UNIVERSITÉ DE LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

U.F.R. Sciences et Techniques de la Matière et des Procédés
Ecole Doctorale Lorraine de Chimie et Physique Moléculaires (SESAMES)
Département de Formation Doctorale en Chimie Informatique et Théorique

THÈSE

présentée à l'Université Henri Poincaré - Nancy I
en vue de l'obtention du grade de

Docteur de l'Université Henri Poincaré, Nancy I

en Chimie Informatique et Théorique

par

Eddy THIRIOT

Modélisation de la reconnaissance et de la catalyse enzymatiques : développements méthodologiques et détermination du mécanisme des Méthionine Sulfoxyde Réductases de classe A

soutenue publiquement le 4 juin 2009, salle Jean Barriol, Faculté des Sciences et Techniques

Membres du jury :

Président :

M. Guy BRANLANT	Professeur, Université Henri Poincaré, Nancy

Rapporteurs :

M. Serge ANTONCZAK	Maître de Conférences (HDR), Université Nice Sophia Antipolis
Mme. Anne MILET	Professeur, Université Joseph Fourier, Grenoble

Examinateurs :

Mme. Sandrine BOSCHI-MULLER	Professeur, Université Henri Poincaré, Nancy
M. Eric HENON	Professeur, Université de Reims Champagne-Ardenne

Directeurs de thèse :

M. Gérald MONARD	Professeur, Université Henri Poincaré, Nancy
M. Manuel F. RUIZ-LOPEZ	Directeur de recherche, CNRS, Nancy

Equipe de Chimie et Biochimie Théoriques, Unité Mixte de Recherche CNRS-UHP 7565
Faculté des Sciences et Techniques, BP 239 - 54506 Vandœuvre-lès-Nancy Cedex

Nancy-Université
Université Henri Poincaré

U.F.R. Sciences et Techniques de la Matière et des Procédés
Ecole Doctorale Lorraine de Chimie et Physique Moléculaires (SESAMES)
Département de Formation Doctorale en Chimie Informatique et Théorique

THÈSE

présentée à l'Université Henri Poincaré - Nancy I
en vue de l'obtention du grade de

Docteur de l'Université Henri Poincaré, Nancy I

en Chimie Informatique et Théorique

par

Eddy THIRIOT

Modélisation de la reconnaissance et de la catalyse enzymatiques : développements méthodologiques et détermination du mécanisme des Méthionine Sulfoxyde Réductases de classe A

soutenue publiquement le 4 juin 2009, salle Jean Barriol, Faculté des Sciences et Techniques

Membres du jury :

Président :

M. Guy BRANLANT	Professeur, Université Henri Poincaré, Nancy

Rapporteurs :

M. Serge ANTONCZAK	Maître de Conférences (HDR), Université Nice Sophia Antipolis
Mme. Anne MILET	Professeur, Université Joseph Fourier, Grenoble

Examinateurs :

Mme. Sandrine BOSCHI-MULLER	Professeur, Université Henri Poincaré, Nancy
M. Eric HENON	Professeur, Université de Reims Champagne-Ardenne

Directeurs de thèse :

M. Gérald MONARD	Professeur, Université Henri Poincaré, Nancy
M. Manuel F. RUIZ-LOPEZ	Directeur de recherche, CNRS, Nancy

Equipe de Chimie et Biochimie Théoriques, Unité Mixte de Recherche CNRS-UHP 7565
Faculté des Sciences et Techniques, BP 239 - 54506 Vandœuvre-lès-Nancy Cedex

Je remercie toutes les personnes qui, de près ou de loin, m'ont accompagné et m'ont soutenu durant ces cinq années de thèse.

Je remercie tout d'abord le Pr. Gérald Monard et le Dr. Manuel Ruiz-López, mes co-directeurs de thèse, de m'avoir accueilli au sein de l'équipe de Chimie et Biochimie Théoriques et d'avoir accepté d'encadrer mon travail. Je tiens à leur montrer ma profonde reconnaissance pour leur patience et leur disponibilité pour répondre à mes interrogations et résoudre mes problèmes ainsi que pour le travail de relecture de ce manuscrit.

Je remercie le Pr. Guy Branlant et le Pr. Sandrine Boschi-Muller d'avoir accepté de faire partie de mon jury de thèse et pour leurs suggestions concernant le manuscrit. J'étends ces remerciements à tous les membres (et ex-membres) de l'équipe d'Enzymologie Moléculaire et Structurale (AREMS) pour les réunions (officielles ou non) au cours desquelles nous avons pu discuter et échanger nos résultats et nos points de vue concernant le mécanisme des Msr.

Je souhaite aussi remercier le Pr. Anne Milet et le Dr. Serge Antonczak d'avoir accepté de rapporter cette thèse et le Pr. Eric Hénon d'avoir accepté de juger mon travail.

Je tiens à montrer toute ma gratitude au Pr. Eric Hénon et au Pr. Frédéric Bohr pour l'intérêt qu'ils portent à mon travail et leur soutien constant depuis mon stage de maîtrise, il y a sept ans. Je les remercie de m'avoir fait découvrir le monde de la chimie théorique. J'adresse aussi mes remerciements au Dr. Sébastien Canneaux, ancien thésard de la "Dream Team" rémoise, pour sa proposition de stage post-doctoral dans son équipe dès septembre prochain.

J'adresse un grand merci aux membres du CBT et du SRSMC pour leur gentillesse, leur sympathie et surtout pour l'aide qu'ils m'ont apportée pendant mon séjour au laboratoire. Je ne ferai pas la liste de toutes ces personnes, mais qu'elles sachent qu'elles ont été essentielles à l'accomplissement de cette thèse dans un cadre très agréable.

Je ne manquerai pas de remercier le Pr. Pascale Friant-Michel d'avoir accepté d'être ma tutrice de monitorat et de m'avoir proposé de poursuivre mon expérience d'enseignement par deux années d'ATER.

Mes remerciements s'adressent, enfin, à mes amis et à ma famille qui, comme dans les coulisses d'un théâtre, ont joué un rôle tout aussi important que les autres dans la réalisation de mon travail de thèse. Je pense tout particulièrement à Nicolas auprès duquel j'ai trouvé une présence plus que précieuse. Je le remercie d'avoir compris et apaisé mes angoisses et de m'avoir montré un soutien indéniable. J'adresse un immense merci à mes parents à qui je dédie ce manuscrit. Ils n'ont jamais hésité à faire des sacrifices financiers pour me permettre de poursuivre mes études. En acceptant et en soutenant mes choix durant mon parcours scolaire, ils auront aussi su trouver la meilleure façon de m'aider. Pour tout cela, je leur serai toujours reconnaissant.

Il est bien difficile de croire que tant de merveilles, tant d'astuces miraculeuses, tant d'ingéniosité dans le monde soient l'effet du hasard et de la chimie seulement.

La faim du tigre (1966)

René Barjavel
(1911-1985)

Table des matières

Introduction, objectifs **13**

I. Développement d'un programme de docking moléculaire semi-empirique à croissance linéaire **19**

1. Rappels bibliographiques et objectifs **21**
1.1. Utilité des programmes de docking moléculaire 22
1.2. Principe de fonctionnement 22
1.2.1. Modélisation du système protéine-ligand et définition du score 23
1.2.2. Algorithme de recherche 25
1.3. Motivations pour le développement d'un nouveau programme 25
1.4. Objectifs, cahier des charges du nouveau programme 27

2. Développement d'un algorithme génétique de minimisation **29**
2.1. Méthodes de minimisation et atouts des algorithmes génétiques 30
2.2. Principe des algorithmes génétiques 31
2.2.1. Définition et principe général 32
2.2.2. Exemple de minimisation 34
2.3. Programmation de l'algorithme génétique 39
2.3.1. Structure du code 39
2.3.2. Codage des gènes 42
2.3.3. Sélection 44
2.3.4. Reproduction 45
2.3.5. Stratégie d'évolution 49
2.3.6. Population initiale 50
2.3.7. Evaluation 51
2.4. Conclusion 53

3. Interface ALGOGEN-DIVCON **55**
3.1. Présentation générale 56
3.2. Structure du code 57
3.2.1. Module `docking.py` 57
3.2.2. Module `quaternion.py` 59
3.2.3. Module `divcon.py` 60
3.3. Déroulement du docking 60

4. Validation du programme ALGOGEN-DIVCON **65**
4.1. Structure du dimère de l'eau . . . 66
4.2. Docking dans une protéine . . . 70
4.2.1. Docking de la 8-azaxanthine dans le site actif de l'urate oxydase . . . 74
4.2.2. Docking de l'acide oxonique dans le site actif de l'urate oxydase . . . 78
4.2.3. Docking de l'acide urique dans le site actif de l'urate oxydase . . . 80
4.3. Conclusions . . . 82

II. Mécanisme catalytique des méthionine sulfoxyde réductases de classe A **85**

5. Rappels bibliographiques et objectifs **87**
5.1. Stress oxydant et oxydation des résidus méthionine . . . 88
5.1.1. Oxydation des résidus méthionine par les EOA . . . 89
5.1.2. Conséquences de l'oxydation des résidus méthionine . . . 91
5.1.3. Propriétés physico-chimiques de la fonction sulfoxyde . . . 93
5.2. Méthionine sulfoxyde réductases . . . 95
5.2.1. Implication physiologique des méthionine sulfoxyde réductases . . . 96
5.2.2. Mécanisme catalytique . . . 98
5.2.3. Structure primaire des méthionine sulfoxyde réductases . . . 99
5.2.4. Structure tridimensionnelle des méthionine sulfoxyde réductases . . . 102
5.3. Etudes antérieures sur l'étape réductase de MsrA . . . 108
5.3.1. Caractérisation cinétique . . . 108
5.3.2. Identification des résidus impliqués . . . 108
5.3.3. Etat de protonation des résidus du site actif . . . 109
5.3.4. Travaux théoriques sur la réduction d'un sulfoxyde par un thiol . . . 109
5.4. Objectifs et approche méthodologique . . . 114

6. Reconnaissance et ancrage du substrat **117**
6.1. Etat de protonation du site actif . . . 118
6.2. Simulations, protocole opératoire . . . 119
6.3. Enzyme libre . . . 126
6.3.1. Etat de protonation CysH-Glu$^{\ominus}$ . . . 126
6.3.2. Etat de protonation CysH-GluH . . . 129
6.3.3. Etat de protonation Cys$^{\ominus}$-GluH . . . 132
6.4. Complexe enzyme-substrat sulfoxyde . . . 135
6.4.1. Etat de protonation CysH-Glu$^{\ominus}$ . . . 135
6.4.2. Etat de protonation CysH-GluH . . . 141
6.4.3. Etat de protonation Cys$^{\ominus}$-GluH . . . 148
6.5. Conclusions . . . 155

7. Activation du substrat **157**
7.1. Réduction des sulfoxydes : généralités . . . 158
7.2. Mécanisme de transfert de proton . . . 159
7.2.1. Modèle du site actif . . . 160
7.2.2. Méthodologie . . . 161
7.2.3. Chemin de réaction . . . 162
7.3. Dynamique moléculaire du substrat activé . . . 173
7.3.1. Simulation de la forme protonée du DMSO . . . 173
7.3.2. Simulation de l'intermédiaire sulfurane dans le site actif . . . 180
7.4. Conclusions . . . 188

8. Réduction du substrat **189**
8.1. Mécanisme I : Transfert d'un groupement OH . . . 190
8.1.1. Hypothèse de travail . . . 190
8.1.2. Chemin de réaction dans un système isolé . . . 190
8.1.3. Chemin de réaction dans un environnement enzymatique modèle . . . 196
8.2. Mécanisme II : 1 H^+ + 1 e^- . . . 200
8.2.1. Hypothèse de travail . . . 200
8.2.2. Propriétés redox des espèces impliquées . . . 201
8.2.3. Transfert d'électron dans l'environnement enzymatique . . . 204
8.3. Mécanisme III : 2 H^+ + 2 e^- . . . 209
8.3.1. Hypothèse de travail . . . 209
8.3.2. Chemin de réaction dans un environnement enzymatique modèle . . . 209
8.4. Devenir du substrat méthionine réduit . . . 216
8.5. Conclusions . . . 216

Conclusion générale **219**
Docking moléculaire semi-empirique à croissance linéaire . . . 219
Mécanisme catalytique des méthionine sulfoxyde réductases de classe A . . . 220
Perspectives . . . 221

III. Annexes **223**

A. Méthodologie **225**
A.1. Méthodes semi-empiriques à croissance linéaire . . . 227
A.2. Mécanique et dynamique moléculaires . . . 230
A.2.1. Champs de forces . . . 230
A.2.2. Conditions aux limites périodiques . . . 234
A.2.3. Traitement des interactions électrostatiques à longue portée . . . 234
A.2.4. Dynamique moléculaire . . . 236
A.3. PROPKA : Prédiction des valeurs de pK_a des résidus d'une protéine . . . 239

B. Manuel utilisateur d'AlgoGen **245**

C. Manuel utilisateur d'AlgoGen-DivCon **263**

Lexique **269**

Table des figures **273**

Liste des tableaux **277**

Bibliographie **279**

Introduction, objectifs

... les connaissances qui résulteront un jour de l'application de la chimie à l'étude des corps vivants et de leurs fonctions, nous offriront peut-être contre nos maux des ressources que nous n'oserions prévoir aujourd'hui ...

Eloge à M. Margraaf (1776)

Marie Jean Antoine Nicolas de Caritat,
Marquis de Condorcet
(1743-1794)

En 1776, Condorcet écrivait ces mots dans une éloge à M. Margraaf, chimiste d'origine allemande membre de l'Académie Royale des Sciences. Le philosophe et mathématicien reconnait ainsi l'énorme espoir que représente la chimie, science naissante, pour à la fois comprendre le fonctionnement des corps vivants et guérir la maladie. Il déplore, par ailleurs, le manque d'intérêt de ses contemporains à investir ce champ de recherche prometteur dont on peut imaginer l'impact qu'il aura sur l'amélioration des conditions de vie des hommes.

De nos jours, le développement de la biochimie a fait progresser de manière spectaculaire la connaissance générale des processus chimiques qui gouvernent le vivant, notamment dans les domaines de l'enzymologie et de la pharmacologie. Le développement des interfaces avec la physique et l'informatique a donné lieu à la naissance d'autres disciplines telles que la biophysique ou la biochimie théorique qui visent à déterminer la structure et la réactivité des systèmes biologiques par la modélisation informatique des processus au niveau moléculaire.

Cette thèse s'inscrit dans le cadre du projet SIRE (Simulation Informatique de la Réactivité Enzymatique) financé par le programme ACI IMPBio (Informatique, Mathématiques, Physique en Biologie moléculaire) du Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche. Le projet SIRE comporte une partie méthodologique qui vise le développement d'une suite de logiciels de modélisation et de simulation en chimie du vivant. La partie applicative de ce projet porte sur l'étude du mode de fonctionnement des enzymes méthionine sulfoxyde réductases (Msr) par la réalisation d'un travail concomitant de recherche expérimentale et théorique. Il réunit trois équipes de recherche dans différents domaines :

- Equipe de Chimie et Biochimie Théoriques (SRSMC, UMR CNRS-UHP 7565), Université Henri Poincaré Nancy I ;
- Equipe ScalApplix "Schémas et algorithmes hautes performances pour les applications scientifiques" (LaBRI, UMR CNRS-Univ. Bordeaux I-ENSEIRB 5800), Université Bordeaux I ;
- Equipe d'Enzymologie Moléculaire et Structurale (AREMS, UMR CNRS-UHP 7214), Université Henri Poincaré Nancy I.

Avant de décrire les objectifs de cette thèse, nous allons évoquer les étapes inhérentes à la modélisation de processus enzymatiques (dont un schéma général est présenté dans la figure I-1) et les difficultés qui y sont associées : (*i*) le choix de la structure de l'enzyme, (*ii*) l'étude du processus de formation du complexe enzyme-substrat et (*iii*) l'étude du processus catalytique proprement dit.

a. Structure de l'enzyme

Actuellement, la détermination *a priori* de la structure tridimensionnelle d'une protéine n'est pas possible et les études théoriques en catalyse enzymatique ont recours aux données expérimentales[a]. Elles peuvent provenir de différentes techniques : cristallographie par diffraction de rayons X, spectroscopie par résonance magnétique nucléaire, *etc*. La *Protein Data Bank*[2] (PDB) est une source de données fondamentale dans ce sens.

Cependant, les structures déterminées expérimentalement sont fréquemment incomplètes. C'est le cas notamment de celles obtenues par diffraction de rayons X, donnant uniquement la position des atomes lourds et rendant nécessaire l'ajout des atomes d'hydrogène. Cette étape est relativement facile à mettre en œuvre pour la plupart des résidus, mais un problème peut toutefois se poser pour

[a] Les méthodes d'homologie de séquences nécessitent aussi la connaissance de structures expérimentales pour proposer une structure modèle.

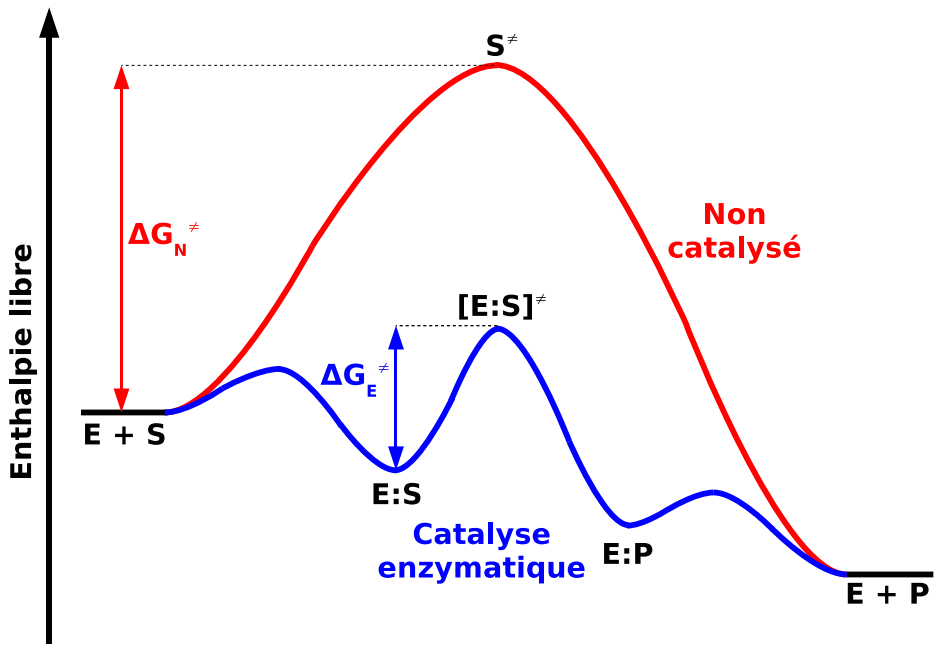


FIG. I-1: *Schéma du bilan énergétique d'une catalyse enzymatique. La formation du complexe enzyme-substrat* $[E:S]$ *permet de diminuer l'énergie d'activation de la réaction* $\Delta G^{\neq}$*. Tiré du livre Biochemistry de Voet et al.*[1]*.*

les résidus ionisables dont les propriétés acido-basiques conditionnent la réactivité de l'enzyme[3]. Ces résidus peuvent parfois être enfouis profondément au sein de l'enzyme ou être en interaction avec d'autres résidus ; leur valeur de pK_a peut ainsi subir de fortes modifications.

Il existe différentes méthodes permettant la prédiction du pK_a des résidus dans les protéines. La plupart d'entre-elles[4–7] sont basées sur des modèles de continuum électrostatique qui résolvent numériquement l'équation de Poisson-Boltzmann linéarisée (LPBE). La méthode de Hammer-Taft[8] est une méthode rapide, principalement utilisée pour la prédiction du pK_a de petits composés organiques acido-basiques (effet de substituants), qui peut donner, dans certains cas, une bonne estimation de l'état de protonation d'un résidu. La méthode PROPKA développée par Jensen et coll.[9,10] est une approche relativement récente qui s'appuie sur la structure de la protéine et sur une paramétrisation empirique des divers termes qui influencent le déplacement de pK_a (liaisons hydrogène, interactions électrostratiques et désolvatation). Les calculs sont fiables comparativement à d'autres méthodes de plus haut niveau[11,12] mais sont réalisés beaucoup plus rapidement. La méthode PROPKA constitue actuellement une méthode de choix pour analyser l'état de protonation des résidus d'un système protéinique. Signalons que certains aspects de cette méthode sont en cours de développement dans notre laboratoire en collaboration avec le Pr. Jensen (Université d'Iowa puis Université de Copenhague).

b. Reconnaissance et ancrage du substrat (docking)

La réaction enzymatique suppose la reconnaissance spécifique du substrat par l'enzyme. Son ancrage dans le site actif correspond à la formation d'un complexe stabilisé par l'existence d'interactions diverses (liaison hydrogène, électrostatique, van der Waals, *etc*). Ce complexe, appelé complexe michaélien (ou complexe de *Michaelis*), conditionne la réactivité de l'enzyme et la détermination de sa structure est primordiale pour comprendre le mécanisme réactionnel.

En principe, les simulations de dynamique moléculaire ou de Monte Carlo peuvent rendre compte de la structure du complexe protéine-ligand et de la valeur énergétique d'affinité[13–17]. Cependant, l'utilisation de ce type d'approche dans le cadre de la reconnaissance moléculaire est relativement récente à cause de certaines limitations. D'une part, les champs de forces utilisés ne permettent pas toujours d'obtenir une valeur précise de l'affinité et, d'autre part, la procédure nécessite la production de trajectoires très longues (pour avoir un échantillonage suffisament complet). Les méthodes de docking moléculaire ont été développées pour traiter plus spécifiquement de l'ancrage d'un ligand. Elles reposent, pour certaines, sur une approche de mécanique moléculaire (comme pour les simulations) avec quelques modifications pour simplifier le traitement. Le chapitre 1 de ce manuscrit traite plus en détail de ces méthodes.

c. Mécanisme de réaction

Pour la détermination du mécanisme réactionnel, l'utilisation des méthodes de la chimie quantique s'avère indispensable car il s'agit de décrire la rupture et la formation de liaisons dans le système biologique. Dans un premier temps, le chemin de réaction est généralement étudié en solution, à l'aide d'un système modèle, puis transposé dans le site actif de l'enzyme. L'environnement enzymatique, *i.e.* les résidus du site actif mais aussi l'environnement de ces résidus, joue un rôle souvent très important qu'il convient de prendre en compte pour pouvoir décrire la catalyse. Cependant, il n'est pas possible en général de décrire tout le système biologique avec les méthodes standards de la chimie quantique et des approximations sont nécessaires.

Plusieurs méthodes ont été développées et utilisent, soit une partition du système complet en deux sous-systèmes classique/quantique comme les méthodes ONIOM[18–24] et QM/MM[25–32], soit des algorithmes à croissance linéaire comme les méthodes Divide & Conquer[33], MOZYME[34] ou Density Matrix Minimization[35] pour ne citer que les plus connues.

d. Objectifs de la thèse

Parmi les objectifs du projet SIRE, ce travail de thèse s'est intéressé aux deux points suivants : (*i*) le développement d'un programme de docking moléculaire exploitant le principe de calcul quantique semi-empirique à croissance linéaire sous architecture parallèle et (*ii*) la détermination du mécanisme réactionnel des méthionine sulfoxyde réductases de classe A. Ce manuscrit est ainsi divisé en deux parties.

Dans la première, le développement du programme de docking moléculaire sera décrit. Nous rappellerons l'utilité de ce type de programme dans le contexte général d'une étude théorique de systèmes biologiques complexes, ainsi que le principe général de leur fonctionnement. Nous exposerons ensuite les motivations qui nous ont menés à développer un nouveau code reposant sur un calcul quantique de l'énergie, ainsi que le cahier des charges que nous nous sommes fixé. Nous décrirons enfin le code informatique développé durant ce travail qui utilise un algorithme de type "génétique". Le programme de docking moléculaire sera validé par l'étude de différents systèmes dont un complexe enzyme-substrat.

La deuxième partie du manuscrit traitera du mécanisme réactionnel des méthionine sulfoxyde réductases de classe A. Cette étude sera menée dans une approche "multi-outils" combinant des méthodes issues de la mécanique statistique (dynamique moléculaire) pour l'analyse des aspects structuraux du système enzymatique et des méthodes issues de la mécanique quantique pour ce qui a trait à la réactivité. Nous ferons, dans un premier temps, un rappel bibliographique portant sur l'importance biologique de cette classe d'enzymes. Nous ferons également un bilan des études (théoriques et expérimentales) antérieures à cette thèse qui fournissent des éléments de connaissance importants vis-à-vis du travail à réaliser. Nous décrirons enfin les résultats de nos travaux qui apportent des informations fondamentales et nous permettent de proposer un mécanisme détaillé de l'étape réductase pour ces enzymes.

Dans un souci de clarté, nous ne ferons volontairement pas, dans le corps de ce manuscrit, de rappels méthodologiques correspondant aux outils utilisés, mais ceux-ci seront présentés sous forme d'annexes. Cependant, la méthode utilisée dans chaque étape du travail sera rappelée dans le texte à chaque fois que cela sera nécessaire.

Nous avons aussi reporté en annexe du manuscrit les deux manuels utilisateur des programmes que nous avons écrits : AlgoGen (algorithme génétique de minimisation) et AlgoGen-DivCon (logiciel de docking).

Première partie

Développement d’un programme de docking moléculaire semi-empirique à croissance linéaire

Chapitre 1

Rappels bibliographiques et objectifs

Sommaire

1.1. Utilité des programmes de docking moléculaire 22
1.2. Principe de fonctionnement 22
1.3. Motivations pour le développement d'un nouveau programme 25
1.4. Objectifs, cahier des charges du nouveau programme 27

Ce chapitre présente de façon générale et succincte les programmes de docking moléculaire et leur utilité dans le contexte d'une étude théorique des systèmes biologiques. De tels programmes (commerciaux ou non) sont déjà disponibles, mais ils conduisent à certaines limitations, principalement liées au traitement classique des interactions, que nous verrons en détail. A partir de ces constatations, nous présenterons les motivations qui nous ont amenés à développer un nouveau programme de docking moléculaire et les améliorations que nous comptons apporter en rapport avec les limitations des autres programmes. Nous terminerons enfin ce chapitre d'introduction en donnant les caractéristiques et les spécificités de notre approche qui seront décrites plus en détail dans les chapitres 2 et 3.

1.1. Utilité des programmes de docking moléculaire

Les interactions protéine-protéine et protéine-ligand jouent un rôle clé dans l'organisation des systèmes biologiques. Elles permettent la régulation de certains processus biologiques[36,37], la transmission des signaux[38] ou encore la catalyse de diverses réactions biochimiques[39–41]. Connaître la façon dont les protéines interagissent avec d'autres entités biochimiques est une étape essentielle pour comprendre les processus biologiques dans lesquels elles sont impliquées. Le développement d'approches prédictives ouvre la voie à la conception assistée par ordinateur de systèmes protéiniques aux propriétés modifiées[42–44] et présente donc un intérêt indéniable pour la recherche et l'industrie phamaceutique et médicale.

Dans le domaine de la chimie théorique, le docking moléculaire est une méthode qui prédit la conformation (position et orientation relatives) la plus favorable de deux molécules en interaction et formant un complexe stable[45–47]. La connaissance de cette conformation préférentielle permet par la suite l'estimation de la force d'association (ou affinité de liaison) entre ces deux molécules.

Il existe deux grands types de docking moléculaire. Le docking rigide consiste à obtenir la conformation préférentielle d'un système protéine-ligand en considérant que chacune des deux molécules conserve une géométrie interne fixe. Dans ce cas, la relaxation de la géométrie interne de chaque entité, en interaction dans le complexe, n'est pas prise en compte. Cependant, il est tout à fait concevable que les structures de la protéine et du ligand soient modifiées durant le processus de docking moléculaire afin d'optimiser au mieux l'interaction entre les deux entités. On parle, dans ce cas, de docking flexible.

1.2. Principe de fonctionnement

Le docking moléculaire peut se rapporter à un problème d'optimisation selon une valeur numérique qui rendrait compte de la conformation plus ou moins favorable de deux entités (*e.g.* l'énergie libre totale du système protéine-ligand). Cette valeur numérique sera appelée score de docking (ou simplement score) par la suite. Sa définition est essentielle à tout programme de docking moléculaire car elle détermine le résultat de la prédiction. En fait, il existe différentes manières d'obtenir ce score en fonction de la qualité du docking moléculaire que l'on souhaite réaliser.

1.2.1. Modélisation du système protéine-ligand et définition du score

Un grand nombre de programmes (commerciaux ou non) de docking moléculaire sont d'ores et déjà disponibles. Parmi ceux-ci, nous citerons par exemple AUTODOCK[48], FLEXX[49], HADDOCK[50], GASDOCK[51], GOLD[52], *etc*. Ils diffèrent les uns des autres sur la manière de réprésenter le système moléculaire et la manière de déterminer le score de docking (fonction de score). Deux approches sont principalement employées pour la modélisation du système protéine-ligand.

a. Représentation géométrique

Cette première approche a conduit au développement de méthodes géométriques dites de complémentarité de forme (*shape complementarity*) où la protéine et le ligand sont décrits par des surfaces de descripteurs[53,54] qu'il suffit d'apparier (*matching*). La construction de la surface de descripteurs de chacune des deux entités prend en compte différentes caractéristiques telles que la surface moléculaire, la surface accessible au solvant, les propriétés hydrophobes, la formation éventuelle de liaisons hydrogène, *etc*. Elles peuvent être construites aussi grâce à la technique des descripteurs de surface de Fourier[55–57]. Dans cette approche, la fonction de score mesure un taux d'appariement correct des descripteurs entre la protéine et le ligand pour chaque conformation à évaluer. Plus ce taux est grand, plus le score de docking est élevé et plus la conformation du système protéine-ligand est optimale. Ces méthodes sont typiquement rapides et robustes. Elles sont choisies préférentiellement, par exemple, dans le cas d'un scan de plusieurs milliers de ligands pharmacophores. Ces méthodes ont le défaut de ne pas pouvoir modéliser exactement les variations dynamiques de configurations protéine-ligand, malgré de récents développements en vue de rendre le ligand flexible[58–60].

b. Représentation atomique

La seconde approche repose, quant à elle, sur une simulation du processus de docking[61] avec une description atomique du système protéine-ligand. La protéine et le ligand sont, au début, éloignés à une certaine distance et le ligand doit, au fil de la simulation, trouver sa position optimale dans le site actif de la protéine après un certain nombre de déplacements dans l'espace configurationnel. Ces déplacements correspondent aux translations et aux rotations du ligand, auxquelles on ajoute les changements de géométrie interne du ligand et de la protéine (principalement des angles de torsion) si on considére un docking flexible. Le traitement explicite de la flexibilité est un avantage indéniable de cette approche par rapport à la précédente. Cette approche s'apparente aussi da-

vantage à un problème de reconnaissance moléculaire. Cependant, son inconvénient majeur réside dans le fait que le score du système protéine-ligand doit être calculé après chaque déplacement, sachant qu'il faut réaliser une exploration relativement grande de l'espace configurationnel.

Il existe ainsi plusieurs fonctions de score qui reposent sur une description atomique du système protéine-ligand. Signalons qu'il existe des fonctions de score liées au traitement spécifique d'un type de docking tel que protéine-ligand[62], protéine-protéine[63] ou protéine-ADN[64]. Ces fonctions calculent généralement le score de docking à partir d'une méthode paramétrée ou empirique. Nous avons regroupé ces méthodes selon trois grandes catégories :

Champ de forces Le score est calculé en faisant la somme des forces intermoléculaires de van der Waals et électrostatiques entre tous les atomes du système protéine-ligand. Les énergies intramoléculaires des deux entités sont parfois incluses dans le score final. Pour tenir compte du solvant, les énergies de désolvatation du ligand et de la protéine sont souvent prises en compte par des méthodes de solvatation implicite du type GBSA ou PBSA[61].

Empirique Ces méthodes se basent sur un décompte du nombre d'interactions dans le système protéine-ligand[65] comme par exemple le nombre de liaisons hydrogène, les contacts hydrophobes, les contacts hydrophiles, les contraintes de torsion dues à la formation du complexe protéine-ligand, *etc*. Le score est finalement obtenu, en général, en utilisant des méthodes de régression linéaire multiple.

Connaissance Ces méthodes sont fondées sur des observations statistiques de contacts intermoléculaires proches dans des bases de données 3D (*Cambridge Structural Database* ou *Protein Data Bank*). Il est supposé que les contacts intermoléculaires proches entre certains types d'atome (ou entre certains groupes fonctionnels) plus fréquents qu'une distribution aléatoire conduisent à des interactions énergétiquement favorables, permettant alors d'estimer une affinité de liaison reliée à un potentiel de force moyenne[66]. Cette valeur estimée de l'affinité de liaison est le score de docking.

Il existe aussi des fonctions de score hybrides dans lesquelles plusieurs méthodes de différents types ont été combinées[50,52].

En plus de la conformation optimale, l'objectif d'un programme de docking moléculaire est idéalement de déterminer l'enthalpie libre (ΔG) de formation en solution du complexe $[P : L]$ entre la protéine P et le ligand L, selon :

$$P_{\text{solv}} + L_{\text{solv}} \xrightarrow{\Delta G} [P : L]_{\text{solv}}$$

Quelle que soit l'approche utilisée, les fonctions précédentes nous donnent le score de docking sous la forme d'une valeur, souvent abstraite, représentative de la conformation obtenue du système protéine-ligand. Cette valeur n'est pas reliée directement à une grandeur énergétique du système, sauf dans le cas de l'utilisation d'un champ de forces où le score est donné par les énergies d'interaction intermoléculaire. Des calculs supplémentaires sont donc nécessaires, à partir des conformations optimales, pour déterminer la valeur de l'enthalpie libre de formation.

1.2.2. Algorithme de recherche

Le deuxième élément d'importance d'un programme de docking moléculaire est l'algorithme de recherche. Comme nous l'avons vu précédemment, l'espace de recherche est constitué de toutes les conformations possibles du ligand par rapport à la protéine, augmenté par toutes les conformations internes du ligand (et éventuellement de la protéine) pour un docking flexible. Malgré les avancées technologiques actuelles du matériel informatique, il reste impossible d'explorer de manière exhaustive cet espace de recherche et plusieurs stratégies ont dû être développées pour l'échantillonner efficacement :

- simulations de dynamique moléculaire gros grains, pour suggérer des conformations convenables de systèmes protéine-ligand ;
- analyse de plusieurs structures connues de la même protéine, pour connaître sa flexibilité ;
- utilisation d'algorithmes dits évolutionnaires (comme les algorithmes génétiques), pour permettre le passage d'une conformation à une autre, représentant une meilleure interaction protéine-ligand.

Nous reviendrons plus en détail sur les algorithmes de recherche dans le prochain chapitre.

1.3. Motivations pour le développement d'un nouveau programme

Les approches utilisant une représentation atomique du système moléculaire sont actuellement limitées par une description trop simple des interactions protéine-ligand, reposant essentiellement sur une paramétrisation du champ de forces. Les limitations concernent notamment (*i*) le manque de paramètres adéquats à la description des ligands complexes, (*ii*) la difficulté de prendre en charge convenablement les effets de polarisation électronique et (*iii*) l'impossibilité de traiter les éventuels changements chimiques dans le processus de docking (état de protonation, par exemple).

En principe, l'utilisation d'une description quantique du système (ou d'une partie du système) ligand-protéine peut être un moyen de s'affranchir de ces problèmes. Cependant, le calcul de l'énergie d'un ligand lié à une protéine peut s'avérer être ardu dans ce cas à cause de la taille du système biomoléculaire. En effet, en utilisant les algorithmes standards, les besoins en temps CPU croissent de manière non linéaire avec la taille du système (au minimum en $O\left(N^{2-3}\right)$, où N est le nombre d'orbitales atomiques dans la base[67]). Ceci limite sévèrement le nombre total d'atomes que l'on peut prendre en compte pour représenter les interactions du système protéine-ligand, moins d'une centaine généralement. La taille minimale d'un système protéine-ligand, en considérant uniquement le ligand et les quelques résidus qui interagissent avec lui, dépasse fréquemment cette limite.

Cependant, depuis ces quinze dernières années, de nouveaux algorithmes de chimie quantique ont vu le jour. Leur objectif est d'aboutir à une croissance linéaire[68–70] du temps de calcul en modifiant les algorithmes standards au niveau de leurs parties non linéaires. Ces nouvelles méthodes permettent de modéliser de très grands systèmes tout en conservant un temps CPU raisonnable. Parmi ces méthodes, nous citerons l'exemple de Divide & Conquer[33], MOZYME[34] ou Density Matrix Minimization[35]. Sans entrer dans les détails, rappelons que la méthode Divide & Conquer dérive du caractère local de la densité électronique dans un système moléculaire. A partir de ce constat, il est alors possible de diviser l'ensemble du système en plusieurs sous-systèmes de taille plus petite. La résolution des équations de Roothaan-Hall de chaque sous-système permet de revenir à l'énergie du système entier, après reconstruction de la matrice densité totale (voir annexe A page 228).

Malgré tout, l'utilisation de ces algorithmes à croissance linéaire dans le cadre des méthodes *ab initio* ou DFT est très peu répandue à cause d'un besoin en temps CPU toujours conséquent. Son utilisation dans le cadre des méthodes semi-empiriques est cependant très prometteuse. Les méthodes semi-empiriques sont plus rapides que les méthodes *ab initio* car on y effectue de nombreuses approximations comme le traitement des électrons de valence uniquement, l'utilisation d'une base minimale d'orbitales atomiques et surtout l'absence du calcul de certaines intégrales (approximation ZDO). L'étape limitante est, dans ce cas, la diagonalisation de la matrice de Fock au lieu du calcul des intégrales biélectroniques comme pour les méthodes *ab initio*. Il convient toutefois de vérifier la qualité de la méthode semi-empirique vis-à-vis du système traité afin de s'assurer de la validité du résultat. Par exemple, certaines de ces méthodes ne reproduisent pas correctement les interactions intermoléculaires[71–73]. Des modifications de fonctions cœur-cœur ont été proposées pour pallier ce défaut et rendre possible l'étude de systèmes multimoléculaires[74–76]. Ces méthodes semi-empiriques modifiées ont été appliquées avec succès à une simulation d'eau liquide utilisant la méthode Divide & Conquer[77].

Ces avancées méthodologiques, visant à décrire au niveau quantique des systèmes de grande taille, rendent possible désormais la définition d'une fonction de score quantique. Ceci nous a conduit à développer un nouveau programme de docking moléculaire utilisant une méthode quantique semi-empirique couplée à un algorithme de croissance linéaire pour la description du système protéine-ligand ainsi qu'un algorithme génétique pour l'exploration de l'espace configurationnel.

1.4. Objectifs, cahier des charges du nouveau programme

Dès la conception, nous souhaitions que notre programme de docking puisse utiliser n'importe quel code de calcul quantique existant, et ce, sans avoir à modifier une partie de ce dernier. Plutôt que de parler de couplage entre deux programmes, nous parlerons alors d'interfaçage.

Le programme de calcul quantique que nous avons choisi pour le développement (et les applications qui ont suivi) est DIVCON[78] développé par Merz *et al.*[79–81]. Il permet de réaliser un calcul quantique semi-empirique par un algorithme de croissance linéaire Divide & Conquer. Plusieurs méthodes semi-empiriques sont disponibles dans le programme : AM1[82], PM3[83–85] et PM3-PIF[74–76, 86]. Cette dernière est la méthode PM3 standard corrigée des fonctions cœur-cœur pour permettre une meilleure description des interactions intermoléculaires. Cette modification a été réalisée par le groupe de chimie théorique de Nancy. Elle n'est pas présente dans la version originale du logiciel.

L'algorithme de recherche est la partie la plus importante du travail de développement. Il est l'interface avec le code quantique (DIVCON dans notre cas) et pilote le processus de docking moléculaire. Nous avons choisi d'utiliser un algorithme génétique pour effectuer cette tâche. Le chapitre suivant donne les raisons qui nous ont amenés à faire ce choix et présente en détail son développement.

D'un point de vue technique, le cahier des charges du nouveau programme de docking reprend celui des programmes développés dans le cadre du projet SIRE :

- structure modulaire ;
- programmation de type *OpenSource* ;
- programme de calcul parallèle ou distribué.

Chapitre 2

Développement d'un algorithme génétique de minimisation

Sommaire

2.1. Méthodes de minimisation et atouts des algorithmes génétiques 30
2.2. Principe des algorithmes génétiques 31
2.3. Programmation de l'algorithme génétique 39
2.4. Conclusion 53

Ce chapitre présente l'algorithme génétique que nous avons développé dans le but d'écrire un programme de docking moléculaire semi-empirique à croissance linéaire. Nous ferons, dans un premier temps, un bref rappel des différentes méthodes de minimisation existantes. En comparant les avantages et les inconvénients de chacune d'entre elles, nous expliquerons le choix d'un algorithme génétique. Puis, après une brève description du vocabulaire propre et des notions importantes inhérantes aux algorithmes génétiques, nous présenterons le code informatique que nous avons écrit.

2.1. Méthodes de minimisation et atouts des algorithmes génétiques

Dans la littérature, il existe trois grandes catégories de méthodes de recherche d'un minimum : (*i*) les méthodes nécessitant le calcul du gradient, (*ii*) les méthodes énumératives et (*iii*) les méthodes aléatoires. Nous allons axer notre comparaison des méthodes de minimisation sur le critère de la robustesse, *i.e.* l'équilibre entre l'efficacité (rapport qualité du résultat / temps de calcul) et la stabilité (adaptation vis-à-vis du problème traité).

a. Méthodes avec gradient

Ces méthodes, très largement employées, sont réparties en deux sous-catégories : les méthodes directes comme la méthode des gradients conjugués[87] ou BFGS[88–91], qui recherchent le minimum en partant d'un point initial et en se déplaçant dans la direction donnée par le gradient, et les méthodes de type Newton-Raphson indirectes[92] qui recherchent le minimum en résolvant un système d'équations non-linéaires.

Bien que l'efficacité de ces méthodes n'ait plus besoin d'être démontrée, elles ne peuvent convenir à l'utilisation souhaitée dans le programme de docking. D'une part, ces méthodes sont locales ; le minimum qu'elles cherchent se situe dans le voisinage du point courant. Elles ne permettent pas de vérifier que le minimum global a été atteint. D'autre part, et plus particulièrement pour les méthodes directes, elles requièrent l'existence d'un gradient dont le calcul n'est pas toujours possible. En effet, les fonctions à minimiser peuvent contenir des discontinuités et/ou être très largement multimodales.

b. Méthodes énumératives

Les méthodes énumératives comme BTD[93] (Backtracking with Tree-Decomposition) ont été développées pour s'affranchir du calcul contraignant d'un gradient et espérer obtenir directement le minimum global. L'espace de recherche est discrétisé en un nombre fini de points dont la valeur est calculée. Le minimum global est obtenu en prenant le point de valeur minimale.

Ici encore, nous ne pouvons retenir ce choix de minimiseur en vue du programme de docking à cause de son manque d'efficacité. En général, les espaces de recherche sont souvent très vastes et il faut évaluer un nombre considérable de points pour obtenir un résultat acceptable.

c. Méthodes aléatoires

Les méthodes aléatoires sont devenues de plus en plus populaires là où les deux précédentes méthodes montraient un manque de robustesse. Parmi ces méthodes, nous pouvons citer le recuit simulé[94,95] et les algorithmes génétiques[96,97], entre autres. Dans cette catégorie, on trouve aussi les méthodes de recherche purement aléatoires du type *random walk*[98], mais celles-ci ne sont pas plus efficaces que les méthodes énumératives. Dans toutes ces méthodes, le hasard est utilisé comme un outil dans l'exploration et la recherche d'un minimum. Contrairement aux deux classes de méthodes précédemment citées, les règles de transition d'un point de l'espace de recherche à un autre ne sont plus déterministes, mais probabilistes.

Dans l'optique du programme de docking envisagé ici, et pour les raisons signalées plus haut, le choix du minimiseur s'est porté sur un algorithme génétique. Une des particularités séduisantes des algorithmes génétiques réside dans l'absence d'hypothèses spécifiques sur la régularité de la fonction à minimiser. Par exemple, aucune hypothèse sur la continuité de cette fonction n'est requise, ses dérivées successives n'étant pas nécessaires. Selon Dorsey et Mayer[99], le processus de minimisation est nettement plus simple et offre des performances plus importantes qu'un recuit simulé. Les principaux logiciels de docking disponibles[48–52] utilisent un algorithme génétique comme algorithme de recherche conformationnelle.

Plusieurs codes d'algorithme génétique existent et sont disponibles. Cependant, nous avons remarqué qu'ils ont souvent le défaut majeur d'être liés au problème de minimisation pour lequel ils ont été créés. L'interfaçage d'un de ces codes avec un autre programme semblait être complexe. Ainsi, nous avons décidé d'écrire notre propre programme de minimisation.

2.2. Principe des algorithmes génétiques

Avant de voir en détail le programme qui a été écrit, nous présenterons ici de façon générale les algorithmes génétiques. Nous nous attacherons à définir les différents termes de vocabulaire et à introduire les notions importantes au travers d'un exemple simple de minimisation. La lecture des ouvrages de Holland[96] et de Goldberg[97] pourra compléter cette partie.

2.2.1. Définition et principe général

Un algorithme génétique (GA) est un processus informatique mimant l'évolution biologique naturelle d'une population d'êtres vivants où chaque individu représente une solution potentielle à un problème donné. Comme l'illustre la figure 2.1, ce processus crée une nouvelle population d'individus par une opération de reproduction où chaque individu a hérité d'une partie du patrimoine génétique de ses parents.

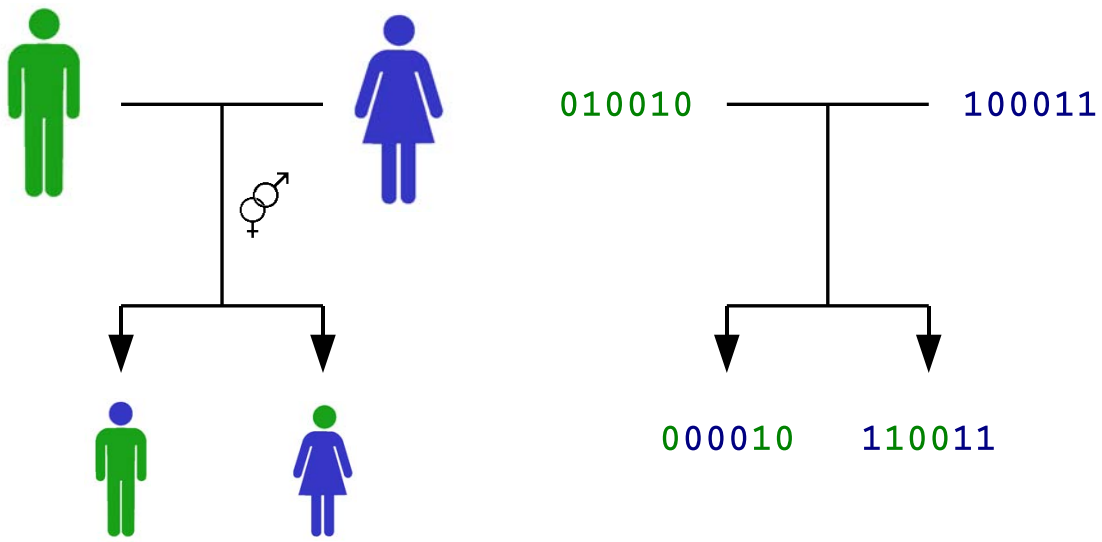


FIG. 2.1: *Analogie entre évolution biologique naturelle et algorithme génétique. La reproduction de deux individus permet de créer deux nouveaux individus où chacun a hérité d'une partie du patrimoine génétique des parents.*

Pour mimer le principe de sélection naturelle, le processus effectue une discrimination entre les individus. Un critère de sélection permet de distinguer ceux qui correspondent à des solutions convenables. Les individus qui ne satisfont pas le critère sont exclus de la population alors que les autres sont combinés pour générer de nouveaux individus.

Le processus de minimisation est illustré par la figure 2.2. Une population initiale d'individus est d'abord évaluée, puis les meilleurs individus sont sélectionnés et, enfin, une nouvelle population est créée par reproduction de ces derniers. Cette population est de nouveau évaluée pour recommencer un autre cycle appelé **cycle de reproduction**. Ainsi, en répétant plusieurs fois les opérations de sélection et de reproduction, nous obtenons successivement des populations d'individus qui convergent vers la solution optimale du problème traité. Pour chaque nouvelle génération produite, il est possible d'extraire le meilleur individu représentant la meilleure solution actuelle au problème.

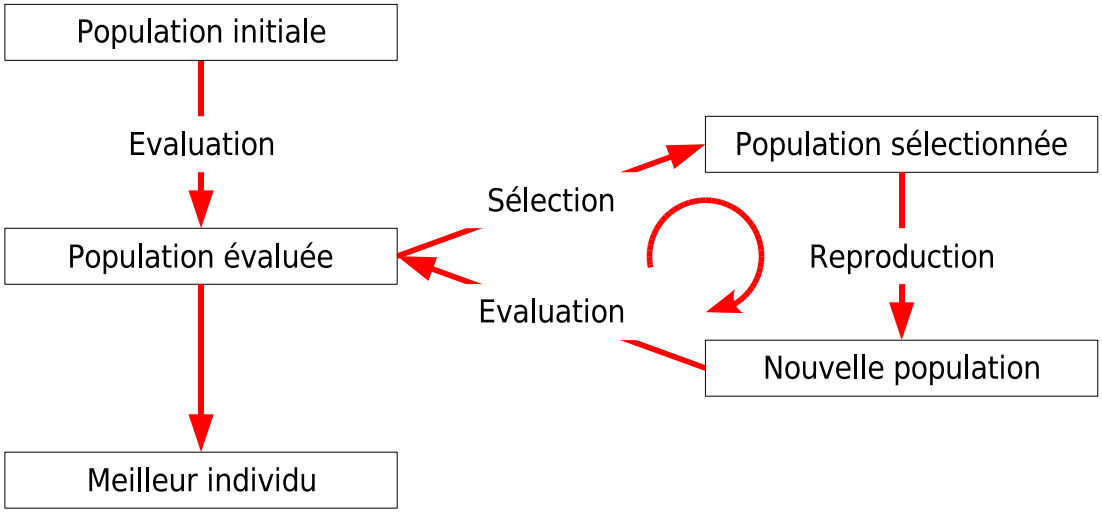


FIG. 2.2: *Schéma illustrant le processus de minimisation dans les algorithmes génétiques. Une population d'individus représente les solutions potentielles d'un problème. Par itération du cycle de reproduction (sélection → reproduction → évaluation), la population converge vers la solution optimale du problème.*

Le matériau de base des algorithmes génétiques est le **gène**. C'est une séquence de symboles qui exprime une caractéristique après lecture de ce code. Ceci conduit à définir deux termes importants, le **génotype** et le **phénotype**. Le premier constitue le code génétique lui-même. Le deuxième est l'expression des caractéristiques ainsi codées. Le passage de l'un à l'autre par une étape essentielle qui est le **décodage**.

Chaque phénotype que l'on souhaite exprimer doit être codé, de manière unique, par l'intermédiaire d'une suite de symboles connus. L'ensemble de ces symboles est appelé **alphabet**. Prenons l'exemple du codage de l'ADN dans le noyau des cellules vivantes. L'alphabet $\mathcal{A}$ est constitué de quatres symboles $\mathcal{A} = \{\mathtt{A}, \mathtt{T}, \mathtt{C}, \mathtt{G}\}$. Ces symboles représentent en fait les quatres nucléotides (ou bases nucléiques) constituant l'ADN : Adénine - Thymine - Cytosine - Guanine. Par la suite et pour plus de simplicité, nous appelerons **bits** les symboles d'un alphabet.

Dans la grande majorité des cas, les caractéristiques d'un individu proviennent de l'expression de plusieurs gènes à la fois. La collection de gènes d'un individu est appelée **génôme**. Ce dernier est représenté physiquement par des entités que l'on appelle **chromosomes**. Un individu peut être constitué d'un grand nombre de chromosomes, l'ensemble des gènes se répartissant sur l'ensemble des chromosomes. Pour des facilités de traitement, l'algorithme génétique que nous considérons dans notre cas suppose que chaque individu ne possède qu'un seul chromosome contenant tous les gènes.

2.2.2. Exemple de minimisation

Nous allons voir d'autres notions importantes des algorithmes génétiques à travers un exemple inspiré du livre de Goldberg[97]. Nous nous proposons de trouver le minimum de la fonction $f(x) = x^2$ sur l'intervalle entier $[0, 31]$ (Fig. 2.3).

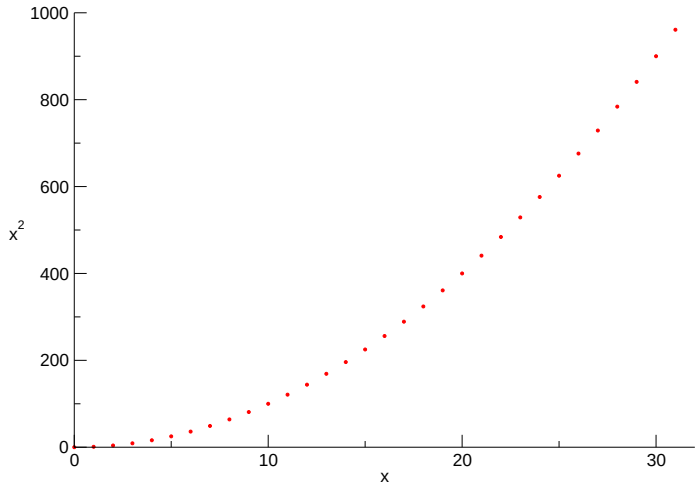


FIG. 2.3: *Fonction $f(x) = x^2$ à minimiser définie dans l'intervalle entier* $[0, 31]$.

a. Fonction de score

Dans un algorithme génétique, il nous faut discerner les individus correspondant à des solutions potentiellement bonnes de ceux à rejeter. Le moyen utilisé pour cela est de donner une valeur numérique d'évaluation à chaque individu. Cette valeur (aussi appelée **score** ou **fitness**) est le résultat d'une fonction, appelée **fonction de score**, appliquée au décodage du chromosome de l'individu. Elle mesure le potentiel de chaque individu à être une bonne solution au problème donné. Ici, dans notre exemple, la fonction de score est considérée comme étant la fonction f à minimiser.

b. Codage

Chaque variable x que nous souhaitons exprimer doit être codée de manière unique dans un gène de taille finie. Il existe plusieurs façons de coder cette variable x comme nous le verrons par la suite. Mais pour le moment, utilisons un codage simple : le codage entier en base 2. De cette façon, chaque valeur entière de l'intervalle $[0, 31]$ est représentée de manière unique par un gène de cinq bits. Par exemple, la valeur 13 sera codée par **01101** car :

$$\mathbf{0} \times 2^4 + \mathbf{1} \times 2^3 + \mathbf{1} \times 2^2 + \mathbf{0} \times 2^1 + \mathbf{1} \times 2^0 = 13$$

Le problème de minimisation que nous avons à traiter ici ne possède qu'une seule dimension. Le chromosome de chaque individu est donc constitué d'un seul et unique gène. Ce dernier est suffisant à l'expression d'une solution potentielle.

c. Population initiale

Avant de démarrer le processus de minimisation, il faut se munir d'une population initiale d'individus. Générons une population aléatoire de quatre individus. Pour cela, tirons à "pile ou face" cinq fois de suite pour construire le chromosome de chaque individu ("pile" $\rightarrow$ **1**, "face" $\rightarrow$ **0**). Le tableau 2.1 montre la population de quatre individus ainsi obtenue et leur évaluation respective.

Individu	Chromosome	x	$f(x)$
1	**01010**	10	100
2	**01100**	12	144
3	**00111**	7	49
4	**10110**	22	484

TAB. 2.1: *Population initiale de quatre individus obtenus aléatoirement et leur évaluation.*

Définissons maintenant les trois opérateurs de base qui permettront de générer les populations successives : (*i*) la sélection, (*ii*) le croisement et (*iii*) la mutation. Nous verrons par la suite qu'il existe différentes manières de réaliser ces opérations mais nous n'utiliserons que la plus simple dans le cadre de cet exemple.

d. Sélection

La sélection est un des éléments les plus importants de tous les algorithmes génétiques. Elle détermine si chaque individu d'une population donnée va pouvoir léguer une partie de son patrimoine génétique à la prochaine génération d'individus. L'objectif principal de toutes les méthodes developpées dans les algorithmes génétiques est de favoriser le passage de ce patrimoine à partir des meilleurs individus. Ce passage est affecté d'une probabilité proportionnelle à l'évaluation propre de chaque individu. En d'autres termes, plus cette valeur est considérée comme bonne, plus il aura de chance d'être choisi et de donner naissance à un individu meilleur. Cet opérateur traduit la sélection naturelle de la théorie de l'évolution des espèces de Darwin[100]. Plus un individu est adapté à l'environnement qui l'entoure, plus il a de chance de survivre et de pérenniser l'espèce.

L'opérateur de sélection peut être construit le plus simplement possible en effectuant une sélection du type "roue de la fortune" (*roulette-wheel selection*) où chaque individu i de la population est associé à une surface de la roue proportionnelle à sa probabilité p_i d'être sélectionné telle que :

$$p_i = \frac{1 - \frac{f(x_i)}{\sum f(x_i)}}{n-1}$$

où $f(x_i)$ est l'évaluation de l'individu i et n le nombre d'individus de la population. La figure 2.4 illustre la répartition de la *roulette-wheel* pour les valeurs calculées dans le tableau 2.2. Sélectionnons trois individus de la population initiale en lançant trois fois cette roue. Les résultats dans le tableau 2.2 sont donnés en nombre de fois où l'individu a été tiré. La population sélectionnée est donc constituée des individus 1, 2 et 3.

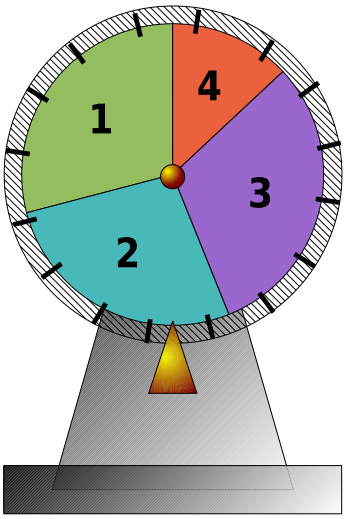


FIG. 2.4: *Roulette-wheel. Chaque individu de la population est représenté sur la roue par une surface proportionnelle à sa probabilité d'être sélectionné. La sélection des individus à reproduire s'effectue en tournant la roue.*

Individu	Chromosome	x	$f(x)$	p_i	Résultat de la sélection par *roulette-wheel*
1	**01010**	10	100	0,29	1
2	**01100**	12	144	0,27	1
3	**00111**	7	49	0,31	1
4	**10110**	22	484	0,13	0
Total			777	1,00	
Moyen			194,25	0,25	
Minimum			49	0,13	
Maximum			484	0,31	

TAB. 2.2: *Evaluation de la population initiale, probabilité p_i pour un individu d'être sélectionné et résultat de la roulette-wheel.*

e. Reproduction

La seconde étape importante de tout algorithme génétique est la reproduction. Durant celle-ci, les individus sélectionnés précédemment engendrent des nouveaux individus. Cette opération se fait en deux sous-étapes : le croisement et la mutation, la plupart du temps combinées.

Durant la phase de croisement, deux individus de la population sélectionnée sont d'abord choisis de façon aléatoire, puis leurs gènes sont copiés et subissent un croisement. Pour cela, une position k, appelée **point de croisement**, est choisie aléatoirement dans l'intervalle $[1, L-1]$, où L est la longueur du gène. Deux nouveaux gènes sont alors créés en permutant tous les bits entre les positions $k+1$ et L inclu.

Par exemple, considérons les individus 1 et 2 de notre population sélectionnée. Supposons que nous obtenons $k = 3$ par tirage aléatoire d'un nombre entre 1 et 4. Nous repérons par une barre verticale rouge cette position dans chaque gène. Le résultat du croisement est la formation de deux nouveaux gènes $1'$ et $2'$.

$$\left.\begin{array}{ll} 1 & \mathtt{010|10} \\ 2 & \mathtt{011|00} \end{array}\right\} \Rightarrow \left\{\begin{array}{ll} \mathtt{010|00} & 1' \\ \mathtt{011|10} & 2' \end{array}\right.$$

Nous remarquons qu'avec un croisement en un seul point, un morceau à l'extrémité des deux gènes est échangé. Il est évident que si nous effectuons une nouvelle fois le croisement de ces deux gènes à une position différente, un morceau à l'intérieur est alors échangé tout en conservant les extrémités initiales des gènes. Cette opération est appelée *2-points crossover*.

La mutation est une opération unaire qui ne prend en compte qu'un seul gène. Elle transforme un bit de celui-ci en tout autre bit de l'alphabet de codage à une position k, appelée **point de mutation**, choisie aléatoirement dans l'intervalle $[1, L]$, L étant toujours la longueur du gène.

Par exemple, considérons l'individu $2'$ obtenu précédemment et un point de mutation $k = 4$ choisi au hasard. Nous repérons par un soulignement rouge cette position. Le résultat de la mutation est la formation d'un nouveau gène $2''$.

$$2' \quad \mathtt{011\underline{1}0} \Rightarrow \mathtt{011\underline{0}0} \quad 2''$$

Par rapport au croisement, l'opérateur de mutation joue un rôle secondaire, mais il est tout de même nécessaire. En effet, les opérateurs de sélection et de croisement suffisent à faire converger la population d'individus vers la solution optimale du problème. Mais il se peut qu'il y ait, au travers de cette population, un manque d'information génétique (défaut de **0** ou de **1**) à certaines positions des gènes. Cette diminution de la diversité génétique peut induire une inefficacité dans la recherche

du minimum[101]. L'opérateur de mutation est donc présent pour pallier cette perte. La mutation est une modification aléatoire et occasionnelle, *i.e.* avec une très faible probabilité de réalisation, d'un bit dans un gène.

A partir des données issues de l'étape de sélection, disponibles dans le tableau 2.2, appliquons les deux opérateurs de reproduction. Nous avons formé deux couples d'individus au hasard (1-3 et 2-3) et choisi aléatoirement un point de croisement pour chaque couple (1-3 en position 2 et 2-3 en position 4). En ce qui concerne la mutation, nous considérons que la probabilité d'effectuer cette opération est très faible, de l'ordre de 1 bit modifié pour 100 ($p_M = 0,01$). Comme il y a 20 bits au total pour les quatre individus, nous aurons $20 \times 0,01 = 0,02$ bits modifiés par cette opération. Ceci implique qu'aucun bit ne soit modifié lors de l'étape de mutation. Une fois l'étape de reproduction effectuée, la nouvelle population d'individus est évaluée. Les résultats sont reportés dans le tableau 2.3.

Individu	Parents	Position	Enfants	x	$f(x)$
1	**01010**	2	**01111**	15	225
3	**00111**		**00010**	2	4
2	**01100**	4	**01101**	13	169
3	**00111**		**00110**	6	36
Total					434
Moyen					108,5
Minimum					4
Maximum					225

TAB. 2.3: *Création d'une nouvelle population d'individus. Les chromosomes de chaque couple d'individus subissent un croisement pour former les individus de la nouvelle génération.*

Après un seul cycle de reproduction, nous observons que le minimum de la fonction $f(x)$ est passé de 49 à 4. De même, la valeur moyenne a diminué de 194,25 à 108,5. Bien que la recherche du minimum n'ait été en fait qu'un tirage de nombres aléatoires, nous remarquons que l'ensemble de la population converge vers le minimum de la fonction.

Les fondements théoriques des algorithmes génétiques ont été mis en équations par J. Holland[96] sous le nom de *Schema theorem*. Ce théorème permet de démontrer qu'il est possible d'atteindre la solution optimale de tout problème par un tel processus.

2.3. Programmation de l'algorithme génétique

Bien que conçu initialement dans le cadre du développement du programme de docking, nous avions le souhait que l'algorithme génétique en soit totalement dissocié. En d'autres termes, ce programme ne s'occupe que de la procédure de minimisation, la partie concernant l'évaluation des individus faisant appel à un autre programme par l'intermédiaire d'un script d'interface. Une conséquence avantageuse de ce principe est que notre code peut être réutilisé de façon simple et transparente dans d'autres applications que le docking. Un grand nombre de paramètres et de fonctionnalités sont aisément redéfinissables par le développeur.

Le choix du langage de programmation s'est porté sur Python[102]. Ce langage fut créé en 1990 par Guido van Rossum et les principales caractéristiques sont :

- langage interprété ;
- multi-paradigme : programmation impérative structurée, orientée objet et fonctionnelle ;
- typage dynamique fort ;
- gestion automatique de la mémoire par l'intermédiaire d'un *garbage collector*.

Le langage Python est placé sous une licence libre proche de la licence BSD et fonctionne sur la plupart des plates-formes informatiques de Linux à Unix en passant par Windows et MacOS. Il est conçu pour optimiser la productivité des programmeurs en offrant des outils de haut-niveau et une syntaxe simple à utiliser. Python possède une grande bibliothèque standard, fournissant des outils convenant à de nombreuses tâches diverses. Les modules de la bibliothèque standard peuvent être augmentés avec des modules spécifiques écrits en C ou en Python.

2.3.1. Structure du code

L'algorithme génétique que nous avons écrit est un module Python nommé `genetic.py` d'environ 1500 lignes. Il comporte 9 classes d'objets, 3 objets et 1 fonction. La figure 2.5 présente le diagramme UML simplifié de ce module.

Les classes `Gene`, `Chromosome`, `Individual` et `Population` représentent respectivement la structure d'un gène, d'un chromosome, d'un individu et d'une population d'individus. Un gène possède un attribut appelé `BinaryCode` qui est une chaîne de caractères correspondant au code génétique. On définit pour cet objet deux fonctions `encode()` et `decode()`, respectivement pour coder et décoder le code génétique. Un chromosome est structuré comme étant une liste de gènes. Nous avons représenté un individu par un objet contenant un chromosome et une valeur numérique d'éva-

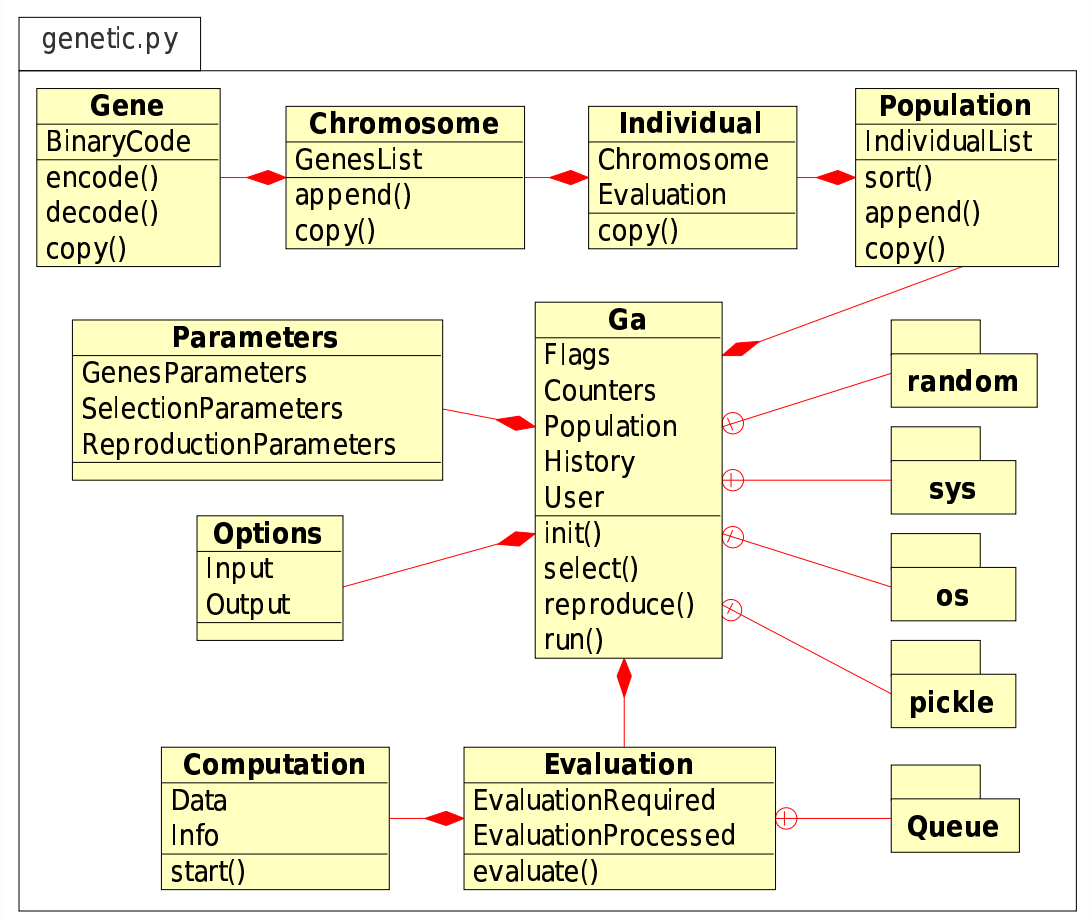


FIG. 2.5: *Diagramme UML simplifié du module* `genetic.py`.

luation, résultat de la fonction de score. Une population est un objet structuré comme étant une liste d'individus et doté d'une fonction `sort()` pour trier les individus selon leur évaluation croissante.

La classe `Ga` est la classe principale de ce module. Pour fonctionner, elle nécessite l'importation de librairies Python standards telles que `random` pour la génération de nombres aléatoires, `os` et `sys` pour la communication avec le système d'exploitation et `pickle` nécessaire à la création de points de sauvegarde. Cette classe contient différents attributs comme `Population` pour la population courante, `History` pour l'historique des individus, *etc*. Il y a aussi les fonctions nécessaires au processus de minimisation comme `init()` pour l'initialisation d'une population, `select()` et `reproduce()` pour la sélection et la reproduction des individus.

Dans le cadre que l'on s'est fixé dès la conception du programme, l'évaluation est totalement dissociée du processus de minimisation. Cette tâche est dévolue à la fonction `evaluate()` de la classe `Evaluation`. Cette dernière est en fait une interface au processus d'évaluation. Elle est constituée de deux queues `EvaluationRequired` et `EvaluationProcessed` issues de la librairie Python standard `Queue`. Pour chaque évaluation requise, un objet `Computation` contenant toutes les informations (`Info`) et données (`Data`) nécessaires au calcul, est créé. Cet objet est échangé avec le script d'interface au programme de calcul pour être traité.

L'ensemble des paramètres et des options nécessaires à la classe `Ga` est stocké dans les deux classes `Parameters` et `Options`. La structure informatique de celles-ci, tout comme celle de la classe `Ga`, est particulière ; il s'agit de singletons. C'est un motif de conception (ou *design pattern*) dont le but est de restreindre l'instanciation d'une classe à un seul objet. Il est utilisé lorsque l'on a besoin d'exactement un objet pour coordonner des opérations dans un système. En d'autres termes, il existe dans ce module un objet `parameters` (instancié de `Parameters`), un objet `options` (instancié de `Options`) et un objet `ga` (instancié de `Ga`), tous trois uniques.

Enfin, la fonction `minimize()` de ce module procède à une minimisation selon les paramètres et options définis et retourne la solution obtenue.

Notre programme, tout en respectant les principes fondamentaux des algorithmes génétiques décrits précédemment, diffère légérement sur certains points, en particulier sur le codage des gènes et les méthodes de sélection et de reproduction. Nous présentons maintenant les principaux points de divergence par rapport à l'algorithme initial et quelques fonctionnalités diverses que nous avons ajoutées. Les différents mots-clés et quelques exemples d'utilisation de ce programme sont présentés dans le manuel utilisateur en annexe B de ce manuscrit.

2.3.2. Codage des gènes

Le codage que nous avons vu précédemment était bien adapté à l'exemple de minimisation, mais il ne pouvait pas convenir à notre programme. Le défaut est qu'il est directement relié au problème de minimisation ; les variables codées par les gènes étaient les variables du problème. De ce fait, il faudrait changer le système de codage pour chaque minimisation à réaliser. Nous décrivons dans la suite le système de codage utilisé par notre programme et les améliorations qui ont été apportées.

a. Système de codage

Avant toute chose, nous souhaitions utiliser un alphabet réduit à un nombre minimum d'élements pour en faciliter la manipulation. Notre choix s'est porté sur l'alphabet binaire $\{\mathbf{0},\mathbf{1}\}$. Partant de cela, il fallait définir une méthode de codage qui puisse convenir à la représentation de valeurs réelles quelconques et faire en sorte que la procédure de décodage soit rapide et fiable. Notre choix s'est porté sur un codage réalisant une homothétie entre un espace réel $\mathbb{R}$, *i.e.* celui des variables à minimiser, et un espace de codage $\mathbb{C}$, *i.e.* celui des variables codées par le chromosome. En d'autres termes, les valeurs codées par les gènes ne sont pas directement les variables du problème. Nous passons de l'une à l'autre par une simple équation linéaire.

La représentation choisie pour les variables de codage est la somme des puissances de 1/2. Soit la variable de codage $x \in \mathbb{C}$ représentée par le gène $\mathbf{a_0 a_1 a_2 \ldots a_n}$ de longueur n avec $\mathbf{a_i} \in \{\mathbf{0},\mathbf{1}\}$, alors :

$$x = \sum_{i=0}^{n} \mathbf{a_i} \frac{1}{2^i} = \mathbf{a_0} \times 1 + \mathbf{a_1} \times \frac{1}{2} + \mathbf{a_2} \times \frac{1}{4} + \ldots + \mathbf{a_n} \times \frac{1}{2^n}$$

Cette écriture nous permet d'obtenir des variables de codage x toujours comprises entre 0 et 2 exclu. Ce principe de codage a aussi l'avantage d'être extensible à l'infini. En effet, la variable de codage est toujours comprise entre 0 et 2 exclu, quelle que soit la valeur de n, la longueur du gène. Nous avons donc, de fait, un moyen simple et pratique de modifier la précision des solutions dans l'espace de recherche. Après le décodage de la variable de codage $x \in \mathbb{C}$, la variable réelle $y \in \mathbb{R}$ est obtenue par une homothétie selon :

$$\mathbb{C} \to \mathbb{R} : \qquad y = \alpha x + \beta \qquad \text{avec} \qquad \begin{cases} \alpha = \dfrac{y^+ - y^-}{2} \\ \beta = y^- \end{cases} \qquad \text{où} \qquad \mathbb{R} = \left[y^-, y^+\right[$$

Le programme réalise principalement des procédures de décodage $\mathbb{C} \to \mathbb{R}$, essentiellement lors de l'étape d'évaluation de la population. Le codage $\mathbb{R} \to \mathbb{C}$ se déroule selon une procédure inverse et

a été développé dans le code du programme. Il est utilisé uniquement dans le cas où un gène doit coder une valeur donnée (contrainte).

b. Code Gray

L'inconvénient majeur du codage précédent est que la modification d'un seul bit (**0** $\leftrightarrow$ **1**) peut produire un changement important de la valeur de codage x exprimée.

Exemple Modification du troisième bit d'un gène de longueur 6.

Gène	**100101**	$\longleftrightarrow$	**101101**
x	1,15625		1,40625

La distance de Hamming[103] est définie comme étant la valeur absolue de la différence entre les deux valeurs exprimées avant et après modification d'un bit. Ce changement important provient du fait que tous les bits n'ont pas le même poids à l'intérieur du code binaire. L'influence de la modification d'un bit sur la distance de Hamming dépend donc du poids du bit considéré.

L'utilisation d'un code Gray[104] permet de remédier en partie à ce problème. Il s'agit d'un code binaire, totalement artificiel, où la modification d'un seul bit entraîne un changement plus petit de la valeur exprimée. Ce code est obtenu par le résultat de la somme de deux bits contigüs du code binaire initial modulo 2, le premier bit du code Gray étant le premier bit du code binaire. Plus concrètement, soit un code binaire A tel que $A = \mathbf{a_0 a_1 a_2} \ldots \mathbf{a_n}$, alors le code Gray $G = \mathbf{g_0 g_1 g_2} \ldots \mathbf{g_n}$ correspondant sera tel que $\mathbf{g_0} = \mathbf{a_0}$, $\mathbf{g_1} = \mathbf{a_0} \oplus \mathbf{a_1}$, $\mathbf{g_2} = \mathbf{a_1} \oplus \mathbf{a_2}$, ... , $\mathbf{g_n} = \mathbf{a_{n-1}} \oplus \mathbf{a_n}$ où $\oplus$ est l'opérateur logique "ou exclusif". Le passage d'un code Gray à un code binaire s'effectue selon les mêmes opérations.

Exemple Modification du troisième bit d'un gène de longueur 6 utilisant un code Gray.

Gène	**100101**		**101010**
Code Gray	**110111**	$\longleftrightarrow$	**111111**
x	1,15625		1,3125

L'exemple précédent illustre bien l'utilité de passer par un code Gray pour le codage des gènes. Ainsi, l'opération de mutation entraîne un plus petit changement de la valeur exprimée par le gène.

2.3.3. Sélection

Dans la méthode de sélection présentée à l'origine par Goldberg[97], la probabilité qu'un individu soit sélectionné est proportionnelle à son évaluation. Bien que cette méthode soit très efficace, elle peut être considérée comme insatisfaisante. En se souvenant du fonctionnement de la *roulette-wheel*, il paraît évident que si un individu possède une probabilité de sélection p_i très grande par rapport à celle de ses congénères, la population sélectionnée peut ne contenir que cet individu. Ceci induirait sans doute une perte importante d'information génétique, assez difficilement compensable par mutation. L'opérateur de sélection doit donc tenir compte des deux principes suivants : (*i*) permettre la convergence de la population vers la solution optimale du problème et (*ii*) éviter le phénomène d'eugénisme (reproduction de génotypes identiques) en sélectionnant un maximum d'individus différents.

La méthode de sélection choisie dans cet algorithme génétique est la combinaison de deux méthodes : l'élitisme et les tournois. La première élimine, de manière grossière, les individus les moins intéressants car étant très loin de la solution optimale du problème. La deuxième constitue des petits groupes d'individus et sélectionne le meilleur de chaque groupe. Nous espèrons convenir de cette manière à ce qui a été énoncé précédemment.

a. Sélection par élitisme

L'élitisme est une suppression dirigée de certains individus de la population. Seuls les individus ayant le meilleur score sont conservés, les autres sont supprimés de la population. En effet, la probabilité qu'un individu soit sélectionné est d'autant plus grande que son score est bon. Nous éliminons donc, de façon arbitraire, tous les individus jugés inaptes à engendrer de bonnes solutions. Cependant, l'élitisme utilisé seul n'est pas une méthode de sélection satisfaisante. S'il est trop important, nous nous acheminons rapidement vers un eugénisme défavorable à l'efficacité de l'algorithme génétique.

Le nombre d'individus conservés est exprimé en pourcentage de la population totale, le taux d'élitisme. Pour obtenir ces individus, il suffit de trier la population dans l'ordre croissant des valeurs d'évaluation et de conserver uniquement la quantité suffisante.

b. Sélection par tournois

Cette méthode de sélection est inspirée de celle développée par Miller et Goldberg[105] en 1995. La population pré-sélectionnée par élitisme est répartie aléatoirement dans un nombre de groupes égal au nombre d'individus à obtenir en fin de sélection. Le meilleur de chaque groupe (ou tournoi) est placé dans la sélection finale des individus à reproduire.

La figure 2.6 illustre le principe de cette sélection avec une population initiale de 12 individus numérotés de 1 à 12. Pour obtenir au final trois individus à reproduire, les trois groupes d'individus suivant sont formés : {9, 7, 3, 6}, {2, 8, 5, 11} et {1, 12, 10, 4}. Dans chaque tournoi, les individus sont triés selon l'ordre croissant des évaluations et l'individu correspondant à la plus petite évaluation est placée dans la sélection finale. Nous obtenons ainsi la sélection des individus 7, 2 et 10.

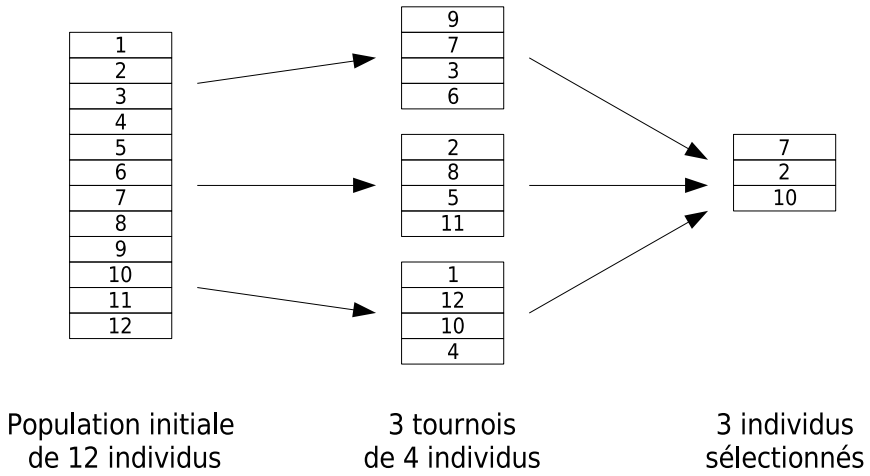


FIG. 2.6: *Schéma de fonctionnement de la sélection par tournois. Les 12 individus de la population initiale sont répartis aléatoirement dans 3 groupes différents. Le meilleur de chacun de ces tournois est sélectionné pour les opérations de reproduction.*

Nous constatons que cette méthode est en adéquation avec notre volonté de conserver une certaine diversité génétique au sein de la population. Si le nombre de groupes créés est suffisamment grand, les individus dont l'évaluation n'est pas excellente, sans être forcément très mauvaise, peuvent être sélectionnés pour la reproduction.

2.3.4. Reproduction

Comme nous l'avons vu précédemment, la reproduction est essentielle à l'algorithme génétique pour explorer de nouvelles régions de l'espace de recherche. Plus cette reproduction est efficace,

i.e. création à chaque génération d'individus nouveaux ayant une évaluation meilleure que celle des parents, plus la recherche de la solution optimale est rapide.

Une des conclusions du *Schema theorem* de Holland[96] est que la recherche d'un minimum par un algorithme génétique s'effectue de manière optimale si les opérations de reproduction ne perturbent pas de façon trop importante le processus global de minimisation. En d'autres termes, ces opérations doivent s'effectuer sur des gènes de la plus petite taille possible.

Pour s'affranchir de cette règle, Goldberg[97] propose alors de ne plus considérer les opérations de reproduction sur le gène en entier, mais sur des partitions de celui-ci. Il nomme cette manière de rendre efficaces les opérateurs de reproduction pour les gènes de grande taille sous le terme de *Building blocks hypothesis*. Nous effectuons la reproduction de deux individus (croisement et mutation) en divisant les gènes en plusieurs parties dans le but d'être en accord avec le principe qui vient d'être énoncé.

a. Croisement

Nous effectuons un ou plusieurs croisements sur des parties du gène grâce au partitionnement de celui-ci et chaque croisement est affecté d'une probabilité p_C de se réaliser. Prenons, par exemple, deux gènes A et B de longueur $L = 8$. Nous décidons de créer deux partitions de longueur $L_i = 4$ sur chaque gène que nous notons $[\quad]_i$ avec i, le numero de partition. En déterminant un point de croisement k choisi au hasard pour chaque partition i dans l'intervalle $[1, L_i - 1]$, nous obtenons les deux nouveaux gènes A' et B'.

$$\left.\begin{array}{llll} A & \mathtt{01101000} & \rightarrow & [\mathtt{0110}]_1\,[\mathtt{1000}]_2 \\ B & \mathtt{10011101} & \rightarrow & [\mathtt{1001}]_1\,[\mathtt{1101}]_2 \end{array}\right\} \Rightarrow \left\{\begin{array}{llll} [\mathtt{0001}]_1\,[\mathtt{1100}]_2 & \rightarrow & \mathtt{00011100} & A' \\ [\mathtt{1110}]_1\,[\mathtt{1001}]_2 & \rightarrow & \mathtt{11101001} & B' \end{array}\right.$$

b. Mutation par permutation

Cette mutation correspond à celle utilisée dans l'exemple de minimisation. Elle est réalisée sur chaque gène A' et B' après l'opération de croisement en effectuant la permutation de bit $\mathtt{0} \leftrightarrow \mathtt{1}$. La façon de procéder est la même que pour le croisement, *i.e.* une ou plusieurs partitions sont créées pour chaque gène sur lesquelles une ou plusieurs mutations peuvent se réaliser avec une probabilité p_M.

c. Mutation par addition

Nous avons apporté à ce programme une amélioration concernant l'opération de mutation. Dans cette approche, nous ne considérons plus la mutation comme une permutation de bit **0** $\leftrightarrow$ **1**, mais comme une opération d'arithmétique binaire appliquée au code binaire du gène. Les conditions de réalisation de l'opération restent les mêmes dans le sens où il est toujours nécessaire de déterminer aléatoirement un point de mutation k et une probabilité de réalisation. Cependant, il ne faut pas introduire de partionnement du gène pour que cette opération conserve un sens.

La mutation par soustraction est le complément de cette méthode. Elle a aussi été aussi développée dans le programme, mais elle ne sera pas décrite ici par souci de simplicité.

Voyons au travers d'un exemple, l'effet de cette mutation sur un gène A de longueur $L = 9$. La valeur de $k = 5$ obtenue aléatoirement entraîne la création d'un gène A_k, de même longueur que A, où la position k est **1**, les autres positions étant **0**. Le gène muté A' est obtenu en faisant l'addition binaire des deux codes génétiques selon :

$$\begin{array}{rcl} & A & \mathbf{01101000} \\ + & A_k & \mathbf{00001000} \\ \hline = & A' & \mathbf{01110000} \end{array}$$

Nous constatons, sur cet exemple, que cette technique de mutation peut s'avérer avantageuse. Une seule mutation a permis de modifier localement deux bits du code génétique, là où la méthode par permutation aurait demandé deux mutations successives.

Le développement de ce nouvel opérateur de mutation visait, d'une part, à introduire les propriétes de périodicité de certaines fonctions mathématiques qui ne pouvaient être respectées compte tenu du codage utilisé et, d'autre part, à éviter l'utilisation d'un code Gray pour le codage des gènes. Nous intégrons, dans le cas présent, une notion de métrique entre les gènes. Prenons l'exemple des deux gènes A et B ci-dessous :

	Code	Variable de codage
A	**010000**	$0,5$
B	**001111**	$0,46875$

Nous constatons que les deux variables de codage sont relativement proches (elles sont contigües dans l'espace de recherche) alors que les deux codes génétiques sont relativement différents. Pour passer de B à A, il existe deux solutions : (i) faire cinq mutations successives sur les cinq derniers

bits de B ou (ii) ajouter au code de B le codage de la différence existante entre A et B, en termes de variable de codage selon :

$$\begin{array}{rrll} & B & \mathtt{001111} & 0,46875 \\ + & & \mathtt{000001} & 0,03125 \\ \hline = & A & \mathtt{010000} & 0,5 \end{array}$$

Nous remarquons que cette méthode de mutation permet l'exploration de l'espace de recherche en passant par des points contigüs de celui-ci. Mais, l'avantage le plus important de ce nouvel opérateur est de pouvoir respecter la périodicité de certaines fonctions. En fait, la périodicité est induite par le codage en lui-même, *i.e.* chaque variable de codage est définie modulo α. Cette quantité α est la borne supérieure de $\mathbb{C}$, l'espace des variables de codage :

$$\alpha = \sum_{i=0}^{n} \frac{1}{2^i} + \frac{1}{2^n} = 2$$

quel que soit n la taille du gène considéré. En effet, nous pouvons écrire en démonstration que :

$$\begin{array}{rll} & \sum_{i=0}^{n} \frac{1}{2^i} & \mathtt{111...11} & 1,999...99 \\ + & \frac{1}{2^n} & \mathtt{000...01} & 0,000...01 \\ \hline = & & \mathtt{000...00} & 0,0 \end{array}$$

Donc, l'ajout du codage maximal **`111...11`** au code d'un gène A quelconque est équivalent à la soustraction du codage **`000...01`** :

$$\begin{array}{rrl} & A & \mathtt{011010} \\ + & & \mathtt{111111} \\ \hline = & & \mathtt{011001} \end{array}$$

La méthode de mutation par addition est aussi utilisée pour une recherche affinée de la solution optimale du problème, une fois le minimum global localisé.

L'utilisation de la méthode de mutation par addition ne convient pas à l'écriture du gène sous la forme d'un code Gray. Ces deux fonctionnalités sont exclusives l'une de l'autre, *i.e.* elles ne peuvent pas être utilisées ensemble dans le programme.

Pour permettre l'exploration de l'espace de recherche par des points contigüs, nous avons développé un sélecteur de position de mutation spécifique afin que la mutation par addition s'effectue plus fréquemment sur les bits de poids faible. Ce sélecteur suit la pondération des bits du gène. Soit p la probabilité de mutation, la position k d'un gène de longueur n est choisie si $\frac{\sum_{i=1}^{n-k} 2^i}{\sum_{i=1}^{n} 2^i} < p$.

d. Inversion

L'inversion est un troisième opérateur possible de reproduction. Il est mentionné par Goldberg[97], mais n'a pas été utilisé dans son algorithme car, selon lui, il crée une rupture dans la recherche de la solution optimale. Nous l'avons tout de même implanté dans notre algorithme afin de pallier une perte de matériel génétique qui, nous le savons, favorise l'eugénisme. Cette fonction est typiquement dévolue à l'opérateur de mutation. Mais, du fait que cette dernière est affectée d'une probabilité très faible, il est difficilement concevable qu'elle puisse combler des lacunes souvent importantes. L'utilité de l'inversion est donc de fournir très occasionnellement un bloc de bits provenant d'un autre gène afin de combler ces lacunes.

Dans le processus de reproduction, l'inversion intervient juste avant le croisement. Après le choix des deux individus à reproduire, nous effectuons la permutation du codage entier de deux gènes choisis au hasard sur l'un des deux individus. Cette opération est affectée d'une probabilité d'inversion p_I très faible.

2.3.5. Stratégie d'évolution

La nature et l'efficacité d'un algorithme génétique dans la recherche de la solution optimale dépend de la stratégie d'évolution employée. Chaque stratégie détermine la façon dont la nouvelle population est créée à chaque cycle de reproduction. Nous avons intégré quatre stratégies d'évolution dans notre algorithme génétique.

La première stratégie, appelée `NewOne`, est celle que nous avons utilisée dans l'exemple de minimisation. Elle correspond à la manière la plus simple de créer une nouvelle population. Les individus de la nouvelle génération sont obtenus par croisement et mutation du chromosome des parents. La deuxième stratégie, appelée `NewUnknown`, est une amélioration du principe d'évolution précédent. Elle consiste à sélectionner les individus considérés bons d'une population et à s'en servir pour créer une nouvelle population par reproduction. On s'assure durant cette dernière phase de ne produire que des individus originaux, *i.e.* n'ayant jamais été produits dans les générations précédentes. Le but de cette stratégie est d'offrir à l'évaluation une population d'individus avec des caractéristiques nouvelles pour converger plus rapidement vers la solution optimale. La troisième stratégie, appelée `UnknownSurvivor`, fait intervenir la notion de pérénnité des individus. Elle repose sur la stratégie précédente, sauf que les individus sélectionnés sont placés dans la nouvelle population. Cette popu-

lation est complétée par des individus nouveaux issus de la reproduction des individus sélectionnés. Cette stratégie a pour but de conserver les meilleurs individus de génération en génération. Elle prévient ainsi contre des opérations de sélection et de reproduction néfastes, *i.e.* conduisant à la création d'une population d'individus moins bons que les parents. La dernière stratégie, appelée `OldSurvivor`, est une amélioration de la stratégie `UnknownSurvivor`. Elle applique la notion d'âge et de vieillissement à chaque individu. A chaque génération, un individu survivant voit son âge augmenté. Si cet âge dépasse une valeur donnée, l'individu est supprimé. Cette amélioration permet d'éviter une trop grande pérénnité des meilleurs individus dans les nouvelles générations.

D'autres stratégies ont été développées par la communauté scientifique[106–108]. Elles miment pour la plupart des phénomènes existant dans le monde vivant. L'algorithme génétique que nous avons développé permet à l'utilisateur et aux futurs développeurs de programmer leur(s) propre(s) stratégie(s) grâce à l'utilisation des primitives de sélection et de reproduction.

2.3.6. Population initiale

La création de la population initiale d'individus est une opération importante de l'algorithme génétique. L'efficacité de la recherche dépend de la diversité génétique des individus de cette première population[101]. Nous présentons les trois méthodes développées pour la création de la population initiale.

a. Population aléatoire

Cette méthode, décrite à l'origine par Goldberg[97], consiste à construire les chromosomes des individus de la population initiale de façon totalement aléatoire. Cette solution permet d'obtenir une population initiale ayant une grande diversité génétique. Mais ceci nécessite d'avoir une taille de population assez importante pour espérer avoir une distribution (pratiquement) uniforme de points dans l'espace de recherche.

Les différents bits du code génétique d'un individu sont déterminés de manière équiprobable. Cette opération est répétée jusqu'à atteindre la taille du chromosome, somme de la taille des gènes qu'il contient. Nous construisons ainsi autant d'individus qu'il est nécessaire pour former la population initiale.

b. Solution approchée

Cette deuxième possibilité d'initier une population est contraire au principe qui veut que la population initiale représente la plus grande diversité génétique possible. Nous connaissons parfois une solution approchée à un problème, solution qui peut provenir par exemple, d'un précédent calcul de minimisation. Il devient donc inutile de chercher une solution dans des régions potentiellement mauvaises de l'espace de recherche. La finalité de cette démarche est donc de créer une population d'individus localisés dans une région de l'espace de recherche et de trouver la solution optimale près de ces points. Une solution possible serait de restreindre le domaine de variation de chaque variable du problème et de générer une population aléatoire dans cet espace de recherche. Cette solution ne peut être satisfaisante car l'exploration de l'espace est elle-aussi limitée.

La méthode que nous proposons consiste à coder la solution approchée au problème dans le chromosome d'un individu. Les autres individus de la population initiale sont générés par une légère mutation des gènes de ce premier individu. Nous construisons de cette manière une population d'individus localisés autour d'un point tout en s'assurant de l'exploration de l'espace de recherche en entier.

c. Sauvegarde de la population

Durant le processus de minimisation, un fichier de sauvegarde est créé à intervalles réguliers. Ce fichier contient la population en cours d'existence au moment de la sauvegarde. Il suffit alors d'indiquer le nom du fichier de sauvegarde pour reprendre une minimisation à partir de la population sauvegardée.

2.3.7. Evaluation

Dans le but de rendre cet algorithme génétique de minimisation totalement indépendant du problème traité, nous avons opté pour un processus d'évaluation séparé du processus de production des individus. En d'autres termes, l'évaluation d'un individu s'effectue dans un autre processus, sans lien direct avec l'individu concerné. Cette partie présente l'interface d'évaluation qui a été développée et différentes fonctionnalités ajoutées au programme concernant l'évaluation.

a. Interface d'évaluation

Le module d'évaluation de l'algorithme génétique est conçu pour permettre l'évaluation de la population par un programme externe. Lors de l'étape de décodage du chromosome de chaque individu de la population, les valeurs décodées sont placées dans une première queue de calcul appelée `EvaluationRequired` et associées à un identifiant numérique. L'interface entre l'algorithme génétique et le programme externe s'effectue par l'intermédiaire d'instances de la classe `Computation`. Elles permettent de stocker les valeurs décodées du chromosome de chaque individu, le résultat de l'évaluation et diverses informations concernant le calcul à réaliser. Une instance de cette classe est créée pour chacun des calculs mis en attente d'exécution dans la queue de calcul. Le processus de production des individus est à ce moment suspendu, *i.e.* en attente que l'évaluation de tous les individus soit réalisée.

Le programme externe récupère toutes les demandes d'évaluation présentes dans la queue de calcul. Il traite le calcul selon les données et les informations qui sont contenues dans l'objet `Computation` et y place le résultat à être retourné à l'algorithme génétique. Une fois toutes les évaluations effectuées, l'algorithme génétique reprend la main. Tous les résultats des calculs effectués sont placés dans une deuxième queue appelée `EvaluationProcessed`, toujours associés avec l'identifiant numérique. Finalement, les individus de la population reçoivent leurs évaluations respectives en leur attribuant le résultat du calcul associé à leur identifiant.

b. Calcul distribué

Les algorithmes génétiques sont, par nature, adaptés à l'exécution de calculs distribués. Lors de la phase d'évaluation, une grande quantité de calcul doit être effectuée. Il est alors astucieux de répartir ces calculs sur plusieurs processeurs en même temps dans le but de gagner du temps. L'interface d'évaluation que nous venons de voir possède l'avantage de pouvoir facilement répartir plusieurs processus d'évaluation en simultané sur une plateforme de calcul distribué (clusters de PC, nœuds de processeurs, ...) sans que cela n'interfère avec le processus de production des individus.

Il existe une grande diversité des plateformes informatiques et la gestion des calculs distribués est souvent liée à l'architecture propre du système d'exploitation. Nous n'avons donc pas intégré directement la distribution des calculs dans l'interface d'évaluation. Cette tâche est dévolue à un programme spécialisé externe. Nous avons utilisé SGE[109] (*Sun Grid Engine*) pour les différents tests que nous avons réalisés. Il s'agit d'un programme libre (*OpenSource* en licence SISSL) d'ordonnan-

cement de tâches informatiques (*batch-queuing system*) soutenu par la société Sun Microsystems. Il permet l'ordonnancement, la répartition et la gestion des tâches séquentielles, parallèles ou interactives sur des clusters de PC. L'algorithme génétique ne gère pas directement le programme SGE. Une fonction Python externe réalisant cette tâche est interfacée à l'algorithme génétique par l'intermédiaire de l'attribut `ResultsGrabber`.

c. Historique des individus

Pour diminuer au maximum le nombre d'évaluations à effectuer lors des cycles de reproduction, il peut être utile de stocker les calculs déjà réalisés. De cette manière, si nous avons à évaluer un individu dont le chromosome est connu, il suffit de rechercher l'évaluation correspondante dans une base de données.

Cette base de données est implémentée dans notre algorithme génétique et prend le nom d'historique des individus. Il contient l'ensemble des chromosomes produits par l'algorithme et l'évaluation correspondante. Cette recherche est effectuée à la création de l'individu, à la fin des opérations de reproduction.

Lors de la préparation de la queue de calcul, les individus déjà évalués (récupérés dans l'historique) sont stockés dans une structure temporaire. On les ajoute ensuite aux individus fraichement évalués en fin de processus d'évaluation.

2.4. Conclusion

Toutes les fonctionalités décrites ci-dessus ont été programmées et intégrées dans le logiciel que nous avons créé : AlgoGen. L'écriture de celui-ci, en Python, est suffisament générale pour être utilisée dans d'autres cadres que celui de l'optimisation de ligands dans le site actif d'enzymes. Le manuel d'utilisation du programme est reporté en annexe B.

Chapitre 3

Interface AlgoGen-DivCon

Sommaire

3.1. Présentation générale 56
3.2. Structure du code 57
3.3. Déroulement du docking 60

Ce chapitre présente le programme de docking moléculaire semi-empirique à croissance linéaire que nous avons développé en interfaçant l'algorithme génétique (AlgoGen) décrit dans le chapitre précédent et le programme DivCon. Nous nous sommes attachés, dans cette phase de développement, à réaliser un programme de docking rigide, plus simple à mettre en œuvre. Après avoir décrit les spécificités de notre programme, nous parlerons plus en détail de la partie développement informatique dans laquelle nous présenterons sa structure. Nous terminerons cette présentation par une description du fonctionnement du programme.

3.1. Présentation générale

Le programme de docking semi-empirique à croissance linéaire que nous avons développé combine ALGOGEN, pour la partie minimisation, et DIVCON, pour la partie énergie potentielle du système moléculaire. Il réalise le docking d'un ligand rigide dans le site actif rigide d'une protéine, *i.e.* les géométries internes de la protéine et du ligand sont gelées durant le processus. Il a été décidé, en outre, que la position des atomes de la protéine soit fixe. La position et l'orientation du ligand par rapport à la protéine sont ainsi optimisées en minimisant l'énergie totale du système. De ce fait, nous avons un système moléculaire qui possède six degrés de liberté :

- trois degrés pour la translation du ligand, décrite par un vecteur (x, y, z) ;
- trois degrés pour la rotation du ligand, décrite par les trois angles d'Euler ϕ, θ et ψ.

Dans l'algorithme génétique, chaque individu possède donc un chromosome contenant six gènes correspondant à ces six degrés de liberté. Les domaines de définition de chaque variable sont :

- $\left[-L/2, L/2\right[$ pour les variables de translation, où L est la taille de la boîte (cubique) où s'effectue le docking (boîte de docking) ;
- $[-\pi, \pi[$ pour les variables de rotation ϕ et ψ, $\left[-\pi/2, \pi/2\right[$ pour la variable de rotation θ.

La fonction de score de notre programme de docking rigide est l'énergie d'interaction du système protéine-ligand. Elle est calculée comme étant l'énergie totale de celui-ci à laquelle on retranche l'énergie de chaque entité isolée.

Le programme effectue les tâches suivantes :

- la lecture de la structure de la protéine et du ligand sous la forme de deux fichiers distincts au format PDB ;
- l'extraction de la partie d'intérêt de la protéine, *i.e.* le site actif de celle-ci et son environnement plus ou moins proche ;
- le réglage des différents paramètres et options de l'algorithme génétique ;
- le réglage des différents paramètres et mots-clés du programme DIVCON ;
- la création des fichiers d'input de DIVCON et la lecture des fichiers de résultat ;
- la création des fichiers de log et de visualisation ;
- l'exécution du docking.

3.2. Structure du code

Le programme AlgoGen-DivCon se compose de plusieurs modules Python (aussi appelés scripts) dont nous faisons la liste :

- `genetic.py` : algorithme génétique de minimisation ;
- `docking.py` : outils de construction des systèmes moléculaires et de définition de la fonction de score de docking ;
- `quaternion.py` : outils de manipulation des systèmes moléculaires ;
- `divcon.py` : interface du programme DivCon.

Cette conception en structure modulaire est un avantage de ce programme. Les différentes fonctions sont séparées dans des modules différents : `genetic.py` pour l'algorithme de recherche conformationnelle (minimisation), `docking.py` et `quaternion.py` pour la gestion du système protéine-ligand et `divcon.py` pour le calcul de l'énergie potentielle de ce système. Il en résulte une certaine flexibilité d'utilisation et de développement ultérieur. Il est possible de remplacer chaque module fonctionnel par un autre de même type. Par exemple, le calcul de l'énergie du système protéine-ligand peut être effectué par un autre programme au choix de l'utilisateur à condition d'en écrire l'interface adéquate.

3.2.1. Module `docking.py`

Ce module Python d'environ 1000 lignes contient trois classes d'objet. La figure 3.1 présente le diagramme UML simplifié de ce module.

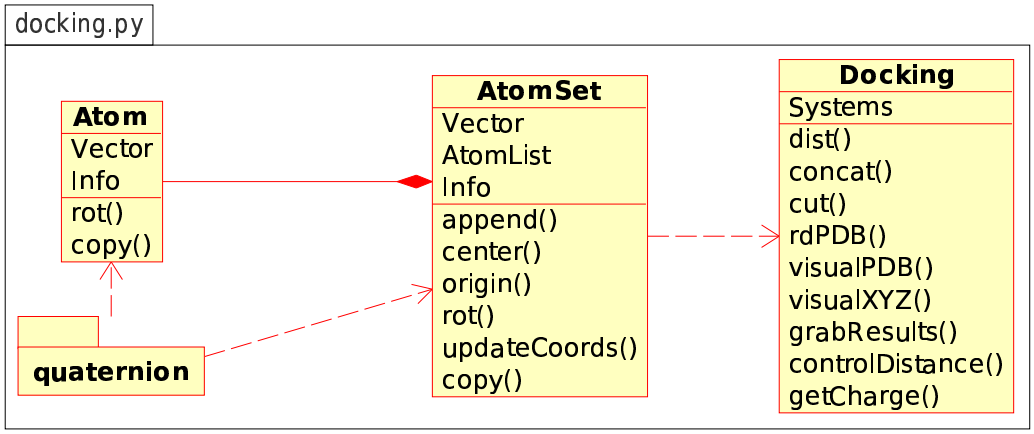


Fig. 3.1: *Diagramme UML simplifié du module `docking.py`.*

La classe `Atom` représente la structure d'un atome. Un objet instancié de cette classe possède un attribut `Vector` qui est son vecteur position sous la forme d'un quaternion imaginaire pur et un attribut `Info` sous la forme d'un dictionnaire où sont stockées toutes les autres informations relatives à la description de cet atome (élément, masse, type, numéro de résidu, ...). Il possède également les fonctions `rot` pour effectuer une rotation et `copy` pour obtenir une copie de l'atome.

La classe `AtomSet` représente un système d'atomes. Il est structuré comme étant une liste d'atomes (`AtomList`) et possède deux autres attributs qui sont `Vector`, le vecteur position du centre géométrique sous la forme d'un quaternion imaginaire pur et `Info`, un dictionnaire où sont stockées les informations relatives à la description du système d'atomes. Plusieurs fonctions sont disponibles pour chacun des objets instanciés de cette classe :

- `append` : ajoute un nouvel atome dans le système ;
- `center` : calcule le centre géométrique du système et met à jour les coordonnées des atomes ;
- `origin` : place l'origine du repère orthonormé du système sur un atome donné et met à jour les coordonnées des atomes ;
- `rot` : effectue la rotation de tous les atomes du système par un quaternion donné ;
- `updateCoord` : effectue la translation de tous les atomes du système par rapport à un nouveau centre géométrique du système ;
- `copy` : renvoie une copie du système.

La classe `Docking` permet la manipulation des systèmes moléculaires et le traitement du processus de docking. Elle possède un attribut `Systems` qui garde en mémoire les structures des systèmes moléculaires. Plusieurs fonctions ont été ajoutées à ce module dont les principales sont :

- `dist` : calcule la distance entre deux atomes ;
- `concat` : fusionne deux systèmes d'atomes ;
- `cut` : élimine les atomes du système n'étant pas contenu dans une sphère de coupure ;
- `rdPDB` : crée un système d'atomes à partir d'un fichier PDB ;
- `visualPDB` et `visualXYZ` : écrit un fichier PDB ou XYZ correspondant à un système d'atomes ;
- `grabResults` : décrit la façon dont les calculs d'énergie potentielle sont effectués ;
- `controlDistance` : vérifie la structure des systèmes à évaluer ;
- `getCharge` : calcule la charge totale des systèmes.

Ce module importe un autre module Python, appelé `quaternion.py`, que nous avons écrit pour gérer les transformations relatives aux systèmes d'atomes.

3.2.2. Module `quaternion.py`

Le formalisme utilisé dans le programme pour effectuer la rotation du ligand est celui des quaternions. Il s'agit d'un type de nombres hypercomplexes à quatre composantes, constituant une extension des nombres complexes à deux composantes. Soit $Q = a + \mathbf{i}b + \mathbf{j}c + \mathbf{k}d$ un quaternion, on appelle a le scalaire de Q et le triplet $\{b, c, d\}$ le vecteur de Q, où $\mathbf{i}$, $\mathbf{j}$ et $\mathbf{k}$ des nombres imaginaires tels que $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$.

Tout vecteur $\mathbf{v}(x, y, z)$ de l'espace réel à trois dimensions peut s'écrire sous la forme d'un quaternion imaginaire pur tel que $\mathbf{v} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$. Un quaternion Q est la représentation complexe d'une matrice de rotation à trois dimensions. Donc, la transformée $\mathbf{v}'(x, y, z)$ d'un vecteur $\mathbf{v}(x, y, z)$ selon une rotation de quaternion Q est donnée par la multiplication de quaternions suivante :

$$\mathbf{v}'(x, y, z) = Q \times \mathbf{v}(x, y, z) \times Q^* \tag{3.1}$$

où Q est un quaternion normé et Q^* le quaternion conjugué de Q.

Il est possible d'écrire le quaternion Q correspondant à une rotation à partir des angles d'Euler (ϕ, θ, ψ) décrivant cette rotation selon :

$$Q = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \cos(\phi/2)\cos(\theta/2)\cos(\psi/2) & + & \sin(\phi/2)\sin(\theta/2)\sin(\psi/2) \\ \sin(\phi/2)\cos(\theta/2)\cos(\psi/2) & - & \cos(\phi/2)\sin(\theta/2)\sin(\psi/2) \\ \cos(\phi/2)\sin(\theta/2)\cos(\psi/2) & + & \sin(\phi/2)\cos(\theta/2)\sin(\psi/2) \\ \cos(\phi/2)\cos(\theta/2)\sin(\psi/2) & - & \sin(\phi/2)\sin(\theta/2)\cos(\psi/2) \end{bmatrix} \tag{3.2}$$

Le module `quaternion.py` est un module Python d'environ 200 lignes. Il est constitué de la classe d'objet `Quaternion` et de la classe `Qvector` héritée de ce dernier. Un objet `Quaternion` contient un attribut `Internal` qui représente ses quatre composantes sous la forme d'une liste et plusieurs fonctions réalisant les diverses opérations d'algèbre quaternionique. Il existe aussi, dans ce module, la fonction `fromEuler` qui permet de créer un quaternion à partir de trois angles d'Euler et la fonction `dist` qui donne la distance entre deux vecteurs position.

Ce module fournit au programme de docking : (*i*) la répresentation d'un vecteur position sous la forme d'un vecteur-quaternion (ou q-vecteur) et (*ii*) les différentes opérations d'algèbre quaternionique relatives aux rotations à effectuer.

3.2.3. Module `divcon.py`

Le module `divcon.py` est un module écrit en Python d'environ 100 lignes. Il permet de gérer essentiellement les entrées-sorties propres au programme DIVCON. Il contient une classe `Divcon` écrite sous la forme d'un singleton. L'objet `divcon`, instancié de ce dernier, possède différents attributs :

- `Method` : méthode semi-empirique utilisée dans le calcul de l'énergie ;
- `Charge` : charge totale du système moléculaire ;
- `Cluster` : paramètres Divide & Conquer ;
- `ProgramPath` : répertoire de l'exécutable DIVCON ;

et deux fonctions :

- `writeDivconFile` : écriture du fichier d'entrée ;
- `readEnergy` : lecture de l'énergie du système moléculaire dans le fichier de sortie.

3.3. Déroulement du docking

Nous allons, à présent, voir en détail les tâches se succèdant durant l'exécution du programme de docking. Elles ont été résumées dans la figure 3.2, sous la forme d'un diagramme UML de séquence. Ce dernier permet aussi de montrer les diverses dépendances des modules entre eux.

Le lancement du programme de docking s'effectue par l'intermédiaire d'un script Python exécutable, représenté par la colonne **Input** sur le diagramme. L'utilisateur y indique les différents paramètres et spécifications du docking qu'il souhaite réaliser. La colonne **AlgoGen** représente l'algorithme de recherche (`genetic.py`), **Docking** le traitement du système moléculaire (`docking.py` et `quaternion.py`), **Divcon** l'interface au programme et **SGE** (*Sun Grid Engine*) le gestionnaire de cluster pour le calcul distribué.

Le partie **Docking** récupère, dans un premier temps, les structures de la protéine et du ligand sous la forme de deux fichiers distincts au format PDB (**1.1**) et des informations telles que la taille de la boîte de docking ou les résidus de la partie d'intérêt de la protéine (**1.2**), *i.e.* le site actif où le ligand va être docké. Il procède ensuite à la préparation des systèmes moléculaires (**2**). Pour chaque fichier PDB, un système d'atomes est créé. L'origine des coordonnées est définie par le centre géométrique de chaque système. Nous appelerons ces systèmes PROT et LIG par la suite,

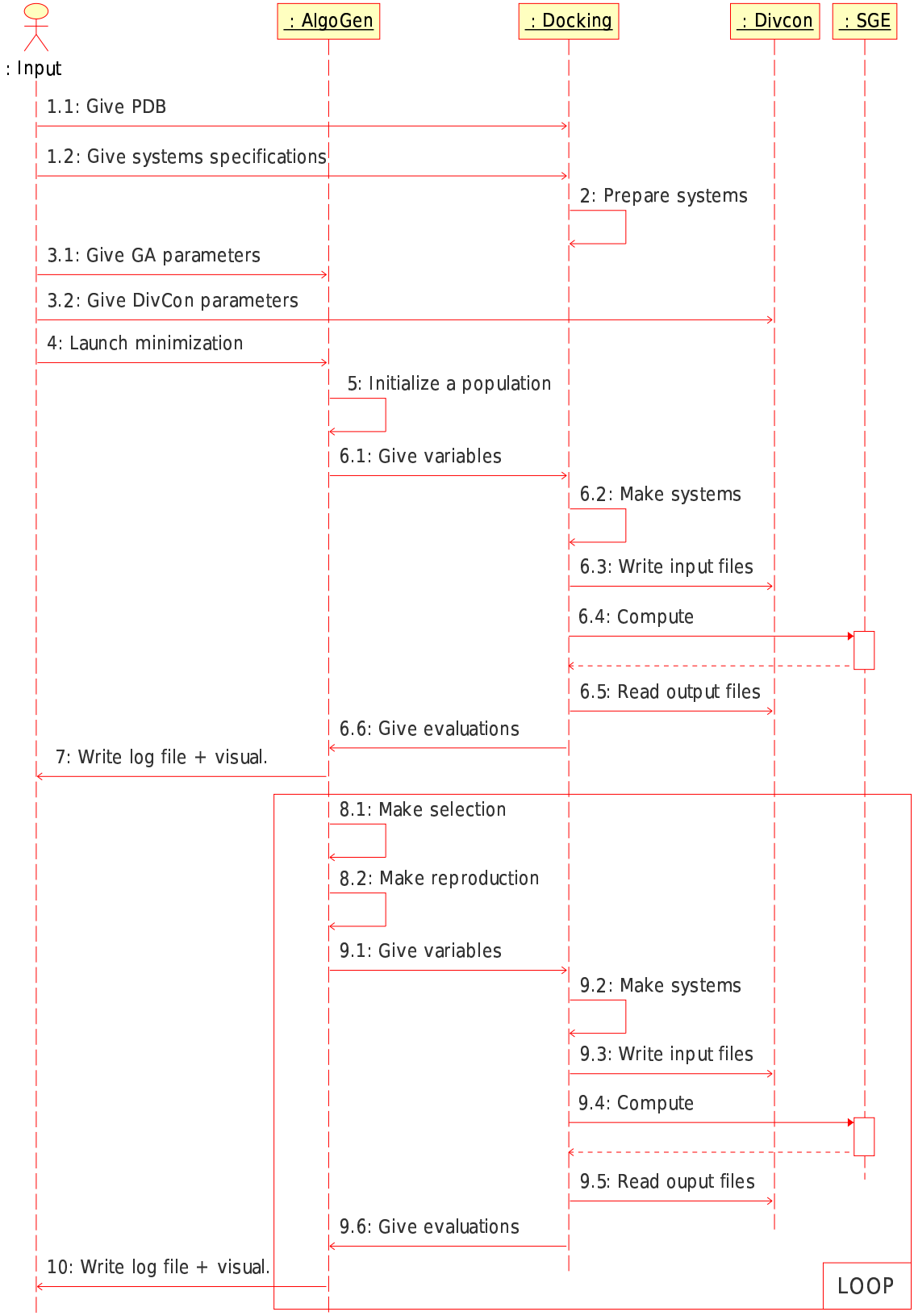


FIG. 3.2: *Diagramme UML de séquence des procédures du programme* ALGOGEN-DIVCON

respectivement pour la protéine et le ligand. L'origine du système PROT est redéfinie, le cas échéant, comme étant le centre géométrique des résidus d'intérêt et le centre de la boîte de docking est placé à cette origine. Les résidus du système PROT, dont au moins un atome se situe à l'intérieur de la sphère circonscrite à la boîte de docking, sont conservés et pris en compte dans le calcul de l'énergie potentielle. Il y a un traitement automatique des éventuelles liaisons peptidiques rompues. Les liaisons entre le carbone C_α et le groupement NH, d'une part, et le groupe CO, d'autre part, sont remplacées par des liaisons C-H de 1 Å de distance.

Le programme récupère ensuite les paramètres relatifs à l'algorithme génétique (**3.1**) et au programme DIVCON (**3.2**). La dernière instruction du script utilisateur est le lancement de la minimisation (**4**) par la partie **AlgoGen**. Une première population d'individus est créée (**5**). Lors de la phase d'évaluation de cette population, le chromosome de chaque individu est décodé et les variables sont transmises à la partie **Docking** (**6.1**). Un système protéine-ligand est créé pour chaque individu (**6.2**) en effectuant une copie du système PROT et du système LIG. Les trois dernières variables (angles d'Euler) servent à créer un quaternion de rotation et on effectue la rotation du système LIG. Les trois premières variables créent ensuite un vecteur de translation des atomes de ce système. Les deux systèmes PROT et LIG sont finalement fusionnés en un seul système, transmis à la partie **Divcon** pour l'écriture du fichier d'entrée correspondant (**6.3**).

Le nombre d'évaluations à effectuer à chaque génération est souvent important, du fait de la taille de la population de l'algorithme génétique. Aussi, il est possible de distribuer l'ensemble des calculs sur un ensemble de processeurs, *e.g.* un cluster de PC. Cette technique a été intégrée dans le programme de docking en combinaison avec le logiciel SGE. Elle est gérée par la fonction `grabResults` du module `docking.py`. Lors de chaque phase d'évaluation de l'algorithme génétique, les différents calculs sont soumis à un cluster de machines grâce à un script SGE (**6.4**). Le programme de docking se met alors en pause, attendant que tous les calculs aient été réalisés.

Une fois les évaluations achevées, la partie **Docking** demande à la partie **Divcon** de lire l'énergie du système protéine-ligand de chaque calcul réalisé (**6.5**) et envoie ces valeurs à la partie **AlgoGen** afin de poursuivre le docking (**6.6**). L'algorithme génétique se charge de l'écriture des fichiers de sortie, de sauvegarde et de visualisation du docking (**7**).

Le programme entre ensuite dans une boucle pour produire les populations successives de l'algorithme génétique. Cette boucle se termine lorsque le nombre maximal de générations voulu par l'utilisateur a été atteint. Le programme peut aussi s'arrêter si le nombre maximal d'évaluations (appels au programme DIVCON) a été atteint ou par tout autre critère d'arrêt défini par l'utilisateur. Dans

la partie **AlgoGen**, chaque cycle de reproduction commence par la sélection des individus (**8.1**) et leur reproduction pour créer une nouvelle population (**8.2**). La procédure d'évaluation de cette population (**9.x**) est identique à celle en dehors de la boucle (**6.x**). Les fichiers de sortie, de sauvegarde et de visualisation du docking sont écrits en fin de chaque cycle de reproduction (**10**).

Une fonctionnalité développée dans ce programme permet d'éliminer, lors de la phase de reproduction de l'algorithme génétique, les individus nouvellement produits dont on sait qu'il ne seront pas de bonnes solutions au problème de docking. Il s'agit de la fonction `controlDistance` du module `docking.py`. Si l'individu en question correspond à un système dans lequel un atome du ligand se situe à moins de 0,75 Å d'un atome de la protéine, il est rejeté et un nouvel individu est produit. Cette technique permet de ne pas effectuer de calculs d'énergie inutiles.

Plusieurs tests ont été réalisés afin de valider le développement du logiciel de docking semi-empirique à croissance linéaire AlgoGen-DivCon. Les résultats sont reportés dans le chapitre suivant. Un manuel d'utilisation, reporté en annexe C de ce manuscrit, est fourni avec le logiciel. Il regroupe la liste des fonctions et un exemple de script de lancement.

Chapitre 4

Validation du programme ALGOGEN-DIVCON

Sommaire

4.1. Structure du dimère de l'eau 66
4.2. Docking dans une protéine 70
4.3. Conclusions 82

Le programme ALGOGEN-DIVCON a fait l'objet d'une phase de validation dont les résultats sont présentés dans ce chapitre. Nous avons effectué différents tests qui nous ont permis de vérifier le bon fonctionnement et l'efficacité de ce programme : (*i*) l'optimisation de la structure du dimère de l'eau *via* différentes méthodes semi-empiriques et (*ii*) le docking de différents ligands dans le site actif d'une enzyme (urate oxydase). Ces tests permettent de valider différentes caractéristiques de notre programme telles que la description quantique d'un système moléculaire ou l'utilisation d'un algorithme génétique pour la minimisation. Des comparaisons vis-à-vis de structures cristallographiques connues de complexes enzyme-ligand viendront valider les résultats de docking obtenus pour l'enzyme.

4.1. Structure du dimère de l'eau

Le premier test réalisé permet de vérifier si le programme que nous venons de développer est capable d'optimiser la géométrie d'une structure simple comme celle du dimère de l'eau. Dans ce système, une première molécule d'eau est fixe, *i.e.* elle joue le rôle de la protéine, et la deuxième est mobile. L'optimisation de géométrie du dimère est réalisée en utilisant différents hamiltoniens semi-empiriques : AM1, PM3 et PM3-PIF.

En comparaison, la même optimisation de géométrie est réalisée par le programme GEOMOP[110] au niveau AM1 et PM3 *via* une méthode des gradients conjugués pour la minimisation. Ici aussi, les géométries internes de chaque molécule d'eau sont gelées, une molécule est fixe et l'autre mobile, ayant donc les mêmes six degrés de liberté que l'algorithme génétique. Il nous est possible alors de valider spécifiquement l'utilisation de notre logiciel pour la minimisation dans le cadre d'un docking moléculaire.

a. Paramètres du docking

Concernant l'algorithme génétique, nous utilisons une population de 32 individus dont le chromosome est constitué de 3 gènes de longueur 10 pour les variables de translation et de 3 gènes de longueur 8 pour la rotation. La taille de la boîte de docking est fixée à 10 Å, centrée sur la molécule d'eau fixe. Compte-tenu de la taille des gènes, nous avons une précision de 0,01 Å pour chacune des trois dimensions du vecteur translation et de 1,4° pour chacun des angles d'Euler. Le croisement est affecté d'une probabilité de 100%. Du fait de la taille relativement petite des gènes, un seul croisement est réalisé sur le gène en entier. Il en est de même pour la mutation, où la probabilité vaut 20%. L'inversion de gènes n'est pas autorisée. La sélection par élitisme n'est pas utilisée. La population est sélectionnée par le biais de 16 tournois. Nous interdisons l'eugénisme des individus, tous les individus sont donc différents dans chaque population. La périodicité est appliquée aux variables de rotation. La stratégie de reproduction utilisée est celle par laquelle la nouvelle génération contient les 16 individus sélectionnés de la génération précédente et des individus totalement nouveaux obtenus par reproduction.

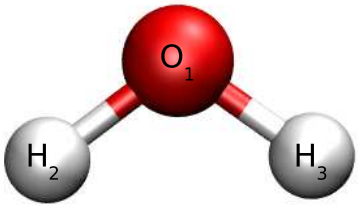
Nous appliquons également, pour chacune des méthodes semi-empiriques, l'approche Divide & Conquer en divisant le système total en sous-systèmes, chacun étant constitué d'une molécule d'eau (cœur) et de son environnement étendu à 6 Å. Chaque cœur est ainsi entouré de deux régions-tampons de taille 4 Å et 2 Å respectivement. Il a été montré que ces paramètres définissant chaque

sous-système permettent une détermination relativement correcte des énergies semi-empiriques de systèmes moléculaires traités par l'approche Divide & Conquer[77,111].

b. Construction des systèmes moléculaires

La géométrie d'une molécule d'eau est pleinement optimisée par la méthode des gradients conjugués au niveau AM1 et PM3. Le tableau 4.1 reporte les valeurs de distance et d'angle des géométries optimisées en fonction de la méthode semi-empirique utilisée. La géométrie optimisée AM1 sert à décrire la géométrie interne de chacune des deux molécules d'eau du dimère traité en AM1. De même, le dimère traité en PM3 et PM3-PIF est construit à partir de la géométrie optimisée PM3. Il n'est pas nécessaire d'optimiser la géométrie d'une molécule d'eau en PM3-PIF puisque cette méthode ne modifie que les fonctions d'interaction cœur-cœur intermoléculaires (*i.e.* les géométries optimisées d'une molécule d'eau en PM3 et PM3-PIF sont identiques).

Méthode	O_1-H_2 (Å)	O_1-H_3 (Å)	H_2-O_1-H_3 (°)
AM1	0,96	0,96	103,6
PM3	0,95	0,95	107,6



TAB. 4.1: *Structure de la géométrie optimisée d'une molécule d'eau par la méthode des gradients conjugués avec les méthodes semi-empiriques AM1 et PM3. La figure ci-contre indique la nomenclature utilisée pour les atomes de la molécule.*

c. Résultat du docking

L'optimisation est lancée sur 500 générations de l'algorithme génétique, et ce, pour chacune des méthodes semi-empiriques. La figure 4.1 montre l'évolution de l'énergie d'interaction du dimère de l'eau, pour l'individu représentant la meilleure solution potentielle de chaque génération, en fonction de la méthode semi-empirique utilisée. Nous constatons que les énergies d'interaction convergent, plus ou moins rapidement, vers une valeur donnée. Il ne nous est pas possible, de prime abord, de déterminer si l'algorithme génétique a permis de faire converger l'énergie d'interaction vers son minimum global. Cependant, nous considérons que les très faibles variations énergétiques observées après un grand nombre de générations sont dues à de petites variations de la structure du dimère, ne changeant en rien sa structure globale d'un point de vue qualitatif. En d'autres termes, nous considérons que la structure du dimère a convergé vers sa structure optimale lorsque les variations

de l'énergie d'interaction tendent à être plus faibles et plus rares. Il est tout à fait possible de prolonger l'algorithme génétique afin de s'assurer de sa bonne convergence vers la structure optimale. Nous avons vérifié avec succès cette affirmation.

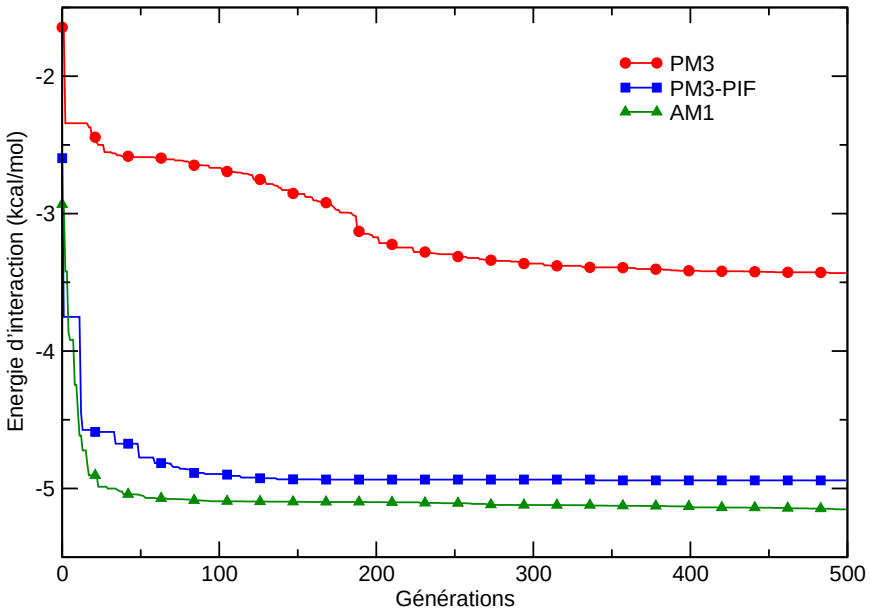


Fig. 4.1: *Evolution de l'énergie d'interaction du dimère de l'eau en fonction de la méthode semi-empirique utilisée. L'énergie d'interaction représentée à chaque génération est celle correspondant à la meilleure solution potentielle donnée par l'algorithme génétique.*

Sur un plan structural, nous observons une différence entre le dimère de l'eau obtenu avec la méthode AM1, d'une part, et les méthodes PM3 et PM3-PIF, d'autre part. La première nous donne une structure "bifurquée" (type C_2v) et la deuxième, une structure "linéaire" (type Cs). Ces deux structures sont présentées dans la figure 4.2. Cette différence de la structure du dimère en AM1 est connue[71,112]. La structure "linéaire" existe aussi en AM1. Elle se situe à environ 6 kcal/mol au dessus de la structure "bifurquée" (cas d'une optimisation complète de géométrie) qui est celle de plus basse énergie. Nous n'observons pas de structure AM1 "linéaire" durant le processus de docking. Cette remarque nous amène à penser que l'algorithme génétique utilisé permet d'atteindre préférentiellement le minimum global de l'énergie d'interaction.

Le tableau 4.2 reporte les énergies d'interaction et les géométries des dimères de l'eau que nous avons obtenus avec AlgoGen-DivCon et avec Geomop pour les différentes méthodes semi-empiriques. Nous constatons que les structures obtenues avec notre programme sont proches de celles obtenues avec Geomop, tant au niveau de la géométrie que de l'énergie d'interaction. Nous obtenons une énergie d'interaction du dimère de l'eau valant -5,15 kcal/mol en AM1 et -3,43 kcal/mol en PM3. La correction des fonctions cœur-cœur de la méthode PM3-PIF améliore la valeur de

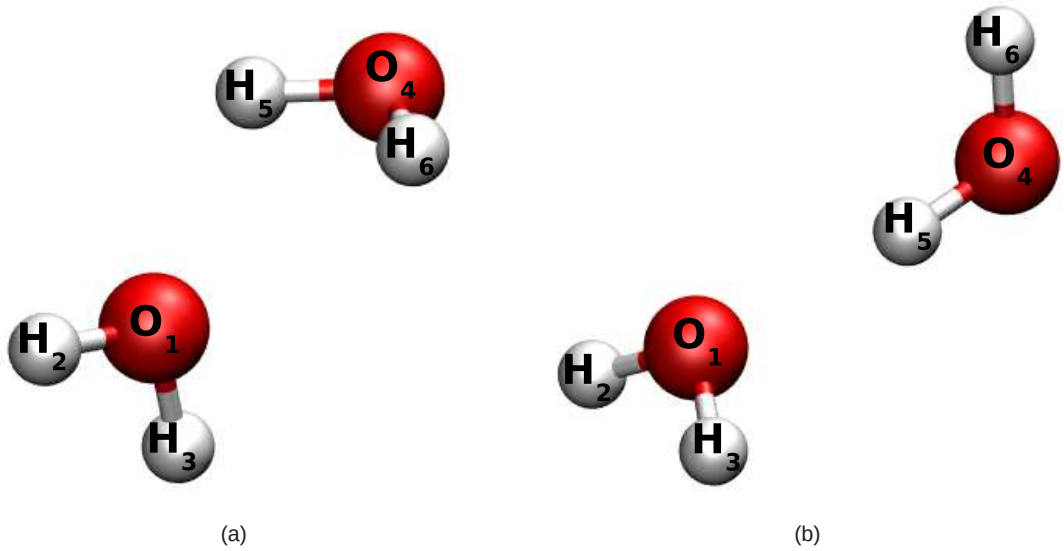
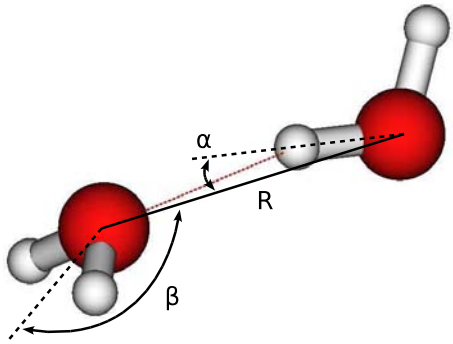


FIG. 4.2: *Structure obtenue après 500 générations de l'algorithme génétique.*
(a) La structure AM1 est de type C_2v. Les plans de symétrie sont définis par les atomes O_1, H_2 et H_3, d'une part, et O_4, H_5 et H_6, d'une part.
(b) Les structures PM3 et PM3-PIF sont toutes les deux de type Cs. Le plan de symétrie est défini par les atomes O_4, H_5 et H_6.



				AlgoGen-DivCon		
Méthode	R (Å)	α (°)	β (°)	ΔE (kcal/mol)	Générations	Evaluations
AM1	2,55	55,5	148,6	-5,15	487	9408
PM3	2,76	2,7	111,2	-3,43	482	7792
PM3-PIF	2,86	3,8	119,7	-4,94	340	5472

				Geomop
Méthode	R (Å)	α (°)	β (°)	ΔE (kcal/mol)
AM1	2,49	58,5	161,3	-5,17
PM3	2,76	0,4	111,0	-3,44

TAB. 4.2: *Energies d'interaction (ΔE) et paramètres structuraux du dimére de l'eau optimisé par* AlgoGen-DivCon *(algorithme génétique) et* Geomop *(gradients conjugués) avec les méthodes semi-empiriques AM1, PM3 et PM3-PIF. Les nombres de générations et d'évaluations réalisées par l'algorithme génétique sont indiqués pour les différents cas. La figure ci-dessus donne la définition de la distance R et des angles α et β dans le dimère de l'eau.*

l'énergie d'interaction obtenue en PM3, passant de -3,43 à -4,94 kcal/mol. On se rapproche ainsi de la valeur expirimentale[113] de -5,4 kcal/mol et de la valeur consensus de plusieurs méthodes *ab initio* de haut niveau[114] de -5,0 kcal/mol.

Le nombre de générations de l'algorithme génétique indiqué dans le tableau 4.2 est celui à partir duquel la structure d'énergie minimale a été obtenue. Il n'est pas surprenant que cette valeur, et le nombre d'évaluations correspondant, soit grand pour chacun des cas. Ceci est principalement dû à la taille relativement grande de la boîte de docking dans laquelle la deuxième molécule d'eau peut se déplacer. Il est certain que, pour ce système moléculaire de petite taille, une minimisation par la méthode des gradients conjugués sera toujours plus rapide et plus efficace qu'un algorithme génétique. Néanmoins, ce test illustre la capacité du programme ALGOGEN-DIVCON à optimiser un système moléculaire simple formé de deux molécules en interaction.

4.2. Docking dans une protéine

Après avoir montré l'efficacité du programme de docking dans l'obtention de la structure du dimère de l'eau, nous avons voulu le tester dans une application de docking moléculaire d'un système protéine-ligand. Notre choix de système d'étude s'est porté sur l'urate oxydase (UOx) qui a fait l'objet de plusieurs travaux dans notre laboratoire.

a. Description de l'urate oxydase

L'urate oxydase (EC 1.7.3.3) est une enzyme de la classe des oxydoréductases catalysant l'oxydation de l'acide urique en 5-hydroxyisourate, un composé instable qui se décompose ultérieurement en allantoïne[115–117], comme l'illustre la figure 4.3. Cette enzyme est présente chez la plupart des organismes vivants à l'exception notable des humains et des primates supérieurs[118]. Elle fait partie de la voie de dégradation de la purine et prévient l'accumulation d'acide urique dans le sang.

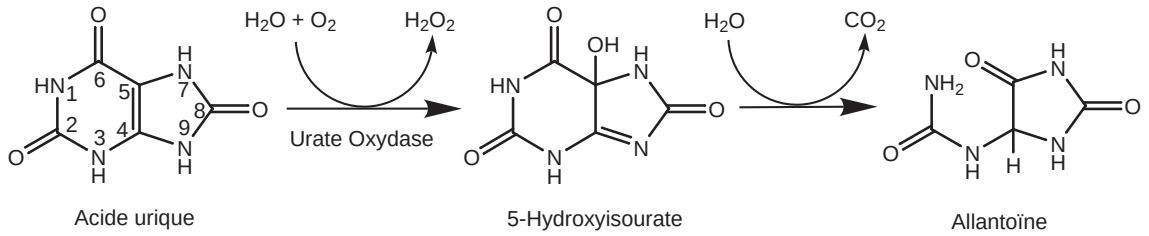


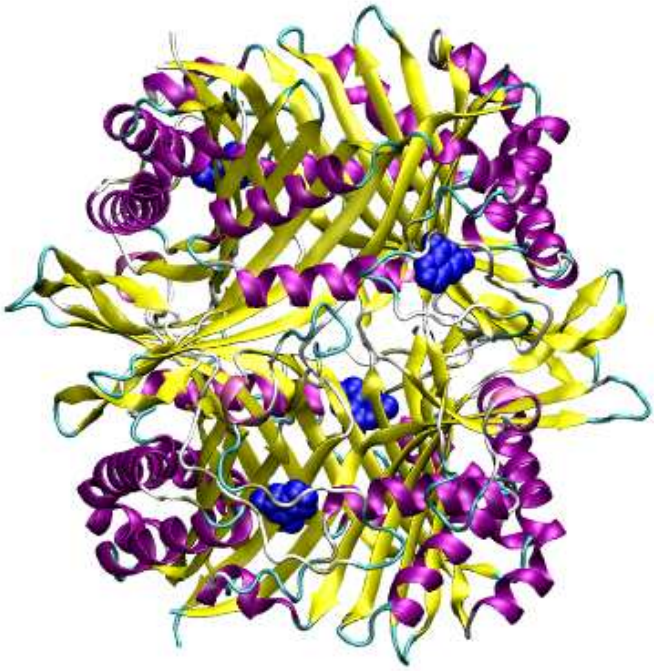
FIG. 4.3: *Décomposition de l'acide urique en 5-hydroxyisourate, puis en allantoïne par la catalyse enzymatique de l'urate oxydase.*

L'urate oxydase d'*Aspergillus flavus* a été cristallisée en complexe avec différents inhibiteurs compétitifs[119–122]. La forme biologiquement active est un homo-dimère ou un homo-tétramère. Le site actif est formé de plusieurs résidus à la jonction de deux monomères : Lys10*, Thr57*, Asp58*, Phe159, Arg176, Val227, Gln228 et Asn254. L'astérisque indique une localisation des résidus sur le deuxième monomère. Deux molécules d'eau (W1 et W2) ont aussi été déterminées à proximité de ces résidus. Elles joueraient un rôle important dans le mécanisme catalytique. La structure de l'enzyme avec l'inhibiteur 8-azaxanthine, connue sous le code PDB 2IBA[122], nous informe des principales interactions qu'il se forme entre ces résidus du site actif et le ligand. La figure 4.4 montre la structure globale de l'enzyme en complexe avec son inhibiteur, ainsi qu'un zoom sur le site actif où sont représentées les principales interactions. Plusieurs liaisons hydrogène sont impliquées dans la stabilisation du ligand :

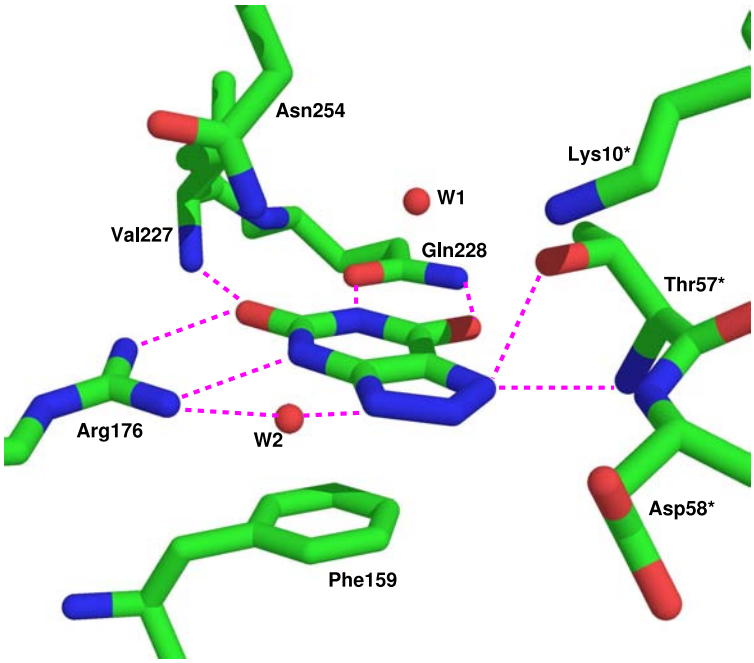
- entre Gln228-$O_{\varepsilon 1}$ et Aza-N_1 ;
- entre Gln228-$N_{\varepsilon 2}$ et Aza-O_6 ;
- entre Val227-N et Aza-O_2 ;
- entre Arg176-$N_{\eta 2}$ et Aza-O_2 ;
- entre Arg176-$N_{\eta 1}$ et Aza-N_3 ;
- entre Thr57-N et Aza-N_7 ;
- entre Thr57-O_γ et Aza-N_7 ;
- entre Wat2-O et Aza-N_9.

Compte-tenu de la nature des résidus impliqués dans ces interactions et de la structure du complexe enzyme-inhibiteur, l'état de protonation est imposé par l'enzyme. L'inhibiteur doit ainsi se présenter sous la forme d'un monoanion, protoné en position 1 et 9.

Les tests du programme AlgoGen-DivCon que nous avons réalisés avec l'urate oxydase ont pour but de vérifier la capacité de ce code à reproduire la structure du complexe de l'enzyme avec les différents substrats/inhibiteurs dont nous connaissons la structure tridimensionnelle. Ces ligands sont la 8-azaxanthine (AZA), l'acide oxonique (OXC) et l'acide urique (URA), représentés dans la figure 4.5. La structure utilisée pour l'urate oxydase est celle du complexe UOx-AZA (2IBA). Nous avons effectué le docking de ces ligands selon une approche semi-empirique à croissance linéaire avec PM3 et PM3-PIF.



(a)



(b)

FIG. 4.4: *Structure de l'urate oxydase d'Aspergillus flavus (2IBA[122]) et de son site actif.*
(a) Structure tridimensionnelle de l'urate oxydase. Le tétramère de l'urate oxydase forme une structure de type T-fold (Tunnelling-fold[123]). Les hélices α sont colorées en violet, les feuillets β en jaune et les coudes β en cyan. L'inhibiteur 8-azaxanthine, en bleu, indique la position des quatre sites actifs du tétramère.
(b) Site actif de l'enzyme en complexe avec la 8-azaxanthine. Le site actif se situe à l'interface de deux monomères. Il est constitué des résidus de Phe159, Arg176, Val227, Gln228 et Asn254 pour la première chaîne et de Lys10, Thr57 et Asp58 pour la seconde. Les principales liaisons hydrogène entre l'inhibiteur et les résidus du site actif sont indiquées en pointillés.

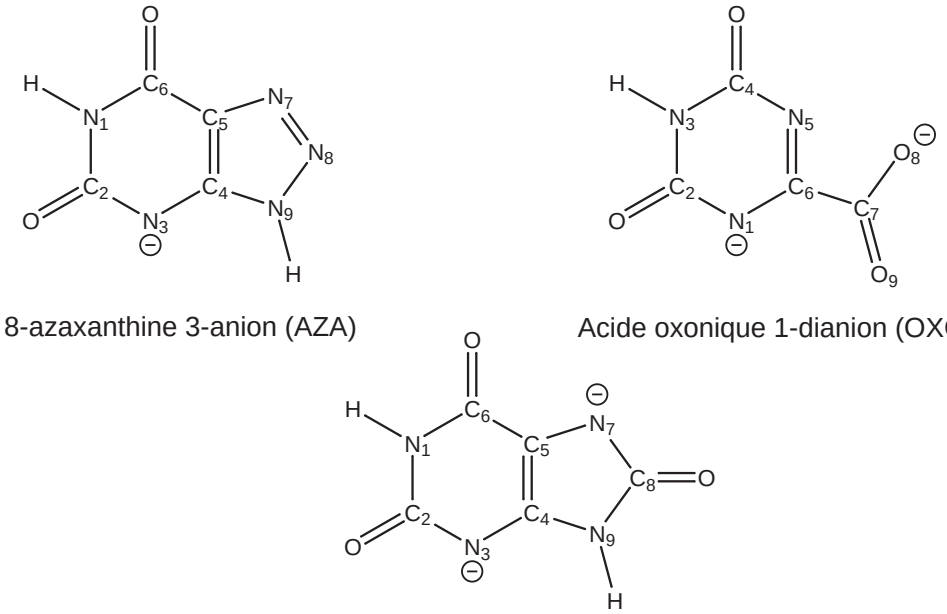


FIG. 4.5: *Structures chimiques de la 8-azaxanthine, de l'acide oxonique et de l'acide urique utilisés comme ligand dans le docking avec l'urate oxydase. La numérotation des atomes est indiquée pour chaque composé.*

b. Construction des systèmes moléculaires

Pour préparer la structure de l'enzyme, nous sommes partis de la structure PDB 2IBA. Ce fichier contient uniquement un monomère. Dans un premier temps, l'inhibiteur AZA et les autres composés de cristallisation sont retirés de la structure. Les molécules d'eau sont toutefois conservées. Les hydrogènes ont été ensuite ajoutés à la structure, en respectant l'état de protonation des résidus, par l'intermédiaire du programme TLEAP d'AMBER[124]. Ces atomes ont fait l'objet d'une première optimisation de géométrie *via* le programme SANDER d'AMBER associé au champ de forces *amber99*[125]. Enfin, le tétramère a été construit et les hydrogènes optimisés une nouvelle fois selon le même procédé. En fin de compte, tous les atomes de l'enzyme ont été conservés à leurs positions cristallographiques, à l'exception des atomes d'hydrogène.

Pour les mêmes raisons que celles mentionnées pour la 8-azaxanthine, l'état de protonation des molécules suivantes est imposé par l'enzyme. L'acide oxonique se trouve sous une forme de dianion oxonate-1,8 et l'acide urique en dianion urate-3,7. La géométrie de chacune de ces trois molécules est optimisée en phase gazeuse avec la méthode PM3.

4.2.1. Docking de la 8-azaxanthine dans le site actif de l'urate oxydase

Le docking de la 8-azaxanthine dans le site actif de l'urate oxydase a fait l'objet de deux tests distincts. Dans un premier temps, nous avons voulu comparer, de la même manière que pour le dimère de l'eau, l'utilisation d'une minimisation par les gradients conjugués et par un algorithme génétique pour le docking de ce système. Le nombre d'atomes est alors réduit au minimum pour pouvoir être traité par GEOMOP. Dans un deuxième temps, un docking moléculaire sur un système enzyme-ligand de plus grande taille a été entrepris pour tester l'efficacité de notre programme dans les conditions "réelles" d'utilisation.

a. Docking dans le petit système

Le petit système est construit par extraction d'une petite partie de la structure de l'enzyme préparée ci-dessus. Il contient tous les résidus et les molécules d'eau compris dans une boîte cubique de 5 Å de côté placée au centre géométrique du site actif, soit un total d'environ 140 atomes pour la partie enzyme.

Le docking avec ALGOGEN-DIVCON a été réalisé en utilisant une population de 32 individus. Chaque chromosome est constitué de 3 gènes de longueur 10 pour la translation et 3 gènes de longueur 8 pour la rotation. La taille de la boîte de docking est de 5 Å. Compte-tenu de la taille réduite de la partie enzyme, nous avons utilisé un processus de docking en deux étapes successives. Par la première, nous voulons explorer grossièrement l'espace de recherche pour localiser la zone du minimum global. Les paramètres de l'algorithme génétique sont alors volontairement importants : 20 individus sélectionnés uniquement par tournois, une probabilité de croisement de 100% et une probabilité de mutation et d'inversion de 20%. Après une centaine de générations produites de cette manière, la solution optimale est déjà proche du minimum global. Les paramètres de l'algorithme génétique sont alors modifiés pour permettre une recherche plus fine de ce minimum : 16 individus sélectionnés par tournois, une probabilité de croisement et de mutation inchangée et une probabilité d'inversion réduite à 0%.

Le docking de ce système est réalisé en plusieurs vagues successives de minimisation de 50 générations. Nous arrêtons le docking après 50 générations produites sans changement de l'énergie d'interaction de la meilleure solution potentielle. Avec les hamiltoniens semi-empiriques PM3 et PM3-PIF, le complexe docké est obtenu après 538 et 149 générations, correspondant à 6476 et

1808 évaluations, respectivement. Les résultats sont reportés dans le tableau 4.3, dans la colonne "Petit système". Les distances en PM3 sont identiques entre les résultats de AlgoGen-DivCon et de Geomop, et très proches de la structure cristallographique. Les RMS de déviation de la position de la 8-azaxanthine dans le site actif, entre la structure 2IBA et AlgoGen-DivCon, sont de 0,12 Å en PM3 et de 0,30 Å en PM3-PIF, respectivement.

	Petit système			Grand système		
	AlgoGen-DivCon		Geomop	AlgoGen-DivCon		PDB
Distances (Å)	PM3-PIF	PM3	PM3	PM3-PIF	PM3	2IBA
Aza-N_1 - Gln228-$O_{\varepsilon 1}$	3,10	2,94	2,95	3,06	2,91	2,97
Aza-O_6 - Gln228-$N_{\varepsilon 2}$	2,89	2,80	2,81	2,93	2,80	2,91
Aza-O_2 - Val227-N	3,06	2,82	2,82	3,03	2,77	2,91
Aza-O_2 - Arg176-$N_{\eta 2}$	2,77	2,72	2,71	2,96	2,77	2,81
Aza-N_3 - Arg176-$N_{\eta 1}$	2,84	2,82	2,81	2,95	2,85	2,99
Aza-N_7 - Thr57-N	2,88	2,85	2,85	2,86	2,84	2,75
Aza-N_7 - Thr57-O_γ	3,59	3,19	3,18	3,60	3,08	3,21
Aza-N_9 - Wat2-O_w	2,87	2,76	2,76	2,85	2,75	2,69
ΔE (kcal/mol)	-138,7	-85,7	-85,7	-171,5	-74,5	—

Tab. 4.3: *Energies d'interaction (ΔE) et distances des interactions du complexe entre l'urate oxydase et la 8-azaxanthine optimisé par* AlgoGen-DivCon *(algorithme génétique) et* Geomop *(gradients conjugués) avec les méthodes semi-empiriques PM3 et PM3-PIF. Les valeurs sont données pour le petit et le grand système. Les distances de la structure cristallographique 2IBA sont reportées dans ce tableau pour comparaison.*

b. Docking dans le grand système

Le même docking a été réalisé en prenant en compte un plus grand système pour tester les aptitudes en croissance linéaire du programme AlgoGen-DivCon. Le grand système est toujours construit à partir de la structure préparée de l'enzyme, mais en prenant tous les atomes compris dans une boîte cubique de 12 Å de côté placée au centre géométrique du site actif. Il contient environ 700 atomes dont une petite partie représente des molécules d'eau cristallographique. Ce système est trop important pour pouvoir être traité par Geomop. Nous validerons ce docking uniquement par rapport à la structure cristallographique.

Les paramètres utilisés pour l'algorithme génétique sont : population de 32 individus dont chaque chromosome contient 3 gènes de taille 10 et 3 gènes de taille 8, 16 individus sélectionnés par tournois, croisement à 100%, mutation à 20% et pas d'inversion. Compte-tenu de la taille du système, il n'a pas été nécessaire de passer par une étape de recherche grossière du minimum global. Ces paramètres ont suffit à le localiser. Les calculs ont été effectués en PM3 et PM3-PIF avec croissance

linéaire. Nous appliquons la méthode Divide & Conquer avec les paramètres suivants : 1 résidu (acide aminé, ligand molécule d'eau) par cœur et 2 régions-tampons de taille 4 et 2 Å, respectivement. Le docking est réalisé en plusieurs vagues successives de minimisation de 50 générations. Nous considérons qu'un système protéine-ligand présente sa conformation optimale lorsque nous observons 50 générations sans changement notable de l'énergie d'interaction. Nous utiliserons, par la suite, ces mêmes paramètres pour les tests de docking avec d'autres ligands.

Les structures dockées en PM3 et PM3-PIF sont obtenues après 238 et 188 générations, correspondant à 3744 et 3024 évaluations, respectivement. Les distances et les énergies d'interaction sont reportées dans le tableau 4.3, dans la colonne "Grand système". Nous constatons, ici aussi, que les structures dockées en PM3 et PM3-PIF sont très similaires à la structure cristallographique. Les RMS de déviation du ligand AZA par rapport à la cristallographie sont de 0,14 Å en PM3 et de 0,23 Å en PM3-PIF. La figure 4.6 présente une comparaison de la structure cristallographique avec la structure dockée en PM3-PIF. Nous n'observons que très peu de différences.

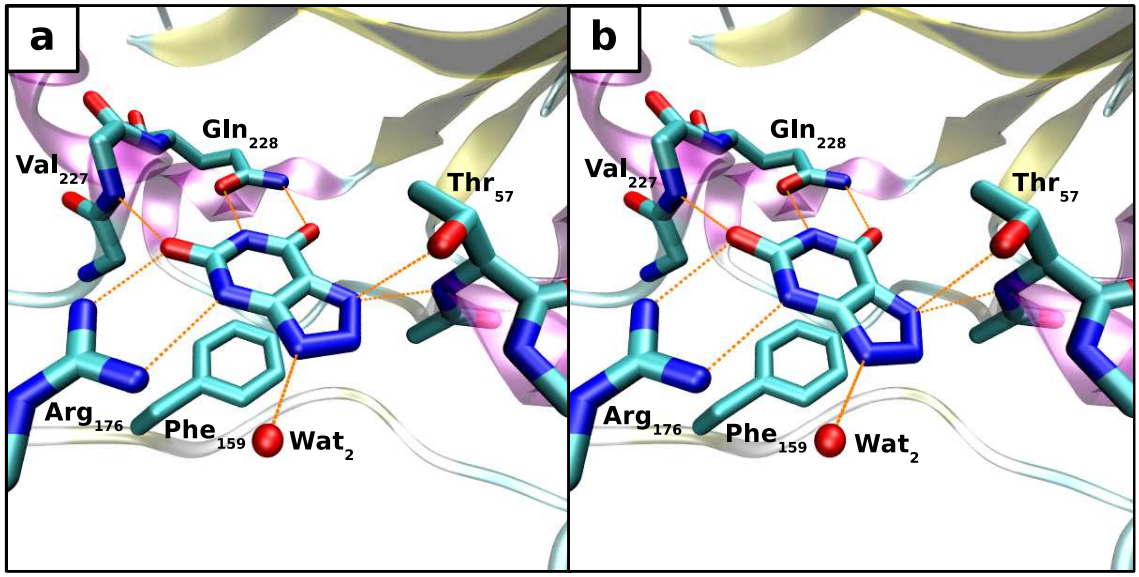


FIG. 4.6: *Structure du complexe de la 8-azaxanthine dans le site actif de l'urate oxydase : a) Structure cristallographique 2IBA et b) Structure dockée par* ALGOGEN-DIVCON *en PM3-PIF pour le grand système, à partir de la structure 2IBA.*

c. Discussion

Le tableau 4.3 regroupe toutes les principales distances d'interaction entre la 8-azaxanthine et l'urate oxydase, soit pour la structure cristallographique 2IBA, soit pour nos résultats de docking moléculaire en PM3 et PM3-PIF, et ce, pour le petit et le grand système. Les structures dockées sont toutes globalement très similaires comparées à la structure cristallographique de référence. Les interactions entre le ligand et les résidus du site actif sont de deux types : liaisons hydrogène et interaction électrostatique globale. Les résidus du site actif ont une charge globale de +1 u.a. (petit et grand système) et le ligand AZA, une charge de -1 u.a. Ceci implique une interaction intermoléculaire négative, stabilisée davantage par l'établissement de plusieurs liaisons hydrogène dans le complexe. L'état de protonation particulier du ligand AZA oriente correctement celui-ci dans le site actif de l'enzyme, comme on peut le voir sur la figure 4.6. L'urate oxydase possède la propriété de sélectionner l'état de protonation approprié de son substrat[126]. Tout autre état de protonation conduirait à un résultat de docking totalement différent, correspondant à une énergie d'interaction plus élevée (résultats non présentés).

Nous constatons aussi qu'il n'y a pas de grandes différences entre le petit et le grand système concernant la structure dockée. Avec l'hamiltonien PM3, les structures dockées sont très similaires. Les RMS de déviation du ligand AZA entre 2IBA et AlgoGen-DivCon sont de 0,12 Å et 0,14 Å dans le petit et le grand système, respectivement. Avec PM3-PIF, ces RMS sont de 0,30 Å et 0,23 Å respectivement. Ce dernier cas semble nous indiquer que l'augmentation de la taille du système tend à améliorer le résultat du docking.

Les énergies d'interaction sont très différentes, d'une part, entre PM3 et PM3-PIF et, d'autre part, entre le petit et le grand système. L'énergie d'interaction en PM3-PIF est toujours plus grande en valeur absolue à celle en PM3. Ce résultat n'est pas surprenant vis-à-vis de ce qui a pu être constaté pour le dimère de l'eau (Tab. 4.2). Dans ce système, une seule liaison hydrogène est présente et implique une différence d'énergie de 1,5 kcal/mol, soit environ 40% de l'énergie d'interaction totale. Considérant l'existence d'un grand nombre de liaisons hydrogène dans le système UOx-AZA, il est acceptable de constater cette grande différence d'énergie entre PM3 et PM3-PIF. Sachant que le système enzyme-inhibiteur est chargé (-1 u.a. pour UOx et +1 u.a. pour AZA), l'énergie d'interaction inclut l'énergie due à une interaction électrostatique charge-charge au premier ordre attractive. Ainsi, une petite déviation du centre de masse dans la partie enzyme (entre petit et grand système) conduit *a fortiori* à une variation de l'énergie d'interaction. Elle est d'autant plus grande que, dans notre cas, ce centre de masse se situe à l'intérieur d'une poche correspondant au site actif de l'enzyme.

4.2.2. Docking de l'acide oxonique dans le site actif de l'urate oxydase

L'acide oxonique (OXC) est un autre inhibiteur compétitif connu de l'urate oxydase[127]. Une structure tridimensionnelle de son complexe dans l'enzyme a été résolue par cristallographie X en 2004 par Retailleau *et al.* et déposée à la Protein Data Bank sous le code 1R4U[120]. L'acide oxonique a une structure relativement différente de celle de la 8-azaxanthine (Fig. 4.5). Il présente seulement un cycle hétéroatomique à six atomes tandis que le ligand AZA est la fusion de deux cycles hétéroatomiques à cinq et six atomes. Les interactions entre le ligand OXC et l'enzyme sont donc assez différentes. Cependant, l'observation de la structure 1R4U indique la présence de quelques interactions similaires à celles rencontrées pour le ligand AZA, essentiellement portées par le cycle à six atomes. La composition et la structure du site actif impliquent que l'acide oxonique se présente sous la forme d'un dianion dans le site actif de l'enzyme. Il est uniquement protoné en position 3.

Le docking moléculaire de l'acide oxonique a été réalisé dans le site actif de l'urate oxydase en PM3 et PM3-PIF. La structure de l'enzyme est celle préparée auparavant à partir de la structure cristallographique 2IBA. La boîte de docking est toujours une boîte cubique de 12 Å de côté placée au centre géométrique du site actif. De cette manière, nous disposons des mêmes résidus et molécules de solvant que le grand système présenté dans le docking avec le ligand AZA, soit environ 700 atomes.

Les structures dockées sont obtenues après 205 générations en PM3 et 160 générations en PM3-PIF, correspondant à 3296 et 2512 évaluations, respectivement. Les résultats sont résumés dans le tableau 4.4, en comparaison avec la structure 1R4U. Nous constatons que les deux structures dockées sont globablement similaires à la structure cristallographique. Les RMS de déviation sont de 0,51 Å en PM3 et de 0,71 Å en PM3-PIF. Ces valeurs sont relativement grandes par rapport à celles que nous avions obtenues précédemment avec le ligand AZA. Ceci s'explique en partie par le fait que nous avons effectué un docking rigide du ligand OXC dans le site actif de l'enzyme 2IBA qui contenait à l'origine le ligand AZA. La structure 2IBA présente un site actif adapté au complexe avec AZA. Sachant que la structure des ligands AZA et OXC est différente, la structure dockée du ligand OXC dans le site actif 2IBA est différente de celle de ce ligand dans le site actif 1R4U. La figure 4.7 montre cette différence entre la structure cristallographique 1R4U et la structure dockée en PM3-PIF dans 2IBA.

Distances (Å)	AlgoGen-DivCon PM3-PIF	AlgoGen-DivCon PM3	PDB 1R4U
Oxc-N_3 - Gln228-$O_{\varepsilon 1}$	3,05	2,76	2,99
Oxc-O_4 - Gln228-$N_{\varepsilon 2}$	2,83	2,74	2,92
Oxc-O_2 - Val227-N	3,43	2,86	2,93
Oxc-O_2 - Arg176-$N_{\eta 2}$	3,08	2,87	2,84
Oxc-N_1 - Arg176-$N_{\eta 1}$	3,73	3,35	3,18
Oxc-N_5 - Thr57-N	3,34	3,99	3,18
Oxc-N_5 - Thr57-O_γ	3,53	3,73	3,48
Oxc-O_9 - Wat2-O_w	2,55	2,06	2,41
ΔE (kcal/mol)	-110,4	-55,1	—

Tab. 4.4: *Energies d'interaction (ΔE) et distances des interactions du complexe entre l'urate oxydase et l'acide oxonique optimisé par* AlgoGen-DivCon *avec les méthodes semi-empiriques PM3 et PM3-PIF. Les distances de la structure cristallographique 1R4U sont reportées dans ce tableau pour comparaison.*

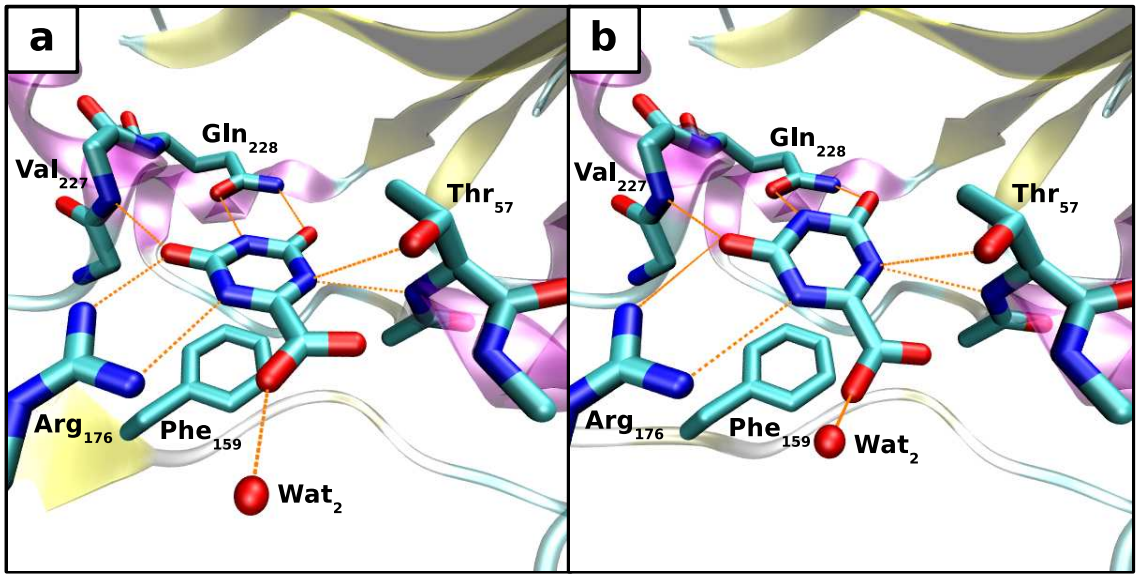


Fig. 4.7: *Structure du complexe de l'acide oxonique dans le site actif de l'urate oxydase : a) Structure cristallographique 1R4U et b) Structure dockée par* AlgoGen-DivCon *en PM3-PIF à partir de la structure 2IBA.*

Nous observons sur la figure 4.7 que le ligand OXC se trouve légèrement incliné dans la structure dockée avec AlgoGen-DivCon en comparaison avec la structure 1R4U. L'analyse des différentes interactions présentes dans la structure dockée montre une forte interaction stérique et électrostatique entre le groupe carboxylate de l'acide oxonique et la molécule d'eau W2. Dans 2IBA et 1R4U, cette molécule est toujours placée dans le plan des interactions, défini par la chaîne latérale des résidus Arg176 et Gln228 et l'atome d'azote de la chaîne principale de Thr57. Les sites actifs 2IBA et 1R4U se révèlent être grandement similaires sauf au niveau de cette molécule W2. Elle est éloignée du centre du site actif d'environ 0,5 Å dans 1R4U par rapport à 2IBA, donnant ainsi assez d'espace à la formation du complexe avec OXC. La molécule d'eau W2 fait partie de la partie enzyme et sa position est fixe dans le processus de docking rigide. Il n'existe donc pas de relaxation de géométrie au niveau de cette molécule, empêchant ainsi le positionnement correct du ligand OXC comme dans la structure 1R4U. D'un point de vue énergétique, la forte interaction électrostatique existante implique une énergie d'interaction plus faible en valeur absolue que celle obtenue pour le ligand AZA.

4.2.3. Docking de l'acide urique dans le site actif de l'urate oxydase

L'acide urique (URA) est le substrat naturel de l'urate oxydase. Récemment, une structure cristallographique de son complexe avec l'enzyme a été publiée sous le code PDB 3BJP[121]. Ce composé possède une structure analogue à celle de la 8-azaxanthine, où l'azote en position 8 est remplacé par un groupement carbonyle (Figure 4.5). La structure anionique de l'acide urique est imposée par la composition et la structure du site actif de l'urate oxydase. Le ligand se présente donc sous la forme d'un dianion 3,7[126].

Nous avons réalisé le docking de l'acide urique dans le site actif de l'urate oxydase avec les méthodes PM3 et PM3-PIF et l'approche Divide & Conquer. Le système enzymatique est construit en prenant tous les atomes à l'intérieur d'une boîte cubique de 12 Å de côté placée au centre du site actif, à partir de la structure préparée 2IBA. La partie enzyme est identique à celle du docking des ligands AZA et OXC.

La structure dockée du complexe UOx-URA a été obtenue après 228 générations en PM3 et 97 générations en PM3-PIF, soit 3568 et 1584 évaluations respectivement. Les résultats sont regroupés dans le tableau 4.5. Les structures dockées sont toutes les deux très proches de la structure cristallographique. Les RMS de déviation du ligand URA sont de 0,16 Å en PM3 et 0,36 Å en PM3-PIF par rapport à la structure 3BJP. La figure 4.8 montre une comparaison du complexe UOx-URA entre la structure cristallographique 3BJP et la structure dockée en PM3-PIF avec AlgoGen-DivCon.

Nous avons réalisé le docking du ligand URA dans la structure de site actif 2IBA. Celle-ci est adaptée normalement à un complexe avec le ligand AZA. Cependant, les ligands AZA et URA sont très similaires. Le positionnement de la molécule d'eau W2 n'est donc pas un problème, comparé au cas du docking de l'acide oxonique. Les énergies d'interaction du complexe entre l'acide urique et l'urate oxydase sont, en PM3 et PM3-PIF, plus grandes en valeur absolue que celle obtenues respectivement pour le ligand AZA. Dans ce cas présent, l'acide urique est sous une forme dianionique. L'énergie d'interaction électrostatique au premier ordre est donc plus importante pour un système chargé +2 u.a. pour le ligand et -1 u.a. pour l'enzyme que pour un système chargé +1 u.a. pour le ligand et -1 u.a. pour l'enzyme.

	AlgoGen-DivCon		PDB
Distances (Å)	PM3-PIF	PM3	3BJP
Ura-N_1 - Gln228-$O_{\varepsilon 1}$	2,99	2,93	2,94
Ura-O_6 - Gln228-$N_{\varepsilon 2}$	2,79	2,72	2,81
Ura-O_2 - Val227-N	2,83	2,79	2,80
Ura-O_2 - Arg176-$N_{\eta 2}$	2,76	2,72	2,99
Ura-N_3 - Arg176-$N_{\eta 1}$	2,90	2,86	2,93
Ura-N_7 - Thr57-N	2,69	2,72	2,72
Ura-N_7 - Thr57-O_γ	3,05	2,97	3,10
Ura-N_9 - Wat2-O_w	2,89	2,75	2,91
ΔE (kcal/mol)	-252,4	-170,8	—

Tab. 4.5: *Energies d'interaction (ΔE) et distances des interactions du complexe entre l'urate oxydase et l'acide urique optimisé par* AlgoGen-DivCon *avec les méthodes semi-empiriques PM3 et PM3-PIF. Les distances de la structure cristallographique 3BJP sont reportées dans ce tableau en comparaison.*

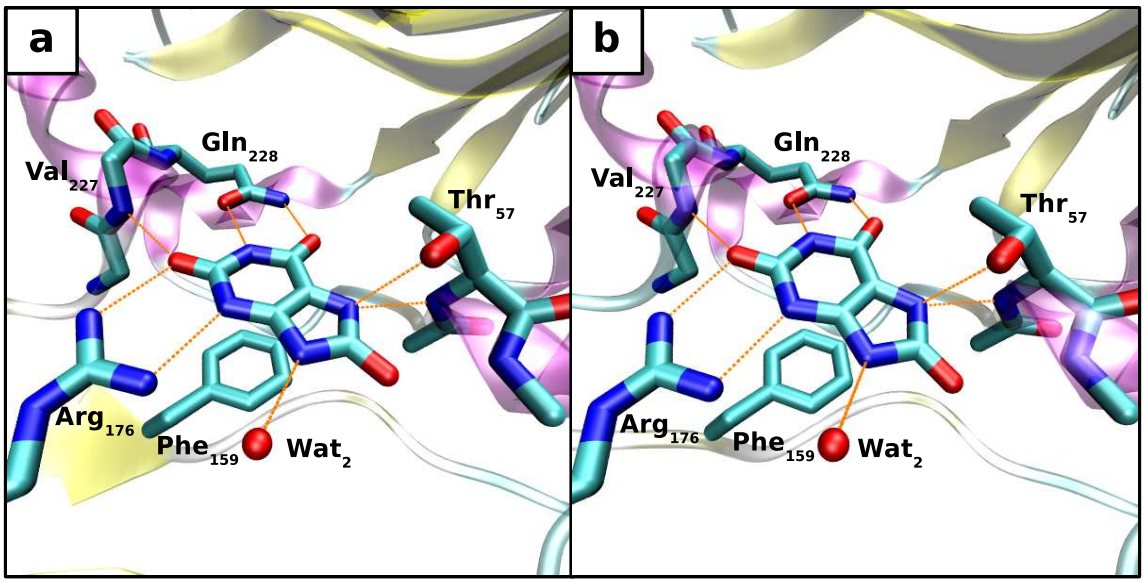


Fig. 4.8: *Structure du complexe de l'acide urique dans le site actif de l'urate oxydase : a) Structure cristallographique 3BJP et b) Structure dockée par* AlgoGen-DivCon *en PM3-PIF à partir de la structure 2IBA.*

4.3. Conclusions

Nous avons développé ALGOGEN-DIVCON, un programme de docking moléculaire en combinant un algorithme génétique pour la recherche conformationnelle et une méthode semi-empirique à croissance linéaire pour le calcul des énergies d'interaction. Ce programme permet de réaliser le docking rigide d'un système protéine-ligand de grande taille avec une description quantique de l'énergie. Les tests que nous venons de présenter valident le programme ALGOGEN-DIVCON dont les atouts et les limitations sont donnés à présent.

a. Atouts du programme ALGOGEN-DIVCON

Comparé aux autres programmes de docking que l'on peut trouver dans la littérature, le principal intérêt à l'utilisation du programme ALGOGEN-DIVCON réside dans le fait que l'énergie d'interaction entre la protéine et le ligand est calculée par une approche quantique semi-empirique à croissance linéaire. Il n'est plus nécessaire, par ce biais, de posséder les paramètres spécifiques à la description du ligand et de ses interaction avec la protéine.

Le programme ALGOGEN-DIVCON optimise l'orientation et la position d'un ligand par rapport à une protéine par l'intermédiaire d'un algorithme génétique. Nos deux premiers exemples, la structure du dimère de l'eau et le docking de la 8-azaxanthine dans le site actif de l'urate oxydase "Petit système", montrent l'aptitude du programme dans l'obtention d'un minimum global. ALGOGEN-DIVCON converge vers la même géométrie que celle obtenue par la méthode des gradients conjugués du programme GEOMOP. Les structures dockées en PM3 correspondent à la même énergie d'interaction. Tous les tests de docking de ligand dans le site actif de l'urate oxydase (petit et grand système) ont montré que les structures dockées sont très proches des structures cristallographiques connues.

Les méthodes semi-empiriques PM3 et PM3-PIF ont été utilisées dans le cadre de cette étude. Aucune conclusion concernant les performances relatives d'une approche par rapport à l'autre ne peut être extraite des différents tests. Il semblerait que l'hamiltonien PM3-PIF permette une convergence plus rapide vers la structure optimale que PM3, mais des tests complémentaires sur d'autres systèmes moléculaires doivent être réalisés pour vérifier cette affirmation.

D'après nos tests de docking moléculaire de petits ligands dans le site actif de l'urate oxydase "Grand système" ($\simeq$ 700 atomes), l'utilisation de la croissance linéaire permet un *speed-up* supérieur à 4 par rapport aux calculs semi-empiriques standards avec diagonalisations complètes de la matrice de Fock. De plus, comme le programme ALGOGEN-DIVCON requiert l'évaluation simulta-

née de chaque individu de la population à chaque génération, l'ensemble des évaluations peut être répartie sur un cluster de PC. Afin de gagner encore plus en efficacité, il est possible d'utiliser les propriétés intrinsèques de parallélisme de la méthode Divide & Conquer en réalisant un calcul parallèle de l'évaluation de chaque individu sur plusieurs processeurs (cette propriété n'a pas été mise en œuvre dans le cadre de cette étude).

Bien que certaines techniques de traitement partiellement quantique soient apparues récemment en vue de réaliser un docking[128–130], AlgoGen-DivCon reste, à notre connaissance, le premier programme réalisant un docking moléculaire rigide avec une description totalement quantique d'un système protéine-ligand de grande taille.

Dans les tests que nous avons présentés, l'état de protonation du système protéine-ligand était connu par avance car celui de l'urate oxydase impose celui du ligand. Malheureusement, ce n'est pas toujours le cas et plusieurs états de protonation doivent souvent être pris en compte pour les différentes simulations de docking. Cette caractéristique est commune à tous les programmes de docking, Nous ne l'avons pas montré ici, mais notre programme, par son traitement quantique du système protéine-ligand, permet de considérer différents états de protonation en cours de docking. Le développement de ce point en fera un atout majeur de notre logiciel.

b. Limitations du programme AlgoGen-DivCon

Le programme AlgoGen-DivCon souffre aussi de certaines limitations que nous avons mentionnées lors de nos tests. La première est liée au coût des calculs. Bien qu'une méthode à croissance linéaire ait été utilisée, le calcul de l'énergie d'interaction avec DivCon est bien plus long que ce qu'il serait avec une approche de type empirique ou champ de forces. C'est le coût de la description quantique. L'amélioration de ce point réside dans l'utilisation des propriétés de calcul parallèle-distribué du programme.

Dans nos tests, le nombre d'évaluations effectuées par AlgoGen-DivCon est relativement élevé. Ceci est la contre-partie de l'utilisation d'un algorithme génétique pour la recherche conformationnelle. Ce type d'outil a besoin d'évaluer un nombre important d'individus avant de trouver la solution optimale. Cette caractéristique est présente aussi dans les autres programmes de docking utilisant un algorithme génétique, mais représente cependant un inconvénient mineur car l'évaluation des individus est un processus rapide. Sachant que, dans notre cas, l'évaluation quantique représente presque toute la durée du processus de docking, il est important d'en tenir compte au niveau des paramètres de l'algorithme génétique. En effet, il est démontré qu'une population de grande taille tend à

faire converger le docking rapidement (*i.e.* nombre de générations réduit) vers la structure optimale, mais le nombre d'évaluations est important à chaque génération en contre-partie[97]. Inversement, une population de taille minimale demande plus de générations pour atteindre cette structure optimale. Il est donc nécessaire de trouver un jeu de paramètres de l'algorithme génétique donnant une valeur optimale du rapport entre la taille de la population et le nombre total d'évaluations.

Le docking réalisé par le programme ALGOGEN-DIVCON s'effectue en phase gazeuse. Ceci n'est pas représentatif du processus réel de docking car le solvant peut y jouer un rôle important. L'ajout de termes de solvatation[130,131] dans l'énergie calculée au niveau semi-empirique peut être un moyen de prendre en compte ce point.

Le relaxation de géométrie de la protéine et du ligand n'est pas prise en compte dans notre programme durant le processus de docking. Cette limitation est partagée par tous les programmes de docking rigide. Dans nos exemples, la flexibilité du ligand n'est pas nécessaire car la 8-azaxanthine, l'acide oxonique et l'acide urique sont les molécules planes conjuguées sans angle dièdre mobile. Cependant, la flexibilité de l'enzyme aurait été nécessaire dans le cas du docking de l'acide oxonique. Le déplacement de la molécule d'eau W2 aurait permis la formation d'un complexe stable avec l'enzyme. Il est possible de s'affranchir de ces problèmes de rigidité en effectuant un docking rigide sur plusieurs conformations de la protéine. Cependant, la grande taille des systèmes moléculaires implique un nombre important de degrés de liberté à prendre en compte dans la construction de ces différentes conformations.

c. Evolution du programme

Cette première version du programme ALGOGEN-DIVCON a fait ses preuves en montrant qu'il est désormais possible de réaliser un docking rigide avec une description quantique du système moléculaire. Nous indiquons les prochaines améliorations que nous souhaitons apporter au programme pour le faire évoluer :

- amélioration de l'algorithme génétique pour accélérer la convergence vers la solution optimale (stratégie d'évolution) ;
- ajout des effets du solvant dans l'énergie d'interaction ;
- flexibilité du ligand et des résidus d'intérêt de la protéine ;
- prise en compte de l'état de protonation du système protéine-ligand en considérant, pour chaque entité, l'existence de différents états de protonation durant le processus de docking.

Deuxième partie

Mécanisme catalytique des méthionine sulfoxyde réductases de classe A

Chapitre 5

Rappels bibliographiques et objectifs

Sommaire

5.1. Stress oxydant et oxydation des résidus méthionine 88
5.2. Méthionine sulfoxyde réductases 95
5.3. Etudes antérieures sur l'étape réductase de MsrA 108
5.4. Objectifs et approche méthodologique 114

Ce chapitre, essentiellement bibliographique, présente le phénomène biologique d'oxydation des protéines, et plus particulièrement celui impliquant les résidus méthionine, conduisant à leur transformation en sulfoxydes de méthionine. La description faite ici est principalement inspirée d'articles de revue[132,133] et de la thèse d'Antoine[134]. Nous décrirons comment de telles entités peuvent être produites *in vivo* et quelles sont les conséquences de cette modification biochimique vis-à-vis du métabolisme. Puis, nous verrons que le vivant possède des "outils" pour lutter contre ces phènomènes d'oxydation. En ce qui concerne l'oxydation des résidus méthionine, il existe une famille d'enzymes appelées méthionine sulfoxyde réductases (Msr) capables de les réparer. Les résultats des différentes études entreprises antérieurement seront exposés avant d'introduire les objectifs de ce travail de thèse et l'approche méthodologique qui a été choisie.

5.1. Stress oxydant et oxydation des résidus méthionine

Dans les conditions physiologiques, l'oxygène atmosphérique donne naissance à des espèces oxygénées activées (EOA) à la suite (*i*) d'une réduction monoélectronique au niveau des mitochondries[135] lors du métabolisme respiratoire ou (*ii*) de mécanismes enzymatiques[136–140]. Parmi ces EOA figurent des radicaux libres comme l'anion superoxyde $O_2^{\bullet -}$ ou le radical hydroxyle $HO^{\bullet}$ et d'autres entités non radicalaires de l'oxygène comme le peroxyde d'hydrogène H_2O_2 ou l'oxygène singulet $^1O_2$. Ces espèces sont particulièrement toxiques pour l'intégrité cellulaire. Elles sont dotées de propriétés oxydantes qui les amènent à réagir, dans l'environnement où elles sont produites, avec toute une série de substrats biologiques tels que les lipides[141], les protéines[142], l'ADN[143,144], les polysaccharides[145], *etc*. Au niveau moléculaire, les EOA peuvent aussi agir comme messagers secondaires[146] et activer différents facteurs ou gènes impliqués dans le développement de diverses pathologies[147,148]. Le tableau 5.1 regroupe une liste non-exhaustive des EOA classées selon leur caractère d'oxydant à un ou à deux électrons et impliquées dans l'oxydation de substrats biologiques.

Oxydants à deux électrons		**Oxydants à un électron**	
Oxygène singulet	$^1O_2$	Anion superoxyde	$O_2^{\bullet -}$
Peroxydes	H_2O_2 ROOH	Radicaux peroxyl	$HO^{\bullet}$ $RO^{\bullet}$
Acide peroxynitreux	ONOOH	Radical nitro	$O_2N^{\bullet}$

TAB. 5.1: *Liste non-exhaustive des principales espèces oxygénées activées (EOA) impliquées dans l'oxydation de substrats biologiques.*

Pour lutter contre ces effets toxiques, il existe dans l'organisme des systèmes de défense qui permettent de réguler la quantité d'EOA présentes ou de réparer les dommages causés sur les substrats biologiques. Ces systèmes sont constitués par (*i*) des anti-oxydants[149,150] tels que les vitamines A, C et E, les oligo-éléments ou diverses biomolécules qui agissent comme des pièges à radicaux, (*ii*) des enzymes comme la superoxyde dismutase[151] qui dégradent les EOA, (*iii*) des enzymes telles certaines oxydoréductases, et notamment les méthionine sulfoxyde réductases dont nous parlerons par la suite, qui réduisent les substrats oxydés et, en dernier recours, (*iv*) des enzymes protéolytiques[152], souvent sous la forme d'un couple ubiquitine-protéasome, qui dégradent les substrats biologiques irréversiblement oxydés.

Le stress oxydant a été défini par Helmut Sies[153] comme étant le résultat d'un déséquilibre de la balance entre les pro-oxydants (EOA) et les anti-oxydants (systèmes de défense) avec, comme conséquence, l'apparition de dégâts souvent irréversibles pour la cellule.

5.1.1. Oxydation des résidus méthionine par les EOA

Les résidus méthionine sont connus pour être très sensibles au stress oxydant[154,155]. Le soufre de la fonction thioéther s'oxyde spontanément pour former une série de dérivés à des degrés d'oxydation divers comme les sulfoxydes ou les sulfones. Selon la nature de l'espèce oxydante, un résidu méthionine peut subir une oxydation à un ou deux électrons pour conduire à la formation d'un sulfoxyde de méthionine (MetSO). L'oxydation à deux électrons d'un sulfure organique R_2S conduit à la formation d'un dication sulfure[156] R_2S^{2+} (Fig. 5.1-(1)). En milieu aqueux, celui-ci se transforme en sulfoxyde R_2SO (Fig. 5.1-(2)). Plusieurs espèces oxydantes telles que des complexes métallo-bioorganiques[157,158] oxydent directement un résidu méthionine en sulfoxyde de méthionine par un transfert formel d'oxygène (Fig. 5.1-(3)). L'oxydation à un électron d'une méthionine conduit à un radical cation sulfure[159] $R_2S^{\bullet+}$ (Fig. 5.1-(4)). C'est une espèce chimique extrêmement instable, qui peut se convertir en une série d'intermédiaires et de produits parmi lesquels plusieurs amenent, à terme, à la dégradation du sulfure initial et à la perte de la fonction biologique de la protéine (Fig. 5.1-(5)). Une deuxième oxydation du radical cation sulfure $R_2S^{\bullet+}$ en dication sulfure R_2S^{2+} est possible (Fig. 5.1-(6)). On revient alors à la formation d'un sulfoxyde de méthionine par hydratation.

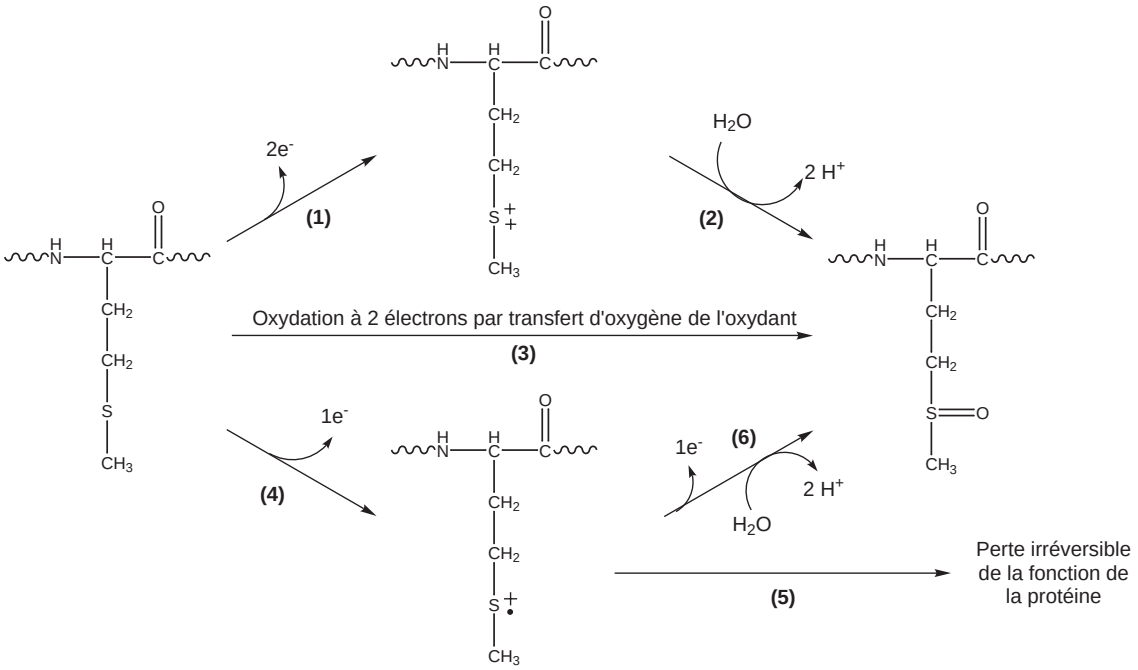


FIG. 5.1: *Différentes voies d'oxydation d'un résidu méthionine.*
(1) Oxydation à deux électrons et formation d'un dication méthionine.
(2) Hydratation du dication méthionine et formation du sulfoxyde.
(3) Oxydation "à deux électrons" par transfert de l'oxygène de l'oxydant.
(4) Oxydation à un électron et formation d'un radical cation méthionine.
(5) Transformation irréversible du sulfure et perte de la fonction biologique de la protéine.
(6) Deuxième oxydation à un électron et formation du sulfoxyde par hydratation.

L'ensemble des mécanismes d'oxydation d'une méthionine par les EOA ne sera pas détaillé. Nous verrons simplement une brève description de la réactivité des oxydants les plus fréquents. L'oxydation d'une méthionine en sulfone R_2SO_2 ne sera pas traitée. Cette réaction est possible mais nécessite des conditions drastiques. La lecture de la revue de Schöneich[132] pourra compléter cette partie.

a. Oxydation par le radical hydroxyl $HO^\bullet$

In vivo, les radicaux hydroxyl sont produits de différentes manières. La principale source provient, néanmoins, de la rupture homolytique de la liaison O–O de H_2O_2 sous la catalyse d'un centre métallique comme Fe^{II} ou Cu^{I} [160].

La première étape de la réaction entre $HO^\bullet$ et un sulfure est son addition sur l'atome de soufre, conduisant à la formation d'un radical hydroxysulfuranyl[161, 162] $R_2S^\bullet - OH$. Ces derniers sont des composés relativement instables[132] - à moins qu'ils ne soient stabilisés par des liaisons hydrogène internes[163] - et se décomposent unimoléculairement en radical cation $R_2S^{\bullet+}$ et HO^-, ou H_2O en milieu acide[162], selon :

$$HO^\bullet + R_2S \quad \rightarrow \quad R_2S^\bullet - OH \quad \rightarrow \quad R_2S^{\bullet+} + HO^-$$

Au sein d'une protéine, le radical cation $R_2S^{\bullet+}$ peut être stabilisé s'il trouve un doublet d'électrons non-liants - provenant d'un atome d'oxygène ou d'azote - avec lequel il pourra former un radical cation du type $^+S \therefore O$ ou $^+S \therefore N$[164]. Ces composés présentent une liaison de deux centres à trois électrons[165] (deux électrons dans une orbitale σ liante et un électron dans une orbitale σ^* anti-liante). L'évolution de cet intermédiaire peut conduire à la destruction de la protéine par clivage du squelette carboné de celle-ci[159].

Un sulfoxyde n'est formé qu'en présence d'O_2 à partir de ces intermédiaires radicaux cations. Pour l'intermédiaire radicalaire de type $^+S \therefore N$, l'oxygène moléculaire réagit pour former un azasulfonium ($R_2S^+ - NHR$), hydrolysable en sulfoxyde[166, 167] :

$$R_2S\,^+\!\therefore NH_2R + O_2 \quad \rightarrow \quad {}^\bullet O - O - R_2S - {}^+NH_2R \quad \rightarrow \quad R_2S^+ - NHR + HO_2^\bullet$$

$$R_2S^+ - NHR + H_2O \quad \rightarrow \quad R_2S = O + RNH_2 + H^+$$

Le radical hydroxysulfuranyl peut aussi réagir directement avec O_2 pour former l'intermédiaire radical peroxyl $R - (HO)S(OO^\bullet) - R$. Ce dernier se réarrange *via* une oxydation à un électron pour former

un sulfoxyde[168,169] :

$$R_2S^{\bullet} - OH + O_2 \quad \rightarrow \quad R - (HO)S(OO^{\bullet}) - R \quad \rightarrow \quad R_2S = O + HO_2^{\bullet}$$

b. Oxydation par le peroxyde d'hydrogène H_2O_2

La réaction de H_2O_2 avec une méthionine est un processus d'oxydation à deux électrons, comparable à un transfert formel d'oxygène selon :

$$R_2S + H_2O_2 \quad \rightarrow \quad R_2SO + H_2O$$

Selon une étude théorique de Chu *et al.*[170,171], ce mécanisme est considéré comme une réaction de type S_N2, accompagnée de la rupture hétérolytique de la liaison O–O du peroxyde. La figure 5.2 montre la structure de l'état de transition de cette réaction. L'allongement de la liaison O–O est concerté avec le raccourcissement de la liaison S–O. Deux à trois molécules d'eau stabilisent la séparation de charge présente au niveau de l'état de transition. Le transfert de l'hydrogène encore porté par l'oxygène du sulfoxyde vers l'anion HO^- intervient après cet état de transition.

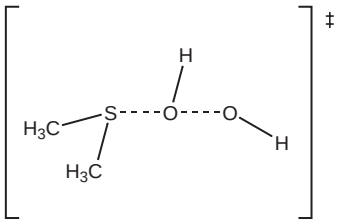


FIG. 5.2: *Etat de transition de l'oxydation d'un sulfure par le peroxyde d'hydrogène. La réaction est de type S_N2 avec rupture hétérolytique de la liaison $O - O$.*

In vivo, un rôle catalytique important a été attribué au dioxyde de carbone. En effet, la préhydratation réversible du CO_2 par H_2O_2 génère des ions peroxomonocarbonates HCO_4^- capables d'oxyder les résidus méthionine cent fois plus rapidement que H_2O_2[172].

5.1.2. Conséquences de l'oxydation des résidus méthionine

L'oxydation des résidus méthionine au sein d'une protéine entraîne une modification des propriétés physico-chimiques de celle-ci. Nous allons voir que ce phénomène, potentiellement néfaste pour la structure de la protéine, peut se révéler être physiologiquement utile pour le métabolisme.

a. Modification de la structure des protéines

Il est bien établi que l'oxydation des chaînes latérales de résidus au sein d'une protéine peut entraîner des changements de conformation ou un dépliement de la protéine, et avoir des conséquences importantes sur sa fonction biologique. Ces observations s'appliquent également pour les méthionines, dont la chaîne latérale est particulièrement hydrophobe. Ainsi, les méthionines peuvent être présentes au sein des régions hydrophobes des protéines et leur oxydation en sulfoxydes, fonction polaire et hydrophile, crée une perturbation déstabilisatrice tendant à modifier la structure de la protéine afin d'amener ces sulfoxydes dans un environnement plus polaire[173].

b. Modulation de la fonction des protéines et implication pathologique

L'oxydation des résidus méthionine en sulfoxyde de méthionine peut être considérée comme une modification post-traductionnelle non enzymatique des protéines. Cette modification est impliquée notamment dans la modulation, voire même la perte de l'activité de la protéine. L'α1-antitrypsine fait partie de la famille des serpines, protéines inhibitrices de protéases. L'oxydation de deux résidus méthionine est responsable de la diminution de l'activité inhibitrice de la protéine vis-à-vis de sa cible, l'élastase, conduisant à une dérégulation de l'activité de la protéase[174]. Cette dérégulation est impliquée dans des phénomènes inflammatoires comme l'emphysème pulmonaire.

Récemment, il a été proposé que l'oxydation des résidus méthionine puisse aussi jouer un rôle dans le mécanisme de maladies neurodégénératives telles que la maladie d'Alzheimer. Cette pathologie est associée à la formation et au dépot de peptides amyloïdes dans le cerveau. La toxicité de ces peptides proviendrait de leur capacité à produire des EOA radicalaires dont la formation serait médiée par un résidu méthionine oxydé par un atome de cuivre lié à un peptide amyloïde[175,176]. L'oxydation du peptide amyloïde inhibe sa fibrillation et augmente sa solubilité[177], *i.e.* il est alors capable de catalyser la formation d'EOA et conduire à d'autres modifications du peptide[159].

c. Lutte contre les EOA

Le fait que les méthionines aient une grande sensibilité à l'oxydation et que, dans certains cas, leur conversion en sulfoxydes de méthionine n'entraîne aucune perturbation de la fonction de la protéine amène à penser que ces résidus peuvent servir de pièges à radicaux libres. En d'autres termes, les méthionines protègeraient les autres résidus de la protéine, essentiels à son activité. Pour exemple, nous citerons le cas de la glutamine synthétase[178].

d. Adressage vers les voies de dégradation

L'oxydation des méthionines fait partie des signaux d'adressage des protéines suroxydées vers les voies de dégradation. Pour exemple, nous savons que l'oxydation des résidus méthionine en sulfoxydes de méthionine peut conduire à des changements de conformation entraînant une exposition au milieu extérieur de zones hydrophobes normalement enfouies au sein de la protéine. Ces protéines sont dégradées par le protéasome 20S, à partir de la reconnaissance spécifique de ces zones hydrophobes exposées[179].

Le phénomène d'oxydation des méthionines est peut-être à mettre en correspondance avec celui de déamidation des résidus asparagine et glutamine dans le mécanisme d'adressage des substrats biologiques vers les voies de dégradation. En effet, les résidus Asn et Gln sont les deux seuls résidus thermodynamiquement instables dans les conditions physiologiques[180, 181]. Ils tendent à perdre la fonction amide à l'extrémité de leur chaîne latérale par un processus de dégradation non-enzymatique[182] qui a été caractérisé comme étant un signal d'adressage vers les voies de dégradation[183, 184]. La cinétique de la réaction de déamidation varie, non seulement, en fonction de la séquence de résidus adjacents dans la structure primaire, mais dépend aussi des structures secondaires, tertiaires et quaternaires[185]. Ce constat permet à Robinson *et al.*[186] de soutenir l'hypothèse, appelée *Molecular Clocks*, selon laquelle le rôle majeur des résidus Asn et Gln serait de mesurer le temps de vie d'une protéine. La dégradation par déamidation des résidus Asn et Gln d'une protéine marquerait la fin de vie de celle-ci et signalerait à l'organisme la dégradation de la protéine et la formation d'une nouvelle entité en remplacement.

5.1.3. Propriétés physico-chimiques de la fonction sulfoxyde

La fonction sulfoxyde est une fonction particulière dont les propriétés physico-chimiques diffèrent de celles habituellement connues pour des analogues comme la fonction carbonyle. Nous allons faire un rappel bibliographique des principales caractéristiques de cette fonction.

a. Aspects structuraux

Le diméthylsulfoxyde (DMSO) est le composé de référence de la famille des sulfoxydes. Plusieurs structures tridimensionnelles du DMSO ont été résolues par cristallographie[187]. La géométrie adoptée par le DMSO dans ces structures est tétraédrique. Elle est caractérisée par une distance moyenne entre l'atome de soufre et l'atome d'oxygène de 1,49 Å, un angle C–O–S de 107° et un

angle C–S–C de 100°. Une statistique faite sur 119 structures de sulfoxydes non coordinés conduit à une longueur de liaison S–O "référence" de 1,492 Å[187].

b. Nature de la liaison S–O

La nature de la liaison entre les atomes de soufre et d'oxygène des composés sulfoxyde est un sujet majeur de controverse. En effet, l'écriture formelle de la double liaison S = O covalente de la fonction sulfoxyde est en violation avec la règle de l'octet, en faisant passer le nombre d'électrons de valence de l'atome de soufre de 8 à 10. Pour cette raison, une structure a été proposée dans laquelle l'oxygène n'est lié au soufre que par une seule paire d'électrons partagés et provenant de l'atome de soufre. Ce type de liaison implique la perte d'un électron par le soufre et le gain d'un électron par l'oxygène. Il en résulte alors une charge positive portée par le soufre et une charge négative portée par l'oxygène. Ainsi, cette liaison pourrait être décrite comme la combinaison d'une liaison covalente simple et d'une liaison ionique. Ce type de liaison est appelé liaison semi-polaire[188]. Un équilibre de résonance entre les structures double liaison covalente et liaison semi-polaire peut être considéré (Fig. 5.3).

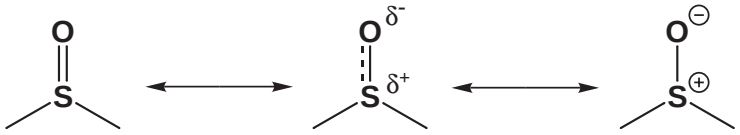


FIG. 5.3: *Formes limites de mésomérie de la fonction sulfoxyde. La liaison entre le soufre et l'oxygène peut être décrite sous deux formes limites : une double liaison (à gauche) ou une simple liaison avec des atomes chargés (à droite). La description couramment retenue est celle d'une liaison partiellement double (au centre), résultant d'un équilibre de résonance entre les deux formes limites.*

De nombreuses études expérimentales ont été entreprises afin de caractériser les propriétés physiques de la liaison S–O. La majorité des données expérimentales acquises est en faveur d'une structure hybride entre les deux formes de résonance, dominée par une forte contribution de la forme semi-polaire. La longueur de la liaison S–O de 1,492 Å observée dans les sulfoxydes est plus courte que la distance de 1,57 Å attendue pour une liaison simple S–O, suggérant un caractère de double liaison[189]. La spectroscopie par rayons X[190] a mis en évidence la polarisation des sulfoxydes et la présence d'une charge positive sur l'atome de soufre. La mesure d'un moment dipolaire de 3,96 D[191] conforte cette observation du caractère ionique des sulfoxydes. La force de la liaison S–O, calculée à partir de spectres IR, est en faveur d'une liaison semi-polaire.

Plusieurs études[192–194] de chimie théorique ont permis de rationaliser ces données expérimentales en montrant clairement que la liaison S–O des composés sulfoxyde est décrite comme la

contribution d'une liaison covalente σ partiellement ionique et fortement polarisée, d'une part, et d'une liaison π presque entièrement ionique, d'autre part. La densité de charge dans cette liaison π étant portée à plus de 90 % par l'atome d'oxygène, la participation de la liaison π peut être également représentée comme une paire d'électrons non partagée et portée par l'oxygène. Cette représentation conduit à décrire la liaison sulfoxyde comme une simple liaison σ polarisée dont la force dépend principalement des interactions électrostatiques. Les paramètres géométriques calculés à partir de cette description de la fonction sulfoxyde sont en excellent accord avec les mesures expérimentales, en particulier celles concernant la longueur de la liaison S–O, son moment dipolaire et sa fréquence de vibration IR.

c. Liaisons hydrogène et sulfoxyde protoné

L'oxygène de la fonction sulfoxyde est un bon accepteur de liaison hydrogène[195]. Cette propriété explique les excellentes capacités de solvatation du DMSO. Elle est, par ailleurs, indicative du caractère polaire de la liaison S–O de la fonction sulfoxyde. La capacité de la fonction sulfoxyde à accepter des liaisons hydrogène peut être reliée à son comportement de base faible. La valeur de pK_a de la fonction sulfoxyde de différents méthylsulfoxydes a été déterminée expérimentalement par titration dans l'acide sulfurique ou l'acide perchlorique aqueux, le degré de protonation du sulfoxyde étant mesuré par RMN et/ou par spectroscopie UV dans le cas de sulfoxydes aromatiques. Les valeurs de pK_a des différents sulfoxydes testés sont extrêmement basses et se situent toutes entre -2 et -5[196].

Des données structurales relatives à des sulfoxydes engagés dans des liaisons hydrogène et à des sulfoxydes protonés ont pu être obtenues par cristallographie de rayons X[187]. D'une longueur "référence" de 1,492 Å, la liaison S–O passe à 1,513 Å pour un sulfoxyde engagé dans une liaison hydrogène et à 1,587 Å pour un sulfoxyde protoné.

5.2. Méthionine sulfoxyde réductases

Les méthionine sulfoxyde réductases (Msr) sont des enzymes de la classe des oxydoréductases (EC 1.8.4.11 à 1.8.4.14) catalysant la réduction des sulfoxydes de méthionine (MetSO), libres ou inclus au sein d'une protéine. Ces enzymes sont présentes chez de nombreuses espèces vivantes, tant chez les animaux que les végétaux et les bactéries. Elles sont localisées à la fois au niveau mitochondrial, au niveau du noyau et dans le cytoplasme[197–199].

L'oxydation de la fonction thioéther de la méthionine conduit à la formation de deux stéréoisomères *R* et *S* de la fonction sulfoxyde au niveau de l'atome de soufre. Les Msr de classe A (MsrA) réduisent l'isomère *S* de la fonction sulfoxyde et les Msr de classe B (MsrB), l'isomère *R*, comme illustré par la figure 5.4.

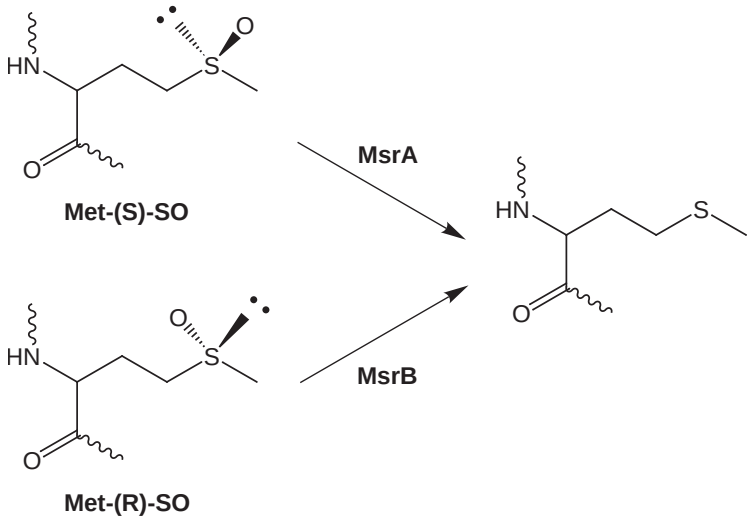


FIG. 5.4: *Stéréospécificité de la réduction enzymatique des sulfoxydes de méthionine. L'isomère S est réduit par les Msr de classe A (MsrA) et l'isomère R par les Msr de classe B (MsrB).*

Les premières évidences de l'activité Msr ont été rapportées en 1979 par l'équipe de Brot et Weissbach[200] pour MsrA et en 2001 par l'équipe de Grimaud *et al.*[201] pour MsrB.

5.2.1. Implication physiologique des méthionine sulfoxyde réductases

L'importance physiologique des Msr a été particulièrement documentée au cours de ces dix dernières années. Sans entrer dans les détails, nous présenterons ici différentes implications de ces enzymes et notamment leur rôle dans la lutte contre le stress oxydant, dans la régulation de la fonction des protéines et dans le vieillissement cellulaire.

a. Lutte contre le stress oxydant

La première mise en évidence de la fonction MsrA *in vivo* a été obtenue à partir de l'étude d'une souche d'*Escherichia coli* n'exprimant plus l'enzyme[202]. Cette souche présente une plus grande sensibilité à H_2O_2. Des études ultérieures ont confirmé le fait que la délétion ou l'inactivation du gène *msrA* implique une plus grande sensibilité des microorganismes à l'oxydation par les EOA[203–208]. Des résultats obtenus chez les organismes eucaryotes supérieurs sont en accord avec les obser-

vations précédentes[209]. D'autre part, une surexpression de MsrA dans les lymphocytes T humains augmente la résistance des cellules au stress oxydant induit par H_2O_2[210].

Des travaux de même type sur les cellules épithéliales de la lentille humaine ont montré que la classe d'enzyme MsrB joue le même rôle que la classe MsrA. La réduction de l'expression du gène *msrB* par l'utilisation de petits ARN interférants conduisant à une diminution de la viabilité des cellules et de leur résistance au stress oxydant[211].

b. Régulation de la fonction des protéines

L'oxydation des résidus méthionine en sulfoxydes de méthionine peut entraîner la modulation de l'activité biologique d'une protéine. La possibilité qu'offrent les Msr de réduire ces résidus oxydés a conduit à proposer que ce mécanisme d'oxydation/réduction peut constituer, dans une certaine mesure, une voie de régulation. Dans une étude *in vitro*[212], la réparation par MsrA (ou MsrB) d'une calmoduline oxydée a entrainé la désinhibition de la fonction activatrice de cette enzyme vis-à-vis de la Ca-ATPase. Cet exemple illustre bien l'implication du cycle **Oxydation des Met par EOA** $\leftrightarrows$ **Réduction des MetSO par Msr** dans des mécanismes rédox régulant le métabolisme cellulaire, *i.e.* une accumulation de calmoduline oxydée lors d'un stress oxydant sévère entraîne une répression du métabolisme de la cellule afin de diminuer la génération d'EOA.

c. Vieillissement cellulaire

Le vieillissement cellulaire est caractérisé par l'accumulation de protéines oxydées dans les cellules. Initialement, la cause de cette accumulation a été attribuée à une baisse de l'activité du protéasome[213] et de la protéase mitochondriale Lon[214] au cours de la vie d'une cellule. La possibilité que des systèmes de réparation des protéines oxydées, comme les Msr, puissent également être impliqués dans la diminution, liée à l'âge, de l'activité des protéines a été récemment envisagée. Chez le rat[215], une diminution significative de l'activité Msr a été décelée au niveau du foie, du rein et du cerveau durant le vieillissement cellulaire. Cette même diminution a également été observée dans des fibroblastes humains au cours de leur vieillissement[216], corrélée avec la diminution de l'expression des gènes codant l'enzyme MsrA et une enzyme MsrB. Ces données indiquent que les capacités de réparation des Msr sont diminuées lors du vieillissement, contribuant probablement à l'accumulation de protéines oxydées.

5.2.2. Mécanisme catalytique

Les Msr sont des enzymes ne nécessitant aucun cofacteur rédox pour la réduction des sulfoxydes de méthionine. Boschi-Muller *et al.*[217] ont démontré que le mécanisme se déroule en trois étapes en passant par la formation d'un intermédiaire acide sulfénique, produit d'oxydation d'une cystéine. La régénération de l'enzyme vers sa forme réduite nécessite la présence d'un réducteur qui est la thiorédoxine (Trx) *in vivo*. Toutes les MsrA et les MsrB présentent un même mécanisme catalytique de type "ping-pong"[217,218] assimilable aux deux équations suivantes :

$$\mathrm{Msr_{réd}} + \mathrm{MetSO} \rightarrow \mathrm{Msr_{ox}} + \mathrm{Met}$$
$$\mathrm{Msr_{ox}} + \mathrm{Trx_{réd}} \rightarrow \mathrm{Msr_{réd}} + \mathrm{Trx_{ox}}$$

La réduction des sulfoxydes de méthionine catalysée par les Msr appartient à un cycle catalytique formé de plusieurs réactions d'oxydoréduction, comme l'illustre la figure 5.5. La thiorédoxine oxydée ($\mathrm{Trx_{ox}}$) est réduite par le nicotinamide adénine dinucléotide phosphate (NADPH) *via* la thiorédoxine réductase (TrxR).

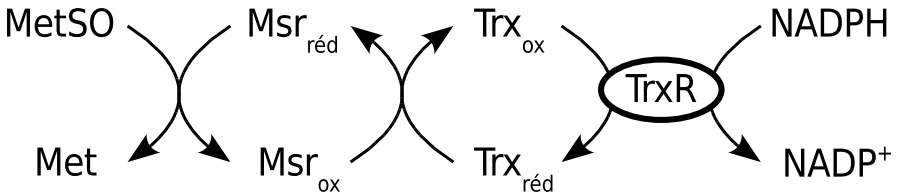


FIG. 5.5: *Cycle catalytique des méthionine sulfoxyde réductases. La méthionine sulfoxyde (MetSO) est réduite en méthionine (Met) par la méthionine sulfoxyde réductase réduite ($\mathrm{Msr_{réd}}$). La méthionine sulfoxyde réductase oxydée ($\mathrm{Msr_{ox}}$) est réduite par la thiorédoxine réduite ($\mathrm{Trx_{réd}}$). La thiorédoxine oxydée ($\mathrm{Trx_{ox}}$) est réduite par le nicotinamide adénine dinucléotide phosphate (NADPH), catalysé par la thiorédoxine réductase (TrxR).*

Les travaux de Boschi-Muller *et al.*[217] pour la classe MsrA et d'Olry *et al.*[218] pour la classe MsrB ont montré que les Msr sont des enzymes à cystéines essentielles. Plusieurs résidus cystéine sont impliqués dans le mécanisme catalytique. La figure 5.6, tirée des travaux de Boschi-Muller *et al.* et d'Olry *et al.*, présente un mécanisme à deux cystéines.

Lors de la première étape du mécanisme (**I**), appelée étape réductase, le substrat sulfoxyde de méthionine est réduit en méthionine dans un rapport stœchiométrique 1 : 1. Une première cystéine, appelée cystéine catalytique ($\mathrm{Cys_C}$), est oxydée en acide sulfénique. Les deux étapes suivantes sont dévouées à la régénération de l'enzyme. Une deuxième cystéine, appelée cystéine de recyclage ($\mathrm{Cys_R}$), se lie avec $\mathrm{Cys_C}$ par formation d'un pont disulfure et élimination d'une molécule d'eau (**II**). Finalement, la Trx vient réduire le pont disulfure de la Msr oxydée (**III**). La formation de l'intermédiaire

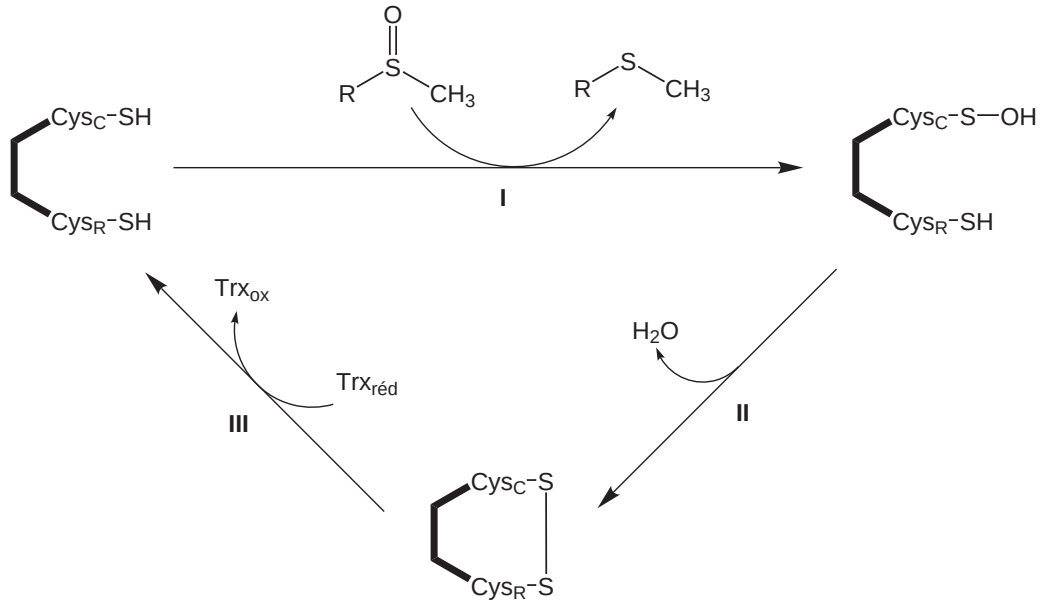


FIG. 5.6: *Mécanisme catalytique des Msr faisant intervenir deux résidus cystéine. Le substrat MetSO est réduit en méthionine lors de l'étape réductase (**I**) avec oxydation de Cys_C en intermédiaire acide sulfénique. La régénération de l'enzyme s'effectue par formation d'un pont disulfure entre Cys_C et Cys_R avec l'élimination d'une molécule d'eau (**II**) et la réduction du pont disulfure par la Trx (**III**), d'après Boschi-Muller et al.[217] et Olry et al.[219].*

acide sulfénique a été mise en évidence à partir d'une mutation de l'enzyme. La substitution de la cystéine de recyclage par une sérine permet l'arrêt du mécanisme après l'étape réductase et l'accumulation de la forme oxydée de la Msr dans le milieu[217,219].

5.2.3. Structure primaire des méthionine sulfoxyde réductases

L'alignement des séquences de différentes MsrA et MsrB montre que la cystéine catalytique Cys_C est strictement conservée au sein de ces enzymes. Cependant, le nombre de cystéines de recyclage Cys_R (et leurs positions respectives) peut varier d'un organisme à l'autre. Certaines Msr présentent des cystéines additionnelles dans leurs structures primaires. Ceci a permis de définir des sous-classes au sein des MsrA et MsrB, d'après la revue de Kauffmann *et al.*[133].

a. Sous-classes de MsrA

L'alignement des séquences de différentes MsrA est reporté dans la figure 5.7. Pour chacune d'entre elles, Cys_C est localisée à la position 51 au sein du motif conservé $_{50}GCFW_{53}$. Cette séquence est caractéristique des MsrA. La numérotation des résidus des MsrA est basée sur la séquence référence de la MsrA d'*Escherichia coli*.

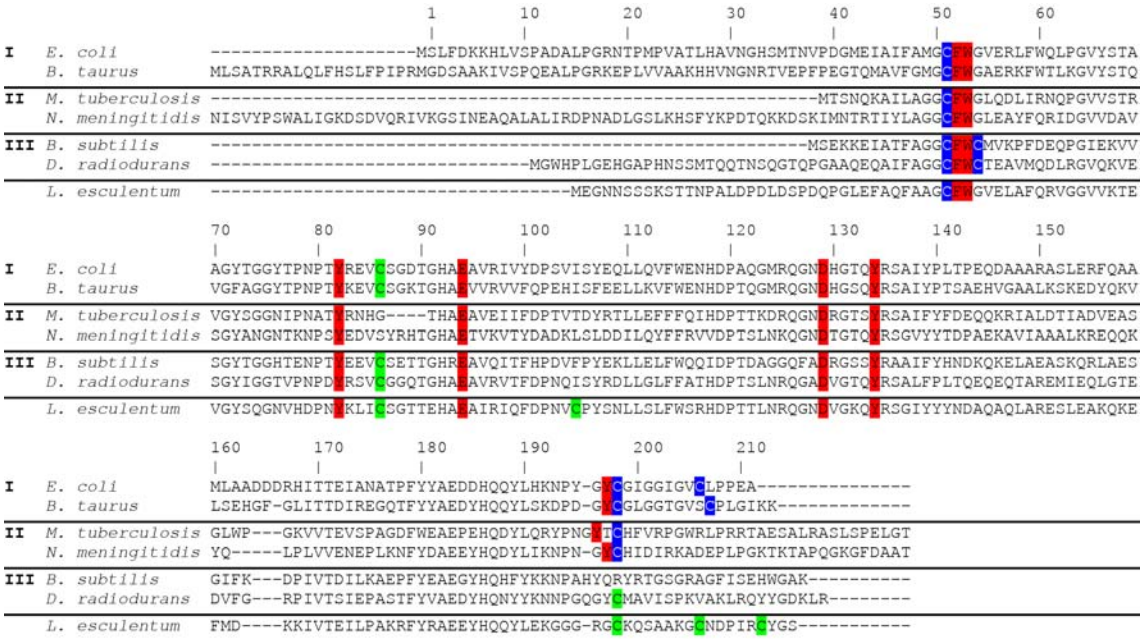


FIG. 5.7: *Alignement des séquences des MsrA de E. coli, B. taurus, M. tuberculosis, N. meningitidis, D. radiodurans, B. subtilis et L. esculentum. Les séquences sont regroupées par sous-classe. La première ligne indique la numérotation de séquence basée sur E. coli. Les cystéines impliquées dans le mécanisme catalytique sont surlignées en bleu, les autres en vert et les résidus du site actif en rouge. Tiré de Kauffmann et al.*[133].

MsrA$_{I}$ Cette sous-classe est caractérisée par la présence de deux Cys$_R$ en position 198 et 206 (ou 207). La première appartient au motif conservé $_{196}$GYCG$_{199}$ et est séparée de la seconde par une séquence riche en Gly. Lors de l'étape de régénération de l'enzyme, deux ponts disulfure sont formés consécutivement : 51–198, puis 198–206. Ce dernier est réduit préférentiellement par la Trx. Les MsrA d'*E. coli*, humaine et bovine (*Bos taurus*) sont représentatives de cette sous-classe.

MsrA$_{II}$ Cette sous-classe est caractérisée par la présence d'une Cys$_R$ en position 198. Pour la plupart des MsrA, cette Cys$_R$ appartient au motif conservé $_{196}$GYCH$_{199}$, tel est le cas de la MsrA de *Neisseria meningitidis*. Certains organismes comme *Mycobacterium tuberculosis* présentent l'insertion d'un résidu avant Cys$_R$, selon le motif GYxCH.

MsrA$_{III}$ Cette sous-classe est caractérisée par la présence d'une Cys$_R$ en position 54. Le motif $_{50}$GCFWG$_{54}$ commun aux autres sous-classes devient donc $_{50}$GCFWC$_{54}$ pour celle-ci. La MsrA de *Bacillus subtilis* est représentative de cette sous-classe. On remarque aussi que certaines MsrA de cette sous-classe comme celle de *Deinococcus radiodurans* possèdent une autre cystéine en position 198. L'implication de celle-ci dans le processus catalytique reste inconnu.

La sous-classe des MsrA végétales, comme celle de *Lycopersicon esculentum*, se caractérise par la présence du motif conservé $_{50}$GCFWG$_{54}$ contenant Cys$_C$ et de cinq résidus cystéine supplémentaires. Deux d'entre eux sont situés dans la chaîne C-terminale et sont conservés à des positions identiques. Il se pourrait que ces cystéines participent au processus catalytique[220,221]. Récemment, une enzyme MsrA active d'*Alkaliphilus oremlandii* à sélénocystéine a été découverte[222].

b. Sous-classes de MsrB

L'alignement des séquences de différentes MsrB est reporté dans la figure 5.8. Pour chacune d'entre elles, Cys$_C$ est localisée dans la chaîne C-terminale de la protéine, à la position 117 au sein du motif conservé $_{116}$YCxN$_{119}$. La numérotation des résidus des MsrB est basée sur la séquence référence de la MsrB d'*E. coli*. Les mammifères ont deux MsrB où Cys$_C$ est un résidu Cys et une MsrB où Cys$_C$ est remplacée par une sélénocystéine. Ces protéines à SeCys ayant une activité MsrB sont appelées SelX[223] ou SelR[224,225].

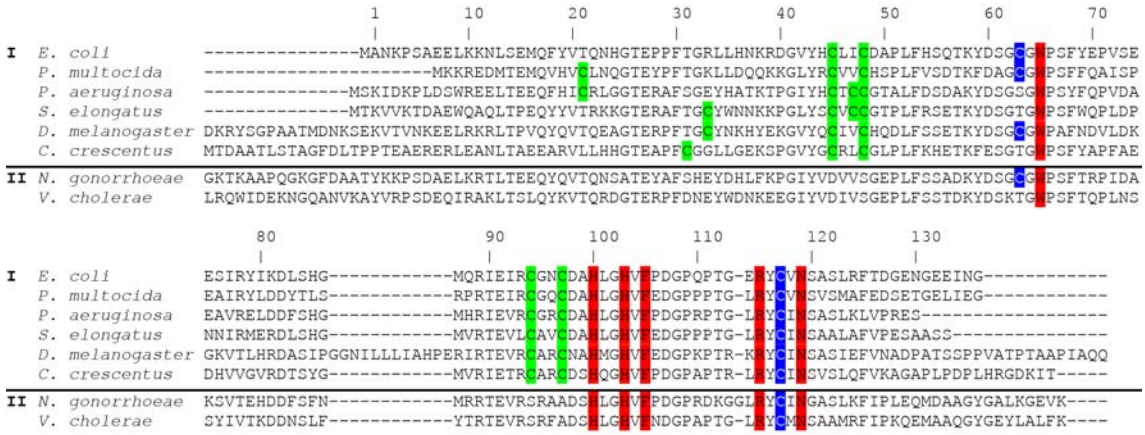


FIG. 5.8: *Alignement des séquences des MsrB de E. coli, P. multocida, P. aeruginosa, S. elongatus, D. melanogaster, C. crescentus, N. gonorrhoeae et V. cholerae. Les séquences sont regroupées par sous-classes. La première ligne indique la numérotation de séquence basée sur E. coli. Les cystéines impliquées dans le mécanisme catalytique sont surlignées en bleu, les autres en vert et les résidus du site active en rouge. Tiré de Kauffmann et al.[133].*

MsrB$_I$ Cette sous-classe est caractérisée par la présence de deux motifs CxxC additionels. Ces quatres résidus conservés aux positions 45-48 et 94-97 ont été mis en évidence dans la fixation d'un cation métallique Zn^{2+} nécessaire à la structuration de l'enzyme[225]. Ce centre métallique n'est pas impliqué dans le mécanisme de réduction. Les SelX, SelR et les MsrB d'*E. coli* font partie de cette sous-classe.

$MsrB_{II}$ Cette sous-classe regroupe toutes les MsrB non-métalliques. Elles représentent près de 60% des MsrB connues. Les enzymes exprimées par les bactéries de type *Neisseria* sont classées dans ce groupe.

En plus de Cys_C, la plupart des enzymes possèdent une autre cystéine conservée en position 63 et considérée comme étant Cys_R. Au cours de la régénération de Msr_{ox}, elle forme un pont disulfure avec Cys_C qui est, par la suite, réduit par la Trx[219,225]. Cependant, certaines organismes comme *Pseudomonas aeruginosa* ou *Caulobacter crescentus* ont une MsrB ne présentant pas cette Cys_R en position 63. Si ces enzymes s'avéraient être actives, ce sont d'autres cystéines (non conservées) qui pourraient jouer le rôle de Cys_R. Elles sont situées dans la chaîne N-terminale, aux alentours des positions 20-30. En cas d'absence de cystéine autre que Cys_C au sein de l'enzyme, comme pour *Vibrio cholerae*, le processus de régénération doit être réalisé selon un mécanisme encore inconnu à ce jour.

5.2.4. Structure tridimensionnelle des méthionine sulfoxyde réductases

Bien que partageant un mécanisme catalytique similaire, nous venons de voir que les MsrA et les MsrB ne présentent aucune identité ni aucune homologie de structure primaire. Leurs structures tridimensionnelles se sont également révélées être différentes. A ce jour (mars 2009), les structures de MsrA de quatres organismes distincts et les structures de MsrB de deux organismes distincts ont été résolues par cristallographie de rayons X et déposées dans la Protein Data Bank[2] (PDB). Des informations les concernant sont regroupées dans le tableau 5.2. D'autres structures ont été déposées à la PDB mais ne donnent que des structures partielles de Msr.

	Organisme	Code PDB	Référence	Résolution (Å)
MsrA	*Escherichia coli*	1FF3	Tête-Favier *et al.*, 2000[226]	1,9
	Bos taurus	1FVA	Lowther *et al.*, 2000[227]	1,7
		1FVG	Lowther *et al.*, 2000[227]	1,6
	Mycobacterium tuberculosis	1NWA	Taylor *et al.*, 2003[228]	1,5
	Neisseria meningitidis	3BQE	Ranaivoson *et al.*, 2008[229]	2,0
		3BQF	Ranaivoson *et al.*, 2008[229]	2,0
		3BQG	Ranaivoson *et al.*, 2008[229]	2,0
		3BQH	Ranaivoson *et al.*, 2008[229]	2,0
MsrB	*Neisseria gonorrhoeae*	1L1D	Lowther *et al.*, 2002[230]	1,85
	Burkholderia pseudomallei	3CEZ	Staker *et al.*, 2008*	2,1
		3CXK	Lovell *et al.*, 2008*	1,7

TAB. 5.2: *Structures de méthionine sulfoxyde réductases déposées à la PDB.*
* *En cours de publication.*

Dans les structures 3BQE, 3CEZ et 3CXK, les enzymes ont été cristallisées seules, sous forme d'enzyme libre. Les autres structures présentent des ligands, des modifications de l'enzyme ou des interactions utiles à la compréhension du mécanisme catalytique de l'enzyme :

1FF3 Cys_C51 a été modifié en S-(diméthylarsénic)cystéine (CAS).

1FVA Un résidu alanine appartenant à un monomère voisin se retrouve à l'intérieur du site actif de l'enzyme sous l'effet de la cristallisation (*crystal packing*).

1FVG Une molécule de dithiothréitol (DTT) est liée de façon covalente à Cys_C51 et Cys_R198.

1NWA Un résidu méthionine appartenant à un monomère voisin se retrouve à l'intérieur du site actif de l'enzyme sous l'effet de la cristallisation (*crystal packing*).

3BQF Cys_C51 a été muté en sérine. Un peptide MetSO est présent dans le site actif.

3BQG L'acide sulfénique est formé sur Cys_C51 après diffusion de DMSO dans le cristal.

3BQH Un pont disulfure est formé entre Cys_C51 et Cys_R198.

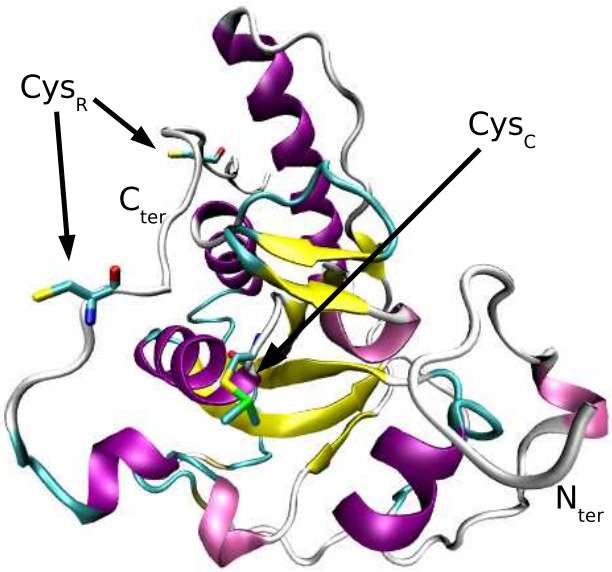
1L1D Une molécule de cacodylate (CAC) est présente dans le site actif.

a. Repliement et structure du site actif des MsrA

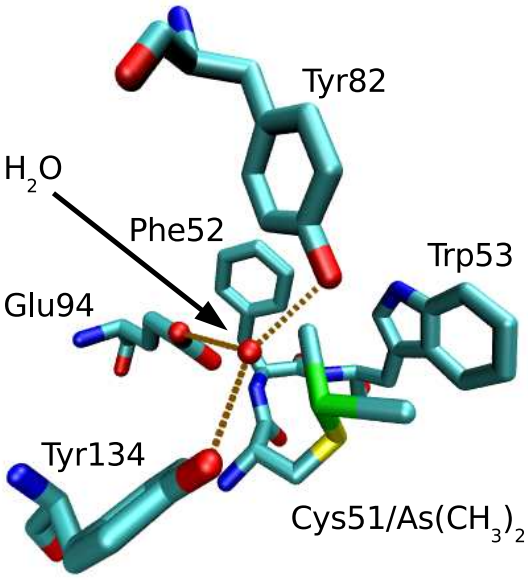
La figure 5.9 représente la structure tridimensionnelle de la MsrA d'*E. coli*. La partie centrale est composée de 140 résidus, entre les positions 42 et 192 chez *E. coli*. Elle adopte un repliement mixte α/β unique[226], *i.e.* ne partageant aucune similarité avec les repliements déjà répertoriés dans les banques de données structurales. Ce repliement est constitué d'un feuillet β antiparallèle à cinq brins, flanqué de quatres hélices α et d'un petit feuillet β antiparallèle à 2 brins.

L'extrémité N-terminale ne présente aucun élément de structure secondaire. Cependant, son positionnement est fortement stabilisé par la formation de nombreuses liaisons hydrogène ou de type van der Waals avec la partie centrale. Pour s'en convaincre, les structures de la PDB se superposent remarquablement bien au niveau de cette extrémité. La longueur de cette extrémité est très variable : de 5 résidus chez *N. meningitidis* à une quarantaine chez *E. coli*. Peu de résidus sont conservés dans cette partie et aucun rôle catalytique spécifique n'a été attribué à cette partie, mis à part de diminuer légèrement la cinétique de réduction[231].

L'extrémité C-terminale ne présente pas non plus d'élément de structure secondaire. Par contre, à la différence de la partie N-terminale, sa conformation varie légèrement selon les structures considérées. Les cystéines de recyclage, aux positions 198 et 206, se situent sur cette extrémité.



(a)



(b)

FIG. 5.9: *Structure tridimensionnelle de la MsrA d'E. coli*[226].
(a) Architecture globale de l'enzyme. Les structures secondaires ont été colorées en violet pour les hélices α, en jaune pour les feuillets β et en bleu clair pour les coudes β. La cystéine catalytique Cys_C a été représentée en bâtons, ainsi que les cystéines de recyclage Cys_R de l'extrémité C-terminale.
(b) Structure du site actif. Ce site actif est composé des résidus conservés Phe52, Trp53, Tyr82, Glu94 et Tyr134. L'adduit diméthylarséniate cristallographique $As(CH_3)_2$ présent sur la Cys_C51 a été représenté (atome d'arsenic en vert). La position de celui-ci, ainsi que la molécule d'eau en interaction avec les résidus du site actif, mime la position du substrat MetSO. Les différentes interactions possibles entre résidus ont été symbolisées en marron.

Le site actif est organisé autour de Cys_C51, située à l'extrémité N-terminale d'une hélice α. Les résidus qui l'entourent - tous conservés dans l'alignement de séquences des MsrA - forment une large poche ouverte qui offre une grande accessibilité vers Cys_C, adaptée à la réduction de substrats enfouis au sein de protéines oxydées. Il présente une grande quantité de chaînes latérales aromatiques (Phe52, Trp53, Tyr82, Tyr134, Tyr197), ce qui lui confère un caractère partiellement hydrophobe, et plusieurs donneurs/accepteurs de liaisons hydrogène.

Le site actif de MsrA est spécifique de la réduction de l'isomère *S* de MetSO. Certaines structures cristallographiques de MsrA, comme celle de la mutation C51S (3BQF), permettent d'avoir une structure approchée du complexe entre le sulfoxyde et le site actif de l'enzyme. Ceci nous permet d'obtenir des indications vis-à-vis du positionnement du substrat ou, au moins, vis-à-vis des résidus impliqués dans sa reconnaissance[227,230]. Ces hypothèses de reconnaissances structurales ont été confirmées expérimentalement par Gand *et al.*[232] :

- le groupement ε-méthyle de MetSO est stabilisé au sein d'une poche hydrophobe formée par les invariants Phe52 et Trp53. Le noyau indole de ce dernier établit lui-même une liaison hydrogène avec le noyau imidazole de His186 et une interaction hydrophobe de type π-stacking avec le noyau phénol de Tyr189 ;
- la chaîne latérale de Asp129 peut établir une liaison hydrogène avec le groupement NH de la chaîne principale du substrat MetSO ;
- Tyr197 peut interagir avec la chaîne principale du substrat MetSO *via* une interaction de type π-π entre son noyau aromatique et la liaison peptidique du côté N-terminal ;
- les résidus Tyr82, Glu94 et Tyr134 interagissent fortement *via* des liaisons hydrogène avec une molécule d'eau présente dans chaque structure. L'oxygène de cette molécule d'eau est bien positionné pour mimer l'oxygène de la fonction sulfoxyde du substrat ;
- de manière annexe, la distance mesurée entre la cystéine catalytique en position 51 et la (ou les) cystéine(s) de recyclage en position 198 et/ou 206 est trop importante pour permettre la formation "directe" du pont disulfure lors de l'étape de régénération. Ces résultats couplés avec les résultats enzymatiques suggèrent fortement l'implication d'une réorganisation conformationnelle locale lors de la formation de ce pont disulfure, successive ou simultanée avec la formation de l'intermédiaire acide sulfénique.

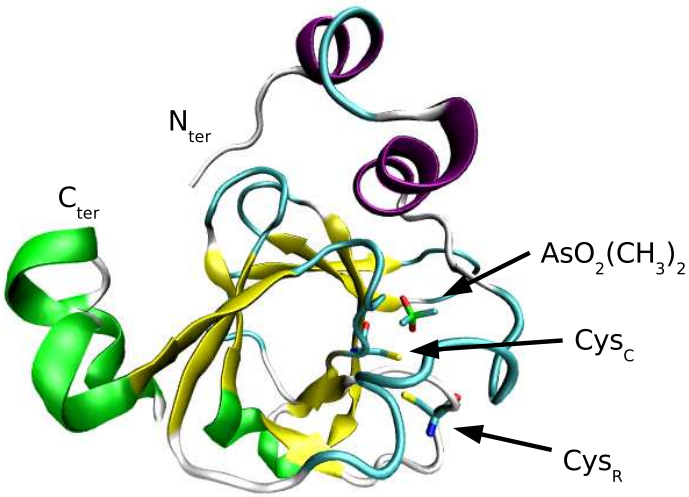
b. Repliement et site actif de MsrB

La figure 5.10 représente la structure tridimensionnelle de la MsrB de *N. gonorrhoeae*. La structure globale des MsrB est totalement différente de celle des MsrA. Elle est caractérisée par un repliement essentiellement de type β[230]. La partie centrale de l'enzyme est formée par deux feuillets β antiparallèles se faisant face pour donner une structure proche du tonneau β et flanquée de trois hélices α et de deux hélices 3_{10}. Les boucles qui relient les brins β sont relativement courtes et adoptent un repliement en coude β. Le cœur de l'enzyme est constitué d'une quarantaine de résidus situés sur des chaînes peu structurées, essentiellement des coudes β. L'extrémité C-terminale constituée de deux petites hélices 3_{10} antiparallèles est plus courte. Deux hélices α structurent la chaîne N-terminale. Les deux extrémités de l'enzyme ne possèdent aucun résidu impliqué dans le mécanisme catalytique. Contrairement aux MsrA, cette structure de MsrB possède des homologues dans les bases de données structurales comme, par exemple, les protéines Mss4 et TCTP.

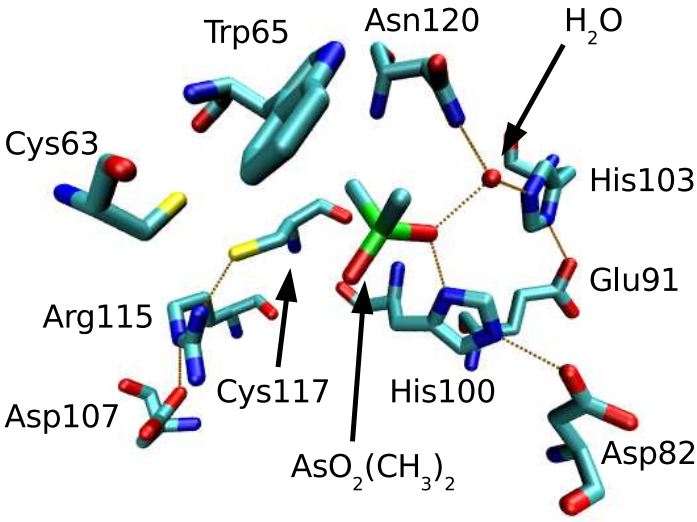
Le site actif de MsrB est, de la même façon que celui de MsrA, structuré autour de Cys_C117 localisée sur un brin β. La chaîne latérale de cette cystéine n'est pas orientée vers le tonneau β qui offre pourtant une large cavité accessible. Elle se situe au fond d'une poche secondaire à la surface de l'enzyme qui constitue le site actif. L'architecture du site actif des MsrA et MsrB ne présente aucune similarité. La proportion de résidus aromatiques est largement inférieure. Les résidus conservés Trp65, His100, His103, Arg115 et Asn120 sont présents dans le site actif.

Dans les conditions utilisées pour effectuer la cristallographie de rayons X de la MsrB de *N. meningitidis*[230], une molécule du tampon cacodylate $AsO_2(CH_3)_2$ a été complexée dans le site actif de l'enzyme. Ce ligand peut jouer le rôle de mime du substrat MetSO, permettant de proposer quelques hypothèses quant à la reconnaissance du subtrat par le site actif[230] :

- un azote du noyau imidazole de His100, d'une part, et une molécule d'eau en liaison hydrogène avec Asn120 et His103, d'autre part, forment un réseau de liaisons hydrogène avec l'oxygène du sulfoxyde ;
- le groupement ε-methyle du sulfoxyde est stabilisé dans une poche hydrophobe formée par le résidu Trp65 ;
- un réseau important d'interactions électrostatiques entre Asp107, Arg115 et Cys63 permet de stabiliser Cys_C sous une forme thiolate.



(a)



(b)

FIG. 5.10: *Structure tridimensionnelle de la MsrB de N. gonorrhoeae*[230].

(a) Architecture globale de l'enzyme. Les structures secondaires ont été colorées en violet pour les hélices α, en vert pour les hélices 3_{10}, en jaune pour les feuillets β et en bleu clair pour les coudes β. La cystéine catalytique Cys_C, la cystéine de recyclage Cys_R et la molécule de cacodylate $AsO_2(CH_3)_2$ complexée dans le site actif ont été représentées en bâtons.

(b) Structure du site actif. Ce site actif est composé des résidus conservés Cys_R63, Trp65, Asp82, Glu91, His100, His103, Asp 107, Arg115, Cys_C117 et Asn120. La molécule de cacodylate $AsO_2(CH_3)_2$ complexée a été représentée (atome d'arsenic en vert). Sa position mime celle du substrat MetSO. Les différentes interactions possibles entre résidus ont été symbolisées en marron.

5.3. Etudes antérieures sur l'étape réductase de MsrA

Un schéma général du mécanisme catalytique des Msr a déjà été présenté dans la section 5.2.2 (page 98). Nous allons nous intéresser de plus près maintenant à l'étape réductase du cycle catalytique (**I**) pour la classe d'enzyme MsrA et nous allons résumer l'état de connaissances actuelles.

5.3.1. Caractérisation cinétique

Les travaux expérimentaux d'Antoine *et al.*[233] et d'Olry *et al.*[218] sur le mécanisme catalytique de la MsrA et de la MsrB de *N. meningitidis* respectivement ont montré que la cinétique de l'étape réductase (**I**) est limitante dans le processus de formation du pont disulfure (**I** + **II**). Rappelons que l'étape limitante du processus catalytique global est l'étape de réduction de ce pont disulfure par Trx (**III**). Des résultats similaires ont été obtenus pour d'autres Msr.

5.3.2. Identification des résidus impliqués

Les travaux de Gand *et al.*[232] mettent en évidence les résidus responsables de la reconnaissance du substrat MetSO par le site actif de la MsrA. De plus, les structures cristallographiques[226–229] de MsrA montrent que les résidus invariants Glu94, Tyr82 et Tyr134 sont en interaction de type liaison hydrogène avec une molécule d'eau, mime de l'oxygène de la fonction sulfoxyde.

Dans une étude menée par Antoine *et al.*[234], les mutations E94A et E94D de la MsrA de *N. meningitidis* diminuent de façon drastique la constante cinétique relative à l'étape réductase. Ces résultats suggèrent que Glu94 est impliqué dans la catalyse de la réduction du sulfoxyde (k_{cat} différent), mais pas dans la formation du complexe michaélien (K_M identique). Ce résidu a été identifié comme étant le catalyseur acide du mécanisme. Cette conclusion est renforcée par le fait qu'aucun autre résidu à caractère acide ne se situe à proximité de la fonction sulfoxyde. La mutation E94Q implique aussi une diminution de la constante cinétique, mais dans une moindre mesure.

La substitution d'une des deux tyrosines (Tyr82 ou Tyr134) par une phénylalanine a une influence modérée sur le mécanisme. La constante cinétique diminue légèrement. Par contre, cette constante diminue de façon importante lors de l'absence des deux tyrosines. Il a été avancé par Antoine *et al.* que, lorsqu'une seule tyrosine est enlevée, Glu94 est sollicité afin de compenser la perte de stabilisation. Le principal rôle de ces tyrosines semble être celui de stabiliser la formation du complexe

michaélien entre le substrat oxydé et l'enzyme et, par la suite, de stabiliser les éventuels intermédiaires de réaction lors de leurs formations.

5.3.3. Etat de protonation des résidus du site actif

Cette même étude d'Antoine *et al.*[234] a permis de déterminer la valeur de pK_{app} des différents résidus du site actif. Dans l'enzyme MsrA libre de *N. meningitidis*, les deux résidus cystéine (Cys_C51 et Cys_R198) ont un pK_{app} d'environ 9,5. Cette valeur se trouve être légèrement supérieure à celle généralement constatée de 8,8 dans d'autres protéines[235].

En phase réductase, la plus grande constante cinétique est observée à un pH de 8. La valeur du pK_a de Cys_C51 passe d'environ 9,5 à moins de 7. La dépendance de la constante cinétique vis-à-vis du pH fait apparaître un pK_{app} de 5,7. Cependant, les mutations E94Q et Y82F/Y134F/E94Q de cette même enzyme montrent toujours la contribution d'un seul pK de 8 et 9,5 respectivement. Ces résultats permettent d'attribuer la valeur pK_{app} de 5,7 à Cys_C51, et non à Glu94. De ce fait, la formation du complexe michaélien MsrA-MetSO permet l'activation de Cys_C51 en abaissant son pK_{app} de 3,8 unités. La cystéine catalytique est ainsi activée par déprotonation afin d'initier la réduction du sulfoxyde.

5.3.4. Travaux théoriques sur la réduction d'un sulfoxyde par un thiol

Balta *et al.*[236] ont publié une étude théorique du mécanisme de réduction d'un sulfoxyde par un thiol en solution. Cette étude fournit un mécanisme réactionnel potentiellement transférable à la catalyse enzymatique de réduction des MetSO. Parmi les différents chemins réactionnels considérés, Balta *et al.* décrivent un chemin conduisant à la formation d'un acide sulfénique, comme il a été décrit par Boschi-Muller *et al.*[217]. Ce travail apporte des informations concernant la structure et l'énergie relative des différentes espèces caractérisées le long du chemin. Les résultats de cette étude seront volontairement détaillés ci-dessous car ils apportent certaines informations importantes vis-à-vis de l'étude enzymatique menée durant ce travail de thèse.

a. Système étudié et méthodologie

Le travail vise la réduction d'un sulfoxyde par un thiol en solution de diméthylsulfoxyde (DMSO). Les auteurs ont considéré une molécule de méthanethiol CH_3SH et une molécule de DMSO. Ces deux molécules jouent le rôle de Cys_C et du substrat oxydé respectivement dans le cas de la réaction

enzymatique. Le solvant DMSO est modélisé en utilisant un modèle mixte discret-continuum. Une deuxième molécule de DMSO (solvant) est donc introduite explicitement dans le système. L'insertion de cette molécule explicite de solvant vise la prise en compte des interactions spécifiques qui pourraient jouer un rôle important dans la stabilisation des intermédiaires de réaction en formant des liaisons hydrogène. Le modèle de continuum utilisé[237–240] est celui développé dans notre laboratoire ; on utilise la constante diélectrique du DMSO ($\varepsilon = 46,7$). Toutes les géométries ont été optimisées au niveau de théorie B3LYP/6-311G(d,p) et caractérisées par un calcul de fréquences. Les minima ont été connectés aux états de transition (TS) par l'intermédiaire d'un calcul d'IRC (*intrinsic reaction coordinate*). Un calcul plus poussé de l'énergie a été effectué au niveau MP2/6-311+G(3d2f,2df,2p) sur les géométries optimisées.

b. Résumé des résultats et discussion

La première partie du chemin réactionnel est illustrée par la figure 5.11. La structure **1** représente le réactif méthanethiol en interaction avec deux molécules explicites du solvant DMSO. Cette structure est prise comme référence des énergies. La structure **2** représente un système où un transfert de proton a été réalisé entre le méthanethiol et une molécule de DMSO. La deuxième molécule de DMSO vient former une liaison hydrogène forte avec ce proton. A partir du complexe ionique **2**, la formation de l'acide sulfénique **4** peut être envisagée mais semble très difficile, la barrière énergétique pour atteindre l'état de transition $\mathbf{TS_{2-4}}$ étant de 45,5 kcal/mol. Un autre chemin conduit à une espèce de type sulfurane **3** suite au réarrangement du complexe ionique **2**. Cette étape implique une toute petite barrière énergétique (20 kcal/mol). Signalons que les auteurs ont également considéré le passage direct du système initial **1** à l'acide sulfénique **4** selon un mécanisme concerté mais il s'avère peu probable. L'état de transition correspondant se situe à 46,3 kcal/mol au-dessus de **1** (chemin non montré).

Le solvant joue un rôle crucial dans l'énergétique de ce processus compte tenu du caractère polaire des espèces en présence. L'existence d'espèces chargées dans le système implique la présence d'interactions électrostatiques importantes. La deuxième molécule explicite de solvant intervient dans une grande proportion dans la stabilisation des différents intermédiaires de réaction, comme on peut le voir pour les structures **2** à **4**. La création de liaisons hydrogène entre les différentes espèces du système est très importante dans l'abaissement des barrières énergétiques. Tout au long du chemin réactionnel, la molécule de DMSO se repositionne afin de stabiliser la séparation de charge.

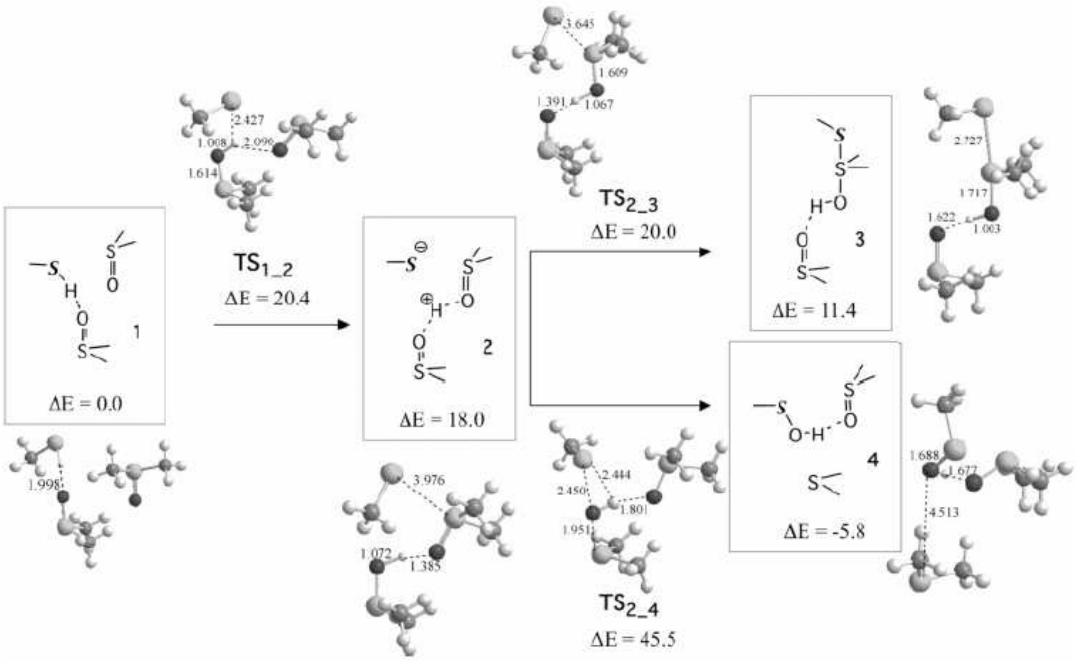


FIG. 5.11: *Chemin réactionnel de la réduction du DMSO par le méthanethiol : formation de l'intermédiaire sulfurane. Les géométries sont optimisées au niveau B3LYP/6-311G(d,p) et les énergies calculées au niveau MP2/6-311+G(3d2f,2df,2p). Le terme ΔE inclut l'énergie libre électrostatique de solvatation dans le continuum diélectrique, relativement au complexe **1**. Les valeurs données sont en Å et kcal/mol. Tiré de Balta et al.*[236]

Les sulfuranes[241–246] sont des intermédiaires réactionnels dans de nombreuses réactions d'oxydoréduction[242,243,246,247]. Ce sont des composés comportant un atome de soufre tétravalent au centre d'une bipyramide à base trigonale où les ligands les plus électronégatifs se situent aux positions axiales[241,248–250]. La détermination de leur structure électronique a donné lieu à des études théoriques[251–253]. En général, l'atome de soufre est supposé être hybridé sp^2. L'orbitale *p* restante et les orbitales des ligands axiaux se combinent pour former une orbitale moléculaire liante occupée, une non-liante et une anti-liante vacante. La géométrie et la structure électronique sont très sensibles à l'inclusion d'orbitales *d* dans les bases d'orbitales atomiques. Dans le cas du sulfurane **3**, les liaisons axiales s'allongent légèrement de 0,15 Å en l'absence d'orbitales *d*. Notons que le sulfurane **3** présente ses deux groupements les plus électronégatifs (SCH_3 et OH) en position axiale. Une structure ayant le groupement SCH_3 en position équatoriale a été caractérisée comme étant un minimum d'énergie à 21,2 kcal/mol au-dessus de **3**. En revanche, l'optimisation d'une structure ayant le groupement OH en position équatoriale conduit à la dissociation spontanée du système.

L'influence du solvant sur la géométrie du sulfurane **3** est remarquable. Entre la phase gazeuse et le solvant DMSO, la liaison S–S est passée de 2,410 à 2,727 Å et S–O de 1,905 à 1,717 Å,

préfigurant une possible dissociation du sulfurane entre les deux atomes de soufre. L'analyse des ordres de liaison en solution montre que S–S (0,22) est plus faible que S–O (1,24). Ainsi, pour les deux processus de dissociation suivants :

$$CH_3S-S(CH_3)_2-OH \quad \rightarrow \quad \left[CH_3S-S(CH_3)_2\right]^+ + OH^-$$

$$CH_3S-S(CH_3)_2-OH \quad \rightarrow \quad CH_3S^- + \left[S(CH_3)_2-OH\right]^+$$

on obtient une énergie libre de réaction en solution (continuum ; DMSO) de 16,9 kcal/mol pour la première et 2,7 kcal/mol pour la deuxième. Ces résultats sont à mettre en confrontation avec les mêmes processus de dissociation en milieu acide, *i.e.* sous l'influence d'un donneur de proton :

$$CH_3S-S(CH_3)_2-OH + H^+ \quad \rightarrow \quad \left[CH_3S-S(CH_3)_2\right]^+ + H_2O$$

$$CH_3S-S(CH_3)_2-OH + H^+ \quad \rightarrow \quad CH_3SH + \left[S(CH_3)_2-OH\right]^+$$

où les énergies libres de réaction sont de -21,3 et -14,1 kcal/mol respectivement, inversant donc la nature de la liaison préférablement dissociée. De ces résultats, il est donc possible de conclure que la stabilité du sulfurane **3** est très largement influencée par les conditions de pH. Un milieu acide ou la présence d'un donneur de proton favorise la rupture de la liaison S–O du sulfurane.

La dernière étape du mécanisme est la formation d'un acide sulfénique à partir de l'intermédiaire sulfurane. Cette étape est plus complexe à mettre en œuvre. Balta *et al.* proposent le mécanisme décrit dans la figure 5.12. Le complexe **5**, formé à partir du sulfurane **3** et d'une molécule de méthanethiol additionnelle, se dissocie pour former le complexe **6** où sont présents les ions sulfuranyle ($\left[CH_3S-S(CH_3)_2\right]^+$) **7** et thiolate ($CH_3S^-$) **8**. Une molécule d'eau, où l'atome d'oxygène est celui initialement porté par la fonction sulfoxyde, a été formée. L'énergie d'activation correspondant à cette dissociation vaut 13,3 kcal/mol.

Suite à cette dissociation, plusieurs produits de réaction ont été envisagés. Nous ne présenterons ici que le chemin réactionnel conduisant à la formation de l'acide sulfénique, illustré par la figure 5.13. Le complexe **6** mène à la formation de l'acide sulfénique **10** et du diméthylsulfure **12**, produit de la réduction du DMSO. Dans le même temps, la molécule de méthanethiol additionnelle est reprotonnée. Le calcul de l'IRC de l'état de transition $\mathbf{TS_{6-10}}$ vers les réactifs montre que cette transformation n'est pas directe. Elle nécessite la réorganisation du système réactif. La molécule d'eau, initialement présente au milieu des ions **7** et **8** et permettant d'assurer la séparation de charge dans le complexe **6**, doit migrer à proximité de l'atome de soufre conduisant à l'acide sulfénique. L'énergie nécessaire à cette transformation est de 0,9 kcal/mol.

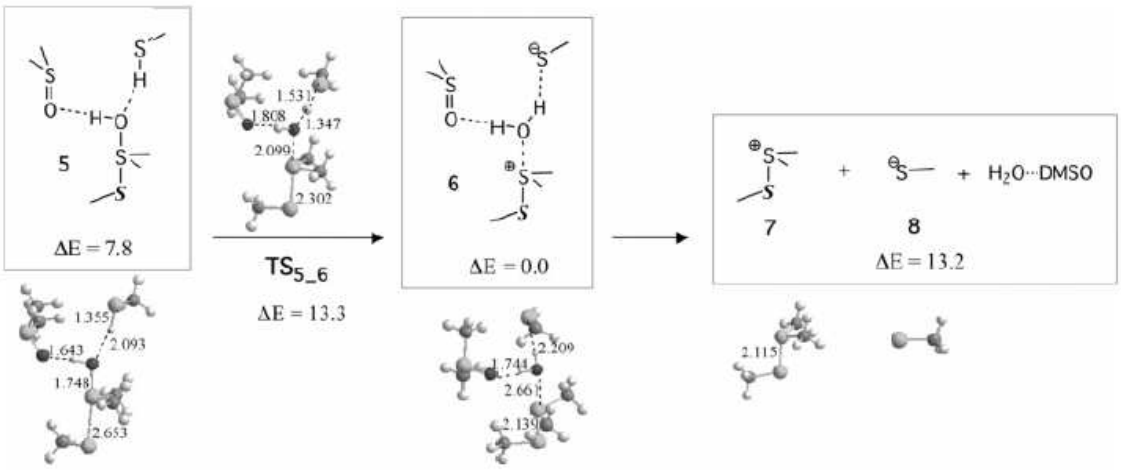


FIG. 5.12: *Chemin réactionnel de la réduction du DMSO par le méthanethiol : dissociation de l'intermédiaire sulfurane. Les géométries sont optimisées au niveau B3LYP/6-311G(d,p) et les énergies calculées au niveau MP2/6-311+G(3d2f,2df,2p). Le terme ΔE inclut l'énergie libre électrostatique de solvatation dans le continuum diélectrique, relativement au système* ***1*** *+ CH_3SH. Les valeurs données sont en Å et kcal/mol. Tiré de Balta et al.*[236]

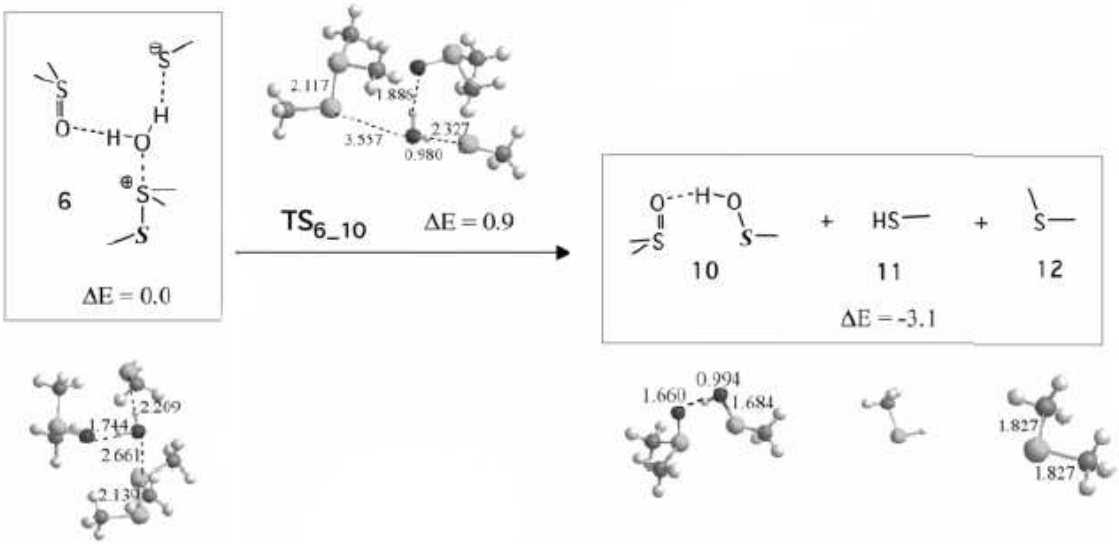


FIG. 5.13: *Chemin réactionnel de la réduction du DMSO par le méthanethiol : formation de l'acide sulfénique. Les géométries sont optimisées au niveau B3LYP/6-311G(d,p) et les énergies calculées au niveau MP2/6-311+G(3d2f,2df,2p). Le terme ΔE inclut l'énergie libre électrostatique de solvatation dans le continuum diélectrique, relativement au système* ***1*** *+ CH_3SH. Les valeurs données sont en Å et kcal/-mol. Tiré de Balta et al.*[236]

Bien que cette dernière étape du mécanisme de réduction ne semble pas limitée par des considérations énergétiques, sa validité dans le milieu enzymatique n'est pas évidente et fera l'objet d'une discussion détaillée dans notre étude.

Parallèlement à la dernière étape du mécanisme que nous venons d'évoquer, Balta *et al.* ont envisagé une autre voie de formation de l'acide sulfénique à partir de l'intermédiaire sulfurane **3**. Cette hypothèse propose que l'acide sulfénique soit formé directement à partir de l'intermédiaire sulfurane par un repliement de l'angle S–S–O en passant par un état de transition dont la géométrie est illustrée par la figure 5.14. Cette transformation pourrait éventuellement être catalysée par un acide afin d'aider à la dissociation de la liaison S–O. Une exploration de la surface d'énergie potentielle sur des structures proches de cet état de transition supposé montre que l'énergie d'activation serait de 41,6 kcal/mol au-dessus de l'intermédiaire sulfurane **3**. Cette valeur importante est due en grande partie à l'interaction très défavorable du doublet libre de l'atome de soufre central avec les deux autres atomes. Il semble difficilement concevable que la barrière énergétique soit abaissée de manière conséquente dans le site actif de l'enzyme. Cette hypothèse se retrouve donc, *a priori*, exclue du mécanisme de réduction.

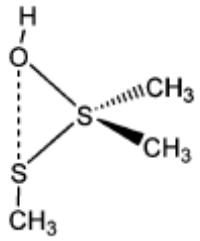


FIG. 5.14: *Chemin réactionnel de la réduction du DMSO par le méthanethiol : hypothèse de formation de l'acide sulfénique à partir de l'intermédiaire sulfurane par repliement de l'angle S–S–O. Tiré de Balta et al.*[236]

5.4. Objectifs et approche méthodologique

L'objectif principal de cette thèse est d'apporter des éléments nouveaux concernant le mécanisme de réduction enzymatique des sulfoxydes de méthionine par les méthionine sulfoxyde réductases de classe A. Nous nous sommes intéressé plus spécialement à l'étape réductase du mécanisme catalytique.

Dans un premier temps, nous étudierons la phase de reconnaissance et d'ancrage du substrat sulfoxyde dans le site actif de l'enzyme (docking). En principe, la méthode de docking développée dans la première partie de cette thèse aurait pu être utilisée. Cependant, la description des états hypervalents du soufre par les méthodes semi-empiriques est peu fiable avec les paramètrisations actuelles et nous avons préféré faire appel aux techniques de dynamique moléculaire. Une comparaison entre les résultats expérimentaux disponibles et les résultats théoriques permettra de valider la modélisation du complexe michaélien d'un sulfoxyde de méthionine dans le site actif de l'enzyme MsrA.

Nous nous intéresserons, dans une deuxième partie, à l'activation du substrat MetSO préalable à la réaction d'oxydoréduction. Sur la base des travaux de Balta *et al.*, cette activation pourrait avoir lieu par protonation du substrat sulfoxyde. Plusieurs outils théoriques seront utilisés dans le cadre de cette étude. La détermination du mécanisme de protonation du sulfoxyde sera effectuée dans une approche quantique de type DFT qui permettra d'évaluer le coût énergétique du processus. Des simulations de dynamique moléculaire seront réalisées ensuite pour estimer la stabilité de la forme activée du sulfoxyde dans le site actif.

Enfin, nous traiterons du mécanisme de réduction du substrat conduisant à la formation de l'intermédiaire acide sulfénique. Les méthodes issues de la chimie quantique seront utilisées à nouveau dans cette partie pour pouvoir rendre compte de la réactivité du système moléculaire. Après avoir étudié la réactivité sur un système modèle isolé, nous décrirons le processus dans un modèle du site actif de l'enzyme afin de prendre en compte les interactions avec l'environnement protéinique.

Chapitre 6

Reconnaissance et ancrage du substrat

Sommaire

6.1. Etat de protonation du site actif 118
6.2. Simulations, protocole opératoire 119
6.3. Enzyme libre 126
6.4. Complexe enzyme-substrat sulfoxyde 135
6.5. Conclusions 155

Le but de l'étude présentée dans ce chapitre est de modéliser la reconnaissance du substrat sulfoxyde par l'enzyme afin de déterminer la structure du complexe michaélien. La problématique abordée présente quelques difficultés qui sont liées principalement à l'état de protonation des résidus du site actif et à leur influence vis-à-vis de l'ancrage du substrat. Nous verrons que plusieurs états de protonation sont envisageables sur la base des travaux expérimentaux. Nous aborderons chacun de ces cas, à la fois, dans la modélisation de l'enzyme libre et dans celle du complexe michaélien.

6.1. Etat de protonation du site actif

Le site actif de MsrA présente une faible proportion de résidus possédant des propriétés acido-basiques. Les travaux d'Antoine *et al.*[234] ont montré que deux d'entre eux se révèlent être fortement impliqués dans le mécanisme catalytique : Cys51 (cystéine catalytique) et Glu94. La connaissance de leur état de protonation respectif, à la fois dans l'enzyme libre et le complexe enzyme-substrat, est donc importante pour comprendre le mécanisme de reconnaissance, puis la réactivité du système.

a. Enzyme libre

Par différentes titrations, Antoine *et al.*[234] ont obtenu une valeur de pK_{app} de 9,5 pour Cys51. La détermination expérimentale de la valeur de pK_{app} de Glu94 n'a pas été possible dans les mêmes conditions. Une prédiction des valeurs de pK_a a été réalisée dans le cadre de cette thèse en utilisant le programme PROPKA[9,10]. Les valeurs de pK_a prédites pour les résidus Cys51 et Glu94 sont regroupées dans le tableau 6.1 pour différentes structures PDB d'une MsrA d'*E. coli*. L'adduit S-(diméthylarséniate)-cystéine dans 1FF3 et le résidu sérine dans 3BQF, tous les deux à la position 51, ont été modifiés en un résidu cystéine pour le calcul. Les valeurs prédites pour Cys51 (entre 8,7 et 9,6) sont en accord avec la valeur de 9,5 obtenue expérimentalement. La valeur de pK_a prédite pour Glu94 est environ de 6,4, qui est supérieure à la valeur de référence 4,5.

Structure PDB	1FF3	3BQE	3BQF
Cys51	8,75	9,36	9,66
Glu94	6,48	6,41	5,75

TAB. 6.1: *Analyse PROPKA des résidus Cys51 et Glu94 de MsrA. Les trois structures correspondent à des MsrA d'E. coli. Les résidus S-(diméthylarsénic)Cys51 dans 1FF3 et Ser51 dans 3BQF ont été modifiés en Cys51.*

b. Complexe enzyme-substrat sulfoxyde

Durant l'étape réductase du mécanisme catalytique (*i.e.* lorsque l'enzyme est en interaction avec le substrat sulfoxyde), Antoine *et al.*[234], en prenant en compte la dépendance au pH de la constante cinétique, ont montré que la valeur de pK_a de Cys51 diminue drastiquement jusqu'à une valeur de pK_{app} de 5,7. La valeur de pK_a de Glu94 n'a pas pu être déterminée expérimentalement. L'analyse des structures cristallographiques de différentes MsrA[226–229] montre que la distance entre l'atome

d'oxygène mimant celui du sulfoxyde et l'atome d'oxygène de Glu94 est en moyenne de 2,6 Å. Ceci suggère qu'un atome d'hydrogène pourrait se situer entre ces deux atomes pour permettre la formation d'une liaison hydrogène.

c. Proposition d'état de protonation

D'après les résultats expérimentaux et les valeurs prédites par PROPKA, les états de protonation suivants sont envisageables pour décrire les résidus Cys51 et Glu94 du site actif :

- Cys51 : CysH dans l'enzyme libre et $Cys^{\ominus}$ dans le complexe enzyme-substrat ;
- Glu94 : $Glu^{\ominus}$ ou GluH dans l'enzyme libre ou dans le complexe enzyme-substrat.

Concernant le résidu glutamate, plusieurs hypothèses sont envisageables lors de l'étape de reconnaissance allant du système enzyme libre jusqu'au complexe enzyme-substrat (complexe michaélien) :

- Glu94 est déprotoné à la fois dans l'enzyme libre et le complexe enzyme-substrat. Dans ce cas, Cys51 transfère son proton au substrat (formation d'un sulfoxyde protoné) ;
- lors de la formation du complexe, Cys51 transfère son proton à Glu94. On a alors les couples CysH-$Glu^{\ominus}$ dans l'enzyme libre et $Cys^{\ominus}$-GluH dans le complexe michaélien ;
- Glu94 reste dans une forme protonée tout au long du processus de reconnaissance du substrat. On a, dans ce cas, GluH dans l'enzyme libre et dans le complexe enzyme-substrat.

Au total, trois états de protonation principaux pour le couple Cys51-Glu94 sont envisageables : CysH-$Glu^{\ominus}$, CysH-GluH et $Cys^{\ominus}$-GluH, à la fois pour l'enzyme libre et le complexe enzyme-substrat[a].

6.2. Simulations, protocole opératoire

Dans un premier temps, nous avons réalisé trois simulations de dynamique moléculaire de l'enzyme libre correspondant aux trois états de protonation mentionnés ci-dessus. Cette étude se justifie par le fait que la structure cristallographique utilisée n'est pas celle d'une enzyme native, *i.e.* l'enzyme a été modifiée (mutations éventuelles de résidus, formation d'adduit, *etc*) pour les besoins de la cristallisation et/ou de la diffraction de rayons X. Les structures tridimensionnelles disponibles ne comportant pas d'hydrogène, il est nécessaire d'envisager les trois états de protonation. Nous avons, ensuite, réalisé les simulations en présence du substrat sulfoxyde. Parmi les trois états de

[a]Un quatrième état, $Cys^{\ominus}$-$Glu^{\ominus}$ et sulfoxyde protoné, sera abordé au prochain chapitre.

protonation, nous souhaitons identifier celui (ou ceux) permettant de réaliser la reconnaissance et l'ancrage du substrat sulfoxyde. Les informations concernant ces six simulations sont résumées dans le tableau 6.2. Le substrat sulfoxyde que nous avons simulé est le diméthylsulfoxyde (DMSO). Nous n'avons pas simulé un peptide sulfoxyde de méthionine pour des raisons de simplicité. Les Msr sont connues pour réduire le DMSO, bien qu'elles aient une moins bonne affinité pour ce substrat[218]. Les simulations de dynamique moléculaire ont été réalisées avec les programmes SANDER (AMBER6[254]) et PMEMD (AMBER9[124]).

Nom de la simulation	Cys51	Glu94	Substrat
Free-CysH-Glu$^{\ominus}$	Protoné	Déprotoné	Non
Free-CysH-GluH	Protoné	Protoné	Non
Free-Cys$^{\ominus}$-GluH	Déprotoné	Protoné	Non
DMSO-CysH-Glu$^{\ominus}$	Protoné	Déprotoné	Oui
DMSO-CysH-GluH	Protoné	Protoné	Oui
DMSO-Cys$^{\ominus}$-GluH	Déprotoné	Protoné	Oui

TAB. 6.2: *Résumé des simulations réalisées pour l'étude de la reconnaissance par l'enzyme de son substrat. Les informations concernent l'état de protonation des résidus Cys51 et Glu94 et la présence ou non du substrat DMSO.*

a. Champ de forces

L'ensemble des atomes de l'enzyme et du solvant a été traité par l'intermédiaire du champ de forces *amber94*[255]. Pour la molécule de DMSO, plusieurs champs de forces sont disponibles dans la littérature[256–260]. Toutefois, les paramètres existants ont été ajustés afin de reproduire le comportement du DMSO en tant que solvant. Quelques uns d'entre eux ont été développés pour une utilisation avec l'algorithme SHAKE[261] ou avec une approche United-Atoms. Ces différents jeux de paramètres ne sont pas adaptés à la description de la molécule de DMSO dans notre étude. Ceci nous a conduit à développer nos propres paramètres selon le protocole de paramétrisation habituel lié à l'utilisation de la suite logicielle AMBER.

Ce protocole consiste à utiliser les paramètres du champ de forces *gaff*[262] (*General Amber Force Field*) pour les atomes de type inconnu. Les charges de ces atomes sont obtenues par le biais d'un calcul du potentiel électrostatique. Pour cela, la géométrie du composé est optimisée par un calcul quantique (programme GAUSSIAN03[263]) au niveau de théorie HF/6-31G*. Les charges issues du potentiel électrostatique sont ensuite recalculées par rapport à une contrainte donnée (programme RESP[264,265]). L'ajout d'une contrainte sur les charges atomiques permet, par exemple, de fixer la

valeur de la charge d'un atome ou d'obtenir la même valeur pour des atomes équivalents. Nous avons utilisé un script Perl nommé RED II[266] pour faciliter les traitements de données entre les programmes. La nomenclature utilisée pour désigner les atomes du substrat DMSO est indiquée dans la figure 6.1.

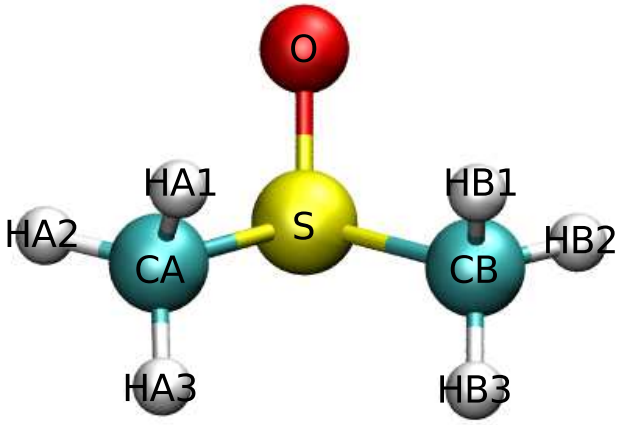


FIG. 6.1: *Nomenclature des atomes du substrat DMSO. Les atomes sont colorés en cyan pour le carbone, en blanc pour l'hydrogène, en bleu pour l'azote, en rouge pour l'oxygène et en jaune pour le soufre.*

Les paramètres de champ de forces du substrat DMSO sont regroupés dans le tableau 6.3. Nous avons utilisé le type s4 du champ de forces *gaff* pour décrire l'atome de soufre. Ce type représente un atome de soufre trivalent. Lors du calcul RESP, nous avons demandé que C_α et C_β, d'une part, et que $H_{\alpha1}$, $H_{\alpha2}$, $H_{\alpha3}$, $H_{\beta1}$, $H_{\beta2}$ et $H_{\beta3}$, d'autre part, portent la même charge. Les charges portées par les atomes de soufre et d'oxygène valent respectivement 0,27 et -0,47 u.a. Le moment dipolaire résultant de la distribution de charge vaut 4,59 D, un peu supérieur à la valeur de 4,00 D obtenue avec un calcul de haut niveau B3PW91/6-311++G(3df,3dp) en phase gazeuse[267].

b. Préparation des simulations

La structure de l'enzyme utilisée pour chacune des simulations est celle de MsrA d'*Escherichia coli* déterminée par cristallographie X par Tête-Favier *et al.* et déposée à la Protein Data Bank[2] sous le code 1FF3[226]. Cette structure a été obtenue à une résolution de 1,9 Å à partir d'un seul cristal à une température de 100 K. La cristallisation a été réalisée à pH 7 (tampon cacodylate 0,1 M) en présence de 0,35 M de sulfate d'ammonium et de 20% de PEG par diffusion de vapeur à 277 K. La cystéine catalytique en position 51 de cet échantillon a été modifiée en S-(dimethylarsénic)cystéine (CAS). Pour rappel, ce résidu permet de créer une espèce analogue au substrat sulfoxyde naturel

Atome	C∗	H∗	S	O
Type	CT	H1	s4	O
Masse (a.m.u.)	12,0	1,0	32,1	16,0
Charge (a.u.)	-0,2891	0,1301	0,2719	-0,4742
σ_{LJ} (Å)	1,9080	1,3870	2,000	1,6612
ε_{LJ} (kcal/mol)	0,1094	0,0157	0,2500	0,2100

Liaison	CT-H1	s4-CT	s4-O
Constante de force (kcal/mol/Å^2)	340,00	233,80	448,70
Distance d'équilibre (Å)	1,090	1,807	1,497

Angle	H1-CT-H1	s4-CT-H1	CT-s4-O	CT-s4-CT
Constante de force (kcal/mol/deg^2)	35,000	42,900	65,600	62,100
Angle d'équilibre (deg)	109,50	108,66	106,21	96,82

Angle dièdre	H1-CT-s4-O	H1-CT-s4-CT
Constante de force (kcal/mol)	0,200	0,200
Phase (rad)	0,00	0,00
Période	3	3

TAB. 6.3: *Paramètres du champ de force du substrat DMSO : types, charges, masses, paramètres de Lennard-Jones, potentiels de liaison, d'angle de valence et d'angle dièdre en fonction du type de l'atome. C∗ représente C_α et C_β. H∗ représente $H_{\alpha 1}$, $H_{\alpha 2}$, $H_{\alpha 3}$, $H_{\beta 1}$, $H_{\beta 2}$ et $H_{\beta 3}$.*

de la Msr. La structure est composée de 3 chaînes de 211 résidus. Certains résidus des chaînes B et C n'ont pu être localisés lors de la diffraction. Une superposition des 3 chaînes ne montre aucune différence au niveau structural. Nous choisissons la chaîne A pour effectuer les différentes simulations de dynamique moléculaire.

Avant le démarrage des simulations, nous enlevons de cette structure les molécules d'eau et les anions sulfates qui sont présents. Le résidu CAS est transformé en cystéine. Le cas échéant, nous positionnons le substrat dans le site actif de telle sorte qu'il se confonde avec le résidu CAS, *i.e.* l'oxygène et les groupements méthyles sont aux mêmes positions respectives. Nous identifions l'état de protonation des résidus ionisables tels que His, Lys, *etc* en effectuant une analyse PROPKA. Une analyse structurale des résidus ionisables en interaction permet de compléter cette identification. Le tableau 6.4 regroupe les résultats de ces analyses. L'ajout des hydrogènes est réalisé par le programme TLEAP du package AMBER en appliquant les états de protonation que nous avons déterminés. La position de ces hydrogènes est optimisée une première fois par minimisation dans le vide en utilisant le programme SANDER du package AMBER. Lors de cette étape, nous fixons la position de tous les atomes lourds pour conserver la structure globale de l'enzyme.

Résidus et Positions		Protoné	Type AMBER
Arg	18,57,83,89,97,125,135,150,155,167	Non	ARG
Asp	4,13,40,101,119,129,146,164,165,166,184,185	Non	ASP
Cys	86,198,206	Oui	CYS
Glu	43,56,84,108,116,144,154,172,183,210	Non	GLU
His	33,118,130	Oui en N_δ	HID
His	7,28,92,186	Oui en N_ε	HIE
His	168,191	Oui en N_δ et N_ε	HIP
Lys	5,6,192	Oui	LYS
Tyr	66,72,76,82,100,107,134,139,180,181,189,195,197	Oui	TYR

TAB. 6.4: *Etat de protonation des résidus ionisables de la MsrA (1FF3). On donne pour chaque résidu de l'enzyme l'état de protonation de sa chaîne latérale et le type* AMBER *associé au résidu, à partir d'une analyse PROPKA complétée par une analyse des résidus en interaction.*

c. Mise en solution et minimisation

La mise en solution est réalisée à la suite de la première minimisation avec le programme TLEAP. Nous ajoutons, dans un premier temps, des cations sodium pour assurer la neutralité de l'ensemble. Puis, nous ajoutons des molécules d'eau (TIP3P[268]) à cette structure afin de créer au final une boîte de simulation orthorhombique contenant l'enzyme entourée par 10 Å de solvant dans chaque direction. La dimension de cette boîte est $80 \times 70 \times 80$ Å et elle contient environ 35 000 atomes dont environ 10 000 molécules d'eau. La quantité de solvant ajoutée dans la boîte permet au système simulé d'être proche de la concentration expérimentale en Msr ($\simeq$ 0,1 mmol/L).

L'optimisation de la position des molécules de solvant se fait par minimisations successives. A chaque fois, nous réalisons une minimisation en appliquant les conditions limites périodiques à la boîte de simulation (coupure des interactions à 12 Å). La succession de minimisations a pour but d'installer le solvant autour de l'enzyme sans endommager sa structure. Les atomes lourds ont leurs positions fixes et nous relachons les contraintes petit à petit selon l'ordre :

- hydrogènes de l'enzyme + hydrogènes du solvant ;
- hydrogènes de l'enzyme + hydrogènes et oxygènes du solvant ;
- tous les atomes libres, mais en mettant une contrainte de 10 kcal/mol/Å^2 par rapport à la structure cristallographique (sous la forme d'un potentiel harmonique) sur les atomes lourds de l'enzyme et du substrat.

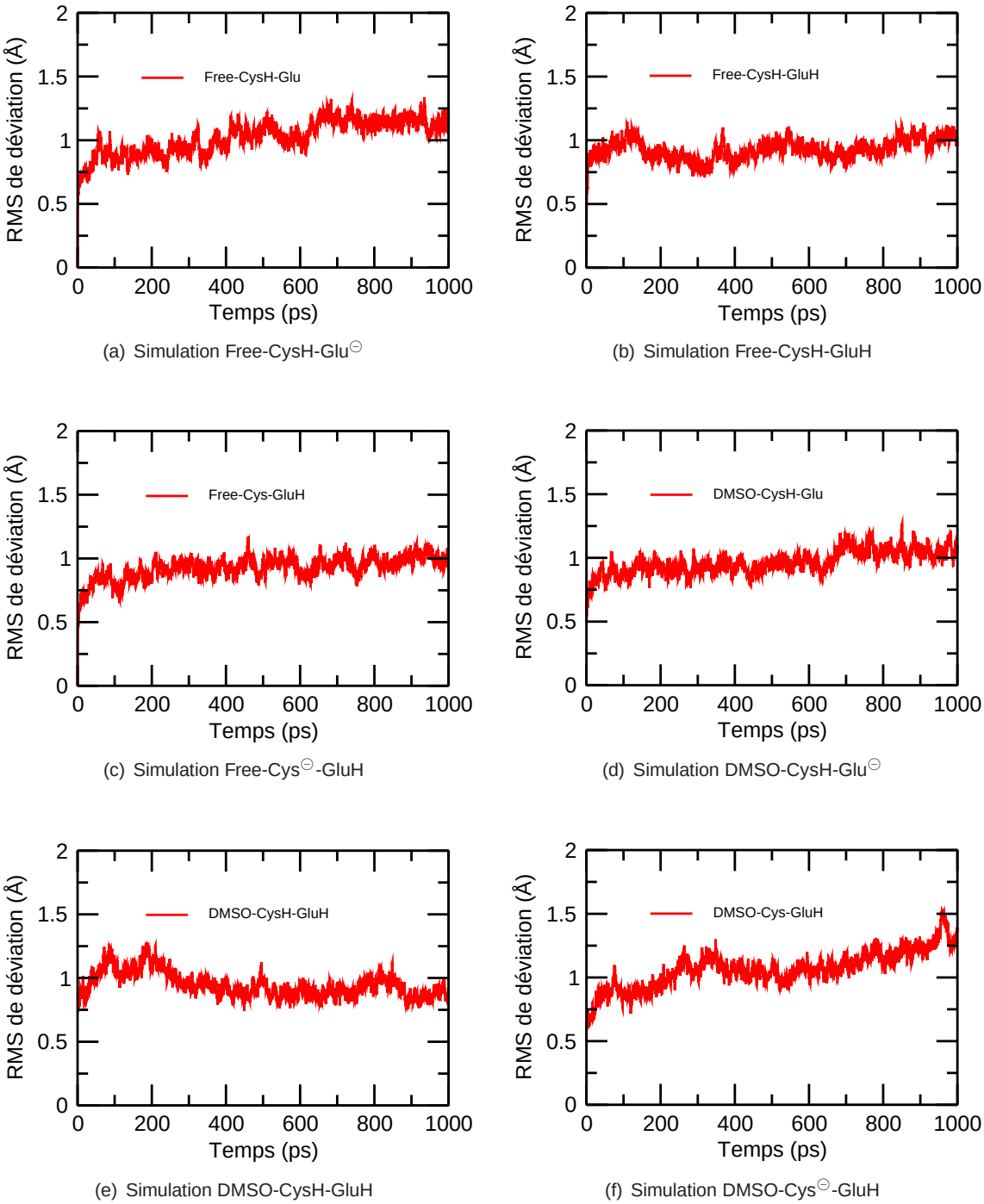
d. Equilibration thermodynamique

Chaque système étudié a fait l'objet d'une phase d'équilibration thermodynamique dans l'ensemble NPT avec l'utilisation des conditions limites périodiques. Comme pour la minimisation lors de la mise en solution, le protocole d'équilibration a été conçu afin de ne pas endommager la structure de l'enzyme et la même contrainte est appliquée aux atomes lourds de l'enzyme (potentiel harmonique de 10 kcal/mol/Å^2 par rapport à la structure cristallographique). Nous avons effectué une montée en température de 10 K à 300 K par paliers de 50 K tous les 10 ps. Le système est ensuite équilibré à 300 K. La pression est maintenue fixe à 1 atm. Nous avons produit ainsi un total de 200 000 pas de dynamique avec un pas d'intégration de 0,5 fs (soit 100 ps). A la fin de cette étape, nous vérifions que les grandeurs thermodynamiques et physiques des différents systèmes sont voisines des valeurs cibles. Nous obtenons, par exemple, une densité d'environ 1,05 pour chaque simulation.

e. Production des trajectoires

Pour chacun des systèmes, nous avons produit une trajectoire de 1 ns. Le pas d'intégration utilisé est de 0,5 fs. La température est fixée à 300 K et la pression à 1 atm. Nous sauvegardons les coordonnées et les vitesses tous les 200 pas, soit toutes les 0,1 ps. Nous avons calculé le RMS atomique de déviation de la chaîne principale de l'enzyme par rapport à la structure cristallographique pour chaque pas sauvegardé. La valeur du RMS est proche de 1 Å pour chaque simulation, comme le montre la figure 6.2. Cette valeur indique une déformation relativement faible de la structure de l'enzyme. La visualisation de chaque trajectoire ne montre pas de modification notable de la structure globale de l'enzyme.

Toutes ces trajectoires ont été réalisées en utilisant l'algorithme Particule-Mesh Ewald[269,270] (PME) pour le traitement des interactions électrostatiques à longue portée. La boîte de simulation de chaque système est ainsi divisée en 96 fragments selon chaque direction, *i.e.* on obtient un découpage en $96 \times 96 \times 96$ sous-boîtes. Ceci permet de convenir à la règle d'un atome par sous-boîte pour une utilisation optimale de la méthode.



(a) Simulation Free-CysH-Glu$^{\ominus}$

(b) Simulation Free-CysH-GluH

(c) Simulation Free-Cys$^{\ominus}$-GluH

(d) Simulation DMSO-CysH-Glu$^{\ominus}$

(e) Simulation DMSO-CysH-GluH

(f) Simulation DMSO-Cys$^{\ominus}$-GluH

FIG. 6.2: *Variation du RMS atomique de déviation de la chaîne principale (atomes lourds) de l'enzyme en fonction du temps de simulation, pour les six systèmes simulés.*

6.3. Enzyme libre

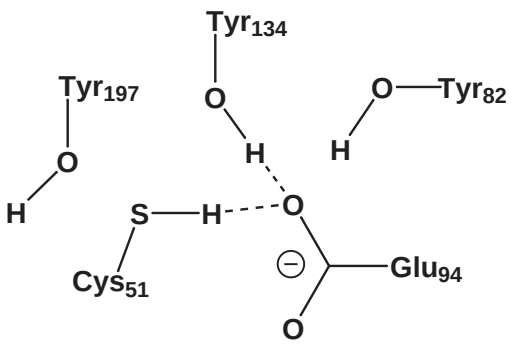
6.3.1. Etat de protonation CysH-Glu$^{\ominus}$

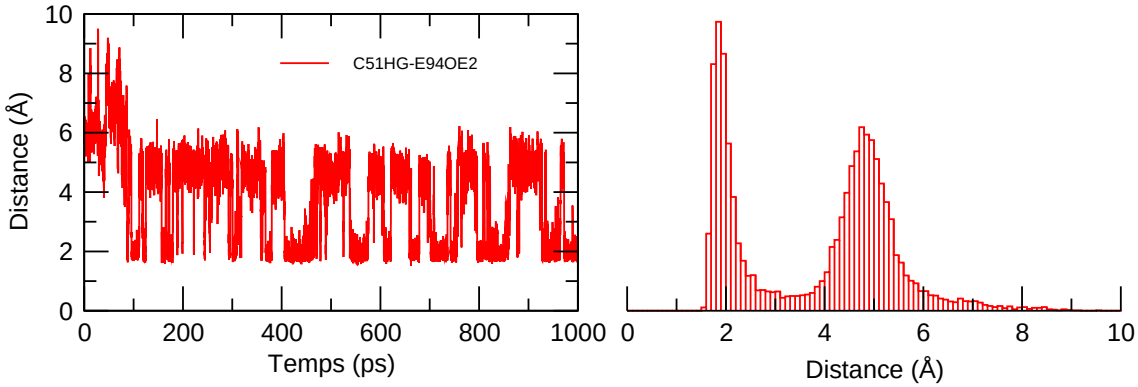
Quelques résultats de la simulation sont présentés dans la figure 6.3. Nous observons une interaction relativement forte entre l'hydrogène H_γ de Cys51 et l'oxygène $O_{\varepsilon 2}$ de Glu94 (Fig. 6.3(a)). L'histogramme montre deux maxima de probabilité autour de 2 Å et 5 Å. En fait, on observe des changements fréquents entre ces deux distances au cours de la simulation. Par ailleurs, la figure 6.3(b) montre que la distance Cys51-H_γ – Tyr197-O_η oscille autour de 4 Å et de 6 Å. Ces interactions s'expliquent par une variation de l'angle dièdre de Cys51 autour de la liaison $C_\beta - S_\gamma$ (Fig. 6.3(c)). Lorsque cet angle est autour de 60°, il y a une liaison hydrogène forte entre CysH et Glu$^{\ominus}$. Lorsque cet angle augmente à 180°, cette liaison est rompue et Cys51 interagit avec Tyr197 par l'intermédiaire d'une molécule d'eau.

La simulation montre aussi une interaction importante entre Tyr134 et Glu94. Les distances Glu94-$O_{\varepsilon 2}$ – Tyr134-H_η et Glu94-$O_{\varepsilon 1}$ – Tyr134-H_η valent entre 2 et 3 Å sur toute la simulation à partir de 150 ps (Fig. 6.3(d) et 6.3(e)). Auparavant, on remarque une rotation libre de l'extrémité de la chaîne latérale de Glu94 autour de la liaison $C_\gamma - C_\delta$ (Fig. 6.3(f)). Nous n'avons pas observé ce type d'interaction entre Tyr82 et Glu94 (Fig. 6.3(d) et 6.3(e)).

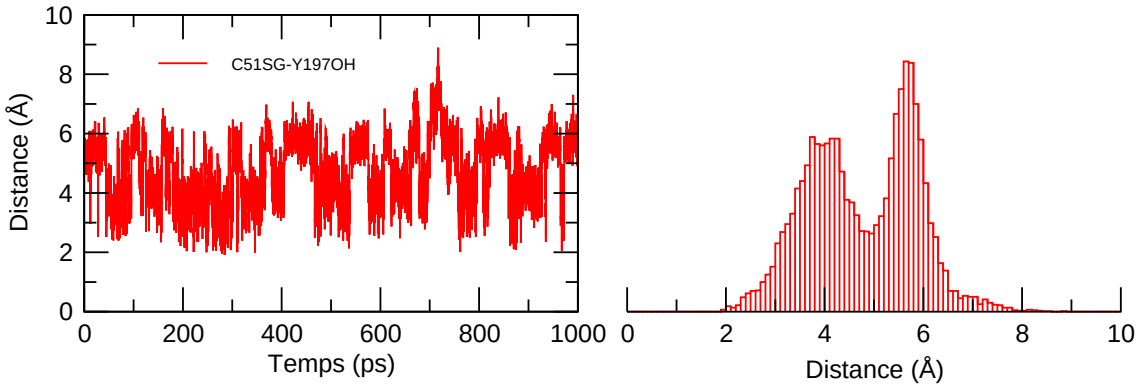
La solvatation des résidus du site actif a été analysée. Les résultats montrent que le résidu Cys51 est solvaté par 4 à 5 molécules de solvant (molécules à une distance inférieure à 3,4 Å) et Glu94 par 2 à 3 molécules (résultats non montrés).

D'après cette simulation donc, un état de protonation CysH-Glu$^{\ominus}$ pour l'enzyme libre conduirait à une interaction stabilisante entre l'hydrogène H_γ porté par Cys51 catalytique et l'oxygène $O_{\varepsilon 2}$ de Glu94 et l'existence de liaisons hydrogène entre les oxygènes de Glu94 et l'hydrogène de Tyr134. Un schéma des interactions entre les résidus du site actif est illustré dans la figure ci-dessous.

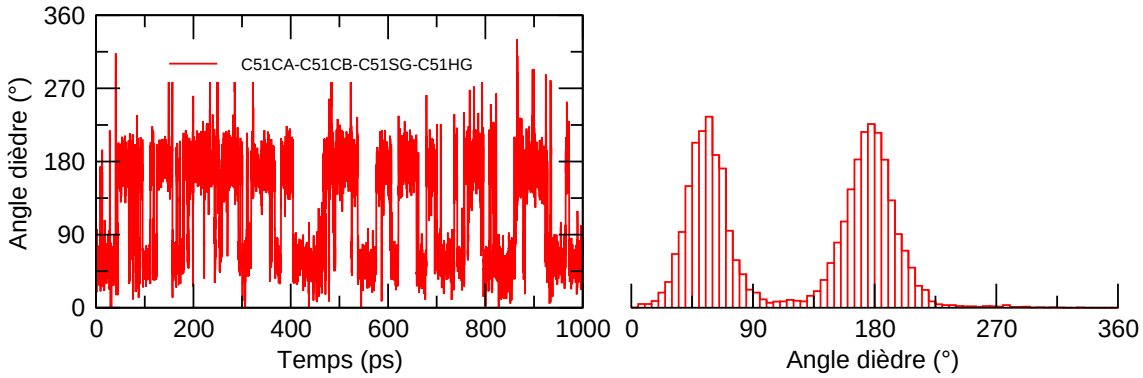




(a) Distance Cys51-H_γ – Glu94-$O_{\varepsilon 2}$

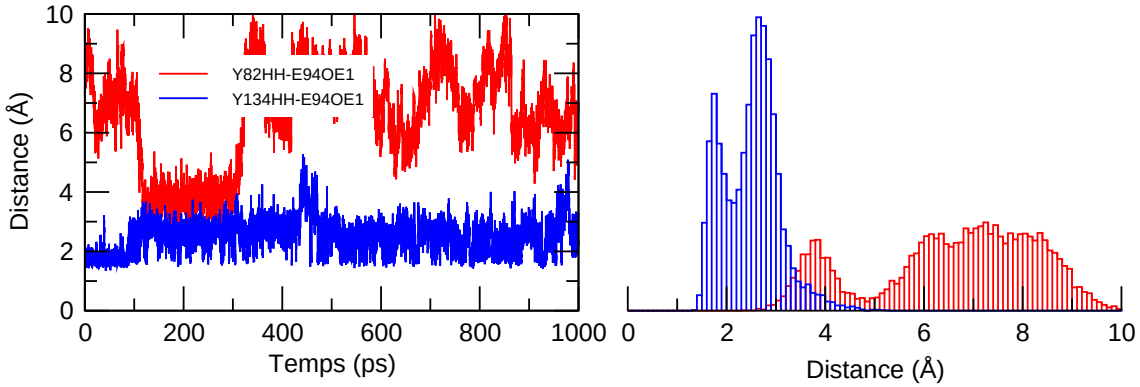


(b) Distance Cys51-H_γ – Tyr197-O_η

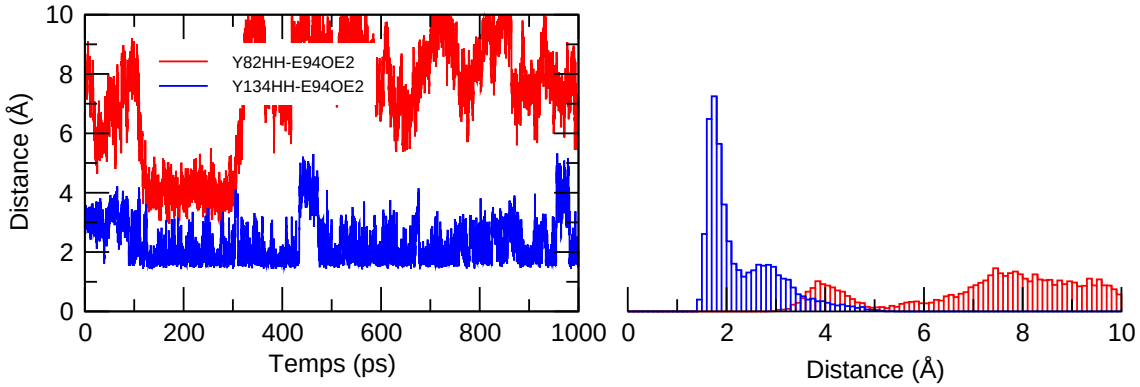


(c) Angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ

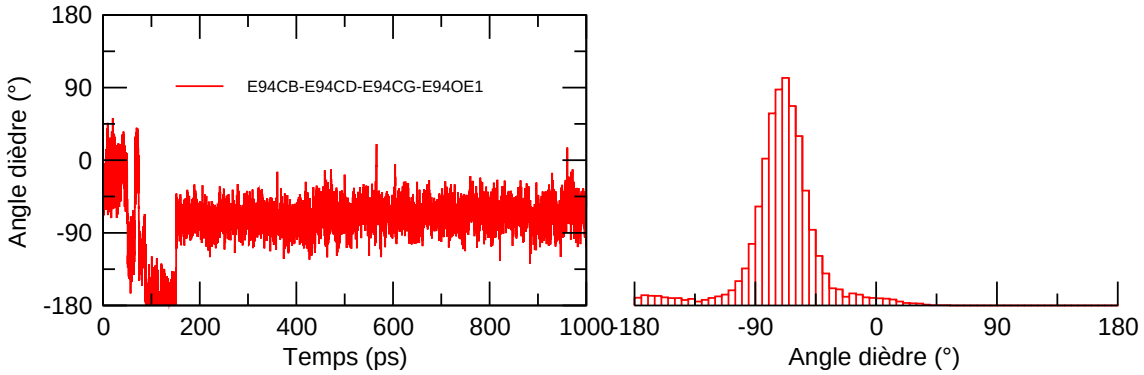
FIG. 6.3: *Variation de la distance Cys51-H_γ – Glu94-$O_{\varepsilon 2}$ (a), de la distance Cys51-H_γ – Tyr197-O_η (b) et de l'angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ (c) en fonction du temps dans la simulation Free-CysH-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(d) Distance Glu94-$O_{\varepsilon 1}$ – Tyr82-H_{η} (rouge) et Glu94-$O_{\varepsilon 1}$ – Tyr134-H_{η} (bleu)



(e) Distance Glu94-$O_{\varepsilon 2}$ – Tyr82-H_{η} (rouge) et Glu94-$O_{\varepsilon 2}$ – Tyr134-H_{η} (bleu)



(f) Angle dièdre Glu94-C_{β} – Glu94-C_{γ} – Glu94-C_{δ} – Glu94-$O_{\varepsilon 1}$

FIG. 6.3: *(suite) Variation de la distance Glu94-$O_{\varepsilon 1}$ – Tyr82-H_{η} et Glu94-$O_{\varepsilon 1}$ – Tyr134-H_{η} (d), de la distance Glu94-$O_{\varepsilon 2}$ – Tyr82-H_{η} et Glu94-$O_{\varepsilon 2}$ – Tyr134-H_{η} (e) et de l'angle dièdre Glu94-C_{β} – Glu94-C_{γ} – Glu94-C_{δ} – Glu94-$O_{\varepsilon 1}$ (f) en fonction du temps dans la simulation de Free-CysH-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*

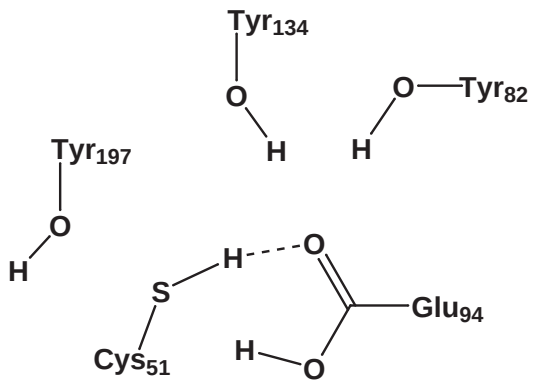
6.3.2. Etat de protonation CysH-GluH

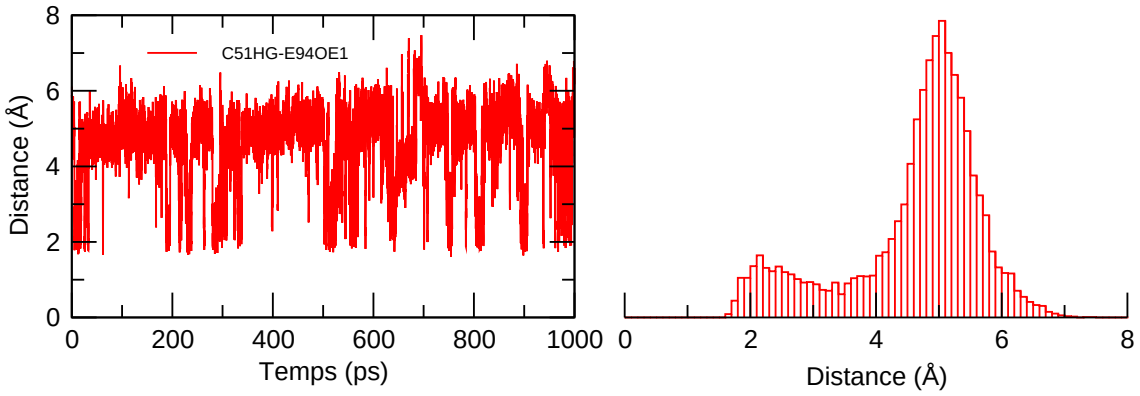
Les résultats pour cet état de protonation sont représentés dans la figure 6.4. Le proton du résidu Glu94 est porté par l'oxygène $O_{\varepsilon 2}$. Au cours de la simulation et pendant des courtes durées, l'hydrogène H_γ du résidu Cys51 est impliqué dans une liaison hydrogène soit avec Glu94 (Fig. 6.4(a)) soit avec Tyr197 (Fig. 6.4(c)), mais l'analyse détaillée de la trajectoire montre que la plupart du temps ce proton est solvaté par une molécule d'eau.

Nous n'avons pas observé de liaison hydrogène notable entre Cys51-S_γ et Glu94-$H_{\varepsilon 2}$, la distance reste supérieure à 4 Å tout au long de la trajectoire (Fig. 6.4(b)). Cet atome d'hydrogène forme une liaison hydrogène avec une molécule d'eau du solvant pendant toute la durée de la simulation.

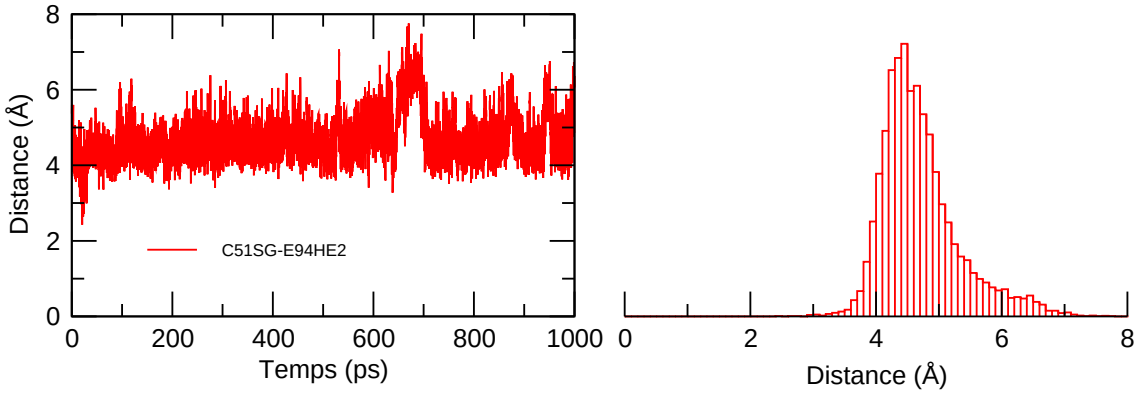
L'analyse des couches de solvatation montre que le résidu Cys51 est globalement bien solvaté par des molécules d'eau (Fig. 6.4(d)). On compte en moyenne 3 à 5 molécules de solvant à une distance inférieure à 3,4 Å. Le résidu Glu94, plus profondément enfoui dans le site actif, est moins solvaté avec 1 molécule d'eau à proximité (distance < 3,4 Å, Fig. 6.4(e)). La neutralité du résidu Glu94 peut expliquer la légère diminution de solvatation constatée par rapport à la simulation précédente. Par ailleurs, il n'existe plus de liaison hydrogène entre celui-ci et le résidu Tyr134, ni avec Tyr82 (Fig. 6.4(f)).

Dans cet état de protonation donc, une liaison hydrogène faible se forme entre l'hydrogène H_γ de Cys51 et l'oxygène $O_{\varepsilon 1}$ de Glu94. En comparaison avec l'état de protonation précédent, les interactions sont globalement moins fortes, ce qui n'est pas étonnant compte tenu de la neutralité du résidu Glu94. La solvatation relativement faible de ce résidu suggère que CysH-GluH peut être un état de protonation envisageable pour l'enzyme libre. La figure ci-dessous présente un résumé des interactions observées dans cette simulation.

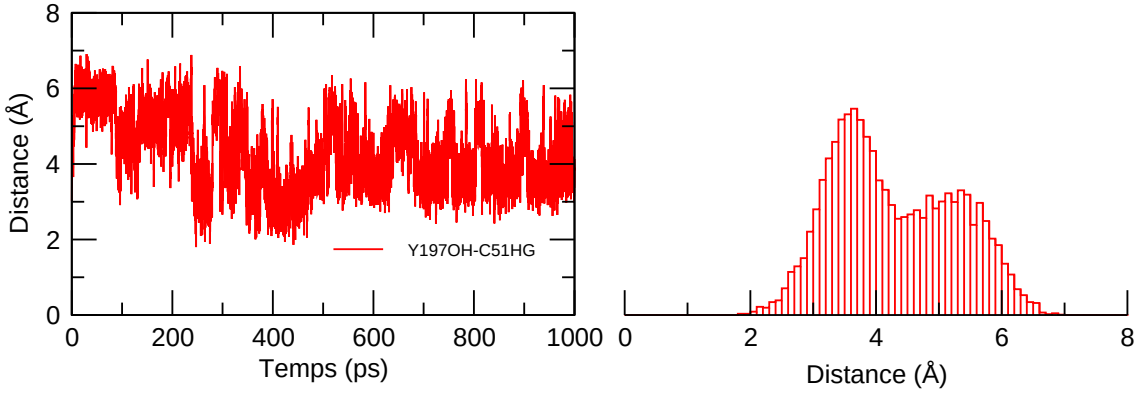




(a) Distance Cys51-H_γ – Glu94-$O_{\varepsilon 1}$

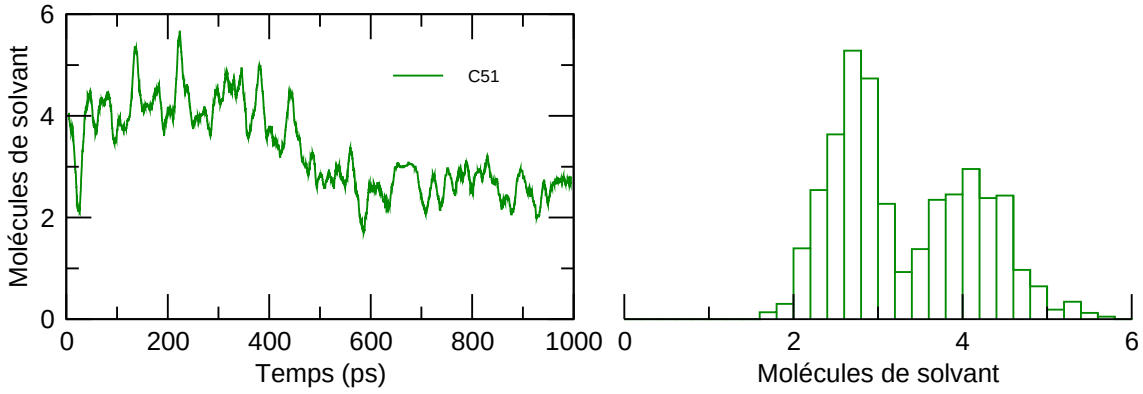


(b) Distance Cys51-S_γ – Glu94-$H_{\varepsilon 2}$

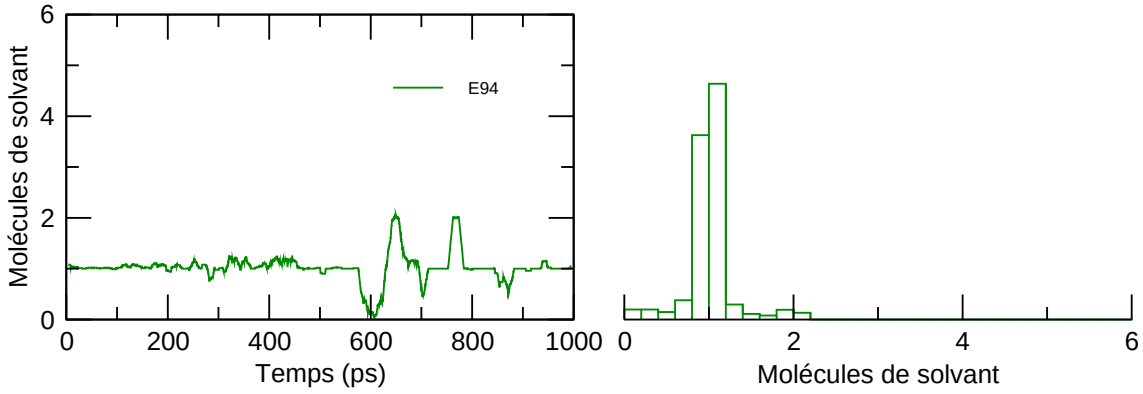


(c) Distance Cys51-H_γ – Tyr197-O_η

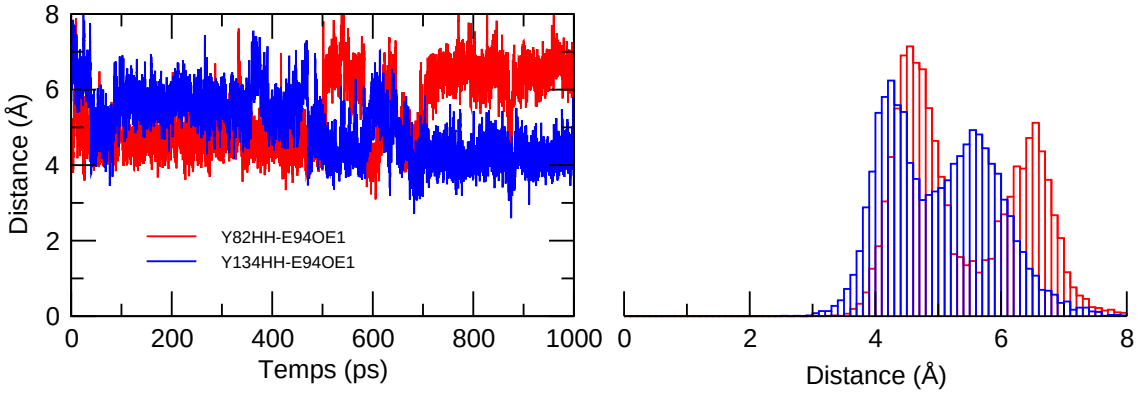
FIG. 6.4: *Variation de la distance Cys51-H_γ – Glu94-$O_{\varepsilon 1}$ (a), de la distance Cys51-S_γ – Glu94-$H_{\varepsilon 2}$ (b) et de la distance Cys51-H_γ – Tyr197-O_η (c) en fonction du temps dans la simulation Free-CysH-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(d) Nombre de molécules d'eau à proximité du résidu Cys51 (distance < 3,4 Å)



(e) Nombre de molécules d'eau à proximité du résidu Glu94 (distance < 3,4 Å)



(f) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (bleu)

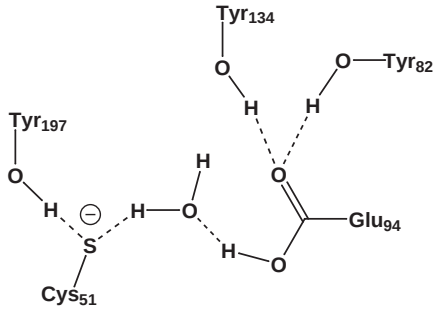
FIG. 6.4: *(suite) Variation du nombre de molécules d'eau dans la solvatation des résidus Cys51 (d) et Glu94 (e) et des distances Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (f) en fonction du temps dans la simulation de Free-CysH-GluH. La solvatation est moyennée sur 10 ps. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*

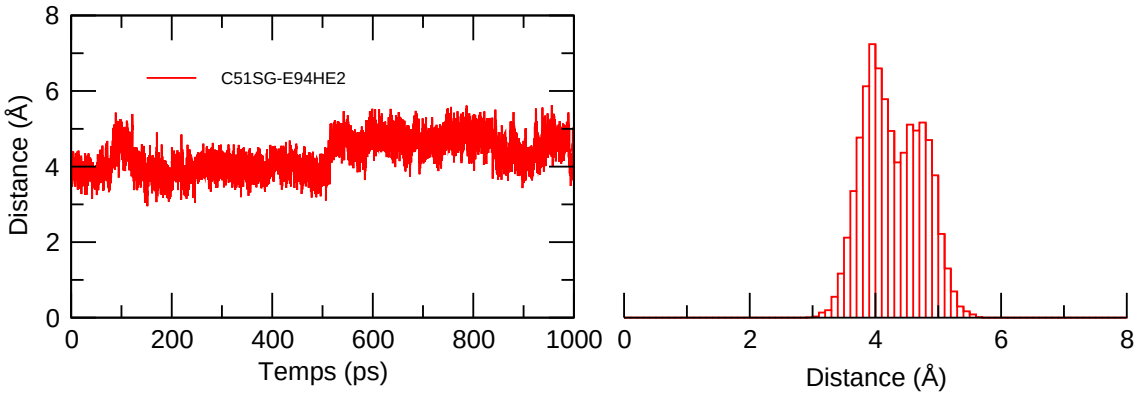
6.3.3. Etat de protonation Cys$^{\ominus}$-GluH

Les résultats de la simulation sont présentés dans la figure 6.5. Les interactions entre résidus sont assez peu nombreuses dans cette simulation. En effet, les résidus du site actif sont, pour la plupart, en interaction avec des molécules d'eau du solvant. Ainsi, par exemple, nous n'observons pas d'interaction entre l'hydrogène $H_{\varepsilon 2}$ de Glu94 et le soufre S_{γ} de Cys51. La distance entre ces deux atomes présente un maximum de probabilité à 4 Å (Fig. 6.5(a)), une molécule d'eau étant intercalée. A partir de t = 600 ps, le soufre S_{γ} de Cys51 entre en interaction avec l'hydrogène H_{η} de Tyr197, formant une liaison hydrogène avec une distance autour de 2 Å (Fig. 6.5(b)). Une molécule de solvant est initialement présente entre ces deux résidus. L'atome Cys51-S_{γ} est peu mobile, comme le montre la variation de l'angle dièdre H_{α}–C_{α}–C_{β}–S_{γ} du résidu (Fig. 6.5(c)).

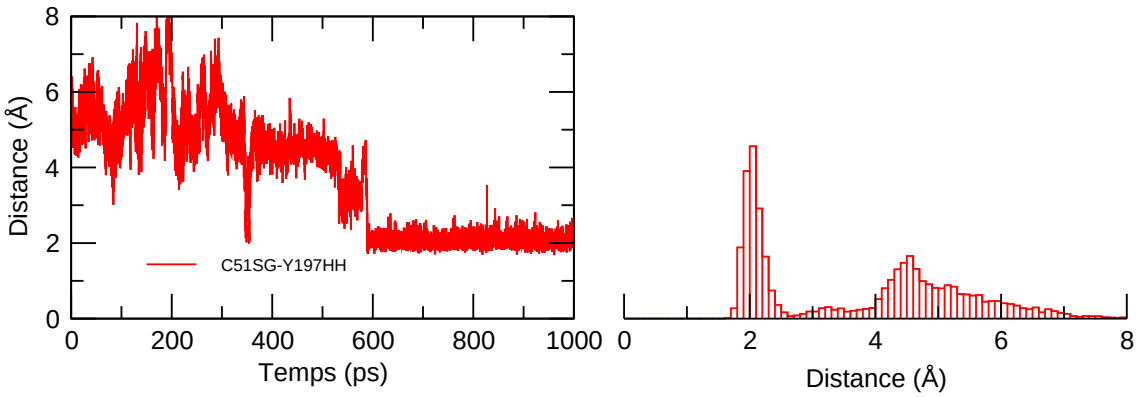
Nous constatons aussi que l'extrémité de la chaîne latérale de Glu94 est peu mobile (Fig. 6.5(d)). Les atomes d'hydrogène H_{η} de Tyr82 et de Tyr134 forment une liaison hydrogène assez faible avec Glu94-$O_{\varepsilon 1}$ (Fig. 6.5(e) et 6.5(f)). La solvatation de ce résidu est toujours relativement faible avec en moyenne 1 à 2 molécules d'eau à proximité (résultats non montrés).

Dans cette simulation, on observe une rigidification globale du système : Cys51 forme une liaison hydrogène stable avec Tyr197 (Fig. 6.5(b)) tandis que Glu94 reste en interaction quasi fixe (Fig. 6.5(d)) avec une molécule d'eau intercalée entre Cys51 et Glu94. Ainsi, la protonation de Glu94 semble adéquate à l'intérieur du site actif car elle stabilise le mouvement de celui-ci sans augmenter sa solvatation. L'augmentation de pK_a de Glu94 prédite par PROPKA, par rapport à sa valeur en solution, est ici confirmée par la désolvatation globale du résidu observée tout au long de la simulation. Au contraire, dans cette simulation, Cys51 tend à maximiser ses interactions avec des molécules d'eau ou avec Tyr197. Sa solvatation augmente. Cela indique donc que Cys51 préfère, dans le site actif libre de l'enzyme, être neutre plutôt que déprotoné. Ceci est en accord avec les résultats expérimentaux d'Antoine *et al.* assignant à Cys51 un pK_a de 9,5. Les interactions entre les résidus du site actif sont représentés dans le schéma ci-dessous.

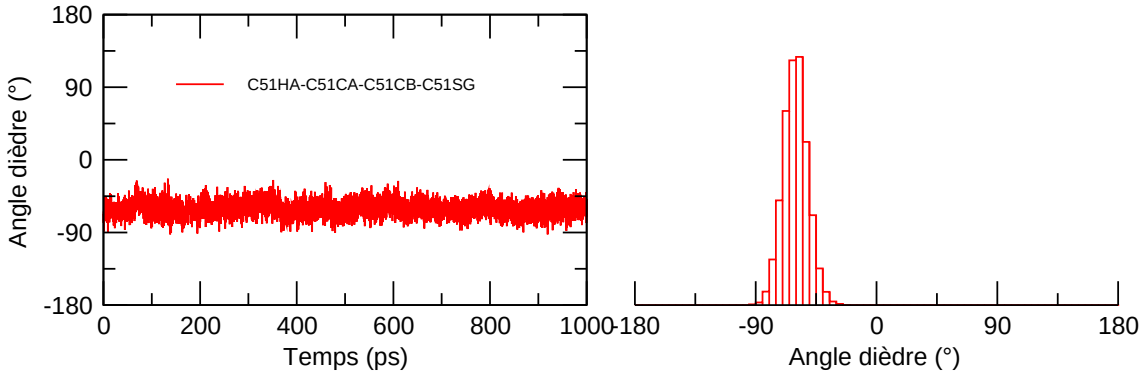




(a) Distance Cys51-S_γ – Glu94-$H_{\varepsilon 2}$

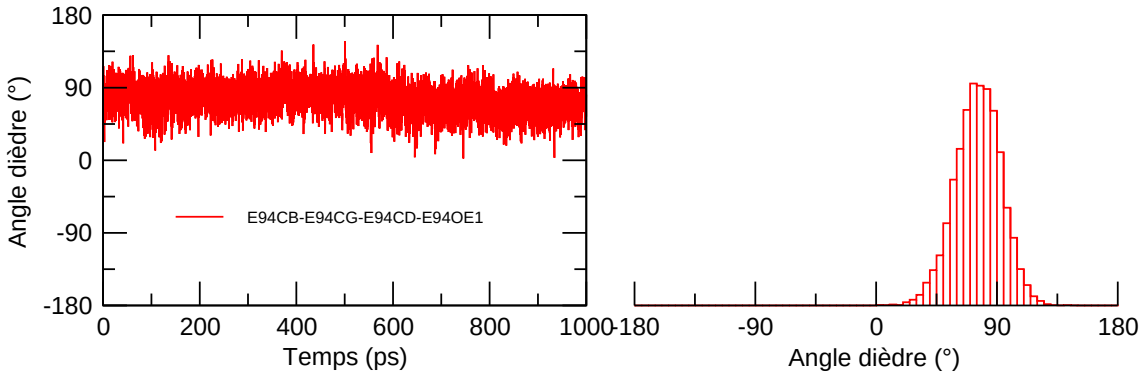


(b) Distance Cys51-S_γ – Tyr197-H_η

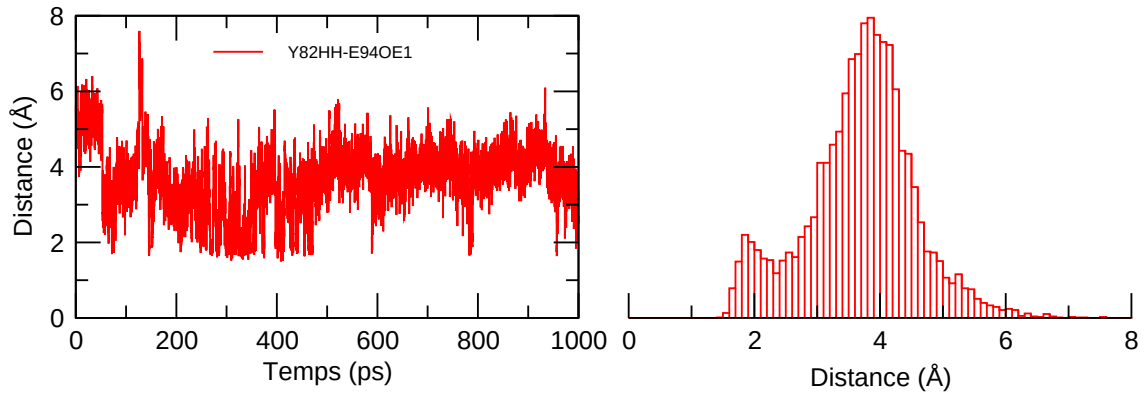


(c) Angle dièdre Cys51-H_α – Cys51-C_α – Cys51-C_β – Cys51-S_γ

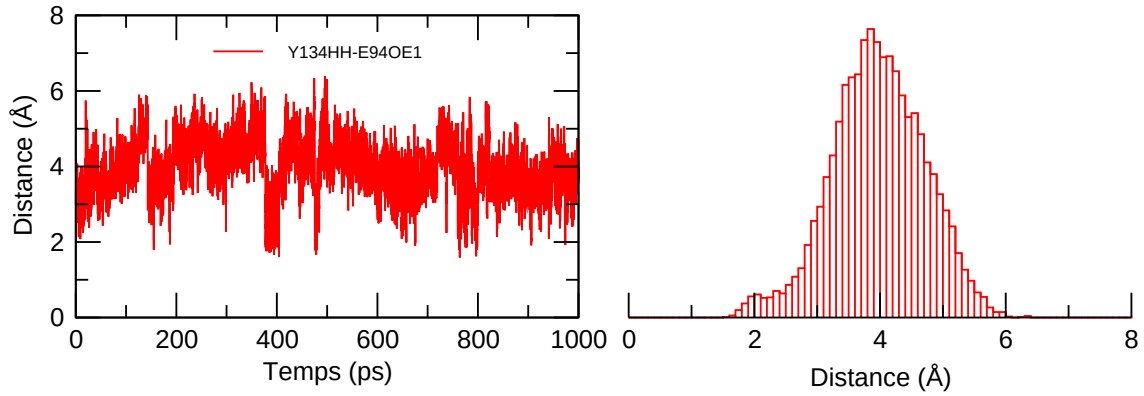
FIG. 6.5: *Variation de la distance Cys51-S_γ – Glu94-$H_{\varepsilon 2}$ (a), de la distance Cys51-S_γ – Tyr197-H_η (b) et de l'angle dièdre Cys51-H_α – Cys51-C_α – Cys51-C_β – Cys51-S_γ (c) en fonction du temps dans la simulation de Free-Cys$^\ominus$-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(d) Angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon 1}$



(e) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 1}$



(f) Distance Tyr134-H_η – Glu94$O_{\varepsilon 1}$

FIG. 6.5: *(suite) Variation de l'angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon 1}$ (d), de la distance Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ (e) et de la distance Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (f) en fonction du temps dans la simulation Free-Cys$^{\ominus}$-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*

6.4. Complexe enzyme-substrat sulfoxyde

6.4.1. Etat de protonation CysH-Glu$^{\ominus}$

Dans une première trajectoire, nous avons constaté que le substrat DMSO est sorti du site actif au bout d'environ 200 ps pour se retrouver dans le solvant. Une deuxième trajectoire a été simulée après avoir prolongé la phase initiale d'équilibration de 100 ps supplémentaires (voir page 124) en appliquant une contrainte entre les atomes de soufre Cys51-S_γ et DMSO-S sous la forme d'un potentiel harmonique de 10 kcal/mol/Å^2 à 3,5 Å, pour maintenir le substrat dans le site actif. A la suite, nous avons effectué notre simulation de 1 ns. Ci-dessous, nous présentons l'analyse de cette seconde simulation qui n'a pas permis de maintenir davantage le sulfoxyde au sein du site actif. Les résultats sont présentés dans la figure 6.6.

La visualisation de la trajectoire montre, au début de la simulation, que le sulfoxyde est entouré d'une grande quantité d'eau avec 10 molécules à proximité (distance < 3,4 Å) (Fig. 6.6(j)). Ces molécules forment des liaisons hydrogène avec les résidus du site actif et permettent au substrat de demeurer à l'intérieur pendant les 200 premières picosecondes de la simulation. Ceci est visible sur le tracé des distances Tyr82-H_η – DMSO-O (Fig. 6.6(a)), Tyr134-H_η – DMSO-O (Fig. 6.6(b)) et Cys51-H_γ – DMSO-O (Fig. 6.6(c)) où nous observons un palier de 0 à 200 ps avec une distance d'environ 6 Å dans tous les cas. Le nombre de molécules de solvant va croissant dans la suite de la simulation. Au delà de 200 ps, l'absence d'interactions spécifiques du substrat avec les résidus du site actif entraîne son départ vers le solvant.

Au début de la simulation, il y a une interaction stabilisante entre le soufre du DMSO et le soufre S_γ de Cys51. Ces deux atomes, de charges opposées, sont séparés d'une distance moyenne de 3,5 Å durant les 100 premières picosecondes (Fig. 6.6(d)). Il s'agit d'une interaction spécifique entre l'enzyme et le substrat mais qui ne permet pas de fixer le DMSO dans le site actif. Rappelons qu'un élément de reconnaissance du substrat, décrit par les travaux de Gand *et al.*[232], est l'interaction de type hydrophobe d'un groupement méthyle du substrat avec la poche hydrophobe formée par les résidus Phe52 et Trp53. Nous n'observons pas ce type d'interaction dans notre simulation (Fig. 6.6(e)). La distance Trp53-$N_{\varepsilon 1}$ – DMSO-C_β est supérieure à 5 Å ; clairement, le site de reconnaissance hydrophobe n'est pas comblé.

Nous considérons à présent les interactions entre les résidus du site actif. Durant la première moitié de la simulation, nous observons une liaison hydrogène entre H_γ de Cys51 et $O_{\varepsilon 2}$ de Glu94 (Fig. 6.6(g)) avec une distance de 2 Å présentant des fortes fluctuations. Pendant la même période,

les oxygènes $O_{\varepsilon 1}$ et $O_{\varepsilon 2}$ de Glu94 forment des liaisons hydrogène fortes avec les hydrogènes H_{η} des résidus Tyr82 et Tyr134 (Fig. 6.6(h) et 6.6(i)), permettant la stabilisation de la charge négative portée par Glu94. Après 500 ps de simulation, toutes ces interactions deviennent plus faibles ou sont rompues pour certaines. L'interaction entre Cys51 et Glu94 est plus ponctuelle et moins longue en durée. Nous retrouvons donc globalement les interactions que nous avons décrites dans le cas de l'enzyme libre pour ce même état de protonation (Free-CysH-Glu$^{\ominus}$). A partir de 500 ps, le substrat se situe dans le solvant et donc il n'a plus d'influence sur les résidus du site actif. Nous vérifions ce changement de comportement dans l'analyse de l'angle dièdre de Cys51 autour de la liaison C_{β}–S_{γ} (Fig. 6.6(f)), dont les variations sont corrélées à la distance Cys51-H_{γ} – Glu94-$O_{\varepsilon 2}$.

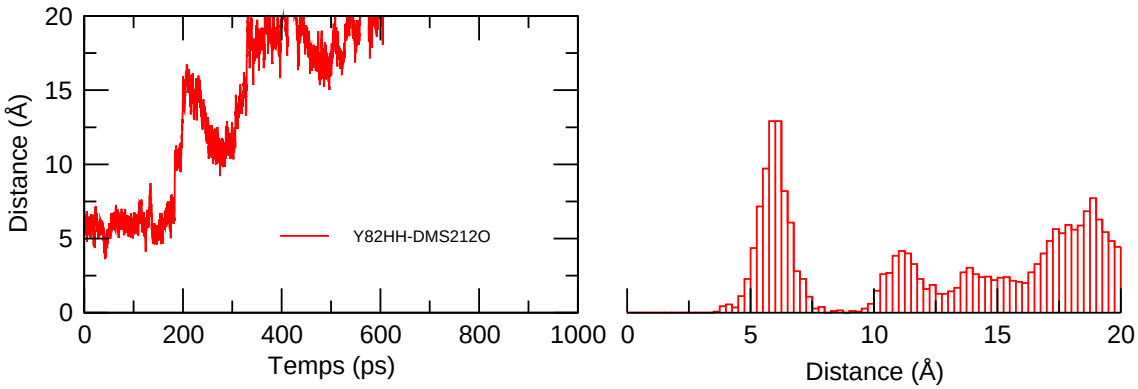
L'analyse de la solvatation de Cys51 et Glu94 (Fig. 6.6(k) et 6.6(l)) montre que le nombre de molécules d'eau autour de ces résidus augmente au cours du temps. Au départ, 2 molécules d'eau se situent autour de Cys51 (distance < 3,4 Å). En fin de simulation, ce nombre a doublé. Pour le résidu Glu94, la solvatation augmente de façon importante au cours des 200 premières picosecondes de simulation, passant de 0 à 2 molécules de solvant (distance < 3,4 Å).

Cette simulation montre donc que l'état de protonation étudié (CysH-Glu$^{\ominus}$) ne convient pas à l'ancrage du substrat sulfoxyde dans le site actif. Le substrat initialement placé dans le site actif s'échappe comme conséquence d'un bilan très défavorable des interactions à la fois entre le substrat et l'enzyme, et entre les résidus du site actif de l'enzyme. Le substrat interagit plus fortement avec le solvant qu'avec les résidus du site actif. Une analyse des charges atomiques peut expliquer, qualitativement, ce comportement (Tab. 6.5), la liaison O–H de la molécule d'eau étant beaucoup plus polaire que la liaison S–H de la cystéine.

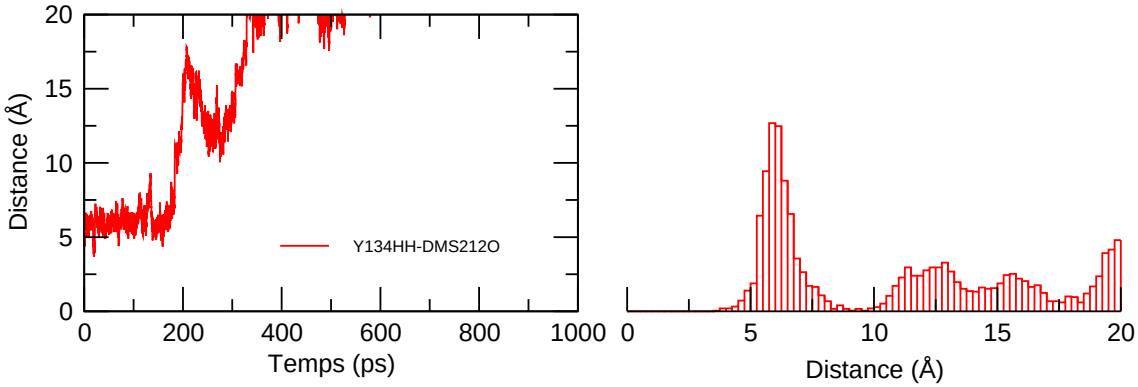
Atomes	DMSO-O	DMSO-S	Cys51-S_{γ}	Cys51-H_{γ}	WAT-O	WAT-H
Charge (u.a.)	-0,47	0,27	-0,31	0,19	-0,83	0,42

TAB. 6.5: *Valeurs des charges atomiques DMSO-O, DMSO-S, Cys51-S_{γ}, Cys51-H_{γ}, WAT-O et WAT-H utilisées pour le calcul du terme d'interaction coulombienne du champ de forces correspondant à la simulation DMSO-CysH-Glu$^{\ominus}$.*

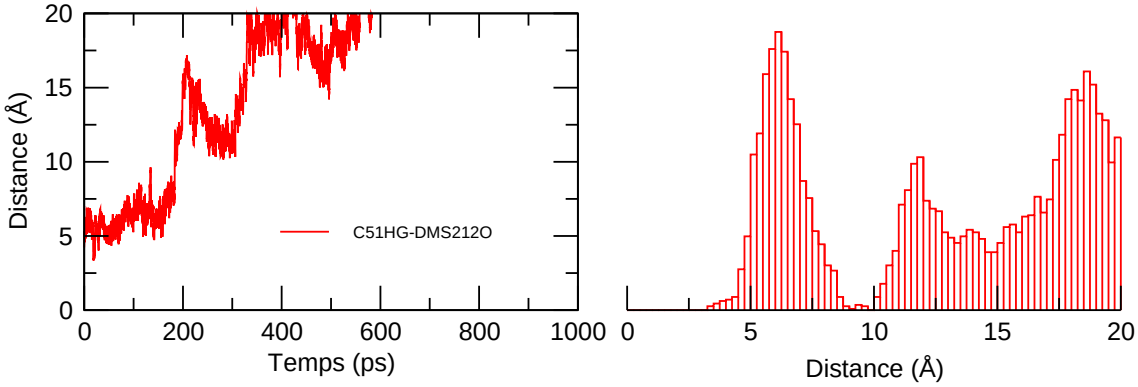
D'une manière plus globale, on peut dire que le grand moment dipolaire porté par la molécule de DMSO (4,59 D), lié à une liaison S=O à fort caractère ionique (S^{+}–O^{-}), n'est pas aisément stabilisé dans le site actif avec un état de protonation CysH-Glu$^{\ominus}$. La charge négative portée par Glu$^{\ominus}$ implique certainement une répulsion importante vis-à-vis de l'atome d'oxygène du DMSO qui tend à réorienter la molécule de substrat d'une façon peu favorable à la reconnaissance par les autres résidus, soit polaires (Cys51, Tyr82 et Tyr134) soit hydrophobes (Phe52 et Trp53).



(a) Distance Tyr82-H_η – DMSO-O

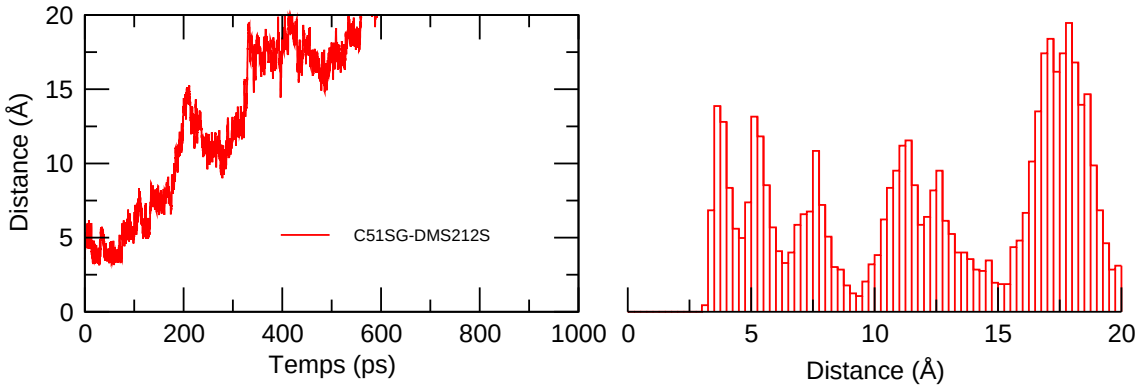


(b) Distance Tyr134-H_η – DMSO-O

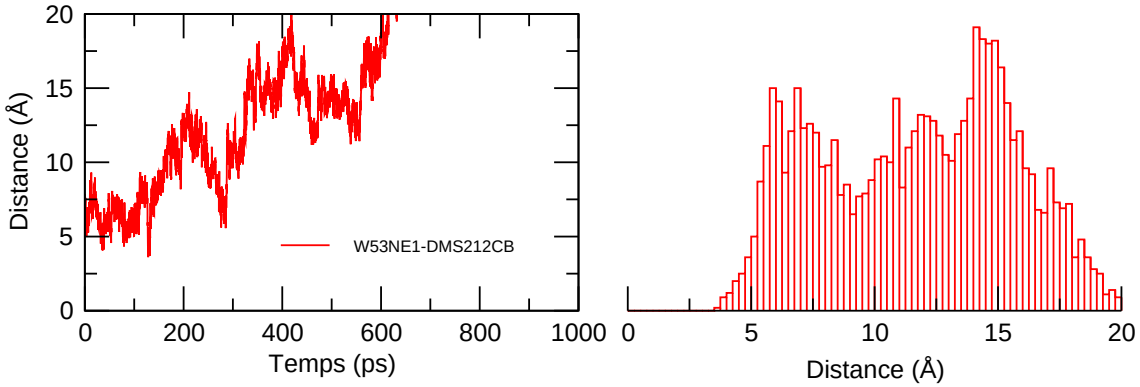


(c) Distance Cys51-H_γ – DMSO-O

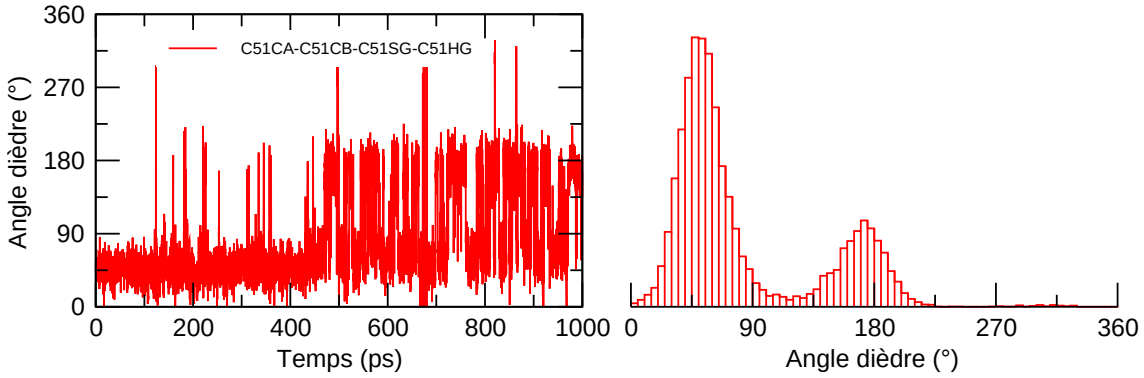
FIG. 6.6: *Variation de la distance Tyr82-H_η – DMSO-O (a), de la distance Tyr134-H_η – DMSO-O (b) et de la distance Cys51-H_γ – DMSO-O (c) en fonction du temps dans la simulation DMSO-CysH-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(d) Distance Cys51-S_γ – DMSO-S

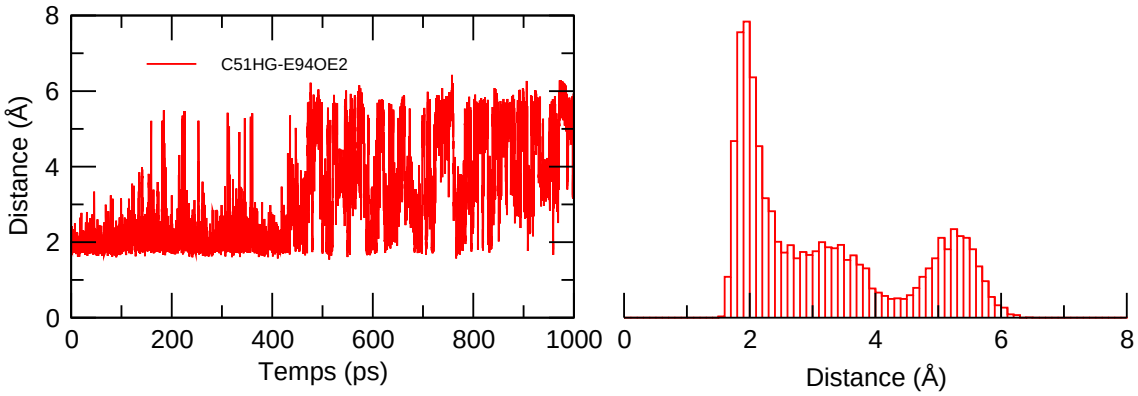


(e) Distance Trp53-$N_{\varepsilon 1}$ – DMSO-C_β

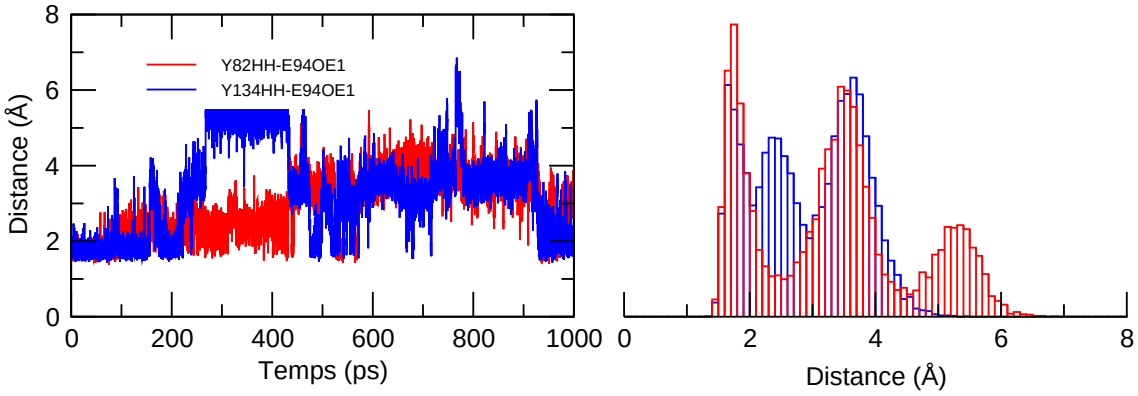


(f) Angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ

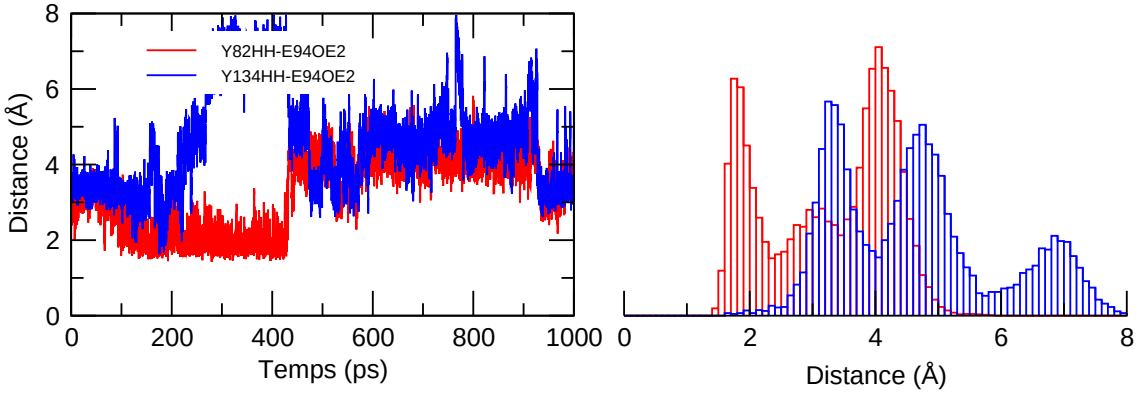
FIG. 6.6: *(suite) Variation de la distance Cys51-S_γ – DMSO-S (d), de la distance Trp53-$N_{\varepsilon 1}$ – DMSO-C_β (e) et de l'angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ (f) en fonction du temps dans la simulation DMSO-CysH-Glu$^\ominus$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(g) Distance Cys51-H_γ – Glu94-$O_{\varepsilon 2}$

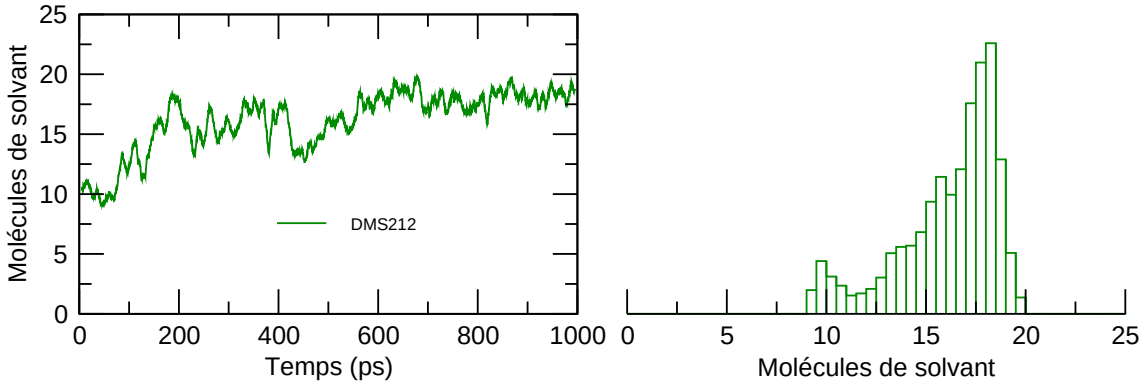


(h) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (bleu)

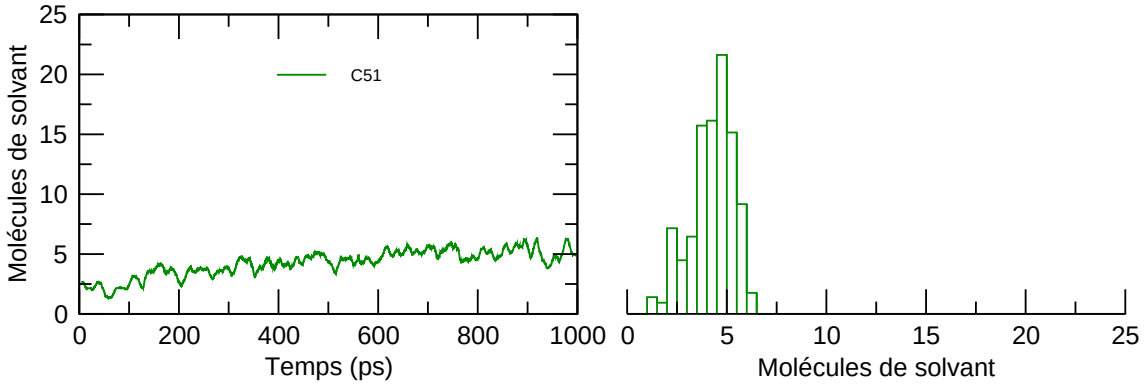


(i) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 2}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ (bleu)

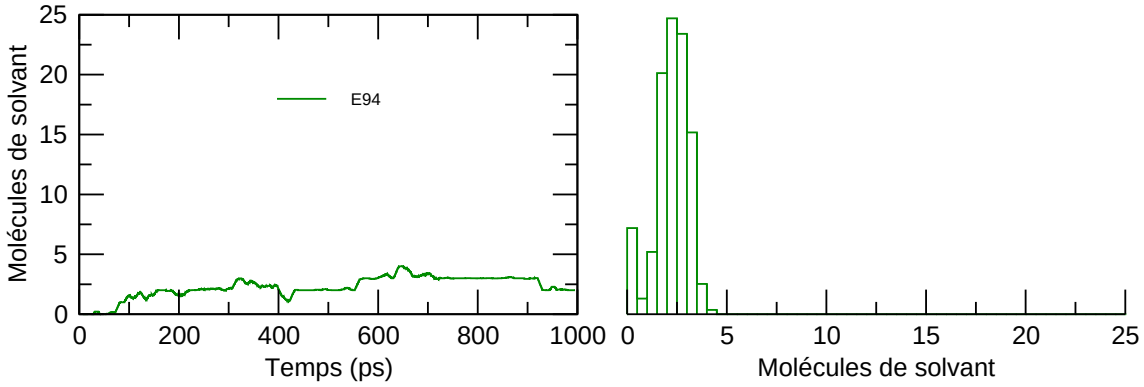
FIG. 6.6: *(suite) Variation de la distance Cys51-H_γ – Glu94-$O_{\varepsilon 2}$ (g), des distances Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (h) et des distances Tyr82-H_η – Glu94-$O_{\varepsilon 2}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ (f) en fonction du temps dans la simulation DMSO-CysH-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(j) Nombre de molécules d'eau à proximité du substrat DMSO (distance < 3,4 Å)



(k) Nombre de molécules d'eau à proximité du résidu Cys51 (distance < 3,4 Å)



(l) Nombre de molécules d'eau à proximité du résidu Glu94 (distance < 3,4 Å)

FIG. 6.6: *(suite) Variation du nombre de molécules d'eau dans la solvatation du substrat DMSO (j), du résidu Cys51 (k) et du résidu Glu94 (l) en fonction du temps dans la simulation DMSO-CysH-Glu$^{\ominus}$. Les valeurs sont moyennées sur 10 ps. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*

6.4.2. Etat de protonation CysH-GluH

Les résultats de la simulation sont présentés dans la figure 6.7. Dans ce cas, le substrat reste dans le site actif de l'enzyme durant toute la durée de la simulation. La protonation de Glu94 permet l'établissement d'une liaison hydrogène très forte avec l'atome d'oxygène du substrat (Fig. 6.7(a)). La distance Glu94-$H_{\varepsilon 2}$ – DMSO-O vaut 2 Å environ, la plupart du temps, avec quelques très faibles fluctuations. Au delà de son interaction avec Glu94, l'atome d'oxygène du DMSO se trouve au centre d'un réseau de liaisons hydrogène avec les résidus Tyr82 et Tyr134 situés à proximité (Fig. 6.7(b) et 6.7(c)). L'interaction avec l'hydrogène H_η de Tyr82 semble être particulièrement forte. L'hydrogène H_γ de Cys51 est aussi en interaction avec l'oxygène du DMSO, sur une période de 50 ps (aux alentours de 200 ps, Fig. 6.7(d)). Le reste du temps, cet hydrogène interagit avec Tyr197 comme le confirme l'analyse de l'angle dièdre de Cys51 autour de la liaison C_β – S_γ (Fig. 6.7(e)) et la distance Cys51-H_γ – Tyr197-O_η (Fig. 6.7(f)). Globalement, les interactions entre les résidus du site actif sont peu nombreuses. Nous n'avons pas observé d'interaction entre Tyr82 ou Tyr134 et les atomes d'oxygène de Glu94 (Fig. 6.7(k) et 6.7(l)), tous ces résidus stabilisant principalement le substrat.

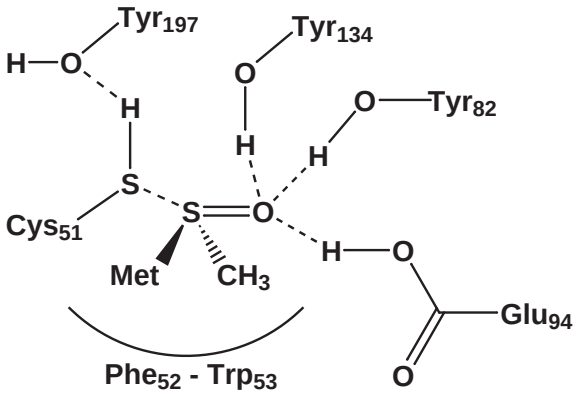
D'autre part, une interaction électrostatique existe entre les atomes de soufre portés par Cys51 et DMSO. Cette interaction était déjà présente dans la simulation précédente, mais comme nous l'avons vu, elle était rapidement rompue par la migration du sulfoxyde vers le solvant. Dans le cas présent, le substrat est maintenu dans le site actif par un réseau de liaisons hydrogène au niveau de l'atome d'oxygène, tel que nous venons de le décrire. L'interaction électrostatique entre les deux atomes de soufre peut ainsi s'établir de manière durable, ce que nous observons. La distance entre ces atomes varie peu autour de 3,5 Å (Fig. 6.7(g)). L'orientation spécifique du substrat dans le site actif a aussi pour effet de permettre l'existence de l'interaction de type hydrophobe avec les résidus Phe52 et Trp53 avec une distance entre le groupement méthyle C_β du DMSO et l'azote $N_{\varepsilon 1}$ de Trp53 variant peu autour de 3,5 Å (Fig. 6.7(h)). Cette interaction hydrophobe semble fournir un point d'ancrage supplémentaire au substrat à l'intérieur du site actif afin d'en diminuer la mobilité. Cette hypothèse est vérifiée par l'analyse de l'angle dièdre Cys51-C_β – Cys51-S_γ – DMSO-S – DMSO-C_β (Fig. 6.7(i)) qui reste globalement constant avec une valeur fluctuant autour de 225°. L'angle diminue à 90° pendant environ 50 ps (aux alentours de 200 ps) ce qui est à mettre en correspondance avec l'interaction de l'hydrogène H_γ de Cys51 et de l'oxygène du DMSO. Pendant cette période, l'orientation du substrat dans le site actif change, puis revient à sa valeur initiale. L'analyse de l'angle Cys51-S_γ – DMSO-S – DMSO-O (Fig. 6.7(j)) confirme également ce mouvement. Plus généralement, on peut dire que la structure globale du substrat dans le site ressemble fortement

à la structure cristallographique 1FF3 dans laquelle le résidu S-(diméthylarsénic)-cystéine mime le substrat sulfoxyde (S–S–O = 159,1°).

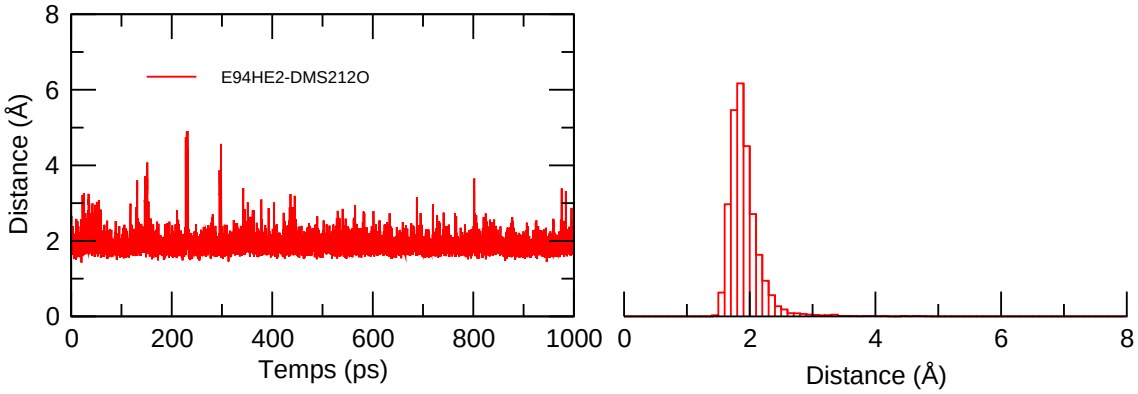
Enfin, la solvatation du DMSO et des résidus Cys51 et Glu94 a été analysée. Le nombre de molécules d'eau autour du substrat est constante durant toute la simulation avec 5 molécules de solvant à proximité (distance < 3,4 Å, Fig. 6.7(m)). Les résidus Cys51 et Glu94 sont très peu solvatés. Nous comptons 2 molécules de solvant pour Cys51 (Fig. 6.7(n)) et aucune pour Glu94 (Fig. 6.7(o)). Le site actif de l'enzyme est désolvaté de manière importante par la présence du substrat comparativement à la simulation Free-CysH-GluH analogue pour l'enzyme libre.

La présence d'un proton sur Glu94 apparaît donc comme un facteur déterminant permettant d'expliquer l'ancrage du substrat sulfoxyde. Il permet l'établissement d'une liaison hydrogène forte avec l'atome d'oxygène du DMSO ainsi que l'établissement d'un réseau assez important de liaisons hydrogène avec les résidus Tyr82 et Tyr134. Au delà, cette interaction permet d'autres points d'ancrage soit par l'intermédiaire de l'interaction électrostatique entre le soufre de Cys51, chargé négativement, et le soufre du DMSO, chargé positivement, soit par l'interaction de type hydrophobe d'un groupement méthyle du DMSO avec Phe52 et Trp53, qui contraint toute rotation du substrat.

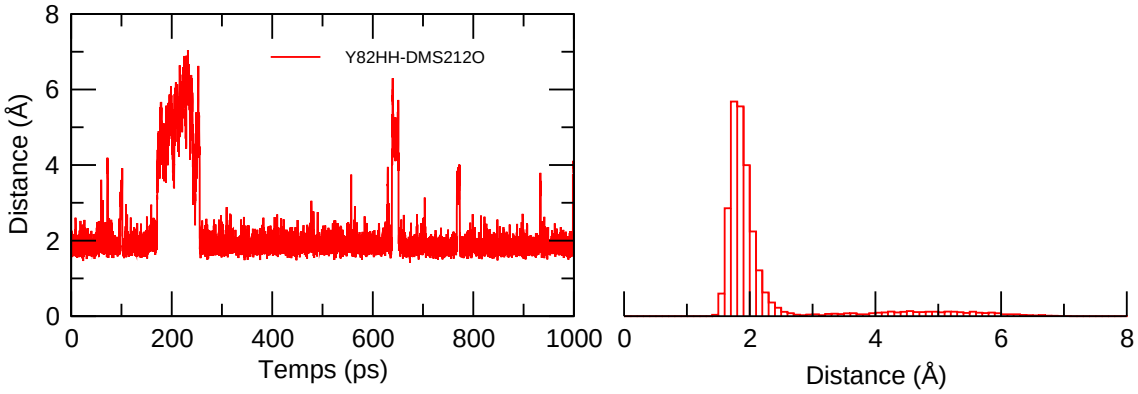
L'implication des résidus Tyr82 et Tyr134 dans la stabilisation et le maintien du substrat sulfoxyde est compatible avec les résultats expérimentaux de mutagénèse dirigée mentionnant que l'absence des deux tyrosines conduit à une perte totale de l'activité alors que l'enzyme reste active en l'absence d'une seule tyrosine[a]. Les différentes interactions observées entre les résidus du site actif et le substrat sulfoxyde sont résumées dans le schéma ci-dessous.



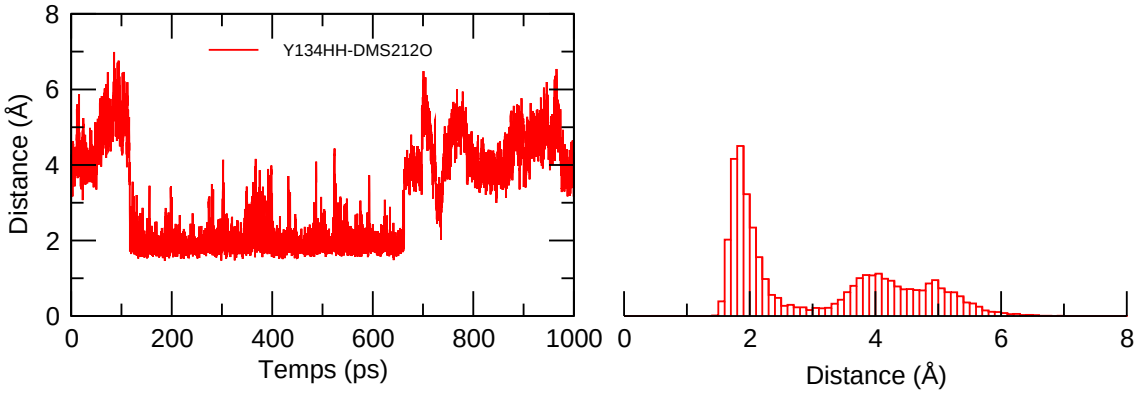
[a]Nous poursuivrons la description du rôle de ces tyrosines dans le chapitre 8.



(a) Distance Glu94-$H_{\varepsilon 2}$ – DMSO-O

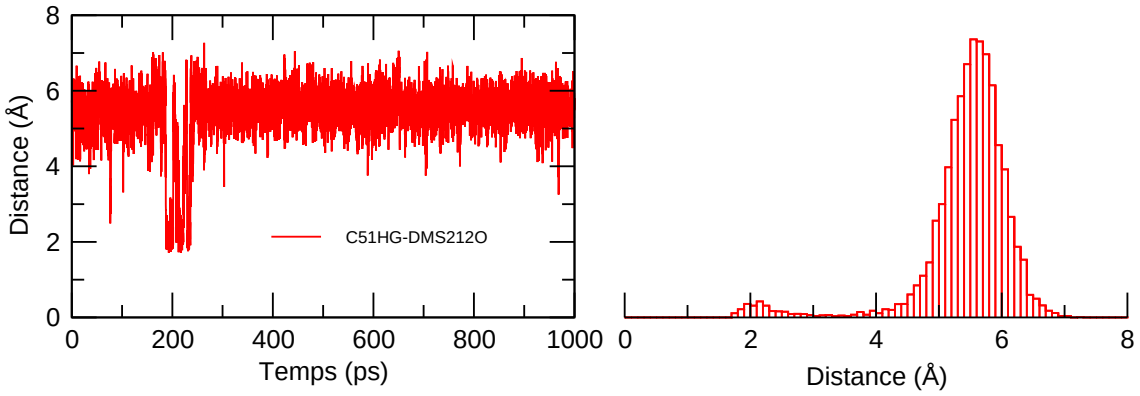


(b) Distance Tyr82-H_{η} – DMSO-O

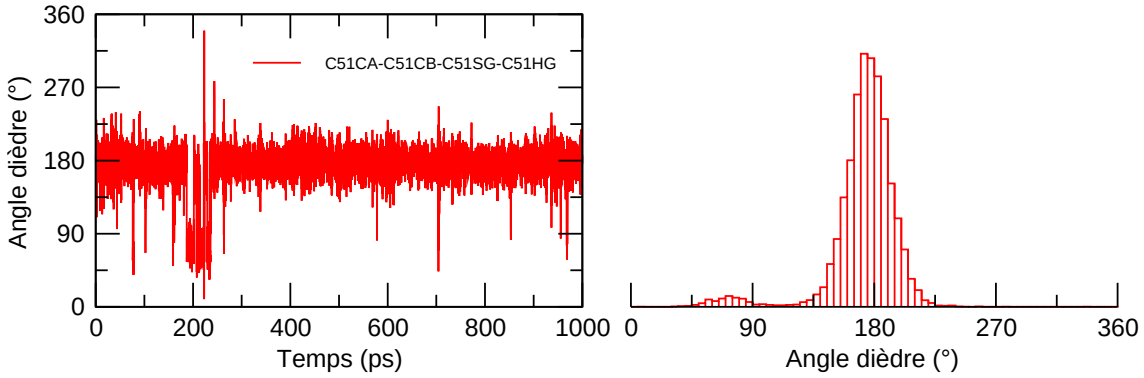


(c) Distance Tyr134-H_{η} – DMSO-O

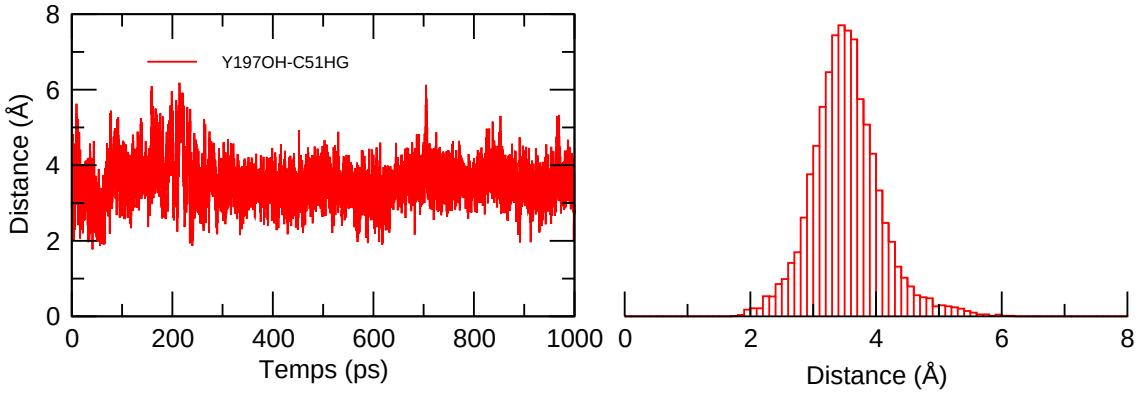
FIG. 6.7: *Variation de la distance Glu94-$H_{\varepsilon 2}$ – DMSO-O (a), de la distance Tyr82-H_{η} – DMSO-O (b) et de la distance Tyr134-H_{η} – DMSO-O (c) en fonction du temps dans la simulation DMSO-CysH-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(d) Distance Cys51-H_γ – DMSOO

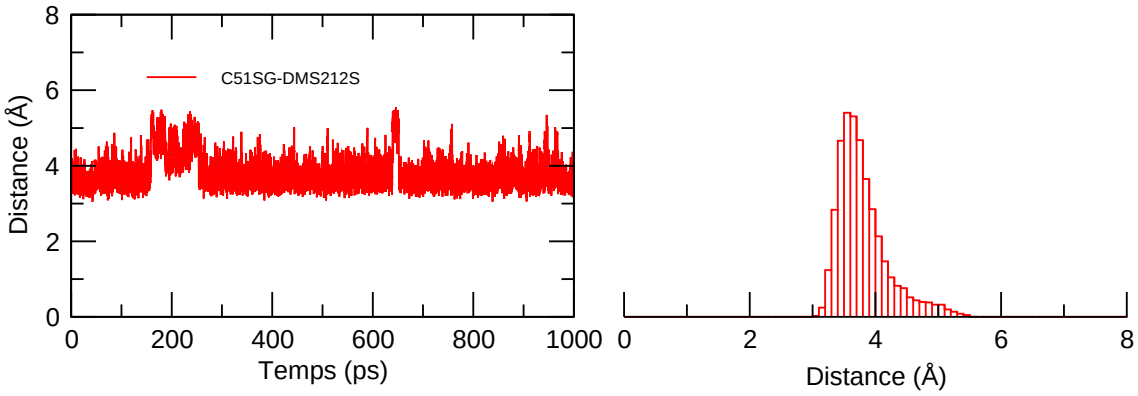


(e) Angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ

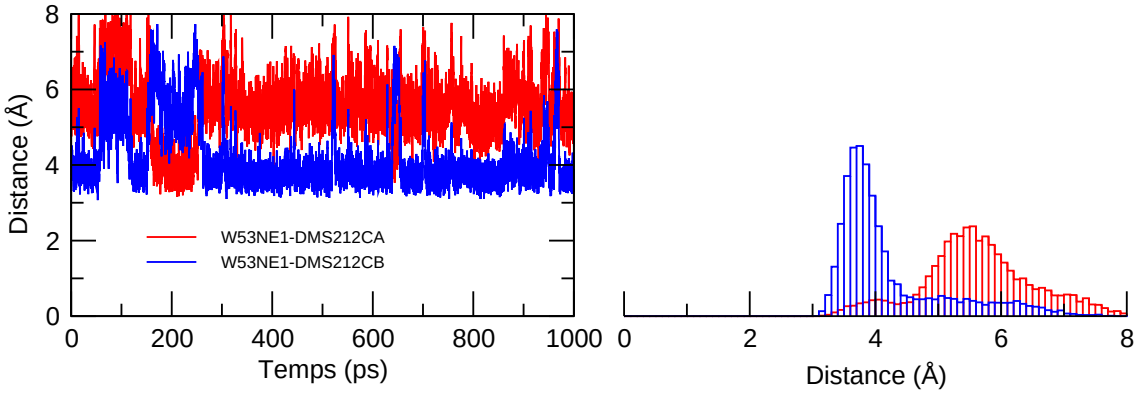


(f) Distance Cys51-H_γ – Tyr197-O_η

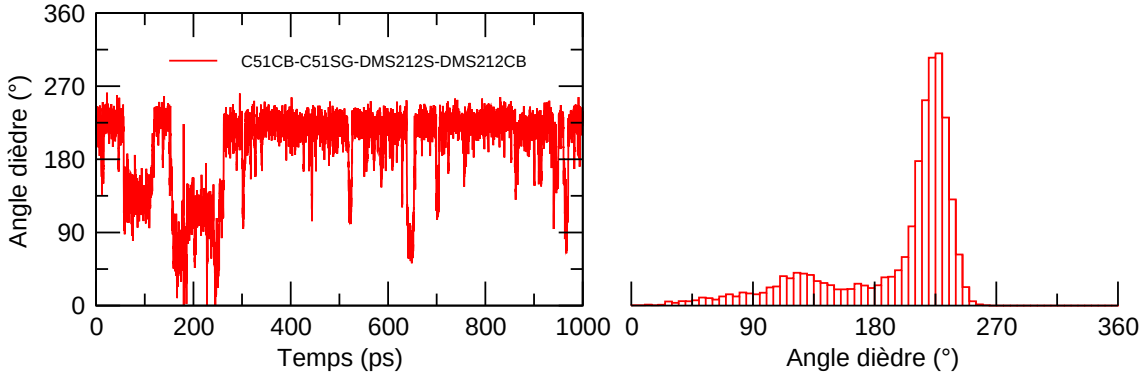
FIG. 6.7: *(suite) Variation de la distance Cys51-H_γ – DMSO-O (d), de l'angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ (e) et de la distance Cys51-H_γ – Tyr197-O_η (f) en fonction du temps dans la simulation DMSO-CysH-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(g) Distance Cys51-S_γ – DMSO-S

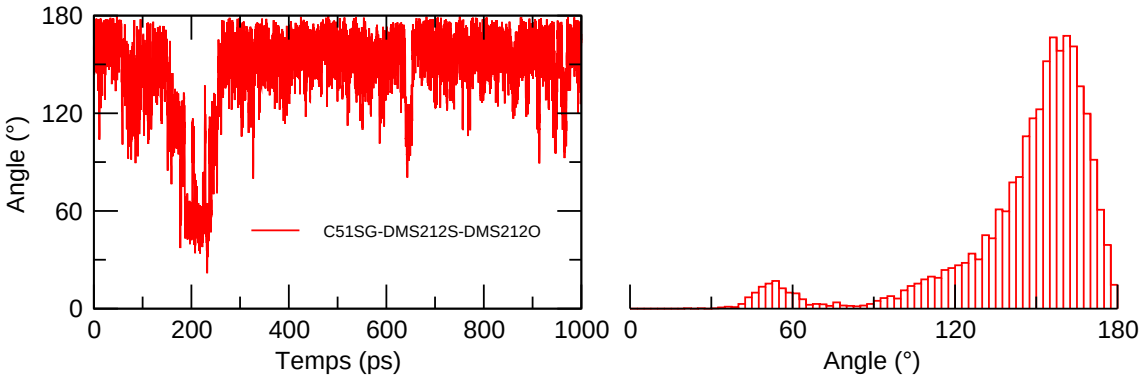


(h) Distance Trp53-$N_{\varepsilon 1}$ – DMSO-C_α (rouge) et Trp53-$N_{\varepsilon 1}$ – DMSO-C_β (bleu)

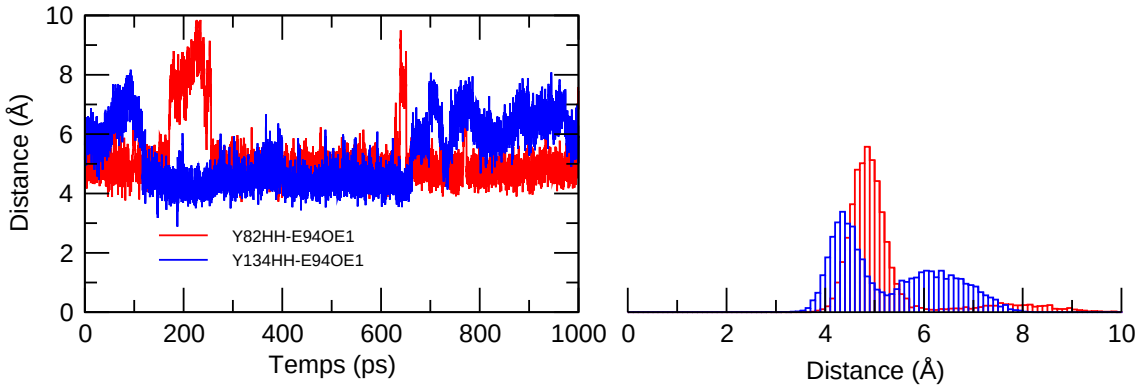


(i) Angle dièdre Cys51-C_β – Cys51-S_γ – DMSO-S – DMSO-C_β

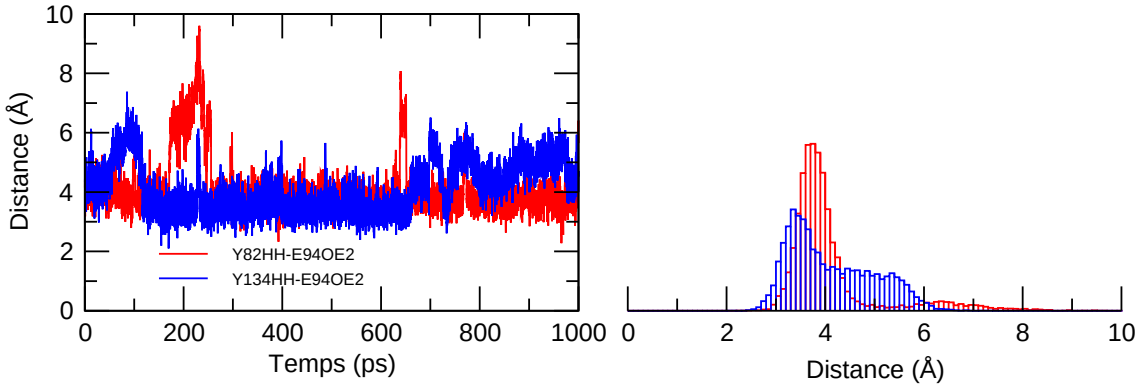
FIG. 6.7: *(suite) Variation de la distance Cys51-S_γ – DMSO-S (g), de la distance Trp53-$N_{\varepsilon 1}$ – DMSO-C_α et Trp53-$N_{\varepsilon 1}$ – DMSO-C_β (h) et de l'angle dièdre Cys51-C_β – Cys51-S_γ – DMSO-S – DMSO-C_β (i) en fonction du temps dans la simulation DMSO-CysH-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(j) Angle Cys51-S_γ – DMSO-S – DMSO-O

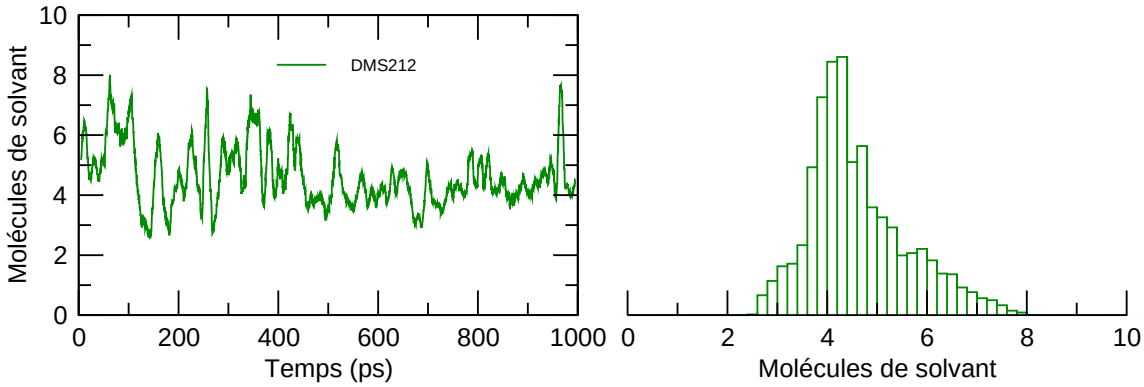


(k) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (bleu)

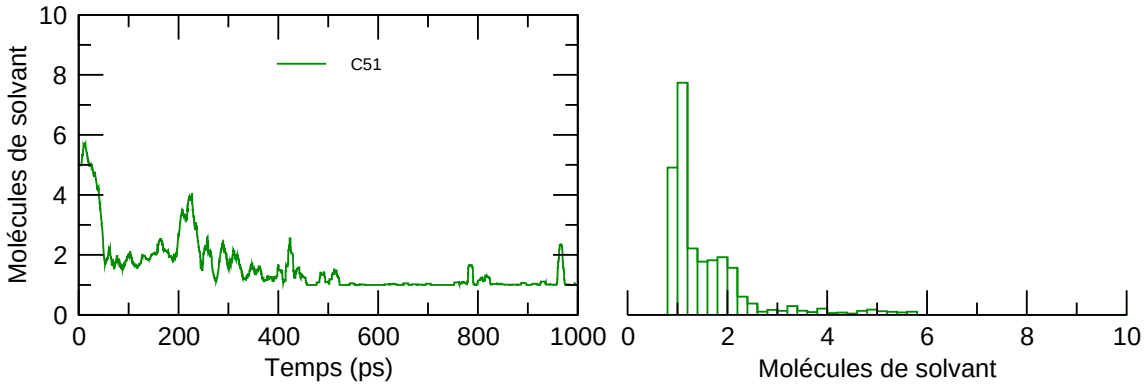


(l) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 2}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ (bleu)

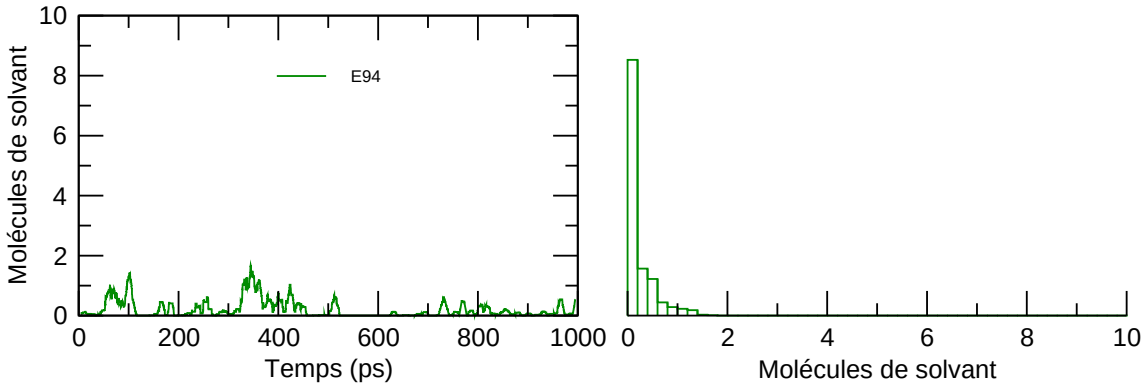
FIG. 6.7: *(suite) Variation de l'angle Cys51-S_γ – DMSO-S – DMSO-O (j), de la distance Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (k) et de la distance Tyr82-H_η – Glu94-$O_{\varepsilon 2}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ (l) en fonction du temps dans la simulation DMSO-CysH-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(m) Nombre de molécules d'eau à proximité du substrat DMSO (distance < 3,4 Å)



(n) Nombre de molécules d'eau à proximité du résidu Cys51 (distance < 3,4 Å)



(o) Nombre de molécules d'eau à proximité du résidu Glu94 (distance < 3,4 Å)

FIG. 6.7: *(suite) Variation du nombre de molécules d'eau dans la solvatation du substrat DMSO (j), du résidu Cys51 (k) et du résidu Glu94 (l) en fonction du temps dans la simulation DMSO-CysH-GluH. Les valeurs sont moyennées sur 10 ps. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*

6.4.3. Etat de protonation Cys$^{\ominus}$-GluH

Les résultats de la simulation sont présentés dans la figure 6.9. Nous retrouvons la grande majorité des interactions que nous avons décrites dans la simulation précédente. Il se forme tout au long de la simulation un réseau important de liaisons hydrogène autour de l'atome d'oxygène du substrat sulfoxyde. Une liaison hydrogène forte à une distance de 2 Å se forme entre l'oxygène du DMSO et l'hydrogène $H_{\varepsilon 2}$ de Glu94 (Fig. 6.9(a)) et entre l'oxygène du DMSO et l'hydrogène H_{η} de Tyr82 (Fig. 6.9(b)). Une liaison hydrogène se forme aussi avec Tyr134H_{η} (Fig. 6.9(c)), mais de manière moins importante. Nous n'avons pas observé de liaison hydrogène se formant entre les résidus Tyr82, Tyr134, d'une part, et Glu94, d'autre part, durant la simulation (Fig. 6.9(k) et 6.9(l)). Ici, les tyrosines stabilisent donc principalement le substrat.

D'autre part, l'interaction électrostatique entre les deux atomes de soufre existe toujours avec une distance Cys51-S_{γ} – DMSO-S variant peu autour de 4 Å (Fig. 6.9(d)). Cette distance d'équilibre est la même que celle observée dans la précédente simulation malgré la déprotonation de Cys51. La charge négative portée par le soufre S_{γ} de Cys51 est stabilisée par l'hydrogène H_{η} de Tyr197 avec une distance fréquente de 2 Å entre ces deux atomes (Fig. 6.9(e)). Cette interaction avait déjà été observée dans la simulation de l'enzyme libre Free-Cys$^{\ominus}$-GluH. Le substrat est encore moins mobile dans cet état de protonation que dans la simulation précédente. L'interaction hydrophobe du groupement méthyle du DMSO avec Trp53 est relativement importante avec une distance variant très peu autour de 4 Å (Fig. 6.9(f)). Ceci se vérifie par l'observation de l'angle dièdre Cys51-C_{β} – Cys51-S_{γ} – DMSO-S – DMSO-C_{β} (Fig. 6.9(g)) qui varie très peu durant la simulation. A noter que la variation subite de -110° à 0° vers 300 ps est principalement due au changement de l'angle dièdre de Cys51 autour de la liaison C_{α}–C_{β} (Fig. 6.9(h)). Signalons aussi que la variation de l'angle dièdre de Cys51 s'accompagne d'une modification de la solvatation de Cys51 (Fig. 6.9(i)). Cependant, cette variation n'entraîne aucune modification de la structure du complexe du substrat DMSO dans le site actif, comme nous pouvons le constater sur les autres courbes. L'angle Cys51-S_{γ} – DMSO-S – DMSO-O s'équilibre autour de la valeur de 120° (Fig. 6.9(j)). Cette valeur est inférieure à celle de 160° obtenue dans l'état de protonation précédent et dans l'analyse des structures cristallographiques. Nous montrerons par la suite que la géométrie optimisée au niveau quantique de ce système donne un angle proche aussi de 160°.

Concernant la solvatation du site actif, nous retrouvons des valeurs proches pour le substrat DMSO et les résidus Cys51 et Glu94 par rapport à la simulation précédente. Le substrat est entouré de 5 molécules de solvant en moyenne (distance < 3,4 Å, courbe non présentée). Les résidus Cys51

et Glu94 sont relativement peu solvatés. La présence d'une charge négative sur Cys51 entraîne une plus grande solvatation de ce résidu par rapport à la simulation précédente, dans laquelle il est protoné, avec 3 molécules de solvant (Fig. 6.9(i)). A nouveau, aucune molécule d'eau n'interagit directement avec Glu94 (courbe non présentée). La faible solvatation globale du site actif traduit la stabilité de la configuration $Cys^{\ominus} \cdots DMSO \cdots GluH$ du complexe enzyme-substrat.

L'état de protonation considéré ici est donc celui qui permet de stabiliser le substrat sulfoxyde dans le site actif de la manière la plus efficace. Le système que nous venons de simuler peut donc être associé au complexe michaélien dont la structure est illustrée par la figure 6.8.

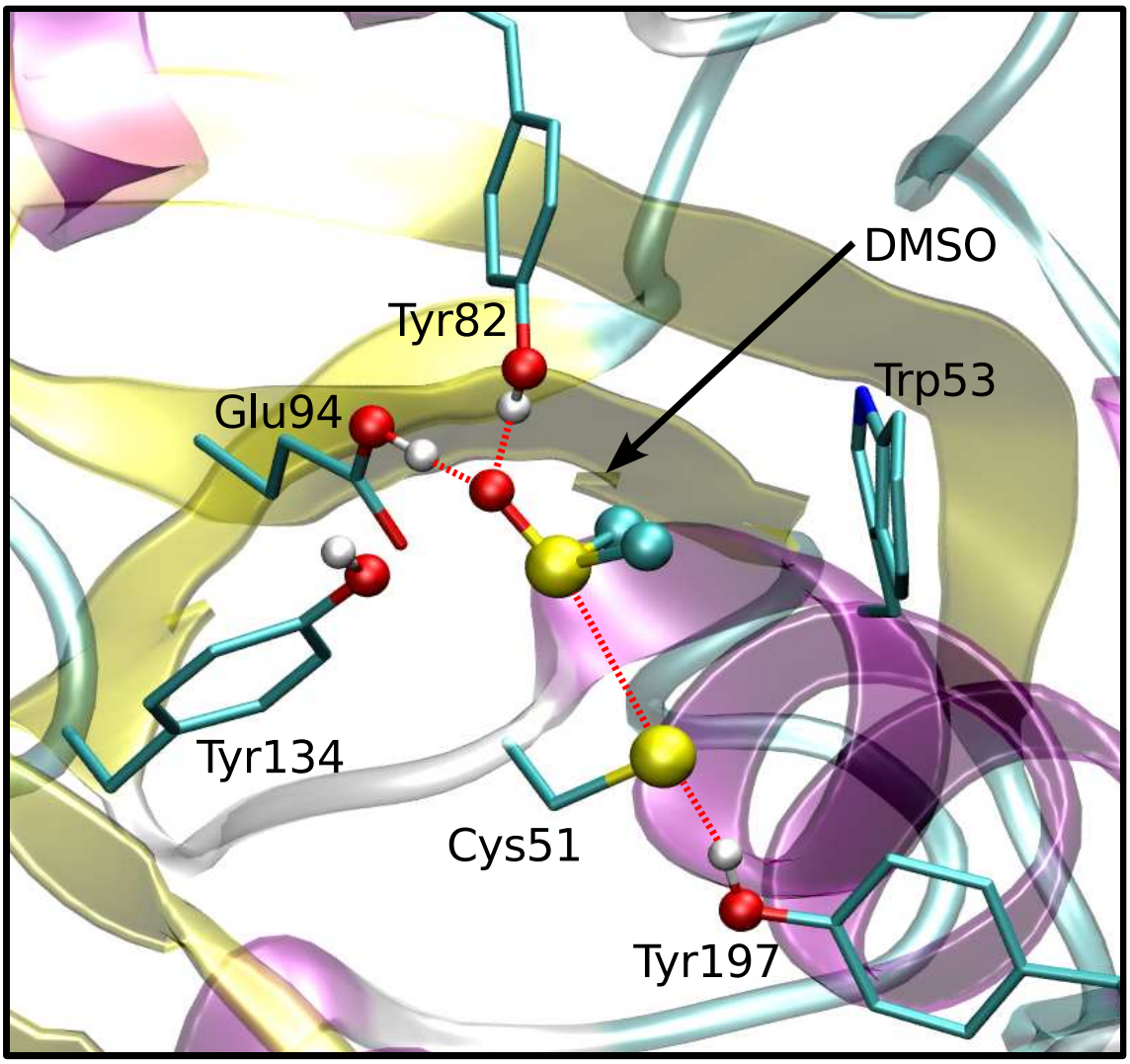
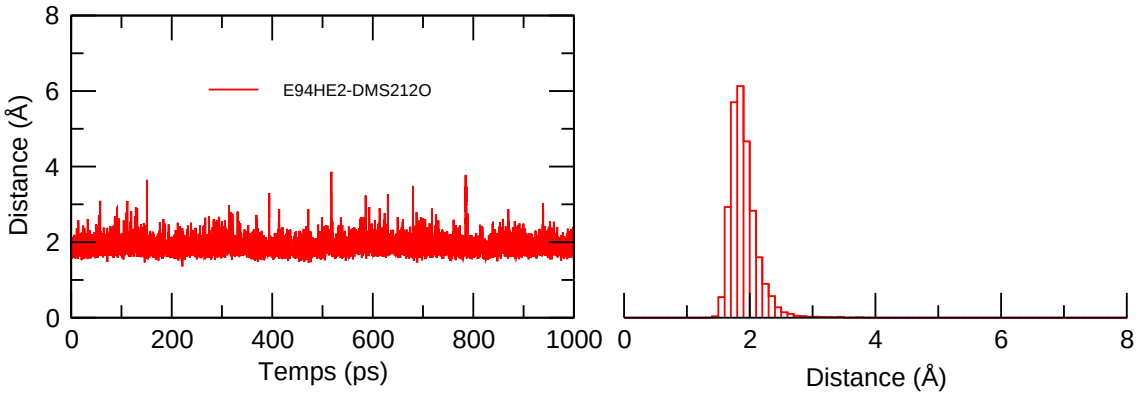
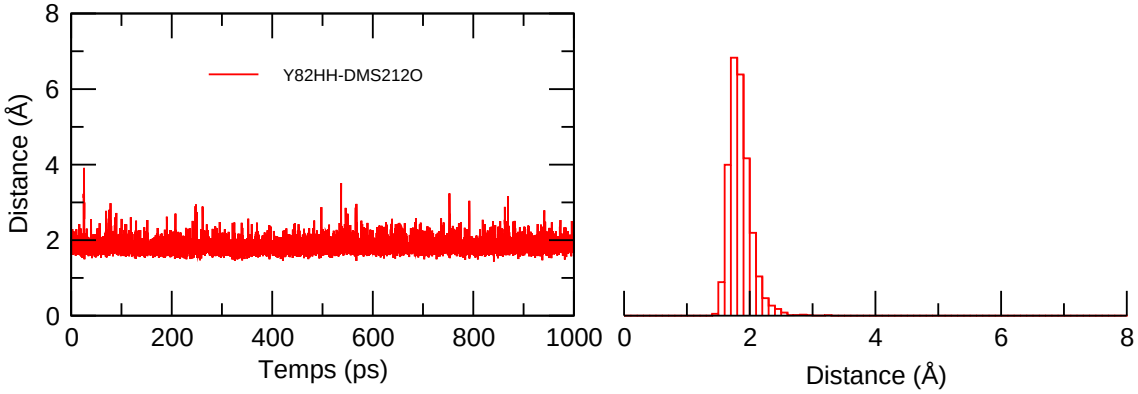


FIG. 6.8: *Structure du complexe michaélien du DMSO dans le site actif de MsrA. Les résidus du site actif impliqués dans les interactions avec le substrat sulfoxyde sont représentés. Seuls les hydrogènes participant aux interactions ont été matérialisés. Les hélices α sont colorées en violet et les feuillets β en jaune.*

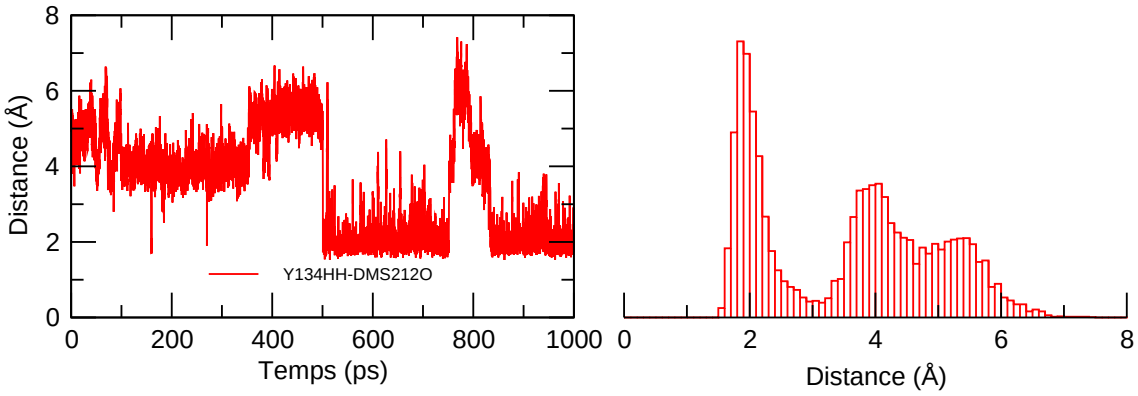
Par rapport à l'enzyme libre où l'état de protonation le plus stable est vraisemblablement CysH-Glu$^{\ominus}$ ou CysH-GluH, le complexe michaélien impliquerait donc une déprotonation de la cystéine. Les simulations avec un état de protonation Cys$^{\ominus}$-GluH pour l'enzyme libre et pour le complexe enzyme-substrat montrent que Tyr197 interagit avec Cys51 par une forte interaction de l'hydrogène H_{η} avec le soufre S_{γ}. De même, il a été mentionné que, dans les simulations avec l'état de protonation CysH-GluH, l'hydrogène H_{γ} de Cys51 s'oriente fréquemment vers cette tyrosine. Par ailleurs, cette partie du site actif offre une grande accessibilité au solvant. On peut ainsi formuler l'hypothèse que Tyr197 participe au processus de déprotonation de la cystéine catalytique, probablement en réalisant un relais de proton. Cette hypothèse pourrait être vérifiée par une expérience de mutagénèse dirigée.



(a) Distance Glu94-$H_{\varepsilon 2}$ – DMSO-O

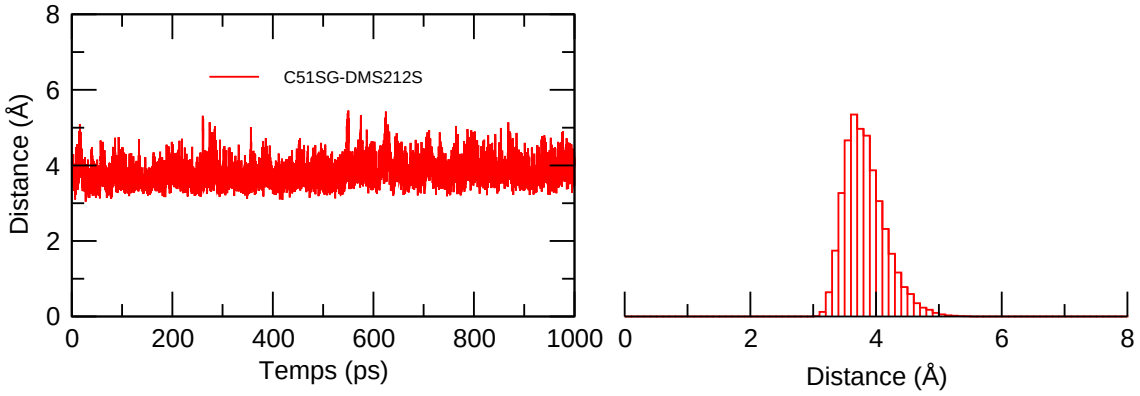


(b) Distance Tyr82-H_{η} – DMSO-O

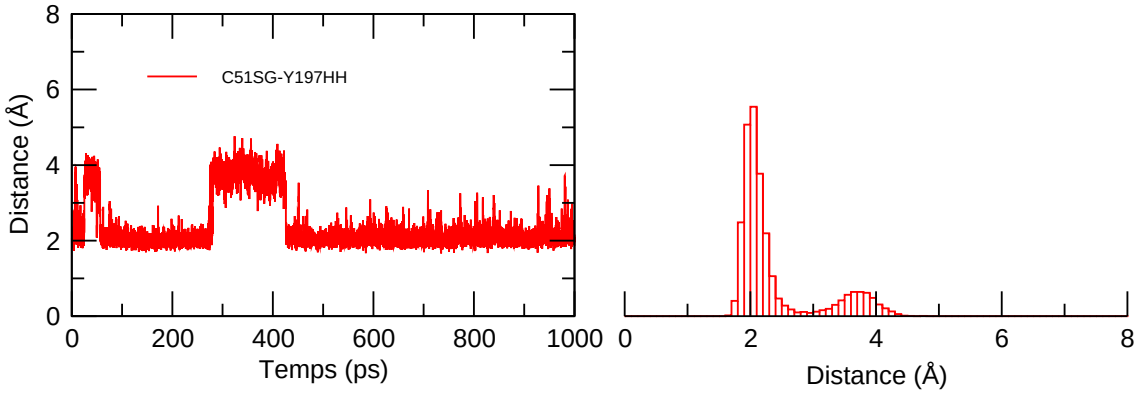


(c) Distance Tyr134-H_{η} – DMSO-O

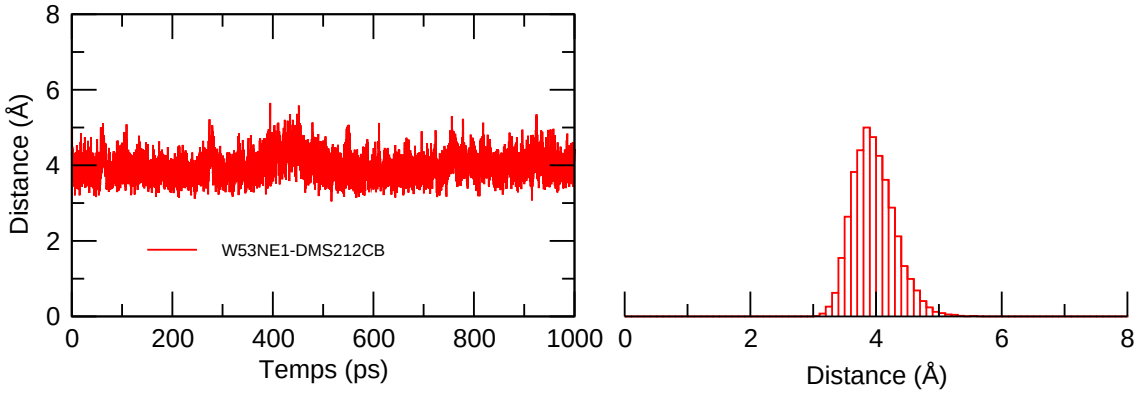
FIG. 6.9: *Variation de la distance Glu94-$H_{\varepsilon 2}$ – DMSO-O (a), de la distance Tyr82-H_{η} – DMSO-O (b) et de la distance Tyr134-H_{η} – DMSO-O (c) en fonction du temps dans la simulation DMSO-Cys$^{\ominus}$-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(d) Distance Cys51-S_γ – DMSO-S

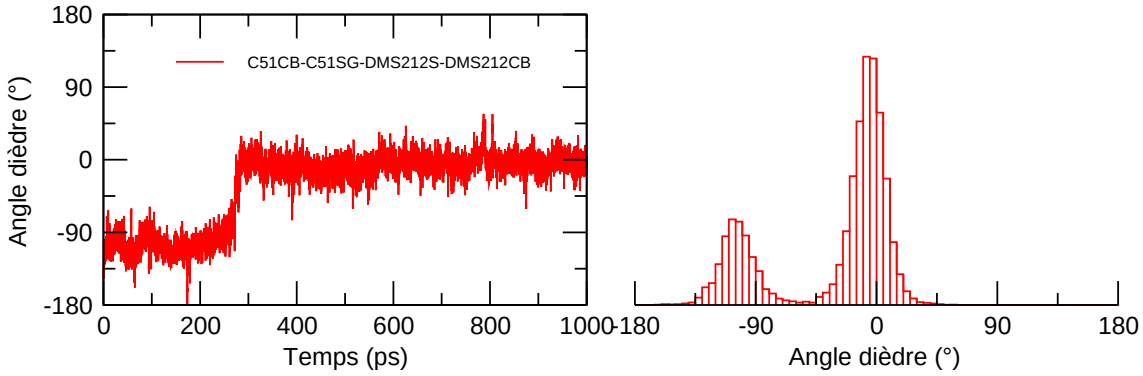


(e) Distance Cys51-S_γ – Tyr197-H_η

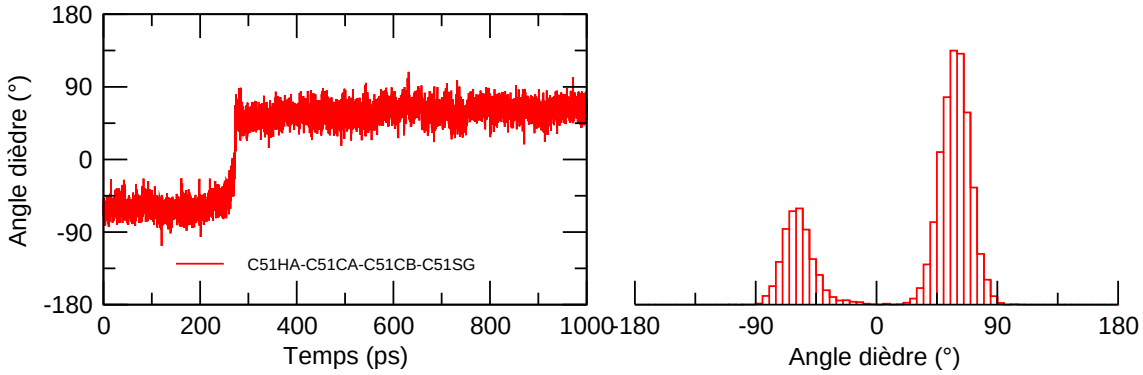


(f) Distance Trp53-$N_{\varepsilon 1}$ – DMSO-C_β

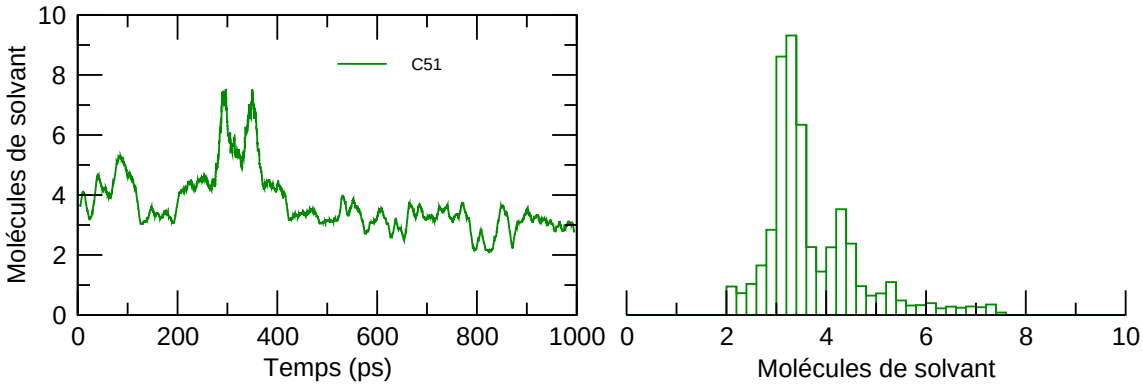
FIG. 6.9: *(suite) Variation de la distance Cys51-S_γ – DMSO-S (d), de la distance Cys51-S_γ – Tyr197-H_η (e) et de la distance Trp53-$N_{\varepsilon 1}$ – DMSO-C_β (f) en fonction du temps dans la simulation DMSO-Cys$^{\ominus}$-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(g) Angle dièdre Cys51-C_β – Cys51-S_γ – DMSO-S – DMSO-C_β

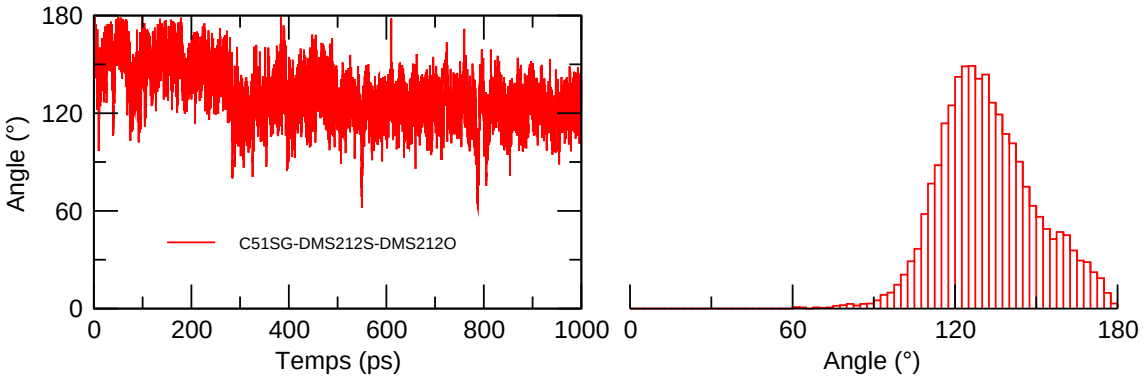


(h) Angle dièdre Cys51-H_α – Cys51-C_α – Cys51-C_β – Cys51-S_γ

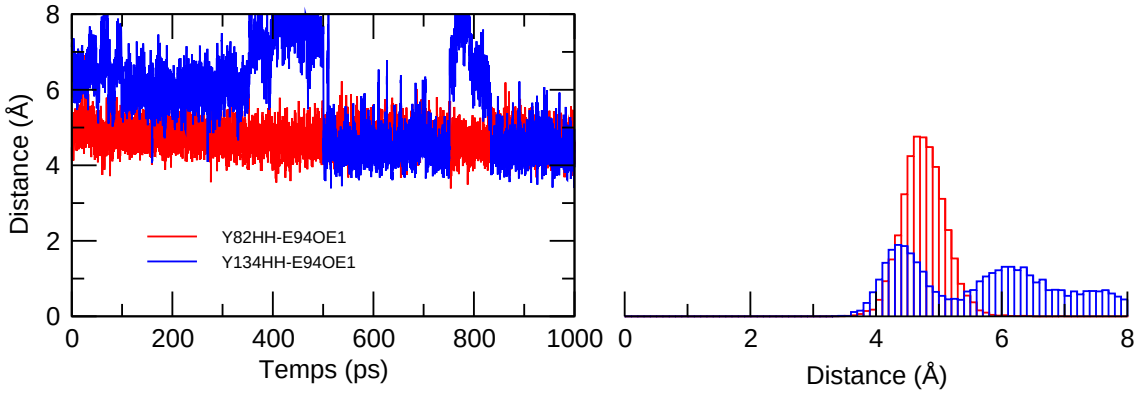


(i) Nombre de molécules d'eau à proximité du résidu Cys51 (distance < 3,4 Å)

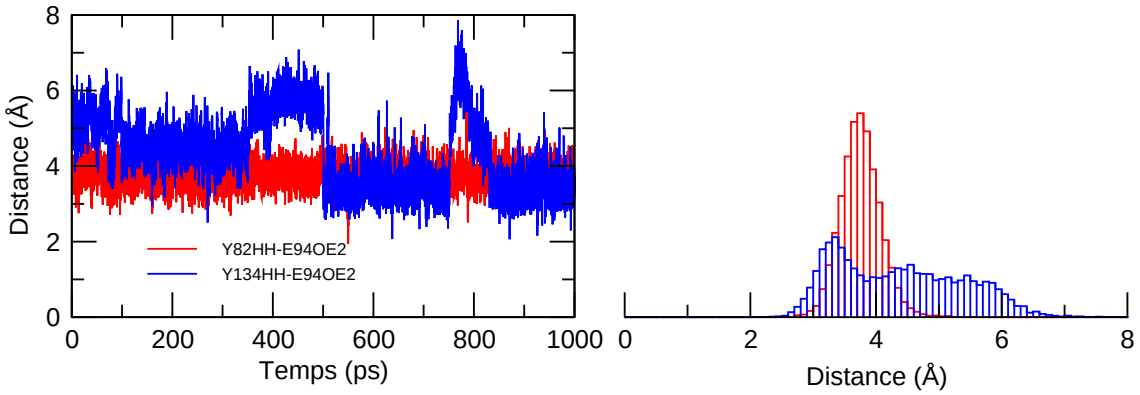
FIG. 6.9: *(suite) Variation de l'angle dièdre Cys51-C_β – Cys51-S_γ – DMSO-S – DMSO-C_β (g), de l'angle dièdre Cys51-H_α – Cys51-C_α – Cys51C_β – Cys51-S_γ (h) et du nombre de molécules d'eau (valeurs moyennées sur 10 ps) dans la solvatation de Cys51 (i) en fonction du temps dans la simulation DMSO-Cys$^{\ominus}$-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(j) Angle Cys51-S_γ – DMSO-S – DMSO-O



(k) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (bleu)



(l) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 2}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ (bleu)

FIG. 6.9: *(suite) Variation de l'angle Cys51-S_γ – DMSO-S – DMSO-O (j), des distances Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (k) et des distances Tyr82-H_η – Glu94-$O_{\varepsilon 2}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ (l) en fonction du temps dans la simulation DMSO-Cys$^{\ominus}$-GluH. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*

6.5. Conclusions

Les différentes simulations que nous venons de présenter nous ont permis d'obtenir des renseignements précieux sur le fonctionnement de l'enzyme tels que (*i*) la reconnaissance et l'ancrage du substrat sulfoxyde en identifiant les résidus impliqués et (*ii*) la structure du complexe michaélien enzyme-substrat.

La présence d'un proton sur le résidu Glu94 est essentielle pour la reconnaissance et l'ancrage du substrat. La prédiction d'une valeur de pK_a proche de 7 par le programme PROPKA indique la possibilité que la forme protonée de ce résidu existe dans l'enzyme libre. Cet état de protonation serait la conséquence d'une plutôt faible ou très faible accessibilité de l'eau au résidu Glu94, constatée dans toutes les simulations (phénomène de désolvatation qui a pour conséquence d'augmenter le pK_a de Glu94). La stabilité du complexe est principalement due à la liaison hydrogène entre GluH et l'oxygène du sulfoxyde. Des liaisons hydrogène avec Tyr82 et Tyr134, l'interaction électrostatique entre l'atome de soufre de Cys51 et l'atome de soufre du DMSO ainsi que l'interaction hydrophobe d'un groupement méthyle avec la poche hydrophobe formée par Phe52 et Trp53 contribuent à stabiliser le système. La nécessité d'un proton sur Glu94 pour la reconnaissance est compatible avec une expérience de mutagénèse dirigée montrant l'ancrage du substrat dans le site actif de MsrA lorsque Glu94 est muté en glutamine.

L'utilisation du DMSO comme substrat de l'enzyme dans nos simulations ne nous a pas permis de vérifier les autres sites de reconnaissance des peptides sulfoxydes de méthionine, déterminés expérimentalement par Gand *et al.*[232], au niveau de la chaîne latérale de ceux-ci.

Ainsi, l'étude des différentes états de protonation possibles pour le couple Cys51-Glu94, à la fois dans l'enzyme libre et le complexe michaélien avec le DMSO, montre que, à l'état libre, Cys51 est protoné (CysH) alors que Glu94 peut l'être ou non (GluH ou $Glu^{\ominus}$). Par contre dans le complexe enzyme-substrat, Glu94 doit obligatoirement être protoné pour assurer l'ancrage du substrat alors que la déprotonation de Cys51 n'est pas nécessaire lors de cette étape de reconnaissance. Cependant, une déprotonation de Cys51, éventuellement *via* Tyr197 vers le solvant, assure une stabilité supérieure du substrat dans le site actif en augmentant l'interaction entre le soufre de Cys51 et le moment dipolaire de la liaison S–O de la fonction sulfoxyde.

Chapitre 7

Activation du substrat

Sommaire

7.1. Réduction des sulfoxydes : généralités 158
7.2. Mécanisme de transfert de proton 159
7.3. Dynamique moléculaire du substrat activé 173
7.4. Conclusions 188

Comme nous venons de le voir, lors de l'insertion du substrat dans le site actif de l'enzyme, la cystéine catalytique est activée par déprotonation. Nous traiterons dans ce chapitre de l'étape suivante qui consiste en une protonation du substrat sulfoxyde permettant de l'activer. Nous étudierons par quel moyen cette protonation peut s'effectuer au sein de l'enzyme. Une étude du mécanisme de transfert de proton utilisant les méthodes de la chimie quantique sera réalisée afin d'évaluer l'énergétique liée à ce processus. Elle sera menée sur un modèle du site actif afin de prendre en compte l'effet de l'environnement enzymatique. Nous verrons que la protonation du substrat conduit à la formation d'une espèce de type sulfurane. Ce composé a déjà été mentionné précédemment comme étant un intermédiaire possible dans le mécanisme de réduction. Ici, il sera caractérisé dans le milieu enzymatique par une simulation de dynamique moléculaire. Nous discuterons de la stabilisation de cette espèce par des interactions avec les résidus du site actif.

7.1. Réduction des sulfoxydes : généralités

La réduction de sulfoxydes est un processus important en chimie et un certain nombre de réactions catalytiques ont été développées dans la littérature[247,271–279]. La réaction d'un sulfoxyde avec un thiol, en particulier, a été étudiée par plusieurs auteurs. Dans ce cas, et en absence de catalyseurs spécifiques, les travaux théoriques de Balta *et al.*[236] ont montré que le sulfoxyde est d'abord activé par protonation sur l'atome d'oxygène. Compte tenu des conclusions obtenues dans le chapitre précédent, on peut ainsi supposer que les principales étapes de la phase réductase du mécanisme enzymatique sont celles schématisées dans la figure 7.1. La réduction du sulfoxyde de méthionine en méthionine conduit à la formation de l'intermédiaire acide sulfénique (**A** → **D**). Dans le chapitre précédent, nous avons mis en évidence que la déprotonation de la cystéine catalytique intervient lors de l'ancrage du substrat sulfoxyde dans le site actif de l'enzyme. Le complexe michaélien du substrat sulfoxyde dans le site actif présente ainsi la cystéine catalytique sous une forme thiolate **B**. A partir du complexe michaélien, l'activation du substrat par protonation créerait par conséquent une paire d'ions dans le site actif **C** qui pourrait servir de force motrice pour le processus de réduction.

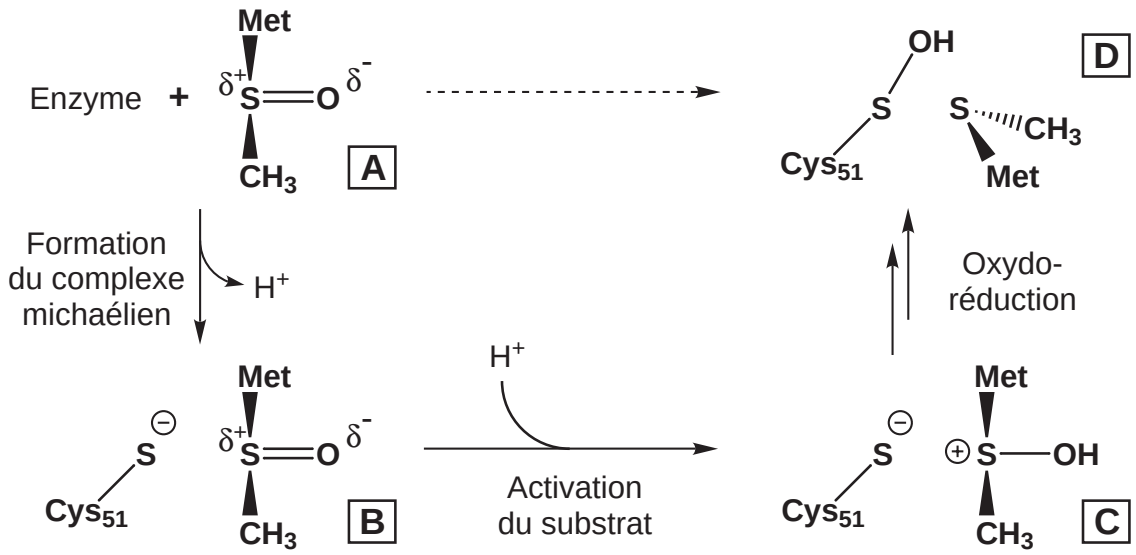


FIG. 7.1: *Mécanisme supposé de l'étape réductase. La déprotonation de la cystéine catalytique est obtenue lors de l'ancrage du substrat dans le site actif de l'enzyme (**A** → **B**). On peut envisager ensuite une activation du substrat par protonation (**B** → **C**) facilitant la réduction du sulfoxyde de méthionine et la formation de l'intermédiaire acide sulfénique **D**.*

En effet, la protonation de la fonction sulfoxyde modifie fortement les propriétés physico-chimiques de ce substrat et notamment sa structure électronique. Pour rappel, la nature de la liaison S – O dans

la fonction sulfoxyde est caractérisée de semi-polaire. Elle est la contribution d'une liaison σ covalente fortement polarisée et d'une liaison π fortement ionique. La protonation entraîne la perte de la liaison π et la diminution de l'ordre de liaison $S-O$. En ce qui concerne la structure géométrique, la longueur de cette liaison augmente ; les données cristallographiques sont en moyenne de 1,492 Å pour le sulfoxyde et 1,524 Å pour le sulfoxyde protoné.

Formellement, la réduction d'une espèce sulfoxyde par un thiol peut être décrite comme le bilan des deux demi-équations électroniques suivantes :

$$
\begin{array}{lll}
 & R_2S{=}O + 2\,e^- + 2\,H^+ & \rightarrow \quad R_2S + H_2O \\
(2H^+, 2e^-) : & RSH + H_2O & \rightarrow \quad RSOH + 2\,e^- + 2\,H^+ \\
\hline
 & R_2S{=}O + RSH & \rightarrow \quad R_2S + RSOH
\end{array}
$$

Cette réaction d'oxydoréduction, médiée par une molécule d'eau qui joue le rôle d'oxydant intermédiaire, correspond au mécanisme décrit dans les travaux de Balta *et al.*[236]. Un mécanisme de réduction mettant en jeu le transfert d'un seul électron peut également être envisagé, l'écriture des deux demi-équations électroniques étant modifiée selon :

$$
\begin{array}{lll}
 & R_2S{=}O + e^- + H^+ & \rightarrow \quad R_2S + HO^\bullet \\
(1H^+, 1e^-) : & RSH + HO^\bullet & \rightarrow \quad RSOH + e^- + H^+ \\
\hline
 & R_2S{=}O + RSH & \rightarrow \quad R_2S + RSOH
\end{array}
$$

La réaction est ici médiée par un radical $HO^\bullet$ jouant le rôle d'agent oxydant de la cystéine. Quel que soit le type de mécanisme considéré, la protonation du substrat sulfoxyde devrait être le facteur déclencheur de la réaction d'oxydoréduction.

Bien que les sulfoxydes soient connus pour avoir une très grande affinité protique en phase gazeuse[280], la protonation du substrat sulfoxyde est quasiment impossible en solution aqueuse compte-tenu de sa valeur de pK_a d'environ -2. Selon le mécanisme proposé par Balta *et al.*[236], la protonation en solution de DMSO est possible grâce à la formation d'un intermédiaire sulfurane, avec création d'une liaison S–S entre l'atome de soufre d'un thiolate et l'atome de soufre du sulfoxyde. Nous allons étudier ensuite comment ce processus peut avoir lieu dans le site actif de MsrA.

7.2. Mécanisme de transfert de proton

Le seul résidu ayant *a priori* des propriétés acido-basiques et ayant un proton facilement mobilisable à proximité du substrat est le résidu Glu94. Nous avons donc fait l'hypothèse d'un transfert de

proton à partir de ce résidu pour arriver au sulfoxyde protoné. Le résidu Cys51 reste sous une forme $Cys^{\ominus}$. Nous allons évaluer à présent l'aspect énergétique de ce processus. Pour cela, nous allons choisir un modèle du site actif et nous allons analyser le mécanisme de la réaction à l'aide des outils de la chimie quantique.

7.2.1. Modèle du site actif

Le modèle de site actif que nous avons utilisé est constitué des résidus susceptibles d'avoir une importance dans le mécanisme catalytique soit par leur participation directe, soit par leur contribution essentielle à la stablisation du système chimique. Nous nous sommes reposé sur les observations des simulations de dynamique moléculaire pour établir la liste des résidus nécessaires :

- Cys51 : cystéine catalytique, représentée par une molécule de méthanethiol ;
- Glu94 : donneur de proton et de liaison hydrogène, représenté par une molécule d'acide acétique ;
- Tyr134 : donneur de liaison hydrogène, représenté par une molécule de phénol ;
- Tyr82 : donneur de liaison hydrogène, représenté par une molécule d'eau ;
- substrat sulfoxyde : représenté par une molécule de diméthylsulfoxyde ;
- H_2O : une molécule explicite d'eau-solvant est incorporée afin de stabiliser les ions présents dans le site actif.

Nous nous sommes limité à un nombre d'atomes relativement petit pour construire ce modèle de site actif, afin de pouvoir explorer différentes possibilités de mécanisme avec des méthodes de chimie quantique. Pour ces raisons, en particulier, les résidus Phe52 et Trp53, formant une interaction de type hydrophobe avec un groupement CH_3 du substrat, ont été volontairement écartés de cette liste.

Nous sommes partis des structures cristallographiques afin de positionner correctement les différents résidus de ce modèle. Les structures PDB 3BQE, 3BQF et 3BQG[229] ont été superposées. Elle correspondent respectivement à la structure de l'enzyme libre, du complexe michaélien avec le substrat sulfoxyde et de l'intermédiaire acide sulfénique en fin d'étape réductase. Il est remarquable de voir que les sites actifs des trois structures se superposent convenablement avec une faible déviation. Le système modèle que nous avons utilisé dans cette étude a été construit à partir de la position moyenne des atomes de cette superposition. La figure 7.2 illustre le modèle et précise la numérotation des atomes qui sera utilisée.

Dans les calculs qui seront présentés ci-dessous, nous avons été amenés à fixer la position de certains atomes lors des phases d'optimisation de géométrie pour assurer la conservation de la

structure globale du site actif. Les atomes gelés sont Cys51-C_β (C_1), Tyr134-C_γ (C_{14}), Glu94-C_γ (C_6) et Tyr82-O_η (O_{27}). Le système quantique porte une charge totale de -1 u.a.

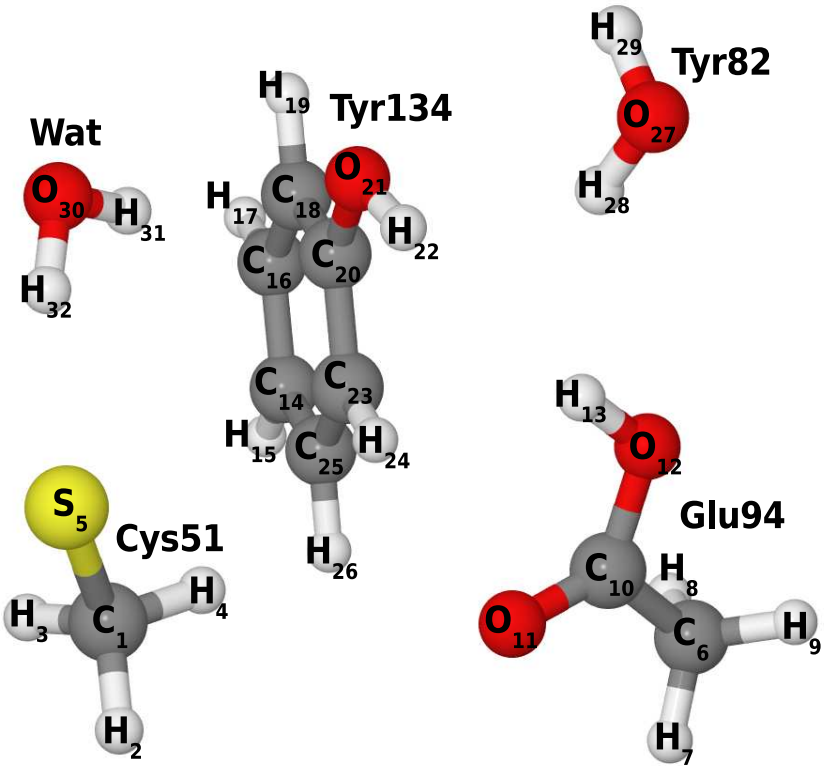


FIG. 7.2: *Structure du modèle quantique de site actif utilisé lors des calculs de chimie quantique. Il est constitué d'une molécule de méthanethiol (atomes 1 à 5) pour Cys51, une molécule d'acide acétique (atomes 6 à 13) pour Glu94, une molécule de phénol (atomes 14 à 26) pour Tyr134, une molécule d'eau (atomes 27 à 29) pour Tyr82 et une molécule d'eau (atomes 30 à 32) correspondant à une molécule de solvant (Wat). L'état de protonation des résidus représentés ici correspond à celui du complexe michaélien. Les atomes dont la position est fixée dans les calculs quantiques sont Cys51-C_β (C_1), Tyr134-C_γ (C_{14}), Glu94-C_γ (C_6) et Tyr82-O_η (O_{27}).*

7.2.2. Méthodologie

Toutes les structures présentées dans cette étude ont été optimisées au niveau B3LYP/6-31G(d), puis évaluées énergétiquement au niveau B3LYP/6-311++G(2df,2p). Cette méthodologie a été choisie afin d'avoir un bon compromis entre précision et temps de calcul.

Toutes les structures ont fait l'objet d'une optimisation de géométrie complète (sauf les quelques atomes gelés) dans le vide et d'un calcul des fréquences. La recherche de l'état de transition pour

le transfert de proton entre Glu94 et le substrat sulfoxyde a été effectuée selon le protocole suivant : (*i*) scan relaxé de la surface d'énergie potentielle selon une variable interne au système, (*ii*) recherche d'une fréquence imaginaire dont le mode normal de vibration correspond au transfert de proton, (*iii*) optimisation de la géométrie de l'état de transition à partir de cette fréquence et (*iv*) calcul des fréquences de l'état de transition optimisé. Le calcul du chemin de réaction intrinsèque (IRC) à partir de l'état de transition vers les réactifs ou les produits n'a pas pu être possible techniquement à cause du gel de la position de certains atomes dans le modèle de site actif. Néanmoins, deux optimisations de géométrie à partir de la structure de l'état de transition ont permis de montrer la connexion de celui-ci avec les réactifs ou les produits.

7.2.3. Chemin de réaction

L'exploration de la surface d'énergie potentielle au moyen du scan relaxé de la distance H–O_{Glu} à partir du complexe michaélien **1** montre que l'espèce sulfoxyde protoné n'est pas formée. Lorsque le proton est transferé sur le sulfoxyde, une liaison S–S avec la cystéine est créée spontanément, conduisant à un intermédiaire sulfurane **2**. Nous avons localisé l'état de transition $\mathbf{TS_{1-2}}$ reliant ces deux structures. Le tableau 7.1 regroupe les valeurs géométriques de ces espèces qui sont représentées dans les figures 7.3(a) et 7.4. Les ordres de liaison et les charges sont reportées dans le tableau 7.2 et le bilan énergétique du chemin de réaction dans le tableau 7.3.

a. Complexe michaélien 1

L'optimisation de géométrie de la molécule de DMSO dans le modèle de site actif a conduit à une structure de complexe michaélien proche de celle décrite dans le chapitre précédent. Nous retrouvons la grande majorité des interactions que nous avons décrit précédemment. Une réseau important de liaisons hydrogène se forme autour de l'oxygène du sulfoxyde avec les résidus Tyr82, Glu94 et Tyr134. Les distances O–H sont respectivement de 1,83 Å pour Tyr134 (O_{38}–H_{22}), 1,61 Å pour Glu94 (O_{38}–H_{13}) et 1,71 Å pour Tyr82 (O_{38}–H_{28}). Cet oxygène O_{38} porte une charge négative relativement importante de -0,54 u.a. Les deux atomes de soufre S_5 et S_{37} portent des charges opposées importantes, respectivement de -0,79 et 0,74 u.a. Il y a donc une forte interaction électrostatique entre ces deux atomes à une distance de 2,97 Å. La molécule d'eau du solvant se positionne à proximité du substrat sulfoxyde et forme un pont entre les résidus modèle Tyr134 et Cys51.

Distance (Å)	**DMSO**	**1**	**TS_{1-2}**	**2**	**DMSOH**
C_1–S_5	—	1,85	1,84	1,84	—
O_{27}–H_{28}	—	0,99	0,99	0,99	—
O_{27}–O_{38}	—	2,80	2,80	2,79	—
C_6–C_{10}	—	1,51	1,53	1,53	—
C_{10}–O_{12}	—	1,33	1,33	1,28	—
C_{10}–O_{11}	—	1,22	1,28	1,25	—
O_{12}–H_{13}	—	1,02	1,25	1,44	—
O_{12}–O_{38}	—	2,62	2,48	2,51	—
C_{20}–O_{21}	—	1,36	1,37	1,37	—
O_{21}–H_{22}	—	1,00	1,00	1,00	—
O_{21}–O_{38}	—	2,71	2,72	2,72	—
O_{21}–H_{31}	—	2,58	2,07	2,05	—
H_{28}–O_{38}	—	1,83	1,83	1,82	—
S_5–S_{37}	—	2,97	2,61	2,58	—
S_5–H_{32}	—	2,39	2,64	2,67	—
H_{22}–O_{38}	—	1,71	1,73	1,73	—
H_{13}–O_{38}	—	1,61	1,10	1,08	0,98
S_{37}–O_{38}	1,51	1,63	1,79	1,82	1,64
S_{37}–C_{33}	1,84	1,83	1,83	1,83	1,81
S_{37}–C_{39}	1,84	1,82	1,83	1,83	1,81
Angle (°)	**DMSO**	**1**	**TS_{1-2}**	**2**	**DMSOH**
C_1–S_5–S_{37}	—	88,1	93,3	93,7	—
S_5–S_{37}–O_{38}	—	166,2	170,5	170,9	—
S_{37}–O_{38}–H_{13}	—	117,5	111,0	110,1	109,5
O_{38}–H_{13}–O_{12}	—	172,0	172,5	172,6	—
C_{10}–O_{12}–H_{13}	—	109,8	112,2	112,2	—
O_{11}–C_{10}–O_{12}	—	124,9	126,2	126,3	—
C_{20}–O_{21}–H_{22}	—	112,1	111,6	111,5	—
O_{21}–H_{22}–O_{38}	—	172,4	169,7	169,7	—
C_{20}–O_{21}–O_{38}	—	116,4	116,2	116,1	—
H_{22}–O_{38}–S_{37}	—	102,0	95,6	95,6	—
C_{33}–S_{37}–O_{38}	107,5	99,4	91,9	91,3	100,4
C_{39}–S_{37}–O_{38}	107,5	110,8	95,1	94,5	100,4
C_{33}–S_{37}–C_{39}	95,8	99,1	100,9	101,1	102,6
O_{27}–H_{28}–O_{38}	—	169,6	165,8	165,5	—
H_{28}–O_{38}–S_{37}	—	119,8	123,4	123,0	—
C_{20}–O_{21}–H_{31}	—	96,9	108,2	109,1	—
O_{21}–H_{31}–O_{30}	—	153,1	168,6	169,1	—
H_{31}–O_{30}–H_{32}	—	104,1	104,6	104,6	—
S_5–H_{32}–O_{30}	—	166,6	158,6	157,4	—
C_1–S_5–H_{32}	—	101,2	99,4	99,1	—
C_1–S_5–H_{30}	—	102,7	103,2	103,2	—
O_{27}–O_{38}–S_{37}	—	118,1	120,8	120,3	—
C_6–C_{10}–O_{11}	—	122,9	119,0	118,7	—
C_6–C_{10}–O_{12}	—	112,2	114,8	115,0	—
C_{10}–O_{12}–O_{38}	—	113,6	115,4	115,3	—

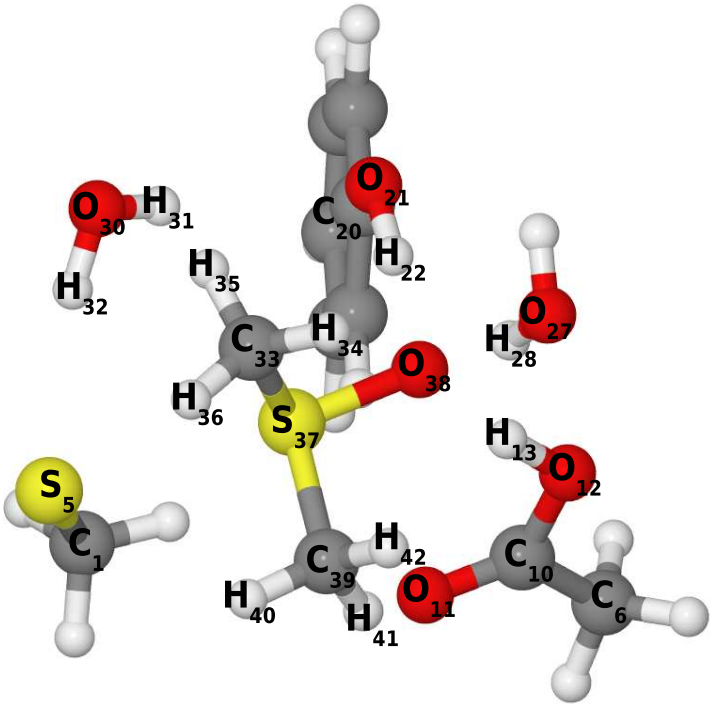
TAB. 7.1: *Valeurs des distances et des angles du complexe michaélien **1**, de l'état de transition **TS_{1-2}** et du sulfurane **2** dans le modèle quantique de site actif. La géométrie est optimisée au niveau B3LYP/6-31G(d). En comparaison, les valeurs pour le DMSO non protoné (**DMSO**) et protoné (**DMSOH**) sont indiquées (isolé, phase gazeuse).*

Ordre de liaison	**DMSO**	**GluH**	**1**	**TS_{1-2}**	**2**	**DMSOH**	**Glu**
C_1–S_5	—	—	1,04	1,03	1,03	—	—
S_5–S_{37}	—	—	0,21	0,46	0,49	—	—
C_6–C_{10}	—	1,00	0,99	0,97	0,96	—	0,92
C_{10}–O_{12}	—	1,04	1,14	1,34	1,36	—	1,49
C_{10}–O_{11}	—	1,77	1,68	1,52	1,51	—	1,48
O_{12}–H_{13}	—	0,72	0,57	0,26	0,23	—	—
H_{13}–O_{38}	—	—	0,13	0,46	0,49	0,69	—
S_{37}–O_{38}	1,28	—	0,88	0,61	0,59	0,93	—
S_{37}–C_{33}	0,90	—	0,97	0,98	0,98	1,00	—
S_{37}–C_{39}	0,90	—	0,97	0,98	0,99	1,00	—
Charge (u.a.)	**DMSO**	**GluH**	**1**	**TS_{1-2}**	**2**	**DMSOH**	**Glu**
C_1	—	—	-0,04	0,01	-0,03	—	—
S_5	—	—	-0,79	-0,64	-0,61	—	—
C_6	—	-0,47	-0,40	-0,42	-0,40	—	-0,29
C_{10}	—	0,73	0,71	0,73	0,73	—	0,74
O_{11}	—	-0,53	-0,55	-0,62	-0,63	—	-0,76
O_{12}	—	-0,56	-0,59	-0,68	-0,69	—	-0,75
H_{13}	—	0,40	0,38	0,40	0,40	0,45	—
C_{20}	—	—	0,52	0,37	0,36	—	—
O_{21}	—	—	-0,57	-0,52	-0,51	—	—
H_{22}	—	—	0,31	0,30	0,29	—	—
O_{27}	—	—	-0,78	-0,78	-0,78	—	—
H_{28}	—	—	0,34	0,37	0,36	—	—
O_{30}	—	—	-0,78	-0,76	-0,76	—	—
H_{31}	—	—	0,34	0,37	0,37	—	—
H_{32}	—	—	0,35	0,33	0,33	—	—
C_{33}	-0,36	—	-0,37	-0,32	-0,30	-0,49	—
S_{37}	0,28	—	0,74	0,59	0,56	0,46	—
O_{38}	-0,42	—	-0,54	-0,57	-0,55	-0,38	—
C_{39}	-0,44	—	-0,44	-0,39	-0,41	-0,50	—

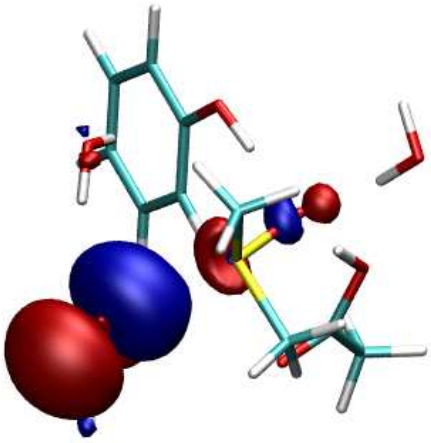
TAB. 7.2: *Valeurs des ordres de liaison et des charges du complexe michaélien **1**, de l'état de transition **TS_{1-2}** et du sulfurane **2** dans le modèle quantique de site actif. La géométrie est optimisée au niveau B3LYP/6-31G(d). Les charges sont dérivées du potentiel électrostatique. En comparaison, les valeurs pour le DMSO non protoné (**DMSO**) et protoné (**DMSOH**) et pour le résidu Glu protoné (**GluH**) et non protoné (**Glu**) sont indiquées (isolé, phase gazeuse).*

ΔE (kcal/mol)	**1**	**TS_{1-2}**	**2**
B3LYP/6-31G(d)	0,0	+ 4,5	+ 4,4
B3LYP/6-311++G(2df,2p)	0,0	+ 5,6	+ 5,6

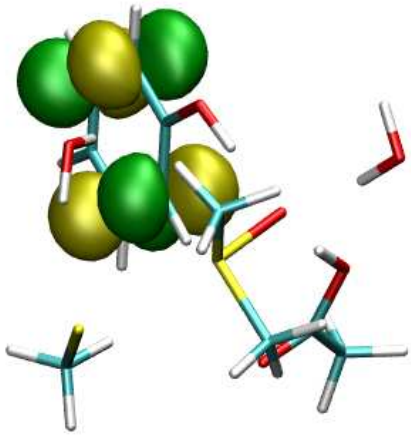
TAB. 7.3: *Bilan énergétique du chemin réactionnel correspondant au transfert de proton entre le résidu Glu94 et le substrat sulfoxyde dans le modèle quantique de site actif. Les espèces ont été optimisées au niveau B3LYP/6-31G(d). Les énergies ont été recalculées au niveau B3LYP/6-311++G(2df,2p). L'énergie de l'espèce **1** est la référence des valeurs d'énergie.*



(a) Complexe michaélien **1**



(b) HOMO : E = -0,04080 Hartree



(c) LUMO : E = 0,09826 Hartree

FIG. 7.3: *Structure du complexe michaélien **1** dans le modèle quantique de site actif après optimisation de géométrie au niveau B3LYP/6-31G(d) (a) et tracé des orbitales HOMO et LUMO ((b) et (c)). Les tracés correspondent à une valeur d'isosurface de 0,05.*

L'ordre de liaison entre S_5 et S_{37} vaut 0,21 et indique une interaction orbitalaire non négligeable entre ces deux atomes, bien qu'ils soient séparés d'une distance d'environ 3 Å. Cette interaction a pour conséquence d'affaiblir de façon importante la liaison entre le soufre et l'oxygène du DMSO. L'ordre de cette liaison vaut 0,88, comparé à la valeur de 1,28 dans le DMSO isolé en phase gazeuse (**DMSO**). Au niveau géométrique, l'allongement de la distance S–O de 1,51 à 1,63 Å, respectivement entre **DMSO** et **1** est significative et confirme l'affaiblissement de la liaison. Par conséquent, les atomes de soufre S_{37} et d'oxygène O_{38} portent une charge relativement élevée, passant de 0,28 à 0,74 u.a. pour S_{37} et de -0,42 à -0,54 u.a. pour O_{38} entre **DMSO** et **1** respectivement. La forte liaison hydrogène entre O_{38} du DMSO et H_{38} porté par Glu94 se traduit par un ordre de liaison non négligeable de 0,13 entre ces deux atomes. La liaison de ce proton avec O_{12} de Glu94 est affaiblie en conséquence, avec un ordre de liaison passant de 0,72 à 0,57 respectivement entre le modèle de glutamate protoné **GluH** et le complexe michaélien **1**.

Une analyse de la fonction d'onde du complexe michaélien **1** dans le modèle de site actif a été réalisée. Le tracé des orbitales HOMO et LUMO est présenté dans la figure 7.3. L'orbitale occupée de plus haute énergie (HOMO, Fig 7.3(b)) est portée principalement par l'atome de soufre S_5, correspondant au résidu Cys51 catalytique, sous la forme d'une orbitale de type *p* dirigée selon l'axe S–O du DMSO. Une petite délocalisation de cette orbitale moléculaire est observée sur les atomes S_{37} et O_{38} de la fonction sulfoxyde du DMSO. L'orbitale vacante de plus basse énergie (LUMO, Fig 7.3(c)) est localisée au niveau du système π de la molécule de phénol imitant le résidu Tyr134.

Nous avons comparé **1** avec la structure cristallographique 3BQE dans laquelle l'atome d'oxygène d'une molécule d'eau du solvant mime l'atome d'oxygène du sulfoxyde. Le tableau 7.4 reporte les distances entre atomes d'oxygène dans le modèle de site actif et dans la structure cristallographique. Nous constatons globalement que les deux structures sont similaires, et plus particulièrement en ce qui concerne la position de l'oxygène du sulfoxyde. Ceci nous permet de valider cette espèce **1** comme point de départ du mécanisme de transfert de proton et, par extension, du modèle de site actif utilisé.

Distance (Å)	O_{38}-O_{21}	O_{38}-O_{12}	O_{38}-O_{27}
Modèle de site actif (**1**)	2,71	2,62	2,80
Cristallographie (3BQE)	2,68	2,63	2,69

TAB. 7.4: *Comparaison de la structure du complexe michaélien **1** dans le modèle quantique de site actif par rapport à la structure cristallographique 3BQE. La géométrie **1** est optimisée au niveau B3LYP/6-31G(d).*

b. Etat de transition pour le transfert de proton $\mathbf{TS_{1-2}}$

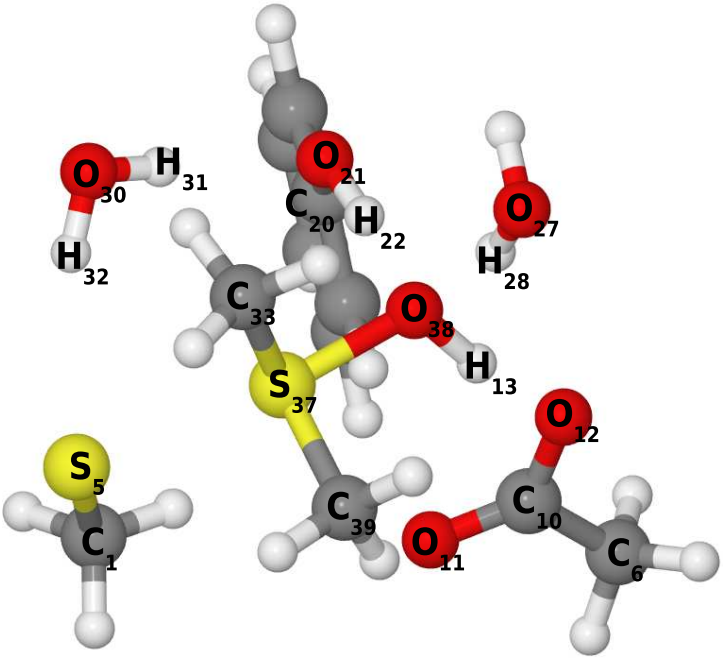
L'état de transition $\mathbf{TS_{1-2}}$ (Fig. 7.4(a)) est caractérisé par une fréquence imaginaire (79**i** cm^{-1}) dont le mode normalde vibration correspond à l'échange de l'hydrogène H_{13} entre les oxygènes O_{12} et O_{38}. Ce mouvement est couplé avec une vibration d'élongation de la liaison S_5–S_{37}. Il montre que la protonation du substrat sulfoxyde se fait de manière concertée avec un rapprochement des deux atomes de soufre S_5 et S_{37}.

Le transfert de proton est pratiquement achevé au niveau de cet état de transition, la longueur de la liaison H_{13}–O_{38} qui se forme étant égale à 1,10 Å. De ce fait, les distances S_5–S_{37} et S_{37}–O_{38} se trouvent fortement modifiées. La première a diminué de 2,97 à 2,61 Å. La seconde s'est allongée, passant de 1,63 à 1,79 Å. Il existe encore le réseau de liaisons hydrogène autour de l'oxygène O_{38} du sulfoxyde avec les résidus Tyr82 et Tyr134. Le système adopte une géométrie de type bipyramide à base trigonale dans laquelle le plan équatorial est formé par les deux atomes de carbone C_{33} et C_{39} et par la paire libre de l'atome de soufre S_{37}. Ceci préfigure la structure de l'intermédiaire sulfurane **2** que nous allons décrire ensuite. Signalons enfin que l'augmentation de l'ordre de liaison de 0,21 à 0,46 entre les atomes S_5 et S_{37}, d'une part, et de 0,13 à 0,46 entre les atomes O_{38} et H_{13}, d'autre part, ainsi que la redistribution de la charge, confirment le mécanisme concerté de formation d'une liaison S_5–S_{37} lors du transfert du proton entre O_{12} et O_{38}.

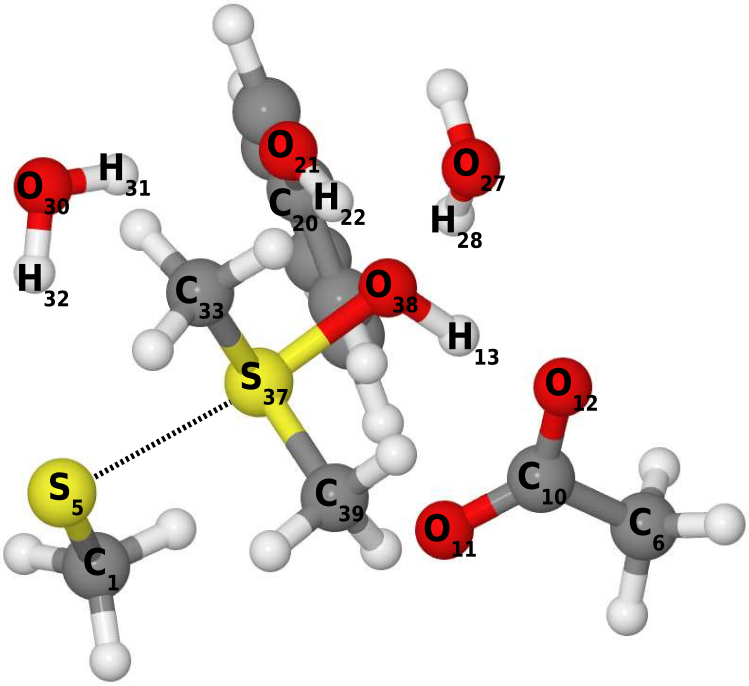
c. Intermédiaire sulfurane 2

L'optimisation de géométrie conduit à l'espèce **2**, illustrée dans la figure 7.4(b). Nous retrouvons donc une structure de type sulfurane proche de celle décrite dans les travaux de Balta *et al.*[236]. Comme nous l'avons dit, la géométrie de **2** est très proche de celle de l'état de transition $\mathbf{TS_{1-2}}$. Nous retrouvons les liaisons hydrogène autour de l'oxygène O_{38} du substrat sulfoxyde avec les résidus modèle Tyr82 et Tyr134 (H_{28}–O_{38} = 1,82 Å, H_{22}–O_{38} = 1,73 Å). Le proton porté par l'oxygène du sulfoxyde crée une liaison hydrogène très forte avec un oxygène de Glu94 (H_{13}–O_{12} = 1,44 Å).

La distance (d_{S-S} = 2,58 Å) entre les deux atomes de soufre est intermédiaire de celles obtenues par Balta *et al.* pour un sulfurane en phase gazeuse (d_{S-S} = 2,41 Å) et en solvant DMSO (d_{S-S} = 2,73 Å). L'ordre de la liaison S_5–S_{37} vaut 0,49, montrant une certaine fragilité de cette liaison (0,63 en phase gazeuse et 0,22 en solvant DMSO dans les travaux de Balta *et al.*). La liaison S–S du sulfurane **2** semble posséder un caractère ionique relativement important. La principale différence de nos résultats par rapport à ceux de Balta *et al.* concerne la liaison S_{37}–O_{38}. La distance obtenue



(a) Etat de transition $\mathbf{TS_{1-2}}$



(b) Intermédiaire sulfurane **2**

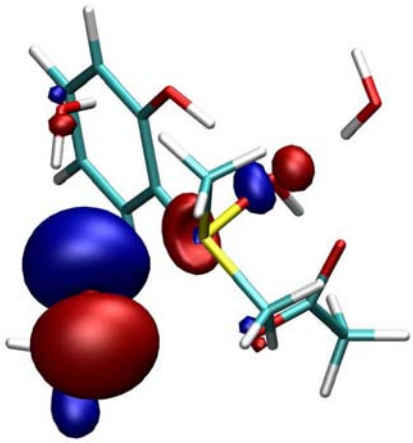
FIG. 7.4: *Structure de l'état de transition* $\boldsymbol{TS_{1-2}}$ *correspondant au transfert de proton entre Glu94 et le DMSO et de l'intermédiaire sulfurane* ***2*** *dans le modèle quantique de site actif après optimisation de géométrie au niveau B3LYP/6-31G(d).*

ici, 1,82 Å, est intermédiaire entre celle du sulfurane en phase gazeuse ($d_{S–O}$ = 1,91 Å) et celle en solution de DMSO ($d_{S–O}$ = 1,72 Å). En revanche, l'ordre de la liaison (0,59) est bien plus faible que la valeur de 0,93 en phase gazeuse et de 1,24 en solution de DMSO. Ceci impliquerait que, à l'instar de la liaison S–S, la liaison S–O du sulfurane présente un caractère ionique important dans le site actif de l'enzyme, probablement comme conséquence de l'établissement d'une interaction forte entre l'atome d'hydrogène H_{13} du sulfurane et l'atome d'oxygène O_{12} de Glu94.

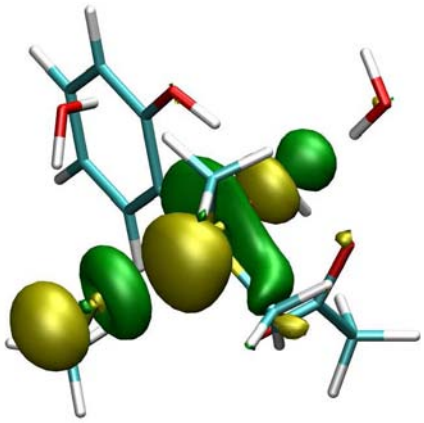
Les commentaires ci-dessus sont également cohérents avec le calcul des charges atomiques. Nous observons une variation de -0,79 à -0,61 u.a. pour S_5, de 0,74 à 0,56 u.a. pour S_{37} et de -0,54 à -0,55 u.a. pour O_{38}, entre **1** et **2** respectivement. De manière générale, l'influence de Cys51 et de Glu94 sur le DMSO a fait diminuer de près d'une unité l'ordre de liaison S–O (de 1,28 pour DMSO à 0,59 pour **2**).

Une analyse de la fonction d'onde de l'intermédiaire sulfurane **2** dans le modèle quantique de site actif a été réalisée. Le tracé des orbitales HOMO et LUMO est présenté dans la figure 7.5. Comme dans **1**, l'orbitale HOMO (Fig 7.5(a)) est portée majoritairement par l'atome de soufre S_5 du résidu Cys51, sous la forme d'une orbitale de type *p* (paire libre). Une petite délocalisation de cette orbitale moléculaire est observée sur les atomes S_{37} et O_{38} provenant de la fonction sulfoxyde. L'orbitale LUMO (Fig. 7.5(b)) se présente comme une combinaison d'orbitales atomiques *p* des atomes S_5, S_{37} et O_{38} de type σ^*. Pour comparaison, nous avons tracé les orbitales HOMO et LUMO de ce même intermédiaire sulfurane **2** (*i*) en supprimant les atomes correspondant au résidu Glu94 (Fig. 7.5(c) et 7.5(d)) et (*ii*) pour une structure de sulfurane optimisée en phase gazeuse (Fig. 7.5(e) et 7.5(f)). La principale différence se situe au niveau de l'orbitale HOMO. En l'absence du résidu Glu94, nous n'observons plus de délocalisation de cette orbitale moléculaire sur les atomes S_{37} et O_{38} de la fonction sulfoxyde. Notons que la suppression d'un ou des deux résidus tyrosine n'a pas d'influence sur la délocalisation de cette orbitale moléculaire (résultats non montrés).

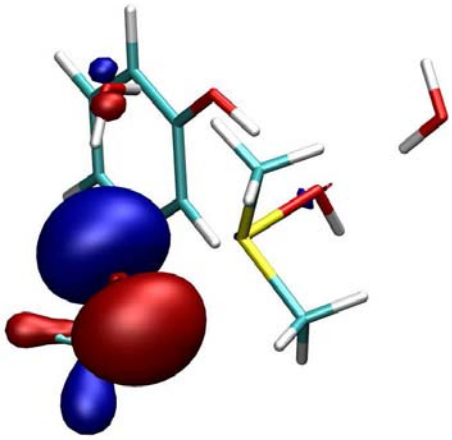
Enfin, la figure 7.6(a) présente l'isosurface de densité électronique (ρ = 0,05) de l'intermédiaire sulfurane **2** dans le modèle de site actif, colorée selon la valeur du potentiel électrostatique (rappelons que la charge nette du système est -1 u.a.). La couleur bleue correspond à un potentiel négatif de -0,2 V et la couleur rouge à un potentiel positif de 0,2 V. Le potentiel négatif est essentiellement concentré au niveau des oxygènes et en partie sur S_5. En fait, nous observons une distribution du potentiel en accord avec les charges ponctuelles calculées. L'atome S_{37} (q = 0,56 u.a.) présente



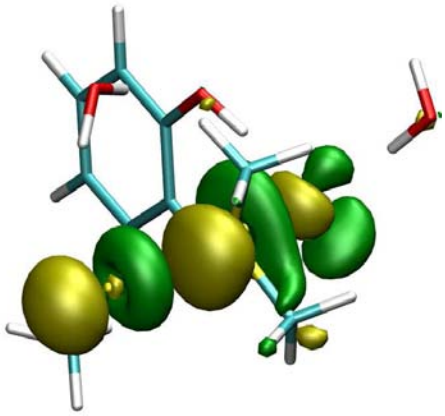
(a) HOMO : E = -0,08409 Hartree



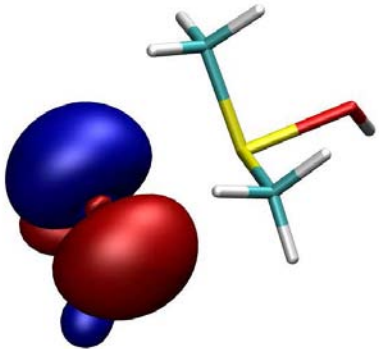
(b) LUMO : E = 0,09127 Hartree



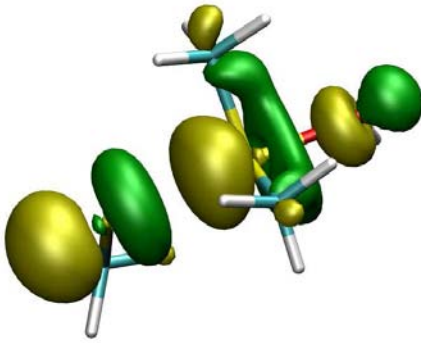
(c) HOMO : E = -0,21501 Hartree



(d) LUMO : E = -0,06205 Hartree

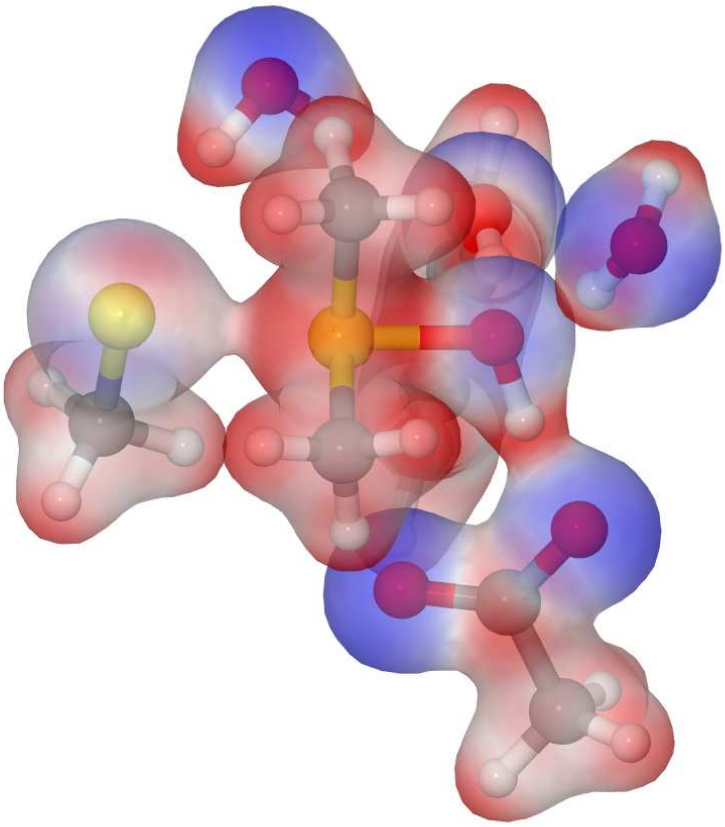


(e) HOMO : E = -0,19774 Hartree

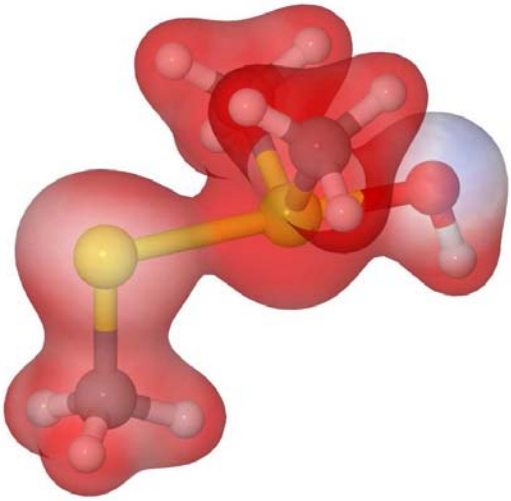


(f) LUMO : E = -0,01425 Hartree

FIG. 7.5: *Tracé des orbitales HOMO et LUMO pour l'intermédiaire sulfurane **2** dans le modèle quantique de site actif ((a) et (b)), pour l'intermédiaire sulfurane **2** dans le modèle quantique de site en absence du résidu Glu94 ((c) et (d)) et pour un composé sulfurane isolé en phase gazeuse ((e) et (f)). Les tracés correspondent à une valeur d'isosurface de 0,05.*



(a)



(b)

FIG. 7.6: *Tracé de l'isosurface de densité électronique (ρ = 0,05) colorée du potentiel électrostatique (bleu : -0,2 V, rouge : 0,2 V) pour l'intermédiaire sulfurane* ***2*** *dans le modèle quantique de site actif (a) et un composé sulfurane isolé en phase gazeuse (b).*

un potentiel positif. L'intensité faible de la couleur bleue au niveau de l'atome S_5, malgré sa charge négative importante (-0,61 u.a.) comparée par exemple à O_{38} (-0,55 u.a.), peut s'expliquer par la différence de taille et la densité électronique plus diffuse sur l'atome de soufre. Nous avons effectué le même tracé pour un composé sulfurane optimisé en phase gazeuse (Fig. 7.6(b)). La différence de charge nette entre le modèle de site actif (-1 u.a.) et la phase gazeuse (0 u.a.) implique une teinte plus rouge de l'ensemble de l'isosurface rendant la comparaison difficile. On peut noter cependant l'augmentation de densité selon la liaison S–S lorsque l'effet des résidus n'est pas pris en compte.

d. Bilan énergétique

Les valeurs d'énergie de chacune des trois espèces que nous venons de présenter sont regroupées dans le tableau 7.3. L'état de transition $\mathbf{TS_{1-2}}$ se situe à 5,6 kcal/mol au dessus du complexe michaélien **1** et nous constatons que l'énergie de l'intermédiaire sulfurane **2** est très proche de celle de l'état de transition (différence inférieure à 0,1 kcal/mol). Ceci est en accord avec le fait que les structures de ces deux espèces sont similaires.

L'énergie d'activation nécessaire à la protonation du substrat est, d'après ces calculs, relativement faible. Ce résultat n'était pas nécessairement attendu compte-tenu des propriétés acido-basiques du sulfoxyde, comparable à un acide fort. Le principal facteur expliquant cette valeur est l'intervention de la cystéine catalytique dans le mécanisme de transfert de proton. D'une certaine manière, la formation de la liaison S–S et le transfert de proton sont deux processus coopératifs et indissociables, l'un impliquant l'autre. Nous avons déjà dit que l'espèce $DMSOH^+$ n'est pas stable en présence de CH_3S^- (formation spontanée de la liaison S–S). De même, la déprotonation du sulfurane conduit à la dissociation hétérolytique de la liaison S–S (résultats non présentés).

L'espèce sulfurane apparaît donc comme un intermédiaire extrêmement labile du mécanisme catalytique. Sa faible stabilité par rapport à l'état de transition $\mathbf{TS_{1-2}}$ ne permet pas d'envisager une mise en évidence expérimentale, la cinétique de retour à l'état initial **1** (complexe michaélien) étant particulièrement rapide. Insistons sur le caractère relativement ionique de ce sulfurane en présence notamment du résidu Glu94 dans le site actif.

Notons enfin que dans ces calculs, nous n'avons pas pris en compte l'effet éventuel de la protéine (au-delà des résidus dans le site actif). Ce point sera évoqué dans le chapitre suivant, lorsque nous ferons une discussion du mécanisme de réduction global.

7.3. Dynamique moléculaire du substrat activé

Nous allons analyser maintenant en détail les différentes interactions du substrat activé dans le site actif par une simulation de dynamique moléculaire. Les calculs quantiques présentés dans la section précédente ont montré que le transfert de proton Glu → DMSO conduit à l'espèce sulfurane avec la formation spontanée d'une liaison S–S entre Cys51 et DMSO. Cette liaison n'est cependant pas très forte et la dissociation peut se produire facilement d'après les résultats en solution de Balta *et al.*[236]. Donc, même si formellement $DMSOH^+$ ne peut pas être considéré comme un intermédiaire réactionnel, il nous a paru interessant d'étudier cette espèce au sein du site actif et de comparer les résultats de dynamique moléculaire avec ceux obtenus pour l'espèce sulfurane. Ces deux simulations sont discutées maintenant.

7.3.1. Simulation de la forme protonée du DMSO

a. Paramètres du champ de forces du substrat DMSO protoné

Une paramétrisation du champ de forces du substrat DMSO protoné a dû être effectuée pour conduire cette étude de dynamique moléculaire car aucun jeu de paramètres spécifiques à cette espèce n'est disponible à l'heure actuelle. Le protocole de paramétrisation est identique à celui décrit dans le chapitre précédent pour le substrat DMSO (voir page 119). Les paramètres du champ de forces du substrat DMSO protoné sont reportés dans le tableau 7.5. La nomenclature des atomes de cette espèce est donnée dans la figure 7.7.

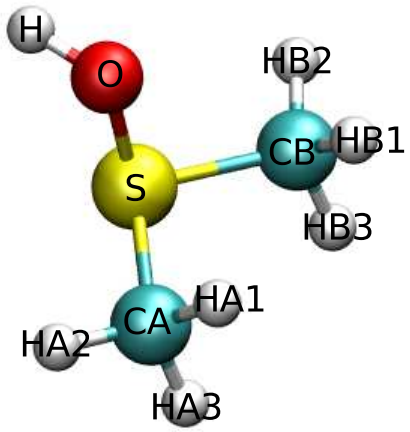


FIG. 7.7: *Nomenclature des atomes du substrat DMSO protoné ($DMSOH^+$). Les carbones sont colorés en cyan, les hydrogènes en blanc, les azotes en bleu, les oxygènes en rouge et les soufres en jaune.*

Atome	C∗	H∗	S	O	H
Type	CT	H1	s4	oh	ho
Masse (a.m.u.)	12,0	1,0	32,1	16,0	1,0
Charge (a.u.)	-0,2335	0,1694	0,4196	-0,4606	0,4918
σ_{LJ} (Å)	1,9080	1,3870	2,000	1,7210	0,0000
ε_{LJ} (kcal/mol)	0,1094	0,0157	0,2500	0,2104	0,0000

Liaison	CT-H1	oh-ho	s4-CT	s4-oh
Constante de force (kcal/mol/Å^2)	340,00	369,60	233,80	274,30
Distance d'équilibre (Å)	1,090	0,974	1,807	1,670

Angle	H1-CT-H1	s4-CT-H1	s4-oh-ho	CT-s4-oh	CT-s4-CT
Constante de force (kcal/mol/deg^2)	35,000	42,900	43,6000	67,300	62,100
Angle d'équilibre (deg)	109,50	108,66	106,31	93,88	96,82

Angle dièdre	H1-CT-s4-oh	CT-s4-oh-ho	H1-CT-s4-CT
Constante de force (kcal/mol)	0,200	5,000	0,200
Phase (rad)	0,00	0,00	0,00
Période	3	1	3

TAB. 7.5: *Paramètres du champ de forces du substrat DMSO protoné : types, charges, masses, paramètres de Lennard-Jones, potentiels de liaison, d'angle de valence et d'angle dièdre en fonction du type de l'atome. C∗ représente C_α et C_β. H∗ représente $H_{\alpha 1}$, $H_{\alpha 2}$, $H_{\alpha 3}$, $H_{\beta 1}$, $H_{\beta 2}$ et $H_{\beta 3}$.*

L'atome de soufre S est représenté par un atome de type s4, l'atome d'oxygène O par un type oh et l'hydrogène H par un type ho du champ de forces *gaff*. Nous avons contraint les carbones, d'une part, et les hydrogènes, d'autre part, à avoir les mêmes charges. Les charges portées par S, O et H sont respectivement de 0,42, -0,46 et 0,49 u.a. Dans le champ de forces *gaff*, l'hydrogène de type ho a des paramètres de Lennard-Jones nuls. Bien que les types d'atomes de *gaff* utilisés ici n'aient pas été optimisés pour décrire la géométrie du DMSO protoné, les distances de liaison et les angles de valence correspondent à ceux de la structure de ce ligand optimisée au niveau quantique.

b. Protocole de simulation

Le protocole de simulation que nous avons suivi pour réaliser cette simulation est identique à celui présenté précédemment :

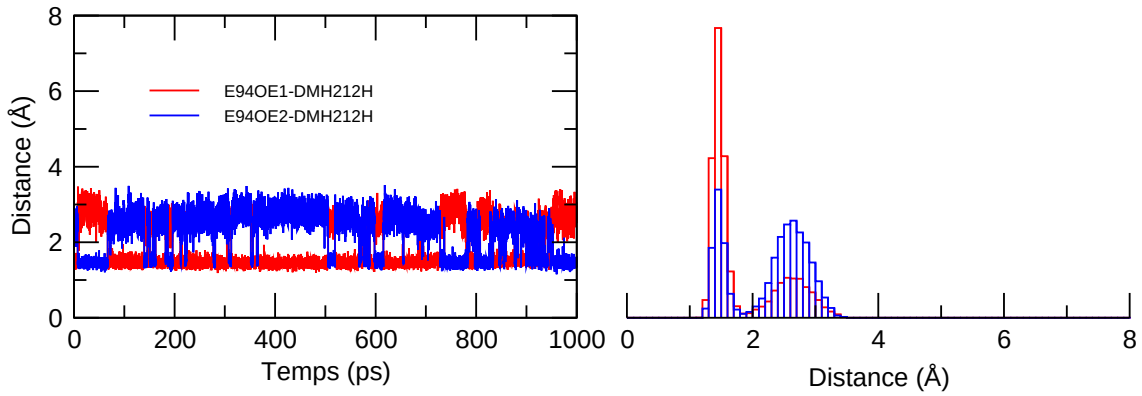
- ajout des hydrogènes de l'enzyme à partir de la structure PDB en accord avec l'état de protonation des résidus ionisables ;
- solvatation de l'enzyme dans une boîte d'eau et minimisations successives pour placer les molécules de solvant ;

- équilibration thermodynamique du système pendant 100 ps avec une montée progressive en température et contrainte sur la chaîne principale de l'enzyme par rapport à la cristallographie ;
- production de la trajectoire dans l'ensemble NPT (P = 1 atm, T = 300 K) sur une durée de 1 ns avec un pas de temps de 0,5 fs.

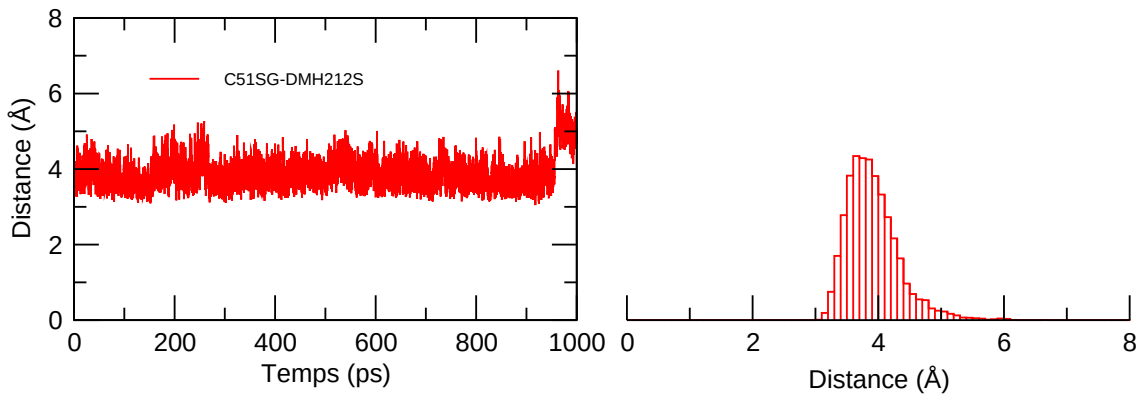
c. Analyse de la trajectoire

Les résultats de la simulation, qui sera identifiée par le sigle DMSOH-Cys$^{\ominus}$-Glu$^{\ominus}$, sont présentés dans la figure 7.8. Le substrat DMSO protoné est désigné dans cette analyse par le terme DMSOH. La plupart des interactions décrites dans le chapitre précédent pour le complexe michaélien sont présentes dans cette simulation, mais elles sont généralement plus faibles. Nous observons une liaison hydrogène à une distance d'environ 2 Å entre l'hydrogène DMSOH-H et les oxygènes $O_{\varepsilon 1}$ et $O_{\varepsilon 2}$ de Glu94 (Fig. 7.8(a)). Cette interaction est alternée entre ces deux atomes tout au long de la simulation. L'extrémité de la chaîne latérale de Glu94 est très mobile, comme le montre la variation de l'angle dièdre autour de la liaison C_{γ}–C_{δ} (Fig. 7.8(d)) sans doute à cause de l'absence de liaisons hydrogène avec les résidus Tyr82 et Tyr134 (Fig. 7.8(e) et 7.8(f)). Les distances entre les hydrogènes H_{η} de Tyr82 et Tyr134 et les oxygènes $O_{\varepsilon 1}$ et $O_{\varepsilon 2}$ de Glu94 sont supérieures à 3 Å. Nous constatons que les résidus Tyr82 et Tyr134 ne forment pas de liaisons hydrogène avec l'oxygène du substrat (Fig. 7.8(g) et 7.8(h)) contrairement à ce qui a été observé pour DMSO. La distance H_{η}–O est supérieure à 4 Å dans les deux cas. Ceci s'explique en partie par le fait que le groupement OH du sulfoxyde protoné possède maintenant un caractère de donneur de liaison hydrogène. Par ailleurs, nous avons observé durant une courte durée de 100 ps, que l'atome d'hydrogène Tyr134-H_{η} vient interagir avec l'atome de soufre Cys51-S_{γ} à une distance d'environ 2 Å (Fig. 7.8(i)).

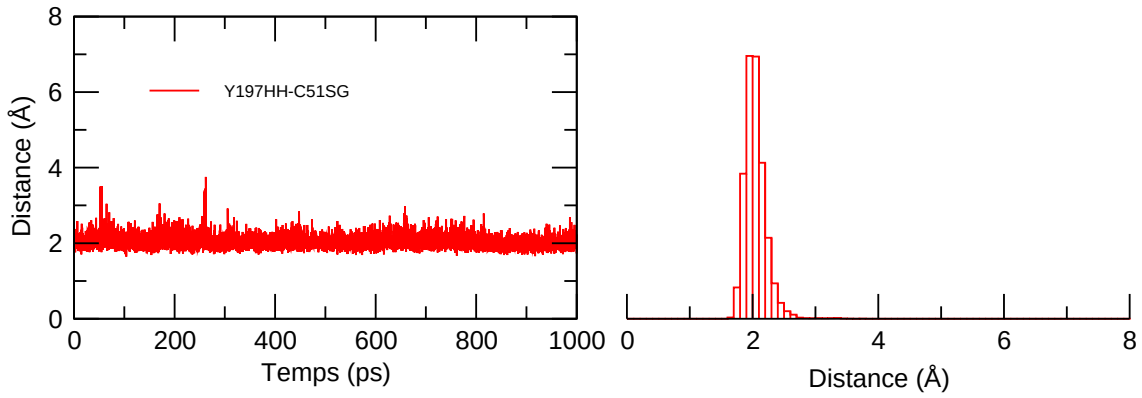
Un point important dans ce cas est l'interaction forte entre les atomes de soufre de Cys51 et du DMSO protoné à une distance moyenne de 3,5 Å (Fig. 7.8(b)). La charge négative portée par Cys51-S_{γ} est toujours stabilisée par l'hydrogène H_{η} de Tyr197 (Fig. 7.8(c)). L'interaction de type hydrophobe du groupement méthyle du substrat avec Trp53 semble peu importante (Fig. 7.8(j)). La distance Trp53-$N_{\varepsilon 1}$ – DMSOH-C_{β} oscille beaucoup autour de la valeur moyenne de 4,5 Å. Cette interaction plutôt faible a pour conséquence d'augmenter la mobilité du substrat dans le site actif. L'angle Cys51-S_{γ} – DMSOH-S – DMSOH-O varie dans des proportions assez importantes, de 100° à 180° (Fig. 7.8(k)). Il en est de même pour l'angle dièdre Cys51-C_{β} – Cys51-S_{γ} – DMSOH-S – DMSOH-C_{β} qui oscille entre 180° et 270° (Fig. 7.8(l)).



(a) Distance Glu94-$O_{\varepsilon 1}$ – DMSOH-H

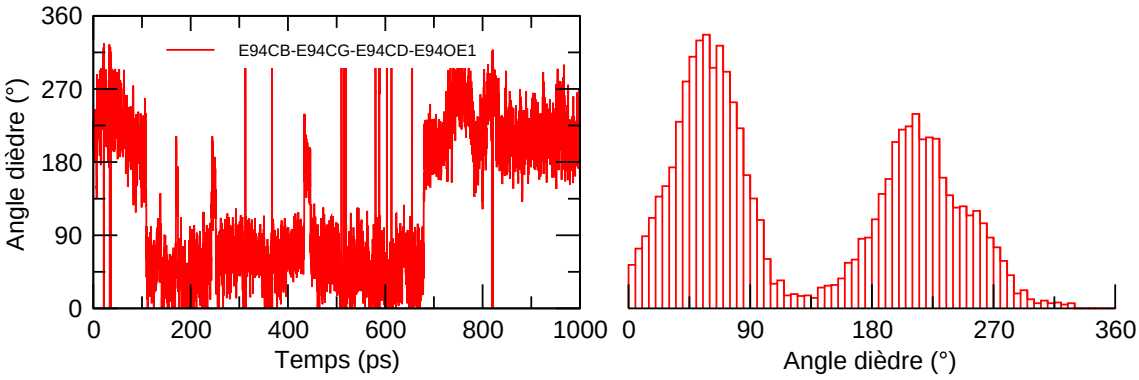


(b) Distance Cys51-S_{γ} – DMSOH-S

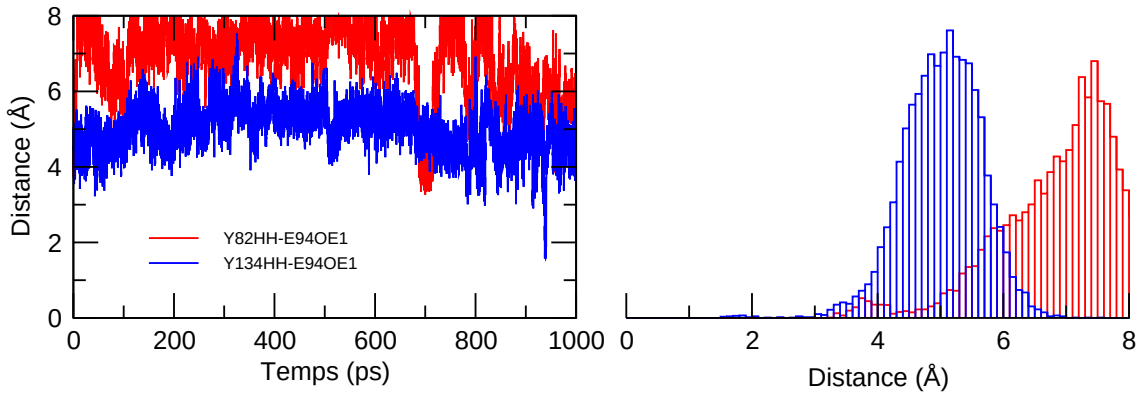


(c) Distance Cys51-S_{γ} – Tyr197-H_{η}

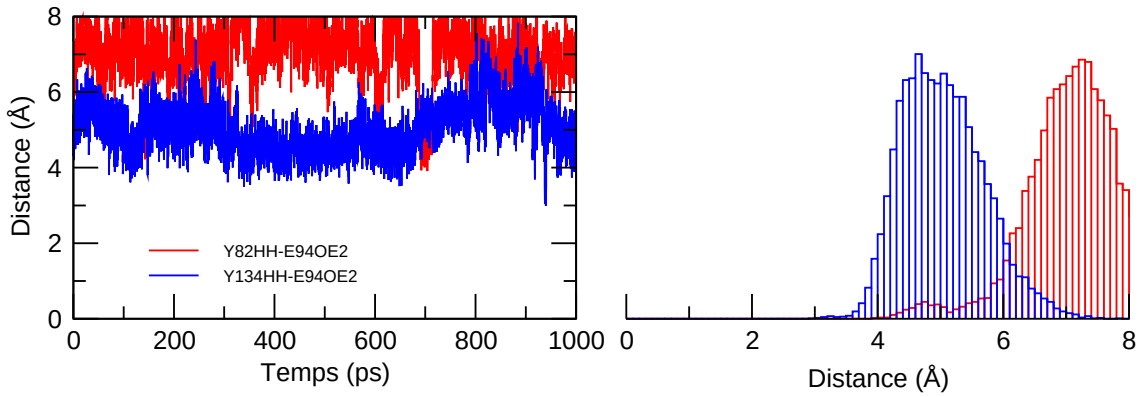
FIG. 7.8: *Variation de la distance Glu94-$O_{\varepsilon 1}$ – DMSOH-H (a), de la distance Cys51-S_{γ} – DMSOH-S (b) et de la distance Cys51-S_{γ} – Tyr197-H_{η} (c) en fonction du temps dans la simulation DMSOH-Cys$^{\ominus}$-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(d) Angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon 1}$

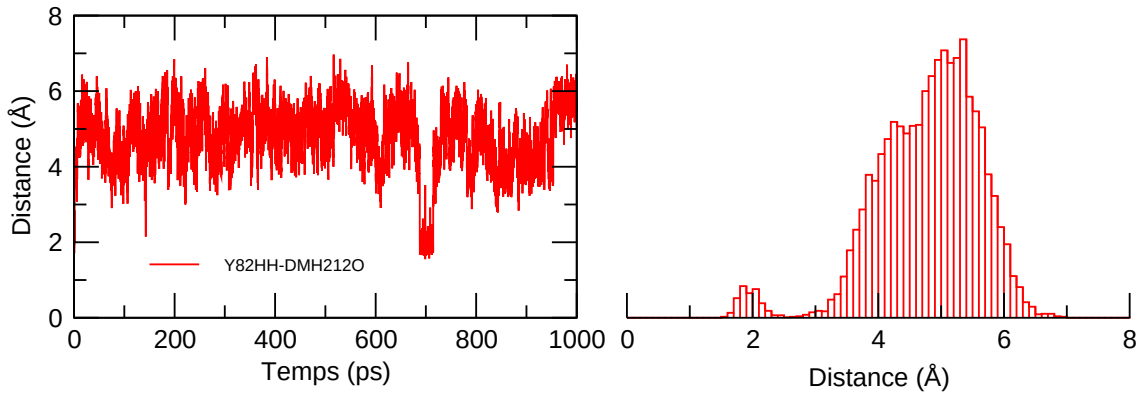


(e) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (bleu)

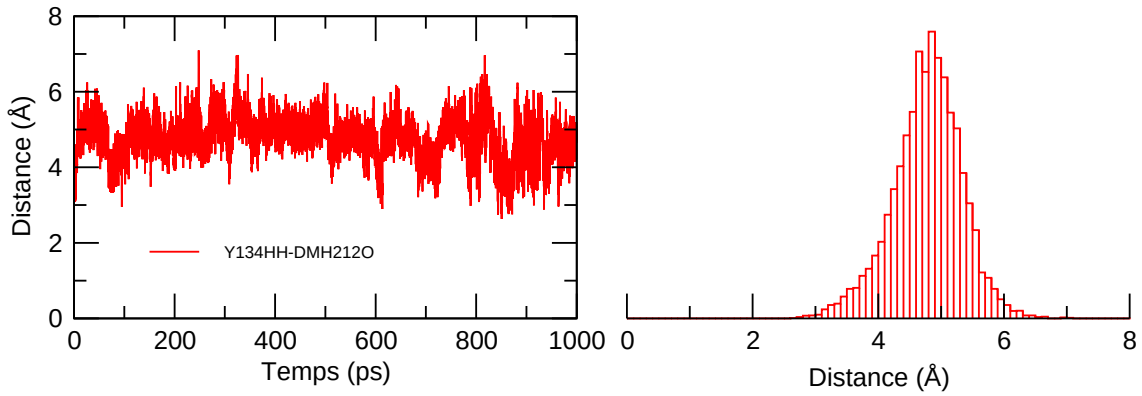


(f) Distance Tyr82-H_η – Glu94-$O_{\varepsilon 2}$ (rouge) et Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ (bleu)

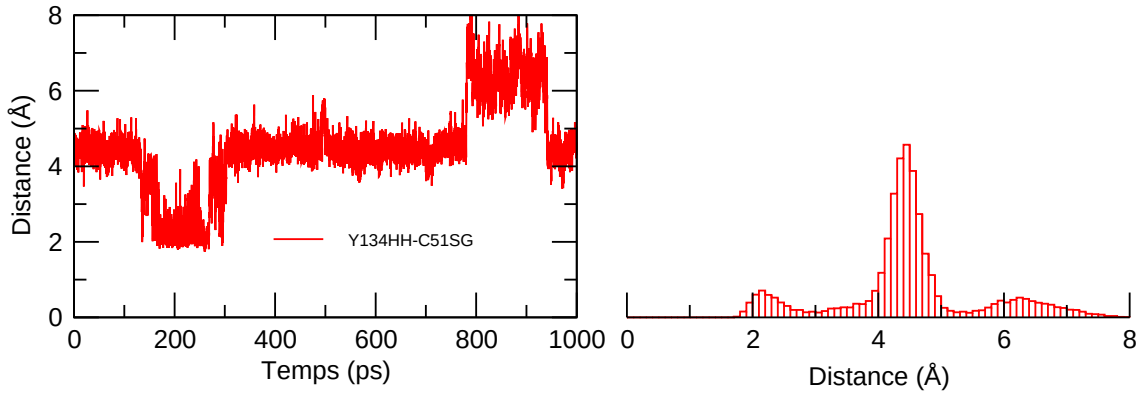
FIG. 7.8: *(suite) Variation de l'angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon 1}$ (d), des distances Tyr82-H_η – Glu94-$O_{\varepsilon 1}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 1}$ (e) et des distances Tyr82-H_η – Glu94-$O_{\varepsilon 2}$ et Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ (f) en fonction du temps dans la simulation DMSOH-Cys$^\ominus$-Glu$^\ominus$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(g) Distance Tyr82-H_η – DMSOH-O

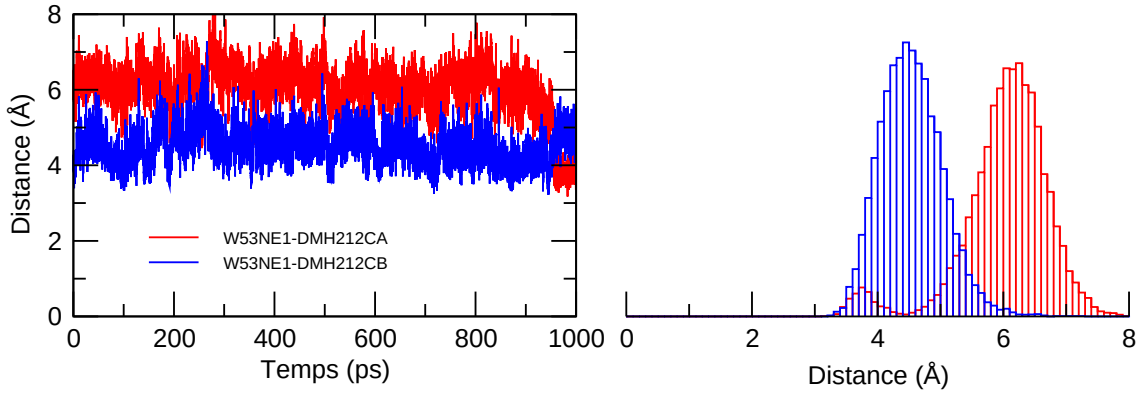


(h) Distance Tyr134-H_η – DMSOH-O

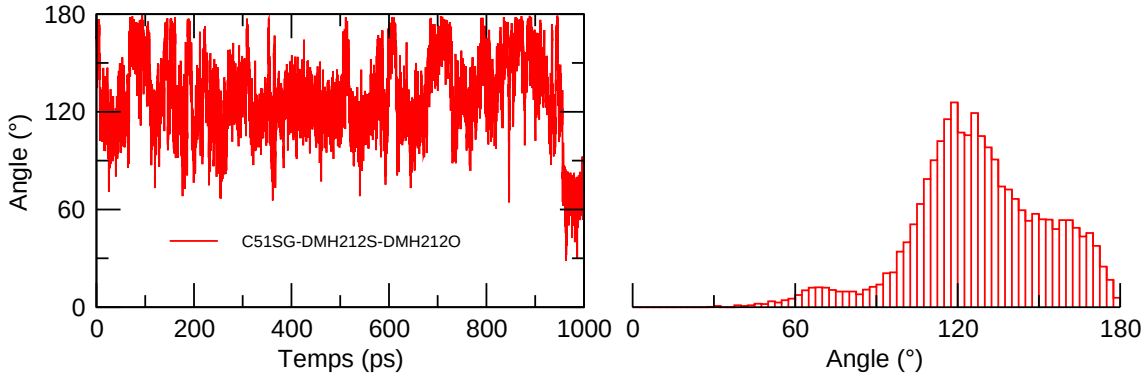


(i) Distance Tyr134-H_η – Cys51-S_γ

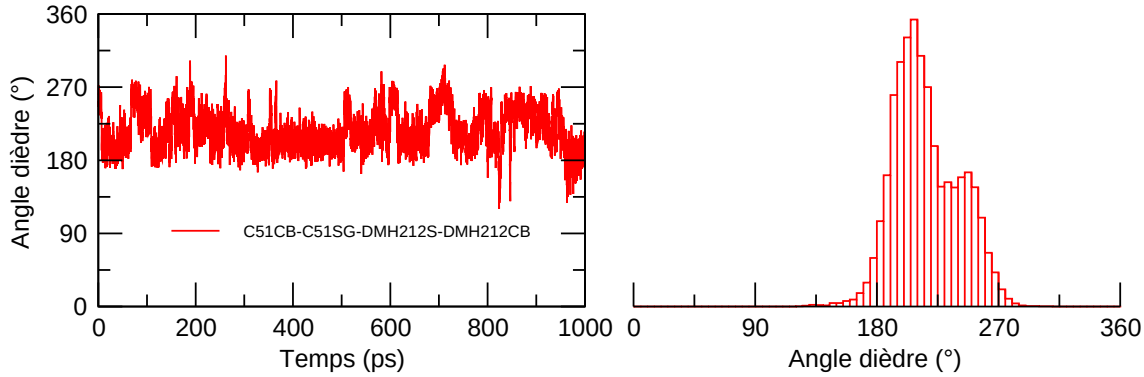
FIG. 7.8: *(suite) Variation de la distance Tyr82-H_η – DMSOH-O (g), de la distance Tyr134-H_η – DMSOH-O (h) et de la distance Tyr134-H_η – Cys51-S_γ (i) en fonction du temps dans la simulation DMSOH-Cys$^{\ominus}$-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(j) Distance Trp53-$N_{\varepsilon 1}$ – DMSOH-C_{α} (rouge) et Trp53-$N_{\varepsilon 1}$ – DMSOH-C_{β} (bleu)



(k) Angle Cys51-S_{γ} – DMSOH-S – DMSOH-O



(l) Angle dièdre Cys51-C_{β} – Cys51-S_{γ} – DMSOH-S – DMSOH-C_{β}

FIG. 7.8: *(suite) Variation des distances Trp53-$N_{\varepsilon 1}$ – DMSOH-C_{α} et Trp53-$N_{\varepsilon 1}$ – DMSOH-C_{β} (j), de l'angle Cys51-S_{γ} – DMSOH-S – DMSOH-O (k) et de l'angle dièdre Cys51-C_{β} – Cys51-S_{γ} – DMSOH-S – DMSOH-C_{β} (l) en fonction du temps dans la simulation DMSOH-Cys$^{\ominus}$-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*

Globalement, nous remarquons donc que le DMSO protoné placé dans le site actif de MsrA est relativement mobile. Les interactions qui permettent de stabiliser le substrat sont plus faibles que dans le cas du DMSO non protoné. C'est essentiellement la liaison hydrogène forte de Glu94 avec le DMSO protoné qui permet de maintenir celui-ci dans le site actif, ainsi que l'interaction électrostatique stabilisante entre les deux atomes de soufre.

7.3.2. Simulation de l'intermédiaire sulfurane dans le site actif

a. Paramètres de champ de forces du résidu sulfurane

Un résidu "sulfurane", appelé Sulf, a été construit. Les paramètres du champ de forces correspondant sont reportés dans les tableaux 7.6 et 7.7. La nomenclature des atomes est donnée dans la figure 7.9. La structure de base de ce résidu est celle d'un résidu cystéine. Lors du calcul des charges RESP, nous avons contraint les charges des atomes de la chaîne principale (N, H, C, O, C_α et H_α) aux valeurs correspondantes à un résidu CYS (*amber94*). Les charges des deux H_β, des deux C_ε et des six H_ε sont contraintes à être identiques. Les charges obtenues pour les atomes S_γ, S_δ et O_ε valent respectivement -0,75, 0,75 et -0,74 u.a. Elles sont relativement importantes, mais l'ordre de grandeur est similaire à celui des charges rapportées dans les travaux de Balta *et al.*[236] et dans notre étude précédente.

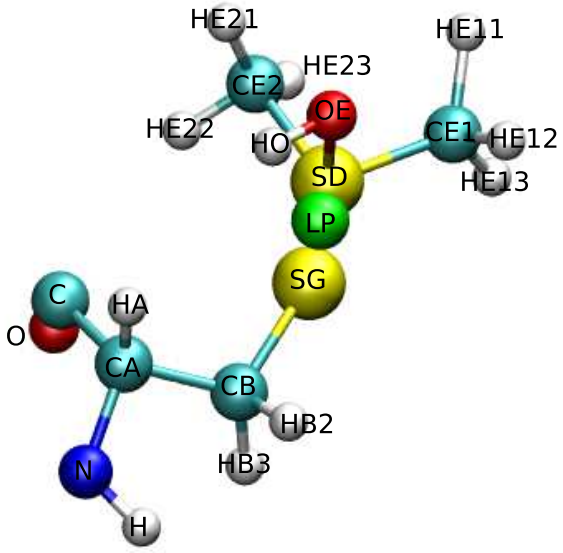


FIG. 7.9: *Nomenclature des atomes du résidu sulfurane (Sulf). Les carbones sont colorés en cyan, les hydrogènes en blanc, les azotes en bleu, les oxygènes en rouge, les soufres en jaune et la paire libre en vert.*

Atome	N	H	C_α	H_α	C_β	$H_\beta*$
Type	N	H	CT	H1	CT	H1
Masse (a.m.u.)	14,0	1,0	12,0	1,0	12,0	1,0
Charge (a.u.)	-0,4157	0,2719	0,0213	0,1124	-0,0336	0,1041
σ_{LJ} (Å)	1,8240	0,6000	1,9080	1,3870	1,9080	1,3870
ε_{LJ} (kcal/mol)	0,1700	0,0157	0,1094	0,0157	0,1094	0,0157
Atome	S_γ	S_δ	$C_\varepsilon*$	$H_\varepsilon*$	O_ε	HO
Type	ss	s6	CT	H1	oh	ho
Masse (a.m.u.)	32,1	32,1	12,0	1,0	16,0	1,0
Charge (a.u.)	-0,7495	0,7470	-0,2128	0,1265	-0,7432	0,4504
σ_{LJ} (Å)	2,0000	2,0000	1,9080	1,3870	1,7210	0,0000
ε_{LJ} (kcal/mol)	0,2500	0,2500	0,1094	0,0157	0,2104	0,0000
Atome	LP	C	O			
Type	LP	C	O			
Masse (a.m.u.)	3,0	12,0	16,0			
Charge (a.u.)	-0,2322	0,5973	-0,5679			
σ_{LJ} (Å)	0,0000	1,9080	1,6612			
ε_{LJ} (kcal/mol)	0,0000	0,0860	0,2100			

Liaison	ce-H1	CT-H1	oh-ho	N-H	C-N
Constante de force (kcal/mol/Å^2)	335,90	340,00	369,60	434,00	490,00
Distance d'équilibre (Å)	1,094	1,090	0,974	1,010	1,335
Liaison	ss-CT	s6-oh	s6-LP	s6-ss	CT-CT
Constante de force (kcal/mol/Å^2)	225,80	190,0	600,00	139,6	310,00
Distance d'équilibre (Å)	1,821	1,812	0,700	2,400	1,526
Liaison	C-O	s6-ce	CT-C	CT-N	
Constante de force (kcal/mol/Å^2)	570,00	254,00	317,00	337,00	
Distance d'équilibre (Å)	1,229	1,774	1,522	1,449	

TAB. 7.6: *Paramètres du champ de forces du résidu sulfurane : types, charges, masses, paramètres de Lennard-Jones et potentiels de liaison en fonction du type de l'atome. $H_\beta*$ représente $H_{\beta 2}$ et $H_{\beta 3}$. $C_\varepsilon*$ représente $C_{\varepsilon 1}$ et $C_{\varepsilon 2}$. $H_\varepsilon*$ représente $H_{\varepsilon 11}$, $H_{\varepsilon 12}$, $H_{\varepsilon 13}$, $H_{\varepsilon 21}$, $H_{\varepsilon 22}$ et $H_{\varepsilon 23}$.*

Angle	C-N-H	H1-ce-H1	s6-ce-H1	s6-oh-ho
Constante de force (kcal/mol/deg^2)	30,000	39,200	43,900	45,500
Angle d'équilibre (deg)	120,00	109,55	108,11	109,20
Angle	H1-CT-CT	H1-CT-C	H-N-CT	N-CT-H
Constante de force (kcal/mol/deg^2)	50,000	50,000	30,000	50,000
Angle d'équilibre (deg)	109,50	109,50	118,04	109,50
Angle	CT-C-N	oh-s6-LP	ce-s6-oh	ce-s6-LP
Constante de force (kcal/mol/deg^2)	70,000	150,000	68,600	150,000
Angle d'équilibre (deg)	116,60	90,00	95,95	130,00
Angle	ss-s6-oh	ss-s6-LP	CT-CT-C	CT-ss-s6
Constante de force (kcal/mol/deg^2)	62,100	150,000	63,000	60,900
Angle d'équilibre (deg)	180,00	90,00	111,10	96,76
Angle	H1-CT-ss	H1-CT-H1	O-C-N	C-N-CT
Constante de force (kcal/mol/deg^2)	42,400	35,000	80,000	50,000
Angle d'équilibre (deg)	109,34	109,50	122,90	121,90
Angle	ce-s6-ce	ss-s6-ce	CT-CT-ss	CT-C-O
Constante de force (kcal/mol/deg^2)	90,000	61,8000	61,100	80,000
Angle d'équilibre (deg)	100,00	90,00	112,69	120,40
Angle	N-CT-CT	N-CT-C		
Constante de force (kcal/mol/deg^2)	80,000	63,000		
Angle d'équilibre (deg)	109,70	110,10		

Angle dièdre	O-C-N-CT	O-C-N-$*$	CT-C-N-$*$	H-N-CT-C
Constante de force (kcal/mol)	10,500	2,000	2,500	1,100
Phase (rad)	π	0,00	π	π
Périodicité	2	1	2	2
Angle dièdre	C-N-CT-CT	C-N-CT-C	$*$-N-CT-$*$	H1-CT-C-O
Constante de force (kcal/mol)	0,530	0,200	0,000	0,800
Phase (rad)	π	π	0,00	0,00
Périodicité	2	1	2	1
Angle dièdre	N-CT-C-N	CT-CT-C-N	$*$-CT-C-$*$	ho-oh-s6-$*$
Constante de force (kcal/mol)	0,750	0,070	0,000	9,500
Phase (rad)	0,00	0,00	0,00	π
Périodicité	1	2	2	1
Angle dièdre	H1-ce-s6-$*$	$*$-CT-CT-$*$	H1-CT-ss-s6	CT-ss-s6-$*$
Constante de force (kcal/mol)	0,144	0,156	0,333	3,067
Phase (rad)	0,00	0,00	0,00	π
Périodicité	3	3	3	2

TAB. 7.7: *Paramètres du champ de forces du résidu sulfurane (suite) : potentiels d'angle de valence et d'angle dièdre en fonction du type de l'atome. Le symbole $*$ représente n'importe quel type d'atome.*

Nous avons modélisé la paire libre du soufre central S_δ par un atome de type LP. Elle est positionnée de manière optimale à une distance de 0,7 Å de l'atome de soufre, comme proposé dans les travaux de Dixon et Kollman[281]. Les autres paramètres du champ de forces pour ce type d'atome sont une charge de -0,23 u.a., une masse de 3,0 a.m.u., des paramètres de Lennard-Jones nuls - comme pour les hydrogènes de groupements hydroxyles - et des constantes de force valant respectivement 600 kcal/mol et 150 kcal/mol pour les potentiels de liaison et d'angle de valence.

Le champ de forces *gaff* ne possède pas les types d'atomes appropriés permettant de décrire une structure de type sulfurane. Nous avons choisi d'utiliser les types d'atome ss et s6 pour décrire respectivement les atomes S_γ et S_δ. Le premier type correspond à un atome en liaison avec un autre soufre (du type disulfure). Le deuxième correspond à un soufre tétravalent à symétrie tétraédrique (du type sulfone). Nous avons modifié les valeurs des distances et des angles d'équilibre de ces deux types d'atome afin de retrouver la structure bipyramidale à base trigonale du sulfurane. Les valeurs utilisées proviennent en majorité des travaux de Balta *et al.*

b. Protocole de simulation

Le protocole de simulation est identique à celui présenté dans le chapitre précédent : (*i*) protonation de l'enzyme, (ii) solvatation de l'enzyme et minimisations successives pour placer les molécules de solvant et (*iii*) équilibration thermodynamique du système. Il s'en suit une phase de production d'une durée de 650 ps dans l'ensemble NPT (P = 1 atm, T = 300 K) avec un pas de temps de 0,5 fs. La trajectoire n'a pas pu être poursuivie au-delà car à partir de 650 ps, nous remarquons une augmentation brutale de l'énergie cinétique du résidu Sulf. L'origine de ce phénomène n'a pas été élucidée. Des tests ont été réalisés en utilisant d'autres paramètres, notamment en y incorporant un potentiel de Lennard-Jones pour les atomes HO et LP, mais sans résultats probants. Une simulation de ce résidu Sulf seul dans une boîte d'eau a reproduit le même phénomène.

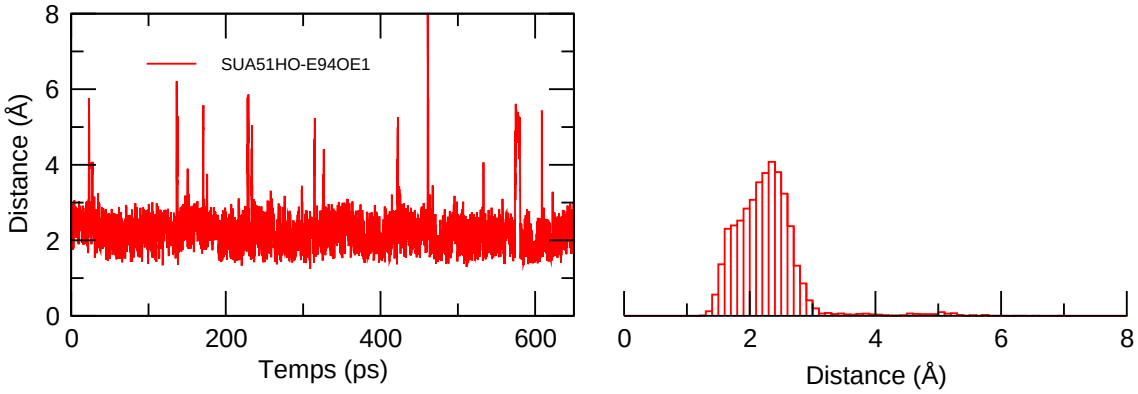
c. Analyse de la trajectoire

Les résultats de la simulation, nommée Sulf-Glu$^{\ominus}$, sont regroupés dans la figure 7.10. Nous observons une interaction par liaison hydrogène de l'hydrogène HO de Sulf avec les oxygènes $O_{\varepsilon 1}$ et $O_{\varepsilon 2}$ de Glu94. Cette interaction est partagée à la fois par les deux atomes (Fig. 7.10(a) et 7.10(b)). L'atome d'hydrogène se trouve à équidistance de 2 Å. L'extrémité de la chaîne latérale de Glu94 est peu mobile comme on peut le constater sur la variation de l'angle dièdre autour de la liaison

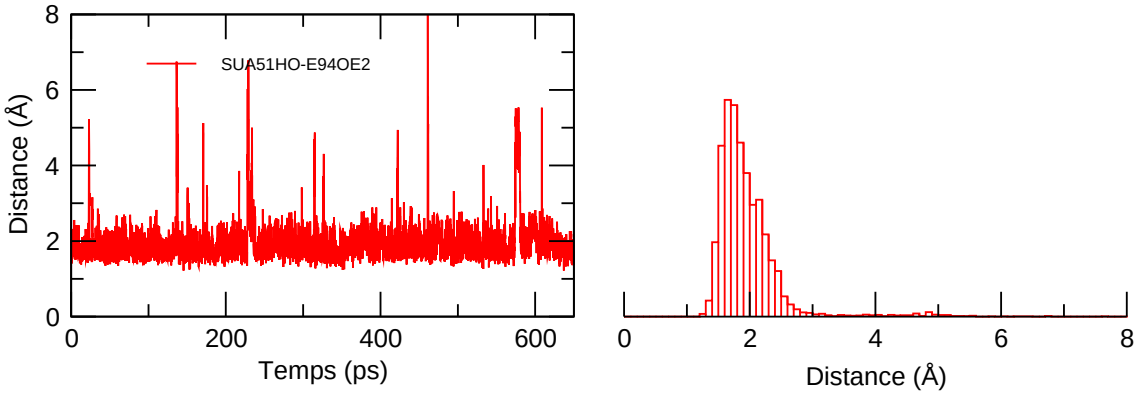
C_γ–C_δ (Fig. 7.10(c)). L'oxygène O_ε du sulfurane forme des liaisons hydrogène fortes avec Tyr82-H_η et Tyr134-H_η (Fig. 7.10(d) et 7.10(e)), à une distance fluctuant principalement autour de 1,8 Å. L'interaction avec Tyr134 est plus faible, avec des fluctuations importantes. Le résidu Tyr134 interagit aussi faiblement avec l'oxygène $O_{\varepsilon 2}$ de Glu94 (Fig. 7.10(i)).

Nous retrouvons l'interaction électrostatique stabilisante entre le soufre S_γ de Sulf et l'hydrogène H_η de Tyr197 avec une distance majoritairement à 2 Å (Fig. 7.10(f)). Un des groupements méthyle du sulfurane est en interaction hydrophobe avec Trp53. La valeur moyenne de la distance Sulf-$C_{\varepsilon 2}$ – Trp53-$N_{\varepsilon 1}$ vaut 3,5 Å (Fig. 7.10(g)). Cette interaction permet au sulfurane de rester peu mobile dans le site actif de l'enzyme. Ceci est vérifié par l'angle dièdre Sulf-C_β – Sulf-S_γ – Sulf-S_δ – Sulf-$C_{\varepsilon 1}$ qui reste relativement constant (Fig. 7.10(h)). Signalons enfin que nous n'avons pas observé d'interaction particulière des résidus du site actif, ni avec le soufre S_δ, ni avec sa paire libre.

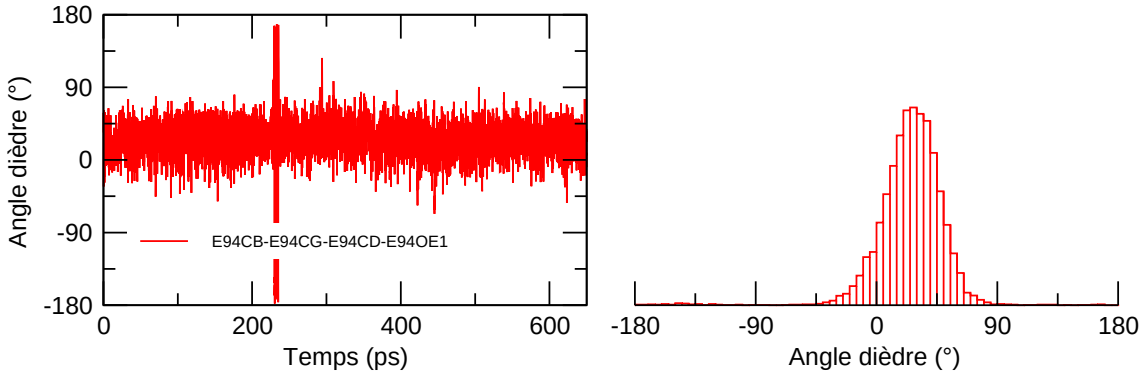
Si nous mettons de côté les problèmes de simulation en fin de trajectoire probablement liés à un souci de paramétrisation qui devra être améliorée, nous remarquons que le site actif de l'enzyme est parfaitement capable de stabiliser le sulfurane. Nous retrouvons la plupart des interactions que nous avons décrites pour le complexe michaélien.



(a) Distance Sulf-HO – Glu94-$O_{\varepsilon 1}$

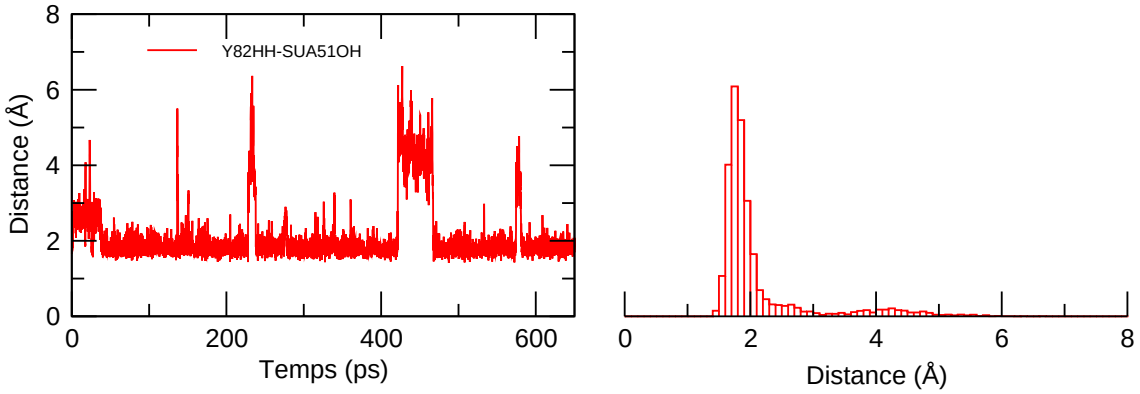


(b) Distance Sulf-HO – Glu94-$O_{\varepsilon 2}$

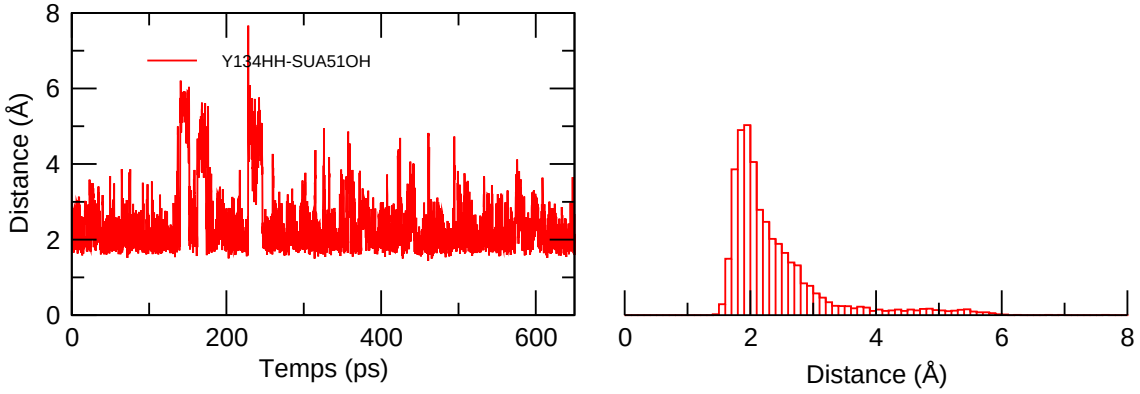


(c) Angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon 1}$

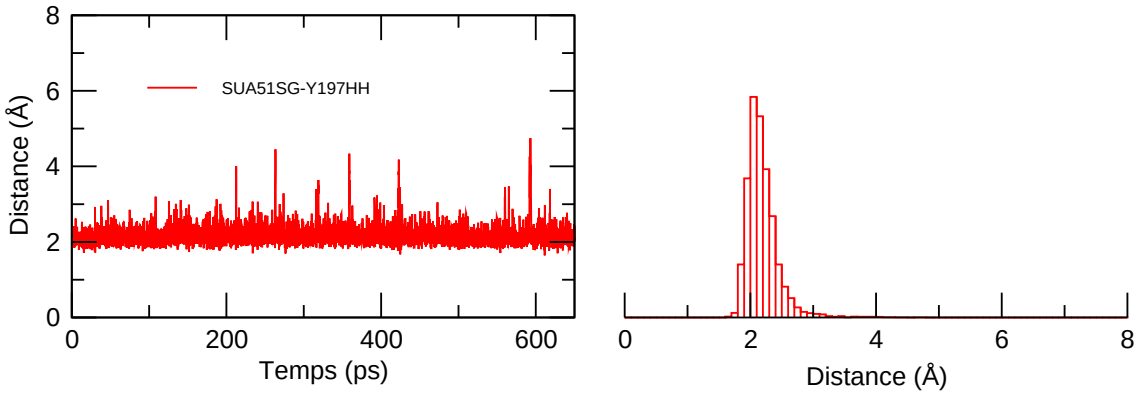
FIG. 7.10: *Variation de la distance Sulf-HO – Glu94-$O_{\varepsilon 1}$ (a), de la distance Sulf-HO – Glu94-$O_{\varepsilon 2}$ (b) et de l'angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon 1}$ (c) en fonction du temps dans la simulation Sulf-Glu$^\ominus$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(d) Distance Sulf-O_ε – Tyr82-H_η

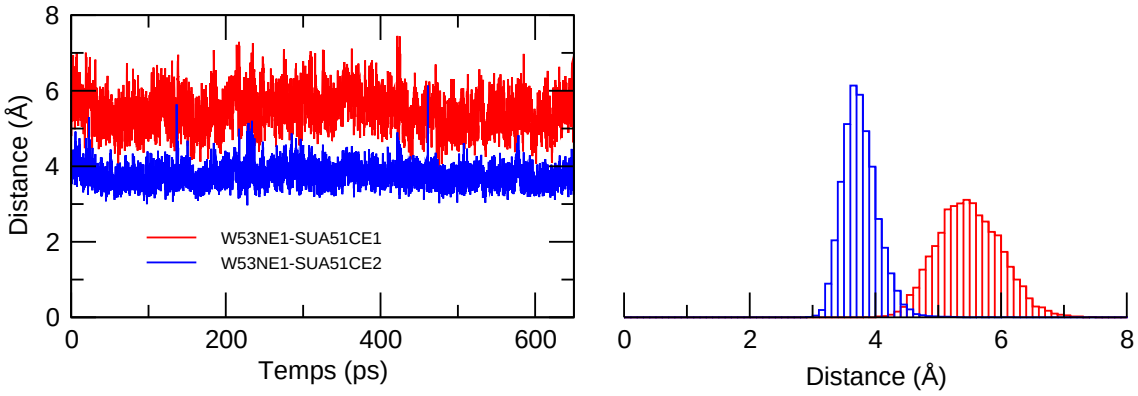


(e) Distance Sulf-O_ε – Tyr134-H_η

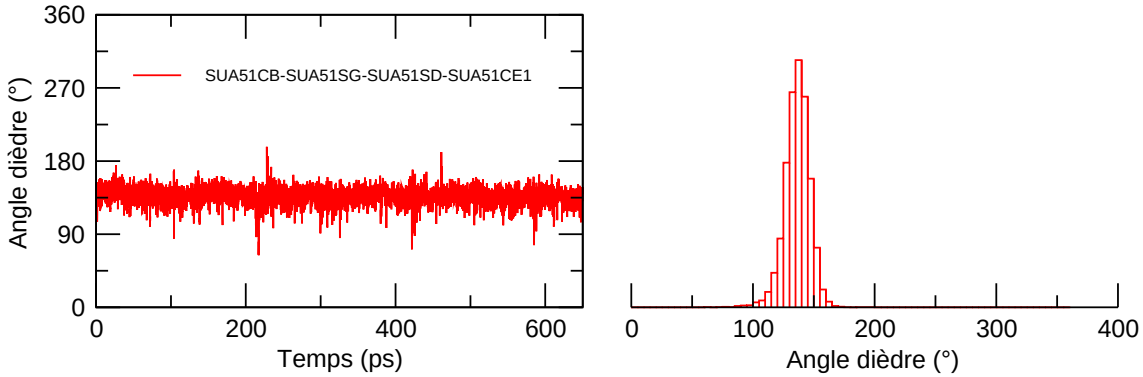


(f) Distance Sulf-S_γ – Tyr197-H_η

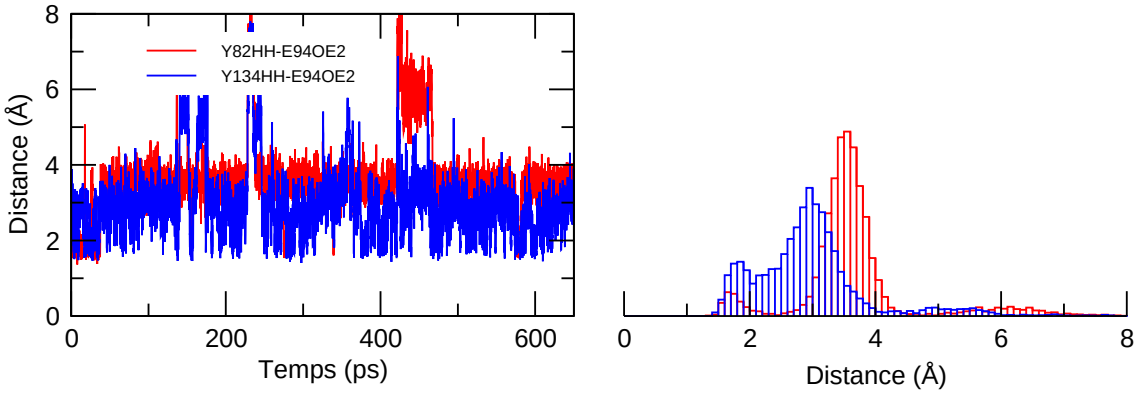
FIG. 7.10: *(suite) Variation de la distance Sulf-O_ε – Tyr82-H_η (d), de la distance Sulf-O_ε – Tyr134-H_η (e) et de la distance Sulf-S_γ – Tyr197-H_η (f) en fonction du temps dans la simulation Sulf-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*



(g) Distance Trp53-$N_{\varepsilon 1}$ – Sulf-$C_{\varepsilon 1}$ (rouge) et Trp53-$N_{\varepsilon 1}$ – Sulf-$C_{\varepsilon 2}$ (bleu)



(h) Angle dièdre Sulf-C_{β} – Sulf-S_{γ} – Sulf-S_{δ} – Sulf-$C_{\varepsilon 1}$



(i) Distance Tyr82-H_{η} – Glu94-$O_{\varepsilon 2}$ (rouge) et Tyr134-H_{η} – Glu94-$O_{\varepsilon 2}$ (bleu)

FIG. 7.10: *(suite) Variation des distances Trp53-$N_{\varepsilon 1}$ – Sulf-$C_{\varepsilon 1}$ et Trp53-$N_{\varepsilon 1}$ – Sulf-$C_{\varepsilon 2}$ (g), de l'angle dièdre Sulf-C_{β} – Sulf-S_{γ} – Sulf-S_{δ} – Sulf-$C_{\varepsilon 1}$ (h) et des distances Tyr82-H_{η} – Glu94-$O_{\varepsilon 2}$ et Tyr134-H_{η} – Glu94-$O_{\varepsilon 2}$ (i) en fonction du temps dans la simulation Sulf-Glu$^{\ominus}$. Les histogrammes à droite montrent la répartition des valeurs pour chacune des courbes.*

7.4. Conclusions

Cette étude a permis de caractériser l'étape d'activation du substrat sulfoxyde après son ancrage dans le site actif de l'enzyme. L'hypothèse d'une protonation du substrat semble la plus raisonnable mais nous avons constaté que, au sein du site actif, l'espèce $DMSOH^+$ n'est pas stable et a tendance à former une liaison S–S à caractère partiellement covalent avec la cystéine catalytique.

L'identification de l'intermédiaire sulfurane comme forme protonée du substrat nous amène à modifier légèrement le mécanisme d'activation présenté en début de l'étude. Dans le mécanisme, illustré par la figure 7.11, la forme cystéine déprotonée – sulfoxyde protoné est remplacée par l'intermédiaire sulfurane **C**. Les calculs de Balta *et al.*[236] ont cependant montré que la dissociation d'un sulfurane selon la liaison S–S est tout à fait envisageable en solution et on peut penser qu'il en sera de même dans l'enzyme, même si, formellement, nous n'avons pas trouvé un minimum d'énergie associé à cette configuration (la présence de $Glu^{\ominus}$ en étant probablement la cause). Nous avons donc réalisé deux simulations de dynamique moléculaire pour le substrat activé, sous forme sulfoxyde protoné ($DMSOH\text{-}Cys^{\ominus}\text{-}Glu^{\ominus}$) et sous forme sulfurane ($Sulf\text{-}Glu^{\ominus}$).

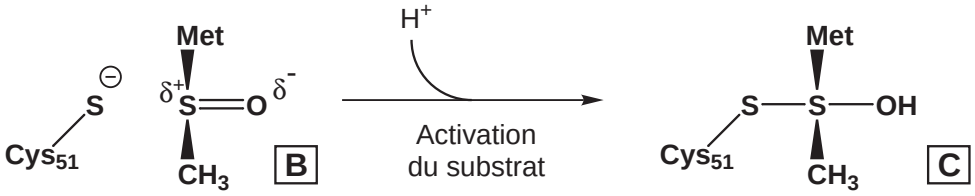


FIG. 7.11: *Mécanisme d'activation de l'étape réductase du mécanisme catalytique. L'activation du substrat (* **B** → **C** *) conduit à la formation de l'intermédiaire sulfurane.*

Les résultats de la simulation pour $DMSOH\text{-}Cys^{\ominus}\text{-}Glu^{\ominus}$ montrent que le sulfoxyde protoné est relativement mobile et forme un faible nombre d'interactions avec les résidus du site actif. La simulation de l'intermédiaire sulfurane ($Sulf\text{-}Glu^{\ominus}$) indique en revanche que celui-ci est relativement bien stabilisé par les résidus du site actif. Il est peu mobile et sa structure est proche de celle du complexe michaélien présenté dans le chapitre précédent.

Chapitre 8

Réduction du substrat

Sommaire

8.1. Mécanisme I : Transfert d'un groupement OH 190
8.2. Mécanisme II : 1 H^+ + 1 e^- 200
8.3. Mécanisme III : 2 H^+ + 2 e^- 209
8.4. Devenir du substrat méthionine réduit 216
8.5. Conclusions 216

Dans ce chapitre, nous explorons un certain nombre de mécanismes réactionnels permettant d'obtenir l'intermédiaire acide sulfénique et libérer le substrat sulfoxyde réduit. Cette partie du mécanisme prend place après l'étape d'activation par protonation de la fonction sulfoxyde, tel que nous venons de le voir dans le chapitre précédent. Nous examinerons trois types de mécanismes possibles. Le premier peut être représenté par un transfert d'un groupement OH, de manière similaire à ce qui a été proposé pour la réaction d'un thiolate avec les peroxydes. En second lieu, nous étudierons une réaction de transfert d'électron qui globalement peut donc être représentée par un mécanisme radicalaire à un proton et un électron ($1H^+$, $1e^-$). Enfin, un mécanisme impliquant un deuxième proton et un deuxième électron ($2H^+$, $2e^-$) sera considéré.

Les différentes études présentées ont été réalisées à l'aide des méthodes de la chimie quantique et en utilisant plusieurs modèles permettant de rendre compte de l'influence de l'environnement enzymatique sur les chemins réactionnels.

8.1. Mécanisme I : Transfert d'un groupement OH

8.1.1. Hypothèse de travail

Ce mécanisme s'inspire des études précédentes sur la réaction de transfert d'un groupement OH d'un peroxyde vers un thiolate[282, 283]. Il a été décrit également par Balta *et al.*[236] pour la réduction du DMSO par un thiol. Notre point de départ ici sera l'espèce sulfurane décrite dans le chapitre précédent. En effet, un transfert du groupement OH à partir du cation $MetSOH^+$ vers $Cys^{\ominus}$ ne paraît pas possible. Ce mécanisme, illustré dans la figure 8.1, implique un réarrangement géométrique préalable de l'intermédiaire sulfurane (pseudorotation) dans lequel l'atome de soufre de Cys51 passerait en position équatoriale. Ensuite, la principale coordonnée de réaction serait l'angle S–S–O qui aurait une valeur d'environ 60° au niveau de l'état de transition.

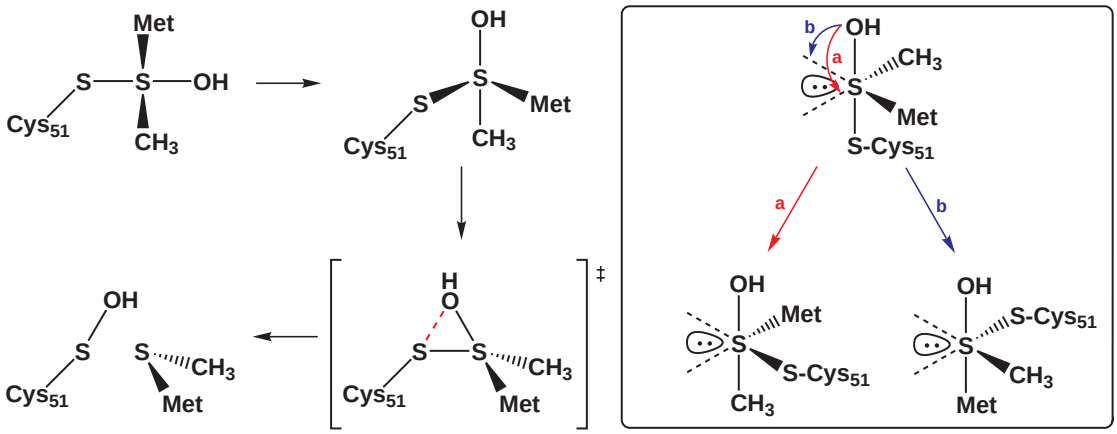


FIG. 8.1: *Hypothèse de travail envisagée pour la formation de l'acide sulfénique à partir de l'intermédiaire sulfurane dans le site actif de l'enzyme (Mécanisme I : Transfert d'un groupement OH). Le mécanisme implique une pseudorotation au niveau de l'atome de soufre tétravalent de l'intermédiaire sulfurane pour placer la liaison S–S en position équatoriale et un repliement de l'angle de liaison S–S–O jusqu'à un état de transition permettant la formation de la liaison S–O de l'acide sulfénique. La pseudorotation conduit à deux stéréoisomères : le groupement OH est positionné sur l'axe S–CH_3 (**a**) ou sur l'axe S–Met (**b**).*

8.1.2. Chemin de réaction dans un système isolé

Nous avons considéré, dans un premier temps, le système modèle illustré par la figure 8.2(a). La chaîne latérale de la cystéine catalytique est simplement représentée ici par un groupement CH_3 (atomes 1 à 5). Le modèle avait déjà été utilisé par Balta *et al.*[236] en solution de DMSO à un niveau de calcul poussé. Nous le décrirons ici à un niveau de théorie plus simple (B3LYP/6-31G(d)) mais de façon plus détaillée. Nous verrons que ce niveau permet de reproduire l'essentiel des résultats de Balta *et al.*

Les différentes espèces présentes sur le chemin réactionnel ont fait l'objet d'une optimisation de géométrie complète dans le vide et d'un calcul de fréquences. La structure des états de transition a été optimisée après une analyse de la surface d'énergie potentielle (scan selon une coordonnée interne correspondant à un angle de liaison). Nous avons vérifié l'existence d'une seule fréquence imaginaire dont le mode normal de vibration correspond à ce repliement. Un calcul du chemin de réaction intrinsèque (IRC) a permis de vérifier la nature des intermédiaires connectés par l'état de transition. Les structures obtenues dans ce chemin de réaction sont décrites ensuite. Nous utiliserons la notation suivante :

A : Intermédiaire sulfurane avec le groupement S-CH_3 en position axiale

B : Intermédiaire sulfurane avec le groupement S-CH_3 en position équatoriale

C : Intermédiaire acide sulfénique

Ils sont reliés par deux états de transition notés $\mathbf{TS_{A\text{-}B}}$ et $\mathbf{TS_{B\text{-}C}}$. Il faut noter ici que, compte tenu de notre modèle (Met = CH_3 dans la figure 8.1), les deux structures possibles pour l'intermédiaire **B** sont énergétiquement équivalentes (énantiomères). Seule une de ces formes (**a**, figure 8.1) sera donc considérée dans les calculs. Le tableau 8.1 regroupe les valeurs géométriques de ces espèces ainsi que les ordres de liaison et les charges. Le bilan énergétique y est également présenté. Les valeurs correspondant à l'intermédiaire sulfurane **2** dans le modèle quantique de site actif utilisé dans le chapitre précédent ont été reportées dans ce tableau pour comparaison. Les structures sont rassemblées dans la figure 8.2.

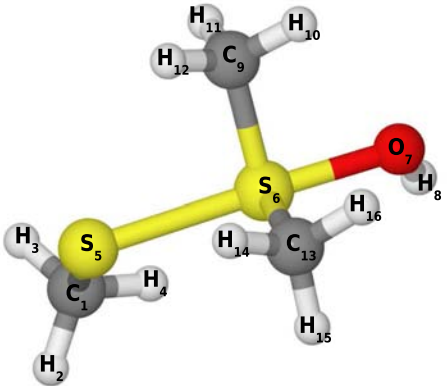
a. Intermédiaire sulfurane A

L'intermédiaire sulfurane **A** (Fig. 8.2(a)) présente une structure similaire à celle que nous obtenons en fin d'étape de protonation dans le modèle de site actif (**2**, fig. 7.4(b)). L'atome de soufre central S_6 adopte une géométrie de type bipyramide à base trigonale où les atomes de soufre S_5 et d'oxygène O_7 (les atomes les plus électronégatifs) sont situés sur l'axe de la bipyramide. Le plan équatorial est occupé par les deux groupements méthyles (C_9 et C_{13}) et la paire libre du soufre S_6.

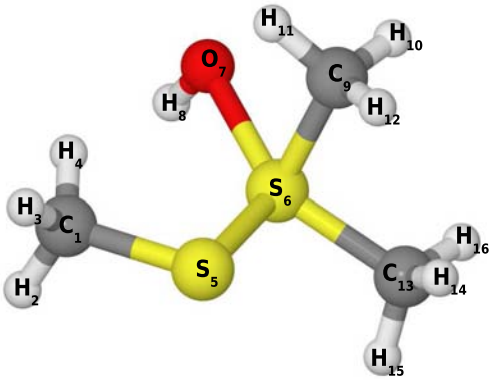
La distance entre les deux atomes de soufre S_5–S_6 vaut 2,43 Å. Elle est plus courte de 0,15 Å que son analogue dans le modèle de site actif. L'absence d'un environnement (enzyme ou solvant) implique une augmentation du caractère covalent de cette liaison, comme le montre les ordres de liaison S–S avec respectivement 0,49 pour **2** et 0,61 pour **A**. Nous observons aussi que la distance S_6–O_7 passe de 1,82 Å pour le modèle de site actif à 1,87 Å pour le vide. Ces valeurs sont en accord avec celles obtenues dans les travaux de Balta *et al.* Les charges des atomes S_5, S_6 et O_7 valent

Distance (Å)	**2**	**A**	**TS_{A-B}**	**B**	**TS_{B-C}**	**C**
C_1–S_5	1,84	1,83	1,84	1,81	1,84	1,82
S_5–S_6	2,58	2,43	2,26	2,18	2,27	4,33
S_6–O_7	1,82	1,87	2,03	2,07	2,20	3,25
O_7–H_8	1,08	0,97	0,97	0,97	0,97	0,98
S_6–C_9	1,83	1,83	1,84	1,83	1,82	1,83
S_6–C_{13}	1,83	1,83	1,92	1,92	1,87	1,83
S_5–O_7	—	4,30	3,76	3,12	2,40	1,69
Angle (°)	**2**	**A**	**TS_{A-B}**	**B**	**TS_{B-C}**	**C**
C_1–S_5–S_6	93,7	95,6	94,8	104,0	132,5	111,2
S_5–S_6–O_7	170,9	178,5	122,4	94,8	64,8	20,1
S_6–O_7–H_8	110,1	102,9	101,5	101,8	106,7	20,9
S_5–S_6–C_9	—	92,4	95,9	106,4	108,6	100,8
S_5–S_6–C_{13}	—	90,8	92,4	86,7	94,7	75,6
C_9–S_6–C_{13}	101,1	101,6	93,0	92,6	96,4	99,5
O_7–S_6–C_9	91,3	88,9	79,8	82,5	82,2	81,1
O_7–S_6–C_{13}	94,5	89,6	144,9	175,1	157,4	82,7
Ordre de liaison	**2**	**A**	**TS_{A-B}**	**B**	**TS_{B-C}**	**C**
C_1–S_5	1,03	1,04	1,03	1,04	1,01	1,01
S_5–S_6	0,49	0,61	0,79	0,88	0,78	0,00
S_6–O_7	0,59	0,53	0,46	0,38	0,31	0,02
O_7–H_8	0,49	0,78	0,80	0,80	0,81	0,69
S_6–C_9	0,98	0,98	0,96	0,96	0,99	1,02
S_6–C_{13}	0,99	0,98	0,86	0,82	0,89	1,02
S_5–O_7	—	0,21	0,18	0,11	0,25	0,89
Charge (u.a.)	**2**	**A**	**TS_{A-B}**	**B**	**TS_{B-C}**	**C**
C_1	-0,03	0,13	-0,13	0,08	-0,38	-0,28
S_5	-0,61	-0,55	-0,24	-0,17	-0,09	-0,09
S_6	0,56	0,55	0,39	0,30	0,17	-0,17
O_7	-0,55	-0,66	-0,69	-0,71	-0,66	-0,40
H_8	0,40	0,35	0,31	0,32	0,33	0,36
C_9	-0,30	-0,44	-0,54	-0,38	-0,25	-0,26
C_{13}	-0,41	-0,53	-0,51	-0,60	-0,52	-0,35
ΔE (kcal/mol)	**2**	**A**	**TS_{A-B}**	**B**	**TS_{B-C}**	**C**
B3LYP/6-31G(d)	—	0,00	+ 25,4	+ 14,9	+ 35,2	-30,2

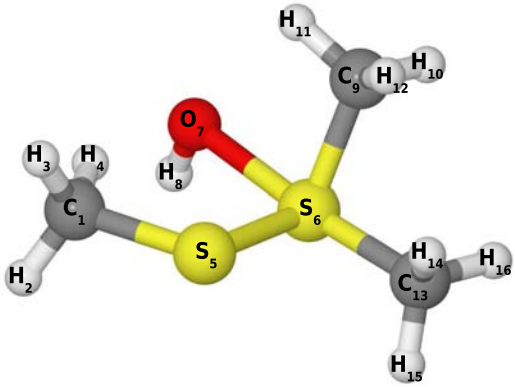
TAB. 8.1: *Valeurs des distances, des angles, des ordres de liaison, des charges et bilan énergétique de l'intermédiaire sulfurane **A**, de l'état de transition **TS_{A-B}**, de l'intermédiaire sulfurane **B**, de l'état de transition **TS_{B-C}** et de l'intermédiaire acide sulfénique **C** dans le système isolé. La géométrie est optimisée au niveau B3LYP/6-31G(d). Les charges sont calculées à partir du potentiel électrostatique. En comparaison, les valeurs correspondant à l'intermédiaire sulfurane **2** dans le modèle quantique de site actif ont été reportées.*



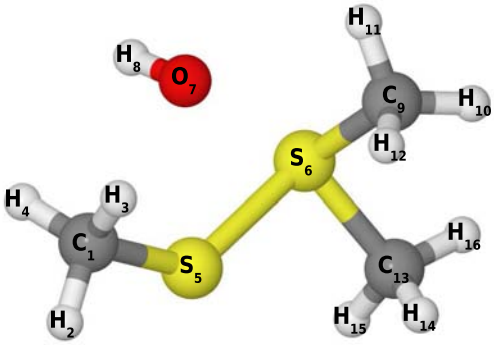
(a) Intermédiaire sulfurane **A**



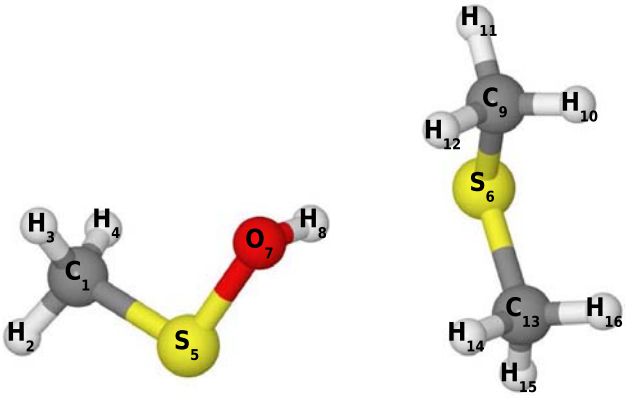
(b) Etat de transition **TS$_{A\text{-}B}$**



(c) Intermédiaire sulfurane **B**



(d) Etat de transition **TS$_{B\text{-}C}$**



(e) Intermédiaire acide sulfénique **C**

FIG. 8.2: *Structures des intermédiaires et états de transition obtenus le long du chemin réactionnel relatif au Mécanisme I (Transfert d'un groupement OH) dans un système isolé après optimisation de géométrie au niveau B3LYP/6-31G(d).*

respectivement -0,55, 0,55 et -0,66 u.a., à comparer avec -0,61, 0,56 et -0,55 u.a. dans le modèle de site actif. L'environnement entraînerait ainsi une augmentation de la charge négative portée par l'atome de soufre de Cys51, au détriment de l'atome d'oxygène.

b. Etat de transition $TS_{A\text{-}B}$

L'état de transition $\mathbf{TS_{A\text{-}B}}$ (Fig. 8.2(b)) est caractérisé par l'existence d'une fréquence imaginaire de 187**i** cm^{-1} dont le mode normal de vibration correspond à la pseudorotation avec changement de position axiale $\leftrightarrow$ équatoriale de l'atome de soufre S_5. Par rapport à **A**, la distance S_5–S_6 raccourcit (de 2,43 à 2,26 Å) et la distance S_6–O_7 s'allonge (de 1,87 à 2,03 Å). Ceci est à mettre en correspondance avec la variation des ordres de liaison pour S_5–S_6 (de 0,61 à 0,79) et S_6–O_7 (de 0,53 à 0,46). On peut noter une perte importante de charge négative (environ 0,3 u.a.) sur l'atome de soufre S_5 au niveau de cet état de transition.

c. Intermédiaire sulfurane B

Dans cet intermédiaire sulfurane, l'axe de la bipyramide est occupé par l'atome d'oxygène O_7 et un groupement méthyle (C_{13}) et le plan équatorial par l'atome de soufre S_5, l'autre groupement méthyle (C_9) et la paire libre de S_6 (Fig. 8.2(c)). Cette structure est la seule alternative possible à la structure **A** si l'on se réfère aux travaux de Balta *et al.* Un sulfurane dans lequel l'atome d'oxygène serait en position équatoriale conduirait à la dissociation de la molécule. Dans la structure obtenue ici (**B**), l'échange du groupement S-CH_3 par un groupement moins électronégatif (CH_3) induit des perturbations importantes dans la structure du système. La distance S_5–S_6 vaut 2,18 Å, soit une diminution de 0,25 Å par rapport au sulfurane **A**. La distance S_6–O_7 augmente de 1,87 à 2,07 Å. L'ordre de liaison augmente de 0,61 à 0,88 pour S_5–S_6 et diminue de 0,53 à 0,38 pour S_6–O_7. Ceci s'explique par le fait que la liaison S_5–S_6, située dans le plan équatorial, est maintenant décrite par une orbitale moléculaire liante. Les charges des deux atomes de soufre et de l'atome d'oxygène sont très différentes dans les sulfuranes **A** et **B**. Respectivement, ces charges sont passées de -0,55 à -0,17 u.a. pour S_5, de 0,55 à 0,30 u.a. pour S_6 et de -0,66 à -0,71 u.a. pour O_7. Un transfert de densité électronique relativement important s'est donc produit lors du passage de la position axiale à la position équatoriale du groupement S-CH_3. Celui-ci va être accentué dans l'étape suivante.

d. Etat de transition $TS_{B\text{-}C}$

Le deuxième état de transition $\mathbf{TS_{B\text{-}C}}$ (Fig. 8.2(d)) de ce mécanisme permet de passer du sulfurane **B** à l'acide sulfénique **C**. Il est caractérisé par une fréquence imaginaire de 391**i** cm^{-1} dont le mode normal de vibration correspond au transfert du groupement OH du soufre S_6 au soufre S_5. Cet état de transition a un caractère relativement précoce avec une distance S_6–O_7 de 2,20 Å et S_5–O_7 de 2,40 Å (ordre de liaison de 0,31 et 0,25 respectivement). L'angle S_5–S_6–O_7 à l'état de transition vaut 64,8°. L'analyse des charges atomiques montre une continuité du transfert électronique déjà observé au niveau du sulfurane **B**. La charge atomique passe de -0,17 à -0,09 u.a. pour le soufre S_5 et diminue de 0,30 à 0,17 u.a. pour le soufre S_6.

e. Intermédiaire acide sulfénique C

Le chemin de réaction aboutit au produit de la réaction **C**, illustré dans la figure 8.2(e), constitué par l'intermédiaire acide sulfénique et le diméthylsulfure. On peut noter la valeur de l'angle dièdre C_1–S_5–O_7–H_8 de -91,3° similaire à celle observée dans les peroxydes et les disulfures, et dont l'origine se trouve dans un effet d'hyperconjugaison.

f. Bilan énergétique

Les valeurs d'énergie de chacune des espèces présentées précédemment sont regroupées dans le tableau 8.1. Elles ont été calculées relativement à l'intermédiaire sulfurane **A** pris comme référence des énergies. Les variations d'énergie mises en jeu dans ce chemin réactionnel sont relativement importantes. L'intermédiaire sulfurane **B** se situe à 14,9 kcal/mol au dessus de son homologue **A** (à comparer avec la valeur de 21,2 kcal/mol dans les travaux de Balta *et al.*[236]). Comme nous l'avons déjà dit, cette structure est plus haute en énergie parce que l'axe de la bipyramide n'est pas occupée par les atomes les plus électronégatifs. L'état de transition $\mathbf{TS_{A\text{-}B}}$ se situe à 25,4 kcal/mol et traduit la difficulté importante pour le réarrangement des orbitales de l'atome de soufre central S_6.

Le deuxième état de transition $\mathbf{TS_{B\text{-}C}}$ permettant de former l'intermédiaire acide sulfénique **C** se situe à 35,2 kcal/mol (20,3 kcal/mol relativement au sulfurane **B**). Ces énergies d'activation suggèrent que la formation de l'acide sulfénique serait éventuellement possible à partir d'une structure sulfurane où le groupement S-CH_3 est en position équatoriale mais deviendrait inaccessible à partir d'une structure avec S-CH_3 en position axiale. Le rôle que l'environnement enzymatique peut jouer dans la différence d'énergie entre les formes **A** et **B** du sulfurane va être analysé ensuite.

8.1.3. Chemin de réaction dans un environnement enzymatique modèle

Cette seconde étude a pour but de transposer le chemin de réaction dans l'environnement enzymatique. Nous discuterons de l'existence ou pas des différents intermédiaires ou états de transition et de l'influence éventuelle des résidus du site actif. Compte-tenu des résultats de l'étude précédente, nous voulons savoir principalement si l'environnement enzymatique favorise la formation du sulfurane de type **B** permettant ainsi d'abaisser l'énergie d'activation du processus de réduction considéré.

Dans un premier temps, une simulation de dynamique moléculaire de l'intermédiaire sulfurane avec le groupement S-Cys en position équatoriale a été réalisée de manière analogue à celle réalisée dans le chapitre précédent pour le groupement en position axiale. Les mêmes paramètres de champ de forces (potentiels de liaison, d'angle, d'angle dièdre et de Lennard-Jones) ont été utilisés. Les charges atomiques ont été recalculées pour décrire correctement les termes d'interaction électrostatique. Le même protocole de simulation a été mis en œuvre (Voir page 119).

Nous constatons que dans cette configuration, le sulfurane n'a pas du tout le même comportement dans le site actif. En particulier, nous observons très peu d'interactions de cet intermédiaire potentiel avec les résidus du site actif. Le groupement OH ne forme aucune liaison hydrogène avec Glu94, Tyr82 et Tyr134 mais essentiellement avec des molécules d'eau du solvant. D'autre part, nous observons une grande mobilité globale de ce sulfurane dans le site actif. Ceci peut s'expliquer en partie par une moins bonne interaction de type hydrophobe entre le groupement ε-méthyle et les résidus Phe52 et Trp53. Cette simulation suggère donc que le sulfurane avec le groupement S-Cys en position équatoriale n'est pas favorisé par l'enzyme et que le mécanisme de transfert du groupement OH, très défavorable en phase gazeuse, le restera dans le cas de la réaction enzymatique.

Afin de vérifier cette analyse qualitative, un traitement quantique du mécanisme intramoléculaire dans l'environnement enzymatique a été mené en utilisant le modèle quantique de site actif que nous avons décrit dans le chapitre précédent (Fig. 7.2). La structure de départ est celle de l'intermédiaire sulfurane **2** dans le site actif. La méthode quantique utilisée est, comme auparavant, B3LYP/6-31G(d).

L'approche quantique montre en fait qu'il n'est pas possible de localiser un minimum dans la surface d'énergie potentielle correspondant à la structure de l'intermédiaire sulfurane avec S-CH_3

équatorial. Lorsque, partant de la structure **2**, l'angle S–S–O est diminué, nous observons la déprotonation spontanée de l'intermédiaire sulfurane et la dissociation de sa liaison S–S. Nous retrouvons, dans ce cas, le substrat sulfoxyde non protoné (DMSO) dans le site actif, c'est-à-dire la structure de complexe michaélien décrite dans le chapitre 6.

Nous avons également optimisé une structure de sulfurane avec S-CH_3 en position équatoriale, en imposant deux contraintes géométriques (distance S–S proche de 2,2 Å et angle S–S–O proche de 90° par analogie avec la structure **B**). A la libération des contraintes géométriques, l'optimisation de géométrie conduit à la structure **3** représentée dans la figure 8.3. Les valeurs géométriques, les ordres de liaison et les charges sont donnés dans le tableau 8.2.

Une dissociation de la liaison S_{37}–O_{38} est observée suite à une protonation du groupement OH facilitée par la molécule de phénol mimant le résidu Tyr134 du site actif. Nous avons déjà établi que la liaison S–O du sulfurane est relativement faible, notamment lorsque le groupement S-CH_3 est en position équatoriale (ordre de liaison de 0,38 dans **B**, par exemple). La présence d'un donneur de proton, même aussi faible qu'un phénol, déstabilise davantage la liaison S–O et conduit à sa dissociation. Ce résultat est en accord avec les travaux de Balta *et al.* mentionnant une rupture facile de la liaison S–O du sulfurane sous catalyse acide par un thiol. Nous reviendrons plus loin sur le rôle joué par le résidu Tyr134 dans le mécanisme de réduction.

La structure **3** se situe à 14,2 kcal/mol au-dessus de l'intermédiaire sulfurane **2** ce qui est comparable à la différence d'énergie entre les espèces **A** et **B** décrites précédemment. Signalons qu'il n'a pas été possible, pour l'instant, de localiser une structure d'état de transition permettant de connecter **2** et **3**. Malgré cela, l'ensemble de ces résultats montre que ce mécanisme ne peut être invoqué pour décrire le processus de réduction dans l'enzyme. Compte-tenu des énergies calculées (et les valeurs rapportées par Balta *et al.* en solution), l'analyse de ce mécanisme à un niveau de théorie plus poussé ne nous a pas paru pertinent.

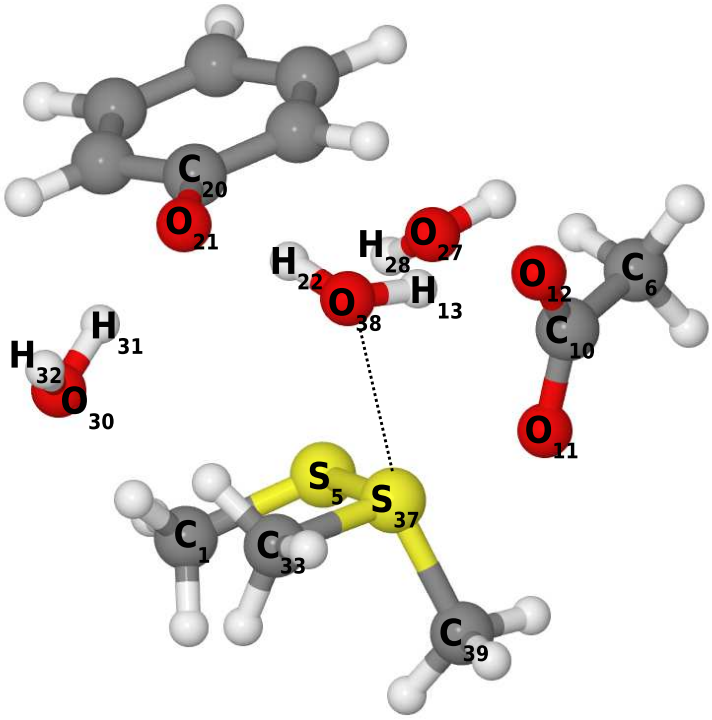


FIG. 8.3: *Structure de l'intermédiare **3** obtenu par optimisation d'une structure sulfurane avec le groupement S-CH_3 en position équatoriale dans le modèle quantique de site actif au niveau B3LYP/6-31G(d).*

Distance (Å)	**2**	**3**	**B**	Distance (Å)	**2**	**3**	**B**
C_1–S_5	1,84	1,83	1,81	O_{21}–H_{31}	2,05	1,60	—
O_{27}–H_{28}	0,99	0,99	—	H_{28}–O_{38}	1,82	1,82	—
O_{27}–O_{38}	2,79	2,78	—	S_5–S_{37}	2,58	2,12	2,18
C_6–C_{10}	1,53	1,54	—	S_5–H_{32}	2,67	4,83	—
C_{10}–O_{12}	1,28	1,25	—	H_{22}–O_{38}	1,73	1,00	—
C_{10}–O_{11}	1,25	1,28	—	H_{13}–O_{38}	1,08	0,99	0,97
O_{12}–H_{13}	1,44	1,80	—	S_{37}–O_{38}	1,82	2,79	2,07
O_{12}–O_{38}	2,51	2,79	—	S_{37}–C_{33}	1,83	1,84	1,83
C_{20}–O_{21}	1,37	1,31	—	S_{37}–C_{39}	1,83	1,83	1,92
O_{21}–H_{22}	1,00	1,68	—	S_{37}–O_{30}	3,46	4,28	—
O_{21}–O_{38}	2,72	2,67	—	S_5–O_{38}	—	—	3,12
Angle (°)	**2**	**3**	**B**	Angle (°)	**2**	**3**	**B**
C_1–S_5–S_{37}	93,7	104,8	104,0	H_{28}–O_{38}–S_{37}	123,0	126,6	—
S_5–S_{37}–O_{38}	170,9	101,2	94,8	C_{20}–O_{21}–H_{31}	109,1	125,8	—
S_{37}–O_{38}–H_{13}	110,1	73,4	101,8	O_{21}–H_{31}–O_{30}	169,1	168,6	—
O_{38}–H_{13}–O_{12}	172,6	176,9	—	H_{31}–O_{30}–H_{32}	104,6	102,2	—
C_{10}–O_{12}–H_{13}	112,2	128,0	—	S_5–H_{32}–O_{30}	157,4	35,1	—
O_{11}–C_{10}–O_{12}	126,3	127,5	—	C_1–S_5–H_{32}	99,1	56,2	—
C_{20}–O_{21}–H_{22}	111,5	124,5	—	C_1–S_5–H_{30}	103,2	50,3	—
O_{21}–H_{22}–O_{38}	169,7	167,8	—	O_{27}–O_{38}–S_{37}	120,3	124,4	—
C_{20}–O_{21}–O_{38}	116,1	126,3	—	C_6–C_{10}–O_{11}	118,7	115,5	—
H_{22}–O_{38}–S_{37}	95,6	107,1	—	C_6–C_{10}–O_{12}	115,0	116,9	—
C_{10}–O_{12}–O_{38}	115,3	127,1	—	C_{33}–S_{37}–O_{38}	91,3	82,3	82,5
S_5–S_{37}–O_{30}	71,4	70,1	—	C_{39}–S_{37}–O_{38}	94,5	156,2	175,1
O_{27}–H_{28}–O_{38}	165,5	165,6	—	C_{33}–S_{37}–C_{39}	101,1	99,0	92,9
Ordre de liaison	**2**	**3**	**B**	Ordre de liaison	**2**	**3**	**B**
C_1–S_5	1,03	1,01	1,04	S_5–S_{37}	0,49	0,97	0,88
C_6–C_{10}	0,96	0,97	—	C_{10}–O_{12}	1,36	1,50	—
C_{10}–O_{11}	1,51	1,37	—	H_{13}–O_{38}	0,49	0,66	0,80
O_{12}–H_{13}	0,23	0,08	—	S_{37}–O_{38}	0,59	0,04	0,38
S_{37}–C_{33}	0,98	0,94	0,96	S_{37}–O_{39}	0,99	0,96	0,82
Charge (u.a.)	**2**	**3**	**B**	Charge (u.a.)	**2**	**3**	**B**
C_1	-0,03	-0,51	0,08	O_{27}	-0,78	-0,84	—
S_5	-0,61	0,03	-0,17	H_{28}	0,36	0,39	—
C_6	-0,40	-0,44	—	O_{30}	-0,76	-0,89	—
C_{10}	0,73	0,83	—	H_{31}	0,37	0,46	—
O_{11}	-0,63	-0,72	—	H_{32}	0,33	0,39	—
O_{12}	-0,69	-0,66	—	C_{33}	-0,30	-0,59	-0,38
H_{13}	0,40	0,29	0,32	S_{37}	0,56	0,58	0,30
C_{20}	0,36	0,61	—	O_{38}	-0,55	-0,76	-0,71
O_{21}	-0,51	-0,80	—	C_{39}	-0,41	-0,66	-0,60
H_{22}	0,29	0,43	—	—	—	—	—

TAB. 8.2: *Valeurs des distances, des angles, des ordres de liaison et des charges de l'intermédiaire sulfurane **3** dans le modèle quantique de site actif. En comparaison, nous avons reporté les valeurs correspondantes pour l'intermédiaire sulfurane **2** dans le modèle quantique de site actif et pour le sulfurane **B** dans le système isolé. La géométrie est optimisée au niveau B3LYP/6-31G(d). Les charges sont issues du potentiel électrostatique.*

8.2. Mécanisme II : 1 H^+ + 1 e^-

8.2.1. Hypothèse de travail

Nous allons considérer maintenant un mécanisme d'oxydoréduction de type ($1H^+$, $1e^-$) dans lequel, après le transfert de proton permettant l'activation du substrat, celui-ci capture l'électron excédentaire de la cystéine catalytique. Selon cette hypothèse, illustrée par la figure 8.4, deux espèces radicalaires sont créées à la suite du transfert d'électron : un radical cystéine et un radical instable (voir plus loin) de type $DMSOH^\bullet$ qui conduit spontanément à la dissociation du radical hydroxyle. Les deux espèces radicalaires se combinent ensuite pour former l'intermédiaire acide sulfénique, le substrat méthionine réduit étant éjecté du site actif de l'enzyme durant ce processus. Nous décrirons l'état initial sous forme sulfurane (liaison covalente S–S) ou zwitterionique (liaison ionique S^- S^+) car, comme il a été discuté dans le chapitre précédent, ces deux structures formelles peuvent être envisagées dans le processus d'activation du substrat.

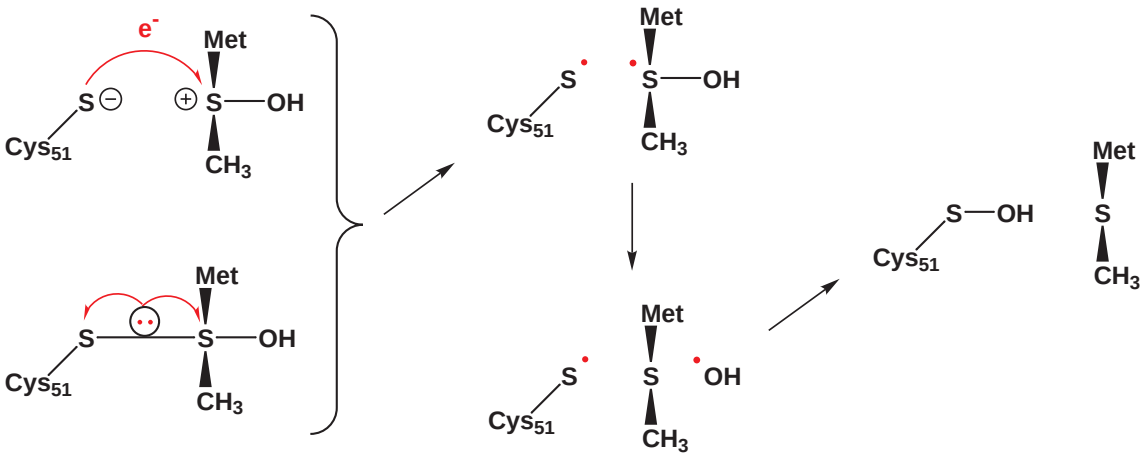


FIG. 8.4: *Hypothèse de travail envisagée pour la formation de l'intermédiaire acide sulfénique à partir de l'intermédiaire sulfurane dans le site actif de l'enzyme (Mécanisme II : 1 H^+ + 1 e^-). Le transfert d'électron peut être décrit formellement comme tel entre la cystéine déprotonée et le substrat sulfoxyde protoné (forme ionique S^- S^+) ou comme une rupture homolytique de la liaison S–S du sulfurane (forme covalente, excitation HOMO → LUMO).*

La présence d'espèces radicalaires a été testée par les expérimentateurs lors de la détermination du mécanisme catalytique. Un trappeur de radicaux libres (2,6-di-*tert*-butyl-4-methylphenol) a été placé pour cela dans le milieu réactionnel. Les résultats ont montré qu'il n'y a pas modification de ce composé et donc que la formation de radicaux semble exclue. Cependant, nous ne pouvons pas exclure d'emblée le mécanisme décrit ci-dessus. En effet, dans l'hypothèse que nous formulons, les radicaux sont formés *in situ* et se recombinent rapidement. De plus, l'écriture de la figure 8.4 est

formelle et chaque configuration correspond à une répartition bien déterminée des électrons (état *"valence bond"*) du système chimique sulfurane dont le poids sera fonction principalement de la valeur de la distance S–S considérée et de l'influence de l'environnement plus ou moins polaire.

Le transfert d'électron peut être décrit de deux manières différentes, chacune liée à la nature de liaison S–S des deux formes limites, ionique (S^- S^+) ou covalente (S–S). Dans le premier cas, il s'agit du transfert effectif d'un électron de l'atome de soufre de la cystéine vers l'atome de soufre du sulfoxyde. Dans le deuxième cas, le transfert d'électron peut être vu comme une excitation HOMO $\rightarrow$ LUMO. Conformément à ce qui a été dit dans le chapitre précédent, l'orbitale LUMO (Fig. 7.5) se présente sous forme d'une combinaison anti-liante d'orbitales atomiques des atomes S–S–O. Le transfert d'un électron de l'orbitale HOMO, portée majoritairement par le soufre de Cys51, dans l'orbitale LUMO conduit assurément à la dissociation des liaisons S–S et S–O.

8.2.2. Propriétés redox des espèces impliquées

Avant de poursuivre avec l'étude de ce mécanisme, il est important de faire une description globale de la réaction d'oxydoréduction selon laquelle un résidu cystéine est oxydé en acide sulfénique et un composé sulfoxyde de méthionine est réduit en méthionine. Cette réaction peut être décrite formellement par deux demi-équations électroniques relatives à chaque espèce. En considérant la forme ionique du sulfurane (S^- S^+), le processus élémentaire de transfert d'électron est décrit par les deux demi-équations suivantes :

$$\begin{array}{lcl} RS^- & \rightarrow & RS^\bullet + e^- \\ R_2S^+\text{–}OH + e^- & \rightarrow & R_2S^\bullet\text{–}OH \end{array}$$

Dans le but de décrire ces processus élémentaires en fonction de l'environnement, nous avons calculé le potentiel d'ionisation (PI) de la cystéine déprotonée et l'affinité électronique (AE) du substrat sulfoxyde protoné. Les méthodes quantiques de haut-niveau comme G3 ou CBS-Q sont connues pour donner des valeurs précises de ces grandeurs[284], mais elles sont excessivement coûteuses. Notre choix s'est porté une fois encore sur la méthode B3LYP mais afin d'obtenir des valeurs suffisament précises, l'utilisation d'une base étendue est nécessaire[285]. Pour l'optimisation des géométries, nous utiliserons la base B3LYP/6-31+G(d), et pour le calcul des énergies, la base 6-311++G(2df,2p). Le résidu cystéine a été modélisé par une molécule de méthanethiol (CH_3–SH) et le substrat sulfoxyde par une molécule de DMSO.

Le tableau 8.3 regroupe les valeurs du potentiel d'ionisation et de l'affinité électronique calculées pour différents états de protonation des composés en phase gazeuse. Les valeurs sans relaxation (ΔE_{vert}) et avec relaxation (ΔE_{adia}) de la géométrie ont été calculées. Les résultats montrent l'effet fondamental de l'état de protonation sur le transfert électronique entre ces deux espèces. Pour le processus CH_3–SH + $(CH_3)_2S{=}O \rightarrow CH_3{-}SH^+ + (CH_3)_2S{=}O^-$, le bilan énergétique est fortement positif ($\delta\Delta E_{\text{vert}}$ = +279,1 kcal/mol) tandis que pour le processus $CH_3{-}S^- + (CH_3)_2S^+{-}OH \rightarrow CH_3{-}S^\bullet + (CH_3)_2S^\bullet{-}OH$, il est fortement négatif ($\delta\Delta E_{\text{vert}}$ = -60,5 kcal/mol).

PI (kcal/mol)	CH_3–SH	CH_3–S^-	AE (kcal/mol)	$(CH_3)_2S{=}O$	$(CH_3)_2S^+$–OH
ΔE_{vert}	217,9	43,3	ΔE_{vert}	61,2	-103,8
ΔE_{adia}	217,0	42,8	ΔE_{adia}	—	-135,7

TAB. 8.3: *Potentiels d'ionisation de la cystéine (CH_3–SH) et de la cystéine déprotonée (CH_3–S^-) et affinités électroniques du sulfoxyde protoné ($(CH_3)_2S^+$–OH) et du sulfoxyde ($(CH_3)_2S{=}O$). Les valeurs d'énergie indiquées correspondent à la variation d'énergie électronique lors de la perte (PI) ou du gain (AE) d'un électron, sans relaxation (ΔE_{vert}) et avec relaxation (ΔE_{adia}) de géométrie. Elles sont obtenues au niveau de théorie B3LYP/6-311++G(2df,2p) sur géométries optimisées B3LYP/6-31+G(d). L'affinité électronique adiabatique du DMSO n'a pas pu être obtenue, l'optimisation de géométrie n'ayant pas convergé.*

Le gain d'un électron par $(CH_3)_2S^+$–OH est particulièrement exothermique. En fait, l'optimisation de géométrie de l'espèce radicalaire $(CH_3)_2S^\bullet$–OH conduit à un allongement de la liaison S–O (de 1,64 à 2,35 Å) et donc à la formation d'un complexe dans lequel on trouve la forme réduite du sulfoxyde, *i.e.* le thioéther :

$$(CH_3)_2S^\bullet{-}OH \quad \rightarrow \quad (CH_3)_2S\cdots{}^\bullet OH$$

Cette relaxation de géométrie représente une diminution de 31,9 kcal/mol par rapport à l'affinité électronique verticale. Le complexe $(CH_3)_2S\cdots{}^\bullet OH$ est une espèce connue du mécanisme d'oxydation des sulfures par le radical hydroxyl. Plusieurs études théoriques[286–289] ont été menées pour déterminer les structures géométriques et électroniques de ce complexe. La forme d'énergie minimale présente une liaison hydrogène très forte entre l'atome d'hydrogène du radical et l'atome de soufre. La forme décrite ici est un peu différente car elle correspond à une liaison S···O du type trois électrons à deux centres.

Pour tenir compte de l'environnement moléculaire, le potentiel d'ionisation de CH_3–S^- et l'affinité électronique de $(CH_3)_2S^+$–OH ont été recalculés dans un milieu de constante diélectrique ε en utilisant un modèle de continuum CPCM[290, 291]. Nous avons considéré trois solvants de polarité différente : Cyclohexane (ε = 2,02), DMSO (ε = 46,7) et Eau (ε = 78,39).

Les résultats, regroupés dans le tableau 8.4, montrent que le potentiel d'ionisation et l'affinité électronique sont logiquement très sensibles à l'environnement. L'augmentation de la polarité du solvant implique une augmentation de ces énergies en valeur algébrique. L'affinité électronique de $(CH_3)_2S^+$–OH reste négative quel que soit le solvant considéré. Par contre, l'augmentation du potentiel d'ionisation de CH_3–S^- entraîne un bilan énergétique du transfert d'électron (PI + AE) tendant très rapidement vers des valeurs positives, avec l'augmentation de la constante diélectrique ε du solvant.

PI CH_3–S^- (kcal/mol)	Gaz	Cyclohexane	DMSO	Eau
ΔE_{vert}	43,3	76,0	108,3	109,0
ΔE_{adia}	42,8	74,8	107,2	108,9
AE $(CH_3)_2S^+$–OH (kcal/mol)	Gaz	Cyclohexane	DMSO	Eau
ΔE_{vert}	-103,8	-75,9	-48,9	-48,3
ΔE_{adia}	-135,7	-108,7	-83,2	-82,8
PI + AE (kcal/mol)	Gaz	Cyclohexane	DMSO	Eau
ΔE_{vert}	-60,5	0,1	59,4	60,7
ΔE_{adia}	-92,9	-33,9	24,0	26,1

TAB. 8.4: *Potentiel d'ionisation de CH_3–S^- et affinité électronique de $(CH_3)_2S^+$–OH en fonction du solvant. Les valeurs sont obtenues au niveau de théorie B3LYP/6-311++G(2df,2p) sur géométries optimisées B3LYP/6-31+G(d), avec un modèle de solvant CPCM (Cyclohexane : ε = 2,02, DMSO : ε = 46,7 et Eau : ε = 78,39). La dernière partie du tableau donne la somme des énergies PI + AE correspondant au bilan énergétique du transfert d'un électron entre CH_3–S^- et $(CH_3)_2S^+$–OH pour des espèces séparées.*

Pour relier les valeurs du potentiel d'ionisation et de l'affinité électronique à une grandeur thermodynamique, nous avons calculé les potentiels d'oxydoréduction des couples CH_3–$S^\bullet$/CH_3–S^- et $(CH_3)_2S^+$–OH/$(CH_3)_2S^\bullet$–OH selon la méthode proposée par Fu *et al.*[292]. La valeur de potentiel rédox absolu (en V) d'un couple (Ox/Red) en fonction du solvant est obtenue selon l'équation :

$$E^\circ(\text{Ox/Red})_{\text{abs}} = \text{PI}_{\text{gaz Red}} + \frac{1}{\mathcal{F}}(-T\Delta S + \Delta G_{\text{solv Ox}} - \Delta G_{\text{solv Red}})$$

où $\text{PI}_{\text{gaz Red}}$ est le potentiel d'ionisation de la forme réduite en phase gazeuse (définie comme différence d'enthalpie ΔH adiabatique en eV), $\mathcal{F}$ la constante de Faraday ($\mathcal{F}$ = 23,06 kcal/mol/eV), T la température (T = 298 K), ΔS la différence d'entropie entre la forme oxydée et la forme réduite (en kcal/mol/K), $\Delta G_{\text{solv Ox}}$ et $\Delta G_{\text{solv Red}}$ l'énergie de solvatation de la forme oxydée et de la forme réduite (en kcal/mol). Ces valeurs sont calculées au niveau de théorie B3LYP/6-311++G(2df,2p) sur géométries optimisées B3LYP/6-31+G(d). Fu *et al.*[292] proposent de corriger la valeur de potentiel d'ionisation obtenue à ce niveau d'un facteur 0,28 eV. Le potentiel rédox d'un couple relativement à

l'électrode normale à hydrogène (NHE) est :

$$E^\circ(\text{Ox/Red})_{/\text{NHE}} = E^\circ(\text{Ox/Red})_{\text{abs}} - E^\circ(\text{H}_2/\text{H}^+)_{\text{abs}}$$

où $E^\circ(\text{H}_2/\text{H}^+)_{\text{abs}}$ est le potentiel rédox absolu associé à la demi-équation $\text{H}^+ + \text{e}^- \longrightarrow \frac{1}{2}\,\text{H}_2$. Il vaut 4,44 V dans un milieu aqueux[293]. Le tableau 8.5 donne le potentiel rédox absolu de chaque couple calculé en considérant trois solvants différents (Cyclohexane, DMSO et Eau). Le potentiel rédox relatif $E^\circ(\text{Ox/Red})_{/\text{NHE}}$ calculé dans l'eau est de 0,57 V et -0,46 V, respectivement pour les couples $\text{CH}_3\text{–S}^\bullet/\text{CH}_3\text{–S}^-$ et $(\text{CH}_3)_2\text{S}^+\text{–OH}/(\text{CH}_3)_2\text{S}^\bullet\text{–OH}$. En comparaison, la détermination expérimentale du potentiel rédox d'un couple thiyl/thiolate conduit à une valeur comprise entre 0,7 et 1,0 V en milieu aqueux, selon le thiol considéré[294,295]. Une mesure expérimentale du potentiel redox du couple DMSO/DMS dans l'eau donne une valeur de -0,574 V[296]. Les potentiels que nous obtenons ont donc le bon ordre de grandeur. La courbe dans la figure du tableau 8.5 montre la variation de la différence de potentiel associé à la réaction d'oxydoréduction

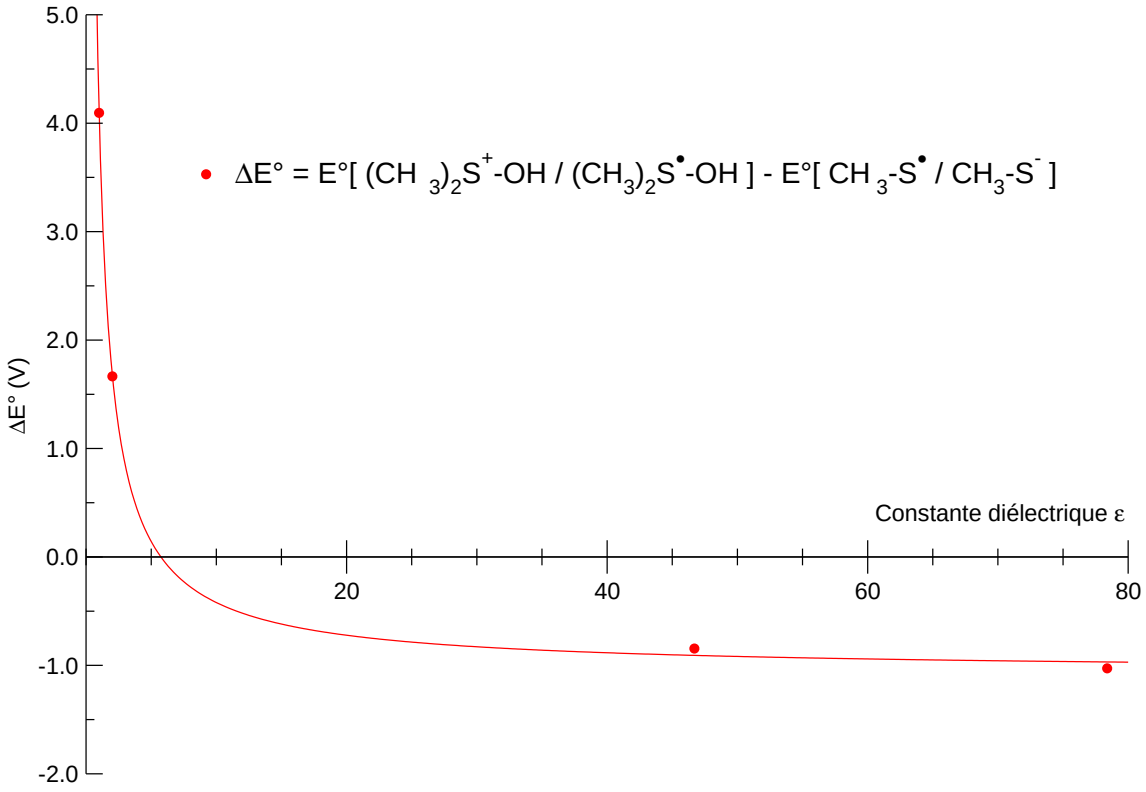
$$\text{CH}_3\text{–S}^- + (\text{CH}_3)_2\text{S}^+\text{–OH} \longrightarrow \text{CH}_3\text{–S}^\bullet + (\text{CH}_3)_2\text{S}^\bullet\text{–OH}$$

$$\Delta E^\circ = E^\circ\left[(\text{CH}_3)_2\text{S}^+\text{–OH}/(\text{CH}_3)_2\text{S}^\bullet\text{–OH}\right] - E^\circ\left[\text{CH}_3\text{–S}^\bullet/\text{CH}_3\text{–S}^-\right]$$

en fonction de la constante diélectrique du solvant. Cette différence devient rapidement négative, à partir d'une constante diélectrique d'environ 5 et on peut donc penser que dans le contexte du site actif de l'enzyme, cette réaction d'oxydoréduction sera thermodynamiquement défavorable. Cependant, cette estimation ne tient pas compte de l'énergie coulombienne d'interaction entre les ions qui est présente dans l'enzyme. Nous allons donc discuter de ce cas maintenant.

8.2.3. Transfert d'électron dans l'environnement enzymatique

Dans l'enzyme, l'énergétique liée au processus de transfert d'électron a été estimée en déterminant l'énergie de deux états électroniques : un état singulet couche fermée (Ψ_{ion}) correspondant à une forme ionique ou partiellement ionique du système moléculaire et un état singulet couche ouverte à caractère diradicalaire (Ψ_{rad}) correspondant formellement au transfert d'un électron de la cystéine vers le sulfoxyde. Cette étude a été réalisée sur deux géométries du système moléculaire sulfurane : une structure à une distance d'équilibre ($d_{\text{S–S}} \simeq 2{,}5$ Å) et une structure dissociée ($d_{\text{S–S}} \simeq 4{,}0$ Å). Le deuxième cas nous permet d'envisager la possibilité d'un mécanisme de réduction qui couplerait le transfert d'électron et l'activation du substrat par transfert de proton Glu $\longrightarrow$ DMSO. Comme nous l'avons déjà dit, le caractère plus ou moins ionique du système dépend de la distance S–S. A la distance d'équilibre ($d_{\text{S–S}} \simeq 2{,}5$ Å), la liaison possède un caractère covalent notable.



E°(Ox/Red)$_{abs}$	Solvant	PI$_{gaz\ Red}$ (kcal/mol) ΔE	ΔH	ΔG	$\Delta G_{solv\ Ox}$ (kcal/mol)	$\Delta G_{solv\ Red}$ (kcal/mol)	E° (V)
CH_3–S•/ CH_3–S$^-$	Gaz	42,82	44,03	43,68	0,00	0,00	2,17
	Cyclo.				-3,87	-33,43	3,46
	DMSO				-4,85	-66,52	4,85
	Eau				-2,00	-67,30	5,01
$(CH_3)_2S^+$–OH/ $(CH_3)_2S^\bullet$–OH	Gaz	135,73	137,03	138,82	0,00	0,00	6,30
	Cyclo.				-30,95	-3,77	5,12
	DMSO				-61,18	-8,22	4,00
	Eau				-62,00	-8,48	3,98

TAB. 8.5: *Potentiel rédox absolu des couples CH_3–$S^\bullet$/CH_3–S^- et $(CH_3)_2S^+$–OH/$(CH_3)_2S^\bullet$–OH en fonction du solvant. Les valeurs sont obtenues au niveau de théorie B3LYP/6-311++G(2df,2p) sur géométries optimisées B3LYP/6-31+G(d), avec un modèle de solvant CPCM (Cyclohexane : ε = 2,02, DMSO : ε = 46,7 et Eau : ε = 78,39), selon la méthode de Fu et al.[292]. Le graphique ci-dessus indique la variation de la différence de potentiel correspondant à la réaction d'oxydoréduction : CH_3–S^- + $(CH_3)_2S^+$–OH → CH_3–$S^\bullet$ + $(CH_3)_2S^\bullet$–OH.*

Dans ce cas, le processus de transfert d'électron est associé à un processus d'excitation électronique HOMO $\rightarrow$ LUMO. Lorsque la distance est plus longue ($d_{S-S} \simeq 4{,}0$ Å), nous avons réalisé un calcul SCF de type "restreint" pour obtenir la forme limite ionique et un calcul "non-restreint" pour obtenir la forme limite diradicalaire (avec une option guess=mix, sous GAUSSIAN03, pour briser la symétrie α/β au départ). Pour chacune de ces structures, nous avons réalisé un calcul pour le système isolé en phase gazeuse puis pour un système en interaction avec l'environnement. Le choix des structures géométriques et du modèle a été réalisé de la manière suivante :

- pour le système à la distance d'équilibre ($d_{S-S} \simeq 2{,}5$ Å), on considère la structure de l'intermédiaire sulfurane **2** dans le modèle quantique de site actif décrit page 167 ;
- pour le système dissocié ($d_{S-S} \simeq 4{,}0$ Å), nous avons utilisé un modèle quantique/classique dans lequel le système quantique est composé du substrat sulfoxyde protoné et des résidus Cys51 et Glu94 déprotonés, le reste de l'enzyme étant représenté par un ensemble de charges ponctuelles. Les résidus ont été coupés au niveau de la liaison C_α–C_β pour Cys51 et de la liaison C_β–C_γ pour Glu94. Un atome d'hydrogène a été placé à une distance de 1,08 Å du carbone de coupure, dans l'axe de la liaison coupée. Les atomes restants de Cys51 et Glu94 (*i.e.* chaîne principale) ont été supprimés. Le calcul a été effectué pour trois configurations différentes issues de la simulation de dynamique moléculaire pour le système DMSOH-$Cys^{\ominus}$-$Glu^{\ominus}$ et une valeur moyenne a été obtenue.

Les énergies des différents états électroniques ont été calculées au niveau B3LYP/6-31G(d). L'énergie pour la transition HOMO $\rightarrow$ LUMO est obtenue par un calcul de type TDDFT (premier état excité du sulfurane). La figure 8.5 présente les diagrammes énergétiques des états électroniques obtenus pour chacun des cas considérés. Pour comparaison, nous incluons les calculs pour un système Cys^-/$DMSOH^+$ séparé ($d_{S-S} = \infty$) ; dans ce cas l'effet d'environnement correspond à l'effet du solvant (Eau) obtenu par un calcul utilisant le modèle du continuum.

Les calculs pour une structure sulfurane à une distance d'équilibre (**1a** et **1b**) montrent que l'état à couche ouverte est situé à un niveau énergétique très au-dessus de l'état fondamental. L'influence de l'environnement, pris en compte dans le diagramme **1b** (sulfurane **2**, modèle de site actif), se manifeste par une très légère diminution de l'énergie relative des états électroniques.

Les diagrammes **2a** et **2b** représentent les états électroniques d'une espèce sulfurane dissociée, respectivement pour un système isolé et dans l'enzyme. Dans le cas du système isolé, l'état singulet couche fermée (Ψ_{ion}, forme ionique) se situe à 6 kcal/mol au-dessus de l'état à couche ouverte diradicalaire (Ψ_{rad}). Cette différence d'énergie est le résultat de deux contributions. D'une part, la différence de potentiel rédox, que nous avons vue plus haut, suggère que le transfert d'électron pour

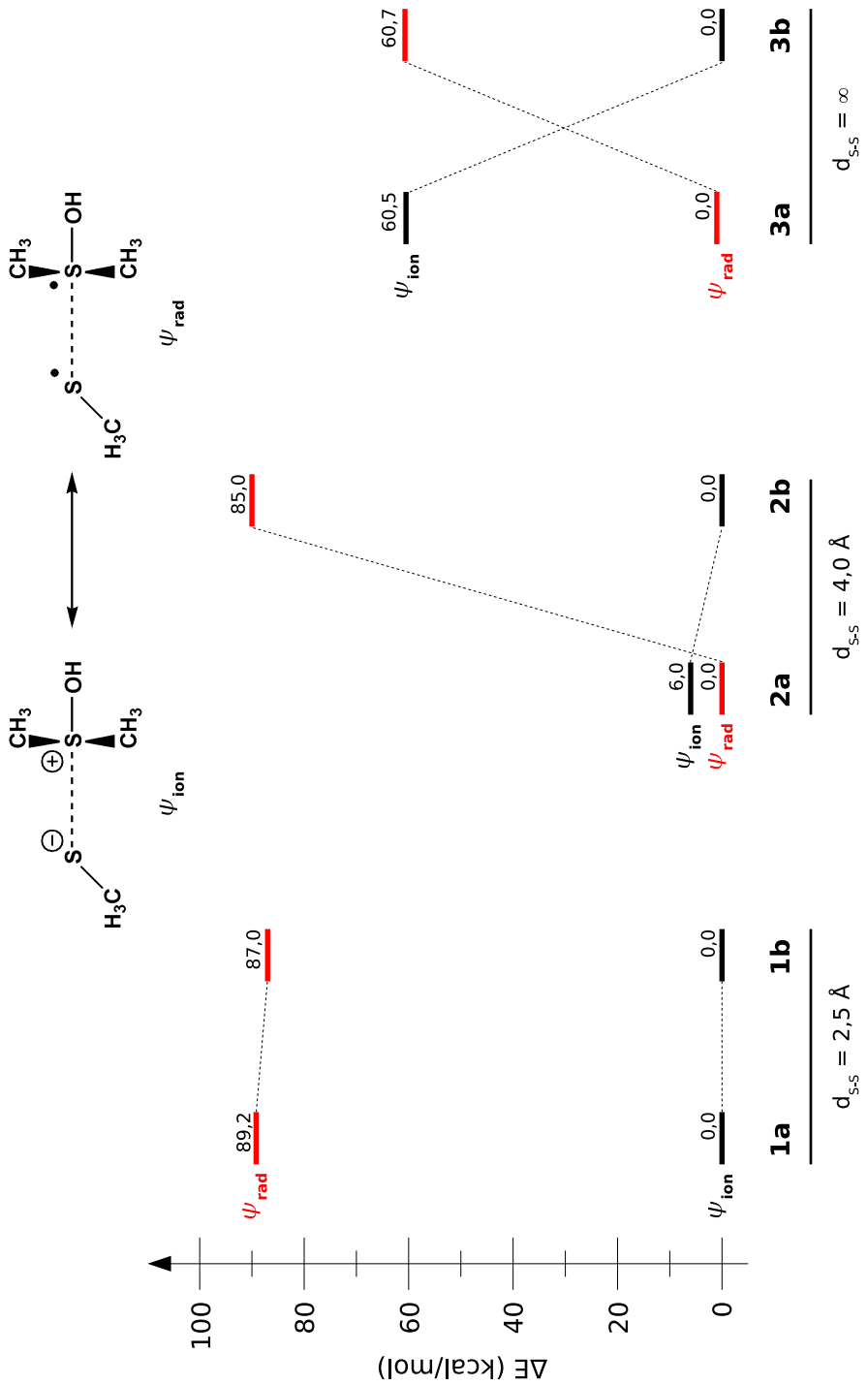


FIG. 8.5: *Diagrammes des états électroniques singulet couche fermée (Ψ_{ion}) et singulet couche ouverte (Ψ_{rad}) correspondant à trois situations : (**1**) sulfurane avec une liaison S–S à la distance d'équilibre, (**2**) sulfurane avec une liaison S–S dissociée à environ 4 Å et (**3**) sulfurane avec une liaison S–S dissociée à l'infini. Résultats pour (**a**) sans environnement et (**b**) avec environnement. Dans le cas **1** et **2**, l'environnement est représenté par un modèle de site actif de l'enzyme. Dans le cas **3**, l'environnement est le solvant eau. Les calculs ont été réalisés au niveau B3LYP/6-31G(d) pour **1** et **2** et B3LYP/6-311++G(2df,2p) pour **3**. (Voir le texte pour plus d'information)*

le couple $Cys^{-}/DMSOH^{+}$ est très favorable. D'autre part, l'interaction coulombienne entre la paire d'ions s'oppose fortement à ce transfert. Les deux contributions ont un ordre de grandeur comparable et par conséquent la différence d'énergie finale entre les deux états électroniques n'est pas très grande. Lorsque l'environnement est pris en compte, la configuration ionique est fortement stabilisée par les résidus du site actif et cette fois, le système diradicalaire se situe très haut en énergie, avec des valeurs conparables à celles obtenues pour le sulfurane à une distance d'équilibre. Le processus n'est donc pas favorable dans l'enzyme. Compte-tenu des différences d'énergie obtenues, l'utilisation d'une méthode de calcul plus poussée n'a pas été envisagée.

Les états électroniques d'une espèce dissociée à l'infini, déjà discutés dans la section précédente, sont représentés par les diagrammes **3a** et **3b**. Ils correspondent au bilan énergétique du transfert d'électron (PI + AE) entre les deux espèces isolées, respectivement en phase gazeuse et dans l'eau. La comparaison des diagrammes **2a** et **3a** permet d'estimer la contribution coulombienne dans la stabilisation de la forme ionique de **2a** lorsque $d_{S-S} \simeq 4{,}0$ Å (environ 54 kcal/mol). La comparaison des diagrammes **2b** et **3b**, en outre, permet de constater que, selon les modèles simples utilisés ici, le processus de transfert d'électron dans l'enzyme est encore moins favorable que dans l'eau.

8.3. Mécanisme III : 2 H^+ + 2 e^-

8.3.1. Hypothèse de travail

Dans les travaux de Balta *et al.*[236], le mécanisme de formation de l'acide sulfénique à partir d'un intermédiaire sulfurane en solution de DMSO a été décrit comme un processus assisté par un donneur de proton qui implique : (*i*) protonation du groupement OH avec formation d'une molécule d'eau et (*ii*) attaque de l'ion sulfuranyle par cette molécule d'eau avec formation d'un acide sulfénique et un sulfure. Le processus est schématisé dans la figure 8.6. Globalement donc, la réduction du DMSO selon ce mécanisme peut être décrit comme un processus à deux protons et deux électrons ($2H^+$, $2e^-$).

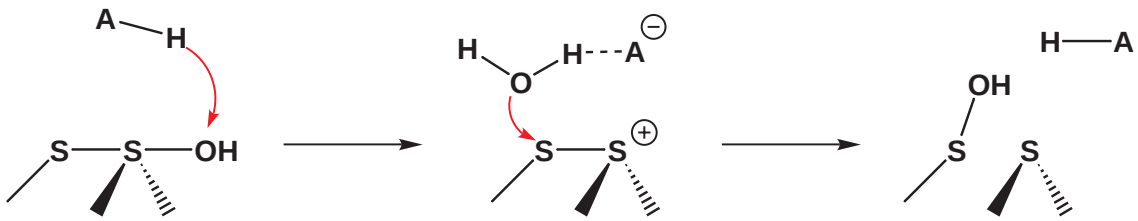


FIG. 8.6: *Hypothèse de travail envisagée pour la formation de l'intermédiaire acide sulfénique à partir de l'intermédiaire sulfurane dans le site actif de l'enzyme (Mécanisme III : 2 H^+ + 2 e^-). La liaison S–O du sulfurane est dissocié par protonation du groupement OH, sous catalyse acide (composé A–H), avec formation d'une molécule d'eau. Une molécule d'eau attaque ensuite l'ion sulfuranyle (hydrolyse) pour conduire à l'acide sulfénique et le sulfure.*

8.3.2. Chemin de réaction dans un environnement enzymatique modèle

Afin d'étudier un mécanisme de ce type dans l'enzyme, nous allons considérer à nouveau le modèle quantique de site actif (voir page 160) utilisé précédemment pour déterminer la formation de l'intermédiaire sulfurane **2** (Fig. 7.4(b), page 168). L'optimisation de la géométrie a été réalisée au niveau B3LYP/6-31G(d). L'énergie du système a été recalculée à un niveau B3LYP/6-311++G(2df,2p).

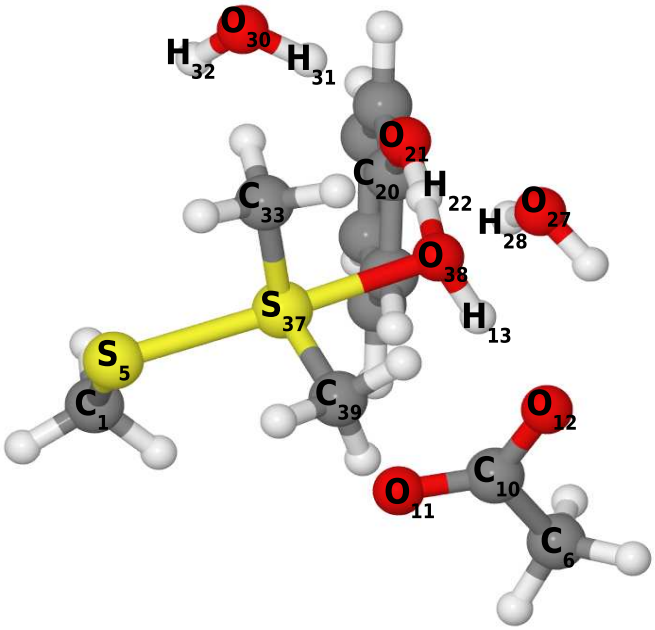
L'exploration de la surface d'énergie potentielle à partir de la structure de l'intermédiaire sulfurane **2** a permis de localiser l'état de transition $\mathbf{TS_{2-4}}$ correspondant au transfert de proton entre le résidu Tyr134 (phénol) et le groupement OH du sulfurane. La figure 8.7(a) montre la structure de cet état de transition qui est caractérisé par l'existence d'une seule fréquence imaginaire (532**i** cm^{-1}) dont le mode normal de vibration correspond à l'échange du proton H_{22} entre les atomes d'oxygène O_{21} et O_{38}. La protonation du groupement OH conduit à la dissociation de la liaison S_{37}–O_{38} du sulfurane, sous la forme d'un ion sulfuranyl et d'une molécule d'eau (**4**), comme l'illustre la figure 8.7(b).

Le tableau 8.6 regroupe les valeurs géométriques de ces deux espèces. Les ordres de liaison et les charges sont reportés dans le tableau 8.7. Rappelons que les charges atomiques ont été obtenues à partir du potentiel électrostatique et que dans le modèle utilisé, la charge nette du système est -1 u.a. Nous nous intéresserons donc à la variation de charges dans le processus plutôt qu'aux valeurs absolues de celles-ci. Le bilan énergétique est donné par le tableau 8.8. Nous avons ajouté à ces tableaux les valeurs relatives aux structures **1** et $\mathbf{TS_{1-2}}$ afin de montrer le mécanisme depuis le complexe michaélien dans le site actif.

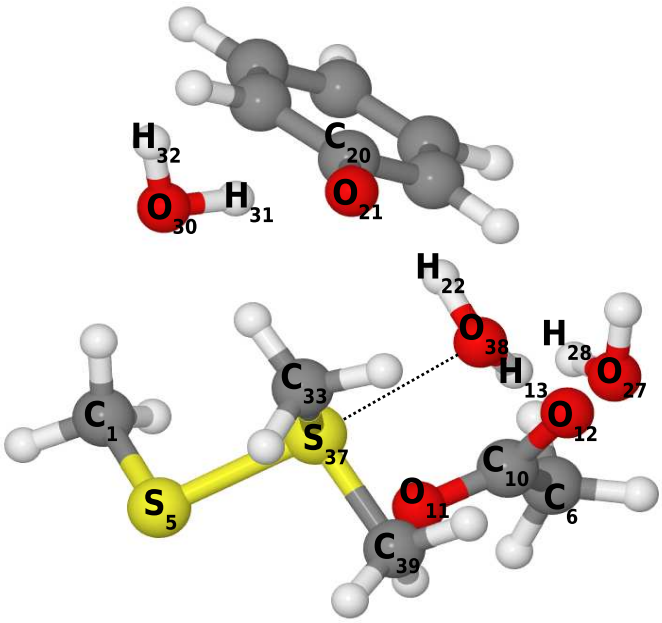
Dans l'état de transition $\mathbf{TS_{2-4}}$, la liaison S_{37}–O_{38} a augmenté de 1,82 à 2,15 Å, tandis que la liaison S_5–S_{37} a diminué de 2,58 à 2,31 Å, respectivement entre **2** et $\mathbf{TS_{2-4}}$. Ces variations sont en accord avec celles des ordres de liaison passant de 0,59 à 0,32 pour S_{37}–O_{38} et de 0,49 à 0,72 pour S_5–S_{37}. La deuxième molécule d'eau (atomes 30 à 32) vient se placer à proximité de la molécule de phénol. La distance O_{21}–H_{31} passe de 2,05 à 1,76 Å entre **2** et $\mathbf{TS_{2-4}}$. Il se forme une liaison hydrogène stabilisant la charge négative de -0,69 u.a. sur l'oxygène O_{21}. Le transfert de densité électronique entre les deux atomes de soufre que nous avons observé précédemment entre le complexe michaélien **1** et le sulfurane **2** est accru au niveau de cet état de transition. Les charges varient de -0,61 à -0,33 u.a. pour S_5 et de 0,56 à 0,48 u.a. pour S_{37}, entre **2** et $\mathbf{TS_{2-4}}$.

L'ion sulfuranyle **4** formé lors de ce tranfert de proton, porte une charge globale de +1 u.a. qui est stabilisée par la présence de la molécule d'eau H_{13}–O_{38}–H_{22} à proximité de l'atome de soufre S_{37}. La distance et l'ordre de liaison de S_{37}–O_{38} valent respectivement 2,39 Å et 0,15. L'atome d'oxygène O_{38} se situe dans l'axe de la liaison S_5–S_{37}. La molécule d'eau H_{13}–O_{38}–H_{22} est maintenue par un réseau important de liaisons hydrogène avec les groupements modélisant les résidus Glu94, Tyr82 et Tyr134. Elle participe avec l'autre molécule d'eau présente sur le site H_{31}–O_{30}–H_{32}, à la stabilisation de la charge négative de -0,65 u.a. portée par l'oxygène O_{21} du phénolate. Nous observons aussi que cette molécule d'eau s'est tourné par rapport à sa position dans l'état de transition $\mathbf{TS_{2-4}}$ pour orienter l'atome d'oxygène O_{30} vers l'atome de soufre S_5. Notons également une diminution importante de la densité électronique sur l'atome S_5 (modélisant Cys51) qui change de -0,61 à -0,24 u.a. entre l'intermédiaire sulfurane **2** et l'intermédiaire dissocié **4**.

D'après les calculs (Tab. 8.8) au niveau de théorie le plus poussé, l'état de transition $\mathbf{TS_{2-4}}$ se situe à seulement 1,4 kcal/mol au-dessus de l'intermédiaire sulfurane **2**. L'énergie d'activation de l'état de transition se trouve donc être un peu plus petite que celle obtenue dans les travaux de Balta *et al.* (ΔE = 5,5 kcal/mol) en solution de DMSO, ce qui est probablement dû (en partie au moins) à l'affaiblissement de la liaison S–O dans le sulfurane interagissant avec le carboxylate mimant Glu94.



(a) Etat de transition $\mathbf{TS_{2-4}}$



(b) Intermédiaire sulfurane dissocié **4** : Ion sulfuranyle

FIG. 8.7: *Structure de l'état de transition* $\boldsymbol{TS_{2-4}}$ *et de l'ion sulfuranyle* ***4*** *dans le modèle de site actif après optimisation de géométrie au niveau B3LYP/6-31G(d).*

Distance (Å)	**1**	**TS_{1-2}**	**2**	**TS_{2-4}**	**4**
C_1–S_5	1,85	1,84	1,84	1,83	1,84
O_{27}–H_{28}	0,99	0,99	0,99	0,99	0,99
O_{27}–O_{38}	2,80	2,80	2,79	2,78	2,70
C_6–C_{10}	1,51	1,53	1,53	1,55	1,53
C_{10}–O_{12}	1,33	1,33	1,28	1,27	1,27
C_{10}–O_{11}	1,22	1,28	1,25	1,26	1,26
O_{12}–H_{13}	1,02	1,25	1,44	1,63	1,51
O_{12}–O_{38}	2,62	2,48	2,51	2,65	2,55
C_{20}–O_{21}	1,36	1,37	1,37	1,34	1,32
O_{21}–H_{22}	1,00	1,00	1,00	1,10	1,62
O_{21}–O_{38}	2,71	2,72	2,72	2,41	2,62
O_{21}–H_{31}	2,58	2,07	2,05	1,76	1,61
H_{28}–O_{38}	1,83	1,83	1,82	1,81	1,73
S_5–S_{37}	2,97	2,61	2,58	2,31	2,18
S_5–H_{32}	2,39	2,64	2,67	4,69	4,94
H_{22}–O_{38}	1,71	1,73	1,73	1,31	1,02
H_{13}–O_{38}	1,61	1,10	1,08	1,02	1,04
S_{37}–O_{38}	1,63	1,79	1,82	2,15	2,39
S_{37}–C_{33}	1,83	1,83	1,83	1,83	1,83
S_{37}–C_{39}	1,82	1,83	1,83	1,83	1,83
Angle (°)	**1**	**TS_{1-2}**	**2**	**TS_{2-4}**	**4**
C_1–S_5–S_{37}	88,1	93,3	93,7	98,2	96,4
S_5–S_{37}–O_{38}	166,2	170,5	170,9	177,4	173,1
S_{37}–O_{38}–H_{13}	117,5	111,0	110,1	104,1	99,2
O_{38}–H_{13}–O_{12}	172,0	172,5	172,6	177,3	178,0
C_{10}–O_{12}–H_{13}	109,8	112,2	112,2	113,9	115,6
O_{11}–C_{10}–O_{12}	124,9	126,2	126,3	127,1	127,6
C_{20}–O_{21}–H_{22}	112,1	111,6	111,5	115,8	117,3
O_{21}–H_{22}–O_{38}	172,4	169,7	169,7	172,3	166,0
C_{20}–O_{21}–O_{38}	116,4	116,2	116,1	119,1	119,3
H_{22}–O_{38}–S_{37}	102,0	95,6	95,6	99,7	90,7
C_{33}–S_{37}–O_{38}	99,4	91,9	91,3	85,4	84,4
C_{39}–S_{37}–O_{38}	110,8	95,1	94,5	86,2	87,6
C_{33}–S_{37}–C_{39}	99,1	100,9	101,1	101,6	99,3
O_{27}–H_{28}–O_{38}	169,6	165,8	165,5	165,6	168,0
H_{28}–O_{38}–S_{37}	119,8	123,4	123,0	118,6	120,7
C_{20}–O_{21}–H_{31}	96,9	108,2	109,1	115,2	113,7
O_{21}–H_{31}–O_{30}	153,1	168,6	169,1	165,1	175,8
H_{31}–O_{30}–H_{32}	104,1	104,6	104,6	102,1	103,3
S_5–H_{32}–O_{30}	166,6	158,6	157,4	104,2	14,8
C_1–S_5–H_{32}	101,2	99,4	99,1	85,6	61,1
C_1–S_5–O_{30}	102,7	103,2	103,2	69,9	59,4
O_{27}–O_{38}–S_{37}	118,1	120,8	120,3	114,9	116,5
C_6–C_{10}–O_{11}	122,9	119,0	118,7	117,6	117,9
C_6–C_{10}–O_{12}	112,2	114,8	115,0	115,2	114,4
C_{10}–O_{12}–O_{38}	113,6	115,4	115,3	115,0	116,4

TAB. 8.6: *Valeurs des distances et des angles du complexe michaélien **1**, de l'état de transition **TS_{1-2}**, du sulfurane **2**, de l'état de transition **TS_{2-4}** et de l'intermédiaire sulfurane dissocié **4** dans le modèle quantique de site actif. La géométrie est optimisée au niveau B3LYP/6-31G(d).*

Ordre de liaison	**1**	**TS_{1-2}**	**2**	**TS_{2-4}**	**4**
C_1–S_5	1,04	1,03	1,03	1,03	1,02
S_5–S_{37}	0,21	0,46	0,49	0,72	0,87
C_6–C_{10}	0,99	0,97	0,96	0,96	0,96
C_{10}–O_{12}	1,14	1,34	1,36	1,42	1,38
C_{10}–O_{11}	1,68	1,52	1,51	1,48	1,49
O_{12}–H_{13}	0,57	0,26	0,23	0,14	0,18
H_{13}–O_{38}	0,13	0,46	0,49	0,59	0,54
S_{37}–O_{38}	0,88	0,61	0,59	0,32	0,15
S_{37}–C_{33}	0,97	0,98	0,98	0,99	0,98
S_{37}–C_{39}	0,97	0,98	0,99	0,99	1,00
Charge (u.a.)	**1**	**TS_{1-2}**	**2**	**TS_{2-4}**	**4**
C_1	-0,04	0,01	-0,03	-0,37	-0,22
S_5	-0,79	-0,64	-0,61	-0,33	-0,24
C_6	-0,40	-0,42	-0,40	-0,49	-0,44
C_{10}	0,71	0,73	0,73	0,83	0,79
O_{11}	-0,55	-0,62	-0,63	-0,66	-0,67
O_{12}	-0,59	-0,68	-0,69	-0,73	-0,70
H_{13}	0,38	0,40	0,40	0,41	0,40
C_{20}	0,52	0,37	0,36	0,41	0,40
O_{21}	-0,57	-0,52	-0,51	-0,69	-0,65
H_{22}	0,31	0,30	0,29	0,46	0,31
O_{27}	-0,78	-0,78	-0,78	-0,82	-0,82
H_{28}	0,34	0,37	0,36	0,36	0,33
O_{30}	-0,78	-0,76	-0,76	-0,82	-0,78
H_{31}	0,34	0,37	0,37	0,44	0,38
H_{32}	0,35	0,33	0,33	0,36	0,36
C_{33}	-0,37	-0,32	-0,30	-0,57	-0,32
S_{37}	0,74	0,59	0,56	0,48	0,51
O_{38}	-0,54	-0,57	-0,55	-0,72	-0,60
C_{39}	-0,44	-0,39	-0,41	-0,05	-0,71

TAB. 8.7: *Valeurs des ordres de liaison et des charges du complexe michaélien **1**, de l'état de transition **TS_{1-2}**, du sulfurane **2**, de l'état de transition **TS_{2-4}** et de l'intermédiaire sulfurane dissocié **4** dans le modèle quantique de site actif. La géométrie est optimisée au niveau B3LYP/6-31G(d). Les charges sont dérivées du potentiel électrostatique.*

ΔE (kcal/mol)	**1**	**TS_{1-2}**	**2**	**TS_{2-4}**	**4**
B3LYP/6-31G(d)	0,0	+ 4,5	+ 4,4	+ 8,1	+ 5,8
B3LYP/6-31++G(2df,2p)	0,0	+ 5,6	+ 5,6	+ 7,0	+ 3,8

TAB. 8.8: *Bilan énergétique du chemin réactionnel correspondant au transfert de proton entre le résidu Glu94 et le substrat sulfoxyde et de protonation de l'intermédiaire sulfurane dans le modèle de site actif. Les espèces ont été optimisées au niveau B3LYP/6-31G(d). L'énergie a été recalculée au niveau B3LYP/6-31++G(2df,2p). L'énergie du complexe michaélien **1** est la référence des valeurs d'énergie.*

Ce chemin de réaction conduisant au final à la dissociation de la liaison S–O existant initialement dans le substrat sulfoxyde, présente donc un profil énergétique qui semble, au niveau de théorie utilisé ici, très favorable au processus de réduction. L'analyse géométrique et électronique des espèces impliquées montre qu'elles sont relativement bien stabilisées par les résidus du site actif, notamment par plusieurs liaisons hydrogène et par des interactions électrostatiques.

Il peut paraître étonnant qu'un résidu tyrosine (le phénol dans notre modèle) puisse jouer le rôle de catalyseur acide, c'est-à-dire de donneur de proton permettant la dissociation de la liaison S–O, compte tenu du pK_a de ce type de groupement ($pK_{a\,Tyr}$ = 10,9). Nos résultats suggèrent que ce rôle est possible grâce à deux facteurs spécifiques au système considéré. D'une part, on sait que le sulfurane a une relative facilité à se dissocier en ions sulfuranyle ($S–S^+$) et hydroxyle (OH^-)[236], à condition de stabiliser convenablement les produits du processus (solvant, donneurs de proton). D'autre part, le processus conduit à la formation d'une paire d'ions (sulfuranyl, phénolate) séparés par une molécule d'eau, dont l'interaction coulombienne contribue notamment à la stabilisation de l'ensemble. Comme nous l'avons dit au chapitre 5, la mutation des deux résidus Tyr82 et Tyr134 produit un effet majeur sur la cinétique de réduction, la mutation d'une seule conduisant à un effet modéré. On peut donc penser que l'une ou l'autre des tyrosines participe activement dans le processus et nos calculs fournissent une possible explication des phénomènes observés.

A partir de l'intermédiaire sulfurane dissocié **4** (ion sulfuranyle), la poursuite du chemin réactionnel devrait conduire sans difficulté à la formation de l'intermédiaire acide sulfénique. En effet, selon les travaux de Balta *et al.*, l'hydrolyse en solution d'un ion sulfuranyle conduit à la dissociation de la liaison S–S (en formant l'acide sulfénique et le sulfure) avec une énergie d'activation extrêmement faible de 0,9 kcal/mol. Nous admettons ici que le devenir de l'ion sulfuranyle formé dans le site actif de l'enzyme lors de notre étude sera celui-ci puisque, d'après l'analyse des structures cristallographiques, cette zone du site actif est assez bien solvatée. On peut raisonnablement penser qu'une molécule d'eau du solvant puisse aisément hydrolyser l'ion sulfuranyle et finisse ainsi le processus de réduction du substrat sulfoxyde. Le mécanisme devrait impliquer également l'ion phénolate selon un processus assisté comme celui schématisé dans la figure 8.8.

Bien entendu, un mécanisme concerté de l'ensemble des transferts de proton peut être imaginé mais d'autres calculs seront nécessaires pour vérifier cette hypothèse qui, en outre, devront prendre en compte la modification du profil énergétique obtenu ici sous l'influence de l'interaction avec le reste de l'enzyme. Une estimation de l'influence de l'environnement a été faite en réalisant un calcul énergétique en solution aqueuse (modèle du continuum) pour les structures **1**, **2**, **4** et les états de

transition les reliant au niveau B3LYP/6-311++G(2df,2p). Dans ces conditions, les espèces **TS_{1-2}**, **2** et **TS_{2-4}** ont des énergies relatives très proches (9,1, 9,0 et 6,9 kcal/mol respectivement), ce qui a tendance à confirmer la possibilité d'un mécanisme concerté asynchrone dans lequel l'état de transition correspondrait principalement au transfert de proton de Glu94 vers le substrat et aurait ainsi une géométrie proche de l'espèce sulfurane **2**.

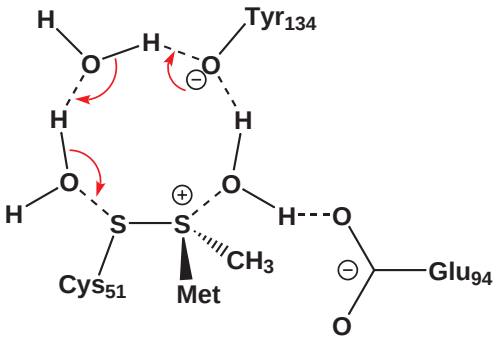


FIG. 8.8: *Hypothèse de mécanisme concerté imaginé pour la formation de l'intermédiaire acide sulfénique à partir de l'intermédiaire sulfurane dissocié dans le site actif de l'enzyme. Le processus d'hydrolyse de l'ion sulfuranyle est assisté par Tyr134 sous forme phénolate.*

Selon notre schéma réactionnel, l'atome d'oxygène de l'intermédiaire acide sulfénique ne serait pas celui présent initialement dans le substrat sulfoxyde, ce qui contraste avec des hypothèses proposées précédemment. Encouragée par nos travaux théoriques, une étude expérimentale de marquage isotopique a été réalisée par Antoine[134]. Les résultats sont regroupés dans le tableau 8.9. Trois séries d'analyses ont été réalisées avec des substrats marqués : (*i*) AcMetSO^{16}NHMe ou AcMetSO^{18}NHMe en présence d'H_2O^{16}, (*ii*) AcMetSO^{16}NHMe en présence d'H_2O^{16} ou d'H_2O^{18} et (*iii*) AcMetSO^{16}NHMe en présence d'H_2O^{18} suivi d'une dilution par H_2O^{18} ou H_2O^{16}. Les cystéines de recyclage ont été mutées en résidus sérine pour arrêter la réaction au niveau de l'intermédiaire acide sulfénique. Ce dernier a été analysé par spectrométrie de masse.

MsrA *E. coli*	Substrat	Solvant	Résultat
C86S/C198S/C206S	AcMetSO^{16}NHMe	H_2O^{16}	Cys51-SO^{16}H
C86S/C198S/C206S	AcMetSO^{18}NHMe	H_2O^{16}	Cys51-SO^{16}H
C86S/C198S/C206S	AcMetSO^{16}NHMe	H_2O^{18}	Cys51-SO^{18}H
C86S/C198S/C206S	AcMetSO^{16}NHMe	H_2O^{18}, dilution H_2O^{18}	Cys51-SO^{18}H
C86S/C198S/C206S	AcMetSO^{16}NHMe	H_2O^{18}, dilution H_2O^{16}	Cys51-SO^{16}H
C86S/C198S/C206S/E94A	AcMetSO^{16}NHMe	H_2O^{16}	Cys51-SO^{16}H
C86S/C198S/C206S/E94A	AcMetSO^{16}NHMe	H_2O^{18}	Cys51-SO^{16}H

TAB. 8.9: *Récapitulatif des conclusions des analyses de marquage isotopique de l'intermédiaire acide sulfénique pour les mutations C86S/C198S/C206S et C86S/C198S/C206S/E94A de la MsrA d'E. coli. Résultats obtenus par ESI-MS en conditions natives. Tiré d'Antoine[134].*

Les résultats de ces expériences ne sont pas concluants car il existe un échange rapide entre l'intermédiaire acide sulfénique et le solvant. Dans le cas d'une mutation du résidu Glu94 par une alanine, cet échange n'est plus observé, ce qui suggère que l'atome d'oxygène présent sur l'acide sulfénique serait celui initialement présent sur le substrat MetSO. Mais le rôle d'un résidu Ala à la place de Glu94 sur le mécanisme de réaction devra être clarifié, les résultats obtenus dans cette thèse ayant souligné l'importance de l'interaction Glu-Substrat à la fois sur la reconnaissance et la réaction enzymatique.

8.4. Devenir du substrat méthionine réduit

Nous allons discuter maintenant le processus de libération du substrat méthionine du site actif de l'enzyme une fois la réduction effectuée. Sur la base des simulations de dynamique moléculaires que nous avons effectuées précédemment, on peut tenter de décrire un scénario plausible de l'éjection de la méthionine réduite du site actif. Nous savons que l'ancrage du substrat oxydé de la méthionine s'effectue essentiellement grâce à l'interaction spécifique par liaison hydrogène d'un proton porté par Glu94 et l'atome d'oxygène de la fonction sulfoxyde. Cette interaction permet l'établissement d'autres interactions entre le substrat et les résidus du site actif (liaisons hydrogène avec Tyr82 et Tyr134, interaction hydrophobe avec Phe52 et Trp53, ...), assurant la stabilité du complexe michaélien, de même que celle de l'intermédiaire sulfurane. La réduction du sulfoxyde entraîne une dissociation des liaisons S–S et S–O. Le substrat méthionine se trouve alors entouré d'un acide sulfénique, qui potentiellement peut former une liaison hydrogène, et un $Glu^{\ominus}$ qui devrait aussi avoir un effet répulsif significatif. La méthionine est donc vraisemblablement en défaut de stabilisation et est éjectée naturellement du site actif vers le solvant.

8.5. Conclusions

Dans ce chapitre, nous avons analysé plusieurs mécanismes de réaction possibles permettant de relier l'espèce intermédiaire sulfurane, identifié dans nos calculs, à l'acide sulfénique mis en évidence expérimentalement. Trois mécanismes ont été explorés. Le premier implique un transfert de groupement OH similaire à celui observé dans l'oxydation de thiolate par les peroxydes. Le deuxième mécanisme correspond à un processus d'oxydoréduction de type ($1H^+$, $1e^-$). Il met en jeu un transfert d'électron du thiolate vers le sulfoxyde protoné, qui représente une forme limite de l'espèce sulfurane. Enfin, dans le troisième mécanisme de type ($2H^+$, $2e^-$), nous avons considéré

l'hypothèse d'un transfert de proton d'une tyrosine vers le sulfurane, conduisant à une dissociation en ion sulfuranyle et une molécule d'eau, suivi d'une hydrolyse de ce même ion.

Parmi tous ces mécanismes, le seul qui présente des barrières d'activation compatibles avec l'activité enzymatique est le troisième. Ce mécanisme est comparable à celui qui a été décrit par Balta *et al.*[236] en solution. Dans le cas d'un processus en solution, le donneur de proton est la molécule de thiol et on prévoit une barrière d'environ 5,5 kcal/mol. Dans l'enzyme, malgré le pK_a plus élevé d'un résidu Tyr par rapport au thiol, la présence de $Glu^{\ominus}$ dans le site actif a réduit la barrière considérablement à 1,4 kcal/mol. Les calculs quantiques pour l'ensemble du processus **1** $\rightarrow$ **2** $\rightarrow$ **4**, et l'estimation de l'effet d'environnement par un modèle de continuum, suggèrent la possibilité d'un mécanisme concerté **1** $\rightarrow$ **4** dans lequel l'état de transition aurait une structure proche de l'espèce sulfurane **2**.

Conclusion générale

Cette thèse avait comme objectif le développement d'un code de docking moléculaire utilisant une méthode quantique pour le calcul de l'énergie et une étude théorique du mécanisme catalytique des enzymes MsrA, qui font l'objet de recherches expérimentales dans l'équipe AREMS au sein de notre université. Les principales conclusions dérivées de notre travail sont présentées ci-dessous.

Logiciel de docking moléculaire semi-empirique à croissance linéaire

Le logiciel de docking moléculaire AlgoGen-DivCon a été développé dans le but de décrire un système protéine-ligand de grande taille selon une approche de mécanique quantique. Il se présente sous la forme d'une interface entre un algorithme génétique (AlgoGen) pour la partie minimisation, entièrement développé par nos soins, et le programme DivCon qui effectue l'évaluation énergétique du système moléculaire par une méthode semi-empirique couplée à l'algorithme de croissance linéaire Divide & Conquer. Les tests que nous avons effectués ont montré que :

- le programme est capable de déterminer la conformation favorable d'un système moléculaire. Ont été validés : (*i*) l'algorithme génétique en tant qu'algorithme de minimisation et (*ii*) l'approche semi-empirique à croissance linéaire pour l'évaluation des structures ;
- la conception du programme permet une utilisation dans un cadre d'architecture de calcul distribué-parallèle ;
- la description quantique du système moléculaire permet de prendre en compte les effets électroniques absents dans les autres méthodes de docking.

L'ensemble des travaux de développement et de validation a été publié sous la référence :

Eddy Thiriot and Gérald Monard,
Combining a genetic algorithm with a linear scaling semiempirical method for protein-ligand docking
J. Mol. Struct.-THEOCHEM, **898**, 31–41 (2009)

Mécanisme catalytique des méthionine sulfoxyde réductases de classe A

L'étape réductase du mécanisme catalytique des méthionine sulfoxyde réductases de classe A a été caractérisée depuis l'ancrage du substrat dans le site actif jusqu'au mécanisme de réduction par une approche théorique "multi-outils".

La reconnaissance et l'ancrage du substrat sulfoxyde (DMSO) ont été étudiés par des simulations de dynamique moléculaire en abordant en détail la problématique de l'état de protonation des résidus Cys51 et Glu94 du site actif. Nos résultats indiquent que :

- l'état de protonation de l'enzyme libre correspond vraisemblablement à CysH-GluH ou CysH-Glu$^{\ominus}$;
- la protonation de Glu94 est essentielle à l'ancrage du substrat ;
- le substrat est stabilisé par (*i*) la formation de liaisons hydrogène avec Glu94, Tyr82 et Tyr134 au niveau de l'oxygène du sulfoxyde, (*ii*) une interaction électrostatique entre le soufre de Cys51 et le soufre du sulfoxyde et (*iii*) une interaction de type hydrophobe d'un groupement méthyle du substrat avec les résidus Phe52 et Trp53 ;
- la déprotonation de la cystéine catalytique (Cys51) est induite par l'ancrage du substrat sulfoxyde, selon un mécanisme pouvant être assisté par Tyr197.

La protonation du substrat dans le site actif constitue l'étape d'activation du mécanisme nécessaire à la réaction d'oxydoréduction. Le mécanisme de transfert de proton a été décrit au moyen de calculs quantiques dans un modèle de site actif. Nos résultats ont montré que :

- l'espèce sulfoxyde protonée (ionique) est relativement peu envisageable dans le site actif de l'enzyme : elle ne correspond pas à un minimum d'énergie ;
- la protonation de l'oxygène du sulfoxyde conduit à la formation spontanée d'une liaison entre l'atome de soufre de MetSO et celui de Cys51, *i.e.* à la formation d'un intermédiaire sulfurane ;
- l'intermédiaire sulfurane est une espèce labile, très proche structurellement de l'état de transition de protonation, qui en outre présente des liaisons S–S et S–O très affaiblies ;
- le site actif de l'enzyme semble être parfaitement adapté pour stabiliser cette espèce sulfurane et donc l'état de transition de protonation.

Différents mécanismes ont été envisagés pour décrire la réduction enzymatique du substrat sulfoxyde à partir de l'intermédiaire sulfurane. Ces travaux se résument ainsi :

- le transfert du groupement OH du sulfurane sur le soufre de Cys51 *via* une pseudorotation ne semble pas envisageable dans l'environnement du site actif ;
- un mécanisme de transfert d'un électron de Cys51 vers le sulfoxyde ou de rupture homolytique de la liaison S–S, envisageable en phase gazeuse, met en jeu des énergies trop importantes pour être considéré dans l'enzyme ;
- la protonation du groupement OH de l'intermédiaire sulfurane par un résidu tyrosine conduisant à la formation d'un ion sulfuranyle et d'une molécule d'eau est énergétiquement envisageable dans le site actif, si l'on admet (sur la base de travaux antérieurs) que l'ion sulfuranyle est aisément hydrolysable en acide sulfénique et en substrat réduit. Ce mécanisme correspond à un mécanisme de réduction de type ($2H^+$, $2e^-$).

Nos études ont ainsi révélé que les résidus Glu94, Tyr82 et Tyr134 ont une importance fondamentale dans chacune des étapes que nous venons de décrire. Glu94 est impliqué dans la reconnaissance et l'ancrage (liaison hydrogène essentielle) et dans le mécanisme d'activation du substrat sulfoxyde (première protonation). Tyr82 et Tyr134 (l'une ou l'autre) sont impliquées dans l'ancrage (liaisons hydrogène) et dans le mécanisme de réduction (deuxième protonation). Les résidus Cys51, Glu94, Tyr134 (ou Tyr82) sont donc nécessaires à l'activité de l'enzyme MsrA et leur mécanisme d'action peut être rapproché de celui d'une triade catalytique.

Perspectives

Les travaux présentés dans cette thèse pourront se prolonger par un certain nombre d'études.

Algorithmes de docking

En perspective de ce travail de développement et de validation du programme d'AlgoGenDivCon, on peut envisager la mise au point d'une version qui permettrait la modification éventuelle de l'état de protonation des résidus dans le complexe protéine-ligand lors du processus de docking. Comme nous l'avons vu, cet aspect est rendu possible grâce à la description au niveau quantique du système moléculaire et pourrait être incorporé facilement dans notre algorithme génétique par la définition d'un gène supplémentaire dans le chromosome des individus.

Par ailleurs, la méthode semi-empirique utilisée pourra être améliorée pour mieux décrire les interactions moléculaires. Pour ce faire, on peut envisager l'implantation de diverses approches parues (ou à paraître) dans la littérature, dont celle proposée dans notre laboratoire (PM3-MAIS[74, 75]).

Mécanisme des méthionine sulfoxyde réductases

Deux aspects semblent importants à approfondir : (*i*) l'utilisation d'un niveau de calcul plus poussé, y compris pour l'optimisation de la géométrie moléculaire, et (*ii*) l'amélioration du modèle de l'enzyme. D'après les résultats présentés dans cette thèse, on peut penser que l'environnement protéinique, c'est-à-dire l'ensemble des résidus au-delà du site actif, joue un rôle important dans le mécanisme de réduction du substrat sulfoxyde par MsrA. Cet effet joue certainement sur l'asynchronicité des transferts de proton et pourrait donc modifier la nature de l'état de transition. Un objectif prioritaire serait donc d'étudier le mécanisme proposé dans ce manuscrit en considérant l'ensemble des atomes de l'enzyme, ce qui pourrait être réalisé dans une approche hybride de type QM/MM.

Une autre perspective intéressante serait d'étudier le mécanisme de réduction de MsrB. L'identification des résidus du site actif de MsrB indique que l'équivalent de Glu94 serait His103 ; en revanche, il n'y a pas d'équivalent pour Tyr82 et Tyr134. La transposition du mécanisme de MsrA à MsrB ne peut donc pas se faire directement. Bien entendu, les autres étapes du mécanisme de MsrA devront être abordées.

Enfin, nous formulons quelques propositions d'études expérimentales qui pourraient apporter des informations complémentaires à nos résultats :

- mutagénèse de Tyr197, effet vis-à-vis du mécanisme de déprotonation de Cys51 ;
- mutagénèse de Tyr82-134 en conjonction avec Glu94 : une double mutation du type Y82F/E94Q ou Y134F/E94Q devrait avoir un effet plus important sur la catalyse enzymatique (k_{cat}) que sur la reconnaissance (K_{M}) ;
- marquage isotopique (études déjà réalisées à compléter) pour identifier la provenance de l'atome d'oxygène sur l'intermédiaire acide sulfénique (sulfoxyde ou solvant).

De manière plus générale, on peut remarquer qu'un effort théorique conséquent est nécessaire si l'on veut mieux comprendre la chimie du soufre au niveau biologique, dont seulement quelques aspects ont été abordés ici. Cet effort concerne à la fois les approches dites classiques (développement de champs de forces adaptés) et les méthodes quantiques, dont relativement peu d'applications de haut niveau ont été consacrées à cet élément aux propriétés électroniques singulières.

Troisième partie

Annexes

Annexe A

Méthodologie

Sommaire

A.1. Méthodes semi-empiriques à croissance linéaire 227
A.2. Mécanique et dynamique moléculaires 230
A.3. PROPKA : Prédiction des valeurs de pK_a des résidus d'une protéine 239

Dans cette annexe, nous présentons quelques détails sur la méthodologie employée dans nos calculs. Nous parlerons essentiellement des approches semi-empiriques à croissance linéaire, utilisées dans la partie I de notre thèse, puis des méthodes de mécanique et dynamique moléculaires, mises en œuvre dans la partie II. Nous finirons par une description de la méthode PROPKA, à laquelle nous avons fait appel, par exemple, pour la détermination du pKa des groupements ionisables des chaînes latérales de MsrA.

Nous ne parlerons pas ici des méthodes quantiques autres que les approches semi-empiriques mentionnées plus haut. Nous avons eu recours à des méthodes de type DFT (*Density Functional Theory*) pour les études de réactivité (Chapitres 7 et 8). Ces méthodes sont maintenant bien connues et leur efficacité n'est plus à démontrer. Elles représentent un excellent rapport qualité/prix et compte tenu de la taille relativement importante des systèmes que nous avons eu à traiter, leur choix s'est imposé. Parmi les différentes fonctionnelles disponibles dans la théorie DFT, nous avons sélectionné la fonctionnelle hybride B3LYP qui est sans aucun doute la plus répandue actuellement. Différentes bases d'orbitales atomiques ont été utilisées en fonction du problème étudié. En règle générale, nous avons utilisé la base 6-31G(d) pour les calculs exploratoires et pour les optimisations de géométrie, la base 6-311++G (2df,2p) ayant été réservée pour le calcul des énergies.

Pour plus de détails sur l'ensemble des méthodes, le lecteur peut consulter les ouvrages suivants :

Jean-Louis Rivail,
Eléments de chimie quantique à l'usage des chimistes,
CNRS Editions (1999)

Christopher J. Cramer,
Essentials of Computational Chemistry : Theories and Models,
Wiley (2006)

Frank Jensen,
Introduction to Computational Chemistry,
Wiley (1999)

John A. Pople and David L. Beveridge,
Approximate Molecular Orbital Theory,
McGraw-Hill (1986)

Attila Szabo and Neil S. Ostlund,
Modern Quantum Chemistry,
Courier Dover Publications (1989)

A.1. Méthodes semi-empiriques à croissance linéaire

Les méthodes *ab initio* et DFT ont une demande très importante en temps de calcul qui, de plus, augmente très rapidement avec le nombre d'atomes, ou plus exactement avec le nombre de fonctions de base. Afin de réduire de façon drastique ce temps de calcul, la stratégie commune employée par les méthodes semi-empiriques[297,298] repose sur les approximations suivantes :

- utiliser une base d'orbitales atomiques minimale ;
- négliger les intégrales de recouvrement dans la résolution des équations de Roothaan-Hall ;
- négliger un grand nombre d'intégrales biélectroniques (approximation ZDO ou *Zero Differential Overlap*) ;
- mettre les autres intégrales sous forme paramétrée et ajuster les paramètres avec des données expérimentales ou des calculs *ab initio* ;
- limiter la description du système aux électrons de valence.

Dans le formalisme ZDO, les intégrales biélectroniques faisant intervenir le recouvrement différentiel de deux orbitales atomiques différentes sont négligées. Selon le degré d'utilisation de cette approximation, on a développé plusieurs classes de méthodes semi-empiriques, citées ici dans l'ordre décroissant du nombre d'approximations effectuées :

- CNDO[299] : *Complete Neglect of Differential Overlap* ;
- INDO[300] : *Intermediate Neglect of Differential Overlap* ;
- NDDO[301] : *Neglect of Diatomic Differential Overlap*.

Dans ces dernières, on considère que le recouvrement différentiel de deux orbitales centrées sur deux atomes différents est nul. Parmi les méthodes semi-empiriques existantes, les plus utilisées sont AM1[82] et PM3[84,85], toutes les deux de type NDDO.

Dans ce travail, nous avons utilisé une variante de la méthode PM3 développée au laboratoire. Dans cette version, appelée PM3-PIF[74–76,86] (PIF pour *Parametrizable Interaction Function*), on élimine un certain nombre d'artéfacts liés à l'utilisation de fonctions gaussiennes dans le calcul de l'énergie d'interaction cœur-cœur de la méthode standard. Ces fonctions sont remplacées par une somme de termes de type $\frac{1}{r^n}$ ayant le comportement physique correct. Ce traitement s'applique uniquement aux interactions intermoléculaires.

Divide & Conquer

Malgré toutes ces approximations, l'étude de systèmes très grands peut être coûteuse car l'étape limitante est la diagonalisation de la matrice de Fock au cours de la procédure SCF. Ainsi, plusieurs algorithmes de calcul ont été imaginés afin d'obtenir une linéarité entre le temps de calcul et la taille du système traité. Nous présentons ici la méthode Divide & Conquer qui permet de s'approcher de cette croissance linéaire.

La méthode Divide & Conquer fut d'abord développée dans le cadre de la DFT[33, 302], puis reformulée selon l'approche semi-empirique[79–81] dont nous allons utiliser le formalisme par la suite. Le postulat de base est la localité de la densité électronique (dans le cas des systèmes non métalliques). A partir de ce constat, le système moléculaire d'intérêt peut être divisé en α sous-systèmes $\mathcal{R}^\alpha$ qui se recouvrent et dont on résout, pour chacun d'entre eux, l'équation de Roothaan-Hall localisée suivante :

$$\mathbf{F}^\alpha \mathbf{C}^\alpha = \mathbf{C}^\alpha \mathbf{E}^\alpha \tag{A.1}$$

Dans cette équation, $\mathbf{C}^\alpha$ est la matrice locale des coefficients, $\mathbf{F}^\alpha$ la matrice locale de Fock et $\mathbf{E}^\alpha$ la matrice diagonale locale des énergies ε_i^α des orbitales i du sous-système α. La matrice densité locale est construite grâce aux coefficients de la matrice $\mathbf{C}^\alpha$ selon :

$$P_{\mu\nu}^\alpha = \sum_i^{N^\alpha} n_i^\alpha c_{\mu i}^{\alpha *} c_{\nu i}^\alpha \tag{A.2}$$

où N^α est le nombre d'orbitales dans le sous-systèmes α, n_i^α est le nombre d'occupation dépendant de l'énergie de Fermi ε_F et de l'énergie ε_i^α de l'orbitale moléculaire i selon :

$$n_i^\alpha = \frac{2}{1 + \exp\left[\left(\frac{\varepsilon_i^\alpha - \varepsilon_F}{k_B T}\right)\right]} \tag{A.3}$$

Dans cette expression, k_B est la constante de Boltzmann et T représente une température absolue qui n'a pas de signification physique réelle et vaut habituellement 1000 K[80]. L'énergie de Fermi est déterminée en utilisant une technique de bisection itérative sur les valeurs propres du sous-système, sous la contrainte que le nombre d'électrons soit conservé. La matrice de Fock locale est construite à partir de la matrice de Fock globale selon :

$$F_{\mu\nu}^\alpha = \begin{cases} F_{\mu\nu} & \text{si les fonctions de base } \chi_\mu \text{ et } \chi_\nu \text{ sont dans le même sous-système } \mathcal{R}^\alpha \\ 0 & \text{sinon} \end{cases} \tag{A.4}$$

Par analogie avec les méthodes semi-empiriques traditionnelles, la matrice de Fock globale est construite à partir de la matrice $H_{\mu\nu}$ à un électron, des intégrales coulombiennes et d'échange et de

la matrice densité globale $P_{\mu\nu}$:

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda}\sum_{\sigma}\left[\left(\mu\nu|\,\lambda\sigma\right) - \frac{1}{2}\left(\mu\sigma|\,\lambda\nu\right)\right]P_{\mu\nu} \tag{A.5}$$

Chaque sous-système est constitué d'un cœur entouré par une[80] ou deux[81] régions-tampons. Ces régions-tampons déterminent quelle information doit être utilisée pour construire la matrice densité globale à partir des matrices densité locales :

$$P_{\mu\nu} = \sum_{\alpha}^{n_{\text{sub}}} D^{\alpha}_{\mu\nu} P^{\alpha}_{\mu\nu} \tag{A.6}$$

La valeur de $D^{\alpha}_{\mu\nu}$ est donnée selon le nombre de tampons :

- Un tampon :

$$D^{\alpha}_{\mu\nu} = \begin{cases} \dfrac{1}{n_{\mu\nu}} & \text{si } \chi_{\mu} \text{ et } \chi_{\nu} \text{ sont dans le cœur} \\ 0 & \text{sinon} \end{cases} \tag{A.7}$$

- Deux tampons :

$$D^{\alpha}_{\mu\nu} = \begin{cases} \dfrac{1}{n_{\mu\nu}} & \text{si } \chi_{\mu} \text{ est dans le cœur et } \chi_{\nu} \text{ dans le cœur ou le tampon interne, et inversement} \\ 0 & \text{sinon} \end{cases} \tag{A.8}$$

où $n_{\mu\nu}$ est le nombre total de sous-systèmes se recouvrant et ayant une contribution non nulle à la matrice densité globale $P_{\mu\nu}$. Il a été constaté que la procédure avec deux tampons accroît significativement l'efficacité de la méthode, avec seulement une faible augmentation du temps de calcul.

Dans un calcul standard, la part la plus importante du temps de calcul est la diagonalisation de la matrice de Fock, croissant en $O\left(N^3\right)$, N étant la dimension de la matrice de Fock. Dans le formalisme Divide & Conquer, n_{sub} matrices locales de Fock de dimension N^{α} $\left(N^{\alpha} \ll N\right)$ sont diagonalisées. Pour un système bien équilibré, *i.e.* un système dans lequel tous les sous-systèmes ont un nombre égal de fonctions de base, le coût d'un calcul Divide & Conquer est proportionnel à $n_{\text{sub}} O\left(N^{\alpha 3}\right)$, *i.e.* linéaire au nombre de sous-systèmes. Le temps de calcul peut être encore réduit en utilisant un cut-off explicite pour les matrices de Fock et densité[81].

La méthode Divide & Conquer couplée à des méthodes semi-empiriques a été développée dans le programme DIVCON[78].

A.2. Mécanique et dynamique moléculaires

Ce chapitre décrit succinctement les principes de la modélisation de systèmes chimiques en mécanique moléculaire, puis la technique de simulation de dynamique moléculaire classique que nous avons utilisée pour notre étude. Une description plus complète est donnée par la lecture de livres tels que ceux d'Allen et Tildesley[303], de Leach[304] et de Rapaport[305]. La mécanique moléculaire décrit l'état d'un système moléculaire uniquement à partir de la position des atomes le constituant. Ces atomes interagissent entre eux à l'aide de potentiels prédéfinis simples qui constituent un champ de forces paramétré à partir des résultats expérimentaux et/ou des calculs quantiques poussés. La dynamique moléculaire utilise les champs de forces et les principes de la dynamique newtonienne pour prédire l'évolution dans le temps du système et en déduire ses propriétés thermodynamiques.

A.2.1. Champs de forces

Le rôle du champ de forces est de reproduire le mieux possible les interactions intra- et intermoléculaires réelles du système. En principe, le potentiel d'interaction $V\left(\mathbf{r}^N\right)$ d'un système à N particules s'écrit comme une somme de potentiels à n-corps ($n = 1, \ldots, N$) :

$$V\left(\mathbf{r}^N\right) = \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2\left(\mathbf{r}_i, \mathbf{r}_j\right) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3\left(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k\right) + \ldots \quad (i = 1, \ldots, N) \tag{A.9}$$

où les $\{\mathbf{r}_i\}$ correspondent aux coordonnées cartésiennes des particules et $v_1(\mathbf{r}_i)$, $v_2\left(\mathbf{r}_i, \mathbf{r}_j\right)$, ... représentent le potentiel intramoléculaire (à 1 corps), le potentiel d'interaction de paire (à 2 corps), *etc*. Afin de parvenir à des temps de calcul raisonnables, la plupart des champs de forces considèrent l'approximation suivante :

$$\begin{aligned} V\left(\mathbf{r}^N\right) &= V_{\text{liés}} + V_{\text{non-liés}} \\ &= V_{\text{élongation}} + V_{\text{flexion}} + V_{\text{torsion}} + V_{\text{torsion impropre}} \\ &\quad + V_{\text{électrostatique}} + V_{\text{van der Waals}} \end{aligned} \tag{A.10}$$

Le potentiel $V_{\text{liés}}$ contient les contributions des liaisons covalentes ($V_{\text{élongation}}$) et des effets d'hybridation dans les déformations angulaires (V_{flexion}) et diédrales (V_{torsion} et $V_{\text{torsion impropre}}$). Les termes du potentiel $V_{\text{non-liés}}$ comprennent les interactions de van der Waals ($V_{\text{van der Waals}}$) et électrostatiques ($V_{\text{électrostatique}}$) entre atomes non-liés (en général, limités aux atomes séparés par au moins trois liaisons chimiques). Ces différents termes sont représentés de manière schématique dans la figure A.1.

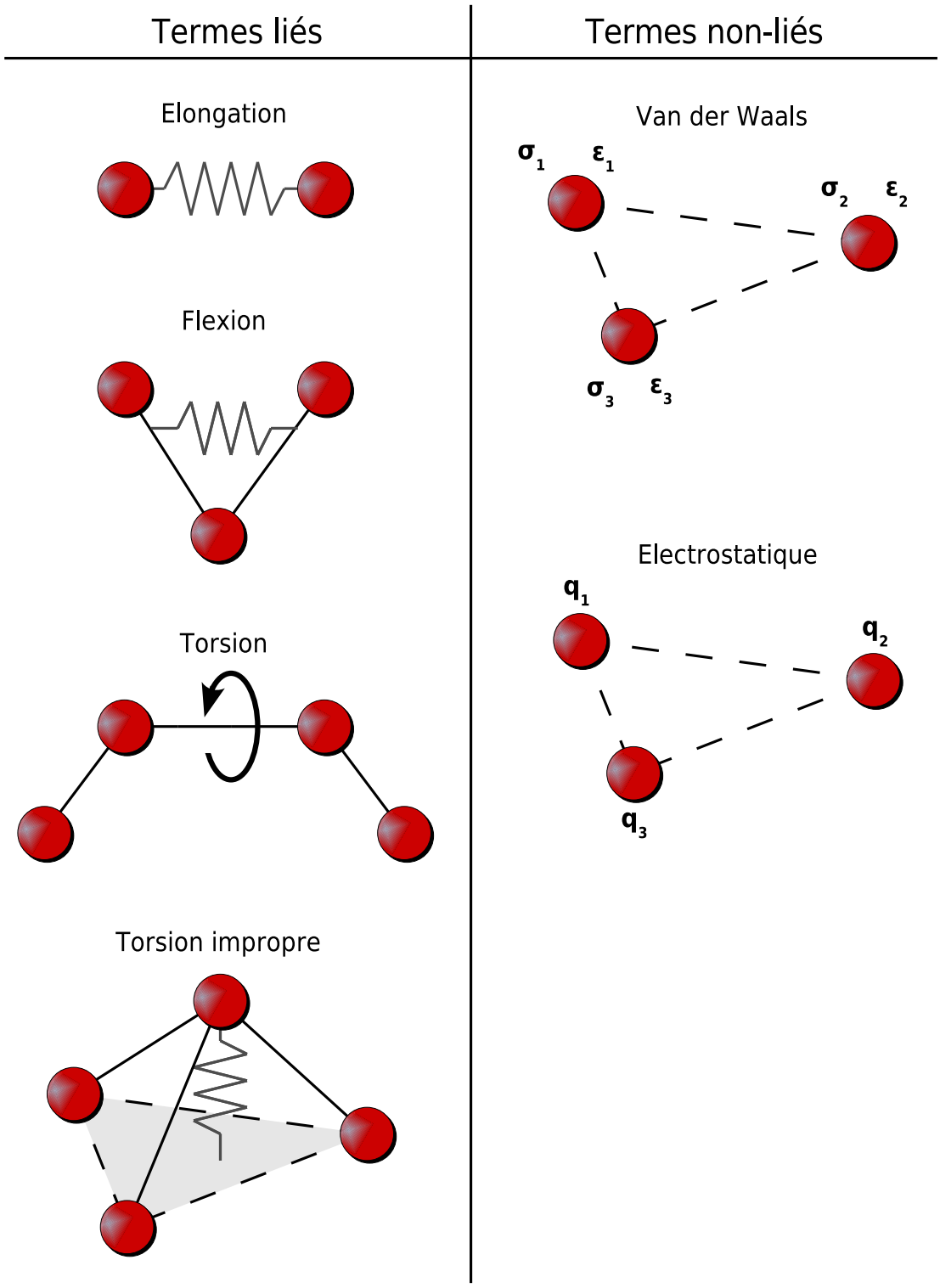


FIG. A.1: *Représentation schématique des différents termes intervenant dans un champ de forces : (i) termes entre atomes liés : élongation, flexion, torsion et torsion impropre et (ii) termes entre atomes non-liés : interaction de van der Waals et interaction électrostatique.*

a. Termes liés

Elongation Le potentiel d'élongation $V_{\text{élongation}}$ des liaisons covalentes est généralement représenté par un simple potentiel harmonique :

$$V_{\text{élongation}} = k_{\text{élongation}} \left(r - r_{\text{eq}}\right)^2 \tag{A.11}$$

où r est la distance entre les deux atomes, r_{eq} la distance d'équilibre de la liaison et $k_{\text{élongation}}$ la constante de force (ou de rappel) du potentiel harmonique.

Dans le but de se rapprocher le plus possible du comportement de la liaison chimique, *i.e.* modélisation de la dissociation moléculaire (anharmonicité du potentiel), le potentiel d'élongation peut aussi être représenté par

- un potentiel de Morse :

$$V_{\text{élongation}} = k_{\text{élongation}} \left[1 - \mathrm{e}^{-A(r - r_{\text{eq}})}\right]^2 \tag{A.12}$$

- un potentiel anharmonique :

$$V_{\text{élongation}} = \frac{1}{2}k_2(r - r_2)^2 + \frac{1}{3}k_3(r - r_3)^3 + \frac{1}{4}k_4(r - r_4)^4 \tag{A.13}$$

Flexion Le terme de flexion décrit la force résultant de la déformation des angles de valence entre trois atomes liés :

$$V_{\text{flexion}} = k_{\text{flexion}} \left(\theta - \theta_{\text{eq}}\right)^2 \tag{A.14}$$

où θ est l'angle de valence entre les atomes, θ_{eq} l'angle d'équilibre et k_{flexion} la constante de force. Les constantes de force de flexion sont nettement plus faibles que celles d'élongation, conformément à ce que l'on peut observer lors d'études de spectroscopie IR ou de calculs de fréquences en chimie quantique.

Torsion Les termes de torsion (ou diédraux propres) sont généralement plus faibles que ceux d'élongation et de flexion. Ils décrivent les barrières de rotation existantes entre quatre atomes liés et sont décrits par une fonction périodique :

$$V_{\text{torsion}} = k_{\text{torsion}} \left[1 + \cos\left(m\varphi - \phi\right)\right] \quad \text{avec} \quad m = 1, 2, \ldots, 6 \tag{A.15}$$

où φ est l'angle dièdre formé par les deux plans d'atomes participant à la rotation considérée. Trois paramètres caractérisent ce potentiel : ϕ l'angle de phase, k_{torsion} la constante d'interaction et m la multiplicité du potentiel de torsion.

Torsion impropre D'autres termes diédraux dits impropres sont parfois utilisés pour maintenir la chiralité ou la planarité autour de certains atomes (*e.g.* planarité des carbones hybridés sp^2). Ces termes diédraux impropres sont représentés par un potentiel harmonique :

$$V_{\text{impropre}} = k_{\text{impropre}} \left(\omega - \omega_{\text{eq}}\right)^2 \tag{A.16}$$

où ω est l'angle entre le plan formé par l'atome central et les deux atomes périphériques, d'une part, et le plan formé par les atomes périphériques, d'autre part.

b. Termes non-liés

Ces termes permettent de prendre en compte les interactions de van der Waals et électrostatiques entre les atomes. En général, elles s'appliquent seulement aux paires d'atomes qui sont séparées par au moins trois liaisons chimiques.

Interaction de van der Waals En mécanique moléculaire, la répulsion à courte distance et la dispersion sont modélisées par un terme d'interaction de van der Waals, le plus souvent exprimé par un potentiel de type Lennard-Jones[306] :

$$V_{\text{van der Waals}} = \frac{B}{r^{12}} - \frac{A}{r^6} = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] \tag{A.17}$$

où r est la distance entre les deux atomes. Les deux paramètres sont ε, la profondeur du puits de potentiel, et σ, la paramètre de collision. Le premier terme rend compte de la répulsion des nuages électroniques à courte distance. Le deuxième terme décrit l'attraction à longue distance due à la dispersion.

Interaction électrostatique Les interactions électrostatiques entre deux atomes i et j sont décrites par la loi de Coulomb :

$$V_{\text{électrostatique}} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{A.18}$$

où q_i et q_j sont les charges des atomes i et j séparés d'une distance r_{ij} et ε_0 la constante diélectrique du vide.

A.2.2. Conditions aux limites périodiques

Dans un système réel, le nombre d'atomes est de l'ordre du nombre d'Avogadro (10^{23}). La modélisation de systèmes aussi grands n'est pas envisageable. Cependant, si on confine le système d'étude dans une boîte de petite dimension, on voit apparaître des effets de surface, certaines entités se trouvant à l'interface matière/vide. Afin de simuler un milieu fini et éviter les effets de surface, la solution réside généralement en l'utilisation de conditions aux limites périodiques (PBC pour *Periodic Boundary Conditions*). Celles-ci consistent à répliquer dans les trois directions de l'espace, la boîte originale à l'infini afin de paver complètement l'espace. Chaque particule de la boîte interagit avec les autres particules, mais également avec leurs images qui se situent dans les boîtes répliquées. Cependant, le caractère pseudo-infini du système ainsi généré nécessite certaines approximations. En particulier, l'approximation dite de l'image minimale suppose que chaque particule i de la cellule centrale interagit uniquement avec l'image la plus proche de toutes les autres particules j. Par ailleurs, l'introduction d'une sphère de troncature (*cut-off*) permet de supprimer les interactions au-delà d'une distance arbitraire, généralement définie comme étant inférieure ou égale à la moitié du plus petit côté de la boîte de simulation. Ces approximations sont d'autant plus valables que la portée des interactions est réduite. Ainsi, les interactions de van der Waals ne soulèvent en général aucune difficulté. Par contre, le traitement des interactions électrostatiques est plus problématique.

A.2.3. Traitement des interactions électrostatiques à longue portée

La manière d'approximer les interactions électrostatiques influe significativement sur le résultat des simulations, notamment pour celles de systèmes chimiques contenant des ions. Par exemple, il s'avère que les artéfacts liés à la troncature simple des interactions électrostatiques par un cut-off peuvent se manifester par des minima ou des maxima prononcés au rayon de coupure r_c dans les fonctions de distributions radiales[307].

a. Somme d'Ewald

La somme d'Ewald[308] permet de calculer précisément les interactions électrostatiques des systèmes périodiques. Elle consiste à évaluer les interactions entre une particule et toutes les autres particules de la boîte de simulation, ainsi que leurs images, dans une matrice infinie de cellules

périodiques. Le potentiel électrostatique s'écrit alors :

$$V_{\text{Coulomb}} = \frac{1}{4\pi\varepsilon_0} \sum_{\mathbf{n}}^{\prime} \sum_{i<j}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \tag{A.19}$$

où N est le nombre de particules dans la boîte de simulation, $\mathbf{n} = n_x L_x + n_y L_y + n_z L_z$ le vecteur réplication de la boîte de dimension (L_x, L_y, L_z), $\mathbf{r}_{ij}$ le vecteur dont $|\mathbf{r}_{ij}|$ est la distance entre les particules i et j de charges respectives q_i et q_j. Cette somme converge lentement et conditionnellement. Elle peut être décomposée en une somme de deux termes :

$$V_{\text{Coulomb}} = \frac{1}{4\pi\varepsilon_0} \sum_{\mathbf{n}}^{\prime} \sum_{i<j}^{N} \frac{q_i q_j S\left(\alpha |\mathbf{r}_{ij} + \mathbf{n}|\right)}{|\mathbf{r}_{ij} + \mathbf{n}|} + \frac{1}{4\pi\varepsilon_0} \sum_{\mathbf{n}}^{\prime} \sum_{i<j}^{N} \frac{q_i q_j \left(1 - S\left(\alpha |\mathbf{r}_{ij} + \mathbf{n}|\right)\right)}{|\mathbf{r}_{ij} + \mathbf{n}|} \tag{A.20}$$

où S est une fonction qui décroît rapidement, par exemple :

$$S(x) = \operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \mathrm{e}^{-u^2} \mathrm{d}u$$

et α, un coefficient d'écrantage choisi de sorte que la première somme converge dans un rayon r tel que $r < \min\left(L_x, L_y, L_z\right)/2$. Ainsi, le premier terme, de courte portée, converge rapidement au sein de la boîte de simulation centrale et est donc sommé dans l'espace direct. Le second contient les contributions à longue portée, convergeant lentement. Sa transformée de Fourier, en revanche, converge rapidement. C'est la raison pour laquelle il est sommé dans l'espace réciproque. Dans le cas où $S(x) = \operatorname{erfc}(x)$, V_{Coulomb} s'écrit :

$$V_{\text{Coulomb}} = \frac{1}{4\pi\varepsilon_0} \sum_{\mathbf{n}}^{\prime} \sum_{i<j}^{N} \frac{q_i q_j \operatorname{erfc}\left(\alpha |\mathbf{r}_{ij} + \mathbf{n}|\right)}{|\mathbf{r}_{ij} + \mathbf{n}|} + \frac{1}{\pi\varepsilon_0} \frac{1}{L^3} \sum_{i<j} \sum_{\mathbf{k} \neq 0}^{\infty} \frac{\mathrm{e}^{\frac{-k^2}{4\alpha^2}}}{k^2} \left| \sum_{i=1}^{N} q_i \mathrm{e}^{-\mathrm{i}\mathbf{k}.\mathbf{r}_i} \right|^2 - \frac{\alpha}{\sqrt{\pi}} \sum_i q_i^2 \tag{A.21}$$

Le dernier terme de l'équation (A.21) est constant et correspond au terme d'auto-couplage.

Le coût en terme de temps de calcul de la somme d'Ewald croît en $O\left(N^2\right)$ où N est le nombre d'atomes dans la boîte de simulation. Ceci nous oblige donc à limiter la taille des systèmes à simuler.

b. Particle-Mesh Ewald

La méthode Particule-Mesh Ewald[269, 270] (PME) a pour but de réduire le temps de calcul des interactions électrostatiques dans le cadre de la somme d'Ewald. Dans cette dernière, l'étape limitante est la transformée de Fourier qui permet de faire converger le deuxième terme de l'équation (A.20) dans l'espace réciproque. Le but de la méthode PME est d'utiliser des transformées de Fourier rapides[309] (FFT). Pour cela, on doit discrétiser le problème ; le volume de la boîte de simulation est divisé en sous-boîtes et on interpole les charges de chacune d'elles sur une grille tridimensionnelle. Ainsi, le temps de calcul des interactions électrostatiques est diminué et croît maintenant en $O(N \log(N))$.

A.2.4. Dynamique moléculaire

La dynamique moléculaire consiste à étudier l'évolution d'un système moléculaire au cours du temps en résolvant les équations de Newton relatives à ce système :

$$m_i \, \mathbf{a}_i = \mathbf{f}_i \tag{A.22}$$

où m_i est la masse d'un atome i, $\mathbf{a}_i$ son accélération et $\mathbf{f}_i$ la force qui lui est appliquée du fait de son interaction avec les autres atomes et l'environnement[304]. Suivant l'hypothèse ergodique, l'étude d'une trajectoire infiniment longue d'un système par dynamique moléculaire revient à échantillonner tout l'espace des phases de ce système. Il est alors possible d'accéder à des grandeurs thermodynamiques telles que les coefficients de diffusion, les fonctions de distributions radiales, l'énergie libre, *etc*.

a. Intégration des trajectoires

La trajectoire de chaque atome i du système est obtenue à partir de l'intégration des équations du mouvement de Newton qui s'écrivent :

$$\begin{cases} \dfrac{\partial \mathbf{r}_i(t)}{\partial t} = \mathbf{v}_i(t) \\[2ex] m_i \dfrac{\partial \mathbf{v}_i(t)}{\partial t} = \mathbf{f}_i(t) \qquad\qquad (i = 1, \ldots, N) \\[2ex] \mathbf{f}_i(t) = -\dfrac{\partial V\left(\mathbf{r}^N\right)}{\partial \mathbf{r}_i}(t) \end{cases} \tag{A.23}$$

où $\mathbf{r}_i$, $\mathbf{v}_i$, m_i et $\mathbf{f}_i$ sont respectivement la position, la vitesse, la masse et la force agissant sur l'atome i et $V\left(\mathbf{r}^N\right)$ la fonctionnelle d'énergie potentielle (champ de forces) du système contenant N atomes. A partir de la connaissance des positions et des vitesses des atomes à l'instant t, l'intégration des équations du mouvement (A.23) permet de déterminer les positions et les vitesses de ceux-ci à l'instant $t + \delta t$. Les algorithmes d'intégration utilisent une méthode des différences finies. La modification des positions de l'ensemble des atomes i est obtenue à partir d'un développement en série de Taylor de la position au temps t :

$$\begin{aligned} \mathbf{r}_i\left(t+\delta t\right) &= \mathbf{r}_i(t) + \delta t \, \mathbf{v}_i(t) + \frac{\delta t^2}{2} \mathbf{a}_i(t) + \ldots \\ &= \mathbf{r}_i(t) + \delta t \, \mathbf{v}_i(t) + \delta t^2 \frac{\mathbf{f}_i(t)}{2 m_i} + \ldots \end{aligned} \tag{A.24}$$

où $\mathbf{r}_i$, $\mathbf{v}_i$, $\mathbf{a}_i$ et $\mathbf{f}_i$ sont les vecteurs position, vitesse, accélération et force de la particule i.

Divers algorithmes d'intégration existent et diffèrent par la manière dont l'expansion de Taylor est effectuée. Les intégrateurs les plus couramment employés en dynamique moléculaire sont :

- Verlet[310] :

$$\mathbf{r}_i\left(t+\delta t\right) = 2\mathbf{r}_i(t) - \mathbf{r}_i\left(t-\delta t\right) + \frac{\mathbf{f}_i(t)}{m_i}\delta t^2 \tag{A.25}$$

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i\left(t+\delta t\right) - \mathbf{r}_i\left(t-\delta t\right)}{2\,\delta t} \tag{A.26}$$

- Leap-Frog[311] (utilisé dans le programme AMBER) :

$$\mathbf{v}_i\left(t+\frac{\delta t}{2}\right) = \mathbf{v}_i\left(t-\frac{\delta t}{2}\right) + \mathbf{f}_i(t)\,\delta t \tag{A.27}$$

$$\mathbf{v}_i(t) = \frac{1}{2}\left(\mathbf{v}_i\left(t+\frac{\delta t}{2}\right) + \mathbf{v}_i\left(t-\frac{\delta t}{2}\right)\right) \tag{A.28}$$

$$\mathbf{r}_i\left(t+\delta t\right) = \mathbf{r}_i(t) + \mathbf{v}_i\left(t+\frac{\delta t}{2}\right)\delta t \tag{A.29}$$

- Predictor-Corrector[312] :

$$\mathbf{r}_i\left(t+\frac{\delta t}{2}\right) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\,\frac{\delta t}{2} \tag{A.30}$$

$$\mathbf{v}_i\left(t+\delta t\right) = \mathbf{v}_i(t) + \mathbf{f}_i\left(t+\frac{\delta t}{2}\right)\frac{\delta t}{m_i} \tag{A.31}$$

$$\mathbf{r}_i\left(t+\delta t\right) = \mathbf{r}_i(t) + \frac{\mathbf{v}_i(t) + \mathbf{v}_i\left(t+\delta t\right)}{2}\delta t \tag{A.32}$$

b. Ensembles thermodynamiques

Les trajectoires de dynamique moléculaire sont généralement réalisées dans l'un des ensembles thermodynamiques suivants :

- microcanonique (NVE) : nombre de particules, volume et énergie du système conservés ;
- canonique (NVT) : nombre de particules, volume et température conservés ;
- isobare-isotherme (NPT) : nombre de particules, pression et température conservés.

Pour reproduire le plus exactement possible les conditions expérimentales, nous avons réalisé des trajectoires de dynamique moléculaire dans l'ensemble NPT. La température et la pression instantanées ont pour expressions :

$$T(t) = \frac{1}{(3N-n)\,k_\mathrm{B}}\sum_{i=1}^{N} m_i\left|\mathbf{v}_i(t)\right|^2 \tag{A.33}$$

$$P(t) = \frac{N}{V(t)}k_\mathrm{B}T(t) + \frac{1}{3V(t)}\sum_{i=1}^{N}\mathbf{r}_i(t)\cdot\mathbf{f}_i(t) \tag{A.34}$$

où N est le nombre d'atomes, $(3N-n)$ le nombre de degrés de liberté du système, k_B la constante de Boltzmann, $V(t)$ le volume instantané de la boîte de simulation et $\mathbf{v}_i(t)$ la vitesse de l'atome i

de masse m_i à l'instant t. Pour contraindre la température et la pression à rester constantes, il suffit d'agir sur les positions et les vitesses des atomes au cours de la trajectoire. Nous détaillons les méthodes du thermostat et du barostat de Berendsen[313] qui sont celles implantées dans AMBER.

c. Thermostat de Berendsen

Dans cette méthode, la vitesse des atomes est modifiée de sorte à relaxer la température instantanée $T(t)$ vers la température T_0 désirée selon :

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{T_0 - T(t)}{\tau_T} \tag{A.35}$$

où τ_T désigne le temps de relaxation associé aux fluctuations de température. Durant un pas d'intégration δt, l'énergie cinétique est modifiée d'une quantité ΔI donnée par :

$$\Delta I = \frac{1}{2}\left(\chi^2 - 1\right) N k_B T(t) \tag{A.36}$$

où χ est le facteur de réajustement des vitesses tel que :

$$\chi = \left[1 + \frac{\delta t}{\tau_T}\left(\frac{T_0}{T(t)} - 1\right)\right]^{1/2} \tag{A.37}$$

d. Barostat de Berendsen

De manière analogue à la température, on souhaite relaxer la pression instantanée $P(t)$ vers la pression de référence P_0 selon :

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = \frac{P_0 - P(t)}{\tau_P} \tag{A.38}$$

où τ_P est le temps de relaxation associé aux fluctuations de pression. En réajustant les coordonnées atomiques et la taille de la cellule périodique par un facteur ς, le volume total se voit modifié de $\Delta V = \left(\varsigma^3 - 1\right) V$ et entraîne une variation de pression :

$$\Delta P = \frac{\Delta V}{\beta_I V} \tag{A.39}$$

où β_I est la compressibilité isotherme. ς s'écrit :

$$\varsigma = \left[1 - \beta_I \delta t \frac{P_0 - P(t)}{\tau_P}\right]^{1/3} \tag{A.40}$$

A.3. PROPKA : Prédiction des valeurs de pK_a des résidus d'une protéine

Les résidus ionisables jouent un rôle important au niveau des propriétés physico-chimiques de la protéine, notamment dans les interactions protéine-ligand ou dans les mécanismes catalytiques[3]. La prédiction théorique de l'état de protonation d'une protéine est rendue nécessaire par le fait que les données cristallographiques ne contiennent pas d'information sur les atomes d'hydrogène. L'utilisation de valeurs modèles de pK_a est fortement déconseillée car les variations peuvent être importantes en fonction des interactions entre résidus.

Il existe différentes méthodes permettant la prédiction de pK_a de protéines. La plupart d'entre-elles[4–7] reposent sur des modèles de continuum électrostatique qui résolvent numériquement l'équation de Poisson-Boltzmann linéarisée (LPBE). Sans entrer dans le détail de ces méthodes, nous pouvons dire que la protéine est décrite par un ensemble d'atomes représentés uniquement par des charges ponctuelles et insérée dans un continuum diélectrique uniforme double (constantes diélectriques différentes pour le solvant et le noyau de la protéine). Le déplacement de pK_a d'un résidu est alors calculé par différence de l'énergie électrostatique entre sa forme chargée et sa forme neutre. Ce déplacement est ajouté à sa valeur modèle de pK_a.

Ces méthodes de prédiction ne sont pas complèment satisfaisantes. Bien qu'elles donnent des valeurs prédites relativement proches des valeurs expérimentales (déviation moyenne inférieure à 1,0 sur un benchmark), il a été montré qu'elles tendaient à sur-estimer les interactions charge-charge intra-protéine et à sous-estimer les effets dus aux liaisons hydrogène et à la désolvatation. Nous pouvons ajouter aussi que la prédiction peut prendre de quelques minutes à plusieurs heures de temps de calcul.

Une autre méthode souvent utilisée pour la prédiction de pK_a de petits composés organiques acido-basiques est la méthode de Hammer-Taft[8]. Elle se base sur une prédiction empirique du déplacement de pK_a vis-à-vis des substituants. Cette méthode présente l'avantage d'être relativement rapide et efficace pour les molécules ayant des similarités avec l'ensemble de paramétrisation. Toutefois, elle ne peut pas être appliquée aux protéines car le déplacement de pK_a des résidus n'est pas dû aux seuls effets des subtituants.

La méthode que nous avons utilisée pour la détermination de l'état de protonation de l'enzyme étudiée est PROPKA[9]. Tout comme la méthode de Hammer-Taft, celle-ci se base sur des relations empiriques entre la structure et le déplacement de pK_a. Les résultats obtenus sont qualitativement

meilleurs car elle prend en compte les effets qui affectent significativement le déplacement de pK_a des résidus tels que la désolvatation, l'interaction charge-charge ou les liaisons hydrogène.

La valeur du pK_a d'un groupe ionisable dans une protéine est prédit en appliquant une perturbation environnementale ΔpK_a à la valeur intrinsèque non-perturbée pK_{Model} de ce groupe.

$$pK_a = pK_{Model} + \Delta pK_a \tag{A.41}$$

Les différentes valeurs de pK_{Model} sont déterminées empiriquement (Tableau A.1) et sont proches de celles utilisées dans d'autres études[314, 315]. Il est connu que pour les groupes ammonium N-terminaux, les groupes carboxyle C-terminaux et les résidus Asp, les effets des substituants du squelette peptidique font diminuer significativement leurs valeurs de pK_a[8]. Ces effets sont déjà inclus dans le terme pK_{Model}. Nous allons voir maintenant les différentes relations empiriques qui interviennent dans le terme ΔpK_a de déplacement du pK_a.

Groupe ionisable	pK_{Model}	Centre du groupe	R_{Local}
C-ter	3,20	$(r_O + r_{OXT})/2$	4,5
Asp	3,80	$(r_{OD1} + r_{OD2})/2$	4,5
Glu	4,50	$(r_{OE1} + r_{OE2})/2$	4,5
His (surface)	6,50	$(r_{CG} + r_{ND} + r_{CE} + r_{NE} + r_{CD})/5$	4,0
His (enfouie)	6,50	$(r_{CG} + r_{ND} + r_{CE} + r_{NE} + r_{CD})/5$	6,0
N-ter	8,00	r_N	4,5
Cys	9,00	r_{SG}	3,5
Tyr	10,00	r_{OH}	3,5
Lys	10,50	r_{NZ}	4,5
Arg	12,50	r_{CZ}	5,0

TAB. A.1: *Valeur de pK_{Model}, définition du centre géométrique et rayon R_{Local} utilisé pour les effets de désolvatation pour chaque type de groupe ionisable.*

Liaisons hydrogène Il a été montré dans des études précédentes[316] que les liaisons hydrogène jouent un rôle déterminant dans la variation du pK_a des résidus Asp et Glu. La relation la plus simple entre le déplacement du pK_a d'un groupe carboxyle et les liaisons hydrogène peut être exprimée par l'équation suivante en considérant un déplacement de pK_a constant (C_{HB}) pour chaque liaison hydrogène :

$$\Delta pK_{HB} = N_{HB} \times C_{HB} \tag{A.42}$$

où N_{HB} est le nombre de liaisons hydrogène et ΔpK_{HB} est le déplacement de pK_a total dû aux liaisons hydrogène. Cependant, l'équation (A.42) ne peut pas toujours être appliquée car la force des liaisons hydrogène et, ainsi, son effet sur le déplacement du pK_a dépendent de la distance et de

l'angle. L'introduction de liaisons hydrogène à caractère faible peut entraîner une sur-estimation du déplacement, et vice-versa. De part les différentes interactions de liaisons hydrogène qui peuvent exister, il est nécessaire d'avoir différentes valeurs de C_{HB}.

Pour prendre en compte les corrections de distance et d'angle pour les effets de liaison hydrogène, une fonction de distance peut être utilisée pour les déplacements dus aux liaisons hydrogène de chaînes latérales (SDC-HB) :

$$\Delta pK_{SDC\text{-}HB} = \begin{cases} C_{HB} & \text{si} \quad D \leq d_1 \\ C_{HB} \times \dfrac{D - d_2}{d_1 - d_2} & \text{si} \quad d_1 < d_2 \\ 0 & \text{si} \quad d_2 \leq D \end{cases} \tag{A.43}$$

et une fonction de distance/angle pour ceux dus aux liaisons hydrogène du squelette de la protéine (BKB-HB) :

$$\Delta pK_{BKB\text{-}HB} = \begin{cases} -\cos\theta \times C_{HB} & \text{si} \quad D \leq d_1 \,,\, \theta > 90^\circ \\ -\cos\theta \times C_{HB} \times \dfrac{D - d_2}{d_1 - d_2} & \text{si} \quad d_1 < d_2 \,,\, \theta > 90^\circ \\ 0 & \text{si} \quad d_2 \leq D \,,\, \theta \leq 90^\circ \end{cases} \tag{A.44}$$

où D est la distance entre les atomes de la liaison hydrogène. Cette distance est définie comme étant celle entre les atomes d'oxygène d'un groupe carboxyle et les protons pour les liaisons hydrogène entre les groupes carboxyles et les groupements de chaîne latérale de type Asn, Gln, Trp, His, Arg et les groupes amides du squelette. Pour les autres liaisons hydrogène, elle est définie comme la distance entre les atomes d'oxygène et les autres atomes lourds (O, S et N). Le paramètre d_1 est la distance optimale de liaison hydrogène pour laquelle la valeur de ΔpK_{HB} est maximale. Généralement, cette distance vaut 2,0 Å si D représente la liaison hydrogène et 3,0 Å si D est la distance entre atomes lourds. Le paramètre d_2 est la distance pour laquelle l'interaction de liaison hydrogène n'existe plus, elle vaut en principe 1,0 Å de plus que la distance d_1. Ces deux paramètres d_1 et d_2 ont été ajustés pour reproduire les valeurs de pK_a expérimentales. Pour la description des liaisons hydrogène avec le squelette de la protéine, θ est défini comme étant le plus grand des deux angles $N\hat{H}O$.

Désolvatation Les groupes carboxyles enfouis à l'intérieur de la protéine montrent souvent des valeurs de pK_a supérieures à pK_{Model}. Il en est de même pour les résidus His qui voient ces valeurs diminuées lorsqu'ils sont enfouis. La désolvatation est le facteur majeur d'un tel déplacement de pK_a. Pour les carboxyles C-terminaux, les résidus Asp, Glu, Cys et Tyr, la désolvatation augmente préférentiellement l'énergie de la forme chargée négativement, et ainsi, accroît la valeur du pK_a. Pour

les amines N-terminales, les résidus His, Lys et Arg, la désolvatation augmente préférentiellement l'énergie de la forme chargée positivement, et donc, abaisse la valeur de pK_a.

Les effets de désolvatation dépendent du degré d'enfouissement du résidu. Pour déterminer celui-ci, la règle suivie est de compter le nombre d'atomes de la protéine entourant le résidu et d'établir alors une relation empirique qui relie le déplacement de pK_a à ce degré d'enfouissement. Deux types de désolvatation sont pris en compte. Une première région autour de 4 à 5 Å (valeurs R_{Local} du tableau A.1) du résidu est d'abord définie. Si aucun atome de la protéine n'est présent, il s'agit d'un groupe ionisable majoritairement entouré de molécules de solvant ; la désolvatation est alors nulle. Par le même raisonnement, plus le nombre d'atomes dans cette région est grand, plus la surface accessible au solvant est petite et donc, plus l'effet de désolvatation se fait ressentir. Cette désolvatation qui se base sur le nombre d'atomes proches d'un résidu est appelée désolvatation locale. La relation empirique qui relie le déplacement de pK_a en désolvatation locale $\Delta pK_{LocalDes}$ et le nombre d'atomes N_{Local} de la région est une constante de déplacement C_{Local} pour chaque atome lourd de la protéine telle que donnée dans l'équation suivante :

$$\Delta pK_{LocalDes} = N_{Local} \times C_{Local} \tag{A.45}$$

La relation (A.45), combinée avec la relation (A.43), permet de donner des prédictions en accord avec l'expérience pour les résidus proches de la surface, mais il n'en est pas de même pour ceux enfouis plus profondemment dans la protéine. Un second effet de désolvatation à plus longue portée doit alors être pris en compte. De façon empirique, un groupe ionisable est considéré comme "enfoui" si le nombre d'atomes $N_{15,5\,Å}$ l'entourant dans une distance de 15,5 Å est supérieur ou égal à 400 ; il est définit comme étant en "surface" dans le cas contraire. De la même manière que pour la désolvatation locale, il est possible de déterminer le déplacement de pK_a dû à cette désolvatation globale $\Delta pK_{GlobalDes}$. La relation empirique est la suivante :

$$\Delta pK_{GlobalDes} = \left(N_{15,5\,Å} - 400\right) \times C_{Global} \tag{A.46}$$

où $\left(N_{15,5\,Å} - 400\right)$ est le nombre d'atomes de la protéine "en excès" et C_{Global} est la constante de déplacement de pK_a définie pour chaque résidu enfoui.

L'effet de désolvatation sur le déplacement de pK_a est la somme des déplacements dus aux désolvatations locales et globales.

$$\Delta pK_{Des} = \Delta pK_{GlobalDes} + \Delta pK_{LocalDes} \tag{A.47}$$

Interactions charge-charge Les interactions charge-charge entre les groupes ionisables interviennent de manière déterminante dans la valeur du pK_a de ces résidus. Cet effet induit un déplacement de pK_a relativement important pour les résidus enfouis ; ceux en surface sont peu affectés par cette interaction. La relation empirique donnant le déplacement de pK_a pour les résidus enfouis en interaction charge-charge ΔpK_{chgchg} ressemble à celle établie précédemment pour les liaisons hydrogène :

$$\Delta pK_{chgchg} = \begin{cases} C_{chgchg} & \text{si} \quad D \leq d_1 \\ C_{chgchg} \times \dfrac{D - d_2}{d_1 - d_2} & \text{si} \quad d_1 < d_2 \\ 0 & \text{si} \quad d_2 \leq D \end{cases} \tag{A.48}$$

où C_{chgchg} est le déplacement de pK_a maximal dû aux interactions charge-charge et D, la distance entre les centres géométriques des groupes ionisables. Les paramètres d_1 et d_2 sont respectivement les distances pour lesquelles cette interaction est maximale et nulle.

Comme il l'a été mentionné précédemment, ce déplacement ΔpK_{chgchg} n'est appliqué que pour les résidus enfouis. On parle alors de paires enfouies. Un critère est donné afin de déterminer si deux résidus (**1** et **2**) sont considérés comme étant une paire enfouie :

$$\left(N_{15,5\,\text{Å}}(\mathbf{1}) \geq 400 \quad \textbf{et} \quad N_{15,5\,\text{Å}}(\mathbf{2}) \geq 400\right) \quad \textbf{ou} \quad \left(\left[N_{15,5\,\text{Å}}(\mathbf{1}) + N_{15,5\,\text{Å}}(\mathbf{2})\right] \geq 900\right) \tag{A.49}$$

Le but de ce critère est de pouvoir prendre en compte les groupes qui ne seraient pas considérés comme étant enfouis ($N_{15,5\,\text{Å}}(\mathbf{1}) = 350$) mais possèdant un voisin plus profondemment enfoui ($N_{15,5\,\text{Å}}(\mathbf{2}) = 550$).

Prise en compte du ligand PROPKA permet désormais de prendre en compte l'effet des ligands présents dans une structure cristallographique[10]. Il faut pour cela identifier les groupes fonctionnels correspondants dans le ligand et, le cas échéant, déterminer les paramètres nécessaires pour le calcul du déplacement de pK_a.

Le protocole retenu consiste à identifier les groupes ionisables tels que les carboxylates ou les azotes de type pyridinium, ammonium, guanidinium et amidinium éventuellement présents sur les ligands. Les valeurs modèles de pK_a de ces groupes sont issues de la littérature ou, à défaut, d'une prédiction empirique. Certains groupes non-ionisables tels que les amines, les alcools, les éthers, les halogènes ou les ions monoatomiques sont aussi identifiés afin de traiter les différentes interactions électrostatiques et/ou par liaison hydrogène. Le traitement des différents termes entrant dans le calcul du déplacement de pK_a (Eq. A.41) se fait de manière similaire à ce que nous avons décrit

précédemment pour la prédiction sans ligand et avec l'utilisation des paramètres spécifiques aux interactions traitées.

Comparaison des méthodes de prédiction

Des tests de prédiction de valeurs de pK_a de différents résidus sur un ensemble de structures cristallographiques ont été réalisés en comparant plusieurs méthodes de prédiction[11,12]. Les résultats montrent que PROPKA est la méthode la plus précise pour tous les types de résidus sauf Glu et His. Les valeurs pour les résidus en surface sont également mieux prédites, contrairement à celles des résidus enfouis qui sont globalement moins bonnes. Aucune méthode ne permet néanmoins de prédire correctement ces valeurs. La rapidité et l'obtention de résultats en accord avec les valeurs expérimentales sont notées comme étant les atouts majeurs de la méthode PROPKA, comparativement aux autres méthodes.

La méthode PROPKA a été codée dans un programme FORTRAN. Il effectue le calcul à partir d'un fichier au format PDB de la protéine, avec ou sans la prise en compte du ligand. Le programme est librement utilisable *via* une interface web à l'adresse `http ://propka.ki.ku.dk/`.

Annexe B

Manuel utilisateur d'AlgoGen

Table des matières

B.1. Lancer une minimisation 246
B.2. Paramètres et options 247
B.2.1. Codage 247
B.2.2. Reproduction 247
B.2.3. Sélection 248
B.2.4. Options de minimisation 248
B.2.5. Options de sortie 252
B.2.6. Options de sauvegarde 252
B.3. Propriétés et méthodes de l'algorithme génétique 253
B.3.1. Propriétés et méthodes de `ga` 253
B.3.2. Propriétés et méthodes des objets `<Population>` 255
B.3.3. Propriétés et méthodes des objets `<Individual>` 255
B.3.4. Propriétés et méthodes des objets `<Chromosome>` 256
B.3.5. Propriétés et méthode des objets `<Gene>` 256
B.4. Définition de fonction 257
B.4.1. Définition de la fonction d'évaluation 257
B.4.2. Définition de la fonction d'arrêt 257
B.4.3. Définition de la fonction de contrôle 257
B.4.4. Définition de la fonction de contrôle de reproduction 258
B.4.5. Définition de la fonction de calcul distribué 258
B.4.6. Définition de la fonction de stratégie d'évolution 259
B.4.7. Définition de la fonction d'affichage 260
B.5. Exemple de script de minimisation 260

B.1. Lancer une minimisation

Cette partie présente le script Python minimal nécessaire pour lancer la minimisation de la fonction de *Schwefel* à trois variables avec $0 \leqslant x_i < 500$.

$$f(x_1, x_2, x_3) = \sum_{i=1}^{3} -x_i \sin\left(\sqrt{|x_i|}\right)$$

Pour cela, il suffit d'affecter à `EvaluationFunction` la fonction à minimiser et de donner à `Ranges` une liste des domaines de définition de chacune des variables comme il est écrit dans le script. La commande `ga.run()` lance la minimisation.

```
# Importation du module genetic.py et math.py
from genetic import parameters, options, ga
import math

# Définition de la fonction à minimiser
def schwefel((x, y, z)):
    f = lambda x: -x * math.sin(math.sqrt(abs(x)))
    return f(x) + f(y) + f(z)

# Définition de l'espace de recherche
parameters.Ranges = [[.0, 500.], [.0, 500.], [.0, 500.]]

# Fonction à minimiser
options.EvaluationFunction = schwefel

# Lancement de la procédure de minimisation
ga.run()
```

Chaque script de minimisation utilisant l'algorithme génétique doit importer les trois objets du module `genetic.py` :

`parameters` contient les informations propres au déroulement de la minimisation telles que le codage des gènes, le mode de reproduction et le mode de sélection.

`options` contient les informations concernant tout ce qui est lié au processus de minimisation avec, entre autres, la définition de la fonction d'évaluation, le contrôle de la minimisation ou la création de fichiers de sortie.

`ga` contient les fonctions et les procédures nécessaires à la minimisation.

L'importation du module `math` est nécessaire à l'écriture de la fonction de *Schwefel*. L'exécution de cette minimisation a été facilitée grâce à l'existence de valeurs par défaut pour un certain nombre de paramètres et options.

B.2. Paramètres et options

B.2.1. Codage

`GeneSize` Taille des gènes

- Liste d'entiers définissant la longueur d'un gène en nombre de bits.
  La taille de la liste doit être égale au nombre de gènes, *i.e.* égale au nombre de variables à minimiser.

```
parameters.GeneSize = <list>
```

⇒ Valeur par défaut = `[16,...,16]`

`Ranges` Intervalles de définition

- Liste définissant les intervalles de définition de chaque variable codée par le chromosome d'un individu.
  Chaque élément de la liste est une liste de deux réels correspondant respectivement à la borne inférieure et supérieure de l'intervalle de définition de chaque variable à coder.
  Le nombre d'éléments dans la liste impose le nombre de variables à optimiser.

```
parameters.Ranges = <list>
```

⇒ Pas de valeur par défaut ⇒ Paramètre obligatoire

B.2.2. Reproduction

`CrossoverPartNumber` Nombre de partitions pour le croisement

- Entier définissant le nombre de partitions créées dans chaque gène lors du croisement.

```
parameters.CrossoverPartNumber = <int>
```

⇒ Valeur par défaut = 1

`CrossoverProbability` Probabilité de croisement

- Réel définissant la probabilité de croisement.
  Cette valeur doit être comprise entre 0 et 1.

```
parameters.CrossoverProbability = <float>
```

⇒ Valeur par défaut = `1.`

`CrossoverMaxNumber` Nombre maximal de croisements par partition

- Entier définissant le nombre maximal de croisements à effectuer sur chaque partition.

```
parameters.CrossoverMaxNumber = <int>
```

⇒ Valeur par défaut = 1

`MutationPartNumber` Nombre de partitions pour la mutation

- Entier définissant le nombre de partitions créées dans chaque gène lors de la mutation.

```
parameters.MutationPartNumber = <int>
```

⇒ Valeur par défaut = 1

`MutationProbability` Probabilité de mutation

- Réel définissant la probabilité de mutation.
  Cette valeur doit être comprise entre 0 et 1.

```
parameters.MutationProbability = <float>
```

⇒ Valeur par défaut = `.02`

`MutationMaxNumber` Nombre maximal de mutations par partition

- Entier définissant le nombre maximal de mutations à effectuer sur chaque partition.

```
parameters.MutationMaxNumber = <int>
```

⇒ Valeur par défaut = 1

`InversionProbability` Probabilité d'inversion

- Réel définissant la probabilité d'inversion.
  Cette valeur doit être comprise entre 0 et 1.

```
parameters.InversionProbability = <float>
```

⇒ Valeur par défaut = `.01`

`PopulationSize` Taille de la population

- Entier définissant la taille de la population à construire lors de la reproduction.

```
parameters.PopulationSize = <int>
```

⇒ Valeur par défaut = 32

B.2.3. Sélection

`ElitismRatio` Taux d'élitisme

- Réel définissant la fraction conservée des meilleurs individus de la population.

```
parameters.ElitismRatio = <float>
```

⇒ Valeur par défaut = `.5`

`TournamentNumber` Nombre de tournois

- Entier définissant le nombre de tournois à créer lors de la sélection par tournois, *i.e.* le nombre d'individus sélectionnés au final.

```
parameters.TournamentNumber = <int>
```

⇒ Valeur par défaut = 8

B.2.4. Options de minimisation

`Initialization` Création de la population initiale

- Mot-clé `'random'`.
  Le programme crée aléatoirement une population initiale d'individus. La taille de cette population est égale à **`PopulationSize`**. Les chromosomes des individus sont représentatifs de points dans tout l'espace de recherche défini par **`Ranges`**.

```
options.Initialization = 'random'
```

- Nom du fichier de sauvegarde.
  On peut redémarrer une minimisation à partir d'une sauvegarde effectuée précédemment. On fournit alors à **`Initialization`** le nom du fichier, sous la forme d'une chaîne de caractères, contenant cette sauvegarde.

```
options.Initialization = <string>
```

- Liste de réels correspondant à une solution approchée.
  La taille de la liste doit être égale au nombre de variables du problème à traiter. Chaque réel appartient au domaine de définition défini dans **`Ranges`**.

```
options.Initialization = <list>
```

⇒ Valeur par défaut = `'random'`

`MaxGenerations` Nombre maximal de générations

- Entier correspondant au nombre maximal de générations à produire, *i.e.* nombre maximal de cycles de reproduction. Dès que le nombre de générations produites est égal à ce nombre, le cycle de reproduction est arrêté et le programme exécute les procédures de terminaison. Ce test est effectué après évaluation de la population.
  Les options **`MaxEvaluations`** et **`MaxGenerations`** fonctionnent en parallèle. Le cycle de reproduction s'arrête au premier critère atteint.

```
options.MaxGenerations = <int>
```

⇒ Valeur par défaut = `100`

`MaxEvaluations` Nombre maximal d'évaluations

- Entier correspondant au nombre maximal d'évaluations à effectuer durant la minimisation.
  Dès que le nombre d'évaluations effectuées est supérieur ou égal à ce nombre, le cycle de reproduction est arrêté et le programme exécute les procédures de terminaison. Ce test est effectué après évaluation de la population.
  Les options **`MaxEvaluations`** et **`MaxGenerations`** fonctionnent en parallèle. Le cycle de reproduction s'arrête au premier critère atteint.

```
options.MaxEvaluations = <int>
```

⇒ Valeur par défaut = `10000`

`EvaluationFunction` Fonction d'évaluation

- Fonction externe définie par l'utilisateur (Voir section B.4.1).
  Cette fonction est appelée pour évaluer chaque individu.
  La fonction fournie à **`EvaluationFunction`** prend en argument un tuple de longueur égale au nombre de variables codées par le chromosome et retourne le résultat de l'évaluation de celui-ci.
  Les options **`EvaluationFunction`** et **`ResultsGrabber`** sont exclusives. Il n'est pas nécessaire de renseigner l'option **`EvaluationFunction`** si l'option **`ResultsGrabber`** est utilisée.

```
options.EvaluationFunction = <function>
```

⇒ Pas de valeur par défaut ⇒ Paramètre obligatoire

StopFunction Fonction d'arrêt

- Fonction externe définie par l'utilisateur (Voir section B.4.2).
  L'appel à cette fonction se fait après l'évaluation des individus pour permettre de stopper la minimisation.
  La fonction fournie à **StopFunction** doit retourner `False` tant que les critères d'arrêt que l'on se donne ne sont pas atteints. Si ces critères sont atteints, la fonction doit retourner `True`. Le cycle de reproduction est arrêté et le programme exécute les procédures de terminaison.

```
options.StopFunction = <function>
```

⇒ Valeur par défaut = `False`

ControlFunction Fonction de contrôle

- Fonction externe définie par l'utilisateur (Voir section B.4.3).
  L'appel à cette fonction se fait avant et après l'évaluation des individus à chaque génération, selon la valeur prise par la propriété **EvaluationFlag** de l'algorithme génétique.
  L'option **ControlFunction** introduit une fonction externe que l'utilisateur définit comme il le souhaite à l'intérieur du cycle de reproduction pour contrôler les populations générées.
  Selon la définition qu'on lui donne, cette fonction de contrôle peut servir au calcul de statistiques sur une population évaluée ou à placer des contraintes de certaines variables avant l'évaluation.
  La fonction ne retourne rien.

```
options.ControlFunction = <function>
```

⇒ Valeur par défaut = `pass`

ReproductionControl Fonction de contrôle de reproduction

- Fonction externe définie par l'utilisateur B.4.4).
  L'appel à cette fonction se fait pendant les opérations de reproduction, à chaque fois qu'un individu est créé (Voir section.
  L'option **ReproductionControl** introduit une fonction externe que l'utilisateur définit comme il le souhaite afin de contrôler les opérations de reproduction.
  Cette fonction de contrôle sert essentiellement à supprimer les nouveaux individus produits avant la phase d'évaluation. Elle renvoie `True` si l'individu est conservé, `False` sinon.

```
options.ReproductionControl = <function>
```

⇒ Valeur par défaut = `True`

ResultsGrabber Fonction de calcul distribué

- Fonction externe définie par l'utilisateur (Voir section B.4.5).
  L'appel à cette fonction se fait au moment de l'évaluation des individus à chaque génération de l'algorithme génétique.
  L'option **ResultsGrabber** introduit une fonction externe que l'utilisateur définit comme il le souhaite pour permettre la distribution des calculs d'évaluation sur plusieurs machines d'un cluster.

```
options.ResultsGrabber = <function>
```

⇒ Valeur par défaut = Utilisation de **EvaluationFunction**

Unicity Unicité des individus

- Booléen conditionnant le phénomène d'eugénisme au sein d'une population.

Si **Unicity** vaut True, alors dans une génération donnée, les chromosomes de chaque individu sont tous distincts. Si **Unicity** vaut False, alors une même génération peut comporter des individus de même chromosome (ils sont alors identiques en terme de génotype).

```
options.Unicity = <boolean>
```

⇒ Valeur par défaut = True

RandomSeed Racine du générateur de nombres aléatoires

- Entrée (nombre, liste, chaîne de caractères, *etc*) déterminant la racine du générateur, *i.e.* le premier terme de la suite de récurrence utilisée pour la production des nombres aléatoires.
  Si la racine du générateur **RandomSeed** est la même lors du lancement de plusieurs minimisations, les suites des nombres aléatoires produites sont identiques. Si **RandomSeed** vaut None, le temps courant est pris comme racine du générateur.

```
options.RandomSeed = <HashableType>
```

⇒ Valeur par défaut = None

Periodicity Application de l'opérateur de mutation par addition

- Liste d'entiers indiquant les index des gènes sur lesquels il faut appliquer l'opérateur de mutation par addition.
  L'opérateur de mutation par addition remplace celui de mutation par permutation sur les gènes considérés.

```
options.Periodicity = <list>
```

⇒ Valeur par défaut = [1,2,3,...]

Strategy Stratégie d'évolution de l'algorithme génétique

- Fonction externe définie par l'utilisateur (Voir section B.4.6).
  L'appel à cette fonction se fait au début de chaque cycle de reproduction, avant les phases de sélection et de reproduction. Elle permet de concevoir le processus d'obtention de la nouvelle génération.
  Cette fonction sert principalement à définir quelle part des individus sélectionnés est réintégrée à la nouvelle génération et de quelle manière les nouveaux sont formés à partir des parents.

```
options.Strategy = <function>
```

⇒ Fonctions pré-définies :

ga.newOne Création d'individus
: Les nouveaux individus sont créés par reproduction des parents sélectionnés.

ga.newUnknown Création d'individus originaux
: Les nouveaux individus créés par reproduction n'ont jamais été produits dans les générations précédentes.

ga.unknownSurvivor Création d'individus originaux et survie des parents
: Les nouveaux individus créés par reproduction n'ont jamais été produits dans les générations précédentes. Les parents (individus sélectionnés) sont intégrés à la nouvelle population.

ga.oldSurvivor Création d'individus originaux, survie et vieillissement des parents
: Les nouveaux individus créés par reproduction n'ont jamais été produits dans les générations précédentes. Les parents (individus sélectionnés) sont intégrés à la nouvelle population, à condition qu'ils ne dépassent pas une limite d'âge.

⇒ Valeur par défaut = ga.unknownSurvivor

Gray Conversion en code Gray du code binaire des gènes

- Booléen conditionnant la conversion en code Gray du code binaire des gènes.
  Si **Gray** vaut `True`, le code Gray est appliqué.
  Les options **Gray** et **Periodicity** sont exclusives. Il est impossible d'appliquer ces deux options en même temps.

```
options.Gray = <boolean>
```

⇒ Valeur par défaut = `False`

B.2.5. Options de sortie

OutputFile Fichier de sortie

- Nom du fichier de sortie.
  L'affichage de sortie de la minimisation est enregistré dans un fichier dont on donne le nom sous la forme d'une chaîne de caractères.

```
options.OutputFile = <string>
```

- Mot-clé `'stdout'`.
  L'affichage de sortie de la minimisation se fait directement sur l'écran de l'ordinateur (sortie standard).

```
options.OutputFile = 'stdout'
```

⇒ Valeur par défaut = `'stdout'`

OutputPrint Fonction d'affichage de sortie

- Fonction externe définie par l'utilisateur (Voir section B.4.7).
  La fonction fournie à **OutputPrint** génère l'affichage de sortie. Elle doit retourner une chaîne de caractères formatée qui sera écrite sur le media sélectionné.

```
options.OutputPrint = <function>
```

⇒ Affichage par défaut :
- Numéro de génération.
- Compteur des évaluations (évaluation courante et totale) et longueur de l'historique des individus.
- Liste des individus de la population courante triés par ordre croissante de leur évaluation.
- Meilleur individu de la population courante.
- Meilleure solution actuelle.

OutputFrequency Fréquence d'écriture en sortie

- Entier définissant la fréquence, en terme de générations, à laquelle un affichage de sortie est généré par le programme.

```
options.OutputFrequency = <int>
```

⇒ Valeur par défaut = 1

B.2.6. Options de sauvegarde

BackupFile Fichier de sauvegarde

- Nom du fichier de sauvegarde.

La population courante et l'historique des individus sont enregistrées dans un fichier dont on donne le nom sous la forme d'une chaîne de caractères.

```
options.BackupFile = <string>
```

⇒ Valeur par défaut = `'ga.bkp'`

`BackupFrequency` Fréquence d'écriture en sauvegarde

- Entier définissant la fréquence, en terme de générations, à laquelle un point de sauvegarde est créé par le programme.

```
options.BackupFrequency = <int>
```

⇒ Valeur par défaut = 5

B.3. Propriétés et méthodes de l'algorithme génétique

Pour l'écriture de quelques fonctions présentées dans la section précédente, il est nécessaire d'avoir accès à certaines propriétés contenues dans l'algorithme génétique. Ces propriétés sont accessibles depuis l'objet `ga` qui contient l'intégralité du processus de minimisation du programme. Nous voyons maintenant les différentes propriétés et méthodes qu'il est possible d'obtenir.

B.3.1. Propriétés et méthodes de `ga`

`Population` Population courante

- La propriété **`Population`** retourne un objet `<Population>` qui contient le liste des individus de la population courante (Voir section B.3.2).

```
ga.Population -> <Population>
```

- On accède à un individu de la population par l'intermédiaire de son index. On récupère ainsi un individu de type `<Individual>` (Voir section B.3.3).

```
ga.Population[<int>] -> <Individual>
```

- On obtient la taille de la population par la fonction `len()`.

```
len(ga.Population) -> <int>
```

`History` Historique des individus

- La propriété **History** retourne un dictionnaire qui contient l'ensemble des chromosomes, ainsi que leur évaluation, générés depuis le début du cycle de reproduction.

```
ga.History -> <dict>
```

- La clé d'accès à ce dictionnaire est la propriété **`BinaryCode`** d'un chromosome. On obtient en retour l'évaluation correspondante à celui-ci.

```
ga.History[<Chromosome>.BinaryCode] -> <float>
```

- On accède à la taille de l'historique par la fonction `len()`.

```
len(ga.History) -> <int>
```

RunFlag Drapeau de minimisation en cours

- La propriété **RunFlag** retourne True si la minimisation est en cours d'exécution, sinon elle retourne False.

```
ga.RunFlag -> <boolean>
```

EvaluationFlag Drapeau d'évaluation de la population

- La propriété **EvaluationFlag** retourne True si la population a été entièrement évaluée, sinon elle retourne False.

```
ga.EvaluationFlag -> <boolean>
```

- La propriété **EvaluationFlag** doit être mentionnée lors de la définition de la fonction de contrôle (voir section B.4.3).

User Espace utilisateur

- La propriété **User** retourne un dictionnaire initialement vide.

```
ga.User -> <dict>
```

- C'est un espace de nom réservé à l'utilisateur. Il sert à stocker des variables de différents types. On peut y affecter ou récupérer les données par l'intermédiaire d'une chaîne de caractères.

```
ga.User[<string>] = <UserType>
ga.User[<string>] -> <UserType>
```

Index Numero de la génération courante

- La propriété **Index** retourne le numero de la génération en cours d'existence.

```
ga.Index -> <int>
```

CurrentEvaluationNumber Evaluations courantes

- La propriété **CurrentEvaluationNumber** retourne le nombre d'évaluations effectuées pour obtenir la population courante.

```
ga.CurrentEvaluationNumber -> <int>
```

TotalEvaluationNumber Evaluations totales

- La propriété **TotalEvaluationNumber** retourne le nombre d'évaluations effectuées depuis la première génération.

```
ga.TotalEvaluationNumber -> <int>
```

BestIndividual Meilleur individu

- La propriété **BestIndividual** retourne un individu de type <Individual> ayant la meilleure évaluation de la génération courante.

```
ga.BestIndividual -> <Individual>
```

B.3.2. Propriétés et méthodes des objets `<Population>`

`Evaluations` Evaluations de la population

- La propriété **`Evaluations`** retourne une liste contenant les évaluations des individus de la population.

```
<Population>.Evaluations -> <list>
```

`RandomIndividual` Individu aléatoire

- La propriété **`RandomIndividual`** retourne un individu de type `<Individual>` choisit aléatoirement dans la population.

```
<Population>.RandomIndividual -> <Individual>
```

`sort()` Tri de la population

- La méthode **`sort()`** trie les individus de la population dans l'ordre croissant de leur évaluation.

```
<Population>.sort()
```

`copy()` Copie de la population

- La méthode **`copy()`** retourne une copie de la population.

```
<Population>.copy() -> <Population>
```

B.3.3. Propriétés et méthodes des objets `<Individual>`

`Evaluation` Evaluation de l'individu

- La propriété **`Evaluation`** retourne la valeur d'évaluation de l'individu.

```
<Individual>.Evaluation -> <float>
```

`Chromosome` Chromosome de l'individu

- La propriété **`Chromosome`** retourne un objet `<Chromosome>` correspondant au chromosome de l'individu (Voir section B.3.4).

```
<Individual>.Chromosome -> <Chromosome>
```

- On accède à un gène du chromosome par l'intermédiaire de son index. On obtient alors un objet <Gene> (Voir section B.3.5).

```
<Individual>.Chromosome[<int>] -> <Gene>
```

- Le nombre de gènes contenus dans le chromosome est obtenu par la fonction `len()`.

```
len(<Individual>.Chromosome) -> <int>
```

`copy()` Copie de l'individu

- La méthode **`copy()`** retourne une copie de l'individu.

```
<Individual>.copy() -> <Individual>
```

B.3.4. Propriétés et méthodes des objets `<Chromosome>`

`BinaryCode` Code binaire du chromosome

- La propriété **`BinaryCode`** renvoie la chaîne de caractères correspondant au code binaire du chromosome.

```
<Chromosome>.BinaryCode -> <string>
```

`HexCode` Code hexadécimal du chromosome

- La propriété **`HexCode`** renvoie la chaîne de caractères correspondant au code hexadécimal du chromosome.

```
<Chromosome>.HexCode -> <string>
```

`decode()` Décodage du chromosome

- La méthode **`decode()`** retourne une liste contenant le décodage des variables codées par le chromosome.

```
<Chromosome>.decode() -> <list>
```

`recode(index, newReal)` Recodage d'un gène du chromosome

- La méthode **`recode()`** recode un gène du chromosome, repéré par son index (**`index`**), avec une nouvelle valeur (**`newReal`**).

```
<Chromosome>.recode(<int>, <float>)
```

`copy()` Copie du chromosome

- La méthode **`copy()`** retourne une copie du chromosome.

```
<Chromosome>.copy() -> <Chromosome>
```

B.3.5. Propriétés et méthode des objets `<Gene>`

`BinaryCode` Code binaire du gène

- La propriété **`BinaryCode`** retourne la chaîne de caractères correspondant au code binaire du gène.

```
<Gene>.BinaryCode -> <string>
```

`HexCode` Code hexadécimal du gène

- La propriété **`HexCode`** retourne la chaîne de caractères correspondant au code hexadécimal du gène.

```
<Gene>.HexCode -> <string>
```

`copy()` Copie du gène

- La méthode **`copy()`** retourne une copie du gène.

```
<Gene>.copy() -> <Gene>
```

B.4. Définition de fonction

B.4.1. Définition de la fonction d'évaluation

La fonction d'évaluation (**`EvaluationFunction`**) doit prendre en argument un tuple des *n* variables codées par le chromosome et renvoyer la valeur prise par cette fonction pour les *n* variables considérées.

Exemple Définition de la fonction `carre` réalisant le calcul de $f(x,y) = x^2 + y^2$

```
def carre((x, y)) :
    return x**2 + y**2
```

B.4.2. Définition de la fonction d'arrêt

La fonction d'arrêt (**`StopFunction`**) doit renvoyer `True` si les critères d'arrêt sont atteints, sinon elle doit renvoyer `False`.

Exemple Définition de la fonction `stop` pour arrêter le cycle de reproduction de l'algorithme si le chromosome du meilleur individu de la génération en cours est présent dans au moins 90 % des individus de la population et pendant au moins 5 générations successives.

```
def stop():
    def _individualCount(individual, population):
        count = 0
        for ind in population:
            if repr(ind) == repr(individual):
                count += 1
        return count
    if ga.Index == 0:
        ga.User['StopFlag'] = 0
    if _individualCount(ga.BestIndividual, ga.Population) > int(parameters.
        PopulationSize * .9):
        ga.User['StopFlag'] += 1
    else:
        ga.User['StopFlag'] = 0
    if ga.User['StopFlag'] == 5:
        return True
    else:
        return False
```

Remarque Nous avons utilisé une variable nommée `'StopFlag'` de l'espace de nom réservé aux utilisateurs pour créer un compteur propre à cette fonction d'arrêt.

B.4.3. Définition de la fonction de contrôle

La fonction de contrôle (**`ControlFunction`**) exécute des instructions propres à l'utilisateur à l'intérieur même du cycle de reproduction des individus. Cette fonction ne doit rien retourner au programme.

Exemple Définition de la fonction `constraint` pour contraindre un gène de tous les individus de la population à exprimer une seule et même valeur fixée (contrainte) après reproduction et calculer la moyenne des évaluations après la phase d'évaluation.

```
def constraint():
    if ga.EvaluationFlag == False:
        for individual in ga.Population:
            individual.Chromosome.recode(1, 100.)
    else:
```

```
        sum = .0
        for value in ga.Population.Evaluations:
            sum += value
        import sys
        sys.stdout.write('%15.3f' % (sum / len(ga.Population.Evaluations)))
```

Remarque Nous rappelons que cette fonction de contrôle sera appelée par le programme après l'étape de reproduction (individus en attente d'évaluation) et après l'étape d'évaluation (individus évalués). Pour distinguer les opérations à effectuer par cette fonction, nous utilisons donc la propriété **`EvaluationFlag`** qui indique l'état de la population vis à vis de l'évaluation.

B.4.4. Définition de la fonction de contrôle de reproduction

La fonction de contrôle de reproduction (**`ReproductionControl`**) permet la suppression d'individus produits pendant la phase de reproduction, avant la phase d'évaluation

Exemple Définition de la fonction `controlRoot` qui permet de supprimer les individus représentatifs de $\sqrt{\sum x_i^2} = 1$.

```
def controlRoot(individual):
    (x, y, z) = individual.Chromosome.decode()
    import math
    if math.sqrt(x**2 + y**2 + z**2) == 1. :
        return False
    return True
```

- La fonction prend en argument un objet `<Individual>`.
- Après décodage du chromosome de l'individu, la fonction renvoie `False` si la condition de suppression est vérifiée.

B.4.5. Définition de la fonction de calcul distribué

La fonction de calcul distribué (**`ResultsGrabber`**) permet la soumission de calculs sur un cluster de machines, éventuellement par l'intermédiaire d'un gestionnaire de queue de calculs.

Exemple Définition de la fonction `grabResults`, utilisée dans le cadre du docking, pour l'évaluation de l'énergie des systèmes protéine-ligand correspondant à chaque individu d'une population. Elle fait appel au programme SGE de gestion de queue de calculs pour la distribution des différentes évaluations sur un cluster de machines.

```
def grabResults(compList):
    dirList = []

# Creation du systeme Prot-Lig pour chaque individu
    for comp in compList:
        (x, y, z, phi, theta, psi) = comp.Values
        quat = quaternion.fromEuler((phi, theta, psi))
        newLigand = Ligand.copy()
        newLigand.rot(quat)
        newLigand.Vector = quaternion.Qvector([x, y, z])
        newLigand.updateCoords()

# Creation d'un repertoire temporaire
        tmpDir = tempfile.mkdtemp('_divcon', 'tmp_',scrDir)
        dirList.append((tmpDir,comp.CompID))
        os.chdir(tmpDir)

# Ecriture du script de soumission SGE
        divcon_sub = open('divcon_sub.sh','w')
        divcon_sub.write("""#!/bin/sh
## Deplace le fichier divcon dans le TMPFS
```

```
fromDir=$PWD
toDir=/tmp/tmpfs/$JOB_ID
mkdir $toDir
cp $PWD/divcon.in $toDir
cd $toDir
## Run Divcon
$HOME/bin/divcon-mais
## Recupere le fichier divcon
cp $toDir/divcon.out $fromDir
rm -fr $toDir
""" % ())
        divcon_sub.close()

# Ecriture du fichier d'entree DivCon
        divcon.writeDivconFile(concat(Prot, newLigand))

# Soumission SGE
        import os
        osReturn = os.system('qsub -q juliet -S /bin/sh -cwd divcon_sub.sh')
        time.sleep(1.)

# Recuperation des evaluations
    IDList = {}
    while len(dirList):
        time.sleep(20.)
        os.system('qstat -u thiriot')
        othList = []
        for (tmpDir,ID) in dirList:
            print tmpDir
            try:
                outFile = open('%s/divcon.out' % (tmpDir),'r')
            except IOError:
                othList.append((tmpDir,ID))
            else:
                content = outFile.read()
                outFile.close()
                if content.find('TOTAL TIME') > 0:
                    IDList[ID] = divcon.readEnergy(content)
                    os.system('rm -fr %s' % (tmpDir))
                else:
                    othList.append((tmpDir,ID))
        dirList = othList
    for comp in compList:
        comp.Result = IDList[comp.CompID]
```

- La fonction prend en argument une liste d'objets `<Computation>` (`compList`). Elle est fournie par l'algorithme génétique après l'étape de décodage des individus. Chaque objet de type `<Computation>` contient les valeurs décodées dans la propriété `Values` et l'évaluation à renvoyer à l'algorithme génétique dans la propriété `Result`.
- L'algorithme génétique appelle la fonction **`ResultsGrabber`** lors de la demande des calculs d'évaluation et se met en attente de la fin d'exécution de celle-ci pour poursuivre la minimisation. L'écriture de cette fonction repose ainsi sur une boucle interne `while` : une fois que tous les calculs d'évaluation ont été réalisés, les résultats peuvent être retournés.

B.4.6. Définition de la fonction de stratégie d'évolution

La fonction de stratégie d'évolution (**`Strategy`**) indique le processus par lequel la population de la nouvelle génération est construite à partir des individus sélectionnés.

Exemple Définition de la fonction `unknownSurvivor`, issue de l'objet `ga`, permettant de réintégrer les individus sélectionnés et d'obtenir des individus originaux dans la nouvelle population.

```
def unknownSurvivor(self):
    def _isReallyNew(ind):
```

```
        for individual in ga.Population:
            if ind.Chromosome.BinaryCode == individual.Chromosome.BinaryCode:
                return False
        return True
    oldPopulationSize = parameters.PopulationSize
    selectedPopulation = ga._select(self.Population)
    parameters.PopulationSize = oldPopulationSize - parameters.TournamentNumber
    newPopulation = ga._reproduce(selectedPopulation)
    ga.Population = selectedPopulation.copy()
    while parameters.PopulationSize:
        for individual in newPopulation:
            if (not ga.History.has_key(individual.Chromosome.BinaryCode) and
                _isReallyNew(individual)):
                ga.Population.append(individual)
                parameters.PopulationSize -= 1
        newPopulation = ga._reproduce(selectedPopulation)
    parameters.PopulationSize = oldPopulationSize
```

- La fonction ne prend pas d'argument et ne renvoie rien.
- La population est sélectionnée par l'intermédiaire de la primitive de sélection `_select()` et la nouvelle population est obtenue par la primitive de reproduction `_reproduce()`.

B.4.7. Définition de la fonction d'affichage

La fonction d'affichage (**`OutputFunction`**) doit renvoyer une chaîne de caractères formatée à enregistrer sur le media de sortie.

Exemple Définition de la fonction `printString` pour afficher le numéro de génération, les différents compteurs, la liste des individus de la génération en cours, le meilleur individu et son décodage.

```
def printString():
    string = ""
    string += "\n---------------\n"
    string += "Generation %s :\n" % (ga.Index)
    string += "---------------\n\n"
    string += "Current evaluations       : %6d\n" % (ga.CurrentEvaluationNumber)
    string += "Total evaluations         : %6d\n" % (ga.TotalEvaluationNumber)
    string += "Individual history length : %6d\n" % (len(ga.History))
    string += "\n---------------\n\n"
    string += "Population :\n\n"
    ga.Population.sort()
    for individual in ga.Population:
        string += repr(individual) + "\n"
    string += "\nBest individual :\n\n"
    string += repr(ga.BestIndividual) + "\n"
    string += "\nBest solution :\n\n"
    string += str(ga.BestIndividual.Chromosome.decode()) \
                + " | " + str(ga.BestIndividual.Evaluation) + "\n"
    return string
```

B.5. Exemple de script de minimisation

Nous reprenons l'exemple de la minimisation de la fonction de *Schwefel* à 3 variables avec $0 \leq x_i < 500$ où figurent la plupart des paramètres et options de l'algorithme génétique :

```
# Importation des modules
from genetic import parameters, options, ga
import math

# Definition de la fonction d'evaluation
```

```
def schwefel((x, y, z)):
    f = lambda x: -x * math.sin(math.sqrt(abs(x)))
    return f(x) + f(y) + f(z)

# Definition de la fonction d'arret qui renvoie toujours False
stop = lambda: False

# Definition de la fonction de contrele qui ne fait rien
def control():
    pass

# Definition de la fonction d'affichage
def schPrint():
    string = ""
    string += "%s = %15.8f\n" % (ga.BestIndividual.Chromosome.decode(), ga.BestIndividual.
        Evaluation)
    return string

# Codage des genes sur 8 bits, chacun defini sur [0,500[
parameters.GeneSize = [8, 8, 8]
parameters.Ranges = [[.0, 500.], [.0, 500.], [.0, 500.]]

# 1 croisement avec une probabilite de 90% sur un chromosome divise en 3 partitions
parameters.CrossoverPartNumber = 3
parameters.CrossoverProbability = .9
parameters.CrossoverMaxNumber = 1

# 4 mutations avec une probabilite de 2% sur un chromosome divise en 5 partitions
parameters.MutationPartNumber = 5
parameters.MutationProbability = .02
parameters.MutationMaxNumber = 4

# Probabilite d'inversion de 1%
parameters.InversionProbability = .01

# Population de 32 individus
parameters.PopulationSize = 32

# Après avoir conserve 80% de la population, on organise 8 tournois
parameters.ElitismRatio = .8
parameters.TournamentNumber = 8

# Population initiale aleatoire
options.Initialization = 'random'

# Maximum de 200 generations et de 1000 evaluations
options.MaxGenerations = 200
options.MaxEvaluations = 1000

# Minimisation de la fonction de Schwefel
options.EvaluationFunction = schwefel

# Arret par la fonction stop et controle par la fonction control
options.StopFunction = stop
options.ControlFunction = control

# Eugenisme interdit et pas de code Gray
options.Unicity = False
options.Gray = False

# Mutation par addition sur le 3eme gene
options.Periodicity = [2]

# Strategie d'evolution newUnknown
options.Strategy = ga.newUnknown

# Pas de racine spécifique au générateur aléatoire
options.RandomSeed = None

# Ecriture de la chaîne de caractères produite par la fonction schPrint dans le fichier
    schwefel.out à toutes les générations
```

```
options.OutputFile = 'schwefel.out'
options.OutputPrint = schPrint
options.OutputFrequency = 1

# Sauvegarde de la population dans le fichier schwefel.bkp toutes les 20 générations
options.BackupFile = 'schwefel.bkp'
options.BackupFrequency = 20

# Lancement de la minimisation
ga.run()
```

Annexe C

Manuel utilisateur d'AlgoGen-DivCon

Table des matières

C.1. Exemple de script de docking **264**
C.2. Importation des modules **266**
C.3. Fonctions et attributs de `docking` **266**
C.3.1. Fonctions 266
C.3.2. Attributs 266
C.4. Attributs de `divcon` **267**

C.1. Exemple de script de docking

Le script Python suivant est un exemple de script de lancement du logiciel ALGOGEN-DIVCON correspondant au docking du dianion urate dans le site actif de l'urate oxydase. Les différentes commandes et options sont présentées ensuite.

```
# Import des modules
import math
import docking
from genetic import ga, options, parameters
from divcon import divcon

#================#
# Initialisation #
#================#

# Definition des repertoires de travail et de scratch
wrkDir = ''
scrDir = '/auto/juliet7/thiriot/scratch/URATE'

######

if not len(wrkDir):
    wrkDir = os.getcwd()
if not len(scrDir):
    scrDir = wrkDir

#=======================================#
# Creation du systeme Proteine - Ligand #
#=======================================#

# Identifiant Docking
title = 'UOX-URA_PM3-2'

# Taille de la boite de docking
boxSize = 12.

# Residus Site actif
actSite = [228,229,255,177,160,355]

# PDB de la proteine
protPDB = 'uox.pdb'

# PDB du ligand
ligPDB = 'urate.pdb'

# Charge du ligand
ligChrg = -2

#####

# Lecture de la PDB de la proteine
prot = docking.rdPDB('%s/%s' % (wrkDir,protPDB))
# Centre sur le site actif
prot.center(actSite)
# Decoupe des residus
prot = docking.cut(prot,prot.Vector,boxSize*math.sqrt(.75))
# Lecture de la PDB du ligand
ligand = docking.rdPDB('%s/%s' % (wrkDir,ligPDB))
# Creation de la boite cubique de docking
box = [boxSize]*3

#======================#
# Configuration DivCon #
#======================#

# Methode semi-empirique
divcon.Method = 'PM3'
```

```
# Charge totale du systeme
divcon.Charge = getCharge(prot) + ligChrg

# Parametres Divide&Conquer
divcon.Cluster = [1, 4., 2.]

#=====================================#
# Configuration Algorithme genetique #
#=====================================#

# Parametres des genes
parameters.GeneSize = [10, 10, 10, 8, 8, 8]
parameters.Ranges = [[-box[0]/2, box[0]/2], [-box[1]/2, box[1]/2],
    [-box[2]/2, box[2]/2], [0., math.pi], [0., 2*math.pi], [0., 2*math.pi]]

# Parametres de selection
parameters.ElitismRatio = 1.
parameters.TournamentNumber = 16

# Parametres de reproduction
parameters.CrossoverProbability = 1.
parameters.CrossoverMaxNumber = 1
parameters.CrossoverPartNumber = 1
parameters.MutationProbability = .5
parameters.MutationMaxNumber = 1
parameters.MutationPartNumber = 1
parameters.InversionProbability = .0
parameters.PopulationSize = 32

# Options de l'algorithme génétique
options.Initialization = 'UOX-URA_PM3-1.bkp'
options.MaxGenerations = 50
options.Unicity = True
options.OutputPrint = docking.printing
options.OutputFile = '%s/%s.out' % (wrkDir, title)
options.Gray = False

options.BackupFile = '%s/%s.bkp' % (wrkDir, title)
options.BackupFrequency = 1

options.Strategy = ga.unknownSurvivor
options.Periodic = [3,4,5]
options.ReproductionControl = docking.controlDistance

options.ResultsGrabber = docking.grabResults

#===========#
# Execution #
#===========#

# Lancement du docking
ga.run()
```

C.2. Importation des modules

En préambule du script doit figurer l'importation des modules Python suivants :

`docking` contient les fonctions et les procédures propres au déroulement du docking moléculaire.

`divcon` contient les procédures relatives aux entrées/sorties du programme DIVCON.

`genetic` contient les procédures relatives aux entrées/sorties du programme ALGOGEN.

Lors de l'importation des modules `divcon.py` et `genetic.py`, il faut importer les singletons `divcon`, d'une part, et les singletons `ga`, `parameters` et `options`, d'autre part. L'importation du module `math.py` permet d'accéder à la constante π nécessaire pour la définition du domaine de variation des angles.

C.3. Fonctions et attributs de `docking`

C.3.1. Fonctions

`rdPDB(pdbFile)` Lecture d'un fichier PDB

- Crée un système d'atomes à partir de la lecture d'un fichier au format PDB nommé **`pdbFile`**. La fonction renvoie un objet **`<AtomSet>`** correspondant aux atomes du fichier PDB.

```
docking.rdPDB(pdbFile) -> <AtomSet>
```

`cut(atomSet, center, radius)` Extraction des atomes

- Crée un système d'atomes en conservant les résidus d'un autre système d'atomes **`atomSet`** dont les atomes sont à l'intérieur d'une sphère de rayon **`radius`** centrée à la position **`center`**. La fonction renvoie un objet **`<AtomSet>`**.

```
docking.cut(atomSet, center, radius) -> <AtomSet>
```

`getCharge(atomSet)` Charge d'un système d'atomes

- Calcule la charge d'un système d'atomes **`atomSet`** (protéine) en se basant sur le nom des résidus.
  La fonction renvoie un réel.

```
docking.getCharge(atomSet) -> <float>
```

C.3.2. Attributs

`Prot` Partie protéine

- Système d'atomes correspondant à la partie protéine.
  Cet attribut permet d'indiquer quel système d'atomes représente la partie fixe du docking.

```
docking.Prot = <AtomSet>
```

$\Rightarrow$ Pas de valeur par défaut

Lig Partie ligand

- Système d'atomes correspondant à la partie ligand.
  Cet attribut permet d'indiquer quel système d'atomes représente la partie mobile du docking.

```
docking.Lig = <AtomSet>
```

⇒ Pas de valeur par défaut

C.4. Attributs de divcon

Method Méthode semi-empirique

- Nom de la méthode semi-empirique utilisée pour l'évaluation énergétique des conformations protéine-ligand.

```
divcon.Method = <string>
```

⇒ Fonctions pré-définies :
- PM3
- PM3 PIF
- PM3 PIF WALID
- AM1

⇒ Pas de valeur par défaut.

Charge Charge du système protéine-ligand

- Réel donnant la charge nette du système protéine-ligand pour l'évaluation énergétique des conformations.

```
divcon.Charge = <float>
```

⇒ Pas de valeur par défaut.

Cluster Paramètres Divide & Conquer

- Liste d'un entier et de deux réels donnant respectivement le nombre de résidus par cœur, la taille de la première région-tampon et la taille de la deuxième région-tampon par l'algorithme à croissance linéaire Divide & Conquer.

```
divcon.Cluster = [<int>, <float>, <float>]
```

⇒ Valeur par défaut = Calcul standard (Pas Divide & Conquer)

Lexique

Outils méthodologiques

AMBER *Assisted Model Building and Energy Refinement*
Ensemble de programmes permettant la réalisation de simulations de dynamique moléculaire et ensemble de champs de force de mécanique moléculaire.

B3LYP *Becke 3-Parameter, Lee, Yang and Parr*
Fonctionnelle d'échange-correlation hybride constituée de la fonctionnelle d'échange (B3) et de la fonctionnelle de corrélation (LYP).

DFT *Density Functional Theory*
Méthodes utilisant la densité électronique pour le calcul des propriétés moléculaires.

HF *Hartree-Fock*
Equations et méthode pour le calcul d'orbitales moléculaires.

MD *Molecular Dynamics*
Méthode permettant de décrire l'évolution dans le temps d'un système moléculaire.

MP2 *2nd order Møller-Plesset*
Méthode perturbative pour le calcul de l'énergie de corrélation au 2ème ordre.

NPT Ensemble thermodynamique isobare-isotherme
Nombre de particules, pression et température du système moléculaire conservés.

NVE Ensemble thermodynamique microcanonique
Nombre de particules, volume et énergie du système moléculaire conservés.

NVT Ensemble thermodynamique canonique
Nombre de particules, volume et température du système moléculaire conservés.

PME *Particle-Mesh Ewald*
Algorithme de traitement des termes électrostatiques à longue portée par combinaison de la somme d'Ewald et de transformées de Fourier rapides.

SCF *Self-Consistent Field*
Procédure itérative de résolution des équations Hartree-Fock.

Composés chimiques

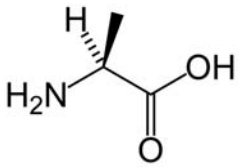
DMSO Diméthylsulfoxyde

MetSO Sulfoxyde de méthionine

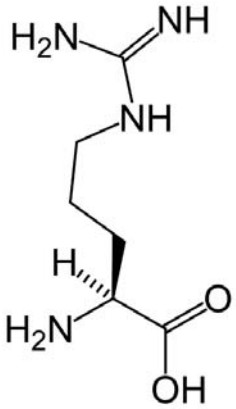
Msr Méthionine sulfoxyde réductase
Enzyme de réduction catalytique des sulfoxydes de méthionine.

Acides aminés

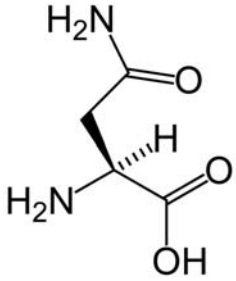
Acide aminé	Code à 3 lettres	Code à 1 lettre	pK_a de la chaîne latérale
Alanine	Ala	A	—
Arginine	Arg	R	12,48
Asparagine	Asn	N	—
Aspartate	Asp	D	3,90
Cystéine	Cys	C	8,18
Glutamate	Glu	E	4,07
Glutamine	Gln	Q	—
Glycine	Gly	G	—
Histidine	His	H	6,04
Isoleucine	Ile	I	—
Leucine	Leu	L	—
Lysine	Lys	K	10,54
Méthionine	Met	M	—
Phénylalanine	Phe	F	—
Proline	Pro	P	—
Sérine	Ser	S	—
Thréonine	Thr	T	—
Tryptophane	Trp	W	—
Tyrosine	Tyr	Y	10,46
Valine	Val	V	—



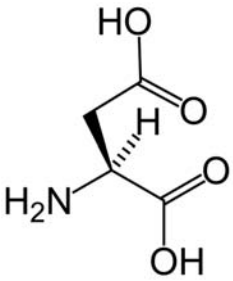
Alanine



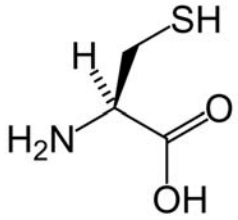
Arginine



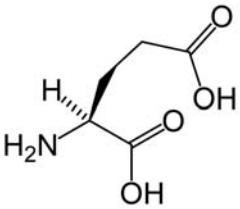
Asparagine



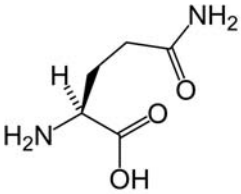
Aspartate



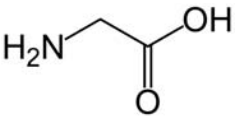
Cystéine



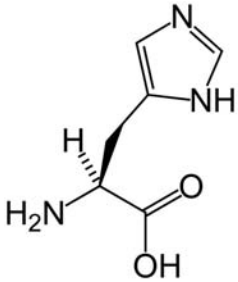
Glutamate



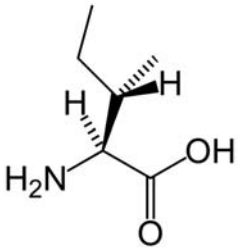
Glutamine



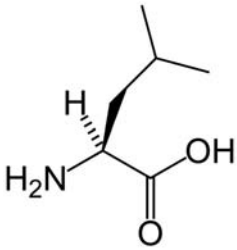
Glycine



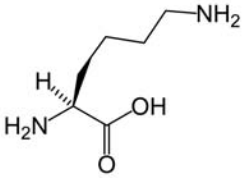
Histidine



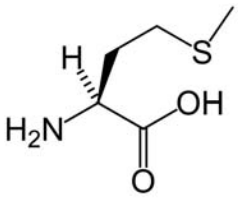
Isoleucine



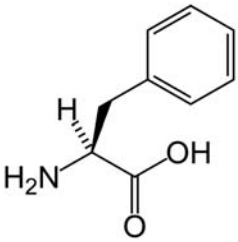
Leucine



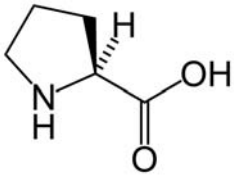
Lysine



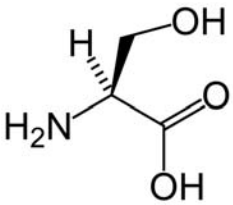
Méthionine



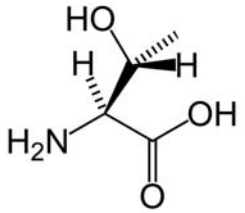
Phénylalanine



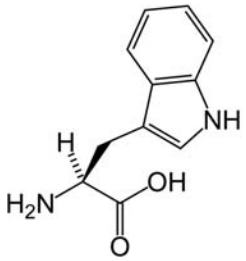
Proline



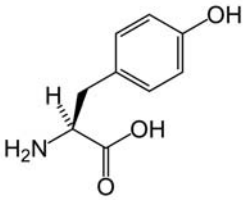
Sérine



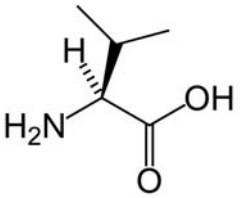
Thréonine



Tryptophane



Tyrosine



Valine

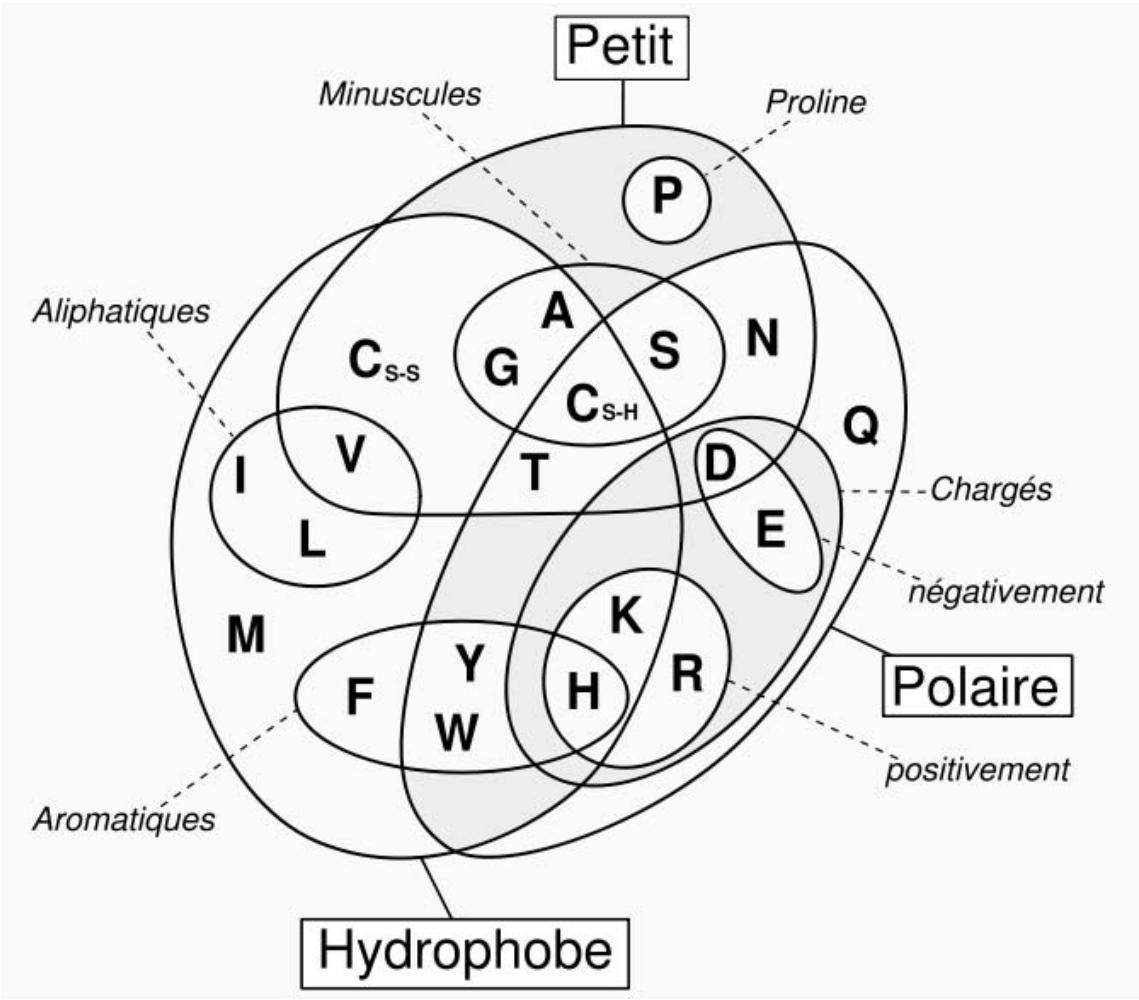


Diagramme de Venn des propriétés physico-chimiques des acides aminés.

Table des figures

I-1. Schéma du bilan énergétique d'une catalyse enzymatique 15

2.1. Analogie entre évolution biologique naturelle et algorithme génétique 32
2.2. Processus de minimisation dans les algorithmes génétiques 33
2.3. Fonction $f(x) = x^2$ à minimiser définie dans l'intervalle entier $[0, 31]$. 34
2.4. Schéma de la sélection de type *roulette-wheel* 36
2.5. Diagramme UML simplifié du module `genetic.py`. 40
2.6. Schéma de fonctionnement de la sélection par tournois 45

3.1. Diagramme UML simplifié du module `docking.py`. 57
3.2. Diagramme UML de séquence des procédures du programme AlgoGen-DivCon 61

4.1. Evolution de l'énergie d'interaction du dimère de l'eau en fonction de la méthode semi-empirique utilisée 68
4.2. Structures optimisées du dimere de l'eau 69
4.3. Catalyse enzymatique d'oxydation de l'acide urique en allantoïne 70
4.4. Structure de l'urate oxydase d'*Aspergillus flavus* 72
4.5. Structures chimiques de la 8-azaxanthine, de l'acide oxonique et de l'acide urique 73
4.6. Structure du complexe UOx-AZA 76
4.7. Structure du complexe UOx-OXC 79
4.8. Structure du complexe UOx-URA 81

5.1. Différentes voies d'oxydation d'une méthionine 89
5.2. Etat de transition de l'oxydation d'un sulfure par le peroxyde d'hydrogène 91
5.3. Formes limites de mésomérie de la fonction sulfoxyde 94
5.4. Réduction stéréospécifique de MetSO par les Msr de classe A ou B 96
5.5. Cycle catalytique des méthionine sulfoxyde réductases 98
5.6. Mécanisme catalytique des méthionine sulfoxyde réductases 99
5.7. Alignement des séquences de différentes MsrA 100
5.8. Alignement des séquences de différentes MsrB 101
5.9. Structure cristallographique de MsrA 104
5.10. Structure cristallographique de MsrB 107
5.11. Chemin réactionnel de la réduction du DMSO par le méthanethiol (1) 111
5.12. Chemin réactionnel de la réduction du DMSO par le méthanethiol (2) 113
5.13. Chemin réactionnel de la réduction du DMSO par le méthanethiol (3) 113
5.14. Chemin réactionnel de la réduction du DMSO par le méthanethiol (4) 114

6.1. Nomenclature des atomes du substrat DMSO 121
6.2. RMS de déviation de la chaîne principale de l'enzyme 125
6.3. Distances Cys51-H_γ – Glu94-$O_{\varepsilon 2}$, Cys51-H_γ – Tyr197-O_η et angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ dans Free-CysH-Glu$^\ominus$ 127

6.3. (suite) Distances Glu94-$O_{\varepsilon1}$ – Tyr82-H_η, Glu94-$O_{\varepsilon1}$ – Tyr134H_η, Glu94-$O_{\varepsilon2}$ – Tyr82-H_η, Glu94-$O_{\varepsilon2}$ – Tyr134-H_η et angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon1}$ dans Free-CysH-Glu$^\ominus$ . . . 128
6.4. Distances Cys51-H_γ – Glu94-$O_{\varepsilon1}$, Cys51-S_γ – Glu94-$H_{\varepsilon2}$ et Cys51-H_γ – Tyr197-O_η dans Free-CysH-GluH . . . 130
6.4. (suite) Solvatation des résidus Cys51 et Glu94 et distances Tyr82-H_η – Glu94-$O_{\varepsilon1}$ et Tyr134-H_η – Glu94-$O_{\varepsilon1}$ dans Free-CysH-GluH . . . 131
6.5. Interaction de Cys51-S_γ avec Glu94-$H_{\varepsilon2}$ et Tyr197-H_η dans Free-Cys$^\ominus$-GluH . . . 133
6.5. (suite) Interaction de Tyr82-H_η et Tyr134-H_η avec Glu94-$O_{\varepsilon1}$ dans Free-Cys$^\ominus$-GluH . . . 134
6.6. Distances Tyr82-H_η – DMSO-O, Tyr134-H_η – DMSO-O et Cys51-H_γ – DMSO-O dans DMSO-CysH-Glu$^\ominus$ . . . 137
6.6. (suite) Distances Cys51-S_γ – DMSO-S et Trp53-$N_{\varepsilon1}$ – DMSO-C_β et angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ dans DMSO-CysH-Glu$^\ominus$ . . . 138
6.6. (suite) Distances Cys51-H_γ – Glu94-$O_{\varepsilon2}$, Tyr82-H_η – Glu94-$O_{\varepsilon1}$, Tyr134-H_η – Glu94-$O_{\varepsilon1}$, Tyr82-H_η – Glu94$O_{\varepsilon2}$ et Tyr134-H_η – Glu94-$O_{\varepsilon2}$ dans DMSO-CysH-Glu$^\ominus$ . . . 139
6.6. (suite) Nombre de molécules d'eau dans la solvatation de DMSO, Cys51 et Glu94 dans DMSO-CysH-Glu$^\ominus$ . . . 140
6.7. Distances Glu94-$H_{\varepsilon2}$ – DMSO-O, Tyr82-H_η – DMSO-O et Tyr134-H_η – DMSO-O dans DMSO-CysH-GluH . . . 143
6.7. (suite) Distances Cys51-H_γ – DMSO-O, Cys51-H_γ – Tyr197-O_η et angle dièdre Cys51-C_α – Cys51-C_β – Cys51-S_γ – Cys51-H_γ dans DMSO-CysH-GluH . . . 144
6.7. (suite) Distances Cys51-S_γ – DMSO-S, Trp53-$N_{\varepsilon1}$ – DMSO-C_α, Trp53-$N_{\varepsilon1}$ – DMSO-C_β et angle dièdre Cys51-C_β – -Cys51-S_γ – DMSO-S – DMSO-C_β dans DMSO-CysH-GluH . . . 145
6.7. (suite) Distances Tyr82-H_η – Glu94-$O_{\varepsilon1}$, Tyr134-H_η – Glu94-$O_{\varepsilon1}$, Tyr82-H_η – Glu94-$O_{\varepsilon2}$, Tyr134-H_η – Glu94-$O_{\varepsilon2}$ et angle Cys51-S_γ – DMSO-S – DMSO-O dans DMSO-CysH-GluH . . . 146
6.7. (suite) Nombre de molécules d'eau dans la solvatation de DMSO, Cys51 et Glu94 dans DMSO-CysH-GluH . . . 147
6.8. Structure du complexe michaélien du DMSO dans le site actif de MsrA . . . 149
6.9. Distances Glu94-$H_{\varepsilon2}$ – DMSO-O, Tyr82-H_η – DMSO-O et Tyr134-H_η – DMSO-O dans DMSO-Cys$^\ominus$-GluH . . . 151
6.9. (suite) Distances Cys51-S_γ – DMSO-S, Cys51-S_γ – Tyr197-H_η et Trp53-$N_{\varepsilon1}$ – DMSO-C_β dans DMSO-Cys$^\ominus$-GluH . . . 152
6.9. (suite) Angles dièdres Cys51-C_β – Cys51-S_γ – DMSO-S – DMSO-C_β, Cys51-H_α – Cys51-C_α – Cys51-C_β – Cys51-S_γ et nombre de molécules d'eau dans la solvatation de Cys51 dans DMSO-Cys$^\ominus$-GluH . . . 153
6.9. (suite) Distances Tyr82-H_η – Glu94-$O_{\varepsilon1}$, Tyr134-H_η – Glu94-$O_{\varepsilon1}$, Tyr82-H_η – Glu94-$O_{\varepsilon2}$, Tyr134-H_η – Glu94-$O_{\varepsilon2}$ et angle Cys51-S_γ – DMSO-S – DMSO-O dans DMSO-Cys$^\ominus$-GluH . . . 154

7.1. Mécanisme supposé de l'étape réductase (Activation) . . . 158
7.2. Structure du modèle quantique de site actif . . . 161
7.3. Complexe michaélien **1** : Structure et tracé des orbitales HOMO et LUMO . . . 165
7.4. Structure de l'état de transition $\mathbf{TS_{1-2}}$ et de l'intermédiaire sulfurane **2** . . . 168
7.5. Tracé des orbitales HOMO et LUMO pour l'intermédiaire sulfurane **2** . . . 170
7.6. Tracé de l'isosurface de densité électronique de **2** . . . 171
7.7. Nomenclature des atomes du substrat DMSO protoné . . . 173
7.8. Distances Glu94-$O_{\varepsilon1}$ – DMSOH-H, Cys51-S_γ – DMSOH-S et Cys51-S_γ – Tyr197-H_η dans DMSOH-Cys$^\ominus$-Glu$^\ominus$ . . . 176
7.8. (suite) Distances Tyr82-H_η – Glu94-$O_{\varepsilon1}$, Tyr134-H_η – Glu94-$O_{\varepsilon1}$, Tyr82-H_η – Glu94-$O_{\varepsilon2}$, Tyr134-H_η – Glu94-$O_{\varepsilon2}$ et angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon1}$ dans DMSOH-Cys$^\ominus$-Glu$^\ominus$ . . . 177
7.8. (suite) Distances Tyr82-H_η – DMSOH-O, Tyr134-H_η – DMSOH-O et Tyr134-H_η – Cys51-S_γ dans DMSOH-Cys$^\ominus$-Glu$^\ominus$ . . . 178
7.8. (suite) Distances Trp53-$N_{\varepsilon1}$ – DMSOH-C_α, Trp53-$N_{\varepsilon1}$ – DMSOH-C_β, angle Cys51-S_γ – DMSOH-S – DMSOH-O et angle dièdre Cys51-C_β – Cys51-S_γ – DMSOH-S – DMSOH-C_β dans DMSOH-Cys$^\ominus$-Glu$^\ominus$ . . . 179
7.9. Nomenclature des atomes du résidu sulfurane . . . 180

7.10. Distances Sulf-HO – Glu94-$O_{\varepsilon 1}$, Sulf-HO – Glu94-$O_{\varepsilon 2}$ et angle dièdre Glu94-C_β – Glu94-C_γ – Glu94-C_δ – Glu94-$O_{\varepsilon 1}$ dans Sulf-Glu$^\ominus$ . . . 185
7.10. (suite) Distances Sulf-O_ε – Tyr82-H_η, Sulf-O_ε – Tyr134-H_η et Sulf-S_γ – Tyr197-H_η dans Sulf-Glu$^\ominus$ 186
7.10. (suite) Distances Trp53-$N_{\varepsilon 1}$ – Sulf-$C_{\varepsilon 1}$, Trp53-$N_{\varepsilon 1}$ – Sulf$C_{\varepsilon 2}$, Tyr82-H_η – Glu94-$O_{\varepsilon 2}$, Tyr134-H_η – Glu94-$O_{\varepsilon 2}$ et angle dièdre Sulf-C_β – Sulf-S_γ – Sulf-S_δ – Sulf-$C_{\varepsilon 1}$ dans Sulf-Glu$^\ominus$ . . . 187
7.11. Mécanisme d'activation de l'étape réductase du mécanisme catalytique . . . 188

8.1. Hypothèse de travail : Transfert d'un groupement OH (Mécanisme I) . . . 190
8.2. Structures **A**, $\mathbf{TS_{A\text{-}B}}$, **B**, $\mathbf{TS_{B\text{-}C}}$ et **C** . . . 193
8.3. Structure de l'intermédiaire **3** . . . 198
8.4. Hypothèse de travail : 1 H^+ + 1 e^- (Mécanisme II) . . . 200
8.5. Structure de l'intermédiaire sulfurane équatorial dissocié dans le modèle de site actif (**3**) . . . 207
8.6. Hypothèse de travail : 2 H^+ + 2 e^- (Mécanisme III) . . . 209
8.7. Structure de l'état de transition $\mathbf{TS_{2-4}}$ et de l'intermédiaire sulfurane dissocié **4** dans le modèle de site actif . . . 211
8.8. Hypothèse de mécanisme concerté imaginé pour la formation de l'intermédiaire acide sulfénique à partir de l'intermédiaire sulfurane dissocié . . . 215

A.1. Représentation schématique des différents termes intervenant dans un champ de forces . . . 231

Liste des tableaux

2.1. Population initiale de quatre individus . . . 35
2.2. Résultat de la sélection d'individus . . . 36
2.3. Création d'une nouvelle population d'individus . . . 38

4.1. Géométrie optimisée d'une molécule d'eau en PM3 et AM1 . . . 67
4.2. Energies d'interaction et paramètres structuraux du dimére de l'eau . . . 69
4.3. Energies d'interaction et paramètres structuraux du complexe UOx-AZA . . . 75
4.4. Energies d'interaction et paramètres structuraux du complexe UOx-OXC . . . 79
4.5. Energies d'interaction et paramètres structuraux du complexe UOx-URA . . . 81

5.1. Principales espèces oxygénées activées . . . 88
5.2. Structures de méthionine sulfoxyde réductases . . . 102

6.1. Analyse PROPKA des résidus Cys51 et Glu94 de MsrA . . . 118
6.2. Résumé des simulations réalisées . . . 120
6.3. Paramètres du champ de force du substrat DMSO . . . 122
6.4. Analyse PROPKA des résidus Cys51 et Glu94 de MsrA . . . 123
6.5. Valeurs des charges atomiques pour DMSO-CysH-Glu$^{\ominus}$ . . . 136

7.1. Valeurs des distances et des angles pour les structures **1**, $\mathbf{TS_{1-2}}$ et **2** . . . 163
7.2. Valeurs des ordres de liaison et des charges pour les structures **1**, $\mathbf{TS_{1-2}}$ et **2** . . . 164
7.3. Bilan énergétique du transfert de proton dans le modèle quantique de site actif . . . 164
7.4. Comparaison entre la structure **1** et la cristallographie . . . 166
7.5. Paramètres du champ de forces du substrat DMSO protoné . . . 174
7.6. Paramètres du champ de forces du résidu sulfurane . . . 181
7.7. Paramètres du champ de forces du résidu sulfurane (suite) . . . 182

8.1. Valeurs des distances, des angles, des ordres de liaison, des charges et bilan énergétique pour les structures **A**, $\mathbf{TS_{A\text{-}B}}$, **B**, $\mathbf{TS_{B\text{-}C}}$ et **C** . . . 192
8.2. Valeurs des distances, des angles, des ordres de liaison et des charges pour la structure **3** . . . 199
8.3. Potentiels d'ionisation et affinités électroniques du méthanethiol et du DMSO . . . 202
8.4. Potentiel d'ionisation de CH_3–S^- et affinité électronique de $(CH_3)_2S^+$–OH en fonction du solvant 203
8.5. Potentiel rédox absolu des couples CH_3–$S^\bullet$/CH_3–S^- et $(CH_3)_2S^+$–OH/$(CH_3)_2S^\bullet$–OH en fonction du solvant . . . 205
8.6. Valeurs des distances et des angles pour les structures **1**, $\mathbf{TS_{1-2}}$, **2**, $\mathbf{TS_{2-4}}$ et **4** . . . 212
8.7. Valeurs des ordres de liaison et des charges pour les structures **1**, $\mathbf{TS_{1-2}}$, **2**, $\mathbf{TS_{2-4}}$ et **4** . . . 213
8.8. Bilan énergétique du transfert de proton dans le modèle de site actif . . . 213
8.9. Expérience de marquage isotopique du substrat sulfoxyde de méthionine . . . 215

A.1. Valeur de pK_{Model}, définition du centre géométrique et rayon R_{Local} utilisé pour les effets de désolvatation pour chaque type de groupe ionisable. . . . 240

Bibliographie

1. Voet D. et Voet J. G., *Biochemistry ;* Wiley International Editions : New York, 4th ed. ; 2004.
2. Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N. et Bourne P. E., *Nucleic Acids Res.* **28**, 235–242 (2000).
3. Warshel A., *Acc. Chem. Res.* **14**, 284–290 (1981).
4. Bashford D. et Karplus M., *Biochemistry* **29**, 10219–10225 (1990).
5. Yang A.-S., Gunner M. R., Sampogna R., Sharp K. et Honig B., *Proteins* **15**, 252–265 (1993).
6. Yang A.-S. et Honig B., *J. Mol. Biol.* **231**, 459–474 (1993).
7. Antosiewicz J., Briggs J. M., Elcock A. H., Gilson M. K. et McCammon J. A., *J. Comput. Chem.* **17**, 1633–1644 (1996).
8. Perrin D. D., Dempsey B. et Serjeant E. P., *pK_a Prediction for Organic Acids and Bases ;* Chapman & Hall : London, 1981.
9. Li H., Robertson A. D. et Jensen J. H., *Proteins* **61**, 704–721 (2005).
10. Bas D. C., Rogers D. M. et Jensen J. H., *Proteins* (2008).
11. Davies M. N., Toseland C. P., Moss D. S. et Flower D. R., *BMC Biochemistry* **7**, 18 (2006).
12. Stanton C. L. et Houk K. N., *J. Chem. Theory Comput.* **4**, 951–966 (2008).
13. Kollman P. A., Massova I., Reyes C., Kuhn B., Huo S., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D. A., et Cheatham, III T. E., *Acc. Chem. Res.* **33**, 889–897 (2000).
14. Kuhn B. et Kollman P. A., *J. Med. Chem.* **43**, 3786–3791 (2000).
15. Huo S., Massova I. et Kollman P. A., *J. Comput. Chem.* **23**, 15–27 (2001).
16. Wang J., Kang X., Kuntz I. D., et Kollman P. A., *J. Med. Chem.* **48**, 2432–2444 (2005).
17. Mobley D. L., Graves A. P., Chodera J. D., McReynolds A. C., Shoichet B. K. et Dill K. A., *J. Mol. Biol.* **371**, 1118–1134 (2007).
18. Feliu Maseras K. M., *J. Comput. Chem.* **16**, 1170–1179 (1995).
19. Humbel S., Sieber S. et Morokuma K., *J. Chem. Phys.* **105**, 1959–1967 (1996).
20. Matsubara T., Sieber S. et Morokuma K., *Int. J. Quant. Chem.* **60**, 1101–1109 (1996).
21. Svensson M., Humbel S., Froese R. D. J., Matsubara T., Sieber S. et Morokuma K., *J. Phys. Chem.* **100**, 19357–19363 (1996).
22. Svensson M., Humbel S. et Morokuma K., *J. Chem. Phys.* **105**, 3654–3661 (1996).
23. Dapprich S., Komáromi I., Byun K. S., Morokuma K. et Frisch M. J., *J. Mol. Struct.-THEOCHEM* **462**, 1–21 (1999).
24. Vreven T. et Morokuma K., *J. Comput. Chem.* **21**, 1419–1432 (2000).
25. Warshel A. et Levitt M., *J. Mol. Biol.* **103**, 227–249 (1976).
26. Field M. J., Bash P. A. et Karplus M., *J. Comput. Chem.* **11**, 700–733 (1990).

27. Monard G., Loos M., Théry V., Baka K. et Rivail J.-L., *Int. J. Quant. Chem.* **58**, 153–159 (1996).
28. Antonczak S., Monard G., Ruiz-López M. F. et Rivail J.-L., *J. Am. Chem. Soc.* **120**, 8825–8833 (1998).
29. Ferré N., Assfeld X. et Rivail J.-L., *J. Comput. Chem.* **23**, 610–624 (2002).
30. Monard G., Prat-Resina X., González-Lafont A. et Lluch J. M., *Int. J. Quant. Chem.* **93**, 229–244 (2003).
31. Amara P. et Field M. J., *Theor. Chem. Acc.* **109**, 43–52 (2003).
32. Martins-Costa M. T. C. et Ruiz-López M. F., *Chem. Phys.* **332**, 341–347 (2007).
33. Yang W. et Lee T.-S., *J. Chem. Phys.* **103**, 5674–5678 (1995).
34. Stewart J. J. P., *Int. J. Quant. Chem.* **58**, 133–146 (1996).
35. Li X.-P., Nunes R. W. et Vanderbilt D., *Phys. Rev. B* **47**, 10891–10894 (1993).
36. Rosen E. D., O'Donnell A. L. et Koenig R. J., *Mol. Cell. Endocrinol.* **78**, C83-C88 (1991).
37. Roberts M. R., *Curr. Opin. Plant Biol.* **3**, 400–405 (2000).
38. Pawson T. et Nash P., *Genes Dev.* **14**, 1027–1047 (2000).
39. Filman D. J., Bolin J. T., Matthews D. A. et Kraut J., *J. Biol. Chem.* **257**, 13663–13672 (1982).
40. Héroux A., White E. L., Ross L. J., Kuzin A. P. et Borhani D. W., *Structure* **8**, 1309–1318 (2000).
41. Zacharias N. et Dougherty D. A., *Trends Pharmacol. Sci.* **23**, 281–287 (2002).
42. Manssour Fraga C. A. et Barreiro E. J., *Curr. Drug Ther.* **3**, 1–13 (2008).
43. Marco E. et Gago F., *ChemMedChem* **2**, 1388–1401 (2007).
44. Fox T., Brennan D., Austen D. A., Swalley S. E., Coll J. T., Raybuck S. A. et Chambers S. P., *Curr. Protein Pept. Sc.* **8**, 439–445 (2007).
45. Vakser I. A., *Biopolymers* **39**, 455–464 (1996).
46. Lengauer T. et Rarey M., *Curr. Opin. Struct. Biol.* **6**, 402–406 (1996).
47. Kitchen D. B., Decornez H., Furr J. R. et Bajorath J., *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
48. Morris G. M., Goodsell D. S., Halliday R. S., Huey R., Hart W. E., Belew R. K. et Olson A. J., *J. Comput. Chem.* **19**, 1639–1662 (1998).
49. Rarey M., Kramer B., Lengauer T. et Klebe G., *J. Mol. Biol.* **261**, 470–489 (1996).
50. Dominguez C., Boelens R. et Bonvin A. M., *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
51. Li H., Li C., Gui C., Luo X., Chen K., Shen J., Wang X. et Jiang H., *Bioorg. Med. Chem. Lett.* **14**, 4671–4676 (2004).
52. Jones G., Willett P., Glen R. C., Leach A. R. et Taylor R., *J. Mol. Biol.* **267**, 727–748 (1997).
53. Shoichet B. K., Kuntz I. D. et Bodian D. L., *J. Comput. Chem.* **13**, 380–397 (2004).
54. Meng E. C., Shoichet B. K. et Kuntz I. D., *J. Comput. Chem.* **13**, 505–524 (2004).
55. Cai W., Shao X. et Maigret B., *J. Mol. Graph. Model.* **20**, 313–328 (2002).
56. Morris R. J., Najmanovich R. J., Kahraman A. et Thornton J. M., *Bioinformatics* **21**, 2347–2355 (2005).
57. Kahraman A., Morris R. J., Laskowski R. A. et Thornton J. M., *J. Mol. Biol.* **368**, 283–301 (2007).
58. Gabb H. A., Jackson R. M. et Sternberg M. J. E., *J. Mol. Biol.* **272**, 106–120 (1997).
59. Norel R., Petrey D., Wolfson H. J. et Nussinov R., *Proteins* **36**, 307–317 (1999).
60. Inbar Y., Schneidman-Duhovny D., Halperin I., Oron A., Nussinov R. et Wolfson H. J., *Proteins* **60**, 217–223 (2005).
61. Feig M., Onufriev A., Lee M. S., Im W., Case D. A. et Brooks III C. L., *J. Comput. Chem.* **25**, 265–284 (2004).
62. Jain A. N., *Curr. Protein Pept. Sc.* **7**, 407–420 (2006).
63. Lensink M. F., Méndez R. et Wodak S. J., *Proteins* **69**, 704–718 (2007).
64. Robertson T. A. et Varani G., *Proteins* **66**, 359–374 (2007).
65. Böhm H.-J., *J. Comput. Aid. Mol. Des.* **12**, 309–323 (1998).

66. Muegge I., *J. Med. Chem.* **49**, 5895–5902 (2006).
67. Strout D. L. et Scuseria G. E., *J. Chem. Phys.* **102**, 8448–8452 (1995).
68. Goedecker S., *Rev. Mod. Phys.* **71**, 1085–1123 (1999).
69. Goedecker S. et Scuseria G. E., *Comp. Sci. Eng.* **5**, 14–21 (2003).
70. Shao Y., Molnar L. F., Jung Y., Kussmann J., Ochsenfeld C., Brown S. T., Gilbert A. T., Slipchenko L. V., Levchenko S. V., ONeill D. P., DiStasio Jr R. A., Lochan R. C., Wang T., Beran G. J., Besley N. A., Herbert J. M., Lin C. Y., Van Voorhis T., Chien S. H., Sodt A., Steele R. P., Rassolov V. A., Maslen P. E., Korambath P. P., Adamson R. D., Austin B., Baker J., Byrd E. F. C., Dachsel H., Doerksen R. J., Dreuw A., Dunietz B. D., Dutoi A. D., Furlani T. R., Gwaltney S. R., Heyden A., Hirata S., Hsu C.-P., Kedziora G., Khalliulin R. Z., Klunzinger P., Lee A. M., Lee M. S., Liang W., Lotan I., Nair N., Peters B., Proynov E. I., Pieniazek P. A., Rhee Y. M., Ritchie J., Rosta E., Sherrill C. D., Simmonett A. C., Subotnik J. E., Woodcock III H. L., Zhang W., Bell A. T. et Chakraborty A. K., *Phys. Chem. Chem. Phys.* **8**, 3172–3191 (2006).
71. Ventura O. N., Coitiño E. L., Lledós A. et Berteán J., *J. Mol. Struc.-THEOCHEM* **187**, 55–68 (1989).
72. Csonka G. I., *J. Comput. Chem.* **14**, 895–898 (1993).
73. Csonka G. I. et Ángyán J. G., *J. Mol. Struc.-THEOCHEM* **393**, 31–38 (1997).
74. Bernal-Uruchurtu M. I., Martins-Costa M. T. C., Millot C. et Ruiz-López M. F., *J. Comput. Chem.* **21**, 572–581 (2000).
75. Bernal-Uruchurtu M. I. et Ruiz-López M. F., *Chem. Phys. Lett.* **330**, 118–124 (2000).
76. Harb W., Bernal-Uruchurtu M. I. et Ruiz-López M. F., *Theor. Chem. Acc.* **112**, 204–216 (2004).
77. Monard G., Bernal-Uruchurtu M. I., van der Vaart A., Merz, Jr. K. M. et Ruiz-López M. F., *J. Phys. Chem. A* **109**, 3425–3432 (2005).
78. Wang B., Raha K., Liao N., Peters M. B., Kim H., Westerhoff L. M., Wollacott A. M., van der Vaart A., Gogonea V., Suarez D., Dixon S. L., Vincent J. J., Brothers E. N. et K. M. Merz J., “DivCon”, QuantumBio Inc., State College, PA 16803, 2005.
79. Lee T.-S., York D. M. et Yang W., *J. Chem. Phys.* **105**, 2744–2750 (1996).
80. Dixon S. L. et Merz K. M. J., *J. Chem. Phys.* **104**, 6643–6649 (1996).
81. Dixon S. L. et Merz K. M. J., *J. Chem. Phys.* **107**, 879–893 (1997).
82. Dewar M. J. S., Zoebisch E. G., Healy E. F. et Stewart J. J. P., *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
83. Stewart J. J. P., *J. Comput. Chem.* **10**, 209–220 (1989).
84. Stewart J. J. P., *J. Comput. Chem.* **10**, 221–264 (1989).
85. Stewart J. J. P., *J. Comput. Chem.* **12**, 320–341 (1989).
86. Harb W., *Méthodes de modélisation en synthèse asymétrique. Etude de la réduction d’alpha-cétoesters glycosylés,* Ph.D. thesis, Université Henri Poincaré, Nancy, France, 2003.
87. Hestenes M. R. et Stiefel E., *J. Res. Nat. Bur. Stand.* **49**, 409–436 (1952).
88. Broyden C. G., *IMA J. Appl. Math.* **6**, 76–90 (1970).
89. Fletcher R., *Comput. J.* **13**, 317–322 (1970).
90. Goldfarb D., *Math. Comp.* **24**, 23–26 (1970).
91. Shanno D. F., *Math. Comp.* **24**, 647–656 (1970).
92. Press W. H., Teukolsky S. A., Vetterling W. T. et Flannery B. P., *Numerical Recipes : The Art of Scientific Computing ;* Cambridge University Press : Cambridge, Third ed. ; 2007.
93. Jégou P. et Terrioux C., *Artif. Intel.* **146**, 43–75 (2003).
94. Kirkpatrick S., C. D. Gelatt J. et Vecchi M. P., *Science* **220**, 671–680 (1983).
95. Kirkpatrick S., *J. Stat. Phys.* **34**, 975–986 (1984).
96. Holland J. H., *Adaptation in Natural and Artificial Systems : An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence ;* University of Michigan Press : Ann Arbor, 1975.

97. Goldberg D. E., *Genetic Algorithms in Search, Optimization and Machine Learning ;* Kluwer Academic Publishers : Boston, MA, 1989.
98. Barber M. N. et Ninham B. W., *Random and Restricted Walks : Theory and Applications ;* Gordon and Breach : New York, 1970.
99. Dorsey R. E. et Mayer W. J., *J. Bus. Eco. Stat.* **13**, (1995).
100. Darwin C. R., *On the Origin of Species by Means of Natural Selection ;* John Murray : London, 1859.
101. Etaner-Uyar A. S. et Harmanci A. E., Preserving diversity through diploidy and meiosis for improved genetic algorithm performance in dynamic environments. In *Advances in Information Systems (Lecture Note in Computer Science)*, Vol. 2457 ; Yakhno T., Ed. ; Springer : New York, 2002.
102. van Rossum G. et Drake F. L., *Python Reference Manual ;* PythonLabs : Virginia, USA, 2008 Available at `http ://www.python.org`.
103. Hamming R. W., *Bell System Technical Journal* **26**, 147–160 (1950).
104. Gray F., “Pulse code communication”, U.S. Patent 2,632,058, 1953.
105. Miller B. L. et Goldberg D. E., “Genetic Algorithms, Tournament Selection and the Effects of Noise”, Technical Report 95006, Department of General Engineering, University of Illinois, Urbana-Champaign, 1995.
106. Kubota R., Horio K. et Yamakawa T., *Int. Cong. Ser.* **1291**, 169–172 (2006).
107. Jiao L. et Wang L., *IEEE Trans. on Systems, Man and Cybernetics A* **30**, 552–561 (2000).
108. Ross B. J., A Lamarckian Evolution Strategy for Genetic Algorithms. In *Practical Handbook of Genetic Algorithms : Complex Coding Systems*, Vol. 3, CRC Press ed. ; Chambers L. D., Ed. ; Boca Raton : London, 1999 ; Chapter 1, pages 1–16.
109. `http ://gridengine.sunsource.net`.
110. Rinaldi D., Hoggan P. E., Cartier A., Baka K., Monard G., Loos M., Reuter N., Mokrane A., Dillet V. et Théry V., “The GEOMOP Program”, 1989 Université Henri Poincaré, Nancy I.
111. van der Vaart A., Suárez D. et Merz, Jr. K. M., *J. Chem. Phys.* **113**, 10512–10523 (2000).
112. Dannenberg J. J., *J. Phys. Chem.* **92**, 6869–6871 (1988).
113. Odutola J. A. et Dyke T. R., *J. Chem. Phys.* **72**, 5062–5070 (1980).
114. Halkier A., Koch H., Jørgensen P., Christiansen O., Beck Nielsen I. M. et Helgaker T., *Theor. Chem. Acc.* **97**, 150–157 (1997).
115. Kahn K., Serfozo P. et Tipton P. A., *J. Am. Chem. Soc.* **119**, 5435–5442 (1997).
116. Sarma A. D., Serfozo O., Kahn K. et Tipton P. A., *J. Biol. Chem.* **274**, 33863–33865 (1999).
117. Tipton P. A., *Nat. Chem. Biol.* **2**, 124–125 (2006).
118. Wu X. W., Muzny D. M., Lee C. C. et Caskey C. T., *J. Mol. Evol.* **34**, 78–84 (1992).
119. Colloc’h N., El Hajji M., Bachet B., L’Hermite G., Schiltz M., Prangé T., Castro B. et Mornon J.-P., *Nat. Struct. Biol.* **4**, 947–952 (1997).
120. Retailleau P., Colloc’h N., Vivarès D., Bonneté F., Castro B., Hajji M. E., Mornon J.-P., Monard G. et Prangé T., *Acta Cryst. D* **60**, 453–462 (2004).
121. Gabison L., Prangé T., Colloc’h N., Hajji M. E., Castro B. et Chiadmi M., *BMC Struct. Biol.* **8**, 32 (2008).
122. Colloc’h N., Sopkova-de Oliveira Santos J., Retailleau P., Vivarès D., Bonneté F., Langlois d’Estainto B. B., Gallois B., Brisson A., Risso J.-J., Lemaire M., Prangé T. et Abraini J.-H., *Biophys. J.* **92**, 217–224 (2007).
123. Colloc’h N., Poupon A. et Mornon J.-P., *Proteins* **39**, 142–154 (2000).
124. Case D., Darden T., Cheatham, III T., Simmerling C., Wang J., Duke R., Luo R., Merz K., Pearlman D., Crowley M., Walker R., Zhang W., Wang B., Hayik S., Roitberg A., Seabra G., Wong K., Paesani F., Wu X., Brozell S., Tsui V., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Beroza P., Mathews D., Schafmeister C., Ross W. et Kollman P., “AMBER 9”, University of California, San Francisco, 2006.
125. Duan Y., Wu C., Chowdhury S., Lee M. C., Xiong G., Zhang W., Yang R., Cieplak P., Luo R., Lee T., Caldwell J., Wang J. et Kollman P., *J. Comput. Chem.* **24**, 1999–2012 (2003).

126. Altarsha M., Monard G. et Castro B., *Int. J. Quant. Chem.* **107**, 172–181 (2007).
127. Stavric B. et Nera E. A., *Clin. Toxicol.* **13**, 47–74 (1978).
128. Vasilyev V. et Bliznyuk A., *Theor. Chem. Acc.* **112**, 313– (2004).
129. Raha K. et Merz, Jr K. M., *J. Med. Chem.* **48**, 4558– (2005).
130. Zhou T., Huang D. et Caflisch A., *J. Med. Chem.* **51**, 4280–4288 (2008).
131. Gogonea V. et Merz, Jr. K. M., *J. Phys. Chem. A* **103**, 5171–5188 (1999).
132. Schöneich C., *Biochim. Biophys. Acta* **1703**, 111–119 (2005).
133. Kauffmann B., Aubry A. et Favier F., *Biochim. Biophys. Acta* **1703**, 249–260 (2005).
134. Antoine M., *Les Méthionine Sulfoxyde Réductases de classe A : catalyse, spécificité structurale et relations structure-fonction,* Ph.D. thesis, Université Henri Poincaré, Nancy, France, 2006.
135. Turrens J. F., *J. Physiol.* **552**, 335–344 (2003).
136. Babior B. M., *UIBMB Life* **50**, 267–269 (2001).
137. Vignais P. V., *Cell. Mol. Life Sci.* **59**, 1428–1459 (2002).
138. Babior B. M., Lambeth J. D. et Nauseef W., *Arch. Biochem. Biophys.* **397**, 342–344 (2002).
139. Coon M. J., Ding X. X., Pernecky S. J. et Vaz A. D., *FASEB J.* **6**, 669–673 (1992).
140. Yokoyama Y., Beckman J. S., Beckman T. K., Wheat J. K., Cash T. G., Freeman B. A. et Parks D. A., *Am. J. Physiol.* **258**, G564–G570 (1990).
141. Rubbo H., Radi R., Trujillo M., Telleri R., Kalyanaraman B., Barnes S., Kirk M. et Freeman B. A., *J. Biol. Chem.* **269**, 26066–26075 (1994).
142. Stadtman E. R. et Levine R. L., *Ann. NY Acad. Sci.* **899**, 191–208 (2000).
143. Richter C., Park J.-W. et Ames B. N., *Proc. Natl. Acad. Sci. USA* **85**, 6465–6467 (1988).
144. LeDoux S. P., Driggers W. J., Hollensworth B. S. et Wilson G. L., *Mutat. Res.* **464**, 149–159 (1999).
145. Kaur H. et Halliwell B., *FEBS Lett.* **350**, 9–12 (1994).
146. Dröge W., *Physiol. Rev.* **82**, 47–95 (2002).
147. Le Bras M., Clément M.-V., Pervaiz S. et Brenner C., *Histol. Hispathol.* **20**, 205–220 (2005).
148. Nakao N., Kurokawa T., Nonami T., Tumurkhuu G., Koide N. et Yokochi T., *Innate Immunity* **14**, 190–196 (2008).
149. Krinsky N. I., *Free Rad. Biol. Med.* **7**, 617–635 (1989).
150. Bors W., Heller W., Michel C. et Saran M., *Methods in Enzymology* **186**, 343–355 (1990).
151. Zelko I. N., Mariani T. J. et Folz R. J., *Free Rad. Biol. Med.* **33**, 337–349 (2002).
152. Davies K. J., *J. Biol. Chem.* **262**, 9895–9901 (1987).
153. Sies H., Oxidative stress : introduction. In *Oxidative stress, oxidants and antioxidants* ; Sies H., Ed. ; London Academic Press : London, 1991.
154. Shechter Y., *J. Biol. Chem.* **261**, 66–70 (1986).
155. Vogt W., *Free Rad. Biol. Med.* **18**, 93–105 (1995).
156. Musker W. K., Wolford T. L. et Roush P. B., *J. Am. Chem. Soc.* **100**, 6416–6421 (1978).
157. Honga J. et Schöneich C., *Free Rad. Biol. Med.* **31**, 1432–1441 (2001).
158. Venkataramanan N. S., Rajagopal S. et Vairamani M., *J. Inorg. Biochem.* **101**, 274–282 (2007).
159. Schöneich C., Pogocki D., Hug G. L. et Bobrowski K., *J. Am. Chem. Soc.* **125**, 13700–13713 (2003).
160. Yamazaki I. et Piette L. H., *J. Biol. Chem* **265**, 13589–13594 (1990).
161. Bonifačić M., Möckel H., Bahnemann D. et Asmus K.-D., *J. Chem. Soc., Perkin Trans. 2* 675–685 (1975).
162. Schöneich C. et Bobrowski K., *J. Am. Chem. Soc.* **115**, 6538–6547 (1993).
163. Bobrowski K. et Schöneich C., *J. Chem. Soc, Chem. Comm.* 795–797 (1993).
164. Hiller K.-O., Masloch B., Goebl M. et Asmus K.-D., *J. Am. Chem. Soc.* **103**, 2734–2743 (1981).

165. Asmus K.-D., *Acc. Chem. Res.* **12**, 436–442 (1979).
166. Miller B. L., Kuczera K. et Schöneich C., *J. Am. Chem. Soc.* **120**, 3345–3356 (1998).
167. Schöneich C., Aced A. et Asmus K.-D., *J. Am. Chem. Soc.* **115**, 11376–11383 (1993).
168. Schöneich C. et Bobrowski K., *J. Phys. Chem.* **98**, 12613–12620 (1994).
169. Merényi G., Lind J. et Engman L., *J. Phys. Chem.* **100**, 8875–8881 (1996).
170. Chu J.-W. et Trout B. L., *J. Am. Chem. Soc.* **126**, 900–908 (2004).
171. Chu J.-W., Yin J., Wang D. I. C. et Trout B. L., *Biochemistry* **43**, 1019–1029 (2004).
172. Richardson D. E., Regino C. A. S., Yao H. et Johnson J. V., *Free Radic. Biol. Med.* **35**, 1538–1550 (2003).
173. Chao C.-C., Ma Y.-S. et Stadtman E. R., *Proc. Natl. Acad. Sci. USA* **94**, 2969–2974 (1997).
174. Taggart C., Cervantes-Laurean D., Kim G., McElvaney N. G., Wehr N., Moss J. et Levine R. L., *J. Biol. Chem.* **275**, 27258–27265 (2000).
175. Schöneich C., *Arch. Biochem. Biophys.* **397**, 370–376 (2002).
176. Varadarajan S., Yatin S., Kanski J., Jahanshahi F. et Butterfield D. A., *Brain Res. Bull.* **50**, 133–141 (1999).
177. Hou L., Kang I., Marchant R. E. et Zagorski M. G., *J. Biol. Chem.* **277**, 40173–40176 (2002).
178. Haghighi A. Z. et Maples K. R., *J. Neurosci.* **43**, 107–111 (1996).
179. Davies K. J. A., *Biochimie* **83**, 301–310 (2001).
180. Charles C. A. et Gordon W. R., *Biochem. J.* **26**, 122–132 (1932).
181. Gilbert J. B., Price V. E. et Greenstein J. P., *J. Biol. Chem.* **180**, 209–218 (1949).
182. Flatmark T. et Sletten K., *J. Biol. Chem.* **243**, 1623–1629 (1968).
183. Patel K. et Borchardt R. T., *Pharm. Res.* **7**, 787–793 (1990).
184. Lerm M., Pop M., Fritz G., Aktories K. et Schmidt G., *Infect. Immun.* **70**, 4053 ?4058 (2002).
185. Robinson A. B., McKerrow J. H. et Cary P., *Proc. Natl. Acad. Sci. USA* **66**, 753–757 (1970).
186. Robinson N. E. et Robinson A. B., *Molecular Clocks : Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins ;* Althouse Press : Cave Junction, OR, USA, CNRS Editions ed. ; 2004.
187. Calligaris M., *Coord. Chem. Rev.* **248**, 351–375 (2004).
188. Price C. C. et Oae S., Sulfur Bonding. In ; Ronald Press : New York, 1962.
189. Kucsman A. et Kapovits I., . In *Organic Sulphur Chemistry : Theoretical and Experimental Advances* ; Bernardi F., Csizmadia I. G. et Mangini A., Eds. ; Elsevier : 1985.
190. Davies J. A., *Adv. Inorg. Chem. Radiochem.* **24**, 115–187 (1981).
191. Dreizler H. et Dendle G., *Z. Naturforsch.* **19a**, 512 (1964).
192. Cioslowski J. et Mixon S. T., *Inorg. Chem.* **32**, 3209–3216 (1993).
193. Dobado J. A., Martínez-García H., Molina J. M. et Sundberg M. R., *J. Am. Chem. Soc.* **121**, 3156–3164 (1999).
194. Reed A. E. et von Ragué Schleyer P., *J. Am. Chem. Soc.* **112**, 1434–1445 (1990).
195. Renault E. et Le Questel J.-Y., *J. Phys. Chem. A* **108**, 7232–7240 (2004).
196. Landini D., Modena G., Scorrano G. et Taddei F., *J. Am. Chem. Soc.* **91**, 6703–6707 (1969).
197. Hansel A., Kuschel L., Hehl S., Lemke C., Agricola H.-J., Hoshi T. et Heinemann S. H., *FASEB J.* **16**, 911–913 (2002).
198. Vougier S., Mary J. et Friguet B., *Biochem. J.* **373**, 531 ?537 (2003).
199. Kim H.-Y. et Gladyshev V. N., *Mol. Biol. Cell* **15**, 1055–1064 (2004).
200. Ejiri S.-I., Weissbach H. et Brot N., *J. Bacteriol.* **139**, 161–164 (1979).
201. Olry A., Boschi-Muller S., Yu H., Burnel D. et Branlant G., *Protein Sci.* **14**, 2828–2837 (2005).

202. Moskovitz J., Rahman M. A., Strassman J., Yancey S. O., Kushner S. R., Brot N. et Weissbach H., *J. Bacteriol.* **177**, 502–507 (1995).
203. St.-John G., Brot N., Ruan J., Erdjument-Bromage H., Tempst P., Weissbach H. et Nathan C., *Proc. Natl. Acad. Sci. USA* **98**, 9901–9906 (2001).
204. Alamuri P. et Maier R. J., *Mol. Microbiol.* **53**, 1397–1406 (2004).
205. Douglas T., Daniel D. S., Parida B. K., Jagannath C. et Dhandayuthapani S., *J. Bacteriol.* **186**, 3590–3598 (2004).
206. Moskovitz J., Berlett B. S., Poston J. M. et Stadtman E. R., *Proc. Natl. Acad. Sci. USA* **94**, 9585–9589 (1997).
207. Singh V. K. et Moskovitz J., *Microbiology* **149**, 2739–2747 (2003).
208. Vattanaviboon P., Seeanukun C., Whangsuk W., Utamapongchai S. et Mongkolsuk S., *J. Bacteriol.* **187**, 5831–5836 (2005).
209. Moskovitz J., Bar-Noy S., Williams W. M., Requena J., Berlett B. S. et Stadtman E. R., *Proc. Natl. Acad. Sci. USA* **98**, 12920–12925 (2001).
210. Moskovitz J., Flescher E., Berlett B. S., Azare J., Poston J. M. et Stadtman E. R., *Proc. Natl. Acad. Sci. USA* **95**, 14071–14075 (1998).
211. Marchetti M. A., Pizarro G. O., Sagher D., DeAmicis C., Brot N., Hejtmancik J. F., Weissbach H. et Kantorow M., *Invest. Ophthalmol. Vis. Sci.* **46**, 2107–2112 (2005).
212. Sun H., Gao J., Ferrington D. A., Biesiada H., Williams T. D. et Squier T. C., *Biochemistry* **38**, 105–112 (1999).
213. Friguet B., Bulteau A.-L., Chondrogianni N., Conconi M. et Petropoulos I., *Ann. N. Y. Acad. Sci.* **908**, 143–154 (2000).
214. Bakala H., Delaval E., Hamelin M., Bismuth J., Borot-Laloi C., Corman B. et Friguet B., *Eur. J. Biochem.* **270**, 2295–2302 (2003).
215. Petropoulos I., Mary J., Périchon M. et Friguet B., *Biochem. J.* **355**, 819–825 (2001).
216. Picot C. R., Perichon M., Cintrat J.-C., Friguet B. et Petropoulos I., *FEBS Lett.* **558**, 74–78 (2004).
217. Boschi-Muller S., Azza S., Sanglier-Cianferani S., Talfournier F., van Dorsselear A. et Branlant G., *J. Biol. Chem.* **275**, 35908–35913 (2000).
218. Olry A., Boschi-Muller S. et Branlant G., *Biochemistry* **43**, 11616–11622 (2004).
219. Olry A., Boschi-Muller S., Marraud M., Sanglier-Cianferani S., van Dorsselear A. et Branlant G., *J. Biol. Chem.* **277**, 12016–12022 (2002).
220. Rouhier N., Vieira Dos Santos C., Tarrago L. et Rey P., *Photosynth. Res.* **89**, 247–262 (2006).
221. Rouhier N., Kauffmann B., Tête-Favier F., Palladino P., Gans P., Branlant G., Jacquot J.-P. et Boschi-Muller S., *J. Biol. Chem.* **282**, 3367–3378 (2007).
222. Kim H.-Y., Zhang Y., Lee B. C., Kim J.-R. et Gladyshev V. N., *Proteins* **74**, 1008–1017 (2009).
223. Lescure A., Gautheret D., Carbon P. et Krol A., *J. Biol. Chem.* **274**, 38147–38154 (1999).
224. Kryukov G. V., Kumar R. A., Koc A., Sun Z. et Gladyshev V. N., *Proc. Natl. Acad. Sci. USA* **99**, 4245–4250 (2002).
225. Kumar R. A., Koc A., Cerny R. L. et Gladyshev V. N., *J. Biol. Chem.* **277**, 37527–37535 (2002).
226. Tête-Favier F., Cobessi D., Boschi-Muller S., Azza S., Branlant G. et Aubry A., *Structure Fold. Des.* **8**, 1167–1178 (2000).
227. Lowther W. T., Brot N., Weissbach H. et Matthews B. W., *Biochemistry* **39**, 13307–13312 (2000).
228. Taylor A. B., David M. Benglis J., Dhandayuthapani S. et Hart P. J., *J. Bacteriol.* **185**, 4119–4126 (2003).
229. Ranaivoson F. M., Antoine M., Kauffmann B., Boschi-Muller S., Aubry A., Branlant G. et Favier F., *J. Mol. Biol.* **377**, 268–280 (2008).
230. Lowther W. T., Weissbach H., Etienne F., Brot N. et Matthews B. W., *Nat. Struct. Biol.* **9**, 348–352 (2002).

231. Boschi-Muller S., Azza S. et Branlant G., *Protein Sci.* **10**, 2272–2279 (2001).

232. Gand A., Antoine M., Boschi-Muller S. et Branlant G., *J. Biol. Chem.* **282**, 20484–20491 (2007).

233. Antoine M., Boschi-Muller S. et Branlant G., *J. Biol. Chem.* **278**, 45352-45352 (2003).

234. Antoine M., Gand A., Boschi-Muller S. et Branlant G., *J. Biol. Chem.* **281**, 39062–39070 (2006).

235. Marchal S. et Branlant G., *Biochemistry* **38**, 12950–12958 (1999).

236. Balta B., Monard G., Ruiz-López M. F., Antoine M., Gand A., Boschi-Muller S. et Branlant G., *J. Phys. Chem. A* **110**, 7628–7636 (2006).

237. Rivail J.-L. et Rinaldi D., *Chem. Phys.* **18**, 233–242 (1976).

238. Rinaldi D., Ruiz-López M. F. et Rivail J.-L., *J. Chem. Phys.* **78**, 834–838 (1983).

239. Dillet V., Rinaldi D. et Rivail J.-L., *J. Phys. Chem.* **98**, 5034–5039 (1994).

240. Rinaldi D., Bouchy A., Rivail J.-L. et Dillet V., *J. Chem. Phys.* **120**, 2343–2350 (2004).

241. Tolles W. M. et Gwinn W. D., *J. Chem. Phys.* **36**, 1119–1121 (1962).

242. Owsley D. C., Helmkamp G. K. et Rettig M. F., *J. Am. Chem. Soc.* **91**, 5239–5242 (1969).

243. Johnson C. R. et Rigau J. J., *J. Am. Chem. Soc.* **91**, 5398–5399 (1969).

244. Martin J. C. et Arhart R. J., *J. Am. Chem. Soc.* **93**, 2339–2341 (1971).

245. Martin J. C. et Arhart R. J., *J. Am. Chem. Soc.* **93**, 2341–2342 (1971).

246. Young P. R. et Hsieh L.-S., *J. Am. Chem. Soc.* **100**, 7121–7122 (1978).

247. Madesclaire M., *Tetrahedron* **44**, 6537–6580 (1988).

248. Paul I. C., Martin J. C. et Perozzi E. F., *J. Am. Chem. Soc.* **94**, 5010–5017 (1972).

249. Martin J. C. et Perozzi E. F., *Science* **191**, 154–159 (1976).

250. Downs A. J., McGrady G. S., Barnfield E. A., Rankin D. W. H., Robertson H. E., Boggs J. E. et Dobbs K. D., *Inorg. Chem.* **28**, 3286–3292 (1989).

251. Koutecký V. B. et Musher J. I., *Theor. Chem. Acc.* **33**, 227–238 (1974).

252. Schwenzer G. M. et Schaefer H. F., *J. Am. Chem. Soc.* **97**, 1388–1393 (1975).

253. Chen M. M. L. et Hoffmann R., *J. Am. Chem. Soc.* **98**, 1647–1653 (1976).

254. Case D., Pearlman D., Caldwell J., Cheatham III T., Ross W., Simmerling C., Darden T., Merz K., Stanton R., Cheng A., Vincent J., Crowley M., Tsui V., Radmer R., Duan Y., Pitera J., Massova I., Seibel G., Singh U., Weiner P. et Kollman P., “AMBER 6”, University of California, San Francisco, 1999.

255. Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W. et Kollman P. A., *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).

256. Rao B. G. et Singh U. C., *J. Am. Chem. Soc.* **112**, 3803–3811 (1990).

257. Vaisman I. I. et Berkowitz M. L., *J. Am. Chem. Soc.* **114**, 7889–7896 (1992).

258. Luzar A. et Chandler D., *J. Chem. Phys.* **98**, 8160–8173 (1993).

259. Liu H., Mueller-Plathe F. et van Gunsteren W. F., *J. Am. Chem. Soc.* **117**, 4363–4366 (1995).

260. Skaf M. S., *J. Chem. Phys.* **107**, 7996–8002 (1997).

261. Ryckaert J.-P., Ciccotti G. et Berendsen H. J. C., *J. Comput. Phys.* **23**, 327–341 (1977).

262. Wang J., Wolf R. M., Caldwell J. W., Kollman P. A. et Case D. A., *J. Comput. Chem.* **25**, 1157–1174 (2004).

263. Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Montgomery Jr.,J. A., Vreven T., Kudin K. N., Burant J. C., Millam J. M., Iyengar S. S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G. A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J. E., Hratchian H. P., Cross J. B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi R., Pomelli C., Ochterski J. W., Ayala P. Y., Morokuma K., Voth G. A., Salvador P., Dannenberg J. J., Zakrzewski V. G., Dapprich S., Daniels A. D., Strain M. C., Farkas O., Malick D. K., Rabuck A. D., Raghavachari K., Foresman J. B., Ortiz J. V., Cui Q.,

Baboul A. G., Clifford S., Cioslowski J., Stefanov B. B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R. L., Fox D. J., Keith T., Al-Laham M. A., Peng C. Y., Nanayakkara A., Challacombe M., Gill P. M. W., Johnson B., Chen W., Wong M. W., Gonzalez C. et Pople J. A., “Gaussian 03, Revision C.02”, Gaussian, Inc., Wallingford, CT, 2004.

264. Bayly C. I., Cieplak P., Cornell W. et Kollman P. A., *J. Phys. Chem.* **97**, 10269–10280 (1993).
265. Cornell W. D., Cieplak P., Bayly C. I. et Kollmann P. A., *J. Am. Chem. Soc.* **115**, 9620–9631 (1993).
266. Pigache A., Cieplak P. et Dupradeau F.-Y., Automatic and highly reproducible RESP and ESP charge derivation : Application to the development of programs RED and X RED. In *227th ACS National Meeting* ; 2004.
267. Typke V. et Dakkouri M., *J. Mol. Struct.* **599**, 177–193 (2001).
268. Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W. et Klein M. L., *J. Chem. Phys.* **79**, 926–935 (1983).
269. Darden T., York D. et Pedersen L., *J. Chem. Phys.* **98**, 10089–10092 (1993).
270. Essmann U., Perera L., Berkowitz M. L., Darden T., Lee H. et Pedersen L. G., *J. Chem. Phys.* **103**, 8577–8593 (1995).
271. Drabowicz J., Dudziński B. et Mikołajczyk M., *Synlett* **1992**, 252–254 (1992).
272. Lee G. H., Choi E. B., Lee E. et Pak C. S., *Tetrahedron Lett.* **35**, 2195–2196 (1994).
273. Mohanazadeh F., Momeni A. R. et Ranjbar Y., *Tetrahedron Lett.* **35**, 6127–6128 (1994).
274. Zhang Y., Yu Y. et Bao W., *Synth. Commun.* **25**, 1825–1830 (1995).
275. Wang J. Q. et Zhang Y. M., *Synth. Commun.* **25**, 3545–3547 (1995).
276. Nicolás E., Vilaseca M. et Giralt E., *Tetrahedron* **51**, 5701–5710 (1995).
277. Fujiki K., Kurita S. et Yoshida E., *Synth. Commun.* **26**, 3619–3626 (1996).
278. Wang Y. et Koreeda M., *Synlett* **1996**, 885–886 (1996).
279. Black S., Harte E. M., Hudson B. et Wartofsky L., *J. Biol. Chem.* **235**, 2910–2916 (1960).
280. Denisov G. S., Koll A., Lobadyuk V. I., Schreiber V. M., Shurukhina A. V. et Spevak V. N., *J. Mol. Struct.* **605**, 221–226 (2002).
281. Dixon R. W. et Kollman P. A., *J. Comput. Chem.* **18**, 1632–1646 (1997).
282. Cardey B. et Enescu M., *Chem. Phys. Chem.* **6**, 1175–1180 (2005).
283. Cardey B. et Enescu M., *J. Phys. Chem. A* **111**, 673–678 (2007).
284. Jalbout A. F., *Chem. Phys. Lett.* **340**, 571–580 (2001).
285. Jursic B. S., *J. Mol. Struct.-THEOCHEM* **452**, 145–152 (1998).
286. McKee M. L., *J. Phys. Chem.* **97**, 10971–10976 (1993).
287. Turecek F., *J. Phys. Chem.* **98**, 3701–3706 (1994).
288. Wang B., Hou H. et Gu Y., *J. Mol. Struct.-THEOCHEM* **505**, 241–246 (2000).
289. Aloisio S., *Chem. Phys.* **326**, 335–343 (2006).
290. Barone V. et Cossi M., *J. Phys. Chem. A* **102**, 1995–2001 (1998).
291. Cossi M., Rega N., Scalmani G. et Barone V., *J. Comput. Chem.* **24**, 669–681 (2003).
292. Fu Y., Liu L., Yu H.-Z., Wand Y.-M. et Guo Q.-X., *J. Mol. Struct.-THEOCHEM* **127**, 7227–7234 (2005).
293. Trasatti S., *Pure Appl. Chem.* **58**, 955–966 (1986).
294. Forni L. G. et Willson R. L., *Biochem. J.* **240**, 897–903 (1986).
295. Vorontsov A. V., *Russ. Chem. Rev.* **77**, 909–926 (2008).
296. Wood P. M., *FEBS Lett.* **124**, 11–14 (1981).
297. Stewart J. J. P., Review in Computational Chemistry. In , Vol. 1 ; K.B. Lipkowitz and D.B. Boyd ed., VCH : New York, 1990 ; Chapter 2, pages 45-81.
298. Zerner M. C., Review in Computational Chemistry. In ; K.B. Lipkowitz and D.B. Boyd ed., VCH : New York, 1991 ; Chapter 8, pages 313-365.

299. Pople J. A. et Segal G. A., *J. Chem. Phys.* **44**, 3289–3296 (1966).
300. Pople J. A., Beveridge D. L. et Dobosh P. A., *J. Chem. Phys.* **47**, 2026–2033 (1967).
301. Pople J. A., Santry D. P. et Segal G. A., *J. Chem. Phys.* **43**, S129-S135 (1965).
302. Yang W., *Phys. Rev. Lett.* **66**, 1438–1441 (1991).
303. Allen M. P. et Tildesley D. J., *Computer Simulations of Liquids ;* Oxford Clarendon Press : Oxford, 1989.
304. Leach A. R., *Molecular Modelling : Principles and Applications, second edition ;* Prentice Hall : 2001.
305. Rapaport D. C., *The Art of Molecular Dynamics Simulations ;* Cambridge University Press : 2004.
306. Lennard-Jones J. E., *Proc. R. Soc. London A* **106**, 463–477 (1924).
307. Patra M., Karttunen M., Hyvönen M. T., Falck E., Lindqvist P. et Vattulainen I., *Biophys. J.* **84**, 3636–3645 (2003).
308. Ewald P. P., *Ann. Phys.* **369**, 253–287 (1921).
309. Cooley J. W. et Tukey J. W., *Math. Comput.* **19**, 259–301 (1965).
310. Verlet L., *Phys. Rev.* **159**, 98–103 (1967).
311. Hockney R. W. et Eastwood J. W., *Computer Simulations Using Particules ;* McGraw-Hill : 1981.
312. Gear C. W., *Numerical Initial Value Problems in Ordinary Differential Equations ;* Prentice-Hall : 1971.
313. Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A. et Haak J. R., *J. Chem. Phys.* **81**, 3684–3690 (1984).
314. Mehler E. L. et Guarnieri F., *Biophys. J.* **77**, 3–22 (1999).
315. Demchuk E. et Wade R. C., *J. Phys. Chem.* **100**, 17373–17387 (1996).
316. Li H., Robertson A. D. et Jensen J. H., *Proteins* **55**, 689–704 (2004).

Résumé

La modélisation de la reconnaissance et de la catalyse enzymatiques a été abordée dans ce travail.

Dans une première partie, un logiciel de docking (AlgoGen-DivCon), déterminant la structure optimale d'un système protéine-ligand, a été développé en combinant une méthode semi-empirique couplée à un algorithme à croissance linéaire (Divide & Conquer) pour la description du système et un algorithme génétique (AlgoGen) pour la recherche conformationnelle. Des tests sur différents systèmes, dont des complexes enzyme-substrat, ont été effectués pour valider l'approche.

La seconde partie concerne l'étude du mécanisme (phase réductase) des méthionine sulfoxyde réductases de classe A (MsrA), catalysant la réduction de résidus méthionine oxydés (MetSO). A l'aide de simulations de dynamique moléculaire pour différents états de protonation du site actif de l'enzyme, nous montrons que l'ancrage du substrat est particulièrement favorable pour un état de protonation Glu94H et Cys51$^{\ominus}$. L'étude au niveau quantique du mécanisme de réduction, à partir du complexe michaélien, indique la formation d'une espèce "sulfurane" par protonation de la fonction sulfoxyde et création d'une liaison S–S avec Cys51. Le transfert d'un second proton provenant de Tyr134 (ou Tyr82) vers le groupement OH du sulfurane induit la formation d'une molécule d'eau et la dissociation de la liaison S–O. L'ion sulfuranyle résultant est enfin hydrolysé pour former un intermédiaire acide sulfénique et le substrat réduit. D'après le profil d'énergie, ce chemin pourrait être simplement décrit par un mécanisme concerté et asynchrone impliquant deux transferts de proton.

Mots-clés

Catalyse enzymatique - Reconnaissance moléculaire - Dynamique moléculaire - Chimie quantique - Méthionine sulfoxyde réductases - Mécanisme réactionnel - Algorithme génétique - Acide sulfénique

---

Abstract

The modeling of enzymatic recognition and catalysis has been examined in this work.

In the first part, a docking software (AlgoGen-DivCon) to determine the optimal structure of a protein-ligand system, has been developed. The molecular system has been described using a semiempirical method combined with a linear scaling algorithm (Divide & Conquer), while the conformational search has been performed by a genetic algorithm (AlgoGen). Tests on different systems, including enzyme-substrate complexes, have been conducted to validate the approach.

The second part concerns the study of the mechanism (reductase phase) of class A methionine sulfoxide reductases (MsrA), which catalyzes the reduction of oxidized methionine residues (MetSO). By means of molecular dynamics simulations for different protonation states of the enzyme active site, we have shown that the substrate docking is particulary favourable for a Glu94H and Cys51$^{\ominus}$ protonation state. The quantum chemistry study of the reduction mechanism, starting from the *Michaelis* complex, indicates that the process is initiated by the formation of a "sulfurane" species, after protonation of the sulfoxide and simultaneous formation of a S–S bond with Cys51. A second proton transfer from Tyr134 (or Tyr82) to the sulfurane OH group induces the formation of a water molecule and the dissociation of the S–O bond. The resulting sulfuranyl ion is finally hydrolyzed to form a sulfenic acid intermediate and the reduced substrate. According to the energy profile, this path could be simply described as a concerted and asynchronous mechanism involving two proton transfers.

Key words

Enzymatic catalysis - Molecular recognition - Molecular dynamics - Quantum chemistry - Methionine sulfoxide reductases - Reaction mechanism - Genetic algorithm - Sulfenic acid