH (Thyrotropin-releasing hormone) et sous l'effet inhibiteur de la dopamine (secrétée par l'hypothalamus). Outre ses effets sur le développement de la glande mammaire et la lactation, la prolactine agit également sur la sécrétion de progestérone chez la femme, et de testostérone chez l'homme. Avant la puberté, les taux de progestérone et d'estradiol sont faibles (<280 pg/mL pour la progestérone et <20 pg/mL pour l'estradiol). Chez l'homme pubère, ils restent relativement faibles (140-560 pg/mL pour la progestérone et 22-49 pg/mL pour l'estradiol). Chez la femme pubère, les taux de progestérone et d'estradiol sont très dépendants du cycle menstruel. En phase pré-ovulatoire (phase folliculaire), les taux de progestérone et d'estradiol sont normalement de 0,1-1,1 ng/mL et 30-120 pg/mL respectivement (pic ovulatoire d'estradiol à 140-400 pg/mL) et en phase post-ovulatoire (phase lutéale) ils sont de 1,5-22,6 ng/mL pour la progestérone et de 65-220 pg/mL pour l'estradiol. Au cours de la grossesse, les taux circulants de progestérone et d'estradiol augmentent progressivement jusqu'à atteindre,

au terme, 150-400 ng/mL et 6,7-30 ng/mL respectivement. Chez la ratte, les ordres de grandeur des taux de progesterone et d'estradiol sont à peu près comparables et varient également en fonction du cycle oestral (entre 1 et 50 ng/mL de progesterone et entre 20 et 110 pg/mL pour l'estradiol Nequin et al., 1975). Les taux circulants de testostérone sont 5 à 10 fois plus faibles chez la femme que chez l'homme et dépendent de l'âge pour les deux sexes. Chez la femme ils sont inférieurs à 0,15 ng/mL avant l'âge de 10 ans, compris entre 0,15 et 0,90 ng/mL entre 10 et 60 ans et entre 0,10 et 0,50 ng/mL au-delà. Chez l'homme, ils sont inférieurs à 0,20 ng/mL avant 10 ans puis augmentent jusqu'à 45-60 ans (0,5-5 ng/mL entre 10 et 20 ans, 3-8,5 ng/mL entre 20 et 45 ans et 2-8 ng/mL entre 45 et 60 ans) avant de redescendre (1-5 ng/mL entre 60 et 75 ans puis 0,15-2,5 ng/mL au-delà). Chez le rat, les taux circulants de testostérone et leurs variations avec l'âge sont assez comparables à ceux de l'homme (0,3-0,4 ng/mL à 1 jour, 3 ng/mL à 100 jours et 1,9 ng/mL à 240 jours Brown-Grant et al., 1975).

Le cortisol est le principal glucocorticoïde chez l'homme. Il est synthétisé par le cortex des surrénales sous l'influence d'une hormone hypophysaire : l'ACTH (Adrenocorticotropic hormone) qui est elle-même secrétée sous l'effet de la CRH (Corticotropin Releasing Hormone) provenant de l'hypothalamus. Chez l'homme, les taux circulants de cortisol varient fortement au cours de la journée : ils sont faibles à minuit (20-70 ng/mL), atteignent un pic le matin vers 6-8h (100-250 ng/mL) et diminuent progressivement jusqu'au soir (70-170 ng/mL à 12h, 60-110 ng/mL à 16h et 40-90 ng/mL à 20h). Chez le rat et la souris, le principal glucocorticoïde est la corticostérone et, comme l'illustre la [Table 5](#), l'affinité du GR des rongeurs pour la corticostérone est meilleure que pour le cortisol (et inversement pour l'homme ou d'autres espèces secrétant du cortisol comme le lapin, Giannopoulos and Keichline, 1981). Les taux plasmatiques de corticostérone chez le rat et la souris sont du même ordre de grandeur que pour le cortisol chez l'homme mais leurs variations circadiennes sont inversées, en lien avec l'activité majoritairement nocturne de ces rongeurs.

L'aldostérone est le principal minéralocorticoïde et est également synthétisé par la glande corticosurrénale. Sa synthèse et sa sécrétion sont principalement sous le contrôle du système rénine-angiotensine, de la concentration extracellulaire en potassium et de l'ACTH (Connell and Davies, 2005). L'aldostérone circule à des taux 100 à 1000 fois inférieurs à ceux des principaux glucocorticoïdes or le récepteur MR possède une affinité quasiment identique pour les glucocorticoïdes et pour l'aldostérone ([Table 5](#)). Pourtant, au niveau rénal, dans les cellules épithéliales de la partie distale du néphron, l'aldostérone active bien le récepteur MR qui participe à son action sur la réabsorption du sodium. C'est l'action de la 11 $\beta$ -

hydroxystéroïde déshydrogénase 2 qui est en fait responsable localement de la transformation des glucocorticoïdes (cortisol et corticostérone) en leurs métabolites inactifs (cortisone et 11-déshydrocorticostérone respectivement), empêchant ainsi l'accès des glucocorticoïdes au récepteur MR (pour une synthèse voir Rashid and Lewis, 2005). Ainsi le métabolisme intracellulaire mais également d'autres paramètres comme la liaison des hormones circulantes à des protéines de transport (par exemple la CBG -corticosteroid-binding globulin- qui lie préférentiellement les glucocorticoïdes parmi les corticostéroïdes) constituent des déterminants fondamentaux de l'exposition des récepteurs à leurs ligands les plus spécifiques. Un exemple similaire est disponible pour le récepteur aux androgènes. La testostérone et la 5 $\alpha$ -dihydrotestostérone circulent dans le sang à des concentrations 10 à 100 fois supérieures aux concentrations requises pour saturer le récepteur AR mais elles sont étroitement liées à des protéines sériques, en particulier la SHBG (sex hormone-binding globulin). De plus, dans de nombreux tissus sensibles aux androgènes, la testostérone est convertie en 5 $\alpha$ -dihydrotestostérone, un androgène plus puissant et ayant une meilleure affinité pour AR, sous l'effet de la 5 $\alpha$ -réductase (pour une synthèse voir Hiipakka and Liao, 1998). Il ressort de ces différents résultats que ce sont autant les taux circulants que les nombreux déterminants de la concentration locale des ligands qui déterminent leur accessibilité à leur récepteur cible.

Nous avons vu que les hormones stéroïdiennes, qu'elles soient d'origine sexuelle ou corticosurrénale circulent à des taux faibles mais largement en rapport avec les affinités de leurs récepteurs cibles ([Table 5](#)). Il en est de même pour les hormones thyroïdiennes, pour la vitamine D3 et pour les métabolites de la vitamine A pour lesquels les récepteurs correspondants disposent également de très fortes affinités pour ces ligands endogènes. Les hormones thyroïdiennes sont synthétisées et secrétées par la thyroïde sous l'effet de la TSH (thyroid-stimulating hormone) produite par l'hypophyse antérieure sous le contrôle positif de la TRH (Thyrotropin-releasing hormone) provenant de l'hypothalamus et le contrôle négatif des hormones thyroïdiennes (rétrocontrôle négatif) ou de la somatostatine (produite par l'hypothalamus mais également l'estomac, les intestins et les cellules delta du pancréas). En dehors de la période périnatale où les taux d'hormones thyroïdiennes sont plus élevés, les concentrations plasmatiques de T4 totale sont de 40-140 ng/mL et celles de T3 totale de 0,5-2 ng/mL. Dans la circulation, T3 et T4 sont majoritairement liées à la TBG (thyroxine-binding globulin), à la transthyrétine et à l'albumine. Seules les T4 (8-28 pg/mL) et T3 libres (2-6 pg/mL) ont une action hormonale. L'affinité des TR pour la T3 est environ 10 fois plus forte que pour la T4 ([Table 5](#)). Des déiodinases (en particulier la déiodinase de type II : D2)

agissent là encore localement au niveau des cellules cibles pour convertir la T4 en T3 et favoriser ainsi l'activation des récepteurs TR (pour une synthèse voir Bianco and Kim, 2006). Le TRIAC ([Table 5](#)) est un métabolite de la T3 qui possède également une forte affinité pour les récepteurs TR. Il a été utilisé en thérapeutique humaine comme hypocholestérolémiant et dans le traitement de cancers de la thyroïde (Tiratricol) mais a été retiré du marché en octobre 2003 par l'AFSSAPS (Agence Française de Sécurité Sanitaire des Produits de Santé).

Les apports alimentaires mais surtout la synthèse au niveau de la peau sous l'action des rayons ultraviolets constituent les principales sources de vitamine D pour l'organisme. La vitamine D existe sous deux formes : la forme D<sub>2</sub> ou ergocalciférol produite par les végétaux et la forme D<sub>3</sub> ou cholécalciférol produite par les animaux. Chez l'homme ces deux formes sont converties en 1,25-dihydroxyvitamine D<sub>3</sub> ([Table 5](#)), leur principal métabolite actif qui circule à des concentrations de 20-50 ng/mL qui semblent cohérentes par rapport à l'affinité de VDR. Ses principaux rôles biologiques, dont VDR est le médiateur, sont la stimulation de l'absorption intestinale du calcium et du phosphore, le remodelage osseux et la conservation des minéraux au niveau rénal. Des rôles dans la modulation du système immunitaire et dans la différenciation cellulaire ont également été décrits. La déficience en vitamine D ou des mutations sur VDR se traduisent par l'apparition du rachitisme, une maladie dont la plupart des symptômes sont reproduits par la délétion de VDR chez la souris (Yoshizawa et al., 1997).

On appelle vitamine A toute substance naturelle présentant des caractéristiques biologiques similaires à celles du rétinol, la principale forme animale de la vitamine A. Le rétinol est une vitamine liposoluble qui présente des propriétés antioxydantes et joue des rôles majeurs notamment dans la vision, le développement et la croissance. En l'absence de synthèse endogène, la vitamine A doit être apportée par l'alimentation, principalement sous la forme de précurseurs d'origine animale sous forme de rétinol ou d'esters de rétinol et d'origine végétale sous formes de carotènes (provitamine A). Chez l'homme, les taux circulants de rétinol se situent environ entre 400 et 800 ng/mL. Le foie constitue le principal site de stockage de la vitamine A. En fonction des besoins, il libère du rétinol dans le plasma sous forme liée à la RBP (retinol-binding protein). Dans l'organisme, la vitamine A existe également sous d'autres formes issues du métabolisme du rétinol : rétinol, rétinal, acide 13-cis rétinoïque, acide all-trans rétinoïque et acide 9-cis rétinoïque. Ce sont ces deux dernières formes qui constituent les principaux ligands et activateurs des récepteurs RXR et RAR ([Table 5](#)). Chez l'homme, l'acide all-trans rétinoïque circule à des taux d'environ 1 à 4 ng/mL (De Leenheer et al., 1982; Eckhoff and Nau, 1990) soit environ 100 à 400 fois plus faibles

que les taux circulants de rétinol (les taux circulants d'acide 13-cis rétinoïque sont comparables à ceux de l'acide all-trans rétinoïque, Eckhoff and Nau, 1990; Tang and Russell, 1990). L'acide 9-cis rétinoïque circule à des taux encore plus faibles, généralement non détectables par les méthodes analytiques usuelles (<0,2-0,5 ng/mL) sauf en cas d'exposition forte d'origine alimentaire (max. 2,5 ng/mL après consommation d'un foie de dinde frit, Arnhold et al., 1996) ou pharmacologique (Lanvers et al., 1996; Weber and Dumont, 1997). Ces taux relativement faibles soulèvent des questions par rapport aux affinités rapportées des récepteurs RAR et RXR ([Table 5](#)) et il a été montré que des mécanismes intracellulaires interviennent, là aussi, pour augmenter les concentrations locales de ligands à proximité des récepteurs. Dans ce domaine, la protéine de transport intracellulaire de l'acide rétinoïque CRABP-II (Cellular Retinoic Acid-Binding Protein II) semble jouer un rôle important (Budhu and Noy, 2002). Après avoir fixé de l'acide rétinoïque, elle effectue une translocation au noyau et interagit de manière transitoire avec RAR pour lui transmettre son ligand (Budhu and Noy, 2002). Certaines protéines impliquées dans le trafic intracellulaire des acides gras (FABP pour Fatty Acid-Binding Proteins) pourraient également jouer des rôles équivalents vis-à-vis des acides gras et des récepteurs PPAR (Tan et al., 2002). Les enzymes intervenant dans le métabolisme du rétinol en rétinaldéhyde (principalement alcool déshydrogénases : ADH) et surtout celles oxydant le rétinaldéhyde en acide rétinoïque (rétinaldéhyde déshydrogénases : RALDH) ainsi que les enzymes intervenant dans la dégradation de l'acide rétinoïque (CYP26A1, B1 et C1, Abu-Abed et al., 2001; Taimi et al., 2004; White et al., 2000) jouent également des rôles clefs pour déterminer le schéma spatio-temporel de concentration locale des activateurs des récepteurs RAR et RXR au cours du développement ou chez l'adulte (pour une synthèse voir Duester et al., 2003; Petkovich, 2001). Il apparaît une fois de plus qu'au-delà des concentrations circulantes des ligands endogènes potentiels ou avérés, de nombreux acteurs (protéines de transport, enzymes du métabolisme) interviennent au niveau intracellulaire dans la régulation fine de la concentration des ligands les rendant plus ou moins biodisponibles pour leur récepteur. Signalons également que des travaux très récents (Ziouzenkova et al., 2007) suggèrent que le rétinaldéhyde, dont les fonctions biologiques n'étaient jusqu'alors connues que dans l'oeil, jouerait un rôle d'inhibiteur de l'adipogénèse, au moins en partie en inhibant la signalisation de RXR et de PPAR $\gamma$ . Cependant, les affinités du rétinaldéhyde rapportées pour ces récepteurs semblent faibles au regard des concentrations circulantes et tissulaires de ce métabolite de la vitamine A (Ziouzenkova et al., 2007) et des travaux complémentaires s'avèrent nécessaire pour élucider définitivement les cibles de ce composé.

De nombreux acides gras, ainsi que des dérivés d'acides gras (leucotriènes et prostaglandines en particulier) ont été décrits comme de bons ligands et activateurs du récepteur PPAR $\alpha$  (Table 5). De fait, les PPAR, avec PXR qui, lui aussi, semble fixer une grande diversité de ligands, présentent les poches de fixation du ligand les plus volumineuses parmi les récepteurs nucléaires dont la structure du LBD a été analysée (env. 1300-1600 Å<sup>3</sup> pour PXR et les trois isoformes de PPAR). Il semble, de plus, que la forme de leur poche de fixation des ligands soit capable de changer assez facilement, s'adaptant ainsi à une variété de structures. Les acides gras pourraient se fixer à PPAR $\alpha$  sous forme libre mais également sous forme estérifiée au coenzyme A (acyl-coenzyme A) comme le suggèrent les travaux de plusieurs équipes (Fan et al., 1998a; Forman et al., 1997; Hostetler et al., 2006; Hostetler et al., 2005; Qi et al., 1999a). Les lipides circulent dans l'organisme principalement sous forme de lipoprotéines (chylomicrons, VLDL, IDL, LDL et HDL) ou complexés à l'albumine. Les acides gras libres ne représentent qu'un faible pourcentage des lipides circulants. Ils entrent dans la cellule par diffusion passive ou via des transporteurs (en particulier FATP –fatty acid transport protein- et CD36/FAT –fatty acid transporter-) et sont rapidement transformés en acyl-CoA via les FATP ou les acyl-CoA synthétases (Jump et al., 2005). Les acides gras libres et les acyl-CoA intracellulaires sont liés aux FABP précédemment évoquées ou aux ACBP (acyl-CoA binding protein) qui les transportent vers les différents compartiments intracellulaires où ils sont métabolisés, ou vers le noyau où ils interagissent avec des facteurs de transcription. Au niveau du noyau, des études de microscopie utilisant des acides gras fluorescents ont permis de mettre en évidence que les concentrations nucléoplasmiques des acides gras à longue chaîne (plus de 16 carbones) et de leurs acyl-CoA étaient de l'ordre de 40-70 nM et 3 nM respectivement (Huang et al., 2004; Huang et al., 2002). Remarquons par ailleurs qu'il a été montré que certains acides gras, en particulier les très longs polyinsaturés (plus de 20 carbones et de 5 double liaisons) étaient capables de modifier de manière significative la composition du pool intracellulaire d'acides gras libres dans des hépatocytes primaires de rat (Pawar and Jump, 2003).

Les oxystérols, dont certains sont considérés comme les ligands endogènes des récepteurs LXR (Table 5), sont des produits d'oxydation du cholestérol (pour une analyse détaillée de la littérature voir Schroepfer, 2000). Ils constituent généralement des intermédiaires ou des produits finaux dans les voies d'excrétion du cholestérol (acides biliaires notamment) et possèdent ainsi, par rapport au cholestérol, une plus grande capacité à traverser les membranes biologiques. L'exposition de l'organisme aux oxystérols provient de l'oxydation des stérols alimentaires, de la synthèse intracellulaire d'oxycholestérol (par auto-

oxydation du cholestérol ou bien via des réactions enzymatiques spécifiques, généralement catalysées par des cytochromes P450 tels CYP7A1, CYP27A1, CYP46 et CYP3A4) ou de l'oxydation de la fraction lipidique des lipoprotéines. Le dosage des oxystérols est particulièrement difficile en raison des nombreux risques de perte ou de génération d'oxystérols aux différentes étapes de stockage, de manipulation et d'analyse des échantillons (Schroepfer, 2000). Ainsi, des variations de plus de 2 ordres de grandeur sur les concentrations de certains oxystérols ont parfois été rapportées (Bjorkhem and Diczfalusy, 2002). Chez l'homme, certains oxystérols ayant potentiellement des rôles biologiques circuleraient à des taux plasmatiques d'environ 0,01 à 0,5 nmol/mL (Schroepfer, 2000), soit très largement en dessous du taux circulant normal de cholestérol (~5000 nmol/mL). Il semble que les trois principaux oxystérols plasmatiques soient le 27-hydroxycholestérol (également appelé 26-hydroxycholestérol et possédant des stéréoisomères appelés 25R,- ou 25S,26-hydroxycholestérol, Javitt, 2002), produit par CYP27A1 (cytochrome P450 mitochondrial), le 24-hydroxycholestérol, produit par CYP46 (exprimé presque exclusivement dans le cerveau chez l'homme) et le  $7\alpha$ -hydroxycholestérol, produit par CYP7A1 (Bjorkhem and Diczfalusy, 2002; Schroepfer, 2000). Chez des adultes sains, ils circulent respectivement à 0,5-0,6 nmol/mL, 0,3-0,4 nmol/mL et 0,1-0,2 nmol/mL de sérum (Setchell et al., 1998). Les concentrations tissulaires des oxystérols sont encore plus mal connues et des résultats extrêmement variables ont été rapportés, hypothéquant pour l'instant des conclusions fiables sur le sujet (Schroepfer, 2000). Bien qu'il soit concevable, comme illustré précédemment, que divers mécanismes moléculaires soient mis en jeu pour provoquer localement des concentrations fortes d'oxystérols, les données actuellement disponibles et les affinités rapportées ([Table 5](#)) ne permettent pas de conclure de manière définitive quant à leur rôle de véritables ligands endogènes des LXR. Notons que le desmostérol (24-déhydrocholestérol, [Table 5](#)), un précurseur dans la biosynthèse du cholestérol, se fixe également sur les LXR, facilite le recrutement de cofacteurs par ces RNs et régule leurs gènes cibles (Yang et al., 2006a). Bien que présentant une affinité plus faible que certains oxystérols pour les LXR, le desmostérol circule dans l'organisme à des concentrations plus élevées, de l'ordre de 0,8 à 1,4 µg/mL (soit 2,1-3,6 nmol/mL) de sérum (Miettinen et al., 1997; Tammi et al., 2001) et pourrait donc agir comme ligand endogène des LXRs. Enfin, l'acide cholesténoïque (acide  $3\beta$ -hydroxy-5-cholesténoïque, non présenté en [Table 5](#)), un acide biliaire produit de l'oxydation du 27-hydroxycholestérol par CYP27A1, a également été décrit comme un ligand endogène potentiel de LXR (Song and Liao, 2000). Il circule sous forme non estérifiée et non conjuguée à environ 0,3-0,5 nmol/mL (Song and Liao, 2000).

Certains acides biliaires et en particulier l'acide chénodésoxycholique (Table 5), l'acide désoxycholique et l'acide lithocholique sont considérés comme des ligands endogènes du récepteur FXR. L'acide cholique et l'acide chénodésoxycholique (CDCA) sont les acides biliaires primaires, synthétisés par le foie à partir du cholestérol et via divers oxystérols. Ils sont conjugués avec la glycine ou la taurine (un dérivé de la cystéine) pour former les sels biliaires primaires (glycocholate, taurocholate, glycochénate et taurochénate) qui sont excrétés du foie vers la vésicule biliaire via les canaux biliaires. La bile hépatique, produite constamment par le foie est concentrée environ dix fois dans la vésicule biliaire et excrétée rapidement dans le duodénum via le canal cholédoque lorsque la vésicule biliaire se contracte sous l'effet de la cholécystokinine. Dans le duodénum et le jéjunum ils émulsifient les graisses et agissent ainsi comme cofacteurs indispensables à l'action de la lipase pancréatique intervenant dans la digestion des graisses. Dans l'iléon, les sels biliaires primaires sont déconjugués sous l'action de la flore intestinale. L'acide cholique et le CDCA sont partiellement transformés par cette flore en acide désoxycholique et acide lithocholique respectivement (acides biliaires secondaires). Les acides biliaires primaires et secondaires sont majoritairement réabsorbés dans l'iléon (environ 97%, les 3% restant étant éliminés par les fèces) et transportés au foie via la veine porte. Une petite partie de ces acides biliaires traversent le foie et sont éliminés via les urines. Tous les acides biliaires reçus par le foie sont reconjugués comme les acides biliaires primaires, sauf l'acide lithocholique qui est sulfo-conjugué ou oxydé en acide ursodésoxycholique (acide bilaire tertiaire, assez rare chez l'homme mais utilisé pour le traitement des calculs biliaires). Les sels biliaires primaires et secondaires sont excrétés à nouveau dans la bile, bouclant ainsi le cycle entéro-hépatique des acides biliaires. Le sulfolithocholate est systématiquement excrété et n'est ni réabsorbé ni déconjugué. Notons que, chez les rongeurs, d'autres acides biliaires sont également observés comme les acides muricholiques ( $\alpha$ ,  $\beta$  et  $\omega$ ) ou l'acide hyodésoxycholique. Suite à cette description, il n'est pas surprenant de constater que les principaux sites d'expression et d'action de FXR décrits sont le foie, l'intestin (iléon plus particulièrement) et les reins où l'on retrouve la présence des acides biliaires. Les travaux de plusieurs équipes ont mis en évidence que les acides biliaires mais également leurs conjugués (sels biliaires) qui constituent chez l'homme environ 98% des acides biliaires de la bile étaient de bons ligands et activateurs de FXR (Fujino et al., 2003; Makishima et al., 1999; Parks et al., 1999; Wang et al., 1999). Dans certains modèles cellulaires, il est néanmoins nécessaire de surexprimer un transporteur spécifique des sels biliaires afin d'observer, via un gène rapporteur, l'activation de FXR par les conjugués des acides biliaires (Parks et al., 1999; Wang et al., 1999). Le pool total

d'acides biliaire dans l'organisme est assez important (2-4 g) et est maintenu à un niveau relativement constant via la néosynthèse hépatique quotidienne de 200 à 500 mg d'acides biliaires primaires. Chez l'homme, les acides biliaires présents dans le foie représenteraient seulement 1 à 2% du pool total des acides biliaires ( $61,6 \pm 29,7$  nmol/g de foie, Setchell et al., 1997). Les acides choliques et chénodésoxycholique sont les plus représentés dans le foie comme dans la bile extraite de la vésicule biliaire et représentent chacun environ 30 à 40% des acides biliaires totaux (Miettinen et al., 1997; Setchell et al., 1997). L'acide désoxycholique représente environ 15% des acides biliaires (foie et vésicule biliaire) et les acides lithocholique et ursodésoxycholique entre 0,5 et 5%. Des dosages effectués sur noyaux isolés d'hépatocytes de rat suggèrent que moins de 0,1% des acides biliaires hépatiques sont retrouvés au niveau du noyau (Setchell et al., 1997). Les acides biliaires sont également retrouvés au niveau sérique à des concentrations d'environ 2-5 nmol/mL (~1-2 µg/mL) chez l'homme (Angelin et al., 1978; Azer et al., 1997; Smith et al., 2004) et d'environ 6-15 nmol/mL (~3-6 µg/mL) chez le rat (Ando et al., 2006). Alors que le CDCA prédomine chez l'homme au niveau sérique (environ 40 à 50%), c'est l'acide cholique (environ 50%) puis les acides hydrosésoxycholique et  $\beta$ -muricholique (environ 15% chacun) qui prédominent chez le rat. L'ordre de grandeur de ces concentrations d'acides biliaires, en particulier comparativement aux hormones et vitamines évoquées précédemment, semble plutôt bien en accord avec la moindre affinité de FXR pour ses ligands endogènes par rapport aux récepteurs aux hormones stéroïdiennes ou thyroïdiennes.

Ce tour d'horizon des concentrations physiologiques des ligands endogènes des récepteurs nucléaires permet de dégager quelques points importants à souligner :

- Il existe dans une certaine mesure un accord assez étroit entre les concentrations physiologiques des ligands endogènes décrits et les affinités de leurs récepteurs respectifs. Certains récepteurs agissent comme senseurs très spécifiques d'hormones produites en faible quantité et dans un organe parfois éloigné, comme les récepteurs aux hormones stéroïdiennes ou thyroïdiennes. D'autres semblent plutôt agir comme senseurs assez généraux de certaines grandes voies du métabolisme et présentent une moindre spécificité de ligand et/ou des affinités plus faibles comme les récepteurs PPAR, PXR, LXR ou FXR,
- Les concentrations circulantes sont de loin les mieux décrites dans la littérature mais ne sont pas toujours les plus pertinentes à prendre en considération. Les concentrations tissulaires, intracellulaires ou intranucléaires semblent plus adaptées dans le contexte des récepteurs nucléaires mais les données en la matière sont assez fragmentaires, voire totalement absentes. De nombreux travaux semblent encore nécessaires pour s'assurer que les

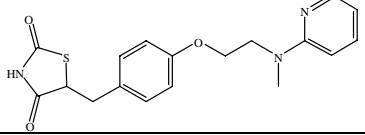
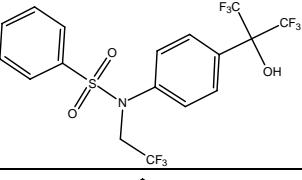
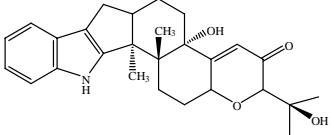
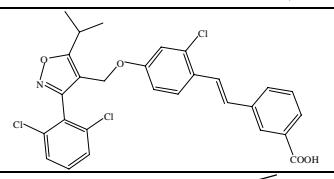
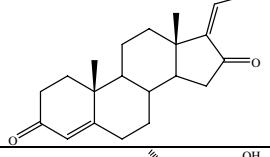
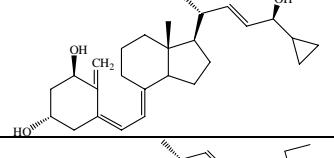
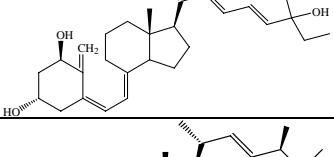
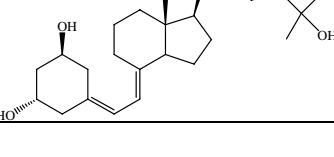
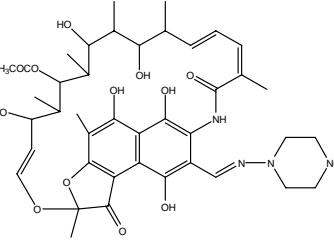
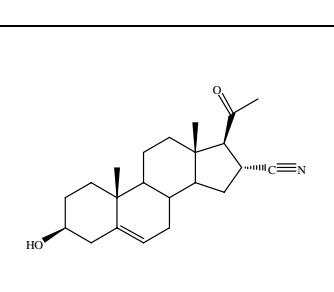
ligands endogènes décrits ([Table 5](#) mais aussi tous ceux qui ne sont pas mentionnés dans ce manuscrit) exercent effectivement leurs effets via les RNs dans des conditions physiologiques.

- Dans plusieurs cas, des mécanismes moléculaires liés au transport, au trafic intracellulaire ou au métabolisme des ligands permettent de générer des zones de plus forte concentration des ligands endogènes au niveau tissulaire, cellulaire ou subcellulaire, favorisant ainsi leur interaction avec les RNs et participant également à une régulation spatio-temporelle très fine de leur activité.

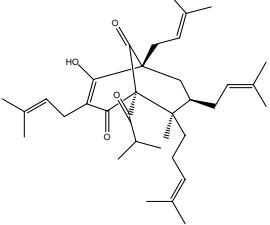
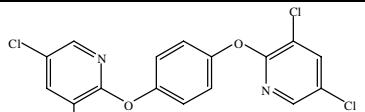
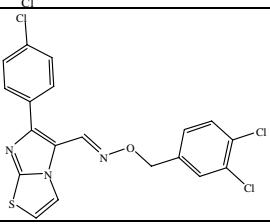
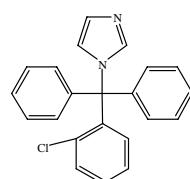
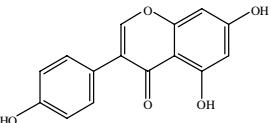
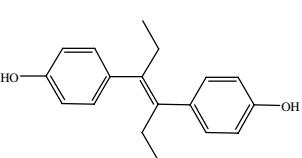
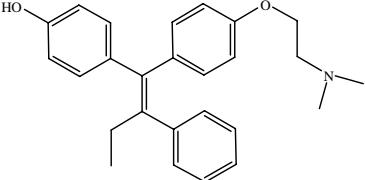
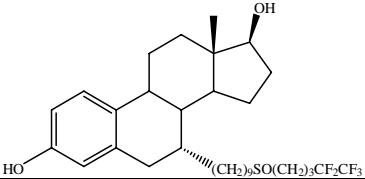
**Table 6 :** Exemples de ligands non endogènes des RN (début)

RN	Ligand	Structure	Remarques	Références
<b>RAR<math>\alpha</math>, <math>\beta, \gamma</math></b>	TTNPB		Kd~2-40nM Pan-agoniste	(Beard et al., 1995; Nagpal et al., 1995; Pignatello et al., 1997)
	Acide Tazaroténique (AGN190299), métabolite actif du Tazarotène (AGN190168)		Kd~40-60 nM ( $\beta, \gamma$ ) Kd~600 nM ( $\alpha$ ) Agoniste $\beta/\gamma$ sélectif Traitement topique de l'acné, du psoriasis et de dommages photosensibles	(Chandraratna, 1996; Nagpal et al., 1995)
<b>RXR<math>\alpha</math>, <math>\beta, \gamma</math></b>	LGD1069 (Bexarotène, Targretin)		Kd~15-30 nM Pan-agoniste Traitement de lymphomes cutanés à cellules T	(Boehm et al., 1994; Zhang et al., 2002a)
	LG100268		Kd~3nM Pan-agoniste	(Boehm et al., 1995; Love et al., 2002)
<b>PPAR <math>\alpha</math></b>	Bezafibrate		Kd~2-10 nM Agoniste double $\alpha/\beta$ Traitement d'hypertriglycéridémies	(Hostetler et al., 2005; Krey et al., 1997; Peters et al., 2003)
	Acide pirinixique (Wy 14,643)		Kd~20-30 nM Agoniste	(Forman et al., 1997; Krey et al., 1997; Lin et al., 1999; Murakami et al., 1999)
	Acide fenofibrate (métabolite du fénofibrate)		Agoniste Traitement d'hypertriglycéridémies	(Krey et al., 1997; Mukherjee et al., 2002; Thomas et al., 2003)
	GW6471		Antagoniste synthétique	(Xu et al., 2002a)

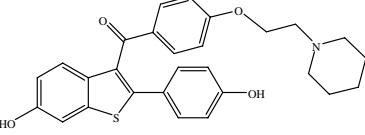
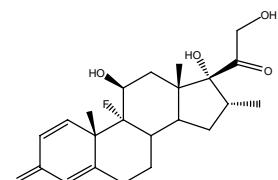
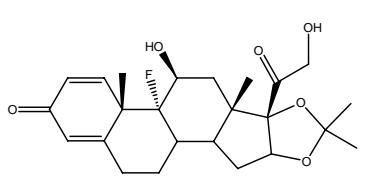
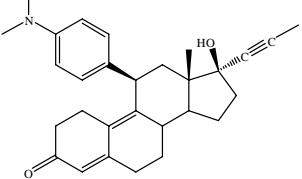
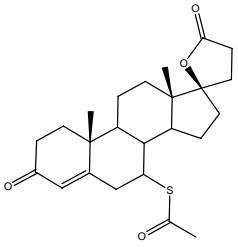
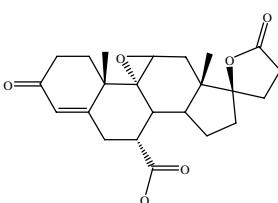
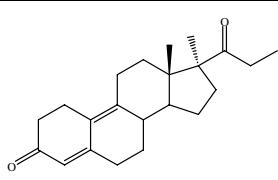
**Table 6 :** Exemples de ligands non endogènes des RN (suite)

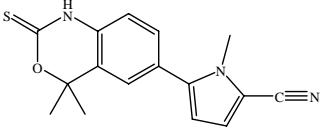
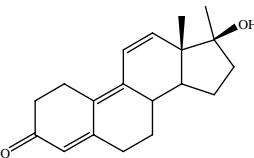
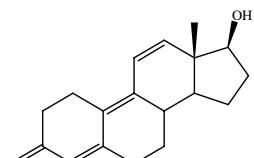
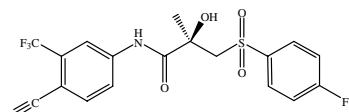
<b>PPAR γ</b>	Rosiglitazone		Kd ~40nM Agoniste Traitement du diabète de type II	(Lehmann et al., 1995)
<b>LXRx, β</b>	T0901317		EC50~20nM Agoniste	(Schultz et al., 2000)
	Paxilline		Ki (25-OH-Chol)~0,6-1,2 μM Agoniste Mycotoxine produite par <i>Penicillium paxilli</i>	(Bramlett et al., 2003)
<b>FXR</b>	GW4064		EC50~40 nM Agoniste	(Maloney et al., 2000)
	Guggulstérolone (Z-guggulstérolone présentée ici)		Ki~μM (ordre de grandeur) Antagoniste Traitement d'hyperlipidémies	(Burris et al., 2005; Urizar et al., 2002; Wu et al., 2002)
<b>VDR</b>	Calcipotriol (MC903)		Agoniste Traitement topique du psoriasis	(Tocchini-Valentini et al., 2004)
	Seocalcitol (EB1089)		Agoniste Anticancéreux en développement	(Tocchini-Valentini et al., 2004)
	Paricalcitol (Zemplar)		Kd~0,15nM Agoniste Traitement de l'hyperparathyroïdie secondaire	(Brown et al., 2002)
<b>PXR</b>	Rifampicine		IC50([3H]SR12813)~6-8μM Agoniste de hPXR (pas mPXR) Antibiotique	(Bertilsson et al., 1998; Blumberg et al., 1998; Chrencik et al., 2005; Jones et al., 2000; Lehmann et al., 1998; Moore et al., 2000b; Zhu et al., 2004)
	Pregnenolone 16α-carbonitrile (PCN)		Agoniste de mPXR (faible pour hPXR)	(Blumberg et al., 1998; Chrencik et al., 2005; Jones et al., 2000; Kliwer et al., 1998; Lehmann et al., 1998; Moore et al., 2000b)

**Table 6 :** Exemples de ligands non endogènes des RN (suite)

PXR	Hyperforine		IC <sub>50</sub> ([ <sup>3</sup> H]SR12813)~30 nM Agoniste Composant extrait du millepertuis ( <i>Hypericum perforatum</i> ) utilisé comme antidépresseur	(Moore et al., 2000a; Watkins et al., 2003b; Zhu et al., 2004)
CAR	TCPOBOP		Agoniste de mCAR (pas hCAR). Agoniste faible de hPXR (pas de mPXR)	(Moore et al., 2000b; Tzameli et al., 2000)
	CITCO		EC <sub>50</sub> (FRET hCAR/SRC1)~50nM Agoniste de hCAR (pas mCAR)	(Maglich et al., 2003)
	Clotrimazole (Trimysten)		IC <sub>50</sub> (FRET hCAR/SRC1)~60nM Antagoniste de hCAR (pas de mCAR), agoniste de PXR Antifongique	(Bertilsson et al., 1998; Jones et al., 2000; Lehmann et al., 1998; Maglich et al., 2003; Moore et al., 2000b; Zhu et al., 2004)
ER $\alpha$ , $\beta$	Génistéine		Ki~0,3-3nM Agoniste Isoflavone issue du soja	(Henley and Korach, 2006; Kuiper et al., 1997; Kuiper et al., 1998; Matthews et al., 2000; Mueller et al., 2004)
	Diethylstilbestrol		Ki~0,05nM Agoniste Traitement de cancers de la prostate métastasés. Interdit depuis les années 70 pour la prévention des avortements et accouchements prématurés	(Henley and Korach, 2006; Kuiper et al., 1997; Kuiper et al., 1998; Matthews et al., 2000; Mueller et al., 2004; Swan, 2000)
	4-hydroxytamoxifène		Ki~0,05-0,1nM Antagoniste partiel / modulateur sélectif Métabolite du tamoxifène (traitement de cancers du sein)	(Kuiper et al., 1997; Kuiper et al., 1998; Matthews et al., 2000; Wijayaratne et al., 1999)
	ICI 182780 (Fulvestran, Faslodex)		Ki~0,05-0,5nM Antagoniste Traitement de cancers du sein	(Kuiper et al., 1998; Wakeling et al., 1991; Wijayaratne et al., 1999)

**Table 6 :** Exemples de ligands non endogènes des RN (suite)

<b>ER<math>\alpha</math>, <math>\beta</math></b>	Raloxifène (LY139481)		Ki~0,1-30 nM Antagoniste partiel / modulateur sélectif Traitement et prévention de l'ostéoporose chez la femme ménopausée	(Kuiper et al., 1998; Wijayaratne et al., 1999)
<b>GR</b>	Dexamethasone		Kd~1-5nM (Forte affinité également pour MR) Agoniste fort de GR et agoniste faible de MR Anti-inflammatoire (et immunosupresseur) très utilisé en médecine humaine et vétérinaire	(Bourgeois et al., 1984; Hellal-Levy et al., 1999; Roux et al., 1996; Rupprecht et al., 1993; Simons et al., 1989)
	Triamcinolone acetonide		Kd~0,5-3nM Agoniste Anti-inflammatoire utilisé en dermatologie, ophtalmologie, rhumatologie et contre les manifestations des rhinites allergiques	(Lind et al., 1996; Lind et al., 2000; Lu et al., 2006b; Zhang et al., 1996b)
	Mifepristone (RU 486)		Kd~0,3-3nM (Forte affinité pour PR également) Antagoniste de GR, PR et AR Abortif (avortement chimique du début de grossesse).	(Agarwal, 1996; Bourgeois et al., 1984; Hurd and Moudgil, 1988; Lu et al., 2006b; Robin-Jagerschmidt et al., 2000; Rupprecht et al., 1993)
<b>MR</b>	Spironolactone		Kd~3nM (Forte affinité pour AR également) Antagoniste de MR, AR et GR, agoniste de PR Traitement de l'hypertension artérielle, de l'insuffisance cardiaque, de l'hirsutisme (antiandrogène) ; diurétique	(Corvol et al., 1975; Couette et al., 1992a; Couette et al., 1992b; Rogerson et al., 2003)
	Eplerenone		Ki~30-300nM Antagoniste de MR (pas d'interactions avec GR et PR et très faible avec AR) Traitement de l'hypertension artérielle, de l'insuffisance cardiaque, de dysfonction ventriculaire gauche ; diurétique	(Brown, 2003; de Gasparo et al., 1987; Garthwaite and McMahon, 2004; Lu et al., 2006b; Rogerson et al., 2004)
<b>PR</b>	Promegestone (R5020)		Kd~0,2-5nM Agoniste Traitement de l'insuffisance lutéale (préménopause principalement)	(Birnbaumer et al., 1983; Fitzpatrick et al., 1999; Hurd and Moudgil, 1988)

<b>PR</b>	Tanaproget		$IC_{50}$ ( $[^3H]R5020$ )~2nM Agoniste Progénétif (essais cliniques en cours)	(Zhang et al., 2005b)
<b>AR</b>	Metribolone (Methyltrienolone, RU1881)		Kd~0,3-1nM Forte affinité pour PR et MR également Agoniste de AR, antagoniste de MR	(Chang et al., 1988b; Freyberger and Ahr, 2004; Kuiper et al., 1993; Ris-Stalpers et al., 1990; Takeda et al., 2007; Toth et al., 1995)
	17 $\beta$ -trenbolone		Ki (r,hAR)~1-5nM Ki (bPR)~1-5nM Agoniste Métabolite actif du trenbolone acetate utilisé comme anabolisant chez le bovin	(Bauer et al., 2000; Wilson et al., 2002)
	Bicalutamide (Casodex)		Ki~10 nM (R-bicalutamide) Antagoniste Traitement de cancers de la prostate	(Fuhrmann et al., 1992; Furr and Tucker, 1996; He et al., 2002; Mukherjee et al., 1996)

**Table 6 :** Exemples de ligands non endogènes (synthétiques et naturels) des récepteurs nucléaires

Lorsqu'elles sont disponibles dans les références, des indications sur l'affinité des ligands sont fournies (Kd, Ki, EC<sub>50</sub> ou IC<sub>50</sub>). Ces affinités sont en général à concevoir en termes d'ordre de grandeur. La fonction d'agoniste, d'antagoniste ou de modulateur spécifique du ligand vis-à-vis de son (ou ses) récepteur(s) cible(s) est précisée. Des informations sont également fournies sur l'origine des ligands naturels ainsi que sur l'utilisation éventuelle en thérapeutique et/ou les propriétés rapportées des différents ligands. Les préfixes m-, r-, h- et b- pour les récepteurs nucléaires correspondent aux espèces étudiées (m : souris, r : rat, h : homme et b : bovin).

#### I.4.4. Ligands synthétiques ou naturels des récepteurs nucléaires

Mis à part ces ligands endogènes, de très nombreux ligands synthétiques ont été développés pour les RNs et plusieurs de ces ligands sont utilisés en thérapeutique humaine pour le traitement de pathologies très diverses comme des cancers, des hypertriglycéridémies ou encore le diabète de type II. Ces ligands synthétiques s'avèrent également extrêmement utiles dans les études plus fondamentales où ils permettent, de part leurs spécificités, leurs activités et leurs affinités une dissection fine du fonctionnement des voies de signalisation des RNs. Nous nous concentrerons dans ce chapitre sur les RNs des classes I et II pour lesquels un nombre plus important de ligands sont connus. Bien entendu, ce manuscrit ne peut

constituer un catalogue exhaustif de tous les ligands synthétiques décrits des RNs. Néanmoins, la présentation de quelques uns de ces ligands synthétiques ([Table 6](#)) qui sont utilisés en thérapeutique humaine, ont été utilisés dans le cadre de ce travail de doctorat ou font l'objet de nombreuses études dans la littérature, nous permet de percevoir la diversité des structures des ligands des RNs et d'entrevoir le large potentiel de ces récepteurs pour le développement de futurs médicaments. Plusieurs xénobiotiques non synthétiques (« naturels ») comme des composés extraits de végétaux ou de moisissures ont également été identifiés comme ligands des RNs et certains sont également présentés en [Table 6](#). L'utilisation de certains de ces ligands, synthétiques ou naturels, pour le traitement de pathologies variées va nous permettre d'évoquer succinctement ci-dessous les principaux rôles joués par ces différents récepteurs dans la physiologie des organismes.

L'industrie de la dermatologie est sans doute l'une de celles qui s'intéresse le plus aux récepteurs RAR et RXR. En effet, ces récepteurs jouent des rôles très importants dans les phénomènes de prolifération et différenciation de manière générale et sont bien exprimés au niveau de la peau. Plusieurs agonistes synthétiques sont utilisés en thérapeutique pour le traitement de divers problèmes dermatologiques comme l'acné, le psoriasis ou certains types de cancers cutanés ([Table 6](#)). Le fait que RXR soit le partenaire de dimérisation de nombreux RNs suggère que sa modulation pourrait également être envisagée dans des pathologies impliquant ses partenaires. La première étude menée dans le cadre de ce doctorat a justement visé à étudier la capacité de ligands de RXR à moduler de manière tissu-spécifique les voies de signalisation de ses partenaires. Au-delà de leurs fonctions dans la peau, les récepteurs RXR et RAR jouent de nombreux autres rôles dans l'organisme, notamment au niveau du développement embryonnaire, du métabolisme et de la balance prolifération/différenciation.

Les récepteurs PPAR jouent des rôles majeurs dans le métabolisme énergétique et en particulier dans le métabolisme des lipides. De manière très schématique, l'activation du récepteur PPAR $\alpha$  favorise le catabolisme et l'utilisation des lipides alors que l'activation du récepteur PPAR $\gamma$ , impliqué notamment dans la différenciation adipocytaire, favorise le stockage des graisses par l'organisme. Les fibrates (ciprofibrate, gemfibrozil, bezafibrate et fenofibrate) sont des activateurs du récepteur PPAR $\alpha$  utilisés depuis les années 1970 pour le traitement de divers types d'hypertriglycéridémies. Les thiazolidinediones, comme la rosiglitazone, activateurs de PPAR $\gamma$ , sont quant à elles utilisées pour rétablir une sensibilité des organes périphériques à l'insuline dans le cadre de traitements du diabète de type II.

Les agonistes de VDR sont utilisés dans diverses pathologies comme le psoriasis, certains cancers ou l'hyperparathyroïdie secondaire à une insuffisance rénale. Ces applications sont en cohérence avec l'expression forte de VDR au niveau des reins, de l'intestin, du tissu osseux et de la parathyroïde ainsi qu'avec ses rôles de régulateur majeur de l'homéostasie du calcium et du phosphate (avec la parathormone qui est régulée négativement par VDR) et de la balance prolifération/différenciation.

Le récepteur PXR fait l'objet d'actives recherches en raison de son implication potentielle dans des interactions médicamenteuses. Il constitue, avec le récepteur CAR, un régulateur puissant du métabolisme des xénobiotiques et en particulier du cytochrome P450 3A (CYP3A) qui est responsable du métabolisme de plus de 60% des médicaments actuellement utilisés (Lehmann et al., 1998). Le cas de l'hyperforine, un composé issu du millepertuis (ou herbe de Saint-Jean) est assez caractéristique (Moore et al., 2000a). Le millepertuis est assez largement utilisé pour ses propriétés d'antidépresseur naturel mais des études ont mis en évidence qu'il augmentait le métabolisme de divers médicaments comme l'indinavir (inhibiteur de protéase utilisé dans le traitement du HIV), la cyclosporine (immunosuppresseur utilisé par exemple lors de greffes) où des contraceptifs oraux. L.B. Moore et collaborateurs ont pu mettre en évidence que l'hyperforine est en fait un ligand du récepteur PXR et induit l'expression du CYP3A via l'activation de PXR (Moore et al., 2000a), conduisant ainsi à un accroissement du métabolisme hépatique de nombreux médicaments.

Les ligands synthétiques des récepteur ER sont majoritairement utilisés dans le traitement de cancers du sein. Bien que les mécanismes moléculaires soient encore mal connus, les estrogènes semblent constituer un facteur déterminant dans le développement de cancers du sein. Cette observation a conduit à la mise en œuvre de stratégies visant à inhiber l'activité des ER via l'utilisation d'antagonistes comme le tamoxifène ou l'ICI 182780 ([Table 6](#)) dans les tumeurs mammaires où l'activation des ER se traduit par une croissance tumorale. Néanmoins, les rôles des estrogènes dans l'organisme sont très nombreux (régulation du cycle menstruel et de la reproduction, modulation de la densité des os ou transport du cholestérol par exemple) et il est connu que la réduction des concentrations des stéroïdes sexuels suite à la ménopause conduit à de multiples désagréments (sueurs nocturnes, bouffées de chaleur) et à des risques accrus d'ostéoporose (diminution de la densité des os favorisant les fractures) et de pathologies cardiaques (Jordan, 2004). Le recours a des thérapies de substitution hormonale post-ménopause semble efficace pour la prévention de cancers du colon, de l'ostéoporose et des symptômes de la ménopause mais accroît par ailleurs les risques de

cancers du sein, de maladie d’Alzheimer, de formations de caillots sanguins (phlébite, embolies pulmonaires) et d’infarctus du myocarde (Jordan, 2004). Ces observations ont conduit de nombreuses équipes à travailler au développement de **modulateurs spécifiques des récepteurs aux estrogènes** (SERM pour Specific Estrogen Receptor Modulators) qui présenteraient des activités agonistes et antagonistes spécifiques en fonction des tissus. Ces recherches ont par ailleurs été stimulées par le clonage du récepteur ER $\beta$  (Kuiper et al., 1996; Tremblay et al., 1997) qui a ouvert la voie au développement de ligands présentant des affinités différentielles pour ces deux récepteurs dont l’expression tissulaire n’est pas identique (Kuiper et al., 1997). Le tamoxifène ([Table 6](#)) est un exemple de SERM car il exerce une activité antiestrogénique au niveau du sein mais présente également une activité partiellement estrogénique dans ce tissu ainsi qu’au niveau de l’utérus et des os. Ces activités d’agoniste partiel conduisent vraisemblablement à une inhibition sub-optimale des ER au niveau des cancers du sein, à un accroissement de l’incidence de cancers de l’endomètre mais, en revanche, à un effet positif sur le maintien de la densité osseuse (Jordan, 2004). Le raloxifène ([Table 6](#)) est un autre exemple de SERM, indiqué, lui, pour la prévention et le traitement de l’ostéoporose (effet estrogénique) et dont l’utilisation s’avère également bénéfique pour la prévention de cancers du sein (effet antiestrogénique) et de l’endomètre (effet antiestrogénique). Enfin, les récepteurs aux estrogènes (avec le récepteur aux androgènes) font également l’objet de nombreuses recherches en raison de leur implication potentielle dans la diminution de la fertilité et l’augmentation de l’incidence de malformations du tractus uro-génital et de cancers des organes liés à la reproduction observés principalement aux Etats-Unis et en Europe (Maffini et al., 2006; Mauduit et al., 2006; Swan, 2000; Swan, 2006). Des composés naturels ou issus de l’industrie pourraient ainsi agir via les voies de signalisation des hormones stéroïdiennes et/ou moduler la synthèse, le métabolisme, le transport ou l’activité de ces hormones et agir ainsi comme **perturbateurs endocriniens**.

Les glucocorticoïdes endogènes, essentiellement régulés en fonction du rythme circadien et du stress, agissent principalement via le récepteur GR mais également via le récepteur MR dans les tissus où celui-ci n’est pas protégé par les voies de dégradation des glucocorticoïdes (voir chapitre sur les ligands endogènes et Funder, 1992). Ils jouent des rôles vitaux et régulent de nombreux processus biologiques dont la croissance, le développement, le métabolisme, le comportement ou l’apoptose. Leurs rôles en situation de stress sont multiples (pour une synthèse voir Sapolsky et al., 2000). En thérapeutique, les glucocorticoïdes sont parmi les médicaments les plus couramment prescrits. Ils sont généralement utilisés pour leurs propriétés anti-inflammatoire et immunsuppressive ([Table 6](#)) dans une grande variété de

pathologies comme l'asthme, les dermatites, la polyarthrite rhumatoïde, la prévention des rejets lors de greffes ou les maladies auto-immunes. Le RU 486 (Mifepristone, [Table 6](#)) est utilisé en thérapeutique humaine pour son action antagoniste de PR (abortif) mais constitue aussi l'antagoniste le mieux caractérisé de GR (Agarwal, 1996).

Outre son rôle pour véhiculer les effets des glucocorticoïdes dans certains tissus, le récepteur MR, activé par l'aldostérone, joue des rôles majeurs dans la réabsorption du sodium et de l'eau et dans l'excrétion du potassium soit, plus généralement, dans le contrôle de l'homéostasie des électrolytes au niveau des cellules épithéliales du colon et de la partie distale des néphrons (Connell and Davies, 2005; Fuller and Young, 2005; Takeda, 2004). De plus, il semble aujourd'hui évident que l'aldostérone exerce, via MR et sans doute via d'autres voies de signalisation encore mal caractérisées, une somme d'actions diverses au niveau de tissus non épithéliaux comme les cardiomycocytes, l'hippocampe, la paroi des vaisseaux (cellules musculaires lisses et cellules endothéliales), les monocytes circulants et le système nerveux central (Connell and Davies, 2005; Fuller and Young, 2005). Ciblant initialement son action sur le transport des électrolytes, les antagonistes de MR (spironolactone et eplerenone, [Table 6](#)) sont utilisés dans le traitement de l'hypertension (Garthwaite and McMahon, 2004). Des études cliniques récentes (RALES et EPHESUS) ont démontré les effets bénéfiques de ces antagonistes de MR sur la réduction de la morbi-mortalité et des hospitalisations chez des sujets atteints d'insuffisance cardiaque (Brown, 2003; Fuller and Young, 2005; Garthwaite and McMahon, 2004; Takeda, 2004). Ces effets des antagonistes de MR semblent indépendants de leurs effets sur l'hypertension, renforçant l'hypothèse selon laquelle l'aldostérone agirait directement comme promoteur de l'hypertrophie et de la fibrose cardiaque et altèrerait la fonction de l'endothélium vasculaire. Les effets secondaires des traitements à la spironolactone, liés en particulier à la structure stéroïdienne de la molécule qui détermine son action sur les autres RNs de classe I ([Table 6](#)) ont conduit au développement de l'eplerenone présentant une affinité plus faible pour MR mais une activité à peine plus réduite et surtout une absence quasi-totale d'affinité pour les autres récepteurs aux hormones stéroïdiennes (Brown, 2003; Garthwaite and McMahon, 2004; Takeda, 2004). Le principal effet secondaire de ce nouvel antagoniste (que l'on retrouve aussi pour la spironolactone) semble être l'augmentation de la concentration plasmatique en potassium (Brown, 2003), pouvant aller, dans quelques cas, jusqu'à l'hyperkaliémie (concentration sérique en ion K<sup>+</sup> supérieure à 5,5 mmol/L) mettant alors en jeu le pronostic vital.

Comme son nom l'indique, la progestérone, via les récepteurs PR, est principalement impliquée dans l'initiation et le maintien de la gestation. Deux récepteurs : PR-A et PR-B sont synthétisés à partir d'un unique ARNm grâce à l'utilisation de deux codons d'initiation de la traduction différents. Ces deux isoformes semblent intervenir à des degrés divers dans les effets de la progestérone au niveau de l'ovulation, de la préparation de l'épithélium utérin et de l'endomètre à la nidation du blastocyste et du développement mammaire associé à la gestation (pour une synthèse voir Conneely et al., 2003). En thérapeutique, les agonistes de PR sont principalement utilisés comme contraceptifs. Ils agissent majoritairement en densifiant la glaire cervicale qui empêche le passage des spermatozoïdes. Les antagonistes de PR comme le RU486 sont principalement utilisés comme abortif en début de grossesse (Spitz, 2003). Ils empêchent le maintien et la densification de la muqueuse utérine pendant la nidation et provoquent l'expulsion de l'embryon, un phénomène qui est facilité par la co-administration d'une prostaglandine.

Les androgènes dont la testostérone est le principal représentant exercent, via le récepteur AR, des actions multiples dans l'organisme, en particulier chez le mâle. Ils participent à la différenciation sexuelle du fœtus (masculinisation à partir de la huitième semaine post-conception), au développement des organes génitaux mâle (conduits, verge, vésicules séminales, prostate) et des caractères sexuels secondaires (pilosité, voix, morphologie, comportement) à la puberté, ils favorisent la spermatogénèse et le maintient des organes sexuels, ils ont un effet anabolisant (accroissement de la fixation des protéines dans les muscles) et agissent sur la croissance (augmentation de la masse musculaire et accélération de la soudure des cartilages de conjugaison), ils favorisent la minéralisation de l'os, augmentent l'hématocrite (stimulation de la sécrétion d'érythropoïétine), la rétention du sodium et la calcémie, ils agissent enfin sur les glandes sébacées (acné) et sur le système nerveux central (rétrocontrôle de l'axe hypothalamo-hypophysaire, libido, agressivité). Les deux principales utilisations des ligands de AR correspondent aux propriétés anabolisantes de ses agonistes et à la capacité de ses antagonistes à inhiber la croissance des tumeurs de la prostate. Les stéroïdes anabolisants (exemple du trenbolone acétate en [Table 6](#)) sont ainsi utilisés pour favoriser la croissance des animaux de rente, en particulier les bovins (légalement dans certains pays dont les Etats-Unis mais illégalement dans l'Union Européenne) ainsi que par certains sportifs pour améliorer leurs performances, leur endurance et leur motivation (Kuhn, 2002). La prostate, glande de l'appareil sexuel masculin, est principalement impliquée dans le stockage et la sécrétion (dans une moindre mesure la synthèse) du liquide séminal, un des constituants du sperme. Le récepteur AR joue des rôles

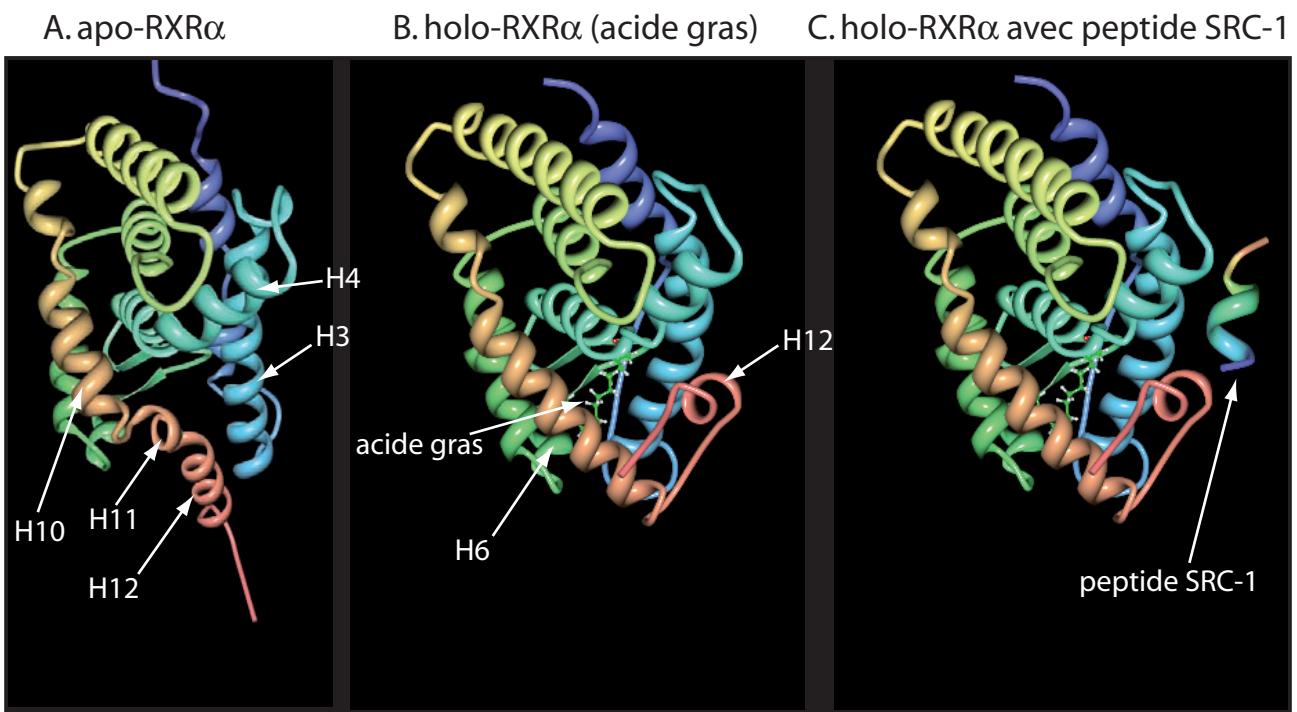


Figure 11 : Changements conformationnels du LBD de RXR $\alpha$  engendrés par la fixation d'un ligand.

En l'absence de ligand (A, PDB : 1LBD, Bourget *et al.*, 1995), l'hélice H12 est orienté vers l'extérieur du LBD. En présence d'un acide gras provenant des bactéries utilisées pour la surexpression (B, PDB : 1XVP, Xu *et al.*, hétérodimère CAR-RXR $\alpha$ , 2004) l'hélice H12 est repliée contre le LBD créant une nouvelle surface avec les hélices H3 et H4. Ces différents changements conformationnels permettent l'interaction d'un co-activateur (ici SRC-1) dont un peptide contenant le motif LXXLL a été co-cristallisé (C).

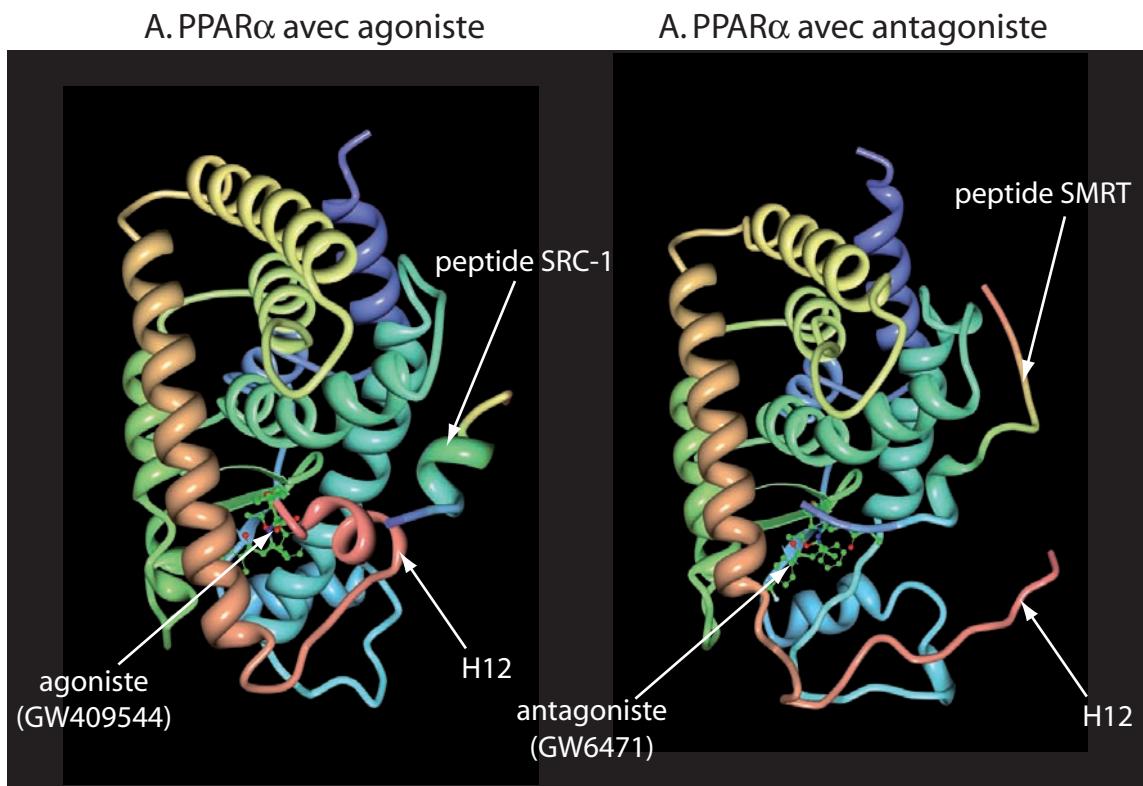


Figure 12 : Comparaison des sites de fixation des coactivateurs et corépresseurs sur le LBD de PPAR $\alpha$

En présence d'un ligand agoniste (A, PDB : 1K7L, Xu *et al.*, 2001), l'hélice H12 adopte une conformation active permettant la fixation d'un peptide contenant le motif LXXLL de SRC-1. En présence d'un antagoniste (B, PDB : 1KKQ, Xu *et al.*, 2002), la fixation d'un peptide contenant le motif LXXXIXXXL de SMRT empêche le positionnement de l'hélice H12 dans sa conformation active.

importants dans le développement de la prostate et dans le maintien de son épithélium sécrétoire qui semble être le premier type cellulaire transformé dans les adénocarcinomes de la prostate (Heinlein and Chang, 2004). Lors du diagnostic initial, environ 80 à 90% des cancers de la prostate sont dépendants des androgènes pour leur croissance (Heinlein and Chang, 2004) et le traitement est donc orienté vers la réduction des androgènes circulants et l'inhibition de l'activité de AR. Le flutamide (et son métabolite actif l'hydroxyflutamide) et le plus récent bicalutamide ([Table 6](#)) sont des ligands antagonistes de AR utilisés pour inhiber AR dans les cas de cancers de la prostate androgènes-dépendants. Malheureusement et comme c'est le cas pour de nombreux cancers du sein traités par le tamoxifène, une proportion importante de cancers de la prostate resurgissent en ayant perdu leur sensibilité au traitement à savoir le retrait des androgènes et/ou l'inhibition de l'activité de AR (Heinlein and Chang, 2004; Hiipakka and Liao, 1998). Des recherches sont en cours pour comprendre les mécanismes qui conduisent au développement de ces tumeurs et à leur insensibilité aux androgènes (Heinlein and Chang, 2004; Hiipakka and Liao, 1998).

Après avoir présenté les ligands endogènes et synthétiques des RNs, voyons à présent quels sont les mécanismes de fixation de ces ligands sur les récepteurs et les conséquences de cette fixation.

#### I.4.5. Mécanismes et conséquences de la fixation des ligands

Les nombreuses études structurales réalisées sur les domaines de fixation des ligands (LBD) des RNs ([Table 3](#)) ont permis de définir les règles générales des mécanismes de fixation des ligands. Ces études ont aboutit à un modèle appelé **modèle du « piège à souris »** pour la fixation des ligands. Comme présenté en [Figure 8](#) avec le cas particulier de RXR, le LBD de la plupart des RNs présente une moindre densité d'hélices  $\alpha$  dans sa partie « basse » définissant ainsi la **poche de fixation du ligand**. La [Figure 11](#) présente la comparaison des structures du LBD de RXR en l'absence de ligand ([Figure 11A](#)), en présence d'un ligand ([Figure 11B](#)) et en présence d'un ligand et d'un peptide issu de SRC-1 ([Figure 11C](#)), une protéine agissant comme coactivateur du récepteur RXR (voir paragraphe I.5.). Le ligand ([Figure 11B-C](#)) est ici un acide gras vraisemblablement issu de la culture de bactéries ayant servie à produire le LBD de RXR (Xu et al., 2004). Cette figure illustre les changements de conformation qui s'opèrent dans le LBD suite à la fixation du ligand (comparer les [Figures 11A et 11B](#)). L'hélice H11 est repositionnée en continuité de l'hélice H10 alors que l'hélice H12 bascule, dépliant ainsi une boucle qui passe sous l'hélice H6 et longe la partie N-terminale de l'hélice H3. Dans sa position finale, l'hélice H12 ferme la poche de fixation du

ligand comme un couvercle. Dans certains cas, l'hélice H12 participe également à stabiliser la fixation du ligand en établissant des interactions protéine-ligand additionnelles par rapport à celles déjà engagées dans la poche de fixation du ligand (Laudet and Gronemeyer, 2002). Le phénomène le plus nettement visible parmi ces changements conformationnels est le repositionnement de l'hélice H12 qui est le support de la fonction de transactivation dépendante du ligand (AF-2). Ce repositionnement, accompagné d'autres modifications plus subtiles comme la torsion de l'hélice H3, conduit à la formation d'une nouvelle surface qui ressemble à un **sillon** sur la partie externe du LBD. C'est la création de cette surface qui va permettre l'interaction du LBD avec les protéines co-activatrices du récepteur comme le coactivateur SRC-1 ([Figure 11C](#), voir paragraphe I.5.). Ces coactivateurs possèdent un motif conservé (motif LXXLL où L est une leucine et X n'importe quel acide aminé), appelé **boîte récepteur nucléaire**, qui permet leur interaction avec le LBD. Dans la structure de la [Figure 11C](#), un peptide issu de SRC-1 et contenant le motif LXXLL a été cristallisé avec le récepteur RXR (Xu et al., 2004). Ce motif adopte une conformation en hélice  $\alpha$  qui fait deux tours complets. La fixation de ce motif sur la surface du LBD est stabilisée par des interactions principalement hydrophobes entre les leucines du motif et les résidus du sillon. De plus, ce motif est maintenu de part et d'autre par deux liaisons hydrogène impliquant des acides aminés conservés de l'hélice H12 (glutamate) et de l'hélice H3 (lysine). Les anglophones parlent de « **charge clamp** » pour qualifier l'effet « d'étau » opéré sur le motif LXXLL par ces deux liaisons hydrogène. En plus de stabiliser l'interaction peptide-récepteur, ces deux liaisons hydrogène déterminent la longueur précise de l'hélice  $\alpha$  qui peut se fixer dans le sillon du LBD (Bourguet et al., 2000a). Lors de la création de cette nouvelle surface sur le LBD, des protéines fixées au LBD et agissant comme corépresseurs sont libérées suite à la fixation d'un ligand agoniste, mais pas nécessairement si le ligand est un antagoniste. Ces corépresseurs peuvent être fixés au LBD en l'absence de ligand ou en présence d'un antagoniste selon les cas. Ils interagissent avec le LBD dans une zone chevauchant la surface d'interaction avec les coactivateurs ([Figure 12](#)) mais ils empêchent l'hélice H12 d'adopter une conformation active (i.e. une conformation permettant la création d'une surface compatible avec la fixation des coactivateurs, Xu et al., 2002a). Les corépresseurs NCoR et SMRT (voir paragraphe I.5.3.) présentent également un motif conservé mais il est plus long (LXXI/HIXXXL/I où L est une leucine, I une isoleucine, H est une histidine, X n'importe quel acide aminé et où le signe / indique une alternative entre les deux acides aminés indiqués) et forme une hélice  $\alpha$  irrégulière à trois tours ([Figure 12](#)). La fixation du ligand et

les changements conformationnels engendrés permettent donc à la fois la libération des corépresseurs et la fixation des coactivateurs sur le LBD ([Figures 11 et 12](#)). Ces éléments permettent de comprendre pourquoi la fonction AF-2 (fonction de transactivation située en partie C-terminale du récepteur, au niveau de l'hélice H12) est dépendante de la fixation du ligand.

Dans les différentes structures de LBD élucidées en présence de ligand ([Table 3](#)), ce dernier est enfoui au sein de la protéine (voir [Figures 11 et 12](#) par exemple) sans qu'un point d'entrée et de sortie manifeste ne soit identifiable. Des changements conformationnels significatifs sont nécessaires pour générer un tel point d'entrée/sortie dans la plupart des structures de LBD disponibles. Le modèle du « piège à souris » évoqué ci-dessus fournit une réponse probable à cette question : ce serait la mobilité de l'hélice H12 qui permettrait l'ouverture d'un canal en ouvrant le couvercle de la poche de fixation du ligand. Les poches de fixation du ligand dont les structures sont disponibles sont bordées d'acides aminés majoritairement hydrophobes. Quelques résidus polaires, situés généralement au fond de la poche, près du feillet  $\beta$ , agissent comme points d'ancrage du ligand et jouent des rôles essentiels dans son positionnement correct et dans la sélectivité de la poche. La plupart des RNs possèdent ainsi une arginine très conservée au sein de l'hélice H5 qui pointe vers l'intérieur de la cavité où se trouve le ligand (Laudet and Gronemeyer, 2002). Ces résidus polaires, souvent conservés au sein d'une sous-famille de RNs, confèrent ainsi des spécificités de groupements chimiques polaires aux ligands des membres de ces sous-familles (exemple : groupement céto pour les stéroïdes et groupement carboxylique pour les rétinoïdes). Enfin, pour certains récepteurs comme les RAR, les RXR ou les récepteurs aux hormones stéroïdiennes, la forme de la poche de fixation du ligand semble plutôt rigide. Dans le cas des RAR et RXR, les structures obtenues ont mis en évidence que c'était le ligand qui adaptait sa conformation à la forme de la poche, maximisant ainsi le nombre des interactions hydrophobes possibles avec la poche. Dans le cas des RNs de classe I, la rigidité des stéroïdes ne permet pas ces adaptations conformationnelles et il a été observé que les poches de ces récepteurs étaient significativement plus grandes que leurs ligands respectifs. Ce sont, pour ces récepteurs, des interactions spécifiques clefs avec des acides aminés précis de la poche qui jouent un rôle prépondérant dans la spécificité de reconnaissance du ligand, plus que la somme des interactions hydrophobes. En revanche, il semble que certains récepteurs comme les PPAR, les LXR ou encore PXR, possédant des poches de fixation plus volumineuses (surtout PPAR et PXR) et présentant, dans l'état actuel des connaissances, une plus grande diversité de ligands, soient capables de modifier la forme générale et le volume de leur poche

de fixation pour s'adapter à une diversité de structure des ligands (Chrencik et al., 2005; Farnegardh et al., 2003; Watkins et al., 2003b).

La [Table 6](#) précise, pour un éventail de ligands synthétiques, si ceux-ci agissent comme agoniste, antagoniste ou modulateur sélectif du récepteur considéré (et éventuellement d'autres récepteurs). Dans la littérature, on rencontre également les notions d'agoniste inverse ou d'agoniste (ou antagoniste) partiel. Afin d'éviter les confusions, revenons sur l'utilisation de ces termes et leur signification. Tout d'abord, rappelons que ces notions ont été initialement définies pour les récepteurs membranaires et qu'elles ne dépendent pas directement de la liaison entre le ligand et le récepteur mais du résultat de cette liaison c'est-à-dire de l'effet induit par cette liaison sur l'activité du récepteur. Lorsque le ligand se fixe à un récepteur, il peut l'activer, le ligand est alors appelé **agoniste** ou bien le bloquer, auquel cas il est appelé **antagoniste**. Un ligand qui se fixe sur le récepteur mais produit un effet inverse de celui d'un agoniste est appelé **agoniste inverse**. Un antagoniste est capable de bloquer l'effet d'un agoniste comme l'effet d'un agoniste inverse. L'effet induit est dépendant de la dose de ligand et est saturable puisque l'effet dépend notamment de la quantité de récepteur présent qui est fixe. En pharmacologie, la relation dose (D)-effet (E) étant de nature hyperbolique, elle est modélisée par la relation  $E=(\alpha E_{max} \times D)/(D+DE_{50})$  où  $E_{max}$  caractérise l'**efficacité** de l'agoniste et  $DE_{50}$  (ou  $CE_{50}$  si on exprime la dose sous la forme de concentrations) caractérise sa **puissance** (plus la  $DE_{50}$  est faible plus la molécule est puissante). Le coefficient  $\alpha$  caractérise l'activité intrinsèque de l'agoniste, c'est-à-dire sa capacité à entraîner le couplage du récepteur à son (ou ses) effecteur(s). S'il vaut 1, l'agoniste est dit entier. S'il est inférieur à 1, on parlera **d'agoniste partiel**. En pratique, cela revient au même de dire qu'un agoniste partiel est un agoniste dont l'efficacité est moindre qu'un agoniste entier et, dans le domaine des récepteurs nucléaires, on parle plutôt **d'agoniste fort** ou **d'agoniste faible** (même si ces notions se réfèrent normalement plutôt à la puissance des ligands qu'à leur efficacité). Dans le domaine des récepteurs nucléaires, nous avons vu que la notion de **modulateur sélectif** a été introduite, notamment dans le cadre des travaux sur les ligands des récepteurs aux estrogènes. Un modulateur sélectif est un ligand qui est capable d'agir comme agoniste dans certains sites d'action et comme antagoniste dans d'autres sites. Remarquons que, dans la littérature sur les récepteurs nucléaires, les notions d'agoniste (ou antagoniste) partiel et de modulateur sélectif sont souvent utilisées comme synonymes, ce qui ne correspond pas aux définitions données ci-dessus. En pratique, pour appliquer ces notions aux récepteurs nucléaires, il est cependant nécessaire de prendre en compte un certains nombre d'éléments spécifiques au mode de fonctionnement de cette famille de facteurs de transcription :

- Un ligand peut interagir positivement ou négativement avec l'activité d'un récepteur nucléaire à différents niveaux. Le ligand peut par exemple affecter 1) la stabilité du complexe RN-protéines chaperones (protéine heat-shock comme Hsp90) situé au niveau du cytosol dans les cellules n'ayant pas été exposées à l'hormone (voir paragraphe I.6.), 2) la durée de demi-vie du récepteur, 3) la capacité d'homo- ou d'hétérodimérisation du récepteur ou encore 4) sa capacité d'interaction avec l'ADN (élément de réponse).
- Les récepteurs nucléaires portent, comme nous l'avons vu, deux fonctions d'activation de la transcription (AF-1 et AF-2) qui fonctionnent de manière dépendante du promoteur et du type cellulaire. Bien que la fonction AF-1 soit qualifiée d'indépendante du ligand car elle peut effectivement fonctionner en l'absence de ligand dans le cadre d'une protéine de fusion contenant un DBD hétérologue, cette fonction est néanmoins, comme AF-2, activée par la présence d'un agoniste dans le contexte du récepteur entier. Un antagoniste peut bloquer l'une, l'autre ou les deux fonctions d'un même récepteur et un antagoniste AF-2 peut éventuellement agir comme agoniste de la fonction AF-1 (exemples : RU486 ou 4-hydroxytamoxifène Laudet and Gronemeyer, 2002).
- Alors que les bases structurales de la fonction AF-1 sont actuellement inconnues, la fonction AF-2 correspond à la formation, dépendante du ligand, d'une surface d'interaction avec des coactivateurs. Par ailleurs, un récepteur non lié à un ligand ou lié à un antagoniste particulier peut présenter une surface permettant l'interaction avec des corépresseurs. Un ligand donné peut générer de manière plus ou moins précise ces différentes surfaces et produire ainsi des efficacités différentes de recrutement des coactivateurs et corépresseurs. On qualifie en général de superagoniste un ligand capable d'augmenter le recrutement de coactivateurs plus efficacement que le ligand endogène du récepteur et d'agoniste inverse un ligand capable de stabiliser le complexe récepteur – corépresseurs.
- Enfin, comme nous le verrons dans le paragraphe I.5., il existe différents types de coactivateurs des récepteurs nucléaires qui ont des actions différentes soit sur la structure locale de la chromatine, soit sur l'interaction récepteur-machinerie transcriptionnelle. Certains ligands semblent capables de promouvoir le recrutement d'un type de coactivateurs au profit d'un autre type comme cela a par exemple été montré pour le récepteur VDR (Yang and Freedman, 1999).

Ces considérations montrent à quel point il est difficile de qualifier sans ambiguïté un ligand d'agoniste ou d'antagoniste tant il y a de facteurs à évaluer. La notion de modulateur

sélectif est sans doute, dans ce cadre, la plus souple puisqu'elle envisage la possibilité que le ligand exerce des effets différentiels en fonction du contexte cellulaire et du promoteur considéré.

Nous avons précisé que ces notions ne sont pas directement reliées aux modifications structurales évoquées précédemment dans le sens où ce sont des études portant sur l'activité du récepteur et non sur sa structure seule qui permettent de qualifier un ligand d'agoniste ou d'antagoniste. Cependant, il est indéniable que les études structurales sur les LBD des RNs ont permis de mieux comprendre les mécanismes moléculaires qui permettent aux différents ligands d'un récepteur de moduler de manière différentielle son activité.

Nous avons déjà évoqué les conséquences de la fixation d'un ligand agoniste qui se traduisent par l'apparition d'une surface permettant l'interaction avec les coactivateurs. Soulignons néanmoins que ce n'est pas toujours le ligand seul qui dicte la conformation « active » de l'hélice H12. Ainsi, il a été observé que le LG100268, un agoniste très spécifique de RXR était incapable de déplacer seul des corépresseurs du récepteur et que l'addition de coactivateurs était nécessaire pour libérer ces corépresseurs et placer le récepteur dans une conformation active (Love et al., 2002). Cette hypothèse est appuyée par l'observation du complexe RXR $\beta$ -LGD100268 où l'hélice H12 ne présente pas sa position active (Love et al., 2002), similaire à celle présentée en [Figure 11B](#). En ce qui concerne les antagonistes, et au regard des spécificités des RNs que nous avons évoquées précédemment, il n'est pas surprenant que plusieurs mécanismes structuraux aient été proposés dans la littérature. Les structures du LBD de ER $\alpha$  en présence des modulateurs sélectifs raloxifène et 4-hydroxytamoxifène (Brzozowski et al., 1997; Shiau et al., 1998) ont mis en évidence une position de l'hélice H12 particulière où celle-ci se loge entre les hélices H3 et H4, dans le sillon normalement occupé par le motif LXXLL des coactivateurs, empêchant ainsi leur interaction avec le LBD. Cette position de l'hélice H12 est permise à la fois par une volumineuse chaîne latérale des antagonistes qui ne peut s'insérer dans la poche de fixation du ligand et empêche ainsi le positionnement « agoniste » de H12 ainsi que par un dépliement de l'hélice H11, lié probablement à des interactions particulières entre H11 et le ligand qui permet à H12 de se positionner dans le sillon où devrait se réaliser l'interaction avec les coactivateurs (Bourguet et al., 2000a; Laudet and Gronemeyer, 2002). Dans d'autres structures (Bourguet et al., 2000b; Pike et al., 1999), il a été montré que certains ligands (génistéine pour ER $\beta$  et acide oléique pour RXR $\alpha$ -F318A) agissant comme agonistes faibles, pouvaient eux aussi engendrer cette même conformation de l'hélice H12 dans les cristaux.

Ces ligands semblent réaliser les interactions nécessaires au dépliement de l'hélice H11 mais ne présentent pas de chaîne latérale volumineuse et n'empêchent donc pas l'hélice H12 de se placer éventuellement dans une position permettant la formation du sillon nécessaire à l'interaction avec les coactivateurs. Ce sont sans doute les concentrations respectives de coactivateurs et de corépresseurs qui dicteront alors l'activité de ces ligands. Alors que le raloxifène et le 4-hydroxytamoxifène sont considérés comme de pures antagonistes de la fonction AF-2 de ER $\alpha$ , la génistéine pour ER $\beta$  et l'acide oléique pour RXR $\alpha$ -F318A sont considérés comme des antagonistes (ou agonistes) partiels sur la base de ces données structurales et des études d'activité. Les structures obtenues sur les PPAR (Oberfield et al., 1999; Xu et al., 2002a) ont permis de proposer de nouvelles hypothèses pour la réalisation des activités antagoniste partielle ou mixte. Ainsi, le GW0072, un agoniste partiel de PPAR $\gamma$ , semble incapable de stabiliser la conformation active de l'hélice H12 en raison de l'absence de contacts entre le ligand et cette hélice (Oberfield et al., 1999). Dans la première structure élucidée en présence d'un peptide issu d'un corépresseur (Xu et al., 2002a), l'hélice H12 adopte encore une nouvelle position, contre l'hélice H3, illustrée par la [Figure 12B](#) qui permet la fixation du motif corépresseur. Enfin, les structures 2OVH et 2OVM de PR en présence de peptides issus de SMRT ou NCoR respectivement qui ont été récemment déposée dans la base de données Protein Data Bank (Madauss et al., 2007) montre une hélice H12 alignée le long de l'hélice H11, une conformation qui semble notamment liée au positionnement fixe du domaine F de PR en présence d'un agoniste ou d'un antagoniste. Il est probable que l'acquisition de nouvelles structures en présence de peptides issus de corépresseurs permettra d'éclairer encore le mode de fonctionnement des antagonistes purs ou partiels.

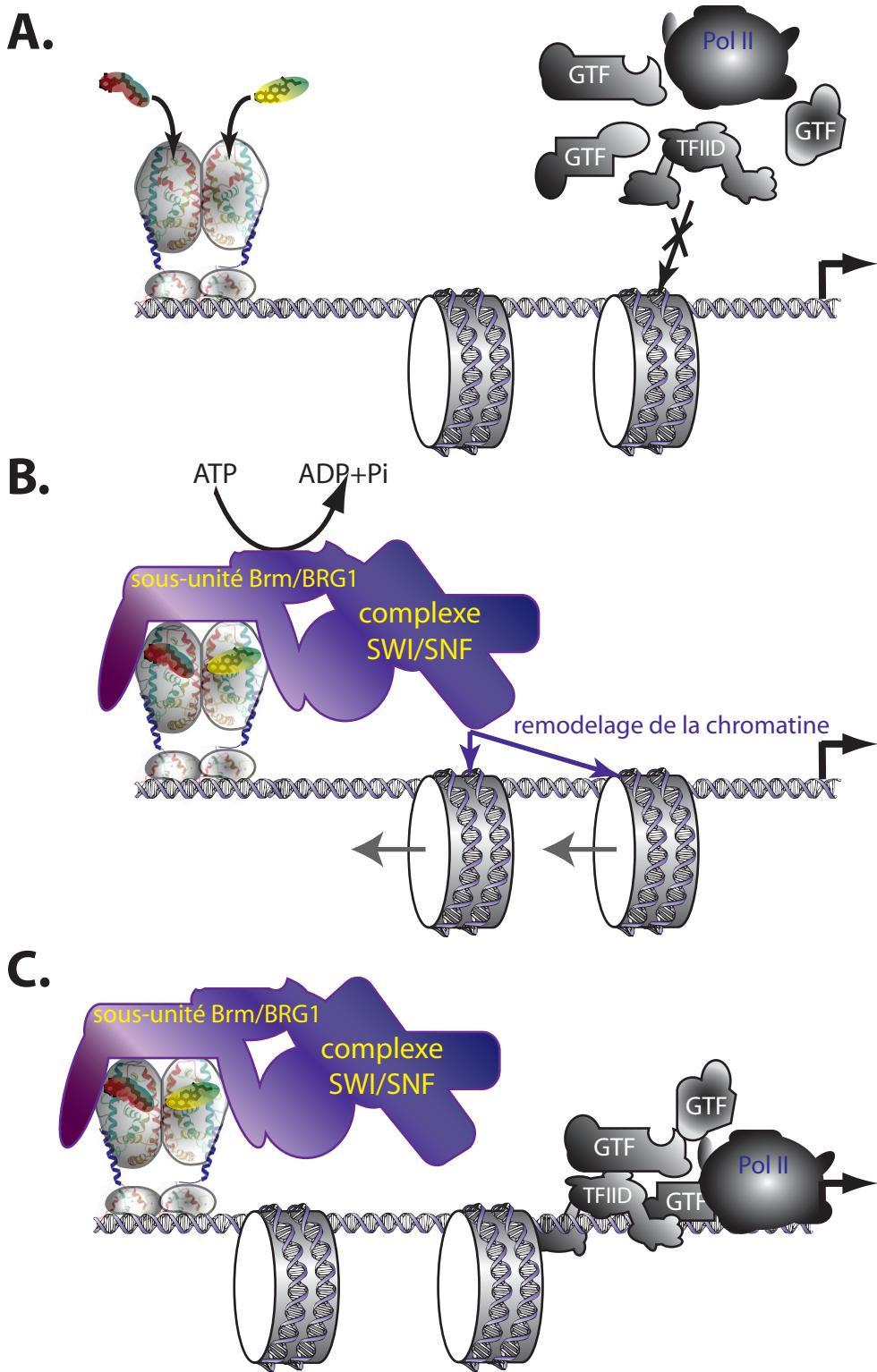
## I.5. COREGULATEURS ET REGULATION DE LA TRANSCRIPTION

De manière générale, la fixation d'un ligand sur un RN se traduit donc par des modifications conformationnelles, assez bien décrites au niveau du LBD, conduisant à la fixation de corégulateurs (ou cofacteurs). Ces corégulateurs, également appelés facteurs intermédiaire de la transcription (transcription intermediary factor ou TIF en anglais) vont nous permettre de décrire les éléments connus du lien entre le RN ayant fixé un ligand (holo-récepteur) et se trouvant sur un élément de réponse et la régulation de la transcription du gène cible correspondant.

### I.5.1. Facteurs généraux de la transcription

La régulation de l'expression des gènes peut s'opérer à divers niveaux : la transcription, la stabilisation, la maturation, le transport ou la dégradation des ARN messagers, la traduction, et les régulations post-traductionnelles. La transcription elle-même peut-être régulée à plusieurs niveaux : régulations épigénétiques via des méthylations par exemple, modifications locales de la structure de la chromatine, assemblage et activité des complexes d'initiation et d'elongation de la transcription. L'activité des RNs a été très majoritairement décrite via leurs effets sur la structure locale de la chromatine et sur l'initiation de la transcription par l'ARN polymérase II. De fait, les RNs appartiennent à la catégorie des facteurs de transcription qui modulent (en plus ou en moins) le niveau transcription. Ils se distinguent ainsi des facteurs d'initiation de la transcription qui déterminent le niveau basal de transcription des gènes. Ces derniers sont parfois qualifiés de machinerie transcriptionnelle basale et appartiennent à la famille plus large des facteurs généraux de transcription (qui comprend les protéines impliquées dans l'initiation, l'elongation et la terminaison de la transcription). La machinerie transcriptionnelle basale est constituée de complexes multiprotéiques qui se fixent sur des éléments spécifiques des promoteurs dont les plus courants sont la boîte TATA (TATAWAW où W est soit A soit T) et une séquence d'initiation riche en bases pyrimidiques (C et T) dénommée Inr (pyrimidine-rich initiator, séquence consensus YYANWYY où Y est une base pyrimidique et N n'importe quelle base). Dans cette machinerie, on distingue en particulier la protéine liant la boîte TATA (TBP pour TATA box-binding protein) et les facteurs associées à TBP (TAFs pour TBP-associated factors) qui forment le facteur de transcription TFIID, d'autres complexes, généralement multiprotéiques, formant les facteurs de transcription TFIIA, IIB, IIF, IIE et IIH et enfin le complexe multiprotéique ARN polymérase II (pour une synthèse voir Roeder, 1996). Cet ensemble forme le complexe de pré-initiation (PIC pour pre-initiation complexe). C'est donc l'interaction directe ou indirecte entre cette machinerie transcriptionnelle basale et les RN activés par leurs ligands et fixés sur leurs éléments de réponse qui est assurée par les cofacteurs des RNs.

Environ 200 coactivateurs des RNs ont été décrits dans la littérature dont une soixantaine ont été caractérisés par plus d'un laboratoire et démontrés comme agissant réellement comme coactivateurs des RN (Lonard and O'Malley, 2006). Par ailleurs, un nombre plus réduit de corépresseurs a été décrit parmi lesquels NCoR (NR-corepressor) et SMRT (silencing mediator for RAR and TR) ont été les plus étudiés. Une description



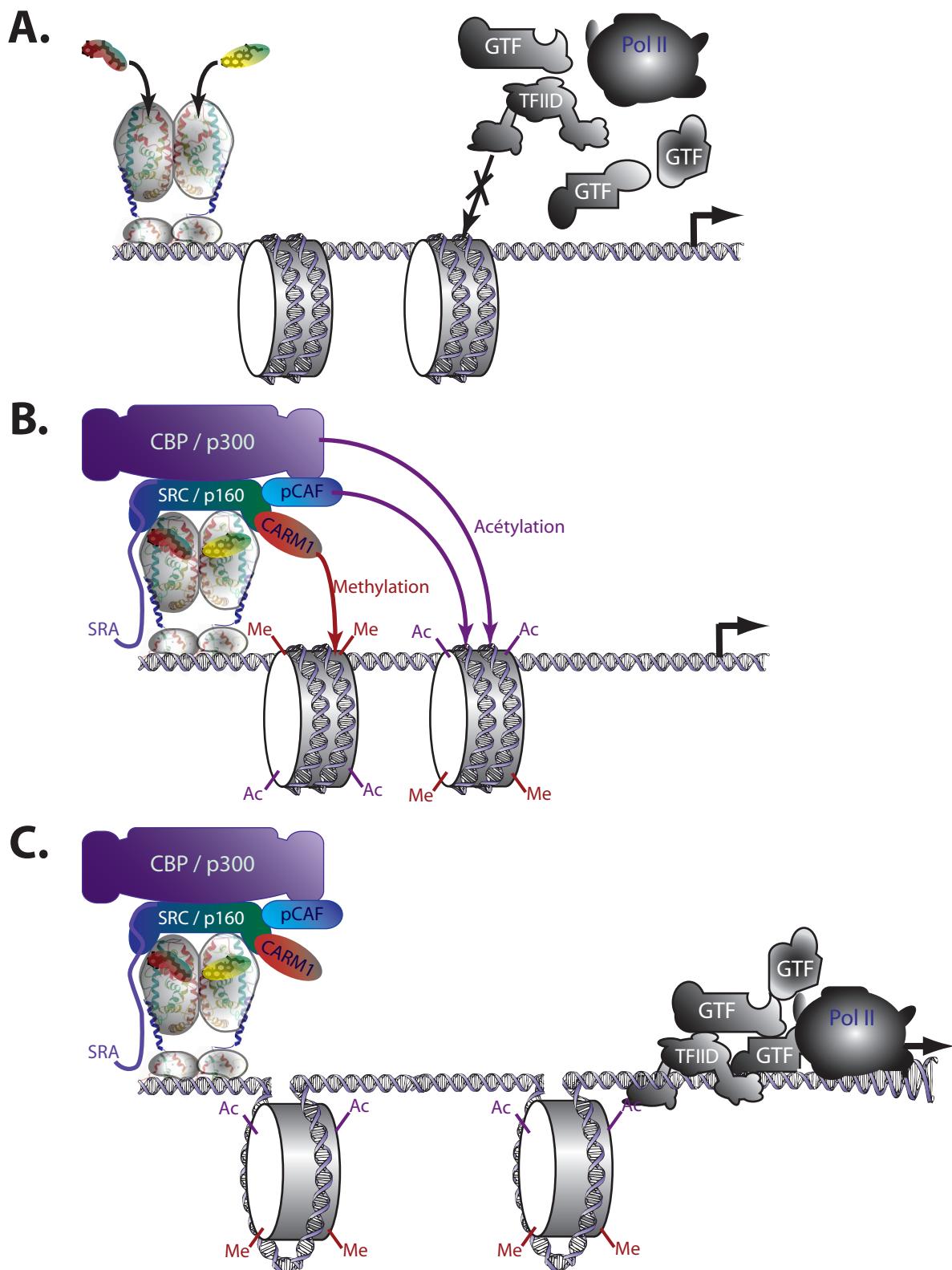
**Figure 13 :** Représentation schématique du fonctionnement des corégulateurs agissant par remodelage de la chromatine

A. L'accès des facteurs généraux de la transcription (GTF, TFIID, Pol II) à l'ADN est limité par sa compaction assurée par les histones. Un hétérodimère de RNs est représenté sur la gauche, fixé à leur élément de réponse et exposé à leurs ligands respectifs. B. La fixation des ligands permet de générer des surfaces d'interaction avec des corégulateurs (voir leur description dans le texte) qui provoquent un remodelage des histones grâce à l'énergie fournie par l'hydrolyse de l'ATP. Le "glissement" des nucléosomes représenté ici n'est qu'une des possibilités de remodelage de la chromatine et la formation, la destruction ou la modification de la conformation d'un nucléosome peuvent également se produire selon les cas. C. Ce remodelage des histones autorise l'accès des facteurs généraux de la transcription à l'ADN. Le complexe de pré-initiation peut ainsi s'assembler et initier la transcription. GTF : General Transcription Factor, TFIID : Transcription Factor II D (contient TBP), Pol II : ARN polymérase II.

régulièrement actualisée des différents cofacteurs décrits dans la littérature peut-être consultée sur [www.nursa.org](http://www.nursa.org). En raison de cette importante diversité de protéines, il reste difficile, à l'heure actuelle, de définir une classification exhaustive des modes d'actions des cofacteurs. Notons également qu'aborder le domaine des corégulateurs des RNs est rendu difficile par l'absence d'une nomenclature officielle qui conduit souvent les corégulateurs à avoir plusieurs noms différents dans la littérature. Nous allons tenter de synthétiser la vaste bibliographie se rapportant aux corégulateurs des RNs en les regroupant en 4 catégories : 1) les facteurs de remodelage ATP-dépendant de la chromatine, 2) les corégulateurs agissant par modification covalente des histones, 3) les complexes Médiateurs permettant l'interaction entre les RNs et la machinerie transcriptionnelle basale et 4) les autres corégulateurs au fonctionnement atypique ou mal connus.

### I.5.2. Corégulateurs et remodelage de la chromatine

Un premier type de corégulateurs agit comme **facteur de remodelage de la chromatine** (pour une synthèse récente, voir Chen et al., 2006). Chez les eucaryotes, l'ADN génomique est empaqueté dans des nucléosomes qui constituent l'unité de base de la structure de la chromatine et sont formés par l'enroulement d'environ 147 paires de bases d'ADN autour d'un octamère constitué de quatre histones différentes : H2A, H2B, H3 et H4 (Kishimoto et al., 2006). Les nucléosomes sont séparés par un court fragment d'ADN auquel est fixée l'histone H1. La structure de la chromatine joue un rôle important dans la régulation de l'activité basale des promoteurs car l'empaquetage de l'ADN sous forme de chromatine a un effet répresseur sur l'expression des gènes et qu'un réarrangement de la structure de la chromatine est nécessaire à l'initiation de la transcription (Li et al., 2007; Urnov and Wolffe, 2001). Notons que certains RNs, contrairement à de nombreux facteurs de transcription, semblent capables de se fixer à leurs éléments de réponse même lorsqu'ils sont enfouis dans la chromatine (Li et al., 2007; Urnov and Wolffe, 2001). L'action des corégulateurs de ce premier groupe consiste à remodeler de larges domaines de la chromatine par des mécanismes ATP-dépendants impliquant une déformation non-covalente des nucléosomes qui a pour conséquence de faciliter l'accès de la machinerie transcriptionnelle basale au promoteur ([Figure 13](#)). On connaît actuellement cinq familles principales de complexes multiprotéiques impliqués dans ces phénomènes ATP-dépendants de remodelage de la chromatine (Chen et al., 2006), caractérisés chacun par la nature de leur sous-unité ATPase. Des interactions avec des RNs ont été démontrées pour deux familles de complexes : SWI/SNF et Mi-2/NuRD. Le



**Figure 14 :** Représentation schématique du fonctionnement des coactivateurs provoquant des modifications covalentes des histones

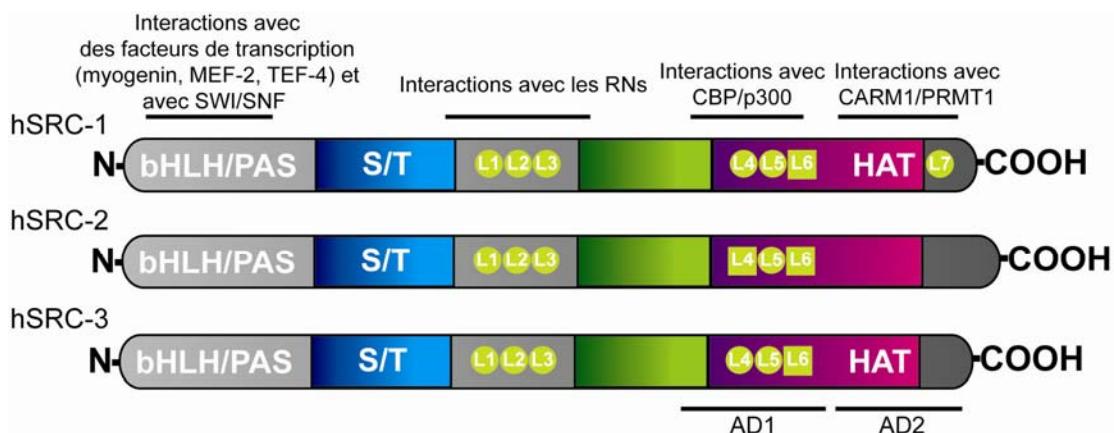
A. L'accès des facteurs généraux de la transcription (GTF, TFIID, Pol II) à l'ADN est limité par sa compaction assurée par les histones. Un hétérodimère de RNs est représenté sur la gauche, fixé à leur élément de réponse et exposé à leurs ligands respectifs. B. La fixation des ligands permet de générer des surfaces d'interaction avec des coactivateurs (voir leur descriptions respectives dans le texte) qui modifient les histones (acétylation et méthylation principalement). C. Ces modifications réduisent l'affinité des octamères d'histone pour l'ADN conduisant ainsi à sa décompaction et facilitant l'accès des facteurs généraux de la transcription à l'ADN. Le complexe de pré-initiation peut ainsi s'assembler et initier la transcription. GTF : General Transcription Factor, TFIID : Transcription Factor II D (contient TBP), Pol II : ARN polymérase II, Me : groupement méthyle, Ac : groupement acétyle

complexe multiprotéique SWI/SNF a été le mieux caractérisé (Chen et al., 2006; Simone, 2006; Urnov and Wolffe, 2001). Initialement découvert chez la levure, ce complexe possède une sous-unité Swi2/snf2, dont l'homme possède deux orthologues appelés Brm et BRG1, qui présente une activité ATPase et un bromodomaine capable de lier les lysines acétylées des histones. Le complexe SWI/SNF interagit, via diverses sous-unités, avec plusieurs récepteurs nucléaires dont PPAR $\gamma$ , VDR, GR, PR, ER et AR (synthétisé dans Chen et al., 2006; Simone, 2006). Il participe généralement à l'activation de l'expression des gènes par ces RNs et peut donc être qualifié de coactivateur mais il a également été impliqué, dans d'autres contextes que celui des RNs, dans des phénomènes de répression de la transcription (Chen et al., 2006; Urnov and Wolffe, 2001). Son action n'est cependant pas indispensable pour l'activation de tous les gènes cibles de ces RNs et son rôle semble dépendre de l'architecture spécifique des promoteurs (Chen et al., 2006). Un autre groupe de complexes : Mi-2/NuRD (nucleosome remodeling and deacetylase) est capable d'interagir avec ER $\alpha$  et ROR $\gamma$  et jouerait un rôle dans la répression (corépresseur) de gènes cibles de ces RNs via sa double activité de désacétylation des histones (HDAC, voir paragraphe suivant) et de remodelage de la chromatine (Chen et al., 2006; McKenna et al., 1999; Privalsky, 2004).

### I.5.3. Corégulateurs et modification covalente des histones

Le deuxième type de corégulateurs est caractérisé par des protéines assurant ou recrutant des fonctions enzymatiques provoquant des **modifications post-traductionnelles (covalentes) des histones** (Figure 14). Il s'agit sans doute du groupe de coactivateurs le mieux caractérisé et qui a fait couler le plus d'encre (et de réactifs !). Des structures tridimensionnelles de nucléosomes ont permis de mettre en évidence que les parties N-terminales des histones saillent hors des régions d'ADN enroulé. Depuis, il a été montré que ces régions des histones étaient sujettes à des modifications post-traductionnelles par acétylation, désacétylation, méthylation, phosphorylation, ubiquitination et sumoylation au niveau d'acides aminés spécifiques (Kishimoto et al., 2006; Li et al., 2007). La combinaison de ces diverses modifications conduirait à des altérations spécifiques de la structure des nucléosomes via une modification des interactions histone-histone ou histone-ADN, permettant ou empêchant l'accès des protéines régulatrices à l'ADN. Le terme de "code des histones" est parfois utilisé dans la littérature pour définir les différentes combinaisons de modifications post-traductionnelles des histones définissant l'état (relâché ou compact) de la chromatine (voir par exemple Privalsky, 2004). Ce code des histones semble particulièrement élaboré et des techniques analytiques permettent aujourd'hui d'en découvrir l'étendue (Garcia

et al., 2007). De manière générale, une acétylation des lysines (chargées positivement) des histones conduit à une neutralisation de leurs charges positives et altère ainsi les interactions électrostatiques entre les histones (chargées positivement) et l'ADN (chargé négativement) responsables du compactage de l'ADN autour du cœur d'histones. Le relâchement du nucléosome ainsi provoqué facilite l'accès des protéines régulatrices à cette séquence de l'ADN. A l'inverse, des protéines disposant d'une activité histone désacétylase (HDAC) privent les lysines des histones de leurs groupements acétyl et renforcent les interactions histone-ADN, conduisant à un renforcement de la compaction locale de l'ADN autour des histones et limitant l'accès des protéines régulatrices à ce segment d'ADN. Typiquement, des corégulateurs possédant (ou recrutant) une activité histone acétyltransférase (HAT) agiront comme coactivateurs et celles possédant (ou recrutant) une activité histone désacétylase (HDAC) agiront comme corépresseurs. L'effet d'autres modifications post-traductionnelles des histones est en général plus complexe. Par exemple, l'effet de la méthylation des lysines de certaines histones peut avoir, selon la lysine et l'histone considérée, un effet positif ou négatif sur le relâchement du nucléosome (Kishimoto et al., 2006; Li et al., 2007; Privalsky, 2004). Il est également important de souligner que ces modifications des histones ne font pas qu'altérer les propriétés biophysiques de la chromatine mais serviraient également de code dictant directement le recrutement ou l'exclusion de diverses protéines régulatrices de la matrice de chromatine (Privalsky, 2004).



**Figure 15 : Domaines fonctionnels et structuraux des coactivateurs SRC/p160**  
 Les boîtes RN (motifs de type LXXLL) sont indiquées par des ronds (motifs LXXLL typiques) ou par des carrés jaunes (motifs LXXLL atypiques comme par exemple L6 de hSRC-1 dont la séquence est IDKLV). Les différents sites d'interaction connus sont décrits dans les légendes du haut. Les domaines AD1 et AD2 décrits dans le texte sont définis dans la légende du bas. bHLH/PAS : domaine basic helix-loop-helix / Per/ARNT/Sim, S/T : région riche en sérine et thréonine, HAT : domaine à activité histone acétyltransférase. Schéma réalisé d'après Xu J. & Li Q., 2003.

Les 3 membres de la famille SRC/p160 : SRC-1/N-CoA1 (steroid receptor coactivator 1 / NR coactivator 1), TIF2/GRIP-1/SRC-2 (Transcriptional intermediary factor 2 / GR interacting protein 1 / steroid receptor coactivator 2) et pCIP/RAC-3/ACTR/AIB1/TRAM-1/SRC-3 (p300/CBP interacting protein / receptor-associated coactivator 3 / activator of TR / amplified in breast cancer 1 / TR activator molecule 1 / steroid receptor coactivator 3) appartiennent tous à ce type de corégulateurs. Ces trois coactivateurs sont capables d'accroître la transcription d'un gène rapporteur par différents RNs de manière dépendante du ligand. Ces trois protéines, d'environ 160 KDa chacune (d'où le nom p160 souvent donné à cette famille), présentent environ 40-50% de similarité de séquence et une organisation semblable, caractérisée par la présence de plusieurs domaines fonctionnels ([Figure 15](#), Leo and Chen, 2000; Xu and O'Malley, 2002). En partie N-terminale, le domaine bHLH-PAS (basic helix-loop-helix / Per/ARNT/Sim) est le mieux conservé. Ce domaine a été initialement identifié dans les protéines de Drosophile Period (Per) et single-minded (Sim) puis dans le récepteur aux dioxines (Aryl hydrocarbon receptor ou AhR) et chez son partenaire ARNT (AhR nuclear translocator) où il joue des rôles dans la fixation à l'ADN et les interactions protéine-protéine. Bien que son rôle soit encore mal caractérisé au sein de la famille SRC/p160, ce domaine semble notamment impliqué dans l'interaction des SRC avec d'autres voies de signalisation que celles des RNs (Wu et al., 2005a) ainsi qu'avec des sous-unités du complexe SWI/SNF (Belandia et al., 2002). Après une région riche en sérines et thréonines, on distingue, en partie centrale des SRC, un domaine appelé domaine interagissant avec les récepteurs (RID pour receptor-interacting domain) qui contient trois motifs de type LXXLL appelés boîtes RN (voir chapitre précédent et Plevin et al., 2005) responsables de l'interaction des SRC avec les RNs liés à leur ligand. Comme nous l'avons vu précédemment, ce sont ces motifs qui forment des hélices  $\alpha$  amphiphiles (les leucines conservées formant une surface hydrophobe sur l'une des faces de l'hélice) capable de se lier au sillon hydrophobe formé à la surface du LBD du complexe RN-ligand. La présence de trois boîtes RNs, leur séquences respectives, mais également les résidus flanquant ces séquences et leur espacement semble générer une certaine spécificité d'interaction des SRC vis-à-vis de différents RNs (pour une synthèse, voir Leo and Chen, 2000). En partie C-terminale des SRC, on distingue deux domaines d'activation de la transcription (AD1 et AD2 où AD signifie transcriptional activation domain) définis grâce à des expériences d'activation de la transcription d'un gène rapporteur par des protéines de fusion contenant différentes régions des SRC associées au domaine de liaison à l'ADN du récepteur Gal4. Le domaine AD1 contient trois autres motifs de type LXXLL qui sont

impliqués dans l'interaction des SRC avec les protéines CBP (CREB-binding protein où CREB signifie cAMP response element binding protein), p300 (adenovirus E1A binding protein p300) et p/CAF (p300/CBP-associated protein, homologue humain du gène GCN5 de levure). CBP, p300 et p/CAF possèdent de fortes activités histone acétyltransférase (HAT) et agissent ainsi comme coactivateurs des RNs (Glass and Rosenfeld, 2000; Xu and O'Malley, 2002). CBP et p300 agissent également comme coactivateurs de nombreux autres facteurs de transcription tels CREB, p53, NF-κB, STAT ou AP-1 et sont donc souvent qualifiés de cointégrateurs de part leur rôle potentiellement limitant dans l'intégration de multiples signaux au niveau des promoteurs. Précisons également que CBP et p300 sont aussi capable d'acétyler certains facteurs de transcription ou corégulateurs comme p53 (Glass and Rosenfeld, 2000), ER $\alpha$  (Kim et al., 2006) ou SRC-3 (Leo and Chen, 2000) par exemple et semblent également, au même titre que SRC-1 d'ailleurs, capable d'interagir avec certaines protéines de la machinerie transcriptionnelle basale comme TBP ou TFIIB (Glass and Rosenfeld, 2000; Leo and Chen, 2000). SRC-1 et SRC-3 présentent également une activité HAT localisée dans leur partie C-terminale mais celle-ci est faible au regard des activités HAT de CBP, p300 et p/CAF, suggérant que les SRC agissent comme coactivateurs de la transcription des gènes cibles principalement via le recrutement de protéines à activité HAT (Glass and Rosenfeld, 2000; Leo and Chen, 2000; Xu and O'Malley, 2002). Enfin, le domaine AD2 localisé à l'extrémité C-terminale des SRC est responsable de leur interaction avec les protéines à activité histone méthyltransférase CARM1 (coactivator-associated arginine methyltransferase 1) et PRMT1 (protein arginine methyltransferase 1). Ces activités enzymatiques agissant, vraisemblablement de manière synergique, sur les histones H3 (CARM1) et H4 (PRMT1) participeraient au relâchement de la chromatine nécessaire à l'activation de la transcription des gènes cibles (Glass and Rosenfeld, 2000; Leo and Chen, 2000; Xu and O'Malley, 2002). De plus, PRMT1 est également capable d'acétyler, dans sa partie C-terminale, le coactivateur PGC-1 $\alpha$  (voir ci-dessous), provoquant ainsi un accroissement de son activité (Handschin and Spiegelman, 2006). En dehors de ces deux protéines à activité méthyltransférase qui interagissent avec les SRC, il existe d'autres corégulateurs possédant ce type d'activité comme par exemple NSD1, (pour NR-binding SET-domain-containing protein 1, Huang et al., 1998), découverte lors d'une recherche de corégulateurs des RNs par la technique de double-hybrid en utilisant le LBD de RAR $\alpha$  et qui semble pouvoir fonctionner soit comme coactivateur soit comme corépresseur des RNs selon les situations. NSD1 possède un domaine SET impliqué dans la méthylation de certaines lysines des histones (Rayasam et al., 2003) mais ne présente

pas d'interaction connue avec les SRC. Les trois membres de la famille SRC sont exprimés dans de nombreux tissus avec des recoupages partiels de leurs sites d'expression (pour une synthèse, voir Xu and Li, 2003). Des lignées murines invalidées pour chaque membre de la famille SRC ont été générées et ont permis de commencer à mettre en évidence un certain niveau de redondance entre ces coactivateurs (comme entre SRC-2 et SRC-1 par exemple) ainsi que les rôles spécifiques de chacun (pour une synthèse, voir Xu and Li, 2003).

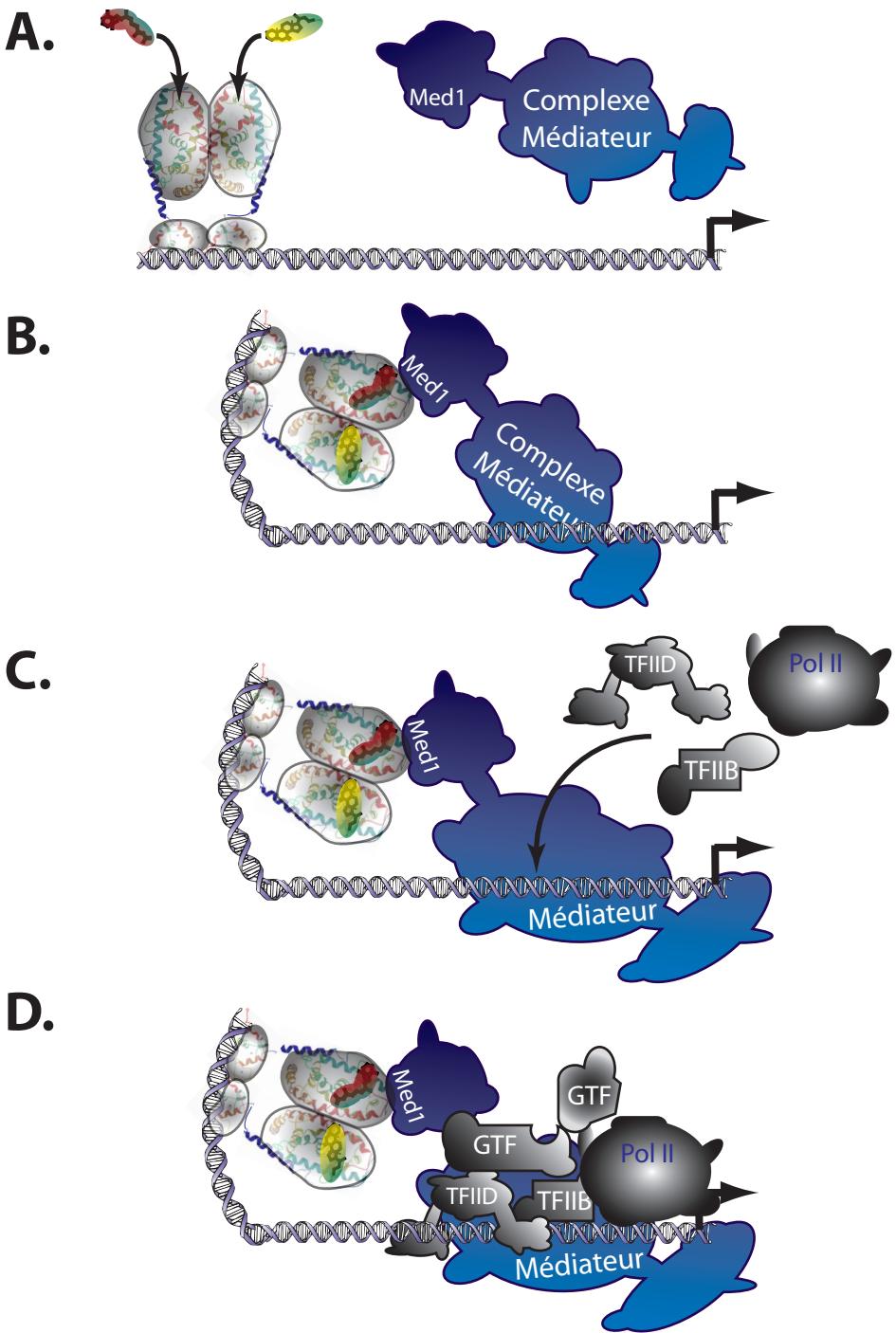
Une autre famille de puissants coactivateurs des RNs qui doit être évoquée dans ce chapitre sont les PGC-1 (PPAR $\gamma$  coactivator 1) dont on connaît actuellement trois membres : PGC-1 $\alpha$ , PGC-1 $\beta$  et PRC (PGC-1-related coactivator) et parmi lesquels PGC-1 $\alpha$  est celui qui a fait l'objet des recherches les plus intenses (pour des synthèses récentes, voir Finck and Kelly, 2006; Handschin and Spiegelman, 2006). En partie N-terminale, PGC-1 $\alpha$  possède un fort domaine d'activation qui interagit avec un complexe à activité HAT contenant SRC-1 et CBP/p300. On distingue ensuite une région inhibitrice constituée d'environ 200 acides aminés qui contient également des sites de fixation à des RNs ainsi qu'à d'autres facteurs de transcription. Enfin, la partie C-terminale de PGC-1 $\alpha$  contient un domaine d'interaction avec le complexe TRAP/DRIP/Mediator (voir paragraphe suivant pour une description de ce complexe qui établit un lien entre RN et machinerie transcriptionnelle basale) ainsi qu'un domaine de reconnaissance des ARNs et un domaine d'épissage (Monsalve et al., 2000). Le double recrutement d'activités HAT et du complexe Médiateur expliquerait l'exceptionnelle capacité de coactivation de PGC-1 $\alpha$  (Handschin and Spiegelman, 2006) et en ferait une sorte de plateforme d'interaction capable de recruter à la fois des activités de modifications des histones et d'établir un lien entre le facteur de transcription activé et la machinerie transcriptionnelle basale (via Médiateur). PGC-1 $\alpha$ , comme PGC-1 $\beta$ , est exprimé dans les tissus présentant une forte capacité oxydative comme le tissu adipeux brun, le cœur, les reins, le muscle squelettique ou le cerveau alors que l'expression de PRC est plus ubiquitaire. Ce coactivateur (PGC-1 $\alpha$ ), initialement identifié comme interagissant avec PPAR $\gamma$  dans le tissu adipeux brun (Puigserver et al., 1998), a depuis été montré comme interagissant avec de nombreux RNs dont les trois PPAR ( $\alpha$ ,  $\beta/\delta$  et  $\gamma$ ), RAR $\alpha$ , TR $\beta$ , FXR, LXR, PXR, GR, ER, HNF-4 $\alpha$  et les ERRs (pour des références, voir Finck and Kelly, 2006) ainsi qu'avec d'autres facteurs de transcription dont NRF-1 et -2 (nuclear respiratory factor), MEF-2 (myocyte enhancer factor-2) ou FOXO1 (forkhead box O1). Par ailleurs, PGC-1 $\beta$  mais pas PGC-1 $\alpha$ , agit comme coactivateur des facteurs de transcription SREBP (sterol regulatory element-binding protein Lin et al., 2005). L'intérêt particulier porté à PGC-1 $\alpha$  par plusieurs équipes de

recherche renommées semble avoir deux principales origines : 1) son rôle majeur dans la régulation du métabolisme cellulaire et en particulier dans la biogénèse et la fonction mitochondriales et le métabolisme oxydatif et 2) sa capacité à être induit au niveau transcriptionnel (par le froid dans le tissu adipeux brun et le muscle par exemple, Puigserver et al., 1998) et régulé au niveau post-traductionnel par divers signaux (phosphorylation, acétylation, méthylation d'arginines, interaction avec des répresseurs, Finck and Kelly, 2006; Handschin and Spiegelman, 2006) ce qui en fait un candidat intéressant pour des interventions pharmacologiques (pour des synthèses, voir Finck and Kelly, 2006; Handschin and Spiegelman, 2006). PGC-1 $\alpha$  apparaît ainsi comme un coactivateur limitant dont la régulation permettrait de relier divers signaux physiologiques ou alimentaires (principalement les situations nécessitant un accroissement de la production d'énergie par les mitochondries comme le froid, l'exercice ou le jeûne) à la fonction des RNs. Il est également intéressant de souligner que les coactivateurs PGC-1 semblent conférer aux RNs avec lesquels ils interagissent une certaine spécificité en termes de gènes cibles régulés. Les mécanismes générant cette spécificité de régulation ne sont pas encore complètement élucidés (Finck and Kelly, 2006; Handschin and Spiegelman, 2006).

Les différents corégulateurs que nous avons évoqués dans ce deuxième groupe possèdent ou recrutent des activités HAT ou de méthylation des histones conduisant généralement à une activation de la transcription mais il existe également des corépresseurs des RNs possédant ou recrutant des activités enzymatiques modifiant les histones. Alors que les récepteurs aux hormones stéroïdiennes (classe I) résident en général dans le cytoplasme sous la forme de complexes avec les protéines hsp70 et hsp90 (voir paragraphe I.6.), certains récepteurs de classe II, comme RAR $\alpha$  et TR $\alpha$ , sont capables de réprimer l'expression des gènes en l'absence de ligand. Ce phénomène, qualifié de transrépression ou de répression active, implique le recrutement de corépresseurs qui créent un environnement incompatible avec l'assemblage et le fonctionnement du complexe de pré-initiation. Il est distinct du phénomène de répression passive qui est basé sur la compétition pour la fixation à l'ADN ou bien pour le recrutement de partenaire de dimérisation ou de coactivateurs (McKenna et al., 1999). Depuis les travaux initiaux reconnaissant ce phénomène de transrépression, il a été montré que les RNs de classe I (ainsi que d'autres RNs) sont aussi capables de recruter des corépresseurs suite à la fixation d'antagonistes et que la fixation d'agonistes sur certains RNs pouvaient également, en fonction de multiples paramètres dont la séquence promotrice, les modifications post-traductionnelles et le niveau d'expression des partenaires, provoquer le recrutement de corépresseurs (Fernandes and White, 2003; Jepsen and Rosenfeld, 2002;

Privalsky, 2004). Les premiers corépresseurs qui ont été identifiés sont SMRT/TRAC (*silencing mediator for RAR and TR / TR-associated cofactor*) et NCoR/RIP13 (*NR-corepressor / receptor interacting protein 13*) qui appartiennent à la même famille de protéines et possèdent une architecture moléculaire commune. Globalement, on distingue leur partie N-terminale qui présente 3 à 4 domaines de répression de la transcription (RD pour *repression domain*) et leur partie C-terminale constituée de 2 ou 3 domaines d’interaction avec les RNs (RID, Jepsen and Rosenfeld, 2002; Privalsky, 2004). Les domaines RD agissent comme plateformes d’interaction avec des protéines possédant des activités histones désacétylases dont HDAC3, HDAC4, HDAC5 et HDAC7 mais également avec des protéines recrutant de telles activités et/ou agissant comme « charpente » du complexe corépresseurs dont TBL-1 (transducin-like protein 1), GPS2 (G protein pathway suppressor 2) ou encore mSin3 (mammalian switch-independent 3 protein) qui est capable de recruter, parmi d’autres partenaires, les protéines HDAC1 et HDAC2 (Jepsen and Rosenfeld, 2002; Privalsky, 2004). Outre le recrutement de ces protéines, NCoR et SMRT semblent également capables d’activer, au sein de ces complexes corépresseurs, certaines des activités histone désacétylases et peuvent également établir des contacts inhibiteurs directs avec certaines sous-unités du complexe de pré-initiation dont TFIIB et TAF30 (Privalsky, 2004). Les domaines RID (2 pour SMRT et 3 pour NCoR) contiennent chacun une boîte CoRNR (pour *corepressor of nuclear receptor*) constituée par le motif L/IXXI/VI qui est inclus dans un motif plus large que nous avons évoqué précédemment : le motif LXXXIXXXL/I. Ce motif forme un domaine à hélice  $\alpha$  étendue qui se fixe dans un sillon formé par les hélices 3/4/5/6 du LBD des RNs (Figure 12). La séquence de ces hélices, des motifs CoRNR et des acides aminés adjacents contribuent à déterminer l’affinité des divers RNs pour les différents motifs RID des corépresseurs.

De nombreux autres corépresseurs des RNs ont été identifiés qui sont capables de recruter des protéines HDAC mais également d’inhiber l’expression des gènes via des mécanismes indépendants de l’activité histone désacétylase. Citons par exemple le corépresseur LCoR (ligand-dependent corepressor, Fernandes et al., 2003; Fernandes and White, 2003) dont la fixation sur certains RNs apparaît dépendante d’un ligand agoniste. Ce corépresseur interagit via une unique boîte RN (motif LXXLL) avec un site du LBD chevauchant le site de fixation des SRC (Fernandes et al., 2003). Il semble agir, à différents degrés selon les récepteurs, via le recrutement de protéines HDAC (HDAC3 et HDAC6) mais également via une interaction avec les corépresseurs CtBP (C-terminal binding protein) connus pour inhiber la transcription via des mécanismes HDAC-dépendants et HDAC-



**Figure 16 : Représentation schématique du fonctionnement des complexes Médiateur**

A. Un hétérodimère de RNs est fixé sur un élément de réponse situé en amont du site d'initiation de la transcription et est exposé aux ligands.B.La fixation des ligands permet de générer des surfaces d'interaction avec un complexe Médiateur (voir la description de ces complexes dans le texte), principalement via la sous-unité Med1. Une courbure de l'ADN a été représentée pour illustrer la possibilité d'interactions entre RNs et machinerie transcriptionnelle basale y compris lorsque l'élément de réponse des RNs est situé loin en amont du site d'initiation de la transcription. L'induction de courbures de l'ADN suite à la fixation de certains RNs a par ailleurs été décrite dans la littérature. C. Le complexe Médiateur subit des modifications conformationnelles lui permettant de servir de plateforme pour la fixation des protéines du complexe de préinitiation (en particulier TFIID,TFIIB et Pol II).D. Le complexe de pré-initiation est assemblé au site actif de l'ARN polymérase II et peut initier la transcription. Le complexe Médiateur participe également à la modulation de l'activité du complexe de pré-initiation et vraisemblablement aussi à la réinitiation de la transcription. GTF : General Transcription Factor, TFIID : Transcription Factor II D (contient TBP), TFIIB : Transcription Factor II B, Pol II : ARN polymérase II.,

indépendants (Fernandes et al., 2003; Fernandes and White, 2003). Un autre corépresseur : RIP140 (receptor interacting protein of 140 KDa) est également capable de se fixer à de nombreux RNs activés par un ligand via plusieurs motifs LXXLL (pour une synthèse, voir Christian et al., 2006). Comme LCoR, ce corépresseur interagit à la fois directement avec des protéines HDAC et indirectement via les corépresseurs CtBP (Christian et al., 2006; Fernandes and White, 2003). RIP140 est fortement impliqué dans la régulation de nombreux processus métaboliques et il est intéressant de noter qu'il intervient dans la régulation, en sens inverse, de nombreux gènes modulés par PGC-1 $\alpha$  (Christian et al., 2006). Pour conclure et synthétiser la présentation de ce deuxième groupe de corégulateurs des RNs, rappelons qu'ils se caractérisent par leur capacité à modifier de manière covalente les histones, qu'ils fonctionnent généralement sous la forme de larges complexes multiprotéiques et que, très souvent, ils sont régulés au niveau transcriptionnel ou post-traductionnel, constituant ainsi des plateformes d'intégration de multiples signaux et participant à l'interaction entre les RNs et les autres voies de signalisation.

#### I.5.4 Complexes Médiateur

Le troisième groupe de corégulateurs des RNs qui peut être distingué correspond aux complexes regroupés sous le nom de Médiateur (Mediator en anglais) qui participent principalement au **recrutement du complexe ARN polymérase II** et plus généralement à **promouvoir l'assemblage et l'activité du complexe de pré-initiation de la transcription** ([Figure 16](#)). Médiateur est un volumineux (environ 1,2 MDa) complexe multiprotéique qui peut prendre différentes formes en fonction de sa composition en sous-unités (plus de 30 sous-unités parfois) et qui est largement conservé chez les eucaryotes (Belakavadi and Fondell, 2006; Blazek et al., 2005). Chez les mammifères, il a été isolé par diverses approches (synthétisé dans Blazek et al., 2005; Conaway et al., 2005; Malik and Roeder, 2005) qui ont conduit à lui donner plusieurs noms différents (les liens entre les différentes nomenclatures peuvent être consulté dans Belakavadi and Fondell, 2006) : TRAP (TR-associated protein complex Fondell et al., 1996), SMCC (Srb and Mediator protein-containing complex, Gu et al., 1999) ou encore DRIP (VDR-interacting protein complex, Rachez et al., 1998). Médiateur semble avoir une structure modulaire consistant en une tête constituante le principal site de contact avec l'ARN polymérase II (Pol II), un corps participant également à la fixation du complexe à Pol II et qui servirait de site de contact avec la partie C-terminale de Pol II et une queue probablement impliquée aussi bien dans l'activation que dans la répression de la transcription (Blazek et al., 2005; Chadick and Asturias, 2005; Malik and Roeder, 2005). Un

module supplémentaire contenant le dimère Cdk8-Cyclin C ainsi que les grosses sous-unités MED12 et MED13 et qui jouerait un rôle dans la répression de la transcription (Blazek et al., 2005; Conaway et al., 2005; Malik and Roeder, 2005) est présent dans certains complexes Médiateur qualifiés de « grands Médiateurs » (également nommés NAT / negative regulator of activated transcription, ARC / activator-recruited coactivator ou Med) et absent dans les complexes qualifiés de « petits Médiateurs » (comme PC2 / positive coactivator 2 ou CRSP / cofactor required for Sp1). Depuis l'identification de complexes Médiateur interagissant avec TR (Fondell et al., 1996) et VDR (Rachez et al., 1998) et stimulant leur capacité d'activation de la transcription, des interactions entre des sous-unités de Médiateur et de nombreux RNs ont été démontrées (pour une synthèse, voir Blazek et al., 2005). Il semble que la sous-unité MED1 (également appelée TRAP220, DRIP205 ou PBP pour PPAR-binding protein) soit majoritairement impliqué, via deux boîtes RN (motifs LXXLL), dans l'interaction entre les RNs activés par un ligand et le complexe Médiateur (Belakavadi and Fondell, 2006; Blazek et al., 2005; Malik and Roeder, 2005). D'autres sous-unités dont MED14 (TRAP170, DRIP150) et MED21 (SRB7) interagissent également avec certains RNs, mais les interactions avec MED14 ne semblent pas dépendre de la fonction AF-2 des RNs (Belakavadi and Fondell, 2006; Blazek et al., 2005). MED1, MED14 ainsi que d'autres sous-unités interviennent également dans l'interaction des complexes Médiateur avec d'autres facteurs de transcription (Blazek et al., 2005; Conaway et al., 2005; Malik and Roeder, 2005) dont AhR (MED1), p53 (MED1 et MED17) et SREBP-1a (MED14). Les mécanismes précis par lesquels Médiateur régule la transcription par l'ARN polymérase II ne sont pas complètement élucidés. Néanmoins, des approches biochimiques et structurales ont permis de démontrer définitivement que ce complexe interagit avec plusieurs sous-unités de l'ARN polymérase II et avec d'autres facteurs généraux de la transcription et on considère actuellement que son rôle majeur consiste à initier l'assemblage du complexe de pré-initiation (Belakavadi and Fondell, 2006; Blazek et al., 2005; Malik and Roeder, 2005). Une fois ce complexe assemblé, Médiateur connaît des modifications structurales et topologiques qui le conduiraient à moduler la fonction de l'ARN polymérase II, notamment en stimulant l'activité de TFIIH qui agit comme kinase sur la partie C-terminale de l'ARN polymérase II, un mécanisme nécessaire à l'initiation de la transcription (Belakavadi and Fondell, 2006; Malik and Roeder, 2005). Des hypothèses restent encore ouvertes sur cet aspect du fonctionnement de Médiateur, ainsi que sur ses rôles éventuels dans la ré-initiation de la transcription, dans les modifications post-traductionnelles qu'il engendrerait sur les facteurs généraux de la transcription, les facteurs de transcription en amont ou leurs corégulateurs et dans la répression de la

transcription (Belakavadi and Fondell, 2006; Blazek et al., 2005; Conaway et al., 2005; Malik and Roeder, 2005). Il est également important de rappeler à ce stade qu'un certain nombre d'interactions directes entre certains RNs et des composants de la machinerie transcriptionnelle basale ont également été décrits (pour une synthèse, voir Laudet and Gronemeyer, 2002; McKenna et al., 1999), suggérant que Médiateur ne serait pas seul responsable de ce type d'interactions.

### I.5.5. Autres corégulateurs

Enfin, on peut regrouper dans un quatrième groupe les nombreux corégulateurs des RNs dont la fonction n'est pas (ou peu) élucidée ou qui fonctionnent de manière atypique par rapport aux corégulateurs évoqués précédemment. Des effets dépendants du type cellulaire, du RN, du ligand ou du promoteur étudiés ou encore une expression restreinte à de rares tissus ou dans des conditions physiologiques particulières rendent parfois difficile l'identification des corégulateurs et la caractérisation de leur fonction. Un aperçu très complet des coactivateurs de ce groupe (et des trois groupes précédemment évoqués) peut être consulté dans (McKenna et al., 1999), dans (Laudet and Gronemeyer, 2002) ou bien sur [www.nursa.org](http://www.nursa.org). Nous n'évoquerons dans ce paragraphe que deux exemples.

Le premier correspond au coactivateur SRA (steroid receptor RNA activator, voir [Figure 14](#)) qui a la particularité (unique à l'heure actuelle pour un corégulateur) d'être un ARN fonctionnant directement sous la forme d'un transcrit (et non d'une protéine) comme coactivateur des RNs de classe I (Lanz et al., 1999). Bien que la démonstration ait été apportée que SRA fonctionne bien comme coactivateur sous sa forme ARN, des transcrits alternatifs produits par épissage et codant des protéines ont également été rapportés (Emberley et al., 2003; Kawashima et al., 2003). SRA agit comme coactivateur de ER, AR, GR et PR de manière dépendante de leur fonction AF-1 (Lanz et al., 1999) et pourrait jouer un rôle dans le développement de certaines tumeurs du sein, de l'utérus ou des ovaires où il est surexprimé (Lanz et al., 2003). SRA agit également comme coactivateur de TR $\alpha$ 1, TR $\beta$ 1 (Xu and Koenig, 2004) et RAR $\gamma$  (Zhao et al., 2004). Il forme un complexe ribonucléoprotéique stable avec SRC-1 qui est recruté par des RNs de classe I et II suite à leur activation par un ligand (Lanz et al., 1999) et interagit également avec les corépresseurs SHARP (SMRT/HDAC1-associated repressor protein, Shi et al., 2001) et SLIRP (SRA stem-loop interacting RNA binding protein, Hatchell et al., 2006) qui réduisent sa capacité de coactivation. Des travaux récents ont également mis en évidence que SRA faisait l'objet de modifications post-transcriptionnelles (pseudouridylation) par des pseudouridine synthases

(Pus1p et Pus3p, Zhao et al., 2004; Zhao et al., 2007) qui interagissent avec des RNs et que ces modifications étaient nécessaires à l'activité de SRA, probablement en affectant sa structure secondaire (Lanz et al., 2002) et sa capacité d'interaction avec des corégulateurs. Bien qu'on ne connaisse pas précisément les mécanismes impliqués dans la régulation de l'activité des RNs par SRA, on considère actuellement qu'il pourrait agir comme composant structural (une « charpente ») du complexe de coactivateurs et spécifier le recrutement de certaines protéines possédant des domaines de reconnaissance des ARNs.

Le deuxième exemple qui sera évoqué dans ce paragraphe est le cas de SHP (NR0B2, [Table 1](#), Seol et al., 1996). Ce RN a la particularité de ne pas posséder de domaine de fixation à l'ADN (DBD) ce qui en a fait un objet de nombreuses recherches pour sa capacité à inhiber les voies de signalisation des autres RNs. SHP possède deux boîtes RN (motifs de type LXXLL) et interagit avec de très nombreux RNs via leur domaine AF-2. Il apparaît donc comme un RN singulier (absence de DBD) mais peut également être considéré comme un corépresseur atypique (interaction avec les RNs via des motifs LXXLL plutôt que par dimérisation des LBDs) car c'est lui-même un RN. Des interactions entre SHP et plus de 20 RNs différents ont été rapportées (pour une synthèse, voir Bavner et al., 2005) et les mieux caractérisées sont sans doute celles concernant LRH-1, les LXR, les ERR, HNF4, les PPAR, GR, les ER et les RXR. SHP semble également réprimer l'activité d'autres facteurs de transcription dont AhR via une interaction directe avec son partenaire ARNT (Klinge et al., 2001). Les mécanismes moléculaires impliqués dans la répression par SHP sont encore mal définis, probablement en partie car il existe plusieurs et qu'ils semblent dépendants du promoteur étudié (Boulias et al., 2005). Deux principaux mécanismes de répression ont été proposés (pour une synthèse, voir Bavner et al., 2005) : 1) une compétition entre SHP et les coactivateurs pour la fixation à la fonction AF-2 des RNs (voir par exemple Lee and Moore, 2002), 2) une répression active qui impliquerait le recrutement d'activités de modification (HDAC et méthyltransférases) et/ou de remodelage des histones par SHP (Boulias and Talianidis, 2004; Kemper et al., 2004). Notons que d'autres mécanismes ont également été rapportés comme une inhibition de la fixation des RNs à l'ADN (voir par exemple Boulias et al., 2005; del Castillo-Olivares and Gil, 2001), une modification de la localisation intranucléaire des RNs (Borgius et al., 2002) ou une fixation directe de SHP sur l'histone H3 modifiée par des corépresseurs recrutés via SHP (Boulias and Talianidis, 2004). L'étude des gènes dont la régulation est affectée par SHP a conduit à identifier de nombreuses voies régulées par ce corépresseur, principalement liées au métabolisme du cholestérol et des acides biliaires, à la lipogenèse, à la néoglucogenèse, à la stéroidogénèse et au métabolisme des

xénobiotiques (pour une liste des gènes modulés et les références correspondantes, consulter Bavner et al., 2005). Une série d'études qui a permis des avancées significatives dans la compréhension du rôle de SHP concerne son implication dans la boucle de régulation de la synthèse des acides biliaires. La synthèse d'acides biliaires à partir du cholestérol est initiée soit par la cholestérol 7 $\alpha$ -hydroxylase (CYP7A1, voie de synthèse dite « classique » ou « neutre »), soit par l'enzyme mitochondriale stérol 27-hydroxylase (CYP27A1, voie de synthèse dite « acide ») qui constituent les enzymes limitantes de ces deux voies de synthèse respectives. En réponse à une augmentation des acides biliaires, l'expression de CYP7A1 est inhibée (rétrocontrôle négatif). Ce mécanisme implique l'activation de FXR par les acides biliaires, conduisant à une induction de l'expression de SHP qui, ensuite, interagit avec LRH-1 (pour une synthèse sur LRH-1, voir Fayard et al., 2004) et inhibe son activité sur les promoteurs de CYP7A1 et de SHP lui-même (Goodwin et al., 2000; Lu et al., 2000). La génération de souris déficientes pour SHP (Kerr et al., 2002b; Wang et al., 2002) a permis de démontrer *in vivo* l'existence de ce rétrocontrôle négatif et de découvrir que d'autres voies, indépendantes de SHP, agissent également pour contrôler de manière fine l'expression des enzymes clefs de la synthèse des acides biliaires (pour des synthèses, voir Bavner et al., 2005; Davis et al., 2002). Il est également intéressant de souligner le parallèle qui existe entre l'interaction SHP/LRH-1 et l'interaction Dax-1/SF-1 (Goodwin et al., 2000; Lu et al., 2000). SF-1 (NR5A1) est le RN le plus proche de LRH-1 (NR5A2) en termes de séquence et régule notamment l'expression de nombreuses enzymes impliquées dans la biosynthèse des hormones stéroïdiennes (pour une synthèse, voir Hammer and Ingraham, 1999). Dax-1 (NR0B1) présente, lui, une séquence proche de celle de SHP (NR0B2) et, comme lui, ne possède pas de domaine de fixation à l'ADN. Il a été montré que, comme pour le couple LRH-1/SHP, l'activité transcriptionnelle de SF-1 est inhibée via une interaction directe avec Dax-1 (Hammer and Ingraham, 1999). D'autres corégulateurs dont le fonctionnement semble atypique dans l'état actuel des connaissances pourraient également être évoqués dans ce paragraphe et des informations complémentaires peuvent être trouvées sur [www.nursa.org](http://www.nursa.org) ou dans diverses synthèses incluant de larges inventaires de corégulateurs (voir par exemple Laudet and Gronemeyer, 2002; Privalsky, 2004; Robyr et al., 2000). Par ailleurs, une excellente synthèse de l'ensemble des cofacteurs (appartenant aux quatre groupes que nous avons évoqués ici) interagissant avec les récepteurs PPAR (Chapitre II) a été récemment publiée (Yu and Reddy, 2007).

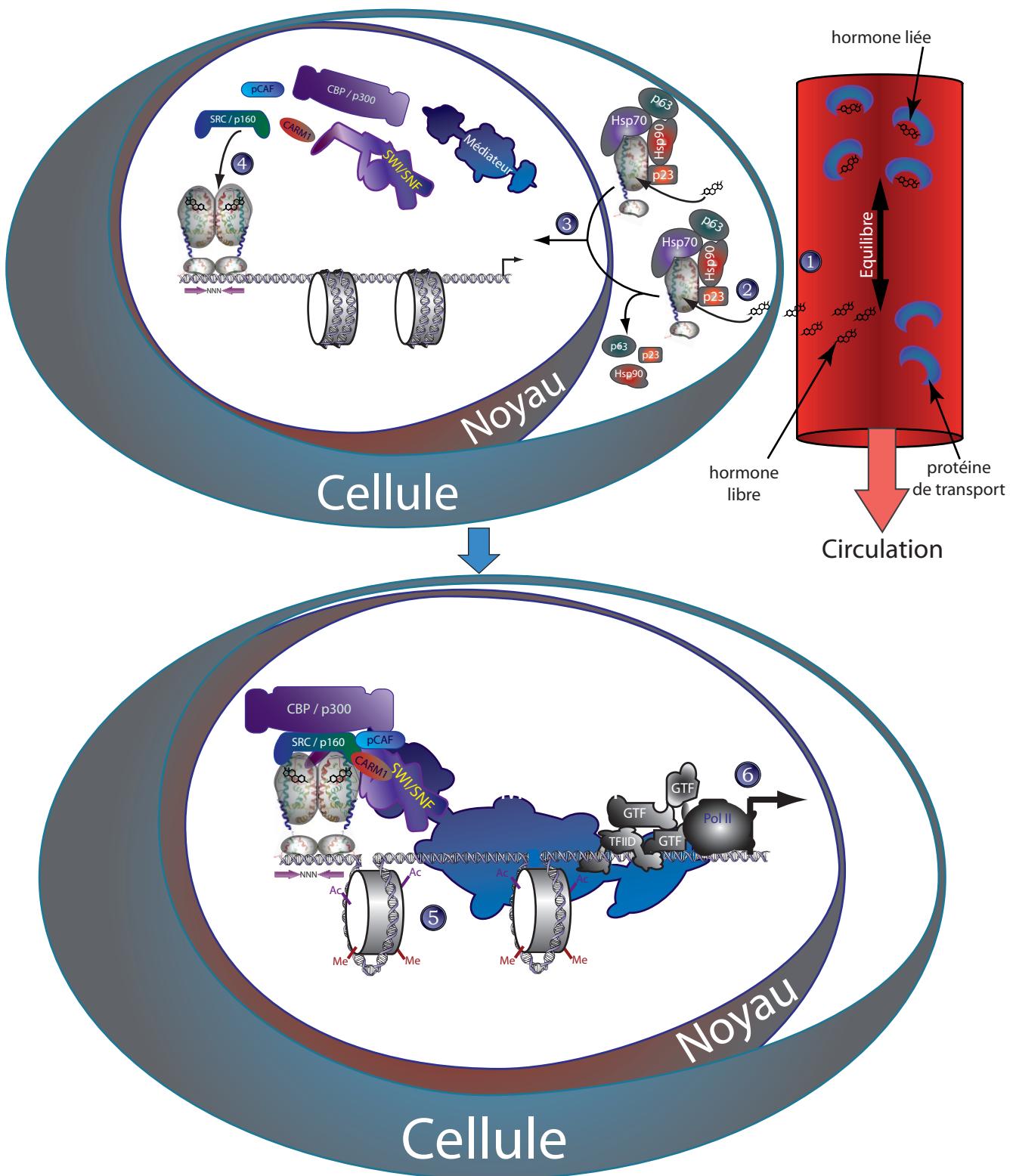


Figure 17 : Schéma de fonctionnement des voies de signalisation des RNs de classe I

- Dans la circulation, il existe un équilibre entre forme liée aux protéines de transport et forme libre des hormones stéroïdiennes. Suite à la pénétration de l'hormone dans la cellule, l'hormone se fixe à son récepteur nucléaire cible qui est localisé dans le cytoplasme, sous la forme d'un complexe avec diverses protéines chaperones et/ou de choc thermique. Cette fixation engendre des changements conformationnels dont les premières conséquences sont de libérer les RNs de leurs complexes cytoplasmiques, de provoquer leur migration vers le noyau, leur homodimérisation et leur fixation sous la forme d'homodimères à leurs éléments de réponse situés dans les promoteurs de leurs gènes cibles. Les modifications conformationnelles engendrées par la fixation du ligand génèrent, sur les RNs, des surfaces d'interaction avec divers corégulateurs (ici, dans le cas d'un ligand agoniste, principalement des coactivateurs). Les effets de ces coactivateurs sur la chromatine (remodelage ATP-dépendant et modifications covalentes des histones) facilitent voire permettent l'accès des facteurs généraux de transcription au promoteur et le complexe Médiateur participe à l'interaction entre l'homodimère de RNs et la machinerie transcriptionnelle basale. L'accès à l'ADN, l'assemblage et le fonctionnement du complexe de préinitiation sont facilités, ce qui se traduit par une augmentation de la transcription du gène cible

### I.5.6. Dynamique de fonctionnement des corégulateurs

Ainsi, divers types de corégulateurs agissent, généralement sous la forme de complexes, pour assurer la transmission du signal reçu par les RN via la fixation du ligand, à la machinerie transcriptionnelle basale. Comme pour les RNs eux-mêmes, le niveau d'expression des corégulateurs, leur protéolyse, diverses modifications post-traductionnelles (phosphorylation, acétylation, méthylation, sumoylation) et/ou leur localisation intracellulaire participent également à la régulation fine de l'activité des RNs en fonction du contexte cellulaire et offrent des possibilités supplémentaires d'interaction avec d'autres voies de signalisation (Hermanson et al., 2002; Privalsky, 2004; Wu et al., 2005a). Un schéma général de l'ordre de recrutement et de fonctionnement de ces différents corégulateurs est encore difficile à établir et dépend vraisemblablement du récepteur étudié, du contexte cellulaire et de l'architecture du promoteur considéré (Urnov and Wolffe, 2001). Cependant, des travaux récents, utilisant notamment les techniques de FRAP (fluorescence recovery after photobleaching, étude de la dynamique des protéines dans le noyau à l'échelle de la seconde) et d'immunoprecipitation de la chromatine (ChIP, étude des cycles de fixation des protéines à l'ADN à l'échelle de quelques minutes) semblent dégager petit à petit un schéma de fixation cyclique des RNs et de leurs coactivateurs sur les promoteurs des gènes cibles (pour une présentation et une discussion du modèle, voir Gelman et al., 2006; Hager et al., 2004; Metivier et al., 2006). Ce modèle réconcilie des événements de nature stochastique, tenant compte de la dynamique des protéines dans le noyau et de leur fixation rapide, le plus souvent non productive, sur le promoteur, et des événements de nature déterministe, tenant compte des associations productives de protéines au promoteur permettant de faire avancer le système de manière orienté vers la fixation ultérieure d'autres protéines régulatrices.

## I.6. MODELES GENERAUX DU FONCTIONNEMENT DES VOIES DE SIGNALISATION DES RECEPTEURS NUCLEAIRES

A partir des informations synthétisées précédemment, deux principaux modèles de fonctionnement des voies de signalisation des RNs ont été proposés dans la littérature. Le premier modèle ([Figure 17](#)) s'applique assez bien aux RNs de classe I, c'est à dire aux récepteurs ER, PR, AR, GR et MR. La plupart de ces récepteurs sont, en l'absence d'hormone, localisés dans le cytoplasme ou plutôt, comme nous l'avons évoqué précédemment, effectuent des navettes entre le noyau et le cytoplasme. Ils se présentent sous la forme de complexes multiprotéiques contenant des protéines agissant comme chaperones et

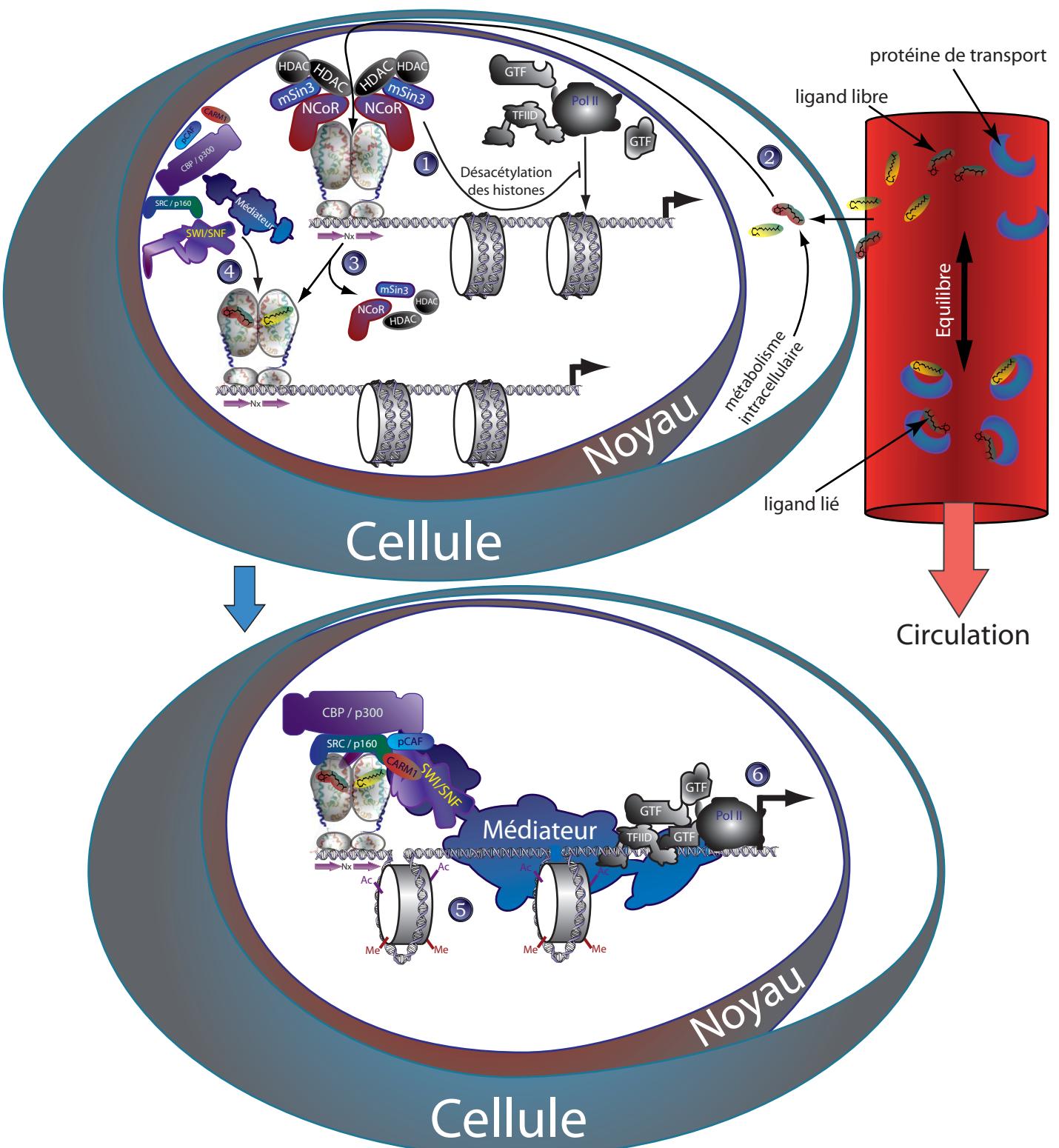


Figure 18 : Schéma de fonctionnement des voies de signalisation des RNs de classe II

① En l'absence de ligand, les RNs fixés sous la forme d'un hétérodimère avec RXR sur leur élément de réponse sont associés à un complexe de corépresseurs. L'action des corépresseurs sur les histones (principalement désacétylation) conduit à un compactage de l'ADN, illustré par les nucléosomes, qui réduit l'accessibilité des facteurs généraux de transcription au site d'initiation de la transcription. ② Les ligands, provenant du métabolisme intracellulaire ou de la circulation (l'équilibre entre forme libre et forme liée est une large simplification en ce qui concerne les ligands des RNs de classe II et ne prend notamment pas en compte le transport des ligands par les lipoprotéines et l'existence de transporteurs plus ou moins spécifiques), se fixent alors de manière spécifique à leurs récepteurs cibles. ③ Cette fixation engendre des changements conformationnels sur les RNs qui réduisent leur affinité pour les corépresseurs et conduisent ④ à générer des surfaces d'interaction avec divers coactivateurs. ⑤ Les effets de ces coactivateurs sur la chromatine facilitent voire permettent l'accès des facteurs généraux de transcription au promoteur et le complexe Médiateur participe à l'interaction entre l'hétérodimère de RNs et la machinerie transcriptionnelle basale. ⑥ L'accès à l'ADN, l'assemblage et le fonctionnement du complexe de préinitiation sont facilités, ce qui se traduit par une augmentation de la transcription du gène cible.

participant au trafic intracellulaire des RNs comme les protéines de choc thermique (Hsp, pour heat shock protein) Hsp90 et Hsp70 ainsi que d'autres protéines comme des immunophilines (pour une synthèse, voir Pratt and Toft, 1997). Sous la forme de ces hétérocomplexes, les récepteurs aux hormones stéroïdiennes possèdent la capacité de fixer leurs ligands respectifs mais sont incapables de se lier à l'ADN. L'exposition des récepteurs à leurs hormones respectives, se traduit par la dissociation de ces complexes, la translocation des RNs au noyau, leur fixation sur des éléments de réponse spécifiques et leur homodimérisation. La fixation du ligand permet, comme nous l'avons vu, la génération de surfaces d'interaction avec des coactivateurs dont les diverses actions (simultanées ou successives) au niveau de la chromatine et des facteurs généraux de la transcription permettent un accroissement de la fonction du complexe de pré-initiation se traduisant par une augmentation de la transcription des gènes cibles. En présence d'un antagoniste des RNs de classe I, ce sont des corépresseurs qui peuvent être recrutés et induire une diminution de la transcription des gènes cibles. Le second modèle ([Figure 18](#)) correspond plutôt au mode de fonctionnement de la voie de signalisation des RNs de classe II. Certains de ces RNs et notamment TR et RAR agissent comme répresseurs de l'expression de leurs gènes cibles en l'absence de ligand activateur. Ils forment un hétérodimère sur les promoteurs de leurs gènes cibles qui est associé à un complexe corépresseur dont la principale action (la désacétylation des histones) conduit à un compactage de l'ADN sous la forme de nucléosomes s'opposant à une transcription efficace des gènes cibles. La fixation des ligands conduit à une dissociation du complexe de corépresseurs et au recrutement de coactivateurs dont l'action a été évoquée précédemment et qui se traduit, là aussi, par un accroissement de la transcription des gènes cibles.

Ces deux modèles ne prennent pas en compte les aspects dynamiques et cycliques évoqués à la fin du paragraphe précédent et peuvent donc être qualifiés de modèles « statiques » ou généraux. Naturellement, ces modèles ne sont pas systématiquement valides tels quels. Dans des cas particuliers, de nombreuses exceptions à ces règles canoniques sont décrites dans la littérature. Citons par exemple le récepteur CAR (NR1i3, [Table 1](#)), appartenant à la classe II car il fonctionne en hétérodimère avec RXR mais qui réside dans le cytoplasme au niveau du foie (mais pas dans la plupart des lignées cellulaires, Kawamoto et al., 1999; Zelko et al., 2001) sous la forme d'un complexe avec Hsp90 et CCRP (cytoplasmic CAR retention protein, Kobayashi et al., 2003; Yoshinari et al., 2003). Ce RN possède une activité constitutive et peut ainsi être activé par des molécules comme le phénobarbital (Kawamoto et al., 1999; Wei et al., 2000; Zelko et al., 2001) ou la phénytoïne (Jackson et al.,

2004) qui provoquent sa migration au noyau sans pour autant se fixer au niveau de son LBD. CAR possède néanmoins aussi des ligands activateurs comme le TCPOBOP ([Table 6](#), spécifique de la forme murine de CAR, Tzameli et al., 2000) ou le CITCO ([Table 6](#), spécifique de la forme humaine de CAR, Maglich et al., 2003) qui peuvent accroître son activité sur les promoteurs de ses gènes cibles. Il possède également des agonistes inverses comme certains métabolites de l'androstane ([Table 5](#), Forman et al., 1998), capables d'inhiber son activité constitutive. Citons également le cas de VDR qui appartient également à la classe II des RNs car il fonctionne en hétérodimère avec RXR mais semble également capable de fonctionner sous la forme d'homodimères (voir par exemple Freedman et al., 1994; Seoane and Perez-Fernandez, 2006; Takeshita et al., 2000). Comme TR et RAR, VDR peut se fixer à l'ADN en l'absence de ligand (Ross et al., 1993) mais une fraction non négligeable de VDR est également localisée dans le cytoplasme et subit une translocation au noyau suite à une exposition à la vitamine D3 (Michigami et al., 1999; Prufer and Barsony, 2002; Yasmin et al., 2005). En conclusion, malgré leurs imperfections, ces deux modèles relativement simples présentent l'avantage d'intégrer une grande part de l'information sur le fonctionnement global de ces voies de signalisation et leur caractère didactique facilite grandement la diffusion des connaissances acquises sur les RNs.

## Chapitre II : Le récepteur PPAR $\alpha$

Après avoir présenté la superfamille des RNs, ce chapitre présente plus spécifiquement le récepteur PPAR $\alpha$  (*peroxisome proliferator-activated receptor  $\alpha$*  ou NR1C1) qui a été au cœur de la plupart des études menées dans le cadre de mon doctorat. Nous évoquerons également au début de ce chapitre les récepteurs PPAR $\beta/\delta$  et PPAR $\gamma$  qui, avec le récepteur PPAR $\alpha$ , constituent la sous-famille 1C des RNs (Table 1). Après avoir présenté les sites d'expression et les ligands des trois PPARs, le fonctionnement de la voie de signalisation de PPAR $\alpha$  sera présenté plus en détail. Ensuite, nous aborderons les principales fonctions et cibles de ce récepteur qui concernent particulièrement le métabolisme énergétique mais également le métabolisme des xénobiotiques et la régulation de l'inflammation.

### II.1. CLONAGE, LOCALISATION CHROMOSOMIQUE ET EXPRESSION

#### II.1.1. Clonage des récepteurs PPAR

PPAR $\alpha$  a été le premier des trois PPAR à être identifié, chez la souris en 1990 (Issemann and Green, 1990). Ultérieurement, l'équipe du Pr Wahli, identifia les trois PPAR chez le xénope (Dreyer et al., 1992) puis l'isoforme  $\alpha$  fut clonée chez le rat (Gottlicher et al., 1992) et chez l'homme (Sher et al., 1993). Depuis, le récepteur PPAR $\alpha$  a été cloné chez de nombreuses espèces (pour une synthèse, voir Laudet and Gronemeyer, 2002). La première identification du récepteur PPAR $\beta$  (alors dénommé NUC1) chez un mammifère a été réalisée chez l'homme (Schmidt et al., 1992). Peu après son clonage chez la souris (Kliewer et al., 1994), l'équipe du Dr Grimaldi montra que PPAR $\beta$  (alors dénommé FAAR) était exprimé au cours de la différenciation d'adipocytes en culture, était induit sous l'effet d'un traitement par des acides gras et agissait comme médiateur des effets de ces acides gras sur l'expression de deux gènes impliqués dans leur transport (Amri et al., 1995). Le récepteur PPAR $\gamma$  de mammifère a été initialement cloné chez la souris par les équipes respectives des Dr Reddy (Zhu et al., 1993) et O'Malley (Chen et al., 1993). Suite aux travaux de l'équipe du Dr Spiegelman, l'isoforme PPAR $\gamma$ 2, exprimée presque exclusivement dans le tissu adipeux et contenant 30 acides aminés supplémentaires en partie N-terminale par rapport à PPAR $\gamma$ 1 (28 acides aminés supplémentaires chez l'homme), a été rapidement impliquée, sous la forme d'un dimère avec RXR, comme régulateur majeur d'ap2 (également appelée ALBP pour *adipocyte lipid-binding protein* ou FABP4 pour *fatty acid binding protein 4*), un gène

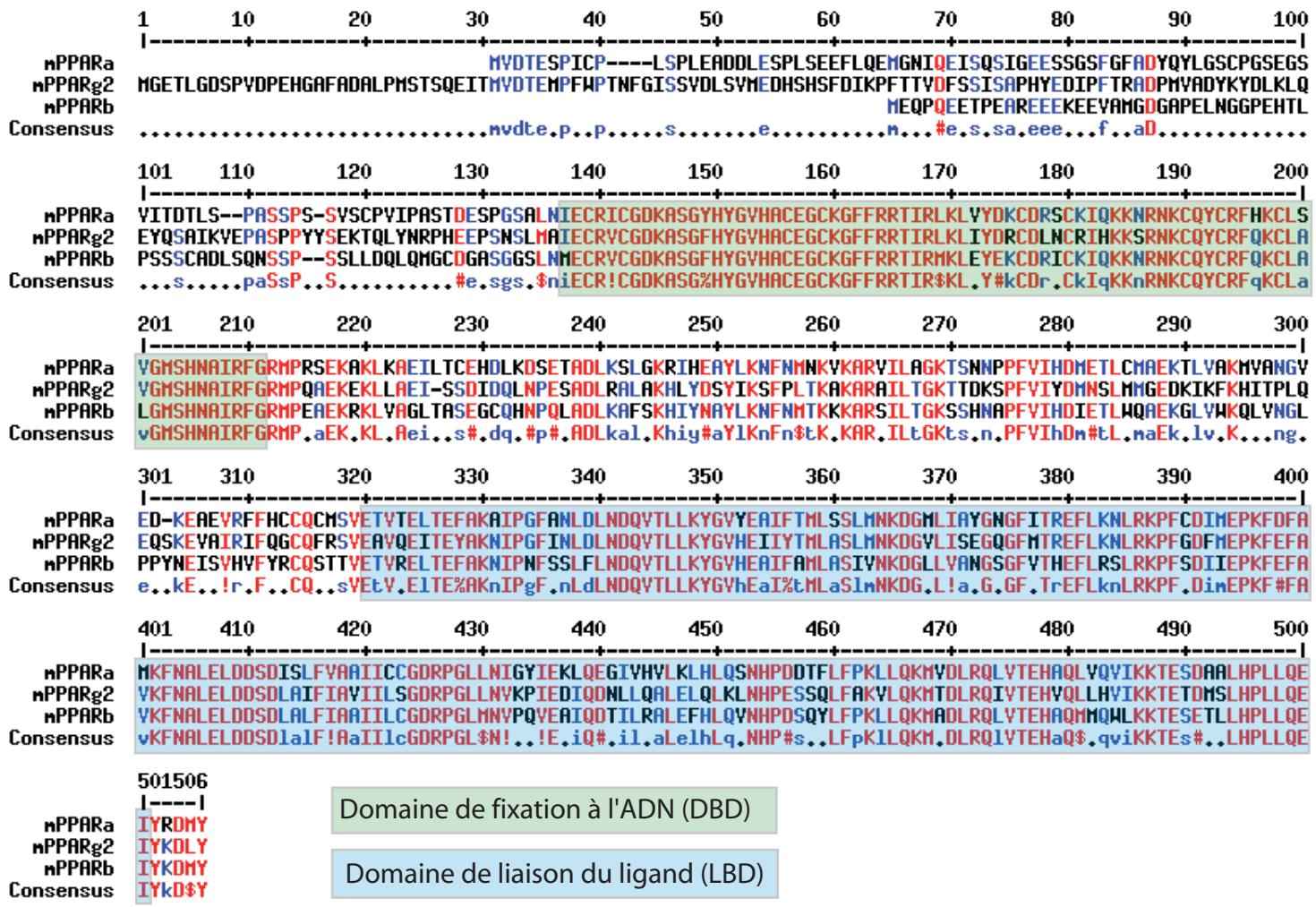


Figure 19 : Alignement des séquences protéiques des PPAR murins

Les références des trois séquences protéiques sont présentées en Table 1. La séquence de PPARγ1 n'est pas présente car elle correspond simplement à la séquence de PPARγ2 sans les 30 premiers acides aminés. L'alignement a été réalisé sous Multalin (<http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html>) avec les paramètres par défaut. Les domaines de fixation à l'ADN (domaine C ou DBD) et de liaison du ligand (domaine E ou LBD) sont surlignés en vert et bleu respectivement. Leur définition est basée ici sur les annotations des séquences protéiques dans la base de données GenBank.

marqueur de la différenciation adipocytaire (Tontonoz et al., 1994a; Tontonoz et al., 1994b) et comme stimulateur de l’adipogénèse (Tontonoz et al., 1994c).

Les trois PPAR présentent entre 78 et 86% d’identité dans leur domaine de fixation à l’ADN et entre 63 et 71% dans leur domaine de fixation du ligand ([Figure 19](#)). De manière surprenante, il semble que les PPAR aient connu une vitesse d’évolution particulièrement rapide car, pour une isoforme donnée, il existe des différences conséquentes entre les séquences issues de diverses espèces. Il est tentant de rapprocher ce taux d’évolution rapide de la large diversité de ligands des PPAR et des différences de reconnaissance de certains ligands entre les PPAR de différentes espèces (pour une synthèse, voir Desvergne and Wahli, 1999).

### **II.1.2. Localisation chromosomique et organisation des gènes PPAR**

Les localisations chromosomiques des trois gènes PPAR ont été identifiées chez l’homme et la souris (Desvergne and Wahli, 1999). Chez la souris, les gènes PPAR $\alpha$ , PPAR $\beta/\delta$  et PPAR $\gamma$  sont situés respectivement sur les chromosomes 15, 17 et 6. Chez l’homme, ils sont localisés respectivement sur les chromosomes 22, 6 et 3.

La structure générale des trois gènes PPAR est assez semblable. Globalement, un exon code la région A/B, deux exons codent le domaine C de liaison à l’ADN, chacun de ces deux exons codant l’un des deux doigts à atome de zinc, un exon code la région charnière D et deux exons codent le domaine E de liaison du ligand. La localisation des introns est bien conservée dans les différentes structures des gènes obtenus chez la souris, l’homme et le xénope (Beamer et al., 1997; Fajas et al., 1997; Gearing et al., 1994; Krey et al., 1993; Zhu et al., 1995). Le gène PPAR $\alpha$  murin, qui couvre environ 30 Kbp d’ADN génomique (Gearing et al., 1994), est plus petit que le gène PPAR $\gamma$  qui s’étend, chez l’homme comme chez la souris, sur plus de 100 Kbp d’ADN génomique (Fajas et al., 1997; Zhu et al., 1995). Le gène murin PPAR $\beta$  couvre, lui, environ 70 Kb ([www.ensembl.org](http://www.ensembl.org)).

Le gène mPPAR $\alpha$  comprend 8 exons au total, les deux premiers correspondant à la partie 5'-non traduite et le dernier exon codant le LBD comprenant la partie 3'-non traduite du transcript (Gearing et al., 1994). De manière similaire, le gène mPPAR $\beta/\delta$  comporte également 8 exons dont les deux premiers constituent, avec un morceau du troisième exon, la partie 5'-non traduite et dont le dernier exon code une partie du LBD et comporte la partie 3'-non traduite. Le gène PPAR $\gamma$ , chez l’homme ou chez la souris, génère trois transcrits différents, notés PPAR $\gamma$ 1, PPAR $\gamma$ 2 et PPAR $\gamma$ 3 qui diffèrent par leur partie 5' du fait de l’utilisation de

divers promoteurs et d'épissages alternatifs (pour une comparaison des trois transcrits, voir Desvergne and Wahli, 1999). Six exons sont en commun entre ces trois transcrits. PPAR $\gamma$ 1 possède deux exons supplémentaires qui composent la partie 5'-non traduite. Le premier exon de PPAR $\gamma$ 2 comporte la partie 5'-non traduite et code les 30 acides aminés (28 chez l'homme) supplémentaires que possède cette isoforme en partie N-terminale. Le troisième transcrit, PPAR $\gamma$ 3, code la même protéine que PPAR $\gamma$ 1 mais est contrôlé par un promoteur alternatif situé entre les deux premiers exons de PPAR $\gamma$ 1.

Signalons qu'une forme tronquée de hPPAR $\alpha$ , apparemment absente chez les rongeurs, a été décrite. Elle serait issue d'un épissage alternatif et pourrait interférer avec la voie PPAR $\alpha$  via une activité de dominant négatif (Gervois et al., 1999b). Enfin, une protéine reliée immunologiquement à PPAR $\gamma$ 2 mais de taille plus réduite a été identifiée dans la mitochondrie (Casas et al., 2000a). Elle serait régulée par certains proliférateurs de peroxysomes (voir paragraphe II.2.2.), sans doute via PPAR $\alpha$  au niveau hépatique, et pourrait jouer un rôle, avec p43, une forme tronquée du récepteur TR $\alpha$  dans la régulation de l'expression du génome mitochondrial (Casas et al., 2000b).

### **II.1.3. Expression et régulation des PPAR**

Le niveau d'expression des récepteurs PPAR dans l'organisme adulte et au cours du développement a été très étudié, notamment car leurs profils d'expression tissulaire respectifs, en particulier pour PPAR $\alpha$  et PPAR $\gamma$ , reflètent dans une certaine mesure leurs fonctions biologiques. Une synthèse complète de l'expression des trois PPARs chez le xénope, les rongeurs et l'humain peut être trouvé dans (Desvergne and Wahli, 1999). Par ailleurs, un criblage par PCR en temps réel de l'expression de tous les RNs dans une diversité de tissus murins (Bookout et al., 2006) ainsi qu'une analyse de leur expression circadienne dans le foie, les tissus adipeux blanc et brun et le muscle (Yang et al., 2006b) ont été récemment publiés et les données issues de ces deux études sont accessibles sur [www.nursa.org](http://www.nursa.org).

Le récepteur PPAR $\beta$  est le plus ubiquitaire des trois PPARs. Son expression débute tôt au cours de l'embryogénèse chez le rat (E8.5), présente un pic d'expression dans le tube neural entre E13.5 et E15.5 puis son expression décroît et devient ubiquitaire, comme observé chez le rat adulte (Braissant et al., 1996; Braissant and Wahli, 1998). Chez le rat adulte, son niveau d'expression est relativement abondant, souvent plus élevé que celui de PPAR $\alpha$  et PPAR $\gamma$  dans de nombreux tissus (Escher et al., 2001). C'est l'isoforme la plus abondante dans le système nerveux central chez l'adulte (Braissant et al., 1996; Cullingford et al., 1998;

Lemberger et al., 1996a). Bien qu'ubiquitaire, PPAR $\beta$  est un peu plus abondamment exprimé dans les reins et le long de l'intestin (muqueuse intestinale mais pas assise musculaire lisse, Braissant et al., 1996) avec une expression qui croît du duodénum au colon en passant par le jejunum et l'iléon et présente une expression relativement faible dans le foie comparativement au tube digestif (Bookout et al., 2006; Escher et al., 2001). Dans le foie et le tissu adipeux brun, PPAR $\beta$  semble présenter une expression qui fluctue (amplitude ~8 fois pour le foie et <2 fois pour le tissu adipeux brun) en fonction du rythme circadien (Yang et al., 2006b) avec une expression maximale au deuxième tiers de la nuit et en fin de nuit respectivement et une expression minimale en fin de journée/début de nuit dans les deux tissus. Une forte expression de PPAR $\beta$  a également été détectée dans les cellules de Sertoli (Braissant et al., 1996; Lemberger et al., 1996a) et dans les ostéoclastes (Mano et al., 2000). L'expression de PPAR $\beta$  est également fortement augmentée dans l'endomètre, au site d'implantation du blastocyste, lors de la nidation (Lim et al., 1999b). Chez l'homme, PPAR $\beta$  semble également ubiquitaire, présentant un niveau d'expression modéré et une répartition tissulaire qui semble globalement en accord avec les résultats obtenus chez les rongeurs avec une expression plus élevée dans le placenta, le gros intestin et le muscle (Auboeuf et al., 1997; Mukherjee et al., 1997b). Le statut nutritionnel semble constituer un régulateur important de l'expression de PPAR $\beta$ . Le jeûne réduit fortement son expression dans les reins et le foie (Escher et al., 2001) et l'augmente fortement dans le muscle squelettique (Holst et al., 2003). L'exercice physique a également été montré comme un stimulateur important de l'expression de PPAR $\beta$  au niveau musculaire (Luquet et al., 2003).

A l'opposé de PPAR $\beta$ , PPAR $\gamma$  présente un profil d'expression beaucoup plus restreint, les tissus adipeux blancs et bruns constituant ses principaux sites d'expression. Au cours du développement chez le rat, PPAR $\gamma$  présente un pic d'expression transitoire dans le système nerveux central à E13.5 puis disparaît complètement de l'organisme avant d'être à nouveau détecté à E18.5 dans le tissu adipeux brun à des niveaux élevés (Braissant and Wahli, 1998). Chez les rongeurs adultes, PPAR $\gamma$  est essentiellement exprimé dans les tissus adipeux blanc et brun (Bookout et al., 2006; Tontonoz et al., 1994b). Son expression est également détectée à des niveaux conséquents dans la muqueuse intestinale du colon et du caecum (Bookout et al., 2006; Lefebvre et al., 1998; Mansen et al., 1996; Saez et al., 1998) ainsi que certains tissus lymphoïdes comme la rate (chez les rongeurs uniquement) et les plaques de Peyer (Braissant et al., 1996; Lemberger et al., 1996a) et à des niveaux plus faibles dans la rétine, le muscle squelettique, le foie, l'os ou la peau. Son expression fluctue avec le cycle circadien dans le

foie et dans le tissu adipeux blanc (Yang et al., 2006b). Dans ce dernier tissu, l'expression des gènes de l'adiponectine, de la leptine et de SREBP-1c fluctuent de manière similaire à celle de PPAR $\gamma$  (Yang et al., 2006b). Chez l'homme, PPAR $\gamma$ 1 et PPAR $\gamma$ 2 sont exprimés à des niveaux élevés dans le tissu adipeux et à de faibles niveaux dans le muscle et PPAR $\gamma$ 1 est également détecté dans le foie, le cœur, les reins et le gros intestin (Fajas et al., 1997; Mukherjee et al., 1997b; Vidal-Puig et al., 1997). PPAR $\gamma$  est également exprimé à des niveaux modestes dans divers tissus ou populations cellulaires et notamment dans les monocytes et les macrophages (Nagy et al., 1998; Ricote et al., 1998; Tontonoz et al., 1998). Enfin, l'isoforme PPAR $\gamma$ 3 présente une expression limitée au tissu adipeux, au colon et aux macrophages (Fajas et al., 1998). La régulation de l'expression de PPAR $\gamma$  a fait l'objet de nombreuses recherches. Tout d'abord, PPAR $\gamma$  est induit, via les facteurs de transcription C/EBP $\alpha$  (CCAAT enhancer binding protein  $\alpha$ , Wu et al., 1999) et C/EBP $\beta$  (Wu et al., 1995), au cours de la différenciation adipocytaire et est un acteur important de la différenciation terminale des adipocytes (Chawla and Lazar, 1994; Tontonoz et al., 1994b; Tontonoz et al., 1994c). Les promoteurs des isoformes PPAR $\gamma$ 1 et PPAR $\gamma$ 3 sont également régulées par les facteurs de transcription SREBP (Fajas et al., 1999) parmi lesquels SREBP-1 semble, de plus, capable d'induire la synthèse de ligands endogènes spécifiques de PPAR $\gamma$  (Kim et al., 1998b). L'expression de PPAR $\gamma$ 1 et PPAR $\gamma$ 2 est réprimée par le jeûne et le diabète induit par la streptozotocine et elle est induite par un régime riche en graisses dans le tissu adipeux des rongeurs (Vidal-Puig et al., 1996). L'insuline et les glucocorticoïdes augmentent l'expression de PPAR $\gamma$  (Vidal-Puig et al., 1996) alors que le TNF- $\alpha$  (tumor necrosis factor  $\alpha$ ) qui est capable de provoquer la dédifférenciation d'adipocytes matures, réduit son expression (Zhang et al., 1996a). Ces différentes données ont suggéré que PPAR $\gamma$  pouvait jouer un rôle dans l'obésité chez l'homme. Plusieurs études ont permis de mettre en évidence un lien entre l'indice de masse corporelle (masse en Kg divisée par la taille en mètres élevée au carré) et le niveau d'expression de PPAR $\gamma$ 2 chez l'homme (pour des synthèses voir Desvergne and Wahli, 1999; Vidal-Puig et al., 1997), confirmant son implication dans le contrôle de la fonction adipocytaire. D'autres études se sont également attachées à mettre en évidence l'influence de polymorphismes du gène PPAR $\gamma$  sur l'obésité ou le diabète de type 2 mais le caractère multifactoriel de ces pathologies et la taille limitée des études conduites ne permettent pas de conclure définitivement sur l'ampleur du rôle joué par PPAR $\gamma$  dans ces pathologies (pour des synthèses voir Desvergne and Wahli, 1999; Semple et al., 2006).

L'expression de PPAR $\alpha$  est tardive (E13,5 chez le rat) au cours du développement chez le rongeur (Braissant and Wahli, 1998; Issemann and Green, 1990). Une expression transitoire de PPAR $\alpha$  est détectée dans le système nerveux central en développement et dans l'épiderme embryonnaire (Braissant and Wahli, 1998). Chez l'adulte, les principaux sites d'expression de PPAR $\alpha$  sont le foie, le cœur, les reins, le tissu adipeux brun, la muqueuse intestinale (expression majoritaire au sommet des villosités intestinales), la rétine, la glande surrénale, le muscle squelettique, les îlots du pancréas et les cellules musculaires lisses (Bookout et al., 2006; Braissant et al., 1996; Issemann and Green, 1990; Kliwer et al., 1994; Lemberger et al., 1996a; Staels et al., 1998; Zhou et al., 1998b). Son expression fluctue en fonction du rythme circadien dans le foie et les tissus adipeux blancs et bruns (Lemberger et al., 1996b; Yang et al., 2006b). Dans le foie, son expression est maximale en fin de journée et minimale en fin de nuit (Lemberger et al., 1996b; Yang et al., 2006b) ce qui semble cohérent avec l'activité nocturne des rongeurs (voir ci-dessous). Chez l'homme, le niveau d'expression de PPAR $\alpha$  dans le foie semble nettement inférieur au niveau d'expression observé chez les rongeurs (Palmer et al., 1998). De plus, un transcrit alternatif issu de l'épissage de l'exon 6 codant potentiellement un récepteur sans LBD est détecté, en plus du transcrit normal, dans tous les échantillons de foie testés (Palmer et al., 1998). Chez l'homme, PPAR $\alpha$  a également été détecté dans le cœur, les reins, le muscle squelettique et le gros intestin (Auboeuf et al., 1997; Mukherjee et al., 1997b). De manière générale, l'expression tissulaire et histologique de PPAR $\alpha$  apparaît très cohérente avec son rôle majeur dans l'activation de l'oxydation mitochondriale et peroxysomale des acides gras. Les principaux organes qui catalysent des acides gras (cœur, muscle squelettique, foie, muqueuse intestinale, reins) expriment des niveaux élevés de PPAR $\alpha$ . Au niveau histologique, les cardiomycocytes, les cellules des tubules proximaux du rein et les entérocytes situés au sommet des villosités de l'intestin sont les principaux sites d'expression de PPAR $\alpha$  dans leurs organes respectifs et utilisent également des acides gras comme principale source d'énergie. Enfin, au cours de la journée, l'activité réduite des souris peut s'apparenter à une phase de quasi-jeûne qui nécessite un recours aux réserves énergétique de l'organisme, principalement sous la forme d'acides gras issus de l'hydrolyse des triglycérides qui sont oxydés au niveau hépatique. La nuit, l'alimentation importante et régulière des souris leur apporte majoritairement des sucres, utilisés comme source d'énergie. Il est donc là encore intéressant de constater que le profil hépatique d'expression de PPAR $\alpha$  apparaît cohérent avec ce besoin d'oxydation des acides gras pendant la phase diurne du cycle circadien. En fait, l'expression hépatique de PPAR $\alpha$  est

augmentée par les glucocorticoïdes (Lemberger et al., 1994; Steiniger et al., 1994) et son niveau d'expression suit les variations circadiennes de la corticostérone plasmatique (maximale quelques heures avant la transition jour-nuit). Plus récemment, il a été montré que l'expression circadienne de PPAR $\alpha$  était liée à sa régulation par le dimère Clock-Bmal1 via des éléments de réponse situés dans son deuxième intron (Oishi et al., 2005). L'expression de PPAR $\alpha$  est par ailleurs diminuée par l'insuline (Steiniger et al., 1994) dont la sécrétion est maximale en situation d'alimentation. Les situations de stress et de jeûne, qui augmentent les niveaux plasmatiques de glucocorticoïdes et inhibent la sécrétion d'insuline, se traduisent, de fait, par une induction de PPAR $\alpha$  au niveau hépatique (Escher et al., 2001; Lemberger et al., 1996b). L'hormone de croissance (GH) réduit l'effet des proliférateurs de peroxysomes sur l'induction de la  $\beta$ -oxydation peroxysomale, un effet qui pourrait être lié à une inhibition de l'expression de PPAR $\alpha$  par la GH (Yamada et al., 1995). Une induction de l'expression de PPAR $\alpha$  par les fibrates, ses principaux activateurs pharmacologiques, a également été rapportée (Sterchele et al., 1996; Zhou et al., 1998b). Enfin, une exposition au froid se traduit par une réduction transitoire de l'expression de PPAR $\alpha$  et PPAR $\gamma$  et une augmentation de PPAR $\beta$  dans le tissu adipeux brun qui accompagne la phase initiale de prolifération des préadipocytes (Guardiola-Diaz et al., 1999). Ensuite, lors de la phase de différenciation des adipocytes, l'expression de PPAR $\alpha$  et de PPAR $\gamma$  ré-augmente, en parallèle de l'expression de leurs gènes cibles et de la prolifération des peroxysomes du tissu adipeux brun (Guardiola-Diaz et al., 1999).

## II.2. LIGANDS DES PPARs

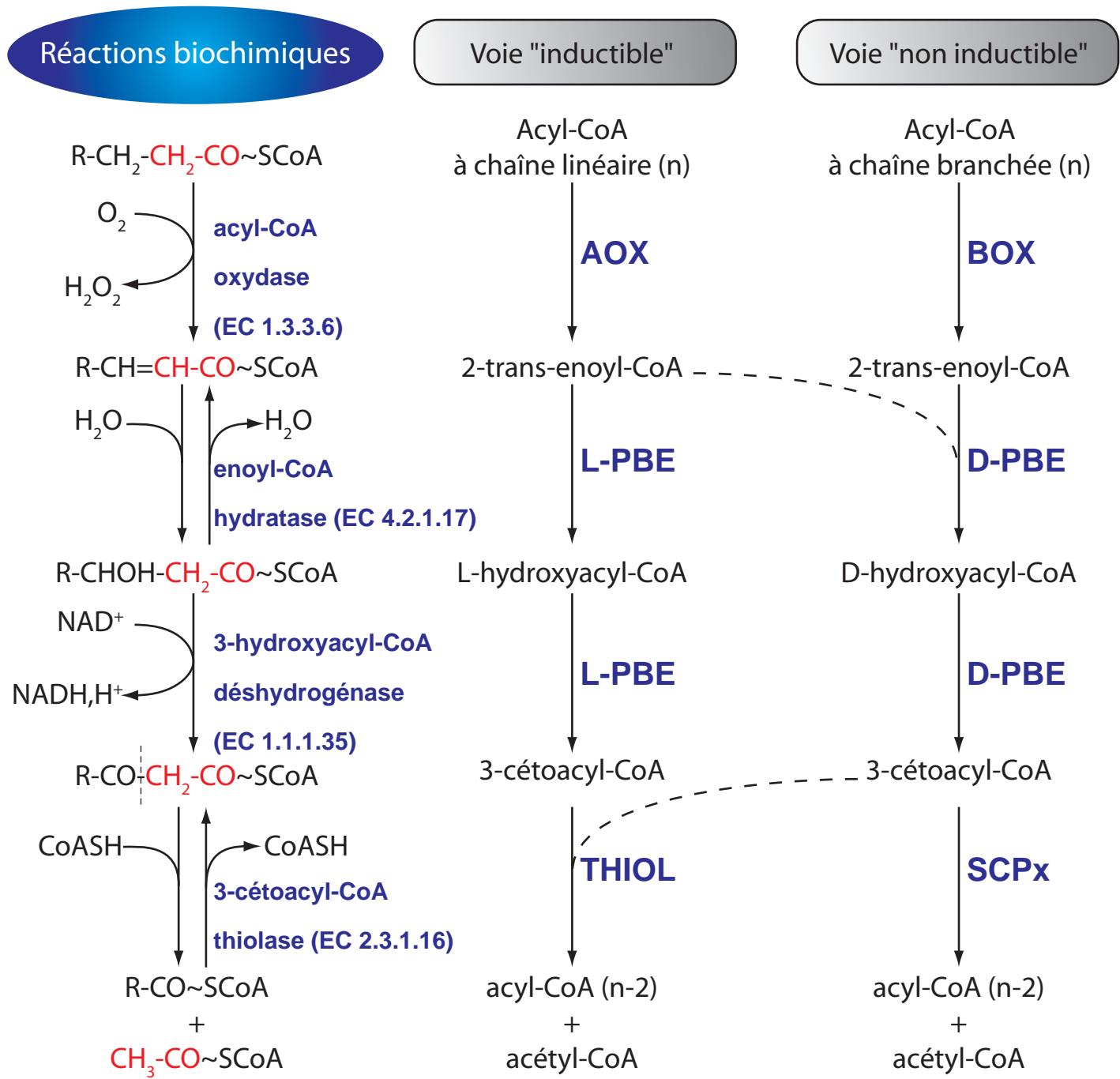
Dans ce chapitre nous évoquerons les principaux ligands endogènes et synthétiques des PPARs (dont certains ont été présentés en [Tables 5 et 6](#)) en nous concentrant prioritairement sur le récepteur PPAR $\alpha$ . Dans une troisième partie, nous évoquerons des stimuli physiopathologiques se traduisant par une activation de PPAR $\alpha$ .

### II.2.1. Ligands endogènes potentiels

Assez rapidement après la découverte des PPARs, une importante diversité d'**acides gras polyinsaturés** furent identifiés comme de bons ligands activateurs des trois PPARs (Forman et al., 1997; Gottlicher et al., 1992; Kliewer et al., 1997; Krey et al., 1997; Murakami et al., 1999). Parmi les PPAR, PPAR $\alpha$  semble avoir la meilleure affinité pour les acides gras

polyinsaturés et parmi les acides gras, les acides gras saturés apparaissent comme des ligands de moindre affinité des PPAR (pour une synthèse, voir Desvergne and Wahli, 1999). Des travaux plus récents ont montré que plusieurs acides gras présentaient des affinités de l'ordre du nM pour PPAR $\alpha$  (Ellinghaus et al., 1999; Hostetler et al., 2006; Hostetler et al., 2005; Lin et al., 1999), largement compatibles avec les concentrations circulantes mais surtout compatibles avec les concentrations intracellulaires et intranucléaires estimées de ces acides gras. Ces travaux récents (Hostetler et al., 2006; Hostetler et al., 2005) ainsi que ceux, plus anciens de Forman *et al.* (Forman et al., 1997) suggèrent que les acyl-CoA pourraient constituer de meilleurs ligands et activateurs que les acides gras correspondants. Cette hypothèse est également renforcée par les observations de l'équipe de J.K. Reddy sur les souris déficientes pour l'acyl-CoA oxydase (Fan et al., 1998a; Hashimoto et al., 2000; Hashimoto et al., 1999) et pour l'enzyme bifonctionnelle (Jia et al., 2003; Qi et al., 1999a), les deux premières enzymes impliquées dans la  $\beta$ -oxydation peroxysomale des acides gras à longue chaîne (Figure 20). Chez les souris AOX-/- (acyl-CoA oxydase 1 -/-), la voie PPAR $\alpha$  est constitutivement active au niveau hépatique, provoquant une prolifération des peroxysomes, une induction des gènes cibles de PPAR $\alpha$  et, à long terme, l'apparition d'hépatocarcinomes (Fan et al., 1998a; Hashimoto et al., 2000). Les doubles mutants AOX-/-PPAR $\alpha$ -/- ne présentent plus de prolifération spontanée des peroxysomes et d'induction des gènes cibles de PPAR $\alpha$  impliqués dans le catabolisme des acides gras (Hashimoto et al., 1999). Ces données suggèrent qu'un substrat de l'acyl-CoA oxydase qui s'accumulerait chez les souris AOX-/- pourrait servir de ligand endogène de PPAR $\alpha$  *in vivo*. L'entrée des acides gras à longue chaîne dans la voie de la  $\beta$ -oxydation peroxysomale nécessitant leur activation sous la forme d'acyl-CoA par l'acyl-CoA synthétase, il est donc probable que le ligand potentiel soit la forme acyl-CoA d'un acide gras à longue chaîne. Il existe deux voies de  $\beta$ -oxydation des acides gras dans le peroxysome (Figure 20), l'une prenant en charge les acyl-CoA à chaîne linéaire et qualifiée de voie « *inductible* » et l'autre, qualifiée de « *non inductible* », prenant en charge les acyl-CoA à chaîne branchée (portant un groupement méthyl en position 2) comme par exemple l'acide pristanique, issu de l'alimentation ou de l' $\alpha$ -oxydation de l'acide phytanique, ainsi que les précurseurs des acides biliaires (acides di- et tri-hydroxycholestanoïques) chenodéoxycholate et cholate (pour une synthèse, voir Wanders et al., 2001). L'invalidation de la seconde enzyme de la voie « *inductible* » chez la souris (souris L-PBE-/-) ne se traduit pas par une prolifération spontanée des peroxysomes ou une induction des gènes cibles de PPAR $\alpha$  (Qi et al., 1999a), suggérant que ce sont bien les

## La $\beta$ -oxydation peroxysomale des acides gras



**Figure 20 : La  $\beta$ -oxydation peroxysomale des acides gras à longue chaîne**

Les réactions biochimiques conduisant au raccourcissement de 2 carbones d'un acyl-CoA (chaîne linéaire) sont représentées à gauche du schéma. Les noms des fonctions enzymatiques intervenant dans ces réactions sont indiquées en bleu. Au centre et à droite, les noms des intermédiaires et les abréviations des noms des enzymes impliquées dans les deux voies de  $\beta$ -oxydation peroxysomale sont représentées. La voie "inductible" (ou L-hydroxy-spécifique) concerne les acides gras à chaîne linéaire, alors que la voie "non inductible" (ou D-hydroxy-spécifique) concerne principalement les acides gras branchés (fonction méthyle sur le carbone 2) et les précurseurs d'acides biliaires. Les traits en pointillés illustrent les échanges potentiels de substrats entre ces voies et sont en partie décrits dans le texte. AOX (ou Acox1: Acyl-CoA oxidase 1), BOX (ou Acox2 : branched acyl-CoA oxidase 2), L-PBE (L-peroxisomal bifunctional enzyme ou enoyl-CoA hydratase/L-3-hydroxyacyl-CoA dehydrogenase, également appelée MFP-1 pour multifunctional protein 1) , D-PBE (D-peroxisomal bifunctional enzyme ou D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase ou MFP-2 ou encore Hsd17b4 pour 17 $\beta$ -hydroxystéroid dehydrogenase) , THIOL (3-cétoacyl-CoA thiolases A et B), SCPx (Sterol Carrier Protein X).

substrats de l'AOX qui sont les ligands endogènes de PPAR $\alpha$ . Cependant, l'absence de stéatose hépatique chez les jeunes souris L-PBE-/-, contrairement aux souris AOX-/-, suggère par ailleurs que les substrats de la L-PBE pourraient être métabolisés via la D-PBE de la voie « non inducible » (Figure 20, Qi et al., 1999a), une hypothèse renforcée par l'observation d'une accumulation hépatique d'acides gras à longue chaîne chez les souris D-PBE-/- (Baes et al., 2000). Ces dernières présentent aussi une induction, moins marquée que chez les souris AOX-/-, de certains gènes cibles de PPAR $\alpha$  qui pourrait être liée à cette accumulation d'acides gras à longue chaîne (Baes et al., 2000; Jia et al., 2003). De manière intéressante, lorsqu'elles sont traitées avec un activateur pharmacologique de PPAR $\alpha$ , les souris L-PBE-/- ne présentent pas de prolifération des peroxysomes malgré une induction des gènes cibles de PPAR $\alpha$  (Qi et al., 1999a), suggérant un rôle clef de l'induction de L-PBE dans le phénomène de prolifération des peroxysomes. Finalement, les souris déficientes pour les deux enzymes bifonctionnelles de la  $\beta$ -oxydation peroxysomale (L-PBE-/- / D-PBE-/-) présentent une induction des gènes cibles de PPAR $\alpha$  en l'absence de prolifération des peroxysomes (Jia et al., 2003), élargissant ainsi la liste des ligands endogènes potentiels de PPAR $\alpha$  aux enoyl-CoA à longue chaîne, substrats de la deuxième étape de la  $\beta$ -oxydation peroxysomale et soulignant à nouveau le rôle important de la L-PBE dans le phénomène de prolifération des peroxysomes. Une souris déficiente pour l'un des deux gènes codant une 3-cétoacyl-CoA thiolase (la thiolase B) de la voie « inducible » a été également rapportée (Chevillard et al., 2004b) mais aucun phénotype particulier n'a été décrit concernant la modulation des gènes cibles de PPAR $\alpha$ . En revanche, l'inactivation du gène SCP2/SCPx, codant deux protéines dont l'une (SCPx) possède une activité thiolase et intervient dans la voie « non inducible » (Figure 20), se traduit notamment par une induction des gènes cibles de PPAR $\alpha$  et une prolifération des peroxysomes (Seedorf et al., 1998). Il a pu être montré que l'acide phytanique, un composé fortement présent dans les produits laitiers et dont l'apport quotidien chez l'homme est en moyenne d'environ 100 mg/jour, était métabolisé via son  $\alpha$ -oxydation en acide pristanique (également apporté par l'alimentation) qui subit ensuite une  $\beta$ -oxydation peroxysomale dont la dernière étape est catalysée par SCPx (pour une synthèse, voir Wanders et al., 2001). Ce composé, qui s'accumule chez les souris SCPx-/-, constitue un ligand de bonne affinité ( $K_d \sim 10\text{nM}$ ) et un activateur de PPAR $\alpha$  comparable aux fibrates (Ellinghaus et al., 1999; Hostetler et al., 2006). En conclusion, il apparaît très vraisemblable que PPAR $\alpha$ , et peut-être également PPAR $\beta$  et PPAR $\gamma$  dans une moindre mesure, soient liés et activés *in vivo* par des acides gras ou des intermédiaires de leur métabolisme. Ils agiraient ainsi comme

« senseurs » des niveaux intracellulaires d'acides gras et réguleraient au niveau transcriptionnel, parmi une grande diversité de gènes cibles, les niveaux d'expression de nombreuses enzymes impliquées dans le métabolisme des acides gras et plus généralement dans le métabolisme énergétique (voir paragraphe II.4.).

Il existe d'autres **acides gras particuliers** qui ont également été montrés comme étant de bons ligands et/ou activateurs des PPARs. On peut citer notamment les conjugués de l'acide linoléique, apportés par l'alimentation via les produits laitiers, les produits carnés bovins ou ovins et les huiles végétales partiellement hydrogénées. Ils activent PPAR $\alpha$ , dans une moindre mesure PPAR $\beta$  et inhibent l'activation de LXR (Clement et al., 2002; Moya-Camarena et al., 1999; Peters et al., 2001). Ces composés présentent une variété d'effets sur l'organisme dont une perte de poids et l'apparition d'une stéatose hépatique mais ces effets semblent néanmoins largement indépendants de leur capacité à activer PPAR $\alpha$  (Clement et al., 2002; Peters et al., 2001). L'oléyléthanolamide, un acide gras éthanolamide structuralement proche du cannabinoïde endogène anandamide et qui régule la satiété et la prise de poids a également été rapporté comme liant et activant PPAR $\alpha$  (et dans une moindre mesure PPAR $\beta$ , Fu et al., 2003). Ses effets sur la diminution de la prise alimentaire et de la prise de poids (Fu et al., 2003) ainsi que sur l'induction de l'oxydation des acides gras dans le muscle et de la lipolyse dans le tissu adipeux (Guzman et al., 2004) semblent s'opérer via PPAR $\alpha$ . Cet acide gras particulier qui n'active pas les récepteurs aux cannabinoïdes CB1 et CB2 est notamment produit par l'intestin grêle en situation d'alimentation mais sa production diminue en situation de jeûne (Rodriguez de Fonseca et al., 2001). Bien que cette observation semble en désaccord avec un rôle de ligand endogène de PPAR $\alpha$  dans la mesure où ce récepteur est activé, au moins au niveau hépatique, en situation de jeûne, une étude récente a montré que la biosynthèse de l'oléyléthanolamide était régulée de manière tissu-spécifique et que le foie et le tissu adipeux blanc (ainsi que d'autres organes comme le pancréas et la rate) voyaient leurs niveaux d'oléyléthanolamide augmentés suite à une épreuve de jeûne (Fu et al., 2007). Le résultat net de ces régulations tissu-spécifiques est une augmentation limitée des niveaux plasmatiques d'oléyléthanolamide en situation de jeûne qui ne semblent néanmoins pas atteindre les niveaux requis pour activer PPAR $\alpha$  (Fu et al., 2007), suggérant que ce ligand pourrait agir plutôt de façon paracrine. L'étude de l'activation de PPAR $\alpha$  en situation de jeûne chez des souris déficientes pour des enzymes de biosynthèse de l'oléyléthanolamide (dont certaines restent à découvrir), pourrait permettre de préciser le rôle joué par ce ligand endogène potentiel dans la régulation physiologique de l'activité de PPAR $\alpha$ . Enfin, signalons

également qu'un autre acide gras éthanolamide, le palmitoyléthanolamide a également été décrit comme exerçant son action anti-inflammatoire via l'activation de PPAR $\alpha$  (Lo Verme et al., 2005).

Une autre catégorie de ligands qui a été largement explorée est formée par les dérivés d'acides gras à 20 carbones regroupés sous le nom d'**éicosanoïdes** (du grec *eicosa* signifiant vingt). Les éicosanoïdes regroupent les prostaglandines, la prostacycline, les thromboxanes, les leucotriènes et les lipoxines et proviennent presque exclusivement du métabolisme de l'acide arachidonique (dont la nomenclature officielle : C20:4 $\omega$ 6, précise qu'il dispose de 20 atomes de carbones, de 4 doubles liaisons conjuguées dont la première est située en position  $\omega$ 6 c'est à dire à 6 carbones de l'atome de carbone le plus éloigné de la fonction acide) via l'action de lipooxygénases ou des cyclooxygénases (COX1 et COX2) suivie de l'action de synthases spécifiques des thromboxanes, des prostaglandines, de la prostacycline ou des leucotriènes (pour une synthèse, voir Folco and Murphy, 2006). Les prostaglandines, les leucotriènes et les thromboxanes agissent via des récepteurs membranaires couplés aux protéines G (pour une synthèse, voir Funk, 2001) mais des actions de ces dérivés d'acides gras et d'autres éicosanoïdes sur les récepteurs nucléaires ont également été rapportées dans des contextes généralement assez spécifiques. Le plus puissant éicosanoïde agissant comme ligand activateur de PPAR $\alpha$  est sans doute l'acide 8(S)hydroxyéicosatetraénoïque ou 8(S)HETE ([Table 5](#) et références incluses), un produit de l'action de la 8S-lipooxygénase sur l'acide arachidonique, qui est synthétisé dans la peau en réponse à une inflammation provoquée expérimentalement par les esters de phorbol ainsi que dans certains psoriasis et certaines tumeurs de la peau. PPAR $\alpha$  participerait ainsi à la promotion de la différenciation des kératinocytes provoquée par une surexpression de la 8S-lipooxygénase en régulant notamment l'expression de la kératine-1 (Muga et al., 2000). Le leucotriène B<sub>4</sub> (LTB<sub>4</sub>) a également été proposé comme un ligand endogène du récepteur PPAR $\alpha$  (Devchand et al., 1996) par le biais duquel il régulerait sa propre dégradation via la  $\beta$ -oxydation et l' $\omega$ -hydroxylation et, ainsi, la durée du processus inflammatoire auquel il participe. Cependant, l'affinité de PPAR $\alpha$  pour le LTB<sub>4</sub> apparaît généralement plus faible que son affinité pour de nombreux acides gras (Forman et al., 1997; Krey et al., 1997; Lin et al., 1999; Murakami et al., 1999) et certaines études semblent en désaccord avec une activation de la dégradation du LTB<sub>4</sub> via PPAR $\alpha$  (Fiedler et al., 2001). Bien que relativement peu de données aient été publiées sur une activation potentielle de PPAR $\alpha$  par le LTB<sub>4</sub>, la synthèse de ce leucotriène au niveau de l'enveloppe nucléaire (Funk, 2001) suggère que des concentrations

intranucléaires conséquentes pourraient être atteintes. La prostacycline ( $\text{PGI}_2$ ) a, elle, été proposée comme étant un activateur endogène de PPAR $\beta$  jouant un rôle important lors de l'implantation de l'embryon (Lim et al., 1999b) mais sa fixation au récepteur et sa capacité d'activation semblent faibles en comparaison de ses analogues (Forman et al., 1997). L'éicosanoïde le plus largement impliqué comme ligand endogène potentiel de PPAR $\gamma$  est un métabolite de la prostaglandine J2 : la 15-déoxy- $\Delta^{12,14}$ -prostaglandin J2 (15d-PGJ<sub>2</sub>). Son identification initiale en tant que ligand et activateur de PPAR $\gamma$  capable d'induire la différenciation de préadipocytes (Forman et al., 1995) a pu ensuite être confirmée par d'autres équipes (Kliwer et al., 1997; Krey et al., 1997). Cependant, son rôle comme ligand endogène de PPAR $\gamma$  reste controversé. Une étude a notamment montré que son niveau de synthèse était très en dessous des niveaux nécessaires à l'activation de PPAR $\gamma$  et n'était pas modulé lors de diverses situations impliquant l'activation de PPAR $\gamma$  (Bell-Parikh et al., 2003). Une étude récente utilisant un test ELISA pour mesurer la 15d-PGJ<sub>2</sub> dans des préadipocytes 3T3-L1 semble indiquer au contraire que sa synthèse augmente au cours de la différenciation adipocytaire (Mazid et al., 2006). Cependant, cette augmentation apparaît seulement tardivement (à partir du huitième jour environ, Mazid et al., 2006), or l'utilisation d'une lignée de préadipocytes 3T3-L1 exprimant une construction sensible à l'activation de PPAR $\gamma$  par un ligand endogène a permis de mettre en évidence qu'un ligand endogène de PPAR $\gamma$  était produit de manière transitoire très tôt au cours de l'adipogénèse (premier et deuxième jour) et disparaissait ensuite rapidement (dès le troisième jour) au cours de la maturation des adipocytes (Tzameli et al., 2004). Très récemment, un métabolite de la prostaglandine E<sub>2</sub> (PGE<sub>2</sub>) : le 15-céto-PGE<sub>2</sub> a été proposé comme un nouveau ligand endogène de PPAR $\gamma$  et il est intéressant de constater que l'expression de l'enzyme responsable de sa dégradation est observée à partir du troisième jour de différenciation des préadipocytes 3T3-L1 (Chou et al., 2007). En conclusion, l'activation des PPARs par des éicosanoïdes semble relativement bien démontrée mais concerne vraisemblablement des conditions très particulières notamment en termes de tissu ou de sous-population cellulaire concernée et, en cohérence avec l'action généralement locale (autocrine ou paracrine) des éicosanoïdes (Folco and Murphy, 2006; Funk, 2001), explique difficilement des effets systémiques (action de type endocrine) conduisant à l'activation simultanée des PPARs dans divers tissus (au cours du jeûne par exemple). La capacité des PPARs à lier une grande diversité de ligands semble suggérer qu'ils pourraient avoir de multiples ligands endogènes et répondre à la fois à des signaux très spécifiques, produits au niveau local, comme les éicosanoïdes ainsi qu'à des signaux moins

spécifiques, comme les acides gras en agissant comme senseur du métabolisme énergétique. Dans ce cadre, des résultats obtenus dans le cadre de ce doctorat suggèrent que même à faible dose, les acides gras alimentaires présentent une capacité à moduler l'expression hépatique de plusieurs gènes de manière partiellement dépendante de PPAR $\alpha$  (Martin et al., 2007). De nombreuses recherches apparaissent encore nécessaires pour définir de manière indubitable la nature exacte des ligands endogènes des PPARs.

### **II.2.2. Ligands synthétiques**

La prévalence de l'obésité a augmenté de manière alarmante dans les pays industrialisés au cours des 30 dernières années, conduisant en particulier environ un tiers de la population américaine (hommes et femmes) à être obèse (Ogden et al., 2006). Cette augmentation s'accompagne d'un accroissement équivalent des désordres associés à l'obésité comme l'intolérance au glucose, l'insulino-résistance, les dyslipidémies ou l'hypertension qui constituent des facteurs prédisposant aux pathologies cardiovasculaires et dont l'association caractérise le syndrome métabolique ou syndrome X (pour une définition précise, voir Desvergne et al., 2006). Les solutions à ce problème sont assez largement connues (diminution des portions alimentaires, modification des choix alimentaires, exercice physique régulier) mais sont difficiles à appliquer à l'échelle d'une population entière. Dans ce contexte, nous allons voir que l'action pharmacologique sur les récepteurs PPARs présente un intérêt particulier pour l'industrie pharmaceutique dans la lutte contre le syndrome métabolique, en raison de leurs rôles majeurs dans la régulation du métabolisme. Nous n'évoquerons dans ce paragraphe que les principaux ligands synthétiques, à savoir les **fibrates** et les **thiazolidinediones** utilisés en thérapeutique humaine ainsi que quelques molécules présentant un intérêt en pharmacologie et toxicologie. Un inventaire très complet des nombreux ligands synthétiques décrits dans la littérature a été récemment publiée (Michalik et al., 2006) ainsi qu'une synthèse des mécanismes d'action des agonistes des PPARs utilisés en thérapeutique humaine (Gervois et al., 2007).

Les ligands synthétiques de PPAR $\alpha$  sont regroupés sous le terme de **proliférateurs de peroxysomes (PPs)**, en raison de leur capacité à induire, chez les rongeurs essentiellement, une prolifération des peroxysomes au niveau hépatique. Les peroxysomes sont des organites cellulaires principalement impliqués dans le métabolisme des lipides. Les activateurs de PPAR $\alpha$  provoquent une augmentation de leur taille et de leur nombre, associé à une hépatomégalie, et régulent fortement l'expression de nombreux gènes impliqués dans le

catabolisme ou le transport des lipides dans les peroxysomes. Leur administration à long terme se traduit, chez les rongeurs, par l'apparition d'hépatocarcinomes bien qu'ils ne soient pas directement génotoxiques (Reddy et al., 1986). Ces différentes réponses sont absentes chez les souris déficientes en récepteur PPAR $\alpha$  (Lee et al., 1995) et la pharmacovigilance n'a pas révélé d'incidence accrue d'hépatocarcinomes chez l'homme suite aux traitements par les agonistes de PPAR $\alpha$  utilisés en thérapeutique humaine. Parmi les PPs, les **fibrates** (pour phenoxyisobutyrates) sont utilisés depuis les années 70 dans le traitement des hyperlipidémies. Ils présenteraient également des activités anti-inflammatoires et anti-athérosclérose. Cette famille de molécules comporte notamment le Wy14643, le gemfibrozil, le ciprofibrate, le clofibrate, le bezafibrate et le fenofibrate dont certains sont illustrés en [Table 6](#). Les premiers criblages de ligands des PPARs publiés ont permis de démontrer la liaison spécifique des fibrates à PPAR $\alpha$  avec des affinités néanmoins relativement faibles (Forman et al., 1997; Krey et al., 1997; Lin et al., 1999; Murakami et al., 1999). Des travaux plus récents suggèrent que les formes thioester du coenzyme A (bezafibroyl-CoA) pourraient être de meilleurs ligands (Hostetler et al., 2005). Les fibrates sont généralement administrés par voie orale sous forme ester et subissent une saponification au niveau intestinal qui génère la forme acide : par exemple, le fenofibrate (désigne la forme ester) est saponifié en acide fénofibrique (forme acide illustrée en [Table 6](#)). Ce sont donc les formes acides, d'ailleurs bien plus solubles, qui sont les activateurs de PPAR $\alpha$  comme constaté dans une étude récente qui a par ailleurs suggéré que les formes esters pourraient constituer des antagonistes de LXR (Thomas et al., 2003). Les principaux effets des fibrates sont une diminution marquée des triglycérides, une légère augmentation du cholestérol-HDL et une diminution plus modérée et moins robuste du cholestérol-LDL qui sont expliquées par leur action, via PPAR $\alpha$ , sur l'expression de nombreux gènes impliqués dans le transport, le captage et le métabolisme des lipides (voir paragraphe II.4.). Leur indication est donc principalement orientée vers les fortes hypertriglycéridémies associées à des niveaux de cholestérol-HDL faibles et un cholestérol-LDL normal. Les fibrates semblent plus efficaces à réduire le risque cardiovasculaire chez les patients dyslipidémiques présentant également un diabète ou une hyperinsulinémie, des situations qui favorisent des conditions proinflammatoires (Michalik et al., 2006). De fait, en plus de leurs effets sur le métabolisme des lipides, les fibrates agissent directement, via PPAR $\alpha$  et son interaction avec d'autres voies de signalisation, sur l'inflammation aux niveaux vasculaire et hépatique (voir paragraphe II.5 et Marx et al., 2004). De nombreuses études cliniques ont mis en évidence la capacité des fibrates à normaliser les niveaux plasmatiques

de lipides et à réduire les risques cardiovasculaires (pour des synthèses récentes voir Gervois et al., 2007; Wierzbicki, 2006) mais, contrairement aux statines qui sont largement utilisées pour le traitement de l'hypercholestérolémie, aucune de ces études n'a pu mettre en évidence une réduction globale de la morbi-mortalité (toutes causes confondues). Actuellement, un intérêt particulier est porté sur le potentiel d'association entre fibrates et statines pour le traitement des dyslipidémies (Wierzbicki, 2006). A priori, ces deux types de molécules ont des modes d'action différents et complémentaires, les statines agissant comme inhibiteurs de la synthèse endogène de cholestérol et les fibrates agissant majoritairement sur les triglycérides. Cependant, leur association présente également des risques en termes d'effets secondaires, notamment au niveau musculaire.

Les **thiazolidinediones** (TZDs ou glitazones) comme la pioglitazone ou la rosiglitazone ([Table 6](#)) sont des analogues structuraux des fibrates qui lient et activent spécifiquement PPAR $\gamma$  (Forman et al., 1995; Lehmann et al., 1995). Elles sont utilisées en thérapeutique humaine pour le traitement du diabète de type II car leur effet consiste principalement à restaurer la sensibilité à l'insuline des tissus périphériques. Cet effet global reste encore mal compris mais l'action des TZDs au niveau du tissu adipeux pourrait être un mécanisme central de leur action sur l'insulino-résistance (pour des synthèses, voir Desvergne et al., 2006; Gervois et al., 2007; Walczak and Tontonoz, 2002). En augmentant la capacité du tissu adipeux à stocker des lipides via son action adipogénique (augmentation du captage des acides gras et de la lipogénèse), PPAR $\gamma$ , drainerait les tissus périphériques et notamment le foie, les muscles et le pancréas de leur graisse en excès. Ces graisses accumulées dans les tissus, liées notamment à des niveaux circulants élevés d'acides gras libres en raison d'un stockage défectueux dans les adipocytes, provoquent une induction de la néoglucogenèse hépatique et une diminution de l'utilisation de glucose par les muscles. En redirigeant ces lipides vers leur lieu de stockage privilégié, les TZDs agiraient positivement sur le métabolisme glucidique dans ces tissus. De plus, l'activation de PPAR $\gamma$  par les TZDs favoriserait le stockage des graisses dans le tissu adipeux sous-cutané qui est plus sensible à l'insuline que le tissu adipeux viscéral et augmenterait le nombre des petits adipocytes (par opposition aux gros adipocytes gorgés de lipides), eux aussi plus sensibles à l'action de l'insuline. La fonction endocrine du tissu adipeux (pour une synthèse, voir Trujillo and Scherer, 2006), qui consiste en la synthèse et la sécrétion de nombreuses protéines regroupées sous le nom d'adipokines, est également affectée par les TZDs. En particulier, PPAR $\gamma$  induit l'expression et facilite la sécrétion de l'adiponectine (Iwaki et al., 2003), une adipokine qui

circule à des concentrations sériques élevées (2-20 µg/mL) dont le niveau baisse dans les cas d'obésité et de résistance à l'insuline (Kadowaki et al., 2006; Oh et al., 2007). Cette hormone agit via deux récepteurs membranaires (AdipoR1 principalement exprimé dans le muscle et AdipoR2 principalement exprimé dans le foie) sur diverses voies dont celles de l'AMP-activated protein kinase (AMPK) et de PPAR $\alpha$  (pour une synthèse voir Kadowaki et al., 2006) qui conduisent à une stimulation de l'oxydation des acides gras et du captage du glucose par le muscle et à une réduction de la production de glucose par le foie. Par ailleurs, il a été montré qu'une masse corporelle élevée et la présence de gros adipocytes (caractéristiques de l'obésité) étaient corrélées à l'accumulation de macrophages dans le tissu adipeux (Weisberg et al., 2003) qui induisent des voies inflammatoires en secrétant notamment du TNF- $\alpha$  (tumor necrosis factor  $\alpha$ ) et de l'IL-6 (interleukine-6), participant vraisemblablement à la résistance à l'insuline en générant un statut d'inflammation chronique et en favorisant la lipolyse (Xu et al., 2003). Les TZDs apparaissent également capables de diminuer l'expression du TNF- $\alpha$  et de l'IL-6 dans les macrophages issus du tissu adipeux (Xu et al., 2003), ce qui participerait à leur action globale sur la résistance à l'insuline. Enfin, les TZDs agiraient également directement au niveau des différentes cellules de la paroi des vaisseaux (macrophages impliqués dans la formation de cellules spumeuses lorsqu'ils accumulent des lipides, recrutement des monocytes par les cellules endothéliales de la paroi des vaisseaux et cellules musculaires lisses), ce qui participerait à leurs effets anti-athérogéniques (pour une synthèse, voir Walczak and Tontonoz, 2002).

Aucun agoniste spécifique de PPAR $\beta$  n'est actuellement utilisé en thérapeutique humaine mais plusieurs études suggèrent qu'il pourrait également constituer une cible thérapeutique intéressante pour le traitement des désordres métaboliques en raison de son implication dans la régulation du métabolisme en particulier dans le tissu adipeux, l'intestin, le muscle, le cœur et les macrophages (Barish et al., 2006). Par exemple, un traitement par un agoniste spécifique de PPAR $\beta$  provoque une augmentation du cholestérol-HDL chez les souris db/db (souris obèses en raison d'une déficience en récepteur à la leptine, Leibowitz et al., 2000), améliore le transport inverse du cholestérol (retour du cholestérol au foie via les HDL), améliore le profil des LDL (réduction du LDL cholestérol et augmentation de la taille des particules) et corrige l'hyperinsulinémie de singes rhésus obèses (Oliver et al., 2001). Les travaux réalisés sur PPAR $\beta$  dans le muscle (synthétisés dans Gaudel and Grimaldi, 2007) suggèrent qu'il agit sur le remodelage et la capacité oxydative du muscle squelettique provoqué notamment par l'exercice physique. Ses effets musculaires pourraient expliquer, au

moins en partie, son action positive sur le profil lipidique dans la mesure où sa surexpression au niveau du muscle se traduit par une réduction du tissu adipeux (et de la taille des adipocytes), une diminution du contenu en lipides de plusieurs organes et une amélioration de la réponse à l'insuline.

Actuellement, de nombreux laboratoires publics et privés travaillent au développement et à l'étude de modulateurs sélectifs des PPARs dans le but d'activer ces voies de signalisation de manière tissu-spécifique et de conserver ainsi leurs effets bénéfiques tout en limitant leurs effets délétères (Marx et al., 2004). La recherche est sans doute au moins aussi active sur le développement d'agonistes doubles PPAR $\alpha$ /PPAR $\gamma$  (dénommés glitazars), voire triple PPAR $\alpha/\beta/\gamma$  (panagonistes) dans le but d'obtenir des effets complémentaires et synergiques de l'activation de ces trois voies de signalisation (Gervois et al., 2007).

D'autres médicaments sur le marché ou en développement activent également les PPARs. On peut citer notamment les anti-inflammatoires non stéroïdiens (indométhacine, acide flufenamique, fenoprofène ou ibuprofène par exemple) dont le mode d'action principal consiste à inhiber la synthèse de prostaglandines via une inhibition des cyclooxygénases (COX1 et COX2) et qui activent également à des degrés divers PPAR $\alpha$  et PPAR $\gamma$  mais à des doses bien supérieures à celles nécessaires à leur action sur les COX (Lehmann et al., 1997). Des polluants potentiels de l'environnement comme des pesticides ou des plastifiants ont été identifiées comme ligands et activateurs des PPARs (Maloney and Waxman, 1999; Takeuchi et al., 2006). Un exemple intéressant est celui des phtalates (esters de l'acide phtalique), des assouplissants des plastiques (en particulier des PVC, poly-vinyl-chloride) pouvant entrer dans la composition de multiples produits finis (emballages alimentaires, tubulures plastiques, revêtements, produits cosmétiques, peintures, colles, etc.) jusqu'à hauteur de 50% en masse et représentant un volume de production considérable (1,3 Kg/habitant/an dans l'union européenne soit environ 600 000 tonnes). Identifiés depuis longtemps comme des proliférateurs de peroxysomes chez le rongeur (Reddy et al., 1986) et considérés comme des disrupteurs endocriniens (actions anti-androgénique et estrogéno-mimétique), les phtalates mais surtout leurs métabolites activent PPAR $\alpha$  et PPAR $\gamma$  (dans une moindre mesure PPAR $\delta$ ) via une fixation directe aux récepteurs (Lapinskas et al., 2005). L'étude des phtalates a connu récemment un regain d'intérêt suite à l'établissement chez l'homme de corrélations entre des perturbations du développement de l'appareil génital masculin et l'exposition prénatale aux phtalates (Swan et al., 2005) et plus récemment entre l'exposition aux phtalates chez l'adulte et des paramètres liés au syndrome métabolique (Stahlhut et al., 2007). Ces résultats ont

participé à promouvoir la mise en œuvre d'études plus poussées sur les conséquences de l'exposition prénatale et postnatale aux phtalates au niveau moléculaire (voir par exemple Feige et al., 2007; Plummer et al., 2007).

En conclusion l'étude des ligands synthétiques des PPARs intéresse fortement l'industrie pharmaceutique pour qui le développement de médicaments contre les différentes composantes du syndrome métabolique représente un marché croissant et largement solvable mais concerne également le domaine de la toxicologie même si les données dans ce domaine apparaissent globalement plus parcellaires, peut-être en raison d'une plus grande difficulté à obtenir des financements pour ce type de travaux de recherche.

### **II.2.3. Autres situations d'activation de PPAR $\alpha$**

L'étude des RNs de classe II s'est rapidement heurtée à la question de savoir si un hétérodimère RN-RXR pouvait être activé par un ligand de l'un, de l'autre ou des deux partenaires engagés. En ce qui concerne les PPAR, les premières études ont mis en évidence que le dimère pouvait être qualifié de « permissif » c'est à dire qu'il pouvait être activé aussi bien par des agonistes de PPAR que par des agonistes de RXR (Keller et al., 1993; Kliewer et al., 1992; Schulman et al., 1998). Ces travaux ont conduit à démontrer que les agonistes de RXR pouvait présenter, *in vivo*, une activité normolipémante comparable aux activateurs de PPAR $\alpha$  (Mukherjee et al., 1998) et une activité antidiabétique comparable aux activateurs de PPAR $\gamma$  (Mukherjee et al., 1997a). Des travaux issus de notre laboratoire (Ouamrane et al., 2003), dont certains font partie intégrante de ce travail de doctorat (Martin et al., 2005) ont confirmés une partie de ces observations en démontrant par ailleurs l'existence de régulations tissus-spécifiques et PPAR $\alpha$ -indépendantes des activateurs spécifiques de RXR. Ainsi, l'activation de RXR constitue une autre situation d'activation des PPARs. Aucune étude, à ma connaissance, n'a cependant évalué la capacité différentielle des trois PPARs à être activés via les ligands de RXR ainsi que le recrutement différentiel de cofacteurs que ce mode d'activation est susceptible d'engendrer.

Il existe également un nombre conséquent de molécules qui activent les voies PPARs sans que leur fixation aux récepteurs ait été démontrée. La DHEA (dehydroépiandrostérone) en est un exemple. Cette hormone, produite par la glande surrénale en quantités importantes (concentrations circulantes de l'ordre du  $\mu$ M) et constituant un précurseur des androgènes et des estrogènes, présente un métabolite (la DHEA-3 $\beta$ -sulfate) capable d'activer le récepteur PPAR $\alpha$  *in vivo* dans la mesure où son administration (à fortes doses ~100-150 mg/Kg) à des

souris de type sauvage produit une hépatomégalie et une induction des gènes cibles de PPAR $\alpha$  qui ne sont pas observées chez les souris PPAR $\alpha$ -/- (Peters et al., 1996). En revanche, la DHEA et plusieurs de ses métabolites semblent incapables d'agir directement comme ligands et activateurs de PPAR $\alpha$  *in vitro* comme cela a été démontré via l'utilisation d'un système de gène rapporteur (Peters et al., 1996). A ma connaissance, le mécanisme précis de prolifération des peroxysomes induit par la DHEA (activation de PPAR $\alpha$  par d'autres métabolites non identifiés, génération d'un autre ligand endogène, nécessité d'un transporteur spécifique non exprimé dans les cultures cellulaires, modification post-traductionnelle de PPAR $\alpha$ , autres mécanismes ?) reste actuellement inconnu. En revanche, la DHEA a depuis été identifiée comme un activateur des récepteurs PXR (Ripp et al., 2002) et CAR (Kohalmy et al., 2007) sans que sa fixation ou celle de ses métabolites à ces récepteurs ait néanmoins été étudiée.

Enfin, il existe des situations physiopathologiques où PPAR $\alpha$  est également activé sans que l'on connaisse avec précision le ou les ligands impliqués. En particulier, le jeûne (Kersten et al., 1999; Kroetz et al., 1998), le diabète induit par la streptozotocine (Kroetz et al., 1998) et un régime riche en graisses (Patsouris et al., 2006) se traduisent par une nette activation de PPAR $\alpha$  au niveau hépatique. Ces trois situations se traduisent par une augmentation des acides gras libres circulants et il est très probable que ce soient ces lipides, ou bien leurs dérivés, qui constituent les ligands endogènes de PPAR $\alpha$  dans ces situations. Un argument en cette faveur a été fourni par une étude de jeûne réalisée chez des souris préalablement alimentées avec un régime normal ou avec un régime riche en acides gras polyinsaturés qui sont de meilleurs activateurs de PPAR $\alpha$  et qui a montrée que l'induction par le jeûne de l'expression de PPAR $\alpha$  et de l'un de ses gènes cibles (L-FABP) était plus forte dans le groupe préalablement alimenté par le régime riche en polyinsaturés (Kersten et al., 1999). En effet, les lipides alimentaires vont s'accumuler dans le tissu adipeux des souris dont la composition en acides gras reflètera bien celle des régimes. En situation de jeûne, ces lipides seront mobilisés du tissu adipeux et parviendront au foie où, selon leur composition en acides gras, ils activeront plus ou moins fortement PPAR $\alpha$ . Dans le but de préciser les effets du jeûne sur l'activation de PPAR $\alpha$  et éventuellement de renforcer ces hypothèses, nous avons conduit une étude de jeûne sous la forme d'une cinétique de jeûne allant de 0 à 72h de jeûne chez des souris de type sauvage ou déficientes en récepteur PPAR $\alpha$ . Cette étude est présentée dans le Chapitre VI.

## **II.3. FONCTIONNEMENT DE LA VOIE DE SIGNALISATION PPAR $\alpha$**

Le principal objet de ce paragraphe est de situer le récepteur PPAR $\alpha$  par rapport aux données présentées dans le Chapitre I et en particulier par rapport aux deux modèles présentés au paragraphe I.6. Pour cela, nous nous appuierons en particulier sur la structure des récepteurs PPARs et la fonction de leurs différents domaines.

### **II.3.1. Interactions avec les protéines chaperones**

Une interaction entre PPAR $\alpha$  et les protéines chaperones Hsp70 (in vitro, Huang et al., 1994) et Hsp90 (in vitro et in vivo, Sumanasekera et al., 2003a; Sumanasekera et al., 2003b) a été décrite. Hsp90 a de plus été proposée comme un inhibiteur de l'activité de PPAR $\alpha$  (Sumanasekera et al., 2003a). Néanmoins, au regard de la localisation principalement cytoplasmique de Hsp90 et exclusivement nucléaire de PPAR $\alpha$  (Feige et al., 2005) et le peu de données produites sur les conséquences fonctionnelles de ces associations, des recherches complémentaires semblent nécessaires pour corroborer ces observations préliminaires.

### **II.3.2. Domaines A/B et phosphorylation de PPAR $\alpha$**

Les PPARs possèdent une structure similaire à celle décrite dans le paragraphe I.2. Une fonction AF-1 est localisée dans leur domaine A/B (Desvergne and Wahli, 1999). Elle participerait vraisemblablement à leur activité indépendante du ligand mais régulerait également, via des interactions intramoléculaires, leur activité dépendante du ligand (Burns and Vanden Heuvel, 2007; Diradourian et al., 2005; Warnmark et al., 2003). L'activité de la fonction AF-1 de PPAR $\alpha$  semble plus forte que celle des fonctions AF-1 de PPAR $\beta$  et de PPAR $\gamma$  (Hi et al., 1999), ce qui expliquerait, au moins en partie et dans certains modèles cellulaires, l'activité notable d'une transfection transitoire de PPAR $\alpha$  en l'absence de ligand (Hi et al., 1999; Juge-Aubry et al., 1999). Notons que la notion d'absence de ligand, ne tient pas ici compte de la présence systématique d'acides gras dans les cellules qui constituent malgré tout des ligands des trois PPARs. La structure, vraisemblablement en  $\alpha$ -hélice, de la fonction AF-1 de PPAR $\alpha$  serait importante pour son activité (Hi et al., 1999). De plus, le domaine A/B de PPAR $\alpha$  est phosphorylé par diverses voies des MAPK. Sous l'effet de l'insuline, la fonction AF-1 de PPAR $\alpha$  est phosphorylée sur les séries 12 et 21 via la voie des MAPK (mitogen-activated protein kinase) p42/ERK2 (extracellular signal-regulated kinase 2) et p44/ERK1 ce qui conduit à un fort accroissement de l'activité de PPAR $\alpha$  (Juge-

Aubry et al., 1999). Dans le cœur, ces deux séries ainsi que la sérine 6 sont phosphorylées par la MAPK p38 qui est activée par des cytokines et divers stress comme l’ischémie, l’hypoxie ou la croissance hypertrophique, ce qui conduit à un accroissement de l’activité de PPAR $\alpha$  dépendante du ligand (Barger et al., 2001). La sérine 73 du domaine A/B de PPAR $\alpha$  est également phosphorylée par la glycogène synthase kinase 3 (GSK3) ce qui déstabilise la protéine PPAR $\alpha$  vraisemblablement via sa dégradation par le système ubiquitine-protéasome (Burns and Vanden Heuvel, 2007). Relativement peu de données sont disponibles sur les conséquences de ces différentes phosphorylations *in vivo* mais il existe un exemple où certaines conséquences pathologiques d’une mutation ont été attribuées à une altération de la phosphorylation des PPARs. Une mutation de l’hélicase XPD appartenant au complexe TFIIH et affectant l’activité de la cyclin-dependent kinase 7 (cdk7) semble réduire la phosphorylation et l’activité de PPAR $\alpha$  (sérines 12 et 21) et PPAR $\gamma$  (sépine 112, Compe et al., 2005). Néanmoins, les conséquences de cette mutation sur l’activité globale de la voie PPAR $\alpha$  nous ont semblé très limitées, voire imperceptibles lorsque nous avons eu l’occasion de tester des échantillons d’ARN issus de ces souris traitées ou non avec du fenofibrate sur les puces à ADN développées dans le cadre de ce doctorat ([Figure 21](#), JM Egly et PG. Martin, données non publiées). Le domaine A/B n’est cependant pas le seul domaine des PPARs à être phosphorylé. Les activateurs de la protéine kinase A (PKA) comme la toxine du choléra (PKA est également activée en situation de stress, de jeûne ou d’exercice physique) activent PPAR $\alpha$  en l’absence, comme en présence d’un ligand exogène vraisemblablement via des phosphorylations du domaine C (DBD) qui affectent la liaison PPAR $\alpha$ -ADN (Lazennec et al., 2000). La voie des protéines kinase C phosphoryle PPAR $\alpha$  sur diverses séries et thréonines situées dans les domaines C et D et affecte ainsi positivement plusieurs aspects de l’activation de PPAR $\alpha$  comme sa dimérisation avec RXR ou sa fixation à l’ADN (Blanquart et al., 2004; Gray et al., 2005). La voie de l’AMPK (5’-AMP-activated protein kinase) qui régule notamment l’oxydation des acides gras dans le muscle, active également PPAR $\alpha$  et augmente son expression mais le mécanisme impliqué n’a pas été détaillé (Lee et al., 2006). Récemment, il a été proposé que l’adiponectine, qui présente la capacité d’activer l’oxydation des acides gras dans les tissus périphériques comme le foie et les muscles, activerait la voie PPAR $\alpha$  dans le muscle via une activation séquentielle de la voie AMPK et de la MAPK p38, suggérant que les effets de la voie AMPK sur PPAR $\alpha$  nécessitent l’activation de la MAPK p38 par l’AMPK (Yoon et al., 2006). Plus récemment encore, il a été montré que la leptine (et l’adiponectine également), qui stimule également l’oxydation des acides gras dans le muscle,

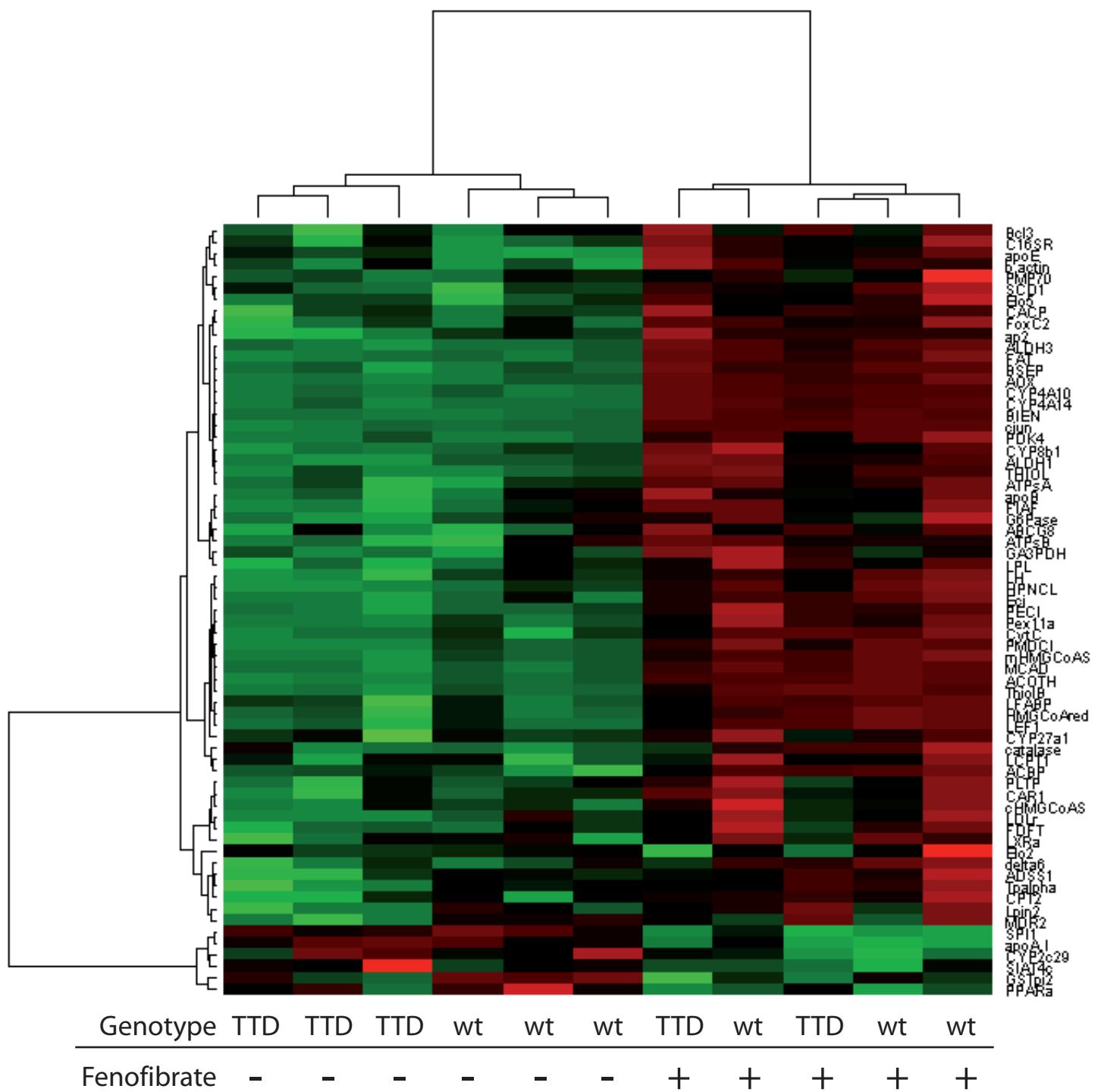


Figure 21 : Effets du fénofibrate chez les souris TTD sur l'expression hépatique d'un panel de gènes

Des souris de type sauvage (wt) ou de type mutante pour l'hélicase XPD (TTD) ont été traitées au fénofibrate (100 mg/Kg/j, 8 jours) ou par le véhicule seul. Les puces INRAArray 01.3 ont été utilisées pour analyser les répercussions de ce traitement dans les deux génotypes sur l'expression d'un panel de gènes cibles des récepteurs nucléaires de classe II. Cette figure a été obtenue par classification ascendante hiérarchique (distance 1-cor où cor est le coefficient de corrélation linéaire de Pearson, et critère d'agglomération de Ward) des échantillons puis des gènes et par représentation de la matrice de données (gènes centrés et réduits) réordonnée par ces classifications sous la forme d'une image en couleur (heatmap) où les fortes intensités sont codées en rouge, les faibles en vert et les intensités moyennes en noir. Cette figure illustre que la réponse des souris TTD au fénofibrate est globalement indiscernable de celle des souris de type sauvage. On distingue aisément un large groupe de gènes dont l'expression est induite par le fénofibrate et un groupe de quelques gènes dont l'expression est réprimée par le fénofibrate. Ces régulations sont observées chez les souris des deux génotypes avec des intensités globalement comparables.

provoquait une translocation nucléaire de la sous-unité  $\alpha$ 2 de l'AMPK et induisait l'expression de PPAR $\alpha$  (Suzuki et al., 2007). Des travaux complémentaires apparaissent nécessaires pour préciser les mécanismes impliqués dans la régulation de l'activité et de l'expression de PPAR $\alpha$  par la voie de l'AMPK. Enfin, PPAR $\beta$  et PPAR $\gamma$  font également l'objet de phosphorylations qui participent à la régulation de leur activité (pour des synthèses récentes, voir Burns and Vanden Heuvel, 2007; Diradourian et al., 2005). La fonction du domaine A/B n'est cependant pas uniquement de rendre les PPARs sensibles à des voies de signalisation conduisant à sa phosphorylation. Par exemple, ce domaine semble indispensable à l'inhibition réciproque des voies de signalisation de PPAR $\alpha$  et de STAT5b (signal transducer and activator of transcription 5b), cette dernière voie étant activée par diverses cytokines et facteurs de croissance dont l'hormone de croissance (Shipley and Waxman, 2003; Zhou and Waxman, 1999). De plus, une étude récente, utilisant des criblages par microarrays de cultures cellulaires exprimant de manière stable les trois PPARs, ou des mutants dont certains domaines ont été échangés entre les isoformes, suggère que le domaine A/B déterminerait la spécificité des gènes cibles des trois isoformes en inhibant la capacité de transactivation des PPARs vis-à-vis de certains gènes cibles (Hummasti and Tontonoz, 2006). Ainsi, un mutant de PPAR $\beta$  dont le domaine A/B a été supprimé, est capable d'activer les gènes cibles spécifiques de PPAR $\gamma$  et d'induire la différenciation adipocytaire de cellules NIH-3T3 (Hummasti and Tontonoz, 2006). Le domaine A/B des PPARs semble également impliqué dans des interactions avec certains corégulateurs bien que les données dans ce domaine restent limitées et que le domaine E (LBD) semble être le plus impliqué dans l'interaction entre les PPARs et leurs corégulateurs (Yu and Reddy, 2007). Une interaction indépendante du ligand entre p300 et PPAR $\gamma$ 2 a été démontrée (Gelman et al., 1999). Certains auteurs suggèrent que la fonction AF-1 de PPAR $\alpha$  interviendrait dans la fixation de corépresseurs (Juge-Aubry et al., 1999) mais peu de données actuellement disponibles permettent de confirmer définitivement ces hypothèses. À ma connaissance, la seule interaction physique rapportée pour la fonction AF-1 de PPAR $\alpha$  est une interaction avec l'enzyme bifonctionnelle peroxyxsomale (Juge-Aubry et al., 2001) qui semble capable d'activer cette fonction, créant ainsi une boucle d'autorégulation positive. Signalons cependant que le domaine A/B de PPAR $\alpha$  semble bien intervenir dans la fixation de corégulateurs en l'absence, comme en présence de ligand exogène comme cela a été récemment démontré par le biais de mesure du coefficient de diffusion d'une forme tronquée de PPAR $\alpha$  par spectroscopie à corrélation de fluorescence (Tudor et al., 2007).

### **II.3.3. Domaine C et éléments de réponse de PPAR $\alpha$**

Le domaine C (DBD) des PPARs leur confère la capacité de reconnaître des éléments de réponse de type DR1, ainsi que certains éléments de réponse de type DR2, appelés PPRES (PPAR response elements). Les PPARs se fixent toujours à l'ADN sous la forme d'un dimère avec le récepteur RXR (Kliewer et al., 1992) et sont incapables de se fixer sous la forme d'homodimères ou de monomères (Desvergne and Wahli, 1999; Hsu et al., 1998). Les premiers éléments de réponse de type DR1 observés dans les gènes cibles identifiés des PPARs (IJpenberg et al., 1997; Juge-Aubry et al., 1997) révèlèrent quelques éléments supplémentaires déterminant le fait qu'un DR1 se comporte comme un PPRE : 1) une extension 5' riche en bases A et T (IJpenberg et al., 1997; Palmer et al., 1995) de manière similaire à ce que nous avons décrit pour les récepteurs se fixant sous la forme de monomères (voir paragraphe I.2.2), 2) un motif de fixation (AGGTCA) imparfait (Juge-Aubry et al., 1997) et 3) une adénine comme base séparant les deux motifs (IJpenberg et al., 1997). En se basant sur ces critères, le DR1 / PPRE typique présente la séquence suivante (les deux motifs ont été soulignés) : 5'-AACTAGGNCAAAGGTCA-3' (Desvergne and Wahli, 1999; IJpenberg et al., 1997). Les caractéristiques d'un tel élément de réponse participeraient à la sélectivité de reconnaissance par le dimère PPAR-RXR par rapport à d'autres homo- ou hétérodimères de RNs (Desvergne and Wahli, 1999). Sur un tel élément de réponse, PPAR occupe le motif situé en 5' et RXR le motif situé en 3' (DiRenzo et al., 1997; IJpenberg et al., 1997) contrairement aux dimères RXR-RAR, RXR-TR et RXR-VDR où RXR interagit avec le motif situé en 5' des éléments de réponse de type DR5, DR4 et DR3 respectivement. Il a été montré que l'extension C-terminale (CTE, voir [Figure 5](#)) du DBD de PPAR $\alpha$  était responsable de la reconnaissance des bases qui jouxtent le DR1 en 5' (Hsu et al., 1998). De plus, le DBD de PPAR $\alpha$  semble peu ou pas impliqué dans la dimérisation avec RXR (Hsu et al., 1998), contrairement à ce qui est observé avec les dimères de DBD entre RXR et RAR (sur un DR5) ou TR (sur un DR4). L'incapacité de PPAR $\alpha$  à se fixer sur l'ADN sous la forme d'un monomère a été attribuée à son domaine A/B dans la mesure où sa délétion permet au récepteur tronqué de se fixer sur un PPRE sous la forme de monomère (Hsu et al., 1998). Certains DR2 capables de fixer des monomères ou des homodimères des récepteurs Rev-Erb et nommés Rev-DR2 sont également capables de fixer PPAR $\alpha$  (Gervois et al., 1999a; Hsu et al., 1998). Ces DR2 se caractérisent par une extension en 5' riche en A/T et par la présence des bases C et T entre les deux motifs AGGTCA. Ces observations suggèrent qu'une

interaction est possible entre les voies PPAR $\alpha$  et Rev-Erb pour la régulation des gènes contenant de tels éléments de réponse, comme cela a été montré sur le Rev-DR2 présent dans le promoteur de Rev-ErbA- $\alpha$  lui-même (Gervois et al., 1999a). Les PPARs semblent également capable de se fixer à certains éléments de réponse des ERs (nommés ERE et constitués d'un ER3) et de réprimer l'action des estrogènes sur une cible classique comme le promoteur de la vitellogénine A2 (Keller et al., 1995). Via leur capacité à se fixer sur des éléments de réponse de type DR1, les homodimères de RXR, HNF-4, TR2, TR4 ou COUP-TF, ainsi que les hétérodimères RAR-RXR et COUP-TF-RXR seraient susceptibles d'interagir avec la fixation des dimères PPAR-RXR sur leurs éléments de réponse (Desvergne and Wahli, 1999). Chacune de ces interactions potentielles a d'ores et déjà fait l'objet d'une démonstration *in vitro* et/ou *in vivo* (Baes et al., 1995; DiRenzo et al., 1997; Hertz et al., 1995; Hertz et al., 1996; IJpenberg et al., 2004; Inui et al., 2003; Miyata et al., 1993; Rodriguez et al., 1998; Winrow et al., 1994; Yan et al., 1998). Dans ce contexte, la séquence précise du DR1 et du promoteur dans son ensemble (Nakshatri and Bhat-Nakshatri, 1998) ainsi que la concentration relative des récepteurs et de leurs ligands éventuels détermineraient vraisemblablement l'établissement de telles interactions. Enfin, signalons qu'aucune structure tridimensionnelle du DBD des PPARs n'a actuellement été publiée.

#### **II.3.4. Domaines D et E et corégulateurs de PPAR $\alpha$**

Le domaine D (domaine charnière) des PPARs contient une partie de l'extension C-terminale du DBD que nous venons d'évoquer et participe donc à la spécificité des éléments de réponse reconnus par ces récepteurs. De plus, il semble que ce domaine participe, principalement avec le domaine E (LBD) mais également avec le domaine C (DBD), à l'interaction entre les PPARs et divers corégulateurs. Par exemple, ce domaine est nécessaire mais pas suffisant pour l'interaction entre PPAR $\alpha$  et le corépresseur NCoR (Dowell et al., 1999). Il est également impliqué dans l'interaction entre p300 et PPAR $\alpha$  (Dowell et al., 1997) qui s'opère en l'absence comme en présence d'un ligand (IJpenberg et al., 2004; Tudor et al., 2007). Le coactivateur PRIC285/PDIP1 (PPAR $\alpha$ -interacting cofactor of 285 KDa / PPAR $\gamma$ -DBD-interacting protein 1) a été isolé indépendamment par deux équipes différentes (Surapureddi et al., 2002; Tomaru et al., 2006) qui ont démontré son interaction notamment avec les récepteurs PPAR $\alpha$  (Surapureddi et al., 2002) et PPAR $\gamma$  (Tomaru et al., 2006). Initialement isolé *in vivo* au sein d'un large complexe multiprotéique contenant de nombreux coactivateurs de PPAR $\alpha$  (Surapureddi et al., 2002), il a depuis été montré que PRIC285 se

fixait de manière spécifique à PPAR $\gamma$  au niveau d'une région comprenant le domaine C et le domaine D de PPAR $\gamma$  (Tomaru et al., 2006). La région de PPAR $\alpha$  impliquée dans l'interaction avec PRIC285 n'a pas encore été précisée. Il est possible qu'elle soit distincte de celle impliquée dans l'interaction avec PPAR $\gamma$  car PRIC285 contient plusieurs motifs LXXLL (Surapureddi et al., 2002; Tomaru et al., 2006) et que la fixation d'un corégulateur sur des domaines différents de ces deux PPARs a déjà été décrite pour PGC-1 $\alpha$  qui se fixe sur les domaines C et D de PPAR $\gamma$ , y compris en l'absence de ligand (Puigserver et al., 1998), mais se fixe principalement sur le domaine E de PPAR $\alpha$  en présence de ligand (en l'absence de ligand une faible fixation est observée qui semble également impliquer le domaine D, Vega et al., 2000).

Le domaine E (LBD) est impliqué dans la reconnaissance des ligands que nous avons abordée au paragraphe II.2. ainsi que dans la dimérisation avec RXR et l'interaction avec de nombreux corégulateurs. De nombreuses structures des LBD des PPARs, en l'absence de ligand, en présence d'un agoniste ou d'un antagoniste ont été élucidées ([Table 3](#)) et ont été récemment synthétisée (Zoete et al., 2007), nous permettant de ne souligner ici que les principaux aspects tirés de l'analyse de ces structures. La structure globale du LBD des PPARs est similaire à celle des autres RNs (sandwich d'hélices  $\alpha$ ). Elle comporte 13 hélices (une hélice H2' supplémentaire est située entre le premier feuillet  $\beta$  S1 et H3 et les hélices H10 et H11 forment en fait une unique hélice continue) et 4 feuillets  $\beta$  relativement courts notés S1 à S4 (pour une correspondance entre la séquence des LBDs des PPAR et ces éléments structuraux, voir Desvergne and Wahli, 1999). Le site de fixation du ligand est une large cavité (1300-1400 Å<sup>3</sup>) en forme de Y délimitée par les hélices H2', H3, H4, H5, H7, H10/H11 et H12 et par les feuillets  $\beta$  S3 et S4 et dont 80% des 34 acides aminés la délimitant sont conservés entre les trois isoformes de PPAR. L'un des bras du Y constitue une cavité polaire où se fixe généralement la tête polaire des ligands (groupement carboxyle des acides gras par exemple) et les deux autres branches, dont la cavité d'entrée du ligand, constituent des poches hydrophobes où se fixe généralement la queue hydrophobe des ligands (chaîne carbonée des acides gras par exemple). La boucle située entre les hélices H2' et H3, localisée à l'entrée de la poche de fixation du ligand, est très flexible et apparaît parfois désordonnée dans certaines structures ce qui participerait à la capacité des PPARs à fixer des ligands volumineux, sans trop affecter la structure globale du LBD. Le LBD des PPARs apparaît comme une structure hautement dynamique en l'absence de ligand (signalons que la ré-analyse de la structure 2GXW de PPAR $\beta$  a néanmoins conduit à l'identification d'un ligand

naturel issu des bactéries utilisées pour produire la protéine, Fyffe et al., 2006a; Fyffe et al., 2006b) et la fixation du ligand agoniste se traduit par une stabilisation globale de la structure (qui devient plus compacte et rigide) dans une conformation favorisant le recrutement de coactivateurs. En l'absence de ligand (du moins de ligand volontairement ajouté), l'hélice H12 (fonction AF-2) des PPARs semble pouvoir malgré tout prendre une conformation active grâce à une série de quatre résidus qui établissent un réseau de liaisons hydrogènes stabilisant l'hélice H12 dans sa conformation active (pour PPAR $\alpha$ , voir notamment Michalik et al., 2007). Un de ces résidus, avec trois autres résidus, participent à la stabilisation de l'hélice H12 en présence d'un ligand via des interactions avec des résidus adjacents du LBD et avec le ligand. Ces données expliqueraient la capacité des PPARs à interagir avec des coactivateurs et à réguler l'expression génique y compris en l'absence de ligand et permettent d'envisager la génération de souris transgéniques exprimant des formes de PPARs dissociant leurs activités dépendantes ou indépendantes des ligands. Par ailleurs, le LBD des PPARs constitue la principale interface de dimérisation avec le récepteur RXR. Plusieurs structures de PPAR $\gamma$  en dimère avec RXR $\alpha$  et en présence de peptides issus de coactivateurs ont été élucidées ([Table 3](#)) et montrent que le LBD de PPAR $\gamma$  est tourné de 10° par rapport à l'axe de symétrie C2 (axe autour duquel une rotation de  $\pi$  de tout élément E d'une des molécules donne l'élément E' correspondant à E dans l'autre molécule), conduisant à générer une interface de dimérisation asymétrique (Gampe et al., 2000a). Cette interface est composée d'un réseau d'interactions polaires et hydrophobes impliquant les hélices H7, H9, la boucle entre H8 et H9 et tout particulièrement l'hélice H10 qui pourrait expliquer notamment la capacité de PPAR $\gamma$  à être activé par des ligands de RXR (positionnement de H12 de PPAR $\gamma$  sous forme active suite à la fixation d'un ligand de RXR) et l'activation synergique du dimère par les activateurs des deux partenaires du dimère (Zoete et al., 2007). La dimérisation avec RXR pourrait également influencer le positionnement du ligand de PPAR $\gamma$  qui se présente sous deux conformations différentes dans les structures en présence (Gampe et al., 2000a) ou en l'absence (Nolte et al., 1998) de RXR. Enfin, le LBD des PPARs constitue le site principal de fixation des coactivateurs et corépresseurs, des fixations qui s'opèrent essentiellement comme nous l'avons décrit au paragraphe I.4.5. De nombreux coactivateurs ont été identifiés qui interagissent avec les PPARs via leurs motifs LXXLL et de manière dépendante ou indépendante du ligand (ou accrue en présence de ligand) en impliquant la fonction AF-2 des PPARs (pour des synthèses récentes sur les corégulateurs des PPARs, voir Feige et al., 2006; Yu and Reddy, 2007) dont certains ont déjà été évoqués comme PGC-1 $\alpha$  ou p300. Parmi les

protéines isolées dans un complexe de coactivateurs de PPAR $\alpha$  in vivo (complexe PRIC pour PPAR $\alpha$ -interacting cofactor complex, Surapureddi et al., 2002), on trouve trois protéines (TOG, CHD5 et TIP120) possédant des domaines LXXLL et appartenant à la famille SWI/SNF intervenant dans le remodelage ATP-dépendant des histones (voir paragraphe I.5.2.), suggérant que ces complexes interviennent in vivo dans l'activité de PPAR $\alpha$ . Parmi les coactivateurs du deuxième groupe définit au paragraphe I.5.3., PPAR $\alpha$  est capable d'interagir avec SRC-1 et SRC-3 (Surapureddi et al., 2002) mais son affinité pour SRC-1 en présence de plusieurs ligands semble plus faible qu'avec CBP/p300 ou PGC-1 $\alpha$  (Mukherjee et al., 2002). De plus, les souris déficientes pour SRC-1 présentent une réponse inaltérée aux agonistes de PPAR $\alpha$  (Qi et al., 1999b) qui pourrait néanmoins être due à une surexpression compensatoire d'autres corégulateurs p160 chez ces souris (Xu and Li, 2003). Ces données suggèrent que les coactivateurs p160, s'ils appartiennent bien au complexe de coactivateurs des PPARs, ne jouent pas un rôle déterminant dans l'activité de ces facteurs de transcription (Feige et al., 2006). En revanche, les coactivateurs CBP et p300 interagissent plus fortement avec les PPARs (Dowell et al., 1997; Feige et al., 2006; Gelman et al., 1999; Mukherjee et al., 2002; Zhou et al., 1998a). CBP a été également identifié dans le complexe PRIC (Surapureddi et al., 2002). De plus, p300 interagit avec PPAR $\alpha$  y compris en l'absence de ligand et cette association est renforcée par l'ajout d'un agoniste et légèrement réduite par l'ajout d'un antagoniste (Tudor et al., 2007). Enfin, la démonstration du rôle important de CBP comme coactivateur des PPARs a été faite dans des fibroblastes embryonnaires de souris CBP $^{+/-}$  qui présentent une activité transcriptionnelle réduite des PPARs (Yamauchi et al., 2002). Nous avons déjà évoqué le rôle important de PGC-1 $\alpha$  comme coactivateur des PPARs et son interaction différente entre PPAR $\gamma$  et PPAR $\alpha$ . Lui aussi a été identifié dans le complexe PRIC (Surapureddi et al., 2002). Une interaction entre PPAR $\alpha$  et PGC-1 $\beta$  a également été démontrée (Lin et al., 2003) mais les conséquences fonctionnelles de cette interaction ont été peu étudiées. D'autres protéines identifiées dans le complexe PRIC (Surapureddi et al., 2002) n'étaient pas connues pour être des coactivateurs des RNs et certaines, comme PRIC285 (Surapureddi et al., 2002) et PRIC320 (Surapureddi et al., 2006), ont commencé à être étudiées et montrent bien une interaction avec PPAR $\alpha$  et une fonction de coactivateur de ce récepteur. Des études complémentaires comprenant la génération de lignées murines déficientes pour ces coactivateurs récemment identifiés devraient permettre de préciser l'étendue de leurs fonctions vis-à-vis des PPAR et des RNs en général. Les coactivateurs PRIP (PPAR $\alpha$ -interacting protein, Zhu et al., 2000a) et PIMT (PRIP-interacting protein with

methyltransferase domain, Zhu et al., 2001b) agissent de manière synergique pour activer la transcription suite à l'activation de PPAR $\gamma$  ou RXR. Il a été proposé que ces corégulateurs agissent comme un pont entre le complexe CBP/p300 à activité HAT et le complexe Médiateur (Misra et al., 2002; Qi et al., 2003). La délétion de PRIP chez la souris est létale au niveau embryonnaire (Antonson et al., 2003; Zhu et al., 2003), démontrant sa fonction indispensable et non redondante au cours du développement. Les études sur fibroblastes embryonnaires issus de ces souris ont démontré le rôle de PRIP comme coactivateur de PPAR $\gamma$  (Antonson et al., 2003; Zhu et al., 2003) nécessaire à la différenciation adipocytaire PPAR $\gamma$ -dépendante (Qi et al., 2003). Cependant, malgré la présence de PRIP et PIMT dans le complexe PRIC (Surapureddi et al., 2002), la délétion spécifique de PRIP dans le foie ne semble pas altérer la réponse de cet organe aux agonistes de PPAR $\alpha$  et de CAR (Guo et al., 2006a; Sarkar et al., 2007), suggérant que PRIP n'est pas indispensable à la fonction de ces deux RNs. En revanche, il semble que le complexe Médiateur joue un rôle important comme coactivateur de PPAR $\alpha$ , comme l'illustre la délétion spécifique de Med1 (également appelé PBP pour PPAR $\gamma$ -binding protein) dans le foie qui mime la délétion de PPAR $\alpha$  dans la mesure où la réponse au Wy14643 en termes de prolifération des peroxysomes et d'induction des gènes cibles de PPAR $\alpha$  est abrogée (Jia et al., 2004). Plus récemment, l'équipe du Dr JK Reddy a également montré que la délétion de Med1 au niveau hépatique bloquait, dans les hépatocytes Med1 $^{-/-}$ , la formation d'hépatocarcinomes provoquée par l'administration à long terme chez les rongeurs d'agonistes de PPAR $\alpha$  (Matsumoto et al., 2007). De plus, ces animaux présentent également une altération de la réponse aux activateurs de CAR (Guo et al., 2006a; Jia et al., 2005; Matsumoto et al., 2007) démontrant un rôle clef de Med1 dans la translocation nucléaire et l'activité de ce RN. La délétion complète de Med1 chez la souris est également létale au niveau embryonnaire (Crawford et al., 2002; Ito et al., 2000; Zhu et al., 2000b) et, comme pour PRIP, les fibroblastes embryonnaires immortalisés Med1 $^{-/-}$  perdent la capacité à se différencier en adipocytes et à induire les gènes cibles de PPAR $\gamma$  sous l'action des thiazolidinediones (Ge et al., 2002). Signalons également que quatre protéines appartenant au complexe Médiateur (Med1, Med12, Med23 et Med24) ont été identifiées dans le complexe PRIC (Surapureddi et al., 2002). Quelques autres coactivateurs interagissant avec PPAR $\alpha$  ont été identifiés (pour des synthèses, voir Feige et al., 2006; Yu and Reddy, 2007) mais leurs fonctions et leur pertinence au plan physiologique restent pour l'instant obscures et nécessitent des investigations complémentaires. Outre ces coactivateurs, les PPARs interagissent également avec des corépresseurs. Une structure de PPAR $\alpha$  en présence d'un

antagoniste et d'un peptide corépresseur a d'ailleurs été élucidée (Xu et al., 2002a). En l'absence de ligand agoniste, les corépresseurs NCoR et SMRT sont capables d'interagir avec les trois PPARs (Dowell et al., 1999; Hu et al., 2001; Krogsdam et al., 2002; Stanley et al., 2003; Yu et al., 2005). Il est vraisemblable que la dynamique de l'hélice H12, dont le déplacement en position active est nécessaire pour permettre la fixation des coactivateurs et dont la délétion accroît généralement l'interaction avec les corépresseurs (voir par exemple Hu et al., 2001), permette ces interactions. En présence d'un ligand antagoniste qui empêche l'hélice H12 de se placer en position active, l'interaction avec les corépresseurs est maintenue, voire accrue (Tudor et al., 2007; Xu et al., 2002a) ce qui participe aussi à bloquer le déplacement de H12 dans sa position active et donc, le recrutement de coactivateurs à la surface du LBD (voir Figure 12 et Xu et al., 2002a). En présence d'un ligand agoniste, qui opère un réseau d'interactions avec l'hélice H12 et favorise ainsi son déplacement en position active, les corépresseurs sont libérés (Dowell et al., 1999; Krogsdam et al., 2002; Stanley et al., 2003; Yu et al., 2005) et les coactivateurs peuvent se fixer à la surface du LBD. Il est probable qu'il s'agisse d'un phénomène moins séquentiel que ce que nous présentons ici et que la fixation de chaque élément (ligand, coactivateur et corépresseur) sur le LBD influence positivement ou négativement la probabilité de fixation des autres éléments. Initialement, il a été montré que PPAR $\gamma$  était capable d'interagir avec NCoR et SMRT en solution mais que sa fixation au promoteur de l'acyl-CoA oxydase empêchait cette interaction, suggérant que PPAR $\gamma$  ne pouvait réprimer l'expression des gènes en l'absence de ligand (Zamir et al., 1997). D'autres études ont néanmoins identifié *in vitro* un complexe PPAR $\gamma$ -RXR-NCoR sur ce même promoteur (Krogsdam et al., 2002) ainsi qu'un complexe similaire mais contenant PPAR $\beta$ . Plus récemment, il a été montré que l'inhibition de NCoR et SMRT par des siRNA dans des cellules 3T3-L1 provoquait une augmentation de l'expression des gènes spécifiques de l'adipocyte facilitant ainsi la différenciation adipocytaire (Yu et al., 2005). Finalement, la capacité de PPAR $\gamma$  à inhiber l'expression de certains de ses gènes cibles en l'absence de ligand et de manière différentielle en fonction du contexte du promoteur a pu être démontrée récemment (Chui et al., 2005; Guan et al., 2005). En l'absence de ligand exogène, PPAR $\gamma$  est associé à un complexe corépresseur contenant NCoR et/ou SMRT sur les PPRE des gènes de la glycérol kinase (GyK, Guan et al., 2005) et sur le récepteur 1 aux LDL oxydés (OLR1 pour oxidized LDL receptor 1, Chui et al., 2005) mais pas sur le promoteur d'aP2 où le dimère PPAR $\gamma$ -RXR est associé avec des coactivateurs, participant ainsi à l'activation constitutive de ce promoteur. Suite à l'activation de PPAR $\gamma$  par la rosiglitazone, les corépresseurs sont libérés

des promoteurs GyK et OLR1 et des coactivateurs sont recrutés, induisant ainsi l'expression de ces gènes. L'observation de ces différences entre GyK et OLR1 d'une part et aP2 d'autre part dans les mêmes cellules, suggère que ce sont le contexte du promoteur (et pas uniquement la séquence du DR1, Guan et al., 2005) et la fixation à l'ADN qui imposent une conformation spécifique du RN déterminant l'affinité différentielle pour les coactivateurs ou les corépresseurs. Contrairement au cas de PPAR $\gamma$ , beaucoup moins de données sont disponibles sur ces aspects de la régulation négative de l'activité de PPAR $\alpha$  ou PPAR $\beta$  et des études supplémentaires sont nécessaires en prenant en compte le fait que l'action des corépresseurs apparaît nettement spécifique du promoteur étudié (Chui et al., 2005; Guan et al., 2005). Enfin, d'autres corépresseurs semblent pouvoir se fixer sur les PPARs mais leurs rôles précis ont été nettement moins étudiés que ceux de NCoR et SMRT (pour une synthèse, voir Feige et al., 2006).

### **II.3.5. Dynamique de fonctionnement de la voie PPAR $\alpha$**

La somme des informations que nous avons présentée dans ce chapitre permet d'entrevoir le mode de fonctionnement vraisemblable de la voie de signalisation de PPAR $\alpha$ . D'importantes informations sur le caractère dynamique de fonctionnement de la voie PPAR $\alpha$  ont été récemment apportées par l'utilisation de diverses techniques de microscopie sur cellules vivantes (Feige et al., 2005; Tudor et al., 2007). Ces études révèlent que PPAR $\alpha$  est exclusivement localisé dans le noyau où il s'y trouve présent de manière diffuse et est exclu des nucléoles (il en est de même pour PPAR $\beta$ , PPAR $\gamma$  et RXR $\alpha$ , Feige et al., 2005). En l'absence de ligand exogène, PPAR $\alpha$  est extrêmement mobile dans le noyau où il se trouve sous la forme d'un hétérodimère avec RXR et est associé, principalement via ses domaines A/B et E, à un ou des volumineux complexes de corégulateurs (Feige et al., 2005; Tudor et al., 2007). Ces données semblent cohérentes avec une activité constitutive de ces récepteurs, qui pourrait d'ailleurs se traduire par une activation ou une répression constitutive en fonction des promoteurs considérés et des corégulateurs associés. Malgré sa forte mobilité, il a été montré que le coefficient de diffusion de PPAR $\alpha$  était cependant nettement plus faible que si le récepteur (ou le récepteur dimérisé avec RXR) diffusait de manière libre dans le noyau. Ceci semble lié très majoritairement à la présence des complexes de corégulateurs et non à la fixation transitoire sur la chromatine (éléments de réponse ou séquences proches des éléments de réponse) qui n'affecte vraisemblablement qu'une très faible proportion du pool total de récepteurs présents dans le noyau à un instant donné. La fixation d'un ligand agoniste réduit

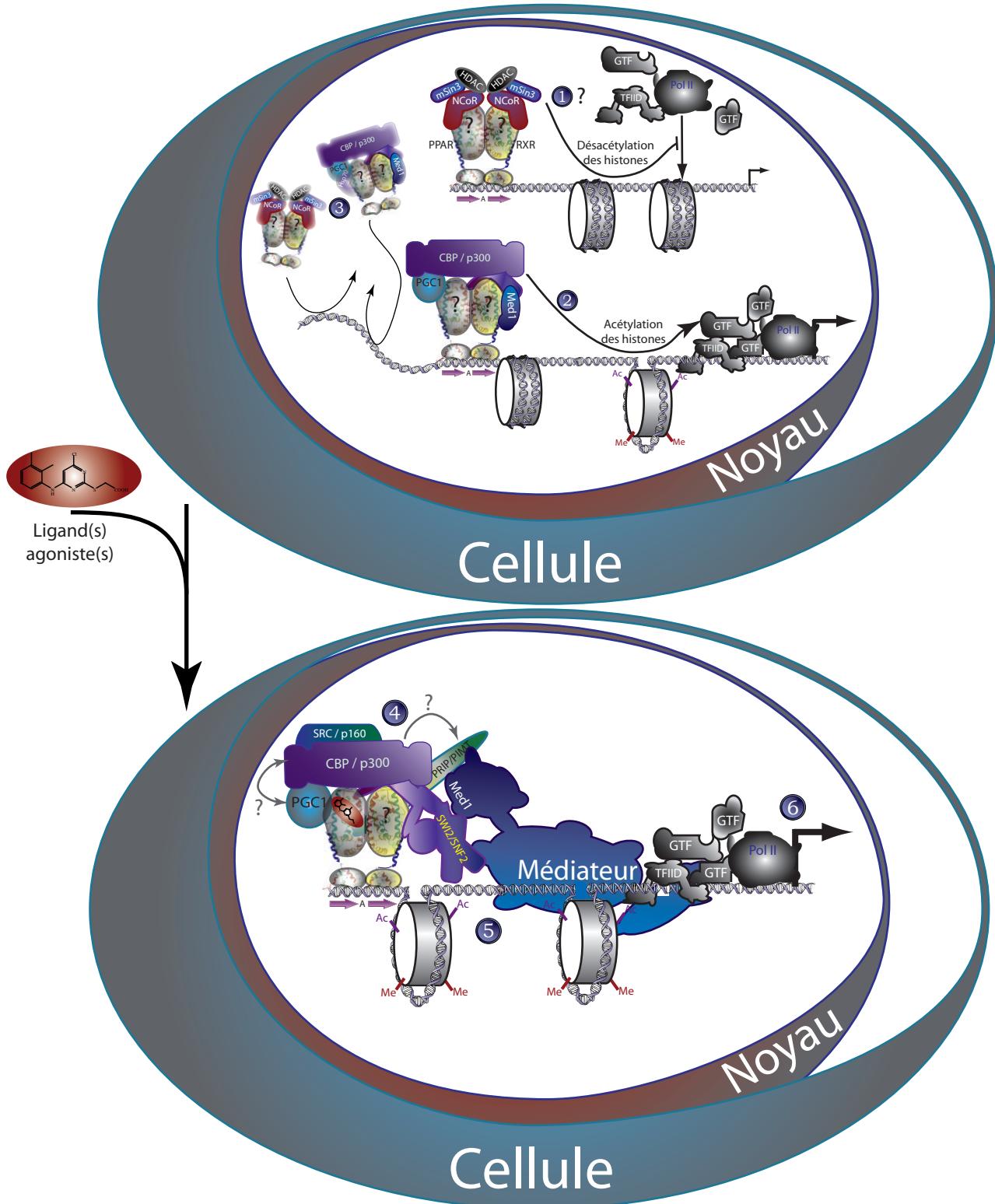


Figure 22 : Fonctionnement général de la voie PPAR $\alpha$

- ① En l'absence de ligand (les ? dans les LBD des RNs illustrent que la présence éventuelle de ligands endogènes n'est généralement pas maîtrisée dans les expériences sur les PPARs), il est possible que PPAR $\alpha$  soit capable de réprimer constitutivement l'expression de certains gènes via son association avec des corépresseurs. ② Son association avec des coactivateurs lui permet d'activer constitutivement l'expression de certains gènes cibles. ③ De nombreux dimères PPAR $\alpha$ -RXR, associés à des complexes corégulateurs, se déplacent rapidement dans le noyau et certains interagissent de manière transitoire avec l'ADN. ④ Lors de l'ajout d'un agoniste, l'association du dimère PPAR $\alpha$ -RXR avec des complexes corépresseurs est déstabilisée et d'autres coactivateurs sont recrutés (les flèches et ? gris illustrent que les interactions précises et éventuellement les modifications post-traductionnelles entre coactivateurs sont actuellement mal connues), formant ainsi un volumineux complexe qui réduit la mobilité des dimères PPAR $\alpha$ -RXR dans le noyau, ⑤ agit sur la chromatine (modification post-traductionnelles, remodelage ATP-dépendant, autres?) et facilite la formation et la fonction du complexe de préinitiation de la transcription via un complexe Médiateur. ⑥ Le résultat est un accroissement notable du niveau de transcription des gènes cible.

la mobilité du dimère PPAR $\alpha$ -RXR ce qui est, là encore, principalement lié au recrutement d'un complexe de corégulateurs de taille plus importante (Feige et al., 2005; Tudor et al., 2007). A l'opposé, la fixation d'un ligand antagoniste accroît la mobilité de PPAR $\alpha$  via la libération des coactivateurs et le recrutement de complexes corépresseurs de taille plus réduite. Un modèle reprenant les résultats de leurs études et ceux issus d'autres auteurs a été proposé par l'équipe du Pr W. Wahli (Feige et al., 2006).

La [Figure 22](#) résume le fonctionnement général de la voie PPAR $\alpha$  soumise ou non à une exposition à un agoniste mais ne peut naturellement pas prendre en compte tous ces aspects dynamiques. Elle met néanmoins en évidence l'activité constitutive du récepteur et l'échange des corégulateurs ainsi que quelques questions qui restent pour l'instant en suspens comme la présence éventuellement constitutive de ligands de PPAR $\alpha$ , y compris en l'absence de ligand volontairement ajouté, la capacité du récepteur à réprimer constitutivement l'expression de certains gènes ou encore le rôle d'interactions et de modifications post-traductionnelles des coactivateurs dans l'échange des corégulateurs. Finalement, dans ce paragraphe nous n'avons que ponctuellement abordé l'interaction entre PPAR $\alpha$  et d'autres facteurs de transcription mais ces aspects seront plus largement évoqués dans les paragraphes suivants.

## II.4. REGULATION DU METABOLISME PAR PPAR $\alpha$

Le rôle de PPAR $\alpha$  dans la régulation du métabolisme énergétique est sans doute la plus reconnue de ses fonctions. Comme nous allons le voir, PPAR $\alpha$  régule de multiples aspects du métabolisme des lipides : leur transport par les lipoprotéines, le transport et le trafic intracellulaire des acides gras, le catabolisme des lipides et la cétogenèse. De plus, PPAR $\alpha$  semble également intervenir dans la régulation du métabolisme des glucides et des acides aminés.

### II.4.1. Régulation du métabolisme des lipoprotéines par PPAR $\alpha$

Dans l'organisme, les lipides sont véhiculés par le plasma et la lymphé où ils sont associés à des protéines (apolipoprotéines) pour former les lipoprotéines (voir [Figure 23](#) pour le transport du cholestérol par les lipoprotéines, Toussaint et al., 2003). Un faible pourcentage des lipides circule aussi de manière libre ou lié à l'albumine. On distingue classiquement les lipoprotéines contenant de l'apolipoprotéine B (apoB) et dénommées Lp-B (chylomicrons, very low density lipoprotein ou VLDL, intermediate density lipoprotein ou IDL et low density

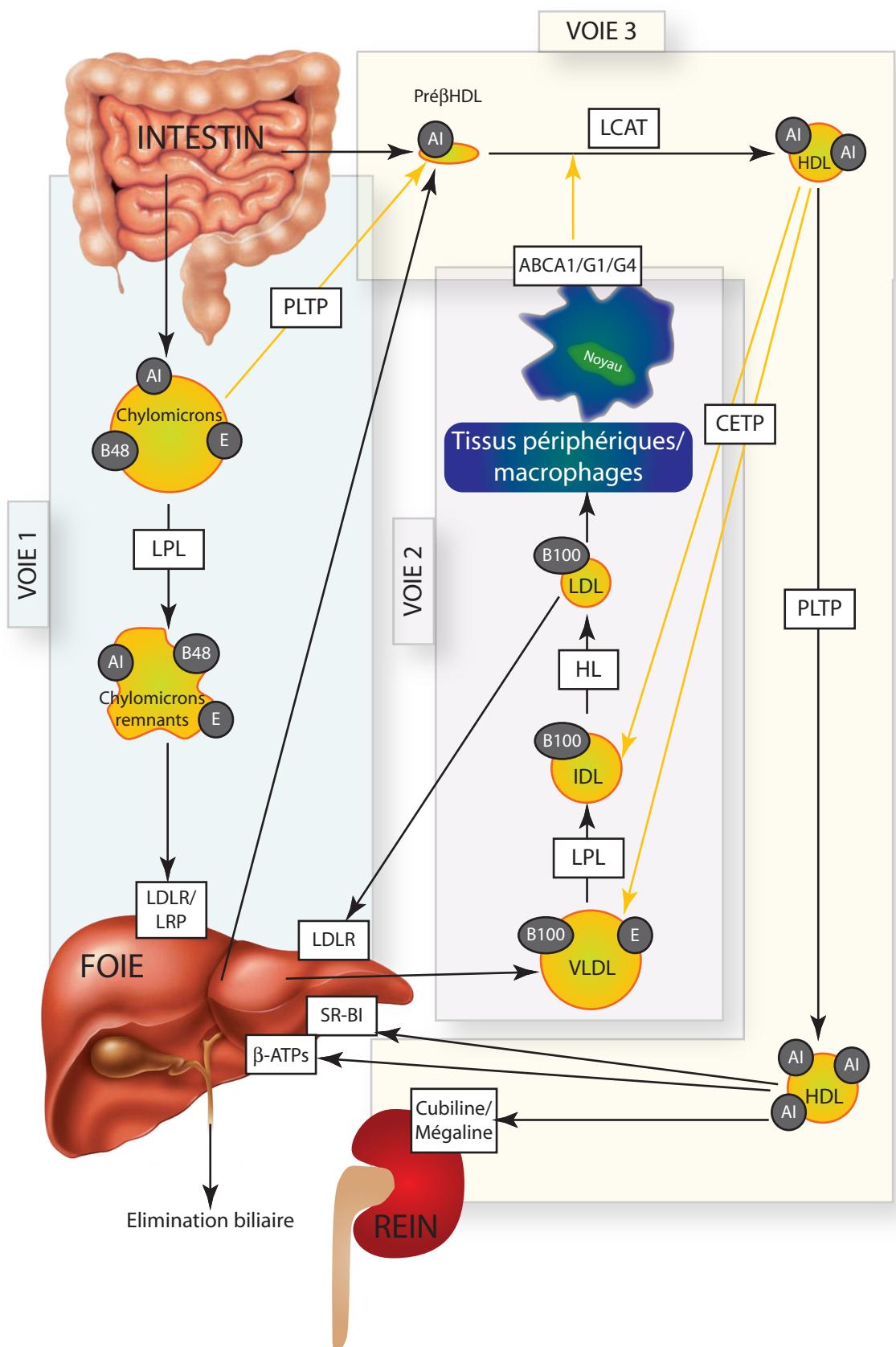


Figure 23 : Schéma général du transport du cholestérol

Voie 1 : voie entéro-hépatique. Voie 2 : Voie endogène d'apport aux tissus périphériques. Voie 3 : Voie de retour ou transport inverse du cholestérol. Ces différentes voies sont décrites au paragraphe II.4.1. Schéma inspiré de Toussaint JF, Jacob MP, Lagrost L, Chapman J (2003), "l'athérosclérose -physiologie, diagnostics, thérapeutiques", Masson, Paris. Les flèches jaunes correspondent à des échanges de lipides et/ou d'apolipoprotéines. LCAT : lecithine:cholesterol acyltransferase, CETP : cholesteryl ester transfer protein, PLTP : phospholipid transfer protein, LPL : lipoprotein lipase, HL : hepatic lipase, LDLR : LDL receptor, LRP : LDL receptor-related protein, SR-BI : scavenger receptor class B type I, ABCA1 : ATP-binding cassette type A1

lipoprotein ou LDL) et les HDL (high density lipoprotein) qui sont dépourvues d'apoB et contiennent principalement de l'apoA-I (voir [Table 7](#)).

	<b>Chylomicrons</b>	<b>VLDL</b>	<b>LDL</b>	<b>HDL</b>
<b>Densité (g/mL)</b>	0,94	0,94-1,006	1,006-1,063	1,063-1,21
<b>Triglycérides (% masse)</b>	80-95	45-65	4-8	2-7
<b>Phospholipides (% masse)</b>	3-6	15-20	18-24	26-32
<b>Cholestérol libre (% masse)</b>	1-3	4-8	6-8	3-5
<b>Esters de cholestérol (% masse)</b>	2-4	16-22	45-50	15-20
<b>Protéines (% masse)</b>	1-2	6-10	18-22	45-55
<b>Taille (nm)</b>	75-1200	30-80	18-35	<7-12
<b>Mobilité électrophorétique</b>	Cathode B-48, A-I, A-IV, C-I, C-III, E	pré-β B-100, E, C-I, C-II, C-III	β B-100	α A-I, A-II, E, A-IV
<b>Principales apolipoprotéines</b>				A-V, C-I, C-II, C-III, D, E, G, H, J, L, SAA
<b>Apolipoprotéines mineures</b>	A-II, C-II, SAA	A-I, A-II, A-IV, A-V, D	C-I, C-II, C-III, D, E, F	

**Table 7 :** Composition biochimique et propriétés des principales lipoprotéines plasmatiques humaines

Les IDL ne sont pas présentées dans ce tableau car leurs caractéristiques correspondent à un intermédiaire entre les VLDL et les LDL (pour les propriétés physico-chimiques, les IDL ont été regroupées avec les LDL dans ce tableau). Les mobilités électrophorétiques des lipoprotéines se rapportent à celles des globulines  $\alpha_1$  et  $\beta$  mais aujourd'hui, l'utilisation de l'ultracentrifugation pour la séparation des lipoprotéines, conduit à prendre plutôt en considération leurs densités respectives. Dans l'intestin grêle qui produit les chylomicrons, l'ARNmessager issu du gène de l'apoB subit une édition (un codon pour une glutamine CAA est modifié en codon stop UAA) via l'APOBEC-1 (apolipoprotein B mRNA editing catalytic component 1) ce qui conduit à la production de l'apoB-48 représentant 48% de la partie N-terminale de l'apoB-100. SAA : serum amyloid A protein (impliquée dans la phase aiguë de l'inflammation).

De manière schématique, les chylomicrons transportent les lipides d'origine alimentaire (principalement triglycérides, phopholipides et esters de cholestérol) de la muqueuse intestinale vers les différents tissus de l'organisme via la lymphe puis la circulation sanguine (**voie entéro-hépatique**). Les triglycérides (TG) sont ensuite hydrolysés par la lipoprotéine lipase (LPL), enzyme lipolytique ancrée par des protéoglycanes à l'endothélium des capillaires sanguins de nombreux tissus, et les acides gras (AG) résultant de cette hydrolyse sont captés par les tissus. Les remnants de chylomicrons sont ensuite captés par le foie via les récepteurs aux LDL (LDLR pour LDL receptor et LRP pour LDL receptor-related

protein). Les VLDL sont également responsables de l'apport d'AG aux divers tissus (**voie d'apport endogène**). Synthétisées par le foie, elles transportent les lipides d'origine hépatique vers les autres tissus de l'organisme. Au cours de l'hydrolyse de leurs triglycérides et d'échanges de lipides et d'apolipoprotéines avec les autres lipoprotéines, elles sont transformées progressivement en IDL puis en LDL. Au niveau hépatique, les LDL sont captées par endocytose via le récepteur aux LDL qui reconnaît les apolipoprotéines B et E, permettant ainsi un recyclage de leurs constituants. D'autres récepteurs appartenant à la famille du récepteur aux LDL comme LRP (LDL receptor-related protein) participent à ces phénomènes d'endocytose. Au niveau des tissus périphériques, l'endocytose des VLDL, IDL et LDL fait également intervenir le récepteur aux LDL, le LRP ainsi que d'autres récepteurs (récepteur aux VLDL, récepteur à l'apolipoprotéine E, mégaline). Enfin, les HDL sont responsables du retour au foie des lipides d'origine extra-hépatique et notamment du cholestérol. Les HDL naissantes (Pré $\beta$ HDL de forme discoïdale et constituées essentiellement d'apoA-I associée à quelques phospholipides), produites par le foie et l'intestin, captent les lipides (principalement du cholestérol et des phospholipides) des tissus périphériques, notamment via les ABC-transporteurs ABCA1, ABCG1 et ABCG4. Puis, sous l'action de la LCAT (lecithine:cholesterol acyltransférase), le cholestérol est estérifié, modifiant ainsi la forme des HDL qui deviennent plus sphériques en raison de la migration en leur centre de lipides apolaires. Les HDL, qui subissent également des remodelages sous l'action de la PLTP (phospholipid transfer protein) et de la CETP (cholesteryl ester transfer protein qui n'est pas exprimée chez la souris), retournent ensuite au foie (une partie des HDL est également captée au niveau rénal via les récepteurs mégaline/gp330/LRP2 et cubilin, Hammad et al., 2000). L'absence d'expression de la CETP chez la souris, conduit les HDL à assurer la majeure partie du transport du cholestérol dans l'organisme alors que chez l'homme, ce sont les LDL (voir [Table 7](#)). La lipase hépatique (LH) hydrolyse les TG issus des HDL et semble faciliter également le captage sélectif des esters de cholestérol issus des HDL par le récepteur scavenger SR-BI (scavenger receptor class B type I) par les cellules hépatiques. La sous unité  $\beta$  de l'ATP synthase a également été identifiée comme un récepteur aux HDL qui participerait à leur endocytose par les hépatocytes (Martinez et al., 2003) et permettrait vraisemblablement le recyclage de leurs constituants. Cette voie de retour du cholestérol au foie par les HDL est connue sous le nom de **transport reverse du cholestérol** ou **voie de retour** et constitue un bénéfice pour l'organisme vis-à-vis des pathologies cardiovasculaires puisque le taux de cholestérol-HDL (le « bon » cholestérol) est corrélé négativement avec l'incidence des maladies cardiovasculaires, alors que le taux plasmatique de cholestérol-LDL (le « mauvais »

**Table 8** : Panel d'éléments de réponse fonctionnels des récepteurs PPAR $\alpha$  et Rev-erb $\alpha$  (début)

RN	Gène cible	Gene ID	Element de réponse	Remarque	Références
PPAR $\alpha$	Lipoprotéine lipase	16956	AAGAA <u>ACGGGAAAGGGCAAGGA</u> -164                                    -184	conservé chez l'homme	(Schoonjans et al., 1996a)
	Apolipoprotéine C-III	11814	GAGCTGGCAAAGGTCAOCTG -71                                    -91	conservé chez l'homme, pas d'activation via PPAR $\alpha$ mais compétition avec HNF-4	(Hertz et al., 1995)
	Human apoAV	116519	GGGA <u>AGGTAAAGGTCAATGGG</u> -275                                    -255	Un autre PPRE est également présent dans le promoteur	(Prieur et al., 2003; Vu-Dac et al., 2003)
	Human Rev-erb $\alpha$	9572	AAAAGT <u>GTTCACTGGGCAOGAG</u> -616                                    -623	Rev-DR2, bien conservé chez la souris	(Gervois et al., 1999a)
	Human apoA-I	335	CTGCC <u>AGGGCAAGGGTCAAAGGGTCAGTGG</u> -420                                    -448	DR1+DR2, non conservé chez le rat	(Vu-Dac et al., 1998)
	Human apoA-II	336	CTACC <u>AGGGTAACAGTTCAAGGCAACG</u> -946                                    -967	DR1. Régulation conservée chez la souris	(Vu-Dac et al., 1995)
	FATP-1 / Slc27a1	26457	GAGGAAG <u>TGGGCAAAAGGGCAACCG</u> -2291                                    -2267	DR1	(Frohnert et al., 1999)
	Human FAT / CD36	948	AAGTA <u>AGTCAGGGCAGGGCAGAAA</u> -8356                                    -8376	DR1 bien conservé chez la souris	(Tontonoz et al., 1998)
	Rat ACSL-1	252288	TGCCA <u>AGGTCTTCAGGGCATCAGTCACATG</u> -1766                                    -1796	DR1+DR3 assez bien conservé chez la souris et l'homme	(Juge-Aubry et al., 1997; Schoonjans et al., 1995)
	Fabp1 / L-FABP	14080	ATATA <u>GGGCATAGGTCAAGTGA</u> -99                                    -119	DR1 identique chez le rat mais moins conservé chez l'homme	(Issemann et al., 1992; Schachtrup et al., 2004)
ACBP		13167	CACT <u>GGCACAGGGTCAAOCC</u> +517                                    +537	DR1 conservé chez le rat et l'homme	(Helledie et al., 2002)
M-CPTI / Cpt1b		12895	ATGTA <u>GGGAAAGGTCAACCAAG</u> -765                                    -785	DR1 conservé chez l'homme	(Brandt et al., 1998; Mascaro et al., 1998)
Human L-CPT1/Cpt1a		600528	AAAAAGT <u>AGGGAAAGGTCAACCAAG</u> -23919                                    -23942	DR1 bien conservé chez le rat et la souris (la base soulignée est un T chez ces espèces)	(Napal et al., 2005)
Human MCAD		34	TCT <u>GGGTAAACGTGAAGGC</u> -523                                    -503	DR1	(Gulick et al., 1994)
HMG-CoA synthase mitochondriale		15360	AACT <u>GGGCCAAAGGTCTCAGA</u> -138                                    -158	DR1	(Rodriguez et al., 1994)

RN	Gène cible	Gene ID	Element de réponse	Remarque	Références
<b>PPAR<math>\alpha</math></b>	rat Acox 1	50681	<u>GACCAGGACAAAGGTCA</u> <b>CAGCT</b> -138	DR1 (la base soulignée est un A chez la souris)	(Osumi et al., 1991; Tugwood et al., 1992)
L-PBE / Ehhadh		74147	<u>ATGTAGGTAA</u> <b>TAGTT</b> <u>CAGTAG</u> -2829	DR1 (identique chez le rat)	(Bardot et al., 1993; Marcus et al., 1993; Zhang et al., 1993; Zhang et al., 1992)
Cyp8b1		13124	<u>AGCAAAAGTOCA</u> <b>AAGGCCAGGAA</b> -110	DR1 identique chez le rat	(Hunt et al., 2000b)
Malic enzyme		17436	<u>TTC</u> <b>TGGTCA</b> <u>AAGTTGA</u> <b>TOC</b> -383	DR1 identique chez le rat	(Castelein et al., 1994; Ippenbergs et al., 1997)
SCD1		20249	<u>GGGTAGATGTGAAGTTAGA</u> <b>CC</b> -819	DR1 identique chez le rat	(Miller and Nambi, 1996)
Fads2		56473	<u>GGAGGGAGGAGGAGGTCA</u> <b>GACA</b> -590	DR1 bien conservé et fonctionnel chez l'homme (fonctionnalité chez la souris non testée)	(Tang et al., 2003)
Bmal1 / Arntl		11865	<u>GGGCTGGACATGGT</u> <b>TCATGGCC</b> -1487	DR1	(Canaple et al., 2006)
<b>Rev-erba</b> $\alpha$ $\beta$	Human Rev-erba	9572	<u>GGAAAAGGTGTCA</u> <b>ACTGGGCA</b> <u>CG</u> -648	Principalement monomère Rev-erba + peu d'homodimère Rev-erba sur RevDR2	(Adelmant et al., 1996)
Human apoC-III		345	<u>GATATAAAACAGGTCA</u> <b>A</b> <u>GAACCTA</u> -28	Monomère Rev-erba (et ROR $\alpha$ 1 et RVR). Séquence -28/-18 conservée chez la souris	(Coste and Rodriguez, 2002)
Rat apoA-I		25081	<u>ACATATATAGGTCA</u> <b>GGGAAGA</b> -39	Monomère Rev-erba. Non conservé chez l'homme et partiellement chez la souris	(Vu-Dac et al., 1998)
Mouse Bmali1		11865	<u>GAAGGCAGAAAAGTAGGTCA</u> <b>AGGGCA</b> <u>CG</u> +65	Monomère Rev-erba. Séquence conservée chez le rat et l'homme.	(Pfeitner et al., 2002)

**Table 8 :** Panel d'éléments de réponse fonctionnels des récepteurs PPAR $\alpha$  et Rev-erba (fin)

Le récepteur PPAR $\alpha$  régulant l'expression de Rev-erba, les éléments de réponse de quelques gènes cibles de ce récepteur sont également présentés. Dans la mesure du possible, le site d'initiation de la traduction a été choisi pour établir la position des séquences sur les promoteurs. Parfois, c'est cependant le site d'initiation de la transcription qui sert de point de départ. Les éléments de réponse sont tous présents dans l'orientation et à partir du brin qui permet de comparer au mieux la séquence fournie dans le tableau au motif consensus AGGTCA. La numérotation des bases flanquantes permet de visualiser l'orientation de l'élément de réponse par rapport au site d'initiation de la transcription et/ou de la traduction. En l'absence de précision, c'est la séquence murine qui est présentée. Parfois, certaines bases sont soulignées pour mettre en évidence des différences de séquences entre l'homme et les rongeurs.

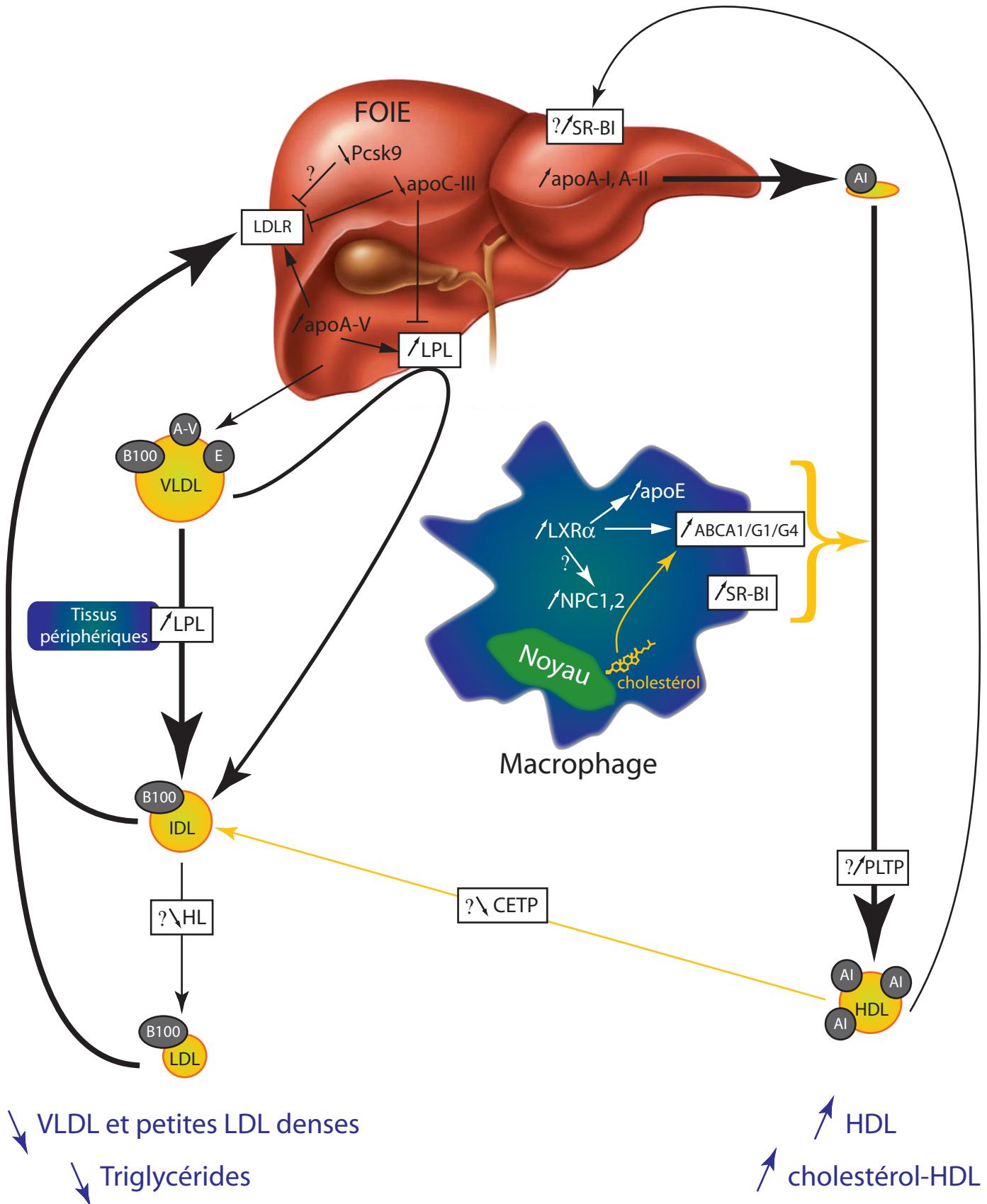
cholestérol) est corrélé positivement avec l'incidence de ces pathologies. En réalité, ce sont principalement les LDL de petites tailles, plus sujettes à l'oxydation et à leur captation par les macrophages via divers récepteurs (SR-A pour scavenger receptor class A, FAT/CD36 pour fatty acid transporter, OLR1 pour oxidized LDL receptor 1, etc.), qui constituent le principal risque au niveau vasculaire. Les macrophages gorgés de lipides (cellules spumeuses) pénètrent dans l'endothélium vasculaire, surtout en situation d'inflammation, et initient ainsi la formation de la plaque d'athérome. Signalons également qu'outre leur rôle bénéfique dans l'élimination du cholestérol excédentaire, les HDL présentent également un pouvoir antioxydant via leur activité paraoxonase et leur apolipoprotéine J.

Les bénéfices thérapeutiques des fibrates s'expliquant, au moins en partie, par leurs capacités à réduire la production de VLDL, à accroître le catabolisme des particules riches en TG ce qui diminue indirectement les petites LDL denses et à augmenter la formation des HDL et l'élimination hépatique du cholestérol en excès, de nombreuses recherches ont été conduites pour mieux comprendre les rôles de PPAR $\alpha$  dans la régulation du métabolisme des lipoprotéines. Chez l'homme comme chez la souris, PPAR $\alpha$  se lie à un PPRE (voir [Table 8](#) pour une présentation de différents PPRE) présent dans le gène de la lipoprotéine lipase (LPL) et active sa transcription hépatique (Schoonjans et al., 1996a) ce qui augmente l'hydrolyse des TG issus des VLDL et des chylomicrons. L'apolipoprotéine C-III est un inhibiteur du catabolisme des Lp-B via la LPL et les récepteurs aux LDL (Frenkel et al., 1994). Les fibrates inhibent l'expression hépatique de l'apoC-III chez l'homme et chez les rongeurs (Staels et al., 1995) de manière dépendante de PPAR $\alpha$  car cet effet n'est pas observé chez les souris PPAR $\alpha$ -/- (Peters et al., 1997). L'inhibition de la transcription de l'apoC-III semble liée en partie à une compétition exercée par PPAR $\alpha$  sur la fixation de HNF-4 (hepatic nuclear factor 4) à un DR-1 situé dans le promoteur de l'apoC-III (Hertz et al., 1995). Par ailleurs, chez l'homme et la souris les fibrates augmentent l'expression hépatique de Rev-erb $\alpha$ , un RN qui réprime l'expression de ses gènes cibles, via la fixation de PPAR $\alpha$  sur un élément de réponse de type DR2 (Gervois et al., 1999a; Vu-Dac et al., 1998), réprimé d'ailleurs par Rev-erb $\alpha$  lui-même (voir [Table 8](#), Adelman et al., 1996). Or, il a été montré que Rev-erb $\alpha$  régulait négativement l'expression de l'apoC-III (Coste and Rodriguez, 2002; Raspe et al., 2002) via sa fixation sur un élément de réponse qui est activé par ROR $\alpha$ 1 (Raspe et al., 2001). Il est donc très probable que cette régulation additionnelle, dépendante également de PPAR $\alpha$ , soit impliquée dans la répression de l'apoC-III par les fibrates. Cette inhibition de l'apoC-III conduirait notamment à favoriser la lipolyse des VLDL et à générer de larges particules de

LDL qui sont mieux captées par le récepteur aux LDL. L'apolipoprotéine A-V a été récemment découverte et a suscité de nombreuses recherches depuis que des corrélations négatives ont été identifiées chez la souris entre son niveau d'expression et les taux circulants de TG et que des associations ont été identifiées chez l'homme entre certains de ses polymorphismes ponctuels (SNPs pour single nucleotide polymorphisms) et le taux de TG circulants. L'apoAV est synthétisée par le foie, associée avec les VLDL et les HDL ([Table 7](#)) et favoriserait l'action de la lipoprotéine lipase et l'endocytose via le récepteur aux LDL conduisant à une diminution des taux circulants de TG. Chez l'homme, l'apoAV est induite au niveau hépatique par les fibrates via la fixation de PPAR $\alpha$  sur un PPRE (Prieur et al., 2003; Vu-Dac et al., 2003), ce qui participerait vraisemblablement fortement à la diminution des TG plasmatiques provoquée par ces médicaments. Enfin, il a été récemment montré que la PCSK9 (proprotein convertase subtilisin kexin type 9), qui inhibe l'expression du récepteur aux LDL au niveau hépatique probablement en augmentant sa dégradation et promeut ainsi l'hypercholestérolémie et l'hypertriglycéridémie, était régulée négativement par le fenofibrate dans le foie de souris de type sauvage mais pas dans le foie de souris PPAR $\alpha$ -/- (Lambert et al., 2006).

Outre ses rôles dans la diminution des lipoprotéines riches en triglycérides, PPAR $\alpha$  influe également positivement sur plusieurs aspects du métabolisme des HDL et du transport reverse du cholestérol. Chez l'homme, les fibrates régulent la production des HDL via une induction PPAR $\alpha$ -dépendante des apoA-I (Vu-Dac et al., 1998) et apoA-II (Vu-Dac et al., 1995). En revanche, chez le rat et la souris, les fibrates répriment l'expression de l'apoA-I en raison d'un PPRE non fonctionnel et de la présence d'un élément de réponse à Rev-erba (Staels and Auwerx, 1998; Vu-Dac et al., 1998), un RN inhibiteur de l'expression génique qui est, comme nous l'avons précisé précédemment, induit par les fibrates via PPAR $\alpha$  (Gervois et al., 1999a). Le rôle de PPAR $\alpha$  dans la régulation de l'apoA-II et de l'apoA-I a été confirmé chez les rongeurs où les effets des fibrates sur l'expression de ces apolipoprotéines ne sont plus observés chez les souris PPAR $\alpha$ -/- (Peters et al., 1997). De plus, la régulation de ces apolipoprotéines est également possible via l'activation de RXR (Vu-Dac et al., 1996), soulignant la convergence entre les voies de signalisation de PPAR $\alpha$  et de RXR. Les agonistes de PPAR $\alpha$  et de PPAR $\gamma$  sont également capables de promouvoir l'efflux de cholestérol par les macrophages (Chinetti et al., 2001), ce qui pourrait jouer un rôle dans leur capacité à réduire l'athérosclérose en inhibant la formation de cellules spumeuses dans la paroi vasculaire (Chinetti et al., 2001; Li et al., 2004).

Ces effets pourraient être dus à une induction de l'expression de LXR $\alpha$  dans les macrophages. Le promoteur murin de LXR $\alpha$  répond à une activation de PPAR $\alpha$  par des acides gras dans le foie (Tobin et al., 2000) et plusieurs PPRE ont été identifiés mais ils ne sont pas conservés chez l'homme (Laffitte et al., 2001a). Un autre PPRE est néanmoins présent dans le promoteur de LXR $\alpha$  chez l'homme et, chez l'homme comme chez la souris, LXR $\alpha$  est induit par PPAR $\gamma$  dans les macrophages et les adipocytes (Chawla et al., 2001; Laffitte et al., 2001a). De plus, LXR $\alpha$  semble autoréguler sa propre expression dans plusieurs tissus (macrophages, foie, tissu adipeux) chez l'homme uniquement (Laffitte et al., 2001a) ce qui peut expliquer son induction dans les macrophages chargés de cholestérol dont les dérivées peuvent activer LXR. Bien que les mécanismes impliquant LXR dans la réduction de l'athérosclérose ne soit pas complètement élucidés, sa capacité à induire l'expression de ABCA1 (Costet et al., 2000; Venkateswaran et al., 2000), ABCG1 (Sabol et al., 2005; Venkateswaran et al., 2000; Wang et al., 2004), ABCG4 (Engel et al., 2001; Wang et al., 2004) et de l'apoE (Laffitte et al., 2001b) dans les macrophages et à promouvoir ainsi l'efflux de cholestérol vers les HDL est actuellement considéré comme un mécanisme central de son action bénéfique au niveau vasculaire (voir [Table 9](#) pour quelques éléments de réponse naturels de LXR). En plus de ses actions vraisemblables via LXR $\alpha$ , les activateurs de PPAR $\alpha$  augmentent in vitro et in vivo l'expression du récepteur SR-BI dans les monocytes et les macrophages (Chinetti et al., 2000) vraisemblablement via un PPRE fonctionnel qui a été identifié chez le rat dans le promoteur de SR-BI (Lopez and McLean, 2006). En dehors de son rôle dans le captage sélectif des esters de cholestérol issus des HDL au niveau du foie et des organes assurant la stéroidogénèse, SR-BI pourrait également participer à l'efflux du cholestérol vers les HDL dans les macrophages gorgés de lipides (Chinetti et al., 2000). Plus récemment, il a également été proposé que l'effet bénéfique des fibrates sur l'efflux de cholestérol des macrophages pouvait être lié à la régulation du trafic intracellulaire du cholestérol (Chinetti et al., 2006). L'expression des gènes NPC1 (Niemann Pick type C 1) et NPC2 qui sont impliqués dans ce processus est modérément induite par les activateurs de PPAR $\alpha$  (Chinetti-Gbaguidi et al., 2005) et de LXR (Rigamonti et al., 2005) dans des macrophages humains mais, à ma connaissance, aucun élément de réponse n'a encore été rapporté. Enfin, les fibrates régulent également, via PPAR $\alpha$ , le remodelage des HDL par la PLTP. L'action de la PLTP consiste à faciliter l'échange de phospholipides entre lipoprotéines, conduisant à l'apparition à la fois de particules HDL de grande taille ( $\alpha$ HDL larges) et de petites particules HDL pauvres en lipides (pré $\beta$ HDL), ces dernières constituant



**Figure 24 : Principaux effets de PPAR $\alpha$  sur le métabolisme des lipoprotéines**

Dans un soucis de simplification, la voie entéro-hépatique (chylomicrons) n'a pas été représentée. Bien que représentés ici dans le foie, il est évident que les rôles des apoC-III et A-V s'exercent dans le cadre de l'interaction entre les lipoprotéines et les récepteurs ou enzymes localisés à la surface des cellules. Les ? indiquent des modulations ou des mécanismes qui ne sont pas encore entièrement démontrés (voir le texte correspondant). Les principaux effets de l'administration de fibrates sur le profil des lipoprotéines plasmatiques sont indiqués en bleu en bas du schéma.

les premiers accepteurs de l'efflux du cholestérol cellulaire. Cependant, malgré son rôle dans le maintien des HDL et la génération de pré $\beta$ HDL, la PLTP est généralement considérée comme pro-athérogène ce qui pourrait être lié à son rôle éventuel dans la sécrétion hépatique de Lp-B (Lewis, 2006). Chez la souris, l'activité et l'expression hépatique de la PLTP sont augmentées par les fibrates mais pas chez les souris PPAR $\alpha$ -/- (Bouly et al., 2001). Des séquences homologues à des PPRE ont été identifiées dans les promoteurs de la PLTP murine (Tu and Albers, 2001) et humaine (Tu and Albers, 1999). Des données contradictoires ont été obtenues chez l'homme, montrant une nette augmentation de l'activité PLTP par le fenofibrate chez des patients présentant un syndrome métabolique en l'absence de diabète (Watts et al., 2006) ou une légère diminution chez des patients hypertriglycéridémiques traités par du bezafibrate (Jonkers et al., 2003). Des études complémentaires apparaissent nécessaires pour préciser le rôle de PPAR $\alpha$  dans la régulation de la PLTP mais peut-être aussi des autres enzymes intervenant dans le remodelage des lipoprotéines comme la CETP, la LCAT ou la lipase hépatique (LH). Les quelques données disponibles indiquent que chez le rat, les fibrates régulent négativement l'expression de la LCAT (Staels et al., 1992b) et de la LH (Staels et al., 1992a). Une légère réduction de l'activité CETP a également été observée chez des patients hypercholestérolémiques (Franceschini et al., 1995; Jonkers et al., 2003) ou présentant un syndrome métabolique (Watts et al., 2006) suivant un traitement par un fibrat. La [Figure 24](#) tente de résumer les principales régulations exercées par PPAR $\alpha$  sur le métabolisme des lipoprotéines. Par ailleurs, le rôle éventuel de PPAR $\alpha$  dans l'absorption intestinale des lipides a été très peu étudié mais quelques données (Bunger et al., 2007; Knight et al., 2003), ainsi que son expression appréciable au niveau intestinal et son induction par des lipides alimentaires (Mochizuki et al., 2001), suggèrent qu'il ne serait pas négligeable.

#### **II.4.2. Régulation du transport et du trafic intracellulaire des acides gras par PPAR $\alpha$**

La fonction la mieux appréciée de PPAR $\alpha$  à l'heure actuelle est sans doute son rôle dans la régulation du catabolisme des acides gras. La première étape requise pour l'oxydation des acides gras est leur transport à travers la membrane plasmique qui est augmenté au niveau hépatique par les fibrates de manière concomitante à l'accroissement du catabolisme des lipoprotéines riches en TG (Frenkel et al., 1994).

Gène cible	Gene ID	Element de réponse	Remarque	Référence
Mouse ABCA1	11303	CGAGGGCAG <u>AGGTAA</u> CTAT <u>GGTCA</u> AAGCCAGT -55                            -70	LXR $\alpha$ ou $\beta$ /RXR sur un DR4	(Costet et al., 2003; Costet et al., 2000)
Mouse LPL	16956	AGGG <u>AGGTCA</u> OCAC <u>GGTC</u> ACTAC +650                            +635	LXR $\alpha$ /RXR sur un DR4	(Zhang et al., 2001)
Mouse SREBP-1c	20787	mLXREa : CGCT <u>GGGTAA</u> CTGG <u>GGTCA</u> ACTGTC -214                            -239 mLXREb : GGG <u>GGGTAA</u> CTAG <u>GGGG</u> GTC -163                            -188	LXR $\alpha$ ou $\beta$ /RXR sur deux DR4 partiellement conservés chez l'homme	(Repa et al., 2000; Yoshikawa et al., 2001)
Human FAS	2194	Human : GGG <u>GGGTAA</u> CTGG <u>GGTCA</u> TGTC -738                            -761 Mouse : GGG <u>GGGTAA</u> CTAC <u>GGGTCA</u> TGTC -510                            -533	LXR $\alpha$ /RXR sur un DR4 conservé chez les rongeurs et le poulet	(Joseph et al., 2002)
Mouse LRH-1	26424	AGACAT <u>GGTTAACAGCAGGTCA</u> TAC -79                            -55	LXR $\alpha$ /RXR sur un DR4 reconnu également par TR/RXR	(Pare et al., 2004)

**Table 9 :** Exemples d'éléments de réponse naturels du récepteur LXR.

Dans la mesure du possible, le site d'initiation de la traduction a été choisi pour établir la position des séquences sur les promoteurs. Parfois, c'est cependant le site d'initiation de la transcription qui sert de point de départ. Les éléments de réponse sont tous présentés dans l'orientation et à partir du brin qui permet de comparer au mieux la séquence fournie dans le tableau au motif consensus AGGTCA. La numérotation des bases flanquantes permet de visualiser l'orientation de l'élément de réponse par rapport au site d'initiation de la transcription et/ou de la traduction.

Le transport des acides gras à longue chaîne est facilité par des transporteurs membranaires : **FATP** (fatty acid transport protein) et **FAT/CD36** (fatty acid translocase), ainsi que par l'activité acyl-CoA synthétase (**ACS**, EC 6.2.1.3) qui estérifie les acides gras libres en dérivés acyl-CoA. Cette activation des acides gras libres prévient leur efflux et permet leur métabolisme ultérieur (catabolisme ou synthèse des lipides cellulaires, Martin et al., 1997). Le gène FATP-1/Slc27a1 (solute carrier family 27 member 1), impliqué dans le transport des acides gras à longue chaîne et des LDL oxydées et possédant une activité acyl-CoA synthétase (Coe et al., 1999) est un gène cible direct de PPAR $\alpha$ , induit dans le foie et l'intestin par les fibrates ou les agonistes de RXR (Martin et al., 2000), et possède un PPRE dans son promoteur (Frohnert et al., 1999; Martin et al., 1997). Le gène FATP-2/Slc27a2, initialement identifié comme une acyl-CoA synthétase spécifique des acides gras à longue chaîne (VLACS pour very long chain fatty acyl-CoA synthétase) située dans la membrane des peroxysomes et des microsomes (Berger et al., 1998), est également induit chez la souris de type sauvage mais pas chez la souris PPAR $\alpha$ -/- suite à l'administration d'un proliférateur de peroxysomes (Aoyama et al., 1998). Le gène FAT/CD36 est induit de manière PPAR $\alpha$ -dépendante dans le foie, l'intestin et les macrophages par les proliférateurs de peroxysomes, dans le tissu adipeux et les macrophages par les agonistes de PPAR $\gamma$  et dans le muscle par le jeûne et les activateurs de PPAR $\beta$  (Holst et al., 2003; Jedidi et al., 2006; Motojima et al., 1998; Sato et al., 2002; Yu et al., 2001). Il semble donc que FAT/CD36 réponde bien et de manière peu différenciée aux trois PPARs comme cela a été démontré récemment dans des lignées de cellules NIH-3T3 exprimant chacune l'un des trois PPARs (Humasti and Tontonoz, 2006). Il a été montré que le gène CD36 possédait plusieurs promoteurs différents, utilisés de manière différentielle en fonction des tissus et répondant de manière différentielle aux agonistes des PPARs (Sato et al., 2002; Sato et al., 2007), cependant aucun PPRE fonctionnel n'a été rapporté par cette équipe. En revanche, un PPRE fonctionnel vis-à-vis de PPAR $\alpha$  (Jedidi et al., 2006) et de PPAR $\gamma$  (Tontonoz et al., 1998) et bien conservé chez la souris a été identifié dans le promoteur proximal de FAT/CD36 (présenté en [Table 8](#)). L'étape d'activation des acides gras pour leur métabolisme, qui s'opère à de multiples localisations intracellulaires, est sous le contrôle des long chain fatty acyl-CoA synthétases (Acsl). L'**Acsl1** est régulée par les fibrates dans le foie et le rein (Martin et al., 1997) via un PPRE situé dans son promoteur (Schoonjans et al., 1995). Bien qu'elle ne soit, à ma connaissance, pas directement impliquée dans le captage et la rétention des acides gras dans la cellule, il me

semble important de mentionner que la réaction inverse de celle catalysée par l'activité ACS est également régulée par PPAR $\alpha$ . Cette réaction inverse qui consiste à hydrolyser des acyl-CoA en acide gras et coenzyme A non estérifiés est catalysée par une famille de 12 enzymes connues sous le nom d'acyl-CoA thioestérases (Acot1-12, Hunt et al., 2005), dont la plupart des membres sont induits par le jeûne ou les proliférateurs de peroxyosomes, le plus souvent de manière PPAR $\alpha$ -dépendante (Aoyama et al., 1998; Hunt et al., 2000a; Hunt et al., 1999; Hunt et al., 2002; Lindquist et al., 1998; Svensson et al., 1995a; Svensson et al., 1998; Svensson et al., 1995b; Westin et al., 2004). Récemment, un PPRE fixant également HNF-4 a été identifié dans le promoteur de l'**Acot1** (Dongol et al., 2007) et un PPRE potentiel a été identifié dans le promoteur de l'**Acot8** (Hunt et al., 2002). L'**Acot6** est également régulée de manière PPAR $\alpha$ -dépendante par les fibrates et interviendrait dans les peroxyosomes sur le métabolisme des dérivés du phytol (Westin et al., 2007). Les diverses fonctions de ces enzymes sont encore mal connues (Maeda et al., 2006) mais elles pourraient être impliquées dans de multiples aspects du métabolisme des lipides comme la synthèse des acides biliaires, des stéroïdes ou des éicosanoïdes, l'élimination sous forme libre de nombreux éicosanoïdes, l'échange d'acides gras entre les peroxyosomes et les mitochondries dans le cadre de leur oxydation, la fourniture de substrats pour l' $\omega$ -hydroxylation où encore le contrôle de la disponibilité de coenzyme A non estérifié pour la dernière étape de la  $\beta$ -oxydation (voir [Figure 20](#)) via l'hydrolyse d'acides gras qui constituent de mauvais substrats des enzymes de la  $\beta$ -oxydation et bloquent ainsi la chaîne de réaction en séquestrant le coenzyme A sous forme estérifiée (pour une synthèse, voir Hunt and Alexson, 2002).

Outre le captage des acides gras, PPAR $\alpha$  régule également le trafic intracellulaire des acides gras via sa capacité d'induction des protéines de fixation des acides gras (**FABP** pour fatty acid binding protein) ou des acyl-CoA (**ACBP** pour acyl-CoA binding protein). Plusieurs FABP ont été identifiées et nommées sur la base de leur expression plus ou moins tissu-spécifique dans le foie (Fabp1 ou L-FABP), dans l'intestin (Fabp2), dans le cœur et le muscle (Fabp3), dans le tissu adipeux (Fabp4 ou encore ALBP ou aP2, une cible de PPAR $\gamma$  que nous avons déjà évoquée), dans l'épiderme (Fabp5), dans l'iléon (Fabp6 ou i-BABP pour iléum bile acid binding protein), dans le cerveau (Fabp7 et Fabp8 ou PMP2 pour peripheral myelin protein 2) ou dans le testicule (Fabp9). Ces transporteurs intracellulaires semblent présenter une large diversité de substrat tels des acides gras, des acyl-CoA, du cholestérol, des phospholipides mais également des proliférateurs de peroxyosomes. Les études sur Fabp1 démontrent qu'en situation de jeûne (et peut-être plus généralement en situation de forte

demande vis-à-vis de l’oxydation hépatique des acides gras), cette protéine semble jouer un rôle important dans le captage, l’oxydation et l’incorporation dans les TG des acides gras à longue chaîne (Antonenkov et al., 2006; Erol et al., 2004; Martin et al., 2003; Newberry et al., 2003). De plus, comme nous l’avons évoqué au paragraphe I.4.3., en situation d’alimentation, les FABP joueraient vraisemblablement un rôle dans l’accessibilité de PPAR $\alpha$  à ses ligands en transportant ces ligands vers le noyau et en interagissant directement avec PPAR $\alpha$  (Tan et al., 2002; Wolfrum et al., 2001). Parmi les Fabp, les principales cibles des PPARs semblent être Fabp4 pour PPAR $\gamma$  et Fabp1 pour PPAR $\alpha$  (Schachtrup et al., 2004). Un PPRE a été identifié dans le promoteur d’ap2 qui semble favoriser une activation dans le tissu adipeux par les dimères PPAR $\gamma$ -RXR (Schachtrup et al., 2004; Tontonoz et al., 1994b) et un PPRE qui semble favoriser une activation dans le foie par les dimères PPAR $\alpha$ -RXR est présent dans le promoteur de Fabp1 (Issemann et al., 1992; Kaikaus et al., 1993; Schachtrup et al., 2004; Simon et al., 1993). L’expression de Fabp1 est activée dans le foie de manière PPAR $\alpha$ -dépendante par les proliférateurs de peroxysomes et les acides gras à longue chaîne mais, dans les mêmes conditions, elle semble activée de manière indépendante de PPAR $\alpha$  (vraisemblablement via PPAR $\beta$ ) dans l’intestin (Lee et al., 1995; Nakagawa et al., 1994; Poirier et al., 2001). Dans le foie, les statines, des inhibiteurs de l’HMG-CoA réductase (enzyme limitante de la synthèse endogène de cholestérol) largement utilisés dans le traitement des hypercholestérolémies, semblent capable d’induire l’expression de PPAR $\alpha$ , conduisant, en co-administration avec un fibrate, à une activation synergique de l’expression de la L-FABP de manière dépendante de PPAR $\alpha$  (Landrier et al., 2004). Des PPRES potentiels ont également été identifiés dans les promoteurs de Fabp3 et Fabp5 mais leur fonctionnalité n’a pas pu être démontrée dans des cellules HepG2 transfectées avec différentes combinaisons des PPARs et des RXRs (Schachtrup et al., 2004). L’acyl-CoA binding protein (ACBP) présente une moins grande diversité de substrat et semble spécifique des Acyl-CoA à chaîne moyenne et longue. Elle est exprimée dans la plupart des tissus mais particulièrement dans les hépatocytes, les adipocytes et les cellules stériodogéniques où elle intervient notamment pour accroître le pool d’acyl-CoA, pour transporter et fournir des acyl-CoA à la  $\beta$ -oxydation mitochondriale, à la synthèse microsomale de glycérolipides et à la synthèse de phospholipides et enfin pour lever les inhibitions par les acyl-CoA de diverses enzymes dont les acyl-CoA synthétases. L’expression d’ACBP est régulée à la fois par PPAR $\alpha$  (et PPAR $\gamma$ ) et par les SREBPs (Neess et al., 2006; Sandberg et al., 2005). Le résultat est qu’en situation de jeûne, ces deux types de signaux s’opposent (activation par PPAR $\alpha$  et

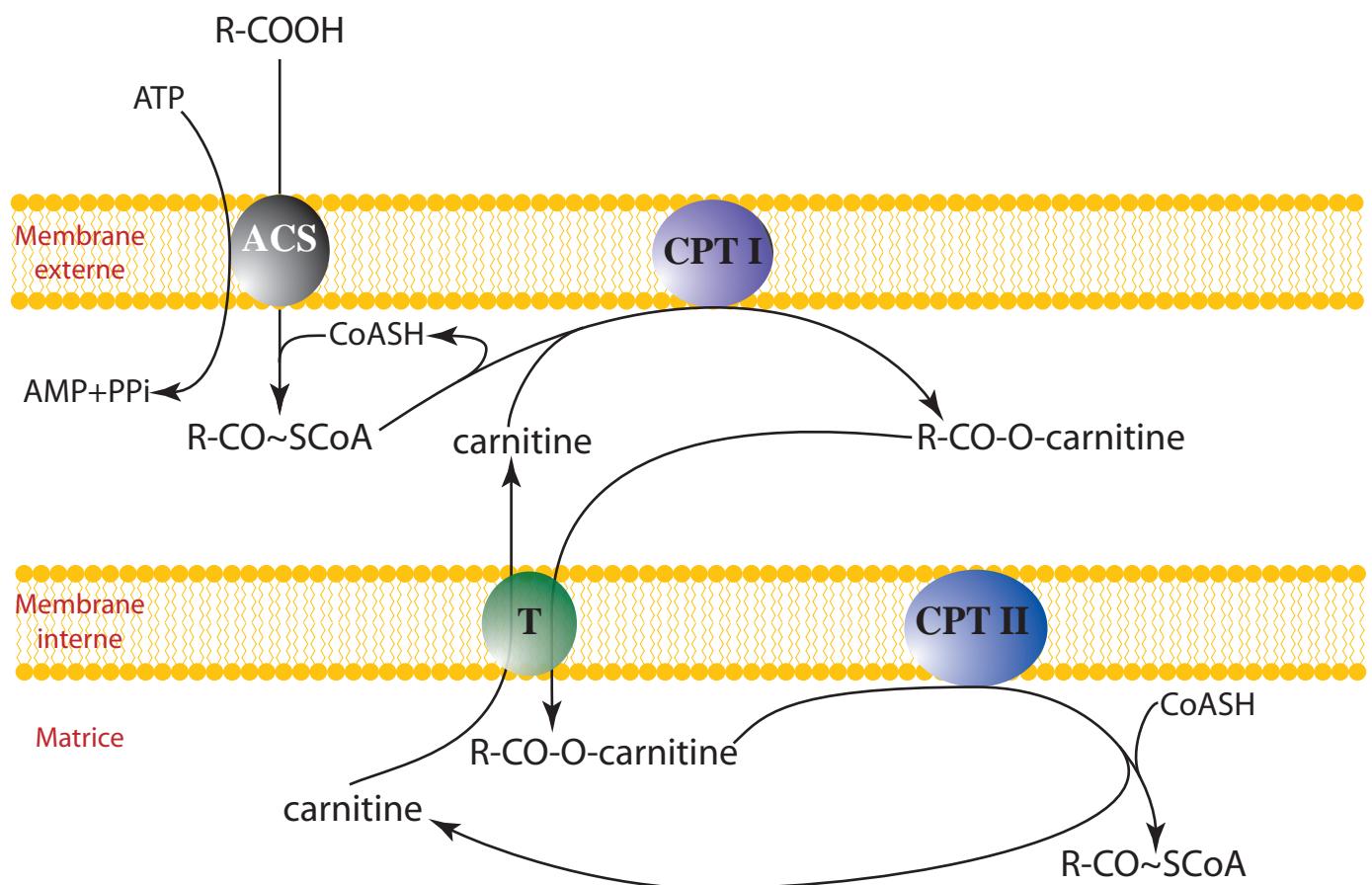
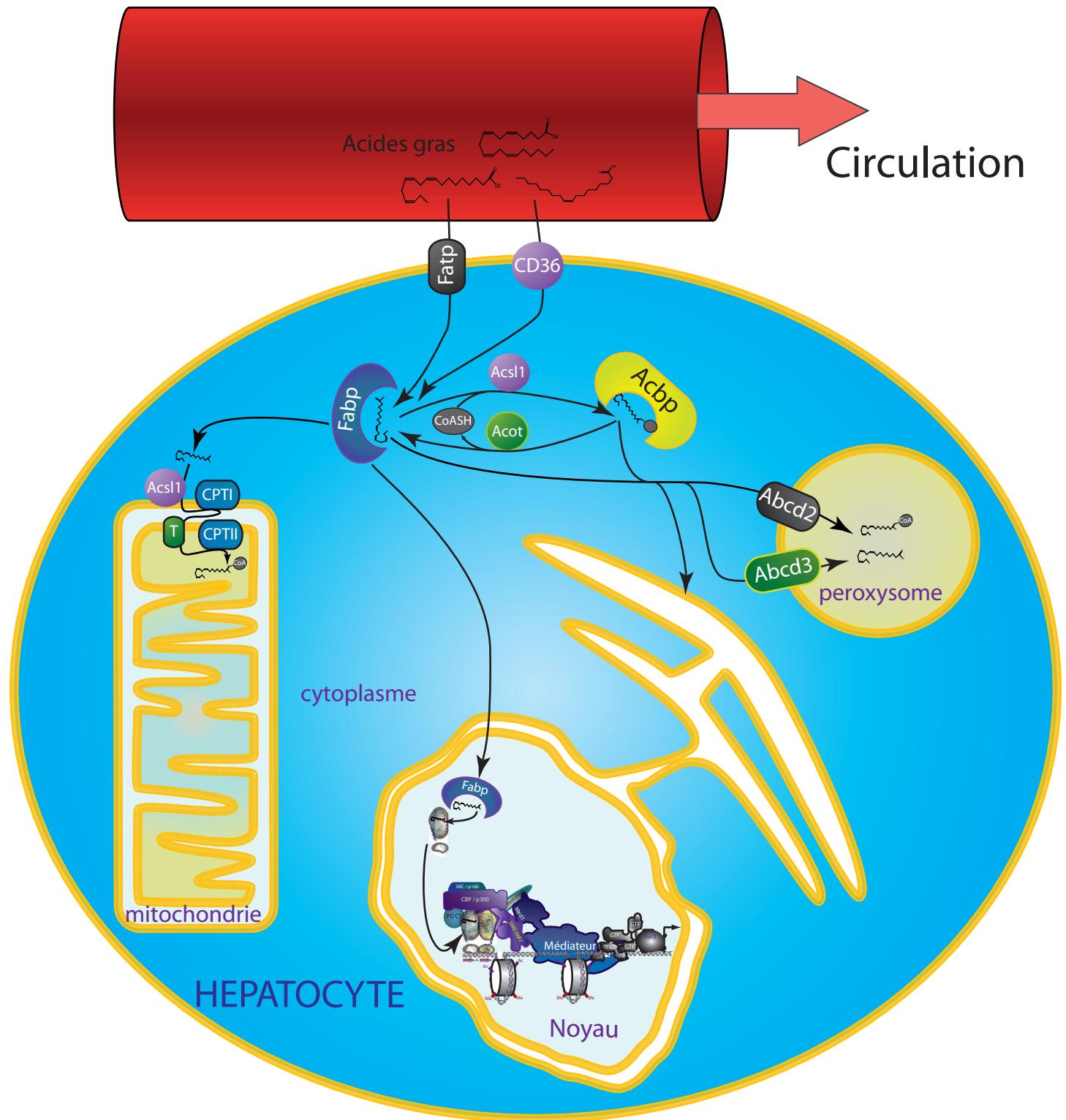


Figure 25 : Entrée dans la mitochondrie et activation des acides gras à longue chaîne  
 ACS : Acyl-CoA synthétase, CPT : carnitine palmitoyltransférase, T : translocase

diminution de l'activation par SREBP) et conduisent à une légère répression d'ACBP. En situation d'activation de la lipogénèse, comme par exemple après une réalimentation suivant un jeûne, l'expression d'ACBP augmente via son induction par SREBP. Les proliférateurs de peroxysomes activent, eux, nettement l'expression d'ACBP dans le foie de manière dépendante de PPAR $\alpha$  (Lee et al., 1995; Sandberg et al., 2005; Vanden Heuvel et al., 1993). Un PPRE fonctionnel, conservé chez les rongeurs et chez l'homme, a été identifié dans le premier intron de l'ACBP, alors qu'un autre PPRE, identifié dans le promoteur de l'ACBP ne semble pas fonctionnel (Elholm et al., 1996; Helledie et al., 2002).

Un traitement par les proliférateurs de peroxysomes se traduit également par un accroissement de l'influx d'acides gras à longue chaîne dans la mitochondrie (Pande and Parvin, 1980). Le passage des acides gras à longue chaîne dans la mitochondrie est considéré comme l'étape limitante de leur  $\beta$ -oxydation mitochondriale et de la cétogenèse. Il est assuré par un système comportant quatre enzyme (voir [Figure 25](#)) : 1) l'acyl-CoA synthétase (**ACS**) converti l'acide gras en thioester du coenzyme A qui franchit la membrane mitochondriale externe. 2) La carnitine palmitoyltransférase I (**CPT I**, EC 2.3.1.21) fixe le groupement acyl de l'acyl-CoA sur la carnitine ce qui est considéré comme l'étape limitante de ce système de transport. 3) L'acyl-carnitine franchit la membrane interne grâce à une carnitine:acylcarnitine **translocase**. 3) L'acyl-CoA est reformé sur la face interne de la membrane interne sous l'action de la carnitine palmitoyltransférase II (**CPT II**). Nous avons déjà évoqué la régulation de l'Acs11, présente dans les mitochondries, par PPAR $\alpha$ . Il existe deux gènes codant respectivement pour la L-CPTI (ou CPT1A) et la M-CPT1 (ou CPTIB). La L-CPTI (pour liver CPTI) est exprimée dans de nombreux tissus mais surtout dans le foie, le cœur et les cellules  $\beta$  du pancréas sous la forme de deux transcrits alternatifs différenciés par leur premier exon non traduit (exons 1a et 1b) alors que la M-CPTI (pour muscle CPTI) est exprimée dans le muscle squelettique, le cœur, le testicule et les tissus adipeux blanc et bruns également sous la forme de deux transcrits ne présentant pas le même premier exon (exons 1a et 1b). Un PPRE fonctionnel vis-à-vis du dimère PPAR $\alpha$ -RXR $\alpha$  a été identifié dans le promoteur de la M-CPTI humaine (Mascaro et al., 1998) qui est impliqué dans la régulation de l'expression de la M-CPT1 dans des cellules musculaires cardiaques humaines suite à leur exposition à l'acide oléique (Brandt et al., 1998). In vivo, l'expression constitutive de la M-CPTI est réduite d'environ 50% dans le cœur des souris PPAR $\alpha$ -/- et les activateurs de PPAR $\alpha$  n'induisent pas l'expression de la M-CPT1 dans le cœur de ces souris (Brandt et al., 1998). En situation de jeûne, l'expression hépatique de la L-CPTI semble augmentée de manière



**Figure 26 :** Principaux effets de PPAR $\alpha$  sur le captage et le trafic intracellulaire des acides gras  
 Par soucis de simplification, seules les protéines évoquées dans le paragraphe II.4.2 sont illustrées (a part dans le noyau). De même, les lipoprotéines n'ont pas été représentées même si la plupart des acides gras ne circulent pas sous forme libre mais proviennent majoritairement de l'hydrolyse des triglycérides par la lipoprotéine lipase et par la lipase hépatique.

indépendante de PPAR $\alpha$  (Leone et al., 1999). Par contre, des activateurs de PPAR $\alpha$  sont capables d'accroître nettement son expression dans des cultures primaires d'hépatocytes humains et dans des cellules HepG2 (Lawrence et al., 2001). Chez la souris, les fibrates induisent l'expression de la L-CPTI de manière PPAR $\alpha$ -dépendante dans des hépatocytes primaires en culture via un PPRE localisé 4090 pb en amont du site d'initiation de la transcription (Louet et al., 2001) qui est bien conservé chez le rat mais ne semble pas être présent chez l'homme (d'après les différents alignements que j'ai pu faire). Récemment, un autre PPRE fonctionnel a été identifié dans le premier intron (intron 1b) du gène humain de la L-CPTI qui est très bien conservé chez le rat et la souris et confère à un gène rapporteur une nette réponse aux activateurs de PPAR $\alpha$ , à PGC-1 et aux acides gras (Table 8, numérotation à partir du site d'initiation de la traduction situé dans l'exon 2, Napal et al., 2005). Le malonyl-CoA est considéré comme le principal inhibiteur physiologique de l'activité CPT-I. Sa production est assurée à partir de l'acétyl-CoA par les acétyl-CoA carboxylase (ACC1 et ACC2) dont l'expression est contrôlée par SREBP et qui présentent une expression hépatique légèrement réduite en phase nocturne chez les souris PPAR $\alpha$ -/- (Patel et al., 2001) sans pour autant que cela semble affecter leur activité constitutive dans le foie (Aoyama et al., 1998) comme dans le cœur (Campbell et al., 2002). Leur activité est légèrement augmentée (25-30%) par le Wy14643 mais de manière indépendante de PPAR $\alpha$  (Aoyama et al., 1998). A l'opposé, la dégradation du malonyl-CoA en acétyl-CoA est assurée par la **malonyl-CoA décarboxylase** (MCD) dont la régulation positive par PPAR $\alpha$  dans le cœur et le muscle a été identifiée (Campbell et al., 2002; Young et al., 2001a). Deux PPRE fonctionnels ont également été identifiés dans le promoteur de la MCD de rat (Lee et al., 2004a). Ainsi, en plus de réguler directement l'expression des CPTI, PPAR $\alpha$  augmenterait aussi leur activité en augmentant la dégradation du malonyl-CoA, leur inhibiteur physiologique, via l'induction de la MCD. La CPT-II semble également être un cible de PPAR $\alpha$  car elle est induite dans le foie par les fibrates de manière PPAR $\alpha$ -dépendante (Aoyama et al., 1998; Hashimoto et al., 1999). En cohérence avec ces observations, un PPRE fonctionnel a été identifié dans le promoteur humain de la CPT-II (Barrero et al., 2003). Finalement, l'administration de clofibrate à des rats se traduit par une augmentation du taux de translocation carnitine-acylcarnitine dans la mitochondrie (Pande and Parvin, 1980) et une induction PPAR $\alpha$ -dépendante de la carnitine:acylcarnitine translocase Slc25a20 a été rapportée suite à un traitement par le Wy14643 (Knight et al., 2005) mais, à ma connaissance, le mécanisme de régulation de la carnitine:acylcarnitine translocase par PPAR $\alpha$  n'a pas été rapporté.

L'entrée des acides gras dans le peroxysome est vraisemblablement, au moins en partie assurée par des membres de la famille des ABC transporteurs comme Abcd1 (ALD), Abcd2 (ALDR) et Abcd3 (PMP70). L'expression d'**Abcd2** et d'**Abcd3** est induite par les fibrates (Fourcade et al., 2001; Hashimoto et al., 1999) mais aucun PPRE n'a pu être pour l'instant identifié malgré l'étude de 2Kb du promoteur proximal d'Abcd2 chez le rat (Fourcade et al., 2001).

En conclusion PPAR $\alpha$  régule, généralement de manière tissu-spécifique et en coordination avec ses effets sur le métabolisme des lipoprotéines : 1) le captage des acides gras à travers la membrane plasmique (FATP1, CD36, Acs1l), 2) le trafic intracellulaire des acides gras (FABP, ACBP), et 3) l'apport des acides gras aux divers compartiments cellulaires et leur activation pour leur métabolisme ultérieur (FATP2, Acot, Acs1l, CPT-I et -II, MCD, Abcd2 et d3). La [Figure 26](#) tente de résumer les principales régulations opérées par PPAR $\alpha$  sur le captage et le trafic intracellulaire des acides gras. Nous allons voir que ces effets sont également coordonnés avec une induction majeure de l'expression de la plupart des enzymes impliquées dans le catabolisme intracellulaire des acides gras.

#### **II.4.3. Régulation du métabolisme intracellulaire des lipides par PPAR $\alpha$**

Le catabolisme intracellulaire des acides gras (AG) fait intervenir trois voies métaboliques : la  $\beta$ -oxydation mitochondriale, la  $\beta$ -oxydation peroxysomale et l' $\omega$ -hydroxylation microsomale. Ces différentes voies présentent des préférences en matière de substrats. La voie peroxysomale métabolise préférentiellement les AG à longue chaîne (plus de 20 carbones) et présente une faible activité sur les substrats à chaîne courte ou moyenne. Elle génère donc des intermédiaires d'acides gras incomplètement dégradés qui sont classiquement catabolisés dans la mitochondrie. L' $\omega$ -hydroxylation microsomale permet la dégradation de divers acides gras et de dérivés d'acides gras comme les prostaglandines. Elle génère des acides dicarboxyliques qui sont des substrats pour les  $\beta$ -oxydations mitochondriale et surtout peroxysomale (Johnson et al., 1996).

##### **$\beta$ -OXYDATION MITOCHONDRIALE :**

Lors d'un traitement par les proliférateurs de peroxysomes (PPs), l'entrée massive d'AG dans la mitochondrie (voir paragraphe précédent) est accompagnée d'un accroissement du flux oxydatif mitochondrial des acides gras, via l'induction par PPAR $\alpha$  de l'expression de plusieurs gènes codant des enzymes clefs de cette voie catabolique (Aoyama et al., 1998). La

$\beta$ -oxydation mitochondriale chez les animaux supérieurs est assurée par la succession de quatre réactions enzymatiques (Figure 27, Uchida et al., 1992). Plusieurs enzymes de cette voie présentent des spécificités de substrat liées à la longueur des chaînes carbonées. De plus, une enzyme fixée à la membrane interne de la mitochondrie, dite « **protéine trifonctionnelle** » (TP), constituée de deux sous-unités (TP $\alpha$  ou Hadha et TP $\beta$  ou Hadhb) et présentant la triple activité enzymatique enoyl-CoA hydratase/3-hydroxyacyl-CoA déshydrogénase/3-cétoacyl-CoA thiolase a été purifiée (Uchida et al., 1992). L'expression de cette enzyme est induite par un fibrate de manière PPAR $\alpha$ -dépendante (Aoyama et al., 1998). La première réaction de la  $\beta$ -oxydation mitochondriale est donc catalysée par quatre enzymes caractérisées par des spécificités de longueur de chaîne carbonée de leur substrat : **SCAD** (ou Acads, EC 1.3.99.2), **MCAD** (ou Acadm, EC 1.3.99.3), **LCAD** (ou Acadl, EC 1.3.99.13) et **VLCAD** (ou Acadvl, EC 1.3.99.-) respectivement pour short, medium, long et very long chain acyl-CoA dehydrogenase. De nombreuses autres déshydrogénases ont été identifiées dans des criblages protéomiques des protéines mitochondrielles (Mootha et al., 2003a; Taylor et al., 2003) et il est donc possible que certaines interviennent aussi dans la  $\beta$ -oxydation mitochondriale d'acides gras particuliers (acides gras branchés par exemple) mais pour la plupart leurs substrats n'ont pas été décrits. Dans la base de données Gene du NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), on trouve notamment Acad8, Acad9, Acad10 et Acadsb dont la localisation mitochondriale est connue et qui présentent plus ou moins d'homologie avec les acyl-CoA déshydrogénases « classiques ». L'expression des quatre enzymes SCAD, MCAD, LCAD et VLCAD est augmentée chez des souris de type sauvage mais pas chez les souris PPAR $\alpha$ -/- suite à un traitement par un fibrate (Aoyama et al., 1998). De plus, l'expression constitutive des LCAD et VLCAD est réduite chez les souris PPAR $\alpha$ -/- (Aoyama et al., 1998). L'induction de la MCAD dans le foie et le cœur des souris par le jeûne dépend également de l'expression de PPAR $\alpha$  (Leone et al., 1999) et un PPRE fonctionnel a été identifié dans son promoteur humain (Gulick et al., 1994). A ma connaissance, aucun PPRE n'a encore été rapporté dans les promoteurs des autres acyl-CoA déshydrogénases mais il est très probable qu'ils en contiennent également. La seconde étape de la  $\beta$ -oxydation mitochondriale fait intervenir au moins deux hydratases : la short chain (SCEH ou Echs1, EC 4.2.1.17) et la long chain enoyl-CoA hydratase (LCEH, EC 4.2.1.74) mais le rôle de PPAR $\alpha$  dans la régulation de ces enzymes n'a pas fait l'objet de publication. Cependant, une légère induction de l'activité enoyl-CoA hydratase mitochondriale sur les substrats à courte chaîne carbonée a été observée suite à l'administration d'un phtalate chez le rat mais cet effet est

## La $\beta$ -oxydation mitochondriale des acides gras

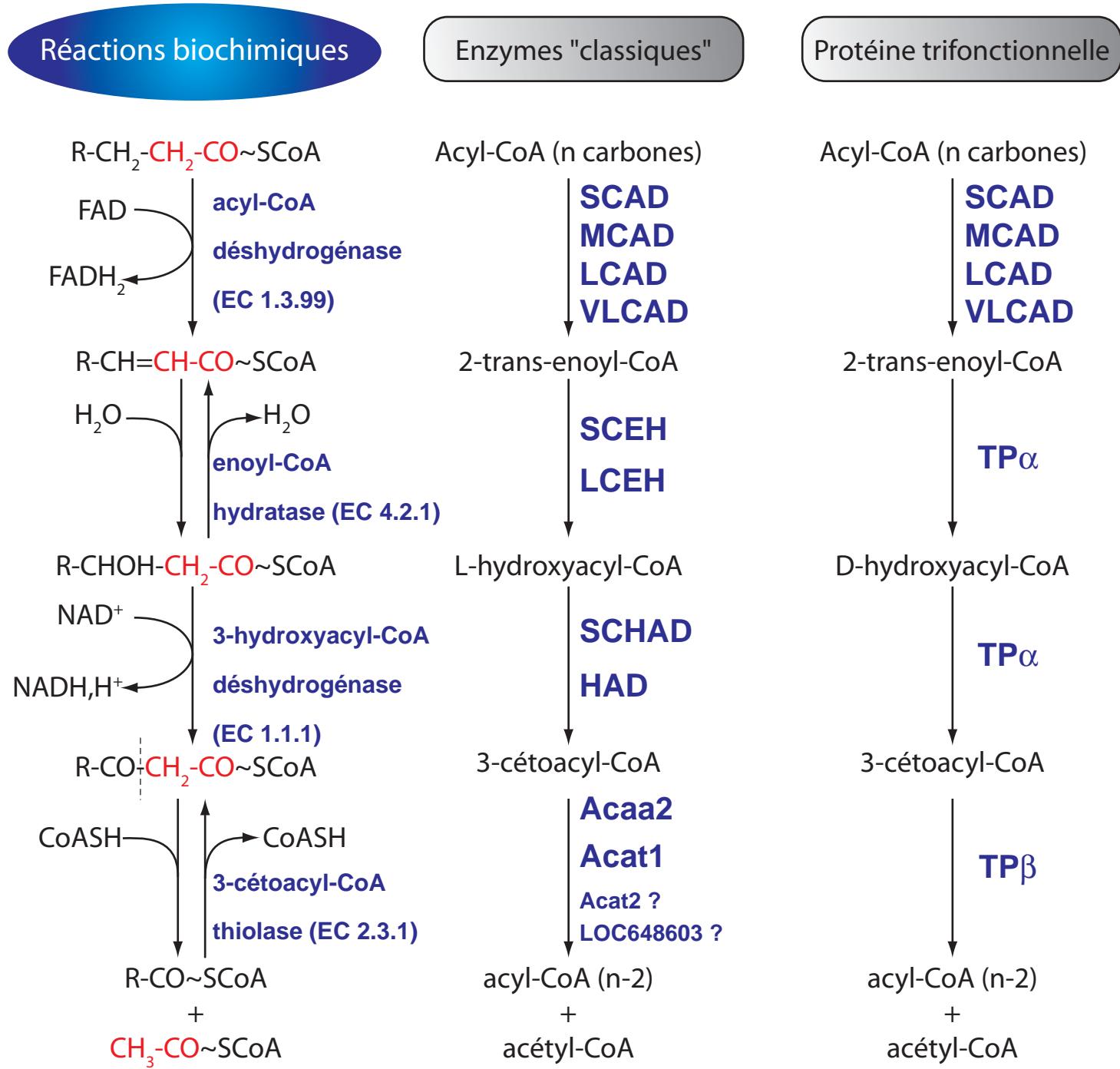


Figure 27 : La  $\beta$ -oxydation mitochondriale des acides gras

Les réactions biochimiques conduisant au raccourcissement de 2 carbones d'un acyl-CoA sont représentées à gauche du schéma. Les noms des fonctions enzymatiques intervenant dans ces réactions sont indiquées en bleu. Au centre et à droite, les noms des intermédiaires et les abréviations des noms des enzymes impliquées dans les deux voies de  $\beta$ -oxydation mitochondriale sont représentées. La protéine trifonctionnelle (TP $\alpha$ / $\beta$  ou Hadha/b) réalise les trois dernières étapes de la  $\beta$ -oxydation mitochondriale. S, M, L et VLCAD (ou Acads, Acadm, Acadl et Acadvl) : short, medium, long et very long chain acyl-CoA dehydrogenase, S et LCEH : short et long chain enoyl-CoA hydratase, SCHAD (ou Hadh2 ou Hsd17b10 pour 17 $\beta$ -hydroxysteroid dehydrogenase) : short chain 3-hydroxyacyl-CoA dehydrogenase, HAD (ou Hadh) : 3-hydroxyacyl-CoA dehydrogenase, Acaa2 : acetyl-CoA acyltransferase 2 (ou T1 pour 3-cétoacyl-CoA thiolase 1), Acat1 : acetyl-CoA acetyltransferase 1 (ou T2 pour acétoacétyl-CoA thiolase). La localisation subcellulaire d'Acat2 et de LOC648603 n'étant pas connue (ainsi que l'activité de LOC648603), il n'est pas certain que ces enzymes interviennent dans la  $\beta$ -oxydation mitochondriale.

néanmoins beaucoup plus marqué sur les activités hydratases microsomale et peroxysonomale (Cook et al., 1986). La troisième réaction de la  $\beta$ -oxydation mitochondriale est assurée par la short chain 3-hydroxyacyl-CoA dehydrogenase (SCHAD également nommé Hadh2 ou Hsd17b10 pour 17 $\beta$ -hydroxystéroïde déshydrogénase 10, EC 1.1.1.178/239/159/150) et par la 3-hydroxyacyl-CoA déshydrogénase (HAD ou Hadh, EC 1.1.1.35, Yang et al., 2005). L'enzyme trifonctionnelle (TP) porte elle une activité long chain 3-hydroxyacyl-CoA dehydrogenase que l'on trouve parfois dans la littérature sous l'abréviation LCHAD (EC 1.1.1.211). Signalons également que certaines enzymes considérées initialement comme peroxysonomales ont également été identifiés dans le protéome des mitochondries (Mootha et al., 2003a). De manière surprenante, l'abondance hépatique de la protéine **SCHAD** est environ quatre fois supérieure chez les souris PPAR $\alpha$ -/- par rapport aux souris contrôles bien que les niveaux d'ARNm correspondant soit comparables entre les deux génotypes (Aoyama et al., 1998). Dans cette même étude, une induction de la SCHAD par un fibrate, dépendante de l'expression de PPAR $\alpha$ , est rapportée chez la souris. Un traitement par le bezafibrate est également capable d'induire l'expression de plusieurs gènes impliqués dans la  $\beta$ -oxydation mitochondriale, dont la MCAD et la protéine trifonctionnelle (sous unité  $\alpha$ ) dans des fibroblastes de patients atteints d'une déficience en VLCAD (Djouadi et al., 2005), illustrant ainsi le potentiel des fibrates dans les pathologies liées à des déficiences en certaines enzymes de l'oxydation des acides gras. Enfin, la dernière étape de la  $\beta$ -oxydation mitochondriale est catalysée par au moins deux thiolases mitochondrielles distinctes : l'acetyl-CoA acyltransferase 2 (Acaa2 ou T1, EC 2.3.1.16, Arakawa et al., 1987), et l'acétoacétyl-CoA thiolase (**T2** ou **Acat1** pour acetyl-CoA acyltransferase 1, EC 2.3.1.9 qui ne doit pas être confondue avec Soat1 pour sterol O-acyltransferase et généralement appelée ACAT1 pour acyl-CoA:cholestérol acyltransférase intervenant dans la synthèse des esters de cholestérol). L'Acat1 n'agit que sur les substrats à courte chaîne comme le 3-oxohexanoyl-CoA ou l'acétoacétyl-CoA. Par ailleurs, elle catalyse également la réaction inverse de condensation de deux acétyl-CoA en acétoacétyl-CoA qui constitue la première étape de la cétogenèse (voir ci-dessous et Fukao et al., 1990; Huth et al., 1975). L'Acat2, dont la localisation subcellulaire n'est pas connue, catalyse les mêmes réactions qu'Acat1 mais elle ne doit pas non plus être confondue avec Soat2 (Sterol-O-acyltransférase 2, EC 2.3.1.26) qui est également souvent appelée ACAT2 (pour acyl-CoA:cholestérol acyltransferase) et intervient aussi dans la formation des esters de cholestérol. Dans les bases de données, on trouve également chez l'homme un troisième gène (LOC648603) qui présente une forte homologie avec les 3-

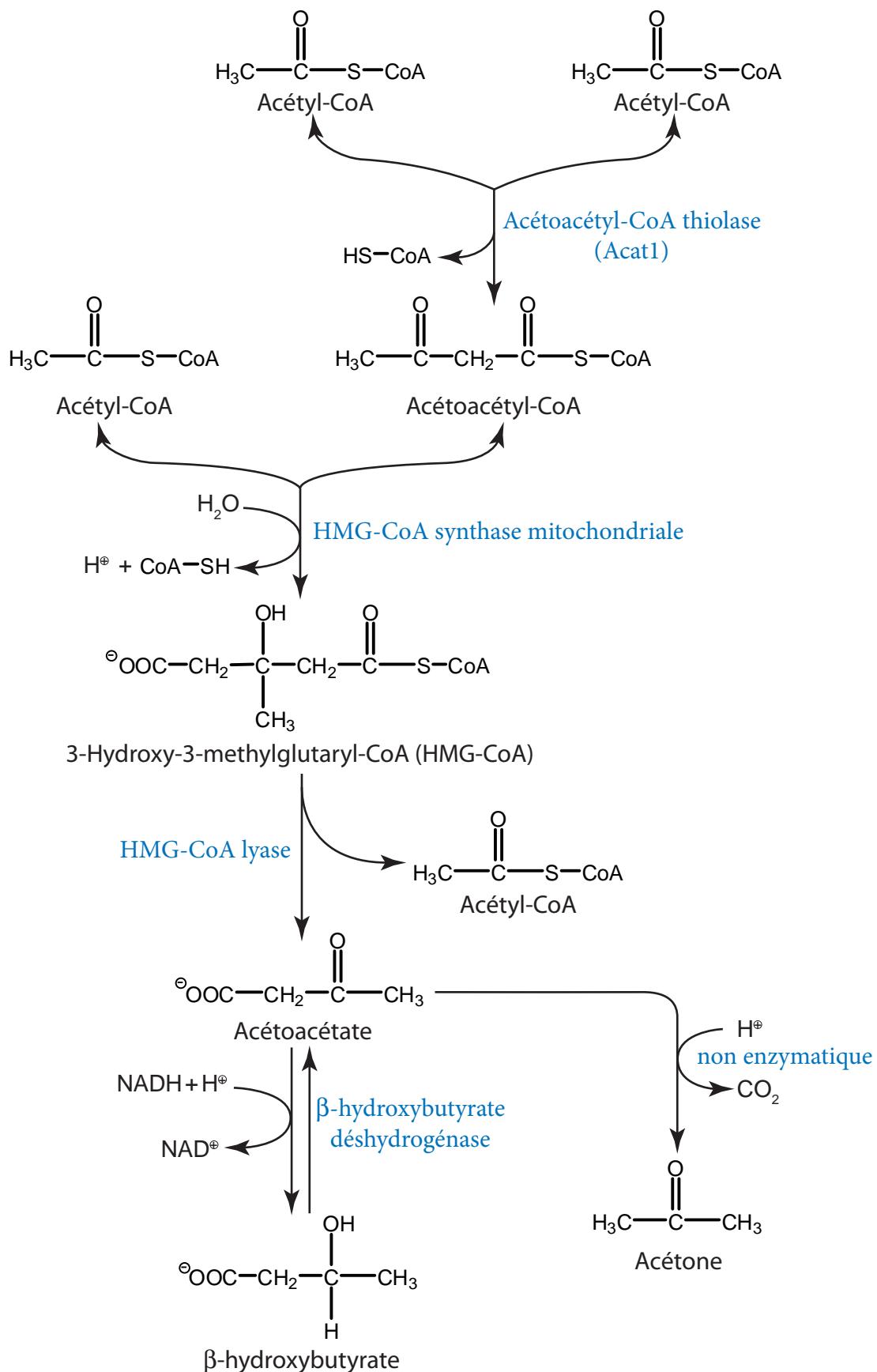


Figure 28 : Biosynthèse des corps cétoniques (cétogenèse)

Les corps cétoniques principaux sont le  $\beta$ -hydroxybutyrate et l'acetoacétate. Les petites quantités d'acétone sont dues à une décarboxylation non enzymatique de l'acetoacétate. L'acétyl-CoA provient essentiellement de la  $\beta$ -oxydation des acides gras et de l'oxydation du pyruvate. L'ensemble de ces réactions se déroulent dans la matrice mitochondriale.

cétoacyl-CoA thiolases mais dont l'ADNc n'a pas été cloné dans son intégralité et dont la fonction et la localisation subcellulaire n'ont pas été étudiées. L'expression de l'Acat1 et de l'Acaa2 est légèrement induite au niveau protéique par un fibrate dans le foie de souris de type sauvage mais pas chez les souris PPAR $\alpha$ -/- (l'Acat1 est nommée T2 et l'Acaa2 est nommée T1 dans la publication, Aoyama et al., 1998). L'expression constitutive de l'Acat1 n'est cependant pas diminuée chez les souris PPAR $\alpha$ -/- contrairement à l'Acaa2 (Aoyama et al., 1998). Les acides gras polyinsaturés comme l'acide linoléique (C18:2 $\omega$ 6) et l'acide  $\alpha$ -linolénique (C18:3 $\omega$ 3) nécessitent des enzymes supplémentaires connues sous le nom d'enoyl-CoA isomérasées pour être dégradées par la  $\beta$ -oxydation (Janssen and Stoffel, 2002). La régulation de ces enzymes par PPAR $\alpha$  n'a pas fait l'objet de publication spécifique à ma connaissance mais au moins l'une d'entre elles est fortement surexprimée dans le tissu adipeux brun au cours de l'acclimatation au froid, un processus au cours duquel PPAR $\alpha$  jouerait un rôle (Guardiola-Diaz et al., 1999). De plus, dans le cadre de ce doctorat, nous avons identifié une nette sous-expression hépatique de ces enzymes chez les souris PPAR $\alpha$ -/- qui pourrait expliquer, au moins en partie, l'accumulation préférentielle des acides gras essentiels (acides linoléique et  $\alpha$ -linolénique) dans leur foie (Martin et al., 2007). Une altération de l'expression d'une enoyl-CoA isomérase (Eci ou Dci pour dodecanoyl-CoA  $\delta$ -isomérase) chez les souris PPAR $\alpha$ -/- en situation de jeûne a également été rapportée (résultats non publiés rapportés dans Mandard et al., 2004) ainsi qu'une induction PPAR $\alpha$ -dépendante de ce même gène par le Wy14643 (Knight et al., 2005).

Ainsi, PPAR $\alpha$  contrôle l'expression de nombreuses enzymes clefs de la  $\beta$ -oxydation mitochondriale. Bien que la plupart de ces régulations aient été étudiées dans le foie, c'est sans doute dans les organes « gourmands » en énergie provenant majoritairement de l'oxydation mitochondriale des acides gras en situation normale comme le cœur que le rôle majeur de PPAR $\alpha$  dans ces régulations se révèle le plus nettement (pour une synthèse, voir Huss and Kelly, 2004). L'expression constitutive des gènes impliqués dans le captage des acides gras, leur transport dans la mitochondrie et leur  $\beta$ -oxydation mitochondriale est diminuée dans le cœur des souris PPAR $\alpha$ -/- (Djouadi et al., 1999; Watanabe et al., 2000) conduisant à une diminution des taux de captage et d'oxydation des acides gras (Campbell et al., 2002; Watanabe et al., 2000). Malgré ces altérations métaboliques, la fonction cardiaque est maintenue chez la souris adulte en l'absence de stress. Chez des souris relativement âgées (9 mois), une dysfonction cardiaque est néanmoins déjà observée qui pourrait être liée à des défauts mécaniques au niveau des chaînes de myosine subissant des dommages oxydatifs

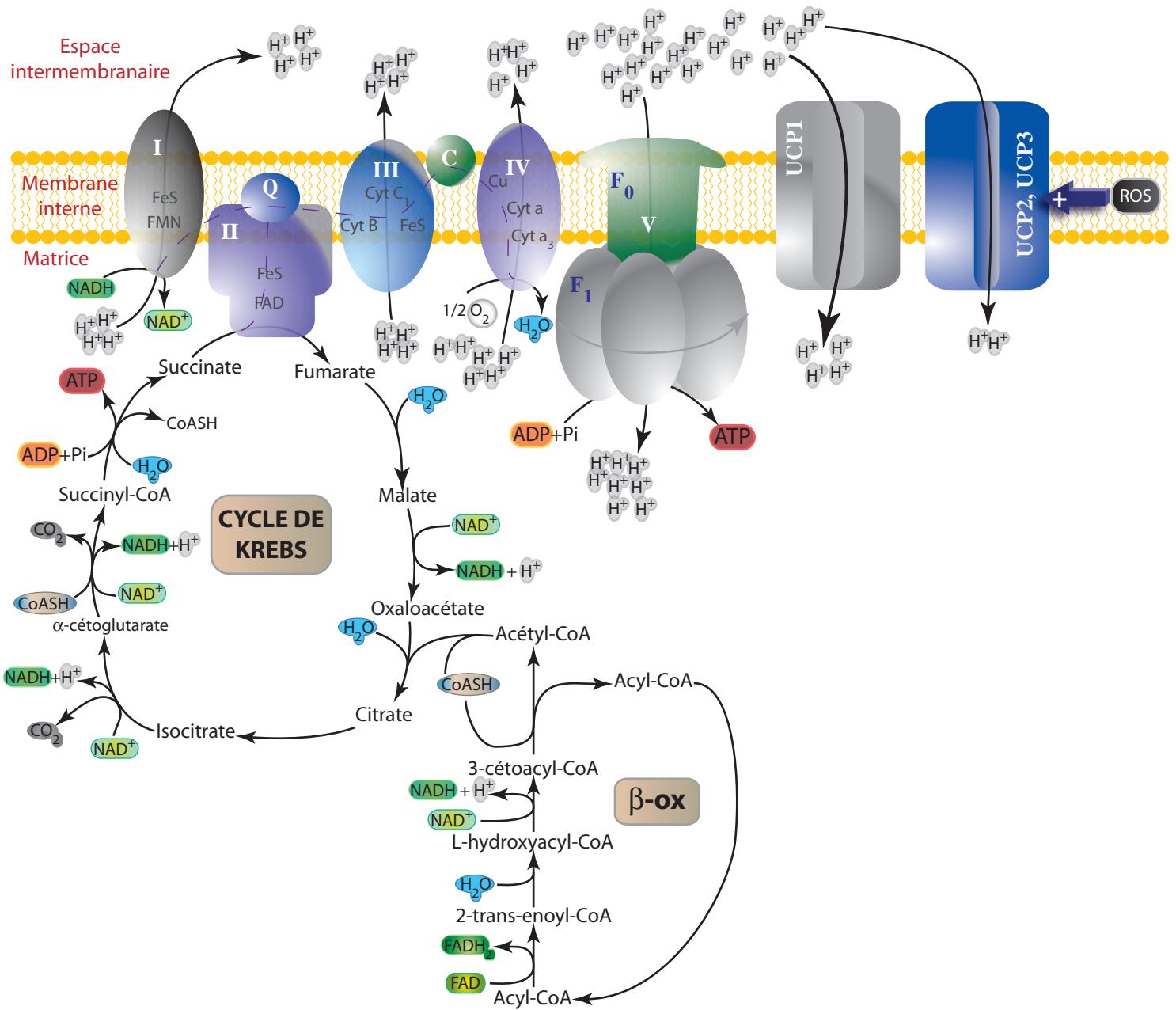
(Guellich et al., 2007) mais également à des défauts du métabolisme des acides gras. Bien qu’ayant une durée de vie normale, les souris PPAR $\alpha$ -/- développent progressivement une fibrose cardiaque et une fragmentation des myofibrilles associées à une ultrastructure anormale des mitochondries (Watanabe et al., 2000). En situation de jeûne, ce qui induit normalement la  $\beta$ -oxydation mitochondriale cardiaque chez la souris de type sauvage, les souris PPAR $\alpha$ -/- développent une hypoglycémie et accumulent des TG dans le foie et le cœur (Leone et al., 1999), illustrant à nouveau l’importance de la régulation de l’oxydation des acides gras par PPAR $\alpha$ .

#### CETOGENESE :

Au cours d’un jeûne, la stimulation lipolytique du tissu adipeux conduit à une augmentation des acides gras libres circulants qui sont rapidement captés par le foie et s’y accumulent, constituant un pool d’activateurs potentiels de PPAR $\alpha$ . De plus, l’augmentation des glucocorticoïdes et la diminution de l’insuline accroissent l’expression de PPAR $\alpha$ . Ces effets combinés conduisent à une activation de l’oxydation des acides gras via PPAR $\alpha$ . Dans une telle situation, les réserves de glycogène s’épuisent rapidement et certains tissus comme le cerveau ou les muscles nécessitent un apport d’énergie sous la forme de corps cétoniques (acétone mais surtout acétoacétate et  $\beta$ -hydroxybutyrate). La **cétogenèse** ([Figure 28](#)), s’opère au niveau hépatique, dans la mitochondrie via la condensation de deux acétyl-CoA provenant de l’oxydation des acides gras ou du pyruvate (pour une synthèse, voir Hegardt, 1998). L’enzyme limitante de cette synthèse est la 3-hydroxy-3-methylglutaryl-CoA synthase mitochondriale (**HMGCoA synthase mitochondriale** ou mHMGCoA synthase) qui est également un gène cible de PPAR $\alpha$  (Le May et al., 2000). Les promoteurs murins (Rodriguez et al., 1994) et porcins (Ortiz et al., 1999) de l’HMG-CoA synthase mitochondriale possèdent un PPRE fonctionnel. Comme dans le cas de l’enzyme bifonctionnelle L-PBE évoqué au paragraphe II.3.2., il a été proposé que l’HMG-CoA synthase mitochondriale autorégule sa propre transcription en interagissant physiquement avec PPAR $\alpha$ , agissant ainsi comme un coactivateur de PPAR $\alpha$  sur son propre promoteur (Meertens et al., 1998). L’HMG-CoA lyase qui agit sur l’HMG après sa synthèse par l’HMG-CoA synthase, serait également régulée de manière PPAR $\alpha$ -dépendante au cours du jeûne (Mandard et al., 2004) et d’un traitement par le Wy14643 (Knight et al., 2005).

### *PROTEINES DE DECOUPLAGE (UCP ou Uncoupling proteins) :*

Les acides gras catabolisés par la  $\beta$ -oxydation et le pyruvate, provenant principalement du glucose ou du lactate, et oxydé par le complexe pyruvate déshydrogénase (localisé dans la membrane interne de la mitochondrie) fournissent de l'acétyl-CoA qui est oxydé par le cycle de Krebs (ou cycle des acides tricarboxyliques ou cycle du citrate) dans la mitochondrie. Les NADH et FADH<sub>2</sub> générés par le cycle de Krebs sont utilisés par la chaîne de transport d'électron (ou chaîne respiratoire) pour établir un gradient de proton à travers la membrane interne de la mitochondrie qui, lorsqu'il est couplé à la phosphorylation oxydative, assure la formation d'ATP par l'ATP synthase (voir [Figure 29](#)). Les protéines **UCP** (Uncoupling proteins) localisées dans la membrane interne de la mitochondrie sont capables d'agir comme des canaux à protons et de découpler la formation du gradient de proton de la synthèse d'ATP, conduisant ainsi à une forte consommation d'oxygène et à la production de chaleur ([Figure 29](#)). L'UCP1 est la principale isoforme impliquée dans la thermogénèse adaptative qui s'opère dans le tissu adipeux brun suite à une exposition au froid. Dans le tissu adipeux brun du rat, UCP-1 semble régulée par PPAR $\alpha$  via un PPARE partiellement conservé chez l'homme et la souris mais uniquement dans des situations où le tissu est supposé pauvre en acides gras (lactation ou nouveau-né à la neutralité thermique) et donc où l'expression d'UCP1 est faible (Barbera et al., 2001). Dans ce tissu, PPAR $\alpha$  intervient également dans la régulation des gènes impliqués dans l'oxydation des acides gras suite à une exposition au froid (Guardiola-Diaz et al., 1999). Cependant, l'expression d'UCP1 n'est pas altérée dans le tissu adipeux brun des souris PPAR $\alpha$ -/- en situation d'alimentation ou de jeûne (Kersten et al., 1999), suggérant que la régulation d'UCP1 par PPAR $\alpha$  n'est pas pertinente dans ce contexte. Les UCP2 et UCP3, exprimées de manière beaucoup plus ubiquitaire qu'UCP1, sont présentent à des niveaux moins importants qu'UCP1 dans la membrane mitochondriale et assureraient donc une régulation fine du découplage entre oxydation des substrats, fonctionnement de la chaîne respiratoire et formation du gradient de proton (Brand and Esteves, 2005). Leurs fonctions physiologiques diffèreraient notamment de celle, principalement thermogénique, d'UCP1 et seraient de plus sous le contrôle d'activateurs comme les acides gras ou les espèces réactives de l'oxygène (ou radicaux libres, ROS en anglais pour reactive oxygen species) via la lipoperoxydation des acides gras insaturés en dérivés réactifs comme le 4-hydroxynonenal (pour une synthèse, voir Brand and Esteves, 2005). UCP2 et UCP3 intervientraient notamment pour limiter la production des espèces réactives de l'oxygène qui est élevée au cours de l'oxydation des acides gras et pour protéger les organes des dommages oxydatifs



**Figure 29 : Chaîne respiratoire mitochondriale et protéines de découplage**

Cette figure met particulièrement l'accent sur les cofacteurs des principales réactions biochimiques mitochondrielles évoquées dans le texte ( $\beta$ -oxydation ou  $\beta$ -ox et cycle de Krebs), sur l'établissement d'un gradient de protons ( $H^+$ ) utilisé par l'ATP synthase pour la production d'ATP ou diminué par les protéines de découplage UCP (uncoupling proteins). La chaîne de transfert d'électron ou chaîne respiratoire comporte 5 complexes. Complexe I : NADH-ubiquinone réductase, Complexe II : Succinate-quinone réductase, Complexe III : Ubiquinone-cytochrome C réductase, Complexe IV : Cytochrome oxydase et Complexe V : ATP synthase. Q : coenzyme Q ou ubiquinone, C : cytochrome C. Le trajet des électrons est illustré par le trait pointillé mauve et les principaux cofacteurs impliqués dans ces transferts sont indiqués en gris au sein de chaque complexe. La moindre capacité de transport de protons des UCP2 et 3 par rapport à l'UCP1 ainsi que la nécessité de leur activation par l'intermédiaire de la génération d'espèces réactives de l'oxygène (ROS) produites en particulier par la chaîne respiratoire sont également illustrés.

causés par ces radicaux libres. Elles interviendraient peut-être aussi pour favoriser l'oxydation des acides gras au niveau mitochondrial bien que les données concernant ces mécanismes semblent moins probants (Brand and Esteves, 2005). Au niveau du pancréas, elles inhiberaient la sécrétion d'insuline en réponse au glucose. L'expression d'UCP2 et d'UCP3 est induite par les acides gras et les PPARs pourraient jouer un rôle dans ces régulations. Des agonistes de PPAR $\alpha$  ou un jeûne augmentent l'expression d'UCP3 de manière PPAR $\alpha$ -dépendante dans le cœur et dans le muscle alors qu'UCP2 n'est que légèrement augmentée dans le muscle (Brun et al., 1999; Young et al., 2001b). Plusieurs PPRE potentiels ont été identifiés dans le promoteur de l'UCP3 humaine mais leur fonctionnalité n'a pas été évaluée (Tu et al., 2000). En revanche, dans des cardiomyocytes de nouveau-né, dans des cultures primaires d'hépatocytes de rongeurs, *in vivo* dans le foie de souris et dans une lignée cellulaire d'insulinome, les agonistes de PPAR $\alpha$  induisent l'expression d'UCP2 (Armstrong and Towle, 2001; Nakatani et al., 2002; Tordjman et al., 2002; Tsuboyama-Kasaoka et al., 1999), suggérant que PPAR $\alpha$  jouerait bien un rôle dans les régulations d'UCP2 et d'UCP3. Cependant, tant que les rôles précis de ces deux UCP n'ont pas été fermement établis, il semble difficile d'appréhender les conséquences physiologiques de ces régulations.

#### *$\beta$ -OXYDATION PEROXYSOMALE :*

Le catabolisme des acides gras à longue chaîne est assuré principalement dans les peroxysomes par les deux voies de  $\beta$ -oxydation peroxysomale que nous avons décrites au paragraphe II.2.1. et qui sont illustrées en [Figure 20](#). La régulation par PPAR $\alpha$  de l'expression des enzymes impliquées dans ces voies a été largement étudiée (pour une synthèse, voir Osumi, 1993). L'acyl-CoA oxydase (ACO ou AOX ou **Acox1**, EC 1.3.3.6) est l'enzyme limitante de la voie dite « inductible » et constitue un marqueur très utilisé de la prolifération des peroxysomes. Son promoteur chez le rat contient deux PPRE fonctionnels de type DR1 (dont l'un est illustré en Table 8, Dreyer et al., 1992; dont l'un est illustré en Table 8, Osumi et al., 1991; Tugwood et al., 1992) et un PPRE a également été identifié dans le promoteur humain de l'Acox1 (Varanasi et al., 1996). Ce PPRE dans le promoteur humain fixe bien le dimère PPAR $\alpha$ -RXR et permet à un gène rapporteur d'être induit par les PPs ou par l'acide 9-cis rétinoïque *in vitro*. Pourtant, même en surexprimant hPPAR $\alpha$  dans des cellules HepG2, l'Acox1 humaine n'est pas induite par les agonistes de PPAR $\alpha$  (Lawrence et al., 2001) suggérant que des séquences adjacentes au PPRE pourraient inhiber son activation via PPAR $\alpha$  chez l'homme. L'enzyme bifonctionnelle (L-PBE, EC 4.2.1.17/EC 1.1.1.35) qui

catalyse les deuxième et troisième étapes de la voie « inductible » de la  $\beta$ -oxydation peroxy somale est un gène cible de PPAR $\alpha$  qui répond fortement aux PPs chez la souris. Trois PPRE différents ont été identifiés dans le promoteur de la **L-PBE** mais un seul semble permettre une transactivation efficace du gène L-PBE par PPAR $\alpha$  in vitro (Bardot et al., 1993; Bardot et al., 1995; Chu et al., 1995; Marcus et al., 1993; Zhang et al., 1993; Zhang et al., 1992) comme in vivo (Alvares et al., 1994). Comme pour l'Acox1, les fibrates n'activent pas la transcription de la L-PBE dans des cellules HepG2, même lorsque PPAR $\alpha$  est surexprimé (Lawrence et al., 2001). Enfin, la dernière étape de la  $\beta$ -oxydation peroxy somale est catalysée par deux thiolases différentes dont les gènes ont été récemment clonés (Chevillard et al., 2004a) : La thiolase A et la thiolase B. L'expression de ces deux gènes est induite par les fibrates mais la thiolase B semble répondre mieux à ce type de traitement et est généralement qualifiée de « thiolase inductible » (Chevillard et al., 2004a; Hijikata et al., 1990). L'induction de la thiolase B est largement dépendante de PPAR $\alpha$  mais une légère induction résiduelle est néanmoins observée chez les souris PPAR $\alpha$ -/- traitées par un fibrat (Chevillard et al., 2003; Lee et al., 1995). Plusieurs séquences susceptibles de constituer des PPRE ont été identifiées dans le promoteur de la thiolase B (Hijikata et al., 1990) et la fixation du dimère PPAR $\alpha$ -RXR a été démontré sur l'une d'elles constituant un DR1 (Kliewer et al., 1992). Cependant, cet élément de réponse ne semble pas fonctionnel en sois vis-à-vis de PPAR $\alpha$  mais il fixe HNF-4 et semble fonctionnel vis-à-vis de ce RN (Nicolas-Frances et al., 2000).

La seconde voie de  $\beta$ -oxydation peroxy somale dite « non inductible » ou D-hydroxy-spécifique permet la dégradation des AG à chaîne ramifiée (Qi et al., 1999a) et intervient également dans l'étape finale de synthèse des principaux acides biliaires chenodéoxycholate et cholate (Figure 20 et 30, Ferdinandusse et al., 2005; Figure 20 et 30, Wanders et al., 2001). L'expression constitutive de la D-PBE est réduite d'environ 50% chez les souris PPAR $\alpha$ -/- (Aoyama et al., 1998). De plus, son expression est induite chez les souris contrôles (Corton et al., 1997) mais pas chez les souris PPAR $\alpha$ -/- suite à l'administration d'un PP (Aoyama et al., 1998; Fan et al., 1998b). L'expression de SCPx, la thiolase catalysant la dernière étape de cette voie semble également inductible chez la souris et le rat par les PPs (Jia et al., 2003; Mori et al., 1991) et la présence de deux PPRE a été rapportée dans le promoteur de rat de SCPx (Lopez et al., 2003) L'expression hépatique de la D-PBE et de SCPx est également induite par le jeûne de manière dépendante de PPAR $\alpha$  (Hashimoto et al., 2000). Ces résultats illustrent que la qualification « non inductible » est sans doute inappropriée pour cette voie

métabolique. De plus, la régulation de ces enzymes qui interviennent dans l'étape finale de synthèse d'acides biliaires illustre l'implication de PPAR $\alpha$  dans la régulation du métabolisme de ces dérivés du cholestérol indispensables à la digestion des graisses et à l'élimination du cholestérol.

#### *$\omega$ -HYDROXYLATION MICROSOMALE :*

La prolifération des peroxysomes s'accompagne généralement d'une augmentation des concentrations de plusieurs enzymes du réticulum endoplasmique (Schoonjans et al., 1996b). En particulier, les expressions des enzymes de la famille des **cytochromes P450 4A** (CYP4A), sont parmi celles qui augmentent le plus suite à un traitement par des PPs du fait d'une induction au niveau transcriptionnel (Schoonjans et al., 1996b). Ces enzymes interviennent, avec les CYP4B et CYP4F, dans l' $\omega$ -hydroxylation de divers acides gras et particulièrement dans le métabolisme de l'acide arachidonique (C20:4 $\omega$ 6) et de ses dérivés comme les prostaglandines de type 2. Cette voie de catabolisme des acides gras joue un rôle mineur en situation normale mais l' $\omega$ -hydroxylation est fortement accrue en situation de jeûne ou de diabète, ce qui est due à une forte induction PPAR $\alpha$ -dépendante des CYP4A (Kroetz et al., 1998). Un PPRE fonctionnel a été identifié dans le promoteur de CYP4A1 (Aldridge et al., 1995) qui est responsable, chez les rongeurs, de son induction par les activateurs de PPAR $\alpha$  (Hashimoto et al., 2000; Kroetz et al., 1998; Lee et al., 1995; Leone et al., 1999). CYP4A3 est également fortement induit par PPAR $\alpha$  mais aucun PPRE n'a encore été rapporté (Hashimoto et al., 2000; Kroetz et al., 1998). De plus, trois PPRE fonctionnels ont été identifiés dans le promoteur de CYP4A6 de lapin (Muerhoff et al., 1992; Palmer et al., 1994). Chez la souris, nous observons dans toutes nos expériences que les Cyp4a10 et Cyp4a14 constituent les gènes cibles de PPAR $\alpha$  qui répondent le mieux à son activation par divers stimuli (acides gras alimentaires, fibrates, jeûne) dans le foie et le rein (Martin et al., 2007; Martin et al., 2005).

A partir des données que nous venons de présenter, il est clair qu'en régulant de manière coordonnée la  $\beta$ -oxydation mitochondriale, peroxysomale et l' $\omega$ -hydroxylation microsomale, PPAR $\alpha$  constitue un régulateur clef du catabolisme intracellulaire des AG. Bien que la totalité de ces régulations ne soient pas observées de manière systématique chez l'homme, il est tentant de rapprocher ces mécanismes de régulation des effets normolipémiants des fibrates. Par exemple, un accroissement de l'activité de  $\beta$ -oxydation dans des hépatocytes de rat en culture est responsable d'une chute de la sécrétion de triglycérides et par la suite, de VLDL

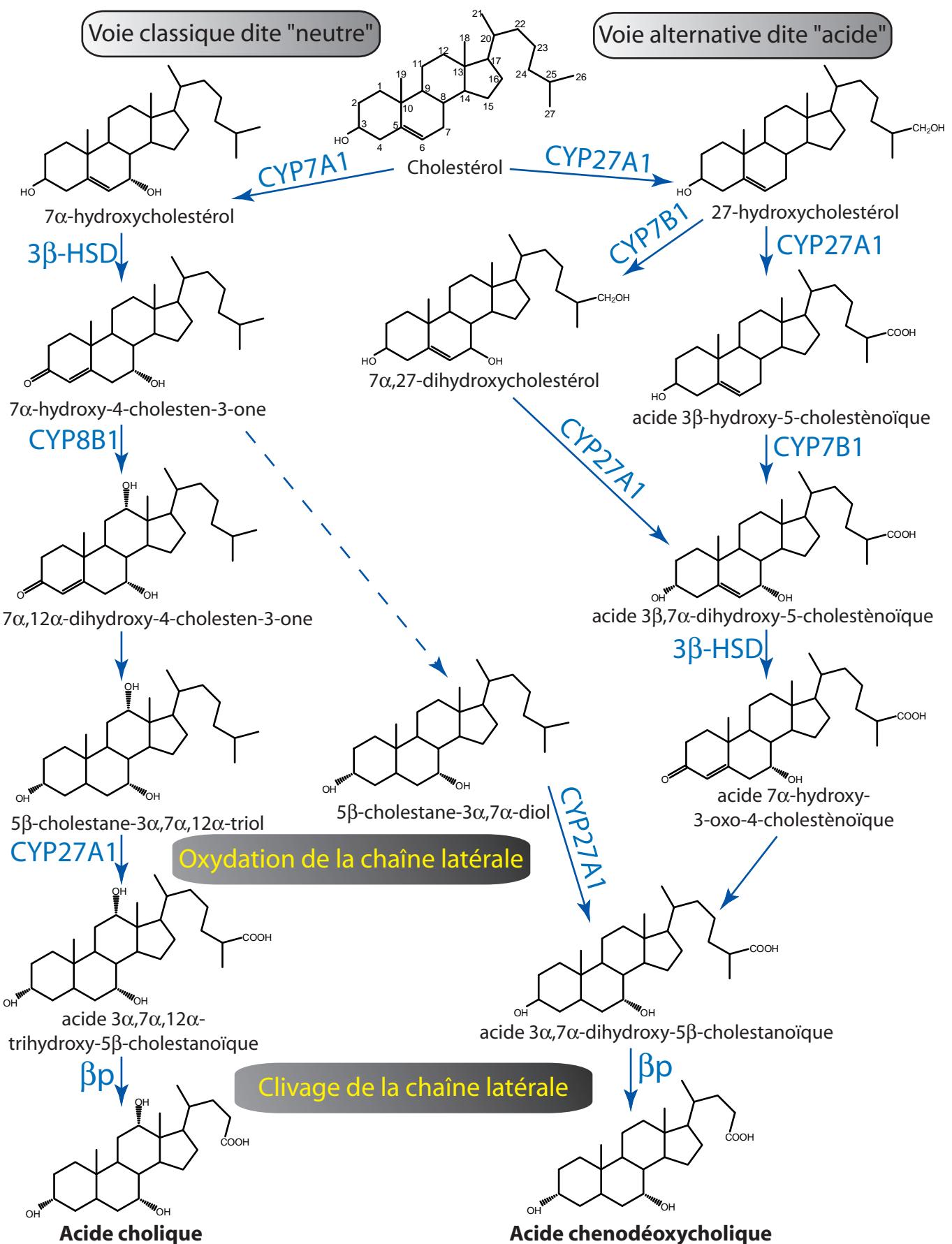


Figure 30 : Principales étapes de la biosynthèse des acides biliaires

Seules les principales enzymes de la voie de biosynthèse ainsi que leurs substrats et produits sont représentés. La dernière étape de clivage de la chaîne latérale est réalisée via un cycle de  $\beta$ -oxydation peroxysomale ( $\beta$ p) par la voie dite "non inducible" (voir Figure 20). Schéma inspiré de JYL Chiang, Endo Rev, 2002, 23(4):443-63. En réalité, acides cholique et chenodéoxycholiques, quittent la  $\beta$ -oxydation peroxysomale sous la forme de thioesters du coenzyme A et forment soit des acide biliaires libres sous l'action de thioestérases, soit des sels biliaires par conjugaison à la taurine ou à la glycine via l'enzyme BAAT (bile acyl-CoA:amino acid N-acetyltransferase).

(Skrede et al., 1994), ce qui participe peut-être, dans une certaine mesure, aux effets normolipémiants des fibrates chez l'homme.

#### *METABOLISME DES ACIDES BILIAIRES :*

Au delà de ses effets sur le catabolisme des acides gras, PPAR $\alpha$  semble également réguler d'autres aspects du métabolisme des lipides et notamment le métabolisme des acides biliaires. La biosynthèse des acides biliaires est une voie métabolique strictement hépatique qui implique de nombreuses enzymes localisées dans les différents compartiments cellulaires ([Figure 30](#), pour une synthèse, voir Russell, 2003). Tout d'abord, le noyau stéroïde apolaire du cholestérol est converti en un stéroïde beaucoup plus polaire par des enzymes du réticulum endoplasmique et du cytosol. Dans la mitochondrie, un groupement carboxyle est formé sur la chaîne aliphatique latérale de ce stérol qui est ensuite raccourcie via un cycle de  $\beta$ -oxydation peroxysomale conduisant à la synthèse de cholate-CoA et de chenodéoxycholate-CoA (CDCA-CoA). Chez la souris, le CDCA est un acide bilaire mineur car la plupart est converti par hydroxylation en acides  $\alpha$ -,  $\beta$ - et  $\gamma$ -muricholiques. Après leur synthèse, les acides biliaires sont conjugués avec de la taurine ou de la glycine par l'enzyme BAAT (bile acyl-CoA:amino acid N-acyltransferase) ou bien excrétés sous forme non conjuguée et non estérifiée au coenzyme A après l'action de thioesterases comme PTE-2 (Hunt et al., 2002). Chez la souris, on ne retrouve presque que des acides biliaires conjugués avec la taurine car la BAAT murine n'utilise pas la glycine comme substrat.

L'enzyme clef de la principale voie de biosynthèse des acides biliaires (dite « voie neutre ») est le cytochrome P450 7A1 (**Cyp7A1**) qui est l'enzyme limitante de cette voie de biosynthèse. Nous avons déjà présenté sa régulation par rétrocontrôle négatif faisant intervenir FXR, SHP et LRH-1 (paragraphe I.5.5, la [Table 10](#) présente des exemples d'éléments de réponse naturels de FXR). Chez les rongeurs, il existe également une boucle de régulation positive qui conduit à l'activation de l'expression de Cyp7A1 par un régime riche en cholestérol et qui implique LXR $\alpha$  (Peet et al., 1998). Globalement, L'expression de Cyp7A1 est influencée par les récepteurs nucléaires RAR, RXR, FXR, LXR, HNF-4, LRH-1, SHP, PXR et PPAR $\alpha$  (pour une synthèse, voir Chiang, 2002) conduisant à une régulation complexe mais fine de son expression.

Gène cible	Gene ID	Element de réponse	Remarque	Référence
Mouse ABCB11 / BSEP	27413	TCTGGACTT <del>AGGCCATTGAACT</del> TATAAGCAAATAGATAGT <sub>+21</sub> +60	FXR/RXR sur un IR1 fonctionnel aussi chez l'homme	(Ananthanarayanan et al., 2001; Ananthanarayanan et al., 2004; Schuetz et al., 2001)
Mouse i-BABP	16204	GT <sub>-203</sub> TTTOCTTA <del>AGGTGA</del> <u>AATAAOC</u> TGGGGCTC <sub>-175</sub>	FXR/RXR sur un IR1 conservé chez l'homme	(Grober et al., 1999; Zhang et al., 2003b)
Mouse SHP / NR0B2	23957	CCT <sub>-315</sub> GGGTTAAC <del>TAAATGAA</del> CCCTGT <sub>-297</sub>	FXR/RXR sur un IR1 conservé chez l'homme	(Goodwin et al., 2000)
Rat ABCC2 / MRP2	25303	CTG <sub>-401</sub> TGA <u>ACTCTTAACCAAGTICA</u> AAC <sub>-376</sub>	FXR/RXR sur un ER8 conservé chez la souris et reconnu aussi par PXR/RXR et CAR/RXR	(Kast et al., 2002)
Human apoA-V	116519	CAGCTG <del>AGGTCA</del> <u>ACTTCCTTTTGAAACT</u> TCAC <sub>-80</sub>	FXR/RXR sur un IR8 mais pas d'induction de l'apoAV par les agonistes de FXR.	(Prieur et al., 2003)
Mouse Sult2a1	20859	TTTG <del>GGGTCA</del> <u>TGAAACT</u> TGGCTCAC <sub>-168</sub> -191	FXR/RXR sur un IRO reconnu aussi par PXR/RXR et VDR/RXR	(Echchgadda et al., 2004; Song et al., 2001; Sonoda et al., 2002)

**Table 10 :** Exemples d'éléments de réponse naturels de FXR.

Dans la mesure du possible, le site d'initiation de la traduction a été choisi pour établir la position des séquences sur les promoteurs. Parfois, c'est cependant le site d'initiation de la transcription qui sert de point de départ. Les éléments de réponse sont tous présentés dans l'orientation et à partir du brin qui permet de comparer au mieux la séquence fournie dans le tableau au motif consensus AGGTCA. La numérotation des bases flanquantes permet de visualiser l'orientation de l'élément de réponse par rapport au site d'initiation de la transcription et/ou de la traduction.

Le rôle de PPAR $\alpha$  dans la régulation de Cyp7A1 apparaît peu clair dans la mesure où des observations conflictuelles ont été rapportées. Dans des cellules d'hépatome humain HepG2, la cotransfection de vecteurs d'expression pour PPAR $\alpha$  et RXR $\alpha$  en présence d'un agoniste de PPAR $\alpha$  provoque une diminution de l'activité du promoteur de Cyp7A1 humain ou de rat (Marrapodi and Chiang, 2000; Patel et al., 2000). Malgré la présence d'un DR1 activé par HNF-4 dans le promoteur de Cyp7A1 chez les deux espèces, aucune fixation directe de PPAR $\alpha$  n'a pu être détectée et il a été proposé que cet effet soit lié à une inhibition de l'expression d'HNF-4 par les fibrates dans ces cellules (Marrapodi and Chiang, 2000). Les variations diurnes de l'expression de Cyp7A1 (élevée le jour et faible la nuit) semblent néanmoins parfaitement maintenues chez les souris PPAR $\alpha$ -/- (Patel et al., 2000), suggérant que cette éventuelle régulation indirecte de Cyp7A1 par PPAR $\alpha$  n'est pas pertinente en situation physiologique normale chez la souris. A l'opposé de ces résultats, Cheema *et al.* observent une stimulation par le Wy14643 et des acides gras des promoteurs murins et humain de Cyp7A1 dans des cellules d'hépatome de rat (Cheema and Agellon, 2000). Le promoteur murin semble répondre mieux à ces stimulations en raison de la présence d'un second DR1 qui chevauche le DR4 activé par LXRx. Les effets sur le promoteur humain sont cependant modestes et l'étude d'un promoteur humain dans le contexte de cellules de rat conduit à douter de la pertinence de ces observations *in vivo* chez l'homme. Cependant, le niveau d'expression de Cyp7A1 semble réduit chez les souris PPAR $\alpha$ -/- suite à un jeûne (Hunt et al., 2000b), suggérant que chez la souris, une régulation positive, PPAR $\alpha$ -dépendante, de l'expression de Cyp7A1 par les acides gras pourrait être pertinente dans certains contextes physiopathologiques de fort afflux d'acides gras au niveau hépatique. Les contradictions apparentes entre ces différentes études ne sont pas résolues à l'heure actuelle. Quoi qu'il en soit, la synthèse des acides biliaires et la taille du pool d'acides biliaires semblent réduites en raison d'une inhibition de l'activité Cyp7A1 chez des patients atteint d'hypercholestérolémie ou de calculs biliaires et traités par les fibrates (Bertolotti et al., 1995; Stahlberg et al., 1995). De plus, une étude *in vivo* montre que les fibrates réduisent l'activité et l'ARNm de Cyp7A1 dans le foie de souris de type sauvage mais pas chez les souris PPAR $\alpha$ -/- (Post et al., 2001), cette régulation au niveau ARNm ayant également été observée par d'autres auteurs (Knight et al., 2005). Ces observations sont également reproduites à des degrés divers avec plusieurs fibrates différents sur des hépatocytes primaires de rat (Le Jossic-Corcos et al., 2004; Post et al., 2001). L'ensemble de ces données prises collectivement soulignent sans doute la régulation fine, via de multiples RNs, de l'expression de Cyp7A1 ce

qui peut conduire à des observations contradictoires en raison de changements ténus dans les conditions expérimentales. D'autres gènes impliqués dans la synthèse des acides biliaires sont également régulés par PPAR $\alpha$ . La stérol 27-hydroxylase (**Cyp27A1**), enzyme limitante de la seconde voie de synthèse des acides biliaires dite « voie acide » est régulée négativement par les fibrates de manière PPAR $\alpha$ -dépendante (Post et al., 2001). Enfin, l'expression de la sterol 12 $\alpha$ -hydroxylase (**Cyp8b1**) qui intervient spécifiquement dans la synthèse de l'acide cholique, est régulée positivement par PPAR $\alpha$  via un PPRE localisé dans son promoteur (Table 8, Hunt et al., 2000b). Nous avons également observé cette régulation PPAR $\alpha$ -dépendante de Cyp8b1 par le fenofibrate grâce à nos puces à ADN mais sa régulation via les activateurs de RXR semble plus complexe et ne fait vraisemblablement pas intervenir que PPAR $\alpha$  (Martin et al., 2005). Cette régulation pourrait être responsable de l'augmentation de la proportion d'acide cholique observée chez les patients traités par des fibrates (Stahlberg et al., 1995), ce qui peut constituer un problème lors de traitements à long terme dans la mesure où l'acide cholique, contrairement à l'acide chénodésoxycholique, augmente le degré de saturation en cholestérol de la bile, un élément déterminant dans la formation des calculs biliaires. Une augmentation du rapport entre acide cholique et (CDCA+ $\beta$ -muricholique acide) a été observée sur des hépatocytes de rat traités au ciprofibrate (Le Jossic-Corcos et al., 2004). Par ailleurs, nous avons déjà vu que la régulation, via PPAR $\alpha$ , des enzymes de la  **$\beta$ -oxydation peroxysomale** est également susceptible de modifier le profil des acides biliaires synthétisés dans la mesure où les acides di- et trihydroxycholestanoïques (DHCA et THCA) doivent subir un cycle de  $\beta$ -oxydation peroxysomale pour former respectivement les acides chénodésoxycholique (CDCA) et cholique (CA). De fait l'oxydation de la chaîne latérale du TCHA est accrue chez le rat ou le lapin suite à une exposition par un fibrate ou un rétinoïde (Farrants et al., 1993). De plus, PPAR $\alpha$  semble également impliqué dans la régulation des enzymes intervenant dans la conjugaison des acides biliaires à la taurine et à la glycine. Deux types d'enzymes entrent en compétition pour leurs substrats communs que sont les acides biliaires activés par le coenzyme A produits par les voies de biosynthèse que nous venons de décrire : des thioestérases comme PTE-2 (Hunt et al., 2002), présentes dans les peroxysomes et qui forment des acides biliaires non conjugués, et l'enzyme BAAT (bile acyl-CoA:amino acid N-acetyltransferase) présente dans les peroxysomes comme dans le cytosol et qui forme les sels biliaires (acides biliaires conjugués à la taurine et à la glycine). Il a été montré que l'activation de PPAR $\alpha$  par le Wy14643 diminuait nettement l'activité et l'expression de la BAAT dans les peroxysomes et induisait l'activité bile acyl-CoA thioestérase dans le cytosol

et les peroxysomes (Solaas et al., 2004) ce qui pourrait conduire à augmenter la sécrétion d'acides biliaires non conjugués et à diminuer celle des sels biliaires. Ces effets n'ont cependant pas été étudiés et la pertinence physiologique de ces régulations reste donc à démontrer. Cependant, chez l'homme, PPAR $\alpha$  régule également via des PPRE fonctionnels l'expression de l'hydroxystéroïde sulfotransférase SULT2A1 (Fang et al., 2005) et de l'UDP-glucuronosyltransferase 2B4 ou UGT2B4 (Barbier et al., 2003a), deux enzymes de phase II qui interviennent respectivement dans la formation de sulfo- et de glucurono-conjugués des acides biliaires, moins toxiques et plus facilement éliminés de l'organisme via la bile ou les urines. Ainsi par son action sur la BAAT et l'activité acyl-CoA thioestérase (permettant l'action des enzymes de phase I comme le Cyp3A) ainsi que via l'induction d'enzymes de phase II, PPAR $\alpha$  pourrait faciliter la détoxicification d'acides biliaires, notamment chez des patients présentant une cholestase. Beaucoup d'autres RNs sont néanmoins impliqués dans la régulation de la détoxicification des acides biliaires (et des xénobiotiques qui empruntent globalement des voies identiques) et des interactions entre ces voies de signalisation peuvent donc être anticipées (pour une synthèse, voir Trottier et al., 2006). Enfin, PPAR $\alpha$  semble également impliqué dans la régulation de transporteurs qui jouent des rôles dans la sécrétion ou la réabsorption des acides biliaires. Les agonistes de PPAR $\alpha$  et le jeûne induisent de manière PPAR $\alpha$ -dépendante l'expression de Mdr2 (ou Abcb4) ce qui se traduit par un accroissement de la capacité de sécrétion des phospholipides dans la bile (Kok et al., 2003a; Kok et al., 2003b). Chez l'homme, un PPRE fonctionnel a également été identifié dans le promoteur du transporteur Slc10a2 (ou ASBT pour apical sodium-dependent bile salt transporter) qui joue un rôle essentiel dans la réabsorption intestinale des sels biliaires (Jung et al., 2002).

L'ensemble de ces données indiquent que PPAR $\alpha$  joue un rôle dans le métabolisme des acides biliaires qui constitue un maillon essentiel de l'élimination du cholestérol et de l'absorption des lipides au niveau intestinal. De plus, les modulations que PPAR $\alpha$  opère sur le métabolisme des acides biliaires pourraient être à la base d'interactions avec le récepteur FXR qui fonctionne comme un senseur des acides biliaires et avec le récepteur PXR qui est également activé par certains acides biliaires.

#### *BIOSYNTHÈSE DES LIPIDES :*

Alors que la plupart des gènes cibles de PPAR $\alpha$  sont impliqués dans le catabolisme des lipides, quelques gènes impliqués dans la biosynthèse des acides gras voient leurs expressions

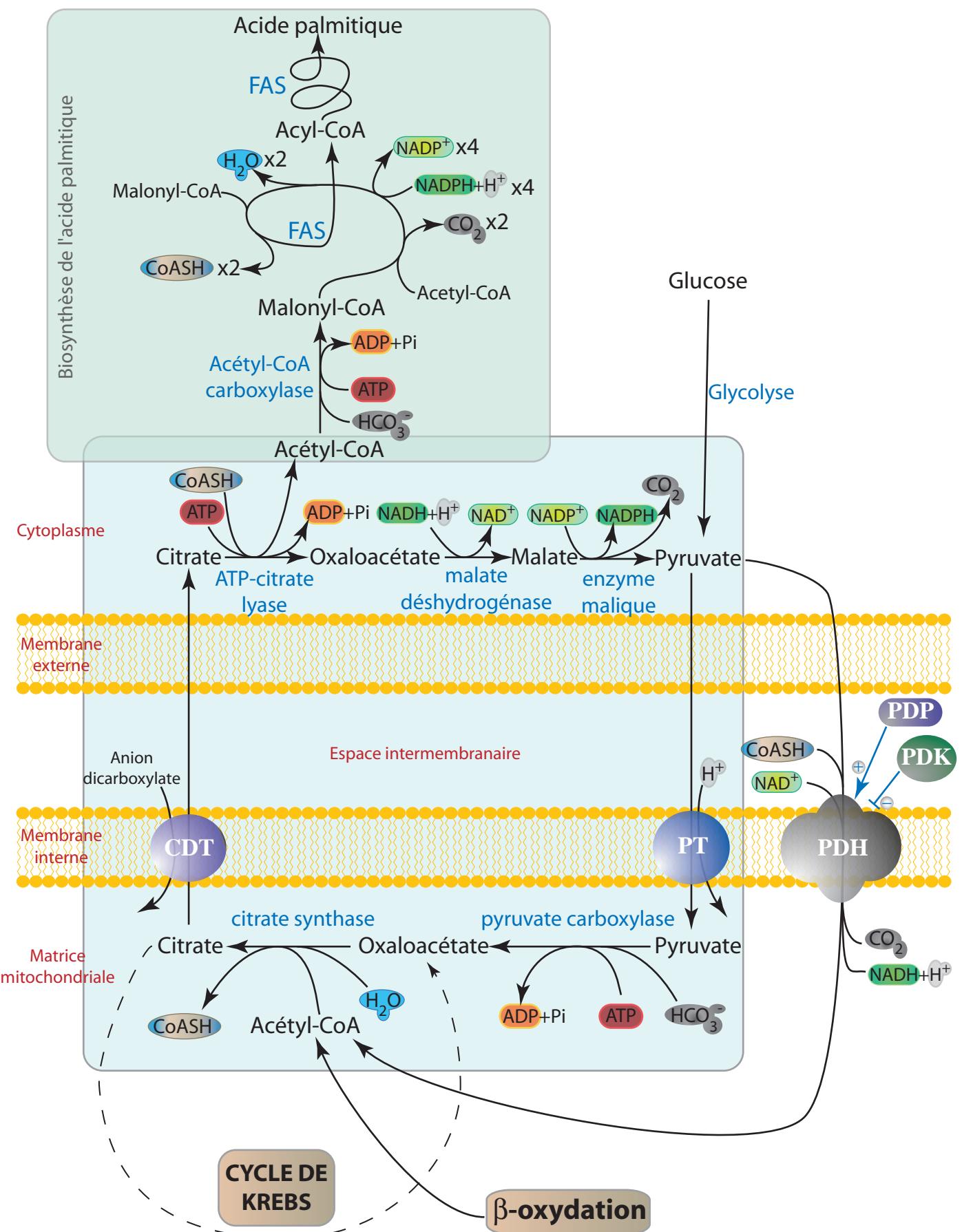
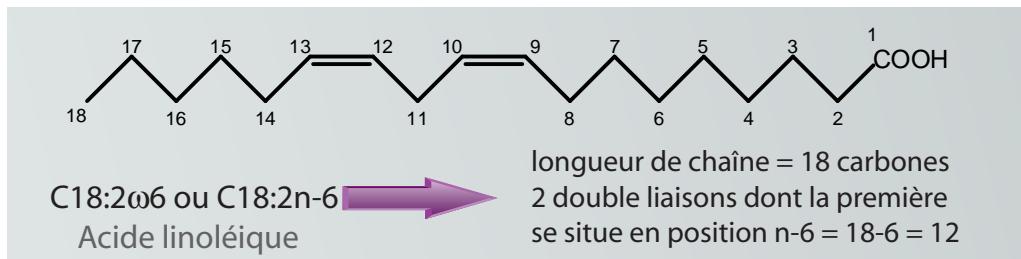
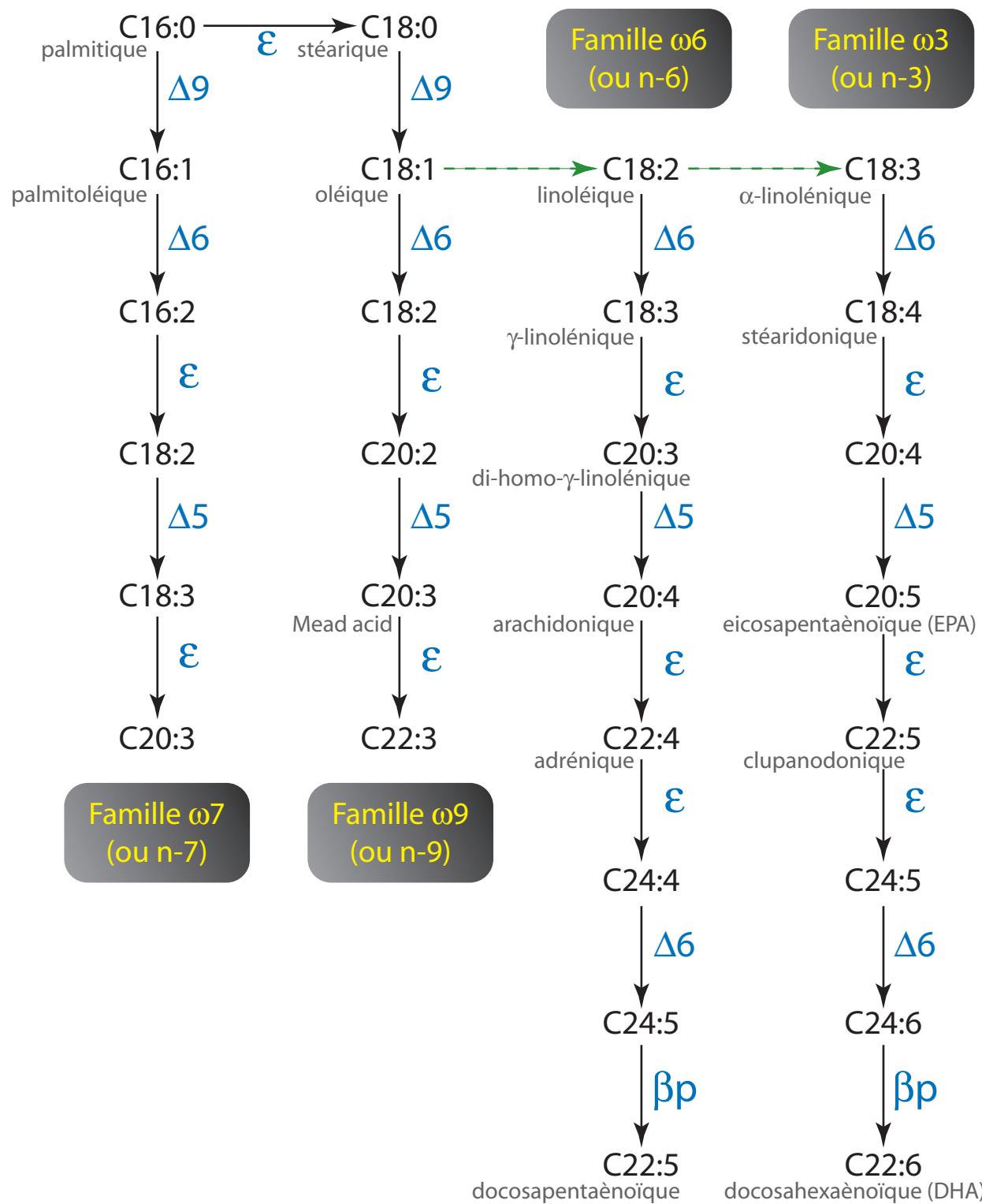


Figure 31 : Biosynthèse des acides gras et origine de l'acétyl-CoA cytosolique

L'origine de l'acétyl-CoA est la  $\beta$ -oxydation des acides gras et la décarboxylation du pyruvate via la PDH (à droite). Un système de navette permet à l'Acétyl-CoA de sortir de la mitochondrie (voir encadré du bas) et d'alimenter ainsi la voie de biosynthèse de l'acide palmitique (voie simplifiée dans l'encadré du haut). CDT : citrate/dicarboxylate transporter, FAS : fatty acid synthase, PT : pyruvate transporter, PDH : pyruvate déshydrogénase, PDK : PDH kinase, PDP : pyruvate phosphatase.

également modulées par PPAR $\alpha$  de manière directe ou indirecte. La lipogénèse désigne les processus biochimiques de synthèse des lipides en général mais le terme est souvent utilisé pour désigner la synthèse de l'acide palmitique à partir de l'acétyl-CoA qui constitue le point de départ de la synthèse de la plupart des autres acides gras. Pour la synthèse de cholestérol, on parlera de cholestérogenèse. Le point de départ de la synthèse des acides gras est l'acétyl-CoA, principalement produit dans la mitochondrie par la  $\beta$ -oxydation des AG et par la décarboxylation oxydative du pyruvate provenant de la glycolyse (complexe pyruvate déshydrogénase). L'acétyl-CoA sort de la mitochondrie par condensation en citrate (première étape du cycle de Krebs), pour lequel il existe un transporteur membranaire (voir [Figure 31](#)). Dans le cytoplasme, l'ATP-citrate lyase reconvertisit le citrate en acétyl-CoA. L'activité ATP-citrate lyase ne semble pas modulée de manière significative par les PPs (Aoyama et al., 1998). La biosynthèse de l'acide palmitique commence par la carboxylation de l'acétyl-CoA en malonyl-CoA, une réaction catalysée par les acétyl-CoA carboxylases (Acc1 et Acc2 ou Acaca et Acacb, [Figure 31](#)). L'activité acétyl-CoA carboxylase n'est pas altérée chez les souris PPAR $\alpha$ -/- et semble légèrement induite de manière indépendante de PPAR $\alpha$  par l'administration d'un PP (Aoyama et al., 1998). La fatty acid synthase (FAS ou Fasn) est un complexe multienzymatique (EC 2.3.1.85/38/39/41, EC 1.1.1.100, EC 4.2.1.61, EC 1.3.1.10, EC 3.1.2.14) permettant la synthèse de l'acide palmitique à partir de l'acétyl-CoA et du malonyl-CoA. Une autre protéine, nommée S14 (ou Thrsp pour thyroid hormone responsive SPOT14 homolog) qui semble nécessaire à la lipogénèse dans la glande mammaire (Zhu et al., 2005) mais pas dans le foie (Zhu et al., 2001a), a été largement étudiée comme modèle pour l'étude de la régulation transcriptionnelle des enzymes de la lipogénèse. L'étude de souris déficientes en S14 (Zhu et al., 2005) a permis de découvrir l'existence d'un autre gène : S14-related qui semble coder une protéine appartenant à la même famille que S14 mais qui, contrairement à S14, est exprimée chez de nombreuses espèces non-mammifères comme la grenouille, le poulet ou le poisson. Ce nouveau gène pourrait être impliqué dans la lipogénèse au niveau du foie et du tissu adipeux (Zhu et al., 2005) où il est bien exprimé (pour une synthèse, voir LaFave et al., 2006). Bien que la transcription de FAS et S14 soit régulée négativement suite à l'ingestion d'acides gras polyinsaturés (AGPI) qui constituent des ligands de PPAR $\alpha$ , ces mécanismes sont indépendants de PPAR $\alpha$ , puisqu'ils sont également observés chez les souris PPAR $\alpha$ -/- (Ren et al., 1996; Ren et al., 1997). L'**enzyme malique** (EM, EC 1.1.1.40) catalyse dans le cytosol la décarboxylation oxydative du malate en pyruvate et produit du NADPH utilisé pour la biosynthèse de l'acide palmitique par la FAS.

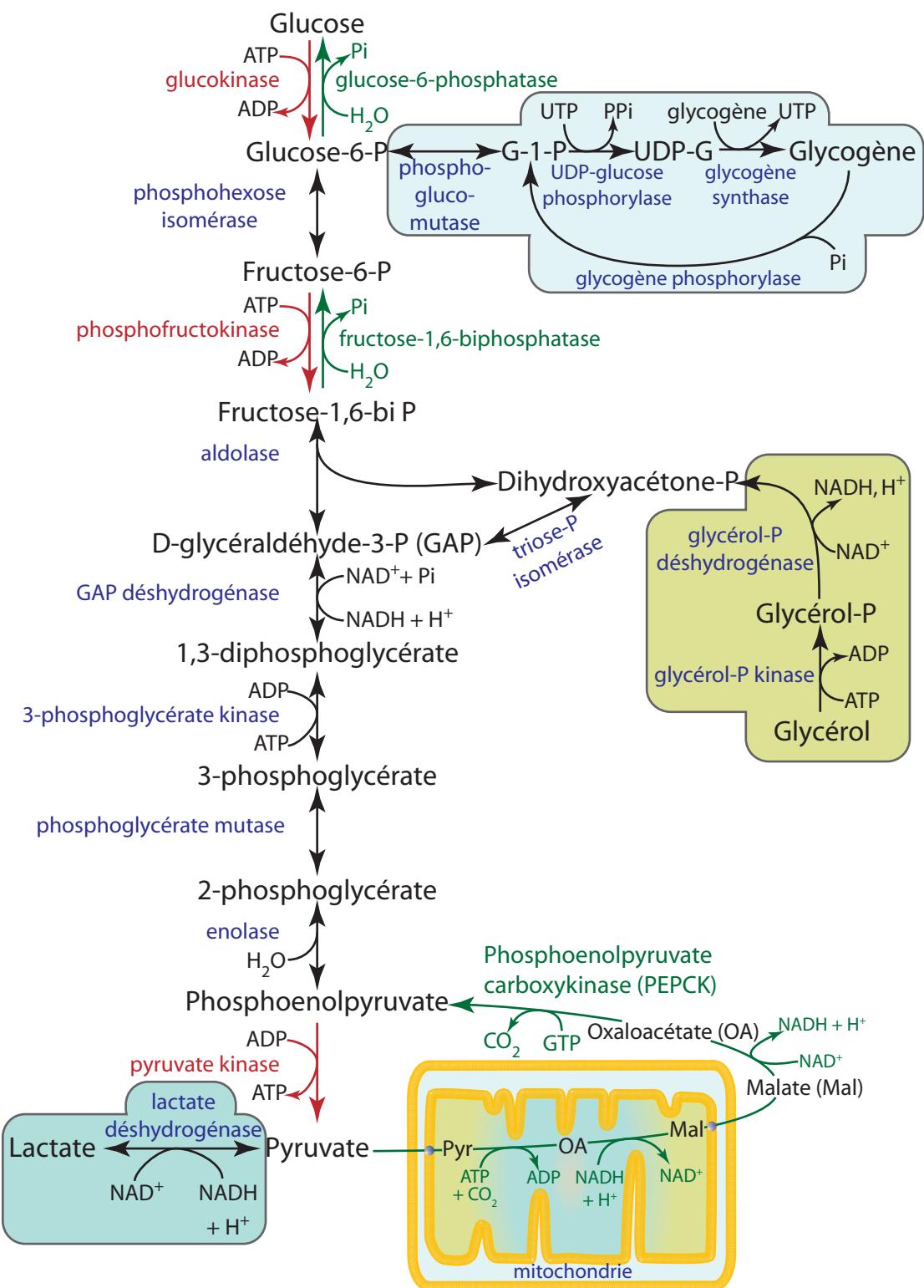
**A.****B.****Figure 32 : Biosynthèse des acides gras longs polyinsaturés**

A. Nomenclature des acides gras. B. Voie de biosynthèse des acides gras longs polyinsaturés. Les flèches vertes représentent des réactions que seules les plantes peuvent accomplir, ce qui rend les acides linoléique (C18:2 $\omega$ 6) et  $\alpha$ -linoléique (C18:3 $\omega$ 3) essentiels chez les animaux car ils doivent être apportés par l'alimentation. Les noms courants de certains acides gras sont indiqués sous leurs formules.  $\Delta$ 9,  $\Delta$ 6 et  $\Delta$ 5 :  $\Delta$ 9-,  $\Delta$ 6- et  $\Delta$ 5-désaturases,  $\epsilon$  : élongation,  $\beta$ p : un cycle de  $\beta$ -oxydation peroxysomale

L'enzyme malique est induite par les PPs via la fixation de PPAR $\alpha$  sur un PPRE situé dans le promoteur du gène codant cette enzyme (Castelein et al., 1994; IJpenberg et al., 1997). L'activité de l'enzyme malique est réduite constitutivement d'environ 50% chez les souris PPAR $\alpha$ -/- et son induction par un fibrate n'est pas observée chez ces souris (Aoyama et al., 1998). La **glucose-6-phosphate déshydrogénase** (G6PDH) fournit également du NADPH utilisé pour la biosynthèse des acides gras. Son activité semble également induite par un fibrate de manière PPAR $\alpha$ -dépendante (Aoyama et al., 1998). Le NADPH constitue également un donneur d'électron clef dans la défense contre les stress oxydants illustrant donc également le rôle de PPAR $\alpha$  dans ce domaine. Alors que le rôle de PPAR $\alpha$  dans la régulation des gènes clefs de la lipogénèse apparaît donc très limité, l'expression de la plupart de ces gènes ainsi que ceux impliqués dans la cholestérogenèse (HMG-CoA synthase cytosolique, HMG-CoA réductase et squalène synthase notamment) est sous le contrôle des facteurs de transcription de la famille SREBP (pour une synthèse sur les SREBP, voir Eberle et al., 2004; et sur leurs gènes cibles, voir Horton et al., 2002). Les SREBP appartiennent à la famille des facteurs de transcription bHLH-LZ pour basic-helix-loop-helix-leucine-zipper et sont synthétisés sous la forme de précurseurs fixés à la membrane du réticulum endoplasmique. Sous l'effet d'une augmentation de la concentration en stérols, ils subissent un clivage protéolytique en deux étapes (via les protéases de site 1 et 2 S1P et S2P), s'associent sous la forme de dimères et subissent une translocation dans le noyau où ils se fixent sur des éléments de réponse de leurs gènes cibles, modulant ainsi leur transcription via le recrutement de corégulateurs. Il existe deux gènes codant des SREBP : Srebf1 et Srebf2. Srebf2 ne produit que la protéine SREBP-2 alors que Srebf1 produit les protéines SREBP-1a et SREBP-1c qui diffèrent par leur premier exon (exon 1a et 1c) du fait de l'utilisation de sites alternatifs d'initiation de la transcription. SREBP-1a est un régulateur transcriptionnel plus puissant que SREBP-1c du fait d'un domaine de transactivation N-terminal plus long mais SREBP-1c est l'isoforme la plus exprimée dans de nombreux tissus comme le foie, le tissu adipeux blanc ou le muscle. SREBP-1a apparaît comme un puissant régulateur de l'expression des gènes de la lipogénèse et de la cholestérogenèse alors que SREBP-1c et SREBP-2 semblent être des régulateurs plus spécifiques de la lipogénèse et de la cholestérogenèse respectivement. La régulation de l'activité des SREBP se situe au niveau transcriptionnel (l'expression de SREBP-1c est notamment augmentée par l'insuline via les PI(3) kinases et par LXRx via deux éléments de réponse de type DR4), au niveau de leur clivage protéolytique (notamment via les protéines SCAP pour SREBP cleavage activating protein et les protéines INSIG pour

insulin-induced genes) ainsi qu'au niveau de modifications post-traductionnelles (phosphorylations par les MAP kinases, ubiquitination, acétylation par des corégulateurs comme p300, et sumoylation). En cohérence avec sa fonction de régulateur clef de la lipogénèse, l'expression de SREBP-1c est diminuée par le jeûne et induite par un régime riche en sucres. L'activité de SREBP-1a et de SREBP-2 semble en revanche plutôt régulée au niveau de leur clivage protéolytique. Les liens entre la voie de signalisation PPAR $\alpha$  et les voies SREBPs sont encore mal connus, pourtant, il apparaît assez cohérent que les régulations du catabolisme des lipides (PPAR $\alpha$ ) d'une part et de la lipogénèse (SREBP-1c) d'autre part soient relativement coordonnées. L'expression de certains gènes cibles des SREBP comme FAS ou l'HMG-CoA réductase suit un rythme circadien avec une expression plus élevée la nuit, alors que les souris s'alimentent par rapport au jour où leur activité locomotrice et alimentaire est réduite. Il a été observé que la régulation de plusieurs de ces gènes était régulée de manière anormale au cours du rythme circadien chez les souris PPAR $\alpha$ -/- (Knight et al., 2005; Patel et al., 2001). L'expression de la FAS est en particulier réduite chez les souris PPAR $\alpha$ -/- (Martin et al., 2007; Patel et al., 2001) ainsi que le niveau basal d'expression de SREBP-1c (Knight et al., 2005). Cependant, il n'est pas impossible que ces effets soient indirectement liés à l'invalidation de PPAR $\alpha$  et à ses effets sur la quantité et le profil des acides gras présents dans le foie des souris. Nous avons notamment observé que les souris PPAR $\alpha$ -/- accumulaient préférentiellement les précurseurs essentiels des acides gras polyinsaturés des familles  $\omega$ 3 et  $\omega$ 6 (acides linoléique et  $\alpha$ -linolénique, voir [Figure 32](#) et Martin et al., 2007), or ces acides gras polyinsaturés semblent capables d'inhiber l'expression (Yoshikawa et al., 2002) et le clivage de SREBP-1c (Sampath and Ntambi, 2005). De plus, l'accumulation progressive de lipides dans le foie des souris mâles PPAR $\alpha$ -/- (Costet et al., 1998) est également susceptible d'influer sur la maturation des SREBPs. Un traitement par le Wy14643 semble se traduire par un accroissement du clivage de SREBP-1c de manière dépendante de PPAR $\alpha$ , ce qui conduit une induction de l'expression de FAS et ACC et à un accroissement de la lipogénèse uniquement chez les souris de type sauvage (Knight et al., 2005). Des effets relativement similaires ont également été rapporté sur la cholestérogenèse (Le Jossic-Corcos et al., 2004) et dans les deux cas, il a été proposé que ces régulations soient dues à des phénomènes de compensation des effets des fibrates soit sur les quantités et les profils d'acides gras au niveau hépatique, soit sur la prolifération des peroxysomes qui nécessite la formation d'une importante quantité de membranes. La première hypothèse apparaît d'autant plus vraisemblable que dans certaines situations, comme par exemple chez

des souris déficientes pour le récepteur aux LDL alimentées avec un régime riche en graisse pour constituer un modèle de syndrome métabolique, l'expression des gènes de la lipogénèse est au contraire abaissée par un traitement par un fibrate (Srivastava et al., 2006), ce qui peut s'expliquer au moins en partie par les effets bénéfiques du traitement sur le profil lipidique (diminution des triglycérides et du cholestérol circulant et diminution de la masse de tissu adipeux notamment). Si la seconde hypothèse (production de lipides en compensation de ceux utilisés pour les membranes des peroxysomes qui prolifèrent) apparaît difficile à évaluer, les effets des fibrates sur les profils d'acides gras hépatiques commencent en revanche à être mieux compris. Les fibrates activent, via la fixation de PPAR $\alpha$  sur un PPRE fonctionnel, la transcription de la  $\Delta$ 9-désaturase (ou SCD1 pour stearoyl-CoA désaturase, EC 1.14.99.5, Miller and Ntambi, 1996) qui catalyse la désaturation en position  $\omega$ 9 des acides palmitique et stéarique (voir [Figure 32](#)) ce qui conduit notamment à un accroissement du rapport C18:1 $\omega$ 9/C18:0 remarqué par plusieurs équipes (Guillou et al., 2002; Knight et al., 2005). De plus, les fibrates régulent également de manière PPAR $\alpha$ -dépendante l'expression des  $\Delta$ 6- et  $\Delta$ 5-désaturases (ou Fads2 et Fads1 respectivement, Guillou et al., 2002; Tang et al., 2003) et un PPRE fonctionnel a été identifié dans le promoteur humain de la  $\Delta$ 6-désaturase (Tang et al., 2003). Bien que sa fonctionnalité n'ait, à ma connaissance, pas encore été testée, j'ai choisi de représenter en [Table 8](#) la séquence du promoteur murin de Fads2 qui se rapproche le plus du PPRE humain fonctionnel et se trouve localisée dans une zone bien conservée du promoteur. Ces deux désaturases sont impliquées dans la biosynthèse des acides gras long polyinsaturés dont les étapes limitantes sont justement les étapes de désaturation (voir [Figure 32](#)). Leurs affinités pour leurs divers substrats suit en général l'ordre suivant : substrat  $\omega$ 3 > substrat  $\omega$ 6 > substrat  $\omega$ 9. Ces régulations par PPAR $\alpha$  des trois  $\Delta$ -désaturases clefs de la synthèse des acides gras polyinsaturés semblent directes, contrairement aux effets que nous avons évoqués sur d'autres enzymes de la lipogénèse ou de la cholestérogenèse. Ces trois désaturases sont de plus également des gènes cibles de SREBP régulés négativement au cours d'un jeûne, alors que PPAR $\alpha$  est activé, probablement par l'afflux d'acides gras provenant du tissu adipeux (Matsuzaka et al., 2002; Tabor et al., 1999). Nous avons proposé que ces régulations directes par PPAR $\alpha$  aient été conservées pour permettre de contrebalancer les effets de l'induction des voies de catabolisme des acides gras par les PPs et maintenir ainsi les niveaux des acides gras polyinsaturés dont les fonctions sont multiples et souvent indispensables (Guillou et al., 2002). Mais il est également envisageable que les effets de ces régulations soient en partie responsables des modulations observées sur les gènes cibles des



BILAN :

### Glycolyse



Figure 33 : Glycolyse et néoglucogenèse

Les voies empruntées par certains substrats de la néoglucogenèse (lactate, glycogène et glycérat) sont également indiquées dans des encadrés de couleur. Les voies irréversibles de la glycolyse sont indiquées en rouge (La glucokinase est spécifique du foie et est une hexokinase dans les autres tissus). Les réactions visant à contourner ces réactions irréversibles au cours de la néoglucogenèse sont indiquées en vert. Les bilans de la glycolyse et de la néoglucogenèse à partir du pyruvate sont également illustrés. G : glucose, P : phosphate, Pi : phosphate inorganique, PPi : pyrophosphate, Pyr : pyruvate

voies de signalisation des SREBPs lors de l'activation de PPAR $\alpha$  par des traitements de type pharmacologique. Des investigations complémentaires, comprenant notamment des dosages de la composition en acides gras (et en cholestérol) au niveau des membranes des différents compartiments cellulaires ainsi que des études de cinétique de ces différentes régulations, apparaissent nécessaires pour confirmer ou non ces hypothèses. Plus récemment, les effets des activateurs de PPAR $\alpha$  ont commencés à être également révélés sur l'expression des élongases (Elovl pour *elongation of very long chain fatty acids*) qui interviennent également dans la biosynthèse des acides gras longs polyinsaturés (voir [Figure 32](#)). Le Wy14643 augmente l'expression des élongases Elovl5 (la principale isoforme exprimée dans le foie) et Elovl6 dans le foie ainsi que l'activité élongase hépatique étudiée sur un panel d'acides gras saturés et monoinsaturés (Wang et al., 2005b). De plus, l'induction des Elovl5 et Elovl6 est dépendante de l'expression de PPAR $\alpha$  (Wang et al., 2006b). Avec nos puces à ADN, nous avons également observé l'induction PPAR $\alpha$ -dépendante de ces deux gènes suite à un traitement par le fénofibrate (données non présentées).

Ces données illustrent bien les rôles principaux joués par PPAR $\alpha$  sur la biosynthèse des acides gras et du cholestérol. PPAR $\alpha$  agit directement sur la biosynthèse des acides gras longs polyinsaturés, en même temps qu'il augmente leur dégradation par ses effets sur la  $\beta$ -oxydation et sur les enoyl-CoA isomérasées nécessaires à la dégradation de ces acides gras particuliers. Par ailleurs, en réponse à l'induction du catabolisme des lipides et peut-être aussi à la prolifération des peroxysomes, la lipogénèse et la cholestérogenèse peuvent être augmentées sous certaines conditions qui restent à préciser mais qui impliqueraient sans doute en particulier le statut physiologique (âge, sexe) et nutritionnel (alimentés/à jeun) des animaux. Naturellement, ces différentes observations sont aussi largement dépendantes de la composition de l'aliment qui est donné aux animaux étudiés, dans la mesure où celle-ci est susceptible d'influer également sur d'autres voies de signalisation (SREBP, LXR, FXR, ChREBP pour carbohydrate response element binding protein, etc.) qui contrôlent également ces mêmes voies métaboliques. Enfin, PPAR $\alpha$  a récemment été identifié comme un régulateur important de l'horloge périphérique du foie via sa régulation du facteur de transcription Bmal1/Arntl (Canaple et al., 2006) qui est lui-même impliqué dans la régulation de l'expression circadienne de nombreux gènes du métabolisme des lipides et des glucides. Ces observations suggèrent notamment que PPAR $\alpha$  pourrait être impliqué dans la régulation circadienne du métabolisme de manière générale et que des études complémentaires prenant

en compte ces phénomènes rythmiques sont nécessaires pour mieux appréhender ses rôles physiologiques.

En conclusion de ce paragraphe, PPAR $\alpha$  apparaît comme un régulateur majeur du métabolisme des lipides dont il régule de multiples aspects de manière directe ou indirecte. Ses actions ont été largement étudiées aux niveaux hépatique et cardiaque, mais la modulation des gènes cibles de PPAR $\alpha$  est au moins aussi forte dans d'autres organes comme le rein ou la muqueuse intestinale qui ont été pour l'instant moins étudiés vis-à-vis de ces régulations. Nous allons maintenant voir que PPAR $\alpha$  est également capable de réguler d'autres aspects du métabolisme et module également le métabolisme des glucides et des acides aminés.

#### **II.4.4. Régulation du métabolisme des glucides et des acides aminés par PPAR $\alpha$**

La glycémie est un paramètre fortement contrôlé par l'organisme malgré des apports nutritionnels discontinus prenant la forme d'alternance entre prise alimentaire et jeûne. Une forte hypoglycémie peut entraîner des dysfonctions au niveau cérébral et conduire au coma ou à la mort. À l'opposé, l'hyperglycémie est une complication sérieuse du diabète qui, dans les cas aigus, s'accompagne d'une cétoacidose et d'une hyperosmolarité hyperglycémique pouvant être fatale en raison de la déshydratation et de déséquilibres des électrolytes et qui, dans les cas chroniques, conduit à des neuropathies et à des atteintes vasculaires. L'homéostasie du glucose est principalement maintenue par deux hormones produites par le pancréas : l'insuline et le glucagon. La synthèse et la sécrétion d'insuline par les cellules  $\beta$  des îlots de Langerhans est stimulée par des niveaux élevés de glucose, observés en particulier après une prise alimentaire, via divers mécanismes impliquant une induction transcriptionnelle de l'expression de l'insuline et une stimulation de sa sécrétion liée à l'augmentation du ratio ATP/ADP (pour une synthèse, voir Desvergne et al., 2006). Les principaux effets de l'insuline sont d'accroître le captage et l'utilisation (glycolyse, voir [Figure 33](#)) du glucose par les tissus périphériques (principalement les muscles et le tissu adipeux) et d'augmenter le stockage d'énergie en stimulant la synthèse de glycogène (glycogénogenèse) par le foie et les muscles et la synthèse d'acides gras (notamment via l'induction de SREBP que nous avons évoquée) par le foie et le tissu adipeux. À l'opposé, la sécrétion de glucagon par les cellules  $\alpha$  des îlots de Langerhans est provoquée par une baisse de la glycémie. Elle est également sensible aux taux circulants d'acides gras et de corps cétoniques et est inhibée par l'insuline et la somatostatine. Son principal effet est d'augmenter la production hépatique de glucose en stimulant la néoglucogenèse ([Figure 33](#)) et la

glycogénolyse ce qui conduit normalement à un rétablissement de la glycémie. Ces différentes régulations provoquées par l’insuline et le glucagon sur les tissus périphériques font intervenir de nombreux facteurs de transcriptions principalement par le biais de phosphorylations initiées au niveau des récepteurs membranaires de ces deux hormones (pour une synthèse, voir Desvergne et al., 2006). Cependant, il est particulièrement difficile de dissocier les effets de ces hormones des effets provoqués par les variations intracellulaires de multiples métabolites (dont le glucose, les acides gras et l’acétyl-CoA, les cofacteurs NADH et l’ATP.) qui elles aussi affectent les fonctions de multiples facteurs de transcription et régulent l’activité de nombreuses enzymes.

Le rôle de PPAR $\alpha$  dans le métabolisme du glucose a commencé à être évalué lorsqu’il a été constaté que les souris PPAR $\alpha$ -/- présentaient une forte hypoglycémie en situation de jeûne (Kersten et al., 1999; Leone et al., 1999). Des études de métabolomique réalisées sur les souris PPAR $\alpha$ -/- ont révélées que cette hypoglycémie se traduisait notamment par des niveaux de glucose faibles dans de nombreux tissus comme le cœur, le muscle, le foie et le tissu adipeux (Atherton et al., 2006). Sous un régime normal, ces souris répondent néanmoins normalement au test de tolérance à l’insuline (Guerre-Millo et al., 2001) et au test de tolérance au glucose (Guerre-Millo et al., 2001; Kersten et al., 1999) et présentent un niveau d’insuline relativement normal, en lien avec leurs niveaux circulants de glucose (Xu et al., 2002b). Les données de la littérature indiquent que l’absence de PPAR $\alpha$  se traduit par une absence de modification (Kersten et al., 1999) voire une amélioration de la sensibilité à l’insuline (Guerre-Millo et al., 2001; Knauf et al., 2006; Patsouris et al., 2004; Tordjman et al., 2001), notamment dans des modèles favorisant l’insulino-résistance (régime riche en graisse, déficience en apolipoprotéine E). Un mécanisme a même été proposé pour expliquer le rôle de PPAR $\alpha$  dans l’insulino-résistance (Koo et al., 2004). PPAR $\alpha$  et le coactivateur PGC-1 $\alpha$ , dont l’expression est induite au cours du jeûne via l’activation du facteur de transcription CREB (cAMP response element binding protein) par le glucagon, augmenteraient l’expression de TRB-3 (tribbles homolog 3) qui se fixe sur Akt/PKB (protéine kinase B), un effecteur d’aval du récepteur à l’insuline, et prévient son activation par phosphorylation, inhibant ainsi l’action de l’insuline au niveau hépatique. Bien que cela puisse paraître contradictoire, des données concluantes indiquent cependant que plusieurs fibrates améliorent l’insulino-sensibilité dans des modèles d’insulino-résistance d’origine génétique où alimentaire (Guerre-Millo et al., 2000; Srivastava et al., 2006; Ye et al., 2001). Le principal mécanisme proposé est une augmentation du catabolisme des lipides au niveau musculaire et hépatique,

conduisant à réduire globalement l'accumulation de graisses et à favoriser ainsi l'insulino-sensibilité. Dans la mesure où le niveau de glucose plasmatique est normalement maintenu au cours du jeûne par l'action combinée de la néoglucogenèse et de la glycogénolyse (Figure 33) au niveau hépatique, les rôles de PPAR $\alpha$  ont particulièrement été étudiés vis-à-vis de ces deux voies.

Tout d'abord, rappelons que la  $\beta$ -oxydation, normalement élevée en situation de jeûne, fournit un certain nombre de cofacteurs (NADH), de substrats (acétyl-CoA) et de l'énergie (ATP) qui sont nécessaires à la néoglucogenèse (Figure 33) mais plus généralement la  $\beta$ -oxydation permet à l'organisme d'utiliser les lipides comme source d'énergie et d'épargner ainsi le glucose. La similitude entre le phénotype des souris PPAR $\alpha$ -/- en situation de jeûne (hypoglycémie et hypothermie en particulier) et celui de patients souffrant de déficiences génétiques touchant certaines enzymes de la  $\beta$ -oxydation a d'ailleurs été rapidement remarquée (Kersten et al., 1999). Une autre hypothèse qui a été avancée pour expliquer l'hypoglycémie des souris PPAR $\alpha$ -/- est une diminution des réserves de glycogène au niveau hépatique. Cependant, alors que certains groupes observent bien une diminution des réserves de glycogène hépatique chez les souris PPAR $\alpha$ -/- en situation alimentées (Kersten et al., 1999; Xu et al., 2002b), d'autres groupes n'observent aucune différence entre les souris des deux génotypes (Hashimoto et al., 2000; Sugden et al., 2002). Des mesures effectuées sur les souris PPAR $\alpha$ -/- entretenues dans notre animalerie montrent même une augmentation des réserves de glycogène hépatique (P. Costet et T. Pineau, observations non publiées). Récemment, il a été montré que la glycogène synthase 2 (Gys-2) était un gène cible des PPARs, régulée par PPAR $\alpha$  dans le foie via un PPRE situé dans son premier intron (Mandard et al., 2007). L'expression basale de Gys-2 semble peu altérée chez les souris PPAR $\alpha$ -/- ainsi que sa surprenante induction au cours du jeûne mais en revanche son niveau d'expression semble plus faible chez les souris PPAR $\alpha$ -/- à la suite d'une réalimentation (Mandard et al., 2007). Il semble que la diminution des réserves de glycogène hépatique chez les souris PPAR $\alpha$ -/- soit principalement observée lors d'une réalimentation suivant une période de jeûne et il ne serait donc pas intéressant d'évaluer les variations des réserves de glycogène au cours du cycle jour/nuit chez ces souris. Globalement, ces résultats indiquent néanmoins que la taille des réserves en glycogène a peu de chance d'être responsable pour une large part de l'hypoglycémie des souris PPAR $\alpha$ -/- en situation de jeûne.

Une autre voie susceptible d'être altérée chez les souris PPAR $\alpha$ -/- est la néoglucogenèse. Les principaux substrats de la néoglucogenèse sont le glycérol provenant de

la lipolyse du tissu adipeux, les acides aminés glucoformateurs dont l'alanine provenant des muscles et le lactate provenant de la glycolyse musculaire (Figure 33). L'hypothèse d'une néoglucogenèse altérée chez les souris PPAR $\alpha$ -/- apparaît difficile à concilier avec l'observation d'une induction normale de la phosphoénolpyruvate carboxykinase (PEPCK, Kersten et al., 1999; Le May et al., 2000; Patsouris et al., 2004; Xu et al., 2002b) et d'une production hépatique de glucose accrue chez ces souris en situation de jeûne (Knauf et al., 2006; Xu et al., 2002b), cette dernière observation n'étant cependant pas reproduite par tous les auteurs (Bandsma et al., 2004; Patsouris et al., 2004). De plus, PPAR $\alpha$  régule l'expression de nombreux gènes impliqués dans le métabolisme hépatique du glycérol (glycérol kinase, glycérol 3-phosphate déshydrogénases mitochondriale et cytosolique, aquaporines 3 et 9) et deux PPREs ont été identifiés dans le promoteur de la glycérol-3-phosphate déshydrogénase cytosolique (Patsouris et al., 2004). L'induction de ces gènes au cours du jeûne, nécessaire à l'utilisation du glycérol pour la néoglucogenèse, est absente chez les souris PPAR $\alpha$ -/- (Patsouris et al., 2004), suggérant qu'une diminution combinée des substrats (glycérol) et des cofacteurs (ATP, NADH) provenant normalement de la  $\beta$ -oxydation des acides gras pourraient être responsable pour une large part d'une altération de la néoglucogenèse chez les souris PPAR $\alpha$ -/. Une autre hypothèse permettant, elle, de concilier l'absence d'effet net sur la néoglucogenèse serait que celle-ci soit tant bien que mal maintenue chez les souris PPAR $\alpha$ -/- grâce à l'utilisation des acides aminés glucoformateurs provenant notamment du catabolisme des protéines musculaires et, moins probablement (voir Knauf et al., 2006; voir Mandard et al., 2004; Xu et al., 2002b), du lactate et qu'en parallèle l'utilisation de glucose par les tissus périphériques, à commencer par les muscles, soit accrue, conduisant ainsi à une hypoglycémie. De fait, PPAR $\alpha$  a été identifié comme un régulateur important du catabolisme des acides aminés (Edgar et al., 1998; Kersten et al., 2001) via la régulation, principalement négative, de nombreux gènes impliqués par exemple dans les transaminations, les désaminations, l'oxydation des  $\alpha$ -céto acides et le cycle de l'urée. Bien que les données biochimiques soient limitées et que l'expression de ces gènes n'ait pas été étudiée dans le muscle des souris PPAR $\alpha$ -/, il semble que le catabolisme des protéines soit accru chez les souris PPAR $\alpha$ -/- en situation de jeûne comme l'illustre la nette augmentation de l'urée dans le sang des souris PPAR $\alpha$ -/- au cours du jeûne (voir Chapitre VI). Les effets du Wy14643 sur l'expression de ces gènes suggèrent que leur régulation par PPAR $\alpha$  est directe et ne constitue donc pas un phénomène compensatoire lié à l'absence d'induction du catabolisme des lipides (Kersten et al., 2001). A ma connaissance, les mécanismes de la régulation de ces gènes n'ont

pas été décrits mais pourraient faire intervenir des interactions avec les voies de signalisation de HNF-4 $\alpha$  et/ou de C/EBP $\alpha$  (Mandard et al., 2004). Un autre rôle de PPAR $\alpha$  dans le métabolisme des acides aminés qui a été rapporté est son action négative sur la régulation par HNF4 de l' $\alpha$ -amino- $\beta$ -carboxymuconate- $\epsilon$ -semialdéhyde décarboxylase (ACMDS, EC 4.1.1.45) impliquée dans la voie de biosynthèse du NAD à partir du tryptophane (Shin et al., 2006). Bien que les conséquences physiologiques de cette régulation ne soient pas complètement élucidées, elle pourrait expliquer pour une large part la présence accrue de certains métabolites du NAD dans les urines de souris de type sauvage par rapport aux souris PPAR $\alpha$ -/- (Zhen et al., 2007). Ces données renforcent l'idée que PPAR $\alpha$  est également un régulateur important du métabolisme des acides aminés mais des études complémentaires apparaissent nécessaires, notamment au niveau musculaire, pour préciser les contours et les mécanismes de son action. En situation de jeûne, le muscle s'adapte normalement à la nécessité d'épargner le glucose en favorisant l'oxydation des acides gras par rapport au glucose. Au cours du jeûne, la  $\beta$ -oxydation au niveau musculaire chez les souris PPAR $\alpha$ -/- est peu altérée (environ 30%) probablement en raison d'une compensation par PPAR $\beta$  (Muonio et al., 2002). Au cours de ces changements métaboliques, le complexe pyruvate déshydrogénase (PDH, [Figure 31](#)) joue un rôle essentiel. L'activité du complexe PDH converti de manière irréversible le pyruvate cytosolique en acétyl-CoA mitochondrial qui, en subissant le cycle de Krebs, conduit à l'oxydation complète du glucose en CO<sub>2</sub> et H<sub>2</sub>O accompagné de production d'ATP. Ce complexe fait donc le lien entre la glycolyse et le cycle de Krebs dans les tissus qui nécessitent un apport important d'énergie et fournit de l'acétyl-CoA au tissu adipeux et au foie pour la synthèse des acides gras et la cétogenèse ([Figure 31](#)). En situation de jeûne, l'activité de ce complexe doit être inhibée dans les tissus qui ne nécessitent pas l'oxydation complète du glucose pour couvrir leurs besoins énergétiques et épargner ainsi le pyruvate pour l'amener au foie afin qu'il serve à la néoglucogenèse. PPAR $\alpha$  est un régulateur important de la pyruvate déshydrogénase kinase 4 (PDK4) qui phosphoryle le complexe PDH conduisant ainsi à son inactivation (Wu et al., 2001). Cependant, si PPAR $\alpha$  est indispensable à l'induction par le jeûne de l'expression de la PDK4 dans le rein (Sugden et al., 2001; Wu et al., 2001), son absence n'altère que modérément l'induction par le jeûne de la PDK4 dans le foie (Sugden et al., 2002), dans les muscles (Holness et al., 2002a) ou dans le cœur (Holness et al., 2002b; Wu et al., 2001). En l'absence de dosages systématiques de l'activité PDH, il reste malgré tout possible que ces faibles altérations conduisent à une utilisation accrue de glucose par les muscles qui participerait à l'hypoglycémie des souris PPAR $\alpha$ -/- à jeun.

Cependant, une étude récente suggère un autre mécanisme, indépendant d'un effet sur les muscles, mais lié lui aussi à une utilisation accrue de glucose par un tissu périphérique : le tissu adipeux (Knauf et al., 2006). Dans cette étude, les auteurs observent un captage de glucose 4 à 5 fois plus important dans le tissu adipeux blanc des souris PPAR $\alpha$ -/- qui est associé à une surexpression du transporteur de glucose Glut4. Leurs résultats suggèrent de plus, que c'est l'absence d'expression de PPAR $\alpha$  au niveau du cerveau qui serait responsable de ces différences d'utilisation du glucose au niveau périphérique. Ces résultats originaux ouvrent donc de nouvelles pistes d'investigations sur le rôle de PPAR $\alpha$  dans le cerveau vis-à-vis du contrôle du métabolisme énergétique par l'organisme.

L'ensemble de ces données indiquent que PPAR $\alpha$  constitue bien un régulateur important du métabolisme des glucides et des acides aminés. Alors que l'importance du cerveau dans la régulation du métabolisme des glucides au niveau des tissus périphériques commence à être décrit, il semble que le rôle de PPAR $\alpha$  dans ce contexte mérite largement d'être évalué (Knauf et al., 2006). De même, le contrôle du catabolisme des protéines au niveau musculaire est encore mal compris (Desvergne et al., 2006) et l'étude du rôle de PPAR $\alpha$  dans ce domaine pourrait permettre de préciser les voies de signalisations et les voies métaboliques impliquées respectivement dans l'épargne et le catabolisme protéique. Je pense que la génération de lignées murines présentant des invalidations tissu-spécifique de PPAR $\alpha$  pourrait faciliter la compréhension des contours précis de son action dans des tissus où le rôle de ce RN est encore peu ou mal compris comme le tissu adipeux, le cerveau ou le muscle.

## II.5. AUTRES FONCTIONS DE PPAR $\alpha$

Alors que PPAR $\alpha$  apparaît comme un puissant régulateur du métabolisme, jouant des rôles majeurs vis-à-vis des lipides, mais également des glucides et des acides aminés, d'autres fonctions de ce RN ont également fait l'objet d'un nombre conséquent d'études. C'est le cas en particulier du rôle de PPAR $\alpha$  dans la régulation du métabolisme des xénobiotiques, de l'inflammation au niveau vasculaire et de la réponse hépatique de phase aiguë.

### II.5.1. Régulation du métabolisme des xénobiotiques par PPAR $\alpha$

Dans l'organisme, les toxines, les drogues, les carcinogènes et les autres xénobiotiques sont éliminés par l'organisme via des séries de réactions souvent regroupées sous le terme de

biotransformations. Le métabolisme des xénobiotiques ou la biotransformation comprend généralement trois étapes :

- 1) la phase I est appelée la fonctionnalisation et consiste à introduire un groupement polaire (généralement un groupement hydroxy -OH) au sein de la molécule xénobiotique. Cette réaction est généralement catalysée par des cytochromes P450 (CYP chez le rat et l'homme et Cyp chez la souris), des hémoprotéines localisées le plus souvent dans la membrane du réticulum endoplasmique ou dans la membrane interne de la mitochondrie.
- 2) la phase II est appelée la phase de conjugaison et consiste à greffer le xénobiotique (via le groupement polaire et réactif qui lui a été ajouté) à une molécule polaire comme le glutathion ou l'acide glucuronique ou à lui fixer un groupement plus polaire comme un sulfate ou un acétyle. Ces réactions sont catalysées par diverses familles d'enzymes caractérisées par ce qu'elles fixent aux xénobiotiques : glutathion S-transférases (GSTs), UDP glucuronosyltransférases (UGTs), sulfotransférases (SULTs), N-acétyltransférases (NATs) ou cathécol O-méthyltransférase (COMT).
- 3) Enfin, au cours de la phase III, le xénobiotique conjugué est excrété par la bile ou les urines via des transporteurs spécifiques dont la plupart appartiennent à la famille des ABC transporteurs (ATP-binding cassette transporters)

Dans le cas, sans doute le plus courant, d'une exposition par voie orale, les principaux organes impliqués dans le métabolisme des xénobiotiques sont l'intestin, le foie et le rein jouant respectivement les rôles de première barrière et de site éventuel d'élimination, de deuxième barrière et d'organe central du métabolisme des xénobiotiques et enfin de site d'élimination par les urines. Il est important de souligner que si ces voies sont essentielles pour le métabolisme des xénobiotiques, elles jouent également des rôles non négligeables sur le métabolisme de nombreuses molécules endogènes que nous avons déjà évoquées pour certaines comme les stéroïdes (synthèse et dégradation), les acides gras et leurs dérivés (CYP4A) ou encore les acides biliaires (synthèse et élimination). Il est également nécessaire d'évoquer le fait que si ces voies métaboliques sont normalement dédiées à la détoxicification des xénobiotiques, elles peuvent dans certains cas et vis-à-vis de certaines molécules, renforcer leur pouvoir toxique (bioactivation) ou pharmacologique (notion de pro-drogue). La modulation et l'étude des polymorphismes de ces systèmes de biotransformation font l'objet de nombreuses études en raison de leur rôle évident dans la pharmacocinétique et la biodisponibilité des médicaments mais également en raison des potentielles interactions médicamenteuses qu'ils peuvent provoquer.

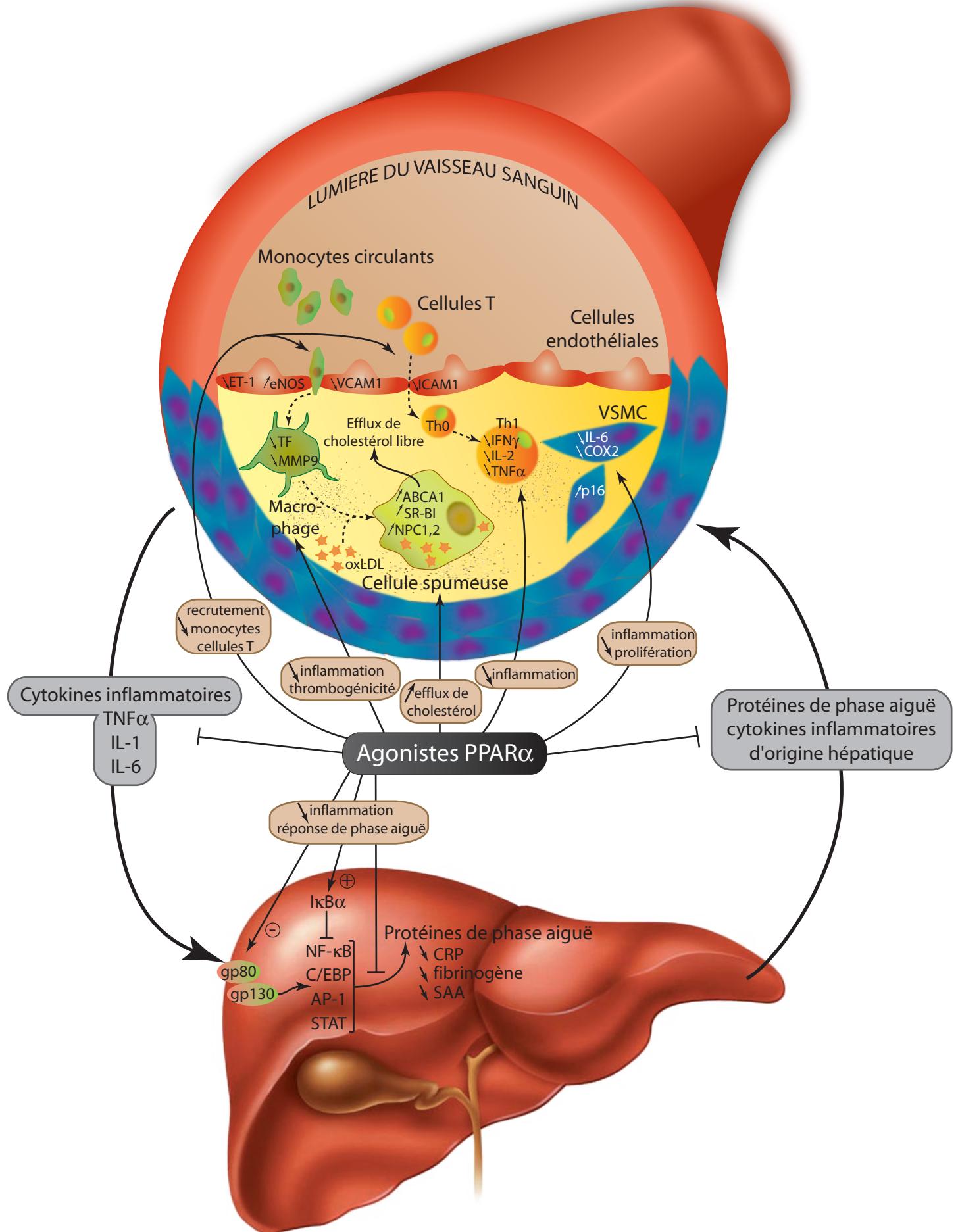
Nous avons déjà évoqué le rôle de PPAR $\alpha$  dans la modulation de l'expression des CYP4A chez les rongeurs, ainsi que son rôle dans la modulation du métabolisme et de la toxicité des acides biliaires via les régulations de Cyp7A1, Cyp27A1, Cyp8b1, Mdr2/Abcb4 et de l'UGT2B4 (pour une synthèse sur ces différents aspects, voir Barbier et al., 2004). Il est probable que parmi les nombreuses régulations rapportées par les PPs de gènes impliqués dans les biotransformations, certaines régulations soient dues à l'exposition du modèle à des doses massives de xénobiotique, éventuellement sur de longues durées, conduisant parfois à des altérations biochimiques des voies de biotransformation pouvant se traduire par des modulations compensatoires au niveau transcriptionnel. La validité des régulations que nous rapportons ici est donc à évaluer avec soin au regard des conditions expérimentales utilisées et de la spécificité des agonistes utilisés pour PPAR $\alpha$  mais éventuellement aussi pour les régulateurs majeurs de l'expression des enzymes de biotransformation que sont PXR et CAR. De nombreux gènes impliqués dans l'hydroxylation (CYP) et la déshydrogénéation (hydroxystéroïde déshydrogénases : HSD chez le rat et l'homme et Hsd chez la souris) des stéroïdes sexuels semblent régulés par certains activateurs de PPAR $\alpha$ , parfois de manière différente en fonction des molécules étudiées. Plusieurs PPs régulent négativement l'expression de CYP2C11 (Corton et al., 1997; Corton et al., 1998; Ripp et al., 2003) au niveau du foie des souris mâles, un CYP impliqué dans l'hydroxylation des stéroïdes (activité 2 $\alpha$ - et 16 $\alpha$ -testostérone et estradiol hydroxylase notamment), l'époxygénéation de l'acide arachidonique et le métabolisme des médicaments. Pour certaines molécules, le mécanisme semble dépendant de l'expression de PPAR $\alpha$  *in vitro* mais n'a pas été élucidé (Ripp et al., 2003). Chez les femelles, CYP2C11 n'est pas exprimé mais CYP2C12, une isoforme exclusivement exprimée chez les femelles et impliquée dans l'hydroxylation de la testostérone en position 15 $\alpha$ , est réprimée par certains fibrates (Corton et al., 1998). L'enzyme bifonctionnelle peroxysomale D-spécifique (D-PBE), également connue sous le nom de 17 $\beta$ -hydroxystéroïde déshydrogénase 4 (Hsd17b4) est impliquée dans la conversion de l'estradiol en estrone. Son induction PPAR $\alpha$ -dépendante (Aoyama et al., 1998; Corton et al., 1997; Fan et al., 1998b) pourrait participer à expliquer la baisse d'estradiol observée chez les rongeurs femelles exposées à certains PPs. Une légère inhibition de l'expression de Cyp3a11 (activité 6 $\beta$ -testostérone hydroxylase notamment, Barclay et al., 1999; Fan et al., 2004b) et de Cyp2c29 (Barclay et al., 1999; Martin et al., 2005) par certains PPs a également été rapportée. Cyp3a11 est cependant induit par le DEHP de manière PPAR $\alpha$ -indépendante (Fan et al., 2004b), vraisemblablement via l'activation de PXR (Hurst and Waxman, 2004). Une

diminution de l'expression de Cyp2A4/5 (les deux isoformes étant difficiles à distinguer par Northern et Western blot) par le clofibrate a été observée (Barclay et al., 1999; Kojo et al., 1996) mais elle serait vraisemblablement spécifique de ce fibrate et indépendante de PPAR $\alpha$  (Barclay et al., 1999; Cai et al., 2002; Kojo et al., 1996). Les PPs inhibent également l'expression hépatique de CYP2C7 chez le rat qui code un CYP à activité 16 $\alpha$ -testostérone hydroxylase et retinoic acid 4-hydroxylase (Fan et al., 2004a). Le DEHP inhibe l'expression des hydroxystéroïde déshydrogénases Hsd3b5 dans le foie et Hsd3b4 dans le rein (Wong et al., 2002), deux enzymes également impliquées dans le métabolisme des estrogènes et des androgènes. Une induction PPAR $\alpha$ -dépendante de l'activité hépatique acyl-CoA:estradiol et acyl-CoA:testostérone acyltransférase a également été rapportée mais la, ou les enzymes impliquées ne sont pas connues (Xu et al., 2001b). En dehors du métabolisme des stéroïdes sexuels, les fibrates inhibent également de manière PPAR $\alpha$ -dépendante l'expression hépatique de l'Hsd11b1 qui converti la cortisone en cortisol actif, constituant un mécanisme potentiel d'interaction entre la signalisation PPAR $\alpha$  et celle des glucocorticoïdes (Hermanowski-Vosatka et al., 2000). De plus, les PPs semblent également avoir des effets sur l'expression des CYP et HSD impliqués dans la biosynthèse des stéroïdes sexuels (stéroidogénèse). Chez le mâle, le ciprofibrate diminue l'ARN et l'activité (mais pas la protéine) HSD3 $\beta$  au niveau testiculaire (Hierlihy et al., 2006). Chez la femelle, une inhibition PPAR $\alpha$ -dépendante de l'expression et de l'activité aromatase (Cyp19a1 qui converti la testostérone en estradiol) et de l'expression de Cyp11a1 (ou P450scc, également impliqué dans la synthèse des stéroïdes sexuels) par le fénofibrate a été observée au bout de 3 semaines de traitement dans l'ovaire (Toda et al., 2003). Ces différentes régulations ainsi que l'observation d'effets variables en fonction des molécules sur le métabolisme de la testostérone notamment (Fan et al., 2004b) ne permettent pas, à mon sens, de dégager actuellement un schéma général du rôle de PPAR $\alpha$  et de son activation dans la régulation du métabolisme des stéroïdes sexuels. Il est néanmoins intéressant de constater que la question d'un lien éventuel entre un traitement par les fibrates et une baisse de la libido/impuissance est posée (Rizvi et al., 2002) et que, d'autre part, les phtalates augmentent les taux circulants de testostérone et d'estriadiol et posent un certain nombre de questions sur le plan de la santé publique en raison de leurs activités de disruption endocrinienne (Akingbemi et al., 2004). Parmi les CYP qui ne sont pas impliqués dans le métabolisme des stéroïdes, on peut également citer la régulation par PPAR $\alpha$ , via sa fixation à deux PPRE, du CYP1A1 humain,

une cible privilégiée du récepteur aux dioxines (AhR) impliquée dans la bioactivation de plusieurs carcinogènes (Seree et al., 2004).

Les activateurs de PPAR $\alpha$ , comme nous l'avons déjà évoqué, régulent également l'expression d'enzymes de phase II et III. Chez l'homme et la souris, PPAR $\alpha$  régule positivement l'expression de l'UGT1A9 via un PPRE identifié (Barbier et al., 2003b). L'UGT1A9 catalyse la glucuronidation de nombreux xénobiotiques et notamment des fibrates ainsi que celle de molécules endogènes comme des éicosanoïdes, suggérant que PPAR $\alpha$  régulerait ainsi l'élimination de ses propres activateurs pharmacologiques et peut-être endogènes. De plus, chez la souris, nous avons vu que PPAR $\alpha$  régule également l'expression de Mdr2/Abcb4 (Kok et al., 2003a), un important transporteur de glucurono-conjugués localisé sur les membranes canaliculaire des hépatocytes, favorisant ainsi d'autant l'élimination de ces composés. Récemment, l'utilisation de souris transgénique exprimant le locus de 220 Kb du chromosome 2 humain contenant les neufs UGT de la famille 1 (UGT1A1 à UGT1A9 qui partagent les mêmes exons 2 à 5 et diffèrent par l'utilisation d'un premier exon alternatif) a permis de démontrer que plusieurs UGT de cette famille étaient régulées par un fibrat de manière tissu-spécifique (Senekeo-Effenberger et al., 2007). Des PPRES ont été identifiés dans les promoteurs des UGT1A3 et 1A6 et un autre PPRE, localisé dans une zone du promoteur qui lui confère également une réponse aux récepteurs AhR, CAR et PXR, a été identifié dans le promoteur de l'UGT1A1. En ce qui concerne les GST, les PPs semblent globalement inhiber leurs activités (Voskoboinik et al., 1996). Les niveaux protéiques des GST $\mu$  et  $\pi$  sont diminués dans le foie de rat suite à un traitement par un fibrat (Foliot and Beaune, 1994). Si nous avons bien observé une diminution PPAR $\alpha$ -dépendante de l'expression de la Gst $\pi$  chez la souris, en revanche, nous constatons une induction PPAR $\alpha$ -dépendante de l'ARNm de la Gst $\mu$  dans le foie (Martin et al., 2005). L'implication de ces différentes enzymes de phase II dans l'élimination de xénobiotiques très divers doit cependant conduire à mener des investigations ciblées des effets de ces modulations sur le métabolisme de chaque composé d'intérêt.

En conclusion, PPAR $\alpha$  agit également comme régulateur de plusieurs systèmes de biotransformation mais les effets physiologiques de ces différentes modulations sont encore mal compris. Les principaux enjeux des investigations dans ce domaines sont : 1) l'effet des activateurs de PPAR $\alpha$  sur le métabolisme des stéroïdes endogènes leur conférant potentiellement un caractère de disrupteurs endocriniens, 2) leurs effets sur le métabolisme des acides biliaires, dans une perspective éventuelle de nouvelle application thérapeutique



**Figure 34 : Principaux effets de PPAR $\alpha$  sur l'inflammation**

La partie haute illustre les effets au niveau vasculaire et la partie basse les effets au niveau de la réponse de phase aiguë hépatique. Les principaux mécanismes d'inhibition de l'inflammation par PPAR $\alpha$  sont illustrés dans le foie uniquement. Schéma inspiré de F Zanbergen & J Plutzky, 2007, BBA, 1771(8):972-82 et de P Lefebvre et al., 2006, JCI, 116(3):571:80.

et/ou de réduction des effets secondaires, 3) l'évaluation du risque d'interactions médicamenteuses potentielles sous un traitement par les fibrates, 4) l'amélioration de la biodisponibilité et de la formulation des fibrates permettant éventuellement de réduire les doses employées et enfin, 5) la compréhension des différences entre espèces en ce qui concerne la réponse aux activateurs de PPAR $\alpha$ , en particulier dans le cadre du développement d'hépatocarcinomes lors de traitements de longue durée (possibilité d'interactions entre polluants environnementaux ou nutriments et activateurs de PPAR $\alpha$ ).

### **II.5.2. Rôles de PPAR $\alpha$ dans l'inflammation au niveau vasculaire**

Comme nous l'avons évoqué, le mode d'action des fibrates semble inclure des effets bénéfiques sur l'inflammation (voir [Figure 34](#)), à la fois au niveau vasculaire et au niveau de la réponse hépatique de phase aiguë (pour des synthèses récentes, voir Marx et al., 2004; Moraes et al., 2006; Zambon et al., 2006; Zandbergen and Plutzky, 2007). L'athérosclérose est une pathologie cardiovasculaire majeure et le rôle de l'inflammation dans son apparition et son développement est largement reconnu (Ross, 1999). Le rôle joué par PPAR $\alpha$  dans l'inflammation a initialement été révélé par la mise en évidence d'une réponse inflammatoire prolongée chez les souris PPAR $\alpha$ -/- (Devchand et al., 1996). Les travaux successifs ont permis de démontrer que l'action de PPAR $\alpha$  sur l'inflammation s'exerce principalement par le biais d'interférences avec les voies de signalisation d'autres facteurs de transcription comme NK- $\kappa$ B, C/EBP, AP-1 (activator protein-1) ou STAT, par exemple via des interactions protéine-protéine ou via des compétitions pour les corégulateurs (un phénomène parfois appelé squelching). La réduction de l'expression des gènes cibles de ces voies de signalisation par le biais de l'activation de PPAR $\alpha$  mais de manière indépendante de sa fixation à l'ADN est souvent qualifiée de transrépression (pour une synthèse sur ces mécanismes, voir Zambon et al., 2006). De fait, l'effet des agonistes de PPAR $\alpha$  est différent en fonction du contexte inflammatoire et les effets anti-inflammatoires de ces molécules sont donc principalement observés en présence d'un stimulus proinflammatoire qui active différentes voies de signalisation avec lesquelles PPAR $\alpha$  interfère. De même, parmi les expériences que nous allons décrire, beaucoup utilisent un prétraitement par les agonistes de PPAR $\alpha$  avant une stimulation par une cytokine proinflammatoire, l'exposition conjointe au stimulus proinflammatoire et à l'agoniste de PPAR $\alpha$  ne fonctionnant pas toujours.

Au niveau vasculaire, l'expression de PPAR $\alpha$  est observée dans les cellules endothéliales (Inoue et al., 1998; Marx et al., 1999), dans les cellules musculaires lisses de la

paroi vasculaire (VSMCs pour vascular smooth muscle cells, Marx et al., 1998; Staels et al., 1998), dans les monocytes et les macrophages (Marx et al., 2001; Neve et al., 2001) ainsi que dans les lymphocytes T et B (Jones et al., 2002; Marx et al., 2002), illustrant le large potentiel de régulation que ce RN peut exercer dans la paroi vasculaire et le processus de développement de la plaque d'athérome (voir Figure 34).

### *EFFETS SUR LES CELLULES ENDOTHELIALES VASCULAIRES*

Dans les cellules endothéliales vasculaires, PPAR $\alpha$  interfère avec les processus impliqués dans le recrutement et l'adhésion des leucocytes (Figure 34). Les leucocytes, principalement monocytes et cellules T, sont attirés sur les sites où se développe une lésion vasculaire par le biais de protéines chimioattractantes (chimiokines) excrétées par les cellules endothéliales et subendothéliales. Une fois attirés par ces chimiokines, les leucocytes se fixent à la paroi via des molécules d'adhésion endothéliale qui facilitent leur attachement et leur migration ultérieure au sein de la paroi des vaisseaux. Le fenofibrate et le Wy14643 inhibent l'induction de l'expression de VCAM-1 (vascular cell adhesion molecule-1) par le TNF- $\alpha$  dans des cellules endothéliales humaines, réduisant ainsi l'adhésion de monocytes aux cellules endothéliales stimulées (Marx et al., 1999). Des effets similaires ont été observés suite à une stimulation des cellules endothéliales par le 12-myristate 13-acétate (PMA) ou le lipopolysaccharide (LPS, Jackson et al., 1999). Cette inhibition semble principalement liée à une inhibition de l'activation de la transcription de VCAM-1 par NF- $\kappa$ B. Le fénofibrate diminuerait également la sécrétion de l'ICAM-1 (intercellular adhesion molecule 1) chez l'homme, comme cela a pu être mis en évidence par des dosages plasmatique chez des sujets obèses (Ryan et al., 2006). Le rôle de PPAR $\alpha$  dans la production de chimiokines par la paroi vasculaire semble moins évident et dépend sans doute largement du type cellulaire étudié, des gènes cibles évalués et des conditions expérimentales (agoniste, dose, durée, etc.). La stimulation de PPAR $\alpha$ , contrairement à celle de PPAR $\gamma$ , dans des cellules endothéliales humaines ne semble pas affecter l'induction de l'expression de plusieurs chimiokines de type CXC (IP-10, Mig, I-TAC) ou de MCP-1 (monocyte chemotactic protein-1) par l'interféron  $\gamma$  (IFN- $\gamma$ , Marx et al., 2000). De plus, un effet proinflammatoire de PPAR $\alpha$  a été proposé car son activation par des phospholipides oxydés ou par un agoniste synthétique dans des cellules endothéliales augmente l'expression de MCP-1 et de l'interleukine 8 (IL-8) dans ces cellules (Lee et al., 2000a). Cependant, via ses effets métaboliques, PPAR $\alpha$  diminue les petites LDL denses, plus sujettes à l'oxydation et il n'est donc pas certain que ces mécanismes soient

pertinents sur le plan thérapeutique. Une étude récente indique que les activateurs de PPAR $\alpha$  inhibent l'induction de l'expression de MCP-1 par des niveaux élevés de glucose dans une lignée de cellules endothéliales humaines (Dragomir et al., 2006). De plus, dans des cellules endothéliales d'aorte murine, PPAR $\alpha$  semble nécessaire à l'inhibition de la voie NF- $\kappa$ B par l'oxEPA (C20:5 $\omega$ 3 oxydé) qui se traduit notamment, dans des cellules endothéliales humaines, par une inhibition par l'oxEPA de l'induction par le TNF- $\alpha$  de MCP-1 et de l'IL8 (Mishra et al., 2004). In vivo, chez la souris soumise à un régime riche en graisses, l'expression de VCAM-1, de MCP-1, du TNF- $\alpha$ , de l'IFN- $\gamma$  et de l'IL-1 $\beta$  est diminuée dans les lésions athérosclérotiques vasculaires par les activateurs de PPAR $\alpha$  (Li et al., 2004). L'ensemble de ces résultats indiquent que PPAR $\alpha$  joue bien un rôle dans le recrutement des leucocytes par l'endothélium vasculaire. Les effets observés dépendent néanmoins certainement du contexte physiologique de l'organisme (hyperglycémie, hypertriglycéridémie, profil lipoprotéique, statut inflammatoire par exemple). Les cellules endothéliales jouent également un rôle important dans le tonus vasculaire (ou tonus vasomoteur). Ce dernier est contrôlé en partie par des stimuli biomécaniques : débit sanguin et pression sanguine. Les cellules endothéliales régulent le tonus vasculaire en équilibrant d'une part la sécrétion de monoxyde d'azote (NO) qui stimule la vasodilatation et accroît ainsi le débit sanguin et d'autre part la sécrétion d'endothéline-1 (ET-1) qui promeut la vasoconstriction et induit la prolifération des VSMCs. La nitric oxide synthase endothéliale (eNOS) est l'enzyme responsable de la synthèse de NO, capable également d'inhiber la voie NF- $\kappa$ B et d'atténuer ainsi les réponses inflammatoires endothéliales. Les fibrates et certains agonistes endogènes (l'EPA mais pas le DHA) de PPAR $\alpha$  semblent favoriser l'expression de eNOS et la production de NO par des cellules endothéliales, suggérant un effet positif sur la vasorelaxation (Goya et al., 2004; Omura et al., 2001). D'autres études montrent que les fibrates inhibent partiellement l'induction de l'ET-1 induite par la thrombine ou des LDL-oxydés (Delerive et al., 1999b; Martin-Nizard et al., 2002), sans doute via une inhibition des voies AP-1 et PKC. Ces effets combinés pourraient participer, avec une amélioration de la fonction des VSMCs, à l'augmentation du tonus vasculaire observée chez des patients hypertriglycéridémiques suite à un traitement par le fénofibrate (Capell et al., 2003).

#### ***EFFETS SUR LES CELLULES MUSCULAIRES LISSES VASCULAIRES***

Au niveau des VSMCs, PPAR $\alpha$  inhibe l'induction par l'IL-1 $\beta$  de l'expression de l'IL-6 et de la production de prostaglandines (Staels et al., 1998). Ce dernier effet serait dû à une

inhibition par PPAR $\alpha$  de l'induction provoquée par l'IL-1 $\beta$  de l'expression de la cyclooxygénase-2 (COX-2) via une inhibition de la voie NF- $\kappa$ B. L'inhibition de l'induction de l'IL-6, également observée dans des aortes de souris stimulées au lipopolysaccharide (LPS), semble liée à des interactions directes PPAR $\alpha$ /c-Jun et PPAR $\alpha$ /p65, des sous-unités des facteurs de transcription AP-1 et NF- $\kappa$ B respectivement (Delerive et al., 1999a). Aux sites d'inflammation, la production de prostaglandines via l'activité COX-2 augmente la production de divers médiateurs inflammatoires, suggérant là encore un rôle anti-inflammatoire du récepteur PPAR $\alpha$ . Lors de l'évolution de la plaque d'athérome vers une lésion athérosclérotique, les VSMCs prolifèrent et migrent pour s'accumuler dans la plaque d'athérosclérose. La production de matrice extracellulaire par les VSMC peut conduire à étendre la lésion et à rendre la plaque plus fibreuse. Les VSMC sont aussi la source du matériel qui forme la chape fibreuse qui sépare de la circulation le cœur nécrotique, riche en lipide, de la lésion. La rupture de cette chape fibreuse expose le cœur pro-coagulant de la plaque d'athérosclérose à la circulation sanguine, activant ainsi la cascade de la coagulation et conduisant à une thrombose. La stabilité de la plaque d'athérosclérose est en particulier liée à l'équilibre entre la formation et la dégradation de la matrice extracellulaire deux mécanismes modulés par l'inflammation et qui font intervenir les VSMCs. Les fibrates ne semblent pas modifier la production par les VSMCs de matrix metalloprotéinases (MMP), des enzymes impliquées dans la dégradation de la matrice extracellulaire nécessaire à la migration des VSMCs (Marx et al., 1998). Leurs effets sur la diminution de la production locale de cytokines proinflammatoires pourraient cependant limiter l'expression des MMPs et augmenter la synthèse de collagène, favorisant ainsi la stabilité de la plaque. De plus, il a été montré que PPAR $\alpha$  inhibe la prolifération des VSMCs (Gizard et al., 2005). Cet effet semble lié à la régulation positive par PPAR $\alpha$  de p16/CDKN2A (cyclin-dependent kinase inhibitor 2A), via un PPRE identifié et en coordination avec le facteur de transcription Sp1. En conséquence, p16 séquestre la cyclin-dependent kinase 4 (CDK4), l'empêchant de phosphoryler le facteur de transcription pRB/RB1 (rétinoblastoma 1) qui contrôle la transition entre les phases G1 et S. Le résultat est qu'en l'absence de PPAR $\alpha$ , les VSMCs prolifèrent significativement plus qu'en sa présence et que son activation par un agoniste inhibe encore plus leur prolifération (Gizard et al., 2005). De plus, il a été rapporté que le DHA (docosahexaenoic acid) induisait l'apoptose des VSMCs par des mécanismes impliquant notamment la MAP kinase p38 (Diep et al., 2000). Ces effets semblent bénéfiques dans le cadre d'interventions chirurgicales sur les vaisseaux, comme la pose d'un stent ou une

angioplastie, qui peuvent être suivies d'une prolifération des VSMCs participant à la resténose des vaisseaux. Cependant, le bénéfice éventuel de ces effets sur la plaque d'athérosclérose n'est pas entièrement démontré car l'inhibition de la prolifération des VSMC pourrait limiter leur capacité à produire la chape fibreuse en même temps qu'elle limiterait l'épaississement de la média au niveau de la plaque (Zandbergen and Plutzky, 2007).

### *EFFETS SUR LES MONOCYTES ET LES MACROPHAGES*

Nous avons déjà largement évoqué l'effet bénéfique des fibrates sur la promotion du transport réverse du cholestérol, limitant ainsi la formation de cellules spumeuses qui constitue l'un des évènements initiant la formation de la plaque d'athérome ([Figure 34](#)). A un stade avancé, les monocytes et les macrophages ont tendance à s'accumuler au niveau de l'épaulement de la plaque d'athérosclérose, fragilisant encore plus cette zone au niveau de laquelle la rupture a souvent lieu. Le fénofibrate inhibe l'expression de MMP9 dans une lignée de monocytes humain activés par le LPS (Shu et al., 2000) par un effet sans doute indirect lié à l'induction de la forme inductible de la NO synthase (iNOS) qui est accrue par l'activation de PPAR $\alpha$ , conduisant ainsi à une augmentation du niveau basal de NO qui réduirait la stabilité de l'ARNm de MMP9 (Eberhardt et al., 2002). Cet effet pourrait participer à limiter la déstabilisation de la plaque d'athérosclérose par les macrophages et monocytes. En plus de sécréter des MMPs, les monocytes et les macrophages produisent du facteur tissulaire (TF pour tissue factor), une glycoprotéine transmembranaire qui est l'un des principaux activateurs de la coagulation, favorisant ainsi la thrombogénérité de la plaque. Les activateurs de PPAR $\alpha$  inhibent l'induction de l'expression du TF dans les monocytes et les macrophages humains stimulés avec du LPS ou de l'IL-1 $\beta$ , sans doute en interférant avec la voie NF- $\kappa$ B (Marx et al., 2001; Neve et al., 2001). De plus, ils inhibent également l'induction de l'expression du TNF- $\alpha$  par le LPS dans ces cellules (Marx et al., 2001). Finalement, l'activation de PPAR $\alpha$  augmente l'apoptose de macrophages humains activés par le TNF- $\alpha$  ou l'IFN- $\gamma$ , vraisemblablement via une inhibition de NF- $\kappa$ B qui joue un rôle anti-apoptotique dans ces cellules (Chinetti et al., 1998). Ces différents effets semblent de nature à favoriser la stabilité de la plaque et à limiter la réponse inflammatoire des monocytes et macrophages.

### *EFFETS SUR LES LYMPHOCYTES T*

Les cellules T et les cellules dendritiques contribuent également à la réponse inflammatoire vasculaire lors des étapes initiales de l'athérogenèse ([Figure 34](#)). Les

lymphocytes T, principalement de type CD4+ (ou T-helper), pénètrent dans la paroi des vaisseaux sous la forme de lymphocytes naïfs (cellules Th0). Dans le subendothélium, ils se différencient en cellules Th1 sous l'influence de divers antigènes, dont les LDL oxydés. Les cellules Th1 sécrètent des cytokines proinflammatoires comme l'IFN- $\gamma$ , le TNF- $\alpha$  et l'IL-2 qui vont ensuite activer les différents types cellulaires impliqués dans l'athérosclérose : monocytes, macrophages, cellules endothéliales et VSMCs. Nous avons vu que les effets des activateurs de PPAR $\alpha$  sur les cellules endothéliales sont susceptibles de limiter le recrutement et la pénétration des lymphocytes T dans la paroi vasculaire. L'expression de PPAR $\alpha$  est maximale dans les lymphocytes T au repos puis diminue suite à leur activation où c'est alors PPAR $\gamma$  dont l'expression prédomine (Jones et al., 2002). Les agonistes de PPAR $\alpha$  inhibent l'expression et la sécrétion d'IFN $\gamma$ , de TNF- $\alpha$  et d'IL-2 par les lymphocytes CD4+ humains et réduisent ainsi leur capacité d'activation de cellules endothéliales et de monocytes (Marx et al., 2002). Ces effets sont d'autant plus importants qu'ils sont susceptibles d'intervenir tôt dans le processus d'athérogenèse. Les mécanismes et notamment le rôle précis joué par PPAR $\alpha$  dans ce contexte ne sont pas connus mais il est probable que des interférences avec les voies AP-1 et/ou NF- $\kappa$ B soit également impliquées dans ces processus.

PPAR $\alpha$  semble donc avoir une action globalement anti-inflammatoire mais régule également le recrutement, l'adhésion, la prolifération et l'apoptose de divers types cellulaires impliqués dans la formation de plaques d'athérosclérose. Ces effets participent sans doute pour partie aux effets bénéfiques des fibrates sur les pathologies cardiovasculaires. Mais le rôle anti-inflammatoire de PPAR $\alpha$  a également été révélé au niveau hépatique et est en lien direct avec ses effets vasculaires (voir [Figure 34](#)).

### **II.5.3. Rôles de PPAR $\alpha$ dans l'inflammation au niveau hépatique**

Le foie est le site de synthèse d'un groupe de protéines appelées protéines de phase aiguë. La réaction de phase aiguë est une réaction généralisée de l'organisme à une perturbation de son homéostasie (blessure, choc, brûlure, inflammation, infection, etc..) qui se traduit par la modulation de la sécrétion d'un grand nombre de protéines par le foie (protéines de phase aiguë négatives si leur sécrétion est inhibée et positive si elle est induite) en réponse à des stress très variés, dont l'exposition à des cytokines (IL-1, IL-6, IL-8, TNF- $\alpha$  en particulier) sécrétées par les neutrophiles granulocytes et les macrophages. Le fibrinogène fait

partie des protéines de phase aiguë positive et est un agent pro-coagulant, favorisant donc, dans le contexte des pathologies cardiovasculaires, la formation de thrombus. La protéine C-réactive (CRP) et la serum amyloïde A (SAA) sont également des protéines de phase aiguë positives et pourraient jouer des rôles également délétères au niveau de la plaque d'athérome. Ces trois protéines, mais surtout le fibrinogène et la protéine C-réactive sont considérées comme des marqueurs de risque cardiovasculaire (Zambon et al., 2006; Zandbergen and Plutzky, 2007). L' $\alpha$ 2-macroglobuline, l'haptoglobuline, la ferritine, le plasminogène et l' $\alpha$ 1-antitrypsine sont d'autres exemples de protéines de phase aiguë positives. Parmi les protéines de phase aiguë négatives on trouve notamment l'albumine et les MUP (major urinary proteins) chez la souris.

Les agonistes de PPAR $\alpha$  inhibent l'induction par l'IL-1 (mais pas par l'IL-6) de l'expression de la protéine C-réactive CRP chez l'homme (Kleemann et al., 2003). Cette protéine n'est pas une protéine de phase aiguë chez la souris mais l'inhibition par les agonistes de PPAR $\alpha$  de l'induction de la CRP par l'IL-1 a également été étudiée sur des souris transgénique exprimant la CRP humaine (Kleemann et al., 2004). Chez ces souris, l'activation de PPAR $\alpha$  réduit également le niveau basal de CRP plasmatique, en l'absence de stimulus proinflammatoire. Au niveau transcriptionnel, il a été montré que l'induction de la CRP humaine par l'IL-1 était due à la fixation au promoteur de la CRP d'un complexe comprenant la sous-unité p50 de NF- $\kappa$ B et le facteur de transcription C/EBP $\alpha$  (Kleemann et al., 2003). L'activation de PPAR $\alpha$  se traduit par l'induction au niveau transcriptionnel de I $\kappa$ B $\alpha$  dans le foie (souris et homme) et dans les VSMCs (Delerive et al., 2000; Kleemann et al., 2003; Kleemann et al., 2004). Le mécanisme de cette induction n'est pas complètement élucidé mais ne semble pas impliquer la fixation de PPAR $\alpha$  sur un PPRE mais plutôt un accroissement de l'induction d'I $\kappa$ B $\alpha$  par les facteurs de transcription p65 et Sp1 via une stabilisation du complexe Médiateur par l'intermédiaire d'une interaction entre PPAR $\alpha$  et Med1 (Delerive et al., 2002). I $\kappa$ B $\alpha$  est un inhibiteur de NF- $\kappa$ B qui se fixe à la sous-unité p50 et inhibe sa translocation nucléaire et sa fixation à l'ADN participant ainsi à l'inhibition de l'expression de nombreux gènes cibles de la voie NF- $\kappa$ B. Chez la souris, l'activation de PPAR $\alpha$  se traduit également par une réduction PPAR $\alpha$ -dépendante des niveaux de protéines p50-NF $\kappa$ B et C/EBP $\beta$  (Kleemann et al., 2003). Une réduction PPAR $\alpha$ -dépendante des protéines C/EBP $\alpha$ ,  $\beta$  et  $\gamma$  et de c-Jun par le fénofibrate a également été observée dans le foie de souris (Gervois et al., 2004). Ces effets combinés conduisent à une réduction par le fénofibrate des niveaux plasmatique de CRP chez l'homme (Gervois et al., 2004; Staels et al.,

1998). Des effets similaires ont également été observés sur les concentrations plasmatiques de fibrinogène (Gervois et al., 2004; Maison et al., 2002; Staels et al., 1998). L'inhibition provoquée par l'activation de PPAR $\alpha$  de l'induction par l'IL-6 des fibrinogènes  $\alpha$ ,  $\beta$  et  $\gamma$  semble faire intervenir les mêmes mécanismes (induction de I $\kappa$ B $\alpha$  et, chez la souris, inhibition de l'expression des C/EBP et de c-Jun) mais il a également été observé, chez la souris, une inhibition PPAR $\alpha$ -dépendante de l'expression hépatique du récepteur membranaire à l'IL-6 gp80 et de la glycoprotéine gp130 intervenant dans la transduction du signal de l'IL-6 et d'autres cytokines de cette famille (Gervois et al., 2004). Cette réduction, spécifique au foie, se traduit notamment par une inhibition de la phosphorylation des facteurs de transcription STAT3 et c-Jun. Cette inhibition de la voie de signalisation de l'IL-6 participe à l'effet d'un traitement chronique par un fibrate sur la réponse inflammatoire hépatique qui conduit à une diminution globale de la réponse de phase aiguë à l'IL-6 chez la souris, comprenant une inhibition de l'induction des fibrinogènes  $\alpha$ ,  $\beta$  et  $\gamma$ , de la serum amyloïde A et de l'haptoglobuline et un blocage de l'inhibition de l'expression de MUP (Gervois et al., 2004). Chez l'homme, un traitement de quatre semaines par le fénofibrate conduit à une diminution des taux plasmatiques de fibrinogène, de SAA, de CRP, de plasminogène et d' $\alpha$ 2-macroglobuline ainsi qu'à une légère augmentation de l'albumine (Gervois et al., 2004). Un autre mécanisme a également été proposé pour expliquer l'inhibition de l'expression du fibrinogène  $\beta$  chez l'homme : l'activation de PPAR $\alpha$  conduirait à une séquestration du coactivateur SRC-2 (ou TIF2/GRIP1) qui ne pourrait plus participer à l'induction du fibrinogène  $\beta$  par C/EBP $\beta$  (phénomène de squelching, Gervois et al., 2001). Enfin, le niveau d'expression basal de SAA et l'activité basale de NF- $\kappa$ B sont élevés dans le foie des souris PPAR $\alpha$ -/- (Han et al., 2006) et un traitement par un agoniste de PPAR $\alpha$  inhibe de manière PPAR $\alpha$ -dépendante l'induction de l'expression de la SAA provoquée par des cytokines proinflammatoires (Gervois et al., 2004; Han et al., 2006). Enfin, les agonistes de PPAR $\alpha$  réduisent également les concentrations plasmatiques des cytokines proinflammatoires elles-mêmes et notamment celles du TNF- $\alpha$ , de l'IFN- $\gamma$  et de l'IL-6 (Madej et al., 1998; Staels et al., 1998).

L'ensemble de ces données illustrent le rôle majeur joué par PPAR $\alpha$  dans la réponse de phase aiguë hépatique qui consiste principalement à limiter la réponse inflammatoire de l'organe en interférant négativement avec plusieurs voies de transcription comme NF- $\kappa$ B, AP-1, C/EBP et STAT. Il est intéressant de constater que ces effets ne semblent pas systématiquement limités à l'activation pharmacologique de PPAR $\alpha$  mais peuvent également

s'appliquer aux mécanismes d'activation endogènes du récepteur comme l'illustre par exemple l'augmentation de l'expression basale de SAA chez les souris PPAR $\alpha$ -/- (Han et al., 2006), suggérant ainsi un véritable rôle physiologique de PPAR $\alpha$  dans le contrôle de l'inflammation. Les modifications que PPAR $\alpha$  provoque au niveau des protéines sécrétées par le foie ont certainement des répercussions systémiques, notamment au niveau de la paroi vasculaire et se potentialisent sans doute avec ses effets métaboliques et sur l'inflammation locale pour déterminer son action bénéfique vis-à-vis des pathologies cardiovasculaires. Mais ses effets sur l'inflammation, ainsi que ses effets métaboliques participent également à son action dans d'autres domaines, comme la réparation et la régénération des tissus après une blessure ou une ischémie/reperfusion ou encore la réponse de l'organisme à un choc septique (pour une synthèse, voir Michalik and Wahli, 2006). La [Figure 34](#) tente de synthétiser les principaux effets de PPAR $\alpha$  sur l'inflammation aux niveaux vasculaire et hépatique.

Dans ce chapitre consacré au récepteur PPAR $\alpha$ , nous avons développé ses principaux sites d'expression, sa régulation au niveau transcriptionnel et la régulation de son activité par ses ligands, ses cofacteurs ou par d'autres voies de signalisations. Nous avons développé ses principaux rôles qui concernent la régulation du métabolisme et la régulation de l'inflammation. Nous avons également régulièrement évoqué ses interactions avec d'autres voies de signalisation. PPAR $\alpha$  remplit une multitude de fonctions dans l'organisme qui se révèlent le plus souvent en réponse à un stimulus particulier comme une infection, une blessure, un régime trop riche en graisse, un jeûne, etc. C'est peut-être aussi ce qui participe à rendre attrayante la recherche sur ce facteur de transcription dont les rôles apparaissent à la fois si nombreux et si cachés. Ce chapitre a cependant ses limites alors que la littérature regorge de bien d'autres informations sur les fonctions de PPAR $\alpha$  qui semblent s'étendre bien au delà de ce qui a été présenté ici. En voici quelques unes qui n'ont pas été abordées dans ce manuscrit et qui me permettront de fournir au lecteur intéressé quelques références bibliographiques complémentaires :

Les rôles de PPAR $\alpha$  dans la prolifération des peroxysomes (Reddy, 2004; Yan et al., 2005), dans le stress oxydatif (Schrader and Fahimi, 2006; Yeldandi et al., 2000) et dans le contrôle de la prolifération cellulaire et de l'apoptose (Roberts et al., 2002; Roberts et al., 2000; Suga, 2004) n'ont pas, ou peu, été abordés. Ces phénomènes sont très vraisemblablement impliqués dans le développement d'hépatocarcinomes (Cattley and Roberts, 2000; Michalik et al., 2004; Peters et al., 2005), ce qui semble être une réponse

relativement spécifique des rongeurs aux activateurs de PPAR $\alpha$  mais pose néanmoins d'importantes questions dans le cadre du développement de nouveaux agonistes de PPAR $\alpha$  à visée thérapeutique. Au niveau du fonctionnement normal et pathologique des organes et de la physiologie des organismes, les rôles de PPAR $\alpha$  dans le tissu adipeux (Mazzucotelli et al., 2007), dans le rein (Guan, 2004; Izzedine et al., 2004), dans la peau (Icre et al., 2006; Michalik and Wahli, 2007), dans le cerveau (Bordet et al., 2006), dans la reproduction (Fournier et al., 2007; Froment et al., 2006), plus particulièrement dans la réponse aux disrupteurs endocriniens de type phtalates (Corton and Lapinskas, 2005), dans la réparation des tissus (Michalik and Wahli, 2006) et dans l'obésité (Costet et al., 1998; Kersten, 2002) notamment n'ont pas ou peu été abordés. Ses rôles dans le cœur (Finck, 2007; Huss and Kelly, 2004; Schiffrian, 2005) n'ont été que brièvement exposés.

Enfin, il y a certainement des rôles du récepteur qui restent encore totalement inconnus, peut-être parce que les stimuli appropriés n'ont pas été évalués sur les modèles animaux disponibles ou bien parce que ces modèles ont leurs limites vis-à-vis des extrapolations à l'homme ou à d'autres espèces d'intérêt.

Après avoir opéré ce tour d'horizon de la bibliographie qui a servie de base à l'élaboration des études conduites dans le cadre de ce doctorat, nous nous intéresserons dans le chapitre suivant aux outils qui ont été utilisés et développés pour la réalisation de ces études en les plaçant dans le cadre plus général des méthodes d'analyse de l'expression des gènes.

## Chapitre III : Méthodes d'analyse de l'expression des gènes

Nous avons vu que l'étude des RNs fait intervenir de multiples champs disciplinaires et une grande variété de techniques. Des études portent sur leur structure, sur leur expression et sa régulation, sur leurs interactions avec diverses protéines (dimérisation, autres facteurs de transcription, corégulateurs...), sur leurs modifications post-traductionnelles, sur leurs ligands et les signaux qui les activent ou les inhibent, sur leur localisation et leur dynamique, sur leurs interactions avec l'ADN ou encore sur les régulations transcriptionnelles qu'ils opèrent. C'est très majoritairement ce dernier type d'études que nous conduisons. PPAR $\alpha$  étant un facteur de transcription activé par divers ligands, nous avons souhaité, dans le cadre de ce doctorat, acquérir une vue un peu plus large des régulations géniques opérées par ce récepteur et ce dans des contextes d'activation variés. Nous nous sommes donc naturellement intéressés aux différentes méthodes qui permettent de mesurer l'expression des gènes, cette donnée constituant la première résultante de l'activité de PPAR $\alpha$ . Dans ce chapitre, nous considérerons que mesurer l'expression d'un gène consiste à mesurer l'abondance de l'ARNm produit à partir de ce gène à un instant donné. Nous n'évoquerons donc pas les méthodes permettant de vérifier qu'une régulation de l'expression d'un gène s'opère bien au niveau transcriptionnel comme la transcription sur noyaux isolés (nuclear run-on) ou l'utilisation d'inhibiteurs de la traduction (cycloheximide par exemple) et de la transcription (actinomycine D par exemple). De même, nous n'évoquerons pas la mesure de l'abondance des protéines qui sont pourtant bien des produits de l'expression des gènes. Néanmoins, entre l'étape de la transcription qui nous intéresse et la protéine, plusieurs étapes peuvent influer sur l'abondance d'une protéine : régulation post-transcriptionnelle par modification de la stabilité de l'ARN, modulation de la traduction, modulations post-traductionnelles par modification de la stabilité de la protéine. Naturellement, les régulations post-transcriptionnelles peuvent également influer sur le lien entre transcription et abondance d'un ARNm mais nous nous affranchissons néanmoins des autres étapes évoquées et nous rapprochons donc ainsi d'une évaluation plus juste de l'activité du récepteur. Bien entendu, en choisissant cette approche, nous acceptons en contrepartie de nous éloigner des conséquences fonctionnelles et physiologiques des modulations provoquées par le récepteur. Mais celles-ci ont fait l'objet, dans certaines de nos études, d'une évaluation spécifique nécessitant la mise en œuvre de techniques et de compétences complémentaires (histologie, dosages de métabolites dans les tissus ou le plasma et mesures d'activités enzymatiques en particulier).

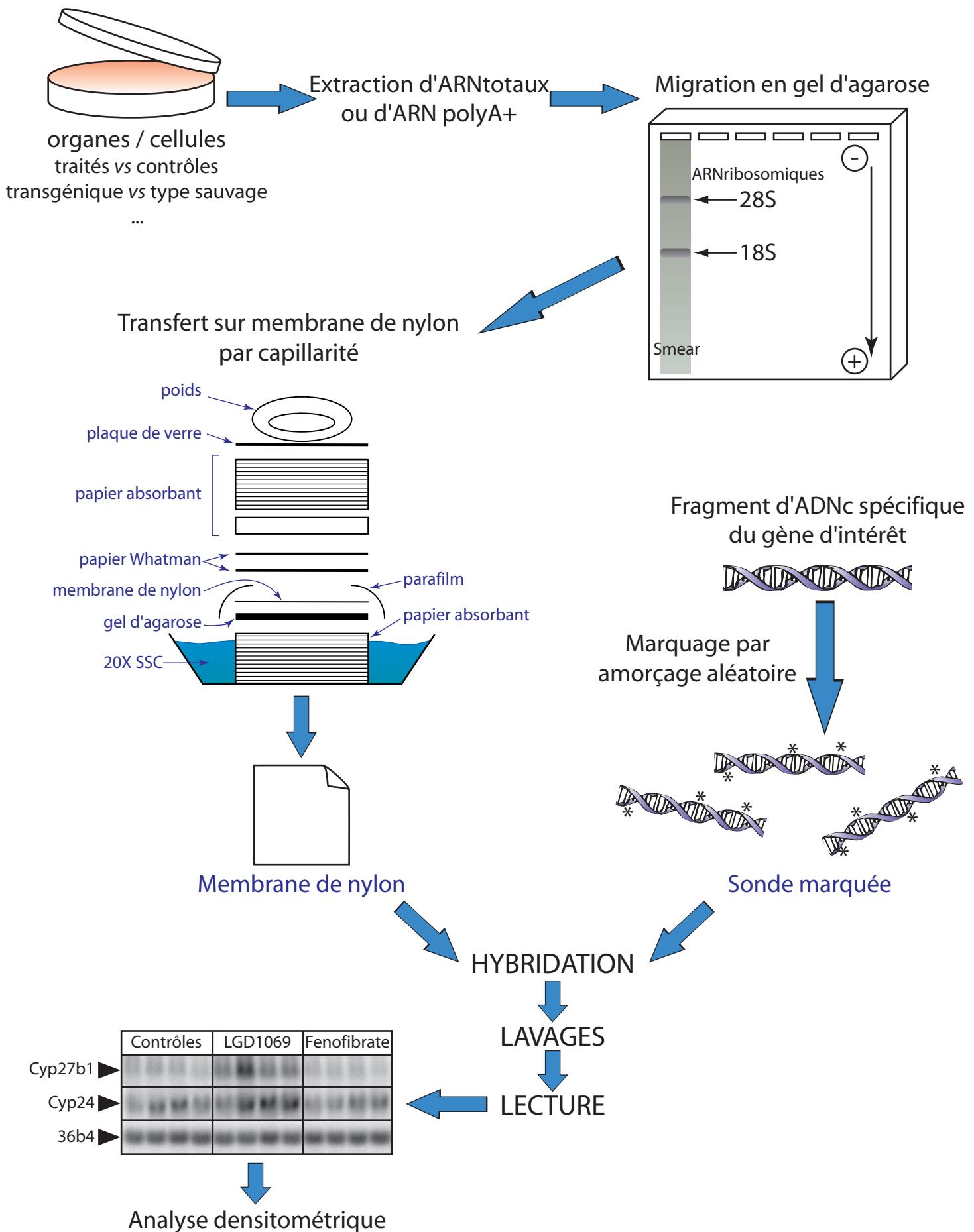


Figure 35 : Principe général du northern blot

Les dot blot et slot blot utilisent globalement le même principe sauf pour l'étape de migration sur gel. Les images de northern blot présentées sont issues d'une étude sur souris C57BL/6J traitées par un agoniste de RXR (LGD1069) ou de PPAR $\alpha$  (fénofibrate) ou par le véhicule seul (contrôles).

Parmi les méthodes de mesure de l'expression génique, on peut distinguer des méthodes qui s'adaptent particulièrement bien à l'étude d'un seul gène, éventuellement sur de nombreux échantillons et que l'on qualifiera ici de techniques à « bas débit », ce terme s'appliquant au nombre de gènes mesurés et pas au nombre d'échantillons et, à l'opposé, les méthodes à « haut débit » plus adaptées à l'étude de nombreux gènes mais en général avec des limitations sur le nombre d'échantillons que l'on peut analyser. Après avoir présenté les principales techniques qui constituent ces deux familles de méthodes en précisant leurs avantages et limites, nous présenterons la mise au point de notre puce à ADN INRAArray en justifiant les choix technologiques réalisés. Nous présenterons ensuite les points essentiels de l'analyse des données d'expression génique. Ce chapitre se terminera par une illustration de la mise en œuvre de plusieurs méthodes d'analyse statistique sur un jeu de données généré dans le cadre de ce doctorat qui a fait l'objet d'une publication de synthèse dans un numéro spécial, dédié aux microarrays, du journal de la société française de statistique.

### **III.1. TECHNIQUES DE MESURE DE L'EXPRESSION GENIQUE A « BAS DEBIT »**

Parmi les méthodes de mesure de l'expression des gènes, les principales méthodes à bas-débit sont le northern blot, la protection aux RNases (RNase protection assay ou RPA en anglais) et la PCR en temps réel. La PCR dite semi-quantitative a perdu beaucoup de terrain dans les laboratoires depuis l'introduction de la PCR en temps réel, principalement pour deux raisons : la première est due au fait qu'elle n'est que semi-quantitative et la seconde est que pour atteindre ce caractère semi-quantitatif, elle nécessite de nombreuses mises au point. Cette technique ne sera donc pas présentée dans ce chapitre. Parmi les trois méthodes présentées dans ce chapitre, le northern blot et la PCR en temps réel ont été utilisés dans le cadre de ce travail de doctorat et sont donc présentées plus en détail.

#### **III.1.1. Le Northern Blot**

##### *PRINCIPE GENERAL :*

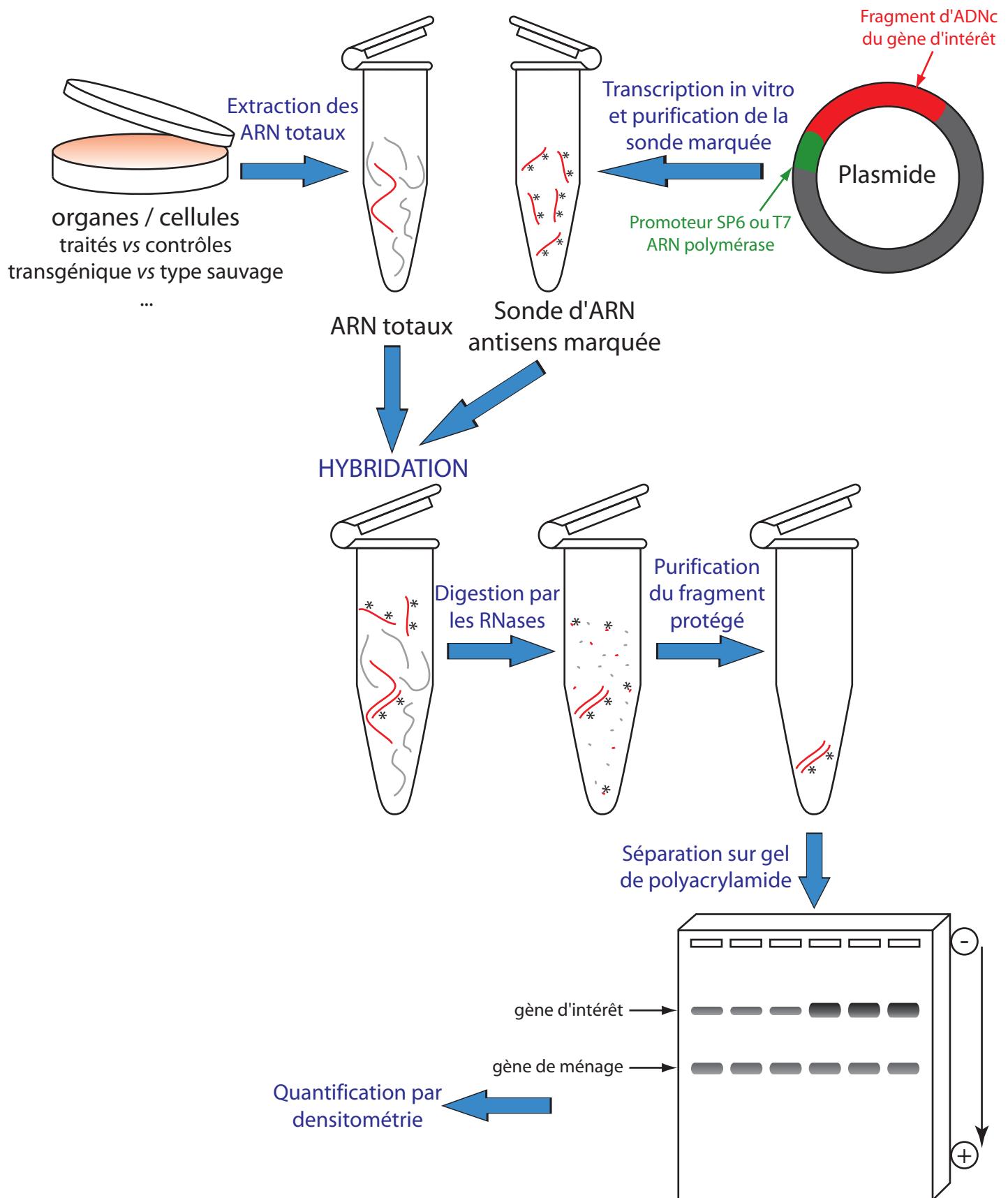
Le principe général du northern blot (NB) est illustré en [Figure 35](#). Son nom lui vient du Dr Southern qui appliqua pour la première fois une technique similaire à l'ADN, et donna ainsi naissance à une famille de méthodes : le southern blot pour l'ADN, le northern blot pour les ARN et le western blot pour les protéines, basées sur une séparation par électrophorèse, un transfert sur membrane et une hybridation/détection à l'aide d'une sonde spécifique. Les ARNs totaux où les ARNm (ou ARN polyA+ car ils comportent une queue polyA) sont extraits des tissus ou cellules étudiées. On réalise ensuite une électrophorèse de ces ARNs en

condition dénaturante (gel d'agarose généralement à 1,2% contenant du formaldéhyde). Le gel est ensuite transféré, le plus souvent par capillarité (ou bien sous vide) sur une membrane de nylon chargé positivement (qui fixe les acides nucléiques chargés négativement). Cette membrane est séchée puis conservée à 4°C au sec jusqu'à son hybridation. Signalons qu'il existe une alternative au NB qui consiste à déposer directement sur la membrane de nylon les échantillons grâce à un appareil sous vide qui porte des puits (dot blot) ou des fentes (slot blot). L'hybridation de la membrane portant les ARNs doit être réalisée à l'aide d'un fragment d'ADNc marqué. La séquence de cet ADNc est choisie pour être parfaitement homologue à l'ARN dont on souhaite mesurer l'abondance mais ne doit pas provoquer d'hybridation croisée avec d'autres transcrits et être donc spécifique de l'ARN étudié. Après avoir sélectionné des amorces adéquates, adjacentes au fragment à amplifier, le fragment choisi est amplifié par PCR (polymerase chain reaction) à partir d'ARN totaux ou polyA+. Il est possible d'effectuer un marquage directement par PCR en incorporant dans la réaction un désoxynucléotide triphosphate (dNTP) marqué (par exemple au  $^{32}\text{P}$  dans le cas d'un radiomarquage). Sinon, le fragment peut être clonné puis extrait du plasmide par digestion à l'aide d'enzymes de restriction et enfin (radio)marqué par amorçage aléatoire (random priming) à l'aide du fragment de Klenow de l'ADN polymérase I. Ce fragment possède l'activité de réparation de l'ADN (5'-3' polymérase et 3'-5'exonucléase) de l'ADN polymérase I et va donc reconstruire les brins complémentaires en partant des amorces aléatoires fixées sur l'ADNc. Les amorces utilisées sont généralement des hexamères dont la séquence est aléatoire. L'incorporation d'un dNTP marqué dans la réaction permet l'obtention d'une sonde double brin marquée. A la suite de la réaction de marquage, les nucléotides non incorporés sont éliminés, généralement par purification sur colonne de Sephadex G50 suivant un principe de chromatographie d'exclusion/diffusion. Une fois la sonde d'ADNc marquée obtenue, la membrane est hydratée (2X SSC ; 0,5% SDS) puis mise en préhybridation, généralement dans un four à hybridation (tubes de borosilicate contenant du tampon d'hybridation et placés en rotation dans un four) pendant quelques heures. La sonde marquée est ensuite ajoutée au milieu d'hybridation et l'hybridation se déroule durant environ 12h. La température d'hybridation dépend généralement du tampon utilisé et peut être ajustée en fonction du signal souhaité : une faible température augmentera le signal mais risquera d'augmenter le bruit de fond et de provoquer des hybridations croisées alors qu'une température plus élevée assurera une meilleure spécificité mais risquera de diminuer le signal. Le milieu d'hybridation est ensuite éliminé puis la membrane lavée, dans un premier temps avec un tampon peu stringent (2X SSC ; 0,5% SDS) puis éventuellement avec un tampon plus

stringent (0,2X SSC ; 0,5% SDS, voire avec de l'eau distillée stérile) si le bruit de fond est encore élevé (dans le cas d'un radiomarquage à l' $[\alpha\text{-}^{32}\text{P}]$ dCTP, on peut, avec l'habitude, évaluer ce bruit de fond au compteur Geiger). La membrane est ensuite placée entre deux feuilles de film Saran et, suivant le type de marquage, lu avec un appareillage adéquat. Dans le cas d'un radiomarquage, l'exposition à un écran au phosphore et l'utilisation d'un scanner de type Storm (imageur au phosphore) permet l'obtention directe d'une image digitalisée. L'image est ensuite analysée par densitométrie pour en extraire une information quantitative de l'abondance du transcrit étudié dans les différents échantillons. La membrane peut être déshybridée (généralement à 70°C en 0,2X SSC ; 0,5% SDS) et réhybridée jusqu'à 5-6 fois. Dans les mêmes échantillons, la mesure de l'expression d'un gène dont on sait qu'il ne varie pas en fonction des conditions expérimentales (« gène de ménage » ou « housekeeping gene ») permet éventuellement de calibrer les données dans les deux cas suivants : 1) dépôt de quantités variables d'ARN d'un puits à l'autre au cours de l'électrophorèse ou 2) dégradation différentielle des échantillons d'ARN. Une alternative à l'utilisation d'un gène de ménage consiste à ajouter au cours de l'électrophorèse d'ARN totaux (dans le gel ou dans les échantillons) du bromure d'éthidium (BET) et à lire le gel sous UV afin de contrôler l'égalité des dépôts et l'absence de dégradation différentielle des échantillons via l'observation des ARN ribosomiques 18S et 28S. Cette méthode reste néanmoins très qualitative et ne permet pas de discerner une faible dégradation des échantillons d'ARN d'une absence de dégradation.

#### AVANTAGES :

Le principal avantage du NB, dont les autres techniques que nous présentons ici ne bénéficient pas, est la possibilité d'évaluer, à l'aide d'un marqueur de taille approprié, la taille du transcrit étudié. Le NB permet également de détecter la présence éventuelle de transcrits alternatifs et d'évaluer leurs tailles, ceci uniquement si la séquence de la sonde marquée est au moins partiellement partagée par ces transcrits alternatifs. L'utilisation d'autres sondes marquées, ciblant d'autres zones des différents transcrits observés, peut également être utile pour préciser les zones des transcrits alternatifs qui leur sont spécifiques et les zones dont la séquence est commune à tous les transcrits. Le NB est relativement peu sensible à une dégradation partielle et modérée des échantillons. De plus, il présente l'avantage de pouvoir visualiser aisément les échantillons dégradés qui apparaissent généralement sous la forme de bandes dont les contours inférieurs sont moins nets, voire « dégoulinent ». Le NB est une méthode quantitative qui a fait ses preuves dans ce domaine, mais il faut veiller de manière



**Figure 36 : Principe général du dosage par protection aux RNases**

L'ARN cible est illustré en rouge. Une sonde d'ARN simple brin antisens de la cible est synthétisée et marquée (\*) par transcription in vitro (ici à partir d'un plasmide contenant le fragment d'intérêt en aval du promoteur d'une ARN polymérase ADN-dépendante). Cette sonde est hybrideée aux ARN totaux avant leur digestion par des RNases spécifiques des ARNs simple brin. L'abondance du fragment protégé des RNases est ensuite analysée par électrophorèse puis densitométrie. Le schéma illustre l'utilisation d'une unique sonde marquée, sauf pour le gel d'électrophorèse où la possibilité d'analyse simultanée de plusieurs transcrits est illustrée.

générale à ne comparer que des échantillons qui ont été traités en parallèle, sur un même gel d'électrophorèse, en raison des multiples étapes susceptibles d'introduire des biais dans les données. Finalement, le NB est une technique peu coûteuse et certains apprécient, peut-être en partie sur des critères esthétiques, sa capacité à permettre une « visualisation » des transcrits.

#### *INCONVENIENTS / LIMITES :*

Les sondes de northern blot font le plus souvent environ 1Kb. Dans ces conditions, l'idéal est donc de travailler sur un organisme dont le génome a été séquencé pour être sûr, a priori, de la spécificité de la sonde. Dans tous les cas, une séquence au moins partielle du transcript à étudier doit être connue. Dans le cas de familles multigéniques, il peut être difficile de trouver un fragment relativement long (de l'ordre de 1Kb) et en même temps spécifique du gène d'intérêt. On peut alors recourir à une hybridation d'un fragment plus court, à condition d'adapter la stringence du tampon d'hybridation. Il est néanmoins beaucoup plus difficile de travailler avec de très courts fragments en NB et on préfèrera se tourner en général vers les techniques de RPA ou de PCR en temps réel, plus adaptées à discriminer des transcrits dont la séquence est très voisine. La sensibilité du NB est relativement bonne mais elle dépend largement du type de marquage effectué, le radiomarquage étant sans doute le plus sensible et, dans tous les cas, la sensibilité du NB n'égale pas la sensibilité de la PCR en temps réel ou du dosage par protection aux RNases (voir paragraphe suivant). Finalement, le NB est une technique relativement lourde à mettre en œuvre qui nécessite au minimum 2,5 jours de manipulation en partant des échantillons d'ARN et avant d'obtenir une image, et qui conduit souvent à manipuler des produits potentiellement dangereux (BET, formaldéhyde, formamide et radioactivité notamment).

#### **III.1.1. La protection aux RNases**

Bien que le dosage par protection aux RNases (RNase protection assay ou RPA) n'ait pas été utilisé dans le cadre de ce doctorat, il est brièvement présenté ci-dessous car il a été assez largement utilisé pour l'étude de l'expression des récepteurs nucléaires (pour des applications sur les PPARs, voir par exemple Escher et al., 2001; Vidal-Puig et al., 1997).

#### *PRINCIPE GENERAL :*

Le RPA est basé sur l'hybridation en milieu liquide d'une sonde d'ARN antisens simple brin de taille déterminée à un échantillon d'ARN total, suivie de la digestion de la sonde non hybridée et des ARNs non protégés puis de l'analyse sur gel de la bande d'ARN protégée par

la sonde. Le principe général du RPA est illustré en [Figure 36](#) et comporte les étapes suivantes :

- La synthèse d'une sonde d'ARN antisens simple brin marquée (souvent par incorporation d' $[\alpha-^{32}\text{P}]UTP$ ). Cette étape est généralement réalisée par transcription *in vitro* à l'aide des ARN polymérases ADN-dépendantes issus des bactériophages T7 ou SP6 et d'un plasmide contenant un ADNc spécifique cloné en aval du promoteur adéquat (SP6 ou T7). La sonde synthétisée couvre généralement quelques centaines de bases de l'ARN ciblé.
- L'extraction des ARN totaux (l'extraction des ARN polyA+ n'étant généralement pas nécessaire) des échantillons à étudier.
- L'hybridation en milieu liquide de la sonde marquée (en excès) aux échantillons d'ARN totaux.
- Le traitement des échantillons avec des RNases spécifiques des ARN simples brins (RNases A et T1 par exemple) afin de digérer les sondes non hybridées et les ARN qui ne sont pas protégés par la sonde hybridée.
- L'inactivation des RNases et la purification du fragment protégé.
- La séparation du ou des fragment(s) protégé(s) par électrophorèse sur gel dénaturant de polyacrylamide
- La détection et la quantification des fragments protégés des RNases par l'hybridation de la sonde marquée. Si la sonde est radiomarquée, le gel peut être séché et exposé à un écran au phosphore pour une quantification à l'aide d'un imageur au phosphore (scanner type Storm).

Comme pour le northern blot, la quantification d'un « gène de ménage » permet de calibrer l'expression du gène d'intérêt mais avec le RPA, cette quantification peut se faire simultanément pour les deux gènes, ce qui n'est pas toujours possible en northern blot si les tailles des deux transcrits sont trop proches. La synthèse d'une sonde synthétique d'ARN sens de concentration connue peut également être utilisée pour réaliser une quantification absolue (réalisation d'une droite de calibration par dilutions séries).

#### AVANTAGES :

L'utilisation d'une hybridation en milieu liquide rend le RPA plus sensible que le NB et donc plus adapté à la quantification des transcrits peu abondants. De plus, elle permet l'utilisation d'importantes quantités d'ARNtotal de départ (jusqu'à 100 µg) alors qu'en NB,

l'utilisateur est souvent limité par la combinaison volume des puits du gel d'électrophorèse / concentration initiale des ARNs. Le dosage RPA présente une sensibilité inférieure à la PCR en temps réel (voir paragraphe suivant) mais nécessite moins de mises au point et moins d'appareillages coûteux que cette dernière. Les dosages RPA sont relativement peu sensibles à une légère dégradation des ARNs dans la mesure où le fragment protégé est de taille réduite. Cette méthode permet d'analyser simultanément plusieurs transcrits (jusqu'à 5-10 transcrits en même temps) à condition de veiller à choisir des sondes de tailles suffisamment différentes pour éviter les chevauchements lors de la migration des fragments protégés. Par rapport au NB, le RPA est plus apte à discriminer des gènes appartenant à une famille multigénique et migrant éventuellement à des tailles similaires en NB, en raison de l'utilisation de sondes plus courtes et d'une bonne spécificité du dosage. Enfin, la résolution des gels de polyacrylamide permet, en adaptant les sondes RPA (oligonucléotide marqués au  $[\gamma-^{32}\text{P}]$ ATP notamment) et les nucléases utilisées (nucléase S1 notamment), de définir les extrémités d'un transcrit (sites d'initiation et de terminaison de la transcription) ainsi que ses jonctions exon/exon.

#### *INCONVENIENTS / LIMITES :*

Contrairement au NB, le RPA ne peut pas fournir d'indication sur la taille du transcrit analysé. Il peut néanmoins être utilisé pour définir les bornes d'un transcrit d'intérêt, à condition de disposer de la séquence génomique correspondante. La mise au point d'un dosage en RPA comprend certes moins d'étapes qu'en PCR en temps réel mais l'obtention d'une sonde de RPA reste une étape relativement longue qui nécessite la manipulation de matériel sensible à la dégradation (ARN simple brin). De plus, les étapes permettant de valider la spécificité d'une sonde de RPA sont bien plus longues que les étapes nécessaires à la validation de la spécificité d'un couple d'amorces en PCR en temps réel. La mise en œuvre d'un dosage RPA validé comporte de nombreuses étapes (hybridation sur plusieurs heures, digestion, purifications, migration, éventuellement transfert sur membrane) et s'avère plus longue que la réalisation d'un dosage validé en PCR en temps réel (voir paragraphe suivant). Enfin, le RPA est une technique qui présente une bonne spécificité, ce qui la rend cependant sensible à de subtiles différences de séquence entre la sonde et la cible.

#### **III.1.3. La PCR en temps réel**

La PCR en temps réel (real-time PCR ou QPCR pour quantitative PCR) a connu un essor considérable au cours de ces dernières années, sans doute en partie en raison de sa simplicité de mise en œuvre et de la publicité importante qui a été faite sur sa précision et sa

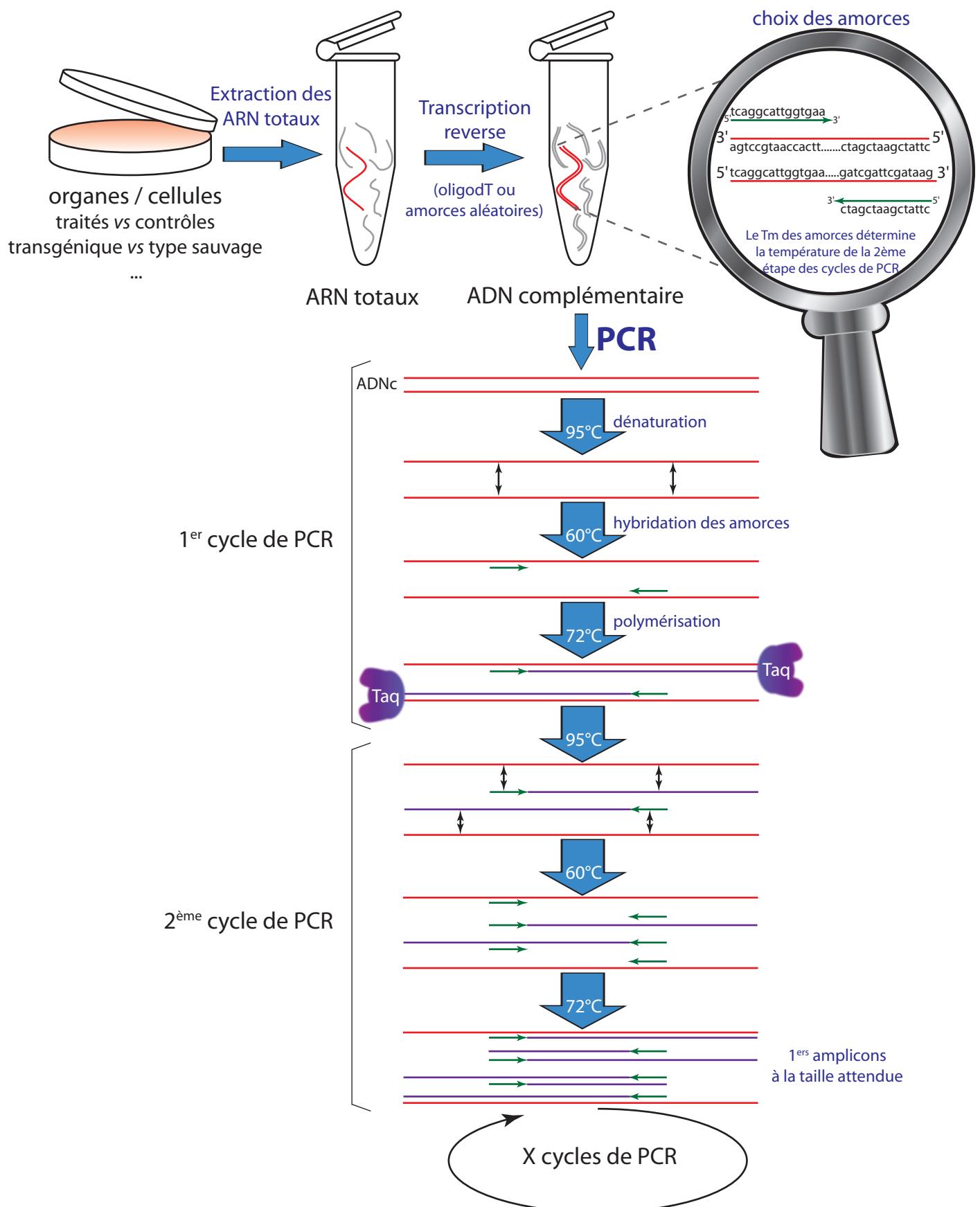
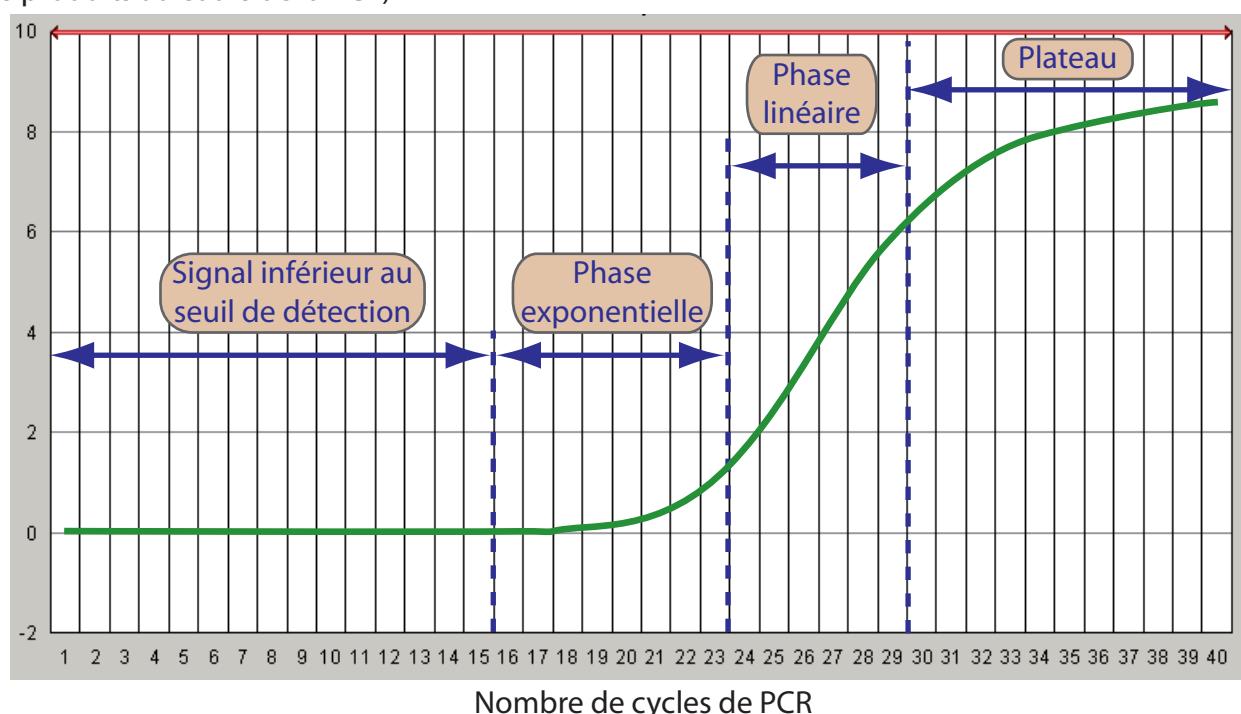


Figure 37 : Principe général de la PCR

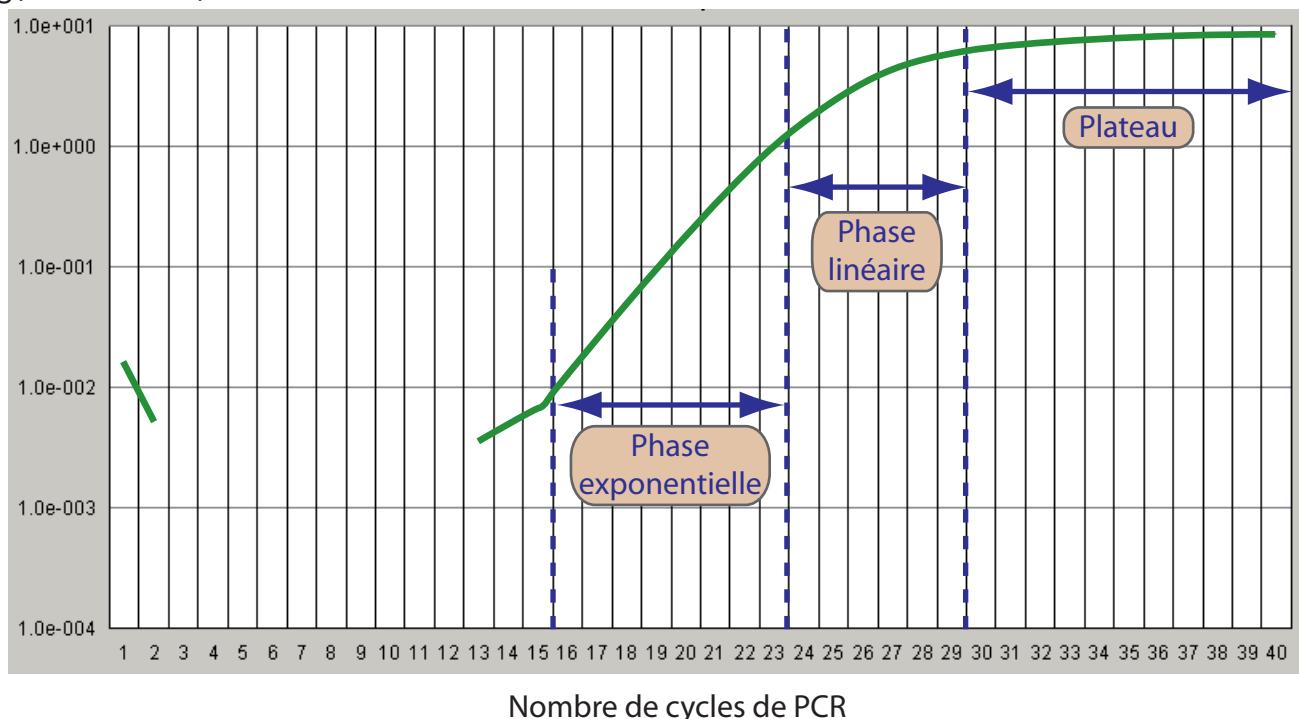
La partie haute du schéma illustre les étapes préliminaires de la PCR, choix des amores compris. Les deux premiers cycles de PCR sont présentés. En général, les cycles de PCR sont précédés d'une dénaturation de quelques minutes à 95°C. Un cycle est constitué de trois étapes : la dénaturation (95°C), l'hybridation des amores (45-65°C à choisir en fonction de la température de fusion ou Tm des amores) et la polymérisation (72°C). PCR : polymerase chain reaction, Taq : polymérase issue de *Thermophilus aquaticus*. En théorie, si aucun réactif n'est limitant, la PCR génère un doublement de la quantité d'amplicon à chaque cycle à partir du 3<sup>ème</sup> cycle.

**A.**

Fluorescence  
(proportionnelle à la quantité  
d'amplicons produits au cours de la PCR)

**B.**

$\log(\text{Fluorescence})$



**Figure 38 : La courbe de PCR**

La courbe de PCR, représentant la quantité d'amplicons produits au cours des cycles de PCR, est présentée en échelle linéaire (A.) et en échelle logarithmique (B.). Différentes phases de production des amplicons sont distinguées sur ces schémas.

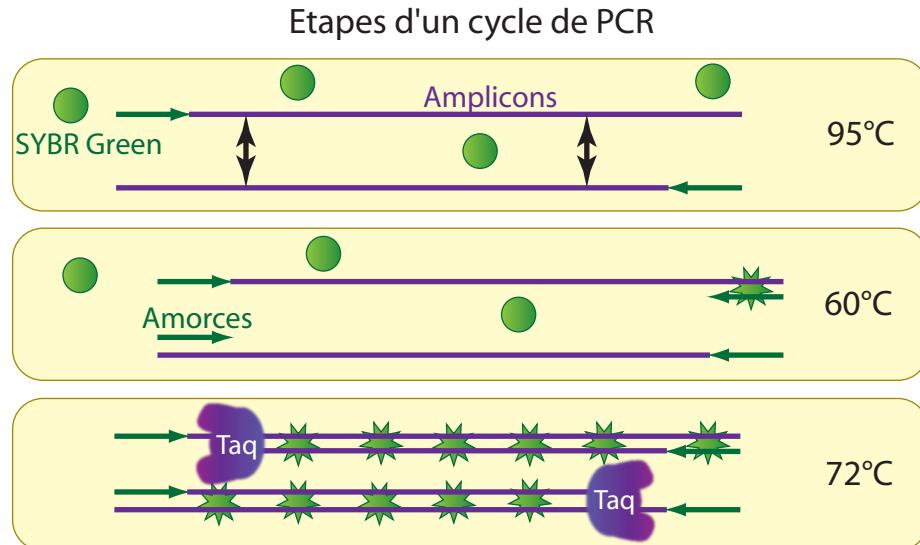
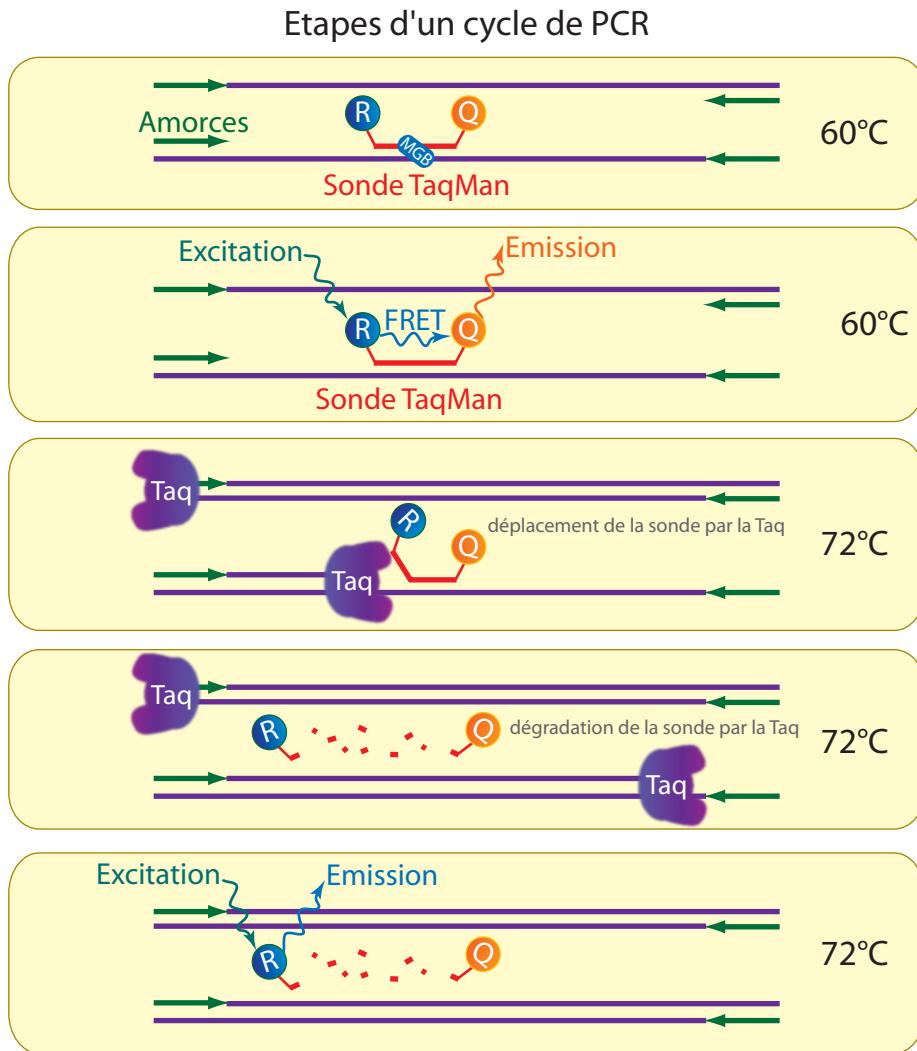
**A.****B.**

Figure 39 : SYBR Green et sondes TaqMan

Le SYBR Green (A.) est un agent intercalant de l'ADN qui émet une intense fluorescence uniquement lorsqu'il se trouve intercalé entre les bases de deux brins d'ADN hybridés. Plus il y a d'amplicons produits, plus il y a de molécules de SYBR Green intercalées et donc plus la fluorescence augmente. La sonde TaqMan (B.) porte deux fluorophores appelés Reporter (R) et Quencher (Q). Le minor groove binder (MGB) augmente artificiellement le  $T_m$  de la sonde TaqMan et assure ainsi sa fixation complète avant celle des amorces de PCR. Lorsque la sonde intacte est excitée, un phénomène de FRET (fluorescence resonance energy transfer) s'opère entre R et Q et l'émission de Q est mesurée (ou aucune émission dans le cas d'un quencher non fluorescent). Lors de la polymérisation, la Taq polymérase déplace et dégrade la sonde TaqMan grâce à son activité 5' exonucléase, conduisant à l'éloignement de R et Q et donc à l'arrêt du phénomène de FRET. L'émission de fluorescence d'un R correspond ainsi à la synthèse de deux nouveaux brins d'ADN.

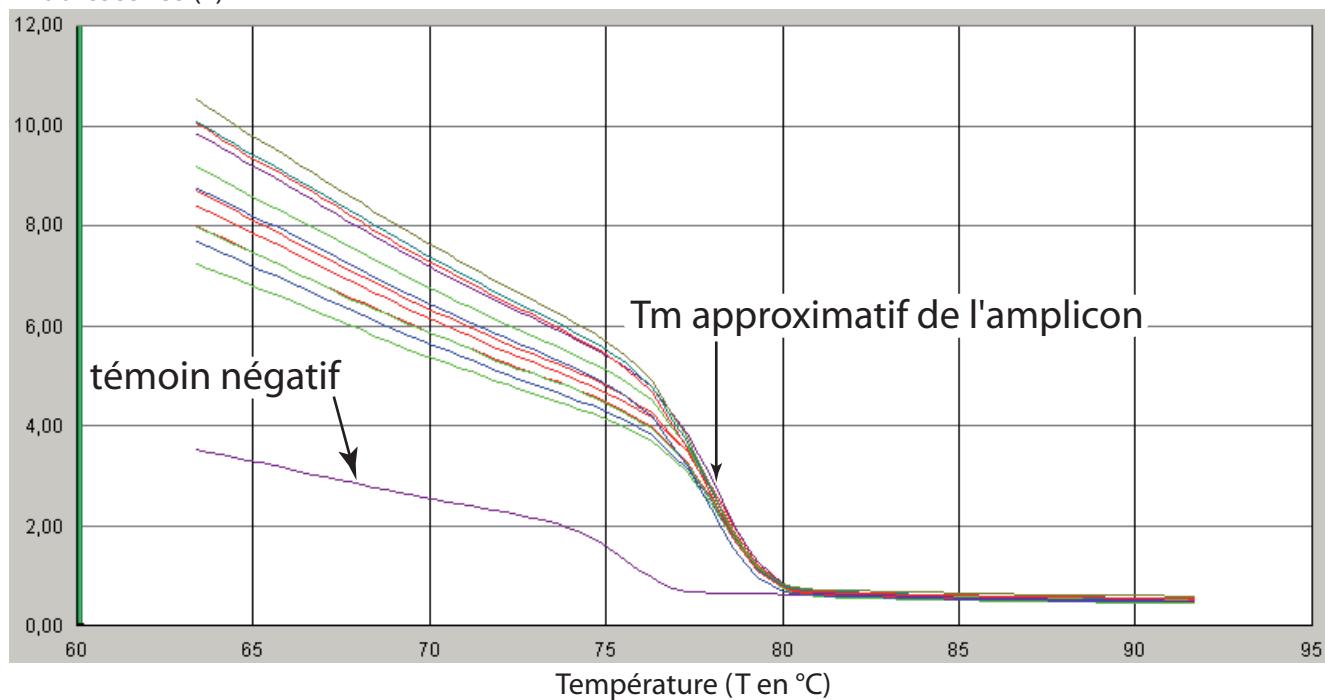
sensibilité. Elle est considérée par de nombreux chercheurs comme la technique de référence de mesure de l'expression des gènes. S'il est vrai que c'est, de loin, la technique à bas débit la plus rapide à réaliser et que sa sensibilité est également la meilleure, je ne connais néanmoins pas d'étude qui ait évalué objectivement et dans des conditions d'utilisation standards par les laboratoires, sa précision par rapport aux autres techniques disponibles. Qualifier cette technique de méthode de référence, risquerait à mon avis de masquer ses quelques défauts et surtout les difficultés qui peuvent être rencontrées par les biologistes pour analyser les résultats qui en sont issus.

#### *PRINCIPE GENERAL :*

Comme son nom l'indique, la PCR en temps réel est basée sur le principe de la PCR ([Figure 37](#)). En fonction du principe présenté en [Figure 37](#), on peut, en théorie, s'attendre, à partir du troisième cycle de la PCR, à observer un doublement de la quantité d'amplicons (fragment d'ADN borné par les deux amorces choisies) à chaque cycle. En réalité (voir [Figure 38](#)), la quantité d'amplicon obtenue à chaque cycle, n'augmente de manière exponentielle que pendant un certain nombre de cycles, ensuite, l'amplification s'amortie (phase « linéaire » de la [Figure 38](#)) en raison de la présence d'un réactif qui commence à être limitant au sein du mélange réactionnel. Cet amortissement de l'amplification se termine par un plateau au cours duquel l'amplification n'a plus lieu en raison de l'absence complète de l'un des réactifs. Le premier principe de base de la PCR en temps réel consiste à permettre une quantification de l'ADN produit dans le tube à chaque cycle et à générer ainsi une courbe de PCR pour chaque échantillon.

Pour atteindre cet objectif, plusieurs techniques ont été développées. Nous ne présenterons ici que deux méthodes ([Figure 39](#)), la première étant sans doute la plus répandue et la seconde ayant également été utilisée dans le cadre de ces travaux de doctorat. La première méthode ([Figure 39A](#)) consiste à utiliser un agent intercalant de l'ADN : le **SYBR Green** dont la fluorescence est très fortement augmentée lorsqu'il est intercalé entre les bases de l'ADN. A chaque cycle de PCR, la quantité d'ADN double brin augmente dans le tube et, pour peu que la mesure de fluorescence dans le tube soit réalisée au moment opportun où les brins d'ADN sont bien associés, le nombre de molécules de SYBR Green intercalées entre les bases de l'ADN augmentera en proportion, fournissant ainsi un signal de fluorescence au cours de la PCR proportionnel au nombre cumulé d'amplicons produits. L'autre méthode ([Figure 39B](#)), développée par Applied Biosystems et nommée **TaqMan**, est basée sur l'utilisation d'une sonde nucléotidique simple brin portant à ses deux extrémités deux

Fluorescence (F)



$$-\frac{dF}{dT}$$

- dérivée

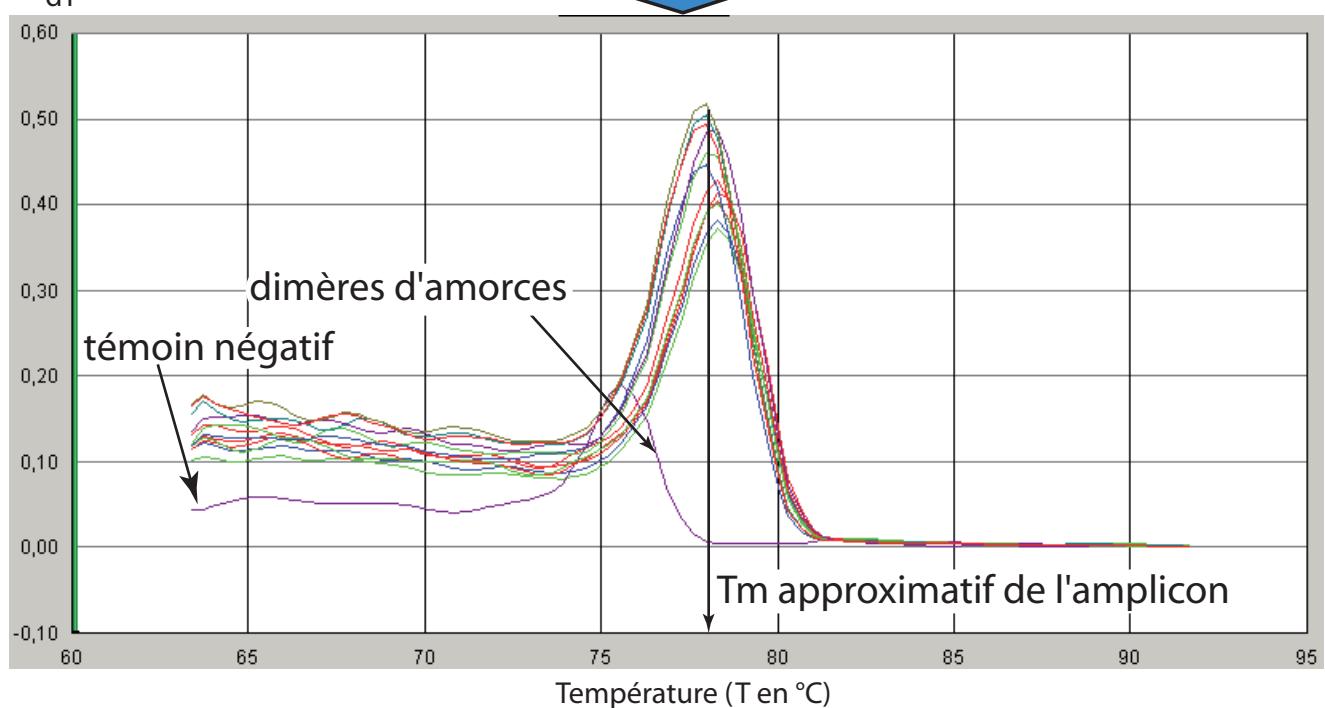


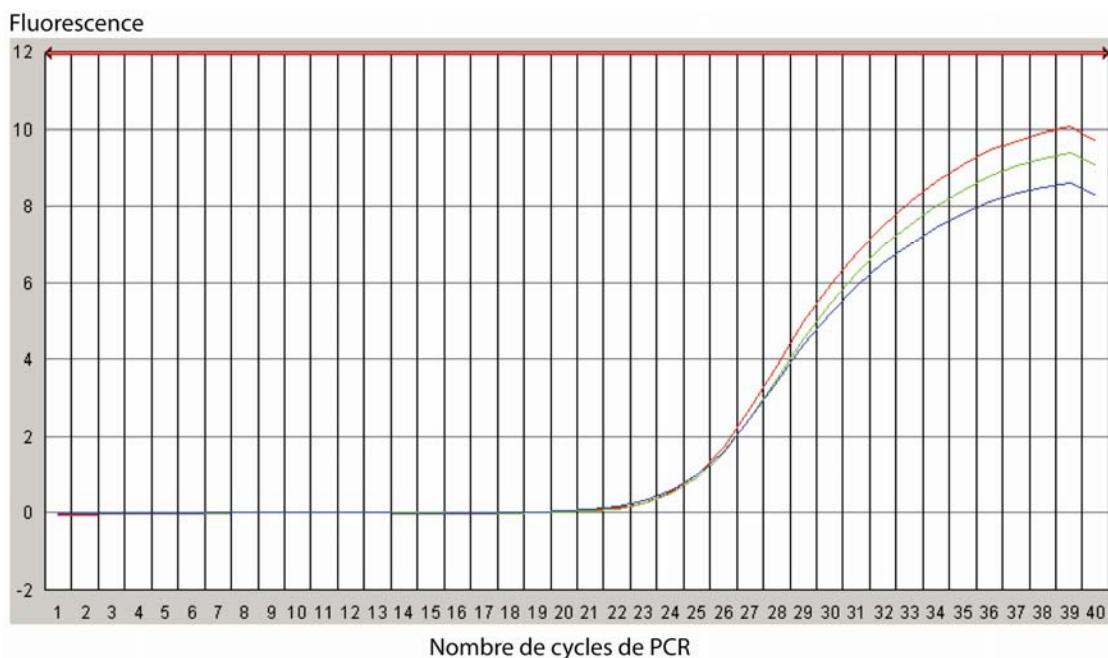
Figure 40 : Courbes de dissociation en SYBR Green

A la fin des cycles de PCR, la température est augmentée progressivement de 60°C à 95°C, provoquant la dénaturation des amplicons qui s'opère principalement autour de la température de fusion (Tm) de l'amplicon (température à laquelle 50% des amplicons sont dénaturés). La représentation de la dérivée négative de la fluorescence donne un pic dont le sommet correspond au Tm de l'amplicon. L'observation de pics précédant le Tm de l'amplicon correspond souvent à la formation de dimères d'amorces (comme présenté ici dans l'échantillon contrôle négatif), alors que l'observation de pics après le Tm de l'amplicon correspond généralement à une amplification d'ADN génomique. L'amplification aspécifique d'un autre fragment d'ADNc peut également donner un pic supplémentaire si son Tm est suffisamment différent du Tm de l'amplicon spécifique. Les données présentées ici correspondent à l'amplification d'un fragment de l'HMG-CoA synthase cytosolique à partir d'ADNc issus de la transcription reverse d'ARN totaux hépatiques de souris. Le contrôle négatif est un pool d'ARN totaux qui n'a pas subi de transcription reverse.

fluorophores appelés Reporter (R) et Quencher (Q). Lorsque la sonde est intacte, un phénomène de FRET (fluorescence resonance energy transfert) s'opère entre ces deux fluorophores judicieusement choisis de manière à ce que la longueur d'onde d'émission de R corresponde à la longueur d'onde d'excitation de Q et que leurs spectres d'excitation et d'émission respectifs se recouvrent le moins possible. Dans ces conditions, si R est excité, sa proximité avec Q va conduire son émission de fluorescence à exciter Q qui va émettre dans sa gamme de longueur d'onde d'émission, non superposée avec celle de R. Des quencher non fluorescent (NFQ) ont été développés de telle sorte qu'aucun signal n'est enregistré lorsque la sonde nucléotidique est intacte. La sonde TaqMan est sélectionnée de manière à ce que sa température de fusion ( $T_m$  pour  $T_{melting}$ ) soit au moins de  $10^{\circ}\text{C}$  supérieure à celle des amorces de PCR. En réalité, la fixation d'une molécule appelée MGB (minor groove binder) à la sonde TaqMan élève artificiellement son  $T_m$ , facilitant ainsi le choix de la sonde sur une gamme de  $T_m$  moins restreinte. Au cours de la deuxième étape d'un cycle de PCR (hybridation des amorces), la sonde TaqMan va se fixer de manière spécifique sur sa séquence complémentaire. Cette fixation sera complète et précèdera toujours la fixation des amorces en raison du  $T_m$  plus élevé de la sonde TaqMan et du fait que l'étape précédente de PCR était à une température supérieure ( $95^{\circ}\text{C}$ ). Lors de l'étape suivante de polymérisation ( $72^{\circ}\text{C}$  pour une Taq polymérase classique), la Taq polymérase va rencontrer sur son chemin la sonde TaqMan et déplacer puis dégrader cette dernière grâce à son activité 5'-exonucléase. Cette dégradation de la sonde TaqMan favorise l'éloignement du reporter et du quencher et élimine ainsi le phénomène de FRET. Lorsque l'étape de polymérisation est complète, l'excitation du reporter se traduit par une émission de fluorescence correspondant au spectre d'émission du reporter. La synthèse de deux nouveaux brins d'ADN, se traduit donc par une augmentation du signal de fluorescence correspondant à l'émission d'une molécule de reporter et le signal enregistré au cours des cycles de PCR est donc bien proportionnel à la quantité d'amplicons synthétisés.

Ces deux méthodes (SYBR Green et sonde TaqMan) ne sont pas parfaitement équivalentes. Le SYBR Green présente l'avantage d'être moins couteux, applicable à tout couple d'amorces et de fournir un signal fort, d'autant plus que la taille des amplicons est importante. Il présente néanmoins le désavantage de n'être pas spécifique de l'amplicon d'intérêt. Si la PCR n'a pas été correctement mise au point, il est possible que les amorces forment des dimères, ou bien que de l'ADN génomique soit amplifié ou encore qu'un amplicon non spécifique soit amplifié en raison d'une faible spécificité des amorces de PCR. Dans ce cas, le signal enregistré ne sera plus proportionnel à la quantité d'amplicons

spécifiques produits et les résultats s'avèreront ininterprétables. Il est donc crucial lorsque l'on travaille en SYBR Green de s'assurer 1) de la spécificité parfaite des amorces, 2) de leur incapacité à amplifier de l'ADN génomique (choix des amorces sur deux exons différents séparés par un intron de taille conséquente) et 3) de leur incapacité à former des dimères (en particulier des dimères 3'-3' qui risquent de provoquer une amplification des amorces par la Taq polymérase). Afin de contrôler ces aspects, il est nécessaire de vérifier l'unicité et la longueur appropriée du produit de PCR par migration sur gel. Une autre possibilité complémentaire consiste à obtenir une **courbe de dissociation** de l'amplicon à la suite de la PCR ([Figure 40](#)). En augmentant progressivement la température dans les tubes de PCR, les amplicons sont progressivement dénaturés à un faible niveau jusqu'à ce que la température s'approche de la température de fusion (Tm) de l'amplicon. À ce moment, la dénaturation des amplicons s'accélère pour atteindre 50% d'amplicons dénaturés lorsque le Tm est atteint. La dénaturation complète des amplicons est ensuite atteinte lorsque la température continue d'augmenter. La dénaturation de l'amplicon provoque la libération du SYBR Green et s'accompagne donc d'une diminution de la fluorescence qui peut être enregistrée ([Figure 40](#)).



[Figure 41](#) : Répliquats techniques en PCR en temps réel

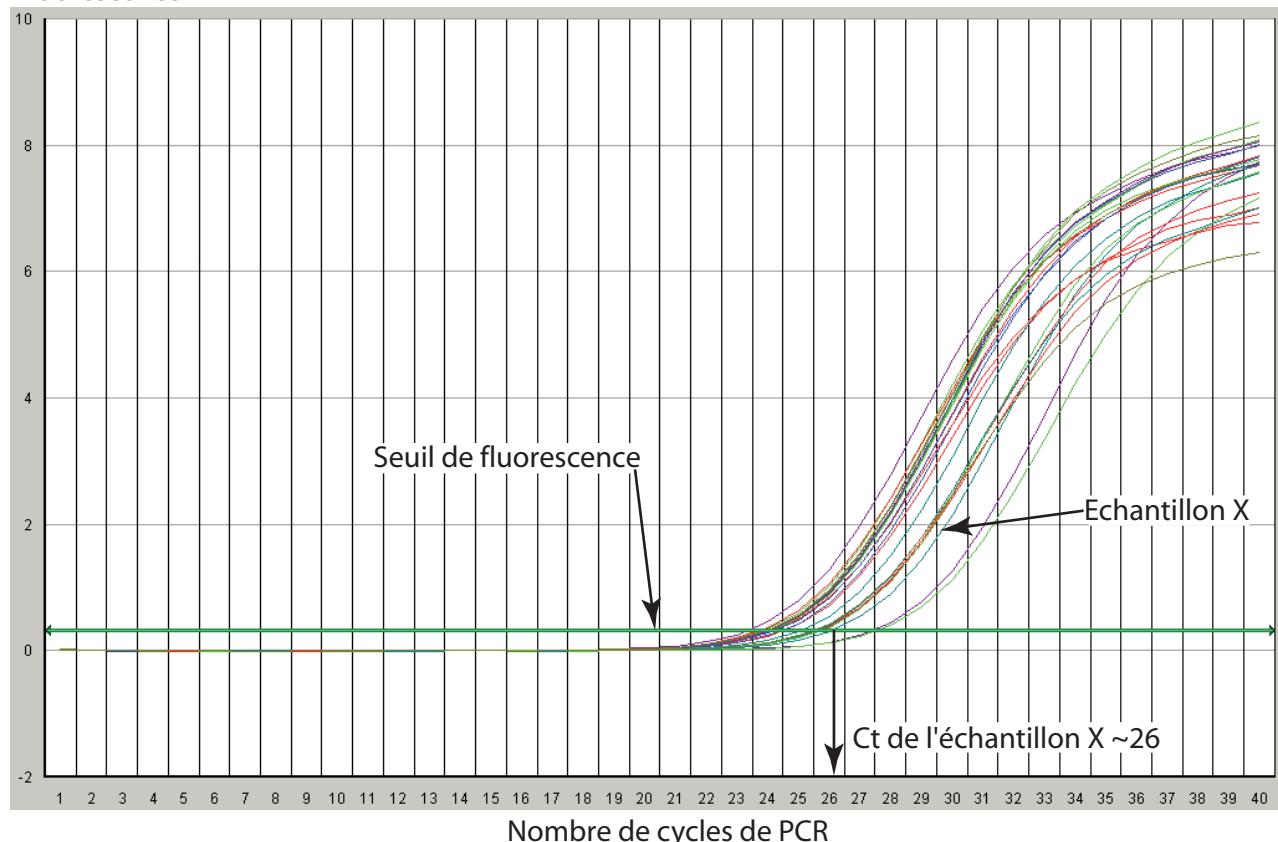
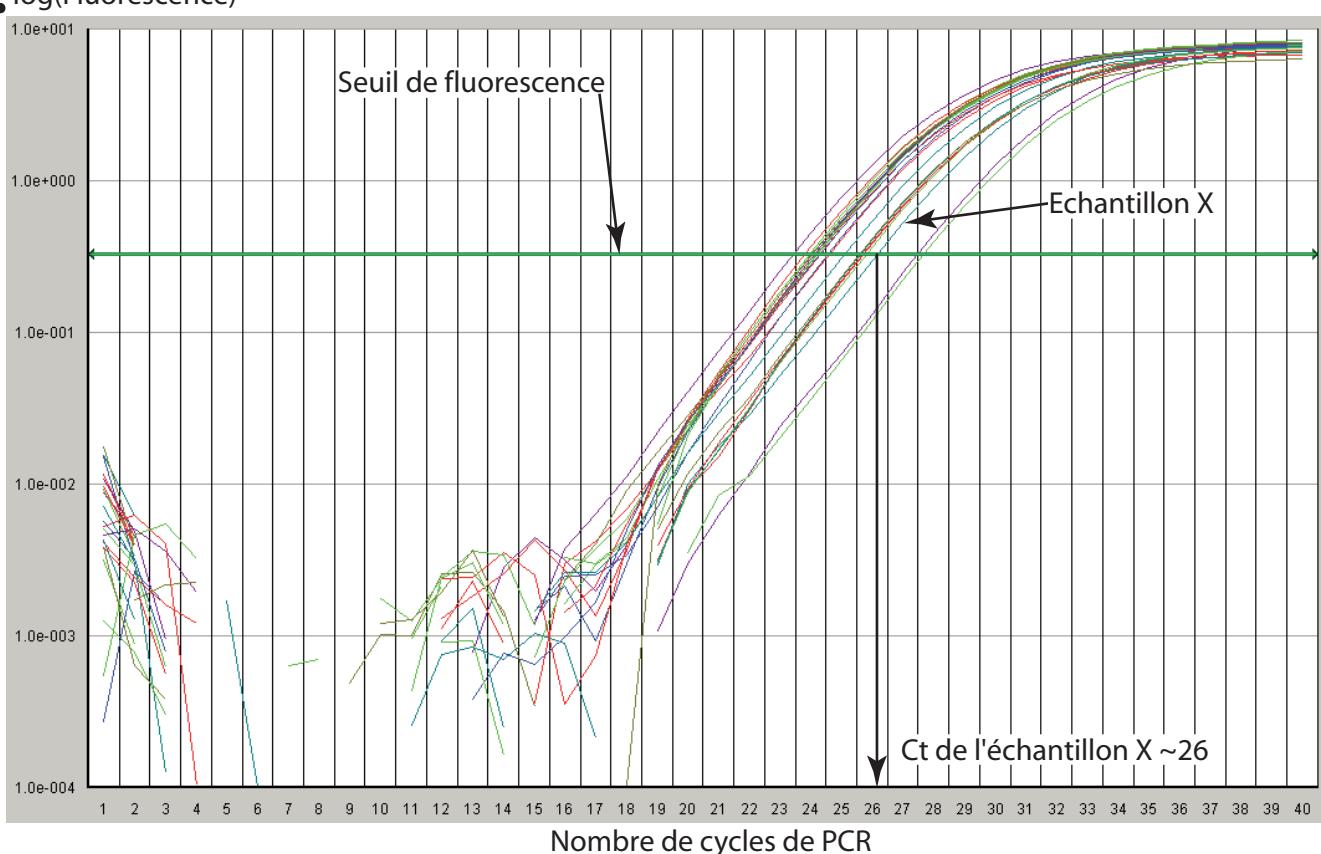
Un même échantillon d'ADNc issu de la transcription reverse d'un ARN total extrait de foie de souris a subi une PCR en temps réel en triplicat avec des amorces spécifiques du gène Abcg5 (méthode SYBR Green). Alors que les trois courbes de PCR sont bien confondues lors de la phase exponentielle d'amplification, on distingue nettement les trois courbes de PCR séparément lors de la phase linéaire et de la phase de plateau. La légère chute de fluorescence observée à la fin de la phase plateau est liée au démarrage d'un protocole de dissociation suite à la PCR (voir [Figure 40](#)).

La présence de dimères d'amorces ou d'amplicons aspécifiques dont le Tm est suffisamment différent du Tm de l'amplicon d'intérêt peut être détectée grâce à cette courbe de dissociation via la présence de pics surnuméraires lors de la représentation de la dérivée négative de la fluorescence ([Figure 40](#)). Contrairement au SYBR Green, si sa séquence a été correctement choisie, la sonde TaqMan assure une parfaite spécificité de quantification de l'amplicon d'intérêt et le choix des amorces est alors un peu moins critique. Les sondes TaqMan sont cependant plus chères et se dégradent plus vite, en particulier lorsqu'elles sont exposées à la lumière. Il en résulte une moins bonne répétabilité des PCRs dans le temps par un effet de dégradation des sondes. Les sondes TaqMan sont également moins sensibles que le SYBR Green car seul un fluorophore émet un signal à chaque synthèse de deux nouveaux brins d'ADN. Signalons également que quelques règles simples doivent impérativement être respectées pour le choix d'une sonde TaqMan comme l'absence de G (guanine) en 5' (cette base risque de quencher la fluorescence du reporter) ou encore un nombre de G inférieur au nombre de C dans la séquence de la sonde (il est toujours possible, le cas échéant, de choisir la sonde sur le brin complémentaire).

En répliquant une même PCR, à partir du même échantillon d'ADNc et des mêmes réactifs (répliqats techniques, voir [Figure 41](#)), on constate que les courbes de PCR générées se superposent bien au niveau de la phase exponentielle mais commencent à diverger les unes des autres à partir de la phase linéaire, les plateaux de ces différentes courbes pouvant également atteindre des niveaux variables. La phase exponentielle est donc l'étape de la PCR à laquelle une quantification fiable peut être réalisée. Dans le cadre de l'étude de l'expression d'un gène, une fois les courbes de PCR générées pour chaque échantillon, on définit à partir de ces courbes la zone (premiers cycles de la PCR) dans laquelle la fluorescence est inférieure au seuil de détection (bruit de fond). Pour chaque courbe, la moyenne des valeurs de fluorescence de cette zone est retranchée aux valeurs brutes, fournissant ainsi des courbes corrigées pour le bruit de fond. Une fois cette étape réalisée, on fixe un seuil de fluorescence qui est choisi de manière à couper toutes les courbes de PCR au niveau de leur phase exponentielle ([Figure 42](#)). L'intersection des courbes de PCR avec ce seuil définit pour chaque échantillon son **cycle seuil** ou **Ct** (pour cycle threshold). C'est cette valeur de Ct qui va être utilisée pour réaliser les calculs nécessaires à l'obtention de données correspondant à l'expression du gène d'intérêt et que nous présentons dans le paragraphe suivant. Comme en NB ou en RPA, on mesure généralement en PCR en temps réel l'expression d'un gène dont l'expression ne varie pas entre les échantillons afin de calibrer les données dans les cas variation de la quantité de matrice de départ (défaut de dosage, dégradation différentielle,

**A.**

## Fluorescence

**B.** $\log(\text{Fluorescence})$ **Figure 42 : Définition des cycles seuils (Ct)**

Une fois les courbes de PCR générées pour chaque échantillon (représentées en échelle linéaire en A et en échelle logarithmique en B), un seuil de fluorescence est défini de manière à couper toutes les courbes au niveau de leur phase exponentielle. La définition du seuil est plus facile en échelle logarithmique (B) où la phase exponentielle apparaît comme une phase linéaire. L'intersection entre chaque courbe de PCR et ce seuil définit, pour chaque échantillon, son cycle seuil (ou Ct pour cycle threshold). Les données présentées sont issues d'une étude de la répression de la gulonolactone L-oxydase (Gulo) par le fénofibrate (sonde TaqMan).

défaut de pipetage, etc.). De plus, les données sont également automatiquement calibrées par rapport à un fluorophore ajouté dans le mélange réactionnel de PCR ce qui permet de calibrer les données en cas de volumes réactionnels légèrement différents d'un tube à l'autre ou de variations spatiales dans la qualité de la mesure de fluorescence.

#### *ANALYSE DES DONNEES DE PCR EN TEMPS REEL :*

Dans le cadre de ce paragraphe, nous nous intéresserons uniquement à l'analyse de données de PCR en temps réel pour la quantification (relative ou absolue) de l'expression des gènes.

Au niveau de la phase exponentielle de la PCR, la courbe de PCR peut être décrite par l'équation suivante :

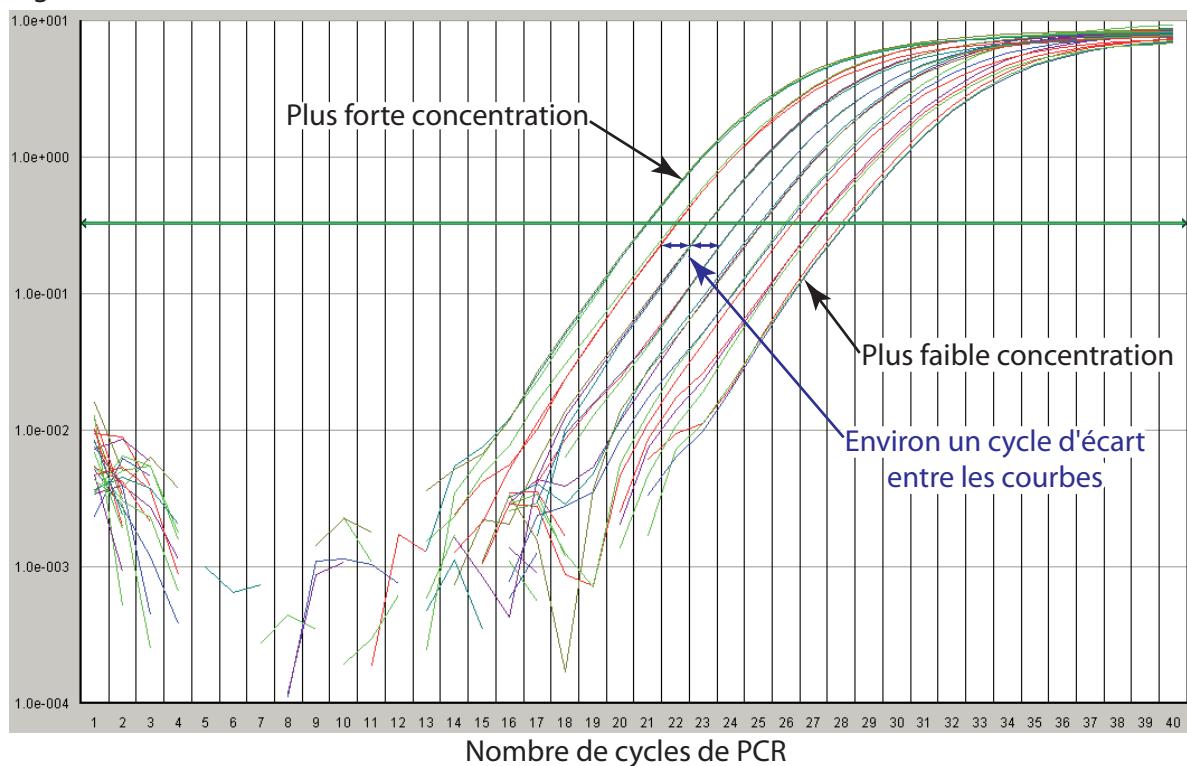
$$X_n = X_0 \times E^n \quad (\text{Eq 1})$$

Où :

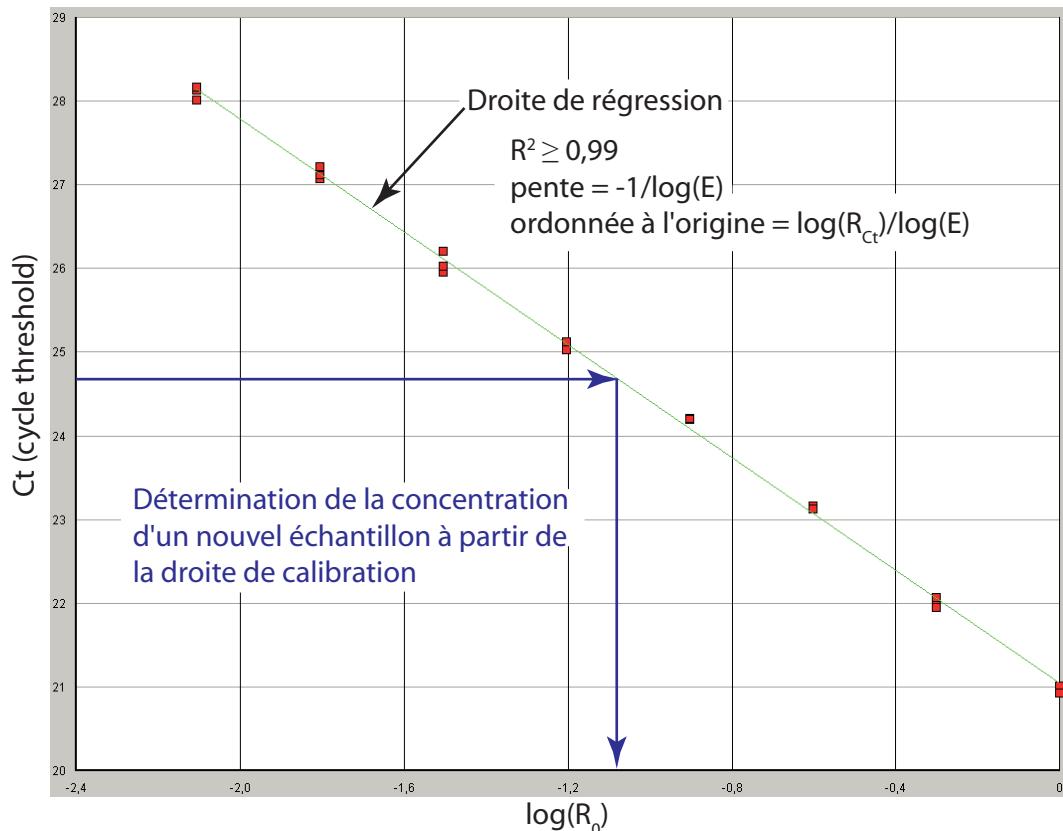
- **n** est le nombre de cycle de PCR (en théorie on devrait plutôt écrire n-2 en raison de la particularité des deux premiers cycles de PCR qui ne produisent qu'un seul amplicon à la taille attendue par transcript initialement présent comme présenté en [Figure 37](#), mais cette approximation ne change rien au raisonnement et aux calculs ultérieurs)
- **X<sub>n</sub>** est le nombre d'amplicons présents dans le tube au cycle n
- **X<sub>0</sub>** est le nombre de copie initiale du transcript d'intérêt et constitue donc la quantité qui nous intéresse en premier lieu.
- **E** est l'efficacité de PCR, comprise entre 1 et 2. Si E=1, l'efficacité est nulle et aucune amplification ne se produit. Si E=2, l'efficacité est de 100% et l'équation 1 donne  $X_{n+1}=2.X_n$  ce qui signifie que, comme la théorie l'indique ([Figure 37](#)), on a bien un doublement de la quantité d'amplicon à chaque cycle de PCR. Dans certains cas, au lieu de considérer E compris entre 1 et 2, on écrit (1+E) où E est compris entre 0 et 1 et représente réellement l'efficacité de PCR. Par soucis de simplification, j'ai choisi la première notation qui limite la taille des formules.

Comme nous l'avons vu précédemment, le seuil de fluorescence ayant été choisi de manière à couper les courbes de PCR au niveau de leur phase exponentielle et l'intensité de fluorescence mesurée étant bien proportionnelle à la quantité d'amplicon produit, l'équation 1

### A. log(Fluorescence)



### B.



**Figure 43 : Utilisation d'une courbe standard en PCR en temps réel**

Un pool d'ADNc issus de la transcription reverse d'ARNs hépatiques de souris a subi une dilution initiale au 1/5ème puis une dilution en série (7 dilutions successives) au 1/2. Une PCR en temps réel sur le gène de la gulonolactone L-oxydase (Gulo) a été réalisée en triplicat sur chacune des huit dilutions. Les courbes de PCR (échelle logarithmique) et le seuil fixé sont illustré en A. On constate qu'il y a environ un cycle d'écart entre chaque dilution au 1/2. Pour la réalisation de la droite de calibration (ou courbe standard, B.), la concentration  $R_0$  de la première dilution (concentration la plus élevée) a été arbitrairement fixée à la valeur 1. La gamme de calibration couvre ici un peu plus de deux ordres de grandeurs mais il est généralement préférable de couvrir au moins 4 ou 5 ordres de grandeurs afin d'avoir la meilleure estimation possible de la pente de la droite de calibration

peut être exprimée en intensité de fluorescence et est valable au niveau du Ct et peut donc s'écrire comme suit :

$$R_{Ct} = R_0 \times E^{Ct} \quad (\text{Eq 2})$$

Où :

- $R_{Ct}$  est la fluorescence sélectionnée comme seuil
- $R_0$  est la fluorescence initiale, non mesurable car située en dessous du seuil de détection
- $E$  est toujours l'efficacité de PCR comprise entre 1 et 2
- $Ct$  est la fraction de cycle à laquelle la courbe de PCR croise le seuil

En retournant l'équation 2 pour exprimer  $R_0$  en fonction de  $R_{Ct}$ , on obtient :

$$R_0 = R_{Ct} \times E^{-Ct} \quad (\text{Eq 3})$$

Et en passant au logarithme on obtient

$$\log(R_0) = \log(R_{Ct}) - Ct \times \log(E) \quad (\text{Eq 4})$$

L'équation 4 illustre le lien linéaire qui existe en le  $\log(R_0)$  et le  $Ct$  et qui est à la base de la construction de **droites de calibration** (ou **courbe standard**) en PCR en temps réel. On peut ainsi construire des droites de calibration en unités relatives comme présenté en [Figure 43](#) ou en unité absolue si l'on dispose d'un standard de concentration connue (idéalement un ARN produit par transcription in vitro). A partir de ces courbes, on estime par régression linéaire les valeurs de la pente et de l'ordonnée à l'origine (et donc de  $R_{Ct}$  et  $E$ ) que l'on utilise ensuite pour calculer les fluorescences initiales  $R_0$ , proportionnelles aux nombres initiaux de copies  $X_0$  du transcript d'intérêt. En appliquant cette méthode au gène d'intérêt et au contrôle endogène (« gène de ménage »), on en déduit aisément la quantité qui nous intéresse à savoir :

$$\frac{R_{0^{\text{target}}}}{R_{0^{\text{endo}}}} = \frac{T_0}{R_0}$$

Où « target » et « endo » représentent respectivement le gène d'intérêt et le contrôle endogène (chacun des gènes devant bien sûr avoir sa propre courbe standard).

L'utilisation de courbes standard est basée sur un principe simple, compréhensible par le plus grand nombre et les manipulations comme les calculs sont relativement faciles à réaliser. De plus, cette méthode permet de gérer le cas où l'efficacité de PCR pour le gène d'intérêt est différente de l'efficacité de PCR pour le contrôle endogène. Néanmoins, l'utilisation de courbes standard présente également quelques limites et inconvénients. Tout d'abord elle fait réaliser un grand nombre de PCR qui n'apportent aucune information scientifique pertinente à l'expérimentateur. Ensuite, on peut rencontrer des problèmes d'estimation des paramètres de la droite de régression (pente et ordonnée à l'origine) comme lors de l'utilisation d'une gamme de concentration trop restreinte (ce qui est courant avec des gènes faiblement exprimés pour lesquels les fortes dilutions atteignent rapidement des Ct élevés et peu reproductibles), l'utilisation d'un nombre insuffisants de réplicats et la présence de points aberrants ou encore la présence d'inhibiteurs de PCR dans les échantillons qui affectent différemment l'efficacité de PCR en fonction de leur dilution. Enfin, l'utilisation d'une courbe standard ne permet pas de repérer les échantillons qui présenteraient une efficacité de PCR aberrante dans la mesure où cette méthode est basée implicitement sur l'hypothèse, relativement forte, que l'efficacité de PCR est la même dans tous les tubes.

Une autre méthode qui s'affranchit de la réalisation d'une courbe standard a été proposée et est connue sous le nom de **méthode du  $2^{-\Delta\Delta Ct}$**  (Livak and Schmittgen, 2001). Nous allons voir que cette méthode est également basée sur l'hypothèse que l'efficacité de PCR est la même dans tous les tubes et, de plus, que cette efficacité de PCR est la même pour le gène d'intérêt et pour le contrôle endogène et est égale à 100% dans les deux cas. Cette méthode présente donc des hypothèses plus fortes que la méthode des courbes standard mais limite la réalisation de ces courbes uniquement à la phase de mise au point (pour s'assurer, au minimum, que les hypothèses avancées sont plausibles). En réécrivant l'équation 3 pour le gène d'intérêt (target) et pour le contrôle endogène (endo) on obtient :

$$T_{Ct_{\text{target}}} = T_0 \cdot E_{\text{target}}^{Ct_{\text{target}}} \text{ pour le gène d'intérêt}$$

$$R_{Ct_{\text{endo}}} = R_0 \cdot E_{\text{endo}}^{Ct_{\text{endo}}} \text{ pour le contrôle endogène}$$

Et en divisant les deux équations ci-dessus l'une par l'autre, on obtient :

$$\frac{T_{Ct_{\text{target}}}}{R_{Ct_{\text{endo}}}} = \frac{T_0}{R_0} \cdot \frac{E_{\text{target}}^{Ct_{\text{target}}}}{E_{\text{endo}}^{Ct_{\text{endo}}}}$$

$T_{Ct}$  et  $R_{Ct}$  sont les seuils de fluorescence qui ont été fixés pour le gène d'intérêt et pour le contrôle endogène respectivement. Ce sont des constantes, leur rapport est donc une constante et si ces seuils de fluorescence ont été choisis égaux pour les deux gènes (ce qui est généralement possible mais pas toujours) alors leur rapport vaut 1 ce qui se traduit par :

$$\frac{T_{Ct_{target}}}{R_{Ct_{endo}}} = \frac{T_0}{R_0} \cdot \frac{E_{target}^{Ct_{target}}}{E_{endo}^{Ct_{endo}}} = K \quad (=1) \quad (\text{Eq 5})$$

Où  $K$  est une constante, éventuellement égale à 1.

De l'équation 5, on peut extraire la quantité qui nous intéresse ( $T_0/R_0$ ), ce qui donne :

$$\frac{T_0}{R_0} = K \cdot \frac{E_{endo}^{Ct_{endo}}}{E_{target}^{Ct_{target}}} \quad (\text{Eq 6})$$

Si les efficacités de PCR pour les deux gènes mesurés sont égales, i.e. si  $E_{target}=E_{endo}=E$ , alors l'équation 6 devient :

$$\frac{T_0}{R_0} = K \cdot E^{Ct_{endo} - Ct_{target}} = K \cdot E^{-\Delta Ct} \quad (\text{Eq 7})$$

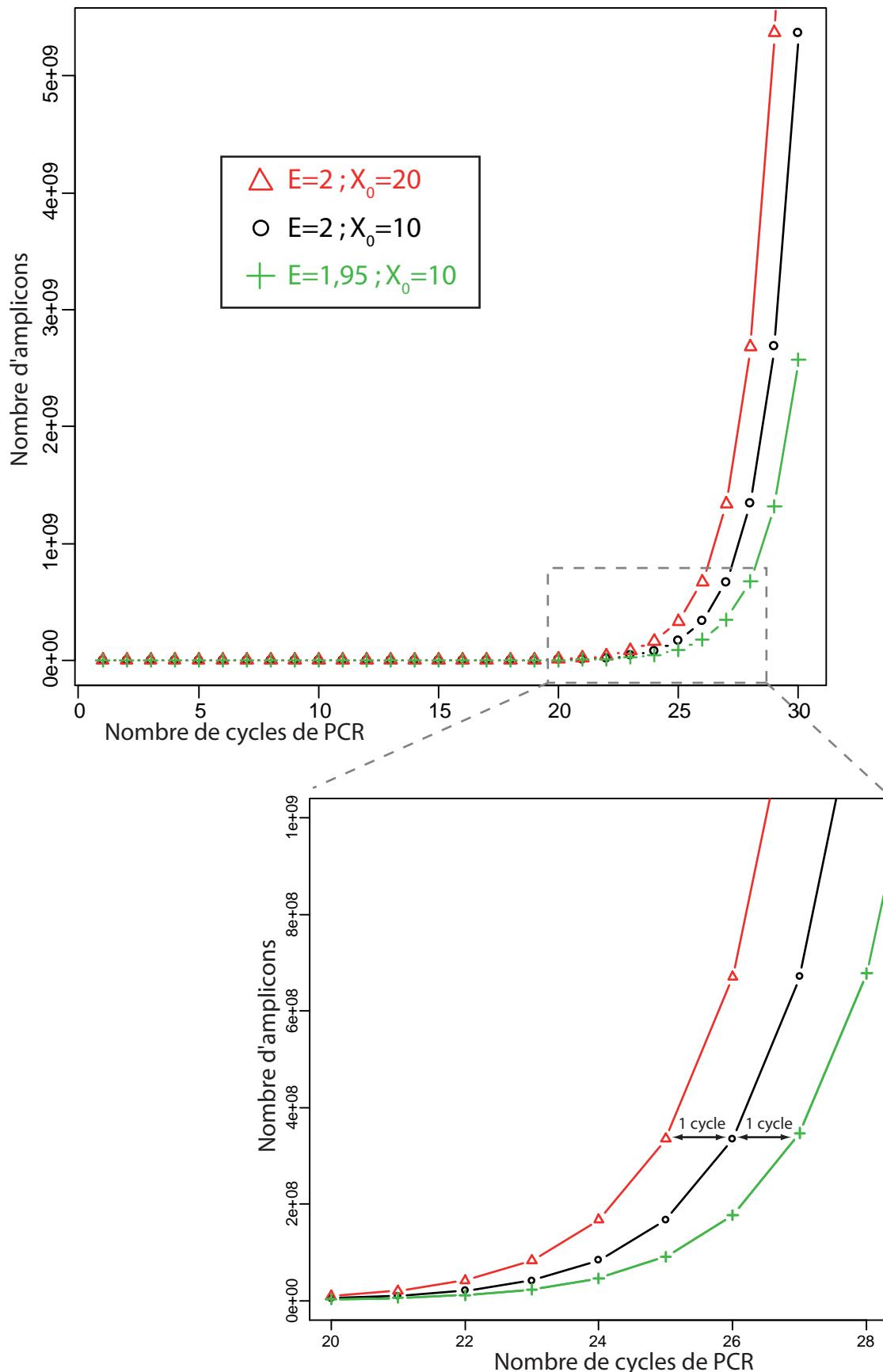
Où  $\Delta Ct = Ct_{target} - Ct_{endo}$

Si, de plus, l'efficacité de PCR vaut 100% et donc que  $E=2$ , alors l'équation 7 devient :

$$\frac{T_0}{R_0} = K \cdot 2^{-\Delta Ct} \quad (\text{Eq 8})$$

Ce qui signifie que l'on peut calculer le rapport  $T_0/R_0$  à une constante près et que cette dernière vaut 1 si les seuils de fluorescence ont été choisis égaux pour les deux gènes mesurés. En utilisant ce calcul (Eq 8) pour chaque échantillon, on obtient donc une mesure de l'expression de notre gène d'intérêt calibré par rapport au contrôle endogène

Dans le cas particulier où l'on dispose de données appariées et de seulement deux groupes d'échantillons (par exemple un plan expérimental de type avant/après une exposition à une perturbation quelconque), alors on peut utiliser la méthode présentée dans (Livak and



**Figure 44 : Effets de l'efficacité sur la courbe de PCR**

A partir de l'équation de la courbe de PCR au cours de la phase exponentielle ( $X_n = X_0 \cdot E^n$ ), trois séries de données ont été générées. Pour une efficacité de PCR de 100% ( $E=2$ ), le nombre de transcript initiaux ( $X_0$ ) a été fixé à 10 (courbe noire) ou à 20 (courbe rouge), conduisant à un écart d'un cycle entre les courbes de PCR rouge et noire. Pour 10 transcripts initiaux ( $X_0=10$ ), l'efficacité de PCR a été fixée à 100% ( $E=2$ , courbe noire) ou à 95% ( $E=1,95$ , courbe verte). On constate qu'au bout de 25-27 cycles, une baisse de 5% de l'efficacité conduit également à un écart d'un cycle entre les courbes de PCR noire et verte (voir l'agrandissement de la zone 20-28 cycles).

Schmittgen, 2001) dont le principe consiste à diviser l'équation 8 pour le groupe « après » par l'équation 8 pour le groupe « avant » pour obtenir :

$$\frac{\left(\frac{T_0}{R_0}\right)_{\text{après}}}{\left(\frac{T_0}{R_0}\right)_{\text{avant}}} = \frac{2^{-\Delta Ct_{\text{après}}}}{2^{-\Delta Ct_{\text{avant}}}} = 2^{-(\Delta Ct_{\text{après}} - \Delta Ct_{\text{avant}})} = 2^{-\Delta\Delta Ct} \quad (\text{Eq 9})$$

Les principaux avantages de la méthode du  $2^{-\Delta\Delta Ct}$  (ou du  $2^{-\Delta Ct}$ ) sont la rapidité des calculs et la facilité à les automatiser ainsi que le fait que les hypothèses qu'elle postule conduisent à ne réaliser qu'une seule fois les courbes standard. Ces hypothèses sont néanmoins très fortes comme nous l'avons souligné, ce qui limite l'applicabilité de la méthode. L'approximation d'une efficacité de 95% par une efficacité de 100% a des conséquences importantes comme l'illustre un exemple simple en [Figure 44](#) ce qui souligne le fait que la vérification expérimentale des hypothèses de cette méthode doit faire l'objet d'une attention rigoureuse. Enfin, comme pour les courbes standard, cette méthode ne permet pas de détecter la présence d'un échantillon qui présenterait une efficacité de PCR aberrante ou extrême.

L'hypothèse que l'efficacité de PCR est la même dans les différents tubes, hypothèses sous-jacente aux deux méthodes que nous venons de présenter, a rapidement fait l'objet de remises en cause, surtout dans la mesure où, comme illustré en [Figure 44](#), une faible variation de l'efficacité (5%) peut avoir des conséquences non négligeables sur la quantification (dans notre exemple, un facteur de 2, soit un cycle d'écart entre les courbes de PCR au bout de 25-27 cycles). De plus, l'enregistrement des données permettant de générer les courbes de PCR (intensité de fluorescence à chaque cycle et pour chaque échantillon) autorise l'**estimation de l'efficacité de PCR pour chaque échantillon** comme nous allons le voir à présent. Si l'on ne considère que la phase exponentielle de la PCR et comme la fluorescence mesurée est bien proportionnelle à la quantité d'amplicons produits, on peut remplacer X (nombre d'amplicons) par R (mesure de fluorescence) dans l'équation 1 pour obtenir :

$$R_n = R_0 \times E^n \quad (\text{Eq 10})$$

Où :

- **n** est le nombre de cycle (n appartenant à la phase exponentielle de la PCR)

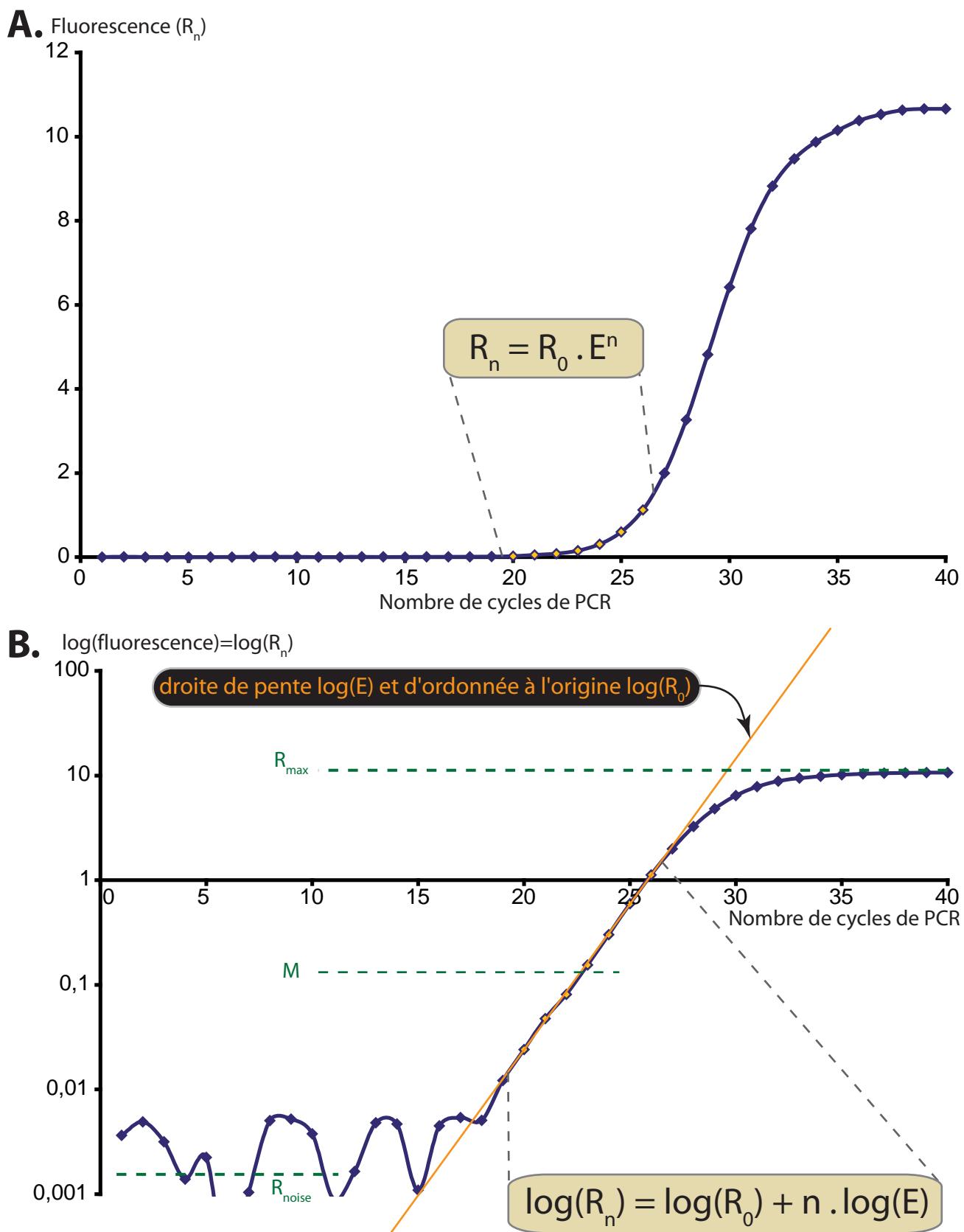


Figure 45 : Estimation de l'efficacité à partir d'une courbe de PCR

Une courbe de PCR est présentée en échelle linéaire (A.) et en échelle logarithmique (B.). Une série de 7 points de mesure appartenant vraisemblablement à la phase exponentielle (sélection manuelle des points) a été colorée en jaune. L'équation de la courbe de PCR au niveau de la phase exponentielle est indiquée sur les deux graphiques. En échelle logarithmique (B.), il est possible d'ajuster une droite sur les points de mesure appartenant à la phase exponentielle dont les estimations de la pente et de l'ordonnée à l'origine fournissent respectivement des estimations de l'efficacité  $E$  et de la fluorescence initiale  $R_0$ . Les traits pointillés verts illustrent la méthode proposée par Peirson et al. (NAR, 2003, 31:e73) pour déterminer le milieu de la phase exponentielle.

- $R_n$  est la fluorescence mesurée au cycle n
- $R_0$  est la fluorescence initiale, non mesurable car située en dessous du seuil de détection
- E est l'efficacité comprise entre 1 et 2

En transformant l'équation 10 en logarithme, on obtient :

$$\log(R_n) = \log(R_0) + n \times \log(E) \quad (\text{Eq 11})$$

L'équation 11 illustre le lien linéaire qui existe entre le log de la fluorescence et le nombre de cycles au niveau de la phase exponentielle. En représentant une courbe de PCR en échelle logarithmique, ce lien est évident comme illustré en [Figure 38](#) ou en [Figure 45](#). Il est donc possible, si l'on connaît les bornes exactes de la phase exponentielle de la PCR, d'ajuster, par régression linéaire par exemple, une droite sur les données de log(fluorescence) en fonction des cycles comme illustré en [Figure 45](#). L'estimation de la pente d'une telle droite donnerait une estimation de  $\log(E)$  et donc de l'efficacité et l'estimation de l'ordonnée à l'origine de cette droite fournirait même directement une estimation de  $\log(R_0)$ , donc de la fluorescence initiale, non mesurable, qui est proportionnelle à la quantité initiale de transcrit  $X_0$ . L'ajustement d'une telle droite ne posant aucun problème en soit, la problématique se résume donc à identifier quels sont les points de mesure qui appartiennent effectivement à la phase exponentielle et donc qui suivent l'équation n°11. Plusieurs approches ont été proposées dans la littérature pour détecter de manière automatique les points de mesure appartenant à la phase exponentielle de la courbe de PCR. L'une d'entre elle a donné lieu au développement d'une macro Excel nommée LinRegPCR (Ramakers et al., 2003). L'algorithme itératif proposé par Ramakers et al. est relativement élaboré et propose de détecter par deux méthodes différentes le point initial de la phase exponentielle et le dernier point appartenant à cette phase afin de définir une « fenêtre de linéarité » contenant au minimum 4 points de mesure et au maximum 6 et fournissant la droite ajustée dont le coefficient de détermination  $R^2$  est maximum et dont la pente est proche du maximum possible (Ramakers et al., 2003). Sans rentrer dans trop de détails techniques, la détection du point initial de la phase exponentielle est basée sur une série d'ajustements linéaires réalisés sur les points initiaux de la PCR (premiers cycles de la PCR) dont les points de mesure sont rajoutés dans le modèle de manière itérative et dont l'influence sur la pente de la droite de régression est mesurée par le biais d'un test sur les résidus studentisés du modèle. La détection du dernier point de la phase exponentielle est, elle, basée sur l'ajustement d'un modèle à quatre paramètres sur la courbe de PCR dont le maximum de la dérivée seconde

indique la fin de la phase exponentielle. Cette méthode relativement compliquée semble donner de bons résultats dans la majeure partie des cas mais une inspection visuelle de la sélection des points et des ajustements des droites restent préconisées par les auteurs (Ramakers et al., 2003). Une autre méthode, bien plus simple et basée sur des observations empiriques a cependant été proposée (Peirson et al., 2003) et semble également donner de bons résultats. Peirson *et al.* (Peirson et al., 2003) proposent de définir le milieu de la phase exponentielle comme étant situé, sur la courbe de PCR en échelle logarithmique, à mi-distance entre le plateau et le bruit de fond ([Figure 45B](#)). Le bruit de fond ( $R_{noise}$ ) est estimé via l'écart-type des signaux issus des dix premiers cycles de PCR et le plateau ( $R_{max}$ ) via le maximum de fluorescence enregistré. Le point M situé entre ces deux valeurs est donc calculé via la formule :

$$M = R_{noise} + \frac{1}{2}(R_{max} - R_{noise}) \quad (\text{Eq 12})$$

Comme les données sont en fait en logarithme, l'équation 12 s'écrit en réalité :

$$\log(M) = \log(R_{noise}) + \frac{1}{2}(\log(R_{max}) - \log(R_{noise})) \quad (\text{Eq 13})$$

Ce qui se simplifie par la formule fournie par les auteurs (Peirson et al., 2003) :

$$M = R_{noise} \times \sqrt{\frac{R_{max}}{R_{noise}}} \quad (\text{Eq 14})$$

Une fois, le milieu de la phase exponentielle défini, les auteurs proposent de sélectionner tous les points dont l'intensité est comprise dans une gamme couvrant un ordre de grandeur autour du point M en imposant la sélection de trois points de mesure au minimum. Cette méthode a également fait l'objet du développement d'un classeur Excel dédié à l'analyse des données de PCR en temps réel et dénommé DART-PCR (Peirson et al., 2003). Bien que ces deux méthodes fournissent des résultats qui semblent le plus souvent satisfaisants d'après notre expérience, l'inspection individuelle des courbes de PCR reste nécessaire dans tous les cas pour s'assurer de la sélection correcte des points de mesure. Une fois les points sélectionnés, une régression linéaire est effectuée sur ces points fournissant ainsi des valeurs de  $R_0$  et de E. L'intérêt évident de ces méthodes est de permettre à l'utilisateur d'identifier des échantillons présentant une efficacité de PCR aberrante ou du moins extrême par rapport aux autres échantillons (présence éventuelle d'inhibiteurs de PCR, problème dans la manipulation des échantillons, etc.). Finalement, la question se pose de savoir si les estimations individuelle de E (ou de  $R_0$ ) doivent être utilisées telles quelles et

donc si elles sont fiables. L'utilisation directe de ces valeurs a été proposée par Ramakers *et al.* (Ramakers et al., 2003) mais les essais que nous avons effectués sur des régulations d'expression provoquées par le fénofibrate et bien décrites dans la littérature (données non présentées) nous ont montré que la fiabilité des estimations de E (et de R<sub>0</sub>) est en réalité très faible, ce à quoi on pouvait s'attendre en considérant le faible nombre de points de mesure (entre 3 et 6-8 points généralement) utilisés pour l'ajustement de la droite de régression. L'utilisation directe des valeurs individuelles conduit à une explosion de la variabilité des données, incompatible avec la détection de la moindre régulation d'expression génique. L'approche proposée par Peirson *et al.* (Peirson et al., 2003) nous semble en revanche plus pragmatique. Les estimations individuelles de l'efficacité de PCR sont utilisées pour tester par analyse de variance s'il existe ou non des différences d'efficacité significatives au sein des groupes (nécessité de réaliser des répliquats techniques pour effectuer ce test qui permet de détecter la présence éventuelle d'échantillons aberrants au sein des groupes) et entre les groupes expérimentaux. En l'absence d'effets intra- et inter-groupes significatifs, les auteurs suggèrent d'utiliser pour tous les échantillons, une valeur d'efficacité unique égale à la moyenne des efficacités de PCR estimées sur chaque courbe. D'après notre expérience, néanmoins limitée à quelques mois, de cette méthode, les résultats apparaissent parfaitement cohérents avec une méthode par courbes standard et avec les données publiées sur plusieurs régulations géniques provoquées par le fénofibrate. L'avantage évident de cette méthode est d'éviter la réalisation de courbes standard par le biais d'une utilisation plus complète des données de fluorescence enregistrées. Un autre avantage est l'applicabilité quasi-universelle de cette méthode dans la mesure où l'efficacité de PCR n'est pas considérée *a priori* comme égale dans tous les tubes.

D'autres méthodes, non évoquées ici, ont également été proposées dans la littérature pour définir les points appartenant à la phase exponentielle et/ou pour déterminer les valeurs individuelles des efficacités de PCR. A ma connaissance, aucune de ces méthodes ne s'affranchit néanmoins de l'inspection individuelle des courbes de PCR et aucune n'a encore fait l'objet du développement d'un package pour un logiciel de statistique comme R ([www.r-project.org](http://www.r-project.org)) par exemple, permettant éventuellement de motiver de nouveaux développements autour de l'analyse des données de PCR en temps réel par la communauté des statisticiens. Nous construisons actuellement un package R implémentant les méthodes présentées dans ce mémoire afin d'engager une réflexion sur l'analyse de ce type de données dans le cadre d'un groupe de travail toulousain sur l'analyse statistique des données de transcriptomiques (plus d'informations sur [www.lsp.ups-tlse.fr/Programmes/Biopuces/](http://www.lsp.ups-tlse.fr/Programmes/Biopuces/) et une série de diapositives sur

le choix des amorces de PCR en temps réel et l'analyse des données disponible sur [www.lsp.ups-tlse.fr/Programmes/Biopuces/Publications/PDF/pmartin\\_26\\_03\\_07.pdf](http://www.lsp.ups-tlse.fr/Programmes/Biopuces/Publications/PDF/pmartin_26_03_07.pdf)).

Comme nous venons de le présenter, plusieurs options s'offrent au biologiste pour analyser ses données de PCR en temps réel. Elles sont généralement basées sur des hypothèses dont il convient d'évaluer la vraisemblance et toutes supposent en particulier 1) que la fluorescence enregistrée est bien proportionnelle à la quantité d'amplicons produits et 2) que le seuil de fluorescence a été correctement fixé de manière à couper toutes les courbes de PCR au niveau de leur phase exponentielle. Alors que la seconde hypothèse relève d'une simple manipulation, généralement triviale, effectuée par l'utilisateur sur son logiciel d'analyse, la première hypothèse relève essentiellement d'un choix judicieux des amorces de PCR (en SYBR Green) ou de la sonde TaqMan. Il convient donc de ne pas sous-estimer l'importance du choix et de l'optimisation de la concentration des amorces lors du développement d'un dosage par PCR en temps réel. Finalement, si les calculs présentés dans ce paragraphe apparaîtront certainement au lecteur mathématicien d'un niveau élémentaire, l'expérience (la mienne du moins) montre que la plupart de ces concepts sont loin d'être maîtrisés par la plupart des utilisateurs de la PCR en temps réel et que cette technique se diffuse bien plus vite que la compréhension des bases théoriques de l'analyse des données qui en sont issues, posant ainsi naturellement des questions sur la validité d'une part non négligeable des observations rapportées dans la littérature.

#### AVANTAGES :

Les principaux avantages de la PCR en temps réel sont sa sensibilité inégalée, liée essentiellement à la sensibilité de la PCR elle-même, et sa simplicité et rapidité de mise en œuvre. En termes concrets, en partant d'un couple d'amorces, même non validé, et d'échantillons d'ARN en nombre raisonnable ( $\leq 96$  si on travaille en plaque 96 puits), il est parfaitement possible en une seule journée de parvenir à un résultat sur la mesure de l'expression d'un gène d'intérêt (à condition que la validation du couple d'amorces se passe sans difficulté majeure, ce qui est le cas dans la majorité des situations lorsque l'on travaille avec un organisme séquencé et que l'on porte une attention particulière au choix des amorces). La PCR en temps réel est par ailleurs relativement peu sensible à une dégradation modérée des échantillons d'ARN dans la mesure où les fragments amplifiés sont de courte taille (généralement inférieurs à 150 pb). Elle est cependant plus sensible à la présence d'inhibiteurs de PCR (impuretés ou traces de solvants dans les ARNs extraits). La PCR en temps réel est la technique de choix pour l'analyse de l'expression de transcrits alternatifs ou

de gènes appartenant à des familles multigéniques puisque une différence d'une seule base localisée à l'extrémité 3' d'une des amores peut être suffisante pour discriminer par PCR deux séquences nucléotidiques très proches. Nous ne l'avons pas évoqué ici, mais la PCR en temps réel est également utile pour d'autres applications comme des études de génotypage (détection de polymorphisme sur un seul nucléotide ou SNPs), ou la mesure du nombre de copie d'un gène dans un génome. Enfin, cette technique a fait l'objet de miniaturisations qui permettent de réaliser des PCR dans un volume de 2 µL, autorisant ainsi son application pour des études à moyen- voire à haut-débit et sur des échantillons rares et disponibles en quantité très limitée.

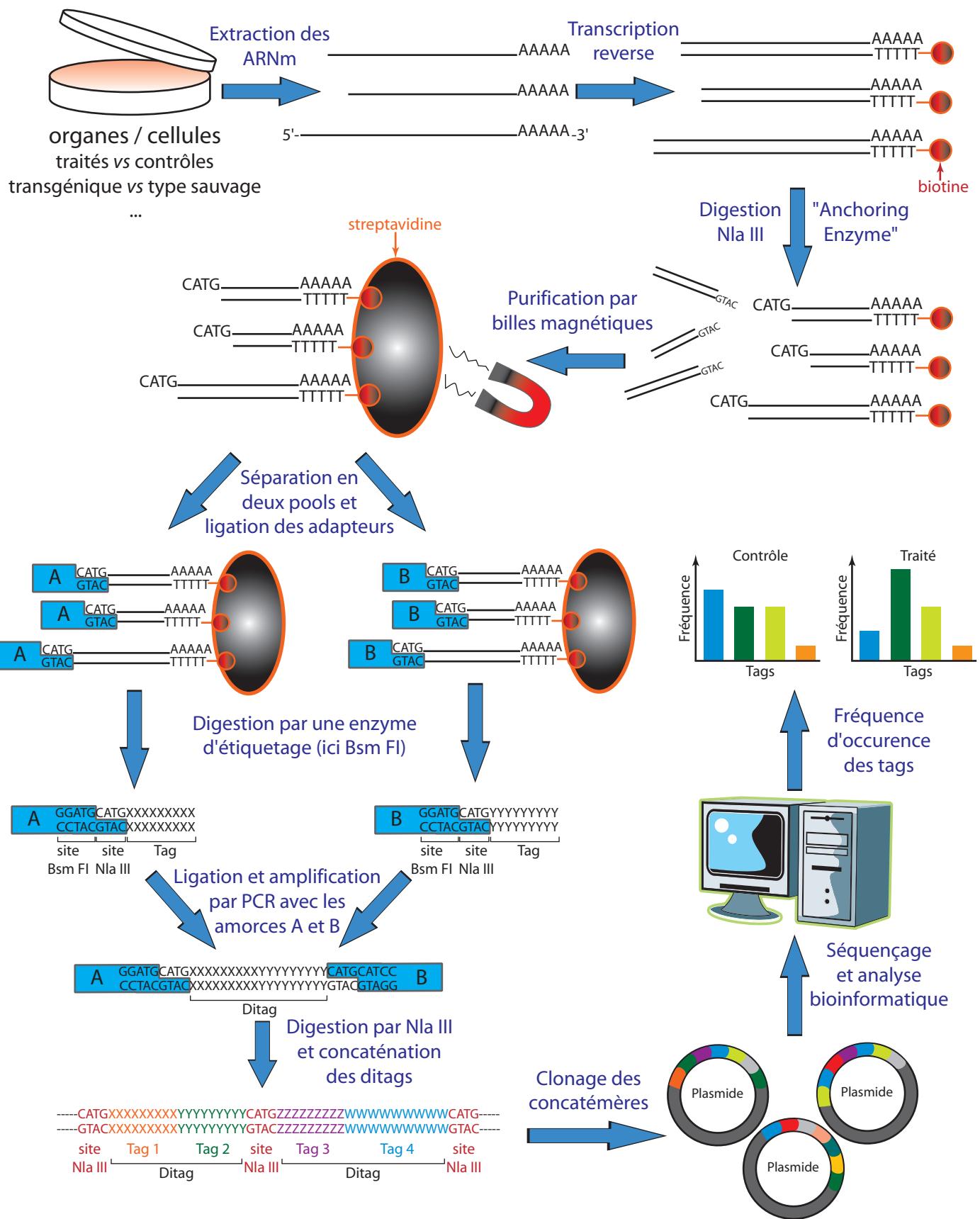
#### *INCONVENIENTS / LIMITES :*

La PCR en temps réel requiert l'achat d'un appareil dédié à cette technique dont le coût n'est pas négligeable. De plus, nous avons souligné que l'analyse des données qui en sont issues n'est pas des plus évidentes et peut être source de confusion. Le choix des amores étant critique en PCR en temps réel, la disponibilité d'une séquence du gène d'intérêt, voire du génome complet de l'organisme étudié constituent un net avantage mais il est toujours possible de pallier à l'absence de ces informations par des approches expérimentales. Enfin, contrairement au NB, la PCR en temps réel ne fournit pas d'indication sur la taille du transcript analysé ou la présence insoupçonnée de transcrits alternatifs.

En conclusion, il existe un assez large panel de méthodes de mesure de l'expression des gènes pour les études à « bas-débit », présentant toutes des avantages et des limites que j'ai tenté de mettre en évidence dans ce paragraphe, mais ces méthodes n'autorisent pas une exploration large, voire quasi-complète, du transcriptome (le premier niveau résultant de l'expression du génome), contrairement aux méthodes que nous allons présenter dans le paragraphe suivant.

### **III.2. TECHNIQUES DE MESURE DE L'EXPRESSION GENIQUE A « HAUT DEBIT »**

Les techniques présentées dans ce paragraphe visent à mesurer de manière simultanée l'abondance d'une large part, voire de la totalité des ARN messagers présents dans un échantillon biologique à un instant donné, c'est à dire à évaluer l'expression des gènes à l'échelle du transcriptome. On peut globalement classer ces techniques en trois familles



**Figure 46 : Principe général de la méthode SAGE**

La méthode SAGE (Serial Analysis of Gene Expression) est basée sur le principe qu'une courte étiquette (ou tag) de 9 à 14 bases de longueur est suffisante pour identifier de manière quasi-certaine la nature du transcript dont cette étiquette provient. La méthode SAGE utilise la concaténation de tags pour permettre une analyse à haut-débit de l'occurrence des tags issus des différents transcrits du génome. La description des différentes étapes est détaillée dans le corps du texte.

(Breyne and Zabeau, 2001) : 1) les méthodes basées sur l'isolation et le comptage de courtes séquences d'ADNc issues des transcrits (les « tags ») comme la méthode SAGE (Serial Analysis of Gene Expression), 2) les méthodes basées sur une amplification par PCR de sous-populations d'ARN messagers comme le differential display et 3) les méthodes basées sur l'hybridation d'un mélange complexe d'ARNs marqués sur un microréseau (microarray) constitué d'un support solide sur lequel des fragments spécifiques de chaque transcrit ont été déposés ou synthétisés *in situ* en des positions définies (les « spots »). Dans ce chapitre, nous présenterons brièvement les deux premières familles de techniques, avec leurs avantages et leurs limites, et nous attarderons sur la technologie des microarrays qui a été utilisée dans le cadre de ce doctorat.

### **III.2.1. Les méthodes basées sur le séquençage d'étiquettes (ou tags)**

Les techniques appartenant à cette famille sont basées sur le principe qu'une courte séquence nucléotidique (une « étiquette » ou un « tag », long d'environ 10-14 bases) contient suffisamment d'information pour identifier de manière quasi-certaine le gène dont cette séquence provient, à condition que le tag provienne toujours de la même position au sein de chaque transcrit. La méthode SAGE (pour Serial Analysis of Gene Expression) est sans doute la plus connue et la plus utilisée des méthodes utilisant ce principe (Velculescu et al., 1995) et sera donc présentée ci-dessous.

#### *PRINCIPE GENERAL DE LA METHODE SAGE :*

La méthode SAGE est illustrée en [Figure 46](#) et se déroule selon les étapes suivantes :

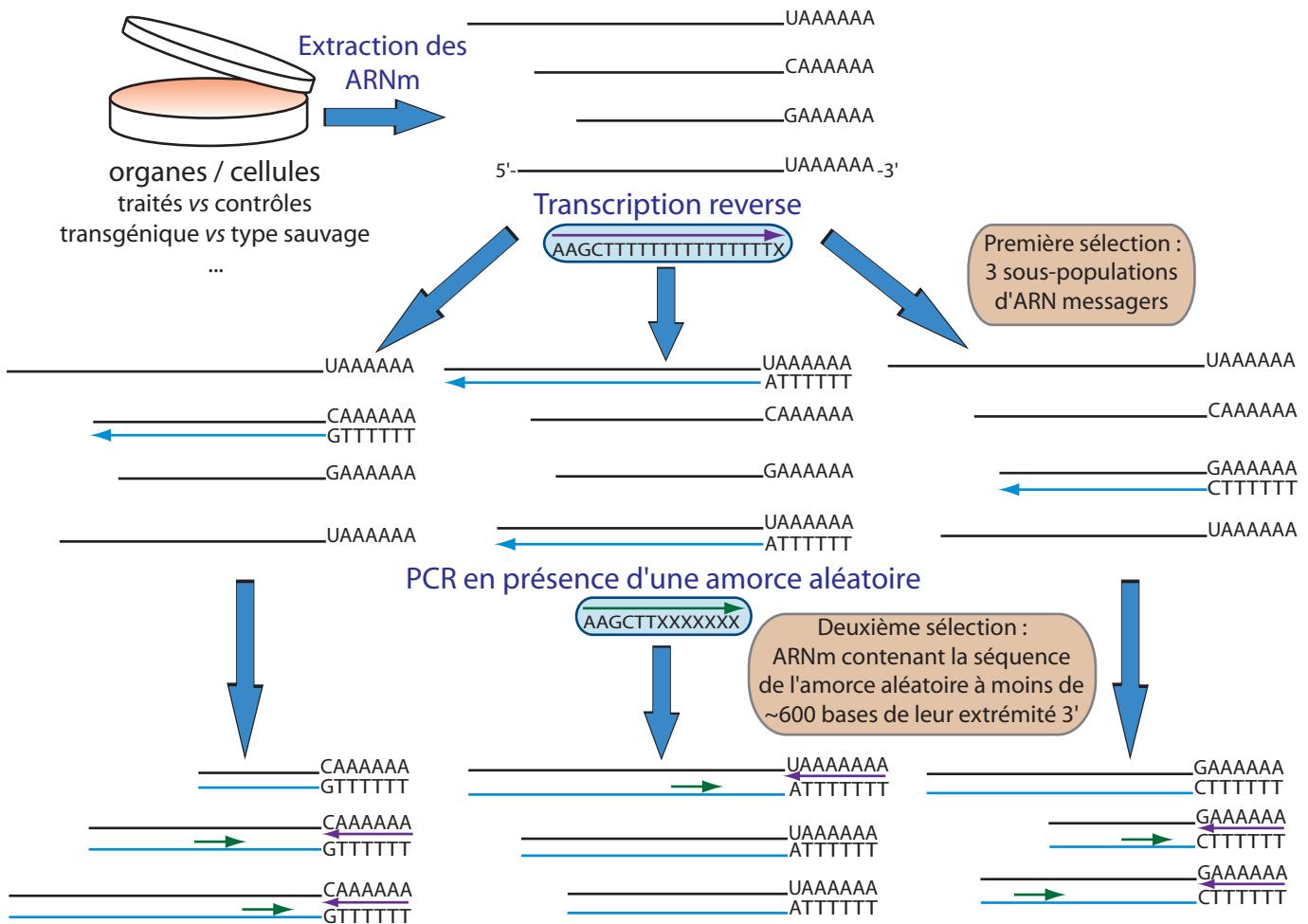
- Les ARNs messagers sont extraits de l'échantillon biologique étudié
- Ils subissent ensuite une transcription reverse en présence d'oligodT biotinylés
- Les ADNc obtenus sont ensuite soumis à une digestion par une endonucléase de restriction particulière appelée « enzyme d'ancre » ou anchoring enzyme (exemples : Nla III, Sau 3A1, Rsa I). Ces enzymes reconnaissent un site de restriction de 4 pb (5'-CATG pour Nla III par exemple) qui est donc fréquent dans les transcrits (en moyenne, un tel site est présent tous les  $4^4 = 256$  pb dans une séquence aléatoire). Il résulte de cette digestion que près de 80% de l'ADNc va être coupé et séparé de la queue polyA/oligodT biotinylée située en 3' et que presque tous les transcrits contenant cette queue vont posséder la séquence CATG à leur extrémité 5' sous la forme d'une extrémité cohésive et toujours localisée au plus près possible de la queue polyA/oligodT.

- Les fragments d'ADNc contenant la queue polyA/oligodT en 3' sont ensuite purifiés à l'aide de billes magnétiques recouvertes de streptavidine et séparés en deux pools.
- Chaque pool subit une ligation à son extrémité cohésive avec un adaptateur (linker en anglais). Le premier pool subit une ligation à l'adaptateur A et le second à l'adaptateur B. Ces adaptateurs possèdent 1) une extrémité cohésive complémentaire de la séquence CATG, reconstituant ainsi le site de restriction de l'enzyme Nla III (ou autre séquence si une autre enzyme est utilisée), 2) une séquence spécifique unique correspondant à une amorce de PCR et 3) la séquence de reconnaissance d'une enzyme d'étiquetage (tagging enzyme en anglais comme par exemple Bsm FI qui reconnaît la séquence 5'-GGGAC).
- Chaque pool subit alors une digestion par l'enzyme d'étiquetage choisie. Ces enzymes (également appelées endonucléases de restriction de type IIS), coupent à une distance définie (jusqu'à 20bp) de leur site de reconnaissance asymétrique en formant des extrémités franches. Cette digestion conduit ainsi à la libération des billes magnétiques de fragments contenant l'adaptateur (A ou B selon le pool) suivi d'un court fragment d'ADNc appelé étiquette ou tag et dont la taille fixe dépend de l'enzyme d'étiquetage utilisée.
- Les deux pools A et B subissent ensuite une ligation de manière à former des « ditags » (ou étiquettes doubles) encadrés à leurs deux extrémités par les adaptateurs A et B respectivement.
- Ces ditags sont ensuite amplifiés par PCR avec des amorces spécifiques des adaptateurs A et B.
- Les ditags amplifiés subissent ensuite une digestion par l'enzyme d'ancre (anchoring enzyme) afin de les libérer de leurs adaptateurs localisés à leurs extrémités.
- Les ditags sont ensuite isolés et concaténés par ligation (jusqu'à une cinquantaine de ditags peuvent être ainsi concaténés) puis clonés.
- Les concatémères sont enfin séquencés et une analyse bioinformatique permet de détecter les séquences de reconnaissance de l'enzyme d'ancre (CATG si Nla III a été utilisée) qui encadre chaque étiquette double, d'identifier les étiquettes et de dénombrer leurs occurrences. La fréquence d'occurrence de chaque tag fournit ainsi un profil d'expression des transcrits qui peut être comparé entre différentes situations (traités vs non traités par exemple).

La méthode longSAGE est une adaptation de la méthode SAGE (Saha et al., 2002) qui génère des tags plus longs (17 pb hors séquence de 4 pb du site de restriction de Nla III) et permet ainsi d'améliorer l'identification des transcrits à partir des bases de données. Par ailleurs, une autre méthode appelée MPSS pour Massively Parallel Signature Sequencing (Brenner et al., 2000a; Brenner et al., 2000b) utilise un principe similaire à la méthode SAGE, basé sur le séquençage de tags localisé dans la partie 3' des transcrits (pour une description animée de cette méthode, voir le site [www.lynxgen.com](http://www.lynxgen.com)). Les principales améliorations apportées par la méthode MPSS sont une augmentation du débit de séquençage lié au développement d'une méthode propriétaire (Brenner et al., 2000a) conduisant à une meilleure détection des gènes dont le niveau d'expression est faible ainsi qu'une augmentation de la longueur des tags, comme en LongSAGE, facilitant ainsi l'identification des transcrits dont les tags sont issus à partir des bases de données (pour une comparaison aux autres méthodes d'étude du transcriptome, voir Reinartz et al., 2002). Une expérience de SAGE génère typiquement plus de 50000 tags alors qu'une expérience de MPSS peut générer plus de 10 millions de tags, fournissant ainsi une couverture bien plus large du transcriptome (Brenner et al., 2000a).

#### AVANTAGES :

Le principal avantage des méthodes basées sur le séquençage de tags est leur capacité à identifier de nouveaux gènes car ces méthodes ne reposent pas sur la connaissance a priori de la séquence complète du génome étudié. C'est particulièrement vrai pour les méthodes LongSAGE et MPSS qui génèrent des tags plus longs. Un tag dont la séquence ne correspond à aucune autre séquence dans les bases de données est vraisemblablement issu d'un transcrit ou d'un exon alternatif qui n'a pas encore été décrit. La séquence du tag peut ainsi être utilisée pour rechercher la séquence complète du transcrit en l'utilisant comme sonde pour cribler une banque d'ADNc ou encore comme amorce dans une expérience de RACE-PCR (Rapid amplification of cDNA ends) ou de GLGI (Generation of longer 3' cDNA fragments from SAGE tags for gene identification, Chen et al., 2000; Silva et al., 2004a). En l'absence d'hybridation dans leurs protocoles respectifs, ces méthodes s'affranchissent également des problèmes d'hybridation croisée, rencontrés fréquemment pour les familles multigéniques ou les transcrits alternatifs. Alors que la méthode SAGE ne peut se targuer d'une sensibilité suffisante pour l'analyse des transcrits faiblement exprimés en raison du séquençage d'un nombre limité de tags, la méthode MPSS assure, elle, une couverture suffisamment large du transcriptome permettant en théorie d'étudier des transcrits présents à seulement 1 à 3 copies



#### Migration par électrophorèse sur gel de polyacrylamide (gel de séquençage)

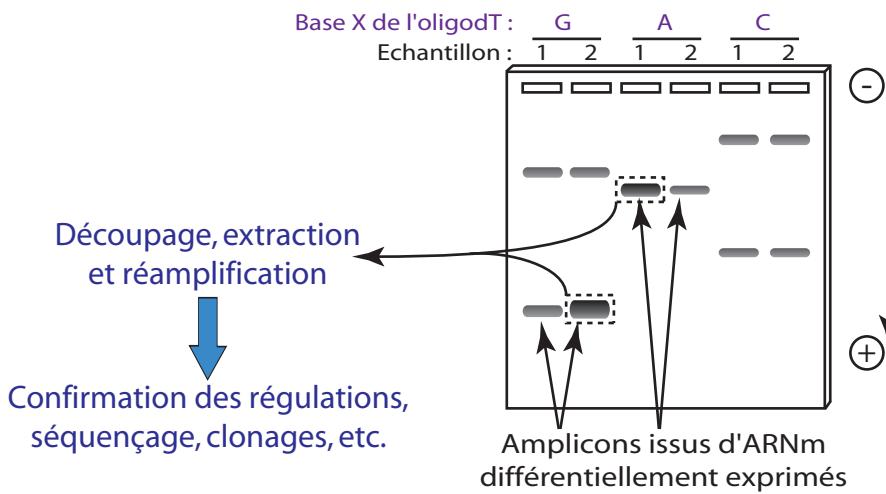


Figure 47 : Principe général du Differential Display

Le differential display est basé sur l'amplification sélective de sous-populations d'ARNmessagers par transcription reverse (RT) puis PCR. Les fragments issus de ces amplifications sont séparés sur gel de séquençage et ceux présentant une abondance différente entre les échantillons sont ensuite analysés individuellement pour établir leur identité et confirmer les régulations observées. Les séquences des amorces classiquement utilisées pour la RT et la PCR sont indiquées dans les cadres bleus. Les deux étapes de sélection de sous-population sont également mises en évidence. La description des différentes étapes est détaillée dans le texte. Par soucis de simplification, un nombre réduit de bandes et d'ARNm ont été représentés sur cette figure.

par cellule. Enfin, un autre avantage majeur de ces méthodes basées sur le séquençage de tags est le format digital dans lequel les données sont générées. Dans la mesure où les données sont exprimées en fréquence d'occurrence des transcrits, elles peuvent être facilement intégrées à des bases de données et comparées directement avec d'autres expériences produites par la communauté des utilisateurs de ces méthodes.

#### *INCONVENIENTS / LIMITES :*

S'il est certain que ces méthodes peuvent s'appliquer à des organismes non séquencés, l'exploitation des données est malgré tout nettement améliorée par la qualité des bases de données génomiques disponibles. Suite à une analyse de type SAGE ou MPSS sur un organisme dont le génome n'est pas séquencé, l'identification des transcrits, la confirmation de leur expression différentielle, et l'étude de l'impact biologique de cette régulation restent généralement fondés sur une approche gène-par-gène. L'identification des transcrits a partir des tags générés par ces méthodes peut également se heurter à de nombreuses barrières comme un tag correspondant à une séquence hautement répétitive dans le génome, la présence de sites alternatifs de polyadénylation des ARNm ou d'épissages alternatifs localisés dans la partie 3' des transcrits ou encore la présence de SNP affectant soit le site de coupure de l'enzyme « d'ancre » soit le tag lui-même. Heureusement, des bases de données facilitant l'identification des transcrits à partir des tags et prenant en compte ces aspects ont été créées (Boon et al., 2002; Silva et al., 2004b), mais elles sont actuellement limitées aux modèles d'étude eucaryotes les plus répandus : l'homme et la souris. Par ailleurs, une proportion, néanmoins faible (de l'ordre de 1%, Boon et al., 2002), de transcrits ne présentent pas de site de coupure pour l'enzyme d'ancre choisie et ne peuvent donc être analysés par ces méthodes. L'utilisation de plusieurs enzymes d'ancre différentes permet, en théorie, de pallier largement à ce problème, mais complique tellement la mise en œuvre des méthodes qu'elle n'est que rarement appliquée. D'autres problèmes techniques peuvent également être rencontrés comme par exemple le fait que la température peut influencer la longueur des fragments générés par les enzymes d'étiquetage. Enfin, les inconvénients majeurs de ces méthodes sont leur coût et le temps nécessaire à leur mise en œuvre. La réalisation de centaines voire de milliers de PCR et de séquençages pour chaque échantillon biologique limite naturellement la diffusion de ces techniques à des laboratoires ou plateformes qui se dédient généralement à ces applications. Le monopole détenu par la société Lynx Therapeutics sur la réalisation de la technique MPSS est un bon exemple de la diffusion très limitée de ces technologies.

### **III.2.2. Les méthodes basées sur l'amplification de sous-populations d'ARNm**

Une autre famille de méthode est basée sur l'amplification par transcription reverse puis PCR de sous-populations d'ARN messagers issus d'un mélange complexe d'ARNm extrait d'un échantillon biologique. Les sous-populations sont ensuite étudiées par une migration sur un gel de séquençage (gel dénaturant de polyacrylamide), offrant une résolution suffisante pour l'étude de larges populations d'ADN (résolution compatible avec l'étude d'environ une centaine de fragment par piste dont la taille ne doit pas excéder 500-600 pb). Les bandes apparaissant à des intensités différentes entre les groupes d'échantillons analysés sont ensuite extraites du gel, clonées et séquencées afin d'identifier les transcrits dont elles proviennent. Les premières méthodes utilisant ce principe ont été introduites en 1992 sous les noms de differential display (ou tri d'ARNm en français, Liang and Pardee, 1992) et de arbitrarily-primed PCR (AP-PCR, Welsh et al., 1992). Plusieurs développements de méthodes basées sur le même principe ont ensuite été entrepris conduisant à l'introduction du cDNA-AFLP (Amplified restriction fragment length polymorphism, Bachem et al., 1996) ou encore de la méthode du GeneCalling par la société CuraGen (Shimkets et al., 1999). Cette dernière technique a notamment été utilisée pour étudier l'effet d'agonistes de PPAR $\alpha$  (Gould Rothberg et al., 2001) et de PPAR $\gamma$  (Way et al., 2001). Nous présenterons ci-dessous le principe général de la technique du differential display et les adaptations apportées par certaines méthodes alternatives puis discuterons les avantages et limites de cette famille de techniques.

#### *PRINCIPE GENERAL DE CES TECHNIQUES :*

Le principe général du differential display est illustré en [Figure 47](#) et comporte les étapes suivantes :

- Les ARNs messagers sont extraits de l'échantillon biologique étudié
- Une transcription reverse est effectuée sur ces ARNm, en présence d'un oligodT comportant deux bases supplémentaires en 3' (amorce de type 5'-T<sub>12</sub>XY-3' où X est l'une des 3 bases A, C ou G et où Y est l'une des quatre bases A, T, G ou C) qui déterminent ainsi la sous-population des ARNm qui vont être reverse transcrits (ceux qui portent les deux bases complémentaires de l'oligodT en position 5' de leur queue polyA). Cette étape permet de sélectionner une sous-population correspondant en moyenne au 1/12ème de la population d'ARNm présente dans l'échantillon biologique

étudié (car il y a  $3 \times 4 = 12$  possibilités pour la combinaison des bases XY). L'utilisation d'oligodT plus longs mais ne contenant qu'une seule base additionnelle en 3', proposée ultérieurement (Liang et al., 1994), semble améliorer les résultats du differential display, notamment en diminuant l'amorçage en plusieurs points de transcrits contenant des séries de A dans leurs séquences et en atténuant le « smear » observé sur les gels de séquençage qui est lié à une contamination par des oligodT des amorces  $T_{12}XY$  où Y est la base T au cours de la synthèse de tels oligonucléotides. De plus, l'utilisation de telles amorces simplifie nettement la procédure de differential display en séparant la population des ARNm en trois sous-populations ([Figure 47](#)) au lieu de douze.

- Les sous-populations d'ADNc ainsi obtenues sont ensuite amplifiées par PCR en présence du même oligodT utilisé à l'étape précédente et d'une amorce aléatoire longue d'une dizaine de bases. L'incorporation d'un nucléotide marqué au cours de cette étape permettra la visualisation des bandes sur le gel de séquençage. La fixation de l'amorce aléatoire doit se faire à une distance inférieure à 500-600 bases de la queue oligodT et cette distance doit être distribuée de manière aléatoire entre les différents ADNc afin que les différents amplicons soient de tailles différentes et puissent être séparés et visualisés sur un gel de séquençage. Pour cela, il faudrait en théorie une amorce longue de 6-7 bases pour s'assurer qu'elle ait une forte probabilité de se fixer à une séquence complémentaire localisée dans les 500 bases de l'extrémité 5' de l'ADNc et qu'elle amplifie des sous-populations de l'ordre de 50 à 100 ADNc (Liang and Pardee, 1992) mais des amorces de taille inférieure à 10 bases fonctionnent mal en PCR. On utilise donc des amorces de 13 bases de long dont seules les 7 bases à l'extrémité 3' sont aléatoires et dont l'extrémité 5' contient une séquence fixe AAGCTT correspondant au site de restriction de HindIII qui peut également être ajouté en 3' de la queue oligodT pour faciliter le clonage ultérieur des fragments identifiés comme différenciellement exprimés ([Figure 47](#), Liang et al., 1994). Cette séquence constante en 5' fournit en moyenne 1,5 bases complémentaires à la séquence à laquelle l'amorce aléatoire va se fixer, assurant ainsi une fixation suffisante (entre 8 et 9 bases) de l'amorce au cours des premiers cycles de la PCR. Il existe environ 16000 séquences différentes de 7 pb ( $4^7 = 16384$  exactement) et l'extrémité 3' d'un ADNc longue de 600 bases contient environ 600 sites de reconnaissance des ~16000 amorces possibles. Chaque amorce a donc une probabilité d'environ 4% ( $600/16000 \sim 0,04$ ) de reconnaître un ARNm donné. Deux amorces aléatoires distinctes

auront une probabilité de reconnaître un ARNm donné de  $1-(0,96)^2=7,8\%$ . L'utilisation de 80 amores distinctes de séquence aléatoire permettent ainsi de détecter un ARNm donné avec une probabilité de  $1-(0,96)^{80}=96\%$ . Pour les trois sous-populations d'ADNc produites au cours de l'étape précédente, cela conduit donc à réaliser un total de  $80 \times 3 = 240$  réactions de PCR par échantillon afin de couvrir la quasi totalité des ARNm du transcriptome.

- Les sous-populations d'amplicons ainsi obtenues subissent ensuite une électrophorèse sur gel de séquençage afin de séparer les amplicons et d'identifier ceux qui présentent une abondance différente entre les échantillons étudiés.
- Les bandes d'intérêt sont ensuite découpées du gel, extraites et réamplifiées avec la même combinaison d'amores que celle utilisée dans les étapes précédentes. Les amplicons obtenus peuvent alors être clonés, séquencés ou utilisés pour cibler une banque d'ADNc ou réaliser un northern blot dans le but de confirmer leur expression différentielle et de déterminer de quels transcrits ils proviennent.

La principale modification introduite à cette méthode par l'AP-PCR (Welsh et al., 1992) est l'utilisation d'amores aléatoires lors de la synthèse de l'ADNc par transcription reverse double brin à faible stringence fournissant ainsi une séquence localisée à l'intérieur de l'ARNm au lieu d'une séquence localisée en partie 3' des transcrits.

La méthode du cDNA-AFLP (pour des protocoles détaillés, voir Bachem et al., 1998; Breyne et al., 2003), également basée sur le même principe général, est réalisée via les grandes étapes suivantes :

- Les ARNm sont extraits des échantillons puis reverse transcrit en présence d'oligodT biotinylés
- Les ADNc obtenus subissent une double digestion par des endonucléases de restriction, libérant ainsi des fragments internes aux transcrits qui subissent ensuite une ligation avec deux adaptateurs permettant une préamplification par PCR de la population de fragments à l'aide d'amores spécifiques de ces adaptateurs.
- Une PCR utilisant des amores homologues aux adaptateurs et contenant une ou deux bases aléatoires à leur extrémité 3' ( $16 \times 16 = 256$  combinaisons possibles d'amores contenant 2 nucléotides aléatoires en 3') permet d'amplifier sélectivement des sous-populations de fragments issus des transcrits qui sont ensuite séparées sur gel de séquençage afin de sélectionner les clones différemment représentés entre les

échantillons pour leur étude ultérieure. L'utilisation d'amorces marquées (radiomarquage) permet la détection des fragments sur le gel.

L'utilisation éventuelle de plusieurs couples différents d'endonucléases de restriction, permet, en théorie, de bien couvrir la quasi-totalité du transcriptome. C'est d'ailleurs la stratégie utilisée par la méthode du GeneCalling (Shimkets et al., 1999) qui utilise également des amorces modifiées (l'une couplée à un fluorophore pour la détection et l'autre à de la biotine pour la purification), homologues aux adaptateurs, pour assurer une détection automatisée des fragments issus des transcrits suite à leur séparation sur gel de polyacrylamide. Le couplage de cette méthode à une base de données des longueurs des fragments attendus par digestion avec les enzymes de restriction utilisées, sur la base des séquences disponibles, permet également une identification directe des transcrits connus, évitant ainsi une part substantielle des étapes d'identification des transcrits différentiellement exprimés et permettant à l'utilisateur de se concentrer sur les nouveaux transcrits.

#### *AVANTAGES :*

Ces différentes méthodes dispose d'une bonne sensibilité (améliorée par la préamplification dans les méthodes cDNA-AFLP et GeneCalling) et d'une bonne spécificité via l'utilisation de la PCR. L'identification des transcrits à partir des bases de données de séquence disponibles est plus facile qu'avec les méthodes basées sur les tags en raison de l'obtention d'une séquence plus longue pour chaque transcrit (de l'ordre de 100 à 500 pb le plus souvent). Ces méthodes ont, de plus, un coût relativement faible, notamment car elles ne nécessitent pas l'achat d'appareillages coûteux mais utilisent essentiellement des techniques couramment pratiquées dans les laboratoires de biologie moléculaire (RT-PCR, digestion enzymatique, ligation, clonage, électrophorèse, séquençage). De manière similaire aux méthodes basées sur les tags, la disponibilité d'informations complètes de séquençage du génome de l'organisme étudié n'est pas un pré-requis pour ces techniques qui permettent théoriquement l'identification de nouveaux transcrits, mais elle améliore néanmoins nettement l'exploitation des résultats.

#### *INCONVENIENTS / LIMITES :*

Bien qu'assez simples à mettre en œuvre en raison de la maîtrise courante des protocoles utilisés par ces technologies, ces méthodes ne permettent d'obtenir un profil du transcriptome de haute résolution qu'au prix d'une multiplication des combinaisons d'amorces et/ou d'enzymes de restriction utilisées. Par exemple, de l'ordre de 150 à 300 PCR

par échantillons sont ainsi nécessaires à l'obtention d'un profil quasi-complet des transcrits par differential display ou cDNA-AFLP. Par ailleurs, bien que des précautions soient normalement prises (nombre limité de cycles de PCR notamment), l'introduction d'étapes de PCR dans ces méthodes risque toujours d'altérer le caractère quantitatif de ces techniques (voir paragraphe III.1.3 sur la PCR en temps réel) et éventuellement de générer plus de faux positifs (présence de SNPs conduisant à l'absence d'amplification d'un fragment dans certains échantillons par exemple). A part peut-être pour la méthode du GeneCalling, la lourdeur des étapes de confirmation des régulations et de l'identification des transcrits conduit généralement à la génération d'un nombre très limité de données, altérant ainsi considérablement le débit et le caractère « pangénomique » de ces méthodes. Lorsque l'on s'adresse, de plus, à un système déjà bien étudié, comme par exemple les régulations géniques opérées par le récepteur PPAR $\alpha$ , une large part du temps de l'expérimentateur sera naturellement consacré à la redécouverte de régulations déjà décrites dans la littérature et la part d'informations nouvelles risque d'être très réduite. Enfin, contrairement aux méthodes présentées dans le paragraphe précédent (III.2.1.), ces techniques ne permettent pas une intégration directe des données générées dans des bases de données et d'exploiter ainsi aisément les informations obtenues par une communauté d'utilisateurs.

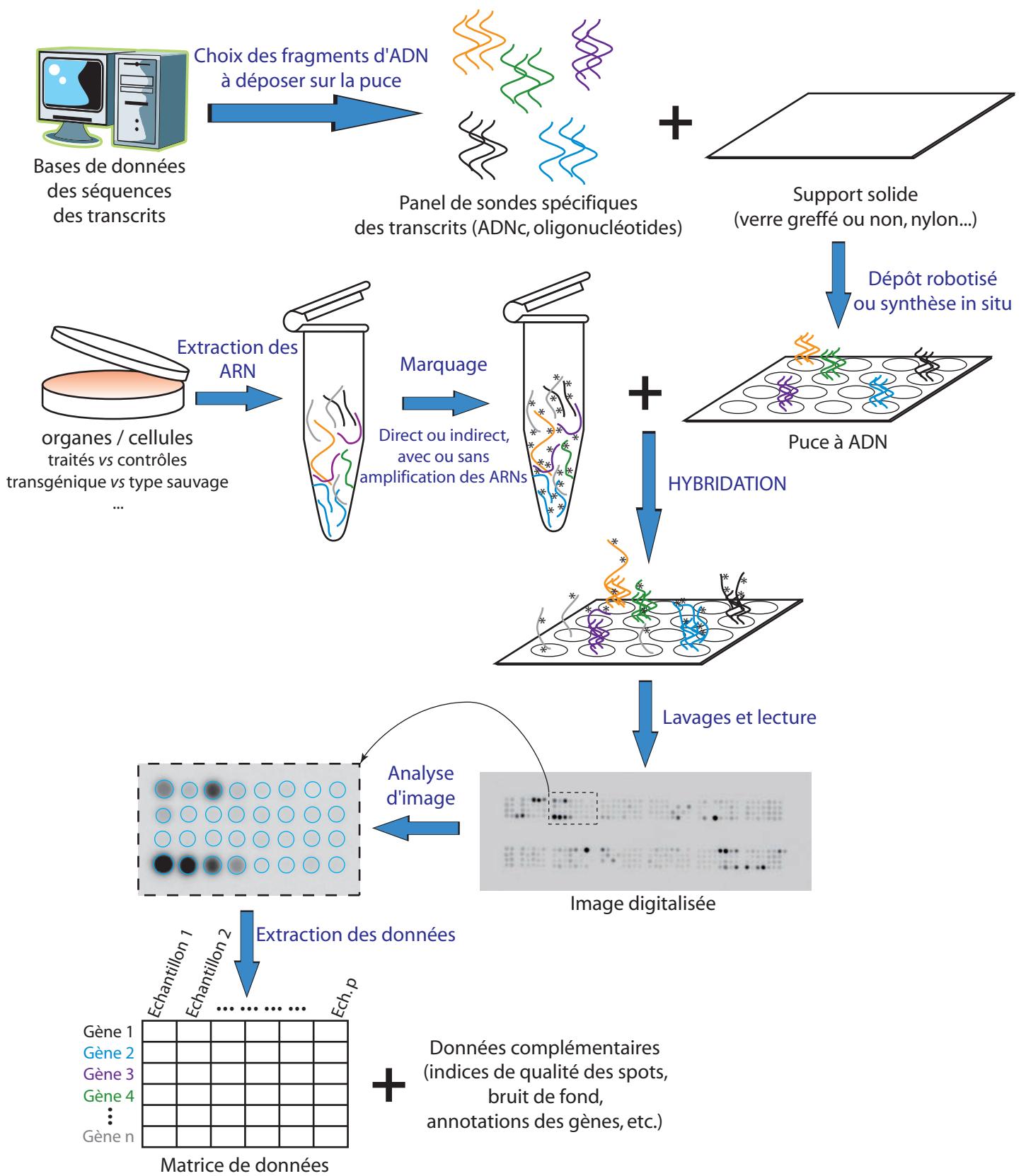
En conclusion, il me semble que l'on peut qualifier les deux familles de méthodes présentées dans les paragraphes III.2.1. et III.2.2. de techniques « exploratoires » du transcriptome dans la mesure où elles permettent à l'utilisateur de s'affranchir d'une connaissance exhaustive du génome de l'organisme étudié et lui permettent de découvrir l'existence et la régulation de gènes totalement nouveaux. D'autres méthodes d'exploration du transcriptome n'ont pas été présentées ici comme l'hybridation soustractive ou encore le séquençage systématique d'ESTs (Expressed Sequence Tags) car ces méthodes me semblent plus limitées dans leur diffusion en raison de la lourdeur de leur mise en œuvre. J'invite donc le lecteur intéressé à consulter des synthèses comparant ces méthodes à celles présentées ici (voir par exemple Ruan et al., 2004; Soares, 1997). Nous allons voir que la troisième famille de méthodes d'étude du transcriptome présentée dans le paragraphe suivant : les puces à ADN, se distingue de ces techniques principalement par la nécessité d'une connaissance a priori des informations de séquence des transcrits que l'on souhaite étudier ainsi que par leur facilité et rapidité d'utilisation conduisant à la génération d'un volume considérable de données obtenues sur des nombres plus conséquents d'échantillons.

### **III.2.3. Les puces à ADN**

Les puces à ADN, apparues aux alentours de 1995 (pour quelques exemples précoce de mise en œuvre, voir notamment Bernard et al., 1996; DeRisi et al., 1997; Gress et al., 1992; Lockhart et al., 1996; Nguyen et al., 1995; Schena et al., 1996; Zhao et al., 1995), ont considérablement modifié la façon dont les chercheurs ont abordé l'étude du transcriptome (pour des synthèses précoce sur le sujet, voir Jordan, 1998; Lipshutz et al., 1999; Schena, 1996). Le développement de ces nouveaux outils n'a été rendu possible que grâce aux vastes projets de séquençage complet des génomes car le principe même de cette technique utilise l'information disponible sur les séquences des transcrits de l'organisme étudié (voir ci-dessous). Ainsi, l'apparition de ces techniques a été considérée comme une avancée majeure dans le domaine que l'on qualifie souvent de post-génomique ou de génomique fonctionnelle. D'une analyse exploratoire du transcriptome, permettant essentiellement à des laboratoires non spécialisés dans des méthodes comme la SAGE, de découvrir quelques régulations géniques, les puces à ADN ont permis à de très nombreuses équipes (souvent par le biais de plateformes dédiées à ces techniques mises en place aux niveaux nationaux ou locaux, en France principalement par le biais des génopoles) de générer des jeux de données couvrant la quasi-totalité du transcriptome et de mesurer ainsi véritablement à haut-débit le niveau d'expression de la majeure partie des gènes du génome. On peut donc considérer qu'il s'agit d'une méthode « d'analyse ou de mesure du transcriptome » par opposition aux méthodes que nous avons précédemment qualifiées « d'exploratoires ».

#### *PRINCIPE GENERAL DES PUCES A ADN :*

La technique des puces à ADN ou microarrays ([Figures 48 et 49](#)) est globalement basée sur un principe opposé à celui du dot blot présenté brièvement au paragraphe III.1.1. Pour chaque gène dont on connaît la séquence, un fragment du transcrit correspondant est sélectionné à partir des bases de données de façon à ce qu'il soit le plus spécifique possible du transcrit. Une ou des sondes sont dessinées dans ce fragment qui peuvent être soit des ADNc (donc double brin), soit des oligonucléotides (simple brin) dont la taille varie d'une vingtaine (puces de la société Affymetrix) à 60-70 bases. Chaque sonde ainsi sélectionnée est soit fixée soit synthétisée *in situ* (puces à ADN des sociétés Affymetrix ou Agilent) sur un support solide (membrane de nylon ou lame de verre greffé le plus souvent). Ce support solide qui comporte des centaines, voire aujourd'hui des milliers ou des dizaines de milliers, de zones bien délimitées et localisées appelées les « spots » contenant chacune un fragment d'ADN



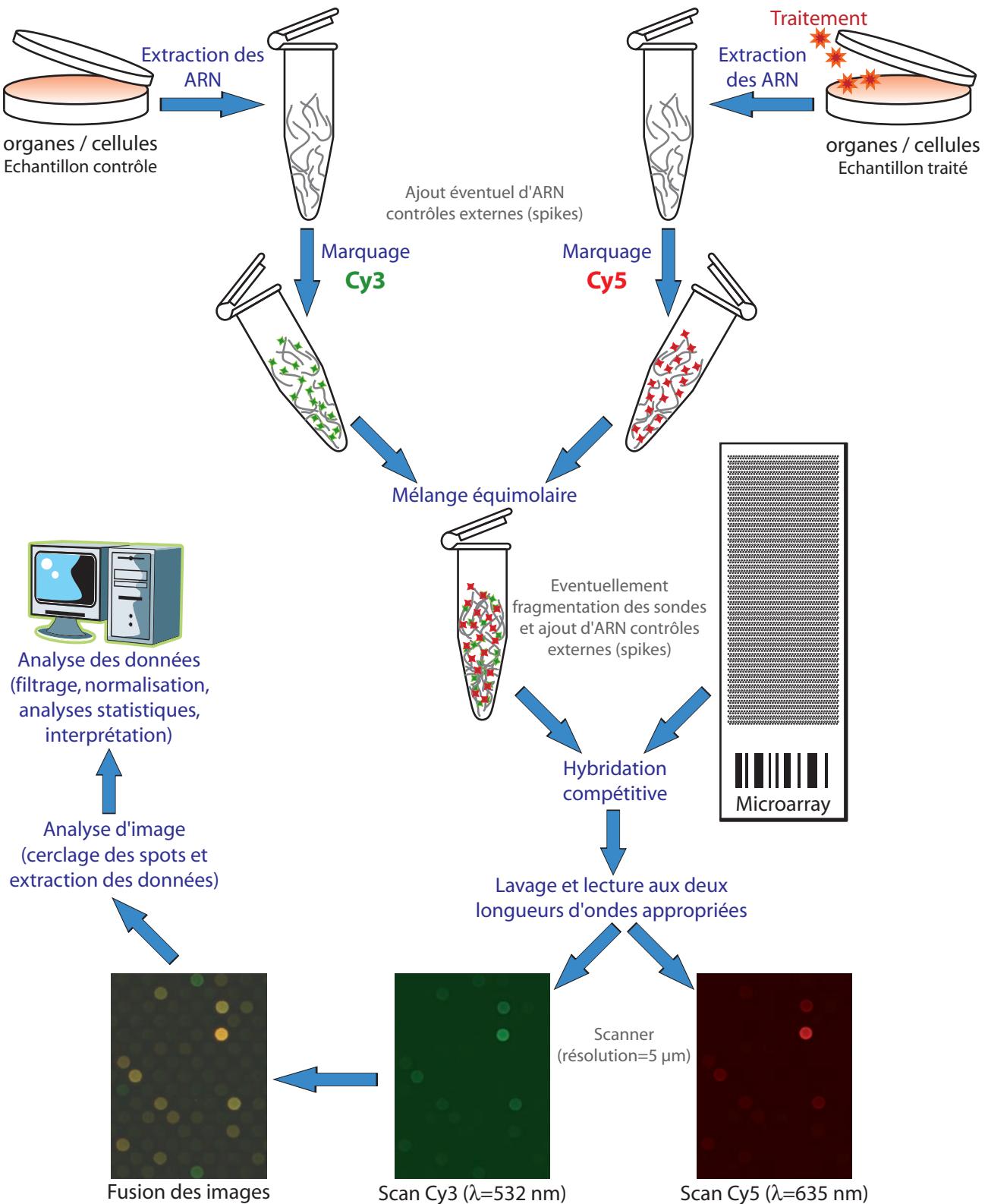
**Figure 48 : Principe général des puces à ADN (simple couleur)**

Les puces à ADN utilisent un principe globalement opposé à celui du dot-blot mais miniaturisé. Un grand nombre de fragments d'ADN spécifiques des différents transcrits du génome sont déposés ou synthétisés *in situ* en des points définis (les spots) d'un support solide (nylon ou lame de verre généralement). Les puces à ADN ainsi produites sont ensuite hybridées avec les échantillons d'ARN qui ont subit un marquage (radiomarquage ou fluorescence le plus souvent) et sont issus des différents échantillons biologiques. Après une série de lavages, les puces sont lues à l'aide d'un scanner de haute résolution et les images résultantes sont analysées pour en extraire l'intensité de chaque spot ainsi qu'une série de données complémentaires.

spécifique d'un transcrit constitue ce que l'on nomme une **puce à ADN**. On distingue classiquement les **microarrays** possédant une très forte densité de spots (plusieurs centaines à plusieurs milliers de spots par cm<sup>2</sup>) des **macroarrays**, généralement réalisés sur membrane de nylon et possédant une moindre densité (de l'ordre de la centaine de spots par cm<sup>2</sup>). Les ARNtotal ou polyA+ extraits des échantillons à étudier sont marqués par diverses méthodes que nous évoquons ci-après de manière à ce que chaque transcrit soit représenté, sous forme marquée (radiomarquage ou utilisation de fluorophores) en proportion de sa quantité initiale dans l'échantillon. Ces échantillons (souvent appelés cibles comme recommandé dans The Chipping forecast, 1999, Nature Genetics, 21, suppl.) sont ensuite hybridés aux puces à ADN qui subissent ensuite des lavages pour réduire les signaux aspécifiques et sont enfin scannées (en fonction de la méthode de marquage des échantillons, les techniques les plus utilisées sont l'écran au phosphore puis l'imageur au phosphore ou l'utilisation d'un scanner de fluorescence à haute résolution). Les images obtenues sont ensuite analysées par des logiciels dédiés pour extraire en particulier les intensités de chaque spot.

#### *ELEMENTS DE DETAIL DES DIFFERENTES ETAPES :*

Les premières puces à ADN étaient en fait constituées de banques de clones bactériens (chaque clone contenant un ADNc cloné dans un plasmide) déposées par un robot sur une membrane de nylon et lysées directement sur la membrane. Par la suite, cette méthode de dépôt robotisé a été conservée mais miniaturisée pour augmenter la densité et réduire la taille des dépôts. L'utilisation de clones bactériens a part ailleurs laissé la place à l'utilisation de produits de PCR issus de l'amplification par des amorces universelles des ADNc clonés dans les plasmides ou bien à l'utilisation d'oligonucléotides longs (60-70mers) synthétisés chimiquement. Le principal défaut de la méthode par dépôt robotisé est la formation de spots de forme irrégulière et présentant souvent un aspect en anneau (parfois appelé effet « donut » en référence au pain brioché du même nom). L'utilisation d'aiguilles fonctionnant par le biais de systèmes piézoélectriques de manière similaire à une imprimante jet d'encre et donc sans contact avec le support solide semble limiter en partie cet effet. Deux sociétés ont également développé des méthodes de synthèse *in situ* des oligonucléotides qui permettent l'obtention de spots bien plus réguliers. La société Affymetrix utilise une méthode de photolithographie pour synthétiser des oligonucléotides courts (~15-20mers) *in situ* sur lame de verre (pour une synthèse, voir Lipshutz et al., 1999). Les puces Affymetrix possèdent une vingtaine d'oligonucléotides (appelés PM pour Perfect Match) pour chaque transcrit ainsi que, pour chacun de ces oligonucléotides, un oligonucléotide de séquence identique dont seule une base



**Figure 49 : Principe des puces à ADN double couleur**

L'utilisation de lames de verre pour le dépôt des sondes spécifiques des transcrits permet la marquage de deux ARNs à comparer à l'aide de fluorophores différents (cyanines 3 -Cy3- et 5 -Cy5-). Les deux échantillons à comparer sont hybridés sur la même puce qui est alors lue aux deux longueurs d'onde d'émission des deux fluorophores utilisés. Les images présentées ici proviennent d'une hybridation de pools d'ARN totaux de foie de souris C57BL/6J traitées (Cy3) ou non (Cy5) avec du fipronil réalisée par notre équipe sur une lame Agilent Whole mouse genome 44K scannée sur un scanner Axon 4100A (résolution 5µm, PMT 532=500, PMT635=600) à la plateforme biopuces du Génopôle de Toulouse et analysée avec le logiciel GenPix Pro 6.0. En haut et en bas de l'image fusionnée, on distingue deux spots nettement verts correspondant à des gènes vraisemblablement induits au niveau hépatique par le Fipronil.

centrale de l'oligonucléotide a été mutée (appelés MM pour MisMatch). Ces derniers sont utilisés pour évaluer la présence d'hybridation non spécifique et les méthodes initiales d'analyse de ces puces suggéraient d'utiliser la somme normalisée des différences de signaux (PM-MM) pour mesurer l'abondance du transcrit correspondant. Le développement de méthodes de modélisation utilisées notamment pour la normalisation et la détection des oligonucléotides aberrants (voir notamment Speed, 2003, pp 1-20) remplacent aujourd'hui ces calculs initiaux et les méthodes les plus récentes s'affranchissent même complètement de l'utilisation des données issues des oligonucléotides MM (implémentée dans le package affy de Bioconductor, Gentleman et al., 2004; méthode RMA pour robust multi-array average Irizarry et al., 2003; Parmigiani, 2003, pp 102-119). Le choix d'une méthode pour rendre compte du niveau d'expression des gènes à partir des données fournies par les puces Affymetrix peut être guidé grâce à la mise en place d'un site internet ([affycomp.biostat.jhsph.edu/](http://affycomp.biostat.jhsph.edu/)) comparant différentes méthodes sur des jeux de données dont les régulations géniques sont connues (Cope et al., 2004) via l'utilisation d'ARN spikes (voir paragraphe III.3.2 pour leur définition). La société Agilent Technologies, issue de la société Hewlett Packard, a, quant à elle, valorisée son expérience dans le domaine des imprimantes à jet d'encre en adaptant cette technologie à la synthèse *in situ* d'oligonucléotides longs sur des lames de verre (Hughes et al., 2001). Leur technologie permet la formation de spots dont l'intensité est bien plus homogène que lors d'un dépôt robotisé et autorise aujourd'hui des densités considérables (plus de 240000 spots actuellement et bientôt de l'ordre de  $10^6$  spots sur une lame de verre de 2,5cm x 7,5cm, voir [www.chem.agilent.com](http://www.chem.agilent.com)).

Plusieurs techniques de marquage des ARNs ont également été développées pour assurer à la fois un lien de proportionnalité avec la composition du mélange complexe d'ARN de départ et un marquage suffisant permettant d'assurer une bonne détection des ARNs, même faiblement exprimés, sans pour autant saturer trop les signaux des gènes fortement exprimés. La technique sans doute la plus simple, utilisée dans le cadre de nos expériences, consiste à partir d'une quantité appréciable d'ARN total (de l'ordre de quelques µg, l'utilisation d'ARN polyA+ est également parfaitement possible, mais d'après les quelques essais que nous avons pu faire cela n'améliore pas sensiblement la qualité des résultats obtenus) et de faire subir à ces ARNs une transcription reverse (RT) en présence d'un nucléotide marqué (de l' $[\alpha-^{33}\text{P}]$ dCTP dans nos expériences mais un marquage fluorescent est, en théorie, également possible). Pour cette étape, il est courant d'utiliser une transcriptase reverse fonctionnant à une température relativement élevée (42°C comme les SuperScript II ou III par exemple commercialisées par Invitrogen et utilisées dans nos protocoles) ce qui

permet de limiter la formation de structures secondaires au sein des ARNs et augmente ainsi le rendement de la transcription reverse. Deux choix pour l'amorçage de la RT s'offre à l'utilisateur : soit des oligodT, soit des amores aléatoires (hexamères aléatoires par exemple). En théorie au moins, l'avantage des oligodT est de ne faire subir une RT qu'aux ARNmessagers car seuls ceux-ci comportent une queue polyA et donc pas aux ARN ribosomiques qui représente la large majorité des ARN totaux ou aux divers petits ARNs (ARN de transfert ou microARN par exemple). En pratique, nous avons constaté qu'une RT en présence uniquement d'oligodT sur un échantillon d'ARN total nous permettait de mesurer l'expression d'un ARN ribosomique mitochondrial : l'ARN ribosomique 16S dont le signal sur nos puces à ADN était encore très fort malgré l'absence d'une queue polyA à l'extrémité 3' de cet ARN (PGP Martin, données non présentées). L'étude de la séquence de cet ARN (région 1094-2675 de la séquence V00711) révèle la présence de plusieurs régions contenant des séries de A (cinq régions contenant 5 A d'affilée, trois régions contenant 6 A d'affilée et une région en contenant 7 d'affilée) susceptibles de constituer des zones d'amorçage par le biais d'oligodT et d'expliquer ainsi la transcription reverse de cet ARN non messager. A l'opposé, l'utilisation d'oligodT est susceptible de poser principalement deux problèmes :

1. le premier est que, en particulier pour les transcrits longs, il n'est pas garanti que la partie 5' du transcrit ait été correctement reverse transcrise en raison de sa distance importante du site d'initiation de la transcription reverse. Ce problème est d'autant plus critique si les fragments d'ADN présents sur la puce à ADN n'ont pas été spécifiquement sélectionnés pour être localisés systématiquement dans la région 3' des transcrits à étudier.
2. le second est lié au fait que l'ensemble des produits issus de cette méthode de marquage contiennent une queue oligodT. Ces produits de marquage vont donc avoir la tendance à se fixer sur toute série de A, même relativement courte, présente sur la puce à ADN, que ce soit au niveau d'une séquence présente dans un spot (hybridation aspécifique) ou au niveau de l'ADN utilisé pour bloquer les sites de fixation aspécifique (classiquement de l'ADN de sperme de saumon) situés autour des spots (formation d'un bruit de fond élevé). Cet effet, qualifié parfois « d'effet polyA » a été rapidement identifié (voir par exemple Nguyen et al., 1995) et peut par exemple être corrigé soit par l'utilisation d'amores aléatoires (voir ci-dessous), soit par une préincubation du produit de marquage avec des oligonucléotides constitués de série de A (dA<sub>80</sub>) qui bloquent ainsi les queues oligodT.

Alternativement, il est possible d'utiliser un mélange complexe d'amorces courtes (6 à 8 bases généralement) dont la séquence est aléatoire. L'amorçage de la transcription reverse se fera alors en différents points sur chaque transcript et produira des fragments plus courts qu'avec des oligodT mais couvrant à coup sûr la totalité de la séquence des transcrits (région 5' comprise). Dans le cadre du développement de la puce à ADN INRAarray, constituant la partie méthodologique de ce doctorat, nous avons conduit plusieurs essais pour évaluer l'intérêt d'utiliser des oligodT et/ou des amorces aléatoires (PGP Martin, résultats non présentés). N'ayant pas sélectionné des sondes nucléotidiques spécifiques localisées dans les régions 3' des transcrits, nous nous sommes naturellement orientés initialement sur l'utilisation d'amorces aléatoires. Nos essais ultérieurs ont mis en évidence que pour certains spots (probablement ceux localisés vers la partie 3' des transcrits comme nous avons pu le noter pour certains d'entre eux sans pour autant réaliser une analyse complète sur toute la puce), les résultats étaient améliorés par l'utilisation d'oligodT. A l'opposé, les signaux d'autres spots étaient fortement diminués par l'utilisation d'oligodT par rapport aux amorces aléatoires. Le dernier protocole de marquage mis en place dans notre équipe (et utilisé en routine depuis 2005) utilise donc finalement à la fois des amorces aléatoires et des oligodT pour la reverse transcription, car nous avons pu montrer que cela nous permettait d'obtenir toujours les signaux maximum (par rapport à l'utilisation de seulement l'un des deux types d'amorces) pour l'ensemble des spots et ce sur des tests portant sur plusieurs échantillons d'ARN hépatiques (PGP Martin, données non présentées).

Initialement, les premières expériences réalisées sur membranes de nylon, qui n'autorisent pas un marquage par fluorescence en raison de la fluorescence intrinsèque du nylon, ont utilisé des nucléotides marqués au  $^{32}\text{P}$  (voir par exemple Bernard et al., 1996; Gress et al., 1992; Nguyen et al., 1995; Zhao et al., 1995). Le principal défaut de ce radioisotope est sa forte énergie d'émission (Emax de l'émission  $\beta$  = 1710 KeV) qui conduit le signal obtenu à diffuser de manière significative (parcours des émissions  $\beta$  dans le verre = 3,4 mm et dans le plexiglas = 6,3 mm), provoquant ainsi des chevauchements de signaux à proximité des spots les plus intenses. L'utilisation du  $^{33}\text{P}$  qui possède une moindre énergie d'émission  $\beta$  (249 KeV) permet de pallier en grande partie à ces défauts et conduit ainsi à une diffusion plus limitée du signal (parcours des émissions  $\beta$  dans le verre = 0,3 mm et dans le plexiglas = 0,5 mm) et donc à une meilleure résolution des images. De plus, la demi-vie plus longue du  $^{33}\text{P}$  (25,6 jours contre 14,3 jours pour le  $^{32}\text{P}$ ) permet son stockage à un peu plus long terme et une perte de signal négligeable lors d'expositions longues (3-5 jours) à l'écran au phosphore.

Ultérieurement, l'utilisation de lames de verre pour déposer ou synthétiser *in situ* les fragments complémentaires des transcrits a permis l'utilisation de nucléotides couplés à des fluorophores dans l'étape de marquage des ARNs. Les plus fréquemment utilisés sont les cyanine 3-CTP ou dCTP (Cy3-CTP : pics d'excitation=550 nm et d'émission=568 nm) et cyanine 5-CTP ou dCTP (Cy5-CTP : pics d'excitation=650 nm et d'émission=668 nm). L'utilisation de ces deux fluorophores dont les spectres ne se chevauchent presque pas permet d'hybrider sur une même lame deux échantillons différents (hybridation compétitive, voir [Figure 49](#)) permettant ainsi à l'utilisateur de s'affranchir en partie de la variabilité liée à la fabrication de la puce et d'éviter les problèmes liés à la présence de matériel limitant au sein des spots. Les lames sont ensuite scannées aux deux longueurs d'onde appropriées (généralement  $\lambda=635$  nm pour la Cy-5 et  $\lambda=532$  nm pour la Cy-3) à l'aide d'un scanner de haute résolution. On parle dans ce cas là de puces en **double couleur** par opposition aux puces en **simple couleur** sur lesquelles un unique échantillon est hybridé.

D'autres améliorations ont également été apportées au protocole de marquage des ARNs. En particulier, il a rapidement été observé que les nucléotides marqués aux Cy3 et Cy5 ne s'incorporaient pas de manière équivalente dans les séquences nucléotidiques notamment en raison de l'encombrement stérique un peu plus important de la Cy5 par rapport à la Cy3. Cet effet, susceptible de générer un biais, peut être corrigé par les techniques dites de marquage indirect où un nucléotide modifié (généralement des amino allyl-dNTP ou -NTP) est incorporé dans les deux échantillons lors du marquage (donc avec le même rendement) et auquel est par la suite fixé par couplage chimique soit de la Cy3 soit de la Cy5 en fonction de l'échantillon. Enfin, l'étude par microarray d'échantillons obtenus en quantité très faibles (quelques dizaines ou quelques centaines de ng d'ARN de départ) a été rendue possible grâce à des protocoles utilisant la synthèse d'ARNcopie (cARN). Les échantillons d'ARN initiaux subissent une transcription inverse double brin en présence d'oligodT contenant le promoteur d'une ARN polymérase (T7 ou SP6 ARN polymérase le plus souvent). Les ADNc obtenus sont ensuite transcrits *in vitro* à l'aide de l'ARN polymérase sélectionnée et en présence d'un nucléotide marqué (ou d'un amino allyl-NTP), conduisant ainsi à une amplification, vraisemblablement linéaire, de la population initiale des ARN messagers ainsi qu'à leur marquage simultané.

De nombreuses autres améliorations que nous ne pourrons développer ont été apportées à chaque étape de la fabrication et de l'utilisation des puces à ADN. Elles concernent le support solide et sa capacité à exposer au mieux les sondes nucléotidiques qui y sont fixées, la sélection des fragments complémentaires des transcrits fixés sur la puce

(nature, localisation, taille, température de fusion, pourcentage de G et de C, structure secondaire, etc.), le marquage des échantillons et la capacité à détecter les transcrits faiblement représentés (sensibilité), la lecture des puces par les scanners (intensité des lasers, résolution et temps de lecture en particulier) en lien notamment avec la miniaturisation des spots et l'accroissement de leur densité et l'analyse des images (reconnaissance automatique des spots, cerclage de diamètre variable des spots, estimation du bruit de fond local, choix de la mesure la plus appropriée pour rendre compte de l'intensité des spots comme le signal médian ou le signal moyen par exemple). Ces développements méthodologiques dans leur ensemble ont conduit à améliorer de manière significative la sensibilité, la spécificité et la reproductibilité de la technique des puces à ADN. De nouvelles applications ont également été développées comme l'étude de l'expression des transcrits alternatifs (choix de sondes appropriées à la détection des différents transcrits alternatifs) ou celle des microARN. Le principe de cette technique a part ailleurs été aujourd'hui étendu à bien d'autres domaines d'applications que l'étude de l'expression des gènes comme l'étude des altérations de l'ADN (délétions, insertions, transposition) au niveau génomique (puces CGH pour comparative genomic hybridization), la détection de polymorphismes affectant une unique base (puces SNPs) ou celle de la fixation de protéines (facteurs de transcription, histones, etc.) à l'ADN (puces ChIP-on-Chip). Le même principe est également utilisé pour l'étude des protéines (puces à protéines sur lesquelles peuvent être déposés des anticorps, des antigènes ou des protéines totales provenant de différents tissus) ou celle de l'hybridation *in situ* à haut-débit (puces de tissus).

#### AVANTAGES :

L'avantage majeur des puces à ADN est leur rapidité et leur facilité d'utilisation. Alors que la mise en place de l'outil peut s'avérer longue car elle nécessite la manipulation d'une grande quantité de données de séquences (aspects bioinformatiques), la synthèse et la fabrication de nombreuses sondes et la standardisation des modes de fabrication pour améliorer la reproductibilité, une fois les puces à ADN disponibles, la quantité de données générées par unité de temps est sans comparaison avec les méthodes présentées précédemment. A titre d'exemple, on peut estimer à environ 2 jours le temps de travail nécessaire à l'obtention des images de puces à ADN commerciales en partant des échantillons biologiques. En termes de reproductibilité, la comparaison de données obtenues à partir d'échantillons identiques sur des plateformes identiques (le terme plateforme désignant ici un type de microarray d'une marque ou d'une fabrication donnée et utilisé selon un protocole

standardisé) donne normalement de bons résultats. En revanche, des questions ont été soulevées dans la littérature sur la capacité de différentes plateformes (et/ou de différents laboratoires utilisant la même plateforme) à obtenir des résultats identiques (voir par exemple Mecham et al., 2004; Rogojina et al., 2003; Shippy et al., 2004; Tan et al., 2003). Des données plus récentes, issues notamment d'un effort collectif visant à évaluer ces technologies (projet MAQC pour MicroArray Quality Control), montrent néanmoins des résultats rassurants (Bosotti et al., 2007; Canales et al., 2006; Guo et al., 2006b; Patterson et al., 2006; Shi et al., 2006; Shippy et al., 2006) et mettent notamment en exergue les rôles importants joués par le filtrage des données, par la procédure de sélection des gènes différentiellement exprimés et par la mise en concordance des annotations des différentes plateformes.

Alors que le caractère pangénomique de ces outils ne peut bien sûr être envisagé qu'en possession de la séquence complète du génome et surtout des transcrits (ce qui est plus difficile à obtenir), il est évident que la couverture du génome assurée par les microarrays disponibles pour les principaux organismes modèles utilisés en recherche animale (homme, souris, rat, drosophile, nématode, levure, etc.) est au moins équivalente à celle des autres méthodes que nous avons évoqué. Le principal problème qui limite le caractère pangénomique de ces outils est la qualité et la sensibilité des mesures. Le signal obtenu pour chaque spot dépend d'une multitude de facteur dont celui qui nous intéresse en premier lieu à savoir le niveau d'expression du gène. La plupart des autres facteurs peuvent être assez aisément standardisés (fabrication des puces, temps et température d'hybridation et de lavage, marquage des échantillons, acquisition et analyse des images, manipulation des données) mais certains et en particulier le choix de la sonde nucléotidique déposée dans chaque spot sont plus difficiles à standardiser car on ne peut être certain de choisir systématiquement la sonde optimale (celle qui fournit un signal spécifique le plus élevé possible) pour chaque transcrit malgré la maîtrise de multiples paramètres (température de fusion des sondes en solution, pourcentage de G et de C, taille, localisation sur les transcrits, structure secondaire des sondes et des transcrits par exemple). La preuve de cette maîtrise difficile du choix optimal des sondes est apportée par la publication régulière de rapports sur le sujet (voir par exemple Held et al., 2003; Held et al., 2006; Letowski et al., 2004; Pozhitkov et al., 2006; Wu et al., 2007; Zhang et al., 2005a; Zhang et al., 2003a) ainsi que par la stratégie adoptée par une société bien placée sur ce marché (Agilent Technologies, voir Paredes et al., 2007), qui a choisi, après avoir dessiné via des outils bioinformatiques plusieurs sondes pour chaque gène d'un génome donné, de tester systématiquement au moins 5 sondes différentes pour chaque transcrit afin d'assurer la sélection de la sonde spécifique fournissant le meilleur signal. Naturellement, le

choix de la sonde optimale devient essentiellement critique lorsque le niveau d'expression du gène est faible car alors, l'influence sur le signal du niveau d'expression est diminué et redonne ainsi de l'importance au choix de la sonde. Outre ce problème de présence systématique au sein des microarrays de sondes suboptimales (et parfois peu spécifiques d'après notre expérience au laboratoire), les mesures issues des puces à ADN souffrent régulièrement de pertes de données liées à des images de moindre qualité (bruit de fond non homogène, présence de traces ou tâches sur les puces, spots irréguliers par exemple). Ces données manquantes, si elles altèrent effectivement dans l'absolu le caractère pangénomique des expériences, n'affectent néanmoins qu'une part très limitée de l'information générée et, moyennant la réalisation d'un nombre suffisant de réplicats, l'utilisation de méthodes d'analyse statistique gérant la présence de données manquantes permet à l'utilisateur de s'affranchir aisément de ce problème.

Contrairement aux méthodes basées sur l'amplification de sous-populations d'ARN (paragraphe III.2.2), l'identification des transcrits régulés est immédiate avec les puces à ADN permettant à l'utilisateur de se concentrer autant sur des ensembles de gènes régulés (groupes de gènes corégulés, identification des principales fonctions biologiques affectées par un traitement à l'aide des bases de données d'annotation des gènes comme Gene Ontology par exemple, voir [www.geneontology.org](http://www.geneontology.org)) que sur des régulations particulières (identification de nouveaux gènes régulés).

Enfin, la diffusion importante de la technique des puces à ADN a conduit à la mise en place de nombreux groupes de réflexion, d'échange et de développement de méthodes pour ces techniques, constituant des communautés actives au sein desquelles une multitude de ressources pour l'obtention et l'analyse des résultats sont disponibles. Par exemple, la société MGED (Microarray and Gene Expression Data, voir [www.mged.org](http://www.mged.org)). vise à faciliter la diffusion et l'échange des données issues des expériences sur le transcriptome et a notamment mis en place des standards pour la diffusion des données (MIAME pour Minimum information about a microarray experiment) ainsi qu'un format (MAGE-ML pour Microarray gene expression – Markup language) et un modèle (MAGE-OM pour Microarray gene expression – object model) d'échange de données. Un autre exemple de communauté très active est celle du projet Bioconductor (voir [www.bioconductor.org](http://www.bioconductor.org) et Gentleman et al., 2004) dédié au développement de méthodes d'analyse statistique et bioinformatique des expériences sur le transcriptome à partir de la mise en place de packages et d'environnements pour le logiciel gratuit R (voir [www.r-project.org](http://www.r-project.org) et Dudoit et al., 2003a). De nombreux autres

exemples peuvent être trouvés sur internet, permettant ainsi à l'utilisateur néophyte de trouver rapidement des réponses à ses questions.

#### *INCONVENIENTS / LIMITES :*

La technique des puces à ADN présente deux inconvénients majeurs : son coût et la nécessité de disposer de bases de données les plus complètes possible sur les séquences des transcrits de l'organisme étudié. Bien qu'il ait considérablement diminué, le coût des expériences de puce à ADN reste néanmoins élevé. L'achat de matériel coûteux (robots de dépôt, scanner, fours ou automates d'hybridation) n'est pas toujours possible pour de petites équipes ce qui a conduit au développement de plateformes au niveau local ou national. En France, la plupart de ces plateformes ont été mises en place dans le cadre des Génopoles (pour un exemple au niveau local, voir <http://genopole-toulouse.prd.fr>) et/ou de grands instituts de recherche (INRA, INSERM, CNRS, IGBMC à Strasbourg, Institut Pasteur) et offre un accès facilité à ces technologies, y compris à des équipes de taille modeste. De nombreuses solutions commerciales sont également disponibles, dont certaines ont déjà été abordées ici, mais leur coût reste encore conséquent.

Contrairement aux méthodes évoquées dans les deux paragraphes précédents, les puces à ADN requièrent une connaissance a priori des séquences des transcrits à étudier, ce qui limite leur utilisation aux génomes séquencés et annotés. Cependant, il existe aujourd'hui un nombre considérable de génomes séquencés (voir notamment [www.ncbi.nlm.nih.gov/Genomes](http://www.ncbi.nlm.nih.gov/Genomes)) et l'amélioration des techniques de séquençage et d'annotation laisse espérer encore une accélération de la production de données génomiques. Les principaux modèles utilisés en pharmacologie et toxicologie étant séquencés, ce domaine d'investigation bénéficie d'une large gamme d'outils d'étude du transcriptome par puces à ADN.

Enfin, une autre limite à l'utilisation des puces à ADN, importante dans un contexte pratique de mise en œuvre, est lié au fait que ces techniques génèrent, comme nous l'avons précisé, une quantité considérable de données. Ces données sont en partie bruitées (variabilité d'origine technique, artéfacts, variabilité biologique) et l'objectif du biologiste est d'en extraire une information pertinente, qui ne représente parfois qu'une très faible proportion de l'information générée (le pourcentage de gènes modulés par un traitement dans un échantillon biologique donné, n'excède que rarement quelques % des gènes du génome). La manipulation, le stockage, le filtrage, la normalisation, l'analyse et l'interprétation de ces données constituent donc autant de challenges qui, selon moi, sont bien plus difficiles et longs

à relever que la simple mise en œuvre des expériences elles-mêmes. Ces étapes nécessitent l’acquisition de compétences en mathématique, statistique, informatique et bioinformatique, relativement peu diffusées et mises en valeur jusqu’à présent dans la communauté des biologistes. L’absence de ces compétences a malheureusement donné lieu à l’abandon de certains projets basés sur les puces à ADN. A Toulouse, nous avons mis en place un groupe de travail sur l’analyse statistique des données issues des expériences sur le transcriptome (voir [www.lsp.ups-tlse.fr/Programmes/Biopuces/](http://www.lsp.ups-tlse.fr/Programmes/Biopuces/)), animé par des statisticiens de l’Institut de Mathématique de Toulouse qui collaborent avec de nombreuses équipes de recherche en biologie. Impliqué précocement dans la mise en place de ce groupe de travail, j’ai eu l’occasion de constater qu’il répondait à une demande forte d’interaction entre la communauté des biologistes et celle des statisticiens et pourrait être amené à se développer en plateforme de services et/ou de conseil dans le proche futur.

En conclusion, les puces à ADN offrent aujourd’hui la possibilité d’analyser et de mesurer l’expression des gènes à l’échelle du génome avec un débit et une rapidité pour l’instant inégalés. La masse de données générée par ces techniques s’accroît à une vitesse impressionnante et leur analyse nécessite la mise en œuvre de méthodes pas toujours triviales que nous présenterons brièvement au paragraphe III.4 ce qui stimule fortement l’interaction du biologiste avec d’autres communautés scientifiques conduisant souvent à un enrichissement mutuel des différents acteurs des projets.

### **III.3. DEVELOPPEMENT DE LA PUCE INRArray**

Dans ce paragraphe, nous exposons les principales étapes qui nous ont conduits au développement de la puce à ADN INRArray dédiée à l’étude des régulations géniques exercée par le récepteur PPAR $\alpha$  et par les autres partenaires du récepteur RXR (RN de classe II). Le développement de cet outil et l’acquisition des compétences nécessaires à l’analyse des données qui en sont issues a constitué le cœur du travail méthodologique de ce doctorat.

#### **III.3.1. Justification des choix technologiques**

Les techniques exploratoires présentées aux paragraphes III.2.1 et III.2.2. ne répondaient pas vraiment à nos objectifs qui consistaient à produire un nombre conséquent de signatures transcriptionnelles dans des contextes d’activation variés du récepteur PPAR $\alpha$  sans forcément chercher à être exhaustif sur les régulations observées mais plutôt à établir des

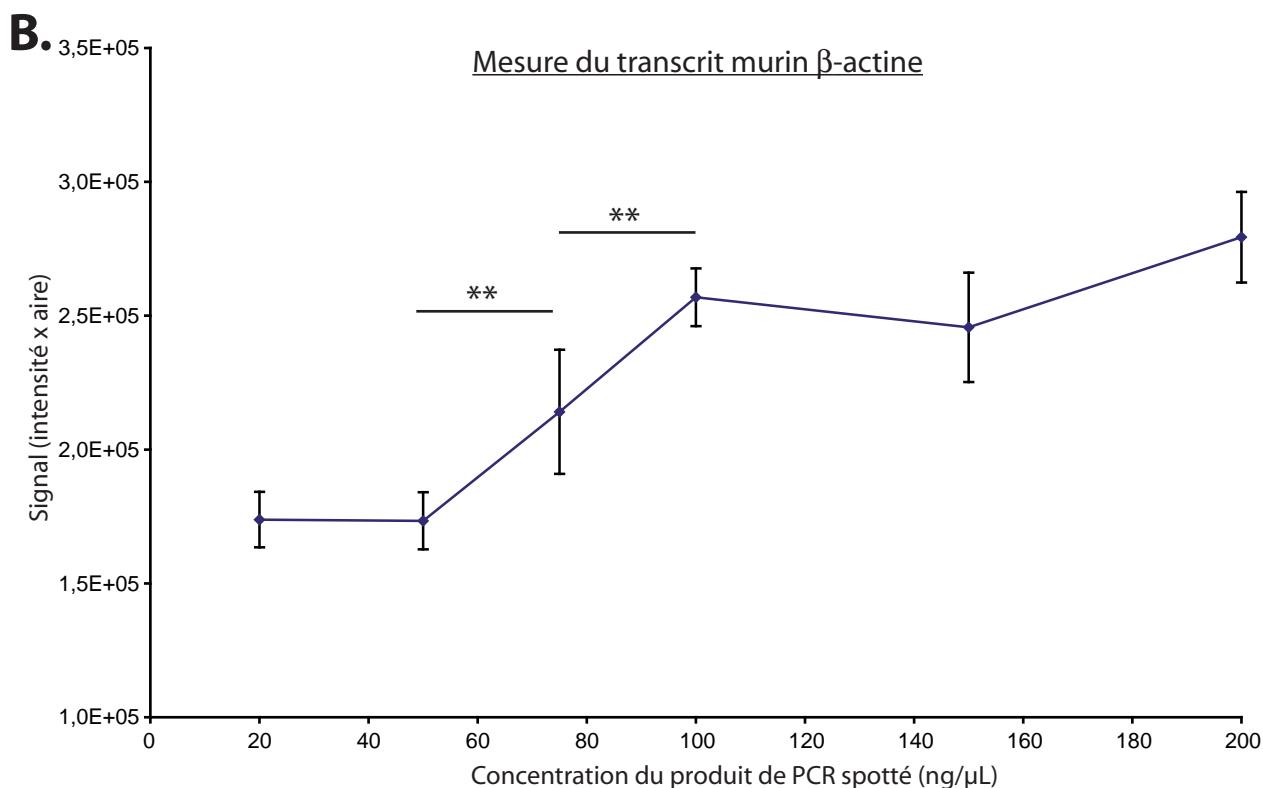
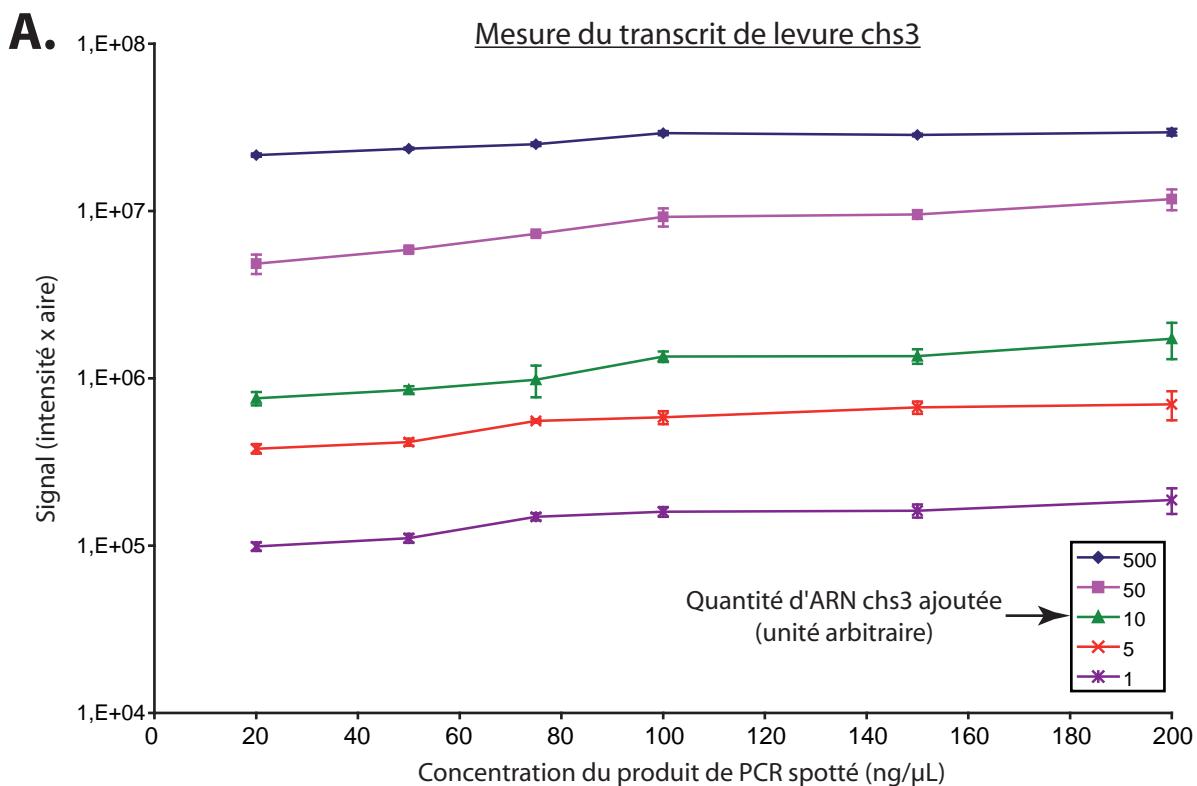
empreintes de l'activation du récepteur qui puissent être comparées qualitativement et quantitativement entre les différentes situations d'activation . Nous nous sommes donc naturellement orientés vers la technologie nous permettant d'atteindre ce but à savoir les puces à ADN. Lorsque ce projet a été initié en 1999/2000 (PGP Martin, stage de DEA et début de doctorat), cette technologie était encore considérée comme récente, les plateformes commerciales ou académiques en étaient encore à leurs débuts et les coûts des outils commerciaux « clef en main » nous étaient pour la plupart inabordables en raison du grand nombre d'échantillons que nous souhaitions analyser. Après des essais préliminaires de fabrication de « puces maison » par sérigraphie d'une encre hydrophobe sur membrane de nylon (PGP Martin, stage de DEA), nous avons pu bénéficier de la mise en place de la plateforme « biopuces » du génopôle de Toulouse (voir [biopuce.insa-toulouse.fr](http://biopuce.insa-toulouse.fr)). Disposant d'une bonne expérience et du matériel adapté à la réalisation d'hybridations de produits radiomarqués sur membrane de nylon, nous nous sommes rapidement orienté vers ce choix technologique (support nylon et radiomarquage). L'utilisation du radiomarquage nous permettait de plus d'obtenir une bonne sensibilité alors que l'utilisation d'un support nylon diminuait les coûts de production des puces. La plupart de nos protocoles expérimentaux étant constitués d'au moins deux facteurs (typiquement génotype et traitement), l'utilisation de puces simple couleur facilitait également grandement la réflexion sur les plans d'hybridation en nous limitant de fait à une approche systématique du type un échantillon/une puce. Ces choix étant effectués, il nous restait à définir ce que nous allions déposer sur nos puces et donc mesurer par leur intermédiaire. Une approche pangénomique nous a rapidement semblé peu rentable en raison de la faible part d'information pertinente que nous allions pouvoir tirer de chaque expérience (la majeure partie des gènes exprimés n'étant pas modulés dans nos expériences) et du nombre plus limité d'échantillon que nous aurions pu nous permettre d'analyser. Nous avons donc opté pour l'étude d'une sélection ciblée de gènes et conçu ainsi un macroarray dédié que nous avons nommé INRArray 01 dans le but éventuel (mais malheureusement jamais réalisé) de fédérer des initiatives similaires au sein de notre institut de recherche. Les principales sources d'information qui ont alors guidé notre sélection de gènes cibles ont été 1) la réalisation de quelques criblages pangénomiques sur puces à ADN commerciales (Martin et al., 2005), 2) l'abondante littérature sur les régulations géniques opérées par PPAR $\alpha$  et les autres partenaires du récepteur RXR et 3) les données non publiées issues de notre laboratoire ou de nos diverses collaborations lorsque l'accord nous en était donné. Un protocole standardisé de clonage des ADNc spécifiques des transcrits d'intérêt a alors été mis en place (voir paragraphe III.3.3) qui a, par la suite, pu être adapté à un débit de

clonage plus élevé (de l'ordre d'une quinzaine de clonages par manipulateur et par semaine dans les périodes où cette activité est prépondérante dans notre équipe). Ces ADNc ont ensuite été regroupés et déposés sur des membranes de nylon grâce à un automate de la plateforme transcriptome du Génopôle de Toulouse pour générer les puces à ADN INRAarray 01. Nous nous sommes équipés du logiciel Image Master Array 2.0 (Amersham Biosciences, France) pour l'analyse des images des puces à ADN obtenues avec notre scanner STORM en vérifiant que ce logiciel prenait bien en compte la transformation racine carrée introduite par ce type de scanner dans les images générées (Ramdas et al., 2001). Dans les trois paragraphes suivants, nous présentons les étapes de mise au point et de validation de cet outil, le clonage des ADNc sélectionnés, puis les différentes générations de puces INRAarray qui se sont succédées.

### **III.3.2. Etapes de mise au point de la puce INRAarray**

Nous avons déjà évoqué dans les paragraphes précédents un certain nombre d'étapes de mise au point et d'optimisation de notre puce à ADN et notamment celles concernant le **marquage des échantillons**. Dans ce cadre, d'autres mises au point ont naturellement été conduites que nous ne pourrons présenter en détail ici. Elles ont porté sur la quantité d'ARN à marquer (essais de 1 à 10 µg en cherchant à obtenir la meilleure activité spécifique pour un marquage total à peu près constant, une quantité de 4-5µg donnant les meilleurs résultats), sur les quantités de dCTP froid et d' $[\alpha\text{-}^{33}\text{P}]$ dCTP à ajouter (nous avons finalement éliminé l'ajout de dCTP froid qui réduisait l'activité spécifique des produits de marquage, le dCTP radiomarqué ne semblant pas limitant pour la synthèse des ADNc), sur les concentrations d'oligodT et d'amorces aléatoires à utiliser (voir paragraphe III.2.3) ou encore sur la reverse transcriptase à utiliser (tests avec diverses enzymes issues de différents fournisseurs nous conduisant à retenir, après analyse de puces hybridées, la SuperScript II, Invitrogen, France). Le protocole de marquage a été maintenant stabilisé depuis environ 2 ans (depuis fin 2005) et le protocole complet d'utilisation des puces INRAarray est présenté en [annexe 1](#) sous la forme qu'il a prise fin 2006 dans la base de données des protocoles sous assurance qualité du Laboratoire de Pharmacologie et Toxicologie de l'INRA.

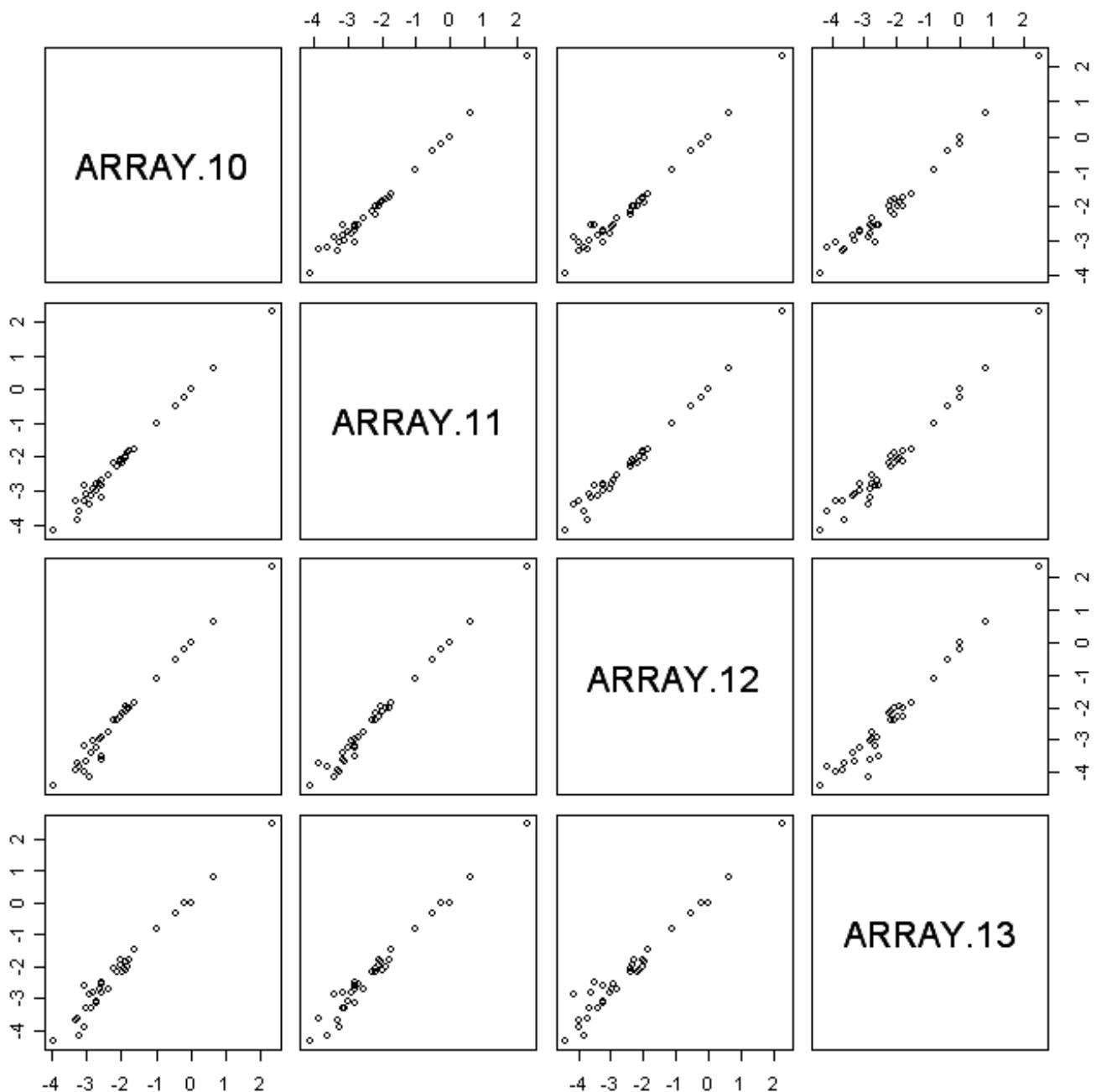
D'autres aspects du protocole ont également fait l'objet de tests visant à choisir les conditions expérimentales nous fournissant les signaux spécifiques les plus élevés et les plus reproductibles. La **quantité de matériel déposé** sur les spots des puces à ADN ne doit pas être limitante de manière à éviter les variations de signaux dues à de légères différences dans les quantités de matériel déposé sur différentes puces (variabilité due à la fabrication des



**Figure 50 :** Mise au point de la quantité de matériel spotté sur les puces INRArray

Une série de puces INRArray 01 contenant des gammes de dépôts des sondes chs3 (A) et  $\beta$ -actine (B) a été spottée (dépôts réalisés en quadruplicats) à la plateforme biopuces du Génopôle de Toulouse (spotteur Eurogridder microarrayer, Virtek/Biorad, 5 allers-retours des aiguilles avec 20 nL de dépôt par contact avec la membrane, la quantité déposée vaut donc  $0,001 \times C \times 5 \times 20$  en ng de sonde où C est la concentration reportée ici en abscisse). Cinq lots de puces ont été hybridées avec un ARN total hépatique de souris additionné d'une quantité variable de l'ARN de levure chs3 couvrant une gamme de 1 à 500 (unité arbitraire, graphique A). Un test de student (\*\* p<0.01) a été réalisé pour comparer les signaux obtenus par les concentrations de dépôts successives (résultats présentés uniquement sur le graphique B mais parfaitement comparables pour le graphique A). Au delà de 100 ng/ $\mu$ L, aucune différence de signal n'est significative.

puces). Nous avons donc réalisé des gammes de concentration d'ADNc déposé pour une sonde ciblant un ARN de levure produit in vitro (chs3) et utilisé comme spike dans nos expériences de microarray récentes (voir ci-dessous). Cet ARN a été incorporé dans des échantillons d'ARN totaux de foie de souris à diverses concentrations couvrant une gamme de 1 à 500 afin d'envisager le caractère limitant de la quantité de matériel déposé dans diverses gammes d'intensité. Une gamme de dépôt similaire a également été réalisée pour un gène endogène de souris : la  $\beta$ -actine, dans le but de confirmer les résultats obtenus avec l'ARN spike. Les résultats obtenus sont présentés en [Figure 50](#). Les courbes de l'intensité en fonction de la quantité de matériel déposé ont été générées et, présentent globalement un plateau apparaissant à la concentration de 100 ng/ $\mu$ L. Des résultats comparables ont été obtenus avec le gène de levure chs3 et avec la  $\beta$ -actine ([Figure 50 A et B](#)). A partir de ces résultats, nous avons choisi de fixer à 200 ng/ $\mu$ L la concentration minimale requise pour les produits de PCR déposés sur nos puces (5 allers-retours de l'aiguille par spot avec 20 nL déposés à chaque aller-retour soit une quantité totale d'ADNc dans le spot de 20 ng) afin que le signal ne dépende pas ou peu de la quantité de matériel déposé. Cette concentration est la concentration maximale testée dans l'expérience présentée en [Figure 50](#) qui semble bien être située dans la phase de plateau de signal. A cette concentration, nous avons observé dans ces expériences une variabilité légèrement accrue du signal dans certains cas (voir [Figure 50A](#)), un effet qui a pu être éliminé par la suite en incorporant une étape d'hybridation-déshybridation à blanc (i.e. sans produit de marquage) des puces nouvellement produites (voir protocole en [annexe 1](#)). Des **lavages** relativement intensifs ont également été adoptés (voir protocole en [annexe 1](#)) car si ceux-ci réduisent en partie le signal, nous avons pu constater qu'ils augmentaient considérablement la reproductibilité des données d'une puce à l'autre, sans doute en éliminant une bonne partie de la fixation aspécifique sur les spots. Ces observations nous semblent en accord avec les résultats obtenu par une étude récente de modélisation des intensités issues des puces à ADN Affymetrix qui soulignent l'importance des étapes de lavages et de la fixation aspécifique sur les signaux obtenus (Held et al., 2006). L'utilisation de dA<sub>80</sub> en préincubation avec les sondes marquées limite également dans notre protocole la fixation aspécifique des produits marqués via les oligodT aux séries de A potentiellement présentes dans nos sondes ou bien dans l'ADN de sperme de saumon utilisé pour bloquer les sites de fixation aspécifiques localisés autour des spots. Une hybridation a également été réalisée avec de l'ADN hautement répétitif (ADN Cot-1) radiomarqué pour contrôler que nos sondes, dans leurs conditions d'utilisation, ne fixaient pas ce type de séquences (PGP Martin, résultats non



**Figure 51 : Exemple d'hybridation d'un ARN en quadruplicat sur les puces INRArray 01**  
 Un ARN hépatique de souris a été séparé en 4 aliquots qui ont été marqués puis hybridés indépendamment sur 4 puces INRArray 01 (version antérieure à l'INRArray 01.1 ne portant que 36 sondes, dont un contrôle externe, spottés en dupliques. Cette matrice de graphiques présente dans ses triangles inférieures et supérieures (qui sont identiques) les graphiques des données transformées en logarithme des puces prises deux-à-deux, les unes contre les autres. Les corrélations linéaires des puces prises deux à deux sont toutes supérieures à 0,97.

présentés). Plusieurs tampons et températures d’hybridation ont enfin été évalués pour aboutir au protocole présenté en [annexe 1](#).

La **reproductibilité** de nos hybridations a été contrôlée précocement au cours du processus de mise au point de l’INRAArray 01 via la réalisation de réplicats techniques (même ARN marqué plusieurs fois et hybridés sur plusieurs puces différentes). Un exemple de résultat obtenu précocement au cours du processus de mise au point et réalisé via le marquage et l’hybridation en quadruplicat d’un ARN total hépatique de souris est présenté en [Figure 51](#). Chacune de nos expériences comporte de plus des réplicats biologiques (ARN issus de souris d’un même groupe d’individus et hybridés sur des puces différentes) et notre procédure de prétraitement des données inclue systématiquement des comparaisons deux à deux des résultats obtenus à partir d’ARN appartenant à un même groupe d’individus. Ce travail, relativement fastidieux dans le cadre de plans expérimentaux élaborés, nous rassure néanmoins sur la reproductibilité des données, permet de détecter aisément des puces aberrantes et d’identifier des erreurs éventuellement commises par le logiciel de quantification dans le cerclage des spots. Les corrélations linéaires moyennes observées entre deux puces hybridées avec des ARNs issus d’un même groupe d’individu sont typiquement supérieures à 0,9 et sont normalement mentionnées dans nos rapports de résultats (Martin et al., 2007; Martin et al., 2005).

Rapidement, nous avons souhaité nous doter d’un outil permettant une normalisation de nos données ainsi que la réalisation de contrôle qualité des données générées. Comme nous le verrons dans le paragraphe III.4., la plupart des méthodes couramment utilisées pour la normalisation des données des puces à ADN pangénomiques ne sont pas appropriées, au moins du point de vue théorique, pour les puces à ADN dédiées. Nous avons donc précocement opté pour le recours à des contrôles externes également appelés « ARN spikes » ou tout simplement « **spikes** » (ou encore ERCs pour External RNA controls en anglais, Tong et al., 2006). Il s’agit d’ARNs, synthétiques ou naturels, provenant d’un organisme étranger à celui que nous étudions (la souris) et ne présentant aucune homologie avec les ARNm de l’organisme étudié. Ils sont ajoutés en quantité fixe aux différents échantillons d’ARN totaux avant leur marquage. Comme les ARN totaux, ils subissent les différentes étapes du protocole (marquage, hybridation, lavage, lecture et analyse d’image) et la présence sur les puces de spots portant des sondes correspondant à ces spikes permettent la quantification de leurs signaux respectifs. Ayant subi les mêmes manipulations que les ARN totaux, les spikes capturent les éventuels biais introduits lors des différentes étapes du protocole. Des signaux faibles pour les spikes, signe par exemple d’un marquage moins efficace, se traduira par des

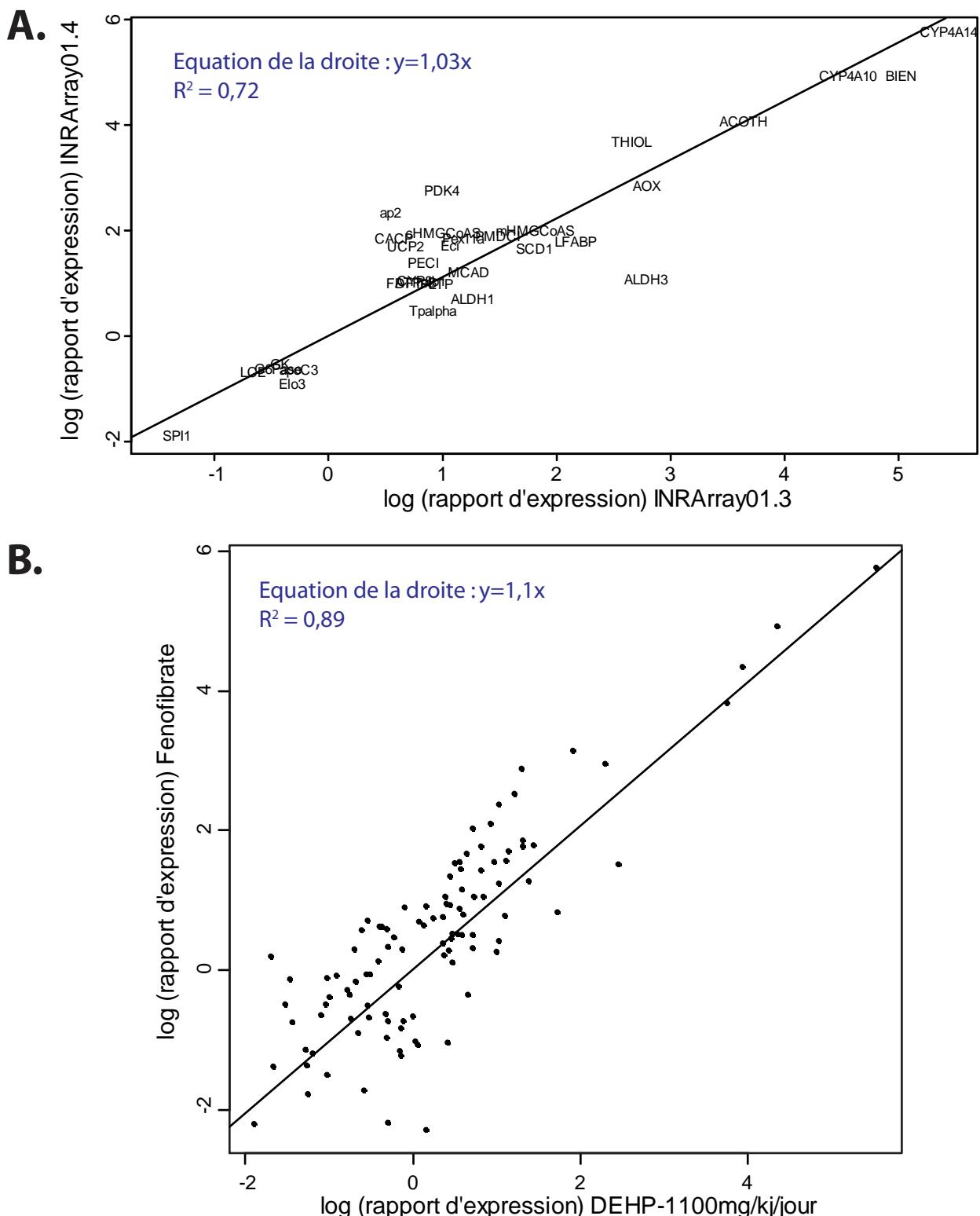


Figure 52 : Comparaisons de résultats issus d'hybridations distinctes sur les puces INRArray 01

Le graphique A compare les rapports d'expression moyens (individus traités/individus contrôles en logarithme base 2) obtenus par hybridations de deux générations différentes de puces (versions 03 et 04) avec des échantillons d'ARN totaux hépatiques issus de deux expériences indépendantes de traitement de souris C57BL/6J par le fénofibrate (100 mg/Kg/j, 14 jours). Les gènes présentant un rapport d'expression supérieur ou égal à 1,3 (en valeur absolue) et présents sur les deux générations de puces sont représentés. Le graphique B compare les rapports d'expression moyens ( $\log_2(\text{traités}/\text{contrôles})$ ) provoqués par deux activateurs de PPAR $\alpha$  (Fénofibrate 100 mg/kg/j, 14 j. et DEHP 1100 mg/Kg/j, 14 j., expériences et hybridations indépendantes) obtenus par hybridation d'ARN totaux hépatiques de souris C57BL/6J traitées ou non sur les puces INRArray 01.4. Chaque point représente un gène dont le rapport d'expression est supérieur ou égal à 1,3 en valeur absolue.

signaux faibles pour les gènes murins également. Etant ajoutés en quantité fixe entre les différents échantillons, ils peuvent également être utilisés pour calibrer les données. Lors de nos expériences initiales (INRAArray 01.1) nous avons utilisé le gène de l'enzyme Rubisco du tournesol (clone généreusement offert par le Dr B. Ranty, INRA, Toulouse) comme spike, puis (INRAArray 01.2), nous avons eu recours à des spikes commerciaux (Lucidea Universal ScoreCard, 23 gènes artificiels de levure, Amersham Biosciences, France). Enfin, à partir de la génération INRAArray 01.3, nous avons développé notre propre panel de spike. Pour cela, nous avons utilisé des puces à ADN de levure portant environ 140 sondes spécifiques (généreusement fournies par Véronique Leberre, plateforme transcriptome du génopole de Toulouse) sur lesquelles nous avons hybride des ARN issus de divers tissus de souris (foie, rein, cœur, tissu adipeux blanc et muqueuse intestinale) afin d'identifier les sondes qui ne croisaient pas avec ces ARN de souris. Une série de 13 spikes a ainsi été sélectionnée et des alignements via l'outil BLAST ont confirmé l'absence d'homologie avec les transcrits murins connus. Les ADNc correspondant ont été clonés dans le vecteur pSP64(polyA) puis les ARN des différents spikes (contenant une queue polyA grâce à sa présence dans le vecteur utilisé) ont été produits par transcription in vitro et contrôlés sur cassette Bioanalyzer 2100 (Agilent Technologies, France). Des sondes correspondant à ces 13 spikes ont également été produites à partir d'ADNc de levure par amplification et clonage (vecteur pCR 2.1 TOPO, Invitrogen, France) de fragment spécifiques de chacun des 13 spikes. Ces sondes ont été déposées en duplicitat sur une puce à ADN qui a de nouveau été hybride avec des ARNs murins issus de divers tissus (pour contrôler l'absence de fixation des ARNs de souris sur les sondes correspondant aux spikes). La génération de puce INRAArray 01.3, qui porte les sondes correspondant aux 13 spikes a, de plus, été hybride avec des ARN totaux hépatiques de souris additionnés de chacun des 13 spikes (un seul spike ajouté par puce hybride) afin de contrôler l'absence d'hybridation croisée de chaque spike avec les sondes qui ne lui correspondent pas sur la puce INRAArray (i.e. les sondes murines et les sondes correspondant aux autres spikes).

La **validation définitive** de nos conditions d'utilisation des puces INRAarray a été réalisée selon deux approches complémentaires. La première consiste globalement en une validation « par la biologie » via l'étude, systématisée à chaque nouvelle génération de la puce INRAarray 01, d'échantillons d'ARN issus de foie de souris de type sauvage ou PPAR $\alpha$ -déficientes traitées ou non à du fénofibrate. Notre puce à ADN comportant de nombreuses cibles du récepteur PPAR $\alpha$ , induites ou réprimées par le fénofibrate uniquement chez les souris de type sauvage, l'étude des résultats issus de ces hybridations de contrôle nous permet

de confirmer notre capacité à détecter la totalité des régulations attendues et ce chez les souris de types sauvage uniquement. C'est donc l'utilisation d'un système biologique que nous connaissons bien qui nous permet de conforter nos conditions expérimentales, ce qui constitue une approche assez classique en biologie lorsque l'on aborde une nouvelle méthodologie (la capacité à reproduire des résultats bien connus et décrits). Un exemple de comparaison des rapports d'expression obtenus avec deux générations de puces INRArray différentes pour l'étude d'un même traitement de souris C57BL/6J par le fénofibrate est présenté en [Figure 52A](#) (dans ce cas précis, les deux générations de puces ont néanmoins été hybridées avec des échantillons d'ARNs issus de deux expériences indépendantes reproduisant cependant le même traitement par le fénofibrate, ce qui peut expliquer en partie la variabilité observée). Une comparaison des rapports d'expression provoqués par deux activateurs différents de PPAR $\alpha$  (le fénofibrate et le DEHP) et mesurés, cette fois, sur une même génération de puces INRArray 01 (version 04) est également présentée en [Figure 52B](#). Cette figure illustre le fait qu'une bonne part des régulations provoquées par ces deux activateurs sont similaires mais que certaines apparaissent plus spécifique d'un traitement que d'un autre. L'analyse complète de ces résultats sera présentée dans la thèse d'Alexandre Eveillard que je co-encadre avec Thierry Pineau. Ces résultats nous rassurent sur la capacité de notre outil à répéter des signatures d'activation du récepteur PPAR $\alpha$  au fur et à mesure des générations de puces, pour autant que les conditions expérimentales et le protocole d'hybridation soient correctement reproduits. La seconde approche a consisté à utiliser les spikes de levure que nous avons évoqués ci-dessus. Dans un premier temps, afin de valider nos conditions expérimentales, nous avons eu recours à un ensemble de spikes commerciaux (Lucidea ScoreCard, Amersham Biosciences, France) qui contenaient 15 ARN incorporés à des concentrations égales entre les échantillons et servant aux contrôles qualité et à la normalisation des données ainsi que 8 ARN incorporés à des concentrations variables entre les échantillons des différents groupes expérimentaux. Ces 8 spikes correspondent à des intensités de signal fortes (de 500 à 5000 pg de spike ajouté par échantillon d'ARN total) ou faibles (de 15 à 150 pg de spike ajouté par échantillon d'ARN total) et, dans chaque étude, sont incorporés dans les groupes d'échantillons selon des rapport de 3:1 (+3), 1:3 (-3), 10:1 (+10) ou 1:10 (-10) selon les spikes afin d'évaluer la capacité du protocole utilisé à détecter des différences de cet ordre de grandeur. La [Table 11](#) présente les résultats obtenus à partir du criblage de 3 organes (24 puces par organes, étude présentée dans Martin et al., 2005) réalisé sur la génération INRArray 01.2 sur les rapports observés entre les groupes d'échantillons. Ces résultats ([Table 11](#)), tous hautement significatifs, confirment la qualité et la répétabilité de notre protocole

d'hybridation des puces INRAArray 01, ainsi que leur capacité à détecter des différences de ces ordres de grandeur. Nous observons aussi que dans la quasi-totalité des cas, les rapports d'expression sont sous-estimés en comparaison des rapports attendus, nous rassurant ainsi sur le caractère conservatif des rapports d'expression que nous rapportons avec cet outil. De multiples étapes ont ainsi été réalisées pour améliorer et standardiser le protocole présenté en annexe 1 qui fournit, dans nos mains, des résultats que nous estimons de qualité.

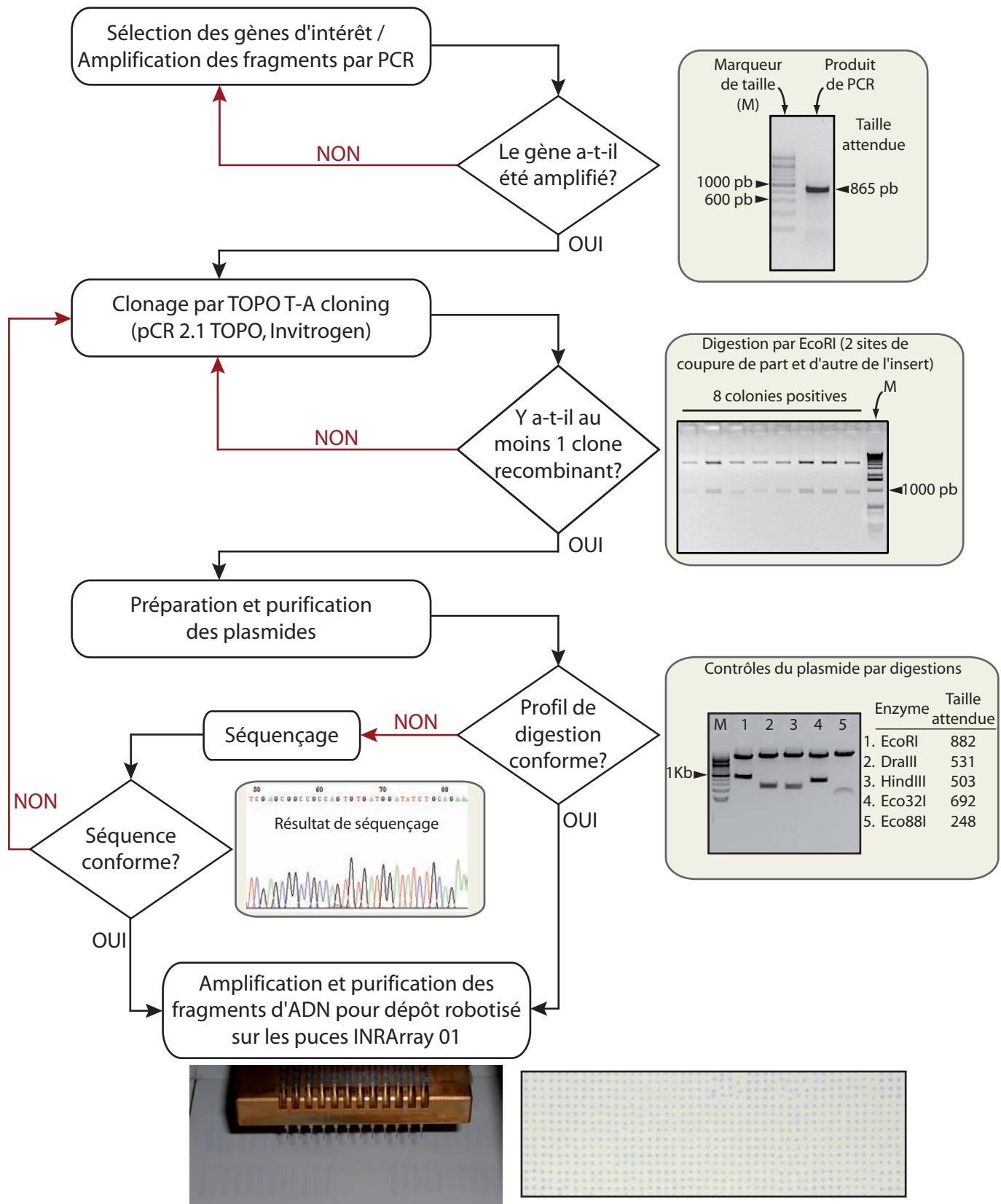
	Gamme d'intensité	Rapport d'expression attendu	Rapport d'expression observé (étude 1)	Rapport d'expression observé (étude 2)	Rapport d'expression observé (étude 3)	Rapport d'expression observé (3 études)
<b>Ratio1</b>	Basse	-3	-2,5 ± 0,4	-2,6 ± 0,2	-2,4 ± 0,4	-2,5 ± 0,3
<b>Ratio2</b>	Basse	3	3,0 ± 0,8	2,2 ± 0,3	2,6 ± 0,4	2,6 ± 0,6
<b>Ratio3</b>	Haute	-3	-3,0 ± 0,6	-2,8 ± 0,2	-2,8 ± 0,4	-2,9 ± 0,4
<b>Ratio4</b>	Haute	3	2,5 ± 0,3	3,0 ± 0,4	3,2 ± 0,4	2,9 ± 0,5
<b>Ratio5</b>	Basse	-10	-3,9 ± 0,8	-4,3 ± 0,4	-4,1 ± 0,5	-4,1 ± 0,6
<b>Ratio6</b>	Basse	10	3,7 ± 0,6	4,2 ± 0,5	5,0 ± 0,6	4,3 ± 0,8
<b>Ratio7</b>	Haute	-10	-5,0 ± 0,5	-9,6 ± 1,1	-7,1 ± 0,6	-7,2 ± 2,1
<b>Ratio8</b>	Haute	10	9,3 ± 1,4	7,8 ± 1,1	9,9 ± 1,2	9,0 ± 1,5

**Table 11 :** Rapports d'expression théoriques et observés de 8 spikes de levures.

Deux mélanges distincts de 8 ARN synthétiques de levure (Lucidea Universal ScoreCard, Amersham Biosciences) sont incorporés dans deux groupes d'échantillons (12 échantillons par groupe et deux groupes par étude). Ces deux mélanges se distinguent par des rapports d'expression connus des différents spikes (colonne « rapport d'expression attendu »). Les spikes sont arbitrairement regroupés en deux gammes d'intensité : basse (de 15 à 150 pg de spike ajouté par échantillon d'ARN) et haute (de 500 à 5000 pg de spike ajouté par échantillon d'ARN). Les rapports d'expression de ces 8 spikes observés dans 3 études (INRAArray 01.2) sont représentés dans ce tableau et sont tous hautement significatifs (test de Student et ajustement des p-values par la méthode de Bonferroni,  $p_{ajusté} < 0,05$ ). La dernière colonne prend en compte les données des 72 puces issues des 3 études. La quasi-totalité des rapports d'expression sont sous-estimés par l'INRAArray 01.2.

### III.3.3. Clonage des ADNc

Le clonage des ADNc déposés sur la puce INRAArray 01 a sans douté été le travail le plus long et le plus répétitif de ce doctorat. Il constitue cependant un élément essentiel de la qualité et de la fiabilité des résultats obtenus et de nombreuses personnes ont été impliquées dans la réalisation effective de ces clonages (citons en particulier Frédéric Lasserre et Arnaud Polizzi pour notre équipe ainsi que Romain Barnouin, Marie Detayrac et Alexandre Eveillard, stagiaires de M2R). J'ai, pour ma part, mis au point les protocoles standardisés utilisés pour la réalisation des clonages, assuré les mises au point nécessaires pour l'accroissement du débit de clonage, réalisé une large part des clonages eux-mêmes et surtout assuré la sélection des gènes et le choix de la plupart des amorces de PCR. La Figure 53 représente les principales



**Figure 53 :** Principales étapes du clonage des ADNc déposés sur les puces INRAarray 01

A gauche, les principales étapes du clonage d'un ADNc sont représentées (rectangles). Les principales questions auquel le manipulateur est confronté sont représentées sous la forme de losanges. Des exemples de résultats obtenus pour l'un de nos clones sont également présentés sous la forme d'insert à proximité de chaque losange.

étapes du clonage des ADNc déposés sur la puce à ADN INRAArray 01 ainsi que quelques images illustrant les principales étapes d'un tel clonage. Nous avons déjà évoqué les critères de choix qui ont guidé notre sélection des gènes présents sur la puce INRAArray 01 (voir paragraphe III.3.1.). La sélection d'une sonde appropriée au sein d'un transcrit sélectionné a été réalisée via l'utilisation des bases de données et d'outils de bioinformatiques standards selon les étapes suivantes :

- Une séquence murine du gène sélectionné est recherchée par mots clefs dans la base de données GenBank ([www.ncbi.nlm.nih.gov/sites/entrez?db=Nucleotide](http://www.ncbi.nlm.nih.gov/sites/entrez?db=Nucleotide)).
- Si un lien est disponible vers la base de données UniGene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene>), le lien est exploré pour rechercher une éventuelle séquence murine de référence (séquences dont l'identifiant commence par NM\_). Dans certains cas, des alignements ont également été réalisés avec les séquences disponibles dans la base de données Ensembl ([www.ensembl.org](http://www.ensembl.org)) pour vérifier leur identité et la présence éventuellement rapportée de transcrits alternatifs.
- Des alignements de la séquence obtenue sont effectués à l'aide de l'outil BLAST (blastn sur [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/), alignements réalisés systématiquement sur les bases de données RefSeq et nr), permettant ainsi d'identifier d'éventuelles zones d'homologies avec d'autres transcrits ou bien la présence non encore détectée de transcrits alternatifs. Les zones d'homologies avec d'autres transcrits seront évitées dans le choix de la sonde. S'il existe des transcrits alternatifs, distingués par des régions suffisamment longues (de l'ordre de 300 à 400 pb minimum) qui ne présentent pas d'homologie avec d'autres transcrits, alors ces régions sont sélectionnées pour dessiner des sondes spécifiques de chaque transcript. Si les régions distinguant les transcrits alternatifs ne sont pas assez longues ou homologues à d'autres transcrits, alors une sonde est sélectionnée qui reconnaît spécifiquement les différents transcrits alternatifs. Une telle sonde mesurera effectivement l'expression du gène considéré (qui reflètera peut-être son activité transcriptionnelle) mais indifféremment en fonction des transcrits produits par ce gène.
- Un couple d'amorce servant à amplifier un fragment spécifique du transcrit est choisi à l'aide de l'application Primer 3 disponible sur internet ([frodo.wi.mit.edu/cgi-bin/primer3/primer3-www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3-www.cgi)), en utilisant généralement les paramètres suivants : Mispriming library=Rodent\_and\_simple ; product size ranges= 800-1200 lorsque c'est

possible et toujours >150 et <1500 sinon ; Max poly-X=3 ainsi que les valeurs par défaut des autres paramètres.

- Les amores et surtout la sonde encadrée par ces amores subissent à nouveau un alignement à l'aide de l'outil BLAST pour s'assurer de manière théorique de l'absence d'hybridation croisée.
- Un profil de restriction de la sonde est ensuite réalisé à l'aide de l'application Webcutter 2 ([www.firstmarket.com/cutter/cut2.html](http://www.firstmarket.com/cutter/cut2.html)) qui servira à valider la sonde une fois produite.

Ensuite, la sonde nucléotidique est amplifiée par PCR à partir d'ADNc murin (des essais sur différents organes et/ou conditions expérimentales d'exposition des animaux ont parfois été nécessaires pour réussir à amplifier certains fragments) en utilisant les amores sélectionnées. La séquence amplifiée est ensuite clonée dans le vecteur pCR 2.1 TOPO (Invitrogen, France) selon les recommandations du fournisseur. Pour les gènes dont la séquence est disponible, des profils de restriction avec au moins 3 enzymes différentes sont réalisés pour valider la séquence clonée. Une proportion appréciable de nos clones ont également fait l'objet d'un séquençage complet et beaucoup des clones les plus anciens ont été utilisés en northern blot, confirmant ainsi leur spécificité. Les clones sont stockés à -80°C dans deux congélateurs séparés, sous la forme de colonies bactériennes et sous la forme de plasmides préparés en quantité importante par purification sur colonne (Plasmid Maxi kit, Qiagen, France).

Des amores de PCR ont été dessinées sur le vecteur pCR 2.1 TOPO et les ADN polymérase de plusieurs fournisseurs ont été évaluées pour pouvoir amplifier aisément, à haut rendement et avec une bonne fidélité (enzymes proofreading) l'ensemble des inserts contenus dans nos clones pCR 2.1 TOPO. Ces PCR sont effectuées en plaque 96 trous puis purifiées et concentrées via un système de colonnes au format 96 trous fonctionnant sous vide (QIAquick 96 PCR purification kit, Qiagen, France). Les produits de PCR obtenus sont dosés au spectrophotomètre et leur taille est contrôlée sur gel d'agarose. La moindre ambiguïté donne lieu à une nouvelle série de réamplification/purification/concentration/contrôle individuel du produit de PCR concerné. De l'ordre de 5% maximum de nos clones sont concernés par ce type de problème (fixation des amores pCR 2.1 TOPO à l'intérieur de la séquence de l'insert, contamination du plasmide, absence d'amplification, concentration insuffisante, etc.). Les produits de PCR sont ensuite dilué en 50% DMSO (additionné d'un peu de bleu de bromophénol pour pouvoir visualiser les spots sur les membranes) à une concentration finale minimale de 200 ng/µL puis déposés (VersArray ChipWriter, Virtek/Biorad, France) sur

membranes de nylon chargé (Millipore, France) et enfin crosslinkés sous UV. Les puces INRArray 01 ainsi produites sont conservées au sec à +4°C avant de subir le protocole présenté en annexe 1.

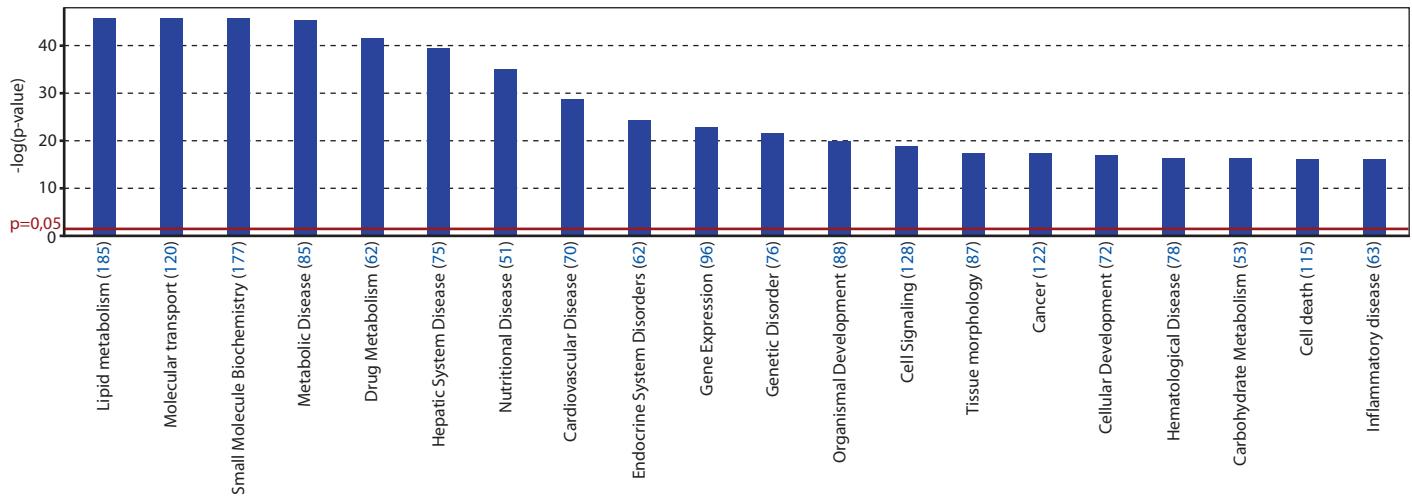
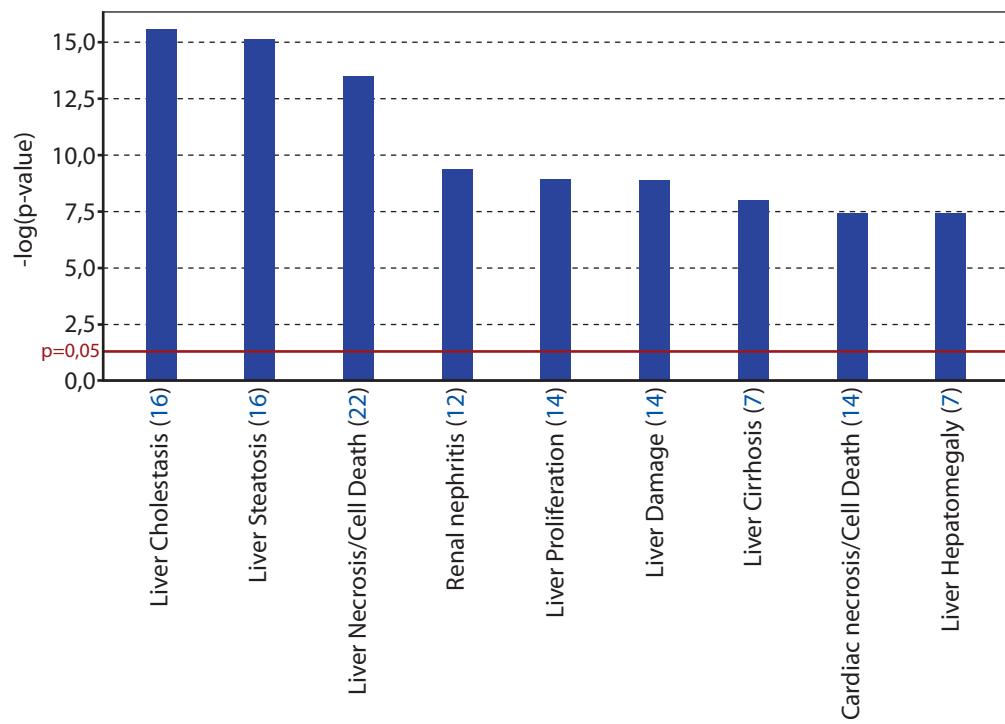
Le protocole de clonage des ADNc a, de plus, fait l'objet d'adaptations permettant d'accroître notre débit de clonage. Les PCR d'amplification des fragments d'ADNc ont ainsi été réalisées par groupe de 15-20 PCR et la préparation des plasmides issus des colonies choisies suite au clonage du fragment dans le vecteur pCR 2.1 TOPO a été réalisée sur colonnes au format 96 puits utilisant un système d'aspiration par le vide (DirectPrep 96 Miniprep kit, Qiagen, France). La plupart des informations présentées dans ce paragraphe et le précédent, ainsi que des informations complémentaires et un protocole illustré sont présentés sur le site web du Laboratoire de Pharmacologie et Toxicologie (UR66) de l'INRA ([www.inra.fr/internet/Centres/toulouse/pharmacologie/pharmaco-moleculaire/technologie/puceADN.html](http://www.inra.fr/internet/Centres/toulouse/pharmacologie/pharmaco-moleculaire/technologie/puceADN.html)).

Enfin, signalons que le panel d'ADNc déposé sur la puce INRArray 01 a fait l'objet d'un dépôt auprès de l'Agence de Protection des Programmes (APP) sous la forme d'une liste de gènes et des amorces ayant servies à amplifier les fragments spécifiques de chaque gène (Référence du dépôt initial de 126 ADNc en 2002 : IDDN.FR.290019.000.D.P.2002.000.40000, mise à jour en 2005 avec 320 ADNc et prévue pour fin 2007 avec environ 450 ADNc).

### III.3.4. Générations de la puce INRArray

Cinq générations de puces INRArray 01 ont été produites jusqu'à présent, au fur et à mesure de l'élargissement du panel d'ADNc constituant la puce à ADN. Ces générations de puces ont été numérotées de la manière suivante : INRArray 01.X où X est le numéro de la version. Un descriptif rapide des différentes versions de l'INRArray 01, de leur composition en ADNc et de l'utilisation qui en a été faite est présenté sur le site internet de l'UR66-INRA ([www.inra.fr/internet/Centres/toulouse/pharmacologie/pharmaco-moleculaire/technologie/puceADN.html](http://www.inra.fr/internet/Centres/toulouse/pharmacologie/pharmaco-moleculaire/technologie/puceADN.html)).

Avant la version **INRArray 01.1**, plusieurs puces à ADN ont été fabriquées en nombre limité, mais elles ont uniquement été utilisées pour la réalisation de tests, de mise au point et de validation du protocole expérimental (voir paragraphes précédents). La version INRArray 01.1 est donc la première version sur laquelle nous avons évalué des échantillons biologiques d'intérêt, et en l'occurrence des ARN totaux issus de foie de souris de type sauvage ou PPAR $\alpha$ -/- soumises ou non à un traitement par le fénofibrate. Un exemple de résultat obtenu est présenté sur le site internet de l'unité. Ces premiers résultats nous ont permis de confirmer

**A.****B.**

**Figure 54 : Principales fonctions biologiques explorées par l'INRAarray 01.4**

La liste des ADNC présents sur l'INRAarray 01.4 (annexe 2) a été chargée dans l'application Ingenuity Pathway Analysis 5.0 (IPA). Une recherche des fonctions biologiques significativement représentées dans la liste des transcrits explorés par l'INRAarray 01.4 est effectuée par cette application en comparant par un test exact de Fisher unilatéral le nombre de gènes représentant une fonction biologique donnée sur la puce par rapport au nombre de gènes représentant cette fonction dans la base de données de connaissances d'IPA (IPA knowledge base). Seules les fonctions pour lesquelles le nombre de gènes explorés par la puce sont sur-représentés par rapport à un tirage aléatoire de gènes sont présentées (test unilatéral). Les 20 fonctions biologiques les plus significativement explorées par notre puce sont représentées en A et ont été sélectionnées car leurs p-values sont toutes hautement significatives et que chacune de ces fonctions est représentée par au moins 50 gènes sur la puce (le nombre de gènes explorant chacune des fonction est précisé en bleu entre parenthèse après la fonction). Les fonctions biologiques liées au domaine de la toxicologie et significativement explorées par la puce INRAarray 01.4 sont représentées en B.

pour la première fois la capacité de notre outil à détecter des régulations géniques provoquées par l'activation pharmacologique de PPAR $\alpha$  bien connues et décrites dans la littérature.

Cent puces **INRArray 01.2** ont été produites et utilisées pour les études publiées dans le cadre de ce doctorat et présentées aux chapitres IV et V de ce manuscrit. La liste complète des 120 ADNc déposés sur cette version est présentée sur le site internet de notre unité (voir ci-dessus).

La version **INRArray 01.3** a été produite à plus grande échelle (500 puces spottées) et a été enrichie de nombreux gènes spécifiques du tissu adipeux blanc via une collaboration avec l'UMR UPS-CNRS 5018 dirigée par Luc Penicaud. Elle permet de mesurer l'expression de 192 gènes et c'est pour cette version qu'ont été développés les 13 ARNs spikes de levure dont l'un est utilisé comme contrôle négatif (l'ARN correspondant n'est pas incorporé dans les échantillons d'ARN totaux avant marquage). Cette version a été utilisée dans le cadre de plusieurs collaborations ainsi que pour l'étude de jeûne présenté au chapitre VI de ce manuscrit.

La version **INRArray 01.4** permet de mesurer l'abondance de 291 transcrits différents (liste des ADNc présentées en [annexe 2](#), l'INRArray 01.4 porte deux sondes différentes pour les gènes PEPCK, PPAR $\alpha$  et Cide-a et Cide-c). Cette version a été enrichie de nombreuses nouvelles sondes grâce notamment à l'important travail de clonage réalisé par Marie De Tayrac et Alexandre Eveillard, alors en stage de M2R au sein de notre équipe (co-encadrement T. Pineau - P.Martin). Les thématiques de notre équipe s'élargissant actuellement à l'étude de disrupteurs métaboliques et endocriniens (notamment les phtalates et le bisphénol A), cette version s'est enrichie de gènes impliqués dans la stéroidogénèse, les phénomènes de division cellulaire (cyclines notamment) et d'apoptose (caspases par exemple) et de gènes cibles des voies de signalisation des estrogènes et des androgènes. Cette version a d'ores et déjà été utilisée pour étudier l'effet de l'administration de différentes doses de DEHP (le Di(2-ethylhexyl)phtalate) chez la souris C57BL/6J dans trois organes différents (le foie, le rein et le testicule, doctorat d'Alexandre Eveillard). La liste complète des ADNc présents sur cette version INRArray 01.4 est présentée en [annexe 2](#) de ce manuscrit. Il n'est pas aisément de regrouper de manière systématique ces différents gènes en termes de fonctions biologiques car certains interviennent dans plusieurs fonctions et d'autres ont des fonctions encore mal décrites. Pour essayer d'illustrer malgré tout les principales fonctions biologiques qui peuvent être explorées à l'aide de l'INRArray 01.4, nous avons conduit une analyse sous l'application Ingenuity Pathway Analysis 5.0 (IPA). Dans cette application, les annotations fonctionnelles des gènes de la puce INRArray 01.4 sont prises en compte et comparées à

celles de tous les gènes contenus dans la base de données IPA knowledge base (IKB) via un test unilatéral exact de Fisher permettant ainsi de déterminer quelles sont les grandes fonctions biologiques significativement explorées par la puce INRArray 01.4. Les principales fonctions biologiques et celles liées au domaine de la toxicologie sont illustrées en [Figure 54](#). Ces différentes fonctions apparaissent parfaitement cohérentes avec l'historique de développement de la puce INRArray 01. Initialement dédiée au récepteur PPAR $\alpha$ , notre puce explore les fonctions biologiques majeures dans lesquelles ce récepteur intervient (métabolisme des lipides, des glucides et des petites molécules en général, métabolisme des xénobiotiques et systèmes de transport, désordres métaboliques, nutritionnels, hépatiques et cardiaques, inflammation). Par rapport aux données présentées dans le chapitre II, cette analyse suggère que l'étude des fonctions « inflammation » et « métabolisme des acides aminés » devrait néanmoins être renforcée sur notre outil, ce qui a été en partie réalisé avec la génération suivante (puces INRArray 01.5, non présentées dans ce manuscrit). Le renforcement de l'étude des fonctions endocriniennes et de prolifération/différenciation/apoptose est également révélé par cette analyse (désordres du système endocrinien, morphologie des tissus, cancer, développement cellulaire, mort cellulaire). La présence de nombreux RNs, de facteurs de transcription avec lesquels ils interagissent et de certains cofacteurs est principalement illustrée par les fonctions « expression des gènes » et « signalisation cellulaire ». Deux fonctions plus inattendues sont également observées : « désordres génétiques » et « maladies hématologiques ». La présence de la première fonction semble liée au fait que des mutations de nombreux gènes présents sur notre puce ont été décrites dans la littérature et en particulier de nombreux ABC transporteurs, cytochromes P450, récepteurs nucléaires et gènes du métabolisme des lipoprotéines (apolipoprotéines, récepteur aux LDL notamment), illustrant l'importance de ces gènes dans des fonctions biologiques importantes en thérapeutique (métabolisme des médicaments en particulier) ou dans des pathologies communes des pays industrialisés (désordres du métabolisme des lipides). La présence de la seconde fonction semble essentiellement refléter notre effort d'enrichissement de notre puce dans le domaine de la division cellulaire et de l'apoptose et de la mort cellulaire en général car les sous-fonctions les plus nettement représentées dans notre panel touchent essentiellement ces aspects de la biologie des cellules sanguines (données non présentées ici). En ce qui concerne le domaine de la toxicologie, les résultats fournis par IPA semblent également largement cohérents avec notre intérêt premier pour la biologie du foie ([Figure 54B](#)). Le rein et le cœur constituant également des cibles que nous explorons dans certaines études (Martin et al., 2005), il est intéressant de les retrouver

ici. Malgré la présence d'un nombre substantiel de sondes ciblant des gènes impliqués dans la biologie du tissu adipeux, cet organe ne constitue pas une cible toxicologique majeure (si ce n'est pour le stockage et donc la biodisponibilité de xénobiotiques) et il est donc normal qu'il n'apparaisse pas dans cette analyse. En revanche, l'orientation de certaines de nos études sur la toxicité de molécules affectant les organes de la reproduction nécessitera manifestement un enrichissement substantiel de notre puce par des sondes plus pertinentes dans ce domaine.

La version **INRAarray 01.5**, récemment produite, s'est enrichie de nouveaux gènes issus notamment de criblages haute-densité réalisés sur des ARN totaux hépatiques de souris C57BL/6J traitées par diverses molécules (mycotoxines, pesticides, médicaments, phtalates) ainsi que d'autres cibles des phtalates récemment décrites dans la littérature. Cette version sera présentée plus en détail dans le manuscrit de doctorat d'Alexandre Eveillard qui a assuré l'essentiel des étapes de sa production et de son utilisation.

En conclusion, le développement de la puce à ADN INRAarray 01 a constitué une étape importante de ce travail de doctorat. Il aura permis à l'équipe de se confronter à de multiples aspects de mise au point technique et de lever, les unes après les autres, une série de barrières conduisant à la fabrication d'un outil fiable et versatile. Il aura également permis à l'équipe de prendre réellement le virage de la biologie à haut-débit car même si nos débits restent modestes en termes de nombre de gènes mesurés, les effectifs de certaines de nos expériences ne sont, encore aujourd'hui, que rarement rencontrés dans les expériences de puces à ADN pangénomiques. S'il n'est pas sûr aujourd'hui que l'avenir de cette puce soit de continuer longtemps son développement en termes de nombre d'ADNc spottés en raison notamment de l'apparition de nouveaux modes de fabrication et de la chute progressive des prix des outils commerciaux, le développement de cet outil aura néanmoins été pour nous d'une part une source d'information scientifique extrêmement importante comme en témoigne les premières valorisation présentées dans le cadre de ce doctorat et d'autre part une source d'expérience riche dans le domaine des puces à ADN et de leur analyse qui se trouve d'ores et déjà valorisée dans le cadre des formations et des collaborations auxquelles nous participons où dont nous sommes à l'initiative. Le paragraphe suivant présente brièvement l'analyse des données issues des puces à ADN, un aspect que nous n'avons qu'évoqué jusqu'à présent mais qui se trouve être sans doute le point critique de l'utilisation des puces à ADN.

### **III.4. ANALYSE STATISTIQUE DES DONNEES ISSUES DES PUCES A ADN**

La génération d'une masse considérable de données par les techniques d'analyse du transcriptome mais également par les autres techniques de la biologie à haut débit conduit aujourd'hui les biologistes à une interaction renforcée avec le domaine de la statistique et de la bioinformatique. L'arrivée des microarrays sur le marché est un excellent exemple qui a fortement stimulé ces différentes communautés scientifiques à interagir, conduisant à l'apparition de multiples outils, bases de données et consortium dédiés aux différents aspects de la génération, de la manipulation, de l'analyse et de l'interprétation des données de transcriptomique. Nous présenterons dans ce chapitre les différentes étapes du traitement des données issues des puces à ADN. Il n'est bien sûr pas possible dans le cadre de ce manuscrit de décrire avec précision l'ensemble des méthodes qui seront évoquées mais elles seront, lorsque mes compétences le permettront, expliquées brièvement en limitant autant que possible le recours aux formules compliquées et en privilégiant des exemples. Après avoir considéré les différents aspects de prétraitement (III.4.1) et de normalisation (III.4.2) des données qui conduisent à générer un jeu de données dont les caractéristiques le rendent exploitable pour répondre aux questions posées par le biologiste, nous présenterons les principaux aspects de l'analyse des données visant précisément à répondre à ces questions. Classiquement, on regroupe en trois catégories, présentées successivement dans les paragraphes III.4.3, III.4.4. et III.4.5, les questions auxquelles les biologistes souhaitent répondre lorsqu'ils réalisent une expérience de puce à ADN. De plus, d'autres types de questionnements ainsi que des outils aidant à une interprétation plus poussée des données sont présentés dans le paragraphe III.4.6. Enfin, les choix réalisés pour le prétraitement des puces INRAarray 01 sont rapidement présentés dans le paragraphe III.4.7. Signalons enfin que la plupart des concepts et outils évoqués dans ces paragraphes sont applicables à bien d'autres domaines de la biologie comme par exemple la protéomique ou la métabolomique.

#### **III.4.1. Qualité des données, filtres et prétraitements**

Lorsque l'on aborde le traitement d'un jeu de données, il est toujours bon de commencer par se remémorer d'où il provient et quelles sont ses principales caractéristiques. En ce qui concerne les puces à ADN, l'une des caractéristiques majeures des données générées est leur **dimension**. Il n'est pas rare que des puces comportant des milliers ou des dizaines de milliers de sondes soient utilisées dans les expériences mais leur coût limite naturellement le nombre d'échantillons analysés, conduisant à la génération de jeux de données contenant des milliers ou des dizaines de milliers de variables mesurées sur quelques

individus uniquement. Cette asymétrie des jeux de données n'est pas triviale pour le statisticien, généralement plus habitué à la manipulation de données contenant quelques variables mesurées sur un grand nombre d'individus. Il n'est donc pas rare que les méthodes statistiques soient conçues pour fonctionner sur des tableaux de données individus  $\times$  variables où les individus sont en lignes et les variables en colonnes, alors que la présentation classique des données de puce à ADN est faite dans le sens inverse. De plus, cette asymétrie génère également parfois un flou sur le rôle précis joué par les individus et par les variables et l'on peut parfois se retrouver dans des situations où ces rôles sont inversés. Tout cela reste néanmoins généralement parfaitement clair lorsque la ou les questions de départ ont été correctement posées et le plan d'expérience conçu en conséquence.

Une autre caractéristique de ces données est liée au fait qu'elles soient relativement **bruitées**. Ces bruits sont d'origine aléatoire comme par exemple la variabilité liée à la technique ou bien la variabilité biologique (que l'on peut d'ailleurs dans de nombreux cas souhaiter estimer) ou bien constituent des sources de variations systématiques comme par exemple le biais introduit par l'utilisation de deux fluorophores différents (Cy3 et Cy5) que nous avons évoquée précédemment. Afin de gérer les sources de variations aléatoires l'utilisateur est tenu de réaliser des réplicats, biologiques (individus ou cultures cellulaires placés dans les mêmes conditions expérimentales) et/ou techniques (les réplicats techniques peuvent être réalisés à de multiples niveaux : plusieurs extractions de l'ARN, plusieurs marquages d'un même ARN, plusieurs hybridations sur différentes puces d'un même produit de marquage par exemple), les premiers étant généralement plus souhaitables lorsque les deux ne sont pas possibles, car ils génèrent le plus souvent un bruit plus important, évitant ainsi nombre de faux positifs mais surtout parce que l'estimation de la variabilité biologique a généralement plus d'intérêt pour le biologiste que l'estimation de la variabilité technique. Nous ne discuterons pas ici des détails liés aux plans d'expérience et aux plans d'hybridation qui déterminent en particulier la puissance statistique des analyses, mais signalons que ce dernier aspect est particulièrement critique pour les puces en double couleur et qu'une riche littérature sur le sujet peut fournir des orientations assez claires à l'utilisateur (pour des synthèses, voir Simon, 2003, pp 11-27; Speed, 2003, pp 35-91; ou encore Bolstad et al., 2004; Churchill, 2002; Dobbin et al., 2003; Kerr and Churchill, 2001a; Kerr and Churchill, 2001b). Ce sont par ailleurs les étapes de normalisation qui doivent corriger les biais systématiques et qui seront abordées au paragraphe suivant. En dehors de ces sources de variations, d'autres artefacts peuvent également ajouter du bruit dans les données comme une tâche ou une poussière sur une puce, un bruit de fond non homogène ou un défaut de cerclage des spots par

exemple. Le seul moyen de résoudre la plupart de ces problèmes est généralement d'apporter un soin particulier à l'étape d'analyse d'image (Simon, 2003, pp 29-38) et de repérer (les utilisateurs utilisent souvent le mot anglais francisé « flagger ») les spots ou régions de la puce problématiques afin de les filtrer lors des étapes préliminaires de l'analyse. La réalisation de contrôles qualités (basés par exemple sur des spikes) et la recherche systématique de valeurs aberrantes doit ainsi impérativement faire partie des étapes préliminaires de l'analyse d'un jeu de données de puce à ADN. Le cas du bruit de fond mérite d'être évoqué à part. Dans le cas d'un bruit de fond non homogène, la solution la plus courante consiste généralement à retrancher aux intensités de chaque spot une valeur de bruit de fond, généralement estimée de manière locale (souvent une valeur d'intensité moyenne ou médiane des pixels situés à la périphérie du spot). Il est cependant important de tenir compte du fait que l'incertitude associée à la mesure du bruit de fond s'additionne à l'incertitude associée à la mesure de l'intensité du spot et qu'une telle correction peut donc ajouter du bruit dans les données. Il paraît donc relativement peu approprié d'appliquer une telle correction dans le cas où le bruit de fond est homogène et à peu près équivalent sur l'ensemble des puces utilisées dans l'expérience car cela ne corrigerait aucun biais tout en rajoutant du bruit dans les données. Par ailleurs, signalons qu'il n'est absolument pas certain que la plupart des valeurs utilisées pour estimer le « bruit de fond » représentent effectivement des valeurs d'intensité aspécifiques prises par les pixels situés à l'intérieur du spot. En effet, pour estimer réellement un bruit de fond, il faudrait idéalement disposer, à la proximité immédiate de chaque spot spécifique, de zones de la puces contenant elles aussi de l'ADN, dont la composition en base et dont les caractéristiques d'hybridation aspécifiques seraient similaires aux sondes présentes dans le spot spécifique mais dont la séquence ne permettrait pas la fixation du transcrit spécifique. Enfin, l'application d'une correction de bruit de fond ne manquera pas de générer des données négatives (qui ne peuvent donc être transformées en logarithme) ou infinitésimales (qui généreront des rapports d'expression aberrants), notamment lorsque le bruit de fond est estimé sur un nombre limité de pixels situés très à proximité des spots et il sera alors essentiel d'identifier et de traiter ces données à part. Loin d'être opposé de manière systématique à l'utilisation de corrections de bruit de fond, il me semble néanmoins important que l'utilisation de cet outil soit faite en connaissance de cause dans la pratique (pour une discussion sur l'utilité des corrections de bruit de fond, voir Parmigiani, 2003, pp 13-14; pour la présentation de différentes méthodes voir Simon, 2003, pp 33-34).

Nous avons évoqué précédemment la nécessité d'appliquer des filtres sur les données obtenues (ou au niveau de l'analyse d'image) pour éliminer les mesures de qualité douteuse, les artéfacts, les spots aberrants, etc. Une autre justification essentielle de l'application de filtres aux puces pangénomiques provient cependant de la biologie elle-même. Si ces puces revendiquent leur capacité à couvrir la quasi-totalité du transcriptome, le biologiste étudiant un type cellulaire ou un organe donné ne s'attend a priori pas à ce que la totalité des gènes soient exprimés dans son échantillon. Il est bien sûr difficile de fournir une échelle fiable du nombre de gènes exprimés dans un tissu donné (le chiffre d'environ 10000 gènes exprimés par type cellulaire est cependant parfois proposé) mais certaines études peuvent néanmoins fournir des indications sur ce sujet. C'est en particulier le cas pour la souris, notre modèle d'intérêt, qui a fait l'objet d'études du transcriptome effectuées sur des variétés de tissus (Bono et al., 2003; Su et al., 2002; Zhang et al., 2004). Zhang et al. (Zhang et al., 2004) détectent ainsi environ 22000 transcrits sur environ 40000 mesurés qui sont considérés comme présents dans au moins un des 55 tissus qu'ils ont étudiés, le nombre de gènes détectés par tissus s'étendant d'environ 5500 à 10000. Ces études incitent le praticien de l'analyse des données de puces à ADN à appliquer des filtres sur l'intensité des signaux des spots (par exemple en comparant ces intensités à celles du bruit de fond, à des témoins négatifs, etc.) pour ne retenir que les gènes effectivement exprimés dans le tissu étudié. Certains gènes susceptibles de n'être exprimés que dans une condition expérimentale donnée, doivent généralement être détectés par le biais d'une analyse qui leur est dédiée.

L'application de ces différents filtres lors du « prétraitement » des données de puces à ADN conduit naturellement à la génération de valeurs manquantes que les méthodes d'analyses développées spécifiquement pour les données de transcriptomique doivent donc gérer.

Enfin, la plupart du temps, les données de puces à ADN nécessitent une **transformation** dont les buts sont souvent multiples (correction d'une asymétrie de distribution des données, homogénéisation des variances par exemple). La transformation logarithme est sans doute la plus utilisée et l'utilisation du logarithme en base 2, notamment pour rapporter des différentiels d'expression (les « log(ratios) » dont on entend souvent parler), est sans doute la pratique la plus répandue. Elle permet généralement de rendre la distribution des données plus symétriques, d'homogénéiser, au moins en partie, les variances de gènes différents ou de conditions expérimentales distinctes, et de fournir des rapports d'expression sur une échelle continue, symétrique et centrée sur 0. Elle présente la propriété de transformer des effets multiplicatifs en effets additifs ( $\log(axb)=\log(a)+\log(b)$ ). Le

principal problème rencontré dans l'utilisation d'une transformation logarithme est sans doute lié à la définition de cette fonction sur  $R^+$  (ensemble des réels strictement positifs), la rendant, de fait, inapplicable sur des valeurs négatives ou nulles et conduisant certains auteurs à lui préférer des transformations définies sur  $R$  (comme par exemple la fonction racine cubique ou encore arcsinh comme dans Zhang et al., 2004). Enfin, il existe des transformations plus « drastiques » comme la transformation en rang qui conduit l'utilisateur à perdre en partie le caractère quantitatif initial de ses données mais lui apporte dans de nombreux cas, une robustesse qui peut être appréciable, surtout dans le cas de données très bruitées.

Ainsi, nous avons vu dans ce paragraphe que les étapes d'acquisition et de prétraitement des données sont indispensables à l'acquisition de jeux de données fiables. Nous allons aborder à présent le rôle plus particulier de la normalisation des données.

### **III.4.2. Normalisation des données**

Comme nous l'avons précisé, l'objectif de l'étape de normalisation des données vise à éliminer la présence d'éventuels biais systématiques qui peuvent provenir de différentes origines (marquage plus intense d'un échantillon par rapport à un autre ou biais dans l'incorporation des fluorophores par exemple). L'objectif premier est donc de rendre les données issues de plusieurs échantillons et de plusieurs puces comparables entre eux. Pour cela, l'idée la plus simple est sans doute de rapporter toutes les données à une quantité dont on sait a priori qu'elle est constante entre les échantillons. Si on note  $K$  cette quantité (appelé le coefficient de normalisation) et  $X$  l'intensité de chaque gène (éventuellement corrigé par rapport au bruit de fond), le calcul est donc  $X/K$  ou, en logarithme,  $\log(X)-\log(K)$ . De plus, si cette quantité  $K$  provient d'un grand nombre de mesures indépendantes (normalisation basée sur des groupes de gènes plutôt que sur un seul gène), la variabilité de l'estimation de cette quantité sera faible (la moyenne de  $n$  mesures suivant une loi de variance  $\sigma^2$  suit une loi normale de variance  $\sigma^2/n$ ) et la normalisation améliorera ainsi la fiabilité des données (correction des biais) sans en faire exploser la variabilité. Une pratique courante, issue des techniques à « bas débit », consiste à utiliser des « **gènes de ménage** » (housekeeping genes en anglais), dont l'expression est généralement observée dans de nombreux types cellulaires (voire idéalement dans tous les types cellulaires) à des niveaux qui sont (idéalement) constants car ces gènes remplissent des fonctions indispensables au niveau cellulaire (facteurs basaux de la transcription ou de la traduction, ARN ribosomiques, gènes de la glycolyse ou du cytosquelette par exemple). L'avantage de l'utilisation de ces gènes de ménage pour la

normalisation est qu'ils présenteront bien, en théorie, le même biais systématique que les autres gènes mesurés. L'inconvénient majeur est qu'il n'existe aucun gène dont le niveau d'expression a effectivement été démontré comme étant parfaitement constant dans tous les types cellulaires et sous toutes les conditions expérimentales. Les gènes de ménages les plus couramment utilisés dans les expériences à bas-débit (la GAPDH, la  $\beta$ -actine ou l'ARN ribosomique 18S par exemple) s'avèrent tous à un moment ou à un autre être modulés sous des conditions expérimentales spécifiques. Il devient donc difficile pour l'utilisateur de définir avec certitude une liste suffisamment longue (pour réduire la variabilité du coefficient de normalisation) de gènes de ménage à utiliser pour normaliser ses données. De fait, cette méthode est peu ou pas utilisée en pratique dans le contexte des puces à ADN.

Une solution alternative consiste à sélectionner *a posteriori*, à partir des données, une liste de gènes qui semble répondre aux critères de définition d'un gène de ménage (stabilité d'expression entre les différents échantillons). Cela peut être par exemple effectué via une transformation des données en rang et une sélection des **gènes** présentant les rangs **les plus stables** parmi les échantillons analysés, à condition de faire attention à ne pas sélectionner ainsi des gènes présents dans des gammes d'intensité où peu de gènes apparaissent (car alors ils auront plus de chance d'avoir un rang stable). Cette technique présente l'avantage de fournir, avec des puces pangénomiques, un nombre conséquent de gènes pour assurer la normalisation, mais elle doit en théorie être répétée à chaque expérience, fournissant à chaque fois des listes de gènes différentes, conduisant à une normalisation inhomogène des données entre plusieurs expériences. De plus, cette méthode n'est généralement pas applicable aux puces à ADN dédiées qui ne portent souvent pas suffisamment de gènes pour que cette méthode fonctionne correctement. Quelle que soit la méthode choisie (gène de ménage ou majorité stable), l'hypothèse sous-jacente à ces normalisations est que ces gènes présentent une expression constante entre les échantillons.

Nous avons déjà évoqué la possibilité d'utiliser des **contrôles externes** (les spikes) au lieu de ces contrôles internes, l'avantage de leur utilisation étant qu'ils présentent généralement de faibles variations et sont utilisables pour des puces à ADN dédiées. Les hypothèses sous-jacentes à cette méthode de normalisation sont que 1) les ARNs spikes ont bien été ajoutés dans les échantillons en quantités égales et 2) les échantillons d'ARN sont tous bien à la même concentration et au même niveau de qualité (ou de dégradation...), i.e. que les quantités d'ARN messagers intacts sont bien égales entre les différents échantillons. Alors que la première hypothèse ne pose pas de problème particulier (aux erreurs et incertitudes de pipetage près), il est plus ardu de s'assurer de la seconde hypothèse. Un dosage

précis (répliquats) des ARNs et une évaluation rigoureuse de leur niveau de dégradation éventuelle (par exemple à l'aide d'un Bioanalyzer 2100, Agilent Technologies, France ou d'un appareil équivalent) est donc indispensable pour utiliser cette normalisation particulièrement sensible aux erreurs de dosage des ARN ou à leur dégradation. Un autre désavantage de cette méthode est qu'elle nécessite la mise au point et la validation de ces spikes (notamment l'absence d'hybridations croisées ou plus généralement d'homologie avec les transcrits de l'organisme étudié) qui, comme précisé précédemment, plus ils sont nombreux, plus ils sont efficaces pour normaliser les données. En dehors de la normalisation, les spikes peuvent cependant se révéler utiles pour la réalisation de contrôles de la qualité des hybridations (Tong et al., 2006) ou encore pour l'alignement des grilles de cerclage des spots (analyse d'image) lorsque les sondes correspondantes sont placées dans les angles des puces à ADN.

Une autre méthode qui a été assez largement utilisée consiste à supposer que l'expression du « **gène moyen** » ou du « gène médian » n'est pas modulée entre les échantillons, ce qui suppose, si on n'exclue pas l'hypothèse qu'il existe malgré tout des modulations d'expression de gènes, que les régulations positives et négatives provoquées par les différentes conditions expérimentales se compensent. Si cette hypothèse est acceptée (elle peut sembler généralement acceptable pour des puces pangénomiques mais bien plus douteuse pour des puces dédiées issues par définition d'une sélection biaisée et susceptibles de présenter une large proportion de gènes dont l'expression est modulée), on peut alors utiliser le signal moyen (ou médian) comme coefficient de normalisation. L'avantage de cette méthode est la stabilité de l'estimation du coefficient de normalisation (qui est basé sur le plus grand nombre de gènes possibles) et son principal désavantage est lié au fait qu'elle n'est pas applicable dans tous les cas.

Dans les expériences sur puces en double couleur, l'étude de graphiques comme les MA-plots (où  $M=\log_2(R/G)$  est représenté en ordonnée et  $A=\log_2((RG)^{1/2})=0,5\cdot\log_2(RG)$  est représenté en abscisse avec R=signal Cy5 dans le rouge et G=signal Cy3 dans le vert) ou les RI-plots (où  $R=ratio=M=\log_2(R/G)$  et  $I=intensity=\log_{10}(RG)$  avec les mêmes définitions de R et G) révèlent une dépendance des rapports d'expression ( $M=R$ ) en fonction de l'intensité des spots (A ou I). Cela est due au fait que le biais introduit par les fluorophores ne semble pas constant en fonction de l'intensité des spots conduisant ainsi à la formation d'un nuage de point qui n'est pas aligné le long d'un axe parallèle à l'axe des abscisses (représentant l'intensité sous la forme A ou I). Il a donc été proposé de modéliser ce lien de dépendance et de soustraire (car les données sont en logarithme) cette fonction aux données (voir notamment

Quackenbush, 2002; Yang et al., 2002). Bien que plusieurs méthodes aient été proposées, la plus couramment utilisée est sans doute la méthode **lowess** ou encore **loess** (locally weighted regression) qui, globalement, consiste en une régression polynomiale réalisée sur une fenêtre mobile d'intensités associée à l'application de poids aux observations pour limiter l'influence des points les plus extrêmes en termes de log(ratios). La taille de la fenêtre utilisée est laissée à l'appréciation de l'utilisateur, sachant que plus la fenêtre est large, plus la courbe résultante sera lisse. Une fois le modèle  $f$  de dépendance  $\log(\text{ratio})=f(\text{intensité})$  déterminé par cette méthode, il ne reste qu'à retrancher pour chaque gène  $i$  la valeur prise par le modèle à l'intensité  $x_i$ , que l'on notera  $y(x_i)$ , au  $\log_2(R_i/G_i)$ . Le calcul est donc le suivant :

$$\log_2\left(\frac{R_i}{G_i}\right) - y(x_i)$$

Ce qui est équivalent à

$$\log_2\left(\frac{R_i}{G_i}\right) - \log_2(2^{y(x_i)})$$

Et peut également s'écrire

$$\log_2\left(\frac{R_i}{G_i} \times \frac{1}{2^{y(x_i)}}\right) = \log_2\left(\frac{R_i}{G_i \times 2^{y(x_i)}}\right) = \log_2(R_i) - \log_2(G_i \times 2^{y(x_i)})$$

Ce qui illustre le fait que cette normalisation consiste à conserver les valeurs correspondant à l'un des canaux (le rouge) et à multiplier les valeurs de l'autre canal (le vert) par un facteur de normalisation dépendant de l'intensité et revient donc, comme pour les autres méthodes de normalisation à multiplier les valeurs initiales par un certain coefficient de normalisation (sauf qu'ici, il dépend de l'intensité).

D'autres méthodes de normalisation ont bien sûr été proposées dans la littérature (voir par exemple la formulation de la normalisation sous la forme d'un modèle d'analyse de variance dans Kerr et al., 2002a; voir par exemple la formulation de la normalisation sous la forme d'un modèle d'analyse de variance dans Kerr et al., 2000) qui ne pourront pas être abordées ici (pour des synthèses sur les méthodes de normalisation, voir Parmigiani, 2003, pp 15-16 et chapitres cités dans ces pages; Simon, 2003, pp 53-64; Speed, 2003, pp 25-29). Citons cependant le fait que les méthodes que nous avons évoquées, ainsi que d'autres, peuvent éventuellement s'appliquer sur des sous-groupes de gènes au sein des puces (comme

par exemple tous les gènes spottés par une même aiguille) dans le but de corriger des biais de type spatial. Des comparaisons de méthodes de normalisation ont également été publiées (voir par exemple Bolstad et al., 2003; Yang et al., 2002) et sont facilitées dans certains cas par la mise à disposition de jeux de données dont les modulations de gènes sont connues a priori (utilisation de spikes par exemple dans les jeux de données appelés « spike-in datasets » ou « Latin square data » de Affymetrix et disponibles sur [www.affymetrix.com/support/technical/sample\\_data/datasets.affx](http://www.affymetrix.com/support/technical/sample_data/datasets.affx)) permettant ainsi de comparer l'efficacité des méthodes de normalisation pour la détection des gènes différentiellement exprimés.

Bien que constituant souvent une réponse décevante pour beaucoup de biologistes, il n'existe pas, à l'heure actuelle, une méthode de normalisation surpassant toutes les autres et dans tous les cas. Chacune est apte à corriger certains biais et s'adapte plus spécifiquement à certains cas particuliers, mais l'utilisateur, surtout lorsqu'il s'agit d'un biologiste peu aguerri à la pratique de l'analyse des données, est souvent plongé dans le flou le plus total quant à la méthode à choisir. Il est bon dans ce cas là de se rappeler que chacune des méthodes proposées est en fait basée sur une ou des **hypothèses** (qui ont été évoquées lors de la présentation des différentes méthodes) vis-à-vis desquelles le biologiste est très souvent en bien meilleure position que le statisticien pour juger de leur caractère acceptable ou non dans son contexte expérimental. En redonnant ainsi une importance comparable aux points de vue du statisticien (qui propose des méthodes) et du biologiste (qui juge de l'acceptabilité des hypothèses sous-jacentes), le dialogue nécessaire sur la méthode de normalisation à appliquer s'en trouve généralement grandement facilité et bien plus fructueux.

Enfin, dans la pratique, il est indispensable de réaliser un certain nombre de **représentations graphiques** pour choisir et évaluer l'effet de la normalisation des données. On peut par exemple utiliser pour cela les MA-plots que nous avons présentés (puces double couleur), les boxplots (ou boîtes à moustaches), particulièrement pratiques pour comparer les distributions de valeurs de multiples puces de manière parallèle, les comparaisons de puces deux à deux ou encore les graphes de distribution des intensités ou des log(ratios).

Maintenant que nous avons évoqué les aspects de prétraitement et de normalisation des données, nous allons évoquer une variété d'outils disponibles pour répondre aux questions posées par l'expérimentateur.

### III.4.3. Comparaison de classes

L'objet de la comparaison de classes est d'identifier les éléments qui diffèrent entre deux (ou plus) classes d'individus. Typiquement, il s'agit, dans le contexte des microarrays, d'identifier les gènes différentiellement exprimés entre deux (ou plus) groupes d'individus. La première approche qui a été utilisée dans ce contexte a été de considérer le **rapport d'expression** entre les individus ou les groupes d'individus (le fameux ratio ou log[ratio]) et de fixer (arbitrairement ou par le biais de l'étude de la distribution des rapports d'expression) **un seuil** afin de déclarer comme « significativement » différentiellement exprimés tous les gènes dont le rapport d'expression (en valeur absolue) dépassait ce seuil. Alors que cette approche a été récemment remise au goût du jour par le biais des études de comparaison de plateformes de puces à ADN (voir par exemple Guo et al., 2006b), elle présente néanmoins deux désavantages. Le premier est qu'étant donné le nombre considérable de gènes mesurés et le nombre généralement limité d'échantillons, il est très probable que les jeux de données contiennent systématiquement un certain nombre de valeurs aberrantes dont l'influence sur le calcul des rapports d'expression correspondants peut être très grande et conduire ainsi à la génération de faux positifs. Cependant, nous avons précisé précédemment que c'était le rôle des étapes d'analyse d'image et de prétraitement des données de repérer et d'éliminer ces valeurs aberrantes. Le deuxième désavantage de cette méthode est qu'elle affecte implicitement à chaque gène la même variance, ce qui constitue une hypothèse forte qui est très peu probable (voire inacceptable) dans la plupart des contextes.

Une autre approche assez naturelle, consiste à utiliser un test statistique dont le test de Student, évoqué ci-dessous, est sans doute le plus utilisé dans la cadre de la comparaison de deux groupes d'échantillons (dont les observations doivent être indépendantes et distribuées selon une loi normale). Le test de Student (notons au passage qu'il requiert également la condition d'homoscédasticité, c'est à dire d'égalité des variances entre les groupes comparés, sinon le test de Welch doit lui être substitué) est basé sur la statistique de test suivante (présentée ici dans un contexte double couleur) :

$$t = \frac{\sqrt{n} \times \bar{M}}{s}$$

Où n est le nombre de puces réalisées,  $\bar{M}$  est la moyenne des log(ratios) des n puces et s est l'écart-type de ces log(ratios). Cette statistique est parfois qualifiée de moyenne standardisée dans la mesure où elle s'écrit comme le rapport d'une moyenne (ou plutôt une différence de moyenne) par son erreur standard (sem pour standard error of the mean qui est égale à l'écart-

type divisé par la racine carrée du nombre d'observations). Sous l'hypothèse  $H_0$  (hypothèse nulle) selon laquelle  $M$  est nul (i.e. que le gène n'est pas différentiellement exprimé), alors cette quantité suit une loi de Student à  $n-1$  degrés de liberté. Un moyen de sélectionner les gènes différentiellement exprimés est donc de calculer les valeurs de  $|t|$  pour chaque gène et de fixer un seuil sur ces valeurs (ou de calculer les p-values ou probabilités critiques correspondant à ces valeurs de  $t$  en utilisant la loi de Student et de fixer un seuil sur les p-values, ce qui est équivalent si ce n'est que l'on n'a pas dans ce cas à recourir aux valeurs absolues). Tout comme la méthode précédente, cette procédure souffre du manque d'observations et de la multiplicité des estimations réalisées qui immanquablement généreront des valeurs de  $t$  extrêmes en raison de termes d'écart-type s infinitésimaux, malgré des valeurs de  $M$  proche de 0. C'est donc essentiellement, à ce stade, un problème de stabilité des estimateurs qui se présente, lié principalement au manque d'observations.

Un moyen pour s'affranchir de ces problèmes consiste à utiliser des estimateurs plus robustes, basés par exemple sur la médiane au lieu de la moyenne, ou sur les rangs au lieu des intensités (test de Mann-Whitney ou de Wilcoxon) mais ces méthodes souffrent néanmoins de problème assez similaires en raison de l'ampleur du problème lié aux estimations. Pour éviter le problème des estimations de variance trop faibles, des filtres peuvent éventuellement être appliqués pour éliminer, par exemple, au moins temporairement, les 1% de variances les plus faibles (Lönnstedt and Speed, 2002). Efron *et al.* ont proposé d'ajouter le 90<sup>ème</sup> percentile des écarts-types de tous les gènes au dénominateur s de la statistique de Student pour stabiliser la variance (Efron et al., 2000). D'autres modifications de la statistique de Student (on parle de **statistique de Student modérée**) ont également été envisagées comme l'ajout d'une constante minimisant le coefficient de variation dans l'approche SAM (Significance analysis of Microarrays, Tusher et al., 2001) ou bien l'approche bayésienne empirique (Smyth, 2004) utilisée dans le package limma (Smyth, 2005) de Bioconductor (Gentleman et al., 2004). Ce package propose également l'utilisation de la statistique B, développée dans un contexte bayésien empirique (Lönnstedt and Speed, 2002). La définition d'une nouvelle statistique de test se heurte souvent au problème de la définition de la loi de cette statistique sous l'hypothèse nulle  $H_0$  mais la disponibilité d'une masse suffisante de données issues de certaines expériences de puces à ADN permet généralement pour cet aspect de recourir à son estimation par le biais de techniques de ré-échantillonnage des données. Ces techniques de ré-échantillonnage ou de permutations peuvent d'ailleurs également être utilisées, lorsque le nombre d'échantillons analysés est suffisant, pour éviter d'émettre des hypothèses sur la distribution des données (voir notamment Ge et al., 2003) souvent nécessaires avec les

statistiques de tests couramment utilisées (hypothèse de normalité pour la statistique de Student par exemple). Nous évoquerons finalement une dernière approche qui consiste à trouver un intermédiaire entre l'estimation d'une variance par gène (dont nous avons vu qu'elle n'était pas idéale en raison de la présence de valeurs extrêmes liée au manque d'observations) et l'estimation d'une variance unique pour tous les gènes (qui considère donc que tous les gènes ont la même variance ce qui ne constitue pas une hypothèse acceptable). Dans le cadre de la mise en œuvre de différentes méthodes sur un jeu de données issu de ce travail de doctorat (Baccini et al., 2005) et présenté au paragraphe III.5. nous avons choisi de définir, de manière arbitraire sur la base d'un histogramme des variances, trois groupes de gènes présentant des variances homogènes. Dans une analyse portant sur 120 gènes au départ, seules 3 variances sont ainsi estimées et utilisées dans le modèle d'analyse de variance comme présenté dans (Baccini et al., 2005). Cette approche a également été exploitée par d'autres équipes et constitue l'un des atouts de la méthode Varmixt (Delmar et al., 2005) qui, dans sa formulation la plus récente (appelée VM), ne classe pas de manière effective les gènes en sous-groupes mais utilise plutôt la probabilité qu'un gène appartienne à un sous-groupe (autorisant ainsi un gène à se situer entre plusieurs sous-groupes) comme pondération dans le calcul de la variance.

Un large panel de méthodes ont ainsi été développées dans le contexte de la comparaison de classes. C'est en particulier vrai pour la recherche de gènes différentiellement exprimés entre deux groupes d'échantillons. Les méthodes applicables à la comparaison de plus de deux groupes d'échantillons sont plus rares mais certaines des méthodes que nous avons évoquées peuvent cependant s'appliquer dans ce contexte. L'analyse de variance (anova) constitue une méthode de choix pour l'étude des effets de facteurs expérimentaux sur l'expression des gènes mais elle nécessite généralement la comparaison ultérieure des groupes deux à deux (en tirant néanmoins bénéfice de la réalisation de l'anova), par le biais de ce que l'on appelle parfois des tests post-hoc, afin de répondre complètement aux questions posées. La présentation détaillée de l'analyse de variance est clairement hors du contexte du présent manuscrit (pour une présentation en détail voir par exemple Dagnelie, 1998), mais rappelons tout de même qu'elle repose sur une modélisation des données (la variable à expliquer) par des facteurs (variables qualitatives explicatives, comme par exemple le traitement ou le génotype), éventuellement en interaction, et une comparaison, par l'intermédiaire d'un test de Fisher (rapport de variances), de la variabilité induite par ces facteurs par rapport à la variabilité non expliquée par le modèle (variabilité résiduelle, qui correspond normalement à la variabilité biologique). Comme le test de Student, l'anova repose sur des hypothèses (par

**A.**

$H_0$  : le gène n'est pas différentiellement exprimé ( $\mu_1 = \mu_2$ )

$H_1$  : le gène est différentiellement exprimé ( $\mu_1 \neq \mu_2$ )

		Décision	
		$H_0$	$H_1$
Réalité	$H_0$	$1-\alpha$	$\alpha$
	$H_1$	$\beta$	$1-\beta$

$\alpha = P(\text{choisir } H_1 \mid H_0 \text{ vraie})$   
 $= P(\text{rejeter } H_0 \mid H_0 \text{ vraie})$   
 $= \text{erreur de type I}$

$\beta = P(\text{choisir } H_0 \mid H_0 \text{ fausse})$   
 $= \text{erreur de type II}$

$1-\beta = \text{Puissance du test}$

**B.****Décision**

		Nombre d' $H_0$ acceptées	Nombre d' $H_0$ rejetées	
Réalité	Nombre d' $H_0$ vraies	$U_{\text{négatifs}}^{\text{vrais}}$	$V_{\text{positifs}}^{\text{faux}}$	$m_0$
	Nombre d' $H_0$ fausses	$T_{\text{négatifs}}^{\text{faux}}$	$S_{\text{positifs}}^{\text{vrais}}$	$m_1$
		$m-R$	R	m

$$\text{FWER} = P(V>0)$$

$$\text{FDR} = E(V/R) \text{ si } R>0 \\ = 1 \text{ sinon}$$

Figure 55 : Risques et taux d'erreurs pour les procédures de test simples et multiples

La Figure A présente les différentes probabilités associées avec une procédure de test simple (test d'une seule hypothèse  $H_0$ ) et indique quelles sont les erreurs de type I et de type II ainsi que la définition de la puissance du test. La Figure B présente une généralisation de ces erreurs au cas des tests multiples. Les notations en noir sont celles utilisées par Benjamini & Hochberg (1995). m est connu puisqu'il représente le nombre de tests réalisés (le nombre de gènes sur lesquels on teste l'hypothèse  $H_0$ ) mais  $m_0$  et  $m_1$  sont inconnus. R est une variable aléatoire observable représentant le nombre de gènes déclarés significativement différemment exprimés. S, T U et V sont des variables aléatoires non observables. Les définitions du Family wise error rate (FWER) et du FDR sont présentées sous la table B. P est une probabilité et E est une espérance.

ordre d'importance : homoscédasticité, indépendance et normalité) dont la véracité doit faire l'objet d'évaluations, souvent via des représentations graphiques réalisées a posteriori (l'indépendance étant liée à la façon dont l'expérience a été menée, elle ne nécessite généralement pas de vérification). Dans nos expériences qui contiennent le plus souvent deux facteurs croisés (typiquement génotype et traitement), nous utilisons souvent l'anova suivie d'un test de Student dont le terme d'écart-type est tiré du modèle d'anova (variance résiduelle de l'anova) et donc calculé sur toutes les données afin d'en améliorer l'estimation. Soulignons à nouveau que l'anova est basée sur des hypothèses, identiques à celles du test de Student.

Quel que soit le contexte de la comparaison de classe effectuée (deux groupes ou plus), ou la méthode choisie pour effectuer cette comparaison, il est nécessaire à un moment donné de **définir un seuil** permettant de scinder la liste des gènes dont l'expression est mesurée en deux groupes : les gènes différentiellement exprimés et ceux qui ne le sont pas. En biologie moléculaire, on a généralement l'habitude d'utiliser un seuil sur les probabilités critiques (p-values qui mesurent le risque qu'un gène soit un faux positif) de 5%. Ce seuil représente, lorsque l'on ne réalise qu'un seul test, le risque de première espèce ou de type I (noté  $\alpha$ ), c'est à dire la probabilité de choisir  $H_1$  (le gène est différentiellement exprimé), alors que c'est  $H_0$  (le gène n'est pas différentiellement exprimé) qui est vraie ([Figure 55A](#)). En choisissant la procédure de test qui possède la meilleure puissance (la puissance est la probabilité de choisir  $H_1$  sachant que  $H_1$  est vraie et s'écrit  $1-\beta$  où  $\beta$  est la probabilité de choisir  $H_0$  sachant que  $H_1$  est vraie), l'utilisateur contrôle ainsi le risque d'obtenir un faux positif (risque  $\alpha$ ) et minimise son risque d'avoir un faux négatif ( $\beta$ , par le biais de la puissance du test qui n'est toutefois pas maîtrisée à un niveau fixe pour tout différentiel d'expression). Si le même risque  $\alpha=5\%$  est maintenant appliqué à 100 tests de différentiels d'expression réalisés simultanément, on s'attend à avoir en moyenne 5 tests déclarant des gènes différentiellement exprimés, alors qu'ils ne le sont pas en réalité et donc à avoir en moyenne 5 faux positifs. L'application de cette idée à 1000 (50 faux positifs), 10000 (500 faux positifs) ou encore 40000 tests (2000 faux positifs) illustre le problème lié à ce que l'on appelle la **multiplicité des tests** : lorsque de nombreuses hypothèses sont testées, la probabilité qu'une erreur de type I soit commise peut augmenter de manière considérable avec le nombre d'hypothèses. Plusieurs généralisations au cas des tests multiples des erreurs de type I ( $\alpha$ ) ou de type II ( $\beta$ ) ont été proposées (pour des présentations complètes voir Dudoit et al., 2003b; Ge et al., 2003) dont la plus utilisée est sans doute le family-wise type I error rate (FWER) qui est la probabilité de commettre une erreur de type I ([Figure 55B](#)). Le contrôle des taux

d'erreur par une procédure de tests multiples donnée est qualifié de fort lorsqu'il est valable sous n'importe quelle combinaison d'hypothèses nulles  $H_0$  vraies ou fausses (c'est à dire pour n'importe quelles valeurs de  $m_0$  et  $m_1$  en utilisant les notations de la Figure 55B) et il est qualifié de faible lorsqu'il n'est valable que dans le cas où toutes les hypothèses  $H_0$  sont vraies (i.e. lorsque  $m_0=m$ ). Dans le contexte des puces à ADN, le biologiste a généralement conçu son expérience dans le but d'observer des modulations d'expression de gènes et un contrôle faible semble donc peu approprié. De nombreuses procédures ont été développées qui contrôlent le FWER dont la plus connue est sans doute la procédure de Bonferroni qui consiste à utiliser un niveau de risque de type I de  $\alpha^*/m$  (où  $m$  est toujours le nombre de tests réalisés) pour chacun des tests réalisés, ce qui assure que le FWER est contrôlé, au sens fort, à un niveau  $\alpha^*$  (i.e.  $\text{FWER} \leq \alpha^*$ ). Notons que cette procédure, comme les suivantes que nous allons évoquer, peuvent être présentées non pas en termes de risque de première espèce utilisé pour chaque test réalisé, mais sous la forme d'ajustement des p-values obtenues à partir de ces tests, les p-values ajustées pouvant être directement comparées au FWER  $\alpha^*$  désiré (Dudoit et al., 2003b; Dudoit et al., 2002; Ge et al., 2003). Ainsi, la procédure de Bonferroni consiste à multiplier toutes les p-values obtenues par  $m$  (si la valeur  $p \times m$  dépasse 1 on prend bien sûr cette valeur de 1 qui est la borne supérieure d'une probabilité critique). Malheureusement, cette procédure simple, si elle est applicable dans certains cas avec des puces à ADN dédiées (Martin et al., 2005) est inutile lorsque le nombre  $m$  de tests réalisés est trop élevé, comme avec des puces pangénomiques. D'autres méthodes permettant de contrôler le FWER ont été développées comme la procédure de Sidak, mais celle-ci suppose que les tests sont indépendants, ce qui n'est pas le cas pour les puces à ADN. Ces procédures, réalisées en une seule étape, tendent à être trop conservatives (c'est à dire à rejeter trop peu d' $H_0$  et donc à être peu puissantes) car elles sont, dans leur esprit, adaptées à la plus petite des p-values. C'est pourquoi des procédures adaptatives (en l'occurrence descendantes) ont été proposées qui ordonnent les p-values de la plus faible à la plus grande et appliquent successivement les procédures de Bonferroni (appelé parfois procédure de Holm sous sa forme adaptative) ou de Sidak aux p-values en partant de la plus faible p-value et en remplaçant à chaque étape  $k$  la valeur  $m$  par  $m-k+1$  (l'étape 1 étant l'application de la procédure à la plus faible p-value). D'après les comparaisons de méthodes que nous avons réalisées (données non présentées) et celles publiées dans la littérature (Dudoit et al., 2003b), ces méthodes restent néanmoins trop conservatives dans le contexte de puces à ADN pangénomiques. Des procédures prenant en compte la dépendance entre les gènes (présence de groupes de gènes dont les niveaux

d'expression sont corrélés entre eux) ont également été proposées comme les méthodes de Westfall & Young minP et maxT (Dudoit et al., 2003b) qui sont basées sur des tests par permutation en appliquant les mêmes permutations à tous les gènes pour préserver leur structure de corrélation. Mais comme la plupart des tests par permutation, ces méthodes ne sont réalisables qu'avec un nombre suffisant de réplicats dans chaque condition expérimentale autorisant ainsi la réalisation d'un grand nombre de permutations (pour deux groupes de taille  $n$ , le nombre de permutations possible est  $\frac{(2n)!}{n!n!}$ ) et donc une borne inférieure très faible pour l'estimation des p-values (pour un algorithme simple d'estimation de p-values non ajustées comme présenté en boîte 1 de Dudoit et al., 2002, la borne inférieure des valeurs prises par les p-values vaut  $2/B$  où  $B$  est le nombre de permutations effectuées).

Finalement, quelle que soit la procédure choisie, le plus important reste malgré tout que l'utilisateur ait bien défini au préalable **l'objectif** qu'il assigne à son opération de sélection de gènes. Ainsi, en discutant avec des utilisateurs, notamment dans le cadre de formations à l'analyse des données des puces à ADN ou de collaborations, de multiples situations différentes peuvent être rencontrées. Il est fréquent que ceux-ci ne souhaitent pas obtenir de listes trop longues de gènes différentiellement exprimés, tout simplement parce qu'ils se sentirraient dans l'incapacité de gérer des listes de plusieurs centaines de gènes d'un point de vue de leur étude bibliographique par exemple, ou encore parce qu'ils travaillent sur un modèle peu étudié et ont utilisé des puces à ADN dans le but unique de découvrir quelques régulations nouvelles et de se concentrer sur celles qui ont la plus forte probabilité d'être vraies. Dans d'autres cas au contraire, les listes les plus longues sont souhaitées dans le but de décrire de manière la plus exhaustive possible le système étudié, ou bien parce qu'il s'agit d'un modèle étudié par de nombreux laboratoires et que les régulations les plus fortes et les plus nettes ont déjà été décrites. Dans les deux cas, la « chasse » aux faux positifs constitue bien un objectif, mais les mêmes outils ne seront pas utilisés. Dans le premier cas, les faux positifs seront de toute façon rares parmi les gènes présentant les plus forts différentiels d'expression et les p-values les plus faibles (quelque soit la procédure utilisée), et ce sont ici les artefacts (défauts au niveau des filtres appliqués aux données ou de l'analyse d'image) qui risquent le plus d'en générer. Une procédure contrôlant le FWER peut sans doute s'appliquer dans ce cas là. Dans le second cas, l'objectif de l'utilisateur est a priori plutôt de contrôler la puissance de sa procédure de sélection puisqu'il souhaite être en mesure de détecter idéalement tous les gènes régulés dans son expérience (les limitations de l'outil de mesure mises de côté dans ce contexte) mais il n'est néanmoins pas prêt à accepter que ses résultats

soient entachés d'une proportion trop importante de faux positifs (qui pourraient, de plus, lui coûter cher s'il envisage de confirmer une partie de ses résultats par une technique indépendante). Il est néanmoins assurément près à en accepter plus d'un seul et unique si cela peut lui permettre d'identifier de nombreux nouveaux gènes régulés. Une procédure contrôlant le FWER (la probabilité qu'il y ait au moins un faux positif c'est à dire le risque de commettre une seule erreur de type I) n'apparaît donc pas appropriée dans ce contexte et ce que l'utilisateur souhaite contrôler, c'est plutôt la proportion de faux positifs parmi ses résultats. C'est sur la base d'exemples similaires que Benjamini & Hochberg (Benjamini and Hochberg, 1995) ont proposé en 1995 l'introduction du concept de false discovery rate (FDR, ou taux de faux positifs en français, voir sa définition exacte [Figure 55B](#)) dont le contrôle est moins conservatif que celui du FWER et permet donc, dans le contexte des puces à ADN, de détecter plus de gènes différentiellement exprimés (voir notamment les estimations de puissance réalisées sur des données simulées dans Dudoit et al., 2003b). La procédure initialement proposée pour contrôler le FDR (Benjamini and Hochberg, 1995) consiste à :

- 1) ordonner les p-values de la plus faible à la plus grande :  $p_{(1)} \leq p_{(2)} \dots \leq p_{(k)} \dots \leq p_{(m)}$
- 2) choisir le plus grand indice  $k$  tel que :  $p_{(k)} \leq \frac{k}{m} \alpha^*$
- 3) déclarer significatives toutes les p-values d'indice inférieur à  $k$ .

Sous l'hypothèse que les tests sont indépendants, cette procédure garantit que :

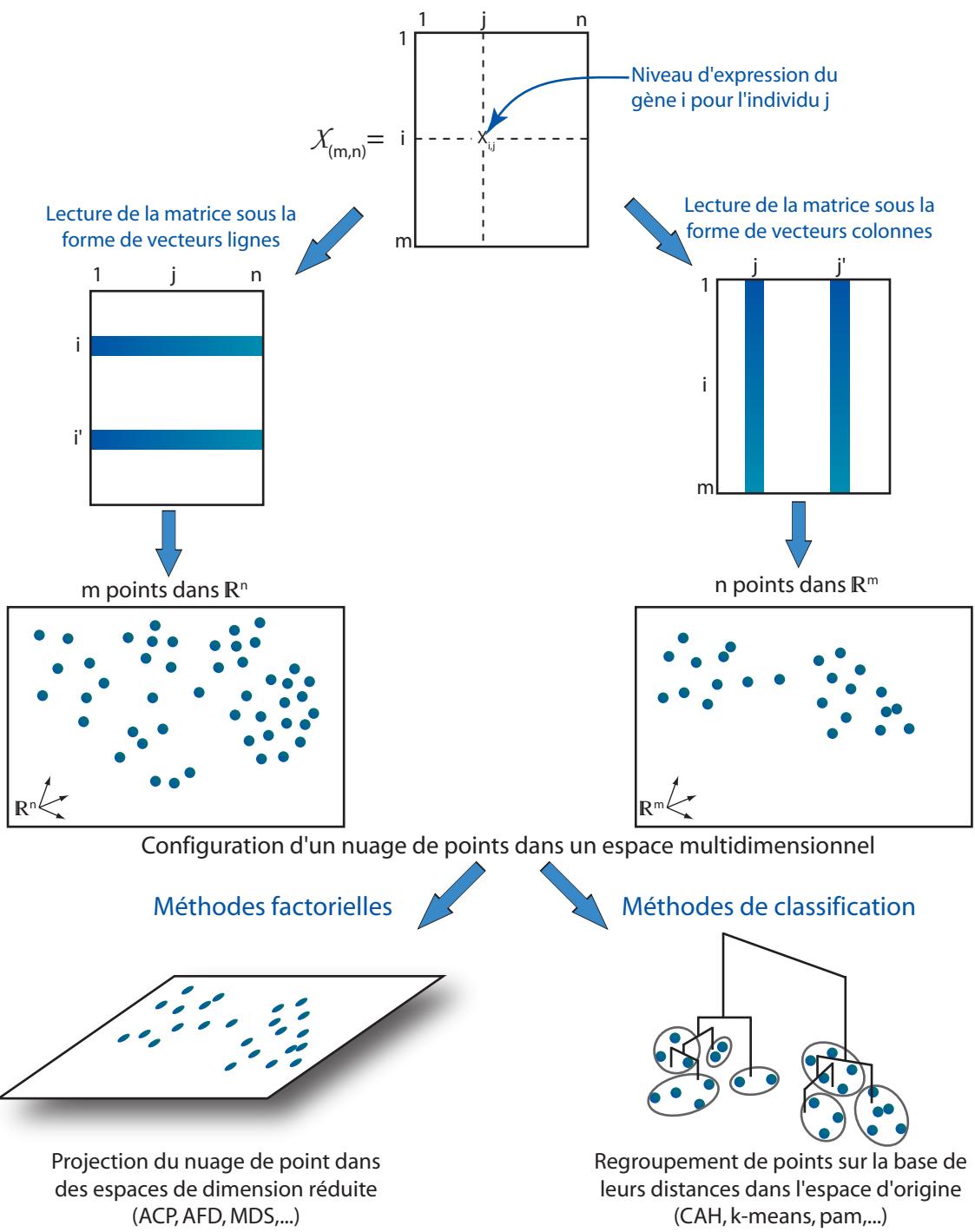
$FDR \leq \frac{m_0}{m} \alpha^* \leq \alpha^*$ . On voit ainsi que l'estimation de  $m_0$  peut constituer un moyen d'améliorer la procédure et des recherches en statistique vont dans ce sens comme par exemple dans le cadre de l'utilisation des q-values (Storey et al., 2004; Storey and Tibshirani, 2003) qui s'expriment à partir des p-values de la manière suivante :

$$q_{(i)} = \frac{\hat{m}_0}{R_i} p_{(i)}$$

Où  $p_{(i)}$  est la p-value ordonnée du gène  $i$ ,  $R_i$  est le nombre total de gènes déclarés différentiellement exprimés dont la p-value est inférieure à  $p_{(i)}$  et  $\hat{m}_0$  est une estimation obtenue à partir de la distribution des p-values (qui nécessite la définition d'un paramètre de réglage  $\lambda$ ). Une autre procédure s'appliquant dans certains cas de corrélations positives entre les tests a également été proposée (procédure BY du package multtest de Bioconductor, Benjamini and Yekutieli, 2001).

Enfin, il est important de comprendre que le FDR, dans sa définition, fournit une idée du nombre de faux positifs que l'on peut obtenir si une expérience est réalisée un nombre infini de fois et ne fourni donc pas strictement le nombre de faux positifs dans une expérience donnée. De plus, le seuil de FDR fixé par l'utilisateur et lui permettant de définir sa sélection de gènes différentiellement exprimés est arbitraire. Des méthodes ont donc été développées pour rendre compte pour chaque gène de la probabilité qu'il a d'être un faux positif (FDR local dans Aubert et al., 2004; q-value dans Storey and Tibshirani, 2003).

Nous avons vu ici qu'il existe un large panel de méthodes et de procédures permettant de définir une liste de gènes différentiellement exprimés entre deux (ou plus) conditions expérimentales. Ce domaine de recherche est particulièrement riche et actif et mes compétences limitées en statistique ne me permettent pas d'en présenter toute l'étendue et la diversité. Nous avons cependant, je pense, évoqué ici les principaux enjeux et les méthodes qui sont les plus utilisées par les biologistes pour analyser leurs données. La méthode SAM (Tusher et al., 2001) arrive sans doute en tête dans ce palmarès, peut-être principalement en raison de sa distribution sous la forme d'une macro Excel et de sa publication initiale dans un grand journal du domaine de la biologie dans la mesure où elle semble être loin de faire l'unanimité auprès de tous les spécialistes en statistique du domaine (voir par exemple Dudoit et al., 2003b). Ce qui est néanmoins frappant dans la mise en œuvre de ces procédures de test, c'est qu'elles sont très souvent combinées à l'application d'autres filtres sur les données (en dehors des filtres de prétraitement que nous avons déjà évoqué). Ainsi, il n'est pas rare que nous (les biologistes pratiquant ces méthodes) combinions l'application d'un seuil en termes de rapport d'expression (on ne conserve que les gènes dont le différentiel d'expression est au moins égal à un facteur 2 par exemple) et d'une procédure de test (SAM par exemple). L'inconvénient pour le statisticien (sans doute dans une moindre mesure pour le biologiste...) est que les propriétés (puissance, contrôle du FDR par exemple) d'une telle procédure sont difficiles à établir et que les démonstrations faites dans le cadre de l'utilisation d'une procédure donnée ne tiennent généralement plus dans ces nouvelles conditions. Le problème vient peut-être du fait que les tests d'hypothèses pour le statisticien sont élaborés pour faire prendre une décision à leur utilisateur alors que le biologiste, à mon sens, conçoit plus cet outil comme une simple aide supplémentaire à la décision. Une illustration de cette idée peut-être aisément trouvée dans les nombreux articles qui commentent de manière quasi-équivalente des résultats déclarés significatifs et ceux que les auteurs appellent souvent des « tendances ». Là encore, je pense que l'essentiel se trouve finalement dans la qualité de la réponse qui est apportée à la question du biologiste, laquelle dépend aussi largement d'une



**Figure 56 : Principe général des analyses statistiques exploratoires multidimensionnelles**

Les données sont représentées sous la forme d'une matrice gène x échantillons de dimension  $m \times n$ . Cette matrice peut être lue sous la forme de vecteurs lignes (les gènes sont des points dans un espace de dimension  $n$  défini par les échantillons et leurs coordonnées sur les axes de cet espace sont fournies par leur niveau d'expression pour chaque échantillon) ou sous la forme de vecteurs colonnes (les échantillons sont des points dans l'espace de dimension  $m$  défini par les gènes). Dans les deux cas, le problème consiste à étudier la configuration d'un nuage de points dans un espace de dimension trop importante pour être représenté directement de manière graphique. Deux familles de méthodes fournissent des résultats complémentaires et/ou adaptés à la réponse à des questions spécifiques : les méthodes factorielles réalisent une projection du nuage de point sur des espaces de dimension réduite en optimisant un critère (comme l'inertie du nuage en ACP par exemple) alors que les méthodes de classification regroupent les points sur la base d'une mesure de leurs distances calculées dans l'espace d'origine. ACP : Analyse en Composantes Principales, AFD : Analyse Factorielle Discriminante, MDS : multidimensional Scaling ou positionnement multidimensionnel, CAH : classification ascendante hiérarchique, pam : partitionning around medoids. Schéma inspiré de Lebart L et al. (2000).

formulation adéquate et précise de la question et des attentes. Nous verrons dans le paragraphe III.4.6. qu'une approche séduisante, répondant en partie au besoin d'intégration de connaissances a priori dans les tests statistiques, consiste à combiner les mesures d'expression géniques à des connaissances a priori de groupes de gènes dans la recherche de fonctions biologiques affectées par les conditions expérimentales.

#### **III.4.4. Découverte de classes**

L'objectif de la découverte de classes consiste à répondre à des questions du type « Peut-on identifier des sous-groupes homogènes de gènes (ou d'échantillons) caractérisés par des profils d'expression similaires ? ». Ce type de questions s'applique à de multiples domaines comme la recherche de gènes corégulés, l'identification et la caractérisation de sous-type de tumeurs, la recherche de points aberrants ou tout simplement l'exploration des données. De part sa formulation, la question renvoie immédiatement à des analyses de type non supervisées, c'est à dire où des connaissances a priori sur la nature et la composition des groupes ou sous-groupes ne sont pas utilisées dans l'analyse. C'est la méthode utilisée qui doit fournir une « cartographie » ou « typologie » des objets (gènes ou échantillons) suggérant ainsi un découpage en classes. Les méthodes utilisées pour répondre à ces questions appartiennent au domaine de « l'analyse des données et de la reconnaissance de formes » ou, formulé autrement, à celui des « analyses statistiques exploratoires multidimensionnelles » (pour des références en français, voir notamment Dagnelie, 1975; pour des références en français, voir notamment Escofier and Pagès, 1998; Lebart et al., 2000).

Le principe général de ces analyses et leur séparation en deux sous-familles de méthodes, les méthodes de classification (méthodes de clustering en anglais) d'une part et les méthodes factorielles d'autres part, permettent d'illustrer de manière compréhensible les principales approches utilisées dans ce domaine ([Figure 56](#)). L'intérêt majeur de ces méthodes et qu'elles tiennent compte directement du caractère multidimensionnel des données et étudient la topologie des gènes et/ou des individus (via leur structure de corrélation et/ou leurs distances), alors que les méthodes de comparaison de classes fonctionnent plutôt sur la base d'une analyse gène-par-gène et que l'intégration des données est généralement réalisée après analyse statistique, lors de l'étape d'interprétation. La formulation géométrique et/ou algorithmique de la plupart de ces méthodes d'analyse favorise également grandement leur diffusion et leur enseignement à des publics non spécialisés en mathématique ou statistique appliquée. De nombreuses méthodes ont été appliquées aux données issues des puces à ADN et certaines améliorées voire développées spécifiquement pour leur analyse (pour des

présentations de différentes méthodes voir Parmigiani, 2003, pp 21-28 et chapitres cités dans ces pages; Simon, 2003, pp 121-155; Speed, 2003, pp 159-200).

Les **méthodes de classification** (clustering en anglais) visent à regrouper les points (échantillons ou gènes) sur la base de leurs distances calculées dans l'espace d'origine. Plus précisément, en partant d'une matrice de distances (ou de dissemblances) entre les objets (échantillons ou gènes), la classification vise à rechercher une typologie, une segmentation ou une partition des objets en classes par optimisation d'un critère. Ce sont parmi les premières méthodes qui ont été utilisées pour visualiser les relations entre des profils d'expression (Eisen et al., 1998). La complexité du problème à traiter (classification de milliers d'objets) impose le recours à un algorithme itératif. De plus, toutes les méthodes de classification nécessitent la définition préalable d'une distance (ou d'une mesure de dissemblance ou d'éloignement) entre les objets à classer et d'autres choix sont également souvent laissés à l'utilisateur à commencer par la méthode de classification mais aussi le critère à optimiser ou encore le nombre de classes. Plusieurs types de distances sont présentées et comparées dans l'article (Baccini et al., 2005) présenté au paragraphe III.5. Dans le contexte de la classification des gènes, il me semble que les distances basées sur la corrélation (du type 1-corrélation) constituent souvent un choix pertinent car nous avons vu que l'intensité d'un spot pouvait dépendre fortement de la nature de la sonde présente sur ce spot et que donc, les intensités des signaux obtenus pour différents gènes ne peuvent être directement interprétés du point de vue biologique comme représentant les niveaux relatifs d'expression des gènes entre eux. C'est donc plutôt la covariation des signaux qui intéresse le biologiste, plutôt que la comparaison de leurs niveaux absolus. Dans le cadre de l'analyse des individus (ou échantillons), la distance euclidienne (ou d'autres distances présentant des comportements relativement similaires) peut être intéressante pour détecter des individus aberrants ou extrêmes mais les distances basées sur la corrélation apportent également ici un point de vue intéressant sur les données. La méthode qui est sans doute la plus répandue dans l'analyse des données issues de puces à ADN est la classification ascendante hiérarchique (CAH) qui est réalisée selon les étapes suivantes :

- Au début chaque individu est une classe
- A chaque étape, on regroupe les deux objets les plus proches puis on recalcule les distances en considérant ces deux objets comme un seul objet
- A la fin, tous les individus sont regroupés dans une classe unique

Le résultat est la construction d'un arbre de classification ou dendrogramme. On voit ici que cet algorithme nécessite la définition d'une distance entre objet mais également d'un moyen

de calculer des distances entre un objet et un groupe d'objets ou encore entre deux groupes d'objets (on appelle cela le critère d'agglomération ou linkage en anglais). Les solutions les plus simples consistent à utiliser le saut minimum (single linkage : distance minimum entre deux groupes de points), le saut maximum (complete linkage : distance maximale entre deux groupes de points) ou le saut moyen (average linkage : distance entre les barycentres des deux groupes de points). D'après notre expérience de la manipulation de ces différents critères, le saut moyen donne généralement des résultats intéressants, le saut maximum peut également être pertinent dans certains cas mais le saut minimum souffre, lui, d'une forte sensibilité aux effets de chaîne, un phénomène auquel on peut s'attendre avec les classifications réalisées sur des milliers de gènes. Le critère de Ward, basé sur la maximisation de l'inertie inter-classes (ou sur la minimisation de l'inertie intra-classe, ce qui est équivalent), tend à fournir des classes de forme sphérique et présente l'avantage de fournir des sorties graphiques (hauteur de coupure en fonction du nombre de classes) qui peuvent aider à fixer le nombre de classes. En effet, la CAH ne fournit qu'une hiérarchie de classifications sous la forme d'un dendrogramme et il revient alors à l'utilisateur de choisir à quel niveau il va « couper » cet arbre pour définir ainsi les classes qu'il cherchera à interpréter. De plus, il est important de concevoir le dendrogramme sous la forme d'un mobile dont les branches peuvent tourner librement autour des fils, fournissant ainsi un nombre considérable de possibilités ( $2^{n-1}$  pour n objets classés) de représenter un même arbre de classification. Il est courant de rencontrer une présentation des données issues des puces à ADN sous la forme d'une matrice de carrés dont la couleur correspond à un codage des données initiales (rouge=forts différentiels d'expression et vert=faibles différentiels d'expression par exemple) et dont les lignes et les colonnes ont été réorganisées par le biais de CAH effectuées sur les échantillons puis sur les gènes. Peut-être encore plus dans cette situation où les couleurs et les éventuels « paquets » qu'elles forment attirent l'œil, cette notion de rotation des branches des dendrogrammes doit être gardée en tête pour assurer une interprétation pertinente des groupes d'objets et de leurs proximités éventuelles. Le principal inconvénient de la CAH reste néanmoins son côté rigide dans le sens où une erreur de regroupement effectuée à une étape ne peut être corrigée au cours des étapes suivantes. Une autre méthode assez souvent mise en œuvre sur les données de puces à ADN et qui ne souffre pas de ce dernier défaut est l'agrégation autour des centres mobiles dont le principe général est une réallocation dynamique des objets à des centres de classes, le nombre de classes  $k$  étant fixé a priori par l'utilisateur. Plusieurs algorithmes légèrement différents sont possibles pour cette méthode dont le plus classique est celui de Forgy :

- La phase d'initialisation consiste à tirer au hasard ou sélectionner, pour des raisons extérieures à la méthode,  $k$  points dans l'espace des objets à classer, en général  $k$  objets appartenant à l'ensemble, appelés centres ou noyaux.
- La phase d'itération consiste à répéter, jusqu'à stagnation du critère de variance inter-classes les deux étapes suivantes : 1) allouer chaque objet à l'un des noyaux (sur la base de sa distance aux noyaux), c'est à dire à une classe et 2) calculer le centre de gravité de chaque classe qui devient le nouveau noyau

On voit ici que si l'initialisation est réalisée par un choix au hasard des  $k$  points, le résultat dépendra de ce choix initial. On a donc l'habitude dans ce cas de répéter ce tirage aléatoire puis l'algorithme un grand nombre de fois dans le but de repérer ce que l'on appelle les « formes fortes », c'est à dire les points qui sont toujours regroupés ensemble (ou bien de sélectionner la classification qui minimise la somme des variances intra-classes par exemple). L'algorithme des k-means est similaire mais le calcul des nouveaux noyaux est effectué à chaque allocation d'un objet à une classe. D'autres approches, comme la méthode pam (partitioning around medoids ou la méthode clara plus adaptée aux grands jeux de données) ou celle des nuées dynamiques, consistent à rechercher  $k$  objets représentatifs (les medoïds) parmi les observations qui doivent représenter la structure des données puis à assigner les objets à ces medoïds. L'objectif de pam est par exemple d'identifier  $k$  objets représentatifs qui minimisent la somme des dissimilarités entre les objets et leur medoïd le plus proche. Cette méthode a de plus l'intérêt de fournir des sorties graphiques (les graphiques des silhouettes) et numériques (les silhouettes moyennes) aidant respectivement à l'interprétation du contenu des classes (détermination de la qualité de l'allocation d'un objet à une classe) et à fixer le nombre de classes (choix de la classification qui a la meilleure silhouette moyenne). Le principal avantage de ces différentes méthodes d'agrégation autour des centres mobiles est leur convergence généralement rapide (et donc leur applicabilité à de très grands jeux de données) et leur principal inconvénient est la nécessité de définir a priori le nombre de classes. Classification ascendante hiérarchique et méthodes d'agrégation autour des centres mobiles peuvent par ailleurs être combinées comme illustré dans l'article présenté en [annexe 4](#), ce qui palie en partie aux principaux défauts de chacune des méthodes (« rigidité » de la classification ascendante hiérarchique et choix arbitraire du nombre de classes pour l'agglomération autour des centres mobiles). Il existe de nombreuses autres méthodes de classification répondant parfois à des objectifs précis et il est donc nécessaire là encore que l'utilisateur définisse au mieux sa question et ses objectifs afin de trouver un conseil approprié auprès de spécialistes de ces méthodes et/ou dans la littérature.

La deuxième catégorie de méthodes exploratoires multidimensionnelles est constituée par les **méthodes factorielles** dont l'analyse en composantes principales (ACP) est la plus connue. L'objectif de l'ACP est de représenter graphiquement plus de 3 variables par la recherche de résumés (nuages de points) pertinents dans le plan respectant les distances entre les individus et les corrélations entre les variables. Concrètement, l'ACP va fournir 1) une représentation graphique « optimale » des individus (ici « optimale » signifie que le nuage de point présente un minimum de déformation ou plus précisément que les axes successifs capture un maximum de sa variance initiale, également appelée inertie totale) dans un sous-espace  $E_q$  de dimension  $q$  (où  $q$  est inférieure à la dimension initiale des données qui est fournie par le nombre de variables) et 2) une représentation graphique des variables dans un sous-espace  $F_q$  explicitant au mieux les liaisons initiales (i.e. les corrélations si les variables sont centrées) entre ces variables. Lorsque les variables sont centrées et réduites, elles sont contenues, dans les représentations graphiques de l'ACP, dans un cercle de rayon 1 appelé le cercle des corrélations. Lorsque les variables sont proches de ce cercle (plus elles en sont proches et mieux elles sont projetées dans le plan considéré), le cosinus de l'angle formé par les vecteurs correspondant à deux variables est environ égal à la corrélation entre ces deux variables. Signalons qu'il est également possible de réaliser des représentations simultanées (souvent appelées biplots) des variables et des individus sur un même graphique moyennant un changement d'échelle sur les axes (pour un exemple de mise en œuvre, voir Chapman et al., 2002) mais dans le contexte des puces à ADN, le nombre important de points à représenter n'est pas toujours favorable à ce type de représentation graphique. L'ACP effectue donc une réduction de dimension (ou une projection du nuage de points sur un espace de dimension réduite) en réalisant une approximation de la matrice de données  $X$  par un tableau de rang  $q$ . Sous cette formulation, l'ACP peut être considéré comme une modélisation des observations de départ, l'estimation des paramètres du modèle étant généralement réalisée par le biais d'une décomposition en valeurs singulières. Globalement, cela consiste à rechercher  $q$  combinaisons linéaires des variables de départ qui soient non corrélées entre elles (pour que les axes de l'ACP appelés composantes principales soient orthogonaux) et dont la somme des variances soit maximale (pour que ces axes capturent les principaux axes d'allongement du nuage de points). En plus de ses représentations graphiques, l'ACP fournit un grand nombre (souvent trop si elles sont toutes calculées...) de mesures de qualité des représentations ou de contributions qui peuvent aider à l'interprétation ou être utilisées pour simplifier ou enrichir les représentations graphiques. L'un des problèmes récurrent en ACP est le choix du nombre de dimension. Plusieurs solutions ont été proposées mais la plus utilisée est sans doute la

recherche d'un « coude » (ou d'une « cassure de pente ») dans l'éboulis des valeurs propres issues de la décomposition aux valeurs singulières (et qui sont liées au % de variance expliquée par chaque axe de l'ACP). La réalisation de boxplots (ou boîtes à moustache) des composantes principales est également un outil pratique qui peut également aider l'identification de points extrêmes. D'après notre expérience de l'utilisation de l'ACP, que nous trouvons particulièrement adaptée à l'exploration des données issues de puces à ADN dédiées, nous constatons que plus le plan expérimental est élaboré (nombre de facteurs et de modalités de ces facteurs notamment mais également plan d'expérience en cinétique), plus il faudra étudier d'axes en ACP car chacun a une chance de capturer une part d'information relative à un aspect du plan d'expérience. Typiquement, dans un plan d'expérience comparant deux conditions, l'ACP ne trouvera son utilité que dans la recherche de points aberrants. Les principales limites de l'ACP sont liées au fait que c'est une technique linéaire (les axes sont des combinaisons linéaires des variables initiales) qui optimise un critère quadratique (l'inertie ou variance) qui ne tient donc pas compte de liaisons non linéaires et présente une forte sensibilité aux valeurs extrêmes (ce qui peut néanmoins être aussi un avantage lorsque l'ACP est justement utilisée pour repérer ces valeurs extrêmes). Dans le contexte de puces pangénomiques, le nombre de variable génère souvent des graphiques illisibles mais l'ACP peut cependant malgré tout trouver son utilité, soit via une sélection de variables préalables, soit via la représentation de variables synthétiques représentant des groupes de gènes issus d'une classification ou encore regroupés par catégories fonctionnelles sur la base de leurs annotations. Des exemples de mise en œuvre et d'interprétation d'ACP peuvent être trouvés dans les publications relatives à ce travail de doctorat (Martin et al., 2007; Martin et al., 2005) et présentées aux chapitres IV et V. Nous évoquerons dans le chapitre suivant une autre méthode factorielle : l'analyse factorielle discriminante (AFD ou analyse linéaire discriminante de Fisher). Citons également ici le positionnement multidimensionnel (MDS pour multidimensional scaling) qui consiste à réaliser une ACP sur un tableau de distances, fournissant ainsi une représentation graphique des objets de dimension réduite (typiquement un plan) qui constitue la meilleure approximation des distances observées entre les objets. Son principal avantage est qu'il permet la comparaison de différentes distances, fournissant ainsi à l'utilisateur des points de vue différents et complémentaires sur son jeu de données.

Il existe ainsi une variété de solutions pour l'exploration des jeux de données multidimensionnels qui trouvent chacun leur pertinence dans l'étude des données de puces à ADN dans un objectif de découverte de classes. Même lorsque la question initiale n'est pas de découvrir des classes au sein des objets analysés, je pense que ces outils sont utiles à la simple

exploration des jeux de données obtenus (identification de points extrêmes ou des principales tendances, hiérarchisation des principaux effets observés sur le profil d'expression des gènes, etc.). Leur interprétation n'est pas toujours simple mais l'habitude et une bonne connaissance du système étudié sont sans doute dans ce cadre les meilleures armes, soulignant ainsi l'avantage qu'il y a à utiliser de manière presque systématique dans le cadre d'une étude préliminaire des données au moins certains de ces outils (comme la CAH et l'ACP par exemple) afin d'en acquérir la maîtrise.

### **III.4.5. Prédiction de classes**

La prédiction de classe vise à répondre à des questions du type « Peut-on définir une règle permettant de séparer (ou discriminer) au mieux mes groupes d'échantillons (ou de gènes) et qui présente un taux d'erreur minimal lorsqu'elle est appliquée pour prédire la classe de nouveaux échantillons ? ». Nous nous limiterons dans ce chapitre à citer les principales méthodes disponibles sans les décrire pour la plupart (pour des descriptions plus complètes et des comparaisons de méthodes, voir Parmigiani, 2003, pp29-33 et chapitres cités dans ces pages; Simon, 2003, pp 95-119; Speed, 2003, pp 93-158) car aucune d'entre elles n'a réellement été utilisée dans un contexte de prédiction de classe dans le cadre de ce travail de doctorat et notre expérience de ces méthodes est donc particulièrement limitée. Signalons tout de suite qu'en anglais les méthodes de discrimination ou de prédiction que nous allons évoquer sont regroupées sous le nom de *classification* (terme qui, en français, se réfère plutôt aux méthodes non supervisées présentées dans le paragraphe précédent). Ces méthodes peuvent être par exemple utilisées pour prédire la fonction de gènes (on établit une règle de discrimination des gènes dont la fonction est connue sur la base de leurs profils d'expression et on l'applique ensuite à des gènes dont la fonction n'est pas connue), pour prédire un type de tumeur (dans un but diagnostique comme dans Golub et al., 1999; pour prédire le tissu d'origine comme dans Ross et al., 2000) ou établir un pronostique d'évolution (prédiction du pouvoir métastatique d'une tumeur, du risque de rejet d'une greffe, d'une espérance de survie, etc.) ou encore pour rechercher des ensembles de biomarqueurs (gènes discriminants au mieux une situation exposée d'une situation non exposée utilisés comme biomarqueur d'exposition par exemple). La formulation du problème de prédiction de classes tel que nous le présentons ici permet de concevoir immédiatement les deux étapes essentielles des différentes méthodes : 1) la construction d'une règle de discrimination (classifier en anglais) et 2) son évaluation sur de nouveaux échantillons. Contrairement aux méthodes que nous avons présenté pour la découverte de classe, les méthodes utilisées ici sont dites supervisées,

dans la mesure où la construction, comme la validation de la règle de discrimination, nécessitent la connaissance a priori des classes auxquelles les objets appartiennent. Idéalement, il faut donc deux jeux de données indépendants pour répondre à ce type de question : l'un est appelé le jeu d'apprentissage (training dataset) et sert à construire le prédicteur, alors que l'autre constitue le jeu de données de validation (test dataset) et sert à évaluer le taux d'erreur du prédicteur précédemment défini. Une alternative consiste à utiliser la validation croisée dont une variante consiste à séparer le jeu de données initial en K sous-parties et à construire K prédicteurs à partir du jeu de données dont on a éliminé l'une des K sous-partie à chaque fois, le taux moyen d'erreur de la méthode de discrimination calculé à partir des K analyses étant une estimation non biaisée du taux réel d'erreur. Dans le cas où le nombre de prédicteurs est grand devant le nombre de cas (ce qui est typiquement le cas dans la prédiction de classes d'échantillons à partir de leurs données d'expression génique), si l'estimation de la précision de la règle de discrimination est réalisée sur le jeu de données qui a servi à construire cette règle (le jeu d'apprentissage), cette précision sera toujours excellente, mais la règle fonctionnera mal sur de nouvelles données. Ce phénomène qualifié de sur-apprentissage (overfitting) est lié au fait qu'un trop grand nombre de paramètres sont ajustés sur trop peu de points de mesures, conduisant à un ajustement sur du bruit aléatoire contenu dans les données (autrement dit, on trouvera toujours quelques gènes qui discrimineront par hasard de manière parfaite les différentes classes, si on mesure suffisamment de gènes et que l'on a peu d'échantillons). Il ressort de cette remarque qu'une étape préliminaire à la construction d'une règle de discrimination doit généralement consister à réduire la dimension des données. Ceci peut être principalement réalisé par deux types de méthodes : 1) par la recherche de « gènes synthétiques » ou « variables latentes » en utilisant par exemple les axes fournis par des méthodes factorielles comme l'ACP ou l'AFD ou encore des noyaux de classes obtenus par une méthode d'agrégation autour des centres mobiles ou 2) par une sélection d'une sous-liste de gènes vraisemblablement discriminants (par exemple en utilisant des approches de type comparaison de classes). Après avoir globalement présenté les principaux enjeux et les grandes étapes impliquées dans la prédiction de classes, citons quelques méthodes utilisées dans ce contexte.

**L'analyse factorielle discriminante** (AFD) est basée, comme l'ACP, sur une projection des nuages de points sur des espaces de dimension réduite (c'est donc une méthode factorielle). Mais, alors que l'ACP optimise la part d'inertie totale capturée par chaque axe, l'AFD optimise, elle, le rapport entre l'inertie inter-groupes et l'inertie intra-groupe, conduisant ainsi à une séparation optimale des groupes dans les premiers axes factoriels. Les

axes étant des combinaisons linéaires des variables initiales, il est possible, à partir des coefficients de cette combinaison linéaire, d'essayer d'interpréter les axes du point de vue biologique. Alors que l'AFD est clairement une méthode de prédiction, elle peut néanmoins également être utile à titre exploratoire pour identifier des groupes de gènes discriminant bien deux (ou plus) conditions expérimentales. D'autres méthodes peuvent également être utilisées pour la prédiction comme le **modèle linéaire généralisé** (ou plus précisément le modèle de discrimination logistique) qui consiste globalement à modéliser le logarithme du rapport des probabilités d'appartenance d'un individu aux différentes classes (le logodds) et à introduire pas-à-pas les prédicteurs par un processus de sélection de modèles. L'un des problèmes de cette approche est le recours à une sélection pas-à-pas qui n'est donc pas exhaustive et ne peut garantir l'obtention d'une règle optimale. Une autre méthode dont le principe est relativement simple est la méthode des **plus proches voisins** (nearest neighbor classification) où un nouvel échantillon est classé en se basant sur les classes auxquelles appartiennent les échantillons du jeu d'apprentissage dont les profils d'expression sont les plus similaires (ou les plus voisins) au nouvel échantillon. Il est nécessaire pour cette méthode de définir une mesure de distance entre profils et un nombre de voisins (méthode des  $k$  plus proches voisins ou  $k$ -nearest neighbor) à considérer ( $k$  devant être impairs pour éviter les égalités entre deux classes prédites). Une autre méthode assez similaire mais utilisant une contraction (shrinkage) des mesures d'expression de centroïdes des classes est implementée dans la méthode **PAM** (Prediction Analysis for Microarrays, Tibshirani et al., 2002). Elle permet notamment de diminuer le nombre de gènes impliqués dans la règle de discrimination et d'opérer ainsi une étape de sélection de variables. Les méthodes basées sur les **arbres de classification** (Classification and Regression Trees) comme les **forêts aléatoires** (random forest) semblent bien adaptées à la gestion d'une grande quantité de variables et fournissent également des sorties numériques et/ou graphiques (les mesures « d'importance » des variables) pouvant permettre une sélection de variables les plus discriminantes. Enfin, les **machines à vecteurs supports** (ou séparateurs à vastes marges, en anglais : SVM pour support vector machines), basées sur des algorithmes de discrimination, ont également été utilisées avec succès, notamment pour la prédiction de la fonction de gènes (Brown et al., 2000; Zhang et al., 2004).

Là encore de nombreux outils sont disponibles qui possèdent chacun leurs avantages, leurs inconvénients et leur capacité à répondre à différentes questions et objectifs. De nombreux logiciels implémentent ces différents outils, mais citons plus particulièrement le package MLInterfaces de Bioconductor (voir [www.bioconductor.org](http://www.bioconductor.org)) qui implémente de

nombreuses méthodes différentes et harmonise dans la mesure du possible leurs sorties numériques.

### III.4.6. Aide à l'interprétation / aspects bioinformatiques

Alors que l'utilité des outils bioinformatiques est évidente dans l'étape de conception des puces à ADN (prédition de gènes à partir des séquences génomiques et sélection de sondes spécifiques notamment), leur utilité est au moins aussi importante dans la phase d'analyse ou plutôt d'interprétation des données. La diversité des questions posées par les biologistes à partir de leurs expériences de puces à ADN rend difficile la définition d'une typologie des problèmes abordés par les multiples outils bioinformatiques mais on peut malgré tout citer à titre d'exemple : **l'analyse des promoteurs** dans le but de comprendre la co-régulation de l'expression de gènes, **l'analyse de voies** (Pathway analysis en anglais) cherchant à comprendre les conséquences des régulations à l'échelle des grandes voies biologiques (voies métaboliques ou de signalisation) et plus généralement **l'intégration de connaissances** (bibliographiques par exemple) et d'annotations sur les gènes pour guider l'interprétation des résultats issus des puces à ADN, **l'inférence de réseaux** dans le but de comprendre les interactions entre les gènes, ou encore la **réconciliation de données** visant à comprendre les liens entre des observations réalisées à différents niveaux (transcriptome/protéome, transcriptome/métabolome par exemple). Nous ne pourrons explorer ici tous ces aspects et nous concentrerons donc sur quelques exemples que nous avons eu l'occasion de rencontrer (mais pas toujours de manipuler) dans le cadre de nos travaux.

La suite Genomatix ([www.genomatix.de](http://www.genomatix.de)) est une suite de logiciel dédiée à l'analyse de promoteurs qui offre la possibilité intéressante de rechercher des groupes de motifs de fixation de facteurs de transcription au lieu de sites de fixation isolés. Ces outils ou encore ceux disponibles dans le domaine public comme TransFac ([www.gene-regulation.com](http://www.gene-regulation.com)) s'avèrent intéressant pour comprendre la corégulation des gènes via l'analyse de leurs promoteurs.

Ce que je regroupe dans « l'intégration de connaissance » constitue un enjeu majeur de l'analyse des données issues des puces à ADN. Ces expériences fournissent classiquement à l'utilisateur des listes de dizaines voire de centaines de gènes dont l'analyse bibliographique individuelle est le plus souvent impossible. Des outils peuvent donc être utilisés pour extraire de ces listes des informations plus globales comme les principales fonctions biologiques représentées ou encore les localisations subcellulaires des produits des transcrits mesurés. L'une des bases de données les plus utilisée dans ce contexte est **Gene Ontology** (GO, voir

[www.geneontology.org](http://www.geneontology.org)) qui est basée sur la définition d'un vocabulaire associé à une structure hiérarchique (les liens de hiérarchie entre les termes utilisés) utilisé ensuite pour annoter les gènes (ou plutôt leurs produits issus de leur transcription/traduction) via trois points de vue : leurs localisations (cellular component), leurs fonctions moléculaires (molecular function) et leurs rôles biologiques (biological processes). Plusieurs outils ont ainsi été développés (Al-Shahrour et al., 2004; Beissbarth and Speed, 2004; Manoli et al., 2006) pour déterminer la sur- ou la sous-représentation de certaines catégories issues de Gene Ontology dans des listes de gènes en les comparant généralement à des données prises comme référence (tous les gènes annotés dans GO ou encore tous les gènes présents sur la puce utilisée). La base de données **KEGG** (Kyoto Encyclopedia of Genes and Genomes, [www.genome.jp/kegg](http://www.genome.jp/kegg)), orientée sur un regroupement des gènes en voies (métaboliques, de signalisation, processus cellulaires, etc.) est également très utilisée. La méthode du **Gene Set Enrichment Analysis** (GSEA, Mootha et al., 2003b; Subramanian et al., 2005) peut utiliser différentes sources de données pour définir des groupes de gènes a priori dont l'analyse GSEA proprement dite recherche ensuite un enrichissement éventuel parmi les gènes régulés par le biais d'un score d'enrichissement basé sur la comparaison de la distribution des rangs des gènes appartenant au groupe prédéfini par rapport à ceux des autres gènes. Cette analyse, qui semble fournir des résultats intéressants, surtout dans les cas où les modulations d'expression sont relativement faibles dans leur amplitude et où la variabilité des échantillons est grande et peu contrôlable, a stimuler d'autres recherches visant à intégrer des connaissances biologiques a priori dans l'identification de groupes de gènes régulés (Tian et al., 2005). La base de données **Ingenuity** ([www.ingenuity.com](http://www.ingenuity.com)), basée elle sur une annotation manuelle des gènes réalisée par des experts à partir de la bibliographie, est accompagnée d'une suite d'outils qui facilitent grandement l'analyse des puces à ADN pangénomiques et l'identification de réseaux ou de voies de gènes corégulés. Enfin, de nombreux outils pour la manipulation et l'analyse des annotations de gènes sont également disponibles sous le logiciel R et ont été développés dans le cadre du projet Bioconductor ([www.bioconductor.org](http://www.bioconductor.org)).

Des travaux importants sont également conduits dans le domaine de la modélisation des systèmes biologiques via l'inférence et l'analyse de réseaux. Etant particulièrement novice dans ce domaine, je me contenterais de signaler l'existence d'un outil accessible librement pour l'analyse et la visualisation de réseaux d'interactions moléculaires nommé Cytoscape ([www.cytoscape.org](http://www.cytoscape.org)).

Enfin, alors que de nombreux autres domaines de la biologie génèrent des jeux de donnée de plus en plus volumineux, des outils sont nécessaires à la **réconciliation** de ces

différentes sources d'information afin d'établir des liens entre les niveaux d'observation. Plusieurs méthodes statistiques existent pour répondre à ce type de problème comme l'analyse canonique ou encore la régression PLS (Partial Least Squares). Dans ce domaine, j'ai eu l'occasion de participer à un projet en collaboration avec Ignacio González, Sébastien Déjean, Olivier Gonçalvez, Alain Baccini et Philippe Besse du Laboratoire de Statistique et Probabilité (Institut de Mathématique de Toulouse) qui consistait à développer, en particulier pour l'analyse des données issues de nos puces à ADN (Martin et al., 2007), une méthode d'analyse canonique pouvant fonctionner sur des jeux de données contenant plus de variables que d'individus (deux articles actuellement soumis sur l'analyse canonique régularisée dont un est présenté en [annexe 3](#)). L'utilisation de l'analyse canonique est illustrée au paragraphe III.5. et la mise en œuvre de l'analyse canonique régularisée est présentée en [annexe 3](#).

Comme on le voit, le domaine de l'étude du transcriptome est particulièrement riche en questionnements et s'enrichit fortement d'autres domaines scientifiques comme ceux de la statistique et de la bioinformatique. En contrepartie, les caractéristiques des données et des questionnements générés stimulent fortement une recherche de qualité en mathématique et informatique appliquées. Ce contexte a constitué pour moi une opportunité majeure de me situer à l'interface entre biologistes et statisticiens dans le cadre du début de ma carrière d'ingénieur, me permettant ainsi d'enrichir mon point de vue sur nos objets de recherche et la façon de les aborder.

### **III.4.7. Analyse des puces INRArry 01**

Après avoir présenté les principaux outils de l'analyse des données issues des puces à ADN, voyons à présent les principales étapes utilisées dans le traitement des puces INRArry 01. Les étapes présentées ci-dessous ont fait l'objet d'une mise au point progressive et une procédure complète a été mise au point et associée à la création de fonctions sous le logiciel d'analyse R dans le cadre du stage de Master 1 d'Alice Vigneron (2006) que j'ai encadré. Ces fonctions R ainsi qu'un fichier d'aide détaillant la procédure de traitement et présentant un exemple issu de nos travaux sur les phthalates est disponible à l'adresse suivante : [www.inra.fr/internet/Centres/toulouse/pharmacologie/pharmaco-moleculaire/technologie/data-functions.html](http://www.inra.fr/internet/Centres/toulouse/pharmacologie/pharmaco-moleculaire/technologie/data-functions.html).

Nous ne présenterons donc ici que les grandes étapes du traitement de ces données en nous concentrant sur les étapes de prétraitement, dans la mesure où les étapes d'analyse dépendent de la question posée et que les outils utilisés pour y répondre sont ceux présentés dans les trois paragraphes précédents.

- Les données sont acquises sur un imageur au phosphore (scanner STORM 840, Amersham Biosciences, France) après un à trois jours d'exposition des puces à ADN à un écran au phosphore.
- L'analyse d'image est réalisée avec le logiciel Image Master Array 2.0 (Amersham Biosciences, France) et les données d'intensité (volume = intensité moyenne des pixels du spot x aire du spot = somme des intensités des pixels du spot) pour chaque puce sont extraites sous la forme d'un fichier Excel. Une correction de bruit de fond local (médiane des intensités des pixels adjacents aux spots) n'est appliquée aux données que si les images en démontrent la nécessité (bruit de fond inhomogène, présence de tâches). L'analyse est parfois conduite en parallèle sur données corrigées et sur données non corrigées en cas de doute. Les données des différentes puces sont combinées sous la forme d'une matrice gène x échantillons dont la dernière colonne spécifie le type de spot (gène, bruit de fond ou spike). Les données sont ensuite importées sous le logiciel R.
- Les puces INRArray 01 (à partir de la version INRArray 01.2) portent autant de spots ne contenant que du tampon de dépôt (50% DMSO) que de spots spécifiques de gènes d'intérêt. Les spots ne contenant que du tampon de dépôt (spots « DMSO ») sont utilisés pour évaluer la distribution du bruit de fond. Les données « DMSO » sont extraites du jeu de données, transformées en log base 2 et un filtre (on élimine les valeurs dépassant la moyenne du bruit de fond + 1,5 fois l'espace interquartiles) est appliqué pour éliminer les valeurs aberrantes (liées à un défaut de lavage des aiguilles du spotteur). La normalité de la distribution du bruit de fond est évaluée via un test de Shapiro-Wilk, éventuellement corrigé par la méthode de Benjamini-Hochberg. Les boxplots du bruit de fond de chaque puce sont également observés. Généralement, le bruit de fond est distribué de manière gaussienne sur plus de 95% des puces, ce qui conduit à accepter l'hypothèse de normalité du bruit de fond.
- La valeur seuil de la moyenne du bruit de fond + 2 fois son écart-type est utilisée sur chaque puce pour déterminer si les intensités des spots dépassent ou non significativement le bruit de fond (ce qui correspond globalement à un test de Student unilatéral à 2,5%). Une matrice de présence-absence est produite (valeur 1 si le spot dépasse la valeur seuil basée sur le bruit de fond et 0 sinon). En nous basant sur une série de représentations graphiques (distribution du nombre de gènes déclarés présents sur les puces notamment), sur la gestion de la multiplicité des tests réalisés (pour

chaque gène on effectue autant de tests qu'il y a de puces dans l'expérience), sur la taille des groupes expérimentaux (certains gènes peuvent n'être exprimés que dans un seul groupe d'individus) et sur des comparaisons de listes de gènes en fonction du filtre appliqué, on détermine le nombre de puces minimum sur lequel un spot doit être déclaré présent pour qu'il soit conservé dans les futures analyses. La présence d'un spike contrôle négatif (ARN non ajouté dans les échantillons) permet de contrôler que le seuil fixé n'est pas trop bas (si le spike est conservé, alors le seuil est trop bas).

- La normalisation est réalisée par trois méthodes différentes : la moyenne globale par puce (spots correspondant à des gènes et à des spikes uniquement), la médiane globale par puce (idem) et la moyenne des signaux des spikes (transformés en log base 2). Ces trois normalisations sont comparées, principalement sur la base de représentations en boxplots et une normalisation est choisie. Par défaut, c'est la normalisation par les spikes qui est choisie car nous avons vu que les hypothèses qui la sous-tendent ont plus de chance d'être vérifiées dans nos expériences. Dans de nombreux cas, les normalisations par les spikes et par le signal moyen donnent néanmoins des résultats proches.
- Dans tous les cas, des analyses en composantes principales et des classifications des gènes et des individus sont réalisées et comparées entre au moins deux types de normalisations comprenant la normalisation par les spikes. Pour chaque groupe expérimental, des graphiques sont produits qui comparent les signaux des puces deux-à-deux dans le but 1) de repérer des puces aberrantes (réalisé en parallèle via les boxplots et les analyses multidimensionnelles) et 2) de repérer des problèmes de quantification (intensité de certains spots aberrants). Une liste des spots présentant des valeurs bizarres est réalisée à partir de l'observation de ces graphiques et un retour vers l'analyse d'image est effectué pour contrôler spécifiquement la quantification des spots incriminés. Dans le cas où des modifications sont apportées à la quantification, le jeu de données est alors réanalysé en partant des premières étapes. De manière générale, nous constatons que deux allers-retours entre analyse préliminaire (telle que présentée ici) et analyse d'image sont nécessaires pour s'assurer de l'obtention d'un jeu de données fiable.
- A ce stade, le jeu de données peut être analysé pour répondre aux questions biologiques spécifiques.

En conclusion de ces sept paragraphes, il est relativement étonnant de constater la vitesse à laquelle le domaine de l'analyse des puces à ADN se saisit des nouveaux outils générés par la recherche en statistique et en bioinformatique. La gestion de la multiplicité des tests statistique en est un bon exemple car elle a connu un renouveau en 1995 avec l'introduction de la notion de FDR (Benjamini and Hochberg, 1995), de manière concomitante à la diffusion des puces à ADN. Très rapidement, ces notions ont été appliquées aux données de puces à ADN et de nouvelles méthodes ont été proposées spécifiquement pour ce domaine (voir par exemple Tusher et al., 2001), ce qui illustre bien l'intérêt et la qualité des relations entre biologistes moléculaires et statisticiens dans le cadre des projets de transcriptomique. En contrepartie, les puces à ADN et les nombreuses nouvelles applications qui en sont faites (ChIP on chip, tilling arrays, puces SNP, puces de tissus, etc.) stimulent une recherche en statistique appliquée diversifiée et de qualité. Le paragraphe suivant est constitué d'une synthèse, publiée dans un numéro spécial du Journal de la Société Française de Statistique (Baccini et al., 2005), qui a été réalisée en collaboration avec une équipe du Laboratoire de Statistique et Probabilités appartenant à l'Institut de Mathématiques de Toulouse (Sébastien Déjean, Alain Baccini et Philippe Besse) et avec des collègues statisticiens de l'INRA de Toulouse (Christelle Robert-Granié et Magali San Cristobal). Elle illustre la mise en œuvre de différentes méthodes sur le jeu de données issu des puces INRArray 01.2 interprété du point de vue biologique dans (Martin et al., 2007) et présenté au chapitre V.

### **III.5. MISE EN ŒUVRE DE DIVERSES METHODES SUR UN MEME JEU DE DONNEES**

**Article n°1 :** Alain Baccini, Philippe Besse, Sébastien Déjean, Pascal G. P. Martin, Christelle Robert-Granié, Magali San Cristobal. Stratégies pour l'analyse statistique de données transcriptomiques. *Journal de la Société Française de Statistique*. 2005 ; 146(1-2): 5-44.



# STRATÉGIES POUR L'ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

Alain BACCINI\*, Philippe BESSE\*, Sébastien DÉJEAN\*,  
Pascal G.P. MARTIN \*\*, Christèle ROBERT-GRANIÉ \*\*\*,  
Magali SAN CRISTOBAL \*\*\*\*

## RÉSUMÉ

Afin d'illustrer la diversité des stratégies applicables à l'analyse de données transcriptomiques, nous mettons d'abord en œuvre des méthodes issues de la statistique exploratoire (ACP, positionnement multidimensionnel, classification), de la modélisation (analyse de variance, modèles mixtes, tests) ou de l'apprentissage (forêts aléatoires), sur un jeu de données provenant d'une étude de nutrition chez la souris. Dans un second temps, les résultats obtenus sont mis en relation avec des paramètres cliniques mesurés sur les mêmes animaux, en utilisant cette fois l'analyse canonique. La plupart des méthodes fournissent des résultats biologiquement pertinents sur ces données.

De cette expérience, nous tirons quelques enseignements élémentaires : il n'y a pas, *a priori*, de meilleure approche ; il faut trouver la « bonne » stratégie associant exploration et modélisation, adaptée tant aux données qu'à l'objectif recherché. Dans cette optique, une collaboration étroite entre statisticien et biologiste est indispensable.

*Mots clés :* biopuces, analyse en composantes principales, classification, modèle linéaire, forêts aléatoires, analyse canonique.

## ABSTRACT

In order to illustrate the variety of strategies applicable to transcriptomic data analysis, we first implement methods of exploratory statistics (PCA, multidimensional scaling, clustering), modelling (ANOVA, mixed models, tests) or learning (random forests), on a dataset coming from a nutrition study for mice. In a second stage, relationships between the previous results and clinical measures are studied through canonical correlation analysis. Most of the methods provide biological relevant results on these data.

---

\* Laboratoire de Statistique et Probabilités - UMR CNRS 5583, Université Paul Sabatier - 31062 Toulouse Cedex 9.  
{alain.baccini,philippe.besse,sebastien.dejean}@math.ups-tlse.fr

\*\* Laboratoire de Pharmacologie et Toxicologie - UR 66 - INRA, 180, chemin de Tournefeuille - B.P. 3 - 31931 Toulouse Cedex.

\*\*\* Station d'Amélioration Génétique des Animaux - INRA

\*\*\*\* Laboratoire de Génétique Cellulaire - INRA, Auzetville B.P. 27 - 31326 Castanet-Tolosan Cedex.  
{pmartin,robert,msc}@toulouse.inra.fr

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

From this experience we conclude that there is not one best approach; we have to find the “good” strategy combining exploration and modelling to fit the data as well as to achieve the biological purpose. From this point of view, a strong collaboration between statistician and biologist is essential.

*Keywords :* microarray, principal component analysis, clustering, linear model, random forests, canonical correlation analysis.

### 1. Introduction

Les données transcriptomiques, ou données d'*expression des gènes*, sont issues de technologies variées (*Polymerase Chain Reaction* – PCR – en temps réel, biopuces constituées d'acide désoxyribonucléique – ADN – complémentaire ou d'oligonucléotides déposés sur membrane de nylon ou sur lame de verre), chacune nécessitant des prétraitements spécifiques : analyse d'image, quantification et normalisation. Ces derniers ne sont pas abordés dans cet article, même si, déterminants dans beaucoup de résultats, ils doivent être envisagés avec prudence et évalués en termes d'impact sur les conclusions retenues.

Les problématiques rencontrées, les conditions expérimentales, ainsi que les objectifs visés varient considérablement : présélection de gènes, prise en compte d'une cinétique, importance de covariables, présence de variables phénotypiques... Néanmoins, la spécificité majeure de ces données, celle qui remet le plus en cause le bon usage et le savoir faire du statisticien, est la très haute dimensionnalité du nombre de gènes par rapport au nombre comparativement très restreint d'échantillons biologiques sur lesquels leur expression est mesurée. De façon formelle, le problème se pose comme l'observation d'une variable, l'*expression* (ou quantité d'acide ribonucléique messager – ARNm – produite), dans des situations expérimentales croisant au moins deux facteurs : le gène et le type d'échantillon biologique (tissus sain ou pathologique, cellule sauvage ou modifiée...). Le premier facteur peut présenter de quelques centaines à quelques dizaines de milliers de niveaux, tandis que le second, pour des raisons évidentes de coût, ne présente en général que quelques dizaines de niveaux au maximum. L'objet de l'analyse statistique est alors d'extraire une information pertinente concernant l'effet de différents facteurs sur l'état fonctionnel de la cellule.

L'objectif de cet article est de montrer, sur un exemple réel, quels sont, dans la vaste panoplie disponible, les outils qui se montrent les plus efficaces pour apporter des solutions pertinentes aux questions du biologiste. Nous verrons que chaque outil et, plus précisément, chaque option (pondération, distance, transformation...) apporte un point de vue différent sur les données. Le défi consiste alors à conduire ces choix en conscience, compte tenu de leurs limites et de l'interprétation biologique qui en est faite, afin de déterminer la technique, ou l'optique, qui offre la représentation la plus éclairante au biologiste. Ainsi, nous apporterons, pour les différentes techniques, quelques commentaires sur l'interprétation biologique des résultats; ceux-ci feront l'objet d'une publication ultérieure en cours de préparation (Martin *et al.*, 2005b).

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

Dans le paragraphe suivant, nous présentons le jeu de données à l'origine de cet article ainsi que la problématique biologique. Le troisième paragraphe est consacré à la démarche exploratoire se focalisant sur l'analyse d'un tableau de données pour lequel on cherche des représentations graphiques pertinentes. Dans le paragraphe 4, nous présentons diverses méthodes de modélisation qui ont permis de confirmer et de compléter la connaissance sur l'expression des gènes acquise avec la démarche exploratoire. L'un des objectifs recherchés, à travers l'expérimentation biologique mise en œuvre, étant de mettre en relation l'expression de certains gènes avec les quantités d'acides gras hépatiques, le paragraphe 5 précise comment l'analyse canonique permet de répondre à cette question. Enfin, conclusions et perspectives sont présentées dans le dernier paragraphe.

## 2. Les données

Les données ont été fournies par l'équipe de pharmacologie moléculaire de l'Unité de Recherche 66 de l'INRA de Toulouse. Elles proviennent d'une étude de nutrition chez la souris. Pour 40 souris, nous disposons :

- des données d'expression de 120 gènes sélectionnés parmi les 30000 de la souris comme étant susceptibles d'être régulés par les conditions nutritionnelles de cette étude; elles ont été recueillies sur membrane nylon avec marquage radioactif;
- des concentrations de 21 acides gras hépatiques obtenues par chromatographie en phase gazeuse au laboratoire de Biochimie ENSAR-INRA de Rennes (pour chaque souris, la somme des 21 mesures est égale à 100).

Les 120 gènes analysés constituent une puce dédiée (*INRArray 01.2*). Ces gènes ont été sélectionnés, d'une part, sur la base d'une étude bibliographique, d'autre part, à partir d'une étude préliminaire réalisée sur 10000 gènes dont l'expression a été mesurée par *microarray* dans des conditions susceptibles de mettre en évidence les gènes cibles de PPAR $\alpha$  (Martin *et al.*, 2005a). Cette biopuce, en développement permanent, compte aujourd'hui (juin 2005) environ 300 gènes.

Par ailleurs, les 40 souris sont réparties selon deux facteurs croisés dans un plan complet, équilibré, à quatre répétitions. Les deux facteurs sont décrits ci-dessous.

- Génotype (2 niveaux). Les souris sont soit de type sauvage (WT) soit de type PPAR $\alpha$ -déficientes (PPAR), avec 20 souris dans chaque cas (Lee *et al.*, 1995).
- Régime (5 niveaux). Les cinq régimes alimentaires sont notés :
  - **dha** : régime enrichi en acides gras de la famille Oméga 3 et particulièrement en acide docosahexaénoïque (DHA), à base d'huile de poisson;
  - **efad** (*Essential Fatty Acid Deficient*) : régime constitué uniquement d'acides gras saturés, à base d'huile de coco hydrogénée;
  - **lin** : régime riche en Oméga 3, à base d'huile de lin;

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

- **ref** : régime dont l'apport en *Oméga 6* et en *Oméga 3* est adapté des Apports Nutritionnels Conseillés pour la population française, sept fois plus d'*Oméga 6* que d'*Oméga 3*;
- **tsol** : riche en *Oméga 6*, à base d'huile de tournesol.

Quatre souris de chaque génotype sont soumises à chaque régime alimentaire.

La question centrale de l'analyse des données d'expression consiste à détecter des gènes qui ont un comportement différent selon les conditions auxquelles ils sont soumis. Dans notre cas, nous cherchons à déterminer les gènes :

- caractéristiques de l'effet génotype,
- caractéristiques de l'effet régime,
- en relation avec des taux élevés d'acides gras hépatiques.

Sans revenir sur la phase de normalisation des données, précisons toutefois qu'elle a été abordée en centrant les valeurs sur l'intensité moyenne de 13 ARN exogènes ajoutés à l'échantillon avant marquage (*spike*). Elle a ensuite été validée par la présence de 8 autres ARN exogènes introduits à des doses parfaitement connues dans la solution initiale (Martin *et al.*, 2005b). Nous vérifions dans la première partie de l'analyse exploratoire qu'aucune membrane ne présente un comportement atypique (voir Fig. 1).

### 3. Démarche exploratoire

Dans ce paragraphe, nous passons en revue diverses techniques exploratoires dans le but, d'une part, de nous familiariser avec les données, d'autre part, de commencer à identifier certains gènes ayant un comportement particulier. Le lecteur souhaitant des développements sur les techniques évoquées ci-dessous pourra les trouver dans Saporta (1990), dans Lebart *et al.* (1995) ou dans Baccini *et al.* (2005). Signalons également l'ouvrage récent de McLachlan *et al.* (2004), plus spécifiquement orienté vers l'analyse des données d'expression.

#### 3.1. Stratégie

Dans une étude statistique sophistiquée, avant d'utiliser des méthodes de modélisation ou des techniques d'apprentissage, il est toujours prudent de commencer par une étude *exploratoire* à l'aide d'outils, certes élémentaires mais robustes, en privilégiant les représentations graphiques. C'est la seule façon de se familiariser avec les données et, surtout, de dépister les sources de problèmes comme les valeurs manquantes, erronées ou atypiques, les distributions «anormales» (dissymétrie, multimodalité, épaisseur des queues de distributions), les liaisons non linéaires...

Ensuite, au vu des résultats précédents, on peut être conduit à mettre en œuvre divers prétraitements des données afin de rendre ces dernières conformes aux hypothèses des techniques de modélisation ou d'apprentissage qu'il sera nécessaire d'utiliser pour atteindre les objectifs fixés. Ces prétraitements peuvent être les suivants :

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

- transformation des variables (logarithme, puissance...), centrage, réduction, passage aux rangs ;
- codage en classes ou recodage de classes ;
- imputation ou non des données manquantes ;
- réduction de dimension, classification et premier choix de variables ;
- classification ou typologie des observations.

Attention, le côté élémentaire des techniques exploratoires ne doit pas conduire à les négliger au profit d'une mise en œuvre immédiate de méthodes beaucoup plus sophistiquées, donc beaucoup plus sensibles aux problèmes cités ci-dessus. Si ces problèmes ne sont pas pris en compte au début de l'analyse, ils réapparaîtront ensuite comme autant d'*artefacts* susceptibles de dénaturer, voire de fausser, toute tentative de modélisation.

Pour l'analyse exploratoire, au-delà des techniques unidimensionnelles, deux grandes familles de méthodes sont utilisées :

- les méthodes factorielles ;
- la classification (*clustering*), ou apprentissage non supervisé.

Dans les deux cas, de nombreux choix sont laissés à l'utilisateur qui doit soit les effectuer en connaissance de cause, soit les tester pour arriver à une représentation satisfaisante compte tenu de ses a priori et des conditions expérimentales. Ces choix doivent bien sûr être connectés à ceux relatifs à la normalisation des données.

### 3.2. Choix méthodologiques

Nous donnons ici une présentation synthétique des choix qui doivent être explicités. Cette liste n'est sans doute pas exhaustive et devra être complétée avec l'approfondissement de l'expertise du traitement de ce type de données. Nous pouvons d'ores et déjà insister sur l'indispensable dialogue entre biologiste et statisticien pour opérer ces choix en connaissance de cause, tant sur les aspects techniques que sur leurs implications biologiques. Ces choix ne pourront bien sûr pas tous être discutés en détail dans le cadre restreint de cet article.

#### 3.2.1 Transformations

Les données traitées sont issues des procédures de normalisation spécifiques à la technologie utilisée pour les produire. Néanmoins, elles peuvent encore subir des transformations dont nous précisons ci-dessous les plus courantes.

**Passage au logarithme.** Cette transformation corrige une distribution trop dissymétrique (*skewness*) et réduit l'influence des grandes valeurs (éventuellement atypiques). Elle est justifiée dans la mesure où, dans certains systèmes naturels, divers effets peuvent être modélisés par des facteurs multiplicatifs plutôt qu'additifs.

**Centrage.** Les données se présentant sous la forme d'une matrice, il est habituel, par exemple dans une analyse en composantes principales (ACP),

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

de centrer les colonnes. L'information liée à la moyenne peut être utile en soi, mais est rarement très informative. Dans les données transcriptomiques, le rôle des lignes et des colonnes, c'est-à-dire la distinction entre individus et variables, n'est pas toujours explicite. Le choix doit être fait en fonction de la nécessité de centrer les données d'expression par gène, par biopuce, voire par gène et par biopuce, ce qui correspond au double centrage (lignes et colonnes).

**Réduction.** Dans les méthodes descriptives multidimensionnelles, il est courant de faire une réduction des variables soit lorsque leurs unités de mesure sont différentes soit, lorsqu'avec la même unité de mesure, on observe des variances très hétérogènes. Dans l'analyse de données transcriptomiques, on peut être conduit à faire une réduction des variables si celles-ci sont bien identifiées. Toutefois, il convient d'être très prudent compte tenu des problèmes suivants. En ramenant à un la variance des gènes, les effets de surexpression ou sous-expression de certains d'entre eux sont éliminés. Cette transformation a donc surtout un sens lors de l'étude d'un sous-ensemble de gènes déjà sélectionnés car différentiellement exprimés. D'autre part, dans une ACP, la réduction peut conduire à un effet taille artificiel sur l'axe 1 dû à un grand nombre de gènes d'expressions voisines et sans particularité. Nous l'avons observé dans l'ACP présentée en 3.4 où le phénomène le plus structurant, la séparation des deux génotypes, apparaît sur l'axe 1 dans l'ACP non réduite et sur l'axe 2 dans l'ACP réduite, le premier axe correspondant à un effet taille dans ce second cas.

**Marges unitaires.** Une autre façon d'éliminer l'influence des unités de mesure consiste à diviser les lignes (ou les colonnes) d'un tableau par ses marges en lignes (ou en colonnes). C'est la pratique courante lorsque le tableau contient des effectifs (table de contingence) et cela conduit à l'analyse des correspondances. Pour les raisons évoquées ci-dessus (surexpression et sous-expression), cette approche ne semble pas bien adaptée à l'analyse des données d'expression.

**Passage aux rangs.** Lorsque les données sont parsemées de valeurs atypiques sans qu'aucune transformation fonctionnelle (logarithme, puissance...) ne puisse en atténuer efficacement les effets, une façon «brutale», ou robuste, de procéder consiste à remplacer une valeur par son rang dans la séquence ordonnée. En pratique, cela revient à calculer le coefficient de Spearman plutôt que celui de Bravais-Pearson.

### 3.2.2 Pondérations et distances

Il arrive que l'on introduise des pondérations sur les lignes ou sur les colonnes du tableau des données. Cette pratique permet de donner plus ou moins d'importance à certains éléments; cela permet de redresser un échantillon. Un autre exemple simple consiste à affecter des poids nuls à certaines lignes ou à certaines colonnes, alors dites supplémentaires : elles n'interviennent pas dans les calculs mais restent représentées dans les graphiques. Par défaut, les poids sont égaux pour les lignes et égaux pour les colonnes. Sauf cas très particulier, on utilise des poids égaux dans l'analyse des données d'expression.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

D'un point de vue mathématique, chaque ligne (chaque colonne) est considérée comme un vecteur d'un espace vectoriel muni d'un produit scalaire induisant une norme euclidienne et une distance (entre ces vecteurs). Par défaut, cette distance est celle, classique, dont le carré est la somme des carrés des écarts entre les coordonnées de deux vecteurs. Introduire des pondérations sur les lignes (sur les colonnes) conduit à pondérer le calcul de cette distance. La matrice associée au produit scalaire est alors une matrice diagonale comportant le carré des pondérations sur la diagonale. Elle remplace la matrice identité associée à la distance classique.

D'autres matrices carrées symétriques et définies positives sont également utilisables de façon plus générale. Citons l'inverse de la matrice des variances-covariances résiduelles (ou intra-classes) en analyse discriminante (métrique de Mahalanobis) ou la matrice diagonale des inverses des fréquences marginales en analyse des correspondances (métrique du khi-deux). Citons encore les matrices largement répandues dans le contexte transcriptomique faisant intervenir la corrélation linéaire usuelle, ou la corrélation par rangs, entre deux variables  $X$  et  $Y$ , et dont le terme général peut être défini par  $1 - \text{cor}(X, Y)$  ou  $\sqrt{1 - \text{cor}(X, Y)^2}$ . Le problème du choix de la distance sera discuté en 3.5.

### 3.2.3 Factorisation et projections

Beaucoup de méthodes proposées recherchent des nouvelles variables non corrélées (*factor scores*) obtenues par combinaisons linéaires des variables initiales et optimisant un critère. Il s'agit de la maximisation de la variance dans le cadre de l'ACP qui conduit à la construction des variables principales. La décomposition ainsi obtenue a-t-elle un sens pour les données considérées ? Combien de composantes sont nécessaires pour résumer l'information et fournir des représentations graphiques pertinentes des nuages de points (individus et variables) ? Seule la discussion entre biologiste et statisticien peut éclairer ces questions.

Sur le plan mathématique, lorsqu'on utilise la métrique identité dans l'espace vectoriel des individus, les variables principales sont obtenues à partir des facteurs principaux, vecteurs propres associés aux plus grandes valeurs propres d'une matrice carrée, symétrique et définie non négative (variances-covariances, corrélations, produits scalaires...). De manière plus générale, les métriques définies dans l'espace des individus et dans celui des variables interviennent dans la définition de la matrice diagonalisée.

### 3.2.4 Classification

Une approche classique dans toute discipline scientifique consiste à faire de la taxinomie, c'est-à-dire à rechercher des classes homogènes des objets étudiés (gènes ou échantillons biologiques dans notre cas) au sens d'un critère défini par une matrice de distances ou de dissemblances. Le choix de ce critère, qui sera précisé au 3.6, est évidemment prépondérant pour la signification et l'interprétation des résultats obtenus.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

### 3.3. Techniques unidimensionnelles

Parmi les outils et indicateurs recensés, les diagrammes en boîtes présentent un grand intérêt pour l'analyse des données d'expression. Eux seuls permettent de représenter simultanément un grand nombre de distributions (quelques centaines) sur un même graphique et d'en analyser rapidement les spécificités en termes de médiane, de dispersion et de valeurs atypiques. Il serait bien entendu impossible de considérer simultanément autant d'histogrammes.

Ainsi, la représentation des diagrammes en boîtes pour les souris, ordonnées selon le génotype et le régime suivi (Fig. 1), ne donne *a priori* aucune tendance spécifique sur le comportement de l'ensemble des gènes. Cette représentation atteste de la qualité de la production et du prétraitement des données. En effet, celles-ci ont été recueillies en utilisant une membrane par souris; ainsi, une quelconque anomalie sur un support, affectant l'ensemble des mesures relatives à une souris particulière, apparaîtrait nécessairement sur cette représentation. Notons seulement que quelques gènes atypiques, facilement repérables sur la figure 2 comme les plus surexprimés, se retrouvent dans les valeurs extrêmes pour chaque souris sur la figure 1.

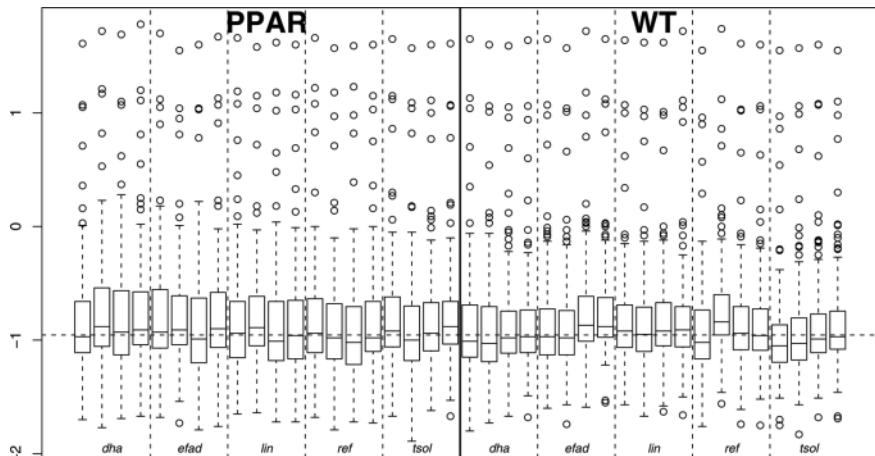


FIG 1. — Diagrammes en boîtes pour les 40 souris. La ligne verticale pleine sépare les souris selon leur génotype. Les lignes verticales en pointillés séparent les souris selon le régime qu'elles ont suivi. La ligne horizontale en pointillés représente la médiane de l'ensemble des valeurs

Les diagrammes en boîtes pour chaque gène (Fig. 2) révèlent des gènes dont l'expression est, sur l'ensemble des souris, nettement différentes des autres (par exemple, 16SR, apoA.I, apoE). Les gènes des ARN ribosomiques, comme le 16SR (ARN 16s ribosomique mitochondrial), présentent, dans toutes les cellules de l'organisme, des niveaux d'expression plus élevés que tous les gènes codant des ARN messagers. Ces ARN servent en effet à la traduction des ARN messagers en protéines. Par ailleurs, on peut constater que les expressions de

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

certains gènes varient beaucoup plus que d'autres sur l'ensemble des souris (par exemple, FAS, S14 et THIOL). Pour ces derniers gènes, on peut supposer qu'une part de cette variabilité est due aux facteurs considérés, ce que nous essaierons de confirmer par la suite au moyen de techniques de modélisation.

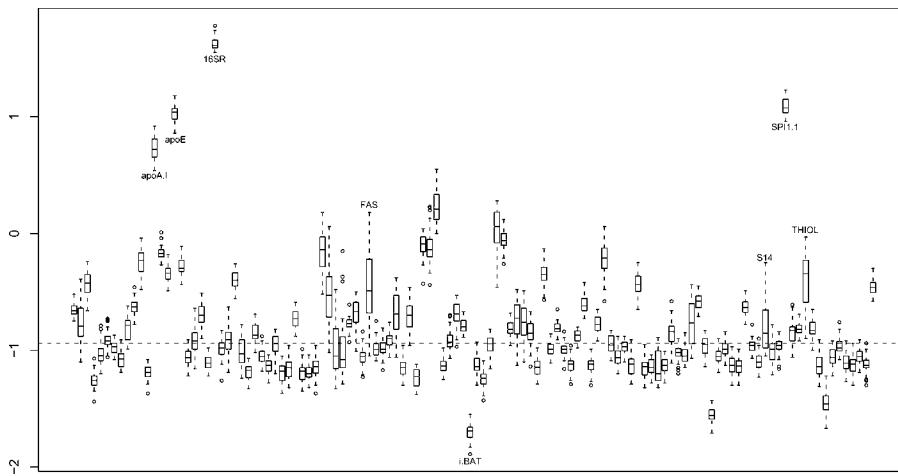


FIG 2. — Diagrammes en boîtes pour les 120 gènes. Quelques gènes particuliers ont été étiquetés

L'intérêt de ces représentations réside davantage dans la vision synthétique qu'elles offrent que dans l'information biologique que l'on peut en extraire. Elles nous orientent également dans les premiers choix méthodologiques à établir avant de poursuivre l'analyse. En effet, les boîtes relatives à la distribution des gènes mettent clairement en évidence un certain nombre de gènes dont l'expression est systématiquement supérieure à celle des autres, quelles que soient les conditions expérimentales. De plus, la variabilité de ces expressions est, le plus souvent, très faible. Ce constat nous conduit à effectuer un centrage des gènes (en colonnes), afin d'éviter un effet taille lors de la mise en œuvre de techniques factorielles. En revanche, rien dans ces représentations ne nous pousse à centrer les échantillons (en lignes), ce qui, par ailleurs, ne se justifierait pas sur le plan biologique. En effet, nous travaillons sur des données acquises via des puces dédiées sur lesquelles les gènes considérés ont été présélectionnés et sont donc, *a priori*, potentiellement différemment exprimés dans les conditions étudiées. Un centrage des échantillons serait susceptible de cacher des phénomènes biologiques. Ce raisonnement ne tiendrait pas pour une expérimentation pangénomique, où l'on pourrait supposer que globalement les gènes s'expriment de la même façon et que les surexprimés compensent les sous-exprimés.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

### 3.4. Analyse en Composantes Principales

Pour la mise en œuvre de l'ACP, nous travaillons sur le tableau croisant les souris en lignes (en tant qu'individus), et les gènes en colonnes (en tant que variables). Comme nous l'avons précisé dans le point précédent, ce tableau est préalablement centré en colonnes. Le centrage évite un effet taille sans grande signification quant à la discrimination des phénomènes observés.

D'un point de vue très technique, dans les logiciels statistiques R et S-PLUS, deux fonctions sont disponibles pour réaliser des ACP :

- `princomp` ne réalise que des ACP centrées et extrait les valeurs propres et vecteurs propres de la matrice des covariances (par défaut) ou des corrélations ; le nombre de colonnes de la matrice des données doit être inférieur ou égal au nombre de lignes ;
- `prcomp` calcule directement la décomposition en valeurs singulières (SVD) de la matrice des données centrées (par défaut, mais avec la possibilité de ne pas centrer) et non réduite (par défaut, mais avec possibilité de réduire) ; cette fonction accepte un nombre de colonnes supérieur au nombre de lignes.

Compte tenu de la particularité des données transcriptomiques, nous avons systématiquement utilisé la procédure `prcomp`.

Nous donnons ci-après le tableau (tableau 1) et le graphique (figure 3) des premières valeurs propres qui nous ont conduit à considérer trois dimensions représentant environ les deux tiers de l'inertie globale.

TABLEAU 1. – Parts de variance expliquée et cumuls pour les six premiers axes principaux

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
Parts de variance expliquée	0.350	0.196	0.124	0.0608	0.0447	0.0289
Cumuls	0.350	0.546	0.670	0.7311	0.7757	0.8046

Les figures 4 et 5 donnent la représentation des souris et celle des gènes, d'abord dans le premier plan principal, ensuite dans celui correspondant aux dimensions 1 et 3. Dans le cadre de cette ACP, il est cohérent de rechercher quels sont les gènes contribuant le plus à la définition des trois premiers axes. Une adaptation de l'expression classique de la contribution d'un individu fournit, pour la  $j^{\text{ème}}$  variable, la quantité :

$$\frac{\sum_{k=1}^3 \lambda_k (u_j^k)^2}{\sum_{k=1}^3 \lambda_k}, \quad (1)$$

où les  $\lambda_k$  sont les valeurs propres de l'ACP et les  $u_j^k$  les coordonnées des vecteurs propres normés correspondants.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

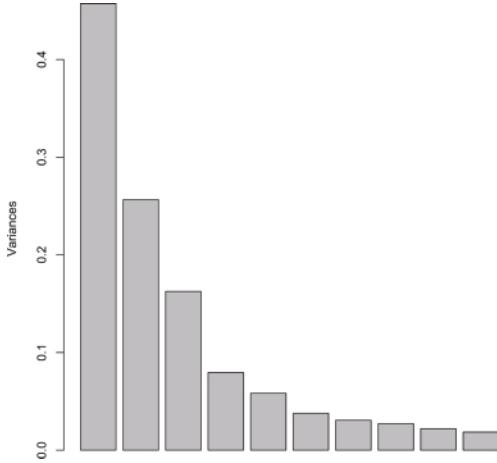


FIG 3. — Éboulis des dix premières valeurs propres de l'ACP

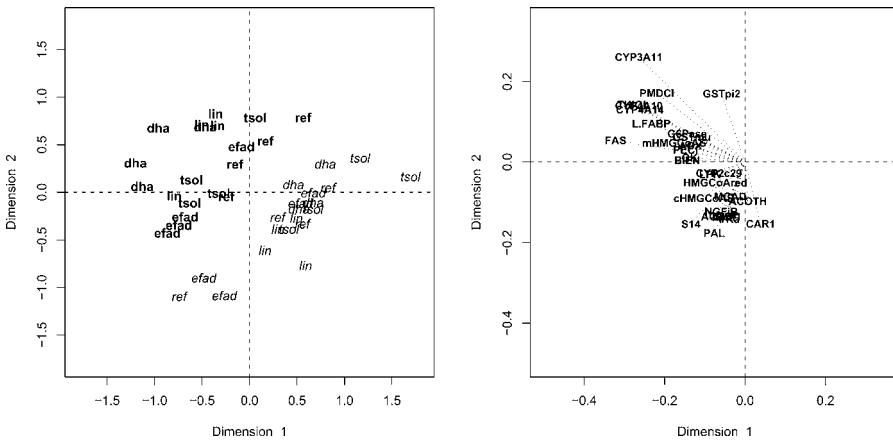


FIG 4. — Représentations sur le premier plan principal de l'ACP. À gauche : individus-souris identifiés par leur génotype (WT en gras, PPAR en italique) et leur régime (**dha**, *efad*, *lin*, *ref*, *tsol*). À droite : les 30 variables-gènes qui contribuent le plus aux trois premiers axes

Ces quantités permettent de rechercher, par exemple, les 25 % des gènes contribuant le plus à la définition de l'espace propre à trois dimensions jugé pertinent. Avec cette sélection, la représentation des variables ainsi restreinte à 30 gènes est plus facilement lisible sur les figures 4 et 5. Toutefois, dans le cas d'une puce pangénomique, avec potentiellement plusieurs milliers de gènes, une telle représentation ne serait pas exploitable.

Le premier plan (Fig. 4) doit être interprété globalement puisque sa première bissectrice sépare exactement les souris WT des souris PPAR. Les gènes à coordonnées négatives sur l'axe 1 et positives sur l'axe 2 sont sensiblement

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

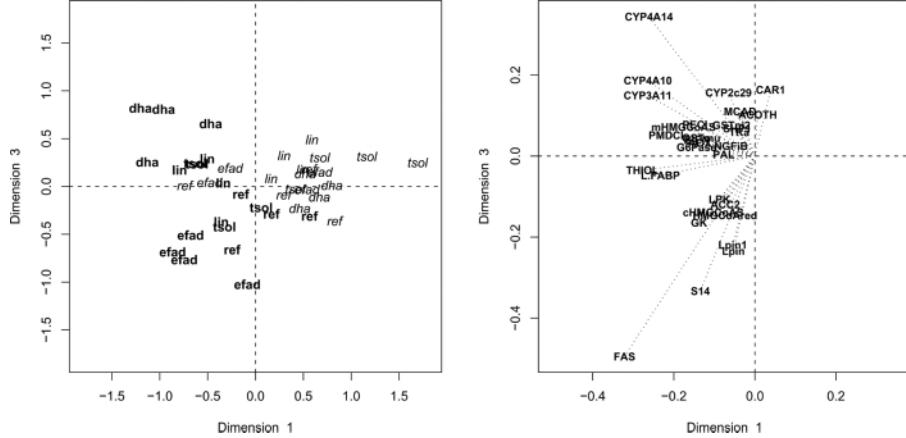


FIG 5. — Représentations sur le plan composé des axes principaux 1 et 3 avec les mêmes conventions que pour la figure 4

plus exprimés chez les souris WT, en particulier CYP3A11, CYP4A10, CYP4A14, THIOL, PMDC1, GSTpi2, L.FABP et FAS. À l'inverse, les gènes à forte coordonnée négative sur l'axe 2 s'expriment davantage chez les souris PPAR, par exemple, S14, PAL et CAR1. Ceci est en partie connu des biologistes (Aoyama *et al.*, 1998).

Le phénomène le plus marquant concernant l'axe 3 (Fig. 5) est l'opposition, chez les souris WT, entre les régimes **dha**, dont les coordonnées sont toutes positives, et **efad**, dont les coordonnées sont toutes négatives. Les gènes les plus exprimés dans le premier cas (régime **dha** chez les souris WT) sont CYP3A11, CYP4A10, CYP4A14, CYP2c29 et CAR1; dans le second cas (régime **efad** chez les mêmes souris), il s'agit des gènes FAS, S14, Lpin et Lpin1. Parmi ces régulations, on note une opposition entre les gènes CYP4A, connus pour être impliqués dans le catabolisme des acides gras, et les gènes FAS et S14, impliqués eux dans la synthèse des lipides. Par ailleurs, la régulation du gène CYP3A11 par l'acide DHA a déjà été décrite dans Berger *et al.* (2002).

### 3.5. Positionnement multidimensionnel

L'ACP considère implicitement des distances euclidiennes entre les lignes et entre les colonnes du tableau, relativement aux métriques associées à des matrices particulières : la matrice identité pour les lignes, la matrice diagonale des poids pour les colonnes. Il peut être intéressant, sur un plan biologique, de choisir d'autres distances. Ce choix est souvent proposé par les logiciels de traitement statistique de données transcriptomiques destinés aux biologistes. Le positionnement multidimensionnel (*multidimensional scaling*, ou MDS), également appelé ACP d'un tableau de distances, permet d'illustrer les effets des différents choix.

L'objectif du MDS (Mardia *et al.*, 1979) est de rechercher, dans un espace euclidien de dimension réduite  $q$  fixée *a priori*, la représentation euclidienne

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

la plus proche, c'est-à-dire respectant au mieux les distances (pas nécessairement euclidiennes) entre les individus décrits dans une matrice symétrique  $D$ , ( $n \times n$ ). Plus précisément, il s'agit de construire une matrice ( $n \times q$ ), dite des coordonnées principales, de sorte que les distances euclidiennes calculées entre les vecteurs lignes de cette matrice soient globalement les plus proches des distances initiales. Les coordonnées principales fournissent alors la représentation graphique recherchée. Dans le cas très particulier où la matrice des distances  $D$  est euclidienne, calculée à partir de la matrice  $X(n \times p)$  des coordonnées des individus, l'ACP de  $X$  coïncide à un facteur près au MDS de  $D$ .

Nous avons procédé au positionnement multidimensionnel des données d'expression génique à partir d'une matrice de distance inter-gènes calculée selon différentes formules :

- distance euclidienne,  $d_1(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$ , positive ou nulle ;
- distance associée à la corrélation carrée,  $d_2(X, Y) = \sqrt{1 - \text{cor}(X, Y)^2}$ , comprise entre 0 et 1 ;
- distance associée à la corrélation,  $d_3(X, Y) = 1 - \text{cor}(X, Y)$ , comprise entre 0 et 2.

Remarquons tout d'abord que, dans les trois cas, plus la valeur est petite plus les gènes dont on mesure l'éloignement sont proches. Ensuite, pour  $d_2$  et  $d_3$ , une valeur proche de 1 caractérise deux gènes non corrélés, ce qui n'est pas nécessairement le cas de la distance euclidienne. Enfin, il est important de noter qu'une corrélation forte et négative entre deux gènes conduit à deux résultats opposés selon  $d_2$  (valeur proche de 0) et  $d_3$  (valeur proche de 2),  $d_2$  ne dépendant pas du signe de la corrélation contrairement à  $d_3$ .

La figure 6 illustre les trois possibilités avec le positionnement multidimensionnel des gènes. L'analyse conjointe de ces trois graphiques conduit à de nombreuses interprétations sur le plan biologique. Sans entrer dans les détails, nous noterons que ces trois graphiques tendent à séparer deux groupes de gènes qui interviennent dans deux fonctions biologiques opposées : les CYP4A, PMDC1, PEC1, AOX, BIEN, THIOL, CPT2, mHMGCoAS, Talpha et Tbeta sont impliqués dans le catabolisme des lipides et la cétogénèse alors que les gènes FAS, S14, ACC2, cHMGCoAS, HMGCoAred et, plus indirectement, GK et LPK sont impliqués dans la synthèse de lipides au niveau hépatique. On observera qu'aucun des trois graphiques de la figure 6, analysé individuellement, ne conduit à la totalité de cette interprétation mais que c'est bien l'analyse conjointe de ces représentations qui permet d'affiner la connaissance du biologiste sur ces données. Succinctement, notons également que d'autres gènes tendent à participer à ces groupes. Par exemple, le gène Lpin1 est proche des gènes impliqués dans la lipogénèse. Bien que sa fonction soit actuellement inconnue, Peterfy *et al.* (2001) ont observé que la lignée de souris déficiente pour Lpin1 présente des altérations du métabolisme des lipides.

Les gènes dont la position sur le graphique sera la plus modifiée en passant de la distance  $d_2$  à la distance  $d_3$  seront ceux présentant des corrélations négatives et importantes avec de nombreux autres gènes. Un cas typique dans notre exemple est celui de CAR1 dont l'ACP (et d'abord la matrice des corrélations)

# ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

a montré qu'il était négativement corrélé avec des gènes tels que GSTpi2, CYP3A11, FAS... La position relative des couples de gènes ainsi obtenus change de façon importante entre les deux graphiques. On observera en particulier le couple CAR1-GSTpi2 totalement opposé sur l'axe 1 selon  $d_3$  et relativement proche selon  $d_2$  (tandis qu'il présente une opposition moins marquée selon  $d_1$ ). La surexpression du gène CAR1 et la sous-expression du gène GSTpi2 chez les souris déficientes en récepteur PPAR $\alpha$  n'ont pas été décrites et constituent l'un des résultats originaux de ce travail. L'étude d'un lien potentiel entre ces deux modifications d'expression nécessitera la mise en œuvre d'expériences complémentaires.

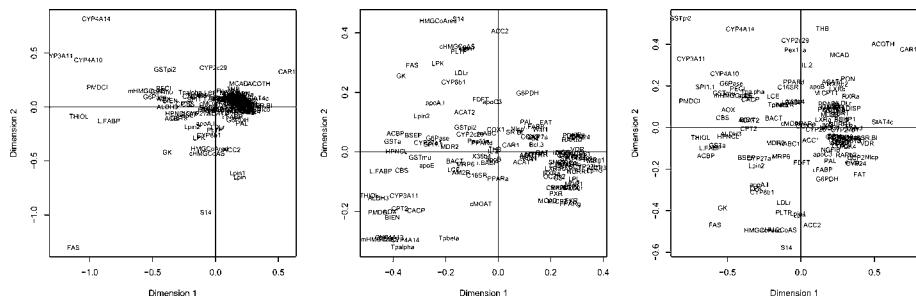


FIG 6. — Positionnement multidimensionnel des gènes sur les axes 1 et 2 selon 3 distances différentes : distance euclidienne ( $d_1$  à gauche), corrélation carrée ( $d_2$  au centre), corrélation ( $d_3$  à droite)

TABLEAU 2. – Comparaison de la distance euclidienne ( $d_1$ ), de la distance basée sur la corrélation carrée ( $d_2$ ) et de la distance basée sur la corrélation ( $d_3$ ) sur un exemple élémentaire. Les calculs sont effectués pour chacune des variables  $Y_i$  ( $i = 1, \dots, 4$ ) par rapport à la variable X

	X	Y1	Y2	Y3	Y4
	1	2	-2	10	2
	2	3	-3	20	3
	3	4	-4	30	5
	4	5	-5	40	4
	5	6	-6	50	1
$d_1(X, Y_i)$	0	2,2	16,9	66,8	4,7
$d_2(X, Y_i)$	0	0	0	0	0,99
$d_3(X, Y_i)$	0	0	2	0	1,1

Ces tendances générales sont illustrées dans le tableau 2 sur la base d'un exemple élémentaire. Ce tableau a été construit de manière à illustrer les comportements différents des trois distances envisagées. Hormis dans le cas de  $Y1$  où  $d_1$ ,  $d_2$  et  $d_3$  donnent la même indication (faible éloignement), les conclusions divergent selon la distance envisagée. Pour  $Y2$ , corrélée négativement avec  $X$ ,  $d_1$  et  $d_3$  sont importantes alors que  $d_2$  est nulle, comme entre

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

$X$  et  $Y1$ . Pour  $Y3$ ,  $d_1$  montre deux variables très éloignées alors que  $d_2$  et  $d_3$  sont nulles. Pour  $Y4$ , non corrélée avec  $X$ ,  $d_1$  est relativement faible alors que  $d_2$  et  $d_3$  sont proches de 1.

D'une manière générale, on peut retenir que l'utilisation de la distance euclidienne tend à rapprocher des gènes dont les expressions sont proches. En revanche, les deux autres indicateurs considèrent que deux gènes sont proches si leur expression varie dans le même sens selon les conditions expérimentales. La corrélation ( $d_3$ ) distingue les gènes corrélés négativement, ce que ne permet pas la corrélation carrée ( $d_2$ ) qui doit donc être utilisée en connaissance de cause.

Notons que la distance  $d_1$  est plus courante en statistique alors que  $d_3$  l'est davantage dans les études relatives aux biopuces. Autant que possible, une comparaison des trois approches est recommandée. On se référera à Draghici (2003, chapitre 11) pour une discussion plus détaillée sur le sujet.

### 3.6. Classification

La mise en œuvre d'une méthode de classification non supervisée vise à obtenir une partition des individus (ou des variables) en classes, sans *a priori* sur le nombre de classes. Toute méthode de ce type est basée sur deux critères de distance : l'un entre individus (ou variables), l'autre entre groupe d'individus (ou de variables). Le premier choix est guidé par les mêmes considérations que dans le cas du MDS : les mêmes distances entre gènes peuvent être considérées avec les mêmes implications biologiques. Depuis les travaux de Eisen *et al.* (1998), la distance la plus fréquemment rencontrée pour l'étude du transcriptome est du type de  $d_3$ , basée sur la corrélation ; il nous semble cependant pertinent d'utiliser les trois types de distances et d'en apprécier la complémentarité quant à l'interprétation des résultats, ce que nous avons fait ci-dessus pour le MDS. En revanche, nous nous contenterons ici de présenter une classification basée sur la distance euclidienne  $d_1$ . Le deuxième choix intervenant en classification concerne le critère d'agglomération, c'est-à-dire la façon dont est définie la distance entre deux groupes, et n'a pas d'interprétation biologique simple. Ce choix a plus une implication géométrique, sur la forme des classes obtenues. Nous avons utilisé le critère de Ward (regroupement des deux classes minimisant la perte d'inertie inter-classes, voir Lebart *et al.*, 1995) parce qu'il favorise la construction de classes relativement « sphériques » et qu'on peut lui associer des critères guidant la détermination du nombre de classes (Baccini *et al.*, 2005).

Une représentation très répandue en analyse des données de biopuces consiste à fournir simultanément les résultats de deux classifications ascendantes hiérarchiques menées indépendamment sur les lignes (individus-souris) et sur les colonnes (variables-gènes) du tableau de données.

L'interprétation de la figure 7 présente des analogies avec celle de l'ACP sur le premier plan principal. Si l'on s'intéresse aux individus-souris, on peut constater que les deux génotypes sont différenciés en deux groupes, à l'exception de trois souris de type PPAR ayant suivi les régimes **efad** (pour deux d'entre elles) et **ref**. Ce sont ces trois mêmes individus que l'on retrouve

# ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

projétés dans la partie négative du premier axe de l'ACP (Fig. 4). Pour les variables-gènes, un groupe attire particulièrement l'attention sur l'image : sur une bande verticale (encadrée en pointillés sur la figure) correspondant à 14 gènes, les niveaux de gris sont nettement plus variables que sur le reste de l'image. Il s'agit des gènes

CYP4A10, CYP4A14, CYP3A11, L.FABP, THIOL, PMDC1, S14, Lpin1, Lpin, FAS, GSTmu, GSTpi2, CYP2c29, G6Pase

qui apparaissent tous parmi les gènes les plus corrélés aux trois premiers axes principaux de l'ACP (Fig. 4 et 5)

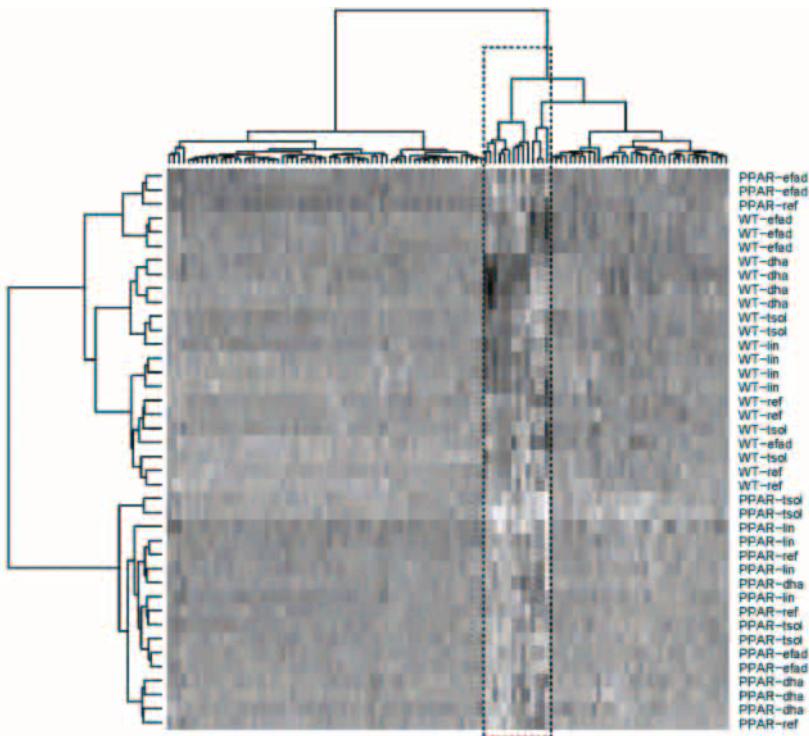


FIG 7. — Représentation simultanée des résultats des classifications ascendantes hiérarchiques des individus-souris et des variables-gènes selon la méthode de Ward, avec la distance euclidienne. Sur l'image du tableau des données, les intensités sont croissantes du blanc vers le noir. Le rectangle en pointillés met en évidence 14 gènes au comportement particulier

MDS et classification apparaissent donc comme des techniques complémentaires, mais elles ne sont pas sensibles de la même façon aux perturbations. La perturbation d'une donnée peut fortement influencer la structure d'un dendrogramme alors qu'en MDS, la prise en compte conjointe de toutes les distances deux à deux assure une certaine robustesse pour le calcul des coordonnées

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

principales. Pour cette raison, il est utile de représenter les classes dans une projection sur les axes factoriels obtenus soit par MDS soit par ACP. L'éboulis des valeurs propres (Fig. 8) nous oriente vers une représentation du MDS en deux dimensions.

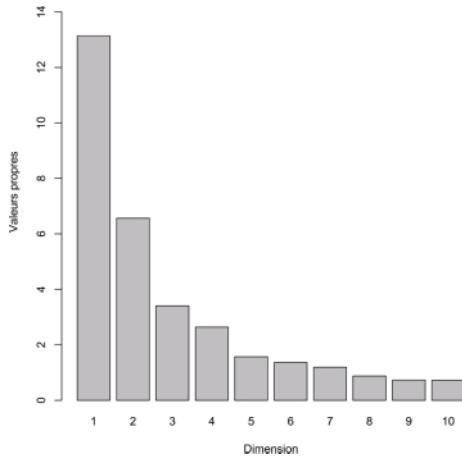


FIG 8. — Éboulis des valeurs propres pour le MDS de la matrice des distances euclidiennes inter-gènes

La représentation de la figure 9 est analogue à celle déjà présentée (Fig. 6, graphique de gauche). Elle est complétée par un codage graphique des gènes, selon leur appartenance à un groupe issu de la classification hiérarchique. Pour cela, nous avons coupé l'arbre afin d'en extraire 5 groupes.

Brièvement, on peut noter que l'axe 1 met en évidence l'opposition précédemment évoquée entre **CAR1** (surexprimé chez les souris PPAR) et un groupe de gènes (**CYP3A10**, **CYP4A10**, **CYP4A14**, **PMDCI**, **THIOL** et **L.FABP**) qui est surexprimé chez les souris **WT**. De manière similaire, l'axe 2 oppose les gènes induits par le régime **dha** (valeurs positives, gènes impliqués dans le catabolisme des lipides et dans le métabolisme des xénobiotiques) aux gènes induits par le régime **efad** (valeurs négatives, gènes principalement impliqués dans la synthèse des lipides). En remontant vers les feuilles de l'arbre de classification, on notera que le groupe des gènes représentés au centre et en italique est séparé en deux sous-groupes qui conservent une cohérence vis-à-vis des fonctions biologiques de catabolisme et de synthèse des lipides respectivement. Une observation des données individuelles révèle que ces régulations opérées par les régimes semblent plus marquées chez les souris **WT**. Nous verrons ultérieurement que d'autres techniques (forêts aléatoires par exemple) permettent de confirmer ces observations de manière plus objective.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

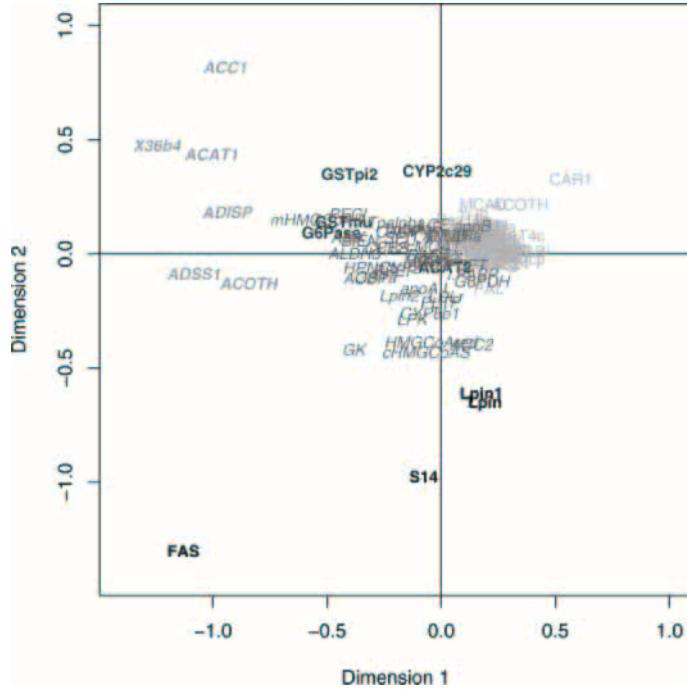


FIG 9. — Représentation par positionnement multidimensionnel (distance euclidienne) des 5 groupes issus de la classification hiérarchique des gènes

## 4. Modélisations

### 4.1. Analyses de variance et modèle mixte

Avant de détailler les modèles utilisés, précisons tout de suite que les tests mis en œuvre en ANOVA ou dans les modèles mixtes supposent des observations indépendantes. Or, pour un génotype donné, un régime donné et une souris donnée, les expressions des différents gènes de cette souris ne sont pas indépendantes, à cause des régulations entre gènes. Dans une expérience comme celle considérée ici, il n'est pourtant pas possible de prendre en compte cette structure de dépendance entre gènes. En effet, la prise en compte, par exemple, d'un modèle mixte avec le facteur gène en effet aléatoire et une structure de covariance « pleine » entraînerait l'estimation de 7260 paramètres  $\binom{120(120 + 1)}{2}$ . Toutefois, certaines simulations ont montré que les tests considérés restent en général acceptables, en dépit de la dépendance des données (voir, par exemple, Nguyen, 2004 et Dudoit *et al.*, 2003).

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

### 4.1.1 Principe des analyses de variance

L'analyse de variance (ANOVA) permet d'apprécier l'effet d'une ou de plusieurs variables qualitatives (les facteurs) sur une variable quantitative (la variable réponse, ici le niveau d'expression des gènes). Dans le domaine de l'analyse transcriptomique, cette approche a été largement développée, en particulier par Kerr *et al.* (2000). Pour l'analyse de nos données, un modèle d'ANOVA à trois facteurs (génotype, régime, gène) permet de mettre en évidence des effets de l'interaction d'ordre 3 très significatifs à l'aide du test de Fisher. Cela signifie qu'il existe des gènes régulés simultanément par le régime et le génotype, les effets du régime et du génotype étant non additifs. Le modèle d'ANOVA considéré s'écrit

$$y_{ijkl} = g_i + r_j + G_k + gr_{ij} + gG_{ik} + rG_{jk} + grG_{ijk} + e_{ijkl} \quad (2)$$

où  $y_{ijkl}$  représente le logarithme du niveau d'expression du gène  $k$  ( $k = 1, \dots, 120$ ), pour le régime  $j$  ( $j = 1, \dots, 5$ ) et le génotype  $i$  ( $i = 1, 2$ ), mesuré chez la souris  $l$  ( $l = 1, \dots, 4$ );  $g_i$  représente l'effet du génotype  $i$ ,  $r_j$  celui du régime  $j$ ,  $G_k$  celui du gène  $k$ ,  $gr_{ij}$  représente l'effet de l'interaction du génotype  $i$  et du régime  $k$ ,  $gG_{ik}$  l'effet de l'interaction du génotype  $i$  et du gène  $k$ ,  $rG_{jk}$  l'effet de l'interaction du régime  $j$  et du gène  $k$  et  $grG_{ijk}$  représente l'interaction d'ordre 3 combinant le génotype  $i$ , le régime  $j$  et le gène  $k$ . On suppose que les résidus  $e_{ijkl}$  du modèle sont indépendants et identiquement distribués suivant une loi normale de moyenne nulle et de variance  $\sigma^2$ . L'écriture d'un tel modèle suppose que les gènes sont tous de même variabilité. Cette hypothèse est discutable (en effet, la figure 2 montre clairement quelques gènes à forte variabilité) ; nous verrons par la suite comment lever cette difficulté.

Une autre approche consiste à écrire un modèle d'ANOVA par gène, sous la forme

$$y_{ijl} = g_i + r_j + gr_{ij} + e_{ijl} \quad (3)$$

où les notations utilisées ici sont identiques à celles du modèle (2). Ici, il est nécessaire de faire autant d'analyses de variance que de gènes étudiés (soit 120 dans notre exemple) mais nous disposerons d'une variance estimée par gène. Toutefois, une telle analyse n'est pas toujours recommandée car, en règle générale, le nombre d'observations par gène est très faible, ce qui conduit à des estimations de variance très peu précises. Notons cependant que ces 120 analyses conduisent à 120 estimations des 10 effets  $genotype_i \times regime_j$ . Un modèle équivalent, mais utilisant simultanément l'ensemble des données pour estimer les paramètres, s'écrit comme le modèle (2) en posant

$$\text{var}(e_{ijkl}) = \sigma_k^2 \quad (k = 1, \dots, 120). \quad (4)$$

D'autre part, entre le modèle (2), supposant toutes les variances des gènes égales, et le modèle (4), supposant une variance différente pour chaque gène, il est possible d'ajuster un modèle intermédiaire prenant en compte les hétérogénéités de variances de l'expression des gènes, en définissant simplement des groupes de gènes de variabilité homogène (Robert-Granié *et al.*, 1999; Foulley *et al.*, 2000; San Cristobal *et al.*, 2002; Delmar *et al.*, 2005).

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

Ainsi, sur les 120 gènes analysés, un histogramme des variances nous a conduit à définir trois groupes de gènes ayant des variabilités très différentes : un groupe contenant les gènes FAS, G6Pase, PAL et S14, présentant des variabilités résiduelles importantes (variances supérieures à 0.02) ; un deuxième groupe à variabilité modérée (variances comprises entre 0.009 et 0.02), comprenant les gènes CYP2c29, CYP3A11, CYP4A10, CYP4A14, CYP8b1, GSTmu, GSTpi2, L.FABP, Lpin, Lpin1, TRa et cHMGCoAS ; enfin un dernier groupe à faible variabilité (variances inférieures à 0.009), contenant l'ensemble des autres gènes. À partir de ces trois groupes de gènes, nous pouvons construire un modèle dont la variance dépend de cette nouvelle variable à trois classes. Le modèle s'écrit encore comme les modèles (2) et (4) en posant cette fois

$$\text{var}(e_{ijkl}) = \sigma_h^2, \quad (5)$$

où  $h = \{1, 2, 3\}$  représente l'indice d'hétérogénéité de variance.

Nous pouvons ainsi comparer les gènes différemment exprimés selon les 3 modèles :

- Modèle (2), modèle d'ANOVA avec une unique variance pour l'ensemble des gènes ;
- Modèle (4), modèle d'ANOVA avec une variance différente par gène ;
- Modèle (5), modèle d'ANOVA avec trois groupes de variances différentes.

Notons que le modèle (4) implique l'estimation de 120 variances différentes, alors que le modèle (5) ne nécessite l'estimation que de trois paramètres de variances ; ce dernier est donc beaucoup plus économique en nombre de paramètres à estimer. Enfin, d'un point de vue technique et opérationnel, la mise en œuvre de ces modèles peut être réalisée en utilisant la fonction `lme` du logiciel statistique R ou la procédure `mixed` du logiciel SAS.

### 4.1.2 Problème des tests multiples

L'objectif de l'analyse statistique est de déterminer quels sont les gènes différemment exprimés entre les 2 génotypes et les 5 régimes. Quelle que soit la méthode statistique utilisée, il existe une probabilité non nulle (risque de première espèce  $\alpha$ ) de détecter des faux positifs (gènes déclarés différemment exprimés alors qu'ils ne le sont pas) et une autre probabilité non nulle (risque de deuxième espèce  $\beta$ ) de ne pas être capable de détecter des gènes réellement différemment exprimés (faux négatifs). Il est bien entendu souhaitable de minimiser ces deux probabilités d'erreur sachant que, toutes choses égales par ailleurs, la seconde augmente quand la première diminue et réciproquement. Le test de Student est couramment utilisé pour tester l'égalité de deux moyennes (l'hypothèse nulle consistant à considérer que les moyennes des intensités des signaux d'un gène donné dans chacune des deux conditions sont égales). Ainsi, quand la statistique de Student excède un certain seuil (dépendant du risque de première espèce  $\alpha$  choisi, généralement 5 %), les niveaux d'expression du gène étudié entre les deux populations testées sont considérés comme significativement différents. Lorsque l'on souhaite tester plus de deux conditions, le test de Fisher, qui est une extension du

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

test de Student, est utilisé. L'hypothèse nulle consiste à supposer l'absence d'expression différentielle d'un gène entre les diverses conditions et l'hypothèse alternative à supposer une différence d'expression.

Bien sûr, prendre un risque de 5 % dans une expérimentation où 10 000 gènes, par exemple, sont étudiés simultanément peut conduire à obtenir autour de 500 faux positifs, ce qui est parfaitement inacceptable. C'est pourquoi ont été proposées des modifications du test de Student adaptées à l'analyse du transcriptome (méthodes de Bonferroni, FWER, FDR...). Le lecteur souhaitant des détails sur ces approches peut se référer, entre autres, à Benjamini & Hochberg (1995), Bland & Altman (1995), Dudoit *et al.* (2002) ou Speed (2003) et plusieurs articles de ce même volume.

La méthode de Bonferroni est une méthode qui ne permet pas un strict contrôle de  $\alpha$ , mais qui en donne une majoration. Pour avoir un risque global  $\alpha$ , il faut que chacune des  $p$  comparaisons soit effectuée avec un risque  $\alpha' = \alpha/p$ . En pratique, Bonferroni fournit une liste de gènes différentiellement exprimés dans laquelle on contrôle le nombre de faux positifs. Mais, lorsque le nombre des gènes est grand, cette liste est souvent vide.

En général, on présente ces taux d'erreurs dans le tableau suivant où  $m$  tests sont effectués :

Réalité	Décision		
	$H_0$ vraie	$H_1$ vraie	Total
$H_0$ vraie	$U$	$V$	$m_0$
$H_1$ vraie	$T$	$S$	$m_1$
	W	R	$m$

Pour une analyse de biopuces dans laquelle on teste les effets différentiels de  $m$  gènes,  $R$  est le nombre de gènes déclarés différentiellement exprimés, alors que  $m_1$  est le nombre réel (mais inconnu) de gènes différentiellement exprimés. Diverses méthodes sont proposées pour contrôler ces divers taux d'erreurs. Nous en précisons deux ci-dessous.

Le *Family Wise Error Rate* (FWER) représente la probabilité d'effectuer au moins une erreur de première espèce sur l'ensemble des comparaisons :

$$P[V \geq 1] \leq m\alpha.$$

On prend donc un seuil nominal de  $\alpha' = \alpha/m$ . Au même titre que Bonferroni, plus il y a de tests (c'est-à-dire de gènes à tester), moins on rejette  $H_0$  (moins il y a de gènes déclarés différentiellement exprimés). La notion suivante est très utile pour pallier cet inconvénient.

Le *False Discovery Rate* (FDR) contrôle l'espérance du taux de faux positifs, ou le nombre de faux positifs parmi les différences déclarées significatives. Pratiquement, on range par ordre croissant les  $m$  *p-values* des  $m$  tests (les gènes), on recherche le plus grand rang  $k$  des *p-values* tel que

$$p-value(k) < \alpha k/m$$

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

et on sélectionne les gènes de rang plus petit que  $k$ .

Il existe d'autres approches récentes, ou en cours de développement, pour contrôler le FDR, le nombre moyen d'erreurs... (voir, par exemple, Bar-Hen et Robin, 2003, Aubert *et al.*, 2004, Broët *et al.*, 2004, Dalmasso *et al.*, 2005, Delmar *et al.*, 2005, ou encore plusieurs articles de ce même volume).

Pour revenir à notre étude, à partir de chaque modèle proposé dans le paragraphe précédent, nous pouvons rechercher les gènes différentiellement exprimés entre les deux génotypes à régime fixé (120 comparaisons pour chacun des 5 régimes) ou entre régimes à génotype fixé (1 200 comparaisons par génotype), ce qui conduit à effectuer 3 000 comparaisons. Le tableau 3 présente le nombre de gènes sélectionnés selon les trois modèles considérés et selon le test ou l'ajustement utilisé (Student, Bonferroni et Benjamini-Hochberg qui correspond à l'approche FDR).

TABLEAU 3. – Nombre de gènes sélectionnés selon le modèle et le test utilisés

Tests	Modèle (2)	Modèle (4)	Modèle (5)
Student à 5 %	85	103	97
Student à 1 %	55	65	67
Benjamini-Hochberg à 5 %	44	56	57
Benjamini-Hochberg à 1 %	35	40	38
Bonferroni à 5 %	26	25	26
Bonferroni à 1 %	20	22	24

On peut remarquer que, pour un mode d'ajustement donné, le nombre de gènes sélectionnés est peu différent selon le modèle utilisé et que, globalement, les trois modèles sélectionnent le même groupe de gènes. Les petites différences sont principalement liées à l'ordre de sélection de ces gènes.

D'autre part, on peut, à partir de critères de sélection de modèle tels que le critère d'Akaike (AIC; Akaike, 1974) ou le critère de Schwarz (BIC; Schwarz, 1978), ou encore en effectuant un test du rapport de vraisemblance, choisir le modèle le plus adéquat.

Le tableau 4 présente les valeurs des critères AIC et BIC pour les trois modèles mis en compétition.

TABLEAU 4. – Valeurs des critères AIC et BIC

Modèles	-2AIC	-2BIC
(2)	-6576.9	-6570.7
(4)	-6946.6	-6612.1
(5)	-7044.5	-7036.2

Le meilleur modèle est celui pour lequel les valeurs des critères -2AIC ou -2BIC sont les plus petites. Dans les deux cas, il s'agit du modèle (5).

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

Le test du rapport de vraisemblance consiste, quant à lui, à comparer deux modèles emboités (par exemple, (2) vs (4) : l'hypothèse nulle considérée suppose alors que toutes les variances sont égales). La statistique du rapport de vraisemblance nécessite de calculer la différence entre les logarithmes des vraisemblances sous chacun des deux modèles. Sous l'hypothèse nulle, cette statistique suit asymptotiquement une loi de khi-deux dont le nombre de degrés de liberté est égal à la différence des nombres de paramètres à estimer sous chacun des deux modèles considérés. Si nous effectuons ces différents tests du rapport de vraisemblance ((2) vs (4), (2) vs (5), (4) vs (5)), il en ressort que le modèle (5), avec trois groupes de variances, est encore le meilleur.

À partir de ce modèle (5), on peut estimer les différents effets du modèle, et s'intéresser aux différences d'expression des gènes entre génotypes à régime fixé ou encore aux différences d'expression des gènes entre régimes à génotype fixé.

En raison de la multiplicité des tests, la correction proposée par Benjamini & Hochberg (1995), souvent considérée comme la plus appropriée, a été utilisée. Lorsque nous considérons les différences d'expression des gènes entre génotypes à régime fixé, l'hypothèse nulle représente l'absence d'expression différentielle d'un gène entre les deux génotypes. On peut visualiser l'ensemble des résultats des *p*-values de ces différents tests en effectuant une ACP centrée sur le tableau contenant, en lignes, les 57 gènes difféntiellement exprimés (selon le modèle (5) et la correction de Benjamini & Hochberg à 5 % (Tab. 3)) et, en colonnes, l'opposé du logarithme de la *p*-value associée au différentiel d'expression entre les deux génotypes pour les cinq régimes.

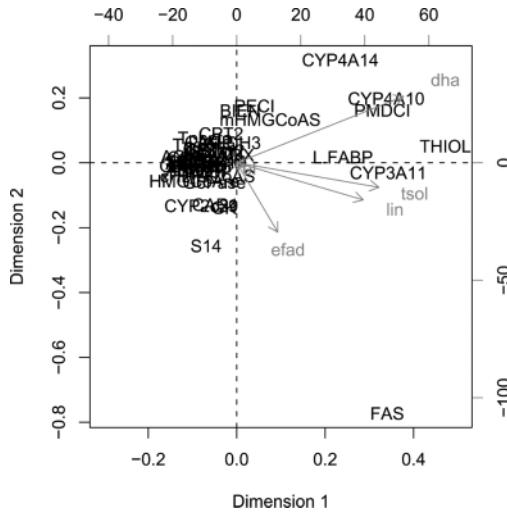


FIG 10. — Représentation sur le premier plan principal de l'ACP de l'opposé du logarithme des *p*-values des gènes difféntiellement exprimés entre les deux génotypes à régime fixé

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

On observe que les gènes CYP3A11, CYP4A10, CYP4A14, L.FABP, PMDCI et THIOL diffèrent les deux génotypes pour les régimes **dha**, **lin** et **tsol**. Certains de ces gènes présentent des expressions constitutives différentielles entre les souris des deux génotypes. De plus, ces gènes sont régulés positivement par ces trois régimes riches en acides gras polyinsaturés d'une famille particulière (*Oméga 3* pour **dha** et **lin** et *Oméga 6* pour **tsol**) chez les souris WT, alors que la régulation de plusieurs de ces gènes est altérée chez les souris PPAR. Les gènes mHMGCoAS, PECI et BIEN apparaissent dans le contraste entre génotypes pour le régime **dha**, alors que les gènes S14 et FAS apparaissent pour le régime **efad**. Les souris des deux génotypes présentent là encore des régulations différentes de ces gènes, soulignant ainsi le rôle du récepteur PPAR $\alpha$  dans ces modulations d'expression provoquées par les régimes alimentaires.

Nous adoptons la même approche que précédemment, cette fois-ci sur les effets différentiels entre couples de régimes, à génotype fixé. L'ACP est donc ici réalisée sur le tableau de l'opposé du logarithme des *p*-values associées au différentiel d'expression des 57 gènes pour les dix différences entre les régimes pris deux à deux, à génotype fixé. La figure 11 présente le premier plan principal des gènes différentiellement exprimés entre régimes pour le génotype WT (à gauche) et pour le génotype PPAR (à droite). Les deux premiers axes, pour chacune des figures, représentent respectivement 79 % et 78 % de la variance totale. Les gènes Lpin et Lpin1 apparaissent dans des contrastes impliquant le régime **efad** pour le génotype WT et le régime **tsol** pour le génotype PPAR. Le gène CYP3A11 est impliqué dans le régime **dha**, quel que soit le génotype. Les gènes FAS et S14 apparaissent dans les contrastes impliquant le régime **efad** pour le génotype WT, alors que le même gène FAS apparaît dans les contrastes impliquant le régime **ref** pour le génotype PPAR. L'ensemble de ces résultats confirme les résultats obtenus avec l'ACP présentée en 3.4.

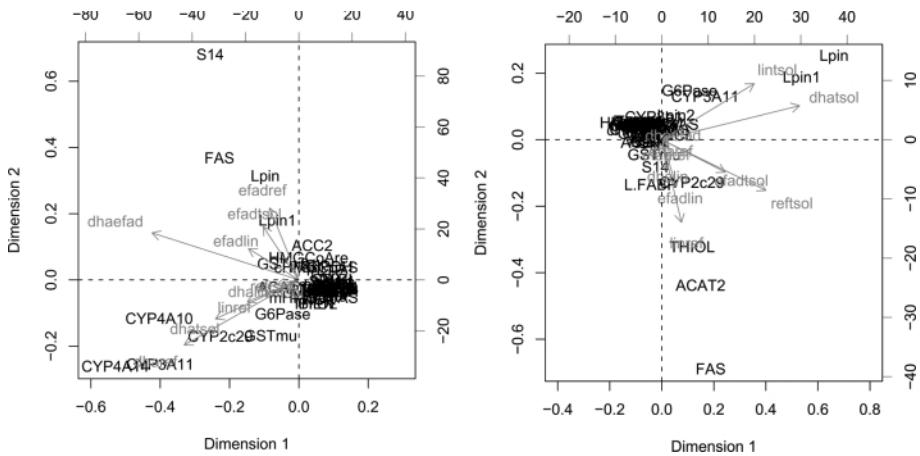


FIG 11. — Représentation sur le premier plan principal de l'ACP de l'opposé du logarithme des *p*-values des gènes différentiellement exprimés entre les régimes pour le génotype WT à gauche et PPAR à droite

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

### 4.1.3 Modèle mixte

Les souris étant issues d'une lignée consanguine, elles ont été considérées, dans un premier temps, comme des répétitions indépendantes et identiquement distribuées. Cependant, à l'aide d'un modèle linéaire mixte, chaque souris peut être considérée comme un tirage aléatoire dans une population plus large de souris. Le modèle linéaire mixte mis en œuvre s'écrit

$$y_{ijkl} = g_i + r_j + G_k + gr_{ij} + gG_{ik} + rG_{jk} + grG_{ijk} + souris_l + e'_{ijkl}, \quad (6)$$

où  $souris_l$  représente l'effet aléatoire de la souris  $l$ , avec  $souris_l \sim \mathcal{N}(0, \sigma_s^2)$ , les différentes réalisations étant indépendantes, et les  $e'_{ijkl}$  représentent les résidus, avec  $e'_{ijkl} \sim \mathcal{N}(0, \sigma_e^2)$ , les résidus étant indépendants entre eux et indépendants de l'effet aléatoire souris.

Dans ce cas, les estimations des composantes de la variance sont pour la variance «souris» de 0.001 et pour la variance résiduelle de 0.007. La variabilité individuelle est donc très faible. La variance des observations est identique à celle obtenue à l'aide d'une ANOVA (modèle à effets fixes) puisque nous sommes dans le cadre d'un plan équilibré et que la méthode d'estimation pour le modèle mixte est la méthode du maximum de vraisemblance restreinte (REML). Pour les mêmes raisons, les tests mis en œuvre dans ce modèle mixte pour tester la significativité des différents effets fixes sont identiques à ceux réalisés avec le modèle (2); les gènes mis en évidence sont donc les mêmes. Notons encore qu'il est possible d'étendre ce modèle aux cas de variances résiduelles hétérogènes, comme c'était le cas dans le modèle (5).

L'application du modèle linéaire mixte est beaucoup plus appropriée lorsque la variabilité due à la technique, à la diversité génétique ou aux gènes de la biopuce, présente un intérêt. C'est le cas dans l'étude transcriptomique décrite dans Bonnet *et al.* (2004) concernant la folliculogénèse chez la truie, dans laquelle le logarithme du signal est modélisé en fonction des facteurs membrane, animal, aiguille (ou bloc), jour d'hybridation, et des covariables logarithme de l'intensité du bruit de fond et de l'hybridation en sonde vecteur. Après une étape de choix de modèle (à l'aide du test de Fisher), le modèle linéaire mixte permet d'appréhender et de quantifier la part de variabilité due aux différentes sources de variation. Dans cet exemple relatif aux truies, la part de variabilité due à la diversité génétique représente 8 %, celle due à la technique 4 % et celle due aux gènes 75 %. Toute inférence basée sur ce modèle sera valide pour tout animal, toute membrane... car l'échantillonnage des animaux, des membranes... de cette étude, dans une population plus large d'animaux, de membranes... est pris en compte. Considérer les membranes (par exemple) comme effets fixes dans ce modèle aurait entraîné des conclusions valides uniquement sur les membranes de l'expérience. De plus, une structure de covariance non diagonale est prise en compte par ce modèle mixte puisque deux signaux d'une même membrane seront corrélés, la corrélation étant égale à  $\sigma_{membrane}^2 / \sigma_{totale}^2$ .

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

### 4.2. Forêts aléatoires

Les techniques précédemment décrites et utilisées (ACP, MDS, modèles gaussiens) sont, par construction, linéaires et, pour les approches de type ANOVA, reposent principalement sur une significativité d'ajustement d'un modèle. Une autre approche (Speed, 2003) est basée sur la recherche d'un sous-ensemble de gènes susceptibles de construire le meilleur modèle de discrimination des génotypes, ou des régimes, à partir des expressions. Parmi les approches non linéaires, certaines semblent relativement prometteuses, à condition de satisfaire, avant tout, une règle incontournable de parcimonie pour éviter un sur-ajustement, ou sur-apprentissage de l'échantillon. Analyse discriminante linéaire ou quadratique, modèles polynomiaux, logistiques, réseaux neuronaux, arbres de classification,  $k$  plus proches voisins... s'en accommodent mal. Parmi les nombreuses approches non linéaires récentes qui se préoccupent de la complexité du modèle (*support vector machine* ou SVM, *boosting*, *bagging*...) nous avons privilégié les forêts aléatoires (*random forest*) proposées par Breiman (2001) qui sont construites par agrégation d'arbres de classification. La même démarche pourrait être mise en œuvre avec les SVM.

Le modèle de forêt aléatoire est un exemple de *bagging*, c'est-à-dire un algorithme d'agrégation de modèles, spécialement adapté aux arbres de décision. Son principe est simple; il associe deux niveaux de randomisation :

1. chaque élément de la forêt est un arbre de décision de type CART (Breiman *et al.*, 1984) estimé sur un échantillon *bootstrap* de l'échantillon initial;
2. au cours de la construction de l'arbre, la dichotomie d'un noeud, qui vise à réduire au mieux l'entropie au sein de ce noeud, est optimisée sur un sous-ensemble restreint des variables explicatives (toujours le même nombre) tiré au hasard.

Par une simple moyenne pondérée dans le cas de la régression et un vote majoritaire dans le cas de la discrimination, la prédiction d'une forêt est calculée à partir de l'ensemble des arbres estimés pour chaque échantillon *bootstrap*. À chaque itération, le calcul d'une erreur *out of bag*, c'est-à-dire estimée sur les observations exclues du tirage *bootstrap*, permet de contrôler l'évolution de l'apprentissage.

Comme pour beaucoup de méthodes « boîte noire » ( $k$  plus proches voisins, réseau de neurones, *bagging*, *boosting*, SVM), une forêt d'arbres n'est pas un modèle explicite dans lequel il est facile d'interpréter le rôle des variables. La stratégie proposée par Breiman (2001) consiste alors à rechercher un indicateur d'importance de chacune des variables explicatives. Le principe est simple et efficace, même s'il requiert des calculs lourds : plus la qualité de la prédiction se dégrade quand on remplace les valeurs d'une variable par un réarrangement aléatoire de ses valeurs, plus cette variable est considérée comme importante. Un tel indicateur possède, bien entendu, des faiblesses. Prenons le cas de deux variables importantes et fortement corrélées; la perturbation d'une de ces variables peut ne conduire qu'à une faible dégradation de la qualité de la prédiction tant que la présence de l'autre variable la supplée.

# ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

En résumé, une forêt aléatoire est un modèle qui approche des surfaces de discrimination non linéaires dans l'espace des individus, sans pour autant sur-ajuster l'échantillon d'apprentissage, tandis que l'indice d'importance souligne les variables, ici les gènes, qui, en moyenne sur l'ensemble des tirages *bootstrap*, contribuent le mieux à discriminer les classes de la variable qualitative à prévoir.

Dans le cas élémentaire de la discrimination des génotypes des souris, les gènes qui apparaissent les plus significatifs sont, par ordre décroissant : PMDC1, CAR1, THIOL, L.FABP, ALDH3, CYP3A11, PEC1, GK, CYP4A10, ACBP, FAS, CPT2, BSEP, mHMGCoAS, ACOTH. Avec ceux-ci, la prédiction des génotypes est presque sûre avec une estimation (*out of bag*) de l'erreur de prédiction de 2 %.

En revanche, la discrimination des régimes, beaucoup plus délicate, a été traitée conditionnellement au génotype. Le régime **ref** est dans les deux cas le plus difficile à reconnaître.

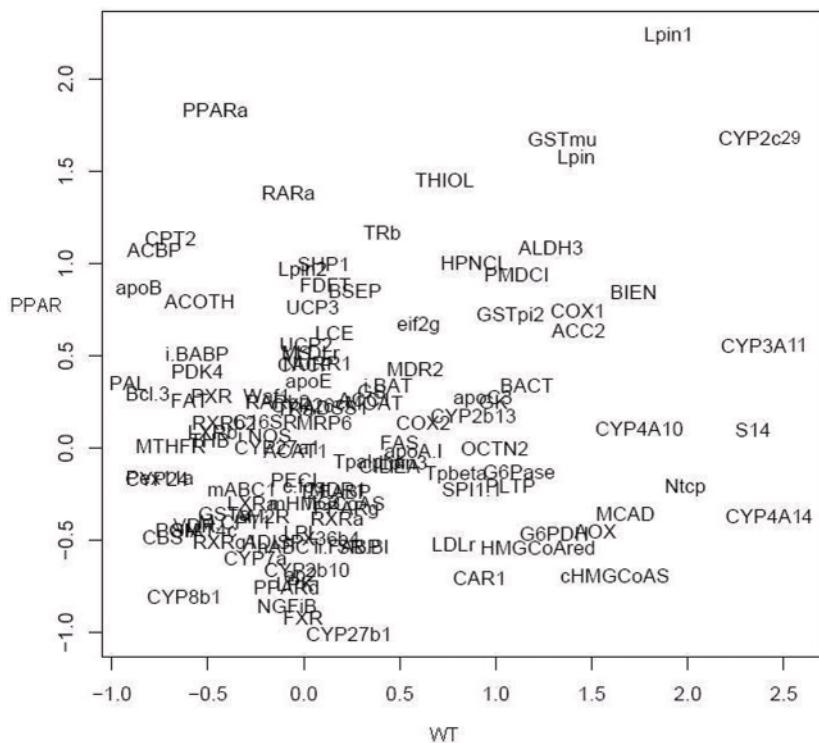


FIG 12. — Représentation des gènes en fonction de leur importance pour la discrimination des régimes à génotype fixé (WT sur l'axe horizontal et PPAR sur l'axe vertical)

La figure 12 représente les gènes en fonction de leur importance pour la discrimination des régimes pour chacun des génotypes. C'est pour les souris

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

PPAR que la discrimination des régimes est la plus difficile, ce qui confirme les observations déjà faites sur les résultats de la classification représentés par MDS (Fig. 9). Ce résultat s'interprète sur le plan biologique comme une implication du récepteur PPAR $\alpha$  dans les régulations géniques provoquées par les régimes alimentaires.

## 5. Relations entre expressions des gènes et variables cliniques

Dans diverses applications biologiques, l'interprétation des données d'expression génique peut être avantageusement complétée par leur rapprochement avec des données mesurées par ailleurs sur les mêmes individus.

L'objectif général de l'analyse canonique (voir, par exemple, Mardia *et al.*, 1979) est d'explorer les relations pouvant exister entre deux groupes de variables quantitatives observées sur le même ensemble d'individus. Nous l'avons mise en œuvre pour étudier les relations entre certains gènes et certains acides gras hépatiques de la souris.

Au préalable, une sélection des gènes et des acides gras a été nécessaire. Elle est décrite dans le point 5.1.

### 5.1. Sélections préliminaires

Les calculs effectués dans l'analyse canonique imposent de disposer d'un nombre de variables dans chaque groupe inférieur au nombre d'individus (40 dans notre exemple). Lors d'études préliminaires, nous avons de plus constaté que le nombre de variables devait être sensiblement inférieur au nombre d'individus. En effet, lorsque le nombre de variables est relativement important (disons supérieur à 25 dans notre cas), nous nous retrouvons avec plusieurs corrélations canoniques égales à 1 : la somme des dimensions des deux sous-espaces considérés dépassant la dimension de l'espace des variables, l'intersection de ces sous-espaces contient nécessairement des dimensions sans signification concrète qui correspondent aux plus grandes corrélations canoniques.

#### 5.1.1 Sélection des gènes

Nous repassons en revue les différentes méthodes mises en œuvre dans les deux paragraphes précédents (3 et 4) afin d'extraire un groupe de gènes particulièrement intéressant vis-à-vis des facteurs expérimentaux. Cette synthèse est une façon indirecte mais, nous semble-t-il, efficace de répondre à la problématique de la détection de gènes différentiellement exprimés.

#### ACP.

Comme nous l'avons déjà évoqué, dans le cadre de l'ACP, il est possible de calculer la contribution des variables-gènes aux axes sélectionnés afin notamment d'alléger les représentations graphiques en ne représentant que les gènes qui contribuent le plus à ces axes. Dans une optique de sélection de

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

gènes, une représentation de la distribution de ces contributions permet dans notre exemple de détecter 12 gènes dont les contributions apparaissent comme *outlier* dans le diagramme en boîte (Fig. 13).

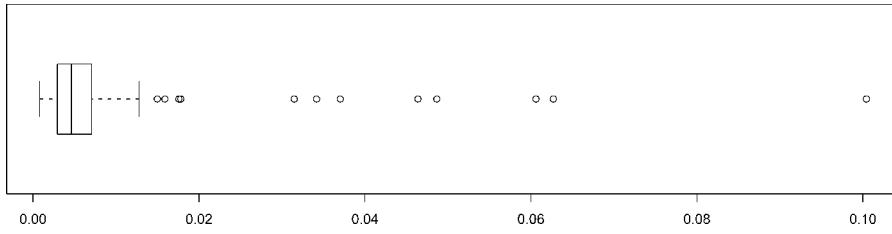


FIG 13. — Diagramme en boîte des contributions des gènes à la définition des trois premières composantes principales

Il s'agit des gènes suivants (dans l'ordre décroissant de leur contribution aux trois premiers axes) :

FAS, CYP4A14, CYP3A11, CYP4A10, THIOL, S14, PMDCI, L.FABP,  
mHMGCoAS, Lpin, Lpin1, GK.

Si on ne se limite pas à ces *outliers*, on peut étendre la liste aux 30 gènes contribuant le plus et ajouter ainsi :

PECI, CAR1, G6Pase, PAL, cHMGCoAS, BIEN, GSTpi2, ACC2, GSTmu,  
AOX, ALDH3, Lpin2, HPNCL, HMGCoAred, Tbeta, TRa, SHP1, ACBP.

### Classifications.

Pour sélectionner des gènes, nous nous basons essentiellement sur une interprétation visuelle de la carte induite par les classifications des lignes et des colonnes (Fig. 7). En effet, il nous semble pertinent de retenir les gènes pour lesquels l'aspect contrasté des colonnes est dominant, révélant ainsi une hétérogénéité des expressions géniques selon le génotype et le régime des individus-souris. Ces gènes sont également caractérisés par un regroupement «tardif» à l'ensemble des autres gènes au cours du déroulement de la classification hiérarchique ascendante, ce qui correspond à une expression sensiblement différente des autres. Ces 14 gènes sont :

PMDCI, THIOL, CYP3A11, CYP4A10, L.FABP, CYP4A14, FAS, Lpin,  
Lpin1, CYP2c29, GSTmu, GSTpi2, G6Pase, S14.

Contrairement aux autres méthodes, la liste des gènes sélectionnés est ici quasiment imposée par la lecture du graphique (Fig. 7). Il ne semblerait donc pas pertinent de vouloir étendre cette liste à des gènes supplémentaires.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

### Modélisation.

En procédant à une ANOVA selon le modèle (5) proposé dans le paragraphe 4, une façon de fournir une vision synthétique consiste à trier les gènes dans l'ordre croissant du minimum des 3 *p-values* associées aux doubles interactions (gène\*génotype et gène\*régime) et à l'interaction d'ordre 3 (gène\*génotype\*régime). Les 30 premiers gènes de ce classement sont :

PMDCI, THIOL, CYP3A11, CYP4A10, ALDH3, CAR1, HPNCL, L.FABP,  
mHMGCoAS, GK, PECI, CYP4A14, CPT2, BIEN, FAS, ACAT2, Lpin,  
Lpin1, Lpin2, CYP2c29, GSTmu, HMGCoAre, ACC2, ACBP, AOX, CACP,  
COX1, GSTpi2, PLTP, S14.

### Forêts aléatoires.

Dans cette partie, nous n'utilisons que les résultats des forêts aléatoires à génotype fixé. En effet, presque tous les gènes mis en évidence par la discrimination des génotypes sont déjà signalés par les autres méthodes.

La sélection de gènes que nous proposons à partir de l'importance calculée par les forêts aléatoires consiste à extraire les gènes au-delà d'un arc de cercle centré en  $(-1; -1)$  et de rayon approximatif 2.7 sur la figure 12. Cela correspond à sélectionner les gènes dont la racine carrée des sommes des carrés des importances est supérieure à un seuil. Par ce moyen, nous sélectionnons les 20 gènes suivants :

PMDCI, THIOL, CYP3A11, CYP4A10, ALDH3, HPNCL, CYP4A14, BIEN,  
Lpin, Lpin1, CYP2c29, GSTmu, COX1, ACC2, GSTpi2, Ntcp, S14,  
PPAR<sub>a</sub>, RAR<sub>a</sub>, TR<sub>b</sub>.

### Synthèse.

La synthèse de la procédure de sélection de gènes selon les différentes méthodes est fournie dans le tableau 5. La première partie de ce tableau correspond aux 10 gènes sélectionnés par les quatre méthodes, la deuxième aux 7 gènes sélectionnés par trois méthodes sur quatre (à l'intérieur de chaque groupe, l'ordre est celui donné par la modélisation).

#### 5.1.2 Sélection des acides gras

Sans reprendre l'intégralité du traitement appliqué aux gènes, nous nous sommes attachés à gommer les fortes corrélations entre acides gras pouvant perturber l'analyse canonique. La démarche adoptée consiste à combiner les résultats d'une classification ascendante hiérarchique et d'une ACP.

Le dendrogramme de la figure 14 a été construit en utilisant la distance basée sur la corrélation ( $d_3$ ) et le critère du saut moyen pour l'agglomération. Nous avons également procédé à l'agglomération selon le saut de Ward ; la classification ainsi produite entraîne les mêmes interprétations. Nous avons choisi ici le critère du saut moyen qui, associé à la distance basée sur la corrélation, permet d'interpréter simplement l'échelle verticale du dendrogramme (un moins la corrélation entre les différents lipides).

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

TABLEAU 5. – Synthèse de la sélection des gènes selon les différentes méthodes. Seuls les gènes détectés par au moins trois méthodes sont présents dans le tableau. Le symbole \* indique que le gène est détecté par la méthode de la colonne correspondante

Gène	ACP	Classification	Forêts aléatoires	ANOVA
PMDCI	*	*	*	*
THIOL	*	*	*	*
CYP3A11	*	*	*	*
CYP4A10	*	*	*	*
CYP4A14	*	*	*	*
Lpin	*	*	*	*
Lpin1	*	*	*	*
GSTMu	*	*	*	*
GSTpi2	*	*	*	*
S14	*	*	*	*
ALDH3	*		*	*
HPNCL	*		*	*
L.FABP	*	*	*	*
BIEN	*		*	*
FAS	*	*	*	*
CYP2c29	*	*	*	*
ACC2	*		*	*

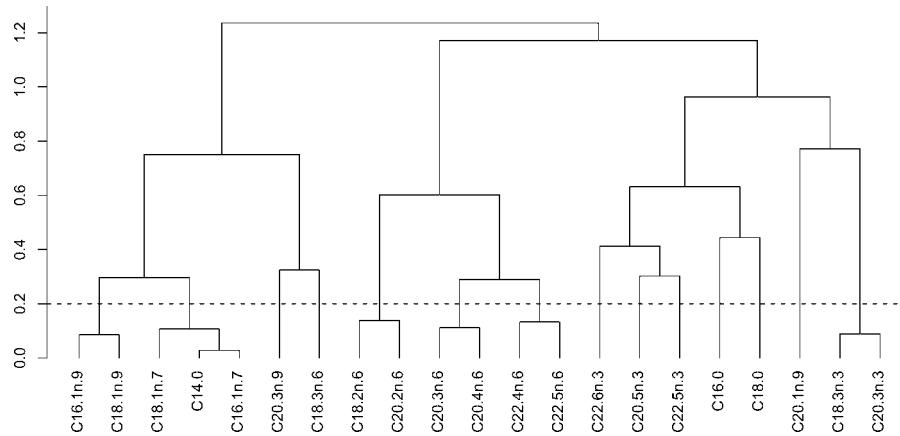


FIG 14. — Classification hiérarchique ascendante des acides gras selon  $d_3$  (basée sur la corrélation) et le critère du saut moyen. Les gènes regroupés sous la ligne horizontale pointillée sont corrélés à plus de 0.8

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

De manière synthétique, on remarquera que cette représentation des acides gras tend à regrouper les principales familles : acides gras monoinsaturés à gauche, famille *Oméga 6* au centre et famille *Oméga 3* à droite. Cela est d'autant plus vrai si l'on fait abstraction de quelques acides gras qui atteignent la limite de détection de la méthode de dosage (chromatographie en phase gazeuse) dans plusieurs groupes de souris (notamment les C20.3n-9 et C18.3n-6) et du C20.1n-9 dont la proportion ne varie pas de manière sensible entre les groupes. Sur le plan biologique, cette représentation reflète bien les connaissances actuelles de la voie de biosynthèse des acides gras polyinsaturés (Nakamura & Nara, 2004).

Coupé au niveau 0.2 pour signifier que les gènes regroupés sous cette limite sont corrélés à plus de 0.8, l'arbre de la figure 14 fournit une classification en 14 groupes. Parmi ces groupes, nous éliminons les acides gras C20.3n-9, C18.3n-6 et C20.1n-9 pour les raisons indiquées ci-dessus. Le principe de notre sélection consiste à ne conserver qu'un seul acide gras pour chacun des 11 groupes restants.

Pour effectuer ce choix, nous nous basons sur les résultats de l'ACP des acides gras.

TABLEAU 6. – Parts de variance expliquée et cumuls pour les quatre premières dimensions de l'ACP des acides gras

	Axe 1	Axe 2	Axe 3	Axe 4
Parts de variance expliquée	0.41	0.29	0.17	0.09
Cumuls	0.41	0.70	0.86	0.96

Le tableau des valeurs propres (Tab. 6) nous conduit à conserver quatre dimensions à partir desquelles ont été calculées les contributions (voir équation (1)) des acides gras que nous avons classées ci-dessous par ordre décroissant :

C18.2n – 6, C18.1n – 9, C18.3n – 3, C22.6n – 3, C20.4n – 6, C16.0,  
 C18.1n – 7, C16.1n – 7, C18.0, C20.5n – 3, C22.5n – 3, C14.0, C22.5n – 6,  
 C20.3n – 9, C20.3n – 6, C18.3n – 6, C16.1n – 9, C22.4n – 6, C20.2n – 6,  
 C20.3n – 3, C20.1n – 9.

Le représentant de chacun des 11 groupes est défini comme celui contribuant le plus aux quatre premiers axes de l'ACP à l'intérieur de son groupe.

Nous retenons finalement les 11 acides gras suivants :

C18.1n – 9, C18.1n – 7, C18.2n – 6, C20.4n – 6, C22.5n – 6, C22.6n – 3,  
 C20.5n – 3, C22.5n – 3, C16.0, C18.0, C18.3n – 3.

Afin de contrôler l'effet des deux facteurs génotype et régime, globalement sur les dix gènes choisis, ainsi que sur les onze acides gras sélectionnés, nous avons réalisé deux analyses de variance multidimensionnelles (MANOVA), à dix et onze dimensions respectivement. Dans chaque cas, on a obtenu des effets extrêmement significatifs, que ce soit pour le génotype, pour le régime ou pour les interactions (*p*-values inférieures à  $10^{-4}$  pour le test de Wilks).

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

### 5.2. Mise en œuvre de l'analyse canonique

Nous avons mis en œuvre deux analyses canoniques (AC) avec les onze acides gras sélectionnés précédemment : l'une avec les dix gènes retenus par les quatre méthodes de sélection, l'autre avec les dix-sept gènes retenus par au moins trois méthodes.

Dans la représentation graphique des variables, nous avons ajouté deux cercles : le cercle des corrélations, de rayon unitaire, justifié par la représentation des variables à partir des corrélations variables-facteurs; le cercle de rayon 0,5, pour faciliter la lecture en mettant en évidence les phénomènes les plus intéressants dans la couronne ainsi définie.

Dans la première analyse, les 10 corrélations canoniques sont les suivantes :

0.96 0.93 0.91 0.86 0.79 0.72 0.61 0.41 0.25 0.04

Les figures 15 et 16 représentent variables et individus dans les trois premières dimensions pour lesquelles les corrélations canoniques sont supérieures à 0.90.

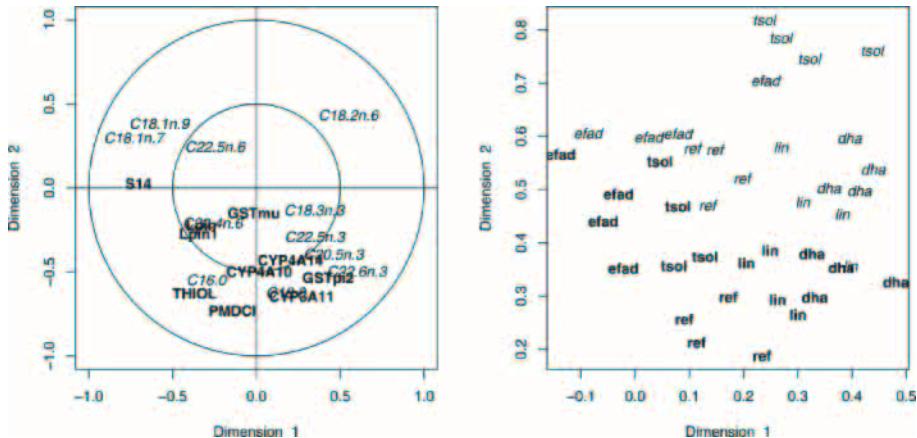


FIG 15. — Représentation des variables (10 gènes en gras, 11 acides gras en italique), à gauche et des individus (WT en gras et PPAR en italique), à droite, dans le premier plan canonique

Le premier plan de cette analyse canonique (Fig. 15) permet de résumer de manière originale les principaux résultats de cette étude. Dans le graphique des individus, on observe la séparation des deux génotypes, l'opposition entre le régime **efad** (faibles coordonnées sur l'axe 1) et les régimes riches en *Oméga 3*, **lin** et **dha** (fortes coordonnées sur l'axe 1), enfin la position particulière du régime **tsol** chez les souris PPAR. On retrouve, sur le graphique des variables, les regroupements de familles d'acides gras évoqués au paragraphe 5.1.2. La séparation des génotypes est principalement liée à l'accumulation préférentielle du **C18.2n-6** chez les souris PPAR au détriment du **C16.0**, du **C18.0** et des acides gras longs polyinsaturés **C20.5n-3** et **C22.6n-3** (DHA), ainsi

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

qu'à la plus forte expression des gènes THIOL, PMDCI, CYP3A11 et GSTpi2 chez les souris WT par rapport aux souris PPAR. On note les proximités entre le C16.0 et le gène THIOL, ainsi que les proximités entre CYP3A11 et GSTpi2 et les acides gras C18.0 et C22.6n-3. L'opposition entre le régime *efad* et les régimes *lin* et *dha* est liée aux particularités suivantes du régime *efad* : accumulation d'acides gras monoinsaturés (C18.1n-9 et C18.1n-7) chez les souris des deux génotypes (mais plus marquée chez les souris PPAR) et surexpression du gène S14 presque exclusivement chez les souris WT (d'où un décalage entre les acides gras monoinsaturés et S14 le long de l'axe 2). Sous régime riche en *Oméga 3* (*lin* et *dha*), on observe une accumulation préférentielle des acides gras C20.5n-3 (surtout pour le régime *lin*), C22.6n-3 (surtout pour le régime *dha*) et C18.0 accompagnée de régulations positives des gènes GSTpi2, CYP3A11 et des CYP4A qui, cependant, se révèlent moins marquées, voire absentes, chez les souris PPAR. Enfin, la position particulière du régime *tsol* chez les souris PPAR est liée à l'accumulation extrêmement marquée de C18.2n-6 dans le foie de ces souris sous le régime *tsol* (sous ce régime, la proportion de C18.2n-6 est presque deux fois plus importante chez les souris PPAR que chez les souris WT), soulignant ainsi le rôle primordial de PPAR $\alpha$  dans la prise en charge de cet acide gras, que ce soit pour sa dégradation ou pour son utilisation dans la biosynthèse des acides gras longs polyinsaturés de la famille *Oméga 6*. Le troisième axe de cette analyse canonique (Fig. 16) n'apporte qu'une information vraiment nouvelle par rapport aux deux premiers axes, c'est la régulation positive du gène GSTmu chez les souris des deux génotypes (à nouveau, elle est moins marquée chez les souris PPAR) soumises au régime *dha*. Cette régulation n'est en revanche pas observée sous le régime *lin*.

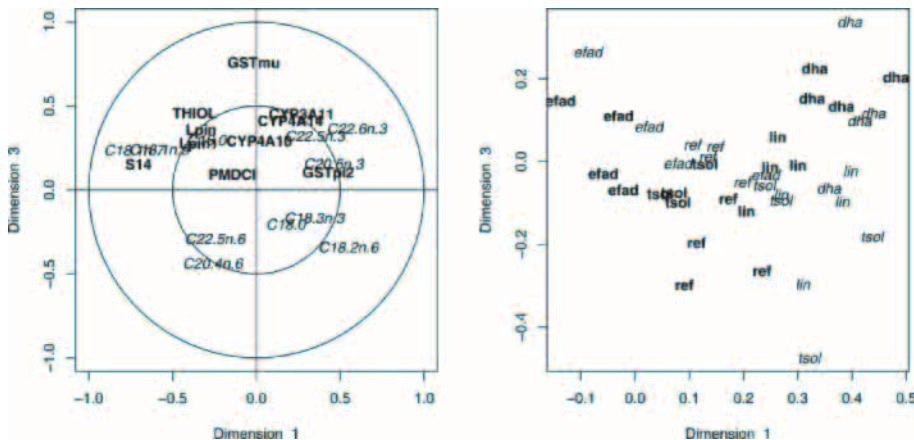


FIG 16. — Représentation de l'AC des 10 gènes et 11 acides gras dans le plan canonique 1-3 avec les mêmes conventions que pour la figure 15

Dans la seconde analyse canonique, les 11 corrélations sont :

0.98 0.98 0.94 0.93 0.93 0.87 0.83 0.76 0.72 0.68 0.53 0.44

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

L'information pertinente se concentre sur les quatre premières dimensions. En particulier, sur les dimensions cinq et six, toutes les variables sont projetées à l'intérieur du petit cercle.

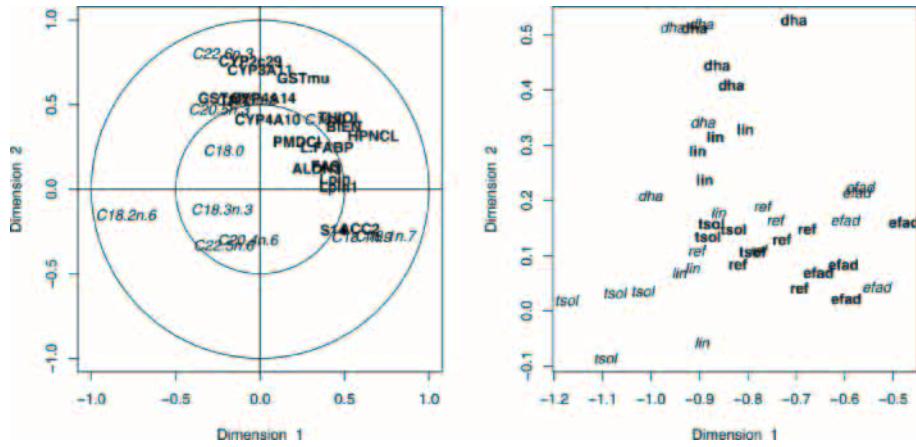


FIG 17. — Représentation de l'AC des 17 gènes et 11 acides gras sur le premier plan canonique avec les mêmes conventions que pour la figure 15

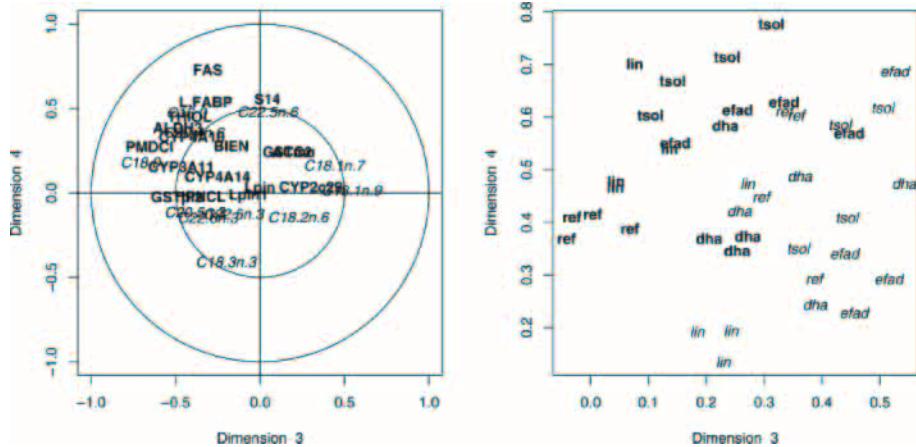


FIG 18. — Représentation de l'AC des 17 gènes et 11 lipides sur le plan 3-4 avec les mêmes conventions que pour la figure 15

Cette deuxième analyse canonique (Fig. 17 et 18) nous apporte des informations complémentaires grâce à l'ajout de plusieurs gènes qui se rapprochent de groupes de gènes et acides gras déjà évoqués lors de l'étude de la première analyse. Ainsi, dans le plan 1-2 (Fig. 17), on note que CYP2c29 se comporte de manière assez similaire à CYP3A11 et GST $\mu$  et, dans une moindre mesure, à GSTpi2 et CYP4A14. Ces gènes codant des cytochromes P450 et des glutathion-S-transférases sont impliqués de manière globale dans les

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

phénomènes de détoxication. Leurs niveaux d'expression semblent reliés dans cette étude aux proportions d'acides gras long polyinsaturés de la famille Oméga 3 comme le C22.6n-3 (DHA), le C22.5n-3 et le C20.5n-3 (acide eicosapentanoïque, EPA). Comme nous l'avons évoqué, les régulations de ces gènes par les régimes alimentaires dha et lin sont généralement plus marquées chez les souris WT que chez les souris PPAR. Il est particulièrement rassurant de constater que plusieurs auteurs ont déjà observé les régulations de certains de ces gènes par ces mêmes acides gras (Berger et al., 2002; Chen et al., 2001) et que ces acides gras sont connus pour être des activateurs du récepteur PPAR $\alpha$  (Forman et al., 1997; Murakami et al., 1999). De manière similaire, les gènes BIEN et HPNCL se rapprochent de THIOL et du C16.0. Ces trois gènes sont impliqués dans le catabolisme des acides gras par  $\alpha$ - ou  $\beta$ -oxydation. En revanche, leur corrélation avec la proportion de C16.0 constitue une donnée nouvelle qui nécessitera des expériences complémentaires pour être complètement comprise. Enfin, le gène ACC2 est proche du gène S14 et tous deux sont impliqués dans la lipogénèse. Leurs niveaux d'expression ainsi que l'abondance des acides gras monoinsaturés C18.1n-7 et C18.1n-9 sont maximaux avec le régime efad. Le plan 3-4 (Fig. 18) nous permet principalement de compléter la liste des gènes présentant des expressions plus fortes chez les souris WT par rapport aux souris PPAR en ajoutant les gènes FAS, L.FABP et ALDH3.

Ainsi, l'analyse canonique fournit au biologiste un outil intéressant pour mieux appréhender de manière conjointe deux jeux de variables sélectionnées. Elle offre des représentations graphiques facilement lisibles dans lesquelles il est possible de puiser des informations pertinentes. Il est important de souligner que l'analyse canonique intervient presque nécessairement après des étapes d'exploration des données puis de sélection de variables. Ainsi, lorsqu'elle est mise en œuvre, le biologiste dispose déjà d'une bonne connaissance de ses deux jeux de variables, facilitant d'autant plus l'interprétation des graphiques. Enfin, notons que l'analyse canonique met en évidence des corrélations qui ne peuvent donc être directement interprétées comme des phénomènes de cause à effet. Dans nos exemples, il nous est impossible de distinguer les effets des variations d'expression génique sur les proportions d'acides gras, les effets des proportions d'acides gras sur l'expression des gènes ou encore la simple concomitance de variations d'expression génique et de proportions d'acides gras. En revanche, nous avons souligné que des données publiées permettaient de conforter certaines proximités de gènes et d'acides gras et que d'autres soulevaient des questions nouvelles nécessitant la mise en œuvre d'expériences complémentaires.

## 6. Conclusion et perspectives

Cette étude, qui se veut essentiellement didactique, permet de mettre en avant le fait qu'il n'existe pas une méthode unique permettant de traiter des données d'expression. La question « Quelle méthode dois-je utiliser pour traiter mes données d'expression ? » n'a pas de sens de façon générale. En revanche, à une question précise du type « Puis-je effectuer une partition des gènes ? »,

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

une méthode statistique (ici la classification) peut apporter des éléments de réponses par des sorties numériques et/ou graphiques. Toutefois, la réponse précise à la question ne peut être apportée que par une collaboration étroite entre le statisticien, pour son expertise des méthodes utilisées, et le praticien, pour son expertise des phénomènes biologiques considérés. Soulignons encore une fois l'importance du choix de la méthode et de celui des options (qui fournissent différentes optiques). Ces choix ne sont pas neutres sur le plan biologique et doivent être pris en compte dans l'interprétation des graphiques et des modèles obtenus selon chaque optique.

Dans cet article, nous avons surtout essayé de répondre aux questions posées par l'expérience biologique mise en œuvre. De nombreux autres problèmes soulevés par l'analyse des données transcriptomiques n'ont pas été abordés. Signalons, pour mémoire, le *design* des oligos, la mise en œuvre de plans d'expériences pour le recueil des données, la normalisation, d'autres gestions des tests multiples, l'étude de cinétiques d'expression... Sur ce dernier point, on trouvera quelques pistes sur la classification de profils temporels d'expression ainsi que des ouvertures vers l'ANOVA fonctionnelle dans Baccini *et al.* (2005).

Dans le cas d'une puce comportant plusieurs milliers de gènes (en particulier une puce pangénomique), les méthodes présentées, notamment les méthodes exploratoires, peuvent s'appliquer de manière analogue, même si leur interprétation est alors plus délicate. En revanche, l'analyse canonique nous semble plus spécifique au cas d'un nombre restreint de gènes.

La démarche adoptée dans ce travail n'a pas été de rechercher la « meilleure » méthode pour détecter les gènes différentiellement exprimés, mais de voir comment chaque approche renseigne sur la problématique biologique. Une synthèse des différentes approches est plus susceptible de faire converger un faisceau de présomptions sur un ensemble de gènes, avec l'idée qu'elle permet d'éviter des artefacts (faux positifs) contrairement à une approche unique.

**Remerciements :** Les auteurs souhaitent remercier Thierry Pineau et Hervé Guillou pour leur participation active à l'acquisition des données et pour les discussions scientifiques concernant leur interprétation. D'autre part, ils remercient les relecteurs qui ont permis de bien améliorer la qualité de l'article.

Cette recherche a été partiellement financée par l'ACI IMPBio *Développement d'un environnement dédié à l'analyse statistique des données d'expression* (ENV-STAT-EXP).

## Compléments

Diverses informations relatives aux activités des auteurs sur la thématique des biopuces sont accessibles en ligne à l'adresse

[www.lsp.ups-tlse.fr/Biopuces](http://www.lsp.ups-tlse.fr/Biopuces)

En particulier, on y trouvera les commandes R ayant permis de mettre en œuvre les analyses présentées dans cet article ainsi qu'un livret regroupant ces mêmes figures en couleur.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

### Références

- AKAIKE H. (1974). A new look at the statistical model identification, *IEEE Transaction on Automatic control*, AC-19, 716-723.
- AOYAMA T., PETERS J.M., IRITANI N., NAKAJIMA T., FURIHATA K., HASHIMOTO T., GONZALEZ F.J. (1998). Altered constitutive expression of fatty acid-metabolizing enzymes in mice lacking the peroxisome proliferator-activated receptor alpha (PPARalpha), *Journal of Biological Chemistry*, 273(10), 5678-84.
- AUBERT J., BAR-HEN A., DAUDIN J.-J., ROBIN S. (2004). Determination of the differentially expressed genes in microarray experiments using local FDR, *BMC Bioinformatics*, 5, 125.
- BACCINI A., BESSE Ph., DÉJEAN S., MARTIN, P., ROBERT-GRANIÉ C., SAN CRISTOBAL M. (2005). *Analyse statistique des données d'expression génomique*, support de formation disponible en ligne, [www.lsp.ups-tlse.fr/Biopuces](http://www.lsp.ups-tlse.fr/Biopuces).
- BACCINI A., BESSE Ph., DÉJEAN S., MARTIN, P., ROBERT-GRANIÉ C., SAN CRISTOBAL M. (2005). Étude de données cinétiques issues de biopuces, *XXXVII<sup>èmes</sup> Journées de Statistique*, juin 2005, Pau.
- BAR-HEN A., ROBIN S. (2003). An iterative procedure for differential analysis of gene expression, *Comptes rendus de l'Académie des sciences*, Série I, 337, 343-346.
- BENJAMINI Y., HOCHBERG Y. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society - Series B*, 85, 289-300.
- BERGER A., MUTCH D.M., GERMAN J.B., ROBERTS M.A. (2002). Dietary effects of arachidonate-rich fungal oil and fish oil on murine hepatic and hippocampal gene expression, *Lipids in Health and Disease*, 1 :2.
- BLAND J., ALTMAN D. (1995). Multiple significance tests : the Bonferroni method, *British medical Journal*, 310, 170.
- BONNET A., BENNE F., DANTEC C., GOBERT N., FRAPPART P.O., SAN CRISTOBAL M., HATEY F., TOSSER-KLOPP G. (2004). Identification of genes and gene networks involved in pig ovarian follicular development, by using c-DNA microarrays, *XIII International Workshop on the Development and Function of Reproductive organs*, 12-15 July 2004, Copenhagen, Denmark.
- BREIMAN L. (2001). Random forest, *Machine Learning*, 45, 5-32.
- BREIMAN L., FRIEDMAN J.H., OLSHEN R.A., STONE C.J. (1984). *Classification and regression trees*, Wadsworth & Brooks, Cole Advanced Books & Software.
- BROËT P., LEWIN A., RICHARDSON S., DALMASSO C., MAGDELENAT H. (2005). A mixture model-based strategy for selecting sets of genes in multiclass response microarray experiments, *Bioinformatics*, 20(16), 2562-2571.
- CHEN H.W., YANG J.J., TSAI C.W., WU J.J., SHEEN L.Y., OU C.C., LII C.K. (2001). Dietary fat and garlic oil independently regulate hepatic cytochrome p(450) 2B1 and the placental form of glutathione S-transferase expression in rats, *The Journal of Nutrition*, 131(5), 1438-43.
- DALMASSO C., BROËT P., MOREAU T. (2005). A simple procedure for estimating the false discovery rate, *Bioinformatics*, 21(5), 660-668.
- DELMAR P., ROBIN S., TRONIK-LEROUX D., DAUDIN J.-J. (2005). Mixture model on the variance for the differential analysis of gene expression data, *Journal of the Royal Statistical Society - Series C, Applied Statistics*, 54, 1-20.
- DRAGHICI S. (2003). *Data Analysis Tools for DNA Microarrays*, Mathematical Biology and Medicine Series, Chapman & Hall/CRC.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

- DUDOIT S., POPPER J., BOLDRICK J.C. (2003), Multiple Hypothesis Testing in Microarray Experiments, *Statistical Science*, 18(1), 71-103.
- DUDOIT S., YANG Y., SPEED T., CALLOW M. (2002), Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments, *Statistica Sinica*, 12(1), 111-139.
- EISEN M.B., SPELLMAN P.T., BROWN P.O., BOTSTEIN D. (1998), Cluster analysis and display of genome-wide expression patterns, *Proceedings of the National Academy of Sciences of the USA*, 95(25), 14863-14868.
- FORMAN B.M., CHEN J., EVANS R.M. (1997). Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta, *Proceedings of the National Academy of Sciences of the USA*, 94(9), 4312-7.
- FOULLEY J.-L., JAFFREZIC F., ROBERT-GRANIÉ C. (2000), EM-REML estimation of covariances parameters in Gaussian mixed models for longitudinal data analysis, *Genetics Selection Evolution*, 32, 129-141.
- KERR K., MARTIN M., CHURCHILL G. (2000). Analysis of variance for gene expression microarray data, *Journal of Computational Biology*, 7, 819-837.
- LEBART L., MORINEAU A., PIROU M. (1995). *Statistique exploratoire multidimensionnelle*, Dunod.
- LEE S.S., PINEAU T., DRAGO J., LEE E.J., OWENS J.W., KROETZ D.L., FERNANDEZ-SALGUERO P.M., WESTPHAL H., GONZALEZ F.J. (1995). Targeted disruption of the alpha isoform of the peroxisome proliferator-activated receptor gene in mice results in abolishment of the pleiotropic effects of peroxisome proliferators, *Molecular and Cellular Biology* 15(6), 3012-22.
- MARDIA K., KENT J., BIBBY J. (1979). *Multivariate analysis*, Academic Press.
- MARTIN P.G.P., LASSERRE F., CALLEJA C., VAN ES A., ROULET A., CONCORDET D., CANTIËLLO M., BARNOUIN R., GAUTHIER B., PINEAU T. (2005a). Transcriptional modulations by RXR agonists are only partially subordinated to PPAR $\alpha$  signaling and attest additional, organ-specific, molecular cross-talks, *Gene Expression*. Voir également le matériel supplémentaire associé à l'article à l'adresse : <http://www.inra.fr/Internet/Centres/toulouse/pharmacologie/pharmaco-moleculaire/valorisations>
- MARTIN P.G.P. et al. (2005b). A nutrigenomic approach in mice reveals new aspects of PPAR $\alpha$ -deficient phenotype with important implications in pharmacology, en préparation.
- MCLACHLAN G.J., DO K.-A., AMBROISE C. (2004). *Analysing microarray gene expression data*, Wiley.
- MURAKAMI K., IDE T., SUZUKI M., MOCHIZUKI T., KADOWAKI T. (1999). Evidence for direct binding of fatty acids and eicosanoids to human peroxisome proliferators-activated receptor alpha, *Biochemical and Biophysical Research Communications*, 260(3), 609-13.
- NAKAMURA M.T., NARA T.Y. (2004). Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases. *Annual Review of Nutrition*, 24, 345-76.
- NGUYEN D.V. (2004). On estimating the proportion of true null hypothesis for false discovery rate controlling procedures in exploratory DNA microarrays studies. *Computational Statistics and Data Analysis*, 47, 611-637.
- PETERFY M., PHAN J., XU P., REUE K. (2001). Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin, *Nature Genetics*, 27(1), 121-4.

## ANALYSE STATISTIQUE DE DONNÉES TRANSCRIPTOMIQUES

- ROBERT-GRANIÉ C., BONAITI B., BOICHARD D., BARBAT A. (1999). Accounting for variance heterogeneity in French dairy cattle genetic evaluation, *Livestock Production Science*, 60, 343-357.
- SAN CRISTOBAL M., ROBERT-GRANIÉ C., FOULLEY J-L. (2002). Hétéroscédasticité et modèles linéaires mixtes : théorie et applications en génétique quantitative, *Journal de la Société Française de Statistique*, 143, 1-2.
- SAPORTA G. (1990). *Probabilités analyse des données et statistique*, Paris : Éditions Technip.
- SCHWARZ G. (1978). Estimating the dimension of a model, *Annals of Statistics*, 6, 461-464.
- SPEED T. (2003). *Statistical Analysis of Gene Expression Microarray Data*, Interdisciplinary Statistics, Chapman & Hall/CRC.

## Chapitre IV : Etude d'agonistes synthétiques de RXR

### IV.1. INTRODUCTION

Ce projet de recherche a été initié dans le cadre d'une collaboration entre l'équipe de Pharmacologie Moléculaire de l'UR66 (Dr T. Pineau, Toulouse) et les laboratoires Galderma (Dr. B. Gauthier, Sophia Antipolis). Ces laboratoires, spécialisés dans le développement de molécules à visée thérapeutique dans le domaine de la dermatologie s'intéressaient de près aux effets des agonistes de RXR (voir chapitre I). Jusqu'alors, les principales applications dermatologiques concernaient des agonistes peu spécifiques des récepteurs RAR (pommade Roaccutane contenant de l'isotrétinoïne pour le traitement de l'acné par exemple, voir aussi [Table 6](#)) mais s'accompagnaient de lourds effets secondaires et de fortes contre-indications, notamment en raison de leurs effets tératogènes. Des agonistes synthétiques spécifiques des RXR ont donc été développés dans le but d'éviter ces effets délétères liés principalement à l'activation des RAR et de bénéficier des effets bénéfiques de l'activation des RXR (effets positifs escomptés sur l'acné, le psoriasis et certains cancers cutanés en raison du potentiel d'inhibition de la prolifération associé à l'activation de ces RNs). La littérature suggérait alors que l'activation de PPAR $\alpha$  par un agoniste spécifique provoquait son association avec RXR (on sait maintenant que cette association est préétablie, voir chapitre II) et des indications suggéraient que l'activation de RXR par des agonistes spécifiques pouvait également conduire au recrutement et à l'activation du récepteur PPAR $\alpha$ , au moins dans certains organes (Mukherjee et al., 1998). Cependant, des observations issues de l'étude de différents activateurs de RXR suggéraient également la capacité de ces molécules à accroître, au moins de manière transitoire, les triglycérides plasmatiques (Ouamrane et al., 2003, et travaux de doctorat non publiés de L. Ouamrane), un mécanisme qui a priori n'était pas dépendant de PPAR $\alpha$  dont l'activation a généralement l'effet inverse. De plus, ces molécules provoquent une hépatomégalie indépendamment de l'expression de PPAR $\alpha$  (Ouamrane et al., 2003), suggérant le recrutement potentiel par RXR d'autres voies de signalisation, au moins au niveau hépatique. Nous avons donc souhaité évaluer quels étaient les effets tissu-spécifiques (foie, rein et cœur) sur l'expression d'un panel de gènes cibles de PPAR $\alpha$  et d'autres partenaires de RXR (puces INRAarray 01.2), d'un agoniste de RXR utilisé en thérapeutique humaine pour le traitement de lymphomes cutanés à cellules T (le LGD1069), en le comparant à un agoniste de PPAR $\alpha$  (le fenofibrate) et en utilisant les souris PPAR $\alpha$ -/- pour discerner aisément les modulations d'expression dépendantes et indépendantes de ce

récepteur. La mise en œuvre et les résultats de cette étude (Martin et al., 2005) sont présentés ci-après.

#### **IV.2. PRESENTATION DES RESULTATS PUBLIES**

**Article n°2 :** Pascal G. P. Martin, Frédéric Lasserre, Cécile Calléja, Armelle van Es, Alain Roulet, Didier Concorde, Michela Cantiello, Romain Barnouin, Béatrice Gauthier, Thierry Pineau. Transcriptional modulations by RXR agonists are only partially subordinated to PPAR $\alpha$  signalling and attest additional, organ-specific, molecular cross-talks. *Gene Expression*. 2005 ; 12: 177-192.

# Transcriptional Modulations by RXR Agonists Are Only Partially Subordinated to PPAR $\alpha$ Signaling and Attest Additional, Organ-Specific, Molecular Cross-Talks

PASCAL G. P. MARTIN,\* FRÉDÉRIC LASSERRE,\* CÉCILE CALLEJA,\*<sup>1</sup> ARMELLE VAN ES,\*<sup>1</sup>  
ALAIN ROULET,\* DIDIER CONCORDET,† MICHELA CANTIELLO,‡ ROMAIN BARNOUIN,\*<sup>1</sup>  
BÉATRICE GAUTHIER,§ AND THIERRY PINEAU\*

\**Laboratoire de Pharmacologie et Toxicologie, I.N.R.A., BP3, Toulouse, France*

†*UMR 181, I.N.R.A./E.N.V.T., Toulouse, France*

‡*Dipartimento di Patologia Animale, Università degli Studi di Torino, Torino, Italy*

§*Galderma R&D, Sophia-Antipolis, France*

Nuclear hormone receptors (NR) are important transcriptional regulators of numerous genes involved in diverse pathophysiological and therapeutic functions. Following ligand activation, class II NR share the ability to heterodimerize with the retinoid X receptor (RXR). It is established that RXR activators, retinoids, transactivate several peroxisome proliferator-activated receptor  $\alpha$  (PPAR $\alpha$ ) target genes in a PPAR $\alpha$ -dependent manner. We hypothesized that, once activated, RXR might signal through quiescent NR other than PPAR $\alpha$ , in an organ-specific manner. To study this putative phenomenon *in vivo*, we developed an array of 120 genes relevant to the class II NR field. The genes were selected using both published data and high-density screenings performed on RXR or PPAR $\alpha$  agonist-treated mice. Wild-type C57BL/6J and PPAR $\alpha$ -deficient mice were treated with fenofibrate (PPAR $\alpha$  activator) or LGD1069 (RXR activator). Using our customized array, we studied the hepatic, cardiac, and renal expression of this panel of 120 genes and compared them in both murine genotypes. The results obtained from this study confirmed the ability of an RXR agonist to modulate PPAR $\alpha$ -restricted target genes in the liver and the kidney. Furthermore, we show that various organ-specific regulations occurring in both genotypes (PPAR $\alpha$   $+/+$  or  $-/-$ ) are highly indicative of the ability of RXR to recruit other class II NR pathways. Further development of this molecular tool may lead to a better understanding of the permissiveness of class II nuclear receptor dimers *in vivo*.

**Key words:** Nuclear receptor; RXR; PPAR; Retinoid; Rexinoid; Fibrate; Microarray; PPAR $\alpha$   $-/-$ ; Cutaneous T-cell lymphoma; Vitamin D

NUCLEAR receptors (NR) are potent transcriptional modulators involved in numerous physiological functions such as embryonic development, cell differentiation, and energy homeostasis. They act as hormones, nutrients, drugs, or pollutants sensors. Upon activation by selective chemicals, NR of the class II trans-

duce their molecular signal by heterodimerizing with the retinoid X receptor (RXR $\alpha$ ,  $\beta$ , and  $\gamma$  isoforms/NR2B). In this molecular network, RXR isoforms hold the remarkably central position of a common partner dimerizing with all class members [thyroid hormone receptor (TR/NR1A), retinoic acid receptor

<sup>1</sup>Present address: IGBMC, Illkirch, France.

Address correspondence to Thierry Pineau, Laboratoire de Pharmacologie et Toxicologie, Institut National de la Recherche Agronomique, 180 Chemin de Tournefeuille, BP3, F 31931 Toulouse, Cedex 9, France. Tel: (33) 561 28 53 95; Fax: (33) 561 28 53 10; E-mail: tpineau@toulouse.inra.fr

(RAR/NR1B), peroxisome proliferator-activated receptor (PPAR/NR1C), liver X receptor (LXR/NR1H2 and NR1H3), farnesoid X receptor (FXR/NR1H4 and NR1H5), vitamin D receptor (VDR/NR1I1), pregnane X receptor (PXR/NR1I2), constitutive androstane receptor (CAR/NR1I3 and NR1I4), etc.].

Several molecular signaling pathways elicited from these receptors are drug targets for therapeutic intervention in humans. Fibrates activate PPAR $\alpha$  and are potent hypolipidemic drugs. Retinoids that activate more or less specifically the RAR and RXR isoforms are beneficially prescribed for dermatological disorders (acne, psoriasis, photodamage, and cancer). Nevertheless, side effects elicited by retinoids (e.g., teratogenicity, hypertriglyceridemia, mucocutaneous toxicity) led to the development of more selective RXR activators termed rexinoids (6,24). Recently, the FDA approved bexarotene and Targretin (LGD1069; Ligand Pharmaceuticals, San Diego, CA) for the treatment of cutaneous T-cell lymphoma (CTCL). Rexinoids such as LGD1069 and the more potent RXR-selective agonist LG100268 (7) sustain beneficial actions for the treatment of cancers (12,13) and promise well for the treatment of type II diabetes (33) and atherosclerosis (10).

The high selectivity of these molecules towards RXR provides novel chemical probes for molecular investigations. Indeed, it could be envisaged that, following its selective activation, RXR could recruit one or several of its quiescent partners, even in the absence of their own xenobiotic activators, to trigger an original and potentially beneficial pharmacological response. Numerous RXR heterodimers have been initially classified as either permissive or nonpermissive based on their ability to respond to RXR ligands (28). RXR is thought to be ligand inducible in its association with PPAR (17,33), LXR (43), FXR (15), NGFI-B and Nurr1 (14,37). It has been considered a silent partner in its association with TR, VDR, and RAR. However, these assumptions are still a matter of debate and the necessary caution in the definition of dimer permissiveness has been highlighted in a recent study of the RXR-CAR $\beta$  heterodimer (41). Considering the multiplicity of the factors affecting the activity of each RXR heterodimer (promoter context, endogenous ligands, interactions with other RXR dimerization partners, cofactors expression, phosphorylations), we can expect that the responses would vary depending on the cellular context. We recently exemplified this concept by showing that rexinoid-induced upregulation of mouse pyruvate dehydrogenase kinase 4 is strictly PPAR $\alpha$  dependent in the liver and the kidney but PPAR $\alpha$  independent in the heart (35).

Considering the theoretical multiple impacts of re-

xinoids, assessing their net molecular effects would obviously require an integrative biology strategy aimed at recording their organ-specific transcriptional signatures. We thought to develop a customized low-density murine cDNA macroarray dedicated to class II NR pathways to gain information relevance and processing efficiency. This molecular tool allowed us to simultaneously record the level of expression of 120 transcripts, in three organs, under conditions of chemical exposure of the mice.

We thus screened wild-type and PPAR $\alpha$   $-/-$  mice treated with LGD1069 (RXR ligand) or fenofibrate (PPAR $\alpha$  ligand) for the expression level of these 120 transcripts in liver, kidney, and heart. Fenofibrate and PPAR $\alpha$   $-/-$  mice were used because rexinoids are known to induce the PPAR $\alpha$  signaling pathway in the liver (32,35). Our results demonstrate a clear activation by LGD1069 of the PPAR $\alpha$  signaling pathway in the liver and the kidney but not in the heart. We further identify new organ-specific targets of LGD1069 treatment. These noticeably include lipogenesis-related genes as well as Cyp2b10, Cyp2c29, and Cyp26 in the liver. In the kidney LGD1069 treatment results in induction of two genes implicated in vitamin D<sub>3</sub> homeostasis independently of PPAR $\alpha$  expression while the apolipoprotein E transcript is induced solely in the absence of PPAR $\alpha$ . These regulations are discussed in light of the actual knowledge about RXR heterodimer permissiveness and the use of rexinoids in cancer therapies.

## MATERIALS AND METHODS

### *Chemicals*

RXR selective agonists, LGD1069 and LG100268, were synthesized and controlled by the Chemistry Department at Galderma R&D. LGD1069 and LG100268 are highly selective RXR agonists with equilibrium dissociation constant ( $K_d$ ) of 14 and 3 nM for RXR $\alpha$ , respectively, and poor agonists for all three RAR isoforms (both molecules have  $K_d > 1 \mu\text{M}$ ) (6,7). Fenofibrate was purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). Suspensions of test compounds were prepared in a 0.5% (w/v) carboxymethyl-cellulose solution in purified water. Biochemical assays were purchased from Sigma-Aldrich.

### *Animals*

PPAR $\alpha$ -deficient male mice on a C57BL/6J genetic background (11,23) were bred at INRA transgenic rodent facility according to European Union guidelines for animal care. Age-matched male C57BL/

6J mice were obtained from Charles River France (Les Oncins, France). In vivo studies were conducted under the E.U. guidelines for the use and care of laboratory animals and all protocols were reviewed and approved by the Ethical Committee for Experimentation on Animals from the "Région Midi Pyrénées" (agreement #MP/03/26/07/03).

#### *Experimental Protocol*

Oral administrations were performed by gavage once daily (0900 h) for 8 consecutive days. Mice were fasted 2 h prior to euthanasia by cervical dislocation. The selected organs were dissected, weighed, snap-frozen in liquid nitrogen, and stored at -80°C until RNA extraction.

#### *Triglycerides and Apolipoproteins Measurements*

Serum triglycerides, apoA-I and apoB measurements were performed as previously detailed (35).

#### *RNA Isolation*

Total RNA was extracted with TRIzol® reagent (Invitrogen, Cergy Pontoise, France) according to the manufacturer's instructions.

#### *Northern Blot Analysis*

The procedures used for RNA electrophoretic separation, capillary transfer onto nylon membranes, and hybridization with <sup>32</sup>P-radiolabeled cDNA probes were previously detailed (35).

#### *Gene Expression Microarray Analysis Using Incyte® Mouse Unigene 1 Templates*

Frozen total RNA samples (200 µg) were provided to Incyte Genomics, Inc. (Palo Alto, CA, USA) for polyA messenger purification. Two samples (LG100268-treated mice and vehicle-treated mice) were analyzed. Taking into account the potential interindividual variability of the response to the drug, each sample consisted of a mix of equal amounts of RNAs extracted from three animals. The quality of the final samples was assessed by electrophoresis on an ethidium bromide-containing agarose gel, prior to shipment. Cy3 and Cy5 fluorescent labeling of the cDNAs was performed while the corresponding messengers were reverse transcribed. Both samples were competitively hybridized on a Mouse Unigene 1 glass template containing 9596 elements from the IMAGE Consortium library. Final data were delivered as a HTML file. Gene names were assigned based on Unigene (build

#136) clusters, which include the corresponding IMAGE clone.

#### *Class II NR-Dedicated Macroarray: INRAARRAY 01.2*

We selected a set of 120 relevant genes (rationale presented in the Results section) to design our original dedicated array. A fragment of each of these genes was amplified by PCR using murine cDNA and selective primer pairs. Each individual fragment was cloned into the pCR®2.1 TOPO® vector (Invitrogen) according to the manufacturer's protocols, thus allowing future cDNAs amplifications using "universal" vector-based primers located on both edges of the cloning site. All plasmids were produced using QIAfilter plasmid maxikit (Qiagen, Courtaboeuf, France). All cloned cDNA were subjected to control analyses (minimum of three restriction fragments analysis and full-length sequencing, especially in the case of failure to integrally produce all expected restriction patterns; 46 clones fully sequenced). A comprehensive list of 120 gene names, corresponding acronyms, GenBank accession numbers, PCR primer sequences, PCR conditions, and probe length can be obtained as supplemental material (<http://www.inra.fr/Internet/Centres/toulouse/pharmacologie/pharmacolemme/valorisations/materiel.html>). In addition to our own controls (housekeeping genes 36B4 and β-actin, sunflower rubisco-protein large subunit), negative, calibration, and ratio controls used in these experiments were provided by the Lucidea Universal Scorecard (23 artificial genes designed from yeast intergenic regions; Amersham Biosciences, Orsay, France).

Finally, individual cDNA solutions [200 ng µl<sup>-1</sup>, 50% (v/v) in DMSO] were prepared and spotted as duplicates onto 19 cm<sup>2</sup> positively charged nylon membranes (Immobilon™-Ny+, Millipore, Guyancourt, France) at the Genopole facility in Toulouse, France (<http://genopole.toulouse.inra.fr/>) with a customized Eurogridder microarrayer (Virtek/Biorad, Marnes-la-Coquette, France).

Quality and concentration of TRIzol®-extracted RNAs were assessed using a 2100 Bioanalyzer (Agilent Technologies, Massy, France). Labeling by random primed reverse transcription was performed on 2 µg of total RNA using the Hotscribe™ kit (Amersham Biosciences, Orsay, France) with 40 µCi of [ $\alpha$ -<sup>32</sup>P]dCTP (MP Biomedicals, Orsay, France). RNA matrix was degraded by alkaline lysis (SDS 0.2%, EDTA 50 mM, and NaOH 300 mM final, 68°C, 45 min). Following neutralization (Tris-HCl, pH 7.5, 500 mM final), unincorporated nucleotides were re-

moved on Sephadex G50 columns (ProbeQuant<sup>TM</sup> G-50 Micro columns, Amersham Biosciences). The activity of an aliquot was measured (Beta V  $\beta$ -counter, Kontron Instruments, Watford, UK) and the remaining probe was heated (95°C, 5 min) before being added to the hybridization solution ( $6 \times 10^6$  cpm ml<sup>-1</sup>). Membranes were prehybridized (65°C, 6 h) and hybridized (65°C, 24 h) in a rotating oven (QBioGene, Illkirch, France) in 60-ml individual polypropylene tubes (Nalge Nunc, Rochester, NY, USA) containing 5 ml of hybridization solution (5 × SSC, 5 × Denhardt's solution, 0.5% SDS, 100  $\mu$ g ml<sup>-1</sup> salmon sperm DNA). Following washing steps (65°C, 30 min twice and 65°C, 1 h once in 10 ml of 0.1 × SSC, 0.1% SDS) the membranes were exposed for 48–72 h to a phosphor screen before being scanned (50  $\mu$ m resolution) using a Storm phosphor-imager (Amersham Biosciences). Image analysis was performed with ImageMaster<sup>TM</sup> Array software v2.0 (Amersham Biosciences) with a background correction calculated on the median of spots edge.

#### *Statistical Analysis of INRARRAY 01.2 Data*

All data analyses were performed using the SPlus<sup>®</sup> 2000 software (Insightful, Toulouse, France) and Multidim library (<http://www.lsp.ups-tlse.fr/Carlier/Logiciel.html>) for multidimensional statistical analyses. The procedure for assessment of significant gene expression ratios described below was also used for Northern blot data analysis.

**Data Normalization and Transformation.** The signals for replicate spots on one microarray were averaged. Following log-transformation, data were normalized using the mean of eight spiked mRNAs (Lucidea Universal Scorecard) selected in the linear segment of the "log(signal)/mRNA amount" curve.

**Assessment of Significant Gene Expression Ratios.** To assess the significance of gene expression ratios, we performed an analysis of variance (ANOVA) for each gene using the following model:  $\log(\text{normalized signal}) = \text{genotype} + \text{treatment} + \text{genotype:treatment} + \epsilon$ . We then performed multiple means comparisons using a two-tailed Student's test with a Bonferroni family-wise error rate protection (comparison-wise error rate set to 5% or 1%). By specifying the linear combinations of the ANOVA model factor levels, the number of comparisons was limited to the five meaningful comparisons presented and discussed in this article. Due to the Bonferroni correction, each ratio is declared significant if its raw  $p$ -value is  $p < 0.002$  (for a 1% comparison-wise error rate) or  $p < 0.01$  (for 5%). Because the number of genes is reduced (120

genes analyzed), the number of falsely significant ratios should not exceed 1.2. Results were further filtered by retaining solely the genes exhibiting at least one modulation exceeding an absolute 1.5-fold expression ratio between two groups and being highly significant (raw  $p < 0.002$ ).

**Multidimensional Statistical Analyses.** For each organ, genes complying with the above criteria (at least one modulation with  $p < 0.002$  and absolute ratio  $\geq 1.5$ ) were selected. Principal Component Analysis and hierarchical clustering were performed using this selection of scaled gene expression levels. The script used in this study (S-language) is available as supplemental material (<http://www.inra.fr/Internet/Centres/toulouse/pharmacologie/pharmaco-moleculaire/valorisations/materiel.html>).

## RESULTS

#### *Identification of Rexinoid-Regulated (LG100268) Murine Hepatic Transcripts by High-Density Microarray Screening*

High-density screenings were performed to complement our literature mining efforts to document a relevant set of class II NR target genes. Here we present the results of the screening aimed at identifying RXR target genes in the liver. We selected the very specific and highly potent RXR activator, LG100268, and studied its effect on the liver transcriptome in wild-type mice. Pools of hepatic total RNA from three males treated with LG100268 (10 mg kg<sup>-1</sup> day<sup>-1</sup>, 8 days), or with the vehicle alone, were compared for transcript expression levels. Both cDNA samples were hybridized to a Mouse Unigene 1 glass slide (9596 IMAGE clones). As presented in Table 1, using the 1.7-fold cut-off value proposed by the manufacturer, we observed 44 upregulated transcripts (including six expressed sequences with no known function) and 27 downregulated transcripts (including one clone that does not belong to any Unigene cluster). Among the upregulated transcripts, 13 are unambiguously related to PPAR $\alpha$  activation (Cyp4a14, bifunctional enzyme, Cyp4a10, Pdk4, cyclin D1, Slc22a5, SCD1, Crat, Facl2, Akp2, Acox1, Hmgcs 1, and Fabp1; full names appear in Table 1) and one (Cyp26a1) is a RAR target in mouse liver (22). Five downregulated transcripts are known PPAR $\alpha$  targets (Mt 1 and 2, Gstp2, Mup1, Serpina1e) (9).

#### *Development of a Customized Class II NR cDNA Macroarray: INRARRAY 01.2*

All available class II NR were identified from extensive literature screening and their corresponding

TABLE 1  
LG100268-INDUCED HEPATIC GENE EXPRESSION MODULATIONS IN WILD-TYPE C57BL/6J MICE

Unigene Cluster Name	Clone Accession	Ratio
<b>Genes induced by LG100268 in C57BL/6J mice</b>		
Cyp4a14: cytochrome P450, family 4, subfamily a, polypeptide 14	AA106365	7.4
Ehhadh: enoyl-coenzyme A, hydratase/3-hydroxyacyl coenzyme A dehydrogenase	AA718155	6.5
Cyp4a10: cytochrome P450, family 4, subfamily a, polypeptide 10	AA755385	5.8
Cyp26a1: cytochrome P450, family 26, subfamily a, polypeptide 1	AA239785	5.4
G0s2: G0/G1 switch gene 2	AA458241	5.1
Pdk4: pyruvate dehydrogenase kinase, isoenzyme 4	AI322278	3.8
cDNA clone IMAGE: 5028619, partial cds	AI391013	3.1
Por: P450 (cytochrome) oxidoreductase	AI322279	2.7
Ccnd1: cyclin D1	AA117547	2.5
Alas1: aminolevulinic acid synthase 1	AI892379	2.3
Slc22a5: solute carrier family 22 (organic cation transporter), member 5	AI608375	2.2
Wbscr14: Williams-Beuren syndrome chromosome region 14 homolog (human)	AA543589	2.2
Sgk2: serum/glucocorticoid regulated kinase 2	AA138663	2.1
1700012B18Rik: RIKEN cDNA 1700012B18 gene	AI595047	2.0
Igfbp2: insulin-like growth factor binding protein 2	AA879643	2.0
Gadd45g: growth arrest and DNA-damage-inducible 45 gamma	W18822	2.0
Insig2: insulin-induced gene 2	AA197454	2.0
Scd1: stearoyl-coenzyme A desaturase 1	AA754922	1.9
Crat: carnitine acetyltransferase	AA473526	1.9
1500041J02Rik: RIKEN cDNA 1500041J02 gene	AI451818	1.8
Facl2: fatty acid coenzyme A ligase, long chain 2	AI789976	1.8
Akp2: alkaline phosphatase 2, liver	AA517588	1.8
Pla2g6: phospholipase A2, group VI	AA271866	1.8
Clu: clusterin	AA210481	1.8
5830435C13Rik: RIKEN cDNA 5830435C13 gene	AA200336	1.8
Cxadr: coxsackievirus and adenovirus receptor	AA560303	1.8
Adam11: A disintegrin and metalloprotease domain 11	AA733811	1.8
Aldh3a2: aldehyde dehydrogenase family 3, subfamily A2	AA821976	1.7
Acox1: acyl-coenzyme A oxidase 1, palmitoyl	AA612012	1.7
Scl27a2: solute carrier family 27 (fatty acid transporter), member 2	AA066694	1.7
Gadd45a: growth arrest and DNA-damage-inducible 45 alpha	AA553242	1.7
Crat: carnitine acetyltransferase	AI607260	1.7
AAss: amino adipate-semialdehyde synthase	AA271121	1.7
Adfp: adipose differentiation related protein	AA624422	1.7
Gstt2: glutathione S-transferase, theta 2	AA511089	1.7
Hdc: histidine decarboxylase	AI322263	1.7
Cct2: chaperonin subunit 2 (beta)	W17932	1.7
Kap: kidney androgen regulated protein	AA106167	1.7
Transcribed sequences	AA242024	1.7
Vnn1: vanin 1	AI597120	1.7
Hsd17b9: hydroxysteroid (17-beta) dehydrogenase 9	AI425345	1.7
AW539457: expressed sequence AW539457	AI509582	1.7
Hmges1: 3-hydroxy-3-methylglutaryl-coenzyme A synthase 1	AA466979	1.7
Fabp1: fatty acid binding protein 1, liver	AI466451	1.7
<b>Genes repressed by LG100268 in C57BL/6J mice</b>		
Ppp1r3c: protein phosphatase 1, regulatory (inhibitor) subunit 3C	AA462037	-2.9
Mt2: metallothionein 2	W36474	-2.8
Trp53inp1: transformation related protein 53 inducible nuclear protein 1	AA437755	-2.6
Mt1: metallothionein 1	AA638765	-2.4
Rgs16: regulator of G-protein signaling 16	AA390086	-2.3
Seph2: selenophosphate synthetase 2	AA414662	-2.1
Fos: FBJ osteosarcoma oncogene	AA002910	-2.1
Aqp1: aquaporin 1	AA241281	-2.0
Gstp2: glutathione S-transferase, pi 2	AA437941	-2.0
No Unigene cluster found	AI121690	-2.0
Got1: glutamate oxaloacetate transaminase 1, soluble	AA415254	-2.0
Similar to alpha-2u-globulin V precursor-mouse (LOC384025), mRNA	AA674270	-1.9
Cyp1a2: cytochrome P450, family 1, subfamily a, polypeptide 2	AA242360	-1.9
Rxra: retinoid X receptor alpha	AA462070	-1.9
Mup1: major urinary protein 1	AA822105	-1.8

TABLE 1  
CONTINUED

Unigene Cluster Name	Clone Accession	Ratio
Serpina3c: serine (or cysteine) proteinase inhibitor, clade A, member 3C	W14912	-1.8
Idb2: inhibitor of DNA binding 2	AI322367	-1.8
Serpina1e: serine (or cysteine) proteinase inhibitor, clade A, member 1e	W13979	-1.8
Ela1: elastase 1, pancreatic	AA717025	-1.8
Dio1: deiodinase, iodothyronine, type I	AA260525	-1.7
El2: elongation factor RNA polymerase II 2	AA545429	-1.7
Cyp2a4: cytochrome P450, family 2, subfamily a, polypeptide 4	AA674177	-1.7
Acta1: actin, alpha 1, skeletal muscle	AA770902	-1.7
Acta2: actin, alpha 2, smooth muscle, aorta	AA624460	-1.7
Tdo2: tryptophan 2,3-dioxygenase	AA572623	-1.7
Cyp7a1: cytochrome P450, family 7, subfamily a, polypeptide 1	AI464796	-1.7
Sema4g: (semaphorin) 4G	AA238294	-1.7

Wild-type C57BL/6J mice were orally administered for 8 consecutive days 10 mg kg<sup>-1</sup> day<sup>-1</sup> of LG100268 or the vehicle alone (*n* = 3 per group). Hepatic total RNA were extracted. Pools of RNA corresponding to the treated and control groups were competitively hybridized to a mouse Unigene 1 microarray (Incyte Genomics, Inc., Palo Alto, CA, USA). Data are expressed as fold change (i.e., ratio of treated samples/control samples). Genes exhibiting an absolute expression ratio between the two groups superior to 1.7 are presented.

cDNA sequences were obtained from GenBank. For NR target genes identification, we relied on three independent data sources: i) published results from high-density screenings, ii) articles reporting gene regulation(s) by any of the class II NR activators, iii) our unpublished results of high-density screenings. The latter includes an additional study performed on Mouse Unigene 1 microarrays, with hepatic RNA from fenofibrate- or vehicle-treated mice of either wild-type (C57BL/6J) or PPAR $\alpha$ -deficient genotype. Thus, we identified a set of PPAR $\alpha$ -regulated hepatic transcripts (data not shown), which is largely redundant with the data obtained by Cherkaoui-Malki et al. [Unigene array (9)] and Yamazaki et al. [Affymetrix U74Av2 array (46)] both using the Wy14,643 PPAR $\alpha$ -activator as a gene expression modulator.

Based on these data sources we developed a first version of a set of gene-specific probes (120 probes) relevant to the field of RXR and its known dimerization partners. We amplified these cDNAs from C57BL/6J mouse reverse transcribed RNAs (liver, kidney, enterocytes, adipose tissue, brain) using synthetic primers. All cDNAs were cloned into a common vector and all underwent a control procedure.

Due to the amount of available literature and to some of our sources (i.e., high-density screening with fibrate-treated livers) the PPAR $\alpha$  signaling pathway holds a prevailing space within the array. The most represented genes belong to the following groups of functions: xenobiotic-metabolizing enzymes (phase I and phase II), ABC transporters, lipid homeostasis (synthesis, transport, catabolism), carbohydrates homeostasis (synthesis, storage, utilization).

INRARRAY 01.2 was used to assess genes expres-

sion in liver, kidney, and heart of wild-type and PPAR $\alpha$   $-/-$  mice treated with fenofibrate, LGD1069, or vehicle alone. On the whole study (total of 72 macroarrays), the mean intragroup Pearson's correlation coefficient was  $0.98 \pm 0.01$  (mean  $\pm$  SD). Deviation from the ideal 1 value accounts for biological interindividual variability, technical variability, as well as presence of a few outlier macroarrays (3 out of 72 macroarrays); those values were subsequently removed from the datasets. Using artificial mRNAs spiked into the samples at known ratios between groups of mice, we showed that our macroarray data and subsequent analysis procedure led to reliable results as well as conservative estimations of expression ratios (Table 2). Theoretical 10-fold ratios appeared more affected by underestimation than three-fold ratios, especially in the low intensity range (Table 2).

#### *LGD1069 Induces a PPAR $\alpha$ -Independent Hepatomegaly in Mice*

Eight-month-old male C57BL/6J (wild-type) and PPAR $\alpha$   $-/-$  mice were orally treated for 8 consecutive days with fenofibrate (100 mg kg<sup>-1</sup> day<sup>-1</sup>), LGD1069 (100 mg kg<sup>-1</sup> day<sup>-1</sup>), or vehicle alone. Fenofibrate triggered a significant 47% increase in relative liver weight (liver weight/body weight, *p* < 0.05) in wild-type mice only. In contrast, the LGD1069-induced hepatomegaly observed in wild-type mice (30% increase, *p* < 0.05) was reproduced in PPAR $\alpha$   $-/-$  mice (22% increase, *p* < 0.05). Similarly, the treatment with LG100268 (10 mg kg<sup>-1</sup> day<sup>-1</sup>, 8 days), applied to wild-type mice, resulted in a significant hepatomegaly (27% increase, *p* < 0.05).

TABLE 2  
THEORETICAL AND OBSERVED EXPRESSION RATIOS  
FOR EIGHT ARTIFICIAL mRNAs SPIKED INTO THE HEPATIC  
TOTAL RNA SAMPLES

Name	Intensity Range	Expected Ratio	Observed Ratio
Ratio 1	low	-3	-2.5 ± 0.4
Ratio 2	low	3	3.0 ± 0.8
Ratio 3	high	-3	-3.0 ± 0.6
Ratio 4	high	3	2.5 ± 0.3
Ratio 5	low	-10	-3.9 ± 0.8
Ratio 6	low	10	3.6 ± 0.6
Ratio 7	high	-10	-5.0 ± 0.5
Ratio 8	high	10	9.3 ± 1.4

Eight artificial mRNAs (Lucidea Universal ScoreCard, Amersham Biosciences, Orsay, France) were spiked into the RNA samples prior to labeling. These spikes were incorporated at known ratios between treated and control groups. Corresponding spots were arbitrarily divided as exhibiting a low (from 15 to 150 pg of spiked mRNA per sample) or high (from 500 to 5000 pg of spiked mRNA per sample) intensity. This table summarizes the expected and observed ratios (mean fold change ± SD) for the hybridization of hepatic RNAs from 24 mice. Similar results were obtained from the hybridizations of cardiac and renal samples (data not shown).

#### *Effects of the Treatments on Serum Triglycerides, apoA-I, and apoB*

The 8-month old PPAR $\alpha$ -deficient mice displayed significantly higher constitutive serum triglycerides (112% increase,  $p < 0.05$ ,  $n = 5$  per genotype), apolipoprotein A-I (apoA-I, 20% increase,  $p < 0.05$ ,  $n = 14$  or 15 per genotype), and apolipoprotein B (apoB, 84% increase,  $p < 0.05$ ,  $n = 14$  or 15 per genotype) levels compared to wild-type littermates. No effects of either LGD1069 or fenofibrate were observed on serum triglycerides (TG) in PPAR $\alpha$   $-/-$  mice. In contrast, LGD1069 induced a significant 76% increase in serum TG of wild-type mice ( $p < 0.05$ ,  $n = 5$  per group). Although nonsignificant in this study, we noticed a 20% decrease of the mean serum TG in wild-type mice treated with fenofibrate. No significant effects of the treatments were observed on serum apoA-I and apoB.

#### *Hepatic, Renal, and Cardiac Effects of LGD1069 and Fenofibrate on the Abundance of a Selection of Class II NR-Related Transcripts in Wild-Type and PPAR $\alpha$ -Deficient Mice*

We used INRARRAY 01.2 to screen the effects of LGD1069 and fenofibrate on our selected genes panel in wild-type and PPAR $\alpha$ -deficient mice liver, kidney, and heart. Significantly modulated genes were identified as described in the Materials and Methods.

As shown in Table 3, using this procedure with a 1% comparison-wise error rate we identified in wild-

type mice liver 14 genes significantly modulated by LGD1069 treatment (11 upregulated and 3 downregulated) and 24 genes significantly modulated by fenofibrate treatment (21 upregulated and 3 downregulated). In PPAR $\alpha$ -deficient mice liver, LGD1069 treatment significantly increased the level of five transcripts and fenofibrate treatment of one transcript. Furthermore, three transcripts showed significantly higher levels and one transcript a lower level in PPAR $\alpha$   $-/-$  mice compared to wild-type mice under the control treatment.

In wild-type mice kidney (Table 4), LGD1069 significantly increased the levels of 15 transcripts while fenofibrate significantly increased the levels of 17 transcripts. In PPAR $\alpha$ -deficient mice kidney, LGD1069 significantly upregulated the cytochromes P450 24 and 27b1 as found in wild-type mice but also induced apoE mRNA expression. Fenofibrate did not trigger any significant modulations in PPAR $\alpha$   $-/-$  mice kidney. Fourteen transcripts were found differentially expressed between control wild-type and PPAR $\alpha$   $-/-$  mice kidney (eight transcripts overexpressed and six transcripts underexpressed in PPAR $\alpha$   $-/-$  mice kidney compared to wild-type littermates).

Screening of wild-type and PPAR $\alpha$   $-/-$  mice heart under vehicle, fenofibrate, or LGD1069 treatment led to a more limited number of significant gene expression modulations. Under the control treatment, the pyruvate dehydrogenase kinase 4 (PDK4), the peroxisomal/mitochondrial dienoyl-CoA isomerase (PMDCI), as well as both subunits of the mitochondrial hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase trifunctional protein (Tp $\alpha$  and Tp $\beta$ ), were downregulated in PPAR $\alpha$   $-/-$  mice compared to wild-type mice heart (-2.2-, -2.9-, -1.6-, and -1.7-fold, respectively, all  $p < 0.002$ ). The only significant regulation triggered by the treatments was the induction of PDK4 mRNA level under LGD1069 treatment. This upregulation was found significant in wild-type (2.4-fold,  $p < 0.002$ ) and in PPAR $\alpha$   $-/-$  mice (3.3-fold,  $p < 0.002$ ) as previously reported (35).

#### *Northern Blot Analysis of Some Gene Expression Modulations*

To confirm some of these gene expression regulations, Northern blots of individual mRNA samples were performed (Figs. 1, 2, and 3). In the liver (Fig. 1), except for Cyp8b1, all regulations observed by Northern blot are in agreement with the macroarray results. Macroarrays as well as Northern blot identified a modestly significant upregulation of Cyp8b1 by fenofibrate (1.8-fold and 2.3-fold, respectively, both  $p < 0.01$ ) in wild-type mice only. While Northern blot data indicated a significant downregulation

TABLE 3  
HEPATIC GENE EXPRESSION MODULATIONS INDUCED BY FENOFIBRATE OR LGD1069 IN WILD-TYPE AND PPAR $\alpha$   $-/-$  MICE

Genes	Wild-Type		PPAR $\alpha$ $-/-$		CT PPAR $\alpha$ $-/-$ /wt	Function
	Feno/CT	LGD1069/CT	Feno/CT	LGD1069/CT		
Cyp4a10	32.9 $\dagger$	7.8 $\dagger$	ns	ns	ns	microsomal FA $\omega$ -hydroxylation
Cyp4a14	44.8 $\dagger$	12.0 $\dagger$	ns	ns	2.4 $\dagger$	microsomal FA $\omega$ -hydroxylation
AOX	6.4 $\dagger$	2.5 $\dagger$	ns	ns	ns	peroxisomal FA oxidation and biogenesis
BIEN	40.0 $\dagger$	5.1 $\dagger$	ns	ns	ns	peroxisomal FA oxidation and biogenesis
THIOL	8.5 $\dagger$	3.8 $\dagger$	ns	ns	ns	peroxisomal FA oxidation and biogenesis
PECI	2.1 $\dagger$	ns	ns	ns	ns	peroxisomal FA oxidation and biogenesis
PMDCI	3.2 $\dagger$	ns	ns	ns	ns	peroxisomal FA oxidation and biogenesis
Pex11a	2.1 $\dagger$	ns	ns	ns	ns	peroxisomal FA oxidation and biogenesis
CACP	1.8 $\dagger$	ns	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
CPT2	1.7 $\dagger$	ns	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
MCAD	2.8 $\dagger$	ns	ns	ns	1.6*	mitochondrial FA oxidation and energy homeostasis
Tp $\alpha$	2.4 $\dagger$	ns	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
mHMGCoAS	3.1 $\dagger$	1.5*	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
PDK4	3.3 $\dagger$	ns	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
ACOTH	6.3 $\dagger$	2.0*	ns	ns	ns	FA metabolism
FAT/CD36	1.9 $\dagger$	ns	ns	ns	ns	FA transport
L-FABP	4.2 $\dagger$	ns	ns	ns	ns	FA trafficking
FAS	ns	3.5 $\dagger$	ns	2.4 $\dagger$	ns	lipogenesis
S14	ns	2.4 $\dagger$	ns	ns	ns	lipogenesis
Cyp26	ns	3.5 $\dagger$	ns	4.4 $\dagger$	ns	other cytochromes P450
Cyp2b10	ns	4.6 $\dagger$	1.7 $\dagger$	3.3 $\dagger$	ns	other cytochromes P450
Cyp2c29	-1.6 $\dagger$	2.2 $\dagger$	ns	1.7 $\dagger$	1.5 $\dagger$	other cytochromes P450
Cyp8b1	1.8*	-2.1 $\dagger$	ns	ns	ns	other cytochromes P450
ALDH3	4.1 $\dagger$	ns	ns	ns	ns	fatty aldehyde dehydrogenase
GST $\mu$	ns	2.2 $\dagger$	ns	ns	ns	glutathione S-transferases
GST $\pi$ 2	ns	-3.6 $\dagger$	ns	ns	-2.9 $\dagger$	glutathione S-transferases
LCE	-1.6 $\dagger$	ns	ns	ns	ns	TG hydrolysis
PLTP	2.0 $\dagger$	ns	ns	ns	ns	lipoprotein metabolism
apoC-III	ns	ns	ns	ns	1.6 $\dagger$	lipoprotein metabolism
PPAR $\alpha$	1.6 $\dagger$	ns	ns	ns	ns	nuclear receptor
SPI1	-2.7 $\dagger$	-2.1 $\dagger$	ns	-1.6*	-1.6*	acute phase response
ADISP/FSP27	2.0 $\dagger$	ns	ns	ns	ns	apoptosis
BSEP	ns	ns	ns	1.9 $\dagger$	ns	bile salt transport

Hepatic total RNA from wild-type and PPAR $\alpha$   $-/-$  mice treated with fenofibrate (Feno), LGD1069, or vehicle alone (CT) were hybridized to INRARRAY 01.2 ( $n = 3$  or 4 per group, total of 22 macroarrays). Raw data were normalized by the mean log(signal) of eight artificial mRNAs spiked into the samples. Normalized data were subjected to an ANOVA followed by a two-tailed Student's test using a Bonferroni comparison-wise error rate protection. The ratios (fold change) of transcript abundance correspond to the comparisons between the groups indicated in the headings of the columns. Genes presented exhibited at least one highly significant ( $p < 0.002$ ) gene expression modulation of 1.5-fold amplitude minimum out of the five comparisons studied. wt: wild-type.

Due to the Bonferroni  $p$ -value adjustments: \*raw  $p < 0.01$  and  $\dagger$ raw  $p < 0.002$ . ns: a nonsignificant expression ratio (raw  $p > 0.01$ ).

of Cyp8b1 mRNA by LGD1069 only in PPAR $\alpha$   $-/-$  mice ( $-2.5$ -fold,  $p < 0.002$ ), macroarrays did not find this regulation significant ( $-1.3$ -fold,  $p > 0.01$ ). Conversely, macroarrays identified a significant LGD1069-induced downregulation of Cyp8b1 only in wild-type mice ( $-2.1$ -fold,  $p < 0.002$ ). Although suggesting the same trend, this downregulation was not found sig-

nificant by Northern blot ( $-1.9$ -fold,  $p > 0.01$ ). We thus confirmed the PPAR $\alpha$ -independent regulations of Cyp2c29, Cyp26, Cyp2b10, FAS (all upregulated), and SPI1 (downregulated) mRNAs by LGD1069. In PPAR $\alpha$   $-/-$  mice liver compared to wild-type littermates, the constitutive underexpression of GST $\pi$ 2 and SPI1 mRNAs as well as a modest overexpression

**TABLE 4**  
RENAL GENE EXPRESSION MODULATIONS INDUCED BY FENOFIBRATE OR LGD1069 IN WILD-TYPE AND PPAR $\alpha$   $-/-$  MICE

Genes	Wild-Type		PPAR $\alpha$ $-/-$		CT PPAR $\alpha$ $-/-$ /wt	Function
	Feno/CT	LGD1069/CT	Feno/CT	LGD1069/CT		
Cyp4a10	5.5 $\dagger$	2.8 $\dagger$	ns	ns	-2.4 $\dagger$	microsomal FA $\omega$ -hydroxylation
Cyp4a14	22.1 $\dagger$	7.7 $\dagger$	ns	ns	ns	microsomal FA $\omega$ -hydroxylation
AOX	2.1 $\dagger$	1.7 $\dagger$	ns	ns	ns	peroxisomal FA oxidation and biogenesis
BIEN	5.0 $\dagger$	2.1 $\dagger$	ns	1.5*	-2.0 $\dagger$	peroxisomal FA oxidation and biogenesis
THIOL	3.0 $\dagger$	2.4 $\dagger$	ns	ns	ns	peroxisomal FA oxidation and biogenesis
PECI	1.6 $\dagger$	1.6 $\dagger$	ns	ns	ns	peroxisomal FA oxidation and biogenesis
PMDCI	1.9 $\dagger$	ns	ns	ns	-1.8 $\dagger$	peroxisomal FA oxidation and biogenesis
HPNCL	2.2 $\dagger$	1.8 $\dagger$	ns	ns	ns	peroxisomal FA oxidation and biogenesis
Pex11a	2.0 $\dagger$	ns	ns	ns	ns	peroxisomal FA oxidation and biogenesis
CACP	1.8 $\dagger$	1.3 $\dagger$	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
CPT2	2.1 $\dagger$	1.5 $\dagger$	ns	ns	-1.3*	mitochondrial FA oxidation and energy homeostasis
Tp $\alpha$	1.6 $\dagger$	1.2*	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
mHMGCoAS	10.4 $\dagger$	3.7 $\dagger$	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
PDK4	5.6 $\dagger$	3.1 $\dagger$	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
UCP2	3.8 $\dagger$	1.6 $\dagger$	ns	ns	ns	mitochondrial FA oxidation and energy homeostasis
ACOTH	4.6 $\dagger$	2.9 $\dagger$	ns	ns	ns	FA metabolism
L-FABP	2.1 $\dagger$	1.3*	ns	ns	1.4 $\dagger$	FA trafficking
Cyp24	ns	3.2 $\dagger$	ns	1.9 $\dagger$	2.2 $\dagger$	other cytochromes P450
Cyp27b1	ns	2.6 $\dagger$	ns	1.9 $\dagger$	2.1 $\dagger$	other cytochromes P450
Cyp2c29	ns	ns	ns	-1.7*	2.0 $\dagger$	other cytochromes P450
LPL	-1.6*	ns	ns	ns	-1.8 $\dagger$	TG hydrolysis
apoB	ns	ns	ns	ns	1.7 $\dagger$	lipoprotein metabolism
apoE	ns	ns	ns	2.9 $\dagger$	2.0 $\dagger$	lipoprotein metabolism
CAR1	ns	ns	ns	ns	1.6 $\dagger$	nuclear receptor
SPI1	ns	ns	ns	ns	-1.9 $\dagger$	acute phase response
i-NOS	1.4*	ns	ns	ns	1.6 $\dagger$	inflammation
cHMGCoAS	ns	ns	ns	ns	-2.1 $\dagger$	cholesterol biosynthesis

Renal total RNA from wild-type and PPAR $\alpha$   $-/-$  mice treated with fenofibrate (Feno), LGD1069, or vehicle alone (CT) were hybridized to INRARRAY 01.2 ( $n = 3$  or 4 per group, total of 23 macroarrays). Raw data were normalized by the mean log(signal) of eight artificial mRNAs spiked into the samples. Normalized data were subjected to an ANOVA followed by a two-tailed Student's test using a Bonferroni comparison-wise error rate protection. The ratios (fold change) of transcript abundance correspond to the comparisons between the groups indicated in the headings of the columns. Genes presented exhibited at least one highly significant ( $p < 0.002$ ) gene expression modulation of 1.5-fold amplitude minimum out of the five comparisons studied. wt: wild-type.

Due to the Bonferroni  $p$ -value adjustment: \*raw  $p < 0.01$  and  $\dagger$ raw  $p < 0.002$ . ns: a nonsignificant expression ratio (raw  $p > 0.01$ ).

of Cyp2c29 mRNA were also confirmed. PPAR $\alpha$ -dependent downregulation of Cyp2c29 and SPI1 by fenofibrate was also found by both techniques. We observed that fenofibrate raised the hepatic expression of Cyp2b10 transcript in PPAR $\alpha$   $-/-$  mice only, with a noticeable interindividual variability of the response to the drug.

In the kidney (Fig. 2), we confirmed the constitutive overexpression of Cyp24 and Cyp27b1 in PPAR $\alpha$   $-/-$  mice compared to wild-type mice. LGD1069 induced in both PPAR $\alpha$   $-/-$  and wild-type mice kidney an upregulation of these two genes. As shown in Figure 3, we further confirmed that PPAR $\alpha$   $-/-$  mice

slightly overexpressed apoE mRNA in the absence of LGD1069 or fenofibrate treatments. A very modest downregulation of renal apoE mRNA was observed in wild-type mice under fenofibrate treatment. We clearly identified that LGD1069 strongly induced the expression of renal apoE mRNA in PPAR $\alpha$   $-/-$  mice only (Fig. 3).

#### *Principal Component Analysis (PCA) of Hepatic and Renal Dedicated Transcriptional Signatures*

To globally represent the transcriptional signatures obtained in the liver and the kidney, we used PCA

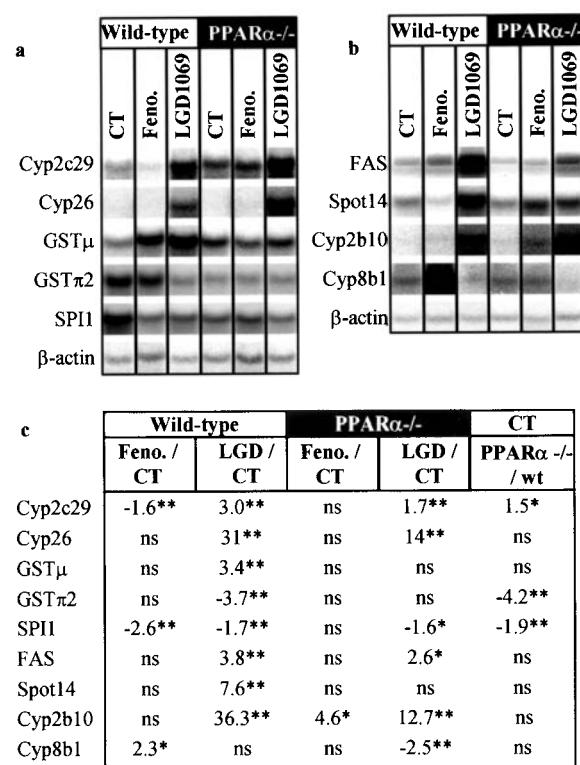


Figure 1. Regulation of transcript abundance in the liver of wild-type and PPAR $\alpha$  -/- mice by LGD1069 and fenofibrate. Northern blot analyses were carried out using 20  $\mu$ g of hepatic total RNA per lane. cDNA fragments spotted on the INRARRAY 01.2 were used as [ $\alpha$ - $^{32}$ P]dCTP-labeled probes. Samples from three animals were analyzed for each condition ( $n = 3 \times 6$  experimental conditions = 18 samples). (a, b) The results of 11 different hybridizations are presented (6 and 5 for a and b, respectively). Pictures obtained for 6 liver samples representative of the 6 groups of samples analyzed are presented. (c) Isotopic signals were quantified for the full set of 18 samples using a phosphorimager. Raw data generated for each gene were normalized according to the corresponding  $\beta$ -actin hybridization data. Normalized data were subjected to an ANOVA followed by a two-tailed Student's test using a Bonferroni comparison-wise error rate protection. The ratios of transcript abundance between the groups are indicated in the headings of the columns. Due to the Bonferroni  $p$ -value adjustments, \*raw  $p < 0.01$  and \*\*raw  $p < 0.002$ . ns: a nonsignificant expression ratio (raw  $p > 0.01$ ). CT: vehicle-treated control group, LGD: LGD1069, Feno.: fenofibrate.

on the subset of scaled variables presented in Tables 3 and 4 (33 and 27 selected genes for liver and kidney, respectively). Figure 4 presents the first three principal components obtained for the kidney. Very similar results were obtained for the liver and are thus not presented in this article. PCA allowed us to observe the hierarchy of variability sources affecting the transcriptional signatures. The biological meaning of these variability sources was extracted from the projections of the genes on the principal components (Fig. 4b and d). Figure 4a corresponds to the projection of the mice on the first two principal components (PC1 and PC2), which extracted 57% and 20% of

the total variability of the initial dataset, respectively. Wild-type mice treated with fenofibrate were markedly separated from wild-type controls along PC1 (Fig. 4a). All LGD1069-treated wild-type mice adopted an intermediate position along this axis. The three groups of PPAR $\alpha$  -/- mice were not separated along PC1. A majority of the 27 genes were highly correlated to PC1 (Fig. 4b). These genes are mostly well-described PPAR $\alpha$ -target genes implicated in various steps and pathways of fatty acid catabolism. PC2 discriminated all groups of PPAR $\alpha$  -/- mice from wild-type controls. This axis was most highly correlated with genes displaying a constitutive differential expression between wild-type and PPAR $\alpha$  -/- mice (Fig. 4b, Table 4, fifth column) such as inducible nitric oxide synthase (i-NOS), constitutive androstane receptor  $\beta$  (CAR1), or lipoprotein lipase (LPL). The third axis (PC3), which is orthogonal to PC1 and PC2, discriminated LGD1069-treated mice of both genotypes from

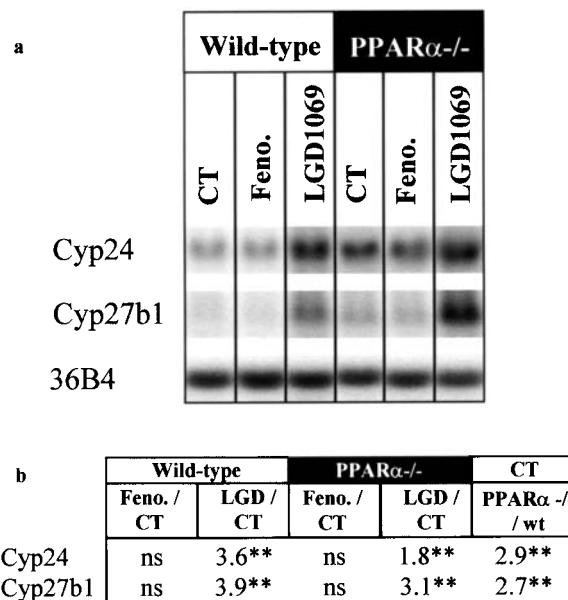


Figure 2. Regulation of Cyp24 and Cyp27b1 transcript abundance in the kidney of wild-type and PPAR $\alpha$  -/- mice by LGD1069 and fenofibrate. The experimental conditions for these Northern blot analyses are identical to those described in the legend of Figure 1, except for the number of animals per experimental group ( $n = 4 \times 6$  experimental conditions = 24 samples). (a) The results of three different hybridizations are presented. Pictures obtained for 6 liver samples representative of the 6 groups of samples analyzed are presented. (b) Isotopic signals were quantified for the full set of 24 samples using a phosphorimager. Raw data generated for each gene were normalized according to the corresponding 36B4-probe hybridization data. Normalized data were subjected to an ANOVA followed by a two-tailed Student's test using a Bonferroni comparison-wise error rate protection. (b) The ratios of transcript abundance between the groups are indicated in the headings of the columns. Due to the Bonferroni  $p$ -value adjustments: \*raw  $p < 0.01$  and \*\*raw  $p < 0.002$ . ns: a nonsignificant expression ratio (raw  $p > 0.01$ ). CT: vehicle-treated control group, LGD: LGD1069, Feno.: fenofibrate.

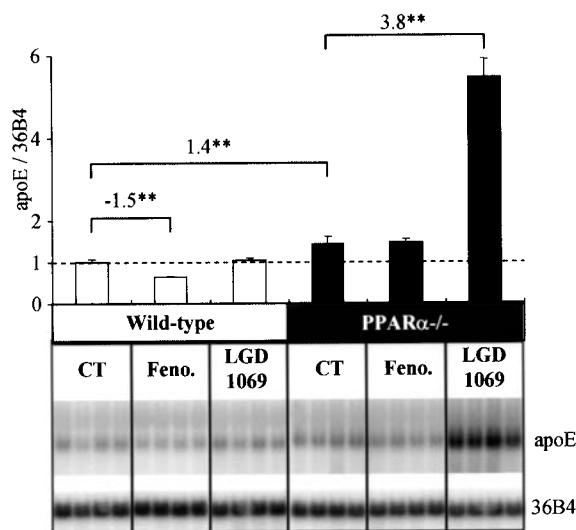


Figure 3. Regulation of apoE transcript abundance in the kidney of wild-type and PPAR $\alpha$  $^{-/-}$  mice by LGD1069 and fenofibrate. The experimental conditions for these Northern blot analyses are identical to those described in the legend of Figure 2. The pictures corresponding to all 24 kidney samples hybridized with apoE or 36B4 radiolabeled probes are presented in the lower part. Isotopic signals were quantified for the full set of 24 samples using a phosphorimager. Raw data generated for apoE mRNA were normalized according to the corresponding 36B4-probe hybridization data. Normalized data were subjected to an ANOVA followed by a two-tailed Student's test using a Bonferroni comparison-wise error rate protection. Histograms represent the mean of normalized data for each of the 6 experimental conditions, compared to the group of control (CT) wild-type mice. Error bars represent SD. Lines above the bars indicate which comparisons gave significant results. Expression ratios (or fold change) are given above these lines. Due to the Bonferroni *p*-value adjustments: \*\*raw *p* < 0.002. CT: vehicle-treated control group, Feno.: fenofibrate.

all other mice (Fig. 4c). This axis was most highly correlated with cytochromes P450 (Cyp) 24, 27b1, and 2c29 and to a lesser extent to apoE.

## DISCUSSION

Several class II NR display a valuable potential as they behave as drug sensors. Through their control on gene expression, they convey the beneficial effects of drugs that prove to be active on dyslipidemia, diabetes, or skin disorders. The lately developed drug family, rexinoids, is unique in that its members selectively activate RXR, the obligatory partner for all class II NRs. Activation of this central communication node could potentially result in modulations of various pathways through differential recruitment of dimeric partners. Thus, multiorgan screening of gene expression modulations is a relevant strategy to anticipate and delineate the tissue-specific pharmacological and/or toxicological impacts of rexinoids. Therefore, we designed and developed a dedicated cDNA

macroarray to implement this study. Our primary motivation was to record true organ-specific "fingerprints" of NR targeting drugs that contribute to accurately reporting their molecular impact on a mammalian organism. Remarkably, this allows considering global sets of data that include all statistically significant transcript modulations, even those displaying weak magnitude and, thus, doubtful biochemical significance when considered individually. Indeed, all these limited regulations retain, and even gain, significance when they are processed as multivariate signatures and when they are analyzed through multidimensional statistical methods such as PCA or hierarchical clustering. We believe that improving the accuracy of organ-specific transcriptional signatures is a relevant mean to focus, secondarily, the investigations on pathways where protein levels are eligible to fluctuate in response to NR activators, although this was not the aim of this initial investigation.

### Hepatic and Renal Transcriptional Signatures Analysis

PCA was used to observe the main effects of the treatments and of PPAR $\alpha$  deletion on the transcriptional signatures. Renal (Fig. 4) and hepatic (data not shown) transcriptional signatures displayed comparable patterns. The first PC was strongly correlated to numerous genes regulated by LGD1069 and fenofibrate in wild-type mice only and should thus be regarded as the illustration of PPAR $\alpha$ -dependent effects of the treatments on gene expression. The intermediate position of wild-type LGD1069-treated mice on this axis revealed the lower activation of the PPAR $\alpha$  signaling pathway by LGD1069 compared to fenofibrate. The second PC highlighted the importance of renal constitutive gene expression alterations following PPAR $\alpha$  abrogation. Finally, PC3 should be predominantly interpreted as PPAR $\alpha$ -independent effects of LGD1069 on Cyp24 and Cyp27b1 expressions. Hierarchical clustering of both genes and samples was also used to analyze these data and led to similar conclusions (data not shown).

### Rexinoid-Induced Hypertriglyceridemia

Although conflicting data have been collated in rodents (35), oral treatment with rexinoids can induce hypertriglyceridemia in humans (39). As previously presented, such hypertriglyceridemia can be observed in PPAR $\alpha$  $^{-/-}$  mice younger than those monitored in the present study (35). Using 8-month-old animals, a significant rise in serum TG was only observed in wild-type mice. At the molecular level, our data indicate that LGD1069 treatment raises the level of two transcripts related to fatty acid synthesis: Spot14

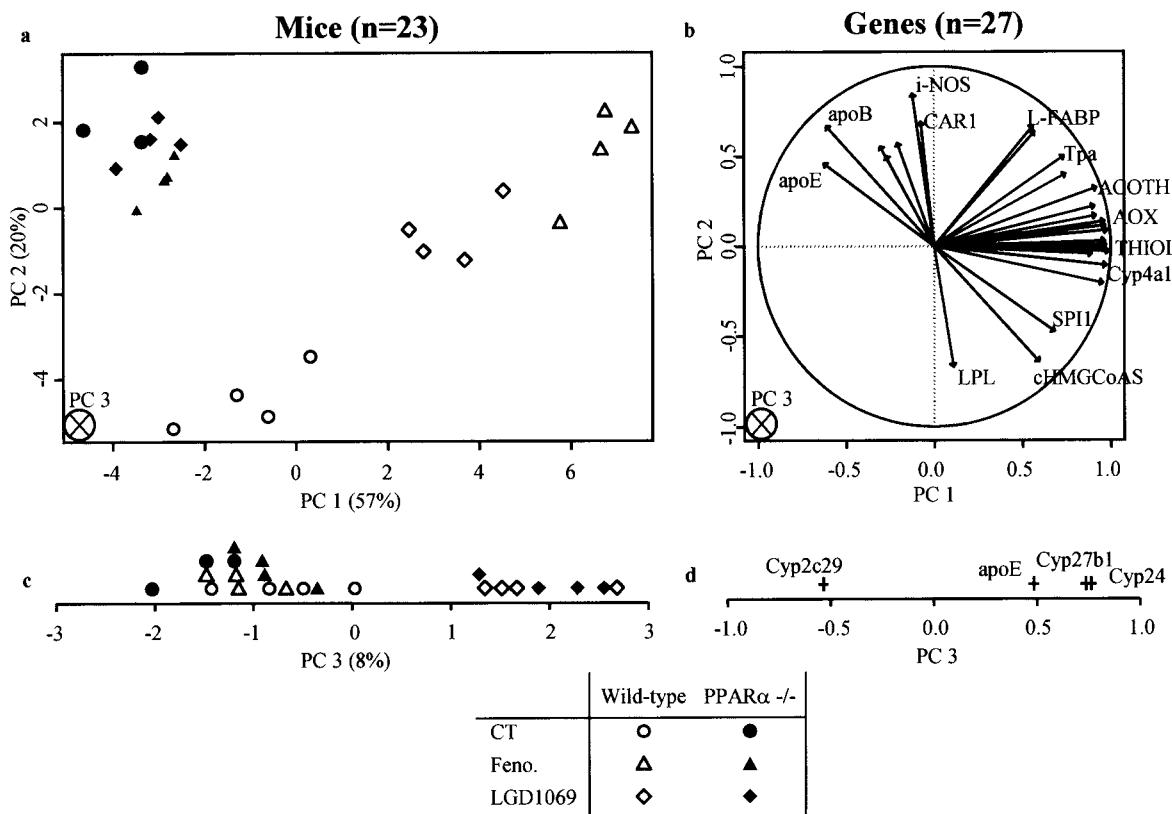


Figure 4. Principal component analysis of renal transcriptional signatures of wild-type and PPAR $\alpha$   $-/-$  treated with LGD1069, fenofibrate, or vehicle alone. Principal component analysis was performed on a data matrix composed of 23 RNA samples (mice) and 27 selected genes presented in Table 4. Thus, only one macroarray, corresponding to one PPAR $\alpha$   $-/-$  control mouse, was identified as outlier by quality controls and removed from the data set. (a) Plot of the 23 samples (mice) on the first two principal components (PC). Numbers in parentheses indicate the percentage of the total variance extracted by the corresponding PC. Open symbols correspond to wild-type mice and filled symbols to PPAR $\alpha$   $-/-$  mice. Circles are for controls (CT), diamonds for LGD1069-treated, and triangles for fenofibrate-treated (Feno.) mice. (b) Plot of genes (variables) on the first two PC. Each vector represents one gene. The names of some genes have been removed to facilitate the reading of the figure. (c) Plot of the 23 samples on the third PC. Points having close coordinates on this axis were piled up to facilitate the reading. (d) Plot of the genes on the third axis. Each gene's coordinate is represented by a cross. Only the four genes whose coordinates on the third PC were superior to 0.4 are represented. The names of the corresponding genes are indicated.

(S14) and fatty acid synthase (FAS) in wild-type mice. Although to a lesser extent, the rise of FAS mRNA is partially sustained in PPAR $\alpha$   $-/-$  mice. Furthermore, our data evidenced a 1.9-fold rise of Stearoyl-CoA desaturase 1 mRNA in wild-type mice following LG100268 treatment (Table 1). Altogether, these observations are consistent with a coordinated action of rexinoids on a set of genes known to promote hypertriglyceridemia. Based on our findings, it would be of great interest to assess the putative roles played by known lipogenesis regulators (e.g., SREBP, TR, or LXR) in this phenomenon, which could be related to a documented side effect of rexinoids.

#### Organ-Specific Rexinoid Targets

Besides renal and hepatic PPAR $\alpha$  activation illustrated by PCA, we observed a more restricted number

of PPAR $\alpha$ -independent gene expression modulations. In liver, LGD1069 induces Cyp2b10, a known target for PXR and CAR $\beta$ . The tandem DR-4 element from mouse Cyp2b10 does not respond to LGD1069 in RXR and/or CAR $\beta$ -transfected HepG2 cells (41). However, this cannot rule out the putative involvement of CAR $\beta$  nor of PXR in the context of our in vivo study (44). In addition, Cyp2c29 is also induced by LGD1069 in a PPAR $\alpha$ -independent manner and has recently been shown to be a target of CAR $\beta$  (20). Of interest is our original observation of the marked rise in hepatic Cyp26 transcript by both rexinoid molecules. Besides its role in the degradative 4-hydroxylation of *all-trans* RA during embryonic development (1,34), Cyp26 is thought to participate in limiting the access of RA to the transcriptional machinery (27). Regulation of Cyp26 by *all-trans* RA has been extensively studied (29,45). The central role RAR $\gamma$  and

RXR $\alpha$  has first been delineated using modified murine F9 cells lacking retinoid receptors (2). Cell line-specific interactions between a RA-response element (RARE-DR5) and other Cyp26 promoter regions have been shown (26). Here, we provide the first evidence that hepatic Cyp26 transcript can be upregulated *in vivo* by highly specific RXR activators. Even partially translated to the protein level, such a marked induction could have a profound repercussion on the cellular bioavailability of RA, thus possibly hampering the activity of the RXR-RAR dimer.

Very interestingly, in the kidney, we observed a significant induction by LGD1069 of Cyp24 and Cyp27b1, whose products are involved in vitamin D<sub>3</sub> homeostasis. Although constitutively overexpressed in PPAR $\alpha$   $-/-$  mice, these two genes are still responsive to LGD1069 treatment in this mouse model. Cyp24 and Cyp27b1 encode a 25-hydroxyvitamin D<sub>3</sub> 24-hydroxylase (19) and a 25-hydroxyvitamin D<sub>3</sub> 1 $\alpha$ -hydroxylase (40), respectively. These enzymes display opposite biosynthetic (Cyp27b1) and catabolic (Cyp24) functions in the homeostasis of the active form of vitamin D<sub>3</sub> (1 $\alpha$ ,25-dihydroxyvitamin D<sub>3</sub>). Induction of Cyp24 transcript in mouse kidney by LG100268 has been previously shown (3). In humans, this induction is mediated by RXR-RAR or RXR-VDR heterodimers binding on two vitamin D-responsive elements (VDREs) located in the promoter of CYP24 (47). Recently, Pascussi et al. identified PXR as a novel transcriptional regulator of both mouse and human Cyp24 (36), thus providing an additional mechanistic hypothesis for the regulation that we observed. Cyp27b1 was shown to be downregulated by liganded VDR (40). Here we show that both Cyp24 and Cyp27b1 are induced to a similar extent by LGD1069 in mouse kidney, independently of PPAR $\alpha$  expression. Consistent with a previously documented rexinoid-induced expression of Cyp24, we could hypothesized that a subsequent renal depletion in active vitamin D derivative would lower the activated proportion of the VDR, thus allowing the upregulation of Cyp27b1 reported here. However, this hypothesis remains to be fully substantiated. The coordinated modulation of two genes oppositely associated to the bioavailability of cellular active vitamin D suggests the existence of a putative feedback loop that could preserve a physiological balance between active and inactive vitamin D metabolites. Interestingly, our study revealed that Cyp27b1 is a novel target for the rexinoid family of xenobiotics, which likely deserves future attention in the skin.

The RXR-RAR, RXR-TR, and RXR-VDR dimers were initially reported to be nonpermissive (14). Consistent observations now support the concept that upon specific activation, RXR could become an ac-

tive contributor to gene expression modulation. Its action towards RAR, TR, or VDR known target genes appears subordinated to favorable cellular conditions that affect its partner (4,18,25,30). Further evidences, consistent with this concept, are provided by our *in vivo* studies. Indeed, the nature of the rexinoid-modulated targets indicate that various signaling pathways deserve further detailed investigations (e.g., RAR and CAR $\beta$  in the liver and VDR in the kidney). In general, this illustrates how cautiously the concept of dimer permissiveness should be viewed and how much it is likely to rely on specific experimental conditions.

This study further identified a puzzling regulation of the renal apolipoprotein E (apoE) transcript. ApoE is a surface constituent of plasma lipoproteins and a high-affinity ligand for the LDL receptor, which mediates the hepatic uptake of remnant lipoproteins. Although mostly produced in the liver, plasma apoE is also synthesized in various extrahepatic tissues including kidney (5,38). Strikingly, renal apoE mRNA is strongly induced by LGD1069 in PPAR $\alpha$   $-/-$  mice only. Response elements for LXR (21) and PPARs (16) have been found in apoE promoter. Our observations suggest that the PPAR $\alpha$  signaling pathway exerts an inhibitory control over renal apoE transcription through a direct or indirect mechanism that remains to be described. Furthermore, besides its clear atheroprotective effects through lipoprotein metabolism control, apoE displays antiproliferative functions in kidney (8) and most interestingly in mitogen-activated T lymphocytes (31), as in endothelial and tumor cells (42). Our work showed that LGD1069 has a capacity to regulate, *in vivo*, the apoE transcript as LG100268 does on the human THP-1 monocytic/macrophage line (21). Thus, we observed a degree of consistency between the pharmacological ability of rexinoids to limit cell proliferation in lymphoproliferative disorders and their potential to regulate the expression of apoE, an antiproliferative mediator. Exploring hypothetical causal links between these observations might represent a promising strategy to further elucidate the molecular events lying behind the beneficial effects of rexinoids in CTCL patients.

Together our results suggest that the PPAR $\alpha$  signaling pathway is the most sensitive pathway to be recruited through specific ligand activation of RXR in liver and kidney. However, the identification of several organ-specific, PPAR $\alpha$ -independent modulations triggered by LGD1069 suggests that different cellular contexts are capable of eliciting partial but active contributions of various NR signaling pathways.

The development of a dedicated set of target genes to study the organ-specific transcriptional signature

of rexinoids has documented the complex nature of their molecular impact *in vivo*. Although the development state of our dedicated macroarray does not yet allow us to fully explore all class II NR pathways, we show that this strategy already provides interesting evidences that rexinoids affect different molecular pathways in an organ-specific manner. These signatures strongly suggest that rexinoids trigger a sophisticated interplay between several NR signaling pathways. Further investigations are required to precisely delineate this network. Ongoing developments of our macroarray and its use in several organs and chemical exposure conditions will provide an additional valuable tool to improve our understanding of the pathways followed by drugs, pollutants, nutrients, or hormones to sustain their biological actions.

## ACKNOWLEDGMENTS

We are indebted to L. Ouamrane-Bernard (Galderma-Inra CIFRE-PhD program) for her expertise in triglycerides and lipoprotein analysis in mice and for monitoring the commercial microarray screening. We thank C. Bétoulières, G. Galy, and A.-M. Caussalter for excellent technical assistance. We are grateful to Dr. F. J. Gonzalez for the generous gift of the PPAR $\alpha$  $-/-$  mouse line. P. G. P. Martin was supported by the Ligue contre le cancer and A. Roulet by the Association pour la Recherche contre le Cancer. This work was supported by a grant from the ATC Nutrition founded by INSERM and INRA (AIP275) and by a grant from the Agence Française de Sécurité Sanitaire de l'Environnement (RD-2003-018/A2049).

## REFERENCES

- Abu-Abed, S.; Dolle, P.; Metzger, D.; Beckett, B.; Chambon, P.; Petkovich, M. The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev.* 15:226–240; 2001.
- Abu-Abed, S. S.; Beckett, B. R.; Chiba, H.; Chithalen, J. V.; Jones, G.; Metzger, D.; Chambon, P.; Petkovich, M. Mouse P450RAI (CYP26) expression and retinoic acid-inducible retinoic acid metabolism in F9 cells are regulated by retinoic acid receptor gamma and retinoid X receptor alpha. *J. Biol. Chem.* 273:2409–2415; 1998.
- Allegretto, E. A.; Shevde, N.; Zou, A.; Howell, S. R.; Boehm, M. F.; Hollis, B. W.; Pike, J. W. Retinoid X receptor acts as a hormone receptor *in vivo* to induce a key metabolic enzyme for 1,25-dihydroxyvitamin D<sub>3</sub>. *J. Biol. Chem.* 270:23906–23909; 1995.
- Bettoun, D. J.; Burris, T. P.; Houck, K. A.; Buck, D. W., 2nd; Stayrook, K. R.; Khalifa, B.; Lu, J.; Chin, W. W.; Nagpal, S. Retinoid X receptor is a nonsilent major contributor to vitamin D receptor-mediated transcriptional activation. *Mol. Endocrinol.* 17:2320–2328; 2003.
- Blue, M. L.; Williams, D. L.; Zucker, S.; Khan, S. A.; Blum, C. B. Apolipoprotein E synthesis in human kidney, adrenal gland, and liver. *Proc. Natl. Acad. Sci. USA* 80:283–287; 1983.
- Boehm, M. F.; Zhang, L.; Badea, B. A.; White, S. K.; Mais, D. E.; Berger, E.; Suto, C. M.; Goldman, M. E.; Heyman, R. A. Synthesis and structure–activity relationships of novel retinoid X receptor-selective retinoids. *J. Med. Chem.* 37:2930–2941; 1994.
- Boehm, M. F.; Zhang, L.; Zhi, L.; McClurg, M. R.; Berger, E.; Wagoner, M.; Mais, D. E.; Suto, C. M.; Davies, J. A.; Heyman, R. A.; et al. Design and synthesis of potent retinoid X receptor selective ligands that induce apoptosis in leukemia cells. *J. Med. Chem.* 38: 3146–3155; 1995.
- Chen, G.; Paka, L.; Kako, Y.; Singhal, P.; Duan, W.; Pillarisetti, S. A protective role for kidney apolipoprotein E. Regulation of mesangial cell proliferation and matrix expansion. *J. Biol. Chem.* 276:49142–49147; 2001.
- Cherkaoui-Malki, M.; Meyer, K.; Cao, W. Q.; Latruffe, N.; Yeldandi, A. V.; Rao, M. S.; Bradfield, C. A.; Reddy, J. K. Identification of novel peroxisome proliferator-activated receptor alpha (PPAR $\alpha$ ) target genes in mouse liver using cDNA microarray analysis. *Gene Expr.* 9:291–304; 2001.
- Claudel, T.; Leibowitz, M. D.; Fievet, C.; Tailleux, A.; Wagner, B.; Repa, J. J.; Torpier, G.; Lobaccaro, J. M.; Paterniti, J. R.; Mangelsdorf, D. J.; Heyman, R. A.; Auwerx, J. Reduction of atherosclerosis in apolipoprotein E knockout mice by activation of the retinoid X receptor. *Proc. Natl. Acad. Sci. USA* 98:2610–2615; 2001.
- Costet, P.; Legendre, C.; More, J.; Edgar, A.; Galtier, P.; Pineau, T. Peroxisome proliferator-activated receptor alpha-isoform deficiency leads to progressive dyslipidemia with sexually dimorphic obesity and steatosis. *J. Biol. Chem.* 273:29577–29585; 1998.
- Duvic, M.; Hymes, K.; Heald, P.; Breneman, D.; Martin, A. G.; Myskowski, P.; Crowley, C.; Yocom, R. C. Bexarotene is effective and safe for treatment of refractory advanced-stage cutaneous T-cell lymphoma: Multinational phase II–III trial results. *J. Clin. Oncol.* 19:2456–2471; 2001.
- Esteva, F. J.; Glaspy, J.; Baidas, S.; Laufman, L.; Hutchins, L.; Dickler, M.; Tripathy, D.; Cohen, R.; DeMichele, A.; Yocom, R. C.; Osborne, C. K.; Hayes, D. F.; Hortobagyi, G. N.; Winer, E.; Demetri, G. D. Multicenter phase II study of oral bexarotene for patients with metastatic breast cancer. *J. Clin. Oncol.* 21: 999–1006; 2003.
- Forman, B. M.; Umesono, K.; Chen, J.; Evans, R. M. Unique response pathways are established by allosteric interactions among nuclear hormone receptors. *Cell* 81:541–550; 1995.

15. Forman, B. M.; Goode, E.; Chen, J.; Oro, A. E.; Bradley, D. J.; Perlmann, T.; Noonan, D. J.; Burka, L. T.; McMorris, T.; Lamph, W. W.; et al. Identification of a nuclear receptor that is activated by farnesol metabolites. *Cell* 81:687–693; 1995.
16. Galetto, R.; Albajar, M.; Polanco, J. I.; Zakin, M. M.; Rodriguez-Rey, J. C. Identification of a peroxisome-proliferator-activated-receptor response element in the apolipoprotein E gene control region. *Biochem. J.* 357: 521–527; 2001.
17. Gampe, R. T., Jr.; Montana, V. G.; Lambert, M. H.; Miller, A. B.; Bledsoe, R. K.; Milburn, M. V.; Kliewer, S. A.; Willson, T. M.; Xu, H. E. Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol. Cell* 5:545–555; 2000.
18. Germain, P.; Iyer, J.; Zechel, C.; Gronemeyer, H. Co-regulator recruitment and the mechanism of retinoic acid receptor synergy. *Nature* 415:187–192; 2002.
19. Itoh, S.; Yoshimura, T.; Iemura, O.; Yamada, E.; Tsujikawa, K.; Kohama, Y.; Mimura, T. Molecular cloning of 25-hydroxyvitamin D-3 24-hydroxylase (Cyp-24) from mouse kidney: Its inducibility by vitamin D-3. *Biochim. Biophys. Acta* 1264:26–28; 1995.
20. Jackson, J. P.; Ferguson, S. S.; Moore, R.; Negishi, M.; Goldstein, J. A. The constitutive active/androstane receptor regulates phenytoin induction of Cyp2c29. *Mol. Pharmacol.* 65:1397–1404; 2004.
21. Laffitte, B. A.; Repa, J. J.; Joseph, S. B.; Wilpitz, D. C.; Kast, H. R.; Mangelsdorf, D. J.; Tontonoz, P. LXRs control lipid-inducible expression of the apolipoprotein E gene in macrophages and adipocytes. *Proc. Natl. Acad. Sci. USA* 98:507–512; 2001.
22. Lampen, A.; Meyer, S.; Nau, H. Effects of receptor-selective retinoids on CYP26 gene expression and metabolism of all-trans-retinoic acid in intestinal cells. *Drug Metab. Dispos.* 29:742–747; 2001.
23. Lee, S. S.; Pineau, T.; Drago, J.; Lee, E. J.; Owens, J. W.; Kroetz, D. L.; Fernandez-Salguero, P. M.; Westphal, H.; Gonzalez, F. J. Targeted disruption of the alpha isoform of the peroxisome proliferator-activated receptor gene in mice results in abolishment of the pleiotropic effects of peroxisome proliferators. *Mol. Cell. Biol.* 15:3012–3022; 1995.
24. Lehmann, J. M.; Jong, L.; Fanjul, A.; Cameron, J. F.; Lu, X. P.; Haefner, P.; Dawson, M. I.; Pfahl, M. Retinoids selective for retinoid X receptor response pathways. *Science* 258:1944–1946; 1992.
25. Li, D.; Li, T.; Wang, F.; Tian, H.; Samuels, H. H. Functional evidence for retinoid X receptor (RXR) as a nonsilent partner in the thyroid hormone receptor/RXR heterodimer. *Mol. Cell. Biol.* 22:5782–5792; 2002.
26. Loudig, O.; Babichuk, C.; White, J.; Abu-Abed, S.; Mueller, C.; Petkovich, M. Cytochrome P450RAI (CYP26) promoter: A distinct composite retinoic acid response element underlies the complex regulation of retinoic acid metabolism. *Mol. Endocrinol.* 14:1483–1497; 2000.
27. Luu, L.; Ramshaw, H.; Tahayato, A.; Stuart, A.; Jones, G.; White, J.; Petkovich, M. Regulation of retinoic acid metabolism. *Adv. Enzyme Regul.* 41:159–175; 2001.
28. Mangelsdorf, D. J.; Evans, R. M. The RXR heterodimers and orphan receptors. *Cell* 83:841–850; 1995.
29. Marikar, Y.; Wang, Z.; Duell, E. A.; Petkovich, M.; Voorhees, J. J.; Fisher, G. J. Retinoic acid receptors regulate expression of retinoic acid 4-hydroxylase that specifically inactivates all-trans retinoic acid in human keratinocyte HaCaT cells. *J. Invest. Dermatol.* 111: 434–439; 1998.
30. Minucci, S.; Leid, M.; Toyama, R.; Saint-Jeannet, J. P.; Peterson, V. J.; Horn, V.; Ishmael, J. E.; Bhattacharyya, N.; Dey, A.; Dawid, I. B.; Ozato, K. Retinoid X receptor (RXR) within the RXR-retinoic acid receptor heterodimer binds its ligand and enhances retinoid-dependent gene expression. *Mol. Cell. Biol.* 17:644–655; 1997.
31. Mistry, M. J.; Clay, M. A.; Kelly, M. E.; Steiner, M. A.; Harmony, J. A. Apolipoprotein E restricts interleukin-dependent T lymphocyte proliferation at the G1A/G1B boundary. *Cell. Immunol.* 160:14–23; 1995.
32. Mukherjee, R.; Strasser, J.; Jow, L.; Hoener, P.; Paterniti, J. R., Jr.; Heyman, R. A. RXR agonists activate PPARalpha-inducible genes, lower triglycerides, and raise HDL levels in vivo. *Arterioscler. Thromb. Vasc. Biol.* 18:272–276; 1998.
33. Mukherjee, R.; Davies, P. J.; Crombie, D. L.; Bischoff, E. D.; Cesario, R. M.; Jow, L.; Hamann, L. G.; Boehm, M. F.; Mondon, C. E.; Nadzan, A. M.; Paterniti, J. R., Jr.; Heyman, R. A. Sensitization of diabetic and obese mice to insulin by retinoid X receptor agonists. *Nature* 386:407–410; 1997.
34. Niederreither, K.; Vermot, J.; Fraulob, V.; Chambon, P.; Dolle, P. Retinaldehyde dehydrogenase 2 (RALDH2)-independent patterns of retinoic acid synthesis in the mouse embryo. *Proc. Natl. Acad. Sci. USA* 99:16111–16116; 2002.
35. Ouamrane, L.; Larrieu, G.; Gauthier, B.; Pineau, T. RXR activators molecular signalling: Involvement of a PPAR alpha-dependent pathway in the liver and kidney, evidence for an alternative pathway in the heart. *Br. J. Pharmacol.* 138:845–854; 2003.
36. Pascussi, J. M.; Robert, A.; Nguyen, M.; Walrant-Debray, O.; Garabedian, M.; Martin, P.; Pineau, T.; Saric, J.; Navarro, F.; Maurel, P.; Villarem, M. J. Drug-activated pregnane X receptor enhances CYP24 expression: Possible involvement in drug-induced osteomalacia. *J. Clin. Invest.* 115:177–186; 2005.
37. Perlmann, T.; Jansson, L. A novel pathway for vitamin A signaling mediated by RXR heterodimerization with NGFI-B and NURR1. *Genes Dev.* 9:769–782; 1995.
38. Reue, K. L.; Quon, D. H.; O'Donnell, K. A.; Dizikes, G. J.; Fareed, G. C.; Lusis, A. J. Cloning and regulation of messenger RNA for mouse apolipoprotein E. *J. Biol. Chem.* 259:2100–2107; 1984.
39. Rizvi, N. A.; Marshall, J. L.; Dahut, W.; Ness, E.; Truglia, J. A.; Loewen, G.; Gill, G. M.; Ulm, E. H.; Geiser, R.; Jaunakais, D.; Hawkins, M. J. A Phase I study of LGD1069 in adults with advanced cancer. *Clin. Cancer Res.* 5:1658–1664; 1999.

40. Takeyama, K.; Kitanaka, S.; Sato, T.; Kobori, M.; Yanagisawa, J.; Kato, S. 25-Hydroxyvitamin D<sub>3</sub> 1alpha-hydroxylase and vitamin D synthesis. *Science* 277: 1827–1830; 1997.
41. Tzameli, I.; Chua, S. S.; Cheskis, B.; Moore, D. D. Complex effects of retinoids on ligand dependent activation or inhibition of the xenobiotic receptor, CAR. *Nucl. Recept.* 1:2; 2003.
42. Vogel, T.; Guo, N. H.; Guy, R.; Drezlich, N.; Krutzsch, H. C.; Blake, D. A.; Panet, A.; Roberts, D. D. Apolipoprotein E: a potent inhibitor of endothelial and tumor cell proliferation. *J. Cell. Biochem.* 54:299–308; 1994.
43. Willy, P. J.; Umesono, K.; Ong, E. S.; Evans, R. M.; Heyman, R. A.; Mangelsdorf, D. J. LXR, a nuclear receptor that defines a distinct retinoid response pathway. *Genes Dev.* 9:1033–1045; 1995.
44. Xie, W.; Barwick, J. L.; Simon, C. M.; Pierce, A. M.; Safe, S.; Blumberg, B.; Guzelian, P. S.; Evans, R. M. Reciprocal activation of xenobiotic response genes by nuclear receptors SXR/PXR and CAR. *Genes Dev.* 14: 3014–3023; 2000.
45. Yamamoto, Y.; Zolfaghari, R.; Ross, A. C. Regulation of CYP26 (cytochrome P450RAI) mRNA expression and retinoic acid metabolism by retinoids and dietary vitamin A in liver of mice and rats. *FASEB J.* 14: 2119–2127; 2000.
46. Yamazaki, K.; Kuromitsu, J.; Tanaka, I. Microarray analysis of gene expression changes in mouse liver induced by peroxisome proliferator-activated receptor alpha agonists. *Biochem. Biophys. Res. Commun.* 290: 1114–1122; 2002.
47. Zou, A.; Elgort, M. G.; Allegretto, E. A. Retinoid X receptor (RXR) ligands activate the human 25-hydroxyvitamin D<sub>3</sub>-24-hydroxylase promoter via RXR heterodimer binding to two vitamin D-responsive elements and elicit additive effects with 1,25-dihydroxyvitamin D<sub>3</sub>. *J. Biol. Chem.* 272:19027–19034; 1997.

### **IV.3. CONCLUSIONS / PERSPECTIVES**

Alors que ces résultats confirment sans ambiguïté l'activation de la voie PPAR $\alpha$  par les agonistes spécifiques de RXR dans deux organes : le foie et le rein, où PPAR $\alpha$  est fortement activable par le fenofibrate également, cette étude met également en évidence des régulations tissu-spécifiques provoquées indépendamment de l'expression de PPAR $\alpha$  par les activateurs de RXR. Plusieurs de ces régulations permettent de suggérer des hypothèses d'activation d'autres RNs expliquant vraisemblablement des effets biochimiques et morphologiques observés dans cette étude : par exemple, une activation de LXR au niveau hépatique pourrait conduire à l'induction des gènes de la lipogénèse et participer à l'accroissement des taux circulant de triglycérides ou encore une activation de CAR, dont l'implication dans l'hépatomégalie provoquée par certains xénobiotiques a depuis été démontrée (Huang et al., 2005), pourrait participer à l'induction d'une hépatomégalie chez les souris des deux génotypes. Le caractère hautement tissu-spécifique de ces régulations semble de plus en accord assez étroit avec les niveaux d'expression des RNs suspectés (Bookout et al., 2006). Le cas du cœur semble néanmoins à part. Malgré un niveau d'expression élevé de PPAR $\alpha$  et RXR dans cet organe, aucun effet transcriptionnel du fenofibrate n'est détecté dans cette étude mais les souris PPAR $\alpha$ -/- présentent quelques altérations constitutives d'expression, suggérant que PPAR $\alpha$  pourrait déjà être hautement et constitutivement activé dans cet organe dont l'apport en énergie repose majoritairement chez l'adulte sur l'oxydation des acides gras.

Cette étude nous ayant permis d'acquérir une bonne confiance dans l'utilisation de notre puce à ADN via l'étude d'activations pharmacologiques, nous avons ensuite conduit une étude nutritionnelle de l'activation de ce RNs par différents acides gras pour nous rapprocher des conditions physiologiques d'activation du récepteur.



# **Chapitre V : Etude des effets de lipides alimentaires sur la voie PPAR $\alpha$**

## **V.1. INTRODUCTION**

Alors que de nombreux acides gras fixent et activent PPAR $\alpha$  (Chapitre II), peu d'études ont été menées *in vivo* sur modèle rongeur dans le but de démontrer le rôle de ce récepteur comme senseur des acides gras alimentaires. L'implication de PPAR $\alpha$  dans la médiation des effets de régimes riches en graisses sur l'expression de nombreux gènes au niveau hépatique avait déjà été démontrée (Patsouris et al., 2006). Nous souhaitions pour notre part évaluer son rôle dans un contexte d'apport en lipides couvrant les besoins de l'espèce sans être pour autant en excès. En nous basant sur les recommandations en termes d'alimentation des rongeurs (Council et al., 1995, p 82), nous avons donc conçu cinq régimes isocaloriques pour souris contenant environ 5% de lipides (poids/poids). Les huiles utilisées dans ces régimes diffèrent nettement par le type d'acides gras qu'elles apportent aux animaux, ce qui nous a offert la possibilité d'évaluer les « préférences » de PPAR $\alpha$  pour certains types d'acides gras. Nos principaux résultats, présentés dans les pages qui suivent sous leur forme publiée, mettent en évidence le rôle joué par PPAR $\alpha$  dans la médiation des effets des acides gras alimentaires et sa « préférence » pour les acides gras longs polyinsaturés de la famille  $\omega$ 3. Ils révèlent également l'impact des lipides alimentaires sur le degré de stockage des lipides au niveau hépatique avec des applications potentiellement intéressantes pour le traitement de stéatoses hépatiques, de plus en plus fréquemment rencontrées dans nos pays industrialisés notamment en raison de leur lien avec les problèmes d'obésité. Enfin, nos résultats nous ont permis de caractériser de manière relativement fine les effets de la délétion du récepteur PPAR $\alpha$  sur le métabolisme des lipides et l'expression de plusieurs gènes hépatiques. Ils nous ont permis de déceler de nouvelles cibles potentielles du récepteur (Lpin1 et les enoyl-CoA isomérasées) qui participent sans doute à façonner le phénotype des souris PPAR $\alpha$ -/. Ces travaux ont de plus révélé une légère surexpression hépatique du récepteur CAR chez les souris PPAR $\alpha$ -/ par rapport aux souris contrôles.

## **V.2. PRESENTATION DES TRAVAUX PUBLIES**

**Article n°3 :** Pascal G. P. Martin, Hervé Guillou, Frédéric Lasserre, Sébastien Déjean, Annaïg Lan, Jean-Marc Pascussi, Magali SanCristobal, Philippe Legrand, Philipe Besse, Thierry Pineau. Novel aspects of PPAR $\alpha$ -mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study. *Hepatology*. 2007 ; 45(3): 767-777

# Novel Aspects of PPAR $\alpha$ -Mediated Regulation of Lipid and Xenobiotic Metabolism Revealed Through a Nutrigenomic Study

Pascal G. P. Martin,<sup>1</sup> Hervé Guillou,<sup>2</sup> Frédéric Lasserre,<sup>1</sup> Sébastien Déjean,<sup>3</sup> Annaig Lan,<sup>2</sup> Jean-Marc Pascussi,<sup>4</sup> Magali SanCristobal,<sup>5</sup> Philippe Legrand,<sup>2</sup> Philippe Besse,<sup>3</sup> and Thierry Pineau<sup>1</sup>

**Peroxisome proliferator-activated receptor- $\alpha$  (PPAR $\alpha$ ) is a major transcriptional regulator of lipid metabolism. It is activated by diverse chemicals such as fatty acids (FAs) and regulates the expression of numerous genes in organs displaying high FA catabolic rates, including the liver. The role of this nuclear receptor as a sensor of whole dietary fat intake has been inferred, mostly from high-fat diet studies. To delineate its function under low fat intake conditions (4.8% w/w), we studied the effects of five regimens with contrasted FA compositions on liver lipids and hepatic gene expression in wild-type and PPAR $\alpha$ -deficient mice. Diets containing polyunsaturated FAs reduced hepatic fat stores in wild-type mice. Only sunflower, linseed, and fish oil diets lowered hepatic lipid stores in PPAR $\alpha^{-/-}$  mice, a model of progressive hepatic triglyceride accumulation. These beneficial effects were associated, in particular, with dietary regulation of  $\Delta 9$ -desaturase in both genotypes, and with a newly identified PPAR $\alpha$ -dependent regulation of lipin. Furthermore, hepatic levels of 18-carbon essential FAs (C18:2 $\omega$ 6 and C18:3 $\omega$ 3) were elevated in PPAR $\alpha^{-/-}$  mice, possibly due to the observed reduction in expression of the  $\Delta 6$ -desaturase and of enoyl-coenzyme A isomerase. Effects of diet and genotype were also observed on the xenobiotic metabolism-related genes Cyp3a11 and CAR. Conclusion:** Together, our results suggest that dietary FAs represent—even under low fat intake conditions—a beneficial strategy to reduce hepatic steatosis. Under such conditions, we established the role of PPAR $\alpha$  as a dietary FA sensor and highlighted its importance in regulating hepatic FA content and composition.

(HEPATOLOGY 2007;45:767-777.)

The nuclear receptor (NR) superfamily in the mouse is composed of 49 transcriptional regulators of gene expression that respond to a wide variety of small lipophilic molecules. The NRs are impli-

cated in numerous biological processes, including development, cell proliferation and differentiation, inflammation, energy, and xenobiotic metabolism. Among them, the peroxisome proliferator-activated receptors (PPAR $\alpha$ ,  $\beta/\delta$ , and  $\gamma$  or NR1C1, C2 and C3) are important modulators of lipid metabolism. In organs displaying high fatty acid (FA) catabolic rates, PPAR $\alpha$  regulates the expression of genes involved in most aspects of FA catabolism (FA uptake, activation, trafficking,  $\alpha$ - and  $\beta$ -oxidation,  $\omega$ -hydroxylation, ketogenesis), polyunsaturated FA (PUFA) biosynthesis (elongation and desaturation), and lipoprotein metabolism (apolipoprotein C-III, lipoprotein lipase). PPAR $\alpha$  also plays important roles in inflammation, glucose and bile acid metabolism, biotransformations, and hepatocarcinogenesis (reviewed by Mandard et al.<sup>1</sup>).

PPAR $\alpha$  is activated by structurally diverse molecules known as peroxisome proliferators, which induce hepatic peroxisome proliferation and modulate the expression of several target genes in rodents.<sup>2</sup> Among the peroxisome proliferators, fibrates are potent hypolipidemic drugs beneficially administered to humans. In vitro, several FAs bind to and activate PPAR $\alpha$  at physiologically relevant concentrations.<sup>3</sup> In vivo, high-fat (24% w/w)<sup>4</sup> and fish oil (10% w/w)<sup>5</sup> diets activate PPAR $\alpha$ . Thus, activation of

**Abbreviations:** CoA, coenzyme A; COC, saturated fatty acid diet; FA, fatty acid; FISH, fish oil diet; LIN,  $\omega$ 3-rich diet; NR, nuclear receptor; PPAR $\alpha$ , peroxisome proliferator-activated receptor- $\alpha$ ; PUFA, polyunsaturated fatty acid; Q-PCR, quantitative polymerase chain reaction; REF, reference diet; SCD1, stearoyl-CoA desaturase-1; SUN,  $\omega$ 6 fatty acid-rich diet; TG, triglyceride.

From the <sup>1</sup>Laboratoire de Pharmacologie et Toxicologie UR66, INRA, F-31931 Toulouse, France; <sup>2</sup>Laboratoire de Biochimie, ENSAR-INRA, Rennes, France; <sup>3</sup>Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France; <sup>4</sup>Laboratoire de Physiopathologie Hépatique, INSERM 632, Montpellier, France; and <sup>5</sup>Laboratoire de Génétique Cellulaire UR444, INRA, F-31326 Toulouse, France.

Received August 3, 2006; accepted November 1, 2006.

Supported by a grant from the ATC Nutrition founded by INSERM and INRA (AIP275) and by a grant from the Agence Française de Sécurité Sanitaire de l'Environnement (RD-2003-018/A2049).

Dr. Hervé Guillou is currently affiliated with Inositide Laboratory, The Babraham Institute, Cambridge, UK.

Address reprint requests to: Thierry Pineau, Laboratoire de Pharmacologie et Toxicologie, Institut National de la Recherche Agronomique (INRA), 180 Chemin de Tournefeuille, BP 3, F 31931 Toulouse, Cedex 9, France. E-mail: Thierry.Pineau@toulouse.inra.fr; fax: (33) 561-28-53-10.

Copyright © 2007 by the American Association for the Study of Liver Diseases. Published online in Wiley InterScience (www.interscience.wiley.com).

DOI 10.1002/hep.21510

Potential conflict of interest: Nothing to report.

Supplementary material for this article can be found on the HEPATOLOGY website (<http://interscience.wiley.com/jpages/0270-9139/suppmat/index.html>).

PPAR $\alpha$  by dietary FAs may represent a promising strategy to counteract obesity-related complications, with the caveat that the benefits of activating this pathway are demonstrated under conditions of moderate fat intake.

It is well known that dietary PUFAs have effects on diverse biological processes such as fuel partitioning, insulin action, cardiovascular function, neuronal development and immune function. The  $\omega 3$  and  $\omega 6$  FAs are the major classes of dietary PUFAs and display differential effects on pathologies such as cancer<sup>6</sup> and chronic inflammation.<sup>7</sup> Additionally, dietary PUFAs activate—both directly and indirectly—other transcription factors such as liver X receptor, hepatocyte nuclear factor-4, and sterol regulatory element binding protein (reviewed by Jump et al.<sup>8</sup>), which mediate, to some extent, their effects on gene expression.

The current study evaluates the role of PPAR $\alpha$  as a dietary FA hepatic sensor by exposing wild-type and PPAR $\alpha$ -deficient mice to five selected FA mixes while maintaining 4.8% total dietary fat (usual fat content for mouse maintenance diets: 4%–6% w/w). We used 5 contrasted diets: an essential FA-deficient diet, a reference diet, an  $\omega 6$  FA-rich diet, an  $\omega 3$  FA-rich diet, and a fish oil-enriched diet, and studied their effects on hepatic lipids and gene expression.

Even under low dietary fat intake, PUFA-containing diets reduced hepatic FA stores in a manner partially independent of PPAR $\alpha$ . Dietary regulations of Lpin1 in wild-type and stearoyl-coenzyme A (CoA) desaturase-1 (SCD1) in both genotypes paralleled these effects on hepatic FA content. Moreover, hepatic FA profiles revealed that linolenic and linoleic acids accumulate in PPAR $\alpha^{-/-}$  livers, suggesting a role for this NR in the degradation and/or use of the essential precursors of  $\omega 3$  and  $\omega 6$  PUFAs. Consistent with this finding, we showed that in PPAR $\alpha^{-/-}$  mice, the hepatic expression of 3 enoyl-CoA isomerasases as well as  $\Delta 6$ -desaturase were markedly reduced. In addition, we observed regulations of the xenobiotic metabolism-related genes Cyp3a11 and CAR. Overall, our study demonstrates that low-fat, PUFA-containing diets exert beneficial effects on hepatic fat stores. Moreover, it clarifies the role of PPAR $\alpha$  in mediating dietary effects on hepatic gene expression, FA content, and composition.

## Material and Methods

**Chemicals.** Chemicals for FA analyses were obtained from Prolabo (Paris, France) or Merck (Darmstadt, Germany). Standards for FA analyses, 1,4-bis [2-(3,5-dichloropyridyloxy)] benzene (TCPOBOP), and fenofibrate were obtained from Sigma (Lyon, France).

**Animal Housing and Treatment.** PPAR $\alpha$ -deficient mice with a C57BL/6J genetic background<sup>2,9</sup> were bred at INRA's transgenic rodent facility at 22 ± 2°C under 12-hour light/dark cycles. Age-matched male C57BL/6J mice were obtained from Charles River (Les Oncins, France) and were acclimated to local animal facility conditions for 2 weeks. In vivo studies were conducted under European Union guidelines for the use and care of laboratory animals and were approved by the institutional ethics committee.

**Diet Study.** Ten to 11-week-old male mice were given experimental diets ad libitum for 8 weeks (pellets prepared by UPAE-INRA, Jouy-en-Josas, France, replaced weekly), with free access to water. Food consumption (groups of 6) and individual body weight were recorded weekly. All diets were isocaloric and contained 4.8% fat (w/w). Oils used for experimental diet preparation were corn and colza oils (50/50) for a reference diet (REF), hydrogenated coconut oil for a saturated FA diet (COC), sunflower oil for an  $\omega 6$  FA-rich diet (SUN), linseed oil for an  $\omega 3$ -rich diet (LIN), and corn/colza/docosahexaenoic acid-enriched fish oils for the fish oil diet (FISH) (42.5/42.5/15). The fish oil (72% of C22:6n-3) was obtained from Polaris (Quimper, France). Other FAs were C22:5n-3 (17%), C22:5n-6 (2%), C22:4n-3 (2.4%), and C20:5n-3 (2%). The detailed composition of the diets is given in Table 1. FA composition was controlled via gas chromatography analysis of organic extracts from manufactured food pellets.

**Compound Administration.** Fenofibrate (100 mg/kg/day for 14 days) was administered via gavage and TCPOBOP (3 mg/kg/day for 3 days) was administered via intraperitoneal injection.

**Tissue Sample Collection.** Selected organs were dissected, rinsed, weighed, snap-frozen in liquid nitrogen, and stored at –80°C until analysis.

**FA Composition.** Hepatic FA composition was assayed as described previously.<sup>10</sup> Triglycerides (TGs) and phospholipids were separated via thin-layer chromatography before FA methyl esters analysis via gas chromatography (Agilent Technologies 6890N, Bios Analytique, Toulouse, France).

**Enzyme Activities.** FA-desaturase activities were assayed on postmitochondrial fractions using <sup>14</sup>C-labeled FAs.<sup>10</sup>

**Northern Blot, Macroarray, and Quantitative PCR Analysis.** Total RNA was extracted with TRIzol reagent (Invitrogen, France). Northern blot analysis was performed as described.<sup>10,11</sup> Production of INRARRAY, which comprises the selection, cloning, amplification and spotting of complementary DNAs onto nylon membranes, has been described previously.<sup>11</sup> Briefly, INRARRAY 01.2

**Table 1. Fatty Acid Composition of the Five Experimental Diets**

	<b>COC</b>	<b>REF</b>	<b>SUN</b>	<b>LIN</b>	<b>FISH</b>
SFA	99.6	10.8	11.8	9.0	9.3
MUFA	0.3	46.0	22.6	18.0	39.0
PUFA	0.1	43.2	65.6	72.9	51.7
$\omega_6$	0.1	37.9	65.5	14.9	33.7
$\omega_3$ (DHA)	0.0 (0.0)	5.3 (0.0)	0.1 (0.0)	58.0 (0.0)	18.0 (10.6)
UFA/SFA	0.0	8.3	7.5	10.1	9.7
$\omega_6/\omega_3$	—	7.1	1091.7	0.3	1.9

NOTE. Results are presented as a percentage of the total FAs present in the oil (except for the calculated ratios). All diets contained 4.8% fat, 45% starch, 22.2% sucrose, 1% cellulose, 21% casein, 4% salt mixture, 1% vitamin mixture (0.02% for vitamin E), and 1% agar.

Abbreviations:  $\omega_3$  or  $\omega_6$ , FAs of the  $\omega_3$  or  $\omega_6$  family, respectively; DHA, docosahexaenoic acid; MUFA, mono-unsaturated fatty acid; SFA, saturated fatty acid; UFA, unsaturated fatty acid.

consists of 120 complementary DNA fragments comprising class II NRs, target genes for each NR, housekeeping genes, and external controls (Lucidea Universal Scorecard, Amersham Biosciences, Orsay, France). Protocols for RNA radiolabeling, INRARRAY hybridization, and image analysis have been described.<sup>11</sup> Forty hepatic RNA samples (4 mice/group) were screened with INRARRAY. For real-time quantitative PCR (Q-PCR), total RNA samples (2  $\mu$ g) were reverse-transcribed with SuperScript II (Invitrogen). All assays were performed on an ABI Prism 7000 (Applied Biosystems, Courtaboeuf, France) using standard PCR conditions. Primers and Taqman probes for TATA-box binding protein (Mm00446973\_m1), Lpin1 (Mm00550511\_m1), CAR $\beta$  (Mm00437986\_m1), PECL (Mm00478725\_m1), ECI (Mm00494452\_m1), FADS1 (Mm00507605\_m1), and FADS2 (Mm00517221\_m1) were purchased from Applied Biosystems Assays-on-Demand. Primers for SYBR Green assays were as follows: SCD1-F, 5'-CCGGAGACCCCTTAGATCGA-3'; SCD1-R, 5'-TAGCCTGTAAAGATTCTGCAAA-CC-3'; PMDCI-F, 5'-GGAAAGATGTTCACTTCA-GGTATTGAC-3'; PMDCI-R, 5'-CGGGCCGCAT-CATCTC-3'. A pool of all complementary DNA samples was used to generate calibration curves. All Q-PCR data were normalized by TATA-box binding protein levels.

**Statistical Analysis.** Data analyses were performed using SPlus 2000 (Insightful, Toulouse, France) and R (<http://www.r-project.org/>) software. Data are expressed as the mean  $\pm$  SD ( $n = 4$ –6 per group). Differential effects were analyzed via ANOVA with the appropriate factors and interactions. When an effect was significant, a post hoc 2-tailed Student *t* test with a pooled variance estimate was used to compare the groups. For the macroarray data, the Benjamini-Hochberg procedure<sup>12</sup> was used to control the false discovery rate at 5% (multitest package from [www.bioconductor.org](http://www.bioconductor.org)). Only the genes displaying a minimum of 1 significant modulation of at least 1.5-fold amplitude are presented.

## Results

**Food Consumption, Body, and Organ Weights.** No significant effect of genotype or diet was observed on body weight gain or food consumption. PPAR $\alpha^{-/-}$  mice displayed significantly higher epididymal WAT (white adipose tissue) weight compared with controls ( $3.1 \pm 1.0\%$  and  $2.3 \pm 0.6\%$  of body weight, respectively;  $n = 30$  per genotype), without significant diet effect. Only the LIN diet significantly reduced liver weight in PPAR $\alpha^{-/-}$  mice (Table 2), which consistently displayed heavier livers than their wild-type counterparts.

**Effect of Diet and Genotype on Hepatic FA Content and Composition.** Hepatic total FA content was significantly higher in PPAR $\alpha^{-/-}$  versus wild-type mice for all diets (Table 2). In wild-type mice fed PUFA-containing diets (REF, SUN, LIN, and FISH), hepatic total FA content was significantly reduced compared with the COC group. Conversely, in PPAR $\alpha^{-/-}$  mice, the REF diet no longer reduced hepatic FA, whereas the SUN, LIN, and FISH diets did efficiently reduce hepatic FA stores. Similar results were obtained for the hepatic TG-FA content, whereas no significant effects were observed on hepatic phospholipid-FA content (Table 2). Histological examination of neutral lipid-staining in frozen liver sections (Fig. 1) confirmed these quantitative data (Table 2). In wild-type mice (Fig. 1A–C), increased TG accumulation was also observed in the COC-fed group. As previously reported,<sup>9</sup> PPAR $\alpha$ -deficiency favored macrovacuolar, centrilobular steatosis. However, the level of hepatic neutral lipid accumulation was reduced in FISH-fed PPAR $\alpha^{-/-}$  mice (Fig. 1F) compared with mice fed the COC (Fig. 1D) and REF (Fig. 1E) diets.

Twenty-one hepatic FAs were assayed via gas chromatography in the total lipid, TG, and phospholipid fractions (Table 3 and supplementary material). Livers from both genotypes displayed specific accumulations of FA families that were present in the diets. Mice fed the COC

**Table 2. Genotype and Diet Effects on Liver Weight and Hepatic FA Content in the Total Lipid, Triglyceride, and Phospholipid Fractions**

PPAR $\alpha$	Diet	Liver Weight (% Body Weight)	Total FA Content ( $\mu\text{g}/\text{mg}$ of Liver)	TG-FA Content ( $\mu\text{g}/\text{mg}$ of Liver)	PL-FA Content ( $\mu\text{g}/\text{mg}$ of Liver)
+/+	COC	4.8 $\pm$ 0.2	60.5 $\pm$ 21.4	16.4 $\pm$ 3.9	13.3 $\pm$ 2.3
	REF	4.1 $\pm$ 0.4 <sup>b</sup>	38.3 $\pm$ 4.0 <sup>b</sup>	9.9 $\pm$ 2.2 <sup>b</sup>	13.9 $\pm$ 1.0
	SUN	4.3 $\pm$ 0.2 <sup>b</sup>	37.6 $\pm$ 5.2 <sup>b</sup>	8.7 $\pm$ 2.5 <sup>b</sup>	14.2 $\pm$ 1.9
	LIN	4.5 $\pm$ 0.3 <sup>c</sup>	31.5 $\pm$ 6.0 <sup>b</sup>	6.2 $\pm$ 2.2 <sup>b,c</sup>	13.0 $\pm$ 0.7
	FISH	4.6 $\pm$ 0.3 <sup>c</sup>	42.3 $\pm$ 6.7 <sup>b</sup>	7.8 $\pm$ 2.0 <sup>b</sup>	15.0 $\pm$ 5.8
-/-	COC	5.3 $\pm$ 0.4 <sup>a</sup>	219.5 $\pm$ 82.3 <sup>a</sup>	96.1 $\pm$ 65.4 <sup>a</sup>	12.6 $\pm$ 2.4
	REF	5.3 $\pm$ 0.7 <sup>a</sup>	212.4 $\pm$ 44.6 <sup>a</sup>	73.1 $\pm$ 26.5 <sup>a</sup>	12.8 $\pm$ 2.9
	SUN	5.0 $\pm$ 0.4 <sup>a</sup>	128.5 $\pm$ 54.0 <sup>a,b,c</sup>	46.4 $\pm$ 32.5 <sup>a,b</sup>	13.8 $\pm$ 2.3
	LIN	4.5 $\pm$ 0.4 <sup>b,c</sup>	96.9 $\pm$ 42.0 <sup>a,b,c</sup>	42.9 $\pm$ 24.2 <sup>a,b</sup>	15.8 $\pm$ 1.3
	FISH	5.6 $\pm$ 0.6 <sup>a</sup>	76.3 $\pm$ 29.2 <sup>a,b,c</sup>	26.8 $\pm$ 12.8 <sup>a,b</sup>	16.1 $\pm$ 2.8

<sup>a</sup>Significant genotype effect (same diet; n = 6 per group).<sup>b,c</sup>Significant difference versus the COC or REF diet, respectively (same genotype; n = 6 per group).

diet preferentially accumulated saturated FA (except C18:0) and especially mono-unsaturated FAs in their livers. Mice fed the SUN diet had livers enriched in  $\omega$ 6 FAs, whereas those fed the LIN and FISH diets preferentially accumulated  $\omega$ 3 FAs. Regarding FA composition, the REF diet corresponded to a balanced profile compared with the other diets (Table 3 and supplementary material). Generally, PPAR $\alpha^{-/-}$  livers when compared with wild-type livers accumulated 16-carbon and 18-carbon monounsaturated FAs and, notably, displayed a robust increase in linoleic (C18:2 $\omega$ 6) and  $\alpha$ -linolenic (C18:3 $\omega$ 3) acid proportions. Conversely, most saturated FAs and 20- to 22-carbon PUFA were reduced in PPAR $\alpha^{-/-}$  livers. Hepatic TG and phospholipid analyses yielded comparable conclusions (see Supplementary material).

**Effect of Diet and Genotype on Hepatic Gene Expression.** Under such low fat intake conditions, we analyzed the expression of genes related to class II NR signaling using INRARRAY01.2.<sup>11</sup> Mean Pearson correlation coefficient was 0.97 between macroarrays (n = 40), indicating good reproducibility and relatively discrete gene expression changes. Table 4 presents the results of the macroarray screening. Thirty-one genes displayed differential expressions between PPAR $\alpha^{-/-}$  and wild-type mice. Minimal effects of genotype were observed for the REF diet. For dietary effects, comparisons to the COC and REF diets provided control situations of PUFA deficiency and intermediate FA profile, respectively (Table 4). For wild-type mice, 26 genes displayed at least one significant modulation out of the 7 comparisons performed. In

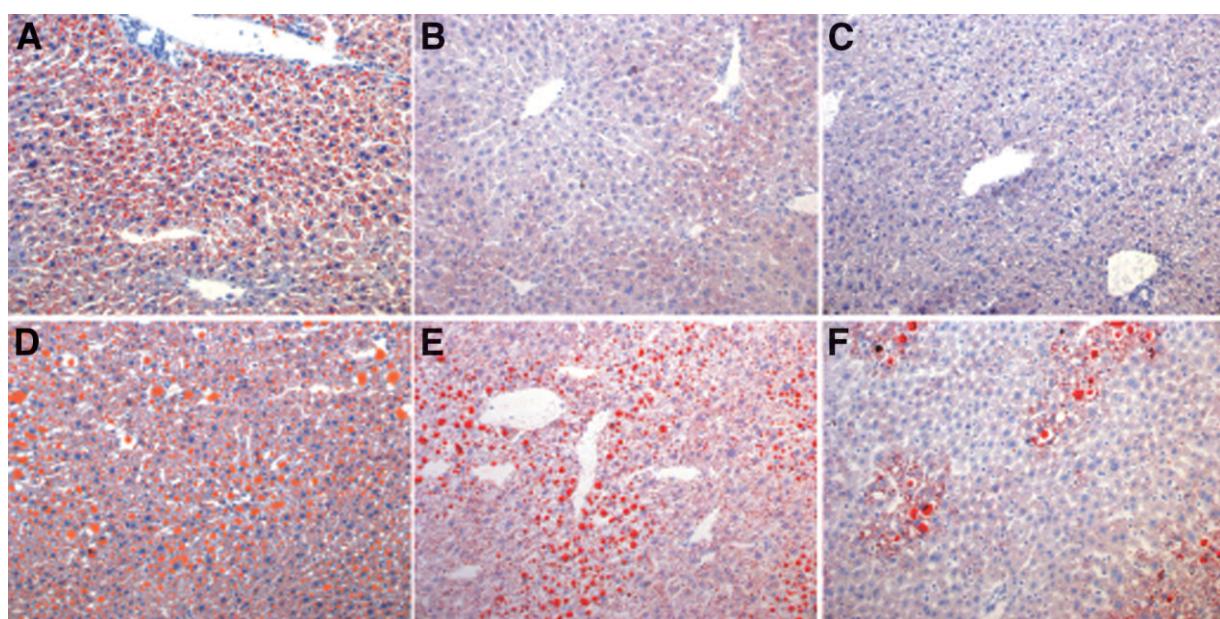


Fig. 1. Liver neutral lipids in mice fed the COC, REF, and FISH diets. Representative Oil Red O-stained frozen sections of liver from (A-C) wild-type and (D-F) PPAR $\alpha^{-/-}$  mice fed the (A, D) COC, (B, E) REF, or (C, F) FISH diet. (Original magnification  $\times 200$ .) Neutral lipids appear in red.

**Table 3. Hepatic FA Composition of Liver Total Lipids**

	Wild-type						PPAR $\alpha$ <sup>-/-</sup>						
	COC	REF	SUN	LIN	FISH	COc	REF	SUN	LIN	FISH			
C14:0	1.4 ± 0.2	0.4 ± 0.0 <sup>b</sup>	0.3 ± 0.0 <sup>b</sup>	0.4 ± 0.2 <sup>b</sup>	0.2 ± 0.0 <sup>b</sup>	2.8 ± 0.4 <sup>a</sup>	0.6 ± 0.1 <sup>a,b</sup>	0.5 ± 0.1 <sup>a,b</sup>	0.5 ± 0.1 <sup>b</sup>	0.3 ± 0.1 <sup>b</sup>	22.5 ± 2.8 <sup>a</sup>	SFA	
C16:0	25.5 ± 2.2	24.8 ± 0.7	24.2 ± 0.6	26.7 ± 1.2 <sup>d</sup>	26.4 ± 1.7	22.6 ± 0.8 <sup>a</sup>	20.2 ± 1.0 <sup>a</sup>	18.9 ± 2.6 <sup>a</sup>	18.9 ± 2.8 <sup>a</sup>	8.3 ± 3.6 <sup>a,c,e</sup>	8.3 ± 2.7 <sup>b</sup>	ω7 FA	
C18:0	5.7 ± 0.5	9.1 ± 0.6 <sup>b</sup>	9.4 ± 0.5 <sup>b</sup>	10.0 ± 0.6 <sup>b</sup>	8.8 ± 0.7 <sup>e</sup>	3.2 ± 0.9 <sup>a</sup>	3.4 ± 0.9 <sup>a</sup>	6.5 ± 3.3	6.3 ± 2.7 <sup>a</sup>	4.3 ± 0.9 <sup>b</sup>	4.3 ± 0.7 <sup>b</sup>	ω7 FA	
C16:1ω7	7.6 ± 1.5	3.1 ± 0.5 <sup>b</sup>	2.6 ± 0.2 <sup>b</sup>	3.4 ± 0.5 <sup>b</sup>	2.6 ± 0.2 <sup>b</sup>	11.5 ± 1.6 <sup>a</sup>	4.1 ± 0.6 <sup>a,b</sup>	2.9 ± 0.7 <sup>b</sup>	2.5 ± 0.6 <sup>b</sup>	3.5 ± 0.6 <sup>b</sup>	1.8 ± 0.4 <sup>a,b,c,e</sup>	ω7 FA	
C18:1ω7	9.6 ± 0.9	3.0 ± 0.6 <sup>b</sup>	3.7 ± 0.2 <sup>b</sup>	2.1 ± 0.2 <sup>b,c,d</sup>	2.2 ± 0.3 <sup>b,d</sup>	11.6 ± 1.8 <sup>a</sup>	4.7 ± 0.6 <sup>a,b</sup>	7.5 ± 2.1 <sup>a</sup>	1.9 ± 0.2 <sup>b,c</sup>	0.6 ± 0.1 <sup>a,b,c</sup>	0.6 ± 0.1 <sup>a,b,c</sup>	ω7 FA	
C16:1ω9	0.8 ± 0.1	0.5 ± 0.1 <sup>b</sup>	0.6 ± 0.1 <sup>b</sup>	0.3 ± 0.0 <sup>b,c,d</sup>	0.4 ± 0.0 <sup>b</sup>	1.2 ± 0.2 <sup>a</sup>	1.1 ± 0.2 <sup>a</sup>	0.7 ± 0.2 <sup>b,c</sup>	0.6 ± 0.1 <sup>a,b,c</sup>	0.6 ± 0.1 <sup>a,b,c</sup>	0.6 ± 0.1 <sup>a,b,c</sup>	ω7 FA	
C18:1ω9	33.4 ± 2.9	22.2 ± 1.8 <sup>b</sup>	19.9 ± 1.9 <sup>b</sup>	17.5 ± 2.2 <sup>b,c</sup>	19.4 ± 2.1 <sup>b</sup>	37.1 ± 3.0	36.2 ± 2.4 <sup>a</sup>	21.7 ± 4.9 <sup>b,c</sup>	21.5 ± 4.1 <sup>b,c</sup>	20.6 ± 5.1 <sup>a,c</sup>	ND	ω9 FA	
C20:1ω9	0.2 ± 0.2 <sup>f</sup>	0.3 ± 0.0 <sup>f</sup>	0.2 ± 0.0 <sup>f</sup>	0.3 ± 0.0 <sup>f</sup>	ND	ND	0.3 ± 0.1	0.3 ± 0.1 <sup>a</sup>	0.4 ± 0.1 <sup>a</sup>	0.3 ± 0.2	0.3 ± 0.2 <sup>a</sup>	ND	ω9 FA
C20:3ω9	2.4 ± 0.3	0.2 ± 0.1 <sup>b</sup>	0.1 ± 0.1 <sup>b</sup>	ND	ND	0.3 ± 0.1 <sup>a</sup>	0.1 ± 0.0 <sup>a,b</sup>	0.0 ± 0.0 <sup>a,b</sup>	ND	ND	ND	ND	ω9 FA
C18:2ω6	2.9 ± 0.5	15.0 ± 1.1 <sup>b</sup>	16.2 ± 1.0 <sup>b</sup>	8.9 ± 0.8 <sup>b,c,d</sup>	14.0 ± 1.5 <sup>b,d,e</sup>	5.5 ± 1.5 <sup>a</sup>	21.4 ± 2.1 <sup>a,b</sup>	31.2 ± 5.4 <sup>a,b,c</sup>	14.8 ± 3.1 <sup>a,b,c,d</sup>	21.3 ± 3.9 <sup>a,b,d,e</sup>	ND	ND	ω9 FA
C18:3ω6	2.0 ± 2.4	0.3 ± 0.1	0.4 ± 0.1	ND	ND	0.3 ± 0.4	0.3 ± 0.1	0.5 ± 0.2 <sup>a</sup>	0.1 ± 0.1 <sup>a,d</sup>	0.1 ± 0.1 <sup>a,d</sup>	0.1 ± 0.1 <sup>a,d</sup>	ND	ω9 FA
C20:2ω6	ND	0.2 ± 0.0	0.3 ± 0.0 <sup>c</sup>	ND	ND	0.0 ± 0.0	0.3 ± 0.1 <sup>a,d</sup>	0.7 ± 0.1 <sup>a,b,d</sup>	0.7 ± 0.1 <sup>a,b,d</sup>	0.1 ± 0.1 <sup>a,d</sup>	0.2 ± 0.1 <sup>a,b,d</sup>	ND	ω9 FA
C20:3ω6	0.7 ± 0.1	1.4 ± 0.1 <sup>b</sup>	1.6 ± 0.1 <sup>b,c</sup>	0.7 ± 0.1 <sup>c,d</sup>	1.1 ± 0.1 <sup>b,c,d,e</sup>	0.3 ± 0.1 <sup>a,d</sup>	0.5 ± 0.1 <sup>a,d</sup>	0.9 ± 0.1 <sup>a</sup>	0.3 ± 0.1 <sup>a,d</sup>	0.5 ± 0.2 <sup>a,b,d,e</sup>	0.5 ± 0.2 <sup>a,b,d,e</sup>	ND	ω9 FA
C20:4ω6	4.2 ± 0.7	11.2 ± 0.6 <sup>b</sup>	14.9 ± 1.1 <sup>b,c</sup>	2.8 ± 0.3 <sup>b,c,d</sup>	4.0 ± 0.2 <sup>c,d,e</sup>	2.0 ± 1.0 <sup>a,d</sup>	2.8 ± 1.1 <sup>a,d</sup>	8.1 ± 4.1 <sup>a</sup>	1.7 ± 0.7 <sup>a,d</sup>	2.4 ± 0.8 <sup>a,d</sup>	ND	ND	ω9 FA
C22:4ω6	ND	0.3 ± 0.0	0.6 ± 0.1 <sup>c</sup>	ND	ND	0.0 ± 0.0 <sup>c,d</sup>	0.2 ± 0.0 <sup>a</sup>	0.6 ± 0.1 <sup>c</sup>	0.0 ± 0.0 <sup>c,d</sup>	0.0 ± 0.0 <sup>c,d</sup>	0.1 ± 0.1 <sup>c</sup>	ND	ω9 FA
C22:5ω6	0.3 ± 0.2	0.4 ± 0.1	2.0 ± 0.3 <sup>b,c</sup>	ND	ND	0.2 ± 0.1 <sup>d</sup>	0.1 ± 0.0 <sup>a,d</sup>	1.3 ± 0.6 <sup>a</sup>	ND	ND	0.1 ± 0.1 <sup>d</sup>	ND	ω9 FA
C18:3ω3	ND	0.5 ± 0.1 <sup>e</sup>	ND	6.7 ± 1.3	0.4 ± 0.2 <sup>e</sup>	0.1 ± 0.1 <sup>a,e</sup>	1.2 ± 0.1 <sup>a,e</sup>	0.1 ± 0.1 <sup>a,e</sup>	1.7 ± 4.8 <sup>a</sup>	0.9 ± 0.5 <sup>e</sup>	ND	ND	ω3 FA
C20:3ω3	ND	0.3 ± 0.0 <sup>e,f</sup>	ND	0.4 ± 0.0	ND	0.1 ± 0.0 <sup>a,e</sup>	ND	0.5 ± 0.1	ND	0.5 ± 0.1	ND	ND	ω3 FA
C20:5ω3	0.1 ± 0.3 <sup>e,f</sup>	1.2 ± 1.3 <sup>d</sup>	0.4 ± 0.1	0.1 ± 0.2	ND	3.0 ± 0.2 <sup>e</sup>	0.1 ± 0.1 <sup>e,f</sup>	0.2 ± 0.0 <sup>e,f</sup>	0.1 ± 0.1 <sup>a,e,f</sup>	4.4 ± 2.4 <sup>a</sup>	3.0 ± 1.1	ND	ω3 FA
C22:5ω3	2.1 ± 0.3 <sup>c,e,f</sup>	6.6 ± 0.4 <sup>f</sup>	2.8 ± 0.3 <sup>c,e,f</sup>	10.0 ± 1.0 <sup>c,f</sup>	16.3 ± 0.9	0.7 ± 0.4 <sup>a,e,f</sup>	1.9 ± 0.8 <sup>a,e,f</sup>	1.4 ± 0.5 <sup>a,e,f</sup>	4.6 ± 1.8 <sup>a,f</sup>	13.4 ± 2.7 <sup>a</sup>	ND	ND	ω3 FA
Total SFA	32.6 ± 2.4	34.3 ± 1.2 <sup>*</sup>	33.9 ± 0.9	37.1 ± 1.4 <sup>b,d,*</sup>	35.5 ± 2.0 <sup>b</sup>	28.6 ± 1.0 <sup>a</sup>	24.3 ± 0.6 <sup>a</sup>	25.9 ± 5.2 <sup>a</sup>	25.9 ± 5.2 <sup>a</sup>	31.2 ± 5.8	31.2 ± 5.8	ω3 FA	
Total MUFA	51.6 ± 4.0	29.0 ± 2.9 <sup>b</sup>	27.0 ± 2.1 <sup>b</sup>	23.6 ± 2.6 <sup>c</sup>	24.6 ± 2.6 <sup>b</sup>	61.7 ± 4.0 <sup>a</sup>	46.5 ± 3.5 <sup>a,b</sup>	29.2 ± 6.0 <sup>c</sup>	29.2 ± 6.0 <sup>c</sup>	25.2 ± 5.7 <sup>b,c</sup>	25.2 ± 5.7 <sup>b,c</sup>	ω3 FA	
Total ω6 FA	10.1 ± 2.6	28.8 ± 1.5 <sup>b</sup>	36.0 ± 1.4 <sup>b,c</sup>	12.4 ± 0.8 <sup>c,d</sup>	19.0 ± 1.7 <sup>b,c,d,e</sup>	8.4 ± 2.7	25.5 ± 3.0 <sup>a,b</sup>	43.4 ± 3.8 <sup>a,b,c</sup>	17.0 ± 3.3 <sup>a,b,c,d</sup>	24.5 ± 3.5 <sup>a,b,d,e</sup>	19.1 ± 3.0 <sup>e</sup>	ω3 FA	
Total ω3 FA	3.4 ± 1.6	7.7 ± 0.4 <sup>b</sup>	2.9 ± 0.2 <sup>c</sup>	26.9 ± 1.8 <sup>b,c,d</sup>	20.8 ± 1.1 <sup>b,c,d,e</sup>	1.1 ± 0.6 <sup>a,e,f</sup>	3.6 ± 0.9 <sup>a,e,f</sup>	1.6 ± 0.4 <sup>a,e,f</sup>	28.1 ± 4.1	22.1 ± 3.7 <sup>a,b,c,d</sup>	22.1 ± 3.7 <sup>a,b,c,d</sup>	ω3 FA	
Total HUFA	13.0 ± 4.3	21.5 ± 1.0 <sup>b</sup>	22.6 ± 1.6 <sup>b</sup>	30.4 ± 2.0 <sup>b,c,d</sup>	25.9 ± 1.1 <sup>b,c,e</sup>	4.1 ± 2.2 <sup>a</sup>	13.1 ± 4.9 <sup>a,b</sup>	30.2 ± 4.2 <sup>b,c,d</sup>	13.1 ± 4.9 <sup>a,b</sup>	13.0 ± 5.1 <sup>a,b</sup>	21.2 ± 4.2 <sup>a,b,c,d,e</sup>	ω3 FA	
Total LHUFA	10.9 ± 2.2	20.7 ± 1.0 <sup>b</sup>	22.2 ± 1.5 <sup>b</sup>	23.7 ± 2.8 <sup>b</sup>	25.5 ± 1.2 <sup>b,c,d</sup>	3.8 ± 1.9 <sup>a</sup>	6.1 ± 2.1 <sup>a</sup>	12.5 ± 5.0 <sup>a,b</sup>	31.3 ± 5.5 <sup>a,b,c,f</sup>	13.0 ± 5.3 <sup>a,b,c,f</sup>	22.2 ± 4.3 <sup>a,b</sup>	ω3 FA	
Total precursors <sup>†</sup>	2.9 ± 0.5	15.5 ± 1.2 <sup>b</sup>	16.2 ± 1.0 <sup>b</sup>	15.6 ± 1.8 <sup>b</sup>	14.4 ± 1.7 <sup>b</sup>	5.6 ± 1.5 <sup>a</sup>	22.6 ± 2.2 <sup>a,b</sup>	31.3 ± 5.5 <sup>a,b,c,f</sup>	31.9 ± 5.3 <sup>a,b,c,f</sup>	22.2 ± 4.3 <sup>a,b</sup>	ND	ω3 FA	

NOTE. Data are expressed as the mean percentage of FA ± SD (n = 6 per group).

Abbreviations: HUFA, highly unsaturated FA (at least 3 double bonds); LHUFA, long-chain highly unsaturated FA (HUFA with more than 20 carbons); ND, not detected in any sample of the group; SFA, saturated fatty acid.

<sup>a</sup>Significant difference versus wild-type mice fed the same diet (Student t test; P < .05).<sup>b,c,d,e,f</sup>Significant difference versus the same genotype fed the COC, REF, SUN, LIN, or FISH diet, respectively (Tukey test; P < .05).<sup>\*</sup>Saturated FA for wild-type mice, REF vs. LIN (P = .0502).<sup>†</sup>C18:2ω6 + C18:3ω3.

**Table 4. Genotype and Diet Effects on Hepatic Gene Expression Monitored with INRARRAY 01.2**

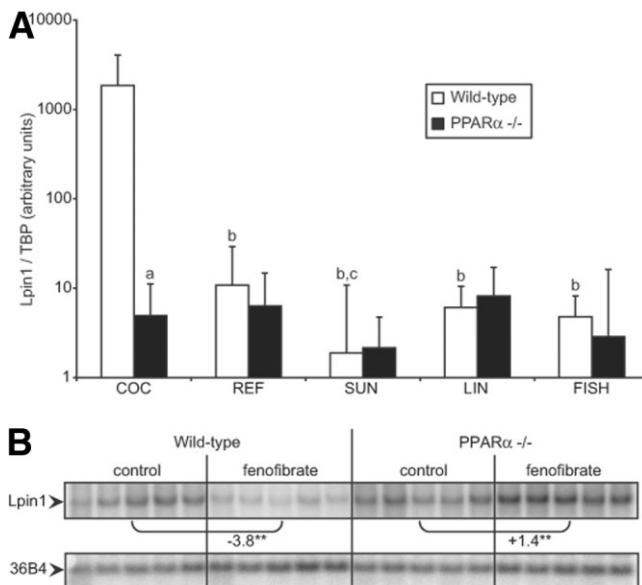
Gene	RefSeq	Genotype Effects				Diet Effects in Wild-type Mice						Diet Effects in PPAR $\alpha^{-/-}$ Mice				Function		
		COC	REF	SUN	LIN	FISH	REF/COC	SUN/COC	LIN/COC	FISH/COC	SUN/REF	LIN/REF	FISH/REF	SUN/COC	FISH/COC	SUN/REF	FISH/REF	
PECI	NM_011868	NS	NS	-1.6	NS	-2.2	NS	NS	NS	1.4	NS	NS	1.5	NS	NS	NS	NS	Enoyl-CoA isomerases
PMDCI	NM_016772	-1.6	-1.9	-2.5	-2.0	-3.1	NS	NS	1.3	1.6	NS	1.3	1.5	NS	NS	NS	NS	
HPNCL	NM_019975	-1.3	NS	-1.6	-1.3	-1.6	NS	-1.4	NS	NS	NS	NS	-1.7	NS	-1.4	NS	FA $\alpha$ -oxidation	
AOX	NM_015729	NS	NS	-1.7	-1.6	-1.7	NS	NS	NS	NS	NS	1.6	NS	NS	NS	NS		
BIEN	NM_023737	NS	NS	-1.5	NS	-2.1	NS	NS	NS	NS	NS	1.7	-1.5	-1.5	NS	NS		
THIOL	NM_130864	-1.5	NS	-3.2	-2.7	-3.2	NS	NS	NS	NS	NS	1.7	-2.0	NS	-1.9	NS	Peroxisomal and mitochondrial FA $\beta$ -oxidation	
CACP	NM_020520	NS	NS	-1.4	NS	-1.6	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS		
CPT2	NM_009949	NS	NS	-1.5	NS	-1.8	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS		
TP $\alpha$	NM_178878	NS	NS	-1.4	NS	-1.5	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS		
TP $\beta$	NM_145558	NS	NS	NS	NS	-1.6	NS	NS	NS	NS	NS	1.5	NS	NS	NS	NS		
mHMGCoAS	NM_008256	NS	NS	-1.9	-1.4	-2.2	NS	NS	1.5	NS	NS	1.6	NS	NS	NS	NS	Ketogenesis	
Cyp4a10	NM_010011	NS	NS	-2.3	-2.4	-4.8	NS	1.6	1.9	2.8	1.8	2.1	3.2	NS	NS	NS	Microsomal FA $\omega$ -hydroxylation	
Cyp4a14	NM_007822	NS	NS	-1.9	-1.8	-5.9	NS	NS	1.8	4.7	1.7	2.0	5.1	NS	NS	NS		
ACBP	NM_007830	-1.3	NS	-1.4	-1.6	-1.3	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	FA transport and trafficking	
L-FABP	NM_017399	NS	NS	-2.2	-2.4	-2.4	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS		
ACOTH	NM_012006	1.3	1.4	NS	1.5	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS		
PLTP	NM_011125	NS	NS	NS	NS	NS	-1.5	NS	-1.5	-1.4	NS	NS	NS	NS	NS	NS	Lipoprotein metabolism	
cHMGCoAS	NM_145942	NS	NS	-1.7	NS	NS	NS	NS	-1.7	NS	NS	NS	NS	NS	NS	NS	Cholesterol synthesis	
HMGCoARed	NM_008255	-1.6	NS	NS	NS	NS	-1.4	-1.4	-1.7	-1.7	NS	NS	NS	NS	NS	NS		
S14	NM_009381	-2.4	NS	NS	NS	NS	-2.9	-2.5	-2.3	-3.4	NS	NS	NS	NS	NS	NS	FA synthesis	
FAS	NM_007988	-4.4	NS	-3.7	-4.0	NS	NS	NS	NS	-3.0	NS	NS	NS	NS	NS	NS		
ACC2	NM_133904	NS	NS	NS	NS	NS	-1.6	-1.5	-1.7	-1.6	NS	NS	NS	NS	NS	NS		
Lpin1	NM_015763	-1.7	NS	NS	NS	1.8	-1.7	-2.5	-2.0	-2.3	NS	NS	-1.9	NS	-2.3	NS	TG synthesis	
G6PDH	NM_019468	NS	NS	NS	NS	NS	-1.5	NS	-1.5	NS	NS	NS	NS	NS	NS	NS	Glucose metabolism	
GK	NM_010292	-1.8	NS	-1.7	-1.5	NS	NS	NS	-1.7	NS	NS	NS	NS	NS	NS	NS		
G6Pase	NM_008061	NS	NS	NS	NS	NS	NS	NS	-2.0	NS	NS	NS	NS	NS	NS	NS		
ALDH3	NM_007437	-1.3	NS	-1.6	-1.3	-1.8	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	Detoxification	
Cyp2c29	NM_007815	1.8	1.8	NS	NS	NS	NS	NS	1.7	2.3	NS	1.9	2.6	NS	NS	NS	Phase I detoxification	
Cyp3a11	NM_007818	-1.9	-1.7	-2.4	-2.6	-2.8	NS	NS	2.0	2.7	NS	2.2	3.0	1.8	NS	1.9	(Cyp)	
GST $\alpha$	NM_010356	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	-1.5	NS	NS	NS	Phase II	
GST $\mu$	NM_010358	NS	NS	NS	NS	-1.6	-1.6	NS	NS	NS	NS	2.4	NS	NS	NS	NS	detoxification (GSTs)	
GST $\pi$	NM_013541	NS	NS	NS	-1.6	-1.6	NS	NS	1.6	2.1	NS	NS	NS	NS	NS	NS		
Slc10a1/Nicp	NM_011387	NS	1.4	NS	1.5	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	Transporter	
NR113/CAR	NM_009803	1.6	1.8	NS	1.8	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	Class II nuclear receptor	
NR0B2/SHP	NM_011850	NS	1.7	NS	1.5	NS	NS	NS	NS	NS	NS	1.5	NS	-1.5	NS	NS		
Lpin2	NM_022882	-1.4	NS	-1.5	NS	-1.4	NS	-1.5	e-1.4	NS	NS	NS	-1.6	NS	-1.5	NS	Unknown	

NOTE. Data are expressed as expression ratios (PPAR $\alpha$ -deficient/wild-type for genotype effects and diet/COC or REF for diet effects). No significant differences were observed for PPAR $\alpha^{-/-}$  mice between the REF, COC, and LIN diets. Shaded areas correspond to transcript ratios that do not reach statistical significance (NS in shaded areas = not significant).

contrast, in PPAR $\alpha^{-/-}$  mice, only 8 genes were modulated. Overall, the FISH diet, followed by the LIN diet, were responsible for the majority of gene expression changes observed in wild-type mice. Except for the induction of Cyp3a11 by the FISH diet, all other modulations seen in wild-type mice were absent in PPAR $\alpha^{-/-}$  mice. Additionally, the SUN diet reduced the expression of genes in PPAR $\alpha^{-/-}$  mice. Table 4 clearly illustrates that dietary effects were more pronounced in wild-type mice than in PPAR $\alpha^{-/-}$  mice, establishing that, even under low fat intake, PPAR $\alpha$  mediates the effects of dietary FAs on the expression of many genes present on the macroarray.

**Analysis of Lipid Metabolism Genes.** To gain insight into the molecular mechanisms underlying the effects observed on hepatic FA content and composition, we further examined genes involved in lipid metabolism. First, we noticed that Lpin1 was induced in the COC diet, suggesting a coregulation with genes involved in FA and cholesterol synthesis (Table 4). Q-PCR confirmed a striking 400-fold increase of Lpin1 messenger RNA un-

der the COC diet in wild-type livers, whereas no effect was observed in PPAR $\alpha^{-/-}$  mice (Fig. 2A). Because this effect appeared to be PPAR $\alpha$ -dependent, we tested the *in vivo* effect of fenofibrate on Lpin1 expression. We observed a PPAR $\alpha$ -dependent downregulation of Lpin1 (Fig. 2B). The expression of stearoyl-CoA desaturase 1 (SCD1), the main hepatic  $\Delta 9$ -desaturase isoform controlling mono-unsaturated FA synthesis, may also influence hepatic lipid content and composition.<sup>13</sup> PPAR $\alpha^{-/-}$  mice expressed lower levels of hepatic SCD1 than wild-type mice except when given the REF diet (Fig. 3A), suggesting that altered FA oxidation rather than increased expression of SCD1 (or other lipogenic enzymes) accounts for the hepatic steatosis developed by PPAR $\alpha$  knockout mice. Hepatic  $\Delta 9$ -desaturase activity was not significantly lowered in PPAR $\alpha^{-/-}$  mice on the REF diet. Whereas all PUFA-containing diets decreased SCD1 expression compared with the COC diet in wild-type mice, only the SUN, LIN, and FISH diets significantly reduced SCD1 expression in PPAR $\alpha^{-/-}$  mice. Changes in C16:



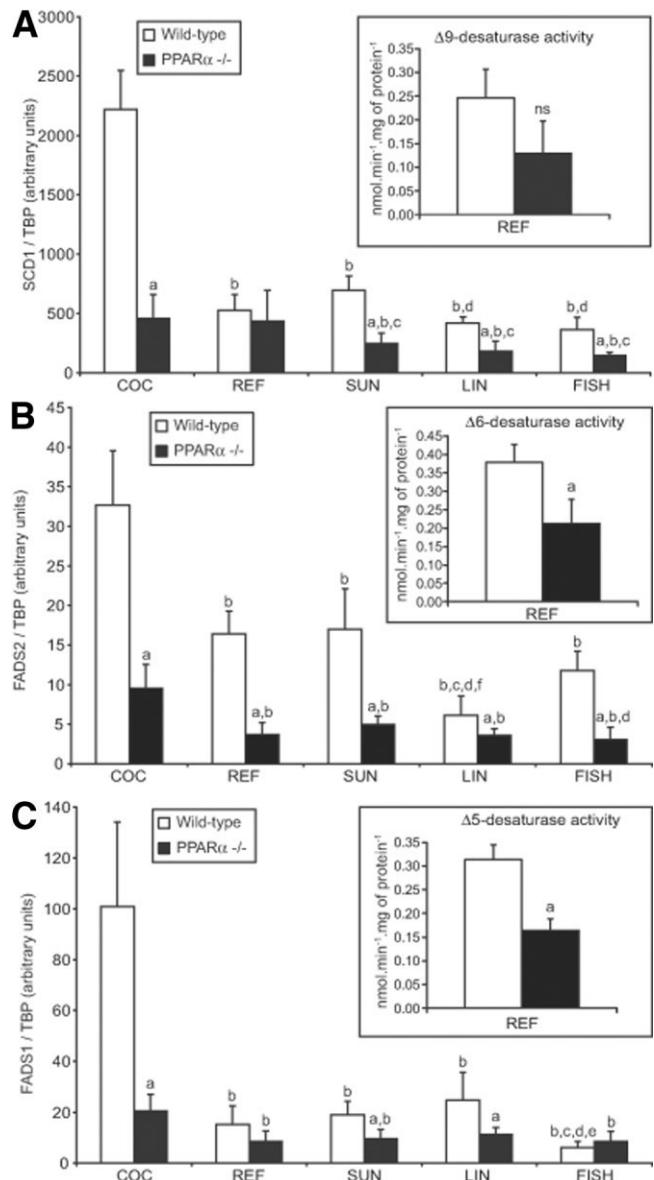
**Fig. 2.** Dietary and pharmacological regulation of Lpin1 messenger RNA levels. (A) Q-PCR quantification of hepatic Lpin1 expression (logarithmic scale) in the diet study. <sup>a</sup>Significant difference between the 2 genotypes (same diet). <sup>b,c</sup>Significant effect of the diet versus the COC or REF diet, respectively (same genotype). (B) Northern blot analysis of hepatic Lpin1 messenger RNA expression in wild-type and PPAR $\alpha^{-/-}$  mice after fenofibrate treatment. Fold changes are indicated. \*\* $P < .01$ .

$1\omega 9$  and  $C18:1\omega 9$  proportions paralleled the diet-induced changes in SCD1 expression within each genotype (Table 3).

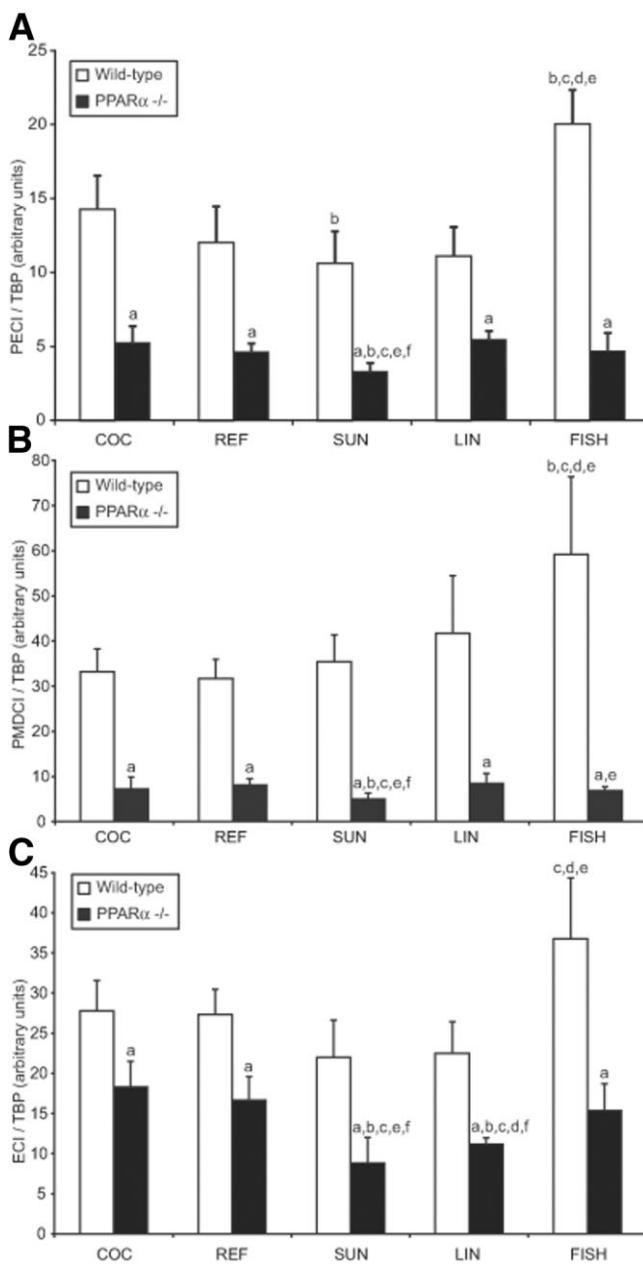
We next studied FADS2 and FADS1, the  $\Delta 6$  and  $\Delta 5$ -desaturases (reviewed by Nakamura and Nara<sup>14</sup>) that are involved in rate-limiting steps of long-chain PUFA biosynthesis from  $C18:2\omega 6$  and  $C18:3\omega 3$  (Fig. 3B-C). Both desaturases displayed reduced expression under PUFA-containing diets compared with the COC diet in the two genotypes. Furthermore, FADS2 and FADS1 were expressed—under most dietary conditions—at significantly lower levels in PPAR $\alpha^{-/-}$  livers. Similarly,  $\Delta 6$ - and  $\Delta 5$ -desaturase activities under the REF diet were significantly reduced in PPAR $\alpha^{-/-}$  mice. Finally, we studied enoyl-CoA isomerase, which allow PUFAs to enter the  $\beta$ -oxidation pathway. In accordance with the macroarray data, PECI, PMDCI, and ECI (monitored by Q-PCR only) always displayed lower expression in PPAR $\alpha^{-/-}$  livers (Fig. 4). The only consistent dietary effects observed were the increased expression of these genes by the FISH diet in wild-type mice and a decrease in their expression by the SUN diet in PPAR $\alpha^{-/-}$  mice.

**Analysis of Genes Related to Xenobiotic Metabolism.** Table 4 and multidimensional exploratory analyses (data not shown) highlighted a major effect of the genotype on the expression of the gene studied. This is well exemplified by the consistent reduction of Cyp3a11 ex-

pression in PPAR $\alpha^{-/-}$  livers (Table 4), which was confirmed via Northern blot analysis (Fig. 5). Furthermore, the FISH and LIN diets increased Cyp3a11 levels in wild-type and, to a lesser extent, PPAR $\alpha^{-/-}$  livers. Interestingly, PPAR $\alpha^{-/-}$  livers displayed a modest but significant overexpression of CAR, a master regulator of xenobiotic-metabolizing enzymes (Table 4, Fig. 6A,B [five alternative transcripts studied<sup>15</sup>]). Two independent experiments using fenofibrate (PPAR $\alpha$  activator) (Fig. 6C) or TCPOBOP (CAR activator) (Fig. 6D) confirmed



**Fig. 3.** Analysis of  $\Delta 9$  (SCD1),  $\Delta 6$  (FADS2), and  $\Delta 5$  (FADS1) FA-desaturase expression. Q-PCR quantification of hepatic (A) SCD1, (B) FADS2, and (C) FADS1 messenger RNA expression in the diet study is shown. <sup>a</sup>Significant difference between the 2 genotypes (same diet). <sup>b,c,d,e,f</sup>Significant effect of the diet versus the COC, REF, SUN, LIN, or FISH diet, respectively (same genotype). Enzyme activities measured under the REF diet are illustrated in the respective insets.



**Fig. 4.** Analysis of three hepatic enoyl-CoA isomerases. Q-PCR quantification of hepatic (A) PECL, (B) PMDCI, and (C) ECI messenger RNA expression in the diet study. <sup>a</sup>Significant difference between the two genotypes (same diet). <sup>b,c,d,e,f</sup>Significant effect of the diet versus the COC, REF, SUN, LIN, or FISH diet, respectively (same genotype).

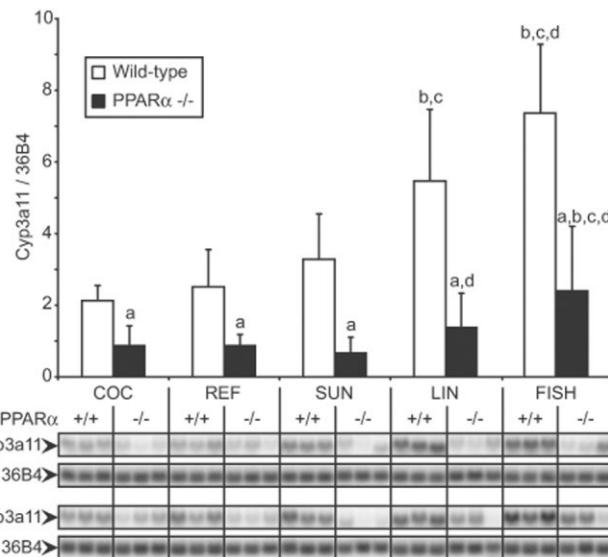
this overexpression. Fenofibrate had no effect on CAR expression, suggesting that CAR overexpression is an indirect consequence of PPAR $\alpha$  deficiency. TCPOBOP reduced hepatic CAR expression in both genotypes. Similarly, an intraperitoneal injection of phenobarbital reduced hepatic CAR expression after 3 hours (data not shown). TCPOBOP induced a more pronounced liver enlargement in PPAR $\alpha$ -/- (73% increase) versus wild-type mice (47%), as described previously.<sup>16</sup> Furthermore,

the Mdm2 transcript, a CAR target implicated in TCPOBOP-induced hepatocyte proliferation,<sup>17</sup> was significantly more induced by TCPOBOP in the PPAR $\alpha$ -/- genotype (Q-PCR, 2.4- vs. 2.0-fold [data not shown]).

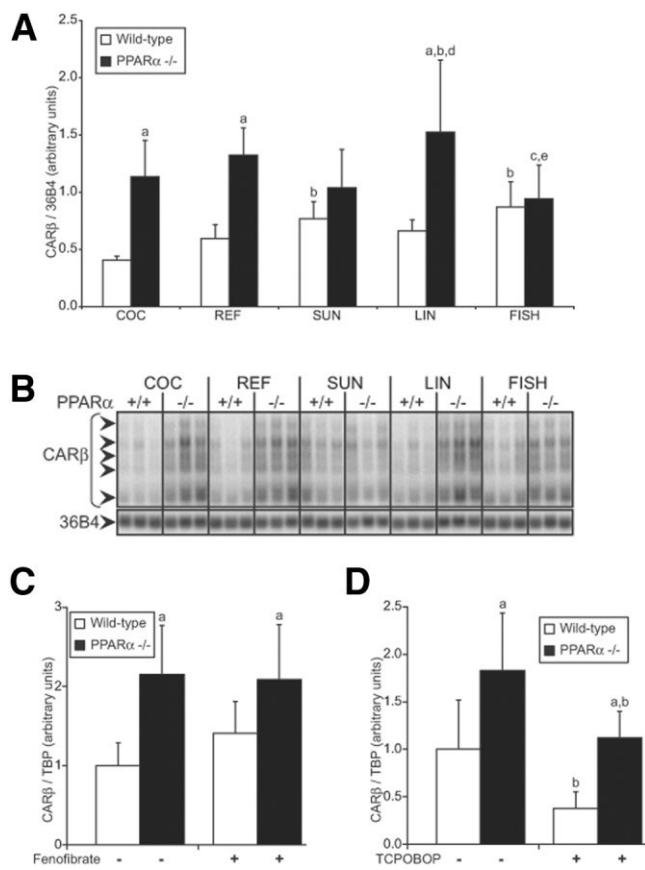
## Discussion

Our current knowledge on FA molecular signaling in physiology relies mostly on high-fat diet exposures. This study was designed to investigate, under conditions of low fat intake, the role of PPAR $\alpha$  in mediating the effects of dietary FA composition on hepatic gene expression, lipid content, and FA profile. Overall, our study, which fully complied with mouse dietary requirements (4%-6% fat intake) and used contrasted dietary FA profiles, highlighted a physiological role of PPAR $\alpha$  as a sensor of the quality of the ingested fat.

Beneficial effects of dietary PUFAs, especially  $\omega$ 3-PUFAs, on hepatic steatosis have been reported in the context of high fat input (59%).<sup>18</sup> This condition triggers the simultaneous transcriptional upregulation of hepatic FA oxidation through PPAR $\alpha$  activation and downregulation of hepatic lipogenesis through the inhibition of sterol regulatory element binding protein-1c.<sup>19,20</sup> Accordingly, under the COC diet in wild-type mice, we observed a higher hepatic FA content than under any other PUFA-containing diet (Table 2, Fig. 1). Consistent with previous reports,<sup>9</sup> we found that PPAR $\alpha$  deficiency led to higher hepatic triglyceride accumulation under all diets. However, in PPAR $\alpha$ -/- mice fed SUN, LIN, and FISH



**Fig. 5.** Northern blot analysis of Cyp3a11 and 36B4. Northern blot analysis (15  $\mu$ g of total RNA per lane) of hepatic Cyp3a11 normalized by 36B4. <sup>a</sup>Significant difference between the 2 genotypes (same diet). <sup>b,c,d</sup>Significant effect of the diet versus the COC, REF, or SUN diet, respectively (same genotype).



**Fig. 6.** Overexpression of CAR in PPAR $\alpha$ <sup>-/-</sup> liver and regulation by fenofibrate and TCPOBOP. (A) Quantification of five CAR transcripts (normalized by 36B4). <sup>a</sup>Significant difference between the two genotypes (same diet). <sup>b,c,d,e</sup>Significant effect of the diet versus the COC, REF, SUN, or LIN diet, respectively (same genotype). (B) Representative blots of hepatic CAR. (C, D) Q-PCR quantification of hepatic CAR expression in wild-type and PPAR $\alpha$ <sup>-/-</sup> mice after fenofibrate (C) or TCPOBOP (D) treatment. <sup>a</sup>Significant difference between the two genotypes. <sup>b</sup>Significant effect of the treatment.

diets, TG accumulation was reduced, which is consistent with the beneficial effect of PUFAs on hepatic steatosis reported in other rodent models.<sup>19,21,22</sup> The absence of an effect of the REF diet in PPAR $\alpha$ <sup>-/-</sup> mice further suggests that when the hepatic FA sensor PPAR $\alpha$  is compromised, some PUFAs can achieve their beneficial effect only when they have reached a critical threshold rate in the diet. Remarkably, in our study, the replacement of 15% of the fat with fish oil (FISH vs. REF) led to a significant reduction of hepatic FA accumulation in both genotypes. This suggests that physiologically relevant levels of dietary FAs present in fish oil have a beneficial impact on hepatic FA accumulation independently of PPAR $\alpha$ . Together, these results further support the hypothesis that fish oil can be used as a therapeutic approach in the treatment of nonalcoholic fatty liver disease.<sup>23</sup> Unexpectedly, our macroarray study revealed that sterol regulatory element binding protein target genes involved in cholesterol and FA syn-

thesis were downregulated by PUFA-containing diets in wild-type mice only (Table 4). Many of these genes display significant circadian variations in expression, some of which are altered in PPAR $\alpha$ <sup>-/-</sup> mice.<sup>24,25</sup> A role for PPAR $\alpha$  in the dietary downregulation of these genes cannot be excluded but requires additional investigation, including a time-course study. SCD1 plays an important role in hepatic cholesterol ester and triglyceride biosynthesis.<sup>13</sup> Its disruption reduces hepatic steatosis in PPAR $\alpha$ <sup>-/-</sup> mice.<sup>26</sup> In our study, all the diets that reduced SCD1 expression (Fig. 3A) in a given genotype also reduced hepatic fat stores (Table 2) and proportion of C18:1 $\omega$ 9, the major hepatic mono-unsaturated FA (Table 3). Thus, downregulation of SCD1 may contribute to the reduction of hepatic lipid stores in both genotypes. Moreover, we report the regulation of lipid expression by dietary FAs. The mouse null mutation of lipin accounts for the fatty liver dystrophy phenotype,<sup>27</sup> which combines metabolic disorders, lipodystrophy, and neuropathy.<sup>28</sup> Lipin1 has been related to adipocyte differentiation,<sup>29</sup> obesity, energy expenditure, and fuel partitioning,<sup>30</sup> and very recently has been identified as a phosphatidate phosphatase,<sup>31</sup> the penultimate enzyme in TG synthesis. The Lipin1 hepatic transcript is negatively regulated by PUFA-containing diets or fenofibrate in wild-type mice only (Fig. 2). Its coregulation with cholesterol and FA synthesis enzymes (Table 4) is consistent with its enzymatic function.<sup>31</sup> Recently, Lipin1 was described as an enhancer of the PGC-1 $\alpha$ /PPAR $\alpha$  pathway.<sup>32</sup> Adenoviral overexpression of Lipin1 upregulated PPAR $\alpha$ , increased FA oxidation and decreased lipogenesis while increasing liver TG content.<sup>32</sup> The elevated Lipin1 expression under our COC diet (wild-type) (Fig. 2) was consistently associated with increased hepatic TG stores. However, we observed a different gene expression pattern than Finck et al.<sup>32</sup> This discrepancy may be due to the specific hepatic FA profile of COC-fed mice, which may not provide appropriate PPAR $\alpha$  ligands and/or differentially affect other transcriptional pathways. Further studies are required to understand the interplay between Lipin1 biochemical and transcriptional functions and its potential as a drug target.

Collectively, our results suggest that reduction of SCD1 expression may contribute to the reduction of hepatic TG content by PUFA-containing diets. In wild-type mice, additional mechanisms include PPAR $\alpha$ -dependent induction of FA catabolism, reduction of lipogenesis, and TG synthesis (including Lipin1). Further studies in PPAR $\alpha$ <sup>-/-</sup> mice are underway to evaluate the possible role of TG secretion and intestinal lipid absorption in these beneficial effects of PUFAs.

In addition, dietary FA composition markedly impacted the hepatic FA profile in both genotypes. Consis-

tent with other reports,<sup>33</sup> we observed that PPAR $\alpha^{-/-}$  livers accumulate the  $\omega 3$  (C18:3 $\omega 3$ ) and  $\omega 6$  (C18:2 $\omega 6$ ) FA precursors at the expense of long-chain PUFAs (Table 2). First, this may reflect an altered ability to synthesize long-chain PUFAs through the elongase/desaturase pathways. Indeed, PPAR $\alpha$  is implicated in the transcriptional control of FA desaturases,<sup>10</sup> and Li et al.<sup>34</sup> have reported the role of PPAR $\alpha$  in the nutritional regulation of PUFA synthesis. We observed reduced expression and activity of  $\Delta 6$ - and  $\Delta 5$ -desaturases in PPAR $\alpha^{-/-}$  livers (Fig. 3), which is likely to reduce long-chain PUFA synthesis and contribute to the precursor accumulations. Interestingly, alteration in  $\Delta 6$ - and  $\Delta 5$ -desaturase levels developed with age and were not observed in younger PPAR $\alpha^{-/-}$  mice<sup>10</sup> (data not shown). Second, PPAR $\alpha^{-/-}$  livers displayed reduced expression of three enoyl-CoA isomerases (Table 4, Fig. 4), which may also contribute to precursor accumulations, as enoyl-CoA isomerase have been shown *in vivo* to be rate-limiting for linoleic and  $\alpha$ -linolenic acids to enter  $\beta$ -oxidation.<sup>35</sup> Thus, impairments of both PUFA-desaturases and enoyl-CoA isomerases likely play a role in the combined depletion of long-chain PUFAs and the converse accumulation of their precursors in PPAR $\alpha^{-/-}$  livers.

To evaluate the safety of a therapeutic intervention with dietary FAs, potential drug–diet interactions involving xenobiotic-metabolizing enzymes should also be investigated. As a key factor of drug metabolism, Cyp3a11 induction under diets containing high  $\omega 3$ -FA levels (Fig. 5) may raise concerns. However, recent studies identified PUFAs as inhibitors of rodent and human CYP3A activities.<sup>36,37</sup> The induction of Cyp3a11 transcript, also reported by others,<sup>38</sup> may thus be secondary to the reduction in enzymatic activity and restore a constitutive CYP3A hepatic activity. We also observed an overexpression of CAR in PPAR $\alpha^{-/-}$  livers (Fig. 6). Our preliminary results on TCPOBOP-mediated liver enlargement and Mdm2 induction support an increased CAR function in PPAR $\alpha^{-/-}$  livers. Interestingly, CAR activation downregulated its own transcript (Fig. 6D), which could represent a negative feedback loop preventing excessive induction of xenobiotic-metabolizing enzymes after xenochemical exposure.

In conclusion, this study further defines the role of PPAR $\alpha$  as a hepatic dietary FA sensor under low fat intake conditions and delineates its importance in regulating hepatic FA content and composition.

**Acknowledgment:** We are grateful to Dr. Daniel Catheline for expertise and assistance in analytical procedures. We thank Colette Bétoulières and Gérard Galy for excellent technical assistance. We thank Talal Al Saati and

Florence Capilla (Histopathology core facility of IFR30, Toulouse, France) for histology studies. We are grateful to Dr. Franck Jess Gonzalez for the generous gift of the PPAR $\alpha^{-/-}$  mouse line. We thank Dr. Suhasini Kulkarni, Pr. Alain Baccini, and Dr. Christelle Robert-Granié for critical review of the manuscript.

## References

- Mandard S, Muller M, Kersten S. Peroxisome proliferator-activated receptor alpha target genes. *Cell Mol Life Sci* 2004;61:393-416.
- Lee SS, Pineau T, Drago J, Lee EJ, Owens JW, Kroetz DL, et al. Targeted disruption of the alpha isoform of the peroxisome proliferator-activated receptor gene in mice results in abolition of the pleiotropic effects of peroxisome proliferators. *Mol Cell Biol* 1995;15:3012-3022.
- Hostettler HA, Petrescu AD, Kier AB, Schroeder F. Peroxisome proliferator-activated receptor alpha interacts with high affinity and is conformationally responsive to endogenous ligands. *J Biol Chem* 2005;280:18667-18682.
- Patsouris D, Reddy JK, Muller M, Kersten S. Peroxisome proliferator-activated receptor alpha mediates the effects of high fat diet on hepatic gene expression. *Endocrinology* 2006;147:1508-1516.
- Ren B, Thelen AP, Peters JM, Gonzalez FJ, Jump DB. Polyunsaturated fatty acid suppression of hepatic fatty acid synthase and S14 gene expression does not require peroxisome proliferator-activated receptor alpha. *J Biol Chem* 1997;272:26827-26832.
- Davidson LA, Nguyen DV, Hokanson RM, Callaway ES, Isett RB, Turner ND, et al. Chemopreventive n-3 polyunsaturated fatty acids reprogram genetic signatures during colon cancer initiation and progression in the rat. *Cancer Res* 2004;64:6797-6804.
- Simopoulos AP. Omega-3 fatty acids in inflammation and autoimmune diseases. *J Am Coll Nutr* 2002;21:495-505.
- Jump DB, Botolin D, Wang Y, Xu J, Christian B, Demeure O. Fatty acid regulation of hepatic gene transcription. *J Nutr* 2005;135:2503-2506.
- Costet P, Legendre C, More J, Edgar A, Galtier P, Pineau T. Peroxisome proliferator-activated receptor alpha-isoform deficiency leads to progressive dyslipidemia with sexually dimorphic obesity and steatosis. *J Biol Chem* 1998;273:29577-29585.
- Guillou H, Martin P, Jan S, D'Andrea S, Roulet A, Catheline D, et al. Comparative effect of fenofibrate on hepatic desaturases in wild-type and peroxisome proliferator-activated receptor alpha-deficient mice. *Lipids* 2002;37:981-989.
- Martin PG, Lasserre F, Calleja C, Van Es A, Roulet A, Concorde D, et al. Transcriptional modulations by RXR agonists are only partially subordinated to PPARalpha signaling and attest additional, organ-specific, molecular cross-talks. *Gene Expr* 2005;12:177-192.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289-300.
- Miyazaki M, Kim YC, Gray-Keller MP, Attie AD, Ntambi JM. The biosynthesis of hepatic cholesterol esters and triglycerides is impaired in mice with a disruption of the gene for stearoyl-CoA desaturase 1. *J Biol Chem* 2000;275:30132-30138.
- Nakamura MT, Nara TY. Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases. *Annu Rev Nutr* 2004;24:345-376.
- Choi HS, Chung M, Tzameli I, Simha D, Lee YK, Seol W, et al. Differential transactivation by two isoforms of the orphan nuclear hormone receptor CAR. *J Biol Chem* 1997;272:23565-23571.
- Columbano A, Ledda-Columbano GM, Pibiri M, Concas D, Reddy JK, Rao MS. Peroxisome proliferator-activated receptor-alpha mice show enhanced hepatocyte proliferation in response to the hepatomitogen 1,4-bis [2-(3,5-dichloropyridyloxy)] benzene, a ligand of constitutive androstane receptor. *HEPATOLOGY* 2001;34:262-266.
- Huang W, Zhang J, Washington M, Liu J, Parant JM, Lozano G, et al. Xenobiotic stress induces hepatomegaly and liver tumors via the nuclear receptor constitutive androstane receptor. *Mol Endocrinol* 2005;19:1646-1653.

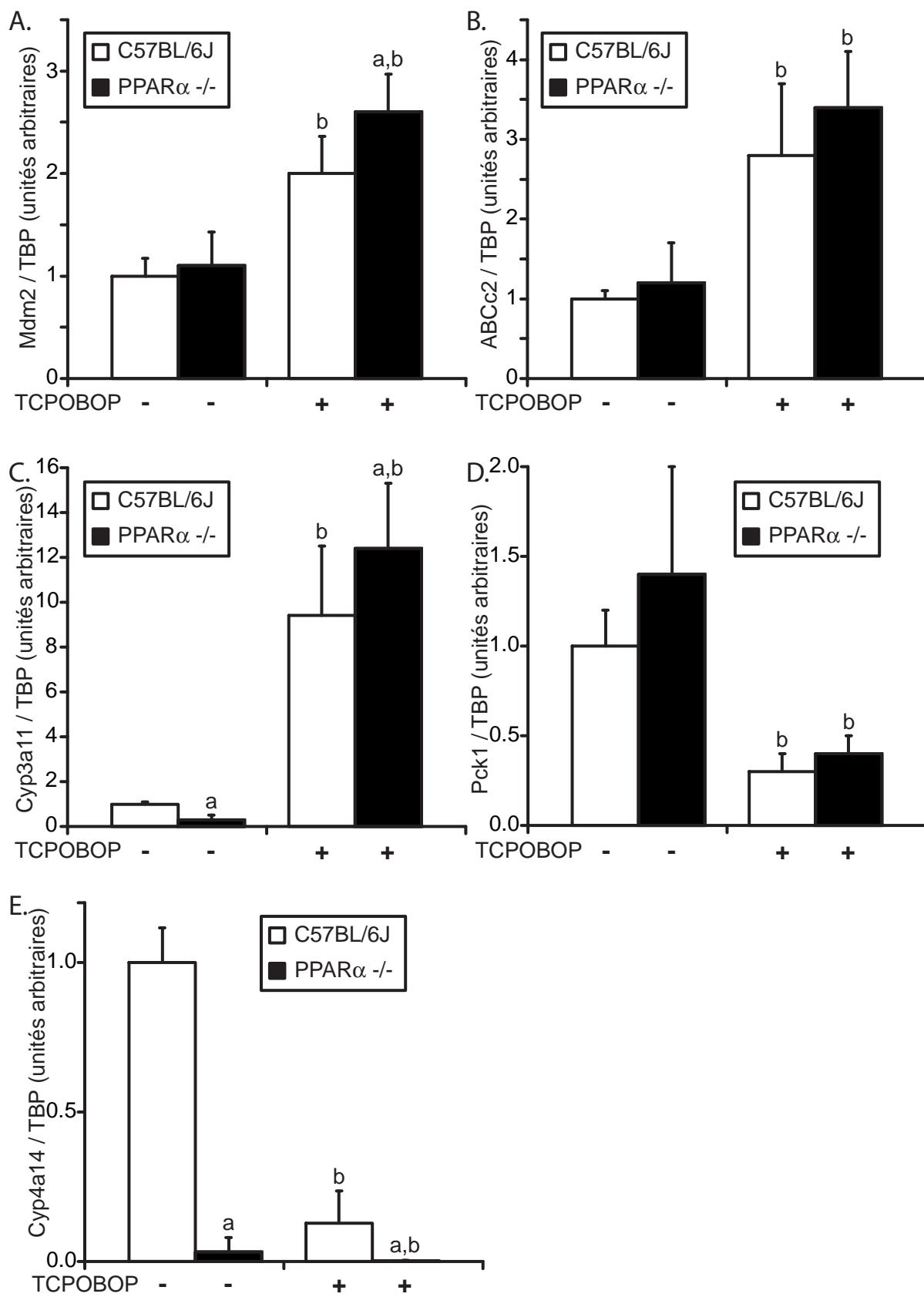
18. Neschen S, Moore I, Regitnig W, Yu CL, Wang Y, Pypaert M, et al. Contrasting effects of fish oil and safflower oil on hepatic peroxisomal and tissue lipid content. *Am J Physiol Endocrinol Metab* 2002;282:E395-E401.
19. Sekiya M, Yahagi N, Matsuzaka T, Najima Y, Nakakuki M, Nagai R, et al. Polyunsaturated fatty acids ameliorate hepatic steatosis in obese mice by SREBP-1 suppression. *HEPATOLOGY* 2003;38:1529-1539.
20. Xu J, Teran-Garcia M, Park JH, Nakamura MT, Clarke SD. Polyunsaturated fatty acids suppress hepatic sterol regulatory element-binding protein-1 expression by accelerating transcript decay. *J Biol Chem* 2001;276:9800-9807.
21. Alway IP, Gura K, Nose V, Zausche B, Javid P, Garza J, et al. Omega-3 fatty acid supplementation prevents hepatic steatosis in a murine model of nonalcoholic fatty liver disease. *Pediatr Res* 2005;57:445-452.
22. Levy JR, Clore JN, Stevens W. Dietary n-3 polyunsaturated fatty acids decrease hepatic triglycerides in Fischer 344 rats. *HEPATOLOGY* 2004;39:608-616.
23. Capanni M, Calella F, Biagini MR, Genise S, Raimondi L, Bedogni G, et al. Prolonged n-3 polyunsaturated fatty acid supplementation ameliorates hepatic steatosis in patients with non-alcoholic fatty liver disease: a pilot study. *Aliment Pharmacol Ther* 2006;23:1143-1151.
24. Patel DD, Knight BL, Wiggins D, Humphreys SM, Gibbons GF. Disturbances in the normal regulation of SREBP-sensitive genes in PPAR alpha-deficient mice. *J Lipid Res* 2001;42:328-337.
25. Knight BL, Hebbachi A, Hauton D, Brown AM, Wiggins D, Patel DD, et al. A role for PPARalpha in the control of SREBP activity and lipid synthesis in the liver. *Biochem J* 2005;389:413-421.
26. Miyazaki M, Dobrzyn A, Sampath H, Lee SH, Man WC, Chu K, et al. Reduced adiposity and liver steatosis by stearoyl-CoA desaturase deficiency are independent of peroxisome proliferator-activated receptor-alpha. *J Biol Chem* 2004;279:35017-35024.
27. Peterfy M, Phan J, Xu P, Reue K. Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nat Genet* 2001;27:121-124.
28. Rehnmark S, Giometti CS, Slavin BG, Doolittle MH, Reue K. The fatty liver dystrophy mutant mouse: microvesicular steatosis associated with altered expression levels of peroxisome proliferator-regulated proteins. *J Lipid Res* 1998;39:2209-2217.
29. Phan J, Peterfy M, Reue K. Lipin expression preceding peroxisome proliferator-activated receptor-gamma is critical for adipogenesis in vivo and in vitro. *J Biol Chem* 2004;279:29558-29564.
30. Phan J, Reue K. Lipin, a lipodystrophy and obesity gene. *Cell Metab* 2005;1:73-83.
31. Han GS, Wu WI, Carman GM. The *Saccharomyces cerevisiae* Lipin homolog is a Mg<sup>2+</sup>-dependent phosphatidate phosphatase enzyme. *J Biol Chem* 2006;281:9210-9218.
32. Finck BN, Gropler MC, Chen Z, Leone TC, Croce MA, Harris TE, et al. Lipin 1 is an inducible amplifier of the hepatic PGC-1alpha/PPARalpha regulatory pathway. *Cell Metab* 2006;4:199-210.
33. Lee SS, Chan WY, Lo CK, Wan DC, Tsang DS, Cheung WT. Requirement of PPARalpha in maintaining phospholipid and triacylglycerol homeostasis during energy deprivation. *J Lipid Res* 2004;45:2025-2037.
34. Li Y, Nara TY, Nakamura MT. Peroxisome proliferator-activated receptor alpha is required for feedback regulation of highly unsaturated fatty acid synthesis. *J Lipid Res* 2005;46:2432-2440.
35. Janssen U, Stoffel W. Disruption of mitochondrial beta-oxidation of unsaturated fatty acids in the 3,2-trans-enoyl-CoA isomerase-deficient mouse. *J Biol Chem* 2002;277:19579-19584.
36. Hirunpanich V, Katagi J, Sethabouppha B, Sato H. Demonstration of docosahexaenoic acid as a bioavailability enhancer for CYP3A substrates: in vitro and in vivo evidence using cyclosporin in rats. *Drug Metab Dispos* 2006;34:305-310.
37. Yao HT, Chang YW, Lan SJ, Chen CT, Hsu JT, Yeh TK. The inhibitory effect of polyunsaturated fatty acids on human CYP enzymes. *Life Sci* 2006;79:2432-2440.
38. Berger A, Mutch DM, German JB, Roberts MA. Dietary effects of arachidonate-rich fungal oil and fish oil on murine hepatic and hippocampal gene expression. *Lipids Health Dis* 2002;1:2.



### V.3. RESULTATS COMPLEMENTAIRES

A la suite de ces travaux, nous avons notamment souhaité clarifier si la légère surexpression du récepteur CAR observée dans cette étude avait ou non des conséquences fonctionnelles et s'il existait une possibilité d'interaction entre les voies de signalisation de CAR et de PPAR $\alpha$ . Des données de la littérature suggéraient, de manière intéressante, que le TCPOBOP, un activateur pharmacologique de CAR, provoquait une prolifération accrue des hépatocytes chez les souris PPAR $\alpha$ -/- (Columbano et al., 2001). De plus, des interactions entre les voies de signalisation PPAR $\alpha$  et CAR au niveau des promoteurs de certains gènes cibles étaient suggérées par certains auteurs (Kassam et al., 2000; Stoner et al., 2007). Un criblage par puces à ADN des effets du phénobarbital (un activateur de CAR qui provoque la translocation nucléaire de ce RN sans se fixer sur CAR) a été réalisé sur des souris de type sauvage et de type CAR-/- (Ueda et al., 2002), et montre une diminution de l'expression de certains gènes cibles de PPAR $\alpha$ , dont une enoyl-CoA isomérase et la carnitine palmitoyltransférase I chez les souris de type sauvage uniquement. De plus, le phenobarbital augmente l'expression des deux gènes cibles répondant le plus fortement à PPAR $\alpha$  : Cyp4a10 et Cyp4a14 chez les souris de type CAR-/- uniquement (Ueda et al., 2002). Ces résultats suggèrent que l'activation de CAR par le phenobarbital pourrait inhiber l'action de PPAR $\alpha$  au niveau hépatique. Enfin, Med1 constitue un coactivateur indispensable au fonctionnement des voies de signalisation de PPAR $\alpha$  et de CAR dans le foie (Jia et al., 2005; Jia et al., 2004), suggérant une possibilité d'interaction entre ces deux voies via la séquestration éventuelle de ce cofacteur. Nous avons donc réalisé deux études différentes : la première reproduisant les conditions d'exposition au TCPOBOP de souris de type sauvage et de type PPAR $\alpha$ -/- (injection intrapéritonéale, 3 mg/Kg/j, 3 jours, n=5 par groupe) décrites dans (Columbano et al., 2001) et la seconde constituant une activation à court terme du récepteur CAR par le phenobarbital (euthanasie et dissection des animaux 3h après une injection intrapéritonéale correspondant à 80 mg/Kg, n=6 par groupe). Le phénobarbital, comme la phénoxytoïne également utilisée dans une autre expérience préliminaire (non présentée ici), provoquent la translocation nucléaire de CAR *in vivo*, déjà observable 3 heures après l'administration par voie i.p. des composés (Jackson et al., 2004; Kawamoto et al., 1999).

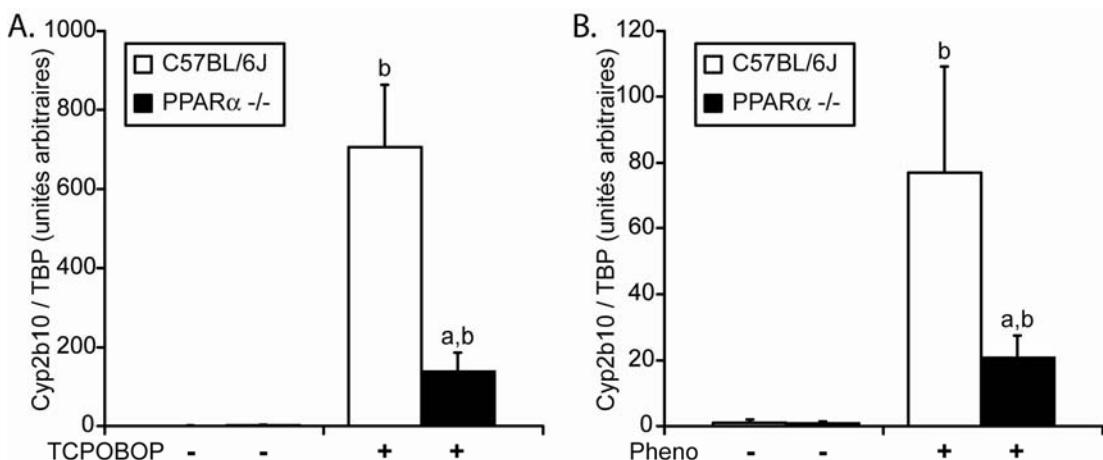
Dans un premier temps, nous avons confirmé que l'hépatomégalie provoquée par le TCPOBOP est significativement plus marquée chez les souris PPAR $\alpha$ -/- que chez les souris de type sauvage (augmentation de la masse relative du foie de 73% chez les souris PPAR $\alpha$ -/-



**Figure 57 : Effets du TCPOBOP sur quelques gènes cibles hépatiques**

Des échantillons d'ARN totaux hépatiques issus de souris de type sauvage (C57BL/6J) ou PPAR $\alpha$ -déficiente (PPAR $\alpha$  -/-) ayant subi un traitement par le TCPOBOP (3 mg/Kg/j, i.p., 3 jours, n=5/groupe) ou par le véhicule seul, ont été reverse transcrits (Superscript II, Invitrogen, France). Les ADNC correspondants ont subi des PCR en temps réel (approche TaqMan, Assays-on-Demand, Applied Biosystems, France) pour mesurer l'expression de Mdm2 (A), Abcc2 (B), Cyp3a11 (C), Pepck1 (D) et Cyp4a14 (E). Dans tous les cas, les données sont calibrées par la mesure de l'expression de la TATA-box binding protein (TBP) et exprimées relativement aux souris de type sauvage traitées par le véhicule seul. a : effet significatif du génotype, b : effet significatif du traitement

contre 47% chez les souris de type sauvage). Nous avons ensuite évalué l'expression hépatique de plusieurs gènes modulés positivement ou négativement par le TCPOBOP à partir des échantillons de cette étude. Les résultats sont présentés en [Figure 57](#). Les modulations d'expression observées se sont toutes révélées être au minimum d'une ampleur équivalente chez les souris PPAR $\alpha$ -/- par rapport aux souris de type sauvage et parfois les modulations se sont révélées légèrement supérieures chez les souris PPAR $\alpha$ -/- (voir notamment Mdm2, [Figure 57A](#) et Cyp3a11, [Figure 57C](#)). De manière intéressante, malgré la faible expression constitutive de Cyp4a14 chez les souris PPAR $\alpha$ -/-, le TCPOBOP semble malgré tout capable de réprimer l'expression de ce gène chez ces souris ([Figure 57E](#)). Deux études (TCPOBOP et phenobarbital) réalisées sur les puces INRAarray 01.3 ont permis de confirmer ces principales observations et d'identifier de nouvelles cibles potentielles des activateurs de CAR au niveau hépatique (PGP Martin, résultats non présentés). L'une des cibles hépatiques les mieux décrites de CAR chez la souris est sans doute le gène Cyp2b10 (Honkakoski et al., 1998). Nous avons donc également mesuré l'expression de ce gène dans le foie des souris de type sauvage et de type PPAR $\alpha$ -/- soumises soit à un traitement à moyen terme par le TCPOBOP (3 jours), soit un traitement à court terme par le phenobarbital (3 heures).



[Figure 58](#) : Induction de Cyp2b10 par les activateurs de CAR

Des échantillons d'ARN totaux hépatiques issus de souris de type sauvage (C57BL/6J) ou PPAR $\alpha$ -déficiente (PPAR $\alpha$ -/-) ayant subi un traitement par le TCPOBOP (3 mg/Kg/j, i.p., 3 jours, n=5/groupe, figure A), par le phenobarbital (80 mg/Kg, 1 injection i.p. unique, suivie par le prélèvement des échantillons 3 heures après, figure B) ou par le véhicule seul (contrôles propres à chaque expérience), ont été reverse transcrits (Superscript II, Invitrogen, France). Les ADNC correspondants ont subi des PCR en temps réel (approche TaqMan, Assays-on-Demand, Applied Biosystems, France) pour mesurer l'expression de Cyp2b10. Dans tous les cas, les données sont calibrées par la mesure de l'expression de la TATA-box binding protein (TBP) et exprimées relativement aux souris de type sauvage traitées par le véhicule seul. a : effet significatif du génotype, b : effet significatif du traitement.

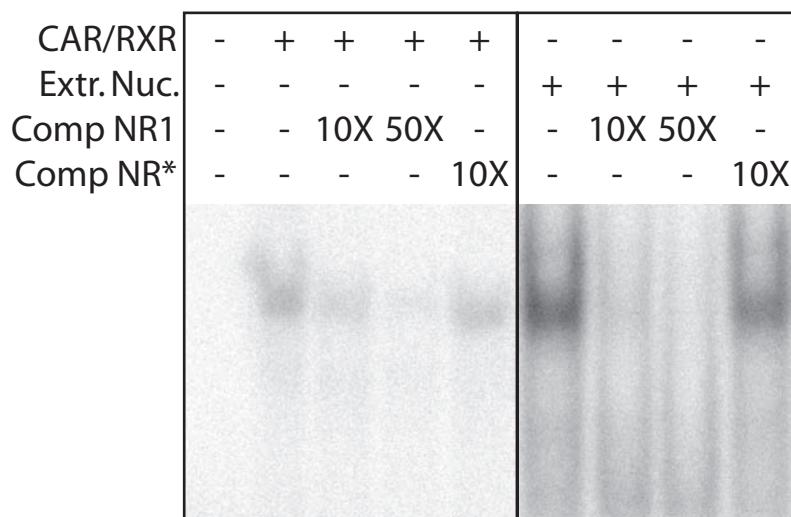
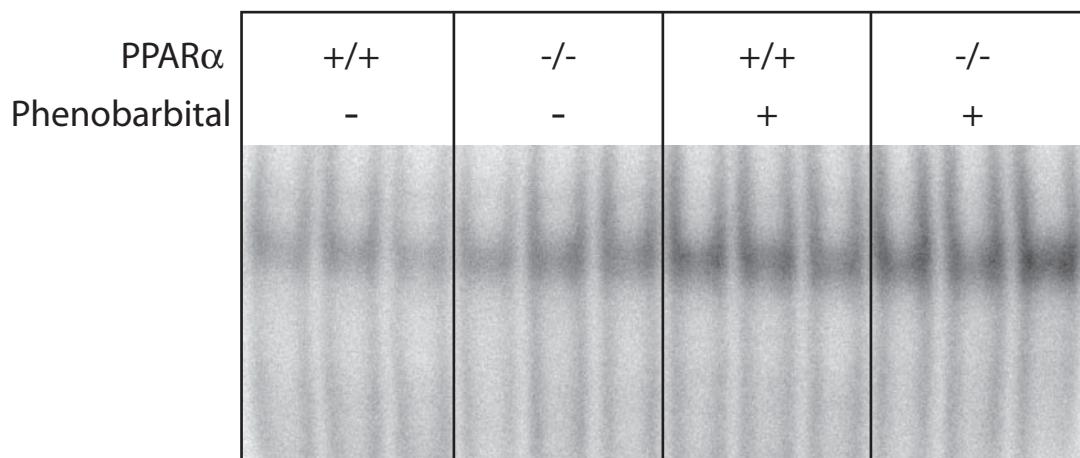
**A.****B.**

Figure 59 : Fixation de CAR sur le promoteur de Cyp2b10

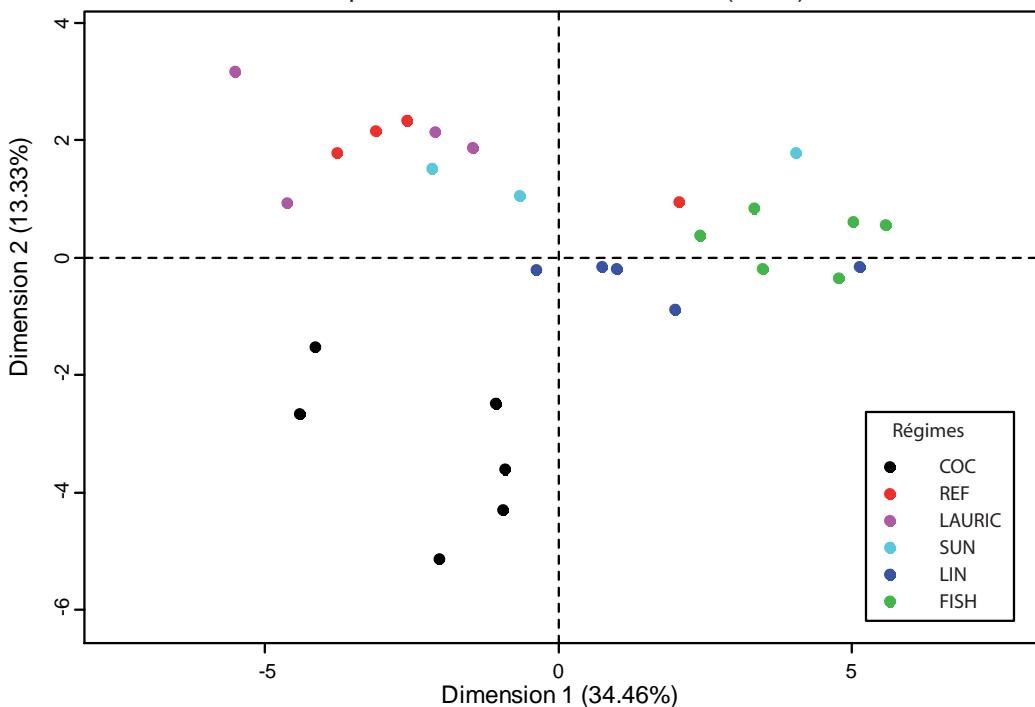
Des expériences de retard sur gel ont été réalisées avec la sonde NR1 (gccAGGTCAgaaAGTACAgatt) issue du promoteur de Cyp2b10 et radiomarquée au  $^{32}\text{P}$  à son extrémité (T4 polynucléotide kinase, Invitrogen). CAR et RXR (plasmides générueusement fournis par le Dr M. Negishi et le Pr P. Chambon) ont été produits in vitro (TNT T7 quick coupled Transcription/translation system, Promega) et utilisés pour contrôler la spécificité de la fixation à la sonde NR1 (NR\* : gtcaAGATCAggaaAGTACTggta, bases mutées en rouge, Comp : compétiteur froid) et la co-migration avec le complexe observé dans les extraits nucléaires (Extr. Nuc.) de foie de souris (Figure A). Les extraits nucléaires issus de foie de souris de type sauvage ou de type PPAR $\alpha$ -/- ont ensuite été utilisés pour évaluer l'accroissement de la fixation du complexe CAR/RXR au promoteur de Cyp2b10 suite à un traitement pas le phénobarbital (Figure B).

Les résultats présentés en [Figure 58](#) montrent que l'induction de Cyp2b10 est fortement altérée chez les souris PPAR $\alpha$ -/- par rapport aux souris de type sauvage et ce que ce soit dans le cadre d'une activation durant plusieurs jours par un ligand agoniste du récepteur CAR murin ([Figure 58A](#)) ou dans celui d'une translocation rapide du récepteur par un activateur comme le phenobarbital ([Figure 58B](#)). Il pourrait être pertinent de rapprocher ces résultats de ceux présentés dans l'article présenté au chapitre IV (Martin et al., 2005) où nous avons observé une induction de Cyp2b10 par le fenofibrate chez les souris PPAR $\alpha$ -/- uniquement. Bien qu'il soit difficile à ce stade d'envisager le mécanisme précis expliquant ces différents résultats, il semble que PPAR $\alpha$ , ou plutôt son absence, ait une influence notoire sur l'induction de Cyp2b10 par les activateurs de CAR (et peut-être par d'autres molécules également). L'une des idées que nous avons souhaité évaluer était de contrôler si un traitement à court terme par le phenobarbital provoquait bien un accroissement similaire chez les souris des deux génotypes de la fixation de CAR sur son élément de réponse NR1 localisé dans le promoteur de Cyp2b10. Les résultats présentés en [Figure 59](#) ne semblent pas montrer de différence nette de la fixation de CAR sur son élément de réponse et suggère donc qu'une altération de la fixation du dimère CAR/RXR n'est pas à l'origine de la moindre induction de Cyp2b10 par les activateurs de CAR chez les souris PPAR $\alpha$ -/. Ces résultats préliminaires, nous ont conduit à initier un projet de recherches visant à étudier les éventuelles interactions entre les voies de signalisation des récepteurs CAR et PPAR $\alpha$ .

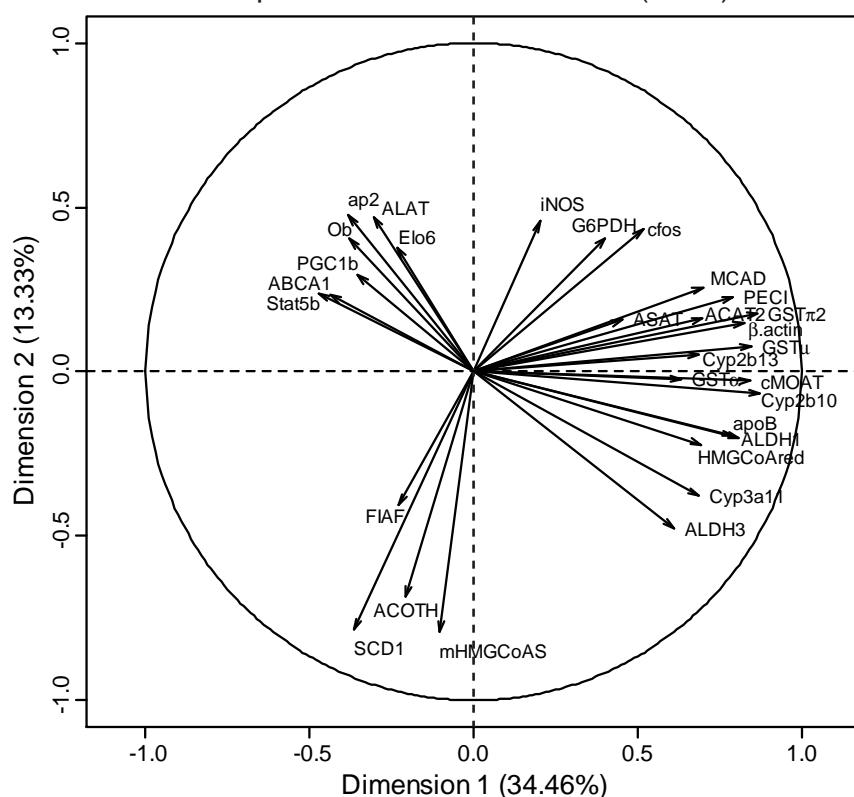
Le prélèvement de différents organes dans cette étude de nutrition (14 organes prélevés au total), nous a permis de poursuivre les investigations initiées au niveau hépatique sur d'autres organes d'intérêt. Nous avons ainsi analysé les modulations d'expression géniques provoquées par les régimes chez les souris de type sauvage au niveau de la muqueuse intestinale en utilisant l'INRAarray 01.3. Une analyse en composante principale réalisée sur une sélection de 31 gènes présentant un différentiel d'expression entre l'un des régimes et les régimes REF ou COC est présentée en [Figure 60](#). Dans le premier plan factoriel, une séparation assez nette des régimes REF, COC et FISH est mise en évidence ([Figure 60A](#)), le régime LAURIC étant globalement groupé avec le régime REF, deux des trois individus soumis au régime SUN sont également relativement proches du régime REF alors que les individus soumis au régime LIN sont globalement situés entre le régime REF et le régime FISH. Une analyse rapide de la représentation des variables ([Figure 60B](#)) montrent que trois gènes principalement (et FIAF dans une moindre mesure) sont surexprimés suite à l'exposition au régime COC (SCD1, ACOTH et mHMGCoAS) dont l'un (SCD1) est

**A.**

Représentation des échantillons (n=28)

**B.**

Représentation des variables (n=31)

**Figure 60 : ACP des profils d'expression intestinaux dans l'étude de nutrition**

Les conditions de l'expérimentation et les notations des régimes sont décrites dans PGP Martin *et al.*, Hepatology, 2007. Les données correspondant à un régime basé sur le régime de référence mais additionné de 10% de trilaurate sont également représentées (LAURIC). Seules les données correspondant aux animaux de type sauvage sont présentées ici. Les intestins des souris ont été grattés sur toute leur longueur à l'aide d'une lame de verre pour récupérer majoritairement la muqueuse intestinale. Les ARN totaux ont été extraits et analysés avec les puces INRAarray 01.3. Une liste de 31 gènes présentant au moins une régulation significative par rapport aux régimes REF ou COC (anova + test de Dunnett) a été utilisée pour réaliser l'analyse en composante principale (ACP réalisée sous R avec le package FactoMineR). Les individus (A) et les variables (B) sont représentés dans le premier plan factoriel qui explique  $34,5+13,3=47,8\%$  de la variance totale.

également surexprimé dans le foie sous ce même régime (Martin et al., 2007). Les gènes surexprimés dans le régime FISH (et dans une moindre mesure dans le régime LIN apportant également beaucoup d'acides gras de la famille  $\omega$ 3) sont plus nombreux et on note notamment la présence de plusieurs gènes impliqués dans le métabolisme des xénobiotiques (les GST $\alpha$ ,  $\mu$  et  $\pi$ 2, les Cyp2b10, 2b13 et 3a11, et les aldéhyde déshydrogénases ALDH1 et 3) et d'au moins deux gènes cibles de PPAR $\alpha$  (MCAD et PECI). A l'opposé, certains gènes comme STAT5b et Abca1 semblent moins exprimés dans le régime FISH que dans le régime de référence ([Figure 60B](#)). Ces résultats préliminaires illustrent bien le fait que même apportés à des doses couvrant juste les besoins physiologiques de l'organisme (5% de lipides dans les régimes), les acides gras alimentaires modulent l'expression de gènes dans plusieurs organes clefs impliqués dans leur propre métabolisme (dans cette étude : le foie et l'intestin).

Enfin, la disponibilité de deux jeux de données multidimensionnels acquis sur les mêmes individus (mesures des acides gras hépatiques et mesure d'expression hépatiques de gènes), nous a conduit à envisager la possibilité de croiser ces données pour éventuellement observer quels pouvaient être les liens entre les niveaux d'acides gras et l'expression des gènes. Les acides gras, notamment via les récepteurs PPAR et SREBP sont susceptibles de moduler l'expression des gènes et à l'opposé, de nombreux gènes présents sur notre puce sont impliqués dans le métabolisme des lipides (voir [Figure 54](#)). Il nous a donc semblé que les méthodes de modélisation comme la régression PLS, visant à expliquer l'un des deux jeux de données par l'autre, n'étaient pas forcément les plus appropriées dans notre contexte. Nous nous sommes donc tournés vers une analyse visant à mettre en évidence des corrélations (donc sans lien explicatif préétabli) entre ces deux jeux de variables et avons dans un premier temps évalué l'analyse canonique. Cette analyse consiste globalement à identifier des combinaisons linéaires des variables issues des deux jeux de données, fortement corrélées entre elles (cette recherche itérative aboutie à l'obtention de plusieurs couples de combinaisons linéaires indépendants entre eux, constituant les axes canoniques) et permettant ainsi de représenter les deux jeux de variables dans un même sous-espace (si les corrélations au sein des couples sont suffisamment grandes). Une mise en œuvre de l'analyse canonique sur une sélection de variables issues de l'étude présentée dans ce chapitre est illustrée dans la synthèse du paragraphe III.5. La principale limite de l'analyse canonique est le fait que la méthode nécessite un nombre d'individus supérieur au nombre de variables, ce qui rendait l'applicabilité de cette méthode plus que limitée dans le contexte de l'analyse de données issues des puces à ADN. Dans le cadre du doctorat d'Ignacio González (encadré par le Pr.

Alain Baccini), j'ai donc participé à l'évaluation sur deux expériences de biologie (dont l'une correspond à l'étude présentée dans ce chapitre) d'une version régularisée de l'analyse canonique permettant l'analyse de jeux de données comprenant plus de variables que d'individus. Les résultats de cette mise au point de méthode et de sa mise en œuvre sur des données réelles sont présentés en détail dans le manuscrit de doctorat d'Ignacio Gonzalez et l'article correspondant, actuellement soumis à une revue internationale à comité de lecture, est présenté en [annexe 3](#) de ce manuscrit.

#### V.4. CONCLUSIONS / PERSPECTIVES

Les principales conclusions et perspectives scientifiques de notre étude étant déjà largement explicitées dans la discussion de l'article présenté au paragraphe V.2., nous nous intéresserons donc ici principalement aux conclusions et perspectives méthodologiques de ce travail. Cette étude de nutrition a constitué pour nous l'occasion de vérifier que notre outil était adapté à la mise en évidence d'effets transcriptionnels d'ampleur globalement modérée mais coordonnés sur plusieurs gènes ou groupes de gènes par des composés auxquels les animaux ont été exposés à des doses relativement faibles par rapport aux études rapportées dans la littérature. Bien que très pertinente dans le domaine de la nutrition, la disponibilité d'un tel outil dans les domaines de la pharmacologie et de la toxicologie, qui constituent le cœur de métier de notre laboratoire, semble au moins autant valorisable. En effet, il est aisément de constater que de très nombreuses études expérimentales dans ces deux domaines sont réalisées à des doses qui sont largement supérieures (souvent éloignées de plusieurs ordres de grandeur) aux doses d'exposition observées chez l'homme. La mesure d'un grand nombre de variables sur des individus exposés, même à des doses faibles et en rapport avec les expositions de terrain, permet de détecter avec plus de chance des variables affectées mais plus encore, d'appréhender de manière globale des ensembles de modulations coordonnées qui, même de faible ampleur prises individuellement, peuvent, en se combinant, conduire à faire émerger un phénotype. Plusieurs exemples sont aujourd'hui disponibles dans la littérature (voir par exemple Mootha et al., 2003b) et des méthodes d'analyse dédiées à ce type de recherche, comme la méthode GSEA que nous avons évoquée au chapitre III.4.6., ont même été développées. Partant de ce constat, nous avons initié un projet de recherche visant à évaluer les effets transcriptionnels dans plusieurs organes et sous-populations cellulaires d'un phthalate : le di(2-ethylhexyl)phtalate administré à des doses faibles aux animaux. Cet assouplissant des plastiques auquel nous sommes, semble-t-il, tous exposés a été récemment

incriminé dans des problèmes de malformation génitale des nouveau-nés de sexe masculin (Swan et al., 2005) et suspecté d'un lien potentiel avec certains paramètres liés au syndrome métabolique (Stahlhut et al., 2007).

Le développement d'outils d'analyse permettant de confronter des données acquises à des niveaux d'observations différents constitue également une avancée importante de ce travail. Nous pensons que ces outils peuvent trouver des applications dans de multiples domaines, comme nous avons tenté de l'illustrer dans l'article présenté en [annexe 3](#) par le biais de la mise en œuvre de l'analyse canonique régularisé dans deux contextes bien distincts (transcriptome/lipidome dans un contexte de nutrition *in vivo* et transcriptome/activités de xénobiotiques dans un contexte de cellules cancéreuses en culture). Afin de faciliter la diffusion de cette méthode, un package pour le logiciel R a été créé et un article décrivant l'utilisation de ce package a été soumis à un journal de statistique appliquée (article non présenté ici). D'autres méthodes se développant, notamment sous l'impulsion de questionnements spécifiques comme cela a été le cas pour nous, nous envisageons que le niveau d'intégration des données et des connaissances pourrait s'accroître grandement dans les années à venir. Cela est d'autant plus probable que les technologies à haut-débit se diffusent de plus en plus rapidement au sein des laboratoires et que la constitution de plateformes technologiques offre un accès facilité aux compétences nécessaires à leur mise en œuvre.

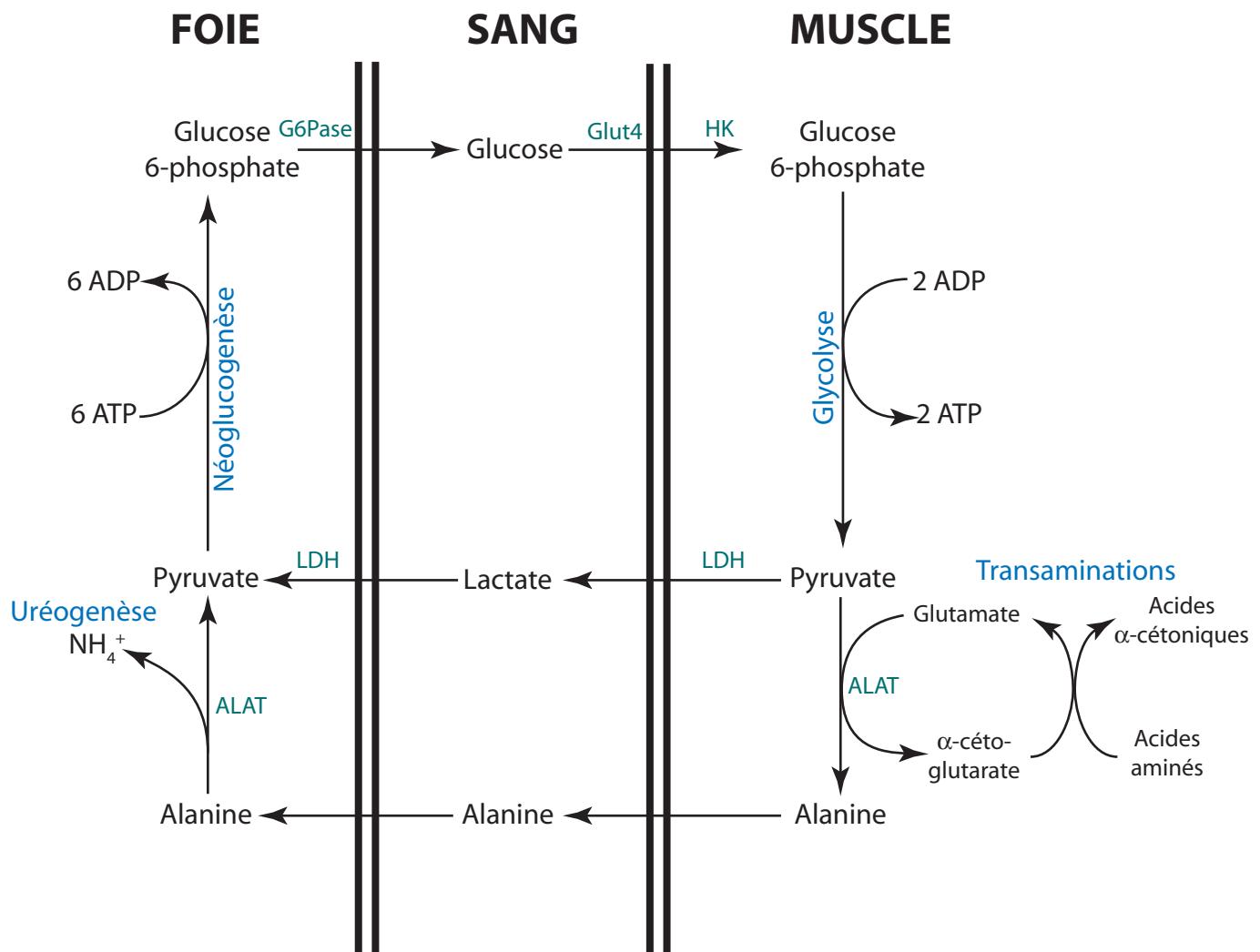


Figure 61 : Le cycle des Cori

La glycolyse musculaire fournit du pyruvate qui peut soit être réduit en lactate via l'action de la lactate déshydrogénase (LDH) en présence d'un excès de NADH (effort physique par exemple), soit transaminé en alanine sous l'action de l'alanine aminotransférase (ALAT) lorsque les acides aminés musculaires sont catabolisés (situation de jeûne prolongé). Ces deux substrats sont ensuite acheminés au foie via la circulation sanguine et utilisés comme substrats de la néoglucogenèse hépatique. G6Pase : Glucose 6-phosphatase, Glut4 : Glucose transporter 4 (régulé par l'insuline), HK : hexokinase.

# **Chapitre VI : Etude cinétique des effets du jeûne sur l'expression génique**

## **VI.1. INTRODUCTION**

Après avoir étudié l'activation pharmacologique et nutritionnelle du récepteur PPAR $\alpha$  (Chapitres IV et V), nous avons souhaité évaluer de manière plus large qu'auparavant, à l'aide de notre puce à ADN, les modulations d'expressions dépendantes et indépendantes de ce RN au cours d'une épreuve de jeûne. En parallèle, nous avons souhaité acquérir une vision générale de la cinétique des adaptations métaboliques opérées au cours du jeûne chez la souris.

Alors que l'utilisation de substrats énergétiques par l'organisme pour maintenir ses fonctions vitales est permanente, l'apport de ces substrats par l'alimentation est irrégulier. L'organisme dispose donc de systèmes de stockage de l'énergie absorbée en excès lors des prises alimentaires et de libération de cette énergie lors des phases de jeûne. Chez l'homme, la période de jeûne dépasse rarement les douze heures de jeûne nocturne dans les pays industrialisés, mais famines, disettes et jeûnes prolongés non volontaires restent malheureusement encore aujourd'hui des situations fréquentes dans certains pays. Chez l'animal, la diminution ou l'arrêt de la prise alimentaire peut être fortuite, suivant une recherche de nourriture infructueuse, ou bien associée à certains processus physiologiques (reproduction, mue ou hibernation par exemple) ou comportementaux (migration par exemple). Il existe cependant une différence fondamentale de contexte entre le jeûne naturel et expérimental dans la mesure où le premier est un phénomène perçu et généralement anticipé par un accroissement préalable des réserves énergétiques de l'organisme. Cependant, il a été montré que les mécanismes physiologiques mis en place au cours du jeûne sont très similaires dans les deux situations (Le Maho and Groscolas, 1990). Les réserves énergétiques de l'organisme sont constitués 1) de glucides, principalement sous la forme de glycogène hépatique et musculaire (environ 60 et 6mg/g d'organe respectivement chez les rongeurs) représentant moins de 1% des réserves énergétiques totales, 2) de lipides, principalement sous la forme de triglycérides stockés dans le tissu adipeux blanc qui constituent la majeure partie des réserves énergétiques de l'organisme (1 g de tissu adipeux fournissant 8 à 10 fois plus d'énergie qu'1 g de tissu glycogénique) et constituant 9 à 16% de la masse corporelle chez les rongeurs et 3) de protéines, réparties de manière hétérogène mais enrichies dans certains tissus comme les muscle (environ 40% des protéines de l'organisme), la peau (env. 20%), le

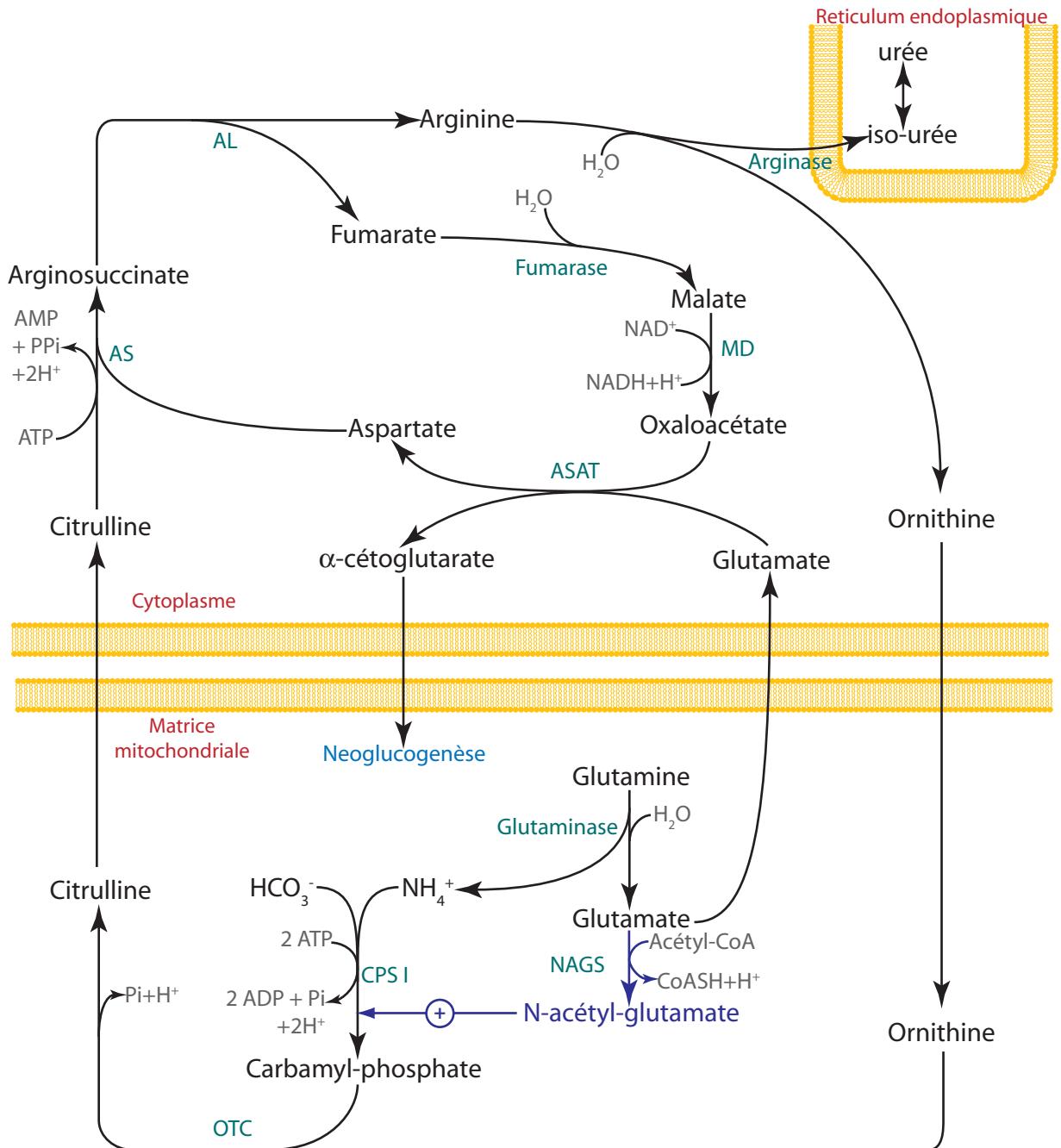


Figure 62 : Le cycle de l'urée

L'apport d'azote aux mitochondries, sous la forme d'ammoniaque, est assuré par la glutamine lors de sa conversion en glutamate sous l'action de la glutaminase. La carbamyl-phosphate synthétase I (CPSI) produit le carbamyl-phosphate à partir du NH<sub>4</sub><sup>+</sup> et du CO<sub>2</sub>. La CPSI doit obligatoirement être activée par son effecteur allostérique positif : le N-acétyl-glutamate (en bleu) produit à partir du glutamate sous l'action de la N-acétyl glutamate synthétase (NAGS). Le carbamyl-phosphate est condensé avec l'ornithine par l'ornithine transcarbamoylase (OTC) pour donner la citrulline qui sort de la mitochondrie par diffusion facilitée. L'arginine se condense avec un aspartate qui apporte le second atome d'azote de l'urée sous l'action de l'arginosuccinate synthase (AS) pour former l'arginosuccinate. Celui-ci est converti en arginine et fumarate sous l'action de l'arginosuccinate lyase (AL). L'arginase catalyse enfin l'hydrolyse de l'arginine en ornithine et iso-urée, laquelle est rapidement convertie dans le réticulum endoplasmique en urée. L'ornithine ainsi régénérée peut alors regagner la mitochondrie pour capter une nouvelle molécule de carbamyl-phosphate. Le fumarate est converti en oxaloacétate sous les actions successives de la fumarase et de la malate déshydrogénase (MD). L'aspartate aminotransférase (ASAT) transfère la fonction amine du glutamate à l'oxaloacétate pour former de l'aspartate, substrat de l'AS, et de l'α-cétoglutamate qui peut participer à la néoglucogenèse ou être aminé par la glutamate déshydrogénase et ressortir de la mitochondrie pour former de l'aspartate. Pi : phosphate inorganique, PPi : pyrophosphate.

foie et le tube digestif (env. 10% pris ensemble). Les protéines subissent en réalité un renouvellement constant dans l'organisme, rendant difficile l'évaluation de leur importance du point de vue énergétique. Mais certaines protéines, dites labiles et représentant 15 à 25% des protéines totales, peuvent néanmoins être mobilisées et servir de substrats énergétiques. Le recours massif aux protéines est néanmoins limité par les rôles fonctionnels qu'elles jouent. Naturellement, les trois formes de réserves énergétiques que nous avons évoquées ne sont pas indépendantes et des interactions s'établissent au cours du jeûne dans la gestion et l'utilisation de ces différentes sources d'énergie.

Le jeûne se traduit par une inévitable perte de masse corporelle qui suit généralement deux phases : une perte initiale rapide suivie d'un ralentissement de la perte de masse qui diminue ensuite de manière régulière. Une troisième phase caractérisée par une accélération de la perte de masse, dont l'issue est généralement fatale, peut être observée chez les modèles animaux de laboratoire lors de jeûnes très prolongés (Goodman et al., 1984). Lors de la phase initiale du jeûne, les réserves en glycogène s'épuisent rapidement conduisant à une chute de la glycémie (Hashimoto et al., 2000; Kersten et al., 1999; Leone et al., 1999; Liang et al., 2002). L'apport de glucose est alors assuré par la néoglucogenèse (principalement hépatique et dans une moindre mesure rénale, voir [Figure 33](#)) à partir des acides aminés issus de la protéolyse musculaire (alanine et glutamine principalement), du glycérrol issu de l'hydrolyse des triglycérides et du lactate issu de la glycolyse anaérobie musculaire et érythrocytaire (voir le cycle des Cori en [Figure 61](#)). L'utilisation des acides aminés pour la synthèse de glucose active conjointement l'uréogenèse hépatique (ou cycle de l'urée, [Figure 62](#)). Cependant, une épargne des protéines musculaires est mise en place par l'organisme afin d'éviter la perte de plus de 35 à 50% des protéines qui serait létale (Cherel et al., 1992). Cette épargne protéique est particulièrement précoce chez la souris et une baisse de l'urée plasmatique est observée dès les premières heures de jeûne (Kersten et al., 2001). En parallèle, les tissus non glucodépendants, largement privés de glucose, utilisent les acides gras issus de la lipolyse du tissu adipeux comme substrats énergétiques via les  $\beta$ -oxydations mitochondriale et peroxysomale et via l' $\omega$ -oxydation microsomale (voir chapitre II.4.3). Une augmentation des acides gras libres circulants est constatée en situation de jeûne (Hashimoto et al., 2000; Kersten et al., 1999; Liang et al., 2002) qui résulte de cette mobilisation active des réserves en triglycérides de l'organisme. A ce stade, l'oxydation des acides gras n'est pas encore suivie par la cétogenèse mais la biosynthèse des acides gras est déjà inhibée. Le métabolisme basal se réduit ainsi que la température corporelle qui, chez la souris peut chuter de 1,6°C après 24h de jeûne (Kersten et al., 1999). A la fin de cette première phase de l'adaptation au jeûne, la

part de l'apport énergétique due à la néoglucogenèse diminue, la protéolyse musculaire est faible ainsi que le taux circulant d'alanine, l'organisme s'est placé dans une situation d'économie d'énergie signant le début d'une nouvelle phase d'adaptation que l'on peut considérer comme celle d'un jeûne prolongé. Au cours de la deuxième phase du jeûne, l'organisme tend à stabiliser son métabolisme. Il utilise principalement l'oxydation des acides gras et les corps cétoniques résultant de la cétogenèse comme substrats énergétiques, le glucose étant réservé aux organes glucodépendants comme le cerveau (qui utilise également des corps cétoniques), la moelle épinière, les glandes surrénales, les erythrocytes et les leucocytes. Les taux circulant d'acides gras libres et de  $\beta$ -hydroxybutyrate sont élevés. L'épargne protéique se poursuit, mais ne bloque pas complètement la participation de la protéolyse à la néoglucogenèse qui se poursuit tout au long de cette phase d'adaptation au jeûne. Cette deuxième phase se poursuit tant que les réserves lipidiques ne sont pas épuisées. Lorsqu'un état critique est atteint (déplétion d'environ 80% des réserves lipidiques), l'utilisation des protéines corporelles reprend subitement, ce qui correspond à la transition vers la troisième phase de l'adaptation au jeûne. Cette phase se caractérise par la déplétion complète des lipides, l'augmentation de l'excrétion azotée et du taux plasmatique d'alanine et une nouvelle accélération de la perte de poids, conduisant rapidement à mettre en jeu le pronostique vital. Des animaux obèses semblent généralement mieux supporter une épreuve de jeûne car leurs importantes réserves lipidiques leur permettent de prolonger la deuxième phase et de préserver ainsi leurs protéines (Cherel et al., 1992).

Ces adaptations métaboliques au jeûne sont accompagnées, et sans doute largement orchestrées, par de nombreuses modulations hormonales. La chute du glucose sanguin entraîne une chute rapide de l'insulinémie et provoque la sécrétion de glucagon par le pancréas. Les niveaux d'expression et les taux plasmatiques de plusieurs adipokines diminuent également au cours du jeûne. La concentration sérique de leptine baisse ainsi dès les 12 premières heures de jeûne chez le rat (Friedman and Halaas, 1998; Kim et al., 1998a). L'expression de la résistine dans le tissu adipeux blanc diminue mais une baisse significative de sa concentration sérique n'est observée qu'au bout de 24h de jeûne chez la souris (Rajala et al., 2004). L'expression et la concentration plasmatique de l'apeline, une adipokine récemment identifiée chute également fortement après 24h de jeûne chez la souris (Boucher et al., 2005). En revanche, les niveaux circulants d'adiponectine sont peu affectés (une légère baisse est parfois rapportée) par 48h de jeûne chez les rats malgré une réduction significative de son expression dans le tissu adipeux blanc (Gui et al., 2003; Zhang et al., 2002b). Le jeûne constituant un stress indéniable pour l'organisme, les concentrations plasmatiques de

plusieurs « hormones de stress » comme les catécholamines et les glucocorticoïdes sont augmentées et participent au contrôle des phénomènes de lipolyse et de protéolyse. De manière surprenante, la sécrétion par l'hypophyse antérieure de l'hormone de croissance (GH), une hormone qui stimule notamment la croissance musculaire et osseuse, est augmentée en situation de jeûne (Thissen et al., 1994), un effet peut-être en partie lié à l'augmentation de la concentration plasmatique de ghreline (Ariyasu et al., 2002; Ariyasu et al., 2001; Tschop et al., 2000), une hormone produite principalement par l'estomac et favorisant la prise alimentaire (Nakazato et al., 2001) et la sécrétion de GH. Cependant, le taux d'IGF-I (Insulin like growth factor I) diminue pendant le jeûne (Thissen et al., 1994) alors que la GH active normalement sa sécrétion par le foie, indiquant que l'activité de la GH pourrait être diminuée en situation de jeûne. Enfin, signalons également que les taux des hormones thyroïdiennes T3 et T4, régulant en particulier le métabolisme basal, diminuent également au cours du jeûne (Maglich et al., 2004).

Le jeûne constitue ainsi une situation d'adaptation métabolique complexe au cours de laquelle de multiples signaux sont intégrés pour assurer la régulation du métabolisme, à court terme via les régulations biochimiques des enzymes du métabolisme énergétique et à plus long terme principalement via des régulations transcriptionnelles faisant intervenir une variété de facteurs de transcription. Les rôles du récepteur PPAR $\alpha$  au cours du jeûne ont été évalués dans plusieurs études (Hashimoto et al., 2000; Kersten et al., 2001; Kersten et al., 1999; Kroetz et al., 1998; Le May et al., 2000; Lee et al., 2004b; Sugden et al., 2002) qui ont notamment mis en évidence sa fonction importante de régulateur transcriptionnel de gènes du métabolisme des lipides, des glucides et des protéines et son rôle dans le contrôle de la quantité et de la qualité des lipides stockés au niveau hépatique. Néanmoins, la plupart de ces études ont été réalisées avec une ou deux durées de jeûne, ne permettant pas une réelle vue de la cinétique des modulations d'expression et des changements métaboliques opérés par l'organisme au cours de l'adaptation à la privation de nourriture. Nous avons donc conduit une expérience de jeûne que nous qualifions de postprandiale (jeûne initié 1h avant le début de la phase diurne) chez des souris de type sauvage ou de type PPAR $\alpha$ -/- et avons mesuré à 11 temps de jeûne différents, les concentrations de certains métabolites et les niveaux hépatiques d'expression d'environ 200 gènes à l'aide notre puce INRArray 01.3. Les résultats obtenus illustrent les grandes étapes de l'adaptation métabolique au jeûne chez la souris et démontrent le rôle clef de PPAR $\alpha$  dans la régulation de nombreux gènes impliqués dans ces adaptations. Elles mettent également en évidence, grâce à la réalisation d'une cinétique, le

carrefour clef de l'implication de PPAR $\alpha$  et d'autres facteurs de transcription dans l'adaptation du transcriptome hépatique à la privation de nourriture.

## VI.2. MATERIELS ET METHODES

*Animaux, procédures expérimentales et prélèvement des échantillons :*

Les souris PPAR $\alpha$ -/- (fond génétique C57BL/6J) évoluent dans l'environnement contrôlé de notre animalerie. Les animaux sont soumis à des cycles jour/nuit de 12h/12h (6h-18h) et à une température de 22 + 2°C et reçoivent un régime alimentaire (Harlan Teklad TRM T-2018-520 souris, Gannat, France) avec accès libre à l'eau et à la nourriture. Des lots de souris de type sauvage (C57BL/6J, Charles River, Les Oncins, France) et de souris PPAR $\alpha$ -/-, âgées de 8 semaines, ont été constitués de manière à avoir des masses moyennes voisines et des écart-types faibles. Pour chaque génotype, 11 lots ont été ainsi constitués comprenant 4 souris par lot, sauf pour les lots soumis la plus longue durée de jeûne (72h) et non soumis au jeûne (0h) qui contenaient 6 souris pour chaque génotype afin de pallier une éventuelle mortalité des souris et d'assurer une bonne estimation des paramètres en situation basale ( $t=0h$ ) respectivement. A 5h du matin (soit 1h avant le début de la phase diurne), les animaux ont été transférés dans de nouvelles cages propres et privés d'aliment mais en maintenant un libre accès à l'eau. Les différents lots ont subis les temps de jeûne suivants (horaires des prélèvements indiqués en italique entre parenthèses) : 0h (*5h*), 3h (*8h*), 6h (*11h*), 9h (*14h*), 12h (*17h*), 18h (*23h*), 24h (*5h*), 36h (*17h*), 48h (*5h*), 60h (*17h*), 72h (*5h*). Ayant constaté le décès d'une souris PPAR $\alpha$ -/- dans le lot soumis à 60h de jeûne, deux souris du lot soumis à 72h de jeûne ont été transférées dans ce lot. Aucun autre décès n'a été constaté à l'issue de l'épreuve de jeûne. Avant l'euthanasie des souris, un prélèvement de sang a été effectué au sinus rétro-orbitaire après une brève anesthésie (isoflurane/Forène, Abbott, Rungis, France) et une mesure de glycémie a été réalisée sur une goutte de sang total (glucomètre One Touch profile, Lifescan, France). Les souris euthanasiées par dislocation cervicale ont été immédiatement disséquées par 4 à 6 manipulateurs expérimentés afin d'assurer de coordonner les temps de prélèvements des échantillons à chaque point de la cinétique. Le foie, les reins, le cerveau, le quadriceps et les tissus adipeux blanc péri-épididymaires et inguinaux ont été prélevés, rincés et saisis à l'azote liquide et conservés à -80°C jusqu'à leur utilisation. Le foie et les territoires adipeux ont été pesés avant leur rinçage et leur congélation.

*Dosages biochimiques plasmatiques :*

Après prélèvement, les échantillons de sang total ont subi une centrifugation de 15 min à 13 000g à 4°C. Les plasmas obtenus ont été conservés à -80°C. Les mesures des teneurs plasmatiques en β-hydroxybutyrate et en urée ont été réalisées par dosage colorimétrique utilisant des réactifs commerciaux (R-Biopharm, Saint Didier au Mont D'Or, France).

*Analyse des acides gras du foie et du tissu adipeux péri-épididymaire :*

L'analyse des lipides contenus dans le foie et le tissu adipeux péri-épididymaire a été réalisée au Laboratoire de Biochimie ENSAR-INRA de Rennes (USC758) de manière conforme aux protocoles précédemment décrits (Guillou et al., 2002; Martin et al., 2007).

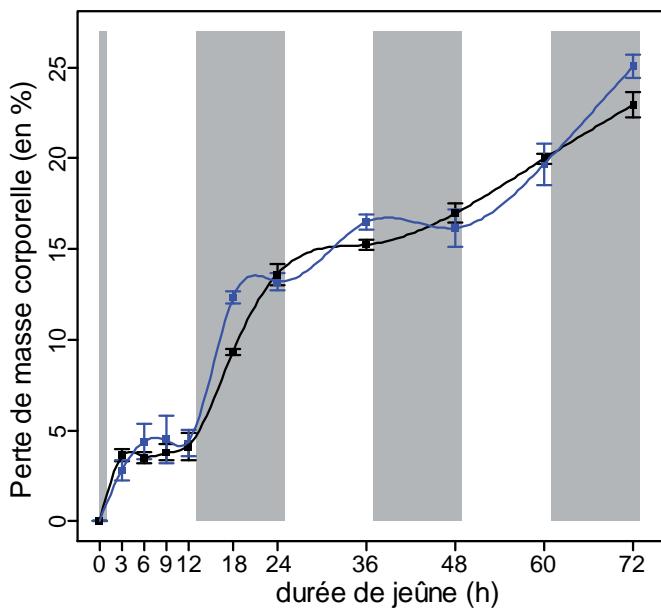
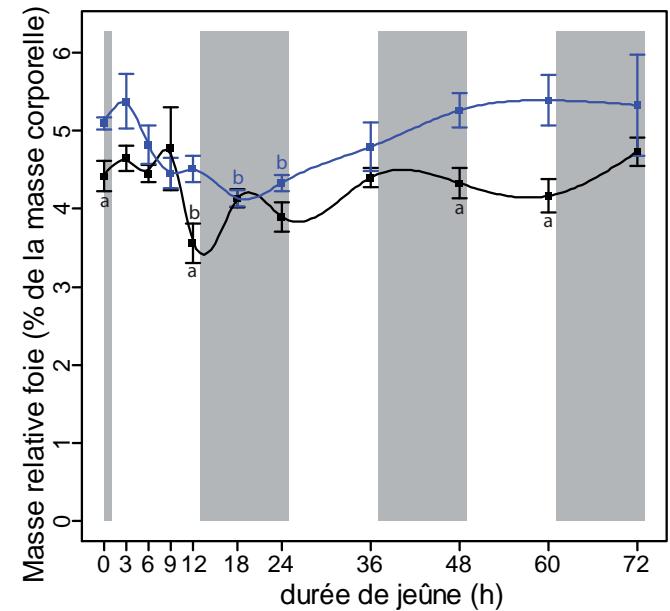
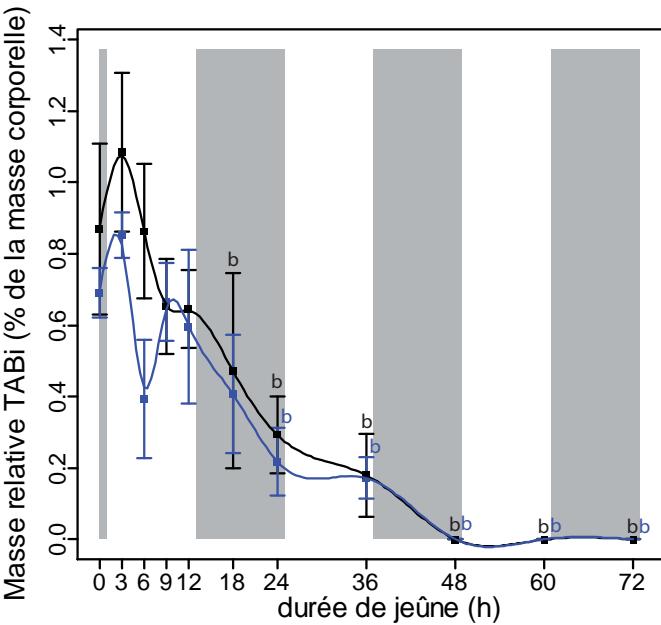
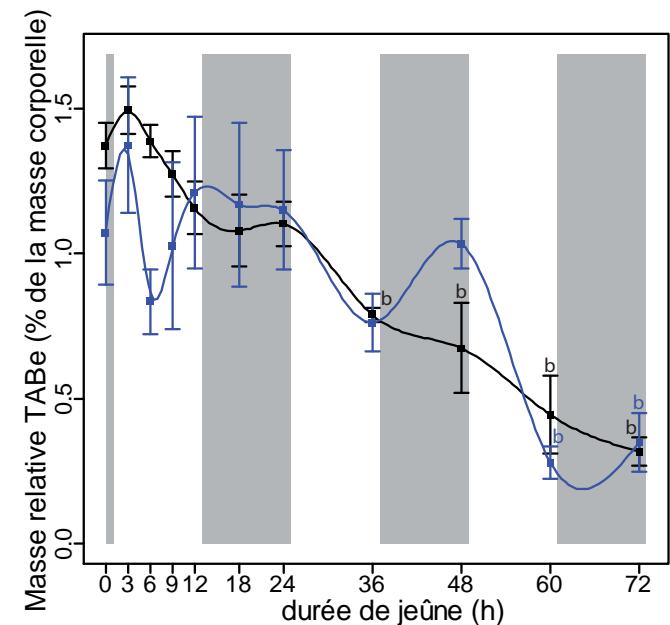
*Analyse de l'expression d'un panel de gènes hépatiques :*

Une extraction des ARN totaux a été réalisée à partir de 100 mg de foie en utilisant 1mL de Trizol (Invitrogen, Cergy Pontoise, France). Les échantillons ont été dosés au spectrophotomètre et la qualité des ARNs a été contrôlée sur un appareil Bioanalyzer 2100 (Agilent Technologies, Massy, France).

Les sondes nucléotidiques déposées sur la puce INRArray 01.3 ont été produites selon le protocole présenté au paragraphe III.3.3. et la liste des clones est disponible à l'adresse internet suivante : [www.inra.fr/internet/Centres/toulouse/pharmacologie/pharmacologie/technologie/cloneslist\\_INRArray01-3.pdf](http://www.inra.fr/internet/Centres/toulouse/pharmacologie/pharmacologie/technologie/cloneslist_INRArray01-3.pdf). Les puces à ADN ont été produites sur la plateforme « biopuces » du génopôle de Toulouse avec un spotteur VersArray ChipWriter (Virtek/Biorad, Marnes-La-Coquette, France). Le protocole de marquage des ARNs et d'hybridation des puces INRArray 01.3 est décrit en détail en annexe 1 de ce manuscrit. Quatre échantillon de chaque lot a été analysé (soit 11 temps x 2 génotypes x 4 individus /lot = 88 puces à ADN utilisées)

*Analyse statistique des résultats :*

Les données issues des puces à ADN INRArray 01.3 ont été prétraitées selon les étapes décrites au paragraphe III.4.7. Sur les 88 puces analysées, 4 puces présentaient des résultats aberrants, liés à un bruit de fond élevé (une puce pour laquelle l'ADN de sperme de saumon n'avait pas été ajouté dans le milieu d'hybridation) ou a des signaux trop faibles (3 puces dont les données n'ont pu être correctement normalisées par les spikes). Les données correspondant à ces 4 puces ont été éliminées du jeu de données. Seuls les spots dont le signal dépassait significativement le bruit de fond sur au moins 6 puces à ADN parmi les 84 puces

**A.****B.****C.****D.**

■ Souris PPAR $\alpha$ -/-
■ Souris C57BL/6J

**Figure 63 : Evolution pondérale des souris et de leurs organes au cours du jeûne**

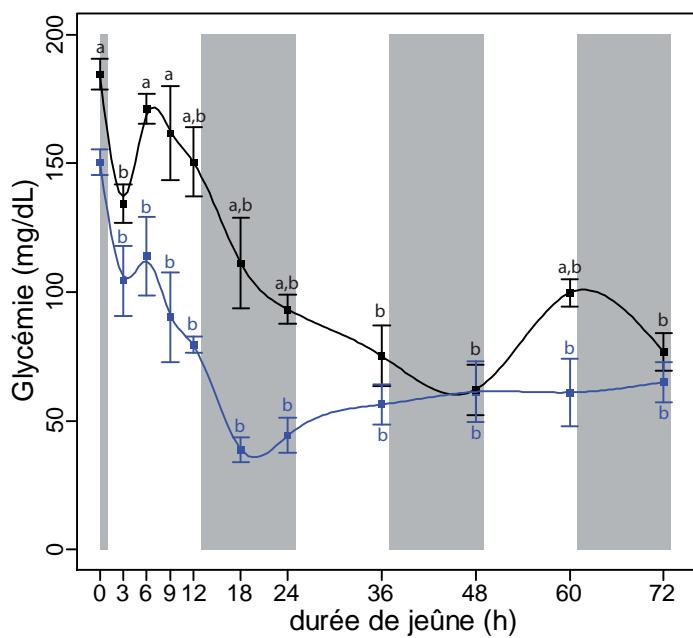
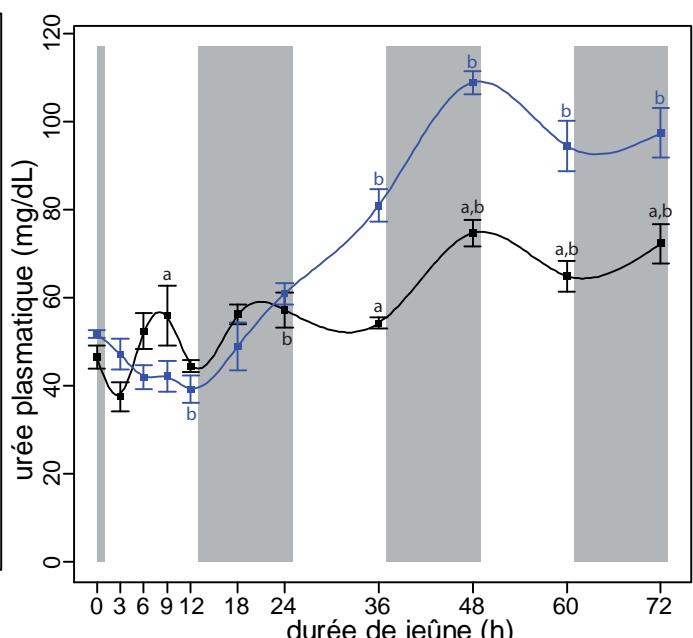
La perte de masse corporelle des souris (A), la masse relative du foie (B), la masse relative du tissu adipeux blanc inguinal (TABi, C) et la masse relative du tissu adipeux blanc péri-épididymaire (TABe, D) ont été représentées en fonction de la durée de jeûne. Les périodes nocturnes sont présentées sur chaque graphique sous la forme de rectangles gris. Des courbes de tendance obtenues via un lissage par des splines cubiques (paramètre de lissage=0,1) sont également représentées. Les souris PPAR $\alpha$ -/- sont représentées en bleu et les souris de type sauvage C57BL/6J sont représentées en noir. Les données sont représentées sous la forme moyenne +/- erreur standard de la moyenne. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis tests de Student lorsque l'effet d'un facteur est significatif). Les effets temps (<sup>b</sup>) ne sont pas représentés sur la figure A mais les pertes de poids moyennes de chaque temps (sauf t=0h) sont toutes significativement différentes de 0.

analysées ont été conservés. Cette procédure a conduit à sélectionner 130 gènes parmi les 192 mesurés sur la puce INRAarray 01.3, un résultat relativement cohérent en raison de la présence d'un nombre conséquent de gènes spécifiques du tissu adipeux blanc sur cette version de notre puce à ADN. La matrice des données analysées est donc constituée de 84 échantillons  $\times$  130 gènes. Sauf indication contraire, les données sont présentées sous la forme moyenne  $\pm$  erreur standard de la moyenne et l'analyse de la significativité des comparaisons de moyennes ont été réalisées à l'aide d'un test de Student suivant une analyse de variance (facteurs génotype, temps et genotype:temps). Pour les données de puces à ADN INRAarray 01.3 et pour les mesures de composition en acides gras, les p-values sont ajustées par la méthode de Benjamini-Hochberg (Benjamini and Hochberg, 1995) pour chaque comparaison et le taux de fausses découvertes (FDR) est fixé arbitrairement à 5%. L'ensemble des analyses et des représentations graphiques ont été réalisées sous les logiciels R et S-plus 2000 à l'aide de divers packages (En particulier multidim pour S-Plus, FactoMineR, multtest, marray, geneplotter et limma pour R).

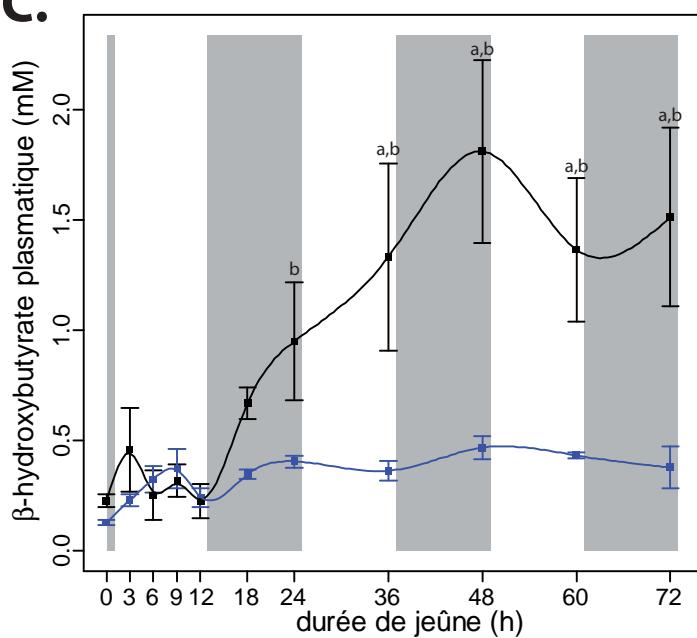
### VI.3. RESULTATS

#### *Evolution pondérale des souris et de leurs organes au cours du jeûne :*

Les souris ont été pesées la veille de leur mise à jeun puis juste avant leur sacrifice. Avant la mise à jeun, les souris PPAR $\alpha$ -/- présentaient une masse légèrement supérieure aux souris de type sauvage ( $26,8 \pm 0,3$  g contre  $24,2 \pm 0,2$  g, n=47 et 48 respectivement,  $p < 0,05$ , test bilatéral de Welch). Les pesées effectuées permettent d'évaluer la perte de poids des animaux au cours de l'épreuve de jeûne (Figure 63A). La perte de poids suit une évolution comparable entre les souris des deux génotypes (anova,  $p > 0,05$  pour les facteurs génotype et génotype:temps). Au cours des douze premières heures de jeûne, la perte de poids n'excède pas 5% de la masse corporelle initiale des animaux. Entre 12 et 18-24h de jeûne, l'amaigrissement s'accélère chez les souris des deux génotypes pour atteindre, au bout de 24h de jeûne, 13,6 et 13,2% de perte de masse corporelle pour les souris de type sauvage et pour les souris PPAR $\alpha$ -/- respectivement. Au delà de 24h de jeûne, la perte de masse corporelle est plus régulière, de l'ordre de 0,2% par heure. A l'issue des 72h de jeûne, les souris de type sauvage et de type PPAR $\alpha$ -/- ont perdu respectivement 23 et 25% de leur masse corporelle. L'évolution de la masse relative du foie (masse du foie/masse corporelle  $\times 100$ ) est présentée en Figure 63B. Avant le début de l'épreuve de jeûne, les souris PPAR $\alpha$ -/- présentent un foie

**A.****B.****C.**

■ Souris PPAR $\alpha$ -/-  
■ Souris C57BL/6J



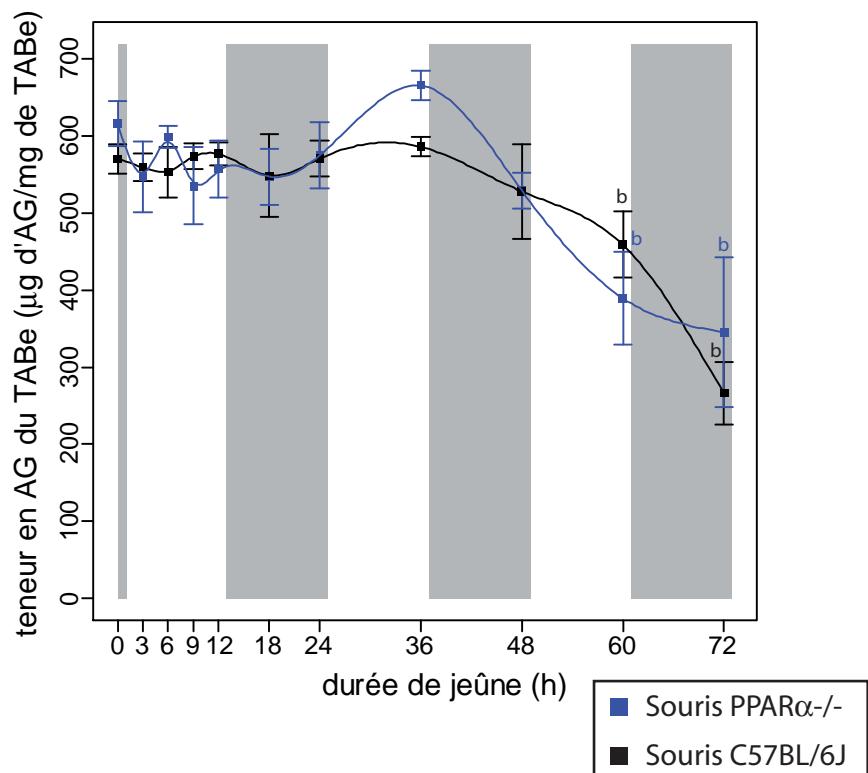
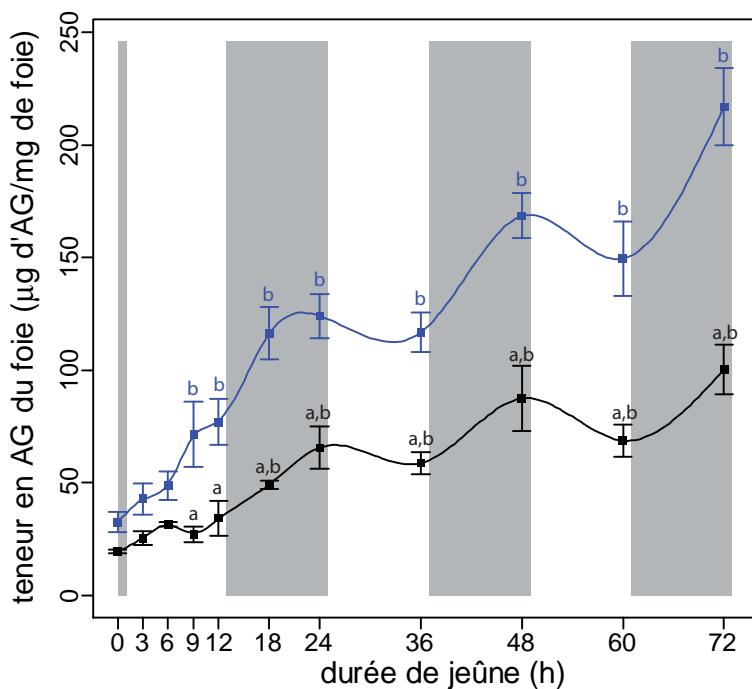
**Figure 64 : Dosages de glucose, d'urée et de  $\beta$ -hydroxybutyrate circulants**

Les concentrations de glucose dans le sang (A), d'urée (B) et de  $\beta$ -hydroxybutyrate (C) dans le plasma ont été mesurées après différentes durées de jeûne chez les souris de type sauvage (C57BL/6J) ou de type PPAR $\alpha$ -/- ( $n=4$  à 6 par points de mesure). Les périodes nocturnes sont présentées sur chaque graphique sous la forme de rectangles gris. Des courbes de tendance obtenues via un lissage par des splines cubiques (paramètre de lissage=0,1) sont également représentées. Les souris PPAR $\alpha$ -/- sont représentées en bleu et les souris de type sauvage C57BL/6J sont représentées en noir. Les données sont représentées sous la forme moyenne +/- erreur standard de la moyenne. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis tests de Student lorsque l'effet d'un facteur est significatif).

de masse relative significativement plus élevée que pour les souris de types sauvage ( $5,1 \pm 0,1$  % contre  $4,4 \pm 0,2$  %, n=6 par lot, p<0,05). Cet effet est globalement maintenu au cours du jeûne (anova, facteur génotype, p<0,05) sans qu'il y ait d'interaction entre les facteurs génotype et temps (anova, terme d'interaction, p>0,05). Au cours des premières 18 à 24h de jeûne, la masse relative de foie a tendance à diminuer chez les souris des deux génotypes. Au delà de 24h, la masse relative du foie varie de manière moins rapide. Elle se stabilise voire augmente légèrement chez les souris de type sauvage et augmente assez nettement chez les souris de type PPAR $\alpha$ -/. Pour les deux génotypes, la masse relative moyenne du foie en fin d'épreuve de jeûne (72h) n'est pas significativement différente de la masse relative moyenne du foie avant l'épreuve de jeûne (0h). Les tissus adipeux blancs péri-épididymaires (TABe) et inguinaux (TABi) ont également été pesés ([Figure 63 C et D](#)). Les souris des deux génotypes présentaient, avant le début de l'épreuve de jeûne, des masses relatives de TABe et de TABi comparables (p>0,05). L'effet du génotype sur les masses relatives de TABe et de TABi au cours du jeûne est également non significatif (anova, facteur génotype, p>0,05, terme d'interaction non significatif). Par rapport aux paramètres précédemment évoqués, on note cependant une plus grande dispersion des valeurs. Il apparaît clairement sur la [Figure 63C-D](#) que la période de jeûne se traduit par la fonte des tissus adipeux blancs, allant même jusqu'à l'épuisement du TABi pour les deux génotypes au bout d'environ 48h de jeûne.

#### *Dosages plasmatiques de glucose, d'urée et de $\beta$ -hydroxybutyrate :*

Le glucose sanguin ainsi que l'urée et le  $\beta$ -hydroxybutyrate (principal corps cétonique) plasmatiques ont été mesurés au cours de l'épreuve de jeûne ([Figure 64](#)). Pour ces trois paramètres, la durée de jeûne n'a pas le même effet chez les souris des deux génotypes (anova, terme d'interaction génotype:temps, p<0,05). Avant le début de l'épreuve de jeûne, les souris PPAR $\alpha$ -/ présentent déjà une hypoglycémie comparées aux souris de type sauvage ( $150 \pm 5$  mg/dL contre  $185 \pm 6$  mg/dL respectivement). Chez les souris PPAR $\alpha$ -/, la glycémie chute rapidement pour atteindre son plus bas niveau au bout de 18h-24h de jeûne. Au-delà de 24h, la glycémie semble se maintenir voire très légèrement augmenter. Chez les souris de type sauvage, la glycémie chute également rapidement, malgré un rebond entre 6 et 9h de jeûne, mais n'atteint son plus bas niveau, équivalent à celui atteint par les souris PPAR $\alpha$ -/, qu'au bout de 36h-48h de jeûne. On observe ensuite un rebond à 60h de jeûne avant un retour au niveau le plus bas après 72h de jeûne.

**A.****B.**

**Figure 65 : Evolution des teneurs en acides gras du TABe et du foie au cours du jeûne**  
Les teneurs en acides gras du foie (A) et du tissu adipeux blanc péri-épididymaire (TABe, B) ont été mesurées au cours de l'épreuve de jeûne. Les périodes nocturnes sont présentées sur chaque graphique sous la forme de rectangles gris. Des courbes de tendance obtenues via un lissage par des splines cubiques (paramètre de lissage=0,1) sont également représentées. Les souris PPAR $\alpha$ -/- sont représentées en bleu et les souris de type sauvage C57BL/6J sont représentées en noir. Les données sont représentées sous la forme moyenne +/- erreur standard de la moyenne. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis tests de Student lorsque l'effet d'un facteur est significatif).

L'urée plasmatique (Figure 64B) constitue globalement le reflet du catabolisme des protéines et provient du fonctionnement du cycle de l'urée au niveau hépatique (Figure 62). Chez les souris de type sauvage, la concentration plasmatique d'urée varie peu au cours des premières 48h de jeûne, malgré une légère augmentation transitoire significative au bout d'environ 9h de jeûne. Ensuite, après 48h de jeûne, l'urémie augmente de manière nette, reflétant vraisemblablement un recours plus important aux protéines pour assurer la néoglucogenèse. Il est intéressant de constater qu'il semble également y avoir une composante circadienne qui se superpose à l'évolution globale de l'urémie au cours du jeûne avec une urémie qui tend à baisser en phase diurne et à augmenter en phase nocturne. Chez les souris PPAR $\alpha$ -/-, on constate une baisse initiale de l'urémie qui n'est cependant significative qu'un bout de 12h de jeûne. Au-delà de 12h de jeûne, le taux d'urée dans le sang augmente nettement chez ces souris et dépasse significativement le taux initial ( $t=0h$ ) au bout de 36h de jeûne, une durée à partir de laquelle il reste très élevé par rapport aux souris de type sauvage.

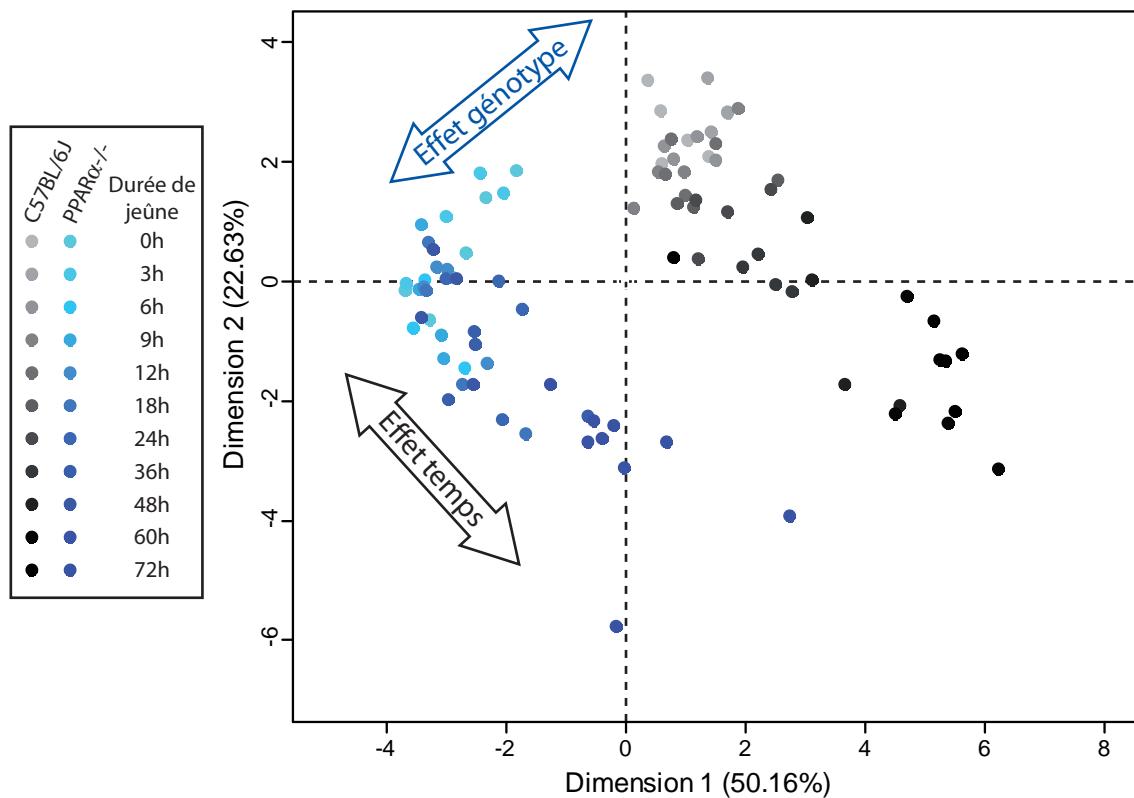
Le  $\beta$ -hydroxybutyrate (Figure 64C) est issu de la cétogenèse (Figure 28) qui est alimentée en acétyl-CoA par l'oxydation des acides gras et dont l'étape limitante est catalysée par l'HMG-CoA synthase mitochondriale. Alors que la concentration plasmatique en  $\beta$ -hydroxybutyrate est constante chez les souris PPAR $\alpha$ -/-, elle augmente nettement chez les souris de type sauvage à partir de 12h de jeûne pour être significativement supérieure à la concentration initiale ( $t=0h$ ) à partir de 24h de jeûne.

#### *Evolution du contenu en acides gras du foie et du TABe au cours du jeûne :*

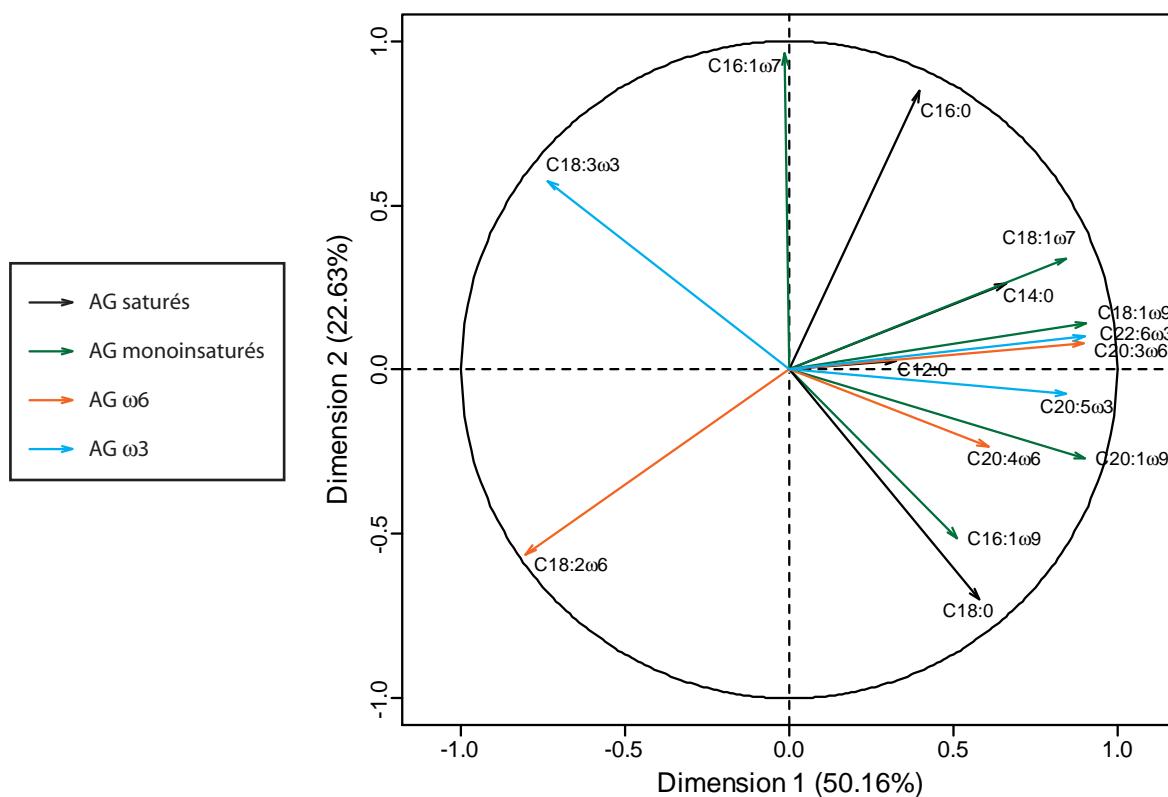
Au cours du jeûne, les lipides sont mobilisés du tissu adipeux blanc et vont alimenter divers tissus et en particulier le foie qui, de plus, est capable de transformer ces acides gras en corps cétoniques utilisables comme source énergétique par divers tissus tels le cerveau ou le cœur. Nous avons donc mesuré la teneur en acides gras de deux tissus, le TABe et le foie au cours de l'épreuve de jeûne (Figure 65). Les souris des deux génotypes présentent, en début d'épreuve de jeûne, comme au cours de cette épreuve, des teneurs en acides gras du TABe comparables (anova, interaction génotype:temps et facteur génotype non significatifs). La fonte du tissu adipeux (Figure 63) ne se traduit pas par un appauvrissement significatif du TABe en acides gras jusqu'à la 48<sup>ème</sup> heure de jeûne. Entre 48 et 72h de jeûne, le tissu s'appauvrit cependant nettement en acides gras dont la teneur tissulaire est alors presque divisée par deux. Au niveau hépatique, les acides gras s'accumulent nettement au cours du jeûne chez les souris des deux génotypes mais cette accumulation est à la fois plus précoce

**A.**

Représentation des échantillons (n=93)

**B.**

Représentation des variables (n=15)

**Figure 66 : ACP de la composition en acides gras du TABe**

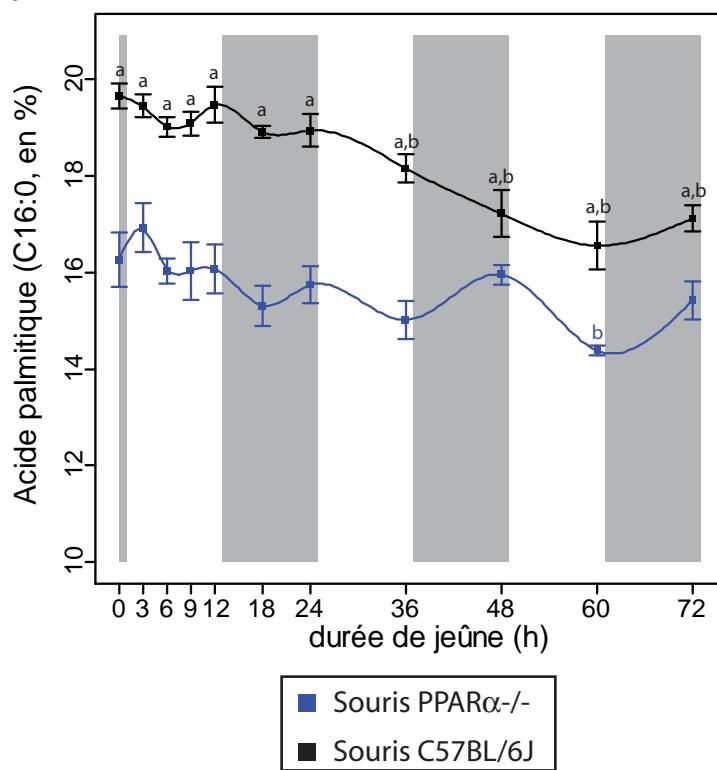
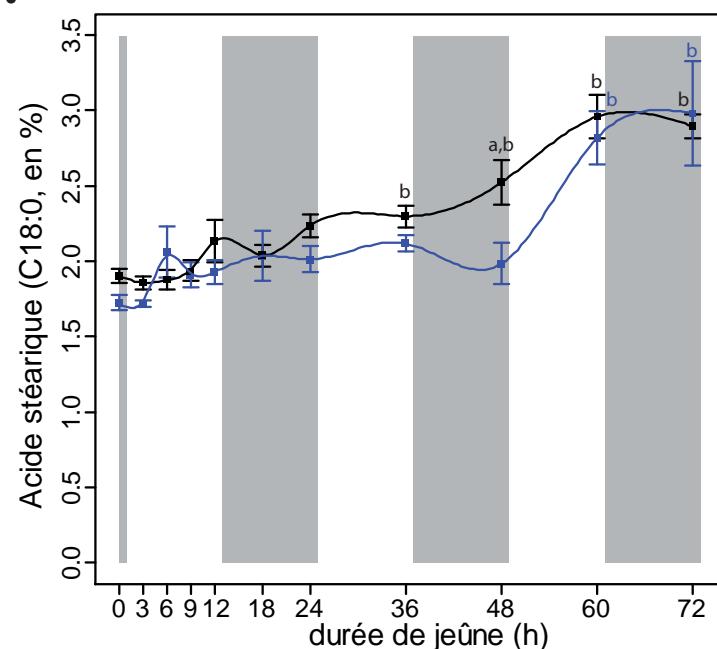
L'abondance relative de 15 acides gras a été mesurée par chromatographie en phase gazeuse dans les tissus adipeux blancs péri-épididymaires (TABe) des souris de type sauvage et de type *PPAR $\alpha$ -/-* au cours de l'épreuve de jeûne. Les pourcentages des 15 acides gras pour 93 souris ont été analysés par analyse en composante principale (ACP) à l'aide du package FactoMineR de R. Les individus (A) et les variables (B) sont représentés dans le premier plan factoriel qui explique  $50,2+22,6=72,8\%$  de la variance totale. Sur la figure A, des flèches doubles illustrant les effets génotype et temps ont été placées manuellement sur le graphique.

(une différence significative par rapport au temps t=0h apparaît dès la 9<sup>ème</sup> heure de jeûne chez les souris PPAR $\alpha$ -/-) et plus marquée (une différence significative entre les deux génotypes est constatée dès 9h de jeûne) chez les souris PPAR $\alpha$ -/- par rapport aux souris de type sauvage. Là encore, il est intéressant de constater que des variations circadiennes sont clairement superposées aux modulations de la teneur hépatique en acides gras liées au jeûne ([Figure 65B](#)). A partir de 24h de jeûne, les acides gras ne s'accumulent au niveau hépatique que lors des phases nocturnes et ce phénomène est observé chez les souris des deux génotypes.

#### *Evolution des profils d'acides gras du TABe et du foie au cours du jeûne :*

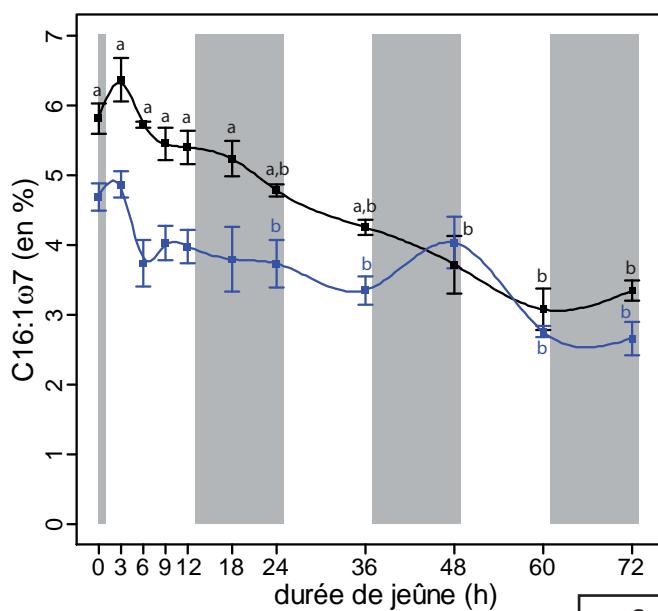
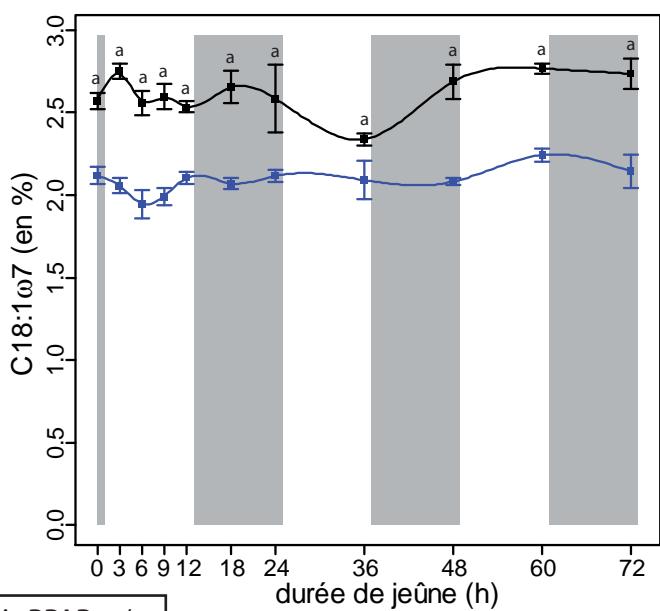
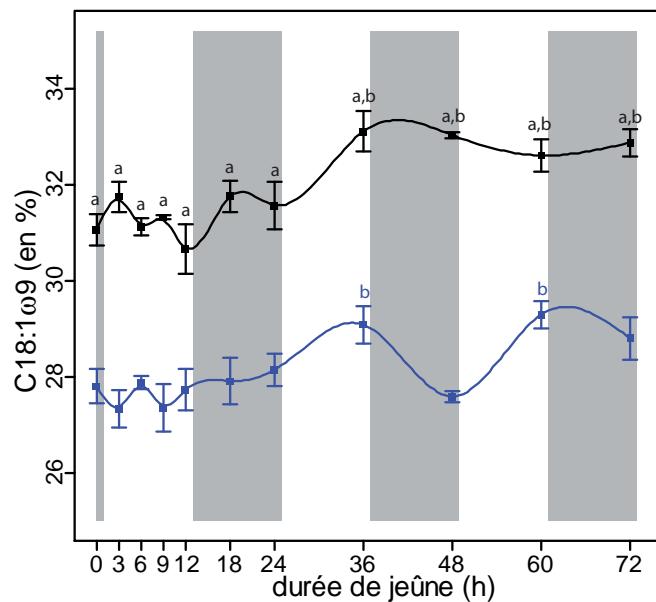
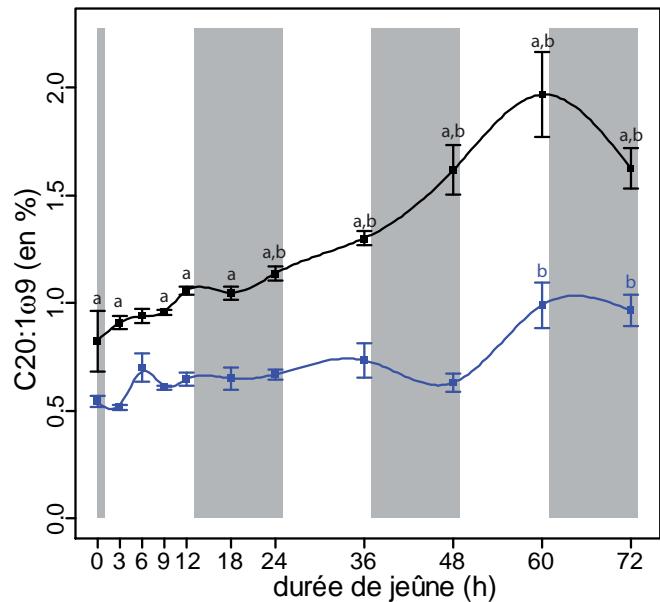
Afin d'évaluer si les différents acides gras, dont les capacités d'activation du récepteur PPAR $\alpha$  et les fonctions biologiques de manière générale peuvent être assez différentes, présentent des modulations distinctes au cours du jeûne, nous avons mesuré par chromatographie en phase gazeuse l'abondance relative de 15 acides gras dans le TABe et le foie.

Les données pour le TABe (93 échantillons  $\times$  15 acides gras exprimés en % des acides gras totaux) ont été dans un premier temps étudiées de manière globale par une analyse en composante principale ([Figure 66](#)). L'interprétation de cette ACP étant facilitée par l'observation conjointe d'un certain nombre de profils individuels d'AG, les profils de deux AG saturés ([Figure 67](#)), de quatre AG monoinsaturés ([Figure 68](#)) et de quatre AG polyinsaturés des familles  $\omega$ 3 et  $\omega$ 6 ([Figure 69](#)) sont également présentées. Les profils des acides gras C12:0, C14:0, C16:1 $\omega$ 9 et C20:4 $\omega$ 6 ne sont pas représentés car ces AG sont moins bien représentés dans le premier plan principal (le cosinus carré des angles formés par ces vecteurs et le premier plan principal est inférieur à 0,6 ce qui peut être visualisé en comparant la longueur des vecteurs correspondants à ces quatre AG aux autres vecteurs sur la [Figure 66B](#)). Deux effets apparaissent nettement sur le premier plan principal : l'effet génotype et l'effet temps ([Figure 66A](#)). Les différentiels les plus marqués de proportions d'acide gras entre les deux génotypes concernent : 1) l'acide palmitique (C16:0, [Figure 67A](#)) et plusieurs acides gras monoinsaturés (principalement les C16:1 $\omega$ 7, C18:1 $\omega$ 7 et C18:1 $\omega$ 9, [Figure 68](#)) qui sont préférentiellement accumulés par les souris de type sauvage, et 2) l'acide linoléique (C18:2 $\omega$ 6, [Figure 69A](#)) qui est préférentiellement accumulé chez les souris PPAR $\alpha$ -/-, vraisemblablement au dépend d'acides gras plus longs et plus insaturés comme l'illustre la [Figure 69B](#). Une tendance comparable est observée avec l'acide  $\alpha$ -linolénique (qui tend à

**A.****B.**

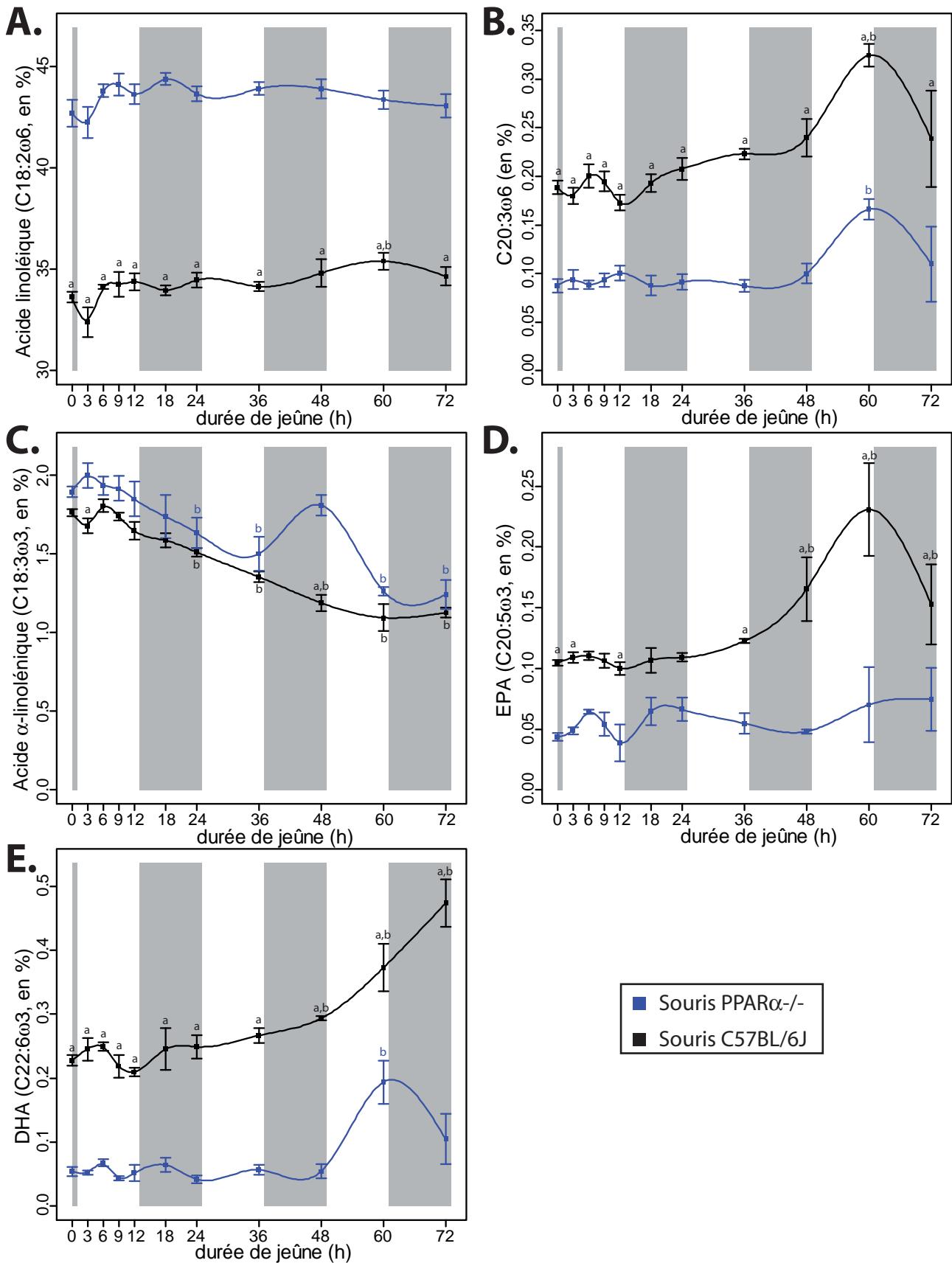
**Figure 67 :** Profils des proportions d'acides palmitique et stéarique dans le TABe

Les évolutions au cours du jeûne des proportions (exprimées en % des acides gras totaux) d'acide palmitique (C16:0, A) et d'acide stéarique (C18:0, B) du tissu adipeux blanc péri-épididymaire (TABe) sont présentées. Les périodes nocturnes sont représentées par des rectangles gris. Des courbes de tendance obtenues via un lissage par des splines cubiques (paramètre de lissage=0,1) sont également représentées. Les données sont représentées sous la forme moyenne +/- erreur standard de la moyenne. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).

**A.****B.****C.****D.**

**Figure 68 : Profils des proportions d'acides gras monoinsaturés dans le TABe**

Les évolutions au cours du jeûne des proportions (exprimées en % des acides gras totaux) de C16:1ω7 (A), de C18:1ω7 (B), de C18:1ω9 (C) et de C20:1ω9 (D) du tissu adipeux blanc péri-épididymaire (TABe) sont présentées. Les périodes nocturnes sont représentées par des rectangles gris. Des courbes de tendance obtenues via un lissage par des splines cubiques (paramètre de lissage=0,1) sont également représentées. Les données sont représentées sous la forme moyenne +/- erreur standard de la moyenne. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).

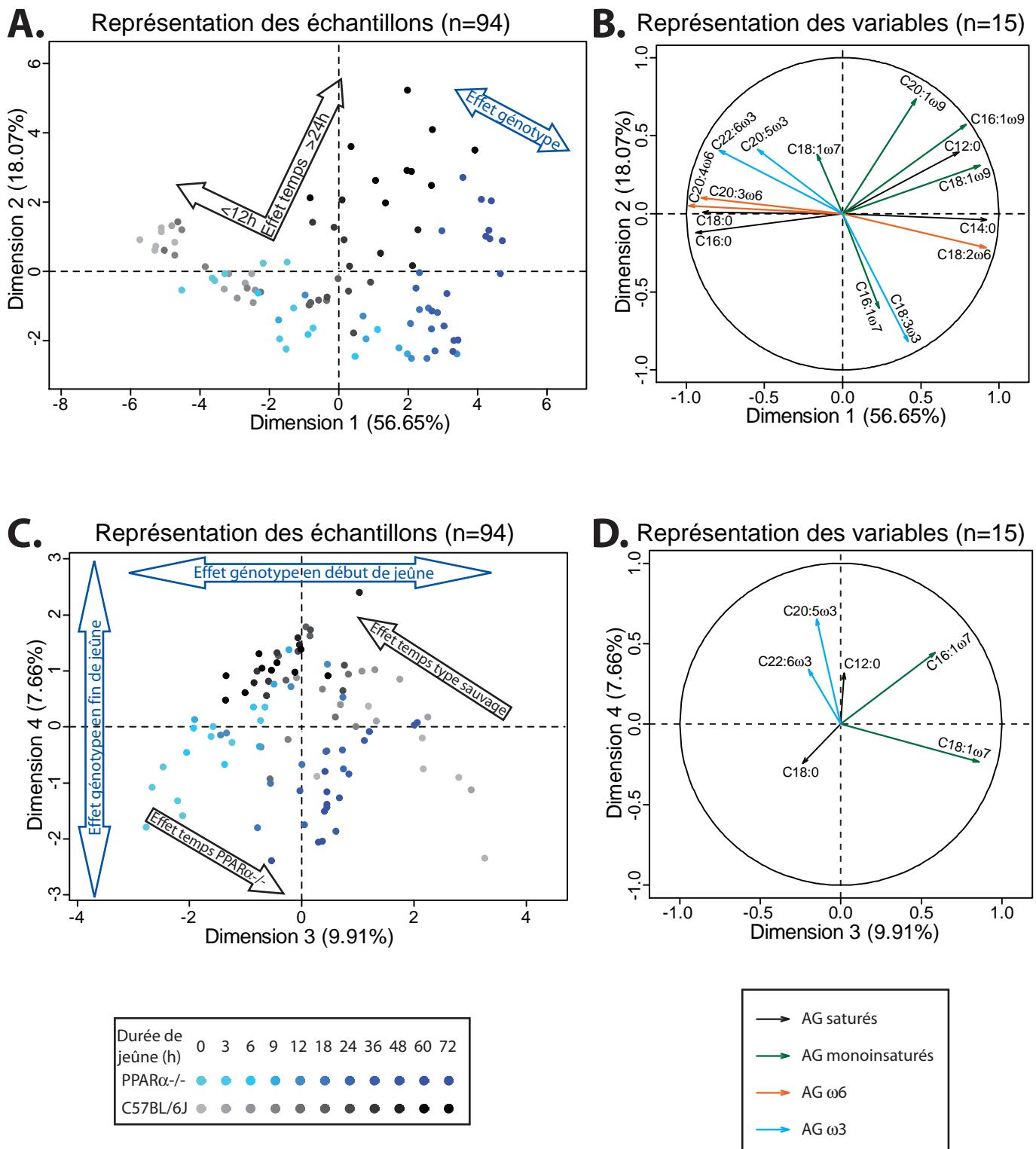


**Figure 69 : Profils des proportions d'acides gras polyinsaturés dans le TABe**

Les évolutions au cours du jeûne des proportions (exprimées en % des acides gras totaux) des acides linoléique (C18:2 $\omega$ 6, A), di-homo- $\gamma$ -linolénique (C20:3 $\omega$ 6, B),  $\alpha$ -linolénique (C18:3 $\omega$ 3, C), eicosapentaénoïque (C20:5 $\omega$ 3, D) et docosahexaénoïque (C22:6 $\omega$ 3, E) du tissu adipeux blanc péri-épididymaire (TABe) sont présentées. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).

s'accumuler chez les souris PPAR $\alpha$ -/- par rapport aux souris de type sauvage) et ses produits d'elongation/désaturation que sont l'EPA et le DHA ([Figure 69D-E](#), qui sont plus fortement accumulés par les souris de type sauvage). De manière générale, l'effet temps semble très comparable entre les deux génotypes ([Figure 66A](#)), ce qui est largement confirmé par les anova réalisées sur chaque acide gras (le terme d'interaction génotype:temps n'étant significatif au seuil de 5% que pour 6 acides gras sur les 15 et au seuil de 1% pour 3 acides gras sur les 15) et sur les graphiques présentés en [Figures 67, 68 et 69](#). Jusqu'à 18h de jeûne, on ne note aucune modulation significative des proportions d'AG dans le TABe des souris des deux génotypes. Les animaux ayant déjà perdu environ 10% de leur masse corporelle à ce temps de jeûne et le TABe ayant participé en proportion à cette perte de masse ([Figure 63A et E](#)), cela signifie que les acides gras n'ont pas été mobilisés de manière préférentielle à ce stade du jeûne. A partir de 24h quelques variations de la composition en acides gras du TABe sont observées chez les souris des deux génotypes : 1) la proportion d'acide palmitique (C16:0, [Figure 67A](#)) tend à diminuer alors que celle de l'acide stéarique augmente (C18:0, [Figure 67B](#)), 2) pour les AG monoinsaturés, on observe une baisse de la proportion de C16:1 $\omega$ 7 et une légère augmentation des C18:1 $\omega$ 9 et C20:1 $\omega$ 9 ([Figure 68](#)), enfin 3) pour les polyinsaturés, on observe une diminution de la proportion d'acide  $\alpha$ -linolénique (C18:3 $\omega$ 3, [Figure 69C](#)) et une augmentation des acides gras longs polyinsaturés de la famille  $\omega$ 3 (DHA et EPA, [Figure 69D et E](#)). Il est intéressant de constater que les modifications de la composition en acides gras du TABe semblent largement coïncider avec une participation accrue de ce tissu à la perte de masse corporelle ([Figure 63D](#)) ce qui pourrait correspondre à la mise en place de mécanismes de lipolyse affectant de manière différentielle les différents acides gras du TABe.

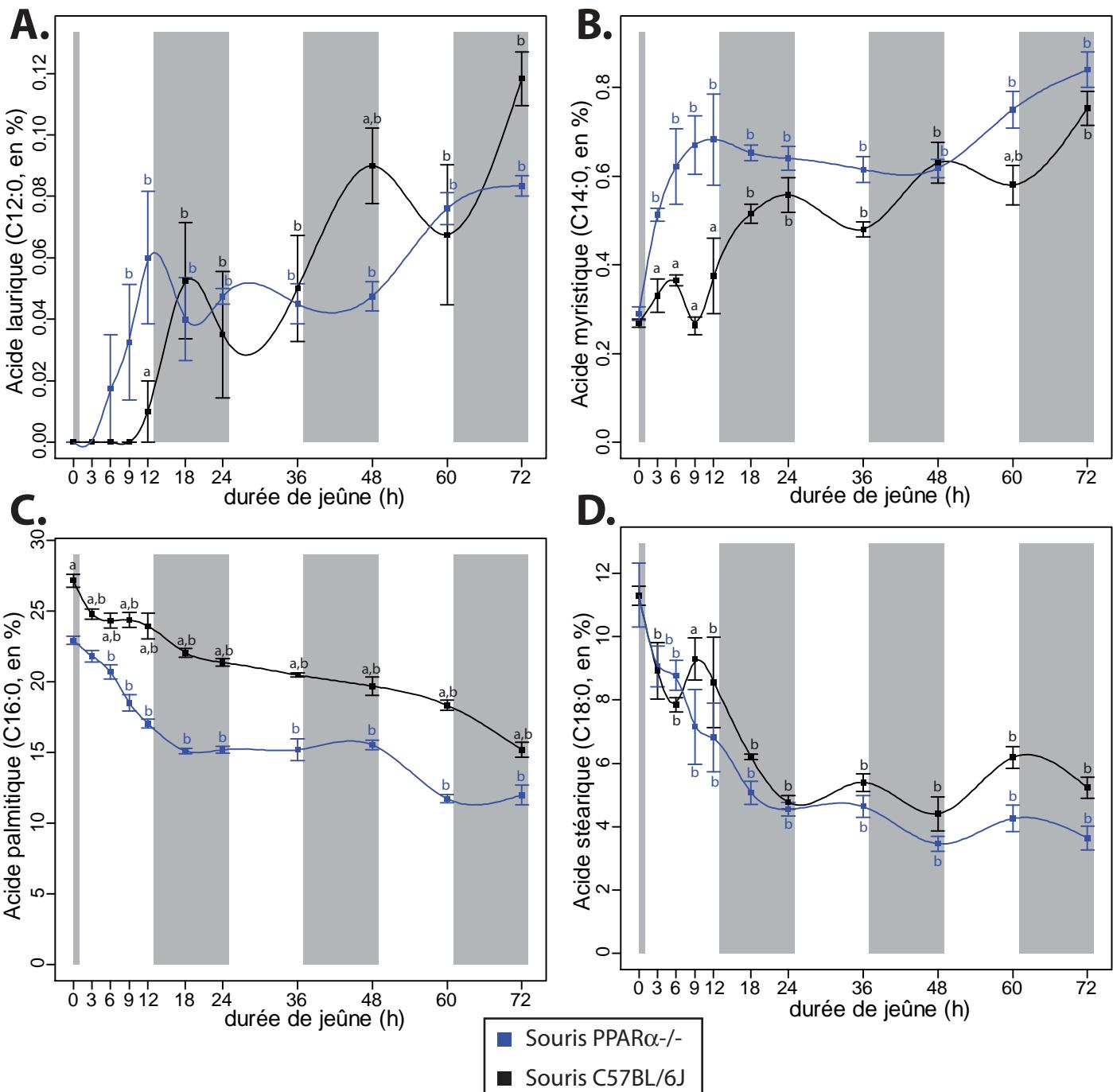
De la même manière, nous avons étudié la composition en acides gras du foie au cours de l'épreuve de jeûne. Une ACP des pourcentages d'acides gras du foie est présentée en [Figure 70](#). Contrairement à l'ACP réalisée sur la composition en AG du TABe où les axes 3 et 4 étaient générés par la présence d'échantillons présentant des valeurs extrêmes pour certains acides gras, l'ACP réalisée sur la composition hépatique en AG fournit des informations biologiquement pertinentes et interprétables jusqu'à l'axe 4 (voir [Figure 70C-D](#) pour le plan généré par les axes 3 et 4). Pour faciliter l'interprétation de cette ACP, les profils individuels des proportions hépatiques des acides gras saturés ([Figure 71](#)), monoinsaturés ([Figure 72](#)) et polyinsaturés des familles  $\omega$ 3 et  $\omega$ 6 ([Figure 73](#)) sont également présentées. Dans le premier plan principal ([Figure 70A](#)), on constate que les individus des deux génotypes se séparent



**Figure 70 : ACP de la composition hépatique en acides gras**

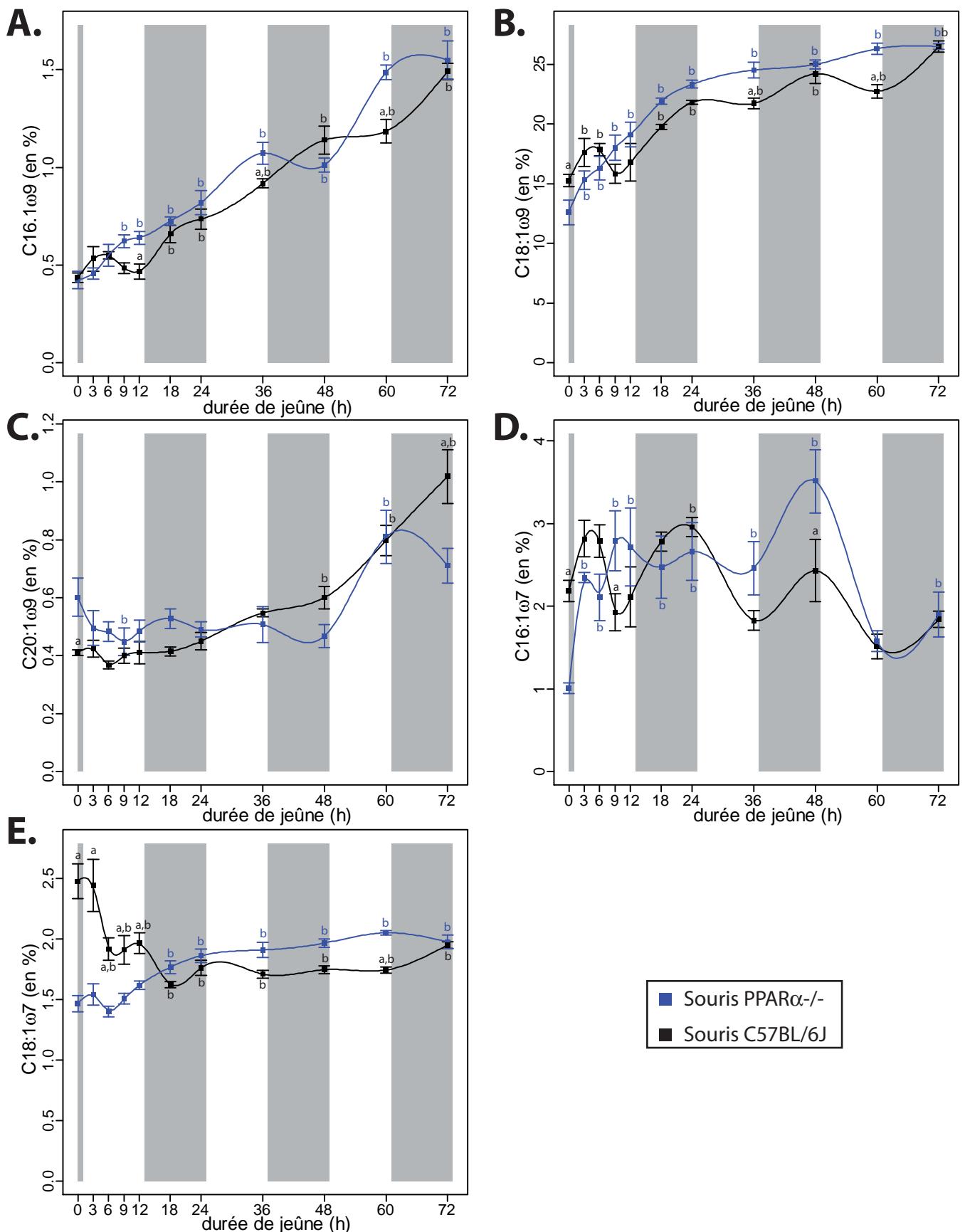
L'abondance relative de 15 acides gras a été mesurée par chromatographie en phase gazeuse dans le foie des souris de type sauvage et de type PPAR $\alpha$ -/- au cours de l'épreuve de jeûne. Les pourcentages des 15 acides gras pour 94 souris ont été analysés par analyse en composante principale (ACP) à l'aide du package FactoMineR de R. Les individus (A et C) et les variables (B et D) sont représentés dans le premier plan factoriel (A et B) qui explique  $56.6+18.1=74.7\%$  de la variance totale et dans le plan défini par les composantes principales 3 et 4 (C et D) qui explique  $9.9+7.7=17.6\%$  de variance supplémentaire (soit au total 92.3% de variance expliquée pour la représentation prise dans son ensemble). Sur les figures A et C, des annotations manuelles ont été rajoutées sur les graphiques pour en faciliter l'interprétation.

selon un axe légèrement diagonal qui apparaît un peu plus corrélé à l'axe 1 de l'ACP qu'à l'axe 2. Cette séparation des deux génotypes dans ce premier plan principal ne semble pas dépendre de la durée de jeûne car la distance et l'orientation de la séparation entre les deux nuages de points correspondant aux deux génotypes sont à peu près similaires aux faibles et aux fortes durées de jeûne ([Figure 70A](#)). La représentation des variables (AG) dans le premier plan principal ([Figure 70B](#)) permet de repérer les AG dont les proportions permettent d'expliquer majoritairement cette séparation des génotypes. Les acides linoléique (C18:2 $\omega$ 6) et arachidonique (C20:4 $\omega$ 6) sont les deux plus corrélés à l'axe 1 de l'ACP (positivement pour le linoléate et négativement pour l'arachidonate). De fait, le terme d'interaction genotype:temps de l'anova pour ces deux acides gras est non significatif et les effets génotypes et temps sont donc purement additifs expliquant en partie pourquoi la séparation des deux nuages de point correspondant aux deux génotypes reste à peu près identique au cours du jeûne. Les souris PPAR $\alpha$ -/- accumulent fortement l'acide linoléique dans leur foie (C18:2 $\omega$ 6, [Figure 73A](#)) et présentent une proportion hépatique en acide arachidonique (C20:4 $\omega$ 6, [Figure 73C](#)) significativement plus faible que les souris de type sauvage. D'autres acides gras participent également à cette séparation des deux génotypes comme l'acide  $\alpha$ -linolénique (C18:3 $\omega$ 3, [Figure 73D](#)) qui s'accumule préférentiellement dans le foie des souris PPAR $\alpha$ -/- et est vraisemblablement largement responsable de la séparation partielle des deux génotypes le long de l'axe 2 de l'ACP, ainsi que les acides palmitique (C16:0, [Figure 71C](#)) et di-homo- $\gamma$ -linolénique (C20:3 $\omega$ 6, [Figure 73B](#)) qui s'accumulent préférentiellement dans le foie des souris de type sauvage par rapport aux souris PPAR $\alpha$ -/. Ces données sont en accord étroit avec les observations que nous avions réalisées précédemment lors de l'étude de nutrition (Chapitre V et Martin et al., 2007). De plus, sur le plan défini par les axes 3 et 4 ([Figure 70C](#)), on note une séparation des deux génotypes aux faibles durées de jeûne (durée inférieure ou égale à 12h de jeûne, qualifiée de « début de jeûne » sur la [Figure 70C](#)) le long de l'axe 3 et une séparation des deux génotypes aux durées de jeûne supérieures ou égale à 18h (qualifiées de « fin de jeûne sur la [Figure 70C](#)») le long de l'axe 2. Ce plan de l'ACP met donc en exergue des acides gras qui présentent des comportements différents entre les deux génotypes au cours du jeûne. Le C18:1 $\omega$ 7 ([Figure 72E](#)) est un bon exemple des phénomènes mis en exergue par ce plan de l'ACP. De proportion plus élevée chez les souris de type sauvage en début de jeûne, son pourcentage diminue chez ces souris au cours du jeûne, alors qu'il augmente chez les souris PPAR $\alpha$ -/- conduisant les courbes à se rejoindre, voire à se croiser. Cet AG sépare donc bien les deux génotypes en début de jeûne mais pas au delà de la douzième heure. A l'inverse,



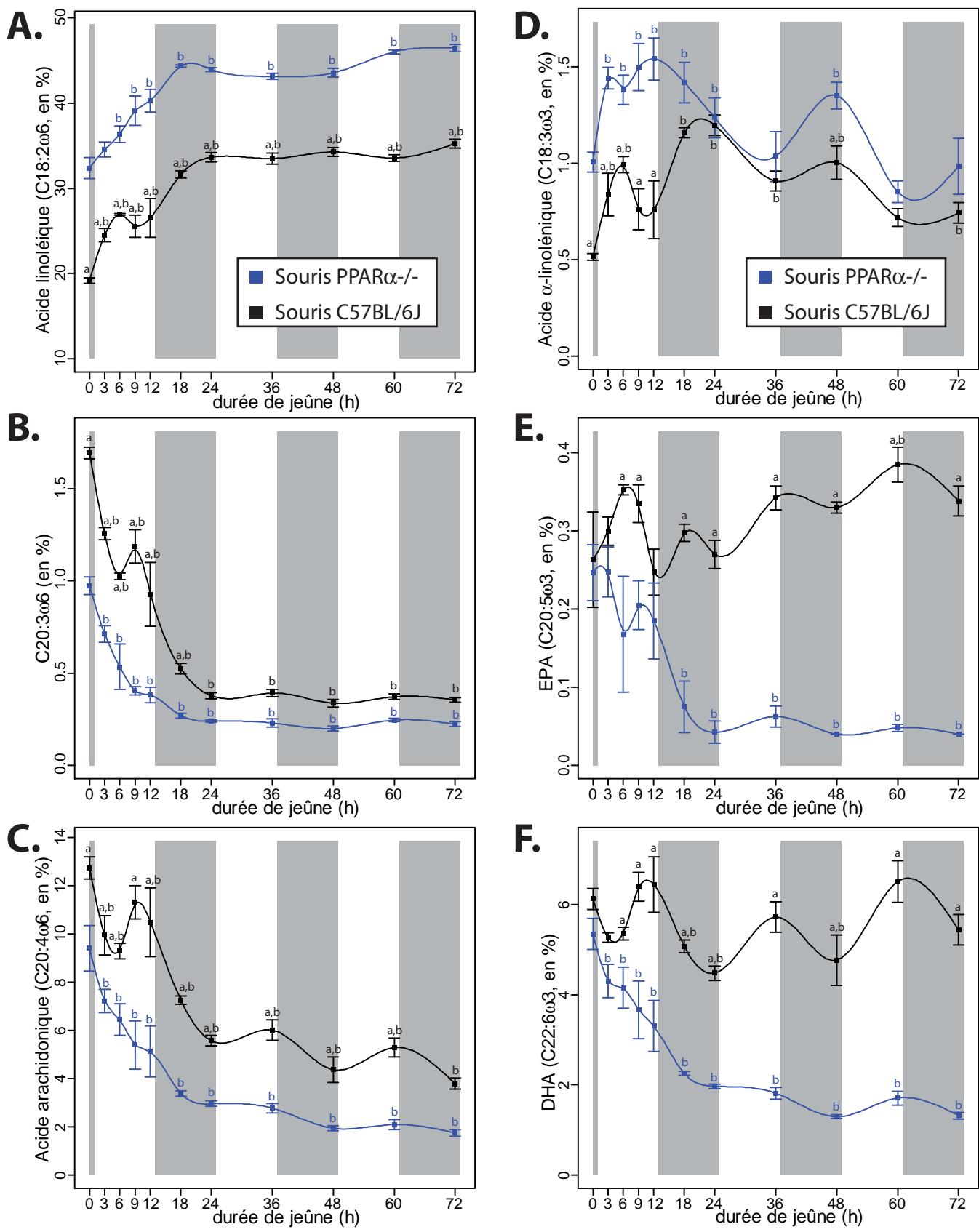
**Figure 71 : Profils des proportions d'acides gras saturés dans le foie**

Les évolutions au cours du jeûne des proportions hépatiques (exprimées en % des acides gras totaux) des acides laurique (C12:0, A), myristique (C14:0, B), palmitique (C16:0, C) et stéarique (C18:0, D) sont présentées. Les périodes nocturnes sont représentées par des rectangles gris. Des courbes de tendance obtenues via un lissage par des splines cubiques (paramètre de lissage=0,1) sont également représentées. Les données sont représentées sous la forme moyenne +/- erreur standard de la moyenne. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).



**Figure 72 : Profils des proportions d'acides gras monoinsaturés dans le Foie**

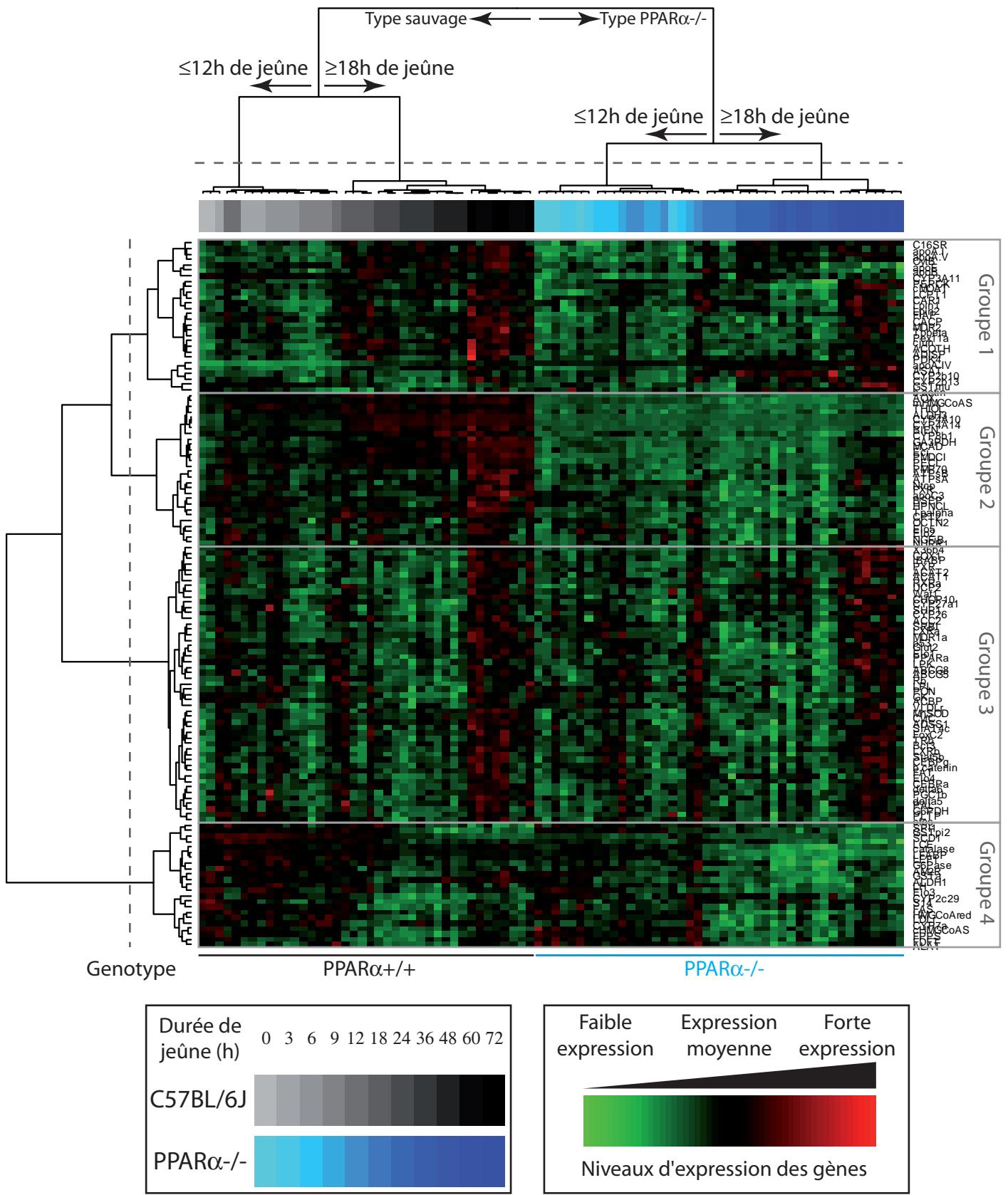
Les évolutions au cours du jeûne des proportions hépatiques (en % des acides gras totaux) de C16:1 $\omega$ 9 (A), de C18:1 $\omega$ 9 (B), de C20:1 $\omega$ 9 (C), de C16:1 $\omega$ 7 (D) et de C18:1 $\omega$ 7 (E) sont présentées. Les périodes nocturnes sont représentées par des rectangles gris. Des courbes de tendance obtenues via un lissage par des splines cubiques (paramètre de lissage=0,1) sont également représentées. Les données sont représentées sous la forme moyenne +/- erreur standard de la moyenne. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).



**Figure 73 : Profils des proportions d'acides gras polyinsaturés dans le foie**

Les évolutions au cours du jeûne des proportions hépatiques (en % des acides gras totaux) des acides linoléique (C18:2 $\omega$ 6, A), di-homo- $\gamma$ -linolénique (C20:3 $\omega$ 6, B), arachidonique (C20:4 $\omega$ 6, C),  $\alpha$ -linolénique (C18:3 $\omega$ 3, D), eicosapentaénoïque (C20:5 $\omega$ 3, E) et docosahexaénoïque (C22:6 $\omega$ 3, F) sont présentées. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).

l'EPA (C20:5 $\omega$ 3, [Figure 73E](#)) présente un pourcentage équivalent chez les souris des deux génotypes en début d'épreuve de jeûne, mais une évolution différente de sa proportion entre les deux génotypes explique qu'il participe à séparer les deux génotypes à partir de 18h de jeûne. Contrairement à ce que nous avons observé dans le TABe où une unique phase de modulation de la composition en AG est observée (débutant au bout de 18-24h de jeûne), la composition hépatique en AG est modifiée dès les premières heures de jeûne et les modifications de cette composition s'opèrent, pour les deux génotypes, en deux phases distinctes (voir la forme en L inversé des deux nuages de points sur la [Figure 70A](#) et la flèche illustrant l'effet du temps sur les deux génotypes). Sur plusieurs courbes individuelles, et particulièrement celles des AG de la famille  $\omega$ 6 ([Figure 73A-C](#)), on remarque aisément la présence de deux phases de modulation de la composition hépatique en AG. La première phase correspond globalement à la période 0h-24h de jeûne où les individus se répartissent principalement le long de l'axe 1 de l'ACP ([Figure 70A](#)). Pour les acides gras saturés ([Figures 70B et 71](#)), cette phase se caractérise par une diminution des proportions des acides palmitique (C16:0) et stéarique (C18:0) au profit d'une augmentation des proportions des acides laurique (C12:0) et myristique (C14:0). Les proportions des acides gras monoinsaturés C16:1 $\omega$ 9 ([Figure 72A](#)) et C18:1 $\omega$ 9 ([Figure 72B](#)) augmentent également de manière significative au cours de cette première phase. Pour les acides gras polyinsaturés (AGPI) de la famille  $\omega$ 6, on constate une augmentation très importante de la proportion hépatique du précurseur de la famille : l'acide linoléique (C18:2 $\omega$ 6, [Figure 73A](#)) accompagnée d'une forte diminution des AG plus longs et plus insaturés que sont les acides di-homo- $\gamma$ -linolénique (C20:3 $\omega$ 6, [Figure 73B](#)) et arachidonique (C20:4 $\omega$ 6, [Figure 73C](#)). Chez les souris PPAR $\alpha$ -/-, on constate un comportement à peu près similaire des AGPI de la famille  $\omega$ 3 : accumulation du précurseur l'acide  $\alpha$ -linolénique (C18:3 $\omega$ 3, [Figure 73D](#)) et diminution nette des acides eicosapentaénoïque (EPA, C20:5 $\omega$ 3, [Figure 73E](#)) et docosahexaénoïque (DHA, C22:6 $\omega$ 3, [Figure 73F](#)). En revanche, chez les souris de type sauvage, si le précurseur tend bien à s'accumuler préférentiellement au cours des premières 24h de jeûne (C18:3 $\omega$ 3, [Figure 73D](#)), les AGPI EPA et DHA ([Figure 73E-F](#)) sont globalement maintenus à des niveaux proches de celui précédent la période de jeûne. La seconde phase de modulation de la composition hépatique en lipides s'étend globalement de 24h à 72h de jeûne ([Figure 70A](#)). D'après la [Figure 70B](#), cette phase est principalement caractérisée par l'accumulation hépatique préférentielle d'acides gras monoinsaturés et notamment le C20:1 $\omega$ 9 ([Figure 72C](#)) mais également les C16:1 $\omega$ 9 ([Figure 72A](#)) et C18:1 $\omega$ 9 ([Figure 72B](#)) dont l'augmentation en



**Figure 74 : Classification des gènes et des échantillons sur la base des profils d'expression**

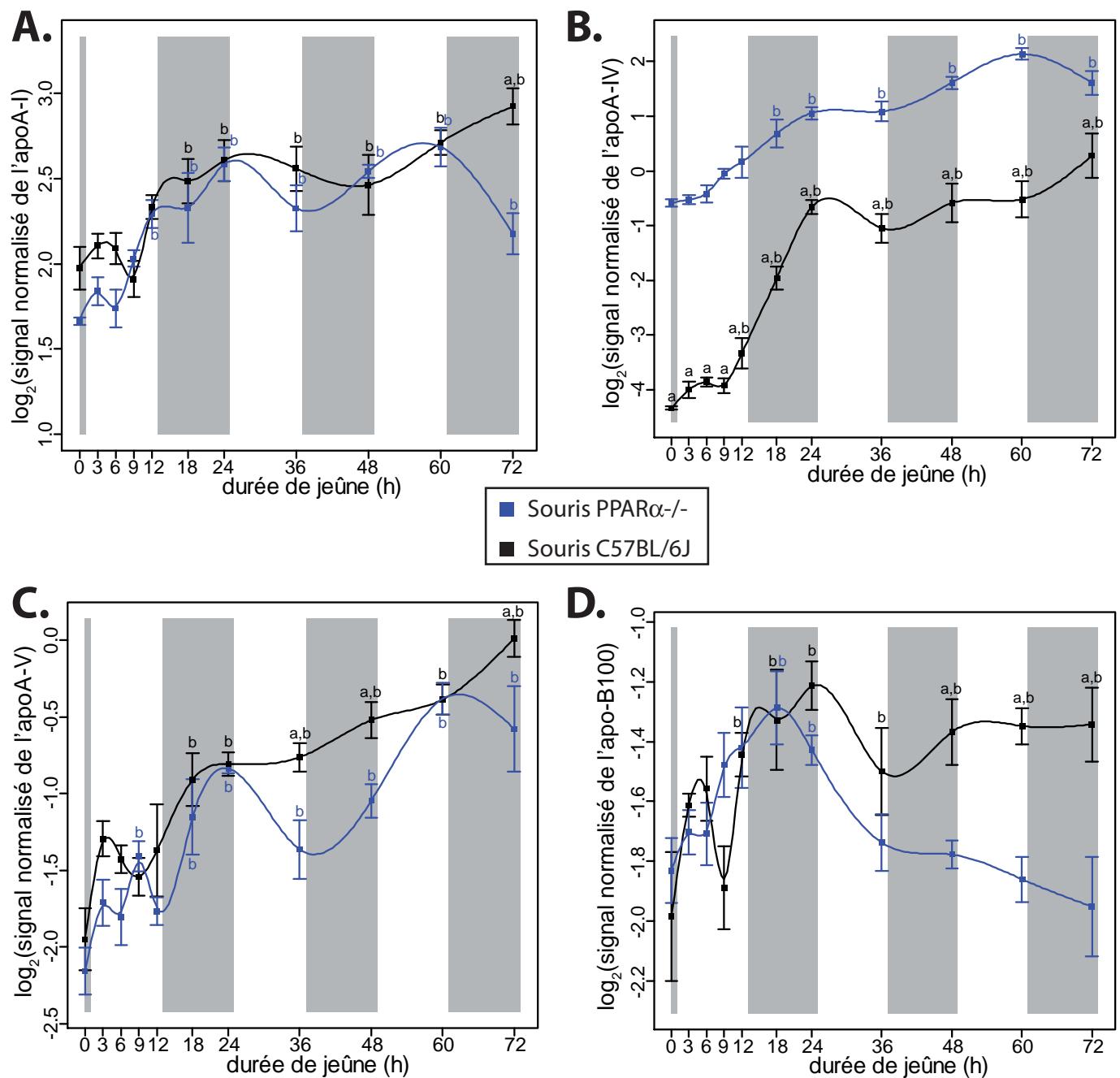
Une classification ascendante hiérarchique (distance= 1-coefficient de corrélation de Pearson, critère d'agglomération de Ward) a été réalisée sur les individus (dendrogramme du haut) et sur les gènes (dendrogramme de gauche). Les gènes ont été centrés et réduits et la matrice de données à été codée selon le code couleur indiqué en bas à droite de la figure. Les génotypes et les temps de jeûne ont également été codés afin de faciliter l'interprétation de la figure (code indiqué en bas à gauche). Quatre groupes de gènes et d'individus sont définis par les coupures des dendogrammes illustrées par des pointillés. Des éléments d'interprétation des 4 groupes d'individus ont également été ajoutés sur le dendrogramme correspondant.

proportion se poursuit après 24h de jeûne (à une vitesse moindre pour le C18:1 $\omega$ 9 cependant). Durant cette même phase, la proportion d'acide  $\alpha$ -linolénique (C18:3 $\omega$ 3, [Figure 73D](#)) diminue pour retrouver en fin de jeûne des valeurs proches de celles observées avant le début de l'épreuve de jeûne. Alors que ces modulations sont pour la plupart très comparables entre les souris PPAR $\alpha$ -/- et les souris de type sauvage, nous avons déjà évoqué le fait que le plan 3-4 de l'ACP ([Figure 70C-D](#)) met en lumière des effets du jeûne dépendant du génotype des souris.

#### *Evolution des profils d'expression hépatiques d'un panel de gènes au cours du jeûne :*

Ayant caractérisé un certain nombre de paramètres biochimiques et leurs évolutions au cours du jeûne chez les souris de type sauvage et de type PPAR $\alpha$ -/-, nous avons souhaité évaluer quel était le rôle précis joué par le récepteur PPAR $\alpha$  dans les modulations d'expressions géniques accompagnant l'adaptation du métabolisme de l'organisme à la privation d'aliment. Nous avons donc mesuré les niveaux d'expression d'un panel de gènes au niveau hépatique chez les souris des deux génotypes à l'aide de la puce INRAarray 01.3 qui comporte de nombreux gènes cibles du récepteur PPAR $\alpha$  impliqués notamment dans le métabolisme des lipides et des glucides. Sur les 192 gènes que peut potentiellement mesurer cette version de notre puce à ADN, 130 ont été retenus comme mesurés à des niveaux acceptables dans cette expérience et sur les 88 échantillons analysés (2 génotypes  $\times$  11 temps de jeûne  $\times$  4 individus par lot) 84 puces ont donné des résultats satisfaisant (voir Matériels et Méthodes).

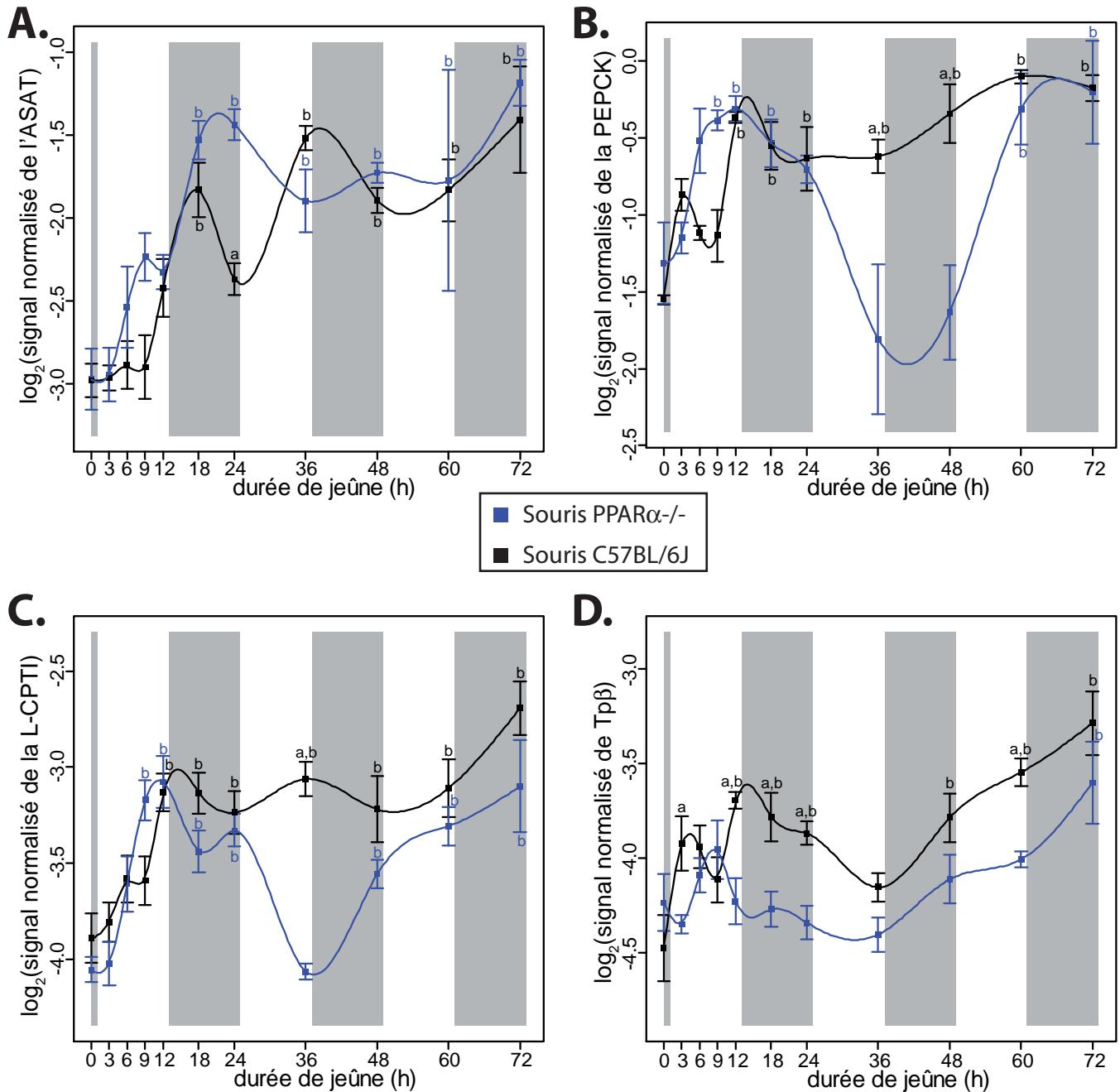
Une première analyse a consisté à réaliser une classification des échantillons et des gènes sur la base des profils d'expression obtenus. Une représentation sous la forme de codes couleurs (heatmap) de la matrice de données dont les lignes et les colonnes ont été réorganisées par le biais d'une classification ascendante hiérarchique est présentée en [Figure 74](#). Sur la base de graphiques de la hauteur de coupure en fonction du nombre de groupes définis (on recherche une cassure de pente dans ce type de graphique), les dendrogrammes des individus et des gènes ont été coupés de manière à définir 4 groupes ([Figure 74](#)). Les individus sont nettement séparés en fonction du génotype, illustrant le fait que les souris PPAR $\alpha$ -/- présentent des modulations constitutives de l'expression de nombreux gènes présents sur la puce INRAarray 01.3. De manière surprenante, au sein de chaque génotype, deux groupes sont définis qui correspondent respectivement i) à des animaux ayant subit une durée de jeûne inférieure ou égale à 12h et ii) à des animaux ayant subit une durée de jeûne



**Figure 75 : Profils d'expression hépatique de 4 apolipoprotéines**

Les évolutions au cours du jeûne de l'expression hépatique de l'apoA-I (A), de l'apoA-IV (B), de l'apoA-V (C) et de l'apo-B100 (D) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%). apo : apolipoprotéine.

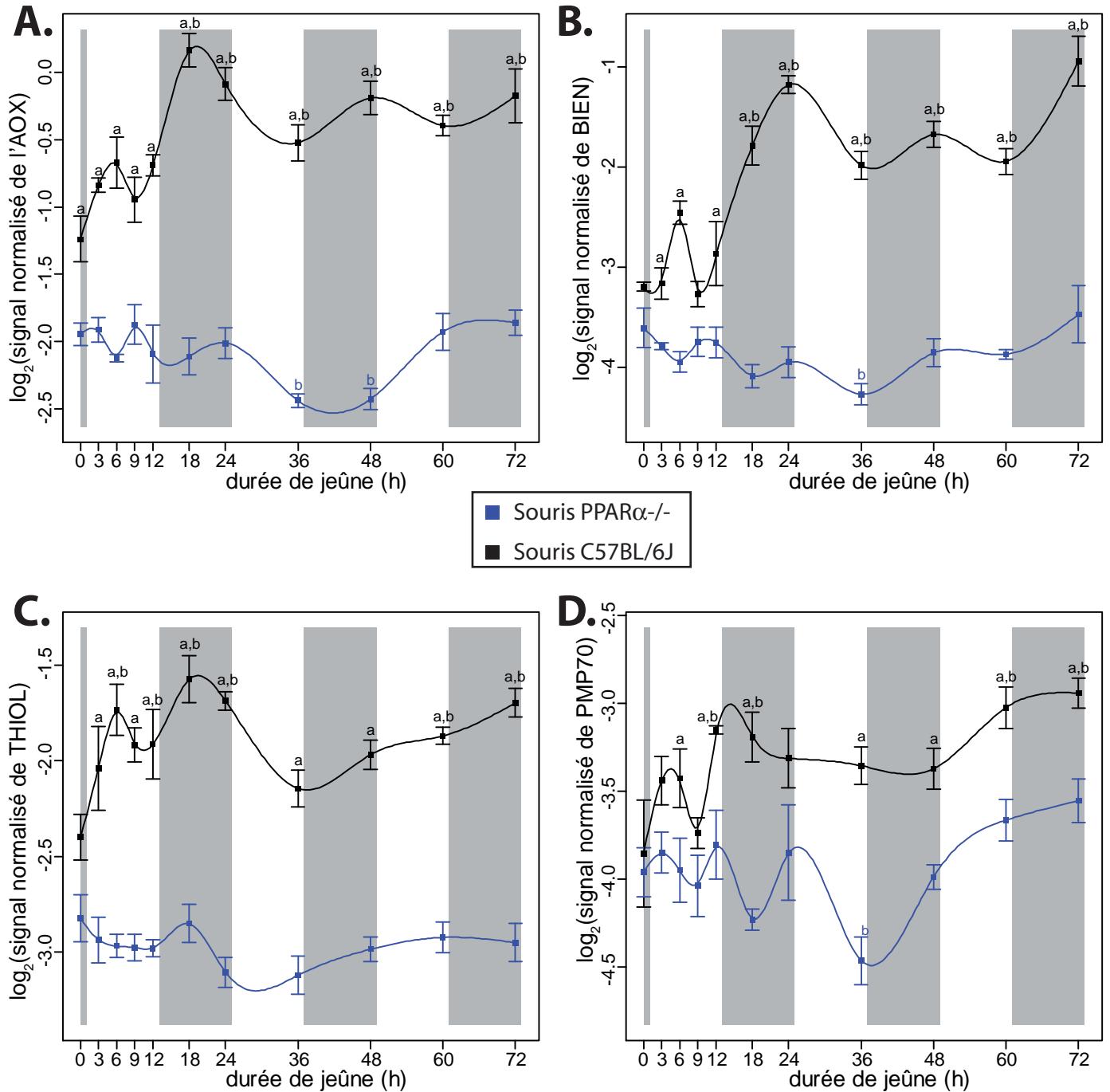
supérieure ou égale à 18h ([Figure 74](#)). Cette séparation stricte illustre le fait que c'est entre 12h et 18h de jeûne que les profils d'expression des gènes mesurés présentent leurs modulations les plus notables. Cette période de modification maximale du profil d'expression des gènes est maintenue chez les souris des deux génotypes mais il apparaît cependant nettement que la séparation des individus de type sauvage en fonction de la période de jeûne est plus franche que celle des souris PPAR $\alpha$ -/- (en coupant l'arbre plus haut, on aurait obtenu un seul groupe pour les souris PPAR $\alpha$ -/- mais toujours deux pour les souris de type sauvage). Ceci illustre le fait que le récepteur PPAR $\alpha$  participe, chez les souris de type sauvage, à moduler fortement les profils d'expression de certains gènes mesurés en particulier entre 12h et 18h de jeûne. Les comportements généraux des quatre groupes de gènes définis en [Figure 74](#) peuvent être évalués visuellement à l'aide des couleurs de la matrice de données dans la mesure où les échantillons sont globalement classés, par génotype, de gauche à droite en fonction de l'augmentation de la durée de jeûne (l'ordre des individus n'est pas parfait mais le dendrogramme a volontairement été réorganisé de manière à se rapprocher au mieux de l'ordre des individus en fonction de leur durée de jeûne tout en respectant les contraintes du dendrogramme initial). Le premier groupe contient 28 gènes dont les fonctions sont relativement variées. Il comprend néanmoins plusieurs apolipoprotéines (apoA-I, A-IV, A-IV, apoB et apoE), quelques gènes impliqués dans le métabolisme intracellulaire des lipides (ACOTH, CACP, L-CPTI, Lpin1 et Tp $\beta$ ) et plusieurs gènes impliqués dans le métabolisme des xénobiotiques ou son contrôle (Cyp3a11, Cyp2b10 et 2b13, GST $\mu$ , Abcb4/MDR2, Abcc2/cMOAT et CAR). Au temps t=0h et aux faibles temps de jeûne les expressions de plusieurs gènes de ce groupe semblent plus faibles chez les souris PPAR $\alpha$ -/- par rapport aux souris de type sauvage sans que cela soit cependant une règle absolue au sein du groupe (l'apoA-IV constituant le meilleur contre-exemple). Ces gènes voient majoritairement leur expression augmenter au cours du jeûne chez les souris des deux génotypes. On distingue sur la [Figure 74](#) un sous-groupe (situé à peu près au milieu du Groupe 1) pour lequel les souris PPAR $\alpha$ -/- semblent présenter une induction retardée par rapport aux souris de type sauvage (il s'agit majoritairement de gènes pour lesquels des régulations PPAR $\alpha$ -dépendantes ont été décrites mais qui présentent malgré tout une induction de leur expression chez les souris PPAR $\alpha$ -/- au bout de 48h-72h de jeûne) et un autre sous-groupe (en bas du groupe 1) où au contraire l'augmentation de l'expression des gènes semble plus précoce chez les souris PPAR $\alpha$ -/. Le deuxième groupe ([Figure 74](#)) comporte également 28 gènes. Ce groupe est particulièrement riche en cibles bien décrites du récepteur PPAR $\alpha$  comme les Cyp4a10 et



**Figure 76 : Profils d'expression hépatique de l'ASAT, de la PEPCK, de la L-CPTI et de Tp $\beta$**   
 Les évolutions au cours du jeûne de l'expression hépatique de l'ASAT(A), de la PEPCK (B), de la L-CPTI (C) et de Tp $\beta$  (D) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%). ASAT : Aspartate aminotransférase, PEPCK : phosphoenolpyruvate carboxykinase, L-CPTI : Liver carnitine palmitoyltrasférase I, Tp $\beta$  : sous-unité  $\beta$  de l'enzyme mitochondriale trifonctionnelle.

Cyp4a14, les trois enzymes de la  $\beta$ -oxydation peroxyxsomale AOX, BIEN et THIOL, des gènes impliqués dans la  $\beta$ -oxydation mitochondriale des AG (CPT-II, Tp $\alpha$ ) ou dans la cétogenèse (HMGCoA synthase mitochondriale) ainsi que d'autres cibles aux fonctions plus diverses (Cyp8b1, PMP70, les enoyl-CoA isomérasées, FXR et ALDH3 par exemple). De manière cohérente, on observe en [Figure 74](#) que ces gènes sont majoritairement induits chez les souris de type sauvage au cours du jeûne mais pas chez les souris PPAR $\alpha$ -/. Le troisième groupe ([Figure 74](#)) comporte 51 gènes dont les profils d'expression semblent relativement identiques entre les souris des deux génotypes. Une augmentation de l'expression de la majorité de ces gènes est observée à la fin de l'épreuve de jeûne (t=60h et 72h) et quelques variations d'expression, pas toujours significatives, sont également constatées entre 12 et 24h de jeûne. Le caractère hétérogène de la composition de ce groupe qui contient beaucoup de gènes aux fonctions différentes et aux profils d'expression relativement variables rend difficile son interprétation globale. D'après l'observation qui a été faite des profils d'expression individuels, ce troisième groupe constitue vraisemblablement un regroupement des gènes dont les profils d'expression ne présentent pas une tendance nette de variation au cours du jeûne. L'augmentation de l'expression de la plupart de ces gènes en fin de jeûne, pourrait être liée à un biais imparfaitement corrigé par la normalisation (on constate sur une représentation en boxplot que les signaux normalisés des puces correspondant à ces temps de jeûne sont généralement supérieurs à ceux des autres puces). Alternativement, il est possible que ces temps de jeûne conséquents pour des souris s'accompagnent de modulations globales de la transcription des gènes mais la confirmation de cette hypothèse nécessiterait objectivement une évaluation indépendante. Enfin, le quatrième groupe ([Figure 74](#)) comporte 23 gènes qui sont majoritairement impliqués dans la biosynthèse des acides gras (FAS, S14, SCD1, Elo3), du cholestérol (HMG-CoA synthase cytosolique, HMG-CoA réductase, FDFT, FPP synthase) et des acides biliaires (Cyp7a1), dans le captage des acides gras par les hépatocytes (récepteur aux LDL, lipase hépatique, Lrp-1) et dans le métabolisme des xénobiotiques (ALDH1, Cyp2c29, GST $\alpha$ , GST $\pi$ 2 et LCE). L'expression de la plupart de ces gènes diminue nettement au cours du jeûne chez les souris des deux génotypes ([Figure 74](#)). De manière surprenante, on trouve également dans ce groupe quelques gènes dont l'expression est généralement induite par les activateurs pharmacologiques de PPAR $\alpha$ , comme L-FABP, la catalase et LEF1.

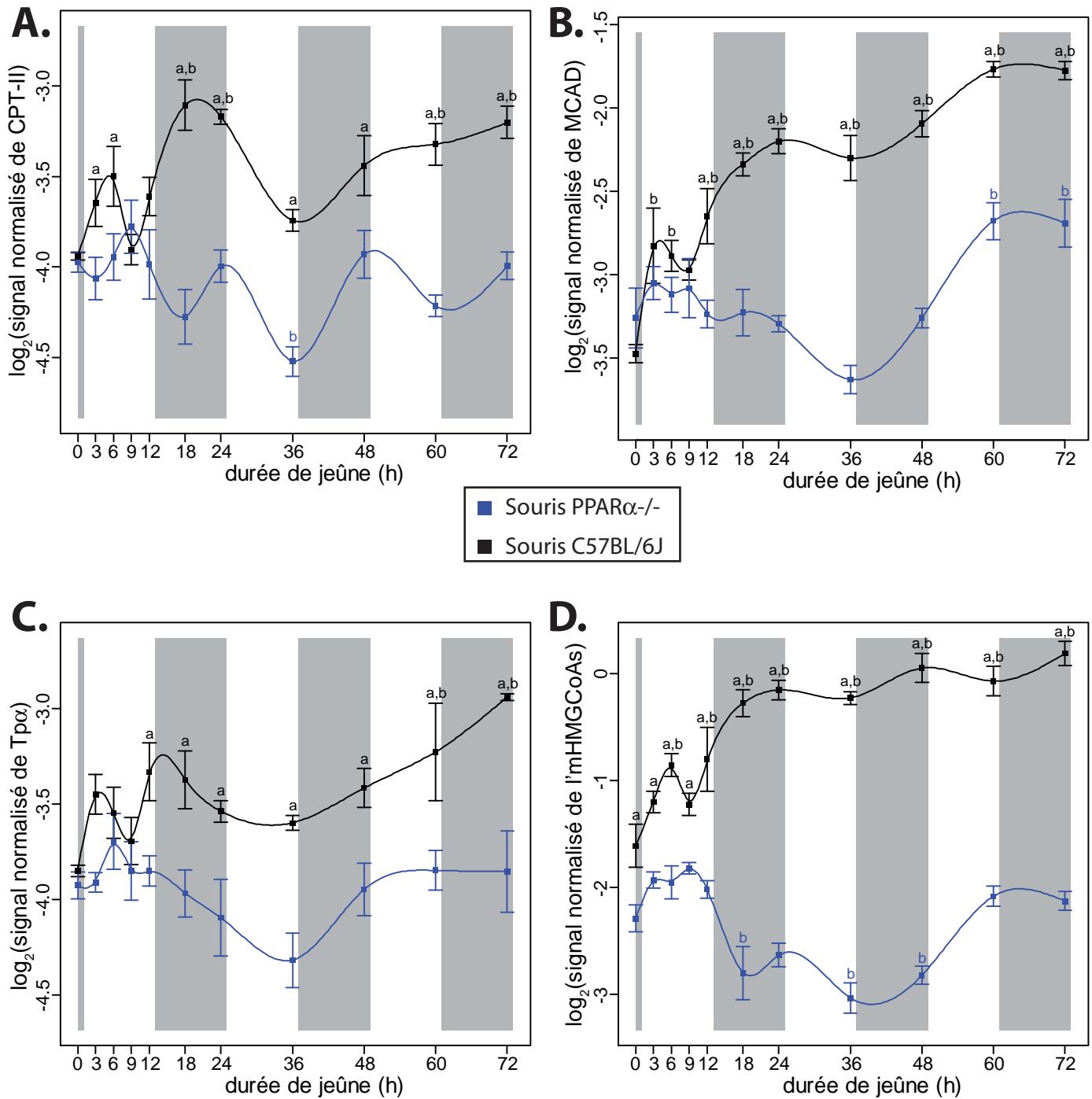
Afin de faciliter 1) l'établissement de liens potentiels entre les variations des paramètres biochimiques que nous avons observés et les profils d'expression des gènes et 2) l'interprétation des quatre groupes de gènes identifiés précédemment, nous présentons ci-



**Figure 77 : Profils d'expression hépatique de gènes impliqués dans l'oxydation peroxyxsomale**

Les évolutions au cours du jeûne de l'expression hépatique de l'acyl-CoA oxydase (AOX, A), de l'enzyme bifonctionnelle (BIEN, B), de la 3-cétoacyl-CoA thiolase B (THIOL, C) et de l'Abc transporteur D3 (PMP70/Abcd3, D) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).

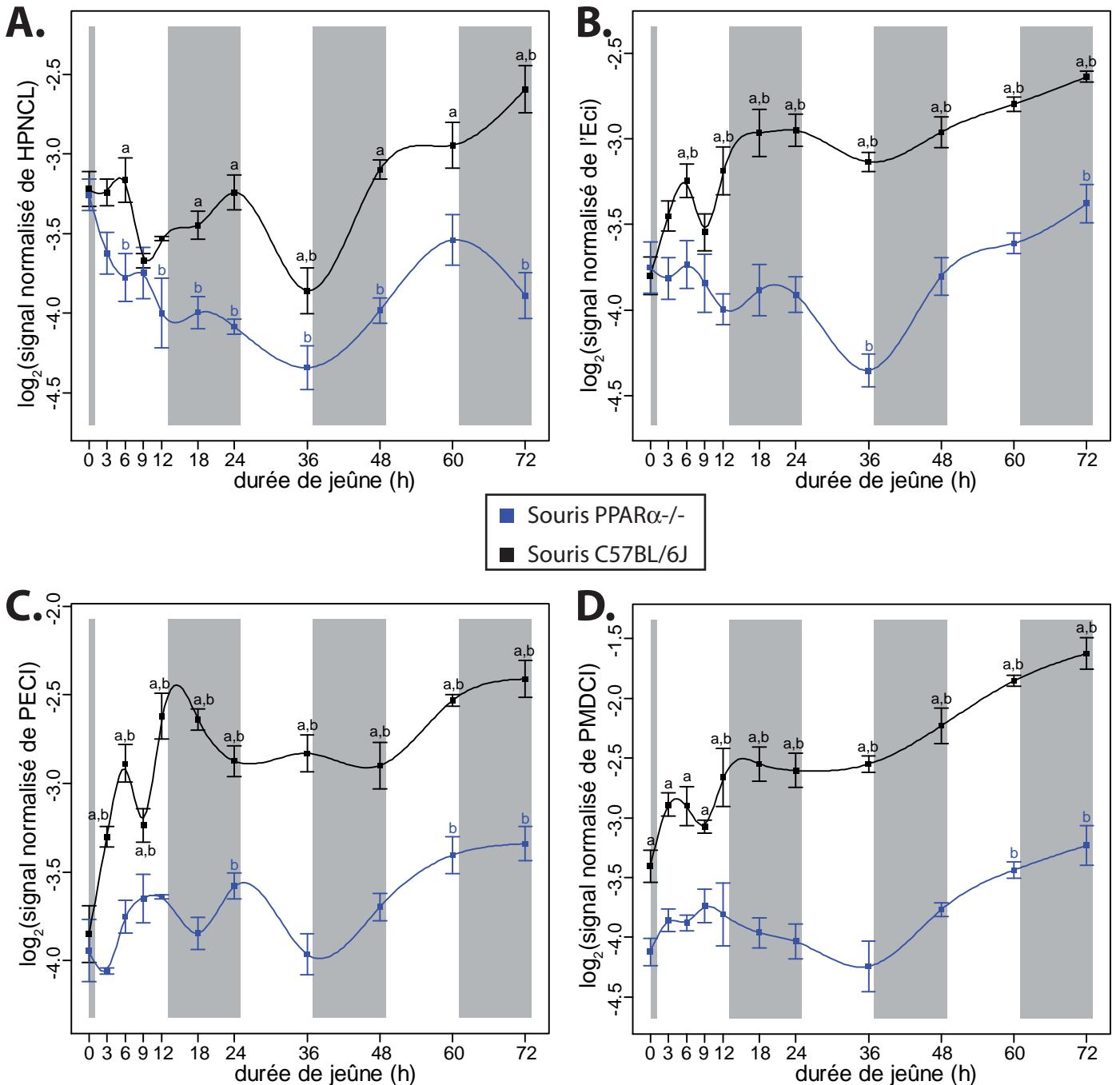
dessous certains profils individuels en respectant la structure des groupes que nous venons de définir (par souci de simplification les gènes du groupe 3 seront omis). La [Figure 75](#) présente les profils d'expression de quatre apolipoprotéines appartenant au groupe 1. L'expression de ces apolipoprotéines augmente au cours du jeûne chez les souris des deux génotypes. Pour l'apolipoprotéine B100, les souris PPAR $\alpha$ -/- semblent néanmoins incapables de maintenir un niveau d'expression comparable à celui des souris de type sauvage au delà de 24h de jeûne. Pour l'apolipoprotéine A-IV, on constate une forte surexpression constitutive de ce gène chez les souris PPAR $\alpha$ -/- (que nous avons également observée dans plusieurs expérimentations indépendantes de celle-ci) mais une régulation au cours du jeûne assez comparable aux souris de type sauvage. Enfin, l'apoAV humaine a été décrite comme une cible de PPAR $\alpha$  (Prieur et al., 2003; Vu-Dac et al., 2003) mais est également régulée négativement par SREBP-1c (Jakel et al., 2004). Nos données suggèrent que le gène murin est bien régulé par SREBP (dont l'activité au cours du jeûne diminue comme l'atteste plusieurs régulations de gènes du groupe 4, voir ci-dessous) au cours du jeûne et que le rôle de PPAR $\alpha$  dans la régulation de l'apoAV est sans doute limité chez la souris (on observe néanmoins une différence significative d'expression de l'apoAV entre les souris des deux génotypes après 36h, 48h et 72h de jeûne, [Figure 75](#)). Ce premier groupe de gène contient également plusieurs gènes impliqués dans le métabolisme énergétique. La [Figure 76](#) présente les profils d'expression de l'ASAT impliquée dans le cycle de l'urée ([Figure 62](#)), de la PEPCK, impliquée dans la néoglucogenèse ([Figure 33](#)) et de deux enzymes impliquées dans la  $\beta$ -oxydation mitochondriale qui ont été décrites comme des gènes cibles de PPAR $\alpha$  : la L-CPTI impliquée dans l'entrée des AG dans la mitochondrie ([Figure 25](#)) et la sous-unité  $\beta$  de l'enzyme trifonctionnelle (Tp $\beta$ ) catalysant la dernière étape de la  $\beta$ -oxydation mitochondriale ([Figure 27](#)). Malgré quelques différences entre les deux génotypes ne concernant que un à deux points de mesure, les régulations positives de l'ASAT, de la PEPCK et de la L-CPTI semblent s'opérer de manière comparable entre les deux génotypes ([Figure 76](#)). Il est intéressant de constater que la principale différence entre les deux génotypes en ce qui concerne l'ASAT est observée au bout de 24h de jeûne, reproduisant ainsi une observation publiée (Kersten et al., 2001), mais illustrant l'intérêt additionnel d'une approche en cinétique du jeûne car cette différence n'est pas observée pour toutes les durées de jeûne. Le parallèle entre le niveau d'expression de l'ASAT et la concentration plasmatique d'urée ([Figure 64](#)) n'apparaît pas évident, suggérant qu'une analyse plus approfondie de l'expression d'autres enzymes impliquées dans le métabolisme des acides aminés est nécessaire pour mieux appréhender cet aspect des adaptations métaboliques au



**Figure 78 : Profils d'expression hépatique de gènes impliqués dans l'oxydation mitochondriale**

Les évolutions au cours du jeûne de l'expression hépatique de la carnitine palmitoyltransférase II (CPT-II, A), de la medium chain acyl-CoA dehydrogenase (MCAD, B), de la sous-unité  $\alpha$  de l'enzyme mitochondriale trifonctionnelle (Tp $\alpha$ , C) et de l'HMG-CoA synthase mitochondriale (mHMGCoAs, D) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).

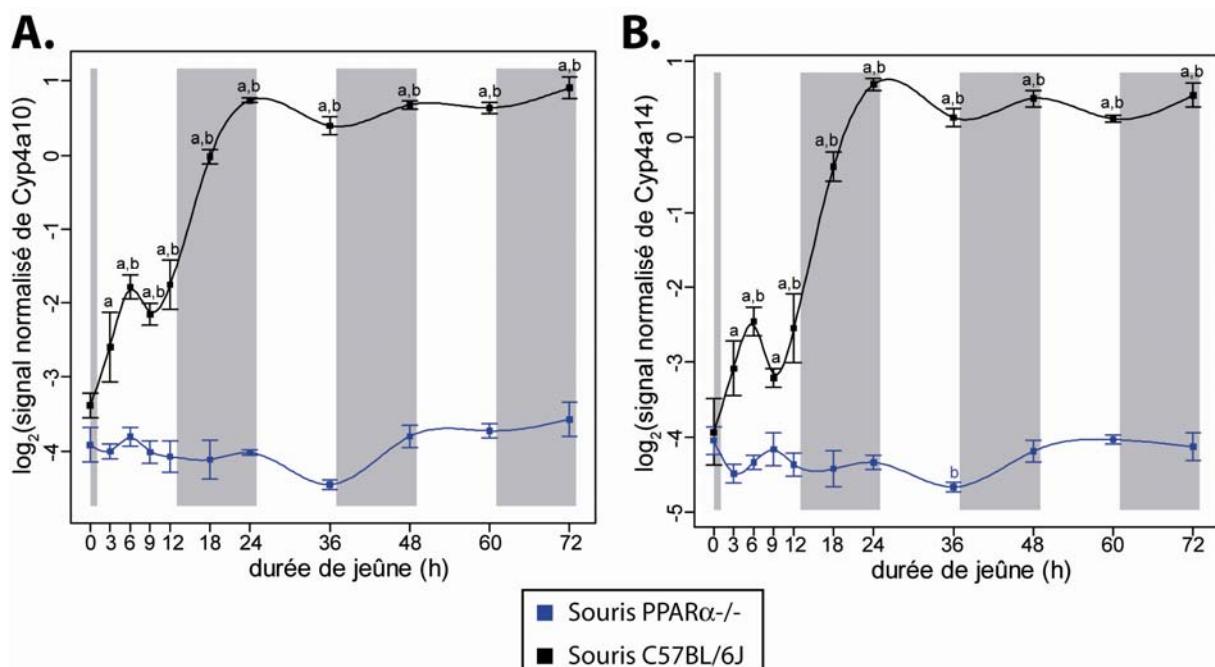
jeûne. Alors que l'induction de la PEPCK chez les souris de type sauvage semble s'accompagner d'un léger fléchissement de la diminution du glucose sanguin ([Figure 64](#), autour de t=18h), les modulations de la PEPCK observées chez les souris PPAR $\alpha$ -/- à 36h et 48h de jeûne, ne semblent pas correspondre à des modifications substantielles de la glycémie et nécessiteront donc d'être confirmer par une technique indépendante. Finalement, le gène Tp $\beta$  ([Figure 76](#)) semble présenter un comportement relativement classique pour un gène cible de PPAR $\alpha$  : régulé positivement chez les souris de type sauvage uniquement au cours du jeûne, illustrant ainsi les limites des méthodes de mesure (la sonde utilisée pour ce gène n'est peut-être pas optimale ou bien le gène est peut-être trop faiblement exprimé) et de classification des gènes utilisées qui sont sans doute atteintes dans des cas comme celui-ci où la modulation est de relativement faible ampleur et où une régulation est également observée chez les souris PPAR $\alpha$ -/- à la fin de l'épreuve de jeûne ([Figure 76](#)). Pour le deuxième groupe de gènes, les profils d'expression de gènes impliqués dans l'oxydation des AG au niveau peroxysomal sont présentés en [Figure 77](#). Les trois enzymes de la  $\beta$ -oxydation peroxysomale (AOX, BIEN et THIOL) sont induites au cours du jeûne chez les souris de type sauvage uniquement. Le transporteur peroxysomal PMP70 présente un profil d'expression similaire. De légères diminutions transitoires de l'expression de ces gènes sont également constatées chez les souris PPAR $\alpha$ -/. Il est intéressant de constater que le niveau maximal d'expression de ces gènes s'établit bien entre 12h et 24h de jeûne ce qui semble correspondre, comme nous l'avons vu, à la phase de modulation maximale des signatures transcriptionnelles mais également à la période d'intervention clef du récepteur PPAR $\alpha$ . La [Figure 78](#) présente les profils d'expression de gènes du groupe 2 impliqués dans la  $\beta$ -oxydation mitochondriale des AG (CPT-II, MCAD, Tp $\alpha$ ) et dans la cétogenèse (HMG-CoA synthase mitochondriale). Ces gènes présentent des profils similaires à ceux que nous venons de décrire. Comme précédemment, nous constatons, aux alentours de 36h de jeûne, une diminution parfois significative de l'expression de certains gènes cibles de PPAR $\alpha$  chez les souris PPAR $\alpha$ -/. De plus, les niveaux d'induction observés chez les souris de type sauvage restent de manière générale assez modestes ([Figures 77 et 78](#)) par rapport à ce qui peut être observé avec un activateur pharmacologique de type fibrate (Martin et al., 2005, et données non présentées). Il est intéressant de constater le parallèle assez frappant entre le profil d'expression de l'HMG-CoA synthase mitochondriale ([Figure 78](#)) et le profil de concentration plasmatique en  $\beta$ -hydroxybutyrate ([Figure 64](#)), suggérant que la régulation transcriptionnelle de ce gène constitue un déterminant majeur de la régulation d'activité de son produit. Les profils



**Figure 79 :** Profils d'expression hépatique de gènes impliqués dans l'oxydation des AGPI

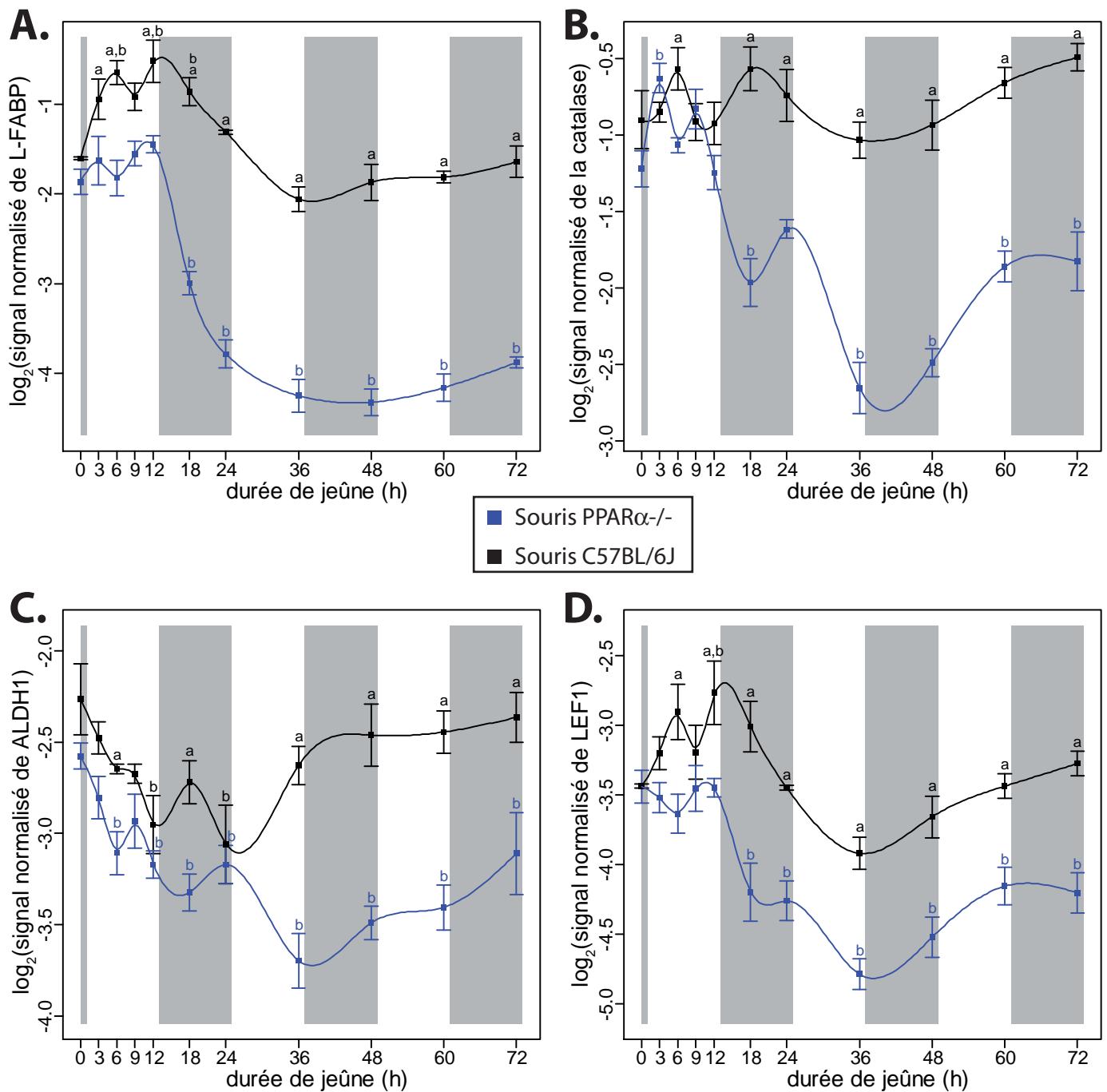
Les évolutions au cours du jeûne de l'expression hépatique de l'hydroxyphytanoyl-CoA lyase (HPNCL, A), de l'enoyl-CoA isomérase (Eci, B), de la peroxisomal enoyl-CoA isomérase (PECl, C) et de la peroxisomal/mitochondrial dienoyl-CoA isomérase (PMDCI, D) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%). AGPI : acides gras polyinsaturés.

d'expression de gènes impliqués dans l' $\alpha$ -oxydation de l'acide phytanique (HPNCL) et dans l'oxydation des acides gras polyinsaturés (enoyl-CoA isomérasées Eci, PECI et PMDCI) sont présentés en [Figure 79](#). L'hydroxyphytanoyl-CoA lyase (HPNCL, [Figure 79A](#)) est régulée négativement au cours du jeûne chez les souris PPAR $\alpha$  $^{-/-}$  mais son niveau est maintenu chez les souris de type sauvage. Nos données indiquent que l'HPNCL constitue une cible vraisemblable du récepteur PPAR $\alpha$  car ce gène est fortement induit chez les souris de type sauvage uniquement suite à un traitement par un fibrate ou un agoniste de RXR, notamment au niveau rénal (Martin et al., 2005). Nous verrons qu'un certain nombre de gènes cibles de PPAR $\alpha$  présentent des comportements similaires à l'HPNCL au cours du jeûne mais ils ont été pour la plupart, regroupés dans le quatrième groupe par la classification. De même, les enoyl-CoA isomérasées sont vraisemblablement des gènes cibles de PPAR $\alpha$  (Martin et al., 2007; Martin et al., 2005) même si aucun élément de réponse n'a encore été rapporté dans les promoteurs de ces gènes.



**Figure 80 : Profils d'expression hépatique des cytochromes P450 4a**

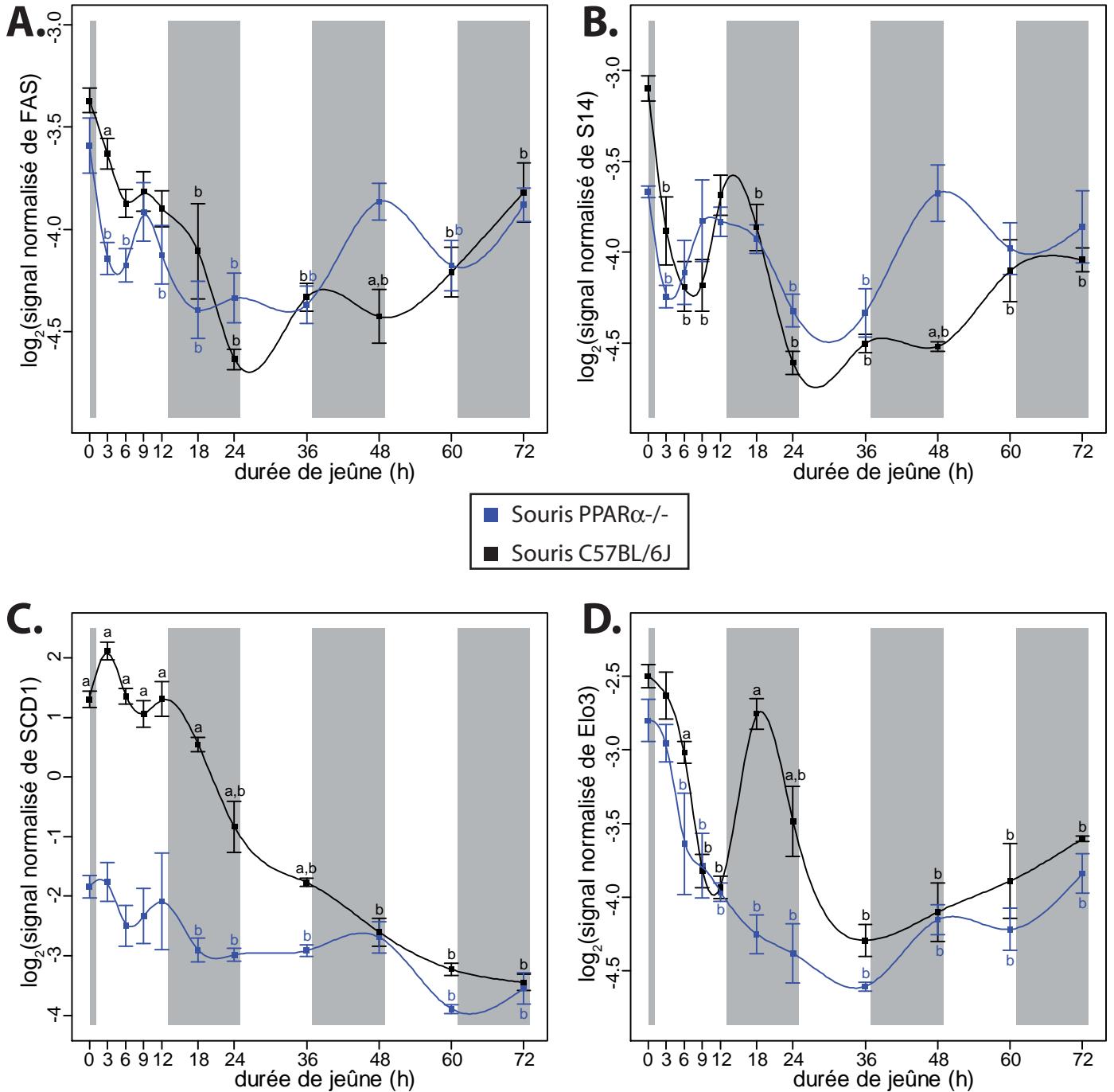
Les évolutions au cours du jeûne de l'expression hépatique de Cyp4a10 (A) et de Cyp4a14 (B) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%). Cyp : cytochrome P450.



**Figure 81 :** Profils d'expression hépatique de L-FABP, catalase, ALDH1 et LEF1

Les évolutions au cours du jeûne de l'expression hépatique de la liver fatty acid-binding protein (L-FABP, A), de la catalase (B), de l'aldéhyde déshydrogénase 1 (ALDH1, C) et du lymphoid enhancer binding factor 1 (LEF1, D) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%).

Ces enzymes, nécessaires à l'entrée des acides gras polyinsaturés dans les cycles de  $\beta$ -oxydation, présentent des profils d'expression très similaires au cours du jeûne et qui sont caractéristiques des autres gènes cibles du récepteur PPAR $\alpha$  que nous venons d'évoquer ([Figure 79B-D](#)). On constate cependant une légère induction de ces trois gènes chez les souris PPAR $\alpha$ -/- lors des plus longues durées de jeûne de cette épreuve. Il est probable que ces gènes participent à l'établissement et à la régulation de la composition hépatique en acides gras en situation normale d'alimentation (Martin et al., 2007). Cependant, le fait que l'évolution de la composition hépatique en lipides au cours du jeûne soit assez similaire pour les souris des deux génotypes ([Figure 70](#)) alors que la régulation de ces gènes est très différente entre les deux génotypes ([Figure 79B-D](#)) suggère qu'un réseau plus large de gènes et de phénomènes biochimiques sont impliqués dans la régulation de la composition hépatique en lipides au cours du jeûne. En revanche, il est probable que la régulation différentielle de ces gènes participe à déterminer les différences quantitatives d'accumulation hépatique des AG ([Figure 65B](#)) entre les deux génotypes. Finalement, les gènes cibles de PPAR $\alpha$  qui répondent sans doute le plus nettement à l'activation du récepteur au cours du jeûne sont les Cyp4a ([Figure 80](#)) dont le niveau maximal d'expression au cours du jeûne chez les souris de type sauvage atteint environ 20-25 fois le niveau initial d'après nos mesures par puce à ADN. Cette observation souligne que le récepteur PPAR $\alpha$  est bien fortement activé au cours du jeûne mais il est possible que d'autres voies de signalisation inhibent l'expression de certains gènes cibles de PPAR $\alpha$  au cours du jeûne ce qui permettrait d'expliquer 1) les diminutions d'expressions observées sur plusieurs gènes chez les souris PPAR $\alpha$ -/- soumises au jeûne, 2) des ampleurs d'inductions atténées pour certains gènes cibles de PPAR $\alpha$  chez les souris de type sauvage et 3) parfois l'absence quasi-complète chez les souris de type sauvage d'induction de gènes constituant pourtant de bonnes cibles du récepteur. L'un des exemples que nous avons présenté est le gène HPNCL ([Figure 79A](#)) mais d'autres gènes, appartenant principalement au groupe 4 présentent des profils encore plus caractéristiques de cette situation. La [Figure 81](#) présente notamment les profils d'expression de gènes du groupe 4 : la L-FABP, la catalase, ALDH1 et LEF1 qui se caractérisent par une chute nette de l'expression au cours du jeûne chez les souris PPAR $\alpha$ -/- et un maintien global de l'expression chez les souris de type sauvage. La L-FABP (Issemann et al., 1992; Kaikaus et al., 1993; Schachtrup et al., 2004; Simon et al., 1993) et la catalase (Chu et al., 2004; Girnun et al., 2002) constituent des gènes cibles bien caractérisés de PPAR $\alpha$  et nos données non publiées de criblages des effets du fénofibrate suggèrent que ALDH1 et LEF1 pourraient également être

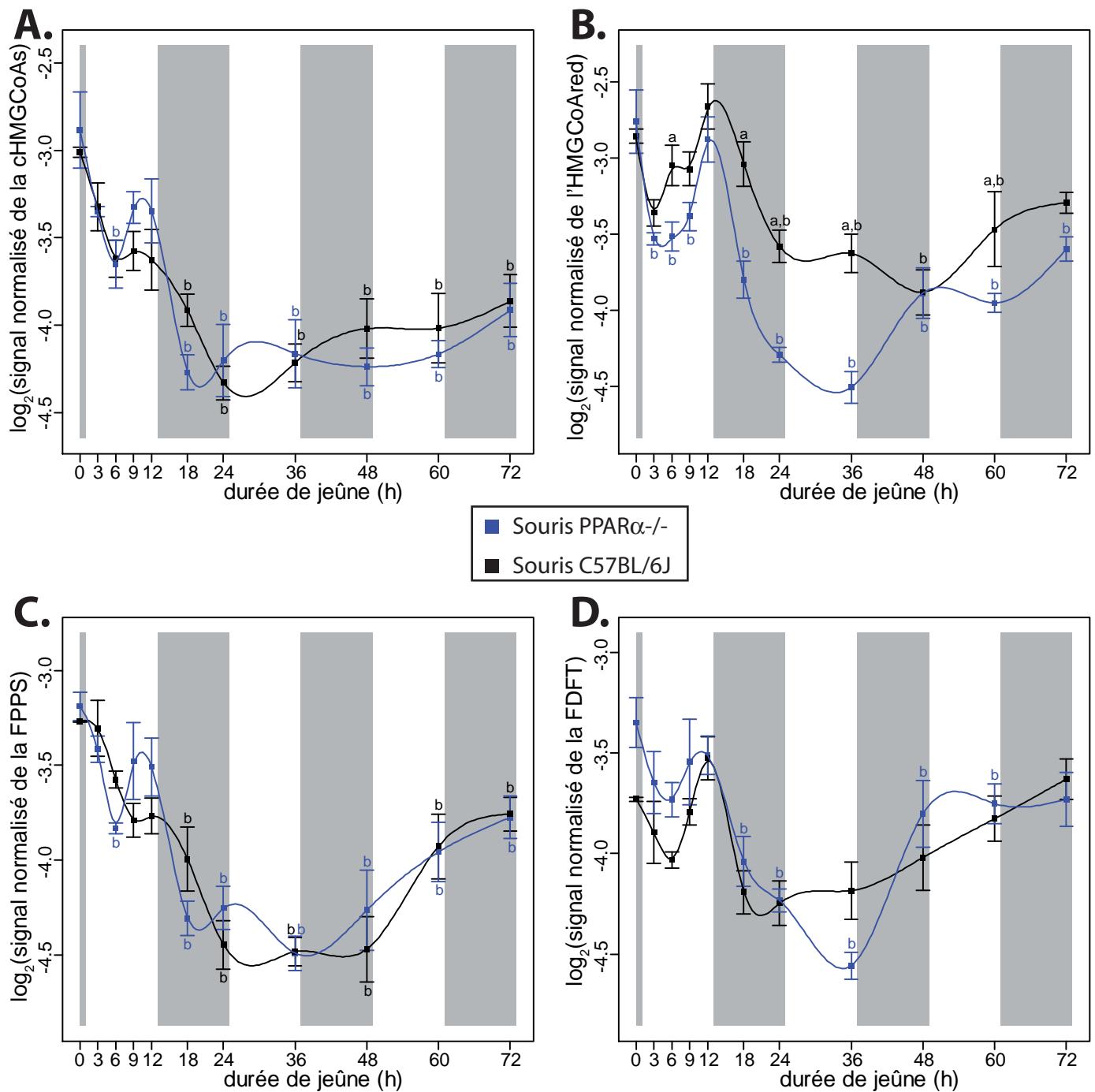


**Figure 82 : Profils d'expression hépatique de gènes impliqués dans la biosynthèse des AG**

Les évolutions au cours du jeûne de l'expression hépatique de la fatty acid synthase (FAS, A), de Spot14 (S14, B), de la  $\Delta 9$ -désaturase (SCD1, C) et de l'élongase 3 (Elo3, D) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%). AG : acides gras

des gènes cibles de PPAR $\alpha$ . Le regroupement de ces gènes dans le quatrième groupe est lié à leur répression au cours du gène chez les souris PPAR $\alpha$ . Comme l'illustrent les profils d'expression de gènes impliqués dans la biosynthèse des acides gras ([Figure 82](#)) et du cholestérol ([Figure 83](#)) qui constituent des cibles des récepteurs SREBP et appartiennent à ce quatrième groupe de gènes, les niveaux d'expression des gènes du groupe 4 diminuent nettement au cours du jeûne. Cependant, les diminutions d'expression pour les gènes présentés en [Figures 82 et 83](#) sont, de manière générale, observées chez les souris des deux génotypes, illustrant là encore les limites de la classification qui a été réalisée.

Au delà des interprétations biologiques de ces différents résultats qui seront discutés dans le paragraphe suivant, ces données nous ont conduits à nous interroger sur la façon de classer des profils d'expression en fonction du temps. Rapidement, il nous est apparu que le paramètre pertinent n'était pas le profil d'expression en lui-même (dont nous avons vu que le niveau en particulier pouvait être influencé par la qualité de la sonde nucléotidique déposée sur la puce) mais sa dérivée qui, si l'on admet que le profil initial est le reflet de l'abondance de l'ARNmessager, représente alors les variations de l'abondance de cet ARN en fonction du temps, ou, par analogie au domaine de la physique, représente une vitesse de production ou de disparition de cet ARN. L'intérêt d'un travail sur les dérivées consiste donc principalement 1) à s'affranchir des niveaux absolus des courbes (mais ceci est également réalisé dans la classification des gènes présentée en [Figure 74](#) par le biais d'un centrage des gènes) mais surtout 2) à travailler sur une quantité (la dérivée du profil d'expression) qui, si la demi-vie de l'ARN messager reste constante, représente en théorie directement le taux de transcription du gène considéré. Afin d'obtenir un profil d'expression qui soit dérivable en tout point, nous avons utilisé un lissage par des splines cubiques, similaire à celui qui a été utilisé pour tracer les courbes de tendances des profils présentés dans de nombreuses figures de ce paragraphe. Le principal problème auquel nous avons été confrontés a consisté à déterminer la valeur du paramètre de lissage  $\lambda$ . Dans les courbes qui sont présentées dans ce manuscrit, j'ai choisi un paramètre  $\lambda$  faible ( $\lambda=0,1$  pour toutes les courbes présentées), de manière à ce que les courbes splines passent au plus près des moyennes. Un tel lissage représente néanmoins selon moi une vision quelque peu optimiste de la qualité des mesures et il est probable qu'un certain nombre de moyennes aient été mal estimées (en raison du caractère bruité de la technique de mesure et de l'utilisation de seulement 4 animaux par lot) et que le profil d'expression réel ne passe pas strictement par ces points. L'utilisation d'un paramètre de lissage aussi faible, conduit à générer des dérivées qui oscillent fortement et un grand nombre de fois autour de 0. Si l'on



**Figure 83 : Profils d'expression hépatique de gènes de la biosynthèse du cholestérol**

Les évolutions au cours du jeûne de l'expression hépatique de l'HMG-CoA synthase cytosolique (cHMGCoAs, A), de l'HMG-CoA réductase (HMGCoAred, B), de la farnesyl diphasosphate synthétase (FPPS, C) et de la squalène synthase (FDFT, D) sont présentées pour les souris des deux génotypes. Les rectangles gris représentent les périodes nocturnes et les courbes de tendance sont obtenues par lissage spline (paramètre de lissage=0,1). Les données sont représentées sous la forme moyenne +/- sem. <sup>a</sup> indique une différence significative entre les deux génotypes et <sup>b</sup> indique une différence significative par rapport au temps t=0h (anova puis, lorsque l'effet d'un facteur est significatif, tests de Student et correction de Benjamini-Hochberg pour contrôler le FDR à 5%). HMG : 3-hydroxy-3-méthylglutaryl.

conçoit aisément qu'il faille au moins 3 points de mesures (dans notre cas 3 temps de jeûne) pour observer un changement de signe d'une dérivée, avec 11 points de mesure, nous ne pouvons en l'occurrence nous permettre d'observer plus de 9 changements de signe de la dérivée, le profil correspondant oscillant déjà de manière importante. Si l'on prend en plus en considération le caractère bruité de la méthode de mesure, le nombre limité d'échantillons par lot et, dans notre cas, la proximité des points de mesure du début de l'épreuve de jeûne, il semble raisonnable de se limiter à un nombre inférieur de changements de signes de la dérivée. Dans un article récemment publié (Dejean et al., 2007) et présenté en [annexe 4](#), nous proposons une méthode basée notamment sur ces réflexions et sur la réalisation d'analyses en composantes principales sur les dérivées des profils lissés pour fixer le paramètre de lissage. Des méthodes de classification sont ensuite appliquées pour classer les dérivées des profils lissées et représenter les résultats. Nous constatons d'ailleurs que la lecture d'une heatmap représentant la classification des dérivées des courbes lissées s'avère bien plus facile à interpréter que celle présentée en [Figure 74](#) (voir l'article en annexe 4). Un autre intérêt de cette méthode qui n'est pas illustré dans (Dejean et al., 2007) est qu'elle permet éventuellement de travailler sur les différences de dérivées entre souris de type sauvage et souris de type PPAR $\alpha$ -/. En réalisant ces analyses sur nos données (résultats non présentés dans le cadre de ce manuscrit), nous avons constaté que certains gènes comme L-FABP ou la catalase (groupe 4) mais également l'acyl-CoA thioestérase cytosolique ACOTH (groupe 1) se retrouvaient correctement classés parmi les gènes cibles connus de PPAR $\alpha$  (comme AOX ou mHMGCoAs par exemple). Malgré des profils d'expression très différents pour l'AOX ([Figure 77](#)) et pour la L-FABP ([Figure 81](#)) par exemple, la différence des dérivées du profil des souris PPAR $\alpha$ -/ et du profil des souris de type sauvage est à peu près identique pour ces deux gènes car les modulations s'opèrent à des temps identiques et dans des amplitudes comparables, indiquant très vraisemblablement une intervention du récepteur PPAR $\alpha$  dans les deux cas (même si dans le cas de L-FABP, un signal inhibiteur de l'expression joue manifestement également un rôle). Ces travaux préliminaires se poursuivent actuellement et le développement d'un package R pour l'analyse de données acquises en cinétique est envisagé.

## VI.4. DISCUSSION

*Une épreuve de jeûne schématiquement séparée en deux phases :*

Cette étude nous a permis d'acquérir une vision assez large et surtout dynamique des modulations de paramètres biochimiques et d'expression hépatique d'un panel de gènes au cours d'une épreuve de jeûne longue pour un organisme comme la souris. Bien que la mise en place de la plupart des adaptations à la privation d'aliment s'opère de manière continue, il nous semble que notre épreuve de jeûne peut être schématiquement séparée en deux phases à partir des données qui ont été recueillies. La première phase correspond au début du jeûne et s'étend à peu près jusqu'à 12h de jeûne. Une transition s'opère vers la deuxième phase entre 12h et 24h de jeûne puis la deuxième phase s'étend de 24h à 72h de jeûne.

Lors de la première phase, la perte de masse corporelle est rapide ([Figure 63A](#)) mais semble concerner de manière comparable des organes comme le foie ([Figure 63B](#)) ou les tissus adipeux blancs ([Figure 63C-D](#)). Le glucose semble constituer une source d'énergie majeure lors de cette phase initiale : il s'épuise rapidement ([Figure 64A](#)), conduisant à l'induction des enzymes de la néoglucogenèse ([Figure 76B](#)), vraisemblablement principalement sous l'effet combiné de la chute d'insuline et de l'augmentation du glucagon. Les protéines sont peu utilisées lors de cette phase initiale ([Figure 64B](#)) et les acides gras ne s'accumulent pas encore significativement au niveau hépatique ([Figure 65B](#)), probablement en raison d'un niveau de  $\beta$ -oxydation suffisant chez les souris de type sauvage ([Figure 77](#)). Les niveaux constitutivement plus faibles d'expression des gènes de la  $\beta$ -oxydation peroxysomale chez les souris PPAR $\alpha$ -/- ([Figure 77](#)) conduisent vraisemblablement, vers la fin de cette première phase, à une augmentation significative de la teneur hépatique en acides gras chez ces souris ([Figure 65B](#)). A la fin de cette première phase, les corps cétoniques ne sont pas encore augmentés ([Figure 64C](#)), probablement en raison d'une induction encore limitée de l'HMG-CoA synthase mitochondriale ([Figure 78D](#)). La composition en acides gras du TABe évolue peu lors de cette première phase, indiquant une mobilisation non spécifique des AG à ce stade du jeûne. En revanche, la composition hépatique en AG est nettement modifiée et de manière à peu près identique pour les deux génotypes ([Figure 70A-B](#)). Les principales modifications concernent 1) les acides gras polyinsaturés des familles  $\omega$ 3 et surtout  $\omega$ 6 dont les précurseurs essentiels s'accumulent (l'acide linoléique, [Figure 73A](#) et dans une moindre mesure l'acide  $\alpha$ -linolénique, [Figure 73D](#)) au détriment des acides gras plus longs et plus insaturés ([Figure 73B-C](#)) et 2) les acides gras saturés dont la longueur de chaîne diminue ([Figure 71](#)). Ces effets pourraient être due à la combinaison de multiples phénomènes dont

certains, mais pas tous, ont été observés dans notre expérience : 1) un captage sélectif par le foie de certains acides gras en provenance du tissu adipeux (non étudié dans notre étude en raison notamment du manque de données disponibles dans la littérature sur les spécificités de substrats de différents transporteurs hépatiques d'acides gras), 2) une inhibition de l'expression ([Figure 82](#)) et de l'activité des gènes de la lipogénèse, vraisemblablement due en bonne partie à une chute de l'activité des facteurs de transcription SREBP en raison du début de l'accumulation des acides gras dans le foie et de la chute de l'insulinémie et 3) une régulation différentielle de l'expression de gènes du catabolisme ([Figure 79](#)) ou de la biosynthèse ([Figure 82D](#)) des acides gras présentant des spécificités de substrats (qui sont encore peu décrites dans la littérature). Nous envisageons d'approfondir ce dernier point en étudiant notamment les spécificités de substrats des différentes enoyl-CoA isomérasées et en mesurant avec une meilleure précision les profils d'expression des  $\Delta$ -désaturases et des élongases afin de mieux comprendre les déterminants clefs de ces modifications importantes de la composition hépatique en AG.

Entre 12h et 24h de jeûne, on assiste à une phase transition de la plupart des paramètres mesurés. De nombreux paramètres fluctuent de manière importante au cours de cette transition comme la perte de masse corporelle ([Figure 63A](#)), la glycémie ([Figure 64A](#)), l'urée et le  $\beta$ -hydroxybutyrate plasmatiques qui amorcent leurs augmentations ([Figure 64B-C](#)), la teneur hépatique en AG qui commence également à augmenter de manière significative chez les souris de type sauvage ([Figure 65B](#)), la composition du TABe en AG qui commence à connaître quelques modifications sensibles ([Figure 66](#), [Figure 68A et D](#) et [Figure 69C](#)) et la composition hépatique en AG qui connaît la transition entre ses deux phases de modulation que nous avons décrites en détail dans la partie résultats ([Figure 70](#)). Cette phase de transition correspond également à la mise en place d'un nouveau « programme transcriptionnel » au niveau hépatique comme en témoigne la parfaite séparation, pour les deux génotypes, des profils d'expression correspondant à des durées de jeûne inférieures à 12h d'une part et supérieures à 18h d'autre part ([Figure 74](#)). C'est également au cours de cette transition, que l'activité maximale du récepteur PPAR $\alpha$  est mise en place comme l'illustre la séparation moins nette des profils transcriptionnels « avant 12h » et « après 18h » chez les souris déficientes pour ce récepteur ([Figure 74](#)) et les nombreuses régulations de gènes cibles observées individuellement ([Figure 77, 78, 79 et 80](#)). D'autres voies de signalisation participent néanmoins aussi à assurer cette transition du « programme transcriptionnel hépatique » car les souris PPAR $\alpha$  présentent malgré tout une séparation assez nette de leurs

profils d'expression hépatiques « avant 12h » et « après 18h » (Figure 74). Parmi les voies mises en jeu, la diminution de l'activité des SREBP affecte très probablement, chez les souris des deux génotypes, l'expression de nombreux gènes impliqués dans la biosynthèse des AG (Figure 82) et du cholestérol (Figure 83). Ces régulateurs transcriptionnels et/ou d'autres facteurs de transcription affectent aussi très vraisemblablement l'expression de nombreux gènes cibles de PPAR $\alpha$  dans le sens d'une diminution de l'expression comme en témoigne les inductions limitées (Figures 77, 78 et 79) voire absentes (Figure 81) de plusieurs de ces gènes chez les souris de type sauvage, associées à des diminutions fréquentes et parfois d'ampleur conséquente (Figure 78D, 79A et 81 notamment) de l'expression de ces mêmes gènes observées chez les souris PPAR $\alpha$ -/. La voie de signalisation de CAR, activée au cours du jeûne dans le foie (Maglich et al., 2004) et qui présente la capacité de réguler négativement un certain nombre de gènes cibles de PPAR $\alpha$  lorsqu'elle est activée (Figure 57E, données additionnelles non présentées et Ueda et al., 2002) pourrait constituer un candidat intéressant à évaluer lors d'investigations futures.

La seconde phase du jeûne qui s'amorce alors se caractérise par une perte pondérale plus lente des animaux (Figure 63A), probablement, en partie, en raison d'une diminution de leur métabolisme basal (Kersten et al., 1999), notamment sous l'effet de la chute des hormones thyroïdiennes (Maglich et al., 2004), accompagnée d'une diminution progressive de l'activité locomotrice des souris (données non présentées dans ce manuscrit mais paramètre mesuré dans une étude indépendante réalisée sur 9 animaux par génotype) et d'une chute de leur température corporelle (Kersten et al., 1999). Les tissus adipeux blancs (TABs) dont la masse relative chute nettement au cours de cette deuxième phase du jeûne (Figure 63C-D) sont fortement impliqués dans la perte de masse corporelle à cette étape du jeûne ce qui suggère, à ce stade, une mobilisation tissu-spécifique des réserves énergétiques au sein de l'organisme par le biais de la lipolyse des TABs. Cette lipolyse au cours de la deuxième phase de jeûne semble affecter de manière différentielle, et indépendante du génotype, les différents acides gras du TABe (Figure 66). Cependant, il est également possible que des modulations biochimiques et/ou transcriptionnelles au sein du TABe et affectant des enzymes du métabolisme des AG à spécificités de substrats marquées expliquent aussi en partie les modifications de la composition du TABe au cours du jeûne. L'acide palmitique (Figure 67A), le C16:1 $\omega$ 7 (Figure 68B) et l'acide  $\alpha$ -linolénique (Figure 69C) semblent notamment mobilisés de manière préférentielle du TABe alors que l'acide stéarique (Figure 67B), ses principaux produits de désaturation et d'elongation (C18:1 $\omega$ 9 et C20:1 $\omega$ 9, Figure 68C-D) et

les acides gras longs polyinsaturés (Figure 69B, D et E) semblent au contraire être mobilisés dans une moindre mesure du TABe et voir ainsi leurs teneurs respectives augmenter au sein de ce tissu. Sous l'effet probable d'un influx d'AG important (Figure 65B), essentiellement lors des phases nocturnes, la masse relative du foie tend à retrouver sa valeur initiale (Figure 63B) au cours de cette phase du jeûne. L'accumulation massive d'AG dans le foie des souris PPAR $\alpha$ -/- (Figure 65B) ne semble pas due à une incapacité à réguler négativement les gènes de la lipogénèse (Figure 82) mais à l'absence quasi-totale d'induction ou de maintien de l'expression des gènes du catabolisme des acides gras chez ces souris (Figures 77, 78, 79 et 80). La composition du foie en AG connaît à ce stade une nouvelle étape de modification caractérisée essentiellement par l'accumulation préférentielle des acides gras monoinsaturés C16:1 $\omega$ 9 et C20:1 $\omega$ 9 (Figure 72A et C) et un retour progressif du pourcentage d'acide  $\alpha$ -linolénique à son niveau initial (Figure 73D). Ces modifications de la composition hépatique en lipides sont cependant nettement moins importantes quantitativement que celles qui s'opèrent lors de la première phase du jeûne comme en témoigne leur plus forte corrélation avec le deuxième axe de l'ACP qui n'explique que 18% de la variance totale (contre plus de 56% pour le premier axe, Figure 70A-B). En termes d'utilisation des substrats énergétiques, cette phase de jeûne est manifestement caractérisée, chez les souris de type sauvage, par une utilisation accrue des lipides et des corps cétoniques (Figure 64C), ces derniers étant produits par le foie grâce à l'induction, généralement maximale à cette étape, des gènes du catabolisme des lipides (Figures 77, 78, 79 et 80) et de la cétogenèse (Figure 78D). A ce stade, on note également une légère augmentation de l'urée plasmatique chez les souris de type sauvage, surtout après 48h de jeûne (Figure 64B), signant vraisemblablement un recours accru au catabolisme des protéines pour maintenir la glycémie qui atteint alors ses niveaux les plus bas (Figure 64A). Chez les souris PPAR $\alpha$ -/-, en l'absence d'induction des enzymes du catabolisme des acides gras et de la cétogenèse (Figures 77 à 80), les corps cétoniques ne peuvent servir de substrat aux tissus qui peuvent les utiliser (Figure 64). Cela impose à ces souris un recours massif au catabolisme des protéines se traduisant par une augmentation considérable de la concentration plasmatique d'urée (Figure 64B) qui s'accompagne d'une stabilisation de la glycémie à partir de 18h de jeûne (Figure 64A). Une étude plus poussée des niveaux d'expression de gènes impliqués dans le catabolisme des protéines au niveau musculaire et dans le cycle de l'urée au niveau hépatique nous semble nécessaire pour mieux comprendre les mécanismes précis impliqués dans ces différences majeures d'utilisation des substrat par les souris des deux génotypes au cours de cette deuxième phase du jeûne. Enfin,

au cours de cette deuxième phase du jeûne, l'expression hépatique des gènes étudiés se stabilise pour la plupart d'entre eux ([Figures 74 à 83](#)). Beaucoup de gènes connaissent, lors des deux derniers temps de jeûne de notre étude ( $t=60\text{h}$  et  $72\text{h}$ , [Figure 74](#)), une légère augmentation de leur expression chez les souris des deux génotypes. Si nous ne pouvons exclure à l'heure actuelle qu'il s'agisse d'un artéfact de nos mesures d'expression par puces à ADN (une éventuelle incapacité de la normalisation par les spikes à corriger un biais d'intensité trop fort peut être envisagée par exemple), il est également possible qu'il s'agisse là d'une nouvelle phase de transition vers une troisième phase de jeûne. Il est probable que des confirmations d'un certain nombre de modulations observées dans cette étude par PCR en temps réel nous permettront de fournir prochainement une réponse à ces hypothèses.

Pensant initialement que des modifications importantes allaient se produire au cours des premières heures de jeûne, nous avions opté pour un échantillonnage plus resserré (intervalle de temps de 3h) au cours des 12 premières heures de jeûne. S'il est vrai que sur un certain nombre de paramètres comme la composition hépatique en lipides par exemple, cet échantillonnage s'avère parfaitement adapté, la réalisation de nos futures épreuves de jeûne destinées à l'étude d'effets transcriptionnels et réalisées dans des conditions similaires (horaire de début de jeûne en particulier), se fera sans doute avec un échantillonnage plus resserré entre 12 et 24h de jeûne qui nous apparaît, à l'issue de cette étude, comme la principale phase de transition au niveau transcriptionnel mais également pour d'autres paramètres comme nous venons de le discuter. Nous avons constaté dans nos mesures qu'un rythme circadien semblait parfois se superposer aux effets du jeûne. C'est sans doute sur l'accumulation des lipides au niveau hépatique que cet effet est le plus net ([Figure 65B](#)) mais ce type d'oscillations est également perçu dans certains graphiques de proportions d'acides gras ([Figure 73 C-F](#) par exemple) ou d'expression de gènes ([Figure 77A-B](#) ou [Figure 80](#) par exemple). Nous constatons de plus, que la période de transition (12h-24h) que nous avons définie, s'étend durant la première période nocturne qui suit la mise à jeun des animaux. Les souris s'alimentant essentiellement la nuit, il est possible que leur organisme réagisse différemment au jeûne s'il est initié de jour (globalement les 12 premières heures de notre épreuve de jeûne) ou de nuit. Nous ne pouvons donc conclure sur le fait que les phases de jeûne que nous avons définies ici sont effectivement valables quelque soit l'heure de la mise à jeun des animaux. Des études complémentaires pourraient permettre de préciser cet aspect et de révéler éventuellement des liens entre la réponse au jeûne et les horloges centrales ou périphériques, avec la difficulté liée au fait que la mise à jeun des animaux de jour nécessite à

priori un entraînement des animaux à s'alimenter uniquement le jour pour s'assurer que les animaux du temps t=0h sont bien en situation alimentée.

#### *Eléments complémentaires sur le phénotype des souris PPAR $\alpha$ -/-*

Cette étude nous a également permis de mieux caractériser le phénotype des souris PPAR $\alpha$ -/-, notamment en ce qui concerne leur profil d'AG dans le tissu adipeux blanc, et de mieux caractériser son rôle dans les régulations d'expression qui s'opèrent au cours du jeûne. Concernant ce deuxième aspect qui a été largement évoqué ci-dessus, nos observations suggèrent en particulier que le rôle de ce récepteur au cours du jeûne consiste autant à assurer l'induction de certains de ses gènes cibles qu'à maintenir à un niveau stable l'expression d'autres de ses gènes cibles en s'opposant vraisemblablement à d'autres voies de signalisation. Des effets spécifiques des promoteurs, notamment de recrutement différentiels de corégulateurs, peuvent également être envisagés pour expliquer l'induction de certains gènes cibles alors que d'autres restent stables mais les régulations négatives de gènes cibles chez les souris PPAR $\alpha$ -/- plaident néanmoins en faveur de l'implication d'autres voies de signalisation. Le jeûne constitue dans ce cadre une situation physiopathologique particulièrement pertinente pour l'étude des interactions entre de nombreuses voies de signalisation en raison des changements hormonaux importants qu'il provoque et de la modulation conséquente des activités de facteurs de transcription qui en résulte.

Concernant les compositions tissulaires en AG, les données de cette étude sur le foie sont en parfait accord avec nos observations récemment publiées (Martin et al., 2007) et montrent en particulier l'accumulation hépatique majeure des précurseurs des AGPI des familles  $\omega$ 3 et  $\omega$ 6 chez les souris PPAR $\alpha$ -/- au détriment des AGPI plus longs et plus insaturés ([Figure 73](#)). La différence de composition en AG du TABe entre les souris des deux génotypes est encore plus marquée ([Figure 66](#)) et la quasi-totalité des courbes de pourcentages d'AG ([Figures 67, 68 et 69](#)) présentent une différence nette de niveau entre les souris des deux génotypes conduisant les courbes à ne presque jamais se croiser en raison de modulations globalement similaires au cours du jeûne chez les souris des deux génotypes. Cette observation suggère que la différence de composition en AG du TABe est le reflet d'altérations spécifiques du métabolisme des AG chez les souris PPAR $\alpha$ -/- (sans doute principalement au niveau du foie) qui, en s'intégrant au fur et à mesure du vieillissement des souris par le biais du stockage d'AG dans le tissu adipeux, conduit à accroître les différences de composition en AG dans ce tissu par rapport au foie dont on peut penser qu'il présente un

turnover plus important des AG au cours du temps. Il est donc probable qu'avec l'âge, la composition en AG du TAB des souris PPAR $\alpha$ -/- soit de plus en plus différente de celle des souris de type sauvage, ce qui mériterait sans doute d'être évalué lors de futures investigations. Plus spécifiquement, les différences de composition du TABe entre les souris des deux génotypes peuvent notamment suggérer les altérations suivantes du métabolisme des AG : 1) un déficit de synthèse de l'acide palmitique ([Figures 67A et 71C](#)), issu de l'action de la fatty acid synthase ([Figure 31](#)) dont nous constatons qu'elle est, en situation d'alimentation, généralement sous-exprimée chez les souris PPAR $\alpha$ -/- par rapport aux souris de type sauvage ([Figure 82A](#) et Martin et al., 2007) et ce, en accord avec des données publiées (Patel et al., 2001). 2) Un déficit de synthèse des AG monoinsaturés ([Figure 68 et 72B, D et E](#)) lié sans doute à une déficience en  $\Delta 9$ -désaturase ([Figure 82C](#) et Martin et al., 2007) dont nous avions étudié précédemment la régulation par un activateur pharmacologique de PPAR $\alpha$  (Guillou et al., 2002) et pour laquelle une déficience chez les souris PPAR $\alpha$ -/- pourrait s'accroître avec l'âge des souris (H. Guillou et PGP Martin, observations non publiées). 3) une déficience dans le catabolisme ([Figure 79B, C et D](#) et Martin et al., 2007) et la conversion des précurseurs essentiels des AGPI des familles  $\omega 3$  et  $\omega 6$ . Cette conversion est assurée par les  $\Delta 6$ - puis  $\Delta 5$ -désaturases et par les élongases. Nous avons précédemment montré que les  $\Delta 6$ - et  $\Delta 5$ -désaturases sont régulées par le fenofibrate (Guillou et al., 2002) et que les souris PPAR $\alpha$ -/- présentent une déficience constitutive d'expression et d'activité (Martin et al., 2007) de ces enzymes limitantes dans la voie de biosynthèse des AGPI. Les données récemment publiées sur les élongases (Wang et al., 2005b; Wang et al., 2006b) indiquent que PPAR $\alpha$  jouerait un rôle dans la régulation de certaines isoformes mais nos mesures effectuées par puce à ADN dans cette étude ([Figure 82D](#) et courbes individuelles non présentées dans ce manuscrit) ne montrent pas d'effet net du génotype sur l'expression constitutive de ces enzymes. Un étude par PCR en temps réel serait néanmoins nécessaire en raison de niveaux d'expression souvent faibles des gènes correspondants.

#### *Vers une identification du ligand endogène du récepteur PPAR $\alpha$ au cours du jeûne ?*

L'intérêt d'une approche en cinétique du jeûne a été bien illustré, je pense, par la mise en évidence des différentes phases d'adaptation au jeûne et l'identification « périodes charnières » de ces adaptations. Par ailleurs, nous avons vu précédemment qu'il était possible d'établir des liens de corrélations entre des variables issues de deux jeux d'observations de nature différente mais obtenues sur les mêmes individus (voir [annexe 3](#)). En théorie au moins,

la disponibilité de mesures réalisées en cinétique offre potentiellement la possibilité supplémentaire (moyennant le développement d'outils d'analyse appropriés) de suggérer des liens de cause à effet entre les modulations de ces deux types de variables par le biais de l'observation de légers décalages dans le temps entre les différentes modulations. Il pourrait ainsi être envisagé d'essayer de déterminer avec plus de précision quels sont les AG (ou les autres métabolites si les données étaient disponibles) qui constituent les meilleurs activateurs endogènes du récepteur PPAR $\alpha$  au cours du jeûne. Nous travaillons donc aujourd'hui à l'évaluation de l'analyse canonique régularisée (peut-être associée à la méthode de lissage des profils proposée dans Dejean et al., 2007) pour répondre à cet enjeu. En parallèle, nous souhaitons également acquérir des signatures transcriptionnelles dans d'autres organes, et notamment le rein qui répond bien à l'activation de PPAR $\alpha$  (Martin et al., 2005), dans le but de vérifier la concomitance dans différents organes des modulations d'expression des gènes cibles de PPAR $\alpha$  au cours du jeûne et démontrer ainsi le caractère circulant de l'activateur (ou des activateurs) endogène(s) de ce RN.

## CONCLUSION

Ce travail de doctorat visait à évaluer le potentiel de régulation génique de PPAR $\alpha$ . L'activation de ce récepteur a donc été étudiée dans différents contextes : pharmacologique (activateurs de PPAR $\alpha$  et de RXR utilisés en thérapeutique humaine : Chapitre IV), nutritionnel (acides gras alimentaires, Chapitre V) et physiopathologique (jeûne, Chapitre VI) par le biais d'une puce à ADN développée pour ce programme de recherche et largement dédiée à PPAR $\alpha$ . Nos travaux ont notamment mis en évidence des rôles de PPAR $\alpha$  comme médiateur central d'effets hépatiques et rénaux des activateurs pharmacologiques de RXR et dans la régulation de gènes déterminant la composition tissulaire en AG (enoyl-CoA isomérasées et  $\Delta$ -désaturases). Ils ont également permis de préciser la période d'intervention du récepteur au cours du jeûne, de mieux apprécier son rôle de régulateur transcriptionnel dans ce contexte et de suggérer son interaction avec d'autres voies de signalisation.

Au delà des résultats et conclusions propres à chaque étude, ces travaux ont engagés notre équipe de recherche dans le domaine de la transcriptomique à moyen et haut-débit. la question qui se pose aujourd'hui est de savoir comment ces outils peuvent servir les objectifs du pôle de Toxicologie alimentaire de Toulouse auquel appartient le laboratoire et comment l'expertise et le réseau de collaborations développé par l'équipe dans le cadre de ce doctorat peuvent être mis au service d'un projet fédérateur répondant à ces objectifs. Dans ce cadre, il est intéressant de constater que de nombreuses molécules, contaminants alimentaires ou environnementaux, empruntent les voies des récepteurs nucléaires pour exercer leurs effets, parfois délétères. Cependant, nous avons vu que la plupart des recherches dans ce domaine étaient conduites à des doses fortes, souvent sur de courtes durées, et avec des composés évalués individuellement et non sous la forme de mélanges auxquels les organismes sont pourtant exposés. Nous avons apporté la démonstration que les outils de transcriptomique, associés en particulier à ceux de l'analyse statistique multidimensionnelle, peuvent détecter des effets liées à de faibles doses d'exposition. Il nous semble donc qu'un projet d'envergure peut maintenant être envisagé répondant en particulier à l'objectif d'études plus en rapport avec les conditions réelles d'exposition des organismes.

Un tel projet pourrait se concentrer initialement sur le foie qui constitue un organe fortement exposé aux contaminants absorbés par voie orale et joue des rôles majeurs à la fois dans le métabolisme de ces contaminants mais également dans le métabolisme énergétique et dans celui de nombreuses hormones endogènes, constituant par là un organe clef des

phénomènes de disruption endocrinienne. L'idée centrale d'un tel projet serait de constituer, grâce aux outils de transcriptomique, une base de données des signatures transcriptionnelles de l'activation des différents récepteurs nucléaires impliqués dans la physiopathologie hépatique. Il existe aujourd'hui pour la plupart des récepteurs nucléaires des modèles murins d'invalidation et des agonistes synthétiques spécifiques et de haute affinité. Ces modèles, combinés à la réalisation de criblages transcriptomiques pangénomiques, peuvent permettre d'acquérir une somme d'information suffisante permettant de définir avec précision la signature d'activation des différents récepteurs. Ces informations, éventuellement combinées à d'autres sources d'informations provenant des bases de données ou de la bibliographie, peuvent être utilisées pour construire un modèle de prédiction de l'activité des récepteurs nucléaires, une donnée qu'il est actuellement impossible d'évaluer de manière globale *in vivo*, par exemple en utilisant une approche par PLS (Boulesteix and Strimmer, 2005). De nouvelles molécules d'intérêt, voire des mélanges de molécules, pourraient alors être ciblés sur modèles murins et l'utilisation de ce modèle pourrait permettre de prédire le ou les récepteurs nucléaires qui sont affectés ainsi que l'ampleur de leur activation.

Naturellement, un tel projet ne peut être conduit qu'en collaboration avec plusieurs laboratoires dans le cadre d'un programme de recherches de taille conséquente, comme un projet européen par exemple. Un tel modèle constituerait un outil particulièrement pertinent pour le criblage de ligands des récepteurs nucléaires et l'évaluation vis-à-vis de ces récepteurs nucléaires d'un choix raisonné de composés entrants dans le cadre du programme européen REACH par exemple. De plus, une somme de connaissances nouvelles serait également acquise sur chacune des voies de signalisation étudiées (identification de nouvelles cibles des récepteurs nucléaires ou d'éventuelles interactions entre ces différentes voies de signalisation par exemple) qui pourrait alors faire l'objet de sous-projets visant à évaluer les conséquences fonctionnelles des modulations transcriptionnelles observées. L'utilisation d'outils visant à réconcilier différents niveaux d'observations (par exemple étude des corrélations entre données de transcriptomique et évaluations de la toxicité par la mesure de divers paramètres) pourrait dans ce cadre permettre de faire le lien entre une recherche cognitive sur les récepteurs nucléaires et une recherche appliquée visant à définir les niveaux de risque associés à l'exposition à diverses substances. Entre ce qui pourrait être expliqué (la variation d'expression d'un groupe de gènes explique tel aspect de la toxicité de la molécule par exemple) et ce qui ne le pourrait pas (en raison principalement des limites naturelles d'investigations de tout projet de recherche), le modèle pourrait petit à petit s'enrichir

d'informations non redondantes et améliorer ainsi sa précision au fur et à mesure de l'acquisition de nouvelles connaissances.

Dans la perspective d'un tel projet, les compétences et le réseau de collaborations qui ont été développés dans le cadre de ce doctorat apparaissent pertinents et pourraient être mis au service de tels objectifs. Enfin, à titre de première étape, une preuve de concept pourrait être apportée en commençant de telles investigations sur modèles cellulaires, par exemple dans le cadre d'un projet soutenu par l'Agence Nationale de la Recherche. L'arrivée récente dans notre équipe de Laila Lakhal nous a permis d'initier la mise en place de modèles cellulaires dans le but de réaliser des démonstrations des mécanismes impliqués dans les modulations géniques que nous observons. L'arrivée prochaine d'un nouveau CR2 disposant de compétences en biochimie et en physiologie pourrait permettre d'accroître notre capacité à faire le lien entre modulations transcriptionnelles et conséquences physiopathologiques. Ces éléments constituent autant de garanties de la faisabilité d'un tel projet qui constituerait à la fois une excellente prolongation de ce travail de doctorat et un moyen pour l'équipe, le laboratoire et le pôle de toxicologie alimentaire de l'INRA d'élargir le champ de ses investigations et de ses collaborations.

## BIBLIOGRAPHIE

### Nombre de références bibliographiques : 1128 références

- Abu-Abed, S., Dolle, P., Metzger, D., Beckett, B., Chambon, P., and Petkovich, M. (2001). The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev* 15, 226-240.
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185-2195.
- Adelman, G., Begue, A., Stehelin, D., and Laudet, V. (1996). A functional Rev-erb alpha responsive element located in the human Rev-erb alpha promoter mediates a repressing activity. *Proc Natl Acad Sci U S A* 93, 3553-3558.
- Agarwal, M. K. (1996). The antiglucocorticoid action of mifepristone. *Pharmacol Ther* 70, 183-213.
- Akingbemi, B. T., Ge, R., Klinefelter, G. R., Zirkin, B. R., and Hardy, M. P. (2004). Phthalate-induced Leydig cell hyperplasia is associated with multiple endocrine disturbances. *Proc Natl Acad Sci U S A* 101, 775-780.
- Al-Shahrour, F., Diaz-Uriarte, R., and Dopazo, J. (2004). Fatigo: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* 20, 578-580.
- Aldridge, T. C., Tugwood, J. D., and Green, S. (1995). Identification and characterization of DNA elements implicated in the regulation of CYP4A1 transcription. *Biochem J* 306 ( Pt 2), 473-479.
- Allenby, G., Bocquel, M. T., Saunders, M., Kazmer, S., Speck, J., Rosenberger, M., Lovey, A., Kastner, P., Grippo, J. F., Chambon, P., and et al. (1993). Retinoic acid receptors and retinoid X receptors: interactions with endogenous retinoic acids. *Proc Natl Acad Sci U S A* 90, 30-34.
- Alvares, K., Fan, C., Dadras, S. S., Yeldandi, A. V., Rachubinski, R. A., Capone, J. P., Subramani, S., Iannaccone, P. M., Rao, M. S., and Reddy, J. K. (1994). An upstream region of the enoyl-coenzyme A hydratase/3-hydroxyacyl-coenzyme A dehydrogenase gene directs luciferase expression in liver in response to peroxisome proliferators in transgenic mice. *Cancer Res* 54, 2303-2306.
- Amri, E. Z., Bonino, F., Ailhaud, G., Abumrad, N. A., and Grimaldi, P. A. (1995). Cloning of a protein that mediates transcriptional effects of fatty acids in preadipocytes. Homology to peroxisome proliferator-activated receptors. *J Biol Chem* 270, 2367-2371.
- Ananthanarayanan, M., Balasubramanian, N., Makishima, M., Mangelsdorf, D. J., and Suchy, F. J. (2001). Human bile salt export pump promoter is transactivated by the farnesoid X receptor/bile acid receptor. *J Biol Chem* 276, 28857-28865.
- Ananthanarayanan, M., Li, S., Balasubramanyan, N., Suchy, F. J., and Walsh, M. J. (2004). Ligand-dependent activation of the farnesoid X-receptor directs arginine methylation of histone H3 by CARM1. *J Biol Chem* 279, 54348-54357.
- Ando, M., Kaneko, T., Watanabe, R., Kikuchi, S., Goto, T., Iida, T., Hishinuma, T., Mano, N., and Goto, J. (2006). High sensitive analysis of rat serum bile acids by liquid chromatography/electrospray ionization tandem mass spectrometry. *J Pharm Biomed Anal* 40, 1179-1186.
- Angelin, B., Bjokhem, I., and Einarsson, K. (1978). Individual serum bile acid concentrations in normo- and hyperlipoproteinemia as determined by mass fragmentography: relation to bile acid pool size. *J Lipid Res* 19, 527-537.
- Antonenkov, V. D., Sormunen, R. T., Ohlmeier, S., Amery, L., Fransen, M., Mannaerts, G. P., and Hiltunen, J. K. (2006). Localization of a portion of the liver isoform of fatty-acid-binding protein (L-FABP) to peroxisomes. *Biochem J* 394, 475-484.
- Antonson, P., Schuster, G. U., Wang, L., Rozell, B., Holter, E., Flodby, P., Treuter, E., Holmgren, L., and Gustafsson, J. A. (2003). Inactivation of the nuclear receptor coactivator RAP250 in mice results in placental vascular dysfunction. *Mol Cell Biol* 23, 1260-1268.
- Aoyama, T., Peters, J. M., Iritani, N., Nakajima, T., Furihata, K., Hashimoto, T., and Gonzalez, F. J. (1998). Altered constitutive expression of fatty acid-metabolizing enzymes in mice lacking the peroxisome proliferator-activated receptor alpha (PPARalpha). *J Biol Chem* 273, 5678-5684.
- Arakawa, H., Takiguchi, M., Amaya, Y., Nagata, S., Hayashi, H., and Mori, M. (1987). cDNA-derived amino acid sequence of rat mitochondrial 3-oxoacyl-CoA thiolase with no transient presequence: structural relationship with peroxisomal isozyme. *Embo J* 6, 1361-1366.
- Ariyasu, H., Takaya, K., Hosoda, H., Iwakura, H., Ebihara, K., Mori, K., Ogawa, Y., Hosoda, K., Akamizu, T., Kojima, M., et al. (2002). Delayed short-term secretory regulation of ghrelin in obese animals: evidenced by a specific RIA for the active form of ghrelin. *Endocrinology* 143, 3341-3350.

- Ariyasu, H., Takaya, K., Tagami, T., Ogawa, Y., Hosoda, K., Akamizu, T., Suda, M., Koh, T., Natsui, K., Toyooka, S., et al. (2001). Stomach is a major source of circulating ghrelin, and feeding state determines plasma ghrelin-like immunoreactivity levels in humans. *J Clin Endocrinol Metab* 86, 4753-4758.
- Armstrong, M. B., and Towle, H. C. (2001). Polyunsaturated fatty acids stimulate hepatic UCP-2 expression via a PPARalpha-mediated pathway. *Am J Physiol Endocrinol Metab* 281, E1197-1204.
- Arnhold, T., Tzimas, G., Wittfoht, W., Plonait, S., and Nau, H. (1996). Identification of 9-cis-retinoic acid, 9,13-di-cis-retinoic acid, and 14-hydroxy-4,14-retro-retinol in human plasma after liver consumption. *Life Sci* 59, PL169-177.
- Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B. L., Housman, D. E., and Evans, R. M. (1987). Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocorticoid receptor. *Science* 237, 268-275.
- Aschrafi, A., Meindl, N., Firla, B., Brandes, R. P., and Steinhilber, D. (2006). Intracellular localization of RORalpha is isoform and cell line-dependent. *Biochim Biophys Acta*.
- Atherton, H. J., Bailey, N. J., Zhang, W., Taylor, J., Major, H., Shockcor, J., Clarke, K., and Griffin, J. L. (2006). A combined 1H-NMR spectroscopy- and mass spectrometry-based metabolomic study of the PPAR-alpha null mutant mouse defines profound systemic changes in metabolism linked to the metabolic syndrome. *Physiol Genomics* 27, 178-186.
- Aubert, J., Bar-Hen, A., Daudin, J. J., and Robin, S. (2004). Determination of the differentially expressed genes in microarray experiments using local FDR. *BMC Bioinformatics* 5, 125.
- Auboeuf, D., Rieusset, J., Fajas, L., Vallier, P., Frerling, V., Riou, J. P., Staels, B., Auwerx, J., Laville, M., and Vidal, H. (1997). Tissue distribution and quantification of the expression of mRNAs of peroxisome proliferator-activated receptors and liver X receptor-alpha in humans: no alteration in adipose tissue of obese and NIDDM patients. *Diabetes* 46, 1319-1327.
- Azer, S. A., Klaassen, C. D., and Stacey, N. H. (1997). Biochemical assay of serum bile acids: methods and applications. *Br J Biomed Sci* 54, 118-132.
- Baccini, A., Besse, P., Déjean, S., Martin, P. G. P., Robert-Granié, C., and San Cristobal, M. (2005). Stratégies pour l'analyse statistique de données transcriptomiques. *J Soc Fr Statistique* 146, 5-44.
- Bachem, C. W., Oomen, R. J., and Visser, R. G. (1998). Transcript imaging with cDNA-AFLP: A step-by-step protocol. *Plant Mol Biol Rep* 16, 157-173.
- Bachem, C. W., van der Hoeven, R. S., de Bruijn, S. M., Vreugdenhil, D., Zabeau, M., and Visser, R. G. (1996). Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J* 9, 745-753.
- Baes, M., Castelein, H., Desmet, L., and Declercq, P. E. (1995). Antagonism of COUP-TF and PPAR alpha/RXR alpha on the activation of the malic enzyme gene promoter: modulation by 9-cis RA. *Biochem Biophys Res Commun* 215, 338-345.
- Baes, M., Huyghe, S., Carmeliet, P., Declercq, P. E., Collen, D., Mannaerts, G. P., and Van Veldhoven, P. P. (2000). Inactivation of the peroxisomal multifunctional protein-2 in mice impedes the degradation of not only 2-methyl-branched fatty acids and bile acid intermediates but also of very long chain fatty acids. *J Biol Chem* 275, 16329-16336.
- Bain, D. L., Franden, M. A., McManaman, J. L., Takimoto, G. S., and Horwitz, K. B. (2000). The N-terminal region of the human progesterone A-receptor. Structural analysis and the influence of the DNA binding domain. *J Biol Chem* 275, 7313-7320.
- Baker, A. R., McDonnell, D. P., Hughes, M., Crisp, T. M., Mangelsdorf, D. J., Haussler, M. R., Pike, J. W., Shine, J., and O'Malley, B. W. (1988). Cloning and expression of full-length cDNA encoding human vitamin D receptor. *Proc Natl Acad Sci U S A* 85, 3294-3298.
- Bandsma, R. H., Van Dijk, T. H., Harmsel At, A., Kok, T., Reijngoud, D. J., Staels, B., and Kuipers, F. (2004). Hepatic de novo synthesis of glucose 6-phosphate is not affected in peroxisome proliferator-activated receptor alpha-deficient mice but is preferentially directed toward hepatic glycogen stores after a short term fast. *J Biol Chem* 279, 8930-8937.
- Barbera, M. J., Schluter, A., Pedraza, N., Iglesias, R., Villarroya, F., and Giralt, M. (2001). Peroxisome proliferator-activated receptor alpha activates transcription of the brown fat uncoupling protein-1 gene. A link between regulation of the thermogenic and lipid oxidation pathways in the brown fat cell. *J Biol Chem* 276, 1486-1493.
- Barbier, O., Duran-Sandoval, D., Pineda-Torra, I., Kosykh, V., Fruchart, J. C., and Staels, B. (2003a). Peroxisome proliferator-activated receptor alpha induces hepatic expression of the human bile acid glucuronidating UDP-glucuronosyltransferase 2B4 enzyme. *J Biol Chem* 278, 32852-32860.
- Barbier, O., Fontaine, C., Fruchart, J. C., and Staels, B. (2004). Genomic and non-genomic interactions of PPARalpha with xenobiotic-metabolizing enzymes. *Trends Endocrinol Metab* 15, 324-330.

- Barbier, O., Villeneuve, L., Bocher, V., Fontaine, C., Torra, I. P., Duhem, C., Kosykh, V., Fruchart, J. C., Guillemette, C., and Staels, B. (2003b). The UDP-glucuronosyltransferase 1A9 enzyme is a peroxisome proliferator-activated receptor alpha and gamma target gene. *J Biol Chem* 278, 13975-13983.
- Barclay, T. B., Peters, J. M., Sewer, M. B., Ferrari, L., Gonzalez, F. J., and Morgan, E. T. (1999). Modulation of cytochrome P-450 gene expression in endotoxemic mice is tissue specific and peroxisome proliferator-activated receptor-alpha dependent. *J Pharmacol Exp Ther* 290, 1250-1257.
- Bardot, O., Aldridge, T. C., Latruffe, N., and Green, S. (1993). PPAR-RXR heterodimer activates a peroxisome proliferator response element upstream of the bifunctional enzyme gene. *Biochem Biophys Res Commun* 192, 37-45.
- Bardot, O., Clemencet, M. C., Passilly, P., and Latruffe, N. (1995). The analysis of modified peroxisome proliferator responsive elements of the peroxisomal bifunctional enzyme in transfected HepG2 cells reveals two regulatory motifs. *FEBS Lett* 360, 183-186.
- Barger, P. M., Browning, A. C., Garner, A. N., and Kelly, D. P. (2001). p38 mitogen-activated protein kinase activates peroxisome proliferator-activated receptor alpha: a potential role in the cardiac metabolic stress response. *J Biol Chem* 276, 44495-44501.
- Barish, G. D., Narkar, V. A., and Evans, R. M. (2006). PPAR delta: a dagger in the heart of the metabolic syndrome. *J Clin Invest* 116, 590-597.
- Barrera, M. J., Camarero, N., Marrero, P. F., and Haro, D. (2003). Control of human carnitine palmitoyltransferase II gene transcription by peroxisome proliferator-activated receptor through a partially conserved peroxisome proliferator-responsive element. *Biochem J* 369, 721-729.
- Bauer, E. R., Daxenberger, A., Petri, T., Sauerwein, H., and Meyer, H. H. (2000). Characterisation of the affinity of different anabolics and synthetic hormones to the human androgen receptor, human sex hormone binding globulin and to the bovine progestin receptor. *Apmis* 108, 838-846.
- Baumann, H., Paulsen, K., Kovacs, H., Berglund, H., Wright, A. P., Gustafsson, J. A., and Hard, T. (1993). Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry* 32, 13463-13471.
- Bavner, A., Sanyal, S., Gustafsson, J. A., and Treuter, E. (2005). Transcriptional corepression by SHP: molecular mechanisms and physiological consequences. *Trends Endocrinol Metab* 16, 478-488.
- Beamer, B. A., Negri, C., Yen, C. J., Gavrilova, O., Rumberger, J. M., Durcan, M. J., Yarnall, D. P., Hawkins, A. L., Griffin, C. A., Burns, D. K., et al. (1997). Chromosomal localization and partial genomic structure of the human peroxisome proliferator activated receptor-gamma (hPPAR gamma) gene. *Biochem Biophys Res Commun* 233, 756-759.
- Beard, R. L., Chandraratna, R. A., Colon, D. F., Gillett, S. J., Henry, E., Marler, D. K., Song, T., Denys, L., Garst, M. E., Arefieg, T., and et al. (1995). Synthesis and structure-activity relationships of stilbene retinoid analogs substituted with heteroaromatic carboxylic acids. *J Med Chem* 38, 2820-2829.
- Beissbarth, T., and Speed, T. P. (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 20, 1464-1465.
- Belakavadi, M., and Fondell, J. D. (2006). Role of the mediator complex in nuclear hormone receptor signaling. *Rev Physiol Biochem Pharmacol* 156, 23-43.
- Belandia, B., Orford, R. L., Hurst, H. C., and Parker, M. G. (2002). Targeting of SWI/SNF chromatin remodelling complexes to estrogen-responsive genes. *Embo J* 21, 4094-4103.
- Bell-Parikh, L. C., Ide, T., Lawson, J. A., McNamara, P., Reilly, M., and FitzGerald, G. A. (2003). Biosynthesis of 15-deoxy-delta12,14-PGJ2 and the ligation of PPARgamma. *J Clin Invest* 112, 945-955.
- Benecke, A., Chambon, P., and Gronemeyer, H. (2000). Synergy between estrogen receptor alpha activation functions AF1 and AF2 mediated by transcription intermediary factor TIF2. *EMBO Rep* 1, 151-157.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* 57, 289-300.
- Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29, 1165-1188.
- Berger, J., Truppe, C., Neumann, H., and Forss-Petter, S. (1998). cDNA cloning and mRNA distribution of a mouse very long-chain acyl-CoA synthetase. *FEBS Lett* 425, 305-309.
- Bernard, K., Auphan, N., Granjeaud, S., Victorero, G., Schmitt-Verhulst, A. M., Jordan, B. R., and Nguyen, C. (1996). Multiplex messenger assay: simultaneous, quantitative measurement of expression of many genes in the context of T cell activation. *Nucleic Acids Res* 24, 1435-1442.
- Bertilsson, G., Heidrich, J., Svensson, K., Asman, M., Jendeberg, L., Sydow-Backman, M., Ohlsson, R., Postlind, H., Blomquist, P., and Berkenstam, A. (1998). Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. *Proc Natl Acad Sci U S A* 95, 12208-12213.
- Bertolotti, M., Concari, M., Loria, P., Abate, N., Pinetti, A., Guicciardi, M. E., and Carulli, N. (1995). Effects of different phenotypes of hyperlipoproteinemia and of treatment with fibric acid derivatives on the rates of cholesterol 7 alpha-hydroxylation in humans. *Arterioscler Thromb Vasc Biol* 15, 1064-1069.

- Bertrand, S., Brunet, F. G., Escriva, H., Parmentier, G., Laudet, V., and Robinson-Rechavi, M. (2004). Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. *Mol Biol Evol* 21, 1923-1937.
- Bianco, A. C., and Kim, B. W. (2006). Deiodinases: implications of the local control of thyroid hormone action. *J Clin Invest* 116, 2571-2579.
- Birnbaumer, M., Schrader, W. T., and O'Malley, B. W. (1983). Photoaffinity labeling of the chick progesterone receptor proteins. Similar hormone binding domains detected after removal of proteolytic interference. *J Biol Chem* 258, 1637-1644.
- Bitsch, F., Aichholz, R., Kallen, J., Geisse, S., Fournier, B., and Schlaeppi, J. M. (2003). Identification of natural ligands of retinoic acid receptor-related orphan receptor alpha ligand-binding domain expressed in Sf9 cells--a mass spectrometry approach. *Anal Biochem* 323, 139-149.
- Bjorkhem, I., and Diczfalussy, U. (2002). Oxysterols: friends, foes, or just fellow passengers? *Arterioscler Thromb Vasc Biol* 22, 734-742.
- Black, B. E., Holaska, J. M., Rastinejad, F., and Paschal, B. M. (2001). DNA binding domains in diverse nuclear receptors function as nuclear export signals. *Curr Biol* 11, 1749-1758.
- Blanquart, C., Mansouri, R., Paumelle, R., Fruchart, J. C., Staels, B., and Glineur, C. (2004). The protein kinase C signaling pathway regulates a molecular switch between transactivation and transrepression activity of the peroxisome proliferator-activated receptor alpha. *Mol Endocrinol* 18, 1906-1918.
- Blazek, E., Mittler, G., and Meisterernst, M. (2005). The mediator of RNA polymerase II. *Chromosoma* 113, 399-408.
- Bledsoe, R. K., Madauss, K. P., Holt, J. A., Apolito, C. J., Lambert, M. H., Pearce, K. H., Stanley, T. B., Stewart, E. L., Trump, R. P., Willson, T. M., and Williams, S. P. (2005). A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. *J Biol Chem* 280, 31283-31293.
- Bledsoe, R. K., Montana, V. G., Stanley, T. B., Delves, C. J., Apolito, C. J., McKee, D. D., Consler, T. G., Parks, D. J., Stewart, E. L., Willson, T. M., et al. (2002). Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* 110, 93-105.
- Blizzard, T. A., Dininno, F., Morgan, J. D., 2nd, Chen, H. Y., Wu, J. Y., Kim, S., Chan, W., Birzin, E. T., Yang, Y. T., Pai, L. Y., et al. (2005). Estrogen receptor ligands. Part 9: Dihydrobenzoxathiin SERAMs with alkyl substituted pyrrolidine side chains and linkers. *Bioorg Med Chem Lett* 15, 107-113.
- Blumberg, B., Sabbagh, W., Jr., Jugulon, H., Bolado, J., Jr., van Meter, C. M., Ong, E. S., and Evans, R. M. (1998). SXR, a novel steroid and xenobiotic-sensing nuclear receptor. *Genes Dev* 12, 3195-3205.
- Bocquel, M. T., Kumar, V., Stricker, C., Chambon, P., and Gronemeyer, H. (1989). The contribution of the N- and C-terminal regions of steroid receptors to activation of transcription is both receptor and cell-specific. *Nucleic Acids Res* 17, 2581-2595.
- Boehm, M. F., Zhang, L., Badea, B. A., White, S. K., Mais, D. E., Berger, E., Suto, C. M., Goldman, M. E., and Heyman, R. A. (1994). Synthesis and structure-activity relationships of novel retinoid X receptor-selective retinoids. *J Med Chem* 37, 2930-2941.
- Boehm, M. F., Zhang, L., Zhi, L., McClurg, M. R., Berger, E., Wagoner, M., Mais, D. E., Suto, C. M., Davies, J. A., Heyman, R. A., and et al. (1995). Design and synthesis of potent retinoid X receptor selective ligands that induce apoptosis in leukemia cells. *J Med Chem* 38, 3146-3155.
- Bolstad, B. M., Collin, F., Simpson, K. M., Irizarry, R. A., and Speed, T. P. (2004). Experimental design and low-level analysis of microarray data. *Int Rev Neurobiol* 60, 25-58.
- Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185-193.
- Bommer, M., Benecke, A., Gronemeyer, H., and Rochette-Egly, C. (2002). TIF2 mediates the synergy between RARalpha 1 activation functions AF-1 and AF-2. *J Biol Chem* 277, 37961-37966.
- Bono, H., Yagi, K., Kasukawa, T., Nikaido, I., Tominaga, N., Miki, R., Mizuno, Y., Tomaru, Y., Goto, H., Nitanda, H., et al. (2003). Systematic expression profiling of the mouse transcriptome using RIKEN cDNA microarrays. *Genome Res* 13, 1318-1323.
- Bookout, A. L., Jeong, Y., Downes, M., Yu, R. T., Evans, R. M., and Mangelsdorf, D. J. (2006). Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. *Cell* 126, 789-799.
- Boon, K., Osorio, E. C., Greenhut, S. F., Schaefer, C. F., Shoemaker, J., Polyak, K., Morin, P. J., Buetow, K. H., Strausberg, R. L., De Souza, S. J., and Riggins, G. J. (2002). An anatomy of normal and malignant gene expression. *Proc Natl Acad Sci U S A* 99, 11287-11292.
- Bordet, R., Ouk, T., Petralt, O., Gele, P., Gautier, S., Laprais, M., Deplanque, D., Duriez, P., Staels, B., Fruchart, J. C., and Bastide, M. (2006). PPAR: a new pharmacological target for neuroprotection in stroke and neurodegenerative diseases. *Biochem Soc Trans* 34, 1341-1346.
- Borgius, L. J., Steffensen, K. R., Gustafsson, J. A., and Treuter, E. (2002). Glucocorticoid signaling is perturbed by the atypical orphan receptor and corepressor SHP. *J Biol Chem* 277, 49761-49766.

- Borngraeber, S., Budny, M. J., Chiellini, G., Cunha-Lima, S. T., Togashi, M., Webb, P., Baxter, J. D., Scanlan, T. S., and Fletterick, R. J. (2003). Ligand selectivity by seeking hydrophobicity in thyroid hormone receptor. *Proc Natl Acad Sci U S A* **100**, 15358-15363.
- Bosotti, R., Locatelli, G., Healy, S., Scacheri, E., Sartori, L., Mercurio, C., Calogero, R., and Isacchi, A. (2007). Cross platform microarray analysis for robust identification of differentially expressed genes. *BMC Bioinformatics* **8 Suppl 1**, S5.
- Boucher, J., Masri, B., Daviaud, D., Gesta, S., Guigne, C., Mazzucotelli, A., Castan-Laurell, I., Tack, I., Knibiehler, B., Carpene, C., et al. (2005). Apelin, a newly identified adipokine up-regulated by insulin and obesity. *Endocrinology* **146**, 1764-1771.
- Boulesteix, A. L., and Strimmer, K. (2005). Predicting transcription factor activities from combined analysis of microarray and ChIP data: a partial least squares approach. *Theor Biol Med Model* **2**, 23.
- Boulias, K., Katrakili, N., Bamberg, K., Underhill, P., Greenfield, A., and Talianidis, I. (2005). Regulation of hepatic metabolic pathways by the orphan nuclear receptor SHP. *Embo J* **24**, 2624-2633.
- Boulias, K., and Talianidis, I. (2004). Functional role of G9a-induced histone methylation in small heterodimer partner-mediated transcriptional repression. *Nucleic Acids Res* **32**, 6096-6103.
- Bouly, M., Masson, D., Gross, B., Jiang, X. C., Fievet, C., Castro, G., Tall, A. R., Fruchart, J. C., Staels, B., Lagrost, L., and Luc, G. (2001). Induction of the phospholipid transfer protein gene accounts for the high density lipoprotein enlargement in mice treated with fenofibrate. *J Biol Chem* **276**, 25841-25847.
- Bourgeois, S., Pfahl, M., and Baulieu, E. E. (1984). DNA binding properties of glucocorticosteroid receptors bound to the steroid antagonist RU-486. *Embo J* **3**, 751-755.
- Bourguet, W., Germain, P., and Gronemeyer, H. (2000a). Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. *Trends Pharmacol Sci* **21**, 381-388.
- Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., and Moras, D. (1995). Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. *Nature* **375**, 377-382.
- Bourguet, W., Vivat, V., Wurtz, J. M., Chambon, P., Gronemeyer, H., and Moras, D. (2000b). Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. *Mol Cell* **5**, 289-298.
- Braissant, O., Foufelle, F., Scotto, C., Dauca, M., and Wahli, W. (1996). Differential expression of peroxisome proliferator-activated receptors (PPARs): tissue distribution of PPAR-alpha, -beta, and -gamma in the adult rat. *Endocrinology* **137**, 354-366.
- Braissant, O., and Wahli, W. (1998). Differential expression of peroxisome proliferator-activated receptor-alpha, -beta, and -gamma during rat embryonic development. *Endocrinology* **139**, 2748-2754.
- Bramlett, K. S., Houck, K. A., Borchert, K. M., Dowless, M. S., Kulanthaivel, P., Zhang, Y., Beyer, T. P., Schmidt, R., Thomas, J. S., Michael, L. F., et al. (2003). A natural product ligand of the oxysterol receptor, liver X receptor. *J Pharmacol Exp Ther* **307**, 291-296.
- Brand, M. D., and Esteves, T. C. (2005). Physiological functions of the mitochondrial uncoupling proteins UCP2 and UCP3. *Cell Metab* **2**, 85-93.
- Brandt, J. M., Djouadi, F., and Kelly, D. P. (1998). Fatty acids activate transcription of the muscle carnitine palmitoyltransferase I gene in cardiac myocytes via the peroxisome proliferator-activated receptor alpha. *J Biol Chem* **273**, 23786-23792.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., et al. (2000a). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18**, 630-634.
- Brenner, S., Williams, S. R., Vermaas, E. H., Storck, T., Moon, K., McCollum, C., Mao, J. I., Luo, S., Kirchner, J. J., Eletr, S., et al. (2000b). In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci U S A* **97**, 1665-1670.
- Breyne, P., Dreesen, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inze, D., and Zabeau, M. (2003). Quantitative cDNA-AFLP analysis for genome-wide expression studies. *Mol Genet Genomics* **269**, 173-179.
- Breyne, P., and Zabeau, M. (2001). Genome-wide expression analysis of plant cell cycle modulated genes. *Curr Opin Plant Biol* **4**, 136-142.
- Brown-Grant, K., Fink, G., Greig, F., and Murray, M. A. (1975). Altered sexual development in male rats after oestrogen administration during the neonatal period. *J Reprod Fertil* **44**, 25-42.
- Brown, A. J., Dusso, A. S., and Slatopolsky, E. (2002). Vitamin D analogues for secondary hyperparathyroidism. *Nephrol Dial Transplant* **17 Suppl 10**, 10-19.
- Brown, M. P., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., Jr., and Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A* **97**, 262-267.
- Brown, N. J. (2003). Eplerenone: cardiovascular protection. *Circulation* **107**, 2512-2518.

- Brun, S., Carmona, M. C., Mampel, T., Vinas, O., Giralt, M., Iglesias, R., and Villarroya, F. (1999). Activators of peroxisome proliferator-activated receptor-alpha induce the expression of the uncoupling protein-3 gene in skeletal muscle: a potential mechanism for the lipid intake-dependent activation of uncoupling protein-3 gene expression at birth. *Diabetes* 48, 1217-1222.
- Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engstrom, O., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389, 753-758.
- Budhu, A. S., and Noy, N. (2002). Direct channeling of retinoic acid between cellular retinoic acid-binding protein II and retinoic acid receptor sensitizes mammary carcinoma cells to retinoic acid-induced growth arrest. *Mol Cell Biol* 22, 2632-2641.
- Bunger, M., van den Bosch, H. M., van der Meijde, J., Kersten, S., Hooiveld, G. J., and Muller, M. (2007). Genome-wide analysis of PPARalpha activation in murine small intestine. *Physiol Genomics* 30, 192-204.
- Burgermeister, E., Schnoebelen, A., Flament, A., Benz, J., Stihle, M., Gsell, B., Rufer, A., Ruf, A., Kuhn, B., Marki, H. P., et al. (2006). A novel partial agonist of peroxisome proliferator-activated receptor-gamma (PPARgamma) recruits PPARgamma-coactivator-1alpha, prevents triglyceride accumulation, and potentiates insulin signaling in vitro. *Mol Endocrinol* 20, 809-830.
- Burns, K. A., and Vanden Heuvel, J. P. (2007). Modulation of PPAR activity via phosphorylation. *Biochim Biophys Acta*.
- Burris, T. P., Montrose, C., Houck, K. A., Osborne, H. E., Bocchinfuso, W. P., Yaden, B. C., Cheng, C. C., Zink, R. W., Barr, R. J., Hepler, C. D., et al. (2005). The hypolipidemic natural product guggulsterone is a promiscuous steroid receptor ligand. *Mol Pharmacol* 67, 948-954.
- Cai, Y., Konishi, T., Han, G., Campwala, K. H., French, S. W., and Wan, Y. J. (2002). The role of hepatocyte RXR alpha in xenobiotic-sensing nuclear receptor-mediated pathways. *Eur J Pharm Sci* 15, 89-96.
- Campbell, F. M., Kozak, R., Wagner, A., Altarejos, J. Y., Dyck, J. R., Belke, D. D., Severson, D. L., Kelly, D. P., and Lopaschuk, G. D. (2002). A role for peroxisome proliferator-activated receptor alpha (PPARalpha) in the control of cardiac malonyl-CoA levels: reduced fatty acid oxidation rates and increased glucose oxidation rates in the hearts of mice lacking PPARalpha are associated with higher concentrations of malonyl-CoA and reduced expression of malonyl-CoA decarboxylase. *J Biol Chem* 277, 4098-4103.
- Canales, R. D., Luo, Y., Willey, J. C., Austermiller, B., Barbacioru, C. C., Boysen, C., Hunkapiller, K., Jensen, R. V., Knight, C. R., Lee, K. Y., et al. (2006). Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* 24, 1115-1122.
- Canaple, L., Rambaud, J., Dkhissi-Benyahya, O., Rayet, B., Tan, N. S., Michalik, L., Delaunay, F., Wahli, W., and Laudet, V. (2006). Reciprocal regulation of brain and muscle Arnt-like protein 1 and peroxisome proliferator-activated receptor alpha defines a novel positive feedback loop in the rodent liver circadian clock. *Mol Endocrinol* 20, 1715-1727.
- Capell, W. H., DeSouza, C. A., Poirier, P., Bell, M. L., Stauffer, B. L., Weil, K. M., Hernandez, T. L., and Eckel, R. H. (2003). Short-term triglyceride lowering with fenofibrate improves vasodilator function in subjects with hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 23, 307-313.
- Carter, M. E., Gulick, T., Moore, D. D., and Kelly, D. P. (1994). A pleiotropic element in the medium-chain acyl coenzyme A dehydrogenase gene promoter mediates transcriptional regulation by multiple nuclear receptor transcription factors and defines novel receptor-DNA binding motifs. *Mol Cell Biol* 14, 4360-4372.
- Casas, F., Domenjoud, L., Rochard, P., Hatier, R., Rodier, A., Daury, L., Bianchi, A., Kremerik-Bouillaud, P., Becuwe, P., Keller, J., et al. (2000a). A 45 kDa protein related to PPARgamma2, induced by peroxisome proliferators, is located in the mitochondrial matrix. *FEBS Lett* 478, 4-8.
- Casas, F., Pineau, T., Rochard, P., Rodier, A., Daury, L., Dauca, M., Cabello, G., and Wrutniak-Cabello, C. (2000b). New molecular aspects of regulation of mitochondrial activity by fenofibrate and fasting. *FEBS Lett* 482, 71-74.
- Castelein, H., Gulick, T., Declercq, P. E., Mannaerts, G. P., Moore, D. D., and Baes, M. I. (1994). The peroxisome proliferator activated receptor regulates malic enzyme gene expression. *J Biol Chem* 269, 26754-26758.
- Cattley, R. C., and Roberts, R. A. (2000). Peroxisome proliferators and carcinogenesis: editorial perspectives. *Mutat Res* 448, 117-119.
- Chadick, J. Z., and Asturias, F. J. (2005). Structure of eukaryotic Mediator complexes. *Trends Biochem Sci* 30, 264-271.
- Chandraratna, R. A. (1996). Tazarotene--first of a new generation of receptor-selective retinoids. *Br J Dermatol Suppl* 49, 18-25.
- Chang, C. S., Kokontis, J., and Liao, S. T. (1988a). Molecular cloning of human and rat complementary DNA encoding androgen receptors. *Science* 240, 324-326.
- Chang, C. S., Kokontis, J., and Liao, S. T. (1988b). Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors. *Proc Natl Acad Sci U S A* 85, 7211-7215.

- Chao, E. Y., Collins, J. L., Gaillard, S., Miller, A. B., Wang, L., Orband-Miller, L. A., Nolte, R. T., McDonnell, D. P., Willson, T. M., and Zuercher, W. J. (2006). Structure-guided synthesis of tamoxifen analogs with improved selectivity for the orphan ERR $\gamma$ . *Bioorg Med Chem Lett* 16, 821-824.
- Chapman, S., Schenk, P., Kazan, K., and Manners, J. (2002). Using biplots to interpret gene expression patterns in plants. *Bioinformatics* 18, 202-204.
- Chawla, A., Boisvert, W. A., Lee, C. H., Laffitte, B. A., Barak, Y., Joseph, S. B., Liao, D., Nagy, L., Edwards, P. A., Curtiss, L. K., et al. (2001). A PPAR gamma-LXR-ABCA1 pathway in macrophages is involved in cholesterol efflux and atherogenesis. *Mol Cell* 7, 161-171.
- Chawla, A., and Lazar, M. A. (1994). Peroxisome proliferator and retinoid signaling pathways co-regulate preadipocyte phenotype and survival. *Proc Natl Acad Sci U S A* 91, 1786-1790.
- Cheema, S. K., and Agellon, L. B. (2000). The murine and human cholesterol 7alpha-hydroxylase gene promoters are differentially responsive to regulation by fatty acids mediated via peroxisome proliferator-activated receptor alpha. *J Biol Chem* 275, 12530-12536.
- Chen, F., Law, S. W., and O'Malley, B. W. (1993). Identification of two mPPAR related receptors and evidence for the existence of five subfamily members. *Biochem Biophys Res Commun* 196, 671-677.
- Chen, H., and Privalsky, M. L. (1995). Cooperative formation of high-order oligomers by retinoid X receptors: an unexpected mode of DNA recognition. *Proc Natl Acad Sci U S A* 92, 422-426.
- Chen, J., Kinyamu, H. K., and Archer, T. K. (2006). Changes in attitude, changes in latitude: nuclear receptors remodeling chromatin to regulate transcription. *Mol Endocrinol* 20, 1-13.
- Chen, J. D., and Evans, R. M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* 377, 454-457.
- Chen, J. J., Rowley, J. D., and Wang, S. M. (2000). Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. *Proc Natl Acad Sci U S A* 97, 349-353.
- Chen, M., Beaven, S., and Tontonoz, P. (2005). Identification and characterization of two alternatively spliced transcript variants of human liver X receptor alpha. *J Lipid Res* 46, 2570-2579.
- Cherel, Y., Robin, J. P., Heitz, A., Calgari, C., and Le Maho, Y. (1992). Relationships between lipid availability and protein utilization during prolonged fasting. *J Comp Physiol [B]* 162, 305-313.
- Chevillard, G., Clemencet, M. C., Etienne, P., Martin, P., Pineau, T., Latruffe, N., and Nicolas-Frances, V. (2003). Tissue-specific expression of two peroxisomal 3-ketoacyl-CoA thiolase genes in wild and PPAR alpha-null mice and induction by fenofibrate. *Adv Exp Med Biol* 544, 55-56.
- Chevillard, G., Clemencet, M. C., Etienne, P., Martin, P., Pineau, T., Latruffe, N., and Nicolas-Frances, V. (2004a). Molecular cloning, gene structure and expression profile of two mouse peroxisomal 3-ketoacyl-CoA thiolase genes. *BMC Biochem* 5, 3.
- Chevillard, G., Clemencet, M. C., Latruffe, N., and Nicolas-Frances, V. (2004b). Targeted disruption of the peroxisomal thiolase B gene in mouse: a new model to study disorders related to peroxisomal lipid metabolism. *Biochimie* 86, 849-856.
- Chiang, J. Y. (2002). Bile acid regulation of gene expression: roles of nuclear hormone receptors. *Endocr Rev* 23, 443-463.
- Chinetti-Gbaguidi, G., Rigamonti, E., Helin, L., Mutka, A. L., Lepore, M., Fruchart, J. C., Clavey, V., Ikonen, E., Lestavel, S., and Staels, B. (2005). Peroxisome proliferator-activated receptor alpha controls cellular cholesterol trafficking in macrophages. *J Lipid Res* 46, 2717-2725.
- Chinetti, G., Fruchart, J. C., and Staels, B. (2006). Transcriptional regulation of macrophage cholesterol trafficking by PPARalpha and LXR. *Biochem Soc Trans* 34, 1128-1131.
- Chinetti, G., Gbaguidi, F. G., Griglio, S., Mallat, Z., Antonucci, M., Poulain, P., Chapman, J., Fruchart, J. C., Tedgui, A., Najib-Fruchart, J., and Staels, B. (2000). CLA-1/SR-BI is expressed in atherosclerotic lesion macrophages and regulated by activators of peroxisome proliferator-activated receptors. *Circulation* 101, 2411-2417.
- Chinetti, G., Griglio, S., Antonucci, M., Torra, I. P., Delerive, P., Majd, Z., Fruchart, J. C., Chapman, J., Najib, J., and Staels, B. (1998). Activation of proliferator-activated receptors alpha and gamma induces apoptosis of human monocyte-derived macrophages. *J Biol Chem* 273, 25573-25580.
- Chinetti, G., Lestavel, S., Bocher, V., Remaley, A. T., Neve, B., Torra, I. P., Teissier, E., Minnich, A., Jaye, M., Duverger, N., et al. (2001). PPAR-alpha and PPAR-gamma activators induce cholesterol removal from human macrophage foam cells through stimulation of the ABCA1 pathway. *Nat Med* 7, 53-58.
- Chou, W. L., Chuang, L. M., Chou, C. C., Wang, A. H., Lawson, J. A., Fitzgerald, G. A., and Chang, Z. F. (2007). Identification of a Novel Prostaglandin Reductase Reveals the Involvement of Prostaglandin E2 Catabolism in Regulation of Peroxisome Proliferator-activated Receptor {gamma} Activation. *J Biol Chem* 282, 18162-18172.
- Chrencik, J. E., Orans, J., Moore, L. B., Xue, Y., Peng, L., Collins, J. L., Wisely, G. B., Lambert, M. H., Kliewer, S. A., and Redinbo, M. R. (2005). Structural disorder in the complex of human pregnane X receptor and the macrolide antibiotic rifampicin. *Mol Endocrinol* 19, 1125-1134.

- Christian, M., White, R., and Parker, M. G. (2006). Metabolic regulation by the nuclear receptor corepressor RIP140. *Trends Endocrinol Metab* 17, 243-250.
- Chu, R., Lim, H., Brumfield, L., Liu, H., Herring, C., Ulitz, P., Reddy, J. K., and Davison, M. (2004). Protein profiling of mouse livers with peroxisome proliferator-activated receptor alpha activation. *Mol Cell Biol* 24, 6288-6297.
- Chu, R., Lin, Y., Rao, M. S., and Reddy, J. K. (1995). Cooperative formation of higher order peroxisome proliferator-activated receptor and retinoid X receptor complexes on the peroxisome proliferator responsive element of the rat hydratase-dehydrogenase gene. *J Biol Chem* 270, 29636-29639.
- Chui, P. C., Guan, H. P., Lehrke, M., and Lazar, M. A. (2005). PPARgamma regulates adipocyte cholesterol metabolism via oxidized LDL receptor 1. *J Clin Invest* 115, 2244-2256.
- Churchill, G. A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 32 Suppl, 490-495.
- Clement, L., Poirier, H., Niot, I., Bocher, V., Guerre-Millo, M., Krief, S., Staels, B., and Besnard, P. (2002). Dietary trans-10,cis-12 conjugated linoleic acid induces hyperinsulinemia and fatty liver in the mouse. *J Lipid Res* 43, 1400-1409.
- Coe, N. R., Smith, A. J., Frohnert, B. I., Watkins, P. A., and Bernlohr, D. A. (1999). The fatty acid transport protein (FATP1) is a very long chain acyl-CoA synthetase. *J Biol Chem* 274, 36300-36304.
- Columbano, A., Ledda-Columbano, G. M., Pibiri, M., Concas, D., Reddy, J. K., and Rao, M. S. (2001). Peroxisome proliferator-activated receptor-alpha mice show enhanced hepatocyte proliferation in response to the hepatomitogen 1,4-bis [2-(3,5-dichloropyridyloxy)] benzene, a ligand of constitutive androstane receptor. *Hepatology* 34, 262-266.
- Committee, N. R. N. (1999). A unified nomenclature system for the nuclear receptor superfamily. *Cell* 97, 161-163.
- Compe, E., Drane, P., Laurent, C., Diderich, K., Braun, C., Hoeijmakers, J. H., and Egly, J. M. (2005). Dysregulation of the peroxisome proliferator-activated receptor target genes by XPD mutations. *Mol Cell Biol* 25, 6065-6076.
- Conaway, R. C., Sato, S., Tomomori-Sato, C., Yao, T., and Conaway, J. W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem Sci* 30, 250-255.
- Conneely, O. M., Mulac-Jericevic, B., and Lydon, J. P. (2003). Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptor isoforms. *Steroids* 68, 771-778.
- Connell, J. M., and Davies, E. (2005). The new biology of aldosterone. *J Endocrinol* 186, 1-20.
- Cook, L., Nagi, M. N., Piscatelli, J., Joseph, T., Prasad, M. R., Ghesquier, D., and Cinti, D. L. (1986). Hepatic subcellular distribution of short-chain beta-ketoacyl coenzyme A reductase and trans-2-enoyl coenzyme A hydratase: 25- to 50-fold stimulation of microsomal activities by the peroxisome proliferator, di-(2-ethylhexyl)phthalate. *Arch Biochem Biophys* 245, 24-36.
- Cope, L. M., Irizarry, R. A., Jaffee, H. A., Wu, Z., and Speed, T. P. (2004). A benchmark for Affymetrix GeneChip expression measures. *Bioinformatics* 20, 323-331.
- Corton, J. C., Bocos, C., Moreno, E. S., Merritt, A., Cattley, R. C., and Gustafsson, J. A. (1997). Peroxisome proliferators alter the expression of estrogen-metabolizing enzymes. *Biochimie* 79, 151-162.
- Corton, J. C., Fan, L. Q., Brown, S., Anderson, S. P., Bocos, C., Cattley, R. C., Mode, A., and Gustafsson, J. A. (1998). Down-regulation of cytochrome P450 2C family members and positive acute-phase response gene expression by peroxisome proliferator chemicals. *Mol Pharmacol* 54, 463-473.
- Corton, J. C., and Lapinskas, P. J. (2005). Peroxisome proliferator-activated receptors: mediators of phthalate ester-induced effects in the male reproductive tract? *Toxicol Sci* 83, 4-17.
- Corvol, P., Michaud, A., Menard, J., Freifeld, M., and Mahoudeau, J. (1975). Antiandrogenic effect of spiro lactones: mechanism of action. *Endocrinology* 97, 52-58.
- Coste, H., and Rodriguez, J. C. (2002). Orphan nuclear hormone receptor Rev-erbalpha regulates the human apolipoprotein CIII promoter. *J Biol Chem* 277, 27120-27129.
- Costet, P., Lalanne, F., Gerbod-Giannone, M. C., Molina, J. R., Fu, X., Lund, E. G., Gudas, L. J., and Tall, A. R. (2003). Retinoic acid receptor-mediated induction of ABCA1 in macrophages. *Mol Cell Biol* 23, 7756-7766.
- Costet, P., Legendre, C., More, J., Edgar, A., Galtier, P., and Pineau, T. (1998). Peroxisome proliferator-activated receptor alpha-isoform deficiency leads to progressive dyslipidemia with sexually dimorphic obesity and steatosis. *J Biol Chem* 273, 29577-29585.
- Costet, P., Luo, Y., Wang, N., and Tall, A. R. (2000). Sterol-dependent transactivation of the ABC1 promoter by the liver X receptor/retinoid X receptor. *J Biol Chem* 275, 28240-28245.
- Couette, B., Lombes, M., Baulieu, E. E., and Rafestin-Oblin, M. E. (1992a). Aldosterone antagonists destabilize the mineralocorticosteroid receptor. *Biochem J* 282 ( Pt 3), 697-702.

- Couette, B., Marsaud, V., Baulieu, E. E., Richard-Foy, H., and Rafestin-Oblin, M. E. (1992b). Spironolactone, an aldosterone antagonist, acts as an antiglucocorticosteroid on the mouse mammary tumor virus promoter. *Endocrinology* 130, 430-436.
- Council, N. R., Agriculture, B. o., Nutrition, C. o. A., and Nutrition, S. o. L. A. (1995). Nutrient Requirements of Laboratory Animals, 4th rev. ed. edn: National Academy Press).
- Craig, T. A., and Kumar, R. (1996). Synthesis and purification of soluble ligand binding domain of the human vitamin D<sub>3</sub> receptor. *Biochem Biophys Res Commun* 218, 902-907.
- Crawford, S. E., Qi, C., Misra, P., Stellmach, V., Rao, M. S., Engel, J. D., Zhu, Y., and Reddy, J. K. (2002). Defects of the heart, eye, and megakaryocytes in peroxisome proliferator activator receptor-binding protein (PPB) null embryos implicate GATA family of transcription factors. *J Biol Chem* 277, 3585-3592.
- Cronet, P., Petersen, J. F., Folmer, R., Blomberg, N., Sjöblom, K., Karlsson, U., Lindstedt, E. L., and Bamberg, K. (2001). Structure of the PPARalpha and -gamma ligand binding domain in complex with AZ 242; ligand selectivity and agonist activation in the PPAR family. *Structure* 9, 699-706.
- Cullingford, T. E., Bhakoo, K., Peuchen, S., Dolphin, C. T., Patel, R., and Clark, J. B. (1998). Distribution of mRNAs encoding the peroxisome proliferator-activated receptor alpha, beta, and gamma and the retinoid X receptor alpha, beta, and gamma in rat central nervous system. *J Neurochem* 70, 1366-1375.
- Dagnelie, P. (1975). Analyse statistique à plusieurs variables (Gembloix: Les Presses agronomiques de Gembloix).
- Dagnelie, P. (1998). Statistique théorique et appliquée Tome 2 Inférence statistique à une et à deux dimensions (Paris Bruxelles: De Boeck université).
- Dahlman-Wright, K., Baumann, H., McEwan, I. J., Almlöf, T., Wright, A. P., Gustafsson, J. A., and Hard, T. (1995). Structural characterization of a minimal functional transactivation domain from the human glucocorticoid receptor. *Proc Natl Acad Sci U S A* 92, 1699-1703.
- Darimont, B. D., Wagner, R. L., Apriletti, J. W., Stallcup, M. R., Kushner, P. J., Baxter, J. D., Fletterick, R. J., and Yamamoto, K. R. (1998). Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev* 12, 3343-3356.
- Davis, R. A., Miyake, J. H., Hui, T. Y., and Spann, N. J. (2002). Regulation of cholesterol-7alpha-hydroxylase: BAREly missing a SHP. *J Lipid Res* 43, 533-543.
- de Gasparo, M., Joss, U., Ramjoue, H. P., Whitebread, S. E., Haenni, H., Schenkel, L., Krahenbuehl, C., Biollaz, M., Grob, J., Schmidlin, J., and et al. (1987). Three new epoxy-spirolactone derivatives: characterization in vivo and in vitro. *J Pharmacol Exp Ther* 240, 650-656.
- De Leenheer, A. P., Lambert, W. E., and Claeys, I. (1982). All-trans-retinoic acid: measurement of reference values in human serum by high performance liquid chromatography. *J Lipid Res* 23, 1362-1367.
- de The, H., Vivanco-Ruiz, M. M., Tiollais, P., Stunnenberg, H., and Dejean, A. (1990). Identification of a retinoic acid responsive element in the retinoic acid receptor beta gene. *Nature* 343, 177-180.
- Dejean, S., Martin, P. G., Baccini, A., and Besse, P. (2007). Clustering time-series gene expression data using smoothing spline derivatives. *EURASIP J Bioinform Syst Biol*, 70561.
- del Castillo-Olivares, A., and Gil, G. (2001). Suppression of sterol 12alpha-hydroxylase transcription by the short heterodimer partner: insights into the repression mechanism. *Nucleic Acids Res* 29, 4035-4042.
- Delerive, P., De Bosscher, K., Besnard, S., Vanden Berghe, W., Peters, J. M., Gonzalez, F. J., Fruchart, J. C., Tedgui, A., Haegeman, G., and Staels, B. (1999a). Peroxisome proliferator-activated receptor alpha negatively regulates the vascular inflammatory gene response by negative cross-talk with transcription factors NF-kappaB and AP-1. *J Biol Chem* 274, 32048-32054.
- Delerive, P., De Bosscher, K., Vanden Berghe, W., Fruchart, J. C., Haegeman, G., and Staels, B. (2002). DNA binding-independent induction of IkappaBalphagene transcription by PPARalpha. *Mol Endocrinol* 16, 1029-1039.
- Delerive, P., Gervois, P., Fruchart, J. C., and Staels, B. (2000). Induction of IkappaBalphagene expression as a mechanism contributing to the anti-inflammatory activities of peroxisome proliferator-activated receptor-alpha activators. *J Biol Chem* 275, 36703-36707.
- Delerive, P., Martin-Nizard, F., Chinetti, G., Trottein, F., Fruchart, J. C., Najib, J., Duriez, P., and Staels, B. (1999b). Peroxisome proliferator-activated receptor activators inhibit thrombin-induced endothelin-1 production in human vascular endothelial cells by inhibiting the activator protein-1 signaling pathway. *Circ Res* 85, 394-402.
- Delmar, P., Robin, S., and Daudin, J. J. (2005). VarMixt: efficient variance modelling for the differential analysis of replicated gene expression data. *Bioinformatics* 21, 502-508.
- Denson, L. A., Auld, K. L., Schiek, D. S., McClure, M. H., Mangelsdorf, D. J., and Karpen, S. J. (2000). Interleukin-1beta suppresses retinoid transactivation of two hepatic transporter genes involved in bile formation. *J Biol Chem* 275, 8835-8843.
- DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680-686.

- Desvergne, B., Michalik, L., and Wahli, W. (2006). Transcriptional regulation of metabolism. *Physiol Rev* 86, 465-514.
- Desvergne, B., and Wahli, W. (1999). Peroxisome proliferator-activated receptors: nuclear control of metabolism. *Endocr Rev* 20, 649-688.
- Devchand, P. R., Keller, H., Peters, J. M., Vazquez, M., Gonzalez, F. J., and Wahli, W. (1996). The PPARalpha-leukotriene B4 pathway to inflammation control. *Nature* 384, 39-43.
- Dhe-Paganon, S., Duda, K., Iwamoto, M., Chi, Y. I., and Shoelson, S. E. (2002). Crystal structure of the HNF4 alpha ligand binding domain in complex with endogenous fatty acid ligand. *J Biol Chem* 277, 37973-37976.
- Diep, Q. N., Touyz, R. M., and Schiffzin, E. L. (2000). Docosahexaenoic acid, a peroxisome proliferator-activated receptor-alpha ligand, induces apoptosis in vascular smooth muscle cells by stimulation of p38 mitogen-activated protein kinase. *Hypertension* 36, 851-855.
- Diradourian, C., Girard, J., and Pegorier, J. P. (2005). Phosphorylation of PPARs: from molecular characterization to physiological relevance. *Biochimie* 87, 33-38.
- DiRenzo, J., Soderstrom, M., Kurokawa, R., Ogliastro, M. H., Ricote, M., Ingrey, S., Horlein, A., Rosenfeld, M. G., and Glass, C. K. (1997). Peroxisome proliferator-activated receptors and retinoic acid receptors differentially control the interactions of retinoid X receptor heterodimers with ligands, coactivators, and corepressors. *Mol Cell Biol* 17, 2166-2176.
- Djouadi, F., Aubey, F., Schlemmer, D., Ruiter, J. P., Wanders, R. J., Strauss, A. W., and Bastin, J. (2005). Bezafibrate increases very-long-chain acyl-CoA dehydrogenase protein and mRNA expression in deficient fibroblasts and is a potential therapy for fatty acid oxidation disorders. *Hum Mol Genet* 14, 2695-2703.
- Djouadi, F., Brandt, J. M., Weinheimer, C. J., Leone, T. C., Gonzalez, F. J., and Kelly, D. P. (1999). The role of the peroxisome proliferator-activated receptor alpha (PPAR alpha) in the control of cardiac lipid metabolism. *Prostaglandins Leukot Essent Fatty Acids* 60, 339-343.
- Dobbin, K., Shih, J. H., and Simon, R. (2003). Questions and answers on design of dual-label microarrays for identifying differentially expressed genes. *J Natl Cancer Inst* 95, 1362-1369.
- Dongol, B., Shah, Y., Kim, I., Gonzalez, F. J., and Hunt, M. C. (2007). The acyl-CoA thioesterase I is regulated by PPAR{alpha} and HNF4{alpha} via a distal response element in the promoter. *J Lipid Res* 48, 1781-1791.
- Dowell, P., Ishmael, J. E., Avram, D., Peterson, V. J., Nevrivy, D. J., and Leid, M. (1997). p300 functions as a coactivator for the peroxisome proliferator-activated receptor alpha. *J Biol Chem* 272, 33435-33443.
- Dowell, P., Ishmael, J. E., Avram, D., Peterson, V. J., Nevrivy, D. J., and Leid, M. (1999). Identification of nuclear receptor corepressor as a peroxisome proliferator-activated receptor alpha interacting protein. *J Biol Chem* 274, 15901-15907.
- Downes, M., Verdecia, M. A., Roecker, A. J., Hughes, R., Hogenesch, J. B., Kast-Woelbern, H. R., Bowman, M. E., Ferrer, J. L., Anisfeld, A. M., Edwards, P. A., et al. (2003). A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. *Mol Cell* 11, 1079-1092.
- Dragomir, E., Tircol, M., Manduteanu, I., Voinea, M., and Simionescu, M. (2006). Aspirin and PPAR-alpha activators inhibit monocyte chemoattractant protein-1 expression induced by high glucose concentration in human endothelial cells. *Vascul Pharmacol* 44, 440-449.
- Dreyer, C., Krey, G., Keller, H., Givel, F., Helftenbein, G., and Wahli, W. (1992). Control of the peroxisomal beta-oxidation pathway by a novel family of nuclear hormone receptors. *Cell* 68, 879-887.
- Duarte, J., Perriere, G., Laudet, V., and Robinson-Rechavi, M. (2002). NUREBASE: database of nuclear hormone receptors. *Nucleic Acids Res* 30, 364-368.
- Duda, K., Chi, Y. I., and Shoelson, S. E. (2004). Structural basis for HNF-4alpha activation by ligand and coactivator binding. *J Biol Chem* 279, 23311-23316.
- Dudoit, S., Gentleman, R. C., and Quackenbush, J. (2003a). Open source software for the analysis of microarray data. *Biotechniques Suppl*, 45-51.
- Dudoit, S., Shaffer, J. P., and Boldrick, J. C. (2003b). Multiple Hypothesis Testing in Microarray Experiments. *Statistical Science* 18, 71-103.
- Dudoit, S., Yang, Y. H., Callow, M. J., and Speed, T. P. (2002). Statistical Methods for Identifying Differentially Expressed Genes in Replicated cDNA Microarray Experiments. *Statistica Sinica* 12, 111-139.
- Duester, G., Mic, F. A., and Molotkov, A. (2003). Cytosolic retinoid dehydrogenases govern ubiquitous metabolism of retinol to retinaldehyde followed by tissue-specific metabolism to retinoic acid. *Chem Biol Interact* 143-144, 201-210.
- Durand, B., Saunders, M., Leroy, P., Leid, M., and Chambon, P. (1992). All-trans and 9-cis retinoic acid induction of CRABPII transcription is mediated by RAR-RXR heterodimers bound to DR1 and DR2 repeated motifs. *Cell* 71, 73-85.

- Dussault, I., Yoo, H. D., Lin, M., Wang, E., Fan, M., Batta, A. K., Salen, G., Erickson, S. K., and Forman, B. M. (2003). Identification of an endogenous ligand that activates pregnane X receptor-mediated sterol clearance. *Proc Natl Acad Sci U S A* **100**, 833-838.
- Ebdrup, S., Pettersson, I., Rasmussen, H. B., Deussen, H. J., Frost Jensen, A., Mortensen, S. B., Fleckner, J., Pridal, L., Nygaard, L., and Sauerberg, P. (2003). Synthesis and biological and structural characterization of the dual-acting peroxisome proliferator-activated receptor alpha/gamma agonist ragaglitazar. *J Med Chem* **46**, 1306-1317.
- Eberhardt, W., Akool el, S., Rebhan, J., Frank, S., Beck, K. F., Franzen, R., Hamada, F. M., and Pfeilschifter, J. (2002). Inhibition of cytokine-induced matrix metalloproteinase 9 expression by peroxisome proliferator-activated receptor alpha agonists is indirect and due to a NO-mediated reduction of mRNA stability. *J Biol Chem* **277**, 33518-33528.
- Eberle, D., Hegarty, B., Bossard, P., Ferre, P., and Foufelle, F. (2004). SREBP transcription factors: master regulators of lipid homeostasis. *Biochimie* **86**, 839-848.
- Echchgadda, I., Song, C. S., Roy, A. K., and Chatterjee, B. (2004). Dehydroepiandrosterone sulfotransferase is a target for transcriptional induction by the vitamin D receptor. *Mol Pharmacol* **65**, 720-729.
- Eckhoff, C., and Nau, H. (1990). Identification and quantitation of all-trans- and 13-cis-retinoic acid and 13-cis-4-oxoretinoic acid in human plasma. *J Lipid Res* **31**, 1445-1454.
- Edgar, A. D., Tomkiewicz, C., Costet, P., Legendre, C., Aggerbeck, M., Bouquet, J., Staels, B., Guyomard, C., Pineau, T., and Barouki, R. (1998). Fenofibrate modifies transaminase gene expression via a peroxisome proliferator activated receptor alpha-dependent pathway. *Toxicol Lett* **98**, 13-23.
- Eelen, G., Verlinden, L., Rochel, N., Claessens, F., De Clercq, P., Vandewalle, M., Tocchini-Valentini, G., Moras, D., Bouillon, R., and Verstuyf, A. (2005). Superagonistic action of 14-epi-analogs of 1,25-dihydroxyvitamin D explained by vitamin D receptor-coactivator interaction. *Mol Pharmacol* **67**, 1566-1573.
- Efron, B., Tibshirani, R., Goss, B., and Chu, G. (2000). Microarrays and Their Use in a Comparative Experiment (Stanford: Stanford University), pp. 1-38.
- Egea, P. F., Mitschler, A., and Moras, D. (2002). Molecular recognition of agonist ligands by RXRs. *Mol Endocrinol* **16**, 987-997.
- Egea, P. F., Mitschler, A., Rochel, N., Ruff, M., Chambon, P., and Moras, D. (2000). Crystal structure of the human RXRalpha ligand-binding domain bound to its natural ligand: 9-cis retinoic acid. *Embo J* **19**, 2592-2601.
- Eiler, S., Gangloff, M., Duclaud, S., Moras, D., and Ruff, M. (2001). Overexpression, purification, and crystal structure of native ER alpha LBD. *Protein Expr Purif* **22**, 165-173.
- Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-14868.
- Elholm, M., Bjerking, G., Knudsen, J., Kristiansen, K., and Mandrup, S. (1996). Regulatory elements in the promoter region of the rat gene encoding the acyl-CoA-binding protein. *Gene* **173**, 233-238.
- Ellinghaus, P., Wolfrum, C., Assmann, G., Spener, F., and Seedorf, U. (1999). Phytanic acid activates the peroxisome proliferator-activated receptor alpha (PPARalpha) in sterol carrier protein 2-/ sterol carrier protein x-deficient mice. *J Biol Chem* **274**, 2766-2772.
- Emberley, E., Huang, G. J., Hamedani, M. K., Czosnek, A., Ali, D., Grolla, A., Lu, B., Watson, P. H., Murphy, L. C., and Leygue, E. (2003). Identification of new human coding steroid receptor RNA activator isoforms. *Biochem Biophys Res Commun* **301**, 509-515.
- Engel, T., Lorkowski, S., Lueken, A., Rust, S., Schluter, B., Berger, G., Cullen, P., and Assmann, G. (2001). The human ABCG4 gene is regulated by oxysterols and retinoids in monocyte-derived macrophages. *Biochem Biophys Res Commun* **288**, 483-488.
- Erol, E., Kumar, L. S., Cline, G. W., Shulman, G. I., Kelly, D. P., and Binas, B. (2004). Liver fatty acid binding protein is required for high rates of hepatic fatty acid oxidation but not for the action of PPARalpha in fasting mice. *Faseb J* **18**, 347-349.
- Escher, P., Braissant, O., Basu-Modak, S., Michalik, L., Wahli, W., and Desvergne, B. (2001). Rat PPARs: quantitative analysis in adult rat tissues and regulation in fasting and refeeding. *Endocrinology* **142**, 4195-4202.
- Escofier, B., and Pagès, J. (1998). Analyses factorielles simples et multiples objectifs, méthodes et interprétation, 3e Ed edn (Paris: Dunod).
- Estebanez-Perpina, E., Moore, J. M., Mar, E., Delgado-Rodrigues, E., Nguyen, P., Baxter, J. D., Buehrer, B. M., Webb, P., Fletterick, R. J., and Guy, R. K. (2005). The molecular mechanisms of coactivator utilization in ligand-dependent transactivation by the androgen receptor. *J Biol Chem* **280**, 8060-8068.
- Eul, J., Meyer, M. E., Tora, L., Bocquel, M. T., Quirin-Stricker, C., Chambon, P., and Gronemeyer, H. (1989). Expression of active hormone and DNA-binding domains of the chicken progesterone receptor in *E. coli*. *Embo J* **8**, 83-90.
- Evans, R. W., and Braverman, L. E. (1986). Use of <sup>125</sup>I-triiodothyroacetic acid to measure nuclear thyroid hormone receptor. *Endocr Res* **12**, 37-47.

- Fagart, J., Huyet, J., Pinon, G. M., Rochel, M., Mayer, C., and Rafestin-Oblin, M. E. (2005). Crystal structure of a mutant mineralocorticoid receptor responsible for hypertension. *Nat Struct Mol Biol* 12, 554-555.
- Fajas, L., Auboeuf, D., Raspe, E., Schoonjans, K., Lefebvre, A. M., Saladin, R., Najib, J., Laville, M., Fruchart, J. C., Deeb, S., et al. (1997). The organization, promoter analysis, and expression of the human PPARgamma gene. *J Biol Chem* 272, 18779-18789.
- Fajas, L., Fruchart, J. C., and Auwerx, J. (1998). PPARgamma3 mRNA: a distinct PPARgamma mRNA subtype transcribed from an independent promoter. *FEBS Lett* 438, 55-60.
- Fajas, L., Schoonjans, K., Gelman, L., Kim, J. B., Najib, J., Martin, G., Fruchart, J. C., Briggs, M., Spiegelman, B. M., and Auwerx, J. (1999). Regulation of peroxisome proliferator-activated receptor gamma expression by adipocyte differentiation and determination factor 1/sterol regulatory element binding protein 1: implications for adipocyte differentiation and metabolism. *Mol Cell Biol* 19, 5495-5503.
- Fan, C. Y., Pan, J., Usuda, N., Yeldandi, A. V., Rao, M. S., and Reddy, J. K. (1998a). Steatohepatitis, spontaneous peroxisome proliferation and liver tumors in mice lacking peroxisomal fatty acyl-CoA oxidase. Implications for peroxisome proliferator-activated receptor alpha natural ligand metabolism. *J Biol Chem* 273, 15639-15645.
- Fan, L. Q., Brown-Borg, H., Brown, S., Westin, S., Mode, A., and Corton, J. C. (2004a). PPARalpha activators down-regulate CYP2C7, a retinoic acid and testosterone hydroxylase. *Toxicology* 203, 41-48.
- Fan, L. Q., Cattley, R. C., and Corton, J. C. (1998b). Tissue-specific induction of 17 beta-hydroxysteroid dehydrogenase type IV by peroxisome proliferator chemicals is dependent on the peroxisome proliferator-activated receptor alpha. *J Endocrinol* 158, 237-246.
- Fan, L. Q., You, L., Brown-Borg, H., Brown, S., Edwards, R. J., and Corton, J. C. (2004b). Regulation of phase I and phase II steroid metabolism enzymes by PPAR alpha activators. *Toxicology* 204, 109-121.
- Fang, H. L., Strom, S. C., Cai, H., Falany, C. N., Kocarek, T. A., and Runge-Morris, M. (2005). Regulation of human hepatic hydroxysteroid sulfotransferase gene expression by the peroxisome proliferator-activated receptor alpha transcription factor. *Mol Pharmacol* 67, 1257-1267.
- Farnegardh, M., Bonn, T., Sun, S., Ljunggren, J., Ahola, H., Wilhelmsson, A., Gustafsson, J. A., and Carlquist, M. (2003). The three-dimensional structure of the liver X receptor beta reveals a flexible ligand-binding pocket that can accommodate fundamentally different ligands. *J Biol Chem* 278, 38821-38828.
- Farrants, A. K., Nilsson, A., Troen, G., Blomhoff, R., and Pedersen, J. I. (1993). The effect of retinoids and clofibrate acid on the peroxisomal oxidation of palmitic acid and of 3 alpha,7 alpha,12 alpha-trihydroxy-5 beta-cholestanoic acid in rat and rabbit hepatocytes. *Biochim Biophys Acta* 1168, 100-107.
- Fayard, E., Auwerx, J., and Schoonjans, K. (2004). LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. *Trends Cell Biol* 14, 250-260.
- Feige, J. N., Gelman, L., Michalik, L., Desvergne, B., and Wahli, W. (2006). From molecular action to physiological outputs: peroxisome proliferator-activated receptors are nuclear receptors at the crossroads of key cellular functions. *Prog Lipid Res* 45, 120-159.
- Feige, J. N., Gelman, L., Rossi, D., Zoete, V., Metivier, R., Tudor, C., Anghel, S. I., Grosdidier, A., Lathion, C., Engelborghs, Y., et al. (2007). The Endocrine Disruptor Monoethyl-hexyl-phthalate Is a Selective Peroxisome Proliferator-activated Receptor {gamma} Modulator That Promotes Adipogenesis. *J Biol Chem* 282, 19152-19166.
- Feige, J. N., Gelman, L., Tudor, C., Engelborghs, Y., Wahli, W., and Desvergne, B. (2005). Fluorescence imaging reveals the nuclear behavior of peroxisome proliferator-activated receptor/retinoid X receptor heterodimers in the absence and presence of ligand. *J Biol Chem* 280, 17880-17890.
- Fejes-Toth, G., Pearce, D., and Naray-Fejes-Toth, A. (1998). Subcellular localization of mineralocorticoid receptors in living cells: effects of receptor agonists and antagonists. *Proc Natl Acad Sci U S A* 95, 2973-2978.
- Ferdinandusse, S., Denis, S., Overmars, H., Van Eckhoudt, L., Van Veldhoven, P. P., Duran, M., Wanders, R. J., and Baes, M. (2005). Developmental changes of bile acid composition and conjugation in L- and D-bifunctional protein single and double knockout mice. *J Biol Chem* 280, 18658-18666.
- Fernandes, I., Bastien, Y., Wai, T., Nygard, K., Lin, R., Cormier, O., Lee, H. S., Eng, F., Bertos, N. R., Pelletier, N., et al. (2003). Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms. *Mol Cell* 11, 139-150.
- Fernandes, I., and White, J. H. (2003). Agonist-bound nuclear receptors: not just targets of coactivators. *J Mol Endocrinol* 31, 1-7.
- Fiedler, J., Simon, F. R., Iwahashi, M., and Murphy, R. C. (2001). Effect of peroxisome proliferator-activated receptor alpha activation on leukotriene B4 metabolism in isolated rat hepatocytes. *J Pharmacol Exp Ther* 299, 691-697.
- Figueira, A. C., Dias, S. M., Santos, M. A., Apriletti, J. W., Baxter, J. D., Webb, P., Neves, F. A., Simeoni, L. A., Ribeiro, R. C., and Polikarpov, I. (2006). Human thyroid receptor forms tetramers in solution, which dissociate into dimers upon ligand binding. *Cell Biochem Biophys* 44, 453-462.

- Finck, B. N. (2007). The PPAR regulatory system in cardiac physiology and disease. *Cardiovasc Res* 73, 269-277.
- Finck, B. N., and Kelly, D. P. (2006). PGC-1 coactivators: inducible regulators of energy metabolism in health and disease. *J Clin Invest* 116, 615-622.
- Fitzpatrick, S. L., Berrodin, T. J., Jenkins, S. F., Sindoni, D. M., Deecher, D. C., and Frail, D. E. (1999). Effect of estrogen agonists and antagonists on induction of progesterone receptor in a rat hypothalamic cell line. *Endocrinology* 140, 3928-3937.
- Flaig, R., Greschik, H., Peluso-Iltis, C., and Moras, D. (2005). Structural basis for the cell-specific activities of the NGFI-B and the Nurr1 ligand-binding domain. *J Biol Chem* 280, 19250-19258.
- Folco, G., and Murphy, R. C. (2006). Eicosanoid transcellular biosynthesis: from cell-cell interactions to in vivo tissue responses. *Pharmacol Rev* 58, 375-388.
- Foliot, A., and Beaune, P. (1994). Effects of microsomal enzyme inducers on glutathione S-transferase isoenzymes in livers of rats and hamsters. *Biochem Pharmacol* 48, 293-300.
- Fondell, J. D., Ge, H., and Roeder, R. G. (1996). Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc Natl Acad Sci U S A* 93, 8329-8333.
- Forman, B. M., Chen, J., and Evans, R. M. (1997). Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta. *Proc Natl Acad Sci U S A* 94, 4312-4317.
- Forman, B. M., Tontonoz, P., Chen, J., Brun, R. P., Spiegelman, B. M., and Evans, R. M. (1995). 15-Deoxy-delta 12, 14-prostaglandin J2 is a ligand for the adipocyte determination factor PPAR gamma. *Cell* 83, 803-812.
- Forman, B. M., Tzameli, I., Choi, H. S., Chen, J., Simha, D., Seol, W., Evans, R. M., and Moore, D. D. (1998). Androstane metabolites bind to and deactivate the nuclear receptor CAR-beta. *Nature* 395, 612-615.
- Fourcade, S., Savary, S., Albet, S., Gauthe, D., Gondcaille, C., Pineau, T., Bellenger, J., Bentejac, M., Holzinger, A., Berger, J., and Bugaut, M. (2001). Fibrate induction of the adrenoleukodystrophy-related gene (ABCD2): promoter analysis and role of the peroxisome proliferator-activated receptor PPARalpha. *Eur J Biochem* 268, 3490-3500.
- Fournier, T., Tsatsaris, V., Handschuh, K., and Evain-Brion, D. (2007). PPARs and the placenta. *Placenta* 28, 65-76.
- Franceschini, G., Lovati, M. R., Manzoni, C., Michelagnoli, S., Pazzucconi, F., Gianfranceschi, G., Vecchio, G., and Sirtori, C. R. (1995). Effect of gemfibrozil treatment in hypercholesterolemia on low density lipoprotein (LDL) subclass distribution and LDL-cell interaction. *Atherosclerosis* 114, 61-71.
- Freedman, L. P. (1992). Anatomy of the steroid receptor zinc finger region. *Endocr Rev* 13, 129-145.
- Freedman, L. P., Arce, V., and Perez Fernandez, R. (1994). DNA sequences that act as high affinity targets for the vitamin D3 receptor in the absence of the retinoid X receptor. *Mol Endocrinol* 8, 265-273.
- Freedman, N. D., and Yamamoto, K. R. (2004). Importin 7 and importin alpha/importin beta are nuclear import receptors for the glucocorticoid receptor. *Mol Biol Cell* 15, 2276-2286.
- Frenkel, B., Bishara-Shieban, J., and Bar-Tana, J. (1994). The effect of beta,beta'-tetramethylhexadecanedioic acid (MEDICA 16) on plasma very-low-density lipoprotein metabolism in rats: role of apolipoprotein C-III. *Biochem J* 298 ( Pt 2), 409-414.
- Freyberger, A., and Ahr, H. J. (2004). Development and standardization of a simple binding assay for the detection of compounds with affinity for the androgen receptor. *Toxicology* 195, 113-126.
- Friedman, J. M., and Halaas, J. L. (1998). Leptin and the regulation of body weight in mammals. *Nature* 395, 763-770.
- Frohnert, B. I., Hui, T. Y., and Bernlohr, D. A. (1999). Identification of a functional peroxisome proliferator-responsive element in the murine fatty acid transport protein gene. *J Biol Chem* 274, 3970-3977.
- Froment, P., Gizard, F., Defever, D., Staels, B., Dupont, J., and Monget, P. (2006). Peroxisome proliferator-activated receptors in reproductive tissues: from gametogenesis to parturition. *J Endocrinol* 189, 199-209.
- Fu, J., Astarita, G., Gaetani, S., Kim, J., Cravatt, B. F., Mackie, K., and Piomelli, D. (2007). Food intake regulates oleoylethanolamide formation and degradation in the proximal small intestine. *J Biol Chem* 282, 1518-1528.
- Fu, J., Gaetani, S., Oveisi, F., Lo Verme, J., Serrano, A., Rodriguez De Fonseca, F., Rosengarth, A., Luecke, H., Di Giacomo, B., Tarzia, G., and Piomelli, D. (2003). Oleoylethanolamide regulates feeding and body weight through activation of the nuclear receptor PPAR-alpha. *Nature* 425, 90-93.
- Fuhrmann, U., Bengtson, C., Repenthal, G., and Schillinger, E. (1992). Stable transfection of androgen receptor and MMTV-CAT into mammalian cells: inhibition of cat expression by anti-androgens. *J Steroid Biochem Mol Biol* 42, 787-793.
- Fujino, T., Sato, Y., Une, M., Kanayasu-Toyoda, T., Yamaguchi, T., Shudo, K., Inoue, K., and Nishimaki-Mogami, T. (2003). In vitro farnesoid X receptor ligand sensor assay using surface plasmon resonance and based on ligand-induced coactivator association. *J Steroid Biochem Mol Biol* 87, 247-252.

- Fukao, T., Yamaguchi, S., Kano, M., Orii, T., Fujiki, Y., Osumi, T., and Hashimoto, T. (1990). Molecular cloning and sequence of the complementary DNA encoding human mitochondrial acetoacetyl-coenzyme A thiolase and study of the variant enzymes in cultured fibroblasts from patients with 3-ketothiolase deficiency. *J Clin Invest* 86, 2086-2092.
- Fuller, P. J., and Young, M. J. (2005). Mechanisms of mineralocorticoid action. *Hypertension* 46, 1227-1235.
- Funder, J. W. (1992). Glucocorticoid receptors. *J Steroid Biochem Mol Biol* 43, 389-394.
- Funk, C. D. (2001). Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science* 294, 1871-1875.
- Furr, B. J., and Tucker, H. (1996). The preclinical development of bicalutamide: pharmacodynamics and mechanism of action. *Urology* 47, 13-25; discussion 29-32.
- Fyffe, S. A., Alphey, M. S., Buetow, L., Smith, T. K., Ferguson, M. A., Sorensen, M. D., Bjorkling, F., and Hunter, W. N. (2006a). Recombinant human PPAR-beta/delta ligand-binding domain is locked in an activated conformation by endogenous fatty acids. *J Mol Biol* 356, 1005-1013.
- Fyffe, S. A., Alphey, M. S., Buetow, L., Smith, T. K., Ferguson, M. A., Sorensen, M. D., Bjorkling, F., and Hunter, W. N. (2006b). Reevaluation of the PPAR-beta/delta ligand binding domain model reveals why it exhibits the activated form. *Mol Cell* 21, 1-2.
- Gampe, R. T., Jr., Montana, V. G., Lambert, M. H., Miller, A. B., Bledsoe, R. K., Milburn, M. V., Kliewer, S. A., Willson, T. M., and Xu, H. E. (2000a). Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol Cell* 5, 545-555.
- Gampe, R. T., Jr., Montana, V. G., Lambert, M. H., Wisely, G. B., Milburn, M. V., and Xu, H. E. (2000b). Structural basis for autorepression of retinoid X receptor by tetramer formation and the AF-2 helix. *Genes Dev* 14, 2229-2241.
- Gangloff, M., Ruff, M., Eiler, S., Duclaud, S., Wurtz, J. M., and Moras, D. (2001). Crystal structure of a mutant hERalpha ligand-binding domain reveals key structural features for the mechanism of partial agonism. *J Biol Chem* 276, 15059-15065.
- Garcia, B. A., Pesavento, J. J., Mizzen, C. A., and Kelleher, N. L. (2007). Pervasive combinatorial modification of histone H3 in human cells. *Nat Methods* 4, 487-489.
- Garlatti, M., Daheshia, M., Slater, E., Bouquet, J., Hanoune, J., Beato, M., and Barouki, R. (1994). A functional glucocorticoid-responsive unit composed of two overlapping inactive receptor-binding sites: evidence for formation of a receptor tetramer. *Mol Cell Biol* 14, 8007-8017.
- Garthwaite, S. M., and McMahon, E. G. (2004). The evolution of aldosterone antagonists. *Mol Cell Endocrinol* 217, 27-31.
- Gaudel, C. L., and Grimaldi, P. A. (2007). Metabolic Functions of Peroxisome Proliferator-Activated Receptor beta/delta in Skeletal Muscle. *PPAR Res* 2007, 86394.
- Ge, K., Guermah, M., Yuan, C. X., Ito, M., Wallberg, A. E., Spiegelman, B. M., and Roeder, R. G. (2002). Transcription coactivator TRAP220 is required for PPAR gamma 2-stimulated adipogenesis. *Nature* 417, 563-567.
- Ge, Y., Dudoit, S., and Speed, T. P. (2003). Resampling-based Multiple Testing for Microarray Data Analysis. *Test* 12, 1-77.
- Gearhart, M. D., Holmbeck, S. M., Evans, R. M., Dyson, H. J., and Wright, P. E. (2003). Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. *J Mol Biol* 327, 819-832.
- Gearing, K. L., Crickmore, A., and Gustafsson, J. A. (1994). Structure of the mouse peroxisome proliferator activated receptor alpha gene. *Biochem Biophys Res Commun* 199, 255-263.
- Gelman, L., Feige, J. N., Tudor, C., Engelborghs, Y., Wahli, W., and Desvergne, B. (2006). Integrating nuclear receptor mobility in models of gene regulation. *Nucl Recept Signal* 4, e010.
- Gelman, L., Zhou, G., Fajas, L., Raspe, E., Fruchart, J. C., and Auwerx, J. (1999). p300 interacts with the N- and C-terminal part of PPARgamma2 in a ligand-independent and -dependent manner, respectively. *J Biol Chem* 274, 7681-7688.
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5, R80.
- Georget, V., Terouanne, B., Lumbroso, S., Nicolas, J. C., and Sultan, C. (1998). Trafficking of androgen receptor mutants fused to green fluorescent protein: a new investigation of partial androgen insensitivity syndrome. *J Clin Endocrinol Metab* 83, 3597-3603.
- Germain, P., Kammerer, S., Perez, E., Peluso-Iltis, C., Tortolani, D., Zusi, F. C., Starrett, J., Lapointe, P., Daris, J. P., Marinier, A., et al. (2004). Rational design of RAR-selective ligands revealed by RARbeta crystal structure. *EMBO Rep* 5, 877-882.

- Gervois, P., Chopin-Delannoy, S., Fadel, A., Dubois, G., Kosykh, V., Fruchart, J. C., Najib, J., Laudet, V., and Staels, B. (1999a). Fibrates increase human REV-ERB $\alpha$  expression in liver via a novel peroxisome proliferator-activated receptor response element. *Mol Endocrinol* 13, 400-409.
- Gervois, P., Fruchart, J. C., and Staels, B. (2007). Drug Insight: mechanisms of action and therapeutic applications for agonists of peroxisome proliferator-activated receptors. *Nat Clin Pract Endocrinol Metab* 3, 145-156.
- Gervois, P., Kleemann, R., Pilon, A., Percevault, F., Koenig, W., Staels, B., and Kooistra, T. (2004). Global suppression of IL-6-induced acute phase response gene expression after chronic in vivo treatment with the peroxisome proliferator-activated receptor-alpha activator fenofibrate. *J Biol Chem* 279, 16154-16160.
- Gervois, P., Torra, I. P., Chinetti, G., Grotzinger, T., Dubois, G., Fruchart, J. C., Fruchart-Najib, J., Leitersdorf, E., and Staels, B. (1999b). A truncated human peroxisome proliferator-activated receptor alpha splice variant with dominant negative activity. *Mol Endocrinol* 13, 1535-1549.
- Gervois, P., Vu-Dac, N., Kleemann, R., Kockx, M., Dubois, G., Laine, B., Kosykh, V., Fruchart, J. C., Kooistra, T., and Staels, B. (2001). Negative regulation of human fibrinogen gene expression by peroxisome proliferator-activated receptor alpha agonists via inhibition of CCAAT box/enhancer-binding protein beta. *J Biol Chem* 276, 33471-33477.
- Gewirth, D. T., and Sigler, P. B. (1995). The basis for half-site specificity explored through a non-cognate steroid receptor-DNA complex. *Nat Struct Biol* 2, 386-394.
- Giannopoulos, G., and Keichline, D. (1981). Species-related differences in steroid-binding specificity of glucocorticoid receptors in lung. *Endocrinology* 108, 1414-1419.
- Girnun, G. D., Domann, F. E., Moore, S. A., and Robbins, M. E. (2002). Identification of a functional peroxisome proliferator-activated receptor response element in the rat catalase promoter. *Mol Endocrinol* 16, 2793-2801.
- Gizard, F., Amant, C., Barbier, O., Bellosta, S., Robillard, R., Percevault, F., Sevestre, H., Krimpenfort, P., Corsini, A., Rochette, J., et al. (2005). PPAR alpha inhibits vascular smooth muscle cell proliferation underlying intimal hyperplasia by inducing the tumor suppressor p16INK4a. *J Clin Invest* 115, 3228-3238.
- Glass, C. K. (1994). Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. *Endocr Rev* 15, 391-407.
- Glass, C. K., and Rosenfeld, M. G. (2000). The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev* 14, 121-141.
- Glickman, J. F., Wu, X., Mercuri, R., Illy, C., Bowen, B. R., He, Y., and Sills, M. (2002). A comparison of ALPHAScreen, TR-FRET, and TRF as assay methods for FXR nuclear receptors. *J Biomol Screen* 7, 3-10.
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-537.
- Goodman, M. N., Lowell, B., Belur, E., and Ruderman, N. B. (1984). Sites of protein conservation and loss during starvation: influence of adiposity. *Am J Physiol* 246, E383-390.
- Goodwin, B., Gauthier, K. C., Umetani, M., Watson, M. A., Lochansky, M. I., Collins, J. L., Leitersdorf, E., Mangelsdorf, D. J., Kliewer, S. A., and Repa, J. J. (2003). Identification of bile acid precursors as endogenous ligands for the nuclear xenobiotic pregnane X receptor. *Proc Natl Acad Sci U S A* 100, 223-228.
- Goodwin, B., Jones, S. A., Price, R. R., Watson, M. A., McKee, D. D., Moore, L. B., Galardi, C., Wilson, J. G., Lewis, M. C., Roth, M. E., et al. (2000). A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. *Mol Cell* 6, 517-526.
- Gottlicher, M., Widmark, E., Li, Q., and Gustafsson, J. A. (1992). Fatty acids activate a chimera of the clofibrate acid-activated receptor and the glucocorticoid receptor. *Proc Natl Acad Sci U S A* 89, 4653-4657.
- Gould Rothberg, B. E., Sundseth, S. S., DiPippo, V. A., Brown, P. J., Winegar, D. A., Gottshalk, W. K., Shenoy, S. G., and Rothberg, J. M. (2001). The characterization of PPAR alpha ligand drug action in an in vivo model by comprehensive differential gene expression profiling. *Funct Integr Genomics* 1, 294-304.
- Goya, K., Sumitani, S., Xu, X., Kitamura, T., Yamamoto, H., Kurebayashi, S., Saito, H., Kouhara, H., Kasayama, S., and Kawase, I. (2004). Peroxisome proliferator-activated receptor alpha agonists increase nitric oxide synthase expression in vascular endothelial cells. *Arterioscler Thromb Vasc Biol* 24, 658-663.
- Gray, J. P., Burns, K. A., Leas, T. L., Perdew, G. H., and Vanden Heuvel, J. P. (2005). Regulation of peroxisome proliferator-activated receptor alpha by protein kinase C. *Biochemistry* 44, 10313-10321.
- Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J. M., Argos, P., and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature* 320, 134-139.
- Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., and Shine, J. (1986). Sequence and expression of human estrogen receptor complementary DNA. *Science* 231, 1150-1154.
- Greschik, H., Flaig, R., Renaud, J. P., and Moras, D. (2004). Structural basis for the deactivation of the estrogen-related receptor gamma by diethylstilbestrol or 4-hydroxytamoxifen and determinants of selectivity. *J Biol Chem* 279, 33639-33646.

Greschik, H., Wurtz, J. M., Sanglier, S., Bourguet, W., van Dorsselaer, A., Moras, D., and Renaud, J. P. (2002). Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3. *Mol Cell* 9, 303-313.

Gress, T. M., Hoheisel, J. D., Lennon, G. G., Zehetner, G., and Lehrach, H. (1992). Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm Genome* 3, 609-619.

Grober, J., Zagħini, I., Fujii, H., Jones, S. A., Kliewer, S. A., Willson, T. M., Ono, T., and Besnard, P. (1999). Identification of a bile acid-responsive element in the human ileal bile acid-binding protein gene. Involvement of the farnesoid X receptor/9-cis-retinoic acid receptor heterodimer. *J Biol Chem* 274, 29749-29754.

Gronemeyer, H., and Laudet, V. (1995). Transcription factors 3: nuclear receptors. *Protein Profile* 2, 1173-1308.

Gronemeyer, H., Turcotte, B., Quirin-Stricker, C., Bocquel, M. T., Meyer, M. E., Krozowski, Z., Jeltsch, J. M., Lerouge, T., Garnier, J. M., and Chambon, P. (1987). The chicken progesterone receptor: sequence, expression and functional analysis. *Embo J* 6, 3985-3994.

Gu, W., Malik, S., Ito, M., Yuan, C. X., Fondell, J. D., Zhang, X., Martinez, E., Qin, J., and Roeder, R. G. (1999). A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol Cell* 3, 97-108.

Guan, H. P., Ishizuka, T., Chui, P. C., Lehrke, M., and Lazar, M. A. (2005). Corepressors selectively control the transcriptional activity of PPARgamma in adipocytes. *Genes Dev* 19, 453-461.

Guan, Y. (2004). Peroxisome proliferator-activated receptor family and its relationship to renal complications of the metabolic syndrome. *J Am Soc Nephrol* 15, 2801-2815.

Guardiola-Diaz, H. M., Rehnmark, S., Usuda, N., Albrektsen, T., Feltkamp, D., Gustafsson, J. A., and Alexson, S. E. (1999). Rat peroxisome proliferator-activated receptors and brown adipose tissue function during cold acclimatization. *J Biol Chem* 274, 23368-23377.

Guellich, A., Damy, T., Lecarpentier, Y., Conti, M., Claes, V., Samuel, J. L., Quillard, J., Hebert, J. L., Pineau, T., and Coirault, C. (2007). Role of oxidative stress in cardiac dysfunction of PPAR{alpha}-- mice. *Am J Physiol Heart Circ Physiol* 293, H93-H102.

Guerre-Millo, M., Gervois, P., Raspe, E., Madsen, L., Poulain, P., Derudas, B., Herbert, J. M., Winegar, D. A., Willson, T. M., Fruchart, J. C., et al. (2000). Peroxisome proliferator-activated receptor alpha activators improve insulin sensitivity and reduce adiposity. *J Biol Chem* 275, 16638-16642.

Guerre-Millo, M., Rouault, C., Poulain, P., Andre, J., Poitout, V., Peters, J. M., Gonzalez, F. J., Fruchart, J. C., Reach, G., and Staels, B. (2001). PPAR-alpha-null mice are protected from high-fat diet-induced insulin resistance. *Diabetes* 50, 2809-2814.

Gui, Y., Silha, J. V., Mishra, S., and Murphy, L. J. (2003). Changes in adipokine expression during food deprivation in the mouse and the relationship to fasting-induced insulin resistance. *Can J Physiol Pharmacol* 81, 979-985.

Guillou, H., Martin, P., Jan, S., D'Andrea, S., Roulet, A., Catheline, D., Rioux, V., Pineau, T., and Legrand, P. (2002). Comparative effect of fenofibrate on hepatic desaturases in wild-type and peroxisome proliferator-activated receptor alpha-deficient mice. *Lipids* 37, 981-989.

Guiochon-Mantel, A., Lescop, P., Christin-Maitre, S., Loosfelt, H., Perrot-Applanat, M., and Milgrom, E. (1991). Nucleocytoplasmic shuttling of the progesterone receptor. *Embo J* 10, 3851-3859.

Guiochon-Mantel, A., Loosfelt, H., Lescop, P., Sar, S., Atger, M., Perrot-Applanat, M., and Milgrom, E. (1989). Mechanisms of nuclear localization of the progesterone receptor: evidence for interaction between monomers. *Cell* 57, 1147-1154.

Gulick, T., Cresci, S., Caira, T., Moore, D. D., and Kelly, D. P. (1994). The peroxisome proliferator-activated receptor regulates mitochondrial fatty acid oxidative enzyme gene expression. *Proc Natl Acad Sci U S A* 91, 11012-11016.

Guo, D., Sarkar, J., Ahmed, M. R., Viswakarma, N., Jia, Y., Yu, S., Sambasiva Rao, M., and Reddy, J. K. (2006a). Peroxisome proliferator-activated receptor (PPAR)-binding protein (PBP) but not PPAR-interacting protein (PRIP) is required for nuclear translocation of constitutive androstane receptor in mouse liver. *Biochem Biophys Res Commun* 347, 485-495.

Guo, L., Lobenhofer, E. K., Wang, C., Shippy, R., Harris, S. C., Zhang, L., Mei, N., Chen, T., Herman, D., Goodsaid, F. M., et al. (2006b). Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol* 24, 1162-1169.

Guzman, M., Lo Verme, J., Fu, J., Oveisi, F., Blazquez, C., and Piomelli, D. (2004). Oleoylethanolamide stimulates lipolysis by activating the nuclear receptor peroxisome proliferator-activated receptor alpha (PPAR-alpha). *J Biol Chem* 279, 27849-27854.

- Haffner, C. D., Lenhard, J. M., Miller, A. B., McDougald, D. L., Dwornik, K., Ittoop, O. R., Gampe, R. T., Jr., Xu, H. E., Blanchard, S., Montana, V. G., *et al.* (2004). Structure-based design of potent retinoid X receptor alpha agonists. *J Med Chem* *47*, 2010-2029.
- Hager, G. L., Nagaich, A. K., Johnson, T. A., Walker, D. A., and John, S. (2004). Dynamics of nuclear receptor movement and transcription. *Biochim Biophys Acta* *1677*, 46-51.
- Hammad, S. M., Barth, J. L., Knaak, C., and Argraves, W. S. (2000). Megalin acts in concert with cubilin to mediate endocytosis of high density lipoproteins. *J Biol Chem* *275*, 12003-12008.
- Hammer, G. D., and Ingraham, H. A. (1999). Steroidogenic factor-1: its role in endocrine organ development and differentiation. *Front Neuroendocrinol* *20*, 199-223.
- Hammer, G. D., Krylova, I., Zhang, Y., Darimont, B. D., Simpson, K., Weigel, N. L., and Ingraham, H. A. (1999). Phosphorylation of the nuclear receptor SF-1 modulates cofactor recruitment: integration of hormone signaling in reproduction and stress. *Mol Cell* *3*, 521-526.
- Han, C. Y., Chiba, T., Campbell, J. S., Fausto, N., Chaisson, M., Orasanu, G., Plutzky, J., and Chait, A. (2006). Reciprocal and coordinate regulation of serum amyloid A versus apolipoprotein A-I and paraoxonase-1 by inflammation in murine hepatocytes. *Arterioscler Thromb Vasc Biol* *26*, 1806-1813.
- Handschin, C., and Spiegelman, B. M. (2006). Peroxisome proliferator-activated receptor gamma coactivator 1 coactivators, energy homeostasis, and metabolism. *Endocr Rev* *27*, 728-735.
- Hangeland, J. J., Doweyko, A. M., Dejneca, T., Friends, T. J., Devasthale, P., Mellstrom, K., Sandberg, J., Grynfarb, M., Sack, J. S., Einspahr, H., *et al.* (2004). Thyroid receptor ligands. Part 2: Thyromimetics with improved selectivity for the thyroid hormone receptor beta. *Bioorg Med Chem Lett* *14*, 3549-3553.
- Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J. A., and Kaptein, R. (1990). Solution structure of the glucocorticoid receptor DNA-binding domain. *Science* *249*, 157-160.
- Harding, H. P., Atkins, G. B., Jaffe, A. B., Seo, W. J., and Lazar, M. A. (1997). Transcriptional activation and repression by RORalpha, an orphan nuclear receptor required for cerebellar development. *Mol Endocrinol* *11*, 1737-1746.
- Harding, H. P., and Lazar, M. A. (1993). The orphan receptor Rev-ErbA alpha activates transcription via a novel response element. *Mol Cell Biol* *13*, 3113-3121.
- Harding, H. P., and Lazar, M. A. (1995). The monomer-binding orphan receptor Rev-Erb represses transcription as a dimer on a novel direct repeat. *Mol Cell Biol* *15*, 4791-4802.
- Hashimoto, T., Cook, W. S., Qi, C., Yeldandi, A. V., Reddy, J. K., and Rao, M. S. (2000). Defect in peroxisome proliferator-activated receptor alpha-inducible fatty acid oxidation determines the severity of hepatic steatosis in response to fasting. *J Biol Chem* *275*, 28918-28928.
- Hashimoto, T., Fujita, T., Usuda, N., Cook, W., Qi, C., Peters, J. M., Gonzalez, F. J., Yeldandi, A. V., Rao, M. S., and Reddy, J. K. (1999). Peroxisomal and mitochondrial fatty acid beta-oxidation in mice nullizygous for both peroxisome proliferator-activated receptor alpha and peroxisomal fatty acyl-CoA oxidase. Genotype correlation with fatty liver phenotype. *J Biol Chem* *274*, 19228-19236.
- Hatchell, E. C., Colley, S. M., Beveridge, D. J., Epis, M. R., Stuart, L. M., Giles, K. M., Redfern, A. D., Miles, L. E., Barker, A., MacDonald, L. M., *et al.* (2006). SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. *Mol Cell* *22*, 657-668.
- He, B., Gampe, R. T., Jr., Kole, A. J., Hnat, A. T., Stanley, T. B., An, G., Stewart, E. L., Kalman, R. I., Minges, J. T., and Wilson, E. M. (2004). Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell* *16*, 425-438.
- He, Y., Yin, D., Perera, M., Kirkovsky, L., Stourman, N., Li, W., Dalton, J. T., and Miller, D. D. (2002). Novel nonsteroidal ligands with high binding affinity and potent functional activity for the androgen receptor. *Eur J Med Chem* *37*, 619-634.
- Hegardt, F. G. (1998). Transcriptional regulation of mitochondrial HMG-CoA synthase in the control of ketogenesis. *Biochimie* *80*, 803-806.
- Heinlein, C. A., and Chang, C. (2004). Androgen receptor in prostate cancer. *Endocr Rev* *25*, 276-308.
- Held, G. A., Grinstein, G., and Tu, Y. (2003). Modeling of DNA microarray data by using physical properties of hybridization. *Proc Natl Acad Sci U S A* *100*, 7575-7580.
- Held, G. A., Grinstein, G., and Tu, Y. (2006). Relationship between gene expression and observed intensities in DNA microarrays--a modeling study. *Nucleic Acids Res* *34*, e70.
- Hellal-Levy, C., Couette, B., Fagart, J., Souque, A., Gomez-Sanchez, C., and Rafestin-Oblin, M. (1999). Specific hydroxylations determine selective corticosteroid recognition by human glucocorticoid and mineralocorticoid receptors. *FEBS Lett* *464*, 9-13.
- Helledie, T., Grontved, L., Jensen, S. S., Kiilerich, P., Rietveld, L., Albrektsen, T., Boysen, M. S., Nohr, J., Larsen, L. K., Fleckner, J., *et al.* (2002). The gene encoding the Acyl-CoA-binding protein is activated by peroxisome proliferator-activated receptor gamma through an intronic response element functionally conserved between humans and rodents. *J Biol Chem* *277*, 26821-26830.

- Henke, B. R., Consler, T. G., Go, N., Hale, R. L., Hohman, D. R., Jones, S. A., Lu, A. T., Moore, L. B., Moore, J. T., Orband-Miller, L. A., et al. (2002). A new series of estrogen receptor modulators that display selectivity for estrogen receptor beta. *J Med Chem* 45, 5492-5505.
- Henley, D. V., and Korach, K. S. (2006). Endocrine-disrupting chemicals use distinct mechanisms of action to modulate endocrine system function. *Endocrinology* 147, S25-32.
- Hermanowski-Vosatka, A., Gerhold, D., Mundt, S. S., Loving, V. A., Lu, M., Chen, Y., Elbrecht, A., Wu, M., Doepper, T., Kelly, L., et al. (2000). PPARalpha agonists reduce 11beta-hydroxysteroid dehydrogenase type 1 in the liver. *Biochem Biophys Res Commun* 279, 330-336.
- Hermanson, O., Glass, C. K., and Rosenfeld, M. G. (2002). Nuclear receptor coregulators: multiple modes of modification. *Trends Endocrinol Metab* 13, 55-60.
- Hertz, R., Bishara-Shieban, J., and Bar-Tana, J. (1995). Mode of action of peroxisome proliferators as hypolipidemic drugs. Suppression of apolipoprotein C-III. *J Biol Chem* 270, 13470-13475.
- Hertz, R., Seckbach, M., Zakin, M. M., and Bar-Tana, J. (1996). Transcriptional suppression of the transferrin gene by hypolipidemic peroxisome proliferators. *J Biol Chem* 271, 218-224.
- Hi, R., Osada, S., Yumoto, N., and Osumi, T. (1999). Characterization of the amino-terminal activation domain of peroxisome proliferator-activated receptor alpha. Importance of alpha-helical structure in the transactivating function. *J Biol Chem* 274, 35152-35158.
- Hierlihy, A. M., Cooke, G. M., Curran, I. H., Mehta, R., Karamanos, L., and Price, C. A. (2006). Effects of ciprofibrate on testicular and adrenal steroidogenic enzymes in the rat. *Reprod Toxicol* 22, 37-43.
- Hiipakka, R. A., and Liao, S. (1998). Molecular Mechanism of Androgen Action. *Trends in Endocrinology and Metabolism* 9, 317-324.
- Hijikata, M., Wen, J. K., Osumi, T., and Hashimoto, T. (1990). Rat peroxisomal 3-ketoacyl-CoA thiolase gene. Occurrence of two closely related but differentially regulated genes. *J Biol Chem* 265, 4600-4606.
- Hoerer, S., Schmid, A., Heckel, A., Budzinski, R. M., and Nar, H. (2003). Crystal structure of the human liver X receptor beta ligand-binding domain in complex with a synthetic agonist. *J Mol Biol* 334, 853-861.
- Hoffmann, B., Lehmann, J. M., Zhang, X. K., Hermann, T., Husmann, M., Graupner, G., and Pfahl, M. (1990). A retinoic acid receptor-specific element controls the retinoic acid receptor-beta promoter. *Mol Endocrinol* 4, 1727-1736.
- Holmbeck, S. M., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J., and Wright, P. E. (1998). High-resolution solution structure of the retinoid X receptor DNA-binding domain. *J Mol Biol* 281, 271-284.
- Holness, M. J., Bulmer, K., Gibbons, G. F., and Sugden, M. C. (2002a). Up-regulation of pyruvate dehydrogenase kinase isoform 4 (PDK4) protein expression in oxidative skeletal muscle does not require the obligatory participation of peroxisome-proliferator-activated receptor alpha (PPARalpha). *Biochem J* 366, 839-846.
- Holness, M. J., Smith, N. D., Bulmer, K., Hopkins, T., Gibbons, G. F., and Sugden, M. C. (2002b). Evaluation of the role of peroxisome-proliferator-activated receptor alpha in the regulation of cardiac pyruvate dehydrogenase kinase 4 protein expression in response to starvation, high-fat feeding and hyperthyroidism. *Biochem J* 364, 687-694.
- Holst, D., Luquet, S., Nogueira, V., Kristiansen, K., Leverve, X., and Grimaldi, P. A. (2003). Nutritional regulation and role of peroxisome proliferator-activated receptor delta in fatty acid catabolism in skeletal muscle. *Biochim Biophys Acta* 1633, 43-50.
- Honkakoski, P., Zelko, I., Sueyoshi, T., and Negishi, M. (1998). The nuclear orphan receptor CAR-retinoid X receptor heterodimer activates the phenobarbital-responsive enhancer module of the CYP2B gene. *Mol Cell Biol* 18, 5652-5658.
- Horlein, A. J., Naar, A. M., Heinzel, T., Torchia, J., Gloss, B., Kurokawa, R., Ryan, A., Kamei, Y., Soderstrom, M., Glass, C. K., and et al. (1995). Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* 377, 397-404.
- Horton, J. D., Goldstein, J. L., and Brown, M. S. (2002). SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J Clin Invest* 109, 1125-1131.
- Hostetler, H. A., Kier, A. B., and Schroeder, F. (2006). Very-long-chain and branched-chain fatty acyl-CoAs are high affinity ligands for the peroxisome proliferator-activated receptor alpha (PPARalpha). *Biochemistry* 45, 7669-7681.
- Hostetler, H. A., Petrescu, A. D., Kier, A. B., and Schroeder, F. (2005). Peroxisome proliferator-activated receptor alpha interacts with high affinity and is conformationally responsive to endogenous ligands. *J Biol Chem* 280, 18667-18682.
- Hsieh, R. W., Rajan, S. S., Sharma, S. K., Guo, Y., DeSombre, E. R., Mrksich, M., and Greene, G. L. (2006). Identification of ligands with bicyclic scaffolds provides insights into mechanisms of estrogen receptor subtype selectivity. *J Biol Chem* 281, 17909-17919.

- Hsu, M. H., Palmer, C. N., Song, W., Griffin, K. J., and Johnson, E. F. (1998). A carboxyl-terminal extension of the zinc finger domain contributes to the specificity and polarity of peroxisome proliferator-activated receptor DNA binding. *J Biol Chem* 273, 27988-27997.
- Htun, H., Barsony, J., Renyi, I., Gould, D. L., and Hager, G. L. (1996). Visualization of glucocorticoid receptor translocation and intranuclear organization in living cells with a green fluorescent protein chimera. *Proc Natl Acad Sci U S A* 93, 4845-4850.
- Hu, X., Li, Y., and Lazar, M. A. (2001). Determinants of CoRNR-dependent repression complex assembly on nuclear hormone receptors. *Mol Cell Biol* 21, 1747-1758.
- Huang, H., Starodub, O., McIntosh, A., Atshaves, B. P., Woldegiorgis, G., Kier, A. B., and Schroeder, F. (2004). Liver fatty acid-binding protein colocalizes with peroxisome proliferator activated receptor alpha and enhances ligand distribution to nuclei of living cells. *Biochemistry* 43, 2484-2500.
- Huang, H., Starodub, O., McIntosh, A., Kier, A. B., and Schroeder, F. (2002). Liver fatty acid-binding protein targets fatty acids to the nucleus. Real time confocal and multiphoton fluorescence imaging in living cells. *J Biol Chem* 277, 29139-29151.
- Huang, N., vom Baur, E., Garnier, J. M., Lerouge, T., Vonesch, J. L., Lutz, Y., Chambon, P., and Losson, R. (1998). Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *Embo J* 17, 3398-3412.
- Huang, Q., Alvares, K., Chu, R., Bradfield, C. A., and Reddy, J. K. (1994). Association of peroxisome proliferator-activated receptor and Hsp72. *J Biol Chem* 269, 8493-8497.
- Huang, W., Zhang, J., Washington, M., Liu, J., Parant, J. M., Lozano, G., and Moore, D. D. (2005). Xenobiotic stress induces hepatomegaly and liver tumors via the nuclear receptor constitutive androstane receptor. *Mol Endocrinol* 19, 1646-1653.
- Huber, B. R., Desclozeaux, M., West, B. L., Cunha-Lima, S. T., Nguyen, H. T., Baxter, J. D., Ingraham, H. A., and Fletterick, R. J. (2003a). Thyroid hormone receptor-beta mutations conferring hormone resistance and reduced corepressor release exhibit decreased stability in the N-terminal ligand-binding domain. *Mol Endocrinol* 17, 107-116.
- Huber, B. R., Sandler, B., West, B. L., Cunha Lima, S. T., Nguyen, H. T., Apriletti, J. W., Baxter, J. D., and Fletterick, R. J. (2003b). Two resistance to thyroid hormone mutants with impaired hormone binding. *Mol Endocrinol* 17, 643-652.
- Hughes, T. R., Mao, M., Jones, A. R., Burchard, J., Marton, M. J., Shannon, K. W., Lefkowitz, S. M., Ziman, M., Schelter, J. M., Meyer, M. R., et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19, 342-347.
- Humasti, S., and Tontonoz, P. (2006). The peroxisome proliferator-activated receptor N-terminal domain controls isotype-selective gene expression and adipogenesis. *Mol Endocrinol* 20, 1261-1275.
- Hummel, C. W., Geiser, A. G., Bryant, H. U., Cohen, I. R., Dally, R. D., Fong, K. C., Frank, S. A., Hinklin, R., Jones, S. A., Lewis, G., et al. (2005). A selective estrogen receptor modulator designed for the treatment of uterine leiomyoma with unique tissue specificity for uterus and ovaries in rats. *J Med Chem* 48, 6772-6775.
- Hunt, M. C., and Alexson, S. E. (2002). The role Acyl-CoA thioesterases play in mediating intracellular lipid metabolism. *Prog Lipid Res* 41, 99-130.
- Hunt, M. C., Lindquist, P. J., Peters, J. M., Gonzalez, F. J., Diczfalusy, U., and Alexson, S. E. (2000a). Involvement of the peroxisome proliferator-activated receptor alpha in regulating long-chain acyl-CoA thioesterases. *J Lipid Res* 41, 814-823.
- Hunt, M. C., Nousiainen, S. E., Huttunen, M. K., Orii, K. E., Svensson, L. T., and Alexson, S. E. (1999). Peroxisome proliferator-induced long chain acyl-CoA thioesterases comprise a highly conserved novel multi-gene family involved in lipid metabolism. *J Biol Chem* 274, 34317-34326.
- Hunt, M. C., Solaas, K., Kase, B. F., and Alexson, S. E. (2002). Characterization of an acyl-coA thioesterase that functions as a major regulator of peroxisomal lipid metabolism. *J Biol Chem* 277, 1128-1138.
- Hunt, M. C., Yamada, J., Maltais, L. J., Wright, M. W., Podesta, E. J., and Alexson, S. E. (2005). A revised nomenclature for mammalian acyl-CoA thioesterases/hydrolases. *J Lipid Res* 46, 2029-2032.
- Hunt, M. C., Yang, Y. Z., Eggertsen, G., Carneheim, C. M., Gafvels, M., Einarsson, C., and Alexson, S. E. (2000b). The peroxisome proliferator-activated receptor alpha (PPARalpha) regulates bile acid biosynthesis. *J Biol Chem* 275, 28947-28953.
- Hurd, C., and Moudgil, V. K. (1988). Characterization of R5020 and RU486 binding to progesterone receptor from calf uterus. *Biochemistry* 27, 3618-3623.
- Hurst, C. H., and Waxman, D. J. (2004). Environmental phthalate monoesters activate pregnane X receptor-mediated transcription. *Toxicol Appl Pharmacol* 199, 266-274.
- Huss, J. M., and Kelly, D. P. (2004). Nuclear receptor signaling and cardiac energetics. *Circ Res* 95, 568-578.

- Huth, W., Jonas, R., Wunderlich, I., and Seubert, W. (1975). On the mechanism of ketogenesis and its control. Purification, kinetic mechanism and regulation of different forms of mitochondrial acetoacetyl-CoA thiolases from ox liver. *Eur J Biochem* 59, 475-489.
- Icre, G., Wahli, W., and Michalik, L. (2006). Functions of the peroxisome proliferator-activated receptor (PPAR) alpha and beta in skin homeostasis, epithelial repair, and morphogenesis. *J Investig Dermatol Symp Proc* 11, 30-35.
- IJppenberg, A., Jeannin, E., Wahli, W., and Desvergne, B. (1997). Polarity and specific sequence requirements of peroxisome proliferator-activated receptor (PPAR)/retinoid X receptor heterodimer binding to DNA. A functional analysis of the malic enzyme gene PPAR response element. *J Biol Chem* 272, 20108-20117.
- IJppenberg, A., Tan, N. S., Gelman, L., Kersten, S., Seydoux, J., Xu, J., Metzger, D., Canaple, L., Chambon, P., Wahli, W., and Desvergne, B. (2004). In vivo activation of PPAR target genes by RXR homodimers. *Embo J* 23, 2083-2091.
- Inoue, I., Shino, K., Noji, S., Awata, T., and Katayama, S. (1998). Expression of peroxisome proliferator-activated receptor alpha (PPAR alpha) in primary cultures of human vascular endothelial cells. *Biochem Biophys Res Commun* 246, 370-374.
- Inui, S., Lee, Y. F., Chang, E., Shyr, C. R., and Chang, C. (2003). Differential and bi-directional regulation between TR2/TR4 orphan nuclear receptors and a specific ligand mediated-peroxisome proliferator-activated receptor alpha in human HaCaT keratinocytes. *J Dermatol Sci* 31, 65-71.
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249-264.
- Issemann, I., and Green, S. (1990). Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. *Nature* 347, 645-650.
- Issemann, I., Prince, R., Tugwood, J., and Green, S. (1992). A role for fatty acids and liver fatty acid binding protein in peroxisome proliferation? *Biochem Soc Trans* 20, 824-827.
- Ito, M., Yuan, C. X., Okano, H. J., Darnell, R. B., and Roeder, R. G. (2000). Involvement of the TRAP220 component of the TRAP/SMCC coactivator complex in embryonic development and thyroid hormone action. *Mol Cell* 5, 683-693.
- Iwaki, M., Matsuda, M., Maeda, N., Funahashi, T., Matsuzawa, Y., Makishima, M., and Shimomura, I. (2003). Induction of adiponectin, a fat-derived antidiabetic and antiatherogenic factor, by nuclear receptors. *Diabetes* 52, 1655-1663.
- Izzedine, H., Launay-Vacher, V., Baumelou, A., and Deray, G. (2004). Renal effects of PPARalpha-agonists. *Minerva Urol Nefrol* 56, 339-342.
- Jackson, J. P., Ferguson, S. S., Moore, R., Negishi, M., and Goldstein, J. A. (2004). The constitutive active/androstane receptor regulates phenytoin induction of Cyp2c29. *Mol Pharmacol* 65, 1397-1404.
- Jackson, S. M., Parhami, F., Xi, X. P., Berliner, J. A., Hsueh, W. A., Law, R. E., and Demer, L. L. (1999). Peroxisome proliferator-activated receptor activators target human endothelial cells to inhibit leukocyte-endothelial cell interaction. *Arterioscler Thromb Vasc Biol* 19, 2094-2104.
- Jakel, H., Nowak, M., Moitrot, E., Dehondt, H., Hum, D. W., Pennacchio, L. A., Fruchart-Najib, J., and Fruchart, J. C. (2004). The liver X receptor ligand T0901317 down-regulates APOA5 gene expression through activation of SREBP-1c. *J Biol Chem* 279, 45462-45469.
- Janowski, B. A., Grogan, M. J., Jones, S. A., Wisely, G. B., Kliewer, S. A., Corey, E. J., and Mangelsdorf, D. J. (1999). Structural requirements of ligands for the oxysterol liver X receptors LXRAalpha and LXRBeta. *Proc Natl Acad Sci U S A* 96, 266-271.
- Janssen, U., and Stoffel, W. (2002). Disruption of mitochondrial beta -oxidation of unsaturated fatty acids in the 3,2-trans-enoyl-CoA isomerase-deficient mouse. *J Biol Chem* 277, 19579-19584.
- Javitt, N. B. (2002). 25R,26-Hydroxycholesterol revisited: synthesis, metabolism, and biologic roles. *J Lipid Res* 43, 665-670.
- Jaye, M. C., Krawiec, J. A., Campobasso, N., Smallwood, A., Qiu, C., Lu, Q., Kerrigan, J. J., De Los Frailes Alvaro, M., Laffitte, B., Liu, W. S., et al. (2005). Discovery of substituted maleimides as liver X receptor agonists and determination of a ligand-bound crystal structure. *J Med Chem* 48, 5419-5422.
- Jedidi, I., Couturier, M., Therond, P., Gardes-Albert, M., Legrand, A., Barouki, R., Bonnefont-Rousselot, D., and Aggerbeck, M. (2006). Cholestryler ester hydroperoxides increase macrophage CD36 gene expression via PPARalpha. *Biochem Biophys Res Commun* 351, 733-738.
- Jepsen, K., and Rosenfeld, M. G. (2002). Biological roles and mechanistic actions of co-repressor complexes. *J Cell Sci* 115, 689-698.
- Jia, Y., Guo, G. L., Surapureddi, S., Sarkar, J., Qi, C., Guo, D., Xia, J., Kashireddi, P., Yu, S., Cho, Y. W., et al. (2005). Transcription coactivator peroxisome proliferator-activated receptor-binding protein/mediator 1 deficiency abrogates acetaminophen hepatotoxicity. *Proc Natl Acad Sci U S A* 102, 12531-12536.

- Jia, Y., Qi, C., Kashireddi, P., Surapureddi, S., Zhu, Y. J., Rao, M. S., Le Roith, D., Chambon, P., Gonzalez, F. J., and Reddy, J. K. (2004). Transcription coactivator PBP, the peroxisome proliferator-activated receptor (PPAR)-binding protein, is required for PPARalpha-regulated gene expression in liver. *J Biol Chem* 279, 24427-24434.
- Jia, Y., Qi, C., Zhang, Z., Hashimoto, T., Rao, M. S., Huyghe, S., Suzuki, Y., Van Veldhoven, P. P., Baes, M., and Reddy, J. K. (2003). Overexpression of peroxisome proliferator-activated receptor-alpha (PPARalpha)-regulated genes in liver in the absence of peroxisome proliferation in mice deficient in both L- and D-forms of enoyl-CoA hydratase/dehydrogenase enzymes of peroxisomal beta-oxidation system. *J Biol Chem* 278, 47232-47239.
- Johnson, E. F., Palmer, C. N., Griffin, K. J., and Hsu, M. H. (1996). Role of the peroxisome proliferator-activated receptor in cytochrome P450 4A gene regulation. *Faseb J* 10, 1241-1248.
- Jones, D. C., Ding, X., and Daynes, R. A. (2002). Nuclear receptor peroxisome proliferator-activated receptor alpha (PPARalpha) is expressed in resting murine lymphocytes. The PPARalpha in T and B lymphocytes is both transactivation and transrepression competent. *J Biol Chem* 277, 6838-6845.
- Jones, S. A., Moore, L. B., Shenk, J. L., Wisely, G. B., Hamilton, G. A., McKee, D. D., Tomkinson, N. C., LeCluyse, E. L., Lambert, M. H., Willson, T. M., et al. (2000). The pregnane X receptor: a promiscuous xenobiotic receptor that has diverged during evolution. *Mol Endocrinol* 14, 27-39.
- Jonkers, I. J., Smelt, A. H., Hattori, H., Scheek, L. M., van Gent, T., de Man, F. H., van der Laarse, A., and van Tol, A. (2003). Decreased PLTP mass but elevated PLTP activity linked to insulin resistance in HTG: effects of bezafibrate therapy. *J Lipid Res* 44, 1462-1469.
- Jordan, B. R. (1998). Large-scale expression measurement by hybridization methods: from high-density membranes to "DNA chips". *J Biochem (Tokyo)* 124, 251-258.
- Jordan, V. C. (2004). Selective estrogen receptor modulation: concept and consequences in cancer. *Cancer Cell* 5, 207-213.
- Joseph, S. B., Laffitte, B. A., Patel, P. H., Watson, M. A., Matsukuma, K. E., Walczak, R., Collins, J. L., Osborne, T. F., and Tontonoz, P. (2002). Direct and indirect mechanisms for regulation of fatty acid synthase gene expression by liver X receptors. *J Biol Chem* 277, 11019-11025.
- Juge-Aubry, C., Pernin, A., Favez, T., Burger, A. G., Wahli, W., Meier, C. A., and Desvergne, B. (1997). DNA binding properties of peroxisome proliferator-activated receptor subtypes on various natural peroxisome proliferator response elements. Importance of the 5'-flanking region. *J Biol Chem* 272, 25252-25259.
- Juge-Aubry, C. E., Hammar, E., Siegrist-Kaiser, C., Pernin, A., Takeshita, A., Chin, W. W., Burger, A. G., and Meier, C. A. (1999). Regulation of the transcriptional activity of the peroxisome proliferator-activated receptor alpha by phosphorylation of a ligand-independent trans-activating domain. *J Biol Chem* 274, 10505-10510.
- Juge-Aubry, C. E., Kuenzli, S., Sanchez, J. C., Hochstrasser, D., and Meier, C. A. (2001). Peroxisomal bifunctional enzyme binds and activates the activation function-1 region of the peroxisome proliferator-activated receptor alpha. *Biochem J* 353, 253-258.
- Jump, D. B., Botolin, D., Wang, Y., Xu, J., Christian, B., and Demeure, O. (2005). Fatty acid regulation of hepatic gene transcription. *J Nutr* 135, 2503-2506.
- Jung, D., Fried, M., and Kullak-Ublick, G. A. (2002). Human apical sodium-dependent bile salt transporter gene (SLC10A2) is regulated by the peroxisome proliferator-activated receptor alpha. *J Biol Chem* 277, 30559-30566.
- Kadowaki, T., Yamauchi, T., Kubota, N., Hara, K., Ueki, K., and Tobe, K. (2006). Adiponectin and adiponectin receptors in insulin resistance, diabetes, and the metabolic syndrome. *J Clin Invest* 116, 1784-1792.
- Kaikaus, R. M., Chan, W. K., Lysenko, N., Ray, R., Ortiz de Montellano, P. R., and Bass, N. M. (1993). Induction of peroxisomal fatty acid beta-oxidation and liver fatty acid-binding protein by peroxisome proliferators. Mediation via the cytochrome P-450IVA1 omega-hydroxylase pathway. *J Biol Chem* 268, 9593-9603.
- Kallen, J., Schlaeppli, J. M., Bitsch, F., Delhon, I., and Fournier, B. (2004a). Crystal structure of the human RORalpha Ligand binding domain in complex with cholesterol sulfate at 2.2 Å. *J Biol Chem* 279, 14033-14038.
- Kallen, J., Schlaeppli, J. M., Bitsch, F., Filipuzzi, I., Schilb, A., Riou, V., Graham, A., Strauss, A., Geiser, M., and Fournier, B. (2004b). Evidence for ligand-independent transcriptional activation of the human estrogen-related receptor alpha (ERRalpha): crystal structure of ERRalpha ligand binding domain in complex with peroxisome proliferator-activated receptor coactivator-1alpha. *J Biol Chem* 279, 49330-49337.
- Kallen, J. A., Schlaeppli, J. M., Bitsch, F., Geisse, S., Geiser, M., Delhon, I., and Fournier, B. (2002). X-ray structure of the hRORalpha LBD at 1.63 Å: structural and functional data that cholesterol or a cholesterol derivative is the natural ligand of RORalpha. *Structure* 10, 1697-1707.
- Kassam, A., Winrow, C. J., Fernandez-Rachubinski, F., Capone, J. P., and Rachubinski, R. A. (2000). The peroxisome proliferator response element of the gene encoding the peroxisomal beta-oxidation enzyme enoyl-

CoA hydratase/3-hydroxyacyl-CoA dehydrogenase is a target for constitutive androstane receptor beta/9-cis-retinoic acid receptor-mediated transactivation. *J Biol Chem* 275, 4345-4350.

Kast, H. R., Goodwin, B., Tarr, P. T., Jones, S. A., Anisfeld, A. M., Stoltz, C. M., Tontonoz, P., Kliewer, S., Willson, T. M., and Edwards, P. A. (2002). Regulation of multidrug resistance-associated protein 2 (ABCC2) by the nuclear receptors pregnane X receptor, farnesoid X-activated receptor, and constitutive androstane receptor. *J Biol Chem* 277, 2908-2915.

Katahira, M., Knegtel, R. M., Boelens, R., Eib, D., Schilthuis, J. G., van der Saag, P. T., and Kaptein, R. (1992). Homo- and heteronuclear NMR studies of the human retinoic acid receptor beta DNA-binding domain: sequential assignments and identification of secondary structure elements. *Biochemistry* 31, 6474-6480.

Kauppi, B., Jakob, C., Farnegardh, M., Yang, J., Ahola, H., Alarcon, M., Calles, K., Engstrom, O., Harlan, J., Muchmore, S., et al. (2003). The three-dimensional structures of antagonistic and agonistic forms of the glucocorticoid receptor ligand-binding domain: RU-486 induces a transconformation that leads to active antagonism. *J Biol Chem* 278, 22748-22754.

Kawamoto, T., Sueyoshi, T., Zelko, I., Moore, R., Washburn, K., and Negishi, M. (1999). Phenobarbital-responsive nuclear translocation of the receptor CAR in induction of the CYP2B gene. *Mol Cell Biol* 19, 6318-6322.

Kawana, K., Ikuta, T., Kobayashi, Y., Gotoh, O., Takeda, K., and Kawajiri, K. (2003). Molecular mechanism of nuclear translocation of an orphan nuclear receptor, SXR. *Mol Pharmacol* 63, 524-531.

Kawashima, H., Takano, H., Sugita, S., Takahara, Y., Sugimura, K., and Nakatani, T. (2003). A novel steroid receptor co-activator protein (SRAP) as an alternative form of steroid receptor RNA-activator gene: expression in prostate cancer cells and enhancement of androgen receptor activity. *Biochem J* 369, 163-171.

Keidel, S., Rupp, E., and Szardenings, M. (1992). Recombinant human retinoic acid receptor alpha. Binding of DNA and synthetic retinoids to the protein expressed in Escherichia coli. *Eur J Biochem* 204, 1141-1148.

Keller, H., Dreyer, C., Medin, J., Mahfoudi, A., Ozato, K., and Wahli, W. (1993). Fatty acids and retinoids control lipid metabolism through activation of peroxisome proliferator-activated receptor-retinoid X receptor heterodimers. *Proc Natl Acad Sci U S A* 90, 2160-2164.

Keller, H., Givel, F., Perroud, M., and Wahli, W. (1995). Signaling cross-talk between peroxisome proliferator-activated receptor/retinoid X receptor and estrogen receptor through estrogen response elements. *Mol Endocrinol* 9, 794-804.

Kemper, J. K., Kim, H., Miao, J., Bhalla, S., and Bae, Y. (2004). Role of an mSin3A-Swi/Snf chromatin remodeling complex in the feedback repression of bile acid biosynthesis by SHP. *Mol Cell Biol* 24, 7707-7719.

Kerr, M. K., Afshari, C. A., Bennett, L., Bushel, P., Martinez, J., Walker, N. J., and Churchill, G. A. (2002a). Statistical Analysis of a Gene Expression Microarray Experiment with Replication. *Statistica Sinica* 12, 203-217.

Kerr, M. K., and Churchill, G. A. (2001a). Experimental design for gene expression microarrays. *Biostatistics* 2, 183-201.

Kerr, M. K., and Churchill, G. A. (2001b). Statistical design and the analysis of gene expression microarray data. *Genet Res* 77, 123-128.

Kerr, M. K., Martin, M., and Churchill, G. A. (2000). Analysis of variance for gene expression microarray data. *J Comput Biol* 7, 819-837.

Kerr, T. A., Saeki, S., Schneider, M., Schaefer, K., Berdy, S., Redder, T., Shan, B., Russell, D. W., and Schwarz, M. (2002b). Loss of nuclear receptor SHP impairs but does not eliminate negative feedback regulation of bile acid synthesis. *Dev Cell* 2, 713-720.

Kersten, S. (2002). Peroxisome proliferator activated receptors and obesity. *Eur J Pharmacol* 440, 223-234.

Kersten, S., Dong, D., Lee, W., Reczek, P. R., and Noy, N. (1998). Auto-silencing by the retinoid X receptor. *J Mol Biol* 284, 21-32.

Kersten, S., Kelleher, D., Chambon, P., Gronemeyer, H., and Noy, N. (1995a). Retinoid X receptor alpha forms tetramers in solution. *Proc Natl Acad Sci U S A* 92, 8645-8649.

Kersten, S., Mandard, S., Escher, P., Gonzalez, F. J., Tafuri, S., Desvergne, B., and Wahli, W. (2001). The peroxisome proliferator-activated receptor alpha regulates amino acid metabolism. *Faseb J* 15, 1971-1978.

Kersten, S., Pan, L., Chambon, P., Gronemeyer, H., and Noy, N. (1995b). Role of ligand in retinoid signaling. 9-cis-retinoic acid modulates the oligomeric state of the retinoid X receptor. *Biochemistry* 34, 13717-13721.

Kersten, S., Pan, L., and Noy, N. (1995c). On the role of ligand in retinoid signaling: positive cooperativity in the interactions of 9-cis retinoic acid with tetramers of the retinoid X receptor. *Biochemistry* 34, 14263-14269.

Kersten, S., Reczek, P. R., and Noy, N. (1997). The tetramerization region of the retinoid X receptor is important for transcriptional activation by the receptor. *J Biol Chem* 272, 29759-29768.

Kersten, S., Seydoux, J., Peters, J. M., Gonzalez, F. J., Desvergne, B., and Wahli, W. (1999). Peroxisome proliferator-activated receptor alpha mediates the adaptive response to fasting. *J Clin Invest* 103, 1489-1498.

- Kim, J. B., Sarraf, P., Wright, M., Yao, K. M., Mueller, E., Solanes, G., Lowell, B. B., and Spiegelman, B. M. (1998a). Nutritional and insulin regulation of fatty acid synthetase and leptin gene expression through ADD1/SREBP1. *J Clin Invest* 101, 1-9.
- Kim, J. B., Wright, H. M., Wright, M., and Spiegelman, B. M. (1998b). ADD1/SREBP1 activates PPARgamma through the production of endogenous ligand. *Proc Natl Acad Sci U S A* 95, 4333-4337.
- Kim, M. Y., Woo, E. M., Chong, Y. T., Homenko, D. R., and Kraus, W. L. (2006). Acetylation of estrogen receptor alpha by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor. *Mol Endocrinol* 20, 1479-1493.
- Kim, S., Wu, J. Y., Birzin, E. T., Frisch, K., Chan, W., Pai, L. Y., Yang, Y. T., Mosley, R. T., Fitzgerald, P. M., Sharma, N., et al. (2004). Estrogen receptor ligands. II. Discovery of benzoxathiins as potent, selective estrogen receptor alpha modulators. *J Med Chem* 47, 2171-2175.
- Kishimoto, M., Fujiki, R., Takezawa, S., Sasaki, Y., Nakamura, T., Yamaoka, K., Kitagawa, H., and Kato, S. (2006). Nuclear receptor mediated gene regulation through chromatin remodeling and histone modifications. *Endocr J* 53, 157-172.
- Klaholz, B., and Moras, D. (2002). C-H...O hydrogen bonds in the nuclear receptor RARgamma--a potential tool for drug selectivity. *Structure* 10, 1197-1204.
- Klaholz, B. P., Mitschler, A., Belema, M., Zusi, C., and Moras, D. (2000a). Enantiomer discrimination illustrated by high-resolution crystal structures of the human nuclear receptor hRARgamma. *Proc Natl Acad Sci U S A* 97, 6322-6327.
- Klaholz, B. P., Mitschler, A., and Moras, D. (2000b). Structural basis for isotype selectivity of the human retinoic acid nuclear receptor. *J Mol Biol* 302, 155-170.
- Klaholz, B. P., Renaud, J. P., Mitschler, A., Zusi, C., Chambon, P., Gronemeyer, H., and Moras, D. (1998). Conformational adaptation of agonists to the human nuclear receptor RAR gamma. *Nat Struct Biol* 5, 199-202.
- Kleemann, R., Gervois, P. P., Verschuren, L., Staels, B., Princen, H. M., and Kooistra, T. (2003). Fibrates down-regulate IL-1-stimulated C-reactive protein gene expression in hepatocytes by reducing nuclear p50-NFkappa B-C/EBP-beta complex formation. *Blood* 101, 545-551.
- Kleemann, R., Verschuren, L., de Rooij, B. J., Lindeman, J., de Maat, M. M., Szalai, A. J., Princen, H. M., and Kooistra, T. (2004). Evidence for anti-inflammatory activity of statins and PPARalpha activators in human C-reactive protein transgenic mice in vivo and in cultured human hepatocytes in vitro. *Blood* 103, 4188-4194.
- Kliewer, S. A., Forman, B. M., Blumberg, B., Ong, E. S., Borgmeyer, U., Mangelsdorf, D. J., Umesono, K., and Evans, R. M. (1994). Differential expression and activation of a family of murine peroxisome proliferator-activated receptors. *Proc Natl Acad Sci U S A* 91, 7355-7359.
- Kliewer, S. A., Moore, J. T., Wade, L., Staudinger, J. L., Watson, M. A., Jones, S. A., McKee, D. D., Oliver, B. B., Willson, T. M., Zetterstrom, R. H., et al. (1998). An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell* 92, 73-82.
- Kliewer, S. A., Sundseth, S. S., Jones, S. A., Brown, P. J., Wisely, G. B., Koble, C. S., Devchand, P., Wahli, W., Willson, T. M., Lenhard, J. M., and Lehmann, J. M. (1997). Fatty acids and eicosanoids regulate gene expression through direct interactions with peroxisome proliferator-activated receptors alpha and gamma. *Proc Natl Acad Sci U S A* 94, 4318-4323.
- Kliewer, S. A., Umesono, K., Noonan, D. J., Heyman, R. A., and Evans, R. M. (1992). Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors. *Nature* 358, 771-774.
- Klinge, C. M., Jernigan, S. C., Risinger, K. E., Lee, J. E., Tyulmenkov, V. V., Falkner, K. C., and Prough, R. A. (2001). Short heterodimer partner (SHP) orphan nuclear receptor inhibits the transcriptional activity of aryl hydrocarbon receptor (AHR)/AHR nuclear translocator (ARNT). *Arch Biochem Biophys* 390, 64-70.
- Knauf, C., Rieusset, J., Foretz, M., Cani, P. D., Uldry, M., Hosokawa, M., Martinez, E., Bringart, M., Waget, A., Kersten, S., et al. (2006). Peroxisome proliferator-activated receptor-alpha-null mice have increased white adipose tissue glucose utilization, GLUT4, and fat mass: Role in liver and brain. *Endocrinology* 147, 4067-4078.
- Knight, B. L., Hebbachi, A., Hauton, D., Brown, A. M., Wiggins, D., Patel, D. D., and Gibbons, G. F. (2005). A role for PPARalpha in the control of SREBP activity and lipid synthesis in the liver. *Biochem J* 389, 413-421.
- Knight, B. L., Patel, D. D., Humphreys, S. M., Wiggins, D., and Gibbons, G. F. (2003). Inhibition of cholesterol absorption associated with a PPAR alpha-dependent increase in ABC binding cassette transporter A1 in mice. *J Lipid Res* 44, 2049-2058.
- Kobayashi, K., Sueyoshi, T., Inoue, K., Moore, R., and Negishi, M. (2003). Cytoplasmic accumulation of the nuclear receptor CAR by a tetratricopeptide repeat protein in HepG2 cells. *Mol Pharmacol* 64, 1069-1075.
- Kohalmy, K., Tamasi, V., Kobori, L., Sarvary, E., Pascussi, J. M., Porrogi, P., Rozman, D., Prough, R. A., Meyer, U. A., and Monostory, K. (2007). Dehydroepiandrosterone induces human CYP2B6 through the constitutive androstane receptor. *Drug Metab Dispos*.

- Kojo, A., Pellinen, P., Juvonen, R., Raunio, H., Pelkonen, O., and Pasanen, M. (1996). Distinct responses of mouse hepatic CYP enzymes to corn oil and peroxisome proliferators. *Biochem Pharmacol* 51, 1137-1143.
- Kok, T., Bloks, V. W., Wolters, H., Havinga, R., Jansen, P. L., Staels, B., and Kuipers, F. (2003a). Peroxisome proliferator-activated receptor alpha (PPARalpha)-mediated regulation of multidrug resistance 2 (Mdr2) expression and function in mice. *Biochem J* 369, 539-547.
- Kok, T., Wolters, H., Bloks, V. W., Havinga, R., Jansen, P. L., Staels, B., and Kuipers, F. (2003b). Induction of hepatic ABC transporter expression is part of the PPARalpha-mediated fasting response in the mouse. *Gastroenterology* 124, 160-171.
- Kong, E. H., Heldring, N., Gustafsson, J. A., Treuter, E., Hubbard, R. E., and Pike, A. C. (2005). Delineation of a unique protein-protein interaction site on the surface of the estrogen receptor. *Proc Natl Acad Sci U S A* 102, 3593-3598.
- Koo, S. H., Satoh, H., Herzig, S., Lee, C. H., Hedrick, S., Kulkarni, R., Evans, R. M., Olefsky, J., and Montminy, M. (2004). PGC-1 promotes insulin resistance in liver through PPAR-alpha-dependent induction of TRB-3. *Nat Med* 10, 530-534.
- Krey, G., Braissant, O., L'Horset, F., Kalkhoven, E., Perroud, M., Parker, M. G., and Wahli, W. (1997). Fatty acids, eicosanoids, and hypolipidemic agents identified as ligands of peroxisome proliferator-activated receptors by coactivator-dependent receptor ligand assay. *Mol Endocrinol* 11, 779-791.
- Krey, G., Keller, H., Mahfoudi, A., Medin, J., Ozato, K., Dreyer, C., and Wahli, W. (1993). Xenopus peroxisome proliferator activated receptors: genomic organization, response element recognition, heterodimer formation with retinoid X receptor and activation by fatty acids. *J Steroid Biochem Mol Biol* 47, 65-73.
- Kroetz, D. L., Yook, P., Costet, P., Bianchi, P., and Pineau, T. (1998). Peroxisome proliferator-activated receptor alpha controls the hepatic CYP4A induction adaptive response to starvation and diabetes. *J Biol Chem* 273, 31581-31589.
- Krogsdam, A. M., Nielsen, C. A., Neve, S., Holst, D., Helledie, T., Thomsen, B., Bendixen, C., Mandrup, S., and Kristiansen, K. (2002). Nuclear receptor corepressor-dependent repression of peroxisome-proliferator-activated receptor delta-mediated transactivation. *Biochem J* 363, 157-165.
- Krylova, I. N., Sablin, E. P., Moore, J., Xu, R. X., Waitt, G. M., MacKay, J. A., Juzumiene, D., Bynum, J. M., Madauss, K., Montana, V., et al. (2005). Structural analyses reveal phosphatidyl inositol as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* 120, 343-355.
- Kuhn, C. M. (2002). Anabolic steroids. *Recent Prog Horm Res* 57, 411-434.
- Kuiper, G. G., Carlsson, B., Grandien, K., Enmark, E., Hagglad, J., Nilsson, S., and Gustafsson, J. A. (1997). Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology* 138, 863-870.
- Kuiper, G. G., de Ruiter, P. E., Trapman, J., Jenster, G., and Brinkmann, A. O. (1993). In vitro translation of androgen receptor cRNA results in an activated androgen receptor protein. *Biochem J* 296 ( Pt 1), 161-167.
- Kuiper, G. G., Enmark, E., Pelto-Huikko, M., Nilsson, S., and Gustafsson, J. A. (1996). Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci U S A* 93, 5925-5930.
- Kuiper, G. G., Lemmen, J. G., Carlsson, B., Corton, J. C., Safe, S. H., van der Saag, P. T., van der Burg, B., and Gustafsson, J. A. (1998). Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. *Endocrinology* 139, 4252-4263.
- Kumar, R., Baskakov, I. V., Srinivasan, G., Bolen, D. W., Lee, J. C., and Thompson, E. B. (1999). Interdomain signaling in a two-domain fragment of the human glucocorticoid receptor. *J Biol Chem* 274, 24737-24741.
- Kumar, R., and Thompson, E. B. (2003). Transactivation functions of the N-terminal domains of nuclear hormone receptors: protein folding and coactivator interactions. *Mol Endocrinol* 17, 1-10.
- LaFave, L. T., Augustin, L. B., and Mariash, C. N. (2006). S14: insights from knockout mice. *Endocrinology* 147, 4044-4047.
- Laffitte, B. A., Joseph, S. B., Walczak, R., Pei, L., Wilpitz, D. C., Collins, J. L., and Tontonoz, P. (2001a). Autoregulation of the human liver X receptor alpha promoter. *Mol Cell Biol* 21, 7558-7568.
- Laffitte, B. A., Repa, J. J., Joseph, S. B., Wilpitz, D. C., Kast, H. R., Mangelsdorf, D. J., and Tontonoz, P. (2001b). LXRs control lipid-inducible expression of the apolipoprotein E gene in macrophages and adipocytes. *Proc Natl Acad Sci U S A* 98, 507-512.
- Lambert, G., Jarnoux, A. L., Pineau, T., Pape, O., Chetiveaux, M., Laboisson, C., Krempf, M., and Costet, P. (2006). Fasting induces hyperlipidemia in mice overexpressing proprotein convertase subtilisin kexin type 9: lack of modulation of very-low-density lipoprotein hepatic output by the low-density lipoprotein receptor. *Endocrinology* 147, 4985-4995.
- Landrier, J. F., Thomas, C., Grober, J., Duez, H., Percevault, F., Souidi, M., Linard, C., Staels, B., and Besnard, P. (2004). Statin induction of liver fatty acid-binding protein (L-FABP) gene expression is peroxisome proliferator-activated receptor-alpha-dependent. *J Biol Chem* 279, 45512-45518.

- Lanvers, C., Hempel, G., Blaschke, G., and Boos, J. (1996). Simultaneous determination of all-trans-, 13-cis- and 9-cis-retinoic acid, their 4-oxo metabolites and all-trans-retinol in human plasma by high-performance liquid chromatography. *J Chromatogr B Biomed Appl* 685, 233-240.
- Lanz, R. B., Chua, S. S., Barron, N., Soder, B. M., DeMayo, F., and O'Malley, B. W. (2003). Steroid receptor RNA activator stimulates proliferation as well as apoptosis in vivo. *Mol Cell Biol* 23, 7163-7176.
- Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97, 17-27.
- Lanz, R. B., Razani, B., Goldberg, A. D., and O'Malley, B. W. (2002). Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc Natl Acad Sci U S A* 99, 16081-16086.
- Lapinskas, P. J., Brown, S., Leesnitzer, L. M., Blanchard, S., Swanson, C., Cattley, R. C., and Corton, J. C. (2005). Role of PPARalpha in mediating the effects of phthalates and metabolites in the liver. *Toxicology* 207, 149-163.
- Laudet, V., and Gronemeyer, H. (2002). The nuclear receptor : factsbook (San Diego: Academic Press).
- Lawrence, J. W., Li, Y., Chen, S., DeLuca, J. G., Berger, J. P., Umbenhauer, D. R., Moller, D. E., and Zhou, G. (2001). Differential gene regulation in human versus rodent hepatocytes by peroxisome proliferator-activated receptor (PPAR) alpha. PPAR alpha fails to induce peroxisome proliferation-associated genes in human cells independently of the level of receptor expression. *J Biol Chem* 276, 31521-31527.
- Lazennec, G., Canaple, L., Saugy, D., and Wahli, W. (2000). Activation of peroxisome proliferator-activated receptors (PPARs) by their ligands and protein kinase A activators. *Mol Endocrinol* 14, 1962-1975.
- Le Jossic-Corcos, C., Duclos, S., Ramirez, L. C., Zaghini, I., Chevillard, G., Martin, P., Pineau, T., and Bournot, P. (2004). Effects of peroxisome proliferator-activated receptor alpha activation on pathways contributing to cholesterol homeostasis in rat hepatocytes. *Biochim Biophys Acta* 1683, 49-58.
- Le Maho, Y., and Groscolas, R. (1990). Modalités et limites d'utilisation des réserves énergétiques au cours du jeûne prolongé. *Cah Nutr Diét* 25, 181-186.
- Le May, C., Pineau, T., Bigot, K., Kohl, C., Girard, J., and Pegorier, J. P. (2000). Reduced hepatic fatty acid oxidation in fasting PPARalpha null mice is due to impaired mitochondrial hydroxymethylglutaryl-CoA synthase gene expression. *FEBS Lett* 475, 163-166.
- Lebart, L., Morineau, A., and Piron, M. (2000). Statistique exploratoire multidimensionnelle, 3e Ed edn (Paris: Dunod).
- Leduc, A. M., Trent, J. O., Wittliff, J. L., Bramlett, K. S., Briggs, S. L., Chirgadze, N. Y., Wang, Y., Burris, T. P., and Spatola, A. F. (2003). Helix-stabilized cyclic peptides as selective inhibitors of steroid receptor-coactivator interactions. *Proc Natl Acad Sci U S A* 100, 11273-11278.
- Lee, C. H., and Wei, L. N. (1999). Characterization of an inverted repeat with a zero spacer (IR0)-type retinoic acid response element from the mouse nuclear orphan receptor TR2-11 gene. *Biochemistry* 38, 8820-8825.
- Lee, C. H., and Wei, L. N. (2000). Characterization of the mouse nuclear orphan receptor TR2-11 gene promoter and its potential role in retinoic acid-induced P19 apoptosis. *Biochem Pharmacol* 60, 127-136.
- Lee, G. Y., Kim, N. H., Zhao, Z. S., Cha, B. S., and Kim, Y. S. (2004a). Peroxisomal-proliferator-activated receptor alpha activates transcription of the rat hepatic malonyl-CoA decarboxylase gene: a key regulation of malonyl-CoA level. *Biochem J* 378, 983-990.
- Lee, H., Shi, W., Tontonoz, P., Wang, S., Subbanagounder, G., Hedrick, C. C., Hama, S., Borromeo, C., Evans, R. M., Berliner, J. A., and Nagy, L. (2000a). Role for peroxisome proliferator-activated receptor alpha in oxidized phospholipid-induced synthesis of monocyte chemotactic protein-1 and interleukin-8 by endothelial cells. *Circ Res* 87, 516-521.
- Lee, M. S., Kliewer, S. A., Provencal, J., Wright, P. E., and Evans, R. M. (1993). Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding. *Science* 260, 1117-1121.
- Lee, S. S., Chan, W. Y., Lo, C. K., Wan, D. C., Tsang, D. S., and Cheung, W. T. (2004b). Requirement of PPARalpha in maintaining phospholipid and triacylglycerol homeostasis during energy deprivation. *J Lipid Res* 45, 2025-2037.
- Lee, S. S., Pineau, T., Drago, J., Lee, E. J., Owens, J. W., Kroetz, D. L., Fernandez-Salguero, P. M., Westphal, H., and Gonzalez, F. J. (1995). Targeted disruption of the alpha isoform of the peroxisome proliferator-activated receptor gene in mice results in abolishment of the pleiotropic effects of peroxisome proliferators. *Mol Cell Biol* 15, 3012-3022.
- Lee, W. J., Kim, M., Park, H. S., Kim, H. S., Jeon, M. J., Oh, K. S., Koh, E. H., Won, J. C., Kim, M. S., Oh, G. T., *et al.* (2006). AMPK activation increases fatty acid oxidation in skeletal muscle by activating PPARalpha and PGC-1. *Biochem Biophys Res Commun* 340, 291-295.

- Lee, Y. K., Dell, H., Dowhan, D. H., Hadzopoulou-Cladaras, M., and Moore, D. D. (2000b). The orphan nuclear receptor SHP inhibits hepatocyte nuclear factor 4 and retinoid X receptor transactivation: two mechanisms for repression. *Mol Cell Biol* 20, 187-195.
- Lee, Y. K., and Moore, D. D. (2002). Dual mechanisms for repression of the monomeric orphan receptor liver receptor homologous protein-1 by the orphan small heterodimer partner. *J Biol Chem* 277, 2463-2467.
- Lees, J. A., Fawell, S. E., and Parker, M. G. (1989). Identification of two transactivation domains in the mouse oestrogen receptor. *Nucleic Acids Res* 17, 5477-5488.
- Lefebvre, A. M., Chen, I., Desreumaux, P., Najib, J., Fruchart, J. C., Geboes, K., Briggs, M., Heyman, R., and Auwerx, J. (1998). Activation of the peroxisome proliferator-activated receptor gamma promotes the development of colon tumors in C57BL/6J-APCMin/+ mice. *Nat Med* 4, 1053-1057.
- Lehmann, J. M., Lenhard, J. M., Oliver, B. B., Ringold, G. M., and Kliewer, S. A. (1997). Peroxisome proliferator-activated receptors alpha and gamma are activated by indomethacin and other non-steroidal anti-inflammatory drugs. *J Biol Chem* 272, 3406-3410.
- Lehmann, J. M., McKee, D. D., Watson, M. A., Willson, T. M., Moore, J. T., and Kliewer, S. A. (1998). The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J Clin Invest* 102, 1016-1023.
- Lehmann, J. M., Moore, L. B., Smith-Oliver, T. A., Wilkison, W. O., Willson, T. M., and Kliewer, S. A. (1995). An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). *J Biol Chem* 270, 12953-12956.
- Leibowitz, M. D., Fievet, C., Hennuyer, N., Peinado-Onsurbe, J., Duez, H., Bergera, J., Cullinan, C. A., Sparrow, C. P., Baffic, J., Berger, G. D., et al. (2000). Activation of PPARdelta alters lipid metabolism in db/db mice. *FEBS Lett* 473, 333-336.
- Lemberger, T., Braissant, O., Juge-Aubry, C., Keller, H., Saladin, R., Staels, B., Auwerx, J., Burger, A. G., Meier, C. A., and Wahli, W. (1996a). PPAR tissue distribution and interactions with other hormone-signaling pathways. *Ann N Y Acad Sci* 804, 231-251.
- Lemberger, T., Saladin, R., Vazquez, M., Assimacopoulos, F., Staels, B., Desvergne, B., Wahli, W., and Auwerx, J. (1996b). Expression of the peroxisome proliferator-activated receptor alpha gene is stimulated by stress and follows a diurnal rhythm. *J Biol Chem* 271, 1764-1769.
- Lemberger, T., Staels, B., Saladin, R., Desvergne, B., Auwerx, J., and Wahli, W. (1994). Regulation of the peroxisome proliferator-activated receptor alpha gene by glucocorticoids. *J Biol Chem* 269, 24527-24530.
- Leo, C., and Chen, J. D. (2000). The SRC family of nuclear receptor coactivators. *Gene* 245, 1-11.
- Leone, T. C., Weinheimer, C. J., and Kelly, D. P. (1999). A critical role for the peroxisome proliferator-activated receptor alpha (PPARalpha) in the cellular fasting response: the PPARalpha-null mouse as a model of fatty acid oxidation disorders. *Proc Natl Acad Sci U S A* 96, 7473-7478.
- Letowski, J., Brousseau, R., and Masson, L. (2004). Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *J Microbiol Methods* 57, 269-278.
- Lewis, G. F. (2006). Are the lipid-modifying effects of fibrates mediated by alterations in plasma lipid transfer activity? *Clin Sci (Lond)* 111, 185-187.
- Li, A. C., Binder, C. J., Gutierrez, A., Brown, K. K., Plotkin, C. R., Pattison, J. W., Valledor, A. F., Davis, R. A., Willson, T. M., Witztum, J. L., et al. (2004). Differential inhibition of macrophage foam-cell formation and atherosclerosis in mice by PPARalpha, beta/delta, and gamma. *J Clin Invest* 114, 1564-1576.
- Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell* 128, 707-719.
- Li, Y., Choi, M., Cavey, G., Daugherty, J., Suino, K., Kovach, A., Bingham, N. C., Kliewer, S. A., and Xu, H. E. (2005a). Crystallographic identification and functional characterization of phospholipids as ligands for the orphan nuclear receptor steroidogenic factor-1. *Mol Cell* 17, 491-502.
- Li, Y., Choi, M., Suino, K., Kovach, A., Daugherty, J., Kliewer, S. A., and Xu, H. E. (2005b). Structural and biochemical basis for selective repression of the orphan nuclear receptor liver receptor homolog 1 by small heterodimer partner. *Proc Natl Acad Sci U S A* 102, 9505-9510.
- Li, Y., Suino, K., Daugherty, J., and Xu, H. E. (2005c). Structural and biochemical mechanisms for the specificity of hormone binding and coactivator assembly by mineralocorticoid receptor. *Mol Cell* 19, 367-380.
- Liang, G., Yang, J., Horton, J. D., Hammer, R. E., Goldstein, J. L., and Brown, M. S. (2002). Diminished hepatic response to fasting/refeeding and liver X receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c. *J Biol Chem* 277, 9520-9528.
- Liang, P., and Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967-971.
- Liang, P., Zhu, W., Zhang, X., Guo, Z., O'Connell, R. P., Averboukh, L., Wang, F., and Pardee, A. B. (1994). Differential display using one-base anchored oligo-dT primers. *Nucleic Acids Res* 22, 5763-5764.

- Lim, C. S., Baumann, C. T., Htun, H., Xian, W., Irie, M., Smith, C. L., and Hager, G. L. (1999a). Differential localization and activity of the A- and B-forms of the human progesterone receptor using green fluorescent protein chimeras. *Mol Endocrinol* 13, 366-375.
- Lim, H., Gupta, R. A., Ma, W. G., Paria, B. C., Moller, D. E., Morrow, J. D., DuBois, R. N., Trzaskos, J. M., and Dey, S. K. (1999b). Cyclo-oxygenase-2-derived prostacyclin mediates embryo implantation in the mouse via PPARdelta. *Genes Dev* 13, 1561-1574.
- Lin, J., Tarr, P. T., Yang, R., Rhee, J., Puigserver, P., Newgard, C. B., and Spiegelman, B. M. (2003). PGC-1beta in the regulation of hepatic glucose and energy metabolism. *J Biol Chem* 278, 30843-30848.
- Lin, J., Yang, R., Tarr, P. T., Wu, P. H., Handschin, C., Li, S., Yang, W., Pei, L., Uldry, M., Tontonoz, P., et al. (2005). Hyperlipidemic effects of dietary saturated fats mediated through PGC-1beta coactivation of SREBP. *Cell* 120, 261-273.
- Lin, K. H., Parkison, C., McPhie, P., and Cheng, S. Y. (1991). An essential role of domain D in the hormone-binding activity of human beta 1 thyroid hormone nuclear receptor. *Mol Endocrinol* 5, 485-492.
- Lin, Q., Ruuska, S. E., Shaw, N. S., Dong, D., and Noy, N. (1999). Ligand selectivity of the peroxisome proliferator-activated receptor alpha. *Biochemistry* 38, 185-190.
- Lind, U., Carlstedt-Duke, J., Gustafsson, J. A., and Wright, A. P. (1996). Identification of single amino acid substitutions of Cys-736 that affect the steroid-binding affinity and specificity of the glucocorticoid receptor using phenotypic screening in yeast. *Mol Endocrinol* 10, 1358-1370.
- Lind, U., Greenidge, P., Gillner, M., Koehler, K. F., Wright, A., and Carlstedt-Duke, J. (2000). Functional probing of the human glucocorticoid receptor steroid-interacting surface by site-directed mutagenesis. Gln-642 plays an important role in steroid recognition and binding. *J Biol Chem* 275, 19041-19049.
- Lindquist, P. J., Svensson, L. T., and Alexson, S. E. (1998). Molecular cloning of the peroxisome proliferator-induced 46-kDa cytosolic acyl-CoA thioesterase from mouse and rat liver--recombinant expression in Escherichia coli, tissue expression, and nutritional regulation. *Eur J Biochem* 251, 631-640.
- Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nat Genet* 21, 20-24.
- Little, T. H., Zhang, Y., Matulis, C. K., Weck, J., Zhang, Z., Ramachandran, A., Mayo, K. E., and Radhakrishnan, I. (2006). Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. *Mol Endocrinol* 20, 831-843.
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402-408.
- Lo Verme, J., Fu, J., Astarita, G., La Rana, G., Russo, R., Calignano, A., and Piomelli, D. (2005). The nuclear receptor peroxisome proliferator-activated receptor-alpha mediates the anti-inflammatory actions of palmitoylethanolamide. *Mol Pharmacol* 67, 15-19.
- Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14, 1675-1680.
- Lonard, D. M., and O'Malley, B. W. (2006). The expanding cosmos of nuclear receptor coactivators. *Cell* 125, 411-414.
- Lönnstedt, I., and Speed, T. P. (2002). Replicated microarray data. *Statistica Sinica* 12, 31-46.
- Lopez, D., Irby, R. B., and McLean, M. P. (2003). Peroxisome proliferator-activated receptor alpha induces rat sterol carrier protein x promoter activity through two peroxisome proliferator-response elements. *Mol Cell Endocrinol* 205, 169-184.
- Lopez, D., and McLean, M. P. (2006). Activation of the rat scavenger receptor class B type I gene by PPARalpha. *Mol Cell Endocrinol* 251, 67-77.
- Loudig, O., Babichuk, C., White, J., Abu-Abed, S., Mueller, C., and Petkovich, M. (2000). Cytochrome P450RAI(CYP26) promoter: a distinct composite retinoic acid response element underlies the complex regulation of retinoic acid metabolism. *Mol Endocrinol* 14, 1483-1497.
- Louet, J. F., Chatelain, F., Decaux, J. F., Park, E. A., Kohl, C., Pineau, T., Girard, J., and Pegorier, J. P. (2001). Long-chain fatty acids regulate liver carnitine palmitoyltransferase I gene (L-CPT I) expression through a peroxisome-proliferator-activated receptor alpha (PPARalpha)-independent pathway. *Biochem J* 354, 189-197.
- Love, J. D., Gooch, J. T., Benko, S., Li, C., Nagy, L., Chatterjee, V. K., Evans, R. M., and Schwabe, J. W. (2002). The structural basis for the specificity of retinoid-X receptor-selective agonists: new insights into the role of helix H12. *J Biol Chem* 277, 11385-11391.
- Lu, I. L., Huang, C. F., Peng, Y. H., Lin, Y. T., Hsieh, H. P., Chen, C. T., Lien, T. W., Lee, H. J., Mahindroo, N., Prakash, E., et al. (2006a). Structure-based drug design of a novel family of PPARgamma partial agonists: virtual screening, X-ray crystallography, and in vitro/in vivo biological activities. *J Med Chem* 49, 2703-2712.

- Lu, N. Z., Wardell, S. E., Burnstein, K. L., Defranco, D., Fuller, P. J., Giguere, V., Hochberg, R. B., McKay, L., Renoir, J. M., Weigel, N. L., *et al.* (2006b). International Union of Pharmacology. LXV. The pharmacology and classification of the nuclear receptor superfamily: glucocorticoid, mineralocorticoid, progesterone, and androgen receptors. *Pharmacol Rev* 58, 782-797.
- Lu, T. T., Makishima, M., Repa, J. J., Schoonjans, K., Kerr, T. A., Auwerx, J., and Mangelsdorf, D. J. (2000). Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. *Mol Cell* 6, 507-515.
- Lubahn, D. B., Joseph, D. R., Sullivan, P. M., Willard, H. F., French, F. S., and Wilson, E. M. (1988). Cloning of human androgen receptor complementary DNA and localization to the X chromosome. *Science* 240, 327-330.
- Ludwig, M. G., Bassett, P., and Anglard, P. (2000). Multiple regulatory elements in the murine stromelysin-3 promoter. Evidence for direct control by CCAAT/enhancer-binding protein beta and thyroid and retinoid receptors. *J Biol Chem* 275, 39981-39990.
- Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., and Sigler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352, 497-505.
- Luquet, S., Lopez-Soriano, J., Holst, D., Fredenrich, A., Melki, J., Rassoulzadegan, M., and Grimaldi, P. A. (2003). Peroxisome proliferator-activated receptor delta controls muscle development and oxidative capability. *Faseb J* 17, 2299-2301.
- Madan, A. P., and DeFranco, D. B. (1993). Bidirectional transport of glucocorticoid receptors across the nuclear envelope. *Proc Natl Acad Sci U S A* 90, 3588-3592.
- Madauss, K. P., Deng, S. J., Austin, R. J., Lambert, M. H., McLay, I., Pritchard, J., Short, S. A., Stewart, E. L., Uings, I. J., and Williams, S. P. (2004). Progesterone receptor ligand binding pocket flexibility: crystal structures of the norethindrone and mometasone furoate complexes. *J Med Chem* 47, 3381-3387.
- Madauss, K. P., Grygielko, E. T., Deng, S. J., Sulpizio, A. C., Stanley, T. B., Wu, C., Short, S. A., Thompson, S. K., Stewart, E. L., Laping, N. J., *et al.* (2007). A structural and in vitro characterization of asoprisnil: a selective progesterone receptor modulator. *Mol Endocrinol* 21, 1066-1081.
- Madej, A., Okopien, B., Kowalski, J., Zielinski, M., Wysocki, J., Szygula, B., Kalina, Z., and Herman, Z. S. (1998). Effects of fenofibrate on plasma cytokine concentrations in patients with atherosclerosis and hyperlipoproteinemia IIb. *Int J Clin Pharmacol Ther* 36, 345-349.
- Maeda, I., Delessert, S., Hasegawa, S., Seto, Y., Zuber, S., and Poirier, Y. (2006). The peroxisomal Acyl-CoA thioesterase Pte1p from *Saccharomyces cerevisiae* is required for efficient degradation of short straight chain and branched chain fatty acids. *J Biol Chem* 281, 11729-11735.
- Maffini, M. V., Rubin, B. S., Sonnenschein, C., and Soto, A. M. (2006). Endocrine disruptors and reproductive health: the case of bisphenol-A. *Mol Cell Endocrinol* 254-255, 179-186.
- Maglich, J. M., Parks, D. J., Moore, L. B., Collins, J. L., Goodwin, B., Billin, A. N., Stoltz, C. A., Kliewer, S. A., Lambert, M. H., Willson, T. M., and Moore, J. T. (2003). Identification of a novel human constitutive androstane receptor (CAR) agonist and its use in the identification of CAR target genes. *J Biol Chem* 278, 17277-17283.
- Maglich, J. M., Sluder, A., Guan, X., Shi, Y., McKee, D. D., Carrick, K., Kamdar, K., Willson, T. M., and Moore, J. T. (2001). Comparison of complete nuclear receptor sets from the human, *Caenorhabditis elegans* and *Drosophila* genomes. *Genome Biol* 2, RESEARCH0029.
- Maglich, J. M., Watson, J., McMillen, P. J., Goodwin, B., Willson, T. M., and Moore, J. T. (2004). The nuclear receptor CAR is a regulator of thyroid hormone metabolism during caloric restriction. *J Biol Chem* 279, 19832-19838.
- Mahindroo, N., Wang, C. C., Liao, C. C., Huang, C. F., Lu, I. L., Lien, T. W., Peng, Y. H., Huang, W. J., Lin, Y. T., Hsu, M. C., *et al.* (2006). Indol-1-yl acetic acids as peroxisome proliferator-activated receptor agonists: design, synthesis, structural biology, and molecular docking studies. *J Med Chem* 49, 1212-1216.
- Maison, P., Mennen, L., Sapinho, D., Balkau, B., Sigalas, J., Chesnier, M. C., and Eschwege, E. (2002). A pharmacoepidemiological assessment of the effect of statins and fibrates on fibrinogen concentration. *Atherosclerosis* 160, 155-160.
- Makishima, M., Okamoto, A. Y., Repa, J. J., Tu, H., Learned, R. M., Luk, A., Hull, M. V., Lustig, K. D., Mangelsdorf, D. J., and Shan, B. (1999). Identification of a nuclear receptor for bile acids. *Science* 284, 1362-1365.
- Malamas, M. S., Manas, E. S., McDevitt, R. E., Gunawan, I., Xu, Z. B., Collini, M. D., Miller, C. P., Dinh, T., Henderson, R. A., Keith, J. C., Jr., and Harris, H. A. (2004). Design and synthesis of aryl diphenolic azoles as potent and selective estrogen receptor-beta ligands. *J Med Chem* 47, 5021-5040.
- Malik, S., and Roeder, R. G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci* 30, 256-263.
- Maloney, E. K., and Waxman, D. J. (1999). trans-Activation of PPARalpha and PPARgamma by structurally diverse environmental chemicals. *Toxicol Appl Pharmacol* 161, 209-218.

- Maloney, P. R., Parks, D. J., Haffner, C. D., Fivush, A. M., Chandra, G., Plunket, K. D., Creech, K. L., Moore, L. B., Wilson, J. G., Lewis, M. C., *et al.* (2000). Identification of a chemical tool for the orphan nuclear receptor FXR. *J Med Chem* 43, 2971-2974.
- Manas, E. S., Unwalla, R. J., Xu, Z. B., Malamas, M. S., Miller, C. P., Harris, H. A., Hsiao, C., Akopian, T., Hum, W. T., Malakian, K., *et al.* (2004a). Structure-based design of estrogen receptor-beta selective ligands. *J Am Chem Soc* 126, 15106-15119.
- Manas, E. S., Xu, Z. B., Unwalla, R. J., and Somers, W. S. (2004b). Understanding the selectivity of genistein for human estrogen receptor-beta using X-ray crystallography and computational methods. *Structure* 12, 2197-2207.
- Mandard, S., Muller, M., and Kersten, S. (2004). Peroxisome proliferator-activated receptor alpha target genes. *Cell Mol Life Sci* 61, 393-416.
- Mandard, S., Stienstra, R., Escher, P., Tan, N. S., Kim, I., Gonzalez, F. J., Wahli, W., Desvergne, B., Muller, M., and Kersten, S. (2007). Glycogen synthase 2 is a novel target gene of peroxisome proliferator-activated receptors. *Cell Mol Life Sci* 64, 1145-1157.
- Mangelsdorf, D. J., and Evans, R. M. (1995). The RXR heterodimers and orphan receptors. *Cell* 83, 841-850.
- Mangelsdorf, D. J., Umesono, K., Kliewer, S. A., Borgmeyer, U., Ong, E. S., and Evans, R. M. (1991). A direct repeat in the cellular retinol-binding protein type II gene confers differential regulation by RXR and RAR. *Cell* 66, 555-561.
- Mano, H., Kimura, C., Fujisawa, Y., Kameda, T., Watanabe-Mano, M., Kaneko, H., Kaneda, T., Hakeda, Y., and Kumegawa, M. (2000). Cloning and function of rabbit peroxisome proliferator-activated receptor delta/beta in mature osteoclasts. *J Biol Chem* 275, 8126-8132.
- Manoli, T., Gretz, N., Grone, H. J., Kenzelmann, M., Eils, R., and Brors, B. (2006). Group testing for pathway analysis improves comparability of different microarray datasets. *Bioinformatics* 22, 2500-2506.
- Mansen, A., Guardiola-Diaz, H., Rafter, J., Branting, C., and Gustafsson, J. A. (1996). Expression of the peroxisome proliferator-activated receptor (PPAR) in the mouse colonic mucosa. *Biochem Biophys Res Commun* 222, 844-851.
- Marcus, S. L., Miyata, K. S., Zhang, B., Subramani, S., Rachubinski, R. A., and Capone, J. P. (1993). Diverse peroxisome proliferator-activated receptors bind to the peroxisome proliferator-responsive elements of the rat hydratase/dehydrogenase and fatty acyl-CoA oxidase genes but differentially induce expression. *Proc Natl Acad Sci U S A* 90, 5723-5727.
- Marrapodi, M., and Chiang, J. Y. (2000). Peroxisome proliferator-activated receptor alpha (PPAR $\alpha$ ) and agonist inhibit cholesterol 7alpha-hydroxylase gene (CYP7A1) transcription. *J Lipid Res* 41, 514-520.
- Martin-Nizard, F., Furman, C., Delerive, P., Kandoussi, A., Fruchart, J. C., Staels, B., and Duriez, P. (2002). Peroxisome proliferator-activated receptor activators inhibit oxidized low-density lipoprotein-induced endothelin-1 secretion in endothelial cells. *J Cardiovasc Pharmacol* 40, 822-831.
- Martin, G., Poirier, H., Hennuyer, N., Crombie, D., Fruchart, J. C., Heyman, R. A., Besnard, P., and Auwerx, J. (2000). Induction of the fatty acid transport protein 1 and acyl-CoA synthase genes by dimer-selective retinoids suggests that the peroxisome proliferator-activated receptor-retinoid X receptor heterodimer is their molecular target. *J Biol Chem* 275, 12612-12618.
- Martin, G., Schoonjans, K., Lefebvre, A. M., Staels, B., and Auwerx, J. (1997). Coordinate regulation of the expression of the fatty acid transport protein and acyl-CoA synthetase genes by PPAR $\alpha$  and PPAR $\gamma$  activators. *J Biol Chem* 272, 28210-28217.
- Martin, G. G., Danneberg, H., Kumar, L. S., Atshaves, B. P., Erol, E., Bader, M., Schroeder, F., and Binas, B. (2003). Decreased liver fatty acid binding capacity and altered liver lipid distribution in mice lacking the liver fatty acid-binding protein gene. *J Biol Chem* 278, 21429-21438.
- Martin, P. G., Guillou, H., Lasserre, F., Dejean, S., Lan, A., Pascussi, J. M., Sancristobal, M., Legrand, P., Besse, P., and Pineau, T. (2007). Novel aspects of PPAR $\alpha$ -mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study. *Hepatology* 45, 767-777.
- Martin, P. G., Lasserre, F., Calleja, C., Van Es, A., Roulet, A., Concorde, D., Cantiello, M., Barnouin, R., Gauthier, B., and Pineau, T. (2005). Transcriptional modulations by RXR agonists are only partially subordinated to PPAR $\alpha$  signaling and attest additional, organ-specific, molecular cross-talks. *Gene Expr* 12, 177-192.
- Martinez, L. O., Jacquet, S., Esteve, J. P., Rolland, C., Cabezon, E., Champagne, E., Pineau, T., Georgeaud, V., Walker, J. E., Terce, F., *et al.* (2003). Ectopic beta-chain of ATP synthase is an apolipoprotein A-I receptor in hepatic HDL endocytosis. *Nature* 421, 75-79.
- Marx, N., Duez, H., Fruchart, J. C., and Staels, B. (2004). Peroxisome proliferator-activated receptors and atherosclerosis: regulators of gene expression in vascular cells. *Circ Res* 94, 1168-1178.

- Marx, N., Kehrle, B., Kohlhammer, K., Grub, M., Koenig, W., Hombach, V., Libby, P., and Plutzky, J. (2002). PPAR activators as antiinflammatory mediators in human T lymphocytes: implications for atherosclerosis and transplantation-associated arteriosclerosis. *Circ Res* 90, 703-710.
- Marx, N., Mach, F., Sauty, A., Leung, J. H., Sarafi, M. N., Ransohoff, R. M., Libby, P., Plutzky, J., and Luster, A. D. (2000). Peroxisome proliferator-activated receptor-gamma activators inhibit IFN-gamma-induced expression of the T cell-active CXC chemokines IP-10, Mig, and I-TAC in human endothelial cells. *J Immunol* 164, 6503-6508.
- Marx, N., Mackman, N., Schonbeck, U., Yilmaz, N., Hombach, V., Libby, P., and Plutzky, J. (2001). PPARalpha activators inhibit tissue factor expression and activity in human monocytes. *Circulation* 103, 213-219.
- Marx, N., Schonbeck, U., Lazar, M. A., Libby, P., and Plutzky, J. (1998). Peroxisome proliferator-activated receptor gamma activators inhibit gene expression and migration in human vascular smooth muscle cells. *Circ Res* 83, 1097-1103.
- Marx, N., Sukhova, G. K., Collins, T., Libby, P., and Plutzky, J. (1999). PPARalpha activators inhibit cytokine-induced vascular cell adhesion molecule-1 expression in human endothelial cells. *Circulation* 99, 3125-3131.
- Mascaro, C., Acosta, E., Ortiz, J. A., Marrero, P. F., Hegardt, F. G., and Haro, D. (1998). Control of human muscle-type carnitine palmitoyltransferase I gene transcription by peroxisome proliferator-activated receptor. *J Biol Chem* 273, 8560-8563.
- Matias, P. M., Donner, P., Coelho, R., Thomaz, M., Peixoto, C., Macedo, S., Otto, N., Joschko, S., Scholz, P., Wegg, A., et al. (2000). Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J Biol Chem* 275, 26164-26171.
- Matsumoto, K., Yu, S., Jia, Y., Ahmed, M. R., Viswakarma, N., Sarkar, J., Kashireddy, P. V., Rao, M. S., Karpus, W., Gonzalez, F. J., and Reddy, J. K. (2007). Critical role for transcription coactivator peroxisome proliferator-activated receptor (PPAR)-binding protein/TRAP220 in liver regeneration and PPARalpha ligand-induced liver tumor development. *J Biol Chem* 282, 17053-17060.
- Matsuzaka, T., Shimano, H., Yahagi, N., Amemiya-Kudo, M., Yoshikawa, T., Hasty, A. H., Tamura, Y., Osuga, J., Okazaki, H., Iizuka, Y., et al. (2002). Dual regulation of mouse Delta(5)- and Delta(6)-desaturase gene expression by SREBP-1 and PPARalpha. *J Lipid Res* 43, 107-114.
- Matthews, J., Celius, T., Halgren, R., and Zacharewski, T. (2000). Differential estrogen receptor binding of estrogenic substances: a species comparison. *J Steroid Biochem Mol Biol* 74, 223-234.
- Mauduit, C., Florin, A., Amara, S., Bozec, A., Siddeek, B., Cunha, S., Meunier, L., Selva, J., Albert, M., Vialard, F., et al. (2006). [Long-term effects of environmental endocrine disruptors on male fertility]. *Gynecol Obstet Fertil* 34, 978-984.
- Mazid, M. A., Chowdhury, A. A., Nagao, K., Nishimura, K., Jisaka, M., Nagaya, T., and Yokota, K. (2006). Endogenous 15-deoxy-Delta(12,14)-prostaglandin J(2) synthesized by adipocytes during maturation phase contributes to upregulation of fat storage. *FEBS Lett* 580, 6885-6890.
- Mazzucotelli, A., Viguerie, N., Tiraby, C., Annicotte, J. S., Mairal, A., Klimcakova, E., Lepin, E., Delmar, P., Dejean, S., Tavernier, G., et al. (2007). The Transcriptional Coactivator PGC-1{alpha} and the Nuclear Receptor PPAR{alpha} Control the Expression of Glycerol Kinase and Metabolism Genes Independently of PPAR{gamma} Activation in Human White Adipocytes. *Diabetes*.
- McDevitt, R. E., Malamas, M. S., Manas, E. S., Unwalla, R. J., Xu, Z. B., Miller, C. P., and Harris, H. A. (2005). Estrogen receptor ligands: design and synthesis of new 2-arylidene-1-ones. *Bioorg Med Chem Lett* 15, 3137-3142.
- McKenna, N. J., Lanz, R. B., and O'Malley, B. W. (1999). Nuclear receptor coregulators: cellular and molecular biology. *Endocr Rev* 20, 321-344.
- Mecham, B. H., Klus, G. T., Strovel, J., Augustus, M., Byrne, D., Bozso, P., Wetmore, D. Z., Mariani, T. J., Kohane, I. S., and Szallasi, Z. (2004). Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. *Nucleic Acids Res* 32, e74.
- Meertens, L. M., Miyata, K. S., Cechetto, J. D., Rachubinski, R. A., and Capone, J. P. (1998). A mitochondrial ketogenic enzyme regulates its gene expression by association with the nuclear hormone receptor PPARalpha. *Embo J* 17, 6972-6978.
- Meinke, G., and Sigler, P. B. (1999). DNA-binding mechanism of the monomeric orphan nuclear receptor NGFI-B. *Nat Struct Biol* 6, 471-477.
- Metivier, R., Petit, F. G., Valotaire, Y., and Pakdel, F. (2000). Function of N-terminal transactivation domain of the estrogen receptor requires a potential alpha-helical structure and is negatively regulated by the A domain. *Mol Endocrinol* 14, 1849-1871.
- Metivier, R., Reid, G., and Gannon, F. (2006). Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep* 7, 161-167.

- Metzger, D., Ali, S., Bornert, J. M., and Chambon, P. (1995). Characterization of the amino-terminal transcriptional activation function of the human estrogen receptor in animal and yeast cells. *J Biol Chem* 270, 9535-9542.
- Mewshaw, R. E., Edsall, R. J., Jr., Yang, C., Manas, E. S., Xu, Z. B., Henderson, R. A., Keith, J. C., Jr., and Harris, H. A. (2005). ERbeta ligands. 3. Exploiting two binding orientations of the 2-phenylnaphthalene scaffold to achieve ERbeta selectivity. *J Med Chem* 48, 3953-3979.
- Mi, L. Z., Devarakonda, S., Harp, J. M., Han, Q., Pellicciari, R., Willson, T. M., Khorasanizadeh, S., and Rastinejad, F. (2003). Structural basis for bile acid binding and activation of the nuclear receptor FXR. *Mol Cell* 11, 1093-1100.
- Michalik, L., Auwerx, J., Berger, J. P., Chatterjee, V. K., Glass, C. K., Gonzalez, F. J., Grimaldi, P. A., Kaduwaki, T., Lazar, M. A., O'Rahilly, S., et al. (2006). International Union of Pharmacology. LXI. Peroxisome proliferator-activated receptors. *Pharmacol Rev* 58, 726-741.
- Michalik, L., Desvergne, B., and Wahli, W. (2004). Peroxisome-proliferator-activated receptors and cancers: complex stories. *Nat Rev Cancer* 4, 61-70.
- Michalik, L., and Wahli, W. (2006). Involvement of PPAR nuclear receptors in tissue injury and wound repair. *J Clin Invest* 116, 598-606.
- Michalik, L., and Wahli, W. (2007). Peroxisome proliferator-activated receptors (PPARs) in skin health, repair and disease. *Biochim Biophys Acta*.
- Michalik, L., Zoete, V., Krey, G., Grosdidier, A., Gelman, L., Chodanowski, P., Feige, J. N., Desvergne, B., Wahli, W., and Michelin, O. (2007). Combined simulation and mutagenesis analyses reveal the involvement of key residues for peroxisome proliferator-activated receptor alpha helix 12 dynamic behavior. *J Biol Chem* 282, 9666-9677.
- Michigami, T., Suga, A., Yamazaki, M., Shimizu, C., Cai, G., Okada, S., and Ozono, K. (1999). Identification of amino acid sequence in the hinge region of human vitamin D receptor that transfers a cytosolic protein to the nucleus. *J Biol Chem* 274, 33531-33538.
- Miettinen, T. E., Tarpila, S., and Gylling, H. (1997). The effects of ursodeoxycholic acid on serum and biliary noncholesterol sterols in patients with gallstones. *Hepatology* 25, 514-518.
- Miller, C. W., and Ntambi, J. M. (1996). Peroxisome proliferators induce mouse liver stearoyl-CoA desaturase 1 gene expression. *Proc Natl Acad Sci U S A* 93, 9443-9448.
- Mishra, A., Chaudhary, A., and Sethi, S. (2004). Oxidized omega-3 fatty acids inhibit NF-kappaB activation via a PPARalpha-dependent pathway. *Arterioscler Thromb Vasc Biol* 24, 1621-1627.
- Misra, P., Qi, C., Yu, S., Shah, S. H., Cao, W. Q., Rao, M. S., Thimmapaya, B., Zhu, Y., and Reddy, J. K. (2002). Interaction of PIMT with transcriptional coactivators CBP, p300, and PBP differential role in transcriptional regulation. *J Biol Chem* 277, 20011-20019.
- Miyamoto, T., Kakizawa, T., Ichikawa, K., Nishio, S., Takeda, T., Suzuki, S., Kaneko, A., Kumagai, M., Mori, J., Yamashita, K., et al. (2001). The role of hinge domain in heterodimerization and specific DNA recognition by nuclear receptors. *Mol Cell Endocrinol* 181, 229-238.
- Miyata, K. S., Zhang, B., Marcus, S. L., Capone, J. P., and Rachubinski, R. A. (1993). Chicken ovalbumin upstream promoter transcription factor (COUP-TF) binds to a peroxisome proliferator-responsive element and antagonizes peroxisome proliferator-mediated signaling. *J Biol Chem* 268, 19169-19172.
- Mochizuki, K., Suruga, K., Kitagawa, M., Takase, S., and Goda, T. (2001). Modulation of the expression of peroxisome proliferator-activated receptor-dependent genes through disproportional expression of two subtypes in the small intestine. *Arch Biochem Biophys* 389, 41-48.
- Monsalve, M., Wu, Z., Adelman, G., Puigserver, P., Fan, M., and Spiegelman, B. M. (2000). Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol Cell* 6, 307-316.
- Moore, L. B., Goodwin, B., Jones, S. A., Wisely, G. B., Serabjit-Singh, C. J., Willson, T. M., Collins, J. L., and Kliewer, S. A. (2000a). St. John's wort induces hepatic drug metabolism through activation of the pregnane X receptor. *Proc Natl Acad Sci U S A* 97, 7500-7502.
- Moore, L. B., Parks, D. J., Jones, S. A., Bledsoe, R. K., Consler, T. G., Stimmel, J. B., Goodwin, B., Liddle, C., Blanchard, S. G., Willson, T. M., et al. (2000b). Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. *J Biol Chem* 275, 15122-15127.
- Mootha, V. K., Bunkenborg, J., Olsen, J. V., Hjerrild, M., Wisniewski, J. R., Stahl, E., Bolouri, M. S., Ray, H. N., Sihag, S., Kamal, M., et al. (2003a). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 115, 629-640.
- Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., et al. (2003b). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34, 267-273.
- Moraes, L. A., Piqueras, L., and Bishop-Bailey, D. (2006). Peroxisome proliferator-activated receptors and inflammation. *Pharmacol Ther* 110, 371-385.

- Mori, T., Tsukamoto, T., Mori, H., Tashiro, Y., and Fujiki, Y. (1991). Molecular cloning and deduced amino acid sequence of nonspecific lipid transfer protein (sterol carrier protein 2) of rat liver: a higher molecular mass (60 kDa) protein contains the primary sequence of nonspecific lipid transfer protein as its C-terminal part. *Proc Natl Acad Sci U S A* 88, 4338-4342.
- Motojima, K., Passilly, P., Peters, J. M., Gonzalez, F. J., and Latruffe, N. (1998). Expression of putative fatty acid transporter genes are regulated by peroxisome proliferator-activated receptor alpha and gamma activators in a tissue- and inducer-specific manner. *J Biol Chem* 273, 16710-16714.
- Moya-Camarena, S. Y., Vanden Heuvel, J. P., Blanchard, S. G., Leesnitzer, L. A., and Belury, M. A. (1999). Conjugated linoleic acid is a potent naturally occurring ligand and activator of PPARalpha. *J Lipid Res* 40, 1426-1433.
- Mueller, S. O., Simon, S., Chae, K., Metzler, M., and Korach, K. S. (2004). Phytoestrogens and their human metabolites show distinct agonistic and antagonistic properties on estrogen receptor alpha (ERalpha) and ERbeta in human cells. *Toxicol Sci* 80, 14-25.
- Muerhoff, A. S., Griffin, K. J., and Johnson, E. F. (1992). The peroxisome proliferator-activated receptor mediates the induction of CYP4A6, a cytochrome P450 fatty acid omega-hydroxylase, by clofibrate. *J Biol Chem* 267, 19051-19053.
- Muga, S. J., Thuillier, P., Pavone, A., Rundhaug, J. E., Boeglin, W. E., Jisaka, M., Brash, A. R., and Fischer, S. M. (2000). 8S-lipoxygenase products activate peroxisome proliferator-activated receptor alpha and induce differentiation in murine keratinocytes. *Cell Growth Differ* 11, 447-454.
- Mukherjee, A., Kirkovsky, L., Yao, X. T., Yates, R. C., Miller, D. D., and Dalton, J. T. (1996). Enantioselective binding of Casodex to the androgen receptor. *Xenobiotica* 26, 117-122.
- Mukherjee, R., Davies, P. J., Crombie, D. L., Bischoff, E. D., Cesario, R. M., Jow, L., Hamann, L. G., Boehm, M. F., Mondon, C. E., Nadzan, A. M., et al. (1997a). Sensitization of diabetic and obese mice to insulin by retinoid X receptor agonists. *Nature* 386, 407-410.
- Mukherjee, R., Jow, L., Croston, G. E., and Paterniti, J. R., Jr. (1997b). Identification, characterization, and tissue distribution of human peroxisome proliferator-activated receptor (PPAR) isoforms PPARgamma2 versus PPARgamma1 and activation with retinoid X receptor agonists and antagonists. *J Biol Chem* 272, 8071-8076.
- Mukherjee, R., Strasser, J., Jow, L., Hoener, P., Paterniti, J. R., Jr., and Heyman, R. A. (1998). RXR agonists activate PPARalpha-inducible genes, lower triglycerides, and raise HDL levels in vivo. *Arterioscler Thromb Vasc Biol* 18, 272-276.
- Mukherjee, R., Sun, S., Santomenna, L., Miao, B., Walton, H., Liao, B., Locke, K., Zhang, J. H., Nguyen, S. H., Zhang, L. T., et al. (2002). Ligand and coactivator recruitment preferences of peroxisome proliferator activated receptor alpha. *J Steroid Biochem Mol Biol* 81, 217-225.
- Muoio, D. M., MacLean, P. S., Lang, D. B., Li, S., Houmard, J. A., Way, J. M., Winegar, D. A., Corton, J. C., Dohm, G. L., and Kraus, W. E. (2002). Fatty acid homeostasis and induction of lipid regulatory genes in skeletal muscles of peroxisome proliferator-activated receptor (PPAR) alpha knock-out mice. Evidence for compensatory regulation by PPAR delta. *J Biol Chem* 277, 26089-26097.
- Murakami, K., Ide, T., Suzuki, M., Mochizuki, T., and Kadowaki, T. (1999). Evidence for direct binding of fatty acids and eicosanoids to human peroxisome proliferators-activated receptor alpha. *Biochem Biophys Res Commun* 260, 609-613.
- Nagpal, S., Athanikar, J., and Chandraratna, R. A. (1995). Separation of transactivation and AP1 antagonism functions of retinoic acid receptor alpha. *J Biol Chem* 270, 923-927.
- Nagy, L., Tontonoz, P., Alvarez, J. G., Chen, H., and Evans, R. M. (1998). Oxidized LDL regulates macrophage gene expression through ligand activation of PPARgamma. *Cell* 93, 229-240.
- Nakagawa, S., Kawashima, Y., Hirose, A., and Kozuka, H. (1994). Regulation of hepatic level of fatty-acid-binding protein by hormones and clofibrate acid in the rat. *Biochem J* 297 ( Pt 3), 581-584.
- Nakai, A., Seino, S., Sakurai, A., Szilak, I., Bell, G. I., and DeGroot, L. J. (1988). Characterization of a thyroid hormone receptor expressed in human kidney and other tissues. *Proc Natl Acad Sci U S A* 85, 2781-2785.
- Nakajima, S., Hsieh, J. C., MacDonald, P. N., Haussler, C. A., Galligan, M. A., Jurutka, P. W., and Haussler, M. R. (1993). Purified human vitamin D receptor overexpressed in *E. coli* and baculovirus systems does not bind 1,25-dihydroxyvitamin D<sub>3</sub> hormone efficiently unless supplemented with a rat liver nuclear extract. *Biochem Biophys Res Commun* 197, 478-485.
- Nakatani, T., Tsuboyama-Kasaoka, N., Takahashi, M., Miura, S., and Ezaki, O. (2002). Mechanism for peroxisome proliferator-activated receptor-alpha activator-induced up-regulation of UCP2 mRNA in rodent hepatocytes. *J Biol Chem* 277, 9562-9569.
- Nakazato, M., Murakami, N., Date, Y., Kojima, M., Matsuo, H., Kangawa, K., and Matsukura, S. (2001). A role for ghrelin in the central regulation of feeding. *Nature* 409, 194-198.
- Nakielny, S., and Dreyfuss, G. (1999). Transport of proteins and RNAs in and out of the nucleus. *Cell* 99, 677-690.

- Nakshatri, H., and Bhat-Nakshatri, P. (1998). Multiple parameters determine the specificity of transcriptional response by nuclear receptors HNF-4, ARP-1, PPAR, RAR and RXR through common response elements. *Nucleic Acids Res* 26, 2491-2499.
- Napal, L., Marrero, P. F., and Haro, D. (2005). An intronic peroxisome proliferator-activated receptor-binding sequence mediates fatty acid induction of the human carnitine palmitoyltransferase 1A. *J Mol Biol* 354, 751-759.
- Nascimento, A. S., Dias, S. M., Nunes, F. M., Aparicio, R., Ambrosio, A. L., Bleicher, L., Figueira, A. C., Santos, M. A., de Oliveira Neto, M., Fischer, H., et al. (2006). Structural rearrangements in the thyroid hormone receptor hinge domain and their putative role in the receptor function. *J Mol Biol* 360, 586-598.
- Neess, D., Kiilerich, P., Sandberg, M. B., Helledie, T., Nielsen, R., and Mandrup, S. (2006). ACBP--a PPAR and SREBP modulated housekeeping gene. *Mol Cell Biochem* 284, 149-157.
- Nequin, L. G., Alvarez, J., and Schwartz, N. B. (1975). Steroid control of gonadotropin release. *J Steroid Biochem* 6, 1007-1012.
- Neve, B. P., Corseaux, D., Chinetti, G., Zawadzki, C., Fruchart, J. C., Duriez, P., Staels, B., and Jude, B. (2001). PPAR $\alpha$  agonists inhibit tissue factor expression in human monocytes and macrophages. *Circulation* 103, 207-212.
- Newberry, E. P., Xie, Y., Kennedy, S., Han, X., Buhman, K. K., Luo, J., Gross, R. W., and Davidson, N. O. (2003). Decreased hepatic triglyceride accumulation and altered fatty acid uptake in mice with deletion of the liver fatty acid-binding protein gene. *J Biol Chem* 278, 51664-51672.
- Nguyen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P., and Jordan, B. R. (1995). Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics* 29, 207-216.
- Nicolas-Frances, V., Dasari, V. K., Abruzzi, E., Osumi, T., and Latruffe, N. (2000). The peroxisome proliferator response element (PPRE) present at positions -681/-669 in the rat liver 3-ketoacyl-CoA thiolase B gene functionally interacts differently with PPAR $\alpha$  and HNF-4. *Biochem Biophys Res Commun* 269, 347-351.
- Nolte, R. T., Wisely, G. B., Westin, S., Cobb, J. E., Lambert, M. H., Kurokawa, R., Rosenfeld, M. G., Willson, T. M., Glass, C. K., and Milburn, M. V. (1998). Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* 395, 137-143.
- Oberfield, J. L., Collins, J. L., Holmes, C. P., Goreham, D. M., Cooper, J. P., Cobb, J. E., Lenhard, J. M., Hull-Ryde, E. A., Mohr, C. P., Blanchard, S. G., et al. (1999). A peroxisome proliferator-activated receptor gamma ligand inhibits adipocyte differentiation. *Proc Natl Acad Sci U S A* 96, 6102-6106.
- Ogden, C. L., Carroll, M. D., Curtin, L. R., McDowell, M. A., Tabak, C. J., and Flegal, K. M. (2006). Prevalence of overweight and obesity in the United States, 1999-2004. *Jama* 295, 1549-1555.
- Oh, D. K., Ciaraldi, T., and Henry, R. R. (2007). Adiponectin in health and disease. *Diabetes Obes Metab* 9, 282-289.
- Oishi, K., Shirai, H., and Ishida, N. (2005). CLOCK is involved in the circadian transactivation of peroxisome-proliferator-activated receptor alpha (PPAR $\alpha$ ) in mice. *Biochem J* 386, 575-581.
- Oliver, W. R., Jr., Shenk, J. L., Snaith, M. R., Russell, C. S., Plunket, K. D., Bodkin, N. L., Lewis, M. C., Winegar, D. A., Sznajzman, M. L., Lambert, M. H., et al. (2001). A selective peroxisome proliferator-activated receptor delta agonist promotes reverse cholesterol transport. *Proc Natl Acad Sci U S A* 98, 5306-5311.
- Omura, M., Kobayashi, S., Mizukami, Y., Mogami, K., Todoroki-Ikeda, N., Miyake, T., and Matsuzaki, M. (2001). Eicosapentaenoic acid (EPA) induces Ca(2+)-independent activation and translocation of endothelial nitric oxide synthase and endothelium-dependent vasorelaxation. *FEBS Lett* 487, 361-366.
- Ortiz, J. A., Mallolas, J., Nicot, C., Bofarull, J., Rodriguez, J. C., Hegardt, F. G., Haro, D., and Marrero, P. F. (1999). Isolation of pig mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase gene promoter: characterization of a peroxisome proliferator-responsive element. *Biochem J* 337 (Pt 2), 329-335.
- Ortlund, E. A., Lee, Y., Solomon, I. H., Hager, J. M., Safi, R., Choi, Y., Guan, Z., Tripathy, A., Raetz, C. R., McDonnell, D. P., et al. (2005). Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nat Struct Mol Biol* 12, 357-363.
- Ostberg, T., Svensson, S., Selen, G., Uppenberg, J., Thor, M., Sundbom, M., Sydow-Backman, M., Gustavsson, A. L., and Jendeberg, L. (2004). A new class of peroxisome proliferator-activated receptor agonists with a novel binding epitope shows antidiabetic effects. *J Biol Chem* 279, 41124-41130.
- Osumi, T. (1993). Structure and expression of the genes encoding peroxisomal beta-oxidation enzymes. *Biochimie* 75, 243-250.
- Osumi, T., Wen, J. K., and Hashimoto, T. (1991). Two cis-acting regulatory sequences in the peroxisome proliferator-responsive enhancer region of rat acyl-CoA oxidase gene. *Biochem Biophys Res Commun* 175, 866-871.

- Ouamrane, L., Larrieu, G., Gauthier, B., and Pineau, T. (2003). RXR activators molecular signalling: involvement of a PPAR alpha-dependent pathway in the liver and kidney, evidence for an alternative pathway in the heart. *Br J Pharmacol* 138, 845-854.
- Palmer, C. N., Hsu, M. H., Griffin, H. J., and Johnson, E. F. (1995). Novel sequence determinants in peroxisome proliferator signaling. *J Biol Chem* 270, 16114-16121.
- Palmer, C. N., Hsu, M. H., Griffin, K. J., Raucy, J. L., and Johnson, E. F. (1998). Peroxisome proliferator activated receptor-alpha expression in human liver. *Mol Pharmacol* 53, 14-22.
- Palmer, C. N., Hsu, M. H., Muerhoff, A. S., Griffin, K. J., and Johnson, E. F. (1994). Interaction of the peroxisome proliferator-activated receptor alpha with the retinoid X receptor alpha unmasks a cryptic peroxisome proliferator response element that overlaps an ARP-1-binding site in the CYP4A6 promoter. *J Biol Chem* 269, 18083-18089.
- Pande, S. V., and Parvin, R. (1980). Clofibrate enhancement of mitochondrial carnitine transport system of rat liver and augmentation of liver carnitine and gamma-butyrobetaine hydroxylase activity by thyroxine. *Biochim Biophys Acta* 617, 363-370.
- Pare, J. F., Malenfant, D., Courtemanche, C., Jacob-Wagner, M., Roy, S., Allard, D., and Belanger, L. (2004). The fetoprotein transcription factor (FTF) gene is essential to embryogenesis and cholesterol homeostasis and is regulated by a DR4 element. *J Biol Chem* 279, 21206-21216.
- Paredes, C. J., Senger, R. S., Spath, I. S., Borden, J. R., Sillers, R., and Papoutsakis, E. T. (2007). A general framework for designing and validating oligomer-based DNA microarrays and its application to *Clostridium acetobutylicum*. *Appl Environ Microbiol* 73, 4631-4638.
- Parks, D. J., Blanchard, S. G., Bledsoe, R. K., Chandra, G., Consler, T. G., Kliewer, S. A., Stimmel, J. B., Willson, T. M., Zavacki, A. M., Moore, D. D., and Lehmann, J. M. (1999). Bile acids: natural ligands for an orphan nuclear receptor. *Science* 284, 1365-1368.
- Parmigiani, G. (2003). The analysis of gene expression data : methods and software (New York: Springer).
- Patel, D. D., Knight, B. L., Soutar, A. K., Gibbons, G. F., and Wade, D. P. (2000). The effect of peroxisome-proliferator-activated receptor-alpha on the activity of the cholesterol 7 alpha-hydroxylase gene. *Biochem J* 351 Pt 3, 747-753.
- Patel, D. D., Knight, B. L., Wiggins, D., Humphreys, S. M., and Gibbons, G. F. (2001). Disturbances in the normal regulation of SREBP-sensitive genes in PPAR alpha-deficient mice. *J Lipid Res* 42, 328-337.
- Patsouris, D., Mandard, S., Voshol, P. J., Escher, P., Tan, N. S., Havekes, L. M., Koenig, W., Marz, W., Tafuri, S., Wahli, W., et al. (2004). PPARalpha governs glycerol metabolism. *J Clin Invest* 114, 94-103.
- Patsouris, D., Reddy, J. K., Muller, M., and Kersten, S. (2006). Peroxisome proliferator-activated receptor alpha mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology* 147, 1508-1516.
- Patterson, T. A., Lobenhofer, E. K., Fulmer-Smentek, S. B., Collins, P. J., Chu, T. M., Bao, W., Fang, H., Kawasaki, E. S., Hager, J., Tikhonova, I. R., et al. (2006). Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol* 24, 1140-1150.
- Pawar, A., and Jump, D. B. (2003). Unsaturated fatty acid regulation of peroxisome proliferator-activated receptor alpha activity in rat primary hepatocytes. *J Biol Chem* 278, 35931-35939.
- Peet, D. J., Turley, S. D., Ma, W., Janowski, B. A., Lobaccaro, J. M., Hammer, R. E., and Mangelsdorf, D. J. (1998). Cholesterol and bile acid metabolism are impaired in mice lacking the nuclear oxysterol receptor LXR alpha. *Cell* 93, 693-704.
- Peirson, S. N., Butler, J. N., and Foster, R. G. (2003). Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res* 31, e73.
- Pemberton, L. F., and Paschal, B. M. (2005). Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* 6, 187-198.
- Pereira de Jesus-Tran, K., Cote, P. L., Cantin, L., Blanchet, J., Labrie, F., and Breton, R. (2006). Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein Sci* 15, 987-999.
- Peters, J. M., Aoyama, T., Burns, A. M., and Gonzalez, F. J. (2003). Bezafibrate is a dual ligand for PPARalpha and PPARbeta: studies using null mice. *Biochim Biophys Acta* 1632, 80-89.
- Peters, J. M., Cheung, C., and Gonzalez, F. J. (2005). Peroxisome proliferator-activated receptor-alpha and liver cancer: where do we stand? *J Mol Med* 83, 774-785.
- Peters, J. M., Hennuyer, N., Staels, B., Fruchart, J. C., Fievet, C., Gonzalez, F. J., and Auwerx, J. (1997). Alterations in lipoprotein metabolism in peroxisome proliferator-activated receptor alpha-deficient mice. *J Biol Chem* 272, 27307-27312.
- Peters, J. M., Park, Y., Gonzalez, F. J., and Pariza, M. W. (2001). Influence of conjugated linoleic acid on body composition and target gene expression in peroxisome proliferator-activated receptor alpha-null mice. *Biochim Biophys Acta* 1533, 233-242.

- Peters, J. M., Zhou, Y. C., Ram, P. A., Lee, S. S., Gonzalez, F. J., and Waxman, D. J. (1996). Peroxisome proliferator-activated receptor alpha required for gene induction by dehydroepiandrosterone-3 beta-sulfate. *Mol Pharmacol* 50, 67-74.
- Petkovich, M., Brand, N. J., Krust, A., and Chambon, P. (1987). A human retinoic acid receptor which belongs to the family of nuclear receptors. *Nature* 330, 444-450.
- Petkovich, P. M. (2001). Retinoic acid metabolism. *J Am Acad Dermatol* 45, S136-142.
- Petrescu, A. D., Hertz, R., Bar-Tana, J., Schroeder, F., and Kier, A. B. (2002). Ligand specificity and conformational dependence of the hepatic nuclear factor-4alpha (HNF-4alpha). *J Biol Chem* 277, 23988-23999.
- Pignatello, M. A., Kauffman, F. C., and Levin, A. A. (1997). Multiple factors contribute to the toxicity of the aromatic retinoid, TTNPB (Ro 13-7410): binding affinities and disposition. *Toxicol Appl Pharmacol* 142, 319-327.
- Pike, A. C., Brzozowski, A. M., Hubbard, R. E., Bonn, T., Thorsell, A. G., Engstrom, O., Ljunggren, J., Gustafsson, J. A., and Carlquist, M. (1999). Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. *Embo J* 18, 4608-4618.
- Pike, A. C., Brzozowski, A. M., Walton, J., Hubbard, R. E., Thorsell, A. G., Li, Y. L., Gustafsson, J. A., and Carlquist, M. (2001). Structural insights into the mode of action of a pure antiestrogen. *Structure* 9, 145-153.
- Plevin, M. J., Mills, M. M., and Ikura, M. (2005). The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem Sci* 30, 66-69.
- Plummer, S., Sharpe, R. M., Hallmark, N., Mahood, I. K., and Elcombe, C. (2007). Time-dependent and compartment-specific effects of in utero exposure to Di(n-butyl) phthalate on gene/protein expression in the fetal rat testis as revealed by transcription profiling and laser capture microdissection. *Toxicol Sci* 97, 520-532.
- Pogenberg, V., Guichou, J. F., Vivat-Hannah, V., Kammerer, S., Perez, E., Germain, P., de Lera, A. R., Gronemeyer, H., Royer, C. A., and Bourguet, W. (2005). Characterization of the interaction between retinoic acid receptor/retinoid X receptor (RAR/RXR) heterodimers and transcriptional coactivators through structural and fluorescence anisotropy studies. *J Biol Chem* 280, 1625-1633.
- Poirier, H., Niot, I., Monnot, M. C., Braissant, O., Meunier-Durmort, C., Costet, P., Pineau, T., Wahli, W., Willson, T. M., and Besnard, P. (2001). Differential involvement of peroxisome-proliferator-activated receptors alpha and delta in fibrate and fatty-acid-mediated inductions of the gene encoding liver fatty-acid-binding protein in the liver and the small intestine. *Biochem J* 355, 481-488.
- Post, S. M., Duez, H., Gervois, P. P., Staels, B., Kuipers, F., and Princen, H. M. (2001). Fibrates suppress bile acid synthesis via peroxisome proliferator-activated receptor-alpha-mediated downregulation of cholesterol 7alpha-hydroxylase and sterol 27-hydroxylase expression. *Arterioscler Thromb Vasc Biol* 21, 1840-1845.
- Pozhitkov, A., Noble, P. A., Domazet-Loso, T., Nolte, A. W., Sonnenberg, R., Staehler, P., Beier, M., and Tautz, D. (2006). Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res* 34, e66.
- Pratt, W. B., and Toft, D. O. (1997). Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr Rev* 18, 306-360.
- Preitner, N., Damiola, F., Lopez-Molina, L., Zakany, J., Duboule, D., Albrecht, U., and Schibler, U. (2002). The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* 110, 251-260.
- Prieur, X., Coste, H., and Rodriguez, J. C. (2003). The human apolipoprotein AV gene is regulated by peroxisome proliferator-activated receptor-alpha and contains a novel farnesoid X-activated receptor response element. *J Biol Chem* 278, 25468-25480.
- Privalsky, M. L. (2004). The role of corepressors in transcriptional regulation by nuclear hormone receptors. *Annu Rev Physiol* 66, 315-360.
- Prufer, K., and Barsony, J. (2002). Retinoid X receptor dominates the nuclear import and export of the unliganded vitamin D receptor. *Mol Endocrinol* 16, 1738-1751.
- Puigserver, P., Wu, Z., Park, C. W., Graves, R., Wright, M., and Spiegelman, B. M. (1998). A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* 92, 829-839.
- Qi, C., Surapureddi, S., Zhu, Y. J., Yu, S., Kashireddy, P., Rao, M. S., and Reddy, J. K. (2003). Transcriptional coactivator PRIP, the peroxisome proliferator-activated receptor gamma (PPARgamma)-interacting protein, is required for PPARgamma-mediated adipogenesis. *J Biol Chem* 278, 25281-25284.
- Qi, C., Zhu, Y., Pan, J., Usuda, N., Maeda, N., Yeldandi, A. V., Rao, M. S., Hashimoto, T., and Reddy, J. K. (1999a). Absence of spontaneous peroxisome proliferation in enoyl-CoA Hydratase/L-3-hydroxyacyl-CoA dehydrogenase-deficient mouse liver. Further support for the role of fatty acyl CoA oxidase in PPARalpha ligand metabolism. *J Biol Chem* 274, 15775-15780.
- Qi, C., Zhu, Y., Pan, J., Yeldandi, A. V., Rao, M. S., Maeda, N., Subbarao, V., Pulikuri, S., Hashimoto, T., and Reddy, J. K. (1999b). Mouse steroid receptor coactivator-1 is not essential for peroxisome proliferator-activated receptor alpha-regulated gene expression. *Proc Natl Acad Sci U S A* 96, 1585-1590.
- Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat Genet* 32 Suppl, 496-501.

- Rachez, C., Suldan, Z., Ward, J., Chang, C. P., Burakov, D., Erdjument-Bromage, H., Tempst, P., and Freedman, L. P. (1998). A novel protein complex that interacts with the vitamin D<sub>3</sub> receptor in a ligand-dependent manner and enhances VDR transactivation in a cell-free system. *Genes Dev* 12, 1787-1800.
- Raisher, B. D., Gulick, T., Zhang, Z., Strauss, A. W., Moore, D. D., and Kelly, D. P. (1992). Identification of a novel retinoid-responsive element in the promoter region of the medium chain acyl-coenzyme A dehydrogenase gene. *J Biol Chem* 267, 20264-20269.
- Rajala, M. W., Qi, Y., Patel, H. R., Takahashi, N., Banerjee, R., Pajvani, U. B., Sinha, M. K., Gingerich, R. L., Scherer, P. E., and Ahima, R. S. (2004). Regulation of resistin expression and circulating levels in obesity, diabetes, and fasting. *Diabetes* 53, 1671-1679.
- Ramakers, C., Ruijter, J. M., Deprez, R. H., and Moorman, A. F. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 339, 62-66.
- Ramdas, L., Coombes, K. R., Baggerly, K., Abruzzo, L., Highsmith, W. E., Krognann, T., Hamilton, S. R., and Zhang, W. (2001). Sources of nonlinearity in cDNA microarray expression measurements. *Genome Biol* 2, RESEARCH0047.
- Rashid, S., and Lewis, G. F. (2005). The mechanisms of differential glucocorticoid and mineralocorticoid action in the brain and peripheral tissues. *Clin Biochem* 38, 401-409.
- Raspe, E., Duez, H., Gervois, P., Fievet, C., Fruchart, J. C., Besnard, S., Mariani, J., Tedgui, A., and Staels, B. (2001). Transcriptional regulation of apolipoprotein C-III gene expression by the orphan nuclear receptor RORalpha. *J Biol Chem* 276, 2865-2871.
- Raspe, E., Duez, H., Mansen, A., Fontaine, C., Fievet, C., Fruchart, J. C., Vennstrom, B., and Staels, B. (2002). Identification of Rev-erbalpha as a physiological repressor of apoC-III gene transcription. *J Lipid Res* 43, 2172-2179.
- Rastinejad, F., Perlmann, T., Evans, R. M., and Sigler, P. B. (1995). Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* 375, 203-211.
- Rastinejad, F., Wagner, T., Zhao, Q., and Khorasanizadeh, S. (2000). Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1. *Embo J* 19, 1045-1054.
- Rayasam, G. V., Wendling, O., Angrand, P. O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G. L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *Embo J* 22, 3153-3163.
- Reddy, J. K. (2004). Peroxisome proliferators and peroxisome proliferator-activated receptor alpha: biotic and xenobiotic sensing. *Am J Pathol* 164, 2305-2321.
- Reddy, J. K., Reddy, M. K., Usman, M. I., Lalwani, N. D., and Rao, M. S. (1986). Comparison of hepatic peroxisome proliferative effect and its implication for hepatocarcinogenicity of phthalate esters, di(2-ethylhexyl) phthalate, and di(2-ethylhexyl) adipate with a hypolipidemic drug. *Environ Health Perspect* 65, 317-327.
- Reid, J., Kelly, S. M., Watt, K., Price, N. C., and McEwan, I. J. (2002). Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. *J Biol Chem* 277, 20079-20086.
- Reinartz, J., Bruyns, E., Lin, J. Z., Burcham, T., Brenner, S., Bowen, B., Kramer, M., and Woychik, R. (2002). Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Brief Funct Genomic Proteomic* 1, 95-104.
- Ren, B., Thelen, A., and Jump, D. B. (1996). Peroxisome proliferator-activated receptor alpha inhibits hepatic S14 gene transcription. Evidence against the peroxisome proliferator-activated receptor alpha as the mediator of polyunsaturated fatty acid regulation of s14 gene transcription. *J Biol Chem* 271, 17167-17173.
- Ren, B., Thelen, A. P., Peters, J. M., Gonzalez, F. J., and Jump, D. B. (1997). Polyunsaturated fatty acid suppression of hepatic fatty acid synthase and S14 gene expression does not require peroxisome proliferator-activated receptor alpha. *J Biol Chem* 272, 26827-26832.
- Renaud, J., Bischoff, S. F., Buhl, T., Floersheim, P., Fournier, B., Geiser, M., Halleux, C., Kallen, J., Keller, H., and Ramage, P. (2005). Selective estrogen receptor modulators with conformationally restricted side chains. Synthesis and structure-activity relationship of ERalpha-selective tetrahydroisoquinoline ligands. *J Med Chem* 48, 364-379.
- Renaud, J., Bischoff, S. F., Buhl, T., Floersheim, P., Fournier, B., Halleux, C., Kallen, J., Keller, H., Schlaeppi, J. M., and Stark, W. (2003). Estrogen receptor modulators: identification and structure-activity relationships of potent ERalpha-selective tetrahydroisoquinoline ligands. *J Med Chem* 46, 2945-2957.
- Renaud, J. P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., and Moras, D. (1995). Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature* 378, 681-689.
- Repa, J. J., Liang, G., Ou, J., Bashmakov, Y., Lobaccaro, J. M., Shimomura, I., Shan, B., Brown, M. S., Goldstein, J. L., and Mangelsdorf, D. J. (2000). Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRAalpha and LXRBeta. *Genes Dev* 14, 2819-2830.
- Ricote, M., Huang, J., Fajas, L., Li, A., Welch, J., Najib, J., Witztum, J. L., Auwerx, J., Palinski, W., and Glass, C. K. (1998). Expression of the peroxisome proliferator-activated receptor gamma (PPARGamma) in

human atherosclerosis and regulation in macrophages by colony stimulating factors and oxidized low density lipoprotein. *Proc Natl Acad Sci U S A* 95, 7614-7619.

Rigamonti, E., Helin, L., Lestavel, S., Mutka, A. L., Lepore, M., Fontaine, C., Bouhlel, M. A., Bultel, S., Fruchart, J. C., Ikonen, E., et al. (2005). Liver X receptor activation controls intracellular cholesterol trafficking and esterification in human macrophages. *Circ Res* 97, 682-689.

Ripp, S. L., Falkner, K. C., Pendleton, M. L., Tamasi, V., and Prough, R. A. (2003). Regulation of CYP2C11 by dehydroepiandrosterone and peroxisome proliferators: identification of the negative regulatory region of the gene. *Mol Pharmacol* 64, 113-122.

Ripp, S. L., Fitzpatrick, J. L., Peters, J. M., and Prough, R. A. (2002). Induction of CYP3A expression by dehydroepiandrosterone: involvement of the pregnane X receptor. *Drug Metab Dispos* 30, 570-575.

Ris-Stalpers, C., Kuiper, G. G., Faber, P. W., Schweikert, H. U., van Rooij, H. C., Zegers, N. D., Hodgins, M. B., Degenhart, H. J., Trapman, J., and Brinkmann, A. O. (1990). Aberrant splicing of androgen receptor mRNA results in synthesis of a nonfunctional receptor protein in a patient with androgen insensitivity. *Proc Natl Acad Sci U S A* 87, 7866-7870.

Rizvi, K., Hampson, J. P., and Harvey, J. N. (2002). Do lipid-lowering drugs cause erectile dysfunction? A systematic review. *Fam Pract* 19, 95-98.

Roberts, R. A., Chevalier, S., Hasmall, S. C., James, N. H., Cosulich, S. C., and Macdonald, N. (2002). PPAR alpha and the regulation of cell division and apoptosis. *Toxicology* 181-182, 167-170.

Roberts, R. A., James, N. H., Hasmall, S. C., Holden, P. R., Lambe, K., Macdonald, N., West, D., Woodyatt, N. J., and Whitcome, D. (2000). Apoptosis and proliferation in nongenotoxic carcinogenesis: species differences and role of PPARalpha. *Toxicol Lett* 112-113, 49-57.

Robin-Jagerschmidt, C., Wurtz, J. M., Guillot, B., Gofflo, D., Benhamou, B., Vergezac, A., Ossart, C., Moras, D., and Philibert, D. (2000). Residues in the ligand binding domain that confer progestin or glucocorticoid specificity and modulate the receptor transactivation capacity. *Mol Endocrinol* 14, 1028-1037.

Robyr, D., Wolffe, A. P., and Wahli, W. (2000). Nuclear hormone receptor coregulators in action: diversity for shared tasks. *Mol Endocrinol* 14, 329-347.

Rochel, N., Wurtz, J. M., Mitschler, A., Klaholz, B., and Moras, D. (2000). The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol Cell* 5, 173-179.

Rodriguez de Fonseca, F., Navarro, M., Gomez, R., Escuredo, L., Nava, F., Fu, J., Murillo-Rodriguez, E., Giuffrida, A., LoVerme, J., Gaetani, S., et al. (2001). An anorexic lipid mediator regulated by feeding. *Nature* 414, 209-212.

Rodriguez, J. C., Gil-Gomez, G., Hegardt, F. G., and Haro, D. (1994). Peroxisome proliferator-activated receptor mediates induction of the mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase gene by fatty acids. *J Biol Chem* 269, 18767-18772.

Rodriguez, J. C., Ortiz, J. A., Hegardt, F. G., and Haro, D. (1998). The hepatocyte nuclear factor 4 (HNF-4) represses the mitochondrial HMG-CoA synthase gene. *Biochem Biophys Res Commun* 242, 692-696.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 21, 327-335.

Rogerson, F. M., Yao, Y., Smith, B. J., and Fuller, P. J. (2004). Differences in the determinants of eplerenone, spironolactone and aldosterone binding to the mineralocorticoid receptor. *Clin Exp Pharmacol Physiol* 31, 704-709.

Rogerson, F. M., Yao, Y. Z., Smith, B. J., Dimopoulos, N., and Fuller, P. J. (2003). Determinants of spironolactone binding specificity in the mineralocorticoid receptor. *J Mol Endocrinol* 31, 573-582.

Rogojina, A. T., Orr, W. E., Song, B. K., and Geisert, E. E., Jr. (2003). Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines. *Mol Vis* 9, 482-496.

Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., et al. (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 24, 227-235.

Ross, R. (1999). Atherosclerosis is an inflammatory disease. *Am Heart J* 138, S419-420.

Ross, T. K., Darwish, H. M., Moss, V. E., and DeLuca, H. F. (1993). Vitamin D-influenced gene expression via a ligand-independent, receptor-DNA complex intermediate. *Proc Natl Acad Sci U S A* 90, 9257-9260.

Ross, T. K., Prahl, J. M., and DeLuca, H. F. (1991). Overproduction of rat 1,25-dihydroxyvitamin D<sub>3</sub> receptor in insect cells using the baculovirus expression system. *Proc Natl Acad Sci U S A* 88, 6555-6559.

Roux, S., Terouanne, B., Balaguer, P., Jausons-Loffreda, N., Pons, M., Chambon, P., Gronemeyer, H., and Nicolas, J. C. (1996). Mutation of isoleucine 747 by a threonine alters the ligand responsiveness of the human glucocorticoid receptor. *Mol Endocrinol* 10, 1214-1226.

Ruan, Y., Le Ber, P., Ng, H. H., and Liu, E. T. (2004). Interrogating the transcriptome. *Trends Biotechnol* 22, 23-30.

Ruau, D., Duarte, J., Ourjdal, T., Perriere, G., Laudet, V., and Robinson-Rechavi, M. (2004). Update of NUREBASE: nuclear hormone receptor functional genomics. *Nucleic Acids Res* 32, D165-167.

- Rupprecht, R., Reul, J. M., van Steensel, B., Spengler, D., Soder, M., Berning, B., Holsboer, F., and Damm, K. (1993). Pharmacological and functional characterization of human mineralocorticoid and glucocorticoid receptor ligands. *Eur J Pharmacol* 247, 145-154.
- Russell, D. W. (2003). The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem* 72, 137-174.
- Ryan, K. E., McCance, D. R., Powell, L., McMahon, R., and Trimble, E. R. (2006). Fenofibrate and pioglitazone improve endothelial function and reduce arterial stiffness in obese glucose tolerant men. *Atherosclerosis*.
- Sablin, E. P., Krylova, I. N., Fletterick, R. J., and Ingraham, H. A. (2003). Structural basis for ligand-independent activation of the orphan nuclear receptor LRH-1. *Mol Cell* 11, 1575-1585.
- Sabol, S. L., Brewer, H. B., Jr., and Santamarina-Fojo, S. (2005). The human ABCG1 gene: identification of LXR response elements that modulate expression in macrophages and liver. *J Lipid Res* 46, 2151-2167.
- Sack, J. S., Kish, K. F., Wang, C., Attar, R. M., Kiefer, S. E., An, Y., Wu, G. Y., Scheffler, J. E., Salvati, M. E., Krystek, S. R., Jr., et al. (2001). Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc Natl Acad Sci U S A* 98, 4904-4909.
- Saez, E., Tontonoz, P., Nelson, M. C., Alvarez, J. G., Ming, U. T., Baird, S. M., Thomazy, V. A., and Evans, R. M. (1998). Activators of the nuclear receptor PPARgamma enhance colon polyp formation. *Nat Med* 4, 1058-1061.
- Saha, S., Sparks, A. B., Rago, C., Akmaev, V., Wang, C. J., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. (2002). Using the transcriptome to annotate the genome. *Nat Biotechnol* 20, 508-512.
- Sampath, H., and Ntambi, J. M. (2005). Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annu Rev Nutr* 25, 317-340.
- Sandberg, M. B., Bloksgaard, M., Duran-Sandoval, D., Duval, C., Staels, B., and Mandrup, S. (2005). The gene encoding acyl-CoA-binding protein is subject to metabolic regulation by both sterol regulatory element-binding protein and peroxisome proliferator-activated receptor alpha in hepatocytes. *J Biol Chem* 280, 5258-5266.
- Sandler, B., Webb, P., Apriletti, J. W., Huber, B. R., Togashi, M., Cunha Lima, S. T., Juric, S., Nilsson, S., Wagner, R., Fletterick, R. J., and Baxter, J. D. (2004). Thyroxine-thyroid hormone receptor interactions. *J Biol Chem* 279, 55801-55808.
- Sap, J., Munoz, A., Damm, K., Goldberg, Y., Ghysdael, J., Leutz, A., Beug, H., and Vennstrom, B. (1986). The c-erb-A protein is a high-affinity receptor for thyroid hormone. *Nature* 324, 635-640.
- Sapolsky, R. M., Romero, L. M., and Munck, A. U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocr Rev* 21, 55-89.
- Sarkar, J., Qi, C., Guo, D., Ahmed, M. R., Jia, Y., Usuda, N., Viswakarma, N., Rao, M. S., and Reddy, J. K. (2007). Transcription coactivator PRIP, the peroxisome proliferator-activated receptor (PPAR)-interacting protein, is redundant for the function of nuclear receptors PParalpha and CAR, the constitutive androstane receptor, in mouse liver. *Gene Expr* 13, 255-269.
- Sato, O., Kuriki, C., Fukui, Y., and Motojima, K. (2002). Dual promoter structure of mouse and human fatty acid translocase/CD36 genes and unique transcriptional activation by peroxisome proliferator-activated receptor alpha and gamma ligands. *J Biol Chem* 277, 15703-15711.
- Sato, O., Takanashi, N., and Motojima, K. (2007). Third promoter and differential regulation of mouse and human fatty acid translocase/CD36 genes. *Mol Cell Biochem* 299, 37-43.
- Sauerberg, P., Pettersson, I., Jeppesen, L., Bury, P. S., Mogensen, J. P., Wassermann, K., Brand, C. L., Sturis, J., Woldike, H. F., Fleckner, J., et al. (2002). Novel tricyclic-alpha-alkyloxyphenylpropionic acids: dual PPARalpha/gamma agonists with hypolipidemic and antidiabetic activity. *J Med Chem* 45, 789-804.
- Schachtrup, C., Emmler, T., Bleck, B., Sandqvist, A., and Spener, F. (2004). Functional analysis of peroxisome-proliferator-responsive element motifs in genes of fatty acid-binding proteins. *Biochem J* 382, 239-245.
- Schena, M. (1996). Genome analysis with gene expression microarrays. *Bioessays* 18, 427-431.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* 93, 10614-10619.
- Schiffrin, E. L. (2005). Peroxisome proliferator-activated receptors and cardiovascular remodeling. *Am J Physiol Heart Circ Physiol* 288, H1037-1043.
- Schmidt, A., Endo, N., Rutledge, S. J., Vogel, R., Shinar, D., and Rodan, G. A. (1992). Identification of a new member of the steroid hormone receptor superfamily that is activated by a peroxisome proliferator and fatty acids. *Mol Endocrinol* 6, 1634-1641.
- Schoonjans, K., Peinado-Onsurbe, J., Lefebvre, A. M., Heyman, R. A., Briggs, M., Deeb, S., Staels, B., and Auwerx, J. (1996a). PPARalpha and PPARgamma activators direct a distinct tissue-specific transcriptional response via a PPRE in the lipoprotein lipase gene. *Embo J* 15, 5336-5348.

- Schoonjans, K., Staels, B., and Auwerx, J. (1996b). The peroxisome proliferator activated receptors (PPARS) and their effects on lipid metabolism and adipocyte differentiation. *Biochim Biophys Acta* *1302*, 93-109.
- Schoonjans, K., Watanabe, M., Suzuki, H., Mahfoudi, A., Krey, G., Wahli, W., Grimaldi, P., Staels, B., Yamamoto, T., and Auwerx, J. (1995). Induction of the acyl-coenzyme A synthetase gene by fibrates and fatty acids is mediated by a peroxisome proliferator response element in the C promoter. *J Biol Chem* *270*, 19269-19276.
- Schrader, M., and Fahimi, H. D. (2006). Peroxisomes and oxidative stress. *Biochim Biophys Acta* *1763*, 1755-1766.
- Schroepfer, G. J., Jr. (2000). Oxysterols: modulators of cholesterol metabolism and other processes. *Physiol Rev* *80*, 361-554.
- Schuetz, E. G., Strom, S., Yasuda, K., Lecureur, V., Assem, M., Brimer, C., Lamba, J., Kim, R. B., Ramachandran, V., Komoroski, B. J., et al. (2001). Disrupted bile acid homeostasis reveals an unexpected interaction among nuclear hormone receptors, transporters, and cytochrome P450. *J Biol Chem* *276*, 39411-39418.
- Schulman, I. G., Shao, G., and Heyman, R. A. (1998). Transactivation by retinoid X receptor-peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ) heterodimers: intermolecular synergy requires only the PPAR $\gamma$  hormone-dependent activation function. *Mol Cell Biol* *18*, 3483-3494.
- Schultz, J. R., Tu, H., Luk, A., Repa, J. J., Medina, J. C., Li, L., Schwendner, S., Wang, S., Thoolen, M., Mangelsdorf, D. J., et al. (2000). Role of LXR $\alpha$ s in control of lipogenesis. *Genes Dev* *14*, 2831-2838.
- Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993a). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* *75*, 567-578.
- Schwabe, J. W., Chapman, L., Finch, J. T., Rhodes, D., and Neuhaus, D. (1993b). DNA recognition by the oestrogen receptor: from solution to the crystal. *Structure* *1*, 187-204.
- Seedorf, U., Raabe, M., Ellinghaus, P., Kannenberg, F., Fobker, M., Engel, T., Denis, S., Wouters, F., Wirtz, K. W., Wanders, R. J., et al. (1998). Defective peroxisomal catabolism of branched fatty acyl coenzyme A in mice lacking the sterol carrier protein-2/sterol carrier protein-x gene function. *Genes Dev* *12*, 1189-1201.
- Semple, R. K., Chatterjee, V. K., and O'Rahilly, S. (2006). PPAR $\gamma$  and human metabolic disease. *J Clin Invest* *116*, 581-589.
- Seneko-Effenberger, K., Chen, S., Brace-Sinnokrak, E., Bonzo, J. A., Yueh, M. F., Argikar, U., Kaeding, J., Trottier, J., Remmel, R. P., Ritter, J. K., et al. (2007). Expression of the human UGT1 locus in transgenic mice by 4-chloro-6-(2,3-xylidino)-2-pyrimidinylthioacetic acid (WY-14643) and implications on drug metabolism through peroxisome proliferator-activated receptor alpha activation. *Drug Metab Dispos* *35*, 419-427.
- Seoane, S., and Perez-Fernandez, R. (2006). The vitamin D receptor represses transcription of the pituitary transcription factor Pit-1 gene without involvement of the retinoid X receptor. *Mol Endocrinol* *20*, 735-748.
- Seol, W., Choi, H. S., and Moore, D. D. (1996). An orphan nuclear hormone receptor that lacks a DNA binding domain and heterodimerizes with other receptors. *Science* *272*, 1336-1339.
- Seree, E., Villard, P. H., Pascussi, J. M., Pineau, T., Maurel, P., Nguyen, Q. B., Fallone, F., Martin, P. M., Champion, S., Lacarelle, B., et al. (2004). Evidence for a new human CYP1A1 regulation pathway involving PPAR-alpha and 2 PPRE sites. *Gastroenterology* *127*, 1436-1445.
- Setchell, K. D., Rodrigues, C. M., Clerici, C., Solinas, A., Morelli, A., Gartung, C., and Boyer, J. (1997). Bile acid concentrations in human and rat liver tissue and in hepatocyte nuclei. *Gastroenterology* *112*, 226-235.
- Setchell, K. D., Schwarz, M., O'Connell, N. C., Lund, E. G., Davis, D. L., Lathe, R., Thompson, H. R., Leslie Tyson, R., Sokol, R. J., and Russell, D. W. (1998). Identification of a new inborn error in bile acid synthesis: mutation of the oxysterol 7alpha-hydroxylase gene causes severe neonatal liver disease. *J Clin Invest* *102*, 1690-1703.
- Shaffer, P. L., and Gewirth, D. T. (2002). Structural basis of VDR-DNA interactions on direct repeat response elements. *Embo J* *21*, 2242-2252.
- Shaffer, P. L., and Gewirth, D. T. (2004). Structural analysis of RXR-VDR interactions on DR3 DNA. *J Steroid Biochem Mol Biol* *89-90*, 215-219.
- Shaffer, P. L., Jivan, A., Dollins, D. E., Claessens, F., and Gewirth, D. T. (2004). Structural basis of androgen receptor binding to selective androgen response elements. *Proc Natl Acad Sci U S A* *101*, 4758-4763.
- Shaffer, P. L., McDonnell, D. P., and Gewirth, D. T. (2005). Characterization of transcriptional activation and DNA-binding functions in the hinge region of the vitamin D receptor. *Biochemistry* *44*, 2678-2685.
- Shan, L., Vincent, J., Brunzelle, J. S., Dussault, I., Lin, M., Ianculescu, I., Sherman, M. A., Forman, B. M., and Fernandez, E. J. (2004). Structure of the murine constitutive androstane receptor complexed to androstenol: a molecular basis for inverse agonism. *Mol Cell* *16*, 907-917.

- Sheppard, K. E., and Funder, J. W. (1987). Equivalent affinity of aldosterone and corticosterone for type I receptors in kidney and hippocampus: direct binding studies. *J Steroid Biochem* 28, 737-742.
- Sher, T., Yi, H. F., McBride, O. W., and Gonzalez, F. J. (1993). cDNA cloning, chromosomal mapping, and functional characterization of the human peroxisome proliferator activated receptor. *Biochemistry* 32, 5598-5604.
- Shi, G. Q., Dropinski, J. F., McKeever, B. M., Xu, S., Becker, J. W., Berger, J. P., MacNaul, K. L., Elbrecht, A., Zhou, G., Doepper, T. W., *et al.* (2005). Design and synthesis of alpha-aryloxyphenylacetic acid derivatives: a novel class of PPARalpha/gamma dual agonists with potent antihyperglycemic and lipid modulating activity. *J Med Chem* 48, 4457-4468.
- Shi, L., Reid, L. H., Jones, W. D., Shippy, R., Warrington, J. A., Baker, S. C., Collins, P. J., de Longueville, F., Kawasaki, E. S., Lee, K. Y., *et al.* (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24, 1151-1161.
- Shi, Y., Downes, M., Xie, W., Kao, H. Y., Ordentlich, P., Tsai, C. C., Hon, M., and Evans, R. M. (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev* 15, 1140-1151.
- Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A., and Greene, G. L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95, 927-937.
- Shiau, A. K., Barstad, D., Radek, J. T., Meyers, M. J., Nettles, K. W., Katzenellenbogen, B. S., Katzenellenbogen, J. A., Agard, D. A., and Greene, G. L. (2002). Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism. *Nat Struct Biol* 9, 359-364.
- Shimkets, R. A., Lowe, D. G., Tai, J. T., Sehl, P., Jin, H., Yang, R., Predki, P. F., Rothberg, B. E., Murtha, M. T., Roth, M. E., *et al.* (1999). Gene expression analysis by transcript profiling coupled to a gene database query. *Nat Biotechnol* 17, 798-803.
- Shin, M., Kim, I., Inoue, Y., Kimura, S., and Gonzalez, F. J. (2006). Regulation of mouse hepatic alpha-amino-beta-carboxymuconate-epsilon-semialdehyde decarboxylase, a key enzyme in the tryptophan-nicotinamide adenine dinucleotide pathway, by hepatocyte nuclear factor 4alpha and peroxisome proliferator-activated receptor alpha. *Mol Pharmacol* 70, 1281-1290.
- Shipley, J. M., and Waxman, D. J. (2003). Down-regulation of STAT5b transcriptional activity by ligand-activated peroxisome proliferator-activated receptor (PPAR) alpha and PPARgamma. *Mol Pharmacol* 64, 355-364.
- Shippy, R., Fulmer-Smentek, S., Jensen, R. V., Jones, W. D., Wolber, P. K., Johnson, C. D., Pine, P. S., Boysen, C., Guo, X., Chudin, E., *et al.* (2006). Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nat Biotechnol* 24, 1123-1131.
- Shippy, R., Sendera, T. J., Lockner, R., Palaniappan, C., Kaysser-Kranich, T., Watts, G., and Alsobrook, J. (2004). Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. *BMC Genomics* 5, 61.
- Shu, H., Wong, B., Zhou, G., Li, Y., Berger, J., Woods, J. W., Wright, S. D., and Cai, T. Q. (2000). Activation of PPARalpha or gamma reduces secretion of matrix metalloproteinase 9 but not interleukin 8 from human monocytic THP-1 cells. *Biochem Biophys Res Commun* 267, 345-349.
- Sierk, M. L., Zhao, Q., and Rastinejad, F. (2001). DNA deformability as a recognition feature in the reverb response element. *Biochemistry* 40, 12833-12843.
- Silva, A. P., Chen, J., Carraro, D. M., Wang, S. M., and Camargo, A. A. (2004a). Generation of longer 3' cDNA fragments from massively parallel signature sequencing tags. *Nucleic Acids Res* 32, e94.
- Silva, A. P., De Souza, J. E., Galante, P. A., Riggins, G. J., De Souza, S. J., and Camargo, A. A. (2004b). The impact of SNPs on the interpretation of SAGE and MPSS experimental data. *Nucleic Acids Res* 32, 6104-6110.
- Simon, R. M. (2003). Design and analysis of DNA microarray investigations (New York: Springer).
- Simon, T. C., Roth, K. A., and Gordon, J. I. (1993). Use of transgenic mice to map cis-acting elements in the liver fatty acid-binding protein gene (Fabpl) that regulate its cell lineage-specific, differentiation-dependent, and spatial patterns of expression in the gut epithelium and in the liver acinus. *J Biol Chem* 268, 18345-18358.
- Simone, C. (2006). SWI/SNF: the crossroads where extracellular signaling pathways meet chromatin. *J Cell Physiol* 207, 309-314.
- Simons, S. S., Jr., Sistare, F. D., and Chakraborti, P. K. (1989). Steroid binding activity is retained in a 16-kDa fragment of the steroid binding domain of rat glucocorticoid receptors. *J Biol Chem* 264, 14493-14497.
- Skrede, S., Bremer, J., Berge, R. K., and Rustan, A. C. (1994). Stimulation of fatty acid oxidation by a 3-thia fatty acid reduces triacylglycerol secretion in cultured rat hepatocytes. *J Lipid Res* 35, 1395-1404.
- Sluder, A. E., Mathews, S. W., Hough, D., Yin, V. P., and Maina, C. V. (1999). The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. *Genome Res* 9, 103-120.

- Smith, J. L., Lewindon, P. J., Hoskins, A. C., Pereira, T. N., Setchell, K. D., O'Connell, N. C., Shepherd, R. W., and Ramm, G. A. (2004). Endogenous ursodeoxycholic acid and cholic acid in liver disease due to cystic fibrosis. *Hepatology* 39, 1673-1682.
- Smith, W. C., Nakshatri, H., Leroy, P., Rees, J., and Chambon, P. (1991). A retinoic acid response element is present in the mouse cellular retinol binding protein I (mCRBPI) promoter. *Embo J* 10, 2223-2230.
- Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3, Article3.
- Smyth, G. K. (2005). Limma: linear models for microarray data, In Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, eds. (New York: Springer), pp. pages 397-420.
- Soares, M. B. (1997). Identification and cloning of differentially expressed genes. *Curr Opin Biotechnol* 8, 542-546.
- Solaas, K., Kase, B. F., Pham, V., Bamberg, K., Hunt, M. C., and Alexson, S. E. (2004). Differential regulation of cytosolic and peroxisomal bile acid amidation by PPAR alpha activation favors the formation of unconjugated bile acids. *J Lipid Res* 45, 1051-1060.
- Solomon, I. H., Hager, J. M., Safi, R., McDonnell, D. P., Redinbo, M. R., and Ortlund, E. A. (2005). Crystal structure of the human LRH-1 DBD-DNA complex reveals Ftz-F1 domain positioning is required for receptor activity. *J Mol Biol* 354, 1091-1102.
- Song, C., and Liao, S. (2000). Cholestenone is a naturally occurring ligand for liver X receptor alpha. *Endocrinology* 141, 4180-4184.
- Song, C. S., Echchgadda, I., Baek, B. S., Ahn, S. C., Oh, T., Roy, A. K., and Chatterjee, B. (2001). Dehydroepiandrosterone sulfotransferase gene induction by bile acid activated farnesoid X receptor. *J Biol Chem* 276, 42549-42556.
- Sonoda, J., Xie, W., Rosenfeld, J. M., Barwick, J. L., Guzelian, P. S., and Evans, R. M. (2002). Regulation of a xenobiotic sulfonation cascade by nuclear pregnane X receptor (PXR). *Proc Natl Acad Sci U S A* 99, 13801-13806.
- Speed, T. P. (2003). Statistical analysis of gene expression microarray data (Boca Raton, FL: Chapman & Hall/CRC).
- Spitz, I. M. (2003). Progesterone antagonists and progesterone receptor modulators: an overview. *Steroids* 68, 981-993.
- Srivastava, R. A., Jahagirdar, R., Azhar, S., Sharma, S., and Bisgaier, C. L. (2006). Peroxisome proliferator-activated receptor-alpha selective ligand reduces adiposity, improves insulin sensitivity and inhibits atherosclerosis in LDL receptor-deficient mice. *Mol Cell Biochem* 285, 35-50.
- Staels, B., and Auwerx, J. (1998). Regulation of apo A-I gene expression by fibrates. *Atherosclerosis* 137 Suppl, S19-23.
- Staels, B., Koenig, W., Habib, A., Merval, R., Lebret, M., Torra, I. P., Delerive, P., Fadel, A., Chinetti, G., Fruchart, J. C., et al. (1998). Activation of human aortic smooth-muscle cells is inhibited by PPARalpha but not by PPARgamma activators. *Nature* 393, 790-793.
- Staels, B., Peinado-Onsurbe, J., and Auwerx, J. (1992a). Down-regulation of hepatic lipase gene expression and activity by fenofibrate. *Biochim Biophys Acta* 1123, 227-230.
- Staels, B., van Tol, A., Skretting, G., and Auwerx, J. (1992b). Lecithin:cholesterol acyltransferase gene expression is regulated in a tissue-selective manner by fibrates. *J Lipid Res* 33, 727-735.
- Staels, B., Vu-Dac, N., Kosykh, V. A., Saladin, R., Fruchart, J. C., Dallongeville, J., and Auwerx, J. (1995). Fibrates downregulate apolipoprotein C-III expression independent of induction of peroxisomal acyl coenzyme A oxidase. A potential mechanism for the hypolipidemic action of fibrates. *J Clin Invest* 95, 705-712.
- Stahlberg, D., Reihner, E., Rudling, M., Berglund, L., Einarsson, K., and Angelin, B. (1995). Influence of bezafibrate on hepatic cholesterol metabolism in gallstone patients: reduced activity of cholesterol 7 alpha-hydroxylase. *Hepatology* 21, 1025-1030.
- Stahlhut, R. W., van Wijngaarden, E., Dye, T. D., Cook, S., and Swan, S. H. (2007). Concentrations of urinary phthalate metabolites are associated with increased waist circumference and insulin resistance in adult U.S. males. *Environ Health Perspect* 115, 876-882.
- Stanley, T. B., Leesnitzer, L. M., Montana, V. G., Galardi, C. M., Lambert, M. H., Holt, J. A., Xu, H. E., Moore, L. B., Blanchard, S. G., and Stimmel, J. B. (2003). Subtype specific effects of peroxisome proliferator-activated receptor ligands on corepressor affinity. *Biochemistry* 42, 9278-9287.
- Stehlin-Gaon, C., Willmann, D., Zeyer, D., Sanglier, S., Van Dorsselaer, A., Renaud, J. P., Moras, D., and Schule, R. (2003). All-trans retinoic acid is a ligand for the orphan nuclear receptor ROR beta. *Nat Struct Biol* 10, 820-825.
- Stehlin, C., Wurtz, J. M., Steinmetz, A., Greiner, E., Schule, R., Moras, D., and Renaud, J. P. (2001). X-ray structure of the orphan nuclear receptor RORbeta ligand-binding domain in the active conformation. *Embo J* 20, 5822-5831.

- Steineger, H. H., Sorensen, H. N., Tugwood, J. D., Skrede, S., Spydevold, O., and Gautvik, K. M. (1994). Dexamethasone and insulin demonstrate marked and opposite regulation of the steady-state mRNA level of the peroxisomal proliferator-activated receptor (PPAR) in hepatic cells. Hormonal modulation of fatty-acid-induced transcription. *Eur J Biochem* 225, 967-974.
- Sterchele, P. F., Sun, H., Peterson, R. E., and Vanden Heuvel, J. P. (1996). Regulation of peroxisome proliferator-activated receptor-alpha mRNA in rat liver. *Arch Biochem Biophys* 326, 281-289.
- Stoner, M. A., Auerbach, S. S., Zamule, S. M., Strom, S. C., and Omiecinski, C. J. (2007). Transactivation of a DR-1 PPRE by a human constitutive androstane receptor variant expressed from internal protein translation start sites. *Nucleic Acids Res* 35, 2177-2190.
- Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society Series B-Methodological* 66, 187-205.
- Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100, 9440-9445.
- Su, A. I., Cooke, M. P., Ching, K. A., Hakak, Y., Walker, J. R., Wiltshire, T., Orth, A. P., Vega, R. G., Sapino, L. M., Moqrich, A., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A* 99, 4465-4470.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545-15550.
- Sucov, H. M., Murakami, K. K., and Evans, R. M. (1990). Characterization of an autoregulated response element in the mouse retinoic acid receptor type beta gene. *Proc Natl Acad Sci U S A* 87, 5392-5396.
- Suga, T. (2004). Hepatocarcinogenesis by peroxisome proliferators. *J Toxicol Sci* 29, 1-12.
- Sugden, M. C., Bulmer, K., Gibbons, G. F., and Holness, M. J. (2001). Role of peroxisome proliferator-activated receptor-alpha in the mechanism underlying changes in renal pyruvate dehydrogenase kinase isoform 4 protein expression in starvation and after refeeding. *Arch Biochem Biophys* 395, 246-252.
- Sugden, M. C., Bulmer, K., Gibbons, G. F., Knight, B. L., and Holness, M. J. (2002). Peroxisome-proliferator-activated receptor-alpha (PPAR $\alpha$ ) deficiency leads to dysregulation of hepatic lipid and carbohydrate metabolism by fatty acids and insulin. *Biochem J* 364, 361-368.
- Suino, K., Peng, L., Reynolds, R., Li, Y., Cha, J. Y., Repa, J. J., Kliewer, S. A., and Xu, H. E. (2004). The nuclear xenobiotic receptor CAR: structural determinants of constitutive activation and heterodimerization. *Mol Cell* 16, 893-905.
- Sumanasekera, W. K., Tien, E. S., Davis, J. W., 2nd, Turpey, R., Perdew, G. H., and Vanden Heuvel, J. P. (2003a). Heat shock protein-90 (Hsp90) acts as a repressor of peroxisome proliferator-activated receptor-alpha (PPAR $\alpha$ ) and PPAR $\beta$  activity. *Biochemistry* 42, 10726-10735.
- Sumanasekera, W. K., Tien, E. S., Turpey, R., Vanden Heuvel, J. P., and Perdew, G. H. (2003b). Evidence that peroxisome proliferator-activated receptor alpha is complexed with the 90-kDa heat shock protein and the hepatitis virus B X-associated protein 2. *J Biol Chem* 278, 4467-4473.
- Surapureddi, S., Viswakarma, N., Yu, S., Guo, D., Rao, M. S., and Reddy, J. K. (2006). PRIC320, a transcription coactivator, isolated from peroxisome proliferator-binding protein complex. *Biochem Biophys Res Commun* 343, 535-543.
- Surapureddi, S., Yu, S., Bu, H., Hashimoto, T., Yeldandi, A. V., Kashireddy, P., Cherkaoui-Malki, M., Qi, C., Zhu, Y. J., Rao, M. S., and Reddy, J. K. (2002). Identification of a transcriptionally active peroxisome proliferator-activated receptor alpha -interacting cofactor complex in rat liver and characterization of PRIC285 as a coactivator. *Proc Natl Acad Sci U S A* 99, 11836-11841.
- Suzuki, A., Okamoto, S., Lee, S., Saito, K., Shiuchi, T., and Minokoshi, Y. (2007). Leptin stimulates fatty acid oxidation and peroxisome proliferator-activated receptor alpha gene expression in mouse C2C12 myoblasts by changing the subcellular localization of the alpha2 form of AMP-activated protein kinase. *Mol Cell Biol* 27, 4317-4327.
- Svensson, L. T., Alexson, S. E., and Hiltunen, J. K. (1995a). Very long chain and long chain acyl-CoA thioesterases in rat liver mitochondria. Identification, purification, characterization, and induction by peroxisome proliferators. *J Biol Chem* 270, 12177-12183.
- Svensson, L. T., Engberg, S. T., Aoyama, T., Usuda, N., Alexson, S. E., and Hashimoto, T. (1998). Molecular cloning and characterization of a mitochondrial peroxisome proliferator-induced acyl-CoA thioesterase from rat liver. *Biochem J* 329 ( Pt 3), 601-608.
- Svensson, L. T., Wilcke, M., and Alexson, S. E. (1995b). Peroxisome proliferators differentially regulate long-chain acyl-CoA thioesterases in rat liver. *Eur J Biochem* 230, 813-820.

- Svensson, S., Ostberg, T., Jacobsson, M., Norstrom, C., Stefansson, K., Hallen, D., Johansson, I. C., Zachrisson, K., Ogg, D., and Jendeberg, L. (2003). Crystal structure of the heterodimeric complex of LXRalpha and RXRbeta ligand-binding domains in a fully agonistic conformation. *Embo J* 22, 4625-4633.
- Swan, S. H. (2000). Intrauterine exposure to diethylstilbestrol: long-term effects in humans. *Apmis* 108, 793-804.
- Swan, S. H. (2006). Does our environment affect our fertility? Some examples to help reframe the question. *Semin Reprod Med* 24, 142-146.
- Swan, S. H., Main, K. M., Liu, F., Stewart, S. L., Kruse, R. L., Calafat, A. M., Mao, C. S., Redmon, J. B., Ternand, C. L., Sullivan, S., and Teague, J. L. (2005). Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environ Health Perspect* 113, 1056-1061.
- Tabor, D. E., Kim, J. B., Spiegelman, B. M., and Edwards, P. A. (1999). Identification of conserved cis-elements and transcription factors required for sterol-regulated transcription of stearoyl-CoA desaturase 1 and 2. *J Biol Chem* 274, 20603-20610.
- Taimi, M., Helvig, C., Wisniewski, J., Ramshaw, H., White, J., Amad, M., Korczak, B., and Petkovich, M. (2004). A novel human cytochrome P450, CYP26C1, involved in metabolism of 9-cis and all-trans isomers of retinoic acid. *J Biol Chem* 279, 77-85.
- Takada, I., Yu, R. T., Xu, H. E., Lambert, M. H., Montana, V. G., Kliewer, S. A., Evans, R. M., and Umesono, K. (2000). Alteration of a single amino acid in peroxisome proliferator-activated receptor-alpha (PPAR alpha) generates a PPAR delta phenotype. *Mol Endocrinol* 14, 733-740.
- Takeda, A. N., Pinon, G. M., Bens, M., Fagart, J., Rafestin-Oblin, M. E., and Vandewalle, A. (2007). The synthetic androgen methyltrienolone (r1881) acts as a potent antagonist of the mineralocorticoid receptor. *Mol Pharmacol* 71, 473-482.
- Takeda, Y. (2004). Pleiotropic actions of aldosterone and the effects of eplerenone, a selective mineralocorticoid receptor antagonist. *Hypertens Res* 27, 781-789.
- Takeshita, A., Ozawa, Y., and Chin, W. W. (2000). Nuclear receptor coactivators facilitate vitamin D receptor homodimer action on direct repeat hormone response elements. *Endocrinology* 141, 1281-1284.
- Takeuchi, S., Matsuda, T., Kobayashi, S., Takahashi, T., and Kojima, H. (2006). In vitro screening of 200 pesticides for agonistic activity via mouse peroxisome proliferator-activated receptor (PPAR)alpha and PPARgamma and quantitative analysis of in vivo induction pathway. *Toxicol Appl Pharmacol* 217, 235-244.
- Tammi, A., Ronnemaa, T., Valsta, L., Seppanen, R., Rask-Nissila, L., Miettinen, T. A., Gylling, H., Viikari, J., Anttolainen, M., and Simell, O. (2001). Dietary plant sterols alter the serum plant sterol concentration but not the cholesterol precursor sterol concentrations in young children (the STRIP Study). Special Turku Coronary Risk Factor Intervention Project. *J Nutr* 131, 1942-1945.
- Tan, N. S., Shaw, N. S., Vinckenbosch, N., Liu, P., Yasmin, R., Desvergne, B., Wahli, W., and Noy, N. (2002). Selective cooperation between fatty acid binding proteins and peroxisome proliferator-activated receptors in regulating transcription. *Mol Cell Biol* 22, 5114-5127.
- Tan, P. K., Downey, T. J., Spitznagel, E. L., Jr., Xu, P., Fu, D., Dimitrov, D. S., Lempicki, R. A., Raaka, B. M., and Cam, M. C. (2003). Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* 31, 5676-5684.
- Tan, Q., Blizzard, T. A., Morgan, J. D., 2nd, Birzin, E. T., Chan, W., Yang, Y. T., Pai, L. Y., Hayes, E. C., DaSilva, C. A., Warrier, S., et al. (2005). Estrogen receptor ligands. Part 10: Chromanes: old scaffolds for new SERAMs. *Bioorg Med Chem Lett* 15, 1675-1681.
- Tanenbaum, D. M., Wang, Y., Williams, S. P., and Sigler, P. B. (1998). Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc Natl Acad Sci U S A* 95, 5998-6003.
- Tang, C., Cho, H. P., Nakamura, M. T., and Clarke, S. D. (2003). Regulation of human delta-6 desaturase gene transcription: identification of a functional direct repeat-1 element. *J Lipid Res* 44, 686-695.
- Tang, G. W., and Russell, R. M. (1990). 13-cis-retinoic acid is an endogenous compound in human serum. *J Lipid Res* 31, 175-182.
- Taylor, S. W., Fahy, E., Zhang, B., Glenn, G. M., Warnock, D. E., Wiley, S., Murphy, A. N., Gaucher, S. P., Capaldi, R. A., Gibson, B. W., and Ghosh, S. S. (2003). Characterization of the human heart mitochondrial proteome. *Nat Biotechnol* 21, 281-286.
- Thissen, J. P., Ketelslegers, J. M., and Underwood, L. E. (1994). Nutritional regulation of the insulin-like growth factors. *Endocr Rev* 15, 80-101.
- Thomas, J., Bramlett, K. S., Montrose, C., Foxworthy, P., Eacho, P. I., McCann, D., Cao, G., Kiefer, A., McCowan, J., Yu, K. L., et al. (2003). A chemical switch regulates fibrate specificity for peroxisome proliferator-activated receptor alpha (PPARalpha) versus liver X receptor. *J Biol Chem* 278, 2403-2410.
- Thompson, C. C., Weinberger, C., Lebo, R., and Evans, R. M. (1987). Identification of a novel thyroid hormone receptor expressed in the mammalian central nervous system. *Science* 237, 1610-1614.
- Tian, L., Greenberg, S. A., Kong, S. W., Altschuler, J., Kohane, I. S., and Park, P. J. (2005). Discovering statistically significant pathways in expression profiling studies. *Proc Natl Acad Sci U S A* 102, 13544-13549.

- Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99, 6567-6572.
- Tilley, W. D., Marcelli, M., Wilson, J. D., and McPhaul, M. J. (1989). Characterization and expression of a cDNA encoding the human androgen receptor. Proc Natl Acad Sci U S A 86, 327-331.
- Tobin, K. A., Steineger, H. H., Alberti, S., Spydevold, O., Auwerx, J., Gustafsson, J. A., and Nebb, H. I. (2000). Cross-talk between fatty acid and cholesterol metabolism mediated by liver X receptor-alpha. Mol Endocrinol 14, 741-752.
- Tocchini-Valentini, G., Rochel, N., Wurtz, J. M., Mitschler, A., and Moras, D. (2001). Crystal structures of the vitamin D receptor complexed to superagonist 20-epi ligands. Proc Natl Acad Sci U S A 98, 5491-5496.
- Tocchini-Valentini, G., Rochel, N., Wurtz, J. M., and Moras, D. (2004). Crystal structures of the vitamin D nuclear receptor liganded with the vitamin D side chain analogues calcipotriol and seocalcitol, receptor agonists of clinical importance. Insights into a structural basis for the switching of calcipotriol to a receptor antagonist by further side chain modification. J Med Chem 47, 1956-1961.
- Toda, K., Okada, T., Miyaura, C., and Saibara, T. (2003). Fenofibrate, a ligand for PPARalpha, inhibits aromatase cytochrome P450 expression in the ovary of mouse. J Lipid Res 44, 265-270.
- Tomaru, T., Satoh, T., Yoshino, S., Ishizuka, T., Hashimoto, K., Monden, T., Yamada, M., and Mori, M. (2006). Isolation and characterization of a transcriptional cofactor and its novel isoform that bind the deoxyribonucleic acid-binding domain of peroxisome proliferator-activated receptor-gamma. Endocrinology 147, 377-388.
- Tong, W., Lucas, A. B., Shippy, R., Fan, X., Fang, H., Hong, H., Orr, M. S., Chu, T. M., Guo, X., Collins, P. J., et al. (2006). Evaluation of external RNA controls for the assessment of microarray performance. Nat Biotechnol 24, 1132-1139.
- Tontonoz, P., Graves, R. A., Budavari, A. I., Erdjument-Bromage, H., Lui, M., Hu, E., Tempst, P., and Spiegelman, B. M. (1994a). Adipocyte-specific transcription factor ARF6 is a heterodimeric complex of two nuclear hormone receptors, PPAR gamma and RXR alpha. Nucleic Acids Res 22, 5628-5634.
- Tontonoz, P., Hu, E., Graves, R. A., Budavari, A. I., and Spiegelman, B. M. (1994b). mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. Genes Dev 8, 1224-1234.
- Tontonoz, P., Hu, E., and Spiegelman, B. M. (1994c). Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor. Cell 79, 1147-1156.
- Tontonoz, P., Nagy, L., Alvarez, J. G., Thomazy, V. A., and Evans, R. M. (1998). PPARgamma promotes monocyte/macrophage differentiation and uptake of oxidized LDL. Cell 93, 241-252.
- Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E., and Chambon, P. (1989). The human estrogen receptor has two independent nonacidic transcriptional activation functions. Cell 59, 477-487.
- Tordjman, K., Bernal-Mizrachi, C., Zemany, L., Weng, S., Feng, C., Zhang, F., Leone, T. C., Coleman, T., Kelly, D. P., and Semenkovich, C. F. (2001). PPARalpha deficiency reduces insulin resistance and atherosclerosis in apoE-null mice. J Clin Invest 107, 1025-1034.
- Tordjman, K., Standley, K. N., Bernal-Mizrachi, C., Leone, T. C., Coleman, T., Kelly, D. P., and Semenkovich, C. F. (2002). PPARalpha suppresses insulin secretion and induces UCP2 in insulinoma cells. J Lipid Res 43, 936-943.
- Toth, I., Faredin, I., Mesko, E., Wolfling, J., and Schneider, G. (1995). In vitro binding of 16-methylated C18 and C19 steroid derivatives to the androgen receptor. Pharmacol Res 32, 217-221.
- Toussaint, J.-F., Jacob, M.-P., Lagrost, L., and Chapman, J. (2003). L'athérosclérose - Physiologie, diagnostics, thérapeutiques (Paris: Masson).
- Trapman, J., Klaassen, P., Kuiper, G. G., van der Korput, J. A., Faber, P. W., van Rooij, H. C., Geurts van Kessel, A., Voorhorst, M. M., Mulder, E., and Brinkmann, A. O. (1988). Cloning, structure and expression of a cDNA encoding the human androgen receptor. Biochem Biophys Res Commun 153, 241-248.
- Tremblay, A., Tremblay, G. B., Labrie, F., and Giguere, V. (1999). Ligand-independent recruitment of SRC-1 to estrogen receptor beta through phosphorylation of activation function AF-1. Mol Cell 3, 513-519.
- Tremblay, G. B., Tremblay, A., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Labrie, F., and Giguere, V. (1997). Cloning, chromosomal localization, and functional analysis of the murine estrogen receptor beta. Mol Endocrinol 11, 353-365.
- Trottier, J., Milkiewicz, P., Kaeding, J., Verreault, M., and Barbier, O. (2006). Coordinate regulation of hepatic bile acid oxidation and conjugation by nuclear receptors. Mol Pharm 3, 212-222.
- Trujillo, M. E., and Scherer, P. E. (2006). Adipose tissue-derived factors: impact on health and disease. Endocr Rev 27, 762-778.
- Tschop, M., Smiley, D. L., and Heiman, M. L. (2000). Ghrelin induces adiposity in rodents. Nature 407, 908-913.
- Tsuboyama-Kasaoka, N., Takahashi, M., Kim, H., and Ezaki, O. (1999). Up-regulation of liver uncoupling protein-2 mRNA by either fish oil feeding or fibrate administration in mice. Biochem Biophys Res Commun 257, 879-885.

- Tu, A. Y., and Albers, J. J. (1999). DNA sequences responsible for reduced promoter activity of human phospholipid transfer protein by fibrate. *Biochem Biophys Res Commun* 264, 802-807.
- Tu, A. Y., and Albers, J. J. (2001). Functional analysis of the transcriptional activity of the mouse phospholipid transfer protein gene. *Biochem Biophys Res Commun* 287, 921-926.
- Tu, N., Chen, H., Winnikes, U., Reinert, I., Pirke, K. M., and Lentes, K. U. (2000). Functional characterization of the 5'-flanking and the promoter region of the human UCP3 (hUCP3) gene. *Life Sci* 67, 2267-2279.
- Tudor, C., Feige, J. N., Pingali, H., Lohray, V. B., Wahli, W., Desvergne, B., Engelborghs, Y., and Gelman, L. (2007). Association with coregulators is the major determinant governing peroxisome proliferator-activated receptor mobility in living cells. *J Biol Chem* 282, 4417-4426.
- Tugwood, J. D., Issemann, I., Anderson, R. G., Bundell, K. R., McPheat, W. L., and Green, S. (1992). The mouse peroxisome proliferator activated receptor recognizes a response element in the 5' flanking sequence of the rat acyl CoA oxidase gene. *Embo J* 11, 433-439.
- Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98, 5116-5121.
- Tzameli, I., Fang, H., Ollero, M., Shi, H., Hamm, J. K., Kievit, P., Hollenberg, A. N., and Flier, J. S. (2004). Regulated production of a peroxisome proliferator-activated receptor-gamma ligand during an early phase of adipocyte differentiation in 3T3-L1 adipocytes. *J Biol Chem* 279, 36093-36102.
- Tzameli, I., Pissios, P., Schuetz, E. G., and Moore, D. D. (2000). The xenobiotic compound 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene is an agonist ligand for the nuclear receptor CAR. *Mol Cell Biol* 20, 2951-2958.
- Uchida, Y., Izai, K., Orii, T., and Hashimoto, T. (1992). Novel fatty acid beta-oxidation enzymes in rat liver mitochondria. II. Purification and properties of enoyl-coenzyme A (CoA) hydratase/3-hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase trifunctional protein. *J Biol Chem* 267, 1034-1041.
- Ueda, A., Hamadeh, H. K., Webb, H. K., Yamamoto, Y., Sueyoshi, T., Afshari, C. A., Lehmann, J. M., and Negishi, M. (2002). Diverse roles of the nuclear orphan receptor CAR in regulating hepatic genes in response to phenobarbital. *Mol Pharmacol* 61, 1-6.
- Umesono, K., and Evans, R. M. (1989). Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* 57, 1139-1146.
- Uppenberg, J., Svensson, C., Jaki, M., Bertilsson, G., Jendeberg, L., and Berkenstam, A. (1998). Crystal structure of the ligand binding domain of the human nuclear receptor PPARgamma. *J Biol Chem* 273, 31108-31112.
- Urizar, N. L., Liverman, A. B., Dodds, D. T., Silva, F. V., Ordentlich, P., Yan, Y., Gonzalez, F. J., Heyman, R. A., Mangelsdorf, D. J., and Moore, D. D. (2002). A natural product that lowers cholesterol as an antagonist ligand for FXR. *Science* 296, 1703-1706.
- Urnov, F. D., and Wolffe, A. P. (2001). Chromatin remodeling and transcriptional activation: the cast (in order of appearance). *Oncogene* 20, 2991-3006.
- van Tilborg, M. A., Bonvin, A. M., Hard, K., Davis, A. L., Maler, B., Boelens, R., Yamamoto, K. R., and Kaptein, R. (1995). Structure refinement of the glucocorticoid receptor-DNA binding domain from NMR data by relaxation matrix calculations. *J Mol Biol* 247, 689-700.
- Vanden Heuvel, J. P., Sterchele, P. F., Nesbit, D. J., and Peterson, R. E. (1993). Coordinate induction of acyl-CoA binding protein, fatty acid binding protein and peroxisomal beta-oxidation by peroxisome proliferators. *Biochim Biophys Acta* 1177, 183-190.
- Vanhooke, J. L., Benning, M. M., Bauer, C. B., Pike, J. W., and DeLuca, H. F. (2004). Molecular structure of the rat vitamin D receptor ligand binding domain complexed with 2-carbon-substituted vitamin D3 hormone analogues and a LXXLL-containing coactivator peptide. *Biochemistry* 43, 4101-4110.
- Varanasi, U., Chu, R., Huang, Q., Castellon, R., Yeldandi, A. V., and Reddy, J. K. (1996). Identification of a peroxisome proliferator-responsive element upstream of the human peroxisomal fatty acyl coenzyme A oxidase gene. *J Biol Chem* 271, 2147-2155.
- Vasios, G., Mader, S., Gold, J. D., Leid, M., Lutz, Y., Gaub, M. P., Chambon, P., and Gudas, L. (1991). The late retinoic acid induction of laminin B1 gene transcription involves RAR binding to the responsive element. *Embo J* 10, 1149-1158.
- Vega, R. B., Huss, J. M., and Kelly, D. P. (2000). The coactivator PGC-1 cooperates with peroxisome proliferator-activated receptor alpha in transcriptional control of nuclear genes encoding mitochondrial fatty acid oxidation enzymes. *Mol Cell Biol* 20, 1868-1876.
- Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* 270, 484-487.
- Venkateswaran, A., Laffitte, B. A., Joseph, S. B., Mak, P. A., Wilpitz, D. C., Edwards, P. A., and Tontonoz, P. (2000). Control of cellular cholesterol efflux by the nuclear oxysterol receptor LXR alpha. *Proc Natl Acad Sci U S A* 97, 12097-12102.

- Vidal-Puig, A., Jimenez-Linan, M., Lowell, B. B., Hamann, A., Hu, E., Spiegelman, B., Flier, J. S., and Moller, D. E. (1996). Regulation of PPAR gamma gene expression by nutrition and obesity in rodents. *J Clin Invest* 97, 2553-2561.
- Vidal-Puig, A. J., Considine, R. V., Jimenez-Linan, M., Werman, A., Pories, W. J., Caro, J. F., and Flier, J. S. (1997). Peroxisome proliferator-activated receptor gene expression in human tissues. Effects of obesity, weight loss, and regulation by insulin and glucocorticoids. *J Clin Invest* 99, 2416-2422.
- Viollet, B., Kahn, A., and Raymondjean, M. (1997). Protein kinase A-dependent phosphorylation modulates DNA-binding activity of hepatocyte nuclear factor 4. *Mol Cell Biol* 17, 4208-4219.
- Vivanco Ruiz, M. M., Bugge, T. H., Hirschmann, P., and Stunnenberg, H. G. (1991). Functional characterization of a natural retinoic acid responsive element. *Embo J* 10, 3829-3838.
- Voskoboinik, I., Drew, R., and Ahokas, J. T. (1996). Differential effect of peroxisome proliferators on rat glutathione S-transferase isoenzymes. *Toxicol Lett* 87, 147-155.
- Vu-Dac, N., Chopin-Delannoy, S., Gervois, P., Bonnelye, E., Martin, G., Fruchart, J. C., Laudet, V., and Staels, B. (1998). The nuclear receptors peroxisome proliferator-activated receptor alpha and Rev-erbalpha mediate the species-specific regulation of apolipoprotein A-I expression by fibrates. *J Biol Chem* 273, 25713-25720.
- Vu-Dac, N., Gervois, P., Jakel, H., Nowak, M., Bauge, E., Dehondt, H., Staels, B., Pennacchio, L. A., Rubin, E. M., Fruchart-Najib, J., and Fruchart, J. C. (2003). Apolipoprotein A5, a crucial determinant of plasma triglyceride levels, is highly responsive to peroxisome proliferator-activated receptor alpha activators. *J Biol Chem* 278, 17982-17985.
- Vu-Dac, N., Schoonjans, K., Kosykh, V., Dallongeville, J., Fruchart, J. C., Staels, B., and Auwerx, J. (1995). Fibrates increase human apolipoprotein A-II expression through activation of the peroxisome proliferator-activated receptor. *J Clin Invest* 96, 741-750.
- Vu-Dac, N., Schoonjans, K., Kosykh, V., Dallongeville, J., Heyman, R. A., Staels, B., and Auwerx, J. (1996). Retinoids increase human apolipoprotein A-11 expression through activation of the retinoid X receptor but not the retinoic acid receptor. *Mol Cell Biol* 16, 3350-3360.
- Wagner, R. L., Apriletti, J. W., McGrath, M. E., West, B. L., Baxter, J. D., and Fletterick, R. J. (1995). A structural role for hormone in the thyroid hormone receptor. *Nature* 378, 690-697.
- Wakeling, A. E., Dukes, M., and Bowler, J. (1991). A potent specific pure antiestrogen with clinical potential. *Cancer Res* 51, 3867-3873.
- Walczak, R., and Tontonoz, P. (2002). PPARadigms and PPARadoxes: expanding roles for PPARgamma in the control of lipid metabolism. *J Lipid Res* 43, 177-186.
- Wanders, R. J., Vreken, P., Ferdinandusse, S., Jansen, G. A., Waterham, H. R., van Roermund, C. W., and Van Grunsven, E. G. (2001). Peroxisomal fatty acid alpha- and beta-oxidation in humans: enzymology, peroxisomal metabolite transporters and peroxisomal diseases. *Biochem Soc Trans* 29, 250-267.
- Wang, H., Chen, J., Hollister, K., Sowers, L. C., and Forman, B. M. (1999). Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. *Mol Cell* 3, 543-553.
- Wang, L., Lee, Y. K., Bundman, D., Han, Y., Thevananther, S., Kim, C. S., Chua, S. S., Wei, P., Heyman, R. A., Karin, M., and Moore, D. D. (2002). Redundant pathways for negative feedback regulation of bile acid production. *Dev Cell* 2, 721-731.
- Wang, L., Zuercher, W. J., Consler, T. G., Lambert, M. H., Miller, A. B., Orband-Miller, L. A., McKee, D. D., Willson, T. M., and Nolte, R. T. (2006a). X-ray crystal structures of the estrogen-related receptor-gamma ligand binding domain in three functional states reveal the molecular basis of small molecule regulation. *J Biol Chem* 281, 37773-37781.
- Wang, N., Lan, D., Chen, W., Matsuura, F., and Tall, A. R. (2004). ATP-binding cassette transporters G1 and G4 mediate cellular cholesterol efflux to high-density lipoproteins. *Proc Natl Acad Sci U S A* 101, 9774-9779.
- Wang, W., Zhang, C., Marimuthu, A., Krupka, H. I., Tabrizizad, M., Shelloe, R., Mehra, U., Eng, K., Nguyen, H., Settachatgul, C., et al. (2005a). The crystal structures of human steroidogenic factor-1 and liver receptor homologue-1. *Proc Natl Acad Sci U S A* 102, 7505-7510.
- Wang, Y., Botolin, D., Christian, B., Busik, J., Xu, J., and Jump, D. B. (2005b). Tissue-specific, nutritional, and developmental regulation of rat fatty acid elongases. *J Lipid Res* 46, 706-715.
- Wang, Y., Botolin, D., Xu, J., Christian, B., Mitchell, E., Jayaprakasam, B., Nair, M. G., Peters, J. M., Busik, J. V., Olson, L. K., and Jump, D. B. (2006b). Regulation of hepatic fatty acid elongase and desaturase expression in diabetes and obesity. *J Lipid Res* 47, 2028-2041.
- Wang, Y., Chirgadze, N. Y., Briggs, S. L., Khan, S., Jensen, E. V., and Burris, T. P. (2006c). A second binding site for hydroxytamoxifen within the coactivator-binding groove of estrogen receptor beta. *Proc Natl Acad Sci U S A* 103, 9908-9911.

- Wang, Z., Benoit, G., Liu, J., Prasad, S., Aarnisalo, P., Liu, X., Xu, H., Walker, N. P., and Perlmann, T. (2003). Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. *Nature* **423**, 555-560.
- Warnmark, A., Treuter, E., Gustafsson, J. A., Hubbard, R. E., Brzozowski, A. M., and Pike, A. C. (2002). Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor alpha. *J Biol Chem* **277**, 21862-21868.
- Warnmark, A., Treuter, E., Wright, A. P., and Gustafsson, J. A. (2003). Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. *Mol Endocrinol* **17**, 1901-1909.
- Warnmark, A., Wikstrom, A., Wright, A. P., Gustafsson, J. A., and Hard, T. (2001). The N-terminal regions of estrogen receptor alpha and beta are unstructured in vitro and show different TBP binding properties. *J Biol Chem* **276**, 45939-45944.
- Watanabe, K., Fujii, H., Takahashi, T., Kodama, M., Aizawa, Y., Ohta, Y., Ono, T., Hasegawa, G., Naito, M., Nakajima, T., et al. (2000). Constitutive regulation of cardiac fatty acid metabolism through peroxisome proliferator-activated receptor alpha associated with age-dependent cardiac toxicity. *J Biol Chem* **275**, 22293-22299.
- Watkins, R. E., Davis-Searles, P. R., Lambert, M. H., and Redinbo, M. R. (2003a). Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. *J Mol Biol* **331**, 815-828.
- Watkins, R. E., Maglich, J. M., Moore, L. B., Wisely, G. B., Noble, S. M., Davis-Searles, P. R., Lambert, M. H., Kliewer, S. A., and Redinbo, M. R. (2003b). 2.1 A crystal structure of human PXR in complex with the St. John's wort compound hyperforin. *Biochemistry* **42**, 1430-1438.
- Watkins, R. E., Wisely, G. B., Moore, L. B., Collins, J. L., Lambert, M. H., Williams, S. P., Willson, T. M., Kliewer, S. A., and Redinbo, M. R. (2001). The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* **292**, 2329-2333.
- Watts, G. F., Ji, J., Chan, D. C., Ooi, E. M., Johnson, A. G., Rye, K. A., and Barrett, P. H. (2006). Relationships between changes in plasma lipid transfer proteins and apolipoprotein B-100 kinetics during fenofibrate treatment in the metabolic syndrome. *Clin Sci (Lond)* **111**, 193-199.
- Way, J. M., Harrington, W. W., Brown, K. K., Gottschalk, W. K., Sundseth, S. S., Mansfield, T. A., Ramachandran, R. K., Willson, T. M., and Kliewer, S. A. (2001). Comprehensive messenger ribonucleic acid profiling reveals that peroxisome proliferator-activated receptor gamma activation has coordinate effects on gene expression in multiple insulin-sensitive tissues. *Endocrinology* **142**, 1269-1277.
- Weber, C., and Dumont, E. (1997). Pharmacokinetics and pharmacodynamics of 9-cis-retinoic acid in healthy men. *J Clin Pharmacol* **37**, 566-574.
- Wei, P., Zhang, J., Egan-Hafley, M., Liang, S., and Moore, D. D. (2000). The nuclear receptor CAR mediates specific xenobiotic induction of drug metabolism. *Nature* **407**, 920-923.
- Weinberger, C., Thompson, C. C., Ong, E. S., Lebo, R., Gruol, D. J., and Evans, R. M. (1986). The c-erb-A gene encodes a thyroid hormone receptor. *Nature* **324**, 641-646.
- Weisberg, S. P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R. L., and Ferrante, A. W., Jr. (2003). Obesity is associated with macrophage accumulation in adipose tissue. *J Clin Invest* **112**, 1796-1808.
- Welsh, J., Chada, K., Dalal, S. S., Cheng, R., Ralph, D., and McClelland, M. (1992). Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* **20**, 4965-4970.
- Westin, M. A., Alexson, S. E., and Hunt, M. C. (2004). Molecular cloning and characterization of two mouse peroxisome proliferator-activated receptor alpha (PPAR $\alpha$ )-regulated peroxisomal acyl-CoA thioesterases. *J Biol Chem* **279**, 21841-21848.
- Westin, M. A., Hunt, M. C., and Alexson, S. E. (2007). Peroxisomes contain a specific phytanoyl-CoA/pristanoyl-CoA thioesterase acting as a novel auxiliary enzyme in alpha- and beta-oxidation of methyl-branched fatty acids in mouse. *J Biol Chem*.
- Whitby, R. J., Dixon, S., Maloney, P. R., Delerive, P., Goodwin, B. J., Parks, D. J., and Willson, T. M. (2006). Identification of small molecule agonists of the orphan nuclear receptors liver receptor homolog-1 and steroidogenic factor-1. *J Med Chem* **49**, 6652-6655.
- White, J. A., Ramshaw, H., Taimi, M., Stangle, W., Zhang, A., Everingham, S., Creighton, S., Tam, S. P., Jones, G., and Petkovich, M. (2000). Identification of the human cytochrome P450, P450RAI-2, which is predominantly expressed in the adult cerebellum and is responsible for all-trans-retinoic acid metabolism. *Proc Natl Acad Sci U S A* **97**, 6403-6408.
- Wierzbicki, A. S. (2006). Fibrates after the FIELD study: Some answers, more questions. *Diab Vasc Dis Res* **3**, 166-171.
- Wijayaratne, A. L., Nagel, S. C., Paige, L. A., Christensen, D. J., Norris, J. D., Fowlkes, D. M., and McDonnell, D. P. (1999). Comparative analyses of mechanistic differences among antiestrogens. *Endocrinology* **140**, 5828-5840.

- Williams, S., Bledsoe, R. K., Collins, J. L., Boggs, S., Lambert, M. H., Miller, A. B., Moore, J., McKee, D. D., Moore, L., Nichols, J., *et al.* (2003). X-ray crystal structure of the liver X receptor beta ligand binding domain: regulation by a histidine-tryptophan switch. *J Biol Chem* 278, 27138-27143.
- Williams, S. P., and Sigler, P. B. (1998). Atomic structure of progesterone complexed with its receptor. *Nature* 393, 392-396.
- Wilson, C. M., and McPhaul, M. J. (1994). A and B forms of the androgen receptor are present in human genital skin fibroblasts. *Proc Natl Acad Sci U S A* 91, 1234-1238.
- Wilson, T. E., Fahrner, T. J., and Milbrandt, J. (1993). The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. *Mol Cell Biol* 13, 5794-5804.
- Wilson, V. S., Lambright, C., Ostby, J., and Gray, L. E., Jr. (2002). In vitro and in vivo effects of 17beta-trenbolone: a feedlot effluent contaminant. *Toxicol Sci* 70, 202-211.
- Winrow, C. J., Marcus, S. L., Miyata, K. S., Zhang, B., Capone, J. P., and Rachubinski, R. A. (1994). Transactivation of the peroxisome proliferator-activated receptor is differentially modulated by hepatocyte nuclear factor-4. *Gene Expr* 4, 53-62.
- Wisely, G. B., Miller, A. B., Davis, R. G., Thornquest, A. D., Jr., Johnson, R., Spitzer, T., Sefler, A., Shearer, B., Moore, J. T., Miller, A. B., *et al.* (2002). Hepatocyte nuclear factor 4 is a transcription factor that constitutively binds fatty acids. *Structure* 10, 1225-1234.
- Wolfrum, C., Borrmann, C. M., Borchers, T., and Spener, F. (2001). Fatty acids and hypolipidemic drugs regulate peroxisome proliferator-activated receptors alpha - and gamma-mediated gene expression via liver fatty acid binding protein: a signaling path to the nucleus. *Proc Natl Acad Sci U S A* 98, 2323-2328.
- Wong, J. S., Ye, X., Muhlenkamp, C. R., and Gill, S. S. (2002). Effect of a peroxisome proliferator on 3 beta-hydroxysteroid dehydrogenase. *Biochem Biophys Res Commun* 293, 549-553.
- Wu, C., Zhao, H., Baggerly, K., Carta, R., and Zhang, L. (2007). Short oligonucleotide probes containing G-stacks display abnormal binding affinity on Affymetrix microarrays. *Bioinformatics*.
- Wu, J., Xia, C., Meier, J., Li, S., Hu, X., and Lala, D. S. (2002). The hypolipidemic natural product guggulsterone acts as an antagonist of the bile acid receptor. *Mol Endocrinol* 16, 1590-1597.
- Wu, P., Peters, J. M., and Harris, R. A. (2001). Adaptive increase in pyruvate dehydrogenase kinase 4 during starvation is mediated by peroxisome proliferator-activated receptor alpha. *Biochem Biophys Res Commun* 287, 391-396.
- Wu, R. C., Smith, C. L., and O'Malley, B. W. (2005a). Transcriptional regulation by steroid receptor coactivator phosphorylation. *Endocr Rev* 26, 393-399.
- Wu, Y. L., Yang, X., Ren, Z., McDonnell, D. P., Norris, J. D., Willson, T. M., and Greene, G. L. (2005b). Structural basis for an unexpected mode of SERM-mediated ER antagonism. *Mol Cell* 18, 413-424.
- Wu, Z., Rosen, E. D., Brun, R., Hauser, S., Adelman, G., Troy, A. E., McKeon, C., Darlington, G. J., and Spiegelman, B. M. (1999). Cross-regulation of C/EBP alpha and PPAR gamma controls the transcriptional pathway of adipogenesis and insulin sensitivity. *Mol Cell* 3, 151-158.
- Wu, Z., Xie, Y., Bucher, N. L., and Farmer, S. R. (1995). Conditional ectopic expression of C/EBP beta in NIH-3T3 cells induces PPAR gamma and stimulates adipogenesis. *Genes Dev* 9, 2350-2363.
- Wurtz, J. M., Bourguet, W., Renaud, J. P., Vivat, V., Chambon, P., Moras, D., and Gronemeyer, H. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. *Nat Struct Biol* 3, 87-94.
- Xu, B., and Koenig, R. J. (2004). An RNA-binding domain in the thyroid hormone receptor enhances transcriptional activation. *J Biol Chem* 279, 33051-33056.
- Xu, H., Barnes, G. T., Yang, Q., Tan, G., Yang, D., Chou, C. J., Sole, J., Nichols, A., Ross, J. S., Tartaglia, L. A., and Chen, H. (2003). Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J Clin Invest* 112, 1821-1830.
- Xu, H. E., Lambert, M. H., Montana, V. G., Parks, D. J., Blanchard, S. G., Brown, P. J., Sternbach, D. D., Lehmann, J. M., Wisely, G. B., Willson, T. M., *et al.* (1999). Molecular recognition of fatty acids by peroxisome proliferator-activated receptors. *Mol Cell* 3, 397-403.
- Xu, H. E., Lambert, M. H., Montana, V. G., Plunket, K. D., Moore, L. B., Collins, J. L., Oplinger, J. A., Kliewer, S. A., Gampe, R. T., Jr., McKee, D. D., *et al.* (2001a). Structural determinants of ligand binding selectivity between the peroxisome proliferator-activated receptors. *Proc Natl Acad Sci U S A* 98, 13919-13924.
- Xu, H. E., Stanley, T. B., Montana, V. G., Lambert, M. H., Shearer, B. G., Cobb, J. E., McKee, D. D., Galardi, C. M., Plunket, K. D., Nolte, R. T., *et al.* (2002a). Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. *Nature* 415, 813-817.
- Xu, J., and Li, Q. (2003). Review of the in vivo functions of the p160 steroid receptor coactivator family. *Mol Endocrinol* 17, 1681-1692.
- Xu, J., and O'Malley, B. W. (2002). Molecular mechanisms and cellular biology of the steroid receptor coactivator (SRC) family in steroid receptor function. *Rev Endocr Metab Disord* 3, 185-192.

- Xu, J., Xiao, G., Trujillo, C., Chang, V., Blanco, L., Joseph, S. B., Bassilian, S., Saad, M. F., Tontonoz, P., Lee, W. N., and Kurland, I. J. (2002b). Peroxisome proliferator-activated receptor alpha (PPARalpha) influences substrate utilization for hepatic glucose production. *J Biol Chem* 277, 50237-50244.
- Xu, R. X., Lambert, M. H., Wisely, B. B., Warren, E. N., Weinert, E. E., Waitt, G. M., Williams, J. D., Collins, J. L., Moore, L. B., Willson, T. M., and Moore, J. T. (2004). A structural basis for constitutive activity in the human CAR/RXRalpha heterodimer. *Mol Cell* 16, 919-928.
- Xu, S., Zhu, B. T., Turan, V., Rusyn, I., Thurman, R., Peters, J. M., Gonzalez, F. J., and Conney, A. H. (2001b). PPARalpha-dependent induction of liver microsomal esterification of estradiol and testosterone by a prototypical peroxisome proliferator. *Endocrinology* 142, 3554-3557.
- Yamada, J., Sugiyama, H., Watanabe, T., and Suga, T. (1995). Suppressive effect of growth hormone on the expression of peroxisome proliferator-activated receptor in cultured rat hepatocytes. *Res Commun Mol Pathol Pharmacol* 90, 173-176.
- Yamauchi, T., Oike, Y., Kamon, J., Waki, H., Komeda, K., Tsuchida, A., Date, Y., Li, M. X., Miki, H., Akanuma, Y., et al. (2002). Increased insulin sensitivity despite lipodystrophy in Crebbp heterozygous mice. *Nat Genet* 30, 221-226.
- Yan, M., Rayapuram, N., and Subramani, S. (2005). The control of peroxisome number and size during division and proliferation. *Curr Opin Cell Biol* 17, 376-383.
- Yan, Z. H., Karam, W. G., Staudinger, J. L., Medvedev, A., Ghanayem, B. I., and Jetten, A. M. (1998). Regulation of peroxisome proliferator-activated receptor alpha-induced transactivation by the nuclear orphan receptor TAK1/TR4. *J Biol Chem* 273, 10948-10957.
- Yang, C., McDonald, J. G., Patel, A., Zhang, Y., Umetani, M., Xu, F., Westover, E. J., Covey, D. F., Mangelsdorf, D. J., Cohen, J. C., and Hobbs, H. H. (2006a). Sterol intermediates from cholesterol biosynthetic pathway as liver X receptor ligands. *J Biol Chem* 281, 27816-27826.
- Yang, S. Y., He, X. Y., and Schulz, H. (2005). 3-Hydroxyacyl-CoA dehydrogenase and short chain 3-hydroxyacyl-CoA dehydrogenase in human health and disease. *Febs J* 272, 4874-4883.
- Yang, W., and Freedman, L. P. (1999). 20-Epi analogues of 1,25-dihydroxyvitamin D3 are highly potent inducers of DRIP coactivator complex binding to the vitamin D3 receptor. *J Biol Chem* 274, 16838-16845.
- Yang, X., Downes, M., Yu, R. T., Bookout, A. L., He, W., Straume, M., Mangelsdorf, D. J., and Evans, R. M. (2006b). Nuclear receptor expression links the circadian clock to metabolism. *Cell* 126, 801-810.
- Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30, e15.
- Yasmin, R., Williams, R. M., Xu, M., and Noy, N. (2005). Nuclear import of the retinoid X receptor, the vitamin D receptor, and their mutual heterodimer. *J Biol Chem* 280, 40152-40160.
- Ye, J. M., Doyle, P. J., Iglesias, M. A., Watson, D. G., Cooney, G. J., and Kraegen, E. W. (2001). Peroxisome proliferator-activated receptor (PPAR)-alpha activation lowers muscle lipids and improves insulin sensitivity in high fat-fed rats: comparison with PPAR-gamma activation. *Diabetes* 50, 411-417.
- Ye, L., Li, Y. L., Mellstrom, K., Mellin, C., Bladh, L. G., Koehler, K., Garg, N., Garcia Collazo, A. M., Litten, C., Husman, B., et al. (2003). Thyroid receptor ligands. 1. Agonist ligands selective for the thyroid receptor beta1. *J Med Chem* 46, 1580-1588.
- Yeldandi, A. V., Rao, M. S., and Reddy, J. K. (2000). Hydrogen peroxide generation in peroxisome proliferator-induced oncogenesis. *Mutat Res* 448, 159-177.
- Yoneda, Y., Han, D., Ogita, K., and Watanabe, A. (1995). Distinction between binding of [<sup>3</sup>H]triamcinolone acetonide to a ligand binding domain on the glucocorticoid receptor complex in cytosol fractions of brain and liver from the rat with intact adrenals. *Brain Res* 685, 105-116.
- Yoon, J. C., Puigserver, P., Chen, G., Donovan, J., Wu, Z., Rhee, J., Adelman, G., Stafford, J., Kahn, C. R., Granner, D. K., et al. (2001). Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* 413, 131-138.
- Yoon, M. J., Lee, G. Y., Chung, J. J., Ahn, Y. H., Hong, S. H., and Kim, J. B. (2006). Adiponectin increases fatty acid oxidation in skeletal muscle cells by sequential activation of AMP-activated protein kinase, p38 mitogen-activated protein kinase, and peroxisome proliferator-activated receptor alpha. *Diabetes* 55, 2562-2570.
- Yoshikawa, T., Shimano, H., Amemiya-Kudo, M., Yahagi, N., Hasty, A. H., Matsuzaka, T., Okazaki, H., Tamura, Y., Iizuka, Y., Ohashi, K., et al. (2001). Identification of liver X receptor-retinoid X receptor as an activator of the sterol regulatory element-binding protein 1c gene promoter. *Mol Cell Biol* 21, 2991-3000.
- Yoshikawa, T., Shimano, H., Yahagi, N., Ide, T., Amemiya-Kudo, M., Matsuzaka, T., Nakakuki, M., Tomita, S., Okazaki, H., Tamura, Y., et al. (2002). Polyunsaturated fatty acids suppress sterol regulatory element-binding protein 1c promoter activity by inhibition of liver X receptor (LXR) binding to LXR response elements. *J Biol Chem* 277, 1705-1711.

- Yoshinari, K., Kobayashi, K., Moore, R., Kawamoto, T., and Negishi, M. (2003). Identification of the nuclear receptor CAR:HSP90 complex in mouse liver and recruitment of protein phosphatase 2A in response to phenobarbital. *FEBS Lett* 548, 17-20.
- Yoshizawa, T., Handa, Y., Uematsu, Y., Takeda, S., Sekine, K., Yoshihara, Y., Kawakami, T., Arioka, K., Sato, H., Uchiyama, Y., *et al.* (1997). Mice lacking the vitamin D receptor exhibit impaired bone formation, uterine hypoplasia and growth retardation after weaning. *Nat Genet* 16, 391-396.
- Young, M. E., Goodwin, G. W., Ying, J., Guthrie, P., Wilson, C. R., Laws, F. A., and Taegtmeyer, H. (2001a). Regulation of cardiac and skeletal muscle malonyl-CoA decarboxylase by fatty acids. *Am J Physiol Endocrinol Metab* 280, E471-479.
- Young, M. E., Patil, S., Ying, J., Depre, C., Ahuja, H. S., Shipley, G. L., Stepkowski, S. M., Davies, P. J., and Taegtmeyer, H. (2001b). Uncoupling protein 3 transcription is regulated by peroxisome proliferator-activated receptor (alpha) in the adult rodent heart. *Faseb J* 15, 833-845.
- Yu, C., Markan, K., Temple, K. A., Deplewski, D., Brady, M. J., and Cohen, R. N. (2005). The nuclear receptor corepressors NCoR and SMRT decrease peroxisome proliferator-activated receptor gamma transcriptional activity and repress 3T3-L1 adipogenesis. *J Biol Chem* 280, 13600-13605.
- Yu, S., Cao, W. Q., Kashireddy, P., Meyer, K., Jia, Y., Hughes, D. E., Tan, Y., Feng, J., Yeldandi, A. V., Rao, M. S., *et al.* (2001). Human peroxisome proliferator-activated receptor alpha (PPAR $\alpha$ ) supports the induction of peroxisome proliferation in PPAR $\alpha$ -deficient mouse liver. *J Biol Chem* 276, 42485-42491.
- Yu, S., and Reddy, J. K. (2007). Transcription coactivators for peroxisome proliferator-activated receptors. *Biochim Biophys Acta*.
- Zambon, A., Gervois, P., Pauletto, P., Fruchart, J. C., and Staels, B. (2006). Modulation of hepatic inflammatory risk markers of cardiovascular diseases by PPAR-alpha activators: clinical and experimental evidence. *Arterioscler Thromb Vasc Biol* 26, 977-986.
- Zamir, I., Zhang, J., and Lazar, M. A. (1997). Stoichiometric and steric principles governing repression by nuclear hormone receptors. *Genes Dev* 11, 835-846.
- Zandbergen, F., and Plutzky, J. (2007). PPAR $\alpha$  in atherosclerosis and inflammation. *Biochim Biophys Acta*.
- Zechel, C., Shen, X. Q., Chen, J. Y., Chen, Z. P., Chambon, P., and Gronemeyer, H. (1994). The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats. *Embo J* 13, 1425-1433.
- Zelent, A., Mendelsohn, C., Kastner, P., Krust, A., Garnier, J. M., Ruffenach, F., Leroy, P., and Chambon, P. (1991). Differentially expressed isoforms of the mouse retinoic acid receptor beta generated by usage of two promoters and alternative splicing. *Embo J* 10, 71-81.
- Zelko, I., Sueyoshi, T., Kawamoto, T., Moore, R., and Negishi, M. (2001). The peptide near the C terminus regulates receptor CAR nuclear translocation induced by xenochemicals in mouse liver. *Mol Cell Biol* 21, 2838-2846.
- Zhang, B., Berger, J., Hu, E., Szalkowski, D., White-Carrington, S., Spiegelman, B. M., and Moller, D. E. (1996a). Negative regulation of peroxisome proliferator-activated receptor-gamma gene expression contributes to the antiadipogenic effects of tumor necrosis factor-alpha. *Mol Endocrinol* 10, 1457-1466.
- Zhang, B., Marcus, S. L., Miyata, K. S., Subramani, S., Capone, J. P., and Rachubinski, R. A. (1993). Characterization of protein-DNA interactions within the peroxisome proliferator-responsive element of the rat hydratase-dehydrogenase gene. *J Biol Chem* 268, 12939-12945.
- Zhang, B., Marcus, S. L., Sajjadi, F. G., Alvares, K., Reddy, J. K., Subramani, S., Rachubinski, R. A., and Capone, J. P. (1992). Identification of a peroxisome proliferator-responsive element upstream of the gene encoding rat peroxisomal enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase. *Proc Natl Acad Sci U S A* 89, 7541-7545.
- Zhang, C., Hazarika, P., Ni, X., Weidner, D. A., and Duvic, M. (2002a). Induction of apoptosis by bexarotene in cutaneous T-cell lymphoma cells: relevance to mechanism of therapeutic action. *Clin Cancer Res* 8, 1234-1240.
- Zhang, L., Hurek, T., and Reinhold-Hurek, B. (2005a). Position of the fluorescent label is a crucial factor determining signal intensity in microarray hybridizations. *Nucleic Acids Res* 33, e166.
- Zhang, L., Miles, M. F., and Aldape, K. D. (2003a). A model of molecular interactions on short oligonucleotide microarrays. *Nat Biotechnol* 21, 818-821.
- Zhang, S., Liang, X., and Danielsen, M. (1996b). Role of the C terminus of the glucocorticoid receptor in hormone binding and agonist/antagonist discrimination. *Mol Endocrinol* 10, 24-34.
- Zhang, W., Morris, Q. D., Chang, R., Shai, O., Bakowski, M. A., Mitsakakis, N., Mohammad, N., Robinson, M. D., Zirngibl, R., Somogyi, E., *et al.* (2004). The functional landscape of mouse gene expression. *J Biol* 3, 21.
- Zhang, Y., Kast-Woelbern, H. R., and Edwards, P. A. (2003b). Natural structural variants of the nuclear receptor farnesoid X receptor affect transcriptional activation. *J Biol Chem* 278, 104-110.

- Zhang, Y., Matheny, M., Zolotukhin, S., Tumer, N., and Scarpase, P. J. (2002b). Regulation of adiponectin and leptin gene expression in white and brown adipose tissues: influence of beta3-adrenergic agonists, retinoic acid, leptin and fasting. *Biochim Biophys Acta* **1584**, 115-122.
- Zhang, Y., Repa, J. J., Gauthier, K., and Mangelsdorf, D. J. (2001). Regulation of lipoprotein lipase by the oxysterol receptors, LX $\alpha$  and LX $\beta$ . *J Biol Chem* **276**, 43018-43024.
- Zhang, Z., Olland, A. M., Zhu, Y., Cohen, J., Berrodin, T., Chippari, S., Appavu, C., Li, S., Wilhem, J., Chopra, R., et al. (2005b). Molecular and pharmacological properties of a potent and selective novel nonsteroidal progesterone receptor agonist tanaproget. *J Biol Chem* **280**, 28468-28475.
- Zhao, N., Hashida, H., Takahashi, N., Misumi, Y., and Sakaki, Y. (1995). High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. *Gene* **156**, 207-213.
- Zhao, Q., Chasse, S. A., Devarakonda, S., Sierk, M. L., Ahvazi, B., and Rastinejad, F. (2000). Structural basis of RXR-DNA interactions. *J Mol Biol* **296**, 509-520.
- Zhao, Q., Khorasanizadeh, S., Miyoshi, Y., Lazar, M. A., and Rastinejad, F. (1998). Structural elements of an orphan nuclear receptor-DNA complex. *Mol Cell* **1**, 849-861.
- Zhao, X., Patton, J. R., Davis, S. L., Florence, B., Ames, S. J., and Spanjaard, R. A. (2004). Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol Cell* **15**, 549-558.
- Zhao, X., Patton, J. R., Ghosh, S. K., Fischel-Ghodsian, N., Shen, L., and Spanjaard, R. A. (2007). Pus3p- and Pus1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling. *Mol Endocrinol* **21**, 686-699.
- Zhen, Y., Krausz, K. W., Chen, C., Idle, J. R., and Gonzalez, F. J. (2007). Metabolomic and Genetic Analysis of Biomarkers for PPAR{alpha} Expression and Activation. *Mol Endocrinol*.
- Zhou, G., Cummings, R., Li, Y., Mitra, S., Wilkinson, H. A., Elbrecht, A., Hermes, J. D., Schaeffer, J. M., Smith, R. G., and Moller, D. E. (1998a). Nuclear receptors have distinct affinities for coactivators: characterization by fluorescence resonance energy transfer. *Mol Endocrinol* **12**, 1594-1604.
- Zhou, Y. C., and Waxman, D. J. (1999). STAT5b down-regulates peroxisome proliferator-activated receptor alpha transcription by inhibition of ligand-independent activation function region-1 trans-activation domain. *J Biol Chem* **274**, 29874-29882.
- Zhou, Y. T., Shimabukuro, M., Wang, M. Y., Lee, Y., Higa, M., Milburn, J. L., Newgard, C. B., and Unger, R. H. (1998b). Role of peroxisome proliferator-activated receptor alpha in disease of pancreatic beta cells. *Proc Natl Acad Sci U S A* **95**, 8898-8903.
- Zhu, Q., Anderson, G. W., Mucha, G. T., Parks, E. J., Metkowski, J. K., and Mariash, C. N. (2005). The Spot 14 protein is required for de novo lipid synthesis in the lactating mammary gland. *Endocrinology* **146**, 3343-3350.
- Zhu, Q., Mariash, A., Margosian, M. R., Gopinath, S., Fareed, M. T., Anderson, G. W., and Mariash, C. N. (2001a). Spot 14 gene deletion increases hepatic de novo lipogenesis. *Endocrinology* **142**, 4363-4370.
- Zhu, X. G., Hanover, J. A., Hager, G. L., and Cheng, S. Y. (1998). Hormone-induced translocation of thyroid hormone receptors in living cells visualized using a receptor green fluorescent protein chimera. *J Biol Chem* **273**, 27058-27063.
- Zhu, Y., Alvares, K., Huang, Q., Rao, M. S., and Reddy, J. K. (1993). Cloning of a new member of the peroxisome proliferator-activated receptor gene family from mouse liver. *J Biol Chem* **268**, 26817-26820.
- Zhu, Y., Kan, L., Qi, C., Kanwar, Y. S., Yeldandi, A. V., Rao, M. S., and Reddy, J. K. (2000a). Isolation and characterization of peroxisome proliferator-activated receptor (PPAR) interacting protein (PRIP) as a coactivator for PPAR. *J Biol Chem* **275**, 13510-13516.
- Zhu, Y., Qi, C., Cao, W. Q., Yeldandi, A. V., Rao, M. S., and Reddy, J. K. (2001b). Cloning and characterization of PIMT, a protein with a methyltransferase domain, which interacts with and enhances nuclear receptor coactivator PRIP function. *Proc Natl Acad Sci U S A* **98**, 10380-10385.
- Zhu, Y., Qi, C., Jia, Y., Nye, J. S., Rao, M. S., and Reddy, J. K. (2000b). Deletion of PBP/PPARBP, the gene for nuclear receptor coactivator peroxisome proliferator-activated receptor-binding protein, results in embryonic lethality. *J Biol Chem* **275**, 14779-14782.
- Zhu, Y., Qi, C., Korenberg, J. R., Chen, X. N., Noya, D., Rao, M. S., and Reddy, J. K. (1995). Structural organization of mouse peroxisome proliferator-activated receptor gamma (mPPAR gamma) gene: alternative promoter use and different splicing yield two mPPAR gamma isoforms. *Proc Natl Acad Sci U S A* **92**, 7921-7925.
- Zhu, Y. J., Crawford, S. E., Stellmach, V., Dwivedi, R. S., Rao, M. S., Gonzalez, F. J., Qi, C., and Reddy, J. K. (2003). Coactivator PRIP, the peroxisome proliferator-activated receptor-interacting protein, is a modulator of placental, cardiac, hepatic, and embryonic development. *J Biol Chem* **278**, 1986-1990.
- Zhu, Z., Kim, S., Chen, T., Lin, J. H., Bell, A., Bryson, J., Dubaque, Y., Yan, N., Yanchunas, J., Xie, D., et al. (2004). Correlation of high-throughput pregnane X receptor (PXR) transactivation and binding assays. *J Biomol Screen* **9**, 533-540.

Ziouzenkova, O., Orasanu, G., Sharlach, M., Akiyama, T. E., Berger, J. P., Viereck, J., Hamilton, J. A., Tang, G., Dolnikowski, G. G., Vogel, S., *et al.* (2007). Retinaldehyde represses adipogenesis and diet-induced obesity. *Nat Med* 13, 695-702.

Zoete, V., Grosdidier, A., and Michelin, O. (2007). Peroxisome proliferator-activated receptor structures: Ligand specificity, molecular switch and interactions with regulators. *Biochim Biophys Acta*.

## **ANNEXE 1 : Protocole d'utilisation des puces à ADN INRArray 01**

Ce protocole figure tel quel dans la base de données qualité du laboratoire de Pharmacologie et Toxicologie (UR66) de l'INRA de Toulouse. Il est reproduit ici avec l'autorisation de Thierry Pineau, directeur du laboratoire.



**INRA  
LABORATOIRE DE PHARMACOLOGIE – TOXICOLOGIE  
UR66**

**Référence : MOLE/OPERAT/019/V04**

**Titre : Hybridation isotopique de macropuces à ADN avec des ARN reverse-transcrits.**

**Objet : rétro-transcrire des populations d'ARN en ADN complémentaires (ADNc) marqués au  $^{33}\text{P}$ , hybrider ces sondes isotopiques au macro-arrangement développé à l'UR66 (INRAarray®), recueillir et organiser les données expérimentales.**

**Mots clés : génomique, transcriptomique, ARN, transcription inverse, RT, marquage isotopique,  $^{33}\text{P}$ , biopuce, INRAarray.**

	Nom	Date	Signature
Rédacteur	Frédéric LASSEUR	24/11/2006	
Vérificateur	Alexandre Eveillard	24/11/2006	
Animateur Qualité	Thierry Hachet	13/12/2005	
Approbateur	Pierre Galtier	13/12/2005	

**Annule et remplace la version précédente**

## 1. Sommaire

Page de garde	Page	1
1. Sommaire	Page	2
2. Objet	Page	2
3. Domaine d'application	Page	2
4. Diffusion	Page	2
5. Matériels et consommables	Page	3
6. Prévention et sécurité	Page	3
7. Documents attachés	Page	3
8. Corps du mode opératoire	Page	4
8.1 Préparation au préalable	Page	4
8.2 Préparation le jour même	Page	4
8.3 Amorçage	Page	5
8.4 Marquage	Page	5
8.5 Pré hybridation	Page	5-6
8.6 Lyse alcaline	Page	6
8.7 Neutralisation	Page	6
8.8 Elimination des nucléotides non incorporés	Page	6
8.9 Hybridation	Page	7
8.10 comptage	Page	7
8.11 Lavages	Page	7
8.12 Contrôles	Page	7
8.13 déshybridation	Page	8
8.14 Première hybridation de membranes neuves	Page	8
8.14 Tampons et milieux	Page	8-9

## 2. Objet

Rétro-transcrire des populations d'ARN en ADN complémentaires (ADNc) marqués au  $^{33}\text{P}$ , hybrider ces sondes isotopiques au macro-arrangement développé à l'UR66 (INRAarray®), recueillir et organiser les données expérimentales.

## 3. Domaine d'application

Ce mode opératoire est réservé au personnel de l'unité habilité à manipuler des produits biologiques isotopiquement marqués et possédant un badge personnel de contrôle.

Chaque utilisateur doit avoir lu et doit respecter en totalité les consignes de la procédure d'utilisation de la zone surveillée de la pièce N° 27, dédiée à l'utilisation de radio-isotopes dont le  $^{32}\text{P}$  et  $^{33}\text{P}$ : [MOLE/UTILIS/001/V01](#).

La manipulation des échantillons contenant du  $^{33}\text{P}$  doit obligatoirement être effectuée dans la zone dédiée ou « box isotopes » (zone surveillée de la pièce 27) à l'exception de la transcription inverse, réalisée dans le thermocycleur 24 puits de l'équipe de pharmacologie moléculaire (geneamp PCR system 2400, pièce n°24) ou du thermocycleur 96 puits de l'équipe de mycotoxicologie (geneamp PCR system 2700, pièce n°2)

Tout pipetage supérieur à 50 $\mu\text{Ci}$  doit être effectué sous la hotte aspirante de la pièce 15-2.

## 4. Diffusion

Personnel de laboratoire

## 5. Matériels et consommables

### Matériel :

Thermocycleur Perkin Elmer Geneamp PCR system 2400 : AQM-INRA-31066-97  
 Thermocycleur Perkin Elmer Geneamp PCR system 2700 : AQM-INRA-31066-194  
 Flacons à hybridation PP Nalgene ref 2105.0002  
 Flacons à scintillation 2ml Sarstedt ref 72.690  
 Bain sec Lab Line multi-block AQM-INRA-31066-78  
 Agitateur Belly Dancer Stovall AQM-INRA-31066-75  
 Thermosoudeur ABgène AQM-INRA-31066-106  
 Compteur béta Perkin Elmer Microbète Trilux (pièce 15-2)  
 Centrifugeuse mini spin Eppendorf AQM-INRA-31066-104, avec réducteurs en plastique  
 Centrifugeuse sigma 4K15 AQM-INRA-31066-81  
 Four à hybridation Appligène AQM-INRA-31066-92  
 Liquide scintillant Ready Safe Beckman ref P/N158735  
 Storm screen KodaK Amersham ref 63.0034.81  
 Imageur au phosphore Storm 840 Molecular Dynamics AQM-INRA-31066-176  
 Compteur béta MCB21 Eurisys measures AQM-INRA-31066-84

### Produits :

Sperme de saumon soniqué (scheared DNA) 10mg/ml euromedex TA202SSP (20x2ml)  
 dCT<sup>33</sup>P stabilisé Perkin Elmer NEG613H  
 Transcriptase inverse : Superscript II Invitrogen 18064014  
 dNTP set 100mM invitrogen 10297018  
 Random primers 1,5mM (3µg/µl) invitrogen 48190011  
 RNase OUT RNase inhibitor invitrogen 10777019  
 dA60 1µMol sigma genosys  
 dT25 1µMol sigma genosys + purification cartouche reverse phase cartridge  
 NaCl Eurobio ref GAUNAC01-66  
 NaOH Fluka ref 71692  
 Citrate tri sodique, dihydraté (C<sub>6</sub>H<sub>5</sub>O<sub>7</sub>Na<sub>3</sub>2H<sub>2</sub>O) Euromedex ref 1126  
 Tris HCl Euromedex ref EU0011C  
 Solution de décontamination isotopique Decon90 Prolabo ref 23.280.296  
 Eau sans nuclease Invitrogen ref 10977015

### Consommables :

Pointes à filtre stériles Axygen (ATGC)  
 Tubes 0,5ml ; 1,5ml et 2ml autoclavés  
 Plaque 96 trous thermofast ABgène AB-1100  
 Mini colonnes de chromatographie (« gel filtration » = exclusion-diffusion) Probequant G50 Amersham  
 27-5335-01

## 6. Prévention et sécurité

Respecter les règles précisées dans :

Procédure d'utilisation du box 32P33P : [MOLE/UTILIS/001/V01](#).

## 7. Documents attachés

- Mode opératoire pour la préparation d'une solution de Denhardts 100X : [MOLE/OPERAT/030/V01](#)
- Mode opératoire pour la préparation d'un litre de solution de SDS à 20% : [MOLE/OPERAT/029/V01](#)
- Mode opératoire pour la préparation d'une solution d'EDTA à 500mM et pH=8,1 : [MOLE/OPERAT/028/V01](#)
- Procédure d'utilisation du box 32P33P : [MOLE/UTILIS/001/V01](#)

## 8.Mode opératoire

### 8.1. Préparation au préalable (par exemple la veille de l'hybridation isotopique) :

- Vérifier l'intégrité des ARN sur gel d'agarose (dépôt 1µl/ARN, gel 1,5% agarose) ou disposer du résultat de leur analyse sur Bioanalyseur – Agilent (appareil disponible sur réservation à la plate forme « Biopuces » du Génopôle Toulouse (INSA), cassette d'analyse disponible dans l'équipe PM de l'UR66). Ne retenir pour une hybridation isotopique sur les macroarray, que les échantillons témoignant d'une intégrité de population d'ARNs, et ne donnant aucun signe de dégradation. Tous les échantillons suspects de dégradation seront repréparés à partir d'échantillons de tissus initiaux.

- Effectuer les calculs pour disposer de solutions d'ARNs totaux à la concentration de 1,25µg/µl. Ne pas opérer ce jour les dilutions nécessaires. Cette opération doit être faite extemporanément, le jour de l'hybridation (voir paragraphe 8.3) afin d'épargner aux échantillons un cycle de congélation-décongélation supplémentaire.

- vérifier que les flacons à hybridation (polypropylène) sont identifiés, lavés et refermés aussitôt secs.

- prévoir des pointes à filtre pour toutes les étapes du protocole.

- préparation des membranes :

- ✓ Membranes déjà hybridées : s'assurer qu'elles ont été correctement déshybridées à la fin de la dernière utilisation (cf. § 8.13) puis conservées humides entre papier 3M et saran à 4°C.
- ✓ Membranes vierges : Faire une 1ere hybridation « à vide » (cf. § 8.14).

- Préparer, identifier et installer sur portoirs :

- ✓ Une série de tubes 0.2ml pour la PCR (si PCR en tubes) ou une plaque 96 puits (si PCR en plaque).
- ✓ Une série de tubes 1.5ml et une de tubes 2ml, pour lesquels il faut séparer le bouchon du tube. Pour cela, couper à mi longueur la languette reliant le tube et le bouchon. Ces tubes serviront pour l'étape de purification sur G50. Identifier les tubes 1,5ml sur le côté et conserver les bouchons.
- ✓ Une série de tubes 2ml (pour l'étape d'incubation du milieu d'hybridation + ADN complémentaire + oligonucléotides dA60).
- ✓ Une série de colonnes G50, à installer sur les tubes de 2ml sans bouchon. Numéro : sur le bouchon
- ✓ Les flacons à hybridation (polypropylène) identifiés pour le suivi individuel des membranes
- ✓ Les flacons à scintillation. Prévoir 1 flacon de plus pour mesurer la radioactivité totale ajoutée par tube de transcription inverse avant incorporation et 1 autre flacon pour la mesure du mélange réactionnel ou « mix » avant RT. Identifier les flacons sur le côté et distribuer 2ml de liquide scintillant par flacon. Fermer les flacons (scintillant = toxique).

- Opérer une vérification de la qualité du dCT<sup>33</sup>P par comptage d'un aliquot dilué au 1/274 dans de l'eau : la source est sortie du réfrigérateur / congélateur situé dans le « box isotopes » et 2 µL sont ajoutés dans 548 µL d'eau. Après homogénéisation, 2 µL de ce mélange aqueux sont ajoutés à 2 mL de liquide scintillant dans une fiole *ad hoc*, et le comptage est effectué (compteur Perkin Elmer, barrette 03). Cette vérification est opérée la veille des marquages pour détecter précocement un éventuel problème de synthèse du dCT<sup>33</sup>P, qui a déjà été observé.

Le résultat attendu d'un tel comptage se situe dans une fourchette de 140 000 à 180 000 ccpm1.

### 8.2. Préparation le jour de l'hybridation isotopique :

✓ Allumer bain sec à 95°C, agitateur Belly Dancer à 65°C, thermocycleur, thermo soudeur (si PCR en plaque).

✓ Décongeler : à 4°C : ARN, dT25, dA60, ADN de sperme de saumon, random primers 3µg/µl, dNTPmix, Spike mix\*

À température ambiante : 5X 1st strand buffer et 0.1M DTT

À 65°C : le milieu d'hybridation.

\* : Spike mix = mélange d'étalons internes = mélange d'ARNs étalons, produits par transcription *in vitro* initialement produits par Romain et Pascal, puis synthétisés en 2005 par Alexandre et Arnaud et conservés à -70°C, dans des tubes verts (50µl ou violet (100µl) contenant le mélange final ou « spike mix »).

*Remarque : l'idéal est de travailler à 2. Le matin, l'un peut s'occuper des dilutions des ARN + amorçage + marquage pendant que l'autre gère la pré-hybridation.*

### 8.3. Amorçage (1<sup>ère</sup> étape de la RT) :

Ajuster la concentration de tous les échantillons d'ARNs totaux à 1,25µg/µl. Opérer les dilutions nécessaires avec une eau sans nucléase. En règle générale, on opère dans l'équipe de pharmacologie moléculaire à partir d'un aliquot de 5 µL, isolé le jour de la préparation des ARNs, et spécifiquement dédié à l'hybridation sur macroarrays (afin de minimiser le nombre des congélations – décongélation).

Préparer un « mix » d'amorçage pour un nombre de réactions égal au nombre d'échantillons + 3.

Mix pour 1 tube :	Spike mix	1µL
	dT25 4µg/µL	0,5µL
	random primers 3µg/µL	0,5µL

Distribuer 2µL de mix par tube de PCR de 0,2ml,

Puis ajouter 4µL d'ARNs totaux spécifiques à 1,25µg/µl dans chaque tube.

Homogénéiser, centrifuger brièvement (pour toutes les étapes, la centrifugation des tubes de 0.2 mL nécessite l'usage de réducteurs de diamètre pour les trous du rotor).

Incuber au thermocycleur choisi, 10mn à 70°C puis 10mn à 42°C puis 5mn à 4°C.

Centrifuger pour reprendre la condensation présente sur les parois et réserver les tubes sur glace.

A ce stade les échantillons initiaux d'ARNs ainsi que le Spike mix peuvent être recongelés.

*Durant l'amorçage (annealing) les ARN perdent leur structure secondaire, les queues polyA permettent l'hybridation des oligos dT25, alors que la séquence propre de l'ARN autorise l'hybridation des primers aléatoires qui sont majoritairement des hexamères.*

### 8.4. Marquage (2<sup>ème</sup> étape de la RT) :

La préparation du « mix » de marquage pour l'ensemble des échantillons considérés nécessite la manipulation en un seul prélèvement d'une partie considérable de la source radioactive. Cette opération est réglementée. Elle sera obligatoirement opérée sous la hotte aspirante (en marche) de la paillasse de la zone surveillée de la pièce N° 15-2.

Préparer sous la hotte de la pièce 15.2 un « mix » de marquage pour un nombre de réactions égal au nombre d'échantillons + 3.

Mix pour 1 tube :	5X 1st strand buffer	4µL
	DTT 0.1M	2
	dNTP mix	2
	RNase OUT	1
	superscriptII	2
	dCT <sup>33</sup> P	4

Homogénéiser ce mélange.

Sous la hotte de la pièce 15.2, ajouter 15µL de ce mix à chacun des tubes issus de l'étape d'amorçage 8.3.

Homogénéiser.

Les prochaines étapes ne nécessitant pas de travailler sous la hotte de la pièce 15.2, revenir dans le « box isotopes » de la pièce N°27 (transport des contenants de <sup>33</sup>P dans une boîte en métacrylate).

Prendre soin de conserver le reste du « mix » de marquage qui vient d'être distribué.

Incuber tous les tubes de réaction au thermocycleur 2h à 42°C puis 5mn à 4°C.

Pendant l'incubation, opérer la vérification suivante : prélever 3µL du reste de « mix » de marquage, ajouter 217µL d'eau. Homogénéiser, puis prélever 2 µL de ce mélange aqueux et les ajouter à 2 mL de liquide scintillant dans une fiole *ad hoc*, et effectuer le comptage (compteur Perkin Elmer, barrette 03). S'il n'y a pas eu d'erreur de préparation, ce comptage doit être rigoureusement équivalent à celui de la veille (cf. § 8.1).

Au terme de la transcription inverse, centrifuger les tubes (condensation).

*Durant l'incubation, le brin d'ADN complémentaire est synthétisé en incorporant le dCT<sup>33</sup>P.*

### **8.5. Pré hybridation (étape à initier durant l'incubation de l'étape précédente [cf. § 8.4]) :**

- ✓ Transférer la quantité nécessaire d'ADN de sperme de saumon en tube de 1,5ml (50 à 80 µl par échantillon, d'une solution mère à 10 mg/ml). Dénaturer 5mn à 95°C puis réserver sur glace.
- ✓ Préparer un « mix » de milieu d'hybridation additionné d'ADN de sperme de saumon pour un nombre d'hybridations égal au nombre d'échantillons + 1 (par échantillon : 5mL de milieu d'hybridation + 50 à 80µl d'ADN de sperme de saumon dénaturé). Homogénéiser.
- Distribuer 4ml du « mix » milieu + ADN par flacon d'hybridation polypropylène.
- Distribuer également 1ml du « mix » milieu + ADN par tube d'une série de tubes de 2ml. Rajouter 2µl d'une solution de dA60 (3µg/µl) dans chacun de ces tubes de 2ml, et les maintenir à 65°C (agitateur Belly Dancer) sous agitation douce, durant au moins 2h.
- ✓ Hydrater les membranes de nylon en solution 2X SSC (température ambiante) et les transférer délicatement en flacon d'hybridation sans toucher à la zone des dépôts, à l'aide d'une pince à extrémités non perforantes. Les membranes vont adhérer à la paroi du flacon (face sans dépôt contre le plastique).
- ✓ Installer les flacons dans le four à hybridation, bouchon à gauche. Pré hybrider au moins 2h à 65°C sous agitation dans le four à hybridation.

### **8.6. Lyse alcaline :**

Ouvrir les tubes issus de la transcription inverse de l'étape 8.4,

Ajouter séquentiellement dans chaque tube, et dans l'ordre :

SDS 5%	1µL
EDTA 0.5M	2,5µL
NaOH 3M	3µL

*Notez S.V.P. : pour homogénéiser, éviter les aspirations refoulements qui génèrent des bulles et potentiellement de nébulisats radioactifs. Préférer des mouvements rotatifs de la pointe.*

Incuber 45mn à 68°C puis 15mn à 20°C puis 5mn à 4°C. Centrifuger (condensation).

Réserver les tubes sur glace.

*Cette étape à pour objet de procéder à la dégradation des chaînes d'ARN ayant servi de matrices à la transcription inverse. A son terme, seuls les brins d'ADNc conserveront leur intégrité dans le mélange réactionnel.*

### **8.7. Neutralisation :**

Ajouter 25µl de Tris HCl 1M pH 7,5 à chacun des tubes issus de l'étape précédente. Homogénéiser par aspirations-refoulements.

Réserver les tubes de produits neutralisés sur glace.

### **8.8. Elimination des nucléotides non incorporés :**

Préparer une colonne de chromatographie G50 par tube

Homogénéiser le contenu de la colonne par une succession de retournements vifs,

Casser l'ergot terminal en plastique au niveau du rétrécissement *ad hoc*, dévisser le bouchon et déposer en équilibre en tube de 2ml sans bouchon

Laisser tomber 1 goutte (vérification d'une liberté d'écoulement)

Centrifuger 1mn à 4500rpm (750G) pour éliminer le tampon de stockage

Installer sur tube de 1.5ml identifié, sans bouchon

Retirer le bouchon à vis des colonnes

Derrière l'écran protecteur en métacrylate, prélever le contenu de chaque tube issu de l'étape de neutralisation (§ 8.7), et le déposer délicatement en surface de la colonne égouttée de gel G50.

Replacer les bouchons à vis sur les colonnes.

Centrifuger 2mn à 4500rpm (750G).

Eliminer (poubelle radioactivité « solide » <sup>33</sup>P) les colonnes qui contiennent encore le dCT<sup>33</sup>P en excès, non incorporé aux chaînes d'ADNc.

Ajouter 50µl d'eau pure par tube et homogénéiser délicatement.

Réserver les tubes sur glace.

*Cette étape à pour objet de procéder à l'élimination du mélange réactionnel de tous les produits, de petites tailles, qui ne sont pas les ADNc radio marqués.*

## 8.9. Hybridation :

Placer sur chaque tube issu de l'étape précédente un collier en plastique coloré (étagère du « box isotopes ») pour éviter les ouvertures intempestives et les risques d'éclaboussures avec des solutions radioactives.

Dénaturer les ADNc marqués, 5mn à 95°C (bain sec dans le « box isotopes ») puis déposer rapidement dans la glace.

Ajouter le contenu de chaque tube à celui du tube correspondant en attente à 65°C (tubes de 2 mL contenant 1 mL de milieu d'hybridation avec ADN, issus de l'étape 8.5).

Incuber 1h à 65°C.

*A cette étape, l'oligonucléotide dA60 s'apparie aux extrémités poly-T des ADNc, inhibant ainsi leur interaction ultérieure, non spécifique, avec d'éventuelles séquences poly-A contenues dans les fragments d'ADN représentés sur la puce.*

Au terme de l'incubation : prélever 2µl du contenu de chaque tube pour un comptage (voir § suivant :8.10).

Derrière l'écran protecteur en métacrylate, ajouter le contenu de chaque tube à celui de chacun des flacons d'hybridation correspondants, qui sont en préhybridation à 65°C, dans le four rotatif.

Incuber une nuit à 65°C sous agitation dans le four à hybridation.

## 8.10. Comptage :

Déposer 2µl d'échantillon marqué par flacon de comptage + 2ml de scintillant. Ne pas laisser le cône. Compter (Barrette 03 au compteur Perkin Elmer).

Après comptage, récupérer les fioles et les stocker dans la poubelle réservée à cet usage sous la paillasse du « box isotopes ».

*Note : le comptage est fait à titre indicatif mais n'est pas nécessaire pour la suite du protocole qui peut se poursuivre pendant le comptage.*

## 8.11. Lavages :

Le lendemain, préchauffer le tampon de lavage à 65°C (0,1X SSC-0,1% SDS).

Derrière l'écran protecteur, vider le contenu de chaque flacon dans l'entonnoir se déversant dans la poubelle « liquides » appropriée ( $^{33}\text{P}$ ).

Faire 1 lavage rapide avec 9 mL/flacon

    2 lavages de ½ h chacun avec 9ml (65°C, + rotation)

    1 lavage de 1h avec 9ml (65°C, + rotation)

Préparer une feuille de papier 3MM de dimension adéquate imbibée de 2X SSC. La déposer sur film saran.

Transférer les membranes en bac contenant du 2X SSC puis les installer sur le papier 3M (face des dépôts en surface).

Recouvrir de film Saran, en évitant la formation de la moindre bulle d'air entre membranes de nylon et film plastique, et fermer hermétiquement.

Exposer contre un écran storm screen : 72h puis éventuellement 48h ou 24h.

Effectuer une lecture à l'imageur Storm à la définition de 50µm. Durant cette lecture aucune opération simultanée avec le lecteur BioRad n'est autorisée. Le fichier source sera immédiatement sauvegardé. Toutes les opérations de traitement d'image (contraste, ...) seront effectuées à partir d'une copie du fichier source.

## 8.12. Contrôles :

A tout moment de la journée et plus particulièrement après la mise en incubation du soir et les lavages du lendemain, rechercher au compteur bêta (position C14) une éventuelle contamination du matériel et des locaux. Décontaminer au Decon 90.

Compléter le classeur vert (inventaire sources) et le classeur violet (inventaire utilisateurs) rangés sur l'étagère du « box isotopes ».

### 8.13. Déshybridation :

Transférer les membranes en 2X SSC pour éviter qu'elles ne sèchent.

Les installer entre 2 résilles, mettre l'ensemble dans un tube *ad hoc*, en borosilicate à bouchon rouge, entreposé sur la paillasse du « box isotopes ».

Effectuer 3 lavages successifs avec rotation :

1h à 55°C : NaOH 0,4M (100mL / tube grand modèle)- vider

1/2h à 55°C : tampon de neutralisation (0.1X SSC - 0.1% SDS – 0.2M tris HCl) - vider

1h à 65°C : 0.1% SDS 1mM EDTA

Vérifier l'efficacité de la déshybridation en exposant à nouveau les membranes 24h.

### 8.14. Première hybridation de membranes neuves :

Allumer le four à hybridation à 65°C.

Découper, randomiser et numérotier les membranes avec un stylo bic bleu (N° en bas à droite, encoche en haut à gauche par exemple).

Poser les membranes en bac contenant du 2X SSC pour les hydrater. Les installer individuellement et délicatement en flacons à hybridation, à l'aide d'une pince adaptée (ne pas rayer ou plier les membranes, ne pas toucher les spots).

- Déposer 5ml de milieu d'hybridation + sperme de saumon par flacon. Installer les flacons dans le four et effectuer un bain de 1h à 65°C

- Vider le milieu, le remplacer par 10ml de tampon de lavage (0,1X SSC, 0,1% SDS) et effectuer un bain de 1h à 65°C

- Vider le tampon, le remplacer par 10ml de tampon de déshybridation (1mM EDTA, 0,1% SDS) et effectuer un bain de 1h à 65°C.

Extraire les membranes des flacons, les déposer en bac + 2X SSC puis les installer sur papier 3M imbibé de 2X SSC. Fermer l'ensemble avec du saran, recouvrir de papier d'aluminium et conserver à 4°C.

### 8.15. Tampons et milieux nécessaires :

#### Milieu d'hybridation

5X SSC, 5X denhardt, 0,5% SDS, soit en détaille :

NaCl 0.75M, citrate tri-sodium dihydraté ( $C_6H_5O_7Na_3 \cdot 2H_2O$ ) 75mM, SDS 0.5%, albumine sérique bovine (BSA) fraction V 1g/l, polyvinylpyrrolidone (PVPL) 1g/l, ficoll 400 1g/l.

20X SSC 250ml

100X denhart 50ml

20% SDS 25ml

eppi qsp 11 675ml

Préparer le milieu puis le conserver à -20°C.

**20X SSC :** concentration finale :

NaCl	175g	3M	Sigma S9625
------	------	----	-------------

Citrate tri sodique, dihydraté	88g	0,3M	( $C_6H_5O_7Na_3 \cdot 2H_2O$ ) euromedex 1126
--------------------------------	-----	------	--

Eau désionisée	qsp 11
----------------	--------

Stockage à température ambiante

**2X SSC :**

20X SSC dilué au 1/10 : 100ml 20X SSC + 900ml eau milliQ

**0,1X SSC 0,1% SDS :**

10ml 20X SSC + 10ml SDS 20% + 1980ml eau milliQ (qsp 2l)

**0,1X SSC - 0,1% SDS – 0,2M tris HCl :**

10ml 20X SSC

10ml SDS 20%

400ml tris HCl 1M

eau milliQ qsp 2l

**0.1% SDS 1mM EDTA**

10ml SDS 20%

4ml EDTA 500mM

eau milliQ qsp 2l

**dNTP mix pour micropuces :** *20mM dATP – 20mM dTTP – 20mM dGTP*

dATP 100mM : 60µl

dGTP 100mM : 60µl

dTTP 100mM : 60µl

Eau 120µl

**NaOH 3M :**

Utiliser du matériel en verre pyrex.

Dissoudre 120g de pastilles de NaOH dans 900ml d'eau désionisée. La réaction étant exothermique, attendre le retour à la température ambiante pour compléter le volume à 1l. Conserver à température ambiante.

**NaOH 0,4M :**

Utiliser du matériel en verre pyrex.

Dissoudre 16g de pastilles de NaOH dans 900ml d'eau désionisée. La réaction étant exothermique, attendre le retour à la température ambiante pour compléter le volume à 1l. Conserver à température ambiante.

**Tris HCl 1M :**

Dans une éprouvette 100ml, verser 15,7g de poudre dans 90ml d'eau désionisée. Après dissolution, compléter le volume à 100ml et transférer en flacon en verre. Conserver à température ambiante.

**SDS 5% :**

Mélanger 10ml de SDS 20% et 30ml d'eau désionisée. Conserver à température ambiante

**dA60 :**

Hydrater à 3µg/µl avec de l'eau milliQ autoclavée.

**dT25 :**

Hydrater à 4µg/µl avec de l'eau milliQ autoclavée.



**ANNEXE 2 : Liste des clones présents sur la puce  
INRArray 01.4**

**Annexe 2 : liste des 295 clones présents sur la puce à ADN INRAArray 01.4**

no	Gene ID	Gene Name	GenBank Acc. Num.	Size (bp)	Locus id
1	<b>36B4</b>	Acidic ribosomal phosphoprotein	NM_007475	1000	11837
2	<b>Abca1 / mABC1</b>	Mouse ATP binding-cassette A1 (ABCa1)	NM_013454	1507	11303
3	<b>Abcb11 / BSEP</b>	Bile Salt Export Pump (ABCb11)	NM_021022	1185	83569
4	<b>Abcb1a / MDR1a</b>	Multidrug resistance 1a (ABCb1a)	NM_011076	390	18671
5	<b>Abcb4 / MDR2</b>	Multidrug resistance protein 2 (ABCb4)	NM_008830	392	18670
6	<b>Abcc1/MRP1</b>	ATP-binding cassette, subfamily C, member 1	NM_008576	1165	17250
7	<b>Abcc2 / cMOAT</b>	Canalicular multispecific organic anion transporter (ABCc2)/(MRP2)	NM_013806	1055	12780
8	<b>Abcc3</b>	ATP-binding cassette, subfamily C, member 3	NM_029600	1135	76408
9	<b>Abcc6 / MRP6</b>	multidrug resistance-associated protein-6 (ABCc6)	NM_018795	1449	27421
10	<b>Abcd2 / ALDR</b>	Adrenoleukodystrophy-related protein (ABC D2)	NM_011994	501	26874
11	<b>Abcd3 / PMP70</b>	Peroxisomal membrane protein 70KDa (ABC D3)	NM_008991	622	19299
12	<b>Abcg1</b>	ATP-binding cassette, subfamily G, member 1	NM_009593	903	11307
13	<b>Abcg2 / BCRP</b>	ATP-binding cassette, subfamily G, member 2	NM_011920	458	26357
14	<b>Abcg4</b>	ATP-binding cassette, subfamily G, member 4	NM_138955	1016	192663
15	<b>Abcg5</b>	ATP-binding cassette transporter G5	NM_031884	995	27409
16	<b>Abcg8</b>	ATP-binding cassette transporter G8	NM_026180	990	67470
17	<b>Acaa1a / ThiolA</b>	Peroxisomal 3-ketoacyl-CoA thiolase A	NM_130864	136	113868
18	<b>Acaa1b / THIOL</b>	Peroxisomal 3-ketoacyl-CoA thiolase B	NM_146230	1004	24157
19	<b>Acadm / MCAD</b>	Medium chain acetyl-CoA dehydrogenase	NM_007382	1095	11364
20	<b>ACAT1 / Soat1</b>	Acyl-CoA:cholesterol acyltransferase 1	NM_009230	1200	20652
21	<b>ACAT2 / Soat2</b>	Acyl-CoA:cholesterol acyltransferase 2	NM_146064	1195	223920
22	<b>ACC1 / Acaca</b>	Acetyl-CoA carboxylase 1	NM_022193	386	60581
23	<b>ACC2 / Acacb</b>	Acetyl-CoA carboxylase 2	NM_133904	554	100705
24	<b>Acdc / AdipoQ</b>	adipocyte, C1Q and collagen domain containing (Acdc) / 30kDa adipocyte complement-related protein (Acp30) / AdipoQ	NM_009605	747	11450
25	<b>ACOTH / Acot1</b>	Cytosolic acyl-CoA thioesterase	NM_012006	880	26897
26	<b>ADSS1</b>	Adenylosuccinate synthetase 1, muscle	NM_007421	1140	11565
27	<b>Alas1</b>	aminolevulinic acid synthase 1	NM_020559	750	11655
28	<b>Aldh1a1 / ALDH1</b>	Aldehyde dehydrogenase 1 (Aldh1a1)	NM_013467	1231	11668
29	<b>Aldh3a2 / ALDH3</b>	Microsomal class 3 aldehyde dehydrogenase	NM_007437	1198	11671
30	<b>AOX / Acox1</b>	Acyl-CoA oxidase	NM_015729	1259	11430
31	<b>apoA-I</b>	apolipoprotein A-I	NM_009692	698	11806
32	<b>apoA-II</b>	apolipoprotein A-II	NM_013474	448	11807
33	<b>apoA-IV</b>	Apolipoprotein A-IV	NM_007468	1049	11808
34	<b>apo-AV</b>	Apolipoprotein A-V	NM_080434	806	66113
35	<b>apoB</b>	apolipoprotein B100	NM_009693	1077	238055
36	<b>apoC-III</b>	apolipoprotein C-III	NM_023114	248	11814
37	<b>apoE</b>	apolipoprotein E	NM_009696	903	11816
38	<b>apoJ / Clu</b>	Clusterin / apoJ / testosterone-repressed prostate message 2 /complement lysis inhibitor / SP-40,40 / sulfated glycoprotein 2	NM_013492	1106	12759
39	<b>Atf4</b>	activating transcription factor	NM_009716	972	11911
40	<b>ATPsA / Atp5a1</b>	ATP synthase alpha subunit	NM_007505	987	11946
41	<b>ATPsB / Atp5b</b>	ATP synthase beta subunit	NM_016774	988	11947
42	<b>BACT / Actb</b>	Beta-actine	NM_007393	1000	11461
43	<b>Bax</b>	Bcl2-associated X protein	NM_007527	700	12028
44	<b>b-catenine</b>	Mouse (clone pMCAT) beta-catenin mRNA	NM_007614	501	12387
45	<b>Bcl2</b>	B-cell leukemia/lymphoma 2	NM_177410	706	12043
46	<b>Bcl3</b>	B-cell leukemia/lymphoma 3	NM_033601	1006	12051
47	<b>BIEN / Ehhadh</b>	Enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase (Bifunctional ENzyme)	NM_133606	1200	171142
48	<b>BUG</b>	Brown Unknown Gene	AF247002	407	NA
49	<b>Bzrp / PBR</b>	peripheral benzodiazepine receptor / Bzrp	NM_009775	701	12257
50	<b>C/EBPalpha</b>	Mus musculus, Similar to CCAAT/enhancer binding protein (C/EBP),alpha	NM_007678	522	12606
51	<b>C/EBPgamma</b>	Mus musculus mRNA for GPE1-BP ( C/EBP gamma)	NM_009884	519	12611
52	<b>C16SR</b>	Chimeric 16S mitochondrial RNA	AF089815	1109	NA
53	<b>C9</b>	Complément C9	NM_013485	1161	12279
54	<b>Casp2</b>	caspase 2	NM_007610	1104	12366
55	<b>Casp3</b>	caspase 3, apoptosis related cysteine protease	NM_009810	604	12367
56	<b>Casp8</b>	caspase 8	NM_009812	1012	12370
57	<b>Casp9</b>	caspase 9	NM_015733	1078	12371
58	<b>catalase</b>	Mus musculus catalase 1	NM_009804	526	12359
59	<b>Cbs</b>	Cystathione beta synthase	NM_012522	1187	24250
60	<b>Ccnd1 / CyclinD1</b>	Cyclin D1	NM_007631	785	12443
61	<b>Ccnd3 / Cyclin D3</b>	Cyclin D3	NM_007632	907	12445
62	<b>Ccng1 / Cyclin G1</b>	Cyclin G1	NM_009831	1006	12450
63	<b>Ccnt2</b>	a amino b carboxymuconate e semialdehyde decarboxylase / cyclin T2	NM_028399	1003	72949
64	<b>Cd36 / FAT</b>	Fatty acid transporter / CD36 antigen	NM_007643	1239	12491
65	<b>Cdkn1a / Waf1</b>	cyclin-dependent kinase inhibitor 1A (p21, Cdkn1a, Cip1, Waf1)	NM_007669	513	12575

66	<b>c-fos / Fos</b>	FBJ osteosarcoma oncogene	NM_010234	1055	14281
67	<b>CHOP10 / Ddit3</b>	Mus musculus, DNA-damage inducible transcript 3	NM_007837	523	13198
68	<b>Cidea</b>	Cell death activator Cide-a	NM_007702	654	12683
69	<b>Cide-a</b>	Cell death-inducing DNA fragmentation factor, alpha subunit-like effector A	NM_007702	566	12683
70	<b>Cide-b</b>	Cell death-inducing DNA fragmentation factor, alpha subunit-like effector B	NM_009894	1004	12684
71	<b>CIDEC / ADISP</b>	Adipocyte specific mRNA (Fat specific gene 27)	NM_178373	1143	14311
72	<b>Cide-c / ADISP</b>	Cell death-inducing DFFA-like effector c	NM_178373	610	14311
73	<b>c-jun / Jun</b>	Mus musculus Jun oncogene (Jun), mRNA.	NM_010591	526	16476
74	<b>COX1 / Ptgs1</b>	Prostaglandin-endoperoxide synthase 1 (cyclooxygenase 1)	NM_008969	829	19224
75	<b>COX2 / Ptgs2</b>	Prostaglandin synthase / cyclooxygenase 2	NM_011198	452	19225
76	<b>Cpt1a / L-CPT1</b>	Liver Carnitine Palmitoyltransferase I	NM_013495	519	12894
77	<b>Cpt1b / M-CPT1</b>	Muscle type carnitine palmitoyltransferase I	NM_009948	1198	12895
78	<b>CPT2</b>	Carnitine palmitoyltransferase II	NM_009949	1084	12896
79	<b>CRABP II</b>	Cellular retinoic acid-binding protein II	NM_007759	790	12904
80	<b>CRBP I / Rbp1</b>	Cellular retinol binding protein 1	NM_011254	571	19659
81	<b>CRBP II / Rbp2</b>	Cellular retinol binding protein 2	NM_009034	436	19660
82	<b>CREB / Creb1</b>	Cyclic AMP Element	NM_133828	508	12912
83	<b>Cyp11a1</b>	P450scc (side chain cleavage enzyme 11a1)	NM_019779	1072	13070
84	<b>Cyp19a1</b>	aromatase	NM_007810	1045	13075
85	<b>Cyp1a1</b>	cytochrome P450 1a1	NM_009992	1091	13076
86	<b>Cyp21a1</b>	Cytochrome P450 21a1	NM_009995	994	13079
87	<b>Cyp24</b>	25-hydroxyvitamin D 24-hydroxylase (cytochrome P450, family 24, subfamily a, polypeptide 1)	NM_009996	1036	13081
88	<b>Cyp26</b>	Retinoic acid 4-hydroxylase (cytochrome P450, family 26, subfamily a, polypeptide 1)	NM_007811	998	13082
89	<b>Cyp27a1</b>	sterol 27-monoxygenase (cytochrome P450, family 27, subfamily a, polypeptide 1)	NM_024264	646	301517
90	<b>Cyp27b1</b>	25-hydroxyvitamin D3 1alpha-hydroxylase (cytochrome P450, family 27, subfamily b, polypeptide 1)	NM_010009	1001	13115
91	<b>Cyp2a5</b>	Cytochrome P450, 2a5	NM_007812	524	13087
92	<b>Cyp2b10</b>	Testosterone 16-alpha hydroxylase (cytochrome P450, family 2, subfamily b, polypeptide 10)	NM_009998	1021	13088
93	<b>Cyp2b13</b>	Testosterone 16-alpha hydroxylase type b (cytochrome P450, family 2, subfamily b, polypeptide 13)	NM_007813	1223	13089
94	<b>Cyp2b19</b>	skin-specific Cytochrome P450 2b19	NM_007814	756	13090
95	<b>Cyp2c29</b>	Microsomal aldehyde oxygenase (cytochrome P450, family 2, subfamily c, polypeptide 29)	NM_007815	1608	13095
96	<b>Cyp2c38</b>	cytochrome P450, family 2, subfamily c, polypeptide 38	NM_010002	947	13097
97	<b>Cyp2c40</b>	cytochrome P450, family 2, subfamily c, polypeptide 40	NM_010004	308	13099
98	<b>Cyp2j5</b>	cytochrome P450, family 2, subfamily j, polypeptide 5	NM_010007	506	13109
99	<b>Cyp3a11</b>	Steroid inducible cytochrome P450 3a11 (cytochrome P450, family 3, subfamily a, polypeptide 11)	NM_007818	1031	13112
100	<b>Cyp4a10</b>	Fatty acid omega-hydroxylase (cytochrome P450, family 4, subfamily a, polypeptide 10)	NM_010011	1287	13117
101	<b>Cyp4a12</b>	cytochrome P450, 4a12	NM_172306	820	13118
102	<b>Cyp4a14</b>	Fatty acid omega-hydroxylase (cytochrome P450, family 4, subfamily a, polypeptide 14)	NM_007822	1247	13119
103	<b>Cyp5a1</b>	Cytochrome P450, family 5, subfamily a, polypeptide 1 / thromboxane A synthase 1, platelet (Tbxas1)	NM_011539	1003	21391
104	<b>Cyp7a</b>	Cholesterol alpha-7-hydroxylase (cytochrome P450, family 7, subfamily a, polypeptide 1)	NM_007824	1200	13122
105	<b>Cyp7b1</b>	cytochrome P450, family 7, subfamily b, polypeptide 1	NM_007825	935	13123
106	<b>Cyp8b1</b>	Sterol 12-alpha-hydrolase (cytochrome P450, family 8, subfamily b, polypeptide 1)	NM_010012	1020	13124
107	<b>cytochrome B / Cybb</b>	cox1 mus musculus mitochondrial genome: cytochromeB	NM_007807	519	13058
108	<b>cytochrome C / Cox4i1</b>	mouse cox4 mRNA for cytochrome C oxydase subunit IV	NM_009941	417	12857
109	<b>Dbi / ACBP</b>	Acyl-CoA binding protein	NM_007830	331	13167
110	<b>Eci / Dci</b>	Enoyl-CoA isomérase	NM_010023	602	13177
111	<b>eif2g / Eif2s3y</b>	Eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked	NM_012011	1197	26908
112	<b>Elo1 / Elov1</b>	Elongase 1	NM_019422	1018	54325
113	<b>Elo2 / Elov2</b>	Elongase 2	NM_019423	978	54326
114	<b>Elo3 / Elov3</b>	Elongase 3	NM_007703	1056	12686
115	<b>Elo4 / Elov4</b>	Elongase 4	NM_148941	944	83603
116	<b>Elo5 / Elov5</b>	Elongase 5	NM_134255	875	68801
117	<b>Elo6 / Elov6</b>	Elongase 6	NM_130450	907	170439
118	<b>Elo7 / Elov7</b>	Elongase 7	NM_029001	989	74559
119	<b>Fabp1 / L-FABP</b>	Liver fatty acid binding protein	NM_017399	358	14080
120	<b>Fabp2 / i-FABP</b>	Intestinal fatty acid binding protein	NM_007980	361	14079
121	<b>Fabp4 / ap2</b>	Adipocyte fatty acid binding protein 4 (Fabp4)	NM_024406	363	11770
122	<b>Fabp6 / i-BABP</b>	Ileal fatty acid binding protein 6 (Bile Acid Binding Protein)	NM_008375	344	16204
123	<b>Fadd</b>	Fas (Tnfrsf6)-associated via death domain	NM_010175	882	14082
124	<b>Fads1 / Δ5</b>	Delta 5 desaturase	NM_146094	205	76267

125	<b>Fads2 / Δ6</b>	Delta 6 desaturase	NM_019699	204	56473
126	<b>FAS / Fasn</b>	Fatty acid synthase	NM_007988	1374	14104
127	<b>Fas / Tnfrsf6</b>	Fas antigen / Tumor necrosis factor receptor superfamily, member 6 (Tnfrsf6)	NM_007987	865	14102
128	<b>Fas-L</b>	Fas-ligand / Tumor necrosis factor (ligand) superfamily, member 6 (Tnfsf6)	NM_010177	804	14103
129	<b>Fbp1</b>	Fructose 1,6-biphosphatase 1 (Liver)	NM_019395	351	14121
130	<b>Fbp2</b>	Fructose 1,6-biphosphatase 1 (Muscle)	NM_007994	350	14120
131	<b>FDFT / Fdft1</b>	Farnesyl-diphosphate farnesyl transferase I (squalene synthase)	NM_010191	952	14137
132	<b>Fhl2</b>	Four and a half LIM domains 2	NM_010212	876	14200
133	<b>FIAF / Angptl4</b>	Fasting Induced Adipose Factor	NM_020581	1173	57875
134	<b>Fibgpp / Fgg</b>	Fibrinogen gamma polypeptide	NM_133862	1079	99571
135	<b>Foxc2</b>	M.musculus MFH-1 gene	NM_013519	575	14234
136	<b>Fpps / Fdps</b>	Farnesyl diphosphate Synthase	NM_134469	1033	83791
137	<b>G6Pase / G6pc</b>	Glucose-6-phosphatase	NM_008061	1005	14377
138	<b>G6PDH / G6pdx</b>	Glucose-6-phosphate dehydrogenase	NM_008062	1289	14381
139	<b>GA3PDH / Gapdh</b>	Glyceraldehyde 3 Phosphate Deshydrogenase	NM_008084	560	14433
140	<b>GCSH / Gclc</b>	g-glutamylcysteine synthetase, heavy chain (= glutamate-cysteine ligase catalytic subunit)	NM_010295	1116	14629
141	<b>GCSL / Gclm</b>	g-glutamylcysteine synthetase, light chain (= glutamate-cysteine ligase modifier subunit)	NM_008129	854	14630
142	<b>GK / Gck</b>	Glucokinase	NM_010292	1195	103988
143	<b>Got1 / ASAT</b>	Aspartate Amino-Transférase	NM_010324	881	14718
144	<b>Gpt1 / ALAT</b>	Alanine Amino-Transférase	NM_182805	883	76282
145	<b>GS / Gys1</b>	Glycogen synthase 1, muscle	NM_030678	1223	14936
146	<b>GSS</b>	glutathione synthetase	NM_008180	1169	14854
147	<b>GSTa3</b>	Glutathione S-transferase, alpha 3	NM_010356	819	14859
148	<b>GSTM1</b>	Glutathione S-transferase, mu 1/Glutathione S-transferase classe mu	NM_010358	993	14862
149	<b>GSTp1</b>	Glutathione S-transferase, pi 2	NM_013541	710	14870
150	<b>GSTt1</b>	glutathione S-transferase, theta 1	NM_008185	423	14871
151	<b>GSTt2</b>	glutathione S-transferase, theta 2	NM_010361	504	14872
152	<b>GSTt3</b>	glutathione S-transferase, theta 3	NM_133994	451	103140
153	<b>Gulo</b>	Gulonolactone L-oxidase	NM_178747	1084	268756
154	<b>Hadha / Tpa</b>	Mitochondrial trifunctional protein, alpha-subunit	NM_178878	1189	170670
155	<b>Hadhb / Tpb</b>	Mitochondrial trifunctional protein, beta-subunit	NM_145558	1200	171155
156	<b>Hamp2</b>	Hepcidin antimicrobial peptide 2	NM_183257	364	66438
157	<b>Hmgcr / HMGCoAred</b>	3-hydroxy 3-methylglutaryl-CoA reductase	NM_008255	974	15357
158	<b>Hmgcs1 / cHMGCoAS</b>	Cytosolic 3-hydroxy 3-methylglutaryl-CoA synthase	NM_017268	1019	29637
159	<b>Hmgcs2 / mHMGCoAS</b>	Mitochondrial 3-hydroxy 3-methylglutaryl-CoA synthase	NM_173094	1089	24450
160	<b>Hmox1</b>	heme oxygenase (decycling) 1	NM_010442	865	15368
161	<b>Hoxa1</b>	Homeobox A1	NM_010449	625	15394
162	<b>Hoxb1</b>	Homeobox B1	NM_008266	416	15407
163	<b>HPNCL / Phyh2</b>	2-hydroxyphytanoyl-CoA lyase	NM_019975	1188	56794
164	<b>Hsd11b1</b>	hydroxysteroid dehydrogenase 11b1	NM_008288	536	15483
165	<b>Hsd17b4</b>	hydroxysteroid dehydrogenase 17b4	NM_008292	1133	15488
166	<b>Hsd3b1</b>	hydroxysteroid dehydrogenase-1, delta<5>-3-beta	NM_008293	351	15492
167	<b>Hsd3b2</b>	hydroxysteroid dehydrogenase-2, delta<5>-3-beta	NM_153193	740	15493
168	<b>Hsd3b4</b>	hydroxysteroid dehydrogenase 3b4	NM_008294	381	15495
169	<b>id2 / hlhid2</b>	inhibitor of DNA binding 2 / helix-loop-helix protein id2	NM_010496	471	15902
170	<b>Ikbkg</b>	Mus musculus NF-κB essential modulator mRNA, complete cds	NM_010547	501	16151
171	<b>IL2</b>	Interleukin 2	NM_008366	489	16183
172	<b>IL6</b>	interleukin 6	NM_031168	651	16193
173	<b>i-NOS / Nos2</b>	Inducible nitric oxide synthase 2	NM_010927	1199	18126
174	<b>Insig1</b>	Insulin-induced gene 1	NM_153526	1146	231070
175	<b>Insig2</b>	Insulin-induced gene 2	NM_133748	1057	72999
176	<b>Jak3 / Insl3</b>	Janus Kinase 3 / insulin-like hormone 3	NM_013564	405	16336
177	<b>Lce / Es1</b>	Liver carboxylesterase (esterase 1)	NM_007954	1223	13884
178	<b>LDLR</b>	Low density lipoprotein receptor	NM_010700	1208	16835
179	<b>LEF1</b>	Lymphoid Enhancer binding Factor 1	NM_010703	510	16842
180	<b>Lep / ob</b>	Leptin	NM_008493	477	16846
181	<b>LH / Lipc</b>	Hepatic Lipase	NM_008280	985	15450
182	<b>Lpin1</b>	Lipin 1	NM_172950	1017	14245
183	<b>Lpin2</b>	Lipin 2	NM_022882	1050	64898
184	<b>Lpin3</b>	Lipin 3	NM_022883	961	64899
185	<b>LPK / Pkrl</b>	L-type pyruvate kinase	NM_013631	1194	18770
186	<b>LPL</b>	Lipoprotein Lipase	NM_008509	867	16956
187	<b>Lrp1 / AM2R</b>	Alpha-2-macroglobulin receptor (Low density lipoprotein receptor-related protein 1 -Lrp1-)	NM_008512	1151	16971
188	<b>Ltf</b>	Lactotransferrin	NM_008522	812	17002
189	<b>Ly6d / THB</b>	Thymocyte B cell antigen (Lymphocyte antigen 6 complex, locus D)	NM_010742	356	17068
190	<b>Mdm2</b>	Transformed mouse 3T3 cell double minute 2	NM_010786	1016	17246
191	<b>MS / Mtr</b>	Methionine synthase	NM_030864	1203	81522

192	<b>MTHFR</b>	5,10-methylenetetrahydrofolate reductase	NM_010840	1181	17769
193	<b>MUP2</b>	Major Urinary Protein 2	NM_008647	717	17841
194	<b>NcoA1 / Src1</b>	Steroid Receptor Coactivator	NM_010881	502	17977
195	<b>Ndufb9</b>	Mus musculus NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9	NM_023172	653	66218
196	<b>Nr0b1 / DAX-1</b>	nuclear receptor subfamily 0, group B, member 1	NM_007430	1022	11614
197	<b>Nr0b2 / SHP1</b>	Small Heterodimer Partner-1 (Nr0b2)	NM_011850	734	23957
198	<b>Nr1a1 / TRa</b>	Thyroid hormone receptor alpha (Nr1a1)	NM_178060	998	21833
199	<b>Nr1a2 / TRb</b>	Thyroid hormone receptor beta (Nr1a2)	NM_009380	1013	21834
200	<b>Nr1b1 / RARa</b>	Retinoic acid receptor alpha (Nr1b1)	NM_009024	1201	19401
201	<b>Nr1b2 / RARB2</b>	Retinoic acid receptor beta 2 (Nr1b2)	NM_011243	1222	218772
202	<b>Nr1b3 / RARg1</b>	Retinoic acid receptor gamma 1	NM_011244	1552	19411
203	<b>Nr1c1 / PPARa</b>	Peroxisome proliferator-activated receptor alpha (Nr1c1)	NM_011144	1211	19013
204	<b>NR1c1 / PPARalpha</b>	Peroxisome Proliferator Activated Receptor alpha	NM_011144	513	19013
205	<b>Nr1c2 / PPARd</b>	Peroxisome proliferator-activated receptor delta (Nr1c2)	NM_011145	1194	19015
206	<b>Nr1c3 / PPARg</b>	Peroxisome proliferator-activated receptor gamma (Nr1c3)	NM_011146	1187	19016
207	<b>Nr1h2 / LXRb</b>	Liver-enriched X receptor beta (Nr1h2)	NM_009473	1147	22260
208	<b>Nr1h3 / LXRa</b>	Liver-enriched X receptor alpha (Nr1h3)	NM_013839	1070	22259
209	<b>Nr1h4 / FXR</b>	Farnesoid X receptor (Nr1h4)	NM_009108	983	20186
210	<b>Nr1i1 / VDR</b>	Vitamin D receptor (Nr1i1)	NM_009504	1007	22337
211	<b>Nr1i2 / PXR</b>	Pregnane X receptor (Nr1i2)	NM_010936	924	18171
212	<b>Nr1i3 / CAR1</b>	Constitutive androstane receptor beta (Nr1i3)	NM_009803	918	12355
213	<b>Nr2b1 / RXRa</b>	Retinoid X receptor alpha (Nr2b1)	NM_011305	1178	20181
214	<b>Nr2b2 / RXRb2</b>	Retinoid X receptor beta (Nr2b2)	NM_011306	974	20182
215	<b>Nr2b3 / RXRg1</b>	Retinoid X receptor gamma (Nr2b3)	NM_009107	1211	20183
216	<b>Nr3a1 / Esr1</b>	estrogen receptor 1 / estrogen receptor alpha	NM_007956	1084	13982
217	<b>Nr3a2 / Esr2</b>	estrogen receptor 2 / estrogen receptor beta	NM_010157	977	13983
218	<b>Nr3c4 / AR</b>	Androgen receptor	NM_013476	975	11835
219	<b>Nr4a1 / NGFIB</b>	Nerve growth factor-induced gene B (Nr4a1)	NM_010444	1002	15370
220	<b>Nr4a2 / NURR1</b>	Nuclear receptor-related factor 1 (Nr4a2)	NM_013613	1099	18227
221	<b>Nr5a1 / SF-1</b>	nuclear receptor subfamily 5, group A, member 1 / Steroidogenic factor 1 (SF-1)	NM_139051	1078	26423
222	<b>Nr5a2 / LRH-1</b>	nuclear receptor subfamily 5, group A, member 2 / LRH-1	NM_030676	661	26424
223	<b>oatp1 / Slco1a1</b>	solute carrier organic anion transporter family, member 1a5	NM_013797	616	28248
224	<b>oatp2 / Slco1a4</b>	solute carrier organic anion transporter family, member 1a4	NM_030687	913	28250
225	<b>oatp5 / Slco1a6</b>	solute carrier organic anion transporter family, member 1a6 (Slco1a6)	NM_023718	460	28254
226	<b>Oaz1</b>	ornithine decarboxylase antizyme 1	NM_008753	627	18245
227	<b>Oaz2</b>	ornithine decarboxylase antizyme 2	NM_010952	1038	18247
228	<b>OPN</b>	Osteopontin	NM_009263	991	20750
229	<b>Oxt</b>	Oxytocin	NM_011025	353	18429
230	<b>p53 / Trp53</b>	Mouse p53 cellular tumor antigen, mRNA	NM_011640	509	22059
231	<b>PAL / Akp2</b>	Liver alkaline phosphatase 2	NM_007431	1325	11647
232	<b>PAPSS1</b>	3'-phosphoadenosine 5'-phosphosulfate synthase 1	NM_011863	803	23971
233	<b>PAPSS2</b>	3'-phosphoadenosine 5'-phosphosulfate synthase 2	NM_011864	656	23972
234	<b>Pbsn</b>	Probasin	NM_017471	490	54192
235	<b>PCNA</b>	proliferating cell nuclear antigen	NM_011045	921	18538
236	<b>PDK4</b>	Pyruvate dehydrogenase kinase, isoenzyme 4	NM_013743	1246	27273
237	<b>PECI</b>	Peroxisomal delta3, delta2-enoyl-CoA isomerase	NM_011868	1181	23986
238	<b>PEPCK / Pck1</b>	cytosolic phosphoenolpyruvate carboxykinase	NM_011044	1024	18534
239	<b>PEPCK1</b>	Phosphoenolpyruvate Carboxykinase	NM_011044	523	18534
240	<b>Pex11a</b>	Peroxisomal biogenesis factor 11a	NM_011068	738	18631
241	<b>PGC-1a / Ppargc1a</b>	Peroxisome Proliferator Activated Receptor Coactivator 1a	NM_008904	505	19017
242	<b>PGC-1b / Ppargc1b</b>	Peroxisome Proliferator Activated Receptor Coactivator 1b	NM_133249	545	170826
243	<b>Pias1</b>	Protein inhibitor of activated STAT 1	NM_019663	1098	56469
244	<b>PLB</b>	Phospholipase B1	NM_001081407	1056	665270
245	<b>PLTP</b>	Phospholipid transfer protein	NM_011125	1000	18830
246	<b>PMDC1 / Ech1</b>	Peroxisomal/mitochondrial dienoyl-CoA isomerase	NM_016772	1017	51798
247	<b>Pomc1 / POMC</b>	Pro-opiomelanocortin-alpha	NM_008895	667	18976
248	<b>Pon1 / PON</b>	Paraoxonase 1	NM_011134	1005	18979
249	<b>Ptpn13</b>	protein tyrosine phosphatase, non-receptor type 13	NM_011204	1073	19249
250	<b>Rb1 / Rb</b>	Retinoblastoma tumor suppressor	NM_009029	517	19645
251	<b>Retn</b>	resistin	NM_022984	510	57264
252	<b>Ripk1</b>	receptor (TNFRSF)-interacting serine-threonine kinase1 (Ripk1),	NM_009068	1075	19766
253	<b>S14 / Thrsp</b>	Spot 14	NM_009381	330	21835
254	<b>Sat1</b>	spermidine/spermine N1-acetyl transferase 1	NM_009121	449	20229
255	<b>Scap</b>	SREBP cleavage activating protein	NM_001001144	1070	235623
256	<b>SCD1</b>	delta 9 desaturase (stearoyl-CoA desaturase)	NM_009127	1068	20249
257	<b>SCP-2 / SCPx</b>	sterol carrier protein-2 / sterol carrier protein X	NM_011327	717	20280
258	<b>Serpina3k</b>	Mus musculus serine (or cysteine) proteinase inhibitor, clade A, member 3K	NM_011458	1113	20714
259	<b>Slc10a1 / Ntcp</b>	Na+/taurocholate cotransporting polypeptide-1 (solute carrier family 10, member 1)	NM_011387	1003	20493

260	<b>Slc10a2 / i-BAT</b>	Ileal Na-dependent Bile Acid Transporter (solute carrier family 10, member 2)	NM_011388	560	20494
261	<b>Slc22a5 / OCTN2</b>	Organic cation/carnitine transporter (solute carrier family 22, member 5)	NM_011396	255	20520
262	<b>Slc25a20 / CACP</b>	Mitochondrial carnitine/acylcarnitine translocase	NM_020520	801	57279
263	<b>Slc27a1</b>	Solute Carrier Family 27, member 1 / Fatty acid transporter 1 (FATP1)	NM_011977	900	26457
264	<b>Slc27a3</b>	Solute Carrier Family 27, member 3 / Fatty acid transporter 3 (FATP3)	NM_011988	898	26568
265	<b>Slc27a4</b>	Solute Carrier Family 27, member 4 / Fatty acid transporter 4 (FATP4)	NM_011989	1048	26569
266	<b>Slc27a5</b>	Solute Carrier Family 27, member 5 / Fatty acid transporter 5 (FATP5)	NM_009512	1086	26459
267	<b>Slc27a6</b>	Solute Carrier Family 27, member 6 / Fatty acid transporter 6 (FATP6)	XM_128932	939	225579
268	<b>Slc2a2 / Glut2</b>	Glucose Transporter 2 (solute carrier family 2 -facilitated glucose transporter-, member 2: Slc2a2)	NM_031197	990	20526
269	<b>Slc2a4 / Glut4</b>	Mus musculus solute carrier family 2 (facilitated glucose transporter), member 4 (Slc2a4), mRNA	NM_009204	510	20528
270	<b>Sod2 / MnSOD</b>	SuperOxide Dismutase	NM_013671	527	20656
271	<b>SPI1 / Serpina1a</b>	Serine (or cysteine) proteinase inhibitor 1, clade A, member 1a	NM_009243	1038	20700
272	<b>SR-BI / Scarb1</b>	Scavenger Receptor, class B, member 1	NM_016741	1195	20778
273	<b>Srd5a2</b>	steroid 5 alpha-reductase 2	NM_053188	408	94224
274	<b>SREBP1 / Srebf1</b>	Sterol Response Element Binding Protein	NM_011480	535	20787
275	<b>SREBP2 / Srebf2</b>	Sterol Regulatory Element Binding Protein 2	NM_033218	1149	20788
276	<b>St3gal4 / SIAT4c</b>	Alpha-2,3-sialyltransferase/ Sialyltransferase 4C (beta-galactoside alpha-2,3-sialyltransferase)	NM_009178	982	20443
277	<b>StAR</b>	Steroidogenic Acute Regulatory protein	NM_011485	876	20845
278	<b>Stat5b</b>	Mammary gland factor	NM_011489	533	20851
279	<b>Sult1b1</b>	Sulfotransferase family 1B, member 1	NM_019878	735	56362
280	<b>Sult1d1/Sultn</b>	sulfotransferase family 1D, member 1 (Sult1d1)	NM_016771	702	53315
281	<b>Sult2a2</b>	Sulfotransferase family 2A, dehydroepiandrosterone (DHEA)-preferring, member 1 (Sult2a2)	NM_009286	451	20865
282	<b>Tnfrsf14</b>	tumor necrosis factor receptor superfamily, member 14 (herpesvirus entry mediator)	NM_178931	504	230979
283	<b>Trf</b>	Transferrin	NM_133977	620	22041
284	<b>Ucp1</b>	Uncoupled Protein 1	NM_009463	547	22227
285	<b>Ucp2</b>	Uncoupling protein 2	NM_011671	886	22228
286	<b>Ucp3</b>	Uncoupling protein 3	NM_009464	796	22229
287	<b>UGT1a1</b>	UDP glycosyl transferase 1 family, member 1	NM_201645	872	22236
288	<b>UGT1a10</b>	UDP-glucuronosyltransferase 1 family, member 10	NM_201642	385	394432
289	<b>UGT1a2</b>	UDP-glucuronosyltransferase 1 family, member 2	NM_013701	531	22236
290	<b>UGT1a5</b>	UDP-glucuronosyltransferase 1 family, member 5	NM_201643	472	394433
291	<b>UGT1a9</b>	UDP glycosyl transferase 1 family polypeptide A9	NM_201410	874	394435
292	<b>UGT2b5</b>	UDP-glucuronosyltransferase 2 family, member 5	NM_009467	459	22238
293	<b>Vanin1 / Vnn1</b>	Vanin-1	NM_011704	1027	22361
294	<b>VLDLr</b>	Very low density lipoprotein receptor	NM_013703	1202	22359
295	<b>Wnt10B</b>	Mus musculus wnt-10B mRNA, complete cds	NM_011718	511	22410

**ANNEXE 3 : Article soumis sur l'analyse canonique  
régularisée**



# Highlighting relationships through Regularized Canonical Correlation Analysis: application to high throughput biology data

I. GONZÁLEZ<sup>(1,2)</sup>, S. DÉJEAN<sup>(1)</sup>, P.G.P. MARTIN<sup>(3)</sup>,  
O. GONÇALVES<sup>(4)</sup>, P. BESSE<sup>(1)</sup>, A. BACCINI<sup>(1)a</sup>

(1) Institut de Mathématiques de Toulouse, UMR CNRS 5219, Université Paul Sabatier

(2) Departamento de Matemática – FACYT, Universidad de Carabobo, Venezuela

(3) Laboratoire de Pharmacologie-Toxicologie, UR 66, INRA, F-31931 Toulouse, France

(4) Laboratoire de Biologie des Protistes, UMR CNRS 6023, Université Blaise Pascal, Clermont-Ferrand II

## Abstract

Biological data produced by high throughput technologies are becoming more and more abundant and are arousing many statistical problems. This paper addresses one of them when gene expression data are jointly observed with other variables with the purpose of highlighting significant relationships between gene expression and these other variables. The relevant statistical method to explore these relationships is Canonical Correlation Analysis (CCA). Unfortunately, in the context of postgenomic data, the number of variables (gene expressions) is usually greater than the number of units (samples) and CCA cannot be directly performed: a regularized version is required. In this paper, we propose adapted graphical displays issued from a Regularized CCA (RCCA) that is applied on data sets coming from two different fields and whose interpretation leads to relevant new hypothesis. The first data set come from a pharmacogenomic study: we identified new ABC transporter candidate substrates which relevancy is illustrated by the concomitant identification of several known substrates. The second one was acquired during a nutrigenomic study: we generated interesting hypothesis on the transcription factor pathways potentially linking hepatic fatty acids and gene expression.

*Keywords:* canonical correlation analysis, regularization, cross-validation, gene expression data, high throughput biology.

## 1 Introduction

During the last decade, many statistical methods have been considered to analyse the large quantity of biological data produced by microarray technology (see, for instance: Parmigiani *and others*, 2003; Speed, 2003; Baccini *and others*, 2005). More recently, some examples of gene expression data jointly observed with other variables (measuring phenotypes, pharmacological activities...) arose. In the context of systems biology, the purpose of such experiments is to highlight significant relationships between gene expression and other variables. The aim of this paper is to build graphical displays adapted to a regularized version of Canonical Correlation Analysis (CCA) in order to analyse such data sets in which the number of variables is greater than the number of units. The relevancy of the results produced by this method is assessed by way of biological and chemical considerations.

The two experiments considered here are examples of joint analysis of gene expression data with another set of variables. The first one comes from a pharmacogenomic study already analysed in Szakács *and others* (2004): expression profiles of the 48 human ABC transporters and 1429 potential anti-cancer drugs efficacy were jointly observed on the NCI-60 cancer cells lines. Using RCCA, we identified new ABC transporter candidate substrates which relevancy is illustrated by the concomitant identification of several known substrates. The second data set comes from a nutrigenomic experiment described in Martin *and others* (2007), in which expressions of 120 genes and percentages of 21 hepatic fatty acids are simultaneously observed on a sample of 40 mice. Note that it is the latter example which has motivated the development of the statistical methodology presented in this paper. Using RCCA, we generated interesting hypothesis on the transcription factor pathways potentially linking hepatic fatty acids and gene expression.

Two kinds of multidimensional statistical methods can be considered to jointly analyse two sets of variables: Partial Least Squares (PLS) Regression, when the objective is to explain variables in one set

<sup>a</sup>To whom correpondance should be addressed, alain.baccini@math.ups-tlse.fr.

by way of variables in the other one (Martens and Naes, 1989; Frank and Friedman, 1993); CCA, when the objective is to explore correlations between two sets of variables which play strictly symmetric roles in the analysis (Mardia *and others*, 1979; Gittins, 1985). PLS Regression has already been used for analysing microarray data (Nguyen & Rocke, 2002; Boulesteix, 2004; Brilli *and others*, 2007) and an R package is described in Mevik and Wehrens (2007). In the two examples considered in this paper, the roles of the two sets of variables were symmetric. Thus, CCA appeared as the most relevant statistical method to analyse their correlations. Moreover, the graphical representations produced by CCA are truly meaningful for biologists.

CCA is an old statistical method first defined by Hotelling (1936), but unused until the sixties for obvious computational reasons. It has known a short renewal of interest around 1960 (Anderson, 1958; Morrison, 1967) with the expansion of computers. However, CCA has never known the success of other multivariate exploratory methods such as Principal Components Analysis (PCA), correspondence analysis or clustering. The reasons for this lack of success were probably of two kinds: firstly the small number of examples corresponding to the context of CCA, secondly some difficulties to perform the method and to accurately interpret its results; for this last aspect, see ter Braak (1990). We believe that CCA will be more and more used in the future because of the increasing production of data on which CCA will apply, in particular with microarray screening, and also because of the development of suitable statistical packages easy to use for performing and interpreting CCA (for example, the CCA package on the Comprehensive R Archive Network available at [cran.r-project.org](http://cran.r-project.org), see González, Déjean, Martin and Baccini, CCA: an R package to extend canonical correlation analysis, submitted).

From a mathematical point of view, CCA needs, as regression or linear discriminant analysis, the computation of the inverse of a matrix (more precisely, the inverse of two matrices for CCA). A necessary condition for these matrices to be regular is that  $n > p$ , if  $p$  denotes the number of considered variables and  $n$  the number of experimental units. Unfortunately, in a microarray data set, the number of variables (gene expressions) is usually greater than the number of units (samples) and the corresponding matrix is singular. Note that similar situations can be found in other fields, for instance chemometrics, with recent developments in spectrometry (Mullen and van Stokkum, 2007). A solution for overcoming this problem is to introduce a regularization step in data processing and to perform a regularized CCA.

The first article in statistics including the idea of regularization (but not the word) is the well known paper by Hoerl and Kennard (1970) defining ridge regression (actually, the concept was first introduced by Tihonov; see, for instance, Tihonov, 1963). The first one defining a so called regularized method is probably the paper by Friedman (1989) defining regularized discriminant analysis. This method was recently applied by Guo *and others* (2007) in the context of microarray data. More generally, reviews concerning regularization in statistics can be found in Hastie *and others* (1995) and in Bickel and Li (2006). Regularized CCA is mentioned in the former (not in the latter), but it is not developed as an applicable method; in particular, CCA is not seen as a generalization of discriminant analysis, that it actually is. In the context of CCA, the idea of regularization was first proposed by Vinod (1976), then developed by Leurgans *and others* (1993); it is also considered in Ramsay and Silverman (2005). However, no statistical software nor application in the context of high throughput biology were available until recently (González, Déjean, Martin and Baccini, CCA: an R package to extend canonical correlation analysis, submitted). Other developments of CCA such as kernel CCA (Yamanishi *and others*, 2004) are not considered here.

In section 2 of the article, the principle of classical CCA is recalled and the associated plots are detailed. Regularized Canonical Correlation Analysis (RCCA) is developed in section 3 with a focus on the estimation of regularization parameters through a cross-validation procedure. The two above mentioned examples are then analysed using RCCA in section 4. In order to highlight the relevancy of the results provided by RCCA biological and chemical considerations complete the interpretation of the statistical analysis. Finally, the last section is devoted to a short discussion.

## 2 Canonical Correlation Analysis (CCA)

The general objective of CCA is to explore linear relationships between two sets of quantitative variables observed on the same experimental units. This is achieved by finding the largest correlation between a linear combination of the variables in the first set and a linear combination of the variables in the second set.

## 2.1 Notation

Let us consider two matrices  $X$  and  $Y$  of order  $n \times p$  and  $n \times q$  respectively. The columns of  $X$  and  $Y$  correspond to variables and the rows correspond to experimental units. The  $j^{th}$  column of the matrix  $X$  is denoted by  $X^j$ , likewise the  $k^{th}$  column of  $Y$  is denoted by  $Y^k$ . Without loss of generality it will be assumed that the columns of  $X$  and  $Y$  are centered and of unit variance. Moreover, it is assumed that  $p \leq q$  (in other words, the group which contains the least variables is denoted by  $X$ ). We denote by  $S_{xx}$  and  $S_{yy}$  the sample correlation matrices for variable sets  $X$  and  $Y$  respectively, and by  $S_{xy} = S'_{yx}$  the sample cross-correlation matrix between  $X$  and  $Y$ . The notation  $A'$  means the transpose of a vector or a matrix  $A$  and  $I_r$  the identity matrix of order  $r$ .

## 2.2 Principle of the classical CCA

Classical CCA assumes that  $p \leq n$ ,  $q \leq n$  and that matrices  $X$  and  $Y$  are of full column rank  $p$  and  $q$  respectively. In the following, the principle of CCA is presented as a problem solved through an iterative algorithm.

The first stage of CCA consists in finding two vectors  $a^1 = (a_1^1, \dots, a_p^1)'$  and  $b^1 = (b_1^1, \dots, b_q^1)'$  that maximize the correlation between the linear combinations

$$U^1 = Xa^1 = a_1^1X^1 + a_2^1X^2 + \dots + a_p^1X^p$$

and

$$V^1 = Yb^1 = b_1^1Y^1 + b_2^1Y^2 + \dots + b_q^1Y^q,$$

assuming that vectors  $a^1$  and  $b^1$  are normalized so that  $\text{var}(U^1) = \text{var}(V^1) = 1$ .

In other words, the problem consists in solving

$$\rho_1 = \text{cor}(U^1, V^1) = \max_{a,b} \text{cor}(Xa, Yb),$$

subject to the constraint  $\text{var}(Xa) = \text{var}(Yb) = 1$ .

The resulting variables  $U^1$  and  $V^1$  are called the first canonical variates and  $\rho_1$  is known as the first canonical correlation.

Higher order canonical variates and canonical correlations can be found as a stepwise problem. For  $s = 1, \dots, p$ , we can find positive correlations  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_p$  with corresponding vectors  $(a^1, b^1), \dots, (a^p, b^p)$ , successively by maximizing

$$\rho_s = \text{cor}(U^s, V^s) = \max_{a^s, b^s} \text{cor}(Xa^s, Yb^s) \quad \text{subject to} \quad \text{var}(Xa^s) = \text{var}(Yb^s) = 1,$$

under the additional restriction  $\text{cor}(U^s, U^t) = \text{cor}(V^s, V^t) = 0$  for  $1 \leq t < s \leq p$ .

From a geometrical point of view, let us define  $P_x = X(X'X)^{-1}X' = \frac{1}{n}XS_{xx}^{-1}X'$  and  $P_y = Y(Y'Y)^{-1}Y' = \frac{1}{n}YS_{yy}^{-1}Y'$  the orthogonal projectors onto the linear spans of the columns of  $X$  and  $Y$  respectively. It is well known that (Mardia *and others*, 1979):

- canonical correlations  $\rho_s$  are the positive square roots of the eigenvalues  $\lambda_s$  of  $P_x P_y$  (which are the same as those of  $P_y P_x$ ):  $\rho_s = \sqrt{\lambda_s}$ ;
- vectors  $U^1, \dots, U^p$  are the standardized eigenvectors corresponding to the decreasing eigenvalues  $\lambda_1 \geq \dots \geq \lambda_p$  of  $P_x P_y$ ;
- vectors  $V^1, \dots, V^p$  are the standardized eigenvectors corresponding to the same decreasing eigenvalues of  $P_y P_x$ .

## 2.3 Graphical representations

As in PCA, two kinds of graphical representations can be displayed to visualize the results of CCA: scatter plots for the initial variables  $X^j$  and  $Y^k$  and scatter plots for experimental units. If  $d$  ( $1 \leq d \leq p$ ) is the selected dimension for the results of CCA (see below for the choice of  $d$ ), graphical representations can be drawn for every pair  $(s, t)$  of axes such that  $1 \leq s < t \leq d$ . For a given pair  $(s, t)$ , variables plots and units plots can be considered with respect either to  $U^s$  and  $U^t$  or to  $V^s$  and  $V^t$ . If the canonical correlations are close to one, then the graphical representations on the axes defined by  $(U^s, U^t)$  and  $(V^s, V^t)$  are similar. In the sequel of this article, graphical representation are described with respect to  $U^s$  and  $U^t$ .

### Representations of variables

The variables plot is of interest because it allows to discern the structure of correlation between the two sets of variables  $X$  and  $Y$ . Coordinates of variables  $X^j$  and  $Y^k$  on the axis defined by  $U^s$  are Pearson correlation coefficients between these variables and  $U^s$ . As variables  $X^j$  and  $Y^k$  are assumed to be of unit variance, their projections onto the plane defined by  $(U^s, U^t)$  are inside a circle of radius 1 centered at the origin, called the correlation circle as in PCA. On this graphic, we suggest to plot two circumferences corresponding to the radius 0.5 and 1 to reveal the most salient patterns in the ring defined between these two circumferences. Variables with a strong relation are projected in the same direction from the origin. The greater the distance from the origin the stronger the relation. The interpretation is similar to the one of “correlation loadings” plot provided by the `pls` package in the context of PLS regression (Mevik and Wehrens, 2007).

### Representation of units

In some cases, the graphical representation of units can be useful to complete the interpretation of correlations between variables. The coordinate of the  $i^{th}$  unit on the axis defined by  $U^s$  is  $U_i^s$  (the  $i^{th}$  coordinate of the  $s^{th}$  canonical variate).

The relationships between the two plots (variables and units) drawn on the matching axes can reveal interesting associations between variables and units.

### Choosing the dimension

Like in PCA, it is advocated to choose a small value for dimension  $d$  ( $1 \leq d \leq p$ ). In practice, this value is very often 2, 3 or 4. Note that small canonical correlations are not relevant: they do not express linear relationships between columns of variables  $X$  and  $Y$  and can be neglected.

For great values of  $p$ , we suggest an empirical approach for choosing the dimension based on the joint examination of two graphical representations: the scree graph of canonical correlation and the scatter plots of variables. The scree graph is the plot of canonical correlations versus the dimension; a clear gap between two successive values suggest to select for  $d$  the rank of the greatest one. On the other hand, for the first values of  $s$  ( $1 \leq s \leq p$ ), we consider the scatter plot of variables according to axes  $(U^s, U^{s+1})$  and we neglect axes  $s$  such that almost all points representing variables  $X^j$  and  $Y^k$  are within the circle of radius 0.5 (that is almost all correlations between variables  $X^j$  or  $Y^k$  and canonical variates  $U^s$  are less than 0.5).

## 2.4 Evaluating relevance

In order to find canonical correlations and canonical variates, we need to compute the inverses of matrices  $S_{XX}$  and  $S_{YY}$ . However, when the number  $p$  of columns of the matrix  $X$  is greater than the number  $n$  of available units, the sample correlation matrix  $S_{XX}$  is singular. Similarly,  $S_{YY}$  is singular as soon as  $q > n$ . Moreover, even if  $p \leq n$ , when some variables  $X^j$  are highly correlated,  $S_{XX}$  tends to be ill-conditioned and its inverse unreliable, and likewise for  $S_{YY}$  (Friedman, 1989; Hastie *and others*, 1995). This phenomenon also occurs when ratio  $p/n$  (or  $q/n$ ) is less than one but not negligible. For these reasons, a standard condition to perform CCA is that  $n > p + q$  (under reasonable conditions, the sample correlation matrix of the concatenated variables  $[X, Y]$  is non-singular; see Eaton and Perlman,

1973). Since in gene expression data  $p + q$  is often greater than  $n$ , classical CCA cannot be directly performed and some form of ridge-type regularization must be considered.

## 2.5 Why is regularization essential?

We addressed this question through a simulated example. We have simulated  $n$  units  $Z_i = [X_i \ Y_i]$  distributed according to a  $(p+q)$ -dimensional normal distribution with zero mean and correlation matrix  $\Sigma_Z$  defined by

$$\Sigma_Z = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$$

where  $\Sigma_{XX} = I_p$ ,  $\Sigma_{YY} = I_q$  and  $\Sigma'_{YX} = \Sigma_{XY} = [D_p \ 0_{p,q-p}]$ , with  $D_p = \text{diag}(0.9, 0.7, 0, \dots, 0)$ , of order  $p \times p$ , and  $0_{p,q-p}$  is the null matrix of order  $p \times (q-p)$ . Then the variables  $X^1$  and  $Y^1$  have a strong relationship on the first dimension and  $X^2$  and  $Y^2$  on the second one. The theoretical canonical correlations for  $X$  and  $Y$  are  $\rho_1 = 0.9$ ,  $\rho_2 = 0.7$  and  $\rho_3 = \dots = \rho_p = 0$ .

The simulation was performed for  $n = 50$ ,  $p = 10$  and  $q = 10, 25, 39$ . The Figure 1 shows the results obtained by CCA in each case. We can see that when  $n/(p+q)$  is close to one the first canonical correlations tend to one and the true linear relationship between  $X$  and  $Y$  is lost. This example highlights that when  $n$  is small compared to  $p$  and  $q$  with  $n > p+q$ , first dimensions in CCA are not determined by the true underlying linear relationship between  $X$  and  $Y$ .

## 3 Regularized CCA (RCCA)

### 3.1 General principle

The principle of ridge regression (Hoerl and Kennard, 1970) were extended to CCA by Vinod (1976), then by Leurgans *and others* (1993). This principle involves regularization of the correlation matrices by adding a multiple of the corresponding identity matrix. In other words, let us consider

$$S_{XX}(\lambda_1) = S_{XX} + \lambda_1 I_p$$

and

$$S_{YY}(\lambda_2) = S_{YY} + \lambda_2 I_q,$$

where  $\lambda_1$  and  $\lambda_2$  are non-negative numbers such as  $S_{XX}(\lambda_1)$  and  $S_{YY}(\lambda_2)$  are regular matrices. Regularized CCA consists in performing classical CCA substituting  $S_{XX}$  and  $S_{YY}$  respectively with  $S_{XX}(\lambda_1)$  and  $S_{YY}(\lambda_2)$ .

This step implies to deal with the core problem of tuning the regularization parameters  $\lambda_1$  and  $\lambda_2$ .

### 3.2 Tuning parameters by $M$ -fold cross-validation

We propose to extend the leave-one-out procedure suggested by Leurgans *and others* (1993) and to select regularization parameters  $\lambda_1$  and  $\lambda_2$  by  $M$ -fold cross-validation ( $2 \leq M \leq n$ ).

Let us denote  $\lambda = (\lambda_1, \lambda_2)$  and consider  $\mathcal{I} : \{1, \dots, i, \dots, n\} \mapsto \{1, \dots, M\}$ , an indexing function that indicates the set  $\mathcal{I}_m$  to which unit  $i$  is allocated by randomization; sets  $\mathcal{I}_m = \{i : \mathcal{I}(i) = m, m = 1, \dots, M\}$  are roughly equal-sized. Let us denote  $X_{\mathcal{I}_m} = \{X_i : i \in \mathcal{I}_m\}$  and  $Y_{\mathcal{I}_m} = \{Y_i : i \in \mathcal{I}_m\}$  (see Fig. 2 for a schematic representation). For a given value of  $\lambda$ , let us denote by  $\rho_\lambda^{(-\mathcal{I}_m)}$  the first canonical correlation of CCA computed from the sample with data  $X_{\mathcal{I}_m}$  and  $Y_{\mathcal{I}_m}$  removed.

Let  $a_\lambda^{(-\mathcal{I}_m)}$  and  $b_\lambda^{(-\mathcal{I}_m)}$  be the corresponding vectors defining the first canonical variates. We do this for  $m = 1, \dots, M$  and obtain  $M$  pairs of vectors  $(a_\lambda^{(-\mathcal{I}_1)}, b_\lambda^{(-\mathcal{I}_1)})$ ,  $\dots$ ,  $(a_\lambda^{(-\mathcal{I}_M)}, b_\lambda^{(-\mathcal{I}_M)})$ . The  $M$ -fold

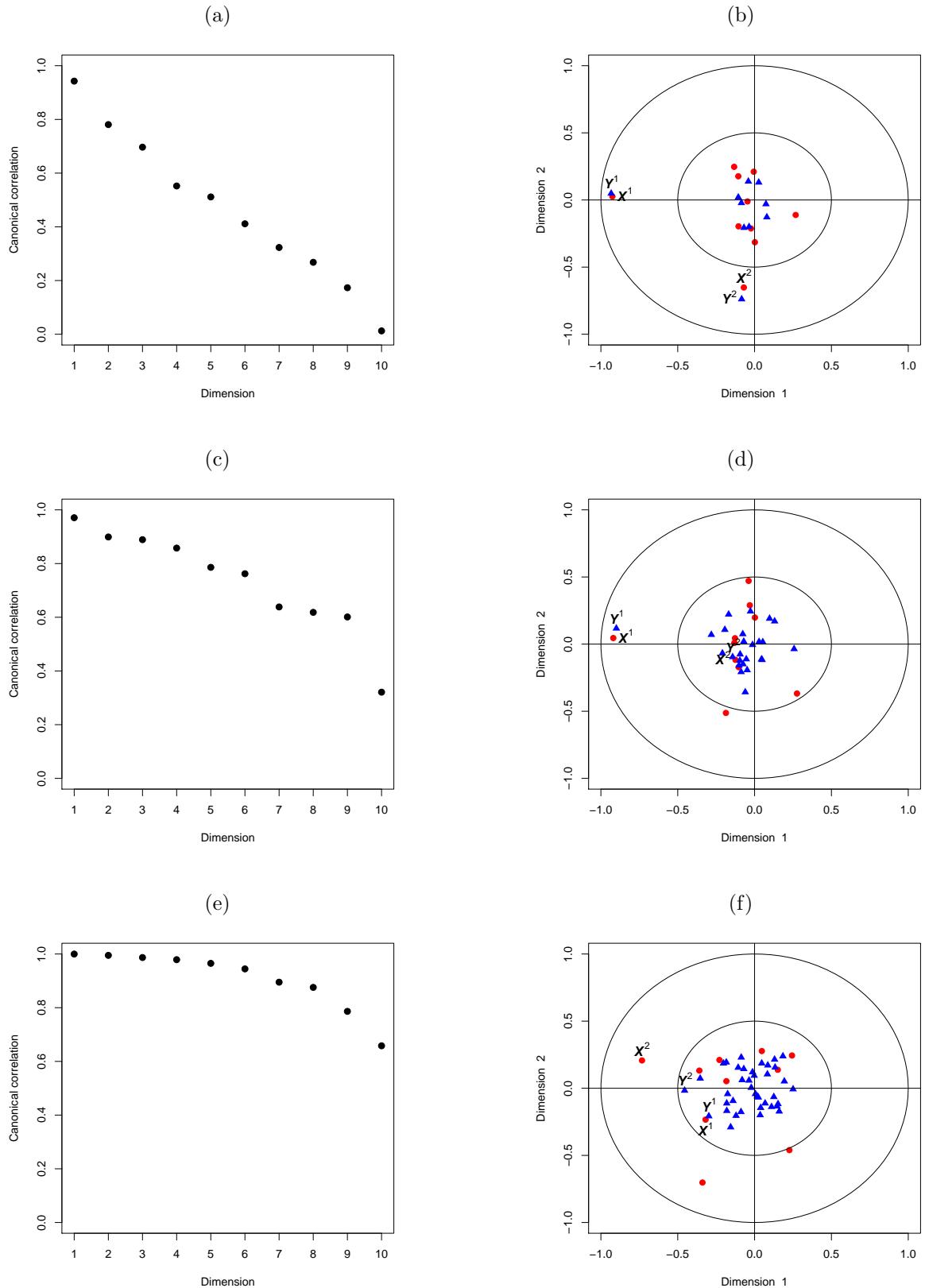


Figure 1: Effect of the ratio  $n/(p+q)$  on the scree graph and variables plots. The panels (a-b), (c-d) and (e-f) correspond to ratio 50/20, 50/35 and 50/49 respectively.  $X$  and  $Y$  variables are represented respectively as points and triangles.

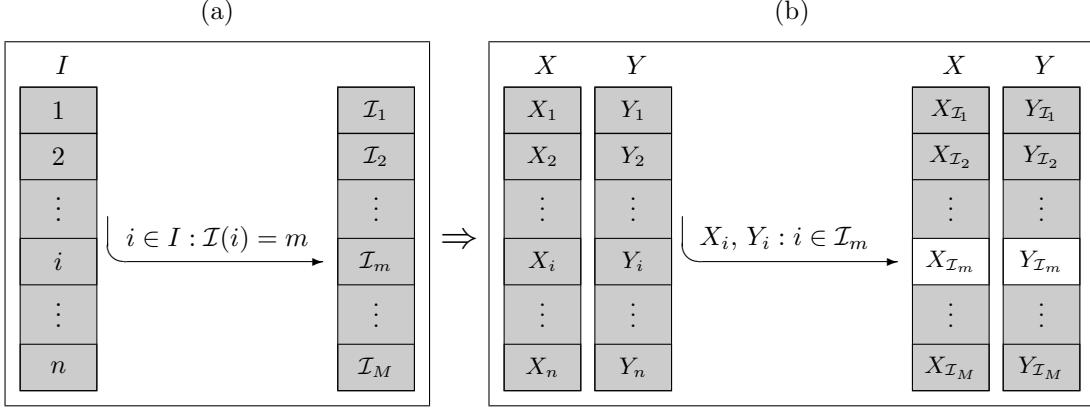


Figure 2: Schematic representation of the  $M$ -fold splitting in both data sets. (a) The units indexes are allocated by randomization in the sets  $\mathcal{I}_1, \dots, \mathcal{I}_M$ . (b) Units in  $X$  and  $Y$  are reordered and regrouped according to  $\mathcal{I}_1, \dots, \mathcal{I}_M$

cross-validation score for  $\lambda = (\lambda_1, \lambda_2)$  is then defined by

$$CV(\lambda_1, \lambda_2) = \text{cor} \left( \left\{ X_{\mathcal{I}_m} a_\lambda^{(-\mathcal{I}_m)} \right\}_{m=1}^M, \left\{ Y_{\mathcal{I}_m} b_\lambda^{(-\mathcal{I}_m)} \right\}_{m=1}^M \right).$$

Then we choose the values of  $\lambda_1$  and  $\lambda_2$  that maximize this correlation:

$$\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2) = \arg \max_{\lambda_1, \lambda_2} CV(\lambda_1, \lambda_2).$$

For  $M = n$ , we obtain the leave-one-out cross-validation. Nevertheless, with  $M = n$ , the computational burden can be considerable on certain data sets, requiring  $n$  applications of the algorithm for each value of  $\lambda$ . On the other hand, a relevant choice of  $M$  ( $M < n$ ) can make faster the algorithm of cross-validation. In practice, it is usual to choose  $M = 5$  or  $M = 10$ . Note that  $M$ -fold cross-validation also implies variance reduction for  $\hat{\lambda}$  in relation to leave-one-out method.

It must be noticed that  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$  are chosen with respect to the first canonical variates and are then fixed for higher order canonical variates.

There are two tuning parameters in the regularized CCA, so the cross-validation is performed on a two-dimensional surface. Directly search for a maximum on the two-dimensional surface ensure to obtain the optimal value for  $\lambda$  but is computationally burdensome. An alternative consists in picking a relatively small grid of “reasonable” values for  $\lambda_1$  and  $\lambda_2$ , to evaluate the cross-validation score at each point of the grid, and then to choose the value of  $\lambda = (\lambda_1, \lambda_2)$  that gives the largest  $CV$ -score. The experience can guide the user to refine the discretization grid, but one can also perform the procedure in a recursive way: first, use a relatively coarse grid and locate an area where the optimal value for  $\lambda_1$  and  $\lambda_2$  could be reached, and then determine a more accurate grid within this area.

### 3.3 Statistical software

To perform RCCA, we used the R package **CCA** (presented in González, Déjean, Martin and Baccini, **CCA**: an R package to extend canonical correlation analysis, submitted). The package provides a set of functions to develop numerical and graphical outputs. It includes the regularized extension of CCA and enables the handling of missing values. The R package **CCA** is freely available from the Comprehensive R Archive Network (CRAN) at [cran.r-project.org](http://cran.r-project.org) and at [www.lps.ups-tlse.fr/CCA](http://www.lps.ups-tlse.fr/CCA).

## 4 RCCA on real data

In this section, we apply RCCA on two gene expression data sets. To highlight the interest of the proposed method, we performed it firstly on a public data set already analysed, then on our own published data. Note that external considerations have been used to emphasize the relevancy of the results obtained on both data sets.

## 4.1 Multidrug resistance data

### Data set

The data come from a pharmacogenomic study (Szakács *and others*, 2004) in which two kinds of measurements acquired on the NCI-60 panel of cancer cell lines are considered:

- the expression of the 48 human ABC transporters measured by real-time quantitative RT-PCR for each cell line;
- the activity of 1429 drugs expressed as  $GI_{50}$  which corresponds to the concentration at which the drug induces 50% inhibition of cellular growth for the cell line tested.

The NCI-60 panel includes cell lines derived from cancers of colorectal (7 cell lines), renal (8), ovarian (6), breast (8), prostate (2), lung (9) and central nervous system origin (6), as well as leukemias (6) and melanomas (8). It was set up by the Developmental Therapeutics Program of the National Cancer Institute ([dtp.nci.nih.gov](http://dtp.nci.nih.gov)) to screen the toxicity of chemical compound repositories. The rationale behind the search for correlations (positive or negative) between ABC transporters and drug activities was reviewed in Szakács *and others* (2006). Briefly, ABC transporters are proteins typically involved in the active translocation of substrates across cell membranes. Cancer cells overexpressing certain ABC transporters can become resistant to chemotherapy because the transporters mediate the efflux of the drug out of the cell thus blocking the access of the drug to its intracellular target. Because of a wide substrate specificity, the overexpression of some transporters such as P-glycoprotein (Pgp / MDR1 / ABCB1) in tumors can confer multidrug resistance, thus strongly limiting the spectrum of anticancer drug that can be beneficially administered to the patients. If a given compound is the substrate of an ABC transporter, one could thus expect a strong negative correlation between the expression of the ABC transporter and the efficacy of the drug. Conversely, a high expression of Pgp can also confer hypersensitivity to certain compounds through mechanisms that are not yet fully understood (reviewed in Szakács *and others* 2006). Such a situation will translate into a strong positive correlation between the expression of the ABC transporter and the efficacy of the drug. As stated by the authors of this study, they “*explored the relationship between ABC transporter expression levels and sensitivity to drugs or drug candidates, asking which of the transporters confer resistance or sensitivity to various classes of agents*”. In this context, RCCA appears as an appropriate tool highlighting the strongest correlations (positive or negative) among the two groups of variables: ABC transporter expression levels and drug efficacy. This approach is an alternative to the one used in Szakács *and others* (2004) which consists in testing by bootstrap (through the construction of a confidence interval) the significance of correlation coefficients calculated between every gene-drug pairs. In the following, we will refer to this method as BCI: Bootstrap Confidence Interval for correlation coefficient. Note that in the context of this specific study and when the problem is presented as previously stated, PLS may also provide interesting results through the explanation of drug activities by ABC transporters expression profiles. However, it is likely that the expression profile of many other genes (the genes encoding the drug targets or proteins involved in drug metabolism for example) may explain the drug activities on this panel of cell lines. Thus, we did not consider PLS on this example but we would recommend its evaluation when analyzing genome-wide gene expression profiles along with drug activities.

### Statistical data analysis

In this example, the matrix  $X$  ( $60 \times 48$ ) contains the expression levels of the 48 ABC transporters while the matrix  $Y$  ( $60 \times 1429$ ) contains the function of the activity of 1429 drugs among the 60 cell lines considered as the experimental units. The database  $X$  and  $Y$  are respectively available in Supplementary Table S1 at [download.cancercell.org/supplementarydata/cell/6/2/129/DC1](http://download.cancercell.org/supplementarydata/cell/6/2/129/DC1) and at [discover.nci.nih.gov/datasetsNature2000.jsp](http://discover.nci.nih.gov/datasetsNature2000.jsp).

RCCA was applied to these data sets and tuning parameters selection by five-fold cross-validation was carried out on a regular grid of size  $31 \times 31$ . A preliminary study enables to focus on the region defined by  $2.5 \leq \lambda_1 \leq 4$ ,  $3.5 \leq \lambda_2 \leq 5$ . For the obtained parameters  $\lambda_1 = 3.05$  and  $\lambda_2 = 3.75$  (Fig. 3(a)), the first ten canonical correlations were 0.81, 0.79, 0.75, 0.72, 0.67, 0.66, 0.63, 0.58, 0.57, 0.56. For graphical representation and analysis of the results of the RCCA, we decided to keep the first four dimensions. Following the procedure suggested in section 2.3, we observed a gap between the fourth and the fifth

canonical correlation and between the seventh and the eighth canonical correlation (Fig. 3(b)), but the scatter plot of variables for axes  $(U^s, U^{s+1})$  suggests to neglect axes (dimensions) for  $s \geq 5$  (results not shown).

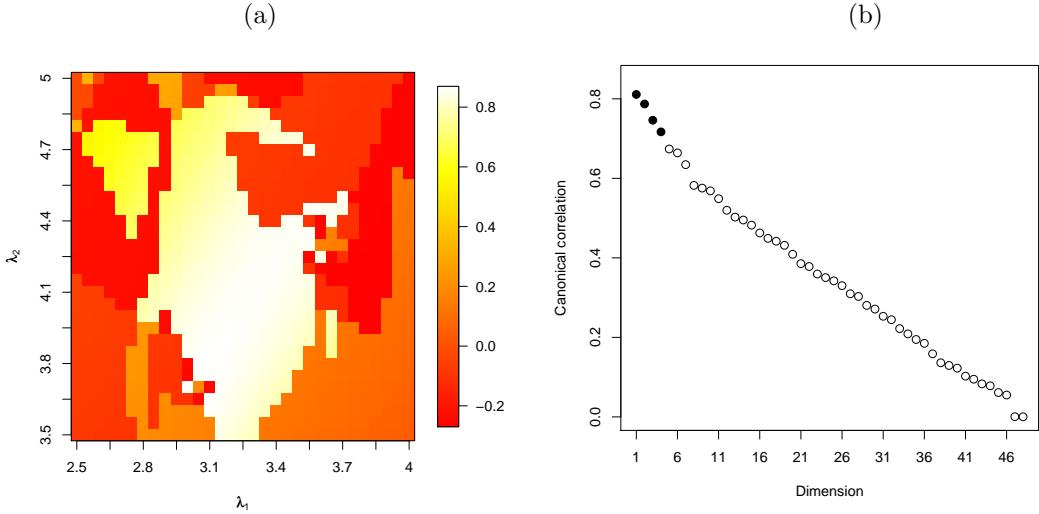


Figure 3: (a) Cross-validation map for  $\lambda_1$  and  $\lambda_2$ . (b) Scree graph for the 48 canonical correlations. Dimensions corresponding to the black thick points were chosen for graphics representation and analysis of the results of the RCCA.

To identify gene-drug pairs showing relevant associations, we calculated a similarity measure between  $X$  and  $Y$  variables in a pair-wise manner: the scalar product value between every pairs of vectors in dimension 4 representing variables  $X$  and  $Y$  on the axis defined by  $(U^1, \dots, U^4)$  (see section 2.3). Variable pairs with a high value (in absolute value) are considered as relevant. Then, given a threshold corresponding to the radius of the small circle in the variables plot, selected variable pairs with a value above this threshold verifies two conditions:

- to be outside the hypersphere (in dimension 4) of radius 0.5;
- to be projected on the canonical variates plane either in the same direction (positive correlation) or in an opposite direction (negative correlation) from the origin.

By changing the threshold (radius), one can tune the relevance of the associations to include or exclude relationships. For analysis of the results we choose the radius equal to 0.5. Gene-drug pairs showing relevant associations are summarized in Tables 1 and 2.

Results of RCCA (Tables 1 and 2 and Figures 4 and 5) reveal some interesting patterns. The most salient feature is the negative correlation between the ABCB1 gene and a group of compounds (see Table 1) represented by dark blue triangles (Fig. 4(a-b-c)). This main observation is consistent with a predominant role of P-glycoprotein, the product of the ABCB1 (or MDR1) gene, in multidrug resistance (for a review see Szakács *and others*, 2006) and is further analysed in this article. We also observe negative correlations between the ABCC3 gene and the compounds represented by green triangles (see Table 1 and Fig. 4(b-d-f)) as well as between the ABCC11 gene and the compounds NSC656159, NSC371010 and NSC665488, represented by yellow triangles (Fig. 4(c-e-f)). These findings are also in general agreement with the findings that these two ABC transporters are involved in resistance to certain anti-cancer drugs (Kool *and others*, 1999; Guo *and others*, 2003). More marginally, there were some evidence of negative correlation between the ABCC2 gene and the compound NSC621812 (Fig. 4(a-d-e)) and between the ABCA3 gene and the compounds NSC403170 and NSC374980 (Fig. 4(a-b-c)).

Positive correlation between anticancer drugs and ABC expression for some of the compounds was also found. The Figure 4(a-b-c) revealed an association between the gene ABCB1 and the NSC693871 compound. A significant observation is the association between the genes ABCC2, ABCA9 and ABCB5

Lab	Gene	Col	Drug (NSC N°)									
C2	ABCC2	[Color]	621812									
A3	ABCA3	[Color]	403170 374980									
B1	ABCB1	[Color]	353076	259968	328426	359449	682066	265450	634791	270693		
		[Color]	646946	337766	640085	645301	163088	38270	604574	630678		
		[Color]	80469	676864	155694	645305	341651	638788	645302	630176		
		[Color]	365360	351710	363997	269756	201241	125973	266763	666608		
C3	ABCC3	[Color]	658450	670762	654236	641240	637693	641594	643915	644751		
		[Color]	654206	690433	641820	641245	642049	627991	628939	670289		
		[Color]	623925	645991								
C11	ABCC11	[Color]	656159	371010	665488							

Table 1: List of gene-drug pairs showing relevant negative linear correlation. Genes and drugs on each row are represented in the Figure 4 according to their label (Lab) and color (Col) respectively.

Lab	Gene	Col	Drug (NSC N°)				
B1	ABCB1	[Color]	693871				
C2	ABCC2	[Color]	627787 602617 690434 690432 644902				
A9	ABCA9	[Color]	627787	602617	690434	690432	644902
		[Color]	670766	670762	637729		
B5	ABCB5	[Color]	627787	602617	690434	690432	644902
		[Color]	670766	670762	637729		
		[Color]	642061	628507			

Table 2: List of gene-drug pairs showing relevant positive linear correlation. Genes and drugs on each row are represented in the Figure 4 according to their label (Lab) and color (Col) respectively.

with the compounds NSC627787, NSC602617, NSC690434, NSC690432 and NSC644902 (see Table 2 and Fig. 4(a-b-c)).

Comparison of the Figures 4(a-b-c-d-e) and 5(a-b-c-d-e) shows that the genes ABCC2, ABCA9 and ABCB5 are highly expressed in the melanoma cell lines, while the ABCB1 gene is highly expressed in cell lines of renal origin. This latter observation is consistent with the rhodamine efflux pattern presented in Lee *and others* (1994) and used to assess P-glycoprotein function in cell lines from the NCI-60 panel. Finally, comparing the Figures 4(b-c-d-e-f) and 5(b-c-d-e-f) evidences that the genes ABCC3 and ABCC11 exhibit a low expression level in leukemia cell lines.

### Biological and chemical considerations

Particular attention was paid to one ABC transporter: the P-glycoprotein (thereafter named ABCB1), which is one of the most important and widely studied members of the ABC transporter superfamily. For this particular gene we analysed only ABCB1-drug pairs displaying negative correlation, which correspond to association detecting drugs potentially transported by ABCB1. Among the 32 drugs potentially transported by ABCB1 as identified by RCCA (Table 1), 10 were also found by Szakács *and others* (2004) using the BCI criterion. Thus, using RCCA, we identified more than 50% (10 out of 18) of the drugs identified by the BCI method and added 22 new candidates to their list of 18 drugs. Additionally, concerning the 8 drugs that were identified by the BCI procedure only, note that they could also have been selected by RCCA using slightly more relaxed selection criterion (see Figure S1 of the supplementary material at the end of the paper). Interestingly, among the 10 drugs found by both RCCA and the BCI procedure, most drugs that were tested by Szakács *and others* (2004) in their follow-up study (4 drugs out of the 6 tested) or which are known substrates of ABCB1 (2 drugs: bouvardin/NSC259968 and phyllanthoside/NSC328426 out of the 3 mentioned by Szakács *and others*, 2004) were identified by RCCA. Thus 6 out of the 10 drugs which are identified by both RCCA and BCI are experimentally

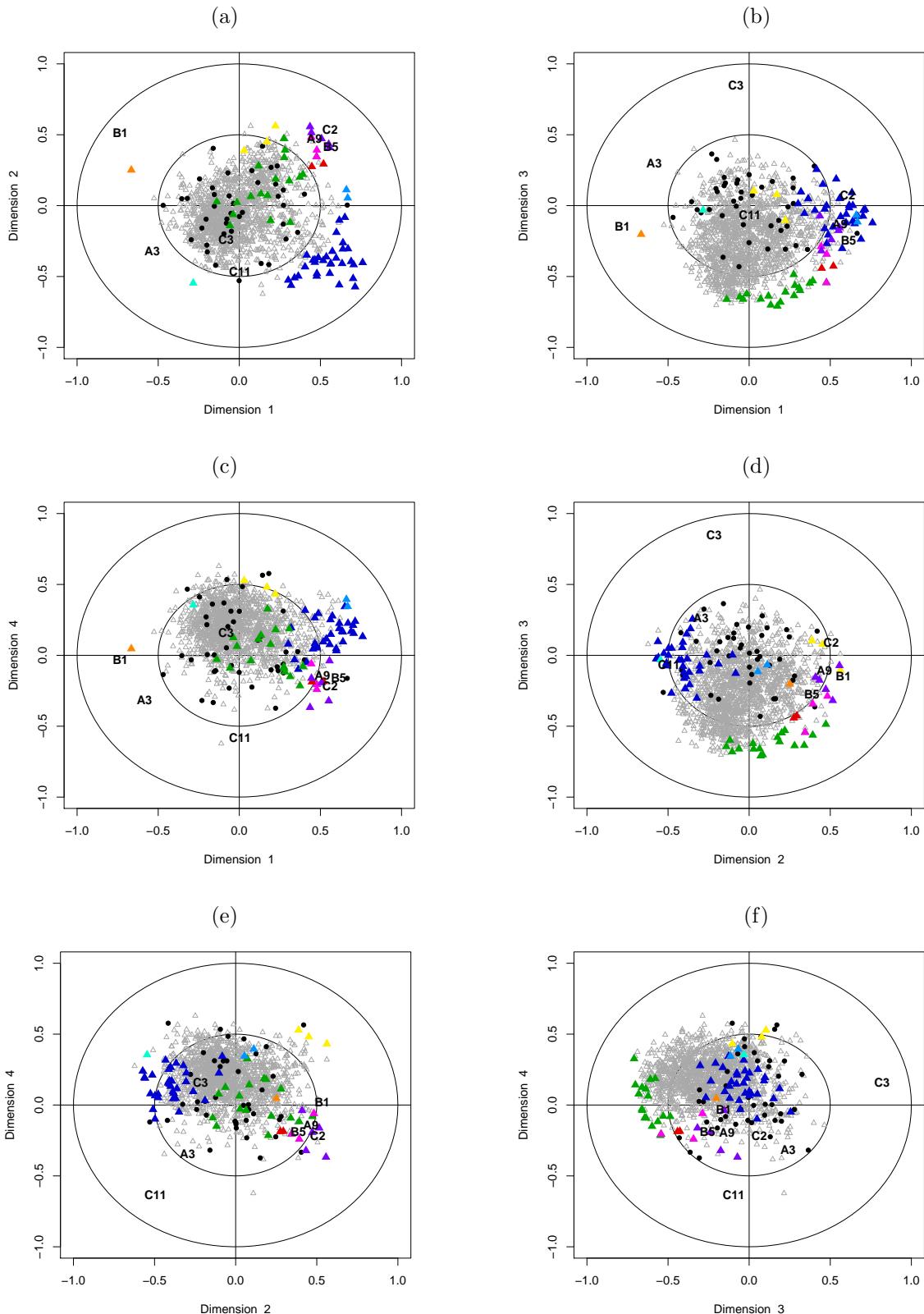


Figure 4: Variables plots for every combination of dimensions from 1 to 4. Genes and compounds are represented by black thick points and gray triangles respectively. Relevant genes and compounds are labelled and colored respectively according to the Tables 1 and 2.

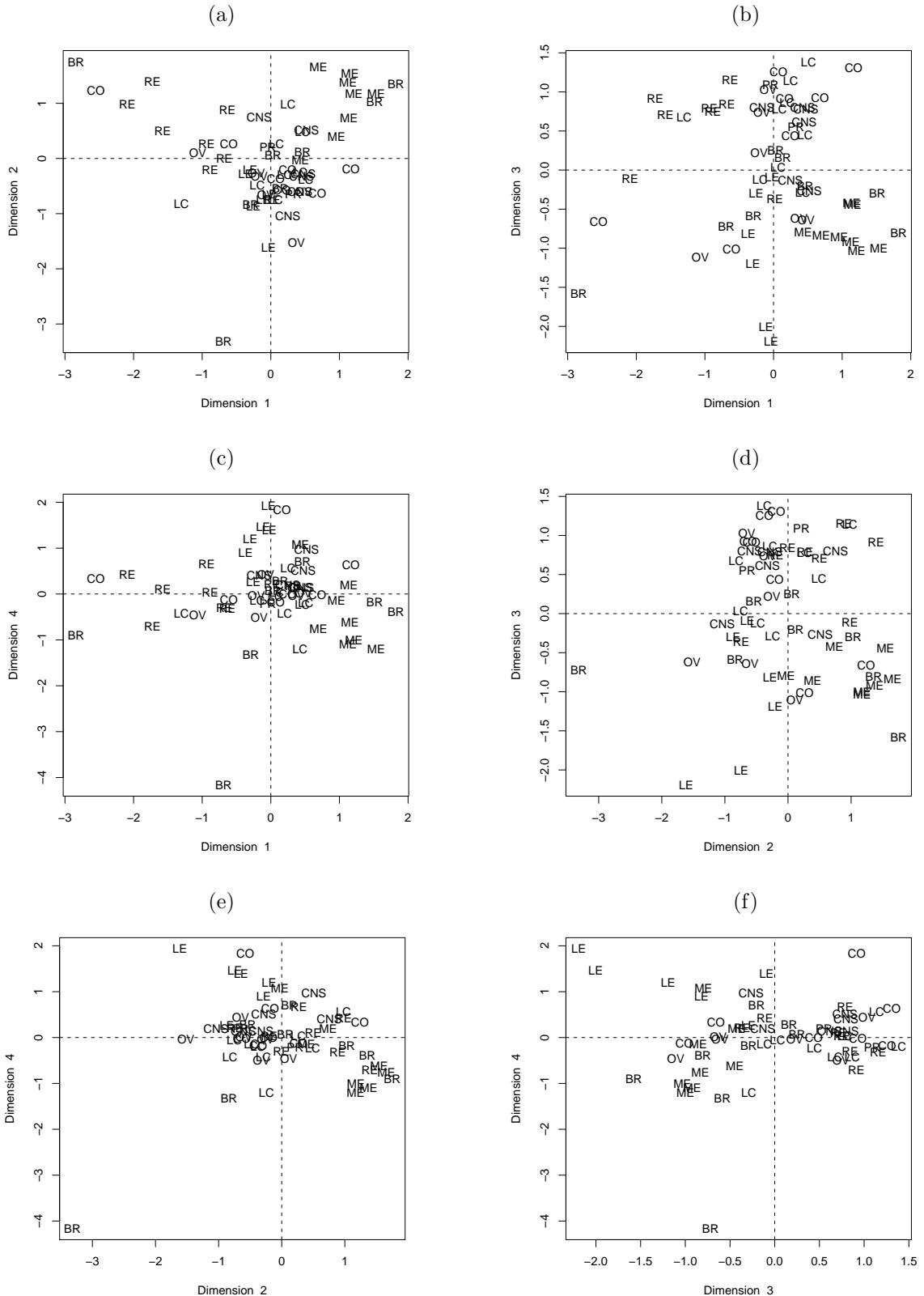


Figure 5: Units plots for the NCI-60 cancer cell lines: colorectal (CO), renal (RE), ovarian (OV), breast (BR), prostate (PR), lung cancer (LC), central nervous system origin (CNS), leukemia (LE) and melanoma (ME).

validated ABCB1 substrates. Additionally, among the 22 drugs which are uniquely identified by RCCA, we found by literature search that for 7 drugs there are experimental evidences or demonstrations that they are substrates of ABCB1. These include olivomycin (NSC38270, Lee *and others*, 1994; Scala *and others*, 1997), paclitaxel (NSC125973, Lee *and others*, 1994; Scala *and others*, 1997; Xing *and others*, 2007), berberine (NSC163088, Shitan *and others*, 2007), bisantrene (NSC337766, Lee *and others*, 1994; Zhang *and others*, 1994), depsipeptide/FK228 (NSC630176, Lee *and others*, 1994; Scala *and others*, 1997; Xiao *and others*, 2005), an ellipticinium (NSC155694, Huang *and others*, 2005) and an aniline mustard derivative (NSC640085, Lee *and others*, 1994; Scala *and others*, 1997). Altogether, 13 drugs among the 32 candidates identified by RCCA are known to be transported by ABCB1, thus demonstrating the relevance of the negative correlations identified between ABCB1 expression and the potential anticancer drugs by RCCA. Finally, among the candidate drugs identified by RCCA, many drugs are of natural origin and display relatively high molecular weights, an observation which has been previously reported in a search for new ABCB1 substrates (Lee *and others*, 1994).

In order to further evaluate the relevance of the 22 additional candidates identified by RCCA, structure similarities were analysed thanks to PubChem webtools. The PubChem Structure Clustering tool was used in order to estimate candidate drugs structure similarity according to the Tanimoto score. This score can be viewed as a similarity index (ranging from 0 to 1) calculated from the structure fingerprints ([pubchem.ncbi.nlm.nih.gov](http://pubchem.ncbi.nlm.nih.gov)). The lists of compounds identified by RCCA and by the BCI procedure were both uploaded to the PubChem website: a total of 40 NSC drugs (32 drugs identified by RCCA, plus 18 drugs identified by BCI, minus 10 duplicate drugs found by both RCCA and BCI). After Compound ID (CIDs) mapping, three drugs: NSC638788, NSC646946 and NSC652903 did not map to any CID in PubChem. The remaining 37 drugs mapped to a total of 63 CIDs, mainly due to redundancy in the PubChem database with sometimes discrete variations between the PubChem entries corresponding to a unique NSC drug. The dendrogram resulting from the structure clustering tool for the 63 PubChem CIDs was retrieved (Fig. 6). Cutting this tree at a Tanimoto score of 0.55 provides 6 clusters, two of which being composed of a single drug, both identified by RCCA only and one of these (bisantrene/NSC337766) being a known ABCB1 substrate (Lee *and others*, 1994; Zhang *and others*, 1994). Candidates identified by RCCA are present in all four remaining clusters and, except for cluster 1, are associated with candidates identified by the BCI procedure, suggesting a certain level of structural similarity between the drugs identified by both methods.

The first cluster (yellow) consists in compact structures with low steric hindrance, non charged, with aliphatic linear or circular chain or with few aromatic cycle. This cluster contains three drugs, all identified uniquely by RCCA, with at least one proven ABCB1 substrate (depsipeptide/NSC630176, Lee *and others*, 1994; Scala *and others*, 1997; Xiao *and others*, 2005). Drugs from the second cluster (orange) share structural features like amphipathic character, large and compact polyaromatic backbone, symmetric arrangements, and are often positively charged. As previously observed (Szakács *and others*, 2004), many of these structural features are shared with well-known ABCB1 substrates. The third cluster of molecules (green), consists in polyaromatic structures with aza cycle derivative, compact with low steric hindrance, and positively charged via a quaternary ammonium. Inside this third cluster, we also observed the presence of a subgroup of 4 ellipticine analogs (see Figure 6) which present very similar structures and were all identified by RCCA. NSC359449 which belongs to this cluster and was identified by both RCCA and BCI have been previously shown to be an ABCB1 substrate (Scala and others, 1997; Szakacs and others, 2004; Huang and others, 2005). RCCA further identifies three closely related ellipticine analogs (NSC351710, NSC604574 and NSC353076, the latter being also identified by BCI) as potential ABCB1 substrates (Figure 6). Note that RCCA also identified a fifth ellipticine analog (NSC155694) which is a reported ABCB1 substrate (Huang *and others*, 2005) presenting lower Tanimoto scores with the other four ellipticines (cluster 3, Figure 6). For the fourth cluster of candidate drugs (blue), identical structural features are also shared between RCCA and BCI selections. Amphipathic character, a majority of aliphatic cycle with a variable proportion of oxo cycle derivative and this time important steric hindrance are found for this group of drugs. One interesting observation is the selection by RCCA of Paclitaxel (see Figure 6) which is a clinically used anti-cancer drug and a known ABCB1 substrate. As illustrated in Figure 6, RCCA further selected NSC666608 which presents a similar structure compared to Paclitaxel.

In conclusion, the high proportion of validated ABCB1 substrates among the candidate drugs identified by RCCA strongly suggests that RCCA is a relevant method to identify potential ABC transporter substrates using these data sets. The analysis of structure similarities and candidate drugs characteristics

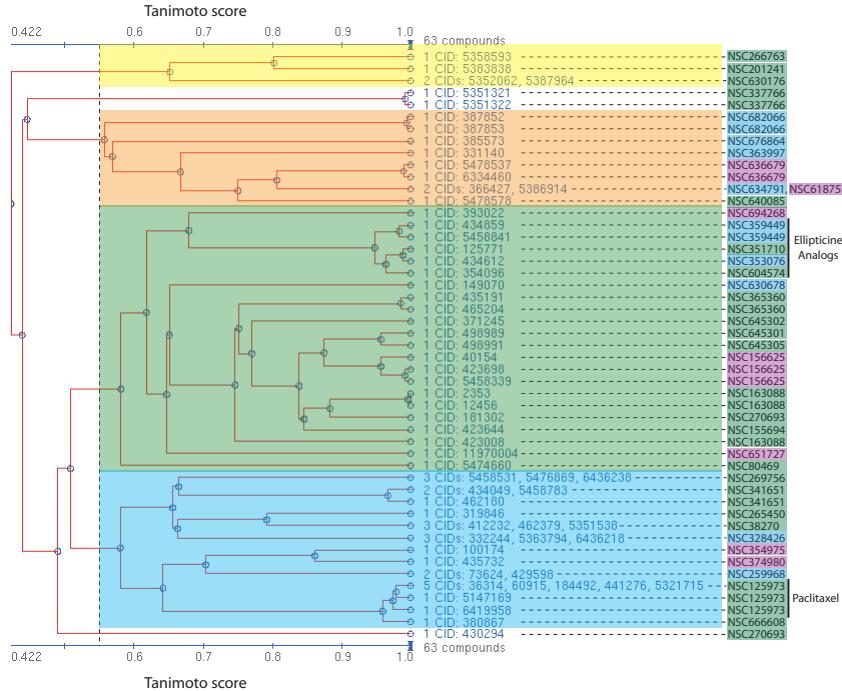


Figure 6: Structure clustering tree for 37 drugs selected by RCCA and BCI. Drugs coming from RCCA selection are colored in green, from BCI in violet and from both methods in blue. Note that the 37 NSC drugs mapped to 63 PubChem CIDs. Cluster 1 is colored in yellow, cluster 2 in orange, cluster 3 in green, and cluster 4 in blue. Isolated structures are not colored.

(high molecular weights and natural origin in particular) also illustrate the relevance of the candidates identified by RCCA. Follow-up studies are now required to fully demonstrate that the new candidates identified by RCCA, and by other published methods, are true ABCB1 substrates. In particular, we envisage to perform docking studies using a refined tridimensional model of ABCB1 (Garrigues *and others*, 2002) in order to compare the molecular fitting between RCCA/BCI-selected drugs versus randomly selected drugs. Such a study may also serve as a second *in silico* screening to identify the best candidates to be experimentally studied.

## 4.2 Nutrigenomic data

### Data set

The data sets come from a nutrigenomic study in the mouse (Martin *and others*, 2007). Forty mice were studied on which two sets of variables were acquired:

- expressions of 120 genes measured in liver cells, selected (among about 30,000) as potentially relevant in the context of the nutrition study. These expressions come from a nylon macroarray with radioactive labelling;
- concentrations (in percentages) of 21 hepatic fatty acids (FA) measured by gas chromatography.

Biological units (mice) are cross-classified according to two factors (Fig. 7):

- genotype: study were done on wild-type (WT) and PPAR $\alpha$ -deficient (PPAR $\alpha$ ) mice;
- diet: oils used for experimental diet preparation were corn and colza oils (50/50) for a reference diet (ref), hydrogenated coconut oil for a saturated FA diet (coc), sunflower oil for an  $\omega$ 6 FA-rich diet (sun), linseed oil for an  $\omega$ 3 FA-rich diet (lin) and corn/colza/fish oils (42.5/42.5/15) for the fish diet.

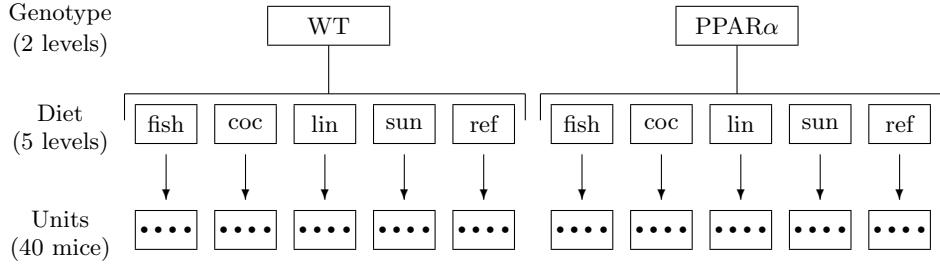


Figure 7: Experimental design of the nutrigenomic study.

In this study, it cannot be assumed that variations in one set of variables cause variations in the other one: we do not know *a priori* if gene expression changes imply fatty acid concentrations changes or inversely. Indeed, the nuclear receptor PPAR $\alpha$ , which acts as a ligand-induced transcriptional regulator, is known to be activated by various fatty acids and regulates the expression of several genes involved in fatty acid metabolism (Martin and others, 2007, and references therein). From an experimental point of view this justifies the use of PPAR $\alpha$ -deficient mice and from an analytical point of view it justifies the use of CCA instead of an explicative method such as PLS regression.

### Statistical data analysis

In this example the matrix  $X$  ( $40 \times 21$ ) contains the concentrations of 21 hepatic fatty acids and the matrix  $Y$  ( $40 \times 120$ ) contains the expressions of 120 genes. The data sets  $X$  and  $Y$  are available at [www.lsp.ups-tlse.fr/CCA](http://www.lsp.ups-tlse.fr/CCA) and into the R package CCA. RCCA was applied to these data set and regularization parameters selection was carried out by leave-one-out cross-validation (the relatively small dimension of this data set enables it). The Figure 8(a) shows the cross-validation score for the nutrigenomic data. The optimization grid of  $(\lambda_1, \lambda_2)$  values was defined on the region  $0 \leq \lambda_1 \leq 0.2$ ,  $10^{-4} \leq \lambda_2 \leq 0.2$  as a regular grid of size  $51 \times 51$ . Since the correlation matrix estimates of  $Y$  is singular, the smallest value of the parameter  $\lambda_2$  was selected just large enough to permit numerically stable inversion. Then the value of  $\lambda_1$  and  $\lambda_2$  that gives the largest  $CV$ -score are 0.064 and 0.008 respectively. For the obtained parameters  $\lambda_1$  and  $\lambda_2$ , the first ten canonical correlations were: 0.96, 0.93, 0.89, 0.84, 0.80, 0.76, 0.71, 0.69, 0.67, 0.60. To choose the dimension, we see that a clear gap does not exist among the first canonical correlations in the scree graph (Fig. 8(b)), however the variables plot revealed, for dimensions greater than 3, that almost all points representing the variables  $X$  and  $Y$  are within the circle of radius 0.5 (results not shown here). Then we chose the dimension  $d = 3$  for graphics representation and analysis of the results.

### Biological considerations

First, it should be noted that the main observations discussed in Martin and others (2007), which were extracted separately from the two datasets by both classical multidimensional tools (hierarchical clustering and PCA) and standard test procedures, are also highlighted by RCCA graphical representations. In particular, the authors observed a strong effect of the genotype as compared to dietary effects, which is emphasized by the separation of the two genotypes along the first canonical variate (Fig. 9(b-d)). We easily identified the specific accumulations of linoleic acid (C18:2 $\omega$ 6) and CAR transcript in PPAR $\alpha$ -deficient livers (Fig. 9(a-b) and (c-d)) which were further analysed (Martin and others, 2007). The constitutive lower expression of PPAR $\alpha$  target genes involved in FA catabolism in PPAR $\alpha$  deficient livers is also revealed along the first dimension (genes represented by orange thick points, Fig. 9(a-c) and Table 3). Furthermore, the accumulations of FA classes which are provided by the diets as well as the specific gene expression changes induced by the diets are correctly highlighted along the second and third canonical variates:

- accumulations of long-chain  $\omega$ 3 FA and induction of GSTpi2 and cytochromes P450 for the fish and lin diets are visualized along the second canonical variate (Fig. 9(a-b-e-f));

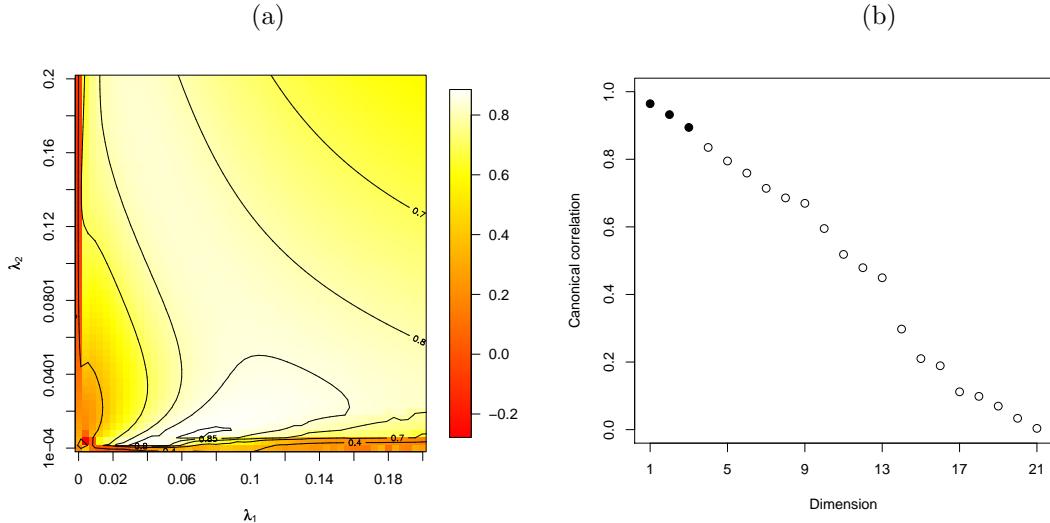


Figure 8: (a) Cross-validation map for  $\lambda_1$  and  $\lambda_2$ . Contour plots are also displayed for values equal to  $\{0 - 0.40 - 0.70 - 0.80 - 0.85 - 0.88\}$ . (b) Scree graph for the 21 canonical correlations of the nutrigenomic data. Dimensions corresponding to the black thick points were chosen for graphics representation and analysis of the results of the RCCA.

- accumulations of long-chain  $\omega 6$  FA and discrete gene expression changes (mainly represions) for the sun diet are well represented along the third canonical variates (Fig. 9(c-d-e-f));
- accumulations of monounsaturated FA and induction in wild-type mice of FA and cholesterol synthesis genes (ACC2, S14, HMGCoAred and cHMGCoAS) are highlighted, to different extents, along the three canonical variates represented (Fig. 9).

Martin *and others* (2007) also reported an altered response of PPAR $\alpha$  deficient mice to diet-induced gene expression changes. Although we still observe in the units plots (Fig. 9(b-d-f)) that the separation between the diets appears less marked for the PPAR $\alpha$  deficient samples than for their wild-type counterparts, this phenomenon seems partially masked through the combined analysis of gene expression and FA profiles. This observation is consistent with the impaired FA metabolism in PPAR $\alpha$  deficient mice which would thus tend to accumulate more directly the FA which are provided through the diets because these FA do not undergo a normal range of metabolism in these animals. Eventually, variables plots (Fig. 9(a-c-e)) illustrate well the correlations observed between FA within each class ( $\omega 3$ ,  $\omega 6$  and monounsaturated FA) and which was highlighted by the authors through PCA and hierarchical clustering (Martin *and others*, 2007, supplementary material). While most, if not all, conclusions drawn by the authors can be readily illustrated through RCCA representations, additional information can be extracted by looking at the correlations between variables of two different origins (gene expression and FA composition). For example, while the fish diet was especially enriched in docosahexaenoic acid (DHA, C22:6 $\omega 3$ ), it seems that DHA but also eicosapentaenoic acid (EPA, C20:5 $\omega 3$ ) and docosapentaenoic acid (DPA, C22:5 $\omega 3$ ) proportions are highly correlated with GSTpi2, Cyp2c29 and Cyp3a11 expression levels (Fig. 9(a-e)). Surprisingly, the proportions of palmitic acid (C16:0) are highly correlated to the expression of numerous PPAR target genes while linoleic acid (C18:2 $\omega 6$ ) proportions are anticorrelated with all these variables (Fig. 9(a-c)). Interestingly, it was recently proposed that newly synthesized palmitic acid (through the FAS pathway) may serve as an endogenous PPAR ligand (Chakravarthy *and others*, 2005). However, it remains unclear how our present observation relates to this hypothesis since the correlation between the FAS gene and palmitic acid is weaker than the correlation between PPAR $\alpha$  target genes (AOX, THIOL, PECL, PMDCI, mHMGCAS...) and palmitic acid. Eventually, a relatively robust correlation is highlighted between the genes involved in FA and cholesterol synthesis and monounsaturated FA with Mead acid (C20:3 $\omega 9$ ) displaying the strongest correlation with this set of genes (Fig. 9(a-c)). Since Sterol Regulatory Element Binding Proteins (SREBP) are transcription factors strongly involved in the regulation

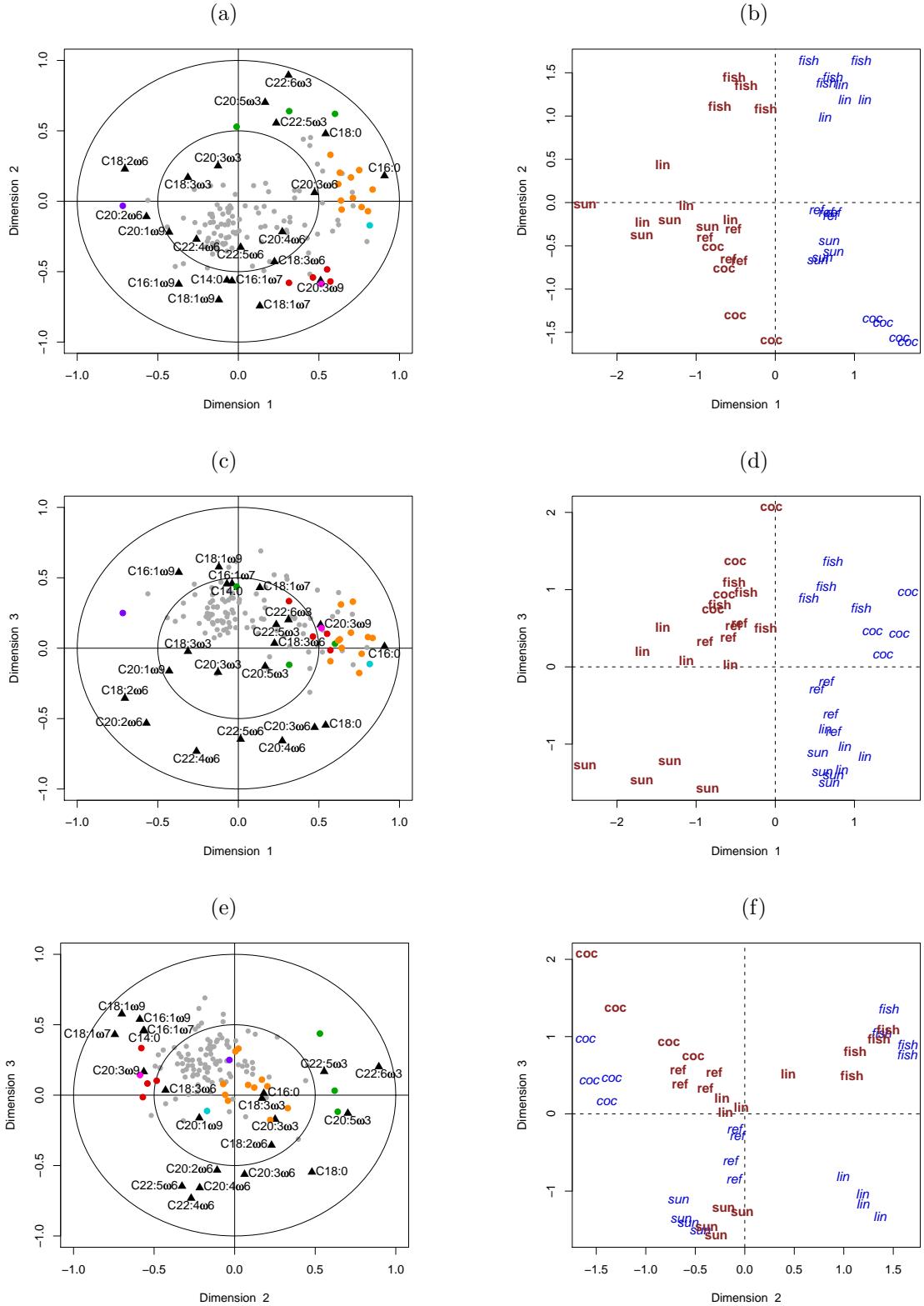


Figure 9: Variables plots and units plots for every combination of dimensions from 1 to 3. In variable plots (panels (a),(c),(e)), relevant genes are represented by thick points colored according to the Table 3, the others genes are colored in gray. The FA are represented by triangles in black and labeled by their name. In units plots (panels (b),(d),(f)), mice are represented according to their genotype (PPAR $\alpha$  in bold and brown, WT in italic and blue) and their diet (fish, coc, lin, sun and ref).

Col	Gene				Function
■	CAR1				Xenobiotic metabolism
■	PLTP				Lipoprotein metabolism
■	GSTpi2	CYP3A11	CYP2c29		Detoxification
■	FAS				Lipid biosynthesis
■	S14	ACC2	cHMGCoAS	HMGCoAred	
■	ACBP	AOX	BIEN	CPT2	
■	CYP4A10	HPNCL	L.FABP	PECI	FA catabolism
■	PMDCI	THIOL	mHMGCoAS		

Table 3: List of relevant genes and their functions. Genes on each row are represented in the Figure 9 according to their color (Col).

of genes implicated in FA and cholesterol synthesis (Jump *and others*, 2005), the study of a potential link between C20:3 $\omega$ 9 and the activity of the SREBP pathways may represent a fruitful approach for future experimental studies.

In conclusion, RCCA analysis and graphical representations of the two groups of variables analysed separately in Martin *and others* (2007) provides an integrated illustration of the effects of the genotype and of the diets on each dataset and further highlights interesting correlations between specific variables from the two datasets which should motivate a second round of biological experiments with cognitive outcomes.

## 5 Discussion

When dealing with the problem of highlighting relationships between several data sets, one answer can be given by the classical Canonical Correlation Analysis in a very particular case:

- only two sets of variables are considered (measured on the same experimental units);
- the number of experimental units must be large enough (greater than the total amount of variables) to allow mathematical computations (typically to invert matrices).

Several directions can be studied to deal with less restricted cases. In this article, we focused on the problem of too many variables in relation to the number of experimental units (further works will address the first point to provide operational tools to jointly examine three or more sets of variables). Here, we used existing methods including a regularization step to overcome the problem of singular matrices to be inverted. Through the implementation of RCCA on two examples, we have shown the efficiency of graphical displays to highlight significant correlations between two sets of variables. In both cases, external considerations (biological as well as chemical) assessed the relevancy of the results obtained.

RCCA can be considered as a very generic method that can be applied in a wide variety of contexts. Indeed, in this paper, we chose to deal with data sets coming from high throughput biology technologies but other collaborations are currently running around RCCA. For instance, in the context of food quality, the first results highlighting relationships between physicochemical measurements of meat quality and sensory evaluation by trained tasters are very promising and will soon be published. Another example comes from the information retrieval field. In collaboration with computer scientists, we are currently studying how information retrieval systems perform according to linguistic features of the submitted queries. Relationships between system performance and linguistic features can be highlighted through RCCA.

Another major interest of this paper is to illustrate the complementarity between statistical and biological approaches. The work we developed around the regularization methodologies was first motivated by a question arising from the nutrigenomic study presented here. Once the method was implemented and applied on the data sets, interesting new biological hypotheses arose from the results and led to further biological investigations. In particular, contacts have been initiated with authors involved in the development of a model of ABCB1 drug binding (Garrigues *and others*, 2002) in order to compare the molecular fitting between RCCA/BCI-selected drugs versus randomly selected drugs.

In conclusion, we feel that the use of RCCA can be beneficial to a number and a diversity of high throughput experiments now performed in biology (and in other fields as well) and we demonstrate here its ability to enhance the range of relevant conclusions that can be drawn from these relatively expensive experiments.

## Acknowledgements

The authors are grateful to Dr. Thierry Pineau for the availability of the data and, together with Dr Nicolas Loiseau and Dr Stephane Orlowski, for interesting discussions about biological interpretation of the results. This work was partially supported by grants from ACI IMPBio and ANR PNRA.

## References

- ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.
- BACCINI, A., BESSE, P., DÉJEAN, S., MARTIN, P. G. P., ROBERT-GRANIÉ, C. AND SAN CRISTOBAL, M. (2005). Stratégies pour l'analyse statistique de données transcriptomiques. *Journal de la Société Française de Statistique* **146**, 5–44.
- BICKEL, P. J. AND LI, B. (2006). Regularization in statistics. *Test* **15**, 271–344.
- BOULESTEIX, A. L. (2004). PLS dimension reduction for classification with microarray data. *Statistical Applications in Genetics and Molecular Biology* **3**, Article 33.
- BRILLI, M., FANI, R. AND LIO, P. (2007). MotifScorer: using a compendium of microarrays to identify regulatory motifs. *Bioinformatics* **23**, 493–495.
- CHAKRAVARTHY, M. V., PAN, Z., ZHU, Y., TORDJMAN, K., SCHNEIDER, J. G., COLEMAN, T., TURK, J. AND SEMENKOVICH, C. F. (2005). “New” hepatic fat activates PPAR $\alpha$  to maintain glucose, lipid, and cholesterol homeostasis. *Cell Metabolism* **1**, 309–322.
- EATON, M. L. AND PERLMAN, M. D. (1973). The non-singularity of generalized sample covariance matrices. *Annals of Statistics* **1**, 470–717.
- FRANK, I. E. AND FRIEDMAN, J. H. (1993). A statistical view of chemometrics regression tools. *Technometrics* **35**, 109–148.
- FRIEDMAN, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association* **84**, 165–175.
- GARRIGUES, A., LOISEAU, N., DELAFORGE, M., FERTE, J., GARRIGOS, M., ANDRE, F. AND ORLOWSKI, S. (2002). Characterization of two pharmacophores on the multidrug transporter P-glycoprotein. *Molecular Pharmacology* **62**, 1288–1298.
- GITTINS, R. (1985). *Canonical Analysis. A Review with Applications in Ecology*. New York: Springer.
- GUO, Y., KOTOVA, E., CHEN, Z. S., LEE, K., HOPPER-BORGE, E., BELINSKY, M. G. AND KRUH, G. D. (2003). MRP8, ATP-binding cassette C11 (ABCC11), is a cyclic nucleotide efflux pump and a resistance factor for fluoropyrimidines 2',3'-dideoxycytidine and 9'-(2'-phosphonylmethoxyethyl)adenine. *Journal of Biological Chemistry* **278**, 29509–29514.
- GUO, Y., HASTIE, T. AND TIBSHIRANI, R. (2007). Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* **8**, 86–100.
- HASTIE, T., BUJA, A. AND TIBSHIRANI, R. (1995). Penalized discriminant analysis. *Annals of Statistics* **23**, 73–102.
- HOERL, A. E. AND KENNARD, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67.
- HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321–377.
- HUANG, Y., BLOWER, P. E., YANG, C., BARBACIORU, C., DAI, Z., ZHANG, Y., XIAO, J. J., CHAN, K. K. AND SADEE, W. (2005). Correlating gene expression with chemical

- scaffolds of cytotoxic agents: ellipticines as substrates and inhibitors of MDR1. *The Pharmacogenomics Journal* **5**, 112–125.
- JUMP, D. B., BOTOLIN, D., WANG, Y., XU, J., CHRISTIAN, B. AND DEMEURE, O. (2005). Fatty acid regulation of hepatic gene transcription. *Journal of Nutrition* **135**, 2503–2506.
- KOOL, M., VAN DER LINDEN, M., DE HAAS, M., SCHEFFER, G. L., DE VREE, J. M., SMITH, A. J., JANSEN, G., PETERS, G. J., PONNE, N., SCHEPER, R. J., ELFERINK, R. P., BAAS, F. AND BORST, P. (1999). MRP3, an organic anion transporter able to transport anti-cancer drugs. *Proceedings of the National Academy of Sciences of the USA* **96**, 6914–6919.
- LEE, J. S., PAULL, K., ALVAREZ, M., HOSE, C., MONKS, A., GREVER, M., FOJO, A. T. AND BATES, S. E. (1994). Rhodamine efflux patterns predict P-glycoprotein substrates in the National Cancer Institute drug screen. *Molecular Pharmacology* **46**, 627–638.
- LEURGANS, S. E., MOYEEED, R. A. AND SILVERMAN, B. W. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society. Series B* **55**, 725–740.
- MARDIA, K., KENT, J. AND BIBBY J. (1979). *Multivariate Analysis*. London: Academic Press.
- MARTENS, H. AND NAES, T. (1989). *Multivariate Calibration*. London: Wiley.
- MARTIN, P. G. P., GUILLOU, H., LASSEURRE, F., DÉJEAN, S., LAN, A., PASCUSSI, J.-M., SAN CRISTOBAL, M., LEGRAND, P., BESSE, P. AND PINEAU, T. (2007). Novel aspects of PPAR $\alpha$ -mediated regulation of lipid and xenobiotic metabolism revealed through a multigenomic study. *Hepatology* **54**, 767–777.
- MEVIK, B.-H. AND WEHRENS, R. (2007). The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software* **18**, issue 2.
- MORRISON, D. F. (1967). *Multivariate Statistical Methods*. New York: McGraw-Hill.
- MULLEN, K. M. AND VAN STOKKUM, I. H. M. (2007). An introduction to the “Special Volume Spectroscopy and Chemometrics in R”. *Journal of Statistical Software* **18**, issue 3.
- NGUYEN, D. V. AND ROCKE, D. M. (2002). Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics* **18**, 39–50.
- PARMIGIANI, G., GARRET, E. S., IRIZARRY, R. A. AND ZEGER, S. L. (2003). *The Analysis of Gene Expression Data. Methods and Software*. New York: Springer.
- RAMSAY, J.O. AND SILVERMAN, B.W. (2005). *Functional Data Analysis*. Springer Series in Statistics. New York: Springer-Verlag.
- SCALA, S., AKHMED, N., RAO, U. S., PAULL, K., LAN, L. B., DICKSTEIN, B., LEE, J.-S., ELGEMEIE, G. H., STEIN, W. D. AND BATES S. E. (1997). P-glycoprotein substrates and antagonists cluster into two distinct groups. *Molecular Pharmacology* **51**, 1024–1033.
- SHITAN, N., TANAKA, M., TERAI, K., UEDA, K. AND YAZAKI, K. (2007). Human MDR1 and MRP1 recognize berberine as their transport substrate. *Bioscience, Biotechnology and Biochemistry* **71**, 242–245.
- SPEED, T. (2003). *Statistical Analysis of Gene Expression Microarray Data*. Boca Raton: Chapman and Hall/CRC.
- SZAKÁCS, G., ANNREAU, J.-P., LABABIDI, S., SHANKAVARAM, U., ARCIELLO, A., BUSSEY, K. J., REINHOLD, W., GUO, Y., KRUH, G. D., REIMERS, M., WEINSTEIN, J. N. AND GOTTESMAN, M. M. (2004). Predicting drug sensitivity and resistance: Profiling ABC transporter genes in cancer cells. *Cancer Cell* **4**, 147–166.
- SZAKÁCS, G., PATERSON, J. K., LUDWIG, J. A., BOOTH-GENTHE, C. AND GOTTESMAN, M. M. (2006). Targeting multidrug resistance in cancer. *Nature reviews. Drug discovery*. **5**, 219–234.

- TER BRAAK, C. J. F. (1990). Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika* **55**, 519–531.
- TIHONOV, A.N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics. Doklady* **4**, 1035–1038.
- VINOD, H. D. (1976). Canonical ridge and econometrics of joint production. *Journal of Econometrics* **6**, 129–137.
- XIAO, J. J., FORAKER, A. B., SWAAN, P. W., LIU S., HUANG, Y., DAI Z., CHEN, J., SADEE, W., BYRD, J., MARCUCCI, G. AND CHAN, K. K. (2005). Efflux of depsipeptide FK228 (FR901228, NSC-630176) is mediated by P-glycoprotein and multidrug resistance-associated protein 1. *Journal of Pharmacology and Experimental Therapeutics* **313**, 268–276.
- XING, H., WANG, S., WENG, D., CHEN, G., YANG, X., ZHOU, J., XU, G., LU, Y. AND MA, D. (2007). Knock-down of P-glycoprotein reverses taxol resistance in ovarian cancer multicellular spheroids. *Oncology Reports* **17**, 117–122.
- YAMANISHI, Y., VERT, J.-P. AND KANEHISA, M. (2004). Heterogeneous data comparison and gene selection with kernel canonical correlation analysis. In: Schölkopf, B., Tsuda, K. and Vert, J.-P. (editors), *Kernel Methods in Computational Biology*. MIT Press, pp. 209-230.
- ZHANG, X. P., RITKE, M. K., YALOWICH, J. C., SLOVAK, M. L., HO, J. P., COLLINS, K. I., ANNABLE, T., ARCECI, R. J., DURR, F. E. AND GREENBERGER, L. M. (1994). P-glycoprotein mediates profound resistance to bisantrene. *Oncology Research* **6**, 291–301.

## Supplementary material

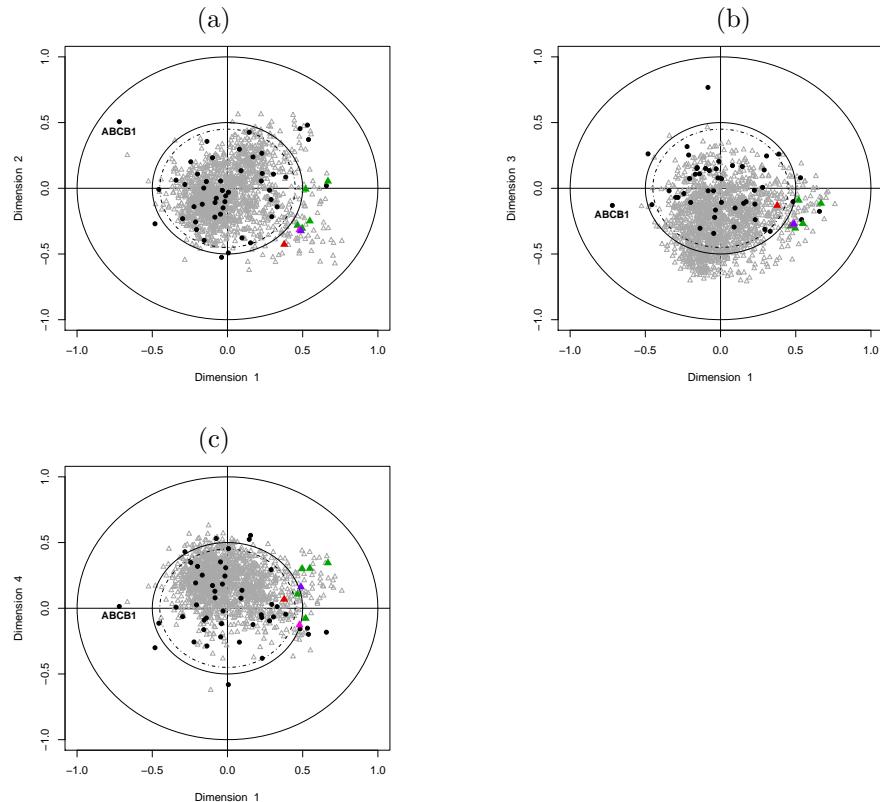


Figure S1: Representation on the first four canonical variates of the 8 drugs identified by the BCI procedure (Szakács *and others*, 2004) but not by RCCA using the threshold of 0.5 described in paragraph 4.1 “Statistical Data Analysis”. The 8 drugs are identified by triangles colored according to Table S1.

Col	Drug (NSC N°)					
[pink]	156625					
[violet]	618757					
[red]	694268					
[green]	354975	374980	636679	651727	652903	

Table S1: List of the 8 drugs identified by the BCI procedure (Szakács *and others*, 2004) but not by RCCA using the threshold of 0.5 described in paragraph 4.1 “Statistical Data Analysis”. NSC156625 (pink) is a known ABCB1 substrate as mentioned in (Szakacs and others, 2004). NSC618757 (violet) and NSC694268 (red) are new ABCB1 substrates which have been validated by the follow-up study in (Szakács *and others*, 2004). The other 5 drugs (green) have not been directly validated as human ABCB1 substrates by Szakacs and others (2004). Using a threshold of 0.45 on the scalar product between the ABCB1 vector and the drug vectors allows RCCA to select 52 drugs, among which 5 of the 8 drugs presented in this Table are present (and in particular the 3 validated ABCB1 substrates are present in this list). NSC374980, NSC651727 and NSC652903 require a threshold lower than 0.45 to be selected by RCCA (threshold=0.44 to select NSC374980 and NSC652903 and threshold=0.36 to additionally select NSC651727).

**ANNEXE 4 : Article publié sur l'utilisation des dérivées de courbes lissées par des splines cubiques pour l'analyse des données de puces à ADN obtenues en cinétique**



## Research Article

# Clustering Time-Series Gene Expression Data Using Smoothing Spline Derivatives

S. Déjean,<sup>1</sup> P. G. P. Martin,<sup>2</sup> A. Baccini,<sup>1</sup> and P. Besse<sup>1</sup>

<sup>1</sup>Laboratoire de Statistique et Probabilités, UMR 5583, Université Paul Sabatier, 31062 Toulouse Cedex 9, France

<sup>2</sup>Laboratoire de Pharmacologie et Toxicologie, UR 66, Institut National de la Recherche Agronomique (INRA), 180 Chemin de Tournesouille, BP 3, 31931 Toulouse Cedex 9, France

Received 14 December 2006; Revised 6 March 2007; Accepted 16 May 2007

Recommended by Stéphane Robin

Microarray data acquired during time-course experiments allow the temporal variations in gene expression to be monitored. An original postprandial fasting experiment was conducted in the mouse and the expression of 200 genes was monitored with a dedicated macroarray at 11 time points between 0 and 72 hours of fasting. The aim of this study was to provide a relevant clustering of gene expression temporal profiles. This was achieved by focusing on the shapes of the curves rather than on the absolute level of expression. Actually, we combined spline smoothing and first derivative computation with hierarchical and partitioning clustering. A heuristic approach was proposed to tune the spline smoothing parameter using both statistical and biological considerations. Clusters are illustrated a posteriori through principal component analysis and heatmap visualization. Most results were found to be in agreement with the literature on the effects of fasting on the mouse liver and provide promising directions for future biological investigations.

Copyright © 2007 S. Déjean et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

In the context of microarray experiments, we focused on the analysis of time-series gene expression data. Our original data were hepatic gene expression profiles acquired during a fasting period in the mouse. Two hundred selected genes were studied through 11 time points between 0 and 72 hours, using a dedicated macroarray.

The literature concerning the analysis of time-series gene expression data mainly addresses two problems: identification of differentially expressed genes over time [1–4] and temporal profile clustering to identify genes which are coordinately regulated during the time course experiment [5–8]. Methods developed to propose solutions to the first problem can be viewed as a preliminary step that filters genes to which a clustering procedure can then be applied [9]. However, since we used a dedicated macroarray with a limited number of genes, we focused directly on the clustering of temporal profiles. In the above-mentioned articles that address the second problem, clustering is based on a set of pre-defined model profiles. This could be relevant when dealing with short time-series, but with 11 time points, we assumed that the information contained in the data was sufficient and that we did not require such prior information.

Since the aim of this paper is not prediction but curve clustering, the approach considered here does not refer to parametric statistical models (such ARMA) used to fit time-series. Furthermore, as mice differ from one point in time to another, models for longitudinal data are not relevant in the present context.

The purpose of the present study was to identify homogeneous clusters of genes. Nevertheless, a relevant clustering method must take into account the data specificity and, in particular, should integrate the temporal aspect. In this context, the absolute level of expression is generally of little interest, mainly because the probes on the microarray can have a significant influence on the measured intensities (see, e.g., [10]). Instead, the shapes of the curves may provide meaningful information on coordinate gene regulations. The suitable mathematical tool to describe this information is the derivative. Therefore, a preliminary stage consists in smoothing the temporal profiles in order to get regular and differentiable functions. The study of functional data is addressed in the statistical literature (see [11], for a survey). In the context of microarray data, Bar-Joseph et al. [12] use splines to provide continuous representations of time-series gene expression profiles, and thus to permit the interpolation of missing

values and dataset alignment. We used the same mathematical tool to propose a methodology for curve clustering.

Our approach is in the framework of functional data analysis [11]. Its main originality lies in its focus on the first derivative of curves by means of a priori spline smoothing. The approach was composed of two steps. The first one can be viewed as a signal extraction method: assuming that gene expression profiles are regular curves, spline smoothing is performed. Tuning the smoothing parameter is a core problem that could not be achieved by the usual cross-validation method because of the poor quality of clustering results. Thus, we propose a heuristic approach that takes into account both statistical and biological considerations. The second step consisted in clustering the derivatives of the smoothed curves after discretization; hierarchical clustering and the  $k$ -means algorithm were used successively in order to obtain robust clusters.

Details of the biological experiment are given in the second section of the paper. Then, statistical methodology is developed with a focus on tuning the smoothing parameter. In the fourth section, clustering results are interpreted, then illustrated *a posteriori* through principal component analysis (PCA) and heatmap visualization of simultaneous clustering of curves and time points. Finally, some elements of discussion about the analysis of times-series gene expression data are given to conclude the paper.

## 2. BIOLOGICAL EXPERIMENT

### 2.1. Experimental design

Ten-week-old male C57BL/6J mice (wild-type) were obtained from Charles River France (Les Oncins, France) and were acclimatized to local animal facility conditions for two weeks prior to the fasting experiment. Mice were housed in groups of four in plastic cages at a temperature of 22°C ( $\pm 2^\circ\text{C}$ ) with a 12/12 hours light/dark cycle. Mice were randomly assigned to the experimental groups. A total of 44 mice (11 cages  $\times$  4 mice/cage) were subjected to 11 different fasting periods ranging from 0 to 72 hours. All mice were moved into clean cages without food at 5 a.m. (2 hours prior to the beginning of the light phase). Since mice mainly eat during the night, this experimental setting corresponded to postprandial fasting. At each of the selected time points (0, 3, 6, 9, 12, 18, 24, 36, 48, 60, and 72 hours), 4 mice were euthanized. The liver was dissected, snap-frozen in liquid nitrogen, and stored at  $-80^\circ\text{C}$  until RNA extraction.

The sampling rate in time-course experiments is discussed in [13]. In our case, gene expression was measured at 11 time points from 0 to 72 hours of fasting with a decreasing sampling rate. It was assumed that most of the gene expression changes would occur at the beginning of fasting. Nevertheless, the number of time points was determined to be able to observe fluctuations in the gene expression profiles, that is, changes in the sign of their derivatives, until the 72nd hour of fasting.

### 2.2. Production of INRAArray 01.3

Selection, cloning, amplification, and spotting of the cDNA fragments onto nylon membranes have been previously described for version 01.2 of INRAArray [14, 15]. The same procedure was followed for INRAArray 01.3. Eighty genes were added to the panel of 120 genes present on INRAArray 01.2, leading to a total of 200 genes. They were mainly genes involved in energy and xenobiotic metabolism. Furthermore, we developed a set of 13 probes and corresponding *in vitro* transcribed polyA-RNAs from yeast to be used as internal controls for normalization purposes (spiked-in RNAs). The full list of clones present on INRAArray 01.3 can be found in [16]. Additionally, the spotting buffer (50% DMSO) was spotted on the macroarray at 200 different locations for the analysis of the background.

### 2.3. RNA extraction and labeling

Total RNA was extracted with TRIzol reagent (Invitrogen, Cergy Pontoise, France) according to the manufacturer's instructions. The integrity of the RNAs was evaluated on a Bioanalyzer 2100 (Agilent Technologies, Massy, France). For each sample, 3  $\mu\text{g}$  of total RNA along with a fixed amount of the 13 spiked-in yeast RNAs were labelled by reverse transcription with Superscript II RT (Invitrogen) in the presence of 40  $\mu\text{Ci}$  of  $[\alpha^{-33}\text{P}]$ dCTP (ICN, Orsay, France). The clean-up of the labelled cDNAs and the hybridization, washing, scanning, and image analysis of INRAArray have been described previously [14].

### 2.4. Data preprocessing

All data were log-transformed. The normality of the background intensities was verified using the Kolmogorov-Smirnov test. Four macroarrays out of 44 exhibited  $P$ -values lower than 0.05. Each gene on each array was declared "present" when its intensity exceeded the mean plus twice the standard deviation of the background intensities. Only the genes declared "present" on a minimum of six macroarrays were retained for further analysis. This procedure yielded a total of 130 genes selected for further analysis. Data were normalized using the average signal of the 13 spiked-in yeast RNAs. Boxplots for the 44 macroarrays led us to declare 4 macroarrays as outliers, which were removed from the dataset. Thus the dataset studied in this paper consists of a matrix of log-transformed normalized intensities for 130 genes  $\times$  40 samples (40 mice).

## 3. STATISTICAL METHODOLOGY

Let us recall that our purpose consisted in clustering temporal profiles according to their shape. In this context, the mathematical tool to be used is the first derivative of the curve. Therefore, the first step aimed at getting one regular curve modeling the evolution of each gene.

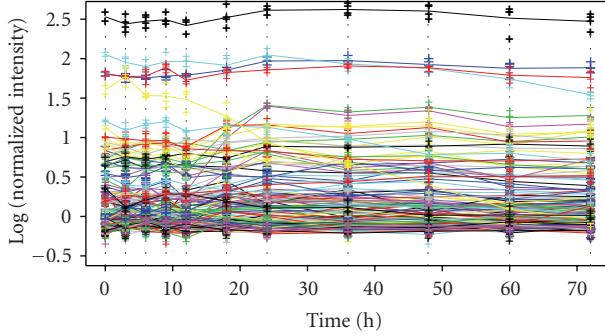


FIGURE 1: Log-normalised intensity versus time for 130 genes. For each gene, the line joins the average value at each time point. Vertical dashed lines indicate time points.

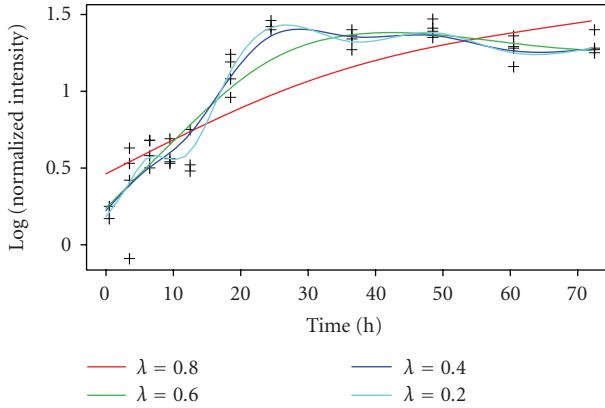


FIGURE 2: Smoothed curves obtained for the gene *Cyp4a10* with  $\lambda = 0.2, 0.4, 0.6$ , and  $0.8$ .

### 3.1. Signal extraction

Rather than directly computing means of the observed values as in Figure 1, we tried a somewhat more realistic approach based on two essential assumptions:

- (i) the values at each time point are noisy observations of the “true” value (obviously unknown),
- (ii) this type of biological phenomenon should be a regular, and so differentiable, function of time. This means for us without singularities or any chaotic behavior. This is a sensible assumption when data are acquired at a macroscopic level; it may be false at a molecular or a single-cell scale. Furthermore, in this study, fasting is typically a progressive stimulus where hormonal changes take place progressively and should not imply biological thresholds.

This led us to consider the following nonparametric model for each gene expression:

$$y_i^j = f(t_j) + \varepsilon_{ij}, \quad i = 1, \dots, 4, \quad j = 1, \dots, 11, \quad (1)$$

where  $y_i^j$  denotes the observation for the  $i$ th mouse ( $i = 1, \dots, 4$ ) at time  $t_j$ ,  $f$  is a continuous and differentiable func-

tion, and  $\varepsilon_{ij}$  are independent and identically distributed random variables satisfying classical assumptions:

$$E(\varepsilon_{ij}) = 0, \quad \text{Var}(\varepsilon_{ij}) = \sigma^2. \quad (2)$$

This problem is classically solved by a nonparametric estimation of  $f$ . Kernel smoothing or spline smoothing both achieve this objective, but we naturally preferred spline smoothing since we needed to estimate both the function and its derivative. This is quite easy using cubic spline smoothing. The estimation of any gene expression curve according to this model is then the solution to the following optimization problem [17]:

$$\min_{f \in H_1} \frac{1}{4 \times 11} \sum_{i=1,4; j=1,11} [y_i^j - f(t_j)]^2 + \lambda \int_{t_1}^{t_{11}} [f''(u)]^2 du, \quad (3)$$

where  $f$  belongs to  $H_1$ , the Sobolev space of continuous functions with integrable squared second derivative, and  $\lambda$  is the smoothing parameter. This parameter balances the influence between the left-hand term of (3), which forces solutions to be close to mean values, and the right-hand one, which controls the regularity of the function.

The solution  $\hat{f}$  of (3) is a piecewise function which is defined on the basis of cubic polynomials. The solution shape and its smoothness depend directly on  $\lambda$ . On the one hand, as  $\lambda$  grows, the solution converges to a trivial linear regression since the integral in the right-hand term of (3) tends to zero (with the second derivative). On the other hand, if  $\lambda$  decreases towards zero, the solution becomes a piecewise polynomial interpolating function of the means of the four values at each time point since the left-hand term reaches its minimum value.

### 3.2. Tuning the smoothing parameter

The estimation of the function  $f$  in model (1) according to formula (3) clearly raises the central problem of how to tune the smoothing parameter  $\lambda$  in order to correctly extract the informative part of the signal. The influence of  $\lambda$  is illustrated with the *Cyp4a10* gene in Figure 2. Depending on the  $\lambda$  value, smoothed profiles exhibit more or fewer fluctuations along the time axis.

We first performed  $\lambda$  tuning by minimizing a generalized cross-validation estimation of a prediction error. Each gene was thus allocated one  $\lambda$  value. Results were disappointing: heterogeneous profiles were clustered together and biological interpretation was very difficult.

Therefore, we adopted another strategy: a unique  $\lambda$  value for all genes. We propose a heuristic approach combining two levels of reflection: eigenelements of the PCA performed *a posteriori* and biological interpretations of results.

#### Scree graph of eigenvalues and eigenvectors smoothness

The PCA computation requires the number of principal components (PC), that is, the projection space dimension, to be chosen. Some subspace stability argumentation is given

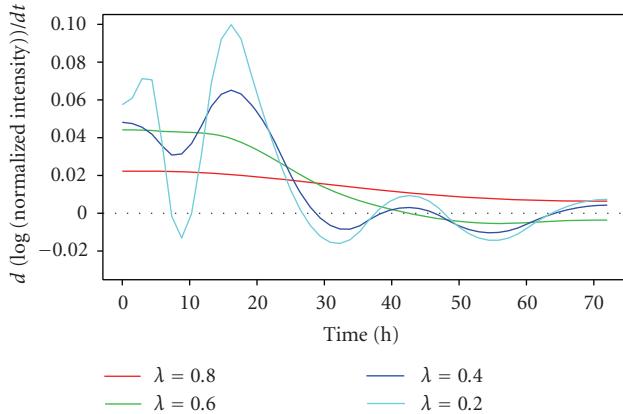


FIGURE 3: First derivatives of smoothed curves obtained for the gene *Cyp4a10* (Figure 2). Horizontal dotted line locates zero.

in [18] to point out the importance of the difference of values between the last eigenvalue kept and the first that is dropped out.

Practically, let us consider the following steps:

- (i) each gene expression profile is smoothed according to the same  $\lambda$  value (Figure 2),
- (ii) first derivatives (Figure 3) are computed and discretized, thus giving a new data matrix on which
- (iii) a PCA is computed, leading to a scree graph (Figure 4) together with eigenvectors (Figure 5) that are also discretized time functions.

These graphs were plotted for different values of  $\lambda$  (Figures 4 and 5). When  $\lambda$  was large, each expression profile was fitted by a linear regression, and so the derivative was constant, equal to the slope. Obviously, a PCA gave only one large eigenvalue (Figure 4(a)) since the data matrix was of rank one. The same computations were run for different decreasing values of  $\lambda$  until a second eigenvalue arose from noise (Figure 4(b)). The eigenvectors associated with the two largest eigenvalues looked regular and led to easy interpretations of approximations of gene profiles which were projected onto the eigenbasis (Figures 5(a) and 5(b)). But as  $\lambda$  continued to decrease, a third eigenvalue arose from noise (Figures 4(c) and 4(d)) and the first two eigenvectors became much more irregular (Figures 5(c) and 5(d)), and thus much more difficult to interpret, with the risk of giving sense to a noise component.

#### *Biological interpretation*

A second consideration which should be addressed is the consistency with biological relevance. For higher  $\lambda$  values, the phenomena highlighted were mainly based on the opposition between the beginning and the end of the experiment. Then, clustering or factorial methods could highlight globally increasing, stationary or decreasing genes without any information about the intermediary period of fasting; two or three time points would have led to the same in-

terpretation. As  $\lambda$  decreased, intermediary time points were integrated (through the second PC) but eigenvectors had to be checked to be smooth enough. Too many oscillations in the eigenvectors could be irrelevant and potentially lead to misinterpretation.

#### *Synthesis*

The two levels of consideration yielded approximately the same value for the parameter  $\lambda \approx 0.6$ . For this value, the detail level of curves was consistent with the number of observations; there were clearly two separate eigenvalues; the corresponding eigenvectors were smooth enough and led to simple and interpretable projection spaces for graphical displays.

#### **3.3. Clustering**

The aim of the analysis of these data was to identify some characteristic evolutions of gene regulation occurring during fasting. More precisely, we intended to obtain a few homogeneous clusters of curves, the curves being summarized by the values of the derivative of smoothed expression profiles at some discretization points. We chose 20 points equally spaced between 0 and 72 hours. This value roughly corresponds to the thinnest interval between two real measurements (3 hours) applied all along the 3-days fasting. Furthermore, let us note that when the smoothing is tuned through a penalization parameter, the number and the positions of the points are not very important; practically, results obtained with values from 10 to 50 discretization points were found to be very stable.

The data to be analyzed can be presented in a table with 130 individual genes in rows and 20 variables dates in columns. The values are the discretized values of the derivative of smoothed curves.

In the context of microarray data analysis, hierarchical clustering is often performed. It was used here in an initial stage. Note that the distance chosen between two curves was the standard Euclidean distance computed between the 20 pairs of coordinates (correlation-based distance would be redundant with the use of the derivative). On the other hand, the criterion chosen to agglomerate two clusters was the Ward criterion, generally advocated by statisticians. It consists in fusing the two clusters that minimize the increase in the total within-cluster sum of squares [19]. We also performed clustering with the information summarized by the first two principal components but, as mentioned by [20], it did not improve the results.

A major weakness of the hierarchical algorithm is that an improper fusion at an early stage cannot be corrected later. In order to correct this weakness, at least partially, we performed a partitioning method (also described as  $k$ -means) in which initialization is given by the  $k$  centroids of the clusters obtained through hierarchical clustering. See, for example, [21] for a survey of  $k$ -means in the context of microarray data.

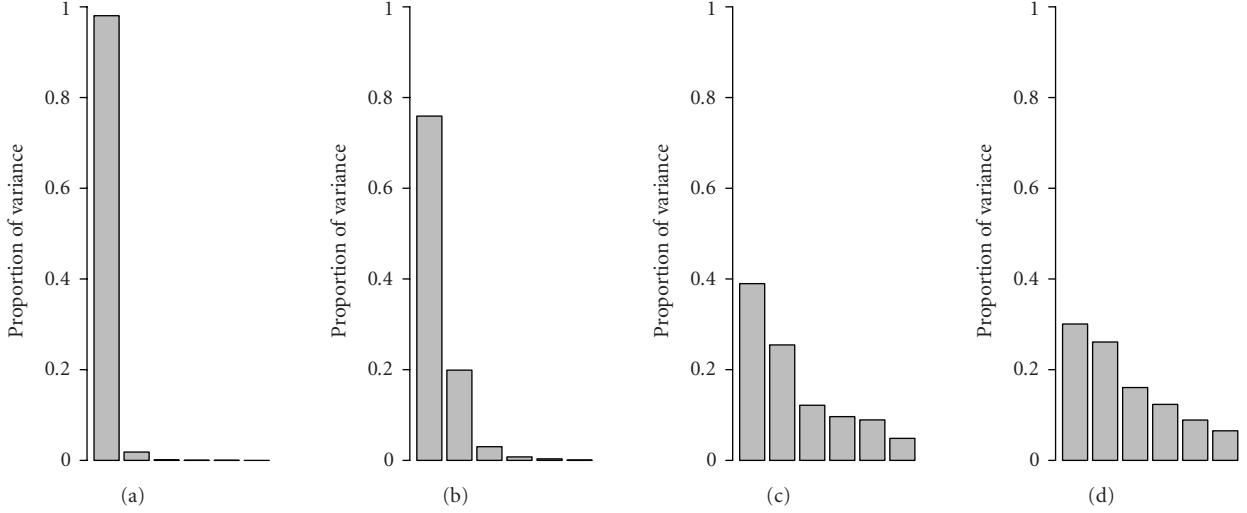


FIGURE 4: Influence of the smoothing parameter  $\lambda$  on the proportion of variance explained by the first six PCs. From left to right,  $\lambda$  equals (a) 0.8, (b) 0.6, (c) 0.4, and (d) 0.2.

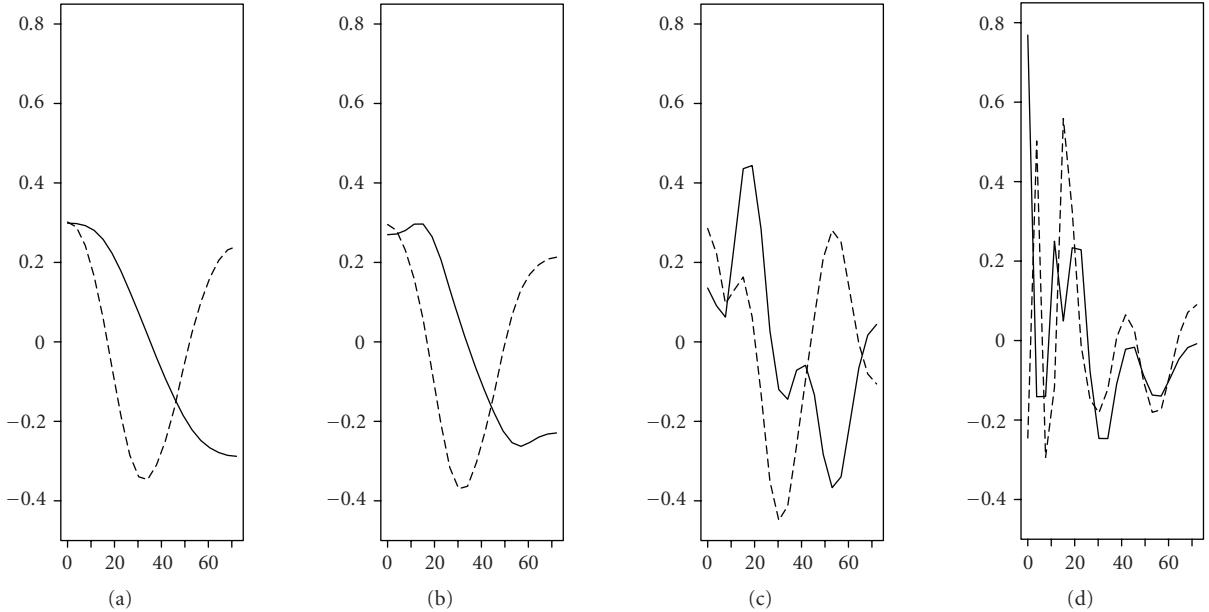


FIGURE 5: Influence of the smoothing parameter  $\lambda$  on the two first eigenvectors (first: full line, second: dashed line) of the PCA. From left to right,  $\lambda$  equals (a) 0.8, (b) 0.6, (c) 0.4, and (d) 0.2.

## 4. RESULTS

### 4.1. Hierarchical clustering

Hierarchical clustering produced a dendrogram (Figure 6) that led to arguable choices between 3 and 8 clusters. Four clusters were considered because they led to a relevant and easily perceived biological interpretation. Analysis of more than 4 clusters provides more precise information to the biologist studying gene expression changes during fasting and will be described elsewhere.

Let us note that the four clusters defined by the dendrogram globally correspond to four temporal expression profiles: decreasing (hc3), stationary (hc2), weakly increasing (hc1), strongly increasing (hc4).

### 4.2. *k*-means partitioning

To make the clustering more robust, we performed the *k*-means algorithm, specifying the initial centers as the centers of the classes obtained when cutting the dendrogram

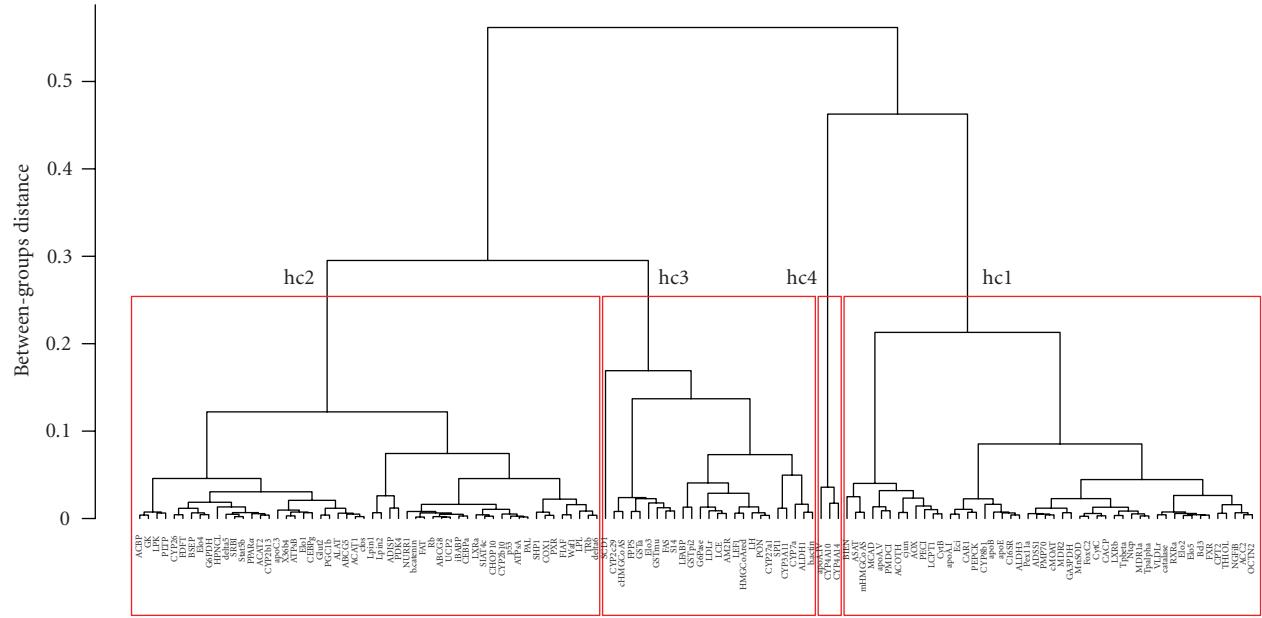


FIGURE 6: Dendrogram representing the result of the hierarchical clustering performed on the value of the first derivative smoothed curves using Euclidean distance and Ward criterion. The horizontal lines locate the cut level identifying 4 clusters ( $hc_1, \dots, hc_4$ ).

TABLE 1: Changes between hierarchical clustering and  $k$ -means clusters.

Clusters	hc1	hc2	hc3	hc4	Sum
km1	26	3	0	0	29
km2	22	48	4	0	74
km3	0	3	21	0	24
km4	0	0	0	3	3
Sum	48	54	25	3	130

(see Figure 6). Changes that occurred during  $k$ -means are summarized in Table 1.

The main event lies in the 22 genes that change from  $hc_1$  (low increasing) to  $km_2$  (stationary). Other changes are minor and the three-gene cluster ( $hc_4$ ) remains unchanged ( $km_4$ ).

The four clusters of curves obtained after  $k$ -means partitioning are displayed in Figure 7; their interpretation is given below.

$km_1$ : the expression of the 29 genes which belong to the first cluster increases during the first half of fasting and then tends to decrease slightly or to stabilize. Most of these genes are involved in lipid catabolism. In particular, this cluster contains the genes encoding the three enzymes involved in fatty acid  $\beta$ -oxidation (Acyl-CoA oxidase, Bifunctional ENzyme, and 3-ketoacyl-CoA thiolase) and the enzyme involved in the rate-limiting step of ketogenesis (mitochondrial HMG-CoA synthase). During fasting, lipids stored in the adipose tissue are mobilized and the liver plays a major role in catabolizing these lipids to provide energy and appropri-

ate substrates to peripheral organs. Peroxisome proliferator-activated receptor alpha (PPAR $\alpha$ ) is an important hepatic transcriptional modulator of lipid catabolism which is activated during fasting [22]. We noticed that most genes in  $km_1$  are well-described PPAR $\alpha$  targets (reviewed in [23]). PPAR $\alpha$  activation and subsequent coordinate induction of  $km_1$  genes likely provide a molecular interpretation of their clustering.

$km_2$ : the second cluster (74 genes) reveals quasi-constant curves. These genes are not regulated during fasting.

$km_3$ : the third one (24 genes) is characterized by a decrease of the gene expression with time. This cluster is mostly composed of genes which are involved in xenobiotic metabolism (the cytochromes P450 3a11, 2c29, and the glutathione-S-transferases  $\alpha$ ,  $\mu$ , and  $\pi$ ), lipogenesis (FAS, S14, SCD1), cholesterol metabolism (FPP synthase, Cyp7a, cytosolic HMG-CoA synthase, and reductase), and glucose metabolism (glucokinase, pyruvate kinase, and glucose 6-phosphatase). Since large amounts of lipids accumulate in mouse liver during fasting (data not shown), it is likely that the activity of the sterol regulatory element binding proteins (SREBP1 and SREBP2) is reduced. These transcription factors regulate numerous genes involved in lipid synthesis. Their reduced activity may provide a rationale for the decreased expression of lipogenesis and cholesterol synthesis genes. One striking observation is that the liver fatty acid-binding protein (L-FABP), a known PPAR $\alpha$  target gene, was also repressed, and is thus found in this third cluster. This result is consistent with a previous report [22] and is currently being investigated.

$km_4$ : the fourth cluster is composed of the most strongly induced genes during fasting: *Cyp4a10* and *Cyp4a14*, the two

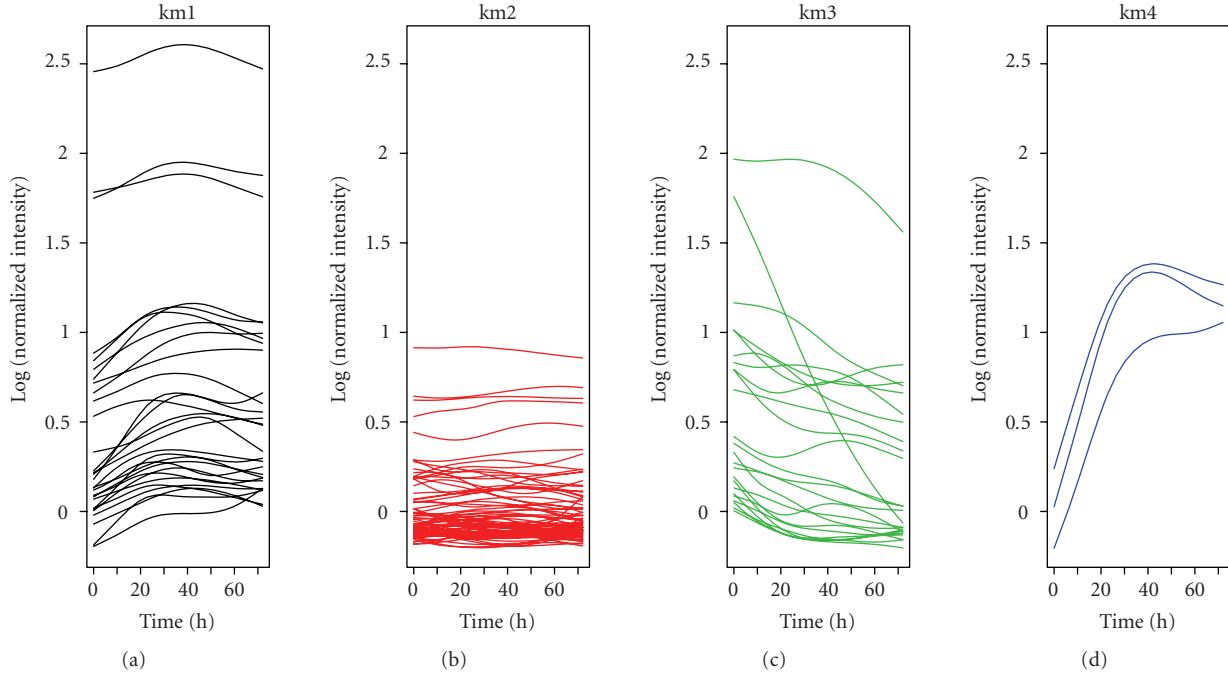


FIGURE 7: Representation of the smooth curves distributed in 4 clusters determined after hierarchical and  $k$ -means classification.

most responsive PPAR $\alpha$  target genes and apoA-IV. Their expression strongly increases until the 40th hour of fasting and then stabilizes.

Overall, these results are consistent with the known hepatic gene expression modulations induced by fasting [24]. Hepatic fatty acid oxidation and fatty acid transport and trafficking are induced (mostly through induction of PPAR $\alpha$  target genes) and allow the liver to manage, at least partially, the large amounts of lipids which are mobilized from the adipose tissue. On the other hand, lipogenesis and cholesterogenesis are decreased, probably due to reduced SREBP activity. Glucose metabolism genes are decreased, probably in parallel with the decrease in plasma glucose (data not shown). Additionally, some novel hypotheses were drawn from this clustering results and are subject to further experimental investigation.

#### 4.3. Graphical display

We used two methods to give graphical evidence of clusters relevance: PCA and heatmap visualization of simultaneous clustering for genes and time points.

##### Principal component analysis

We performed a PCA that checked the relevance of the four clusters. The proportion of variance explained by the first two PCs reached about 96% (85% for the first PC), and thus justified a two-dimensional representation (Figure 8).

Genes are shown with different colors according to their cluster (Figure 8 right). The four clusters are distributed along the first (horizontal) axis in a specific order: from left

to right, gene expression profiles go from a sharply increasing curve (km4, in blue) to a weakly increasing curve (km1, genes in black), then stationary profiles (km2, genes in red), and finally a decreasing curve (km3, genes in green). The second (vertical) axis highlights gene regulations occurring around the 30th hour of fasting. Analysis of more than 4 clusters helps in identifying groups of genes regulated during this intermediary phase of the fasting experiment (data not shown).

The times of discretization are also shown in Figure 8. Their regular pattern indicates the consistency of the smoothed and discretized data. The sort of inverted U formed by the times of discretization recalls well-known situations of variables connected with time.

##### Heatmap visualization

Heatmaps are widely used to graphically represent multidimensional gene expression data which have been subjected to clustering algorithms.

We first compared heatmaps obtained on two different data matrices: the matrix of discretized smoothed gene expression profiles; the matrix of discretized derivatives of the smoothed gene expression profiles. In both cases, we forced a reordering of the time points to follow, as much as the dendrogram allows it, their increase from left to right. Perfectly ordered time points were obtained. Genes were systematically reallocated to four clusters using  $k$ -means algorithm. This explains why a dendrogram cannot be drawn on the left side of the heatmap. Horizontal lines separated the four clusters obtained following  $k$ -means reallocation.

The comparison of the heatmaps obtained (not shown here) clearly highlighted a major advantage of color coding

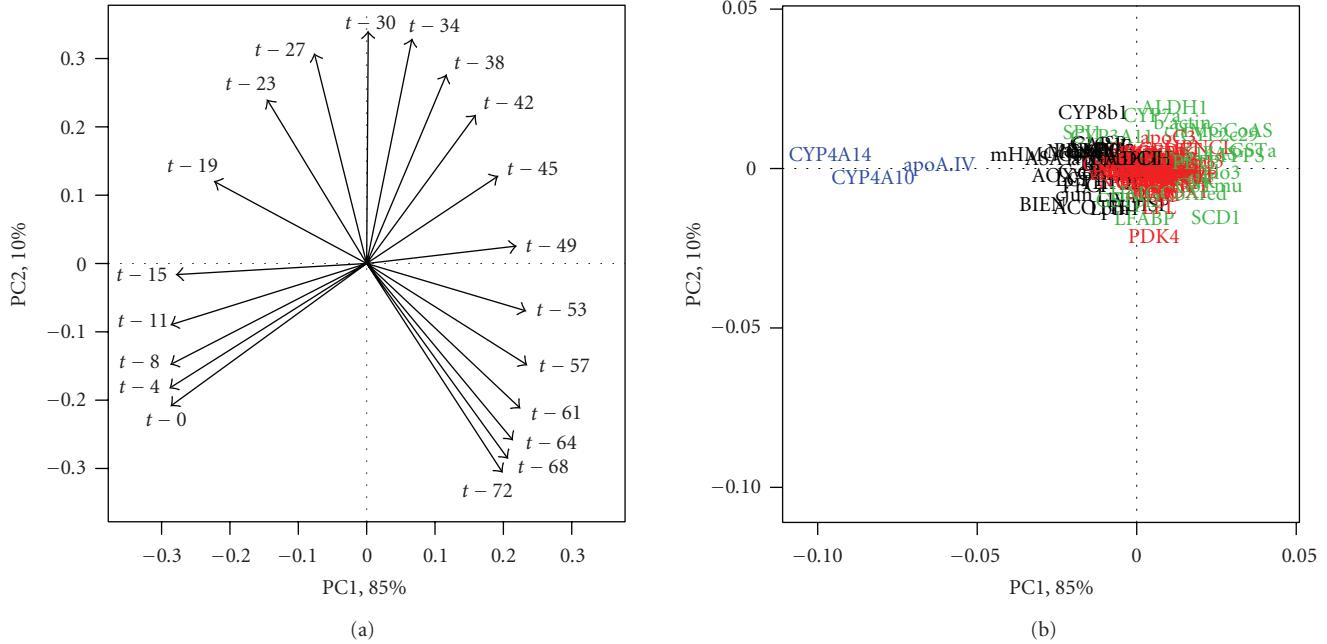


FIGURE 8: Representation of variables (discretized time points, on the left) and individuals (genes, on the right) on the first two principal components. Genes are differentially displayed according to their cluster after  $k$ -means.

the derivatives instead of the profiles themselves. When color coding the profiles themselves, the eye needs to integrate the changes of colors along the ordered time points to extract the direction and the amplitude of the changes in gene expression. Conversely, color coding the derivatives allows a direct extraction of gene expression changes direction and amplitude at the different time points. Consequently, it becomes much easier to identify both the causes of the clustering and the time points at which major transcriptional changes occur.

Here, we present two heatmaps computed on the matrix of discretized derivatives of the smoothed gene expression profiles. The clustering of the gene expression profile derivatives was performed as described in the previous paragraphs. Similarly, the hierarchical clustering of the time points was done with the Euclidean distance and the Ward criterion. The first heatmap was computed with all 130 genes (Figure 9). The most strongly regulated genes are easily visualized: km1 genes at the uppermost and *SCD1* which appears as a green line in the lower quarter of the heatmap. While km4 genes appear most strongly upregulated until the 30th hour of fasting, *SCD1* is negatively regulated in a constant way during all the fasting periods. Thus, by contrast to km4 genes, *SCD1* expression profile could have been equally well modelled by a straight line since its derivative appears constant with fasting time. One obvious drawback of this representation (Figure 9) is that the representation of km4 and *SCD1* gene profile derivatives tend to strongly narrow the color range used to represent the other profile derivatives due to their extreme regulations in mouse liver during fasting. Once interpreted, km4 and *SCD1* genes were thus removed from the dataset and a new heatmap was computed (Figure 10). Genes

belonging to km1 all display a clear increase in their expression until up to 30 hours of fasting. Their expression is stable from 30 to 45 hours. After 45 hours, divergent regulations are observed (stable, increased, or decreased expression) which could have been highlighted through the analysis of more than 4 clusters. A similar interpretation can be drawn for downregulated km3 genes located in the lower part of the heatmap.

Interestingly, time points clustering highlighted that most gene expression changes occur during the first 30 hours of fasting although subtle gene expression modulations are still observed after this time point.

## 5. DISCUSSION

This paper presents an integrated use of statistical tools that provides a framework for the study of time-series data obtained with microarray technology. Before the usual clustering step, we perform spline smoothing as a denoising method. In this context, the quality of the results depends highly on the core problem of tuning the smoothing parameter. For this purpose, we propose an original strategy using both statistical and biological considerations. The procedure is completed by clustering the derivatives of the continuous curves resulting from smoothing, which actually represent the temporal variations of mRNA concentrations.

The main results obtained are clearly in accordance with previous studies on the effects of fasting on hepatic gene expression in the mouse. This study provides a novel time-dependent view of fasting effects on gene expression which are usually studied through 2 or 3 time points only (including a fed state corresponding to time 0). It may thus help in

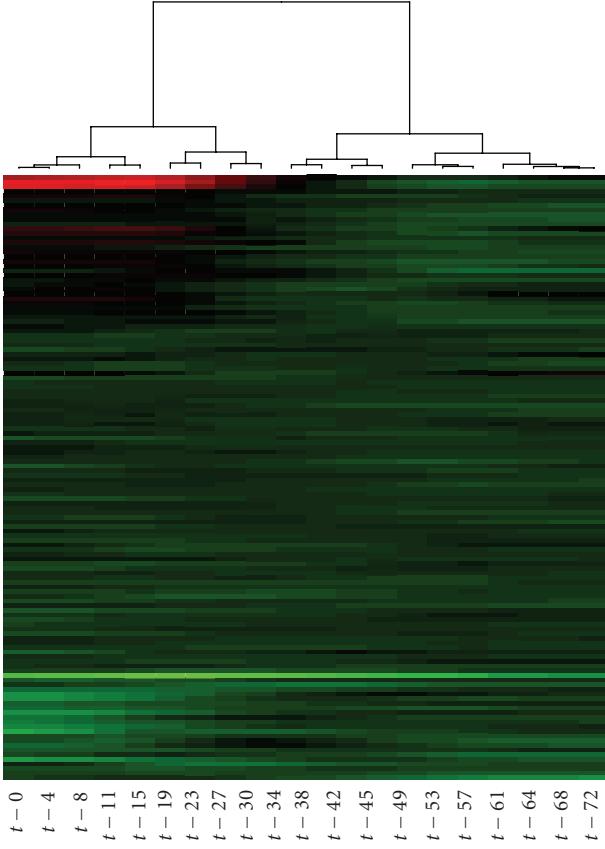


FIGURE 9: Heatmap of smoothed gene expression profiles for the whole dataset. Genes are ordered according to their cluster determined by the  $k$ -means algorithm. Horizontal blue lines separate the 4 clusters. Values increase from green to red via black.

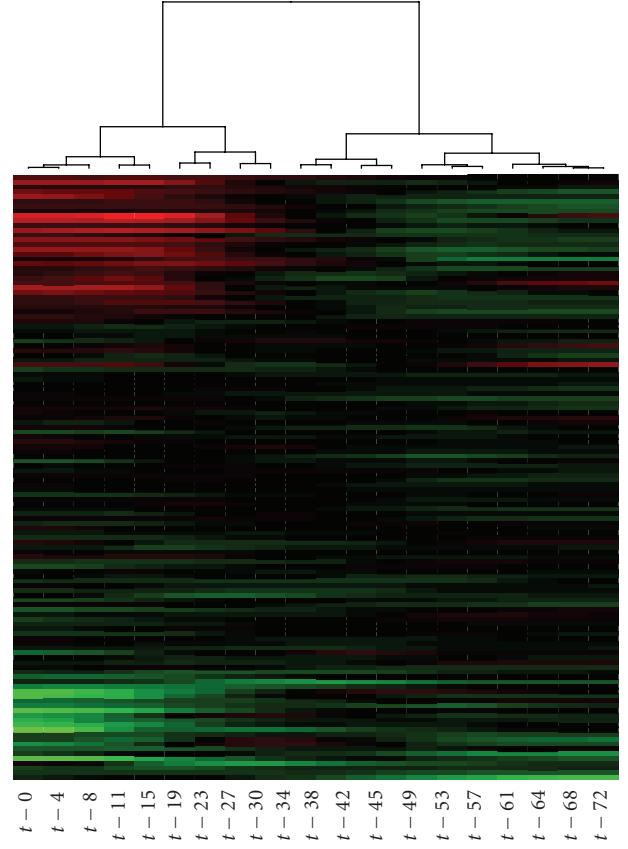


FIGURE 10: Heatmap of smoothed gene expression profiles without SCD1 and km4-genes. Graphical features are the same as Figure 9.

setting up future experiments where time points can be chosen more adequately depending on the scientific aims. Additionally, this work is the starting point of future investigations aiming at delineating the role of various transcription factors such as PPAR $\alpha$  or SREBP in the observed gene expression regulations.

The statistical methodology proposed in the present paper was clearly developed for this specific dataset and its associated scientific aims. Other microarray time-course experiments may benefit from this methodology provided that sufficiently large sample sizes are considered. It is likely that the decreasing cost of microarray technology and the increasing development of cheaper dedicated macroarrays will rapidly yield several suitable time-course datasets.

The dataset studied in this paper and the *R* functions used to perform its analysis are available upon request from the authors.

## ACKNOWLEDGMENTS

The authors are grateful to Thierry Pineau, Romain Barnouin, and Henrik Laurell for interesting discussions about biological interpretation of the results. They thank

Dominique Haughton for critical review of the manuscript and Alice Vigneron for complementary works on this dataset. This work was partially supported by a grant from ACI IMP-Bio.

## REFERENCES

- [1] T. Park, S.-G. Yi, S. Lee, et al., “Statistical tests for identifying differentially expressed genes in time-course microarray experiments,” *Bioinformatics*, vol. 19, no. 6, pp. 694–703, 2003.
- [2] S. D. Peddada, E. K. Lobenhofer, L. Li, C. A. Afshari, C. R. Weinberg, and D. M. Umbach, “Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference,” *Bioinformatics*, vol. 19, no. 7, pp. 834–841, 2003.
- [3] J. D. Storey, W. Xiao, J. T. Leek, R. G. Tompkins, and R. W. Davis, “Significance analysis of time course microarray experiments,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 36, pp. 12837–12842, 2005.
- [4] Y. C. Tai and T. P. Speed, “A multivariate empirical Bayes statistic for replicated microarray time course data,” *The Annals of Statistics*, vol. 34, no. 5, pp. 2387–2412, 2006.

- [5] M. F. Ramoni, P. Sebastiani, and I. S. Kohane, "Cluster analysis of gene expression dynamics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 14, pp. 9121–9126, 2002.
- [6] J. Ernst, G. J. Nau, and Z. Bar-Joseph, "Clustering short time series gene expression data," *Bioinformatics*, vol. 21, supplement 1, pp. i159–i168, 2005.
- [7] C. D. Giurcăneanu, I. Tăbuș, and J. Astola, "Clustering time series gene expression data based on sum-of-exponentials fitting," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 8, pp. 1159–1173, 2005.
- [8] N. A. Heard, C. C. Holmes, D. A. Stephens, D. J. Hand, and G. Dimopoulos, "Bayesian co-clustering of Anopheles gene expression time series: study of immune defense response to multiple experimental challenges," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 47, pp. 16939–16944, 2005.
- [9] A. Conesa, M. J. Nueda, A. Ferrer, and M. Talón, "maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments," *Bioinformatics*, vol. 22, no. 9, pp. 1096–1102, 2006.
- [10] J. Letowski, R. Brousseau, and L. Masson, "Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays," *Journal of Microbiological Methods*, vol. 57, no. 2, pp. 269–278, 2004.
- [11] J. Ramsay and B. Silverman, *Functional Data Analysis*, Springer, New York, NY, USA, 2nd edition, 2005.
- [12] Z. Bar-Joseph, G. K. Gerber, D. K. Gifford, T. S. Jaakkola, and I. Simon, "Continuous representations of time-series gene expression data," *Journal of Computational Biology*, vol. 10, no. 3–4, pp. 341–356, 2003.
- [13] Z. Bar-Joseph, "Analyzing time series gene expression data," *Bioinformatics*, vol. 20, no. 16, pp. 2493–2503, 2004.
- [14] P. G. P. Martin, F. Lasserre, C. Calleja, et al., "Transcriptional modulations by RXR agonists are only partially subordinated to PPAR $\alpha$  signaling and attest additional, organ-specific, molecular cross-talks," *Gene Expression*, vol. 12, no. 3, pp. 177–192, 2005.
- [15] P. G. P. Martin, H. Guillou, F. Lasserre, et al., "Novel aspects of PPAR $\alpha$ -mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study," *Hepatology*, vol. 45, no. 3, pp. 767–777, 2007.
- [16] INRAArray, Laboratoire de Pharmacologie et Toxicologie, INRA, 2005, <http://www.inra.fr/internet/Centres/toulouse/pharmacologie/lpt.htm>.
- [17] B. Silverman, "Some aspects of the spline smoothing approach to non-parametric regression curve fitting," *Journal of the Royal Statistical Society: Series B*, vol. 47, no. 1, pp. 1–52, 1985.
- [18] P. Besse, H. Cardot, and F. Ferraty, "Simultaneous non-parametric regressions of unbalanced longitudinal data," *Computational Statistics & Data Analysis*, vol. 24, no. 3, pp. 255–270, 1997.
- [19] G. A. F. Seber, *Multivariate Observations*, John Wiley & Sons, New York, NY, USA, 1984.
- [20] K. Y. Yeung and W. L. Ruzzo, "Principal component analysis for clustering gene expression data," *Bioinformatics*, vol. 17, no. 9, pp. 763–774, 2001.
- [21] H. Chipman, T. J. Hastie, and T. Tibshirani, "Clustering microarray data," in *Statistical Analysis of Gene Expression Microarray Data*, T. Speed, Ed., pp. 159–200, Chapman & Hall/CRC Press, Boca Raton, Fla, USA, 2003.
- [22] S. Kersten, J. Seydoux, J. M. Peters, F. J. Gonzalez, B. Desvergne, and W. Wahli, "Peroxisome proliferator-activated receptor  $\alpha$  mediates the adaptive response to fasting," *Journal of Clinical Investigation*, vol. 103, no. 11, pp. 1489–1498, 1999.
- [23] S. Mandard, M. Müller, and S. Kersten, "Peroxisome proliferator-activated receptor  $\alpha$  target genes," *Cellular and Molecular Life Sciences*, vol. 61, no. 4, pp. 393–416, 2004.
- [24] M. Bauer, A. C. Hamm, M. Bonaus, et al., "Starvation response in mouse liver shows strong correlation with life-span-prolonging processes," *Physiological Genomics*, vol. 17, no. 2, pp. 230–244, 2004.

**Title:** Evaluation of gene expression regulations induced by the activation of the peroxisome proliferator-activated receptor alpha (PPAR $\alpha$ ).

**Author:** Pascal, GP, MARTIN

### Abstract

The alpha isoform of the peroxisome proliferator-activated receptors (PPAR $\alpha$ ) is a member of the nuclear receptor superfamily (Chapter I). Following its activation by various compounds, PPAR $\alpha$  (Chapter II) regulates the transcription of several target genes, many of which are involved in lipid metabolism. We developed a cDNA macroarray (Chapter III) dedicated to PPAR $\alpha$  and its heterodimerization partner the retinoid X receptor (RXR). This tool was used to study the gene expression regulations induced by the activation of PPAR $\alpha$  through various stimuli which are linked to the fields of : i) pharmacology (PPAR $\alpha$  and RXR-activating drugs, Chapter IV), ii) nutrition (dietary fatty acids, Chapter V) and iii) physiopathology (fasting, Chapter VI). These in vivo studies illustrate the activation of PPAR $\alpha$  by fatty acids provided either by low-fat diets or by adipose tissue lipolysis during fasting. They also demonstrate that the PPAR $\alpha$  signaling pathway is influenced by RXR activation in vivo.

### Keywords

Nuclear receptor, transcriptomic, macroarray, DNA chips, peroxisome proliferator, PPAR, fibrate, rexinoid, fatty acid, fasting, liver

**Titre:** Etude du potentiel de régulation génique exercé par l'isoforme alpha du récepteur activé par les proliférateurs de peroxyosomes (PPAR $\alpha$ )

**Auteur:** Pascal, GP, MARTIN

### Résumé

L'isoforme  $\alpha$  du récepteur activé par les proliférateurs de peroxyosomes (PPAR $\alpha$ ) est un récepteur nucléaire (Chapitre I). Suite à son activation par diverses molécules, PPAR $\alpha$  (Chapitre II) régule l'expression de nombreux gènes impliqués notamment dans le métabolisme des lipides. Grâce au développement d'une puce à ADN (Chapitre III) dédiée à PPAR $\alpha$  et à son partenaire de dimérisation le récepteur X aux rétinoïdes (RXR), nous avons étudié les régulations géniques provoquées par PPAR $\alpha$  lors de son activation par des stimuli i) pharmacologique (activateurs pharmacologiques de PPAR $\alpha$  et RXR, Chapitre IV), ii) nutritionnel (acides gras alimentaires, Chapitre V) et iii) physiopathologique (jeûne, Chapitre VI). Ces travaux montrent que PPAR $\alpha$  est très vraisemblablement l'objet d'une activation physiologique par les acides gras alimentaires ou issus de la lipolyse du tissu adipeux et que ses voies de signalisation subissent l'influence croisée de celles de son partenaire de signalisation RXR.

### Mots clefs

Récepteurs nucléaires, transcriptomique, macroarray, puce à ADN, proliférateur de peroxyosome, PPAR, fibrate, réxinoïde, acide gras, jeûne, foie