UNIVERSITÉ DE FRANCHE-COMTÉ

Université de Franche-Comté

École Doctorale Carnot-Pasteur

Thèse de Doctorat

Spécialité Chimie

présentée par

Guillaume Paris

Effets dynamiques et conformationnels sur le rôle de transport des albumines sériques

Thèse dirigée par Mironel Enescu et Christophe Ramseyer

Laboratoire Chrono-Environnement - UMR 6249
Laboratoire Nanomédecine Imagerie et Thérapeutique - EA 4662

Soutenue le 05 juin 2014 devant la Commission d'Examen

Rapporteurs :	**C. Etchebest**	Professeur, Université Paris Diderot (Paris 7)
	M. Dauchez	Professeur, Université de Reims Champagne-Ardenne (Reims)
Examinateurs :	**P. Senet**	Professeur, Université de Bourgogne (Dijon)
	S. Yesylevskyy	Senior Researcher, Institute of Physics of the NAS of Ukraine (Kiev)
	C. Ramseyer	Professeur, Université de Franche-Comté (Besançon)
	M. Enescu	Professeur, Université de Franche-Comté (Besançon)

Remerciements

Je tiens en premier lieu à adresser mes remerciements les plus sincères à mes directeurs de thèse, les professeurs Mironel Enescu et Christophe Ramseyer. J'ai beaucoup appris à leurs côtés et je suis très honoré de les avoir eu comme encadrants. J'espère avoir été digne de la confiance qu'ils m'ont accordée et que ce travail est à la hauteur de leurs espérances. Merci pour votre patience, votre passion scientifique, votre implication dans ce projet et vos qualités pédagogiques.

Mes remerciements s'adressent ensuite au Pr. Daniel Gilbert qui m'a accueilli au sein du laboratoire Chrono-Environnement, ainsi qu'aux permanents de l'ancien LCPR-AC et à mes collègues doctorants. Je tiens aussi à remercier Sebastien Krazweski de m'avoir initié à la modélisation moléculaire.

J'adresse mes remerciements aux membres du jury d'avoir accepté de juger ce travail et de participer à la soutenance de cette thèse.

Merci également au mésocentre de calcul de Franche-Comté et à leur personnel pour les nombreuses heures de calcul attribuées et leur disponibilité.

Je tiens aussi à remercier tous les amis qui m'ont encouragé et soutenu au long de ces années, sans oublier mes collègues du CAC.

Enfin je tiens à remercier tout particulièrement mes parents, ma soeur, mon frère et leur famille respective pour leur amour, leur confiance et leur soutien inconditionnel. Je vous dois beaucoup.

Résumé

L'albumine sérique humaine (HSA) est une protéine connue pour ses propriétés de transport exceptionnelles et son contenu élevé en ponts disulfure. L'étude de sa dynamique conformationnelle représente un défi important dans la compréhension de ses fonctions physiologiques. Le but de notre travail a été d'étudier cette dynamique conformationnelle et de comprendre le rôle des ponts disulfure dans le maintien de la structure native de la protéine. Notre analyse est basée sur des simulations de dynamique moléculaire couplées à des analyses par composantes principales. Outre la validation de la méthode de simulation les résultats fournissent de nouveaux éclairages sur les principaux effets de la réduction des ponts disulfure dans les albumines sériques. Les processus de dépliement/repliement protéique ont été détaillés. La prédiction de la structure réduite d'équilibre a également fait l'objet d'une attention particulière. Une étude détaillée de la dynamique conformationnelle globale de la protéine ainsi que celle des deux sites principaux de complexation a été effectuée. D'éventuels effets allostériques entre ces deux sites ont été recherchés. Les résultats théoriques obtenus ont été discutés avec les données expérimentales disponibles.

Abstract

Human serum albumin (HSA) is a protein known for its exceptional transport properties and its high content of disulfide bridges. The study of the conformational dynamics represents a major challenge in the comprehension of its physiological functions. The aim of our work was to study the conformational dynamics and to understand the role of disulfide bonds in the stability of the native protein structure. Our analysis is based on simulations of molecular dynamics coupled with principal component analysis. Beyond the validation of the simulation method, the results provide new insights on the main effects of the disulfide bonds reduction in serum albumins. Protein unfolding/refolding processes were detailed. A special attention is paid to the prediction of the reduced structure at the equilibrium. A detailed study of the global protein conformational dynamics as well as the two main binding sites were performed. Possible allosteric effects between these two sites were researched. The theoretical results have been discussed with the available experimental data.

Table des matières

Remerciements **i**

Résumé **iii**

Abstract **vii**

Table des matières **ix**

Liste des figures **xi**

Liste des tableaux **xv**

Liste des principales abréviations **xv**

1 Introduction **1**

1.1 Les protéines, leurs structures et leurs fonctions 2

1.2 L'albumine sérique humaine 4

1.3 Questions traitées dans la thèse 8

1.4 Approche méthodologique 10

1.5 Présentation des chapitres 12

2 Méthodes de modélisation moléculaire **15**

2.1 Dynamique moléculaire classique 17

2.2 Analyse par composantes principales 38

3 Dynamique conformationnelle des sous-domaines et des sites de complexation de HSA$^{(N)}$ **43**

3.1 Introduction 45

3.2 Convergence de la méthode d'analyse 45

3.3 Dynamique globale de la protéine 47

3.4 Dynamique des sites primaires de liaison 55
3.5 Corrélation des mouvements entre les sites 1 et 2 62
3.6 Conclusions 64

4 Rôle structural des ponts disulfure 65
4.1 Introduction 66
4.2 Test de stabilité structurale 66
4.3 Effets de la réduction des ponts sur les structures secondaire et tertiaire . . 68
4.4 Effet de la réduction des ponts S—S sur le rayon de gyration de la protéine 72
4.5 Dissociation des couples thiols 73
4.6 Conclusions 79

5 Dépliement réductif de la HSA : prédiction de la structure réduite 81
5.1 Introduction 82
5.2 Étude de convergence de la méthode d'analyse 82
5.3 Recherche de conformations stables et étude globale du dépliement 84
5.4 Effet du dépliement sur les propriétés de complexation de $HSA^{(R)}$ 89
5.5 Conclusions 91

6 Conclusions 93

Bibliographie 97

Liste des figures

1.1 Structure tridimensionnelle de la HSA. Les 6 sous-domaines protéiques (IA à IIIB) sont illustrés par différentes couleurs. 4
1.2 Représentation schématique de la structure de la HSA avec identification des ponts disulfure en bandes grises. Les atomes, amine et carbone, terminaux sont notés respectivement N et C. 5
1.3 Illustration du processus de dépliement/repliement de la structure protéique induit par la rupture ou la reconstitution des ponts disulfure 6
1.4 Illustration des sites de liaison les plus répandus pour les acides gras de la HSA 7
1.5 Représentation surfacique des deux sites primaires de liaison à travers la structure secondaire de la HSA 7

2.1 Système de simulation typique pour la HSA 19
2.2 Vue bidimensionnelle des conditions périodiques. La boîte mère est située au centre. 20
2.3 Contributions au potentiel d'interaction moléculaire 21
2.4 Représentation des interactions liées 22
2.5 Représentation des interactions non-liées 23
2.6 Représentation de la méthode du cut-off 25
2.7 Représentation de la méthode du double cut-off 26
2.8 Potentiel électrostatique avec (rouge) ou sans (noir) l'application de la fonction de *switching*.En vert, la discontinuité du potentiel électrostatique du à l'utilisation du cut-off. 27
2.9 Défaut de la méthode des gradients : le choix de la configuration initiale influe directement sur la position du minimum 32

3.1 Evolution des RMSD au long des trois trajectoires de simulation : traj1 (noir), traj2 (rouge) traj3 (bleu). 46

3.2 Déplacements (RSD) des atomes de C_α au long de la première PC de chacune des trois trajectoires (noir). On trouve également les déplacements entre la HSA native et complexée avec le myristate (rouge, en haut de la figure). La structure du complexe HSA-myristate provient de : Protein Data Bank RCSB, code ID : 1E7G [1] 48

3.3 Déplacements (RSD) des atomes de C_α le long de PC2 pour chacune des trois trajectoires. 49

3.4 Déplacements (RSD) des atomes de C_α au long de la troisième (à gauche), quatrième (au centre) et cinquième (à droite) PC de chacune des trois trajectoires 50

3.5 Les 4 clusters de conformations (en rouge) obtenus en utilisant la distribution des 3 premières PCs de la traj1 52

3.6 Mouvement des atomes C_α : a) entre les deux valeurs extrêmes de PC1 dérivé de traj1 (bleu pour le minimum, rouge pour le maximum) et b) au cours de la transition entre les conformations native et complexée avec le myristate (rouge pour la native, bleu pour le complexe HSA-myristate). . . 54

3.7 Les différents conformères obtenus pour les sites 1 (à gauche) et 2 (à droite) en utilisant la distribution des trois premières PC des deux trajectoires restreintes. 56

3.8 Conformation relative des principaux résidus d'acides aminés du site 1 dans la structure moyenne de CS5 (bleu), CS6 (rouge) et CS7 (vert). 57

3.9 RMSF calculé entre les conformations CS6 et CS7 (vert) et entre la HSA native et le complexe HSA-thyroxine (PDB ID : 1HK1) (noir). 58

3.10 Mouvement des principaux résidus impliqués dans la formation du site 1 au long de la première composante PC1 de traj1 59

3.11 Conformation relative des principaux résidus d'acides aminés du site 2 dans la structure moyenne des conformères CS8 (bleu), CS9 (rouge), CS10 (jaune) et CS11 (vert). 60

3.12 RMSF calculé entre les conformations CS8 et CS9 (rouge) et entre la HSA native et le complexe HSA-diazepam (PDB ID : 1HA2) (noir). 61

3.13 Le déplacement (RSD) des C_α des trois premières PC de la trajectoire restreinte aux deux sous domaines IIA et IIIA 62

3.14 Le déplacement (RSD) des C_α au cours des transitions CS12→CS13 (a) et CS13→CS14 (b) 63

4.1 RMSD du squelette protéique calculé pour la protéine native (en noir) et réduite (en rouge) avec comme référence la structure de diffraction aux RX. Les simulations sont effectuées à 300 K et 500 K (trois trajectoires différentes pour chaque protéine) avec les champs de force OPLS-AA et AMBER ff 03. 67
4.2 Structure secondaire : évolution du contenu en hélice-α de la protéine native (en noir) et réduite (en rouge). Les simulations de DM sont effectuées à 300 K et 500 K avec les champs de force OPLS-AA et AMBER ff 03. 68
4.3 Structure tertiaire : évolution du pourcentage des contacts natifs dans la protéine native (en noir) et réduite (en rouge). Les simulations de DM sont effectuées à 300 K et 500 K avec les champs de force OPLS-AA et AMBER ff 03. 69
4.4 Distortion de la structure tertiaire de $HSA^{(R)}$. La structure a été extraite de la trajectoire avec le champ de force OLPS-AA simulé à 500 K. Les atomes de soufre sont représentés par des sphères solides et sont connectés par des lignes pointillées lorsqu'ils appartenaient au même pont réduit. 71
4.5 Profils d'énergie libre de Gibbs en rapport avec le rayon de gyration pour les formes native (noir) et réduite (rouge) en utilisant le champ de force OPLS-AA et la méthode Umbrella Sampling. 72
4.6 Evolution des distances S—S pour la HSA réduite : simulations de DM à 300 K (bleu) et 500 K (magenta) réalisées avec le champ de force OPLS-AA 74
4.7 Profils d'énergie libre de Gibbs calculés par rapport à la distance thiol-thiol pour les trois ponts réduits du sous-domaine IA de la HSA. Les calculs ont été effectués sur la conformation de la protéine native. 76
4.8 Illustration du dépliement du pont réduit Cys200-246Cys : les structures ont été extraites de la trajectoire OPLS-AA simulée à 500 K : a) t = 2 ns, b) t = 4 ns , c) t = 8 ns. Les atomes de soufre sont représentés par des sphères solides. 77
4.9 Profils d'énergie libre de Gibbs calculés par rapport à la distance thiol-thiol pour les deux ponts réduits dans une conformation séparée. 79

5.1 Evolution du RMSD (Å) des atomes de C_α par rapport à la structure de référence au long des quatre trajectoires de DM 83
5.2 Evolution de l'hélicité (%) au long des quatre trajectoires de DM 83
5.3 Dépliement de la HSA réduite : séquence de conformères extraits des quatre simulations de DM et ordonnés en fonction de leur hélicité décroissante. . . 86

5.4 Hélicité (%) par sous-domaine des 17 conformères (IA-bleu, IB-cyan, IIA-vert, IIB-jaune, IIIA-rose et IIIB-rouge) 86
5.5 Fraction des contacts natifs tertiaires (bleu) par rapport à l'hélicité globale (noir) pour les 17 conformères 86
5.6 Dépliement de la HSA réduite : hélicité des hélices longues (rouge) par rapport à celle des hélices courtes (vert) 87
5.7 Dépliement de trois hélices longues de type H2 (noir), H3 (bleu) et H4 (rouge) : a) dans le sous-domaine IA et b) dans le sous domaine IIIB. Voir la Figure 1.2 pour l'emplacement des hélices. 88
5.8 Représentation des deux sites principaux de complexation (en vert) à travers la structure native de la protéine (a) et le conformère C4d (b). 90
5.9 Représentation des quatre plus importantes cavités (en vert) de la structure native (a) et du conformère C4d(b). 91

Liste des tableaux

3.1 Contenu en cosinus des 5 premières PC des 3 trajectoires simulées 47
3.2 Distorsion des sous-domaines de la HSA induite par les différentes PC : RMSD (en Å) entre les deux structures extrêmes de la PC 51
3.3 Population (%) des 4 clusters de conformations identifiés en utilisant la distribution des 3 premières PCs de la traj1 52
3.4 Population (%) des différents clusters de conformations identifiés pour les deux sites primaires 56
3.5 Population (%) des différents clusters de conformations identifiés par PCA après projection de traj 1 sur le sous-espace formé par les deux sous-domaines IIA et IIIA 63

4.1 Erreur moyenne (kcal/mol) sur le calcul des profils d'énergie libre de Gibbs des ponts S—S du sous-domaine IA en fonction de la distance SS 75
4.2 Exemples de changements conformationnels dirigeant la dissociation des thiols 78

5.1 Contenu en cosinus des cinq premières PC dérivées des trajectoires de DM simulées à 350, 375, 400 et 425 K. 84
5.2 Conformères identifiés au long des trajectoires de DM 84
5.3 Population (%) des différents conformères identifiés au long des trajectoires de DM 85
5.4 Hélicité des différents conformères identifiés au long des trajectoires de DM 85

Liste des principales abréviations

Pour des raisons de lisibilité la signification d'une abréviation ou d'un acronyme n'est souvent rappelée qu'à sa première apparition dans le texte. Nous présentons ci-après la liste alphabétique des abréviations les plus usuelles, rencontrées dans la thèse, avec leur traduction en anglais.

AMBER	modélisation assistée avec raffinement de l'énergie	Assisted Model Building with Energy Refinement
Å	Angström	Angström
BSA	Albumine Sérique Bovine	Bovine Serum Albumin
CD	Dichroïsme Circulaire	Circular Dichroisme
DM	Dynamique Moléculaire	Molecular Dynamics
fs	femtoseconde	femtosecond
HSA	Albumine Sérique Humaine	Humain Serum Albumin
HSA$^{(N)}$	Albumine Sérique Humaine Native	Native Humain Serum Albumin
HSA$^{(R)}$	Albumine Sérique Humaine Réduite	Reduced Humain Serum Albumin
kcal	kilocalorie	kilocalorie
NAMD	dynamique moléculaire à l'échelle nanométrique	NAnoscale Molecular Dynamics
ns	nanoseconde	nanosecond
OPLS-AA	potentiels optimisés pour les simulations liquides en tout-atome	Optimized Potentials for Liquid Simulations in All Atom
PC	composante principale	Principal Component
PCA	analyse par composante principale	Principal Component Analysis

PME	Sommation d'Ewald	Particle Mesh Ewald
$\mathbf{R}_{gyr}$	Rayon de gyration	Radius of gyration
$\mathbf{R}_H$	Rayon Hydrodynamique	Hydrodynamic Radius
RMSD	écart quadratique moyen	Root Mean Square Deviation
RSD	déplacement quadratique	Root Square Deviation
RX	Rayon X	X-ray
SDS	dodécylsulfate de sodium	Sodium Dodecyl Sulfate
TIP3P	modèle d'eau moléculaire	water molecular model
WHAM	Méthode d'analyse par histogramme pondéré	Weighted Histogram Analysis Method

Chapitre 1

Introduction

Sommaire

1.1 Les protéines, leurs structures et leurs fonctions 2

1.2 L'albumine sérique humaine 4

1.2.1 Structure globale 4

1.2.2 Les ponts disulfure 5

1.2.3 Les sites de complexation 6

1.3 Questions traitées dans la thèse 8

1.3.1 Effets de la dynamique sur la structure et les sites de liaison de la HSA 8

1.3.2 Rôle des ponts disulfure dans le maintien de la structure native de la HSA et le dépliement réductif de la protéine 9

1.4 Approche méthodologique 10

1.4.1 La dynamique moléculaire et l'analyse par composantes principales 10

1.4.2 Les simulations de dépliement 11

1.5 Présentation des chapitres 12

1.1 Les protéines, leurs structures et leurs fonctions

Les protéines sont des acteurs essentiels de tous les processus du vivant. Elles sont utilisées par les organismes vivants pour effectuer une multitude de fonctions comme la catalyse enzymatique, le transport de métabolites (l'hémoglobine, par exemple, permet de distribuer l'oxygène en provenance des poumons dans tout l'organisme), la communication (de nombreuses hormones comme l'insuline permettent de délivrer un message dans tout l'organisme), la reconnaissance (les anticorps du système immunitaire sont des protéines qui permettent la reconnaissance), ou encore la mobilité cellulaire (l'actine et la myosine sont à l'origine du mouvement cellulaire lors de la contraction du muscle). La fonction d'une protéine est intrinsèquement liée à sa structure. Pour décrire cette conformation, différents niveaux de structure sont utilisés :

- la structure primaire décrit les atomes et les liaisons covalentes qui composent la protéine. Les protéines sont des polymères linéaires, leurs structures primaires sont donc souvent assimilées à leurs séquences. Il faut noter que la séquence d'une protéine, contrairement à sa structure primaire, ne renseigne pas sur les ponts disulfure qui peuvent être formés entre les soufres des cystéines.
- la structure secondaire précise les différentes conformations locales de la chaîne polypeptidique. Deux principaux types de structure secondaire sont distingués : les hélices α et les feuillets β.
- la structure tertiaire spécifie l'arrangement des différentes structures secondaires entre elles. Elle donne donc la forme globale de la protéine, les coordonnées de chacun de ses atomes dans un espace tridimensionnel. La fonction d'une protéine est intrinsèquement liée à sa structure tertiaire. La détermination de cette structure est donc d'une importance capitale.
- la structure quaternaire décrit, lorsque la protéine fonctionnelle est composée de plusieurs chaînes polypeptidiques, l'arrangement de ces différentes chaînes entre elles.

Pour assurer toutes ces fonctions, les protéines doivent former une structure bien définie, et doivent donc être «repliées» correctement. Le repliement incorrect de certaines protéines est à l'origine de diverses pathologies comme les maladies neurodégénératives d'Alzheimer et de Parkinson, ou les maladies impliquant la protéine prion (vache folle, Kreutzfeld-Jacob). La connaissance des mécanismes fondamentaux du repliement et de

la stabilité des protéines, ainsi que de la cinétique d'inter-conversion d'un état à l'autre, reste donc un objectif majeur dans la compréhension des bases moléculaires du vivant.

D'un point de vue expérimental, la première structure tridimensionnelle d'une protéine (la myoglobine) a été déterminée par Max Perutz et John Kendrew[2] par diffraction de rayons X (RX) sur un cristal de protéine. Cette méthode reste toujours la plus utilisée aujourd'hui. 90% des structures de protéine ont été résolues par diffraction et sont accessibles sur les banques de données. La structure tertiaire est aussi accessible par résonance magnétique nucléaire (RMN). La RMN multidimensionnelle permet d'étudier les propriétés structurales et dynamiques de protéines en solution. Mais l'étude des processus cinétiques comme le repliement ou le dépliement des protéines par cette technique est limité par les temps de mesure importants qui la caractérisent. On peut également prédire la structure des protéines et leur repliement par simulations numériques. En effet, tout comme la structure secondaire, la structure tertiaire résulte entre autres de liaisons hydrogènes qui sont décrits dans la structure primaire. La structure tertiaire devrait donc pouvoir être déduite de la structure primaire dès lors que l'on connaisse parfaitement les forces entre les différents atomes de la protéine. Cet essor des expériences «in-silico» est remarquable depuis environ une vingtaine d'années. Une preuve irréfutable de la puissance des calculs pour modéliser le vivant à l'échelle moléculaire est sans doute le prix Nobel de chimie attribué en 2013 aux chimistes Martin Karplus, Michael Levitt et Arieh Warshel pour leurs travaux en chimie théorique et les codes numériques adaptés aux biomolécules. On distingue globalement deux types de méthode de prédiction. La première utilise uniquement la structure primaire de la protéine pour essayer de déterminer la structure secondaire. Cette approche est très couteuse en temps de calcul et nécessite des moyens de calcul gigantesque comme *Blue Gene*[3] ou des calculs distribués, comme c'est le cas avec le projet *Folding@home*[4]. La seconde famille est une méthode travaillant par comparaison. Dans ce cas, on utilise la structure primaire, mais aussi des structures tertiaires déjà existantes. Les structures utilisées pour la prédiction peuvent être celles de protéines de la même famille (homology modeling). Une autre méthode consiste à insérer la séquence de la protéine dans les structures déjà connues et à assigner à ces structures un score caractéristique de sa compatibilité avec la séquence à étudier (protein threading). Les méthodes par comparaison sont beaucoup plus répandues que les premières car elles nécessitent beaucoup moins de temps de calcul. Elles donnent généralement de meilleurs résultats, mais elles ne permettent pas de déterminer toutes les structures des protéines et ont besoin de structures connues pour fonctionner. Il existe actuellement une panoplie de méthodes pour trouver les chemins énergétiques et améliorer la statistique d'échantillonnage des configurations

possibles de la protéine. On peut citer, par exemple la métadynamique, les méthodes de dynamique moléculaire biaisée ou adaptative qui connaissent actuellement un essor considérable[5].

1.2 L'albumine sérique humaine

La protéine faisant l'objet de ce travail de thèse est l'albumine sérique humaine (HSA). Il s'agit de la protéine la plus abondante dans le plasma sanguin humain, elle y est présente à une concentration typique comprise entre 35 et 50 $g.L^{-1}$. La HSA est produite par le foie et présente le double intérêt d'être un vecteur de médicaments et d'hormones de tout premier plan [6–8] ainsi que d'avoir une structure native stabilisée par de nombreux ponts disulfure. Ses propriétés de transport sont directement liées à sa structure complexe (Figure 1.1).

1.2.1 Structure globale

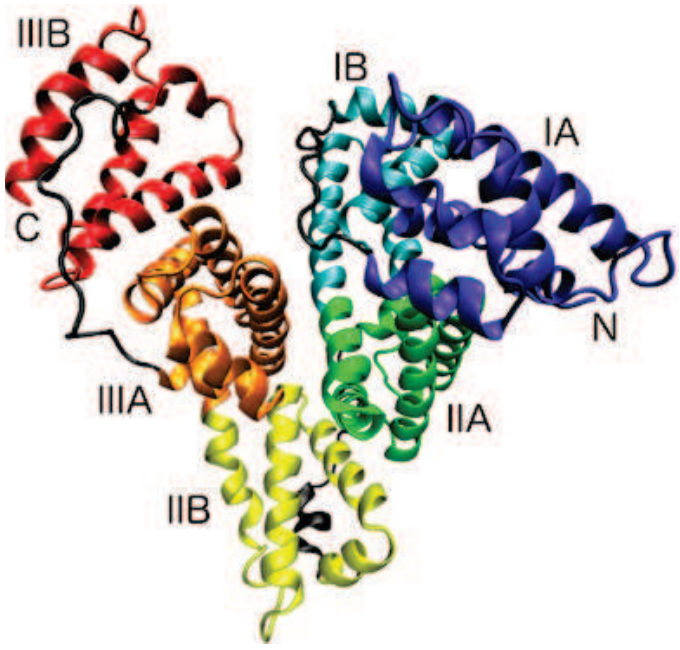

Figure 1.1 – Structure tridimensionnelle de la HSA. Les 6 sous-domaines protéiques (IA à IIIB) sont illustrés par différentes couleurs.

La HSA sous sa forme monomérique est un polypeptide de 585 acides aminés, contenant 17 ponts disulfure, dont les emplacements ont été tout d'abord fournis par des études peptidiques réalisées par Brown [9] et par Sabre et al. [10]. La première détermination de la structure cristalline tridimensionnelle de la HSA a été obtenue par He et Carter à une résolution de 2.8 Angströms (Å) [11]. Pour ce travail, nous avons utilisé une structure plus précise obtenue par Sugio et al. [12] avec une résolution de 2.5 Å. Concernant les structures secondaire et tertiaire, la HSA est organisée en trois domaines α-hélicoïdaux homologues (I à III) [13] à partir de l'amine terminale (Figure 1.1). Chaque domaine est divisé en deux sous-domaines, A et B, composés respectivement de six et quatre hélices α. Les sous-domaines A et B sont reliés par de longues boucles et montrent une structure tridimensionnelle comparable. Les 17 ponts disulfure sont placés à des endroits clés entre les hélices adjacentes (courtes barres verticales grises dans la Figure 1.2) et sont appelés à jouer un rôle conformationnel important.

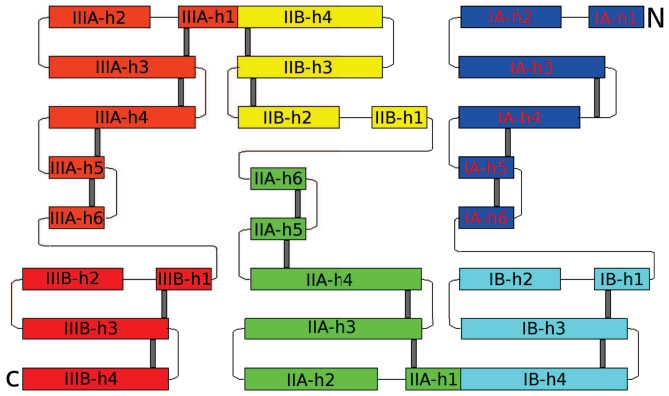

Figure 1.2 – Représentation schématique de la structure de la HSA avec identification des ponts disulfure en bandes grises. Les atomes, amine et carbone, terminaux sont notés respectivement N et C.

1.2.2 Les ponts disulfure

Les liaisons covalentes S—S (ponts disulfure) formées entre les résidus cystéine jouent un rôle particulier dans la stabilité des structures natives de nombreuses protéines. Ceci est un fait clairement établi par des expériences de dépliement/repliement de la protéine réduite/oxydée (Figure 1.3). Des expériences de dénaturation thermique ou en présence d'urée indiquent qu'en moyenne, la stabilité de la conformation de la structure protéique

repliée est réduite d'environ 3 $kcal.mol^{-1}$ lors de la cassure d'un pont S—S [14, 15]. Les ponts S—S stabilisent d'une manière significative la structure native des protéines [15–18]. Dans une interprétation mécanique simple, un pont disulfure est équivalent à une contrainte introduite dans l'espace de phase de la protéine repliée, par rapport à la structure dépliée, réduisant significativement l'entropie de celle-ci [16]. Toutefois, les conséquences structurales de la suppression des liaisons S—S peuvent varier d'une protéine à l'autre. Le dépliement de la protéine induit par la réduction des ponts disulfure affecte principalement les structures tertiaire et quaternaire, la structure secondaire pouvant également être endommagée dans certains cas. Par voie de conséquence, ces modifications structurales peuvent être à l'origine de changements conformationnels importants au niveau des nombreux sites de complexation parcourant la structure de la HSA.

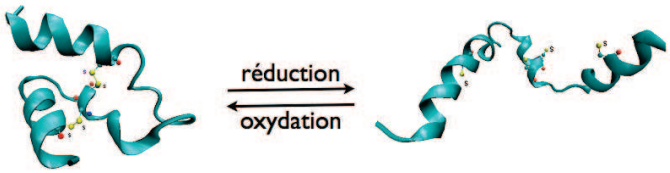

Figure 1.3 – Illustration du processus de dépliement/repliement de la structure protéique induit par la rupture ou la reconstitution des ponts disulfure

1.2.3 Les sites de complexation

De multiples poches hydrophobes et cavités apparaissent dans la structure de la HSA permettant la complexation de nombreux types de ligands. Ces différents sites peuvent accueillir aussi bien des ligands endogènes (comme les acides gras, la bilirubine, l'hémine, la thyroxine, etc...) que des médicaments aux propriétés acides ou électronégatives. Différents sites de complexation de la protéine ont été clairement établi par des études structurales systématiques [6], [1, 13, 19–21]. Sept sites ont été identifiés pour la liaison des acides gras (FA1 à FA7) [1] dont les emplacements sont indiqués dans la Figure 1.4. Les médicaments aux propriétés acides et électronégatives se lient quant à eux principalement dans deux sites principaux, les sites 1 et 2, qui sont situés respectivement dans les sous-domaines IIA et IIIA [22]. Des médicaments comme le paclitaxel, la warfarine, l'azapropazone sont connus pour se fixer préférentiellement au site 1 alors que le diflunisal, l'ibuprofen ou bien encore le diazepam ont une meilleure affinité avec le site 2. Ces deux sites primaires lient aussi les acides gras : en effet, le site 1 est identique au site FA7 alors que le site 2

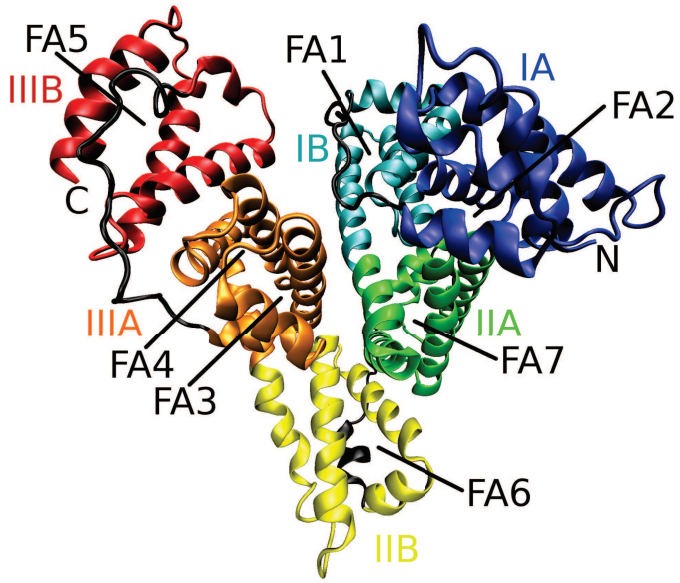

Figure 1.4 – Illustration des sites de liaison les plus répandus pour les acides gras de la HSA

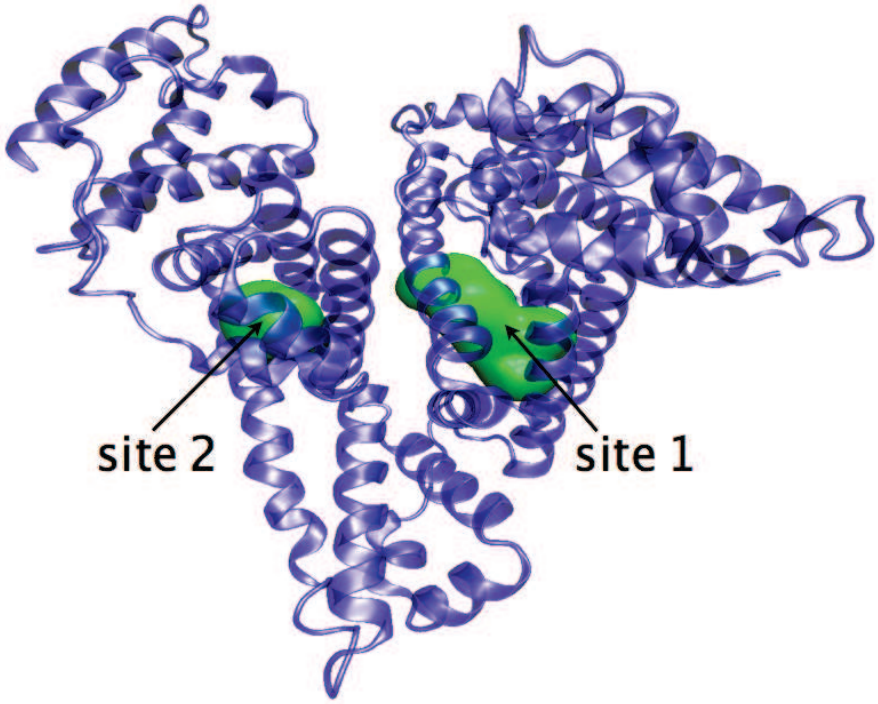

Figure 1.5 – Représentation surfacique des deux sites primaires de liaison à travers la structure secondaire de la HSA

regroupe les sites FA3 et FA4 (Figure 1.5). Fait intéressant, la spécificité du médicament envers les sites de liaison n'est pas très marquée, chaque site étant en mesure de lier des ligands de forme et de taille variables. Ceci est compatible avec une structure protéique flexible où chaque site est à même d'adopter des conformations dynamiques différentes. L'hypothèse actuelle est confortée par une autre propriété de complexation bien connue de la HSA : la modulation allostérique[7]. En d'autres termes, la fixation d'un médicament à un site modifie l'affinité de liaison d'un autre médicament avec un site différent. Par exemple, la liaison du hème au site FA1 réduit la constante de vitesse de liaison au site 1 (FA7), et réciproquement [7]. Des interactions allostériques ont également été signalées entre les sites primaires 1 et 2 [23]. Ce type d' interactions peut être expliqué en supposant que la complexation du ligand induit des changements de conformations qui affectent simultanément au moins deux sites de liaison. Un exemple bien connu est celui de la transition conformationnelle induite par la liaison du myristate [1].

1.3 Questions traitées dans la thèse

1.3.1 Effets de la dynamique sur la structure et les sites de liaison de la HSA

La HSA est la protéine sérique la plus impliquée dans le transport de nombreux composés comme les hormones, les acides gras et les médicaments. De ce fait, cette protéine est l'une des plus étudiées et utilisée expérimentalement. Malgré cela, l'étude de ses conformations natives et de sa dynamique structurale n'a été que récemment envisagée. Pourtant, comme dans le cas de bien d'autres macromolécules biologiques, sa structure n'est pas rigide et des changements dynamiques peuvent jouer un rôle important dans les fonctions physiologiques de la protéine. Par ailleurs, des simulations de dynamique moléculaire (DM) effectuées sur quelques nanosecondes (ns) suggèrent que cette flexibilité structurale est essentielle pour sa fonction de liaison étant donné que les sites de liaison sont préformés par des changements conformationnels structuraux [8]. Des études récentes [8, 24–26] ont apporté des contributions importantes à ce sujet, cependant beaucoup de travail est encore nécessaire pour obtenir une vision globale de la dynamique de la HSA et ses implications dans ses fonctions protéiques. Évidemment, il est très difficile d'obtenir des réponses complètes à ces questions. Une partie importante du travail de thèse est consacrée à ce sujet complexe et permet d'y apporter une contribution intéressante. Nous

avons pour cela mis en place de nouveaux outils d'analyse et améliorer la statistique en effectuant des simulations aux temps longs (centaine de ns).

1.3.2 Rôle des ponts disulfure dans le maintien de la structure native de la HSA et le dépliement réductif de la protéine

Les liaisons covalentes S—S sont connues pour jouer un rôle clé dans la stabilisation de la structure native de diverses protéines. Des mesures de spectroscopie Raman ont montré que la réduction partielle des ponts S—S affecte la structure secondaire de la HSA, comme indiqué par les changements observés dans la bande Raman amide I [27]. Cependant, la manière précise de leur mise en oeuvre n'a pas encore été élucidée bien que plusieurs études expérimentales et théoriques ont été menées jusqu'à présent [28–31]. Il est généralement admis que les ponts S—S agissent comme des contraintes qui réduisent l'espace conformationnel accessible de la protéine résultant à une diminution significative de l'entropie de l'état protéique déplié [28]. Toutefois, les conséquences structurales de la suppression des liaisons S—S peuvent varier d'une protéine à l'autre. Par exemple, la réduction de l'un des quatre ponts S—S de la ribonucléase A produit une espèce qui conserve sa structure native alors que la réduction de deux ponts S—S induit un dépliement spontané de la protéine [16]. Un effet similaire à la ribonucléase A a été observé dans le cas de la restrictocine qui conserve sa structure native lors de la réduction d'un de ses deux ponts S—S mais la perd lors de la réduction simultanée des deux ponts disulfure [32]. D'autre part, la ribonucléase T1 maintient une structure native et une activité enzymatique, même lorsque ses deux ponts S—S sont brisés [14]. Dans le cas des trois ponts S—S de l'interleukine-4, il a été montré que les mutants possédant seulement deux ponts disulfure peuvent conserver ou non leur structure native primitive en fonction de l'ordre selon lequel les ponts ont été réduits [15].

La rupture des ponts donne donc lieu à une structure instable qui peut se déplier spontanément ou non. Le cas de la HSA est particulièrement intéressant car la conformation native de cette protéine est stabilisée par un grand nombre de ponts S—S (17 au total). De plus, le mécanisme de dépliement de cette protéine constituée majoritairement d'hélices α pourrait être un exemple pertinent pour la compréhension du dépliement d'autres protéines dont la structure secondaire est dominé par des hélices α.

Bien qu'aucune structure cristallographique ou à résonance magnétique nucléaire

(RMN) de la HSA réduite n'ait encore été signalée, certaines de ses propriétés générales telles que le contenu en hélices α [14, 15] ou le rayon de gyration [14] ont déjà été décelées par des mesures complémentaires. Par ailleurs, des mesures par dichroïsme circulaire (CD) et de fluorescence du tryptophane indiquent que la HSA réduite se déplie partiellement tout en gardant un état compact. Il a donc été proposé que les structures caractéristiques des protéines réduites sont très proches de celle de l'état *molten globule* [22, 33] : un état où la protéine conserve une forme assez compacte en gardant un pourcentage significatif de sa structure secondaire mais en perdant ses interactions tertiaires spécifiques [34]. Étant donné que la protéine réduite maintient son état compact et conserve une fraction importante de la structure secondaire de la protéine native, il a été proposé que ce conformère partiellement déplié est un état métastable se trouvant au long du chemin de dépliement naturel de la protéine native [14]. Cette hypothèse a également été confortée par des expériences de repliement oxydatif effectuées sur la protéine réduite. En effet, la structure native de la protéine a été récupérée après oxydation de la structure réduite démontrant ainsi la réversibilité du processus de dépliement [14, 32]. De nouveaux éclairages concernant ce processus sont néanmoins nécessaires afin de donner une image claire de la structure de la protéine à l'état réduit. C'est le deuxième objectif de ce travail de thèse.

1.4 Approche méthodologique

1.4.1 La dynamique moléculaire et l'analyse par composantes principales

La flexibilité conformationnelle des sites de liaison de la HSA peut théoriquement être étudiée en effectuant des simulations de DM à une échelle de temps physiquement pertinente. Pour une protéine de la taille de la HSA (67 kDa), des simulations en tout-atome sur des centaines de nanosecondes sont aujourd'hui couramment accessibles. Il s'agit d'une échelle de temps très intéressante pour la dynamique des protéines car elle est assez longue pour comprendre le mouvement relatif des sous-domaines. Cependant, elle reste trop petite pour espérer interpréter le dépliement des protéines. Évidemment, l'analyse de telles trajectoires afin d'en extraire les conformations stables et mettre en évidence les principales caractéristiques de la dynamique protéique est une tâche difficile en raison du grand nombre de degrés de liberté présents. Une dimension réduite pour décrire le mouvement protéique est nécessaire, il faut ainsi avoir recours à un ensemble

de coordonnées réduites conservant les caractéristiques essentielles du mouvement. Une méthode très efficace pour fournir une telle description des changements de conformations des protéines est l'analyse en composantes principales (PCA) [35–41]. Le succès de cette méthode est directement lié aux deux propriétés principales du mouvement interne des protéines : le fait que les mouvements intéressants d'un point de vue fonctionnel sont les déplacements collectifs survenus le long de la direction de certaines coordonnées collectives spécifiques [37] et le fait que très peu de ces coordonnées captent la majeure partie des fluctuations protéiques. Cette dernière propriété, comme l'a déjà souligné Amadei et al. [35], signifie que l'espace configurationel de la protéine peut être divisé en deux sous-espaces : un «essentiel» qui contient les quelques degrés de liberté les plus lents qui composent la fluctuation des atomes la plus globale et le sous-espace restant composé des degrés de liberté rapides qui décrivent de petites fluctuations gaussiennes. Dans le procédé de la PCA, les coordonnées collectives sont obtenues sous forme de vecteurs propres de la matrice de covariance de la fluctuation des positions. Les valeurs propres relatives permettent une identification directe des coordonnées collectives essentielles car elles représentent les poids des différents mouvements collectifs dans la fluctuation globale de la protéine. Il existe plusieurs exemples d'applications réussies de la méthode PCA à extraire le mouvement conformationel essentiel du mouvement protéique produit par une simulation de dynamique moléculaire [36, 39, 42–45]. C'est la méthode que nous avons choisi dans l'étude de la dynamique de la HSA et de ses sites de complexation.

1.4.2 Les simulations de dépliement

Les progrès continus des procédés de simulation de la dynamique moléculaire pour étudier le dépliement des protéines [46–50] en font une alternative intéressante. Cependant, la tâche reste complexe puisque les simulations de dépliement protéique sont confrontées à deux problèmes majeurs.

Le premier provient du fait qu'il n'est pas encore possible de produire un niveau significatif de dépliement des protéines par simulation de DM à température ambiante pour une échelle de temps accessible (à l'exception des peptides courts). En effet, l'échelle de temps de calcul accessible est au moins un ordre de grandeur inférieur à la durée d'un processus de dépliement naturel (millisecondes ou plus) [51]. Plusieurs méthodes d'accélération de la dynamique de simulation ont été proposées [52–55]. Parmi ces techniques, l'élévation de la température de simulation a été systématiquement testée avec des résultats très

encourageants [47, 48, 56–58]. Cette technique est basée sur l'hypothèse que la surface d'énergie libre n'est pas sensiblement affectée par l' augmentation de la température dans un intervalle restreint et donc que le chemin de dépliement est pratiquement indépendant de la température. Cette hypothèse a été vérifiée directement par Daggett et al. [58] dans le cas d'un inhibiteur de la chymotrypsine 2 qui se déplie. Toutefois, les limites de la validité de cette méthode sont encore en discussion. Par exemple, Wang et Wade ont montré que, dans ce type de simulations, l'élévation de la température doit être contrôlée avec soin pour éviter tout dépliement anormal [48]. En outre, il a été montré expérimentalement que le chemin de dépliement de la protéine n'est pas unique, mais très dépendant des conditions initiales [59, 60].

Le second problème est lié à la tendance commune des champs de force de mécanique moléculaire (MM) à stabiliser les structures repliées. En effet la grande majorité de ces champs de force ont été calibrés pour reproduire les structures natives.

Les simulations de DM à haute température ont prouvé être au moins une solution partielle au premier problème [33, 53, 61]. Concernant le second problème, des efforts importants ont été portés sur les paramètres des potentiels de torsion du squelette protéique afin d'harmoniser l'équilibre conformationnel entre les structures en hélices et celles dépliées. De nombreuses simulations de dépliement en DM qui ont été réalisés montrent qu'il est possible de simuler le dépliement/repliement de petites protéines avec un bon accord qualitatif et même quantitatif avec les données expérimentales [48, 50, 62].

Dans ce travail de thèse, nous avons choisi d'accélérer le dépliement simulé en utilisant des températures de simulation supérieures à la température ambiante. Afin d'identifier les étapes principales du dépliement, les conformations moléculaires ainsi générées ont été soumises à la PCA.

1.5 Présentation des chapitres

Le *chapitre 2*, de ce manuscrit est dédié à la présentation générale des méthodes théoriques utilisées dans le présent travail. Les détails les plus importants de ces méthodes sont rappelés dans les chapitres de présentation des résultats afin de faciliter le suivi du travail.

Le *chapitre 3* rapporte une étude détaillée de la dynamique conformationnelle de la HSA basée sur la méthode PCA appliquée à trois trajectoires de DM simulées sur 200 ns chacune. Dans un premier temps, les mouvements d'ensemble de la structure globale de la protéine et de ses conformères les plus stables sont étudiés. Une analyse détaillée des changements structuraux se produisant au niveau des sites primaires 1 et 2 est ensuite effectuée. Les différents mouvements et conformères ainsi trouvés sont comparés aux changements structuraux observés lors de la complexation de différents ligands.

Dans le *chapitre 4*, nous avons d'abord examiné la capacité de deux champs de force de MM couramment utilisés en DM, OPLS-AA [63] et AMBER ff03 [64], à décrire le dépliement de la HSA réduite. Notre analyse se base sur une comparaison entre le dépliement thermique des protéines native et réduite. Les résultats sont discutés en relation avec les données expérimentales disponibles sur la structure de la protéine réduite. Dans ce chapitre, le dépliement des HSA native et réduite a été simulé en effectuant des DM de 20 ns chacune à des températures élevées. L'analyse se concentre principalement sur les changements structuraux affectant les structures tertiaire et secondaire de la protéine réduite. Une attention spéciale est accordée au mécanisme de dissociation des ponts S—S réduits. La comparaison du dépliement des deux structures protéiques permet d'apporter un premier éclairage sur le rôle des ponts S-S dans la stabilisation de la structure native de la HSA.

L'étude systématique du processus de dépliement de la protéine réduite est abordé et fait l'objet du *chapitre 5*. Le chemin de dépliement de la protéine réduite est présenté comme une série ordonnée de conformères protéiques partiellement dépliés. Ces structures ont été obtenues en effectuant des simulations de DM de 160 ns chacune à quatre températures différentes : 350K, 375K, 400K et 425K. Chaque trajectoire est soumise à une analyse en composantes principales [35–38, 65] afin d'identifier les principaux mouvements collectifs de la protéine et d'explorer le regroupement de conformations protéiques donnant lieu à des conformères stables. Sur cette base, nous avons examiné l'évolution de la structure tertiaire de la protéine pendant le dépliement, la stabilité relative des différents types d'hélices α (longues et courtes) et la présence éventuelle d'un état métastable le long du chemin de dépliement. La prédiction d'une structure réduite d'équilibre est largement traitée. Les modifications structurales des sites de liaison font aussi l'objet d'une attention particulière.

Chapitre 2

Méthodes de modélisation moléculaire

Sommaire

2.1 Dynamique moléculaire classique 17
2.1.1 Structure initiale et système de simulation 17
2.1.2 Effets de bord 18
2.1.3 Champ de forces 20
2.1.3.1 Les interactions liées 21
2.1.3.2 Les interactions non-liées 23
2.1.3.3 Les interactions à longue distance 24
2.1.4 Algorithme de la dynamique moléculaire 28
2.1.5 Contrôle de la température 30
2.1.6 Contrôle de la pression 30
2.1.7 Protocoles de simulation 31
2.1.7.1 Minimisation en énergie 31
2.1.7.2 Production 32
2.1.8 Analyse des simulations de dynamique moléculaire 33
2.1.8.1 RMSD et RMSF 33
2.1.8.2 Rayon de gyration 34
2.1.8.3 Hélicité 34

2.1.8.4 Contacts natifs 35
2.1.8.5 Calculs des profils d'énergie libre 35
2.1.9 La visualisation 38
2.2 Analyse par composantes principales 38
2.2.1 Principe 38
2.2.2 Convergence de la méthode 40

2.1 Dynamique moléculaire classique

La dynamique moléculaire est une technique numérique qui permet d'étudier le comportement de la matière à l'échelle atomique. Elle a été introduite dans les années 50 par Alder et Wainwright [66] et est aujourd'hui une discipline en plein essor grâce au développement de la puissance de calcul. Le prix Nobel de chimie 2013 attribué à Martin Karplus, Michaël Levitt et Arieh Warschel pour leurs travaux sur la modélisation moléculaire illustre parfaitement cela. La DM consiste à étudier le mouvement de molécules ou d'atomes en leur appliquant les lois de la mécanique newtonienne. Les interactions entre les particules et l'énergie potentielle sont définies par les champs de force de mécanique moléculaire. Il s'agit ainsi de simuler leurs trajectoires au cours du temps. L'exploitation de ces résultats permet d'approximer les propriétés structurales, dynamiques et thermodynamiques du système simulé. Pour déduire des conclusions pertinentes des simulations de DM, il faut veiller à ce que le temps de simulation corresponde à la cinétique du processus naturel étudié. Une simulation de DM comporte différentes étapes : la construction du système et sa solvatation, sa minimisation en énergie puis la mise en dynamique du système et la phase de production.

2.1.1 Structure initiale et système de simulation

Structure atomique par diffraction de rayons X

Depuis 1958 et la première résolution de la structure 3D de la myoglobine par J. Kendrew [2], le nombre de structures obtenues par diffraction aux rayons X (RX) n'a cessé de croitre. Cette technique permet aujourd'hui d'obtenir des structures de macromolécules d'intéret biologique à résolution quasi-atomique. Le principe consiste à déterminer la position des atomes afin d'en déduire un modèle structural. La première étape est la cristallisation de la protéine, elle consiste à abaisser très graduellement la solubilité de la protéine afin d'éviter la précipitation. Une fois construit, le cristal est ensuite soumis à un faisceau de RX monochromatiques. Les RX interragissent avec le nuage électronique des atomes du cristal et sont diffractés par les plans du réseau cristallin et récupérés sous la forme d'une figure de diffraction. La figure de diffraction est une représentation du réseau cristallin dans l'espace réciproque, mais également la transformée de Fourier de la densité électronique du cristal. La détermination des maxima de densité électronique nous informe

alors sur la position des atomes.

Système de simulation

La structure tridimensionnelle de la HSA sur laquelle nous nous sommes appuyés au cours de ce travail a été déterminée, après cristallisation, par diffraction aux RX avec une résolution de 2.5 Å (pdb 1AO6) [12]. Tout au long de ce manuscrit, les protéines natives c'est à dire celles qui possèdent des liaisons $S-S$ sont nommées HSA$^{(N)}$. La configuration de départ pour la simulation du système réduit HSA$^{(R)}$ possède la même structure cristalline que celle utilisée pour la protéine native (pdb 1AO6). Le protocole de réduction numérique consiste simplement à couper les ponts disulfures $S-S$ et les remplacer par des groupements thiols $-SH$. Afin de diviser le temps de calcul par un facteur 2, nous restreignons le système protéique à un monomère. L' étape suivante consiste à solvater la protéine en créant une boîte d'eau cubique autour de celle-ci. Le modèle retenu pour les molécules d'eau est le TIP3P [67]. Il s'agit d'un modèle semi-empirique simple, basé sur les charges ponctuelles des 3 noyaux atomiques, couramment utilisé en DM dans les milieux biologiques. Afin de neutraliser la solution et de retrouver les conditions du plasma sanguin humain, la dernière étape consiste à ajouter des ions sodium et chlorure à concentration physiologique.

Au final, le système de travail est une boîte cubique de 100 Å de côté, contenant 582 acides aminés, 46 ions sodium pour 33 ions chlorure et 28157 molécules d'eau (Figure 2.1).

2.1.2 Effets de bord

Il faut noter qu'un grand nombre de molécules est généralement située en surface de l'échantillon numérique (boîte de simulation) et des effets de bords ne représentant pas le système peuvent rapidement devenir prépondérants. Obtenir des résultats viables à partir d'un système comme celui-ci implique donc de devoir s'affranchir des effets de bord qui règnent aux extrémités d'une telle boîte de dimension finie. Le recours à des conditions aux limites périodiques s'avère un bon moyen de réduire ces effets indésirables. Les atomes du système simulé sont ainsi répliqués dans des boîtes identiques dans toutes les directions de l'espace. Au final le système complet est composé de 27 boîtes identiques, images l'une de l'autre par translation de la boîte mère dans les 3 directions de l'espace. La périodicité

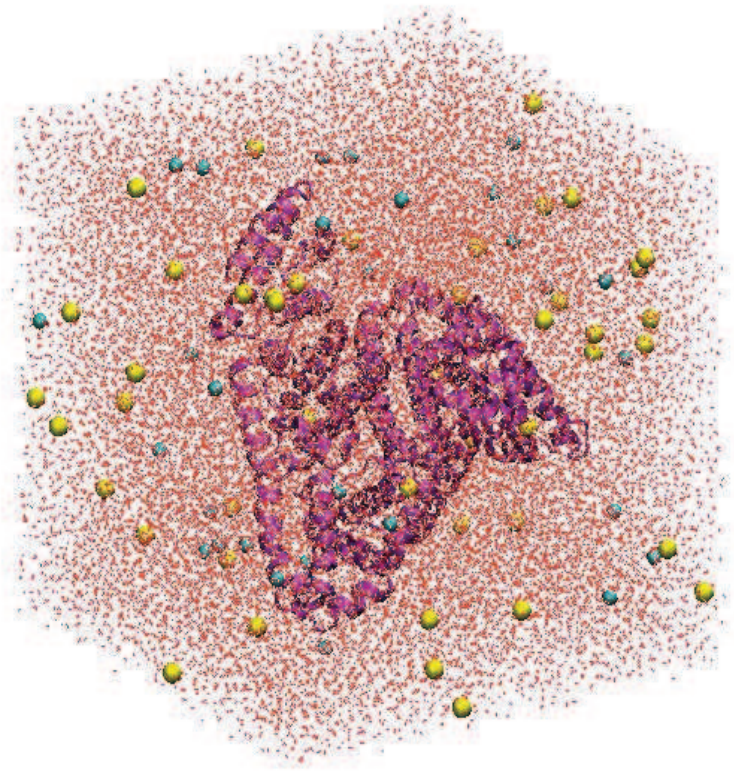

Figure 2.1 – Système de simulation typique pour la HSA

implique que lorsque qu'un atome sort de la boîte «mère» par un côté, ses répliques dans les cellules voisines font de même simultanément. Ceci permet de maintenir le nombre d'atomes dans la cellule constant.

Les atomes proches des bords de la boîte mère sont ainsi entourés de solvant comme un atome situé au centre. L'environnement des atomes devient indépendant de leur position. La taille de la boîte de simulation doit être assez grande afin d'éviter que les artéfacts de périodicité deviennent trop importants. Dans nos simulations, la taille standard est de 100 Å de côté. Durant la simulation, seules les propriétés de la cellule unitaire sont traitées.

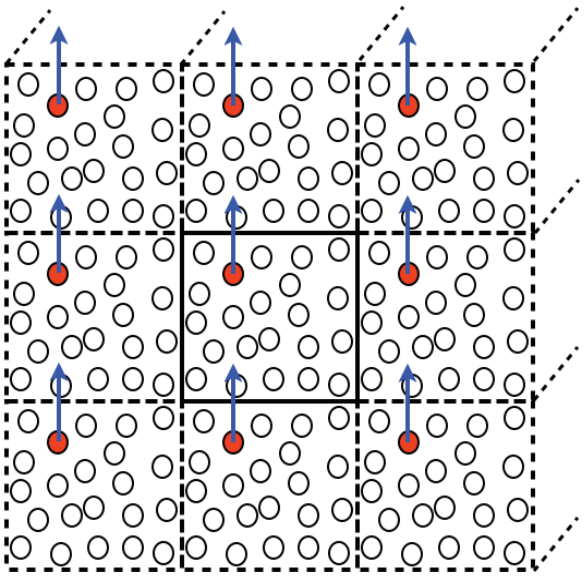

Figure 2.2 – Vue bidimensionnelle des conditions périodiques. La boîte mère est située au centre.

2.1.3 Champ de forces

Le champ de forces est un ensemble de potentiels et de paramètres dont les données permettent de déterminer l'énergie potentielle totale du système. Les champs de forces sont déterminés de façon empirique ou à partir de résultats de calculs quantiques. Un champ de force est, en général, construit et utilisé dans le cadre de l'approximation de Born-Oppenheimer (découplage des mouvements des atomes et des électrons). Ainsi, seules les positions des coeurs atomiques sont nécessaires à la DM. Les paramètres entrant dans le champ de force prennent en compte les interactions des électrons avec les noyaux et les interactions électrons-électrons (liaisons). De nombreux champs de force existent en modélisation de systèmes biologiques et varient fortement en fonction des systèmes étudiés. Le choix d'un champ, plus ou moins efficace, dépend des études que l'on doit produire et des propriétés recherchées.

Considérons un système de simulation, composé de n atomes de masse m_i positionnés en r_i, l'hamiltonien du système peut s'écrire alors sous la forme suivante :

$$H = \sum_{i=1}^{n} \frac{1}{2} m_i \dot{r}_i^2 + V \tag{2.1}$$

où V est le potentiel d'interaction entre atomes. Ces interactions sont de deux types, liées et non liées. Les interactions non-liées s'opèrent entres atomes de différentes molécules mais aussi entre atomes d'une même molécule séparés par quatre liaisons ou plus. Les différentes contributions qui décrivent le potentiel d'interaction sont représentées sur la Figure 2.3 et explicitées dans l'équation suivante

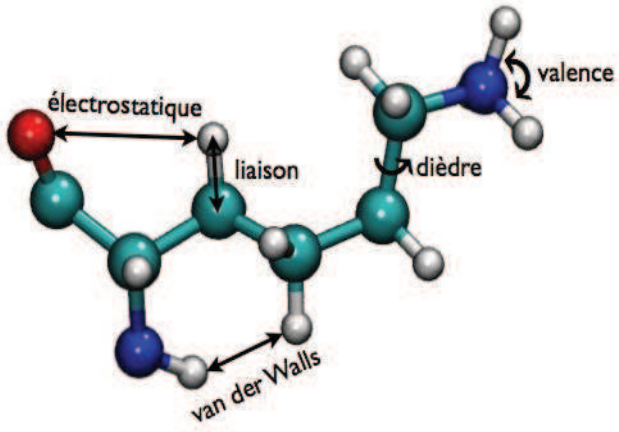

Figure 2.3 – Contributions au potentiel d'interaction moléculaire

$$V = \underbrace{V_{liaison} + V_{valence} + V_{diedre}}_{V_{liées}} + \underbrace{V_{vdW} + V_{elec}}_{V_{non-liées}} \tag{2.2}$$

2.1.3.1 Les interactions liées

Le potentiel des interactions liées prend en compte toutes les interactions de liaison, ce qui correspond à trois composantes énergétiques : l'énergie d'élongation des liaisons covalentes, l'énergie de déformation des angles de valence et l'énergie de torsion des angles dièdraux. Ces énergies décrivent toute modification conformationnelle d'une molécule par rapport à sa structure de plus basse énergie. Les constantes de liaison et les valeurs à l'équilibre sont déterminées empiriquement. Les termes associés aux énergies d'élongation des liaisons covalentes et de déformation des angles de valence sont modélisés par des potentiels harmoniques :

$$V_{liaison} = \sum_{liaisons} K_l(l - l_0)^2 \tag{2.3}$$

$$V_{valence} = \sum_{angles} K_\theta(\theta - \theta_0)^2 \tag{2.4}$$

avec K_l et K_θ les constantes de force associées aux liaisons et aux angles de valence, l et θ la longueur de la liaison et la valeur de l'angle de valence, l_0 et θ_0 leurs valeurs d'équilibre. Ce type de potentiel ne prend évidemment pas en compte la formation ou la rupture de liaisons covalentes qui sont par nature des processus très anharmoniques. L'énergie de torsion des dièdres ne peut être modélisée par un simple potentiel harmonique puisqu'elle présente plusieurs extrema. Le potentiel de torsion des dièdres est généralement défini de la façon suivante :

$$V_{diedre} = \sum_{diedres} K_\phi(1 + cos(n\phi - \gamma)) \tag{2.5}$$

avec $2K_\phi$ qui correspond à la valeur de la barrière de potentiel de l'angle dièdre ϕ, n sa multiplicité et γ sa valeur d' équilibre.

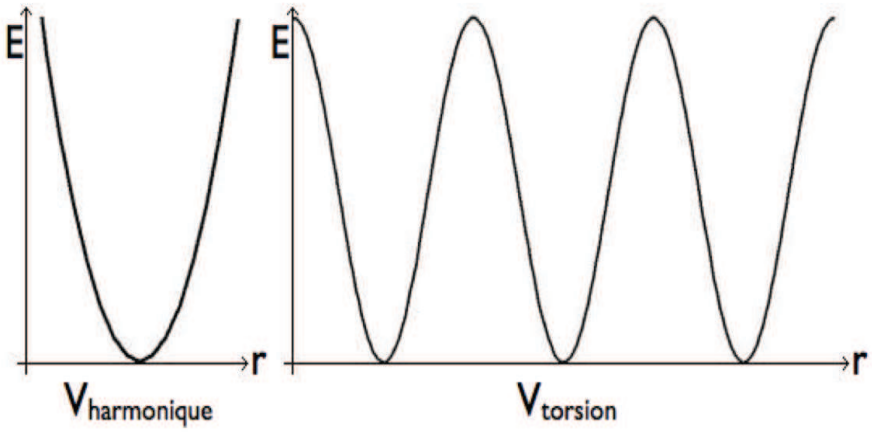

Figure 2.4 – Représentation des interactions liées

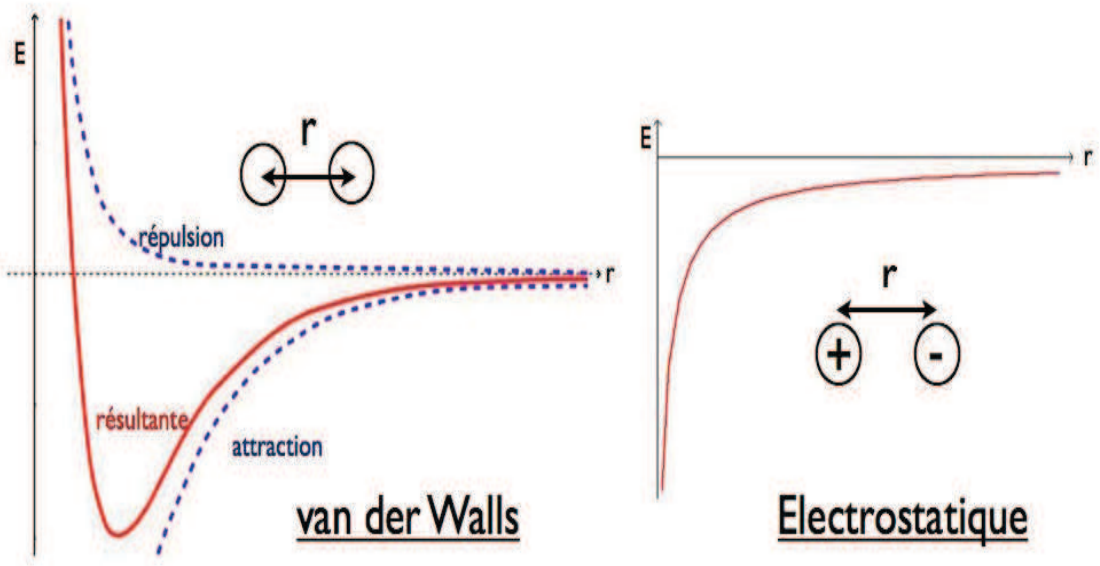

Figure 2.5 – Représentation des interactions non-liées

2.1.3.2 Les interactions non-liées

Le potentiel des interactions non-liées comprend deux types d'interactions : les interactions de van der Waals et les interactions électrostatiques. Les interactions de van der Waals sont souvent modélisées par un potentiel 12-6 de Lennard-Jones :

$$V_{vdW} = \sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.6}$$

où ϵ_{ij} est la valeur absolue de la profondeur du puits de potentiel à la distance d'équilibre, r_{ij} la distance séparant les noyaux des paires d' atomes i et j et σ_{ij} est la distance particulière qui annule le potentiel. À partir de la connaissance des paramètres monoatomiques, on détermine les paramètres du potentiel de Lennard-Jones entre paires d'atomes, ϵ_{ij} et σ_{ij}, en effectuant des moyennes géométriques ou arithmétiques selon les règles de combinaison de Lorentz-Berthelot :

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{2.7}$$

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \tag{2.8}$$

Le terme en $\frac{1}{r^6}$ décrit les interactions attractives dites de dispersion. Ces interactions résultent de l'attraction entre les moments dipolaires instantanés qui sont causés par les fluctuations temporaires de la distribution électronique et les moments dipolaires qu'ils vont induire par effet d'induction. À plus courte distance, cette attraction est fortement dominée par les interactions de répulsion qui deviennent prépondérantes. Cette répulsion, plus connue sous le nom d'interaction d'échange, est la conséquence du principe de Pauli qui empêche les nuages électroniques atomiques de se pénétrer. Elle peut être modélisée par une force en $\frac{1}{r^{12}}$ ou dans certains cas par un terme exponentiel.

Les interactions électrostatiques qui sont dues à l'interaction entre deux particules chargées électriquement sont décrites par un potentiel de Coulomb dans le cadre de l'approximation monopolaire :

$$V_{elec} = \sum_i \sum_{j \neq i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{2.9}$$

où ϵ_0 est la permittivité diélectrique du vide et q_i, q_j les charges atomiques partielles des atomes i et j, calculées à l'aide des méthodes quantiques, séparées par la distance r_{ij}.

2.1.3.3 Les interactions à longue distance

La cellule de simulation construite est pseudo-infinie et le calcul des interactions sur l'atome rouge (Figure 2.6) dans la boîte centrale prend en compte les autres atomes de la boîte ainsi que ceux des boîtes environnantes. Les interactions avec les boites images sont ainsi prises en considération mais les calculs sont effectués uniquement dans la boîte mère. Afin de limiter le coût en calculs, il est nécessaire de faire certaines approximations sur le calcul des interactions atomiques.

Approximation de l'image minimale

La première approximation dite de "l'image minimale" suppose que chaque atome i de la cellule mère interagit seulement avec l'image de l'atome j la plus proche de lui. Cette première approximation en implique une seconde : pour chaque atome i les interactions avec les atomes j sont nulles au delà d'un rayon de troncature R_c.

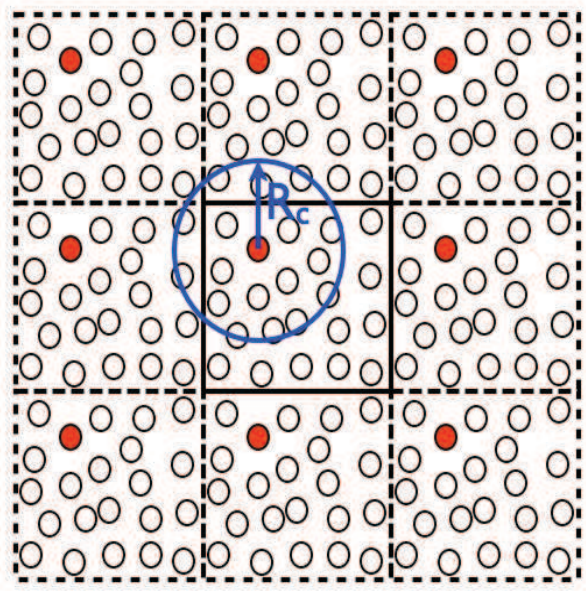

Figure 2.6 – Représentation de la méthode du cut-off

Distance de troncature sphérique

Bien que défini arbitrairement, la convention de l'image minimale implique que ce rayon doit être inférieur à la moitié de la longueur du plus petit côté de la boîte. Dans nos simulations, nous utilisons un cut-off de 11 Å. Ces approximations sont suffisantes lorsque la portée des interactions est faible. Ainsi pour les interactions de type van der Waals, qui sont à courte portée, elles ne posent aucun problème. Ce n'est pas le cas pour les interactions électrostatiques entre charges ponctuelles, où les interactions sont à longue portée (en $\frac{1}{r}$). La taille du système et donc le nombre d'atomes à traiter augmente en r^3 alors que le potentiel électrostatique lui est régi par une loi en $\frac{1}{r}$ ce qui implique une convergence très lente du potentiel. Afin de contourner ce problème, il est préconisé d'utiliser une méthode de type somme de réseau. L'utilisation d'un cut-off ne réduit pas significativement le temps de calcul des interactions non liées. En effet, préalablement à cette étape, le code de calcul utilisé (NAMD (NAnoscale Molecular Dynamics) 2.9b2 [68] en l'occurence) doit lister les paires d'atomes dont il faut calculer les interactions et ceci est une opération coûteuse. Pour cette raison, la mise à jour de la liste est seulement effectuée périodiquement. Pour le calcul des interactions non liées, nous utilisons la méthode du double cut-off (Figure 2.7). Cette méthode a été développée par Berendsen et al. en 1985 [69]. Elle permet de gagner un temps précieux dans le calcul des interactions non-liées. Le

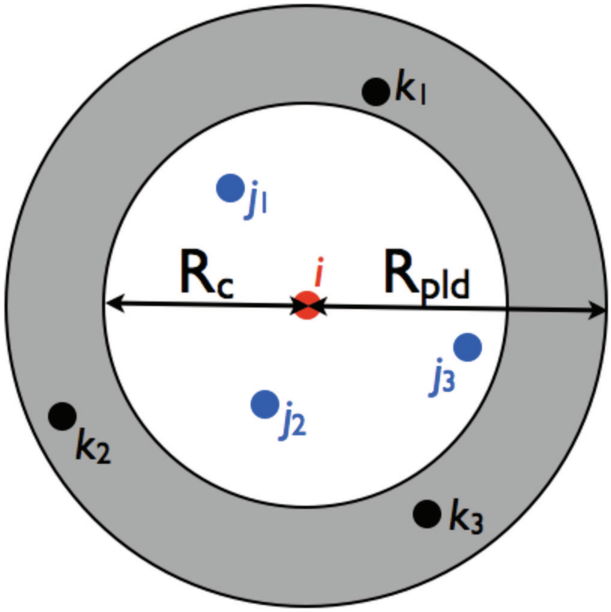

Figure 2.7 – Représentation de la méthode du double cut-off

principe consiste à définir deux rayons de troncature, un premier de rayon R_c et un second de rayon R_{pld} vérifiant la condition $R_{pld}>R_c$. Dans nos simulations, nous utilisons une distance R_{pld} de 13 Å. Pour chaque atome i, les atomes j (interagissant de manière non-liée avec i) qui se trouvent à une distance inférieure à R_c sont stockés dans une liste "*pair-list*" PL_i. Les énergies d'interaction des atomes k qui sont situés dans l'intervalle $R_{pld}>r_{ik}>R_c$ sont stockées dans une seconde variable nommée E_k^{ILD} pour Energie des Interactions à Longues Distances et maintenue constante pour un certain nombre de pas d'intégration. Les énergies d'interactions non liées sont donc la somme des deux contributions.

La discontinuité de l'énergie, à la distance de troncature, est source de nombreux problèmes dans le calcul des énergies et cause des aberrations dans les mouvements protéiques [70–72]. Pour corriger ce problème, il est nécessaire d'estimer les interactions aux environs du cut-off. Pour cela, nous utilisons deux méthodes. Une première qui consiste à faire tendre les interactions vers zéro pour des distances légérement inférieures au cut-off (à l'aide des fonctions de commutation) et une seconde qui consiste à estimer les interactions au delà du cut-off (par la sommation d'Ewald). Ces deux méthodes sont détaillées dans les paragraphes suivants.

Fonction de commutation

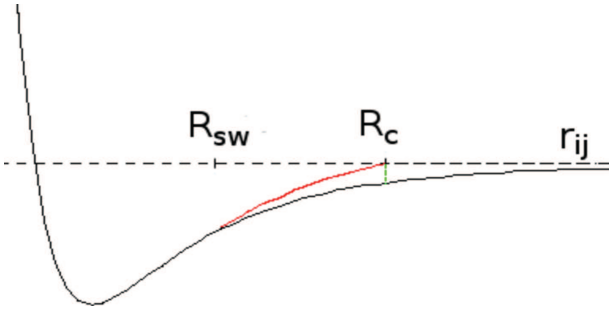

Figure 2.8 – Potentiel électrostatique avec (rouge) ou sans (noir) l'application de la fonction de *switching*.En vert, la discontinuité du potentiel électrostatique du à l'utilisation du cut-off.

.

Cette méthode consiste à lisser les forces et les énergies pour les rendre nulles aux abords du cut-off (R_c). Elle présente l'avantage d'éviter toute discontinuité dans le champ de force. Dans nos simulations, nous utilisons plus précisement la fonction *switch* qui permet de modifier le potentiel uniquement au voisinage de R_c à partir de la distance R_{sw}<R_c (Figure 2.8). Pour l'ensemble des simulations R_{sw} est fixé à 8 Å. Les fonctions d'énergie électrostatique et van der Waals sont alors multipliées par la fonction suivante :

$$S_{switch}(r_{ij}, R_c, R_{sw}) = \begin{cases} 1 & \text{si } r_{ij} < R_c \\ \frac{(R_c^2 - r_{ij}^2)(R_c^2 + 2r_{ij}^2 - 3R_{sw}^2)}{(R_c^2 - R_{sw}^2)^3} & \text{si } R_{sw} < r_{ij} < R_c \\ 0 & \text{si } r_{ij} > R_c \end{cases} \tag{2.10}$$

Méthode de la sommation d'Ewald

La sommation d'Ewald, qui est un cas particulier de la resommation de Poisson, permet d'estimer le potentiel électrostatique. Le traitement des interactions coulombiennes se fait en décomposant le potentiel en deux contributions. Une première qui correspond aux interactions à courte portée dont le calcul se fait dans l'espace réel et une seconde correspondant aux interactions à longue distance dont le traitement s'effectuera dans l'espace réciproque. La somme dans l'espace réciproque étant basée sur la transformée de Fourier, elle converge rapidement lorsque la distance r augmente. L'utilisation de la transformée de Fourier rapide (FFT) optimise considérablement le temps de calcul. La méthode PME (Particle Mesh Ewald) [73] que nous utilisons, repose sur une décomposition

analogue du potentiel mais permet de calculer encore plus efficacement le potentiel. Les charges atomiques sont interpolées tridimensionnellement sur une grille et la distribution de charge correspondante est calculée. Par FFT direct, on évalue tout d'abord le potentiel réciproque puis par inversion du processus, on détermine le potentiel direct. Par dérivation du potentiel ainsi calculé, on détermine les forces électrostatiques. La dernière étape consiste à interpoler les valeurs du champ et du potentiel de la grille vers les atomes et ainsi mettre à jour la position de ceux-ci. Cette méthode permet un traitement efficace des interactions longue distance pour un coût en temps de calcul raisonnable puisqu'il varie en $n \cdot \log(n)$.

2.1.4 Algorithme de la dynamique moléculaire

La DM classique consiste à résoudre numériquement les équations de Newton attachées à chaque atome i. Soit :

$$\overrightarrow{F}_i = m_i \frac{\partial^2 \overrightarrow{r}_i(t)}{\partial t^2} = -\nabla_{r_i} V \tag{2.11}$$

où $\overrightarrow{F}_i$ est la force agissant sur l'atome i exercée par tous les atomes différents de i, m_i la masse de l'atome i et $\overrightarrow{r}_i$ la position de l'atome i. La force $\overrightarrow{F}_i$ exercée sur chaque atome i dérive du potentiel d'interaction V par rapport à la position $\overrightarrow{r}_i$.

L'algorithme préférentiel utilisé pour résoudre les équations du mouvement est l'algorithme de Verlet amélioré. Il permet de déterminer position et vitesse d'un atome i au temps $t + \delta t$ connaissant ses position, vitesse au temps t et son accélération aux temps t et $t + \delta t$. Le principe consiste à écrire le développement en série de Taylor à l'ordre 3 de la position dans les deux directions temporelles :

$$r_i(t + \delta t) = r_i(t) + v_i(t)\delta t + \frac{1}{2}a_i(t)(\delta t)^2 + \frac{(\delta t)^3}{6}\frac{\partial^3 r_i(t)}{\partial t^3} + O((\delta t)^4) \tag{2.12}$$

$$r_i(t - \delta t) = r_i(t) - v_i(t)\delta t + \frac{1}{2}a_i(t)(\delta t)^2 + \frac{(\delta t)^3}{6}\frac{\partial^3 r_i(t)}{\partial t^3} - O((\delta t)^4) \tag{2.13}$$

En les sommant, on obtient l'algorithme de propagation des positions :

$$r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + a_i(t)(\delta t)^2 + O((\delta t)^4) \tag{2.14}$$

En les soustrayant, celui des vitesses :

$$v_i(t) = \frac{r_i(t+\delta t) + r_i(t-\delta t)}{2\delta t} + O((\delta t)^2) \tag{2.15}$$

L'erreur associée à cette dernière expression est en $(\delta t)^2$ et non plus en $(\delta t)^4$. Pour surmonter cette difficulté, plusieurs variantes de cet algorithme existent. En particulier l'algorithme *Leap-frog*. Dans cet algorithme, les vitesses sont calculées pour des demi-entiers :

$$v_i(t + \frac{\delta t}{2}) = \frac{r_i(t+\delta t) - r_i(t)}{\delta t} \tag{2.16}$$

$$v_i(t - \frac{\delta t}{2}) = \frac{r_i(t) - r_i(t-\delta t)}{\delta t} \tag{2.17}$$

permettant d'exprimer les positions de la manière suivante :

$$r_i(t \pm \delta t) = r_i(t) \pm v_i(t + \frac{\delta t}{2})\delta t \tag{2.18}$$

et les vitesses ainsi :

$$v_i(t + \frac{\delta t}{2}) = v_i(t - \frac{\delta t}{2}) + a_i(t)\delta t \tag{2.19}$$

Dans la pratique, la connaissance de la vitesse au temps $t - \frac{\delta t}{2}$ permet le calcul de la vitesse au temps $t + \frac{\delta t}{2}$ via la relation :

$$v_i(t) = \frac{v_i(t - \frac{\delta t}{2}) + v_i(t + \frac{\delta t}{2})}{2} \tag{2.20}$$

Il est alors possible de calculer les postions des atomes au temps t.

Le calcul de la vitesse au temps t_0 nécessite de connaître la vitesse au temps $t = t_0 - \frac{\delta t}{2}$ ce qui n'est pas toujours possible. Le recours à une loi de distribution de Maxwell-Boltzman permet de générer aléatoirement des vitesses initiales à la température désirée.

Le choix du pas d'intégration des équations du mouvement est un compromis entre rapidité du temps de calcul et précision des résultats. La principale contrainte est qu'il doit impérativement être inférieur aux fréquences de vibrations les plus élevées. Les fréquences de vibration les plus élevées sont dues aux liaisons, spécialement celles impliquant l'hydrogène. Par exemple, une liaison C—H vibre avec une période de 10 femtosecondes (fs). Le pas d'intégration δt choisi est ainsi de 2 fs dans nos études.

La résolution des équations de Newton conserve, en principe, l'énergie totale du système. De ce fait, en fixant le volume V de la boîte de simulation et le nombre N

d'atomes, on travaille alors dans l'ensemble microcanonique (N,V,E). Dans nos simulations, nous travaillons préférentiellement à température (ensemble canonique N,V,T) et pression constantes (ensemble isotherme-isobare N,P,T). Pour contrôler ces deux variables, nous couplons le système à un bain thermique et/ou barostatique.

2.1.5 Contrôle de la température

Plusieurs méthodes existent pour réguler la température d'un système à une valeur constante. La méthode que nous avons utlisée consiste à modifier l'équation du mouvement de Newton (équation 2.11) pour chaque atome en y ajoutant des termes stochastiques et de friction pour obtenir une équation de Langevin :

$$m_i \frac{\partial \overrightarrow{v}_i(t)}{\partial t} = \overrightarrow{F}_i - \gamma_i m_i \overrightarrow{v}_i + \overrightarrow{R}_i(t) \tag{2.21}$$

où γ_i est le terme de friction qui s'oppose au mouvement de l'atome i. On peut voir également ce terme comme une fréquence de collisions.
$\overrightarrow{R}_i(t)$ est une force aléatoire due aux collisions associées aux mouvements browniens des atomes environnants vérifiant la relation :

$$< \overrightarrow{R}_i(t) >= 0 \tag{2.22}$$

Les termes $\overrightarrow{R}_i(t)$ et γ_i sont reliés entre eux via la relation :

$$\mid R_i(t) \mid = \sqrt{\frac{2 m_i \gamma_i k_B T}{\Delta t}} \delta(t) \tag{2.23}$$

$\delta(t)$ représente des nombres aléatoires tirés d'une distribution gaussienne à chaque pas Δt. L'équation de Langevin permet de coupler le mouvement de chaque atome à un bain thermique de température T. L'échange est continu. Les atomes sont thermalisés par l'intermédiaire de leurs vitesses, qui sont elles-mêmes couplées aux termes de friction γ_i.

2.1.6 Contrôle de la pression

Pour réguler la pression dans le système de simulation, nous avons eu recours à la méthode du système étendu que développa Andersen en 1980 [74]. Le principe est identique à celui utilisé pour contrôler la température. Il consiste à coupler le système à un volume externe variable V, équivalent au volume de la boîte de simulation. Ce couplage

symbolise l'action qu'excercerait un piston barostatique de masse M_p sur le système, appelé piston de Langevin [75]. Au cours des simulations, la pression du système est controlée par ajustement dynamique de la taille de la cellule unitaire, les coordonnées atomiques sont alors mise à jour.
L'effet du mouvement du piston est décrit par la variation du volume V selon l'équation de Langevin :

$$\ddot{V} = \frac{P(t) - P_{ref}}{M_p} - \Gamma\dot{V} + R_p(t) \tag{2.24}$$

$\ddot{V}$ est assimilable à l'accélération du volume engendré par le piston de masse M_p. $P(t)$ et P_{ref} sont respectivement la pression instantanée et la pression de référence. Γ représente la fréquence de collision et $R_p(t)$ une force stochastique.

2.1.7 Protocoles de simulation

Le protocole typique d'une simulation de DM suit les étapes suivantes :

- minimisation en énergie
- mise en température du système
- équilibration du système à la température référente
- production de la trajectoire
- analyse de la trajectoire

2.1.7.1 Minimisation en énergie

Avant toute simulation de DM, il est nécessaire d'effectuer une optimisation en énergie afin de supprimer toutes conformations non physiques qui généreraient des trajectoires aberrantes. La recherche d'une conformation stable consiste à faire une minimisation de l'énergie potentielle du sytème. Pour cela nous avons utilisé la méthode du gradient conjugué qui est implémentée dans NAMD. Cette méthode est basée sur la dérivée de l'énergie potentielle V par rapport aux coordonnées de chaque atome i :

$$F_i = -\nabla V \tag{2.25}$$

Chaque atome est alors déplacé vers son minimum local en suivant la pente de l'hypersurface énergétique. Cette méthode a le défaut de ne pas permettre de franchir des barrières énergétiques et donc de ne pas faire converger le système vers son minimum global. Elle garantit de trouver le minimum local le plus proche mais le résultat final dépend fortement du choix de la conformation initiale.

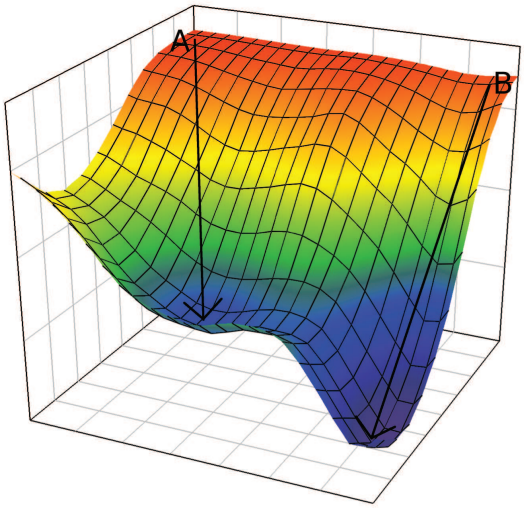

Figure 2.9 – Défaut de la méthode des gradients : le choix de la configuration initiale influe directement sur la position du minimum

Le principe de la méthode consiste à minimiser la fonction énergie potentielle selon une première direction $\overrightarrow{d}_k$. À chaque itération k, la position du système est décrit par un vecteur $\overrightarrow{R}_{k+1}$, de $3N$ dimensions. $\overrightarrow{R}_{k+1}$ est ainsi le minimum le long de cette direction. Les coordonnées $\overrightarrow{R}_{k+1}$ à l'itération $k+1$ s'écrivent alors de la façon suivante :

$$\overrightarrow{R}_{k+1} = \overrightarrow{R}_k + s_k \overrightarrow{d}_k \tag{2.26}$$

s_k est la distance de la descente. L'algorithme s'arrête lorsque les $\overrightarrow{R}_k$ n'évoluent plus ou bien lorsque le gradient atteint un seuil minimal.

2.1.7.2 Production

Les simulations de DM sont menées en deux étapes : une première qui consiste à amener le système vers sa température d'étude et la seconde qui correspond à la phase de production de la trajectoire de simulation.

Mise en température

La mise en température ou thermalisation du système consiste à réaliser une dynamique sur les vitesses des atomes de telle sorte à obtenir une température moyenne proche de la valeur désirée. Dans notre cas, elle a été effectuée dans l'ensemble microcanonique NVT (le nombre de particules N, le volume V et la température T restent constantes au cours du temps). Pour les simulations à haute température, le volume de la boîte de simulation est ajusté afin de reproduire au mieux la densité de l'eau liquide à la température d'étude [76].

Dynamique moléculaire à température et pression constantes

Les simulations de DM à $300K$ ont été réalisées dans l'ensemble thermodynamique isotherme-isobare NPT (le nombre de particules N, la pression P et la température T sont constantes au cours de la dynamique). La méthode du piston et la dynamique de Langevin contrôlent respectivement les deux dernières variables. La température du système a été fixée à 300K et la pression à 1 bar. Les simulations de chauffage et d'équilibration sont effectuées séparement pour chaque trajectoire.

Dans ce travail, nous utilisons les champs de forces AMBER (Assisted Model Building with Energy Refinement) 03 [64] et OPLS-AA (Optimized Potentials for Liquid Simulations - All Atom)[63]. Ce sont des champs dits tout-atome dont les données permettent un traitement complet du système. Les simulations ont été effectuées à l'aide du programme de DM parallèle NAMD. Ce code est conçu pour la simulation haute performance de grands systèmes biomoléculaires. Les liaisons chimiques entre les atomes lourds et les hydrogènes sont contraintes à leur valeur d'équilibre à l'aide de l'algorithme SHAKE [77].

2.1.8 Analyse des simulations de dynamique moléculaire

2.1.8.1 RMSD et RMSF

Le **RMSD** ou écart quadratique moyen est une mesure de la différence qui existe entre 2 structures qui ont, de préférence, été alignées préalablement. On note σ_a et σ_b les 2

structures moléculaires d'intérêt composées de N atomes, le RMSD est calculé en utilisant la formule suivante :

$$RMSD(\sigma_a, \sigma_b) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\overrightarrow{r}_{i,a} - \overrightarrow{r}_{i,b})^2} \tag{2.27}$$

La structure de référence est souvent la structure cristallographique ou la première structure de la trajectoire de simulation. Le **RMSF** ou fluctuation quadratique moyenne mesure la mobilité de chaque atome i au long de la trajectoire de simulation. La formule du calcul du RMSF est la suivante :

$$RMSF(i) = \sqrt{\frac{1}{C}\sum_{k=1}^{C}(\overrightarrow{r}_{i,k} - < \overrightarrow{r}_i >_k)^2} \tag{2.28}$$

Différemment du RMSD, la structure de référence est la structure moyennée sur l'ensemble des C conformations. La principale différence entre le RMSD et le RMSF vient du fait que pour ce dernier la moyenne est effectuée sur le temps différement du RMSD où la moyenne est effectuée sur les atomes.

2.1.8.2 Rayon de gyration

La distribution de masse d'un grand système moléculaire peut être évaluée en calculant son rayon de gyration R_{gyr}. Celui-ci correspond au rayon d'une sphère ayant une distribution de masse uniforme égale à la masse totale de la protéine. R_{gyr} se calcule de la façon suivante :

$$R_{gyr} = \sqrt{\frac{1}{N}\frac{\sum_{i=1}^{N} m_i(\overrightarrow{r}_i - \overrightarrow{r}_{com})^2}{\sum_{i=1} m_i}} \tag{2.29}$$

avec $\overrightarrow{r}_i$ qui correspond à la position du i-ième atome et $\overrightarrow{r}_{com}$ qui est la position du centre de masse de la protéine. De façon analogue au RMSD, le rayon de gyration fournit une information globale sur la déformation structurale de la protéine au cours de la simulation.

2.1.8.3 Hélicité

Les changements dans la structure secondaire sont caractérisés en évaluant la fraction d'hélices α de la protéine. Les calculs reposent sur l'algorithme STRIDE [78] qui est implémenté dans le programme de visualisation VMD[79]

2.1.8.4 Contacts natifs

Les modifications dans la structure tertiaire peuvent être évaluées en suivant l'évolution des contacts natifs au long de la trajectoire de simulation. Par rapport aux contacts formés au sein de la structure secondaire, les contacts tertiaires entre acides aminés sont moins contraints par la rigidité du squelette peptidique. Dans nos analyses, deux résidus sont considérés en contact tertiaire si la distance entre leurs C_α est inférieure à la distance cut-off de 7 Ået si ces deux résidus sont séparés par plus de deux acides aminés dans la séquence primaire, comme le suggère Sardiu et al.[80].

2.1.8.5 Calculs des profils d'énergie libre

Une majorité des propriétés physico-chimiques d'un système peuvent être analysées plus ou moins directement en étudiant les variations de son énergie libre (ou de son enthalpie libre). En particulier, les préférences conformationnelles d'une protéine sont étroitement liées à la différence d'énergie libre qu'il existe entre deux de ces états. Cette variable d'énergie libre est licite dans notre cas car elle prend en compte la cohésion de la macromolécule via son énergie interne et son entropie conformationnelle à la température T.

Définition de l'énergie libre

L'énergie libre de Helmholtz A est une fonction d'état extensive s'appliquant aux systèmes thermodynamiques fermés évoluant à température et volume constant. Elle est définie par :

$$A = -\beta^{-1} ln Q(N, V, T) \tag{2.30}$$

Avec $\beta^{-1} = k_B T$ où k_B est la constante de Boltzmann et T la température du système. $Q(N, V, T)$ est la fonction de partition du système pour un nombre de particules N, un volume V et une température T fixe. La fonction de partition est une grandeur fondamentale qui décrit les propriétés statistiques du système étudié. Elle fait le lien entre les états microscopiques du système et ses observables. L'équation 2.30 en est la preuve directe. C'est la relation fondamentale existante entre la thermodynamique et la physique statistique

dans l'ensemble canonique. Dans cet ensemble, la fonction de partition s'exprime de la façon suivante :

$$Q(N,V,T) = \frac{1}{h^{3N}N!}\int exp[-\beta H(q,p)]dqdp \tag{2.31}$$

$1/h^{3N}$ est un facteur de proportionnalité permettant de retrouver les prédictions de la mécanique quantique aux hautes températures. H est l'hamiltonien total du système qui est fonction des $3N$ coordonnées q et des $3N$ moments conjugués p du système. Il est défini de la façon suivante :

$$H(q,p) = K(p) + U(q) = \frac{p^2}{2m} + U(q) \tag{2.32}$$

En remplaçant cette expression dans l'équation 2.31 et en intégrant sur les vitesses, on obtient directement l'expression suivante :

$$Q(N,V,T) = \frac{1}{\Lambda^{3N}N!}Z(N,V,T) \tag{2.33}$$

Λ est la longueur d'onde thermique de de Broglie : $\Lambda = h\sqrt{\frac{\beta}{2\pi m}}$ et $Z(N,V,T)$ est la fonction de partition configurationnelle :

$$Z(N,V,T) = \int exp(-\beta U(q))dq \tag{2.34}$$

U(q) est l'énergie potentielle du sytème. Evaluer Q ou Z afin de déterminer l'énergie libre absolue est très compliquée. En effet, A est directement liée à la probabilité de trouver le système dans un état donné et dans un certain volume de l'espace des phases.

L'étude de l'évolution d'un système lors d'un processus physico-chimique exige la définition d'une énergie libre partielle qui est une fonction de la coordonnée de réaction ζ_c :

$$F(\zeta_c) = -\beta^{-1}\ln Z(\zeta_c) \tag{2.35}$$

Avec $Z(\zeta_c)$ qui a pour expression :

$$Z(\zeta_c) = \int \delta(\zeta - \zeta_c)\exp(-\beta U(q))d\zeta dq' \tag{2.36}$$

La fonction delta de Dirac $\delta(\zeta-\zeta_c)$ signifie que l'intégration se fait sur toutes les coordonnées q tel que $\zeta = \zeta(q)$. L'enthalpie libre partielle s'écrit alors :

$$G(\zeta_c) = F(\zeta_c) + RT \tag{2.37}$$

Il est possible de calculer l'enthalpie libre partielle à partir d'un ensemble statistique de configurations générées lors d'une simulation de dynamique moléculaire. Les fréquences d'apparition d'une valeur donnée de la coordonnée choisie sont liées à l'enthalpie libre du sous-espace respectif par la relation :

$$n(\zeta_c) = exp(-\frac{G(\zeta_c)}{RT}) \tag{2.38}$$

Soit deux valeurs ζ_1 et ζ_2 de la coordonnée ζ. On peut alors écrire :

$$\frac{n(\zeta_1)}{n(\zeta_2)} = exp\left(\frac{G(\zeta_2) - G(\zeta_1)}{RT}\right) \tag{2.39}$$

Ou bien encore :

$$\Delta G(\zeta) = G(\zeta_2) - G(\zeta_1) = RTln\frac{n(\zeta_1)}{n(\zeta_2)} \tag{2.40}$$

Il est ainsi possible de déterminer la différence d' enthalpie libre par rapport à une valeur de référence $G(\zeta_0)$.

La méthode ***Umbrella Sampling***,qui est basée sur cette théorie, a été proposée en 1977 par Torrie et Valleau [81] et permet d'augmenter l'efficacité de l'échantillonnage. Elle est particulièrement efficace pour échantillonner les systèmes dans lesquelles deux régions de l'espace des configurations sont séparées par une barrière énergétique. En effet, cette méthode permet de forçer le système à évoluer autour d'une certaine valeur ζ_c de la coordonnée de réaction ζ en lui imposant une contrainte harmonique :

$$U(\zeta) \sim -\frac{1}{2}K(\zeta - \zeta_c)^2 \tag{2.41}$$

Ainsi, les régions de l'espace proche de la barrière, là où la probabilité de présence est très faible, seront échantillonnées de la même manière que le reste de l'espace des configurations. Plusieurs fenêtres d'échantillonnage sont générées en déplaçant ζ_c et en modifiant la valeur de la contrainte K. Les effets des contraintes sont corrigés dans le traitement final des résultats. Les différentes séries d'échantillonnages ont été analysées à l'aide du logiciel WHAM [82, 83].

2.1.9 La visualisation

VMD est un logiciel libre conçu pour la modélisation et la visualisation moléculaire. Développé en 1992 par l'université de l'Illinois et l'institut Beckman, il permet principalement de visualiser et d'analyser les résultats des simulations de dynamique moléculaire. L'application de scripts écrits en langage Tcl ou Python fournit un large éventail de possibilités de calculs.

2.2 Analyse par composantes principales

L'analyse en composantes principales (PCA), développée en France depuis les années 60 par J-P. Benzécri est une méthode d'analyse multivariée qui consiste à transformer des variables liées entre elles en nouvelles variables décorrélées les unes des autres. Elle permet principalement de réduire le nombre de variables et de rendre l'information moins redondante. La PCA présente des avantages significatifs par rapport à une analyse en modes normaux basée sur la diagonalisation de la matrice Hessienne du système : elle n'est pas limitée aux systèmes harmoniques (ou aux mouvements), mais prend en compte la contribution de l'entropie de la dynamique de la protéine et les modes sont classés en fonction de leurs amplitudes.

2.2.1 Principe

Les composantes principales (PC) sont les nouvelles coordonnées de la protéine obtenues par transformation orthogonale de l'ensemble des coordonnées cartésiennes x_i ($i = 1, ..., n$ où n est le nombre de degrés de liberté internes de la protéine)[38]. La

transformation orthogonale est choisie de telle sorte que, exprimée dans les nouvelles coordonnées p_i, la matrice de covariance soit diagonale. Cela signifie que les nouvelles coordonnées ne sont pas linéairement dépendantes (ou linéairement corrélées) mais des corrélations d'ordre supérieur peuvent être présentes. On définit C comme la matrice de variance exprimée en utilisant les coordonnées cartésiennes de la protéine. On peut écrire :

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{2.42}$$

où la notation $\langle\rangle$ signifie qu'on effectue la moyenne sur la totalité de la trajectoire de simulation. Trouver les nouvelles coordonnées signifie trouver une matrice orthogonale R qui satisfasse à l'équation :

$$R^T C R = I_n \lambda \tag{2.43}$$

Ici, I_n est la matrice identité de dimension n et λ un vecteur qui contient les n valeurs propres λ_i. Le vecteur r_i forme la colonne i de la matrice R. On peut alors exprimer les nouvelles coordonnées de la façon suivante :

$$p_i = \sum_{j=1}^{n} r_{ji} (x_j - \langle x_j \rangle) \tag{2.44}$$

λ_i est un élément diagonal dans la matrice de covariance transformée, il est donc égal à la moyenne du carré de la fluctuation de la composante principale p_i :

$$\lambda_i = \langle (p_i - \langle p_i \rangle)^2 \rangle \tag{2.45}$$

D'autre part, compte tenu du fait que R est une matrice unitaire, la racine de la moyenne fluctuations carrées calculée sur les composantes principales ou sur les coordonnées cartésiennes est identique. On écrit alors :

$$RMSF^2 = \sum_{i=1}^{n} \langle (x_i - \langle x_i \rangle)^2 \rangle = \sum_{i=1}^{n} \langle (p_i - \langle p_i \rangle)^2 \rangle = \sum_{i=1}^{n} \lambda_i \tag{2.46}$$

Par conséquent, la contribution de la composante principale p_i sur les fluctuations des positions de la protéine est égal à sa valeur propre λ_i. Dans ce travail de thèse, les PCA ont été effectuées à l'aide du programme libre CARMA [84] dans sa version 1.1 développée au sein du groupe de Nicholas M. Glykos à l'université Démocrite de Thrace. CARMA calcule les vecteurs propres et les valeurs propres de la matrice de covariance ainsi que les fluctuations des PC au long de la trajectoire de simulation. Il est également possible d'identifier des conformères protéiques en projetant toutes les structures de la protéine dans les sous-espace formé par les trois premières composantes principales. Pour chaque image

de la trajectoire, les valeurs des fluctuations des trois plus grandes composantes principales sont calculées. Ces valeurs sont utilisées pour remplir une carte en trois dimensions décrivant leur distribution. CARMA identifie alors tous les «pics» isolés de cette carte. Ces regroupements correspondent à des conformations fortement peuplées.

2.2.2 Convergence de la méthode

La pertinence d'une PCA est directement liée à sa convergence. Au sens strict, une PCA est convergente lorsque ses composantes principales sont stables par rapport à la longueur de la trajectoire ou bien en répétant les simulations. Habituellement, cette stabilité est vérifiée pour un sous-espace engendré par les premiers modes. Le recouvrement entre les deux sous-espaces correspondants provenant de deux trajectoires différentes (ou des fragments de trajectoire) est mesuré en calculant leur produit intérieur :

$$RMSIP = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{m} u_i v_j \tag{2.47}$$

Où les vecteurs u_i appartiennent au premier sous-espace, les vecteurs v_j au second et m est la dimension commune des deux sous-espaces. Il est évident que, lorsque les deux sous-espaces sont identiques, le produit scalaire donné par l'équation 2.47 est égal un. Cependant, ce type de convergence est très compliqué à obtenir dans le cas d'un mouvement protéique. En effet, en allongeant la longueur de la trajectoire, la protéine peut changer subitement de conformation ce qui induit un changement de ses composantes principales [65]. Etant donné que les échelles de temps de ses sauts peuvent varier de la picoseconde à la seconde, de très (trop) longues trajectoires sont nécessaires pour obtenir la convergence complète dans le cas de protéines de taille importante [41]. Une autre façon de penser est de cibler la convergence de l'échantillonnage seulement dans un sous espace conformationnel de la protéine. Ce sous espace est en principe associé à une conformation macroscopique donnée de la protéine [65]. Cette convergence est une condition essentielle pour que l'analyse ait un minimum de pertinence. En effet lorsque les modes les plus lents ne sont pas suffisament échantillonnés, les PC correspondantes ne sont pas liées à la dynamique structurale de la protéine mais à une diffusion aléatoire [38]. La variation d'une telle coordonnée le long de la trajectoire est de forme cosinusoïdale avec un nombre de périodes égal à la moitié de l'indice du mode. Le contenu en cosinus d'une composante principale p_k peut être évalué en utilisant la formule suivante [38] :

$$c_k = \frac{2}{T} \left(\int_0^T \cos\left(\frac{k\pi t}{T}\right) p_k(t) dt \right)^2 \left(\int_0^T p_k^2(t) dt \right)^{-1} \tag{2.48}$$

où T est la longueur de la trajectoire. Il varie entre 0 (aucune forme cosinusoïdale) et 1 (cosinus parfait). En fait, pour le mouvement de protéines, un contenu en cosinus inférieur à 0.5 est considéré comme indicatif d'un échantillonnage satisfaisant [41]. De toute évidence une PCA convergente au sens restreint peut ne pas être convergente au sens strict : aux longues durées de simulations, les PC sont susceptibles de changer si de nouveaux et plus lents mouvements apparaissent. Cependant, les PC émises à partir d'une PCA convergente au sens restreint ont un sens physique et expriment les caractéristiques réelles du mouvement protéique se produisant à l'échelle de temps de la trajectoire de simulation. En d'autres termes, ces analyses fournissent une description correcte mais incomplète de la dynamique des protéines.

Chapitre 3

Dynamique conformationnelle des sous-domaines et des sites de complexation de HSA$^{(N)}$

Sommaire

3.1 Introduction 45

3.2 Convergence de la méthode d'analyse 45

3.2.1 Convergence des PCA 45

3.3 Dynamique globale de la protéine 47

3.3.1 Mouvements collectifs associés à une PC 47

3.3.2 Dynamique des sous-domaines 50

3.3.3 Recherche de conformations stables (conformères) 51

3.3.4 Étude des changements conformationnels naturels 52

3.3.5 Étude d'éventuels effets allostériques 54

3.4 Dynamique des sites primaires de liaison 55

3.4.1 Identification des conformères 56

3.4.2 Étude des conformères du site 1 57

3.4.2.1 Changements conformationnels entre les différents conformères 57

3.4.2.2 Une parallèle entre la transition théorique CS6→CS7 et celle observée expérimentalement lors de la complexation de la thyroxine 58
3.4.3 Étude des conformères du site 2 60
3.4.3.1 Changements conformationnels entre les différents conformères 60
3.4.3.2 Une parallèle entre la transition théorique CS8→CS9 et celle observée expérimentalement lors de la complexation du diazépam 61
3.5 Corrélation des mouvements entre les sites 1 et 2 62
3.6 Conclusions 64

3.1 Introduction

Dans ce chapitre, nous proposons une analyse détaillée de la dynamique conformationnelle de la HSA dans sa forme native basée sur la méthode PCA appliquée à trois trajectoires de DM de 200 ns chacune. Une attention spéciale est accordée à la dynamique des deux sites de complexation primaires. Nous nous interrogeons aussi sur le rapport qui existe entre les changements conformationnels spontanés et ceux induits par la complexation des ligands.

3.2 Convergence de la méthode d'analyse

L'évolution de la racine carrée de l'écart-type (RMSD) des positions atomiques de la protéine au long des trois trajectoires de simulation, nommées par la suite traj1, traj2 et traj3, est illustrée sur la Figure 3.1. Les RMSD ont été calculés par rapport aux atomes de carbone α en utilisant la structure de diffraction aux RX comme référence [12]. Dans le cas de traj1 et traj3, le RMSD apparait stabilisé à la fin de la simulation alors qu'il est encore en légère hausse dans la cas de traj2. D'autre part, l'hélicité de la protéine (non représenté, calculée à l'aide de l'algorithme STRIDE [78]) varie peu selon les trois trajectoires (sa valeur moyenne est de 67%, comparable à la valeur de 72% pour la structure de référence).

Ainsi, pour traj1 et traj3, la structure de la protéine semble être globalement stabilisée alors que traj2 montre une relaxation structurale plus importante qui continue encore au-delà des 200ns.

3.2.1 Convergence des PCA

Une analyse plus détaillée des mouvements protéiques au long des trois trajectoires est obtenue en calculant leurs composantes principales. Afin de réduire la dimension de la matrice de covariance à un niveau raisonnable, le calcul a été effectué en prenant en compte uniquement les Cα. Ainsi, la dimension de l'espace configurationnel de la protéine est de 1734. La partie montante rapide de la trajectoire (0-20 ns) n'a pas été inclue dans l'analyse PCA afin d'éviter les instabilités liées à l'équilibration de la protéine. Les vecteurs

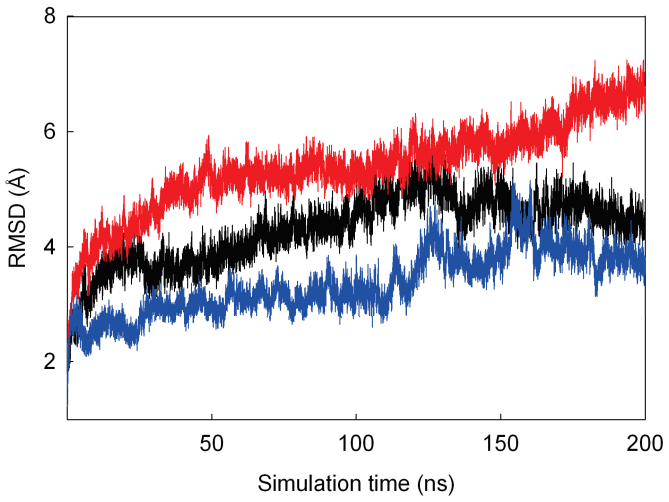

Figure 3.1 – Evolution des RMSD au long des trois trajectoires de simulation : traj1 (noir), traj2 (rouge) traj3 (bleu).

propres calculés sont disposés par ordre décroissant de leurs valeurs propres. Comme prévu, la contribution des différents modes associés aux fluctuations de la protéine diminue très rapidement avec leurs indices. Le mouvement protéique est ainsi concentré uniquement sur quelques modes. Dans les faits, les 10 premières composantes, qui réprésentent seulement 0,6% des degrés de liberté, permettent de capturer environ 81% des mouvements.

La convergence au sens strict (telle que définie dans le chapitre 2) de la PCA a été testée en calculant le recouvrement (équation 2.47) des sous-espaces engendré par les 10 premières PC de chaque trajectoire. Les valeurs obtenues varient entre 0.37 (traj1 contre traj2) et 0.25 (traj2 contre traj3). En d'autres termes, le mouvement protéique le long des trois trajectoires a un degré de similitude d'environ 30% seulement. On en conclut que, dans le cas de la HSA, une PCA basée sur une trajectoire de DM de 200 ns n'est pas convergente au sens strict. Ce résultat est en bon accord avec celui rapporté par Grossfield et al. [85] qui ont comparé 26 trajectoires de 100ns générées pour la rhodopsine. Le recouvrement moyen de covariance entre deux trajectoires était alors de 0.2. Bien que l'échantillonnage de l'espace conformationnel fourni par une simulation de 200ns n'est que partiel, les PC calculées peuvent décrire correctement les caractéristiques réelles de la dynamique de la protéine. C'est le cas lorsque tous les mouvements collectifs qui se produisent au long d'une trajectoire donnée sont correctement échantillonnés. Nous avons testé cette convergence locale en calculant le contenu en cosinus (équation 2.48) des 5 premières PC dérivées de chaque trajectoire. Les résultats sont présentés dans le Tableau

3.1.

	traj1	traj2	traj3
PC1	0.19	0.85	0.90
PC2	0.00	0.53	0.46
PC3	0.15	0.03	0.25
PC4	0.42	0.00	0.06
PC5	0.00	0.01	0.00

Tableau 3.1 – Contenu en cosinus des 5 premières PC des 3 trajectoires simulées

D'après ces données, l'analyse de covariance appliquée à traj1 est convergente au sens restreint. Ce résultat suggère fortement que les cinqs premières PC issues de cette analyse constituent de véritables mouvements collectifs protéiques. D'autre part, les mouvements les plus lents le long de traj2 et traj3 semblent n'être que partiellement échantillonnés puisque le contenu en cosinus de PC1 est supérieur à 0.5 dans les deux cas. Ainsi, un doute subsiste quant à la signification physique du mode PC1 de ces deux trajectoires.

3.3 Dynamique globale de la protéine

3.3.1 Mouvements collectifs associés à une PC

Une description détaillée du mouvement collectif associée à une PC donnée est obtenue en représentant la variation de la racine des déplacements quadratiques (RSD) des atomes de carbone α entre les valeurs extrêmes de la PC. Ces représentations sont données dans les figures 3.2 et 3.3 respectivement pour les modes PC1 et PC2 des trois trajectoires. Il y a très peu de similitudes entre les trois modes PC1 (Figure 3.2).

La PC1 de traj1 affecte principalement les sous-domaines IA et IIIB, alors que celle obtenue pour traj2 est dispersée sur la structure entière de la protéine, sauf pour les sous-domaines IB et IIIA. Dans le cas de traj3, les principaux déplacements induits par le mode PC1 sont situés dans les sous-domaines IB, IIB et IIIB. Certaines ressemblances intéressantes sont observées entre le mode PC1 de traj1 et les modes PC2 des deux autres trajectoires : ces ressemblances concernent le déplacement des sous-domaines IA et IIIB qui sont dominants et de forme similaire dans les trois cas. Une caractéristique

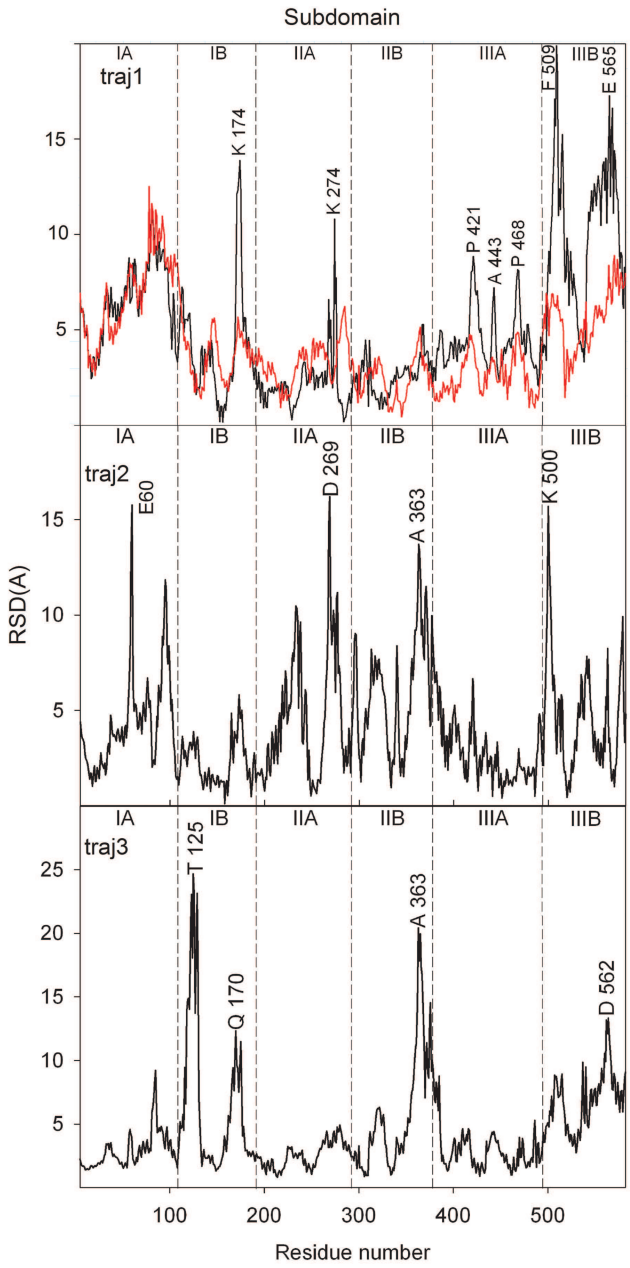

Figure 3.2 – Déplacements (RSD) des atomes de C_α au long de la première PC de chacune des trois trajectoires (noir). On trouve également les déplacements entre la HSA native et complexée avec le myristate (rouge, en haut de la figure). La structure du complexe HSA-myristate provient de : Protein Data Bank RCSB, code ID : 1E7G [1]

commune des profils RSD donnés dans les figures 3.2 et 3.3 est la présence de forts maxima décrivant le mouvement d'un petit groupe d'atomes (par exemple, les maxima centrés sur les résidus K174, K274 et F509 dans le cas du mode PC1 de traj1). Fait intéressant, tous

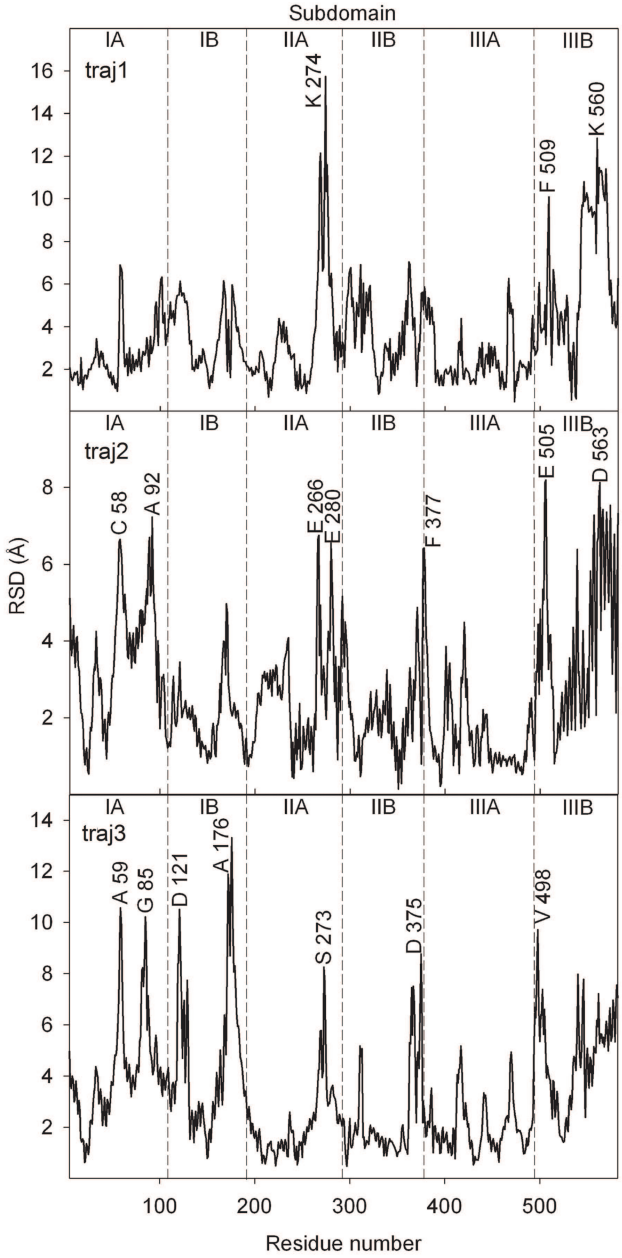

Figure 3.3 – Déplacements (RSD) des atomes de C_α le long de PC2 pour chacune des trois trajectoires.

ces déplacements de résidus bien localisés et importants (environ 10 Å de long) sont situés dans des boucles reliants deux hélices adjacentes. Des maxima similaires sont observés dans le cas des composantes principales PC3, PC4 et PC5 (Figure 3.4).

Il est très probable que l'extension de ces distorsions soit bloquée par les 17 ponts

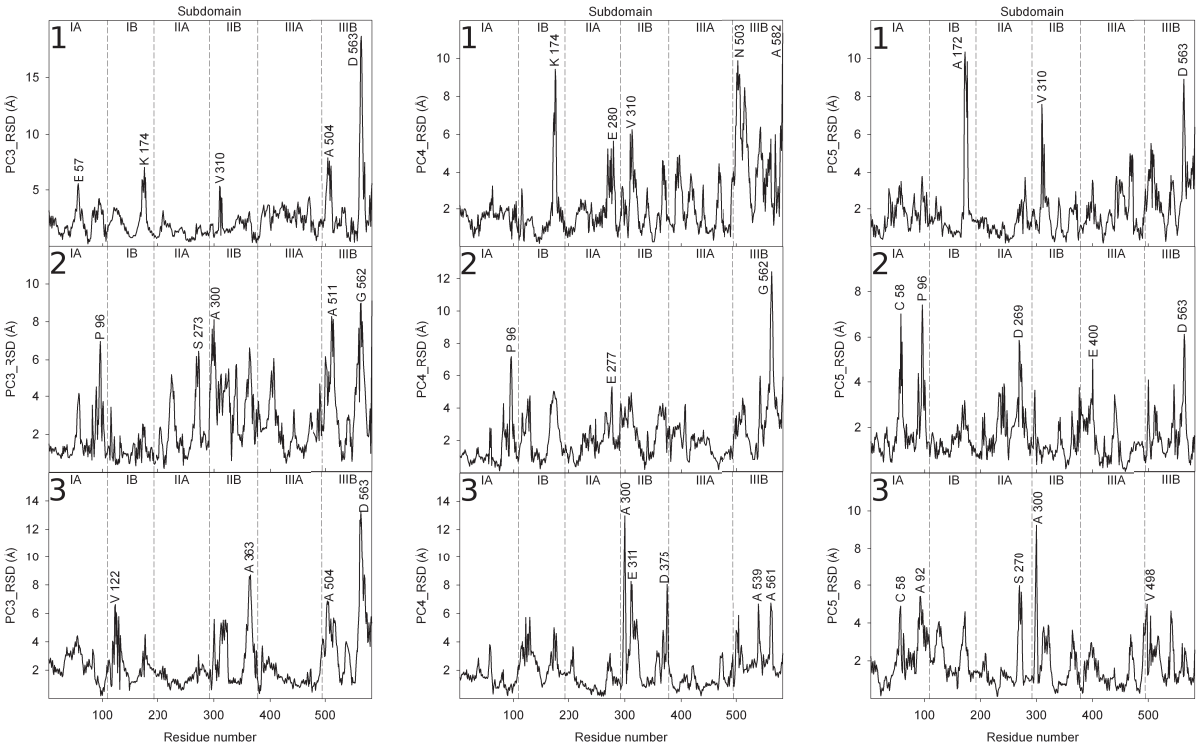

Figure 3.4 – Déplacements (RSD) des atomes de C_α au long de la troisième (à gauche), quatrième (au centre) et cinquième (à droite) PC de chacune des trois trajectoires

disulfures qui relient les extrémités des hélices. Ceci est en accord avec le fait que les ponts disulfure sont des contraintes très efficaces dans les mouvements internes de la HSA[86].

3.3.2 Dynamique des sous-domaines

Une question présentant un intérêt particulier et qui concerne la complexation d'un ligand à la HSA est celui de la déformation interne des sous-domaines induite par la dynamique de la protéine. Chaque PC peut être décrite comme une combinaison entre la distorsion et le mouvement relatif des sous-domaines. La séparation de ces deux contributions peut être obtenue en calculant le RMSD de chaque sous-domaine. Nous avons effectué cette analyse pour les 5 premières PC de traj1. Pour chaque composante, le RMSD de chaque sous-domaine a été calculé après superposition des deux structures extrêmes de la protéine : une correspondante au minimum de la PC et l'autre à la valeur maximale (Tableau 3.2).

RMSD (Å)	IA	IB	IIA	IIB	IIIA	IIIB	ALL
PC1	7.9(2.1)	4.9(1.9)	2.9(2.6)	3.8(2.2)	4.9(2.3)	10.9(4.1)	6.5
PC2	3.7(2.3)	3.8(1.8)	6.1(3.0)	5.2(2.1)	3.5(2.2)	3.5(2.2)	5.3
PC3	3.1(1.7)	4.5(1.7)	3.0(2.4)	2.6(1.7)	2.8(2.0)	8.1(3.4)	4.3
PC4	3.9(2.5)	3.5(2.3)	3.1(2.4)	3.4(2.6)	2.9(2.1)	4.9(2.2)	3.9
PC5	4.1(1.5)	4.4(2.0)	2.6(1.3)	2.4(1.3)	2.3(1.7)	5.2(2.4)	3.7

Tableau 3.2 – Distorsion des sous-domaines de la HSA induite par les différentes PC : RMSD (en Å) entre les deux structures extrêmes de la PC

Chaque conformation est une structure moyennée de 10 structures ayant des valeurs de PC similaires. La contribution du mouvement relatif des sous-domaines au RMSD peut être éliminée en alignant les deux conformations par rapport à un sous-domaine à la fois, et non par rapport à la structure globale de la protéine. Les valeurs du RMSD ainsi obtenues sont données dans le Tableau 3.2, entre parenthèses. En fait, ils représentent la contribution de la distorsion des sous-domaines. Les données du Tableau 3.2 montrent que, dans la plupart des cas, les deux contributions sont équivalentes. Les exceptions notables sont observées pour les sous-domaines IA et IIIB pour la PC1, le sous-domaine IIB pour la PC2 et le sous-domaine IIIB pour la PC4 où la contribution du mouvement relatif est largement répandue. Le sous-domaine IIIB semble être le plus mobile et le moins rigide de la protéine, ce qui est en bon accord avec l'analyse faite sur une PCA basée sur une trajectoire de DM de 10ns [24].

3.3.3 Recherche de conformations stables (conformères)

Le mouvement de la protéine à l'intérieur de l'espace des conformations natives est souvent décrit comme une série de transitions entre les différentes conformations protéiques proches dans l'espace des configurations et d'énergie libre équivalente [87]. Nous avons cherché ces conformations en analysant la répartition de celles-ci dans le sous-espace des trois premières composantes. Quatre groupes de conformations de la protéine (CS1 à CS4) ont ainsi été identifiés pour traj1 (Figure 3.5), trois pour traj2 et trois autres pour traj3. Ces groupes peuvent éventuellement être considérés comme des conformations natives de la HSA. Les dix conformères sont tous différents et leurs RMSD varient entre 2.6 et 4.5 Å. Les valeurs du RMSD ont été calculées en utilisant les structures moyennes fournies par CARMA [84]. Enfin, les dix structures sont assez proches les unes des autres comme

convenu puisque les conformères appartiennent au même état macroscopique. Dans un article très récent, cinq conformères de la HSA ont été identifiés en se basant sur une PCA d'une trajectoire de DM de 250ns [26]. Les structures explicites de ces conformations n'ayant pas été données, il n'est pas possible d'effectuer une comparaison quantitative avec celles reportées ici.

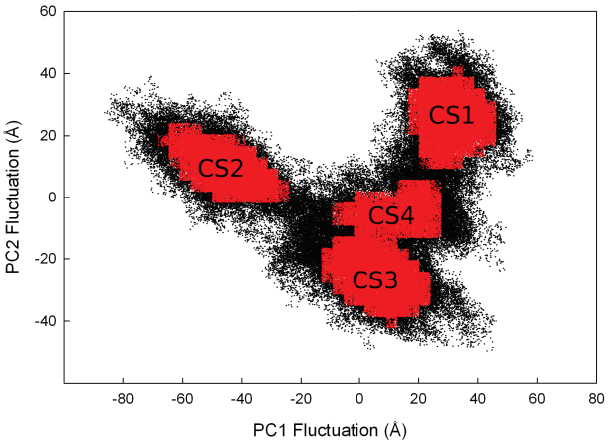

Figure 3.5 – Les 4 clusters de conformations (en rouge) obtenus en utilisant la distribution des 3 premières PCs de la traj1

	CS1	CS2	CS3	CS4
Population (%)	16.4	20.6	15.6	7.4

Tableau 3.3 – Population (%) des 4 clusters de conformations identifiés en utilisant la distribution des 3 premières PCs de la traj1

Bien que nous ayons identifié des transitions réversibles entre certains des dix conformères obtenus ici, la longueur des trajectoires de simulation n'est probablement pas assez longue pour fournir des conformères statistiquement équilibrés. Leurs importances relatives doivent donc être considérées avec prudence.

3.3.4 Étude des changements conformationnels naturels

Les modes principaux sont censés être liés non seulement aux changements de conformations de la protéine qui se produisent librement mais aussi à ceux induits par les

interactions protéine-ligand. Cet aspect présente un intérêt particulier dans le cas de la HSA qui est connue pour ses propriétés liantes. La relation entre les modes principaux de la protéine et les transitions induites par un ligand a déjà été révélée pour l'adénylate kinase d'E-coli. Basée sur une PCA, une transition ouverte-fermée a été identifiée pour la protéine-apo qui était semblable à la transition conformationnelle induite par la complexation de l'AMP [39]. Parmi les changements conformationnels induits par la complexation d'un ligand à la HSA, le plus spectaculaire est celui du à la complexation du myristate. Afin de comparer ce changement de conformation induit par la complexation du médicament avec les mouvements décrits par les PC des figures 3.2, 3.3 et 3.4, nous avons représenté le RSD des C_α entre les HSA native et complexée avec le myristate (Figure 3.2 en haut, en rouge). Des similitudes très intéressantes sont observées entre ce profil RSD et celui du mode PC1 de traj1. Le grand mouvement collectif du domaine I de la protéine a une forme quasi identique, la seule différence notable étant l'amplitude du mouvement du résidu K174 (localisé dans une boucle entre deux hélices) qui est environ trois fois plus importante dans le cas de la transition simulée. Le domaine II apparaît beaucoup moins mobile dans les deux cas. En ce qui concerne le sous-domaine IIIB, le déplacement des C_α présente des modifications structurales similaires, mais très différentes en amplitude. De plus, afin d'avoir une description plus détaillée, le mouvement associé à chaque PC peut être directement imagé en ordonnant les structures au long de la trajectoire. Les structures sont alors classées selon les valeurs croissantes de la PC et nous pouvons représenter la séquence des structures résultantes (Figure 3.6a).

Dans une interprétation rapide, PC1 décrit un déplacement relatif des sous-domaines de la protéine : IA et IIIB s'éloignent l'un par rapport à l'autre par rotation autour de l'axe long de la protéine. À titre de comparaison, une image correspondante au changement conformationnel induit par la complexation du myristate a été obtenue par le biais d'une interpolation linéaire des deux conformations protéiques expérimentales (Figure 3.6b).

Les deux mouvements possèdent des caractéristiques communes importantes avec une ressemblance étroite dans le domaine I de la protéine. Curry et al. [88] ont avancé l'hypothèse que le changement de conformation dans le complexe HSA-myristate est dirigé par la complexation du ligand mais les résultats, présentés ici, suggèrent que cette transition structurale peut précéder la complexation du ligand. Cette idée a déjà été proposée par Walji et al. [89] et réanalysée par Artali et al. [8] en se basant sur une simulation de DM de 2ns.

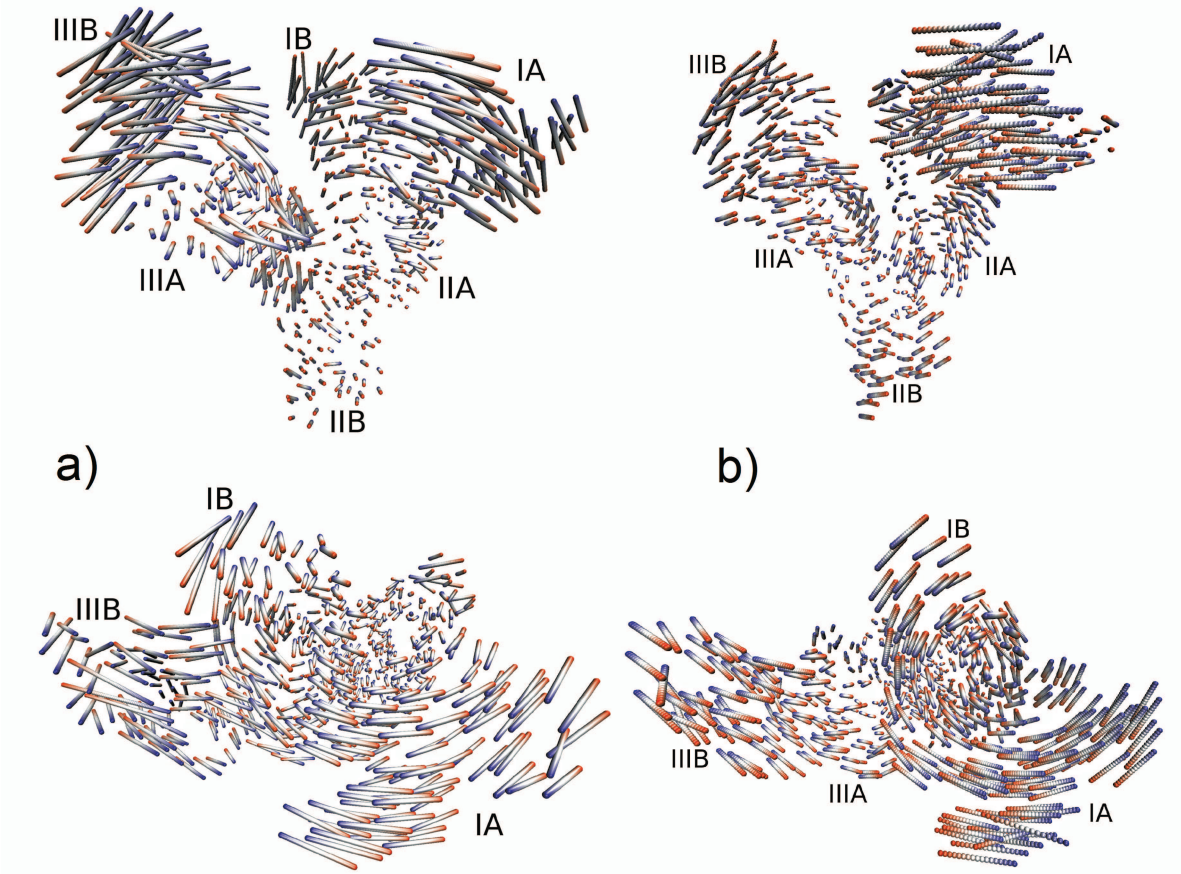

Figure 3.6 – Mouvement des atomes C_α : a) entre les deux valeurs extrêmes de PC1 dérivé de traj1 (bleu pour le minimum, rouge pour le maximum) et b) au cours de la transition entre les conformations native et complexée avec le myristate (rouge pour la native, bleu pour le complexe HSA-myristate).

3.3.5 Étude d'éventuels effets allostériques

Les profils RSD des modes de chaque PC donnent une indication sur le couplage du mouvement des atomes à l'intérieur de la structure protéique. Aucun de ces profils n'est entièrement situé dans un seul domaine, mais plutôt réparti sur la structure entière de la protéine (Figures 3.2, 3.3, 3.4). Ils apparaissent comme des modes stationnaires excités dans une structure compacte. D'un point de vue physique, la corrélation à longue distance observée dans le mouvement interne de la HSA pourrait expliquer le couplage allostérique entre les différents sites de complexation comme celà a déjà été signalé [13, 90]. Parmi les effets allostériques les plus spectaculaires on note la diminution d'un ordre de grandeur de la constante de liaison de l'hème-Fe (III) (site de liaison FA1) lors de la

saturation du site 1 [90]. D'un point de vue mécanique, le couplage allostérique nécessite la présence de transitions conformationnelles délocalisées sur (au moins) deux sites de liaison (sous-domaines), d'où l'apparition de PC délocalisées comme présentées ici.

En résumé, les principaux changements structuraux décrits par les cinq premières PC des trois trajectoires de simulation sont les suivants :

- des déplacements très localisés et intenses (10-15 Å) qui affectent les résidus situés dans les boucles entre deux hélices adjacentes (Figures 3.2, 3.3, 3.4).
- un répartitation des modes des PC à travers la structure protéique qui pourrait être lié au couplage allostérique entre les sites de complexation de la HSA.
- des déplacements relatifs des sous-domaines qui contribuent au RMSD à hauteur de 2-3 Å par sous-domaine et par PC (Tableau 3.2).
- des distorsions internes des sous-domaines dont la contribution moyenne au RMSD (par sous-domaine et par PC) se situe autour de 2.0 Å(Tableau 3.2)

Récemment, Guizado et al. [25] ont rapporté une PCA détaillée du complexe HSA-hème basée sur une trajectoire de DM de 100ns. Bien que les quatres premières PC issues de leur analyse sont distinctes de celles présentées dans ce travail, il existe des similitudes importantes entre les deux descriptions (par exemple la rotation des domaines I et III autour de l'axe long de la HSA observé dans notre cas pour le mode PC1 de traj1). Une fois de plus, les différences entre les deux analyses sont très probablement dues au fait qu'elles ne soient pas convergentes dans le sens strict, mais chacune d'elles n'échantillonnent qu'un sous-espace spécifique de l'espace des configurations natives de la protéine.

3.4 Dynamique des sites primaires de liaison

Dans cette partie, une analyse détaillée de la dynamique des sites primaires de complexation 1 et 2 est effectuée. Compte tenu du rôle des chaînes latérales des résidus dans la configuration des sites de liaison, une étude de la dynamique des sites de complexation basée uniquement sur les C_α ne serait que partiellement pertinente. Par conséquent, dans le cas des sites de complexation, nous avons effectué une analyse de covariance sur tous

les atomes. L'analyse a été basée sur traj1 puisque c'est la trajectoire qui présente la meilleure convergence locale (comme en atteste son contenu en cosinus dans le Tableau 3.1). Deux trajectoires de simulation limitées aux sites 1 ou 2 ont été obtenus en projetant la trajectoire globale sur le sous-domaine IIA (pour le site 1) ou sur le sous-domaine IIIA (pour le site 2). Deux PCA distinctes ont été effectuées pour les deux trajectoires restreintes.

3.4.1 Identification des conformères

Comme dans le cas de l'étude du mouvement global de la protéine, nous avons cherché d'éventuelles conformations stables en effectuant une analyse de cluster basée sur les 3 premières PC. Trois groupes de conformations possibles ont été trouvés pour le site 1 (Figure 3.7, CS5 à CS7) et quatre autres pour le site 2 (Figure 3.7, CS8 à CS11). La corrélation entre les trois premières composantes apparaît en rouge.

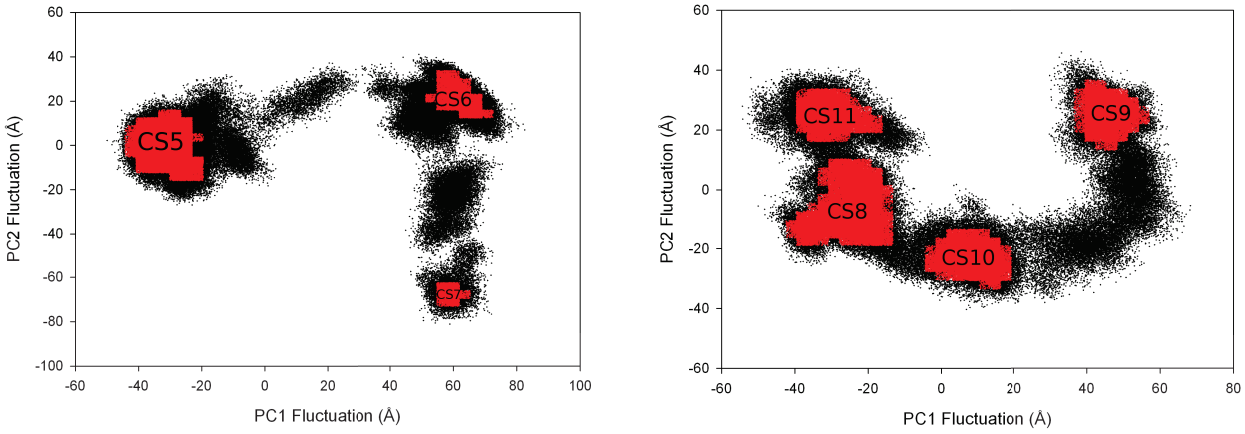

Figure 3.7 – Les différents conformères obtenus pour les sites 1 (à gauche) et 2 (à droite) en utilisant la distribution des trois premières PC des deux trajectoires restreintes.

	CS5	CS6	CS7	CS8	CS9	CS10	CS11
Population (%)	48.5	6.6	1.9	22.9	11.2	16.9	12.3

Tableau 3.4 – Population (%) des différents clusters de conformations identifiés pour les deux sites primaires

3.4.2 Étude des conformères du site 1

3.4.2.1 Changements conformationnels entre les différents conformères

Une représentation du site 1 et de ses conformères a été obtenue en représentant seulement les acides aminés qui sont les plus actifs dans la complexation du ligand (Figure 3.8). Selon cette image, les chaînes latérales les plus mobiles sont celles des résidus W214, H242, R218 et R222 dont les mouvements sont principalement réalisés par rotation des angles de torsion des chaînes latérales. Par exemple, lors du passage de CS6 à CS7, l'angle dièdre $C_\alpha - C_\beta - C_\gamma - C_\delta$ du tryptophane W214 subit une modification d'environ 50°, celui du résidu R218 d'environ 90° tandis que le résidu R222 change d'environ 86°. Il est intéressant de comparer ces mouvements des chaînes latérales avec ceux qui se produisent lors de la complexation d'un ligand. Des médicaments de taille petite et moyenne (tel que le CMPF, la warfarine, la phénylbutazone, etc...) sont connus pour affecter légèrement les conformations des chaînes latérales [6]. La situation est différente pour la liaison de la thyroxine (ou T_4), une hormone de masse moléculaire de 777 Da. Dans ce cas, des changements conformationnels ont été observés principalement pour les chaînes latérales des résidus W214, R218 et R222 [91].

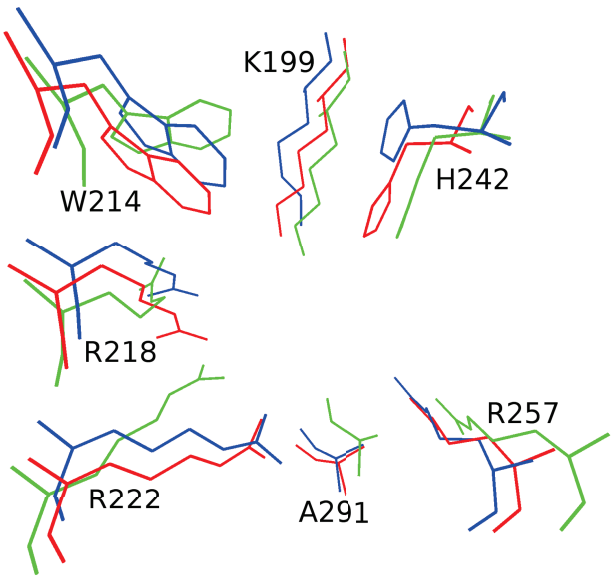

Figure 3.8 – Conformation relative des principaux résidus d'acides aminés du site 1 dans la structure moyenne de CS5 (bleu), CS6 (rouge) et CS7 (vert).

3.4.2.2 Une parallèle entre la transition théorique CS6→CS7 et celle observée expérimentalement lors de la complexation de la thyroxine

En comparant la structure cristalline de la HSA native avec celle du complexe thyroxine-HSA (Protein Data Bank RCSB, code ID :1HK1) on constate que l'angle dièdre $C_\alpha - C_\beta - C_\gamma - C_\delta$ du résidu R218 change d'environ 62° lors de la complexation de la thyroxine et l'angle dièdre $C - C_\alpha - C_\beta - C_\gamma$ de R222 est modifié de 65°. Ces changements présentent des similitudes avec ceux identifiés dans nos simulations, bien que les variations des angles dièdres ne soient pas identiques. Concernant le résidu W214, Abou-Zeid et al. [26] dans une simulation de DM de 250 ns ont rapporté une transition conformationnelle autour de la liaison $C_\beta - C_\gamma$ d'environ 180°. Cette transition est supposée être liée au caractère multi-exponentiel de la fluorescence du tryptophane dans la HSA [92]. Notre simulation prédit une conformation supplémentaire de ce chromophore intrinsèque (correspondant à une rotation de 50° autour de cette même liaison).

Une comparaison détaillée entre la transition prédite théoriquement CS6→CS7 et le changement de conformation observé expérimentalement lors de la complexation de la thyroxine est obtenue en représentant leurs RMSF respectifs calculés sur tous les atomes (Figure 3.9). Il est à noter que lorsqu'on effectue le calcul du RMSF en ne prenant

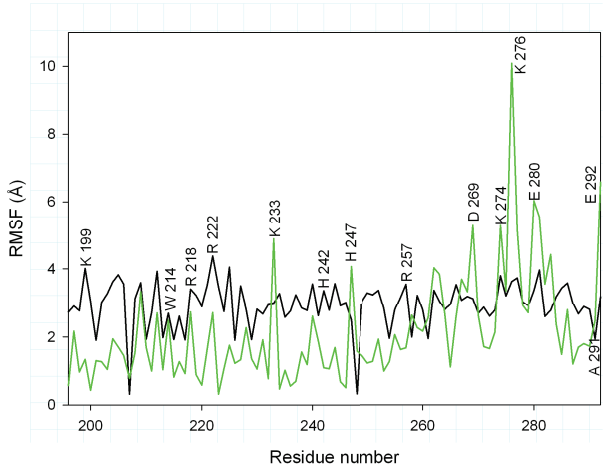

Figure 3.9 – RMSF calculé entre les conformations CS6 et CS7 (vert) et entre la HSA native et le complexe HSA-thyroxine (PDB ID : 1HK1) (noir).

en compte que les atomes de carbone α, les valeurs du RMSF pour la transition HSA native→HSA-thyroxine fluctuent autour de 0.5 Å. Cela signifie que la thyroxine se lie sans distorsion du squelette protéique. Le RMSF de la Figure 3.9 (courbe noire) est donc

principalement dû à la réorganisation des chaînes latérales. Un point commun aux deux courbes de la Figure 3.9 vient du fait que les résidus W214, R218 et R222 possèdent dans les deux cas des maxima locaux mais les deux transitions conformationnelles sont loin d'être équivalentes. Par exemple, on constate que l'amplitude des changements conformationnels lors de la transition CS6→CS7 (courbe verte) est légèrement inférieure à celle induite par la liaison de la thyroxine (courbe noire), sauf pour les résidus K233, H247, D269, K274 , K276, E280 et E292, qui ne sont généralement pas impliqués dans la complexation (parmi ces résidus, les quatre derniers sont situés dans la boucle reliant les hélices 15 et 16). De toute évidence, la transition CS6→CS7 donne seulement un aperçu partiel et statique de la flexibilité du site 1. Il est à noter que les fluctuations autour de la structure moyenne d'un conformère contribuent également à la flexibilité du site. Une illustration de cette flexibilité est obtenue en représentant le mouvement progressif de la protéine au long des différentes PC. L'image correspondante au mouvement de la première composante est donnée à la Figure 3.10, où les déplacements des résidus entre 2 et 3 Å sont respectés.

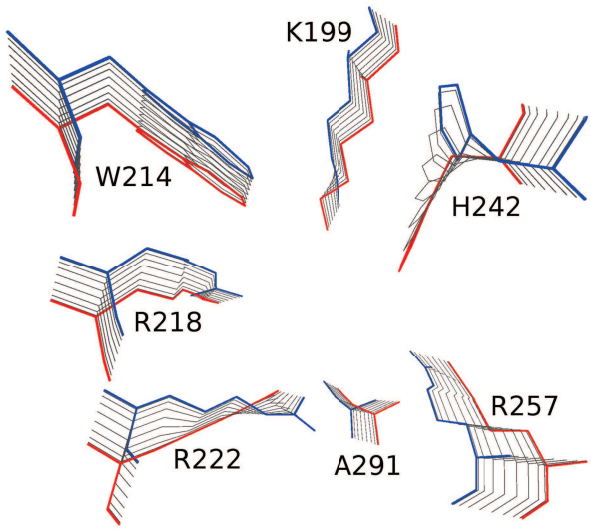

Figure 3.10 – Mouvement des principaux résidus impliqués dans la formation du site 1 au long de la première composante PC1 de traj1

Cette flexibilité dynamique (résultant d'une surface d'énergie libre plate) va certainement faciliter l'arrivée du médicament dans le site de complexation sans impliquer une augmentation significative de l'énergie libre de Gibbs conformationnelle.

3.4.3 Étude des conformères du site 2

3.4.3.1 Changements conformationnels entre les différents conformères

L'image des quatre conformères nommés de CS8 à CS11 est donnée à la Figure 3.11.

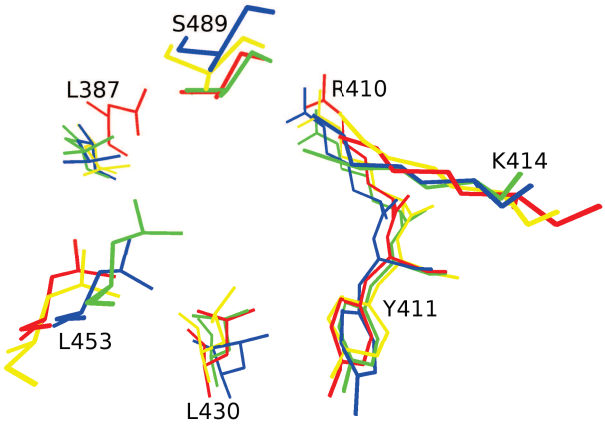

Figure 3.11 – Conformation relative des principaux résidus d'acides aminés du site 2 dans la structure moyenne des conformères CS8 (bleu), CS9 (rouge), CS10 (jaune) et CS11 (vert).

Des changements conformationnels sont détectés dans les chaînes latérales des résidus L387 et L453. Au cours de la transition CS8→CS9, la chaîne latérale du résidu L387 tourne autour de l'axe de la liaison $C_\beta - C_\gamma$ d'environ 70° et autour de l' axe $C_\alpha - C_\beta$ de 20° tandis que lors de la transition CS8→CS10 la chaîne latérale du L453 tourne autour de la liaison $C_\beta - C_\gamma$ de 20°. Ces résidus jouent un rôle actif dans la liaison du diazepam. D'importantes rotations des chaînes latérales ont été observées expérimentalement induisant une augmentation de la distance relative entre les deux chaînes latérales d'environ 2.3 Å [6]. Il est intéressant de constater qu'une augmentation similaire de la distance entre les chaînes latérales des résidus L387-L453 est observée dans le cas de la transition CS8→CS9. Par conséquent, cette transition a été choisie pour une analyse plus détaillée. Le RMSF de la transition CS8→CS9, calculé pour tous les atomes, présente d'importantes fluctuations des résidus L387, Q417 et Q466 tandis que la valeur moyenne du RMSF est d'environ 2 Å (Figure 3.12, rouge). La grande mobilité des résidus Q417 et Q466 s'explique par leur emplacement marginal : le premier est situé dans la boucle de l'hélice 18 tandis que le second est le résidu terminal de la vingt-huitième hélice α.

3.4.3.2 Une parallèle entre la transition théorique CS8→CS9 et celle observée expérimentalement lors de la complexation du diazépam

Une comparaison avec le changement de conformation induit par la complexation du diazépam montre que dans le dernier cas, les valeurs du RMSF sont généralement plus uniforme tandis que sa valeur moyenne est supérieure d'environ 1 Å(Figure 3.12, noir).

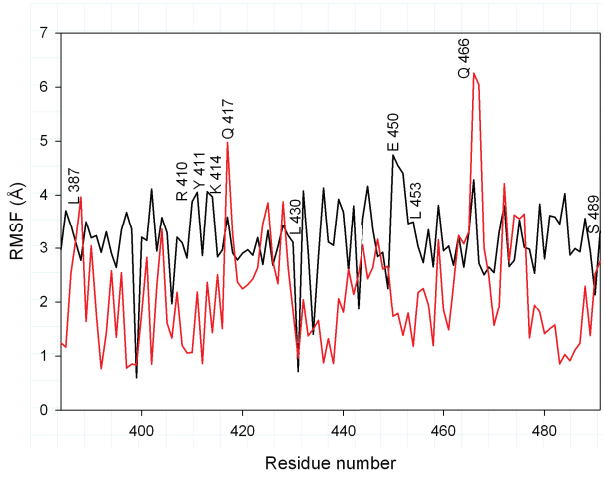

Figure 3.12 – RMSF calculé entre les conformations CS8 et CS9 (rouge) et entre la HSA native et le complexe HSA-diazepam (PDB ID : 1HA2) (noir).

Cette différence peut très probablement être compensée par la flexibilité "dynamique" du site de complexation (fluctuations autour de la structure moyenne des conformations qui ne figurent pas dans la Figure 3.12). La flexibilité du site est également illustrée par les différences qui se produisent spontanément entre les résidus adjacents (courbe rouge dans la Figure 3.12) : des valeurs de 2-3 Å sont observées. D'autre part, dans le cas de la fixation du diazépam, les valeurs du RMSF pour les résidus voisins sont très probablement lissées par l'encombrement stérique imposé par le médicament (les chaînes latérales des résidus s'adaptent à la forme du médicament).

3.5 Corrélation des mouvements entre les sites 1 et 2

Le couplage allostérique entre les sites 1 et 2 a été démontré expérimentalement dans le cas de dasyl-L-asparagine dont la complexation au site 1 a été inhibée par la complexation simultanée d'ibuprofène ou de diazépam sur le site 2 [93]. Une condition nécessaire au couplage est l'apparition de modes de PC délocalisés dans les deux sous-domaines. C'est pourquoi, nous avons analysé la corrélation entre les mouvements des deux sites en effectuant une PCA des sous-domaines IIA et IIIA. La trajectoire utilisée dans l'analyse de covariance est obtenue par projection de la trajectoire de simulation globale de 200ns (traj 1) sur le sous-espace formé par les deux sous-domaines. L'analyse a été effectuée sur les atomes de carbone α.

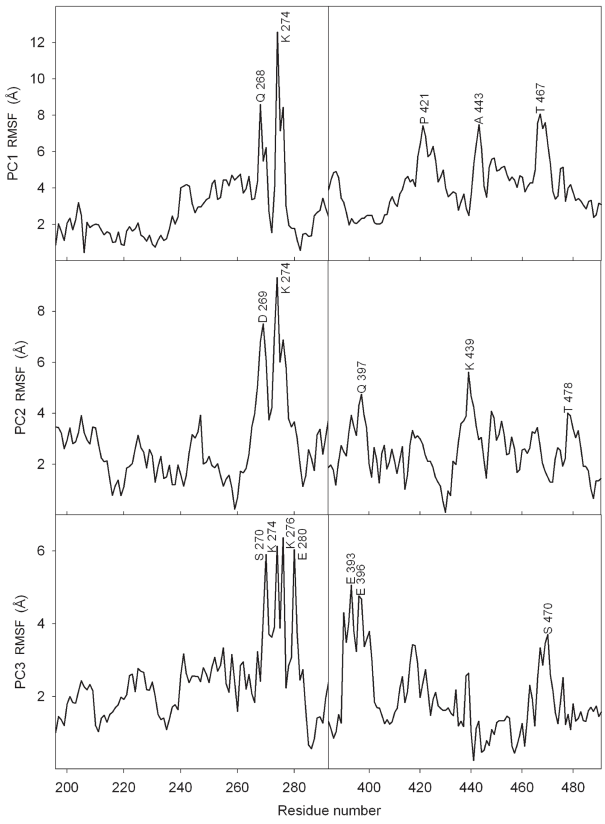

Figure 3.13 – Le déplacement (RSD) des C_α des trois premières PC de la trajectoire restreinte aux deux sous domaines IIA et IIIA

Les trois premières PC obtenues sont clairement délocalisées sur les deux sous-domaines

de la protéine (Figure 3.13) confirmant ainsi le couplage entre les mouvements de ces sous-domaines. Il a déjà été proposé que ce couplage apparaît via l'interface commune des sous-domaines, en incluant les résidus K195, R197, L198, A201, K205, F206 (sous-domaines IIA) et V455, N458, E465, L481, R484 (sous-domaines IIIA)[93]. Cette hypothèse qui paraît logique reste néanmoins difficile à vérifier dans la présente analyse. En se basant sur la distribution des conformations projetée sur le sous-espace des trois premières PC, nous avons identifié 5 conformères : de CS12 à CS16. Parmi eux, les trois premiers sont de loin les plus importants en raison de leur poids statistique :

	CS12	CS13	CS14	CS15	CS16
Population (%)	28.4	14.9	10.3	2.0	1.0

Tableau 3.5 – Population (%) des différents clusters de conformations identifiés par PCA après projection de traj 1 sur le sous-espace formé par les deux sous-domaines IIA et IIIA

Les transitions conformationnelles CS12→CS13 et CS13→CS14 sont illustrées dans la Figure 3.14 en représentant les déplacements correspondants aux C_α.

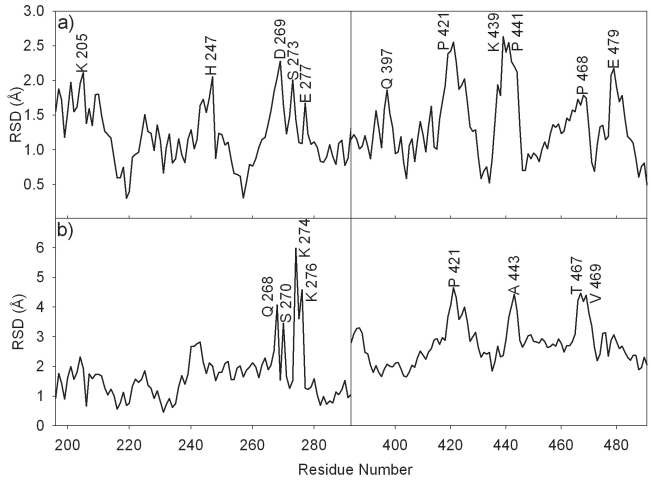

Figure 3.14 – Le déplacement (RSD) des C_α au cours des transitions CS12→CS13 (a) et CS13→CS14 (b)

Il est à noter que la transition CS12→CS13 est certainement réversible car elle se produit plusieurs fois au cours de la trajectoire de simulation. Les transitions sont bien délocalisées sur les deux sites de complexation et pour cette raison, elles peuvent être considérées comme base à de potentiels effets allostériques. Des simulations supplémentaires

en présence de ligands sont nécessaires pour démontrer que ces transitions peuvent être initiées par la complexation d'un médicament.

3.6 Conclusions

D'après les résultats présentés dans ce chapitre, l'échantillonnage d'une DM de la HSA n'est pas entièrement convergente, même pour une longueur de trajectoire de 200ns. Cependant, l'analyse de covariance appliquée à trois de ces trajectoires a donné des informations physiquement pertinentes sur la dynamique conformationnelle de la protéine dans l'intervalle de temps simulé. La première PC dérivée d'une des trois trajectoires ici analysée décrit une transition conformationnelle dans le domaine I proche de celle observée expérimentalement lors de la complexation du myristate. Ce résultat supporte l'hypothèse que dans le cas de la HSA, un changement de conformation favorable à la complexation du médicament peut précèder celle-ci. D'autre part, il a été constaté que les cinq premières PC, obtenues pour chacune des trajectoires, sont étalées sur la structure entière de la protéine indiquant que les mouvements des trois domaines sont fortement couplés. Une PCA détaillée appliquée au mouvement collectif des sous-domaines IIA et IIIA prédit l'existence de transitions conformationnelles délocalisées sur les sites 1 et 2. Ces transitions peuvent être à l'origine d'éventuels effets allostériques entre les deux sites.

D'un point de vue plus général, la relation existante entre les transitions conformationnelles naturelles et celles induites par la complexation d'un médicament dans la HSA est un sujet qui nécessite un travail théorique et expérimental supplémentaire.

Chapitre 4

Rôle structural des ponts disulfure

Sommaire

4.1 Introduction 66

4.2 Test de stabilité structurale 66

4.3 Effets de la réduction des ponts sur les structures secondaire et tertiaire 68

4.4 Effet de la réduction des ponts S—S sur le rayon de gyration de la protéine 72

4.5 Dissociation des couples thiols 73

4.6 Conclusions 79

4.1 Introduction

Les ponts disulfure sont des contraintes physiques connus pour jouer un rôle essentiel dans la stabilité des structures natives des protéines. Toutefois, les conséquences structurales de la suppression des liaisons S—S peuvent varier d'une protéine à l'autre. Ce chapitre a pour vocation d'apporter de nouveaux éclairages concernant le rôle des 17 ponts disulfure de la HSA dans le maintien de sa structure native ainsi que sur le mécanisme de dépliement de la protéine réduite. Deux types de simulations sont ainsi effectuées : sur la structure intacte de la protéine (S—S) et sur sa structure réduite (SH- / -HS). Une étude préliminaire afin de valider l'efficacité des champs de force, OPLS-AA et AMBER ff03, à reproduire le dépliement protéique est réalisée. Après validation de la méthode de simulation, la comparaison systématique du dépliement de la structure réduite avec celui de la structure native permet d'examiner les principaux changements structuraux résultant de la réduction des ponts S—S. La principale contrainte de ce travail résulte du fait que le dépliement des structures protéiques doit se faire dans des temps de calcul accessibles. Dans cette étude, il a été choisi de forçer le dépliement en augmentant la température de simulation. En effet, les simulations de DM à haute température ont prouvé leur efficacité dans l'activation du processus de dépliement [33, 53, 61]. Dans la dernière partie, une attention spéciale est accordée au mécanisme de dissociation des ponts S—S réduits.

4.2 Test de stabilité structurale

Dans un premier temps, la stabilité des structures simulées native et réduite à température ambiante a été testée en utilisant les champs de force AMBER ff03 et OPLS-AA. Les simulations de DM sont effectuées à 300 K pour des temps de 20 ns. Les trois observables retenues pour suivre l'évolution des structures secondaire et tertiaire au cours des trajectoires de simulation sont : le RMSD (Figure 4.1), le pourcentage en hélice-α (Figure 4.2) et le pourcentage des contacts natifs (Figure 4.3).

Pour chaque simulation, les trois observables sont stables à la fin des 20 ns de simulation. Le pourcentage à l'équilibre du contenu en hélice-α pour les deux champs de force est de 70 %, ce qui est pratiquement identique à la structure de référence. Le pourcentage des contacts natifs diminue, quant à lui, à 70 % pour le champ de force AMBER ff 03 et descend même jusqu'à 65% pour le champ de force OPLS-AA. Ces résultats, en

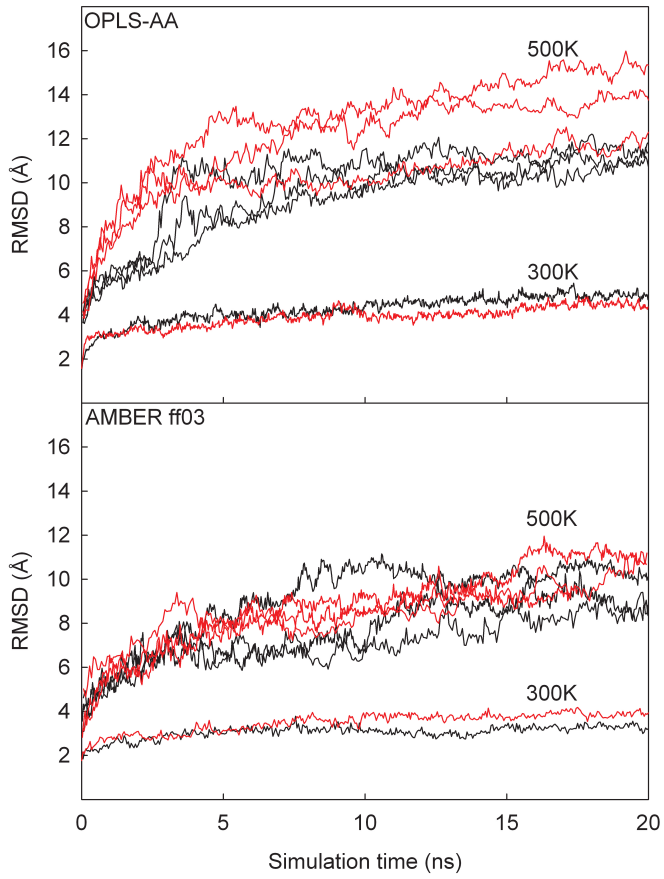

Figure 4.1 – RMSD du squelette protéique calculé pour la protéine native (en noir) et réduite (en rouge) avec comme référence la structure de diffraction aux RX. Les simulations sont effectuées à 300 K et 500 K (trois trajectoires différentes pour chaque protéine) avec les champs de force OPLS-AA et AMBER ff 03.

accord avec la légère hausse du RMSD, indiquent que les structures d'équilibre sont proches de la structure cristallographique. Les résultats obtenus avec la protéine réduite montrent des caractéristiques identiques à celles décrites pour la protéine native. Ce résultat suggère que le processus de dépliement de HSA$^{(R)}$ est trop lent pour être observé à 300 K pour l'échelle de temps simulé ici.

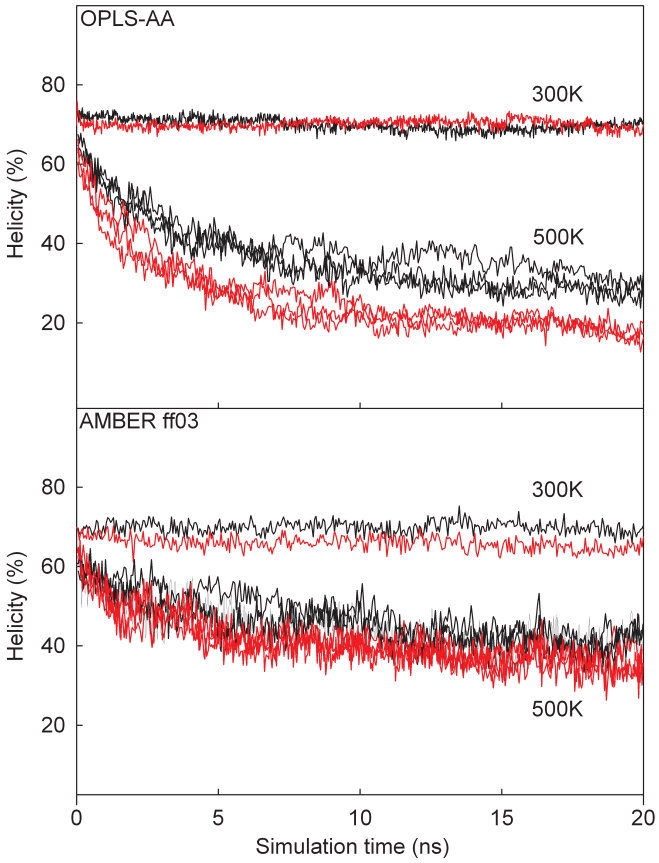

Figure 4.2 – Structure secondaire : évolution du contenu en hélice-α de la protéine native (en noir) et réduite (en rouge). Les simulations de DM sont effectuées à 300 K et 500 K avec les champs de force OPLS-AA et AMBER ff 03.

4.3 Effets de la réduction des ponts sur les structures secondaire et tertiaire

Le dépliement thermique des protéines native et réduite a été comparé pour les températures suivantes : 400, 450 et 500 K. Des différences significatives apparaissent entre les deux formes protéiques principalement à 500 K. L'analyse qui suit a donc été basée uniquement sur cette dernière température pour laquelle ont été générées trois trajectoires différentes pour chaque protéine. Globalement, les différences les plus importantes apparaissent pour les simulations effectuées avec le champ de force OPLS-AA. On peut observer, par exemple, que la hausse du RMSD au long de la trajectoire de simulation pour HSA$^{(R)}$ est bien plus rapide que pour la protéine intacte (Figure 4.1).

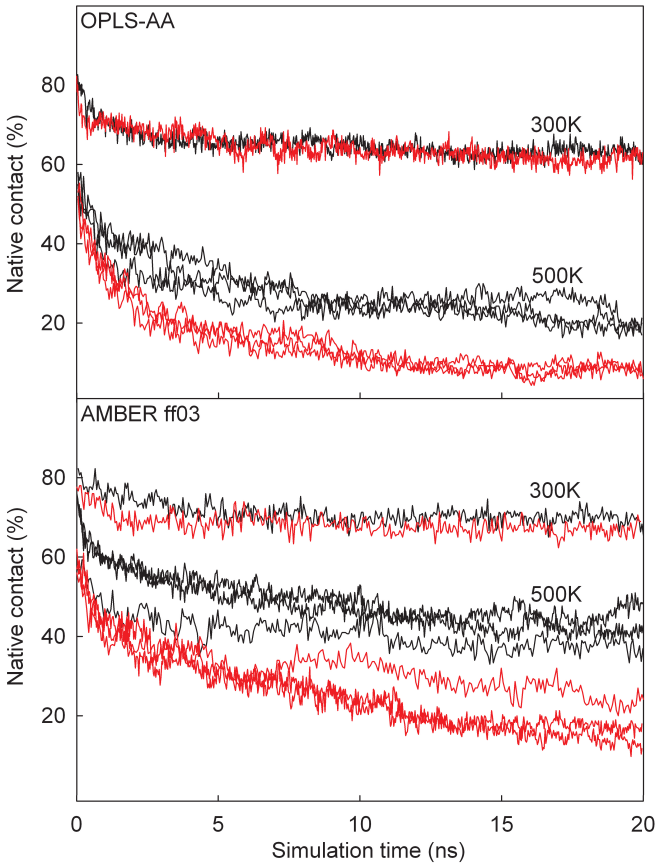

Figure 4.3 – Structure tertiaire : évolution du pourcentage des contacts natifs dans la protéine native (en noir) et réduite (en rouge). Les simulations de DM sont effectuées à 300 K et 500 K avec les champs de force OPLS-AA et AMBER ff 03.

La valeur finale du RMSD moyennée sur les trois trajectoires est approximativement de 13.0 Å pour $HSA^{(R)}$ et de 10.0 Å pour $HSA^{(N)}$. Dans le cas du champ de force AMBER ff 03, les différences entre les deux formes protéiques sont en partie cachées par les fluctuations de leurs RMSD respectifs. Cependant, visuellement on peut observer qu'en moyenne les RMSD de $HSA^{(R)}$ ont systématiquement une valeur supérieure à celle de $HSA^{(N)}$. À la fin de la trajectoire de simulation, la différence entre les deux valeurs moyennes du RMSD est d'environ 1.5 Å.

Évolution du contenu en hélices-α : changements dans la structure secondaire

Une différence encore plus prononcée entre les deux types protéiques est observée pour le dépliement de la stucture secondaire (Figure 4.2). Lors des 5 premières ns avec le champ de force OPLS-AA, le pourcentage en hélice-α de HSA$^{(R)}$ diminue de 70 à 30% alors qu'il se stabilise à environ 45% pour HSA$^{(N)}$. Cette différence absolue de 15% se maintient jusqu'à la fin des trajectoires de simulation. Pour AMBER ff 03, la différence est moins marquée mais elle est également systématique. L'effet déstabilisant de la réduction des ponts disulfure sur la structure secondaire de la HSA est un résultat expérimental bien établi. Expérimentalement, il a été montré que l'amplitude de la bande DC à 208nm, qui est une caractéristique de la structure secondaire en hélice-α, est divisée par un facteur 2 lors de la réduction des 17 ponts disulfure de la protéine [94]. Une description alternative de cet effet peut être obtenue par la spectroscopie Raman. L'intérêt de cette méthode résulte dans le fait que la bande Raman amide I est très sensible à la structure secondaire de la protéine. De plus, celle-ci peut être décomposée en différentes contributions représentant chacune un élément spécifique de la structure secondaire (hélice-α, feuillet β, etc...) [95]. Dans le cas de la HSA, il a été montré que la contribution spécifique aux hélices-α diminue d'environ 30% pour la forme réduite [27]. Nous en concluons que les champs de force OPLS-AA et AMBER ff 03 décrivent au moins qualitativement cet effet déstabilisant. En revanche, une comparaison quantitative directe entre ces résultats théoriques et ceux expérimentaux décrits précédemment ne serait pas suffisamment pertinente. En effet, les simulations de dépliement thermique rapportées ici fournissent une information directe de la cinétique de dépliement des deux formes protéiques alors que les résultats expérimentaux présentés décrivent les structures d'équilibre correspondantes. *Grosso modo*, le dépliement spécifique de la protéine réduite peut être évalué comme étant la différence entre les dépliements thermiques de HSA$^{(N)}$ et HSA$^{(R)}$. Cependant, cette évaluation pourrait sous estimer le dépliement spécifique dès lors que le processus de dépliement n'est pas terminé à la fin des 20 ns de simulation.

Évolution de la fraction des contacts natifs : changements dans la structure tertiaire

La différence la plus marquée entre le dépliement thermique des deux formes protéiques apparaît lorsque la fraction des contacts natifs est prise en considération (Figure 4.4). Sur la deuxième moitié de la trajectoire de simulation, une différence moyenne d'environ 20 unités pour OPLS-AA et 30 unités pour AMBER ff03 est observée entre les formes native et réduite. Par conséquent, les deux champs de force prédisent une déstabilisation

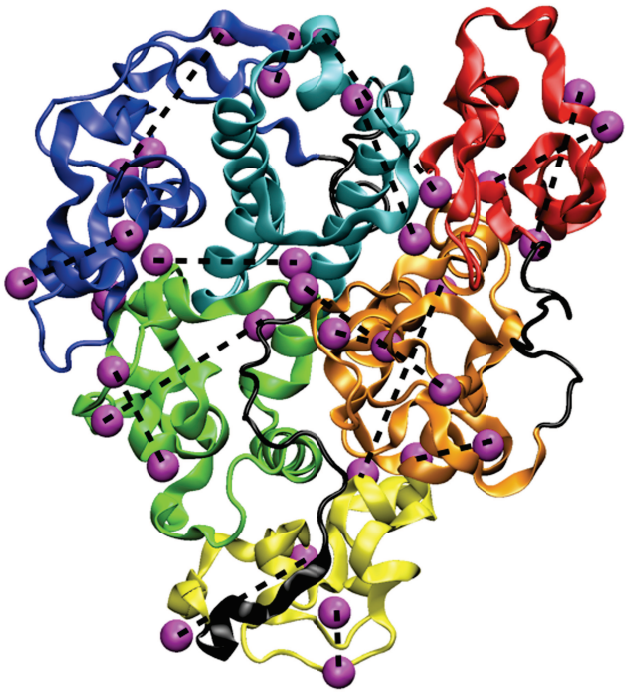

Figure 4.4 – Distortion de la structure tertiaire de HSA$^{(R)}$. La structure a été extraite de la trajectoire avec le champ de force OLPS-AA simulé à 500 K. Les atomes de soufre sont représentés par des sphères solides et sont connectés par des lignes pointillées lorsqu'ils appartenaient au même pont réduit.

significative de la structure tertiaire de la protéine lors de la réduction des ponts disulfure. Expérimentalement, des altérations dans la structure tertiaire de HSA$^{(R)}$ ont déjà été mises en évidence en observant des changements dans la fluorescence du résidu Trp-214 [22]. Il a été détecté, une diminution de 30,6% de l'intensité de la fluorescence par rapport à la forme native et un décalage vers le bleu d'environ 12 nm du maximum de la bande, indiquant un changement important dans l'environnement du résidu. Le désordre prédit théoriquement dans la structure tertiaire lors de la réduction des ponts S—S de la protéine a un support expérimental. Évidemment, ces changements de conformations peuvent avoir des conséquences très importantes sur la fonction de transport de la HSA puisque les sites de liaison devraient être sensibles à ce genre de modifications structurales. Cela a été démontré expérimentalement pour le site de complexation de la warfarine [22]. Une illustration du désordre de la structure tertiaire de la HSA réduite est présentée dans la Figure 4.4. La structure a été extraite à environ 5ns du début de la simulation. Lorsque cette structure est comparée à la structure native de référence, on peut observer que les cavités bien caractérisées sous la forme native ont presque disparu.

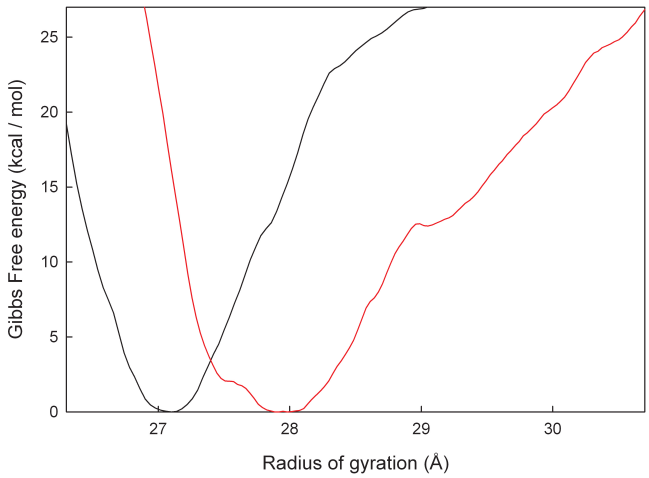

Figure 4.5 – Profils d'énergie libre de Gibbs en rapport avec le rayon de gyration pour les formes native (noir) et réduite (rouge) en utilisant le champ de force OPLS-AA et la méthode Umbrella Sampling.

4.4 Effet de la réduction des ponts S—S sur le rayon de gyration de la protéine

La HSA est une protéine globulaire compacte et, de ce point de vue, une question intéressante à examiner est celle du rôle des ponts disulfure dans le maintien de cette compacité. Nous avons étudié ce rôle en calculant les profils d'énergie libre de Gibbs par rapport au rayon de gyration pour les formes protéiques native et réduite. Pour ces calculs, nous avons choisi le champ de force OPLS-AA puisque, selon les résultats présentés dans la première partie (Figures 4.1, 4.2, 4.3), il semble être plus sensible à reproduire les effets globaux de la réduction des ponts S—S. Les profils d'énergie libre ont été calculés par la méthode *Umbrella Sampling* à T=300 K à partir d'une structure protéique très proche de celle de référence. Le traitement à l'aide du logiciel WHAM s'est effectué sur 51 fenêtres de 0.1 Å. La durée de simulation par fenêtre est de 2ns. L'erreur moyenne sur le calcul du profil d'énergie libre de Gibbs en rapport avec le rayon de gyration pour la forme native est de $\pm$ 0.2162 kcal/mol et de $\pm$ 0.2541 kcal/mol pour la forme réduite. La courbe obtenue pour la protéine native présente un minimum centré à environ 27.1 Å (Figure 4.5). Le minimum pour la protéine réduite est décalé d'environ 1 Å vers les rayons les plus importants. Dans les deux cas, les minima sont très stables et l'augmentation de l'énergie libre de Gibbs nécessaire pour quitter le puits est supérieur à 30 $kcal.mol^{-1}$. On note également que le

profil d'énergie libre, calculé pour la forme réduite est plus large, ce qui indique un rayon de gyration plus fluctuant. Afin de vérifier si le dépliement de la structure secondaire affecte le rayon de gyration, nous avons également calculé le profil d'énergie libre de Gibbs correspondant en commençant par une structure extraite de la trajectoire de DM simulé à 500 K. Le profil (non représenté) était pratiquement identique à celui calculé à partir de la structure réduite repliée (Figure 4.5, rouge). En ce qui concerne le rayon hydrodynamique (R_H) des albumines sériques réduites, les études expérimentales semblent diverger. Lee et Hirose [94] ont utilisé la chromatographie d'exclusion de taille pour mesurer le R_H d'albumines sériques natives et entièrement réduites. Ils ont constaté que la réduction des ponts disulfure induit une augmentation du R_H des albumines sériques de 33.9 Å (protéine native) à 42.0 Å (protéine réduite). Tanner et al. [96] ont étudié l'évolution du R_H de la réduction des BSA complexées avec le dodécylsulfate de sodium (SDS) à l'aide de la diffusion de lumière dynamique. Ils en ont déduit que, dans la gamme des concentrations faibles de SDS, le R_H de la protéine réduite est de 32 Å. Valstar et al. [97] ont confirmé par la même technique que les BSA réduite et native complexées au SDS ont un R_H similaire sur un large domaine de concentration de SDS. Ils ont également constaté que, en l'absence de SDS, la protéine réduite forme des agrégats, le monomère étant pratiquement absent en solution aqueuse. D'autre part, Jachimska et al. [98] ont trouvé pour le R_H du dimère de la HSA une valeur de 41 Å. Dans ces circonstances, la valeur de 42.0 Å mesurée par Lee et Hirose pour la HSA réduite pourrait être plutôt attribuée au dimère de la protéine. Un travail expérimental plus poussé est nécessaire afin de clarifier cet aspect. Les résultats théoriques, présentés ici, prédisent que la réduction des ponts disulfures de la HSA induit une légère augmentation, de l'ordre de 3 %, du rayon de gyration de la protéine. Ceci est en accord avec l'image d'une protéine comprimée par des forces intramoléculaires non covalentes et les interactions soluté-solvant, les liaisons S—S covalentes ne jouant qu'un rôle limité dans cette compression.

4.5 Dissociation des couples thiols

Par rapport à la protéine intacte, la protéine réduite possède 17 degrés de liberté internes supplémentaires qui sont les distances S—S correspondantes aux 17 ponts réduits. De manière générale, le dépliement spécifique de la protéine réduite correspond à la relaxation de ces nouvelles coordonnées sur la surface d'énergie potentielle de la protéine. De toute évidence, les changements spécifiques dans les structures secondaire et tertiaire

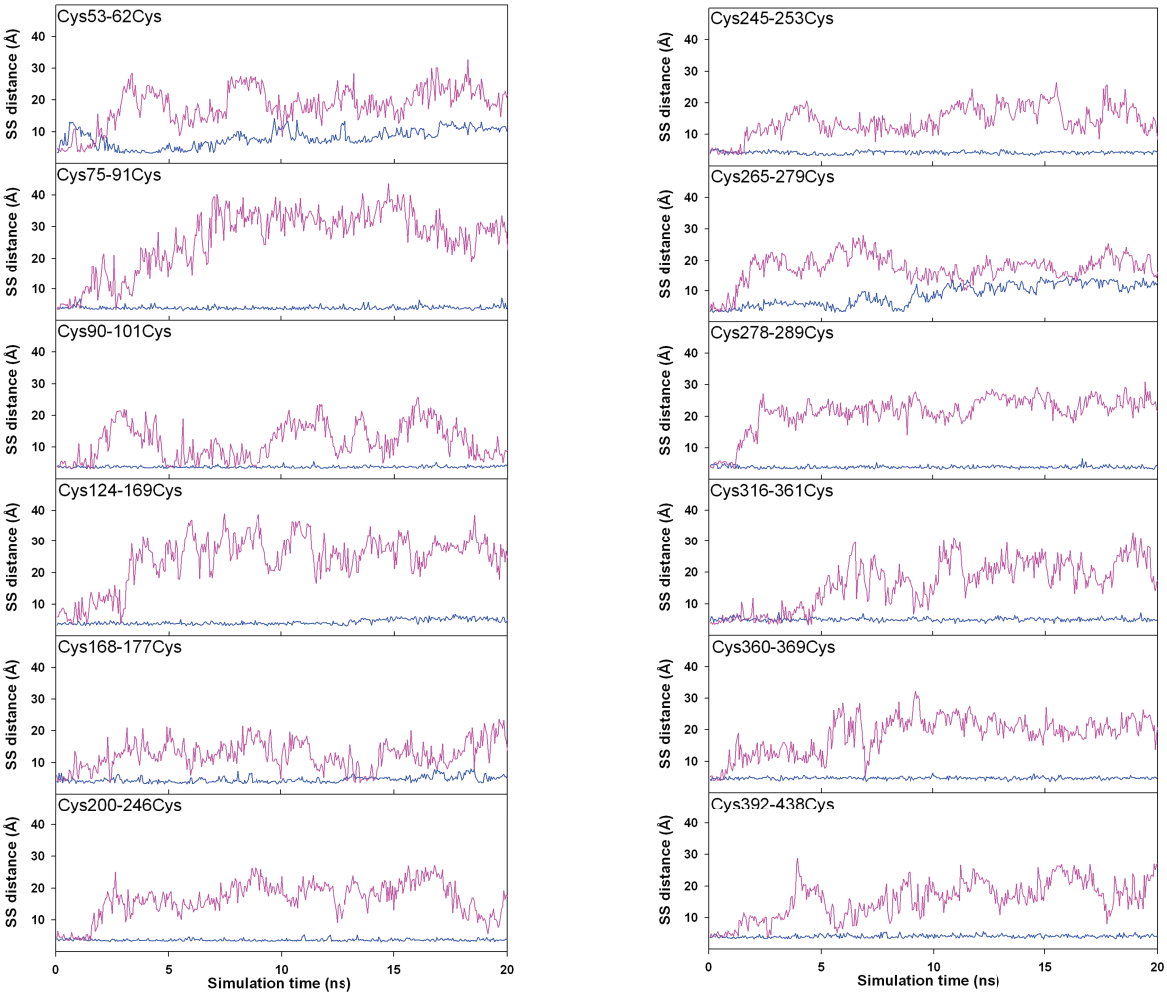

Figure 4.6 – Evolution des distances S—S pour la HSA réduite : simulations de DM à 300 K (bleu) et 500 K (magenta) réalisées avec le champ de force OPLS-AA

de la protéine réduite sont des conséquences de cette détente. Il est donc utile d'analyser la dissociation des couples thiols. Cette analyse a été réalisée sur la trajectoire simulée avec le champ de force OPLS-AA à 500 K (Figure 4.6). A titre de comparaison, la variation des distances S—S des ponts disulfure réduits est également donnée pour la trajectoire simulée à 300 K.

Tout d'abord, on peut noter qu'à 300 K, 13 des 17 ponts disulfure restent en contact (leurs distances S—S varient autour de 4.5 Å). Ceci est une indication claire que dans la conformation native de la protéine, les groupes -SH appartenant à un pont S—S réduit sont situés dans des puits de potentiel qui les maintiennent en contact. La conséquence est une limitation importante du dépliement spécifique de la protéine réduite. La forme de ces puits de potentiel a été calculée en utilisant la méthode *Umbrella Sampling* pour

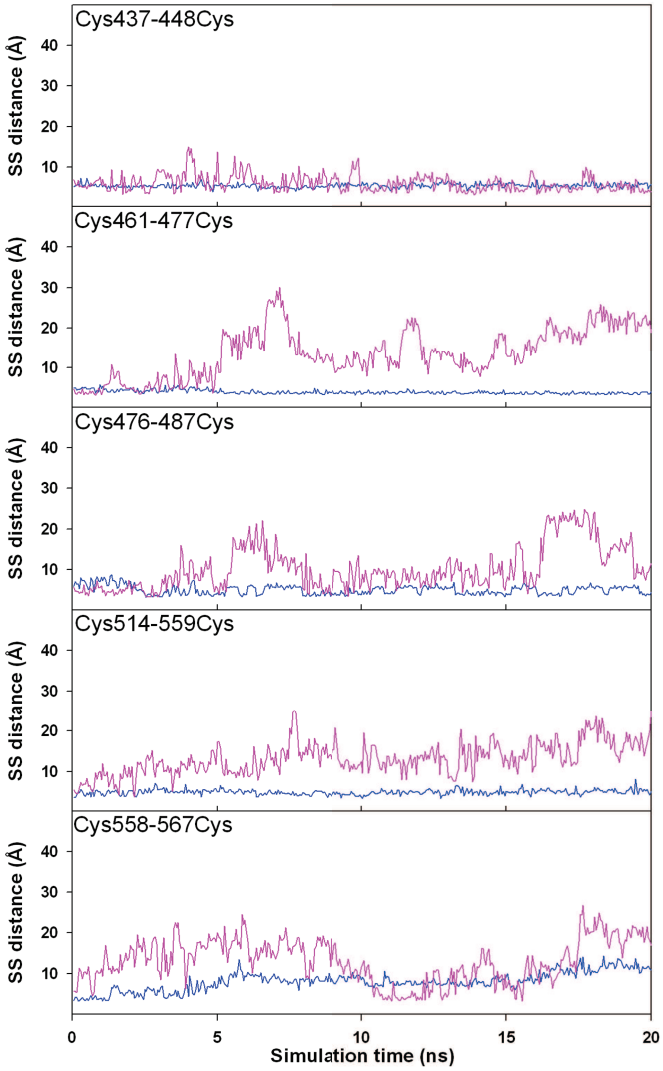

Figure 4.6 (suite)

les trois ponts réduits du sous-domaine IA de la protéine (Figure 4.7). Pour chaque pont, 16 fenêtres de simulations de 0.5 Å chacune ont été réalisées afin de calculer ces profils. L'erreur moyenne sur chaque profil est donné dans le tableau 4.1 :

Erreur moyenne (±kcal/mol)	Cys53-62Cys	Cys75-91Cys	Cys90-101Cys
$HSA^{(N)}$	0.11	0.12	0.19
$HSA^{(R)}$	0.14	0.17	

Tableau 4.1 – Erreur moyenne (kcal/mol) sur le calcul des profils d'énergie libre de Gibbs des ponts S—S du sous-domaine IA en fonction de la distance SS

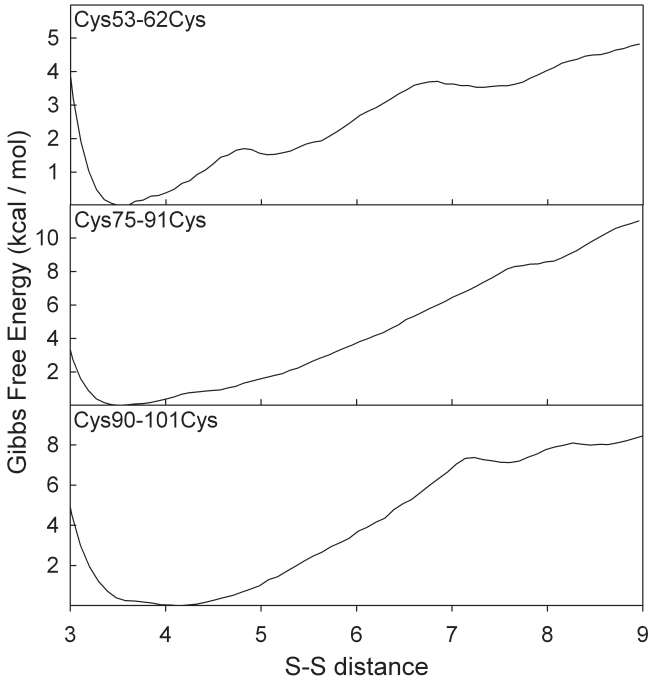

Figure 4.7 – Profils d'énergie libre de Gibbs calculés par rapport à la distance thiol-thiol pour les trois ponts réduits du sous-domaine IA de la HSA. Les calculs ont été effectués sur la conformation de la protéine native.

Pour les ponts réduits Cys53-62Cys et Cys75-91Cys, les minima énergétiques sont localisés aux environs de 3.5 Å alors que pour le pont Cys90-101Cys, il est situé vers 4.0 Å. On peut aussi noter que, dans les trois cas, l'énergie libre de Gibbs augmente relativement lentement lorsque la distance S—S augmente. Ceci indique une certaine flexibilité de la structure protéique. Manifestement, ces puits de potentiel ne sont pas liés à une interaction directe entre les groupements thiols mais ils expriment la réaction de la protéine lors de cette déformation.

L'existence de ces puits de potentiel est également confirmée par l'évolution des distances S—S à 500 K : dans le cas d'au moins 4 ponts réduits (Cys124-169Cys, Cys200-246Cys, Cys245-253Cys et Cys316-361Cys) la dissociation est précédée par un intervalle de simulation de 2-5 ns durant laquelle les thiols restent en contact, même à cette température élevée. Dans le cas de Cys437-448Cys le contact est préservé sur toute la trajectoire de simulation. Cependant, dans la plupart des cas, la dissociation se produit soudainement par une transition prenant moins de 1ns. Un changement de ce genre est clairement lié à une transition conformationnelle. Une analyse minutieuse de la structure protéique montre

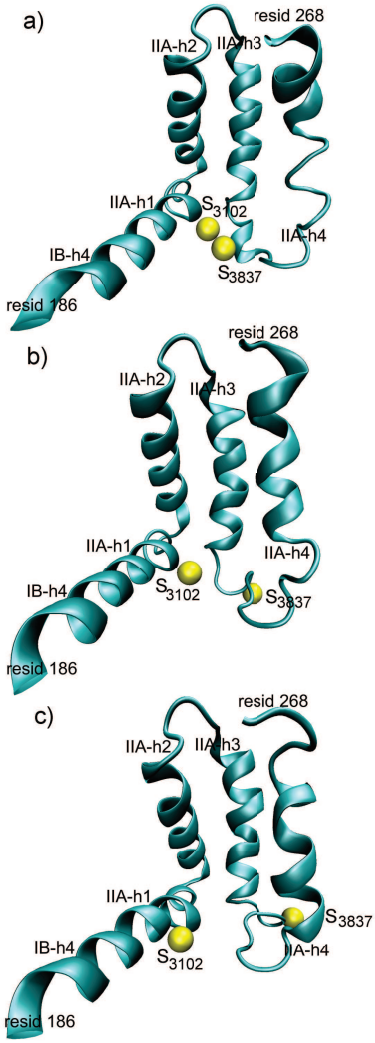

Figure 4.8 – Illustration du dépliement du pont réduit Cys200-246Cys : les structures ont été extraites de la trajectoire OPLS-AA simulée à 500 K : a) t = 2 ns, b) t = 4 ns , c) t = 8 ns. Les atomes de soufre sont représentés par des sphères solides.

que cette transition conformationnelle a lieu autour des résidus cystéine et plus précisement chez ses voisins de premier et second ordre.

Elle se compose principalement d'un changement de conformation dans la chaîne latérale de la cystéine associé à des changements dans les angles dièdres peptidiques ϕ et ψ. Pour illustrer ce type de processus, nous présentons dans la Figure 4.8 la dissociation du pont réduit Cys200-246Cys.

Les structures a, b, c ont été extraites de la trajectoire de DM de la HSA réduite simulée à 500 K. Cet exemple montre également que, pendant le processus de séparation,

les changements conformationnels individuels s'accumulent et le processus inverse devient de moins en moins probable. On constate que, dans le cas présent, la transition conformationnelle responsable de la dissociation des deux thiols est située dans la boucle reliant les hélices IIA-h3 et IIA-h4. Quelques autres exemples de dissociation lors de transitions conformationnelles locales sont données dans le Tableau 4.1. Pour chaque transition seule la contribution majeure a été examinée. Ces données ont été extraites des trajectoires représentées sur la Figure 4.6.

		Angles Dièdres (°)		Distances S-S (Å)	
Pont disulfure	Angle dièdre	Initial	Final	Initial	Final
Cys200-246Cys	Ψ_{244}	-29,21	145,59	3,57	12,27
Cys245-253Cys	Φ_{243}	-60,87	-153,37	4,47	10,88
Cys278-289Cys	Ψ_{279}	-34,03	158,07	11,26	25,15
Cys360-369Cys	Ψ_{360}	-41,71	37,22	7,77	21,01

Tableau 4.2 – Exemples de changements conformationnels dirigeant la dissociation des thiols

Une analyse plus précise de la dissociation des ponts réduits est obtenue en calculant les profils d'énergie libre de Gibbs des distances thiol-thiol pour la conformation de la protéine modifiée. Nous avons effectué cette analyse pour les ponts réduits Cys53-62Cys et Cys75-91Cys appartenant au sous-domaine IA de la protéine (Figure 4.9 : à comparer à la Figure 4.7).

Il semble que dans la conformation modifiée les puits de potentiel thiol-thiol sont maintenus mais leurs minima sont nettement déplacés vers de plus longues distances S—S. Les puits de potentiel modifiés sont relativement larges et l'augmentation de l'énergie libre de Gibbs nécessaire à remettre les deux thiols en contact est plutôt faible, environ 5 $kcal.mol^{-1}$. Cela signifie que la re-formation des ponts S—S ne nécessite pas une séquence inverse des transitions conformationnelles mais est possible par une simple diffusion des thiols. Ce résultat explique pourquoi, dans des conditions oxydantes, la HSA réduite se replie et retrouve sa structure native [94]. En effet, la reconstitution des ponts S—S peut précéder le repliement structural. De plus, lorsque les ponts S—S sont reconstitués, la probabilité de repliement augmente parce que l'entropie de l'état déplié est considérablement réduite par les contraintes S—S.

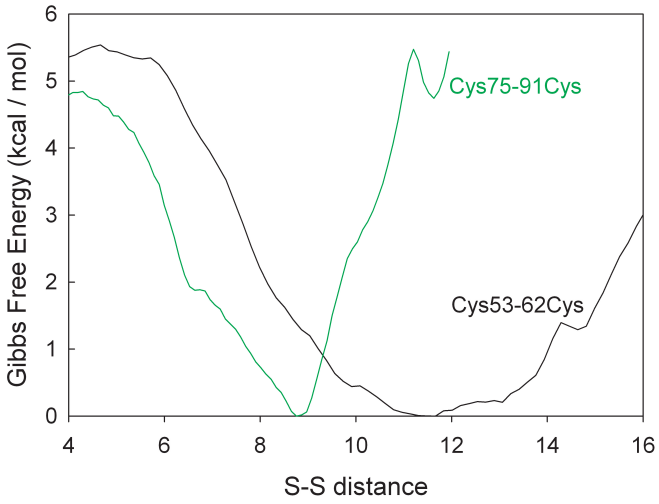

Figure 4.9 – Profils d'énergie libre de Gibbs calculés par rapport à la distance thiol-thiol pour les deux ponts réduits dans une conformation séparée.

4.6 Conclusions

Les simulations de DM présentées ici, prédisent trois effets principaux de la réduction des ponts disulfure dans les albumines sériques : une déstabilisation de la structure secondaire de la protéine, une déformation de la structure tertiaire (prouvée par une diminution significative de la fraction des contacts natifs) et une légère expansion du volume protéique (directement lié au rayon de gyration de la protéine). Ces prédictions sont en accord qualitatif avec les résultats expérimentaux disponibles. Généralement, les deux champs de force testés donnent des résultats similaires, mais le champ de force AMBER ff03 semble sous-estimer le dépliement de la structure secondaire de la protéine (champ de force plus rigide). Au-delà de la validation de la méthode de simulation, les résultats présentés apportent de nouveaux éclairages sur les processus de réduction/oxydation dépliement/repliement de la protéine. Il a été établi que, dans la conformation native de la protéine, les thiols sont maintenus en contact dans des puits de potentiel spécifiques. Le mécanisme de séparation des ponts réduits se produit par des transitions conformationnelles de la chaîne peptidique situées dans les résidus voisins des cystéines impliqués. Les résultats des simulations suggèrent également que le repliement oxydatif se déroule en deux étapes : dans la première le pont S—S se réforme à la suite d'un simple processus diffusif et dans la seconde la protéine se replie et retrouve sa conformation native.

La réduction des ponts S—S touche également de façon très importante la structure des

sites de complexation de la protéine et va perturber ainsi la fonction principale de celle ci qui est le transport. De ce point de vue, l'analyse présentée dans ce chapitre est insuffisante parce qu'elle ne permet pas de prédire la structure d'équilibre de la protéine réduite. Des études de DM supplémentaires sont nécessaires afin de compléter la description théorique de ces changements structuraux.

Chapitre 5

Dépliement réductif de la HSA : prédiction de la structure réduite

Sommaire

5.1 Introduction 82

5.2 Étude de convergence de la méthode d'analyse 82

5.3 Recherche de conformations stables et étude globale du dépliement 84

5.3.1 Dépliement des sous-domaines 86

5.3.2 Dépliement des différents types d'hélices α 87

5.3.3 Prédiction de la structure de la HSA réduite 88

5.4 Effet du dépliement sur les propriétés de complexation de HSA$^{(R)}$ 89

5.5 Conclusions 91

5.1 Introduction

Dans ce chapitre, nous nous intéressons aux effets du dépliement réductif de la HSA sur ses propriétés de transport de médicaments. La première étape du travail consiste alors à prédire la structure de la HSA réduite. Il est évident qu'il est impossible de simuler le processus complet de dépliement à l'aide des moyens de calcul actuels. Nous avons vu précédemment que le recours à des simulations à des températures élevées permettait d'accélérer le processus de dépliement sans pour autant en modifier le chemin de dépliement [58]. Le chemin de dépliement de la HSA réduite est alors présenté comme une série de conformères partiellement dépliés et ordonnés selon leur degré de dépliement. Les structures partiellement dépliées ont été obtenues en effectuant des simulations de DM de 160ns chacune à quatre températures différentes : 350 K, 375 K, 400 K et 425 K. Les simulations sont effectuées avec le champ de force OPLS-AA. Chaque trajectoire est ensuite soumise à une analyse en composantes principales [35–38, 65] afin d'identifier les principaux mouvements collectifs de la protéine et explorer le regroupement de structures donnant lieu à des conformères. Les différentes conformations ainsi identifiées ont été classées en tenant compte de leur contenu en hélice-α par rapport à la structure de référence. Une structure théorique de la HSA réduite à l'état d'équilibre (à température ambiante) est enfin proposée en comparant les paramètres physiques des conformères simulés avec les propriétés de la structure expérimentale.

5.2 Étude de convergence de la méthode d'analyse

Le dépliement de la protéine au long des quatre trajectoires de simulation est illustré sur la Figure 5.1 en représentant le RMSD des positions atomiques des atomes de carbone α par rapport à la structure de référence. Dans tous les cas, nous observons une augmentation rapide du RMSD qui se produit dans les 10 premières ns suivie d'un hausse beaucoup plus lente durant environ 80ns. Les RMSD apparaîssent stabilisés sur la seconde partie de chaque trajectoire.

Une évolution similaire est observée pour la fraction d'hélice-α de la structure protéique (Figure 5.2). Les pourcentages d'équilibre en hélices-α sont respectivement de 61 %, 62 %, 55 % et 42 % pour les trajectoires simulées à 350, 375, 400 et 425 K. Pour comparaison, la valeur correspondante à la protéine native est de 70% [12]. Ces résultats

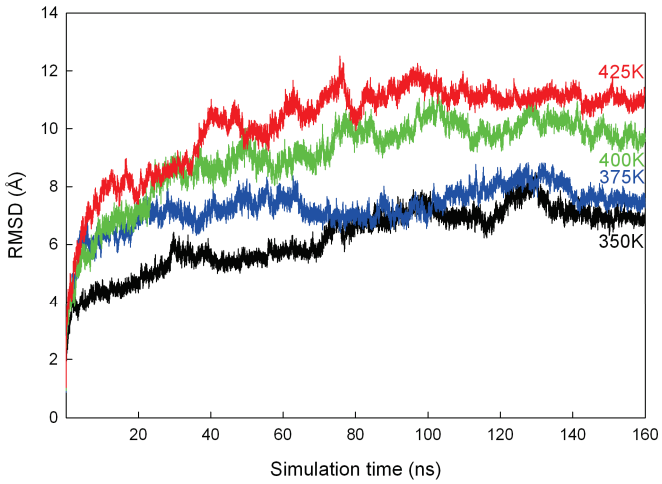

Figure 5.1 – Evolution du RMSD (Å) des atomes de C_α par rapport à la structure de référence au long des quatre trajectoires de DM

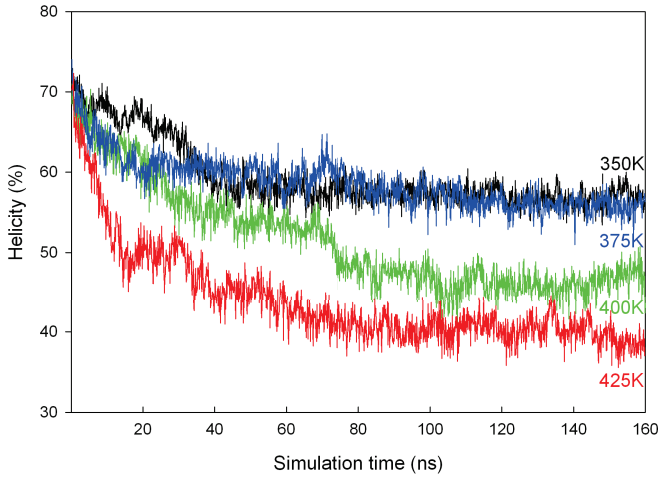

Figure 5.2 – Evolution de l'hélicité (%) au long des quatre trajectoires de DM

correspondent donc à un dépliement moyen de la structure secondaire. Les valeurs les plus faibles se rapprochent des données expérimentales rapportées pour la structure d'équilibre de la HSA réduite [27, 94].

Afin de déterminer les principales étapes impliquées dans le dépliement de la protéine, une analyse en composantes principales a été effectuée pour chaque trajectoire. Le calcul a été effectué en tenant compte uniquement des atomes de C_α.

Pour chaque analyse, la convergence de l'échantillonnage a été vérifiée en calculant le contenu en cosinus des cinq premières PC. Les valeurs données dans le Tableau 5.1 sont en général inférieures à 0.5.

	350K	375K	400K	425K
PC1	0.92	0.75	0.90	0.78
PC2	0.45	0.64	0.41	0.51
PC3	0.32	0.14	0.27	0.58
PC4	0.11	0.12	0.27	0.28
PC5	0.00	0.39	0.51	0.00

Tableau 5.1 – Contenu en cosinus des cinq premières PC dérivées des trajectoires de DM simulées à 350, 375, 400 et 425 K.

Une exception systématique est présente pour la première composante principale, PC1, dont le contenu en cosinus est largement supérieur à 0.5 pour les quatre trajectoires. Ce résultat suggère que, dans les quatre cas de température, le mouvement protéique le plus lent n'est pas suffisament échantillonné. Le problème de convergence actuel est très probablement lié à la spécificité du dépliement de la protéine qui peut impliquer une relaxation par diffusion importante [60, 99]. Il est ainsi possible que les modes les plus lents détectés (PC1) soient liés à cette détente.

5.3 Recherche de conformations stables et étude globale du dépliement

En se basant sur la corrélation entre les trois premières PC, différents groupes de conformations (clusters) ont été identifiés pour chaque trajectoire de simulation (Tableau 5.2).

350K	375K	400K	425K
C1a	C1b	C1c	C1d
C2a	C2b	C2c	C2d
C3a	C3b	C3c	C3d
C4a	C4b	C4c	C4d
		C5c	

Tableau 5.2 – Conformères identifiés au long des trajectoires de DM

Population (%)	350K (a)	375K (b)	400K (c)	425K (d)
C1	43.6	26.5	41.7	42.7
C2	13.1	14.3	15.6	17.1
C3	12.0	12.2	13.5	9.5
C4	4.0	6.0	1.5	6.5
C5			0.6	

Tableau 5.3 – Population (%) des différents conformères identifiés au long des trajectoires de DM

La structure moyenne d'un cluster (tel que prévu par CARMA) définit un conformère de la protéine présentant une stabilité structurale relative. Une série de 17 conformères décrivant le processus de dépliement a ainsi été obtenue. Les 17 conformères ont ensuite été classés en fonction de leur hélicité : C3a, C2a, C4c, C2c, C4b, C4a, C1a, C1b, C3c, C2b, C3b, C5c, C2d, C1c, C3d, C1d, C4d (Tableau 5.3).

	350K (a)	375K (b)	400K (c)	425K (d)
C1	59.41	59.14	50.22	42.44
C2	68.73	56.69	66.17	52.51
C3	69.94	55.72	57.87	47.99
C4	60.57	62.00	66.85	42.18
C5	-	-	55.22	-

Tableau 5.4 – Hélicité des différents conformères identifiés au long des trajectoires de DM

Une illustration du processus de dépliement est présenté dans la Figure 5.3 où huit conformères de cette série sont représentés. La séquence de la Figure 5.3 indique un dépliement progressif des hélices-α associé à un désordre de plus en plus élevé de la structure tertiaire de la protéine. On peut observer également des différences significatives entre la stabilité des sous-domaines. Les sous-domaines IA (bleu) et IIIB (rouge) sont les plus touchés par le dépliement alors que les sous-domaines IIA (vert) et IIB (jaune) sont les plus stables. Une comparaison plus quantitative de la stabilité des sous-domaines est obtenue en représentant l'évolution de la fraction d'hélices-α par sous-domaine.

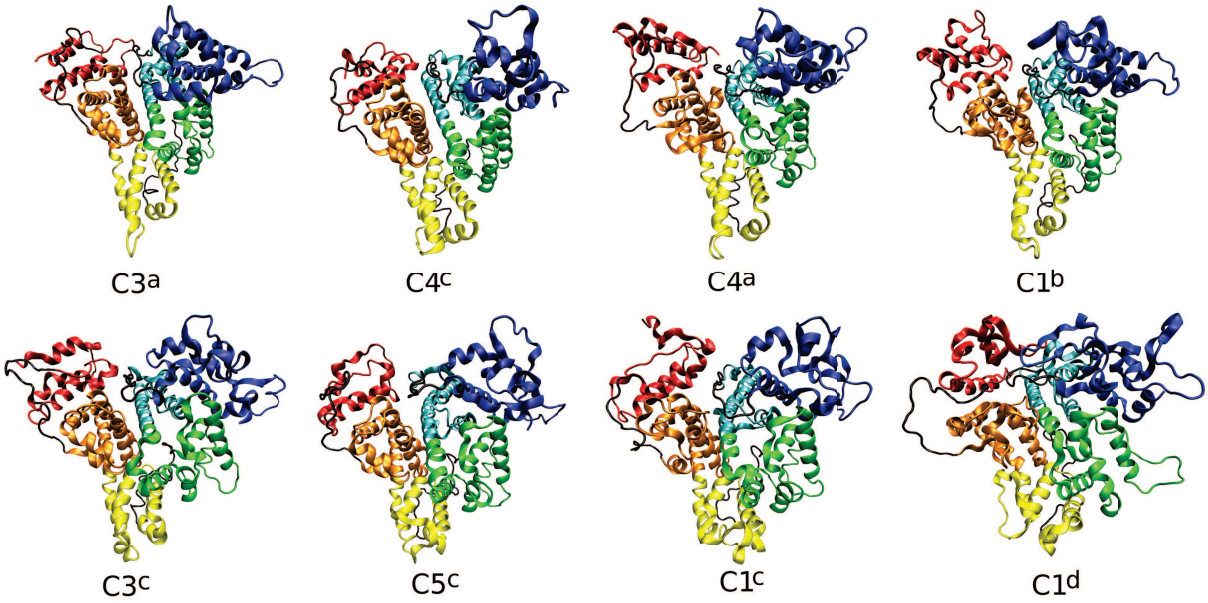

Figure 5.3 – Dépliement de la HSA réduite : séquence de conformères extraits des quatre simulations de DM et ordonnés en fonction de leur hélicité décroissante.

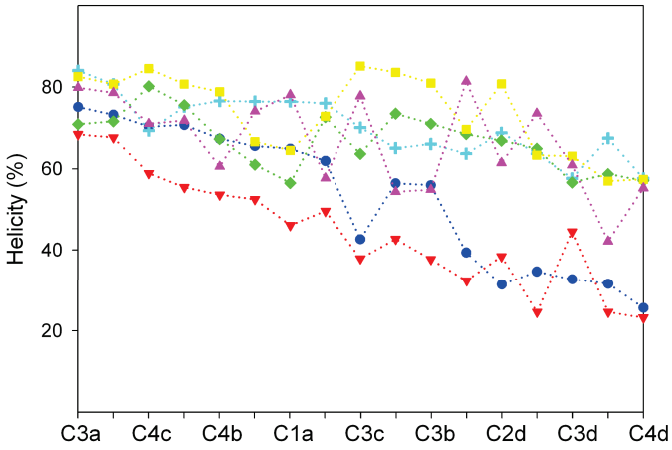

Figure 5.4 – Hélicité (%) par sous-domaine des 17 conformères (IA-bleu, IB-cyan, IIA-vert, IIB-jaune, IIIA-rose et IIIB-rouge)

5.3.1 Dépliement des sous-domaines

Une progression générale et non continue du dépliement est observée pour tous les sous-domaines. Cependant une séparation nette apparaît entre les courbes de dépliement des

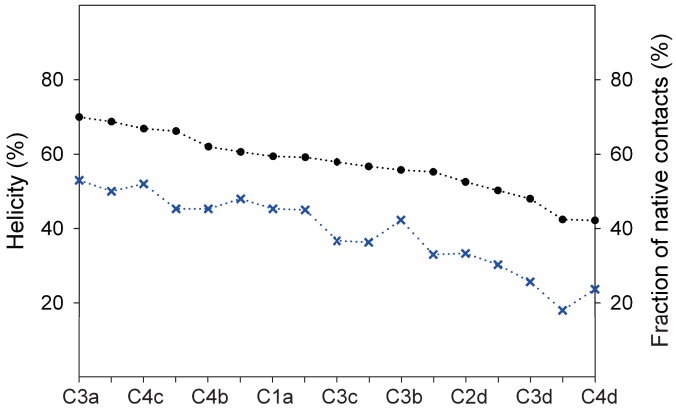

Figure 5.5 – Fraction des contacts natifs tertiaires (bleu) par rapport à l'hélicité globale (noir) pour les 17 conformères

sous-domaines IA et IIIB et les quatre autres (Figure 5.4). Par ailleurs, la corrélation entre le dépliement des structures secondaire et tertiaire de la protéine est mieux illustrée en comparant les variations de l'hélicité avec celle des contacts natifs (Figure 5.5).

Deux courbes similaires sont obtenues indiquant une dépendance linéaire de la fraction des contacts natifs tertiaires par rapport à l'hélicité.

5.3.2 Dépliement des différents types d'hélices α

Les 20 hélices de la HSA (Figure 1.1) peuvent sans ambiguïté être divisées en deux groupes : un groupe pour les hélices longues et un autre pour les hélices courtes. Cette séparation offre une bonne occasion d'examiner une possible relation entre la longueur d'une hélice et sa stabilité. Les changements d'hélicité dans les deux groupes au long de la séquence de conformères sont présentés dans la Figure 5.6.

Les deux courbes montrent une nette différence avec un dépliement plus rapide des hélices courtes. Des détails supplémentaires sur le dépliement des hélices sont obtenus en comparant l'évolution d'hélices longues non équivalentes. En effet, il est possible de classer les hélices longues en deux catégories : chaque sous-domaine de la protéine contient trois longues hélices : l'hélice H2, qui est un peu plus courte, et les hélices H3 et H4 qui sont tenues ensemble par un pont S—S (Figure 1.2). Le dépliement des deux types d'hélices est représenté dans la Figure 5.7 pour les sous-domaines les plus instables : IA et IIIB.

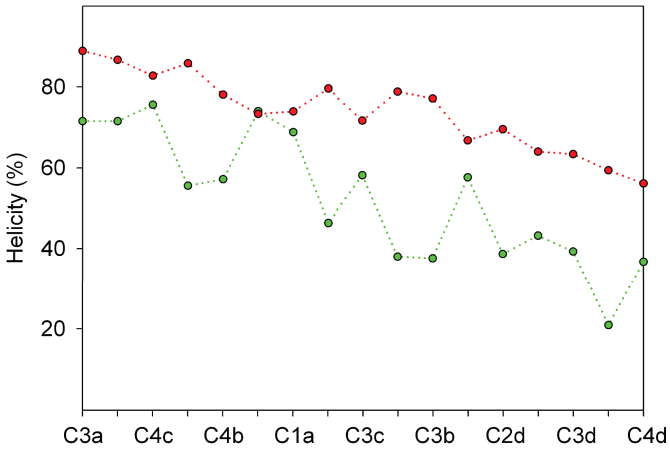

Figure 5.6 – Dépliement de la HSA réduite : hélicité des hélices longues (rouge) par rapport à celle des hélices courtes (vert)

Fait intéressant, dans les deux cas, l'hélice H2 présente une remarquable stabilité, tandis que les hélices H3 et H4 sont complètement dépliées pour les derniers conformères de la séquence.

5.3.3 Prédiction de la structure de la HSA réduite

Les changements structuraux décrits dans les Figures 5.4-5.7 apparaissent comme continu en dépit du fait que les 17 conformères proviennent de quatre simulations réalisées à quatre températures différentes. Ce résultat confirme l'hypothèse selon laquelle le dépliement de la HSA se produit comme une succession stable d'événements qui ne sont pas significativement dépendants de la température de simulation. Un ordre stable des événements qui se déroulent en fonction de la température de simulation a déjà été rapporté par Daggett et al. [58] dans le cas du dépliement d'un inhibiteur de la chymotrypsine 2. Ils ont conclu que la température de simulation modifie l'échelle de temps, mais pas le chemin de dépliement. Toutefois, étant donné la complexité du dépliement des surfaces d'énergie libres [60], cette affirmation est loin d'être démontrée théoriquement et les limites de sa validité doivent être soigneusement examinées. Dans le cas de la HSA, en se basant sur la présente analyse, cette affirmation peut être acceptée au moins d'un point de vue pratique. Un résultat relatif à cette analyse est la confirmation que l'hélicité est une bonne coordonnée pour le dépliement. Cela a été démontré expérimentalement dans des études

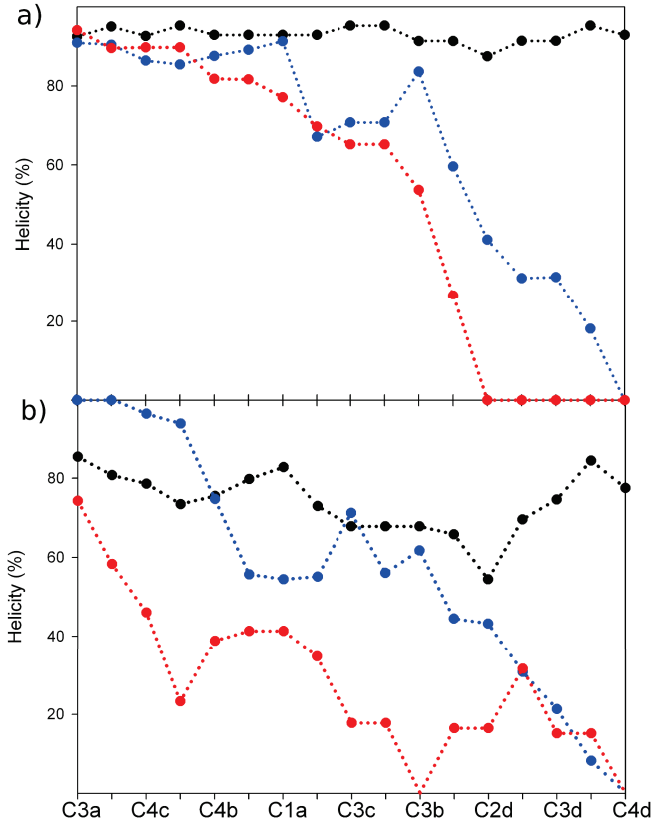

Figure 5.7 – Dépliement de trois hélices longues de type H2 (noir), H3 (bleu) et H4 (rouge) : a) dans le sous-domaine IA et b) dans le sous domaine IIIB. Voir la Figure 1.2 pour l'emplacement des hélices.

de dépliement de la HSA [100, 101]. En se basant sur des résultats de CD [94], on en déduit que la fraction d'hélice-α de la HSA réduite à l'état d'équilibre est d'environ 60% de celle de la forme native. Par conséquent, les conformères C1d et C4d qui ont une hélicité équivalente (Tableau 5.2 et 5.3) apparaissent comme de bons candidats pour représenter l'état d'équilibre déplié de la HSA réduite. Pour être plus précis, nous suggérons que les conformères appartenant à la variété des conformations de la HSA réduite présentent des similarités importantes avec C1d et C4d : une structure globale qui reste compacte, un dépliement plus avancé des sous-domaines IA et IIIB avec les hélices de type H3 et H4 presque complètement dépliées et une structure tertiaire ne conservant que près d'un quart des contacts natifs.

5.4 Effet du dépliement sur les propriétés de complexation de HSA$^{(R)}$

D'un point de vue fonctionnel, il est intéressant d'évaluer l'effet du dépliement sur les propriétés de complexation de la HSA réduite. Une telle évaluation nécessite de réaliser des simulations de complexation de ligand à la fois sur la protéine native et réduite. Cependant, un aperçu qualitatif peut être obtenu par une simple comparaison structurale. Les cavités structurales associées aux deux sites principaux de liaison ont été déterminées à l'aide du programme Fpocket [102]. Leurs formes et leurs emplacements dans la protéine native sont représentés sur la Figure 5.8a.

A titre comparatif, une analyse identique est effectuée sur le conformère C4d. En dépit d'importantes différences dans sa structure tertiaire, le conformère C4d conserve encore ces deux cavités (Figure 5.8b). Cependant, leurs formes et leurs volumes diffèrent des cavités de la structure native. Ils ont très probablement perdu la sélectivité de liaison des sites d'origine. En général, les conformères repliés possèdent encore plusieurs cavités, et celles ci pourraient être actives pour la liaison de médicaments, mais leurs volumes et leurs emplacements sont largement modifiés par rapport à la structure native. Une comparaison entre les quatre plus importantes cavités calculées pour la protéine native et le conformère C4d est présentée dans la Figure 5.9.

Cette comparaison confirme l'hypothèse que la HSA réduite conserve encore une activité de liaison importante mais la spécificité de liaison de la protéine native est clairement perdue.

5.5 Conclusions

La simulation du dépliement de la HSA réduite est étudiée en utilisant différentes températures de simulation. L'analyse PCA donne une séquence de conformères dépliés qui est continue par rapport à plusieurs paramètres, comme l'hélicité totale ou locale ou bien encore la fraction des contacts natifs tertiaires. Ces résultats suggèrent que le chemin de dépliement de HSA$^{(R)}$ est indépendant de la température de simulation. Comme le dépliement local est bien corrélé avec la fraction totale en hélice-α, ces résultats indiquent également que ce dernier paramètre est une bonne coordonnée pour le dépliement.

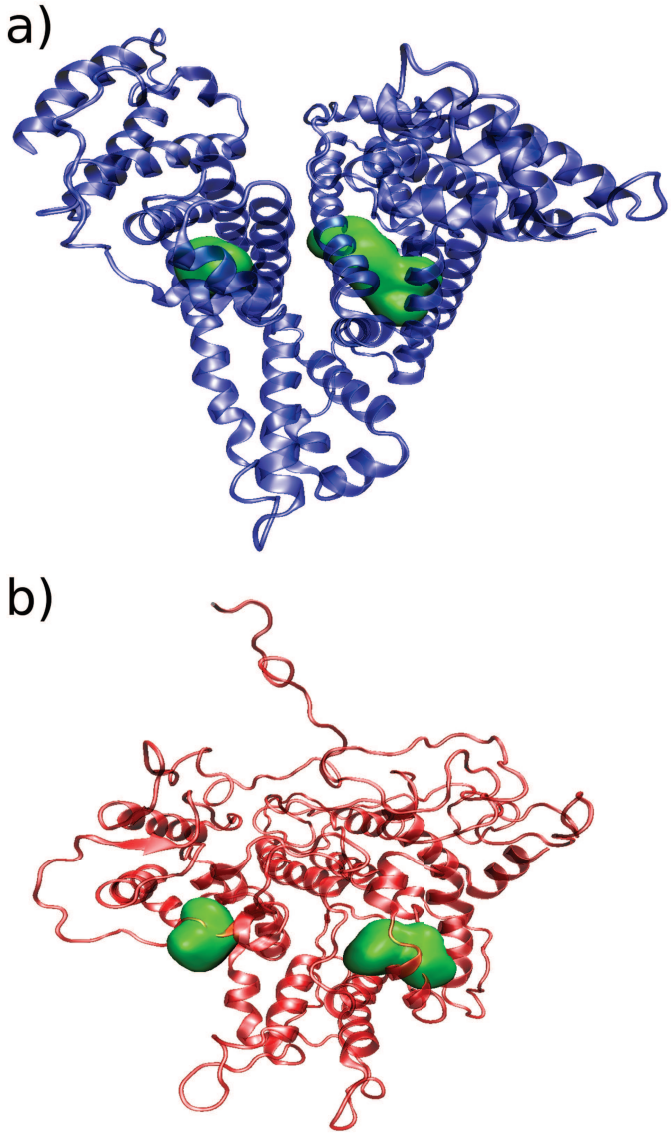

Figure 5.8 – Représentation des deux sites principaux de complexation (en vert) à travers la structure native de la protéine (a) et le conformère C4d (b).

Un point très important à clarifier est de comprendre comment la reproductibilité du chemin de dépliement est liée à la spécificité structurale de la protéine. En effet, une telle reproductibilité pourrait être la conséquence d'une hiérarchie bien établie dans la stabilité des hélices formant la majeure partie de la structure protéique.

Cette étude nous a permis de prédire la structure de la HSA réduite. Une analyse qualitative basée sur l'évolution des cavités de la protéine donne une première idée sur l'effet du dépliement sur les propriétés de transport de la HSA. Cependant des études de

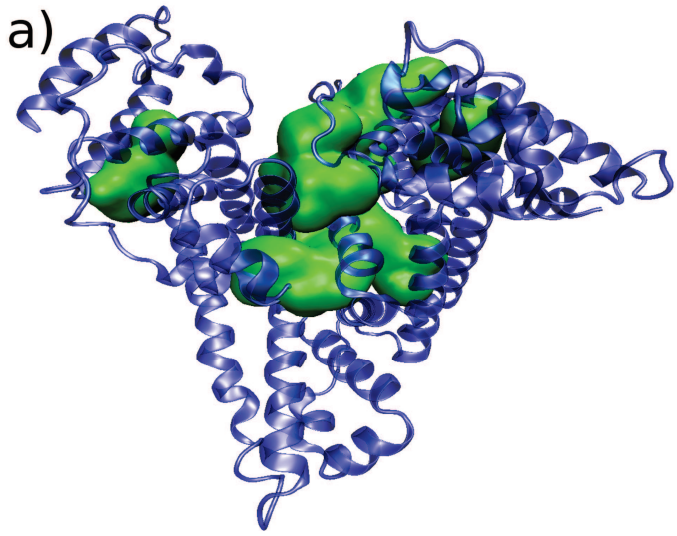

Figure 5.9 – Représentation des quatre plus importantes cavités (en vert) de la structure native (a) et du conformère C4d(b).

simulation directe de la complexation de la HSA réduite avec différents médicaments sont nécessaires pour une description quantitative de ces effets.

Chapitre 6

Conclusions

Dans ce travail de thèse, nous avons réalisé une étude théorique de la dynamique conformationnelle de la HSA et de ses sites de complexation. Notre étude a aussi été orientée sur le rôle structural des ponts disulfure dans le maintien de la stabilité de la structure native de la protéine. La compréhension du processus de dépliement lors de la réduction des ponts S—S a également été traitée.

La méthodologie utilisée repose sur le couplage de la dynamique moléculaire avec l'analyse par composantes principales. La convergence des trajectoires de dynamique moléculaire et des analyses PCA permet d'affirmer qu'elle a été utilisée avec succès et que les résultats obtenus ont une réalité physique pour l'échelle de temps simulé.

La PCA appliquée à trois trajectoires de 200 ns permet d'observer des transitions conformationnelles naturelles qu'on retrouve lors de la complexation de ligands et plus particulièrement celle observée expérimentalement lors de la liaison du myristate à la HSA. Cette observation valide l'hypothèse qu'un changement conformationnel favorable à la liaison d'un ligand peut précéder celle-ci. Une analyse plus fine des mouvements collectifs des sous-domaines IIA et IIIA montre l'existence de transitions conformationnelles délocalisées pouvant être à l'origine d'effets allostériques entre ces deux sites de liaison.

Parallélement, les simulations de dynamique moléculaire prédisent trois effets principaux de la réduction des ponts S—S sur la structure protéique. La déstabilisation de la structure secondaire (perte d'hélicité), la déformation de la structure tertiaire (perte des contacts natifs) et la légère augmentation du volume de la protéine (augmentation du rayon de gyration) sont ces observables. Les deux champs de force testés donnent des résultats quasiment similaires qui sont en accord avec les données expérimentales connues. Cependant le champ de force OPLS-AA semble reproduire au mieux le comportement naturel des mouvements protéiques. Il a été montré que les thiols sont maintenus en contact dans des puits de potentiel spécifiques. Des transitions conformationnelles situées dans les résidus voisins des cystéines impliquées dans le pont S—S sont à l'origine de leur séparation. Les résultats obtenus suggèrent également que le repliement oxydatif se déroule en deux étapes : une première où le pont S—S se réforme à la suite d'un simple processus diffusif et une seconde où la protéine se replie et retrouve sa conformation native.

Notre étude suggère que, pour le protocole de simulation utilisé ici, le chemin de dépliement de la HSA réduite ne dépend pas de la température de simulation. Cela nous a permis de prédire sa structure d'équilibre. Il est alors apparu que la réduction des ponts S—S modifie de façon considérable la structure des sites de complexation de la protéine et

va ainsi perturber sa fonction de transport. En ce point, la structure réduite d'équilibre ainsi prédite nous a permis d'analyser qualitativement l'évolution des cavités de la protéine et de se faire une première idée de l'effet du dépliement sur les propriétés de transport de la HSA.

Des simulations de complexation sur la structure réduite prédite permettraient de décrire de manière quantitative les effets du dépliement sur les propriétés de liaison de la HSA et pourraient être une perspective intéressante dans la poursuite de ce travail. Il est évidemment envisageable de réutiliser la méthodologie employée ici sur d'autres protéines d'intérêt afin d'en extraire leurs propriétés dynamiques ou leur mécanisme de dépliement/repliement.

Bibliographie

[1] Ananyo A Bhattacharya, Tim Grüne, and Stephen Curry. Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *Journal of molecular biology*, 303(5) :721–732, 2000.

[2] John C Kendrew, G Bodo, Howard M Dintzis, RG Parrish, H Wyckoff, and DC Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610) :662–666, 1958.

[3] Frances Allen, G Almasi, Wanda Andreoni, D Beece, Bruce J. Berne, A Bright, Jose Brunheroto, Calin Cascaval, J Castanos, Paul Coteus, et al. Blue gene : a vision for protein science using a petaflop supercomputer. *IBM Systems Journal*, 40(2) : 310–327, 2001.

[4] Michael Shirts, Vijay S Pande, et al. Screen savers of the world unite. *COMPUTING*, 10 :43, 2006.

[5] Christophe Chipot and Andrew Pohorille. *Free energy calculations : theory and applications in chemistry and biology*, volume 86. Springer, 2007.

[6] Jamie Ghuman, Patricia A Zunszain, Isabelle Petitpas, Ananyo A Bhattacharya, Masaki Otagiri, and Stephen Curry. Structural basis of the drug-binding specificity of human serum albumin. *Journal of molecular biology*, 353(1) :38–52, 2005.

[7] Paolo Ascenzi, Alessio Bocedi, Stefania Notari, Gabriella Fanali, Riccardo Fesce, and Mauro Fasano. Allosteric modulation of drug binding to human serum albumin. *Mini reviews in medicinal chemistry*, 6(4) :483–489, 2006.

[8] Roberto Artali, Gabriella Bombieri, Luisella Calabi, and Antonio Del Pra. A molecular dynamics study of human serum albumin binding sites. *Il Farmaco*, 60 (6) :485–495, 2005.

[9] JR Brown. Structure of serum albumin : disulfide bridges. In *Fed. Proc*, volume 33, page 1389, 1974.

[10] MA Saber, P Stöckbauer, L Morávek, and B Meloun. Disulfide bonds in human serum albumin. *Collection of Czechoslovak Chemical Communications*, 42(2) :564–579, 1977.

[11] Xiao Min He and Daniel C Carter. Atomic structure and chemistry of human serum albumin. 1992.

[12] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, and K. Kobayashi. Crystal structure of human serum albumin at 2.5 å resolution. *Protein engineering*, 12(6) :439–446, 1999.

[13] Gabriella Fanali, Riccardo Fesce, Cristina Agrati, Paolo Ascenzi, and Mauro Fasano. Allosteric modulation of myristate and mn (iii) heme binding to human serum albumin. *FEBS Journal*, 272(18) :4672–4683, 2005.

[14] C Nick Pace, GR Grimsley, JA Thomson, and BJ Barnett. Conformational stability and activity of ribonuclease t1 with zero, one, and two intact disulfide bonds. *Journal of Biological Chemistry*, 263(24) :11820–11825, 1988.

[15] Daniela C Vaz, J Rui Rodrigues, Walter Sebald, Christopher M Dobson, and Rui MM Brito. Enthalpic and entropic contributions mediate the role of disulfide bonds on the conformational stability of interleukin-4. *Protein science*, 15(1) :33–44, 2006.

[16] William J Wedemeyer, Ervin Welker, Mahesh Narayan, and Harold A Scheraga. Disulfide bonds and protein folding. *Biochemistry*, 39(15) :4207–4216, 2000.

[17] David P Goldenberg, Lynne S Bekeart, Daniel A Laheru, and Jia Dong Zhou. Probing the determinants of disulfide stability in native pancreatic trypsin inhibitor. *Biochemistry*, 32(11) :2835–2844, 1993.

[18] Martin Karlsson, Lars-Göran Mårtensson, Carin Karlsson, and Uno Carlsson. Denaturant-assisted formation of a stabilizing disulfide bridge from engineered cysteines in nonideal conformations. *Biochemistry*, 44(9) :3487–3493, 2005.

[19] Fabrice Fleury, Anatoli Ianoul, Maurice Berjot, Alexei Feofanov, Alain JP Alix, and Igor Nabiev. Camptothecin-binding site in human serum albumin and protein transformations induced by drug binding. *FEBS letters*, 411(2) :215–220, 1997.

[20] Claire Dufour and Olivier Dangles. Flavonoid–serum albumin complexation : determination of binding constants and binding sites by fluorescence spectroscopy. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1721(1) :164–173, 2005.

[21] JF Neault and HA Tajmir-Riahi. Interaction of cisplatin with human serum albumin. drug binding mode and protein secondary structure. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1384(1) :153–159, 1998.

[22] Naoko Ohta, Danni Chen, Shigekazu Ito, Tomomichi Futo, Toshihisa Yotsuyanagi, and Ken Ikeda. Effect of< i> trans</i>-diamminedichloroplatinum (ii) on human serum albumin : Conformational changes through partial disulfide bond cleavage. *International journal of pharmaceutics*, 118(1) :85–93, 1995.

[23] Ilona Fitos, Julia Visy, and Julianna Kardos. Stereoselective kinetics of warfarin binding to human serum albumin : effect of an allosteric interaction. *Chirality*, 14 (5) :442–448, 2002.

[24] Shin-ichi Fujiwara and Takashi Amisaki. Molecular dynamics study of conformational changes in human serum albumin by binding of fatty acids. *Proteins : Structure, Function, and Bioinformatics*, 64(3) :730–739, 2006.

[25] TR Cuya Guizado, SRW Louro, and C Anteneodo. Dynamics of heme complexed with human serum albumin : a theoretical approach. *European Biophysics Journal*, 41(12) :1033–1042, 2012.

[26] Osama K Abou-Zied, Najla Al-Lawatia, Marcus Elstner, and Thomas B Steinbrecher. Binding of hydroxyquinoline probes to human serum albumin : Combining molecular modeling and forster's resonance energy transfer spectroscopy to understand flexible ligand binding. *The Journal of Physical Chemistry B*, 117(4) :1062–1074, 2013.

[27] Catalina David, Sarah Foley, Christophe Mavon, and Mironel Enescu. Reductive unfolding of serum albumins uncovered by raman spectroscopy. *Biopolymers*, 89(7) : 623–634, 2008.

[28] Marc Langenhof, Susanna SJ Leong, Leonard K Pattenden, and Anton PJ Middelberg. Controlled oxidative protein refolding using an ion-exchange column. *Journal of Chromatography A*, 1069(2) :195–201, 2005.

[29] K Reddy, Ravi Charan, Hauke Lilie, Rainer Rudolph, and Christian Lange. l-arginine increases the solubility of unfolded species of hen egg white lysozyme. *Protein Science*, 14(4) :929–935, 2005.

[30] Steven J BURTON, Alan V QUIRK, and P Carolyn WOOD. Refolding human serum albumin at relatively high protein concentration. *European Journal of Biochemistry*, 179(2) :379–387, 1989.

[31] Floriana Vinci, Margherita Ruoppolo, Piero Pucci, Robert B Freedman, and Gennaro Marino. Early intermediates in the pdi-assisted folding of ribonuclease a. *Protein Science*, 9(3) :525–535, 2000.

[32] Surendra K Nayak, Dharmendar Rathore, and Janendra K Batra. Role of individual cysteine residues and disulfide bonds in the structure and function of aspergillus ribonucleolytic toxin restrictocin. *Biochemistry*, 38(31) :10052–10058, 1999.

[33] Valerie Daggett and Michael Levitt. A model of the molten globule state from molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 89(11) :5142–5146, 1992.

[34] Yuri V Griko. Energetic basis of structural stability in the molten globule state : α-lactalbumin. *Journal of molecular biology*, 297(5) :1259–1268, 2000.

[35] Andrea Amadei, Antonius Linssen, and Herman JC Berendsen. Essential dynamics of proteins. *Proteins : Structure, Function, and Bioinformatics*, 17(4) :412–425, 1993.

[36] Angel E García and James G Harman. Simulations of crp :(camp) 2 in noncrystalline environments show a subunit transition from the open to the closed conformation. *Protein science*, 5(1) :62–71, 1996.

[37] Akio Kitao and Nobuhiro Go. Investigating protein dynamics in collective coordinate space. *Current opinion in structural biology*, 9(2) :164–169, 1999.

[38] Berk Hess. Similarities between principal components of protein dynamics and random diffusion. *Physical Review E*, 62(6) :8438, 2000.

[39] Hongfeng Lou and Robert I Cukier. Molecular dynamics of apo-adenylate kinase : a principal component analysis. *The Journal of Physical Chemistry B*, 110(25) : 12796–12808, 2006.

[40] Oliver F Lange and Helmut Grubmüller. Can principal components yield a dimension reduced description of protein dynamics on long time scales ? *The Journal of Physical Chemistry B*, 110(45) :22842–22852, 2006.

[41] Gia G Maisuradze, Adam Liwo, and Harold A Scheraga. Principal component analysis for protein folding dynamics. *Journal of molecular biology*, 385(1) :312–329, 2009.

[42] Stanislaw T Wlodek, Terry W Clark, L Ridgway Scott, and J Andrew McCammon. Molecular dynamics of acetylcholinesterase dimer complexed with tacrine. *Journal of the American Chemical Society*, 119(40) :9513–9522, 1997.

[43] Caterina Arcangeli, Anna Rita Bizzarri, and Salvatore Cannistraro. Molecular dynamics simulation and essential dynamics study of mutated plastocyanin : structural, dynamical and functional effects of a disulfide bridge insertion at the protein surface. *Biophysical Chemistry*, 92(3) :183–199, 2001.

[44] Lucia D Creveld, Andrea Amadei, René C van Schaik, Henri AM Pepermans, Jacob de Vlieg, and Herman JC Berendsen. Identification of functional and unfolding motions of cutinase as obtained from molecular dynamics computer simulations. *Proteins : Structure, Function, and Bioinformatics*, 33(2) :253–264, 1998.

[45] Alessandro Grottesi, Marc-Antoine Ceruso, Alfredo Colosimo, and Alfredo Di Nola. Molecular dynamics study of a hyperthermophilic and a mesophilic rubredoxin. *Proteins : Structure, Function, and Bioinformatics*, 46(3) :287–294, 2002.

[46] Charles L Brooks. Protein and peptide folding explored with molecular simulations. *Accounts of chemical research*, 35(6) :447–454, 2002.

[47] Mingfeng Yang, Boyan Yordanov, Yaakov Levy, Rafael Brüschweiler, and Shuanghong Huo. The sequence-dependent unfolding pathway plays a critical role in the amyloidogenicity of transthyretin. *Biochemistry*, 45(39) :11992–12002, 2006.

[48] Ting Wang and Rebecca C Wade. On the use of elevated temperature in simulations to study protein unfolding mechanisms. *Journal of Chemical Theory and Computation*, 3(4) :1476–1483, 2007.

[49] Maria Eleftheriou, Robert S Germain, Ajay K Royyuru, and Ruhong Zhou. Thermal denaturing of mutant lysozyme with both the oplsaa and the charmm force fields. *Journal of the American Chemical Society*, 128(41) :13388–13395, 2006.

[50] Dirk Matthes and Bert L De Groot. Secondary structure propensities in peptide folding simulations : a systematic comparison of molecular mechanics interaction schemes. *Biophysical journal*, 97(2) :599–608, 2009.

[51] Marek Cieplak and Trinh Xuan Hoang. Universality classes in folding times of proteins. *Biophysical journal*, 84(1) :475–488, 2003.

[52] David AC Beck and Valerie Daggett. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods*, 34(1) :112–120, 2004.

[53] Lorna J Smith. Computational methods for generating models of denatured and partially folded proteins. *Methods*, 34(1) :144–150, 2004.

[54] Isabella Daidone, Andrea Amadei, Danilo Roccatano, and Alfredo Di Nola. Molecular dynamics simulation of protein folding by essential dynamics sampling : folding landscape of horse heart cytochrome c. *Biophysical journal*, 85(5) :2865–2871, 2003.

[55] Donald Hamelberg, John Mongan, and J Andrew McCammon. Accelerated molecular dynamics : a promising and efficient simulation method for biomolecules. *The Journal of chemical physics*, 120 :11919, 2004.

[56] Steven L Kazmirski and Valerie Daggett. Simulations of the structural and dynamical properties of denatured proteins : the "molten coil" state of bovine pancreatic trypsin inhibitor. *Journal of molecular biology*, 277(2) :487–506, 1998.

[57] Kam-Bo Wong, Jane Clarke, Chris J Bond, José Luis Neira, Stefan Freund, Alan R Fersht, and Valerie Daggett. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *Journal of molecular biology*, 296(5) :1257–1282, 2000.

[58] Ryan Day, Brian J Bennion, Sihyun Ham, and Valerie Daggett. Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *Journal of molecular biology*, 322(1) :189–203, 2002.

[59] Daan Thorn Leeson, Feng Gai, Hector M Rodriguez, Lydia M Gregoret, and R Brian Dyer. Protein folding and unfolding on a complex energy landscape. *Proceedings of the National Academy of Sciences*, 97(6) :2527–2532, 2000.

[60] Ginka S Buchner, Ronan D Murphy, Nicolae-Viorel Buchete, and Jan Kubelka. Dynamics of protein folding : Probing the kinetic network of folding–unfolding transitions with experiment and theory. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1814(8) :1001–1020, 2011.

[61] Kam-Bo Wong, Jane Clarke, Chris J Bond, José Luis Neira, Stefan Freund, Alan R Fersht, and Valerie Daggett. Towards a complete description of the structural

and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *Journal of molecular biology*, 296(5) :1257–1282, 2000.

[62] Seongeun Yang and Minhaeng Cho. Thermal denaturation of polyalanine peptide in water by molecular dynamics simulations and theoretical prediction of infrared spectra : helix-coil transition kinetics. *The Journal of Physical Chemistry B*, 111 (3) :605–617, 2007.

[63] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45) :11225–11236, 1996.

[64] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117 (19) :5179–5197, 1995.

[65] Berk Hess. Convergence of sampling in protein simulations. *Physical Review E*, 65 (3) :031910, 2002.

[66] BJ Alder and TEf Wainwright. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, 27(5) :1208–1209, 1957.

[67] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79 :926, 1983.

[68] Laxmikant Kalé, Robert Skeel, Milind Bhandarkar, Robert Brunner, Attila Gursoy, Neal Krawetz, James Phillips, Aritomo Shinozaki, Krishnan Varadarajan, and Klaus Schulten. Namd2 : Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 151(1) :283–312, 1999.

[69] W.F. van Gunsteren and H.J.C. Berendsen. *Molecular Dynamics and Protein Structure*. Number pp. 5-14. Western Springs, 1985.

[70] Charles L Brooks III, B Montgomery Pettitt, and Martin Karplus. Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *The Journal of chemical physics*, 83 :5897, 1985.

[71] Wilfred F van Gunsteren, Paul K Weiner, and Anthony J Wilkinson. Computer simulation of biomolecular systems. *Theoretical and experimental applications*, 2 : 3–36, 1989.

[72] H Schreiber and O Steinhauser. Molecular dynamics studies of solvated polypeptides : why the cut-off scheme does not work. *Chemical physics*, 168(1) :75–89, 1992.

[73] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald : An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98 :10089, 1993.

[74] Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics*, 72(4) :2384, 1980.

[75] Scott E Feller, Yuhong Zhang, Richard W Pastor, and Bernard R Brooks. Constant pressure molecular dynamics simulation : the langevin piston method. *The Journal of Chemical Physics*, 103 :4613, 1995.

[76] Michael Levitt, Miriam Hirshberg, Ruth Sharon, Keith E Laidig, and Valerie Daggett. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *The Journal of Physical Chemistry B*, 101 (25) :5051–5061, 1997.

[77] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of< i> n</i>-alkanes. *Journal of Computational Physics*, 23 (3) :327–341, 1977.

[78] Dmitrij Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins : Structure, Function, and Bioinformatics*, 23(4) :566–579, 1995.

[79] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd : visual molecular dynamics. *Journal of molecular graphics*, 14(1) :33–38, 1996.

[80] Mihaela E Sardiu, Margaret S Cheung, and Yi-Kuo Yu. Cysteine-cysteine contact preference leads to target-focusing in protein folding. *Biophysical journal*, 93(3) : 938–951, 2007.

[81] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation : Umbrella sampling. *Journal of Computational Physics*, 23(2) :187–199, 1977.

[82] Christian Bartels and Martin Karplus. Multidimensional adaptive umbrella sampling : applications to main chain and side chain peptide conformations. *Journal of computational chemistry*, 18(12) :1450–1462, 1997.

[83] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8) : 1011–1021, 1992.

[84] Nicholas M Glykos. Software news and updates carma : A molecular dynamics analysis program. *Journal of computational chemistry*, 27(14) :1765–1768, 2006.

[85] Alan Grossfield, Scott E Feller, and Michael C Pitman. Convergence of molecular dynamics simulations of membrane proteins. *Proteins : structure, function, and bioinformatics*, 67(1) :31–40, 2007.

[86] Guillaume Paris, Sebastian Kraszewski, Christophe Ramseyer, and Mironel Enescu. About the structural role of disulfide bridges in serum albumins : Evidence from protein simulated unfolding. *Biopolymers*, 97(11) :889–898, 2012.

[87] Hans Frauenfelder, Stephen G Sligar, and Peter G Wolynes. The energy landscapes and motions of proteins. *Urbana*, 51(61801) :61801, 1991.

[88] Stephen Curry, Hendrik Mandelkow, Peter Brick, and Nick Franks. Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nature Structural & Molecular Biology*, 5(9) :827–835, 1998.

[89] F Walji, A Rosen, and RC Hider. The existence of conformationally labile (preformed) drug binding sites in human serum albumin as evidenced by optical rotation measurements. *Journal of pharmacy and pharmacology*, 45(6) :551–558, 1993.

[90] Paolo Ascenzi and Mauro Fasano. Allostery in a monomeric protein : the case of human serum albumin. *Biophysical chemistry*, 148(1) :16–22, 2010.

[91] Isabelle Petitpas, Charles E Petersen, Chung-Eun Ha, Ananyo A Bhattacharya, Patricia A Zunszain, Jamie Ghuman, Nadhipuram V Bhagavan, and Stephen Curry. Structural basis of albumin–thyroxine interactions and familial dysalbuminemic hyperthyroxinemia. *Proceedings of the National Academy of Sciences*, 100(11) : 6440–6445, 2003.

[92] Joseph R Lakowicz. *Principles of fluorescence spectroscopy*. Springer, 2009.

[93] Keishi Yamasaki, Toru Maruyama, Kaori Yoshimoto, Yasuhiro Tsutsumi, Ryuichi Narazaki, Akira Fukuhara, Ulrich Kragh-Hansen, and Masaki Otagiri. Interactive binding to the two principal ligand binding sites of human serum albumin : effect of the neutral-to-base transition. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1432(2) :313–323, 1999.

[94] Jin Young Lee and Masaaki Hirose. Partially folded state of the disulfide-reduced form of human serum albumin as an intermediate for reversible denaturation. *Journal of Biological Chemistry*, 267(21) :14753–14758, 1992.

[95] Samir U Sane, Steven M Cramer, and Todd M Przybycien. A holistic approach to protein secondary structure characterization using amide i band raman spectroscopy. *Analytical biochemistry*, 269(2) :255–272, 1999.

[96] RE Tanner, B Herpigny, S-H Chen, and CK Rha. Conformational change of protein sodium dodecylsulfate complexes in solution : A study of dynamic light scattering. *The Journal of Chemical Physics*, 76 :3866, 1982.

[97] Ank Valstar, Mats Almgren, Wyn Brown, and Marilena Vasilescu. The interaction of bovine serum albumin with surfactants studied by light scattering. *Langmuir*, 16 (3) :922–927, 2000.

[98] Barbara Jachimska, Monika Wasilewska, and Zbigniew Adamczyk. Characterization of globular protein solutions by dynamic light scattering, electrophoretic mobility, and viscosity measurements. *Langmuir*, 24(13) :6866–6872, 2008.

[99] Athi N Naganathan, Urmi Doshi, Adam Fung, Mourad Sadqi, and Victor Muñoz. Dynamics, energetics, and structure in protein folding. *Biochemistry*, 45(28) :8466–8475, 2006.

[100] Mostafa Rezaei-Tavirani, Seyed Hassan Moghaddamnia, Bijan Ranjbar, Mojtaba Amani, and S Marashi. Conformational study of human serum albumin in pre-denaturation temperatures by differential scanning calorimetry, circular dichroism and uv spectroscopy. *Journal of biochemistry and molecular biology*, 39(5) :530, 2006.

[101] Koichi Murayama and Mihoko Tomida. Heat-induced secondary structure and conformation change of bovine serum albumin investigated by fourier transform infrared spectroscopy. *Biochemistry*, 43(36) :11526–11532, 2004.

[102] Vincent Le Guilloux, Peter Schmidtke, and Pierre Tuffery. Fpocket : an open source platform for ligand pocket detection. *BMC bioinformatics*, 10(1) :168, 2009.

Résumé

L'albumine sérique humaine (HSA) est une protéine connue pour ses propriétés de transport exceptionnelles et son contenu élevé en ponts disulfure. L'étude de sa dynamique conformationnelle représente un défi important dans la compréhension de ses fonctions physiologiques. Le but de notre travail a été d'étudier cette dynamique conformationnelle et de comprendre le rôle des ponts disulfure dans le maintien de la structure native de la protéine. Notre analyse est basée sur des simulations de dynamique moléculaire (DM) couplées à des analyses par composantes principales (PCA). Outre la validation de la méthode de simulation les résultats fournissent de nouveaux éclairages sur les principaux effets de la réduction des ponts disulfure dans les albumines sériques. Les processus de dépliement/repliement protéique ont été détaillés. La prédiction de la structure réduite d'équilibre a fait l'objet d'une attention particulière. Une étude détaillée de la dynamique conformationnelle globale de la protéine ainsi que celle des deux sites principaux de complexation a été effectuée. D'éventuels effets allostériques entre ces deux sites ont été recherchés. Les résultats théoriques obtenus ont été discutés avec les données expérimentales disponibles.

Mots clés

Simulations de dynamique moléculaire, analyse en composantes principales, ponts disulfure, albumine sérique humaine, sites de complexation, dépliement protéique.

Abstract

Human serum albumin (HSA) is a protein known for its exceptional transport properties and its high content of disulfide bridges. The study of the conformational dynamics represents a major challenge in the understanding of its physiological functions. The aim of our work was to study the conformational dynamics and understanding the role of disulfide bonds in the stability of the native protein structure. Our analysis is based on simulations of molecular dynamics (MD) coupled with principal component analysis (PCA). Beyond the validation of the simulation method, the results provide new insights into the main effects of the disulfide bonds reduction in serum albumins. Protein unfolding/refolding processes were detailed. A special attention is paid to the prediction of the reduced structure at the equilibrium. A detailed study of the global protein conformational dynamics as well as the two main binding sites were performed. Possible allosteric effects between these two sites were researched. The theoretical results have been discussed with the available experimental data.

Keywords

Molecular dynamics simulations, principal component analysis, disulfide bridge, human serum albumin, binding sites, protein unfolding.