



# Modélisation de l'évolution dirigée de la tyrosyl-ARNt synthétase *in silico*. Introduction d'acides aminés non naturels dans les protéines.

Najette Amara

## ► To cite this version:

Najette Amara. Modélisation de l'évolution dirigée de la tyrosyl-ARNt synthétase *in silico*. Introduction d'acides aminés non naturels dans les protéines.. Bio-Informatique, Biologie Systémique [q-bio.QM]. Ecole Polytechnique X, 2011. Français.

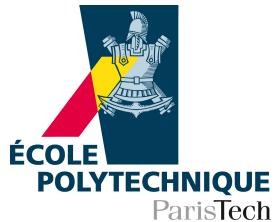
**HAL Id:** pastel-00622571

<https://pastel.archives-ouvertes.fr/pastel-00622571>

Submitted on 12 Sep 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Laboratoire de Biochimie  
Ecole Polytechnique  
91128 Palaiseau CEDEX

THÈSE  
présentée pour obtenir le grade de  
DOCTEUR DE L'ECOLE POLYTECHNIQUE.  
*Discipline : Biologie.*

**Modélisation de l'évolution dirigée de la tyrosyl-ARNt  
synthétase in silico. Introduction d'acides aminés non naturels  
dans les protéines.**

Najette AMARA

Jury composé de

David Perahia	Rapporteur
Yves-Henri Sanejouand	Rapporteur
Catherine Etchebest	Examinateur
Isabelle Andre	Examinateur
Thomas Simonson	Directeur de thèse

---

## Résumé

Les protéines sont synthétisées essentiellement à partir d'acides aminés L. Cependant, les acides aminés D sont naturellement présents dans notre organisme. Il existe plusieurs mécanismes de régulation pour limiter leur présence. Il peut y avoir un intérêt bio-technologique à introduire des acides aminés D dans les protéines pour leur conférer des fonctions chimiques nouvelles. Les protéases, chargées de dégrader les peptides, reconnaissent moins bien les protéines possédant de tels acides aminés. Dans notre cas, présenter la tyrosine sous son énantiomère D donnerait un avantage par rapport aux protéases. En effet, un peptide qui aurait des acides aminés D, resterait plus longtemps actif dans un organisme. La tyrosyl-ARNt synthétase (TyrRS) a déjà été utilisée pour l'incorporation de plus d'une trentaine d'analogues de la tyrosine dans les protéines. Cette enzyme possède déjà une activité naturelle pour la D-tyr, mais celle ci est très faible. L'objectif du présent travail est de l'améliorer par une méthode d'évolution dirigée *in silico*. Cette méthode consiste à muter aléatoirement une protéine puis à sélectionner les mutants qui ont les propriétés désirées, mimant les mécanismes de l'évolution naturelle. L'évolution dirigée *in silico* a déjà été utilisée pour l'ingénierie d'un grand nombre de protéines. On utilise pour cette étude le programme d'évolution dirigée développé au laboratoire qui procède à une optimisation itérative de la séquence et de la structure. Des mutations ont été retenues par cette méthode et proposées aux expérimentalistes du laboratoire de biochimie pour des validations expérimentales. Ils ont conclut que plusieurs mutants présentent une activité faible mais mesurable pour la D-Tyrosine. L'étude a été étendue à d'autres résidus proches de l'ammonium de la L-Tyr. Nous avons identifié d'autres positions qui pourraient être intéressantes à muter.

L'étude, faite dans un premier temps avec le ligand tyrosine, a été ensuite améliorée en considérant le ligand Tyrosyl adénylate (tyrAMP). C'est l'intermédiaire qui va interagir avec l'ARNt pour former l'ARNt<sup>Tyr</sup>. Le L-TyrAMP et le D-TyrAMP ont tous deux été modélisés avec la TyrRS d'*Escherichia coli*. Toutes ces modélisations ont conduit à une liste de séquences mutantes qui ont été prédictes comme favorables à la liaison du D-TyrAMP. Chacune des séquences obtenues pour les deux synthétases a ensuite fait l'objet d'une étude plus approfondie visant à déterminer la spécificité pour le ligand D-TyrAMP. Suite à cela, nous avons sélectionné une dizaine de séquences de TyrRS mutantes d'*Escherichia coli* an de réaliser des calculs de dynamique moléculaire. Nous avons utilisé les ressources du super calculateur du CINES (Centre Informatique Nationale de l'Education Supérieur) pour faire des études de dynamique moléculaire. Une meilleure estimation de l'affinité de ces différents mutants pour le D-TyrAMP a été atteinte.

Un autre but de la thèse était d'améliorer notre protocole de Protein Design en adoptant un champ de force tout atome et un modèle de solvant plus sophistiqué. Pour le solvant, le modèle de Born généralisé que nous avons implémenté dans le protocole est basé sur la théorie de Poisson-Boltzmann, avec un continuum diélectrique qui entoure la protéine. Les validations sur des peptides tests ont été concluantes. L'évolution dirigée de la TyrRS de *Bacillus Stearothermophilus* sert de validation à ce protocole car des mesures expérimentales pour une quinzaine de mutants sont disponibles.

---

## Abstract

Although most proteins are usually synthesized with L-amino acids, D-amino acids are present in organisms in their free states, included in shorts peptides and even in particular proteins. Introducing D-amino acids in proteins in a controlled way can provide a biotechnological advantage, adding new function to proteins. Tyrosyl-tRNA synthetase (TyrRS) has already been used to incorporate more than thirty tyrosine analogues in proteins. Using the D-enantiomer of the tyrosine can give resistance to proteases in cells. Because proteases recognizes less proteins with such amino-acids, peptides bearing those will stay active much longer. TyrRS has a natural though weak activity for D-Tyrosine. The purpose of this thesis is to improve this activity by an *in silico* directed evolution method. The method consists in randomly mutating a protein to select mutants with the desired properties, mimicking mechanism of natural evolution. We use for this study an in house program of directed evolution. It is based on an iterative optimization procedure of both protein sequence and structure. Some mutants have been predicted by this method and proposed to biochemists for experimental *in vivo* validations. They find that some mutants present a weak, but measurable, activity for the D-Tyrosine. The field of study has thus been enlarged to other mutation positions around the tyrosine ammonium group. We study TyrRS and its binding with either L or D tyrosine and furthermore consider an activated form of the aminoacid, the tyrosinyl-adenylate in both geometry L and D. This compound has been modeled with the *E. coli* TyrRS to make molecular dynamic studies. We have use the large ressources of the CINES (Centre Informatique National de l'Education Supérieure) and our laboratory cluster to simulate a maximum of molecular dynamics on our predicted mutants. Another goal of this work was to improve our in house protein design methods with a more sophisticated solvent model. In order to do so, we used Born's model based on the Poisson-Boltzmann theory. Rather than describing solvent as individual molecules, this model considers it as an electrostatic continuum surrounding the protein. We validated that new protocol on samples presenting experimental data to compare with. We successfully tested three small peptides before further simulations were carried out in the same way on a *Bacillus stearothermophilus* TyrRS mutants set.



---

## Abréviations

aa : acide aminé

aaRS : aminoacyl-ARNt synthétase Pour une aaRS donnée, "aa" sera remplacé par le code à trois lettres de l'acide aminé spécifiquement reconnu par l'enzyme

aa-ARNt : aminoacyl-ARNt

aa-AMP : aminoacyl adénylate

ADN : acide desoxyribonucléique

ARN : acide ribonucléique

ARNm : acide ribonucléique messager

ARNt : acide ribonucléique de transfert

ARNt<sup>aa</sup> :acide ribonucléique de transfert spécifique d'un acide aminé aa, ce dernier représenté par son code à trois lettres

ATP : adénosine 5'triphosphate

Kcat : constante catalytique

Km : constante de Michaelis

M : molaire

Vmax : vitesse initiale maximale

Eco-TyrRS : TyrRS d' *Escherichia coli*.

Bst-TyrRS : TyrRS de *Bacillus stearothermophilus*

Mja-TyrRS : TyrRS de *Methanocaldococcus jannaschii*

Sce-TyrRS : TyrRS de *Saccharomyces cerevisiae*

CPD : Computational protein design

PBE : Equation de Poisson-Boltzmann

PBFE : Poisson-Boltzmann Free Energy

GB : Born généralisé

[séquence/structure] : l'ensemble de la séquence d'acide aminé et du jeu de rotamères prédit à l'issu d'un cycle heuristique.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	De l'ADN aux protéines . . . . .	1
1.1.1	Transcription . . . . .	1
1.1.2	Traduction . . . . .	3
1.1.3	Code génétique . . . . .	5
1.1.4	Les ARNt . . . . .	9
1.2	Les aminoacyl ARNt synthétases . . . . .	12
1.2.1	Rôle biologique des aminoacyl-ARNt synthétases . . . . .	12
1.2.2	Deux classes d'aminoacyl-ARNt synthétases . . . . .	17
1.2.3	Evolution des aminoacyl-ARNt synthétases . . . . .	21
1.3	La tyrosyl-ARNt synthétase . . . . .	25
1.3.1	La TyrRS dans la classification des synthétases . . . . .	25
1.3.2	Structures de la TyrRS . . . . .	25
1.3.3	Discrimination des ARNt . . . . .	32
1.3.4	Différences structurales entre les TyrRS de différents règnes . . . . .	33
1.3.5	Spécificité de la TyrRS . . . . .	37
1.3.6	Le rôle du magnésium . . . . .	38
1.3.7	Inhibiteurs . . . . .	39
1.4	Evolution dirigée de la TyrRS . . . . .	40
1.4.1	Charger des acides aminés non naturels. . . . .	40
1.4.2	Acides aminés non naturels déjà incorporés dans des protéines . . . . .	45
1.5	Les acides aminés D . . . . .	49
1.5.1	Présence d'acides aminés D dans les organismes . . . . .	49
1.5.2	Des peptides d'origine animale contenant des acides aminés D . . . . .	50
1.5.3	Acides aminés D sous forme libre dans les cellules vivantes . . . . .	52
1.6	Interactions entre la TyrRS et la D-Tyr . . . . .	55
1.6.1	Etude expérimentale de la TyrRS native liée à la D-Tyr . . . . .	55
1.6.2	Modélisation moléculaire de la TyrRS et de la D-Tyrosine . . . . .	56
<b>2</b>	<b>Le "dessein" computationnel de protéines, ou CPD</b>	<b>61</b>
2.1	La fonction d'énergie . . . . .	66
2.1.1	Mécanique moléculaire, champ de force . . . . .	66
2.1.2	Paramétrisation des principaux champs de forces . . . . .	70
2.1.3	Des approches plus empiriques . . . . .	72
2.1.4	Champ de force à gros grains . . . . .	75
2.2	Modélisation implicite du solvant . . . . .	76
2.2.1	Interactions soluté-solvant . . . . .	77
2.2.2	Modèle de Poisson Boltzmann . . . . .	80

2.2.3	Modèle de Born Généralisé . . . . .	84
2.2.4	Comparaisons de ces modèles . . . . .	85
2.3	L'espace conformationnel . . . . .	87
2.3.1	Représentation de l'état replié . . . . .	87
2.3.2	Représentation de l'état déplié . . . . .	90
2.4	Algorithmes d'optimisation appliqués au CPD . . . . .	90
2.4.1	Méthodes stochastiques . . . . .	91
2.4.2	Méthodes deterministic . . . . .	93
2.5	Dynamique moléculaire . . . . .	95
2.6	Quelques applications du Protein Design . . . . .	97
2.6.1	Etude du repliement . . . . .	98
2.6.2	Protéines entières, mini-protéines . . . . .	98
2.6.3	Interaction protéine-protéine . . . . .	99
2.6.4	Interaction protéine-Ligand . . . . .	100
<b>3</b>	<b>Paramétrisation et validation du modèle de solvant CASA</b>	<b>105</b>
3.1	Paramétrisation du modèle . . . . .	105
3.1.1	Forme de la fonction d'énergie . . . . .	105
3.1.2	Description de l'état déplié . . . . .	106
3.1.3	Optimisation des coefficients de surface : choix des types atomiques	108
3.1.4	Calcul de stabilité : méthode . . . . .	109
3.1.5	Protocole d'optimisation des coefficients surfaciques . . . . .	111
3.1.6	Résultats de l'optimisation des coefficients . . . . .	112
3.1.7	Validation croisée . . . . .	114
3.2	Calcul d'affinité protéine-ligand et protéine-protéine . . . . .	116
3.2.1	Données expérimentales . . . . .	116
3.2.2	Calculs d'affinité protéine-ligand : résultats . . . . .	118
3.2.3	Calculs d'affinité protéine-protéine : résultats . . . . .	121
<b>4</b>	<b>Dessein computationnel de la TyrRS : méthodes</b>	<b>127</b>
4.1	Etape un : construction des systèmes . . . . .	127
4.1.1	Construction du complexe TyrRS:Ligand . . . . .	129
4.1.2	Choix des positions à muter . . . . .	131
4.1.3	Construction des quatres ligands et de leurs rotamères . . . . .	132
4.1.4	Bibliothèque de rotamères pour la protéine . . . . .	134
4.2	Etape deux : calcul de la matrice d'énergie . . . . .	136
4.3	Etape trois : optimisation des séquences et structures . . . . .	140
4.4	Etape quatre : reconstruction des structures 3D et estimation d'affinités . . . . .	140
4.4.1	Reconstruction des structures 3D . . . . .	140
4.4.2	Calcul d'affinité . . . . .	142
4.4.3	Réoptimisation des meilleures séquences et étude de leurs spécificités	142
4.5	Etape cinq : étude par dynamique moléculaire . . . . .	143
4.5.1	Simulations de dynamique moléculaire . . . . .	143
4.5.2	Estimation de l'énergie libre d'association protéine:ligand avec un modèle de Poisson-Boltzmann . . . . .	144

<b>5 CPD de la TyrRS : Caractérisation des séquences et structures obtenues</b>	<b>147</b>
5.1 TyrRS native d' <i>E. coli</i> avec différents ligands . . . . .	147
5.1.1 Comparaison à la structure cristallographique . . . . .	148
5.1.2 Variabilité des structures selon le ligand. . . . .	152
5.2 Evolution dirigée de la TyrRS : analyse générale des séquences obtenues . . . . .	163
5.3 Ré-optimisation de toutes les séquences. . . . .	167
5.3.1 TyrRS d' <i>E. coli</i> . . . . .	167
5.3.2 TyrRS de <i>M. jannaschii</i> . . . . .	170
5.4 Caractérisation des structures obtenues . . . . .	174
5.5 Conclusions . . . . .	194
5.6 Appendice. Résultats supplémentaires pour <i>M. jannaschii</i> . . . . .	195
<b>6 CPD de la TyrRS d'<i>E. coli</i>: Analyse des résultats d'évolution dirigée</b>	<b>203</b>
6.1 Choix des séquences à analyser . . . . .	203
6.1.1 CPD de la TyrRS liée à la D-Tyr . . . . .	203
6.1.2 CPD de la TyrRS liée au L-TyrAMP et au D-TyrAMP . . . . .	206
6.2 Analyse générale des séquences obtenues . . . . .	208
6.2.1 Vue d'ensemble . . . . .	208
6.2.2 Etude de la TyrRS native . . . . .	211
6.2.3 Séquences prédites avec le ligand D-Tyr . . . . .	214
6.2.4 Séquences prédites avec L-TyrAMP ou D-TyrAMP . . . . .	222
6.2.5 Séquences d'affinité moindre . . . . .	227
6.3 Etude par dynamique moléculaire et confrontation aux résultats expérimentaux . . . . .	233
6.3.1 Etude de la TyrRS native . . . . .	234
6.3.2 Séquences prédites avec le ligand D-Tyr . . . . .	238
6.3.3 Comparaison aux résultats expérimentaux . . . . .	246
6.3.4 Séquences prédites avec L- ou D-TyrAMP . . . . .	249
6.3.5 Séquences d'affinité moindre . . . . .	259
6.4 Conclusion . . . . .	261
6.5 Vers de nouvelles positions à muter pour mieux lier la D-Tyrosine . . . . .	264
<b>7 Conclusion</b>	<b>267</b>
<b>8 Annexes</b>	<b>271</b>
8.1 CPD avec le champ de force "tout atome" AMBER et le modèle de solvant GB-HCT . . . . .	271
8.1.1 Champ de force AMBER . . . . .	271
8.1.2 modèle de solvant GB . . . . .	272
8.1.3 Une approximation des rayons de solvatations . . . . .	272
8.1.4 Validations expérimentales . . . . .	274
8.1.5 Conclusion . . . . .	281
<b>9 Remerciements</b>	<b>283</b>

---

*TABLE DES MATIÈRES*

# Chapitre 1

## Introduction

### 1.1 De l'ADN aux protéines

L'information génétique passe successivement de l'ADN à l'ARNm par le mécanisme de transcription, puis de l'ARNm aux protéines par le mécanisme de traduction. Au cours de cette dernière, la lecture des codons individuels est réalisée par les ARNt. Ces derniers doivent être correctement placés en regard des codons de l'ARNm sur le ribosome. Les aa-ARNt aminoacylés permettent de faire le pont entre le monde des acides nucléiques et celui des protéines. Pour comprendre l'importance de la formation de l'aminoacyl-ARNt, ou aa-ARNt, faisons tout d'abord un bref rappel sur cet ensemble de processus biologiques.

#### 1.1.1 Transcription

La transcription utilise les propriétés de complémentarité des bases azotées pour synthétiser un brin d'ARN à partir d'un brin d'ADN. L'ARN polymérase, une hélicase, de nombreux facteurs de transcriptions, ainsi que des séquences consensus sur l'ADN lui-même sont nécessaires à cette opération. La transcription se déroule en trois phases : initiation, elongation et terminaison. Des mécanismes de régulation entrent en jeu à chaque phase.

**Chez les eucaryotes** La transcription se déroule dans le noyau cellulaire. La « boîte TATA » est située environ 30 paires de bases avant l'origine de la transcription ; elle

va fixer l'ARN polymérase II. L'initiation commence avec la protéine TFII D, qui se fixe sur la boîte TATA. Ensuite, différents facteurs de transcription viennent s'assembler sur ce "noyau" : TFII A, TFII B, TFII E, TFII F (hélicase ATP-dépendante) et l'ARN polymérase II. Des séquences d'ADN trans-activatrices ou cis-activatrices sont également requises. L'élongation peut alors commencer. L'élément central de l'élongation est la phosphorylation du domaine CTD (Carboxyl Terminal Domain) de l'ARN polymérase II. Sa phosphorylation (par TFII H) va déplacer l'ARN polymérase jusqu'à l'origine de la transcription. L'addition séquentielle des ribonucléotides peut alors démarrer. La terminaison est assurée par des signaux spécifiques dont le signal de polyadénylation de l'ARNm, "AATAAA". La polymérase continue sa transcription un peu après ce motif puis est libérée sous l'action de divers facteurs. La transcription proprement dite est terminée, mais l'ARN obtenu doit encore être maturé. L'ARN est clivé au niveau du signal de polyadénylation et la polyA polymérase ajoute de nombreux résidus adénine, formant la queue polyA, essentielle à la stabilité de l'ARN. À l'autre extrémité, une coiffe méthylguanosine est ajoutée, nécessaire pour la reconnaissance par les ribosomes lors de l'étape de traduction. L'ARNm subit ensuite une opération d'épissage : un complexe nucléoprotéique (le spliceosome) reconnaît les introns et les élimine.

**Chez les procaryotes** De nombreuses différences existent par rapport à la transcription chez les eucaryotes. La transcription a lieu dans le cytoplasme bactérien puisque c'est là que se trouve l'ADN, sous forme de chromosome ou de plasmide, la bactérie n'ayant pas de noyau. Elle se déroule toujours en trois étapes : l'initiation, l'élongation et la terminaison. L'initiation s'effectue au niveau du promoteur, immédiatement en amont de la séquence codante. Il existe un seul type d'ARN polymérase chez les procaryotes. Elle est constituée de 5 sous unités : deux  $\alpha$ ,  $\beta$ ,  $\beta'$  et  $\sigma$ . La sous unité  $\sigma$  reconnaît une séquence consensus dans le promoteur du gène, classiquement, la "Pribnow-Box" [TATAAT], se fixe sur l'ADN et recrute les autres parties du complexe enzymatique. Une fois l'ARN polymérase mise en place, la sous unité  $\sigma$  se détache du complexe et  $\beta'$  assure la liaison à l'ADN : c'est le complexe "fermé". L'ARN polymérase déroule ensuite la double hélice sur dix-sept paires de bases environ : c'est le complexe "ouvert". La phase d'élongation

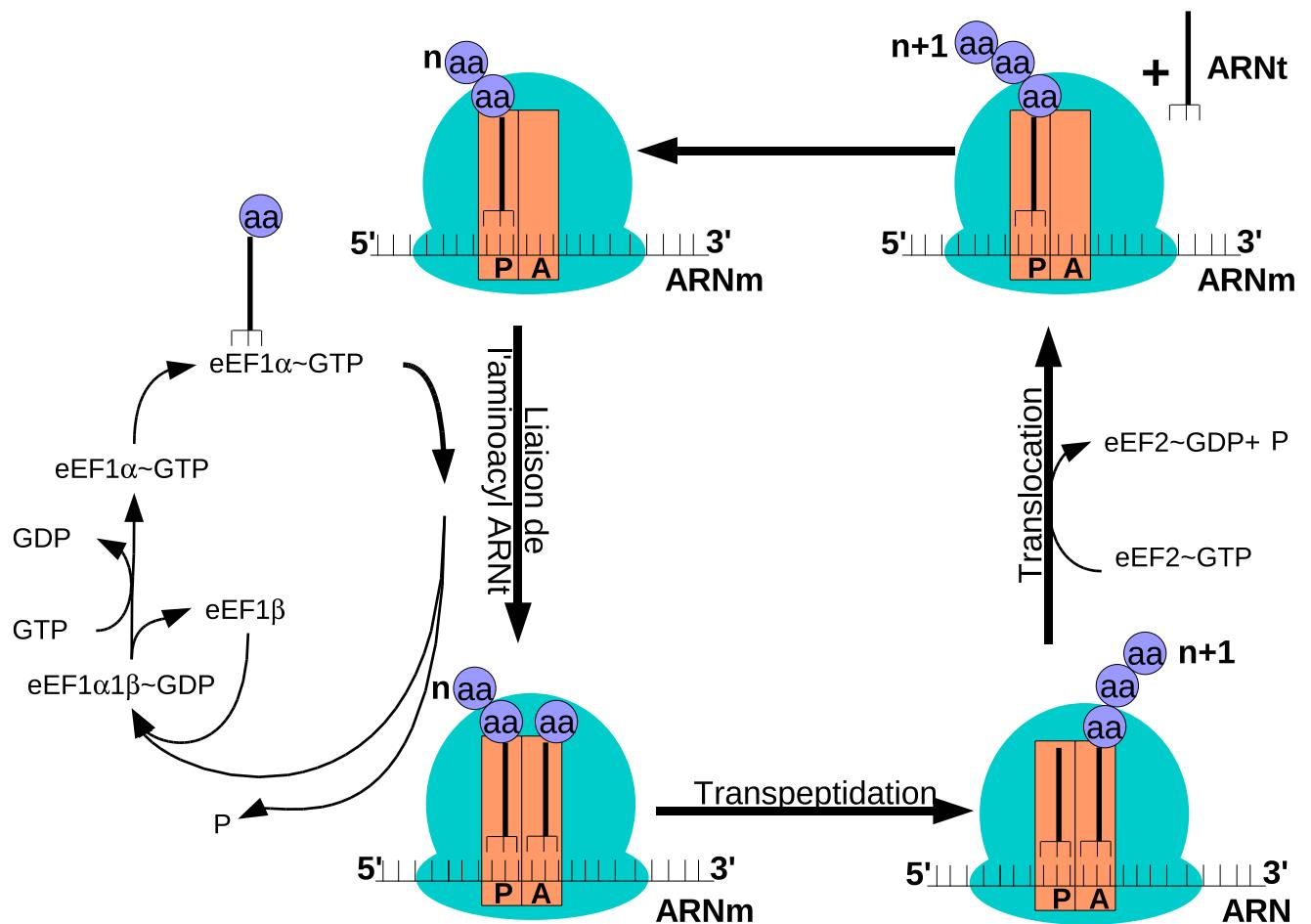
peut commencer. Le complexe enzymatique synthétise alors le brin d'ARN complémentaire du brin codant. La terminaison de la transcription se fait au niveau du terminateur, situé après la séquence codante. Des séquences d'ADN riches en paires G-C ralentissent la progression de l'ARN polymérase. La formation d'une boucle en épingle à cheveux entre deux régions complémentaires de l'ARN bloquent l'enzyme. Enfin une protéine de terminaison (Rho) libère la molécule d'ARN. L'ARNm ainsi produit est directement utilisable par la machinerie cellulaire pour être traduit. Comme la transcription et la traduction se déroulent dans le même compartiment, il est possible que des ribosomes commencent à traduire un ARNm avant que sa transcription ne soit terminée.

### 1.1.2 Traduction

**Chez les eucaryotes** Les sous-unités du ribosome (40S et 60S) s'assemblent au niveau du codon initiateur ( ${}^5\text{AUG}_{3'}$ ), c'est l'initiation, à laquelle participent de nombreuses protéines, ou "facteurs d'initiations". L'ARNt<sup>im</sup> se fixe au niveau du site P (peptidyl) du ribosome. Ensuite vient l'elongation schématisée dans la figure 1.1 : le ribosome se déplace vers l'extrémité 3' de l'ARNm ; les ARNt<sup>aa</sup> se logent successivement dans le site A (aminoacyl) du ribosome, en respectant évidemment la séquence de l'ARNm. Ce décodage de l'information génétique est rendu possible par la correspondance entre l'anticodon de chaque ARNt et l'acide aminé propre à cet ARNt. Cette correspondance est donnée par le code génétique. L'acide aminé fixé sur l'ARNt (site P) se détache et se lie avec l'acide aminé qui est lui-même lié à l'ARNt (site A), c'est l'elongation. L'ARNt du site P est libéré ; celui qui était dans le site A vient dans le site P. Le site A est alors libre et peut accueillir le prochain ARNt<sup>aa</sup>. Arrivé au codon terminateur (UAA, UAG ou UGA), comme il n'existe pas d'ARNt correspondant, le ribosome se disloque : c'est la terminaison. Elle met en jeu des facteurs de terminaison, ou RF (release factor) qui permettent le décrochage du peptide lié à l'ARNt (site P), donnant ainsi la protéine.

**Chez les procaryotes** (figure 1.2) La traduction chez les procaryotes est très proche de celle des eucaryotes. On a 3 facteurs d'initiation : IF1, IF2 : responsable de la fixation

FIGURE 1.1 – Traduction chez les eucaryotes : Aminoacylation de l'ARNt<sup>aa</sup> avec participation du GTP et deux facteurs d'elongation eEF1 (eucaryote elongation factor) et eEF2. Liaison de l'ARNt<sup>aa</sup> dans le site A. Le transfert du peptide est spontané. Translocation grâce à un facteur eEF2, avec consommation d'un GTP. Libération de l'ARNt<sup>aa</sup>. La terminaison se fait grâce à un facteur RF qui a une structure tridimensionnelle très proche de celle d'un tRNA. Le RF ne se fixe que lorsque le codon est un codon stop.



de l'ARNt sur la sous unité 30S du ribosome (IF2 $\alpha$  pour AUG et IF2 $\beta$  pour GUG), IF3 qui aide à fixer l'ARNm sur la sous unité 30S. Tant que l'ARNm n'est pas fixé à la sous unité 30S, il empêche la sous unité 50S de se fixer. L'hydrolyse du GTP (sur IF2) permet l'assemblage des sous-unités 30S et 50S pour donner la particule 70S. La terminaison se fait grâce à 3 facteurs de terminaison : RF1, RF2 et RF3.

### 1.1.3 Code génétique

Le code génétique standard constitue un premier outil pour une compréhension de la biologie moléculaire. La relation entre les triplets de bases de l'ADN et les séquences d'acides aminés dans les protéines a été clairement établie [1, 2] (figures 1.3, 1.4). La synthèse de la chaîne polypeptidique démarre généralement avec une méthionine, codée par un codon AUG, ou bien, chez les bactéries, par un codon GUG, UUG ou CUG. L'arrêt de la synthèse protéique est spécifié par l'un des trois codons stop : UAA, UAG et UGA. Les acides aminés ont plusieurs codons possibles en raison de la dégénérescence du code génétique. Il existe, en effet, 61 codons pour 20 acides aminés. Cette dégénérescence permet de minimiser l'effet de mutations, car une mutation ponctuelle de l'ADN n'engendrera pas forcément un changement d'acide aminé dans la protéine traduite. Même si l'acide aminé change, le nouvel acide aminé sera souvent de même nature chimique, ce qui réduit l'impact sur la protéine. Ceci est possible par une correspondance stricte des bases en 5' et une dégénérescence en 3'. Ainsi, pour huit acides aminés, la troisième base du codon est sans importance (Leu, Val, Ser, Pro, Thr, Ala, Arg, Gly). Les anticodons de l'ARNt doivent donc être capables de reconnaître plusieurs codons différents. Crick émet l'hypothèse de l'interaction "bancale", où Wobble, qui explique le fait qu'un ARNt puisse lire plusieurs codons synonymes [3]. La première base de l'anticodon de l'ARNt en regard de la troisième base du codon de l'ARNm est appelée base wobble ou flottante. On pourrait s'attendre à ce qu'il existe 61 ARNt ; à cause des interactions bancales, il en existe en fait seulement 31.

FIGURE 1.2 – Traduction chez les procaryotes : Aminoacylation de l'ARNt<sup>aa</sup> avec participation du GTP et d'un facteur d'élongation EFTu/Ts. EFTu est lié soit avec GDP soit avec EFTs (Elongation Factor T stable/unstable). Liaison de l'ARNt<sup>aa</sup> au site A, avec hydrolyse du GTP en GDP. Transfert du peptide (spontané). Translocation grâce au facteur EF-G-GTP, consommation d'un GTP par translocation d'un codon. Libération de l'ARNt (site P).

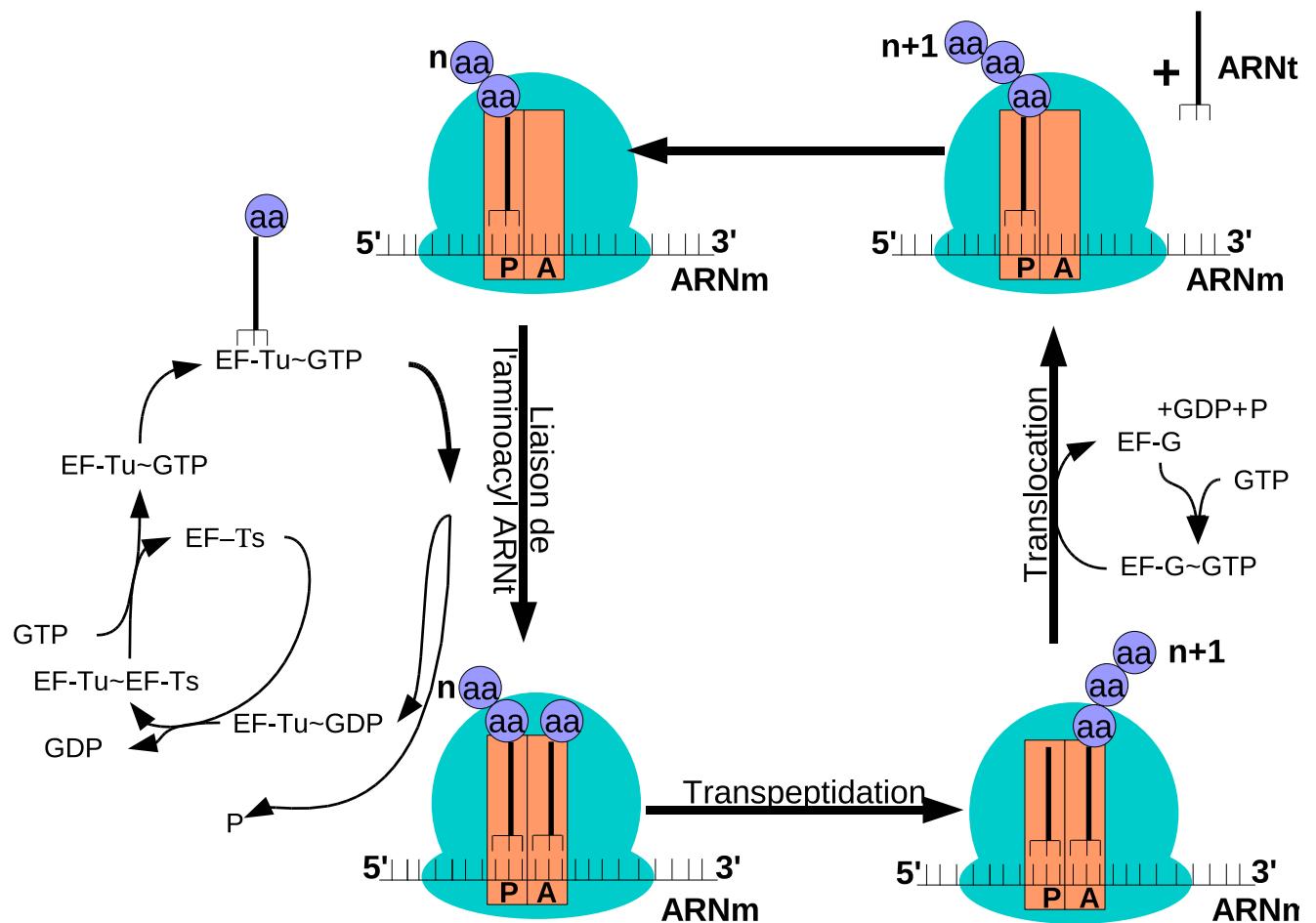
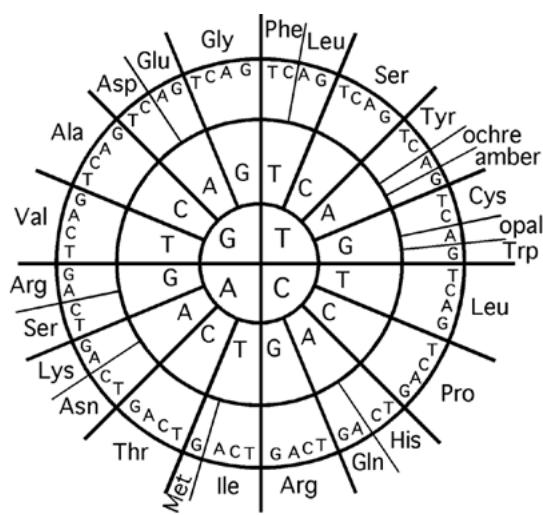


FIGURE 1.3 – Le code génétique.

	Deuxième lettre								
	U		C		A		G		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	<b>UAA</b>	<b>Stop</b>	<b>UGA</b>	<b>Stop</b>	A
	UUG	Leu	UCG	Ser	<b>UAG</b>	<b>Stop</b>	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	UGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	UGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	UGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	UGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	<b>AUG</b>	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
A	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

FIGURE 1.4 – Le code génétique. Représentation circulaire [4].



**Complexité du code génétique** Le code génétique s'avère complexe. Il y a d'une part des exceptions et d'autre part des ambiguïtés. Tout d'abord, un même codon code généralement pour un acide aminé, mais ce n'est pas toujours vrai, comme le montrent les trois exemples ci-dessous :

(1) Le contexte cellulaire est important. Un codon pourra être traduit différemment selon la localisation sub-cellulaire. Le code varie par exemple pour la mitochondrie. Un codon peut aussi être interprété de manière différente d'une espèce à l'autre. Par exemple, le codon AGG peut être traduit en arginine, sérine, glycine, ou même ne pas coder du tout pour un acide aminé [5].

(2) Le contexte moléculaire entre en jeu. Ainsi la sélénocystéine (Sec, ou U) est une modification de la cystéine dans laquelle l'atome de soufre est remplacé par du sélénium. Le codon d'ARNm pour la sélénocystéine est UGA. UGA est en général un codon stop, mais en réalité, il est contexte dépendant et prend une signification différente selon la structure de l'ARNm [6]. La pyrolysine (Pyl, ou O) : utilise un autre codon stop, UAG qui est également contexte dépendant. La méthionine (Met, ou M) est l'exemple le plus familier, le codon AUG représentant dans certains cas un codon start.

(3) Il y a parfois une influence de l'environnement. Ainsi une ambiguïté pour le codon CUG a été détectée dans la levure *Genus candida* [7]. Dans cet organisme, le codon CUG est traduit parfois en sérine et parfois en leucine en réponse à un stress de l'environnement.

**Différence d'utilisation des codons entre organismes** L'abondance des vingt acides aminés varie assez peu entre organismes. En revanche, les fréquences des différents codons varient beaucoup. Dans la liste ci-dessous, on compare trois organismes et leur utilisation de codons codant pour l'arginine. Si tout était égal par ailleurs, on s'attendrait à ce que chacun des six codons Arg soit représenté avec une fréquence d'environ 16,5%. En effet, chez l'être humain, il n'existe qu'un léger biais qui défavorise le codon CGT. Par contre, chez *S. cerevisiae*, les codons CGG, CGC et CGA ne représentent respectivement que 2%, 4% et 5% des codons Arg du génome, et la levure ne fabrique que très peu des ARNt portant les anticodons appropriés (table 1.1). Plusieurs études mettent en évidence une relation entre l'usage des codons dans un gène donné et l'efficacité de la traduction du

gène. En particulier, les codons "optimaux", qui favorisent une traduction rapide, sont ceux qui correspondent à des ARNt abondants. En effet, si un ARNm comporte un codon dont l'ARNt correspondant est très rare, il faudra attendre longtemps pour que cet ARNt vienne se positionner sur le ribosome en face du codon.

Cette diversité d'usage des codons a également d'autres causes possibles [8]. Une étude de Kanaya et coll. met en évidence une corrélation entre la spécificité d'usage des codons et l'efficacité de la traduction. Pour *S. pombe* et *C. elegans*, ils trouvent que les codons optimaux correspondent effectivement au nombre d'ARNt correspondants transcrits. Pour *H. sapiens* et *X. Laevis*, il n'y a pas de biais significatif de l'utilisation des codons dans les gènes codant les protéines ribosomales ou histones, ce qui suggère que le facteur premier influençant la variabilité de l'utilisation des codons pour ces espèces n'est pas l'efficacité de la traduction mais la variation du taux de paires de base C-G.

TABLE 1.1 – Préférence des organismes pour certains codon, l'exemple de l'Arginine. % d'utilisation des codons.

codon	<i>H. sapiens</i>	<i>S. cerevisiae</i>	<i>E. coli</i>
AGG	21,5	17,0	3,2
AGA	21,0	54,0	4,3
CGG	18,2	2,0	9,0
CGA	10,7	5,0	5,0
CGT	9,1	17,0	36,0
CGC	19,5	4,0	40,0

#### 1.1.4 Les ARNt

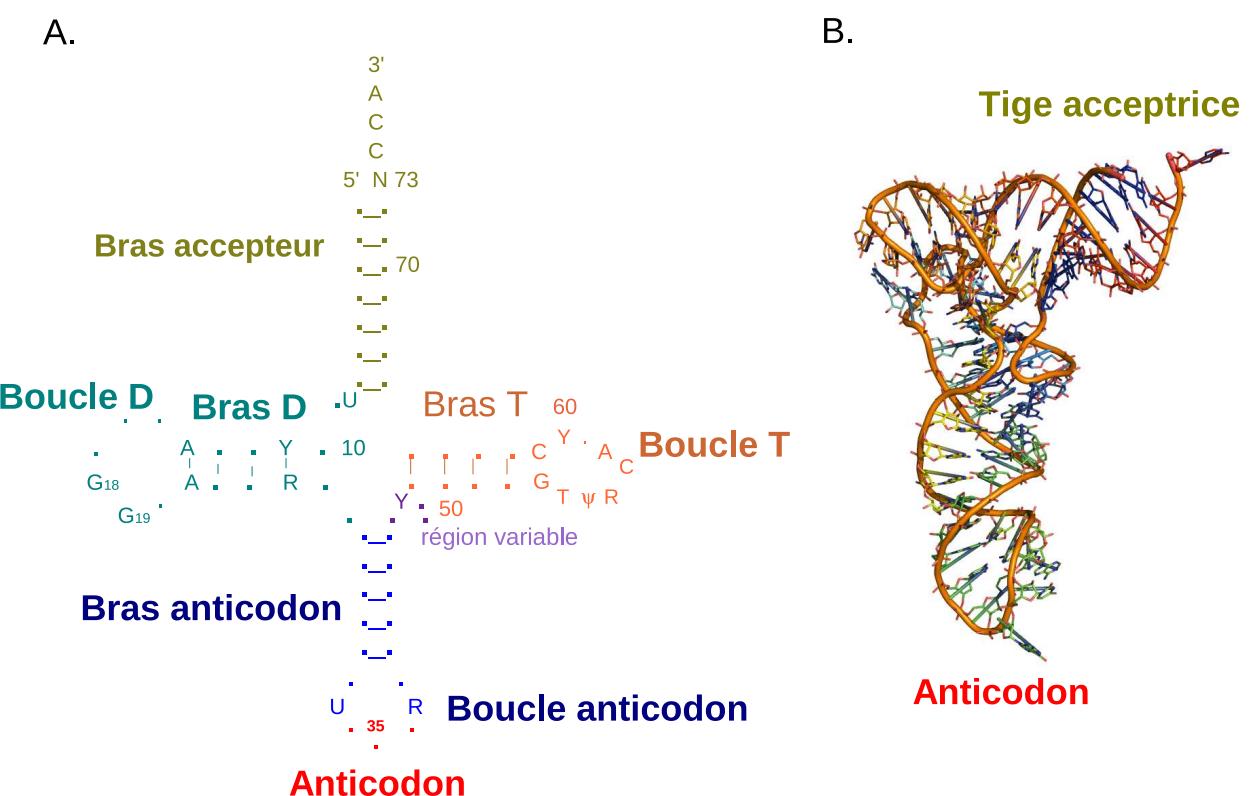
Depuis les années cinquante, on sait que les acides aminés doivent être activés puis liés à une molécule d'ARN de transfert pour jouer leur rôle dans la traduction. L'ARNt est donc une molécule qui joue le rôle d'adaptateur entre les codons de l'ARN messager et les acides aminés. L'existence d'une telle molécule a été postulée par Crick [9] avant même sa découverte effective.

**Structure de l'ARNt** La structure de l'ARNt est complexe. Les ARNt sont de grandes molécules de près de 100 nucléotides. Leur structure secondaire est en forme de trèfle et leur structure tertiaire en forme de L. Le lien entre l'acide aminé et le triplet de nucléotides

correspondant est covalent, en raison de la réaction d'aminoacylation où l'acide aminé est lié à l'ARNt. Les deux composants du code, l'acide aminé et l'anticodon, se trouvent dans des domaines séparés de l'ARNt. La distance entre ces deux éléments est d'environ 75 Å (figure 1.5). 7 paires de bases forment la tige "accepitrice" par l'appariement de nucléotides 5' terminaux et 3' terminaux. L'acide aminé se lie au triplet 5' terminal CCA, situé au bout de la tige accepitrice. Les bras T et D relient la tige accepitrice au domaine de l'anticodon. Le bras D est composé de 4 paires de bases ; le bras T a 5 paires de bases et contient la séquence T $\psi$ C, où  $\psi$  est la pseudouridine (dérivé de l'uracile). Le bras de l'anticodon a 5 paires de bases, il se termine par la boucle anticodon. Des modifications comme la méthylation sur les bases de l'ARNt sont fréquentes. La transcription d'un gène codant pour un ARNt aboutit à un précurseur, qui doit ensuite être clivé pour pouvoir fournir un ou plusieurs ARNt selon l'opéron. Les ARNt ainsi obtenus n'ont pas de 5'CCA<sub>3'</sub> en général. Ils doivent donc subir une maturation par ajout de CCA, ainsi qu'une méthylation.

**Interactions entre acide aminé et ARNt** La tige accepitrice de l'ARNt lie l'acide aminé correspondant à l'anticodon. Le rôle de la tige accepitrice de l'ARNt a été étudié en éliminant l'anticodon. Ceci a permis de créer des "mini-ARNt", formés du seul domaine accepteur. Ces mini-ARNt permettent de poser la question suivante : l'extrémité CCA-3'-OH contient-elle suffisamment d'informations pour conduire à une aminoacylation spécifique ? Sept exemples d'aminoacylations séquence-spécifique de mini-ARNt sont décrits par Shimmel et coll [10]. Ces auteurs décrivent également des "structures minimales" de synthétases capables d'aminoacyler ces mini-substrats. L'interaction entre ARNt minimal et synthétase minimale a dû établir autrefois les bases biochimiques d'un code ARN opérationnel. D'un point de vue historique, ces interactions, bien que parties intégrantes du système actuel, ont dû précéder le code génétique tel qu'il est connu aujourd'hui (figure 1.2). Dans les travaux de Zachau et coll. en 1972 [11], il a été démontré que des fragments d'ARNt<sup>Phe</sup> de levure ne nécessitaient pas l'anticodon pour être aminoacylés. Dans des travaux plus récents, huit acides aminés différents ont été chargés enzymatiquement sur des mini-ARNt de tailles différentes. Certains sont formés de douze paires

FIGURE 1.5 – Structures secondaire (A) et tertiaire (B) d'ARNt classiques. A : séquence consensus d'ARNt canoniques. Y et R indiquent respectivement les pyrimidines et les purines. B : Structure cristallographique de l'ARNt<sup>Tyr</sup> de *M. jannaschii*.



de bases (mini-hélice en structure épingle à cheveux T $\psi$ C) ; d'autres, de sept paires de bases (tige acceptrice en structure épingle à cheveux), d'autres encore, des duplex d'ARN (composés de brins complémentaires de 5 à 12 paires de bases, ou de quatre paires de bases reliées par le motif d'ARN tétra-boucle UUCG). Cette étude a permis de mettre en évidence quelles paires de bases sont importantes pour l'aminoacylation et pour sa spécificité. Ainsi, la base N-76 et les paires de bases 1 – 72 et 4 – 69 sont des "éléments d'identité", importants pour la spécificité (table 1.6).

TABLE 1.2 – Aminoacylation des substrats d'oligonucléotides ARN basée sur l'ARNt accepteur ou sur la tige acceptrice T $\psi$ C

classe I	classe II
Isoleucine	Alanine
Méthionine	Aspartate
Valine	Glycine
	Histidine
	Sérine

## 1.2 Les aminoacyl ARNt synthétases

En 1955, Hoagland décrit les bases du mécanisme d'aminoacylation des ARNt [12]. Par la suite, des enzymes spécifiques de chaque acide aminé ont été identifiées comme responsables de l'attachement entre un acide aminé et l'ARNt correspondant [13]. En 1961, la première purification de ces enzymes fait l'objet d'une publication, elles prennent alors le nom d'aminoacyl-ARNt synthétase (aaRS) [14].

### 1.2.1 Rôle biologique des aminoacyl-ARNt synthétases

**Adénylation des acides aminés** La première étape de la réaction d'activation catalysée par les aaRS est la formation d'un aminoacyl adénylate (aaAMP) à partir de l'acide aminé et de l'ATP :

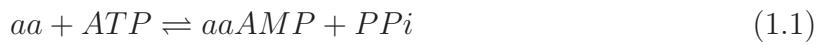
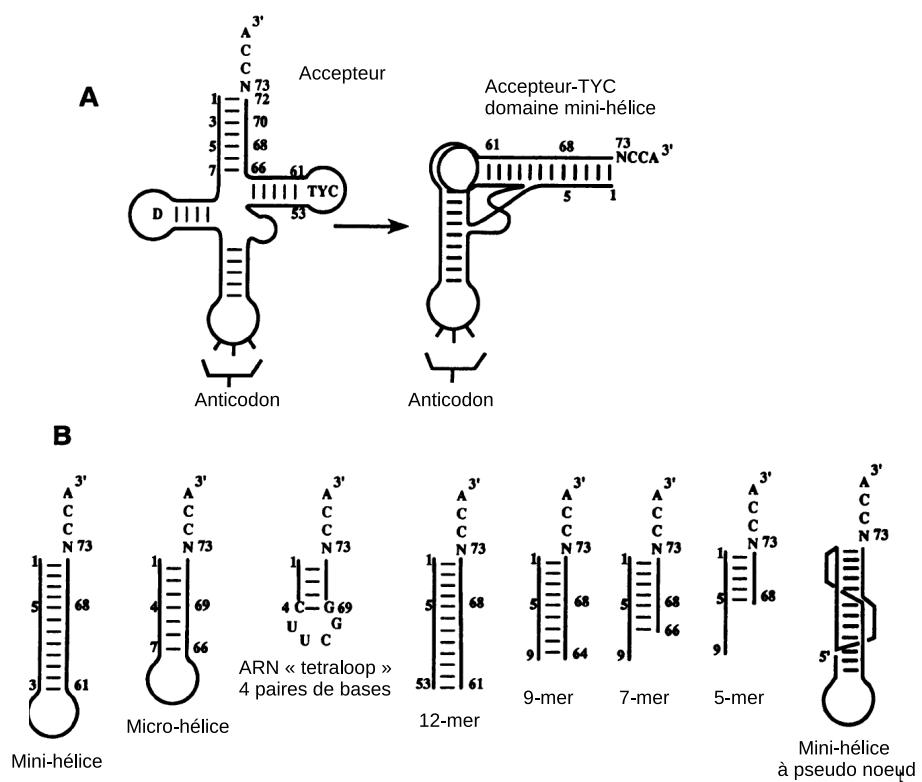
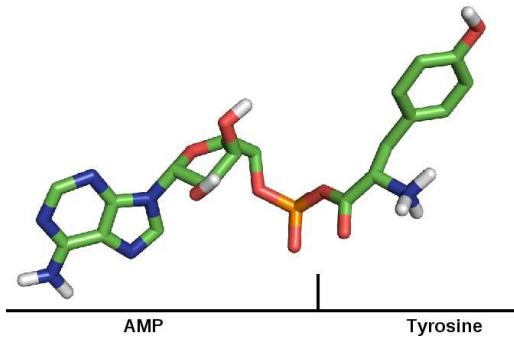


FIGURE 1.6 – A : Structures et représentation de molécule d'ARnt. On voit l'extrémité simple brin NCCA 3' qui est le site d'attachement des acides aminés, le D $\psi$ C et la boucle de l'anticodon. B : Oligonucléotides d'ARN basés sur la tige acceptrice des ARNt, qui sont aminoacylés spécifiquement par des aaRS.

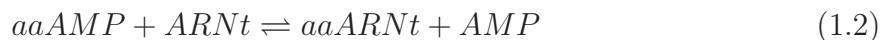


PPi représente le pyrophosphate. Le produit activé est un anhydride mixte dans lequel le groupement carboxyle d'un acide aminé est lié au groupe phosphoryle de l'AMP (figure 1.7).

FIGURE 1.7 – Représentation d'un tyrosyl-adénylate, (ou TyrAMP).



L'étape suivante est le transfert du groupe aminoacyl à une molécule d'ARNt pour former un aminoacyl-ARNt. Celui-ci est un ester d'acide aminé qui lie son groupe carboxyle au groupe hydroxyle 2' ou 3' de l'unité ribose à l'extrémité 3' de l'ARNt.



La somme de ces étapes d'activation et de transfert est :



Ces étapes réactionnelles sont facilement réversibles, car les énergies libres d'hydrolyse des liaisons formées dans l'aminoacyl-adénylate et dans l'aminoacyl-ARNt sont voisines de celle de l'hydrolyse de l'ATP. La réaction est rendue irréversible suite à l'hydrolyse du PPi (formé dans la première réaction) par la pyrophosphatase. La somme des trois réactions est fortement exergonique.

Pour un acide aminé donné, les étapes d'activation et de transfert sont catalysées par la même aminoacyl-ARNt synthétase. En fait, l'aminoacyl-AMP intermédiaire ne se dissocie pas de la synthétase. Il est au contraire étroitement fixé au site actif de l'enzyme par des interactions non covalentes. L'aminoacyl-AMP est normalement un intermédiaire transitoire au cours de la synthèse d'un aminoacyl-ARNt, mais il est très stable et facilement

isolé lorsque l'ARNt est absent du milieu réactionnel.

### 1.2.1.1 Spécificité du site d'activation des acides aminés

Chaque aaRS est spécifique d'un aminoacide donné. Typiquement, une synthétase n'accepte un acide aminé incorrect comme substrat qu'une fois toutes les  $10^4$  à  $10^5$  réactions catalysées *in vitro* [15]. Comment ce niveau de spécificité est-il atteint ? Il y a deux méthodes qui peuvent être conjuguées, selon l'aaRS en question. La première méthode repose sur les propriétés du site catalytique et sa complémentarité avec le substrat. La deuxième méthode fait intervenir un site d'édition, distant du premier site, et chargé de vérifier que l'aminoacylation est correcte. Ce site doit hydrolyser le produit aa-ARNt dans le cas contraire. La majorité des aaRS possède un site d'édition, en plus de leur site d'activation. Ces paires de sites se comportent comme des doubles filtres pour assurer une très haute fidélité.

**Spécificité du site catalytique** En général, le site d'acylation rejette les acides aminés plus gros car il n'y a pas de place pour eux. Le site catalytique utilise également les fonctions chimiques du ligand pour le sélectionner. Considérons le problème auquel doit faire face la ThrRS [16]. La valine a presque la même forme que la thréonine, à ceci près que la valine possède un groupe méthyle à la place du groupe hydroxyle. Comment la ThrRS évite-t-elle de coupler cet acide aminé inapproprié à l'ARNt<sup>Thr</sup> ? Le site actif contient un ion zinc qui lui est relié par l'intermédiaire de deux résidus histidines et d'un résidu cystéine. L'ion zinc est coordonné par le groupement amine de la thréonine, tandis que le groupe hydroxyle de la chaîne latérale de cette dernière forme une liaison H avec un aspartate. Le groupement méthyl de la valine qui se trouve à la place du groupement hydroxyle de la thréonine ne pourrait pas faire une telle interaction. La valine est donc exclue de ce site actif et ne peut être ni adénylée, ni liée à l'ARNt<sup>Thr</sup>. D'autres aaRS ont des stratégies différentes pour reconnaître leur acide aminé. Par exemple, pour l'AspRS, l'aspartate est engagé dans un réseau de liaisons hydrogènes et de ponts salins impliquant les résidus Gln 303[195], Arg 485[489], Glu 344[235], Lys 306[198], Gln 307[199], Asp 342[233], Ser 301[193] et Gln 281[171] (numérotation de *S. cerevisiae* et [*E. coli*]) [26]. Ces

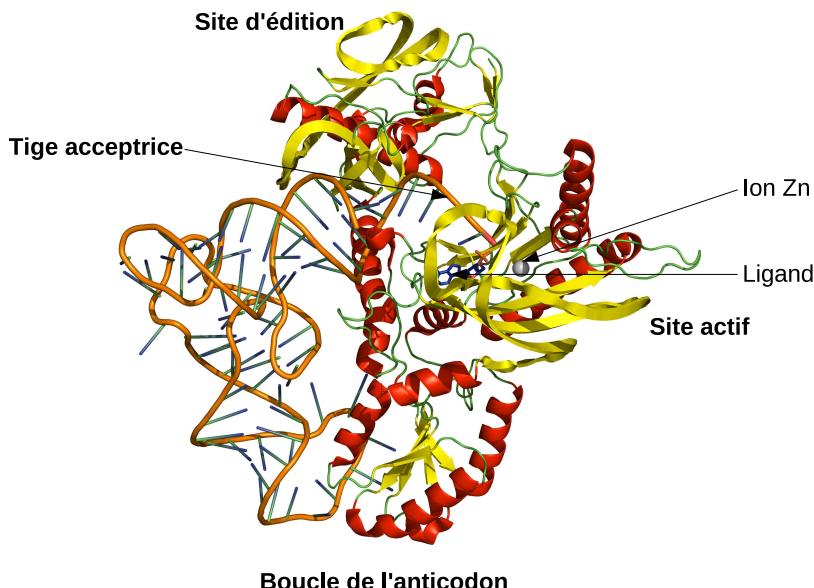
résidus sont strictement invariants chez l'AspRS, à l'exception de ceux interagissant par le biais de leur chaîne principale. Ils sont eux-mêmes maintenus dans des conformations favorables à la reconnaissance de l'aspartate par d'autres résidus légèrement moins bien conservés.

**Spécificité et site d'édition** Le site d'édition clive l'aa-ARNt quand l'acide aminé est plus petit que l'acide aminé correct. IleRS, ValRS, LeuRS sont des synthétases de classe I qui contiennent un domaine d'édition de type "CP1" (Connecting Peptide 1). Ce domaine est inséré dans le domaine d'aminoacylation. Le repliement de CP1 est identique pour ces trois enzymes, quelle que soit l'espèce. Il semble donc que l'insertion de ce domaine ait eu lieu pour un ancêtre commun à ces trois synthétases [17]. Quelques aaRS de classe II, comme ProRS, AlaRS, ThrRS, contiennent un second type de domaine d'édition. Ce domaine est trouvé systématiquement dans toutes les séquences d'AlaRS, ainsi que dans les ThrRS bactériennes et eucaryotes. Les ThrRS d'archéobactérie ont aussi un mécanisme d'édition, mais le domaine en question est différent [18]. Les ProRS sont connues pour mésactiver et mésacyler lalanine et la cystéine. Le domaine d'édition de la ProRS joue un rôle critique dans l'hydrolyse des composés Ala-ARNt<sup>Pro</sup> mais pas pour l'hydrolyse du complexe Cys-ARNt<sup>Pro</sup> pour *E. coli*. Beaucoup de ProRS d'autres organismes ne possèdent pas de domaine d'édition, car ces organismes possèdent d'autres voies de désacylation des ARNt<sup>Pro</sup> mésacylés. Ainsi, les paralogues ProX et YbaK sont capables de désacyler préférentiellement Ala-ARNt<sup>Pro</sup> et Cys-ARNt<sup>Pro</sup>, respectivement [19].

Les aaRSs sont généralement exprimées à partir d'un seul gène, afin d'éviter les mutations qui mèneraient à une moins grande sélectivité du substrat. La bactérie *Deinococcus radiodurans* contient cependant deux versions de la TrpRS : TrpRS I et TrpRS II. La plus divergente des deux, *D. radiodurans*-TrpRS II, interagit avec une synthase d'oxyde nitrique (deiNOS). *In vitro*, *D. radiodurans*-TrpRS II active deiNOS qui produit une forme de nitro-tryptophane, le 4-nitro isomère [20]. Les synthases d'oxyde nitrique de *Streptomyces* produisent pour leur part du 4-nitro tryptophane, utilisé pour la synthèse des toxines de plantes thaxtomine. Tout ceci suggère que les TrpRS II capables de former des intermédiaires ARNt<sup>Trp</sup> modifiés ne contiennent pas de mécanisme de correction qui

éliminerait le substrat modifié [21] (figure 1.8). Les organismes qui perdent leur domaine d'édition pour une synthétase donnée sont sensibles aux variations de concentration de l'acide aminé. Moins l'acide aminé est concentré, moins l'aaRS chargée de l'aminoacyl sera sélective (table 1.3).

FIGURE 1.8 – Représentation Pymol de la structure du complexe ThrRS:ARNt<sup>Thr</sup> d'*E. coli* (code PDB 1QF6). Le bras CCA de l'aminoacyl-ARNt étant mobile, il peut basculer l'acide aminé du site actif vers le site d'édition. Si l'acide aminé n'est pas le bon, il est détaché du site d'édition par hydrolyse.



### 1.2.2 Deux classes d'aminoacyl-ARNt synthétases

Les aminoacyl-ARNt synthétases forment un ensemble d'enzymes hétérogène. La taille des polypeptides, leur séquence, leur état d'oligomérisation peuvent être très différents d'une aaRS à l'autre. En 1983, Blow et coll. mettent en évidence un domaine commun à la TyrRS de *B. stearothermophilus* et à la MetRS d'*E. coli*. Ce domaine est organisé en repliement de Rossmann [23]. D'autre part, la comparaison des séquences de la MetRS et de l'IleRS d'*E. coli*, a permis de mettre en évidence un motif peptidique conservé, le motif HIGH [24]. Cette séquence signature a été retrouvée dans l'ArgRS, la CysRS, la GlnRS,

TABLE 1.3 – La constante de Michaelis,  $K_m$ , est définie comme la concentration de substrat à la moitié de la vitesse maximale de la réaction. On donne ici les  $K_m$  en mM pour la liaison des acides aminés par les aaRS chez *E. coli* (base de données enzymatique BRENDA [22]).

aaRS	ligand	km (mM)
TyrRS	Tyrosine	0.0033
TrpRS	Tryptophan	0.017
ThrRS	Threonine	0.00003
LysRS	Lysine	0.0025
LeuRS	Leucine	0.05
IleRS	Isoleucine	0.0052
AlaRS	Alanine	0.34
ValRS	Valine	0.0019
MetRS	Methionine	7.9
SerRS	Serine	0.3
AspRS	Aspartate	0.06
ProRS	Proline	0.25
GlyRS	Glycine	0.16
CysRS	Cysteine	0.0072
GluRS	Glutamate	0.027
GlnRS	Glutamine	0.11
ArgRS	Arginine	0.009
PheRS	Phenylalanine	0.0103
HisRS	Histidine	0.03
AsnRS	Asparagine	0.015

la GluRS, l'IleRS, la LeuRS, la TyrRS et la ValRS. Une deuxième séquence consensus a été identifiée lors d'expériences de marquage d'affinité sur la MetRS d'*E. coli*; il s'agit du motif KMSKS [25]. Pour les dix autres aaRS ne possédant pas ces séquences consensus, 3 autres motifs qui leurs sont propres ont été identifiés [26]. En 1990, le séquençage de toutes les aaRS d'*E. Coli* et la détermination de quelques structures tridimensionnelles ont permis d'établir une classification des aaRS en deux groupes, ou classes, comprenant dix enzymes chacun. Ces classes diffèrent par leurs structures, leurs modes de fixation à l'ARNt, et leurs modes d'aminoacylation [26] (table 1.4).

TABLE 1.4 – Principales caractéristiques des aaRS et répartition selon leur classe. La structure oligomérique est donnée pour *E. Coli*.

	Classe I	Classe II
	Motifs conservés	
Liaison au substrat	HIGH KMSKS	Motif 2 : ...FRxE Motif 3 : ... $(Gx)_3ER...$
Dimérisation		Motif 1 : ...P...
Repliement du site actif	Feuillet $\beta$ parallèle (Rossman Fold)	Feuillet $\beta$ anti-parallèle
Conformation de l'ATP	Etendue (sauf pour TyrRS)	Repliée
Liaison à l'ARNt		
Bras accepteur	Sillon mineur	Sillon majeur
Boucle variable	Face au solvant	Face à la protéine
Site d'aminoacylation	2'OH	3'OH

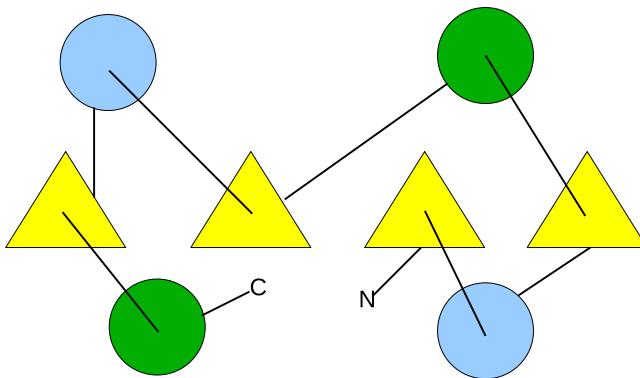
### 1.2.2.1 La classe I

Les enzymes de classe I possèdent toutes un domaine catalytique de 300-400 acides aminés, en repliement de Rossmann, formé de 5 brins  $\beta$  parallèles, associés par des hélices  $\alpha$ . Elles reconnaissent la tige acceptrice de l'ARNt par le petit sillon, tandis que la boucle variable de l'ARNt est dirigée vers le solvant. Elles aminoacylent l'ARNt en 2'OH de la base terminale 76. Elles peuvent être monomériques ou dimériques. La plupart des aaRS de classe I comprennent les deux peptides signatures HIGH et KMSKS [25]. Ces deux motifs conservés interagissent directement avec l'ATP, qui adopte une conformation étendue [27] (figure 1.9).

Trois sous classes ont été déterminées par alignements de séquences d'aaRS de classe

I. La première sous-classe (Ia) fixe essentiellement des acides aminés hydrophobes. La deuxième (Ib) regroupe les aaRS spécifiques d'acides aminés à longues chaînes latérales. La troisième (Ic) active les acides aminés aromatiques.

FIGURE 1.9 – Topologie du domaine catalytique des aaRS de classe I, organisé en Rossman Fold.



### 1.2.2.2 La classe II

Contrairement aux enzymes de classe I, les enzymes de classes II ne possèdent pas de domaine organisé en repliement de Rossmann. Elles possèdent toutes un domaine catalytique organisé autour d'un feuillet  $\beta$  antiparallèle, contenant 3 motifs peptidiques conservés dont la localisation est constante [28]. Elles reconnaissent la tige acceptrice de l'ARNt par le grand sillon, alors que la boucle variable de ce dernier est dirigée vers la synthétase. Elles aminoacylent l'ARNt en 3'OH de la base 76. Les aaRS de classe II sont des dimères ou des tétramères. Le motif 1 est impliqué dans la dimérisation de la protéine, les motifs 2 et 3 appartiennent au site actif. Tout comme la classe I, la classe II peut être subdivisée en trois sous-classes. La première sous classe (IIa) comprend des aaRS présentant des homologies de séquences dans le site actif et dans le domaine C-terminal (à

l’exception de la SerRS pour le domaine C-ter). La région C-terminale est impliquée dans la reconnaissance de l’ARNt. Cette sous-classe charge des acides aminés de petite taille et polaires. La seconde sous-classe (IIb), a un domaine N-terminal organisé en tonneau  $\beta$ , impliqué dans la fixation de l’ARNt. La dernière sous classe (IIc) regroupe toutes les aaRS de classe II dont la structure oligomérique n’est pas conservée [29] (table 1.5).

TABLE 1.5 – Classification des aminoacyl ARNt synthétases

Classe I	structure quaternaire	Classe II	structure quaternaire
Ia		IIa	
ValRS	$\alpha$	SerRS	$\alpha 2$
IleRS	$\alpha$	ThrRS	$\alpha 2$
LeuRS	$\alpha$	ProRS	$\alpha 2$
MetRS	$\alpha, \alpha 2$	GlyRS	$\alpha 2\beta 2$
CysRS	$\alpha$	HisRS	$\alpha 2$
ArgRS	$\alpha$		
Ib		IIb	
GluRS	$\alpha$	AspRS	$\alpha 2$
GlnRS	$\alpha$	AsnRS	$\alpha 2$
LysRS	$\alpha$	LysRS	$\alpha 2$
Ic		IIc	
TyrRS	$\alpha 2$	PheRS	$\alpha 2, \alpha 4$
TrpRS	$\alpha 2$	AlaRS	$\alpha 2, \alpha 2\beta 2$

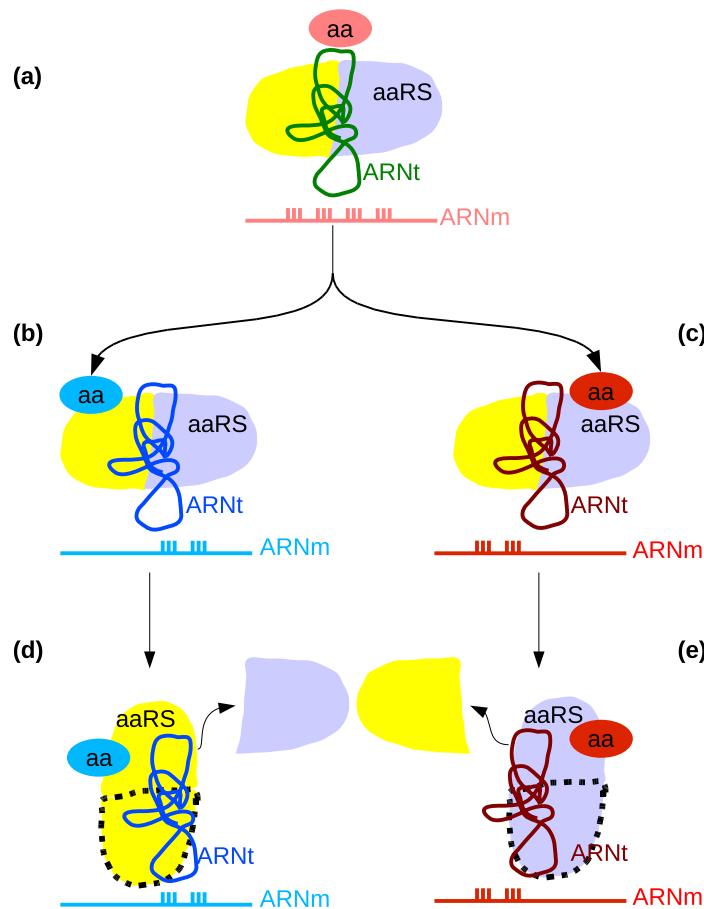
### 1.2.3 Evolution des aminoacyl-ARNt synthétases

Les aaRS auraient émergé à partir du monde des ARN, primitivement sous la forme de ribozymes, puis de ribonucléo-protéines, et enfin de protéines. Les aaRS contemporaines sont équitablement réparties en deux classes distinctes. Cette remarquable symétrie a conduit à l’hypothèse suivante : les deux domaines catalytiques (aaRS primordiales) auraient peut-être évolué à partir de gênes dont le brin sens et antisens auraient simultanément codé pour des ancêtres des deux classes [30]. Une autre théorie suggère que les deux groupes d’enzymes seraient apparus de façon indépendante et auraient fonctionnellement convergé [31]. De façon alternative, les enzymes de classe II seraient les plus anciennes [32]. Les arguments en faveur de cette dernière hypothèse sont entre autres que : (i) les acides aminés reconnus par les aaRS de classe II sont plus simples et plus petits ; (ii) beaucoup

d'entre eux ont des rôles cruciaux dans la structure et/ou activité des protéines (Gly, Pro, Ala, Asp, His, Thr et Ser) et, de façon remarquable, (iii) les substrats minimalistes (minihélices acceptatrices ou mini-ARNt) sont préférentiellement reconnus par les aaRS de classe II [33]. Quoiqu'il en soit, tous les auteurs s'accordent pour dire que la forme primitive des aaRS aurait été réduite au cœur catalytique. Des domaines additionnels se seraient ajoutés de façon concomitante à l'évolution de l'ARNt et à la nécessité d'une spécificité d'aminoacylation accrue (e.g. domaine de liaison à l'anticodon, d'édition,...) ou à l'apparition de nouveaux mécanismes cellulaires (e.g. domaine EMAPII ou "Endothelial monocyte-activating polypeptide" de la TyrRS, impliqué dans l'angiogénèse). Ceci est largement illustré par l'existence de protéines paralogues et par l'organisation modulaire des aaRS contemporaines.

La comparaison des séquences a permis d'établir des liens entre les aaRS, suggérant des ancêtres communs. Par exemple, la duplication et l'évolution du gène d'une AspRS primitive peu spécifique aurait généré l'AspRS spécifique et l'AsnRS. Les empreintes de cette histoire évolutive sont clairement retrouvées dans certains organismes pour lesquels l'AsnRS est manquante et qui conservent alors une AspRS non discriminante, capable d'aminoacyler l'ARNt<sup>Asp</sup> mais aussi l'ARNt<sup>Asn</sup> [34]. Notons la conservation d'une AspRS peu spécifique dans *T. Thermophilus*, alors que l'AsnRS est présente [35]. Ceci pourrait refléter la conservation d'un mécanisme ancestral de biosynthèse d'acides aminés ARNt-dépendante [36]. Une exception à la classification des aaRS est donnée par la LysRS. En effet, une activité LysRS a été purifiée d'extraits de *M. maripaludis* [37], conduisant à l'identification du gène codant pour cette protéine. Celle-ci ne présente aucune homologie avec les LysRS connues (de classe II) mais présente les motifs consensus des synthétases de classe I (HIGH et KMSKS). À l'instar des Arg-, Gln- et GluRS (dont elle se rapproche le plus, d'un point de vue séquence primaire), la LysRS de classe I nécessite l'ARNt<sup>Lys</sup> pour catalyser l'activation de la lysine. Cette LysRS de classe I est retrouvée chez la plupart des archaebactéries et dans quelques bactéries (*Borrelia burgdorferi*, *Rickettsia prowazekii*, *Treponema pallidum*), où elles seraient apparues par transfert horizontal de gène (figure 1.10).

FIGURE 1.10 – Une théorie pour expliquer la séparation des aaRS et des codons au cours de l'évolution du code génétique. La première image (a) décrit une paire de sites actifs de synthétases ancestrales liées à un ARNt (en vert) aminoacylé avec un acide aminé (en rose) codé par un jeu de quatre codons (sur l'ARNt). Le domaine ancestral des classe I est jaune et le domaine ancestral des classe II est violet. L'incorporation d'un nouvel acide aminé dans le code requiert la duplication des éléments de (a), pour générer (b) et (c). Pour préserver la spécificité du code, les deux acides aminés, bleu et rouge, sont pris en charge chacun par un domaine de synthétase. L'identité de l'ARNt change et devient spécifique à un des deux acides aminés, et le jeu de quatre codons original est séparé en deux jeux, chacun étant spécifique d'un des deux ARNt différenciés. Finalement, (d) et (e) représentent l'évolution propre à chaque nouveau complexe après l'étape de spécification pour un acide aminé. Le domaine impliqué dans la reconnaissance de l'acide aminé contient maintenant d'autres domaines (entourés en pointillé) qui lui permettent de mieux reconnaître d'autres régions de l'ARNt, donnant naissance à la structure actuelle des synthétases. Le deuxième composant de la paire ancestrale disparaît.



### 1.2.3.1 D'autres activités des aaRS

Les aaRS ont également des activités secondaires, non reliées à leur fonction catalytique principale (table 1.6). Ces différentes activités dépendent du type d'aaRS et de leur localisation cellulaire [38]. Elles peuvent être impliquées dans certains mécanismes de régulation (ThrRS, AlaRS), dans la fixation à l'ADN (PheRS), dans l'épissage mitochondrial (TyrRS, LeuRS), ou dans l'apoptose (TyrRS humaine). Certaines aaRS sont capables de synthétiser les alarmones Ap4A et Ap3A qui stimulent la synthèse d'enzymes de réparation en réaction à un stress cellulaire [39]. Les aaRS peuvent également aminoacyler des ARNt non canoniques comme les ARN "tRNA-like" des virus de plantes ou le tmRNA (aminoacylé par l'AlaRS).

TABLE 1.6 – Exemples de fonctions d'aaRS autres que l'activité d'aminocylation.

Espèce	aaRS	classes	Localisation	Cible	Activité
<i>H. sapiens</i>	TyrRS	I	Extracellulaire	Endothéliale	cytokine angiogénique
<i>H. sapiens</i>	TrpRS	I	Extracellulaire	Endothéliale	cytokine angiotastique
<i>H. sapiens</i>	LysRS	II	Extracellulaire	Macrophage	cytokine inflammatoire
<i>H. sapiens</i>	LysRS	II	mb Plasmique	HIV Gag	assemblage viral
<i>H. sapiens</i>	GluRS	I	Cytoplasmique	3'UTR	contrôle translationnel
<i>H. sapiens</i>	ProRS	II	Cytoplasmique	3'UTR	contrôle translationnel
<i>H. sapiens</i>	GlnRS	I	Cytoplasmique	ASK1	anti-apoptose
<i>H. sapiens</i>	LysRS	II	Noyau	MITF	contrôle transcriptionnel
<i>H. sapiens</i>	MetRS	I	Noyau	Nucleole	ARNr transcription
<i>N. crassa</i>	TyrRS	I	Mitochondrie	Group I	splicing d'intron
<i>S. cerevisiae</i>	LeuRS	I	Mitochondrie	Group I	splicing d'intron
<i>E. coli</i>	ThrRS	II	Bactérie	5'UTR	contrôle transcriptionnel

## 1.3 La tyrosyl-ARNt synthétase

### 1.3.1 La TyrRS dans la classification des synthétases

La TyrRS appartient aux aminoacyl-ARNt synthétases de classe I. Sa séquence contient les motifs conservés HIGH et KFGKT (qui remplace KMSKS). Fraser et Rich [40] trouvent que le site primaire d’aminoacylation de l’ARNt<sup>Tyr</sup> d’*Eco*-TyrRS est en 2'-OH du ribose de la base terminale A76, comme pour la majorité des aaRS de classe I. Cependant, Sprinzl et Cramer [41] trouvent que l’ARNt peut être aminoacylé en 2'-OH comme en 3'-OH. On note que ces deux groupements hydroxyles sont relativement proches du groupement carboxyle du TyrAMP dans le modèle *Bst*-TyrRS:ARNt<sup>Tyr</sup>. Pour *S. cerevisiae*, l’ARNt<sup>Tyr</sup> peut être chargé en 2'-OH comme en 3'-OH de la base terminale A76. Dans tous ces cas, bien que les  $K_m$  soient proches, c'est la  $V_{max}$  de l’aminoacylation qui est 15 fois supérieure pour une aminoacylation en 2'-OH. D’autres expériences ont montré que pour certaines de ses caractéristiques, la TyrRS est une exception au regard de la classification des aaRS. Ce n'est pas la seule ; par exemple, nous avons vu qu'il existe des LysRS de classe I.

La TyrRS est un enzyme très stable. Pour analyser la stabilité des TyrRS de bactéries, une famille de neuf hybrides entre les TyrRS d'*E. Coli* et *B. stearothermophilus* ont été créées. La stabilité et l'activité de ces hybrides ont été mesurées et leur variation a été analysée quand la position du point de jonction varie sur la séquence. Les résultats ont montré que les deux séquences peuvent être remplacées localement l'une par l'autre et que les hybrides donnent des enzymes TyrRS stables et actives [42].

### 1.3.2 Structures de la TyrRS

Dans la Protein Data Bank (PDB), on trouve 47 structures de TyrRS (en mars 2011). Ces structures proviennent de douze organismes différents, répartis sur les trois règnes du vivant (figure 1.7). La TyrRS a une structure dimérique. Chaque monomère est formé d'un domaine N-terminal très conservé, contenant le site catalytique, et un deuxième domaine qui présente des différences entre eucaryotes/archaeobactéries et bactéries (table 1.11, figure 1.12).

FIGURE 1.11 – Représentation pymol de la structure cristallographique de la tyrosyl-ARNt synthétase d'*E.coli*.

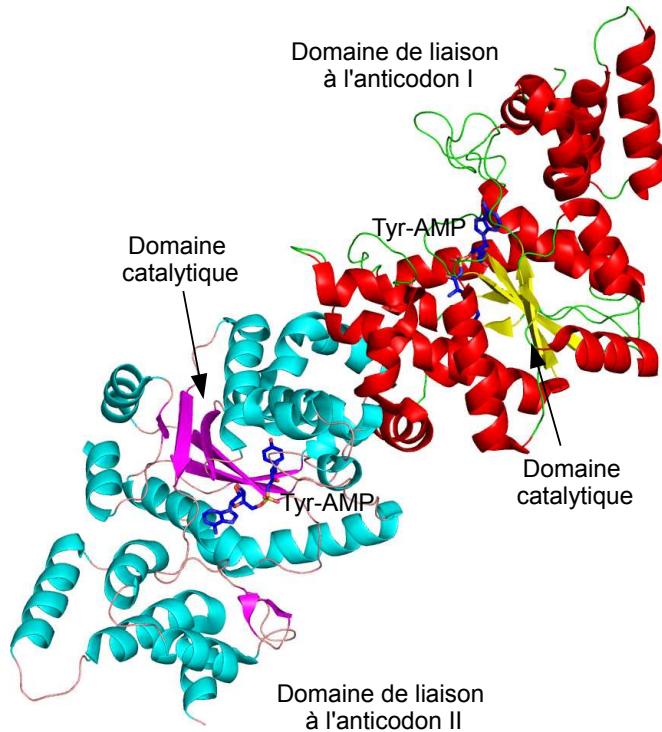


FIGURE 1.12 – Représentation pymol du site actif de la Tyrosyl-ARNt synthétase d'*E.coli*. Le ligand est représenté en bleu, la séquence consensus I (HLGH) est représentée en violet et la séquence consensus II (KFGKT) en vert.

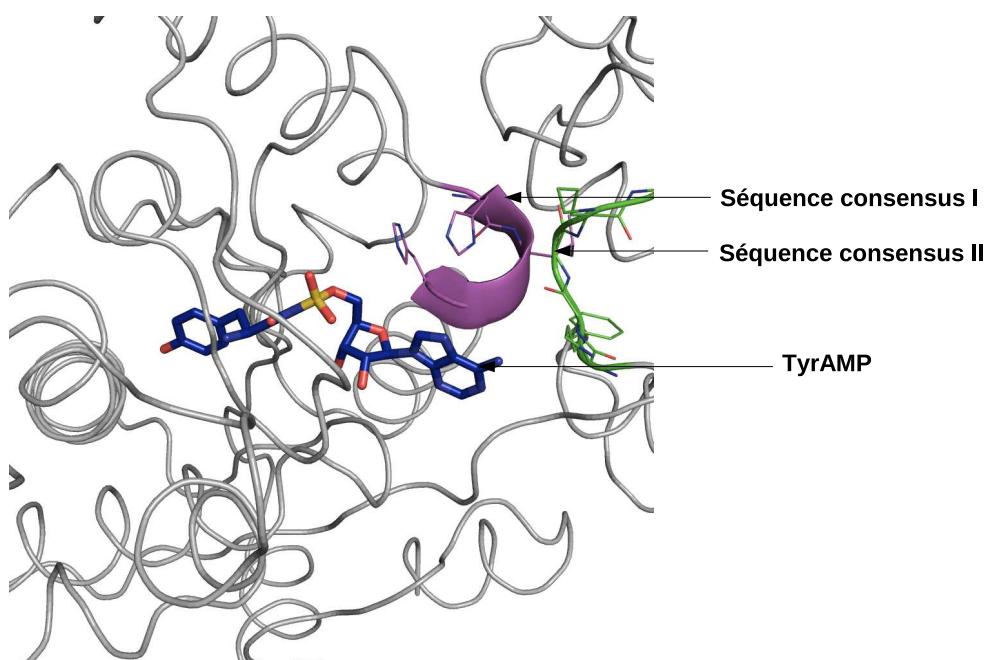


TABLE 1.7 – Quelques structures connues de TyrRS. nb\* : nombre de structures dans la PDB.

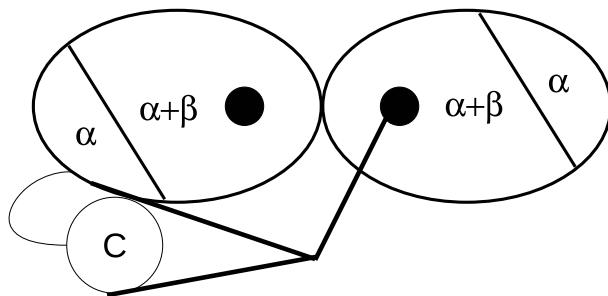
Embranchement	Organisme	nb*	ligands	Référence
Archaeobactérie	<i>M. jannaschii</i>	11	L-Tyrosine Apoenzyme anologue du L-TyrAMP plusieurs ligands	Kobayashi 2003 Zhang 2005 Turner 2006 Liu 2007-2008
Bactérie	<i>B. stearothermophilus</i>	8	L-Tyrosine anologue du L-TyrAMP ARNt <sup>Tyr</sup>	Brick 1989
Eucaryote	<i>H. sapiens</i>	3	anologue du L-TyrAMP Glycerol, SO <sub>4</sub> , PO <sub>4</sub> , K, Phénol	Yang 2004
Bactérie	mitochondrie	1	Apoenzyme	Bonnefond 2007
Bactérie	<i>S. aureus</i>	4	anologue du L-TyrAMP	Qiu 2001
Bactérie	<i>E. coli</i> souche K-12 subs	3	Tyrosine 3-Azido-L-Tyrosine 3-iodo-Tyrosine	Kobayashi 2005 Oki 2008
Bactérie	<i>E. coli</i>	3	L-Tyrosine anologue du L-TyrAMP, SO <sub>4</sub>	Kobayashi 2005
Bactérie	<i>T. thermophilus</i>	3	L-Tyrosine anologue du L-TyrAMP SO <sub>4</sub> , Mg <sup>2+</sup>	Kotik-Kogan 2006
Eucaryote	<i>N. crassa</i>	2	L-Tyrosine	Paukstelis 2005
Archaeobactérie	<i>A. fulgidus</i>	1	L-Tyrosine	Kuratani 2006
Archaeobactérie	<i>P. horikoshii</i>	1	L-Tyrosine	Kuratani 2006
Eucaryote	<i>S. cerevisiae</i>	1	L-Tyrosine et Mg <sup>2+</sup>	Tsunoda 2010
Archaeobactérie	<i>A. pernix</i>	1	SO <sub>4</sub>	Kuratani 2005
Bactérie	<i>M. tuberculosis</i>	1		Hartmann 2010
Virus	<i>A. polyphaga</i> mimivirus	1	Tyrosinal	Aberge 2007

**Compréhension de la structure par des constructions d'hétérodimères** Des hétérodimères de *Bst*-TyrRS (*Bacillus stearothermophilus*-TyrRS) portant différentes mutations ont été construits *in vitro* et utilisés pour prouver que le site de liaison d'une molécule d'ARNt<sup>Tyr</sup> est partagé entre les deux sous unités d'une molécule de TyrRS [43]. Dans la structure cristallographique de *Bst*-TyrRS, les deux sous-unités sont symétriques. Le dimère lie deux molécules de tyrosine et forme deux molécules de TyrAMP. De même, le dimère de *Tth*-TyrRS (*Thermus thermophilus*-TyrRS) lie deux molécules d'ARNt<sup>Tyr</sup> dans le cristal du complexe. Le comportement en solution de *Bst*-TyrRS montre qu'un seul monomère lie l'acide aminé, l'active et aminoacyle l'ARNt. L'autre monomère reste inactif.

Des hétérodimères entre la *Bst*-TyrRS entière et son fragment N-terminal ont été construits in-vitro. De tels hétérodimères comportent deux sites de formation de TyrAMP mais un seul site pour la liaison d'ARNt<sup>Tyr</sup>. Ils ne peuvent transférer la tyrosine à l'ARNt<sup>Tyr</sup> que lorsque le TyrAMP s'est formé dans la sous-unité tronquée. Il a été observé que cet hétérodimère charge le TyrAMP deux fois moins vite que les *Bst*-TyrRS sauvages. La symétrie de *Bst*-TyrRS en solution a été étudiée par l'introduction de différentes mutations dans la protéine entière et dans la protéine tronquée. Les résultats ont montré que chaque dimère est actif en un seul site en solution. Ce site est distribué aléatoirement entre les deux sous-unités. Il n'y a pas d'interconversion détectable entre site actif et inactif au cours du temps. L'enzyme sauvage semble être intrinsèquement asymétrique en solution, même en l'absence de substrats [44] (figure 1.13).

**Détermination des résidus cruciaux pour l'aminoacylation :** Les séquences signatures de la classe I : HIGH, et KMSKS, sont impliquées dans la catalyse de l'aminoacylation de l'ARNt par l'acide aminé activé. D'autres résidus sont également impliqués dans les deux étapes de ce processus, interagissant d'abord avec l'ATP puis avec la base A76 de l'ARNt<sup>Tyr</sup>. La synthèse de TyrAMP à partir de l'ATP et de la tyrosine et son transfert à l'ARNt ont été étudiés en détail pour *Bst*-TyrRS. La création de protéines mutantes et la comparaison au type sauvage ont permis de caractériser l'état stationnaire et pré-stationnaire de la réaction [45]. Ce sont principalement les résidus en contact

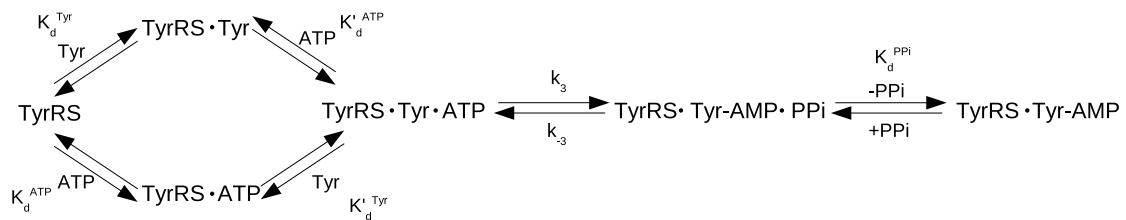
FIGURE 1.13 – Représentation schématique (d'après H. Bebouelle) du complexe entre un héterodimère mutant de TyrRS et d'ARNt<sup>Tyr</sup>. L'hétérodimère est formé d'une sous unité entière de TyrRS et d'une sous unité tronquée de sa partie Cter. Cet hétérodimère contient deux sites potentiels pour la formation de TyrAMP, mais un site seulement pour la liaison de l'ARNt<sup>Tyr</sup>. Seulement une molécule de TyrAMP est formée par hétérodimère de TyrRS. Si du TyrAMP est formé dans la sous unité tronquée de l'hétérodimère, la tyrosine ne pourra pas être transférée du TyrAMP à l'ARNt<sup>Tyr</sup> [43].



avec le TyrAMP dans la structure cristallographique qui ont été mutés. Ainsi, les résidus, Thr40, His45, Lys82, Arg86, Lys230, Lys233 et Thr234 ont été montrés comme jouant un rôle dans la liaison des phosphates  $\beta$  et  $\gamma$  de l'ATP et dans la liaison du pyrophosphate à différents stades de la formation du TyrAMP. Les résidus Thr40, Lys82 et Arg86 ont également un rôle dans l'interaction de l'adénylate avec la base Ade76 de l'ARNt<sup>Tyr</sup> [46]. La structure de *Bst*-TyrRS complexée au tyrosyl-adénylate [47] permet de comprendre le mécanisme réactionnel. Le groupe 2'-OH et le cycle de l'adénine du TyrAMP forment une liaison hydrogène. Cette reconnaissance par liaison hydrogène est très conservée dans les aaRS de classe I. En réponse à la formation de cette liaison hydrogène, la boucle KMSKS se réarrange et adopte une conformation plus compacte. Le motif HIGH reconnaît initialement le phosphate  $\gamma$  puis les autres phosphates de l'ATP, avec un nouveau réarrangement de la structure. Après l'activation de l'acide aminé, la boucle KMSKS adopte à nouveau une configuration semi-ouverte pour accepter l'extrémité CCA de l'ARNt et effectuer la réaction de transfert de l'acide aminé [48] (figure 1.14, table 1.8 d'après [48]).

FIGURE 1.14 – Réactions catalysées par la TyrRS

Première étape: Activation de la Tyrosine



Deuxième étape: Aminoacylation de l'ARNt

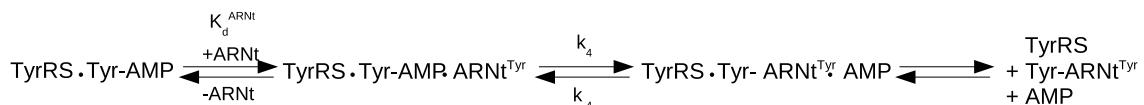


TABLE 1.8 – Enthalpie, entropie, énergie libre de la réaction d’aminoacylation pour la *Bst*-TyrRS.

Complexe Enzymatique	$\Delta H^0$ kJ/mol	$\Delta S^0$ J/mol K	$\Delta G^0$ kJ/mol
TyrRS + Tyr $\rightleftharpoons$ TyrRS:Tyr 5-35 °	103 ( $\pm 2$ )	437 ( $\pm 4$ )	-27.2
TyrRS + ATP $\rightleftharpoons$ TyrRS:ATP 5-35 °	39 ( $\pm 3$ )	180 ( $\pm 10$ )	-14.0
TyrRS:Tyr + ATP $\rightleftharpoons$ TyrRS:Tyr:ATP 5-20 degré	-24 ( $\pm 2$ )	-29 ( $\pm 3$ )	-15.4 ( $\pm 0.1$ )
20-25 °	-144 ( $\pm 6$ )	-438 ( $\pm 6$ )	-13.3
25-35 °	137.9 ( $\pm 0.2$ )	507 ( $\pm 1$ )	-13.3
TyrRS:Tyr:ATP $\rightleftharpoons$ TyrRS:Tyr-ATP 5-29 °	167.6 ( $\pm 0.5$ )	347 ( $\pm 2$ )	64.7 ( $\pm 0.2$ )
29-35 °	-65 ( $\pm 1$ )	-425 ( $\pm 4$ )	61.5 ( $\pm 0.2$ )
TyrRS:Tyr-ATP $\rightleftharpoons$ TyrRS:TyrAMP:PPi 10-35 °	103.2 ( $\pm 0.9$ )	125 ( $\pm 3$ )	-65.7 ( $\pm 0.5$ )
TyrRS:TyrAMP:PPi $\rightleftharpoons$ TyrRS:TyrAMP + PPi			
10-25 °	-18.7 ( $\pm 0.4$ )	-124 ( $\pm 2$ )	18.3
25-35 °	55.8 ( $\pm 0.2$ )	126 ( $\pm 1$ )	18.3
TyrRS:TyrAMP:ARNt <sup>Tyr</sup> $\rightleftharpoons$ TyrRS:TyrAMP:ARNt <sup>Tyr</sup>			
10-35 °	-52 ( $\pm 3$ )	-50 ( $\pm 10$ )	-36.6
10-35 °	53 ( $\pm 4$ )	-40 ( $\pm 10$ )	65.1

### 1.3.3 Discrimination des ARNt

Comme nous l'avons vu avec la construction des mini-ARNt et des mini-aaRS, l'ARNt est également responsable de la spécificité de l'interaction. Les éléments d'identité de l'ARNt<sup>Tyr</sup> sont la base A76, l'anticodon et la paire de bases A1-U72. Les résidus de TyrRS qui interagissent avec l'ARNt<sup>Tyr</sup> ou qui reconnaissent ses éléments spécifiques ont été identifiés par mutagénèse extensive ou étude cinétique de *Bacillus Stearothermophilus*. Un rôle de la paire de base 1-72 dans la reconnaissance de l'ARNt<sup>Tyr</sup> est d'éviter l'aminationoacylation hétérologue : en effet, on trouve G1-C72 chez les bactéries et C1-G72 chez les archéobactéries. Ceci est corrélé à l'existence de deux groupes distincts de TyrRS. Un peptide de 39 acides aminés dans la partie N-terminale du domaine CP1 est responsable de la différence de reconnaissance [49].

Mais une aminoacylation correcte ne dépend pas seulement des éléments d'identité dans l'ARNt et de leur reconnaissance par l'aaRS correspondante. Il y a aussi une compétition entre différentes aaRS pour un ARNt donné ou entre différents ARNt pour une aaRS donnée. Plusieurs études sur la TyrRS reprennent ce concept de compétition. Lorsque *Eco*-GlnRS est surexprimé *in vivo*, elle acyle par erreur l'ARNt<sup>Tyr</sup> avec une glutamine. Cette mésaminoacylation n'a pas lieu si la concentration intracellulaire d'ARNt<sup>Gln</sup> est augmentée, ou si l'*Eco*-TyrRS est surexprimé également. Ce dernier effet peut être reproduit *in vitro* : la TyrRS est en compétition avec la GlnRS pour l'ARNt<sup>Tyr</sup> chargé avec une glutamine. Ces expériences ont montré que la compétition entre deux aaRS se produit au cours de l'étape d'aminationoacylation, et suggèrent qu'elle dépend de l'affinité relative des aaRS pour l'ARNt [50]. De la même manière, le mutant *Eco*-ARNt<sup>Tyr</sup> (G3 :U70) est aminoacylé par lalanine ou la tyrosine en proportions égales *in vitro*. Par contre, la compétition entre *Eco*-AlaRS et *Eco*-TyrRS empêche l'aminationoacylation avec lalanine *in vivo*. Mais l'identité de l'ARNt<sup>Tyr</sup> change de la tyrosine à lalanine avec l'augmentation de la concentration d'AlaRS [51].

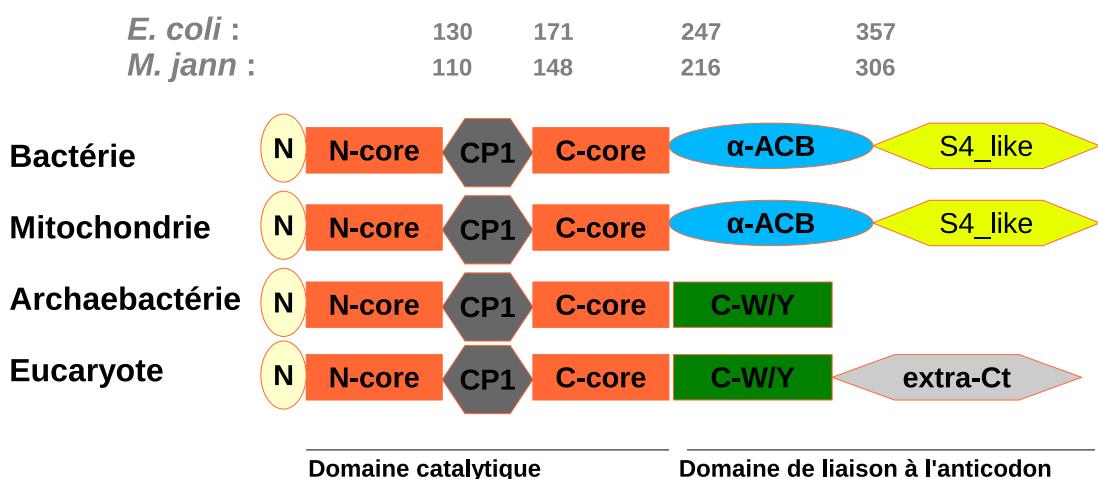
La surproduction d'*Eco*-TyrRS ou de *Bst*-TyrRS est toxique pour les organismes concernés. La cause de ce phénomène a été trouvée en faisant varier la concentration cellulaire de TyrRS et son activité de liaison à l'ARNt<sup>Tyr</sup> par voie génétique. Ces expé-

riences ont montré que la toxicité de la TyrRS surproduite est due à son affinité pour les ARNt. Plus généralement, les expériences ont montré que la quantité d'aaRS et d'ARNt dans la cellule devait respecter un équilibre pour que la cellule soit viable, car leurs concentrations influent sur la précision de traduction du code génétique [52], comme nous venons de le voir.

### 1.3.4 Différences structurales entre les TyrRS de différents règnes

Les structures de TyrRS présentent à la fois des similarités et des différences entre les règnes du vivant [53]. Toutes les TyrRS sont des homodimères avec un site actif compris dans le domaine N terminal. Ce site actif est composé d'un Rossman fold, typique des synthétase de classe I. Le domaine catalytique de la TyrRS contient environ 230 acides aminés et a une structure très conservée. Il contient un domaine d'insertion CP1 d'une cinquantaine d'acides aminés. Contrairement à l'extrémité N terminale, l'extrémité C terminale varie énormément d'un règne à l'autre, en séquence, taille et organisation. Chez toutes les TyrRS eubactériennes, elle est formée d'un domaine hélicoïdal d'une centaine d'acides aminés, suivi d'un domaine qui a une grande homologie avec le domaine C terminal de la protéine ribosomale S4, connue pour lier les ARN. La variabilité de ce domaine C terminale conduit à classer les TyrRS en deux sous groupes. Les TyrRS d'eubactérie ont ce domaine structuré en hélices  $\alpha$ . Les TyrRS d'archaeobactérie ont la particularité d'avoir un domaine C terminal très court et qui a une forte homologie avec la partie C terminal des TrpRS. Ce domaine, appelé C-W/Y, est présent chez les eucaryotes mais est suivi d'une extension de cent soixante acides aminés chez les eucaryotes supérieurs. Dans le cas de la TyrRS humaine, cette extension a 51% d'homologie avec la cytokine humaine EMAP II [54] (figure 1.15). Tsunoda et coll. en 2007 ont résolu la structure cristallographique de *Sce*-TyrRS (*Saccharomyces cerevisiae*-TyrRS) liée à un analogue du TyrAMP et à l'ARNt<sup>Tyr</sup> [55]. A ce moment, les structures de TyrRS des trois domaines du vivant étaient alors résolues, ce qui leur a permis de faire l'analyse suivante. L'aminoacylation spécifique de l'ARNt par la TyrRS repose sur les éléments d'identité de l'ARNt<sup>Tyr</sup>. Les TyrRS eucaryotes et archéabactériennes ne peuvent pas réagir avec un ARNt<sup>Tyr</sup> bacté-

FIGURE 1.15 – Organisation structurale de la TyrRS dans les trois domaines du vivant et dans la mitochondrie. Chaque domaine structural est représenté par une couleur. Les abréviations utilisées sont les suivantes : (N) : N-terminal motif. (N-core) et (C-core) : parties N et C du domaine catalytique. (CP1) : peptide de connection ; ( $\alpha$ -ACB) : domaine  $\alpha$ -hélice de liaison à l'anticodon. (C-W/Y) : domaine C-terminal homologue de la TrpRS. (S4-like) : domaine à forte homologie au domaine S4. (extra-Ct) : domaine C-terminal additionnel [49]. Les acides aminés délimitant ces différents domaines sont donnés pour *E. coli* et *M. jannaschii*.



rien. Les déterminants d'identité de la reconnaissance sont donc supposés semblables pour les TyrRS eucaryotes et archeabactériennes, mais différents pour la TyrRS bactérienne. La reconnaissance de la paire de bases C1-G72 par *Sce*-TyrRS est similaire à celle des archeabactéries, alors que la reconnaissance de la base A73 est la même que celle des TyrRS bactériennes. Donc, l'impossibilité d'aminoacylation croisée entre TyrRS et ARNt<sup>Tyr</sup> archeabactériens/eucaryotes et bactériens, doit venir des différences entre les paires de bases de l'ARNt<sup>Tyr</sup> (C1-G72 et G1-C72, respectivement) [55] (figure 1.16, figure 1.17).

FIGURE 1.16 – Alignement basé sur la structure. TyrRS d'Eucaryote (Levure : 2DLC, Humain : 1N3L), TyrRS d'archeabactérie (*M. Jannaschii* : 1J1U), TyrRS de bactérie (*E. coli* : 1X8X, *T. Thermophilus* : 1H3E). En rouge : résidus reconnaissant le TyrAMP. En vert : résidus reconnaissant la tige accepitrice de l'ARNt. En bleu : résidus reconnaissant l'anticodon de l'ARNt. Les motifs signatures de la classe I des aaRS sont en jaune. [55].

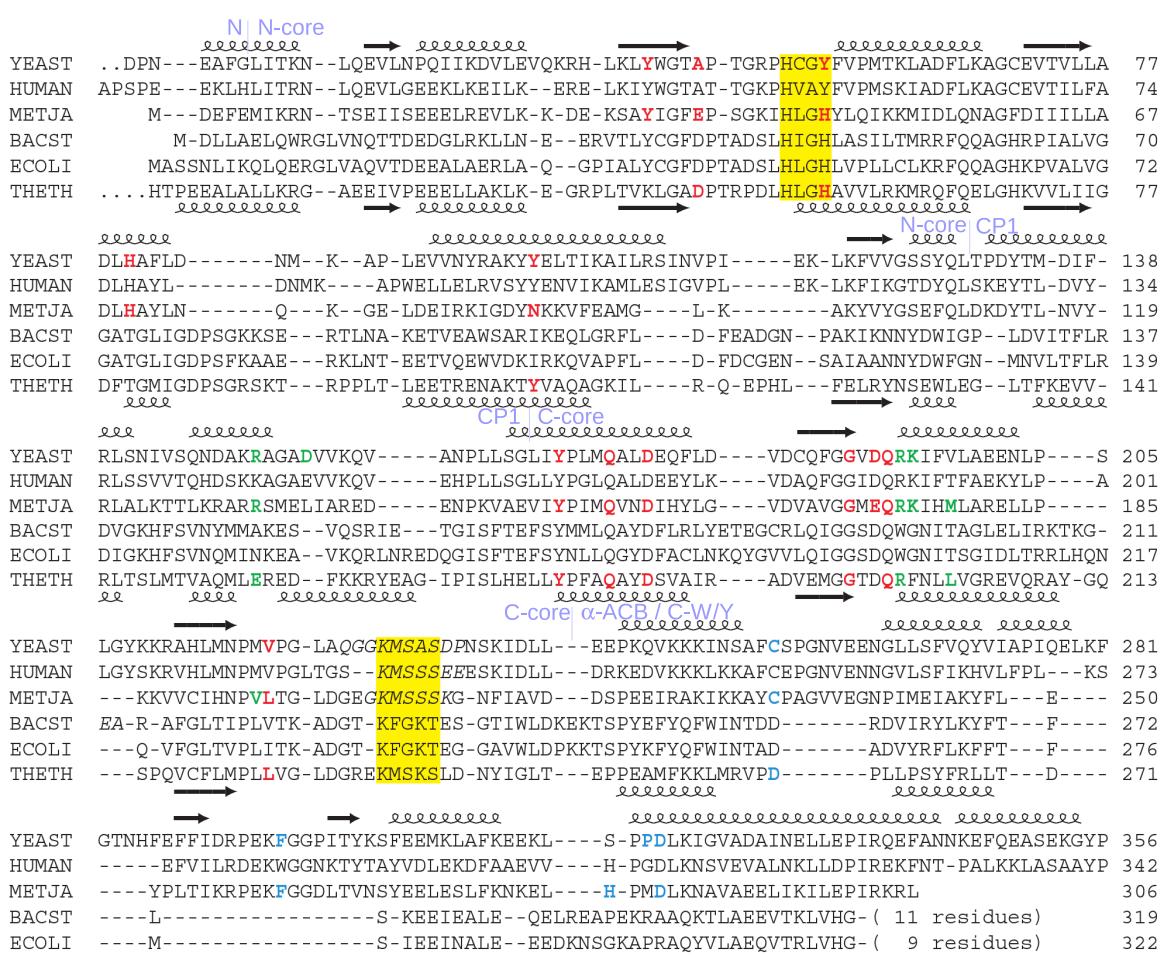
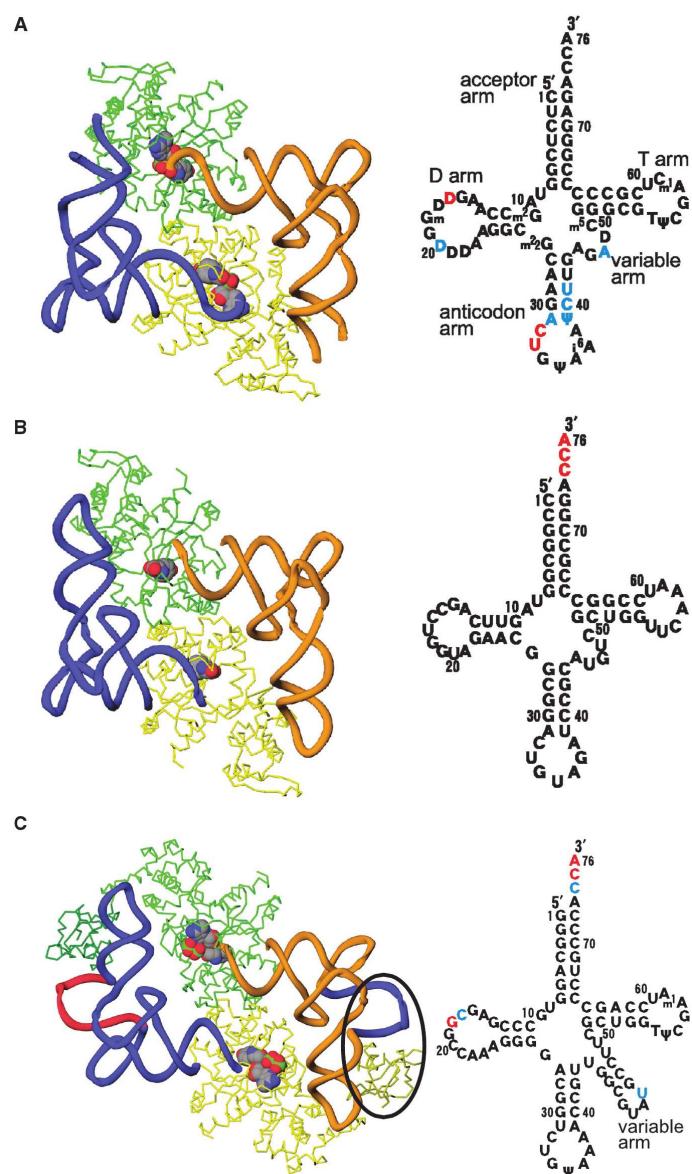


FIGURE 1.17 – (A) TyrRS eukaryote : *Sce*-TyrRS. (B) TyrRS d’archée (*Mja*-TyrRS). (C) TyrRS bacterienne (*Tth*-TyrRS). Un bras variable de l’ARNt et un domaine C-terminal de TyrRS présent seulement chez les bactéries sont entourés. La TyrRS est en jaune et vert, l’ARNt<sup>Tyr</sup> est bleu et orange.[55].



### 1.3.5 Spécificité de la TyrRS

La TyrRS est une enzyme très spécifique, bien qu'elle ne possède pas de site d'édition. Par exemple, elle n'a pas de difficulté à faire la différence entre la tyrosine et la phénylalanine : le groupement hydroxyle permet une fixation  $10^4$  fois plus forte qu'avec la phénylalanine [56]. Chez la levure, la TyrRS est l'aaRS qui a le plus grand facteur de discrimination entre acide aminé correspondant à l'ARNt<sup>Tyr</sup> ou non. Le  $K_m$  pour l'ARNt<sup>Tyr</sup> est de 0.00035 mM [57].

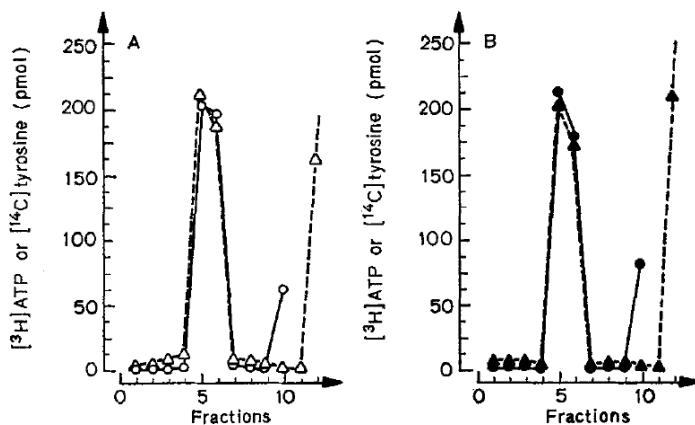
Le mécanisme par lequel *Bst*-TyrRS reconnaît la tyrosine a été étudié par mutagénèse des résidus proches du site actif. Le carboxylate de l'Asp176 fait une liaison hydrogène avec l'hydroxyle de la tyrosine. La mutation de cette position inactive l'enzyme. L'Asn123 et le Trp126 n'interagissent pas directement avec la tyrosine mais forment des liaisons hydrogène avec l'Asp176. Les mutations N123A et N123D affectent fortement la cinétique de l'activation de la tyrosine, et en particulier le  $k_{cat}$  et le  $K_m$  (Tyr). Au contraire, W126F, et W126L affectent à moindre échelle la cinétique. La spécificité pour la tyrosine par rapport à la phénylalanine, déterminée par le ratio  $\frac{k_{cat}}{K_m}$  pour la réaction de l'échange ATP/pyrophosphate, décroît d'un facteur quatre avec la mutation N123A, double avec la mutation W126L, et augmente de sept fois pour la mutation W126F. Donc la spécificité de l'enzyme natif peut être améliorée [58].

La stéréospécificité par rapport au backbone de l'acide aminé est moins forte. Ainsi, les TyrRS d'*E. Coli*, de *B. Subtilis* et de *S. Cerevisiae* peuvent charger l'ARNt<sup>Tyr</sup> avec la D-Tyrosine. Mais le résultat est hydrolysé par la D-Tyr-déacylase, une enzyme qui a été identifiée et caractérisée chez *E. coli* et *S. cerevisiae*. Dans ces deux organismes, la spécificité de la déacylase n'est pas restreinte à l'ARNt<sup>Tyr</sup>; elle peut agir sur d'autres ARNt également. En l'absence de déacylases, certains acides aminés D sont toxiques pour les organismes. Aussi, même si la TyrRS ne possède pas de mécanisme de relecture, les déacylases en apportent un pour les acides aminés D.

### 1.3.6 Le rôle du magnésium

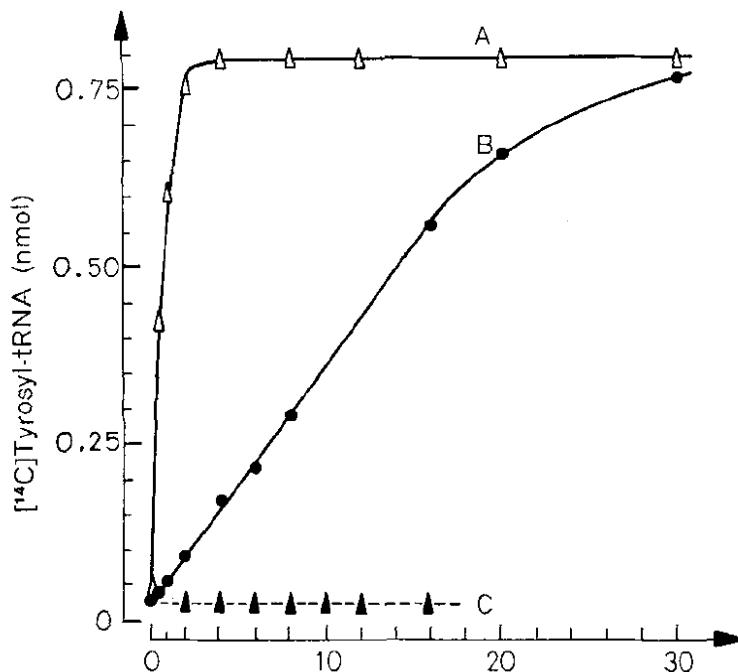
L'ion  $Mg^{2+}$  joue un rôle important dans les réactions catalysées par les aaRS. Il se lie dans le site actif et participe à la catalyse de la réaction d'activation. Sa présence est également essentielle pour l'aminocylation de l'ARNt. Pour étudier l'effet du  $Mg^{2+}$ , Chousterman et coll. ont mené l'expérience suivante [59]. De la TyrRS d'*E. coli* et de l'ARNt sont dialysés avec de l'EDTA. La TyrRS fixe en moyenne 0.8 ions de  $Mg^{2+}$  par molécule pour l'activation de la tyrosine. L'ARNt fixe en moyenne 0.25 ions de  $Mg^{2+}$  par molécule pour la formation de l'ARNt<sup>Tyr</sup>. L'activation de la tyrosine et la formation d'ARNt<sup>Tyr</sup> sont ainsi étudiées en l'absence d'ions  $Mg^{2+}$ . Le complexe TyrRS:TyrAMP se forme toujours, mais la vitesse de la réaction est réduite d'un facteur  $10^4$ . La vitesse de la réaction d'échange ATP-PPi sous ces conditions est réduite d'un facteur  $4 \times 10^4$  (figure 1.18, figure 1.19).

FIGURE 1.18 – Isolation expérimentale du tyrosyl adenylate. Complexe enzymatique préparé en présence (A) et en absence (B) de  $Mg^{2+}$ . L'enzyme a été incubé en présence de [<sup>3</sup>H]ATP et de [<sup>14</sup>C]tyrosine. Le tout a été filtré sur gel et l'éluat fait l'objet de mesures de radioactivité. (A) en absence de  $Mg^{2+}$ , (B) en présence de  $Mg^{2+}$ . D'après Chousterman 1973 [59].



La structure de TyrRS d'*E. coli* a été résolue sans magnésium (code PDB : 1VBM). La structure de *M. jannaschii* résolue avec la L-Tyr en contient à plusieurs endroits mais pas dans le site catalytique (code PDB 1J1U). La TyrRS de *S. cerevisiae* a été résolue en présence de TyrAMP et de magnésium (code PDB 2DLC). On y voit la localisation du magnésium dans le site actif, proche du phosphate du TyrAMP.

FIGURE 1.19 – Formation d'ARNt<sup>Tyr</sup> en présence et en absence d'ions Mg<sup>2+</sup>. La réaction s'est produite en présence de 0.76 mg d'ARNt. (A) : en présence d'ions Mg<sup>2+</sup>, (B) : en absence d'ions Mg<sup>2+</sup> et en présence de 2mM de spermidine. (C) : en absence d'ions Mg<sup>2+</sup>. Triangle pour [<sup>14</sup>C] tyrosine et rond pour [<sup>3</sup>H]ATP. D'après Chousterman 1973 [59].



### 1.3.7 Inhibiteurs

Comme la TyrRS est une protéine essentielle, des inhibiteurs peuvent être utilisés comme antibiotiques s'ils sont spécifiques de la TyrRS d'un organisme pathogène. Des analogues stables du TyrAMP peuvent être des inhibiteurs potentiels de la TyrRS, mais leur polarité est cependant un frein au passage à travers la paroi bactérienne. Plusieurs colorants, dont la triazine ( $C_3H_3N_3$ ), inactivent la *Bst*-TyrRS irréversiblement, en occupant le site de liaison à l'ATP. Ces inhibiteurs ne sont pas spécifiques à la TyrRS et peuvent inactiver également la MetRS ou la TrpRS. Un autre inhibiteur, le tyrosyl-aryl dipeptide (Tyr-Tyr) peut également occuper la poche catalytique de la TyrRS. Des peptides qui lient la TyrRS d'*Haemophilus influenzae* ont été isolés à partir de librairies aléatoires exposées à la surface du phage M13. La plupart de ces inhibiteurs sont spécifiques et se lient dans le site actif [60].

## 1.4 Evolution dirigée de la TyrRS

L'évolution dirigée peut-être utilisée pour modifier la spécificité d'une amino acyl-ARNt synthétase, afin qu'elle lie préférentiellement un acide aminé non naturel. Ceci va permettre d'activer cet acide aminé non-naturel, puis de l'acyler avec un ARNt qui lui est spécifique. L'acide aminé non-naturel, pourra ensuite être incorporé dans une protéine en cours de synthèse, de manière contrôlée. Modifier les aminoacyl-ARNt synthétases permet ainsi à la machinerie de traduction de fabriquer de nouvelles protéines, contenant un ou plusieurs acides aminés non naturels. Ces protéines peuvent avoir une structure nouvelle, ou une fonction originale, ou encore être simplement plus efficaces que la protéine native correspondante. L'incorporation d'acides aminés non-naturels dans les protéines peut-être réalisée *in vitro* ou *in vivo*. Mais l'introduction d'acides aminés non-naturels *in vivo* requiert de nombreux composants, pour que la traduction des ARNm en protéines fonctionne.

### 1.4.1 Charger des acides aminés non naturels.

En général, une extension du code génétique doit répondre à trois conditions :

1. Une aaRS n°21 doit être introduite dans l'organisme pour charger spécifiquement un ARNt, tout en excluant les 20 ARNt endogènes.
2. Un ARNt n°21 doit être introduit dans l'organisme. Il est spécifiquement chargé par l'aaRS n°21 ; il ne réagit pas avec les 20 aaRS endogènes, et il possède un codon différent des autres codons existants.
3. L'aaRS n°21 doit reconnaître spécifiquement l'acide aminé non-naturel, l'activer et le transferer sur l'ARNt n°21, tout en excluant les 20 acides aminés naturels.

RajBhandary et coll. ont construit deux couples d'aaRS/ARNt qui satisfont les deux premières conditions. L'un d'eux est basé sur *Sce(cyt)*-TyrRS, la TyrRS cytoplasmique de levure. L'ARNt porte un anticodon stop AMBER (CUA), ainsi que les éléments d'identité suivants : Cyt1-Gua72 et Ade73, comme chez les ARNt<sup>Tyr</sup> eucaryotes. Mais la surexpression de cette aaRS ne peut pas être maintenue de façon stable dans *E. coli* [61].

Schultz et coll. ont construit un couple aaRS/ARNt chez *E. coli* qui satisfait les trois

conditions . Pour obtenir un tel couple, ils importent dans *E. coli* un couple aaRS/ARNt d'un autre organisme. Un candidat est le couple ARNt<sup>Tyr</sup>/TyrRS de *Methanococcus jannaschii*, une archébactéries dont les éléments d'identité diffèrent de ceux de l'ARNt<sup>Tyr</sup> d'*E. coli* (en particulier, la première paire de base de la tige acceptrice est GC chez *E. coli* et CG chez *Methanococcus jannaschii*). De plus, la TyrRS de *Methanococcus jannaschii* n'a pas de domaine d'édition et a un domaine de liaison à l'anticodon minimalist. Elle n'éliminera donc pas l'acide aminé non naturel lié à l'ARNt.

Ils cherchent dans un premier temps à obtenir un nouvel ARNt qui n'est reconnu par aucune aaRS d'*E. coli*, mais qui fonctionne efficacement pour la translation. Ce nouvel ARNt devra délivrer l'acide aminé non naturel en réponse à un codon non sens, qui ne code pour aucun des 20 acides aminés. Leur point de départ est *Mja*-ARNt<sup>Tyr</sup><sub>CUA</sub>, l'ARNt<sup>Tyr</sup> auquel on a incorporé un anticodon nonsens. La reconnaissance du *Mja*-ARNt<sup>Tyr</sup><sub>CUA</sub> par les aaRS d'*E. coli* est très faible. La TyrRS de *Methanococcus jannaschii* aminoacyle correctement l'ARNt suppresseur Amber dérivé de l'ARNt<sup>Tyr</sup>, mais aucun ARNt d'*E. coli*. De plus, l'ARNt<sup>Tyr</sup><sub>CUA</sub>, mauvais substrat des aaRS d'*E. coli*, fonctionne efficacement pour la translation des ARNm en protéine dans le cytoplasme d'*E. coli*. De plus, une nouvelle aaRS qui aminoacyle l'ARNt modifié est nécessaire. Elle ne devra aminoacyler aucun des ARNt endogènes d'*E. coli*; elle doit reconnaître l'ARNt n°21 avec l'acide aminé non-naturel seulement. Enfin, l'acide aminé non naturel ne devra pas être un substrat des synthétases endogènes. Ils ont donc généré la nouvelle aaRS et le nouveau ARNt par une méthode d'évolution dirigée, schématisée dans la figure 1.20

Pour que la traduction puisse avoir lieu avec le nouvel ARNt, il convient également de muter le ribosome. Wang et coll. ont réalisé l'évolution dirigée du ribosome afin qu'il reconnaisse les anticodons stop Amber qui correspondent à l'ARNt muté [63]. Le ribosome mutant obtenu est tel qu'il affaiblit la liaison au facteur 1, responsable de la terminaison de la chaîne polypeptidique (figure 1.22).

Plusieurs aaRS ont déjà été modifiées pour qu'elles acceptent des aa non-naturels comme substrats. Oue et coll., en 1999, ont mis en oeuvre l'évolution dirigée de l'AspRS pour fixer un substrat non natif, la valine [64]. Ils identifièrent un mutant sur 17 positions

FIGURE 1.20 – Schéma de sélection positive et négative pour la création de variants de TyrRS spécifiques d'un acide aminé non naturel chez *E. coli*. 1 : Les positions à muter sont déterminées par observation de la structure cristallographique. Les mutants sont ensuite créés par PCR mutagène, à l'aide d'amorces spécifiques.

2 : Sélection Positive : Résistance au chloramphenicol conférée par la présence de *Mja*-TyrRS et de l'acide aminé non naturel ou de tout autre acide aminé qui peut être chargé sur *Mja*-ARNt<sub>CUA</sub><sup>Tyr</sup> par un variant de *Mja*-TyrRS. Pour sélectionner les paires TyrRS/ARNt fonctionnelle, la sélection positive est basée sur la capacité des variants *Mja*-ARNt<sub>CUA</sub><sup>Tyr</sup> à supprimer une mutation amber dans le gène de la chloramphenicol acéthyltransféras.

3 : Sélection négative : Gène de Barnase modifié avec mutations Amber. La sélection est basée sur la suppression de mutations non-sens dans le gène de la barnase. La barnase est ici cytotoxique. Cette étape est réalisée en l'absence d'acide aminé non naturel.

Seuls les variants qui peuvent acyler *Mja*-ARNt<sub>CUA</sub><sup>Tyr</sup> avec l'acide aminé non naturel peuvent survivre au deux sélections [62].

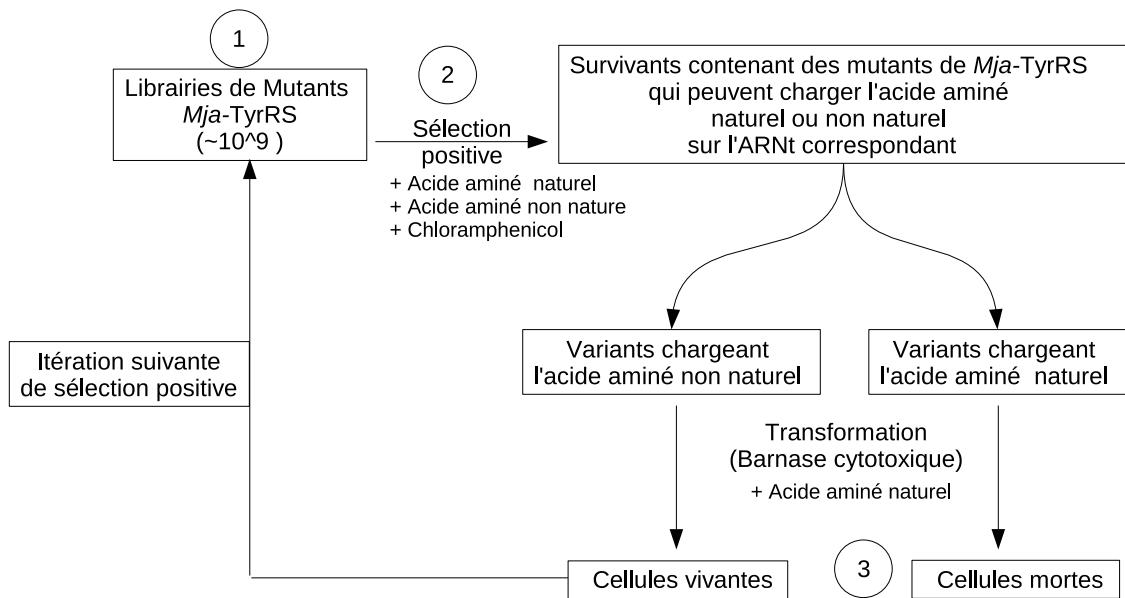


FIGURE 1.21 – Une méthode expérimentale pour incorporer un acide aminé non naturel dans un polypeptide, est basée sur une nouvelle aaRS et un nouvel ARNt, qui vont s'ajouter aux aaRS et ARNt naturels. L'ARNt est modifié pour que son anticodon soit complémentaire d'un codon stop et pour qu'il interagisse favorablement avec la nouvelle aaRS. On mute l'aaRS correspondante pour qu'elle reconnaisse l'ARNt ainsi modifié et l'acide aminé non naturel. Un ou plusieurs codons stop sont introduits dans le gène de la protéine qu'on souhaite modifier. L'acide aminé non naturel sera ainsi incorporé dans la protéine au cours de sa biosynthèse.

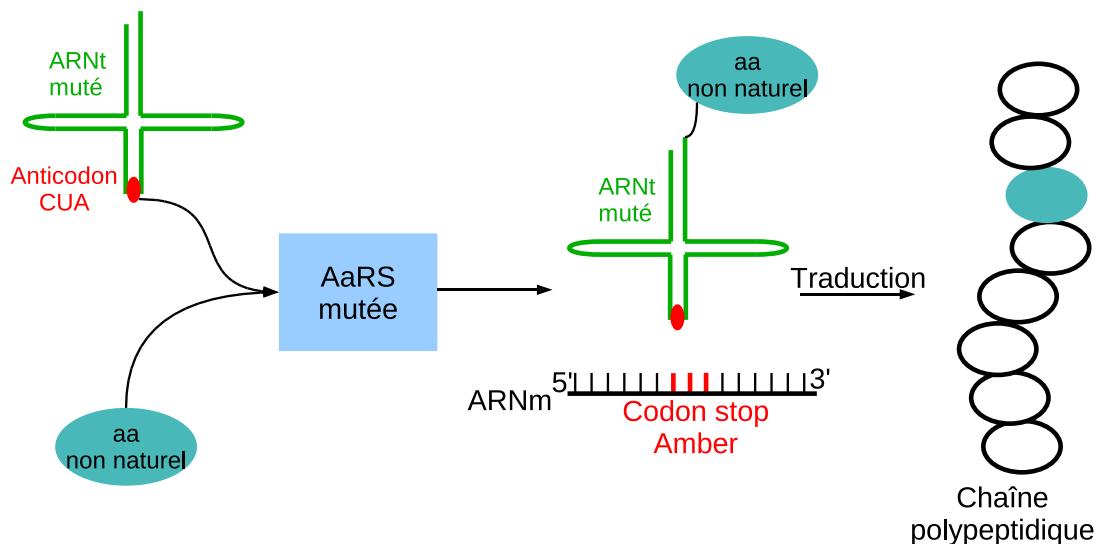
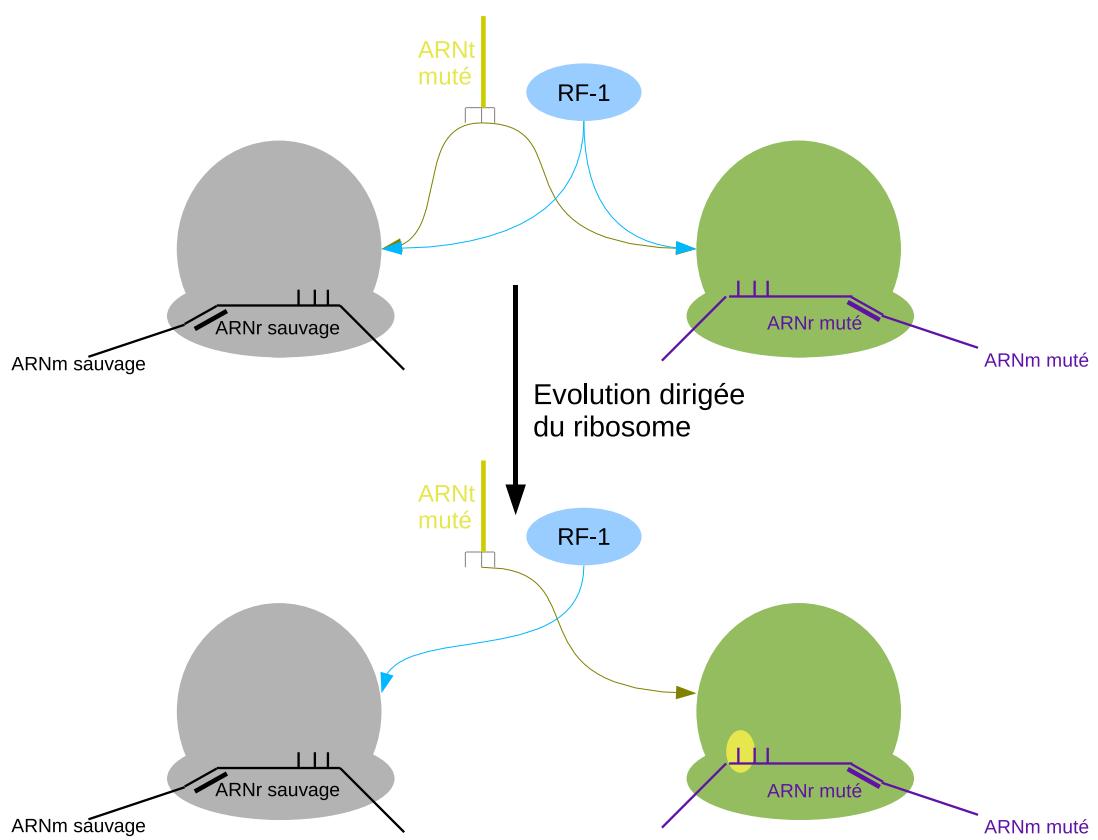


FIGURE 1.22 – Le ribosome naturel (gris) et le ribosome modifié (vert) décodent l'ARNm sauvage (noir) et muté(violet) respectivement. RF-1 (bleue) est en compétition avec l'ARNt amber muté pour le codon UAG sur le site A des deux ribosomes ; l'ARNt va être uniquement aminoacylé par l'acide aminé non naturel.



qui accroît l'efficacité de la catalyse d'un facteur  $10^6$ . Des MetRS mutantes ont également été fabriquées. Enfin, l'aaRS qui a été utilisée le plus souvent dans ce contexte est la TyrRS.

### 1.4.2 Acides aminés non naturels déjà incorporés dans des protéines

Un grand nombre d'acides aminés non naturels ont pu être ponctuellement ajoutés au code génétique d'*E. coli*, de *S. cerevisiae*, de *M. jannaschii* et de mammifères. Ces acides aminés non naturels ont des propriétés chimiques nouvelles, qui peuvent être utiles pour l'étude de la structure et de la fonction des protéines [65].

Ainsi, un analogue de la tyrosine, l'azatyrosine, L-B-(5-hydroxy-2-pyridyl)-alanine peut modifier le phénotype d'une cellule. Une étude a été réalisée avec des mutants d'*Eco-TyrRS* capables de charger l'azatyrosine, trouvés après génération d'une bibliothèque de 1400 mutants, par PCR mutagène. Une TyrRS portant la mutation F130S lie 17 fois mieux l'azatyrosine que la tyrosine native. Le résidu Phe130 interagit avec Asp182, qui fait une liaison hydrogène avec l'hydroxyle du TyrAMP. L'efficacité de l'enzyme peut être évaluée par le ratio  $k_{cat}/K_m$ . Une grande valeur de ce ratio traduit une rapide transformation du substrat en produit, ou une faible valeur de  $K_m$  (grande affinité de l'enzyme pour le substrat). La spécificité pour l'azatyrosine peut être évaluée en calculant un facteur de spécificité pour le chargement de l'ARNt défini par :

$$\frac{(k_{cat}/K_m)_{Tyrozine}}{(k_{cat}/K_m)_{Azatyrozine}} \quad (1.4)$$

Ce facteur de spécificité est de 36 avec la TyrRS native et de 19 avec le mutant F130S. Le mutant discrimine donc moins l'azatyrosine que la native [66]. Les équivalents des résidus Phe130 et Asp182 de *Eco-TyrRS* sont, chez *Bst-TyrRS*, Ile127 et Asp176.

Avec une mutagénèse spécifique du site actif, le mutant *Sce(cyt)-TyrRS* est capable de charger un acide aminé autre que la tyrosine. Ce mutant, peut utiliser trois analogues différents de la tyrosine pour l'aminoacylation, dont la 3-iodo-L-Tyr. L'aminoacylation

avec la tyrosine est alors 400 fois moins efficace que pour le type sauvage. Cette TyrRS mutante peut servir à l'incorporation site-spécifique de nouveaux acides aminés dans les protéines. La structure cristallographique du mutant 37V195C de la TyrRS d'*E. coli*, en interaction avec la 3-iodo-L-tyrosine, a été résolue par Kobayashi et coll. [67]. Cette structure montre que l'iode du ligand est sélectivement reconnu par la valine 37 et la cystéine 195, par des interactions de van der Waals. Ils vérifient également que l'interaction avec l'analogue du tyrosyl-adénylate est conservée.

#### 1.4.2.1 Nouvelles fonctionnalités offertes par l'extension du code génétique

**Les réactifs chimiques** L'ingénierie des aaRS a permis d'incorporer plusieurs acides aminés ayant de nouvelles fonctions chimiques dans des protéines d'*E. coli* et de *S cerevisiae*. Ces acides aminés contenaient des groupes cétones, azides, acetylenes ou thioesters. Les protéines ainsi modifiées peuvent ensuite réagir avec des tags fluorescents et des réactifs exogènes. Par exemple, une série de colorants fluorescents a été introduit sélectivement en un site unique du mutant m-acétyl-Phe de la protéine membranaire LamB de *E. coli*. Un autre exemple concerne une hormone de croissance humaine, modifiée site-spécifiquement avec du polyethylene glycol (PEG), de manière à créer une protéine qui conserve l'activité naturelle mais avec une demi-vie bien plus importante dans le serum. On peut ainsi créer des protéines avec un grand contrôle de leur structure et de leur homogénéité. Ceci peut également être utile pour la synthèse ribosomale de peptides et protéines cycliques. Par exemple, si une cétone est introduite à l'extrémité C terminale d'une protéine, elle peut réagir avec le groupe amine N terminal et cycliser la protéine.

mutant homodimérique de la glutathion S-transférase, avec du p-benzoylphenylalanine incorporé spécifiquement à l'interface du dimère, peut être crosslinké dans le cytoplasme d'*E. coli*. Ce résidu a été récemment incorporé dans le pore central de ClpB, une ATPase qui empêche les protéines de s'aggréger. Le mutant se lie par "photocross-linking" au peptide substrat, ce qui donne des preuves physiques directes du contact entre le pore de ClpB et le substrat. Des cystéines, séries et tyrosines photo-actives ont également été introduites spécifiquement dans les protéines. Le groupement hydroxyl ou thiol de la chaîne latérale de ces acides aminés est bloqué par substitution d'un groupe nitrobenzyl, qui peut être clivé par irradiation à 365 nm *in vitro* ou *in vivo*. Par exemple, le site actif de la cystéine caspase-3, qui est pro-apoptotique, a été substitué avec du nitrobenzyl pour créer une protéine inactive et photo-activable.

**Les sondes biophysiques** Des acides aminés non naturels ont été encodé génétiquement dans *E. coli* et *S. cerevisiae* pour étudier le repliement ou la fonction de protéines. C'est le cas d'acides aminés fluorescents comme la 7-hydroxy-coumarine. Cet acide aminé permet une visualisation directe, par fluorescence, des changements conformationnels de la protéine, de sa localisation, et des interactions intermoléculaires. L'avantage de cette méthode par rapport à la GFP par exemple, est la petite taille du groupement fluorescent. De plus, la GFP est liée aux extrémités N terminales ou C terminales, alors que les acides aminés fluorescents peuvent être introduits partout dans la protéine ou presque. C'est ainsi que des dérivés de la coumarine ont été utilisés pour prouver le dépliement local de la myoglobine. La p-iodo-phenylalanine avec son atome lourd peut servir au phasage des données de diffraction de rayon X. La p-cyanophenylalanine et la m-cyanophenylalanine ont également été incorporées dans les protéines ; comme leur groupement nitrile a une fréquence vibrationnelle différente des autres groupements chimiques présents dans la protéine, ils peuvent être utilisés en spectroscopie infrarouge pour témoigner de l'environnement local ou de la dynamique de la protéine.

**Les acides aminés à activité redox** Les transferts d'électrons sont impliqués dans un grand nombre d'activités biologiques, qui vont de l'activité enzymatique à la pho-

tosynthèse. Deux acides aminés à activités redox ont été incorporés sélectivement dans les protéines : la dihydroxyphenylalanine et la 3-amino-L-tyrosine ; ils peuvent être utilisés pour détecter ou manipuler les transferts d'électrons dans les protéines. La 3-amino L-tyrosine peut éventuellement fonctionner comme une trappe à électrons.

**Les acides aminé chelateurs de métaux** La création de protéines avec des sites de liaison au métal reste un défi, compte-tenu de la difficulté à prédir et contrôler les couches successives d'acides aminés autour de l'ion métallique. Dans cette optique, des chercheurs tentent d'incorporer la bipyridyl-alanine dans les protéines.

**Autres fonctions** Dans d'autres expériences, un analogue de la cystéine à longue chaîne a été incorporé dans des protéines chez *S. cerevisiae* dans le but de stabiliser des conformations particulières, d'étudier la dynamique du repliement des protéines ou d'obtenir des protéines thermostables. D'autres acides aminés non naturels qui ont pu être ajoutés au code génétique de *E. coli* et *S. cerevisiae* sont la  $\beta$ -(2-naphtyl)alanine, l'acide  $\alpha$ -aminocaprylique, la L-homoglutamine et la p-aminophenylalanine. Le grand nombre et la diversité des acides aminés qui ont été génétiquement encodés témoigne d'une grande plasticité de la machinerie traductionnelle concernant les chaînes latérales des acides aminés (figure 1.23).

## 1.5 Les acides aminés D

### 1.5.1 Présence d'acides aminés D dans les organismes

Cette partie présentant les acides aminés D et la suivante concernant les peptides d'origine animale sont un résumé d'un extrait du travail de thèse de Julie Soutourina intitulé : *Les acides aminés de la série D dans la traduction* [68]. Les protéines sont synthétisées essentiellement à partir d'acides aminés L. Cependant, les acides aminés D sont naturellement présents dans les organismes. Les acides aminés D peuvent être sous forme libre ou entrer dans la composition d'autres molécules. Par exemple, les peptidoglycans des parois bactériennes contiennent généralement de la D-alanine et du D-glutamate. Ces acides aminés contribuent à la résistance des structures bactériennes. Chez la levure, la DL-dityrosine a été identifiée dans les parois des ascospores [69]. Les structures contenant de la DL-tyrosine sont spécifiques de la dernière phase de la sporulation. Elles assurent la résistance des ascospores aux enzymes lytiques.

Pour maintenir l'homochiralité des protéines, des mécanismes excluant les acides aminés D de la biosynthèse protéiques ont été sélectionnés au cours de l'évolution. Néanmoins, divers organismes produisent aussi des peptides biologiquement actifs contenant des acides aminés D [70, 71, 72]. Ils sont synthétisés soit par l'assemblage direct d'acides aminés libres, réalisé par des complexes multi-enzymatiques indépendants des ribosomes, soit à partir d'un précurseur formé d'acides aminés L, qui est ensuite modifié post-traductionnellement. Le premier mécanisme est spécifique des bactéries, tandis que le deuxième se rencontre aussi bien chez les procaryotes que chez les eucaryotes.

Parmi les peptides qui contiennent naturellement des acides aminés D, on trouve des antibiotiques, des hormones, des neuropeptides, des peptides opiacés et des hépatotoxines. Les premiers peptides contenant des acides aminés D et possédant une activité antimicrobienne ont été identifiés dans les bactéries. La structure et la synthèse des peptides antibiotiques ont été très étudiées chez les bactéries du genre *Bacillus* [73, 74]. Les antibiotiques cycliques, tels que la gramicidine S et la polymyxine B, contiennent de la D-phénylalanine ; la bacitracine contient de la D-Phe, du D-Glu, de la D-Orn et du D-

Asp. Les tyrocidines contiennent de la D-Phe et du D-Trp. La bacillomicine D contient de la D-Tyr, de la D-Asn et de la D-Ser ; la bacillomicine F contient de la D-Tyr et de la D-Asn. Tous ces peptides sont synthétisés indépendamment des ribosomes. Ils agissent en formant des canaux dans les membranes cellulaires ou en inhibant la synthèse des parois.

Parmi les peptides antibiotiques dont les résidus de la série D sont synthétisés post-traductionnellement, on trouve les lantibiotiques produits par certaines bactéries Gram positives, qui contiennent de la D-alanine et de la S-(2-aminovinyl-D-cystéine) [70].

### 1.5.2 Des peptides d'origine animale contenant des acides aminés D

Des peptides contenant des acides aminés D ont été isolés à partir des sécrétions de peau d'amphibiens et de nombreux invertébrés [75, 76]. Ces peptides ont une activité opiacée ou antimicrobienne. Les sécrétions de la peau des grenouilles du genre *Phyllomedusiae* contiennent de nombreux peptides agissant sur le système nerveux central des vertébrés. Beaucoup d'entre eux sont homologues à des hormones ou à des neurotransmetteurs de mammifères. Les dermophines sont composées de sept acides aminés. Elles contiennent de la D-Ala en deuxième position, indispensable à leur activité biologique. Ces peptides se fixent avec une grande affinité et une grande sélectivité sur les récepteurs opiacés de type  $\mu$ , provoquant une analgésie profonde et prolongée. Des peptides proches, nommés deltorphines, sont spécifiques des récepteurs opiacés de type  $\delta$  qui contiennent de la D-Ala, de la D-Met ou de la D-Leu en deuxième position.

La deuxième famille de peptides animaux contenant des acides aminés D a été découverte dans les sécrétions de la peau de la grenouille européenne, *Bombina variegata* [75, 77, 76]. Ces peptides de 20, 24 ou 27 acides aminés constituent la famille des bombinines. Ils ont une activité antimicrobienne. Ils contiennent de la D-allo-isoleucine ou de la D-leucine, en deuxième position. Des bombinines composées uniquement d'acides aminés L sont également présentes, en petites quantités, dans les sécrétions. Elles possèdent également une activité antibactérienne, mais la vitesse à laquelle elles agissent sur la croissance bactérienne est plus faible.

Des peptides contenant des acides aminés D ont été découverts chez les invertébrés. Deux neuropeptides ont ainsi été isolés chez l'escargot géant africain, *Achatina fulica* [78, 79, 75]. Ils contrôlent ou stimulent la contraction musculaire. Un tri-peptide cardio-excitateur (Asn D-Trp Phe-NH<sub>2</sub>) a été purifié à partir de l'aplysie *Aplysia kurodai* [80]. Le peptide synthétique contenant L-Trp à la place du D-Trp est environ mille fois moins actif que le peptide natif. Les escargots marins du genre *Cornus*, tueurs des poissons, contiennent des toxines peptidiques, nommées contryphanes, qui bloquent la transmission neuromusculaire et immobilisent leurs proies [81, 76].

Divers crustacés, comme les langoustes et le homard contiennent deux isoformes de l'hormone hyperglycémiant des crustacés (CHH) [82]. Ces isoformes peptidiques, composées toutes les deux de 72 résidus, diffèrent uniquement par la conformation L ou D de la phénylalanine en troisième position. L'activité biologique de cette hormone consiste à augmenter la concentration lymphatique du glucose, par exemple en réponse à un stress. Chez la langouste *Procambarus clarkii*, la stéréo-inversion de la phénylalanine en position 3 dans un des deux variants de l'hormone CHH semble importante pour la régulation de la production d'ecdystéroïdes [83]. La présence de deux isoformes de l'hormone hyperglycémante de crustacés est encore inexpliquée. Il est possible que les deux formes D et L soient nécessaires pour assurer une action différentielle de l'hormone dans certains tissus ou organes spécifiques. Le venin de l'araignée *Agelenopsis aperta* contient deux toxines capables de bloquer les canaux calcium [84]. Ces peptides, nommés  $\omega$ -agatoxines IVB et IVC, contiennent chacun 48 acides aminés. Ils se distinguent uniquement par la présence de L ou D-sérine en position 46.

La biosynthèse de la lombricine phosphate par les vers de terre *Lombricus terrestis* et *Megascolides cameroni* se fait à partir de la D-sérine libre [85]. La D-sérine réagit avec la CDP-éthanolamine, dans une réaction enzymatique spécifique de l'acide aminé D, pour former de la D-sérine éthanolamine phosphate. Ce composé réagit alors avec de l'arginine pour former la lombricine, qui est finalement transformée en lombricine phosphate par l'ATP-lombricine phosphotransphérase.

Dans le cas de la lucéférine des lucioles du genre *Photinus*, il a été montré qu'il existait

deux énantiomères, contenant soit un résidu D-cystéine, soit un résidu L-cystéine. Les deux énantiomères sont activés par la luciférase pour former un luciféryl-adénylate. Cependant, seule l’oxydation du D-luciféryl-adénylate est accompagnée d’une émission de lumière [86].

### 1.5.3 Acides aminés D sous forme libre dans les cellules vivantes

Le développement de méthodes analytiques, souvent basées sur l’utilisation d’enzymes spécifiques des acides aminés D, a permis de détecter ces acides aminés sous forme libre dans divers organismes. Chez *E. Coli*, la D-Ala et le D-Glu sont utilisés dans la synthèse du peptidoglycane de la paroi. Il n’est donc pas étonnant que des concentrations élevées de D-alanine et de D-glutamate, 0.5 et 1 mM respectivement, soient présentes chez la bactérie *E. Coli* K12 cultivée en milieu minimum avec glucose [87]. Une étude réalisée sur *E. Coli* a révélé qu’entre 15 et 20% de l’ensemble des acides aminés existait sous forme D. La D-Ala y est le plus abondant des acides aminés D et représente 53% de la quantité totale d’alanine. Par ailleurs, 40% du glutamate, 50% de la méthionine et 5% de la valine se trouveraient sous forme D.

Plus généralement, Preston [88] a étudié la distribution des acides aminés D dans 43 espèces d’invertébrés marins. 18 d’entre elles, appartenant à 6 embranchements différents (*Annelida, Mollusca, Echinodermata, Nemertea, Cnidaria et Arthropoda*) contiennent des acides aminés D. Cette vaste distribution suggère des rôles physiologiques ou nutritionnels des acides aminés D chez ces organismes marins. La plus grande concentration totale d’acides aminés D (plus de 40 mM) a été trouvée dans le polychète *Glycera dibranchiata*. Dans cet organisme, un système de transport spécifique pour les acides aminés D a été identifié.

Pour les vertébrés, diverses études, réalisées essentiellement chez les mammifères, ont montré que des acides aminés D libres étaient présents dans les tissus et dans les liquides physiologiques des vertébrés [89]. Les acides aminés D-Asp et D-Ser sont les plus abondants. Chez le rat, de l’acide D-aspartique a été trouvé dans différentes régions du cerveau et dans les organes périphériques endocriniens [90]. La localisation du D-Asp dans les neurones de certaines régions spécifiques du cerveau, ainsi que dans divers organes en-

docriniens du rat, laisse supposer que cet acide aminé D a un rôle de neurotransmetteur dans la modulation neuroendocrine. On trouve également du D-Asp dans le cerveau humain et dans les embryons de poulet. Plus récemment, il a été confirmé que le D-Asp joue un rôle important dans le système neuroendocrinien, ainsi que dans le développement du système nerveux [91] (table 1.9).

La D-sérine est présente à des concentrations relativement grandes, 32% de la sérine libre totale dans le cerveau de mammifère. La distribution de la D-Ser dans certaines régions spécifiques du cerveau et son affinité envers les récepteurs pour le glutamate qui répondent au N-méthyl-D-aspartate (récepteurs NMDA) indiquent un rôle de neuromodulateur pour cet acide aminé D. Une partie de la D-Ser pourrait être apportée avec la nourriture, car cet acide aminé D franchit facilement la barrière sanguine cérébrale. D'autre part, l'identification d'une D-sérine racémase dans le cerveau de rat suggère qu'il y a aussi une synthèse endogène de D-sérine à partir de L-sérine [92]. L'implication de la D-Ser dans la fonction cérébrale est en bon accord avec l'observation d'effets favorables de l'addition de cet acide aminé aux antipsychotiques utilisés lors des traitements de la schizophrénie [93]. En effet, un fonctionnement insuffisant des récepteurs NMDA semble jouer un rôle dans cette maladie. La D-sérine, en interagissant spécifiquement avec les récepteurs NMDA, pourrait augmenter la neurotransmission associée. D'autres acides aminés D, comme la D-Ala, la D-Leu, la D-Thr, la D-Val, la D-Phe, la D-Tyr, la D-Pro, le D-Trp, le D-Glu, la D-Gln et la D-Met ont été détectés en petites quantités chez les vertébrés. Chez le rat, de la D-Ala et de la D-Leu ont été observées dans certaines régions cérébrales, mais à des concentrations de un à deux ordres de grandeur plus faibles que celles de D-Asp ou de D-Ser [90]. L'analyse des liquides physiologiques humains a révélé la présence de D-Phe, D-Pro, D-Trp et D-Tyr dans le plasma, l'urine, le liquide cérébro-spinal et le liquide amniotique, avec des rapport D/(D+L) compris entre 0.15 et 4% [89]. Les concentrations des acides aminés D sont plus élevées dans le sérum de patients souffrant d'insuffisances rénales, ce qui suggère un lien entre les acides aminés D et la fonction rénale [94]. Une étude sur le venin de mammifère a été menée en 2008 sur l'ornithorinque. Trois types de peptides ont été identifiés : les peptides natriurétiques (OvCNPs), les defensin-like pep-

TABLE 1.9 – D-Asp dans les tissus nerveux et endocriniens dans différentes espèces.

Embranchement	Animal	Tissue	References
Mollusque	<i>Octopus vulgaris</i>	Cerveau	(D'Aniello and Giuditta, 1977 ; D'Aniello and Giuditta, 1978)
	<i>Octopus vulgaris</i>	Rétine	(D'Aniello et al., 2005)
	<i>Octopus vulgaris</i>	Organes reproducteurs	(D'Aniello et al., 1995a)
Opistobranche	<i>Aplysia fasciata</i>	Ganglion oesophagien	(D'Aniello et al., 1992a)
	<i>Aplysia limacina</i>	Ganglion cérébral	(Spinelli et al., 2006)
Urochordés	<i>Ciona intestinalis</i>	Ganglion cérébral	(D'Aniello et al., 1992b ; D'Aniello et al., 2003)
Crustacé	Rock lobster	Cerveau	(Okuma and Abe, 1994)
	Rock lobster	Rétine	(Okuma and Abe, 1994)
Amphibien	<i>Rana esculenta</i>	Organes reproducteurs	(Di Fiore et al., 1998)
		Harderian glande	(Raucci et al., 2005)
Reptile	<i>Podarcis s. sicula</i>	Organes reproducteurs	(Assisi et al., 2001)
Poisson	<i>Merluccius merluccius</i>	Cerveau	(D'Aniello et al., 1995b)
Oiseau	<i>Gallus gallus</i>	Cerveau, rétine (embryo)	(Neidle and Dunlop, 1990)
Mammifère	Rat	Cerveau, rétine (adult)	(Neidle and Dunlop, 1990)
		Cerveau (embryo)	(Dunlop et al., 1986)
		Cerveau (adulte)	(D'Aniello et al., 1993b ; Dunlop et al., 1986 ; Hamase et al., 1996 ; Hashimoto et al., 1993b ; Imai et al., 1995)
Human		Rétine (6 jours)	(Neidle and Dunlop, 1990)
		Rétine (adulte)	(Lee et al., 1999 ; Neidle and Dunlop, 1990)
		Hypophyse (2jours)	(D'Aniello et al., 1996 ; Lee et al., 1999)
		Hypophyse (adulte)	(D'Aniello et al., 2000a ; Hashimoto et al., 1993a,b ; Lee et al., 1999 ; Lee et al., 1997b)
		Glande pineale	(Imai et al., 1995 ; Lee et al., 1997a)
		Organes reproducteurs	(D'Aniello et al., 1996 ; Hashimoto et al., 1993a)
		Foie	(D'Aniello et al., 1993a ; D'Aniello et al., 2000a ; Hashimoto et al., 1995)
		Cerveau (embryon)	(Hashimoto et al., 1995)
		Cerveau (adulte)	(D'Aniello et al., 1998a ; Fisher et al., 1991 ; Hashimoto et al., 1993a)

tides (OvDLPs) et les facteurs de croissance des tissus nerveux (OvNGF). Leurs fonctions exactes ne sont pas encore connues. Un venin-peptide isomérase et une hyaluronidase ont été identifiés mais pas encore séquencés. Ce venin est aussi connu pour avoir une activité protéasique [95]. Chez les animaux et les insectes, certains peptides avec des aa-D sont plus efficaces dans leur fonction que leur isomère de la forme L. Par exemple, la dermorphin, un heptapeptide contenant de la D-alanine en position 2, isolé à partir de la peau d'une grenouille du sud de l'Amérique *Phyllomedusa sauvagei*, est mille fois plus efficace que la morphine pour l'anesthésie quand elle est injectée dans le cerveau. L'existence de venin chez l'ornithorinque suggère la présence d'une isomérase responsable de l'interconversion entre les deux peptides isomères. En 2005, un venin d'ornithorinque de 5 kDa a été découvert. Ce fut le premier venin de mammifère identifié. La structure de ce peptide n'a pas encore été résolue.

## 1.6 Interactions entre la TyrRS et la D-Tyr

### 1.6.1 Etude expérimentale de la TyrRS native liée à la D-Tyr

Calendar et Berg, en 1966, découvrent les premiers que la TyrRS est capable d'ami-noacyler l'ARNt avec la L ou la D-tyrosine [96]. Sheoran et coll. en 2008 ont publié une étude détaillée du mécanisme d'activation de la D-tyrosine par la TyrRS de *B. stearothermophilus*. Ils réalisent une étude de la cinétique en phase pré-stationnaire de la réaction [97]. Bien que le substrat préféré soit la L-Tyr, la TyrRS possède aussi une activité naturelle, plus faible, pour la D-Tyrosine. Ainsi, la *Bst*-TyrRS lie la D-Tyrosine avec 8.5 fois moins d'affinité que la L-Tyrosine :  $K_d^{L-Tyr} = 12 \mu M$  et  $K_d^{D-Tyr} = 102 \mu M$ . La synthétase lie la D-Tyrosine sur l'un des deux monomères, comme avec la L-Tyrosine. Mais d'une manière surprenante, le pyrophosphate se lie à l'intermédiaire TyrRS:D-TyrAMP avec 14 fois plus d'affinité qu'avec le complexe TyrRS:L-TyrAMP ( $K_d^{PPi} = 0.043$  pour TyrRS:D-TyrAMP:PPi). Par ailleurs, l'ARNt<sup>Tyr</sup> se lie au complexe TyrRS:D-TyrAMP avec une affinité 2.3 fois moins forte qu'au complexe TyrRS:L-TyrAMP. La similarité des  $K_d$  pour la D-Tyr et la L-Tyr suggère que leurs chaînes latérales sont dans une orientation très

proche dans la TyrRS, et qu'un oxygène du groupement carboxylate est correctement positionné pour l'attaque du phosphate  $\alpha$  de l'ATP.

**Analyse des profils d'énergie libre pour l'activation de la D-Tyrosine** Les énergies libres standard de la *Bst*-TyrRS et des complexes qu'elle forme avec ses ligands ont été calculées avec les énantiomères L et D de la tyrosine. On peut donc évaluer la déstabilisation des complexes avec la D-Tyrosine par rapport à ceux avec la L-Tyrosine. Ici les  $\Delta\Delta G^0$  représentent les différences d'énergie libre de liaison entre les forme D et L [97] :

$$\Delta\Delta G_{TyrRS:Tyr}^0 = 5.3 \text{ kJ.mol}^{-1},$$

$$\Delta\Delta G_{TyrRS:Tyr:ATP}^0 = 8.7 \text{ kJ.mol}^{-1} \text{ et}$$

$$\Delta\Delta G_{TyrRS:[Tyr-ATP]}^0 = 11.3 \text{ kJ.mol}^{-1}.$$

Au final, l'intermédiaire TyrRS:TyrAMP est presque identique pour la L-Tyr et pour la D-Tyr :  $\Delta\Delta G_{TyrRS:TyrAMP}^0 = 2.2 \text{ kJ.mol}^{-1}$  (figure 1.24).

Dans d'autres organismes, ces valeurs sont un peu différentes. Par exemple, remplacer la L-Tyrosine par la D-Tyrosine diminue les  $V_{max}$  de l'enzyme d'un facteur 10 pour la TyrRS de *B. subtilis*, mais seulement d'un facteur 5 pour celle d'*E. coli*. [98]

Cette même équipe propose un modèle pour la liaison de l'ATP à la TyrRS en présence de D-Tyrosine, illustré dans la figure 1.25. Les monomères fonctionnels et non fonctionnels de la TyrRS sont figurés, respectivement, par des rectangles et des ovales. Dans le cas de l'activation de la D-Tyrosine, la liaison de l'acide aminé diminue l'affinité du monomère fonctionnel pour l'ATP, permettant la liaison de l'ATP dans le monomère non fonctionnel. La liaison de l'ATP dans le second monomère induit un changement de conformation qui mène à un monomère fonctionnel liant l'ATP avec une plus grande affinité.

### 1.6.2 Modélisation moléculaire de la TyrRS et de la D-Tyrosine

Li et coll. en 2008 ont conduit plusieurs études de dynamique moléculaire sur la structure de TyrRS de *S. aureus* dans sa forme libre et complexée à différents ligands : Tyr, ATP, tyrosyl-adenylate et inhibiteurs [99]. Ces modélisations ont pour but d'étudier les changements de stabilité conformationnelle associés au cycle catalytique. Les résultats

montrent que la forme libre de *S. aureus*-TyrRS est dans une conformation plus relaxée. La boucle KMSKS adopte une configuration ouverte dans la forme libre. La tyrosine se lie dans la poche catalytique avec la boucle KMSKS oscillant entre position semi-ouverte et ouverte. La liaison de l'ATP maintient la boucle KMSKS dans sa configuration ouverte. Une fois que le TyrAMP est formée, les deux premiers résidus de la boucle KMSKS s'orientent dans la direction opposée et entraînent la boucle vers une conformation proche de la configuration fermée ; les trois derniers résidus, SKS, adoptent une configuration entre la forme ouverte et semi-ouverte. Cette structuration particulière pourrait être impliquée dans l'initiation de la liaison de l'ARNt (table 1.10).

Thompson et coll. en 2007 ont comparé les affinités de L-Tyr et D-Tyr pour la TyrRS d'*E. coli*, en utilisant des simulations de dynamique moléculaire et des calculs d'énergie libre. Au cours des simulations, le ligand est transformé progressivement de L-Tyr en D-Tyr. On estime le travail nécessaire à la transformation, duquel on peut déduire un changement d'énergie libre [100]. Cette méthode donne une différence d'énergie libre d'association de  $\Delta\Delta G = 3.1 \text{ kcal mol}^{-1}$  en faveur de la L-Tyr, en bon accord avec le ratio des  $K_m$  expérimentaux.

TABLE 1.10 – Systèmes initiaux pour les simulations de dynamique moléculaire de la TyrRS de *S. aureus* [99].

Système	PDB	Construction
Apoenzyme	1J1L	Construit par suppression du ligand de 1J1L
TyrRS + Tyr	4TS1	Construit par superposition de 1JIL et 4TS1 ; TyrRS de 1JIL et ligand Tyr de 4TS1
TyrRS + ATP	1H3E	Construit par superposition de 1JIL et 1H3E ; TyrRS de 1JIL et ligand ATP de 1H3E
TyrRS + TyrAMP	3TS1 + 1I6 K	Ajout d'un oxygène du ligand tryptophanyl-adenylate de 1I6K au tyrosinyl-adenylate de 3TS1 ; en superposant 1JIL avec 3TS1 la TyrRS de 1JIL et TyrAMP de 3TS1
TyrRS + inhibitor (SB284485)	1JIL	TyrRS de 1JIL

FIGURE 1.23 – Des séquences mutantes de la TyrRS de *M. jannaschii*, spécifiques pour des ligands non naturels.

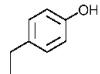
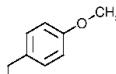
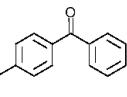
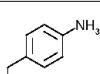
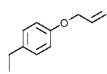
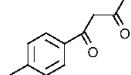
Séquence									Ligand	Ligand	Référence							
32	65	107	108	158	159	162	167											
Y	L	E	F	D	I	L	A	Tyrosine										enzyme sauvage Zhang 2010 Protein Science 14:1340
Q		T		A		P		O-méthyl-Tyr										Wang 2001 Science 292:498
G	S	T	S	T	S			p-benzoyl-Phe										Chin 2002 JACS 124:9026
L		P		A	Q	V		3-(2-naphthyl)Ala										Wang 2002 JACS 124:1836
T	N	P	L	Q				p-azido-Phe										Chin 2002 JACS 124:9026
T	S	P	S	Q				p-iodo-Phe										Xie 2004 Nature Biotechnology 22:1297
L	S	P	L	E				p-acéthyl-Phe										Wang 2003 PNAS 100:56
T	T	P	L	A				p-amino-Phe										Santoro 2002 Nature Biotechnology 20:1044
S	T	T	Y	A				O-allyl-Tyr										Zhang 2003 Biochemistry 42:6735
G	V	T	G		S			$\beta$ -diketone group										Zeng 2006 Bio.Med.Chem.Letters 16:5356
L	S	P	I	E				p-bromophenylalanine										Xie 2006 Bioinformatics 22:188

FIGURE 1.24 – Diagramme d'énergie libre pour l'aminoacylation de l'ARNt<sup>Tyr</sup> par la L- et la D-Tyrosine pour *Bst*-TyrRS : Le trait plein indique l'énergie libre au cours de la réaction d'activation de la L-Tyrosine et son transfert à l'ARNt<sup>Tyr</sup>, le trait en pointillé est pour la D-Tyrosine.

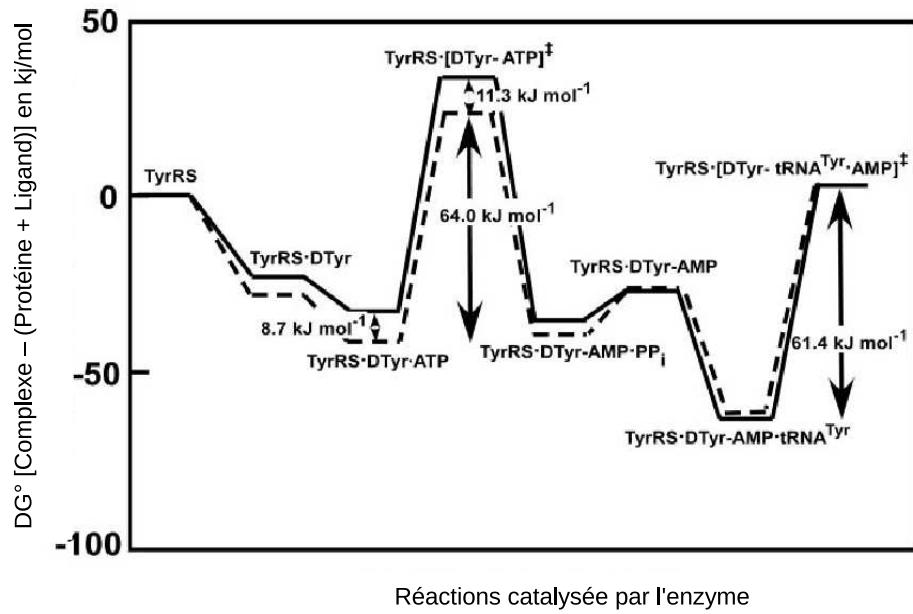
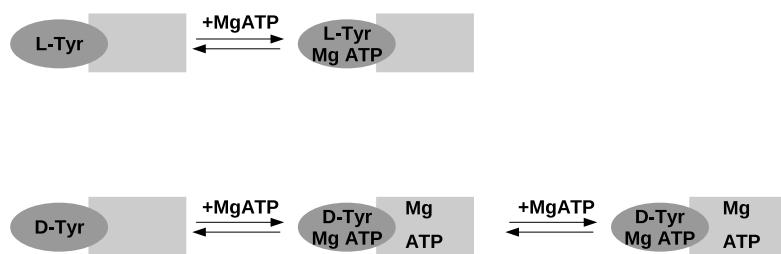


FIGURE 1.25 – Un modèle de la liaison d'ATP à la TyrRS en présence de D-Tyrosine.





# Chapitre 2

## Le "dessein" computationnel de protéines, ou CPD

L'ingénierie des protéines peut être abordée par plusieurs approches. La première est l'ingénierie rationnelle, pour laquelle on identifie des résidus cibles à partir de l'analyse de données de séquences et de structures. Une fois les mutations d'intérêt identifiées, les structures sont fabriquées par mutagénèse dirigée. Un exemple d'application est la comparaison de protéines apparentées. La barnase et la binase par exemple, sont des ribonucléases ayant 85% d'identité de séquence mais qui diffèrent par leur thermostabilité. En utilisant la mutagénèse dirigée, Serrano et coll. en 1993 ont testé l'impact de chacun des 17 acides aminés qui diffèrent entre les deux protéines [101]. Ils ont observé des changements de stabilité de  $\pm 1.1 \text{ kcal.mol}^{-1}$  par rapport à l'enzyme sauvage (figure 2.1). La deuxième approche est l'ingénierie combinatoire, ou évolution dirigée. On crée ici directement une banque de variants. Ceux ayant l'activité recherchée sont isolés par criblage de la banque, puis leur activité est caractérisée. Pour l'ingénierie combinatoire, on introduit les mutations par PCR avec production d'erreurs, par mutagénèse chimique, ou avec des oligonucléotides de synthèse contenant les mutations. L'intérêt de cette méthode est de permettre le criblage d'un grand nombre de mutants. Une autre approche d'ingénierie combinatoire est la recombinaison entre séquences d'ADN, "DNA shuffling". On utilise cette méthode lorsque la protéine cible appartient à une famille connue de protéines. Dans

FIGURE 2.1 – Mutagénèse d'une protéine par une méthode PCR mutagène.

1 : PCR I : On amplifie le gène avec une amorce sens "mutagène" + une amorce reverse normale. Une amorce "mutagène" est une amorce dans laquelle la séquence est spécifiquement modifiée pour provoquer, à la position déterminée, le changement souhaité de la séquence du gène amplifié.

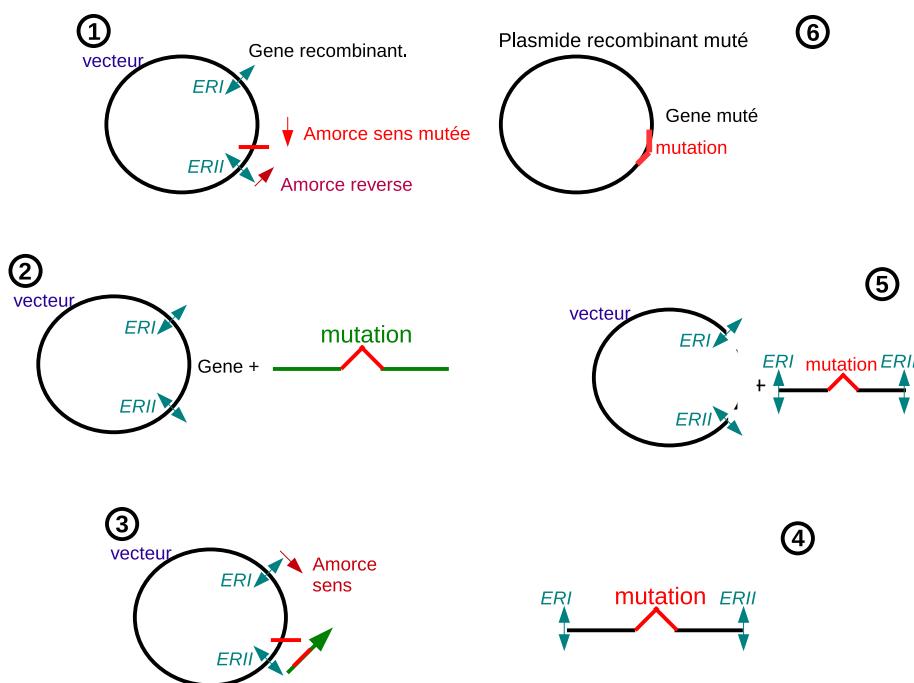
2 : Purification du produit PCR.

3 : PCR II : On amplifie le gène avec une amorce sens normale + le produit PCR I comme amorce reverse "megaprimer".

4 : Purification du produit PCR II.

5 : Digestion de l'ADN plasmidique et du produit PCR II par les enzymes de restriction ERI et ERII.

6 : Ligation de l'insert et du vecteur et transformation d' *E. coli*.



---

ce cas, on isole les gènes des différents membres de la famille pour créer des hybrides artificiels (figure 2.3). L'ingénierie combinatoire a été utilisée pour l'isolement d'un mutant de subtilisine de thermostabilité accrue. Mille clones mutants ont été étalés sur une boîte quadrillée, en deux exemplaires. Une fois qu'ils ont poussé, on incube une des boîtes à température élevée pour inactiver l'enzyme de type sauvage. Puis un test d'activité hydrolytique permet de ne sélectionner que les mutants de stabilité supérieure à celle du type sauvage, car ce sont les seuls à avoir une activité mesurable. Une fois les mutants stables identifiés, il suffit de faire pousser les doubles de ces colonies pour identifier la mutation. Après avoir étudié les changements individuels d'acides aminés, on peut fabriquer une subtilisine extrêmement stable en combinant différentes mutations dans la même molécule [102] (figure 2.2). Pendant ces quinze dernières années, chacune des propriétés de cette protéase à sérine a été modifiée. La majeure partie des modifications avait pour but d'améliorer les capacités de la subtilisine pour l'hydrolyse des protéines lorsqu'elle est incorporée à des détergents [103]. Ainsi, ses caractéristiques cinétiques, sa spécificité de substrat, sa dépendance vis-à-vis du pH, sa stabilité vis-à-vis d'une inactivation par la chaleur ou par les alcalins, ont été étudiés et améliorés. Ces transformations ont nécessité de muter 50% des 275 acides aminés de la subtilisine. Pour certaines positions de la molécule, même les effets du remplacement par les 19 autres acides aminés naturels ont été évalués.

En plus des méthodes expérimentales de génie génétique, les méthodes d'ingénierie de protéines *in silico* se développent. Une méthode d'ingénierie combinatoire est le "dessein computationnel", ou "Computational Protein Design", que nous désignons par "CPD". Elle comporte une série d'ingrédients. La prédiction de structure repose sur une description simple de l'espace conformationnel de la protéine, prenant en compte une ou quelques conformations de la chaîne principale et un jeu discret de rotamères pour les chaînes latérales. La structure la plus stable de cet espace discrétré peut être alors trouvée par des méthodes de recherche exactes ou approchées. La modélisation nécessite une fonction d'énergie, qui prend en compte les effets du solvant ; on parle de fonction d'énergie "effective". En effet, le solvant aqueux joue un rôle dans la structure et la stabilité des protéines.

FIGURE 2.2 – Ingénierie combinatoire des protéines par criblage. Les critères de criblage sont par exemple l'efficacité de l'expression, la thermostabilité, la résistance aux solvants organiques, la résistance au pH ou encore la spécificité modifiée.

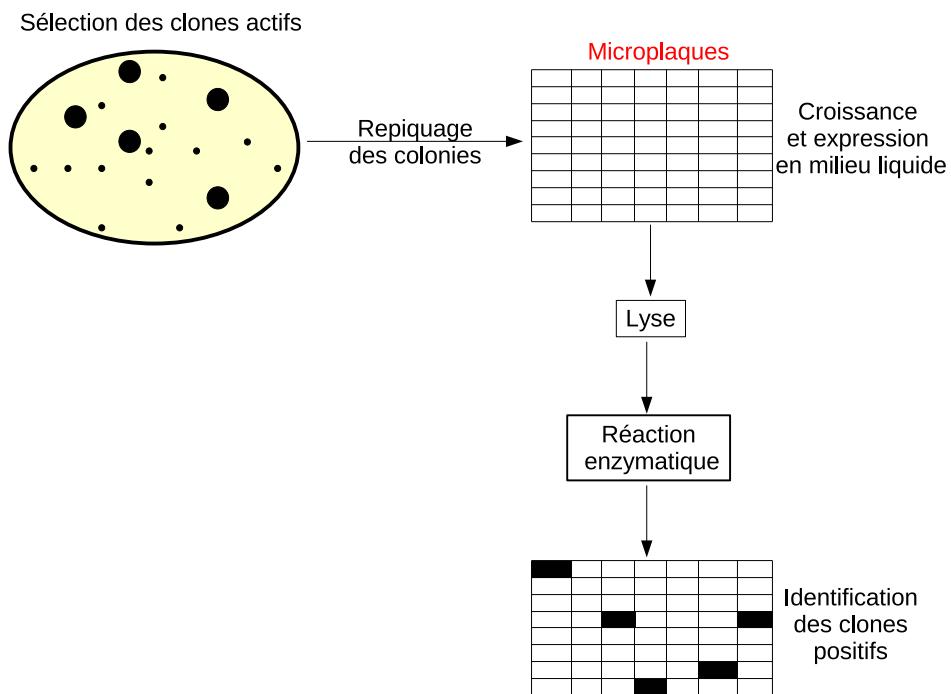
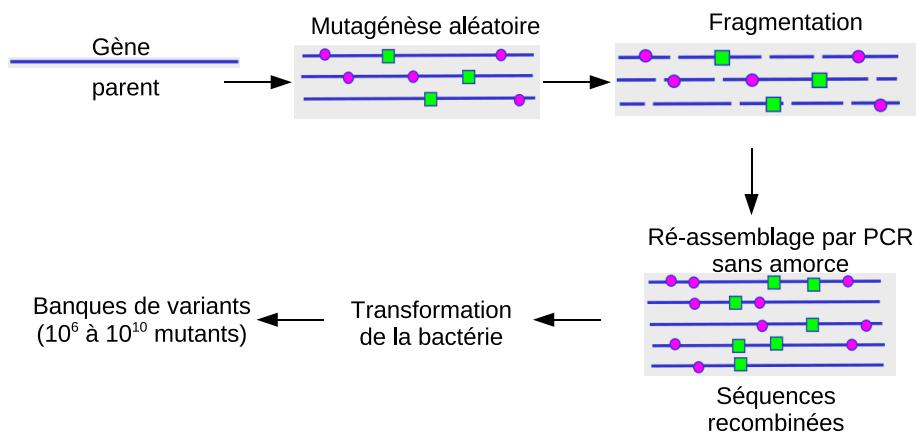


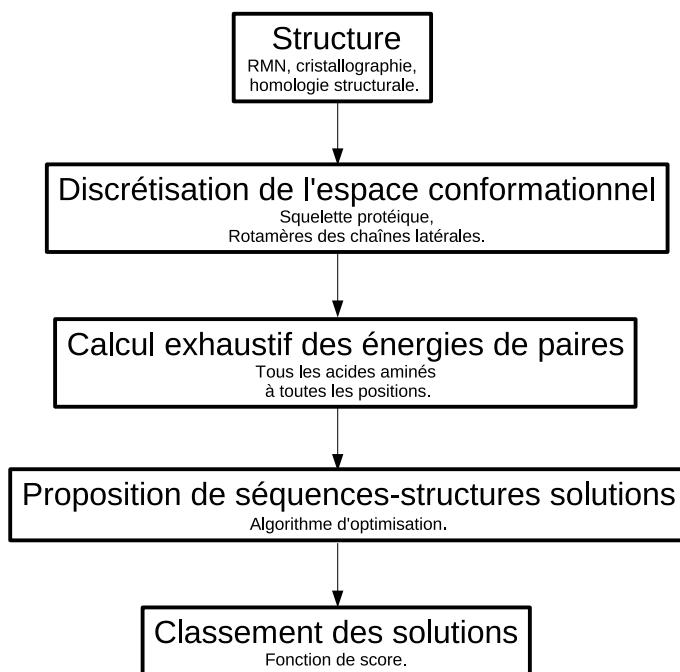
FIGURE 2.3 – Ingénierie combinatoire des protéines par DNA shuffling



---

Pour limiter le coût de la modélisation, ce rôle est pris en compte avec des modèles de solvant "implicite". Les degrés de liberté du solvant ne sont pas explicitement représentés, mais considérés à travers leurs effets sur les interactions intra-protéines. Munis d'une fonction d'énergie effective, les algorithmes de recherche permettent alors d'examiner un large champ combinatoire de séquences et de structures. Le déroulement général d'une procédure de CPD est présenté brièvement dans la figure 2.4.

FIGURE 2.4 – Déroulement de la procédure de protein design, computationnel, ou CPD.



Les buts du CPD sont nombreux. Cette méthode vise à construire des protéines plus stables, à modifier les interactions protéine-protéine, à étudier des conformations alternatives pour une protéine, à modifier la spécificité pour un substrat, à ajouter des sites de liaison d'atomes métalliques dans une protéine, ou encore à étudier le fonctionnement des enzymes. Cette méthode continue d'être améliorée pour accroître sa précision, son

efficacité et la portée de ses applications. Dans ce chapitre de méthodologie, nous décrirons tout d'abord la fonction d'énergie issue d'un champ de force. Nous verrons également comment ces derniers sont paramétrés. Puis nous décrirons différentes modélisations du solvant. Le CPD nécessite de modéliser l'état replié des protéines mais également l'état dénaturée, nous détaillerons les hypothèses faites pour y parvenir. Nous énumérerons ensuite les principaux algorithmes d'optimisation utilisés pour l'optimisation des séquences et des structures. Une fois les prédictions structurales faites, il est courant d'analyser le comportement de la structure à travers le temps sur quelques nanosecondes. La méthode de simulation de dynamique moléculaire par laquelle on y parvient sera décrite ici. Enfin, nous présenterons quelques applications du Protein Design ainsi que ses récents succès.

## 2.1 La fonction d'énergie

Le protein design repose sur une fonction d'énergie de paires permettant d'évaluer l'énergie du système en calculant la somme des interactions entre toutes les paires d'acides aminés. La classe la plus importante est représentée par les fonctions d'énergie de mécanique moléculaire.

### 2.1.1 Mécanique moléculaire, champ de force

La mécanique quantique ne permet pas d'étudier les systèmes de plus de quelques centaines d'atomes. C'est pourquoi l'énergie interne de la protéine est décrite, la plupart du temps, par une fonction empirique issue d'un champ de force comme CHARMM19 ou AMBER, reposant sur des critères de mécanique classique [104, 105]. Dans ce modèle, les atomes sont considérés comme des sphères de rayons et de charges déterminés et les liaisons entre atomes sont considérés comme des ressorts. La fonction d'énergie inclut des termes d'elongation de liaisons, de déformation d'angles, de rotation d'angles dièdres, de torsions "impropres" (maintenant planéité et chiralité), des interactions de Van der Waals, ainsi que des interactions électrostatiques coulombiennes. Au champ de force, il faut aussi ajouter un ou plusieurs termes d'énergie qui décrivent les interactions avec le solvant.

Finalement :

$$E = E_{bonds} + E_{angl} + E_{dih} + E_{impr} + E_{vdw} + E_{coul} + E_{solv} \quad (2.1)$$

Par la suite nous présenterons ces termes plus en détail (figure 2.5).

### 2.1.1.1 Energies liées

Les énergies liées correspondent aux quatre premiers termes de E dans l'équation (2.1).

Les termes  $E_{bonds}$  et  $E_{angl}$  correspondent respectivement aux déformations des liaisons et des angles. Ils sont représentés par des oscillateurs harmoniques. L'énergie est sommée sur toutes les liaisons covalentes et sur tous les angles entre atomes liés par liaisons covalentes.

- Déformation des liaisons :

$E_{bonds}$  est exprimée par la fonction suivante, avec  $i$  la liaison de longueur  $r$ ,  $k$  la constante de force du ressort en  $\text{kcal.mol}^{-1}.\text{\AA}^{-2}$  et  $r^0$  la longueur de liaison optimale :

$$E_{bound} = \frac{1}{2} \sum_{i=1}^n k_{r,i} (r_i - r_i^0)^2 \quad (2.2)$$

L'élongation des liaisons est en général faible, de l'ordre de 0.05 Å.

- Déformation des angles :

$E_{angl}$  est exprimée par la fonction suivante, avec  $\theta$  l'angle entre la liaison  $i$  et la liaison  $j$ ,  $k_\theta$  la constante de force exprimée en  $\text{kcal.mol}^{-1}.\text{degré}^{-2}$  et  $\theta_0$  l'angle optimal :

$$E_{angl} = \frac{1}{2} \sum_{ij} k_{\theta,ij} (\theta_{ij} - \theta_{ij}^0)^2 \quad (2.3)$$

La déformation des angles est également faible, seulement de l'ordre de quelques degrés.

- Déformation des angles dièdres :

Quatre atomes  $a$ ,  $b$ ,  $c$ ,  $d$  reliées par trois liaisons covalentes ( $a-b$ ,  $b-c$ ,  $c-d$ ) définissent

un angle dièdre, correspondant à une rotation des deux liaisons a-b et c-d autour de la liaison centrale b-c. L'énergie de torsion de chaque dièdre i est décrite par une fonction périodique, développée en série de Fourier, où  $A_{i,n}$  est la valeur de la barrière de rotation d'ordre n en  $\text{kcal.mol}^{-1}$ ,  $\gamma$  la phase (en degrés) et  $\phi$  la valeur du dièdre (en degrés) :

$$E_{dih} = \frac{1}{2} \sum_i A_{i,n} [1 + \cos(n\phi_i - \gamma)] \quad (2.4)$$

- Déformation des angles impropre :

Les dièdres impropre permettent de fixer planéité et chiralité d'un centre atomique i, relié à trois autres atomes j, k et l.  $E_{impr}$  a la forme :

$$E_{impr} = \frac{1}{2} A (\phi_{ijkl} - \phi_{ijkl}^0)^2 \quad (2.5)$$

Ici, A est la constante de force exprimée en  $\text{kcal.mol}^{-1}.\text{degré}^{-2}$ .  $\phi$  représente l'angle entre la liaison ij et le plan formé par les atomes i, k et l.

### 2.1.1.2 Energies non liées

Les interactions non liées agissent entre deux atomes qui appartiennent à des molécules différentes, ou qui appartiennent à la même molécule mais sont séparés par plus de trois liaisons covalentes. Ces interactions sont décrites par deux termes : un terme de Lennard-Jones, aussi appelé terme de Van der Waals et un terme électrostatique.

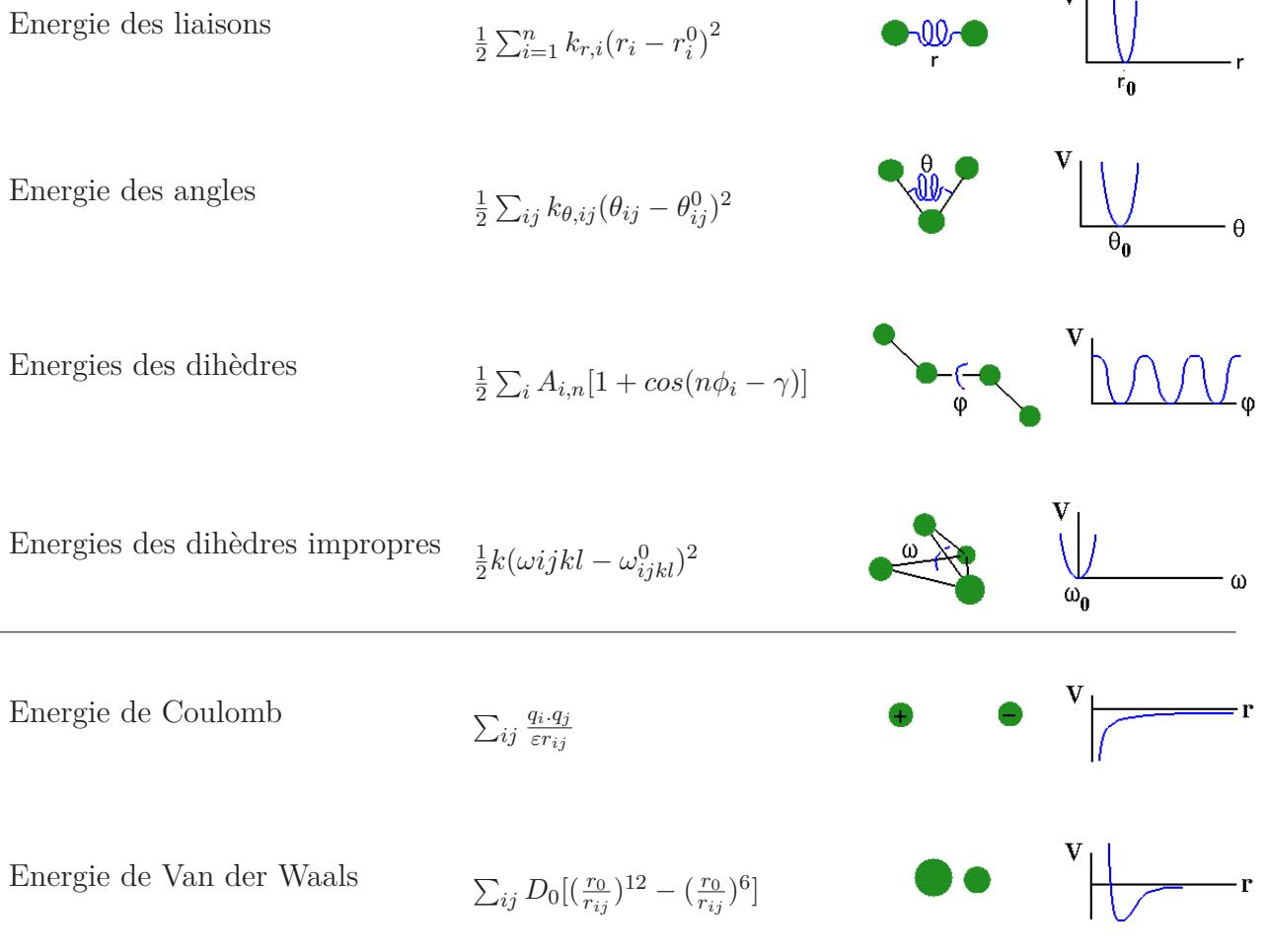
- Energie de Van der Waals :

L'énergie de Van der Waals entre deux atomes i et j a la forme suivante :

$$E_{vdw} = D_0 \left[ \left( \frac{r_0}{r_{ij}} \right)^{12} - \left( \frac{r_0}{r_{ij}} \right)^6 \right] \quad (2.6)$$

Ici,  $D_0$  et  $r_0$  sont des constantes et  $r_{ij}$  représente la distance entre les atomes i et j. Si les atomes sont très proches l'un de l'autre, l'énergie sera défavorable. Ce terme permet d'éviter l'encombrement stérique.

FIGURE 2.5 – Fonction d'énergie de mécanique moléculaire. Les quatre premiers termes représentent les énergies d'interaction entre atomes liés ; les deux derniers termes correspondent aux interactions entre atomes non-liés.



- Energie électrostatique :

L'énergie électrostatique est calculée par un potentiel de Coulomb, qui a la forme suivante.

$$E_{elec} = \frac{q_i \cdot q_j}{\varepsilon r_{ij}} \quad (2.7)$$

Ici,  $q_i$  et  $q_j$  représentent les charges respectives des atomes i et j,  $r_{ij}$  représente la distance entre les deux atomes,  $\varepsilon$  est la constante diélectrique du milieu.

### 2.1.2 Paramétrisation des principaux champs de forces

En plus des coordonnées atomiques de la protéine, la fonction d'énergie dépend de paramètres décrivant la géométrie et les propriétés énergétiques des interactions interatomiques. Contrairement aux coordonnées, ces paramètres sont invariants au cours des simulations. La combinaison de la fonction d'énergie et de ces paramètres est nommé "champ de force". L'optimisation des paramètres du champ de force implique un ajustement de ses paramètres jusqu'à ce qu'il soit capable de reproduire un ensemble de données ciblées au préalable. Ces données sont en général issues de mesures spectroscopiques, thermodynamiques et cristallographiques, ou encore obtenues par calcul de mécanique quantique. Le concept de "type atomique" est utilisé pour différencier des atomes de même numéro atomique mais avec des environnements chimiques différents. Par exemple, un atome d'hydrogène lié à un oxygène n'a pas le même type atomique, et donc pas les mêmes paramètres qu'un atome d'hydrogène lié à un carbone aromatique. Il existe plusieurs champs de force optimisés pour les protéines. Les quatre principaux sont résumés ci-dessous.

**AMBER : Assisted Model Building with Energy Refinement** Ce champ de force a été paramétré sur des données expérimentales de petites molécules et de calculs *ab initio* pour les charges [106]. Plus précisément, AMBER utilise des potentiels de surface électrostatiques calculés par mécanique quantique comme données cibles pour la détermination des charges partielles. Ainsi, les charges partielles d'AMBER reproduisent les potentiels électrostatiques en phase gazeuse. Les ressources informatiques actuelles ont permis d'inclure de plus grandes molécules dans le processus de paramétrisation. Ainsi, les dièdres  $\phi$  et  $\psi$  d'AMBER ont été ajustés pour mieux reproduire les propriétés conformationnelles de peptides structurés.

**CHARMM : Chemistry at HARvard Molecular Mechanics** Les champs de force de CHARMM [105] pour les protéines sont le champ de force Charmm19 à hydrogène polaire et le champ de force Charmm22 incluant tous les atomes. Ce champ de force a été paramétré à partir de calculs quantiques d'énergies d'interaction entre petites molécules et

molécules d'eau. Cette approche a pour but d'équilibrer les interactions eau-protéine, eau-eau et protéine-protéine, en phase condensée. Le champ de force Charmm27 pour l'ADN, l'ARN et les lipides, peut être combiné à Charmm22, par exemple pour les simulations de complexes protéine-ADN.

**OPLS : Optimized Potential for Liquid Simulations** Ce champ de force a été paramétré sur des données expérimentales de liquides [107]. Il existe pour ce champ de force une version "atomes unifiés", où les hydrogènes non polaires sont inclus implicitement dans les paramètres des atomes auxquels ils sont liés, et une version "tout atomes", OPLS-AA, plus récente. Cette approche utilise des simulations en phase condensée pour de nombreux liquides, qui ont tous des similarités avec les groupes chimiques présents dans les protéines, comme des alkanes, des alcools, des amides et des aromatiques. Les paramètres de Lennard-Jones et les charges partielles sont alors ajustés pour reproduire la température de vaporisation expérimentale et la densité du liquide. Ces paramètres seront donc utilisables pour les protéines modélisées dans un solvant.

**GROMOS** La paramétrisation du terme de Lennard-Jones et des charges partielles dans GROMOS utilise également des données expérimentales thermodynamiques, comme OPLS [108]. Les énergies libres de solvatation de petits composés modèles, à la fois dans l'eau et dans le cyclohexane, sont incluses comme données cibles. Le but de cette approche est de modéliser correctement les multiples interfaces de la protéine, qui a des parties exposées au solvant et des parties enfouies.

Il y a de nombreuses similarités, mais aussi des différences entre les différents champs de force. Par exemple, le champ de force CHARMM ajoute un terme d'Urey-Bradley, qui traite les deux atomes qui ne sont pas au sommet de l'angle avec un terme quadratique qui dépend de leur distance. La combinaison des dièdres impropre et du terme d'Urey-Bradley permet une reproduction plus précise des spectres vibrationnels pendant la paramétrisation. Les champs de force OPLS-AA et GROMOS utilisent les règles suivantes pour le terme de Lennard-Jones :  $\varepsilon_{ij} = (\varepsilon_i \varepsilon_j)^{1/2}$  et  $R_{min,ij} = (R_{min,i} R_{min,j})^{1/2}$ . CHARMM et AMBER préfèrent la moyenne arithmétique pour  $R_{min,ij}$  :  $R_{min,ij} = \frac{1}{2}(R_{min,i} + R_{min,j})$ . Une

autre différence concerne les interactions 1-4, entre les atomes a et d du dièdre a-b-c-d. Certains champs de force appliquent ici une constante de proportionnalité. Par exemple, AMBER multiplie les interactions 1-4 Lennard-Jones par un facteur 0.5 et les interactions coulombiennes par 1/1.2. OPLS-AA applique un facteur de proportionnalité de 0.5 pour les deux interactions. GROMOS applique un facteur au cas par cas, en fonction du type atomique. Le champ de force CHARMM aborde le problème plus simplement, en ne multipliant par aucun facteur les interactions 1-4.

**Autres champs de force** De nombreux autres champs de force ont été également paramétrés. Parmis eux, on peut citer :

- ECEPP : "Empirical Conformational Energy Program for Peptides", premier champ de force dédié aux molécules polypeptidiques [109].
- QCFF : " Quantum-mechanical Consistent Force Field" [110].
- CVFF : "consistent valence force-field" [111].
- AMOEBA : "Atomic Multipole Optimized Energetics for Biomolecular Applications". Ce champ de force est "polarisable" : les atomes portent des dipôles induits, qui sont proportionnels au champ électrostatique local et reflètent la polarisation du nuage électronique. [112].

La diversité des traitements suggère qu'il n'existe pas de solution unique au problème.

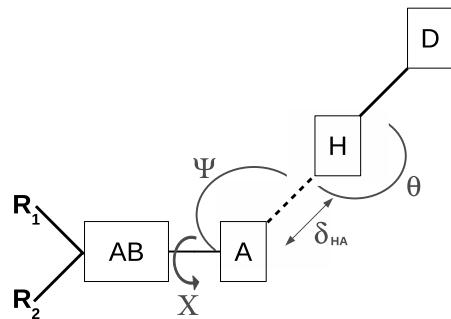
### **2.1.3 Des approches plus empiriques**

Même si les fonctions d'énergie de mécanique moléculaire sont très réalistes pour retranscrire les interactions des petites molécules, l'application au CPD peut nécessiter quelques adaptations. Ainsi, Kortemme et Baker en 2002 [113] mettent au point une fonction d'énergie très simple pour modéliser les interactions protéine-protéine. Cette fonction d'énergie ne prend en compte que les contributions les plus importantes à l'énergie totale : le potentiel de Lennard-Jones, l'énergie de solvatation, l'énergie électrostatique coulombienne et un terme spécifique pour les interactions par liaison hydrogène. L'énergie des

liaisons hydrogène est déterminée par une fonction empirique prenant en compte les distances  $\delta_{HA}$  entre hydrogènes et atomes accepteurs, une pondération  $W_{HB}$  de l'énergie des liaisons hydrogène, l'angle  $\theta$  formé par l'atome accepteur, l'hydrogène et l'atome donneur ainsi que l'angle  $\psi$  qui a pour sommet l'atome accepteur (figure 2.6) :

$$E_{HB} = W_{HB}[E(\delta_{HA}) + E(\theta) + E(\psi) + E(X)] \quad (2.8)$$

FIGURE 2.6 – Représentation schématique des paramètres utilisé pour décrire la liaison H.  $\delta_{HA}$ , distance entre hydrogène et atome accepteur.  $\theta$ , angle ayant pour sommet l'atome d'hydrogène.  $\psi$ , angle ayant pour sommet l'atome accepteur. X, angle dièdre. A, accepteur. D, donneur, H, hydrogène. AB, base accepitrice. R1 R2, atomes liés à la base accepitrice.



Une autre approche empirique est celle de Handel et Pokala qui introduisent dans la modélisation un ajustement empirique de l'énergie de van der Waals pour les interactions protéine-protéine. La fonction de Lennard-Jones a été ici modifiée pour linéariser les énergies répulsives et exagérer les énergies attractives. Ceci permet de réduire l'erreur sur l'encombrement stérique introduite par l'usage de bibliothèques discrètes de rotamères [114]. Ils ont ainsi pu prédire les effets de 400 mutations avec une erreur inférieure à 1.0 kcal.mol<sup>-1</sup>

Il est aussi possible de simplifier la modélisation de l'énergie électrostatique. L'équipe de Mayo, en 2006, a utilisé une fonction empirique pour les liaisons hydrogènes, en plus d'une fonction d'énergie coulombienne qui inclut une "constante diélectrique" dépendant de la distance égale à  $40r_{ab}$ , avec  $r_{ab}$  la distance interatomique [115]. Des charges partielles sont attribuées aux groupes polaires seulement. Une charge de +1 est appliquée à la lysine

et à l'arginine et une charge de -1 pour l'aspartate et le glutamate.

Pour les interactions protéine-ADN, la fonction d'énergie peut aussi être adaptée empiriquement. Une fonction d'énergie pour les calculs d'interactions entre protéine et acide nucléique a été développée par Siggia et coll. L'efficacité de cette fonction est dûe à un terme d'interaction électrostatique intermoléculaire combinée à un terme d'empilement atomique ("packing") et à un terme décrivant la conformation de l'ADN [116].

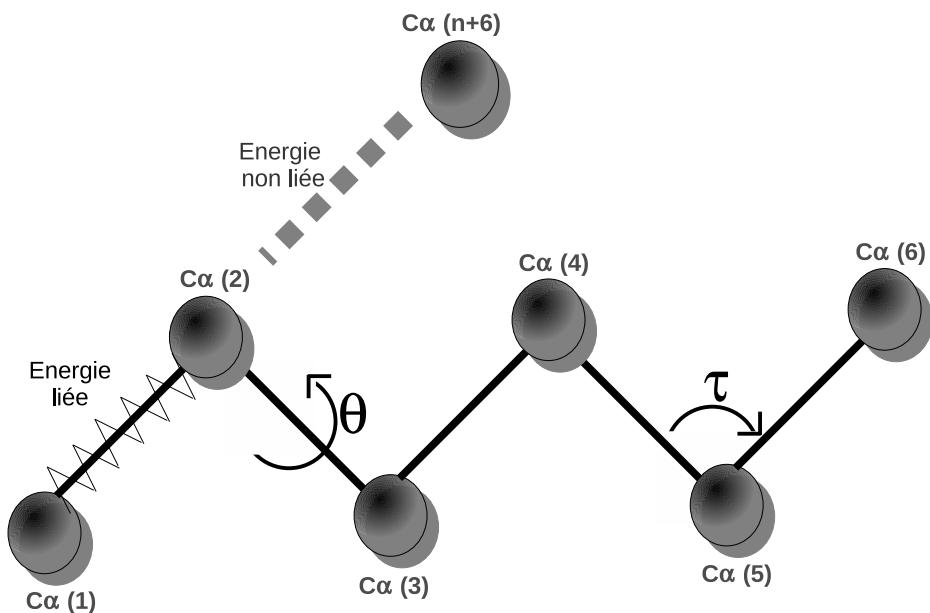
La plupart des fonctions d'énergie du CPD ne prennent pas en compte tous les facteurs possibles. Les énergies de repliement, les énergies libres de liaison, aussi bien que la cinétique de liaison et le taux catalytique résultent d'un ensemble complexe d'effets et incluent des contributions entropiques. De nouvelles études ont étendu la vision du design à la considération d'ensembles de structures, et non plus seulement de structures uniques. L'entropie associée à ces ensembles est explicitement prise en compte. Donald et coll. ont formulé le problème du protein design en utilisant des ensembles conformationnels et ont validé leur méthode en s'intéressant à la spécificité des interactions. Ils ont ainsi étudié la gramicidine synthétase A phénylalanine, en créant un nouveau mutant qui lie mieux la leucine que la phénylalanine. Des validations expérimentales pour ce mutant ont également été faites [117]. Kuhlman et coll. ont utilisé une procédure différente, basée sur un algorithme de recherche de Monte Carlo, pour inclure l'entropie liée à la conformation des chaînes latérales dans le design de 110 squelettes protéiques de protéines natives. Ils n'ont trouvé que de faibles différences dans les résultats selon que l'entropie est incluse dans l'analyse ou pas. Les plus grandes différences étaient remarquées pour les chaînes latérales longues et flexibles [118]. Bien que l'entropie ne soit pas la contribution dominante pour le protein design, l'utilisation d'ensembles semble être une voie intéressante. Dans un autre travail, avec une approche de mécanique classique, les centres réactionnels di-manganèse ont été étudiés par Spiegel et coll. en incluant un traitement implicite des transferts de charges et des effets de polarisation. Ces effets sont importants pour le passage de la forme réduite à la forme oxydée. Le design de ces enzymes a nécessité une stabilisation du centre metalloprotéique par le reste de la protéine, et ce, quel que soit l'état d'oxydation du métal. Ce travail a constitué un pas important dans le design des metalloenzymes

[119].

### 2.1.4 Champ de force à gros grains

L'informatique donne à présent les capacités d'étudier la structure d'une protéine en solution dans un champ de force tout atomes avec un traitement explicite du solvant. Mais les applications comme le docking protéine-protéine, sont trop complexes, l'étude tout atome n'est pas toujours appropriée. D'où l'idée d'intégrer un grand nombre de degrés de liberté dans un champ de force à gros grains qui en présente moins. Dans ces modèles, les groupes fonctionnels sont représentés par des particules de différentes tailles. La plupart des modèles gros grains représentent les acides aminés comme des billes reliées entre elles par des ressorts élastiques [120]. Les modèles basés sur la simplification des résidus à leur atome  $C\alpha$  sont les plus populaires [121] (figure 2.7). D'autres fonctions d'énergie

FIGURE 2.7 – Modélisation gros grain où les résidus sont réduits à leurs atomes  $C\alpha$  [121]. Différentes énergies sont prises en compte : les termes d'énergie liée, les termes d'énergie non liée. Ici,  $\theta$  représente un angle dièdre et  $\tau$  un angle entre deux liaisons covalentes.



utilisent une carte de contact résidu-résidu avec une classification en groupes d'acides aminés [122]. Si deux atomes autres que l'hydrogène appartenant aux deux protéines sont

distsants de moins de quelques Angstroms, les résidus sont considérés comme en contact. Il peut également y avoir d'autres approches ; par exemple, une fonction d'énergie distance-dépendante basée sur les atomes du squelette protéique et sur le Carbone  $\beta$ . C'est le cas dans le potentiel "DIFRE" développé par Zhou et coll. [123]. L'acide aminé peut aussi être représenté par le  $C\alpha$  et le barycentre de la chaîne latérale. Cette dernière approche traite mieux la spécificité des interactions entre acides aminés [124].

## 2.2 Modélisation implicite du solvant

Les protéines se trouvent dans un milieu aqueux, leurs résidus hydrophiles peuvent interagir directement avec les molécules d'eau. Il peut donc y avoir compétition d'interactions entre les groupes polaires de la protéine et les molécules d'eau. Ainsi, l'enfouissement de résidus chargés est généralement défavorable ; même si l'énergie d'interaction directe entre deux résidus chargés enfouis est favorable, cet effet est contré par la perte d'interactions avec le solvant pour ces deux résidus [125, 126]. La première modélisation mathématique de l'eau liquide à l'échelle atomique a été proposée par Bernal et Fowler en 1933 [127]. Depuis, un grand nombre de méthodes ont été développées pour étudier les interactions électrostatiques entre macromolécules biologiques et solvant. Les plus répandues peuvent être classées en deux catégories : les modèles où les atomes du solvant sont pris en compte explicitement, et les modèles de solvant implicites, qui représentent le solvant par ses effets moyens sur le soluté. Dans les modèles de solvatation explicite, les interactions entre les ions mobiles, le solvant et le soluté sont décrites par un champ de force de mécanique moléculaire. Celui-ci a l'avantage de représenter dans le détail les interactions solvant-soluté et solvant-solvant. Ainsi, les interactions entre chaînes latérales peuvent-être mieux retranscrites qu'avec les modèles implicites [128]. Mais le grand nombre de degrés de liberté découlant d'une modélisation explicite du solvant augmente considérablement le temps de calcul.

Dans la modélisation implicite du solvant, le solvant est représenté par son influence moyenne sur le soluté [129]. Ainsi, les interactions solvant-soluté peuvent être caractérisées par l'énergie de solvatation du soluté : l'énergie libre de transfert du soluté dans le solvant

à partir du vide. Il est intéressant de décomposer cette opération en trois étapes :

- 1- les charges du soluté sont réduites à zéro,
- 2- le soluté non-chargé est inséré dans le solvant,
- 3- les charges du soluté sont rétablies à leurs valeurs réelles.

L'énergie libre de la seconde étape est appelée l'énergie de solvatation "non polaire", alors que la somme des énergies des étapes initiale et finale est appelée énergie de solvatation "polaire". Cette dernière énergie représente les effets du solvant sur les charges du soluté (figure 2.8). En général, les énergies de solvatation polaire et non polaire agissent de manières opposées : l'énergie de solvatation non polaire favorise les structures compactes de petit volume alors que l'énergie de solvatation polaire favorise le maximum d'exposition au solvant pour les groupes polaires du soluté.

Pour la solvatation aqueuse, les modèles implicites les plus importants représentent l'énergie de solvatation polaire grâce à un modèle "diélectrique", qui traite le soluté et le solvant comme deux régions distinctes caractérisées par deux constantes diélectriques différentes. L'énergie libre de solvatation polaire est alors obtenue en résolvant l'équation de Poisson-Boltzmann.

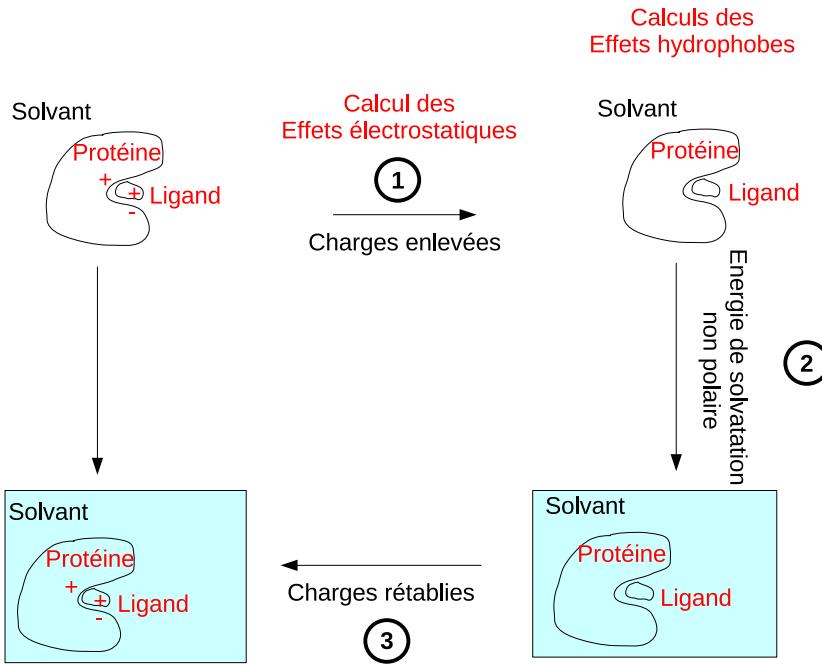
## 2.2.1 Interactions soluté-solvant

Il y a ainsi deux grands types d'interactions du solvant avec le soluté : d'une part l'effet électrostatique, ou polaire, et d'autre part l'effet hydrophobe, ou non polaire. Naturellement, l'introduction ci-dessus d'une protéine sans aucune charge est une étape artificielle, et la séparation entre ces deux effets n'est pas évidente ou complètement rigoureuse dans la réalité. Voyons dans un premier temps la modélisation de l'effet électrostatique.

### 2.2.1.1 Modèle simple d'écrantage diélectrique

Plusieurs modèles représentent le solvant uniquement par ses effets électrostatiques (étapes polaires 1 et 3 ci-dessus). Le solvant a des interactions électrostatiques directes avec le soluté, mais écrante également les interactions entre les charges du soluté. Dans le modèle le plus simple, l'effet d'écrantage est pris en compte : les interactions de Coulomb

FIGURE 2.8 – Schéma de calcul des interactions avec le solvant. Les étapes (1) + (3) donnent l'énergie électrostatique . L'étape (2) correspond à la contribution non polaire, où hydrophobe.



sont réduites par un facteur d'échelle, la constante diélectrique  $\varepsilon$ . L'énergie électrostatique entre deux atomes i et j a la forme :

$$E(r_{ij}) = \frac{q_i q_j}{\varepsilon r_{ij}} \quad (2.9)$$

Le modèle RDIE est très proche, mais ici  $\varepsilon$  dépend de la distance  $r_{ij}$ . Le solvant écrante alors les interactions entre les atomes de la protéine d'autant plus qu'ils sont éloignés [130] ; l'énergie a la forme :

$$E(r_{ij}) = \frac{q_i q_j}{\varepsilon r_{ij}^2} \quad (2.10)$$

Typiquement, on prend  $\varepsilon$  entre 1 et 4. Ces deux modèles sont peu coûteux en temps de calcul mais simplistes.

### 2.2.1.2 Les termes non électrostatiques : modèle surfacique

Le terme "non électrostatique" le plus important est celui qui décrit l'étape de solvation hydrophobe (l'étape 2 ci-dessus). Certains modèles implicites représentent ce terme

en utilisant la surface accessible au solvant.

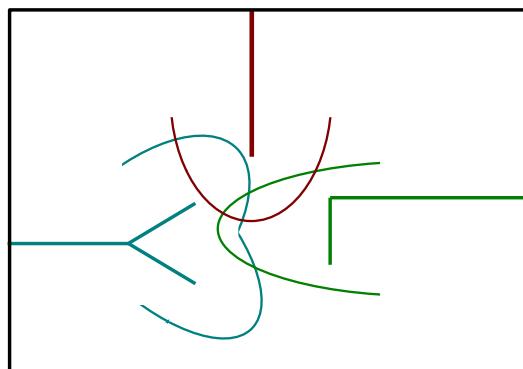
Ainsi, l'énergie libre de solvatation d'un soluté non polaire dans la plus simple des modélisations est la suivante :

$$\Delta G_{solv} = \sum_i \sigma_i A_i \quad (2.11)$$

Ici,  $A_i$  est la surface accessible au solvant de l'atome i et  $\sigma_i$  est un paramètre de solvation pour l'atome i en  $kcal.mol^{-1}.\text{\AA}^{-2}$ . Ce paramètre donne la contribution à l'énergie libre de solvatation de l'atome i par unité de surface. Les paramètres  $\sigma_i$  ont été ajustés pour les différents types atomiques par modélisation d'énergies libres de transfert (*vapeur* → *eau*, par exemple) pour une série de composés organiques.  $\sigma_i$  reflète la préférence d'un atome à être exposé au solvant ou enfoui.

Le calcul des surfaces enfouies dans le contexte du CPD pose un problème, car ces surfaces dépendent de la configuration globale de la protéine. En particulier, un point à la surface d'un atome peut être en contact avec plusieurs autres atomes à la fois. Le terme surfacique est donc un terme "à N corps", coûteux à évaluer. Des approximations courantes le représentent comme une simple somme sur des paires d'atomes. Cette approximation "somme de paires" conduit à une surestimation de la surface enfouie des N atomes. Street et Mayo corrigent cette erreur par une pondération empirique, détaillée plus loin [131] (figure 2.9).

FIGURE 2.9 – Recouplement de surface de trois résidus, illustrant les interactions surfaciques "à trois corps".



**Modèle CASA** Une autre approche empirique combine les deux modèles précédents, incluant l'écrantage des interactions électriques avec un effet de surface exposée au solvant. Ce modèle a été présenté par Wesson et Eisenberg en 1992 [132]. Le solvant est décrit par le terme suivant :

$$E_{solv} = E_{screen} + E_{surf} = \left(\frac{1}{\varepsilon} - 1\right) E_{coul} + \alpha \sum_{i=1}^n A_i \sigma_i \quad (2.12)$$

Ici, le terme  $E_{screen}$  décrit des interactions électrostatiques : c'est un terme d'énergie Coulombienne écranté par les dipôles du solvant.  $E_{surf}$  correspond à la somme des aires atomiques accessibles au solvant  $A_i$ , avec un paramètre de solvatation atomique propre à l'hydrophobicité des atomes,  $\sigma_i$ . Ce dernier coefficient reflète la préférence des types atomiques à être exposés au solvant ou enfouis. Ainsi, pour les atomes polaires et ioniques, plus ils sont exposés au solvant et plus leur contribution à l'énergie  $E_{surf}$  sera négative, donc favorable. On voit que  $E_{surf}$  inclut des effets électrostatiques mais aussi hydrophobes : il est difficile de dissocier complètement les effets polaires et non-polaires.

### 2.2.2 Modèle de Poisson Boltzmann

Le modèle de Poisson-Boltzmann est un des plus utilisés pour les systèmes biomoléculaires, car il a l'avantage de prendre en compte la forme globale de la protéine et aussi l'effet de la force ionique. Ce modèle décrit les variations du potentiel électrostatique dans un milieu diélectrique. L'équation de Poisson a la forme suivante :

$$-\nabla \cdot \varepsilon(x) \nabla \phi(x) = \rho(x) \quad (2.13)$$

où  $\varepsilon(x)$  est un coefficient diélectrique variable,  $\phi(x)$  est le potentiel électrostatique et  $\rho(x)$  est la distribution de charge qui génère  $\phi(x)$ . Le coefficient diélectrique  $\varepsilon(x)$  prend différentes valeurs dans le soluté et dans le solvant, pour refléter la polarisabilité différente des deux milieux. Pour les biomolécules,  $\varepsilon$  prend des valeurs entre 2 et 20 ; une valeur de 80 caractérise le solvant. La distribution de charge  $\rho(x)$  décrit souvent un ensemble de charges atomiques ponctuelles, donc discontinues, ce qui peut poser des difficultés numériques pour

résoudre l'équation de Poisson. En effet, dans ce cas,  $\rho(x)$  est une collection de fonctions de Dirac, qui modélisent les  $N_f$  charges partielles atomiques du soluté :

$$\rho(x) = \rho_f(x) = \sum_i^{N_f} q_i \delta(x - x_i) \quad (2.14)$$

où  $q_i$  est la charge partielle de l'atome i et  $x_i$  sa position. L'équation de Poisson-Boltzmann est une variante de l'équation de Poisson où les charges d'ions mobiles du solvant sont présentes :  $\rho(x) = \rho_f(x) + \rho_{ions}(x)$ . La théorie de Debye-Hückel décrit la distribution de ces ions selon la loi de Boltzmann :

$$\rho_i = \bar{\rho}_i e^{-q_i(\phi(x) + V_i(x))} \quad (2.15)$$

où  $\bar{\rho}_i$  est la concentration volumique des espèces i,  $q_i\phi(x)$  est l'énergie de positionnement d'un ion de charge partielle  $q_i$  à la position x dans le potentiel  $\phi(x)$ , et  $V_i(x)$  est une fonction d'énergie stérique qui empêche les charges mobiles d'entrer à l'intérieur du soluté. Si le potentiel  $\phi(x)$  est partout suffisamment faible, on obtient une simplification de cette équation en développant l'exponentielle en série d'ordre un :

$$-\nabla \cdot \varepsilon(x) \nabla \phi(x) + \bar{\kappa}^2(x) \phi(x) = 4\pi \rho_f \quad (2.16)$$

où  $\kappa$  est le paramètre de Debye-Hückel, qui dépend de la force ionique du solvant.

Une fois que l'équation de Poisson-Boltzmann est résolue, le potentiel électrostatique est connu pour le système entier, et on peut en déduire l'énergie libre électrostatique. En effet, l'énergie libre est une fonction simple du potentiel électrostatique :

$$G = \frac{1}{2} \int \rho_f \phi dx \quad (2.17)$$

Cette expression de l'énergie libre peut-être utilisée pour le calcul de nombreuses propriétés des biomolécules, comme les constantes d'affinité, les pKas et les énergies de solvatation.

### 2.2.2.1 Résolution numérique de l'équation de Poisson

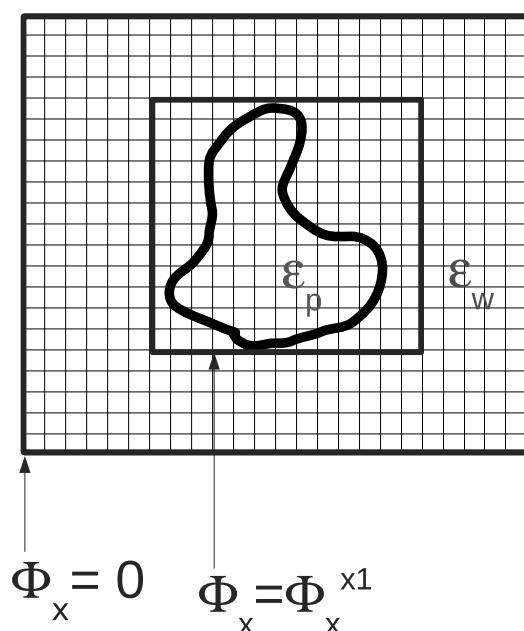
**Principe** Il n'existe pas de résolution analytique de l'équation de Poisson-Boltzmann pour un système moléculaire, car le profil diélectrique autour d'un assemblage d'atomes est complexe. On doit donc la résoudre numériquement. Une méthode très utilisée pour y parvenir est la méthode des différences finies. Une grille tridimensionnelle cubique décompose l'espace en éléments de volume finis [133]. Les charges atomiques sont redistribuées sur les points du maillage. Un maillage cubique de charges est donc obtenu. On dispose maintenant de la charge  $q_x$ , et de la constante diélectrique  $\varepsilon_x$  en chaque point de la grille. Le potentiel électrostatique en chacun des points du maillage est noté  $\phi(x)$ . Les dérivées partielles de l'équation de Poisson sont remplacées par des différences finies. Les six faces k du cube entourant le point de la grille considéré ont une aire de  $l^2$ ; alors l'équation de Poisson prend la forme suivante :

$$\sum_{k=1}^6 \frac{\varepsilon_k (\Phi_k - \Phi_0) l^2}{l} - \kappa_0^2 l^3 = -4\pi q_0 \quad (2.18)$$

Sa résolution consiste à calculer les potentiels  $\Phi_x$  en chacun des points x de la grille, par une méthode itérative.

**Conditions aux limites** Pour définir complètement le problème, il faut fixer les valeurs des  $\phi(x)$  sur la surface extérieure du maillage. La grille étant de grandes dimensions par rapport au soluté, on peut considérer un potentiel nul à toutes les extrémités du maillage, sans que l'erreur induite ne soit pénalisante. Cependant, plus la grille est grande, plus le temps de calcul est important. Une solution couramment utilisée est de calculer successivement des potentiels pour des grilles de plus en plus petites, centrées autour du soluté. A chaque nouvelle grille, les valeurs à la périphérie se voient attribuer celles de la grille précédente qui l'englobe [134] (figure 2.10).

FIGURE 2.10 – Résolution numérique de l'équation de Poisson. L'espace est discréteisé par une grille tridimensionnelle.  $\varepsilon_p$  est la constante diélectrique du soluté ;  $\varepsilon_w$  est la constante diélectrique du solvant et  $\Phi$  est le potentiel électrostatique.



### 2.2.3 Modèle de Born Généralisé

#### 2.2.3.1 Principe

Le modèle de Born généralisé, ou modèle GB, est en fait une approximation de l'équation de Poisson-Boltzmann linéarisée. Dans ce modèle, le soluté est considéré comme un ensemble de charges atomiques partielles situées dans une cavité délimitée par la surface du soluté ; le tout est plongé dans un milieu diélectrique représentant le solvant. L'énergie électrostatique est la somme de deux termes. Le premier décrit les énergies d'interaction coulombienne entre toutes les charges du soluté. Le deuxième est un terme de solvatation  $\Delta E_{solv}$ , qui inclut lui même deux termes : les énergies d'interaction de chaque charge du soluté avec le solvant ( $E_{self}$ ), et une contribution venant de l'écrantage des interactions entre charges du soluté par le solvant ( $E_{int}$ ) [135].

Ainsi,

$$E_{elec} = \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \Delta E_{solv} \quad (2.19)$$

$$\Delta E_{solv} = \sum_i \Delta E_i^{self} + \sum_{i < j} \Delta E_{ij}^{int} \quad (2.20)$$

Les sommes portent sur les atomes i et j du soluté.

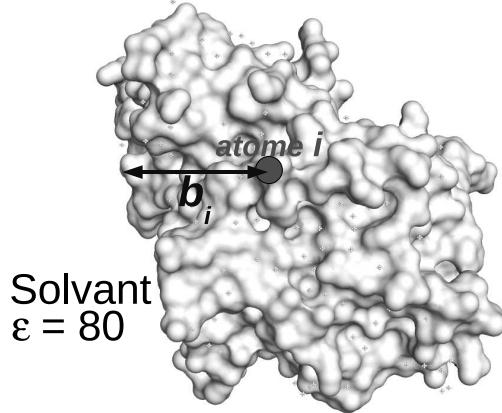
#### 2.2.3.2 Energie self, ou $\Delta E_i^{self}$

L'énergie self représente l'interaction directe d'une charge du soluté avec le solvant. Elle a la forme suivante :

$$\Delta E_i^{self} = -\frac{\theta q_i^2}{2b_i} \quad (2.21)$$

Ici,  $b_i$  est le rayon de solvatation de l'atome i. Pour une charge au centre d'une cavité sphérique,  $b_i$  serait le rayon de la sphère. Dans le cas général, il mesure la distance de l'atome i au solvant. Il dépend donc des positions de tous les autres atomes de la molécule. Ce rayon retranscrit le fait que le solvant est repoussé de l'atome i par les autres atomes de la protéine ; ainsi l'énergie de polarisation du solvant par l'atome i est diminuée (figure 2.11). On voit ici qu'on peut calculer le rayon  $b_i$  à partir de l'énergie self.

FIGURE 2.11 – Le rayon de solvatation pour l'atome i correspond à sa distance au solvant.



### 2.2.3.3 Energie d'interaction, ou $\Delta E_{ij}^{int}$

Dans le modèle GB, la contribution du solvant aux interactions entre charges  $q_i$  et  $q_j$  du soluté est approchée par :

$$\Delta E_{ij}^{int} = \frac{\theta q_i q_j}{(r_{ij}^2 + b_i b_j \exp(-r_{ij}^2/4b_i b_j))^{1/2}} \quad (2.22)$$

où  $r_{ij}$  est la distance entre les charges i et j,  $\theta = -1 + \frac{1}{\varepsilon_w}$ ,  $\varepsilon_w$  est la constante diélectrique du solvant, et  $b_i$  est le "rayon de solvatation" de la charge i. On utilise une valeur de la constante diélectrique de l'ordre de 80.

Quand  $r_{ij} \approx 0$ ,

$$\Delta E_{ij}^{int} \approx \frac{\theta}{2} \frac{q_i q_j}{\sqrt{b_i b_j}} \quad (2.23)$$

Quand  $r_{ij}$  est très grand,

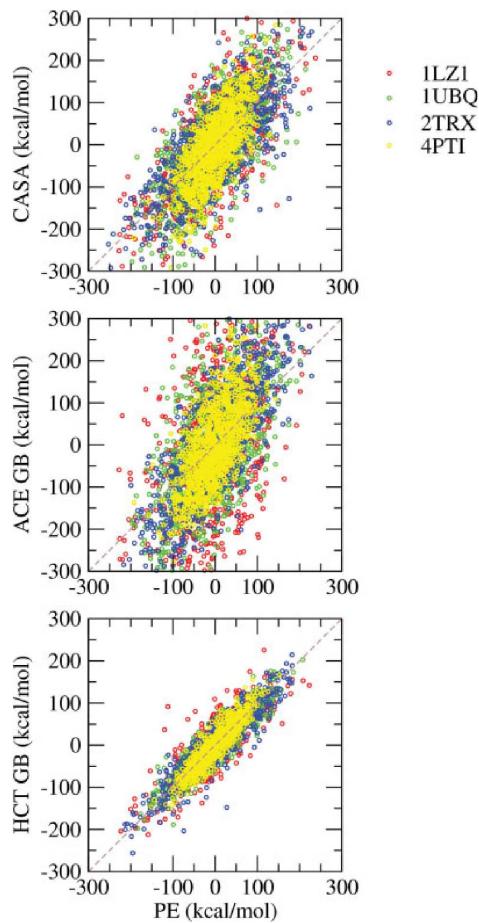
$$\Delta E_{ij}^{int} \approx \frac{\theta}{2} \frac{q_i q_j}{r_{ij}} \quad (2.24)$$

### 2.2.4 Comparaisons de ces modèles

Pour évaluer les solvants CASA et GB, Lopes et coll. ont calculé l'énergie de solvatation de quatre protéines : l'inhibiteur anti-trypsine (4PTI), la thioredoxin (2TRX), le lysozyme

(1LZ1) et l'ubiquitine (1UBQ). Pour chaque protéine, mille structures sont considérées. Les méthodes CASA et GB sont comparées à la méthode "PE", qui correspond à la résolution numérique de l'équation de Poisson, et qui est la plus précise. La corrélation entre CASA et PE montre que CASA reproduit approximativement le modèle de Poisson. Mais la meilleure corrélation est obtenue entre les modèles PE et GB/HCT. La déviation RMS entre ces deux méthodes est faible ( $21\text{-}45 \text{ kcal.mol}^{-1}$ ) [136] (figure 2.12).

FIGURE 2.12 – Energie de solvatation de quatre protéines estimées avec les modèles GB, CASA, et Poisson. On distingue deux variantes de GB : ACE et HCT. Le modèle de Poisson (PE) est pris comme référence. En bleu : 2TRX (thioredoxine). En vert : 1LZ1 (lysozyme). En rouge : 4PTI (BPTI). En jaune : 1UBQ (ubiquitine). Données reproduites à partir de Lopes et coll. (2007) [136].



## 2.3 L'espace conformationnel

Pour un squelette peptidique donné, la séquence d'acides aminés la plus favorable correspond à celle qui maximisera la différence d'énergie libre entre l'état replié et l'état déplié. Pour identifier cette séquence, il faudra donc modéliser ces deux états. Pour que la complexité soit acceptable, on va adopter dans les deux cas des descriptions simplifiées.

### 2.3.1 Représentation de l'état replié

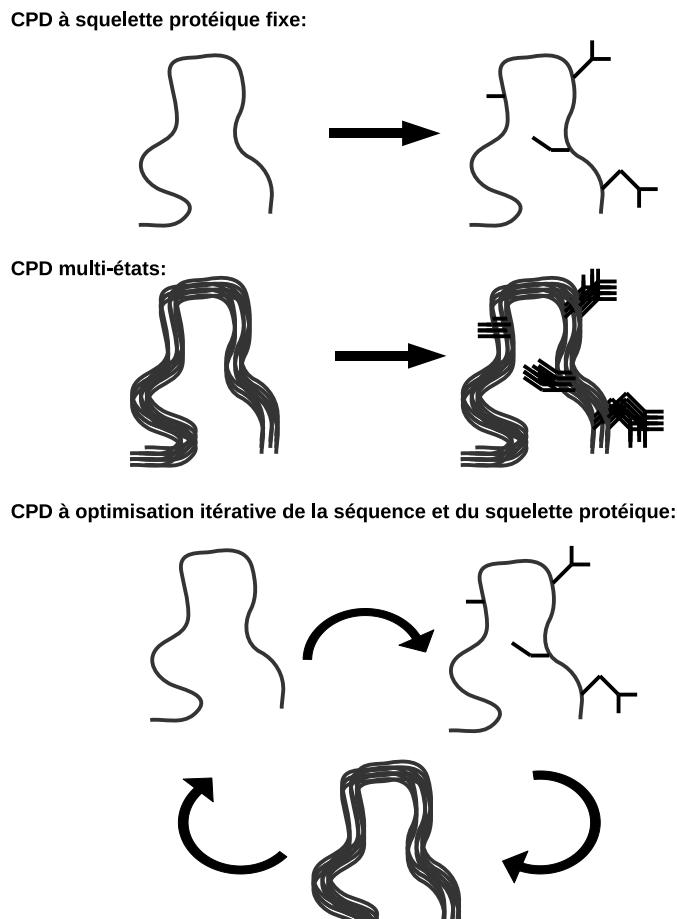
Soit une protéine de  $n$  résidus qu'on peut muter à toutes ses positions par tous les types d'acides aminés. Le nombre de séquences possibles atteint  $20^n$ . De plus, pour chacune, il peut y avoir différentes conformations des chaînes principales et latérales. Pour réduire la complexité de l'espace conformationnel, il sera généralement discréétisé.

#### 2.3.1.1 Discréétisation du squelette peptidique

Dans beaucoup d'applications, la chaîne principale est considérée comme fixe. Il y a plusieurs limitations à cette méthode. La principale est que certaines chaînes latérales peuvent être considérées comme défavorables d'un point de vue de l'énergie, alors qu'une légère adaptation du squelette aurait diminué l'énergie significativement. Une autre approche consiste à générer un ensemble de squelettes et à faire des calculs CPD pour chacun d'eux [137]. Un algorithme de design multi-état, reposant sur la méthode FASTER, a été testé dans des modélisations prenant en compte 128 squelettes protéiques. Dans cet exemple, l'usage d'un design multi-état avec un ensemble de squelettes protéiques, définis à l'avance par dynamique moléculaire, a conduit à une amélioration des résultats dans des tests de stabilité de protéines [138]. Fu et coll. ont réalisé le design de peptides hélicoïdaux à travers un ensemble de squelettes protéiques générés par analyse en mode normaux d'hélices de la PDB ciblant la reconnaissance au ligand Bcl-xL. Plusieurs peptides issus du design se sont montrés sélectifs pour l'interaction avec ce ligand [139]. Une troisième approche consiste à optimiser la conformation du squelette pour un grand nombre de séquences fixes [140]. Enfin, Desjarlais et Handel ont généralisé celà en autorisant des angles  $\phi$ ,  $\psi$  et  $\Omega$  variables dans le squelette protéique. Des vérifications expérimentales

sur des variants de stabilité du lysozyme T4 ont été réalisés [141]. Des algorithmes de Monte Carlo ont également été utilisés pour modéliser la flexibilité de boucles à la surface de la protéine [142] (figure 2.13).

FIGURE 2.13 – Différentes approches pour le traitement du backbone, ou squelette en CPD. Le design "multi-état" fait l'étude d'une séquence sur plusieurs squelettes protéiques. Dans la méthode itérative, il y a une alternance entre les étapes où la séquence d'acide aminé varie et les étapes d'optimisation du squelette protéique, où la meilleure conformation est recherchée pour une séquence donnée.



### 2.3.1.2 Discrétilisations des chaînes latérales

Pour les chaînes latérales, des angles de torsions décrivent la rotation des groupes chimiques autour des liaisons, et l'espace conformationnel des chaînes latérales peut être décrit par ces angles. Les conformations préférentielles constituent les "rotamères". Ce terme a été introduit par Janin et coll. en 1978 [143]. La majorité des méthodes de prédiction de structure reposent sur ce concept de rotamères. On compte en moyenne 10 rotamères

par acide aminé. Cependant, environ 5% des chaînes latérales ne se trouveraient pas dans ces conformations, et seraient mal représentées par les rotamères. Celà introduit une légère erreur dans la méthode. De nombreuses librairies de rotamères ont été construites et détaillées, avec validation statistique à l'appui [143, 144]. La plupart de ces librairies sont indépendantes de la conformation du squelette peptidique. La structure secondaire ou la conformation locale du squelette protéique ne sont donc pas pris en compte. Les premières librairies de rotamères ont été construites à partir d'un petit nombre de structures. Même la plus grande bibliothèque de Ponder et Richards n'est basée que sur 19 structures. Les librairies les plus récentes sont basées sur plus de 600 structures, avec une résolution de 1.8 Å au moins et une identité de séquence inférieure à 50 % entre les protéines utilisées. La librairie de Tufféry repose sur 53 structures, ce qui reste un petit échantillon par rapport au grand nombre de structures actuellement disponibles dans la PDB. Kono et Doi ont publié une librairie de rotamères basée sur l'analyse de 103 structures [145]. Richardson et coll. ont fait de même avec des protéines d'une résolution de 1.7 Å et plus, en excluant les résidus à facteur de température cristallographique élevé ou avec des recouvrements stériques défavorables [146]. Ce procédé restreint la taille de l'échantillon mais la répartition en cluster des angles dièdres est plus nette.

Plusieurs librairies dépendantes du squelette peptidique ont été proposées [147, 148, 147]. Mc Gregor et coll. et Schrauber et coll. ont étudié la probabilité des rotamères ainsi que les moyennes d'angles dièdres dans différentes structures secondaires. Une bibliothèque de rotamères dépendant du squelette protéique à été construite ; elle repose sur les statistiques bayesiennes. On dispose ici d'une moyenne des angles dièdres  $\phi$  et  $\psi$  du squelette protéique par pas de 10 degrés [148, 149]. La version courante de cette bibliothèque est basée sur 699 structures, de résolution supérieure à 1.8 Å et avec moins de 50% d'identité entre les séquences. Certaines études récentes utilisent des librairies avec des milliers de rotamères, pour couvrir l'espace conformationnel avec une grande résolution [150, 151].

### 2.3.2 Représentation de l'état déplié

L'état déplié est un état déstructuré, par conséquent, il est difficile à modéliser. Cette déstructuration conduit à l'exposition au solvant de résidus hydrophobes. Le modèle le plus simple assimile l'état déplié à une chaîne étendue. Les acides aminés interagiraient donc principalement avec le solvant et avec les groupes voisins du squelette peptidique. En revanche, les chaînes latérales interagiraient très peu entre elles. Dans ce modèle, on considère donc que l'énergie libre de l'état déplié est uniquement dépendante de la composition en acides aminés et non de la séquence. Cette représentation est modélisée en pratique par une collection de tripeptides Ala-X-Ala, où X est l'acide aminé courant d'une position donnée [152]. L'état déplié peut aussi être modélisé par une chaîne aléatoire, pour mieux incorporer les effets électrostatiques à longue portée [153]. L'état dénaturé peut également être traité explicitement, en utilisant les mêmes algorithmes que pour l'état replié, ce qui est beaucoup plus coûteux.

## 2.4 Algorithmes d'optimisation appliqués au CPD

Le design de séquences protéiques, pour qu'elles adoptent un repliement particulier ou une fonction différente, implique l'exploration d'un grand nombre de séquences. Les algorithmes d'optimisation décrivent la manière dont le programme va échantillonner l'immense espace de solutions possibles. Tous les algorithmes de recherche font nécessairement des compromis entre vitesse de calcul et perfection. De plus, il y a une grande influence sur les résultats du choix de l'algorithme, de la représentation de l'espace de recherche et de la fonction d'énergie, ou de score.

Dans la plupart des méthodes de CPD, un jeu discret d'états des acides aminés et d'orientations de leurs chaînes latérales (rotamères) est optimisé par rapport à une fonction de score. Le but est alors de trouver la combinaison sequence/rotamères qui donne le meilleur score. Les algorithmes de recherche peuvent être divisés en deux catégories. La première catégorie échantillonne semi-aléatoirement l'espace des configurations spatiales et converge vers une solution possible dépendant du profil énergétique et des règles de l'al-

gorithme lui même. La méthodes de Monte Carlo et les algorithmes génétiques font partie de cette catégorie. Ces méthodes présentent l'avantage de pouvoir être appliquées à des problèmes qui ont une infinité de solutions possibles. Par exemple, les conformations des chaînes latérales et du squelette peptidique peuvent varier de façon continue [154]. Mais, elles ne garantissent pas l'exploration d'une solution proche du minimum global d'énergie. A l'opposé, les algorithmes de la deuxième catégorie vont rechercher le minimum global d'énergie par une méthode déterministe, comparable à la recherche exhaustive. Les recherches exhaustives n'étant possibles que pour de petits espaces conformationnels, cette classe d'algorithme va simplifier l'espace des conformations en le discrétilisant (cf. ci-dessus). On emploie ensuite des critères d'élimination pour écarter la plus grande partie des conformations du champ d'exploration. Le Dead End Elimination est un bon exemple de ce type d'algorithme. Ces deux classes de méthodes sont détaillées dans la suite.

## 2.4.1 Méthodes stochastiques

### 2.4.1.1 Méthode de Monte Carlo

Les algorithmes stochastiques peuvent être utilisés pour changer l'identité des chaînes latérales, leurs orientations, ou encore la structure du squelette peptidique. La procédure la plus simple de ce type est la méthode de Monte Carlo. La stratégie générale de cet algorithme est de modifier itérativement un modèle, et de décider alors si la modification est conservée ou rejetée. Parmi les différentes méthodes de Monte-Carlo existantes, la plus utilisée est l'algorithme de Métropolis :

1. une nouvelle conformation est générée à partir de la conformation courante en modifiant aléatoirement une variable.
2. l'énergie potentielle de la nouvelle conformation est calculée.
3. si l'énergie  $E_n$  de la nouvelle conformation est plus petite que l'énergie courante  $E_c$ , alors la nouvelle conformation est acceptée (i.e., elle devient la conformation courante). Sinon, un facteur de Boltzmann est calculé :  $\exp(-E_c/kT)$ , et un nombre aléatoire compris entre 0 et 1 est tiré et comparé avec ce facteur de Boltzmann. Si le nombre aléatoire est plus grand que le facteur de Boltzmann, la nouvelle conformation est rejetée (i.e.,

on garde la conformation courante) ; sinon, elle est acceptée et devient la conformation courante.

4. on recommence au pas 1.

Le grand succès de cette méthode en mécanique statistique a mené à diverses applications, comme la méthode de recuit simulé pour l'optimisation des structures, dans laquelle la température est initialement élevée puis graduellement diminuée, pour trouver un minimum local d'énergie.

#### **2.4.1.2 Algorithme génétique**

Les algorithmes génétiques sont similaires. La distinction majeure est que la population des modèles évolue au cours de la simulation par générations successives, à l'aide d'opérateurs génétiques, tels que les recombinaisons, utilisées pour créer des modèles différents de ceux existant chez les parents. L'efficacité des algorithmes génétiques repose sur le parallélisme implicite des problèmes de protein design. Différents segments de la structure sont optimisés en parallèle et des recombinaisons sélectives entre les modèles peuvent ensuite réunir plusieurs segments optimisés dans un même modèle.

#### **2.4.1.3 Algorithme heuristique de Wernisch**

En 2000, Wernisch, Hery et Wodak, présentent une procédure automatique pour la prédiction de séquences d'acides aminés compatibles avec une structure cible donnée [155]. Une procédure heuristique d'optimisation est utilisée pour identifier les séquences de haut score. La méthode de Wernisch est itérative ; chaque itération est appelée un cycle heuristique. Un cycle heuristique se déroule comme suit : une séquence d'acides aminés initiale est choisie au hasard, puis optimisée par étapes consécutives. A la position d'acide aminé 1, le meilleur acide aminé et son rotamère sont sélectionnés (le reste de la séquence étant fixe). La même chose est faite pour les résidus suivants : 2, 3, ..., n. On recommence à la position 1 de nouveau, et on effectue des passages multiples sur la séquence d'acides aminés, jusqu'à ce que l'énergie ne s'améliore plus (convergence). Alors, la séquence finale, le jeu de rotamères et l'énergie sont relevés et le cycle est terminé. Pour une protéine donnée,

on réalise un grand nombre de cycles heuristiques, entre  $10^3$  et  $10^6$ , afin de recueillir les solutions d'énergies les plus basses.

## 2.4.2 Méthodes deterministes

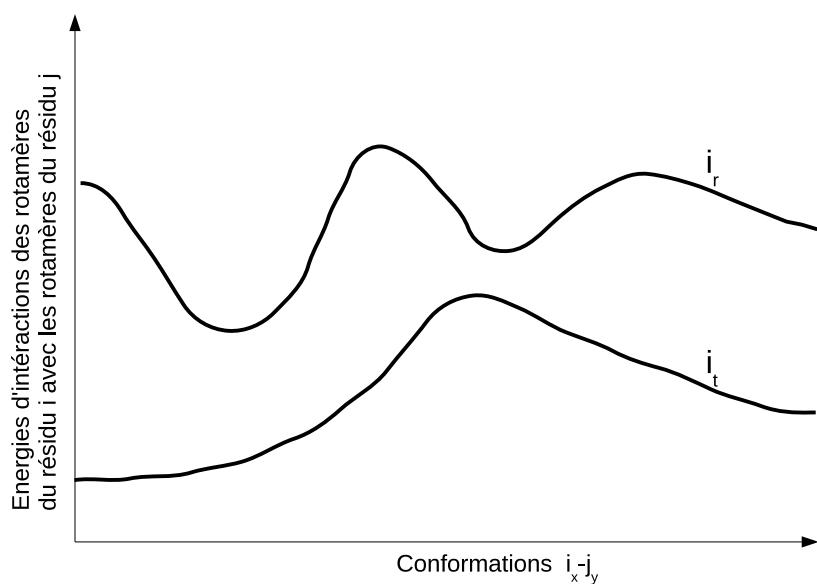
### 2.4.2.1 Dead End Elimination (DEE)

DEE est la méthode la plus efficace pour l'optimisation dans un espace discret, en raison de sa capacité à réduire énormément la complexité combinatoire en utilisant une méthode robuste d'élimination des solutions peu probables. L'implémentation de la DEE requiert une représentation discrète du squelette peptidique et des chaînes latérales. Son utilisation est réduite aux problèmes dont les termes d'énergie peuvent être réduits à la somme de termes d'énergie de paires individuelles. La DEE fait une étude individuelle de chaque chaîne latérale  $i$ , en considérant tous ses types et rotamères possibles. Pour chaque type et rotamère, il est trouvé si cet état de la chaîne latérale est incompatible avec le minimum d'énergie global. Dans sa forme initiale, la DEE utilise le critère suivant : si la structure d'énergie la plus basse trouvée avec un rotamère donné a une énergie plus grande que la plus grande énergie trouvée avec un autre rotamère, alors le premier rotamère peut être éliminé [156] (figure 2.14). Pour appliquer ce critère, on doit analyser les termes d'énergie d'interaction entre  $i$  et tous les autres acides aminés  $j$ . Une réduction encore plus importante des rotamères possibles est atteinte en utilisant la variante de Goldstein : si l'énergie d'une structure possible contenant un rotamère est toujours réduite en le remplaçant par un autre rotamère, alors le premier rotamère peut être éliminé [157]. Pour quelques problèmes simples, l'application de la DDE peut donner une solution unique [158, 152]. La DEE continue d'être couramment utilisée en CPD.

### 2.4.2.2 Champ moyen

Dans cette méthode, chaque rotamère d'une position de la séquence intérage avec tous les rotamères des chaînes latérales qui l'entourent, pondérés cependant par leur probabilité respective. Cet algorithme résume ainsi en une interaction moyenne toutes les interactions possibles entre un rotamère donné et tous les autres rotamères de la protéine. On calcule

FIGURE 2.14 – Dead End Elimination. Les courbes représentent les énergies d’interactions entre deux résidus  $i$  et  $j$ . Ici, l’acide aminé  $i$  peut avoir deux conformations (rotamère  $r$  ou  $t$ ). Le résidu  $j$  peut également prendre différentes conformations. On pose l’équation suivante :  $E(i_r) - E(i_t) + \sum_j \min [E(i_r, j_s) - E(i_t, j_s)] > 0 \quad i \neq j$ . Sur le graphique, nous voyons que l’énergie minimale d’interaction d’ $i_r$  avec les rotamères de  $j$  est supérieure à l’énergie maximale d’interaction d’ $i_t$  avec les rotamères de  $j$ . Ainsi, le rotamère  $i_r$  est éliminé par rapport au rotamère  $i_t$ .



donc toutes les probabilités de Boltzmann  $P(i,a)$  du rotamère a à la position i, à partir de son énergie  $E(i,a)$  :

$$P_{i,a} = \frac{e^{-\frac{E(i,a)}{kT}}}{\sum_{a=1}^{N_i} e^{-\frac{E(i,a)}{kT}}} \quad (2.25)$$

où k est la constante de Boltzmann et T la température.  $E(i,a)$  est l'énergie du rotamère a à la position i, qui peut s'écrire :

$$E(i,a) = E_{BB}(i,a) + \sum_{j \neq i} \sum_l E(ia,jl) P(j,l) \quad (2.26)$$

$E_{BB}$  représente l'énergie d'interaction du rotamère avec le squelette protéique. La double somme représente l'énergie d'interaction entre le rotamère a à la position i et tous les rotamères l à toutes les autres positions j, pondérés par leur probabilité  $P(j,l)$ . On déduit alors de nouvelles probabilités de Boltzmann pour chaque rotamère a à toutes les positions i. Ce protocole est répété jusqu'à ce que les probabilités et les énergies convergent. Pour arriver à cette convergence, on tient compte du cycle précédent pour le recalculation des probabilités :

$$P(i,a)^{(n+1)} = \lambda P(i,a)^{(n)} + (1 - \lambda) P(i,a)^{(n-1)} \quad (2.27)$$

Le calcul du champ moyen aboutit à un état de la structure où chaque rotamère est le plus stable pour la position donnée, compte tenu de son environnement "moyen".

## 2.5 Dynamique moléculaire

Nous étudions jusqu'ici des structures assez figées, à backbone fixe, et entourées de solvant implicite. C'est pourquoi nos prédictions structurales à l'issu du CPD seront ensuite étudiées au cours de simulations de dynamique moléculaire, avec solvant explicite. Nous étudierons la trajectoire d'une protéine en simulant les mouvements atomiques au cours du temps. Dans cette méthode, chaque atome de la molécule est considéré comme une masse ponctuelle dont le mouvement est déterminé par l'ensemble des forces exercées sur lui par les autres atomes en fonction du temps.

**Algorithme** Au temps t, chaque atome i, de coordonnées  $x_i, y_i, z_i$  et de masse  $m_i$  subit une force  $\vec{F}_i$  et possède une accélération  $\vec{\gamma}_i$  telles que :

$$\vec{F}_i = m_i \vec{\gamma}_i \quad (2.28)$$

La force  $\vec{F}_i$  est déterminée par dérivation de la fonction énergie potentielle E.

$$\vec{F}_i = -\text{grad}_i E \quad (2.29)$$

Pour une composante de  $\vec{F}_i$ , par exemple  $F_{ix}$ , cela signifie que :

$$F_{ix} = -\frac{\partial}{\partial x_i} E(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n) \quad (2.30)$$

On peut alors intégrer les équations de mouvement et obtenir une trajectoire pour chaque atome en fonction du temps. L'intégration se fait en subdivisant la trajectoire en une série d'états discrets, séparés par des intervalles de temps très courts, dont la longueur  $\Delta t$  définit le pas d'intégration. Le déplacement d'un atome pendant l'intervalle de temps  $\Delta t$  est relié à la vitesse  $\vec{v}_i(t)$  à l'instant t :

$$\vec{v}_i(t) \Delta t = \vec{r}_i(t + \Delta t) - \vec{r}_i(t) \quad (2.31)$$

De même, il est possible de déterminer la vitesse de chaque atome à partir de l'accélération :

$$\vec{v}_i(t + \Delta t) - \vec{v}_i(t) = \vec{\gamma}_i(t) \Delta t \quad (2.32)$$

La détermination de la vitesse permet d'obtenir la position de l'atome à l'instant  $t + \Delta t$ . La répétition de cette procédure à des intervalles de temps discrets aboutit à l'identification de la trajectoire. Pour que l'énergie totale du système soit conservée durant la simulation, l'intervalle de temps  $\Delta t$  doit être choisi petit (typiquement égal à une femtoseconde). Une nanoseconde de dynamique correspond ainsi à environ  $10^6$  pas.

### **2.5.0.3 Température**

Pour être réaliste, la simulation doit refléter une température physiologique. Or, la température d'une molécule en agitation thermique est proportionnelle à l'énergie cinétique de ses atomes. Pour contrôler la température, la vitesse est donc calculée pour chaque atome, puis ajustée périodiquement pour maintenir une énergie cinétique moyenne et une température constante, proche de la température ambiante.

### **2.5.0.4 Modélisation explicite du solvant**

Le modèle le plus simple considère les trois atomes de la molécule d'eau, soit trois "sites" par molécule. Chaque atome a une charge, et l'oxygène a également un terme d'énergie de Lennard-Jones. Il existe différents modèles à trois sites, dont le modèle TIP3P et le modèle SPC/E. Des modèles à quatre sites introduisent un quatrième pseudoatome, qui représente la charge négative de l'oxygène. Ceci permet d'améliorer la distribution de charge moléculaire. Dans le modèle à cinq sites TIP5P, deux atomes fictifs représentent les doublets non liants de l'oxygène. Ceci améliore la modélisation de la géométrie des dimères de molécules d'eau. Les modèles à six sites prennent en compte tous les atomes et pseudoatomes cités précédemment et sont majoritairement utilisés pour les modélisations dans la glace. Les modèles à trois ou quatre sites sont couramment utilisés pour la dynamique moléculaire de protéines, car les autres sont jugés trop coûteux.

## **2.6 Quelques applications du Protein Design**

Les récents succès du CPD ont prouvé son efficacité pour modifier les protéines naturelles, mais également pour en concevoir de nouvelles. Des structures et des séquences de protéines ayant plus de cent résidus variables ont été réalisées. Les structures d'un grand nombre de protéines issues du CPD ont été résolues expérimentalement.

### **2.6.1 Etude du repliement**

Lumb et Kim soulignèrent l'importance des résidus polaires enfouis dans l'obtention d'une structure unique par l'étude de deux peptides de synthèse [159]. Ces deux peptides se replient en hélices formant un hétérodimère coiled-coil parallèle. Tous les résidus à l'interface sont apolaires à l'exception de deux arginines, situées chacune sur une des hélices et formant une liaison hydrogène entre elles. Quand ces deux arginines sont mutées en résidus apolaires, la structure s'en voit modifiée. Il n'y a plus d'orientation préférentielle des hélices, et celles-ci s'assemblent en tétramère plutôt qu'en dimère. Les interactions électrostatiques sont donc importantes dans la thermodynamique de formation de ce complexe protéine-protéine.

Malgré le grand nombre de termes nécessaires pour avoir une fonction d'énergie précise, des fonctions plus simples reposant sur le potentiel de van der Waals seul ont suffit pour le design de cœurs protéiques. En effet, le cœur est la partie de la protéine la plus simple à décrire d'un point de vue de l'énergie, surtout si on restreint celle-ci uniquement aux résidus hydrophobes [160]. Pour ces cœurs protéiques, des prédictions de structures reposant uniquement sur le terme de van der Waals ont donné des structures très stables [152, 161, 162].

### **2.6.2 Protéines entières, mini-protéines**

#### **2.6.2.1 *De novo* design**

Le *de novo* design correspond au design de séquences et repliements complètement nouveaux. L'un des premiers exemples est un motif  $\beta\alpha\beta$  mimant le module de liaison à l'ADN en doigt de zinc [163]. La molécule créée se replie de façon stable sans l'ion métallique  $Zn^{2+}$  de la structure d'origine. Une  $\alpha\beta$ -protéine de 97 résidus a également été créée avec un repliement non naturel [164]. Sa structure a été résolue expérimentalement. D'autres protéines contenant des assemblages d'hélices en coiled-coil ont été construites [140].

### 2.6.2.2 Design de mini-protéines

Des miniprotéines naturelles comme les toxines ou les inhibiteurs de protéases peuvent avoir une forte activité et grande spécificité en utilisant un motif structural minimal mais stable de moins de 80 acides aminés. Des activités artificielles ont été introduites dans ces structures protéiques en transférant un site actif identifié au préalable dans des régions compatibles, avec introduction des mutations nécessaires. Ces nouvelles mini-proteines présentent une activité spécifique et forte dans une structure tridimensionnelle bien définie et de petite taille. Ceci crée un nouvel outil en biotechnologie. Un premier exemple est la création d'une miniprotéine liant les ions métalliques (MBP, 37 résidus) [165]. Un deuxième exemple est une réplique de CD4 (CD4M33, 27 résidus), qui est un inhibiteur puissant de l'entrée du virus HIV-1 dans les cellules, et qui constitue également une réplique fonctionnelle du récepteur CD4 [166].

### 2.6.3 Interaction protéine-protéine

Schreiber et coll. étudièrent l'interface Barnase-Barstar par mutations ponctuelles et cycles de doubles mutants. Ils ont mis en évidence la relation entre les résidus chargés à l'interface et l'énergie libre de liaison du complexe. Un phénomène de couplage d'énergie des paires de résidus chargés a été mis en évidence [167]. Xu et coll. ont montré que les paires de résidus chargés contribuent plus que les interactions hydrophobes pour la formation d'un complexe protéique; cette étude a été menée sur plusieurs complexes protéine-protéine [168]. Au cours de la formation d'un complexe, les résidus chargés se trouvant sur la surface d'interaction doivent rompre les interactions avec le solvant pour permettre la liaison au ligand, avec lequel ils forment de nouvelles interactions. L'énergie de liaison est déterminée par la balance entre ces deux contributions énergétiques. On voit donc de nouveau que la solvatation et l'électrostatique sont intimement reliés [169].

## **2.6.4 Interaction protéine-Ligand**

### **2.6.4.1 Affinité**

Pour obtenir une liaison protéine-ligand efficace en CPD, la précision de la fonction d'énergie est un facteur critique. L'affinité peut être modifiée de plusieurs ordres de grandeur, ce qui rend le calcul de chaque kcal/mol important. Le redesign de l'affinité du nanomolaire au picomolaire est un défi pour le CPD. Plusieurs groupes ont fait des progrès concernant le re-design de l'affinité protéine-peptide ou protéine-protéine. Song et coll. ont amélioré l'interaction ICAM-I/LFA-I [170]. Cette interaction est de faible affinité. L'étude s'est basée sur des techniques variées. Les résultats en sont une augmentation de l'affinité d'un facteur 20, en combinant plusieurs mutations ponctuelles obtenues par le design. Sood et Baker ont mis au point des extensions N-ter ou C-ter pour augmenter l'interaction protéine-peptide, par une technique qui combine une recherche de conformations du squelette protéique et des chaînes latérales [171]. Leurs résultats sont modestes ; il y a un gain d'affinité d'un facteur 1.3 à 2.3 pour leurs deux systèmes tests. Van Vlijmen et coll. parviennent à un gain d'affinité d'un facteur 8 en combinant les effets de quatres mutations ponctuelles obtenues par des techniques de design variées. Ils ont un taux de succès de 12% sur 83 mutants construits [172]. Dahiyat et coll. ont modifié l'interaction FC/FC $\gamma$ R, ils ont pu augmenter l'affinité d'un facteur 100 [173]. Degrado et coll. ont développé un protocole pour créer une nouvelle protéine capable de lier un cofacteur spécifique donné. Ils ont créé un peptide composé de quatres hélices, liées à un cofacteur métalloporphyrine [174]. Yang et coll. ont réalisé le design d'un site fixant le calcium dans la protéine de liaison cellulaire CD2, avec une approche qui évalue le potentiel de liaison en fonction de la comptabilité géométrique [175]. Tous ces résultats illustrent les difficultés, qui restent considérables en CPD.

### **2.6.4.2 Spécificité**

Dans certains cas, le design consiste à créer une nouvelle reconnaissance, mais en évitant en même temps la reconnaissance d'autres ligands, non souhaités. Sauer et coll. on utilisé une approche basée sur la spécificité pour le redesign d'une protéine adaptatrice

homodimérique, pour la rendre hétérodimérique [176]. Le design sur le seul critère de stabilité n'a pas atteint la spécificité hétérodimérique recherchée, mais à abouti à des structures plus stables que le type sauvage. Le design sur la spécificité à abouti préférentiellement à des structures hétérodimériques, mais moins stables que le type sauvage. Baker et coll. ont exploré deux stratégies pour le design d'un switch de spécificité à une interface protéine-protéine : le complexe colicine E7:DNase-Im7 [177]. Un premier design a d'abord été fait, puis une des 11 structures obtenues a été cristallisée et a servi de base pour un deuxième tour de design. Les auteurs ont réalisé à la fois une stratégie de design positif et une stratégie de design négatif. Leur protocole de design d'affinité a mené à un switch de spécificité de l'ordre de 300 en faveur de l'interaction non naturelle. En dépit du nouveau réseau de liaisons hydrogène crée, la grande partie du gain de spécificité est due à une simple mutation hydrophobe. Le design négatif a été utilisé par Baker et coll. pour modifier la spécificité d'une endonucléase [178], par Keating et coll. pour convertir un homotétramère en hétérotétramère [179], et par Jasanoff et coll. pour altérer la spécificité de la calmoduline [180]. Toutes ces données montrent la nécessité d'intégrer le design négatif pour obtenir une spécificité d'interaction protéine-protéine.

Le design de protéines *de novo* doit conduire à une séquence d'acides aminés ayant un repliement unique et spécifique. Cependant Ambroggio et Kuhlman ont identifié une séquence d'acides aminés qui peut adopter deux plis bien distincts, en fonction des conditions de pH et de la présence de métaux de transition [181]. Ils ont utilisé un design positif sur la stabilité, simultanément pour les deux états. La nouvelle protéine peut passer d'un repliement en doigt de zinc à un coiled-coil trimérique selon les conditions. La protéine à cependant tendance à former des agrégats, à cause de résidus hydrophobes exposés. Ce dernier fait illustre encore une fois la nécessité de faire un design négatif pour éviter les interactions non désirées.

#### 2.6.4.3 Design d'enzymes

Hellinga et coll. en 1991 ont mis au point un programme de modélisation moléculaire qui construit de nouveaux sites catalytiques dans une protéine de structure tridimen-

sionnelle connue [182]. Ce programme n'altère que la séquence d'acides aminés ainsi que la structure des chaînes latérales, laissant le repliement de la protéine intact. Le programme place des chaines latérales de manière qu'elles lient le ligand, en répondant à des contraintes imposées au préalable. Les sites de liaison sont introduits en respectant des règles simples, comme l'encombrement stérique et les réseaux de liaisons hydrogènes connus. Un objectif était l'introduction de site de liaison du cuivre, comme ceux de la plastocyanine, l'azurine ou la cupredoxine au sein de la thioredoxine d'*Escherichia coli*.

Un autre programme de modélisation a été crée en 1995 par Clarke et coll. Ce programme assiste le design de sites de liaison tétravalent de métaux dans des protéines de structures connues [172]. Ce programme utilise les coordonnées du squelette protéique pour proposer quatre positions, susceptibles de former un site de liaison tétrahédrique si leurs chaînes latérales sont remplacées par la cystéine ou l'histidine. Les angles dièdres pour ces chaînes latérales, ainsi que les coordonnées du métal sont également prédits. Ce programme a été utilisé pour l'introduction d'un site de liaison au zinc dans une protéine de synthèse composée de quatre hélices  $\alpha$ , ainsi que dans le domaine B1 de la protéine G du streptocoque. Pour ces deux tests, la liaison d'un ion métallique tétravalent a été confirmé [183, 184].

Les mécanismes de fonctionnement des enzymes sont complexes, ce qui rend les objectifs du design d'enzyme encore plus compliqués à cerner. Les facteurs qui peuvent être importants sont, entre autres : la liaison à l'état de transition, l'accommodation du substrat, le relargage du produit, la flexibilité et la dynamique de la protéine. Chakrabarti et coll. ont pu retrouver la majorité des séquences du type sauvage d'un enzyme en optimisant l'affinité avec son substrat, tout en imposant des contraintes géométriques sur les conformations des chaînes latérales des résidus catalytiques [185]. Mayo et coll. ont exploré le redesign d'enzymes en stabilisant l'état de transition et en ne mutant que des résidus interagissant indirectement avec le ligand. Pour le design de la chorismate mutase d'*Escherichia coli*, une des cinq mutations prédictes augmente l'efficacité de 60% [186]. Le redesign de l'imipenemase IMP-1 prédit une mutation qui enlève un groupe hydroxyle, et une double mutation qui déplace ce même groupe hydroxyle [187]. Le déplacement

du groupe hydroxyle altère la spécificité du substrat ; la présence des deux groupements hydroxyles augmente l'efficacité de la catalyse.

Pour le design d'enzymes *de novo*, la recherche de conformations des résidus du site actif, couplé au placement d'une petite molécule en corps rigide, complique les calculs. Baker et coll. ont développé une méthode pour le placement d'un site actif prédéfini. Ils ont développé une routine pour dix réactions chimiques données [188]. Leur procédure cherche une protéine candidate pour le site actif, puis optimise le reste de la protéine pour améliorer la liaison à l'état de transition. Mayo et coll. ont également développé une méthode pour le placement de petites molécules dans le design d'enzyme. L'incorporation de ces molécules se fait par rotations et translations, en ajoutant un biais énergétique pour favoriser le contact du ligand et des chaînes latérales [189]. Toutes ces applications montrent que l'utilisation du CPD reste un défi.



# Chapitre 3

## Paramétrisation et validation du modèle de solvant CASA

Dans les prochains chapitres, nous décrirons une application du CPD à la tyrosyl-ARNt synthétase et ses interactions avec la D-Tyrosine. Mais dans un premier temps de l'étude, une étape de validation de la fonction d'énergie est nécessaire. L'optimisation des paramètres de solvatation et des énergies de références pour les tripeptides modèles de l'état déplié a été réalisée. Puis ces paramètres ont été testés sur des calculs d'énergies libres de peptides. En confrontant les données expérimentales aux données calculées, il est possible d'évaluer la pertinence du modèle. Enfin, le protocole de calcul d'affinité a été également validé en comparant des énergies de liaison expérimentales et calculées pour les complexes protéine-ligand, d'une part, et protéine-protéine d'autre part. Le travail de ce chapitre a été fait en étroite collaboration avec Christine Bathelt et Marcel Schmidt am Busch.

### 3.1 Paramétrisation du modèle

#### 3.1.1 Forme de la fonction d'énergie

Nous employons une fonction d'énergie issue de la mécanique moléculaire pour décrire les interactions intra-protéine. Le champ de force correspondant est Charmm19, de

Karplus et coll. [105]. Le modèle de solvant implicite est le modèle CASA : Coulombic Accessible Surface Area. L'énergie de solvatation est exprimée par  $E_{solv}$ , somme d'un terme d'écrantage électrique des interactions du soluté par les dipôles du solvant et d'un terme qui dépend de la surface accessible au solvant :

$$E_{solv} = E_{screen} + E_{surf} = \left(\frac{1}{\varepsilon} - 1\right)E_{coul} + \alpha \sum_{i=1}^n A_i \sigma_i \quad (3.1)$$

$A_i$ , est la surface accessible au solvant de l'atome i,  $\alpha$  est une pondération globale pour le terme de surface, et  $\sigma_i$  un paramètre de solvatation atomique qui dépend de l'hydrophobicité de l'atome i. Ce dernier coefficient reflète la préférence des différents types atomiques à être exposés au solvant ou enfouis. Ce sont ces derniers coefficients qui seront ici optimisés.

### 3.1.2 Description de l'état déplié

Pour la protéine dénaturée, on émet l'hypothèse que les chaînes latérales des acides aminés n'interagissent pas entre elles, mais seulement avec le squelette protéique adjacent et le solvant. Ceci peut être modélisé par une collection de structures de tripeptides de séquences Ala-X-Ala, avec un tripeptide pour chaque acide aminé dans la protéine. Pour un acide aminé X, différentes structures possibles du squelette et de la chaîne latérale sont considérées. Ces structures sont issues de tripeptides extraits des structures cristallographiques de six protéines de la PDB : le lysosyme (2LZM), la BPTI (Bovin Pancreatic Trypsin Inhibitor ; 4PTI), la nucléase de staphylocoque (1STN), l' $\alpha$  toxine (1PTX), la ribonucléase A (2RN2), et la cyclophiline (2CPL). Dans chaque tripeptide, la chaîne latérale X est minimisée avec le squelette protéique fixe. Pour choisir la structure de tripeptide optimale pour chaque type X, le critère est l'énergie après minimisation. Ainsi, pour chaque type d'acide aminé X, la structure du tripeptide donnant l'énergie la plus faible est retenue, pour représenter l'acide aminé X dans l'état dénaturé. Pour une protéine, l'énergie libre totale de l'état déplié est alors obtenue en sommant les contributions  $E_X$  des n résidus qui la composent. Nous dirons que  $E_X$  est une "énergie de référence".

FIGURE 3.1 – Cycle thermodynamique.  $\Delta G$  est l'énergie libre de repliement de la protéine native et  $\Delta G'$  est l'énergie libre de repliement de la protéine mutée.

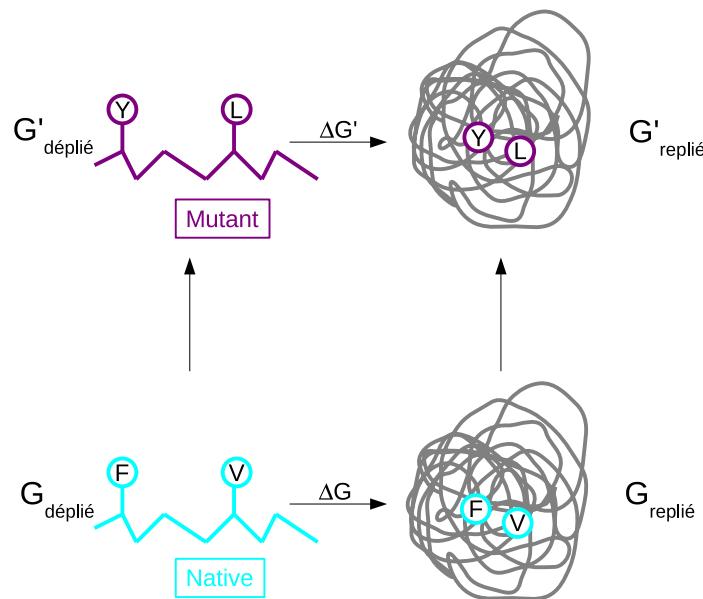


FIGURE 3.2 – Vue schématique d'une "protéine" de trois acides aminés sous forme native et mutée.



La figure 3.1 montre un cycle thermodynamique qui présente le repliement pour une protéine native et mutée. La figure 3.2 montre le cas particulier d'un tri-peptide, où le deuxième acide aminé est muté. La différence entre les énergies libres de repliement peut s'écrire, dans le cas du tri-peptide :

$$G_{\text{déplié}} = G_{\text{ref}}^1 + G_{\text{ref}}^2 + G_{\text{ref}}^3 \quad (3.2)$$

$$G'_{\text{déplié}} = G_{\text{ref}}^1 + G_{\text{ref}}^{2'} + G_{\text{ref}}^3 \quad (3.3)$$

$$\Delta G = G_{\text{replié}} - G_{\text{déplié}} \quad (3.4)$$

$$\Delta G' = G'_{\text{replié}} - G'_{\text{déplié}} \quad (3.5)$$

$$\Delta\Delta G = G'_{\text{replié}} - (G_{\text{ref}}^1 + G_{\text{ref}}^{2'} + G_{\text{ref}}^3) - [G_{\text{replié}} - (G_{\text{ref}}^1 + G_{\text{ref}}^2 + G_{\text{ref}}^3)] \quad (3.6)$$

$$= G'_{\text{replié}} - G_{\text{replié}} + G_{\text{ref}}^{2'} - G_{\text{ref}}^2 \quad (3.7)$$

Pour calculer la différence d'énergie libre entre une protéine native et mutante, on a donc besoin des "énergies de références" de l'acide aminé natif et muté.

### **3.1.3 Optimisation des coefficients de surface : choix des types atomiques**

Dans les modélisations du solvant utilisant la surface accessible au solvant, la contribution  $E_{\text{surf}}$  à l'énergie libre de solvatation est égale à la somme des surfaces atomiques,  $A_i$ , pondérées par leurs coefficients  $\sigma_i$ . La contribution de chaque atome est donc quantifiée par un coefficient qui retranscrit son caractère hydrophile ou hydrophobe. De nombreux jeux de paramètres atomiques de solvatation ont été développés. Eisenberg et McClachan [190] ont utilisé l'énergie de transfert de l'octanol vers l'eau pour les vingt acides aminés, afin d'en déduire des coefficients de solvatation pour cinq classes d'acides aminés. De nombreuses études ont été ensuite menées sur la paramétrisation de ces coefficients atomiques. Les principales différences entre ces méthodes concernent le classement des atomes en groupes ainsi que les données expérimentales utilisées pour ajuster les coefficients. Par exemple, Ooi et coll. utilisent sept groupes atomiques différents, ajustés sur des énergies libres expérimentales de petites molécules organiques [191]. Fraternali et van Gasteren [192] réduisent les types atomiques à deux groupes seulement: un groupe pour les carbones, représentant l'effet hydrophobe, et un pour les oxygènes et azotes, représentant l'effet hydrophile. Ces deux paramètres sont ensuite optimisés pour que les

surfaces accessibles au solvant de protéines obtenues par simulation de dynamique moléculaire correspondent à celles de la structure cristallographique. Ensuite, d'autres modèles de surface accessible au solvant ont été développés. Ils utilisent plus d'une centaine de groupes atomiques et un grand nombre de données expérimentales de solvatation pour diverses molécules organiques [193, 194, 195].

Dans notre modèle, les atomes sont classés en quatre groupes :

- non-polaire : carbones des alcanes et du groupement carbonyl de la chaîne principale.
- aromatique : carbones et azotes des cycles du tryptophane, de la phenylalanine et de la tyrosine.
- polaire : les atomes d'azote et d'oxygène n'appartenant pas à des groupes ionisés et le groupe N-C-N du cycle de l'histidine.
- ionique : le groupe guanidium de l'arginine, le groupe carboxyl de l'aspartate et du glutamate, le groupe N-C-N du cycle de l'histidine doublement protonée, le groupe  $\varepsilon$ -amino de la lysine.

### 3.1.4 Calcul de stabilité : méthode

Introduire une mutation modifie la stabilité de la protéine. La stabilité est calculée comme la différence entre l'énergie libre de l'état replié et celle de l'état dénaturé, ou état de référence. La différence de stabilité entre le mutant et la protéine native peut s'écrire :

$$\Delta\Delta G = \Delta G' - \Delta G \quad (3.8)$$

$$= (G'_{replié} - G_{replié}) - (G'_{ref}(i) - G_{ref}(i)) \quad (3.9)$$

Les notations sont celle de la figure 3.1, i est la position mutée.

#### 3.1.4.1 Données expérimentales utilisées pour l'optimisation

Les structures de trois protéines natives sont utilisées. Il s'agit du lysozyme, de la ribonucléase A, et de la nucléase de staphylocoque, de code PDB : 2LZM, 2RN2, 1ST1,

respectivement. Les chaînes latérales ont été légèrement minimisées avant tout calcul d'énergie (cf. ci-dessous).

Nous avons également considéré huit peptides :

pepT1 : SSDVSTAQ**X**AAYKLHED [196],

KEAKE : YEAAAKEAXAKEAAAKA [197],

K2AE2 : YSEEEEKAKKAXAEEAEKKKK [198],

VAR : KETAAAKFERQHMDS [199],

PAD : YKAAAAKAA**X**AKAAAAAK [200],

KAL : YSEEEEKKKKXSEEEEKKKK [201],

SH1 : AETAAAKFERQHM [202],

SH2 : KETAAAKFERAHA [202]

La séquence des peptides est donnée ici avec un X figurant une position qui sera systématiquement mutée. A cette position, on placera tous les acides aminés possibles, si bien qu'on obtient un peptide natif et dix-neuf mutants. On étudie quelques mutations seulement pour les peptides VAR (M13V, M13A, M13I et M13L), SH1 (A5H) et SH2 (E9L). Pour ces peptides, les structures cristallographiques ne sont pas résolues. Des modèles ont donc été construits dans Swiss PDB Viewer, qui réalise une hélice  $\alpha$  théorique à partir d'une séquence donnée [203]. Après une minimisation des chaînes latérales de ces hélices, ces structures peptidiques seront traitées comme les structures de protéines.

Les mutants correspondant à chaque protéine et peptide sont créés en replaçant les chaînes latérales mutantes aux positions voulues et en maintenant fixe les autres coordonnées. Les coordonnées de la chaîne latérale mutante sont issus de la bibliothèque de rotamères de Tuffery [204]. Ici, nous utilisons un champ de force Charmm19 avec un modèle de solvant CASA. La valeur de la constante diélectrique est 24, celle du coefficient de solvatation, 1. Chaque rotamère de la chaîne latérale mutante est minimisé pendant 30 pas à squelette protéique fixe (minimisation de Powell). L'énergie du mutant est ensuite évaluée. Le rotamère de la chaîne latérale mutante donnant l'énergie la plus basse est sélectionné. Pour la protéine native, la chaîne latérale à la position mutante est gardée dans sa conformation cristallographique, et subit une simple minimisation de 30 pas. Les

TABLE 3.1 – Exploration de différents paramètres de solvatation (en  $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ ).

type atomique	gamme	pas
non polaire	-0.005 à 0.01	0.005
aromatique	-0.08 à 0.01	0.01
polaire	-0.12 à -0.04	0.01
ionique	-0.20 à -0.10	0.01
$\varepsilon$	16 à 32	8

données expérimentales de changement de stabilité pour les protéines sont issues de la base ProTherm [205]. Pour les peptides, des échelles de propension hélicale expérimentale des protéines ont été utilisées. En effet, ces échelles mesurent la tendance des différents types d'acides aminés à former des hélices [206]. Nous disposons donc de 73 mutations dans les trois protéines et de 67 mutations dans les huit peptides, donnant un total de 140 mutations. Les changements d'énergies libre varient entre -2.4 et 3.3  $\text{kcal} \cdot \text{mol}^{-1}$ .

### 3.1.5 Protocole d'optimisation des coefficients surfaciques

Comme le choix des tripeptides modélisant l'état dénaturé dépend également des paramètres utilisés dans le modèle CASA, une optimisation itérative des coefficients de solvatation et des tripeptides modèles a été réalisée. Au point de départ de l'optimisation, on pose les coefficients de solvatation égaux à ceux de Fraternali et van Gasteren [192] ; pour les atomes ioniques, on attribue initialement la même valeur qu'aux atomes polaires [136] ; de même, le groupe d'atomes aromatiques a le même coefficient initial que les atomes apolaires. En utilisant ces paramètres, un premier jeu de structures de tripeptides de référence est choisi. Les changements de stabilité sont alors évalués pour les 140 mutations tests, en utilisant le jeu courant de structures de tripeptides pour modéliser l'état déplié. Ces changements de stabilité sont calculés pour de nombreuses combinaisons de valeurs pour les coefficients de surface (table 3.1).

Pour chaque combinaison, les énergies libres  $\Delta\Delta G$  sont obtenues. Le rms et l'erreur moyenne par rapport aux données expérimentales de changement de stabilité sont calculés. Le rms est utilisé comme un score de performance pour choisir une combinaison de paramètres. On extrait le jeu de paramètres de meilleur rang, qui respecte en même temps l'ordre physique des types atomiques. En effet, il a fallu sélectionner des paramètres

respectant l'ordre logique d'exposition au solvant. L'ordre des coefficients doit être tel que la préférence d'exposition au solvant des types atomiques est la suivante : non-polaire < aromatique < polaire < ionisé.

Le jeu de coefficients de solvatation respectant cet ordre et avec le meilleur rms est utilisé comme nouveau point de départ pour le prochain cycle d'optimisation itérative. Cette procédure est répétée jusqu'à ce qu'il n'y ait plus de changement significatif dans le rang des combinaisons de coefficients de solvatation.

### 3.1.6 Résultats de l'optimisation des coefficients

TABLE 3.2 – Coefficient de solvatation atomique initiaux de Fraternali et van Gunsteren (en kcal.mol<sup>-1</sup>.Å<sup>-2</sup>).

type atomique	valeur initiale
non polaire	0.012
aromatique	0.012
polaire	-0.06
ionique	-0.15

La procédure d'optimisation converge après quatre cycles. Les paramètres retenus sont ceux de la table 3.3. Les nouveaux paramètres montrent une différence entre le coefficient du type apolaire et aromatique alors que ce n'était pas le cas dans le jeu initial. On voit dans ces nouveaux paramètres qu'on respecte bien l'ordre physique pour l'exposition au solvant : non-polaire < aromatique < polaire < ionisé. Le rms final entre  $\Delta\Delta G$  expérimentaux et calculés est de 2.94 kcal.mol<sup>-1</sup>.

Il peut être intéressant de comparer notre modèle au modèle nul, selon lequel l'estimation du changement de stabilité serait systématiquement nulle. Ce modèle produirait une erreur moyenne de -1.07 kcal.mol<sup>-1</sup> sur nos 140 mutants, notre modèle donne une erreur moyenne de 2.28 kcal.mol<sup>-1</sup>, qui n'est pas énorme mais qui est nettement plus mauvaise que le modèle nul. Nous calculons également le coefficient de corrélation de rang de Spearman,

$$rs = \frac{\sum_i (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum (R_i - \bar{R})^2} \sqrt{\sum (S_i - \bar{S})^2}} \quad (3.10)$$

Ici,  $R_i$  et  $S_i$  sont les valeurs d'énergies libres, respectivement expérimentale et calculée,

### 3.1. PARAMÉTRISATION DU MODÈLE

TABLE 3.3 – Coefficient de solvatation atomique optimisés (en  $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ ).

type atomique	valeur optimisée
non polaire	-0.005
aromatique	-0.04
polaire	-0.08
ionique	-0.10
$\epsilon$	24
$\alpha$	1

TABLE 3.4 – Énergie de référence ( $\text{kcal} \cdot \text{mol}^{-1}$ )

résidu	Énergie optimisée
Ala	-11.3
Asp	-19.8
Asn	-17.2
Arg	-25.0
Glu	-21.3
Gln	-17.9
His	-20.4
Ile	-12.3
Leu	-12.6
Lys	-22.2
Met	-13.9
Phe	-17.4
Ser	-13.4
Tyr	-20.3
Thr	-12.6
Trp	-21.0
Val	-11.5

pour un mutant donné.  $\bar{R}$  est la moyenne des énergies libres expérimentales pour les 140 mutants.  $\bar{S}$  est la moyenne des énergies libres calculées pour les 140 mutants. Un paramètre à calculer en parallèle est la valeur  $t$  (dépendant du nombre de degrés de liberté), qui permet d'affirmer ou de rejeter la valeur non nulle de  $rs$  :

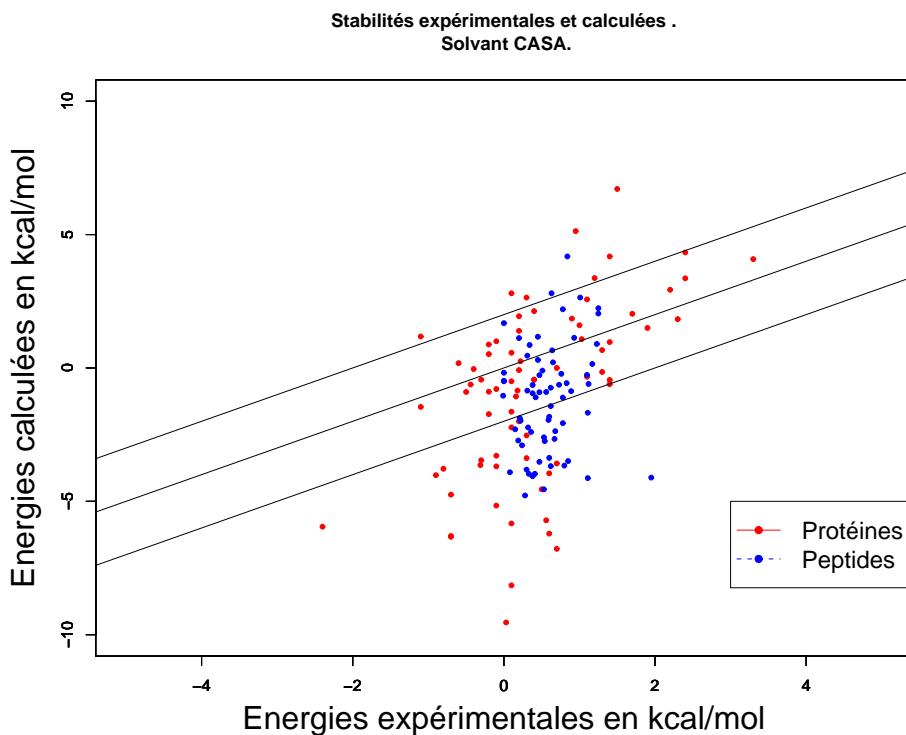
$$t = rs \sqrt{\frac{N - 2}{1 - R_s^2}} \quad (3.11)$$

La variable  $t$  est distribuée selon une distribution de Student. Si le  $t$  calculé est supérieur au seuil de confiance choisi, alors la corrélation de rang est jugée significative. En choisissant un niveau de confiance de 99%, la valeur seuil de  $t$  est de 2.58.

Pour nos 140 mutants de stabilité, les données sont corrélées en rang ( $rs = 0.37$ ,  $t_{calc} = 4.77 > t_{table} = 2.58$  à un niveau de confiance de 99%), ce qui n'est pas le cas pour le modèle nul où  $rs = 0$ . Notre modèle retranscrit donc l'effet stabilisant ou destabilisant d'une mutation, contrairement au modèle nul.

Pour nos trois protéines et huit peptides test, on obtient les énergies libres calculées et expérimentales résumés dans les tables 3.5 pour les protéines et 3.6 pour les peptides.

FIGURE 3.3 – Mutants de stabilités



Un graphique représente l'ensemble des résultats dans la table 3.3.

### 3.1.7 Validation croisée

Pour vérifier que les paramètres ne sont pas biaisés par la procédure d'optimisation, nous avons également réalisé des validations croisées. 30 des 140 mutations étudiées sont choisies au hasard et écartées des cycles d'optimisation. Avec des énergies de référence fixées aux valeurs obtenues au cycle précédent, on optimise les coefficients de surface uniquement. L'erreur moyenne est alors calculée pour les 30 mutations non prises en compte pour l'optimisation. Cette procédure est réalisée deux fois avec deux groupes différents de mutations écartées. Les jeux de paramètres et les niveaux d'erreurs obtenus lors de ces deux tests sont similaires. Ils sont également semblables à ceux obtenus après l'optimisation itérative sur l'ensemble des 140 mutations. On vérifie donc que le protocole d'optimisation n'a pas introduit de biais dans le choix des paramètres.

### 3.1. PARAMÉTRISATION DU MODÈLE

TABLE 3.5 – Stabilité des protéines (kcal.mol<sup>-1</sup>).

Lysozyme mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
D47A	5.13	0.95	4.18
D92N	4.18	1.40	2.78
E45A	-3.69	-0.10	3.59
K43A	1.08	1.03	0.05
K48A	-5.71	0.56	6.27
N144D	-0.90	-0.50	0.40
Q123E	-0.04	-0.40	0.36
D20A	2.64	0.30	2.34
E11A	-1.46	-1.10	0.36
E128A	-1.07	0.16	1.23
N116D	0.18	-0.60	0.78
N40D	-0.62	-0.44	0.18
Q105E	-0.33	1.10	1.43
R119A	-0.85	0.18	1.03
Ribonucléase A mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
K91R	-6.21	0.60	6.81
D94E	-0.44	0.40	0.84
K95A	-0.79	-0.10	0.69
K95N	-4.02	-0.90	3.12
A52E	6.71	1.50	5.21
A52D	1.83	2.30	0.47
D10N	-6.78	0.70	7.48
D10A	-5.95	-2.40	3.55
D10E	1.18	-1.10	2.28
D10S	-4.75	-0.70	4.05
E48Q	-3.46	-0.30	3.16
E48A	2.80	0.10	2.70
E48D	-1.73	-0.20	1.53
D70N	-3.38	0.30	3.68
D70A	-2.23	0.10	2.33
D134N	-8.15	0.10	8.25
D134A	-3.78	-0.80	2.98
D134V	-3.64	-0.31	3.33
H62R	-9.54	0.03	9.57
E135K	0.25	0.22	0.03
Nucléase de staphylocoque mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
D19A	-0.50	0.10	0.60
D21A	-6.33	-0.70	5.63
D40A	0.88	-0.20	1.08
D95A	4.08	3.30	0.78
E101A	1.50	1.90	0.40
E10A	-0.15	1.30	1.45
E122A	-0.44	0.40	0.84
E129A	3.36	2.40	0.96
E135A	-0.00	0.70	0.70
E52A	0.57	0.10	0.47
E57A	1.94	0.20	1.74
E67A	1.60	1.00	0.60
E73A	3.37	1.20	2.17
E75A	2.93	2.20	0.73
K110A	0.67	1.30	0.63
K116A	-6.30	-0.70	5.60
K127A	0.52	-0.20	0.72
K133A	-0.45	1.40	1.85
K136A	1.85	0.90	0.95
K16A	-2.00	0.20	2.20
K24A	1.39	0.20	1.19
K48A	1.00	-0.10	1.10
K49A	-2.53	0.30	2.83
K53A	-0.08	0.20	0.28
K63A	-4.55	0.50	5.05
K64A	-3.29	-0.10	3.19
K6A	-0.44	-0.30	0.14
K70A	-1.64	0.10	1.74
K78A	-3.95	0.60	4.55
K84A	-0.89	-0.20	0.69
K9A	-0.61	1.40	2.01
N118D	4.33	2.40	1.93
R105A	0.97	1.40	0.43
R126A	2.03	1.70	0.33
R81A	2.57	1.10	1.47
K28A	-3.58	0.70	4.28
K97A	-5.83	0.10	5.93
K71A	2.13	0.40	1.73
K134A	-5.16	-0.10	5.06

TABLE 3.6 – Stabilité des peptides (kcal.mol<sup>-1</sup>).

pepT1 mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
a9D	-2.37	0.68	3.05
A9E	-0.85	0.31	1.16
A9F	-1.83	0.60	2.43
A9H	-2.66	0.67	3.33
A9L	-1.98	0.22	2.20
A9M	-2.30	0.15	2.45
A9N	-3.68	0.62	4.30
A9Q	-3.81	0.30	4.11
A9R	-3.52	0.47	3.99
A9S	-0.27	0.47	0.74
A9T	0.21	0.65	0.44
A9V	0.66	0.64	0.02
A9W	-3.91	0.08	3.99
A9Y	-4.06	0.38	4.44
A9I	0.86	0.34	0.52
A9K	-0.95	0.38	1.33
KEAKE mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
A9D	0.9	1.23	0.2
A9E	1.13	0.93	0.74
A9F	1.17	0.45	0.72
A9H	0.15	1.17	1.02
A9I	2.20	0.78	1.42
A9K	1.12	0.20	0.92
A9L	0.46	0.31	0.15
A9M	0.30	0.45	0.15
A9N	-0.60	1.12	1.72
A9Q	-0.74	0.62	1.36
A9R	-1.04	-0.01	1.03
A9S	2.64	1.01	1.63
A9T	2.04	1.25	0.79
A9V	2.24	1.25	0.99
A9W	-1.43	0.62	2.05
A9Y	-1.11	0.78	1.89
K2AE2 mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
A12D	-2.07	0.78	2.85
A12E	-1.88	0.21	2.09
A12F	-1.68	1.11	2.79
A12H	-2.74	0.54	3.28
A12L	-2.40	0.36	2.76
A12M	-2.60	0.53	3.13
A12N	-3.49	0.85	4.34
A12Q	-3.66	0.80	4.46
A12R	-3.97	0.41	4.38
A12S	-0.22	0.76	0.98
A12T	-0.87	0.89	1.76
A12V	-0.63	0.73	1.36
A12W	-4.13	1.11	5.24
A12Y	-4.11	1.95	6.06
A12I	-0.64	0.38	1.02
A12K	-1.95	0.59	2.54
VAR mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
M13V	-0.48	0.00	0.48
M13A	1.68	0.00	1.68
M13I	-0.50	0.00	0.50
M13L	-0.18	0.00	0.18
PAD mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
A10V	-0.57	0.83	1.40
KAL mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
A11L	-2.72	0.19	2.91
A11M	-2.90	0.24	3.14
A11Q	-3.97	0.33	4.30
A11I	-1.10	0.42	1.52
A11V	-0.91	0.47	1.38
A11S	-0.10	0.51	0.61
A11T	-0.90	0.56	1.46
A11N	-3.37	0.60	3.97
A11W	-4.78	0.28	5.06
A11F	-2.23	0.32	2.55
A11Y	-4.55	0.53	5.08
A11H	-2.80	0.63	3.43
SH1 mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
A5H	4.18	0.84	3.34
SH2 mutation	$\Delta\Delta G$ calculé	$\Delta\Delta G$ expérimental	erreur
E9L	-0.26	1.10	0.49

## 3.2 Calcul d'affinité protéine-ligand et protéine-protéine

### 3.2.1 Données expérimentales

Pour valider notre protocole de calcul d'affinité, il est nécessaire de comparer les résultats calculés aux résultats expérimentaux. Pour obtenir les énergies libres expérimentales, considérons l'équilibre suivant :



où E est l'enzyme, S le ligand et ES, le complexe formé. L'affinité du ligand pour son récepteur est caractérisée par la constante de dissociation K :

$$K = \frac{[E][S]}{[ES]} \quad (3.13)$$

Soit  $\Delta G$  l'énergie libre de liaison.

$$\Delta G = E_{complexe} - (E_{protéine} + E_{ligand}) \quad (3.14)$$

Pour relier les constantes d'associations expérimentales aux énergies libres, on utilise la loi d'action de masse :

$$\Delta G_{mut} = \Delta G_{mut}^0 - RT \ln K_{mut} \quad (3.15)$$

$$\Delta G_{nat} = \Delta G_{nat}^0 - RT \ln K_{nat} \quad (3.16)$$

$K_{nat}$  et  $K_{mut}$  sont les constantes de dissociation ; les énergies libres  $\Delta G^0$  correspondent aux conditions standard : concentrations de 1M, température et pression ambiantes.  $R = 8.313 \text{ J.K.mol}^{-1}$  et  $T = 300 \text{ K}$ .

A l'équilibre,

$$\Delta G_{nat} = \Delta G_{mut} = 0 \text{ kcal.mol}^{-1} \quad (3.17)$$

d'où

$$\Delta G_{mut}^0 - RT \ln K_{mut} = \Delta G_{nat}^0 - RT \ln K_{nat} \quad (3.18)$$

### 3.2. CALCUL D'AFFINITÉ PROTÉINE-LIGAND ET PROTÉINE-PROTÉINE

On en déduit :

$$\Delta\Delta G_{mut-nat}^0 = \Delta G_{mut}^0 - \Delta G_{nat}^0 = -RT \ln\left(\frac{K_{mut}}{K_{nat}}\right) \quad (3.19)$$

Pour valider notre modèle, nous comparerons les différences d'énergies libre obtenues à partir des constantes de dissociation expérimentales avec celles calculées par notre protocole. Ceci sera fait pour plusieurs complexes et leurs mutants.

#### **3.2.1.1 Calculs d'affinité : méthode**

Le solvant est modélisé ici avec le modèle CASA décrit précédemment :

$$E_{solv} = E_{screen} + E_{surf} = \left(\frac{1}{\varepsilon} - 1\right)E_{coul} + \alpha \sum_{i=1}^n A_i \sigma_i \quad (3.20)$$

Nous avons testé différentes valeurs de constante diélectrique (12 et 16). Pour chacune d'elles, nous avons testé différentes valeurs du facteur  $\alpha$  du terme d'énergie de surface (0.33, 0.5 et 1). Finalement, ont été utilisés un facteur  $\alpha$  égal à 0.5 et une constante diélectrique  $\varepsilon$  de 16, car ces valeurs donnent ici le meilleur accord entre données expérimentales et calculées. Le champ de force Charmm19 ne représente que les hydrogènes polaires ; les autres hydrogènes sont pris en compte implicitement. quelques molécules d'eau du site actif sont intégrées explicitement dans le système (voir paragraphe suivant). Les histidines sont étudiées comme neutres, avec le proton sur l'atome ND1.

Les calculs d'affinités ont été réalisés sur six complexes différents de la PDB : la tyrosyl-ARNt synthétase complexée à la L-Tyrosine (4TS1), l'aspartyl tRNA-synthétase complexée à l'aspartate (1IL2), le lysozyme complexé à l'anticorps HyHel-10 (3HFM), le complexe formée par la glycoprotéine CD4 avec le composant gp120 du virus HIV (1G9M), le complexe trypsine-BPTI (2PTC), et le complexe chymotrypsine-BPTI (1CBW). Plusieurs de leurs mutants ont été également construits pour en étudier l'affinité. Dans tous les cas, une sphère de résidus est sélectionnée. Elle contient les résidus dont au moins un des atomes est distant de moins de 20 Å du ligand. Les résidus au delà de cette distance n'influencent pas l'interaction protéine-ligand et sont supprimés du système. On considère une molécule d'eau du site actif explicitement pour les complexes suivants : TyrRS:Tyr,

AspRS:Asp et HEL:HyHel10 . Une minimisation de 50 pas avec squelette protéique fixe permet d'obtenir le système natif initial. A partir de ce système natif, tous les atomes sont considérés comme fixes, à l'exception de la chaîne latérale qui sera mutée par la suite. Soit  $n$  la position dans la séquence de l'acide aminé natif à muter. La chaîne latérale native à la position  $n$  est minimisée 50 pas en présence de l'ensemble du système. L'énergie libre du système sous forme complexée enzyme:ligand est calculée. Le ligand est alors éloigné de 1000 Å. La chaîne latérale  $n$  est alors minimisée 50 pas en présence de la protéine seule. L'énergie libre du système [apoenzyme + ligand] est alors calculée. La différence de ces deux énergies libres donne l'énergie de liaison pour le complexe natif.

Les structures des mutants sont créées en remplaçant successivement la chaîne latérale concernée par les rotamères de l'acide aminé mutant provenant de la bibliothèque de Tuffery [204]. Le squelette protéique est considéré comme fixe. Soit  $n$  la position de l'acide aminé muté. Après placement du premier rotamère, la chaîne latérale  $n$  est minimisée 50 pas en présence de l'ensemble du système. L'énergie libre du système sous forme complexée enzyme:ligand est alors calculée. Le ligand est éloigné de 1000 Å. La chaîne latérale  $n$  est alors minimisée 50 pas, en présence de la protéine seule. L'énergie libre du système [apoenzyme + ligand] est alors calculée. On fait les mêmes calculs d'énergie pour tous les rotamères de l'acide aminé mutant. On sélectionne le rotamère ayant conduit à la structure d'énergie libre la plus stable pour la forme complexée. On sélectionne également le rotamère ayant conduit à la structure d'énergie libre la plus stable dans la forme libre. La différence de ces deux énergies libres donne l'énergie de liaison pour le mutant.

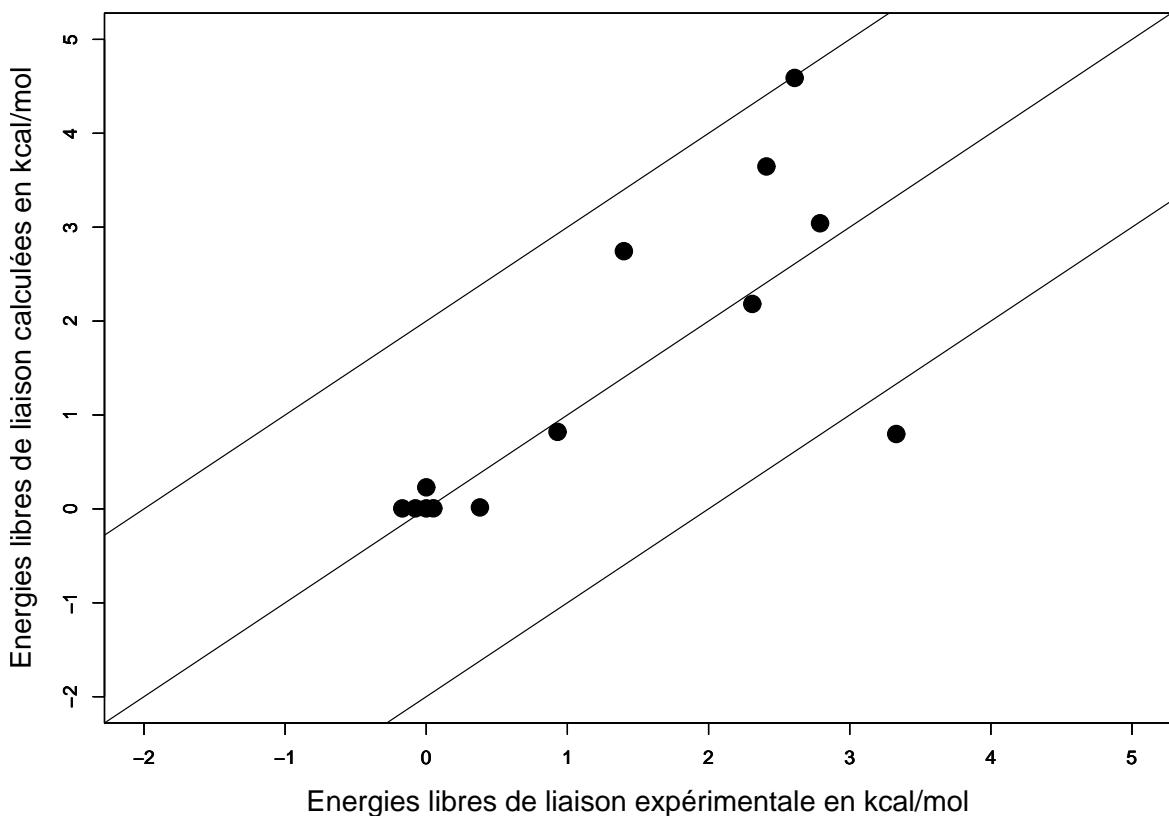
### **3.2.2 Calculs d'affinité protéine-ligand : résultats**

Pour valider le modèle de calcul, nous considérons une série de TyrRS mutantes et nous confrontons le modèle aux résultats expérimentaux pour l'affinité TyrRS :Tyr. La TyrRS d'*E. Coli* n'a pas fait l'objet de nombreuses mesures biochimiques expérimentales. En revanche, la TyrRS de *Bacillus stearothermophilus* a été très étudiée. Ces deux enzymes de TyrRS sont très proches ; l'identité de séquences est de 62%, et la déviation rmsd entre les deux structures n'est que de 3 Å lorsqu'on considère une sphère de 5Å autour du ligand.

### 3.2. CALCUL D'AFFINITÉ PROTÉINE-LIGAND ET PROTÉINE-PROTÉINE

Les énergies libres calculées et expérimentales [207, 58, 208, 46, 209] sont représentées dans la figure 3.4 et données dans la table 3.7. L'erreur moyenne en valeur absolue est de 0.62 kcal.mol<sup>-1</sup> avec la paramétrisation CASA utilisée ici.

FIGURE 3.4 – 13 mutants de Tyr-RS de *B. stearothermophilus* : affinités expérimentales modélisées. La ligne centrale est celle sur laquelle se situeraient les points si l'accord entre données expérimentales et données calculées était parfait. Elle est encadrée par deux lignes qui donnent une marge de  $\pm 2$  kcal.mol<sup>-1</sup>.



L'affinité du mutant N123D n'est pas bien estimée par notre calcul. Ceci montre la difficulté d'estimer des changements de charge important avec ce modèle de solvant.

Les données expérimentales et modélisées présentent un coefficient de corrélation assez élevé,  $r=0.8$ . Pour décider si cette valeur est significative, nous utilisons le test de Student déjà décrit plus haut. En choisissant un niveau de confiance de 99%, la valeur seuil de  $t$  est de 2.95. Si le  $t$  calculé est supérieur au seuil, la probabilité d'observer une telle corrélation de rang sous l'hypothèse nulle est inférieure à 1%. Ici, le calcul du coefficient de corrélation d'ordre de rang de Spearman donne  $rs=0.8$ . Le nombre de degrés de liberté

TABLE 3.7 – affinité protéine-ligand en kcal.mol<sup>-1</sup>.

mutation	TyrRS:L-Tyr		
	$\Delta\Delta G$ expérimental	$\Delta\Delta G$ calculé	erreur
Y34F	1.40	2.74	1.34
C35S	-0.24	0.07	0.31
D38A	2.41	3.60	1.19
T51A	0.0	0.23	0.23
D78A	2.61	4.55	1.94
N123A	0.93	0.82	0.11
N123D	3.33	0.80	2.53
Y169F	2.79	3.21	0.42
Q173A	2.31	2.40	0.09
K230A	0.38	0.02	0.36
F231L	0.00	0.01	0.01
G232A	0.41	0.02	0.39
K233A	-0.08	0.01	0.09
T234A	-0.17	0.01	0.18
T234S	0.05	0.01	0.04
mutation	AspRS:L-Tyr		
	$\Delta\Delta G$ expérimental	$\Delta\Delta G$ calculé	erreur
Q195A	1.59	2.06	0.47
Q195E	1.69	0.85	0.84
Q195N	2.27	-0.97	3.24
K198L	1.00	4.97	3.97
Q199A	1.68	0.05	1.63
Q199E	0.70	-0.15	0.85
Q199N	1.74	0.07	1.67
E235D	1.06	-0.15	1.21
R489H	1.35	4.17	2.82

pour la grandeur t est  $17 - 2 = 15$ . On trouve  $t_{calc}=4,99 > t_{table}=2.95$ . Les différences d'énergies de liaison expérimentales et théoriques entre protéines natives et mutantes sont donc bien classées dans un ordre similaire.

De même, on analyse des données expérimentales pour des énergies de liaisons de l'AspRS native ainsi que des mutants complexés au ligand natif L-Asp [210]. Le calcul des énergies de liaisons est en assez bon accord avec les mesures biochimiques. Avec le modèle de solvant CASA, l'erreur moyenne entre énergies de liaisons modélisées et expérimentales est de 1.86 kcal.mol<sup>-1</sup>.

### **3.2.3 Calculs d'affinité protéine-protéine : résultats**

Nous avons également considéré quatre complexes protéine-protéine. Le premier complexe est celui entre le lysozyme de poulet (HEL) et l'anticorps (HyHel10) de souris. Le lysozyme est constitué de 119 résidus et l'anticorps comporte une chaîne "légère" de 214 résidus et une chaîne "lourde" de 215 résidus. La structure cristallographique de ce complexe de code PDB 3HFM a une résolution de 3 Å. Pour ce complexe, nous disposons de mesures expérimentales pour 10 mutations [211]. 9 seulement ont été traitées ici (table 3.8). La dernière mutation n'a pas pu être construite, en raison de mauvais contacts atomiques menant à une élévation trop importante de l'énergie de van der Waals. Avec la paramétrisation CASA décrite ci-dessus, on trouve une erreur moyenne de 2,6 kcal.mol<sup>-1</sup> pour ce complexe. Cette erreur moyenne est nettement supérieure à celle obtenue pour les affinités TyrRS:Tyr. Il semblerait que les mutations de résidus chargés (7 parmi les 9), soient plus difficiles à modéliser par ce modèle de solvant et ce protocole simples.

Le deuxième complexe est celui entre une glycoprotéine de l'enveloppe du VIH, gp120, et une molécule du complexe majeur d'histocompatibilité, CD4, située sur la paroi des lymphocytes T humains. La protéine CD4 possède 185 résidus et la glycoprotéine gp120 en a 321. La formation du complexe est l'une des premières étapes de l'infection par le virus. La structure cristallographique du complexe natif a une résolution de 2.2 Å (code PDB 1G9M). Les études biochimiques sur cette interaction étant très nombreuses, 25 mutations ont pu être étudiées. Comme pour Hel:HyHel10, la modélisation a générée certains mutants ayant de mauvais contacts inter-atomiques. 22 mutations ont donc été retenues pour les calculs d'affinité [212]. Elles sont détaillées dans la table 3.9. L'erreur moyenne est de 1.3 kcal.mol<sup>-1</sup>.

On étudie également deux autres complexes, le complexe trypsine:BPTI [213] de code PDB 2PTC, et le complexe Chymotrypsine:BPTI [214], de code PDB 1CBW. Le BPTI (Bovine Pancreatic Trypsin Inhibitor) est un inhibiteur de protéase à sérine. Dans ces complexes enzyme-inhibiteur, une boucle contenant un site P1 interagit avec la poche S1 de l'enzyme. 12 et 13 mutations à la position 15 du BPTI sont étudiées, respectivement, dans chaque complexes (tables 3.11 et 3.10). Le résidu natif à la position 15 est une lysine.

TABLE 3.8 – Affinité protéine-protéine : complexe HEL:HyHel10 en kcal.mol<sup>-1</sup>.

mutation	$\Delta\Delta G$ expérimental	$\Delta\Delta G$ calculé	erreur
R21E	2.5	4.39	1.9
R21K	1.9	5.51	3.6
W62Y	-0.10	-0.90	0.8
W63Y	1.70	1.36	0.3
K96M	6.7	2.00	4.7
K97D	6.7	5.13	1.6
K97G	6.4	7.71	1.3
K97M	0.9	5.06	4.2
D101G	0.5	6.24	5.7

Pour ces deux complexes, on trouve une erreur moyenne de 3.4 kcal.mol<sup>-1</sup>. Certaines mutations affectant considérablement l'affinité sont identifiées comme telles par notre protocole. C'est le cas des mutations K15A, K15V, K15T, et K15I, qui changent l'énergie de liaison calculée par plus de 10 kcal.mol<sup>-1</sup> chacune. Le protocole a été modifié pour ces deux complexes : on a permis une flexibilité de tout le BPTI au cours de la minimisation d'énergie, au lieu de laisser uniquement la chaîne latérale mutée libre. L'erreur moyenne donnée par ce protocole modifié est plus faible : 2.68 et 2.81 kcal.mol<sup>-1</sup>, respectivement, pour les complexes trypsine:BPTI et Chymotrypsine:BPTI.

Pour l'ensemble des six systèmes (TyrRS:L-Tyr, AspRS:L-Asp, HEL:HyHel10, gp120:CD4, trypsine:BPTI, Chymotrypsine:BPTI), le coefficient de Spearman pour le calcul d'affinité est égal à 0.70, avec  $t=8.59$ . Au niveau de confiance 99%, la valeur seuil de  $t$  est de 2.68. Globalement, l'erreur moyenne est plus élevée dans le cas d'interactions protéine-protéine que dans le cas d'interactions protéine-ligand. La simplicité du protocole n'autorise pas de variation importante de la structure, car le squelette protéique reste fixe. Ceci peut expliquer en partie la plus grande difficulté à modéliser les interactions protéine-protéine. Ces résultats restent cependant corrects pour l'estimation de l'affinité. Les protocoles exposés ici seront utilisés dans les chapitres suivants pour le CPD de la TyrRS.

### 3.2. CALCUL D'AFFINITÉ PROTÉINE-LIGAND ET PROTÉINE-PROTÉINE

---

TABLE 3.9 – Affinité protéine-protéine : complexe gp120:CD4 en kcal.mol<sup>-1</sup>.

mutation	$\Delta\Delta G$ expérimental	$\Delta\Delta G$ calculé	erreur
S19Y	0.17	0.02	0.15
H27A	0.34	-0.29	0.64
G38A	1.4	0.02	1.38
Q40A	-0.35	3.03	3.38
F43L	0.61	3.12	2.51
F43W	1.71	5.74	4.03
F43Y	2.13	-1.78	3.91
T45A	-0.24	-0.26	0.02
K46A	1.82	1.70	0.12
P48Q	0.96	-1.71	2.67
S49V	1.59	-0.06	1.65
N52A	-0.07	1.31	1.38
A55V	0.58	-0.03	0.61
D56A	0.34	-0.21	0.55
R59A	1.29	7.96	6.67
R59K	-0.24	1.91	2.15
W62Y	3.06	0.07	2.99
E77Q	0.22	0.02	0.20
T81A	0.3	0.02	0.28
E85A	0.22	0.14	0.08
E85Q	0.3	0.00	0.30
Q89L	0.17	2.13	1.96

TABLE 3.10 – Affinité protéine-protéine : complexe Trypsine:BPTI en kcal.mol<sup>-1</sup>.

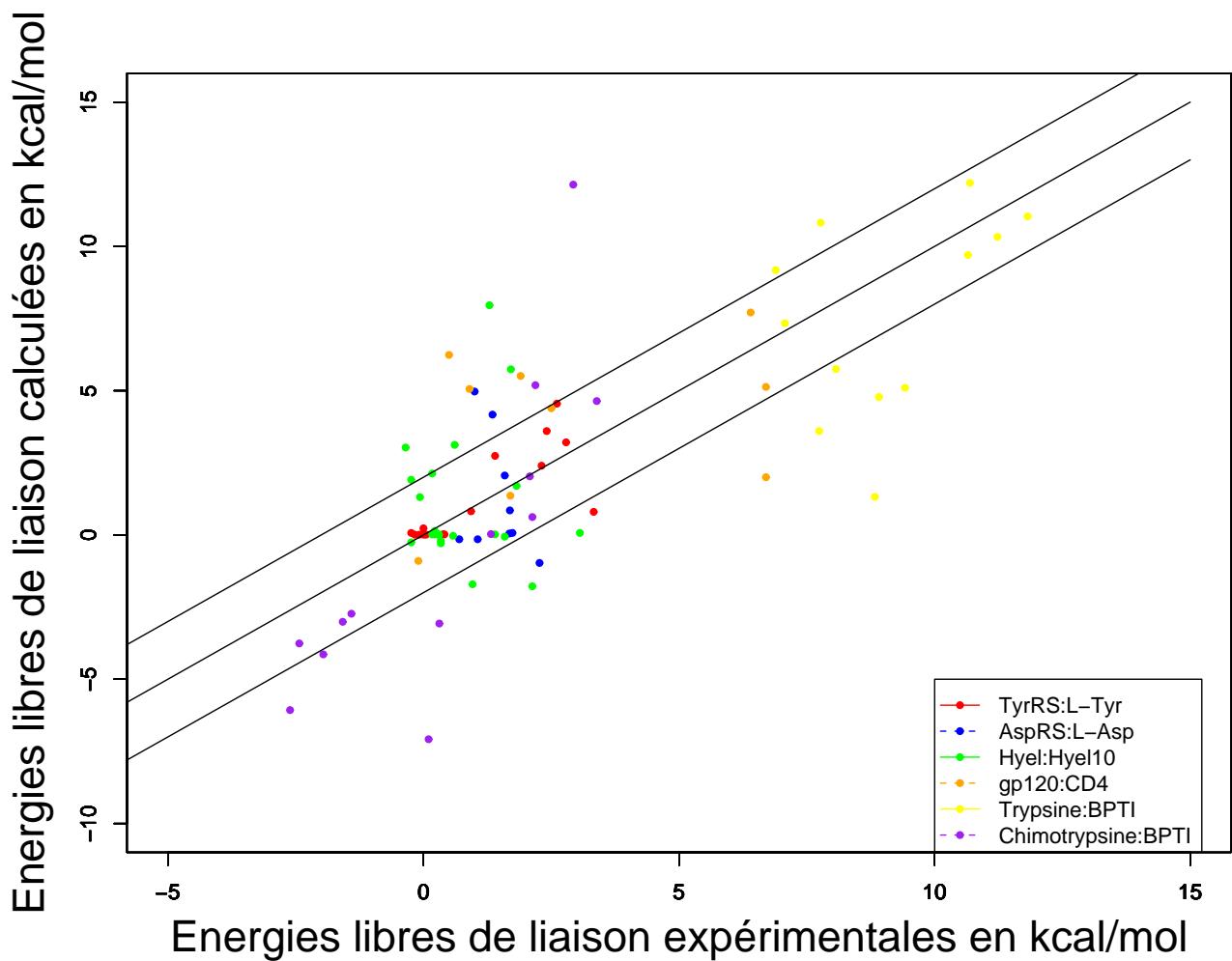
mutation	$\Delta\Delta G$ expérimental	$\Delta\Delta G$ calculé	erreur
K15A	10.69	12.20	1.51
K15S	7.77	10.82	3.05
K15V	11.82	11.04	0.78
K15T	10.65	9.70	0.95
K15L	8.91	4.78	4.13
K15I	11.23	10.33	0.90
K15M	7.74	3.60	4.14
K15N	8.07	5.75	2.32
K15Q	9.42	5.10	4.32
K15H	8.83	1.32	7.51
K15P	7.07	7.34	0.27
K15Y	6.89	9.18	2.29

TABLE 3.11 – Affinité protéine-protéine : complexe Chymotrypsine :BPTI en kcal.mol<sup>-1</sup>.

mutation	$\Delta\Delta G$ expérimental	$\Delta\Delta G$ calculé	$ \text{erreur} $
K15A	2.19	5.19	3.00
K15S	3.39	4.64	1.25
K15V	2.13	0.62	1.51
K15T	2.08	2.03	0.05
K15L	-1.58	-3.01	1.43
K15I	2.93	12.14	9.21
K15M	-1.41	-2.73	1.32
K15N	1.32	0.03	1.29
K15Q	0.31	-3.07	3.38
K15H	0.10	-7.08	7.18
K15P	-1.96	-4.14	2.18
K15Y	-2.61	-6.07	3.46
K15W	-2.43	-3.76	1.33

FIGURE 3.5 – Accord entre données expérimentales et données modélisées pour six complexes.

Modèle de solvant CASA. La ligne centrale est celle sur laquelle se situeraient les points si l'accord entre données expérimentales et données calculées était parfait. Elle est encadrée par deux lignes qui donnent une marge de  $\pm 2 \text{ kcal.mol}^{-1}$ .





# Chapitre 4

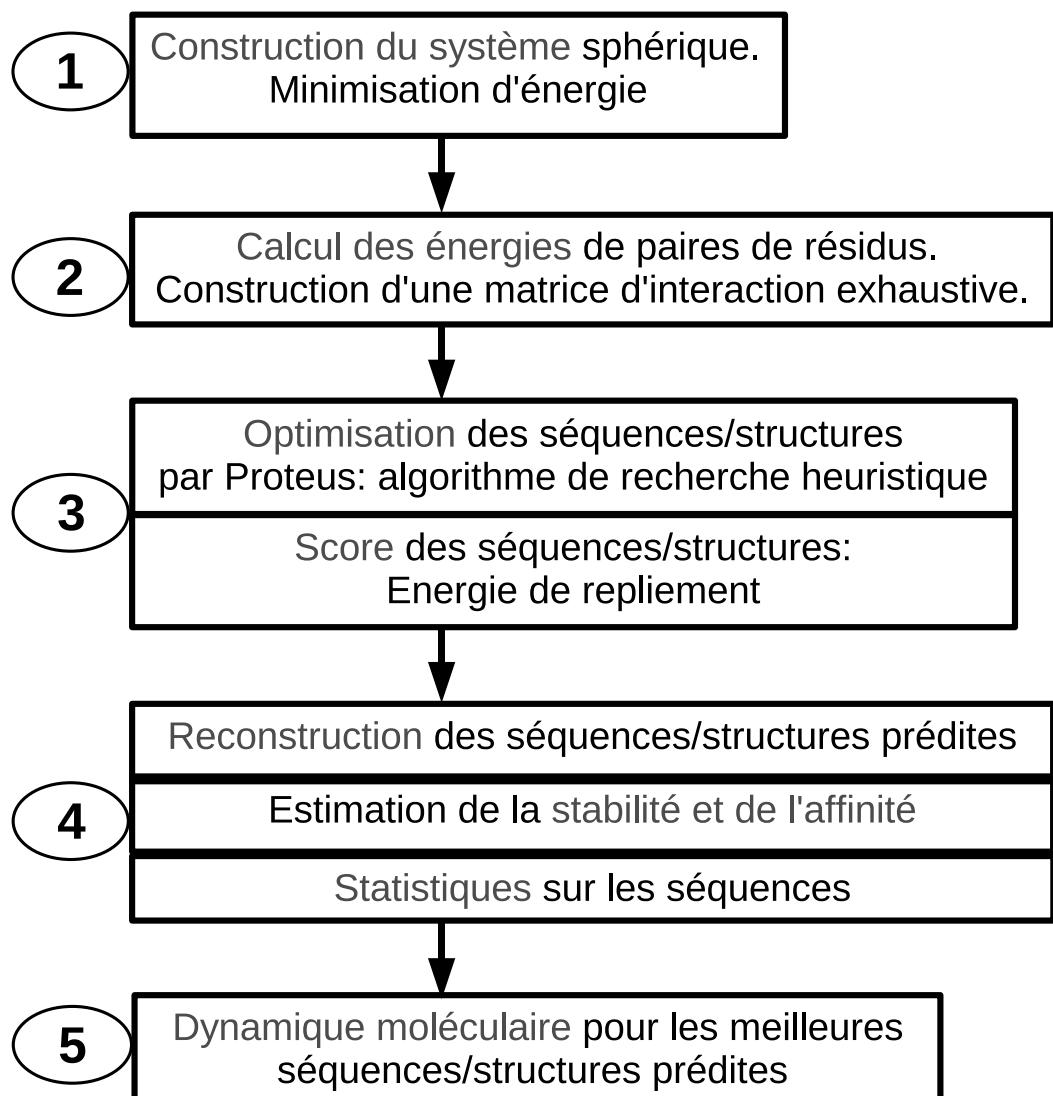
## Dessein computationnel de la TyrRS : méthodes

Nous allons modéliser l'évolution dirigée de la TyrRS d'*E. coli* pour l'amener à lier préférentiellement la D-Tyrosine. Le dessein computationnel, ou CPD de la TyrRS va être mené en cinq étapes. La première est la construction du système. La deuxième est le calcul d'une matrice d'interactions entre les résidus du système. La troisième est l'optimisation de la séquence par un algorithme heuristique exploitant cette matrice d'énergie. Les solutions ainsi optimisées sont classées d'après une fonction de score. Il reste alors, dans la quatrième étape, à reconstruire les structures 3D correspondant aux séquences prédites et à les analyser structuralement et statistiquement. Enfin, dans une cinquième étape, les meilleures séquences et structures seront étudiées par des simulations de dynamique moléculaire et par des estimations d'affinités utilisant un modèle de Poisson-Boltzmann (figure 4.1).

### 4.1 Etape un : construction des systèmes

Pour effectuer le CPD de la TyrRS, nous construirons les quatre complexes suivants : TyrRS:L-Tyr, TyrRS:D-Tyr , TyrRS:L-TyrAMP et TyrRS:D-TyrAMP.

FIGURE 4.1 – Procédure de CPD utilisée.



### 4.1.1 Construction du complexe TyrRS:Ligand

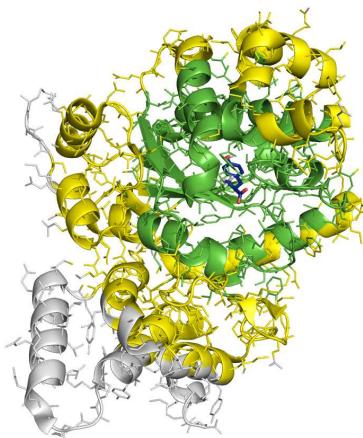
La TyrRS est un dimère symétrique formé de deux chaînes polypeptidiques. Pour modéliser la TyrRS d'*E. coli*, nous avons utilisé la structure cristallographique 1X8X [215], qui contient une chaîne de la TyrRS liée à la tyrosine, et la structure 1VBM [215], qui contient les deux chaînes de la TyrRS en liaison à un analogue du TyrAMP. 1X8X sera utilisé pour les complexes TyrRS:Tyr et 1VBM pour les complexes TyrRS:TyrAMP. Ces deux structures ont une résolution de 2 Å. Le magnésium n'est pas présent dans les modèles cristallographiques, car il n'est pas crucial pour la fixation de la tyrosine [59].

Etudions maintenant la disposition du ligand. Le ligand L-Tyr est présent dans la structure cristallographique de code PDB 1X8X. Le ligand D-Tyr est positionnée de manière à ce que le groupement carboxyle soit superposé à celui de la D-Tyr. Ainsi, ce groupement est dans la même configuration que le ligand natif, ceci permet l'activation de l'acide aminé. Le ligand L-TyrAMP est présent dans la structure cristallographique de code PDB 1VBM. Le ligand D-TyrAMP est construit par superposition au ligand L-TyrAMP. Seuls l'AMP et le groupement C=O sont concernés par cette superposition. La position de l'ammonium et du cycle aromatique de la tyrosine dépend alors de la forme L ou D.

Il est inutile d'étudier la structure entière pour décrire l'interaction protéine-ligand, car au delà d'une distance limite, les interactions avec le ligand sont très faibles et considérées comme négligeables. Nous étudierons donc une sphère de 25 Å de rayon autour du ligand. Plus précisément, nous conservons les résidus dont au moins un atome est distant de moins de 25 Å d'un des atomes du ligand. Les résidus au-delà de cette limite ne seront pas pris en compte. On définit quelques acides aminés proches du ligand qui pourront être mutés. Cet ensemble, nommé "positions actives", pourra adopter toutes les séquences et tous les rotamères possibles. Les autres résidus dans une sphère de 15 Å de rayon autour du ligand pourront prendre toutes les conformations rotamériques possibles, mais ne seront pas mutés. Cet ensemble forme les "positions inactives". Les résidus entre 15 Å et 25 Å sont considérés comme fixes et conservent la structure cristallographique. Le squelette protéique est également considéré comme fixe (figure 4.2).

FIGURE 4.2 – Représentation des systèmes de TyrRS d'*E. coli* étudiés.

(a) Un monomère de la TyrRS d'*E. coli*. code PDB : 1X8X. En bleu le ligand Tyrosine. En vert, la région des positions inactives (sphère de 15 Å). En jaune, la région des acides aminés figés dans la conformation cristallographique (sphère de 25 Å) ; en gris, les acides aminés qui ne sont pas pris en compte dans l'étude.



(b) TyrRS d'*E. coli*. code PDB : 1VBM. En bleu le ligand tyrosyl adénylate. En vert, la région des positions inactives (sphère de 15 Å). En jaune, la région des acides aminés figés dans la conformation cristallographique (sphère de 25 Å) ; en gris, les acides aminés qui ne sont pas pris en compte dans l'étude.

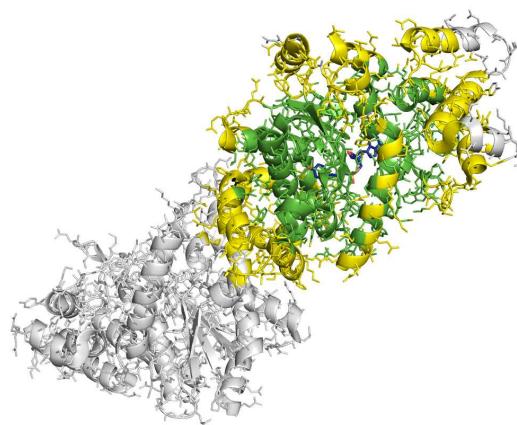
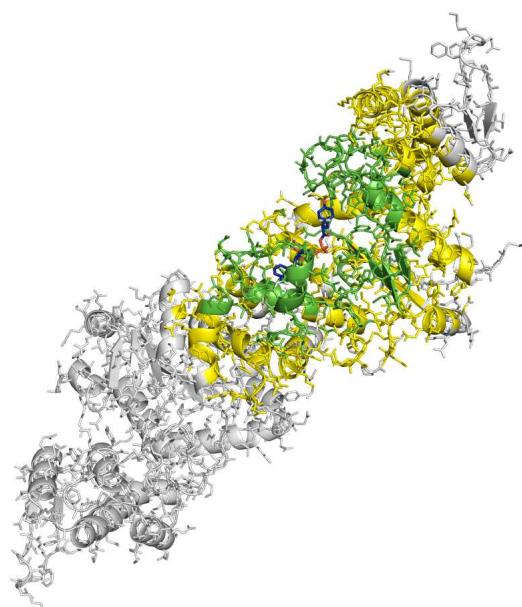


FIGURE 4.3 – Représentation du système de TyrRS d'*M. jannaschii* étudié. code PDB : 1U7D. En bleu le ligand TyrAMP. En vert, la région des positions inactives (sphère de 15 Å). En jaune, la région des acides aminés figés dans la conformation cristallographique (sphère de 25 Å) ; en gris, les acides aminés qui ne sont pas pris en compte dans l'étude.



Nous nous intéressons également à la TyrRS de *M. jannaschii*, liée au TyrAMP sous sa forme L et D, en présence de l'ion Mg<sup>2+</sup>. La structure cristallographique utilisée est celle du dimère de TyrRS, de code 1U7D, résolue à 2.65 Å [216]. L'adénylate et le magnésium proviennent de la structure de TyrRS de *S. cerevisiae*, résolue avec l'adénylate et le Mg<sup>2+</sup> à 2.4 Å de résolution (code PDB 2DLC) [55]. Il existe des résidus manquants dans des régions non structurées dans 1U7D. On obtient leurs coordonnées à partir de la structure d'une TyrRS mutante de code PDB 1U7X [216]. La structure du dimère complet avec le TyrAMP est ensuite raffiné par l'outil Modeller [217]. Comme ci-dessus, les résidus dans une sphère de 15 Å autour du ligand pourront prendre toutes les conformations rotamériques possibles, en conservant leur séquence (positions inactives). Quelques acides aminés pourront muter (positions actives). Les résidus entre 15 Å et 25 Å sont considérés fixes et conservent la structure cristallographique. Le squelette protéique est également considéré comme fixe (figure 4.3).

Ces trois systèmes font tout d'abord l'objet d'une reconstruction des hydrogènes. Après analyse des structures, les histidines sont considérées comme non protonées. On sauvegarde alors les coordonnées des structures initiales complètes. Ensuite, seuls les résidus des sphères de 25 Å autour du ligand sont conservés. Ces trois sphères font l'objet d'une légère minimisation de 50 pas où l'ensemble de la structure est mobile, de façon, à éliminer les encombrements stériques trop importants. Les calculs sont faits avec le logiciel Xplor [218]. Le champ de force utilisé ici est CHARMM19 avec un modèle de solvant CASA. Les interactions non liées sont considérées comme nulles si les atomes considérés sont distants de plus de 9 Å. On obtient alors trois structures initiales stables. A cette étape, on identifie également les résidus exposés au solvant ou enfouis, de manière à leur assigner un coefficient qui nous sera utile plus tard, lors de l'évaluation du terme d'énergie surfacique.

### 4.1.2 Choix des positions à muter

La différence structurale entre la L-Tyr et la D-Tyr est subtile. Il faut donc concentrer les efforts de mutagénèse dirigée sur quelques positions seulement. Dans un premier temps,

on identifie quatre acides aminés très proches du groupement ammonium du ligand (figure 4.6). On choisit de traiter exhaustivement les mutations à ces quatres positions. Dans l'enzyme d'*E. coli*, les quatres acides aminés à muter sont l'aspartate 81, la tyrosine 175, la glutamine 179 et la glutamine 201. Ces positions sont très conservées à travers les différents règnes du vivant (figure 4.5). Seule la position correspondant à Asp81 de la TyrRS d'*E. coli*, varie légèrement ; on y retrouve parfois un acide aminé différent de l'aspartate mais qui reste polaire. Dans l'enzyme de *M. jannaschii*, nous allons considérer des positions un peu différentes : l'isoleucine 137, la glutamine 155, le glutamate 172 et l'aspartate 173. La correspondance des jeux de positions mutantes de ces deux organismes est montrée dans les figures 4.4 et 4.5.

FIGURE 4.4 – Correspondance entre les positions mutantes de TyrRS d'*E. coli* et de *M. jannaschii*. En gras, sont indiquées les positions mutantes dans chaque organisme.

<i>E. coli</i>	<b>81</b> Asp	156 Ala	<b>175</b> Tyr	<b>179</b> Gln	200 Asp	<b>201</b> Gln
<i>M. jannaschii</i>	75 Gln	<b>137</b> Ile	151 Tyr	<b>155</b> Gln	<b>172</b> Glu	<b>173</b> Asp

#### 4.1.3 Construction des quatres ligands et de leurs rotamères

On utilise la librairie de rotamères de Tufféry et coll. de 1995 [204]. Ici, la L-Tyrosine a huit rotamères. Les rotamères de la D-Tyrosine sont déduits de ceux de la L-Tyr : les angles dièdres  $\chi_1$  et  $\chi_2$  des rotamères sont les mêmes que pour la L-Tyrosine, mais l'ammonium de la chaîne principale prend l'orientation D. Il y a donc huit rotamères pour la D-Tyrosine (figure 4.7). Pour le ligand L-TyrAMP, le jeu de rotamères est légèrement enrichi, pour pouvoir avoir un traitement plus fin des orientations de l'hydroxyle de la chaîne latérale. Le jeu initial ne possède que deux orientations du groupement hydroxyle pour chaque position du cycle aromatique. Ces deux orientations possibles de l'hydroxyle forment un angle d'environ  $180^\circ$ . Le nouveau jeu de 32 rotamères comprend quatre orientations pour chaque position du cycle. Dans ce jeu de rotamères on considère aussi la symétrie du cycle aromatique. Pour le D-TyrAMP, on déduit les rotamères de ceux du L-TyrAMP de la manière suivante. On déplace les groupements du carbone  $\alpha$  asymétrique de trois façons possibles, qui font passer de la forme L à la forme D :

FIGURE 4.5 – Alignement de séquences de TyrRS à travers les règnes du vivant. Les flèches rouges indiquent les séquences d' *Escherichia coli*, de *Bacillus stearothermophilus* et de *Methanocaldococcus jannaschii*. Les doubles barres indiquent où la séquence a été tronquée.

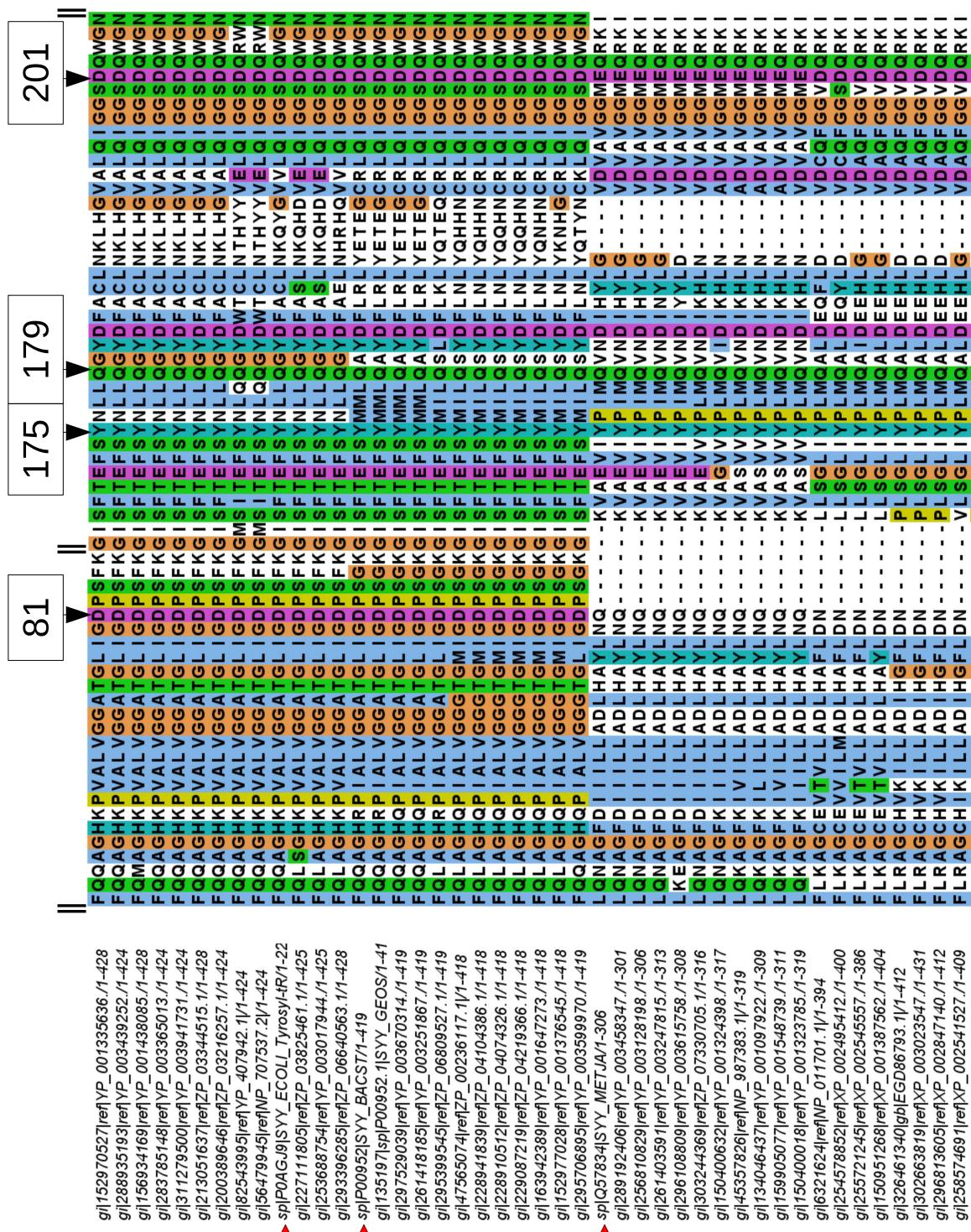
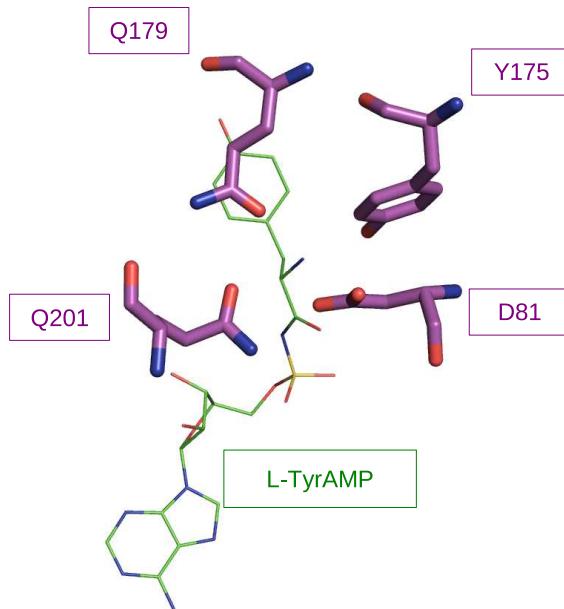


FIGURE 4.6 – Les quatre positions mutées de la TyrRS d'*E. coli*.



- échange de l'hydrogène et du groupement ammonium,
- échange de l'hydrogène et de la chaîne latérale,
- échange du groupement ammonium et de la chaîne latérale.

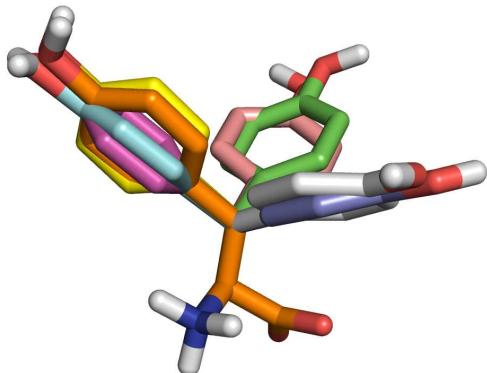
On effectue une légère minimisation d'énergie, et on écarte les rotamères qui ont une très forte énergie ou des structures très similaires entre eux. La procédure conduit finalement à 48 rotamères pour le D-TyrAMP (figure 4.8).

#### 4.1.4 Bibliothèque de rotamères pour la protéine

Dans notre étude, le squelette protéique est considéré comme fixe et l'espace des conformations des chaînes latérales est discréte, à l'aide d'une bibliothèque de rotamères. La bibliothèque utilisée est celle de Tufféry et coll. créée en 1995 [204] (figure 4.9). Elle contient un total de 219 rotamères, soit, en moyenne, 11 par type d'acide aminé.

FIGURE 4.7 – Les rotamères de la L-Tyr et de la D-Tyr.

(a) 8 rotamères de la L-Tyr



(b) 8 rotamères de la D-Tyr

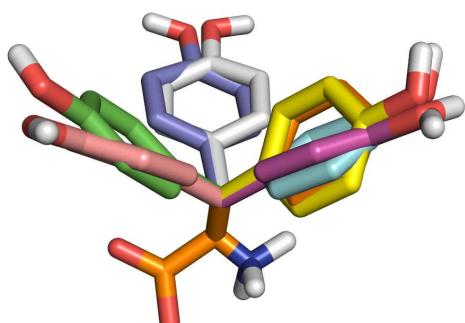
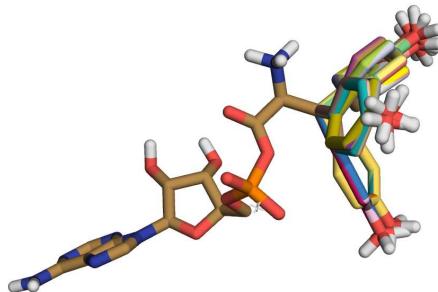


FIGURE 4.8 – Les rotamères du L-TyrAMP et du D-TyrAMP.

(a) 32 rotamères du L-TyrAMP



(b) 48 rotamères du D-TyrAMP

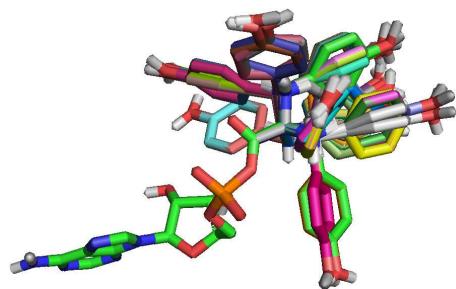


FIGURE 4.9 – Nombre de rotamères par type d’acide aminé. Ici, HSP représente l’histidine doublement protonée sur les atomes ND1 et NE2 alors que le résidu HIS est protonée sur ND1 seulement.

Acide aminé	nombre de rotamères
ALA	1
ASP	5
ASN	11
ARG	39
CYS	3
GLU	12
GLN	19
HIS	9
HSP	9
ILE	7
LEU	9
LYS	49
MET	17
PHE	4
SER	3
TYR	8
THR	3
TRP	8
VAL	3

## 4.2 Etape deux : calcul de la matrice d’énergie.

L’étape deux consiste à calculer toutes les énergies d’interactions entre paires de chaînes latérales de la protéine. Pour chaque paire, on va considérer successivement tous les types d’acides aminés autorisés à ces positions (qui peuvent être actives ou inactives) et tous les rotamères correspondants. On calcule également les énergies d’interaction entre le ligand et les résidus de la protéine, et entre chaque chaîne latérale et l’ensemble du squelette protéique.

La modélisation ne porte ici que sur une sphère composée des résidus distants de moins de 25 Å du ligand. La chaîne principale est maintenue fixe. Dans un premier temps, les acides aminés de la protéine sont classés en trois catégories :

- Actifs : ces résidus devront parcourir tous les types d’acides aminés possibles. Ils pourront adopter tous les rotamères possibles.
- Inactifs : pour ces résidus, la séquence est fixe mais tous les rotamères de l’acide aminé natif sont explorés.

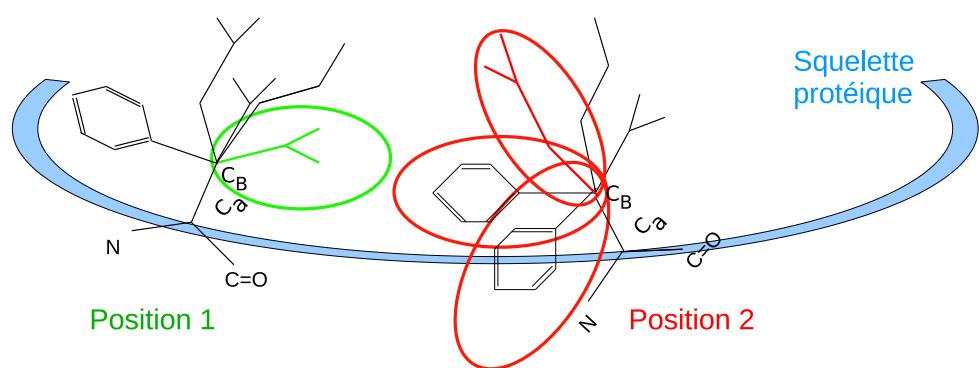
#### 4.2. ETAPE DEUX : CALCUL DE LA MATRICE D'ÉNERGIE.

- Gelés : ces résidus sont fixes tout au long de l'étude (séquence et rotamères). Gly, Cys et Pro sont connus pour modifier la structure de la chaîne principale, ce que nous ne souhaitons pas ; de ce fait ils ne mutent pas et leurs rotamères (Cys) sont maintenus fixes.

Pour une paire d'acides aminés actifs, on considère tous les types possibles de chaînes latérales et tous leurs rotamères, soit  $216 * 216$  possibilités. Pour modéliser une position active, on utilise un résidu "géant" qui a une unique chaîne principale mais possède les chaînes latérales de tous les types d'acides aminés, soit 216 chaînes latérales. On dispose un résidu géant sur chaque position à muter. On sélectionne ensuite une chaîne latérale sur le résidu géant et on positionne un rotamère particulier. On calcule ses interactions avec les autres résidus, puis on passe à un autre rotamère du résidu géant, pour finalement explorer tous les rotamères de tous les acides aminés sur les positions mutées. Pour chaque paire de résidus, qu'ils soient actifs ou inactifs, ainsi que pour le ligand, on calcule les énergies d'interactions chaîne latérale/squelette, et chaîne latérale/chaîne latérale (figure 4.10). Pour les résidus enfouis, une pondération du terme surfacique est appliquée. En effet, les résidus enfouis sont en contact, ils ont une surface en commun. On choisit donc ici de diviser par deux le terme d'énergie surfacique pour les résidus enfouis de manière à ne pas surestimer ce terme dans le calcul des énergies de paires.

Pour une paire individuelle de rotamères, voici comment se déroule le calcul. Tout d'abord, la chaîne latérale à la position 1 est minimisée 15 pas en prenant en compte les interactions suivantes : chaîne latérale 1 contre résidu 1 entier et chaîne latérale 1 contre squelette protéique. L'énergie de la chaîne latérale 1 avec le squelette protéique est alors calculée. Ensuite la chaîne latérale à la position 2 est minimisée 15 pas en tenant compte des interactions suivantes : chaîne latérale 2 contre résidu 2 entier et chaîne latérale 2 contre squelette protéique. Enfin, une dernière minimisation de 15 pas concerne l'interaction entre chaîne latérale 1 et 2 ; les interactions sont les suivantes: chaîne latérale 1 contre résidu 1 entier, chaîne latérale 2 contre résidu 2 entier, chaîne latérale 1 contre squelette protéique, chaîne latérale 2 contre squelette protéique, et chaîne latérale 1 contre chaîne latérale 2. A l'issu de cette dernière minimisation, l'énergie d'interaction entre chaîne latérale 1 et chaîne latérale 2 est calculée. Les énergies calculées sont composées

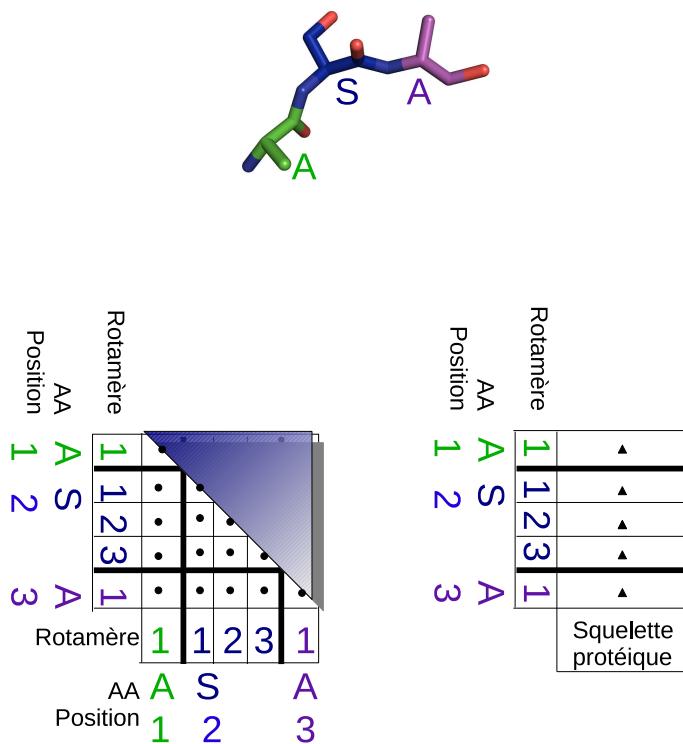
FIGURE 4.10 – Interaction d'une chaîne latérale à une position active (en vert) avec toutes les autres chaînes latérales d'une autre position active (en rouge).



## 4.2. ETAPE DEUX : CALCUL DE LA MATRICE D'ÉNERGIE.

de trois termes : les énergies non électrostatiques (énergie de liaison, d'angle, d'angle dièdre, de van der Waals), les énergies électrostatiques avec le terme de coulomb, enfin l'énergie correspondant à la surface exposée au solvant. Pour cette dernière énergie, on affecte aux résidus enfouis une pondération (cf. ci-dessus). A l'issu de ces calculs, les informations sont stockées dans la matrice d'énergie. On ne calcule qu'une demi-matrice en raison de la symétrie des données (figure 4.11). Au final, la matrice d'énergie contient les énergies de paires pour toutes les combinaisons d'acides aminés et de rotamères possibles dans la protéine, ainsi que l'énergie d'interaction de chaque rotamère avec le squelette peptidique entier. C'est à partir de cette matrice qu'on pourra explorer les meilleurs jeux de séquences/structures.

FIGURE 4.11 – Soit le tripeptide ASA, chaque point de la matrice d'énergie correspond à une énergie d'interaction entre deux chaînes latérales. Chaque triangle correspond à l'interaction entre une chaîne latérale et l'ensemble du squelette protéique.



## 4.3 Etape trois : optimisation des séquences et structures

La troisième étape est l'exploration combinatoire des séquences et structures. Elle est faite avec le programme Proteus, développé au sein du laboratoire par Anne Lopes, Marcel Schmidt am Busch, David Mignon et Thomas Simonson. L'exploration est faite avec l'algorithme heuristique. On réalise 100000 cycles heuristiques pour notre complexe. Le déroulement d'un cycle heuristique est le suivant. On part d'une séquence et d'un jeu de rotamères aléatoires. Tout est fixe sauf la première position. Pour cette position, on choisit à la fois le type d'acide aminé (parmi ceux autorisés) et le rotamère qui minimisent l'énergie libre de repliement. Cette modification est conservée et la position 2 est optimisée de la même façon. L'opération est répétée jusqu'au dernier acide aminé de la séquence. On recommence au début de la séquence, et on itère jusqu'à convergence. La séquence finale, sa structure et son énergie de repliement sont stockées, et le cycle est terminé (figure 4.12).

On dispose maintenant de 100000 jeux de séquences/structures ainsi que de leurs énergies de repliement. On peut déjà évaluer combien de fois chaque séquence a été prédite, sur les 100000 cycles, et ainsi voir si certaines séquences apparaissent fréquemment. Certaines séquences peuvent être prédites avec de nombreux jeux de rotamères différents. Il peut arriver aussi que plusieurs cycles heuristiques conduisent à exactement la même séquence et le même jeu de rotamères. Les solutions complètement redondantes sont éliminées.

## 4.4 Etape quatre : reconstruction des structures 3D et estimation d'affinités

### 4.4.1 Reconstruction des structures 3D

Il faut maintenant "reconstruire" les coordonnées cartésiennes des structures prédites à partir du squelette protéique fixe et des listes de rotamères optimisés. Après lecture du système sphérique initial, construit dans l'étape 1 de la procédure, on boucle sur tous les

FIGURE 4.12 – Déroulement d'un cycle heuristique. X symbolise le ligand.

1 : Une séquence/structure est tirée au hasard.

2 : La séquence et le rotamère de chaque position sont fixes exceptés ceux de la première position.

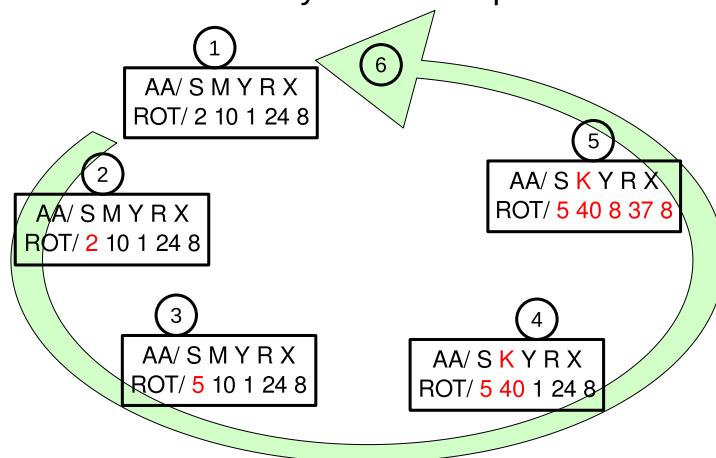
3 : Optimisation à plusieurs pas. Sélection du meilleur acide aminé et du meilleur rotamère pour la première position.

4 : Les modifications précédentes sont conservée. La structure est fixée sauf pour la position 2.

5 : La même chose est faite pour le reste de la séquence jusqu'à convergence.

6 : La séquence finale, les rotamères et l'énergie de repliement sont sauvegardés.

#### Déroulement d'un cycle heuristique:



acides aminés qui ont été optimisés, actifs et inactifs. Pour chacun d'eux, on positionne la chaîne latérale du rotamère prédict, grâce à une superposition de la chaîne principale correspondante avec la chaîne principale dans la protéine. Pour le ligand, on superpose la chaîne principale du rotamère prédict avec celle du ligand de la sphère initiale.

Après avoir positionné tous les rotamères, une légère minimisation de 15 pas à squelette protéique fixé permet de réduire les mauvais contacts entre chaînes latérales. On dispose maintenant de la structure reconstruite avec squelette protéique cristallographique, et séquence et chaînes latérales optimisées.

#### **4.4.2 Calcul d'affinité**

Pour ces structures, on peut calculer l'affinité, selon la procédure détaillée dans le chapitre 3 (Paramétrisation et validation du modèle de solvant CASA). L'énergie du complexe TyrRS:ligand est calculée, puis les énergies de la TyrRS et du ligand séparés sont calculées ; enfin, on estime l'affinité de la structure reconstruite pour le ligand :

$$E_{liaison} = E_{complexe} - (E_{protéine} + E_{ligand}) \quad (4.1)$$

Pour chaque séquence, on estime l'affinité de la TyrRS pour toutes les différentes conformations de chaînes latérales qui ont pu être prédictes. On calcule ensuite la moyenne de l'affinité obtenue pour ces différentes structures, ainsi que la variance et l'écart type. On note également l'affinité maximum et minimum.

#### **4.4.3 Réoptimisation des meilleures séquences et étude de leurs spécificités**

Les séquences de bonne affinité pour la D-Tyr ou le D-TyrAMP font l'objet d'analyses complémentaires. Pour chacune d'elles, on réalise une prédiction de structures à travers 20000 nouveaux cycles heuristiques avec la séquence maintenue fixe ; d'une part avec le ligand natif L-Tyr ou L-TyrAMP, et d'autre part avec la D-Tyr ou le D-TyrAMP. On obtient un grand échantillon de structures possibles pour la séquence étudiée avec les

différents ligands. En évaluant les affinités moyennes sur de nombreuses structures, on peut appréhender la spécificité de l'interaction protéine:ligand. Le  $\Delta\Delta G$  entre complexe TyrRS:L-Tyr et TyrRS:D-Tyr, par exemple, nous donne une mesure de la spécificité. A l'issu de ce processus, on sélectionne la structure de meilleure affinité pour le ligand natif et la structure de meilleure affinité pour le ligand modifié, afin de les étudier par dynamique moléculaire.

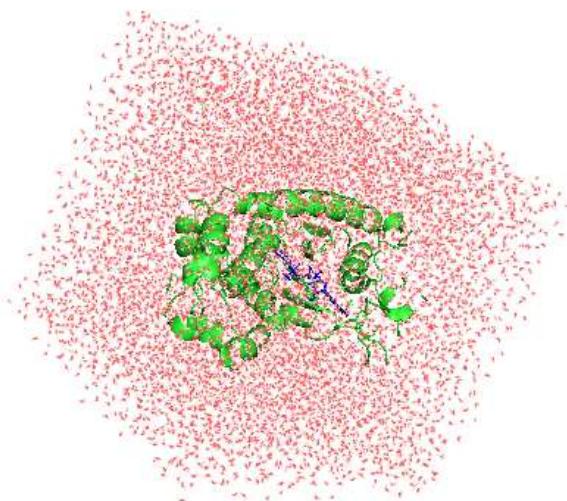
## **4.5 Etape cinq : étude par dynamique moléculaire**

### **4.5.1 Simulations de dynamique moléculaire**

Les dynamiques moléculaires sont réalisées par le programme CHARMM [105]. Le champ de force est celui de CHARMM22. L'algorithme utilisé est celui de Verlet. Les structures utilisées jusqu'ici (fragment sphérique de 25 Å de la TyrRS autour de son ligand) sont plongées dans une boîte de solvant de 68 Å de côté, comportant environ 6000 molécules d'eau et quelques contre-ions. On définit une zone tampon, vers l'extérieur de la sphère, pour y appliquer des contraintes structurales. Ainsi, les atomes de la protéine situées à des distances entre 20 Å et 25 Å du ligand sont soumis à des contraintes harmoniques, qui les retiennent proches de leur positions initiales. Les constantes de forces sont d'autant plus fortes que la distance au ligand est grande.

La structure est minimisée pendant 100 pas dans sa boîte de solvant. On adopte des conditions aux limites périodiques, en considérant un réseau de cubes images par symétrie de translation. La température est maintenue à 295 K par un thermostat de type Nosé-Hoover [219]. Les interactions électrostatiques à grandes distances sont traitées avec la méthode "Particule Mesh Ewald" [220]. La durée totale de la simulation est de 5 ns (figure 4.13).

FIGURE 4.13 – Tyrosyl-ARNt synthétase d'*E. coli* liée au L-Tyrosyl adénylate dans une boîte de solvant explicite pour la dynamique moléculaire. Grâce aux conditions aux limites périodiques, la boîte est répliquée dans toutes les directions. On simule donc en fait un réseau infini de boîtes cubiques.



#### 4.5.2 Estimation de l'énergie libre d'association protéine:ligand avec un modèle de Poisson-Boltzmann

La contribution électrostatique à l'énergie libre de fixation du ligand est obtenue en soustrayant l'énergie libre électrostatique du complexe à celle de la protéine et du ligand seuls. Les contributions non électrostatiques ne sont pas prises en compte. L'expression de l'énergie libre du complexe protéine:ligand en solution,  $G_{pl}$  a la forme suivante :

$$G_{pl} = \frac{1}{2} \sum_i q_i V_i^{pl} = \frac{1}{2} \sum_{i \in lig} q_i V_i^{pl} + \frac{1}{2} \sum_{i \in prot} q_i V_i^{pl} \quad (4.2)$$

Ici, la première somme est sur l'ensemble des atomes du complexe,  $q_i$  est la charge partielle de l'atome  $i$  et  $V_i^{pl}$  est le potentiel électrostatique total à l'atome  $i$  dans le complexe. Les deux termes à droite correspondent à la somme sur les atomes du ligand et à celle sur les atomes de la protéine, respectivement.

Voici comment se déroule l'estimation des potentiels  $V_i$ . On utilise une constante di-

#### *4.5. ETAPE CINQ : ÉTUDE PAR DYNAMIQUE MOLÉCULAIRE*

---

électrique de 4 pour le soluté et de 80 pour le solvant. La force ionique correspond à une concentration de 0.1 M de sel monovalent et à une température de 295 K. La surface séparant la protéine et l'eau est la surface moléculaire, calculée par déplacement d'une sphère de 2 Å de rayon le long de la surface protéique. On extrait plusieurs centaines de structures de la trajectoire de dynamique moléculaire. Pour chacune, on utilise une méthode numérique de résolution de l'équation de Poisson-Boltzmann. Cette méthode considère un maillage cubique qui discrétise l'espace autour de la protéine. La grille est de grandes dimensions par rapport au soluté : on utilise 181 cellules de 0.8 Å d'arrêté dans chaque direction. On impose un potentiel nul à toutes les extrémités du maillage. Le calcul est ensuite répété avec une grille plus fine et plus petite. Pour la nouvelle grille, les valeurs à la périphérie sont celles de la grille précédente qui l'englobe. L'énergie libre est alors calculée.



# Chapitre 5

## CPD de la TyrRS : Caractérisation des séquences et structures obtenues

Au cours de l'optimisation de la séquence et de la structure par CPD, se succèdent des milliers de cycles heuristiques. Nous analysons ici l'ensemble des séquences et structures générées. Nous comparons les ensembles obtenus avec différents ligands. Dans un premier temps, nous étudierons le cas de la séquence native. La séquence est fixée ; l'optimisation ne concerne que la structure. Nous comparerons les structures des complexes générés avec le ligand L-TyrAMP, le ligand D-TyrAMP ou sans ligand. Après avoir détaillé la variabilité structurale à séquence fixe, nous analyserons la variabilité des séquences obtenues. Nous nous intéresserons en particulier à la variabilité des structures prédites pour chaque mutant. Enfin, la synthèse de toutes ces informations nous permettra de caractériser les ensembles structuraux prédits par notre modèle d'évolution dirigée.

### 5.1 TyrRS native d'*E. coli* avec différents ligands

Voyons tout d'abord la variabilité des structures obtenues à séquence native fixe, ainsi que les différences entre structures prédites avec différents ligands.

### 5.1.1 Comparaison à la structure cristallographique

La TyrRS native a été étudiée ici sous plusieurs aspects. Des générations de structures sur 100000 cycles heuristiques ont été conduites sans ligand, avec le ligand natif L-TyrAMP, et avec le ligand modifié D-TyrAMP. De nombreux cycles ont convergé vers les mêmes structures. On obtient finalement 110 structures différentes avec le ligand L-TyrAMP, 68 avec le ligand D-TyrAMP et 842 pour l'apoenzyme. Dans ce dernier cas, l'absence de ligand donne une plus grande liberté de conformations aux acides aminés du site catalytique ce qui explique qu'on puisse prédire beaucoup plus de structures. Etudions à présent les différences entre les structures générées et la structure cristallographique de code PDB 1VBM [215]. Nous regarderons le RMSD des coordonnées, puis les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  (angles de torsion des chaînes latérales). Nous considérerons une sphère de 15 Å, puis de 5 Å autour du ligand, afin d'évaluer l'influence directe du ligand sur les modifications structurales prédictes. Pour ces deux sphères, les complexes et la forme apoenzyme sont comparés à la structure cristallographique.

**Etude d'une sphère de 15 Å** Nous considérons d'abord les déviations structurales observées pour les chaînes latérales situées dans une sphère de 15 Å autour du ligand, représentées dans la figure 5.1. Cette figure montre les déviations sous forme d'histogrammes. Pour la séquence native liée au L-TyrAMP, soit 110 structures, le RMSD des coordonnées montre que les structures générées forment deux groupes. Un groupe de structures est très proche de la structure cristallographique, avec un RMSD entre 0.6 et 1.0 Å. Un autre groupe de structures est un peu différent, son RMSD est entre 1.2 et 1.6 Å. Ce deuxième groupe de structures doit correspondre à un autre minimum d'énergie que celui de la structure cristallographique. Ceci est également visible dans l'étude des angles de torsions  $\chi_1$  et  $\chi_2$ , où l'on distingue également deux groupes de structures distinctes pour  $\chi_2$ .

Dans le cas du ligand D-TyrAMP, les structures générées sont plus éloignées de la structure cristallographique, le RMSD des coordonnées variant de 0.6 à 1.3 Å, et davantage de structures ont un RMSD supérieur à 1 Å. Un premier groupe de structures a un RMSD

entre 0.6 et 0.8 Å, un deuxième groupe plus variable a un RMSD entre 0.8 et 1.3 Å. Les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  ont des valeurs supérieures à celles obtenues avec le ligand natif. On distingue deux groupes structuraux pour  $\chi_1$ .

On considère enfin l'apoenzyme, soit 843 structures. Le RMSD des coordonnés entre structures sans ligand et structure cristallographique est compris entre 0.6 et 1.1 Å. Les structures sont donc également très proches de la structure native. Les valeurs des RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  montrent non seulement qu'il y a plusieurs groupes structuraux, mais aussi que certaines structures ont des angles très différents de la structure cristallographique, avec notamment des valeurs de RMSD de  $\chi_2$  élevées.

**Etude d'une sphère de 5 Å** Voyons maintenant l'influence directe du ligand sur les résidus qui l'entourent (figure 5.2). Avec le ligand L-TyrAMP (110 structures) les RMSD des coordonnées sont du même ordre que pour la grande sphère : entre 0.6 et 1.3 Å. On voit donc que le site actif est le lieu des principales modifications structurales à travers les structures générées. Pour les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$ , ils sont en moyenne plus grands que dans la grande sphère. Ceci montre également que le site actif est le siège principal des modifications structurales proposées par notre optimisation.

Avec le ligand D-TyrAMP (69 structures), le RMSD des coordonnées sur la petite sphère est un peu plus grand que la moyenne sur la sphère de 15 Å : il varie de 0.6 à 1.6 Å. Les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  sont également plus grands que pour la grande sphère.

Pour les structures d'apoenzyme, le RMSD est aussi le même autour du ligand que pour la sphère de 15 Å, il varie entre 0.6 et 1.3 Å, mais la majorité des structures ont un RMSD supérieur à 1.1 Å alors que pour la grande sphère, la majorité des structures était en dessous de ce seuil. Donc, les chaînes latérales les plus modifiées sont celles du site actif. Les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  sont également supérieurs à ceux de la grande sphère et à ceux des structures modélisées avec un ligand. On constate que la structure est plus variable dans cette sphère du site actif pour les structures d'apoenzyme : l'absence de ligand laisse les chaînes latérales se réarranger dans cette région.

FIGURE 5.1 – Sphère de 15 Å autour du ligand. Référence : structure cristallographique (1VBM)

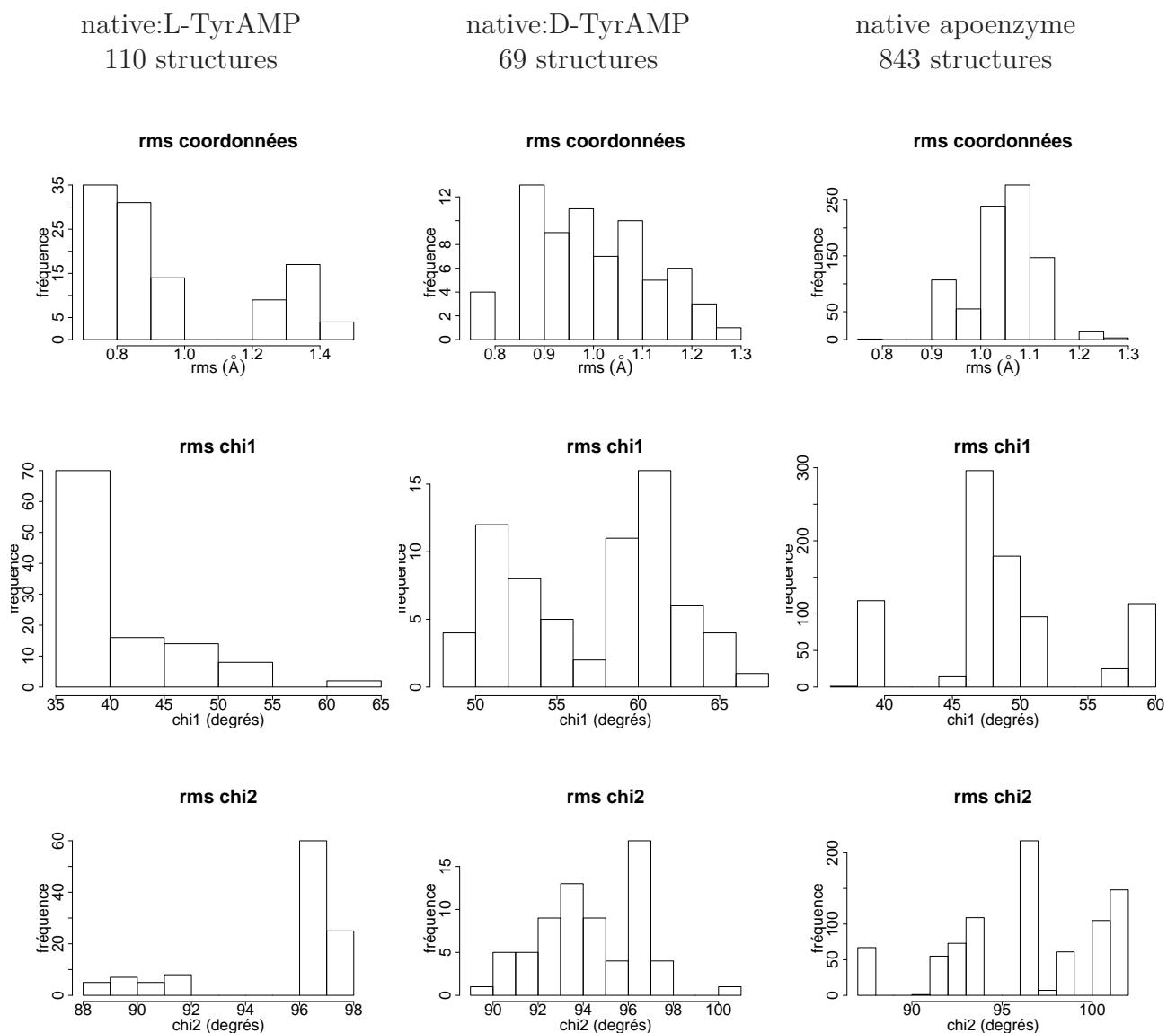
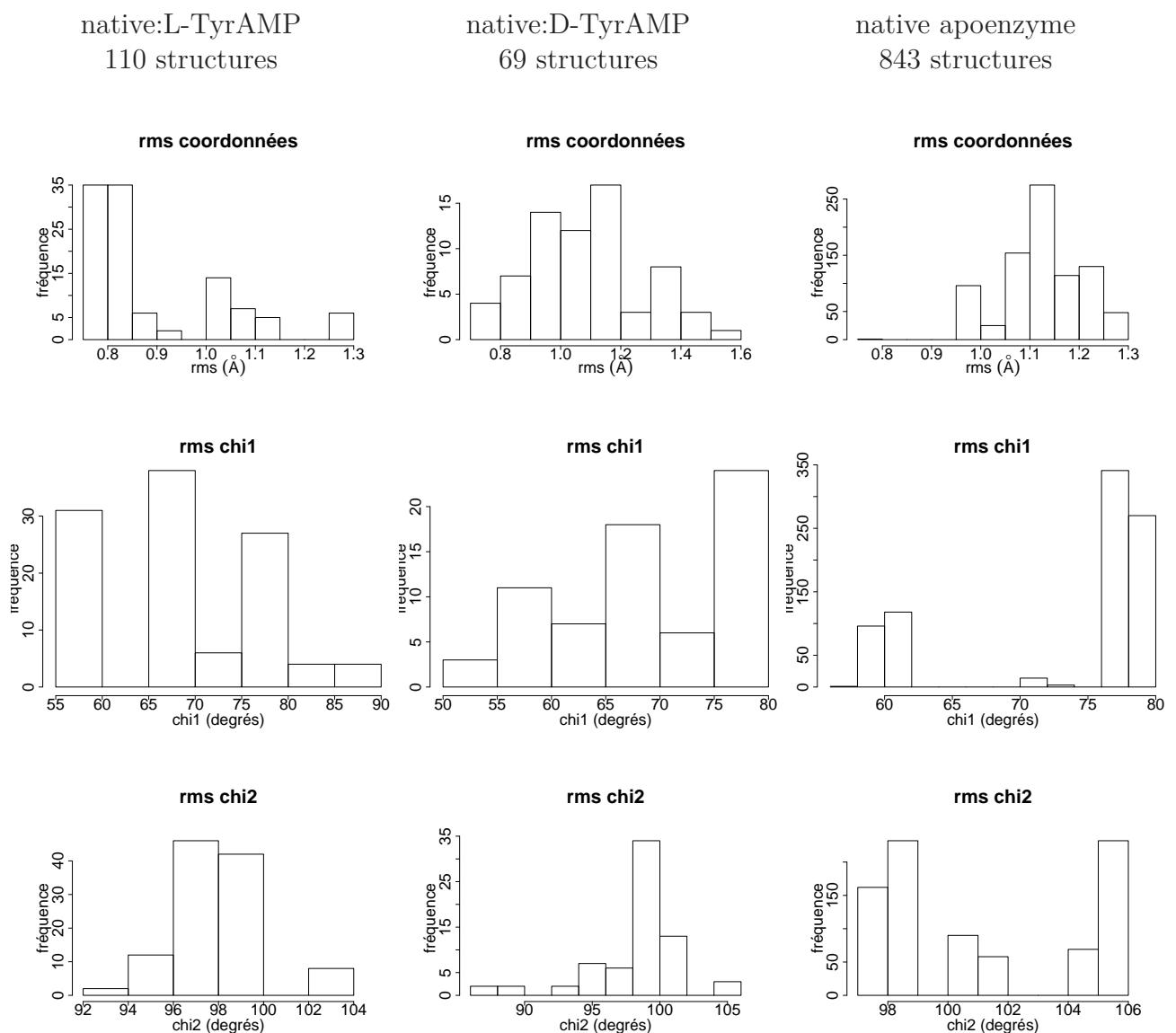


FIGURE 5.2 – Sphère de 5 Å autour du ligand. Référence : structure cristallographique (1VBM)



### 5.1.2 Variabilité des structures selon le ligand.

Pour caractériser l'effet du ligand, nous allons maintenant comparer entre elles les structures générées avec L-TyrAMP, D-TyrAMP, ou sans ligand.

La figure 5.3 donne l'histogramme des RMSD, en prenant comme référence le groupe de structures générées avec le ligand L-TyrAMP. On considère d'abord les chaînes latérales situées à moins de 15 Å du ligand. On compare d'abord toutes les structures de séquence native liées au L-TyrAMP (soit 110 structures), aux autres structures du même groupe. Les RMSD varient de 0 à 1.2 Å, avec trois groupes de structures distincts. Les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  montrent également trois groupes de structures. On compare ensuite les structures de séquence native liées au D-TyrAMP (soit 69 structures), aux 110 structures liées au L-TyrAMP : ici, le RMSD varie de 0.2 à 1.6 Å, il est difficile de déterminer des groupes ; les structures sont plus variables. Les comparaisons des RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  montrent également qu'on ne peut pas distinguer de groupes. Enfin, on compare les structures de séquence native sous forme d'apoenzyme (soit 843 structures), aux 110 structures liées au L-TyrAMP : ici, les différences entre structures sont plus marquées, il y a une plus grande variabilité structurale dans les prédictions. Le RMSD des coordonnées varie entre 0.6 et 1.6 Å et on distingue différents groupes.

Quand on restreint l'étude précédente à une sphère de 5 Å autour du ligand, on peut faire les mêmes observations, ce qui témoigne du fait que le site catalytique est le lieu qui concentre les principales modifications structurales parmi nos ensembles de structures (figure 5.4).

La figure 5.5 donne l'histogramme des RMSD en prenant comme référence le ligand D-TyrAMP. On considère les chaînes latérales à moins de 15 Å du ligand. Nous comparons d'abord les structures de TyrRS de séquence native liées au D-TyrAMP (soit 69 structures) à elles-mêmes. Le RMSD des coordonnées varie de 0.0 à 1.2 Å. Les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  sont répartis uniformément sur l'intervalle. Les structures sont très différentes les unes des autres. On a ainsi généré de nombreuses structures différentes avec un même ligand. On compare ensuite toutes les structures de TyrRS de séquence native

## 5.1. TYRRS NATIVE D'*E. COLI* AVEC DIFFÉRENTS LIGANDS

---

FIGURE 5.3 – Sphère de 15 Å autour du ligand. Référence : [native:L-TyrAMP] (110 structures).

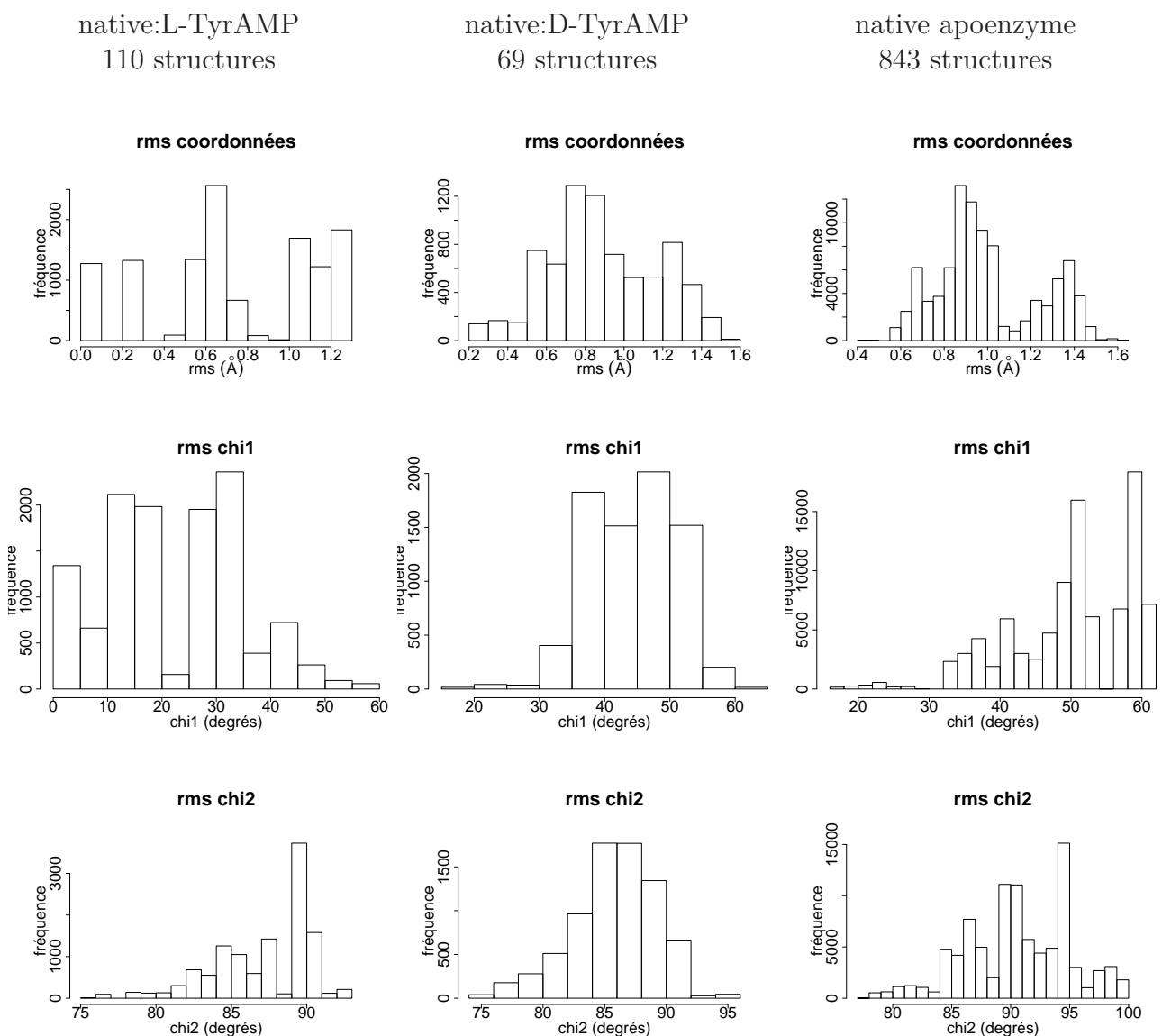
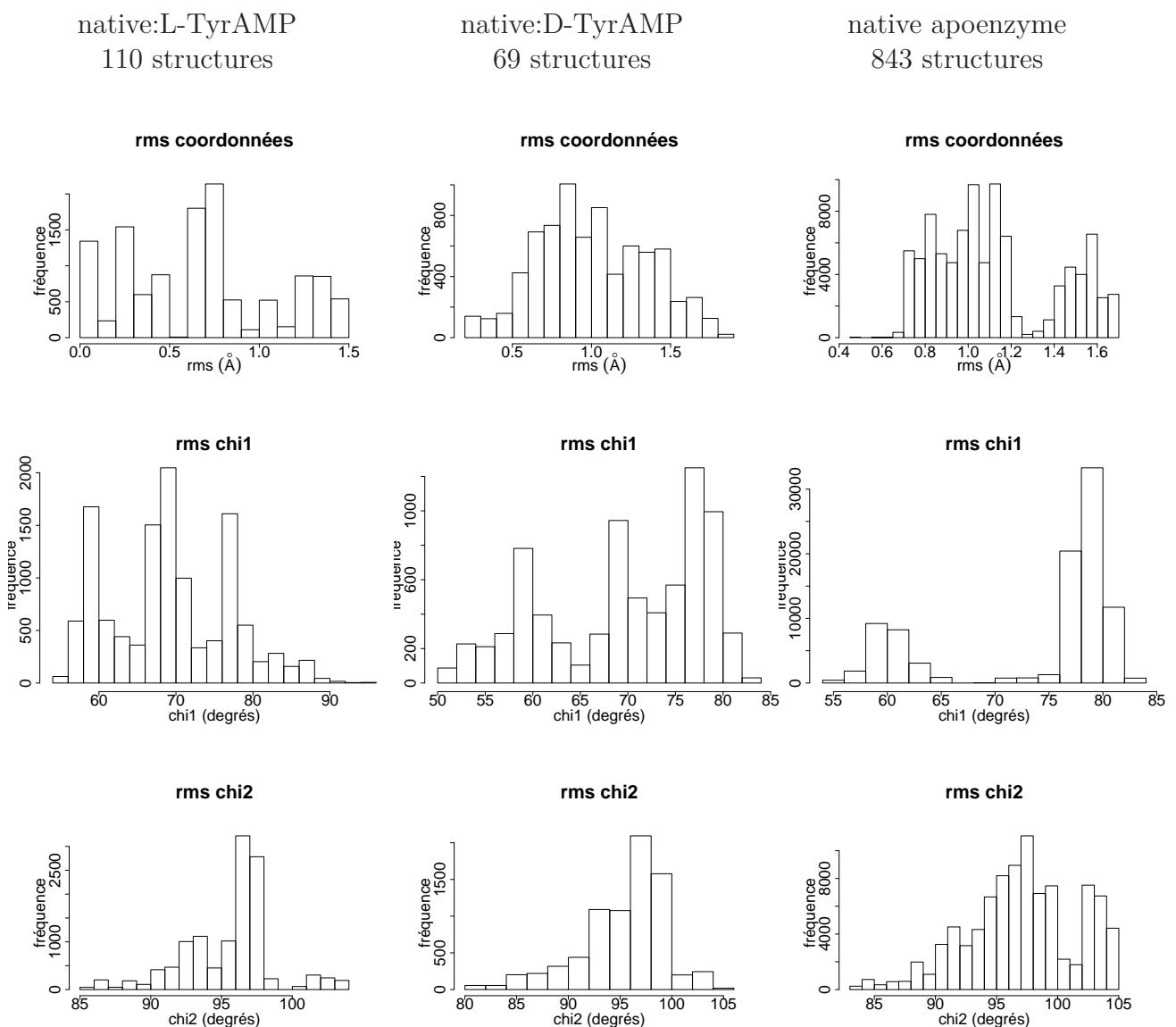


FIGURE 5.4 – Sphère de 5 Å autour du ligand. Référence : [native:L-TyrAMP] (110 structures).



### 5.1. TYRRS NATIVE D'*E. COLI* AVEC DIFFÉRENTS LIGANDS

liées au D-TyrAMP (soit 69 structures) aux 110 structures avec le ligand L-TyrAMP. Le RMSD des coordonnées varie de 0.2 à 1.6 Å ; les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  sont réparties uniformément sur les intervalles 30-70 degrés et 0-65 degrés, respectivement. On compare, enfin, toutes les structures de TyrRS native liées au D-TyrAMP (soit 69 structures), aux 843 structures sans ligand. Le RMSD des coordonnées varie de 0.2 à 1.4 Å ; les RMSD des angles de torsions  $\chi_1$  et  $\chi_2$  sont réparties uniformément sur les intervalles 75-105 degrés et 75-110 degrés, respectivement. Les structures sont donc très différentes selon qu'il y a un ligand ou pas.

Quand on restreint l'étude précédente à une sphère de 5 Å autour du ligand, on peut faire les mêmes observations, ce qui témoigne encore une fois du fait que le site catalytique est le lieu qui concentre les principales modifications de structures (figure 5.6).

Les figures 5.7 et 5.8, enfin, donnent les rmsd en prenant l'apoenzyme comme référence. Les résultats sont similaires à ceux décrits ci-dessus.

En résumé, notre modèle d'évolution dirigée réussit à générer des structures qui s'adaptent effectivement au ligand. Dans le cas de l'apoenzyme, beaucoup plus de structures sont générées, car les acides aminés du site actifs ont plus d'espace, et peuvent prendre différentes conformations. La comparaison des structures de l'apoenzyme et des complexes avec ligand L-TyrAMP montre que des groupes de structures ont été identifiés, centrés autours de quelques minima locaux d'énergie, à la fois pour l'apoenzyme et pour le complexe TyrRS:L-TyrAMP. Il y a plus de variabilité pour les structures des complexes TyrRS:D-TyrAMP ; en revanche, elles semblent être groupées autour d'un seul minimum local d'énergie (figure 5.9, 5.10 et 5.11 ).

FIGURE 5.5 – Sphère de 15 Å autour du ligand. Référence : [native:D-TyrAMP] (69 structures).

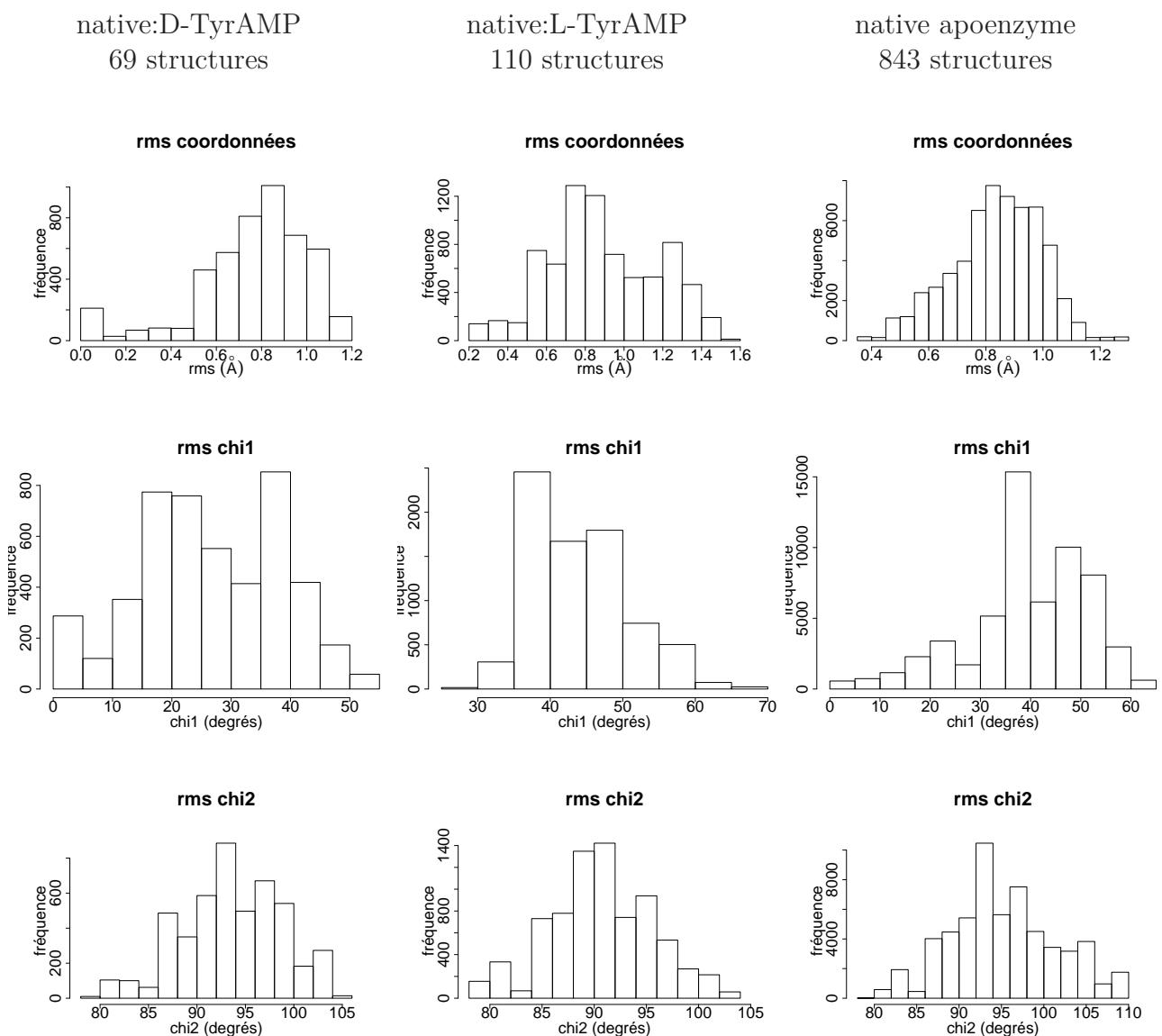


FIGURE 5.6 – Sphère de 5 Å autour du ligand. Référence : [native:D-TyrAMP] (69 structures).

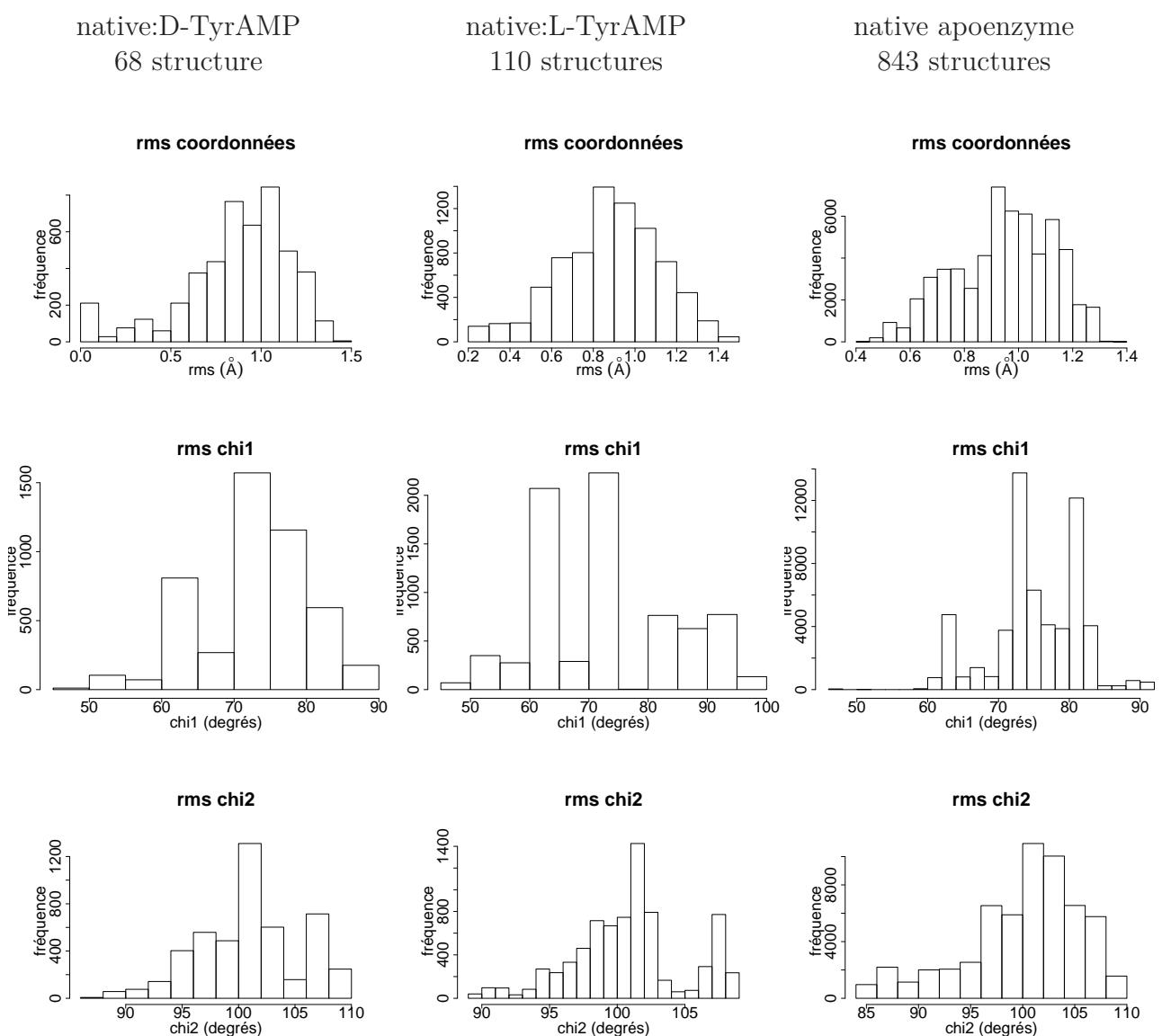
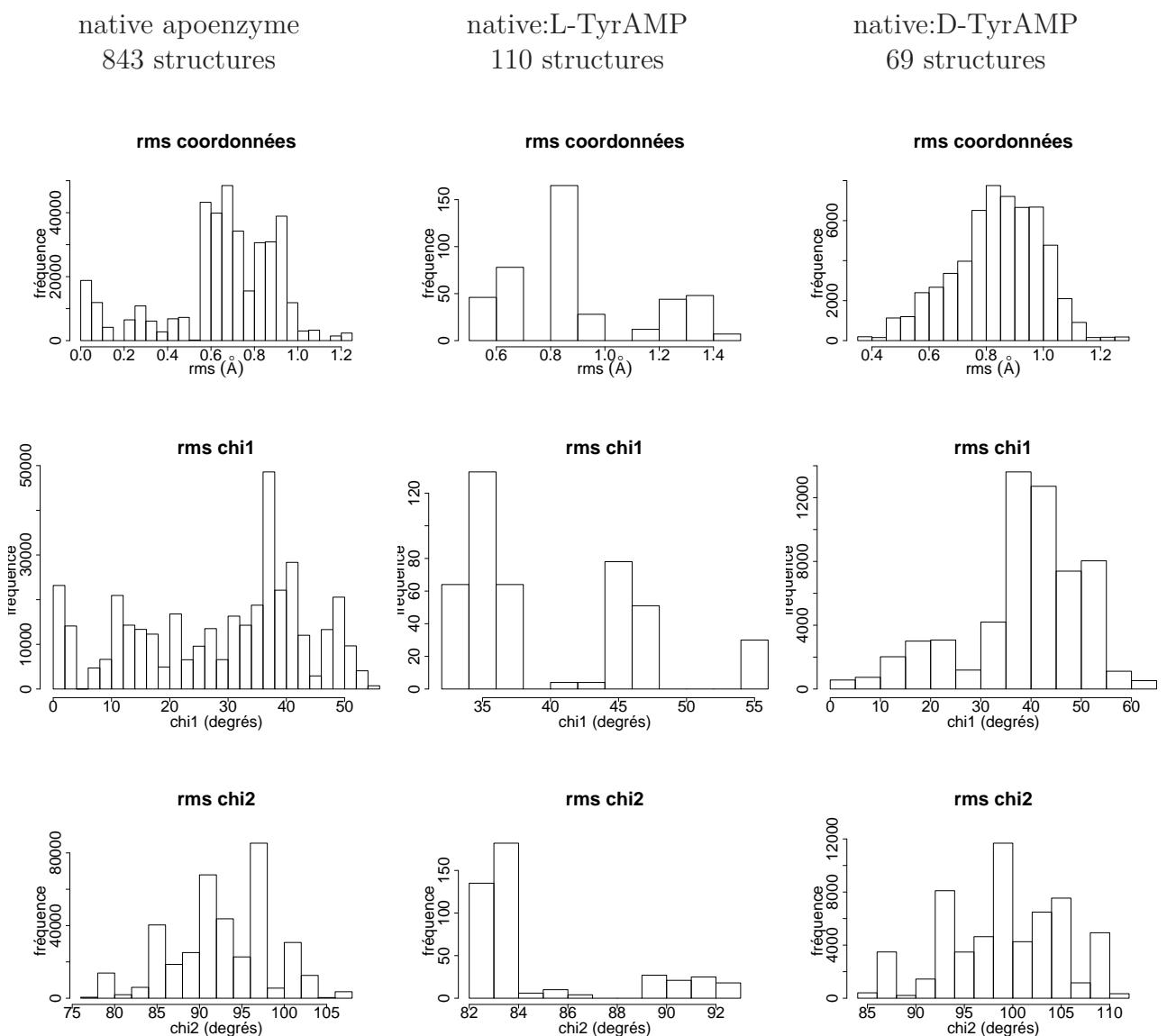


FIGURE 5.7 – Sphère de 15 Å autour du ligand. Référence : apoenzyme native (843 structures).



## 5.1. TYRRS NATIVE D'*E. COLI* AVEC DIFFÉRENTS LIGANDS

---

FIGURE 5.8 – Sphère de 5 Å autour du ligand. Référence : apoenzyme native (843 structures).

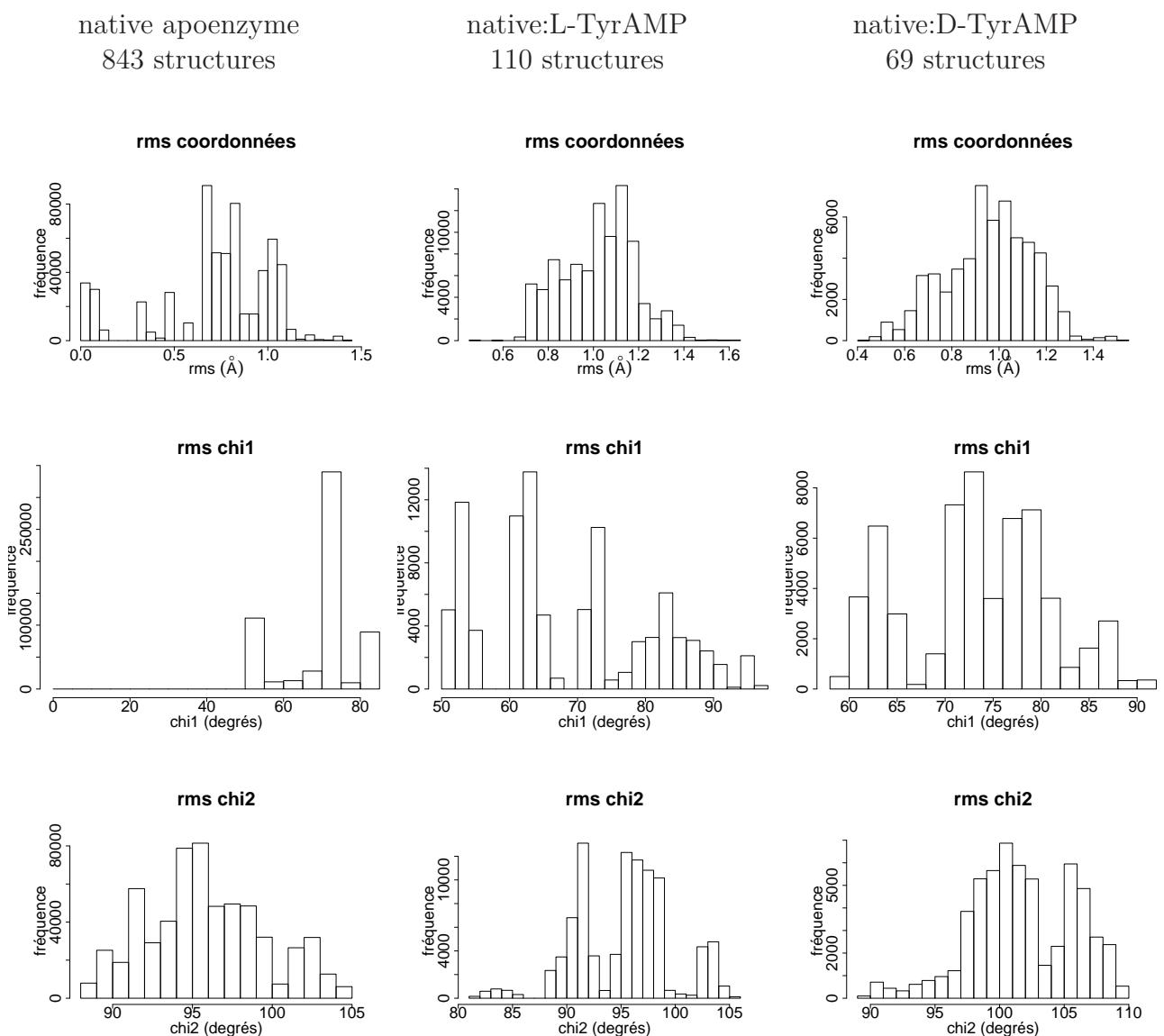


FIGURE 5.9 – Superposition de dix structures reconstruites pour le complexe TyrRS:L-TyrAMP. Sphère de 5Å autour du site catalytique.

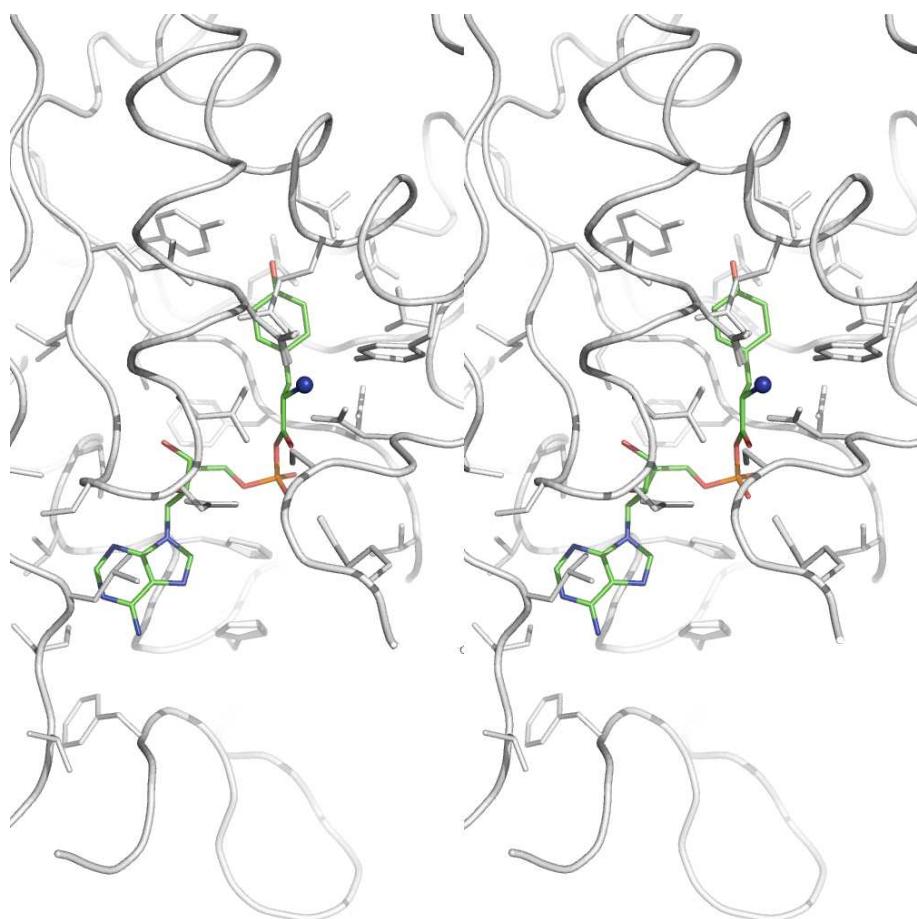


FIGURE 5.10 – Superposition de dix structures reconstruites pour le complexe TyrRS:D-TyrAMP. Sphère de 5Å autour du site catalytique.

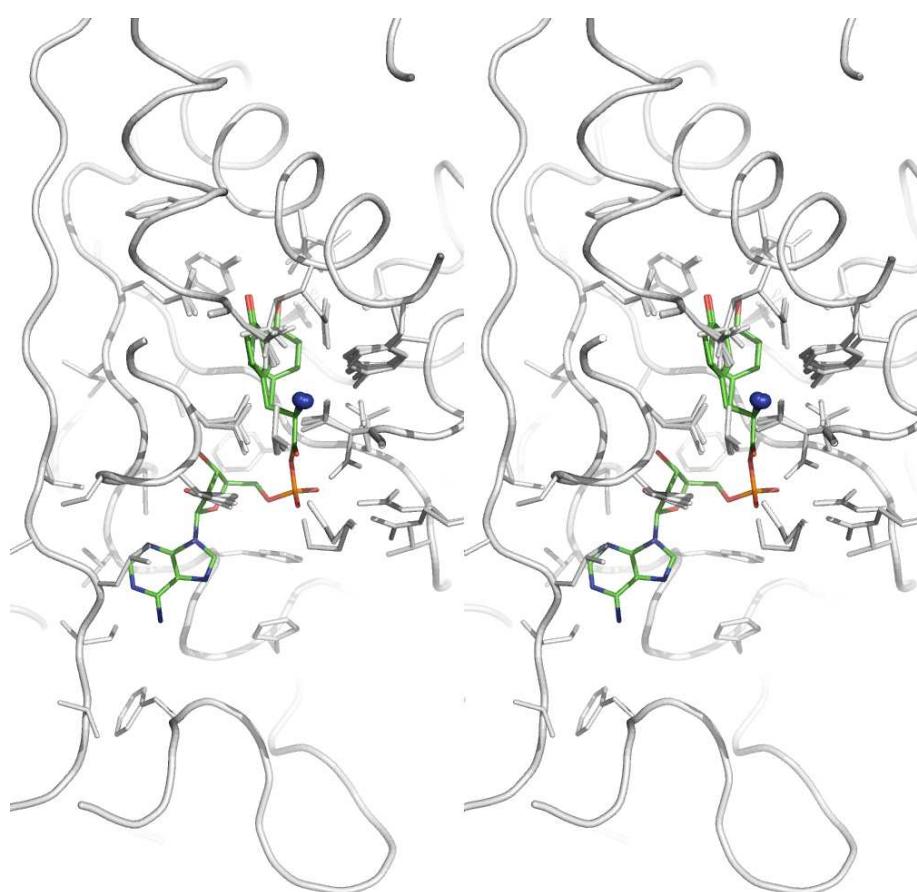
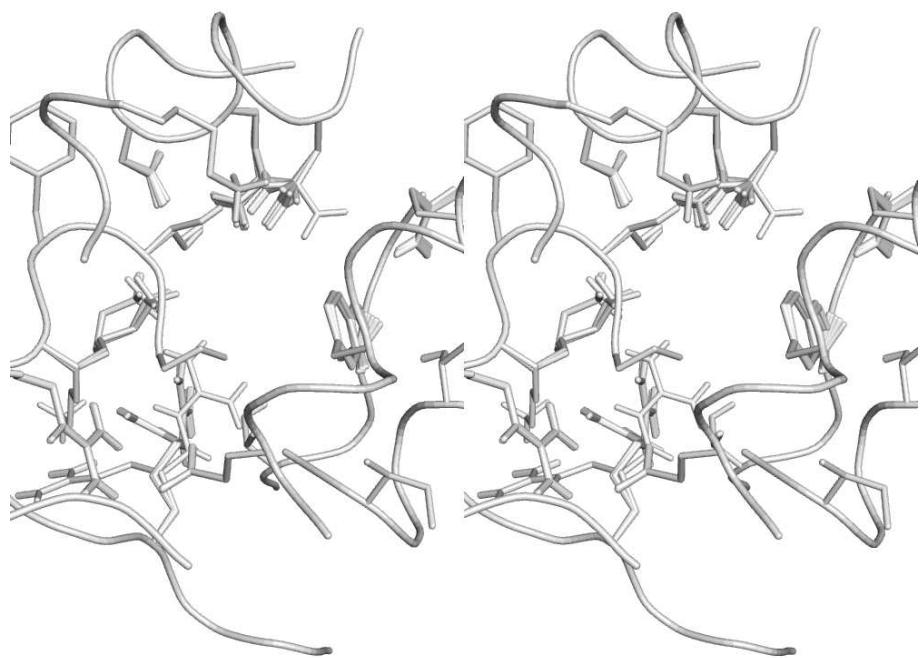


FIGURE 5.11 – Superposition de dix structures reconstruites pour l'apoenzyme. Sphère de 5Å autour du site catalytique.



## 5.2 Evolution dirigée de la TyrRS : analyse générale des séquences obtenues

Nous considérons maintenant les résultats obtenus avec quatre positions "actives", qui sont soumises à une mutagénèse combinatoire. Cette analyse des séquences et structures prédictes est faite pour les TyrRS d'*E. coli* et de *M. jannaschii*, en liaison avec le L-TyrAMP et le D-TyrAMP.

Pour la TyrRS d'*E. coli*, les mutations concernent les positions 81, 175, 179, et 201. Pour la TyrRS de *M. jannaschii*, les mutations concernent les positions 137, 155, 172 et 173. La procédure de CPD utilisée ici est résumée dans la figure 5.12. Les séquences et structures ont été optimisées en générant 100000 cycles heuristiques avec chaque ligand (L-TyrAMP et D-TyrAMP). Après élimination des séquences et une structures redondantes, (prédictes plusieurs fois), toutes les structures uniques sont sélectionnées et analysées. Les coordonnées 3D de chaque séquence/structure optimisée sont reconstruites à partir du squelette protéique (fixe) et des rotamères prédicts. L'affinité de la séquence (moyenne sur plusieurs structures) est estimée. Les différences d'énergies libres de liaison entre complexe avec ligand modifié et complexe avec ligand natif, donne une indication sur la spécificité d'interaction des séquences prédictes. Dans le chapitre suivant, les séquences qui apparaissent à la fois de bonne affinité et de bonne spécificité pour le ligand non naturel (D-TyrAMP) seront étudiées par dynamique moléculaire.

Fig. 5.12.

Les séquences obtenues à l'issue de la mutagénèse combinatoire sont montrées dans la figure 5.13. Les séquences correspondent aux quatres complexes suivant : *Eco*-TyrRS:L-TyrAMP, *Eco*-TyrRS:D-TyrAMP, *Mja*-TyrRS:L-TyrAMP, *Mja*-TyrRS:D-TyrAMP. Pour chaque organisme, la première colonne donne le ligand avec lequel a été prédict le mutant ; la deuxième colonne donne la séquence du mutant aux quatre positions variables. Certaines séquences sont prédictes avec les deux ligands, et apparaissent déjà peu spécifiques. Elles sont indiquées dans la partie centrale de la figure. Les positions des mutations dans la séquence pour chaque organisme sont indiquées en bas de la figure. La séquence native

FIGURE 5.12 – Schéma détaillé de la procédure de CPD appliquée.

Mutagénèse combinatoire à travers 100000 cycles heuristiques

Reconstruction des structures 3D

Calcul d'affinité protéine-ligand

Réoptimisation de toutes les séquences prédites:

- 20000 cycles heuristiques (séquence fixe)
- Reconstruction des structures 3D
- Calcul d'affinité

Sélection des séquences de bonne affinité et spécificité pour le ligand non naturel

Simulation des complexes TyrRS:Ligand par dynamique moléculaire

Affinité estimée avec un modèle diélectrique continu (méthode “PBFE”)



## **5.2. EVOLUTION DIRIGÉE DE LA TYRRS : ANALYSE GÉNÉRALE DES SÉQUENCES OBTENUES**

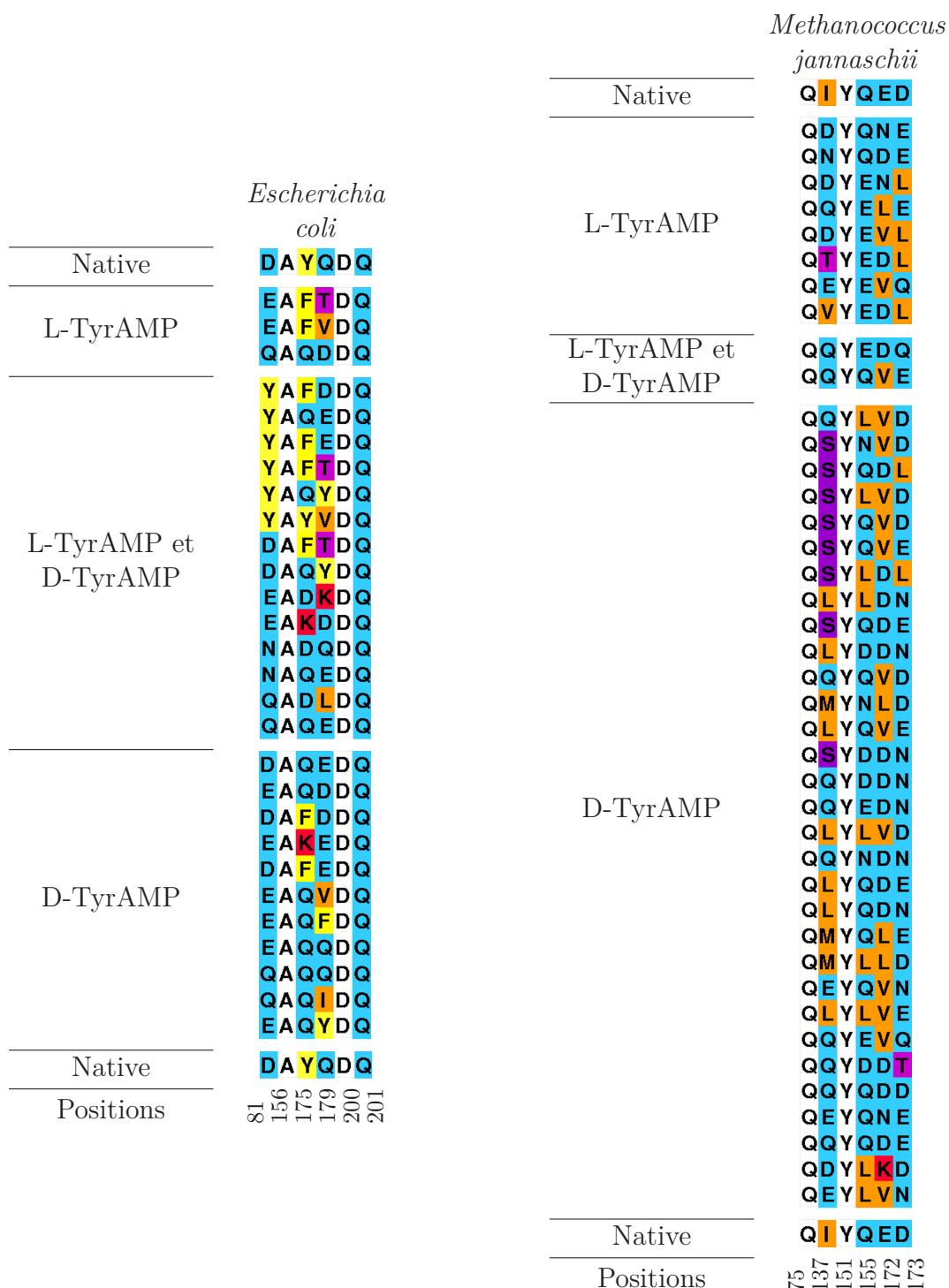
---

est rappelée à titre comparatif.

Dans le cas d'*E. coli*, 28 séquences mutantes ont été identifiées, dont 14 à la fois avec le L-TyrAMP et le D-TyrAMP, ce qui laisse penser qu'elles ne sont pas très spécifiques. Des lysines en positions 175 et 179 apparaissent avec le ligand D-TyrAMP. Des acides aminés hydrophobes L, V, et I à la position 179 apparaissent avec le ligand D-TyrAMP. Des phénylalanines et des tyrosines sont prédictes aux positions 81, 175 et 179, autant avec le L-TyrAMP qu'avec le D-TyrAMP. Les forces hydrophobes stabilisent le cycle aromatique du ligand sans favoriser la forme L ou la forme D. Un fait remarquable est que la position 201 n'est jamais mutée, bien qu'elle soit autorisée à le faire. L'observation des structures a conduit à identifier une interaction très forte entre l'aspartate 200 et la glutamine 201, qui expliquerait que cet acide aminé soit systématiquement prédict sous sa forme native. L'analyse détaillée des structures obtenues sera faite plus loin.

Dans le cas de *M. jannaschii*, 41 séquences sont prédictes. Deux séquences seulement sont prédictes à la fois avec le L-TyrAMP et avec le D-TyrAMP ; on remarque qu'elles sont très proches de la séquence native. 31 séquences sont prédictes avec le D-TyrAMP. À la position 137, des sérines sont prédictes avec le D-TyrAMP, mais une thréonine est prédict également avec le L-TyrAMP. Toujours à la position 137, on trouve de nombreux acides aminés apolaires : leucine, méthionine et valine, qui rappellent l'alanine qu'on trouve à la position équivalente chez *E. coli*. On retrouve parfois des acides aminés polaires à cette position, à la fois avec le L-TyrAMP et le D-TyrAMP, alors que l'acide aminé natif est une isoleucine. Pour la position 155, on retrouve des acides aminés polaires, comme chez la native, mais également des leucines, avec le D-TyrAMP uniquement. Pour la position 172, on trouve des acides aminés polaires, comme chez la native, mais également des leucines et des valines, aussi bien avec le L-TyrAMP qu'avec le D-TyrAMP. Une lysine est prédict une fois à cette position, en liaison avec le D-TyrAMP. On y retrouve également l'aspartate qui est l'acide aminé à la position équivalente chez *E. coli*. À la position 173, on retrouve des acides aminés polaires, comme chez la native, mais aussi apolaires. On retrouve également une thréonine, prédict avec le D-TyrAMP, et la glutamine qui correspond à l'acide aminé à la position équivalente chez *E. coli*.

FIGURE 5.13 – Séquences prédictes en fonction du ligand. Les positions qui mutent sont colorées selon le type physico-chimique de l'acide aminé prédit. Les positions non colorées sont inactives (ne mutent pas), mais sont homologues à une position active dans l'autre organisme.



## 5.3 Ré-optimisation de toutes les séquences.

Ici, nous étudions individuellement chaque séquence prédite. Chacune sera "ré-optimisée" avec soit le ligand L-TyrAMP, soit le ligand D-TyrAMP. La ré-optimisation consiste à faire une exploration de l'espace conformationnel à l'aide de l'algorithme heuristique, qui échantillonne donc les rotamères, à séquence fixe. Ceci permettra d'évaluer plus en détail l'affinité et la spécificité de chacune d'entre elles.

### 5.3.1 TyrRS d'*E. coli*

Pour chaque séquence prédite ci-dessus, soit avec le L-TyrAMP, soit avec le D-TyrAMP, on ré-optimise la structure par 20000 cycles heuristiques, avec les deux ligands (quel que soit le ligand avec lequel la séquence a été prédite). Les résultats sont présentés dans les deux tableaux suivants : table 5.1 pour la TyrRS liée au L-TyrAMP et table 5.2 pour la TyrRS liée au D-TyrAMP. Dans ces deux tableaux, la première colonne correspond au ligand avec lequel la séquence a été prédite initialement. La deuxième colonne correspond aux acides aminés prédits aux 4 positions actives. La troisième colonne donne le nombre de structures différentes obtenues à travers les 20000 cycles heuristiques. La quatrième colonne donne la charge électrostatique sur ces quatre positions. La cinquième colonne est l'énergie de dépliement de la structure, en  $\text{kcal} \cdot \text{mol}^{-1}$ . La sixième colonne est l'écart type de l'énergie de dépliement sur toutes les structures de cette même séquence. La septième colonne donne l'énergie de dépliement maximum pour la séquence ; la huitième donne l'énergie minimum. La neuvième donne l'affinité, moyennée sur toutes les structures échantillonées. La dixième donne l'écart type ; la onzième, l'affinité maximum ; la douzième l'affinité minimum. La treizième donne le numéro de la structure de meilleure affinité, de façon à pouvoir l'identifier et l'étudier ensuite par dynamique moléculaire.

TABLE 5.1 – Ré-optimisation des mutants. TyrRS d'*E. coli* avec L-TyrAMP. Les données sur la séquence native sont rappelées. Un filtre est appliqué pour ce tableau, où figure seulement les complexes qui ont une énergie libre de liaison supérieure à -70 kcal.mol<sup>-1</sup>.

ligand	séquence	occurrence	charge	énergie moyenne	écart -type	énergie max.	énergie min.	affinité moyenne	écart -type	affinité max.	affinité min.	meilleure structure
	DYQQ	110	-1	-713.51	77.81	-641.62	-867.19	-52.85	2.16	-47.74	-56.85	16
D	DFDQ	70	-2	-661.99	17.82	-637.99	-689.78	-58.92	0.89	-56.51	-61.69	7
D	DQEQQ	59	-2	-661.01	17.91	-636.98	-689.33	-58.61	1.14	-56.41	-61.60	5
D	EQDQ	58	-2	-663.23	17.83	-639.49	-689.51	-57.61	2.07	-54.47	-61.16	39
L	EFTQ	100	-1	-647.33	240.35	-635.28	-841.95	-41.75	15.16	-38.84	-60.63	37
D	DFEQ	112	-2	-746.48	80.56	-637.54	-859.52	-56.88	1.51	-53.89	-60.17	6
L	EFVQ	96	-1	-683.66	178.60	-634.89	-841.46	-47.06	11.51	-43.81	-60.06	39
L-D	YFDQ	64	-1	-677.85	170.35	-641.32	-968.57	-47.55	14.55	-25.24	-59.16	7
L-D	NQEQQ	96	-1	-660.12	16.99	-638.71	-688.56	-54.19	1.67	-50.08	-58.43	17
L-D	YFEQQ	149	-1	-789.31	151.07	-641.67	-1030.72	-45.33	12.98	-11.80	-57.21	67
L-D	YQEQQ	90	-1	-696.97	101.56	-641.14	-880.24	-43.44	12.42	-21.25	-57.07	48
D	EQQQQ	116	-1	-718.87	81.26	-639.17	-865.88	-50.92	2.67	-42.19	-55.77	89
D	EKEQQ	129	-1	-736.58	81.27	-639.47	-861.19	-51.10	1.70	-43.95	-55.24	7
L-D	EKDQ	57	-1	-659.50	17.49	-636.81	-688.23	-52.05	1.21	-48.15	-54.14	33
L-D	QQEQQ	125	-1	-742.03	79.82	-641.18	-864.16	-48.76	2.00	-43.56	-53.37	2
D	EQVQ	62	-1	-661.51	18.06	-636.95	-688.69	-50.62	0.98	-48.29	-53.30	11
L	QQDQ	52	-1	-664.48	17.52	-640.00	-689.84	-48.90	2.54	-43.31	-53.00	2
L-D	NDQQ	123	-1	-662.42	16.85	-641.62	-692.11	-47.79	1.62	-44.74	-52.86	95
L-D	DFTQ	87	-1	-657.95	17.19	-636.43	-687.83	-48.57	1.54	-45.24	-52.26	8
D	QQQQQ	139	0	-742.76	86.24	-643.03	-865.99	-46.96	2.46	-38.17	-52.19	91
D	EQYQ	260	-1	-669.93	18.36	-642.88	-705.88	-38.55	6.58	-25.83	-51.28	132
L-D	YYVQ	105	0	-813.52	162.26	-645.88	-1021.58	-34.10	16.04	-5.75	-51.18	13
D	QQIQ	115	0	-665.20	16.85	-642.86	-701.83	-43.82	4.00	-36.43	-50.25	1
L-D	YFTQ	112	0	-670.91	126.62	-642.46	-857.48	-40.89	11.17	-18.75	-48.18	27
L-D	QDLQ	75	-1	-669.18	17.73	-643.65	-701.80	-43.24	2.85	-38.94	-47.93	9
L-D	EDKQ	64	-1	-664.53	17.21	-641.93	-692.99	-40.53	1.37	-37.32	-42.82	64
L-D	DQYQ	57	-1	-659.57	17.99	-634.90	-686.91	-38.34	1.53	-35.21	-41.49	5
D	EQFQ	74	-1	-663.70	17.17	-641.97	-695.01	-34.68	1.90	-30.66	-38.65	1
L-D	YQYQ	90	0	-672.69	21.38	-642.38	-721.41	-30.65	5.83	-15.62	-36.80	18

5.3. RÉ-OPTIMISATION DE TOUTES LES SÉQUENCES.

TABLE 5.2 – Ré-optimisation des mutants. TyrRS d'*E. coli* avec D-TyrAMP. Les données sur la séquence native sont rappelées. Toutes les séquences au dessus de la ligne en pointillée présentent des structures capables de lier le D-TyrAMP avec un meilleure affinité que les structures de séquence native. Ces mutants vont faire l'objet d'analyses plus détaillées. Les données sur la séquence native sont rappelée à titre indicatif. Un filtre est appliqué pour ce tableau, où figure seulement les complexes qui ont une énergie libre de liaison supérieure à -70 kcal.mol<sup>-1</sup>.

ligand	séquence	occurrence	charge	énergie moyenne	écart -type	énergie max.	énergie min.	affinité moyenne	écart -type	affinité max.	affinité min.	meilleure structure
	DYQQ	69	-2	-648.98	8.77	-636.37	-666.56	-44.32	4.68	-34.57	-50.59	37
D	DQEQQ	73	-2	-641.73	7.45	-632.91	-656.72	-54.88	2.31	-49.33	-59.39	23
L-D	NQEQQ	65	-1	-642.95	7.72	-633.62	-657.12	-50.80	2.19	-47.00	-58.21	18
D	DFDQQ	48	-2	-647.43	8.56	-635.25	-667.08	-48.08	5.60	-39.72	-56.98	8
D	EQDQQ	53	-2	-651.13	10.01	-634.46	-675.39	-46.97	5.38	-35.35	-55.35	5
L-D	QQEQQ	54	-1	-648.41	9.75	-635.09	-668.31	-47.17	4.57	-37.51	-55.24	13
L-D	NDQQQ	28	-1	-650.92	10.95	-635.85	-674.20	-43.53	4.64	-36.01	-55.11	22
D	EQVQQ	70	-1	-645.36	8.57	-634.27	-672.79	-46.02	1.89	-38.68	-55.09	42
D	EQQQQ	60	-1	-647.87	8.36	-635.27	-670.65	-46.79	3.32	-38.51	-55.06	14
D	DFEQQ	62	-2	-647.52	8.15	-635.22	-664.97	-47.80	3.12	-33.70	-53.43	21
L-D	YFEQQ	65	-1	-624.43	260.93	-639.13	-1063.79	-34.37	14.65	-8.50	-53.07	32
L-D	YQEQQ	93	-1	-623.85	165.02	-637.82	-703.20	-36.15	14.69	-1.47	-52.82	25
D	EKEQQ	49	-1	-657.23	18.94	-633.40	-690.47	-46.29	2.57	-40.04	-50.56	13
L-D	YFDQQ	75	-1	-742.58	311.35	-638.76	-1060.29	-31.81	12.87	-22.79	-49.47	4
L	QQDQQ	79	-1	-653.49	10.06	-636.61	-672.87	-41.61	3.43	-35.15	-49.17	8
D	QQIQQ	101	0	-654.16	8.46	-641.01	-672.70	-40.05	2.73	-32.24	-48.98	7
L-D	YFTQQ	74	0	-668.81	246.37	-639.67	-1053.48	-31.50	11.55	-19.37	-48.33	46
L-D	EKDQQ	23	-1	-647.97	9.79	-632.33	-667.79	-44.33	2.81	-39.96	-48.32	23
L-D	DQYQQ	81	-1	-651.36	10.47	-631.03	-671.29	-34.00	5.46	-21.70	-48.21	58
L-D	DFTQQ	34	-1	-643.42	7.86	-633.55	-657.13	-42.80	1.43	-40.68	-47.22	3
D	QQQQQ	88	0	-651.90	8.73	-639.06	-672.42	-41.18	3.08	-31.30	-47.07	17
L	EFVQQ	71	-1	-679.44	94.41	-640.33	-1033.26	-37.40	3.57	-28.83	-46.54	69
L-D	QDLQQ	43	-1	-650.27	8.50	-638.31	-670.42	-39.08	2.35	-33.34	-44.71	38
L-D	YYVQQ	141	0	-632.62	369.88	-640.77	-1167.25	-25.28	16.01	-3.01	-44.51	36
D	EQYQQ	48	-1	-649.98	9.53	-633.00	-667.99	-34.25	5.63	-19.10	-44.46	13
L-D	YQYQQ	143	0	-615.62	209.57	-638.45	-857.16	-22.98	11.52	-0.31	-43.79	58
D	EQFQQ	34	-1	-651.12	10.92	-631.92	-670.55	-35.54	5.39	-24.83	-43.51	6
L	EFTQQ	38	-1	-654.84	9.65	-639.59	-669.31	-36.43	2.64	-30.70	-41.70	32
L-D	EDKQQ	63	-1	-662.07	11.09	-635.58	-681.54	-30.56	5.36	-21.74	-39.86	9

Pour *E. coli*, les mutations concernent les positions 81, 175, 179 et 201. La séquence native est DYQQ. On voit que onze mutants possèdent au moins une structure capable de lier le D-TyrAMP avec plus d'affinité que la TyrRS native. Ces mutants ont pour séquences DQEQQ, DFDQ, DFEQQ, NQEQQ, EQDQQ, QQEQQ, NDQQ, EQVQQ, EQQQQ, YFEQQ et YQEQQ. Pour les mutants DQEQQ, DFDQ et DFEQQ, l'acide aminé natif est prédict en première et dernière position, avec une phénylalanine à la place de la tyrosine native dans deux cas sur trois. Pour ces trois mutants, l'affinité pour le D-TyrAMP est améliorée par rapport à la native, mais l'affinité pour le L-TyrAMP apparaît encore meilleure. Pour les mutants NQEQQ, EQDQQ, QQEQQ, NDQQ, EQVQQ et EQQQQ, la première position est polaire mais différente de l'aspartate natif. Pour ces mutants, il pourrait y avoir une augmentation de sélectivité pour le D-TyrAMP. Enfin chez les mutants YFEQQ et YQEQQ, on retrouve une tyrosine à la position 81 au lieu de la Tyr175 native. L'aspartate natif de la position 81 se retrouve approximativement dans le glutamate prédict à la position 179. Il ne semble pas y avoir de changement décisif pour la spécificité de l'enzyme. Ces onze mutants seront analysés par dynamique moléculaire par la suite.

### 5.3.2 TyrRS de *M. jannaschii*

Pour l'évolution dirigée de la TyrRS de *M. jann*, les mutations concernent les positions 137, 155, 172 et 173. La séquence native est IQEQ. Dix mutants donnent des affinités pour le D-TyrAMP comparables ou supérieures, à la TyrRS native. Ces mutants sont : EEVQ, EQNE, LDDN, LQDN, LLDN, LLVE, NQDE, QELE, LLVD, QEDQ (figure 5.14et 5.15).

Les séquences LDDN, LQDN, LLDN, LLVE et LLVD, quant à elles, améliorent la spécificité pour le D-TyrAMP par rapport à l'enzyme natif. A la position 137, la leucine remplace l'isoleucine natif et libère de la place pour l'ammonium du ligand D. A la position 155, on retrouve un aspartate, un glutamate ou une leucine, à la place de la glutamine native. L'acide aminé natif peut aussi être conservé. A la position 172, on retrouve un acide aminé polaire ou apolaire. A la position 173, un acide aminé polaire est prédict. EEVQ, NQDE, QEDQ semblent améliorer la spécificité de l'enzyme pour le D-TyrAMP, et sont très proches de la séquence native IQEQ. Enfin, EQNE et QELE pourraient lier

### 5.3. RÉ-OPTIMISATION DE TOUTES LES SÉQUENCES.

le D-TyrAMP avec plus d'affinité que la native, sans améliorer la spécificité de l'enzyme pour ce ligand.

Globalement, plus de séquences et de structures différentes ont été échantillonées pour la TyrRS de *M. jann* par rapport à celle d'*E. coli*. L'ion magnésium, modélisé dans le cas de la TyrRS de *M. jann*, joue peut-être un rôle dans ces différences.

FIGURE 5.14 – Ré-optimisation des mutants. TyrRS de *M. jannaschii* avec L-TyrAMP. Les données sur la séquence native sont rappelées. Un filtre est appliqué pour ce tableau, où figure seulement les complexes qui ont une énergie libre de liaison supérieure à -70 kcal.mol<sup>-1</sup>.

ligand	séquence	occurrence	charge	énergie moyenne	écart -type	énergie max.	énergie min.	affinité moyenne	écart -type	affinité max.	affinité min.	meilleure structure
	IQEQQ	6137	-1	-979.65	13.06	-958.52	-1061.69	-41.69	2.50	-32.15	-65.66	3566
L	QELE	2958	-2	-962.07	11.84	-945.67	-1029.59	-50.63	5.40	-11.64	-93.14	1420
D	MQLE	1367	-1	-970.13	14.80	-947.92	-1044.35	-49.73	2.10	-42.55	-70.45	604
D	LQDE	2908	-2	-968.43	11.60	-951.13	-1043.59	-46.73	3.49	-35.87	-68.11	1810
D	EQNE	2756	-2	-960.57	7.54	-947.00	-978.01	-53.75	2.49	-42.45	-67.60	1552
D	MNLD	1796	-1	-974.55	14.11	-954.05	-1044.65	-39.03	2.23	-32.47	-64.80	1047
L	DQNE	1495	-2	-962.30	12.86	-944.13	-1026.57	-56.47	1.97	-49.95	-62.07	1483
L	DENL	1737	-2	-963.40	12.44	-945.47	-1027.46	-55.86	2.02	-48.54	-61.84	850
L	DEVL	1485	-2	-964.04	12.53	-946.06	-1027.29	-53.77	1.83	-47.33	-61.27	1484
L	EEVQ	2587	-2	-962.84	11.60	-946.94	-1029.64	-52.20	2.13	-45.05	-61.05	2556
D	QNDN	1698	-1	-971.96	12.39	-953.70	-1040.06	-36.90	2.21	-27.59	-61.02	1070
D	SNVD	1642	-1	-971.10	12.68	-953.69	-1047.95	-36.74	2.16	-30.16	-60.49	1603
D	LQDN	2754	-1	-974.46	11.67	-957.21	-1048.54	-38.73	2.75	-28.78	-59.26	2284
D	QEQQ	1579	-1	-965.18	12.86	-947.08	-1030.98	-49.72	2.47	-41.40	-59.19	1568
L-D	QEDQ	2336	-2	-964.73	12.21	-947.10	-1043.42	-49.43	2.98	-39.68	-58.07	305
L	VEDL	1799	-2	-963.47	12.10	-946.62	-1027.31	-50.06	1.99	-41.52	-57.15	1788
D	LLVE	2416	-1	-968.74	12.32	-951.50	-1038.02	-40.31	4.02	-28.50	-56.61	2404
L	TEDL	1762	-2	-963.44	12.49	-946.35	-1032.39	-49.79	1.87	-43.21	-55.87	1762
D	LQVE	2647	-1	-968.48	11.79	-951.27	-1035.53	-45.38	3.72	-34.04	-55.67	2639
D	LLDN	2038	-1	-975.22	12.29	-956.95	-1045.84	-34.32	3.15	-24.39	-55.47	1990
L-D	QQVE	1398	-1	-962.93	13.58	-943.75	-1034.49	-49.84	2.26	-35.71	-55.24	1378
D	QQDE	2506	-2	-962.98	11.72	-945.18	-1030.63	-47.62	2.95	-37.44	-55.13	2504
D	SQVE	1378	-1	-966.52	12.75	-948.19	-1031.40	-47.39	1.99	-39.79	-54.58	1373
L	NQDE	1378	-2	-962.48	12.87	-944.61	-1027.76	-49.43	1.82	-43.48	-54.31	1356
D	EQVN	2062	-1	-969.28	11.65	-950.17	-1040.90	-44.57	2.71	-35.38	-53.72	2060
D	SQDE	1592	-2	-965.74	12.87	-947.97	-1029.98	-47.74	1.86	-41.26	-53.53	922
D	QEDN	2772	-2	-964.80	8.29	-949.94	-983.83	-46.01	2.68	-32.10	-53.23	330
D	QQDD	2765	-2	-965.90	9.41	-950.24	-993.08	-43.44	4.07	-30.41	-51.86	1717
D	QQVD	1584	-1	-965.50	13.02	-947.69	-1031.67	-43.04	2.42	-35.03	-50.57	1562
D	ELVN	1146	-1	-969.85	13.26	-950.40	-1034.85	-43.02	1.75	-34.79	-49.64	416
D	QLVD	1743	-1	-965.50	12.15	-948.37	-1034.76	-41.00	1.84	-31.88	-48.59	1738
D	QDDT	3066	-2	-971.68	11.01	-952.88	-1036.03	-41.14	2.30	-28.94	-48.46	1395
D	SQVD	1573	-1	-968.49	12.34	-951.37	-1032.28	-40.80	2.00	-33.00	-48.32	1549

FIGURE 5.15 – Ré-optimisation des mutants. TyrRS de *M. jannaschii* avec D-TyrAMP. Les données sur la séquence native sont rappelées. Toutes les séquences au dessus de la ligne en pointillée présentent des structures capables de lier le D-TyrAMP avec un meilleure affinité que les structures de séquence native. Ces mutants vont faire l'objet d'analyses plus détaillées. Les données sur la séquence native sont rappelée à titre indicatif. Un filtre est appliqué pour ce tableau, où figure seulement les complexes qui ont une énergie libre de liaison supérieure à -70 kcal.mol<sup>-1</sup>.

ligand	séquence	occurrence	charge	énergie moyenne	écart -type	énergie max.	énergie min.	affinité moyenne	écart -type	affinité max.	affinité min.	meilleure structure
	IQEQQ	13031	-1	-996.68	12.82	-973.64	-1079.23	-31.12	4.18	-3.43	-52.72	12227
L	EEVQ	3460	-2	-988.55	12.63	-968.15	-1082.04	-31.49	8.07	24.29	-84.07	3637
D	EQNE	5170	-2	-987.41	12.01	-964.11	-1025.34	-31.55	7.77	21.18	-57.94	19
D	LDDN	3313	-2	-980.52	12.44	-961.97	-1062.28	-29.81	4.47	-9.20	-57.31	2927
D	LQDN	3965	-1	-979.64	11.99	-961.75	-1062.69	-29.20	5.80	1.95	-56.51	2027
D	LLDN	3253	-1	-979.97	12.88	-961.34	-1054.53	-26.09	6.03	33.82	-56.44	1361
D	LLVE	4048	-1	-984.47	12.62	-963.06	-1062.07	-32.55	5.81	23.61	-56.21	3305
L	NQDE	4201	-2	-986.43	13.49	-964.63	-1065.38	-27.38	6.77	33.39	-53.26	4086
L	QELE	5227	-2	-986.48	10.88	-967.67	-1019.25	-35.79	7.93	18.86	-53.17	2602
D	LLVD	3668	-1	-980.87	12.01	-962.97	-1047.16	-28.08	5.49	23.13	-52.67	2140
L-D	QEDQ	4885	-2	-984.35	11.93	-965.50	-1052.73	-31.13	7.18	33.69	-50.80	2990
D	QQVD	5333	-1	-978.42	11.08	-962.36	-1052.03	-28.55	6.12	20.95	-49.01	5179
L	DEVL	4109	-2	-987.71	12.33	-969.98	-1067.78	-33.05	7.09	21.43	-48.96	4307
D	LQVE	4020	-1	-985.27	12.92	-962.44	-1056.76	-31.56	6.41	19.70	-48.76	2066
L	DQNE	5321	-2	-985.24	9.52	-967.79	-1006.60	-32.52	7.68	31.48	-48.70	2741
L-D	QQVE	4676	-1	-983.82	13.20	-962.59	-1056.46	-30.23	8.06	22.40	-48.56	4773
D	SQDE	3848	-2	-983.43	12.31	-962.86	-1060.36	-30.39	6.57	22.16	-48.31	2235
L	DENL	4872	-2	-986.66	12.06	-969.59	-1071.32	-33.24	7.34	33.17	-48.09	1992
D	QEDN	2735	-2	-974.67	4.95	-964.27	-984.91	-30.73	6.24	22.67	-48.08	862
L	TEDL	3719	-2	-983.21	12.39	-964.88	-1072.03	-31.91	5.79	17.52	-48.03	474
D	SNVD	2561	-1	-980.35	12.61	-960.51	-1044.61	-21.26	8.09	18.82	-47.64	2418
D	LQDE	3782	-2	-985.93	13.70	-962.82	-1065.56	-31.37	6.70	26.29	-47.62	1566
D	EQVN	3505	-1	-983.11	12.17	-964.74	-1056.85	-28.23	6.23	3.67	-47.49	1613
L	VEDL	3568	-2	-984.47	12.47	-965.79	-1062.25	-31.80	6.02	18.72	-47.31	2765
D	MQLE	3781	-1	-985.21	12.57	-963.28	-1056.18	-31.81	6.49	19.81	-47.27	3956
D	QEVD	3579	-1	-985.55	12.27	-965.85	-1059.56	-29.65	7.20	43.18	-46.15	3685

## 5.4 Caractérisation des structures obtenues

Nous allons nous intéresser tout d'abord à la diversité des séquences et des structures générées par notre méthode de CPD, comme nous l'avons fait plus haut pour la séquence native. Pour chaque mutant, ainsi que pour l'enzyme natif, nous aurons un regard sur toutes les structures prédites (avec ligand natif et non natif). Sur les 20000 cycles heuristiques générés pour chaque séquence mutante (100000 pour la native), beaucoup de structures optimisées sont prédites plusieurs fois. Après éliminations des structures, répétées, on obtient finalement pour chaque séquence une centaine de structures différentes. L'analyse est faite par l'étude des graphiques suivants :

- la densité de répartition des affinités,
- les énergies de dépliement en fonction des affinités,
- les énergies de dépliement calculées par Proteus en fonction des énergies libres calculées par Xplor (à partir des structures reconstruites),
- la densité de répartition des énergies de repliement,
- la densité de répartition des énergies libres calculées par Xplor.

A chaque fois, nous analyserons ces paramètres à la fois pour la TyrRS d'*E. coli* et pour celle de *M. jannaschii*. Nous présenterons ici les douze mutants les plus intéressants, d'après le critère d'affinité pour le D-TyrAMP.

### 5.4.0.1 Répartition des affinités protéine:ligand

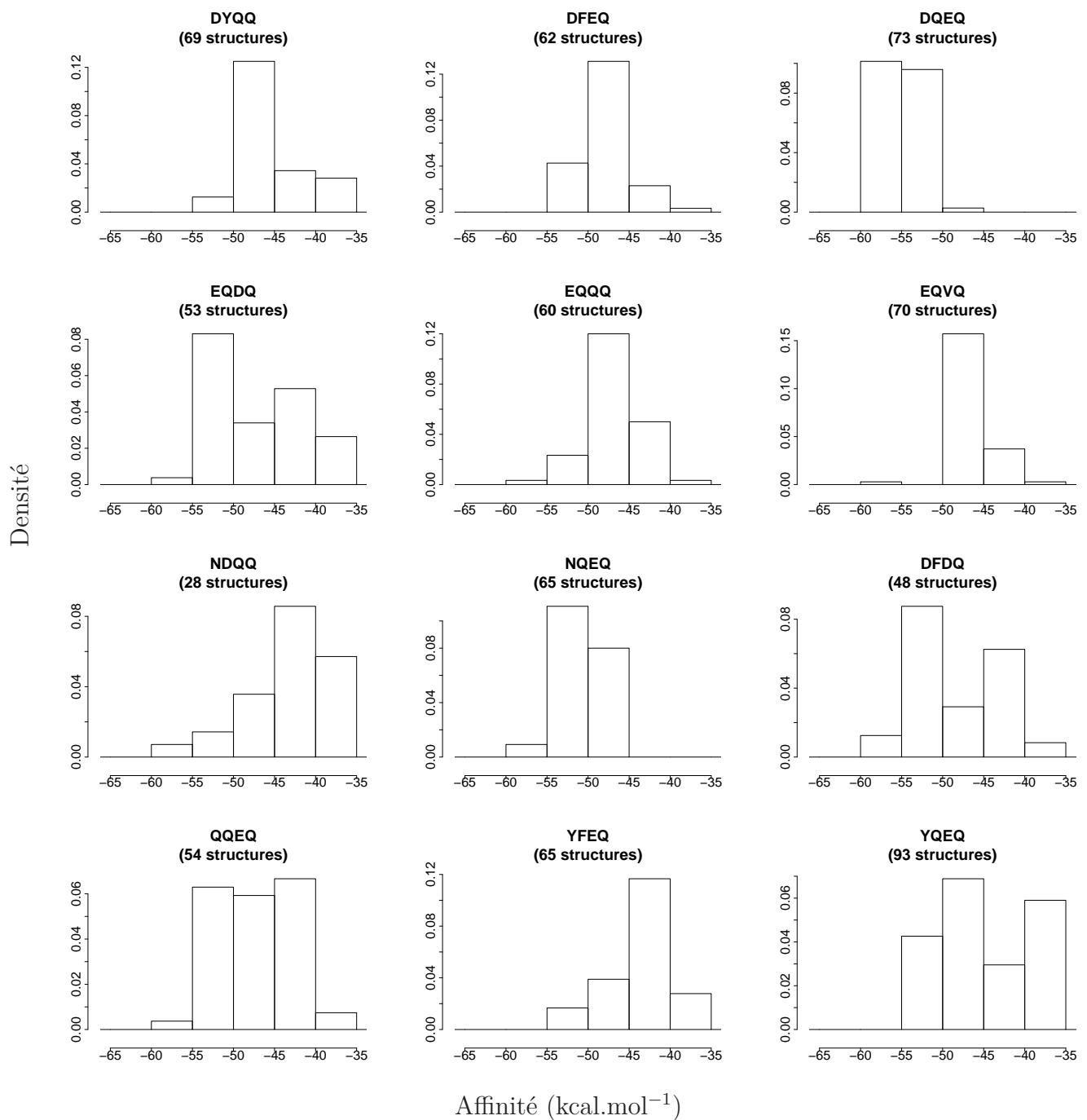
**TyrRS d'*E. coli*** Intéressons nous tout d'abord à la répartition des affinités pour les douze séquences prédites avec le ligand D-TyrAMP, représentées dans la figure 5.16. Il faut noter que les valeurs absolues des affinités ne sont pas significatives, car certaines contributions sont omises (entropie de rotation, translation). Ce sont plutôt les différences entre séquences qui nous intéressent. Les affinités sont globalement réparties entre -65 et -35 kcal.mol<sup>-1</sup>. Pour la native, la majorité des structures ont une affinité entre -50 et -45 kcal.mol<sup>-1</sup>. Il en est de même pour les séquences DFEQ, DFEQ, EQQQ, EQVQ et YQEQQ. Les séquences EQDQ, DFDQ et NQEQQ ont une majorité de structures avec une affinité pour le D-TyrAMP entre -50 et -55 kcal.mol<sup>-1</sup>, ce qui voudrait dire que ces

séquences améliorent l'affinité pour le nouveau ligand. La séquence DQEQQ a des structures majoritairement réparties entre -50 et -60 kcal.mol<sup>-1</sup>, ce qui voudrait dire également que ce mutant améliore l'affinité pour le D-TyrAMP. Pour la séquence QQEQ, la plage d'affinité balayée par l'ensemble des structures est très large, de -55 à -40 kcal.mol<sup>-1</sup>; on ne peut donc pas conclure sur les effets de ce mutant, car autant de structures présentent une affinité supérieure à la native qu'inférieure. Enfin, les séquences YFEQQ et NDQQQ ont une majorité de structures avec une affinité pour le D-TyrAMP située entre -45 et -40 kcal.mol<sup>-1</sup>; elles sont donc moins efficaces que la native pour lier le D-TyrAMP, en moyenne, bien que certaines structures y parviennent (figure 5.16).

Comparons maintenant les affinités obtenues pour ces séquences avec le L-TyrAMP (figure 5.17). Les affinités sont globalement réparties entre -65 et -45 kcal.mol<sup>-1</sup>, comme avec le D-TyrAMP. La native a des structures majoritairement comprises entre -55 et -50 kcal.mol<sup>-1</sup>. Notre modélisation de l'enzyme natif donne ainsi, comme attendue, une affinité moyenne supérieure pour le L-TyrAMP par rapport au D-TyrAMP. Les mutants EQQQQ, EQVQQ, NQEQQ et YFEQQ ont une majorité de structures dans la même plage d'affinité que la native. Les mutants DFEQQ, DQEQQ, EQDQQ, DFDQQ et YQEQQ ont une majorité de structures avec une affinité entre -60 et -55 kcal.mol<sup>-1</sup>; ces mutants lient donc mieux le L-TyrAMP que la native. Enfin, les mutants NDQQQ et QQEQQ ont une majorité de structures entre -50 et -45 kcal.mol<sup>-1</sup>; ces mutants ont donc une affinité pour le L-TyrAMP inférieure à la native.

La fonction d'énergie CASA est très simple, et ces séquences devront être réexaminées à la lumière des simulations de dynamique et des calculs Poisson-Boltzmann, décrits dans le chapitre suivant. Néanmoins, ces résultats nous donnent quelques indications sur la spécificité de ces séquences. Voyons tout d'abord les séquences EQDQQ, DFDQQ et NQEQQ, qui sont prédictées comme améliorant l'affinité pour le nouveau ligand. EQDQQ et DFDQQ améliorent également l'affinité pour le L-TyrAMP; ces mutants ne sont pas spécifiques pour lier le D-TyrAMP. Le mutant NQEQQ, en revanche, a une majorité de structures entre -55 et -45 kcal.mol<sup>-1</sup>, à la fois pour le L-TyrAMP et pour le D-TyrAMP. Le ratio de fixation D/L serait donc amélioré par rapport à la native, mais toujours pas inversé.

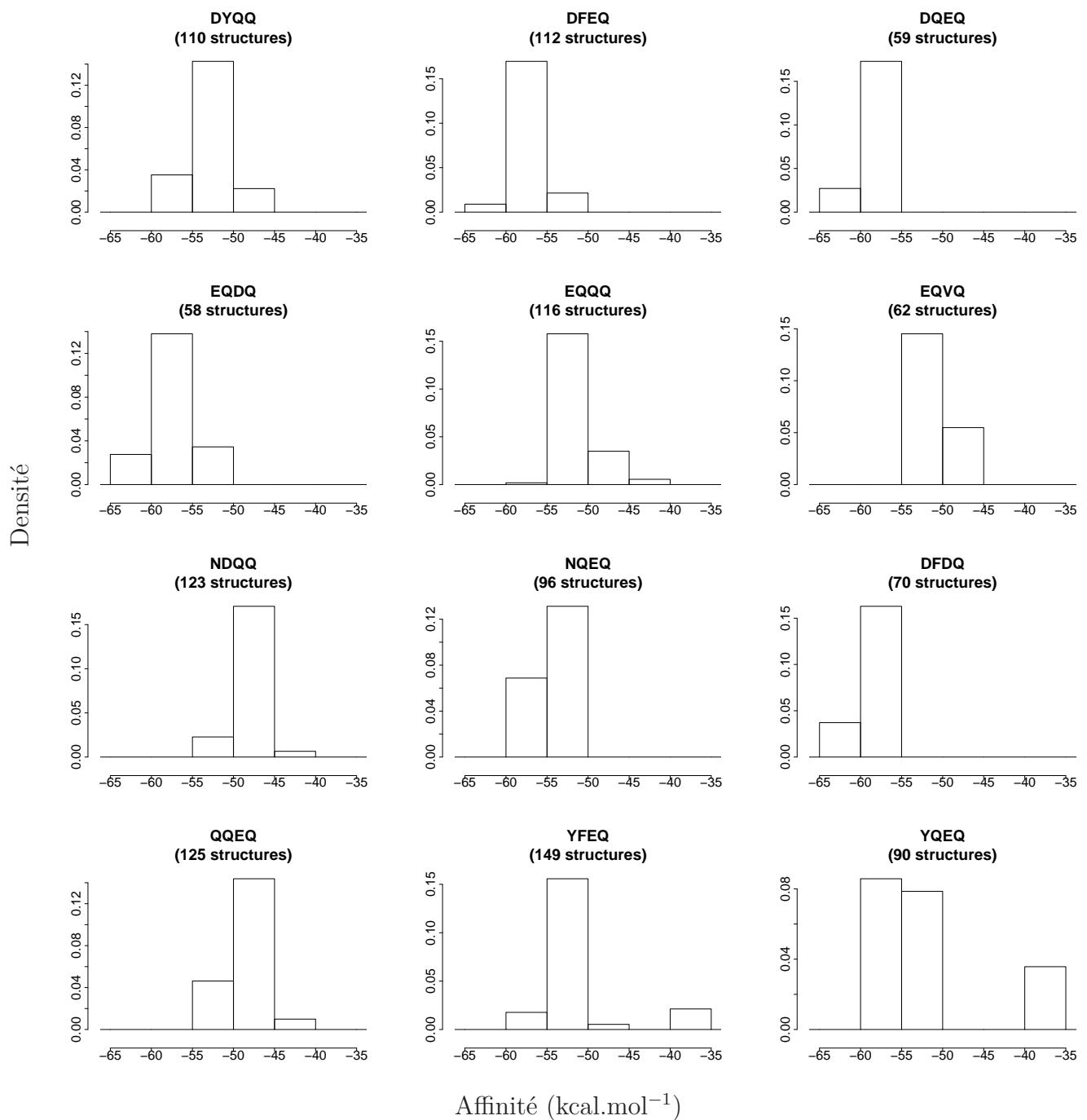
FIGURE 5.16 – Affinités estimées pour les complexes Eco-TyrRS:D-TyrAMP



#### 5.4. CARACTÉRISATION DES STRUCTURES OBTENUES

---

FIGURE 5.17 – Affinités estimées pour les complexes *Eco-TyrRS:L-TyrAMP*



Pour le mutant DQEQQ, de même, on a également amélioré l'affinité à la fois pour le L-TyrAMP et pour le D-TyrAMP, mais le ratio de fixation D/L semble considérablement amélioré même s'il n'est toujours pas inversé. Le mutant QQEQ ne change pas l'affinité pour le L-TyrAMP ; certaines de ces structures lient mieux le D-TyrAMP que la native et d'autres, moins bien ; il est donc difficile de dégager une tendance pour ce mutant. Pour le mutant YFEQQ, l'affinité pour le L-TyrAMP est comparable à la native, et l'affinité pour la D-TyrAMP est diminuée par rapport à la native. C'est donc un moins bon candidat pour notre étude. Enfin, le mutant NDQQQ défavorise à la fois la liaison au L et D-TyrAMP, de manière similaire.

**TyrRS de *M. jannaschii*** Pour la TyrRS de *M. Jannaschii* native, l'affinité pour le D-TyrAMP varie de -45 à -20 kcal.mol<sup>-1</sup>. Pour les mutants EEVQ, EQNE, et QELE, cette affinité est supérieure à la native pour un grand nombre de structures prédites. Pour les autres mutants étudiées, la plupart des structures ont une affinité comparable à la native pour le D-TyrAMP (figure 5.18).

Comparons maintenant avec l'affinité pour le L-TyrAMP. La TyrRS native a une affinité entre -50 et -35 kcal.mol<sup>-1</sup>. L'affinité pour le L-TyrAMP est améliorée par les mutants EEVQ, EQNE, NQDQ, QELE et QEDQ. L'affinité est sensiblement la même que la native pour les mutants LLVE et QQVD. L'affinité pour le L-TyrAMP est diminuée avec les mutants LDDN, LQDN, LLDN et LLVD (figure 5.19).

En conséquence, les mutants qui amélioreraient l'affinité pour le D-TyrAMP apparaissent peu spécifiques, et ceux qui maintiendraient une affinité comparable à la native, tout en diminuant l'affinité pour le L-TyrAMP, paraissent plus intéressants ; c'est le cas des mutants LDDN, LQDN, LLDN et LLVD.

#### 5.4.0.2 Energies de déploiement en fonction des affinités

**TyrRS d'*E. coli*** Voyons maintenant la relation entre les énergies de déploiement et les affinités pour toutes ces séquences. Pour les complexes TyrRS:D-TyrAMP, les énergies de déploiement se trouvent entre -700 et -630 kcal.mol<sup>-1</sup>. Comme pour les affinités, il faut noter que les valeurs absolues ne sont pas directement interprétables, car plusieurs contributions

#### 5.4. CARACTÉRISATION DES STRUCTURES OBTENUES

FIGURE 5.18 – Affinités estimées pour les complexes *Mja*-TyrRS:D-TyrAMP

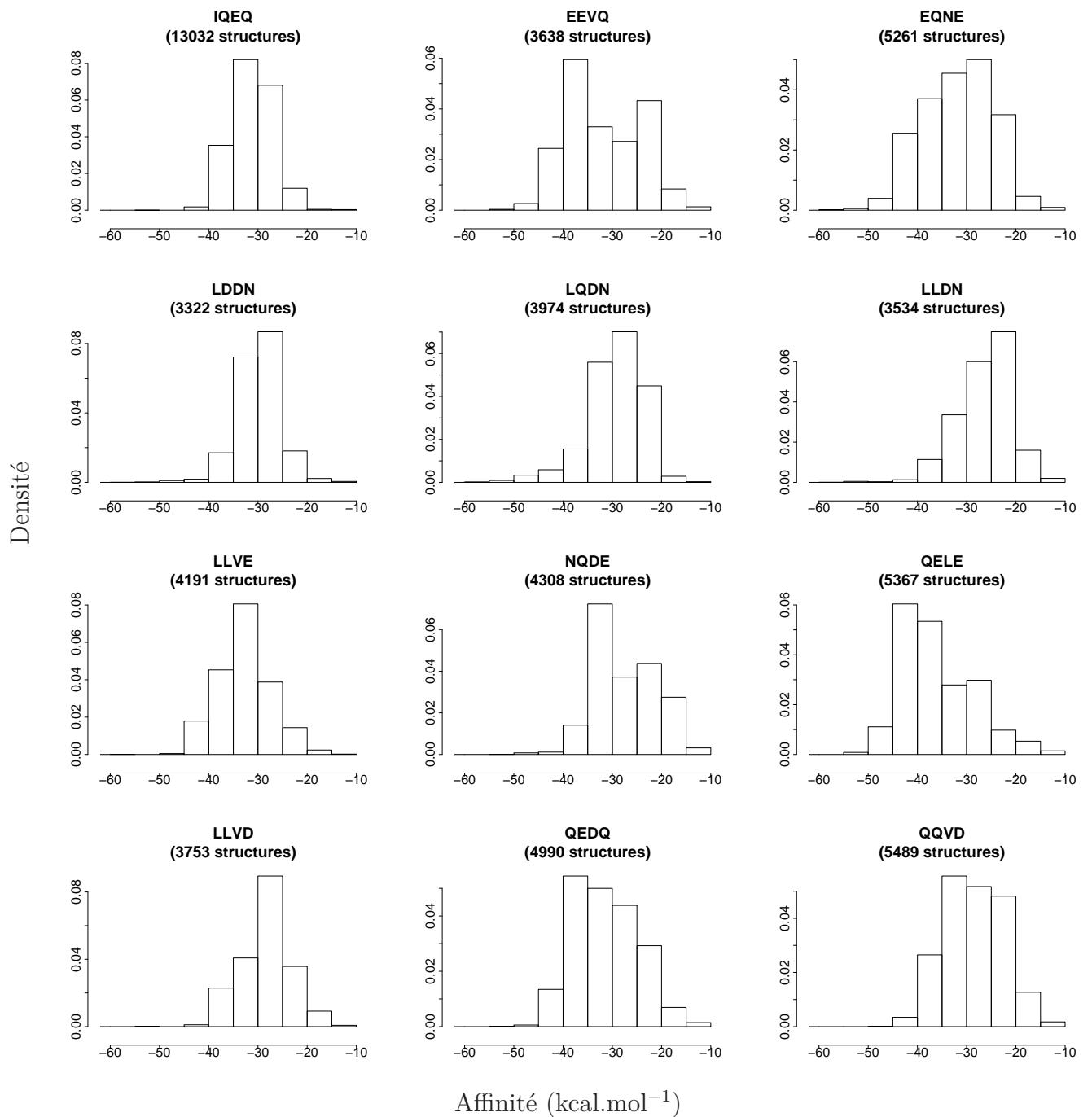
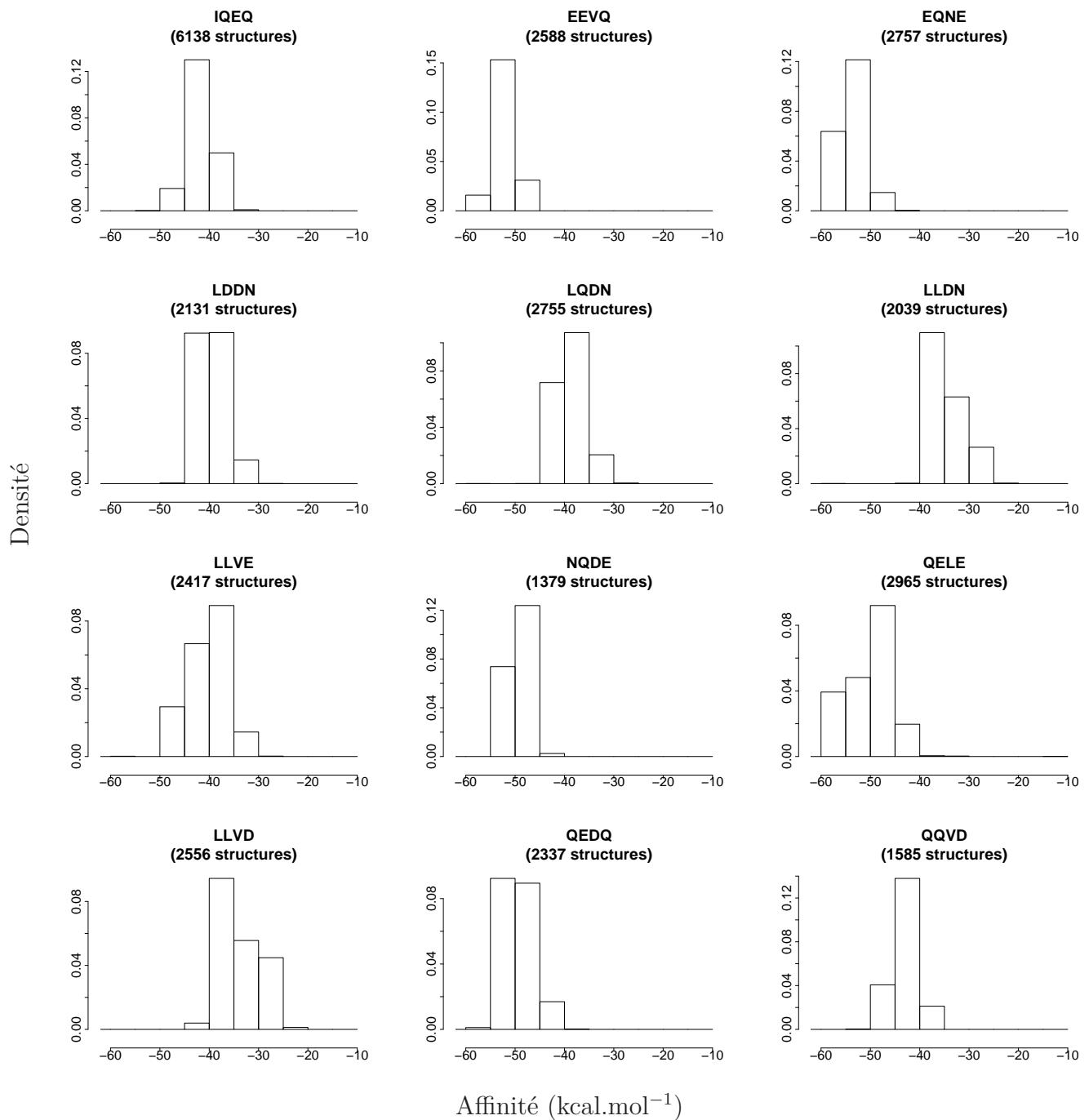


FIGURE 5.19 – Affinités estimées pour les complexes *Mja*-TyrRS:L-TyrAMP



importantes sont omises. Les structures stables sont celles qui ont l'énergie de dépliement la plus grande, elles sont donc situées dans la partie supérieure des graphiques de la figure 5.20. On remarque qu'il y a deux minima locaux d'énergie distincts, donnant deux nuages de points. Pour chaque minimum d'énergie, on obtient des structures de plus ou moins bonne affinité, mais pour toutes les séquences on remarque que les structures de meilleures affinités sont également parmi les plus stables. Les graphiques qui présentent un maximum de points (structures) dans la partie gauche supérieure, correspondent aux mutants qui lient le ligand avec une bonne affinité et qui ont une structure stable. Les mutants DFEQ, DQEQQ, EQDQ, EQQQ, DFDQ, QQEQ et YQEQQ correspondent à ce critère.

Le comportement de ces séquences avec le L-TyrAMP doit également être analysé. Les énergies de dépliement se trouvent entre -850 et -630 kcal.mol<sup>-1</sup>. Ces énergies de dépliement sont, en moyenne, plus faibles qu'avec le D-TyrAMP. On remarque que l'optimisation génère des structures groupées autour de minima d'énergie bien distincts, au nombre de deux dans la plupart des cas, mais pouvant aller jusqu'à quatre ou plus pour les mutants NDQQ, NQEQQ et QQEQ. Ici, la structure de meilleure affinité correspond toujours à une énergie de dépliement élevée, sauf pour les mutants EQDQ, NDQQ, où quelques structures moins stables dérogent à cette règle (figure 5.21).

De cette analyse, il ressort que les mutants DFEQ et EQDQ seraient ceux qui amélieraient le plus la liaison du D-TyrAMP.

**TyrRS de *M. jannaschii*** Pour les complexes TyrRS:D-TyrAMP, les énergies de dépliement se trouvent entre -1040 et -940 kcal.mol<sup>-1</sup>. On constate ici qu'il y a au moins deux minima d'énergie par mutant. La meilleure affinité est toujours donnée par une structure d'énergie de dépliement élevée (figure 5.22). Même constat avec le ligand L-TyrAMP : il y a deux minima locaux d'énergie de dépliement, mais l'affinité maximale n'est pas toujours donnée par la structure qui à la plus grande énergie de dépliement. Pour les mutants LDDN, LLVE, NQDE, QQVD, la structure de TyrRS donnant la meilleure affinité pour le L-TyrAMP n'a pas une énergie de dépliement très élevée ; elle est peu stable (figure 5.23).

FIGURE 5.20 – Energie de repliement en fonction de l'affinité pour les complexes *Eco*-TyrRS:D-TyrAMP

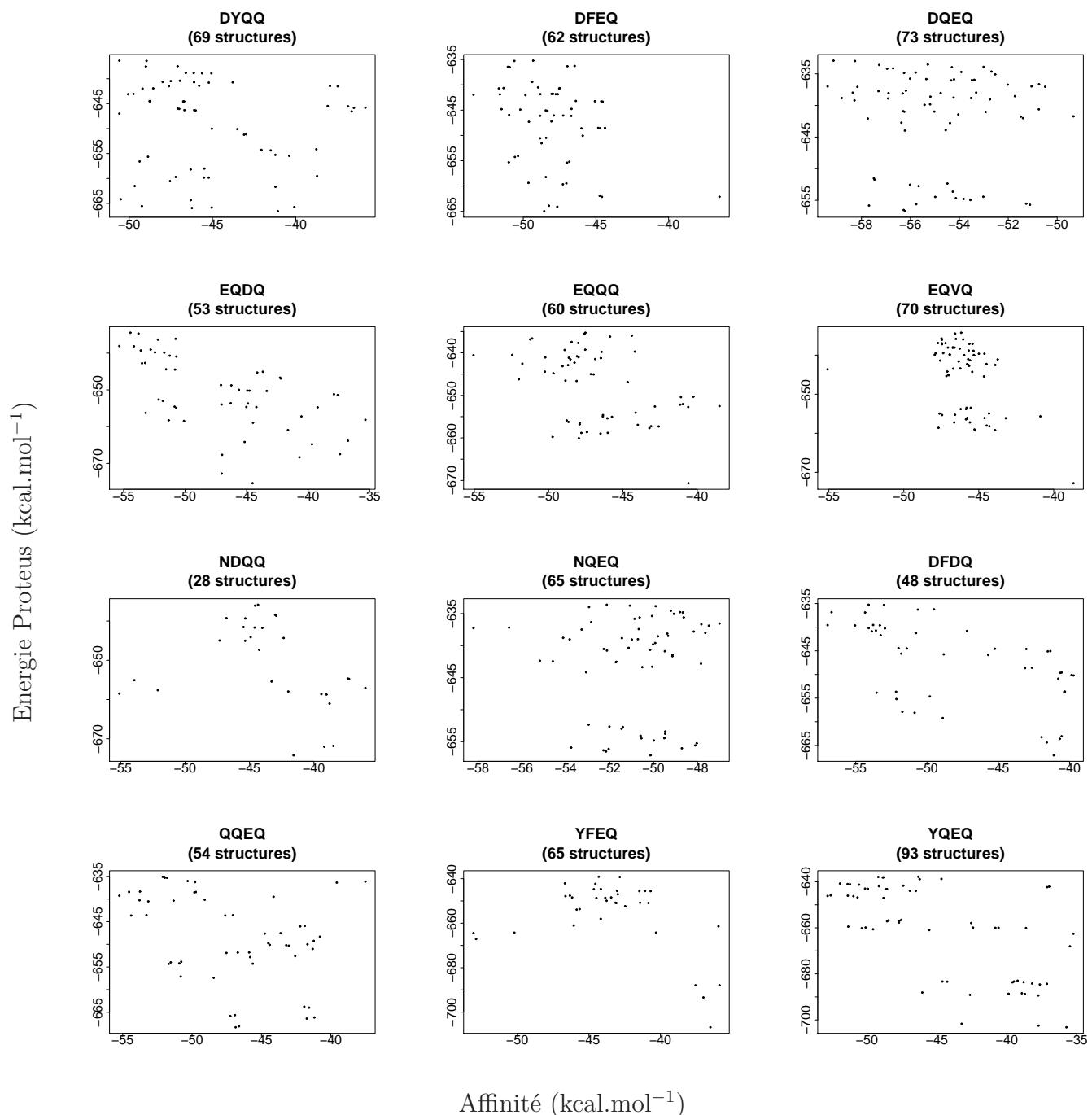


FIGURE 5.21 – Energie de repliement en fonction de l'affinité pour les complexes *Eco-TyrRS:L-TyrAMP*

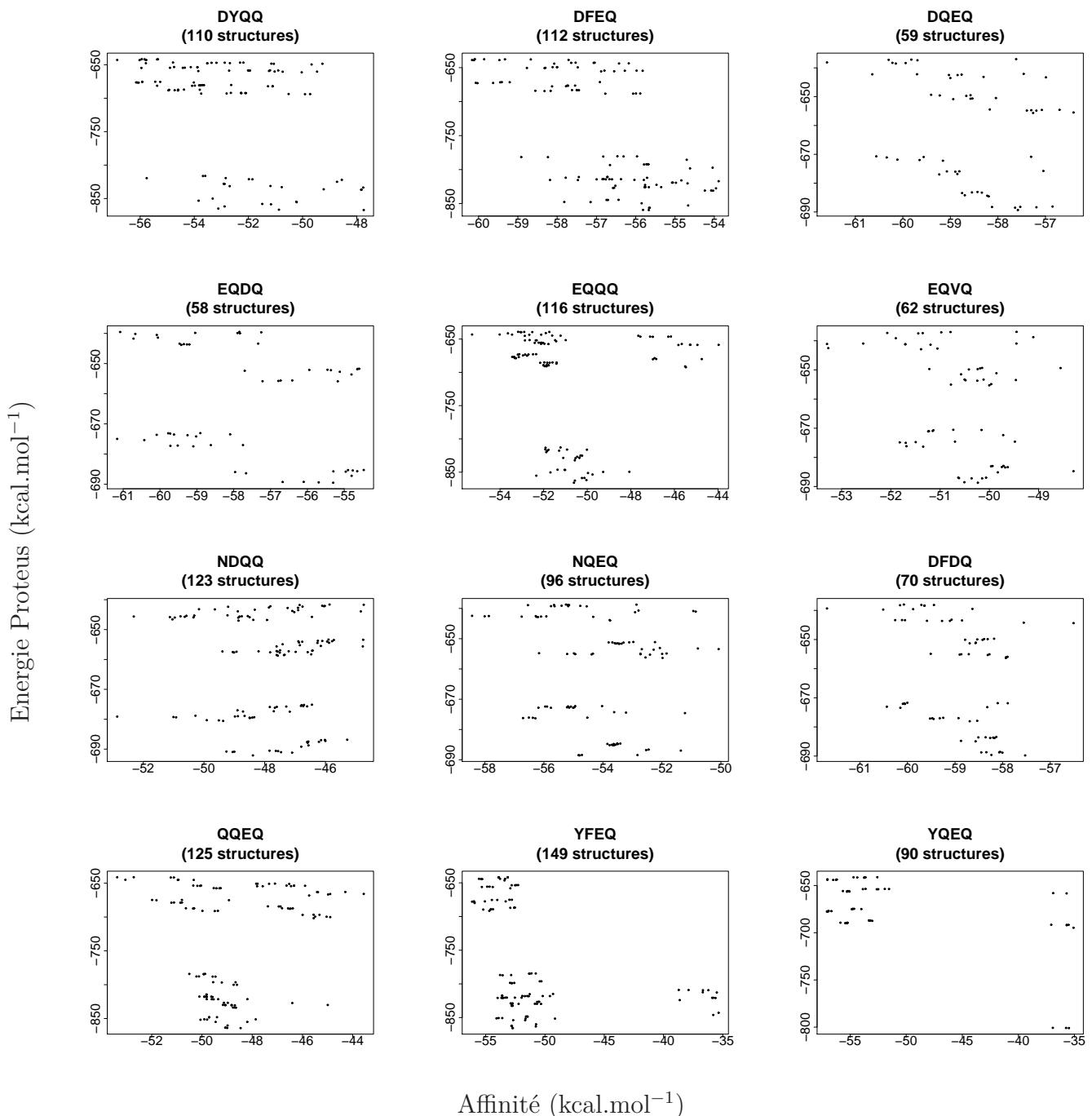


FIGURE 5.22 – Energie de repliement en fonction de l'affinité pour les complexes *Mja*-TyrRS:D-TyrAMP

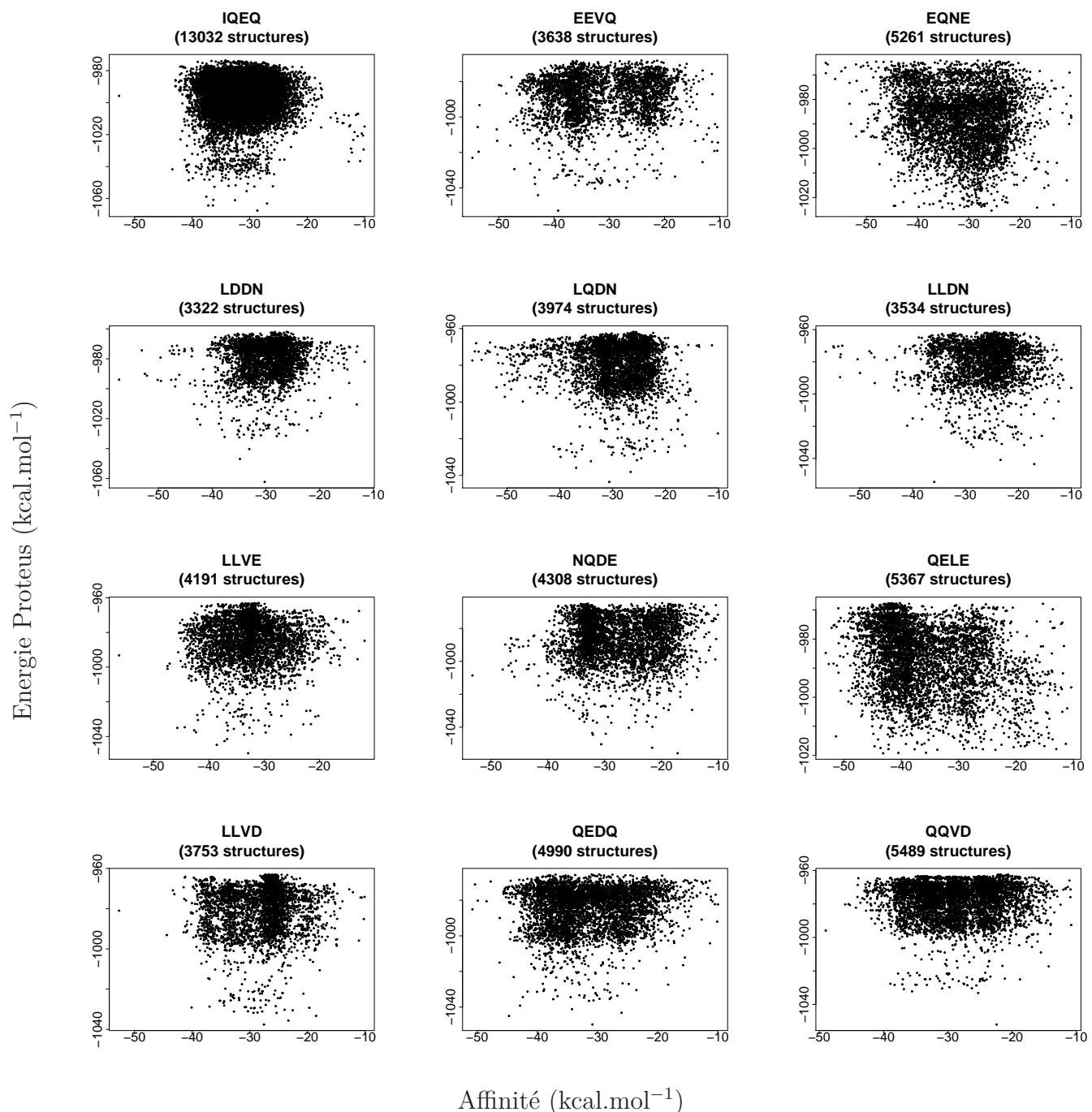
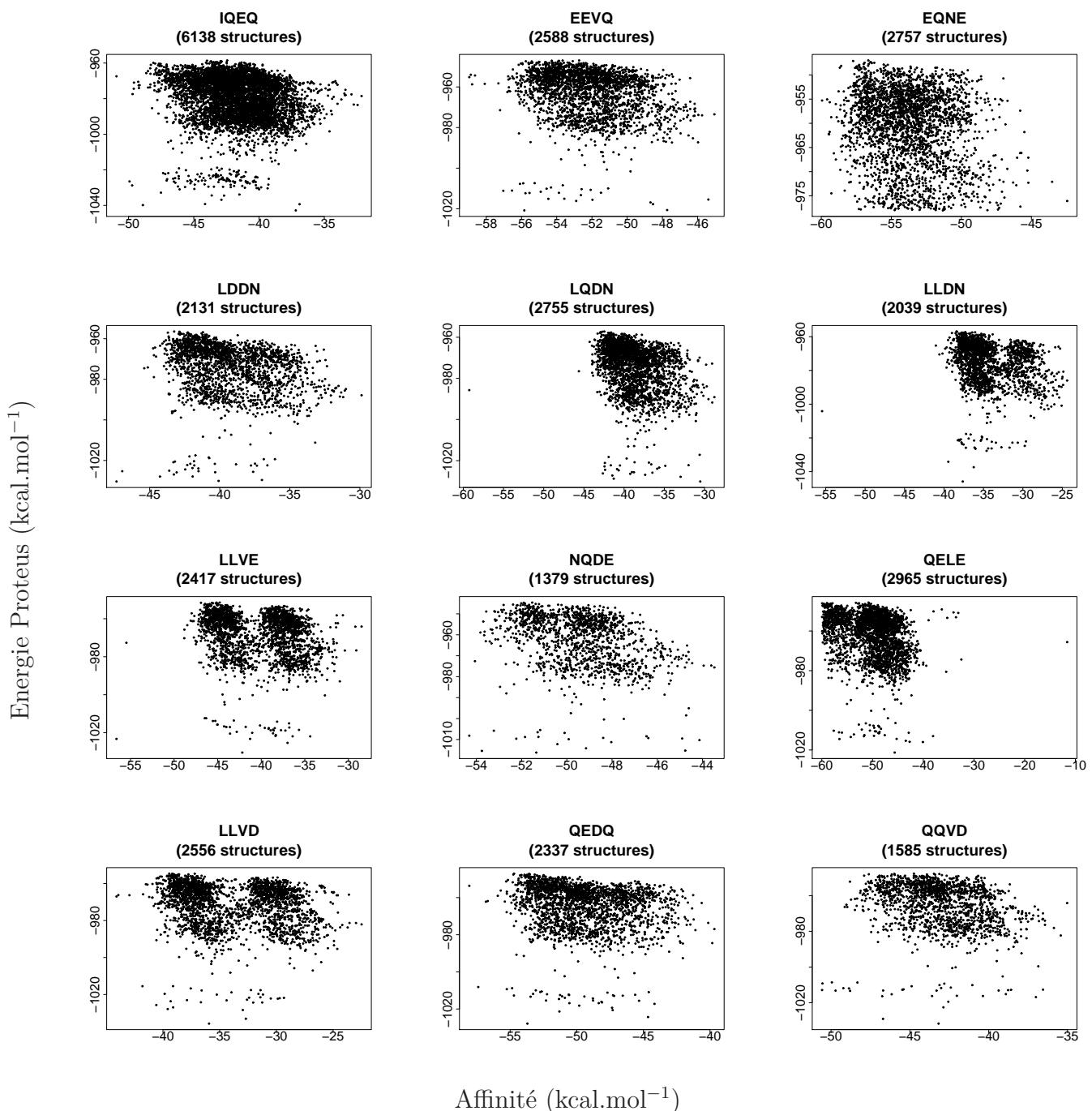


FIGURE 5.23 – Energie de repliement en fonction de l'affinité pour les complexes *Mja*-TyrRS:L-TyrAMP



#### 5.4.0.3 Energies de dépliement comparées aux énergies des structures reconstituées

**TyrRS d'*E. coli*** L'énergie de dépliement est obtenue par Proteus à partir de la matrice d'énergie, en tenant compte également des énergies de référence. En revanche, la reconstruction 3D des structures prédictes permet un calcul direct de l'énergie totale estimée après une légère minimisation d'énergie (cf. Chapitre Dessein computationnel de la TyrRS : méthodes). Il est donc intéressant de vérifier qu'il y a une certaine cohérence entre ces valeurs. Il s'agit donc de comparer les énergies obtenue par Proteus et par Xplor. Avec le D-TyrAMP, on remarque l'existence de deux groupes de structures. Ceci est très visible sur la figure 5.24 pour les mutants YFEQ et YQEQQ. On constate qu'il y a des minima d'énergie donnés par Xplor, comme il y a des minima d'énergie donnés par Proteus, et que ceux-ci semblent corrélés "par groupes". Globalement, plus l'énergie du système est stable, plus l'énergie de dépliement est élevée au sein d'un même bassin d'énergie. Mais d'un groupe à l'autre, on peut avoir une même énergie de dépliement pour des énergies Xplor très différentes (figure 5.24). Pour les complexes L-TyrAMP, la répartition des énergies Xplor autour de minima locaux est plus marquée (figure 5.25).

**TyrRS de *M. jannaschii*** On constate également que les structures les plus stables ont des énergies de dépliement élevées. Il y a donc un bon accord entre les calculs d'énergie Xplor et Proteus. La même observation est faite avec le ligand L-TyrAMP (cf. figures 5.31 et 5.30 en Appendice).

#### 5.4.0.4 Répartition des énergies de repliement

**TyrRS d'*E. coli*** Les énergies de dépliement des structures de TyrRS liées au D-TyrAMP, sont globalement comprises entre -750 et -600 kcal.mol<sup>-1</sup> (figure 5.26). Pour toutes les séquences, la majorité des structures sont dans un intervalle de 20 kcal.mol<sup>-1</sup>. Avec le L-TyrAMP, les structures ont une énergie de dépliement comprise entre -900 et -600 kcal.mol<sup>-1</sup> (figure 5.27). Pour la native et les mutants DFEQ, EQQQ, QQEQ, YFEQ et YQEQQ, on observe deux minima locaux d'énergies, un vers -850 kcal.mol<sup>-1</sup> et l'autre

FIGURE 5.24 – Energies de dépliement comparées aux énergies des structures reconstruites pour les complexes *Eco-TyrRS:D-TyrAMP*

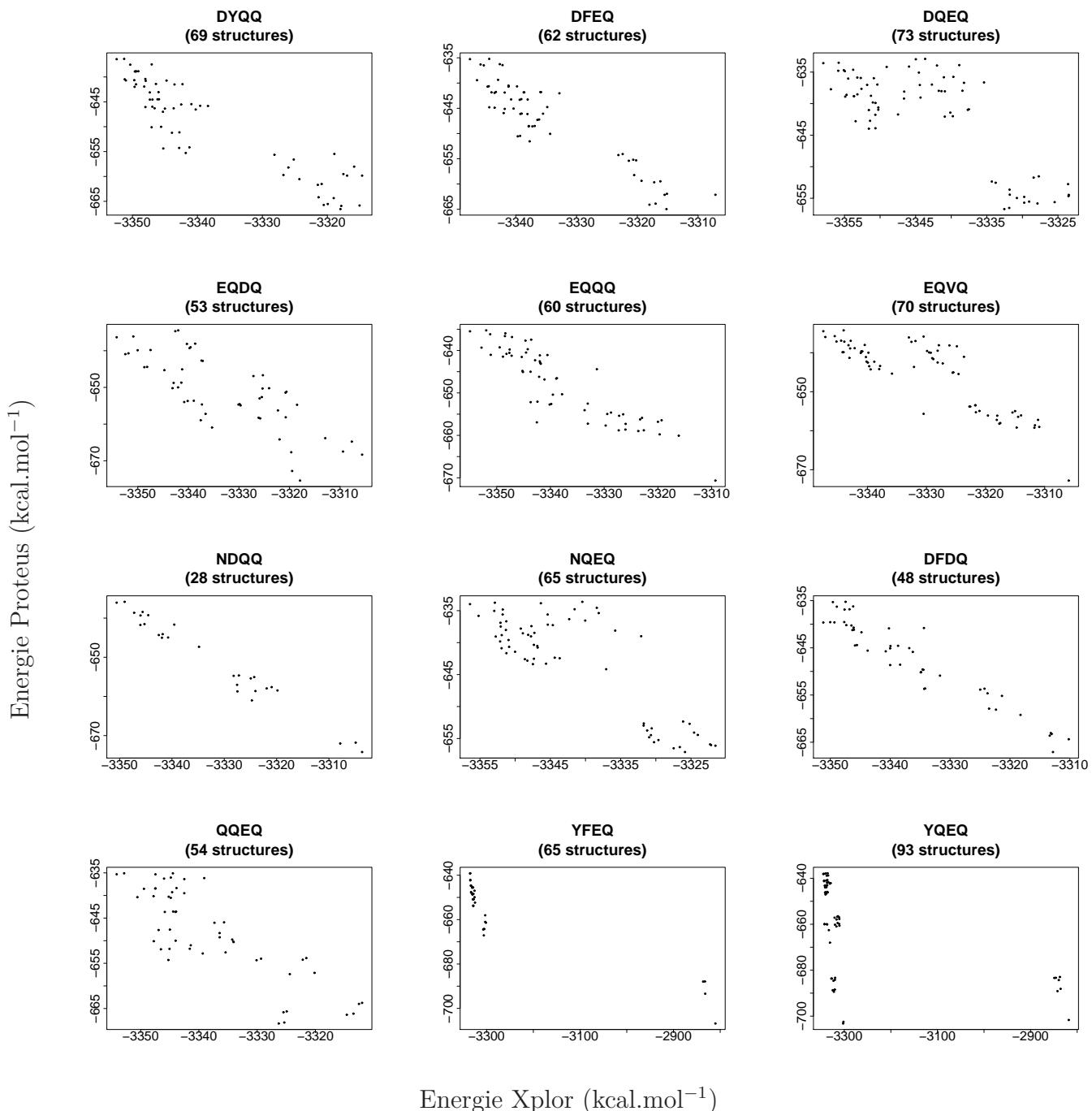
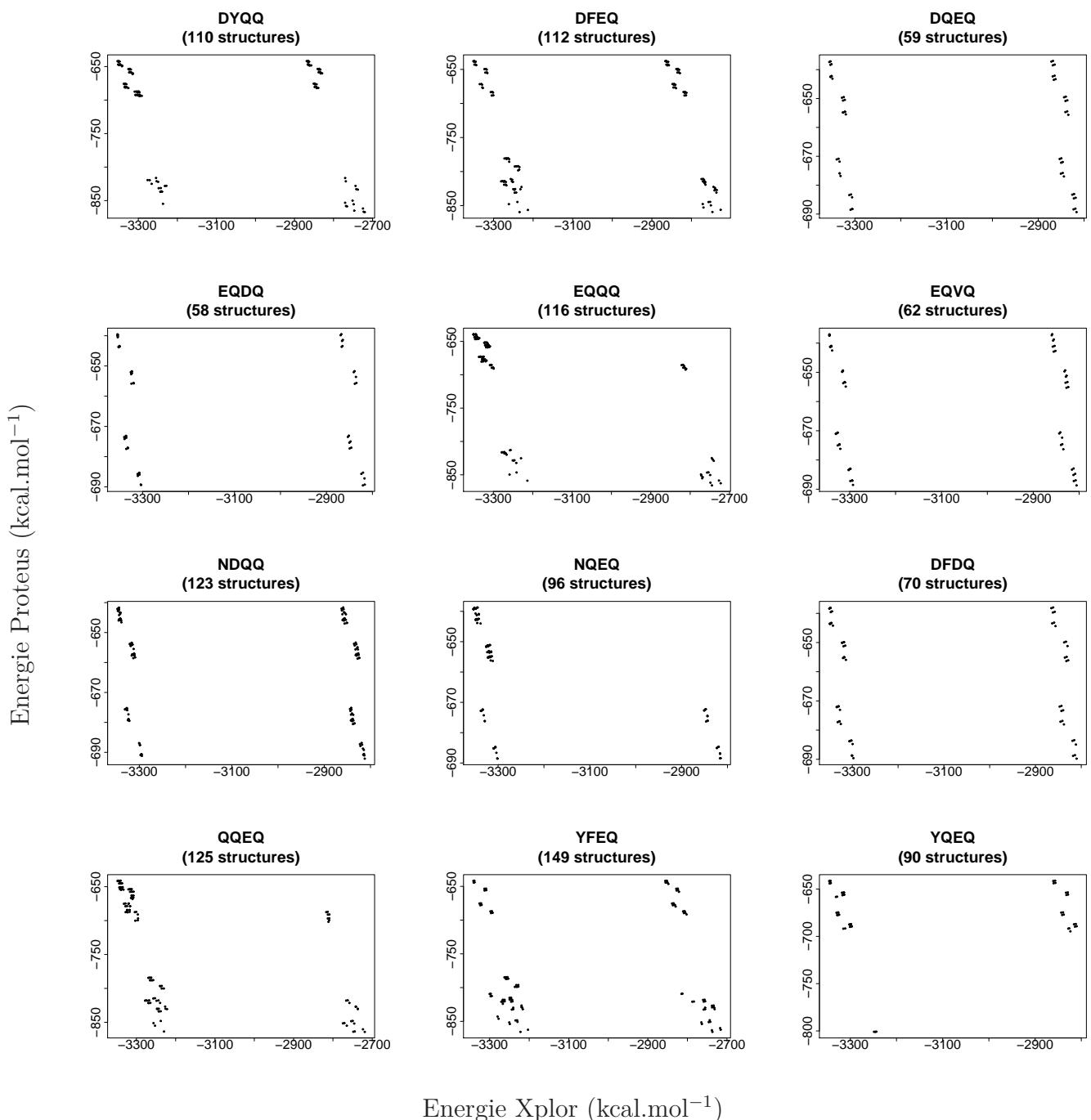


FIGURE 5.25 – Energies de dépliement comparées aux énergies des structures reconstruites pour les complexes *Eco-TyrRS:L-TyrAMP*



vers  $-650 \text{ kcal.mol}^{-1}$ , alors que pour les mutants DQEQQ, EQDQQ, EQVQQ, NDQQQ, NQEQQ et DFDQQ, seul celui vers  $-650 \text{ kcal.mol}^{-1}$  existe.

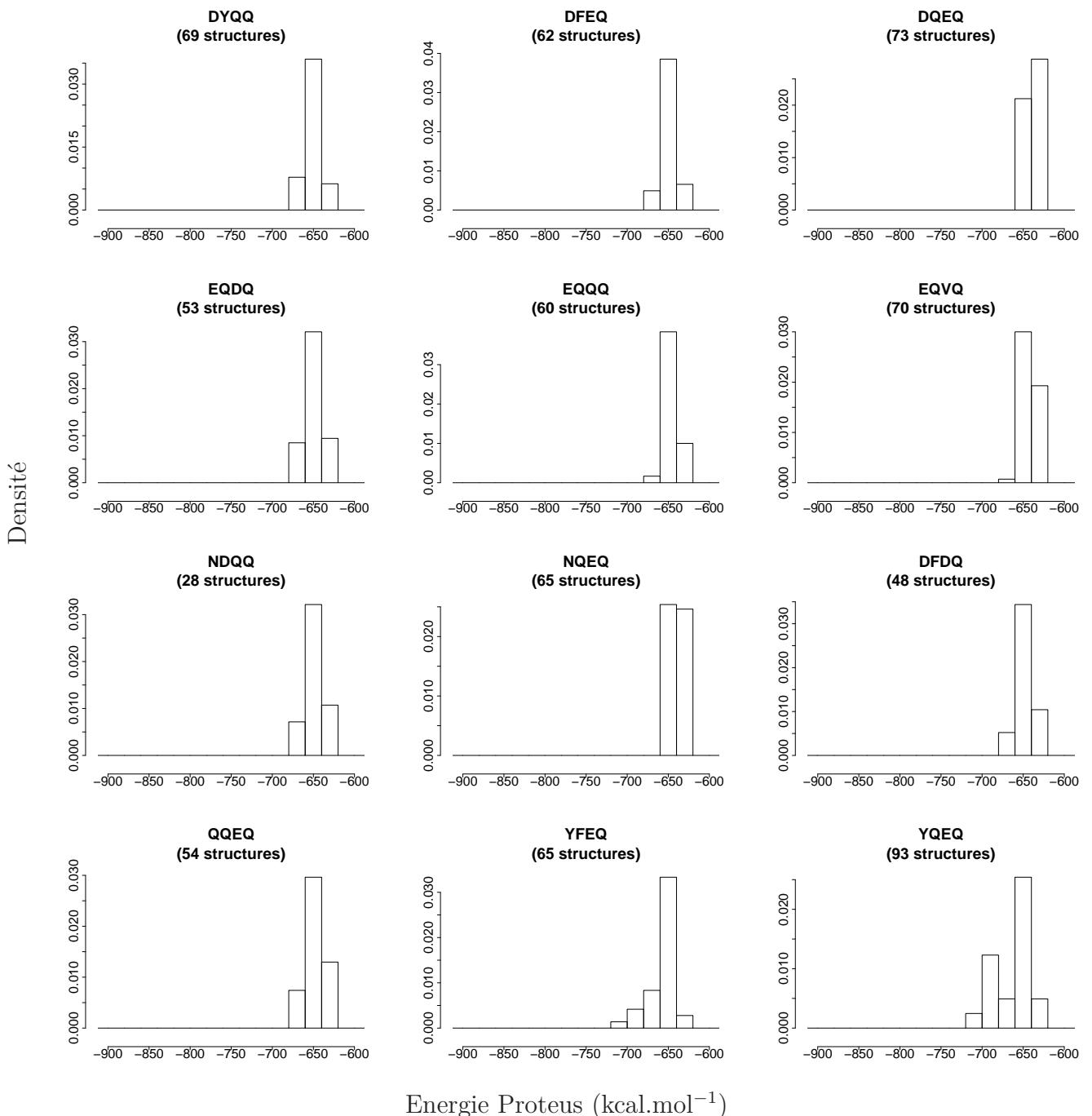
**TyrRS de *M. jannaschii*** Pour le ligand D-TyrAMP, les énergies de dépliement calculées par Proteus sont réparties entre  $-1040$  et  $-960 \text{ kcal.mol}^{-1}$  pour la native. Les mutants ont tous une énergie de dépliement plus grande (cf. figure 5.32 en Appendice). Avec le ligand L-TyrAMP, l'énergie de dépliement est entre  $-1040$  et  $-960 \text{ kcal.mol}^{-1}$ . Tous les mutants augmentent cette énergie de dépliement également, sauf LDDN et LQDN qui conservent des valeurs d'énergies de dépliement comparables à la native (cf. figure 5.33 en Appendice). Ce sont donc ces deux mutants, LDDN et LQDN, qui vont favoriser le ligand D-TyrAMP, car ils ont des structures plus stables que la native avec le D-TyrAMP, et de même stabilité que la native avec le ligand natif.

### 5.4.0.5 Répartition des énergies des structures reconstruites (énergies "Xplor")

**TyrRS d'*E. coli*** Pour les complexes TyrRS:L-TyrAMP, les énergies données par Xplor sont comprises entre  $-3400 \text{ kcal.mol}^{-1}$  et  $-3300 \text{ kcal.mol}^{-1}$  (figure 5.28). Pour les mutants YFEQQ et YQEQQ, il existe un autre minimum local d'énergie, entre  $-2800$  et  $-2600 \text{ kcal.mol}^{-1}$ . Pour les complexes avec D-TyrAMP, on distingue majoritairement deux groupes d'énergie : un premier entre  $-3400$  et  $-3200 \text{ kcal.mol}^{-1}$  et un second entre  $-2900$  et  $-2800 \text{ kcal.mol}^{-1}$ . Ceci est le cas pour la native et l'ensemble des mutants. Pour la native et les mutants DFEQQ, EQQQQ, QQEQQ et YFEQQ, un troisième groupe de structures existe, avec une énergie entre  $-2800$  et  $-2600 \text{ kcal.mol}^{-1}$  (figure 5.29).

**TyrRS de *M. jannaschii*** Les énergies pour les complexes TyrRS:D-TyrAMP sont comprises entre  $-2350$  et  $-2250 \text{ kcal.mol}^{-1}$  pour la TyrRS native. On constate que tous les mutants générés donnent des énergies sensiblement du même ordre de grandeur : les mutants créés sont donc très stables (cf. figure 5.34 en Appendice). De même, les énergies des complexe TyrRS:L-TyrAMP sont comprises entre  $-2400$  et  $-2300 \text{ kcal.mol}^{-1}$ . Les mutants donnent donc des structures d'énergies comparables à la native (cf. figure 5.35 en Appendice).

FIGURE 5.26 – Répartition des énergies de dépliement pour les complexes *Eco-TyrRS:D-TyrAMP*



#### 5.4. CARACTÉRISATION DES STRUCTURES OBTENUES

FIGURE 5.27 – Répartition des énergies de dépliement pour les complexes *Eco-TyrRS:L-TyrAMP*

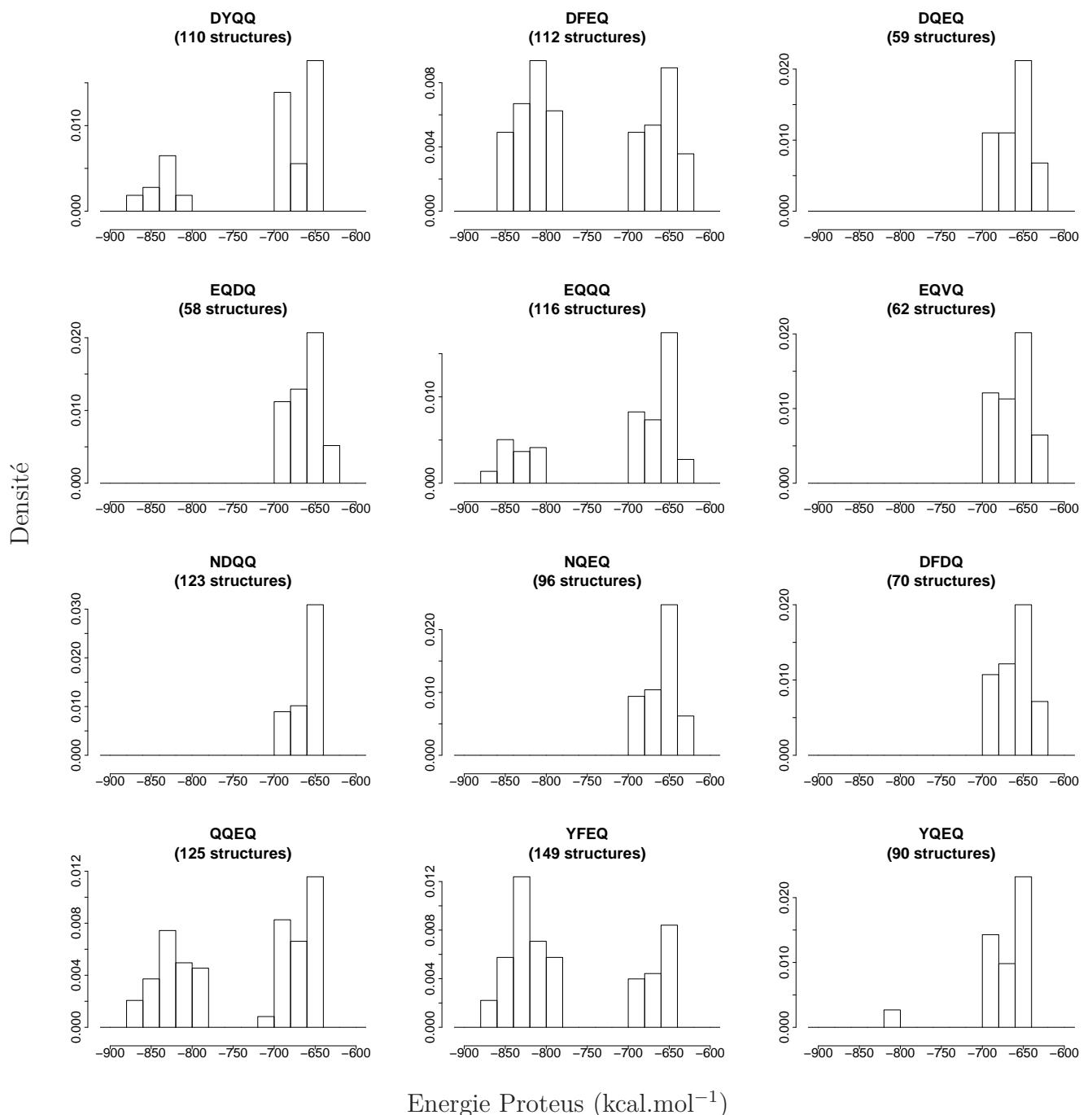
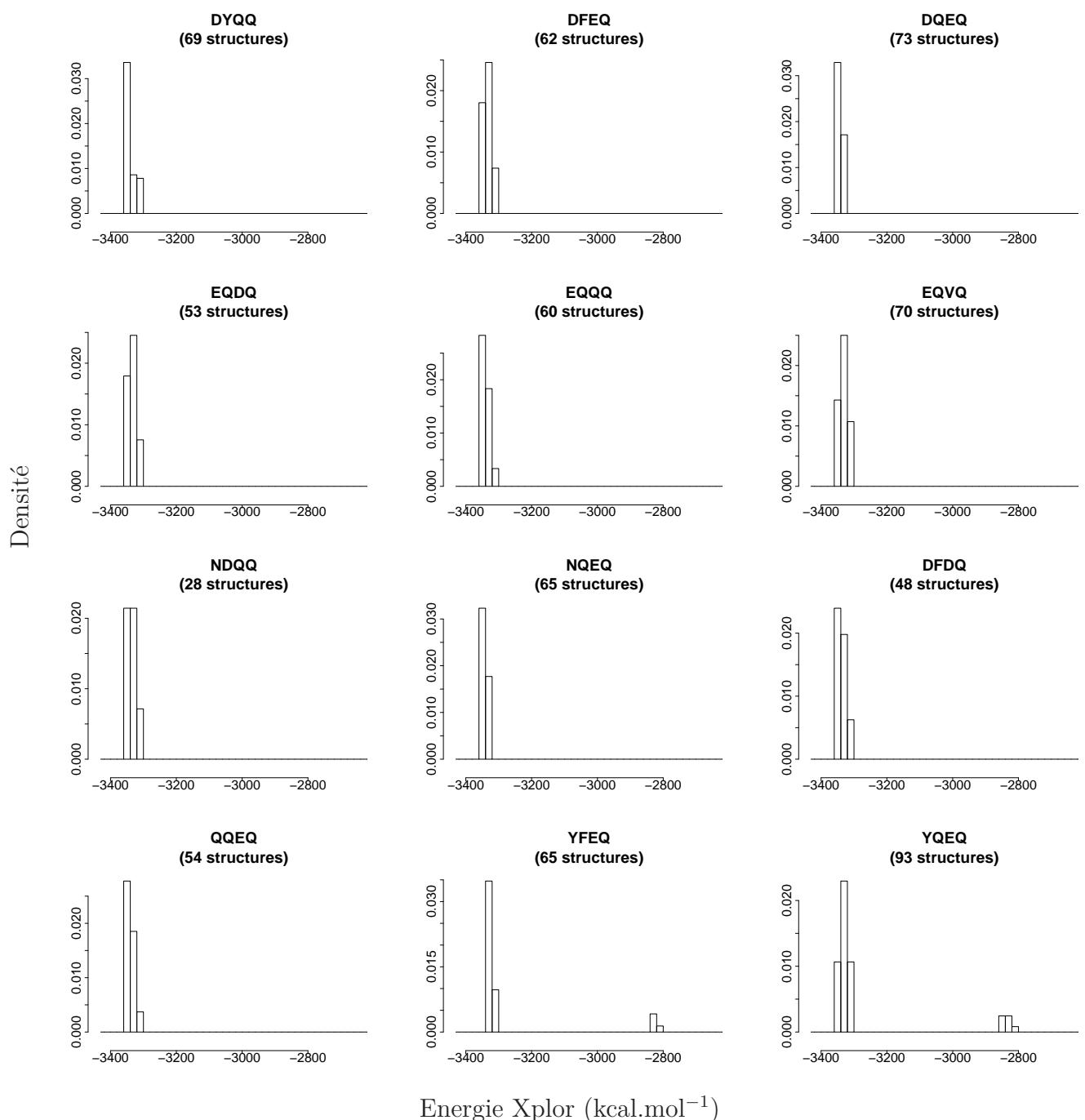
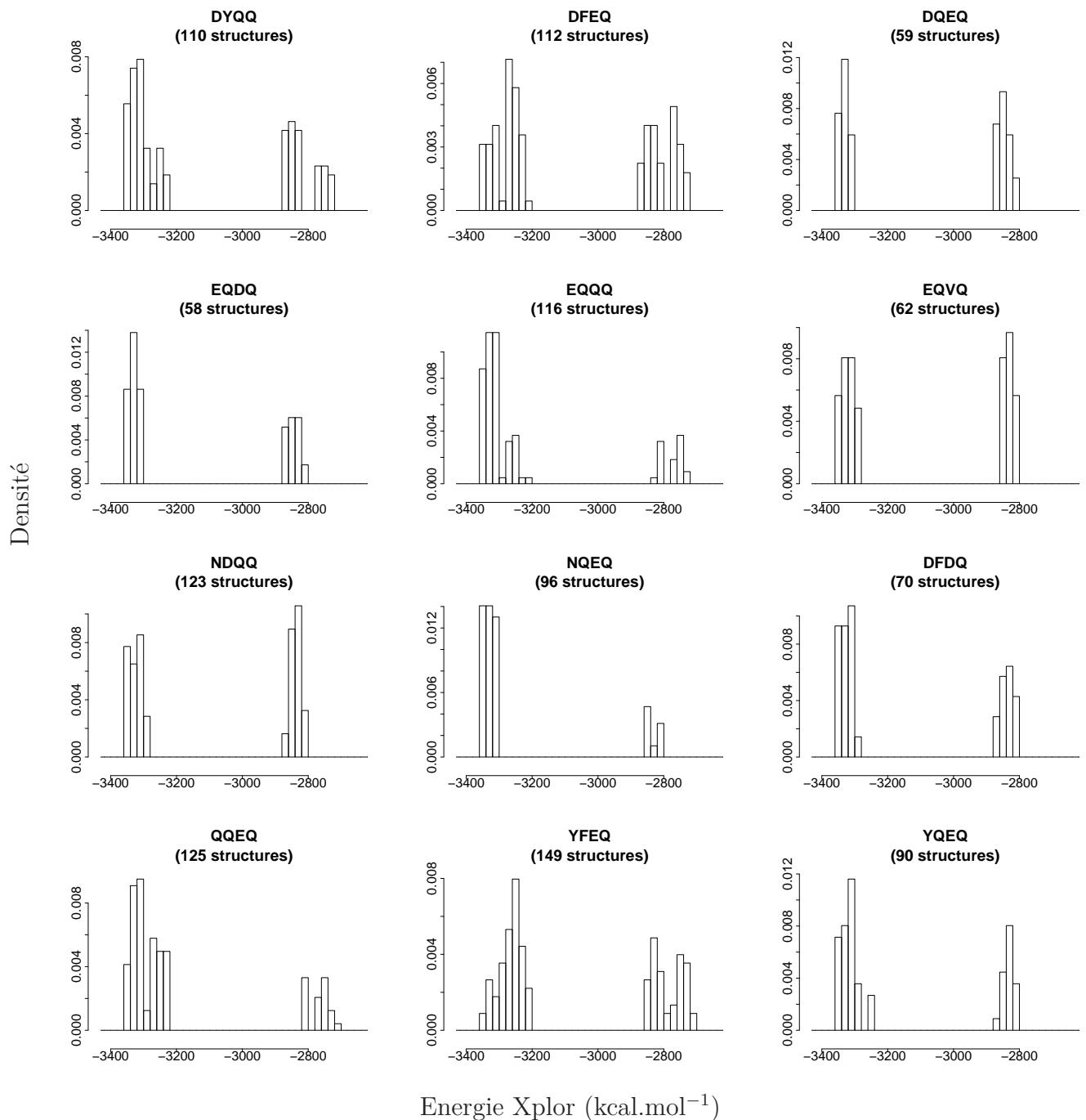


FIGURE 5.28 – Répartition des énergies des structures reconstruites pour les complexes *Eco*-TyrRS:D-TyrAMP



#### 5.4. CARACTÉRISATION DES STRUCTURES OBTENUES

FIGURE 5.29 – Répartition des énergies des structures reconstruites pour les complexes *Eco-TyrRS:L-TyrAMP*



## 5.5 Conclusions

Pour les prédictions de structures de TyrRS native liées au ligand natif, notre méthode a été en mesure de retrouver des structures proches de la structure cristallographique, groupées dans un minimum d'énergie local. Notre méthode a également échantillonné des structures correspondant à d'autres minima locaux d'énergie. Avec le D-TyrAMP, nous avons également généré plusieurs structures, explorant ainsi plusieurs minima d'énergie possibles. Dans le cas de l'apoenzyme, nous avons observé que la liberté de placement des chaînes latérales permet plus de conformations possibles. L'étude de sphères de différentes tailles, centrées sur le ligand, a montré que le site actif est l'endroit de la structure le plus concerné par l'optimisation de séquence et de structure. Aussi, pour une même séquence, les structures de TyrRS proposées par notre méthode sont très différentes selon le ligand.

Pour la génération de séquences mutantes par CPD, il est satisfaisant d'observer que nous retrouvons souvent l'acide aminé natif à une position mutante, ou bien un acide aminé avec des propriétés physico-chimiques proches. Nous générerons aussi parfois des acides aminés aux propriétés très différentes. Aussi, certaines combinaisons de mutations donnent finalement des acides aminés très proches de la séquence native, mais dans un ordre différent. Pourtant, quelques séquences sont très différentes et suscitent d'autant plus notre intérêt. Les séquences générées et leurs structures sont d'une stabilité comparable à la native. Il faut cependant noter que les mutations peuvent entraîner des modifications structurales du squelette, qui ne peuvent pas être retranscrites ici puisque nous maintenons le squelette protéique fixe. Au sujet des affinités des séquences générées, certains mutants peuvent présenter une affinité supérieure pour le ligand non naturel, mais également pour le ligand natif. D'autres sont plus spécifiques. Une telle étude systématique permet ainsi de se pencher à la fois sur le problème de l'affinité et sur celui de la spécificité. Nous verrons dans le chapitre suivant des analyses complémentaires sur les séquences les plus prometteuses.

**5.6 Appendix. Résultats supplémentaires pour *M. jannaschii***

FIGURE 5.30 – Energies de dépliement comparées aux énergies des structures reconstruites pour les complexes *Mja*-TyrRS:D-TyrAMP

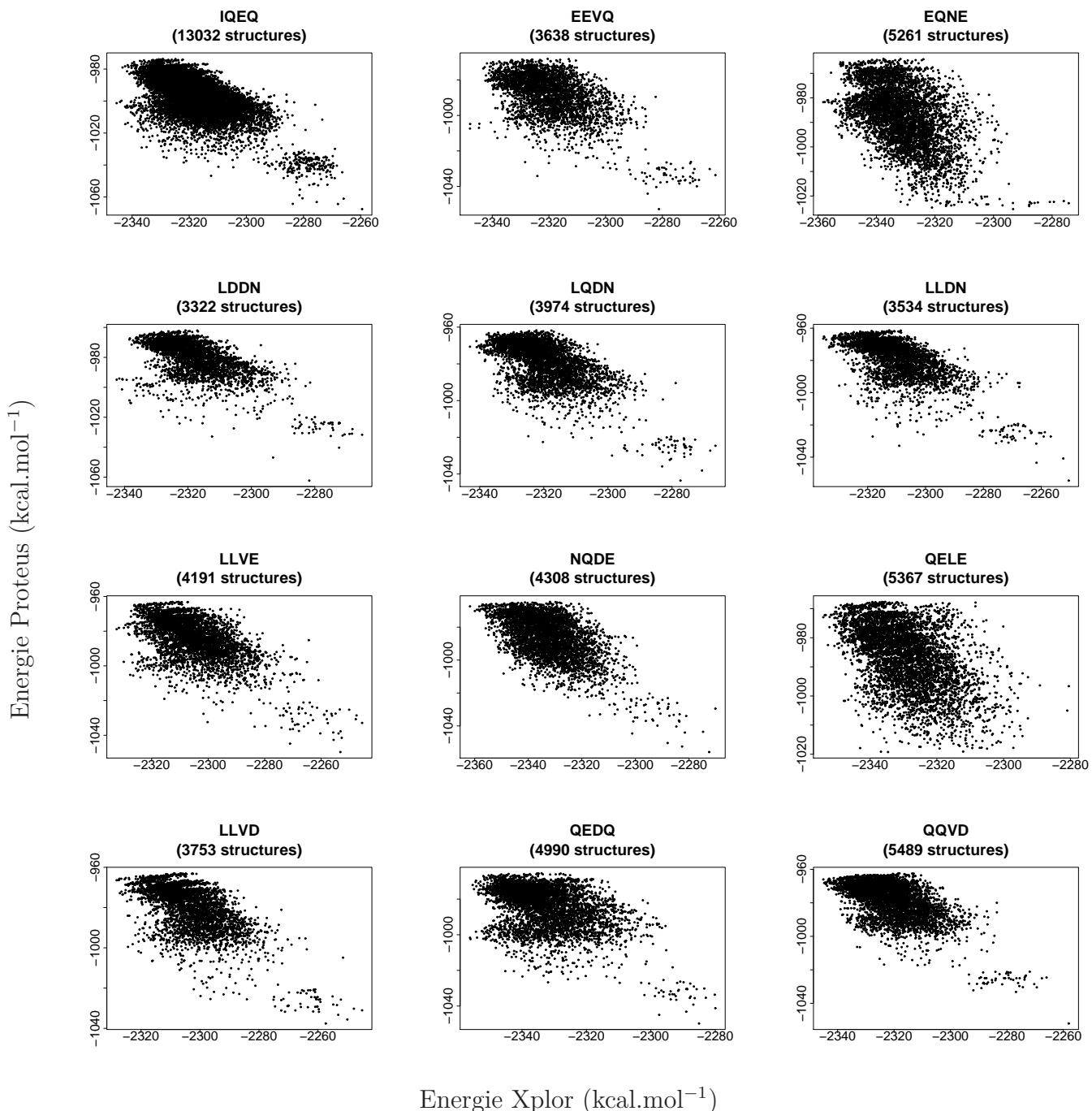


FIGURE 5.31 – Energies de dépliement comparées aux énergies des structures reconstruites pour les complexes *Mja*-TyrRS:L-TyrAMP

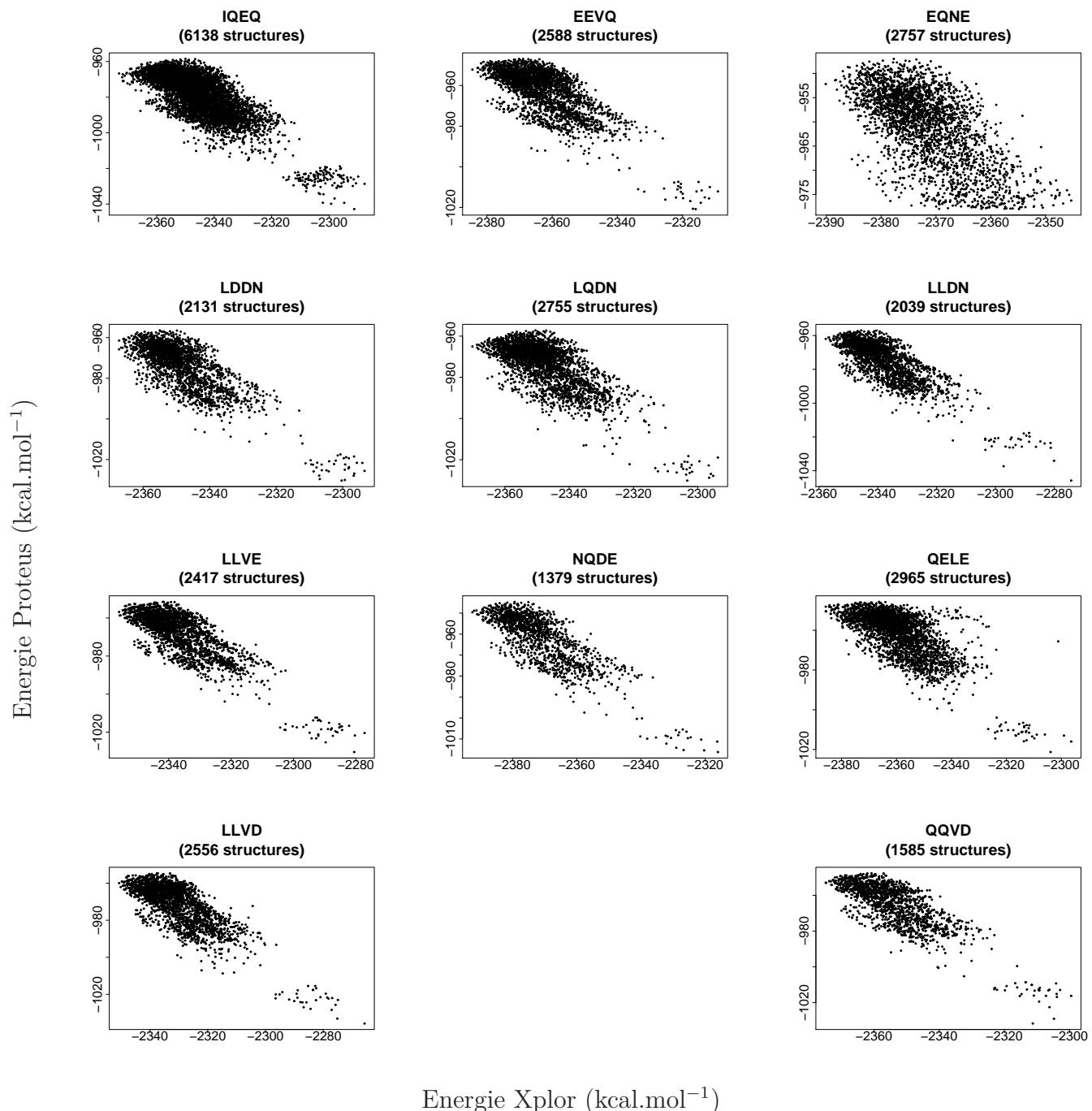


FIGURE 5.32 – Répartition des énergies de dépliement pour les complexes *Mja*-TyrRS:D-TyrAMP

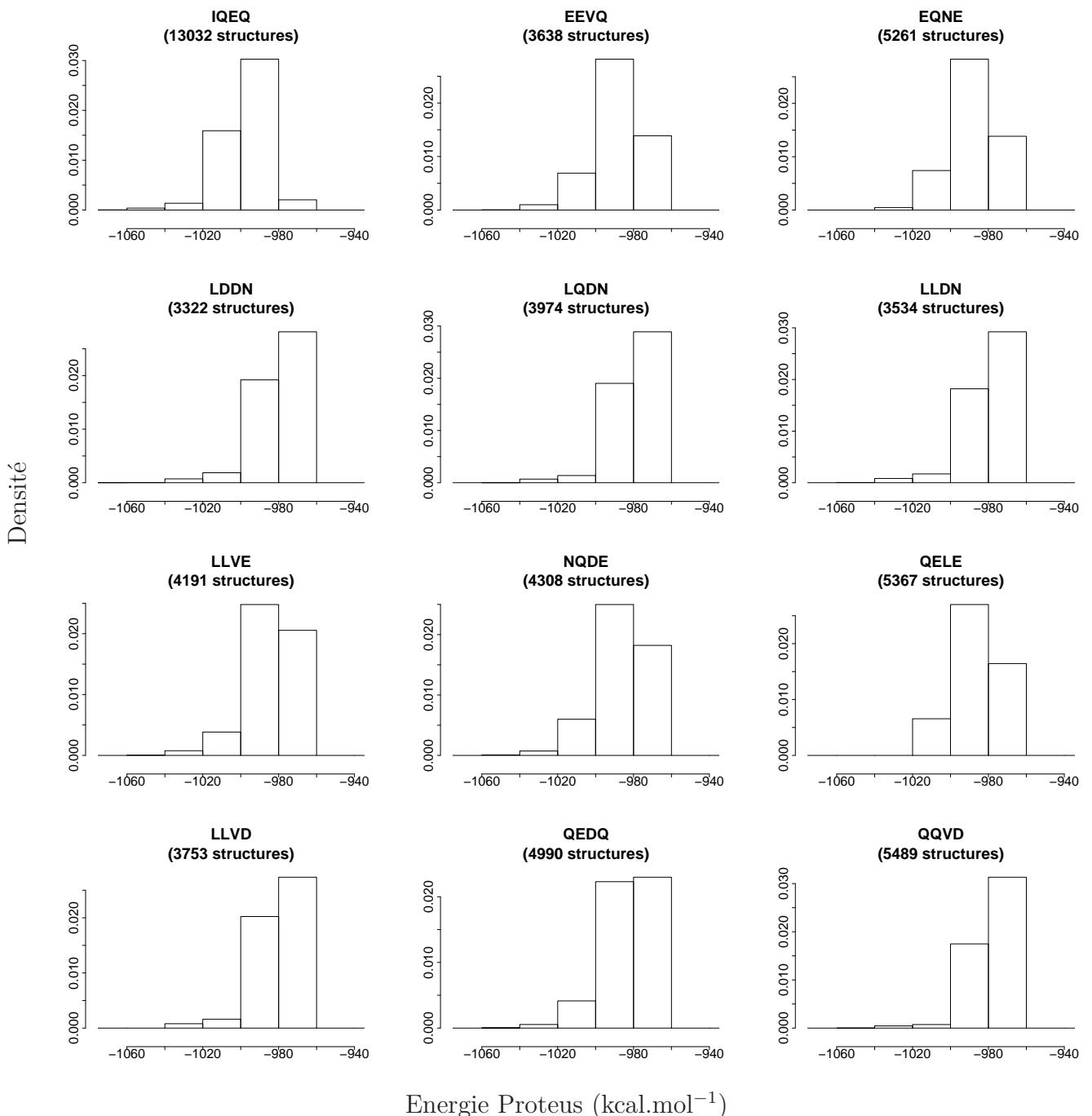


FIGURE 5.33 – Répartition des énergies de dépliement pour les complexes *Mja*-TyrRS:L-TyrAMP

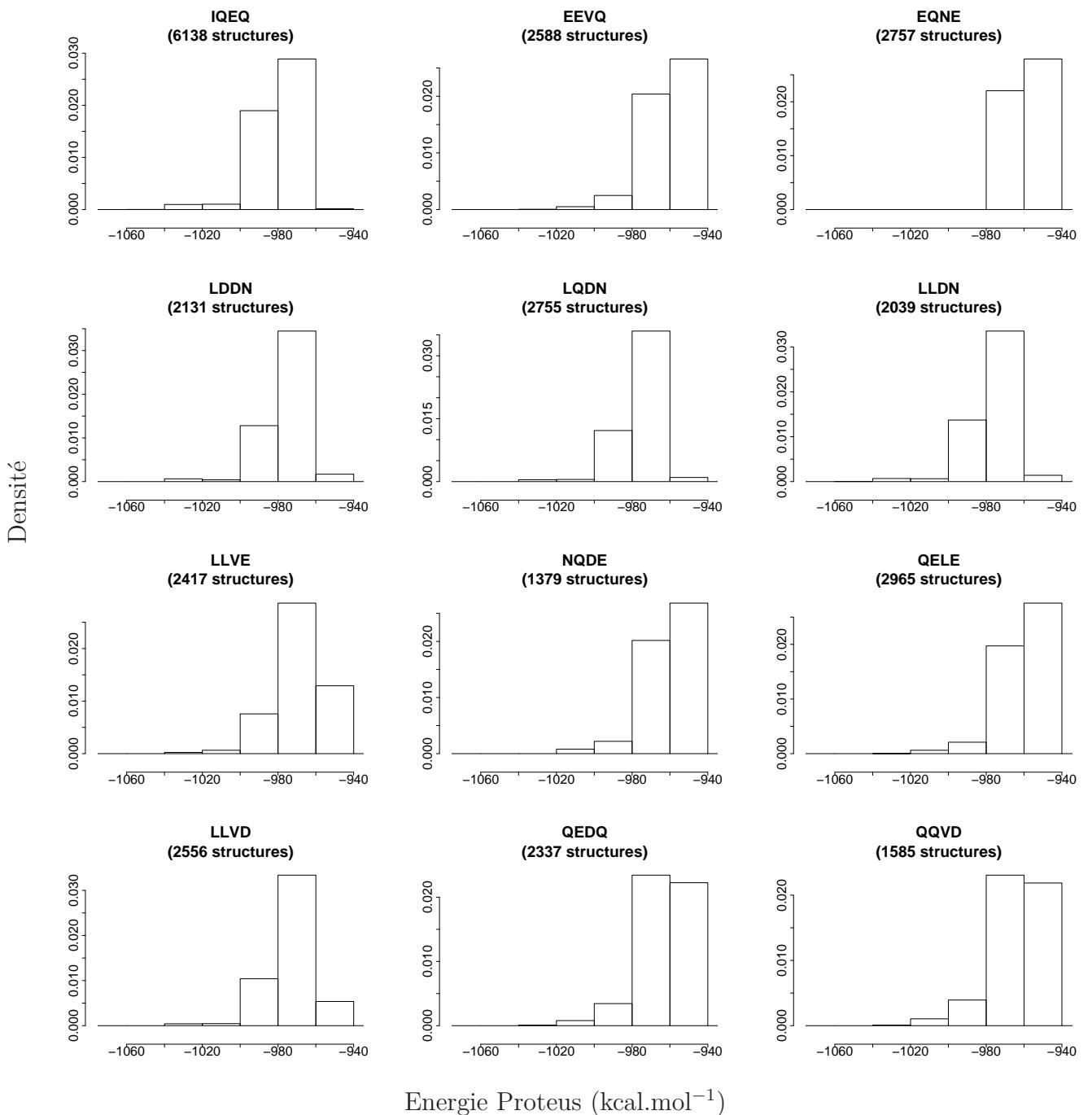


FIGURE 5.34 – Répartition des énergies des structures reconstruites pour les complexes *Mja*-TyrRS:D-TyrAMP

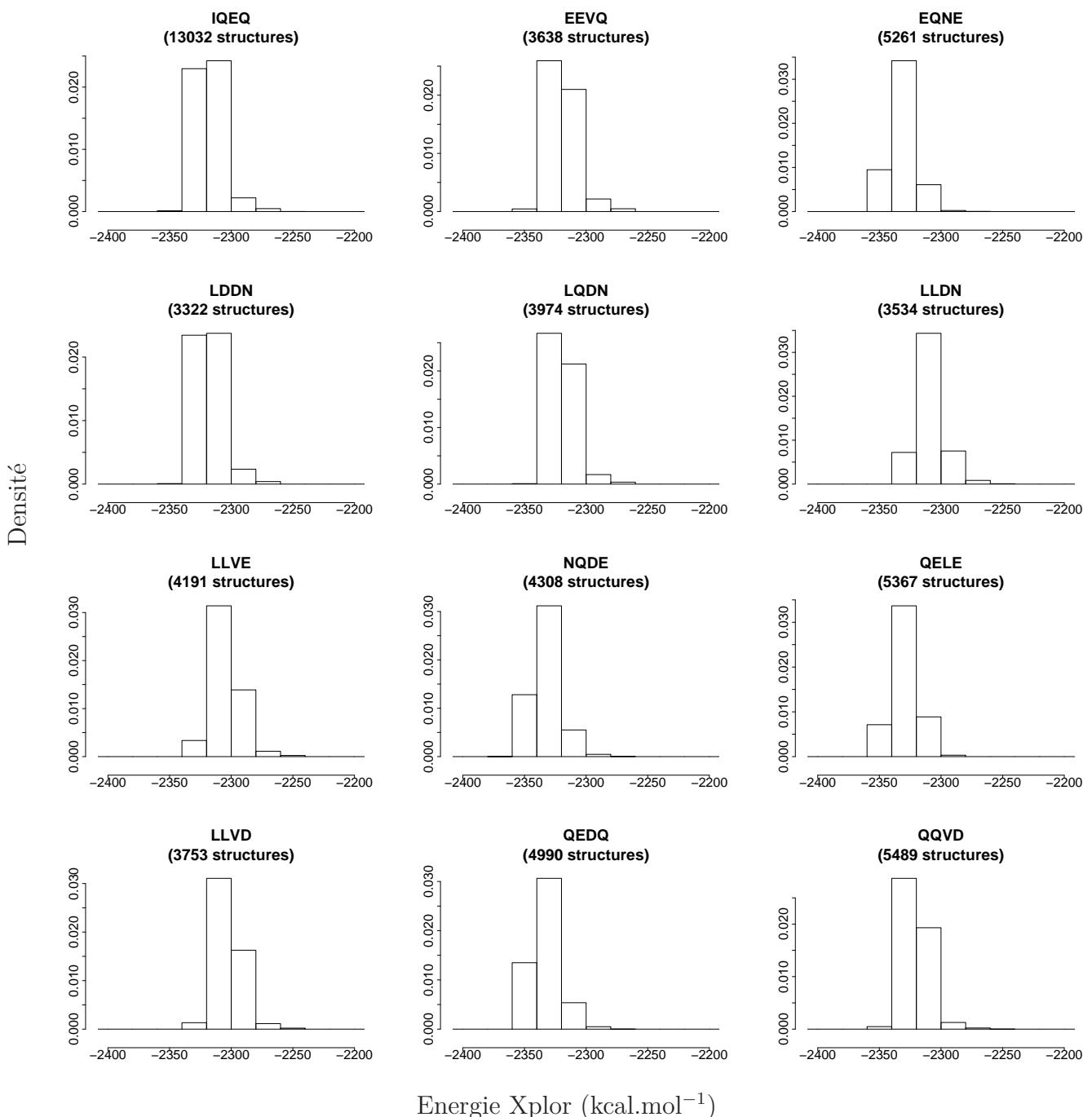
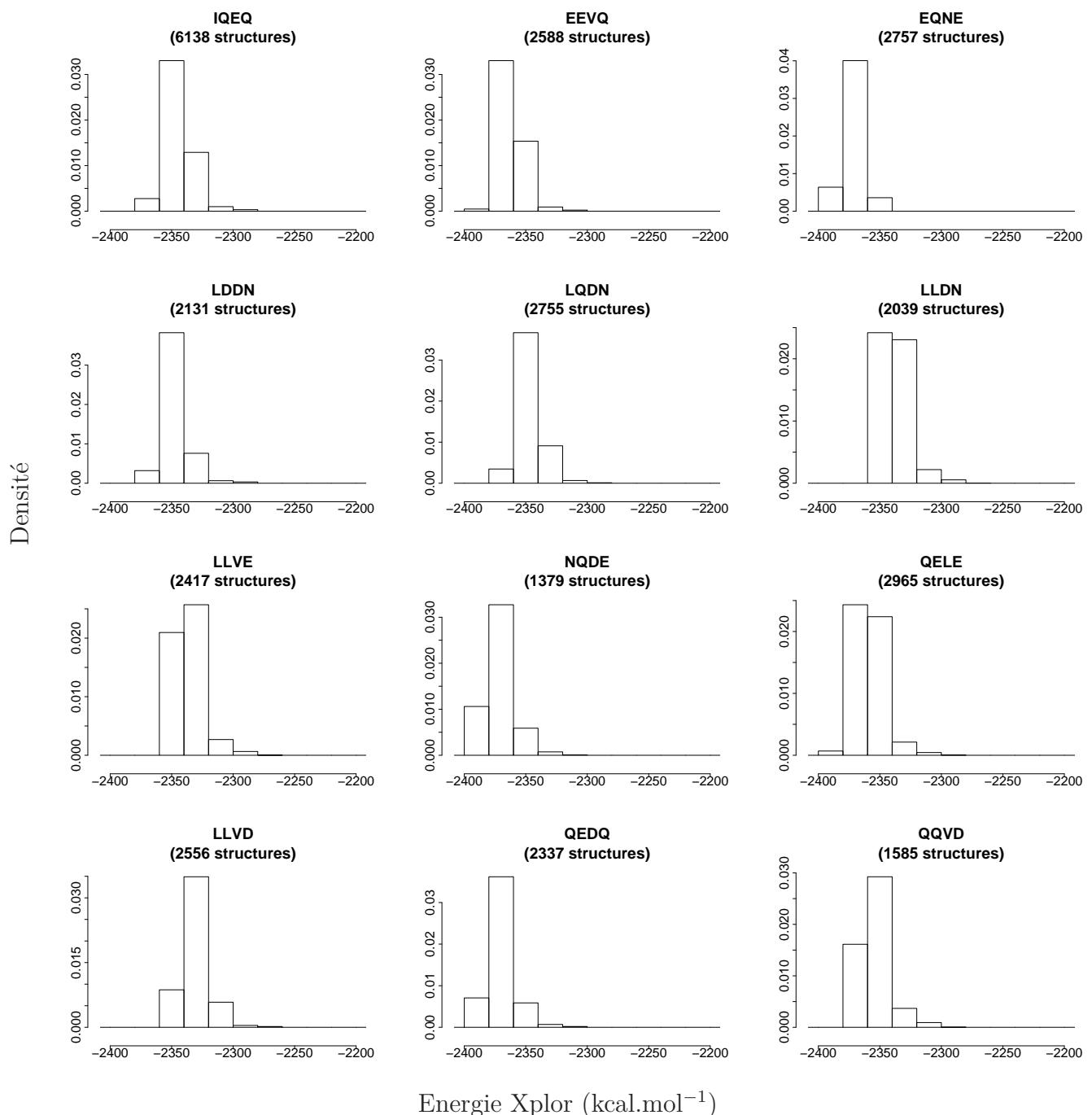


FIGURE 5.35 – Répartition des énergies des structures reconstruites pour les complexes *Mja*-TyrRS:L-TyrAMP





# Chapitre 6

## CPD de la TyrRS d'*E. coli*: Analyse des résultats d'évolution dirigée

Ce chapitre va plus en avant dans l'analyse des meilleures séquences obtenues à l'issue de la modélisation de l'évolution dirigée, à la fois avec les ligands L et D-Tyrosine, et avec les ligands L et D-TyrAMP. Les meilleures séquences prédites ont fait l'objet de calculs d'affinité, de simulations de dynamique moléculaire, puis d'estimations d'affinité par des calculs d'énergies libre par la méthode de résolution de l'équation de Poisson-Boltzmann (PBFE). Nous nous sommes également attachés à l'étude des structures 3D. Certaines séquences, enfin, ont fait l'objet d'analyses biochimiques. Nous recouperons toutes les informations disponibles pour une même séquence. Enfin, nous proposerons d'autres positions de la séquence qu'il pourrait être intéressant de muter.

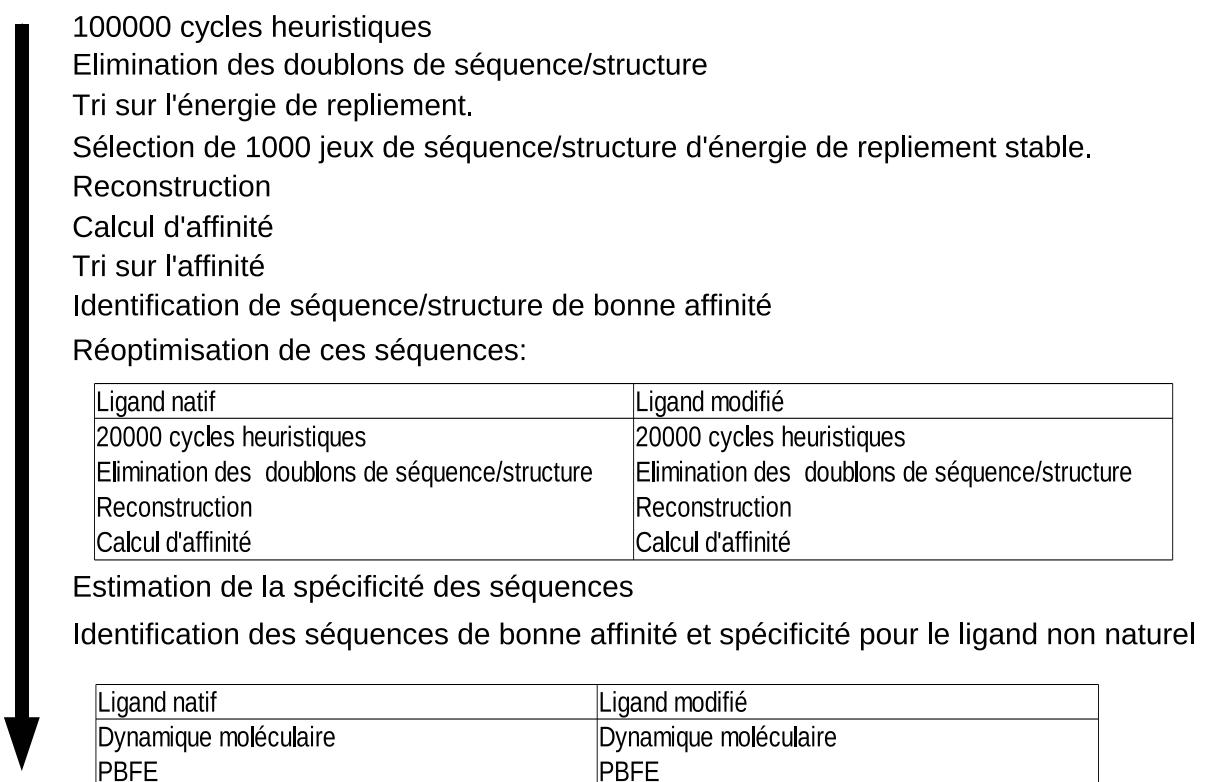
### 6.1 Choix des séquences à analyser

#### 6.1.1 CPD de la TyrRS liée à la D-Tyr

Le protocole qui a été utilisé pour l'évolution dirigée est rappelé dans la figure 6.1. Nous mutons les positions suivantes: D81, Y175, Q179 et Q201. Ce sont elles qui sont jugées responsables de la stéréo-spécificité de l'interaction avec le ligand. A l'issu des 100000 cycles heuristiques, 144 séquences uniques ont été obtenues. Leurs énergies de repliement

sont du même ordre de grandeur, entre -1730 et -1720 kcal.mol<sup>-1</sup>. On sélectionne les mille structures d'énergie de repliement les plus basses. Dans ce jeu de données, on ne retrouve plus que 71 séquences uniques. L'affinité est ensuite calculée pour ces mille structures, recouvrant les 71 séquences. On peut alors comparer la fixation à la D-Tyr de ces mutants par rapport à l'enzyme natif. Dans le tableau 6.1, on indique les mutants prédis donnant au moins une structure ayant une affinité supérieure ou comparable à la TyrRS native pour la D-Tyr.

FIGURE 6.1 – Schéma détaillé de la procédure de CPD appliquée à la TyrRS avec les ligands L-Tyr et D-Tyr. Une séquence/structure: constitue la séquence et le jeu de rotamères prédis à l'issue d'un cycle heuristique.



A partir de ce tableau, deux séquences semblent particulièrement intéressantes. La première est KYED, car elle est prédictive à travers 364 structures différentes, sur les mille sélectionnées. La deuxième est la séquence HDED, qui possède l'affinité moyenne la plus

## 6.1. CHOIX DES SÉQUENCES À ANALYSER

---

TABLE 6.1 – Séquences susceptibles de lier préférentiellement la D-Tyrosine. Energies en kcal.mol<sup>-1</sup>

Séquence	nombre de structures	charge	Energie de dépliement moyenne	écart type	Affinité moyenne	écart type	Meilleure affinité
DYQQ	256	-1	-1729.91	10.26	-7.89	1.78	-11.48
EKDD	85	-2	-1721.12	11.08	-21.88	6.83	-27.42
FYER	13	0	-1721.49	9.80	-19.82	1.66	-24.30
FYMD	2	-1	-1721.28	2.80	-22.00	2.10	-24.10
NWDD	2	-2	-1719.36	0.33	-22.23	1.11	-23.34
<b>HDED</b>	7	-2	-1720.04	2.52	-22.64	0.26	-23.02
KSDD	2	-1	-1722.95	0.76	-21.44	0.26	-21.70
KDED	3	-2	-1727.49	9.70	-20.83	0.23	-21.10
FDHD	2	-1	-1717.83	0.33	-17.50	2.21	-19.70
EKER	8	0	-1721.16	7.70	-18.23	0.28	-18.50
HDHD	1	0	-1717.61	0.00	-18.26	0.00	-18.26
DYER	6	-1	-1719.79	10.19	-12.79	2.44	-18.06
HDER	12	0	-1722.50	8.76	-16.36	1.55	-17.98
DDKE	3	-2	-1724.04	12.13	-17.75	0.16	-17.91
DWDD	18	-3	-1721.08	8.71	-16.09	0.93	-17.71
HDDD	3	-2	-1716.88	1.02	-9.12	5.95	-17.53
EHER	2	0	-1728.30	11.63	-16.60	0.91	-17.51
YWDD	1	-2	-1717.64	0.00	-16.77	0.00	-16.77
KDER	2	0	-1722.47	1.87	-16.57	0.02	-16.59
<b>KYED</b>	364	-1	-1716.96	8.20	-13.65	0.57	-15.44
DWDR	31	-1	-1717.91	4.22	-9.59	2.61	-15.37
DDKD	3	-2	-1732.96	4.15	63.25	54.36	-13.63
DWRD	1	-1	-1718.48	0.00	-12.63	0.00	-12.63
DDRD	9	-2	-1719.36	6.78	-10.33	1.50	-12.05
DDRE	2	-2	-1730.68	12.78	-11.76	0.01	-11.77
DWHD	6	-1	-1717.88	2.35	-9.59	0.67	-10.45

forte. Cependant, nous la remplacerons par HYED, qui est proche de la séquence HDED mais avec la position Y175 native. De plus, nous étudierons la séquence KYEQ, dérivant de la séquence KYED prédictive mais avec l'acide aminé Q201 natif. Ces "retouches", HDED→HYED et KYED→KYEQ, ont été effectuées pour se rapprocher de la séquence native et ne pas trop modifier l'environnement électrostatique du ligand ; en effet, KYED et HYED portent une charge de -1 sur ces quatre positions, comme la séquence native DYQQ.

Une autre séquence, EKDD, est prédictive à travers 87 structures sur mille ; mais l'introduction de trois charges supplémentaires dans le site actif ne paraît pas réaliste, compte tenu du simple changement d'énantiométrie du ligand. Les autres séquences sont également laissées de côté, car on ne veut pas introduire de tryptophane ou de phénylalanine qui favoriseraient à la fois la liaison de la L-Tyr et celle de la D-Tyr par effet hydrophobe avec le cycle aromatique de la tyrosine. On choisit donc de faire l'étude approfondie des trois séquences de TyrRS mutantes : HYED, KYED et KYEQ. Le mutant RYQQ proposé par les expérimentalistes sera également étudié.

### **6.1.2 CPD de la TyrRS liée au L-TyrAMP et au D-TyrAMP**

On veut à présent générer des mutants liant préférentiellement le D-TyrAMP. Pour cela, on réalise par Proteus 100000 cycles heuristiques avec le ligand D-TyrAMP, d'une part, et avec le ligand L-TyrAMP, d'autre part. On obtient 28 séquences prédictives. Pour chacune, on fait un rééchantillonage des rotamères à séquence fixe (figure: 6.2).

A l'issu de cette étude exhaustive, nous sélectionnons onze mutants sur la base de l'affinité au nouveau ligand D-TyrAMP. Sont sélectionnés tous les mutants donnant des structures de meilleures affinités au D-TyrAMP que la structure native. Aux positions 81, 75, 179 et 201, nous étudierons donc la séquence native DYQQ et les mutants sélectionnés: DFEQ, DQEQQ, EQDQQ, EQQQ, EQVQ, NDQQ, NQEQQ, DFDQQ, QQEQ, YFEQ, YQEQQ. La différence d'énergie libre de liaison  $\Delta\Delta G$  entre D-TyrAMP et L-TyrAMP est estimée à 8.53 kcal.mol<sup>-1</sup> en moyenne pour la TyrRS native. Pour tous les mutants, qu'ils aient été prédisits initialement avec le L-TyrAMP ou avec le D-TyrAMP, les énergies libres de

FIGURE 6.2 – Schéma détaillé de la procédure de CPD appliquée à la TyrRS avec les ligands L-TyrAMP et D-TyrAMP.

Mutagénèse combinatoire à travers 100000 cycles heuristiques

Reconstruction des structures 3D

Calcul d'affinité protéine-ligand

Réoptimisation de toutes les séquences prédites:

- 20000 cycles heuristiques (séquence fixe)
- Reconstruction des structures 3D
- Calcul d'affinité

Sélection des séquences de bonne affinité et spécificité pour le ligand non naturel

Simulation des complexes TyrRS:Ligand par dynamique moléculaire

Affinité estimée avec un modèle diélectrique continu (méthode “PBFE”)



liaison indiquent qu'ils favoriseraient la liaison du D-TyrAMP, car leur  $\Delta\Delta G$  est négative (table 6.2).

TABLE 6.2 – Affinités moyennes des mutants et de la native en complexe avec le D-TyrAMP ou avec le L-TyrAMP ( $\text{kcal} \cdot \text{mol}^{-1}$ ). La première colonne indique avec quel ligand le mutant a été prédict initialement. La dernière colonne indique la différence d'affinités entre le D-TyrAMP et le L-TyrAMP.

Ligand	Séquence	Affinité L-TyrAMP moyenne	Affinité L-TyrAMP min	Affinité D-TyrAMP moyenne	Affinité D-TyrAMP min	$\Delta\Delta G$
	DYQQ	-52.85	-56.85	-44.32	-50.59	8.53
D	DQEQQ	-54.88	-59.39	-58.61	-61.60	-3.73
L&D	NQEQQ	-50.80	-58.21	-54.19	-58.43	-3.39
D	DFDQ	-48.08	-56.98	-58.92	-61.69	-10.84
D	EQDQ	-46.97	-55.35	-57.61	-61.16	-10.64
L&D	QQEQ	-47.17	-55.24	-48.76	-53.37	-1.59
L&D	NDQQ	-43.53	-55.11	-47.79	-52.86	-4.26
D	EQVQ	-46.02	-55.09	-50.62	-53.30	-4.6
D	EQQQ	-46.79	-55.06	-50.92	-55.77	-4.13
D	DFEQ	-47.80	-53.43	-56.88	-60.17	-9.08
L&D	YFEQ	-34.37	-53.07	-45.33	-57.21	-10.96
L&D	YQEQQ	-36.15	-52.82	-43.44	-57.07	-7.29

## 6.2 Analyse générale des séquences obtenues

Dans cette section nous considérons l'ensemble des séquences prédictes. Dans les sections suivantes, nous les considérerons individuellement.

### 6.2.1 Vue d'ensemble

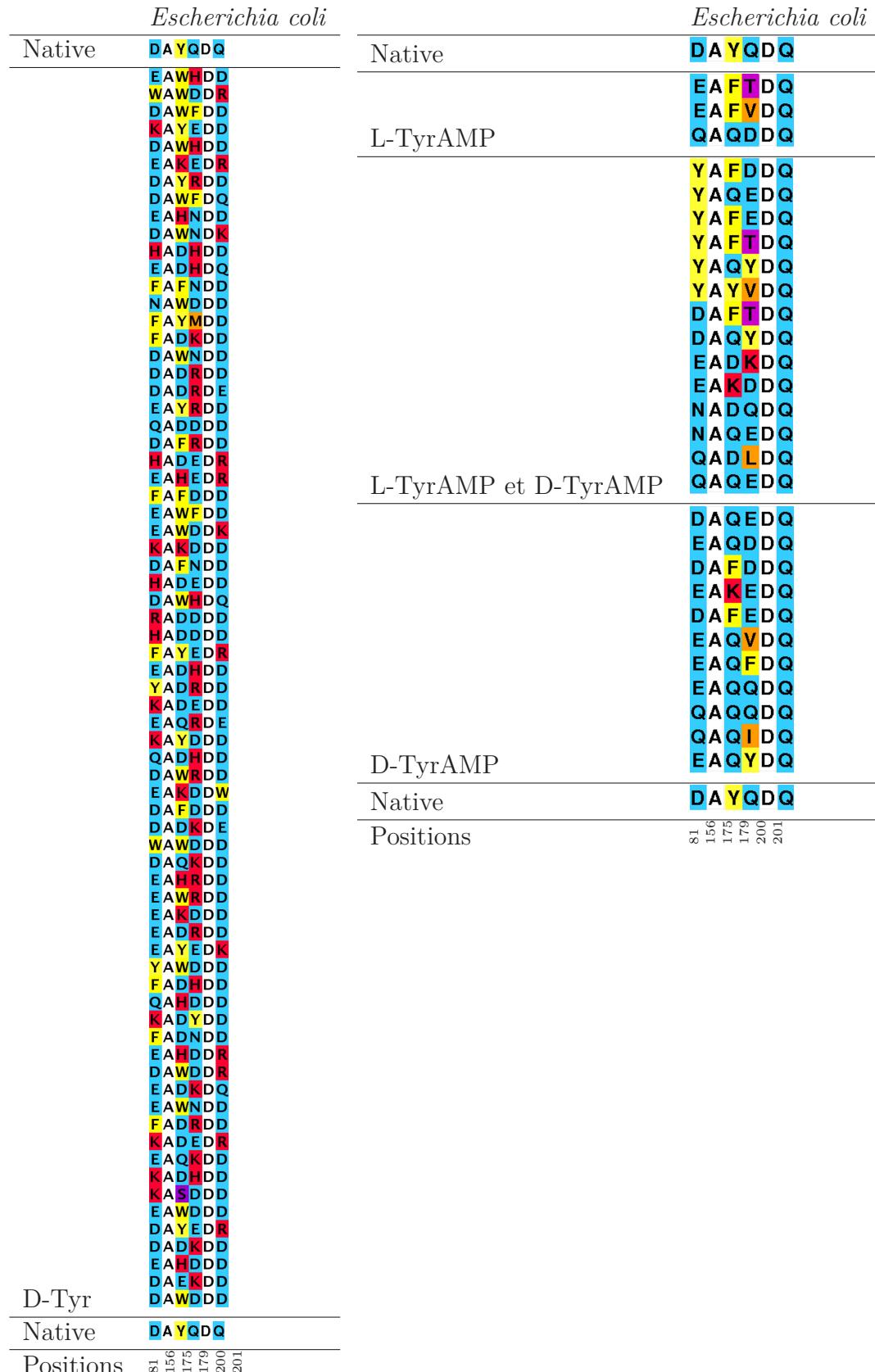
Nous regarderons ici les propriétés des acides aminés prédicts à chacune des quatre positions actives. Les séquences sont indiquées dans la figure 6.3.

Dans le cas de la TyrRS en liaison à la D-Tyr, 71 séquences sont identifiées. A la position 81, on retrouve des acides aminés chargés positivement ou négativement ; on retrouve également quelques résidus aromatiques. Nous avons vu que les résidus aromatiques avaient tendance à stabiliser à la fois la liaison de la L-Tyr et celle de la D-Tyr, car l'effet hydrophobe sur le cycle aromatique du ligand joue un rôle important dans la

## 6.2. ANALYSE GÉNÉRALE DES SÉQUENCES OBTENUES

FIGURE 6.3 – Séquences prédites par notre protocole d'évolution dirigée, en fonction du ligand.

(a) Evolution dirigée avec le ligand  
(b) Evolution dirigée avec les ligands D-TyrAMP et L-TyrAMP



fixation des deux ligands. Les mutations intéressantes à cette position sont donc celles qui transforment le résidu natif chargé négativement en résidu chargé positivement. A la position 175, on retrouve beaucoup d'acides aminés aromatiques, qui rappellent la tyrosine native. Mais on y trouve également des résidus chargés positivement ou négativement. On remarque également qu'une séquence est prédictive avec une sérine à cette position. A la position 175, on retrouve principalement des résidus chargés positivement ou négativement. On trouve dans quatre séquences une tyrosine, alors qu'un autre mutant place une méthionine à cette position. A la position 201, l'acide aminé natif est prédicté par deux séquences ; les autres séquences présentent des acides aminés chargés positivement ou négativement. Un mutant place un tryptophane à cette position, sans doute pour remplacer l'effet de la tyrosine 175 mutée en lysine dans cette séquence.

Dans le cas de la TyrRS d'*E. coli* en liaison au L-TyrAMP ou au D-TyrAMP, 28 séquences mutantes ont été identifiées, dont 14 à la fois avec le L-TyrAMP et le D-TyrAMP, ce qui laisse penser qu'elles ne sont pas très spécifiques. Des lysines en positions 175 et 179 apparaissent avec le ligand D-TyrAMP. Des acides aminés hydrophobes L, V, et I à la position 179 apparaissent avec le ligand D-TyrAMP. Des phénylalanines et des tyrosines sont prédictives aux positions 81, 175 et 179, autant avec le L-TyrAMP qu'avec le D-TyrAMP. Les forces hydrophobes stabilisent le cycle aromatique de la partie tyrosine du ligand sans favoriser ni la forme L ni la forme D. Un fait remarquable est que la position 201 ne soit jamais mutée, bien qu'elle soit autorisée à le faire. L'observation des structures a conduit à identifier une interaction très forte entre l'aspartate 200 et la glutamine 201, qui expliquerait que cet acide aminé soit systématiquement prédicté sous sa forme native.

On prévoit globalement plus de résidus chargés positivement avec l'acide aminé seul qu'avec le ligand adénylé. La présence de l'AMP favorise clairement les acides aminés chargés négativement aux trois premières positions mutées. La position 201, autorisée à muter avec tous les ligands, reste fixée sur le type natif quand le ligand est adénylé. On prévoit un acide aminé chargé à cette position quand la protéine est simulée avec le ligand D-Tyr.

Une autre séquence a été prédictive par observation structurale, par les expérimentalistes,

## 6.2. ANALYSE GÉNÉRALE DES SÉQUENCES OBTENUES

Sandra Wydau et Pierre Plateau. Il s'agit du mutant RYQQ. L'arginine à la position 81, permettrait de défavoriser le positionnement du groupement ammonium de la L-Tyr, et de favoriser celui de la D-Tyr.

### **6.2.2 Etude de la TyrRS native**

On réalise 100000 cycles heuristiques pour explorer les rotamères de la TyrRS native avec les différents ligands. Un filtre est appliqué pour ne garder qu'un seul exemplaire des structures identiques. Toutes les structures différentes sont reconstruites à partir du squelette protéique cristallographique et des prédictions de rotamères. L'affinité moyenne sur toutes ces structures est calculée. Cette moyenne donne une bonne estimation de l'affinité pour un ligand à travers différentes conformations structurales d'une même séquence. Les comparaisons d'affinités d'une même séquence pour différents ligands nous permet d'en estimer la spécificité. Les résultats sont résumés dans la table 6.3.

TABLE 6.3 – TyrRS native. Optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Energie de dépliement moyenne *	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-Tyr	256	-1729.91	-7.89	1.78	-11.48
TyrRS:L-Tyr	678	-1729.40	-14.71	4.04	-22.21
TyrRS:D-TyrAMP	69	-648.98	-44.32	4.68	-50.59
TyrRS:L-TyrAMP	110	-713.51	-52.85	2.16	-56.85

\* : Plus l'énergie de dépliement est élevée, plus il est facile de dénaturer le système.

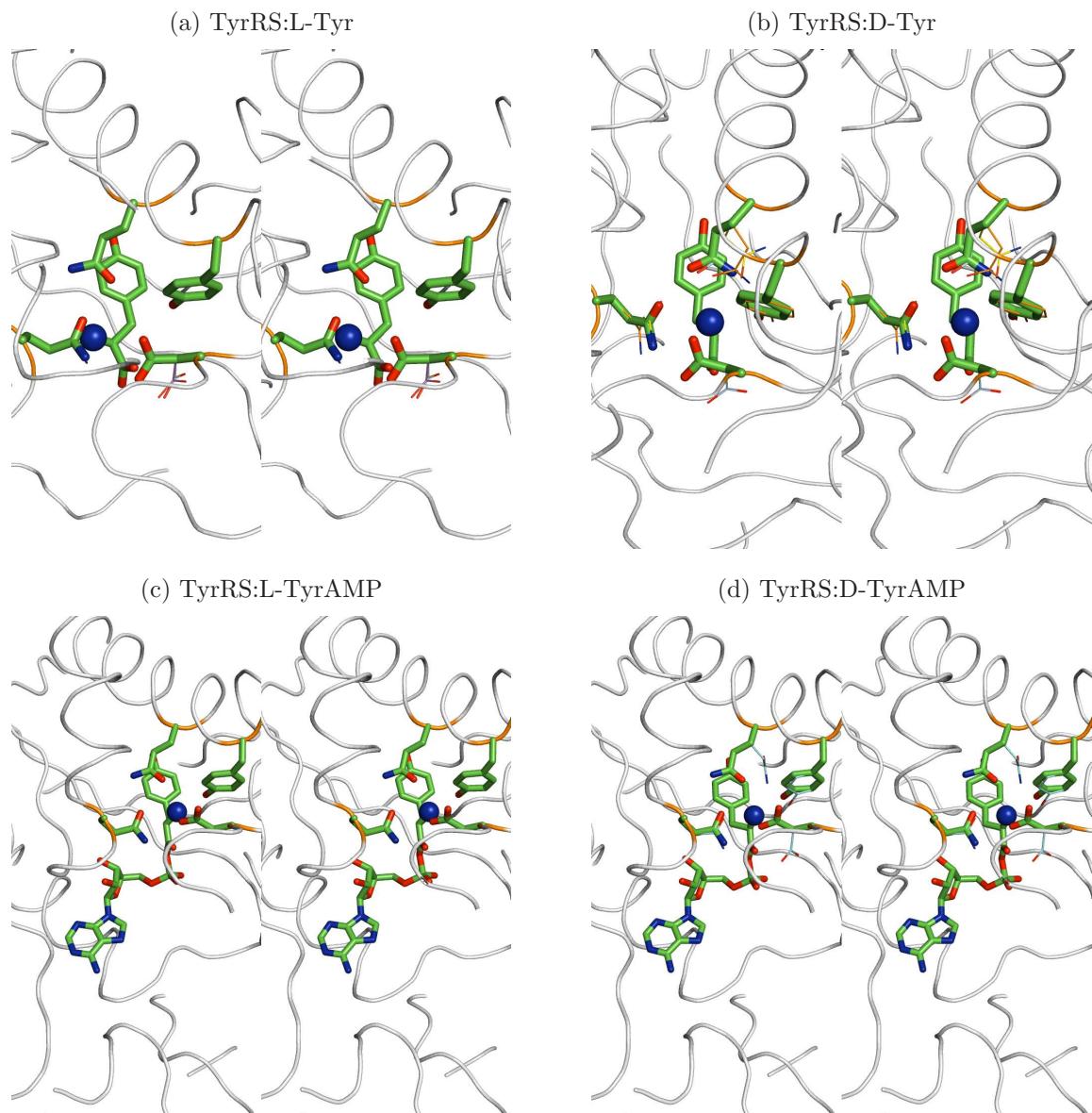
On constate que l'énergie moyenne de dépliement des complexes TyrRS:Tyr est ici très différente de celle des complexes TyrRS:TyrAMP. Ceci s'explique par le fait que le système étudié dans chaque cas est une sphère de 25 Å autour du ligand. La sphère du complexe TyrRS:TyrAMP est composée de plus de résidus que la sphère du complexe TyrRS:Tyr. On constate également qu'il y a une influence de la forme L ou D du ligand sur l'énergie de dépliement du complexe TyrRS:TyrAMP. Le D-TyrAMP donnerait une énergie de dépliement plus stable au complexe. Cette différence n'est pas observée pour le complexe TyrRS:Tyr.

Il a été déterminé expérimentalement chez *Bst*-TyrRS que la différence d'énergie libre de liaison entre D-Tyr et L-Tyr est de 1.9 kcal.mol<sup>-1</sup>. Ici, si nous regardons la meilleure

affinité pour chaque complexe,  $\Delta\Delta G = 10.73 \text{ kcal.mol}^{-1}$ , alors que cet écart n'est plus que de  $6.26 \text{ kcal.mol}^{-1}$  lorsque le ligand est adénylé. Notre méthode reproduit la tendance expérimentale, mais les différences d'énergies libres sont surestimées.

La figure 6.4 montre la superposition des cinq structures prédictes de meilleures affinités avec les quatre ligands. Avec le ligand L-Tyr, seul l'Asp81 présente des conformations différentes, alors qu'avec le ligand D-Tyr, les quatres positions présentent plusieurs rotamères distincts. Les conformations des rotamères restent très proches quel que soit le ligand. On constate cependant une différence qui semble importante pour la spécificité entre L-Tyr et D-Tyr au niveau du résidu Gln179. Les atomes OE1 et NE2 sont inversés avec les deux ligands. Pour le complexe TyrRS:TyrAMP, il n'y a pas d'inversion.

FIGURE 6.4 – TyrRS native. Superposition des 5 structures de meilleures affinités.



### 6.2.3 Séquences prédictes avec le ligand D-Tyr

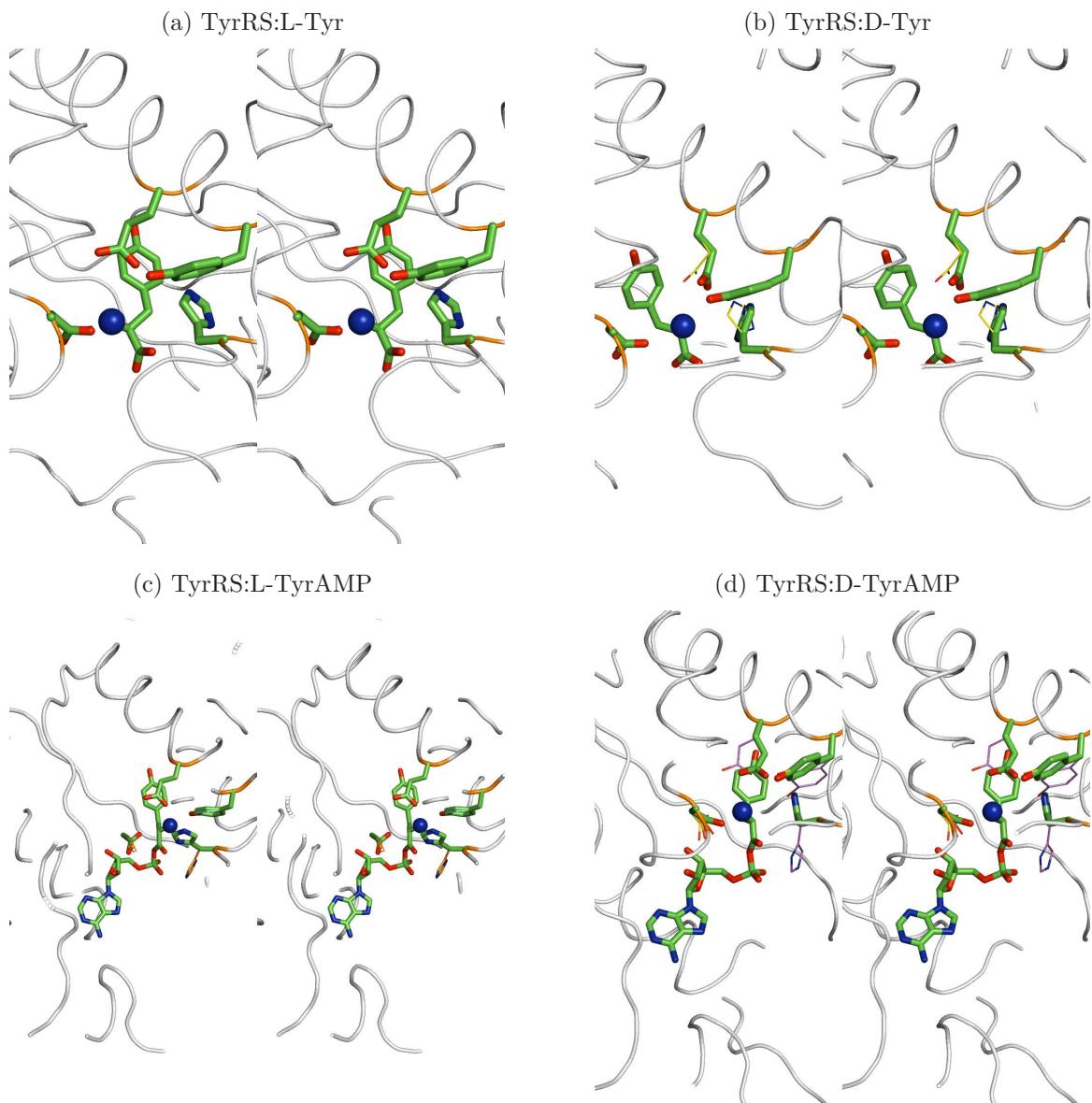
#### 6.2.3.1 Mutant HYED

TABLE 6.4 – Mutant HYED. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>.

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-Tyr	342	-1851.66	-8.68	2.35	-15.85
TyrRS:L-Tyr	977	-1848.75	-18.55	4.20	-27.78
TyrRS:D-TyrAMP	57	-801.25	-40.92	5.89	-46.98
TyrRS:L-TyrAMP	170	-863.84	-47.68	5.95	-55.62

La séquence HYED a été prédictée en liaison avec la D-Tyr. Les rotamères ont été ré-optimisées par 20000 cycles heuristiques avec chaque ligand. Le complexe TyrRS:TyrAMP apparaît beaucoup plus stable que le complexe TyrRS:Tyr (Table 6.4). Ce mutant qui donne une bonne affinité pour la D-Tyr, favoriserait également la L-Tyr. Il n'a donc pas une bonne spécificité pour la D-Tyr. Les structures de meilleures affinités sont montrées dans la figure 6.5. On note pour le complexe TyrRS:TyrAMP que l'histidine peut avoir deux orientations du cycle bien différentes, alors qu'elle reste dans le même emplacement pour le complexe TyrRS:Tyr.

FIGURE 6.5 – Mutant HYED. Superposition des 5 structures de meilleures affinités.



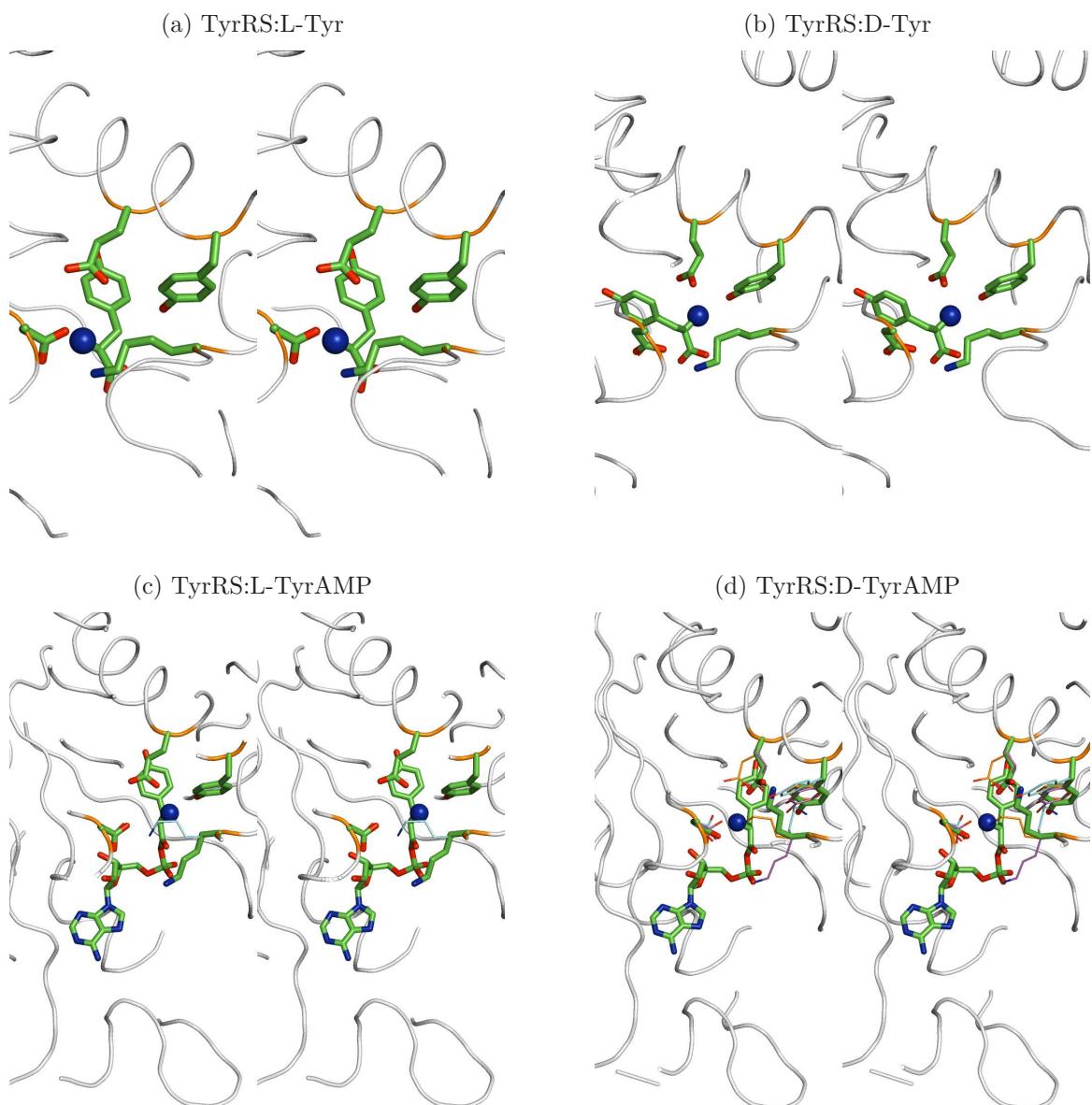
### 6.2.3.2 Mutant KYED

TABLE 6.5 – Mutant KYED. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Ener dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-Tyr	969	-1738.28	-6.98	1.09	-10.20
TyrRS:L-Tyr	993	-1738.15	-20.10	1.79	-24.66
TyrRS:D-TyrAMP	89	-706.36	-28.53	6.29	-45.58
TyrRS:L-TyrAMP	261	-849.19	-39.48	9.91	-54.11

En se basant sur les affinités moyennes (table 6.5), on constate que ce mutant lierait la L-Tyr mieux que la TyrRS native et moins bien la D-Tyr. Aussi, il lierait les deux ligands adénylés moins bien que la native. Sur la figure 6.6, nous voyons que la chaîne latérale du ligand n'est pas orientée de la même manière pour les conformations L et D. Aussi, la chaîne latérale du Glu179 est plus proche de l'ammonium dans les structures avec la forme D du ligand. Avec le D-TyrAMP, la Tyr175 peut également se trouver dans des conformations qui rapprochent son groupe hydroxyle de l'ammonium du ligand. Le positionnement de la Lys81 est singulière. Cette lysine stabilise le groupement carboxylate des tyrosines L et D. Mais en présence de l'AMP, il semble y avoir plusieurs orientations possibles. La première placerait l'extrémité de la chaîne latérale de la lysine vers le phosphate, la deuxième la placerait vers l'ammonium. En présence de D-TyrAMP, on peut avoir un déplacement de la Tyr175 et du Glu179 qui s'adaptent à la conformation D du ligand.

FIGURE 6.6 – Mutant KYED. Superposition des 5 structures de meilleures affinités.



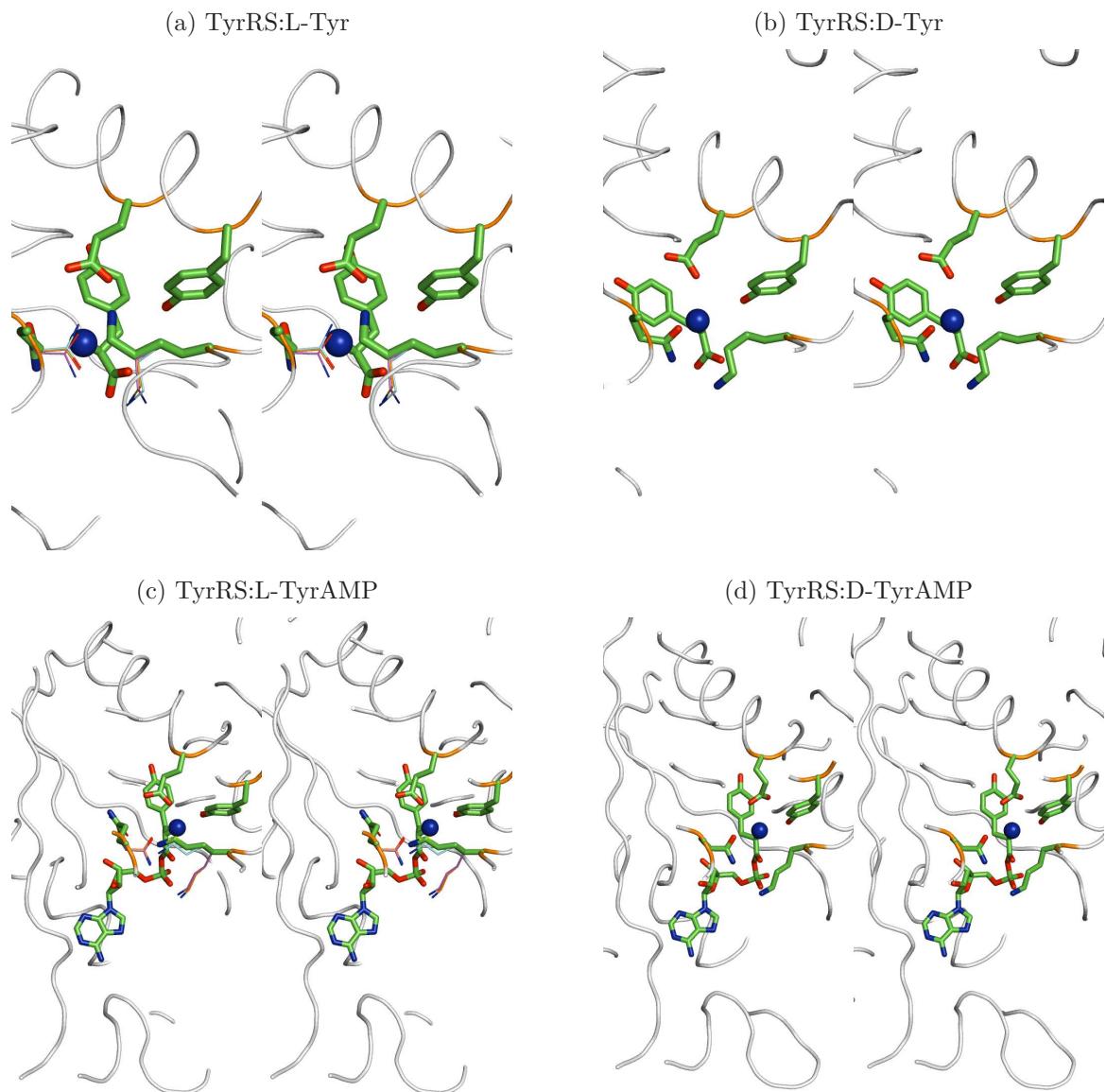
### 6.2.3.3 Mutant KYEQ

TABLE 6.6 – Mutant KYEQ. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Ener dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-Tyr	356	-1745.02	-9.46	3.03	-17.19
TyrRS:L-Tyr	925	-1744.91	-13.85	2.93	-28.19
TyrRS:D-TyrAMP	100	-720.24	-36.60	10.10	-53.98
TyrRS:L-TyrAMP	352	-817.77	-43.46	9.94	-58.05

Ce mutant semble très intéressant. Il augmente d'une part l'affinité moyenne pour la D-Tyr et diminue d'autre part celle pour la L-Tyr. Malheureusement, il diminue les affinités moyennes pour les ligands adénylés par rapport à la native. En revanche, si nous observons les affinités maximales, elles sont supérieures à la native pour tous les ligands (table 6.6). Sur la figure 6.7, nous voyons que pour le complexe TyrRS:L-Tyr, la Lys81 peut adopter deux conformations. Dans un cas, elle stabilise le carboxylate de la tyrosine. Dans l'autre cas, elle interagit avec le Glu179. Pour les complexes avec la D-Tyr, on constate que cette lysine ne présente qu'une orientation possible, celle qui stabilise le carboxylate. Pour le complexe TyrRS:L-TyrAMP, l'extrémité chargée de la chaîne latérale de la lysine peut stabiliser le groupement C=O du ligand ou le phosphate. Dans le cas du complexe TyrRS:D-TyrAMP, il n'y a qu'une seule conformation pour cette lysine, celle qui stabilise le phosphate.

FIGURE 6.7 – Mutant KYEQ. Superposition des 5 structures de meilleures affinités.



#### **6.2.3.4 Mutant RYQQ**

TABLE 6.7 – Mutant RYQQ. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Ener dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-Tyr	379	-1873.55	-7.67	1.94	-15.18
TyrRS:L-Tyr	473	-1872.95	-11.34	2.16	-16.49
TyrRS:D-TyrAMP	162	-853.78	-34.85	7.23	-45.98
TyrRS:L-TyrAMP	353	-855.93	-39.26	7.32	-54.17

La séquence RYQQ a été proposée par les expérimentalistes sur observation de la structure cristallographique. L'étude des affinités moyennes pour le ligand RYQQ apparaît singulière. Ce mutant diminuerait l'affinité des quatres complexes considérés, mais augmenterait le ratio des affinités D-Tyr/L-Tyr ou D-TyrAMP/L-TyrAMP. Ce mutant donne cependant une affinité maximale pour le complexe TyrRS:D-Tyr supérieure à la meilleure affinité de la native. Dans la figure 6.8, nous voyons que pour le complexe TyrRS:L-Tyr, la chaîne latérale de l'arginine 81 peut être proche de celle du ligand ou s'en éloigner. Pour le complexe TyrRS:D-Tyr, la position des chaînes latérales est plus variable. Dans ces deux cas, la Tyr175 s'est écartée du ligand par rapport à la structure native, pour libérer l'espace nécessaire à l'arginine. Pour le complexe TyrRS:L-TyrAMP, L' Arg81 stabilise le groupement phosphate. Pour le complexe TyrRS:D-TyrAMP, la chaîne latérale du ligand adopte une conformation différente de celle du ligand cristallographique. Ceci serait dû à l'encombrement stérique créé par la chaîne latérale de l'Arg81, car ici, la Tyr175 est dans une conformation proche de la structure cristallographique. On note un placement couplé des chaînes latérales de l'Arg81 et de la Gln179. On peut avoir un rapprochement de l'atome OE1 de la Gln179 avec le groupement guanidine de l'arginine 81. Si l'atome OE1 est orienté différemment, cette interaction n'a pas lieu et la chaîne latérale entière de la Gln179 est déplacée.

FIGURE 6.8 – Mutant RYQQ. Superposition des 5 structures de meilleures affinités.

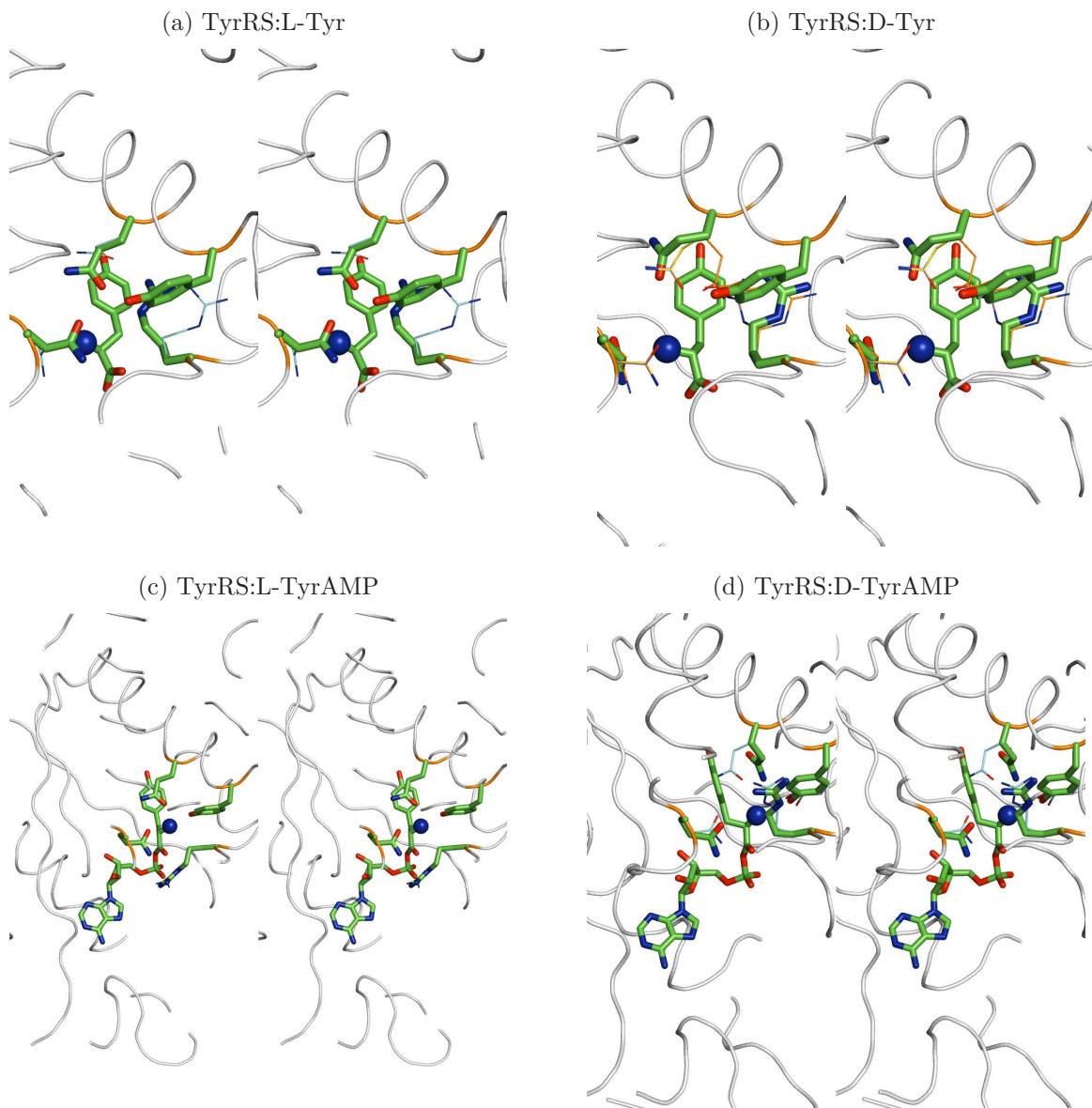
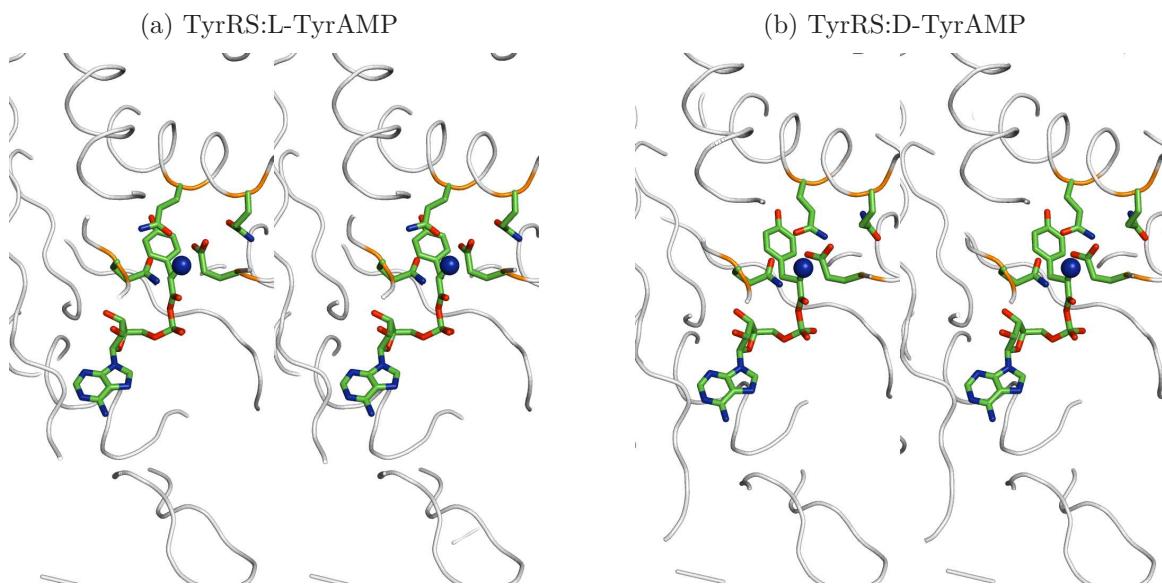


FIGURE 6.9 – Mutant EQQQ. Superposition des 5 structures de meilleures affinités.



## 6.2.4 Séquences prédictes avec L-TyrAMP ou D-TyrAMP

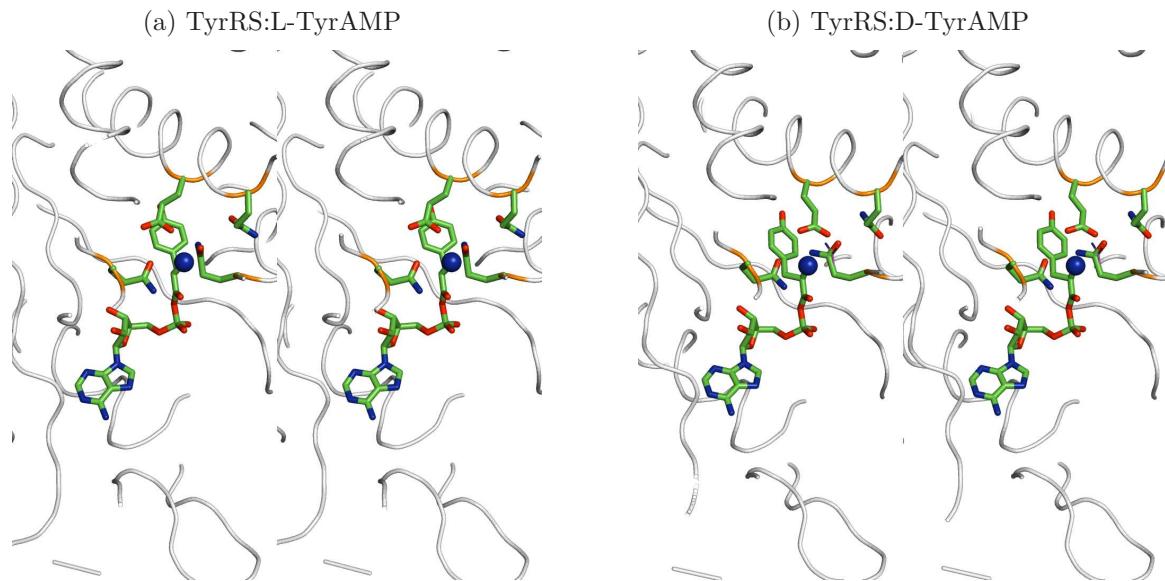
### 6.2.4.1 Mutant EQQQ

TABLE 6.8 – Mutant EQQQ. Ré-optimisation des structures. Energies en  $\text{kcal} \cdot \text{mol}^{-1}$

Complexe	nombre de structures	Ener dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	60	-647.87	-46.79	3.32	-55.06
TyrRS:L-TyrAMP	116	-718.87	-50.92	2.67	-55.77

La séquence EQQQ a été prédite en liaison avec le D-TyrAMP. Ce mutant améliore l'affinité moyenne pour le ligand D-TyrAMP et diminue celle pour le ligand L-TyrAMP ; il en va de même pour les affinités maximales. Ceci en fait un bon candidat pour notre étude. Pour ces deux complexes, les chaînes latérales du Glu81 et de la Gln201 sont placées de manière identique (figure 6.9). La chaîne latérale de la Gln175 s'écarte du site actif également dans les deux cas. Un fait remarquable est le positionnement des atomes NE2 et OE1 du groupement amide de la Gln179. Ils s'inversent selon que ligand est L ou D, comme pour la TyrRS native.

FIGURE 6.10 – Mutant QQEQ. Superposition des 5 structures de meilleures affinités.



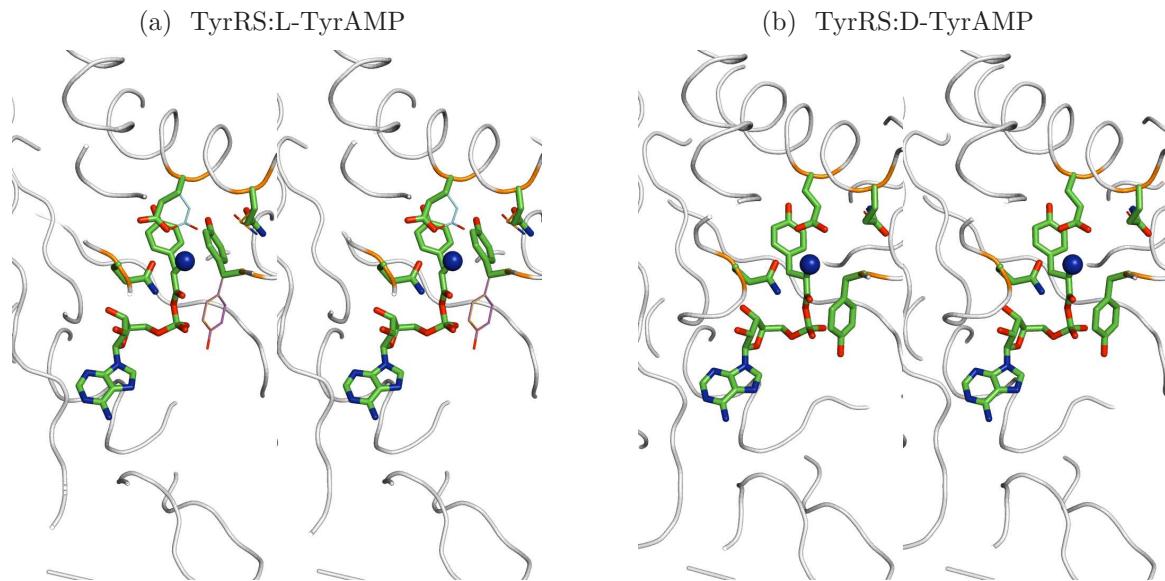
#### 6.2.4.2 Mutant QQEQ

 TABLE 6.9 – Mutant QQEQ. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Ener dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	54	-648.41	-47.17	4.57	-55.24
TyrRS:L-TyrAMP	125	-742.03	-48.76	2.00	-53.37

La séquence QQEQ a été prédite à la fois en présence du ligand L-TyrAMP et du D-TyrAMP. Ce mutant améliore l'affinité moyenne pour le ligand D-TyrAMP et diminue celle pour le ligand L-TyrAMP ; de même pour les affinités maximales. C'est également un bon candidat pour notre étude. Les chaînes latérales du Glu179 et de la Gln201 sont à la même position indifféremment avec la forme L ou D du ligand (figure 6.10). La chaîne latérale de la Gln175 est repoussée du site actif, comme pour le mutant précédent. On observe pour la Gln81 la même inversion de positionnement des atomes OE1 et NE2.

FIGURE 6.11 – Mutant YQEQQ. Superposition des 5 structures de meilleures affinités.



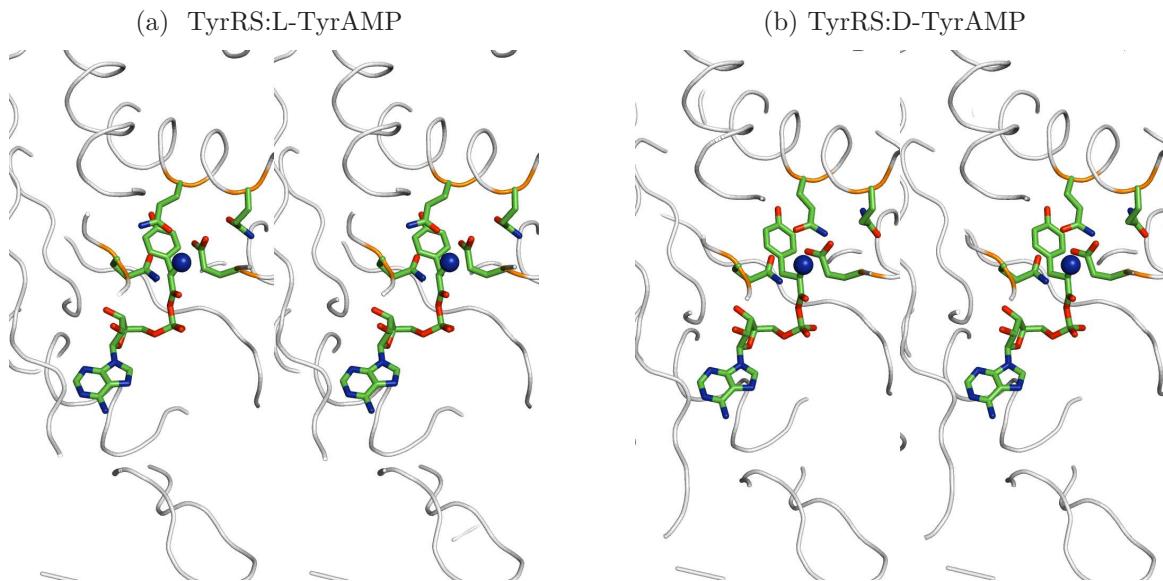
#### 6.2.4.3 Mutant YQEQQ

TABLE 6.10 – Mutant YQEQQ. Ré-optimisation des structures. Energies en  $\text{kcal} \cdot \text{mol}^{-1}$

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	93	-623.85	-36.15	14.69	-52.82
TyrRS:L-TyrAMP	90	-696.97	-43.44	12.42	-57.07

La séquence YQEQQ a été prédite à la fois en présence du ligand L-TyrAMP et D-TyrAMP. Ici, les affinités moyennes sont inférieures à celle de la native pour les deux ligands. On arrive cependant à prédire des structures plus affines pour les deux ligands. Dans la figure 6.11, pour le complexe TyrRS:L-TyrAMP, la Tyr81 peut avoir une orientation comparable à celle de la Tyr175 native, en pointant son groupement hydroxyle vers le ligand. Elle peut s'orienter également vers le bas, alors la chaîne latérale du Glu179 se déplace. Dans le cas du complexe TyrRS:L-TyrAMP, la Tyr81 est orientée vers le bas, la chaîne latérale 179 est déplacée dans le site actif, et ce glutamate interagit plus favorablement avec la D-Tyr. La Gln175 est éloignée du site actif dans les deux cas, mais on observe cependant l'inversion de positionnement des atomes OE1 et NE2 selon la forme L ou D du ligand. Ce mutant semble défavoriser la liaison au L-TyrAMP, plus qu'il ne favorise réellement le D-TyrAMP.

FIGURE 6.12 – Mutant DQEQQ. Superposition des 5 structures de meilleures affinités.



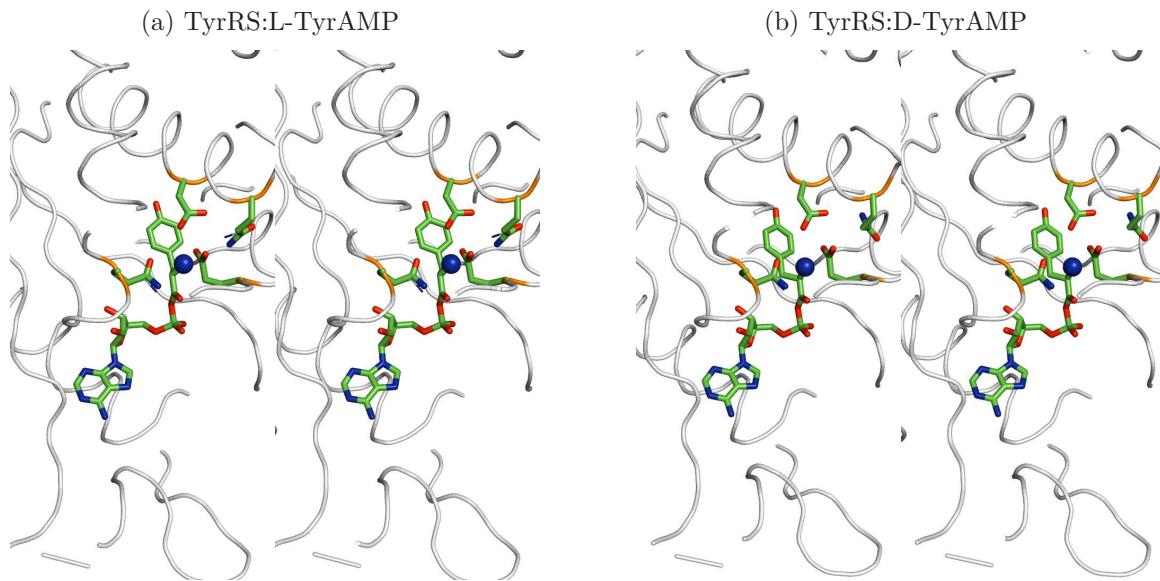
#### 6.2.4.4 Mutant DQEQQ

 TABLE 6.11 – Mutant DQEQQ. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	73	-641.73	-54.88	2.31	-59.39
TyrRS:L-TyrAMP	59	-661.01	-58.61	1.14	-61.60

La séquence DQEQQ a été prédite en liaison avec le D-TyrAMP. Les affinités moyennes et maximales pour les ligands sont bien supérieures à celle de la native, avec toujours plus d'affinité pour le ligand L-TyrAMP. Nous voyons sur la figure 6.12 que pour ce mutant, les chaînes latérales semblent au premier abord positionnées identiquement pour les deux ligands. On remarque pourtant que les atomes OE1 et NE2 des glutamines 175 et 201 sont inversés.

FIGURE 6.13 – Mutant EQDQ. Superposition des 5 structures de meilleures affinités.



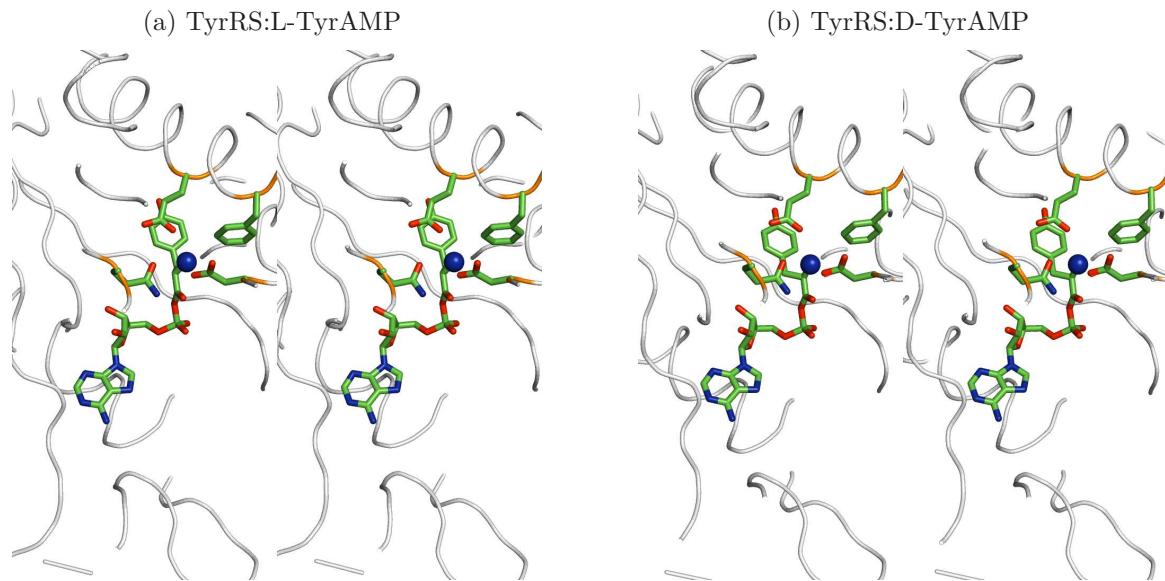
#### 6.2.4.5 Mutant EQDQ

TABLE 6.12 – Mutant EQDQ. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	53	-651.13	-46.97	5.38	-55.35
TyrRS:L-TyrAMP	58	-663.23	-57.61	2.07	-61.16

La séquence EQDQ a été prédite en liaison avec le D-TyrAMP. L'affinité moyenne et l'affinité maximale sont améliorées par rapport à la native pour les deux ligands. Au regard de la figure 6.13, nous voyons que pour ce mutant, les acides aminés aux quatre positions sont placés identiquement. Les deux acides aminés chargés interagissent directement avec l'ammonium du ligand L ou D.

FIGURE 6.14 – Mutant DFEQ. Superposition des 5 structures de meilleures affinités.



### 6.2.5 Séquences d'affinité moindre

 TABLE 6.13 – Mutant DFEQ. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	62	-647.52	-47.80	3.12	-53.43
TyrRS:L-TyrAMP	112	-746.48	-56.88	1.51	-60.17

**Mutant DFEQ** La séquence DFEQ a été prédite en liaison avec le D-TyrAMP. L'affinité moyenne et l'affinité maximale sont améliorées par rapport à la native pour les deux ligands. Sur la figure 6.14, nous voyons que pour ce mutant également, les acides aminés aux quatre positions sont positionnés identiquement. Les deux acides aminés chargés interagissent directement avec l'ammonium du ligand.

FIGURE 6.15 – Mutant EQVQ. Superposition des 5 structures de meilleures affinités.

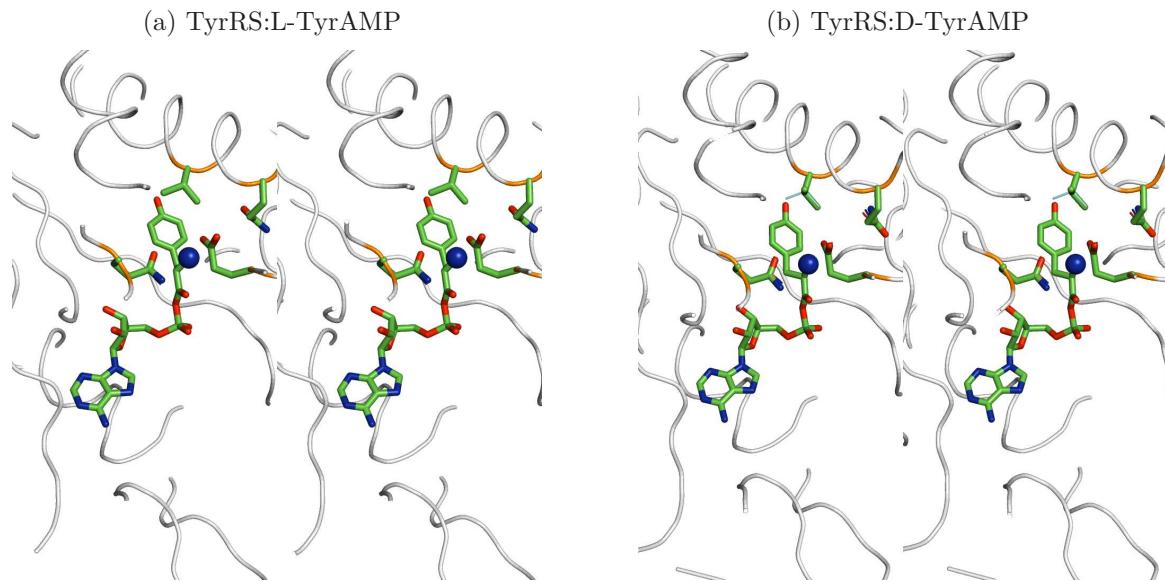


TABLE 6.14 – Mutant EQVQ. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	70	-645.36	-46.02	1.89	-55.09
TyrRS:L-TyrAMP	62	-661.51	-50.62	0.98	-53.30

**Mutant EQVQ** La séquence EQVQ a été prédite en liaison avec le D-TyrAMP. L'affinité moyenne et l'affinité maximale sont améliorées par rapport à la native pour les deux ligands. Pour ce mutant, la Val179, de par sa taille, permet un meilleur positionnement du Glu81, ce dernier interagit plus favorablement avec le D-TyrAMP qu'avec le L-TyrAMP. La Gln175 peut voir ses atomes OE1 et NE2 inversés selon le ligand (figure 6.15).

## 6.2. ANALYSE GÉNÉRALE DES SÉQUENCES OBTENUES

FIGURE 6.16 – Mutant NDQQ. Superposition des 5 structures de meilleures affinités.

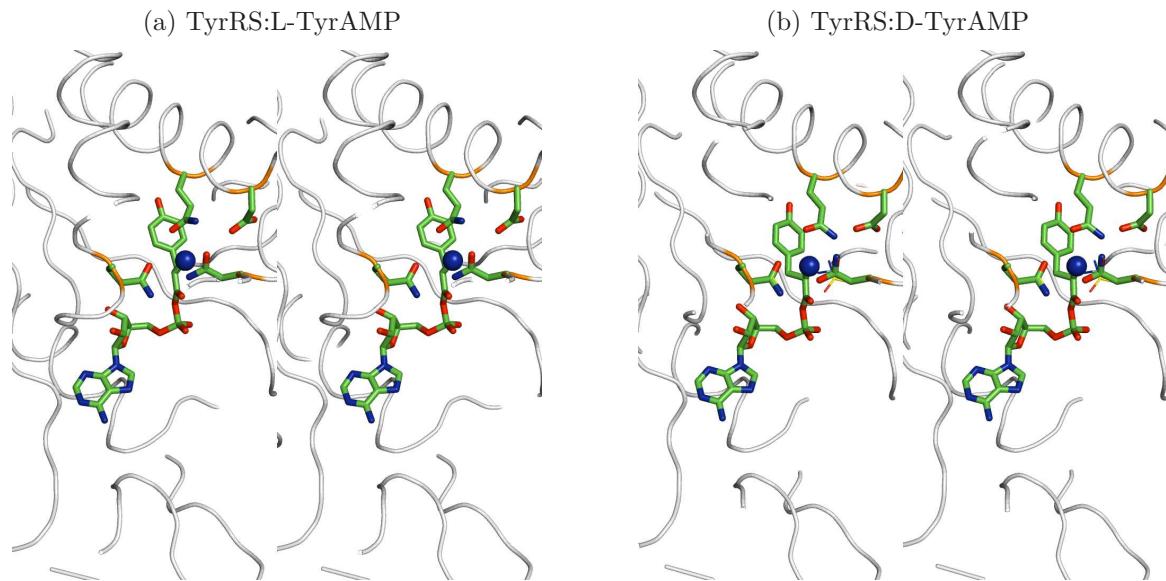


TABLE 6.15 – Mutant NDQQ. Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	28	-650.92	-43.53	4.64	-55.11
TyrRS:L-TyrAMP	123	-662.42	-47.79	1.62	-52.86

**Mutant NDQQ** La séquence NDQQ a été prédite à la fois en présence du ligand L-TyrAMP et D-TyrAMP. L'affinité moyenne est améliorée par rapport à la native pour les deux ligands. Mais lorsqu'on regarde les affinités maximales, seul le complexe TyrRS:D-TyrAMP a été amélioré par rapport à la native. Sur la figure 6.17, l'Asn81 peut voir ses atomes OD1 et ND2 inversé ou pas selon le ligand. Les trois autres acides aminés sont dans la même conformation quel que soit le ligand. Cet ensemble d'acides aminés chargés ou polaires encadre l'ammonium du ligand.

FIGURE 6.17 – Mutant NQEQQ. Superposition des 5 structures de meilleures affinités.

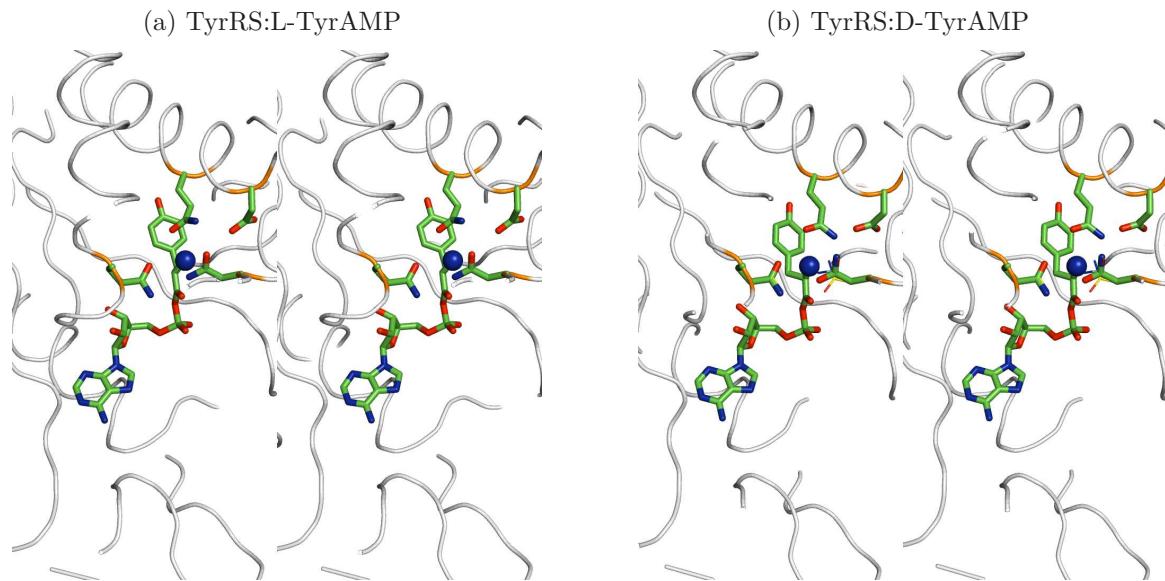


TABLE 6.16 – Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	65	-642.95	-50.80	2.19	-58.21
TyrRS:L-TyrAMP	96	-660.12	-54.19	1.67	-58.43

**Mutant NQEQQ** La séquence NQEQQ a été prédictée à la fois en présence du ligand L-TyrAMP et D-TyrAMP. L'affinité moyenne et l'affinité maximale sont améliorées par rapport à la native pour les deux ligands. Les affinités maximales pour les deux ligands sont semblables. La figure 6.17 nous indique que l'Asn81 peut voir ses atomes OD1 et ND2 inversés selon que le ligand soit L ou D. Les trois autres acides aminés sont dans la même conformation quel que soit le ligand.

## 6.2. ANALYSE GÉNÉRALE DES SÉQUENCES OBTENUES

FIGURE 6.18 – Mutant DFDQ. Superposition des 5 structures de meilleures affinités.

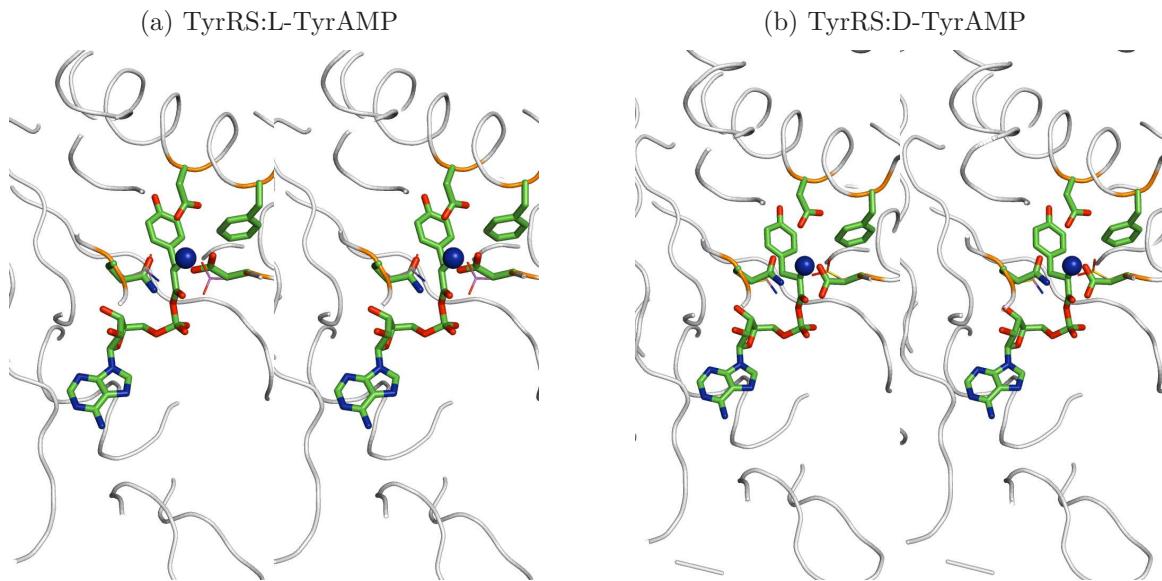


TABLE 6.17 – Mutant DFDQ. Ré-optimisation des structures. Energies en  $\text{kcal} \cdot \text{mol}^{-1}$

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	48	-647.43	-48.08	5.60	-56.98
TyrRS:L-TyrAMP	70	-661.99	-58.92	0.89	-61.69

**Mutant DFDQ** La séquence DFDQ a été prédictée en liaison avec le D-TyrAMP. L'affinité moyenne et l'affinité maximale sont améliorées par rapport à la native pour les deux ligands. La figure 6.18 n'est pas très informative car les conformations des chaînes latérales mutées sont semblables quel que soit le ligand.

FIGURE 6.19 – Mutant YFEQ. Superposition des 5 structures de meilleures affinités.

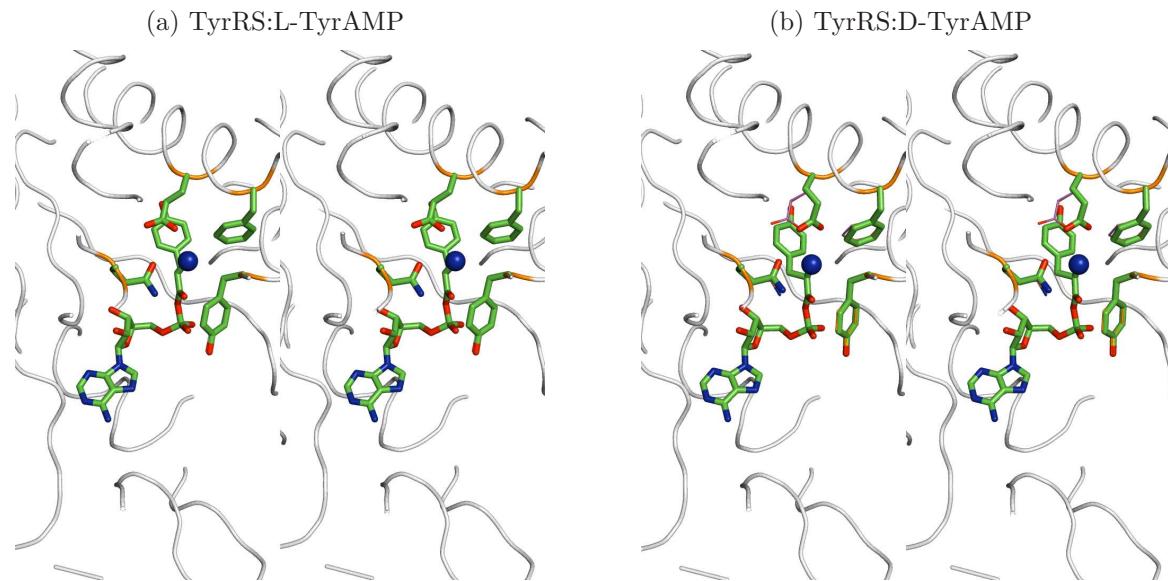


TABLE 6.18 – Ré-optimisation des structures. Energies en kcal.mol<sup>-1</sup>

Complexe	nombre de structures	Energie de dépliement moyenne	affinité moyenne	écart type	Meilleure affinité
TyrRS:D-TyrAMP	65	-624.43	-34.37	14.65	-53.07
TyrRS:L-TyrAMP	149	-789.31	-45.33	12.98	-57.21

**Mutant YFEQ** La séquence YFEQ a été prédictée à la fois en présence du ligand L-TyrAMP et D-TyrAMP. L'affinité moyenne est diminuée par rapport à la native pour les deux ligands. Cependant, les affinités maximales sont supérieures à celle obtenue pour la TyrRS native. Sur la figure 6.19, on retrouve une orientation vers le bas de la Tyr81, comme chez le mutant YQEQQ. Ici, la Phe175 empêche le positionnement vers l'ammonium de la Tyr81. Les acides aminés mutés sont disposés de la même façon quel que soit le ligand. Cependant, le Glu179, s'oriente de manière à interagir avec l'ammonium du ligand.

## **6.3 Etude par dynamique moléculaire et confrontation aux résultats expérimentaux**

A l'issue de l'analyse des résultats d'évolution dirigée, nous voyons que la plupart des mutants prédis permettent d'augmenter la fixation de la D-Tyr ou du D-TyrAMP, sans pour autant inverser le ratio de fixation D/L. Les seuls mutants qui semblent y parvenir sont les mutants : EQQQ, EQDQ et NDQQ. Une propriété structurale est retrouvée dans plusieurs mutants, les glutamines ont tendance à inverser les positions de leurs atomes OE1 et ND2 selon que le ligand soit L ou D. L'atome OE1 s'orientera ainsi au plus près de l'ammonium du ligand. Étudions maintenant plus en détail toutes ces séquences qui paraissent déjà lier la D-Tyr un peu mieux que la native. Nous étudierons l'évolution des structures au cours de simulations de dynamique moléculaire. De plus, pour certaines de ces séquences, nous décrirons des mesures expérimentales d'activité. En effet, au sein du laboratoire, des travaux expérimentaux de mutagénèse combinatoire par oligonucléotides dégénérés ont été menés par Pierre Plateau et Sandra Wydau sur la TyrRS, pour favoriser la liaison de la D-Tyr. Cette mutagénèse aléatoire concerne également les quatre résidus Asp81, Tyr175, Gln179 et Gln201. Les biochimistes ont conduit des expériences d'aminoacylation avec du pyrophosphate inorganique radiomarqué décrites plus loin. Ils ont également mesurés des constantes d'inhibition avec radiomarquage de la tyrosine puis introduction de L-Tyr ou de D-Tyr froide. Ils ont ainsi obtenu une information biochimique sur l'efficacité des séquences prédictives pour lier la L-Tyr et la D-Tyr. Une autre expérience a été réalisée par Sandra Wydau et Pierre Plateau, l'analyse cinétique de la réaction d'échange de phosphate inorganique. Le milieu contient initialement de la tyrosine radiomarquée, puis la réaction est inhibée avec de la L-Tyr ou de la D-Tyr froide. On mesure alors la vitesse Vmax.

Dans ce qui suit, nous revenons donc sur les séquences les plus intéressantes, pour les étudier par dynamique moléculaire et par calculs d'affinité Poisson-Boltzmann et pour donner, quand ils sont disponibles, les résultats expérimentaux.

### 6.3.1 Etude de la TyrRS native

TABLE 6.19 – Affinité protéine-ligand en kcal.mol<sup>-1</sup>). TyrRS native.

Complexe	CASA	PBFE			$\Delta\Delta G$
		3 ns	5 ns		
TyrRS:D-Tyr	-7.9	-4.3	-4.6	+0.5	
TyrRS:L-Tyr	-14.71	-5.0	-5.1		
TyrRS:D-TyrAMP	-44.32	-6.8	-6.6	+ 0.5	
TyrRS:L-TyrAMP	-52.85	-6.7	-7.1		
TyrRS(cristallo):L-TyrAMP	-	-7.2	-7.2		

TABLE 6.20 – Dynamique moléculaire de la TyrRS native. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

\* : Simulations initiées à partir de la structure cristallographique. Les autres sont initialisées à partir de la structure optimisée par Proteus.

Complexe	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
TyrRS:D-Tyr	0.7	1.1	0.9	0.7	1.1	1.0
TyrRS:L-Tyr	0.7	1.2	0.7	0.7	1.2	0.7
TyrRS:D-TyrAMP	0.6	1.0	0.8	0.6	1.1	0.8
TyrRS:L-TyrAMP	0.7	1.0	1.0	0.7	1.0	1.0
*TyrRS:L-TyrAMP	0.7	1.0	1.1	0.7	1.1	0.8
*TyrRS Apoenzyme	0.7	1.3	-	0.7	1.1	-

\* : Simulations initiées à partir de la structure cristallographique.

Les autres sont initialisées à partir de la structure optimisée par Proteus.

L’analyse du rmsd par rapport à la structure initiale (table 6.20) montre que la structure ne se dénature pas au cours de la dynamique, et qu’elle reste proche de la structure initiale. On calcule le rms des coordonnées par rapport à la structure cristallographique sur une sphère de 15  $\text{\AA}$  autour du ligand, en moyenne, sur les 100 dernières ps, après 3 ou 5 ns de dynamique. Les déviations sont de 0.6 à 0.7  $\text{\AA}$  pour la chaîne principale et de 0.8 à 1  $\text{\AA}$  pour le ligand, quelque soit le ligand. Une étude exhaustive des distances inter atomiques permet de savoir quels atomes interagissent avec l’ammonium du ligand (table 6.21). Quelques résidus interagissent directement avec l’ammonium du ligand. L’Asp41 lie l’ammonium, cette position pourrait sembler intéressante à muter, mais elle est en fait impliquée dans la coordination du magnésium nécessaire à l’aminoacylation de l’ARNt. La structure finale du complexe avec le L-TyrAMP est montré dans la figure 6.20. On voit

### **6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX**

---

que la structure évolue peu au cours de la dynamique. Cependant, le cycle de la Tyr175, ainsi que la base adénosine, apparaissent un peu plus mobiles que le reste de la structure.

Une deuxième simulation du complexe TyrRS:L-TyrAMP a été menée en prenant comme structure initiale pour la dynamique celle de meilleure affinité générée par Proteus. Après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est légèrement plus grand que dans la simulation précédente. On voit donc que la structure reste encore stable au cours de la dynamique. Plusieurs résidus interagissent directement avec l'ammonium (table 6.21) ; sans surprise, il s'agit des résidus 81, 175, 179 et 201.

Nous avons également simulé le complexe TyrRS:D-TyrAMP. Après 5 ns de dynamique, le rms moyen obtenu par rapport à la structure cristallographique sur une sphère de 15 Å autour du ligand est de 0.6 Å pour le squelette protéique et de 1.1 Å pour les chaînes latérales. Les atomes OD1 et OD2 de l'aspartate 81 stabilisent le groupement ammonium du TyrAMP par liaison hydrogène. La liaison hydrogène entre l'hydroxyle de la tyrosine 175 et l'ammonium du TyrAMP est également un élément stabilisant, ainsi que l'atome OE1 de la glutamine 179. La glutamine 201 a une action stabilisante pour le TyrAMP, en effet, l'atome OE1 est très proche du groupement ammonium. D'autres interactions sont également importantes. L'arginine 204 interagit à la fois avec la glutamine 179, la glutamine 201, et l'aspartate 81. L'aspartate 200 interagit avec la glutamine 201, de manière à fixer sa conformation et favoriser ainsi son interaction avec l'ammonium du ligand. On constate (table 6.21) que les mêmes atomes des mêmes résidus sont en contact avec l'ammonium des ligand L-TyrAMP et D-TyrAMP. On voit (figure 6.21) que le squelette protéique et les chaînes latérales des résidus mutés ont peu évolué au cours de la dynamique moléculaire. En revanche, la partie tyrosine du ligand s'est déplacée dans le site actif.

A partir des structures obtenues par dynamique, nous estimons l'affinité protéine-ligand avec un traitement de Poisson-Boltzmann, dite méthode PBFE (Poisson Boltzmann Free Energy). On considère la protéine et le ligand comme des corps diélectriques, avec une constante diélectrique  $\epsilon$  qui sera généralement prise égale à 4. Ici, avec  $\epsilon = 4$ , en

moyennant sur 5 ns de dynamique, la différence d'énergie libre d'association  $\Delta\Delta G$  entre L- et D-Tyr est égale à 0.5 kcal.mol<sup>-1</sup> en faveur de la L-Tyr (table 6.19). Ce résultat est proche de la valeur expérimentale de 1.9 kcal.mol<sup>-1</sup>. On retrouve la même valeur quand on étudie les ligands adénylés, ce qui est également en accord avec l'expérience, où la différence d'énergie libre est de 1.03 kcal.mol<sup>-1</sup>.

TABLE 6.21 – Distance des atomes avec le groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du		
		L-TyrAMP cristallographique (Å)	L-TyrAMP (Å)	D-TyrAMP (Å)
OD2	ASP 41	2.8	4.7	4.1
OD1	ASP 81	2.9	4.7	4.7
OD2	ASP 81	4.6	2.6	2.7
OH	TYR 175	2.9	2.7	2.9
OE1	GLN 179	2.9	3.1	3.3
OE1	GLN 201	3.8	2.8	2.6

FIGURE 6.20 – TyrRS native:L-TyrAMP. Superposition de la structure cristallographique (en vert) et de la structure après 5 ns de dynamique moléculaire (en rouge).

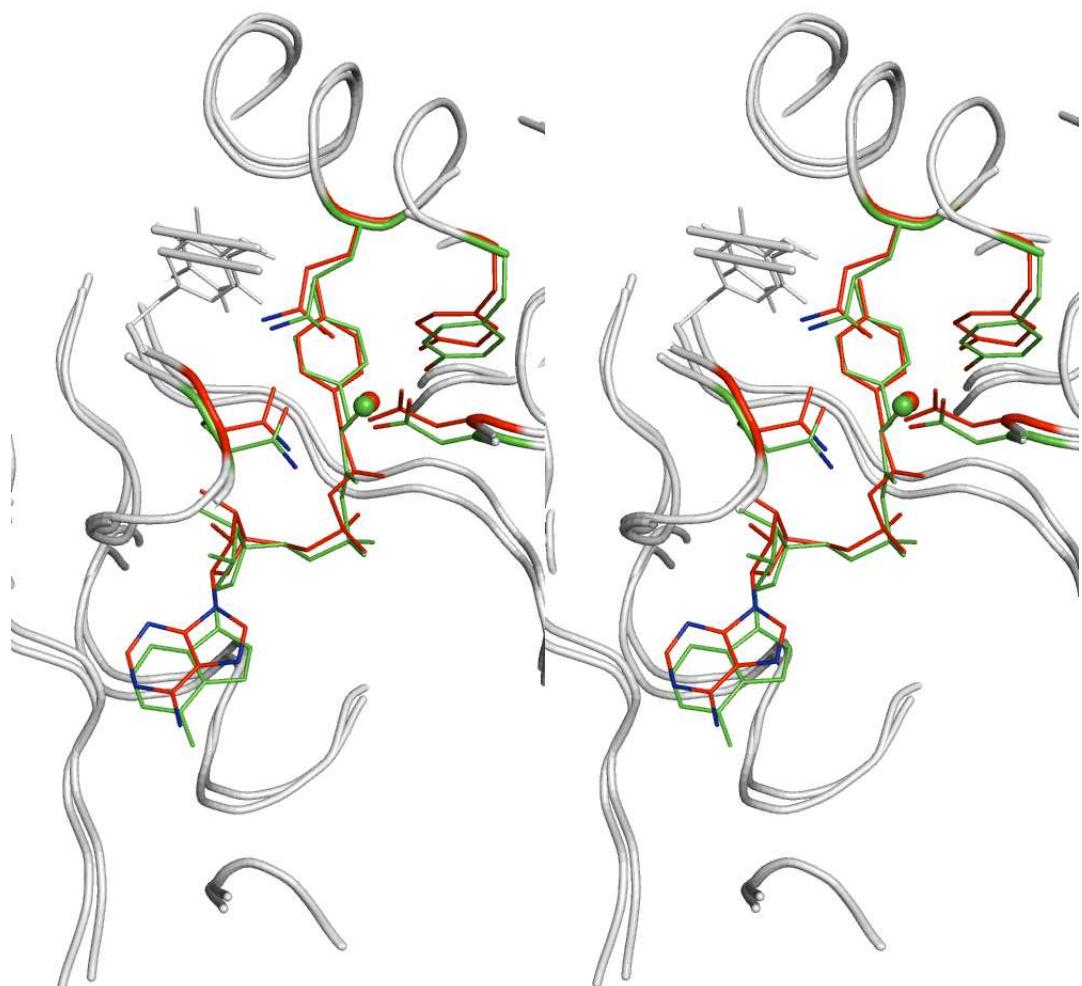
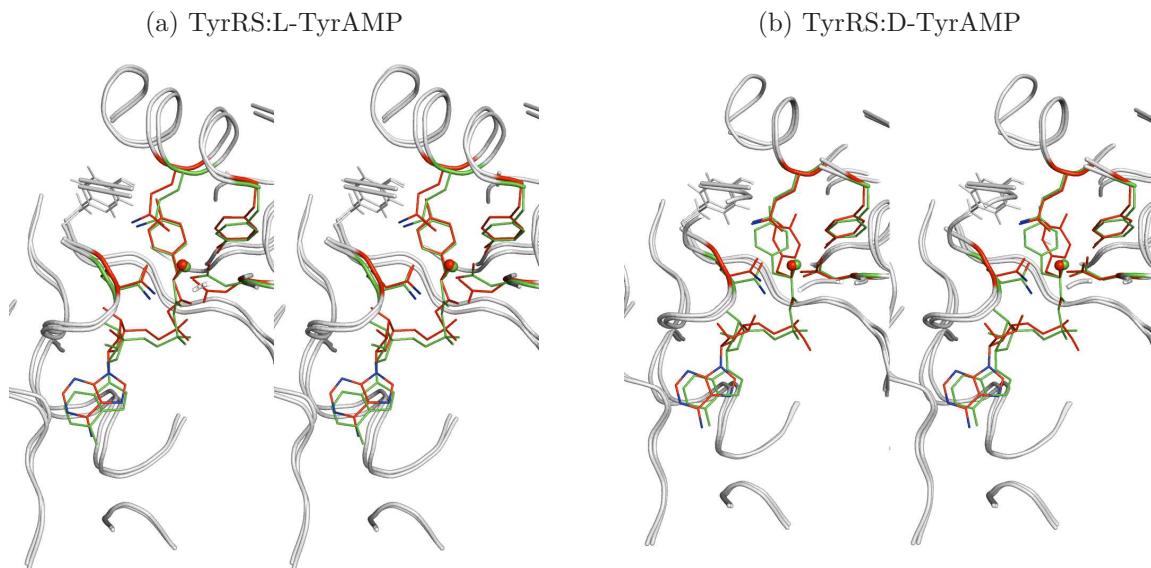


FIGURE 6.21 – TyrRS native. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge).



### 6.3.2 Séquences prédites avec le ligand D-Tyr

#### 6.3.2.1 Mutant HYED

TABLE 6.22 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant HYED.

Complexe	CASA	PBFE		
		3 ns	5 ns	$\Delta\Delta G$
TyrRS:D-Tyr	-8.68	-6.4	-7.3	-7.8 sur 5ns et -6.9 sur 3 ns
TyrRS:L-Tyr	-18.55	0.5	0.5	
TyrRS:D-TyrAMP	-40.92	-4.7	-4.1	6.61 sur 3 ns
TyrRS:L-TyrAMP	-47.68	-11.31	-	

TABLE 6.23 – Dynamique moléculaire du mutant HYED. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

Complexe	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
TyrRS:D-Tyr	0.7	1.2	0.7	0.8	1.2	0.5
TyrRS:L-Tyr	0.6	1.0	1.0	0.73	1.1	1.0
TyrRS:D-TyrAMP	0.7	1.1	0.7	0.6	1.1	0.9
TyrRS:L-TyrAMP						calcul en cours

Avec CASA, ce mutant n'était pas prédict pour lier préférentiellement la D-Tyr (table 6.28). En revanche, selon les calculs PBFE, ce mutant aurait une forte préférence pour la

### *6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX*

---

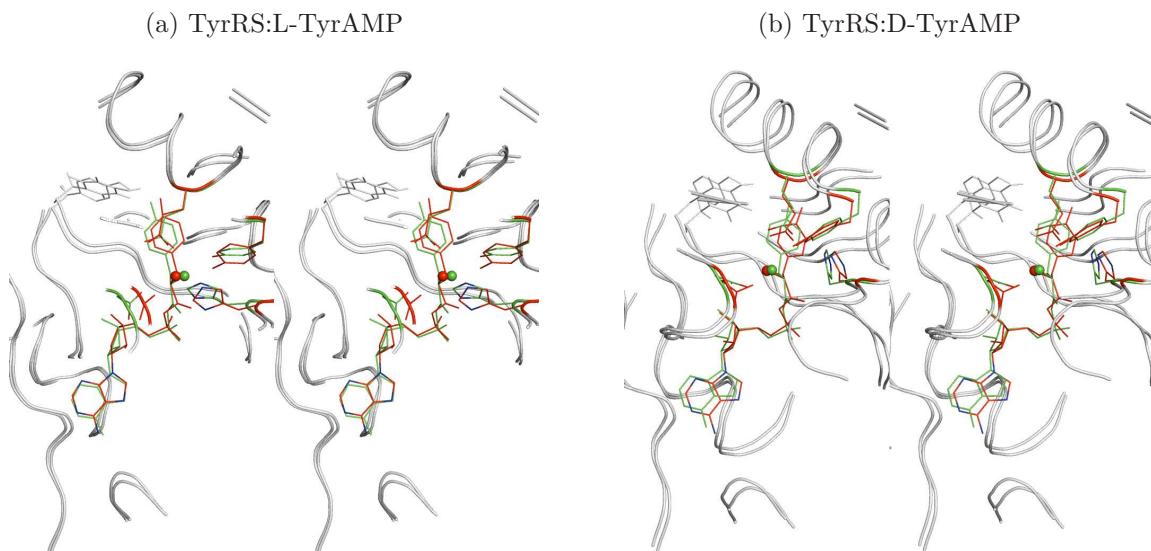
D-Tyr, avec  $\Delta\Delta G = -7.8 \text{ kcal.mol}^{-1}$ . Nous verrons que les expériences (cf. sous chapitre 6.3.3) montrent une légère amélioration pour la liaison de la D-Tyr. Lorsque le ligand est adénylé,  $\Delta\Delta G = 6.61 \text{ kcal.mol}^{-1}$  sur 3ns de dynamique. Le mutant HYED favoriserait la liaison de la D-Tyr mais pas celle du D-TyrAMP.

Les résidus 179 et 201 interagissent directement avec l'ammonium du ligand L-TyrAMP (table 6.24). Pour le complexe TyrRS:D-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure cristallographique est inférieur à 1.1 Å (table 6.24). Les résidus 179 et 201 interagissent également avec l'ammonium du D-TyrAMP, le résidu 204 interagit spécifiquement avec l'ammonium du D-TyrAMP. C'est donc cette dernière interaction qui favoriserait spécifiquement le résidu D-TyrAMP (figure 6.22).

TABLE 6.24 – Distances des atomes au groupement ammonium après 5 ns de dynamique pour le complexe TyrRS:D-TyrAMP et après 2.8 ns pour le complexe TyrRS:L-TyrAMP

Nom de l'atome	Acide aminé	Distance à l'azote du	
		L-TyrAMP (Å)	D-TyrAMP (Å)
OE1	GLU 179	2.5	2.8
OE2	GLU 179	3.9	2.8
OD1	ASP 201	2.6	4.0
OD2	ASP 201	3.8	2.9
HD22	ASN 204	3.6	3.4

FIGURE 6.22 – Mutant HYED. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge) pour le complexe TyrRS:D-TyrAMP. Superposition de la structure de meilleure affinité (en vert), et de la structure après 2.8 ns de dynamique moléculaire (en rouge) pour le complexe TyrRS:L-TyrAMP .



### 6.3.2.2 Mutant KYED

TABLE 6.25 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant KYED.

Complexe	CASA	PBFE		
		3 ns	5 ns	$\Delta\Delta G$
TyrRS:D-Tyr	-6.98	-6.2	-6.8	-9.2
TyrRS:L-Tyr	-20.10	2.8	2.4	
TyrRS:D-TyrAMP	-28.53	-4.9	-4.6	7.1
TyrRS:L-TyrAMP	-39.48	-12.0	-11.7	

TABLE 6.26 – Dynamique moléculaire du mutant KYED. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

Complexe	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
TyrRS:D-Tyr	0.8	1.1	0.7	0.8	1.3	0.5
TyrRS:L-Tyr	0.8	1.2	1.1	0.8	1.3	1.3
TyrRS:D-TyrAMP	0.7	1.0	1.0	0.7	1.0	1.0
TyrRS:L-TyrAMP	0.6	0.9	0.6	0.6	0.9	0.7

De 1 à 5 ns de dynamique moléculaire,  $\Delta\Delta G = -9.2 \text{ kcal.mol}^{-1}$  pour la différence D-Tyr/L-Tyr. Ce mutant lierait donc mieux la D-Tyr que la L-Tyr. Mais ceci n'est plus

### 6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX

---

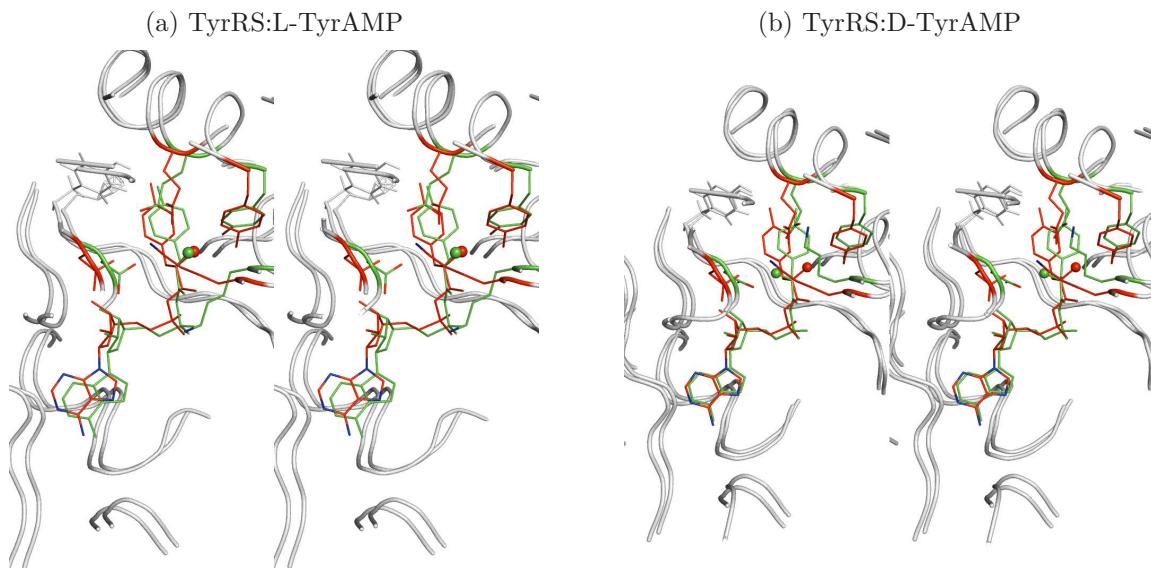
vrai lorsque le ligand est adénylé :  $\Delta\Delta G = 7.1 \text{ kcal.mol}^{-1}$ .

Pour le complexe TyrRS:L-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.3 Å (table 6.26). Les résidus 175, 179 et 201 interagissent avec le ligand L-TyrAMP. Au cours de la dynamique, il y a un réajustement de plusieurs chaînes latérales. Tout d'abord celle du ligand mais également celle de la tyrosine 175. La lysine 81 se rapproche du groupement phosphate du ligand (figure 6.23). Pour le complexe TyrRS:D-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.3 Å (table 6.26). Les résidus 175 et 179 stabilisent les deux ligands, mais les résidus 41 et 76 stabilisent spécifiquement le D-TyrAMP pour ce mutant (table 6.27). Ici, la conformation de la partie tyrosine du ligand est très différente après la dynamique. Le positionnement de la lysine 81 est très différent également par rapport à son positionnement avec le ligand L-TyrAMP. En effet, ici, la chaîne latérale de la lysine 81 se rapproche de l'ammonium du ligand et repousse la chaîne latérale de l'aspartate 201. C'est pour celà que ce dernier résidu interagit avec le L-TyrAMP mais pas avec le D-TyrAMP (figure 6.23).

TABLE 6.27 – Distances des atomes au groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du L-TyrAMP (Å)	Distance à l'azote du D-TyrAMP (Å)
OD1	ASP 41	5.2	2.9
OD2	ASP 41	4.9	2.8
HG1	THR 76	6.1	3.2
OH	TYR 175	2.9	3.1
OE1	GLU 179	3.2	4.7
OE2	GLU 179	2.7	2.8
OD2	ASP 201	2.8	7.6

FIGURE 6.23 – Mutant KYED. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge).



### 6.3.2.3 Mutant KYEQ

TABLE 6.28 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant KYEQ.

Complexe	CASA	PBFE		
		3 ns	5 ns	$\Delta\Delta G$
TyrRS:D-Tyr	-9.46	-6.4	-5.9	1.6
TyrRS:L-Tyr	-13.85	-7.5	-7.5	
TyrRS:D-TyrAMP	-36.60	-9.2	-8.9	5.2
TyrRS:L-TyrAMP	-43.46	-14.8	-14.1	

TABLE 6.29 – Dynamique moléculaire du mutant KYEQ. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

Complexe	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
TyrRS:D-Tyr	0.7	1.2	0.4	0.7	1.2	0.7
TyrRS:L-Tyr	0.7	1.1	0.6	0.7	1.2	1.0
TyrRS:D-TyrAMP	0.7	1.2	0.8	0.7	1.2	0.9
TyrRS:L-TyrAMP	0.6	1.0	0.7	0.6	1.0	0.5

De 1 à 5 ns de dynamique moléculaire, les calculs PBFE donnent  $\Delta\Delta G=1.6 \text{ kcal.mol}^{-1}$  pour la différence D-Tyr/L-Tyr, et  $\Delta\Delta G=5.2 \text{ kcal.mol}^{-1}$  pour la différence D-TyrAMP/L-TyrAMP. Les résultats de dynamique moléculaire et d'évolution dirigée sont en accord.

### **6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX**

---

Ce mutant ne lierait pas mieux les ligands D que le L, mais augmenterait un peu le ratio de fixation en faveur de la D-Tyr.

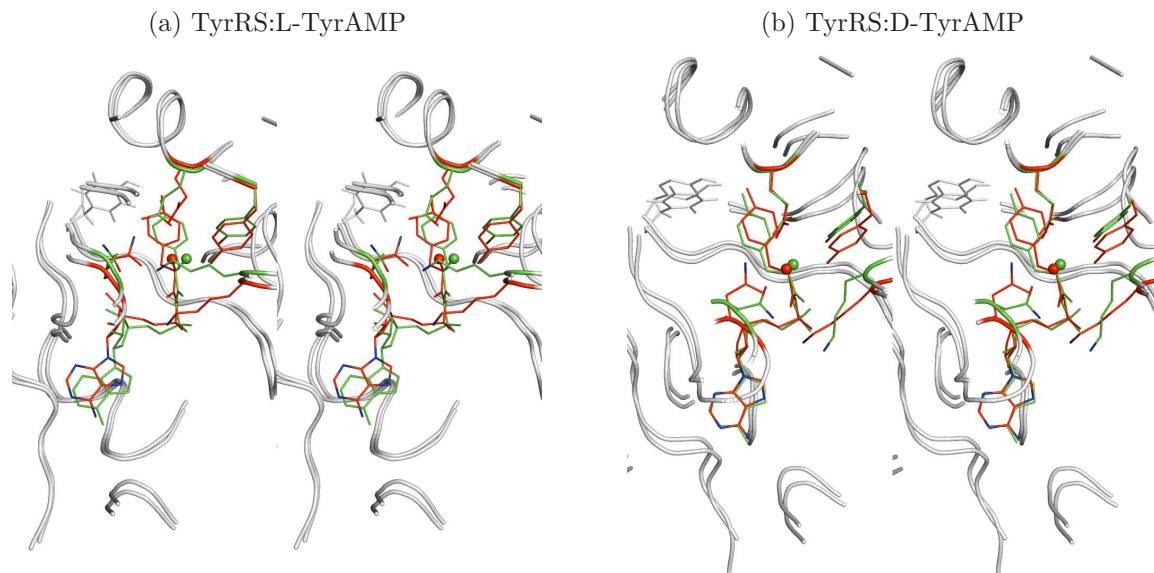
Pour le complexe TyrRS:L-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.2 Å (table 6.30). Les résidus 41, 175 et 179 interagissent avec l'ammonium du ligand L-TyrAMP. On observe également pour ce mutant (figure 6.24), un déplacement de la lysine 81 vers le groupement phosphate du ligand au cours de la dynamique moléculaire ; ceci permet à la glutamine 201 de se rapprocher de l'ammonium, mais pas suffisamment pour créer une liaison hydrogène. La tyrosine 175 rapproche également son groupement hydroxyle de l'ammonium du ligand au cours de la dynamique. La partie tyrosine du ligand s'est également déplacée dans le site actif au cours de la dynamique.

Pour le complexe TyrRS:D-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure cristallographique sur une sphère de 15 Å autour du ligand est inférieur à 1.2 Å (table 6.30). Les résidus 175 et 179 interagissent également avec l'ammonium du D-TyrAMP. La glutamine 201 lie spécifiquement la D-Tyr dans ce mutant. Les structures avant et après dynamique sont semblables d'un point de vue des interactions. Les rotamères des chaînes latérales évoluent peu. Ici (figure 6.24), c'est le squelette protéique qui s'est le plus adapté aux mutations au cours de la dynamique, bien que le rms de la structure après dynamique par rapport à la structure initiale reste faible.

TABLE 6.30 – Distances des atomes au groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du	
		L-TyrAMP (Å)	D-TyrAMP (Å)
OD1	ASP 41	2.8	4.3
OD2	ASP 41	3.2	4.7
OH	TYR 175	2.8	2.8
OE1	GLU 179	2.5	4.5
OE2	GLU 179	4.3	2.6
OE1	GLN 201	4.4	2.7

FIGURE 6.24 – Mutant KYEQ. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge).



#### 6.3.2.4 Mutant RYQQ

La séquence RYQQ a été proposée par les expérimentalistes sur observation de la structure cristallographique. Les résultats des calculs de PBFE entre 1 et 5 ns de dynamique avec  $\varepsilon = 4$  sont :  $\Delta\Delta G = -9.1 \text{ kcal.mol}^{-1}$  pour D-Tyr/L-Tyr, et  $\Delta\Delta G = 11.8 \text{ kcal.mol}^{-1}$  pour la différence D-TyrAMP/L-TyrAMP. Ce mutant favoriserait la liaison de la D-Tyr, mais pas celle du D-TyrAMP. Le protocole d'évolution dirigée nous prédisait que le ligand L était favorisé dans les deux cas.

Pour le complexe avec le L-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.6 Å (table 6.33). Les résidus 41, 75, 179 et 201 stabilisent l'ammonium du L-TyrAMP (table 6.33). La structure de ce complexe n'évolue pas beaucoup au cours de la dynamique (figure 6.25).

Avec le D-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.8 Å (table 6.33). Seuls les résidus 41 et 175 stabilisent l'ammonium du D-TyrAMP (table 6.33). Ici, figure 6.25, la partie tyrosine du ligand a tourné d'environ 90°. Cette conformation du ligand dans le site actif apparaît défavorable : on perd l'interaction entre les groupement hydroxyle du

### 6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX

---

ligand et de la tyrosine 37, dont la chaîne latérale est figurée en gris.

TABLE 6.31 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant RYQQ.

Complexe	CASA	PBFE			$\Delta\Delta G$
		3 ns	5 ns		
TyrRS:D-Tyr	-7.67	-6.0	-5.9	-9.1	
TyrRS:L-Tyr	-11.34	3.7	3.2		
TyrRS:D-TyrAMP	-34.85	3.2	3.1	11.8	
TyrRS:L-TyrAMP	-39.26	-8.8	-8.8		

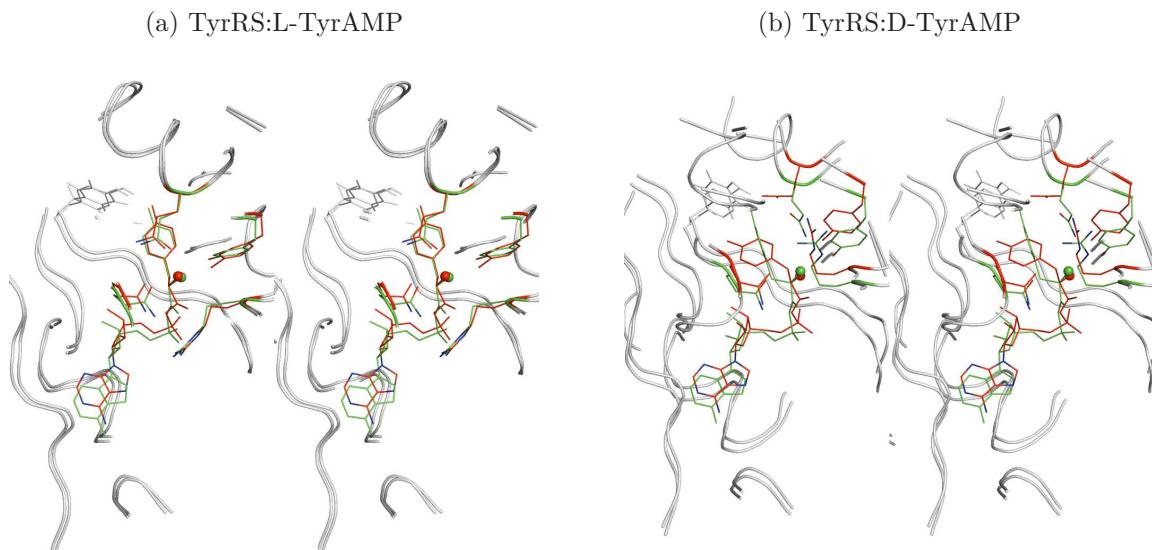
TABLE 6.32 – Dynamique moléculaire du mutant RYQQ. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

Complexe	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
TyrRS:D-Tyr	0.8	1.2	1.4	0.7	1.3	1.6
TyrRS:L-Tyr	0.7	1.1	1.6	0.7	1.1	0.9
TyrRS:D-TyrAMP	0.7	1.0	1.8	0.7	1.1	1.8
TyrRS:L-TyrAMP	0.6	0.8	0.6	0.6	0.8	0.7

TABLE 6.33 – Distances des atomes au groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du	
		L-TyrAMP ( $\text{\AA}$ )	D-TyrAMP ( $\text{\AA}$ )
OD2	ASP 41	3.3	2.7
OH	TYR 175	2.7	3.4
OE1	GLN 179	2.7	6.9
OE1	GLN 201	3.3	6.0

FIGURE 6.25 – Mutant RYQQ. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge).



### 6.3.3 Comparaison aux résultats expérimentaux

Plusieurs expériences ont été menées par les biochimistes Sandra Wydau et Pierre Plateau. Ils ont mené une mutagénèse combinatoire par oligonucléotides dégénérés sur la TyrRS, pour favoriser la liaison de la D-Tyr. Cette mutagénèse aléatoire concerne nos quatre résidus actifs : Asp81, Tyr175, Gln179 et Gln201. Malheureusement, la toxicité relative de la D-Tyr rend difficile la croissance et la détection des colonies de mutants actifs pour la D-Tyr expérimentalement. De plus, il existe une compétition entre l'ARNt chargeant l'acide aminé modifié et le facteur de terminaison de la traduction. Cette compétition s'effectue sur le site de fixation au ribosome en regard du codon stop. En conséquence, de nombreuses protéines tronquées et non actives sont produites lorsqu'on veut produire des protéines avec des acides aminés non naturels. Actuellement, les biologistes de cette équipe travaillent à mettre en place un crible suffisamment pertinent pour palier à ces difficultés.

Les expérimentalistes ont conduit des expériences d'aminoacylation en présence de 50  $\mu\text{M}$  d'ATP avec de la tyrosine radiomarquée, puis, ils ont inhibé le signal par introduction de tyrosine froide. Ces expériences ont été menées à la fois avec la L-Tyr et avec la D-Tyr froide. Les constantes d'inhibition calculées permettent d'avoir une information

### 6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX

---

biochimique sur l'efficacité des séquences prédictes pour lier la L-Tyr et la D-Tyr. Ils ont également mené une étude par fluorescence. Un signal est détecté pour la TyrRS native liée à la L-Tyr, mais pas pour la TyrRS native liée à la D-Tyr. On ne peut donc pas interpréter ces expériences pour la native comme pour les mutants.

Une autre expérience a été réalisée par Sandra Wydau et Pierre Plateau, l'analyse cinétique de la réaction d'échange de pyrophosphate inorganique (PPi). Cette réaction est étudiée en présence de 50  $\mu\text{M}$  d'ATP et de PPi radioactif. Le milieu contient initialement de la tyrosine radiomarquée, puis la réaction est inhibée avec de la L-Tyr ou de la D-Tyr froide. On mesure alors la vitesse  $V_{max}$ .



Dans le cas de la TyrRS native, les  $V_{max}$  ont été calculées pour la réaction d'échange de PPi (table 6.34). La TyrRS a une vitesse d'aminoacylation plus grande avec la L-Tyr qu'avec la D-Tyr. Par le calcul de constante d'inhibition, les expérimentalistes ont montré que la L-Tyr se lie 6 fois mieux que la D-Tyr à la TyrRS native (table 6.35). L'affinité de l'inhibiteur est d'autant plus grande que le  $K_i$  est faible. Ici,  $K_i$  de la L-Tyr froide est bien moindre que celui de la D-Tyr froide. Les résultats de dynamique moléculaire et les résultats expérimentaux sont du même ordre de grandeur pour la native. La dynamique moléculaire prédit en effet que la L-Tyr se lie 2.3 fois mieux que la D-Tyr.

TABLE 6.34 – Cinétique de la réaction d'échange de pyrophosphate inorganique.

sequences	$V_{max}(\text{L-Tyr}) \text{ s}^{-1}$	$V_{max}(\text{D-Tyr}) \text{ s}^{-1}$	$\frac{V_{max}(\text{D-Tyr})}{V_{max}(\text{L-Tyr})}$
WT=DYQQ	1.8	$6.5 * 10^{-3}$	$3.6 * 10^{-3}$
HYED	0.03	$1.7 * 10^{-4}$	$5.7 * 10^{-3}$
RYQQ	0.03	$1.7 * 10^{-4}$	$5.7 * 10^{-3}$

Les expérimentalistes ont également étudié la vitesse d'aminoacylation du mutant HYED. La  $V_{max}(\text{L-Tyr})$  et la  $V_{max}(\text{D-Tyr})$  sont diminuées mais le ratio D-Tyr/L-Tyr est légèrement amélioré (table 6.34). Par le calcul de constante d'inhibition (table 6.35), on voit que le ratio  $K_i^{\text{D-Tyr}}/K_i^{\text{L-Tyr}}$  augmente également. Ce mutant n'inverse donc pas le

TABLE 6.35 – Constantes d'inhibition de la L-Tyr et de la D-Tyr liée à la TyrRS native  
Unité:  $\mu\text{M}$

Séquence	$K_i^{D-\text{Tyr}}$	$K_i^{L-\text{Tyr}}$	$\frac{K_i^{D-\text{Tyr}}}{K_i^{L-\text{Tyr}}}$
DYQQ	210	35	6
KYED	143	14	10
KYEQ	486	8	60

ratio de fixation D-Tyr/L-Tyr, mais le  $K_i$  a quand même été diminué pour la D-Tyr par rapport à la native, donc on a quand même augmenté l'affinité pour la D-Tyr par ces mutations. La méthode d'évolution dirigée in silico est en accord sur ce point : le ratio de fixation D/L est amélioré mais pas inversé. La dynamique, elle, annonce que le ligand D-Tyr se lierait plus facilement, mais la dynamique avec le ligand adénylé donne une estimation de l'affinité qui ne va pas dans ce sens. Les résultats de la dynamique recoupent ainsi les résultats expérimentaux.

Pour le mutant KYEQ, le calcul de constante d'inhibition (table 6.35) montre une augmentation du ratio  $K_i^{D-\text{Tyr}}/K_i^{L-\text{Tyr}}$ . Ce mutant n'inverse donc pas le ratio de fixation D-Tyr/L-Tyr ; il avantage au contraire la fixation de la L-Tyr. La méthode d'évolution dirigée in silico prédisait ce mutant KYEQ avec un peu plus d'affinité pour la D-Tyr, ce qui est en accord avec les simulations de dynamique moléculaire mais n'est pas vérifié par les résultats expérimentaux. Pour le mutant RYQQ (table 6.34), on voit qu'en réalité la  $V_{max}$  est très affectée pour les deux ligands, cependant, le ratio des vitesses est comparable à celui de la native.

Ainsi, nous n'avons pas pour l'instant réussi à identifier de mutants inversant le ratio de fixation D/L. Cependant, les mutants se replient de manière stable et certains peuvent améliorer un peu la fixation de la D-Tyr par rapport à celle de la L-Tyr.

### 6.3.4 Séquences prédictes avec L- ou D-TyrAMP

#### 6.3.4.1 Mutant EQQQ

La séquence EQQQ a été prédictée en liaison avec le D-TyrAMP. La valeur de  $\Delta\Delta G$  calculée par PBFE est de  $-0.9 \text{ kcal.mol}^{-1}$ ; ce mutant favoriserait donc légèrement la liaison du D-TyrAMP. Les résultats du protocole d'évolution dirigée et de la dynamique moléculaire concordent (table 6.36).

Pour le complexe TyrRS:L-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure cristallographique sur une sphère de 15 Å autour du ligand est inférieur à 1.3 Å (table 6.37). Les résidus 41, 81, 175 et 179 interagissent avec l'ammonium du ligand (table 6.38). Au cours de la dynamique, la chaîne latérale de la glutamine 175 s'est réorientée dans le site actif. La chaîne latérale de la glutamine 179 s'est réajustée de manière à interagir avec l'ammonium du ligand (figure 6.26).

Pour le complexe TyrRS:D-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.2 Å. L'atome OE1 du glutamate 81 stabilise l'ammonium du ligand. On perd une interaction par rapport à l'aspartate natif à cette position. La chaîne latérale de la glutamine 175 est orientée de manière à libérer l'espace pour l'ammonium de la D-Tyr. L'atome OE1 de la glutamine 179 stabilise l'ammonium du ligand. Le glutamate 81 fait également une liaison hydrogène avec le résidu 204. La chaîne latérale du glutamate 175 prend moins d'espace dans le site actif que la tyrosine native et libère ainsi la place pour le positionnement de l'ammonium du D-TyrAMP. La structure ne se modifie pas réellement après 5 ns de dynamique moléculaire. Ce mutant semble donc très stable. L'ammonium du ligand s'est tout de même adapté au site catalytique (figure 6.26).

TABLE 6.36 – Affinité protéine:ligand ( $\text{kcal.mol}^{-1}$ ). Mutant EQQQ.

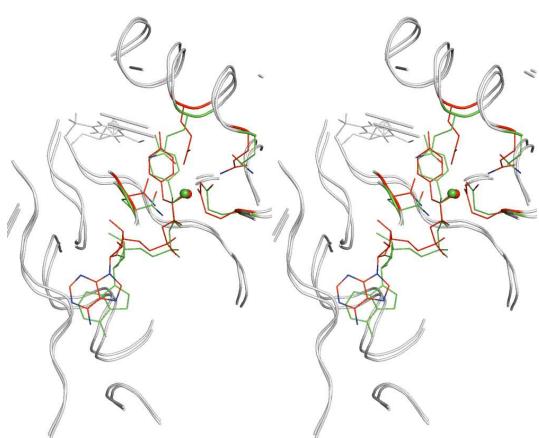
Ligand	CASA	PBFE				$\Delta\Delta G$
		3 ns $\varepsilon = 6$	5 ns $\varepsilon = 4$	5 ns $\varepsilon = 4$	$\Delta\Delta G$	
D-TyrAMP	-50.92	-8.9	-11.2	-11.3	-9.0	-0.9 ( $\varepsilon = 4$ )
L-TyrAMP	-46.79	-8.4	-10.5	-10.4	-8.5	

TABLE 6.37 – Dynamique moléculaire du mutant EQQQ. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

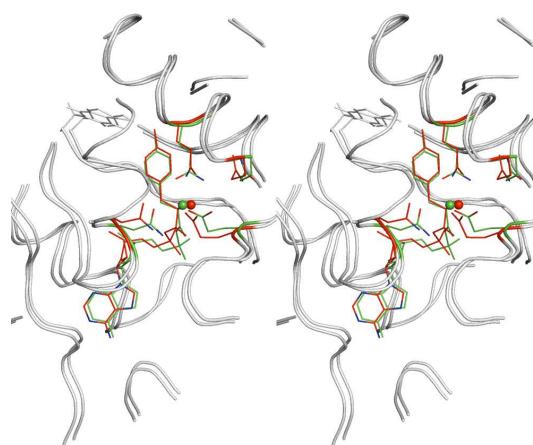
Ligand	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
D-TyrAMP	0.7	1.0	1.2	0.7	1.0	1.2
L-TyrAMP	0.8	1.2	0.7	0.8	1.3	0.7

FIGURE 6.26 – Mutant EQQQ. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge).

(a) TyrRS:L-TyrAMP



(b) TyrRS:D-TyrAMP



### 6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX

---

TABLE 6.38 – Distances des atomes au groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du	
		L-TyrAMP (Å)	D-TyrAMP (Å)
OD1	ASP 41	3.6	2.9
OD2	ASP 41	2.8	2.7
OE1	GLU 81	2.6	2.6
NE2	GLN 175	3.3	4.9
OE1	GLN 179	2.6	2.6

#### 6.3.4.2 Mutant QQEQ

La séquence QQEQ a été prédite à la fois en présence du ligand L-TyrAMP et du D-TyrAMP. Les calculs d'affinité PBFE indiquent que  $\Delta\Delta G = -0.9 \text{ kcal.mol}^{-1}$  pour la différence D-TyrAMP/L-TyrAMP (table 6.39). Ce mutant favorisera donc la liaison du D-TyrAMP, ce qui confirme les résultats d'évolution dirigée.

Pour le complexe TyrRS:L-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure cristallographique sur une sphère de 15 Å autour du ligand est inférieur à 1.2 Å (table 6.40). Les résidus 41 et 179 interagissent directement avec l'ammonium du L-TyrAMP (table 6.41). Sur la figure 6.27, on voit que la partie tyrosine du ligand s'est déplacée dans le site actif; toute la structure suit ce mouvement sans changement des interactions du ligand avec les résidus mutés au cours de la dynamique.

Pour le complexe TyrRS:D-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.1 Å (table 6.40). Le résidu 41 et 179 interagissent également avec le D-TyrAMP, mais plus les résidus 41 et 175 (table 6.41). La chaîne latérale de la glutamine 175 est orientée de manière à libérer l'espace pour le positionnement de l'ammonium du ligand dans le site actif. Le glutamate 179 avec ses atomes OE1 et OE2 stabilise plus l'ammonium du ligand que la glutamine native qui ne possède qu'un atome interagissant avec l'ammonium : OE1 (figure 6.27). Avec le ligand D-TyrAMP, la structure évolue peu au cours de la dynamique. Les interactions des chaînes latérales mutées sont conservées au cours de la dynamique.

TABLE 6.39 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant QQEQ.

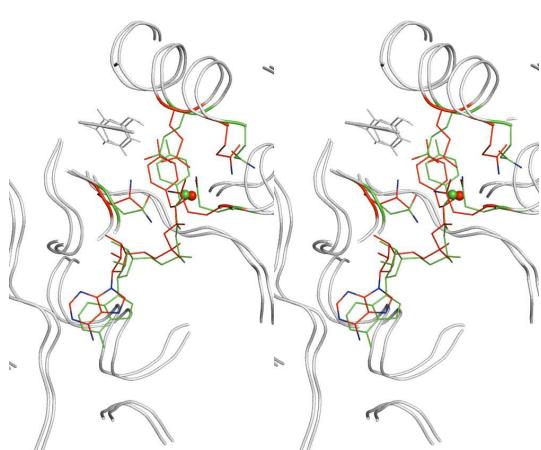
Ligand	CASA	PBFE				$\Delta\Delta G$
		3 ns	$\varepsilon = 6$	$\varepsilon = 4$	5 ns	
D-TyrAMP	-48.76	-9.2	-11.6	-11.5	-9.0	-0.9 ( $\varepsilon = 4$ )
L-TyrAMP	-47.17	-8.4	-10.2	-9.6	-7.9	

TABLE 6.40 – Dynamique moléculaire du mutant QQEQ. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

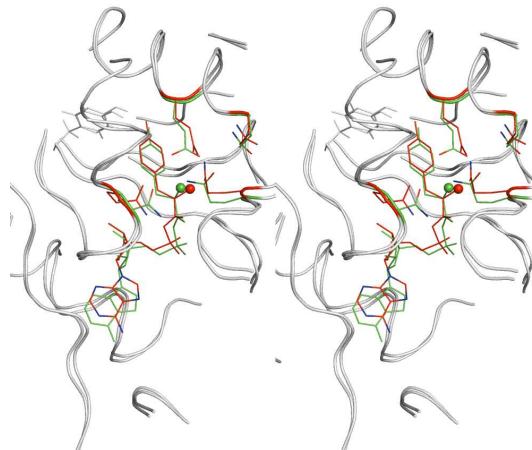
Ligand	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
D-TyrAMP	0.6	1.0	1.0	0.7	1.1	1.0
L-TyrAMP	0.7	1.1	0.8	0.8	1.2	0.9
APO	0.7	1.1	-	0.7	1.1	-

FIGURE 6.27 – Mutant QQEQ. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge).

(a) TyrRS:L-TyrAMP



(b) TyrRS:D-TyrAMP



### 6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX

TABLE 6.41 – Distances des atomes au groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du	
		L-TyrAMP (Å)	D-TyrAMP (Å)
OD1	ASP 41	2.7	2.8
OD2	ASP 41	2.6	2.9
OE1	GLU 179	2.6	3.4
OE2	GLU 179	3.8	2.8

#### 6.3.4.3 Mutant YQEQQ

La séquence YQEQQ a été prédite à la fois en présence du ligand L-TyrAMP et D-TyrAMP. Par PBFE, on obtient  $\Delta\Delta G = -6.3 \text{ kcal.mol}^{-1}$  pour la différence D-TyrAMP/L-TyrAMP. Ce mutant favoriserait donc la liaison du D-TyrAMP, ce qui confirme les résultats d'évolution dirigée.

Avec le ligand L-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieure à 1.1 Å (table 6.43). Les résidus 179, 201 et 204 interagissent avec le ligand. L'Asn 204 apparaît ici comme une position qu'il serait intéressant de muter (table 6.44). Au cours de la dynamique, la tyrosine 81 approche son groupe hydroxyle de l'ammonium du ligand. Le glutamate 179 s'en rapproche également. La chaîne latérale de la glutamine 175 s'est écartée du site actif (figure 6.28).

Avec le ligand D-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.1 Å. Les résidus 179 et 201 stabilisent l'ammonium du D-TyrAMP (table 6.44). Sur la figure 6.28, on voit qu'avec le ligand D-TyrAMP, la tyrosine 81 s'est retournée dans le site actif, pointe son groupe hydroxyle maintenant vers le phosphate. La chaîne latérale de la glutamine 175 s'est écartée du site actif. La nouvelle orientation de la chaîne latérale de la tyrosine 81 libère l'espace pour l'ammonium du ligand. La glutamine 201 stabilise à la fois l'ammonium et le groupe C=O du ligand.

TABLE 6.42 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant YQEQQ.

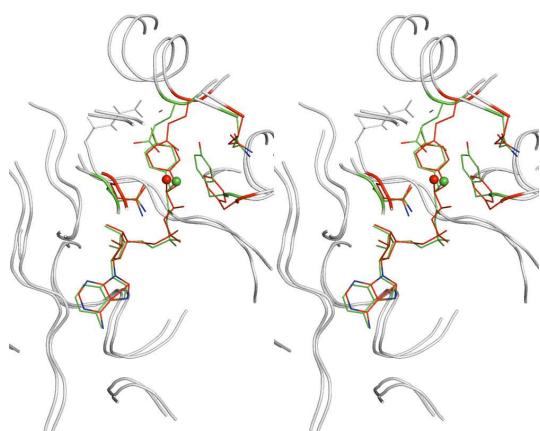
Ligand	CASA	PBFE				$\Delta\Delta G$	
		3 ns		5 ns			
		$\varepsilon = 6$	$\varepsilon = 4$	$\varepsilon = 4$	$\varepsilon = 6$		
D-TyrAMP	-43.44	-6.2	-6.7	-7.6	-6.0	-6.3 ( $\varepsilon = 4$ )	
L-TyrAMP	-36.15	-4.1	-3.1	-2.0	-2.8		

TABLE 6.43 – Dynamique moléculaire du mutant YQEQQ. Déviation RMSD (Å) par rapport à la structure initiale.

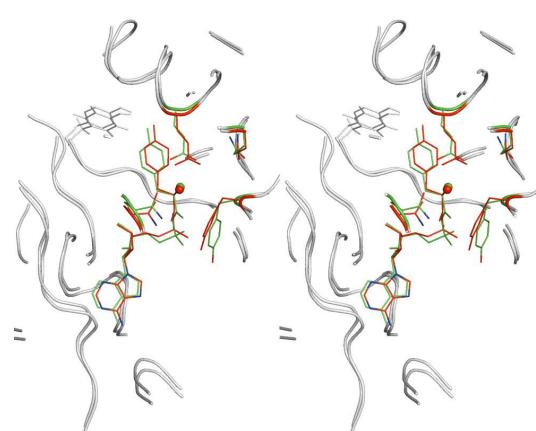
Ligand	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
D-TyrAMP	0.8	1.1	1.0	0.8	1.1	1.0
L-TyrAMP	0.8	1.5	1.5	0.8	1.6	1.1

FIGURE 6.28 – Mutant YQEQQ. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge).

(a) TyrRS:L-TyrAMP



(b) TyrRS:D-TyrAMP



### 6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX

---

TABLE 6.44 – Distances des atomes au groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du	
		L-TyrAMP (Å)	D-TyrAMP (Å)
OE1	GLU 179	2.6	2.7
OE2	GLU 179	2.9	3.7
OE1	GLN 201	2.8	2.7
HD21	ASN 204	3.2	4.4
HD22	ASN 204	3.4	3.4

#### 6.3.4.4 Mutant EQDQ

La séquence EQDQ a été prédite en liaison avec le D-TyrAMP. Par PBFE, on obtient  $\Delta\Delta G = 18.4 \text{ kcal.mol}^{-1}$  pour la différence D-TyrAMP/L-TyrAMP. Ce mutant défavoriserait donc la liaison du D-TyrAMP, ce qui confirme les résultats du protocole d'évolution dirigée (table 6.45).

Pour le complexe TyrRS:L-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 0.9 Å (table 6.46). Les résidus 41, 81 et 179 interagissent avec l'ammonium du L-TyrAMP (table 6.47). Au cours de la dynamique (figure 6.29), il y a un déplacement du ligand et une adaptation du squelette protéique, mais les positions des chaînes latérales sont restées celles prédites.

Pour le complexe TyrRS:D-TyrAMP, après 5 ns de dynamique, le rms moyen par rapport à la structure cristallographique sur une sphère de 15 Å autour du ligand est inférieur à 1.0 Å (table 6.46). Il y a beaucoup d'interactions avec le ligand D-TyrAMP. Les résidus 81, 179, 201 et 204 stabilisent cette forme du ligand (table 6.47). De même que pour le L-TyrAMP, la partie tyrosine du ligand s'est un peu déplacée au cours de la dynamique, mais les chaînes latérales des positions mutées ont conservé leur conformation au cours de la dynamique (figure 6.29).

TABLE 6.45 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant EQDQ.

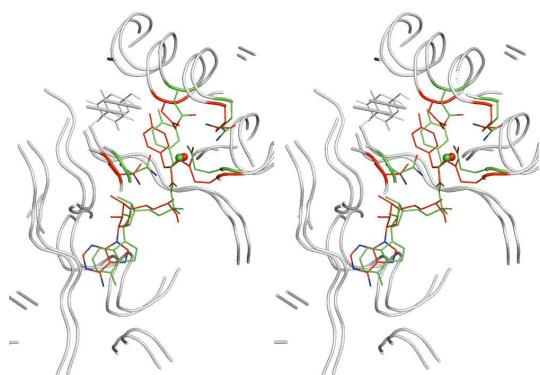
Ligand	CASA	PBFE				$\Delta\Delta G$	
		3 ns		5 ns			
		$\varepsilon = 6$	$\varepsilon = 4$	$\varepsilon = 4$	$\varepsilon = 6$		
D-TyrAMP	-57.61	-6.8	-10.0	-11.3	-	18.4 ( $\varepsilon = 4$ )	
L-TyrAMP	-46.97	-9.7	-16.0	-29.7	-		

TABLE 6.46 – Dynamique moléculaire du mutant EQDQ. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

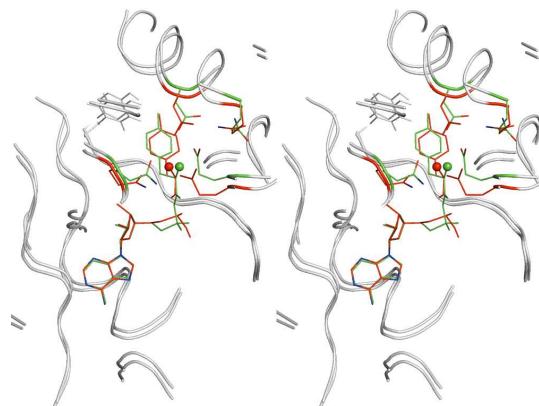
Ligand	3 ns			5 ns		
	chaîne principale	chaîne latérale	ligand	chaîne principale	chaîne latérale	ligand
D-TyrAMP	0.8	1.3	1.3	0.6	1.0	0.9
L-TyrAMP	0.8	1.1	0.6	0.6	0.9	0.7

FIGURE 6.29 – Mutant EQDQ. Superposition de la structure de meilleure affinité (en vert), et de la structure après 5 ns de dynamique moléculaire (en rouge).

(a) TyrRS:L-TyrAMP



(b) TyrRS:D-TyrAMP



### 6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX

---

TABLE 6.47 – Distances des atomes au groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du	
		L-TyrAMP (Å)	D-TyrAMP (Å)
OD1	ASP 41	2.9	5.3
OD2	ASP 41	2.8	4.5
OE1	GLU 81	2.6	2.8
OD1	ASP 179	2.8	4.4
OD2	ASP 179	3.6	2.8
OE1	GLN 201	4.4	2.6
HE22	GLN 201	5.0	3.2
ND2	ASN 204	4.8	3.1

#### 6.3.4.5 Mutant DQEQQ

La séquence DQEQQ a été prédite en liaison avec le D-TyrAMP. Pour la différence D-TyrAMP/L-TyrAMP,  $\Delta\Delta G = 1.0 \text{ kcal.mol}^{-1}$ . Ce mutant ne favoriserait donc pas la liaison du D-TyrAMP, comme prévu par l'évolution dirigée (table 6.48).

Avec le L-TyrAMP, après 3 ns de dynamique, le rms moyen par rapport à la structure de départ sur une sphère de 15 Å autour du ligand est inférieur à 1.1 Å (table 6.49). Les résidus 41, 81, 175 et 179 stabilisent l'ammonium du L-TyrAMP (table 6.50). La structure se réarrange au cours de la dynamique (figure 6.30). La glutamine 179 oriente l'atome OE1 en interaction avec l'ammonium du ligand. La chaîne latérale de la glutamine 201 s'écarte de l'ammonium du ligand.

Pour le D-TyrAMP, après 3 ns de dynamique, le rms moyen par rapport à la structure cristallographique sur une sphère de 15 Å autour du ligand est inférieur à 1.3 Å (table 6.49). Les résidus 81, 179, 201 et 204 interagissent avec l'ammonium du ligand (table 6.50). Ici, la structure évolue peu au cours de la dynamique (figure 6.30).

TABLE 6.48 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant DQEQQ.

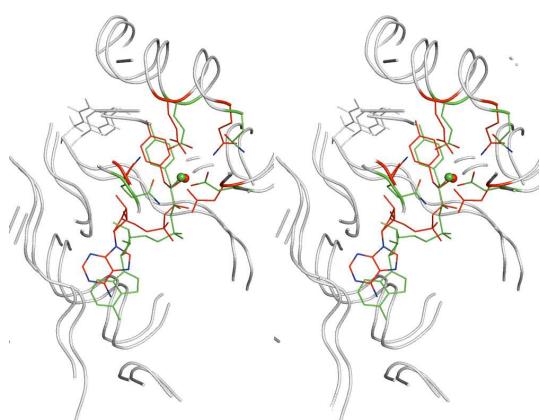
Ligand	CASA	PBFE		$\Delta\Delta G$
		3 ns $\varepsilon = 4$	3 ns $\varepsilon = 6$	
D-TyrAMP	-59.4	-5.5	-6.1	1.0 ( $\varepsilon = 4$ )
L-TyrAMP	-61.6	-6.5	-7.2	

TABLE 6.49 – Dynamique moléculaire du mutant DQEQQ. Déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale.

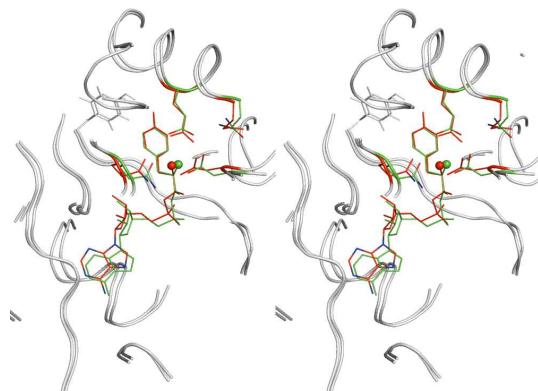
Ligand	chaîne principale	3 ns chaîne latérale	ligand
D-TyrAMP	0.8	1.3	0.9
L-TyrAMP	0.8	1.1	1.0

FIGURE 6.30 – Mutant DQEQQ. Superposition de la structure de meilleure affinité (en vert), et de la structure après 3 ns de dynamique moléculaire (en rouge).

(a) TyrRS:L-TyrAMP



(b) TyrRS:D-TyrAMP



### 6.3. ETUDE PAR DYNAMIQUE MOLÉCULAIRE ET CONFRONTATION AUX RÉSULTATS EXPÉRIMENTAUX

---

TABLE 6.50 – Distances des atomes au groupement ammonium après 5 ns de dynamique.

Nom de l'atome	Acide aminé	Distance à l'azote du	
		L-TyrAMP (Å)	D-TyrAMP (Å)
OD2	ASP 41	2.6	8.2
OD1	ASP 81	4.5	2.8
OD2	ASP 81	2.7	4.2
NE2	GLN 175	3.2	5.9
OE1	GLU 179	2.7	3.0
OE2	GLU 179	4.6	2.6
OE1	GLN 201	5.6	2.7
NE2	GLN 201	4.6	3.3
HD21	ASN 204	6.1	2.8

#### 6.3.5 Séquences d'affinité moindre

##### 6.3.5.1 Mutant DFEQ

La séquence DFEQ a été prédite en liaison avec le D-TyrAMP. On obtient,  $\Delta\Delta G = 5.0 \text{ kcal.mol}^{-1}$  pour la différence D-TyrAMP/L-TyrAMP avec  $\varepsilon = 6$ , en utilisant 3 ns de dynamique. Ce mutant favoriserait donc légèrement la liaison du D-TyrAMP par rapport à l'enzyme natif, comme prévu par l'évolution dirigée (figure 6.51).

TABLE 6.51 – Affinité protéine:ligand ( $\text{kcal.mol}^{-1}$ ) et déviation RMSD (Å) par rapport à la structure initiale. Mutant DFEQ.

Ligand	CASA	PBFE			RMSD sur 3 ns		
		3 ns	$\varepsilon = 6$	$\Delta\Delta G$	chaîne principale	chaîne latérale	ligand
D-TyrAMP	-56.88	-3.3	5.0		0.8	1.3	1.4
L-TyrAMP	-47.80	-8.3			0.8	1.3	0.9

**Mutant EQVQ** La séquence EQVQ a été prédite en liaison avec le D-TyrAMP. On obtient,  $\Delta G_{D-TyrAMP} = -4.2 \text{ kcal.mol}^{-1}$  pour la différence D-TyrAMP/L-TyrAMP avec  $\varepsilon = 6$ , en utilisant 3 ns de dynamique. Ce mutant favoriserait donc la liaison du D-TyrAMP, comme prévu par l'évolution dirigée (figure 6.52).

##### 6.3.5.2 Mutant NQEQ

La séquence NQEQ a été prédite à la fois en présence du ligand L-TyrAMP et D-TyrAMP. Avec le calcul PBFE, on obtient  $\Delta\Delta G = 12.7 \text{ kcal.mol}^{-1}$  pour la différence

TABLE 6.52 – Affinité protéine:ligand ( $\text{kcal} \cdot \text{mol}^{-1}$ ) et déviation RMSD ( $\text{\AA}$ ) par rapport à la structure initiale. Mutant EQVQ.

Ligand	CASA	PBFE		3 ns		
		3 ns	$\varepsilon = 6$	chaîne	chaîne	
		principale	latérale	ligand		
D-TyrAMP	-50.62	-4.2		0.7	1.2	0.9

D-TyrAMP/L-TyrAMP, de 1 à 5 ns de dynamique. Ce mutant ne favorisera donc pas la liaison du D-TyrAMP, comme prévu par l'évolution dirigée (figure 6.53).

TABLE 6.53 – Affinité protéine:ligand ( $\text{kcal} \cdot \text{mol}^{-1}$ ). Mutant NQEQQ.

Ligand	CASA	PBFE				$\Delta\Delta G$	RMSD sur 3 ns			
		3 ns		5 ns			chaîne	chaîne		
		$\varepsilon = 6$	$\varepsilon = 4$	$\varepsilon = 4$	$\varepsilon = 6$		principale	latérale	ligand	
D-TyrAMP	-54.19	-5.3	2.6	-	-	12.7	0.8	1.2	1.1	
L-TyrAMP	-50.80	-9.4	-10.1	-	-13.4	-	0.8	1.2	0.8	

### 6.3.5.3 Mutant DFDQ

La séquence DFDQ a été prédite en liaison avec le D-TyrAMP. Les calculs PBFE donnent  $\Delta\Delta G = 1.9 \text{ kcal} \cdot \text{mol}^{-1}$  en utilisant 3 ns de dynamique, avec  $\varepsilon = 6$ . Ce mutant favorisera donc légèrement la liaison du D-TyrAMP, comme prévu par l'évolution dirigée (figure ??).

captionAffinité protéine:ligand ( $\text{kcal} \cdot \text{mol}^{-1}$ ). Mutant DFDQ.

Ligand	CASA	PBFE		RMSD sur 3 ns		
		3 ns	$\varepsilon = 6$	chaîne	chaîne	
		$\Delta\Delta G$	principale	latérale	ligand	
D-TyrAMP	-58.92	-2.3	1.1	1.9	0.9	1.3
L-TyrAMP	-48.08	-4.2		0.8	1.3	0.9

### 6.3.5.4 Mutant YFEQ

La séquence YFEQ a été prédite à la fois en présence du ligand L-TyrAMP et D-TyrAMP. Les calculs PBFE donnent  $\Delta G = 5.5 \text{ kcal} \cdot \text{mol}^{-1}$  en utilisant 3 ns de dynamique, avec  $\varepsilon = 6$ . Ce mutant favorisera donc très légèrement la liaison du D-TyrAMP, comme prévu par l'évolution dirigée (tablee 6.54).

TABLE 6.54 – Affinité protéine:ligand (kcal.mol<sup>-1</sup>). Mutant YFEQ.

Ligand	CASA	PBFE		RMSD sur 3 ns		
		3 ns	$\varepsilon = 6$	chaîne principale	chaîne latérale	ligand
D-TyrAMP	-34.37	-3.4	-5.47	0.9	1.5	1.1

## 6.4 Conclusion

Nous avons mené ici l'évolution dirigée de la TyrRS d'*E. coli* avec les ligands tyrosine et tyrosyl-adénylate. Plusieurs observations ressortent. En présence du ligand tyrosine, on prédit plus d'acides aminés chargés positivement qu'avec l'adénylate. Ceci peut être dû à l'encombrement stérique. En effet, la partie AMP du ligand rend le placement de grandes chaînes latérales (Arg, Lys) moins facile dans le site actif. Avec le TyrAMP, on prédit au contraire des résidus chargés négativement sur les trois premières positions mutées. Avec ce ligand adénylé, la position 201 ne mute pas, alors qu'elle peut le faire. Cette glutamine 201 réalise une liaison hydrogène avec la glutamine 200. Ainsi, la chaîne latérale de Glu201 est très bien orientée pour interagir avec l'ammonium du ligand. Au cours de l'évolution dirigée, nous nous sommes aperçus qu'un mutant favorisant la liaison de la D-Tyr n'était pas forcément favorable à la liaison du D-TyrAMP. C'est le cas du mutant RYQQ par exemple. On voit qu'il n'est donc pas si aisément de trouver un mutant qui stabilise les différents états du complexe.

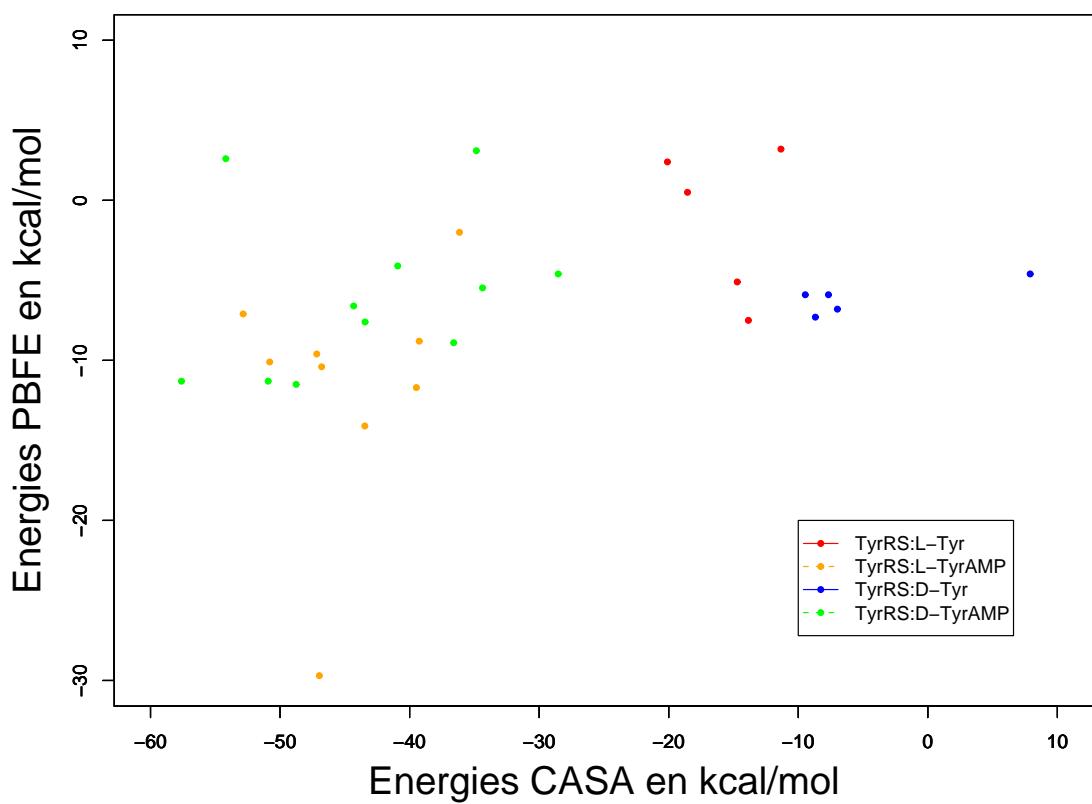
L'observation des énergies de repliement est également intéressante. Pour la TyrRS native et les mutants HYED et KYED, on observe que l'énergie de repliement est plus stable pour le complexe TyrRS:TyrAMP que pour le complexe TyrRS:Tyr. La forme L ou D ne modifie pas cette stabilité pour le complexe TyrRS:Tyr, en revanche, elle stabiliserait davantage le complexe TyrRS:TyrAMP. On note que si les résultats CASA et PBFE donnent souvent la même information sur l'affinité pour un ligand donné, les deux jeux de données ne sont pas très corrélés. Nous représentons dans la figure 6.31 tous les résultats d'affinité CASA et PBFE (avec  $\varepsilon = 4$ ).

Les analyses structurales que nous avons conduites donnent plusieurs informations. Les glutamines natives ou prédictes ont tendance à orienter l'atome OE1 vers l'ammonium

du ligand. Les lysines, prédites souvent avec la D-Tyr, s'avèrent stabiliser le groupement phosphate de l'AMP et sont peu impliquées dans la spécificité. Les résidus aromatiques tyrosine et histidine peuvent déplacer toute leur chaîne latérale en présence de D-Tyr, et ainsi libérer l'espace pour l'ammonium de la D-Tyr dans le site actif du ligand. En dehors des résidus 81, 175, 179 et 201, deux autres résidus sont ici remarqués. La glutamine 200 et l'arginine 204. Ils n'interagissent pas directement avec le ligand, mais appartiennent au réseau d'interactions secondaires stabilisant le ligand.

Quelques séquences obtenues par évolution dirigée ont ensuite fait l'objet d'analyses biochimiques. A l'issue des expériences, le mutant KYED apparaît comme augmentant le ratio de fixation D/L, sans l'inverser, ce qui corrobore les résultats de l'évolution dirigée et de la dynamique moléculaire menée sur ce mutant. Il en est de même pour le mutant HYED. En revanche, pour le mutant KYEQ, les expériences montrent que la fixation de la D-Tyr n'est pas améliorée par rapport à la TyrRS native, alors que les prédictions d'évolution dirigée et les dynamiques prévoient une légère amélioration.

L'évolution dirigée avec le D-TyrAMP propose les mutants EQQQ, QQEQ, YQEQQ, comme favorables à la liaison du D-TyrAMP. Les résultats de dynamique moléculaire confirment ce résultat. Ces séquences restent à tester expérimentalement.

FIGURE 6.31 – Affinités PBFE ( $\varepsilon = 4$ ) en fonction des affinités CASA.

## 6.5 Vers de nouvelles positions à muter pour mieux lier la D-Tyrosine

Pour trouver d'autres résidus de la structure qu'il serait intéressant de muter, nous avons procédé à une recherche exhaustive d'acide aminés de la structure cristallographique sur la base de deux critères: (i) la distance au groupement ammonium de la L-Tyr ou de la D-Tyr doit être inférieure à 10 Å; (ii) la surface accessible au solvant de ce résidu doit également être supérieure à 10% de la surface totale exposée quand la protéine est dépliée. On identifie ainsi treize positions intéressantes: 40F, 41D, 43T, 81D, 85K, 89R, 175Y, 176N, 179Q, 182D, 200D, 201Q, 204N. La position 41 doit être laissée de côté, car l'Asp41 interagit avec l'ion magnésium de manière essentielle pour l'aminoacylation de l'ARNt. On étudie quelques mutations aux autres positions avec un protocole simple. A partir de la structure cristallographique, on teste tous les rotamères à la position modifiée, le reste de la structure étant fixe. Une faible minimisation d'énergie de la chaîne latérale mutée en présence du reste du système fixe permet à celle-ci de s'adapter. Pour chacun de ces rotamères, on estime l'affinité protéine-ligand. On garde alors le rotamère qui donne la meilleure affinité (table 6.55). La seule mutation qui semble favoriser la D-Tyr est la mutation N204R, qu'il serait intéressant d'étudier plus en détail.

Nous avons également appliqué le protocole d'évolution dirigée par Proteus, en définissant quelques-uns des résidus ci-dessus comme "actifs". Les positions autorisées à muter sont les suivantes: 40, 41, 81, 175, 176, 179, 201 (table 6.56). En gras, sont figurés les acides aminés déjà rencontrés aux positions correspondantes dans cette thèse. Bien qu'on parte de séquences et structures aléatoires au début d'un cycle heuristique, on retrouve ici une identité forte avec des séquences que nous avons déjà étudiées. Notre méthode d'évolution dirigée permet donc de converger vers des structures similaires à travers différents essais sur le même système.

**6.5. VERS DE NOUVELLES POSITIONS À MUTER POUR MIEUX LIER LA D-TYROSINE**

---

TABLE 6.55 – Affinités relative D/L pour des mutations choisies de la TyrRS d'*E. coli*. Différences d'énergies libres de liaison (D-Tyr) - (L-Tyr) en kcal.mol<sup>-1</sup>.

Mutation	différence D/L
native	7.05
F40E	5.9
T43E	6.39
D81L	3.74
K85E	6.28
R89E	6.48
Y175F	5.68
N176L	6.25
N176K	6.29
Q179L	4.42
Q179N	4.89
D182L	8.73
D200E	6.59
Q201D	2.42
N204D	6.42
N204R	-1.38

TABLE 6.56 – Evolution dirigée de la TyrRS d'*E. coli* avec le ligand D-Tyr et 6 positions actives.

Positions mutées	40	81	175	176	179	201
Native	F	<b>D</b>	<b>Y</b>	N	<b>Q</b>	<b>Q</b>
Séquences	H	W	<b>E</b>	Y	L	R
Prédites	<b>E</b>	H	W	D	H	L
	D	<b>K</b>	<b>D</b>	H	K	<b>Q</b>
	K	<b>Y</b>	<b>Y</b>		W	H
	R	<b>E</b>	S		F	<b>D</b>
	N		<b>Q</b>		<b>D</b>	E
					T	



# Chapitre 7

## Conclusion

L'introduction d'acides aminés non naturels dans les protéines permet d'accroître les possibilités du Protein Design. Des peptides aux fonctions différentes peuvent réaliser de nouvelles réactions chimiques. Le CPD joue donc un rôle important dans la recherche de peptides pharmaceutiques, et dans l'industrie des biotechnologies. Par exemple, des peptides synthétisés avec des acides aminés D seraient plus résistants aux protéases, qui les reconnaissent mal. De tels peptides auraient une durée de vie prolongée dans la cellule. C'est pourquoi nous nous y intéressons dans cette thèse.

La méthode de CPD présentée dans cette thèse se veut générale. Elle a déjà été mise en œuvre pour la modélisation de l'évolution dirigée de l'AsnRS avec le solvant CASA. Les développements récents apportés ont permis de réaliser le CPD de l'AsnRS avec le solvant GB également. Le but était de lier préférentiellement l'aspartate au lieu du ligand naturel, l'asparagine. La méthode de CPD que nous avons utilisée est simple, mais exhaustive ; elle permet de proposer de nombreux mutants. Les séquences ainsi prédictes doivent ensuite faire l'objet d'une étude sur la spécificité, être étudiées par dynamique moléculaire, et enfin être étudiées par mutagénèse dirigée expérimentale.

Cette méthode présente quelques inconvénients. Tout d'abord, le squelette protéique est fixé dans sa conformation cristallographique et ne s'adapte pas aux mutations introduites. Ensuite, la discréétisation de l'espace conformationnel par l'usage d'une bibliothèque de rotamères introduit un biais, car on ne peut pas analyser l'immense possibilité de conformations comme dans un espace continu. Une autre limitation qui nous intéresse

directement est l'incertitude sur les calculs de stabilités et d'affinités, incertitudes qui sont proches des différences d'énergie libre de liaison entre D-Tyr et L-Tyr. Ces incertitudes rendent difficile l'identification de mutants ayant une affinité pour la D-Tyr supérieure à la native. Si nous avions choisi d'introduire un groupement chimique plus conséquent qu'une simple modification de la stéréochimie, les résultats donnés par la méthode auraient probablement été plus tranchés en faveur de tel ou tel ligand.

Nous avons utilisé au cours de cette étude un programme d'évolution dirigée reposant sur un champ de force et un modèle de solvant simple, dans un premier temps (CHARMM19 et CASA respectivement). Nous avons montré que cette méthode simple permettait d'estimer approximativement la stabilité des protéines ou encore l'affinité des complexes. Dans un second temps, nous avons essayé d'améliorer ce modèle en utilisant un champ de force "tout atomes" et un modèle de solvant plus sophistiqué (respectivement AMBER et GB-HCT). Avec ces deux méthodes, on obtient des résultats cohérents aux résultats expérimentaux. Le facteur limitant de l'étude avec le solvant GB-HCT a été le temps de calcul ; mais l'automatisation des procédures et l'amélioration de l'efficacité des calculs permettra dans un avenir proche d'utiliser le CPD avec le champ de force AMBER et un modèle de solvant GB-HCT, de façon plus automatique. Ceci permettrait sans doute d'obtenir des séquences plus cohérentes d'un point de vue de la modélisation des charges.

Au cours de cette étude, nous avons cherché à caractériser les séquences et structures obtenues par le protocole de CPD. Les structures prédictives sont stables et peuvent présenter des affinités supérieures à la native pour un ligand donné. Puis, nous nous sommes intéressés à la modélisation de l'évolution dirigée de la TyrRS d'*E. coli*., pour qu'elle lie plus favorablement la D-Tyrosine ou le D-Tyrosyl adénylate. Notre protocol de CPD nous permet d'identifier quelques mutants prometteurs. Les tests expérimentaux sont mitigés, mais une activité pour la D-Tyrosine a bien été identifiée. Pour la TyrRS native, les données expérimentales sont en assez bon accord avec les simulations de dynamique moléculaire. Nous sommes donc en mesure de penser que les mutant HYED et KYED favoriseraient la liaison de la D-Tyr. Mais, les simulations de dynamique moléculaire, nous montrent que dans le cas de KYED, le D-TyrAMP n'est pas lié avec plus d'affinité que le

---

L-TyrAMP. Il faut donc rechercher des mutants qui stabilisent également le D-TyrAMP. Le mutant KYEQ qui semblait prometteur par sa bonne affinité à la D-Tyr, apparaît en réalité peu spécifique. Les mutants EQQQ et QQEQ sont prédis avec une bonne affinité et une bonne spécificité pour le D-TyrAMP, et ceci est vérifié par les simulations de dynamique moléculaire. Ils sont également prédis pour être plus spécifiques que les mutants précédents. Ces séquences prometteuses devront faire également l'objet de vérifications expérimentales.



# Chapitre 8

## Annexes

### 8.1 CPD avec le champ de force "tout atome" AMBER et le modèle de solvant GB-HCT

#### 8.1.1 Champ de force AMBER

Jusqu'à présent nous utilisions le champ de force Charmm19 pour notre modélisation de l'évolution dirigée. Dans ce modèle les hydrogènes polaires sont représentés explicitement alors que les hydrogènes non polaires sont combinés aux atomes qui les portent. Autrement dit, les hydrogènes non polaires sont représentés implicitement, par un changement des paramètres des atomes de carbone. Ceci permet d'obtenir des résultats proches de ceux issus d'une représentation tout atomes. Ce champ de force a été développé initialement pour augmenter la rapidité des modélisations. Cet avantage est maintenant plus relatif, compte tenu de l'augmentation des ressources de calculs. Nous cherchons donc à implémenter le champ de force "tout atome" AMBER dans notre procédure de CPD. Ce champ de force a été paramétré sur des données expérimentales de petites molécules et de calcul *ab initio* pour les charges [106].

### 8.1.2 modèle de solvant GB

Nous avons remarqué au cours de notre étude que le modèle de solvant CASA retranscrit mal les interactions électrostatiques. Il a aussi été montré que le champ de force AMBER permet de modéliser les structures d'une manière plus corrélée avec les résultats obtenus par un modèle explicite comme la résolution de l'équation de Poisson-Boltzmann [136]. La modélisation du solvant GB a donc été implémentée dans notre modèle d'évolution dirigée, conjointement au champ de force AMBER tout atome. Cette représentation du solvant apparaît dans la fonction d'énergie à travers une énergie de solvatation :

$$E_{elec} = E_{coul} + E_{solv} \quad (8.1)$$

$$E_{solv} = E_{self} + E_{int} \quad (8.2)$$

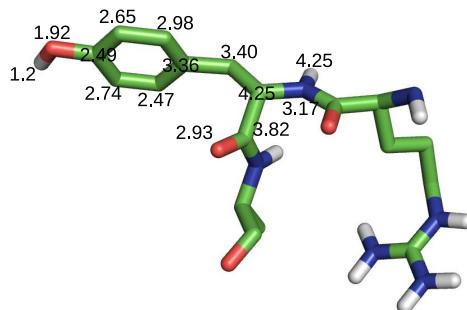
On rappelle que  $E_{self}$  est la somme des énergies self atomiques, c'est à dire les énergies d'interaction des charges i du soluté avec le solvant. Ce terme est inversement proportionnel au rayon de Born,  $B_i$ .  $E_{int}$  est la somme des énergies d'interaction entre les charges i et j du soluté. Ce terme tient compte de l'écrantage des charges par le solvant. Dans cette approche, l'énergie d'interaction entre deux acides aminés dépend de toute la structure protéique, mais peut être calculé approximativement à partir des énergies des paires de résidus [135]. Le grand temps de calcul nécessaire à ce modèle plus précis que CASA, nous oblige à paralléliser les calculs et à utiliser le supercalculateur du CINES (Centre Informatique National de l'Enseignement Supérieur).

### 8.1.3 Une approximation des rayons de solvations

Pour les modélisations avec le solvant GB, une collaboration a été menée avec le Pr. Giorgios Archontis et son doctorant Savvas Polydoridis du département de physique, de l'université de Chypre. Un des paramètres du modèle de Born est le rayon de solvation  $B_i$  de l'atome i. Pour une charge au centre d'une cavité sphérique, ce rayon serait celui de la sphère. Dans notre étude, on construit  $B_{res}$ , une approximation de  $B_i$ . On calcule un rayon de Born moyen pour le squelette protéique et un autre pour la chaîne latérale,

FIGURE 8.1 – représentation des rayon de Born atomique,  $B_{res}$ , et de leurs simplification par des moyennes sur la chaîne latérale et sur le squelette protéique,  $B_{avg}$

$$B_{res(\text{chaîne latérale})} = 2.42 \quad B_{res(\text{squelette protéique})} = 3.59$$



et ce, pour chaque résidu. Les  $B_{res}$  sont calculés dans une étape antérieure au calcul de la matrice d'interaction. Ainsi, pendant l'étape de calcul de la matrice d'énergie, on lit directement ces rayons atomiques moyens.

Le calcul des rayons moyens  $B_{res}$  se déroule comme suit. On utilise une constante diélectrique de 4 pour le soluté et de 80 pour l'eau.

On calcule l'énergie self de tout le système, puis on fixe ce terme pour chaque atome au cours de l'étude. On garde en mémoire toutes les charges des atomes, puis on les rend nulles, sauf pour la chaîne latérale de la position 1, le rayon de Born moyen pour la chaîne latérale est calculée à partir de l'énergie Self.

$$b_i = -\frac{\theta q_i^2}{2\Delta E_i^{self}} \quad (8.3)$$

Ensuite, tous les atomes récupèrent leur charge initiale, la même opération est faite pour calculer le rayon moyen du squelette protéique de la position 1. L'opération est répétée pour tous les acides aminés de la protéine ainsi que pour le ligand.

D'autre part, nous nous intéressons aux énergies de référence des 20 acides aminés, que nous estimons à partir de tri-peptides Ala-X-Ala, avec un modèle de solvant de Born généralisé GB-HCT et le champ de force Amber "tout atome". On utilise le même jeu de

tri-peptides que celui utilisés pour la paramétrisation du modèle CASA. Ces structures sont issues de tri-peptides extraits des structures cristallographiques de six protéines de la PDB : le lysosome (2LZM), la BPTI (Bovin Pancreatic Trypsin Inhibitor ; 4PTI), la nucléase de staphylocoque (1STN), l' $\alpha$  toxine (1PTX), la ribonucléase A (2RN2), et la cyclophiline (2CPL). Pour chaque tri-peptide, l'énergie  $E_{self}$  est calculée pour l'ensemble du tri-peptide, puis fixée pour la suite de l'étude. L'énergie d'interaction de la chaîne latérale X avec elle-même et le squelette protéique est calculée. Pour choisir la structure de tri-peptide optimale pour chaque type X, le critère est l'énergie après minimisation. Ainsi, pour chaque type d'acide aminé X, la structure du tri-peptide donnant l'énergie la plus faible est retenue, pour représenter l'acide aminé X dans l'état dénaturé.

Les énergies obtenues avec les deux modèles de solvant pour chaque acide aminé sont représentées dans la table 8.1. Ces deux modèles donnent des énergies du même ordre de grandeur. Mais avec GB, on remarque que l'effet électrostatique est plus important, et sans doute mieux modélisé, les différences d'énergies entre acides aminés polaires et apolaires étant plus grandes. Les énergies de référence propres au modèle GB nous serviront par la suite dans toutes les modélisations d'évolution dirigée réalisées avec ce solvant.

Voyons maintenant si notre approximation du rayon de solvatation moyen ne conduit pas à une erreur trop grande pour notre modèle. Pour cela nous calculons les énergies de référence avec et sans moyennation des rayons de Born moyen (figure 8.2 et 8.2).

Les énergies de référence modélisées avec ou sans moyennation des rayons de Born atomiques sont très proches. On va pouvoir utiliser cette approximation dans les modélisations qui suivent. Les rayons vont donc pouvoir être pré-calculés et appliqués aux peptides modélisés.

## 8.1.4 Validations expérimentales

### 8.1.4.1 Calcul d'énergie libre sur des peptides

Les peptides utilisés pour ce test sont trois peptides en hélices  $\alpha$ . Une position particulière est mutée en tous les acides aminés possibles. Des données expérimentales pour toutes ces mutations permettent d'extraire l'énergie libre de repliement des hélices natives

**8.1. CPD AVEC LE CHAMP DE FORCE "TOUT ATOME" AMBER ET LE MODÈLE DE SOLVANT GB-HCT**

---

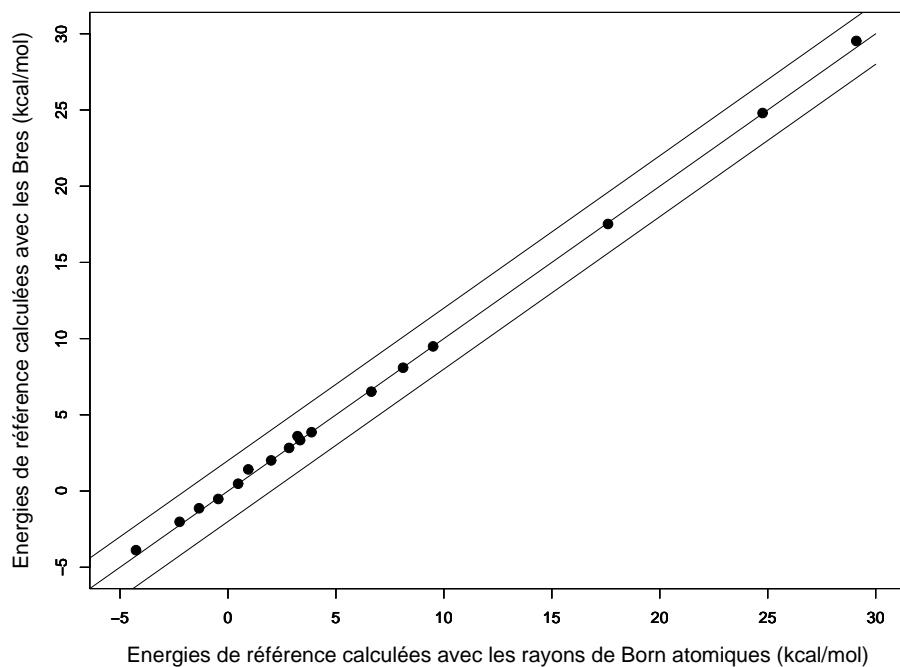
TABLE 8.1 – Énergie de référence (kcal/mol) avec deux modèles de solvant différents.

Acide aminé central du tri-peptide	modèle de solvant CASA dans le champ de force CHARMM19	modèle de solvant GB-HCT dans le champ de force AMBER avec $\epsilon=4$
ALA	-10.89	4.23
ASP	-19.30	-11.15
ASN	-16.40	-11.73
ARG	-24.96	-38.77
GLU	-20.86	-5.38
GLN	-16.70	-9.22
HIS	-18.0	(HIE) 24.37
HSP	-21.88	(HIP) 30.40
ILE	-12.61	10.34
LEU	-12.26	6.48
LYS	-21.64	1.90
MET	-13.30	2.95
PHE	-16.52	7.68
SER	-13.00	-1.10
TYR	-18.62	1.97
THR	-12.13	-2.47
TRP	-20.91	19.38
VAL	-10.50	2.08

TABLE 8.2 – Énergie de référence (kcal/mol) avec le modèle de solvant GB-HCT. Energies calculées avec rayons atomiques dans la colonne de gauche  $B_i$  et rayon moyen dans la colonne de droite  $B_{res}$ . ( $\epsilon = 4$ )

Acide aminé	$B_i$	$B_{res}$
ALA	4.23	4.25
ASP	-11.152	-10.35
ASN	-11.73	-11.26
ARG	-38.77	-37.15
GLU	-5.38	-4.35
GLN	-9.22	-8.77
HIE	24.37	24.44
HIP	30.40	31.33
ILE	10.34	10.33
LEU	6.48	6.46
LYS	1.90	2.69
MET	2.95	2.97
PHE	7.68	7.44
SER	-1.10	-1.24
TYR	1.97	1.95
THR	-2.47	-2.44
TRP	19.38	19.18
VAL	2.08	2.09

FIGURE 8.2 – Calcul des énergies de référence avec et sans l'utilisation de rayons de solvatation moyen.



et mutantes [206, 196] . On ne considérera pas les mutations en cystéine, glycine et proline, qu'on ne peut pas modéliser avec notre protocole. On dispose au final d'informations biochimiques sur 47 mutations de ces trois peptides. K2AE2 est une hélice de 21 acides aminés, qui sera mutée sur la position 12. KEAKE une hélice de 17 résidus, mutée sur la position 9 ; et pepT1 est un fragment d'hélice de la ribonucléase T1, avec 17 résidus, mutée sur la position 9. A la position mutée, on place successivement tous les acides aminés possibles, si bien qu'on obtient une protéine native et dix sept mutants pour chaque peptide. Voici les séquences des trois peptides utilisés :

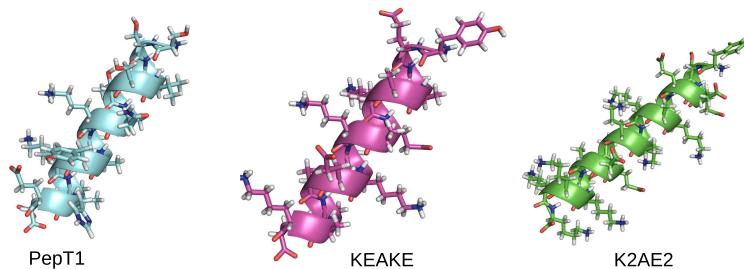
pepT1 : SSDVSTAQXAAYKLHED [196],

KEAKE : YEAAAKEAXAKEAAAKA [197],

K2AE2 : YSEEEEKAKKAXAEEAEKKKK [198],

Le protocole de modélisation est l'optimisation structurale des rotamères par le programme Proteus, décrit précédemment. Pour chaque peptide et pour chaque séquence mutante, 100000 cycles heuristiques sont réalisés ; compte tenu de la petite taille des pep-

FIGURE 8.3 – 3 peptides utilisés pour tester le protocol d'évolution dirigée avec le modèle de solvant GB-HCT



tides, ils ne conduisent qu'à un nombre très faible de structures différentes (inférieur à 5). On reconstruit les structures obtenues. Les programmes ont été modifiés pour prendre en compte la nouvelle modélisation du solvant. Ici on obtient une erreur moyenne de 0.99 kcal.mol<sup>-1</sup> sur 47 mutations. C'est pour l'instant un peu moins bon qu'avec le modèle de Solvant CASA (0.56 kcal.mol<sup>-1</sup> sur le même échantillon) mais l'erreur est encore acceptable (figure 8.3 et 8.4).

#### **8.1.4.2 Validation sur la TyrRS de *B. stearothermophilus***

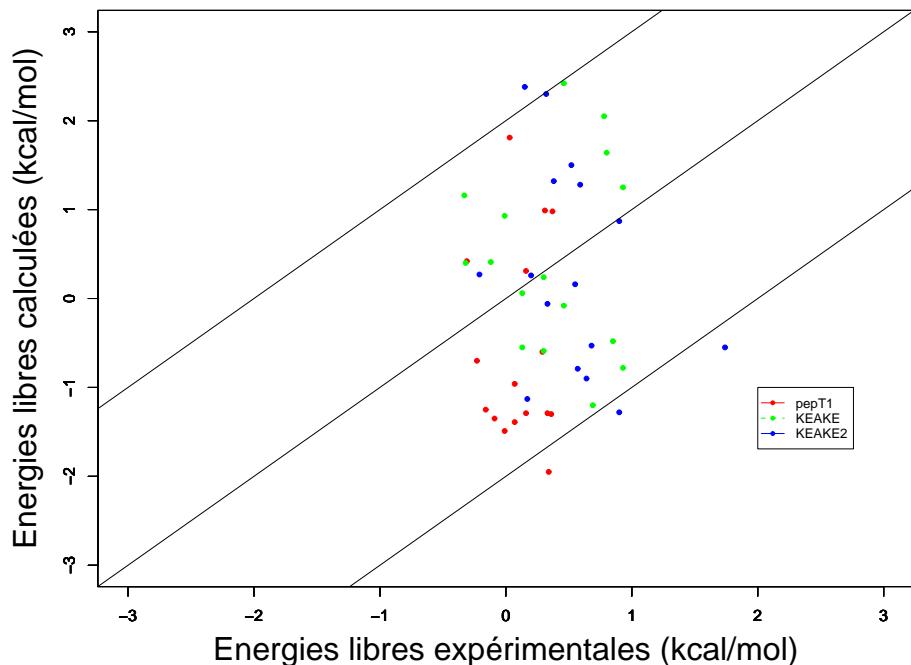
Nous disposons de données expérimentales sur 13 mutants de *Bst*-TyrRS. Il s'agit de constantes de dissociations des complexes TyrRS:Tyr natif et mutants. La même validation du protocole d'évolution dirigée est réalisée, on trouve une erreur moyenne de 1.72 kcal.mol<sup>-1</sup>. Cette erreur est un peu plus grande qu'avec le modèle CASA (0.62 kcal.mol<sup>-1</sup>).

Le protocole de calcul de la matrice d'énergie est similaire à celui décrit avec le solvant CASA. On utilise ici une constante diélectrique de 8 pour la protéine et un coefficient  $\alpha$  de -0.03 pour le terme de surface exposée au solvant. Des énergies de références sont calculées avec cette constante diélectrique. On calcule ainsi un terme d'énergie lié, un terme d'énergie non lié qui inclut les termes  $E_{self}$  et  $E_{int}$ , ainsi qu'un terme d'énergie de surface. Lors de la reconstruction, on ne calcule que les énergies d'interactions protéine-solvant

TABLE 8.3 – Energies libres calculées et expérimentales pour les trois peptides hélicoïaux.

Peptide s Mutation	$\Delta G$ calculé en kcal.mol <sup>-1</sup>	$\Delta G$ expérimental en kcal.mol <sup>-1</sup>
pepT1		
ALA	0.42	-0.31
ARG	0.31	0.16
ASN	0.99	0.31
ASP	0.98	0.37
GLN	-1.49	-0.01
HIE	-1.3	0.36
ILE	1.81	0.03
LEU	-1.35	-0.09
LYS	-0.96	0.07
MET	-1.25	-0.16
PHE	-0.6	0.29
SER	-1.29	0.16
THR	-1.95	0.34
TRP	-0.7	-0.23
TYR	-1.39	0.07
VAL	-1.29	0.33
K2AE2		
ALA	0.27	-0.21
ARG	0.26	0.2
ASN	1.32	0.38
ASP	2.38	0.15
GLN	-0.9	0.64
HIE	-0.79	0.57
ILE	2.3	0.32
LEU	0.87	0.9
LYS	1.5	0.52
MET	-1.13	0.17
PHE	-0.55	1.74
SER	-1.28	0.9
THR	-0.53	0.68
TRP	-0.06	0.33
TYR	0.16	0.55
VAL	1.28	0.59
KEAKE		
ALA	0.4	-0.32
ARG	1.16	-0.33
ASN	1.64	0.8
ASP	2.05	0.78
GLN	0.24	0.3
HIE	-0.48	0.85
ILE	2.42	0.46
LEU	0.93	-0.01
LYS	0.41	-0.12
MET	0.06	0.13
PHE	-0.55	0.13
SER	-1.2	0.69
THR	-0.78	0.93
TRP	-0.59	0.3
TYR	-0.08	0.46
VAL	1.25	0.93

FIGURE 8.4 – Comparaison des données expérimentales et des données modélisées avec le nouveau traitement du solvant sur les trois peptides hélicoïaux.

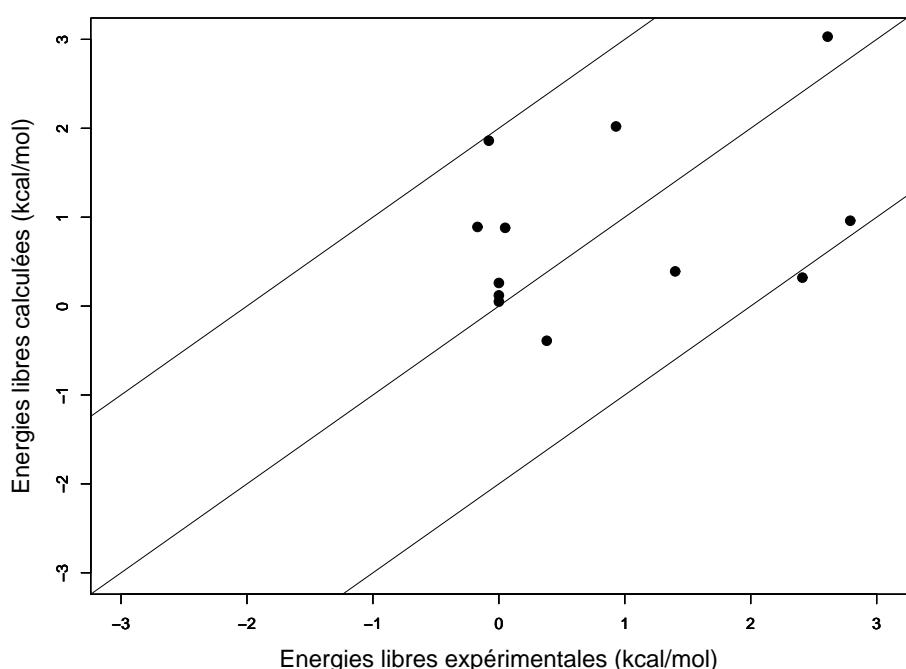


pour estimer l'affinité (figure 8.4 et 8.5). Les énergies d'interaction protéine-protéine et ligand-ligand sont négligées. Un contrôle a été réalisé sur ce protocole, en optimisant la structure d'un mutant, "identité", T234T, pour lequel la variation d'énergie libre de liaison est nulle. La structure est optimisée selon le protocole donné. Nous estimons alors la différence d'énergie libre de liaison à  $0.05 \text{ kcal} \cdot \text{mol}^{-1}$ , ce qui est proche de zéro.

TABLE 8.4 – Energies libre calculées et expérimentales pour 13 mutants de TyrRS de *B. stearothermophilus*.

Mutations	$\Delta\Delta G$ calculé en kcal.mol <sup>-1</sup>	$\Delta\Delta G$ expérimental en kcal.mol <sup>-1</sup>
Y34F	0.39	1.40
D38A	0.32	2.41
T51A	0.12	0.00
D78A	3.03	2.61
N123A	2.02	0.93
N123D	0.04	3.33
Y169F	0.96	2.79
Q173A	-5.35	2.31
K230A	-0.39	0.38
F231L	0.26	0.00
K233A	1.86	-0.08
T234A	0.89	-0.17
T234S	0.88	0.05
contrôle		
T234T	0.05	0.00

FIGURE 8.5 – 13 mutants de *Bst*-TyrRS. Comparaison des données expérimentales et des données modélisées avec le nouveau traitement du solvant.



### **8.1.5 Conclusion**

Bien que les erreurs moyennes par rapport aux données expérimentales soit un peu plus importantes avec le modèle de solvant GB-HCT qu'avec le modèle CASA, elles sont du même ordre. Les interactions électrostatiques semblent cependant mieux modélisées avec le solvant GB-HCT. Ceci porte à croire que les séquences prédictes par un programme d'évolution dirigée utilisant ce dernier modèle de solvant pour le calcul de la matrice d'énergie conduirait à une meilleure prédiction de séquence.



# Chapitre 9

## Remerciements

Je remercie Thomas Simonson d'avoir accepté de m'encadrer. Je le remercie pour ce qu'il m'a apris sur la démarche scientifique et pour tous ces conseils judicieux qui ont donné de la cohérence au projet. Enfin, je le remercie pour sa patience, et son soutien tout au long de la thèse, qui m'ont poussés à ne pas baisser les bras . Il me reste encore beaucoup à apprendre de lui. J'ai déjà changé dans ma manière d'aborder une question scientifique grâce à lui : moins de précipitation, plus d'organisation et d'analyse. Je n'ai plus peur de voir quelque chose dans le détail, et surtout de passer du temps sur la phase de préparation du système !

Je remercie également les rapporteurs David Perahia et Yves-Henry Sanejouand ainsi que les examinateurs Catherine Etchebest et Isabelle Andre, pour le temps consacré à l'évaluation de ce travail.

Je remercie Gorgio Archontis et son doctorant Savvas Polydorides pour le travail que nous avons mené en équipe sur la modélisation du solvant GB.

Je remercie particulièrement l'équipe de bioinformatique du laboratoire de Biochimie. Tout d'abord Merci à Anne Lopes, qui m'a consacré beaucoup de temps lorsque j'étais en stage M2. Merci à David Mignon pour son aide précieuse en informatique, et sur l'utilisation du programme Proteus. Merci à Alexey Aleksandrov, Thomas Gaillard, Priyadarshi Satpati et Guillaume Launay pour leurs conseils en informatique, Merci également à Alfonso Jaramillo, avec qui j'ai encadré le TP de bioinformatique du M2 ISFB et merci à ses doctorants de l'époque : Pablo Tortosa et Maria Suarez. Merci également aux nouveaux collègues Audrey Sedano et Seydou Traore.

Je remercie Yves Mechulam, Pierre Plateau, Emmanuelle Schmidt, Thomas Simonson, Alfonso Jaramillo et Anne Lopes, pour leurs enseignements dans le cadre du Master ISFB, qui m'ont donné envie de poursuivre en stage M2 dans ce laboratoire. Je les remercie également pour leur sympathie et leur accueil dans le laboratoire. Je remercie également Romary Perrin, Solène Macquere, Michel Fromant, Marc Dauplais, Pierre-Damien Coureux, Caroline Aubard, Sandra Wydau, Michel Panvert, Laura Fioni, Catherine Ferrer-Teillet, Myriam Lazard, Laurent Guillot, Laurent Giorgi, Marie Naveau et Sylvain Blanquet.

Je remercie Pierre Plateau et Yves Mechulam de m'avoir permis de participer aux oraux de Modex de biologie, qui ont été très formateurs pour moi, et merci à Catherine Morais pour leur organisation.

Je remercie également Cristine Ferret, Fabrice Baronnet, Claudette Dessertaine et Audrey Lemarechal et Michel Rosso de l'école doctorale pour la vision humaine de leur travail.

Je remercie Pascale et Pascal Delplanche, Christine Lazenec-Schurdevin, Annick Car-

pentier, pour les midis ensembles et je remercie Alain Truong et Guillaume Heugas pour le Viet-vo-dao. Merci à mes amies Christine Bathelt et Salima Hamichi. Merci à Mélanie et Julien Monguillon, pour leur bonne humeur communicative et leur amitié.

Merci à mes amis d'enfance, Suzanne Dos-Santos, Jessica Hygounet, merci également à Edouard Martins, qui a fini sa thèse avant moi, et à Olivier Bernard. Merci aux musiciens du jeudi !

Merci à ma belle famille pour sa présence et pour son aide : Maryse, Etienne et Claire Ged ainsi que Marie et Cédric Lafont.

Merci à ma famille. Merci à mon père d'avoir tant travaillé toute sa vie pour que je puisse faire des études, pour son courage, sa force et son soutien. Merci à mes grandes soeurs Rachida ma petite mère et Fatima, pour leur soutien et leur aide, merci à mon frère ahmed pour son innocence et son sourire. Merci à ma mère pour l'éducation qu'elle m'a donnée et tous ce qu'elle m'a apporté, notamment son goût pour les sciences. J'espère que de là où elle est, elle sera fière de moi, en tout cas, moi, je pense à elle.

Merci à mon mari Guillaume Ged. J'espère qu'un jour nous nous souviendrons en riant franchement de cette période ou nous rédigions nos mémoires en même temps. Merci de m'avoir supportée pendant la rédaction. Merci de m'avoir donné le plus beau cadeau du monde, Jalil. Pour terminer ces remerciements, merci à Jalil pour tous le bonheur qu'il m'apporte.

# Bibliographie

- [1] M. NIRENBERG et P. LEDER, RNA codewords and protein synthesis. the effect of trinucleotides upon the binding of sRNA to ribosomes, *Science*, vol. 145, p. 1399–1407, 1964.
- [2] H. MATTHAEI et M. NIRENBERG, The dependence of cell-free protein synthesis in *E. coli* upon RNA prepared from ribosomes, *Biochemical and Biophysical Research Communications*, vol. 4, p. 404–408, 1961.
- [3] F. CRICK, « codon–anticodon pairing : the wobble hypothesis. », *J. Mol. Biol.*, vol. 19, p. 548–555, 1966.
- [4] K. MILLER et J. LEVINE, Biology, *Prentice Hall*, p. 1146, 2007.
- [5] P. KEELING et W. DOOLITTLE, A non-canonical genetic code in an early diverging eukaryotic lineage, *EMBO J.*, vol. 15, p. 2285–2290, 1996.
- [6] L. DANIELS, Selenium metabolism and bioavailability, *Biological Trace Element Research*, vol. 54, p. 185–199, 1996.
- [7] M. SANTOS, T. UEDA, K. WATANABE et M. TUITE, The non-standard genetic code of *Candida spp.* : An evolving genetic code or a novel mechanism for adaptation ?, *Molecular Microbiology*, vol. 26, p. 423–431, 1997.
- [8] S. KANAYA, Y. YAMADA, M. KINOUCHI, Y. KUDO et T. IKEMURA, Codon usage and tRNA genes in eukaryotes : Correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis, *J. Mol. Evol.*, vol. 53, p. 290–298, 2001.
- [9] F. CRICK, "on protein synthesis.", *Symposium of the Society for Experimental Biology*, vol. 31, p. 3–9, 1966.
- [10] P. SCHIMMEL, R. GISGET, D. MORAT et S. YOKOYAMA, An operational RNA code for amino acids and possible relationship to genetic code, *Proc. Natl. Acad. Sci. USA*, vol. 90, p. 8765–8768, 1993.
- [11] R. THIEBE, K. HARBERS et H. ZACHAU, Aminoacylation of fragment combinations from yeast tRNA<sup>Phe</sup>, *Eur. J. Biochem*, vol. 26, p. 144–152, 1972.
- [12] M. HOAGLAND, An enzymatic mechanism for amino acid activation in animal, *Biochem. Biophys. Acta*, 1955.
- [13] P. BERG et E. OFENBGAND, An enzymatic mechanism for linking amino acids to RNA, *Biochem. Biophys. Acta*, vol. 44, 1958.
- [14] F. BERGMANN, P. BERG et M. DIECKMANN, The enzymic synthesis of amino acyl derivatives of ribonucleic acid, *J. Biol. Chem.*, vol. 236, juin 1961.
- [15] R. LOFTFIELD et D. VANDERJAGT, The frequency of errors in protein biosynthesis, *Biochem. J.*, vol. 128, p. 1353–1356, 1972.

- [16] R. SANKARANARAYANAN, A. DOCK-BREGEON, B. REES, M. BOVEE, J. CAILLET, P. ROMBY, C. FRANCKLYN et D. MORAS, Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase, *Nat. Struct. Biol.*, vol. 7, p. 461–465, 2000.
- [17] L. R. de POUPLANA et P. SCHIMMEL, Aminoacyl-tRNA synthetases : potential markers of genetic code development, *Trends Biochem. Sci.*, vol. 26, p. 591–596, 2001.
- [18] R. GESLAINA et L. R. de POUPLANA, Regulation of RNA function by aminoacylation and editing ?, *Trends in Genetics*, vol. 20, p. 604–610, 2004.
- [19] T. CREPIN, A. YAREMCHUK, M. TUKALO et S. CUSACK, Structures of two bacterial prolyl-tRNA synthetases with and without a cis-editing domain, *Structure*, vol. 14, p. 1511–1525, 2006.
- [20] M. BUDDHA et B. CRANE, Structure and activity of an aminoacyl-trna synthetase that charges trna with nitro-tryptophan, *Nature Structural and Molecular Biology*, vol. 12, p. 274–275, 2005.
- [21] J. KERS, M. WACH, S. KRASNOFF, J. WIDOM, K. CAMERON, R. BUKHALID, D. GIBSON, B. CRANE et R. LORIA, Nitration of a peptide phytotoxin by bacterial nitric oxide synthase, *Nature*, vol. 6, p. 79–82, 2004.
- [22] M. SCHEER, A. GROTE, A. CHANG, I. SCHOMBURG, C. MUNARETTO, M. ROTHER, C. SÖHNGEN, M. STELZER, J. THIELE et D. SCHOMBURG, BRENDa, the enzyme information system in 2011, *Nucl. Acids Res.*, vol. 39, p. 670–676, 2011.
- [23] D. BLOW, T. BAHT, A. METCALFE, J. RISLER, S. BRUNIE et C. ZELWER, Structural homology in the amino-terminal domains of two aminoacyl-tRNA synthetases, *J. Mol. Biol.*, vol. 171, p. 571–576, 1983.
- [24] T. WEBSTER, H. TSAI, M. KULA, G. MACKIE et P. SCHIMMEL, Specific sequence homology and three-dimensional structure of an aminoacyl-tRNA synthetase, *Science*, vol. 226, p. 1315–1317, 1984.
- [25] C. HOUNTONDJI, P. DESSEN et S. BLANQUET, Sequence similarities among the family of aminoacyl-tRNA synthetases, *Biochimie*, vol. 68, p. 1071–1094, 1986.
- [26] G. ERIANI, M. DELARUE, O. POCH, J. GANDLOFF et D. MORAS, Partition of aminoacyl-tRNA synthetase into two classes on the basis of two mutually exclusive sets of sequence motifs, *Nature*, vol. 347, p. 203–206, 1990.
- [27] R. LEATHERBARROW, A. FERSHT et G. WINTER, Transition state stabilisation in the mecanism of tyrosyl-tRNA synthetase revealed by protein engineering, *Proc. Natl. Acad. Sci. USA*, vol. 82, p. 7840–7844, 1985.
- [28] S. CUSACK, M. HÄRTLEIN et R. LEBERMAN, Sequence, structure and evolutionary relationships between classe 2 aminoacyl-tRNA synthetase, *Nucleic Acids Research*, vol. 19, p. 3489–3498, 1991.
- [29] L. R. de POUPLANA et P. SHIMMEL, Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem, *Cell*, vol. 104, p. 191–193, 2001.
- [30] S. RODIN et S. OHNO, Two types of aminoacyl-tRNA synthetases could be originally encoded by complementary strands of the same nucleic acid, *Orig. Life. Evol. Biosph.*, vol. 25, p. 565–589, 1995.
- [31] G. NAGEL et R. DOOLITTLE, Evolution and relatedness in two aminoacyl-tRNA synthetase families, *Proc. Natl. Acad. Sci. USA*, vol. 88, p. 8121–8125, 1991.

- [32] L. KLIPCAN et M. SAFRO, Amino acid biogenesis, evolution of the genetic code and aminoacyl-tRNA synthetases, *Journal of Theoretical Biology*, vol. 228, p. 389–396, 2004.
- [33] P. SCHIMMEL, R. GIEGÉ, D. MORAS et S. YOKOYAMA, An operational RNA code for amino acids and possible relationship to genetic code, *Proc. Natl. Acad. Sci. USA*, vol. 90, p. 8763–8768, 1993.
- [34] H. BECKER et D. KERN, *Thermus thermophilus* - a link in evolution of the tRNA-dependent amino acid amidation pathways, *Proc. Natl. Acad. Sci. USA*, vol. 95, p. 12832–12837, 1998.
- [35] H. BECKER, J. REINBOLT, R. KREUTZER, R. GIEGÉ et D. KERN, Existence of two distinct aspartyl-tRNA synthetases in *Thermus thermophilus*. structural and biochemical properties of the two enzymes, *Biochemistry*, vol. 36, p. 8785–8797, 1997.
- [36] C. FRANKLYN, tRNA synthetase paralogs : evolutionary links in the transition from tRNA-dependent amino acid biosynthesis to de novo biosynthesis, *Proc. Natl. Acad. Sci. USA*, vol. 100, p. 9650–9652, 2003.
- [37] M. IBBA, A. CURNOW et D. SÖLL, Aminoacyl-tRNA synthesis : divergent routes to a common goal, *Trends Biochem. Sci.*, vol. 22, p. 39–42, 1997.
- [38] S. PARK, K. EWALT et S. KIM, Functional expansion of aminoacyl-tRNA synthetases and their interacting factors : new perspectives on housekeepers, *Trends Biochem. Sci.*, vol. 30, p. 569–574, 2005.
- [39] P. PLATEAU, M. FROMANT, A. BREVET et S. BLANQUET, Properties of homogeneous AP4A-specific dinucleotide tetraphosphatase from *Escherichia coli*, *Hoppe-Seylers Zeitschrift fur Physiologische Chemie*, vol. 365, p. 602–602, 1984.
- [40] T. FRASER et A. RICH, Amino acids are not all initially attached to the same position on tRNA molecules, *Proc. Natl. Acad. Sci. USA*, vol. 8, p. 3044–3048, 1975.
- [41] M. SPRINZL et F. CRAMER, Site of aminoacylation of tRNA from *Escherichia coli* with respect to the 2'- or 3'-hydroxyl group of the terminal adenosine, *Proc. Natl. Acad. Sci. USA*, vol. 72, p. 3049–3053, 1975.
- [42] V. GUEZ-IVANIER, M. HERMANN et D. BALDWIN, Mapping the stability determinants of bacterial tyrosyl-tRNA synthetases by an experimental evolutionary approach, *J. Mol. Biol.*, vol. 234, p. 209–221, 1993.
- [43] P. CARTER, H. BEDOUELLE et G. WINTER, Construction of heterodimer tyrosyl-tRNA synthetase shows tRNA<sup>Tyr</sup> interacts with both subunits, *Proc. Natl. Acad. Sci. USA*, vol. 83, p. 1189–1192, 1986.
- [44] H. BEDOUELLE et G. WINTER, A model of synthetase-tRNA interaction as deduced by protein engineering, *Nature*, vol. 320, p. 371–373, 1986.
- [45] A. FERSHT, Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis, *Biochemistry*, vol. 26, p. 8031–8037, 1987.
- [46] E. FIRST et A. FERSHT, Mutational and kinetic analysis of a mobile loop in tyrosyl-tRNA synthetase, *Biochemistry*, vol. 32, p. 13658–13663, 1993.
- [47] P. BRICK, T. BATH et D. BLOW, Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. interaction of the enzyme with the tyrosyl-adenylate intermediate, *J. Mol. Biol.*, vol. 208, p. 83–98, 1989.

- [48] G. SHARMA et E. FIRST, Thermodynamic analysis reveals a temperature-dependent change in the catalytic mechanism of *Bacillus stearothermophilus* tyrosyl-tRNA synthetase, *J. Biol. Chem.*, vol. 284, p. 4179–90, 2008.
- [49] L. BONNEFOND, R. GIEGE et J. RUDINGER-THIRION, Evolution of the tRNA<sup>Tyr</sup>/TyrRS aminoacylation systems, *Biochemistry*, vol. 87, p. 873–883, 2005.
- [50] J. SHERMAN, K. ROGERS, M. ROGERS et D. SÖLL, Synthetase competition and tRNA context determine the *in vivo* identity of tRNA discriminator mutants, *J. Mol. Biol.*, vol. 228, p. 1055–1062, 1992.
- [51] Y. HOU et P. SCHIMMEL, Modeling with in vitro kinetic parameters for the elaboration of tRNA identity *in vivo*, *Biochemistry*, vol. 28, p. 4942–4947, 1989.
- [52] H. BEDOUELLE, V. GUEZ et A. VIDAL-CROS, Overproduction of tyrosyl-tRNA synthetase is toxic to *Escherichia coli* : a genetic analysis, *J Bacteriol*, vol. 172, p. 3940–3945, 1990.
- [53] Y. WOLF, L. ARAVIND, N. GRISHIN et E. KOONIN, Evolution of aminoacyl-tRNA synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events, *Genome Res*, vol. 9, p. 689–710, 1999.
- [54] T. KLEEMAN, D. WEI, K. SIMPSON et E. FIRST, Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine, *J. Biol. Chem.*, vol. 272, p. 14420–14425, 1997.
- [55] M. TSUNODA, Y. KUSAKABE, N. TANAKA, S. OHNO, M. NAKAMURA, T. SENDA, T. MORIGUCHI, N. ASAI, M. SEKINE, T. YOKOGAWA, K. NISHIKAWA et K. NAKAMURA, Structural basis for recognition of cognate tRNA by tyrosyl-tRNA synthetase from three kingdoms, *Nucl. Acids Res.*, vol. 35, p. 4289–4300, 2007.
- [56] J. BERG, J. TYMOCZKO et L. STRYER, Biochemistry, *W.H. Freeman*, vol. 5th edition, 2002.
- [57] Z. KUCAN et R. CHAMBERS, Purification of tyrosine : tRNA ligase from *saccharomyces cerevisiae alphaS288C*, *J. Biochem.*, vol. 73, p. 811–819, 1973.
- [58] G. D. P. GAY, H. DUCKWORTH et A. FERSHT, Modification of the amino acid specificity of tyrosyl-tRNA synthetase by protein engineering, *FEBS Letters*, vol. 318, p. 167–171, 1993.
- [59] S. CHOUSTERMAN et F. CHAPEVILLE, Tyrosyl-tRNA synthetase of *Escherichia coli*. Role of magnesium ion in reaction catalyzed by the enzyme, *Eur. J. Biochem*, vol. 35, p. 46–50, 1973.
- [60] R. HYDE-DERUYSCHER, L. PAIGE, D. CHRISTENSEN, N. HYDE-DERUYSCHER, A. LIM, Z. FREDERICKS, J. KRANZ, P. GALLANT, J. ZHANG, S. ROCKLAGE, D. FOWLKES, P. WENDLER et P. HAMILTON, Detection of small-molecule enzyme inhibitors with peptides isolated from phage-displayed combinatorial peptide libraries, *Chem. Biol.*, vol. 7, p. 17–25, 2000.
- [61] A. KOWAL, C. KOHRER et U. RAJBHANDARY, Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria, *Proc. Natl. Acad. Sci. USA*, vol. 98, p. 2268–2273, 2001.
- [62] J. XIE et P. SCHULTZ, A chemical toolkit for proteins - an expanded genetic code, *Nature Reviews. Molecular Cell Biology*, vol. 7, p. 775–782, 2006.

- [63] K. WANG, H. NEUMANN, S. PEAK-CHEW et J. CHIN, Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion, *Nature Biotechnology*, vol. 25, p. 770–777, 2007.
- [64] S. OUE, A. OKAMOTO, T. YANO et H. KAGAMIYAMA, Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues, *J. Biol. Chem.*, vol. 274, p. 2344–2349, 1999.
- [65] W. XIE et N. SAHINIDIS, Residue-rotamer-reduction algorithm for the protein side-chain conformation problem, *BMC Bioinf.*, vol. 22, p. 188–194, 2006.
- [66] F. HAMANO-TAKAKU, T. IWAMA et S. SAITO-YANO, A mutant *Escherichia coli* tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine, *J. Biol. Chem.*, vol. 275, p. 40324–40328, 2000.
- [67] T. KOBAYASHI, K. SAKAMOTO, T. TAKIMURA, R. SEKINE, K. VINCENT, K. KAMATA, S. NISHIMURA et S. YOKOYAMA, Structural basis of nonnatural amino acid recognition by an engineered aminoacyl-tRNA synthetase for genetic code expansion, *Proc. Natl. Acad. Sci. USA*, vol. 102, p. 1366–1371, 2005.
- [68] J. SOUTOURINA et P. PLATEAU, Les acides aminés de la série D dans la traduction, *Travaux Universitaires - Thèse de doctorat*, 2000.
- [69] P. BRIZA, G. WINKLER, H. KALCHHAUSER et M. BREITENBACH, Dityrosine is a prominent component of the yeast ascospore wall : a proof of its structure, *J. Biol. Chem.*, vol. 261, p. 4288–4294, 1986.
- [70] A. MOR, M. AMICH et P. NICOLAS, Enter a new post-translational modification : D-amino acids in gene-encoded peptides, *Trends Biochem. Sci.*, vol. 17, p. 481–485, 1992.
- [71] G. KREIL, D-amino acids in animal peptides, *Ann. Rev. Biochem.*, vol. 66, p. 337–345, 1997.
- [72] R. JACK et G. JUNG, Natural peptides with antimicrobial activity, *Chimia*, vol. 52, p. 48–55, 1998.
- [73] K. KURAHASHI, M. YAMADA, K. MORI, K. FUJIKAWA, M. KAMBE, Y. IMAE, E. SATO, H. TAKAHASHI et Y. SAKAMOTO, Biosynthesis of cyclic oligopeptide, *Cold Spring Harbor Symp. Quant. Biol.*, vol. 34, p. 815–826, 1969.
- [74] F. PEYPOUX, F. BESSON, G. MICHEL et L. DELCAMBE, Structure of bacillomycin D, a new antibiotic of the iturin group, *Eur. J. Biochem.*, vol. 118, p. 323–327, 1981.
- [75] G. KREIL, Peptides containing a D-amino acid from frogs and molluscs, *J. Biol. Chem.*, vol. 269, p. 10967–10970, 1994.
- [76] R. VOLKMANN et S. HECK, Biosynthesis of D-amino acid-containing peptides : exploring the role of peptide isomerases, *EXS*, vol. 85, p. 87–105, 1998.
- [77] G. MIGNOGNA, M. SIMMACO et D. BARRA, Occurrence and function of D-amino acid-containing peptides and proteins : antimicrobial peptides, *EXS*, vol. 85, p. 29–36, 1998.
- [78] Y. KAMATANI, H. MINAKATA, P. KENNY, T. IWASHITA, K. WATANABE, K. FUNASE, X. SUN, A. YONGSIRI, K. KIM, P. NOVALES-LI, E. NOVALES, C. KANAPI, H. TAKEUCHI et K. NOMOTO, Achatin-i, an endogenous neuroexcitatory tetrapeptide from *Achatina fulica* Férrussac containing a D-amino acid residue, *Biochem. Biophys. Res. Commun.*, vol. 160, p. 1015–1020, 1989.

- [79] K. FUJIMOTO, I. KUBOTA, Y. YASUDA-KAMATANI, K. NOMOTO, M. YOSHIDA, A. HARADA, Y. MUNEKOA et M. KOBAYASHI, Purification of achatin-I from the atria of the african giant snail, achatina fulica, and its possible function, *Biochem. Biophys. Res. Commun.*, vol. 177, p. 847–853, 1991.
- [80] F. MORISHITA, Y. NAKANISHI, S. KAKU, Y. FURUKAWA, S. OHTA, T. HIRATA, M. OHTANI, Y. FUJISAWA, Y. MONEOKA et O. MATSUSHIMA, A novel D-amino acid-containing peptide isolated from aplysia heart, *Biochem. Biophys. Res. Commun.*, vol. 240, p. 354–358, 1997.
- [81] E. JIMENEZ, B. OLIVERA, W. GRAY et L. CRUZ, Contryphan is a D-tryptophan-containing Conus peptide, *J. Biol. Chem.*, vol. 271, p. 28002–28005, 1996.
- [82] D. SOYER, F. V. HERP, J. ROSSIER, J. L. CAER, C. TENSEN et R. LAFONT, Evidence for a conformational polymorphism of invertebrate neurohormones. D-amino acid residue in crustacean hyperglycemic peptides, *J. Biol. Chem.*, vol. 269, p. 18295–18298, 1994.
- [83] A. YASUDA, Y. YASUDA, T. FUJITA et Y. NAYA, Characterization of crustacean hyperglycemic hormone from the crayfish *Procambarus clarkii* : multiplicity of molecular forms by stereoinversion and diverse functions, *Gen. Comp. Endocrin.*, vol. 95, p. 387–398, 1994.
- [84] S. HECK, C. SIOK, K. KRAPCHO, P. KELBAUGH, P. THADEIO, M. WELCH, R. WILLIAMS, A. GANONG, M. KELLY, A. LANZETTI, W. GRAY, D. PHILLIPS, T. PARKS, H. JACKSON, M. AHLIJANIAN, N. SACCOMANO et R. VOLKMANN, Functional consequences of posttranslational isomerization of ser46 in a calcium channel toxin, *Science*, vol. 266, p. 1065–1068, 1994.
- [85] A. ALLEN et H. ROSENBERG, The biosynthesis of D-serine ethanolamine phosphate in the earthworm megascolides cameroni, *Biochem. Biophys. Acta*, vol. 152, p. 208–210, 1968.
- [86] H. SELIGER, W. McELROY, E. WHITE et G. FIELD, Stereospecificity and firefly bioluminescence, a comparison of natural and synthetic luciferins, *Proc. Natl. Acad. Sci. USA*, vol. 47, p. 1129–1134, 1981.
- [87] J. PARK, Murein synthesis in *Escherichia coli* and *Salmonella typhimurium* : cellular and molecular biology, *American Society for Microbiology, Washington, D. C*, p. 663–671, 1987.
- [88] R. PRESTON, Occurrence of D-amino acids in higher organisms : a survey of the distribution of D-amino acids in marine invertebrates, *Comp. Biochem. Physiol.*, vol. 87B, p. 55–62, 1987.
- [89] D. ARMSTRONG, M. GASPER, S. LEE, J. ZUKOWSKI et N. ERCAL, D-amino acid levels in human physiological fluids, *Chirality*, vol. 5, p. 375–378, 1993.
- [90] K. HAMASE, H. HOMMA, Y. TAKIGAWA, T. FUKUSHIMA, T. SANTA et K. IMAI, Regional distribution and postnatal changes of D-amino acids in rat brain, *Biochem. Biophys. Acta*, vol. 1334, p. 214–222, 1997.
- [91] A. D'ANIELLO, d-aspartic acid : An endogenous amino acid with an important neuroendocrine role, *Brain Research Reviews*, vol. 53, p. 215–234, 2007.
- [92] H. WOLOSKER, K. SHETH, M. TAKAHASHI, J. MOTHET, R. BRADY, C. FERRIS et S. SNYDER, Purification of serine racemase :biosynthesis of the neuromodulator d-serine, *Proc. Natl. Acad. Sci. USA*, vol. 96, p. 721–725, 1999.

- [93] Y. TSAI, C. TSENG, K. HSIAO et L. CHEN, Production and purification of D-aminoacylase from *Alcaligenes denitrificans* and taxonomic study of the strain, *Appl. Environ. Microbiol.*, vol. 54, p. 984–989, 1988.
- [94] K. IMAI, T. FUKUSHIMA, T. SANTA, H. HOMMA, Y. HUANG, K. SAKAI et M. KATO, Distribution of free D-amino acids in tissues and body fluids of vertebrates, *Enantiomer*, vol. 2, p. 143–145, 1997.
- [95] C. WHITTINGTON, J. KOH, W. WARREN, A. PAPENFUSS, A. TORRES, P. KUCHEL et K. BELOV, Understanding and utilising mammalian venom via a platypus venom transcriptome, *Journal of Proteomics*, vol. 72, p. 155–164, 2009.
- [96] R. CALENDAR et P. BERG, Purification and physical characterization of tyrosyl ribonucleic acid synthetases from *Escherichia coli* and *Bacillus subtilis*, *J. Mol. Biol.*, vol. 26, p. 39–54, 1967.
- [97] A. SHEORAN et E. FIRST, Activation of D-tyrosine by *Bacillus stearothermophilus* tyrosyl-tRNA synthetase - 1. Pre-steady-state kinetic analysis reveals the mechanistic basis for the recognition of D-tyrosine, *J. Biol. Chem.*, vol. 283, p. 12971–12980, 2008.
- [98] R. CALENDAR et P. BERG, Purification and physical characterization of tyrosyl ribonucleic acid synthetases from *Escherichia coli* and *Bacillus subtilis*, *Biochemistry*, vol. 5, p. 1690 –1695, 1966.
- [99] T. LI, M. FROEYEN et P. HERDEWIJN, Comparative structural dynamics of tyrosyl-tRNA synthetase complexed with different substrates explored by molecular dynamics, *European Biophysics Journal with Biophysics Letters.*, vol. 38, p. 25–35, 2008.
- [100] D. THOMPSON, C. LAZENNEC, P. PLATEAU et T. SIMONSON, Ammonium scanning in an enzyme active site : The chiral specificity of aspartyl-tRNA synthetase, *J. Biol. Chem.*, vol. 282, p. 30858–30870, 2007.
- [101] L. SERRANO, A. DAY et A. FERSHT, Step-wise mutation of barnase to binase. a procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability, *J. Mol. Biol.*, vol. 233, p. 305–312, 1993.
- [102] M. PANTOLIANO, M. WHITLOW, J. WOOD, M. ROLLENCE, B. FINZEL, G. GILLILAND, T. POULOS et P. BRYAN, The engineering of binding affinity at metal ion binding sites for the stabilization of proteins : subtilisin as a test case, *Biochemistry*, vol. 27, p. 8311–8317, 1988.
- [103] P. BRYAN, Protein engineering of subtilisin, *Biochem. Biophys. Acta*, vol. 1543, p. 203–222, 2000.
- [104] F. DONG, B. OLSEN et N. BAKER, Computational methods for biomolecular electrostatics, *Methods Cell Biol.*, 2008.
- [105] B. BROOKS, R. BRUCCOLERI, B. OLAFSON, D. STATES, S. SWAMINATHAN et M. KARPLUS, CHARMM : a program for macromolecular energy, minimization, and molecular dynamics calculations, *J. Comp. Chem.*, 1983.
- [106] W. CORNELL, P. CIEPLACK, C. BAYLY, I. GOULD, K. MERZ, D. FERGUSON, D. SPELLMEYER, T. FOX, J. CALDWELL et P. KOLLMAN, A second force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.*, 1995.
- [107] W. JORGENSEN et J. T. RIVES, The OPLS force field for proteins. energy minimizations for crystals of cyclic peptides and crambin, *J. Am. Chem. Soc.*, vol. 110, p. 1657–1666, 1988.

- [108] M. CHRISTEN, P. HÜNENBERGER, D. BAKOWIES, R. BARON, R. BÜRG, D. GEERKE, T. HEINZ, M. KASTENHOLZ, V. KRÄUTLER, C. OOSTENBRINK, C. PETER, D. TRZESNIAK et W. van GUNSTEREN, The GROMOS software for biomolecular simulation : Gromos05, *J. Comp. Chem.*, vol. 16, p. 1719–1751, 2005.
- [109] G. NEMETHY, M. POTTE et H. SCHERAGA, Energy parameters in polypeptides. 9. updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids, *J. Phys. Chem.*, vol. 87, p. 1883–1887, 1983.
- [110] A. WARSHEL, Quantum mechanical consistent force field (QCFF/PI) method : Calculations of energies, conformations and vibronic interactions of ground and excited states of conjugated molecules, *Israel J. Chem.*, vol. 11, p. 709, 1973.
- [111] A. HAGLERAND, E. HULER et S. LIFSON, Energy functions for peptides and proteins. i. derivation of a consistent force field including the hydrogen bond for amide crystals, *J. Am. Chem. Soc.*, vol. 96, p. 5319–5327, 1973.
- [112] J. PONDER, C. WU, P. REN, V. PANDE, J. CHODERA, M. SCHNIEDERS, I. HAQUE, D. MOBLEY, D. LAMBRECHT, R. DiSTASIO, M. HEAD-GORDON, G. CLARK, M. JOHNSON et T. HEAD-GORDON, Current status of the AMOEBA polarizable force field, *J. Phys. Chem. B*, vol. 114, p. 2549–2564, 2010.
- [113] T. KORTEMME, A. MOROZOV et D. BAKER, An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes, *J. Mol. Biol.*, vol. 326, p. 1239–1259, 2002 ;
- [114] N. POKALA et T. HANDEL, Energy functions for protein design : adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity, *J. Mol. Biol.*, vol. J347, p. 203–227, 2005.
- [115] E. ZOLLARS, S. MARSHALL et S. MAYO, Simple electrostatic model improves designed protein sequences, *Prot. Sci.*, vol. 15, p. 2014–2018, 2006.
- [116] A. MOROZOV, J. HAVRANEK, D. BAKER et E. SIGGIA, Protein-DNA binding specificity predictions with structural models, *Nucl. Acids Res.*, vol. 33, p. 5781–5798, 2005.
- [117] R. LILIEN, B. STEVENS, A. ANDERSON et B. DONALD, A novel ensemble-based scoring and search algorithm for protein redesign and its application to modify the substrate specificity of the gramicidin synthetase a phenylalanine adenylation enzyme, *J. Comp. Chem.*, vol. 12, p. 740–761, 2005.
- [118] X. HU et B. KUHLMAN, Protein design simulations suggest that side-chain conformational entropy is not a strong determinant of amino acid environmental preferences, *Proteins*, vol. 62, p. 739–748, 2006.
- [119] K. SPIEGEL, W. DEGRADO et M. KLEIN, Structural and dynamical properties of manganese catalase and the synthetic protein DF1 and their implication for reactivity from classical molecular dynamics calculations, *Proteins*, vol. 65, p. 317–330, 2006.
- [120] V. TOZZINI, Coarse-grained models for proteins, *Current Opinion in Structural Biology*, vol. 15, p. 144–150, 2005.
- [121] A. KORKUT et W. HENDRICKSON, A force field for virtual atom molecular mechanics of proteins, *Proc. Natl. Acad. Sci. USA*, vol. 106, p. 15667–72, 2009.

- [122] G. LAUNAY, R. MENDEZ, S. WODAK et T. SIMONSON, Recognizing protein-protein interfaces with empirical potentials and reduced amino acid alphabets, *BMC Bioinf.*, vol. 8, p. 270–291, 2007.
- [123] H. ZHOU et Y. ZHOU, Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction, *Prot. Sci.*, vol. 11, p. 2714–2726, 2002.
- [124] B. JERNIGAN, Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation, *J. Mol. Biol.*, vol. 266, p. 195–214, 1997.
- [125] Z. HENDSCH et B. TIDOR, Do salt bridges stabilize proteins ? a continuum electrostatic analysis, *Prot. Sci.*, vol. 3, p. 211–226, 1994.
- [126] C. WALDBURGER, J. SCHILDBACH et R. SAUER, Are buried salt bridges important for protein stability and conformational specificity ?, *Nat. Struct. Biol.*, vol. 2, p. 122–128, 1995.
- [127] J. BERNAL, A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions, *J. Chem. Phys.*, vol. 1, p. 515, 1933.
- [128] A. MASUNOV et T. LAZARIDIS, Potentials of mean force between ionizable amino acid side chains in water, *J. Am. Chem. Soc.*, 2003.
- [129] B. ROUX et T. SIMONSON, Implicit solvent models, *Biophys. Chem.*, 1999.
- [130] B. GELIN et M. KARPLUS, Role of structural flexibility in conformational calculations. application to acetylcholine and  $\beta$ -methylacetylcholine, *J. Am. Chem. Soc.*, vol. 97, p. 6996–7006, 1995.
- [131] A. STREET et S. MAYO, Pairwise calculation of protein solvent-accessible surface areas, *Folding and Design*, 1998.
- [132] L. WESSON et D. EISENBERG, Atomic solvation parameters applied to molecular dynamics of protein in solution, *Prot. Sci.*, 1992.
- [133] J. WARWICKER et H. WATSON, Calculation of the electric potential in the active site cleft due to  $\alpha$ -helix dipoles, *J. Mol. Biol.*, vol. 157, p. 671–679, 1982.
- [134] I. KLAPPER, R. HAGSTROM, R. FINE, K. SHARP et B. HONIG, Focusing of electric fields in the active site of Cu-Zn superoxide dismutase : effects of ionic strength and amino-acid modification, *Proteins*, vol. 1, p. 47–59, 1986.
- [135] G. ARCHONTIS et T. SIMONSON, A residue-pairwise generalized born scheme suitable for protein design calculations, *J. Phys. Chem.*, p. 22667–22673, 1998.
- [136] A. LOPES, A. ALEXANDROV, C. BATHELT, G. ARCHONTIS et T. SIMONSON, Computational sidechain placement and protein mutagenesis with implicit solvent models, *Proteins : Structure, Function, and Bioinformatics*, vol. 67, p. 853–867, 2007.
- [137] A. SU et S. MAYO, Coupling backbone flexibility and amino acid sequence selection in protein design, *Prot. Sci.*, vol. 9, p. 1701–1707, 1997.
- [138] B. ALLEN et S. MAYO, An efficient algorithm for multistate protein design based on FASTER, *J. Comp. Chem.*, vol. 31, p. 904–916, 2010.
- [139] X. FU, J. APGAR et A. KEATING, Modeling backbone flexibility to achieve sequence diversity : the design of novel  $\alpha$ -helical ligands for Bcl-xL, *J. Mol. Biol.*, vol. 371, p. 1099–1117, 2007.

- [140] P. HARBURY, J. PLECS, B. TIDOR, T. ALBER et P. KIM, High-resolution protein design with backbone freedom, *Science*, p. 1462–1467, 1998.
- [141] J. DESJARLAIS et T. HANDEL, Side-chain and backbone flexibility in protein core design, *J. Mol. Biol.*, vol. 290, p. 305–318, 1999.
- [142] D. MANDELL, E. COUTSIAS et T. KORTEMME, Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling, *Nature Methods*, vol. 6, p. 551–552, 2009.
- [143] J. JANIN et S. WODAK, Conformation of amino acid side-chains in proteins, *J. Mol. Biol.*, vol. 125, p. 357–386, 1978.
- [144] J. PONDER et F. RICHARDS, Tertiary templates for proteins : use of packing criteria in the enumeration of allowed sequences for different structural classes, *J. Mol. Biol.*, vol. 193, p. 775–792, 1987.
- [145] H. KONO et J. DOI, Energy minimization method using automata network for sequence and side-chain conformation prediction from given backbone geometry, *Proteins*, vol. 19, p. 244–255, 1994.
- [146] S. LOVELL, J. WORD, J. RIDCHARDSON et D. RIDCHARDSON, Asparagine and glutamine rotamers : B-factor cutoff and correction of amide flips yield distinct clustering., *Proc. Natl. Acad. Sci. USA*, vol. 96, p. 400–405, 1999.
- [147] M. MCGREGOR, S. ISLAM et M. STERNBERG, Analysis of the relationship between sidechain conformation and secondary structure in globular proteins, *J. Mol. Biol.*, vol. 198, p. 295–310, 1987.
- [148] R. DUNBRACK et M. KARPLUS, Backbone-dependent rotamer library for proteins. application to side-chain prediction, *J. Mol. Biol.*, vol. 230, p. 543–574, 1993.
- [149] R. DUNBRACK et F. COHEN, Bayesian statistical analysis of protein sidechain rotamer preferences, *Prot. Sci.*, vol. 6, p. 1661–1681, 1997.
- [150] R. PETERSON, P. DUTTON et A. WAND, Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library, *Prot. Sci.*, vol. 13, p. 735–751, 2004.
- [151] Z. XIANG et B. HONIG, Extending the accuracy limits of prediction for side- chain conformations, *J. Mol. Biol.*, vol. 311, p. 421–430, 2001.
- [152] B. DAHIYAT et S. MAYO, Protein design automation, *Protein Science*, vol. 5, p. 895–903, 1996.
- [153] H. ZHOU, Directed test of the gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins, *J. Am. Chem. Soc.*, vol. 125, p. 2060–2061, 2003.
- [154] H. HELLINGA et F. RICHARDS, Optimal sequence selection in proteins of known structure by simulated evolution, *Proc. Natl. Acad. Sci. USA*, vol. 91, p. 5803–5807, 1994.
- [155] L. WERNISCH, S. HERY et S. WODAK, Automatic protein design with all atom force-fields by exact and heuristic optimization, *J. Mol. Biol.*, vol. 301, p. 713–736, 2000.
- [156] J. DESMET, M. D. MAYER, B. HAZES et I. LASTERS, The dead-end elimination theorem and its use in protein side-chain positioning, *Nature*, vol. 356, p. 539–542, 1992.

- [157] R. GOLDSTEIN, Efficient rotamer elimination applied to protein sidechains and related spin glasses, *Biophys. J.*, vol. 66, p. 1335–1340, 1994.
- [158] B. DAHIYAT et S. MAYO, De novo protein design : fully automated sequence selection, *Science*, vol. 278, p. 82–87, 1997.
- [159] K. LUMB et P. KIM, A buried interaction imparts structural uniqueness in a designed heterodimeric coiled coil, *Biochemistry*, vol. 34, p. 8642–8648, 1995.
- [160] X. JIANG, H. FARID, E. PISTOR et R. FARID, A new approach to the design of uniquely folded thermally stable proteins, *Prot. Sci.*, vol. 9, p. 403–416, 2000.
- [161] E. JOHNSON, G. LAZAR, J. DESJARLAIS et T. HANDEL, Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin, *Structure Fold. Des.*, vol. 7, p. 967–976, 1999.
- [162] G. LAZAR, W. DANG, S. KARKI, O. VAFA, J. PENG, L. HYUN, C. CHAN, H. CHUNG, A. EIVAZI, S. YODER, J. VIELMETTER, D. CARMICHAEL, R. HAYES et B. DAHIYAT, Engineered antibody fc variants with enhanced effector function, *Proc. Natl. Acad. Sci. USA*, vol. 103, p. 4005–4010, 2006.
- [163] B. DAHIYAT et S. MAYO, De novo protein design :fully automated sequence selection, *Science*, vol. 278, p. 82–87, 1997.
- [164] B. KUHLMAN, G. IRETON, G. VARANI, B. STODDARD et D. BAKER, Design of a novel globular protein fold with atomic-level accuracy, *Science*, vol. 302, p. 1364–1368, 2003.
- [165] B. PIERRET, H. VIRELIZIER et C. VITA, Synthesis of a metal binding protein designed on the alpha/beta scaffold of charybdotoxin, *Int. J. Pept. Protein Res.*, vol. 46, p. 471–479, 1995.
- [166] E. DRAKOPOULOU, J. VIZZAVONA et C. VITA, Engineering a CD4 mimetic inhibiting the binding of the human immunodeficiency virus-1 (HIV-1) envelope glycoprotein gp120 to human lymphocyte CD4 by the transfer of a CD4 functional site to a small natural scaffold, *Letters in Peptide Science*, vol. 5, p. 241–245, 1997.
- [167] G. SCHREIBER et A. FERSHT, Energetics of protein-protein interactions : analyse of the barnase-barnase interface by single mutations and double mutant cycles, *J. Mol. Biol.*, 1995.
- [168] D. XU, S. LIN et R. NUSSINOV, Protein binding versus protein folding : The role of hydrophilic bridges in protein associations, *J. Mol. Biol.*, 1997.
- [169] D. BASHFORD et D. CASE, Generalized born models of macromolecular solvation effects, *Ann. Rev. Phys. Chem.*, vol. 51, p. 129–152, 2000.
- [170] G. SONG, G. LAZAR, T. KORTEMME, M. SHIMAOKA, J. DESJARLAIS, D. BAKER et T. SPRINGER, Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion, *J. Biol. Chem.*, vol. 281, p. 5042–5049, 2006.
- [171] V. SOOD et D. BAKER, Recapitulation and design of protein binding peptide structures and sequences, *J. Mol. Biol.*, vol. 357, p. 917–927, 2006.
- [172] L. CLARK, P. BORIACK-SJODIN, J. ELDREDGE, C. FITCH, B. FRIEDMAN, K. HANF, M. JARPE, S. LIPAROTO, Y. LI, A. LUGOVSKOY, S. MILLER, M. RUSHE, W. SHERMAN, K. SIMON et H. V. VLIJMEN, Affinity enhancement of an /textitin vivo matured therapeutic antibody using structure-based computational design, *Prot. Sci.*, vol. 15, p. 949–960, 2006.

- [173] B. DAHIYAT et S. MAYO, Probing the role of packing specificity in protein design, *Proc. Natl. Acad. Sci. USA*, vol. 94, p. 10172–10177, 1997.
- [174] F. COCHRAN, S. WU, W. WANG, V. NANDA, J. SAVEN, M. THERIEN et W. DEGRADO, Computational de novo design and characterization of a four-helix bundle protein that selectively binds a non biological cofactor, *J. Am. Chem. Soc.*, vol. 127, p. 1346–1347, 2005.
- [175] W. YANG, A. L. WILKINS, Y. YE, Z. LIU, S. LI, J. URBAUER, H. HELLINGA, A. KEARNEY, P. V. der MERWE et J. YANG, Design of a calcium-binding protein with desired structure in a cell adhesion molecule, *J. Am. Chem. Soc.*, vol. 127, p. 2085–2093, 2005.
- [176] D. BOLON, R. GRANT, T. BAKER et R. SAUER, Specificity versus stability in computational protein design, *Proc. Natl. Acad. Sci. USA*, vol. 102, p. 12724–12729, 2005.
- [177] L. JOACHIMIAC, T. KORTEMME, B. STODDARD et D. BAKER, Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface, *J. Mol. Biol.*, vol. 361, p. 195–208, 2006.
- [178] J. ASHWORTH, J. HAVRANEK, C. DUARTE, D. SUSSMAN, R. MONNAT, B. STODDARD et D. BAKER, Computational redesign of endonuclease DNA binding and cleavage specificity, *Nature*, vol. 441, p. 656–659, 2006.
- [179] M. ALI, C. TAYLOR, G. GRIGORYAN, K. ALLEN, B. IMPERIALI et A. KEATING, Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed  $\alpha\beta$  structure, *Structure*, vol. 13, p. 225–234, 2005.
- [180] D. GREEN, A. DENNIS, P. FAM, B. TIDOR et A. JASANO, Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide, *Biochemistry*, vol. 45, p. 12547–12559, 2006.
- [181] X. AMBROGGIO et B. KUHLMAN, Computational design of a single amino acid sequence that can switch between two distinct protein folds, *J. Am. Chem. Soc.*, vol. 128, p. 1154–1161, 2006.
- [182] H. HELLINGA et F. RICHARDS, Construction of new ligand binding sites in proteins of known structure. i. computer-aided modeling of sites with predefined geometry, *J. Mol. Biol.*, vol. 222, p. 763–785, 1991.
- [183] M. KLEMBA et L. REGAN, Characterization of metal binding by a designed protein : single ligand substitutions at a tetrahedral Cys2His2 site, *Biochemistry*, vol. 34, p. 10094–10100, 1995.
- [184] L. REGAN, Protein design : novel metal-binding sites, *Trends Biochem. Sci.*, vol. 20, p. 280–285, 1995.
- [185] R. CHAKRABARTI, A. KLIBANOV et R. FRIESNER, Computational prediction of native protein ligand-binding and enzyme active site sequences, *Proc. Natl. Acad. Sci. USA*, vol. 102, p. 10153–10158, 2005.
- [186] J. LASSILA, J. KEEFFE, P. OELSCHLAEGER et S. MAYO, Computationally designed 64 variants of *Escherichia coli* chorismate mutase show altered catalytic activity, *Protein Eng. Des. Sel.*, vol. 18, p. 161–163, 2005.
- [187] P. OESCHLAEGER et S. MAYO, Hydroxyl groups in the beta beta sandwich of metallo-beta-lactamases favor enzyme activity : A computational protein design study, *J. Mol. Biol.*, vol. 350, p. 395–401, 2005.

- [188] A. ZANGHELLINI, L. JIANG, A. WOLLACOTT, G. CHENG, J. MEILLER, E. ALTHOFF, D. RÖTHLISBERGER et D. BAKER, New algorithms and an in silico benchmark for computational enzyme design, *Prot. Sci.*, vol. 15, p. 2785–2794, 2006.
- [189] J. LASSILA, H. PRIVETT, B. ALLEN et S. MAYO, Combinatorial methods for small-molecule placement in computational enzyme design, *Proc. Natl. Acad. Sci. USA*, vol. 103, p. 16710–16715, 2006.
- [190] D. EISENBERG et A. MCCLACHLAN, Solvation energy in protein folding and binding, *Nature*, vol. 319, p. 199–203, 1986.
- [191] T. OOI, M. OOBATAKE, NEMETHY et H. SCHERAGA, Accessible surface areas as a measure of the thermodynamic hydration parameters of peptides, *Proc. Natl. Acad. Sci. USA*, vol. 84, p. 3086–3090, 1987.
- [192] F. FRATERNALI et W. van GUNSTEREN, An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution., *J. Mol. Biol.*, vol. 256, p. 939–948, 1996.
- [193] W. WANG, W. LIM, A. JAKALIAN, J. WANG, R. LUO, C. BAYLY et P. KOLLMAN, An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis, *J. Am. Chem. Soc.*, vol. 123, p. 3986–3994, 2001.
- [194] T. HOU, X. QIAO, W. ZHANG et X. XU, Empirical aqueous solvation models based on accessible surface areas with implicit electrostatics, *J. Phys. Chem. B*, vol. 106, p. 11295–11304, 2002.
- [195] J. PEI, Q. WANG, J. ZHOU et L. LAI, Estimating protein-ligand binding free energy : Atomic solvation parameters for partition coefficient and solvation free energy calculation, *Proteins*, vol. 57, p. 661–664, 2004.
- [196] J. MYERS, C. PACE et J. SCHOLTZ, Helix propensities are identical in proteins and peptides, *Biochemistry*, vol. 36, p. 10923–10929, 1997.
- [197] S. PARK, W. SHALONGO et E. STELLWAGEN, Residue helix parameters obtained from dichroic analysis of peptides of defined sequence, *Biochemistry*, vol. 32, p. 7048–7053, 1993.
- [198] J. YANG, E. SPEK, Y. GONG, H. ZHOU et N. KALLENBACH, The role of context on  $\alpha$ -helix stabilization : host-guest analysis in a mixed background peptide model, *Prot. Sci.*, vol. 6, p. 1–9, 1997.
- [199] R. VARADARAJAN, P. CONNELLY, J. STURTEVANT et F. RICHARDS, Heat capacity changes for protein-peptide interactions in the ribonuclease s system, *Biochemistry*, vol. 31, p. 1421–1426, 1992.
- [200] S. PADMANABHAN, S. MARQSEE, T. RIDGEWAY, T. LAUE et R. BALDWIN, Relative helix-forming tendencies of nonpolar amino acids, *Nature*, vol. 344, p. 268–270, 1990.
- [201] P. LYU, M. LIFF, L. MARKY et N. KALLENBACH, Side chain contributions to the stability of  $\alpha$ -helical structure in peptides, *Science*, vol. 250, p. 669–673, 1990.
- [202] K. SHOEMAKER, P. KIM, D. BREMS, S. MARQSEE, E. YORK, I. CHAIKEN, J. STEWART et R. BALDWIN, Nature of the charged-group effect on the stability of the C-peptide helix, *Proc. Natl. Acad. Sci. USA*, vol. 82, p. 2349–2353, 1985.
- [203] N. GUEX et M. PEITSCH, SWISS-MODEL and the Swiss-Pdb-Viewer : An environment for comparative protein modeling, *Electrophoresis*, vol. 18, p. 2714–2723, 1997.

- [204] P. TUFFERY, C. ETCHEBEST, S. HAZOUT et R. LAVERY, A new approche to the rapid determination of protein side chain conformations., *J. Biomol. Struct. Dynam.*, vol. 8, p. 1267–1289, 1991.
- [205] M. KUMAR, K. BAVA, M. GROMIHA, P. PARABAKARAN, K. KITAJIMA, H. UEDAIRA et A. SARAI, ProTherm and ProNIT : thermodynamic databases for proteins and protein-nucleic acid interactions, *Nucl. Acids Res.*, vol. 34, p. 204–206, 2006.
- [206] C. PACE et J. SCHOLTZ, A helix propensity scale based on experimental studies of peptides and proteins, *Biophys. J.*, vol. 75, p. 422–427, 1998.
- [207] C. HO et A. FERSHT, Internal thermodynamics of position 51 mutants and natural variants of tyrosyl-tRNA synthetase, *Biochemistry*, vol. 25, p. 1891–1897, 1986.
- [208] A. FERSHT, R. LEATHERBARROW et T. WELLS, Structure-reactivity relationships in engineered proteins : analysis of use of binding energy by linear free energy relationships, *Biochemistry*, vol. 26, p. 6030–6038, 1987.
- [209] T. WELLS et A. FERSHT, Use of binding energy in catalysis analysed by mutagenesis of tyrosyl-tRNA synthetase, *Biochemistry*, vol. 25, p. 1881–1886, 1986.
- [210] J. CAVARELLI, G. ERIANI, B. REES, M. RUFF, M. BOEGLIN, A. MITSCHLER, F. MARTIN, J. GANGLOFF, J. THIERRY et D. MORAS, The active site of yeast aspartyl-tRNA synthetase : structural and functional aspects of the aminoacylation reaction, *EMBO J.*, vol. 13, p. 327–337, 1994.
- [211] K. SHARP, Calculation of HyHel10-Lysozyme binding free energy changes : Effect of ten point mutations, *Proteins*, vol. 48, p. 39–48, 1998.
- [212] U. MOEBIUS, L. CLAYTON, S. ABRAHAM, S. HARRISON et E. REINHERZ, The human immunodeficiency virus gp120 binding site on CD4 : delineation by quantitative equilibrium and kinetic binding studies of mutants in conjunction with a high-resolution CD4 atomic structure, *J. Exp. Med.*, vol. 176, p. 507–517, 1992.
- [213] M. ALMLÖF, J. AQVIST, A. SMALAS et B. BRANSDAL, Probing the effect of point mutations at protein-protein interfaces with free energy calculations, *Biophys. J.*, vol. 90, p. 433–442, 2006.
- [214] D. KROWARSCH, M. DADLEZ, O. BUCZEK, I. KROKOSZYNKA, A. SMALAS et J. OTLEWSKI, Interscuffolding additivity : Binding of P1 variants of bovine pancreatic trypsin inhibitor to four serine proteases, *J. Mol. Biol.*, vol. 289, p. 175–186, 1999.
- [215] T. KOBAYASHI, T. TAKIMURA, R. SEKINE, K. VINCENT, K. KAMATA, K. SAKAMOTO, S. NISHIMURA et S. YOKOYAMA, Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase, *J. Mol. Biol.*, vol. 346, p. 105–117, 2004.
- [216] Y. ZHANG, L. WANG, P. SCHULTZ et I. WILSON, Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine, *Prot. Sci.*, vol. 14, p. 1340–1349, 2005.
- [217] N. ESWAR, B. WEBB, M. MARTI-RENOM, M. MADHUSUDHAN, D. ERAMIAN, M. SHEN, U. PIEPER et A. SALI, Comparative protein structure modeling with MODELLER, *Current Protocols in Bioinformatics*, John Wiley and Sons, Inc, vol. Supplement 15, p. 5.6.1–5.6.30, 2006.
- [218] A. BRÜNGER, X-PLOR manual version 3.1., Yale University New Haven, Connecticut., 1992.

- [219] S. NOSE, A unified formulation of the constant temperature molecular-dynamics methods, *J. Chem. Phys.*, vol. 81, p. 511–519, 1984.
- [220] T. DARDEN, L. PERERA, L. LI et L. PEDERSEN, New tricks for modelers from the crystallography toolkit : the particle mesh ewald algorithm and its use in nucleic acid simulations, *Structure*, vol. 7, p. 55–60, 1999.