
Modélisation de réseaux biologiques et
application pour l'analyse de données
expérimentales

Rapport de DEA de Julien OUY

Table des matières

1	Repères biologiques	5
1.1	Biochimie moléculaire	5
1.1.1	Gènes et facteurs de transcription	5
1.1.2	Réactions enzymatiques	5
1.1.3	Voies métaboliques	6
1.2	Fonctionnement du foie	7
1.2.1	Métabolisme lipidique dans le foie chez le poulet.	7
1.2.2	Expression des gènes dans le foie	8
1.3	Puces à ADN	10
1.3.1	Fonctionnement d'une puce à ADN	10
1.3.2	Matrice exploitable numériquement	11
1.4	Objectifs biologiques	12
2	Présentation des informations issues d'expérimentation	15
2.1	GARMeN	15
2.2	Extraction des contradictions sous forme de texte	16
2.3	Représentation graphique des contradictions	17
2.4	Extraction des sous-graphes	19
2.5	Phénomène de Balance	21
2.5.1	Contradictions relatives à $LXR\alpha$	21
2.5.2	Balance	21
2.5.3	Préalable à la mise en évidence des balances	21
3	Interprétation des sous-graphes	23
3.1	Travail d'analyse sur les sous-graphes	23
3.1.1	Relation d'inclusion : \subseteq	23
3.1.2	Relation d'ordre : \sqsubseteq	24
3.2	Représentation hiérarchique des éléments	25
3.2.1	Choix d'une représentation	26
3.2.2	Représentation des sommets porteurs de contradictions	29
3.2.3	Application et interprétation des résultats	30
4	Analyse dynamique du réseau	32
4.1	Définitions des fonctions utilisées	32
4.2	Algorithme de l'analyse dynamique	32

4.3	Applications	34
4.3.1	Simulation d'un réseau suralimenté	35
5	Traitement du métabolisme	37
5.1	Représentation du métabolisme	37
5.1.1	Représentation des réactions enzymatiques	38
5.1.2	Représentation des voies métaboliques	38
5.1.3	Représentation de l'activation d'une protéine	38
5.2	Extrapolation des régulations à partir du métabolisme	39

Introduction

L'objectif de ce stage de DEA est de concevoir une méthode qui permet de traiter et d'analyser une grande quantité de données biologiques. Ces données issues de puces à ADN décrivent l'expression d'une centaine de gènes concernés par le métabolisme hépatique du poulet et la régulation de ce métabolisme.

L'idée sous-jacente de ce stage est d'exploiter ces données en les mettant en relation avec les données bibliographiques amassées depuis des années afin de les compléter et les valider : cette démarche permettra d'extraire des données des connaissances relatives à la dynamique du système qui ne sont pas fournies par les analyses statistiques classiques.

Ces travaux reposeront sur l'utilisation d'outils d'analyse déjà conçus : une partie importante du métabolisme lipidique et de ses régulations a déjà été extraite de la bibliographie et modélisée sous forme de graphes par des membres du projet Symbiose de l'IRISA (Anne Siegel et Michel Le Borgne) et par des biologistes de l'INRA (laboratoire de génétique animale - Sandrine Lagarigue). Le logiciel GARMeN qui réunit le modèle et les outils, est capable de rechercher des chemins de régulations, des successeurs ou des prédécesseurs métaboliques et d'afficher sous forme de graphes d'interactions ces différents chemins.

Ce travail de DEA ajoute à ce modèle les outils qui permettront d'analyser des données d'expressions biologiques et d'en tirer les informations recherchées par les biologistes. Cela passera par une analyse du contenu des données ainsi que de leur organisation. Une étude portera aussi sur la fiabilité des données expérimentales ainsi que sur celle du modèle utilisé.

Le fonctionnement du métabolisme lipidique ainsi que les méthodes utilisées pour fournir les données seront abordés dans un premier chapitre afin de fournir les bases nécessaires à l'étude du sujet.

Le second chapitre introduira la notion de contradictions dans le réseau et définira les sous-graphes créés pour les représenter.

Ces sous-graphes seront utilisés dans le troisième chapitre dans une méthode d'analyse des données de puce. Nous verrons comment les sous-graphes peuvent être organisés dans des diagrammes basés sur des relations binaires et comment ces diagrammes permettent de détecter les contradictions.

Dans le quatrième chapitre, nous étudierons une méthode dynamique qui permet outre la détection, l'interprétation des contradictions.

Le dernier chapitre conclura ce rapport en introduisant une représentation qui apporte une nouvelle dimension aux graphes de régulation en ajoutant des informations qualitatives de régulation sous les réactions métaboliques.

Chapitre 1

Repères biologiques

Ce travail s'appuie sur une collaboration avec des Biologistes de l'INRA-ENSAR¹ (laboratoire de génétique animale). Leur objectif est d'identifier les gènes responsables de la variabilité de l'engraissement dans deux lignées de poulets de chair, l'une grasse, l'autre maigre. Cette identification passe par une compréhension globale du métabolisme hépatique des lipides chez le poulet. Notre objectif est de concevoir des outils capables de modéliser à la fois un réseau métabolique et un réseau de régulation génétique, et d'intégrer des données bruitées comme des résultats d'expérimentations afin de fournir des renseignements enrichis synthétisant toutes les informations disponibles.

1.1 Biochimie moléculaire

1.1.1 Gènes et facteurs de transcription

Dans chaque cellule de tout organisme, on trouve un exemplaire de son code génétique. Ce code génétique, porté par les brins d'ADN des chromosomes est découpé en plusieurs gènes. Les gènes sont les «plans» qui servent à synthétiser des polypeptides : des protéines ou des fragments de protéines. Ces gènes sont inactifs et ils sont activés pour produire des protéines lorsqu'un besoin de l'organisme en déclenchera la transcription.

La synthèse des protéines se fait à l'extérieur du noyau, dans le cytoplasme, grâce aux ribosomes. Ce n'est pas l'ADN qui est utilisé pour cette transcription mais une copie, l'ARN messager. Ce brin ne mesure que quelques centaines de nucléotides, alors que la longueur totale de l'ADN humain est de trois milliards de nucléotides et celui du poulet de cinq cent millions[DDB04]. Sa taille réduite lui permet de sortir du noyau alors que l'ADN reste protégé à l'intérieur.

La transcription d'un gène est déclenchée par une protéine appelée facteur de transcription. La lecture de l'ARN messager est elle aussi effectuée par une protéine nommée ribosome. On observe ici les premiers stades de la régulation du métabolisme : un gène code un facteur de transcription qui déclenche lui même l'expression de plusieurs gènes et ainsi la synthèse de plusieurs protéines.

1.1.2 Réactions enzymatiques

Une réaction enzymatique est la plus commune des réactions métaboliques. Elle met en jeu trois éléments :

¹Institut National de la Recherche Agronomique - École Nationale Supérieure Agronomique de Rennes

- **Une enzyme** est une protéine synthétisée par un être vivant à partir d'un gène afin de catalyser une réaction chimique. Une enzyme est spécifique à une réaction : elle catalyse toujours la même transformation. Elle n'est pas modifiée par la réaction et peut donc catalyser la réaction jusqu'à épuisement des substrats.
- **Un substrat** est une molécule qui entre dans une réaction pour y être transformée grâce à l'action catalytique d'une enzyme. Toutes les molécules qui sont modifiées par une réaction enzymatique sont appelées substrats.
- **Un produit** est une molécule qui apparaît lors d'une réaction enzymatique.

Modélisation mathématique Une seule enzyme catalyse plusieurs fois la même réaction mais elle est limitée par une vitesse de travail maximum obtenue lorsque le substrat est en concentration optimale. L'équation de Michaelis et Menten permet d'obtenir la vitesse réelle de travail liée à la concentration du substrat.

$$\frac{dS}{dt} = -\frac{dP}{dt} = v = V_{max} \frac{[S]}{K_m + [S]} \quad (1.1)$$

Dans cette équation, $[S]$ représente la concentration de substrat ; v et V_{max} sont des vitesses, v est la vitesse réelle pour une concentration donnée, V_{max} est la vitesse maximale théorique, obtenue pour un dosage enzymatique fini et pour une concentration de substrat infinie ; K_m est la constante de Michaelis. Propre à chaque enzyme, elle est inversement proportionnelle à l'affinité de l'enzyme pour le substrat, elle représente la concentration de substrat pour laquelle la vitesse réelle vaut la moitié de la vitesse maximale. Cette équation nous sera utile lorsqu'il sera nécessaire de traiter des informations métaboliques sous forme de régulation.

Activation des enzymes On notera que la transcription d'un gène est un processus lent. Pour utiliser ses enzymes au bon moment, l'organisme a recours à une astuce : la plupart des enzymes sont produites par l'organisme et conservées sous une forme inactive. Un procédé d'activation, processus rapide, est nécessaire pour libérer la forme active des enzymes capables de catalyser une réaction. Ces procédés sont appelés modifications post-transcriptionnelles et sont au nombre d'une centaine. On peut citer l'association avec une molécule tierce (ligand), la coupure d'un bout de l'enzyme (clivage), un changement de conformation spatiale (régulation allostérique) ou l'ajout d'un élément phosphate (phosphorylation).

1.1.3 Voies métaboliques

Une voie métabolique est un ensemble de réactions métaboliques successives aboutissant à la production d'un composé biologique ayant une fonction indispensable pour l'organisme qui les réalise. Elle est constituée de plusieurs réactions catalysées par des enzymes diverses. Chacune de ces enzymes peut éventuellement participer à plusieurs voies métaboliques. On peut entrer dans une voie métabolique à partir d'un substrat initial ou à partir du produit d'une autre voie métabolique. On distinguera alors deux types de produits impliqués dans les réactions métaboliques :

- les produits intermédiaires, il ne sont souvent pas utilisés autrement qu'au sein de leur voie métabolique, on considère de plus qu'ils ne sont pas stockés. On peut donc estimer que leur flux est nul : $\frac{dS}{dt} = 0$.

- les produits finaux ou «carrefours métaboliques», il s’agit de produits susceptibles d’être transportés ou bien de produits qui peuvent être le substrat de plusieurs enzymes appartenant à des voies métaboliques différentes.

Il est important de souligner que ces voies métaboliques sont généralement régulées de manière globale, c’est-à-dire qu’une ou plusieurs protéines, elles-mêmes codées par des gènes, sont des facteurs de transcription qui assurent l’activation ou l’inhibition de l’expression de toutes les enzymes de la voie.

1.2 Fonctionnement du foie

1.2.1 Métabolisme lipidique dans le foie chez le poulet.

Le foie des volailles est un organe de régulation des réserves énergétiques et de synthèse des acides gras. Son fonctionnement est déterminé par les besoins de l’organisme et est illustré très schématiquement par la figure 1.1.

Lorsque le poulet est alimenté, le foie synthétise des acides gras à partir du glucose et les envoie vers les cellules adipeuses où ils seront stockés. Une partie du glucose est aussi transformée en énergie pour les besoins immédiats de l’organisme. Cette voie est décrite en rouge sur le schéma 1.1.

Au contraire, lorsque l’animal n’est plus alimenté, le fonctionnement du foie change : il se met à dégrader les acides gras stockés pour fournir de l’énergie au reste de l’organisme (voie illustrée en vert).

Ces réactions biochimiques sont régulées : si les réactions de stockage et de destockage des acides gras ont lieu simultanément, le cycle ne fournira pas d’énergie et sera inutile voire néfaste. Il a été prouvé que toutes ces réactions sont contrôlées en amont par l’expression de gènes régulateurs (flèches bleues), elle-même déclenchée par la présence de certains métabolites (flèches violettes).

On notera d’une part que l’expression des gènes a une influence sur la synthèse et la dégradation des acides gras et que d’autre part la quantité de ces acides gras détermine en retour l’expression ou non de ces gènes : ce sont des boucles de rétrorégulation.

La figure 1.2 offre une vision plus complète des effets des gènes LXR, SREBP et des acides gras polyinsaturés AGPI sur le métabolisme. Elle schématise une partie des données bibliographiques disponibles sur le sujet.

On remarque en noir les voies métaboliques, processus les plus étudiés par la littérature scientifique et donc bien connus.

En bleu on observe les effets de la protéine LXR α vu comme un facteur de transcription. Deux types de régulations sont à distinguer :

- Des régulations directes, colorées en bleu. Un mécanisme détaillé a été mis en évidence qui prouve que LXR α se fixe sur un promoteur des gènes en question pour activer leur transcription.
- Des régulations indirectes marquées par des \oplus et des \ominus bleus. Des publications ont mis en évidence qu’augmenter la quantité de LXR α provoque une augmentation de la protéine cible. Cependant, aucun mécanisme n’a été mis en évidence pour expliquer ce comportement. En particulier, pour ce genre d’information, il est possible que l’effet de LXR α sur sa cible soit la conséquence d’une succession de régulations.

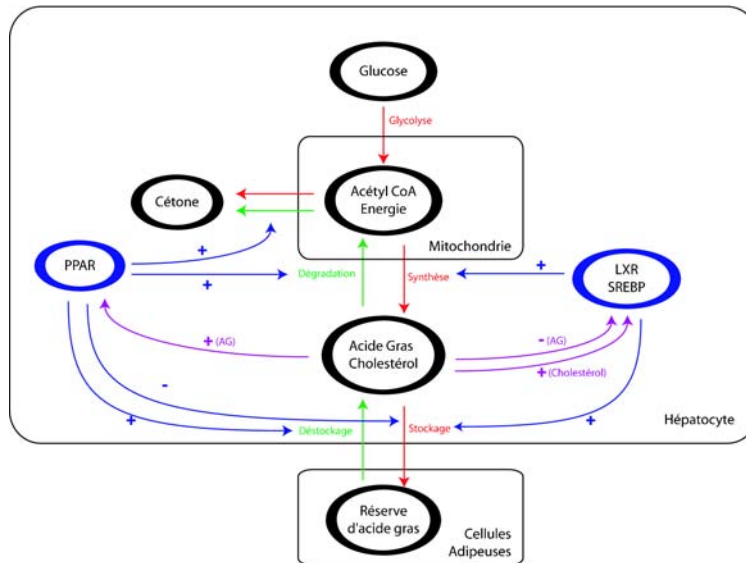


FIG. 1.1 – Schéma simplifié du métabolisme du foie.

De même, les gènes dont les noms sont coloriés en violet sont directement régulés par la protéine SREBP. Ils sont tous impliqués dans la synthèse des acides gras : ACL, ACC, FAS, SCD1. Ceux marqués d'un \oplus violet sont liés indirectement. On retrouve ici une illustration de la notion de régulation indirecte : LXR α agit directement sur SREBP qui agit directement sur ACL, ce qui explique l'action indirecte de LXR α sur ACL.

Les acides gras ont une influence indirecte et inhibitrice sur le processus de leur synthèse représentée par des \ominus orange.

1.2.2 Expression des gènes dans le foie

Des expérimentations ont montré que les différents états du foie (synthèse ou dégradation des acides gras) sont liés à la quantité de certaines protéines dans le foie, c'est-à-dire l'expression des gènes correspondants. Une diète prolongée ou une suralimentation entraîne une augmentation ou une diminution de l'activité de certains gènes, mise en évidence par la présence des ARN messagers correspondants.

L'expérience présentée Figure 1.3 et réalisée au laboratoire INRA-ENSAR de génétique animale présente un de ces liens. Les interdépendances de deux gènes nommés FAS et ApoA1 sont illustrées par ces lignes dont les tâches successives représentent différentes conditions d'alimentation :

- état normal,
- 16 heures de jeun,
- 48 heures de jeun,
- 16 heures de jeun puis 5 heures après réalimentation,
- 16 heures de jeun puis 21 heures après réalimentation,
- 48 heures de jeun puis 24 heures après réalimentation.

On constate que les ARN messagers du gène FAS sont très présents lorsque ceux de ApoA1 sont presque inexistants : ces gènes sont corégulés.

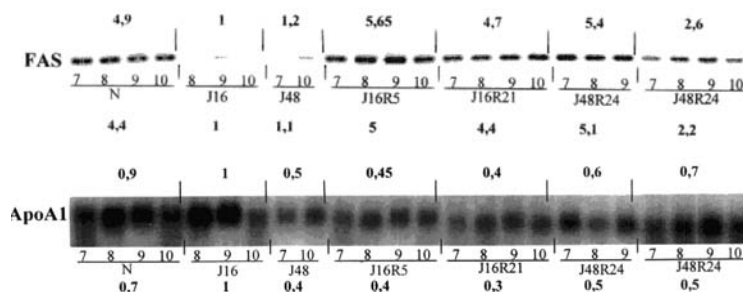


FIG. 1.3 – Résultats d’expériences sur l’expression des gènes FAS et ApoA1 dans une cellule du foie chez le poulet.

Cette expérience nous montre qu’il existe un phénomène de corégulation entre ApoA1 et FAS mais ne donne pas d’explication. Il peut s’agir d’un lien direct : l’un des gènes paraît synthétiser une protéine qui bloque l’expression de l’autre ; ou d’un lien indirect : la présence d’un gène paraît déclencher une cascade de réactions chimiques ou génétiques qui aboutit au blocage de l’autre. Il peut s’agir aussi de deux effets d’une troisième réaction : la présence d’un gène ou d’une protéine pourrait entraîner celle de FAS et ApoA1.

Seul un modèle complet contenant des réactions métaboliques et les régulations génétiques des enzymes qui interviennent dans ces voies peuvent nous renseigner sur la nature exacte de ce lien mais aussi sur des dizaines d’autres présents dans le foie et plus généralement dans l’organisme. L’état actuel des connaissances permet de remplir une partie du modèle mais l’expression des gènes et leur influence restent encore assez incomprises.

La littérature nous renseigne tout de même sur certaines fonctions des gènes : on sait par exemple que l’expression du gène ApoA1, chez l’homme, implique le stockage des acides gras en participant au transport du cholestérol [PRM⁺02].

1.3 Puces à ADN

Au milieu des années quatre-vingt dix sont apparues les puces à ADN qui permettent au biologiste de « relever » l’expression de plusieurs milliers de gènes à un instant donné et ainsi d’observer un processus biologique dans sa globalité. On dispose ainsi de données réelles qui permettent de valider le modèle théorique donné par un réseau d’interactions.

1.3.1 Fonctionnement d’une puce à ADN

Le concept de puce à ADN, s’il rappelle l’informatique, n’est pas un composant électronique. Une puce à ADN est un support semblable aux lamelles utilisées pour les microscopes sur lequel sont placés des milliers de fragments d’ADN. Ces fragments, correspondant chacun à un unique gène, sont placés dans une grille à l’aide d’une micro-pipette robotisée. Lorsque ces fragments sont mis en contact avec l’ARN messager présent dans les échantillons biologiques à tester, l’ADN complémentaire issu de la retranscription de l’ARN messager se fixe au brin d’ADN par hybridation

(figure 1.4). Si l'ARN messager a préalablement été étiqueté à l'aide d'un marqueur fluorescent, la présence du gène sera mise en évidence par la coloration du résultat de l'hybridation dans une matrice de points comme sur la figure 1.5.

Une analyse de l'intensité de la fluorescence en mesurant la quantité d'ARN messager présent dans l'échantillon par rapport à une quantité de référence permet de déterminer l'expression d'un gène [BV00].

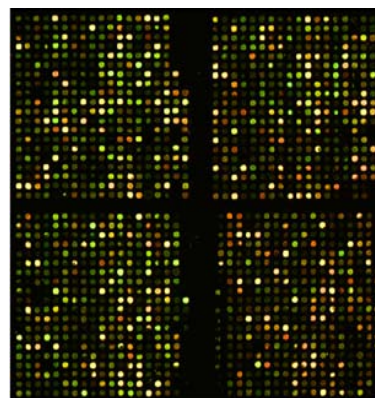
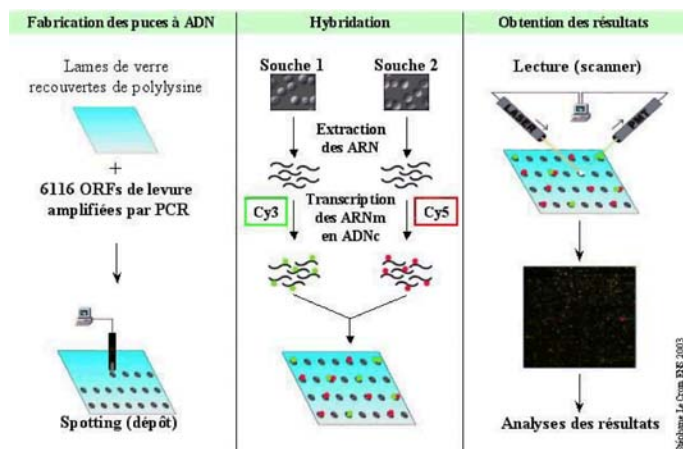


FIG. 1.4 – Principe d'une puce à ADN illustré dans le cas de la levure. FIG. 1.5 – Matrice de points.

1.3.2 Matrice exploitable numériquement

Les valeurs de la matrice de points sont réunies avec les résultats de plusieurs expériences sur les mêmes gènes. On obtient alors une matrice d'expression de gènes dont les milliers de lignes représentent les gènes et les quelques colonnes les expériences (différence de tissus, stade de développement, traitement...). Chaque cellule contient une valeur caractérisant le niveau d'expression d'un gène particulier dans une condition d'expérience particulière.

On peut ensuite soumettre cette matrice à une analyse statistique afin de regrouper les différents gènes par famille et constater des antagonismes ou des liens entre les gènes. L'équipe de Jean Léger de l'Inserm² à Nantes travaille sur une suite logicielle dédiée à cette analyse [TBM⁺03].

Des méthodes de clustering³ [BV00] permettent d'extraire des informations de ces matrices. En regroupant par famille les gènes dont l'expression varie de manière similaire lors des expériences, on peut associer ces familles de gènes à des voies métaboliques. Ces méthodes de classification sont utilisées en médecine pour certains diagnostics. L'observation des familles de gènes actives et inactives permet de différencier certaines maladies et de choisir un traitement adapté. Ces associations sont malheureusement incomplètes : le rôle des gènes dans les voies métaboliques est encore inconnu. De plus, de nombreux gènes restent isolés. Ces techniques, malgré d'excellentes applications médicales, n'apportent pas aux biologistes les résultats qu'ils attendent des puces à ADN. La richesse des données qu'elles apportent laisse en effet supposer une meilleure analyse.

²Institut National de la Santé et de la Recherche Médicale

³Regroupement

La confrontation des matrices d'expression de gènes avec un modèle de réseau génétique devrait fournir différentes informations sur les relations entre les gènes. Dans ce contexte, le modèle peut être vu comme une carte qui peut guider le biologiste ou le médecin lors de la lecture des données de puce. Les données des matrices peuvent aussi être utilisées pour valider un modèle ou pour souligner ses incohérences. On pourra alors répondre aux attentes des biologistes qui souhaitent analyser des gènes isolés comme des familles de gènes et connaître leur rôle sur le métabolisme et les facteurs qui les font s'exprimer.

1.4 Objectifs biologiques

Objectif

L'un des buts de ce travail est de fournir aux biologistes de l'INRA les moyens de déterminer l'origine génétique de l'opposition entre des lignées de poulets gras et maigres. Ces poulets sont issus d'élevages surveillés et dirigés par l'INRA depuis trente ans et une sélection divergente a permis d'obtenir des lignées dont la proportion de masse grasse diffère de plus de trente pourcents.

Données expérimentales à analyser

Dans le cadre de cette collaboration, on dispose de données de puces à ADN relatives à des expérimentations sur des populations de poulets gras ou maigres parmi lesquelles coexistent des populations de poulets nourris ou à jeun. Ces données se présentent sous la forme de tables contenant des listes de gènes et des valeurs issues d'une analyse en composantes principales : cette analyse statistique place les données dans un repère à n dimensions où n est le nombre de variables relatives à une valeur, calcule une direction dans laquelle le nuage de points semble le mieux aligné (c'est l'axe pour lequel la rotation du nuage de points offre le moins de résistance) et fournit des valeurs relatives à la position des points par rapport à cet axe.

L'analyse repose sur une quantification de l'intensité lumineuse des spots de la puce à ADN, elle compare des résultats d'expérimentations relatifs à deux populations : soit des poulets gras par rapport à des poulets maigres ; soit des poulets à jeun par rapport à des poulets nourris. La valeur de la coordonnée sur l'axe principal correspondra donc au degré d'expression des gènes observés. Si la coordonnée est positive, le gène sera considéré activé pour l'expérience, si la coordonnée est négative, le gène sera inactivé.

Ces données sont séparées en trois parties, classées selon une valeur de l'analyse en composante principale, la p -value. Ces trois parties renseignent sur la fiabilité des données : une valeur située dans la première partie sera considérée comme sûre, une valeur de la seconde partie pourra être faillible mais reste relativement crédible, on se méfiera des valeurs situées dans la dernière partie, leur position sur l'axe de l'analyse laissant la place au doute, au bruit.

De ces tables de valeurs, nous retiendrons deux choses :

- le sens de corrélation : il s'agit du signe de la coordonnée principale ;
- la fiabilité, bonne, moyenne ou faible selon la position de la valeur dans le tableau.

Données bibliographiques

Pour interpréter ces données de puces, nous travaillerons principalement sur deux réseaux.

Le premier représente une grande partie de la bibliographie connue sur la régulation du métabolisme lipidique. Ce réseau est de taille importante, ce qui en fait un matériau riche pour les biologistes mais qui handicape lors de la mise au point d'algorithmes et leur description. De plus ce réseau est en perpétuelle évolution et les données qu'il contient ne sont pas systématiquement liées à des données de puce.

Le second réseau est un extrait du premier. Il ne contient qu'une dizaine de produits et possède une table de données de puce adaptée. Ce réseau possède un intérêt biologique limité mais non nul car les produits qu'il fait intervenir sont au cœur de problèmes biologiques encore d'actualité. On possède de plus des données de puce pour chacun de ses sommets. Ce réseau est représenté dans la figure 1.6. Dans ce réseau, on fait aussi figurer des données issues de puces à ADN, provenant de l'expérience Fasting : jeun qui suppose la présence d'acides gras et mesure la présence et l'absence de certains gènes impliqués.

Au sein de ce modèle, on notera que l'expérimentation a détecté une activité de la protéine ACL (flèche $Fasting \rightarrow ACL$) alors que la bibliographie prédit une inhibition d'ACL via $Fasting \rightarrow SREBP \rightarrow ACC$ (comme pour les enzymes ACC et SCD1)

Notre objectif est de retrouver cette erreur grâce à une analyse des données.

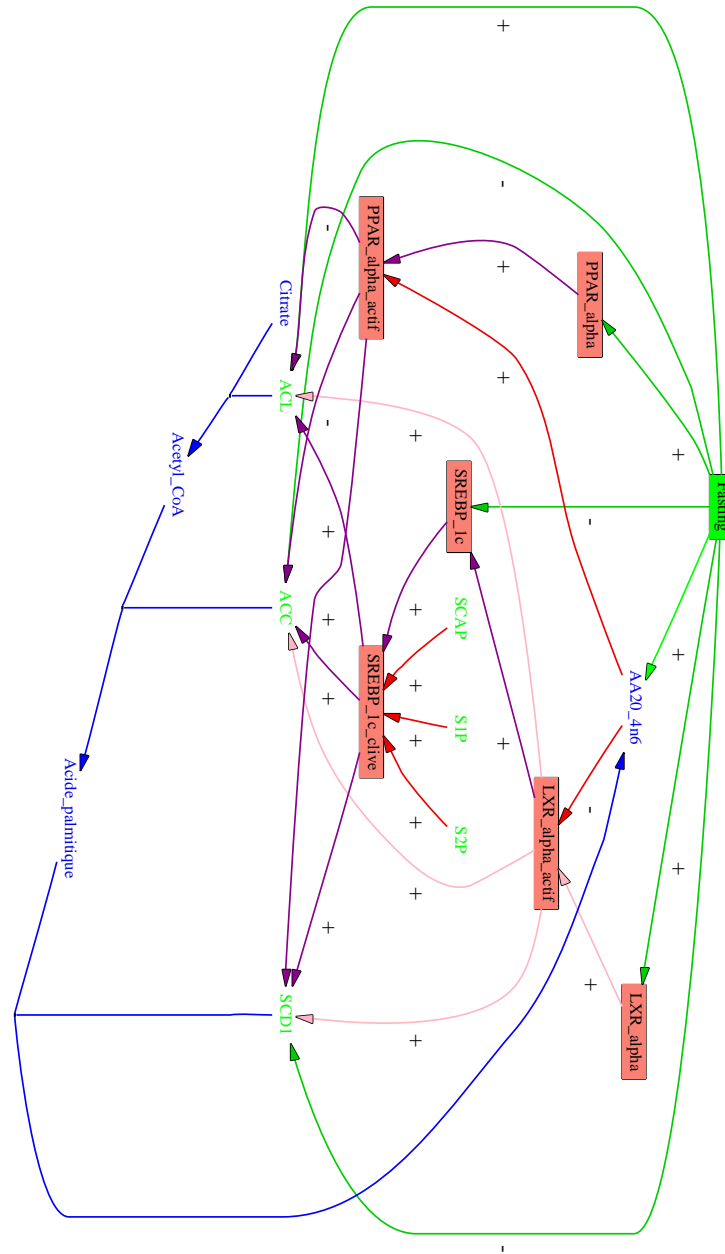


FIG. 1.6 – Réseau simplifié de la synthèse des acides gras.

Chapitre 2

Présentation des informations issues d'expérimentation

2.1 GARMeN

Les données bibliographiques sont de deux sortes :

- **chimie ou biochimie** : mécanisme décrivant la transformation des molécules. On peut citer les réactions enzymatiques, les associations et les transports.
- **comportementale** : mécanisme décrivant l'effet de l'augmentation de la quantité d'un produit ou d'un autre.

Certaines interactions font le lien entre ces deux catégories et peuvent apparaître dans les deux :

- une régulation directe traduit à la fois des informations comportementales et des informations biochimiques (fixation sur un promoteur, déclenchement d'une réaction).
- une transformation post-traductionnelle est un phénomène chimique (ligand, clivage) qui a un effet d'activation d'une protéine.

Ces données bibliographiques sont représentées au sein de la suite logicielle GARMeN développée dans un même réseau sous forme de graphe bipartite, c'est-à-dire un graphe constitué de deux types de sommets.

- Des nœuds produits, relatifs à des métabolites, des protéines inactives ou des protéines actives. Dans une telle structure, on admettra aussi des nœuds «expérimentation». Une expérimentation sera ainsi assimilée à un «meta-produit» ayant des effets qualitatifs sur un grand nombre de cibles.
- Des nœuds interactions classifiés en biochimie, en comportementale ou en expérimentale. Parmi les interactions biochimiques on distinguera les réactions enzymatiques (qui auront l'enzyme pour paramètre), les transports ou des boîtes noires biochimiques (cycle de Crebs). Parmi les réactions comportementales se trouvent des réactions génétiques directes, indirectes et des actions biochimiques indirectes (une réaction génétique agit sur les ARN du produit, une action biochimique agit sur sa forme mature). Enfin, les expérimentations distinguent les effets observés (Fasting augmente ACC) et les hypothèses expérimentales (le jeun augmente les acides gras, non observé mais considéré comme point d'entrée dans le réseau).

Les données étant synthétisées dans une unique modélisation, la question que posent les biologistes réside sur la cohérence des données expérimentales vis à vis des connaissances bibliographiques.

Par exemple, dans le schéma de la figure 1.6 on notera que le jeun fait augmenter ACL alors qu'il fait diminuer AA20, qui fait diminuer LXR, SREBP puis ACL. Il y a là une contradiction qui mérite d'être montrée au biologiste pour en comprendre la nature et identifier éventuellement des mécanismes nouveaux.

Définition 1 (Contradiction) Soit S_1 une expérience et S_2 une cible de cette expérience toutes deux associées à des sommets d'un graphe de régulation. Soit ε le signe d'une arête directe de S_1 vers S_2 . Un chemin qualitatif est dit «*contradictoire*» de S_1 vers S_2 si le produit des signes porté par ce chemin est différent de ε . Il est dit explicatif si le produit des signes porté par ce chemin est égal à ε . Ces chemins sont représentés respectivement en rouge ou en bleu comme présentés sur la figure 2.1.

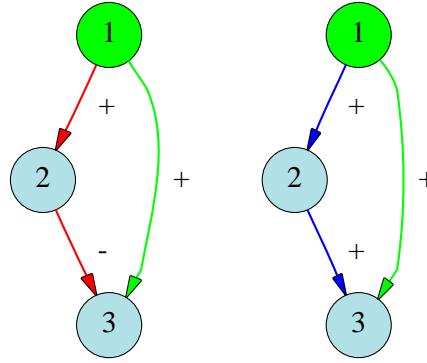


FIG. 2.1 – Gauche : une contradiction. Droite : une explication.

Définition 2 (Produit de signe) On définit le produit de signe, \otimes sur $\Sigma \times \Sigma$ comme suit :

$$\Sigma = \{+, -, 0, 0_{+-}\}$$

$$+ \otimes + = +$$

$$+ \otimes - = -$$

$$- \otimes - = +$$

$$\forall \varepsilon \in \Sigma, 0 \otimes \varepsilon = 0$$

$$\forall \varepsilon \in \Sigma - \{0\}, 0_{+-} \otimes \varepsilon = 0_{+-}$$

Notre objectif est de mettre en avant trois types de contradictions et si possible de les différencier.

- il peut s'agir d'une erreur dans le modèle ;
- d'une erreur dans l'expérimentation ;
- d'un phénomène biologique réel, par exemple une balance : deux comportements contradictoires coexistent et c'est un élément extérieur qui va en privilégier un.

2.2 Extraction des contradictions sous forme de texte

Une représentation sous format texte des contradictions et explications se fait déjà dans GAR-MeN via la méthode «*explication d'expérience*». Il s'agit de parcourir le graphe, de repérer les contradictions et les explications et de les afficher une par une dans un fichier texte.

Pour construire ce fichier, on choisit une donnée expérimentale : $Exp \rightarrow S$, et on recherche le sommet S dans le graphe en commençant par le sommet Exp . On effectue un parcours en profondeur, c'est-à-dire qu'on applique la procédure suivante :

```
Procédure parcourt(s : sommet) ;
var t : sommet ;
Début
    etat(s)=exploré ;
    Pour chaque sommet t dans successeurs[s] faire
        Si etat(t)=inexploré alors
            parcourt(t) ;
fin
```

Pendant ce parcours, une pile conserve la trace de tous les sommets traversés, ainsi lorsque le sommet S est trouvé, on peut écrire dans le fichier de sortie le chemin parcouru.

Un extrait de fichier est présenté figure 2.2. Cet extrait de fichier nous montre tous les chemins composés de régulations partant de l'expérience Fasting (diète) et finissant en ACC. Les deux premiers chemins sont compatibles, c'est-à-dire que le produit des signes composant chacun des chemins est compatible avec le résultat de l'expérience. Le troisième est contradictoire, le chemin calculé s'oppose au chemin référence.

La diminution de l'enzyme ACC peut s'expliquer par une augmentation des Acides gras (AA20) qui provoque une inhibition de $LXR\alpha$ et donc une inhibition de ACC. La même inhibition a aussi lieu avec SREBP comme intermédiaire à la place de $LXR\alpha$.

La contradiction repérée repose sur $LXR\alpha$, on détecte son augmentation dans l'expérience et on sait qu'il active la transcription de ACC. Pourtant l'expérience montre que ACC diminue.

Ce type de fichier remplit sa fonction, il renseigne bien sur les contradictions et les explications. Cependant une telle représentation n'est pas satisfaisante : elle contient des redondances et sous-entend que le biologiste dispose à côté d'une représentation du graphe qu'il étudie. On constate que dans notre exemple $LXR\alpha$ est impliqué à la fois dans une explication et dans une contradiction. On souhaite mettre en valeur cette double implication.

2.3 Représentation graphique des contradictions

Le modèle utilisé reposant sur un graphe, nous proposons une représentation graphique basée sur plusieurs caractéristiques :

- **Division des contradictions** Le fichier texte fournit cinquante explications et dix contradictions, ce qui est inexploitable. On rassemble ces informations en autant de sous-graphes qu'il existe de produits cibles. Chaque sous-graphe contient tous les chemins reliant l'expérimentation au produit cible.
- **Code couleur comportemental** Un exemple de graphe obtenu est donné figure 2.3. Le code couleur utilisé est alors relatif aux types d'interactions (direct, indirect ...). Cependant l'utilisateur est concerné par les cohérences et incohérences plutôt que les types d'interactions. On propose donc de représenter les chemins contradictoires en rouge et les chemins explicatifs en bleu. On conserve le vert pour la donnée de puce qui sert de référence. Un exemple est donné figure 2.4.

```

COMPATIBLE PATH FOR:
.....
Fasting (Experiment)
|
| | [+] (Hypothesis)
| v
| AA20_4n6
| | [-] (direct biochemical regulation)
| v
| LXR_alpha
| | [+] (observed genetic action)
| v
| [-]
|-----> ACC (Enzyme)
(Observation)

COMPATIBLE PATH FOR:
.....
Fasting (Experiment)
|
| | [+] (Hypothesis)
| v
| AA20_4n6
| | [-] (observed genetic action)
| v
| SREBP_1c_clive
| | [+] (direct genetic regulation)
| v
| [-]
|-----> ACC (Enzyme)
(Observation)

POTENTIAL CONTRADICTIONARY PATH FOR:
.....
Fasting (Experiment)
|
| | [+] (Observation)
| v
| LXR_alpha
| | [+] (observed genetic action)
| v
| [-]
|-----> ACC (Enzyme)
(Observation)

```

FIG. 2.2 – Sortie texte de l'analyse de données de puce concernant l'enzyme ACC.

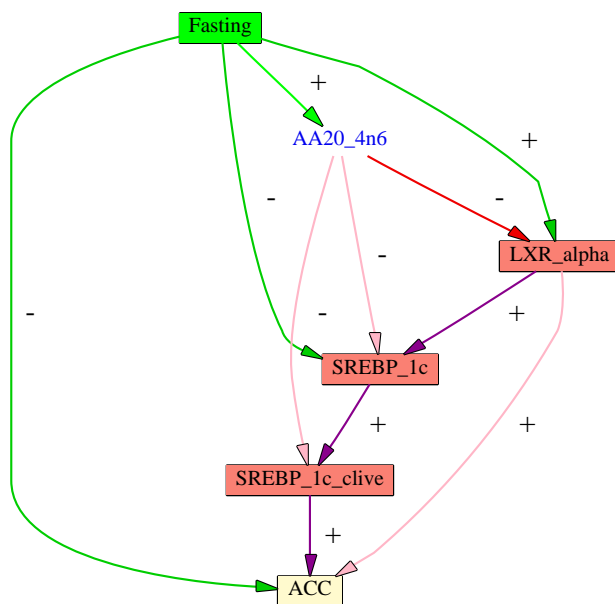


FIG. 2.3 – Graphes des données de puces concernant l'enzyme ACC, représentation type d'interactions.

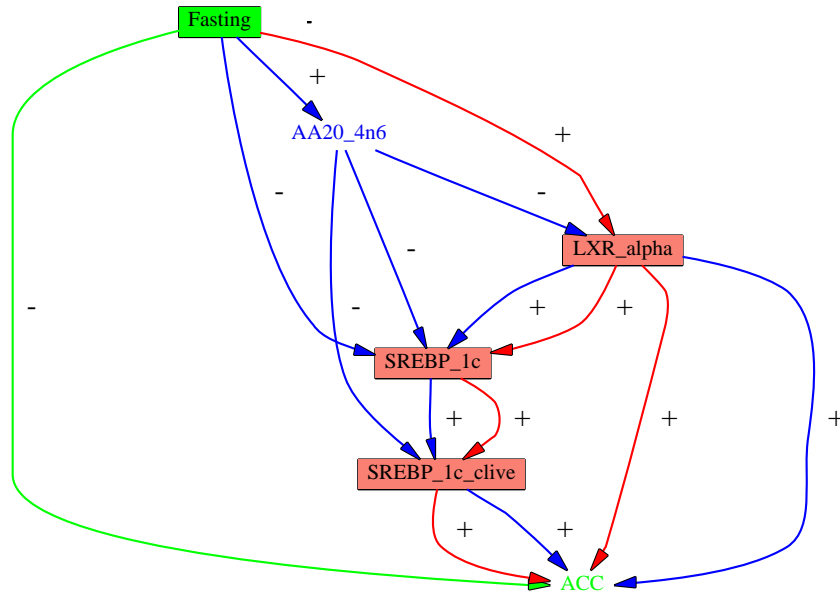


FIG. 2.4 – Graphes des données de puces concernant l'enzyme ACC, représentation des chemins.

Arcs multiples Avec cette représentation on remarque quelque chose d'important : les arcs $LXR\alpha \rightarrow ACC$, $LXR\alpha \rightarrow SREBP$, $SREBP \rightarrow SREBP$ clivé et $SREBP$ clivé $\rightarrow ACC$ sont représentés en double. Cela signifie que, en fonction des antécédents, ces chemins seront empruntés pour des explications ou pour des contradictions. Cela signifie aussi qu'il y a là un phénomène qui mérite que le biologiste s'y attarde.

2.4 Extraction des sous-graphes

Dans le cadre de l'analyse des contradictions, on va extraire un sous-graphe pour chacune des données de l'expérience. Ce sous-graphe conservera sa filiation avec la donnée d'expérience de manière interne : nous parlerons de sous graphe ou de graphe d'explication lié à une paire. En effet, on étudie une contradiction dans un graphe comme une contradiction dans une paire de produits (par exemple $Fasting - > FAS$). Les sous-graphes construits possèdent pour une paire donnée tous les chemins de régulations qui mènent d'un sommet à l'autre.

Dans un souci de clarté, lors de la construction des graphes, on coloriera chaque chemin en rouge s'il s'agit d'une contradiction et en bleu s'il s'agit d'une explication. Le vert sera réservé aux données de l'expérience qui servent de référence aux contradictions et explications.

La construction des sous-graphes repose sur l'utilisation du patron de conception «Visiteur». Un visiteur est un objet qui parcourt un ensemble d'objets prévus à cet effet. Le visiteur doit avoir la méthode `-parcourt(objet)` et l'objet parcouru doit avoir la méthode `-accepte(visiteur)`. Cette méthode se contente simplement d'appeler la méthode `parcourt` du visiteur et de l'appliquer sur l'objet lui-même.

```
Classe Visiteur{
    public void parcourt(Objet O){
```

```

    // travail a effectuer avec l'objet
}
}

Classe Objet{
    public void accepte(Visiteur V){
        V.parcourt(this);
    }
}

```

Le visiteur effectue le parcours du graphe en profondeur, visitant un par un tous les successeurs d'un noeud. Il s'arrête lorsqu'un noeud n'a pas de successeur ou qu'il a déjà été visité. Les noeuds successifs qui forment un chemin sont empilés ainsi que les signes successifs. Lorsqu'un chemin est trouvé, on calcule le signe final et on marque tous les sommets et les arcs qui composent le chemin parcouru. Ces arcs sont créés avec la bonne couleur et l'étiquette correspondant à la régulation qu'ils représentent.

La construction du graphe doit tenir compte des boucles. On fera en sorte qu'un chemin ne passe pas deux fois par le même sommet mais qu'aucun chemin ne soit oublié. Le graphe présenté dans la figure 2.5 compte quatre chemins. Les chemins 1-2-3-2-4 et 1-3-2-3-4 ne sont pas valides. Ce critère s'explique par le fait que ces graphes représentent des systèmes statiques dans lesquels un sommet a un état unique, qu'il soit déterminé ou non.

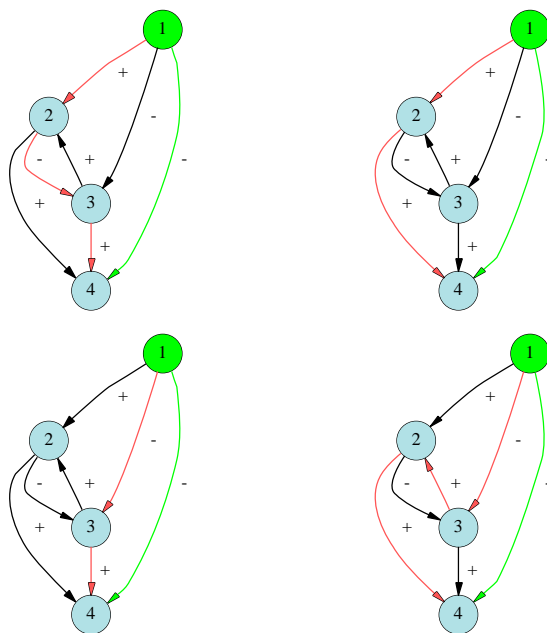


FIG. 2.5 – Y aller par quatre chemins.

L'algorithme extrait donc des sous graphes qui correspondent à une paire donnée et qui contiennent tous les chemins possibles pour aller d'un sommet à l'autre (y compris les boucles éventuelles). Les graphes sont placés dans une nouvelle classe, GraphTree, qui s'occupe aussi du tri des graphes selon une hiérarchie que nous étudierons plus tard.

2.5 Phénomène de Balance

2.5.1 Contradictions relatives à $LXR\alpha$

Dans l'exemple de la figure 2.4, on trouve plusieurs contradictions :

- sur $LXR\alpha$ la donnée de puce (en vert) s'oppose au comportement prédit par le modèle : l'inhibition par les acides gras AA20.
- sur SREBP et ACC, la donnée de puce s'oppose aux chemins positifs issus de $LXR\alpha$.

Le biologiste se concentrera sur la contradiction qui porte sur $LXR\alpha$ puisqu'elle montre une incohérence entre le modèle et les données.

En revanche, les contradictions sur ACC sont moins significatives : elles sont toutes issues de la première contradiction, les deux chemins contradictoires étant des prolongements du chemin contradictoire qui mène à $LXR\alpha$. Ceci laisse penser qu'il vaut mieux favoriser le chemin Fasting \rightarrow AA20 \rightarrow $LXR\alpha$ plutôt que la donnée Fasting \rightarrow $LXR\alpha$.

2.5.2 Balance

Cependant, il s'avère que les chemins Fasting \rightarrow AA20 \rightarrow $LXR\alpha$ et Fasting \rightarrow $LXR\alpha$ sont aussi exacts l'un que l'autre. L'explication biologique repose sur la notion de compétition ou balance : deux effets contradictoires coexistent sur un même produit et un seul sera prépondérant en fonction du contexte.

La mise en évidence des lieux de balance ainsi que leur interprétation est une question non triviale et très intéressante du point de vue des biologistes. Nous allons proposer des méthodes pour les mettre en évidence dans les chapitres suivants.

2.5.3 Préalable à la mise en évidence des balances

Pour mieux expliquer et représenter la balance portant sur $LXR\alpha$, on a précisé le modèle en lui ajoutant un sommet : $LXR\alpha$ actif, (figure 2.6), c'est-à-dire la forme active de $LXR\alpha$. La molécule de $LXR\alpha$ subit en effet une transformation entre le moment où elle est synthétisée et le moment où elle va s'exprimer. La puce à ADN détecte une quantité d'ARN messenger utilisée pour synthétiser la molécule inactive. La détection de l'ARN est suffisante pour que l'on soit sûr de la présence de la molécule mais l'expérience ne nous dit pas si la molécule sera plus tard activée.

On sait par d'autres connaissances bibliographiques que l'activation de $LXR\alpha$ est inhibée par les acides gras. Finalement, le phénomène de compétition porte sur $LXR\alpha$ actif, entre l'effet de Fasting qui augmente la quantité de protéine inactive, et l'effet de Fasting qui augmente la quantité d'acides gras susceptibles de bloquer la transformation de $LXR\alpha$ immature en protéine active. En ajoutant le sommet $LXR\alpha$ actif, on ne supprime pas la contradiction car elle existe réellement mais on peut désormais l'expliquer. *L'élément $LXR\alpha$ est effectivement présent lors de l'expérience mais c'est une forme inactive qui est détectée par la puce à ADN. Son activation n'a pas lieu à cause de la présence des acides gras.*

Affichage des sous graphes

L'affichage des sous-graphes se fait par l'interpréteur comme pour n'importe quel graphe ; en effet, les sous-graphes construits sont conformes à la spécification générale des graphes et les méthodes

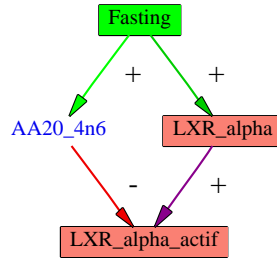


FIG. 2.6 – LXR α scindé en deux sommets.

d’affichage peuvent s’appliquer.

Conclusion

A partir d’une analyse des contradictions, on a ainsi proposé une méthode de représentation graphique des cohérences et incohérences entre les données de puces et le modèle bibliographique sous forme d’un ensemble de graphes introduisant des formes actives et inactives de protéines.

A partir de cette représentation, le chapitre 3 consistera à hiérarchiser les sous-graphes ainsi obtenus de manière à identifier les sommets réellement porteurs de contradictions.

Le chapitre 4 proposera une méthode dynamique pour extraire des données les différents types de contradictions que l’on souhaite mettre en évidence : celles qui correspondent à une erreur du modèle, celles qui sont issues d’une erreur dans l’expérimentation et celles qui traduisent un phénomène réel.

Les deux premières possibilités amèneront l’utilisateur à revoir ses données, la dernière lui indiquera la présence d’un phénomène complexe supposé l’intéresser et qu’il pourra étudier plus précisément.

Pour distinguer ces cas, on doit procéder à une analyse systématique capable de les différencier. Cette analyse reposera sur l’étude du graphe de contradictions et de son contexte.

Chapitre 3

Interprétation des sous-graphes

Dans cette section, nous proposerons une méthode de hiérarchisation des sous-graphes extraits d'une expérience pour mettre en évidence les sommets du graphe porteurs de contradictions ou de balances.

3.1 Travail d'analyse sur les sous-graphes

On a remarqué, en observant des exemples de graphes tels que celui représenté figure 2.3 que les graphes issus d'une analyse sur des données de puce sont naturellement hiérarchisés : ainsi, nous avons vu section 2.5 que les contradictions détectées sur SREBP et ACC étaient issues d'une seule contradiction sur LXR α .

Une manière plus efficace de présenter les sous-graphes extraits par l'analyse de contradictions est donc de les présenter sous forme d'arbre, c'est-à-dire un graphe dans lequel les arcs représentent une relation de filiation.

Ainsi, on va hiérarchiser les sommets concernés par l'expérimentation en fonction des chemins allant des uns aux autres. Pour cela, on définit une relation entre les paires de sommets du graphe et on ordonne ensuite les sommets en utilisant la restriction de cette relation aux paires (Exp, A) où Exp désigne l'expérience à analyser.

3.1.1 Relation d'inclusion : \subseteq

Définition 3 (Cellule) Soit G un graphe orienté. Pour toute paire (A, B) de sommets de G , on définit la cellule $C(A, B)$ qui désigne le triplet $[A, B, \mathcal{G}(A, B)]$, où $\mathcal{G}(A, B)$ est le sous-graphe de G qui contient tous les chemins issus de A et menant à B .

Cette cellule est dite non vide ssi $\mathcal{G}(A, B) \neq \emptyset$.

Définition 4 (Inclusion) Soit E un ensemble de cellules. On appelle inclusion la relation \subseteq définie sur $E \times E$ par : $C(A_1, A_2) \subseteq C(B_1, B_2)$ s'il existe au moins un chemin issu de B_1 menant à B_2 qui contient A_1 et s'il existe au moins un chemin issu de B_1 menant à B_2 qui contient A_2 , c'est-à-dire si A_1 et A_2 sont des sommets du graphe $\mathcal{G}(B_1, B_2)$.

Lemme 1 Soit G un graphe orienté, A_1, A_2, B_1 et B_2 des sommets de G tels que $C(A_1, A_2)$ et $C(B_1, B_2)$ sont des cellules non vides.

Si $C(B_1, B_2) \subseteq C(A_1, A_2)$ alors $\mathcal{G}(A_1, A_2)$ contient un chemin qui parcourt dans l'ordre A_1, B_1, B_2 et A_2 .

Preuve La démonstration s'appuiera sur un schéma : la Figure 3.1.

Si $C(B_1, B_2)$ est non vide, G contient un chemin issu de B_1 et menant à B_2 , en vert sur la figure 3.1.

Si $C(B_1, B_2) \subseteq C(A_1, A_2)$, G contient un chemin de A_1 vers B_1 , en jaune et un chemin de B_2 vers A_2 en rose.

G contient donc un chemin qui parcourt dans l'ordre B_1, B_1, B_2 et A_2 . $\mathcal{G}(A_1, A_2)$ contient tous les chemins de G issus de A_1 et menant à A_2 ; il contient donc le chemin qui parcourt dans l'ordre A_1, B_1, B_2 et A_2 .

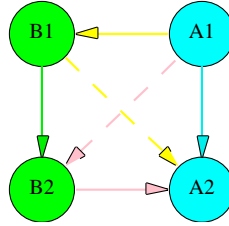


FIG. 3.1 – $\mathcal{G}(A_1, A_2)$ illustrant la démonstration du lemme 1.

■

Propriété 1 La relation d'inclusion sur les cellules est réflexive et transitive mais ni symétrique ni antisymétrique.

Preuve

- *Réflexivité.* Les sommets d'une paire appartiennent par construction au graphe des chemins entre ces sommets : $A_1, A_2 \in \mathcal{G}(A_1, A_2)$. Ainsi, $C(A_1, A_2) \subseteq C(A_1, A_2)$.
- *Transitivité.* Soit trois paires A_1, A_2, B_1, B_2 et C_1, C_2 telles que $C(C_1, C_2) \subseteq C(B_1, B_2)$ et $C(B_1, B_2) \subseteq C(A_1, A_2)$.
D'après la proposition 1, $\mathcal{G}(B_1, B_2)$ contient un chemin qui parcourt B_1, C_1, C_2, B_2 . Un raisonnement similaire à la démonstration de la proposition 1 nous montre que $\mathcal{G}(A_1, A_2)$ contient un chemin qui parcourt $A_1, B_1, C_1, C_2, B_2, A_2$.
Ceci implique que $C_1 \in \mathcal{G}(A_1, A_2)$ et $C_2 \in \mathcal{G}(A_1, A_2)$, i.e. $C(C_1, C_2) \subseteq C(A_1, A_2)$ la relation \subseteq est transitive.
- *Non symétrie.* Soit le graphe G à gauche dans la figure 3.2. $C(B_1, B_2) \subseteq C(A_1, A_2)$ mais $C(A_1, A_2) \not\subseteq C(B_1, B_2)$
- *Non antisymétrie.* Soit le graphe G à droite dans la figure 3.2. $C(A_1, B_1) \subseteq C(A_1, B_2)$ et $C(A_1, B_2) \subseteq C(A_1, B_1)$. Pourtant, $C(A_1, B_1) \neq C(A_1, B_2)$

■

3.1.2 Relation d'ordre : \sqsubseteq

On rend la relation sur les cellules $C(A, B)$ antisymétrique par une opération de quotient.

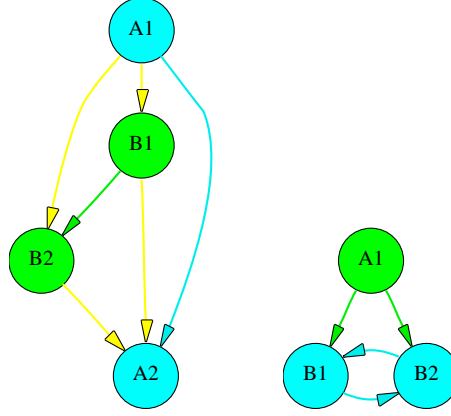


FIG. 3.2 – Graphes illustrant les propriétés de non symétrie et non antisymétrie.

Définition 5 (Relation d'ordre) Pour tous sommets A_1, A_2 d'un graphe G , $\mathfrak{C}(A_1, A_2)$ désigne l'ensemble des Cellules $C(B_1, B_2)$ telles que $\mathcal{G}(B_1, B_2) = \mathcal{G}(A_1, A_2)$, c'est-à-dire telles que $A_1, A_2 \in \mathcal{G}(B_1, B_2)$ et $B_1, B_2 \in \mathcal{G}(A_1, A_2)$.

Propriété 2 La relation définie sur l'ensemble des classes de cellules d'un graphe G par $\mathfrak{C}(A_1, A_2) \sqsubseteq \mathfrak{C}(B_1, B_2)$ si $\mathcal{G}(A_1, A_2) \subseteq \mathcal{G}(B_1, B_2)$ est une relation d'ordre partiel.

Implémentation La classe GraphTree est composée d'éléments de classe Cell, homogènes aux cellules $C(A, B)$ définies précédemment. Chaque ajout d'un nouvel élément nécessite le parcours de tous les éléments déjà présents afin de déterminer si une inclusion existe entre le nouvel élément et un ancien.

La complexité d'un ajout est $p \times a \times b$, où p est le nombre d'éléments déjà présents, a est le nombre de sommets du graphe de l'élément à ajouter et b le nombre moyen de sommets des graphes déjà présents dans la structure.

La possibilité de retirer des éléments n'a pas été implémentée, sa complexité serait $p \times c$ où c est le nombre moyen de sous-graphes inclus dans un graphe.

3.2 Représentation hiérarchique des éléments

On va organiser les différents sous-graphes issus d'une expérience sous la forme d'un diagramme. Cela nous permettra d'apprécier les différentes informations dans leur globalité et de mettre au point des méthode d'analyses liées à l'organisation des différents sous-graphes entre eux.

On doit trouver dans un tel diagramme les informations qui permettent de localiser les chemins contradictoires significatifs et d'éliminer les contradictions apportées par un sous-graphe qui ne font que bruyter l'information.

Le réseau utilisé pour les exemples sera celui de la figure 1.6 ; ce réseau décrit une partie de la régulation de la synthèse des acides gras. C'est une version simplifiée du réseau utilisé par les biologistes mais elle est suffisamment riche pour illustrer les méthodes de recherche et de représentation des contradictions.

Notre objectif sera entre autre de repérer l'erreur qui s'est glissée dans les données de puce. Pour la retrouver, nous allons faire appel à une relation hiérarchique entre les sommets. En effet, la donnée erronée doit apporter une contradiction avec le réseau, en isolant cette contradiction, nous devrions découvrir la donnée erronée.

3.2.1 Choix d'une représentation

Définition 6 (Espace de définition des sommets) Soit G un graphe de régulation et Exp un sommet associé à une expérimentation donnée.

On définit $\mathfrak{S}(Exp)$, l'ensemble des sommets Y tels que $Y \in \mathfrak{G}(Exp, X)$ et X est la cible d'un arc issu de Exp .

L'ensemble $\mathfrak{S}(Exp)$ est l'union des cibles de Exp et des sommets qui leur servent d'intermédiaires ce qui se traduit par l'ensemble des cibles de Exp et leur forme active.

Définition 7 (Relation sur les sommets) On définit la relation \leq sur $\mathfrak{S}(Exp) \times \mathfrak{S}(Exp)$:

$$A \leq B \text{ ssi } C(Exp, A) \subseteq C(Exp, B).$$

On représente graphiquement le diagramme issu de cette relation sur la Figure 3.3, c'est un graphe transitivement fermé (s'il existe un chemin reliant deux sommets en plusieurs sauts alors il existe un chemin reliant ces deux sommets en un saut).

Un tel diagramme permet de représenter de manière exhaustive toutes les interactions présentes dans le modèle étudié.

Cependant, ce diagramme est trop surchargé : certains chemins vont être représentés plusieurs fois. Par exemple, la relation $AA20 \leq SREBP$ existe mais tous les chemins de $AA20$ vers $SREBP$ passent obligatoirement par $LXR\alpha$. On ne souhaite donc pas représenter la relation $AA20 \leq SREBP$ dans le diagramme mais considérer qu'elle est déjà prise en compte par les deux relations $AA20 \leq LXR\alpha$ et $LXR\alpha \leq SREBP$.

Définition 8 (Domine, \triangleleft) Si R est une relation d'ordre sur un ensemble E , on dit qu'un élément x de E domine un élément y de E lorsque xRy , x différent de y , et qu'il n'existe aucun z , différent de x et y , tel que xRz et zRy .

On définit la relation de dominance : \triangleleft sur $\mathfrak{S}(Exp) \times \mathfrak{S}(Exp)$ basée sur la relation d'ordre \sqsubseteq

- $A \triangleleft B$ ssi
- $C(Exp, A) \sqsubseteq C(Exp, B)$
 - $C(Exp, A) \sqsubseteq C(Exp, D) \sqsubseteq C(Exp, B) \Rightarrow D = A \text{ ou } D = B.$

Propriété 3 La clôture transitive de la relation \triangleleft est la relation \leq .

Définition 9 (Diagramme de Hasse) Le graphe de la relation domine est appelé diagramme de Hasse de la relation d'ordre R .

Ce diagramme, représenté par la figure 3.4 possède d'importantes propriétés liées à la puissance de la relation d'ordre mais là encore, un problème se pose quant à l'interprétation des données. Contrairement au diagramme de la relation \leq , il va manquer des chemins. Par exemple, $LXR\alpha_{actif}$ a deux actions sur ACC , l'une est directe et l'autre passe par $SREBP$. Dans le diagramme de Hasse, seule la seconde est représentée.

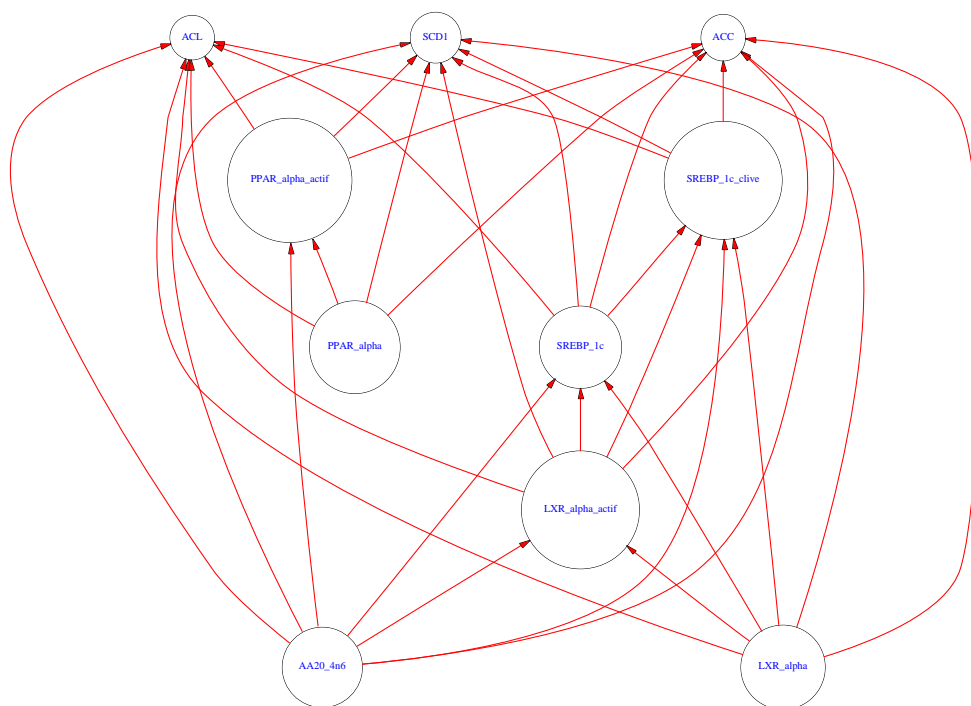


FIG. 3.3 – Diagramme de la relation \leq .

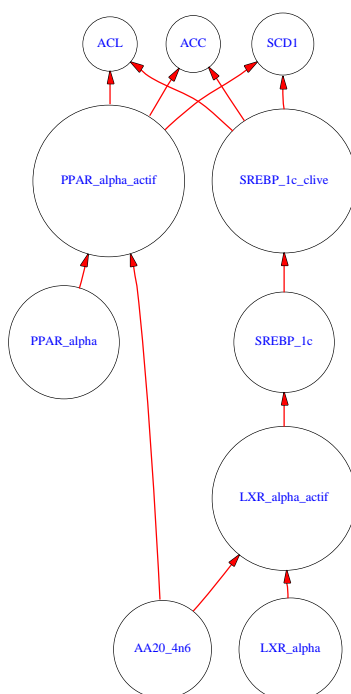


FIG. 3.4 – Diagramme de Hasse de la relation \leq .

Le diagramme que l'on souhaite afficher doit représenter tous les chemins directs entre deux sommets mais uniquement ceux-ci.

Ce choix impose de renforcer la contrainte sur la relation entre les cellules que l'on a définie plus tôt.

Définition 10 (Relation des chemins directs, \mapsto) Soit G un graphe de régulation, et Exp un sommet associé à une expérimentation donnée.

On définit la relation \mapsto sur $\mathfrak{S}(Exp) \times \mathfrak{S}(Exp) : A \mapsto B$ si et seulement s'il existe un chemin direct entre A et B dans G . Un chemin direct étant un chemin qui ne passe par aucun des autres sommets de $\mathfrak{S}(Exp)$.

On obtient le diagramme de la figure 3.5.

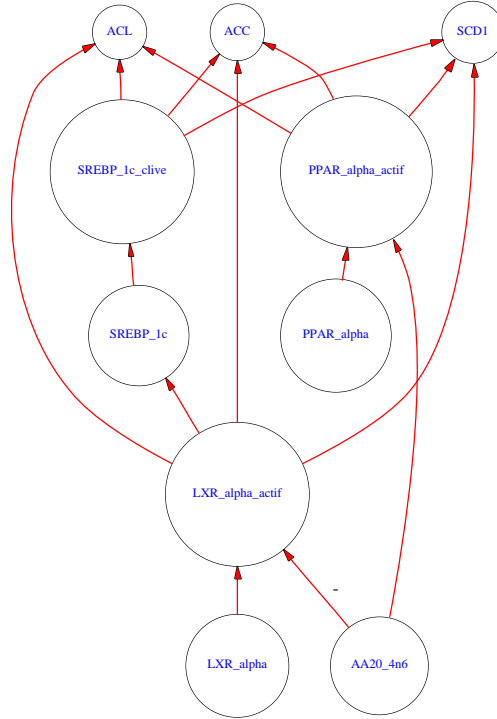


FIG. 3.5 – Diagramme de la relation \mapsto .

Cette nouvelle relation apporte une propriété combinatoire. Si l'on connaît le nombre de chemins directs reliant tous les sommets, il est possible de connaître le nombre total de chemins reliant chaque sommet.

Propriété 4 Soit G un graphe de régulation, Exp un sommet de G associé à une expérience donnée. Pour tout $A, B \in \mathfrak{S}(Exp)$,

$$N(A, B) = N_d(A, B) + N_i(A, B) \quad \text{et} \quad N_i(A, B) = \sum_{C \in \mathfrak{S}} N(A, C) \times N_d(C, B).$$

Dans ces deux formules, $N(A, B)$ est le nombre de chemins total reliant A à B , $N_d(A, B)$ le nombre de chemins directs et $N_i(A, B)$ le nombre de chemins indirects. \mathbf{S} est l'ensemble des sommets C du diagrammes tels que $N_d(C, B) > 0$.

Ce calcul utilise une récurrence et impose que l'ensemble des sommets soit ordonné, c'est-à-dire que le graphe de régulation ne contienne pas de cycles. Les graphes de régulation que nous analyserons contiennent évidemment des cycles qui correspondent aux retro-régulations. Néanmoins en pratique, on peut dénombrer le nombre chemins directs et indirects lors de l'extraction du graphe. Ce calcul n'est alors utile que pour vérifier l'homogénéité de l'ensemble des sous-graphes, vérifier que tous les chemins sont bien présents.

Il est aussi possible lors de la construction des graphes de dénombrer pour chaque paire de sommets les nombres des chemins contradictoires relativement à la donnée de l'expérience. On peut dès lors vérifier pour chaque Cellule si les chemins contradictoires sont issus d'un sommet inférieur ou si la contradiction est propre à ce sommet à l'aide du critère suivant :

Définition 11 (Critère de détection des contradictions propres.) *Un sommet est dit porteur de contradictions ssi*

$$c_p(A) = c(A) - \sum_{B \in \mathbf{S}} c(B) \times N_d(B, A) > 0.$$

$c(A)$ est le nombre total de contradictions aboutissant au sommet A par rapport à la donnée de l'expérience, $c_p(A)$ est le nombre de contradictions propres au sommet A , ce sont celles qui nous intéressent. Les autres variables sont identiques à celles que l'on a définies plus haut. Il est important de noter que cette propriété est valable même lorsque le graphe est cyclique puisque le calcul n'est pas récurrent : seul $c_p(A)$ est une inconnue.

3.2.2 Représentation des sommets porteurs de contradictions

Nous utiliserons pour la suite le diagramme de la relation \mapsto . Sur ce graphe nous ajouterons plusieurs informations : dans chaque sommet nous placerons, s'il y en a, le nombre de contradictions relevées par rapport au signe de référence et le nombre de contradictions attendues venant des sommets inférieurs. Si le nombre de contradictions dénombrées est supérieur au nombre de contradictions calculées à partir des sommets inférieurs, le sommet sera coloré en orange. Nous décorerons chaque arc du signe nécessaire pour passer d'un sommet à l'autre dans le graphe en passant uniquement par les chemins directs. S'il existe plusieurs chemins directs dont les signes se contredisent, on mettra le signe indéterminé (0_+_-).

Pour finir avec la description du protocole utilisé, on notera que le diagramme contient les formes actives des facteurs de transcriptions (LXR α , SREBP, PPAR) alors que ces produits n'ont pas de donnée expérimentale. En effet, ces produits revêtent une importance majeure dans la localisation des contradictions car ils sont placés à des carrefours métaboliques. La figure 2.6 illustre particulièrement le problème : certaines sources de conflits comme les balances sont susceptibles de porter sur des sommets non ciblés par l'expérimentation.

Pour repérer les contradictions et construire les sous graphes et le diagramme on utilise presque exclusivement l'objet visiteur décrit en 2.4.

Le programme effectue un premier passage du visiteur sur les cibles de l'expérimentation qui calcule les graphes et dénombre les contradictions. Un deuxième passage est effectué sur les sommets

sensibles. Ces sommets sont tous ceux qui interviennent à l'intérieur des graphes sans être ciblés par l'expérience. On leur attribue donc le signe marquant l'indétermination ($0_+ _ -$). Les contradictions sont calculées par rapport au signe qui offre le moins de contradictions. C'est là la faiblesse de cette méthode, elle sous-entend que le réseau est globalement homogène, i.e. que les contradictions sont minoritaires sur l'ensemble des chemins.

On effectue ensuite un passage du visiteur sur les sous-graphes intermédiaires, c'est-à-dire sur les graphes issus d'un sommet du diagramme et aboutissant sur un autre sommet du diagramme. Ce passage a pour but de dénombrer les chemins directs reliant ces sommets et l'éventuel signe qui y est attaché.

Le programme a alors toutes les informations requises pour construire le diagramme.

3.2.3 Application et interprétation des résultats

La figure 3.6 nous montre le résultat de cette analyse sur le réseau représenté par la figure 1.6.

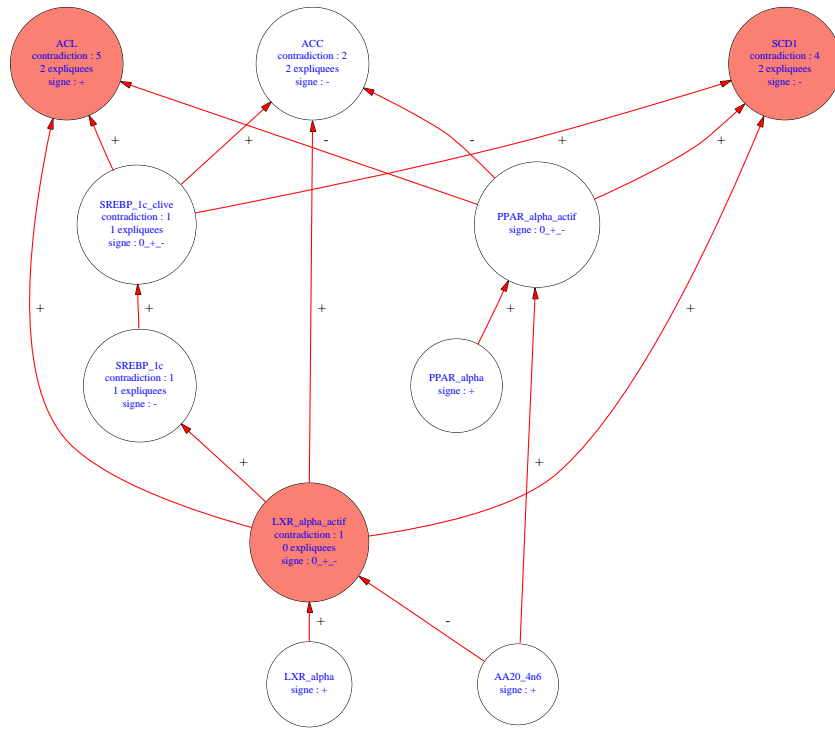


FIG. 3.6 – Résultat de l'analyse des contradictions sur le réseau simplifié de la synthèse des acides gras.

On retrouve colorés en orange les sommets sur lesquels on s'attendait à trouver des contradictions propres :

- sur LXR α actif car sur ce sommet aboutit la balance présentée dans la partie 2.5 ;
- sur le sommet de l'enzyme ACL car on savait la donnée de puce incompatible avec le réseau.

On découvre de plus deux nouvelles contradictions portant sur le sommet SCD1. La figure 3.7 nous montre le sous-graphe associé à ce sommet. On constate que quatre chemins contredisent la

donnée de puce. Sur ces quatre contradictions, deux sont issues de LXR α actif, le graphe associé à cette enzyme doit donc permettre de régler ce problème. De plus comme il s'agit d'une balance qui porte sur LXR α actif, les chemins contradictoires sont doublés de chemins explicatifs. Par contre les deux autres chemins contradictoires ne passent pas par LXR α actif. Ces deux chemins passent par PPAR α actif. Le diagramme n'a pas relevé de contradictions sur PPAR α actif et le réseau est supposé sans erreurs, il s'agit donc d'une balance portant sur SCD1, balance entre l'influence positive sur SCD1 du facteur de transcription PPAR α d'une part et celle négative du facteur de transcription LXR α d'autre part.

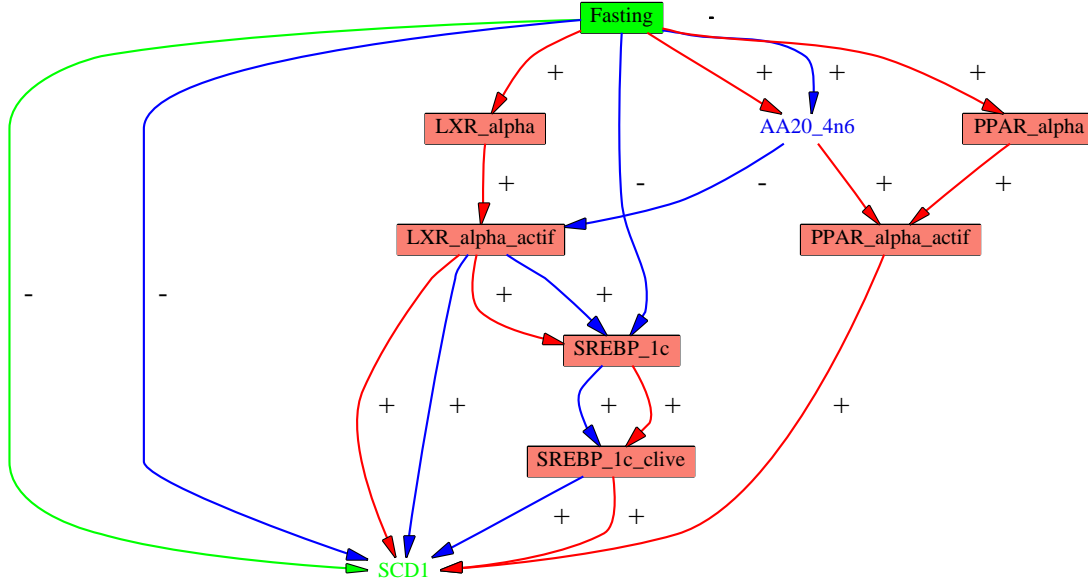


FIG. 3.7 – Résultat de l'analyse des contradictions sur le sous-réseau de régulation de l'enzyme SCD1.

Chapitre 4

Analyse dynamique du réseau

4.1 Définitions des fonctions utilisées

Le diagramme construit précédemment (Figure 3.6) contient des informations qui n'ont pas été utilisées dans l'analyse des contradictions : il s'agit des signes attribués aux sommets (qui correspondent aux données de puce) et des signes attribués aux arcs (qui correspondent aux signes des chemins directs reliant les sommets).

Ces informations ont été recueillies pour un deuxième type d'analyse assimilable à une recherche itérative de point fixe au sein du graphe qualitatif déduit d'une expérimentation.

On peut considérer que le diagramme construit représente une version simplifiée du réseau étudié dont tous les arcs sont signés et où le signe de chacun des sommets est fonction du signe de ses voisins.

Définition 12 (Signes initiaux des sommets) Soit G un graphe de régulation, Exp un sommet associé à une expérience. La fonction ξ est définie sur $\mathfrak{S}(Exp)$ par :

- $\xi(S) = +$ si $Exp \rightarrow^+ S$,
- $\xi(S) = -$ si $Exp \rightarrow^- S$,
- $\xi(S) = 0_{+}-$ si $Exp \nrightarrow S$.

La fonction τ est définie sur $\mathfrak{S}(Exp)$ par :

- $\tau(S) = 100$ si la p -value de la donnée relative à S dans Exp est inférieure à 1,6 %,
- $\tau(S) = 50$ si la p -value est inférieure à 5 %,
- $\tau(S) = 20$ si la p -value est inférieure à 10 %.

Définition 13 (Signes des arêtes) Soit A et B deux sommets du graphe G . La fonction Φ est définie sur $\mathfrak{S}(Exp) \times \mathfrak{S}(Exp)$ par

- $\Phi(A, B) = 0$ s'il n'existe pas de chemin direct de A vers B .
- $\Phi(A, B) = \varepsilon$ si tous les chemins directs de A vers B sont du signe ε .
- $\Phi(A, B) = 0_{+}-$ si deux chemins directs de A vers B de sens différent.

4.2 Algorithme de l'analyse dynamique

On considère que les fonctions ξ , τ et Φ sont fournies.

Le calcul itératif portera sur ξ_k et τ_k . ξ_k et τ_k sont initialisées avec ξ et τ ,

ξ_{k+1} et τ_{k+1} sont calculées comme suit :

Pour chaque sommet I , on effectue les actions suivantes

– **Calcul pour le sommet A de la contribution de chacun de ses voisins**

La contribution du sommet B sur le sommet A est définie par le produit

$$C_k(A, B) = \xi_k(B) \otimes \Phi(A, B)$$

Il s'agit du signe que devrait avoir le sommet A pour que le signe du sommet B soit déduit du signe A via l'action $A \rightarrow B$.

\otimes est l'opérateur de produit de signe utilisé dans GARMeN que l'on a défini dans le chapitre 2.

– **Calcul des éléments suivants**

$$\sigma^+(A) = \sum_{B|C_k(A,B)=+} \tau_k(B) + \sum_{B \text{ t.q. } C_k(B,A)=+} \tau_k(B)$$

$\sigma^+(A)$ est le poids positif envoyé sur A par son environnement.

$$\sigma^-(A) = \sum_{B|C_k(A,B)=-} \tau_k(B) + \sum_{B \text{ t.q. } C_k(B,A)=-} \tau_k(B)$$

$\sigma^-(A)$ est le poids négatif envoyé sur A par son environnement.

$$\rho(A) = \frac{\sigma^+(A)}{l} \quad \eta(A) = \frac{\sigma^-(A)}{m}$$

l est le nombre de contributions positives de B sur A et de A sur B , m est le nombre de contributions négatives de B sur A et de A sur B . ρ et η sont les poids moyens respectivement positif et négatif envoyés sur A .

– **Calcul de la moyenne pondérée**

On va considérer que le signe probable du sommet A est positif si l'environnement envoie plus de poids positif que de poids négatif sur A

$$\xi_{k+1}(A) = + \text{ ssi } \rho(A) > \eta(A)$$

$$\xi_{k+1}(A) = - \text{ ssi } \rho(A) < \eta(A)$$

On considère que la fiabilité de cette information est la moyenne pondérée des poids.

$$\tau_{k+1}(A) = \left| \frac{\sigma_+(A)}{\sigma_+(A) + \sigma_-(A)} \rho(A) - \frac{\sigma_-(A)}{\sigma_+(A) + \sigma_-(A)} \eta(A) \right|$$

On obtient ainsi le signe probable d'un sommet A en fonction de son environnement avec une certaine fiabilité. L'itération de ces calculs permettra de voir si ces signes se stabilisent sur une fiabilité non nulle. Si c'est le cas, on pourra considérer qu'on a trouvé l'état stable du système et on pourra le comparer aux données réelles. On en déduira :

- les données fausses,
- les balances.

4.3 Applications

L'analyse de notre réseau de régulation, poussée jusqu'à la 200^e itération fournit un tableau de valeurs pour chaque sommet du diagramme. Ces valeurs comprises entre -100 et +100 représentent le signe et la fiabilité calculés à chaque itération. La table 4.1 rassemble les dix premières valeurs pour chaque sommet, la figure 4.1 montre les deux cents points obtenus pour chaque sommet.

AA20-4n6	ACL	ACC	SREBP-1c	LXR α
+100.0	+50.0	-100.0	-100.0	+100.0
+100.0	+50.0	-100.0	-100.0	+100.0
+57.57576	-43.896305	-63.574524	-72.33766	+22.171717
+47.149082	-54.02825	-56.839424	-65.9001	-37.434624
+44.703804	-50.04615	-50.44775	-57.882977	-47.00818
+43.20278	-47.38858	-47.445953	-53.884346	-49.394367
+41.149223	-43.941093	-43.949287	-49.332024	-48.171467
+39.4414	-41.491062	-41.49223	-46.13566	-46.359715
+37.450226	-38.870247	-38.870415	-42.752556	-43.731968
+35.70049	-36.730522	-36.73054	-39.98829	-41.24323
PPAR α	SCD1	LXR α -actif	PPAR α -actif	SREBP-1c-clive
+100.0	-100.0	0 ₋ + ₋ -0.0	0 ₋ + ₋ -0.0	0 ₋ + ₋ -0.0
+100.0	-100.0	-52.272728	+41.666668	-78.57143
+61.111115	-50.98957	-50.515507	+38.569324	-78.0659
+46.083252	-41.569603	-51.794956	+36.39002	-59.962444
+39.621094	-33.75757	-50.587467	+35.067585	-55.18191
+36.58542	-30.353935	-47.56002	+33.71166	-48.827866
+34.669575	-26.44757	-45.453842	+32.575043	-45.21928
+33.27322	-23.863466	-42.418095	+31.486765	-41.39547
+32.08225	-21.028097	-39.99886	+30.527252	-38.59559
+31.045586	-18.758743	-37.338917	+29.572817	-35.737583

TAB. 4.1 – Les dix premières valeurs obtenues pour chacun des sommets.

On constate que dès les premières valeurs, chaque sommet réussit à se fixer sur un signe. De plus ce signe est cohérent par rapport à notre analyse : les signes fournis par les données de puces sont conservés à l'exception d'ACL qui change dès la deuxième itération (le signe d'ACL obtenu par les données de puce était faux et sa fiabilité initiale n'était que de 50 %) et de LXR α qui change de valeur à la troisième itération (cela signifie que dans la balance qui porte sur LXR α actif, ce sont les acides gras qui l'emportent).

Cependant, le réseau ne se stabilise pas sur un point fixe. Les fiabilités décroissent lentement jusqu'à zéro alors que l'on aimerait trouver une valeur non nulle.

Deux explications sont à considérer.

- Dès qu'un signe ne peut se stabiliser ou qu'une donnée est fausse, sa fiabilité τ décroît vers 0 et entraîne avec elle celle de ces voisins. On pourra essayer d'introduire un seuil en dessous duquel on ne prendra plus en compte la donnée τ d'un voisin dans le calcul de σ .
- Il manque des informations pour stabiliser le système, le réseau est trop incomplet pour être stable. Les données manquantes pourraient être obtenues dans un réseau qualitatif extrapolé à partir du réseau métabolique existant.

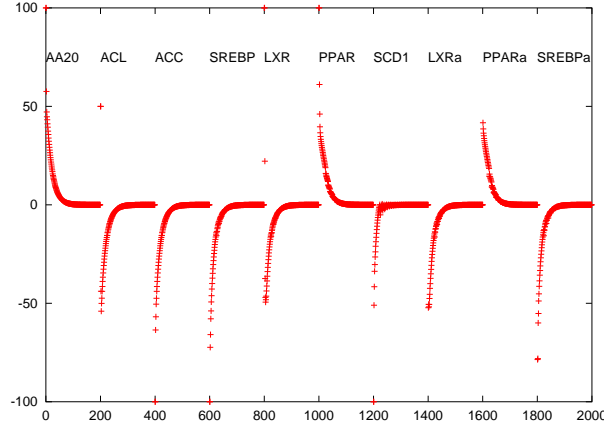


FIG. 4.1 – Les deux cents valeurs obtenues pour chacun des sommets.

4.3.1 Simulation d'un réseau suralimenté

Si l'on bloque une des valeurs, par exemple la valeur des acides gras, on arrive à obtenir un équilibre non nul comme le montre la figure 4.2. On peut justifier le blocage de cette valeur par le fait que lors du jeun, des acides gras sont apportés depuis l'extérieur dans le foie. Le blocage de la valeur simule un apport permanent d'acides gras dans le système.

On obtient alors un réseau qui oscille entre deux états, les valeurs ne parviennent pas à se stabiliser. La plupart des valeurs sont comprises dans une fourchette raisonnable mais la valeur de SCD1 oscille entre des valeurs positives et négatives, il est alors impossible de lui attribuer un signe.

La méthode présentée dans ce chapitre semble efficace pour détecter le signe d'un sommet et vérifier la validité des données de puce mais n'arrive pas à se stabiliser. La convergence est très lente, mais ne trouve pas de point fixe. Si l'on bloque une valeur, le réseau se stabilise mais on ne parvient pas à déterminer le signe de tous les sommets. Une autre solution consiste à considérer qu'une valeur τ est stable dès que sa variation $\delta\tau$ passe sous un certain seuil. Le réseau est alors stable avec des fiabilités non nulles. Pour des seuils de l'ordre de 0,1% et 1%, la stabilité est obtenue est encore trop proche de zéro. Pour un seuil de 5% en revanche, on obtient une stabilité véritablement satisfaisante : figure 4.3.

Une telle méthode de recherche de points fixes est prometteuse mais nécessite d'être affinée dans plusieurs directions :

- Test de différentes formules de calcul de τ et σ pour améliorer la convergence,
- Prise en compte lors des itérations des phénomènes (balances, défaut) déjà détectés.
- Complétion du réseau qualitatif.

On disposera alors d'un état stable du réseau qui pourra être comparé aux données initiales. On disposera ainsi d'un critère de fiabilité des données d'une part et d'extraction des nœuds de balance d'autre part.

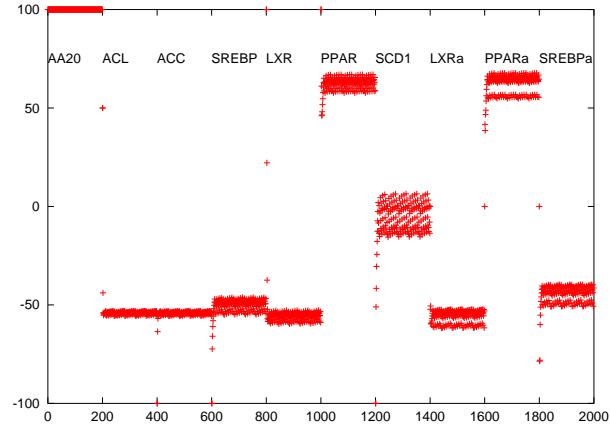


FIG. 4.2 – Valeurs obtenues en bloquant la fiabilité de AA20.

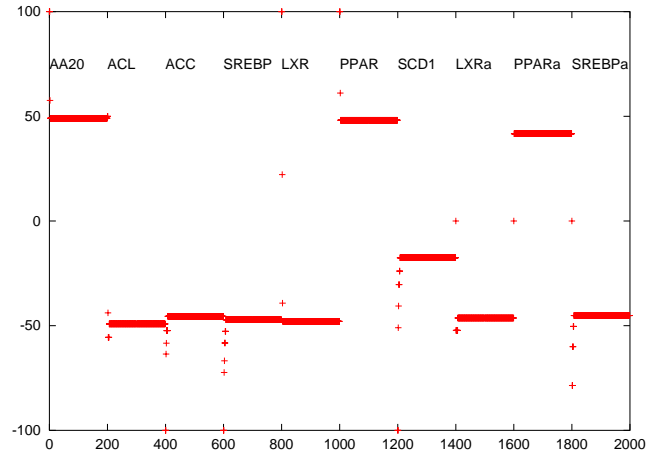


FIG. 4.3 – Valeurs obtenues en négligeant les faibles variations.

Chapitre 5

Traitement du métabolisme

Les voies métaboliques, dans le réseau que nous utilisons ne sont pas modélisées. On ne distingue pas non plus les produits intermédiaires ou finaux. On représente les enzymes connues, ainsi que leur substrat et leur produit mais elles ne sont pas organisées sous forme de voies. Pourtant, la distinction de voies métaboliques permettrait d'ajouter de l'information de type régulation sur toute la masse de données métaboliques. En effet, il est possible d'appliquer des règles de régulation sur les produits finaux en oubliant les produits intermédiaires.

On dispose de plus de données de régulations sur les réactions enzymatiques qui se modélisent par l'équation de Michaelis (cf 1.1.2) : un substrat active souvent une catalyse.

Les processus d'activation ne sont pas non plus représentés de manière exacte. On représente par exemple l'activation de $LXR\alpha$ par une flèche positive sur la forme active de $LXR\alpha$ et son inhibiteur par une flèche négative. Si cette relation est vraie quant aux régulations, elle est inexacte vis à vis du métabolisme. L'activation de $LXR\alpha$ est une seule réaction biochimique qui fait intervenir plusieurs acteurs :

- une forme inactive : $LXR\alpha$;
- une forme active : $LXR\alpha$ actif ;
- un inhibiteur : l'acide gras 20 : 4n – 6 (AA20_4n6 dans les graphes).

Ces réactions, voies métaboliques ou activations, peuvent être représentées de manière similaire : une représentation qui offre un affichage correct des processus métaboliques et qui tient compte des régulations sous-jacentes.

L'introduction d'une nouvelle représentation, tenant compte des régulations du métabolisme permet ainsi d'aller plus loin dans les vérifications de signe et d'homogénéité.

5.1 Représentation du métabolisme

On peut modéliser une réaction métabolique ou une voie métabolique à l'aide d'activateur ou d'inhibiteur. Une réaction enzymatique ainsi modélisée verra son produit lié par une régulation à son substrat et à son enzyme. Dans certains cas, on pourra même ajouter la rétroinhibition du produit sur la réaction. Ce type d'inhibition n'a de valeur que dans une étude dynamique mais il possède une existence réelle qu'il faudra prendre en compte tôt ou tard.

Le type de modélisation qui se prête le mieux aux réactions métaboliques est une représentation de type «boîte noire» : un sommet représente la réaction à modéliser. Il possède une entrée, le

substrat et une sortie, le produit. On peut de plus lui ajouter une quantité variable d'activateurs et d'inhibiteurs.

5.1.1 Représentation des réactions enzymatiques

Dans le cas d'une réaction enzymatique, l'enzyme sera un des activateurs. Si cet enzyme nécessite un ligand, on pourra l'ajouter comme second activateur. On peut voir sur la figure 5.1 un exemple de synthèse enzymatique représentée de trois manières :

- Représentation initiale (avec sommet matérialisé)
- Représentation style biochimique (avec branchement)
- Représentation sous la nouvelle forme (avec sommet matérialisé et branchement)

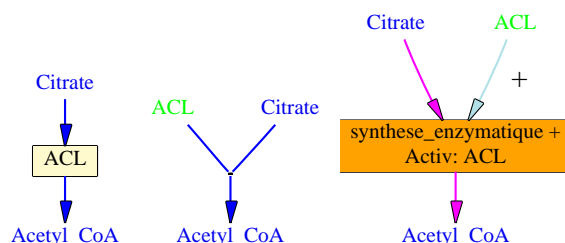


FIG. 5.1 – Représentations d'une réaction enzymatique.

Avec cette représentation, les enzymes sont considérées comme des produits ordinaires, ce qui avait été introduit avec la représentation en «style biochimique». La représentation avec sommet a été réintroduite car elle permet de matérialiser la réaction, de lui attribuer un nom (ici synthèse enzymatique) et de récapituler ses activateurs et inhibiteurs. La représentation en branchement du «style biochimique» posait un problème : il était difficile de sélectionner la réaction car cela supposait de cliquer sur le point situé à l'intersection des branches. Ce problème est résolu quand l'intersection est un sommet.

5.1.2 Représentation des voies métaboliques

La représentation en sommet avec branchement permet bien sûr de modéliser toute une voie métabolique (cf 1.1.3). Il suffit de fournir comme facteur le substrat de la première enzyme et comme cible le produit final de la voie. Exemple : Figure 5.2.

5.1.3 Représentation de l'activation d'une protéine

La modélisation de l'activation d'une protéine comme LXR α fait appel à un inhibiteur. L'inhibition est représentée par une flèche orange et est rappelée dans le sommet : Figure 5.3.

Dans le sommet, la première ligne porte le nom de la réaction, ce nom doit être fourni lors de la fabrication du réseau et n'obéit pas à des critères particuliers. On peut ainsi modéliser toute sorte de réactions même si leur nom n'était pas connu lors de la création de cette modélisation.

Le signe placé à côté du nom sert à l'interprétation du métabolisme en régulation. Il représente l'influence du facteur sur la réaction. Comme pour n'importe quelle régulation, on a le choix, , entre régulation positive +, régulation négative -, pas de régulation 0 ou régulation inconnue +_ _0

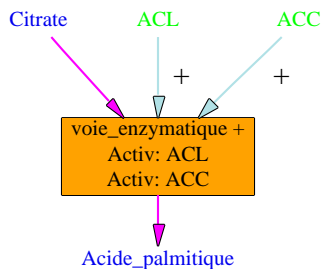


FIG. 5.2 – Représentation d’une voie enzymatique.

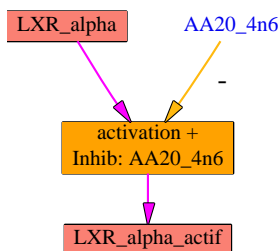


FIG. 5.3 – Représentation de l’activation de LXR α .

5.2 Extrapolation des régulations à partir du métabolisme

Le schéma de la figure 5.4 reprend le réseau de la figure 1.6 en utilisant la nouvelle représentation.

Il contient les mêmes régulations concernant la régulation des facteurs de transcriptions que le schéma d’origine (en particulier sur LXR α , PPAR et SREBP) et en ajoute sur la synthèse même des acides gras. On a maintenant des régulations positives des enzymes ACC, ACL et SCD1 sur ces acides gras.

Conclusion

Il n’existe pas encore de règle qui permet d’associer systématiquement des régulations au métabolisme.

Par exemple, une voie enzymatique est parfois régulée par la première enzyme, d’autres fois, elle est régulée par son produit. Elle peut très bien être régulée par les deux. L’information existe mais elle n’est encore que dans les publications biologiques. De ce fait, les données de régulation seront dans un premier temps entrées manuellement, lors de la construction du réseau. On peut tout de même espérer dégager des heuristiques pour associer automatiquement des données comportementales aux données métaboliques. L’analyse des données de puce, et notamment la recherche de contradictions permettra de vérifier la cohérence de ces nouvelles interactions.

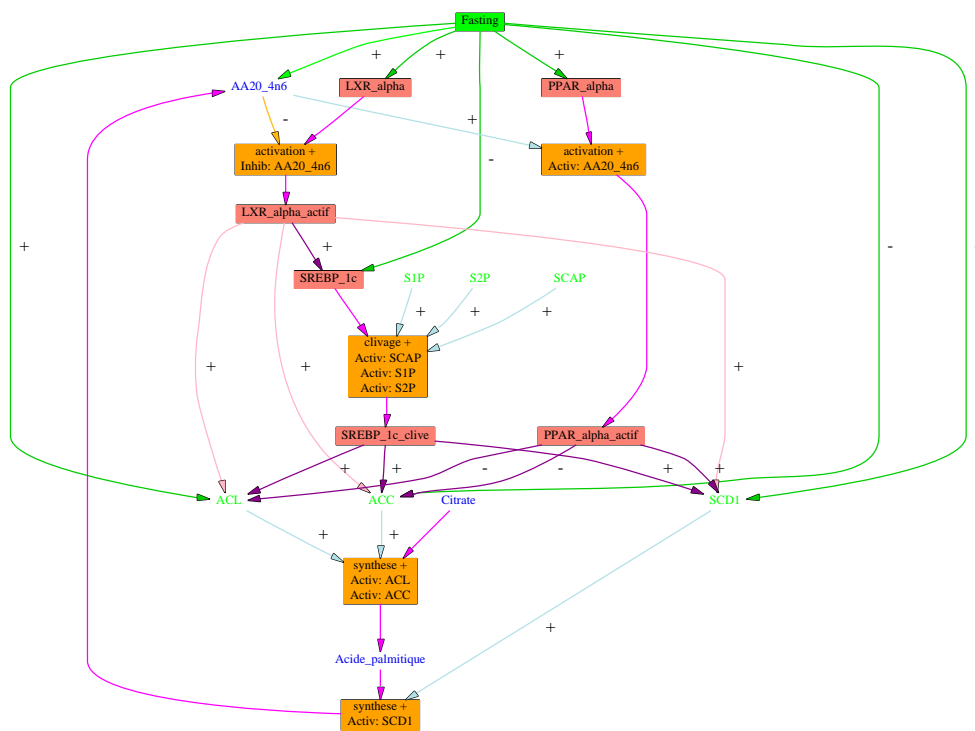


FIG. 5.4 – Réseau simplifié de la synthèse des acides gras.

Conclusion

L'analyse de données de puces peut se faire de plusieurs manières :

- **localement** ; en observant les sous-graphes qui portent des contradictions on peut dégager des renseignements sur le fonctionnement du métabolisme.

Pendant le stage, on a vu par exemple s'éclaircir les phénomènes de balance qui portent autour de LXR et de PPAR, phénomènes très bien représentés dans les sous-graphes.

- **globalement** ; l'étude des diagrammes issus d'une analyse de données de puce permet de localiser et de corriger des erreurs : la procédure qui permet d'obtenir les données de puce à été éprouvée et est reconnue fiable mais elle reste une méthode analogique susceptible d'être bruitée. Ainsi, l'erreur sur la donnée de ACL que nous possédions n'a pas été ajoutée, elle existe réellement et l'analyse par diagramme est tout à fait capable de la repérer.
- **dynamiquement** ; l'analyse dynamique itérative permet de sortir du cadre des données de puce : les données de puce forment un point de départ mais les signes des sommets sont estimés de proche en proche. On peut donc assez vite voir se répandre l'information et les sommets qui n'avaient pas de signes de départ se fixent rapidement sur un signe qui représente leur état lors de l'expérience.

La représentation comportementale du métabolisme permettra d'aller encore plus loin : le réseau métabolique est très important et la plupart des régulations sont rebouclées par des réactions métaboliques. Si l'analyse de ces régulations est non triviale, l'utilisation et l'analyse des données de puce en parallèle avec le réseau métabolique devraient fortement contribuer à simplifier la tâche des biologistes.

Bibliographie

- [BV00] A. Brazma and J. Vilo. Gene expression data analysis. Technical report, European Molecular Biology Laboratory, Outstation Hinxton - the European Bioinformatics Institute, Cambridge, Juin 2000.
- [CB95] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, London, 1995.
- [DDB04] Dna data bank of japan. www.ddbj.nig.ac.jp, mars 2004.
- [DJ02] H. De Jong. Modeling and simulation of genetic regulatory systems : A literature review. Technical report, INRIA, Septembre 2002.
- [HMC02] J. Hasty, D. McMillen, and J.J. Collins. Engineered gene circuits. *Nature*, 420 :224 – 230, Novembre 2002.
- [Kit02] H. Kitano. Computational systems biology. *Nature*, 420 :206 – 210, Novembre 2002.
- [MJ61] J. Monod and F. Jacob. General conclusions : telenomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harbor Symp Quant Biol*, 26, 1961.
- [PRM⁺02] T. Poynard, V. Ratziu, J. Moussalli, C. Regimbeau, V. di Martino, Y. Benhamou, R. Myers, and F. Imbert-Bismut. Biopsie du foie contre prise de sang pour le suivi de l'hépatite C. *Medecine / Sciences*, 3, Mars 2002.
- [Sea02] D. B. Searls. The language of genes. *Nature*, 420 :211 – 217, Novembre 2002.
- [Sur02] C. Surridge. Computational biology. *Nature*, 420 :205, Novembre 2002.
- [TBM⁺03] R. Teusan, A. Bihouee, N. Le Meur, G. Ramstein, and J. Leger. Eccb. In *MADTools : management tool for the mining of microarray data.*, septembre 2003.
- [TDJ02] D. Thieffry and H. De Jong. Modélisation, analyse et simulation des réseaux génétiques. *Medecine / Sciences*, 4 :492 – 502, Avril 2002.
- [TJ00] D. Thieffry and B. Jacq. Modéliser les réseaux de gènes. *Biofutur*, 206, 2000.