



HAL
open science

Étude des propriétés mécaniques des protéines par modélisation moléculaire

Fabien Cailliez

► **To cite this version:**

Fabien Cailliez. Étude des propriétés mécaniques des protéines par modélisation moléculaire. Chimie théorique et/ou physique. Université Paris Diderot Paris 7, 2006. Français. NNT : . tel-01084328

HAL Id: tel-01084328

<https://hal.science/tel-01084328>

Submitted on 19 Nov 2014

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

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

UFR de Biologie
2 place Jussieu, Tour 42
75005 Paris

Laboratoire de Biochimie Théorique
IBPC, CNRS UPR9080
13 rue P. et M. Curie
75005 Paris

Thèse de Doctorat de l'Université Paris 7 - Denis Diderot

École Doctorale B2M - Biochimie et Biologie Moléculaire

Spécialité : Analyse des Génomes et Modélisation Moléculaire

par

Fabien Cailliez

pour obtenir le grade de Docteur de l'Université Paris 7

**Étude des propriétés mécaniques des protéines
par modélisation moléculaire**

Soutenance publique le 29 septembre 2006 devant le jury composé de :

Présidente :	Pr. Catherine Etchebest, Université Paris 7
Rapporteurs :	Pr. Martine Prévost, Université Libre de Bruxelles (Belgique) Pr. Georges Wipff, Université Louis Pasteur - Strasbourg
Examineurs :	Dr. Patrick Fuchs, Université Paris 7 Dr. Yves-Henri Sanejouand, CNRS
Directeur de thèse :	Dr. Richard Lavery, CNRS

Table des matières

Introduction générale	9
I Les protéines : des molécules flexibles	13
1 Relation structure-fonction des protéines	15
1.1 De la séquence à la fonction	15
1.2 Relation structure-fonction	16
1.2.1 Repliement et fonction	16
1.2.2 Sites fonctionnels des protéines	17
1.3 Rôle de la flexibilité des protéines	18
1.3.1 Mise en évidence expérimentale	19
1.3.2 Échelles de temps et importance fonctionnelle	20
2 Caractérisation des propriétés mécaniques des protéines	25
2.1 Mesure expérimentale de la dynamique des protéines	25
2.1.1 Cristallographie aux rayons X	26
2.1.2 Résonance Magnétique Nucléaire	27
2.2 Nanomanipulations	29
2.2.1 Expériences de dépliement	29
2.2.2 Séparation de complexes	31
2.3 Étude théorique des propriétés mécaniques des protéines	32
2.3.1 Analyses par modes normaux	32
2.3.2 Dynamique moléculaire et minimisation d'énergie	35
3 Une nouvelle mesure de la flexibilité des protéines	39
3.1 Limitation des mesures actuelles	39
3.2 Méthodologie	40
3.3 Un nouvel indicateur de rigidité	43

3.3.1	Propriétés mécaniques d'une hélice α	43
3.3.2	Étude d'une protéine globulaire : la nucléase de staphylocoque	44
3.4	Définition de domaines dynamiques	48
3.4.1	Méthodologie	49
3.4.2	Domaines mécaniques de SNase	52
3.4.3	Étude d'une série de protéines	54
3.4.4	Positionnement des résidus rigides	55
3.5	Conclusion	56
3.6	Perspectives	58
 II Propriétés mécaniques et fonction : la E-cadhérine		61
 4 Les molécules de cadhérine		63
4.1	L'adhésion cellulaire	63
4.2	Les cadhérines	65
4.2.1	Cadhérines classiques	67
4.2.2	Autres cadhérines	68
4.3	Propriétés des cadhérines classiques	69
4.3.1	Structure du domaine extracellulaire	69
4.3.2	Le rôle du calcium	69
4.3.3	Interfaces de dimérisation	73
4.3.4	Le rôle du tryptophane 2	77
4.3.5	Les différents modèles d'adhésion	79
4.3.6	Études théoriques des cadhérines	81
 5 Le rôle des ions calcium sur la stabilité des cadhérines		83
5.1	Introduction	83
5.2	Matériel et méthodes	84
5.2.1	Configuration initiale	84
5.2.2	Simulations de dynamique moléculaire	84
5.2.3	Analyse structurale	86
5.2.4	Estimation d'enthalpies libres de dimérisation	88
5.3	Étude des systèmes monomériques	89
5.3.1	Mise en évidence de changements conformationnels	89
5.3.2	Caractérisation des mouvements inter-domaines	91
5.4	Étude du dimère <i>BN</i> de EC1-2	94

5.4.1	Dynamique interne	94
5.4.2	Évolution de l'interface	95
5.4.3	Enthalpies libres de dimérisation	96
5.5	Conclusion	99
6	Dynamique des interfaces	101
6.1	Introduction	101
6.2	Matériel et méthodes	102
6.2.1	Construction des dimères	102
6.2.2	Construction des cellules de simulation	103
6.2.3	Simulations de dynamique moléculaire	103
6.2.4	Analyse conformationnelle et estimation d'enthalpies libres de dimé- risation	104
6.3	Stabilité structurale des dimères	105
6.4	Changements conformationnels du dimère <i>BE</i>	107
6.5	Enthalpies libres de dimérisation	111
6.6	Interfaces et spécificité	112
6.7	Conclusion	115
III	Étude des propriétés des grands systèmes	117
7	Définition d'une représentation multi-échelle	119
7.1	Limitation des représentations actuelles des protéines	119
7.2	Principe d'une représentation multi-échelle	121
7.2.1	Exemple des méthodes QM/MM	121
7.2.2	Principe de la représentation "tout-atome"/"gros grains"	122
7.3	Mise en œuvre	124
7.3.1	Définition de la représentation multi-échelle MM/CG	124
7.3.2	Intégration dans le programme AMBER	125
7.4	Étude d'une petite protéine globulaire	127
7.4.1	Étude dans le vide	127
7.4.2	Incorporation du solvant dans le modèle	132
7.5	Étude d'un homodimère	138
7.5.1	Choix du système	138
7.5.2	Dynamique globale du système	139
7.5.3	Fluctuations atomiques dans la zone MM	142

7.6	Bilan et perspectives	143
Annexes		151
A Structure des protéines		151
A.1	Structure primaire - la chaîne polypeptidique	151
A.2	Les structures secondaires	153
A.3	Structures tertiaire et quaternaire	156
B Champs de force et modèles de représentation		159
B.1	Définition du champ de force	159
B.2	Modèles "tout-atome"	159
B.3	Représentation "gros grains" et réseaux élastiques	161
C Minimisation d'énergie		163
C.1	Surface d'énergie potentielle	163
C.2	Principe de la minimisation d'énergie	164
C.3	Les principales méthodes de minimisation	165
C.4	Minimisation en coordonnées internes, le programme LIGAND	167
D Dynamique moléculaire		171
D.1	Principe de la dynamique moléculaire	171
D.2	Troncatures du potentiel	173
D.3	Conditions aux limites	174
D.4	Ensembles thermodynamiques	178
D.5	Protocole	179
D.6	Limitations	180
E Modes normaux et analyse en composantes principales		183
E.1	Modes normaux	183
E.2	Analyse en composantes principales - modes quasi-harmoniques	186
F Modèles théoriques de solvation		189
F.1	Solvant explicite	189
F.2	Solvant implicite	190
F.2.1	Modèle Poisson-Boltzmann	191
F.2.2	Modèle Born généralisé	192

G	Méthode MM-PBSA	195
G.1	Principe de la méthode	195
G.2	Terme enthalpique	197
G.3	Terme de solvatation	197
G.4	Terme entropique	198
G.5	Limitations	198
	Bibliographie	198

Introduction générale

Contexte du travail

Les protéines représentent plus de 50 % de la masse sèche des cellules (elles tirent d'ailleurs leur nom du mot grec "*prôtos*", qui signifie premier) et jouent un rôle primordial dans le bon fonctionnement des organismes. Elles sont des acteurs indispensables pour la grande majorité des fonctions de la cellule : les enzymes servent de catalyseurs pour les réactions métaboliques ; certaines protéines comme la tubuline ou le collagène assurent la structure des cellules ; d'autres encore, comme les cadhérines, jouent un rôle dans la formation des contacts entre cellule...

Cette grande variété de fonctions est assurée par une non moins grande variété d'architectures protéiques. La structure d'une protéine est en effet adaptée à la fonction qu'elle assure, ce qui a pendant longtemps amené les biologistes à porter leur attention sur la résolution de ces structures. Malheureusement, la connaissance de la structure seule n'est pas toujours suffisante pour appréhender le rôle d'une protéine et différentes observations expérimentales ont montré que la dynamique des protéines est essentielle pour qu'elles puissent assurer leur fonction.

C'est pourquoi l'étude des propriétés mécaniques des protéines est devenue un champ de recherche privilégié, tant au niveau expérimental que théorique. Expérimentalement, le développement depuis une quinzaine d'années de nouvelles techniques comme les nanomanipulations ont permis de compléter les informations déjà disponibles (par cristallographie ou par diverses méthodes spectroscopiques). Dans ces expériences, une force extérieure est appliquée à une biomolécule, permettant de mesurer sa résistance à un stress mécanique. Ceci est particulièrement intéressant pour l'étude de certaines protéines qui subissent de telles forces *in vivo*, comme par exemple les protéines moteurs ou les protéines d'adhésion cellulaire. Parallèlement à ces développements méthodologiques, l'essor de la modélisation a également permis de mieux comprendre ces propriétés au niveau atomique, en reprodui-

sant par exemple les expériences de nanomanipulation et en en donnant une interprétation en termes de modifications structurales. Pour autant, les propriétés mécaniques des protéines sont loin d'être comprises dans leur ensemble.

Dans ce contexte, la modélisation moléculaire est amenée à jouer un rôle majeur dans les années à venir pour tenter d'élucider les relations entre les propriétés mécaniques des protéines et leurs fonctions. Afin d'y parvenir, il est nécessaire à la fois de tirer parti des techniques déjà existantes, mais également de développer de nouvelles méthodologies.

Présentation du mémoire

Au cours de ces quatre dernières années, notre équipe a ainsi cherché à mieux comprendre le fonctionnement des protéines en étudiant leurs propriétés mécaniques. Ce travail se situe dans la continuité des études déjà réalisées au laboratoire sur l'ADN (Lebrun et Lavery, 1998) puis sur les protéines (Navizet *et al.*, 2004*b*) et comporte à la fois un aspect méthodologique et l'application des techniques de dynamique moléculaire à l'étude d'une molécule impliquée dans les mécanismes d'adhésion cellulaire.

Dans une première partie, après avoir fait le point sur les principales méthodes d'étude théorique et expérimentale des propriétés mécaniques des protéines, ce manuscrit décrit un travail visant à définir une nouvelle mesure de la flexibilité des protéines à l'échelle du résidu, dont la mise en place a été guidée par la volonté de dépasser les limitations des mesures disponibles (comme les facteurs de Debye-Waller).

Nous nous sommes ensuite intéressés à la E-cadhérine, molécule qui appartient à la classe des protéines d'adhésion cellulaire. Ces protéines ont en effet des propriétés mécaniques tout à fait particulières qui leur permettent d'assurer leur fonction. Nos études, qui sont détaillées dans la deuxième partie, ont porté plus précisément sur le rôle des ions calcium sur la flexibilité des cadhérines puis sur la stabilité dynamique et thermodynamique des diverses surfaces d'interaction observées expérimentalement pour des dimères de cadhérine.

La troisième partie de ce manuscrit présente des travaux préliminaires, réalisés en collaboration avec le groupe de Steve Harvey au Georgia Institute of Technology à Atlanta (USA), qui visent à créer une représentation multi-échelle des protéines à mi-chemin entre une représentation dans laquelle tous les atomes sont considérés et un modèle plus simple dans lequel la protéine est modélisée par un réseau élastique. La mise en place d'une telle représentation est une réponse possible aux limites imposées par les moyens de calcul. Dans cette optique, elle s'avèrerait très intéressante pour l'étude de systèmes de grande taille.

Enfin, les annexes regroupent un descriptif de la structure des protéines ainsi que l'ensemble des méthodes classiques utilisées au cours du travail présenté dans les trois parties.

Première partie

Les protéines : des molécules flexibles

Chapitre 1

Relation structure-fonction des protéines

Sommaire

1.1	De la séquence à la fonction	15
1.2	Relation structure-fonction	16
1.2.1	Repliement et fonction	16
1.2.2	Sites fonctionnels des protéines	17
1.3	Rôle de la flexibilité des protéines	18
1.3.1	Mise en évidence expérimentale	19
1.3.2	Échelles de temps et importance fonctionnelle	20

1.1 De la séquence à la fonction

Avec les projets de séquençage de génomes entiers, on assiste à une croissance exponentielle du nombre de séquences de protéines disponibles de fonction inconnue. Une telle quantité de données nécessite l'utilisation des outils informatiques afin de déterminer les fonctions de ces protéines. Des méthodes sont ainsi développées qui tentent de relier une séquence d'acides aminés à la fonction de la protéine correspondante, notamment en tirant parti de la comparaison de la nouvelle séquence avec celles de protéines dont la fonction est déjà établie.

Il est en effet clairement établi que les similarités de séquence entre protéines se transmettent à la structure tertiaire (Chothia et Lesk, 1986). Ainsi, deux protéines partageant une forte homologie de séquence partagent également fréquemment le même repliement. Le dernier lien nécessaire à la détermination de la fonction de la protéine est donc la relation entre la structure et la fonction. Mais ce lien existe-t-il ? En d'autres termes, la structure de la protéine impose-t-elle la fonction ?

1.2 Relation structure-fonction

Depuis ces dix dernières années, la motivation de l'obtention des structures tridimensionnelles de protéine s'est modifiée. Dans l'approche traditionnelle, la résolution de la structure sert à comprendre le mécanisme d'action de la protéine dont la fonction est connue. Récemment cependant, la structure est parfois devenue un outil utilisé pour déterminer la fonction de la protéine. Il est de ce fait devenu important d'élucider les rapports qui peuvent exister entre structure et fonction.

1.2.1 Repliement et fonction

1.2.1.1 Banques de données biologiques

Une première façon de classer les protéines est de comparer leurs fonctions. Parmi les nombreuses banques de données biologiques disponibles, l'une des plus usitées est certainement ENZYME¹ (Bairoch, 2000) qui se focalise sur les enzymes. Celles-ci sont classées en fonction de leur code EC (Enzyme Commission), attribué selon le type de réaction catalysée ainsi que les substrats métabolisés.

1.2.1.2 Banques de données structurales

Parallèlement à l'augmentation du nombre de séquences protéiques connues, le nombre de structures tridimensionnelles de protéines a lui aussi augmenté exponentiellement. En 1971, la Protein Data Bank² (Deshpande *et al.*, 2005), notée PDB par la suite, contenait 7 structures, alors qu'elle en compte près de 36000 maintenant.

Il est possible de classer ces protéines de structure connue en fonction de leur repliement. Ceci est réalisé dans plusieurs banques de données qui diffèrent par la méthode

1. <http://www.expasy.org/enzyme/>.

2. <http://www.pdb.org/>.

utilisée pour comparer deux structures. Parmi ces banques de données, les plus utilisées sont certainement FSSP¹ (Holm et Sander, 1998), SCOP² (Murzin *et al.*, 1995), et CATH³ (Pearl *et al.*, 2005). En dépit de leurs différents modes de classification, les hiérarchies de structure définies dans ces trois banques sont souvent cohérentes (Getz *et al.*, 2002).

1.2.1.3 Le repliement ne définit pas la fonction

De nombreux groupes se sont penchés sur la relation entre le repliement d'une protéine et sa fonction en réalisant des analyses systématiques des liens entre les différentes banques de données biologiques et structurales (Martin *et al.*, 1998 ; Hegyi et Gerstein, 1999 ; Thornton *et al.*, 2000 ; Laskowski *et al.*, 2003).

Les conclusions de ces études indiquent qu'il n'existe pas de relation simple entre repliement et fonction. Hegyi et collaborateurs (Hegyi et Gerstein, 1999) ont noté que pour les enzymes il existe en moyenne environ 1,8 fonctions par repliement et 2,5 repliements par fonction. Ainsi, il existe de nombreux exemples de protéines partageant la même fonction mais dont le repliement est très différent (voir figure 1.1 A) ou inversement des repliements pouvant servir à réaliser de multiples fonctions (la figure 1.1 B montre l'exemple de la classe TIM-barrel qui est le support de nombreuses fonctions (Nagano *et al.*, 2002)). Il n'existe donc pas de lien clair entre le repliement et la fonction d'une protéine. Néanmoins, l'utilisation des comparaisons de séquence en plus de l'information sur le repliement permet de préciser la fonction d'une protéine. En effet, deux protéines de repliement similaire et dont l'homologie de séquence est supérieure à 40 % auront très certainement des fonctions identiques. En revanche, pour un taux d'homologie inférieur à 25 %, les fonctions divergeront sûrement.

1.2.2 Sites fonctionnels des protéines

La complexité des relations existant entre le repliement et la fonction nécessite de regarder la structure non plus dans son intégralité mais de façon plus fine et de s'intéresser à des sites fonctionnels localisés (Jones et Thornton, 2004).

1. FSSP : *Families of Structurally Similar Proteins*, <http://swift.cmbi.kun.nl/swift/fssp/>.
2. SCOP : *Structural Classification of Proteins*, <http://scop.berkeley.edu/>.
3. CATH : *Class, Architecture, Topology and Homologous superfamily*,
<http://cathwww.biochem.ucl.ac.uk/latest/index.html>.



FIGURE 1.1 – Exemples de divergences entre repliement et fonction. **A** : Deux protéines partageant la même fonction (anhydrases carboniques) mais possédant un repliement différent. **B** : Le repliement TIM-barrel peut servir de support à de multiples fonctions en fonction de la localisation du site actif (identifiée par les boîtes numérotées 1, 3, 4 et 5), reproduit de (Hegyi et Gerstein, 1999).

Laskowski et collaborateurs (Laskowski *et al.*, 1996) ont ainsi analysé la position des ligands dans les complexes enzyme-ligand. Les résultats ont montré que les sites de fixation se situent très majoritairement dans la plus grande crevasse définie par la surface protéique, et ce d'autant plus que la taille de cette crevasse est très supérieure à celle des autres crevasses. Il est donc souvent possible d'identifier sur la structure la zone correspondant aux sites actifs enzymatiques, sans pour autant pouvoir déterminer avec précision les résidus catalytiques et la fonction de la protéine.

L'analyse de la position des résidus catalytiques dans des familles de protéines a également permis de mettre en évidence que la position spatiale relative de ces résidus étaient très souvent conservée (Wallace *et al.*, 1996 ; Wallace *et al.*, 1997), permettant ainsi de retrouver dans certains cas un site catalytique par la recherche d'un motif structural conservé. La géométrie de la triade catalytique Ser-His-Asp dans la structure est ainsi caractéristique des protéases à sérine. En revanche, la disposition de la dyade catalytique Asp-Glu présente dans les lysozymes n'est pas caractéristique de cette famille d'enzymes.

1.3 Rôle de la flexibilité des protéines

L'existence de mouvements au sein des protéines est connue depuis longtemps. Pourtant, le rôle fonctionnel de cette flexibilité a été largement sous-estimé jusqu'à il y a une dizaine

d'années, au profit de la relation structure-fonction.

1.3.1 Mise en évidence expérimentale

Les preuves de l'importance des mouvements pour la fonction des protéines sont nombreuses. Des études menées sur l'ARNase A par cristallographie aux rayons X ont montré qu'à 228 K, l'enzyme fixe un analogue du substrat non réactif, tandis qu'à 212 K aucune fixation n'est observée. Or il a été démontré qu'en dessous de 220 K environ, les mouvements au sein des protéines sont limités aux fluctuations atomiques¹ (Voet et Voet, 1998). Ceci montre bien que la fixation du substrat par l'ARNase nécessite un changement conformationnel relativement important.

Par ailleurs, il est fréquent que la conformation d'une protéine non liée à son substrat soit très différente de sa conformation liée. Ceci est illustré sur la figure 1.2 avec l'exemple d'une enzyme liant le D-ribose (notée DRBP pour "*D-Ribose Binding Protein*").

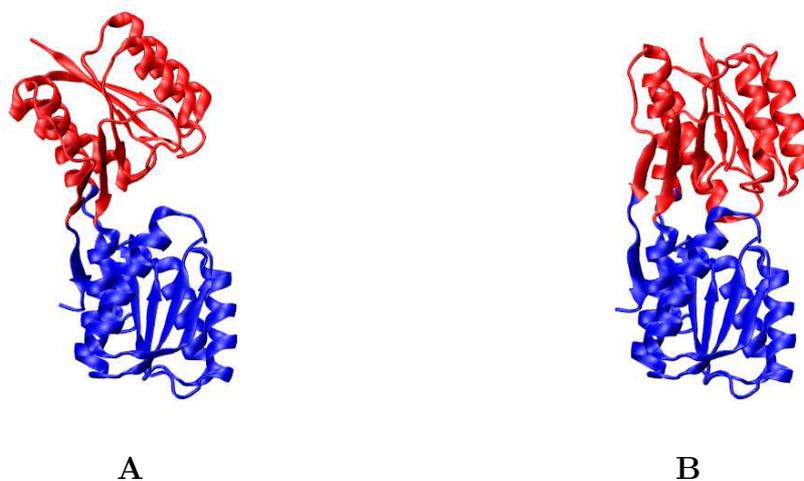


FIGURE 1.2 – Changement conformationnel lors de la fixation d'un ligand. **A** : Forme libre de la DRBP. **B** : Forme liée au β D-ribose de la DRBP.

Il est également possible de suivre des changements conformationnels à l'aide d'études d'échange d'hydrogène, et ce sur plusieurs échelles de temps en fonction de la technique

1. Les fluctuations atomiques (notées RMSF pour "*root mean square fluctuations*") définissent la variabilité de la position d'un atome autour de sa position moyenne. Elles sont calculées pour un ensemble de K structures par la formule : $RMSF_i = \sqrt{\frac{1}{K} \sum_{j=1}^K \left((x_i^j - \langle x_i \rangle)^2 + (y_i^j - \langle y_i \rangle)^2 + (z_i^j - \langle z_i \rangle)^2 \right)}$. Les (x_i^j, y_i^j, z_i^j) représentent les coordonnées de l'atome i dans la structure j et les grandeurs entre crochets représentent des valeurs moyennes de ces grandeurs calculées sur l'ensemble des K structures.

utilisée pour mesurer ce phénomène d'échange (mesures de radioactivité, RMN,...). Dans une protéine, différents atomes d'hydrogène sont labiles (provenant des groupes amines ou hydroxyles par exemple) et peuvent s'échanger avec ceux de l'eau, si toutefois leur accessibilité le permet. Or les protons enfouis au cœur de la structure peuvent également être échangés, bien que moins rapidement. Ceci ne peut s'expliquer que par une mobilité conformationnelle de la protéine modifiant l'accessibilité de certains résidus. Ainsi, la vitesse d'échange d'un proton va refléter la mobilité de son environnement (Závodszky *et al.*, 1998).

Différentes méthodes de spectroscopie ont également permis de mettre en évidence la flexibilité des protéines, et le fait que la structure de celles-ci fluctue entre différents sous-états conformationnels. Ceci a notamment été démontré sur l'exemple de la myoglobine par diverses observations utilisant la photolyse éclair, la spectroscopie aux rayons X ou la spectroscopie Mössbauer à différentes températures (Frauenfelder *et al.*, 1988).

Des expériences de mutations ponctuelles restreignant la flexibilité de certaines parties de la molécule permettent également de mettre en évidence la nécessité de changements conformationnels pour le bon fonctionnement d'une protéine. Dans le cas de la E-cadhérine, protéine d'adhésion cellulaire qui sera étudiée dans la partie II, il est par exemple possible d'annihiler les propriétés adhésives de cette molécule en créant un pont disulfure qui bloque la position du brin N-terminal (Harrison *et al.*, 2005).

1.3.2 Échelles de temps et importance fonctionnelle

Les protéines sont donc des molécules flexibles qui oscillent entre de nombreuses conformations, en fonction de leurs propriétés mécaniques. Les mouvements internes qu'elles subissent se déroulent à différentes échelles de temps, allant de la picoseconde à la seconde, qui correspondent à des fonctionnalités différentes (voir tableau 1.1).

L'importance des propriétés mécaniques pour la fonction des protéines fait ainsi aujourd'hui l'objet d'un large consensus dans la communauté scientifique. Pour autant, les relations qui existent entre ces propriétés et leurs effets ne sont pas toujours claires. À ce titre, on peut citer cinq domaines d'étude dans lesquels les propriétés mécaniques jouent

certainement un rôle essentiel, mais qui reste encore à ce jour à clarifier :

– **lien entre flexibilité et thermostabilité**

De nombreux travaux ont montré qu'il était possible de stabiliser des protéines par rapport à la dénaturation thermiquement induite, en effectuant un petit nombre de mutations ponctuelles (Vieille *et al.*, 1996 ; Haney *et al.*, 1999). La compréhension des raisons de ce phénomène serait d'un intérêt certain au niveau industriel (il serait possible de modifier "à souhait" le domaine de stabilité d'une enzyme). Plusieurs études contradictoires ont tenté de relier la stabilité aux propriétés de flexibilité des protéines : pour Grottesi et collaborateurs (Grottesi *et al.*, 2002), une augmentation de la flexibilité augmente la stabilité tandis que Chakravarty et collaborateurs (Chakravarty et Varadarajan, 2002) arrivent à une conclusion opposée.

– **repliement des protéines**

L'étude du repliement et du dépliement des protéines revêt un intérêt tout particulier de par son importance dans les phénomènes d'agrégation de protéines (formation de fibres amyloïdes).

– **sites catalytiques et surfaces d'interaction**

Certaines protéines possèdent une structure qui semble adéquate pour réaliser une fonction, et qui pourtant sont inactives. Cette inactivité pourrait provenir de propriétés mécaniques rendant l'enzyme inapte à réaliser la fonction voulue.

– **protéines moteurs**

La résolution de plusieurs structures des domaines moteurs de myosine (protéine se déplaçant sur une fibre d'actine) et de kinésine (qui se déplace sur une microtubule) n'a pas pu apporter de réponse définitive quant au mécanisme du moteur. L'analyse des propriétés mécaniques des "bras de levier" et de leur rigidité devrait permettre d'apporter de nouvelles informations sur ce mécanisme.

– **protéines d'adhésion**

Les protéines d'adhésion sont soumises à des forces mécaniques importantes pour maintenir la structure des agrégats cellulaires et des tissus. Elles doivent donc présenter des propriétés mécaniques particulières pour répondre à ces sollicitations.

Type de mouvements	Exemple de fonctionnalité	Échelle de temps et d'amplitude
Mouvements locaux : - fluctuations atomiques		fs - ps moins de 1 Å
Mouvements de petites amplitudes : - mouvements de chaînes latérales	- association d'un ligand	ns - μ s quelques Å
Mouvements d'amplitudes moyennes : - mouvements des boucles - extrémités N/C-terminales - mouvements rigides des hélices	- adaptation du site de liaison - association spécifique	μ s - ms 5 à 10 Å
Mouvements de grandes amplitudes : - mouvements de domaine - mouvements des sous-unités	- déplacements de domaines au sein de la protéine - transition allostérique	ms - s plus de 10 Å
Mouvements globaux : - transition hélice- <i>coil</i> - dépliement ou repliement	- activation des hormones - fonctionnalité de la protéine	s - h plus de 5 Å

TABLE 1.1 – Mouvements internes des protéines, fonctionnalité, et échelles temporelle et spatiale. Inspiré de (McCammon et Harvey, 1987).

Chapitre 2

Caractérisation des propriétés mécaniques des protéines

Sommaire

2.1	Mesure expérimentale de la dynamique des protéines	25
2.1.1	Cristallographie aux rayons X	26
2.1.2	Résonance Magnétique Nucléaire	27
2.2	Nanomanipulations	29
2.2.1	Expériences de dépliement	29
2.2.2	Séparation de complexes	31
2.3	Étude théorique des propriétés mécaniques des protéines . .	32
2.3.1	Analyses par modes normaux	32
2.3.2	Dynamique moléculaire et minimisation d'énergie	35

2.1 Mesure expérimentale de la dynamique des protéines

La dynamique des protéines et leurs propriétés mécaniques peuvent être étudiées expérimentalement. Compte tenu de la grande gamme de temps caractéristiques correspondant aux différents mouvements décrits dans le tableau 1.1, l'utilisation de plusieurs techniques est nécessaire pour recueillir des informations sur les propriétés dynamiques des protéines à ces différentes échelles. Nous nous limiterons ici à décrire deux méthodes fréquemment utilisées en biologie qui apportent à la fois des informations structurales et dynamiques : la cristallographie aux rayons X et la résonance magnétique nucléaire.

2.1.1 Cristallographie aux rayons X

La cristallographie aux rayons X est l'une des deux principales techniques utilisées pour déterminer la structure des protéines. En plus de cette information structurale, elle permet également d'obtenir des informations sur la dynamique des protéines.

Tout d'abord, comme déjà évoqué au paragraphe 1.3, l'obtention de différentes conformations pour une même protéine indique que les protéines peuvent subir des mouvements internes de grande ampleur. La banque de données "Database of Macromolecular Movements"¹ (Gerstein et Krebs, 1998 ; Echols *et al.*, 2003 ; Flores *et al.*, 2006) est ainsi dédiée à l'étude comparative des différentes conformations d'une protéine. Dans cette banque de données sont stockées un nombre important de protéines pour lesquelles plusieurs conformations sont connues (ainsi que quelques protéines au sein desquelles on suspecte un mouvement de domaines), et le mouvement associé au passage d'une conformation à une autre est analysé et classé. La majorité des mouvements observés correspond soit à un mouvement de rotation autour d'une charnière ("*hinge bending*"), soit à un mouvement de cisaillement d'un domaine par rapport à un autre ("*shear motion*"). La figure 2.1 montre un schéma explicitant ces deux types de mouvement. De façon générale, un changement de conformation au sein d'une protéine peut être décrit grossièrement comme la combinaison de mouvements de charnière et de mouvements de cisaillement (Gerstein et Krebs, 1998).

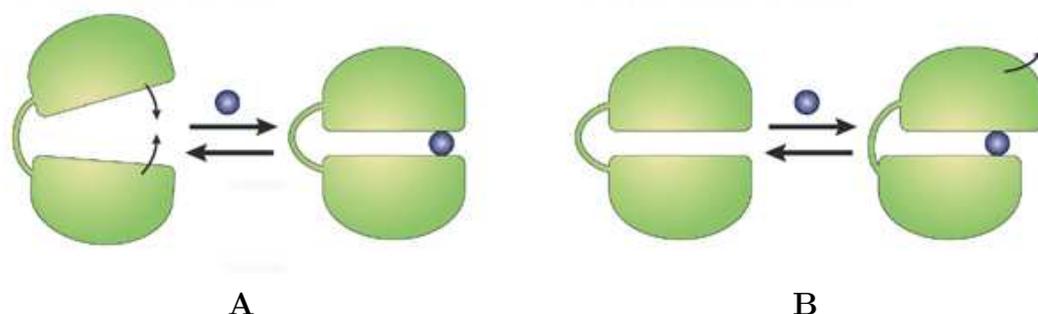


FIGURE 2.1 – Représentation schématique d'un mouvement de rotation autour d'une charnière (A) et d'un mouvement de cisaillement (B) de la protéine (en vert) au cours de l'association avec un ligand (en bleu). Reproduit de (Teague, 2003).

Mais la cristallographie permet également de collecter des informations sur des mouvements plus localisés. Les facteurs de Debye-Waller fournis par cette technique sont une mesure de la mobilité des atomes. En effet, la valeur de ces facteurs de Debye-Waller

1. <http://molmovdb.mbb.yale.edu/MolMovDB>.

(appelés aussi facteurs de température ou B-facteurs) est reliée à la valeur moyenne du déplacement atomique. Pour un atome i , on a :

$$B_i = \frac{8\pi^2}{3} \langle \Delta \vec{r}_i^2 \rangle$$

où B_i est le facteur de température de l'atome i et \vec{r}_i est le vecteur décrivant sa position. Cette formule donne accès au facteur de température isotrope. Dans les cristaux à haute résolution, il est possible de définir des facteurs de température anisotropes fournissant en plus une indication sur la directionnalité des fluctuations.

Une valeur faible du facteur de température indique que l'atome est assez peu mobile. Au contraire, un atome soumis à des déplacements plus importants possède un facteur de température élevé. En ce sens, cette grandeur est souvent considérée comme un indicateur de la flexibilité des différentes zones de la protéine. Il convient néanmoins de rester prudent sur l'interprétation du profil des facteurs de température au sein d'une protéine. En effet, une étude récente a montré que ceux-ci sont dominés par l'empaquetage local du résidu (Halle, 2002) : il existe une corrélation importante entre l'inverse de la densité de contacts d'un atome (définie comme le nombre de voisins non covalents de cet atome) et son facteur de température. De plus, cette mesure étant réalisée dans le cristal, il est difficile d'estimer l'influence de la maille cristalline sur la dynamique de la protéine, et donc les différences qu'elle induit par rapport au comportement de la protéine en solution.

2.1.2 Résonance Magnétique Nucléaire

À l'instar de la dynamique moléculaire, la résonance magnétique nucléaire (RMN) permet de déterminer des structures protéiques et fournit également un grand nombre d'informations sur leurs propriétés dynamiques.

L'un des grands avantages de la RMN est sa capacité à pouvoir mesurer des propriétés physiques sur plusieurs échelles de temps caractéristiques (de la milliseconde à la seconde). Comme évoqué dans la section 1.3, la RMN permet de mesurer des cinétiques d'échange de protons. La mobilité des chaînes latérales peut également être sondée à l'aide d'expériences de RMN, en étudiant la dégénérescence des spectres. Dans l'exemple donné figure 2.2, les protons du cycle aromatique de la phénylalanine sont deux à deux symétriques. Si la chaîne latérale est fixe, le spectre RMN observé pour ces protons est différent selon le caractère isotrope ou non de l'environnement. En revanche, si une rotation rapide autour de la liaison $C_\beta - C_\gamma$ est possible, alors l'environnement moyen vu par les protons symétriques devient

équivalent même dans un environnement anisotrope, et le spectre est alors modifié. On voit sur cet exemple qu'il est possible d'obtenir des informations qualitatives sur les possibilités de rotation dans les chaînes latérales de la protéine par spectroscopie RMN.

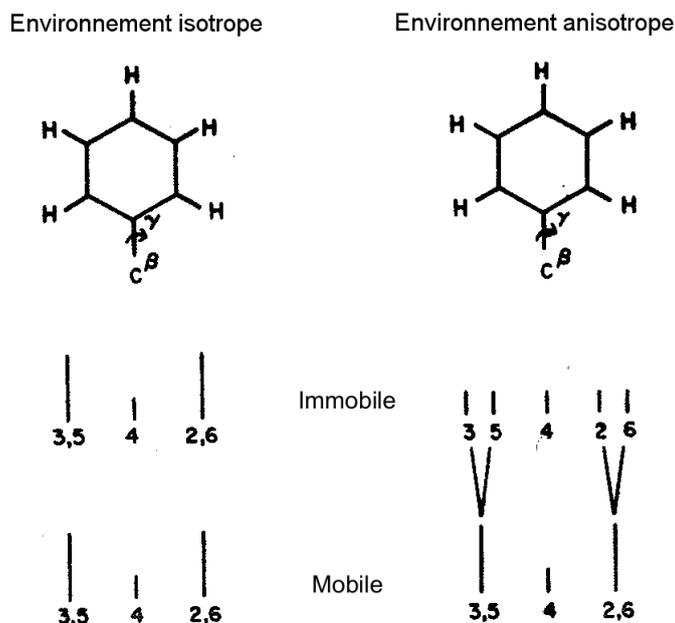


FIGURE 2.2 – Spectre RMN schématique des protons du cycle aromatique d'une phénylalanine en fonction de l'anisotropie de l'environnement et de la mobilité du cycle.

Par ailleurs, la relaxation du signal de RMN de l'azote ^{15}N peut être analysée en utilisant la procédure "*model-free*" de Lipari et Szabo (Lipari et Szabo, 1982), qui aboutit au calcul d'un paramètre d'ordre généralisé (noté S). Cette grandeur, qui prend une valeur comprise entre 0 et 1, est une mesure associée à la liberté de mouvement d'un vecteur (le plus souvent le vecteur $^{15}\text{N} - \text{H}$ du groupement amide de la liaison peptidique) qui permet une relaxation plus ou moins rapide après l'excitation par un champ magnétique. Dans le cadre du modèle de Lipari et Szabo, S est calculé par la relation suivante :

$$S^2 = \lim_{t \rightarrow \infty} C_I(t)$$

où $C_I(t)$ est la fonction de corrélation des mouvements du vecteur unitaire $\vec{\mu}$ colinéaire à $^{15}\text{N} - \text{H}$, donnée par :

$$C_I(t) = \langle P_2(\vec{\mu}(0) \cdot \vec{\mu}(t)) \rangle$$

P_2 est le polynôme de Lagrange d'ordre 2 dont l'expression est :

$$P_2(x) = \frac{3x^2 - 1}{2}$$

Si le vecteur est libre de ses mouvements, alors S prend une valeur proche de 0, tandis que si le mouvement est complètement contraint, S est égal à 1. Cette mesure est donc considérée comme un indicateur qualitatif de la rigidité locale de la protéine.

2.2 Nanomanipulations

Les expériences de nanomanipulation consistent à appliquer des forces sur une molécule ou un édifice moléculaire préalablement ancré à une surface par une extrémité et à un capteur de force par l'autre. Depuis une dizaine d'années, de nombreuses techniques ont été mises au point pour étudier les molécules biologiques à l'échelle de la molécule unique : microscopie à force atomique (AFM¹), pinces optiques ou magnétiques, BFP²,...

La variété des méthodologies qui ont été développées a permis d'examiner le comportement de la majorité des macromolécules biologiques. Des expériences ont en effet déjà été menées sur des brins d'ADN³ (Cluzel *et al.*, 1996 ; Smith *et al.*, 1996 ; Bryant *et al.*, 2003 ; Bustamante *et al.*, 2003), d'ARN⁴ (Liphardt *et al.*, 2001), ou sur des protéines (Smith *et al.*, 2003) pour étudier plus particulièrement la réponse de ces molécules à un stress mécanique. Les nanomanipulations permettent également d'examiner la force des interactions entre molécules en appliquant ces méthodologies à l'étude de complexes ADN-protéine (Dohoney et Gelles, 2001), protéine-ligand (Florin *et al.*, 1994 ; Moy *et al.*, 1994), ou protéine-protéine (Perret *et al.*, 2004).

2.2.1 Expériences de dépliement

Dans ce type de manipulations, les extrémités de la molécule étudiée sont attachées à des surfaces et une force est appliquée entre ces deux extrémités afin d'induire le dépliement de la molécule. L'analyse des résultats permet de déterminer quelles interactions sont responsables de la stabilité mécanique de la molécule et dans quelles proportions.

1. AFM : *Atomic Force Microscopy*.
2. BFP : *Biomembrane Force Probe*.
3. ADN : *Acide Désoxyribo-Nucléique*.
4. ARN : *Acide Ribo-Nucléique*.

Ces expériences de dépliage de molécules uniques sont particulièrement adaptées à l'étude de protéines telles que les protéines musculaires (comme la titine), de la matrice extracellulaire (comme la téna-scine), ou du cytosquelette¹. En effet, ces protéines sont soumises dans leur environnement biologique à des forces importantes et leurs propriétés mécaniques doivent être adaptées à la résistance à de telles tensions, que l'on peut reproduire expérimentalement. De plus, d'un point de vue pratique, la grande taille de ces protéines permet de les manipuler relativement aisément à l'échelle de la molécule unique.

Un résultat typique d'une expérience de dépliage induit par étirement est représenté figure 2.3 sur l'exemple d'un fragment I27 de la titine (Linke *et al.*, 2002). La titine est une protéine musculaire géante constituée de 300 domaines immunoglobuline (Ig) et fibronectine (FN3) connectés par des régions PEVK². La figure 2.3 A montre l'évolution de la force à appliquer pour maintenir une extension donnée de la molécule I27. On distingue clairement un profil caractéristique en dents de scie, dans lequel chaque pic est provoqué par le dépliage d'un domaine Ig : la force à appliquer augmente jusqu'à atteindre une valeur suffisante pour déplier entièrement le domaine, après quoi elle diminue brusquement. Le point 4 correspond à la rupture de la liaison entre le support et la molécule. L'interprétation à l'échelle atomique de ce phénomène est indiquée sur la figure 2.3 B. Il est possible de reproduire, à l'aide de modèles de physique des polymères, la courbe de la figure 2.3 et d'en déduire des informations sur la structure et la rigidité de l'édifice moléculaire. Dans cet exemple, le modèle WLC³ (courbe rouge) reproduit raisonnablement la courbe expérimentale et la distance entre deux sommets permet de déterminer que la longueur d'un module Ig déplié est d'environ 28,1 nm (la distance entre deux pics).

Les expériences de dépliage réalisées permettent ainsi d'évaluer la résistance des protéines à une traction, ainsi que de déterminer les domaines de la protéine les plus sensibles à l'application d'une contrainte mécanique. Couplées à des techniques de modélisation moléculaire, elles permettent de plus d'obtenir des informations sur les modifications structurales qui ont lieu lors de l'application des contraintes de torsion ou de traction (voir paragraphe 2.3.2 et (Lavery *et al.*, 2002 ; Tajkhorshid *et al.*, 2003)).

1. réseau de fibres intracellulaire qui maintient la structure de la cellule.

2. régions composées de 70 % de prolines, d'acides glutamiques, de valines, et de lysines.

3. WLC : *Worm-Like Chain model*.

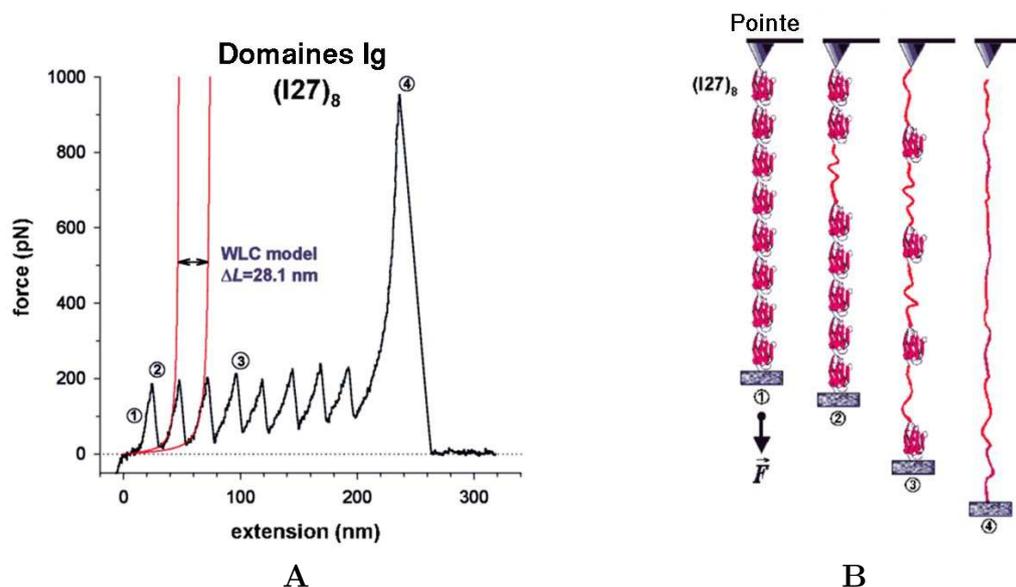


FIGURE 2.3 – Dépliement de la titine par microscopie à force atomique. **A** : Courbe force-extension ; **B** : Interprétation à l'échelle moléculaire. Les figures de 1 à 4 correspondent respectivement aux points 1 à 4 de la figure A. Reproduit de <http://www.uni-muenster.de/Biologie.AllgmZoo/AG-Linke/PAGES/GENERAL/RESEARCH/research6.htm>.

2.2.2 Séparation de complexes

Les premières utilisations de la microscopie à force atomique en biologie ont consisté à étudier la rupture des interactions protéine-ligand entre l'avidine et la biotine (Florin *et al.*, 1994). Depuis, les expériences de nanomanipulation ont été appliquées à l'étude d'autres complexes biologiques comme des complexes antigène-anticorps (Schwesinger *et al.*, 2000).

Les protéines d'adhésion cellulaire sont des candidates idéales à l'utilisation des expériences d'étirement afin de tester la force des interactions qui rendent possible la cohésion entre deux cellules. Dans ce cadre, Perret et collaborateurs ont réalisé des expériences de BFP sur des molécules de E-cadhérine¹ (Perret *et al.*, 2002 ; Perret *et al.*, 2004), dont l'étude théorique sera présentée dans la partie II. Le montage expérimental utilisé est montré sur la figure 2.4 A : on voit sur la gauche, une pipette pressurisée reliée à une cellule de globule rouge qui joue le rôle de transducteur de force (schématisé par un ressort blanc), elle-même attachée à une micro-bille de verre portant les molécules de cadhérine représentées par des petits points bleus ; sur la droite du dispositif, une microbille de verre porte des molécules de cadhérine cibles (en rouge). Dans les expériences de BFP, on mesure

1. E-cadhérine : cadhérine épithéliale.

la force de rupture des interactions entre les molécules situées sur les deux microbilles de verre. Les interactions entre les domaines extracellulaires entiers (EC1-5) des cadhérines ainsi qu'entre les deux domaines EC1-2 les plus éloignés de la membrane (suspectés de jouer un rôle majeur dans la formation de contacts inter-cadhérines) ont été testées (voir figure 2.4 B pour un schéma de ces deux fragments). Les résultats obtenus au cours de cette étude ont révélé une faible interaction entre les domaines EC1-2, ainsi qu'une variété d'interactions plus fortes pour l'ensemble du domaine extracellulaire. Une interprétation possible de ces résultats est que les domaines EC1-2 jouent un rôle important dans les premiers instants de l'adhésion, tandis que les autres domaines sont nécessaires dans un second temps pour renforcer les interactions au sein du complexe.

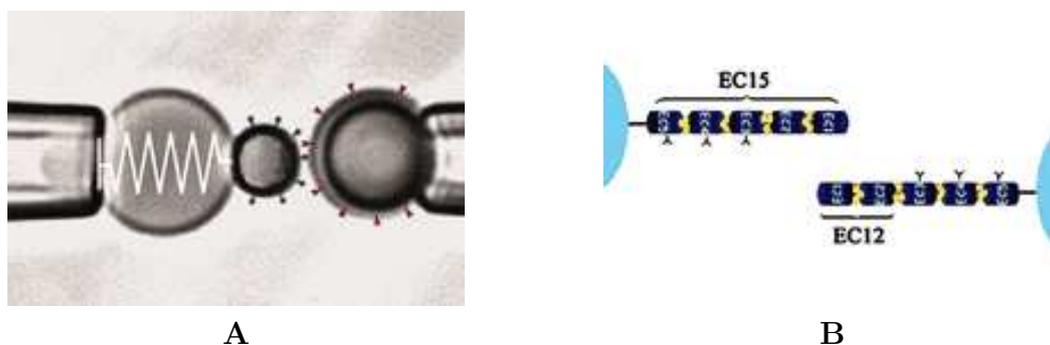


FIGURE 2.4 – Étude des interactions entre molécules de E-cadhérine par BFP. **A** : Montage expérimental ; **B** : Schéma des molécules de E-cadhérine. Reproduit de (Perret *et al.*, 2004).

2.3 Étude théorique des propriétés mécaniques des protéines

2.3.1 Analyses par modes normaux

L'analyse des modes normaux est utilisée depuis le début des années 1980 pour l'étude des macromolécules biologiques (Tama *et al.*, 2000). Cette méthode, qui consiste à déterminer les modes de vibration de la molécule (voir annexe E), a obtenu de grands succès, en permettant de retrouver de nombreux résultats expérimentaux. Initialement utilisée en adoptant une représentation "tout-atome" des protéines très coûteuse, elle est maintenant aujourd'hui fréquemment utilisée en combinaison avec une représentation simplifiée de type réseau de ressorts (voir annexe B) qui permet de retrouver l'essentiel des résultats.

Il est possible, en connaissant l'ensemble des fréquences de vibration d'une molécule, de retrouver les facteurs de température expérimentaux (voir figure 2.5), à partir de la formule suivante :

$$\langle \Delta r_i^2 \rangle = \frac{k_B T}{m_i} \sum_j \frac{a_{ij}^2}{\omega_j^2}$$

qui donne le carré des fluctuations atomiques de i pour une température T . m_i est la masse de l'atome i , j parcourt l'ensemble des modes normaux de la molécule, a_{ij} est la projection du mode normal j sur la coordonnée r_i et ω_j est la fréquence de ce mode j . Il n'est en général pas nécessaire d'utiliser tous les modes normaux pour ce calcul, car les modes normaux de basse fréquence sont responsables de la majeure partie des déplacements atomiques.

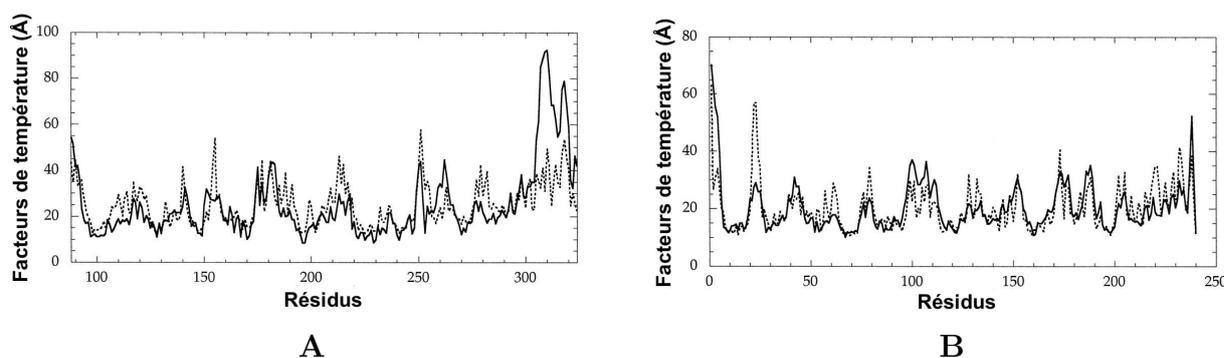


FIGURE 2.5 – Comparaison de facteurs de température obtenus théoriquement par analyse de modes normaux (trait plein) et expérimentaux (trait pointillé) pour un fragment de la protéine CysB liant le cofacteur (A) et la protéine LAO (B). Reproduit de (Keskin *et al.*, 2000).

Les modes normaux ont aussi été utilisés avec succès pour retrouver les changements conformationnels de grande amplitude au sein des protéines observés par cristallographie, comme les mouvements de domaines. Tama et collaborateurs (Tama et Sanejouand, 2001) ont réalisé une étude sur un ensemble de 20 protéines dont différentes conformations (une conformation "ouverte" et une conformation "fermée") étaient connues en comparant les modifications structurales observées expérimentalement avec celles prédites par les modes normaux de plus basses fréquences. Le tableau 2.1, reproduit de cette étude, présente le recouvrement entre le changement conformationnel observé par la cristallographie et le mode normal reproduisant le mieux ce changement. Cette grandeur, qui varie entre 0 et 1, mesure la similarité entre la direction du changement conformationnel et la direction de déplacement le long du mode normal j considéré : plus le recouvrement est proche de 1, plus les directions sont semblables (une valeur de 1 correspondrait à des directions de déplacement

alignées et une valeur de 0 à des déplacements perpendiculaires). Le recouvrement R_j pour le mode j est calculé à l'aide de la formule :

$$R_j = \frac{\left| \sum_{i=1}^{3N} a_{ij} \Delta r_i \right|}{\sqrt{\left[\sum_{i=1}^{3N} a_{ij}^2 \sum_{i=1}^{3N} \Delta r_i^2 \right]}}$$

où $\Delta r_i = r_i^o - r_i^f$, r_i^o et r_i^f étant les i^e coordonnées atomiques de la protéine dans les conformations "ouverte" et "fermée", a_{ij} est la i^e coordonnée du mode normal j , et N est le nombre d'atomes de la molécule.

Protéine	Recouvrement	
	Conformation "ouverte"	Conformation "fermée"
Citrate synthase	0,83	0,57
Calmoduline	0,50	0,37
Protéine Che Y	0,32	0,34
Dihydrofolate réductase	0,72	0,64
Toxine de la diphtérie	0,58	0,37
Énolase	0,33	0,30
LAO	0,84	0,40
Triglycéride lipase	0,30	0,17
Protéine fixant la maltodextrine	0,86	0,77
Thymidilate synthase	0,56	0,40

TABLE 2.1 – Recouvrement entre le changement conformationnel observé par cristallographie et le mode normal reproduisant le mieux ce changement. Reproduit de (Tama et Sanejouand, 2001).

Dans la majorité des cas, il existe un mode normal permettant d'expliquer plus de 50 % du changement conformationnel observé, indiquant que les mouvements globaux dans ces protéines suivent approximativement les directions des modes normaux.

Récemment, les modes normaux ont même permis de prédire l'existence de différents conformères d'une enzyme impliquée dans la synthèse d'acide gras (Brink *et al.*, 2004), montrant par là leur utilisation possible dans les protocoles de raffinement de structures.

2.3.2 Dynamique moléculaire et minimisation d'énergie

2.3.2.1 Dynamiques moléculaires "libres"

La dynamique moléculaire est certainement la méthode de simulation la plus employée actuellement pour étudier les molécules biologiques. Les principes de cette technique sont détaillées plus précisément dans l'annexe D, et nous nous intéresserons ici à quelques-unes de ses applications permettant de retrouver certains résultats expérimentaux.

La première mesure expérimentale accessible à l'aide de la dynamique moléculaire est le facteur de température. Le calcul de ces B-facteurs se fait en utilisant directement la formule donnée dans la section 2.1, après superposition des structures de la trajectoire. Les tendances dans l'évolution des facteurs calculés théoriquement reproduisent de façon raisonnable celles observées expérimentalement, même si les valeurs obtenues par dynamique moléculaire sont sensiblement inférieures aux valeurs expérimentales (voir figure 2.6 A).

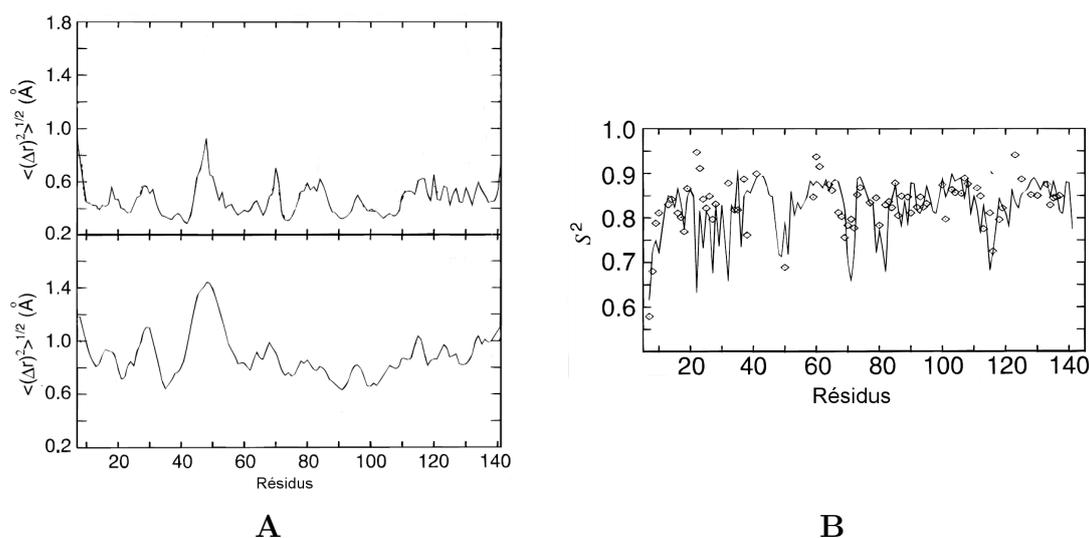


FIGURE 2.6 – Comparaison entre les mesures extraites d'une simulation de dynamique moléculaire et expérimentales sur l'exemple de la nucléase de staphylocoque. **A** : Fluctuations atomiques : le diagramme supérieur montre les fluctuations calculées à partir d'une simulation de dynamique moléculaire et le diagramme inférieur montre celles extraites des données cristallographiques ; **B** : Paramètres d'ordre généralisés des liaisons $N - H$ du squelette peptidique : la courbe correspond aux valeurs calculées par simulation et les diamants représentent celles obtenues par RMN. Reproduit de (Chatfield *et al.*, 1998).

Il est également possible de retrouver théoriquement les valeurs de paramètres d'ordre obtenus par RMN. En suivant l'évolution d'un vecteur joignant deux atomes liés au cours d'une trajectoire, on peut calculer à partir des équations fournies dans la section 2.1 le pa-

ramètre d'ordre associé à ce vecteur. La figure 2.6 B, reproduite de (Chatfield *et al.*, 1998), montre une comparaison entre les paramètres d'ordre issus d'une simulation de dynamique moléculaire et expérimentaux pour la nucléase de staphylocoque. L'accord entre les deux jeux de données est globalement bon avec une valeur moyenne des paramètres d'ordre de 0,82 et 0,83 respectivement pour la simulation et l'expérience (Chatfield *et al.*, 1998).

Les mouvements de domaines au sein d'une protéine sont parfois accessibles par les méthodes de dynamique moléculaire, même si celles-ci souffrent d'un problème d'échelle de temps, puisque les simulations ne couvrent en général que quelques nanosecondes. Cependant, les techniques d'analyse en composantes principales, qui permettent de déterminer les mouvements collectifs de la protéine à partir d'une trajectoire de dynamique moléculaire (voir annexe E), peuvent permettre d'atteindre de telles informations (Hayward *et al.*, 1997).

2.3.2.2 Utilisation de contraintes

Les limitations temporelles imposées par la dynamique moléculaire peuvent être partiellement levées par l'utilisation de contraintes géométriques qui permettent à certains mouvements trop lents de se dérouler à l'échelle de la nanoseconde en abaissant l'énergie d'activation le long du chemin conformationnel. Elles permettent ainsi d'étudier des phénomènes hors de portée des dynamiques moléculaires à l'heure actuelle (dépliement de protéines, changements conformationnels importants,...).

D'un point de vue pratique, l'application d'une contrainte dans une simulation en modélisation moléculaire consiste à ajouter au potentiel énergétique un terme (le plus souvent quadratique) qui va favoriser les structures qui satisfont la contrainte. La nature de cette contrainte peut être variée et impliquer une distance entre atomes ou groupes d'atomes (Keepers *et al.*, 1984), un angle de valence (Ravindranathan *et al.*, 2005), un angle dièdre (Banavali et MacKerell, 2002 ; Giudice *et al.*, 2001), ou encore une mesure plus globale comme le rayon de giration (Hunenberger *et al.*, 1995) ou la distance à rapport à une structure cible (Gilquin *et al.*, 2000 ; Ma *et al.*, 2000). Le plus souvent, l'application de la contrainte se fait par étapes successives afin de ne pas introduire de perturbations trop importantes dans le système et pouvoir observer un chemin réactionnel le plus favorable énergétiquement possible.

L'utilisation des contraintes est également nécessaire pour reproduire les effets des expériences de nanomanipulation à l'échelle de la molécule unique. L'avantage de coupler cette

approche numérique à ce type d'expériences est de pouvoir interpréter au niveau atomique les changements conformationnels induits par l'application d'une force sur la molécule. Dans ce cas, la contrainte utilisée consiste à étirer une distance entre deux atomes, mimant la force à laquelle est soumise la molécule lors des nanomanipulations. Cette contrainte peut être appliquée soit dans le cadre de minimisations d'énergie, soit dans le cadre de simulations de dynamique moléculaire.

Les méthodes de minimisation d'énergie permettent d'obtenir des informations structurales et énergétiques mais ne donnent pas accès à l'aspect temporel de l'étirement de la molécule. Elles sont cependant très utiles pour interpréter les expériences de déformation de l'ADN et ont permis de révéler l'existence de nouvelles structures de la double hélice quand celle-ci est fortement étirée (Cluzel *et al.*, 1996 ; Lebrun et Lavery, 1996 ; Lebrun et Lavery, 1998). En ce qui concerne les protéines, cette technique a également été employée pour étudier de façon théorique les forces nécessaires pour détruire de petits éléments de structure secondaire (Rohs *et al.*, 1999).

La dynamique moléculaire d'étirement¹ (Isralewitz *et al.*, 2001a ; Isralewitz *et al.*, 2001b) a été largement utilisée pour analyser les déformations d'une protéine soumise à une force (Lu *et al.*, 1998 ; Lu et Schulten, 1999 ; Lu *et al.*, 2000), ou la séparation entre un ligand et une protéine (Grubmüller *et al.*, 1996 ; Izrailev *et al.*, 1997 ; Bayas *et al.*, 2003 ; Bayas *et al.*, 2004).

Il existe plusieurs méthodes permettant d'étirer la protéine (ou d'induire la séparation de deux molécules) en dynamique moléculaire :

– **Étirement à vitesse constante :**

Dans ce cas, la force extérieure ajoutée au champ de force est dépendante du temps, et s'exprime comme :

$$F = K(z_0 + vt - z)$$

où z est la coordonnée de réaction selon laquelle l'étirement est réalisé, K une constante de force et v la vitesse constante d'étirement.

– **Étirement à force constante :**

La force extérieure ajoutée est indépendante du temps et conserve une valeur unique tout au long de la simulation.

L'avantage de la simulation à vitesse d'étirement constante est qu'elle reproduit directement les contraintes appliquées lors d'une expérience d'AFM (à l'exception de l'ordre de

1. connue sous l'acronyme SMD pour *Steered Molecular Dynamics*.

grandeur de v , qui est surévaluée dans les simulations du fait des limitations temporelles de la dynamique moléculaire). Les simulations à force constante, en revanche, ont l'avantage de permettre l'observation d'intermédiaires de dépliement qui survivent plus longtemps, sans pour autant altérer qualitativement les chemins réactionnels.

Ces simulations d'étirement permettent de retrouver les profils d'extension observés expérimentalement et d'en donner une interprétation à l'échelle atomique. Le mécanisme de dépliement du domaine I27 de la titine a ainsi été étudié par des simulations de ce type (Lu et Schulten, 2000). Expérimentalement, le processus de dépliement se déroule en trois phases. Au cours de la première, le domaine est étendu d'une longueur de 6 à 12 Å. La deuxième phase est caractérisée par une faible élongation malgré l'application d'une force, jusqu'à un point de rupture à partir duquel on observe une nouvelle élongation de la protéine. Des dynamiques d'étirement réalisées à force constante ont permis de retrouver ce profil (voir figure 2.7 A). De plus, elles ont permis de relier l'existence des trois phases avec la rupture de deux réseaux de liaisons hydrogène qui stabilisent la structure en feuillet β du domaine Ig (figure 2.7 B). La première phase correspond ainsi à la rupture des liaisons hydrogène entre les brins A et B (voir figure 2.7 C) et la seconde élongation à la rupture des six liaisons entre le brin A' et le brin G (figure 2.7 D).

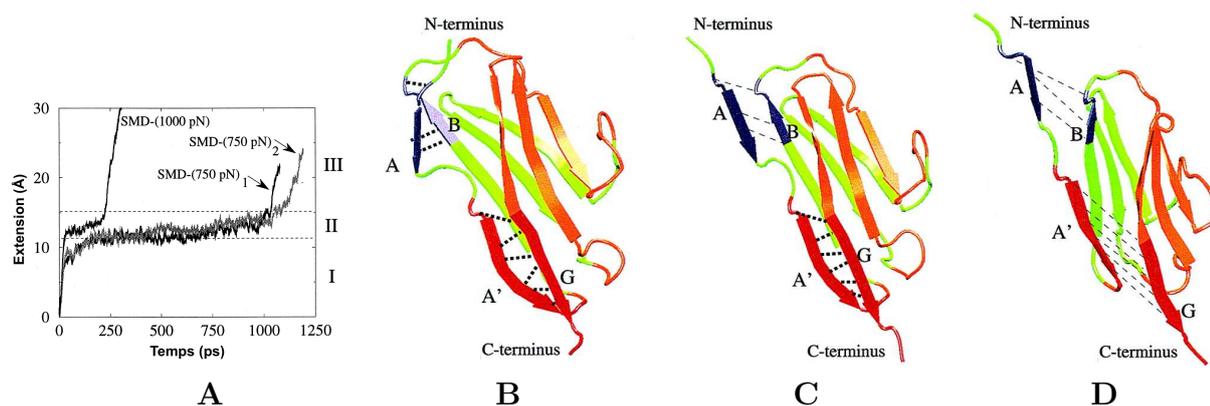


FIGURE 2.7 – Dynamiques d'étirement du domaine I27 de la titine **A** : courbes d'extension du domaine I27 au cours du temps pour 3 simulations d'étirement à force constante; **B** : structure native du domaine I27 montrant les deux réseaux de liaison hydrogène; **C** : structure intermédiaire pour une extension de 10 Å montrant la rupture des liaisons hydrogène entre les brins A et B (traits fins) tandis que les brins A' et G sont toujours reliés entre eux; **D** : structure du domaine I27 pour une extension de 25 Å montrant la rupture des liaisons hydrogène entre les brins A' et G. Figures reproduites de (Lu et Schulten, 2000).

Chapitre 3

Définition d'un nouvel indicateur théorique pour mesurer la flexibilité des protéines

Sommaire

3.1	Limitation des mesures actuelles	39
3.2	Méthodologie	40
3.3	Un nouvel indicateur de rigidité	43
3.3.1	Propriétés mécaniques d'une hélice α	43
3.3.2	Étude d'une protéine globulaire : la nucléase de staphylocoque	44
3.4	Définition de domaines dynamiques	48
3.4.1	Méthodologie	49
3.4.2	Domaines mécaniques de SNase	52
3.4.3	Étude d'une série de protéines	54
3.4.4	Positionnement des résidus rigides	55
3.5	Conclusion	56
3.6	Perspectives	58

3.1 Limitation des mesures actuelles

Bien que les fluctuations atomiques, obtenues expérimentalement (facteurs de température extraits des données cristallographiques) ou théoriquement donnent des indications sur la flexibilité des protéines, cette information est limitée à des changements conformationnels de petite amplitude liés à l'agitation thermique et est dominée par l'effet d'empaquetage

(Halle, 2002). La dynamique moléculaire peut donner accès à des informations sur des mouvements de plus grande amplitude, mais ceux-ci sont toujours limités par l'agitation thermique et confinés à des échelles de temps de l'ordre de quelques nanosecondes. Il est par ailleurs également difficile d'extraire des informations sur la dynamique globale de la protéine d'une trajectoire, bien que des techniques existent qui permettent de déconvoluer les mouvements de domaines et de repérer les résidus charnières au sein de l'édifice structural (Hayward *et al.*, 1997). L'analyse des modes normaux de basse fréquence peut aussi fournir des informations sur les mouvements entre deux domaines (Marques et Sanejouand, 1995 ; Hinsen, 1998 ; Thomas *et al.*, 1999), de même que l'étude des comparaisons entre plusieurs conformations d'une protéine (Lesk et Chothia, 1984 ; Gerstein *et al.*, 1994 ; Wrighers et Schulten, 1997 ; Navizet *et al.*, 2004b).

En dépit des progrès réalisés, notamment grâce à l'utilisation des nanomanipulations, il n'existe donc toujours aucun moyen de sonder les propriétés mécaniques des protéines à l'échelle des résidus. De telles mesures sont pourtant nécessaires afin d'analyser comment les protéines ont acquis leurs comportements dynamiques. Idéalement, celles-ci fourniraient un indicateur mesurant la difficulté à déplacer chaque résidu, ainsi que les directions de moindre résistance lors de ces déformations.

Les expériences de nanomanipulation et les simulations qui leur sont associées sont souvent limitées à l'étude d'un seul aspect de la mécanique des protéines, à savoir l'effet d'un étirement entre les deux extrémités C et N-terminales de la molécule. Bien que certains travaux aient étudié l'application d'une force sur des résidus non terminaux sur de petits peptides, elles ont tout de même été limité à sonder la mécanique globale du système (Rohs *et al.*, 1999 ; Bryant *et al.*, 2000 ; Idiris *et al.*, 2000 ; Masugata *et al.*, 2002). Comment est-il possible d'obtenir des informations sur les propriétés mécaniques d'une protéine à un niveau plus fin, idéalement à l'échelle du résidu ?

3.2 Méthodologie

Plusieurs approches peuvent être envisagées pour sonder la rigidité d'une protéine à l'échelle du résidu à l'aide de simulations numériques. Elles se heurtent cependant toutes à un problème fondamental : pour mesurer une résistance mécanique, il est nécessaire de pousser ou tirer par rapport à quelque chose. La première façon de résoudre ce problème est de réaliser un étirement des distances séparant chaque paire de résidus. Cette approche a cependant deux désavantages majeurs. Tout d'abord, pour une protéine contenant N acides aminés, il est nécessaire de réaliser $O(N^2)$ simulations. De plus, chacune des données obte-

nues se référant à deux résidus, il est difficile de dériver une information sur la déformation locale au niveau d'un des deux résidus en question.

Au contraire, s'il est possible d'agir directement sur un résidu unique, le nombre de simulations nécessaires devient proportionnel à N et la mesure est directement reliée à un seul résidu. La question demeure cependant : comment agir sur un résidu unique ? Un choix envisageable est de déplacer un résidu par rapport au centre de masse de la protéine. Des résultats utilisant cette approche ont montré qu'il n'était pas toujours possible de relier ces expériences avec les propriétés mécaniques locales liées au résidu sondé. Ceci s'explique par le fait qu'une contrainte appliquée sur la distance entre un résidu et le centre de masse de la protéine peut être satisfaite en déplaçant soit le résidu, soit le centre de masse. Ainsi, si l'on considère une protéine possédant une région très flexible ou composée de deux ou plusieurs domaines reliés entre eux par des charnières flexibles, une telle contrainte sera satisfaite principalement par un mouvement relatif des domaines ou de la région flexible, entraînant ainsi un déplacement du centre de masse. La gamme de variation de l'indicateur de flexibilité mesuré est alors réduite ne permettant pas par exemple de différencier les résidus situés au sein de domaines rigides des résidus charnières (voir chapitre 3.3). Il est possible de surmonter cette difficulté en modifiant la contrainte appliquée au résidu sondé. Il faut en fait agir non pas sur une simple distance mais sur la moyenne des distances entre le résidu sondé et les autres résidus de la protéine (la distance entre deux résidus est définie comme la distance séparant les C_α de chacun des résidus). La distance D_i qui sera contrainte aux cours des simulations est définie comme la moyenne des distances r_{ij} entre le résidu sondé i et tous les autres résidus j :

$$D_i = \sum_{j^*} \frac{r_{ij}}{N - 3}$$

N représente le nombre de résidus de la protéine. L'indice j parcourt l'ensemble j^* contenant tous les résidus à l'exception des résidus $(i - 1)$, i , et $(i + 1)$, car la distance entre deux résidus voisins est presque constante.

En contraignant D_i à adopter des valeurs à la fois inférieures et supérieures à la valeur de référence calculée pour la structure native, il est possible d'obtenir des données sur les propriétés mécaniques du résidu i dans son environnement protéique. Nous avons choisi de faire varier D_i sur une gamme allant de $-0,2$ à $+0,2$ Å par rapport à la valeur de référence, par pas de $0,1$ Å. Pour chaque résidu et chaque nouvelle valeur de D_i , l'énergie de la structure est minimisée afin d'adapter la conformation de la protéine à la valeur imposée

pour D_i . La plage de déformation a été choisie afin d'être compatible avec des déformations dues à l'agitation thermique, mais il est également possible d'étudier des mouvements de plus grande amplitude et même de dénaturer partiellement la protéine. Afin de contraindre D_i à adopter une valeur donnée, le champ de force doit être modifié par l'ajout d'un terme quadratique favorisant la valeur D_i choisie :

$$E_{pen} = k (D_i - (D_i^0 + \Delta D_i))^2$$

où D_i^0 est la valeur de D_i dans la structure initiale et ΔD_i la variation imposée à D_i . k est une constante de force qu'il faut ajuster pour que la contrainte sur D_i soit respectée.

Après la minimisation d'énergie, la structure de la protéine déformée peut être superposée sur la structure initiale afin de déterminer la direction de déplacement du résidu sondé. Un avantage de cette contrainte est qu'elle n'est pas de nature vectorielle mais uniquement scalaire. En d'autres termes, aucune direction de déformation n'est imposée *a priori* et les déplacements induits indiquent donc le chemin de moindre résistance. De même, rien n'impose que les déplacements observés pour différentes valeurs de ΔD_i soient alignés. À partir de la courbe de variation d'énergie en fonction de ΔD_i ¹, il est possible de dériver une constante de force en modélisant cette courbe par une fonction quadratique en ΔD_i (voir figure 3.1). Cette constante de force est d'autant plus grande que l'énergie nécessaire pour faire varier D_i d'une valeur donnée est importante. La constante de force ainsi calculée caractérise donc la résistance du résidu à la déformation.

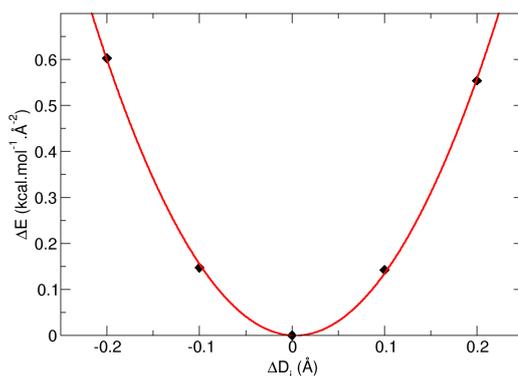


FIGURE 3.1 – Obtention de la constante de force par régression quadratique. La courbe rouge représente la meilleure fonction quadratique pour reproduire les constantes de force obtenues lors des simulations (diamants noirs).

1. L'énergie en question est calculée sans tenir compte du terme de pénalité quadratique en D_i .

Dans un premier temps, la méthode a été testée en utilisant un champ de force "tout-atome" (voir annexe B.2) sur des systèmes simples (une hélice α et une petite protéine globulaire), en utilisant le programme LIGAND, dans lequel la structure de la protéine est définie à l'aide de coordonnées internes (voir annexe C). Dans le but d'étudier un nombre plus important de protéines, nous avons ensuite appliqué la même méthode en combinaison avec une représentation simplifiée des protéines dans laquelle les résidus sont représentés par un seul site (positionné à la place du C_α) et les interactions entre eux modélisées par un réseau de ressorts (voir annexe B.3).

3.3 Un nouvel indicateur de rigidité

3.3.1 Propriétés mécaniques d'une hélice α

Notre premier système test a consisté en une hélice poly-alanine de 13 résidus, dont nous avons préalablement minimisé l'énergie. Nous avons appliqué ensuite notre contrainte à chaque résidu en utilisant le protocole décrit dans la section 3.2. La figure 3.2 montre l'histogramme des constantes de force calculées ainsi qu'une représentation schématique de la déformation du squelette carboné produite lorsque la contrainte est appliquée sur un résidu central de l'hélice ou sur un résidu proche de l'extrémité C-terminale. Dans la gamme relativement restreinte de déformation étudiée, on observe que le déplacement des résidus se produit selon une direction constante quelle que soit la valeur de la contrainte appliquée.

Les résultats de la figure 3.2 montrent que les constantes de force diminuent plus ou moins linéairement du centre de l'hélice vers les extrémités. La valeur maximale est observée pour le résidu central ($1,98 \text{ nN.Å}^{-1}$), approximativement huit fois supérieure à celle des résidus terminaux (environ $0,25 \text{ nN.Å}^{-1}$). Ceci indique, comme attendu, qu'il est plus difficile de perturber une hélice en son centre, du fait du réseau de liaisons hydrogène plus dense à cet endroit.

Si l'on s'intéresse maintenant aux déplacements induits par la contrainte, on peut voir que l'hélice préfère se plier lorsqu'on sonde un résidu central, tandis qu'elle a tendance à s'étirer longitudinalement lorsqu'on sonde un résidu terminal (voir figure 3.2).

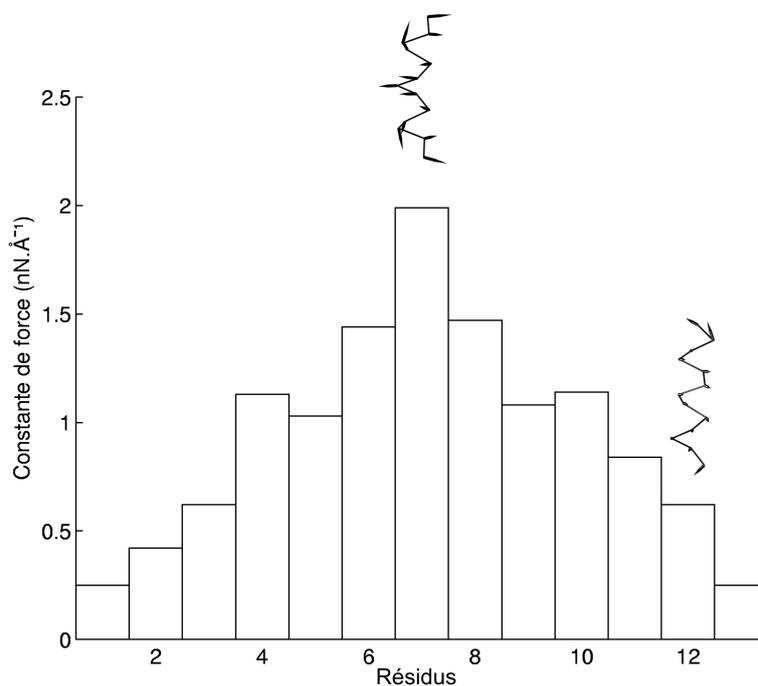


FIGURE 3.2 – Histogramme des constantes de force pour les résidus d’une hélice α (Ala)₁₃. Les schémas insérés indiquent le déplacement des C_α après avoir imposé une contrainte $\Delta D_i = +0,2 \text{ \AA}$ sur le résidu central et sur un résidu proche de l’extrémité C-terminale. La longueur des déplacements a été multipliée par 10 dans un souci de lisibilité.

3.3.2 Étude d’une protéine globulaire : la nucléase de staphylocoque

La première protéine que nous avons étudiée avec cette approche est la nucléase de staphylocoque (notée SNase par la suite). Les raisons de ce choix sont que cette enzyme est de taille relativement petite (149 résidus¹), globulaire, et qu’elle contient à la fois des hélices α et des feuillets β (Chen *et al.*, 2000).

3.3.2.1 Calcul des constantes de force

L’histogramme des constantes de force obtenues pour tous les résidus de la chaîne peptidique est montré sur la figure 3.3. De même que pour l’hélice α , les valeurs observées varient significativement d’un résidu à l’autre, allant de $0,16 \text{ nN.\AA}^{-1}$ pour les résidus les

1. La structure utilisée dans ce travail, extraite de la PDB (code 1EY0), ne contient en fait que les résidus 6 à 141, les deux extrémités N- et C-terminales étant désordonnées dans le cristal. L’ensemble des simulations est donc réalisé sur les 136 résidus définis dans la structure. Il est à noter cependant que ces délétions n’ont pas d’influence sur le repliement et la stabilité de la protéine (Hirano *et al.*, 2002).

plus flexibles à $10,68 \text{ nN.Å}^{-1}$ pour les plus rigides. Il est à noter que cette variation d'un facteur de plus de 50 n'est pas directement corrélée aux structures secondaires indiquées le long de l'abscisse de la figure 3.3, les hélices ou les feuillets ne présentant en effet pas systématiquement des valeurs de constantes de force particulières par rapport aux autres résidus. En fait, les plus grandes constantes de force sont observées pour des résidus positionnés au cœur de la protéine. On peut notamment extraire onze résidus dont la constante de force est supérieure à $3,4 \text{ nN.Å}^{-1}$ (F34, R35, L36, L37, L38, L89, A90, Y91, I92, N100, et L103). Ces résidus sont majoritairement hydrophobes et appartiennent tous à des structures secondaires, mis à part les résidus L37 et L38. Ils ne correspondent pourtant pas exactement à la notion de cœur hydrophobe¹, même si quatre de ces résidus (F34, L36, I92, et L103) font partie du cœur hydrophobe de SNase comme défini dans une étude antérieure (Chen et Stites, 2001). Afin de visualiser la distribution spatiale de la rigidité dans SNase, la figure 3.4 A montre la structure de la protéine, colorée en fonction de la valeur des constantes de force. L'échelle de couleurs bleu \rightarrow vert \rightarrow rouge, qui correspond à une augmentation de la constante de force, indique clairement que les zones les plus rigides se situent au cœur de la protéine dans une zone où les structures secondaires sont bien empaquées.

3.3.2.2 Directions de déformation de la protéine sous contrainte

La figure 3.4 montre le mouvement des C_α soumis à une contrainte $\Delta D_i = +0,2 \text{ Å}$. Dans l'ensemble, les vecteurs ont tendance à pointer vers l'extérieur de la protéine. Ce n'est cependant pas toujours le cas, l'architecture de la protéine amenant à d'importantes déviations dans certaines situations. En effet, si l'on mesure l'angle formé entre la direction de déplacement du C_α et le vecteur reliant le résidu au centre de masse de la protéine, la déviation est en moyenne de 20° , mais peut excéder 50° pour certains résidus situés au centre de la protéine. De plus, contrairement à l'hélice α étudiée précédemment, la quasi-colinéarité des directions de déplacement pour différentes valeurs de la contrainte imposée peut être perdue. Bien que cette déviation soit en moyenne de 10° environ, elle peut en effet atteindre 90° pour certains résidus. Ces deux aspects montrant l'anisotropie des propriétés mécaniques pourraient indiquer des régions d'intérêt spécial dans la structure protéique.

1. Le cœur hydrophobe est souvent défini comme une collection de résidus faiblement accessibles au solvant, appartenant à des régions de structures secondaires et dont les chaînes latérales non polaires interagissent entre elles (Swindells, 1995).

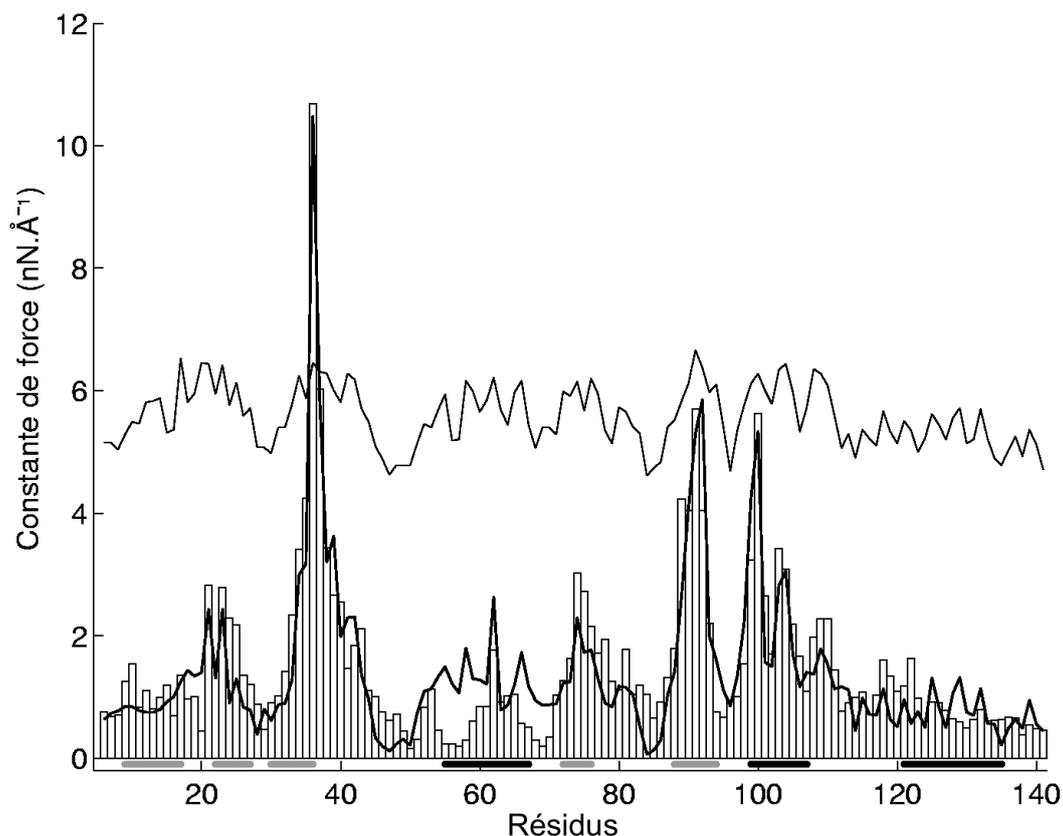


FIGURE 3.3 – Constantes de force calculées pour les résidus de SNase. L'histogramme indique les constantes de force calculées en utilisant une représentation "tout-atome" de la protéine. Les structures secondaires sont indiquées en abscisse par des barres grises pour les feuillets β et noires pour les hélices α . La courbe en trait gras représente les constantes de force calculées en utilisant une représentation granulaire de la protéine, et la ligne en trait fin montre l'inverse des fluctuations atomiques au carré calculées à partir des modes normaux obtenus avec un modèle granulaire de la protéine (ajusté en utilisant un coefficient de proportionnalité et déplacé de $4 \text{ nN} \cdot \text{Å}^{-1}$ selon l'axe des ordonnées pour une meilleure lisibilité).

3.3.2.3 Utilisation d'une représentation simplifiée de la protéine

La méthode développée ici apporte de nouvelles données sur la dynamique des protéines. Cependant, le nombre de simulations à réaliser étant important (en $O(N)$, si N est le nombre de résidus de la protéine), il est difficile de l'appliquer de manière systématique. Il est donc intéressant de se demander si le profil de constantes de force obtenu avec une représentation "tout-atome" de la protéine peut être reproduit à l'aide d'un modèle de représentation protéique plus simple qui permet de diminuer le temps de calcul de chaque simulation. Afin de répondre à cette question, nous avons réalisé les mêmes simulations sur SNase en utilisant une représentation "gros grains" de la protéine, décrite brièvement

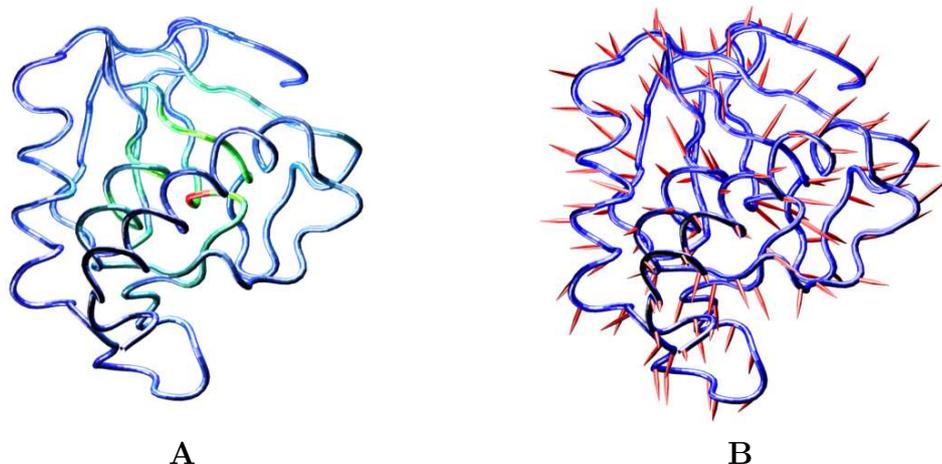


FIGURE 3.4 – Réponse de SNase aux contraintes mécaniques. **A** : Représentation de la chaîne peptidique de SNase colorée en fonction de la constante de force. L'échelle de couleurs bleu \rightarrow vert \rightarrow rouge correspond à une augmentation de la rigidité. **B** : Déplacement des C_α pour une contrainte appliquée de $\Delta D_i = +0,2 \text{ \AA}$. Pour plus de visibilité, la longueur des vecteurs représentant les déplacements (en rouge) a été multipliée par 10.

dans la section 3.2. La ligne en trait gras de la figure 3.3 montre le profil des constantes de force calculées avec ce nouveau modèle, ajusté sur les valeurs obtenues avec le modèle "tout-atome". L'accord entre les deux courbes est remarquablement bon : les différences en valeur absolues entre les deux jeux de valeurs sont, en moyenne, de $0,46 \text{ nN.\AA}^{-1}$ et le coefficient de corrélation est de 0,91 pour 136 points. Les directions de déplacement sont également bien reproduites quelle que soit la valeur de la contrainte appliquée avec une déviation moyenne de seulement 22° . Cependant, des détails plus fins comme la perte de colinéarité des déplacements en fonction de la valeur de la contrainte sont moins bien reproduits par le modèle granulaire. Les résultats sont néanmoins assez encourageants pour justifier l'utilisation de ce modèle nettement plus rapide.

3.3.2.4 Comparaison avec la dynamique moléculaire

Bien que la contrainte utilisée pourrait être appliquée au cours de simulations de dynamique moléculaire, cela serait en pratique beaucoup trop coûteux en temps de calcul. Il est cependant intéressant de se demander s'il existe une corrélation entre les données obtenues à l'aide de notre méthodologie et les fluctuations observées lors de trajectoires de dynamique. Dans ce but, nous avons réalisé une simulation de dynamique moléculaire de

SNase en représentation "tout-atome" et en solvant explicite¹ et comparé nos constantes de force aux fluctuations de la variable D_i au cours de la dynamique. Nous avons pour cela comparé nos constantes de force avec l'inverse du carré des fluctuations de D_i (noté K_i) :

$$K_i = \frac{1}{\langle (D_i - \langle D_i \rangle)^2 \rangle}$$

Le coefficient de corrélation obtenu entre ces deux grandeurs est supérieur à 0,80 pour 136 points, que l'on considère la première ou la seconde moitié de la trajectoire de dynamique (indiquant que ces fluctuations ont bien convergé). Au contraire, il n'existe qu'une très faible corrélation entre les constantes de force et les simples fluctuations des atomes autour de leur position d'équilibre. En effet, le coefficient de corrélation entre les constantes de force et l'inverse des facteurs de température calculés au cours de la dynamique moléculaire n'est que de 0,66. L'accord est encore moins satisfaisant si l'on considère les facteurs de température calculés à l'aide des modes normaux en utilisant un réseau élastique (coefficient de corrélation de 0,47). Ces résultats suggèrent que les propriétés mécaniques mesurées par les constantes de force ne sont pas dues uniquement à l'environnement local de chaque résidu, mais bien liées à l'ensemble de l'architecture protéique.

3.4 Définition de domaines dynamiques

La méthode que nous avons développée avait pour but initial le calcul de constantes de force propres à chaque résidu afin de définir les zones les plus rigides de la protéine. La réponse à la contrainte peut cependant également être analysée plus en détail, en caractérisant la déformation de la structure.

1. La simulation de dynamique moléculaire a été réalisée avec le logiciel AMBER 7.0 (Case *et al.*, 2002) et le champ de force Parm99 (Wang *et al.*, 2000). SNase a été solvatée dans une boîte octaédrique tronquée de molécules d'eau explicites TIP3P (Jorgensen *et al.*, 1983). Le système contient 5655 molécules d'eau et la distance entre les faces de la boîte est d'environ 66 Å. La simulation a été conduite en conditions périodiques en traitant les interactions électrostatiques à l'aide de la méthode PME (Darden *et al.*, 1993 ; Cheatham *et al.*, 1995) en utilisant un rayon de coupure de 9 Å dans l'espace direct et un espacement de grille de 1 Å. L'électroneutralité du système est assurée par l'addition de 14 ions chlorures placés de façon aléatoire dans la cellule de simulation. Le pas d'intégration de la dynamique a été fixé à 2 fs en utilisant des contraintes sur les liaisons impliquant un atome d'hydrogène à l'aide de l'algorithme SHAKE (Ryckaert *et al.*, 1977). Après une minimisation d'énergie, le système est chauffé jusqu'à 300 K en 10 ps, en utilisant des contraintes quadratiques sur les atomes de la protéine, puis gardé à cette température pendant encore 90 ps. Les contraintes sur la protéine sont ensuite relâchées graduellement sur une période de 300 ps, après quoi la trajectoire est continuée pendant 4 ns à température et pression constantes. La stabilisation de la trajectoire est observée au cours de ces 4 ns avec une déviation quadratique moyenne des positions atomiques par rapport à la structure cristallographique d'environ 1,4 Å pour les atomes de la chaîne peptidique et de 1,9 Å pour tous les atomes lourds.

3.4.1 Méthodologie

Si l'on considère une contrainte ΔD_i donnée appliquée au résidu i , on peut représenter les déformations structurales par un vecteur $(|\Delta r_{ij}|)_{j=1,N}$ qui décrit le changement de la distance entre les C_α du résidu i et de tous les résidus j de la protéine par rapport à la structure initiale. En concaténant les différents vecteurs $(|\Delta r_{ij}|)_j$ obtenus à partir des simulations sondant chaque résidu i , on obtient une matrice de $(|\Delta r_{ij}|)_{i,j}$ (voir figure 3.5 A). Cette matrice reflète les liens mécaniques à l'intérieur de la protéine : une petite valeur de $|\Delta r_{ij}|$ indique en effet que les deux résidus i et j sont fortement couplés et se déplacent plus ou moins ensemble lorsque la contrainte est appliquée (il est à noter que cette matrice n'est pas symétrique, c'est-à-dire que $|\Delta r_{ij}|$ peut être différent de $|\Delta r_{ji}|$). De tels résidus sont donc susceptibles d'appartenir à un seul et même domaine "mécanique" ; au contraire, une grande valeur de $|\Delta r_{ij}|$ indique un couplage faible et l'existence de deux domaines distincts. Ceci nous a amené à chercher un moyen de définir des domaines protéiques en se fondant sur la réponse de la protéine à notre contrainte. Les domaines ainsi observés sont à rapprocher des domaines généralement appelés domaines "dynamiques", plutôt que des domaines dits "structuraux", dont la définition repose sur des comparaisons de structures statiques ou de séquences (Lesk et Chothia, 1984 ; Janin et Chothia, 1985 ; Swindells, 1995 ; Orengo *et al.*, 2003). Notre approche est donc reliée aux méthodes qui définissent des domaines sur la base de différentes structures de la même protéine, obtenues expérimentalement ou théoriquement (Wriggers et Schulten, 1997), ou sur la base de données conformationnelles extraites de simulations de dynamique moléculaire (Hayward *et al.*, 1997) ou de calculs de modes normaux (Hinsen *et al.*, 1999). Sur l'exemple de la figure 3.5 A, la matrice de $(|\Delta r_{ij}|)_{i,j}$ montre clairement le découpage en deux domaines mécaniques représentés sur la figure 3.5 B. Le premier domaine est constitué environ des résidus 1 à 100 et 260 à 271, le reste de la structure formant le second domaine. Néanmoins, en vue du traitement d'un nombre important de protéines et pour des cas moins évidents, il est nécessaire d'utiliser des méthodes de regroupement (Jain *et al.*, 1999) afin de définir de manière automatique les domaines dans les matrices de $(|\Delta r_{ij}|)_{i,j}$.

L'algorithme de regroupement que nous avons utilisé pour détecter les domaines regroupe ensemble les résidus pour lesquels les valeurs de $|\Delta r_{ij}|$ sont inférieures à un seuil T . Le premier domaine est constitué en recherchant la plus petite valeur de $|\Delta r_{ij}|$ dans la matrice (dans la mesure où la valeur complémentaire $|\Delta r_{ji}|$ est également inférieure à T). Les résidus i et j forment alors le noyau du premier domaine. Puis nous recherchons le long de la ligne i la prochaine plus petite valeur et ajoutons au domaine le résidu auquel elle correspond. Ce processus est répété jusqu'à ce que plus aucun résidu j ne satisfasse la

relation $|\Delta r_{ji}| \leq T$. Pour chaque nouveau membre ajouté au domaine, il est de nouveau nécessaire qu'à la fois $|\Delta r_{jk}|$ et $|\Delta r_{kj}|$ soient inférieures à T , où k représente chacun des résidus constituant le domaine en question.

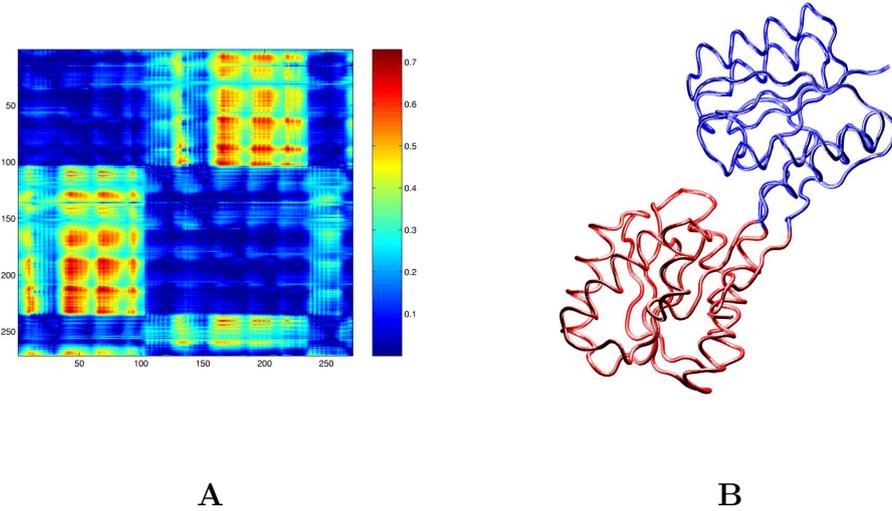


FIGURE 3.5 – Définition de domaines mécaniques sur l'exemple d'une protéine liant le D-ribose. **A** : Matrice de $(|\Delta r_{ij}|)_{i,j}$. **B** : Domaines mécaniques obtenus après tri des colonnes de la matrice.

Bien que la recherche de la valeur minimale de $|\Delta r_{ij}|$ à chaque étape rende cette procédure peu dépendante de l'ordre des résidus, le regroupement est complété en étudiant chaque résidu un par un et en testant s'il ne pourrait pas être inclus dans un autre domaine. Pour un résidu i , un déplacement du domaine A vers le domaine B est accepté si la valeur moyenne $\langle |\Delta r_{ij}| \rangle_{j \in B}$ est inférieure à $\langle |\Delta r_{ij}| \rangle_{j \in A}$ (cette question ne se pose bien sûr que si le résidu i peut effectivement appartenir au domaine B , c'est-à-dire si toutes les valeurs $(|\Delta r_{ij}|)_{i \in B}$ sont inférieures à la valeur seuil T). Cette réaffectation des résidus au sein des différents domaines est répétée cycliquement pour chaque résidu tant que des domaines stables ne sont pas obtenus, ce qui nécessite en général moins de quatre cycles. L'utilisation d'algorithmes de regroupement amène toujours le risque de créer un nombre de groupes trop important. Pour s'affranchir de ce problème et tester la pertinence des domaines définis, une arborescence est calculée qui décrit la hiérarchisation de la structure. La distance entre deux domaines est définie comme la valeur $|\Delta r_{ij}|$ maximale entre les résidus définissant deux domaines distincts (d'après les définitions données plus haut, une telle valeur calculée au sein d'un seul domaine est inférieure à T). Utilisant cette distance, la paire de domaines la plus proche est regroupée au sein d'un "super-domaine" et la procédure est répétée jusqu'à ce que la hiérarchie complète soit construite. En général, comme c'est le cas pour l'exemple de SNase (voir figure 3.6), les branches les plus basses de l'arbre sont séparées par des distances bien supérieures à T .

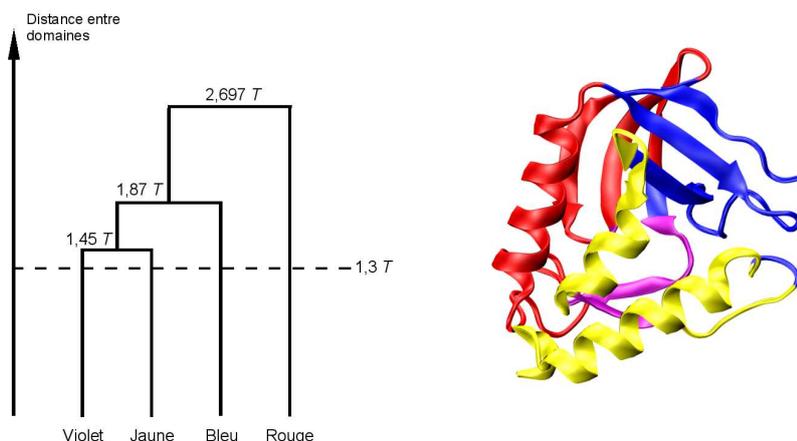


FIGURE 3.6 – Hiérarchisation des domaines mécaniques de SNase.

Si ce n'est pas le cas, les deux domaines sont réunis et la procédure d'affinement décrite au paragraphe précédent est recommencée. Deux groupes sont réunis en un unique groupe si la distance qui les sépare est inférieure à $1,3 \times T$. Cette procédure empêche également la séparation artificielle d'un domaine, qui peut éventuellement survenir lors de la construction initiale des domaines. L'exemple de la DRBP est montré sur la figure 3.7 : après la première phase de regroupement, trois domaines sont définis mais deux d'entre eux sont suffisamment proches pour être fusionnés (domaines rouge et jaune), ce qui mène à deux domaines qui sont bien éloignés après la deuxième phase de raffinement.

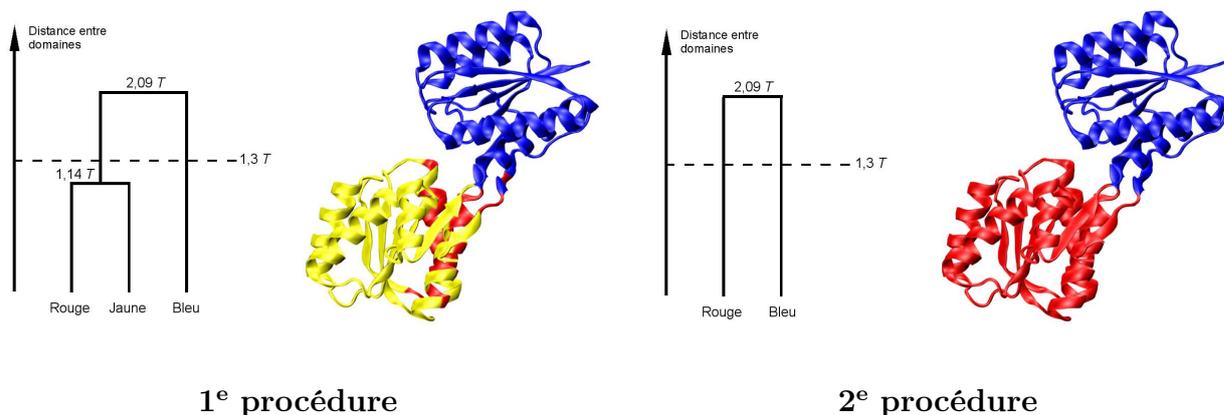


FIGURE 3.7 – Formation des domaines de la DRBP suivant l'algorithme de regroupement. Après la 1^e procédure, trois domaines sont définis mais deux d'entre eux (les domaines rouge et jaune) sont relativement proches et sont donc fusionnés. À la fin de la 2^e procédure, seuls deux domaines sont définis, suffisamment distincts l'un de l'autre.

Afin de déterminer une valeur appropriée pour le seuil T , nous avons observé la variation du nombre de domaines créés en fonction de ce paramètre. Des tests réalisés sur des protéines d'architectures différentes ont montré que ce nombre est à peu près stable pour des valeurs de T comprises entre 0,34 Å et 0,5 Å (voir figure 3.8). En dessous de cette valeur, le nombre de domaines augmente rapidement, conduisant à de petits groupes de résidus sans réelle signification physique. Nous avons donc choisi de fixer la valeur de T à 0,35 Å pour le reste de nos études. Pour quelques protéines (comme la pepsine montrée sur la figure 3.8) le nombre total de domaines peut varier d'une unité pour de petites variations de T . Cependant, de tels changements n'ont qu'un impact réduit sur l'interprétation des propriétés mécaniques.

3.4.2 Domaines mécaniques de SNase

Les résultats de la décomposition en domaines pour la nucléase de staphylocoque sont présentés sur la figure 3.9 A. SNase est découpée en quatre domaines mécaniques, dont trois sont relativement grands (contenant respectivement 36, 46 et 42 résidus) et dont le plus petit ne contient que 12 résidus. Le premier domaine (en bleu) débute à l'extrémité N-terminale et contient la moitié du premier brin β ainsi que la majeure partie des 4^e et 5^e brins et la boucle les reliant. Le 2^e domaine (en rouge) contient le reste du feuillet β à cinq brins, ainsi que la boucle et l'hélice α situées entre le 3^e et le 4^e brin. Le dernier gros domaine (en jaune) est constitué des deux hélices α à l'extrémité C-terminale de la protéine. Le petit domaine restant (en violet) est principalement composé de deux petits brins β situés à l'interface entre le 3^e domaine et le reste de la protéine. Il est à noter que tous les résidus de SNase ont été affectés à l'un des quatre domaines sans modification de la procédure de regroupement. On remarque de plus que les éléments de structure secondaire appartiennent généralement entièrement à un seul domaine, la seule exception notable étant le clivage du feuillet β à cinq brins entre les deux premiers domaines. Ce découpage correspond cependant au réseau de liaisons hydrogène à l'intérieur de ce feuillet.

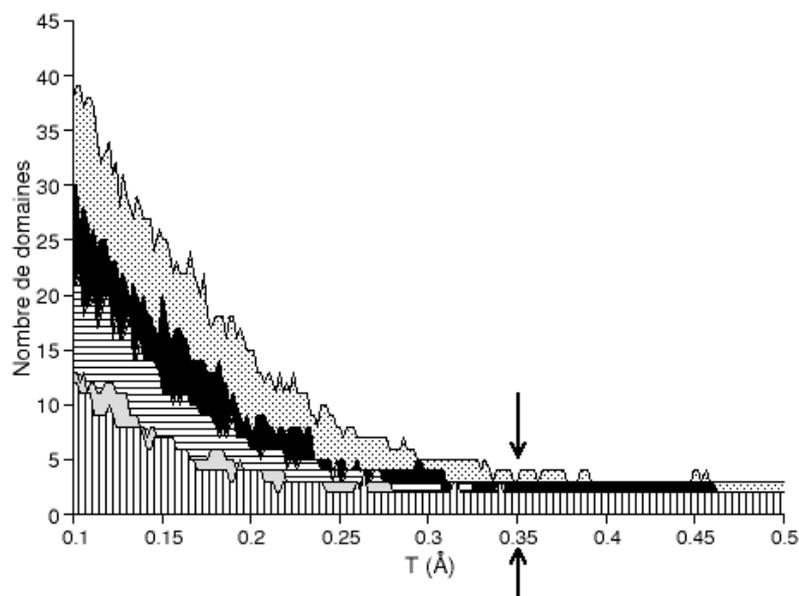


FIGURE 3.8 – Nombre de domaines définis en fonction du paramètre de seuil T . Les courbes du bas vers le haut correspondent respectivement à la caténine, la cadhérine, la DRBP dans sa conformation "ouverte", un dimère de cadhérine, et la pepsine. Les flèches indiquent l'abscisse correspondant à la valeur de 0,35 Å correspondant au seuil utilisé.

Si l'on observe le positionnement des onze acides aminés de grande rigidité définis dans la section 3.3, il est intéressant de noter qu'ils sont situés à l'interface entre les différents domaines (voir figure 3.9 B) :

- les résidus 34 à 38 du troisième brin du feuillet β principal sont situés à l'interface entre les domaines 1 (en bleu) et 2 (en rouge).
- les résidus 89 à 92 du 5^e brin du feuillet β se trouvent à l'interface entre les domaines 1, 2, et 4 (en violet).
- les résidus 100 et 103, qui appartiennent à la première hélice α du domaine 3 (en jaune) sont positionnés à l'interface entre les domaines 2 et 4.

Ce positionnement rejoint les résultats obtenus par l'analyse des facteurs de température calculés à l'aide des modes normaux en représentation "gros grains", qui montrent que les résidus les plus rigides sont situés aux interfaces des domaines dynamiques (Bahar et Jernigan, 1999 ; Isin *et al.*, 2002).

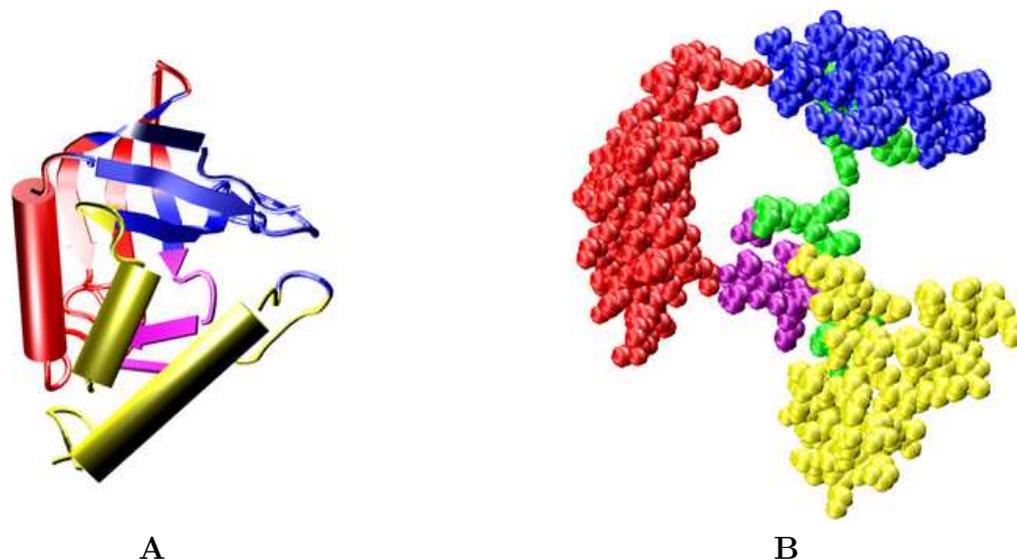


FIGURE 3.9 – Domaines mécaniques de SNase. **A** : Les quatre domaines sont colorés respectivement en bleu, rouge, jaune et violet sur une représentation en structures secondaires de la protéine. **B** : Vue "explodée" de la structure protéique en représentation de van der Waals montrant les différents domaines (colorés de la même façon qu'en **A**). Les résidus possédant les plus grandes constantes de force sont colorés en vert.

3.4.3 Étude d'une série de protéines

Nous avons également appliqué notre méthode à dix protéines de taille variable (de 140 à 420 résidus environ, voir tableau 3.1) afin d'obtenir une vue plus générale de notre approche de détection de domaines mécaniques. L'ensemble des résultats présentés ici ont été obtenus en utilisant une représentation granulaire des protéines.

La figure 3.10 montre la décomposition en domaines obtenue pour chacune de ces protéines. Dans tous les cas, deux à quatre domaines mécaniques sont définis, dont la taille varie entre 14 résidus (hélice α joignant les deux domaines de fixation du calcium de la calmoduline, en rouge dans la figure 3.10 A) et 188 résidus (le plus gros domaine défini pour le dimère de E-cadhérine, en bleu dans la figure 3.10 J). Ces domaines sont généralement composés de segments de la chaîne peptidique séquentiellement contigus, même si un nombre relativement restreint de résidus se retrouvent isolés à l'intérieur de quelques domaines. Nous n'avons pas analysé en détail les domaines obtenus, mais il est clair que lorsque des domaines structuraux visibles existent, ils se comportent comme des domaines mécaniques distincts. Pourtant, dans certains cas, l'origine de l'existence de domaines mécaniques est moins évidente et ces cas nécessiteraient une analyse plus poussée.

Protéine	Fonction	code PDB et référence	Nombre de résidus
Calmoduline	fixation du calcium	1CLL (Chattopadhyaya <i>et al.</i> , 1992)	144
γ B-crystalline	maintien de la transparence du cristallin	4GCR (Najmudin <i>et al.</i> , 1993)	174
Guanylate kinase	transférase	1EX6 (Blaszczyk <i>et al.</i> , 2001)	186
E-cadhérine (monomère)	adhésion cellulaire	1EDH (chaîne A) (Nagar <i>et al.</i> , 1996)	211
LAO	fixation de la lysine, de l'arginine et de l'ornithine	2LAO (Oh <i>et al.</i> , 1993)	238
α -caténine	formation du cytosquelette	1H6G (Yang <i>et al.</i> , 2001)	255
DRBP (forme "ouverte")	transport	1URP (Björkman et Mowbray, 1998)	271
DRBP (forme "fermée")	transport	2DRI (Björkman <i>et al.</i> , 1994)	271
Pepsine	hydrolase	5PEP (Cooper <i>et al.</i> , 1990)	326
E-cadhérine (dimère)	adhésion cellulaire	1EDH (Nagar <i>et al.</i> , 1996)	422

TABLE 3.1 – Protéines étudiées.

3.4.4 Positionnement des résidus rigides

La figure 3.11 montre une vue "explosée" des protéines étudiées dans laquelle les différents domaines sont artificiellement séparés. Les résidus possédant une grande constante de force¹, colorés en vert, sont majoritairement situés à l'interface entre les domaines mécaniques. Comme déjà mentionné précédemment, des résultats similaires ont été obtenus sur la base de calculs de modes normaux utilisant un modèle de représentation granulaire de la protéine (Bahar et Jernigan, 1999 ; Isin *et al.*, 2002). Dans notre cas, ce résultat

1. En nous basant sur l'exemple de SNase, nous avons extrait les 8 % de résidus possédant les plus grandes valeurs de constante de force.

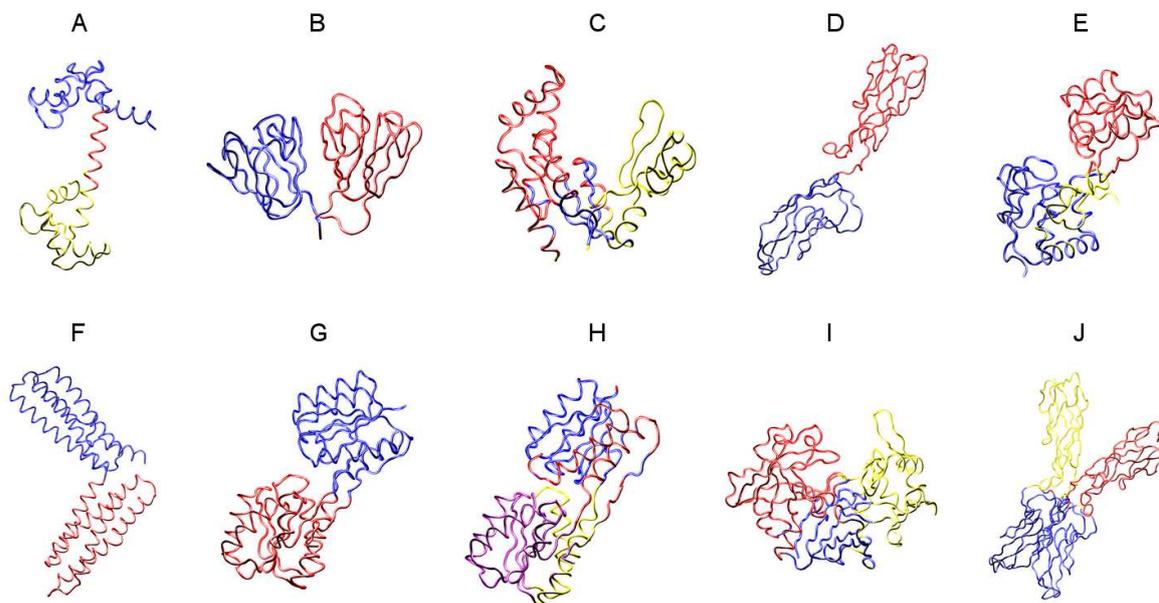


FIGURE 3.10 – Domaines mécaniques de diverses protéines. **A** : calmoduline ; **B** : γ B-crystalline ; **C** : guanylate kinase ; **D** : E-cadhérine ; **E** : LAO ; **F** : α -caténine ; **G** : DRBP (forme "ouverte") ; **H** : DRBP (forme "fermée") ; **I** : pepsine ; **J** : dimère de E-cadhérine. Seule la chaîne peptidique est représentée sous forme de tube.

s'explique en notant que les déplacements les moins défavorables énergétiquement correspondent généralement à des mouvements relatifs de domaines plus ou moins rigides. Ainsi, la réponse de la protéine à une contrainte sur un résidu situé au sein d'un domaine consiste en un mouvement relatif de ce domaine par rapport aux autres domaines constituant la protéine. En revanche, lorsque le résidu i sondé se trouve à l'interface entre plusieurs domaines, un tel mouvement (généralement une rotation) ne se traduit pas par une variation de D_i , puisqu'il ne modifie pas les distances $C_\alpha - C_\alpha$ impliquant les résidus situés aux interfaces entre domaines. Pour satisfaire à la contrainte, il est alors nécessaire de réaliser des déformations intra-domaines beaucoup plus coûteuses du point de vue énergétique.

3.5 Conclusion

La méthodologie que nous avons développée permet d'étudier la réponse d'une protéine à une contrainte locale. En contraignant la distance moyenne entre un résidu donné et tous les autres résidus de la protéine, il est possible de déformer la structure et d'en retirer des informations à la fois scalaires et vectorielles. L'information scalaire est une constante

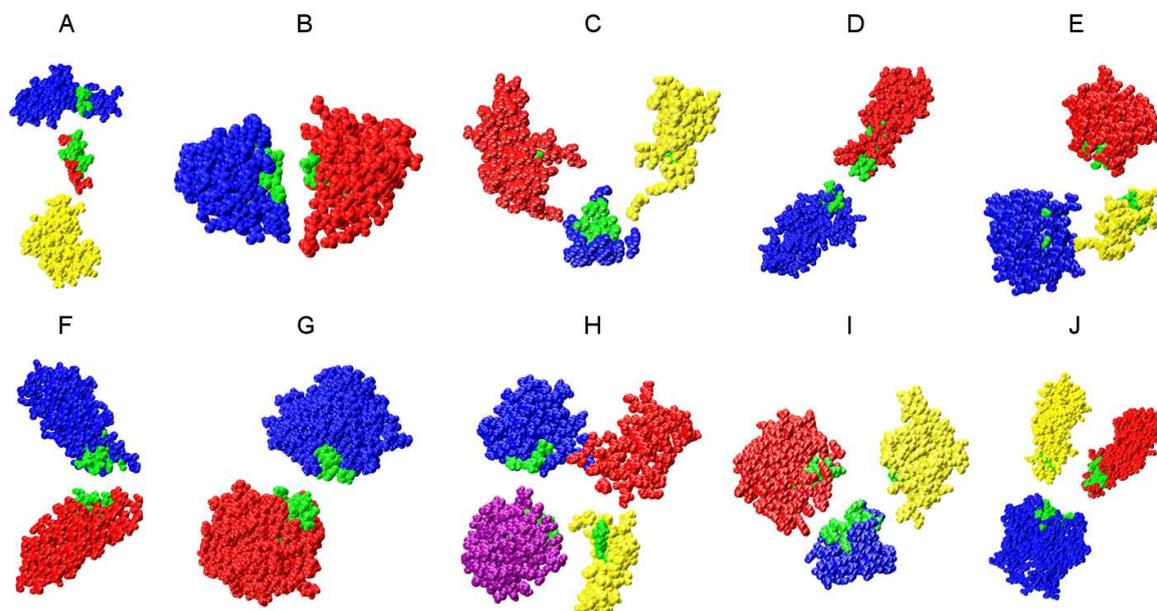


FIGURE 3.11 – Positionnement des résidus rigides à l'intérieur des domaines mécaniques. Les protéines sont vues avec une représentation de van der Waals "explosée" afin de repérer le positionnement des résidus de forte constante de force (en vert) pour les différentes protéines. **A** : calmoduline ; **B** : γ B-crystalline ; **C** : guanylate kinase ; **D** : E-cadhérine ; **E** : LAO ; **F** : α -caténine ; **G** : DRBP (forme "ouverte") ; **H** : DRBP (forme "fermée") ; **I** : pepsine ; **J** : dimère de E-cadhérine. Les couleurs des domaines sont identiques à celles de la figure 3.10.

de force qui mesure la facilité avec laquelle un résidu peut se déplacer par rapport à son environnement protéique. Cette constante de force est sensible à la structure globale de la protéine et, à l'inverse des facteurs de température, n'est pas dominée par l'empaquetage local autour du résidu sondé. L'information vectorielle est tout d'abord la direction de déplacement du résidu sondé, indiquant le chemin de moindre résistance, mais également les modifications des distances $C_{\alpha,i} - C_{\alpha,j}$ qui permettent de satisfaire la contrainte. Regroupées ensemble, ces dernières informations permettent de détecter la présence de zones rigides dans la protéine, ayant tendance à se déplacer les unes par rapport aux autres. Il est alors possible de déterminer des domaines dynamiques à l'aide d'un algorithme de regroupement. En comparaison des autres méthodes de définition de domaines dynamiques fondées sur la comparaison de structures (expérimentales ou théoriques) ou sur la sélection de modes normaux de basse fréquence, notre approche possède l'avantage de reposer sur un sondage uniforme de tous les résidus de la protéine.

L'application de la contrainte définie ici ainsi que l'analyse des résultats des simulations peuvent être réalisées avec une représentation "tout-atome" de la protéine aussi bien qu'avec un modèle "gros grains" dans lequel les interactions entre résidus sont modélisées par un réseau de ressort. La comparaison des résultats obtenus avec ces deux modèles sur l'exemple de la nucléase de staphylocoque suggère que la majorité des informations est conservée en utilisant une représentation simplifiée de la protéine. Il serait cependant nécessaire de réaliser la même comparaison sur d'autres systèmes afin de valider définitivement l'utilisation du modèle granulaire. De plus, l'analyse d'effets plus fins comme l'impact d'une mutation ponctuelle sur les propriétés mécaniques d'une protéine nécessite l'utilisation d'une représentation "tout-atome".

3.6 Perspectives

Lors de l'élaboration de la méthode présentée dans ce chapitre, l'une des motivations était de définir une nouvelle grandeur permettant d'obtenir de nouvelles informations sur la flexibilité de la protéine à l'échelle du résidu dans le but de pouvoir étudier les sites catalytiques des enzymes. Nous sommes en effet persuadés que les résidus impliqués dans les mécanismes enzymatiques présentent des propriétés mécaniques tout à fait particulières indispensables au bon fonctionnement de la protéine. Des travaux récents ont d'ailleurs mis en évidence que ces résidus catalytiques possèdent fréquemment des facteurs de température faibles (Bartlett *et al.*, 2002 ; Yuan *et al.*, 2003 ; Yang et Bahar, 2005), indiquant qu'ils sont en moyenne moins flexibles que les autres résidus de la protéine.

Une étude des propriétés mécaniques des sites fonctionnels des hémoprotéines a été menée récemment au laboratoire en utilisant cette mesure (Sacquin-Mora et Lavery, 2006). La méthode utilisée pour dériver les constantes de force dans ce travail est néanmoins différente de celle présentée ici : elle se base sur l'étude des fluctuations de D_i au cours de simulations de dynamique brownienne et sont données, pour le résidu i , par :

$$k_i = \frac{3k_B T}{\langle (D_i - \langle D_i \rangle)^2 \rangle}$$

où k_B représente la constante de Boltzmann et T la température.

L'accord entre les constantes de force calculées par dynamique brownienne et par minimisation d'énergie sous contrainte est très bon pour différentes architectures protéiques. La figure 3.12 montre l'exemple de la peroxydase du raifort, pour laquelle le coefficient de corrélation entre les deux jeux de valeurs est de 0,967.

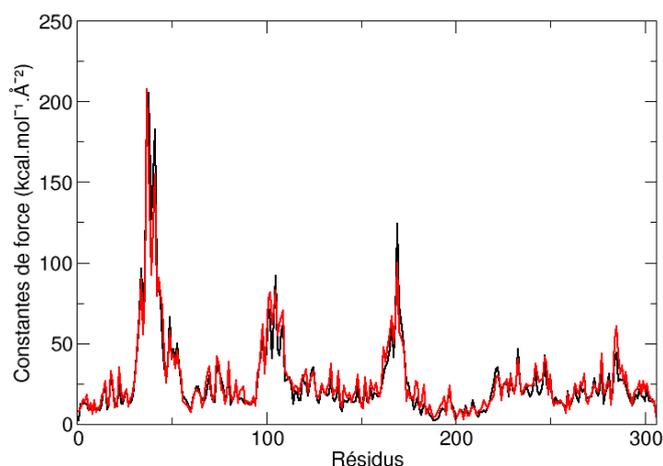


FIGURE 3.12 – Constantes de force calculées pour les 306 résidus de la peroxydase du raifort. La courbe rouge représente les valeurs obtenues par minimisation sous contrainte et la courbe noire celle obtenue par dynamique brownienne. Reproduit de (Sacquin-Mora et Lavery, 2006).

Les résultats obtenus dans l'étude de Sacquin et Lavery montrent qu'en règle générale les résidus fonctionnels sont notablement rigides. Par ailleurs, dans les cas où ceci ne se vérifie pas, la flexibilité observée est reliée à une nécessaire mobilité pour la fonction de la protéine.

Deuxième partie

Propriétés mécaniques et fonction : la E-cadhérine

Chapitre 4

Les molécules de cadhérine

Sommaire

4.1	L'adhésion cellulaire	63
4.2	Les cadhérines	65
4.2.1	Cadhérines classiques	67
4.2.2	Autres cadhérines	68
4.3	Propriétés des cadhérines classiques	69
4.3.1	Structure du domaine extracellulaire	69
4.3.2	Le rôle du calcium	69
4.3.3	Interfaces de dimérisation	73
4.3.4	Le rôle du tryptophane 2	77
4.3.5	Les différents modèles d'adhésion	79
4.3.6	Études théoriques des cadhérines	81

4.1 L'adhésion cellulaire

L'adhésion cellulaire est un processus central de la constitution des organismes évolués. Elle est nécessaire à la formation des tissus au stade embryonnaire et à leur maintien à l'âge adulte, mais joue également un rôle dans la réponse immunitaire ou la progression tumorale.

L'adhésion entre deux cellules ne procède pas toujours du même mécanisme et différents types de contact peuvent être distingués (voir figure 4.1 A) :

- *les jonctions serrées ("tight junctions")*

Ce sont les contacts entre cellules les plus étroits. Leur étanchéité permet de créer une barrière physiologique entre les compartiments extérieurs et intérieurs de l'organisme,

empêchant par exemple que le corps ne se vide de son eau par les tissus en contact avec le milieu extérieur, comme la peau ou le tube digestif. Ces jonctions consistent en une fusion partielle des membranes.

– *les jonctions adhérentes et les desmosomes*

Ces deux types de contact relient le cytosquelette de deux cellules par l'intermédiaire de molécules transmembranaires, les cadhérines (voir figure 4.2). Dans le cas des jonctions adhérentes, les cadhérines "classiques" (voir section 4.2) jouent, via les caténines, le rôle de médiateurs entre les filaments d'actine des deux cellules. Les desmosomes sont des structures en forme de disque dans lesquelles d'autres familles de cadhérines (les desmogléines et desmocollines) sont liées à des filaments intermédiaires.

– *les jonctions communicantes (gap junctions)*

Ces contacts permettent de relier le cytoplasme de deux cellules adjacentes par des molécules de connexine, qui s'arrangent spatialement pour créer des canaux (figure 4.1 B). Ce type de jonction permet en outre le passage de petites molécules entre les cellules.

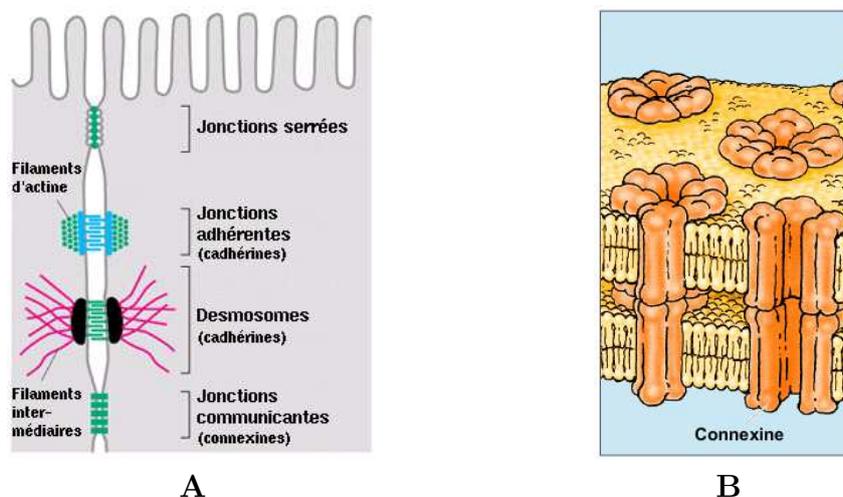


FIGURE 4.1 – Différents types d'adhésion cellulaire. **A** : Schéma des quatre types de jonctions cellulaires ; **B** : Jonctions communicantes. La figure A est reproduite de (Alberts *et al.*, 1994) et la figure B est reproduite de (Cooper et Hausman, 2006).

L'adhésion cellulaire est un phénomène spécifique, dans la mesure où une cellule ne se lie qu'à certains types d'autres cellules. Des études sur le développement embryonnaire ont par exemple montré que des cellules provenant d'un tissu particulier (par exemple des cellules du foie) ont tendance à adhérer à des cellules du même tissu plutôt qu'à des cellules d'un autre tissu (par exemple des cellules du cerveau). Cette spécificité peut s'expliquer

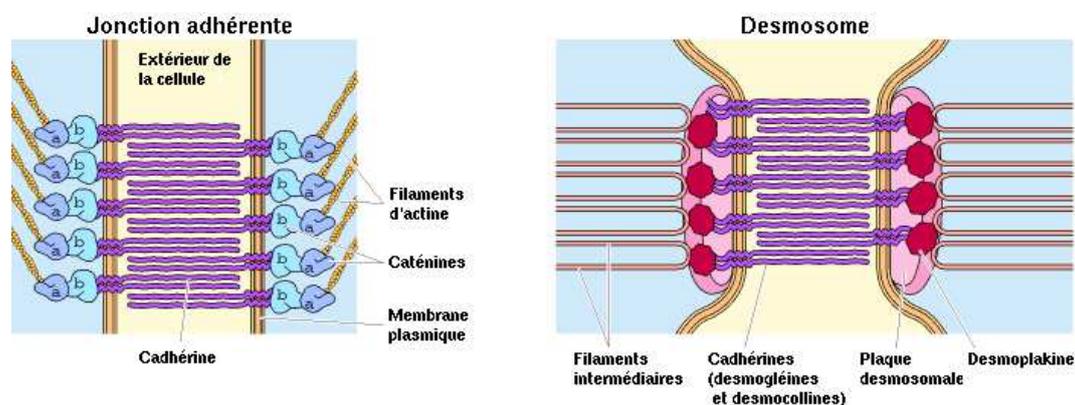


FIGURE 4.2 – Jonctions adhérentes et desmosomes. Adapté de (Cooper et Hausman, 2006).

par le type de molécules d'adhésion exprimées à la surface des cellules et qui interviennent dans la formation des contacts. Il existe quatre grandes familles de molécules d'adhésion impliquées dans les interactions inter-cellulaires, décrites succinctement dans le tableau 4.1.

Molécules d'adhésion cellulaire	Caractéristiques des interactions	Exemples de fonction
sélectines	- dépendantes de Mg^{2+} - hétérophiles - transitoires	- lie les cellules endothéliales aux lymphocytes - réponse inflammatoire
intégrines	- dépendantes de Mg^{2+} - hétérophiles - transitoires	- migration lymphocytaire
superfamille des immunoglobulines	- homophiles / hétérophiles - transitoires	- réponse immunitaire
cadhérines	- dépendantes du Ca^{2+} - homophiles - stables / transitoires	- embryogénèse - morphogénèse des tissus

TABLE 4.1 – Les différentes familles de molécules d'adhésion cellulaire.

4.2 Les cadhérines

Les cadhérines représentent une super-famille de glyco-protéines transmembranaires, qui sont impliquées à différents stades de l'évolution des organismes dans la formation et le maintien des contacts entre les cellules formant des tissus solides comme les épithéliums. Elles sont nécessaires au bon fonctionnement des organismes : une modification de leur

expression ou des mutations ponctuelles peuvent en effet souvent être reliées au développement de tumeurs (Aken *et al.*, 2001 ; Handschuh *et al.*, 2001 ; Christofori, 2003). Parmi les familles de protéines d'adhésion cellulaire, la famille des cadhérines est la seule dont l'activité est régulée par les ions calcium¹. Les cadhérines sont ainsi responsables de l'ensemble des phénomènes d'adhésion Ca^{2+} -dépendants chez presque tous les vertébrés (Takeichi, 1990).

Sous-famille	Exemples	Caractéristiques
Cadhérines classiques	- E-cadhérine ¹ - N-cadhérine ² - C-cadhérine ³ - R-cadhérine ⁴	- 5 domaines EC - domaine transmembranaire conservé - connectées à l'actine via les caténines
Cadhérines desmosomales	- desmogléines - desmocollines	- 5 domaines EC - un domaine transmembranaire - connectées aux filaments intermédiaires via la plaque desmosomale
Protocadhérines	- protocadhérine-1 - protocadhérine-11	- 6 ou 7 domaines EC - un domaine transmembranaire
Cadhérines atypiques	- DN-cadhérine ⁵ - cadhérines Flamingo	- nombre de domaines EC variable - présence de domaines extracellulaires "non-EC" - un ou plusieurs domaines transmembranaires

TABLE 4.2 – Les différentes familles de cadhérines. Inspiré de (Koch *et al.*, 2004).

La super-famille des cadhérines contient un nombre important de protéines (environ 100 gènes codant pour des molécules de cadhérine ont été identifiés dans le génome humain). Ces protéines peuvent être regroupées en quatre sous-familles, en fonction de leurs différences structurales (voir tableau 4.2). Le domaine cytosolique des cadhérines leur permet de se rattacher au cytosquelette et semble également être impliqué dans les phénomènes d'adhésion (Nagafuchi et Takeichi, 1988 ; Ozawa *et al.*, 1990). Le domaine extracellulaire (ou ectodomaine) présente une structure de type modulaire avec une répétition de domaines similaires, notés domaines EC pour "*extra-cellular*". Le nombre de ces domaines est variable entre les différentes familles de cadhérines et parfois au sein d'une même famille (voir figure 4.3).

1. Leur nom provient de la contraction de "*calcium-dependant adherent protein*".

1. Cadhérine épithéliale.

2. Cadhérine neuronale.

3. Cadhérine de compaction.

4. Cadhérine rétinienne.

5. Cadhérine neuronale de la drosophile.

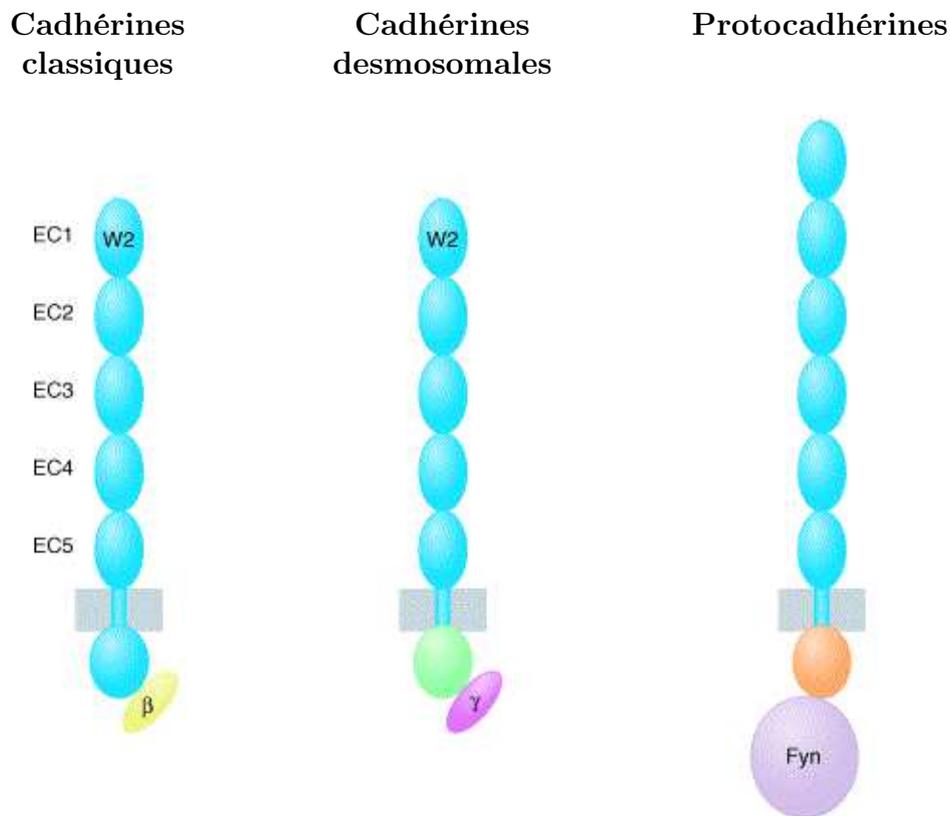


FIGURE 4.3 – Structure schématique de représentants de trois familles de cadhérines. Pour les cadhérines classiques et desmosomales, W2 indique la présence du résidu Trp 2, conservé au sein du domaine EC1. Les ellipsoïdes notés β et γ représentent respectivement des molécules de β - et γ -caténine, qui relie le domaine cytoplasmique des cadhérines au cytosquelette. La sphère notée Fyn représente la Fyn tyrosine kinase, qui est un exemple de protéine liant le domaine cytoplasmique des protocadhérines. Adapté de (Patel *et al.*, 2003).

4.2.1 Cadhérines classiques

Les cadhérines classiques forment la famille la plus étudiée parmi les familles de cadhérines. Les premières cadhérines découvertes font en effet partie de cette famille et ont été dénommées en fonction des tissus dans lesquels elles ont été trouvées :

- la E-cadhérine, présente dans de nombreux types de cellules épithéliales.
- la N-cadhérine, présente dans les cellules nerveuses ou du muscle.
- la P-cadhérine, présente dans les cellules du placenta et de l'épiderme.
- ...

Le domaine extracellulaire des cadhérines classiques comporte cinq domaines EC, dont la séquence est relativement bien conservée à l'intérieur de la famille, et un domaine cytoplasmique très conservé. Un résidu notamment, le tryptophane 2, qui joue un rôle important dans les phénomènes d'adhésion (voir section 4.3), est retrouvé chez toutes les cadhérines classiques.

En dépit de leur fort taux d'homologie de séquence (voir tableau 4.3), les cadhérines classiques interagissent principalement de manière homophile au niveau des jonctions adhérentes : lorsque des cellules exprimant l'une ou l'autre espèce de cadhérine classique sont mises en co-culture, il est observé que celles-ci adhèrent préférentiellement aux cellules de même type (Nose *et al.*, 1990 ; Gumbiner, 1996 ; Tamura *et al.*, 1998). De récentes études ont cependant montré que, si les interactions homophiles restent favorisées, certains types de cadhérines classiques (comme la N- et la E-cadhérine) peuvent adhérer entre elles, tandis que d'autres forment des agrégats séparés (Niessen et Gumbiner, 2002 ; Duguay *et al.*, 2003). En réalité, une cellule exprime fréquemment plusieurs types de cadhérines et la spécificité des interactions pourrait provenir de différences dans les concentrations relatives de chaque type de cadhérine à la surface de la cellule (Duguay *et al.*, 2003 ; Chen *et al.*, 2005).

	E-cadhérine	N-cadhérine	C-cadhérine
E-cadhérine	100	59	71
N-cadhérine		100	56
C-cadhérine			100

TABLE 4.3 – Taux d'homologie des séquences des domaines EC1 de divers représentants de la famille des cadhérines classiques. Adapté de (Patel *et al.*, 2003).

4.2.2 Autres cadhérines

4.2.2.1 Cadhérines desmosomales

Il existe deux types de cadhérines desmosomales (les desmogléines et les desmocollines), qui sont responsables des jonctions au sein des desmosomes. Comme les cadhérines classiques, elles possèdent un domaine extracellulaire composé de cinq modules EC et le résidu Trp 2 au sein du domaine EC1. La principale différence entre les cadhérines desmosomales et les cadhérines classiques réside en fait dans leur domaine cytosolique.

4.2.2.2 Protocadhérines

Les protocadhérines forment la plus grande sous-famille de cadhérines. Au niveau structural, leur domaine extracellulaire contient généralement six à sept modules EC. Les protocadhérines sont impliquées dans la morphogénèse des tissus dans les étapes embryonniques puis plus tard à l'âge adulte dans la différenciation cellulaire (Frank et Kemler, 2002).

4.2.2.3 Cadhérines atypiques

Cette dernière famille de cadhérines contient les molécules adhésives dépendantes du calcium, et qui ne peuvent être classées dans une des trois familles précédemment citées. Les cadhérines atypiques possèdent une grande variété de structures, avec notamment un nombre variable de domaines EC.

4.3 Propriétés des cadhérines classiques

4.3.1 Structure du domaine extracellulaire

Le segment extracellulaire des cadhérines classiques possède une structure modulaire, constituée de cinq domaines EC, notés EC1 pour le domaine le plus distant de la membrane jusqu'à EC5 (voir figure 4.3). Chacun de ces domaines, longs d'environ 45 Å contient environ 110 acides aminés et son repliement correspond à la formation d'un sandwich de sept feuillets β , à la façon des domaines immunoglobulines. La connexion entre les modules EC est assurée par des jonctions d'une dizaine de résidus, et associée à la présence de trois ions Ca^{2+} à l'interface entre les paires de domaines. Contrairement aux premières hypothèses, qui postulaient une forme très rectiligne de la molécule, l'ensemble du segment extracellulaire n'est pas linéaire mais courbé à tel point que, dans la structure cristallographique de l'ectodomaine de la C-cadhérine (Boggon *et al.*, 2002), la direction du domaine EC1 est presque perpendiculaire à celle du domaine EC5, en dépit du fait que l'angle entre chaque paire de domaines est relativement faible (voir figure 4.4).

4.3.2 Le rôle du calcium

Diverses observations ont rapidement mis en avant l'importance du calcium lors des premières études sur les cadhérines :

- la présence d'ions Ca^{2+} rend les cadhérines moins sensibles à l'action des protéases (Takeichi, 1990).

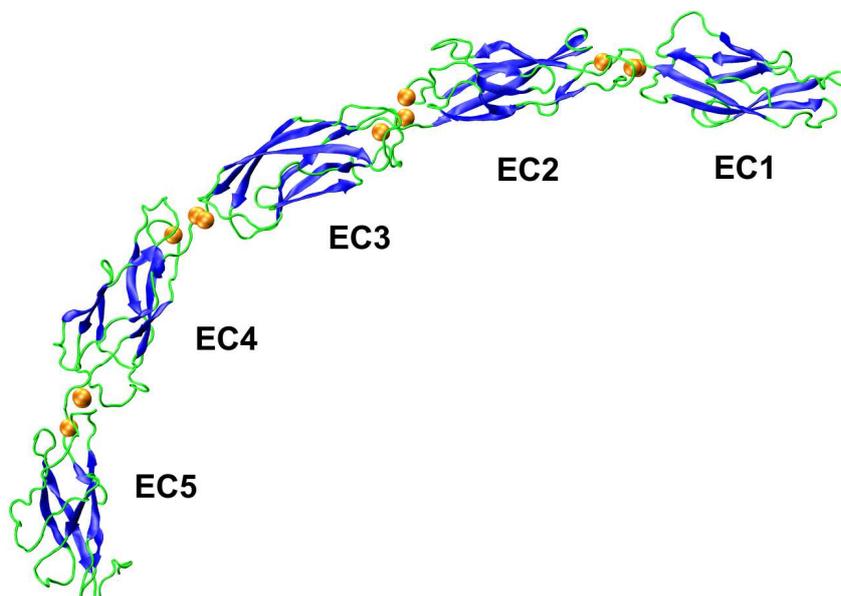


FIGURE 4.4 – Structure du domaine extracellulaire de la C-cadhérine (1L3W). Les ions Ca^{2+} aux interfaces entre les domaines EC sont représentés par des sphères orange.

- la mutation de certains résidus impliqués dans la fixation du calcium, comme par exemple le remplacement de l'aspartate en position 134 par une alanine, abolit les propriétés adhésives des cadhérines (Ozawa *et al.*, 1990).
- en l'absence de calcium, la structure des molécules de cadhérine est complètement modifiée (Pokutta *et al.*, 1994 ; Alattia *et al.*, 1997).

Les structures à haute résolution obtenues depuis ont permis d'interpréter ces observations. La présence de trois sites de fixation pour les ions Ca^{2+} à la jonction entre les domaines EC permet en effet de contrôler la conformation du segment extracellulaire des molécules de cadhérine. Plusieurs études de microscopie électronique ont été réalisées sur l'ensemble de l'ectodomaine (EC1-5), ainsi que sur des fragments constitués des seuls deux domaines N-terminaux (EC1-2) de la E-cadhérine afin de mieux comprendre l'impact structural de ces ions (Pokutta *et al.*, 1994 ; Alattia *et al.*, 1997 ; Pertz *et al.*, 1999). En présence de concentration physiologique de calcium, les micrographes électroniques montrent des molécules allongées présentant une certaine modularité. En revanche, ces structures en "bâtonnet" sont détruites en l'absence de calcium, et l'orientation relative des domaines devient aléatoire (voir figure 4.5).

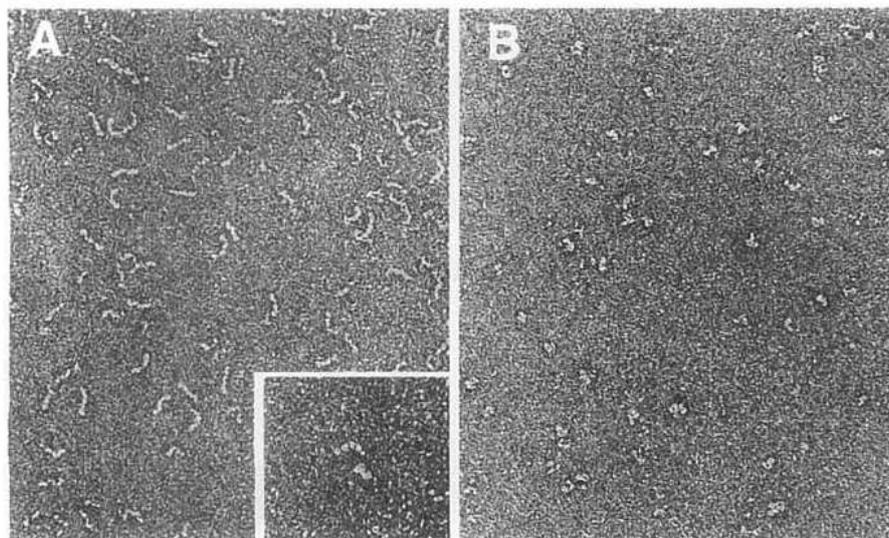


FIGURE 4.5 – Micrographies électroniques montrant l’influence de la présence (A) ou de l’absence (B) de calcium sur la structure du domaine extracellulaire de la E-cadhérine. Les molécules de cadhérine sont représentées en blanc. Dans la figure A, elles ont une géométrie plutôt allongée, tandis que dans la figure B, on observe qu’elles sont complètement repliées. L’encart de la figure A montre un agrandissement d’une molécule particulière de cadhérine permettant de repérer la modularité de la structure. Reproduit de (Pokutta *et al.*, 1994).

Les modifications de déplacement chimique observées en RMN lors de l’ajout d’ions Ca^{2+} à des solutions contenant des fragments EC1-2 confirment cette perte de structure (Häussinger *et al.*, 2002). Ces changements sont localisés à l’interface entre les domaines EC1 et EC2, révélant que ce changement conformationnel conserve la structure de chacun des domaines EC. De plus, l’analyse des temps de relaxation T_1 et T_2 lors d’expériences de RMN ^{15}N confirme la rigidification de la molécule en présence de calcium. Enfin, l’impact des ions Ca^{2+} sur la dimérisation a également été montré dans cette étude, l’association de monomères de EC1-2 n’étant observée qu’en présence de calcium. Il a récemment été proposé que le calcium joue également un rôle dans un changement conformationnel de moindre ampleur mais qui est important pour l’adhésion (Harrison *et al.*, 2005). Harrison et collaborateurs ont en effet montré que le calcium régule l’équilibre d’insertion du Trp 2 dans sa poche hydrophobe (voir les paragraphes suivants).

En dépit de ces nombreuses données, certains points restent non élucidés, comme l’importance relative des différents sites de fixation du calcium à l’interface des domaines EC. Pokutta et collaborateurs ont réalisé des expériences de dichroïsme circulaire et de fluorescence de tryptophane pour observer la fixation du calcium aux cadhérines (Pokutta *et al.*, 1994). Les valeurs de constantes de dissociation K_d obtenues avec les deux techniques sont

significativement différentes ($45 \mu M$ par dichroïsme circulaire et $170 \mu M$ par fluorescence de tryptophane). Cependant, les données de spectroscopie de fluorescence indiquent une forte coopérativité entre les différents sites de fixation¹. Dans une autre étude, Alattia et collaborateurs ont également noté une coopérativité de la fixation des ions calcium après la formation des dimères de cadhérine (Alattia *et al.*, 1997). Finalement, des mesures d'affinité du calcium pour EC1-2 obtenues par des expériences de dialyse à l'équilibre (Koch *et al.*, 1997) soutiennent un modèle dans lequel deux sites de fixation sont équivalents (avec $K_d = 330 \mu M$) et le troisième présente une faible affinité pour le calcium ($K_d = 2 mM$). Les auteurs proposent que le site CA3 (voir figure 4.6) soit le site de plus faible affinité. En effet, ce site, qui est le plus exposé au solvant n'offre pas une sphère de coordination complète à l'ion Ca^{2+} , qui est alors, dans la structure cristallographique, également lié à des molécules d'eau. De plus, tous les ligands du site CA3 sont présents dans le domaine EC1, qui pourtant ne fixe pas le calcium en l'absence de EC2.

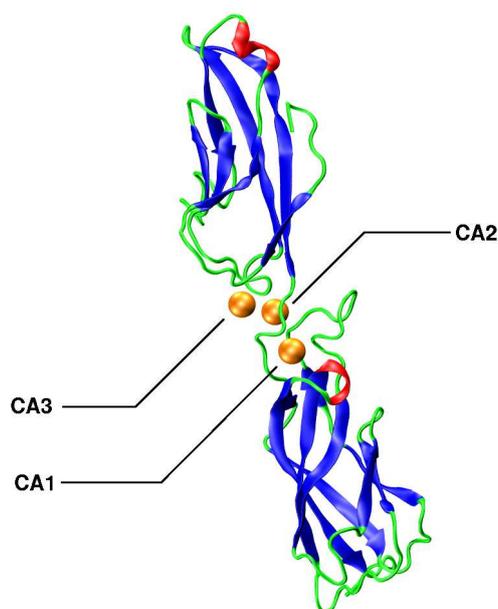


FIGURE 4.6 – Position des ions Ca^{2+} à l'interface entre les domaines EC1 et EC2 dans un monomère de EC1-2, extrait de la structure 1FF5 (Pertz *et al.*, 1999).

1. Lors de l'interprétation des résultats de ces expériences, le nombre d'ions fixés aux cadhérines était encore inconnu, et ces résultats doivent donc être traités avec précaution.

4.3.3 Interfaces de dimérisation

4.3.3.1 Rôle du domaine EC1

Sur les cinq modules EC qui composent le segment extracellulaire des cadhérines classiques, le domaine EC1 (le plus distant de la membrane cellulaire) joue un rôle particulièrement important dans les phénomènes adhésifs. Diverses preuves expérimentales supportent cette hypothèse :

1. une protéine chimérique de E-cadhérine dans laquelle le domaine EC1 N-terminal a été échangé par le domaine EC1 d'une N-cadhérine acquiert la même spécificité d'adhésion que la E-cadhérine, c'est-à-dire qu'elle va se lier avec les E-cadhérines mais également ne plus interagir avec les N-cadhérines (Nose *et al.*, 1990 ; Shan *et al.*, 2000). Ce résultat indique à la fois que le domaine EC1 est important pour l'adhésion mais également que la spécificité des interactions repose au moins en grande partie sur lui. Cette conclusion est renforcée par de récents travaux qui permettent de mieux comprendre l'orthogonalité des interactions entre les cadhérines classiques de type I et II, l'interface de dimérisation dans le cas des cadhérines classiques de type II impliquant non seulement le Trp 2 mais également le Trp 4, conservé dans cette sous-famille mais absent chez les cadhérines classiques de type I (Patel *et al.*, 2006).
2. lors de leur synthèse, les cadhérines possèdent un pré-domaine situé à l'extrémité N-terminale, qui doit être retiré pour activer les propriétés adhésives (Ozawa et Kemler, 1990). La proximité de ce domaine avec le domaine EC1 renforce l'idée que cette région est impliquée dans les interactions nécessaires à l'adhésion.
3. toutes les structures cristallographiques obtenues jusqu'à présent (détaillées par la suite), que ce soit pour le fragment EC1-2 uniquement (Nagar *et al.*, 1996 ; Pertz *et al.*, 1999 ; Häussinger *et al.*, 2004) ou pour le domaine extracellulaire entier (Boggon *et al.*, 2002), et qui présentent des interactions entre cadhérines indiquent une interface au niveau du domaine EC1.
4. des résultats récents obtenus par tomographie électronique¹ par He et collaborateurs sur l'organisation des desmosomes montrent que le site d'interaction prédominant entre deux cadhérines est situé au niveau du domaine EC1 (He *et al.*, 2003).
5. la mutation ponctuelle de certains résidus du domaine EC1 (en particulier le résidu Trp 2) abolit l'adhésion cellulaire engendrée par les cadhérines (Tamura *et al.*, 1998 ; Shan *et al.*, 2000).

1. La tomographie électronique est une technique basée sur la microscopie électronique qui permet d'obtenir des images tridimensionnelles de molécules avec des résolutions allant de 10 à 30 Å.

Le rôle du domaine EC1 pour les interactions entre les cadhérines est désormais bien établi. Il n'en est cependant pas de même de celui de chacun des autres modules EC. La question reste en effet ouverte de savoir combien de domaines EC sont réellement nécessaires pour que les cadhérines puissent engendrer l'adhésion cellulaire. Sur la base de mesures de forces de surface¹, Leckband et Sivasankar ont proposé que les domaines EC1 à EC3 sont nécessaires à l'adhésion (Leckband et Sivasankar, 2000), tandis que Shan et collaborateurs, à partir d'expériences d'agrégation cellulaire, ont observé que les deux premiers domaines EC1-2 permettent l'adhésion, et que le seul domaine EC1 n'était pas suffisant (Shan *et al.*, 2004). Diverses études ont également noté la possibilité de multiples modes d'interaction impliquant un, plusieurs, voire les cinq domaines EC (Chappuis-Flament *et al.*, 2001 ; Zhu *et al.*, 2003 ; Perret *et al.*, 2004).

4.3.3.2 Dimères *cis* et *trans*

Les expériences de microscopie électronique déjà citées précédemment (Tomschy *et al.*, 1996 ; Pertz *et al.*, 1999) ont permis d'obtenir des informations sur le mode de dimérisation des cadhérines. Ces études ont en effet montré que deux types de contacts étaient impliqués dans les phénomènes d'adhésion. Dans un premier temps, il y a formation d'un dimère entre molécules provenant de la même cellule (contact *cis*), puis ces dimères *cis* de chacune des cellules en contact interagissent à leur tour pour former une interaction *trans* adhésive (Brieher *et al.*, 1996 ; Yap *et al.*, 1997 ; Chitaev et Troyanovsky, 1998).

4.3.3.3 Structures à résolution atomique

Dans ce paragraphe, nous décrivons l'ensemble des structures de cadhérines classiques de type I à haute résolution déterminées à ce jour, qui sont listées dans le tableau 4.4.

Les premières structures à haute résolution de fragments de domaine extracellulaire de cadhérines ont été obtenues en 1995. Elles concernent le domaine EC1 de la E-cadhérine (Overduin *et al.*, 1995) et de la N-cadhérine (Shapiro *et al.*, 1995). Ces deux structures montrent des domaines possédant un repliement identique. La structure de la N-cadhérine montre également deux interfaces, décrites comme représentant les interactions *cis* entre molécules de la même cellule et inter-cellulaires *trans*. On observe dans l'interface *cis* que le Trp 2 d'un monomère est inséré dans une poche hydrophobe de l'autre monomère (voir figure 4.7). L'importance de ce motif, nommé échange de brins (ou "*strand-exchange*"), sera discutée dans le paragraphe suivant. Le contact *trans* implique lui la face opposée du

1. Les mesures de forces de surface quantifient les interactions entre deux matériaux en fonction de la distance qui les séparent.

Domaines	Type de cadhérine	Technique	Interfaces modélisées	code PDB Référence	Modèle
EC1	E-cadhérine (souris)	RMN		1SUH (Overduin <i>et al.</i> , 1995)	
EC1	N-cadhérine (souris)	RX	- <i>cis</i> <i>BE</i> - <i>trans</i>	1NCG (Shapiro <i>et al.</i> , 1995)	B et F
EC1-2	E-cadhérine (souris)	RX	- <i>cis</i> <i>BN</i>	1EDH (Nagar <i>et al.</i> , 1996)	C
EC1-2	N-cadhérine (souris)	RX	- <i>cis</i> <i>BN</i>	1NCJ (Tamura <i>et al.</i> , 1998)	C
EC1-2	E-cadhérine (souris)	RX	- <i>cis</i> <i>BN</i>	1FF5 (Pertz <i>et al.</i> , 1999)	C
EC1-2	E-cadhérine (souris)	RX	- <i>trans</i> <i>BE</i>	1Q1P (Häussinger <i>et al.</i> , 2004)	D et F
EC1-5	C-cadhérine (crapaud)	RX	- <i>trans</i> <i>BE</i>	1L3W (Boggon <i>et al.</i> , 2002)	D et F

TABLE 4.4 – Ensemble des structures à haute résolution disponibles pour les cadhérines classiques. Dans la colonne "Interfaces modélisées", *BE* signifie que ce dimère présente le motif d'échange de brin. La colonne "modèle" renvoie aux modèles décrits dans la figure 4.10.

domaine EC1 et contient notamment un motif His-Ala-Val très conservé chez les cadhérines classiques.

Deux structures cristallographiques de fragments EC1-2 de E-cadhérine (codes PDB 1EDH et 1FF5) ont ensuite été résolues (Nagar *et al.*, 1996 ; Pertz *et al.*, 1999), dans lesquelles les interactions entre monomères sont différentes de celles observées dans le cristal 1NCG du module EC1 de la N-cadhérine. Dans ces deux structures, les monomères forment un dimère en forme de croix avec des interactions parallèles entre les domaines EC1, et une interface qui s'étend jusqu'à la jonction inter-domaines fixant le calcium (voir figure 4.8 A). Contrairement à la structure de Shapiro et collaborateurs, l'interface observée ici n'apporte pas d'explication quant au rôle du résidu Trp 2, ce résidu étant soit désordonné (dans la structure 1EDH), soit enfoui dans la poche hydrophobe de son propre monomère. Ce dimère sera ainsi noté dimère *BN* (pour "brins non échangés") par opposition aux dimères présentant le phénomène d'échange de brins (notés dimères *BE*)¹. En revanche, l'implication de la jonction inter-domaines dans l'interface permet d'expliquer l'implication des ions Ca^{2+} dans la dimérisation. Il est à noter que dans les structures 1EDH et 1FF5, le fragment EC1-2 possède quelques résidus supplémentaires à l'extrémité N-terminale (Met-Arg pour

1. Ces deux dimères sont respectivement nommés *non-swapped* et *swapped dimers* en anglais.

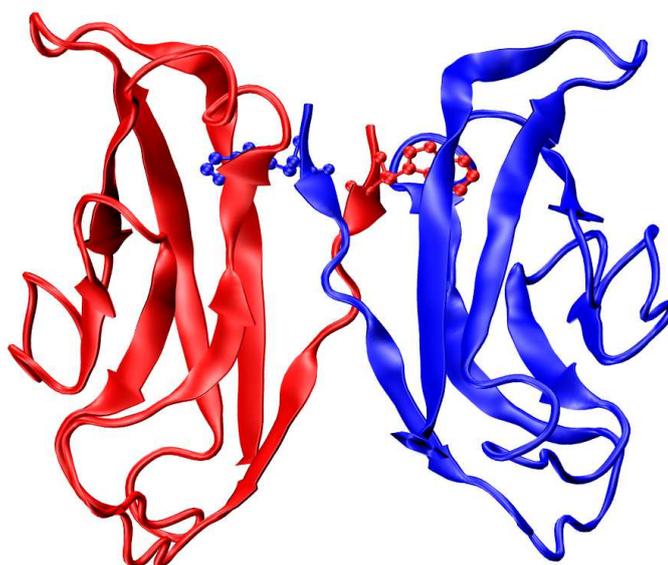


FIGURE 4.7 – Interaction *cis* observée dans le cristal 1NCG de la N-cadhérine (Shapiro *et al.*, 1995). La chaîne latérale des résidus Trp 2 est représentée afin de visualiser le motif d'échange de brins.

1EDH et Met pour 1FF5). Or la présence de résidus N-terminaux supplémentaires peut abolir l'adhésion par les cadhérines (Ozawa et Kemler, 1990). Cependant, une interface impliquant la jonction inter-domaines est également observée dans une structure du fragment EC1-2 de la N-cadhérine (Tamura *et al.*, 1998), dans laquelle l'extrémité N-terminale est celle de la molécule biologiquement active.

Une structure plus récente du fragment EC1-2 de la E-cadhérine (voir figure 4.8 B) présente une interaction des domaines EC1, donnant une représentation possible de l'interaction *trans* (Häussinger *et al.*, 2004). Dans cette structure (notée dimère *BE* par la suite) est retrouvée l'insertion du Trp 2 d'un monomère dans la poche hydrophobe de son partenaire. Il convient de noter toutefois que dans ce cas, ce motif est associé à une interaction *trans*, tandis que dans la structure du domaine EC1 de N-cadhérine (Shapiro *et al.*, 1995), il l'était à une interaction *cis*. Le domaine extracellulaire de la C-cadhérine a été cristallisé dans son ensemble (Boggon *et al.*, 2002), et les contacts inter-moléculaires observés dans cette structure (code PDB 1L3W) sont similaires aux contacts *trans* de la structure de Häussinger et collaborateurs (Häussinger *et al.*, 2004).

En plus de ces interactions entre domaines EC1, une autre interface a été observée entre les fragments EC1-2 dans plusieurs des structures citées précédemment. Celle-ci correspond à un contact parallèle entre le domaine EC1 d'un monomère et le domaine EC2 de son partenaire (voir figure 4.8 C), et le dimère résultant sera noté par la suite dimère *décalé*. Bien que cette interaction ait souvent été considérée comme un contact cristallin en raison de la faible aire de l'interface, elle a également été proposée dans un contexte biologique comme pouvant représenter l'interaction *cis* entre cadhérines provenant de la même cellule (Boggon *et al.*, 2002).

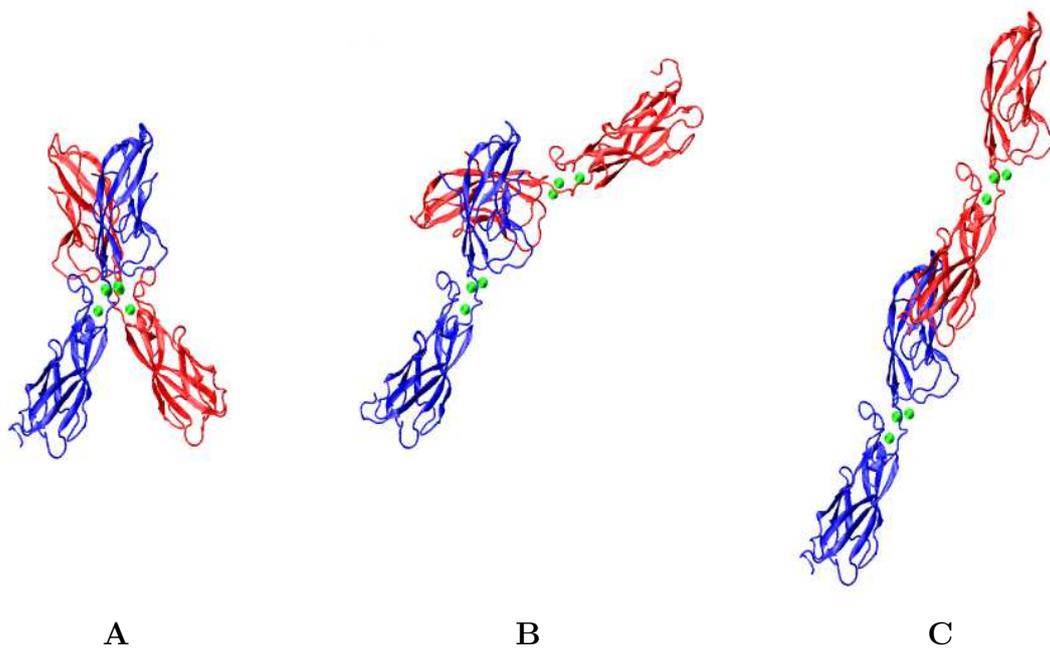


FIGURE 4.8 – Différentes interfaces observées par cristallographie entre des monomères de fragments EC1-2 de E-cadhérine. Les ions Ca^{2+} sont représentés par des sphères vertes. **A** : interaction *cis* dans le dimère *BN* (1FF5); **B** : interaction *trans* avec échange de brins dans le dimère *BE* (1Q1P); **C** : interaction *cis décalée* entre le domaine EC1 d'un monomère et le domaine EC2 de son partenaire (1FF5 et 1Q1P).

4.3.4 Le rôle du tryptophane 2

De nombreuses études ont montré l'importance d'un acide aminé particulier, le tryptophane 2, dans les phénomènes d'adhésion par les cadhérines classiques. Des cadhérines dans lesquelles ce résidu est substitué par une alanine perdent en effet leurs propriétés adhésives. Dans une moindre mesure, le remplacement du tryptophane par une phénylalanine ou une tyrosine, qui possèdent une chaîne latérale aromatique comme le tryptophane, réduit l'adhésion sans toutefois l'abolir (Tamura *et al.*, 1998 ; Shan *et al.*, 2000).

Ce rôle prépondérant du Trp 2 est parfaitement compréhensible au vu de certaines des structures cristallographiques obtenues (Shapiro *et al.*, 1995 ; Boggon *et al.*, 2002 ; Häussinger *et al.*, 2004). Dans ces trois structures, qui représentent soit une interaction parallèle (Shapiro *et al.*, 1995), soit une interaction *anti*-parallèle (Boggon *et al.*, 2002 ; Häussinger *et al.*, 2004) des domaines EC1, le Trp 2 d'un monomère est inséré dans une poche hydrophobe de l'autre monomère (voir figures 4.7 et 4.9 B). Ce motif d'échange de brins correspond au détachement d'un brin du tonneau β vers le monomère opposé. L'importance de ce mécanisme d'échange de brins est révélée par des expériences de mutagenèse dirigée : non seulement, la substitution du Trp 2 abolit l'adhésion, mais la substitution de résidus appartenant à la poche hydrophobe qui accueille ce tryptophane (mutations de la sérine 78 en méthionine ou de l'alanine 80 en méthionine ou en leucine) entraîne une perte des propriétés adhésives des cadhérines (Tamura *et al.*, 1998 ; Laur *et al.*, 2002).

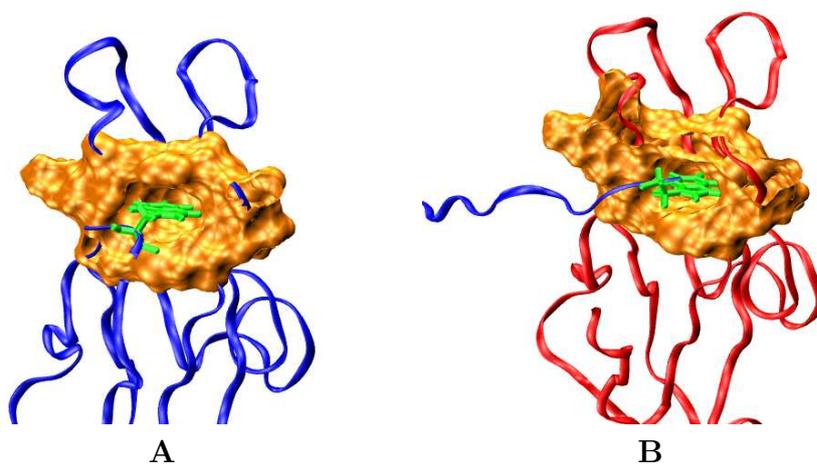


FIGURE 4.9 – Positionnement du Trp 2 dans les dimères de cadhérine. **A** : dans la poche hydrophobe de son propre monomère dans la structure 1FF5; **B** : dans la poche hydrophobe de son partenaire dans la structure 1Q1P.

De façon intéressante, dans les autres structures de cadhérines ainsi que dans la forme monomérique observée par RMN (Häussinger *et al.*, 2002), le Trp 2 est soit désordonné (Nagar *et al.*, 1996 ; Tamura *et al.*, 1998) soit inséré dans la même poche hydrophobe mais cette fois-ci de son propre domaine (Pertz *et al.*, 1999), comme sur la figure 4.9 A. Ceci révèle l'importance de la flexibilité du brin N-terminal chez les cadhérines et indique que la sortie du tryptophane de sa poche hydrophobe est une étape clé du mécanisme de formation des dimères *BE* de cadhérine. Ceci a par ailleurs été démontré dans une étude récente menée par Harrison et collaborateurs (Harrison *et al.*, 2005).

4.3.5 Les différents modèles d'adhésion

L'ensemble des structures de cadhérines obtenues à résolution atomique sont résumées dans le tableau 4.4. Au fur et à mesure que ces structures ont été publiées, différents modèles plus ou moins contradictoires ont été développés pour expliquer les phénomènes d'adhésion par les cadhérines, qui sont représentés sur la figure 4.10.

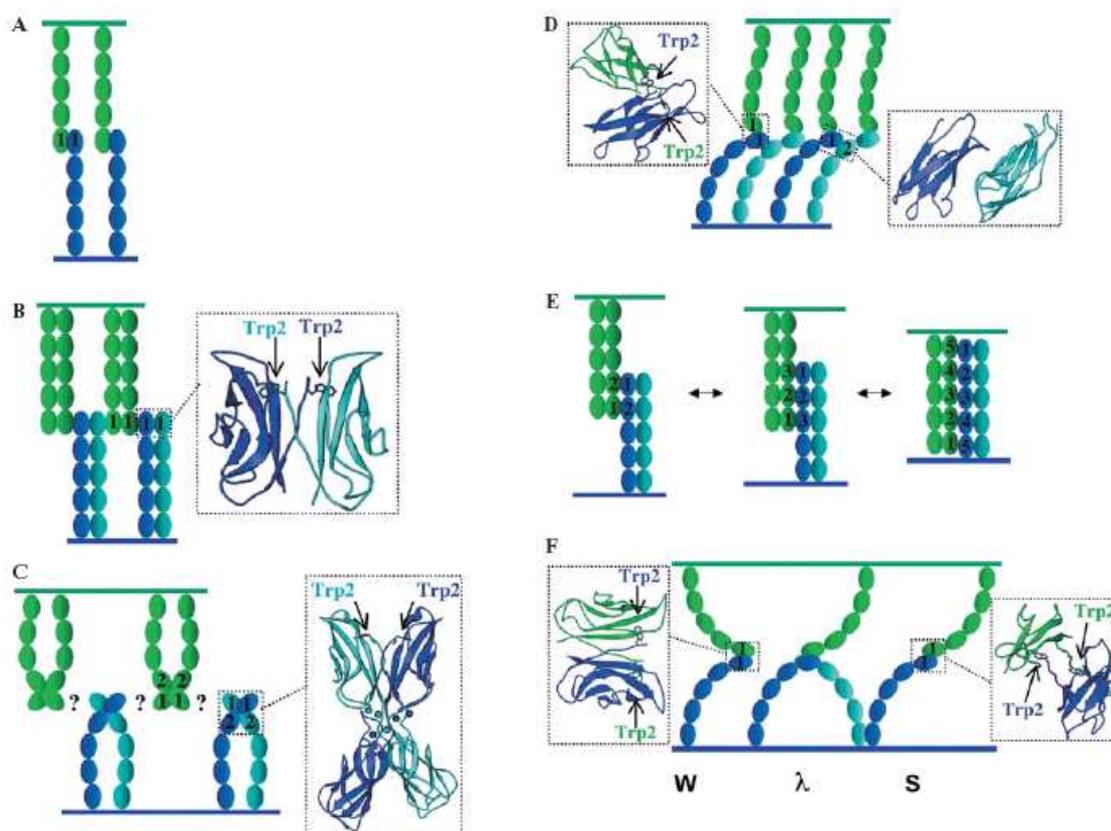


FIGURE 4.10 – Modèles d'adhésion cellulaire par les cadhérines s'appuyant sur les différentes structures obtenues expérimentalement. Reproduit de (Koch *et al.*, 2004).

Le modèle A est le premier modèle d'interactions entre cadhérines, bâti alors qu'aucune structure à haute résolution n'était disponible. Dans ce modèle, les domaines extracellulaires des cadhérines sont vus comme des bâtonnets qui réalisent des interactions homophiles par leur domaine N-terminal EC1.

La structure obtenue par Shapiro et collaborateurs (Shapiro *et al.*, 1995) a engendré une vision complètement différente. Les deux interfaces observées dans cette structure permettent de proposer un modèle de type fermeture éclair ("*zipper*") : les cadhérines provenant de cellules opposées interagissent de manière *anti*-parallèle sous la forme de dimères *cis* stabilisés par le mécanisme d'échange de brins (voir figure 4.10 B).

Les premières structures de dimères des fragments EC1-2 ont remis en question ce modèle, car ces dimères, représentant des interactions parallèles entre les domaines EC1, ne présentent pas le mécanisme d'échange de brins. Le modèle C montre la possible implication de ce dimère dans les phénomènes adhésifs. Ce modèle ne propose cependant aucune interface pour représenter les contacts adhésifs entre cadhérines provenant de cellules opposées.

Le modèle D a été conçu sur la base des dimères observés dans la structure de l'ectodomaine entier de la C-cadhérine (Boggon *et al.*, 2002). Dans cette structure, le mécanisme d'échange de brins se déroule entre deux domaines EC1 de manière plutôt *anti*-parallèle, représentant ainsi un contact *trans* entre cadhérines provenant de cellules opposées. Le contact rendant compte des interactions des molécules de la même cellule est alors un contact *décalé* entre le module EC1 d'une cadhérine et le module EC2 d'une autre molécule.

Les expériences récentes de mesure de forces (Leckband et Sivasankar, 2000 ; Zhu *et al.*, 2003 ; Perret *et al.*, 2004 ; Bayas *et al.*, 2006), montrant l'implication de domaines autres que le domaine EC1 dans l'adhésion cellulaire par les cadhérines ont conduit à l'élaboration du modèle E. Dans ce modèle, une seule interaction n'est pas suffisante pour expliquer les propriétés adhésives des cadhérines, et l'adhésion est en fait due à la somme de plusieurs types d'interactions impliquant de un à cinq des domaines EC.

Enfin, une récente étude basée sur l'utilisation de la tomographie électronique a permis de proposer le modèle F (He *et al.*, 2003). Dans cette étude, He et collaborateurs ont cartographié une jonction desmosomale et reconstruit les interactions en superposant la structure du domaine extracellulaire entier de la C-cadhérine sur les cartes de densité électronique¹. Les auteurs ont observé trois motifs d'interaction différents, baptisés *W*, *S* et λ en raison de leur topologie (voir figure 4.10 F). Ces trois géométries sont compatibles

1. L'utilisation de la structure de la C-cadhérine comme modèle est raisonnable du fait de la forte similarité entre les domaines extracellulaires des cadhérines classiques et desmosomales, qui possèdent 5 domaines et le même résidu tryptophane en position 2 du brin N-terminal.

avec les structures 1NCG et 1Q1P présentant le mécanisme d'échange de brins pour les dimères *cis* et *trans*, mais qui représentent cette fois-ci toujours des interfaces adhésives (contrairement à ce qui avait été proposé par Shapiro et collaborateurs pour la structure 1NCG (Shapiro *et al.*, 1995)). De plus, les interactions observées semblent contredire les expériences de mesure de forces, dans la mesure où toutes les interactions observées impliquent uniquement le domaine EC1.

Depuis une quinzaine d'années, plusieurs modèles ont donc été développés pour tenir compte des différentes structures résolues et des données biochimiques. Ces modèles sont plus ou moins contradictoires et aucune preuve définitive n'a pu être apportée en faveur de l'un ou l'autre. Il se peut cependant que certaines des structures observées expérimentalement n'aient en fait pas d'implication biologique, ou qu'au contraire ces multiples structures soient le signe d'une grande plasticité des molécules de cadhérines qui peuvent interagir de diverses façons.

4.3.6 Études théoriques des cadhérines

Peu d'études théoriques ont été menées sur les cadhérines. À notre connaissance, seules trois telles études ont été publiées en dehors de nos travaux (Makagiansar *et al.*, 2002 ; Bayas *et al.*, 2004 ; Sotomayor *et al.*, 2005). L'une d'elle (Makagiansar *et al.*, 2002) consiste à étudier par dynamique moléculaire la possibilité d'introduire un pont dissulfure pour stabiliser le dimère *BN* de E-cadhérine, en relation avec un travail expérimental. Les deux autres études sont en revanche uniquement tournées vers la simulation. Elles utilisent la technique de dynamique moléculaire d'étirement, décrite dans la section 2.3, pour étudier la dissociation d'une interaction *BE* de C-cadhérine (Bayas *et al.*, 2004), ou la stabilité du module EC2 de la C-cadhérine (Sotomayor *et al.*, 2005).

Les résultats obtenus par Bayas et collaborateurs (Bayas *et al.*, 2004) indiquent que, sous l'action d'une force extérieure, un dimère *BE* constitué uniquement des modules EC1 de deux C-cadhérines se dissocie sans modification structurale significative de chacun des modules (voir figure 4.11). Par ailleurs, le rôle du Trp 2 dans la force de l'interaction entre les deux monomères est démontré par des simulations réalisées sur un mutant dans lequel le tryptophane est remplacé par une alanine et qui se dissocie plus rapidement et pour des forces d'étirement moins importantes.

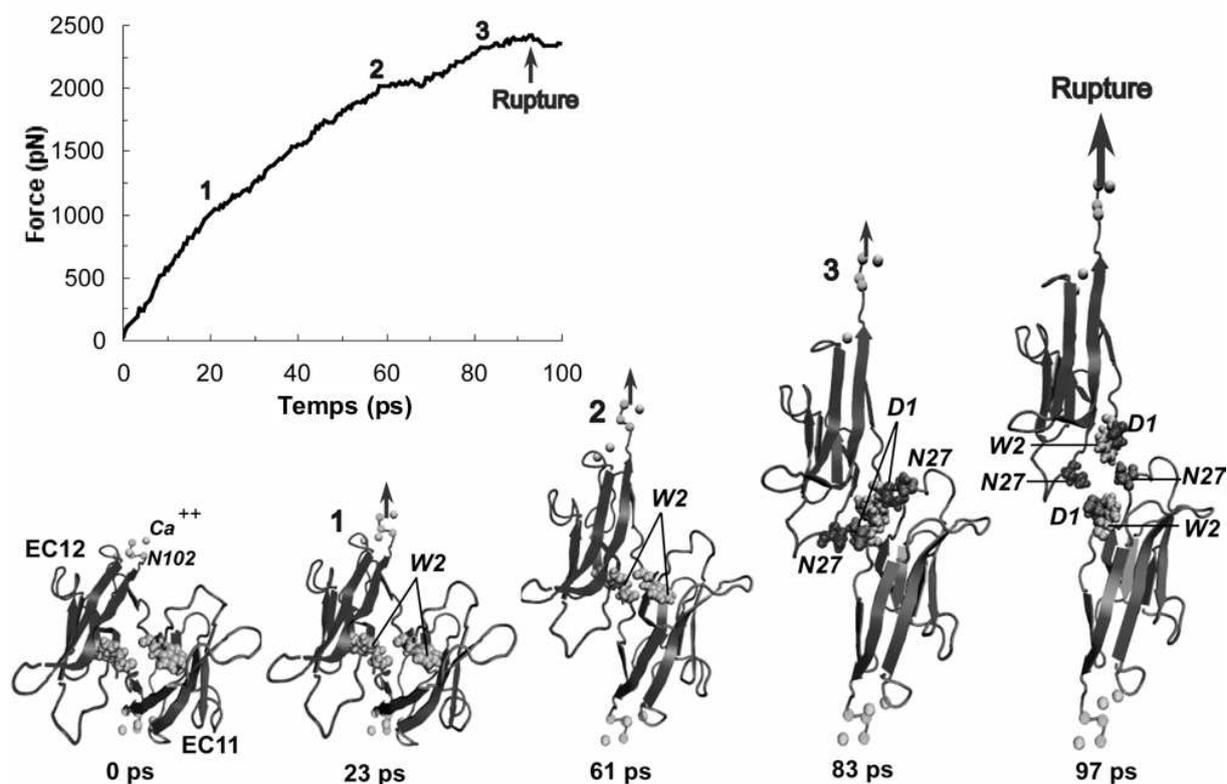


FIGURE 4.11 – Structures issues d’une dynamique moléculaire d’étirement du dimère *BE* de C-cadhérine et profil force-extension au cours de cette simulation. Reproduit de (Bayas *et al.*, 2004).

Dans leurs travaux, Sotomayor et collaborateurs (Sotomayor *et al.*, 2005) ont eux observé que le dépliage d’un module de cadhérine nécessite l’application de forces largement supérieures à celles auxquelles sont soumises ces molécules *in vivo*, en accord avec l’étude précédemment citée, qui montre que la dissociation du complexe formé par deux modules EC1 survient avant le dépliage de ces modules. De plus, ces expériences ont mis en évidence que les ions calcium apportent une stabilité supplémentaire à la structure du domaine.

Chapitre 5

Le rôle des ions calcium sur la stabilité des cadhérines

Sommaire

5.1	Introduction	83
5.2	Matériel et méthodes	84
5.2.1	Configuration initiale	84
5.2.2	Simulations de dynamique moléculaire	84
5.2.3	Analyse structurale	86
5.2.4	Estimation d'enthalpies libres de dimérisation	88
5.3	Étude des systèmes monomériques	89
5.3.1	Mise en évidence de changements conformationnels	89
5.3.2	Caractérisation des mouvements inter-domaines	91
5.4	Étude du dimère <i>BN</i> de <i>EC1-2</i>	94
5.4.1	Dynamique interne	94
5.4.2	Évolution de l'interface	95
5.4.3	Enthalpies libres de dimérisation	96
5.5	Conclusion	99

5.1 Introduction

Il a été démontré expérimentalement que la présence des ions Ca^{2+} permet de rigidifier la structure du domaine extracellulaire des cadhérines (voir section 4.3). Mais quelles sont exactement les conséquences au niveau atomique du retrait de ces ions? Afin de répondre à cette question, nous avons réalisé des simulations de dynamique moléculaire de fragments *EC1-2* de E-cadhérine en faisant varier le nombre d'ions calcium complexés.

L'impact structural de cette variation a été mesuré pour le monomère ainsi que pour le dimère *BN*. Nous avons également estimé la stabilité thermodynamique de ce dimère en présence et en l'absence de calcium.

5.2 Matériel et méthodes

5.2.1 Configuration initiale

Nous avons choisi pour ces simulations la structure 1FF5 du dimère *BN* (Pertz *et al.*, 1999) comme point de départ, principalement parce que la structure alternative 1EDH (Nagar *et al.*, 1996) est incomplète. En effet, l'extrémité N-terminale désordonnée dans le cristal, n'est pas résolue dans cette structure, notamment le résidu Trp 2, dont l'importance a été discutée dans la section 4.3. Nous avons cependant ôté le résidu Met (-1), qui n'est pas présent dans la molécule biologiquement active. La première chaîne du dimère a également été utilisée comme point de départ des simulations monomériques.

Le pKa des résidus ionisables a été calculé à l'aide du programme WHATIF (Vriend, 1990) : tous les résidus se trouvent dans leur état standard de protonation à pH 7, lorsque les trois ions calcium sont complexés. Lorsque ceux-ci sont retirés, certains résidus Glu et Asp de la sphère de coordination des ions calcium possèdent des valeurs de pKa très élevées indiquant qu'ils seraient protonés à pH neutre. Nous avons cependant choisi de les laisser dans leur état standard de protonation, car les sites de fixation se déforment rapidement et deviennent accessibles au solvant au cours de la dynamique moléculaire, permettant de retrouver des valeurs de pKa proches des valeurs standards. Les structures monomériques et dimériques de EC1-2 ont été immergées dans des boîtes octaédriques tronquées de molécules d'eau TIP3P (Jorgensen *et al.*, 1983), avec une distance minimale de 10 Å entre les bords de la boîte et le soluté. Vingt ions K^+ ont été ajoutés pour neutraliser la cellule de simulation et d'autres ions K^+ et Cl^- ont été ajoutés pour reproduire une concentration ionique proche de 0,15 mol.L⁻¹. Ces ions ont été placés aléatoirement dans la boîte d'eau.

5.2.2 Simulations de dynamique moléculaire

Toutes les simulations ont été réalisées avec le champ de force AMBER (Case *et al.*, 2002) et les paramètres parm99 (Wang *et al.*, 2000). Des conditions périodiques ont été utilisées et les interactions électrostatiques ont été traitées à l'aide de la méthode PME (York *et al.*, 1993 ; Cheatham *et al.*, 1995). Un pas d'intégration de 2 fs a été utilisé en appliquant des contraintes SHAKE (Ryckaert *et al.*, 1977) sur les liaisons impliquant un atome d'hydrogène.

L'équilibration des systèmes a été réalisée comme suit :

- 2000 pas de minimisation en restreignant la position des atomes de la protéine et les ions Ca^{2+} par des contraintes harmoniques avec une constante de rappel de $25 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$.
- 10 ps de dynamique moléculaire pendant lesquelles la température est graduellement augmentée pour atteindre 300 K, puis 40 ps de simulation à cette même température à volume et température constants.
- relâchement des contraintes harmoniques de 5 à $0,5 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$ au cours de séries composées de 1000 pas de minimisation et 25 ps de dynamique d'équilibration à pression et température constantes.
- 50 ps de dynamique moléculaire "libre", c'est-à-dire sans aucune contrainte sur les positions des atomes, dans l'ensemble *NPT*.

Après ces étapes d'équilibration, la phase de production est réalisée à pression et température constantes (respectivement fixées à 1 bar et 300 K) pour des durées variant de 5 à 15 ns (voir tableau 5.1).

Nom de la dynamique	Système modélisé	Équilibration	Temps de simulation
C3	monomère + 3 Ca^{2+}	oui	10 ns
C2	monomère + 2 Ca^{2+}	non ¹	6 ns + 5 ns ²
C0	monomère	oui	15 ns
K3	monomère + 3 K^+	non ³	5 ns
D6	dimère <i>BN</i> + 6 Ca^{2+}	oui	10 ns
D0	dimère <i>BN</i>	oui	10 ns

TABLE 5.1 – Récapitulatif des simulations effectuées et de leurs protocoles.

1. Le point de départ de la simulation C2 est la dernière structure issue de la simulation C3. L'ion Ca^{2+} du site CA3 a été retiré et remplacé par une molécule d'eau, de même que deux ions Cl^- afin de conserver l'électroneutralité de la cellule de simulation. Le système a ensuite été soumis à 500 pas de minimisation avec des contraintes de $5 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$ puis à 50 ps de dynamique moléculaire libre avant la phase de production.

2. La simulation de 5 ns a été débutée après 4,7 ns de la première simulation C2 (qui dure 6 ns). L'ion potassium venu s'insérer dans le site CA3 (voir section 5.3) est échangé avec une molécule d'eau. Le système est alors soumis à 1000 pas de minimisation sous contraintes de $5 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$, puis à 100 ps de dynamique moléculaire libre avant la phase de production.

3. Le point de départ de la simulation K3 est la dernière structure issue de la simulation C3. Les trois ions Ca^{2+} sont remplacés par une molécule d'eau ainsi que trois ions Cl^- afin de conserver l'électroneutralité dans la cellule de simulation. Le système a ensuite été soumis à 1000 pas de minimisation avec des contraintes de $5 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$, puis à 100 ps de dynamique moléculaire libre avant la phase de production.

5.2.3 Analyse structurale

L'ensemble des analyses standards des dynamiques moléculaires a été réalisé avec les logiciels *Ptraj* et *Carnal* de la suite AMBER 7 (Case *et al.*, 2002). Les déviations quadratiques moyennes¹ sont calculées, après superposition des structures, sur tous les atomes lourds (c'est-à-dire tous les atomes à l'exception des atomes d'hydrogène). L'évolution des distances entre les ligands opposés liant les ions Ca^{2+} a servi de mesure de la déformation des sites de fixation du calcium (voir figure 5.1). Pour les résidus Asp, Glu ou Gln, cette distance est calculée en se référant à l'atome d'oxygène ou d'azote le plus proche du calcium. Les surfaces accessibles au solvant² ont été calculées en utilisant une grille de 610 points sur une surface de Korobov et une sonde de rayon 1,4 Å correspondant à une molécule d'eau (Lavery *et al.*, 1981). La surface enfouie d'un résidu i (notée BSA_i pour "*Buried Surface Area*") est définie comme la différence entre la surface accessible de ce résidu dans le monomère et la surface accessible dans le dimère :

$$BSA_i = ASA_i^m - ASA_i^d$$

La surface d'interaction entre les deux monomères, notée IA par la suite pour "*Interfacial Area*", est calculée de façon similaire en retranchant la surface accessible du dimère à la somme des surfaces accessibles de chacun des monomères pris séparément :

$$IA = (ASA_1 + ASA_2) - ASA_{dim}$$

Cette grandeur toujours positive est d'autant plus grande que l'interaction entre les deux monomères est importante. Dans les complexes protéine-protéine, cette grandeur caractérise donc un aspect de la stabilité de l'édifice. Dans les complexes protéiques homodimériques, l'aire enfouie est typiquement comprise entre 2000 et 3600 Å² (Bahadur *et al.*, 2003 ; Bahadur *et al.*, 2004).

1. Notées ensuite RMSD pour "*root mean-square deviations*". Le RMSD entre deux structures contenant N atomes équivalents de masses m_i et dont les jeux de coordonnées sont $(x_i^1, y_i^1, z_i^1)_{i=1,N}$ et $(x_i^2, y_i^2, z_i^2)_{i=1,N}$ est défini par : $RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N m_i ((x_i^1 - x_i^2)^2 + (y_i^1 - y_i^2)^2 + (z_i^1 - z_i^2)^2)}$.

2. Notées par la suite ASA pour "*Solvent Accessible Surface*".

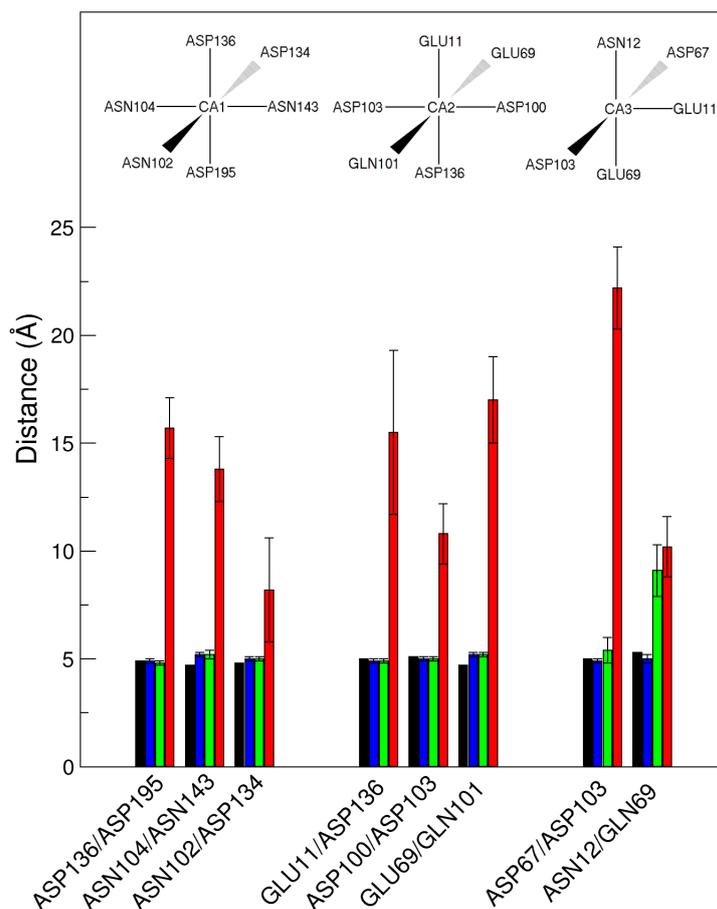


FIGURE 5.1 – Géométrie des sites de fixation du calcium dans les simulations monomériques. L'historgramme montre l'évolution de la géométrie de ces sites en termes de distance entre les atomes opposés dans la sphère de coordination des ions Ca^{2+} . Les quatre jeux de données correspondent à la structure cristallographique 1FF5 (en noir), et aux simulations C3 (en bleu), C2 (en vert) et C0 (en rouge). Pour les valeurs extraites des simulations, les déviations standards sont reportées sur les barres d'erreur.

Nous avons calculé un angle de pliure afin de caractériser la position relative des domaines EC1 et EC2 au cours des dynamiques moléculaires. Pour cela, nous avons défini un axe pour chacun des domaines (voir figure 5.2) en nous basant sur leur géométrie en tonneau β , en utilisant une approche proposée par Rosenberg et collaborateurs (Rosenberg *et al.*, 1976). Un résidu choisi arbitrairement sur l'un des brins β dans chaque domaine nous permet de définir un vecteur perpendiculaire à l'axe du domaine et ainsi de calculer un angle de torsion entre les domaines¹.

1. Les résidus Tyr 77 dans EC1 et Lys 129 dans EC2 ont été choisis pour donner une valeur de cet angle proche de 0 dans la structure cristallographique de référence.

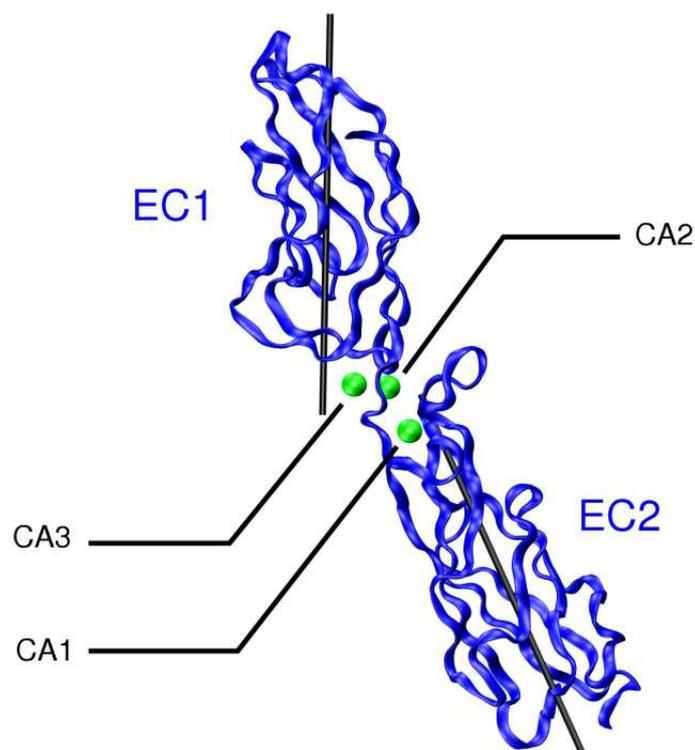


FIGURE 5.2 – Représentation en ruban du monomère EC1-2 de la E-cadhérine. Les trois ions Ca^{2+} sont représentés (par des sphères vertes) ainsi que les axes des deux domaines.

La présence des ions aux sites de liaison de la jonction inter-domaines est mesurée à l'aide de critères géométriques :

- un ion calcium est considéré lié s'il est positionné à moins de 4 Å d'au moins ($N - 1$) des N atomes liant l'ion initialement ($N = 6$ pour les sites CA1 et CA2, et $N = 5$ pour le site CA3).
- un ion potassium est considéré lié s'il est situé à moins de 4 Å d'au moins 2 de ces atomes.

Les critères moins contraignants utilisés pour l'ion potassium permettent de détecter la présence de cet ion, qui n'est initialement pas présent dans les sites de fixation.

5.2.4 Estimation d'enthalpies libres de dimérisation

Nous avons estimé des enthalpies libres de liaison en utilisant l'approche MM-PBSA (Kollman *et al.*, 2000 ; Gohlke et Case, 2004), dont le principe est exposé dans l'annexe G. 500 structures ont été extraites des simulations monomériques et dimériques concernées (échantillonnées toutes les 10 ps au cours des 5 dernières nanosecondes de simulation).

Les énergies conformationnelles ont été calculées avec le champ de force AMBER et les paramètres parm99. Les enthalpies libres de solvatation ont été calculées à l'aide d'un modèle Poisson-Boltzmann de représentation du solvant¹ (avec un rayon de Born de 1,72 Å pour les ions Ca^{2+} qui permet de retrouver une enthalpie libre de solvatation correcte pour cet ion (Babu *et al.*, 2003)) et les effets de solvatation non polaire ont été estimés à l'aide d'un terme de surface. Les entropies conformationnelles ont été estimées à partir de calculs de modes normaux (Gohlke et Case, 2004) pour 10 structures extraites régulièrement des 5 dernières nanosecondes de chaque simulation. Chaque structure a été soumise avant le calcul des modes normaux à une minimisation utilisant la méthode des gradients conjugués jusqu'à obtenir une valeur du gradient inférieur à $0,0001 \text{ kcal.mol}^{-1}.\text{Å}^{-1}$ dans un solvant implicite modélisé par une fonction diélectrique en $4r$.

5.3 Étude des systèmes monomériques

5.3.1 Mise en évidence de changements conformationnels

La figure 5.3 A montre l'évolution des RMSD calculés sur toute la structure EC1-2 au cours des simulations monomériques. On remarque que durant la simulation C3 (ligne bleue), la conformation reste proche de la structure expérimentale, avec un RMSD fluctuant autour de 2-3 Å. Le retrait de l'ion Ca^{2+} du site CA3 ne modifie pas significativement cette situation (ligne verte). Au cours de cette simulation C2, il est intéressant de remarquer que le site de fixation du calcium laissé vacant est spontanément occupé par un ion potassium monovalent (voir figure 5.4). Après 6 ns de simulation, nous avons donc décidé de redémarrer une nouvelle simulation après avoir échangé cet ion K^+ avec une molécule d'eau distante du site CA3 (voir la partie méthodologie 5.2). Après environ 2 ns, un nouvel ion potassium vient se fixer, mais aucune augmentation notable de la fluctuation de la structure n'est observé avant ou après cette seconde fixation, comme indiqué par l'évolution du RMSD figure 5.3 A. De récentes simulations ont montré que la constante de diffusion des ions potassium avec un modèle d'eau TIP3P est égale à deux fois sa valeur expérimentale (Várnai et Zakrzewska, 2004). Ainsi, la vitesse des phénomènes de fixation du potassium observés ici est peut-être surestimée. L'observation de la géométrie des sites de fixation du calcium montrée figure 5.1 appuie les résultats obtenus sur le comportement dynamique global de EC1-2. En effet, le retrait de l'ion calcium du site CA3 n'a qu'un effet limité sur ce site et aucun impact visible sur les deux autres sites de fixation.

1. Un espacement de grille de 0,67 Å a été choisi pour la résolution de l'équation de Poisson-Boltzmann par la méthode des différences finies (voir annexe F), la constante diélectrique du solvant a été fixée à 78,5 et celle du soluté à 1. Une concentration de $0,15 \text{ mol.L}^{-1}$ en sel monovalent a été modélisée.

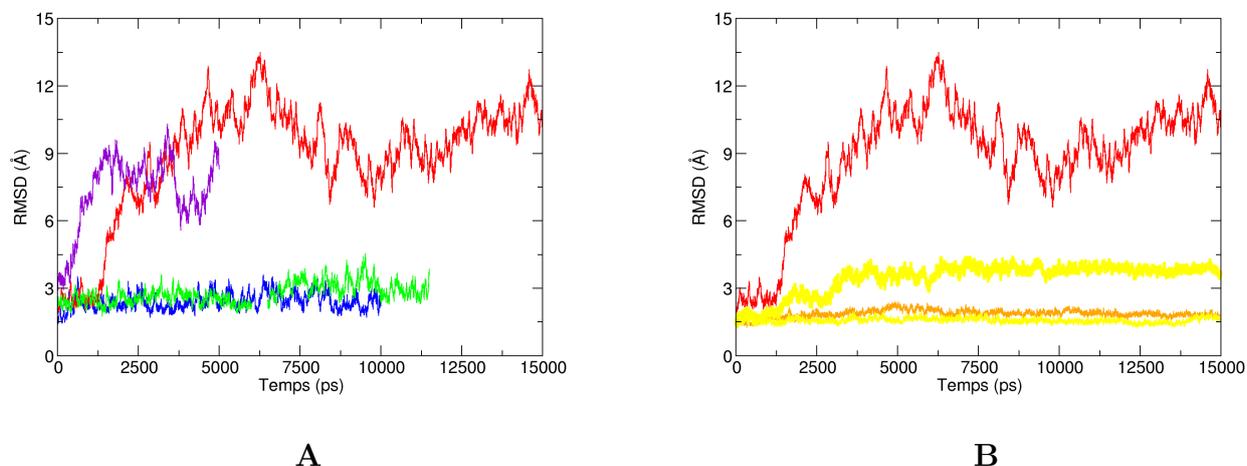


FIGURE 5.3 – RMSD calculé par rapport à la structure cristallographique du monomère EC1-2. **A** : Simulations C3 (en bleu), C2 (en vert), C0 (en rouge), et K3 (en violet). La rupture (pour 6 ns) dans la ligne correspondant à C2 indique le redémarrage de la simulation après le retrait de l'ion potassium du site CA3. **B** : RMSD calculés pour la simulation sur la totalité de EC1-2 (en rouge), et sur chacun des domaines EC1 (en orange) et EC2 (en jaune trait gras). Le trait fin jaune correspond au calcul du RMSD pour le domaine EC2, en excluant les résidus de la boucle flexible 130-145.

Lorsque les trois ions Ca^{2+} sont retirés de leurs sites de fixation, le fragment EC1-2 devient beaucoup plus flexible. Ceci est illustré par les valeurs de RMSD lues sur la figure 5.3 A qui fluctuent entre 7 et 14 Å (ligne rouge). Il est cependant intéressant de constater que ces importantes fluctuations ne commencent qu'après environ 1,5 ns de simulation. Ce retard semble être connecté avec la fixation d'un ion potassium au site CA2 (voir figure 5.4), suggérant que les ions potassium peuvent, au moins temporairement, stabiliser la jonction entre les deux domaines. Afin de déterminer si les ions K^+ peuvent entraîner une stabilisation à long terme, nous avons réalisé une autre simulation (K3) dans laquelle les trois sites sont occupés initialement par des ions K^+ . Les résultats de la figure 5.4 montrent que dans ce cas l'ion du site CA3 se dissocie rapidement, tandis que les deux autres ions restent en place. Ce départ de l'ion potassium du site CA3 peut être attribué à une déformation du site de fixation. Le site CA2 est également déformé à un degré moindre, même si l'ion reste en place¹. La présence des ions K^+ ne permet cependant pas de stabiliser la jonction : on peut observer sur la figure 5.3 A que les valeurs de RMSD dans la simulation K3 (ligne violette) atteignent rapidement celles observées dans la simulation C0 (ligne rouge).

1. Il est à noter néanmoins que la mesure de la fixation de l'ion K^+ au niveau d'un site de liaison est moins contraignante que celle d'un ion Ca^{2+} .

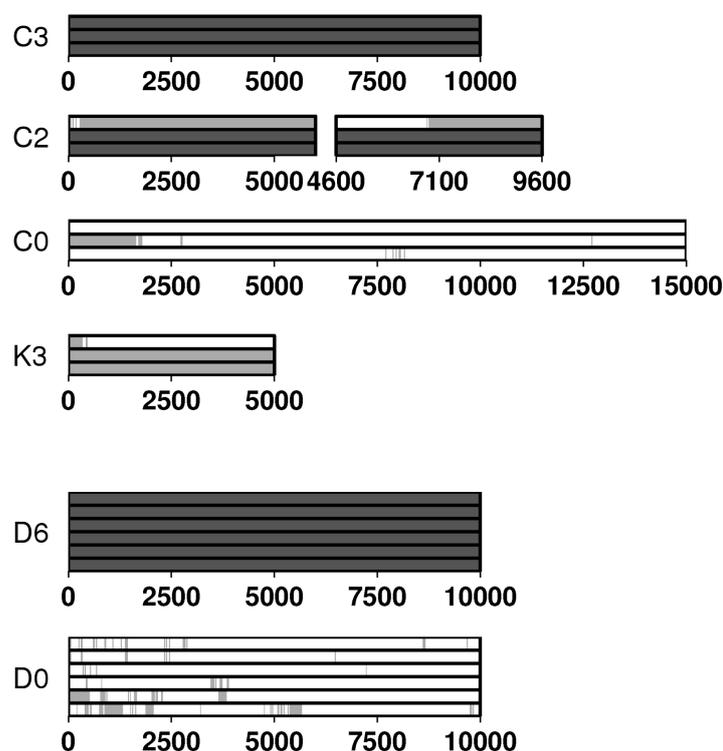


FIGURE 5.4 – Occupation des sites de fixation du calcium au cours des simulations de dynamique moléculaire (temps en ps). Les zones grisées correspondent à la fixation d'un ion Ca^{2+} et les zones ombrées correspondent à la fixation d'un ion K^+ . Chaque groupe de trois barres correspond aux sites CA1, CA2, et CA3 d'un monomère en allant du bas vers le haut.

5.3.2 Caractérisation des mouvements inter-domaines

En l'absence d'ion calcium aux sites de fixation, l'augmentation du RMSD indique un changement conformationnel de grande ampleur. La figure 5.3 B montre l'évolution du RMSD calculé sur chacun des domaines EC1 et EC2 au cours de la simulation C0. On peut voir que ces deux domaines restent stables pendant l'ensemble de la simulation avec des RMSD inférieurs à 4 Å. Le domaine EC2 est cependant moins stable que le domaine EC1, pour lequel le RMSD n'excède que très rarement 2 Å. Cette différence provient d'une boucle flexible dans le domaine EC2, située entre les résidus 130 et 145, qui contient trois résidus (Asp 134, Asp 136, et Asn 143) impliqués dans la sphère de coordination des sites CA1 et CA2. Lorsque cette boucle est retirée du calcul de RMSD, la figure 5.3 B montre que le reste du domaine EC2 est sujet à des fluctuations similaires à celles observées pour le domaine EC1. Puisque la majorité des modifications structurales a lieu au niveau de la jonction inter-domaines, il n'est pas surprenant de constater que la géométrie des sites de fixation du calcium est fortement perturbée (voir figure 5.1).

Afin de caractériser les mouvements relatifs des deux domaines du fragment de cadhérine étudié, nous avons utilisé les angles de pliure et de torsion définis dans la partie 5.2. Les résultats sont donnés pour l'ensemble des simulations dans le tableau 5.2 et illustrés sur la figure 5.5. Les deux angles subissent clairement des fluctuations plus importantes dans les simulations sans calcium, indiquant de plus grands mouvements inter-domaines. Lorsque les ions Ca^{2+} sont présents, les deux domaines restent approximativement coaxiaux, bien que la pliure observée soit légèrement supérieure à celle mesurée dans la structure cristallographique. La torsion entre les deux domaines reste également proche de celle observée dans la structure cristallographique et ne subit que de faibles fluctuations. En accord avec les RMSD montrés dans la figure 5.3 A, les résultats obtenus pour la simulation C2 indiquent que le retrait de l'ion situé dans le site CA3 n'a que peu d'influence sur les propriétés mécaniques de EC1-2 et ne réduit pas significativement la rigidité de la jonction inter-domaines. Au contraire, sans calcium, cette jonction devient particulièrement flexible avec un angle de pliure qui varie entre 60° et 120° . Le mouvement de charnière est également couplé à une liberté de torsion accrue entre les domaines, comme montré sur la figure 5.5 B, avec des valeurs au cours de la trajectoire couvrant une gamme de 120° contre 40° en présence de calcium.

Simulation	Intervalle de mesure (ns)	Pliure ($^\circ$)	Torsion ($^\circ$)
C3	2,5 - 10	137 ± 6	-18 ± 9
C2	2,5 - 11	144 ± 7	-31 ± 9
C0	5 - 15	81 ± 17	27 ± 24
K3	2,5 - 5	84 ± 10	-26 ± 13
D6	2,5 - 10	138 ± 7	-16 ± 6
		147 ± 5	-15 ± 7
D0	5 - 10	107 ± 11	6 ± 9
		139 ± 11	-79 ± 11
Cristal		154	-16
		152	-16

TABLE 5.2 – Valeurs moyennes et écarts types des angles de pliure et de torsion entre les domaines EC1 et EC2 au cours des dynamiques moléculaires. Les données de la figure 5.3 A ont servi à déterminer les intervalles de temps pendant lesquels les RMSD de chaque dynamique sont assez stables pour permettre le calcul de grandeurs moyennes. L'entrée "Cristal" renvoie à la structure cristallographique 1FF5.

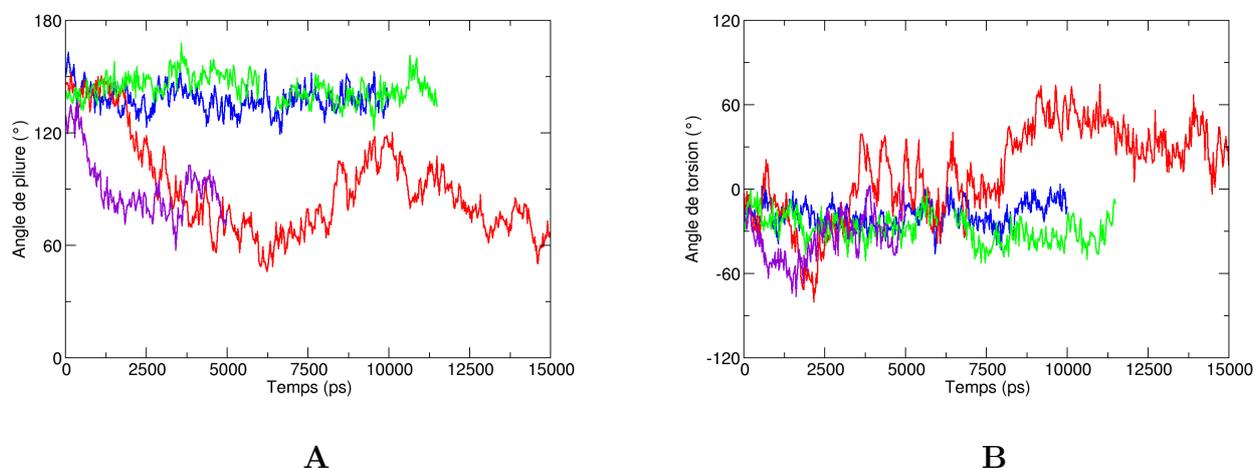


FIGURE 5.5 – Évolution des angles inter-domaines au cours des simulations C3 (en bleu), C2 (en vert), C0 (en rouge), et K3 (en violet). **A** : Angle de pliure ; **B** : Angle de torsion.

La figure 5.6 illustre les variations des angles de pliure et de torsion inter-domaines au cours de la simulation C0, à l'aide de structures extraites à intervalles réguliers de 5 ns. La variation de la position relative de EC2 par rapport à EC1 traduit bien la grande flexibilité de la jonction inter-domaines en l'absence de calcium.

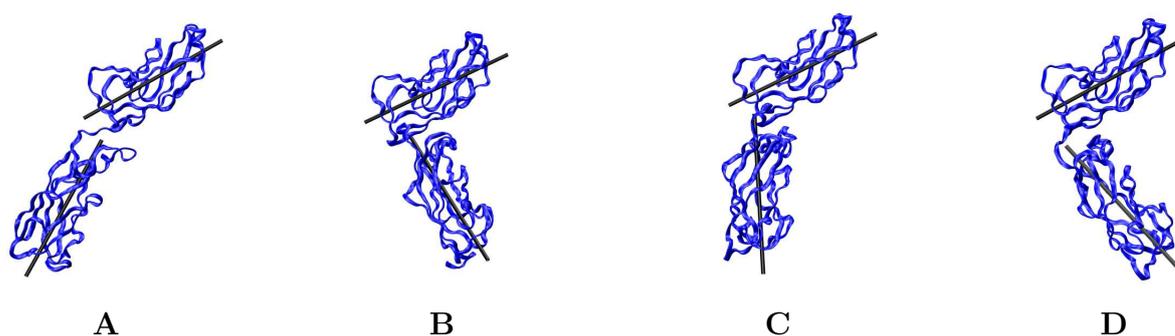


FIGURE 5.6 – Structures extraites de la simulation C0 à différents temps t montrant la flexibilité de la jonction inter-domaines en l'absence d'ion Ca^{2+} complexé. Les domaines supérieurs EC1 des structures ont été placés dans une même orientation. Les axes des domaines sont indiqués en noir. **A** : $t = 0$ ns ; **B** : $t = 5$ ns ; **C** : $t = 10$ ns ; **D** : $t = 15$ ns.

5.4 Étude du dimère *BN* de EC1-2

5.4.1 Dynamique interne

En présence des six ions Ca^{2+} complexés, le dimère *BN* reste stable, comme l'indique l'évolution du RMSD (figure 5.7 A), qui fluctue entre 2 et 3 Å, qu'il soit calculé sur le dimère entier ou sur chacun des monomères. Ces valeurs sont similaires à celles observées dans les simulations du monomère isolé. La situation est complètement différente lorsque les ions calcium sont retirés : les valeurs du RMSD augmentent alors en effet jusqu'à atteindre environ 6 Å (voir figure 5.7 B), traduisant d'importants changements conformationnels. Cette instabilité structurale est répartie au niveau de chacun des monomères qui constituent le dimère *BN* (en orange). En fait, l'un des deux monomères semble être transitoirement plus stable que l'autre, peut-être en raison de fixations plus fréquentes d'ions K^+ au niveau de la jonction inter-domaines. Ces fixations ont principalement lieu dans les sites CA1 et CA2 du monomère, comme illustré sur les trois barres inférieures du diagramme de fixation des ions au cours de la simulation D0 de la figure 5.4.

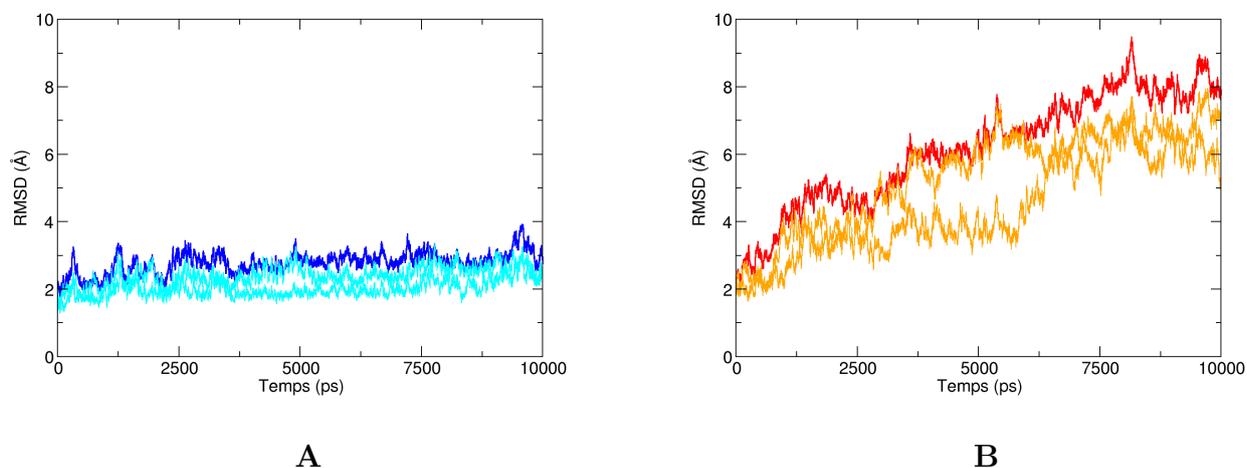


FIGURE 5.7 – RMSD calculé par rapport à la structure cristallographique pour les simulations D6 (A) et D0 (B). Les lignes bleue et rouge correspondent au RMSD calculé pour le dimère et les lignes cyan et orange correspondent aux RMSD calculés sur chacun des monomères.

Il est toujours possible d'étudier les changements conformationnels en termes de modification des angles de pliure et de torsion pour chacun des monomères dans les simulations D6 et D0. Les résultats sont montrés figure 5.8 et résumés dans le tableau 5.2. En présence des six ions Ca^{2+} , à la fois les angles de pliure et de torsion restent proches des valeurs

cristallographiques et montrent de petites fluctuations. Au contraire, en l'absence de ces ions calcium, on observe au cours de la simulation D0 une déviation de ces angles par rapport à leurs valeurs initiales, accompagnée d'une augmentation de leurs variations.

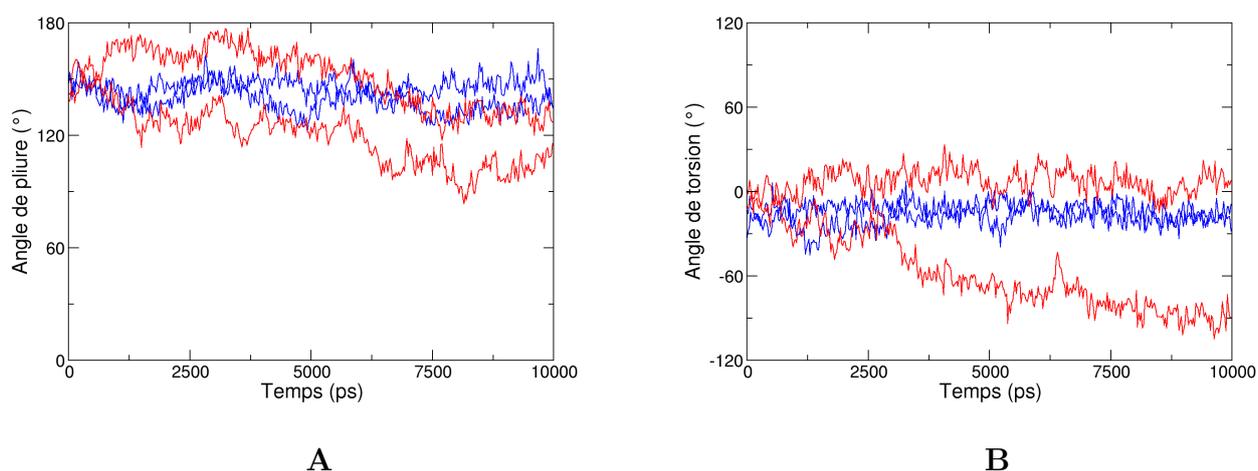


FIGURE 5.8 – Évolution des angles inter-domaines des monomères EC1-2 au cours des simulations D6 (en bleu) et D0 (en rouge). **A** : Angle de pliure ; **B** : Angle de torsion.

En résumé, la fixation des ions calcium rigidifie les monomères et les maintient dans une conformation optimale pour la formation du dimère *BN*. Le départ de ces ions entraîne d'importantes déformations du dimère, bien que les fluctuations au sein des monomères soient atténuées par l'existence de l'interface dimérique.

5.4.2 Évolution de l'interface

L'évolution de l'aire interfaciale lors des simulations D6 et D0 est représentée sur la figure 5.9. Cette mesure atteste la stabilité du dimère *BN* lorsque le calcium est présent, avec une aire de l'interface qui fluctue autour d'une valeur moyenne de 2400 \AA^2 environ. Au contraire, en l'absence des ions Ca^{2+} , de grandes fluctuations sont observées, avec temporairement une diminution de l'aire interfaciale de l'ordre de 35 %. En fait, même lorsque l'aire d'interface augmente de nouveau pour atteindre des valeurs similaires à celles observées dans la simulation D6 (après environ 10 ns de simulation), ceci correspond à une interface dimérique fortement modifiée par rapport à l'interface initiale.

Afin d'analyser ces modifications, nous avons regardé l'évolution des aires enfouies de chacun des acides aminés des monomères EC1-2. L'essentiel de l'interface dimérique initiale

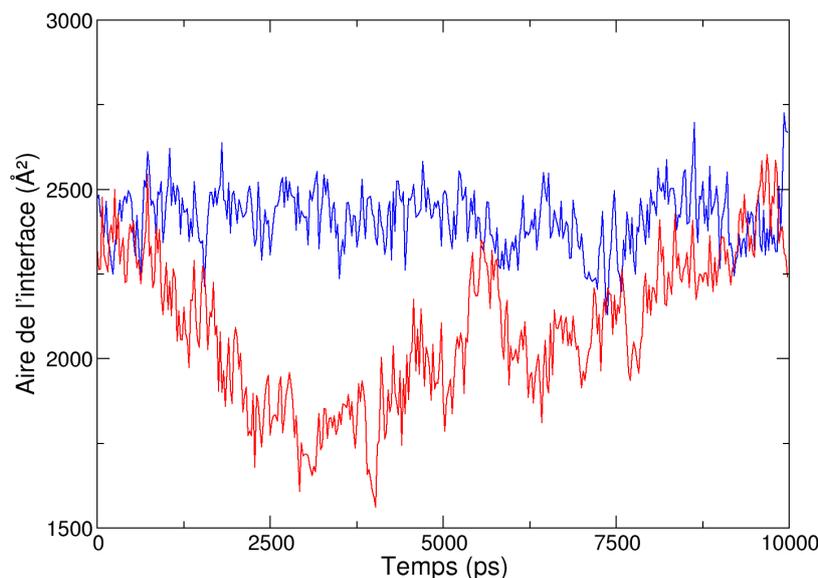


FIGURE 5.9 – Évolution de l'aire de l'interface dimérique au cours des simulations D6 (en bleu) et D0 (en rouge).

est constitué des 25 résidus N-terminaux, ainsi que des résidus autour des positions 100, 140 et 200 dans la séquence (voir figure 5.10). La figure 5.11 présente l'évolution de l'aire enfouie pour les résidus impliqués dans l'interface dimérique. Dans la dynamique D6 en présence de calcium, peu de changements sont observés, tandis qu'il est clair que l'interface est fortement modifiée dans la simulation D0 où les ions Ca^{2+} ont été retirés. Ces modifications consistent principalement en une perte de contact au niveau des brins N-terminaux ainsi qu'autour du résidu 100 (certainement en raison de répulsions électrostatiques, puisque plusieurs de ces résidus normalement impliqués dans la fixation des ions Ca^{2+} possèdent une charge négative). En revanche, on remarque l'apparition de nouveaux contacts autour des résidus 90 et 195, responsables de l'augmentation de l'aire de l'interface dans la deuxième partie de la simulation.

5.4.3 Enthalpies libres de dimérisation

Nous avons voulu déterminer l'influence des ions calcium sur la stabilité thermodynamique du dimère *BN* de EC1-2. Pour cela, nous avons calculé l'enthalpie libre de formation du dimère *BN* dans différents états de complexation à l'aide de la méthode MM-PBSA. Les résultats sont donnés dans le tableau 5.3. En présence de calcium, l'enthalpie libre de formation du dimère est favorable, avec une valeur proche de $-30 \text{ kcal.mol}^{-1}$, causée par un terme enthalpique nettement favorable. En revanche, en l'absence d'ions Ca^{2+} , le dimère

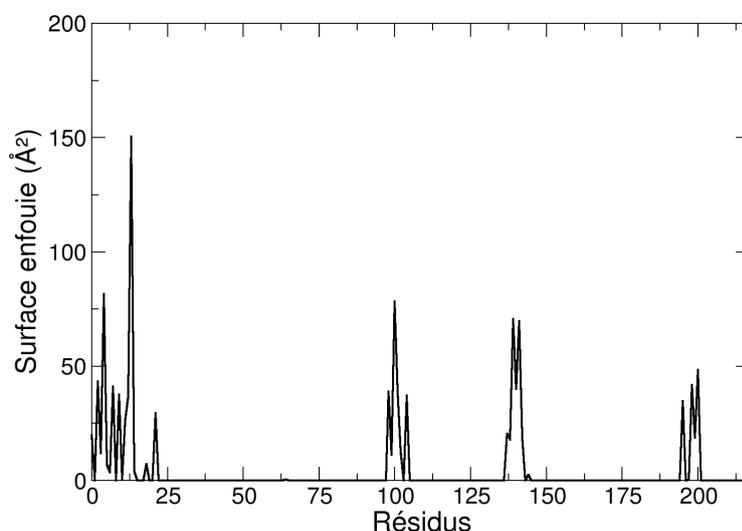


FIGURE 5.10 – Aire enfouie par chacun des résidus du monomère EC1-2 dans la structure 1FF5 du dimère *BN*.

est clairement instable. En effet, malgré un terme entropique légèrement diminué grâce à une flexibilité accrue du dimère, le terme enthalpique est fortement défavorable. Les valeurs obtenues par la méthode MM-PBSA sont largement surestimées en valeur absolue en raison des limitations de la méthode (voir annexe G) et ne peuvent être comparées avec les données expérimentales qui correspondent à une valeur proche de -4 kcal.mol^{-1} pour la dimérisation de fragments EC1-2 de E-cadhérine (Alattia *et al.*, 1997 ; Koch *et al.*, 1997 ; Häussinger *et al.*, 2004).

Dimérisation	ΔH	$-T\Delta S$	ΔG
$2 \times C3 \rightleftharpoons D6$	-84	55	-29
$2 \times C0 \rightleftharpoons D0$	100	48	148
$2 \times C3 \rightleftharpoons D6$ (sans les ions)	-27	-	-
$2 \times C3 \rightleftharpoons D6$ (sans l'ion CA3)	-99	-	-
$2 \times C3 \rightleftharpoons D6$ (avec des ions K^+)	-65	-	-

TABLE 5.3 – Enthalpies libres de formation (en kcal.mol^{-1}) du dimère *BN* de EC1-2 estimées par la méthode MM-PBSA. Les trois dernières lignes du tableau ont été obtenues avec des structures issues des simulations C3 et D6 après avoir retiré tous les ions calcium, seulement l'ion calcium positionné dans le site CA3, ou avoir remplacé tous les calcium par des potassium.

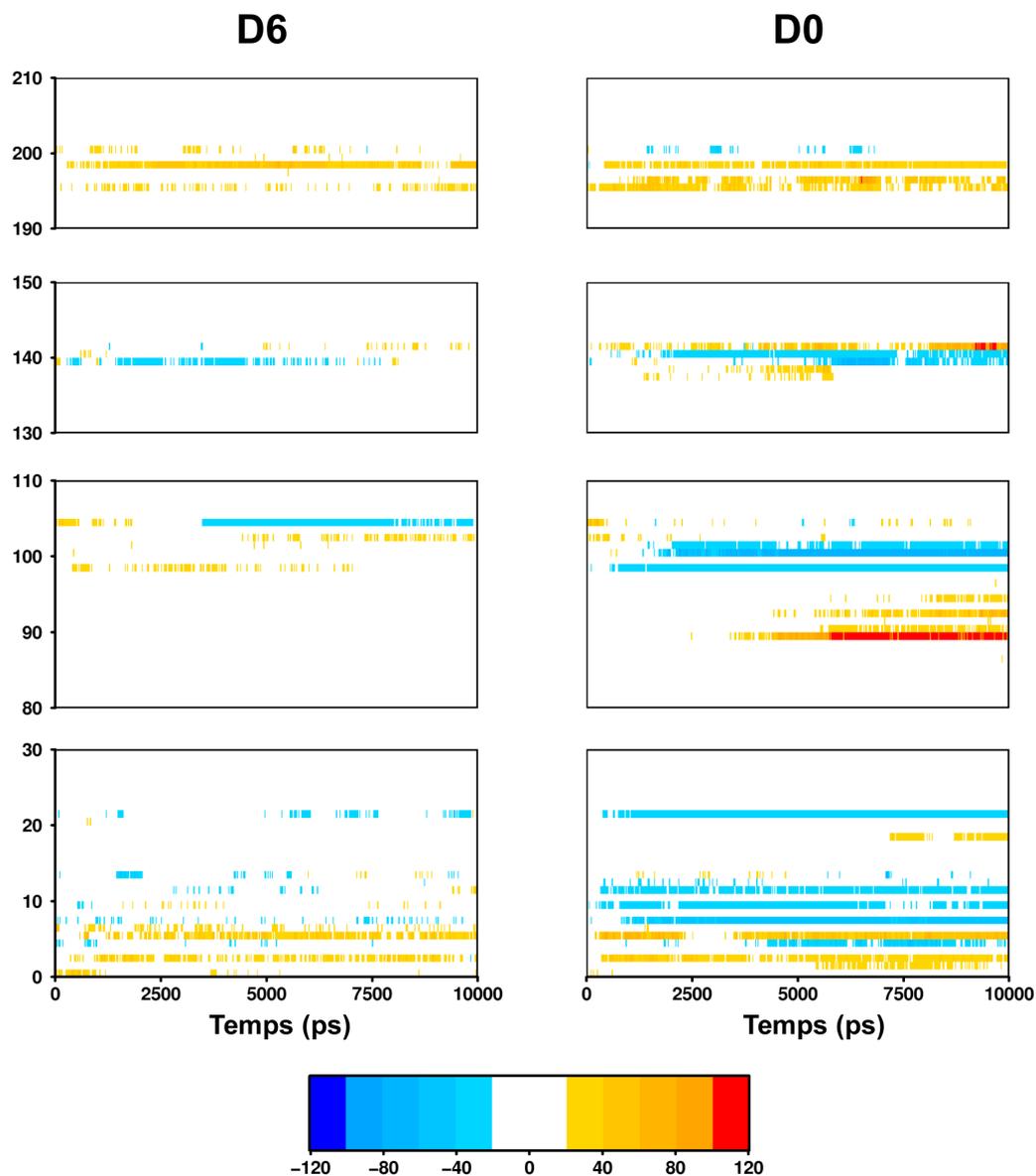


FIGURE 5.11 – Détail de l’interface au cours des simulations D6 et D0 du dimère *BN*. La variation de l’aire enfouie d’un résidu par rapport à la même grandeur mesurée dans la structure cristallographique 1FF5 est repérée par l’échelle de couleur bleu \rightarrow blanc \rightarrow rouge.

La différence de stabilité provenant du retrait des ions calcium peut en partie être expliquée par la répulsion électrostatique entre les résidus anioniques des sites CA1 et CA2, proches de l’interface dimérique. En calculant une enthalpie de dimérisation fictive pour un système sans ion calcium en utilisant les structures issues de C3 et D6 (voir tableau 5.3), on peut estimer cette répulsion à environ 60 kcal.mol^{-1} . Ceci implique que la dimérisation en l’absence d’ions Ca^{2+} est déstabilisée d’encore 70 kcal.mol^{-1} du fait des changements de

conformation du monomère et de la destruction de l'interface dimérique. Il est également intéressant de noter qu'en l'absence d'ion calcium au niveau du site CA3, l'enthalpie de dimérisation calculée à partir des structures issues de C3 et D6 est légèrement plus favorable que lorsque les trois sites sont occupés. Bien que ce terme soit susceptible d'être compensé par une augmentation de la pénalité entropique, cette constatation soutient l'idée que le cation du site CA3 n'est pas essentiel pour la formation du dimère *BN*. La dernière ligne du tableau 5.3 montre enfin qu'un complexe stable pourrait en principe être formé si les ions Ca^{2+} étaient remplacés par des ions K^+ (en dépit de l'augmentation de la charge négative portée par chaque monomère). Ce résultat a cependant été obtenu en utilisant des structures issues des simulations C3 et D6, dans lesquelles la structure des sites de fixation à la jonction des domaines EC1 et EC2 est maintenue. En fait, comme la simulation K3 l'a montré, les ions potassium monovalents ne sont pas capables de stabiliser la structure du fragment EC1-2. La dimérisation serait donc défavorable non pas en raison de l'augmentation des répulsions électrostatiques, mais à cause de la perte de structure.

5.5 Conclusion

Les simulations de dynamique moléculaire ont permis d'obtenir une vue détaillée du rôle des ions calcium sur les propriétés mécaniques des deux domaines extracellulaires les plus importants de la E-cadhérine. En accord avec plusieurs études expérimentales, nous avons constaté que la jonction entre les domaines EC1 et EC2 se comporte comme une région charnière en l'absence d'ions Ca^{2+} . Ce résultat est conforme aux expériences de microscopie électronique discutées dans la section 4.3 et qui montrent l'effondrement de la structure dans ces conditions. La présence des ions calcium rigidifie la jonction en termes de pliure et de torsion ce qui est en accord avec les données obtenues par RMN (Häussinger *et al.*, 2002). Les simulations indiquent par ailleurs que seuls deux des ions calcium semblent jouer un rôle important dans cette rigidification, l'absence de l'ion le plus exposé au solvant (placé au site CA3) n'ayant que peu d'impact sur la dynamique de la jonction inter-domaines. Nous avons de plus observé que les sites de fixation laissés vacants sont susceptibles d'être, au moins temporairement, occupés par des ions potassium. Ce résultat est en accord avec la non-équivalence des sites de fixation du calcium, en ce sens que la fixation d'un ion au site CA3 nécessite que les sites CA1 et CA2 soient correctement formés et occupés.

Nous avons également retrouvé que les ions calcium sont essentiels à la formation du dimère BN , en fixant la conformation des monomères et en limitant les fluctuations au sein du dimère. Les estimations d'enthalpie libre confirment que la dimérisation est fortement défavorisée en l'absence d'ions calcium, en raison de deux effets approximativement égaux : une interaction électrostatique défavorable provenant de la répulsion entre les résidus négativement chargés des sites $CA1$ et $CA2$, et une déformation de l'interface. Ceci est encore une fois en accord avec les observations expérimentales qui montrent que la fixation des ions Ca^{2+} est une étape préalable nécessaire à l'association de deux monomères de $EC1-2$.

Chapitre 6

Dynamique des interfaces

Sommaire

6.1	Introduction	101
6.2	Matériel et méthodes	102
6.2.1	Construction des dimères	102
6.2.2	Construction des cellules de simulation	103
6.2.3	Simulations de dynamique moléculaire	103
6.2.4	Analyse conformationnelle et estimation d'enthalpies libres de dimérisation	104
6.3	Stabilité structurale des dimères	105
6.4	Changements conformationnels du dimère <i>BE</i>	107
6.5	Enthalpies libres de dimérisation	111
6.6	Interfaces et spécificité	112
6.7	Conclusion	115

6.1 Introduction

Différentes interfaces de dimérisation ont été observées pour les cadhérines classiques, qui ont donné lieu à l'élaboration de divers modèles pour expliquer les propriétés adhésives de ces molécules (voir paragraphe 4.3). Aucun de ces modèles n'a cependant été définitivement validé expérimentalement. Bien que les plus récentes études soulignent l'importance du dimère *BE* pour l'adhésion, certaines propriétés sont difficiles à expliquer uniquement à partir de cette structure et il n'est pas clair que les autres dimères observés n'ont aucun rôle à jouer.

En ce qui concerne la spécificité des interactions entre cadhérines, des zones d'ombre subsistent. En dépit du fort taux d'homologie de séquence entre les différentes cadhérines classiques, il a longtemps été admis que les interactions homophiles étaient largement favorisées. De récentes études ont cependant remis cette idée en question, en montrant que des interactions hétérophiles pouvaient se former, et que la sélectivité observée à l'échelle cellulaire pourrait provenir d'un effet d'accumulation (Patel *et al.*, 2003 ; Foty et Steinberg, 2005). Un modèle a d'ailleurs été proposé qui montre que de très petites différences d'affinité pour la formation du dimère *BE* entre des interactions homophiles ou hétérophiles peuvent être amplifiées de cette façon pour conduire à une ségrégation de cellules exprimant des cadhérines différentes (Chen *et al.*, 2005). L'implication du dimère *BE* pour la spécificité des interactions entre cadhérines classiques a été par ailleurs démontrée dans une récente publication de Patel et collaborateurs (Patel *et al.*, 2006). Cette étude apporte également de nouvelles preuves de la participation du domaine EC1 dans la spécificité des interactions, même si, en raison des faibles différences des séquences, la spécificité à l'intérieur des sous-familles de cadhérines de type I ou II reste difficile à expliquer sur un plan purement structural. Dans ce contexte, il est intéressant de noter que des expériences de BFP ont suggéré que la sélectivité pouvait être reliée à l'aspect cinétique des interactions plutôt qu'à l'aspect thermodynamique (Perret *et al.*, 2004 ; Bayas *et al.*, 2006).

Afin d'apporter de nouveaux éléments de réponse à ces questions qui restent posées à l'heure actuelle, nous avons réalisé des simulations de dynamique moléculaire sur les trois dimères de E-cadhérine dont une structure est disponible (voir figure 4.8 page 77). Nous avons ainsi comparé leurs propriétés dynamiques et thermodynamiques pour évaluer leur possible implication dans les phénomènes adhésifs.

6.2 Matériel et méthodes

6.2.1 Construction des dimères

Nous avons étudié trois dimères de EC1-2 dans cette étude :

- le dimère *BN* a été directement extrait de la structure 1FF5 (chaînes A et B). Comme indiqué dans le paragraphe 5.2, nous avons cependant retiré le résidu Met (-1) qui n'est pas présent dans la molécule biologique.

- le dimère *BE* a été construit à partir de la structure cristallographique 1Q1P en appliquant l'opération de symétrie adéquate sur la chaîne A. Dans cette structure, les résidus Asp 1, ainsi que les résidus 214 à 218 sont absents. Nous les avons donc rajoutés à l'aide du module *TLeap* du logiciel AMBER (Case *et al.*, 2004). La conformation des quatre résidus C-terminaux a directement été copiée de celle présente dans le dimère *BN*.
- le dimère *décalé* a été obtenu en appliquant une opération de symétrie à la chaîne A du dimère *BN*. Bien que celui-ci aurait pu être construit à partir de la structure du dimère *BE*¹, nous n'avons pas retenu cette solution en raison des résidus manquant dans cette dernière structure. De plus, les brins N-terminaux des monomères auraient été exposés au solvant dans une telle structure, en raison de l'échange de brins dans le dimère *BE*.

6.2.2 Construction des cellules de simulation

Le pKa de tous les groupes ionisables ont été calculé avec le programme WHATIF (Vriend, 1990). Les valeurs obtenues indiquent que tous les résidus sont dans leur état de protonation standard, donnant lieu à une charge totale de -20 pour chaque dimère (en tenant compte des six ions Ca^{2+} complexés). Chaque structure dimérique a été solvatée par des molécules d'eau explicites TIP3P (Jorgensen, 1981) et 20 ions K^+ ont été ajoutés pour neutraliser les cellules de simulation. Afin d'obtenir une concentration saline proche de $0,15 \text{ mol.L}^{-1}$, des ions K^+ et Cl^- ont été rajoutés et tous les ions monovalents ont été placés aléatoirement autour du soluté. Nous avons utilisé des boîtes octaédriques tronquées pour les dimères *BN* et *BE*, mais une boîte parallélépipédique pour le dimère *décalé* en raison de la forme très allongée de cette conformation (voir figure 4.8 C). Dans tous les cas, une distance minimale de 10 \AA a été imposée entre le bord de la boîte et le soluté. Au final, les systèmes contiennent environ 85 000, 147 000 et 230 000 atomes respectivement pour les dimères *BN*, *décalé* et *BE*.

6.2.3 Simulations de dynamique moléculaire

Les simulations de dynamique moléculaire ont été réalisées avec le module *PMEMD* de la suite AMBER8 (Case *et al.*, 2004) et le champ de force parm99 (Wang *et al.*, 2000). Des conditions périodiques ont été utilisées et les interactions électrostatiques ont été traitées avec la méthode PME (York *et al.*, 1993 ; Cheatham *et al.*, 1995) et un rayon de coupure

1. Les dimères *décalés* construits à partir des structures 1FF5 ou 1Q1P ne diffèrent que par un RMSD calculé sur les C_α de $1,9 \text{ \AA}$.

de 9 Å dans l'espace réel. Des contraintes SHAKE (Ryckaert *et al.*, 1977) ont été appliquées à toutes les liaisons impliquant un atome d'hydrogène, ce qui a permis d'utiliser un pas d'intégration de 2 fs.

Le protocole utilisé pour les simulations est similaire à celui détaillé dans le paragraphe 5.2. Cependant, pour les dimères *décalé* et *BE*, les phases de minimisation et d'équilibration ont été rallongées en raison du nombre d'atomes beaucoup plus importants dans ces systèmes. De plus, dans le cas du dimère *décalé*, des contraintes quadratiques ont été appliquées sur les deux atomes d'une liaison approximativement parallèle au plus grand axe de la cellule de simulation¹. Ceci permet d'éviter une rotation globale de la molécule perpendiculairement à cet axe, ce qui pourrait engendrer des contacts entre la molécule et son image dans la cellule voisine.

Les phases de production représentent 10 ns pour le dimère *BN* et 12,5 ns pour les dimères *décalé* et *BE*. Les comparaisons avec les systèmes monomériques renvoient aux simulations présentées dans le chapitre 5.

6.2.4 Analyse conformationnelle et estimation d'enthalpies libres de dimérisation

Les analyses conformationnelles classiques ont été réalisées à l'aide des modules *Carnal* et *Ptraaj* de la suite AMBER, ainsi qu'avec le logiciel Profit développé par A.C.R. Martin². Les surfaces accessibles et aires d'interface ont été calculées selon la même méthode que celle présentée dans la section 5.2, de même que les angles de pliure et de torsion entre les deux domaines EC d'un monomère.

Le calcul des enthalpies libres de dimérisation a été réalisé avec la méthode MM-PBSA décrite dans l'annexe G et dans la section 5.2. Il est important de rappeler que dans la simulation monomérique, le Trp 2 est inséré dans la poche hydrophobe du monomère. Ainsi, lors du calcul de l'enthalpie libre de formation du dimère *BE*, l'effet de l'échange de brins est pris en compte.

1. Les deux atomes choisis sont les atomes C_α et C' de l'isoleucine 242.

2. <http://www.bioinf.org.uk/software/profit>.

6.3 Stabilité structurale des dimères

Nous avons déjà observé qu’au sein du dimère *BN*, en présence d’ions Ca^{2+} , la conformation des monomères EC1-2 reste stable tout au long de la simulation (voir section 5.4), avec des RMSD par rapport aux structures initiales inférieurs à 2 Å (ou 3 Å pour la simulation du dimère *BE*, en raison du mouvement des résidus N-terminaux). Cette stabilité est également reflétée dans les valeurs moyennes et les fluctuations des angles de pliure et de torsion entre les domaines EC1 et EC2, qui restent proches des valeurs observées dans les simulations monomériques (voir tableau 6.1). On note cependant que les angles de pliure sont plus faibles que ceux mesurés dans les structures cristallographiques, probablement du fait de l’absence des contraintes dues à l’empaquetage cristallin. Ces résultats montrent que les jonctions au sein des monomères EC1-2 sont rigides et non affectées par l’existence ou le type de l’interface dimérique.

Simulation	Intervalle de mesure (ns)	Pliure (°)	Torsion (°)
monomère	2,5 - 10	137 ± 6 (154)	-18 ± 9 (-16)
dimère <i>BN</i>	2,5 - 10	138 ± 7 (154)	-16 ± 6 (-16)
		147 ± 5 (154)	-15 ± 7 (-16)
dimère <i>BE</i>	2,5 - 12,5	141 ± 7 (148)	-18 ± 8 (-21)
		134 ± 9 (148)	-19 ± 8 (-21)
dimère <i>décalé</i>	2,5 - 12,5	138 ± 8 (154)	-29 ± 9 (-16)
		152 ± 9 (154)	-17 ± 9 (-16)

TABLE 6.1 – Valeurs moyennes et déviations standards des angles de pliure et de torsion entre les domaines EC1 et EC2 au cours des dynamiques moléculaires. Les données de la figure 6.1 ont servi à déterminer les intervalles de mesures pendant lesquels les RMSD de chaque dynamique sont assez stables pour permettre le calcul de moyennes. Les valeurs entre parenthèses correspondent aux valeurs mesurées dans les structures cristallographiques.

La figure 6.1 présente l’évolution du RMSD des différents dimères par rapport aux structures cristallographiques au cours des simulations. Les valeurs les plus faibles (en dessous de 3 Å) sont observées pour le dimère *BN*, qui reste très stable tout au long des 10 ns de simulation. Bien que le dimère *décalé* possède une géométrie bien plus étendue (voir figure 4.8), sa structure reste également proche de sa structure initiale, avec des valeurs de RMSD oscillant autour de 4 Å. Le résultat le plus surprenant est obtenu pour le dimère *BE* qui subit un changement conformationnel important avec des valeurs de RMSD atteignant jusqu’à 12 Å.

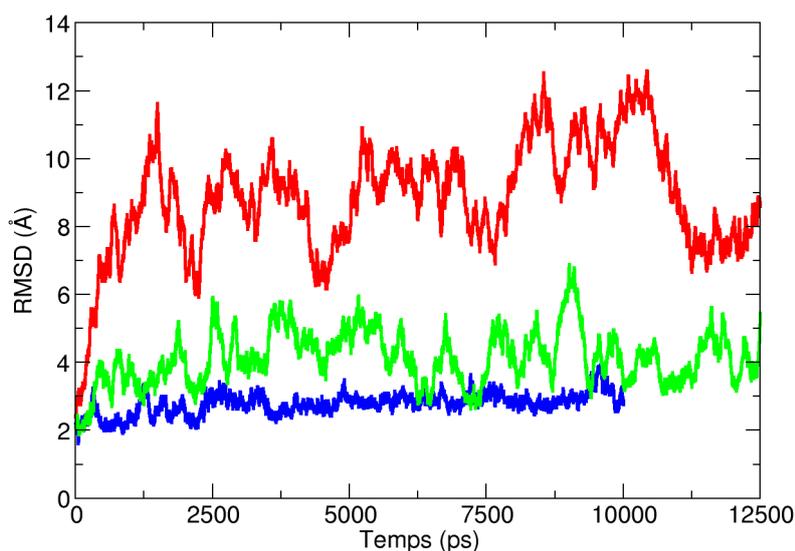


FIGURE 6.1 – Évolution du RMSD dans les simulations des dimères *BN* (en bleu), *BE* (en rouge), et *décalé* (en vert).

L'analyse de l'évolution de l'aire d'interface est une autre mesure de la stabilité d'un dimère. L'évolution des aires interfaciales pour les trois formes dimériques est représentée sur la figure 6.2. Les résultats montrent deux caractéristiques principales :

1. Les aires interfaciales des trois dimères sont très différentes. Tandis que le dimère *BN* présente une interface protéine-protéine typique d'environ 2400 \AA^2 (Chakrabarti et Janin, 2002 ; Bahadur *et al.*, 2004), les dimères *BE* et *décalé* présentent des interfaces relativement petites d'environ 1700 et 1200 \AA^2 respectivement.
2. Alors que l'interface du dimère *BN* reste stable pendant la durée entière de la dynamique, les deux autres interfaces subissent des modifications. En effet, l'aire de l'interface *BE* augmente d'environ 200 \AA^2 au cours des deux dernières nanosecondes de simulation et l'aire de l'interface du dimère *décalé* est diminuée d'environ 200 \AA^2 durant la même période.

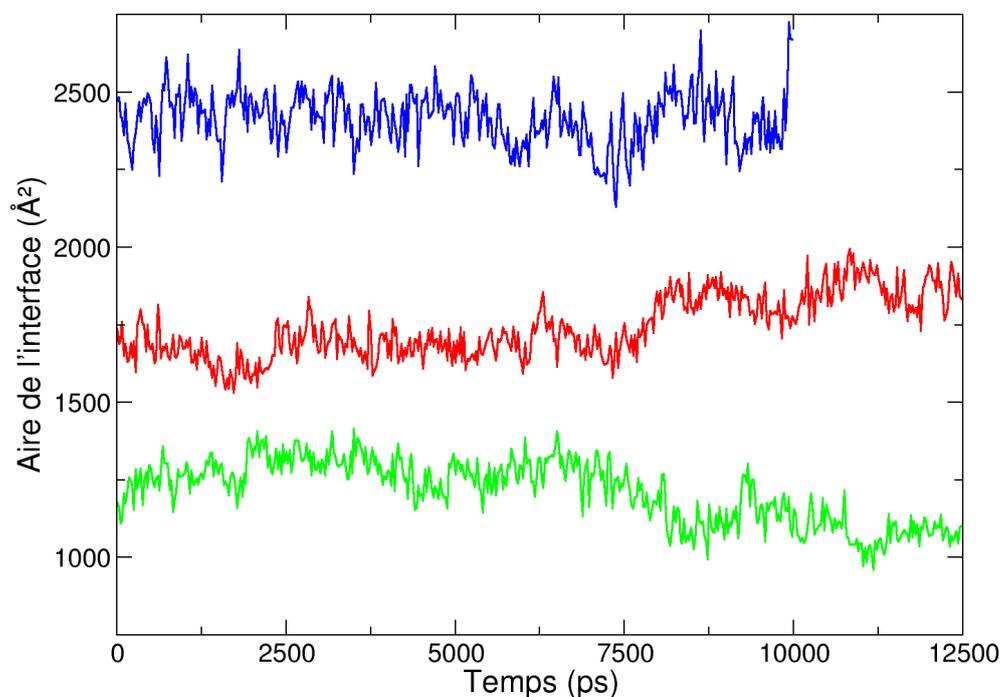


FIGURE 6.2 – Évolution des interfaces dans les simulations des dimères *BN* (en bleu), *BE* (en rouge), et *décalé* (en vert).

6.4 Changements conformationnels du dimère *BE*

Quelle est la nature du changement conformationnel qui intervient lors de la simulation du dimère *BE*? Au vu des résultats présentés dans le paragraphe précédent, il implique clairement l'interface dimérique, plutôt que la conformation de chacun des monomères. L'évolution générale est montrée sur la figure 6.3, qui compare la structure initiale avec la structure moyenne calculée sur les 2,5 dernières nanosecondes de la simulation. L'interface dimérique est significativement modifiée, bien que l'orientation relative des deux domaines EC1 ne soit que peu affectée. Une vue plus détaillée est montrée sur la figure 6.4. Le processus commence par la formation de nouveaux contacts entre les brins N-terminaux des domaines EC1 (impliqués dans l'échange de brins) et leur insertion dans leur propre domaine. Puis ceci est suivi par une nouvelle approche des brins N-terminaux pour créer un court feuillet β parallèle inter-chaînes.

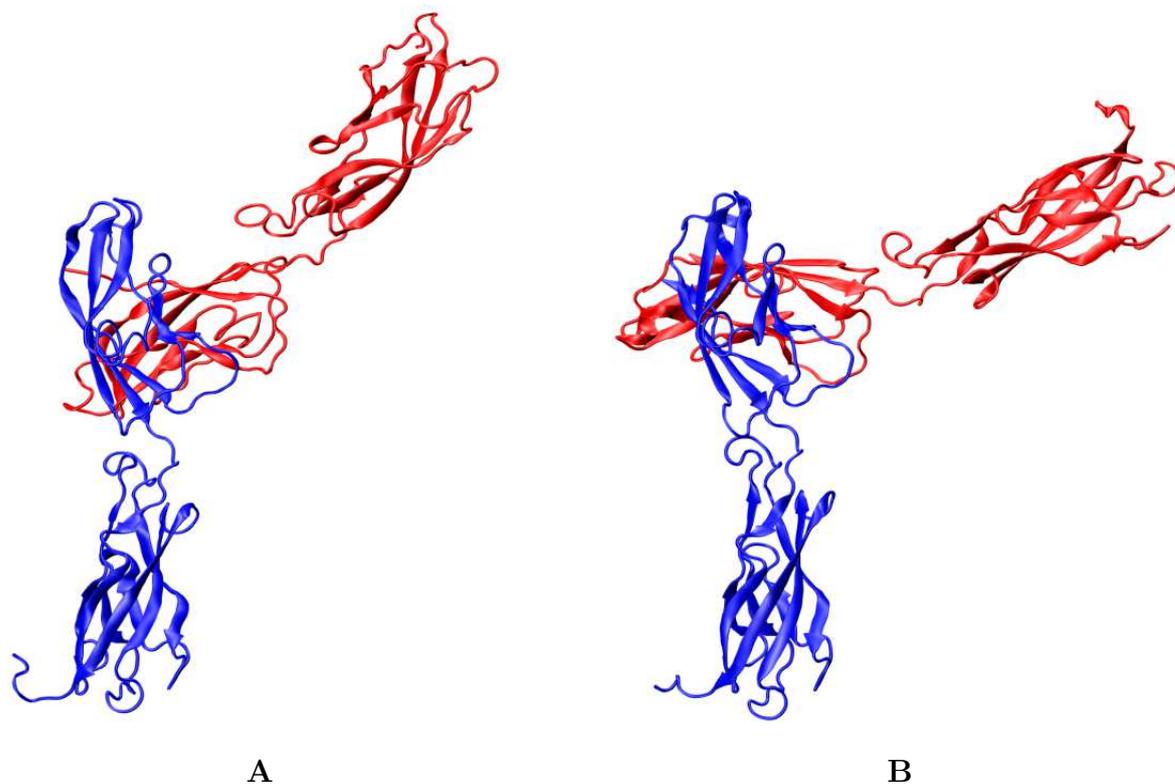


FIGURE 6.3 – Changement conformationnel pendant la simulation du dimère *BE*. **A** : structure cristallographique initiale (1Q1P); **B** : Structure moyenne calculée sur les dernières 2,5 ns de la simulation.

Il est intéressant de noter que la première phase de cette évolution, qui permet l'intégration de chaque brin N-terminal dans son propre tonneau β , ressemble à l'interface cristallographique du dimère *BE* de la C-cadhérine (Boggon *et al.*, 2002). Bien que les séquences de la E- et de la C-cadhérine soient très homologues (65 % pour le fragment EC1-2 et 70 % pour le seul domaine EC1), cette dernière présente une interface un peu plus grande, qui correspond à l'évolution observée au cours de la dynamique moléculaire pour la E-cadhérine. L'évolution du RMSD de l'interface par rapport aux interfaces cristallographiques des dimères *BE* de la E- et de la C-cadhérine est montrée figure 6.5. Celle-ci montre que la structure de la E-cadhérine s'éloigne rapidement de sa conformation initiale et approche celle de la C-cadhérine : ce réarrangement entraîne un RMSD de 5 Å par rapport au point de départ, tandis qu'après les 500 premières picosecondes, le RMSD par rapport à la C-cadhérine reste inférieur à 2,5 Å.

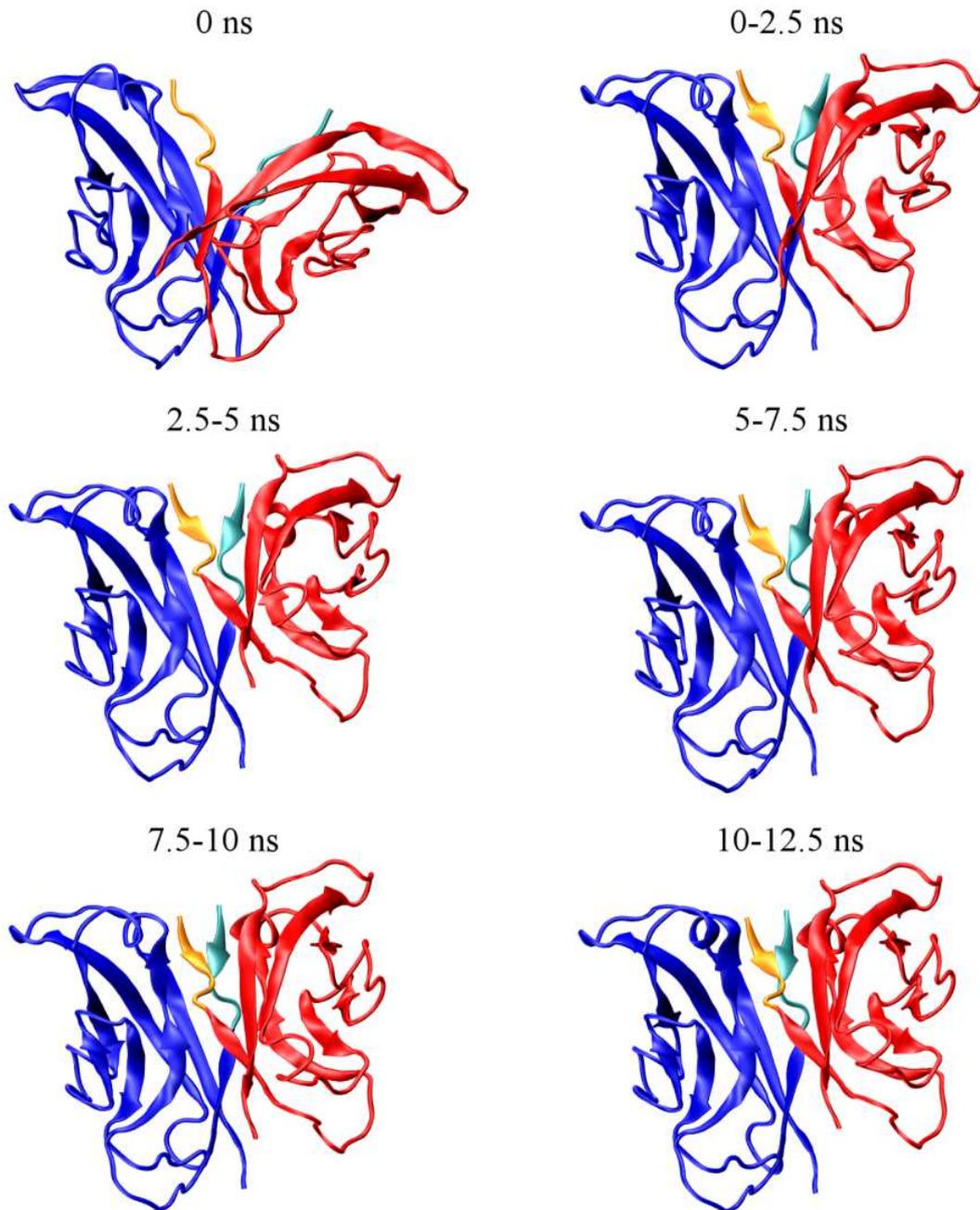


FIGURE 6.4 – Évolution de l'interface entre les domaines EC1 des monomères EC1-2 dans le dimère *BE*. Les six premiers résidus de chaque monomère, impliqués dans l'échange de brins, sont montrés en cyan pour le monomère bleu et en orange pour le monomère rouge. Chaque image, exceptée la structure à 0 ns, est moyennée sur des segments de 2,5 ns de la dynamique moléculaire. Les structures secondaires sont calculées avec l'option *Stride* du programme VMD (Humphrey *et al.*, 1996).

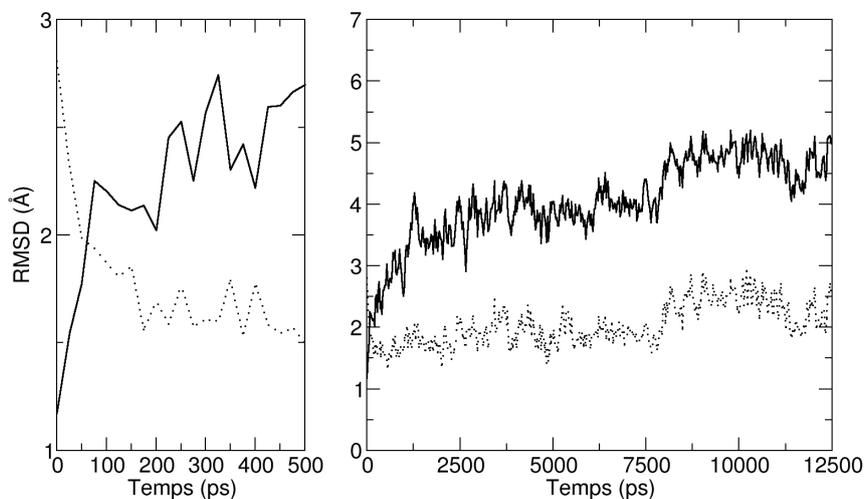


FIGURE 6.5 – RMSD calculé sur le squelette carboné pour les domaines EC1 interagissant au cours de la simulation du dimère *BE*. La ligne pleine montre les valeurs calculées par rapport à la structure cristallographique de la E-cadhérine et la ligne pointillée celles calculées par rapport au dimère *BE* de la C-cadhérine. Le tracé sur la gauche montre un agrandissement de la zone correspondant aux 500 premières picosecondes de la dynamique.

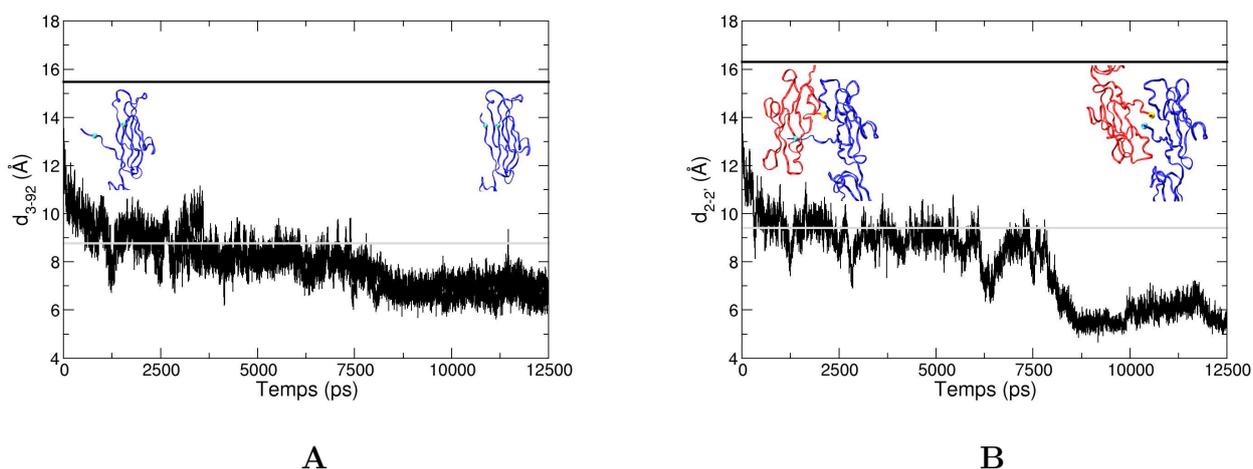


FIGURE 6.6 – Modifications de l'interface du dimère *BE* reflétées par les distances entre les résidus Val 3 et Met 92 (**A**) et entre les résidus Trp 2 des deux domaines (**B**). Les barres horizontales noires et grises montrent les distances correspondantes dans les structures cristallographiques de la E- et de la C-cadhérine. Les schémas montrent les instants initiaux et finaux de la simulation, les C_{α} des résidus entre lesquels sont calculées les distances sont représentés par des sphères cyan pour le monomère bleu et jaunes pour le monomère rouge.

La figure 6.6 montre une autre manière d’analyser ce changement conformationnel en suivant l’évolution de deux distances inter-résidus. La première, entre les résidus Val 3 et Met 92 d’un même domaine EC1 (notée d_{3-92}), reflète l’insertion du brin N-terminal dans le tonneau β de son propre domaine, tandis que la seconde, entre les résidus Trp 2 de chaque domaine (notée $d_{2-2'}$), reflète la formation du brin β inter-chaînes. Ces deux distances adoptent presque immédiatement des valeurs plus faibles que celles observées dans la structure cristallographique de la E-cadhérine. Au cours de la simulation, d_{3-92} et $d_{2-2'}$ restent ensuite proches des valeurs observées dans la C-cadhérine. Cependant, après environ 8 ns de simulation, $d_{2-2'}$ diminue encore, indiquant un renforcement du feuillet β entre les deux domaines EC1 par rapport à celui présent dans la structure de la C-cadhérine.

6.5 Enthalpies libres de dimérisation

Bien que nous ayons vu que les dimères *décalé* et *BE* subissent des changements conformationnels au cours des simulations de dynamique moléculaire, il n’est pas aisé d’en déduire leurs conséquences sur la stabilité thermodynamique de ces complexes. Nous avons pour cela estimé l’enthalpie libre de formation des trois dimères en utilisant l’approche MM-PBSA. Les résultats obtenus sont résumés dans le tableau 6.2. De cette analyse, il ressort que tous les dimères possèdent des enthalpies libres de formation négatives, donc favorables. On peut néanmoins mettre en avant deux constatations :

1. Le dimère *décalé* est de loin la conformation la moins stable, avec une enthalpie libre de formation inférieure de 20 à 30 kcal.mol⁻¹ aux deux autres dimères. Ceci est cohérent avec la faible aire d’interface enfouie dans ce complexe. Bien que son rôle dans l’adhésion ne puisse pas être complètement écarté, cette interface semble bien être nettement plus faible que les deux autres.
2. En dépit du profond changement de sa conformation, le dimère *BE* est plus stable que le dimère *BN*. Ces deux structures restent de bons candidats pour la formation de fortes interactions entre molécules de E-cadhérine.

Il est par ailleurs intéressant de noter que le terme dominant pour le dimère *BN*, qui possède la plus grande interface, correspond aux interactions de van der Waals, tandis que le dimère *BE*, dans lequel on observe le phénomène d’échange de brins est le seul à montrer une forte stabilisation électrostatique. De plus, comme attendu, la perte d’entropie due à la formation du dimère *BN* rigide est plus importante que celle impliquée dans la formation du dimère *BE*, cette grande différence étant cependant partiellement compensée par une plus forte contribution enthalpique favorable pour le dimère *BN*.

Contribution	Dimère <i>BN</i>	Dimère <i>BE</i>	Dimère <i>décalé</i>
ΔH_{int}	-15	-9	-15
ΔH_{vdW}	-103	-28	-50
ΔH_{coul}	482	289	397
ΔG_{solv}	-439	-325	-374
ΔG_{elec}	43	-36	23
ΔG_{surf}	-9	-1	-3
$-T\Delta S$	55	34	34
ΔG	-29	-41	-11

TABLE 6.2 – Enthalpies libres de formation des dimères et les différentes contributions énergétiques calculées à l’aide de la méthode MM-PBSA. Toutes les valeurs sont données en kcal.mol⁻¹.

6.6 Interfaces et spécificité

Il est enfin intéressant de considérer ce que les simulations des différents dimères de E-cadhérine peuvent apporter pour tenter d’expliquer la spécificité des interactions entre cadhérines. Afin d’analyser le rôle que peuvent jouer les différences dans les séquences des C-, E-, et N-cadhérines (voir figure 6.7), nous avons tout d’abord identifié les résidus qui contribuent significativement aux différentes interfaces dans les simulations. Le nombre de résidus formant l’interface n’est pas relié à la stabilité : on peut en effet remarquer que les dimères *décalé* et *BE* sont similaires de ce point de vue.

En supposant que la C- et la N-cadhérine adoptent des conformations identiques à celle de la E-cadhérine, on peut alors comparer la nature des acides aminés importants pour la formation des interfaces dans chacun des cas. Les résultats sont montrés dans le tableau 6.3. Dans ce tableau, les cellules grisées indiquent les résidus contribuant de manière significative à l’interface et qui diffèrent dans au moins deux des trois cadhérines comparées. On observe que de ce point de vue, les dimères *BN* et *décalé* font tous deux intervenir des résidus non conservés. Au contraire, les résidus enfouis dans l’interface *BE* sont majoritairement conservés chez les cadhérines classiques. Il peut même sembler étonnant que, bien que trois résidus diffèrent pour cette interface entre la E- et la N-cadhérine (parmi lesquels le résidu 25 contribue de façon importante à l’interface), aucune différence n’est observée entre la E- et la C-cadhérine. Ces observations sont cependant en accord avec les expériences qui montrent la formation d’agrégats spécifiques lorsque l’on mélange des cellules

exprimant la E-cadhérine à d'autres exprimant la N-cadhérine (Patel *et al.*, 2003) mais pas de discrimination dans le cas où sont exprimées les formes E- et C- de la cadhérine (Niessen et Gumbiner, 2002). Ceci est également en adéquation avec nos résultats de dynamique moléculaire qui montrent que le dimère *BE* de la E-cadhérine évolue pour ressembler à celui de la C-cadhérine.

	10	20	30	40	50	60
E-Cadhérine	DWVIPPISCPENKGEFPK	NLVQIKSNRDKETKVFYSITGQGADKPPVGVFIIERETGWL				
C-Cadhérine	DWVIPPISKVSENERGPF	PKRLVQIKSNKDRFNKVYYSITGQGADNPPQGVFRIEWETGWM				
N-Cadhérine	DWVIPPINLPENSRGPF	QELVRIRSDRDKNLSLRYSVTGPGADQPPTGIFIIINPISGQL				
	70	80	90	100	110	120
E-Cadhérine	KVTQPLDREAI	AKYILYSHAVSSNGEAVEDPMEIVITVTDQNDNRPEFTQEVFEGSVAEG				
C-Cadhérine	LVTRPLDREYDKYVLS	SHAVSENGSPVEEPMETINVIDQNDNRPKFTQDVFRGVSREG				
N-Cadhérine	SVTKPLDRELIARFHL	RAHAVDINGNQVENPIDIVINVIDMNDNRPEFLHQVWNGSVPEG				
	130	140	150	160	170	180
E-Cadhérine	AVPGTSMKVSATD	ADDDVNTYNAAIAYTIVSQDPELPHKNMFTVNRDTGVISVLTSGLD				
C-Cadhérine	VQPGTQVMAVSAT	DEDDNIDSLNGVLSYSILKQDPEEIPNLFTINRETGVISLIGTGLD				
N-Cadhérine	SKPGTYVMTVTAID	ADD-PNALNGMLRYRIVSQAPSTPSPNMFTINNETGDIITVAAGLD				
	190	200	210	220		
E-Cadhérine	RESYPTYTLV	QAADLQGE---GLSTTAKAVITVKDINDNA				
C-Cadhérine	REKFPETYTLV	QATDLEGA---GLSVEGKAIIQITDANDNA				
N-Cadhérine	REKVQYTLI	QATDMEGNPTYGLSNTATAVITVTD-----				

FIGURE 6.7 – Alignement des séquences primaires des fragments EC1-2 de E-, C-, et N-cadhérines. La couleur rouge indique une conservation du résidu dans les trois cadhérines, la couleur bleu indique que deux des cadhérines ont un résidu commun, et la couleur noire indique les résidus différents dans les trois cadhérines. L'alignement a été obtenu avec le programme *ClustalW* disponible sur le serveur NPSA de l'IBCP (<http://npsa-pbil.ibcp.fr/>) en utilisant des paramètres standards.

L'analyse menée ici ne concerne que l'étude des surfaces enfouies sans tenir compte de facteurs énergétiques, et ne tient pas compte également d'éventuels changements structuraux induits par les résidus extérieurs à la surface de contact chez les autres cadhérines. Néanmoins, elle soutient l'idée que la spécificité des interactions entre cadhérines serait reliée à de petites différences thermodynamiques (Chen *et al.*, 2005) et pourrait impliquer des facteurs cinétiques (Perret *et al.*, 2004 ; Bayas *et al.*, 2006).

	<i>BN</i>					<i>BE</i>					<i>décalé</i>				
	Résidu	BSA	E	C	N	Résidu	BSA	E	C	N	Résidu	BSA	E	C	N
M o n o m è r e A	14	156	K	R	R	2	261	W	W	W	175	79	L	I	V
	142	84	Y	L	L	25	90	K	K	R	177	63	S	T	A
	101	81	Q	Q	M	5	90	P	P	P	123	61	P	P	P
	140	74	N	D	N	3	72	V	V	V	124	54	G	G	G
	5	72	P	P	P	26	33	S	S	S	166	41	N	N	N
	3	66	V	V	V	92	31	M	M	I	176	36	T	G	A
	196	55	L	L	M						164	35	T	T	T
	10	51	P	S	P						173	17	S	S	I
	99	43	T	I	I										
	102	40	N	N	N										
	141	35	T	S	A										
	13	28	E	E	S										
	200	24	G	G	G										
	22	20	V	V	V										
	7	15	I	I	I										
143	13	N	N	N											
M o n o m è r e B	14	151	K	R	R	2	254	W	W	W	84	91	N	N	N
	140	91	N	D	N	1	112	D	D	D	55	73	R	W	P
	101	88	Q	Q	M	25	99	K	K	R	85	59	G	G	G
	5	71	P	P	P	5	86	P	P	P	79	58	H	H	H
	201	64	L	L	L	3	71	V	V	V	81	53	V	V	V
	99	40	T	I	I	89	50	E	E	E	83	48	S	E	I
	102	34	N	N	N	23	46	Q	Q	R	37	46	S	S	S
	141	34	T	S	A	92	35	M	M	I	35	44	F	Y	R
	13	34	E	E	S										
	10	31	P	S	P										
	7	30	I	I	I										
	22	26	V	V	V										
	200	24	G	G	G										
	143	16	N	N	N										

TABLE 6.3 – Surface enfouie (BSA) en Å² des résidus appartenant aux monomères formant les dimères *BN*, *BE*, et *décalé* de E-cadhérine, calculée en moyenne sur les 5 dernières nanosecondes des simulations de dynamique moléculaire. Seuls les résidus enfouissant plus de 50 % de leur surface accessible dans le monomère et contribuant à hauteur d’au moins 10 Å² à l’aire de l’interface sont inclus et listés dans l’ordre décroissant de leur BSA. Les résidus équivalents dans la C- et la N-cadhérine sont obtenus à partir d’un alignement multiple réalisé avec le programme *ClustalW*. Les cellules grises indiquent les résidus qui diffèrent dans au moins deux des cadhérines.

6.7 Conclusion

Nous avons utilisé la dynamique moléculaire pour examiner les propriétés mécaniques et thermodynamiques des trois dimères de EC1-2 observés par cristallographie. Dans tous les cas, nous avons pu observer que la dimérisation (quelle que soit l'interface) n'influe pas sur la dynamique des monomères de EC1-2, qui gardent leur conformation initiale en "bâtonnet".

Les différentes interfaces ne présentent toutefois pas la même rigidité. En fait, seul le dimère *BN* conserve sa structure initiale tout au long de la trajectoire. Le dimère *BE* subit un changement conformationnel important qui entraîne un renforcement de l'interface et une conformation proche de celle du dimère *BE* de la C-cadhérine. Enfin, dans le cas du dimère *décalé*, l'aire de l'interface, initialement faible, diminue encore durant la simulation.

Les estimations d'enthalpie libre de formation montrent par ailleurs que les dimères *BN* et *BE* sont bien plus stables que le dimère *décalé*. Ceci conforte l'idée que l'interface *décalée* serait un simple contact cristallin plutôt qu'une interface fonctionnelle. Il convient pourtant de rester prudent car l'ensemble des interactions entre cadhérines est faible (Alattia *et al.*, 1997 ; Koch *et al.*, 1997 ; Patel *et al.*, 2003) et l'adhésion repose sur la formation de multiples contacts (Koch *et al.*, 2004).

Pour finir, une analyse des résidus participant aux interfaces dimériques suggère que les dimères *BN* et *décalé* peuvent raisonnablement expliquer la spécificité des interactions entre domaines EC1, en permettant de discriminer les C-, E-, et N-cadhérines. En revanche, l'étude menée sur l'interface du dimère *BE* semble indiquer que cette interface ne permet pas de distinguer les C- et E-cadhérines et seulement faiblement la N-cadhérine. Ceci suggère que dans le cas où ce dimère *BE* serait impliqué dans la spécificité des interactions, celle-ci reposerait sur de subtils facteurs thermodynamiques ou cinétiques.

Troisième partie

Étude des propriétés des grands systèmes

Chapitre 7

Définition d'une représentation multi-échelle

Sommaire

7.1	Limitation des représentations actuelles des protéines	119
7.2	Principe d'une représentation multi-échelle	121
7.2.1	Exemple des méthodes QM/MM	121
7.2.2	Principe de la représentation "tout-atome"/"gros grains"	122
7.3	Mise en œuvre	124
7.3.1	Définition de la représentation multi-échelle MM/CG	124
7.3.2	Intégration dans le programme AMBER	125
7.4	Étude d'une petite protéine globulaire	127
7.4.1	Étude dans le vide	127
7.4.2	Incorporation du solvant dans le modèle	132
7.5	Étude d'un homodimère	138
7.5.1	Choix du système	138
7.5.2	Dynamique globale du système	139
7.5.3	Fluctuations atomiques dans la zone MM	142
7.6	Bilan et perspectives	143

7.1 Limitation des représentations actuelles des protéines

La variété des systèmes biologiques rend leur modélisation particulièrement délicate. En effet, les échelles de temps et de taille que recouvrent les différents phénomènes biologiques sont très étendues, allant de la femtoseconde (pour les vibrations de liaison) à plusieurs

heures (pour le repliement protéique) et de quelques angströms (pour de petites molécules) à quelques centimètres (pour les fibres de chromatine¹).

La chimie quantique est la méthode de choix pour étudier un système au niveau atomique. Pourtant, l'utilisation de cette méthode est très limitée par la taille du système et confinée à des molécules possédant au maximum quelques dizaines voire centaines d'atomes. On lui préfère donc, pour l'étude des biomolécules, les méthodes de modélisation moléculaire basées sur une représentation classique des particules. Tout particulièrement, il est attirant d'utiliser une représentation "tout-atome" (voir annexe B.2) pour conserver le plus d'information possible. C'est cette représentation que nous avons utilisée dans la majeure partie de notre travail pour l'étude des cadhérines.

Mais la représentation "tout-atome" est elle aussi limitée à l'étude de systèmes comportant quelques milliers à centaines de milliers d'atomes et ne permet pas une application systématique des méthodologies sur un grand nombre de systèmes. Il convient alors d'utiliser une représentation plus simple, comme celle que nous avons utilisée au cours de notre étude sur la mesure de la flexibilité des protéines dans la partie I de ce manuscrit. Cette méthodologie "gros grains" nous a permis d'automatiser le calcul des constantes de force pour les résidus d'une protéine en diminuant le temps de calcul d'un facteur 4000 environ, tout en gardant un profil de constantes de force très similaire à celui obtenu en représentation "tout-atome".

Comme nous l'avons vu, l'utilisation de modèles simplifiés permet donc de reproduire de façon très correcte les résultats obtenus avec une représentation "tout-atome". Néanmoins, il existe une perte d'information non négligeable qui rend ces modèles simplifiés inaptes à l'étude de phénomènes tels que l'effet de mutations ponctuelles ou la reconnaissance moléculaire qui nécessitent généralement une représentation au niveau atomique. Quand de tels problèmes sont posés au modélisateur, celui-ci est obligé d'avoir recours de nouveau au modèle "tout-atome" impliquant des temps de calcul importants.

1. Les fibres de chromatine sont des structures constituées d'ADN enroulé autour d'histones et qui composent les chromosomes.

7.2 Principe d'une représentation multi-échelle

7.2.1 Exemple des méthodes QM/MM

L'une des limitations majeures des représentations utilisées en modélisation moléculaire classique est de ne pas tenir compte des effets électroniques, et donc de n'avoir pas accès à la réactivité chimique des molécules (pas de formation ou de rupture de liaison). Lorsqu'on souhaite s'intéresser à cette réactivité, par exemple au niveau du site actif d'une enzyme, une méthode possible est d'extraire une partie réduite (quelques dizaines d'atomes au maximum) de la structure correspondant à ce site actif et d'appliquer à ce sous-système les méthodes de la chimie quantique. Le problème de ce procédé est qu'il ignore complètement l'effet de l'environnement protéique sur la dynamique des interactions au niveau du site actif.

Le principe des méthodes QM/MM (pour "Quantum Mechanics"/"Molecular Mechanics"), proposées pour la première fois par Warshel et Levitt, est de combiner les représentations quantiques et classiques au sein d'une représentation multi-échelle (Warshel et Levitt, 1976). Dans l'exemple d'un site actif enzymatique, cela revient à considérer les atomes composant le site actif à l'échelle quantique et le reste de la protéine à l'échelle classique. L'énergie du système ainsi construit peut être décomposée en trois termes :

- un terme purement quantique E_{QM} , correspondant aux interactions entre les particules quantiques.
- un terme purement classique E_{MM} , correspondant aux interactions entre les particules classiques.
- un terme de couplage quantique/classique $E_{QM/MM}$, correspondant aux interactions croisées entre les particules classiques et les particules quantiques.

Le cœur du problème dans ce type de représentation est de coupler les zones quantiques et classiques (terme $E_{QM/MM}$), ce qui nécessite une calibration minutieuse. Un autre problème réside dans la définition de la zone quantique : si celle-ci est trop grande, la rapidité des calculs sera affectée ; d'un autre côté, si elle est trop petite, les propriétés électroniques de la partie quantique peuvent être affectées de façon notable par rapport à la situation réelle. Enfin, afin de satisfaire des considérations de valence complète pour la partie quantique, il est également nécessaire d'introduire des atomes fictifs vus uniquement par la partie QM. La représentation QM/MM est fréquemment utilisée en combinaison avec la dynamique moléculaire (Carloni *et al.*, 2002) ou la minimisation d'énergie (Dinner *et al.*, 2001).

7.2.2 Principe de la représentation "tout-atome" / "gros grains"

Les méthodes QM/MM répondent au besoin d'une représentation des phénomènes chimiques comme les formations ou les ruptures de liaison pour des structures de taille relativement réduite. Même dans ces conditions, le coût des simulations limite les temps simulés en dynamique moléculaire à quelques dizaines de picosecondes. Dans notre cas, nous ne nous intéressons pas aux propriétés chimiques des protéines, mais à leurs propriétés mécaniques, et les problèmes que nous rencontrons sont plutôt dus à la taille des systèmes étudiés, qui limite les études réalisées à résolution atomique. Si les méthodes QM/MM se situent à l'interface entre les représentations quantique et "tout-atome", notre problématique se situe ainsi à l'interface entre les représentations "tout-atome" et "gros grains". Nous souhaitons donc définir une représentation multi-échelle couplant ces deux dernières représentations au sein d'un même modèle.

Une telle représentation multi-échelle nous semble particulièrement adaptée à l'étude des sites actifs enzymatiques ou des surfaces d'interaction dans les complexes protéiques (voir figure 7.1). En effet, dans les deux cas, la zone d'intérêt ne représente qu'une petite fraction du système, et il devrait être possible de diminuer les temps de calcul en simplifiant la représentation du reste de la protéine.

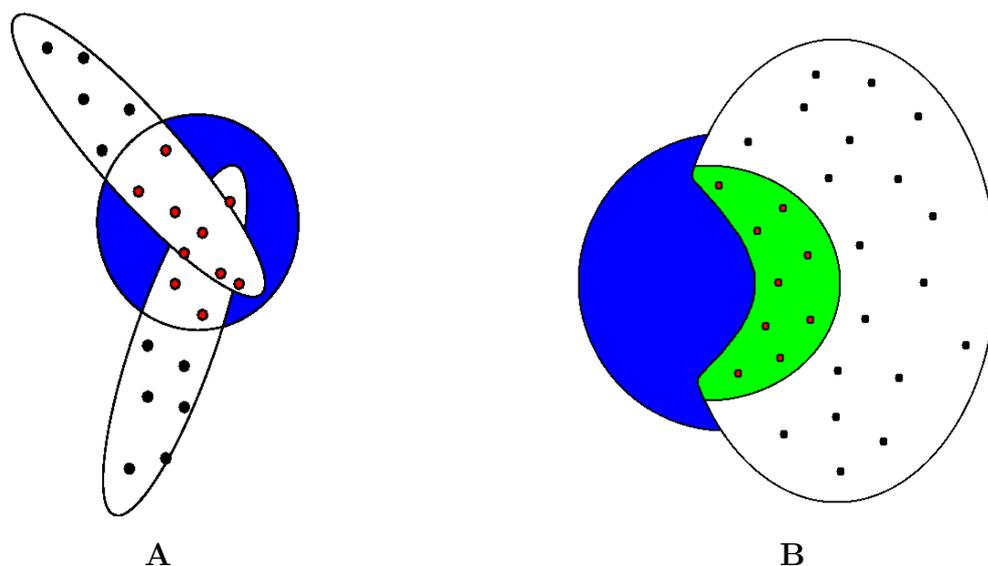


FIGURE 7.1 – Applications possibles du modèle MM/CG dans le cadre de l'étude d'une surface d'interaction protéique (A) ou d'un site actif (B). Les résidus sont représentés par des points noirs (résidus "gros grains") ou rouges (résidus "tout-atome"). Le cercle bleu représente la sphère de solvation nécessaire pour la région "tout-atome".

Au moment où nous avons débuté nos recherches, nous n'avions connaissance que d'une seule étude présentant une représentation multi-échelle impliquant une représentation atomique couplée à une représentation plus grossière (Villa *et al.*, 2005). Cette représentation est appliquée à la modélisation du complexe entre le répresseur LAC et un double brin d'ADN. Tandis que la protéine est représentée classiquement à l'échelle atomique, l'ADN est modélisé à la fois avec une résolution atomique au niveau de la zone de contact avec la protéine, et avec un modèle continu (sous forme de bâtonnet élastique) pour le reste de la molécule.

Pour notre part, nous avons choisi de coupler la représentation atomique avec un modèle granulaire dans lequel les résidus sont reliés par un réseau élastique. Plusieurs éléments ont conditionné notre choix. Tout d'abord, la simplicité des réseaux élastiques rend ce modèle aisé à intégrer dans les programmes existant. De plus, cette représentation est capable de reproduire de façon très raisonnable les mouvements globaux. L'utilisation d'un réseau de ressorts permet également de stabiliser la structure de la partie "gros grains" de telle sorte qu'aucune solvatation n'est nécessaire pour cette partie du système. Enfin, notre groupe possédait déjà une certaine expérience sur les réseaux élastiques, comme indiqué chapitre 3 et dans (Navizet *et al.*, 2004a ; Navizet *et al.*, 2004b).

Dans un souci de lisibilité, ce type de modèle sera abrégé MM/CG (pour "*Molecular Mechanics*" / "*Coarse-Grained*"), par analogie avec la notation QM/MM. Par la suite, les résidus décrits à l'échelle atomique seront notés MM, tandis que les résidus "gros grains" seront notés CG.

Récemment, Neri et collaborateurs ont présenté un modèle similaire à celui que nous mettons en place (Neri *et al.*, 2005). Les différences avec notre représentation tiennent principalement dans le traitement des interactions impliquant les résidus CG ainsi que dans le traitement du solvant. Un potentiel de Go simplifié est en effet utilisé pour les interactions impliquant les résidus CG : dans ce modèle, l'interaction est un potentiel quadratique entre deux résidus CG et un potentiel de Morse autrement. L'effet du solvant, quant à lui, est modélisé par l'introduction d'un terme de viscosité dans les équations du mouvement.

7.3 Mise en œuvre

7.3.1 Définition de la représentation multi-échelle MM/CG

Par analogie avec une méthode QM/MM, trois zones peuvent être distinguées dans la structure protéique :

- une zone MM, de résolution atomique, où toutes les interactions sont gouvernées par un champ de force classique (dans notre cas, nous avons utilisé un champ de force "tout-atome" AMBER).
- une zone CG, dans laquelle chaque résidu est représenté par un site (positionné au niveau du C_α du résidu) et où les interactions sont modélisées par un réseau élastique.
- une zone intermédiaire dans laquelle les résidus représentés à l'échelle atomique sont également en interaction avec des résidus CG.

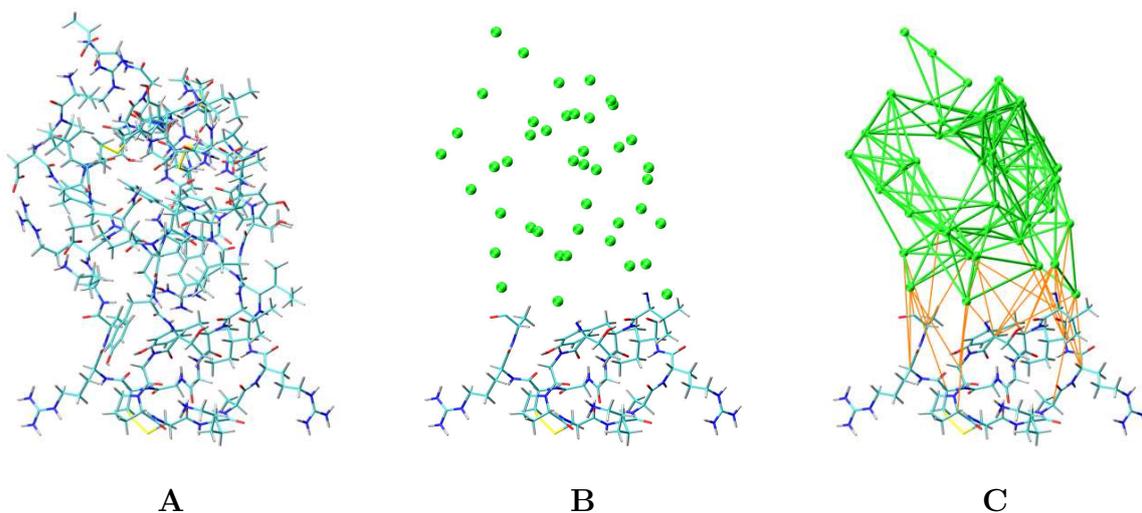


FIGURE 7.2 – Différentes étapes de la construction d'un modèle MM/CG sur l'exemple d'une protéine globulaire. **A** : Représentation "tout-atome" ; **B** : Sélection des résidus CG ; **C** : Création du réseau élastique.

La figure 7.2 montre les différentes étapes de la construction d'un modèle MM/CG. À partir d'une représentation "tout-atome" de la protéine (A), on sélectionne la zone CG (résidus représentés par des sphères vertes) et la zone MM (B). Ensuite, le réseau de ressorts est créé (C). Les traits verts représentent les ressorts créés entre deux résidus CG et les traits orange ceux qui impliquent un résidu CG et un résidu MM (ces résidus sont donc ceux qui constituent la zone intermédiaire définie ci-dessus). La présence d'un ressort entre deux résidus est soumise à une condition de distance entre les deux atomes qui doit être inférieure à une valeur seuil.

Dans le cadre de ce modèle, l'énergie potentielle E_p du système est donnée par :

$$E_p = E_{MM} + E_{MM/CG} + E_{CG}$$

Les différents termes de cette énergie sont :

- l'énergie potentielle E_{MM} de la partie "tout-atome", dont l'expression est celle du champ de force AMBER (voir annexe B.2).
- l'énergie potentielle E_{CG} de la région CG, dont l'expression est celle d'un réseau élastique :

$$E_{CG} = \sum_{i < j} K_{CG} (r_{ij} - r_{ij}^0)^2$$

où K_{CG} est la constante de force des ressorts dans la zone CG, r_{ij} la distance entre les résidus i et j , et r_{ij}^0 cette même distance dans la structure initiale. Il est à noter que la valeur de la constante K_{CG} est indépendante du couple $(i ; j)$.

- le terme de couplage $E_{MM/CG}$ entre les résidus de la zone intermédiaire et les résidus CG, donné par :

$$E_{MM/CG} = \sum_{i < j} K_{MM/CG} (r_{ij} - r_{ij}^0)^2$$

où $K_{MM/CG}$ représente la constante de force des ressorts "inter-zones".

7.3.2 Intégration dans le programme AMBER

Afin de tester notre nouvelle représentation, nous avons voulu réaliser des simulations de dynamique moléculaire. Nous avons donc intégré cette nouvelle représentation dans le logiciel AMBER. Le choix du potentiel d'interaction harmonique pour les interactions CG offre deux options :

1. une option "externe", consistant à appliquer des contraintes harmoniques sur les distances séparant deux atomes liés par un ressort dans le cadre des interactions CG ou MM/CG.
2. une option "interne", consistant à intégrer ces interactions dans le terme d'énergie de liaisons du champ de force.

Les deux options ont été implémentées et utilisées dans les calculs présentés ici.

7.3.2.1 Définition des résidus CGX

Vingt nouveaux types de résidus ont été créés pour remplacer les vingt types d'acides aminés dans la zone CG. Ces résidus, notés CGX où X représente le code à une lettre de

l'acide aminé remplacé (voir annexe A), sont constitués d'un seul pseudo-atome dont la masse est égale à la masse du résidu X. Par la suite, CGX désignera indifféremment le résidu ou le pseudo-atome le constituant en fonction du contexte.

En raison de l'utilisation d'un réseau de ressorts pour représenter les interactions impliquant les résidus CGX, nous avons annulé toutes les interactions non liées impliquant ces résidus. En d'autres termes, une charge nulle est attribuée à chacun des CGX (annulant la partie électrostatique) et les coefficients des interactions de Lennard-Jones sont fixés à 0 pour chaque paire d'atomes de la protéine impliquant au moins un atome CGX. En revanche, nous avons choisi de conserver les interactions répulsives de Lennard-Jones entre les CGX et les molécules d'eau. Nous reviendrons sur ce point dans la section 7.4.2.

Dans un premier temps, nous n'avons pas modifié les listes d'exclusion pour les interactions non liées¹ impliquant les CGX. Celles-ci, bien que nulles en raison de la définition des coefficients explicités ci-dessus, sont donc calculées. Il sera donc possible d'améliorer l'efficacité du modèle en modifiant les listes d'exclusion.

7.3.2.2 Description des paramètres du modèle

Cinq paramètres doivent être calibrés afin de reproduire les grandeurs observées dans des simulations "tout-atome" :

- la constante de force K_{CG} des interactions entre résidus CGX.
- la constante de force $K_{CG/MM}$ des interactions entre résidus CGX et "tout-atome".
- la distance maximale r_{CG} entre deux résidus CG reliés par un ressort.
- la distance maximale $r_{MM/CG}$ entre un résidu CG et un résidu "tout-atome" reliés par un ressort
- le coefficient de l'interaction répulsive $A_{CG/W}$ entre les résidus CGX et les atomes d'oxygène des molécules d'eau.

Dans un souci de simplification, nous avons choisi dans un premier temps de ne pas différencier les ressorts ajoutés au système, c'est-à-dire que nous avons fixé $K_{CG/MM} = K_{CG}$ et $r_{CG/MM} = r_{CG}$. Ceci permet de ramener le nombre de paramètres du modèle à trois. La justification de cette approximation provient de notre expérience sur les modèles de réseau élastique dans lesquels la distinction de différents types de ressorts n'apporte que

1. La liste d'exclusion contient l'ensemble des paires d'atomes pour lesquelles les interactions non liées ne sont pas calculées. Dans les cas standards, elles contiennent l'ensemble des interactions entre atomes séparés par une ou deux liaisons.

peu d'information supplémentaire. Néanmoins, l'effet d'une telle distinction dans notre représentation multi-échelle pourrait être différent et il sera peut-être nécessaire de revenir par la suite sur l'hypothèse d'équivalence des ressorts.

7.4 Étude d'une petite protéine globulaire

Afin de paramétrer le modèle, nous avons décidé de tester notre représentation MM/CG sur l'inhibiteur de la trypsine pancréatique bovine (abréviée par la suite BPTI pour "*bovine pancreatic trypsin inhibitor*"). Cette protéase de 58 résidus pour environ 900 atomes est très utilisée pour tester de nouvelles méthodologies car c'est une protéine globulaire de petite taille et qui possède à la fois des hélices α et des feuillets β .

7.4.1 Étude dans le vide

Les premiers tests ont été menés dans le vide. En dépit des limitations physiques de telles simulations, elles sont très rapides (pour le BPTI en représentation "tout-atome", 1 ns de dynamique moléculaire est obtenue en 10 heures environ) et permettent ainsi de tester rapidement de nombreux paramètres.

7.4.1.1 Protocoles

Nous avons réalisé une série de simulations du BPTI en représentation multi-échelle en faisant varier les paramètres K_{CG} et r_{CG} , et cela pour deux choix de la zone MM différents. La zone MM a été définie comme l'ensemble des résidus dont le C_α se situe à l'intérieur d'une sphère de rayon 10 Å autour du résidu 37 dans un cas puis du résidu 43 dans l'autre cas. Ceci donne lieu à la création de deux zones aux propriétés assez différentes (voir figure 7.3). Dans la première (figure B), il n'existe qu'une seule interface entre la zone MM et la zone CG, tandis que dans la seconde (C), deux zones CG distinctes sont visibles qui possèdent toutes deux des liaisons avec la zone MM.

L'ensemble des simulations MM/CG effectuées est listé dans le tableau 7.1. Elles ont été comparées avec une simulation du BPTI réalisée avec une représentation "tout-atome". Les comparaisons ont porté sur l'évolution du RMSD et sur les fluctuations atomiques calculées sur les 300 dernières picosecondes.

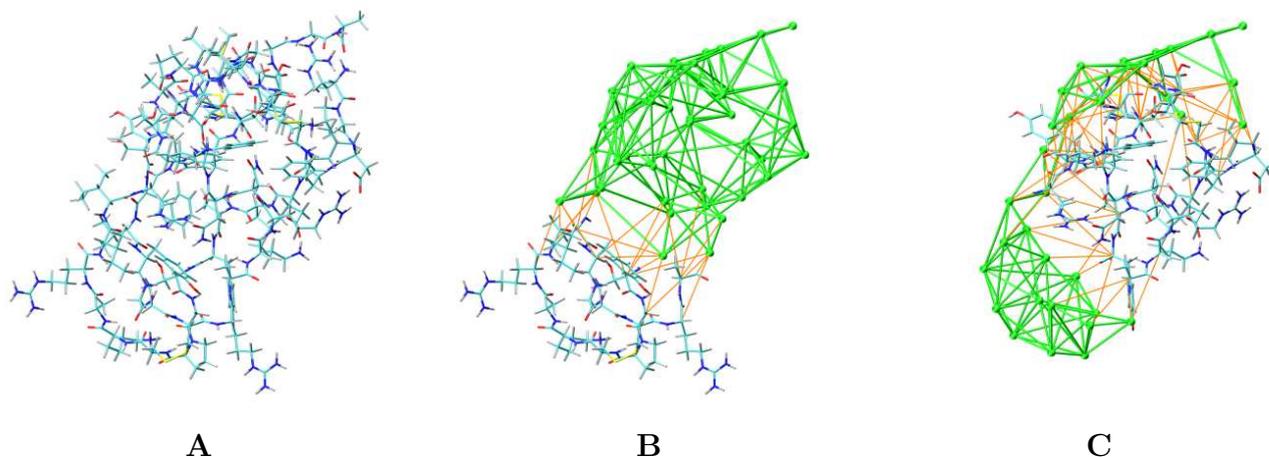


FIGURE 7.3 – Différentes représentations du BPTI étudiées. **A** : Représentation "tout-atome"; **B** : Représentation MM/CG avec la zone MM centrée autour du résidu 37; **C** : Représentation MM/CG avec la zone MM centrée autour du résidu 43.

Toutes les simulations ont été conduites pendant 1 ns à 300 K, la température étant maintenue constante par la méthode du thermostat de Langevin¹. Le pas d'intégration a été fixé à 1 fs et aucune troncature n'a été utilisée pour le calcul des interactions non liées. La structure du BPTI a été extraite du fichier 4PTI de la PDB (Marquart *et al.*, 1983) et soumise à 5000 pas de minimisation dans le vide afin de relâcher les contraintes stériques (conduisant à un RMSD de 0,7 Å par rapport à la structure cristallographique). Cette structure minimisée a servi ensuite de point de départ pour toutes les simulations (après passage ou non à la représentation MM/CG).

7.4.1.2 Résultats

Rôle de la constante de force K_{CG} des ressorts

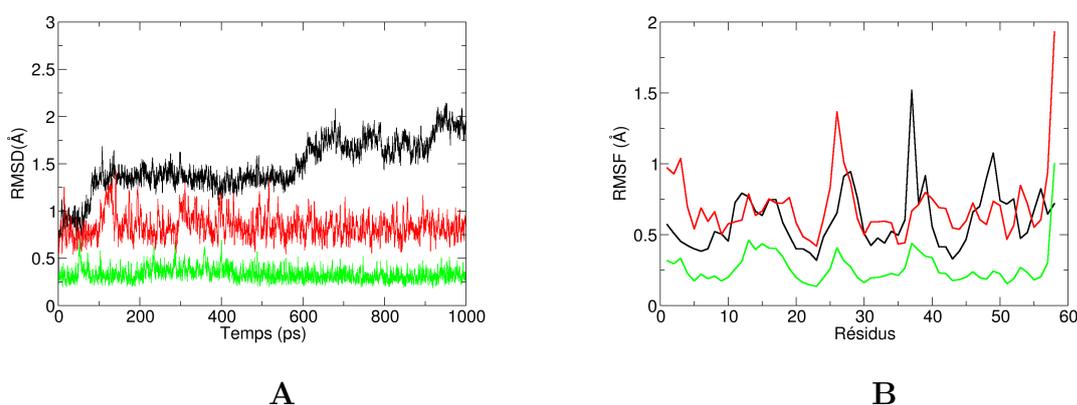
La figure 7.4 montre l'évolution du RMSD et les fluctuations atomiques globales calculés sur l'ensemble des C_α de la structure protéique pour la simulation "tout-atome" (en noir), ainsi que pour deux simulations MM/CG dans lesquelles nous avons fait varier la constante de force K_{CG} des ressorts du réseau élastique (simulations S37K1R8 et S37K10R8 du tableau 7.1).

1. Cette méthode consiste à appliquer l'équation de Langevin à la place de l'équation de Newton du mouvement. L'équation de Langevin contient une force frictionnelle proportionnelle et une force aléatoire qui permettent de maintenir la température du système.

Nom	centre de la zone MM	Résidus de la zone MM	K_{CG} (kcal.mol. \AA^{-2})	r_{CG} (\AA)
S37K1R8	37	11 à 18; 34 à 40	1	8
S37K1R9				9
S37K1R10				10
S37K10R8			10	8
S37K10R9				9
S37K10R10				10
S43K1R8	43	3 à 10; 20 à 24; 30; 40 à 46; 51; 54 à 55	1	8
S43K1R9				9
S43K1R10				10
S43K10R8			10	8
S43K10R9				9
S43K10R10				10

TABLE 7.1 – Paramètres des simulations MM/CG réalisées sur le BPTI dans le vide.

On remarque que l'utilisation d'un réseau de ressort restreint les possibilités de mouvement de la protéine. En effet, les RMSD moyens des simulations MM/CG (0,4 et 0,8 \AA) sont inférieurs à celui de la simulation "tout-atome" (qui augmente jusqu'à environ 1,75 \AA en fin de simulation). Ceci est d'autant plus vrai que la constante K_{CG} est grande. Cependant, les fluctuations atomiques globales observées avec la représentation MM/CG suivent la même tendance que dans la dynamique "tout-atome", notamment avec une valeur de 1 kcal.mol⁻¹. \AA^{-2} pour K_{CG} . L'augmentation de K_{CG} engendre donc une augmentation de la rigidité de la protéine, en diminuant les fluctuations atomiques (voir figure 7.4 B).

FIGURE 7.4 – Influence de K_{CG} sur la dynamique du BPTI dans le vide. Evolution des RMSD (A) et RMSF (B) calculés sur tous les C_α de la protéine pour la simulation "tout-atome" (en noir) et pour des simulations MM/CG avec $r_{CG} = 9$ \AA et K_{CG} prenant les valeurs de 1 kcal.mol⁻¹. \AA^{-2} (en rouge) et 10 kcal.mol⁻¹. \AA^{-2} (en vert).

Il est nécessaire de vérifier que la dynamique de la zone MM est peu perturbée par l'introduction du réseau de ressorts. Pour cela, nous avons comparé les mêmes grandeurs (RMSD et RMSF) calculées uniquement sur la zone MM. Celles-ci sont reportées sur la figure 7.5. Les valeurs du RMSD pour les simulations MM/CG sont en accord avec la simulation "tout-atome", même si en fin de simulation, un changement conformationnel est observé dans la dynamique "tout-atome" mais pas dans les simulations MM/CG. Concernant les fluctuations atomiques, s'il n'y a pas de réel accord entre les différentes simulations, il est clair que la gamme de variation avec la représentation MM/CG (entre 0,2 et 0,7 Å) est proche de celle observée avec la représentation "tout-atome".

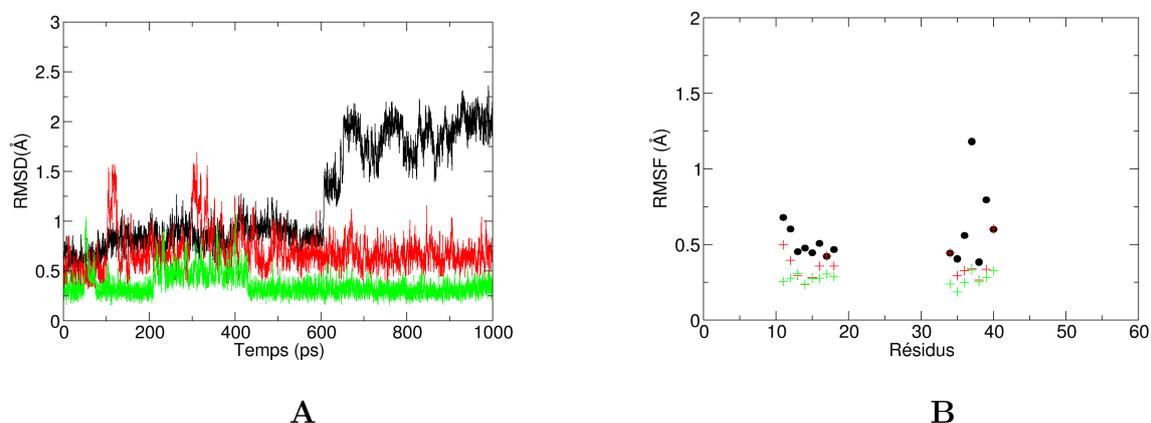


FIGURE 7.5 – Influence de K_{CG} sur la dynamique du BPTI dans le vide. Evolution des RMSD (A) et RMSF (B) calculés sur les C_{α} de la zone MM uniquement pour la simulation "tout-atome" (en noir) et pour des simulations MM/CG avec $r_{CG} = 9 \text{ \AA}$ et K_{CG} prenant les valeurs de $1 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ (en rouge) et $10 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ (en vert).

Les résultats décrits dans ce paragraphe pour une zone MM située autour du résidu 37 sont également observés pour le deuxième modèle MM/CG construit dans lequel le centre de la zone MM est le résidu 43, comme l'indiquent les fluctuations atomiques présentées sur la figure 7.6.

Rôle du rayon de coupure r_{CG}

Dans un deuxième temps, nous avons testé l'influence du rayon de coupure r_{CG} , qui permet de définir l'extension du réseau de ressorts. La figure 7.7 montre les fluctuations atomiques globales pour des valeurs de r_{CG} variables dans des simulations MM/CG et pour des zones MM centrées respectivement sur les résidus 37 et 43.

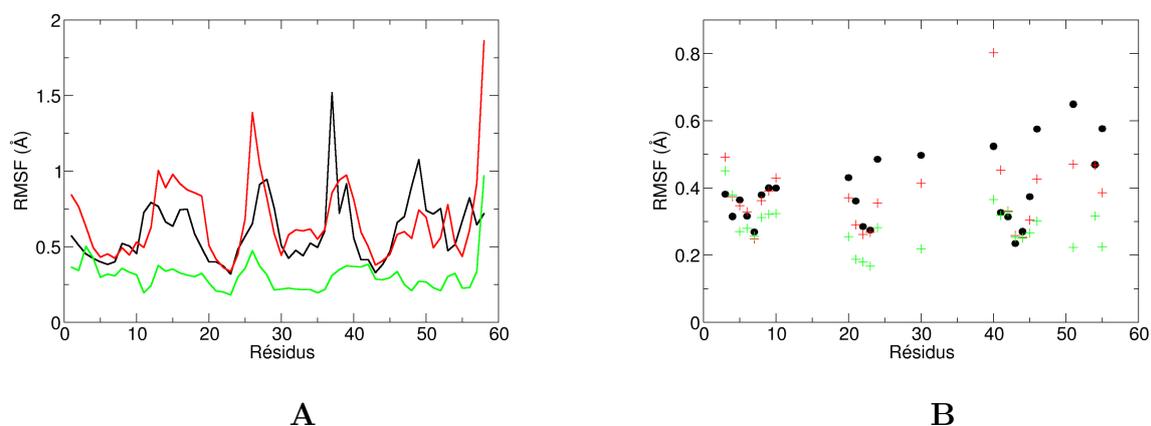


FIGURE 7.6 – Influence de K_{CG} sur la dynamique du BPTI dans le vide. Fluctuations atomiques des C_{α} calculées sur l'ensemble de la structure (A) et sur la région MM uniquement (B). La courbe noire représente la simulation "tout-atome" et les courbes rouge et verte représentent respectivement les simulations MM/CG avec des valeurs de K_{CG} de 1 et 10 kcal.mol⁻¹.Å⁻², la valeur de r_{CG} étant fixée à 9 Å.

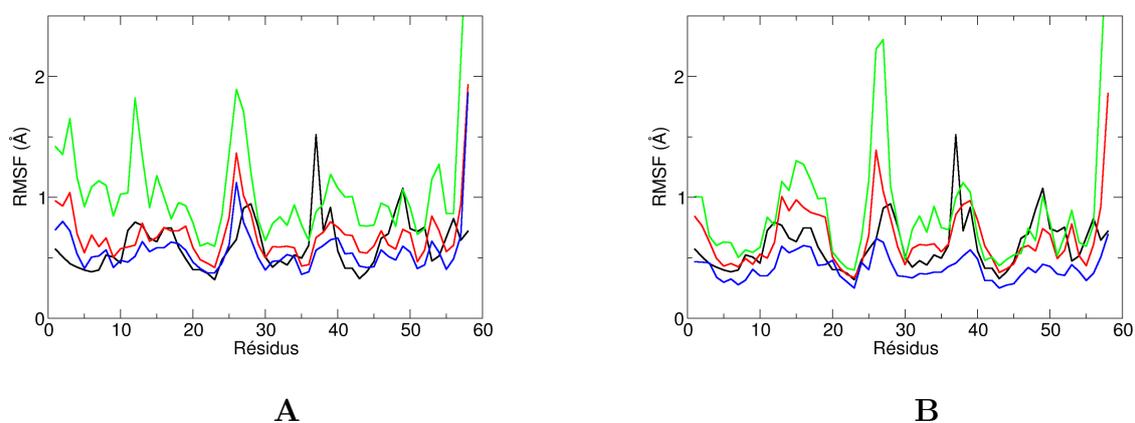


FIGURE 7.7 – Influence de r_{CG} sur la dynamique du BPTI dans le vide. RMSF calculés sur les C_{α} de la protéine pour la simulation "tout-atome" (en noir) et pour des simulations MM/CG avec $K_{CG} = 1$ kcal.mol⁻¹.Å⁻² et r_{CG} prenant les valeurs de 8 (en vert), 9 (en rouge), et 10 Å (en bleu) pour une zone MM centrée sur le résidu 37 (A) et sur le résidu 43 (B).

Comme déjà observé dans le paragraphe précédent, les tendances des fluctuations atomiques globales sont bien reproduites dans notre modèle MM/CG. La simulation dans laquelle la zone MM est centrée sur le résidu 43 reproduit notamment un peu mieux ces RMSF, certainement du fait d'un meilleur empaquetage par rapport à l'autre région MM testée. La figure 7.8 montre les mêmes grandeurs calculées uniquement sur la zone MM,

qui montre encore une fois que la gamme des fluctuations atomiques dans cette zone est conservée pour les différentes valeurs de r_{CG} testées.

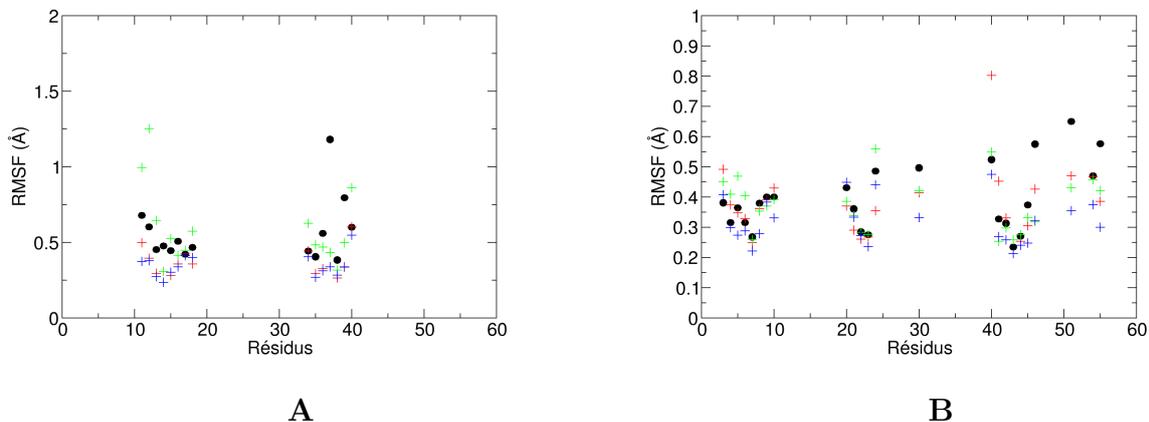


FIGURE 7.8 – Influence de r_{CG} sur la dynamique du BPTI dans le vide. RMSF calculés sur les C_α de la zone MM uniquement pour la simulation "tout-atome" (en noir) et pour des simulations MM/CG avec $K_{CG} = 1 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$ et r_{CG} prenant les valeurs de 8 (en vert), 9 (en rouge), et 10 Å (en bleu) pour une zone MM centrée sur le résidu 37 (A) et sur le résidu 43 (B).

Bilan

Une augmentation de r_{CG} entraîne l'augmentation du nombre de liens entre les résidus et accroît donc la rigidité de la structure, notamment dans la région CG. De même, l'augmentation de K_{CG} rigidifie l'architecture protéique. Au final, une valeur de r_{CG} de 9 ou 10 Å, couplée à une valeur d'environ $1 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$ pour K_{CG} permet d'obtenir un accord satisfaisant au niveau des fluctuations atomiques avec une simulation "tout-atome". Il est intéressant de noter que ces valeurs sont sensiblement les mêmes que celles utilisées dans les représentations "gros grains" (voir chapitre 3 et annexe B.3) pour reproduire les valeurs expérimentales. Les valeurs approchées de ces paramètres ont été obtenues avec de courtes simulations dans le vide et pourraient être affinées par des tests plus poussés. Néanmoins, les résultats obtenus sont assez encourageants pour passer à l'étape d'incorporation du solvant dans le modèle.

7.4.2 Incorporation du solvant dans le modèle

L'utilisation d'un solvant implicite du type Born généralisé aurait été possible mais dans la mesure où nous souhaitons garder un maximum de détail dans la zone MM, il nous

a semblé plus intéressant d'ajouter des molécules d'eau explicites. La région CG ne nécessitant pas de solvatation, nous avons limité l'ajout de molécules de solvant explicites à une sphère centrée sur la zone MM et de rayon suffisamment important pour englober toute cette zone. Cette approche est comparable à certaines études d'enzymes dans lesquelles seule une région située autour du site actif est solvatée et autorisée à se déplacer (Allouche *et al.*, 1999). Dans de telles études, le reste de la protéine est figé, ce qui autorise à ne pas incorporer de solvant autour de ces parties. Cependant, l'effet de la dynamique globale de la protéine sur la dynamique du site actif est alors négligée. Dans notre cas, la flexibilité allouée à la partie CG permet de conserver les vibrations collectives de basse fréquence de la protéine (voir section 7.5).

7.4.2.1 Mise en œuvre

Protocole

Le système a été construit en ajoutant une sphère de molécules d'eau TIP3P de rayon 25 Å autour du résidu central de la zone MM, toutes les molécules d'eau se trouvant à l'intérieur du volume de la protéine étant éliminées. Le système est alors soumis à 1000 pas de minimisation, la position des atomes de la protéine étant soumise à des contraintes harmoniques de $5 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$. Puis le système est chauffé en 10 ps jusqu'à 300 K et encore gardé à cette température pendant 15 ps de dynamique moléculaire sous contraintes. L'équilibration est poursuivie par 25 ps de simulation avec des contraintes diminuées puis 50 ps de dynamique libre. Le pas d'intégration est fixé à 2 fs, la longueur des liaisons impliquant un atome d'hydrogène étant fixée par l'algorithme SHAKE (Ryckaert *et al.*, 1977), et la température maintenue constante à 300 K par couplage avec un bain de température (Berendsen *et al.*, 1984).

Au cours des simulations, les molécules d'eau sont confinées à l'intérieur de la sphère de solvatation par un potentiel harmonique. Cette contrainte ne s'applique sur une molécule d'eau que lorsque celle-ci sort de la sphère. De plus, pour que la zone MM de la protéine reste dans la sphère de solvatation, la position du C_α du résidu central est fixée.

Interactions CGX / molécules d'eau

Il a été nécessaire de paramétrer les interactions entre les résidus CGX et les molécules d'eau. En effet, il faut s'assurer que les molécules d'eau ne pénètrent pas à l'intérieur de la région CG de la structure protéique, afin de ne pas envelopper entièrement la zone MM.

Pour cela, nous avons choisi de conserver le terme répulsif des interactions de Lennard-Jones entre un atome CGX et l'atome d'oxygène des molécules d'eau¹ :

$$E_{CG/W} = \frac{A_{CG/W}}{r^{12}}$$

Différentes valeurs de $A_{CG/W}$ ont été testées sur la période d'équilibration décrite dans le paragraphe précédent et les structures finales comparées afin de détecter la présence d'eau à l'intérieur de la zone CG. La figure 7.9 montre que pour des valeurs de $A_{CG/W}$ supérieures à 10^6 kcal.mol⁻¹.Å¹², les molécules d'eau restent à l'extérieur de la structure protéique.

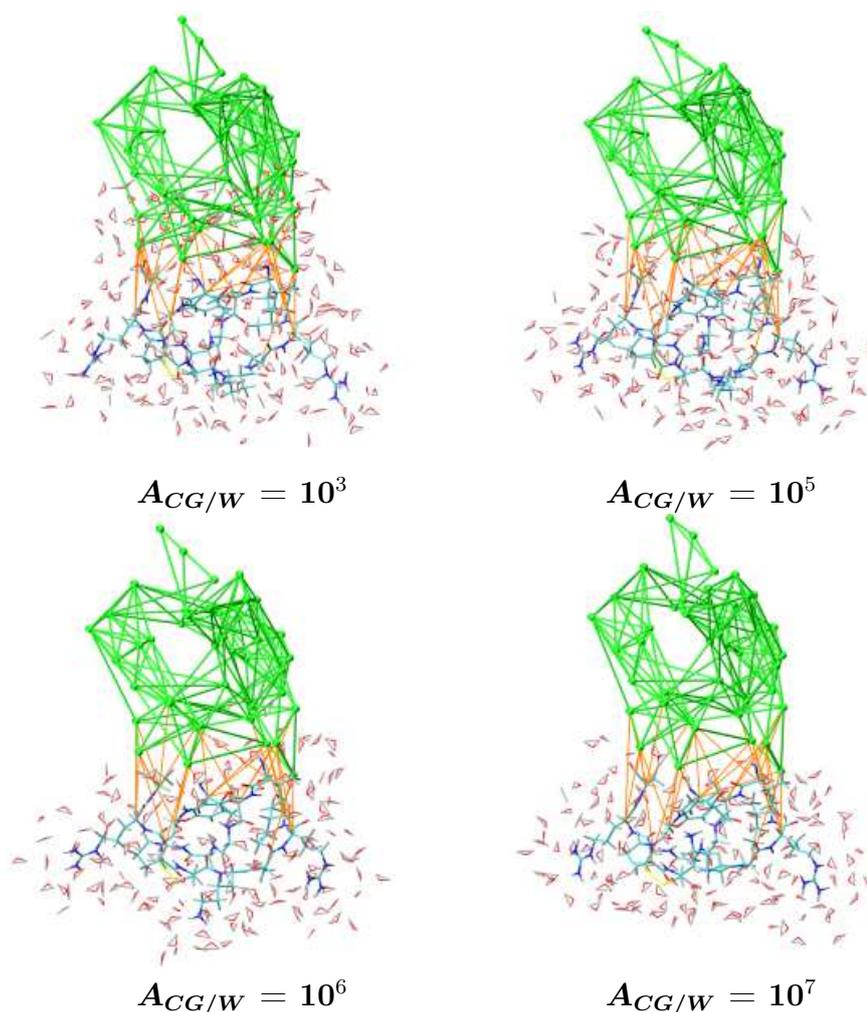


FIGURE 7.9 – Paramétrage des interactions répulsives entre l'eau et les résidus CGX. Les images montrent la position des molécules d'eau situées à moins de 4 Å de la protéine après la phase d'équilibration pour différentes valeurs de $A_{CG/W}$ (données en kcal.mol⁻¹.Å¹²).

1. Dans le modèle d'eau TIP3P, les atomes d'hydrogène ne réalisent pas d'interaction de van der Waals.

Troncature des interactions non liées

L'utilisation de conditions non périodiques lors de nos simulations, si elle permet de réduire considérablement le nombre d'atomes, impose de choisir un rayon de coupure pour les interactions non liées plus grand que celui utilisé avec des conditions périodiques. Les temps de calcul obtenus pour différents choix de ce rayon de coupure sont donnés dans le tableau 7.2.

Type de simulation	Rayon de coupure des interactions non liées (Å)	Nombre d'atomes	Temps CPU pour 500 ps de simulation (en heures)
"tout-atome"	10 (PME)	25873	180
MM/CG	∞	5190	200
MM/CG	20	5190	100
MM/CG	15	5190	55

TABLE 7.2 – Temps CPU pour différentes troncatures des interactions non liées dans le modèle MM/CG. La simulation "tout-atome" a été réalisée avec des conditions périodiques et l'algorithme PME a été utilisé pour traiter les interactions électrostatiques.

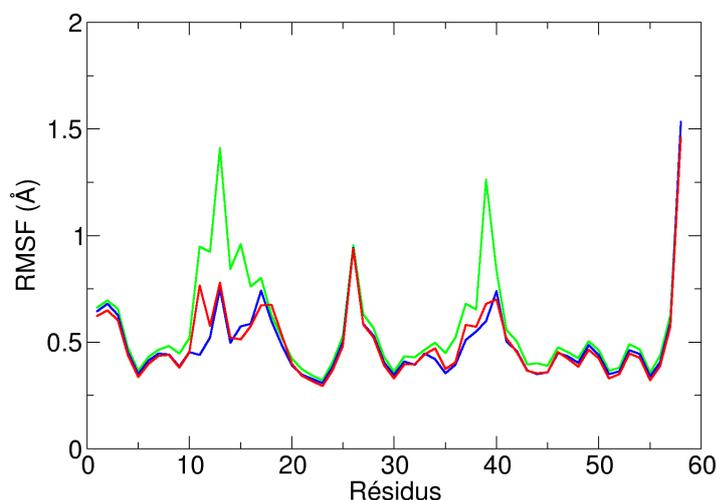


FIGURE 7.10 – Profil des fluctuations atomiques des résidus du BPTI lors de simulations MM/CG réalisées en présence de molécules d'eau explicites sans troncature des interactions non liées (en vert), et avec une troncature fixée à 20 Å (en bleu) ou 15 Å (en rouge). Un paramètre $A_{CG/W}$ de $10^6 \text{ kcal.mol}^{-1}.\text{Å}^{12}$ a été utilisé pour les interactions répulsives entre les molécules d'eau et les résidus CGX.

En utilisant un rayon de coupure infini, la simulation en représentation MM/CG devient plus lente qu'une simulation "tout-atome" conventionnelle utilisant des conditions périodiques. En abaissant la troncature à 15 Å, on diminue le temps de calcul d'un facteur 3 environ par rapport à cette même simulation "tout-atome", tandis que le profil des fluctuations atomiques reste comparable à celui obtenu sans troncature du potentiel (voir figure 7.10). Par la suite, nous fixerons donc le rayon de coupure des interactions non liées à cette valeur de 15 Å.

7.4.2.2 Comparaison avec une simulation "tout-atome" en conditions périodiques

Nous avons comparé nos simulations MM/CG à une simulation utilisant une représentation "tout-atome" et réalisée avec des conditions périodiques¹.

Comme l'indique la figure 7.11 A, les déviations par rapport à la structure initiale sont bien reproduites par le modèle MM/CG, avec des RMSD similaires dans toutes les simulations (entre 1 et 1,25 Å pour tous les C_α de la protéine). Il est à noter qu'au cours de la trajectoire "tout-atome", on observe un petit changement conformationnel après environ 5,5 ns, repéré par une augmentation du RMSD d'environ 0,5 Å.

Les fluctuations atomiques des C_α sont montrées sur la figure 7.11 B. Deux courbes représentent la simulation "tout-atome" en raison du changement conformationnel évoqué plus haut. L'accord entre les simulations mixtes MM/CG et la simulation "tout-atome" est très satisfaisant. En effet, les variations des RMSF avec la représentation MM/CG (en rouge et en bleu) suivent celles de la simulation "tout-atome" pour l'ensemble de la protéine, notamment dans la région représentée à l'échelle atomique (résidus 11 à 18 et 34 à 40). Ce constat est valable également pour le système dans lequel la zone MM du modèle MM/CG est centrée autour du résidu 43 (voir figure 7.12).

1. La structure initiale a été immergée dans une boîte d'eau TIP3P octaédrique tronquée. La distance minimale entre les faces de la boîte et le soluté a été fixée à 15 Å. Six ions Cl^- ont été ajoutés et placés aléatoirement dans la cellule pour assurer l'électroneutralité du système. Après 5000 pas de minimisation sous contraintes, le système est chauffé à 300 K en 10 ps et maintenu à cette température pendant encore 40 ps de dynamique moléculaire dans l'ensemble *NVT*. Puis les contraintes sont progressivement relâchées dans une succession de phases de 2000 pas de minimisation et 25 ps de dynamique moléculaire dans l'ensemble *NPT*. Enfin, 50 ps de simulation sans contrainte sont réalisées avant la phase de production qui couvre 10 ns. La simulation est réalisée à pression et température maintenues constantes respectivement à 1 bar et 300 K par couplage avec des bains de pression et température. Le pas d'intégration est fixé à 2 fs et la longueur des liaisons impliquant un atome d'hydrogène est fixée à l'aide de l'algorithme SHAKE (Ryckaert *et al.*, 1977). L'algorithme PME est utilisé pour le calcul des interactions électrostatiques en utilisant un rayon de coupure de 10 Å dans l'espace réel.

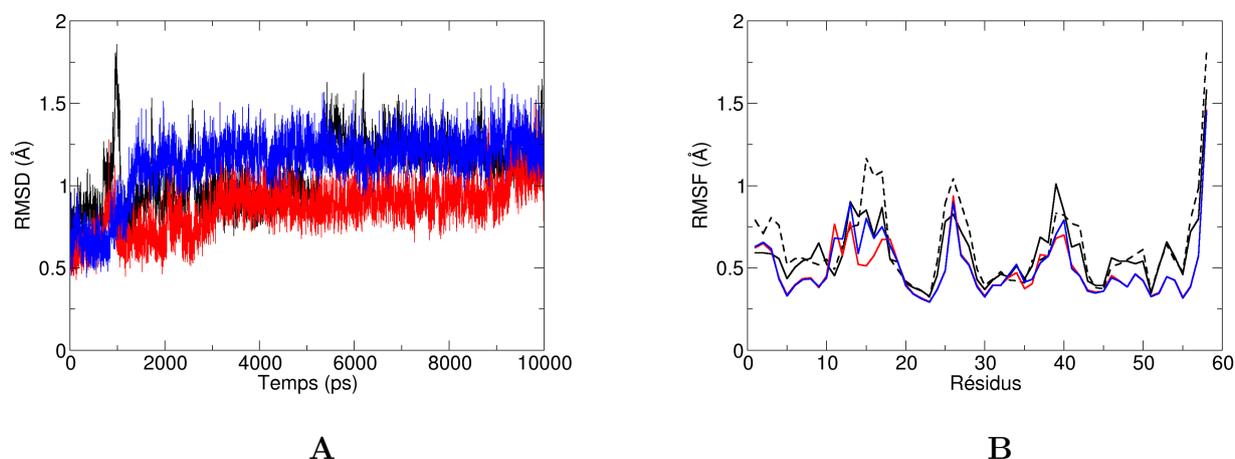


FIGURE 7.11 – RMSD (A) et RMSF (B) calculés sur les C_α dans des simulations du BPTI en modèle "tout-atome" et conditions périodiques (en noir), et en représentation MM/CG avec un paramètre $A_{CG/W}$ de 10^6 (en rouge) et 10^7 kcal.mol⁻¹.Å¹² (en bleu). Dans la figure B, les courbes noires continues et brisées représentent les fluctuations atomiques dans la simulation "tout-atome" calculées respectivement sur les intervalles de temps [2 ns ; 5 ns] et [7 ns ; 10 ns].

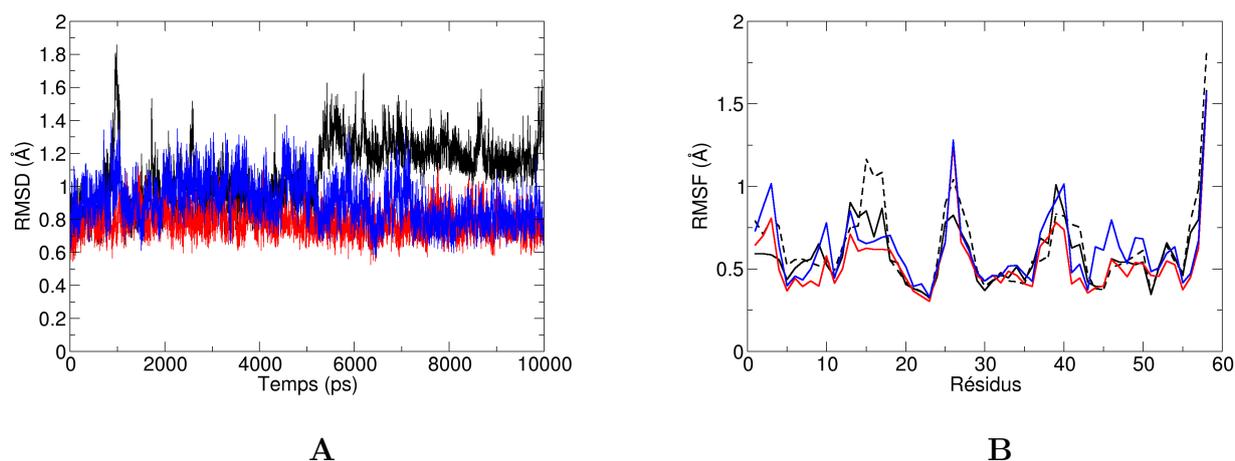


FIGURE 7.12 – RMSD (A) et RMSF (B) calculés sur les C_α dans des simulations du BPTI en modèle "tout-atome" et conditions périodiques (en noir), et en représentation MM/CG centrée autour du résidu 43 (bleu et rouge). Les valeurs des paramètres sont les mêmes que celles données dans la figure 7.11.

7.5 Étude d'un homodimère

7.5.1 Choix du système

Au vu des résultats encourageants obtenus sur le BPTI, nous avons décidé de tester notre représentation MM/CG sur une protéine de plus grande taille. Notre choix s'est porté sur l'étude de l'adénine-phosphoribosyltransférase (notée APRTase par la suite). Cette protéine catalyse la formation de l'adénosine monophosphate et du pyrophosphate à partir de l'adénine et du 5-phospho- α -D-ribose-1-pyrophosphate en présence d'ions magnésium et joue ainsi un rôle de recyclage de l'adénine pour reformer des nucléotides d'adénine. Au niveau structural, l'APRTase se présente sous la forme d'un homodimère dont chaque sous-unité de 178 résidus est composée de cinq hélices α et neuf brins β .

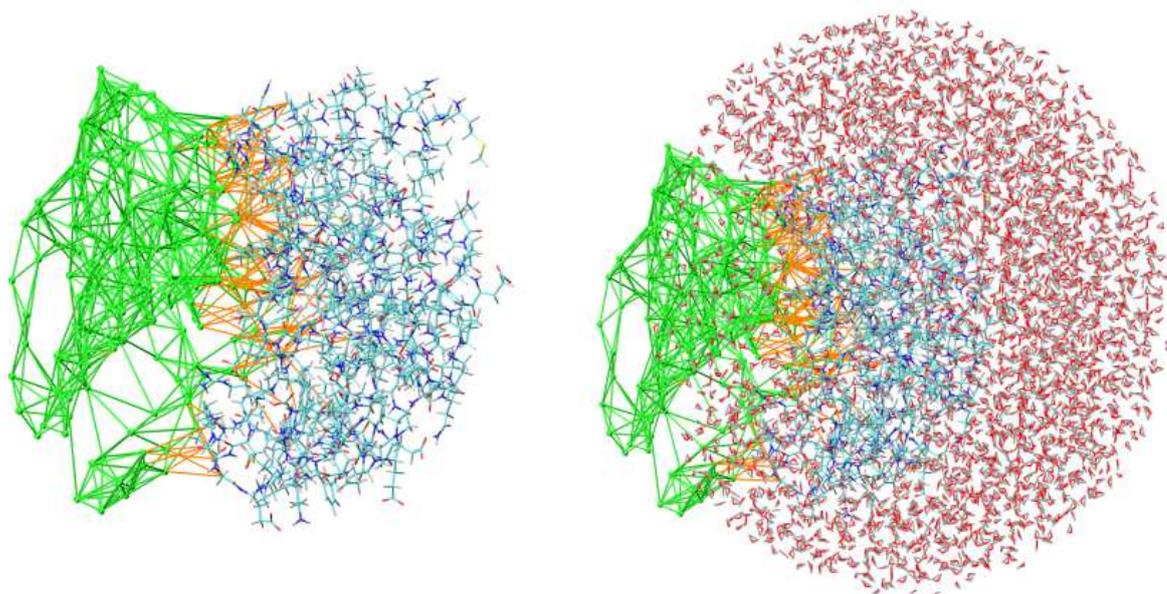


FIGURE 7.13 – Représentation MM/CG de l'APRTase avant et après ajout de la sphère de solvation.

La structure homodimérique de l'APRTase offre la possibilité de construire un modèle MM/CG dans lequel un des monomères est représenté à l'échelle atomique (noté monomère MM), l'autre (noté monomère CG) étant représenté de façon simplifiée (voir figure 7.13). Les paramètres choisis pour le modèle MM/CG sont les suivants :

- $K_{CG} = 1 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$.
- $r_{CG} = 10 \text{ \AA}$.
- $A_{CG/W} = 10^7 \text{ kcal.mol}^{-1}.\text{\AA}^{12}$.

Nous avons également réalisé une simulation de dynamique moléculaire de l'APRTase en représentation "tout-atome" en solvant explicite avec des conditions périodiques (simulation TA). Les conditions de simulation et le protocole sont identiques à ceux utilisés pour le BPTI (voir section précédente 7.4). Trois structures issues de la trajectoire TA ont servi de point de départ pour des simulations MM/CG :

- la structure initiale (simulation T0).
- la structure extraite après 7 ns de simulation "tout-atome" (simulation T7).
- la structure extraite après 10 ns de simulation "tout-atome" (simulation T10).

7.5.2 Dynamique globale du système

7.5.2.1 Comparaison des fluctuations atomiques

La comparaison des fluctuations atomiques est montrée sur la figure 7.14. Sur la figure A sont représentés les RMSF calculés sur le monomère MM des simulations T0, T7 et T10, tandis que la figure B montre les RMSF calculés sur le monomère CG. Sur les deux figures sont également représentées les fluctuations atomiques des deux monomères de la simulation TA (ligne noire et ligne noire brisée). Les tendances observées dans la simulation TA sont retrouvées dans les simulations MM/CG, à la fois pour le monomère MM et pour le monomère CG. L'accord le plus satisfaisant est obtenu pour les simulations T7 et T10, c'est-à-dire lorsque la structure de départ de la simulation MM/CG provient d'une structure pré-équilibrée par une simulation de dynamique moléculaire classique. Ainsi, le coefficient de corrélation entre les fluctuations atomiques de la simulation TA et de la simulation T10 est voisin de 0,9 (voir tableau 7.3). Ceci peut certainement être relié aux contraintes imposées par le réseau de ressort qui limite les possibilités de relaxation dans la simulation T0 (voir par exemple les zones autour des résidus 25 et 165). Par la suite, nous discuterons uniquement de la comparaison entre les simulations TA et T10, car cette dernière est celle qui reproduit au mieux les propriétés de la dynamique TA (l'ensemble des résultats obtenus avec les simulations T0 et T7 sont cependant assez similaires).

Simulation	T0		T7		T10	
	MM	CG	MM	CG	MM	CG
TA "MM"	0,60	0,74	0,78	0,90	0,89	0,91
TA "CG"	0,61	0,72	0,81	0,84	0,85	0,87

TABLE 7.3 – Coefficients de corrélation des fluctuations atomiques entre le modèle MM/CG et le modèle "tout-atome".

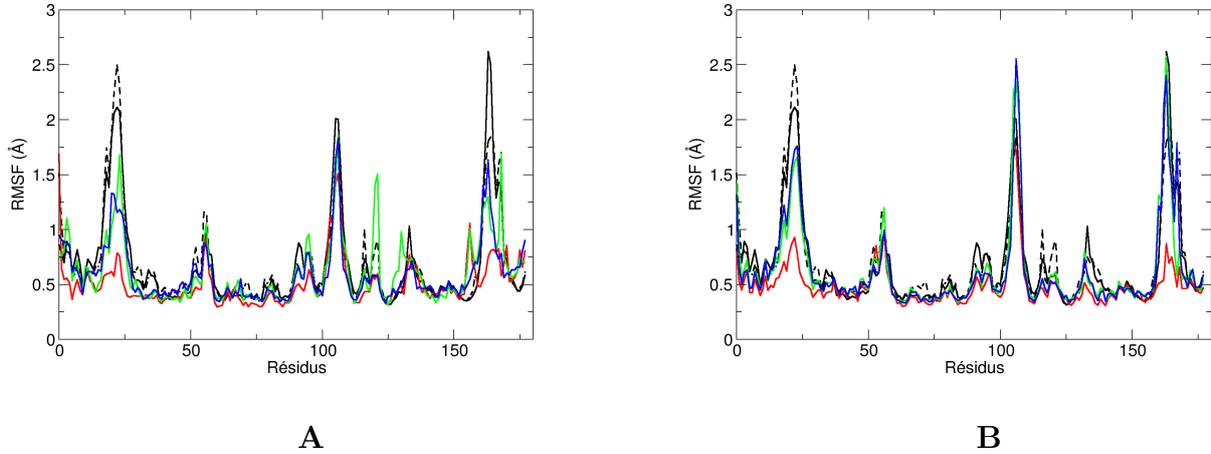


FIGURE 7.14 – Fluctuations atomiques calculées sur les C_α de chaque résidu à l'intérieur d'un monomère dans la simulation TA (en noir) et pour les simulations MM/CG T0 (en rouge), T7 (en vert), et T10 (en bleu).

7.5.2.2 Modes de vibration

Une autre façon de mesurer l'accord entre les dynamiques MM/CG et la dynamique TA est de comparer les modes quasi-harmoniques de vibration de l'APRTase au cours des différentes trajectoire (voir annexe E pour le calcul des modes quasi-harmoniques). Plus précisément, nous avons cherché à mesurer à quel point les modes de plus basse fréquence de la dynamique TA peuvent être reproduits par les modes de basse fréquence des dynamiques MM/CG.

Différents modes de vibration ont été calculés sur l'ensemble des C_α du dimère et également sur les C_α de chacun des monomères pour les cinq dernières nanosecondes de dynamique. Les modes v_i^{TA} obtenus pour la dynamique "tout-atome" sont ensuite projetés sur les modes v_i d'une trajectoire MM/CG :

$$v_i^{TA} = \sum_j c_{ij} v_j$$

Le coefficient c_{ij} représente le produit scalaire de v_i^{TA} et v_j . Les modes quasi-harmoniques utilisés ici étant normalisés, il peut donc prendre des valeurs entre -1 et 1, une valeur absolue de c_{ij} proche de 1 indiquant des directions de déplacement proches pour les deux

vecteurs. Pour comparer les jeux de vecteurs entre eux, nous avons calculé des coefficients C_i , définis par :

$$C_i = \sqrt{\sum_{j=1}^N c_{ij}^2}$$

Ce coefficient C_i (qui peut varier entre 0 et 1) quantifie la capacité des N premiers modes de vibration v_j de la trajectoire MM/CG à reproduire le mode v_i^{TA} de la trajectoire TA. En d'autres termes, plus cette valeur est proche de 1, et plus l'espace conformationnel défini par les N premiers modes v_j recouvre celui défini par le mode v_i^{TA} (si la somme est réalisée sur tous les vecteurs v_j , alors C_i prend une valeur de 1, car les deux bases $\{v_i^{TA}\}$ et $\{v_j\}$ forment une base complète de l'espace accessible aux C_α). Nous avons choisi d'utiliser $N = 30$ car les 30 premiers vecteurs propres de chacune des trajectoires participent pour environ 80 % des fluctuations atomiques totales calculées sur les C_α , et reproduisent notamment les pics majoritaires observés (voir figure 7.15).

	Dimère	Monomère MM	Monomère CG		Dimère	Monomère MM	Monomère CG
C_1	0,77	0,88	0,96	C_{16}	0,72	0,77	0,76
C_2	0,81	0,91	0,83	C_{17}	0,72	0,72	0,77
C_3	0,79	0,94	0,88	C_{18}	0,66	0,66	0,60
C_4	0,86	0,91	0,94	C_{19}	0,65	0,75	0,65
C_5	0,86	0,82	0,92	C_{20}	0,61	0,74	0,63
C_6	0,83	0,88	0,86	C_{21}	0,73	0,74	0,62
C_7	0,80	0,79	0,87	C_{22}	0,57	0,52	0,59
C_8	0,90	0,87	0,82	C_{23}	0,58	0,68	0,52
C_9	0,85	0,87	0,80	C_{24}	0,63	0,63	0,44
C_{10}	0,80	0,84	0,74	C_{25}	0,57	0,61	0,53
C_{11}	0,80	0,84	0,84	C_{26}	0,48	0,57	0,58
C_{12}	0,68	0,83	0,89	C_{27}	0,51	0,62	0,55
C_{13}	0,73	0,78	0,62	C_{28}	0,50	0,77	0,58
C_{14}	0,75	0,85	0,62	C_{29}	0,44	0,62	0,44
C_{15}	0,66	0,66	0,69	C_{30}	0,53	0,51	0,51

TABLE 7.4 – Décomposition des 30 premiers modes de vibration de la dynamique TA dans l'espace des 30 premiers modes de vibration de la dynamique T10.

Les valeurs des coefficients C_i sont données dans le tableau 7.4. La première colonne du tableau 7.4 correspond aux modes de vibration du dimère tandis que les deux autres colonnes se réfèrent aux modes de vibrations au sein de chacun des monomères de l'APRTase. On remarque que les dix premières valeurs des C_i sont toujours supérieures à 0,7 et le plus

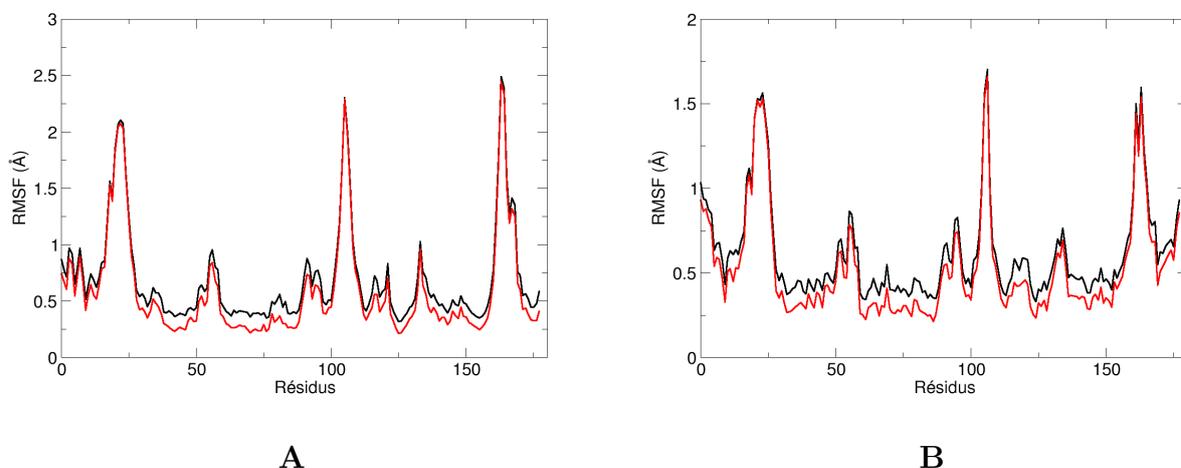


FIGURE 7.15 – Fluctuations atomiques calculées sur les C_α de chaque résidu pour le monomère MM à partir de la dynamique totale (en noir) et en utilisant les 30 modes de vibration de plus basse fréquence (en rouge) dans le cas des simulations TA (**A**) et T10 (**B**).

souvent supérieurs à 0,8. Ceci indique que les 30 premiers modes de basse fréquence de la simulation MM/CG recouvre de façon notable les mouvements globaux de grande amplitude de la dynamique TA. Ceci est d'autant plus vrai que l'on s'intéresse à la dynamique des monomères pris individuellement, avec une moyenne de 0,86 et 0,87 sur les 10 premiers C_i . Au-delà des 10 premiers vecteurs en revanche, l'accord est moins bon (notamment pour le monomère CG). Néanmoins, on peut considérer que l'espace conformationnel défini par les modes de vibration de plus basse fréquence dans la dynamique MM/CG recouvrent de manière significative le même espace défini dans la dynamique TA.

7.5.3 Fluctuations atomiques dans la zone MM

Afin de déterminer si les détails dans la région MM sont conservés lors des simulations MM/CG, nous avons calculé les fluctuations atomiques sur tous les atomes lourds du monomère MM. Ceci donne un indice sur l'amplitude des fluctuations des chaînes latérales au cours des simulations. La comparaison entre la simulation T10 et la simulation TA est montrée sur la figure 7.16. L'accord entre les deux profils est excellent avec un coefficient de corrélation de 0,89 pour 1387 points, indiquant que la représentation MM/CG n'introduit pas de perturbation majeure en ce qui concerne la mobilité au niveau atomique dans la région MM.

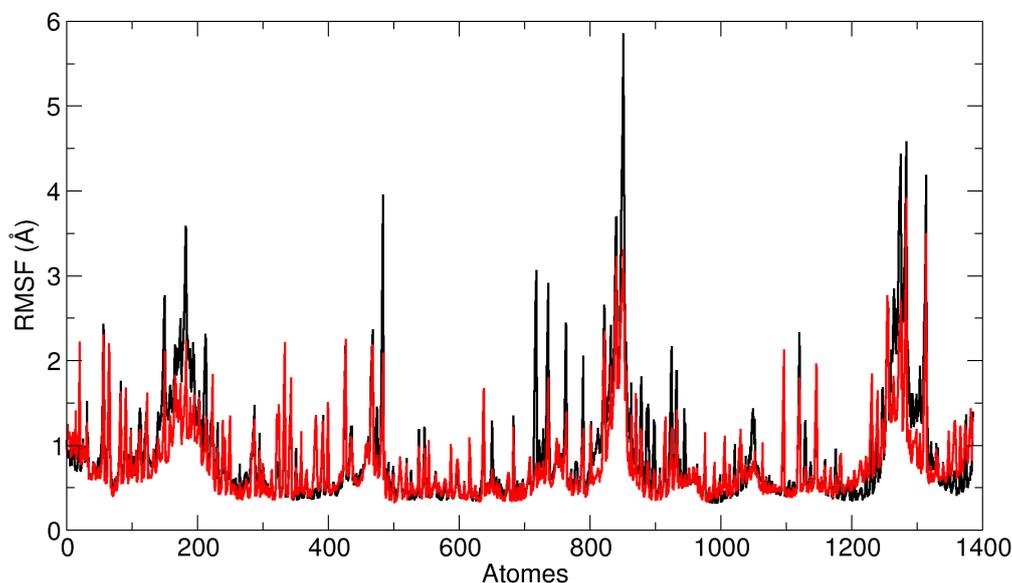


FIGURE 7.16 – Fluctuations atomiques calculées sur tous les atomes lourds du monomère MM dans les dynamiques TA (en noir) et T10 (en rouge).

7.6 Bilan et perspectives

Nous avons défini une nouvelle représentation multi-échelle incorporant une représentation atomique de certaines régions de la protéine, couplée à une représentation "gros grains". Les premiers résultats obtenus sont encourageants : ils indiquent que la dynamique globale des protéines est relativement bien conservée par rapport à une simulation classique "tout-atome" en solvant explicite et que les fluctuations au niveau atomique sont également bien reproduites dans la partie "tout-atome" du système.

Il sera cependant nécessaire de déterminer si les résultats obtenus dans notre étude sont généralisables à d'autres architectures protéiques. De plus, cela permettra de reconsidérer le paramétrage du modèle, qui n'a pas été abordé de manière extensive dans ce travail exploratoire. Enfin, différentes améliorations sont envisageables afin d'améliorer la précision de la représentation MM/CG. Parmi celles-ci, on peut citer :

- Utiliser une représentation plus détaillée pour la zone CG, avec un à trois points par résidus pour mimer la taille des chaînes latérales (voir (Sacquin-Mora et Lavery, 2006)).

- Ajouter des ressorts entre les chaînes latérales des résidus de la zone MM en interaction avec des résidus CG, afin de minimiser la flexibilité artificielle créée par le retrait des chaînes latérales des résidus CG à l'interface entre les deux régions MM et CG.
- Modifier le traitement des interactions électrostatiques. En effet, la troncature des interactions non liées est une troncature brutale. Ceci est assez peu réaliste et peut conduire à des instabilités. Il serait intéressant de tester l'utilisation d'une fonction de "*shift*" (voir annexe D.2). Ceci nécessite l'introduction de la représentation dans un autre programme qu'AMBER¹, qui ne propose pas ce traitement, ou la modification directe du traitement des interactions électrostatiques dans le programme AMBER.
- Une limitation de la représentation MM/CG est que les mouvements des atomes CGX sont restreints à des fluctuations harmoniques autour d'une position initiale, empêchant les changements conformationnels importants. On pourrait peut-être modifier ce comportement en utilisant un potentiel de Morse à la place des ressorts utilisés ici.

1. Le programme NAMD pourrait être facilement utilisé car il est capable d'utiliser les fichiers de topologie d'AMBER.

Conclusion générale

Comme illustré dans plusieurs chapitres de ce manuscrit, les propriétés mécaniques des protéines jouent un rôle fondamental pour leur fonctionnement, et donc, indirectement pour le bon fonctionnement des organismes. Le travail présenté dans ce manuscrit, réalisé au cours des quatre dernières années au Laboratoire de Biochimie Théorique, a donc consisté, à travers différents aspects, à tenter d'améliorer notre connaissance des propriétés mécaniques des protéines.

Au niveau expérimental, le développement des expériences de nanomanipulation de molécules uniques a ouvert de nouvelles perspectives pour l'étude de ces propriétés mécaniques en permettant de mesurer directement la résistance d'une protéine à l'application d'une force. Cette évolution s'est accompagnée d'un développement parallèle de la modélisation moléculaire, qui a montré sa capacité à décrire, à l'échelle atomique, les déformations induites dans les protéines par un stress mécanique. Dans la continuité de ces études théoriques dans lesquelles l'application d'une force sur la protéine est modélisée à l'aide d'une contrainte, nous avons développé un nouvel indice permettant de mesurer la résistance mécanique d'une protéine à l'échelle du résidu. La méthodologie mise en place, fondée sur des minimisations d'énergie sous contrainte, est utilisable à la fois avec une représentation détaillée à l'échelle atomique et avec une représentation simplifiée de la protéine, permettant une étude plus systématique. En plus de l'information obtenue sur les zones de relative flexibilité et rigidité de la protéine, cette méthode permet également de définir, au sein d'une architecture protéique, des domaines dynamiques au sein desquels les atomes se déplacent préférentiellement de manière concertée. La constante de force définie dans notre travail a d'ores-et-déjà été utilisée (bien qu'obtenue par une autre méthode) pour étudier les liens existant entre la flexibilité des résidus catalytiques et leur fonction au sein des protéines à hème (Sacquin-Mora et Lavery, 2006).

Nous avons également étudié *in silico* la dynamique des cadhérines, présentée dans la deuxième partie du manuscrit, à l'aide de simulations de dynamique moléculaire. Ces mo-

lécules d'adhésion cellulaire sont un sujet d'étude idéal pour tenter de mieux comprendre comment les propriétés mécaniques des protéines permettent l'adhésion entre deux cellules. Nos travaux ont porté sur une partie de ces propriétés, à savoir la dépendance de la flexibilité des E-cadhérines vis-à-vis des ions calcium et la stabilité de l'interface entre deux molécules. Les résultats obtenus sont intéressants dans la perspective de la compréhension de l'adhésion impliquant les cadhérines, mais de nombreux travaux supplémentaires sont nécessaires pour avoir une vision aussi complète que possible de ce phénomène. L'étude du mécanisme d'échange de brins notamment a ainsi été débutée récemment au laboratoire par dynamique moléculaire sous contrainte. De plus, l'impact de différentes mutations au niveau des interfaces pourrait apporter de nouveaux indices sur l'origine de la sélectivité des interactions au sein de la famille des cadhérines.

Dans la troisième partie de ce manuscrit est présentée l'ébauche d'une nouvelle représentation des protéines, combinant les avantages d'une représentation "tout-atome" à ceux des représentations "gros grains" en mélangeant ces deux représentations pour un même système. Les résultats préliminaires obtenus avec ce modèle semblent indiquer qu'il est à même de conserver les propriétés dynamiques globales de l'ensemble du système ainsi que les propriétés locales au sein de la région "tout-atome". Ce type de représentation permettra de s'atteler à la résolution de problèmes inaccessibles encore aux simulations "tout-atome" en permettant d'économiser les capacités des ordinateurs et en minimisant les temps de calcul. Une application possible de ce modèle pourrait être l'impact de diverses mutations sur les propriétés des interfaces dans les dimères de cadhérine, en représentant par exemple les deux modules EC en interaction à l'échelle atomique et les autres à l'échelle du résidu. Ainsi, en diminuant de façon importante les temps de calcul de chaque simulation, le nombre d'études menées sur ces systèmes et le nombre de données obtenues pourraient être augmentés.

Au cours de ce travail, nous avons donc à la fois commencé l'étude d'un système biologique dont les propriétés mécaniques jouent un rôle certain pour la fonction et développé des outils dont l'application permettra d'obtenir de nouvelles informations sur la mécanique des protéines, directement à l'aide de la constante de force décrite dans la partie I, ou indirectement en accélérant les calculs par l'utilisation de représentations multi-échelle des protéines. De nombreuses perspectives sont donc envisageables pour ce travail à tous les niveaux, notamment avec l'application des outils développés aux cadhérines.

Au sortir de notre étude, les propriétés mécaniques des protéines restent bien évidemment dans leur grande majorité encore difficiles à cerner. Nous espérons cependant que les progrès réalisés au cours de ce travail auront permis et permettront d'améliorer notre compréhension des liens entre ces propriétés et la fonction des protéines.

Annexes

Annexe A

Structure des protéines

Les protéines sont des polymères biologiques dont les unités structurales (ou résidus) sont les acides aminés. L'agencement dans l'espace de ces acides aminés détermine la structure tridimensionnelle de la protéine. Cet agencement est guidé par les propriétés physico-chimiques des résidus qui composent la protéine, par leur ordre dans la séquence, ainsi que par l'environnement. L'obtention de cette structure généralement unique, appelée structure native, est gouvernée par un contrôle thermodynamique, c'est-à-dire qu'elle est la conformation d'enthalpie libre la plus basse (Anfinsen et Scheraga, 1975). Au sein de la structure protéique, on peut distinguer plusieurs niveaux hiérarchiques de structuration.

A.1 Structure primaire - la chaîne polypeptidique

Le premier niveau hiérarchique correspond à l'enchaînement séquentiel des acides aminés. Les acides aminés qui composent les protéines sont des acides α -aminés dont la formule générale est : $H_2N-C_\alpha HR-CO_2H$. Ils diffèrent par le groupement R porté par le C_α (voir figure A.1). Ce groupement est appelé chaîne latérale et sa nature (hydrophobe, polaire, ou chargé) va conférer à l'acide aminé des propriétés bien particulières, qui vont influencer le repliement et/ou la fonction de la protéine. Il est à noter qu'à l'exception de la glycine ($R = H$), le C_α d'un acide aminé est un centre chiral, dont la configuration fortement majoritaire dans la nature est la configuration L.

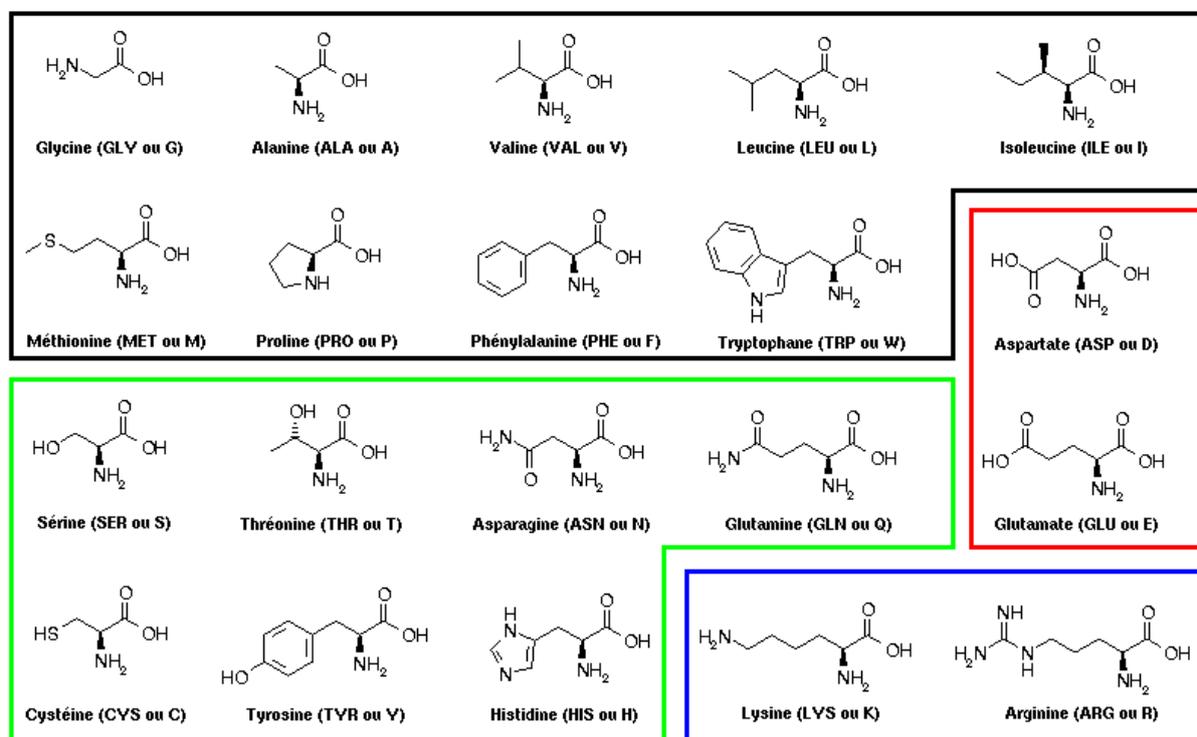


FIGURE A.1 – Structure des vingt acides aminés composant les protéines. Les cadres regroupent les acides aminés selon leurs propriétés physico-chimiques : hydrophobes en noir, polaires en vert, acides en rouge, et basiques en bleu. Les codes à trois et une lettres sont également donnés pour chaque résidu.

À pH physiologique ($\text{pH} \approx 7$), le résidu se trouve en réalité sous forme zwitterionique (la fonction acide carboxylique est déprotonée tandis que la fonction amine est protonée). La liaison entre deux acides aminés i et $(i + 1)$ est assurée par la création d'une liaison peptidique (voir figure A.2) et l'enchaînement ordonné (du résidu N-terminal au résidu C-terminal) de tous les acides aminés forme la séquence primaire de la protéine.

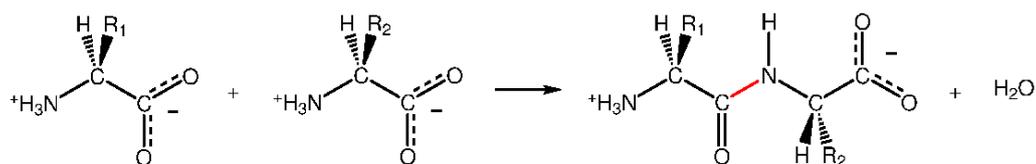


FIGURE A.2 – Formation de la liaison peptidique (en rouge).

A.2 Les structures secondaires

Le deuxième niveau hiérarchique de la conformation spatiale des protéines correspond à la formation de structures locales dont la géométrie est dictée par la formation de réseaux de liaisons hydrogène.

Diagramme de Ramachandran

Trois angles dièdres imposent la géométrie du brin peptidique : les angles ϕ , ψ , et ω , repérés sur la figure A.3 A. Les valeurs prises par ces angles sont soumises à certaines limitations. Ainsi, la planéarité de la liaison peptidique, imposée par la conjugaison entre le doublet non liant de l'azote et la liaison π du groupement carbonyle, ne permet que deux valeurs pour ω : 0° (conformation *cis*) et 180° (conformation *trans*), le profil énergétique pour passer de l'une à l'autre de ces conformations présentant une barrière énergétique élevée (de l'ordre de 20 kcal.mol^{-1})¹. Les angles ϕ et ψ peuvent prendre plus de valeurs mais les combinaisons $(\phi ; \psi)$ accessibles restent cantonnées à des valeurs limitées, que l'on peut représenter sur le diagramme de Ramachandran (voir figure A.3 B).

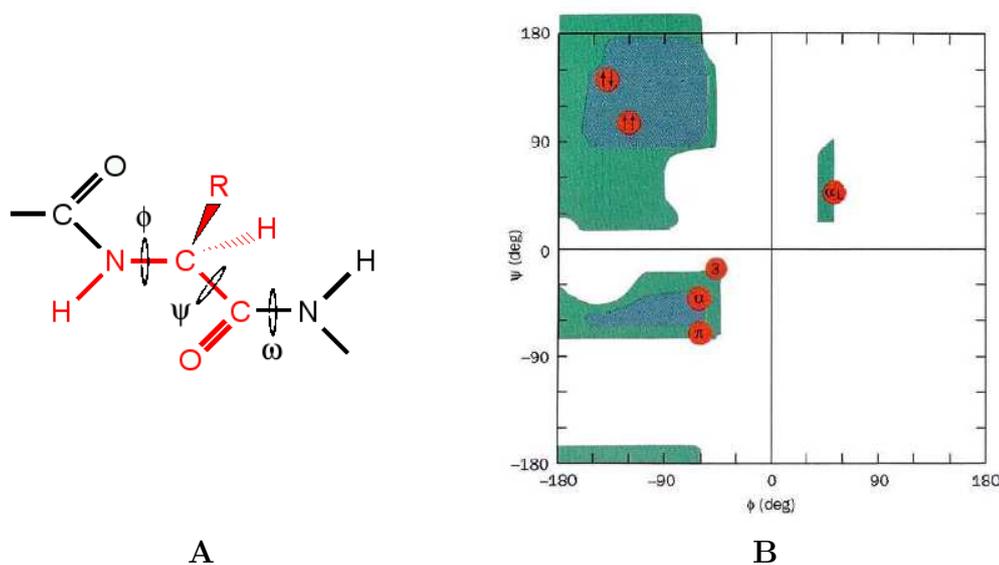


FIGURE A.3 – Conformation du squelette peptidique. **A** : Définition des angles dièdres ϕ , ψ , et ω du squelette peptidique. **B** : Diagramme de Ramachandran présentant les couples $(\phi ; \psi)$ accessibles (en vert). Les zones bleues correspondent aux conformations les plus stables. Les régions $(\phi ; \psi)$ correspondant aux hélices α droites (α), α gauches (α_L), 3_{10} (3), et π (π) sont indiquées, ainsi que les régions correspondant aux feuillets β parallèles ($\uparrow\uparrow$) et *anti*-parallèles ($\uparrow\downarrow$).

1. En raison de la gêne stérique entre les chaînes latérales dans la conformation *cis*, la conformation *trans* ($\omega = 180^\circ$) est la forme majoritaire.

Hélices

La structuration en hélice α (figure A.4 A) du brin peptidique est stabilisée par la formation de liaisons hydrogène entre l'atome d'oxygène du groupement carbonyle du résidu i et le groupement $N-H$ du résidu $(i+4)$. Cette structuration en hélice α correspond à une région réduite du diagramme de Ramachandran, centrée autour des valeurs ($\phi = -57^\circ$; $\psi = -47^\circ$). Un tour d'hélice contient 3,6 résidus et son pas est de 5,4 Å.

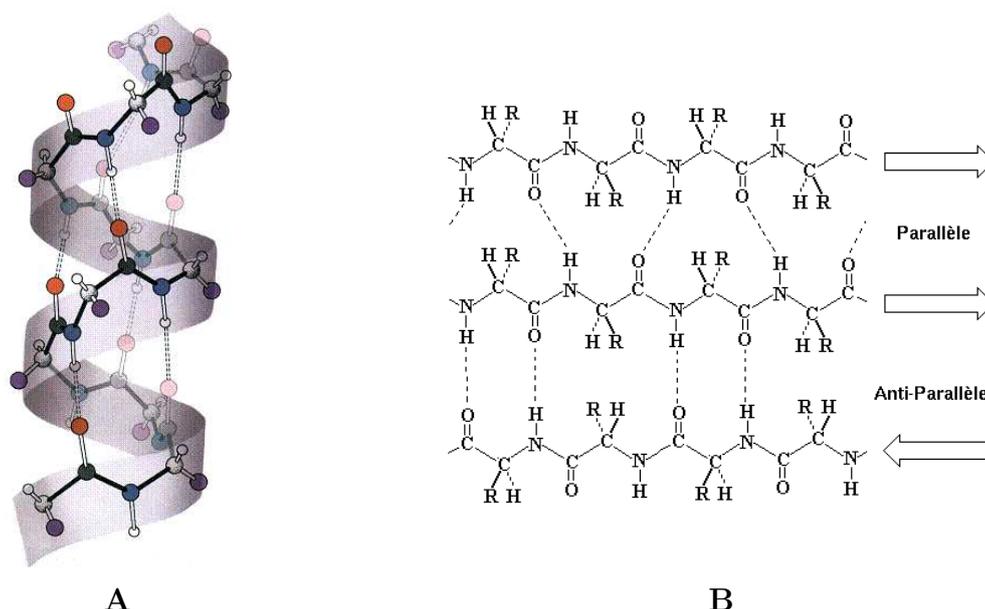


FIGURE A.4 – Structures secondaires dans une protéine. **A** : Hélice α ; **B** : Feuille β composé de trois brins. Le brin supérieur est parallèle au brin du milieu, alors que le brin inférieur est *anti*-parallèle au brin du milieu.

Si l'hélice α (figure A.5 A) est la forme la plus courante d'hélices, d'autres types d'hélices existent également qui diffèrent par le réseau de liaisons hydrogène formé. Dans l'hélice 3_{10} (figure A.5 B), les liaisons hydrogène impliquent l'oxygène du résidu i et le groupement $N-H$ du résidu $(i+3)$, ce qui engendre un rétrécissement du rayon de l'hélice, couplé à un allongement du pas. Au contraire, l'hélice π (figure A.5 C) est caractérisée par des liaisons hydrogène entre le résidu i et le résidu $(i+5)$, conduisant à une hélice plus élargie et un pas plus court. Dans les trois formes d'hélice, les chaînes latérales pointent vers l'extérieur.

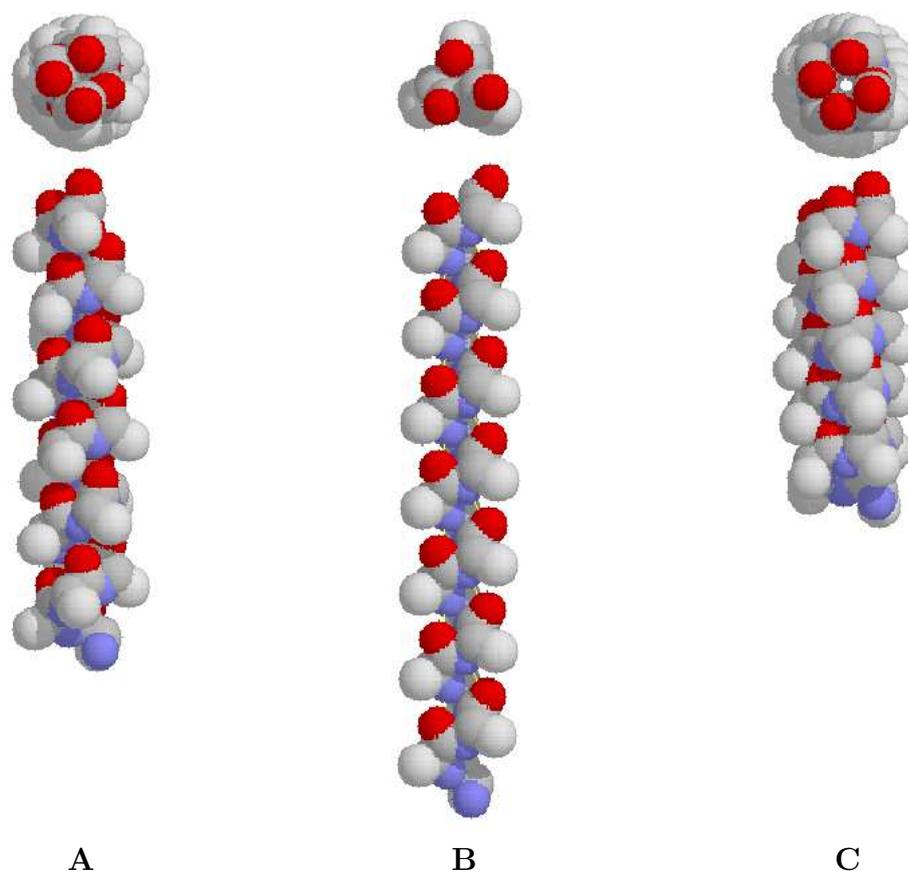


FIGURE A.5 – Différents types d’hélices en représentation de van der Waals, les atomes de carbone étant colorés en gris, les atomes d’azote en bleu, les atomes d’oxygène en rouge, et les atomes d’hydrogène en blanc. Les schémas supérieurs montrent une vue de dessus qui permet de mesurer le rayon de chacun des types d’hélices. **A** : Hélice α ; **B** : Hélice 3_{10} ; **C** : Hélice π .

Brins et feuillets β

La deuxième classe majeure de structures secondaires dans les protéines correspond aux feuillets β . Un feuillet β est un assemblage de brins β , reliés par un réseau de liaisons hydrogène entre les groupements carbonyle d’un brin et les groupements $N-H$ de l’autre. L’agencement de ces brins, caractérisés par une conformation de la chaîne peptidique correspondant à des couples ($-139^\circ \leq \phi \leq -119^\circ$; $113^\circ \leq \psi \leq 135^\circ$) peut se faire de manière parallèle ou *anti*-parallèle, selon le sens de chacun des brins, permettant la formation d’un réseau plus ou moins régulier de liaisons hydrogène (voir figure A.4 B). Dans les deux cas, le feuillet acquiert une forme plissée et les chaînes latérales pointent alternativement d’un côté ou de l’autre du plan médian.

Autres éléments de structure secondaire

Les hélices et les feuillets constituent en moyenne 50 % de l'architecture protéique. Le reste de la chaîne peptidique est soit dans un état désordonné (pelote statistique ou "*random coil*"), soit forme de petits éléments structurés. Parmi ceux-ci, on peut citer les coudes β (voir figure A.6), caractérisés par l'existence d'une liaison hydrogène unique entre les résidus i et $(i + 3)$, qui relie fréquemment deux éléments de structure secondaire répétitive (hélices ou feuillets).

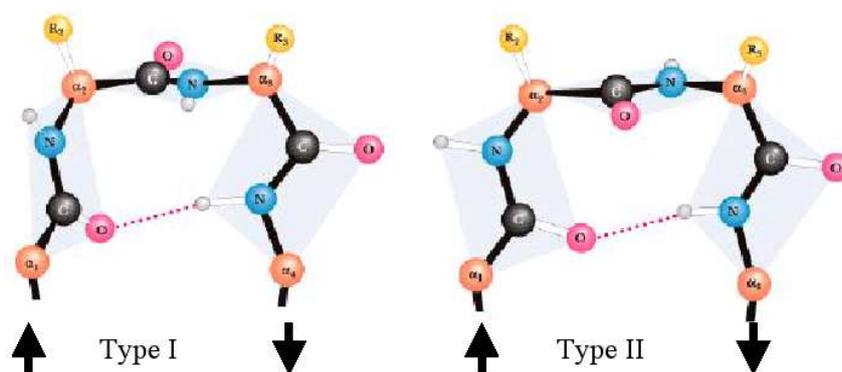


FIGURE A.6 – Schéma de coudes β de types I (à gauche) et II (à droite).

A.3 Structures tertiaire et quaternaire

L'agencement spatial relatif des différentes structures secondaires de la protéine, reliées entre elles par des connexions plus ou moins ordonnées va déterminer la structure tertiaire de la protéine. Cette structure est stabilisée par différentes interactions :

- les ponts disulfures, qui sont des liaisons covalentes créées par oxydation de deux cystéines.
- les liaisons hydrogène, qui peuvent être réalisées par les groupements des chaînes latérales ou du brin peptidique.
- les ponts salins, formés entre les chaînes latérales d'un résidu basique (positivement chargé) et d'un résidu acide (négativement chargé).
- les interactions hydrophobes, qui correspondent à l'enfouissement des chaînes latérales hydrophobes, afin de minimiser la perte entropique due à leur solvatation.

Souvent, la forme biologiquement active d'une protéine n'est pas constituée d'une seule chaîne peptidique. Dans ce cas, les différentes chaînes (appelées alors monomères ou sous-unités) s'agencent dans l'espace pour former un complexe macromoléculaire. Ceci constitue le niveau le plus élevé d'organisation des protéines, dénommé structure quaternaire.

Annexe B

Champs de force et modèles de représentation

B.1 Définition du champ de force

La taille des systèmes biologiques étudiés en simulation, qui comportent typiquement plusieurs milliers à centaines de milliers d'atomes, rend impossible toute modélisation à l'échelle quantique. Les atomes sont donc considérés comme des particules ponctuelles chargées, dont le mouvement obéit aux lois de la mécanique classique. Il est à noter que ceci rend impossible l'étude des processus chimiques, puisque la formation ou la rupture de liaisons nécessitent la prise en compte des effets électroniques.

Le terme "champ de force" désigne l'ensemble des données nécessaires au calcul de l'énergie d'un tel système. En d'autres termes, il contient l'ensemble des fonctions mathématiques correspondant aux différentes interactions entre les atomes ainsi que les paramètres de ces fonctions (qui dépendent du type des atomes).

B.2 Modèles "tout-atome"

Lorsqu'on souhaite représenter une molécule avec un maximum de précision en mécanique moléculaire, on utilise une représentation dite "tout-atome" dans laquelle tous les atomes sont représentés. Différents champs de force existent pour calculer l'énergie d'un système dans cette représentation. Le champ de force "tout-atome" qui a été utilisé au cours de ce travail est le champ de force intégré dans la suite de programme AMBER (Case *et al.*, 2002 ; Case *et al.*, 2004), avec les paramètres parm99 (Wang *et al.*, 2000).

L'expression mathématique de l'énergie potentielle E_p associée est caractéristique des champs de force "tout-atome" couramment utilisés aujourd'hui :

$$E_p = \sum_{\text{liaisons}} K_r (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dièdres}} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ + \sum_{i < j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i < j} \left(\frac{q_i q_j}{\epsilon r_{ij}} \right)$$

Les différents termes de l'énergie potentielle représentent les différentes interactions modélisées par ce champ de force :

$$- E_{\text{liaisons}} = \sum_{\text{liaisons}} K_r (r - r_0)^2$$

Ce terme représente l'énergie de déformation des liaisons. Les énergies de liaison sont modélisées par un potentiel harmonique : la longueur r de la liaison fluctue autour de la longueur de référence r_0 . K_r représente la constante de force associée à la liaison et est d'autant plus grande que la liaison est forte. Ces paramètres sont optimisés pour reproduire les données spectroscopiques.

$$- E_{\text{angles}} = \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2$$

Les variations d'énergie associées à une déformation des angles de valence sont également modélisées par un potentiel harmonique autour d'une valeur θ_0 de référence.

$$- E_{\text{dièdres}} = \sum_{\text{dièdres}} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]$$

La variation d'énergie potentielle due à la rotation autour d'une liaison (mesurée par l'angle dièdre ϕ) est modélisée par une somme de termes sinusoïdaux dont la périodicité n et la phase γ dépendent de l'état de substitution des deux atomes engagés dans la liaison. Ce terme tient également en partie compte des interactions entre atomes non liés, qui seront donc pondérées.

$$- E_{\text{LJ}} = \sum_{i < j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right)$$

Ce terme représente les interactions non liées de type van der Waals (en r_{ij}^{-6}) attractives et de répulsion d'échange (en r_{ij}^{-12}). Les paramètres A_{ij} et B_{ij} sont déterminés de façon empirique et optimisés pour reproduire des données expérimentales (comme des enthalpies de solvation pour les ions (Åqvist, 1990)).

$$- E_{\text{elec}} = \sum_{i < j} \left(\frac{q_i q_j}{\epsilon r_{ij}} \right)$$

Les interactions électrostatiques entre les atomes i et j sont modélisées par un potentiel coulombien, dont les paramètres sont déterminés selon la procédure RESP (Bayly *et al.*, 1993 ; Cornell *et al.*, 1993 ; Cieplak *et al.*, 1995). Dans cette procédure, les charges partielles q_i et q_j des atomes i et j sont optimisées pour reproduire le potentiel électrostatique moléculaire obtenu à l'aide de calculs de chimie quantique.

B.3 Représentation "gros grains" et réseaux élastiques

Représenter tous les atomes d'une protéine afin d'étudier sa dynamique n'est pas toujours nécessaire. En effet, il a été démontré que les modes de vibration correspondant à des déplacements de grande amplitude résultent de fluctuations concertées des C_α . De plus, les modes normaux de basse fréquence (voir annexe E) peuvent être reproduits à l'aide d'un champ de force très simple dans lequel les interactions entre deux atomes sont représentées par un ressort dont la constante de force est indépendante du type des atomes impliqués (Tirion, 1996). De ces deux observations, il ressort qu'il est possible de reproduire la dynamique globale d'une protéine à l'aide d'une représentation très simple dans laquelle un résidu est représenté par un nombre de sites réduit, couplée à un champ de force harmonique.

Ces représentations "gros grains" ont été développées pour les protéines dans les années 1990. Dans le modèle le plus simple, chaque acide aminé est représenté par un site (placé en général au niveau du C_α) et un ressort (de constante de force unique¹) est placé entre deux sites si la distance les séparant est inférieure à une valeur donnée (typiquement, cette valeur est comprise entre 7 et 15 Å). Il est à noter que ce modèle ne tient pas compte de la connectivité des résidus au sein de la protéine. De plus, dans la structure initiale, tous les ressorts sont relâchés, ce qui fait que le système est alors à l'équilibre². Un exemple du réseau obtenu est donné figure B.1. Cette représentation initialement proposée par Haliloglu et collaborateurs (Haliloglu *et al.*, 1997) a connu de grands succès notamment grâce à son utilisation pour le calcul des modes normaux (Bahar *et al.*, 1997 ; Keskin *et al.*, 2000 ; Tama et Sanejouand, 2001 ; Keskin, 2002). Quelques raffinements peuvent être apportés au modèle comme une variation de la constante de force des ressorts en fonction de la distance initiale entre les deux sites (Hinsen, 1998) ou une augmentation du nombre de sites par résidus (de un à trois selon le type de résidu) pour rendre compte grossièrement de la nature de ces résidus (Zacharias, 2003 ; Sacquin-Mora et Lavery, 2006).

L'utilisation d'un modèle granulaire couplé à un réseau de ressorts présente deux intérêts :

- un gain de temps de calcul par rapport aux simulations "tout-atome".
- la possibilité d'étudier la dynamique de très gros systèmes.

1. Le fait que cette valeur soit identique pour tous les ressorts permet de l'ajuster *a posteriori* en comparant les résultats obtenus avec les données expérimentales.

2. Ceci est un avantage certain pour l'utilisation de ce modèle avec la technique des modes normaux (voir annexe E).

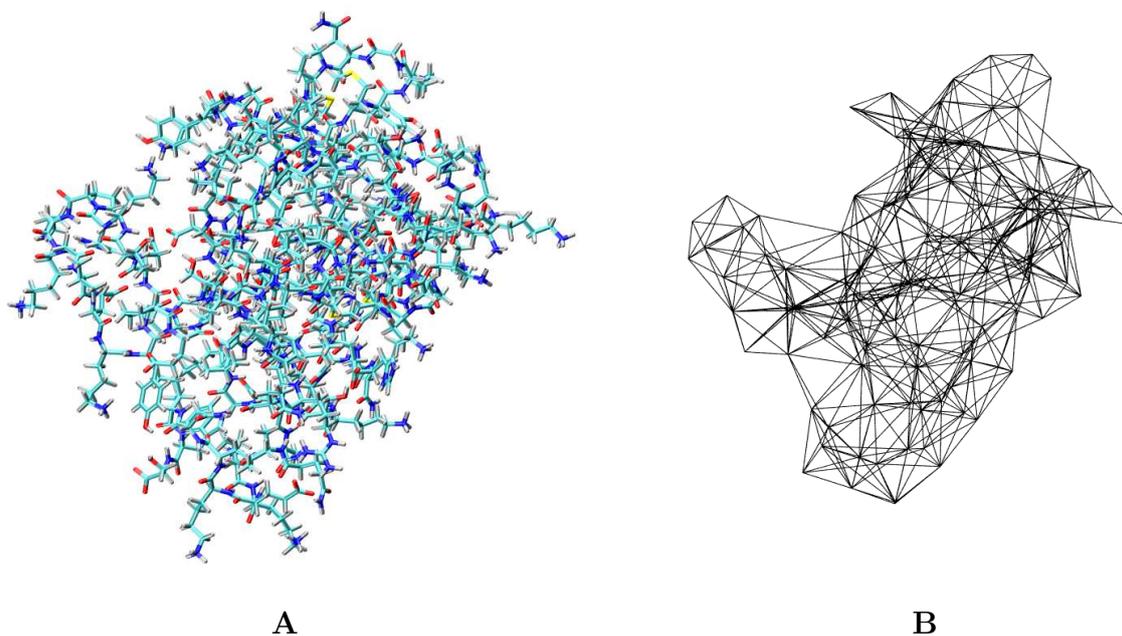


FIGURE B.1 – Représentation des protéines sur l'exemple de la nucléase de staphylocoque. **A** : représentation "tout-atome" ; **B** : représentation granulaire. Les lignes représentent les ressorts reliant deux C_α quand ceux-ci sont distants de moins de 8 Å.

Cependant, ce modèle présente également quelques désavantages par rapport à la représentation "tout-atome" :

- impossibilité d'étudier des phénomènes fins qui dépendent de la nature des résidus mis en jeu, comme l'effet de mutations.
- limitation des mouvements possibles : l'ensemble des connectivités qui sont créées par le réseau de ressort limite en effet le déplacement des atomes à de petites fluctuations autour de leur position initiale.

Annexe C

Minimisation d'énergie

C.1 Surface d'énergie potentielle

L'énergie potentielle V d'une molécule est une fonction des M coordonnées x_i de la molécule. Les valeurs de cette fonction définissent donc une hypersurface (appelée aussi "surface" par simplification) de dimension M . Cette surface est caractérisée par un ensemble de minima et de maxima, représentés dans un cas simple à trois dimensions sur la figure C.1. Les minima correspondent à des structures stables, tandis que les maxima sont des états de transition. Pour passer d'un minimum à un autre, il est nécessaire de passer par un état de transition en apportant l'énergie correspondante à la différence d'énergie potentielle, souvent appelée barrière de potentiel, entre cet état de transition et l'état stable.

Dans le cas des molécules biologiques comme les protéines, la surface de potentiel est très accidentée et le nombre de minima locaux est très important. La recherche du minimum global, qui représente la structure la plus stable de la protéine, est donc une tâche ardue, qui nécessite l'utilisation de méthodes mathématiques sophistiquées.

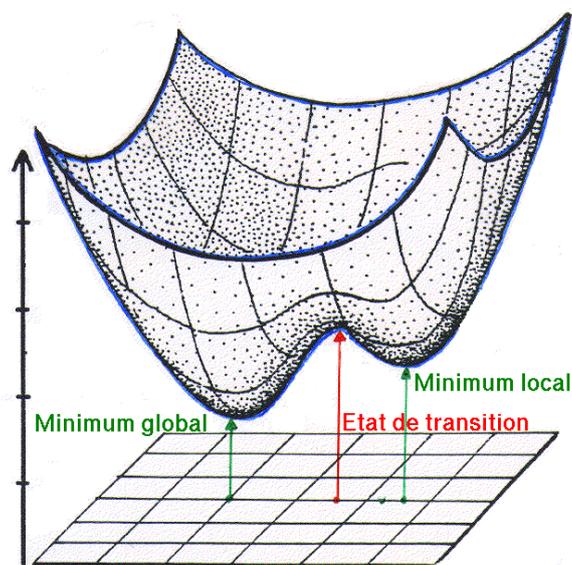


FIGURE C.1 – Schéma d'une surface de potentiel simple présentant un minimum local, un minimum global, et un état de transition.

C.2 Principe de la minimisation d'énergie

La minimisation d'énergie est un outil indispensable en simulation moléculaire. Elle est en effet utilisée de manière isolée pour déterminer les structures les plus stables, mais également pour préparer les systèmes avant l'application d'autres méthodes comme la dynamique moléculaire ou les modes normaux.

Lorsque la minimisation est utilisée en combinaison avec les modes normaux, le but recherché est de s'approcher le plus possible du minimum énergétique, qui est défini par un gradient d'énergie nul. Il sera donc nécessaire de réaliser des minimisations très poussées. En revanche, le but d'une minimisation peut simplement être de réduire les mauvais contacts inter-atomiques, lorsqu'elle est utilisée en préambule de simulations de dynamique moléculaire. Dans tous les cas, la minimisation est stoppée lorsqu'un critère de convergence est atteint. Ce critère peut porter sur le gradient (qui doit être inférieur à une valeur seuil), le déplacement lors de la dernière itération (qui doit être suffisamment petit), sur la variation d'énergie entre les deux dernières itérations (qui doit être inférieure à une valeur choisie par l'utilisateur), ou encore le nombre d'itérations.

La minimisation d'énergie est fondée sur les méthodes de recherche de minima de fonctions en mathématiques (dans notre cas, la fonction est celle du champ de force et les

variables sont les coordonnées). Trois grands types de méthodes (gradient simple, gradient conjugué, Newton-Raphson et quasi-Newton) sont utilisés en modélisation moléculaire. Toutes trois procèdent de façon similaire pour modifier les variables, selon l'équation :

$$\vec{r}(k+1) = \vec{r}(k) + \alpha(k) \cdot \vec{S}(k)$$

où $\vec{r}(k)$ représente le vecteur des coordonnées pour l'itération k , $\alpha(k)$ est le pas de déplacement et $\vec{S}(k)$ est la direction de déplacement. Les différences entre les différentes méthodes reposent dans les définitions de $\alpha(k)$ et $\vec{S}(k)$.

C.3 Les principales méthodes de minimisation

Méthode du gradient simple

Dans cette méthode la plus simple, la direction de déplacement est opposée à celle du gradient $\vec{g}(k)$ de la fonction énergétique :

$$\vec{S}(k) = -\frac{\vec{g}(k)}{\|\vec{g}(k)\|}$$

Le déplacement se fait donc selon la direction pour laquelle la variation d'énergie est la plus importante. Le pas $\alpha(k)$ est ajusté à chaque itération : si l'énergie diminue entre deux itérations le pas est augmenté, tandis qu'il est diminué si l'énergie augmente.

La méthode du gradient simple est très performante lors des premières itérations, quand le système est éloigné du minimum énergétique. En revanche, elle adopte un comportement oscillatoire au fur et à mesure que l'on s'approche de ce minimum. C'est pourquoi cette méthode est souvent utilisée lors des premières étapes d'une minimisation puis remplacée par une autre méthode plus performante, comme la méthode du gradient conjugué.

Méthode du gradient conjugué

Dans la méthode du gradient conjugué, la direction de déplacement n'est plus fonction uniquement du gradient à l'itération k , mais également du gradient à l'itération $(k-1)$ précédente :

$$\vec{S}(k) = -\vec{g}(k) + \gamma_k \vec{S}(k-1)$$

avec (dans la méthode de Fletcher-Reeves) :

$$\gamma_k = \frac{\|\vec{g}(k)\|^2}{\|\vec{g}(k-1)\|^2}$$

Lors des minimisations précédant les simulations de dynamique moléculaire, les deux méthodes du gradient sont souvent utilisées l'une après l'autre : la méthode du gradient simple permet de s'approcher rapidement du minimum, puis la méthode du gradient conjugué est appliquée pour affiner la minimisation.

Méthodes Newton-Raphson et quasi-Newton

Les méthodes de minimisation du gradient n'utilisent que la dérivée première de l'énergie pour définir la direction de déplacement. La méthode Newton-Raphson utilise de plus les dérivées secondes de l'énergie :

$$\vec{S}(k) = -H^{-1}(k)\vec{g}(k)$$

$\vec{g}(k)$ représente toujours le gradient de la fonction énergétique à l'itération k et H^{-1} est l'inverse de la matrice hessienne H , qui est la matrice des dérivées secondes de l'énergie potentielle V , dont les éléments pour l'itération k sont donnés par :

$$[H]_{ij}(k) = \frac{\partial V(k)}{\partial r_i \partial r_j}$$

Le calcul de la matrice hessienne pour les systèmes de grande taille est un facteur limitant de cette méthode. Afin de contourner ce problème, la matrice hessienne peut être approchée à chaque itération. La méthode est alors dénommée quasi-Newton.

Les méthodes Newton-Raphson et quasi-Newton sont très performantes lorsque l'on s'approche des minima d'énergie potentielle, mais peuvent être instables loin de ces minima. Ce sont donc des méthodes de choix pour raffiner une minimisation après avoir au préalable déjà minimisé l'énergie de la structure par une méthode du gradient (simple et/ou conjugué).

C.4 Minimisation en coordonnées internes, le programme LIGAND

Les coordonnées internes

La conformation d'une molécule est repérée par la position dans l'espace de chacun des atomes qui la compose. Il existe deux principaux jeux de coordonnées qui permettent de traduire cette position :

1. Les coordonnées cartésiennes.
2. Les coordonnées internes.

Dans un système de coordonnées cartésiennes, les atomes sont définis par trois coordonnées (x, y, z) relativement à un repère fixe dans l'espace. Ce système particulièrement simple est le plus fréquemment utilisé en mécanique moléculaire.

Dans un système de coordonnées internes, un atome est repéré par rapport aux autres atomes du système. Plus précisément, il est nécessaire de préciser une distance, un angle de valence, et un angle de torsion. Dans l'exemple de la figure C.2, la position de l'atome 4 est ainsi définie par :

- la distance d_{34} de la liaison 3 – 4.
- l'angle de valence A_{234} entre les liaisons 2 – 3 et 3 – 4
- l'angle de torsion T_{1234} autour de la liaison 2 – 3.

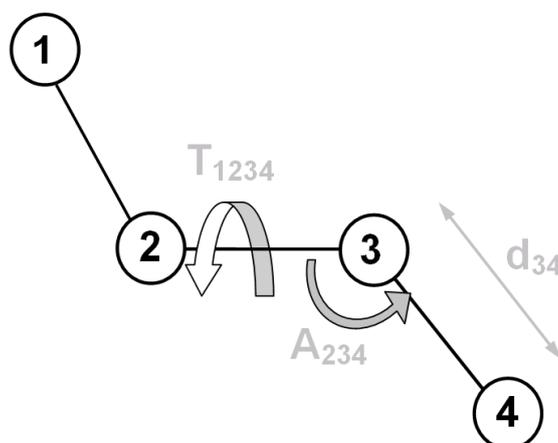


FIGURE C.2 – Définition des coordonnées internes.

Les coordonnées internes sont fréquemment utilisées pour préciser la géométrie des molécules dans les programmes de mécanique quantique mais peu en mécanique moléculaire. Pourtant, cette représentation présente plusieurs avantages :

- les coordonnées internes sont les variables naturelles pour décrire une molécule. En effet, la modification de la valeur d'une seule distance ou d'un seul angle nécessite de faire varier simultanément une combinaison bien précise d'un nombre important de coordonnées cartésiennes, tandis qu'il suffit d'une seule coordonnée interne pour effectuer la même opération. Ceci permet notamment de rendre les minimisations en coordonnées internes plus efficaces que les minimisations en coordonnées cartésiennes.
- il est facile de contraindre une distance ou un angle dans ce système de coordonnées. Ceci se révèle très utile pour minimiser les temps de calcul. En effet, les distances et certains angles de valence peuvent être considérés comme fixes. En coordonnées internes, contraindre ces paramètres revient à fixer la valeur d'une coordonnée, et donc le nombre de variables peut être divisé d'un facteur 10 environ. Ceci rend plus rapide une exploration de l'espace conformationnel, en lissant la surface d'énergie potentielle.

Les coordonnées internes peuvent être utilisées en combinaison avec différentes méthodologies comme les modes normaux (Kitao *et al.*, 1994 ; Duong et Zakrzewska, 1997), la dynamique moléculaire (Schwieters et Clore, 2001 ; Mazur, 2002), la minimisation d'énergie (Li et Scheraga, 1987 ; Navizet *et al.*, 2004a), ou les simulations de Monte Carlo (Fernández-Recio *et al.*, 2002 ; Bastard *et al.*, 2003).

Le programme LIGAND

LIGAND est un programme de modélisation en coordonnées internes dérivé du programme JUMNA (Lavery *et al.*, 1995). Dans ce programme, toutes les liaisons ainsi que les angles de valence des chaînes latérales sont fixés. Seuls les angles dièdres et les angles de la valence de la chaîne principale sont donc autorisés à varier, ce qui réduit le nombre de variables par un facteur 10 environ. Le programme LIGAND offre le choix entre différentes méthodes de modélisation : la minimisation d'énergie (avec un minimiseur de type quasi-Newton), le calcul des modes normaux, ou la technique de Monte-Carlo. Il est spécialement adapté à l'étude des biomolécules (ADN, ARN, et protéines), et permet aussi l'inclusion de molécules d'eau ou d'ions. Initialement conçu pour être utilisé en combinaison avec le champ de force FLEX (Lavery *et al.*, 1986), LIGAND peut également être utilisé avec le champ de force AMBER (Flatters *et al.*, 1997), tandis que le solvant est traité de manière implicite par l'utilisation d'une fonction diélectrique sigmoïdale ou par le modèle Born généralisé (voir annexe F).

L'une des applications privilégiées de LIGAND est la minimisation d'énergie en présence de contraintes. De nombreuses contraintes (sur les distances, les angles, les torsions, le rayon de giration...) sont déjà programmées et l'ajout de nouvelles contraintes est relativement aisé : il suffit pour cela d'ajouter un terme de pénalité dans le champ de force, ainsi que l'expression analytique des forces dérivées de ce nouveau potentiel. Nous avons utilisé LIGAND pour réaliser nos études de minimisation sous contrainte présentées dans le chapitre 3 en introduisant notre nouvelle contrainte sur la distance moyenne entre un résidu et les autres résidus de la protéine dans le programme LIGAND. Pour effectuer nos simulations, nous avons utilisé le champ de force AMBER avec les paramètres parm99 et utilisé le modèle de Born généralisé pour reproduire les effets électrostatiques du solvant.

Annexe D

Dynamique moléculaire

D.1 Principe de la dynamique moléculaire

La dynamique moléculaire est une méthode très utilisée pour étudier les systèmes biologiques en modélisation. Elle consiste à intégrer les équations du mouvement de la mécanique classique afin de déterminer la trajectoire des atomes du système et de suivre ainsi son évolution.

L'équation fondamentale de la dynamique moléculaire provient de la troisième loi de Newton reliant l'accélération \vec{a}_i d'une particule i aux forces \vec{F}_i qui lui sont appliquées :

$$m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$$

où r_i représentent les coordonnées de l'atome i et m_i sa masse. Les forces sont reliées à l'énergie potentielle V du système (définie par le champ de force) par la relation :

$$\vec{F}_i = - \frac{\partial V}{\partial \vec{r}_i}$$

La présence de plusieurs particules dans le système ne permettant pas une intégration analytique de l'équation du mouvement, l'évolution des coordonnées des atomes est calculée par intégration numérique. Plusieurs algorithmes ont été développés pour réaliser cette intégration (Verlet, leap-Frog, Beeman...), qui sont tous fondés sur un développement en série de Taylor des coordonnées :

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \delta t \frac{d\vec{r}_i}{dt} + \frac{\delta t^2}{2} \frac{d^2 \vec{r}_i}{dt^2} + \dots$$

et de même :

$$\vec{r}_i(t - \delta t) = \vec{r}_i(t) - \delta t \frac{d\vec{r}_i}{dt} + \frac{\delta t^2}{2} \frac{d^2\vec{r}_i}{d^2t} + \dots$$

En soustrayant les deux équations ci-dessus et en limitant le développement à l'ordre 2, on obtient l'expression de la vitesse \vec{v}_i à l'instant t en fonction des coordonnées aux instants $(t - \delta t)$ et $(t + \delta t)$:

$$\vec{v}_i(t) = \frac{d\vec{r}_i}{dt} = \frac{\vec{r}_i(t + \delta t) - \vec{r}_i(t - \delta t)}{2\delta t}$$

En les additionnant, on obtient l'expression des coordonnées à l'instant $(t + \delta t)$ en fonction des coordonnées aux instants $(t - \delta t)$ et t et de l'accélération :

$$\vec{r}_i(t + \delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \delta t) + \delta t^2 \frac{d^2\vec{r}_i}{d^2t}$$

Ces deux dernières équations permettent de définir la position des atomes dans le système à l'instant $(t + \delta t)$ en utilisant les coordonnées des deux itérations précédentes et l'accélération au temps t . Elles définissent l'algorithme de Verlet qui est utilisé dans le logiciel AMBER pour déterminer la trajectoire des particules. L'initialisation de l'algorithme se fait par l'attribution aléatoire de vitesses à chacun des atomes du système, l'ensemble des vitesses attribuées suivant une loi de Maxwell-Boltzmann autour d'une valeur correspondant à la température désirée pour le système¹.

Le choix du pas d'intégration δt n'est pas anodin. S'il est trop petit, alors la trajectoire ne couvrira qu'un intervalle de temps réduit en raison des limites informatiques. En revanche, s'il est trop important, il peut amener à des instabilités dans le système, du fait de collisions entre atomes qui font diverger l'énergie. En règle générale, on considère que δt doit être au moins dix fois inférieur au temps caractéristique du mouvement le plus rapide du système étudié. Pour l'étude des macromolécules biologiques avec un champ de force "tout-atome", celui-ci correspond à la vibration des liaisons $X - H$ impliquant un atome d'hydrogène. Le temps caractéristique de ces vibrations étant de l'ordre de 10 fs, le pas d'intégration doit être au maximum de 1 fs. Cependant, il est souvent raisonnable de considérer la longueur des liaisons $X - H$ comme fixe, ce qui permet d'augmenter le pas d'intégration. C'est ce que nous avons réalisé dans la majorité de nos simulations de dynamique moléculaire au cours desquelles la longueur des liaisons $X - H$ est contrainte à l'aide de la procédure SHAKE (Ryckaert *et al.*, 1977), permettant l'utilisation d'un pas d'intégration de 2 fs.

1. La température T d'un système est en effet reliée à son énergie cinétique K , et donc aux vitesses des particules. Si N_l désigne le nombre de degrés de liberté du système (typiquement $3N - 6$ pour un système comprenant N atomes non contraints), on a en effet : $K = \sum_i \frac{1}{2} v_i^2 = N_l \frac{k_B T}{2}$.

D.2 Troncatures du potentiel

Pour les grands systèmes, il est très pénalisant de devoir calculer toutes les interactions non liées, dont le nombre est proportionnel à N^2 , où N représente le nombre d'atomes du système. Afin de limiter les temps de calcul, il est intéressant de ne calculer ces interactions que pour des atomes relativement proches. En d'autres termes, on définit une distance de coupure r_c au-delà de laquelle les interactions non liées (électrostatiques et de Lennard-Jones) ne sont plus calculées.

Si cette approximation est acceptable pour les interactions de Lennard-Jones qui diminuent rapidement avec la distance (en r^{-6} pour la partie attractive et r^{-12} pour la partie répulsive), elle engendre une forte discontinuité pour le potentiel électrostatique dont la décroissance est plus lente (en r^{-1}). Il est possible de surmonter cette difficulté en modifiant la fonction de potentiel afin de l'annuler, ainsi que la force qui en dérive, pour la valeur du rayon de coupure avec une fonction de "shift". Une méthode simple pour y parvenir est de multiplier le potentiel par une fonction dont la valeur et la dérivée sont nulles pour $r = r_c$. On peut ainsi par exemple exprimer le nouveau potentiel comme :

$$\begin{aligned} V_{shift}(r) &= V(r) \left(1 - \left(\frac{r}{r_c}\right)^2\right)^2 && \text{pour } r \leq r_c \\ V_{shift}(r) &= 0 && \text{pour } r \geq r_c \end{aligned}$$

La figure D.1 A montre un exemple de fonction "shift" sur le potentiel électrostatique avec une distance de coupure $r_c = 15$. Cette expression permet de résoudre le problème de discontinuité du potentiel et de la force en $r = r_c$ mais le potentiel obtenu dévie du potentiel "réel" sur l'ensemble des valeurs. Pour éviter cela, on peut utiliser une fonction de "switch", c'est-à-dire une fonction multiplicative qui n'a d'effet que sur une plage de valeurs limitée. Un exemple de l'effet d'une fonction de "switch" sur le potentiel électrostatique est montré figure D.1 B. La fonction de "switch" S utilisée dans cet exemple a pour expression :

$$\begin{aligned} S(r) &= 1 && \text{pour } r \leq r_u = 13 \\ S(r) &= 1 - 10 \left(\frac{r - r_u}{r_c - r_u}\right)^3 + 15 \left(\frac{r - r_u}{r_c - r_u}\right)^4 - 6 \left(\frac{r - r_u}{r_c - r_u}\right)^5 && \text{pour } r_u \leq r \leq r_c = 15 \\ S(r) &= 0 && \text{pour } r \geq r_c \end{aligned}$$

Les fonctions de "switch" introduisent cependant des forces non physiques pour $r_u \leq r \leq r_c$ (voir la modification de la pente de la courbe dans cette portion sur la figure D.1 B).

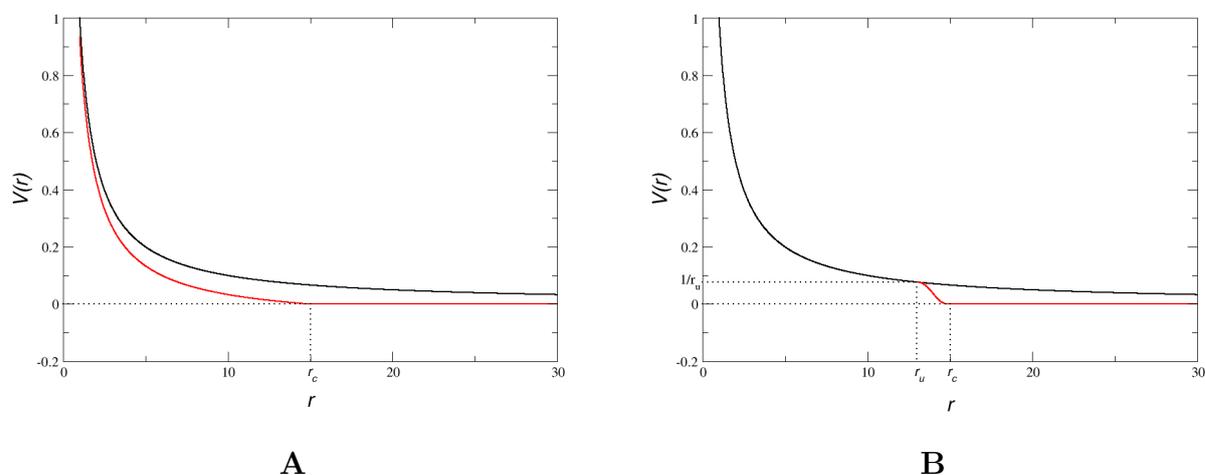


FIGURE D.1 – Exemples de troncature du potentiel électrostatique par des fonctions de "shift" (A) et de "switch" (B). La courbe noire correspond au potentiel coulombien et les courbes rouges aux potentiels tronqués.

Une autre façon de régler ce problème de discontinuité du potentiel due à la troncature est d'utiliser des conditions périodiques et de la sommation d'Ewald (voir section D.3) pour calculer les interactions électrostatiques.

D.3 Conditions aux limites

Les interactions entre le solvant et les macromolécules biologiques jouent souvent un rôle primordial pour la fonction de ces dernières. C'est pourquoi il est indispensable de pouvoir prendre en compte les effets du solvant lors des simulations de dynamique moléculaire. Pour ce faire, il est possible d'utiliser une représentation implicite ou explicite du solvant (voir annexe F).

Dans nos simulations de dynamique moléculaire, nous avons généralement utilisé une représentation explicite du solvant. Dans ces conditions, le soluté est inséré dans une boîte de molécules d'eau. Le système ne pouvant pas être de dimension infinie, on observe des effets de bord, c'est-à-dire que les molécules situées à proximité de l'interface avec le vide sont soumises à des perturbations. Il convient de tenir compte de ces effets pour qu'ils ne perturbent pas la dynamique de la molécule d'intérêt. Ainsi, différentes mesures peuvent être prises suivant les conditions aux limites imposées.

Conditions non-périodiques

Lors de l'utilisation de conditions non-périodiques, il est nécessaire d'ajouter plusieurs couches de solvant autour de la molécule, de telle manière que celle-ci ne soit pas en contact avec le vide. En d'autres termes, le bord de la boîte d'eau ainsi créée doit se situer au minimum à une distance supérieure à la distance de coupure des interactions non liées de tout atome du soluté. Les effets de bord avec le vide impliquent ainsi les molécules d'eau et non le soluté, réduisant l'effet de cette interface sur la molécule d'intérêt, et ce d'autant plus que les couches de solvation sont nombreuses.

Conditions périodiques

Dans l'idéal, le soluté devrait être immergé dans une cellule de dimension quasi-infinie. L'utilisation des conditions périodiques permet de se rapprocher de cette situation et de créer un réseau de grande dimension en utilisant pourtant un nombre réduit d'atomes.

Le principe de cette méthode est de dupliquer la cellule de simulation dans l'espace pour créer un système quasi-cristallin, dont chaque maille est identique à la maille centrale (la figure D.2 montre un exemple de conditions périodiques appliquées à un système à deux dimensions). Pour cela, la cellule de simulation doit avoir une géométrie bien particulière, permettant d'occuper tout l'espace par duplication dans les trois directions. En règle générale, cette géométrie est choisie en fonction de la propre forme du soluté afin de minimiser le nombre de molécules d'eau à ajouter au système : pour un système de géométrie allongée, on utilisera une cellule parallélépipédique, tandis que pour un système de géométrie sphérique, une cellule cubique ou octaédrique tronquée sera préférée.

Au cours du mouvement, les positions des atomes de la cellule centrale sont copiées dans les autres cellules. Si une particule vient à sortir de la cellule de simulation, les conditions périodiques consistent à la faire rentrer par la face opposée de la cellule, permettant de garder le nombre d'atomes constant dans chaque maille (voir figure D.2). De plus, le calcul des interactions pour une particule de la maille centrale est réalisé en tenant compte de l'ensemble des voisins de cette particule, qu'ils appartiennent ou non à la même maille. Il faut toutefois s'assurer que le soluté ne soit pas en interaction avec son image dans une cellule voisine, en utilisant une taille de boîte suffisamment grande. En règle générale, on utilise une boîte dont la plus petite dimension est au moins égale à deux fois la distance de coupure des interactions non liées.

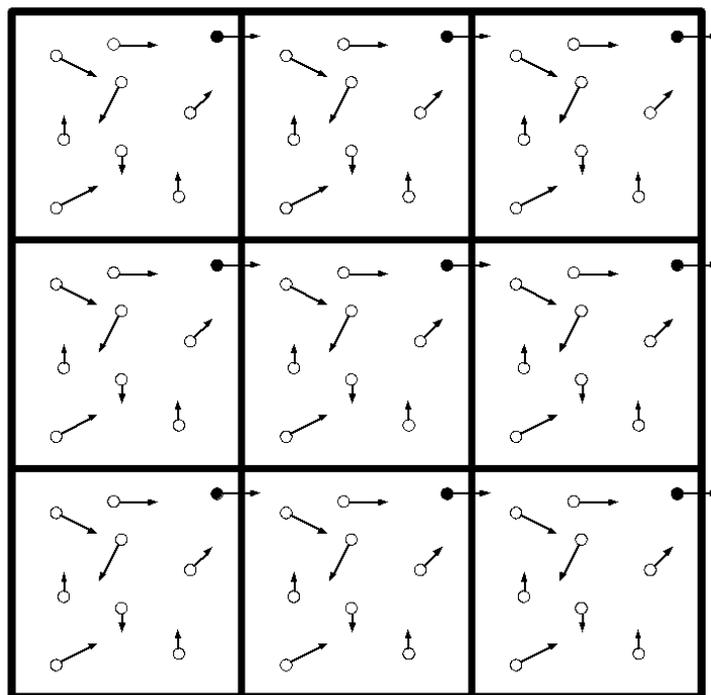


FIGURE D.2 – Exemple de conditions périodiques appliquées à un système en deux dimensions. La cellule centrale est la cellule de simulation. Les particules sont représentées par des cercles, et les flèches représentent leurs déplacements. Le déplacement de la particule noire de la maille centrale l'amène à sortir de la cellule. Elle est alors remplacée par son image dans la maille de gauche.

Lorsque la simulation est réalisée avec des conditions périodiques, il est possible d'utiliser la méthode dite de sommation d'Ewald (Allen et Tildesley, 1987) afin de calculer les interactions électrostatiques. Cette méthode consiste à décomposer la somme des interactions électrostatiques, qui converge lentement, en plusieurs termes dont la convergence est rapide. Pour cela, deux distributions de charges gaussiennes opposées et centrées sur les charges ponctuelles du système sont artificiellement ajoutées dans le calcul : la première "neutralise" les charges "réelles" et la seconde annule la précédente. Le calcul des interactions électrostatiques peut alors être réalisé en trois étapes, la première consistant à calculer les interactions des charges ponctuelles et de la distribution "neutralisante", la seconde consistant à calculer la contribution de la seconde distribution de charges. Ces deux termes ont une convergence rapide qui permet de rendre le calcul des interactions électrostatiques plus efficaces qu'une méthode "directe". La troisième étape consiste à calculer des termes correctifs qui tiennent compte de l'introduction des distributions gaussiennes fictives.

Le temps de calcul des sommes d'Ewald reste toujours proportionnel à N^2 , où N représente le nombre de charges. La méthode *Particle Mesh Ewald*, utilisée dans les modules *Sander* et *PMEMD* de la suite AMBER permet de calculer de façon rapide les sommes d'Ewald (Cheatham *et al.*, 1995) par l'utilisation d'une coupure des interactions non liées pour une valeur de r_c assez faible (de l'ordre de 8 à 10 Å) pour le calcul de la première somme, et par l'utilisation des méthodes de transformées de Fourier rapides pour la seconde, rendant le temps de calcul proportionnel à $N \ln N$.

D.4 Ensembles thermodynamiques

L'ensemble thermodynamique naturel associé à la méthode de la dynamique moléculaire est l'ensemble NVE dans lequel le nombre de particules N , le volume V et l'énergie E sont constants. En effet, l'ensemble des forces qui entrent en jeu dans le champ de force sont conservatives et l'énergie totale du système doit être conservée. Cependant, des fluctuations ou de faibles dérives de l'énergie sont couramment observées lors de simulations dans cet ensemble, causées par l'accumulation d'erreurs numériques ou par les approximations réalisées dans le calcul des interactions non liées (Allen et Tildesley, 1987 ; Leach, 2001 ; Frenkel et Smit, 2002). De plus, il ne correspond pas aux conditions expérimentales, qui sont réalisées dans les ensembles NVT (nombre de molécules, volume et température constants) ou NPT (nombre de molécules, pression et température constants). Dans toutes nos simulations réalisées en conditions périodiques (voir paragraphe D.3), nous sommes placés dans l'ensemble NPT .

L'utilisation de l'ensemble NPT permet de contrôler à la fois la pression et la température pour se placer dans des conditions qui reproduisent celles de la majorité des expérimentations. Dans nos simulations, la régulation de la température est généralement effectuée par couplage avec un bain thermique, en utilisant l'algorithme de Berendsen (Berendsen *et al.*, 1984). Avec ce couplage, la température T du système est reliée à la température T_0 du bain thermique par la relation :

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_T}$$

τ_T représente le temps de relaxation ou taux de couplage de la température. T est calculée à partir des vitesses v_i des particules du système :

$$T = \frac{\sum_i m_i v_i^2}{3Nk_B}$$

À chaque pas d'intégration, la température est alors contrôlée en ajustant les vitesses des atomes par un facteur multiplicatif λ donné par :

$$\lambda = \sqrt{1 + \frac{\delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right)}$$

La pression est également régulée dans nos simulations par un couplage avec un "bain de pression". Cette fois-ci, on a :

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P}$$

P représente la pression du système, P_0 celle du "bain de pression" et τ_P la constante de couplage de la pression. La pression est calculée à partir du théorème du Viriel et est donnée par :

$$P = \frac{2}{3V} \left(\sum_{i=1}^N \frac{1}{2} m_i v_i^2 - \sum_{i=1}^N \vec{r}_i \cdot \vec{F}_i \right)$$

où \vec{r}_i représentent les coordonnées de l'atome i , \vec{F}_i les forces qui lui sont appliquées, et m_i sa masse. De façon analogue au contrôle de la température, la régulation de la pression se fait à chaque pas d'intégration par multiplication du volume de la boîte par un facteur μ donné par :

$$\mu = 1 - \beta(P_0 - P) \frac{\delta t}{\tau_P}$$

où β est la compressibilité isotherme de l'eau : $\beta = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T = 44,75 \cdot 10^{-6} \text{ atm}^{-1}$.

D.5 Protocole

Une dynamique moléculaire nécessite la mise en place d'un protocole pour obtenir une simulation stable. Ce protocole peut être variable d'un système à l'autre mais comporte néanmoins certaines étapes indispensables :

1. Préparation de la structure initiale

La structure de la molécule est souvent extraite de la PDB. Il est ensuite nécessaire d'ajouter l'ensemble des atomes manquants (généralement les atomes d'hydrogène et certains atomes des chaînes latérales). Dans notre cas, ceci est effectué avec le module *TLeap* de la suite AMBER. La conformation ainsi obtenue contient certainement des gênes stériques qu'il est nécessaire de diminuer par une minimisation préalable. Si la simulation doit avoir lieu en présence d'eau explicite, c'est également au cours de cette première étape qu'il faut ajouter les molécules de solvant et les ions éventuels.

2. *Thermalisation*

Lors de l'initialisation d'une trajectoire de dynamique moléculaire, on attribue des vitesses aléatoires à chacun des atomes du système (voir section D.1), dont les valeurs dépendent de la température simulée. Pour ne pas introduire de trop brusques changements structuraux, il convient de débiter la simulation à une température assez faible puis d'augmenter progressivement cette température. Dans cette étape, il est également prudent de se placer dans l'ensemble NVT afin d'éviter une "explosion" du volume de la cellule. De plus, les atomes du soluté sont contraints à rester proches de leur position initiale afin d'éviter des changements conformationnels qui seraient dus à une mauvaise interface initiale avec le solvant.

3. *Équilibration*

Cette étape est indispensable lorsqu'un solvant explicite est employé afin d'équilibrer le solvant. En d'autres termes, il s'agit d'adapter la position des molécules d'eau et des ions à la conformation de la protéine. Cette étape est donc, comme la précédente, réalisée en appliquant des contraintes sur les atomes du soluté. Ces contraintes sont peu à peu diminuées pour permettre une adaptation de la conformation de la protéine.

4. *Production*

Cette phase correspond à la partie de la dynamique qui sera analysée (à partir des coordonnées qui sont sauvegardées à intervalles réguliers). Dans nos simulations, cette phase de production est réalisée dans des conditions de température et de pression constantes (ensemble NPT) lorsque des conditions périodiques ont été appliquées.

D.6 Limitations

La principale limitation de la dynamique moléculaire provient de son coût important en temps de calcul, notamment lorsqu'elle est utilisée en combinaison avec une représentation "tout-atome" (le tableau D.1 présente un ordre de grandeur des temps de calcul pour différents systèmes étudiés en dynamique moléculaire "tout-atome" et présentés dans ce manuscrit). Ainsi, les échelles de temps qui sont accessibles en dynamique moléculaire (aujourd'hui typiquement de l'ordre d'une dizaine à centaine de nanosecondes en fonction de la taille du système) sont réduites et incompatibles avec certains phénomènes biologiques, comme le repliement.

Système	Nombre d'atomes	Temps de calcul pour 100 ps de dynamique moléculaire (en heures)
Dimère <i>BE</i> de cadhérine	230000	320
Dimère <i>décalé</i> de cadhérine	150000	160
Dimère <i>BN</i> de cadhérine	85000	100
APRTase	45000	70
BPTI	25000	40

TABLE D.1 – Ordre de grandeur des temps de simulation pour des systèmes de taille variable et pour 100 ps de dynamique moléculaire réalisés en parallèle sur 16 processeurs SGI Origin 3800 au CINES. Toutes ces simulations ont été réalisées en conditions périodiques.

Ceci entraîne de plus une limitation de l'espace conformationnel échantillonné. Contrairement à la minimisation, la dynamique moléculaire devrait en principe permettre d'explorer plusieurs puits de potentiel en surpassant certaines barrières d'énergie. En pratique, le temps limité de la simulation rend peu probable l'exploration de multiples minima.

Annexe E

Modes normaux et analyse en composantes principales

E.1 Modes normaux

L'analyse des modes normaux consiste à considérer l'ensemble des fluctuations atomiques d'une molécule comme la superposition de modes de vibration harmoniques indépendants. Chacun de ces modes de vibration est caractérisé par une direction de déplacement (appelé mode normal, car les modes de vibration obtenus par cette méthode forment une base orthogonale de l'espace) et une fréquence.

Obtention des modes normaux

Si l'on note x_i les $3N$ coordonnées des atomes d'une molécule, et x_i^0 les coordonnées de référence, l'énergie potentielle peut être décomposée en série de Taylor :

$$V = V_0 + \sum_{j=1}^{3N} \left(\frac{\partial V}{\partial x_j} \right)_0 (x_j - x_j^0) + \frac{1}{2} \sum_{i,j=1}^{3N} \left(\frac{\partial^2 V}{\partial x_i \partial x_j} \right)_0 (x_i - x_i^0)(x_j - x_j^0) + \dots$$

V étant définie à une constante additive près, on peut poser :

$$V_0 = 0$$

De plus, si la structure se trouve être à un minimum d'énergie, on a :

$$\forall i \in \{1, 2, \dots, N\}, \left(\frac{\partial V}{\partial x_i} \right)_0 = 0$$

Dans le cadre de petits déplacements autour de la position d'équilibre, on peut négliger les termes d'ordre supérieur ou égal à 3. Au final, on obtient V sous la forme d'une somme de potentiels harmoniques :

$$V \approx \frac{1}{2} \sum_{i,j=1}^{3N} \left(\frac{\partial^2 V}{\partial x_i \partial x_j} \right)_0 (x_i - x_i^0)(x_j - x_j^0)$$

L'équation du mouvement s'écrit donc, sous forme matricielle :

$$M\ddot{X} = -K(X - X_0)$$

M est la matrice ($3N \times 3N$) diagonale contenant les masses des atomes, K est la matrice des dérivées secondes du potentiel dont les éléments sont donnés par $K_{ij} = \frac{\partial^2 V}{\partial x_i \partial x_j}$, X est le vecteur des coordonnées x_i et X_0 est ce même vecteur dans la structure de référence. En utilisant des coordonnées pondérées par les masses, cette équation devient :

$$\ddot{\tilde{X}} = -\tilde{K}(\tilde{X} - \tilde{X}_0)$$

avec : $\tilde{X} = \sqrt{M}X$, $\tilde{X}_0 = \sqrt{M}X_0$, $\tilde{K} = \sqrt{M}^{-1}K\sqrt{M}^{-1}$

Cette équation classique d'oscillateurs harmoniques à $3N$ dimensions se résout en recherchant des solutions de la forme :

$$x_i(t) = \sum_{k=1}^{3N} C_k a_{ik} \cos(\omega_k t + \phi_k)$$

c'est-à-dire comme une superposition de $3N$ modes normaux de vibration. Chaque mode normal de vibration est défini par :

- une direction de déplacement donnée par le vecteur $A_k = \{a_{ik}\}_{i=1,3N}$.
- une fréquence ω_k .

Ces grandeurs sont obtenues en diagonalisant la matrice \tilde{K} : les vecteurs A_k sont les vecteurs propres de cette matrice et les valeurs propres associées sont égales au carré des fréquences. Il est à noter qu'il existe 6 modes de fréquence nulle (5 pour une molécule linéaire) qui correspondent aux mouvements d'ensemble de la protéine (3 translations et 2 ou 3 rotations).

Limitations

La théorie des modes normaux repose sur diverses approximations sur lesquelles il convient de revenir :

1. L'ensemble des dérivées premières du potentiel doivent être nulles, ce qui exige que la protéine soit dans une conformation d'énergie minimale. Ceci peut nécessiter des étapes de minimisation préalables très intensives qui peuvent déformer notablement la structure initiale. Cette limitation peut cependant être levée en utilisant une représentation granulaire de la protéine couplée à un réseau de ressorts (voir annexe B.3), car la protéine est alors par définition initialement à un minimum d'énergie potentielle.
2. Dans le développement en série de Taylor, tous les termes d'ordre supérieur à 2 sont négligés. Compte tenu de la complexité de la surface d'énergie potentielle des protéines, cette approximation peut ne pas être vérifiée, notamment pour de grands déplacements (Perahia *et al.*, 1990).
3. Les mouvements décrits par les modes normaux sont confinés à l'intérieur d'une surface d'énergie parabolöide multi-dimensionnelle et ne peuvent donc pas servir à étudier des mouvements nécessitant le franchissement de barrières énergétiques.
4. Le nombre d'éléments de la matrice \tilde{K} à diagonaliser croît en $O(N^2)$ (où N est le nombre d'atomes de la molécule), atteignant rapidement les limites mémorielles des ordinateurs. Plusieurs solutions permettent néanmoins de remédier à ce problème comme par exemple l'utilisation de modèles réduits de la protéine (Hinsen, 1998 ; Keskin, 2002), ou l'utilisation de blocs structuraux rigides (Durand *et al.*, 1994 ; Tama *et al.*, 2000).

Applications

L'analyse des modes normaux est une technique très employée pour l'étude des mouvements des protéines. Parmi les applications courantes, on peut citer :

- la reproduction de mouvements de domaines observés en cristallographie (Tama et Sanejouand, 2001) ou la définition de domaines dynamiques à partir de la structure de la protéine (Hinsen, 1998 ; Hinsen *et al.*, 1999).
- l'optimisation des paramètres d'un champ de force, par comparaison des spectres de fréquence calculés avec ceux obtenus expérimentalement par spectroscopie infrarouge ou Raman ou par diffusion de neutrons (Smith, 1991).

- l’affinement cristallographique (Diamond, 1990).
- l’exploration conformationnelle, en utilisant les coordonnées normales comme variables pour des simulations de Monte Carlo ou de dynamique moléculaire (Noguti et Go, 1985).

E.2 Analyse en composantes principales - modes quasi-harmoniques

Les modes quasi-harmoniques peuvent être vus comme les modes normaux d’une approximation harmonique de la surface d’énergie potentielle réelle. À l’instar des modes normaux, ils décrivent le mouvement de la protéine comme une somme de modes de vibration. D’ailleurs, dans le cas où la surface d’énergie serait réellement harmonique, les modes quasi-harmoniques se confondraient avec les modes normaux.

Obtention des modes quasi-harmoniques

Contrairement aux modes normaux, les modes quasi-harmoniques (ou principaux) ne sont pas une technique de simulation à part entière, mais plutôt un outil d’analyse. Leur calcul nécessite en effet de posséder au préalable un échantillonnage de la surface énergétique obtenu par dynamique moléculaire ou simulation de Monte Carlo : le calcul des modes quasi-harmoniques est une analyse en composantes principales (ACP) des mouvements internes de la protéine le long de la trajectoire obtenue par l’une de ces méthodes. Il est ainsi possible de ne calculer les modes quasi-harmoniques de la protéine que sur un sous-ensemble d’atomes (typiquement les C_α du système).

Soit R la matrice de variance-covariance des déplacements internes pondérés par les masses. Les éléments de cette matrice sont donnés par :

$$R_{ij} = \frac{1}{K} \langle \Delta x_i \Delta x_j \rangle$$

où Δx_i représente le déplacement multiplié par la racine carrée de la masse de l’atome i , et K est le nombre de structures, provenant de la simulation d’échantillonnage, utilisées pour effectuer l’analyse.

La diagonalisation de cette matrice R donne les modes principaux : les vecteurs propres U_j sont les directions de ces modes quasi-harmoniques et les valeurs propres associées w_j sont les carrés des fluctuations dues aux modes principaux. La projection de la trajectoire sur le j^e mode principal est donnée par :

$$X_j(t) = \sum_i x_i(t) U_{ij}$$

où $x_i(t)$ est la valeur du déplacement atomique pour la structure t (le temps t si l'ACP a été réalisée sur une trajectoire de dynamique moléculaire), et U_{ij} le i^e élément du mode quasi-harmonique U_j .

Les fréquences sont obtenues à partir des fluctuations des modes principaux par la relation :

$$\omega_j^{eff} = \sqrt{\frac{k_B T}{w_j}}$$

Comparaison avec les modes normaux

L'obtention des modes quasi-harmoniques est indépendante de la simulation effectuée pour échantillonner l'espace conformationnel. Ainsi, il est possible de réaliser une analyse sur une trajectoire de dynamique moléculaire dans laquelle le solvant a été représenté de manière explicite. Les modes principaux alors obtenus tiennent compte des effets de gêne stérique du solvant, au contraire des modes normaux qui sont calculés dans le vide ou en utilisant une représentation implicite du solvant.

De plus, l'analyse en composantes principales permet de définir des modes anharmoniques, modes qui ne sont évidemment pas accessibles par la technique des modes normaux. Hayward et collaborateurs (Hayward *et al.*, 1995) ont démontré que les modes de plus basse fréquence obtenus par ACP correspondent aux modes anharmoniques. Cependant, l'espace conformationnel qu'ils décrivent est contenu en grande partie dans le sous-espace décrit par les modes normaux harmoniques de basse fréquence (Hayward *et al.*, 1994). De plus, les fluctuations dérivées du calcul des modes quasi-harmoniques obtenus par ACP et celles obtenus par les modes normaux sont reliées par un facteur de proportionalité de 1,4 (Hayward *et al.*, 1995). Ceci indique que l'approximation de la surface de potentiel par une surface harmonique est similaire à celle effectuée au niveau du minimum d'énergie à un facteur de proportionnalité près. Cette observation permet notamment de comprendre comment les modes normaux, dont la validité théorique est limitée aux petits déplacements, sont capables de prédire des changements conformationnels de grande ampleur.

Annexe F

Modèles théoriques de solvation

Le rôle de l'eau est primordial pour la structuration et la fonction des édifices biomoléculaires (Nicholls, 2000). Pour tenir compte de ses effets en modélisation moléculaire, deux grandes méthodes sont couramment utilisées. Dans la première, les molécules du solvant sont représentées de manière explicite à l'échelle microscopique, tandis que dans la seconde seuls les effets macroscopiques du solvant sont modélisés.

F.1 Solvant explicite

L'utilisation de molécules d'eau explicites est certainement la meilleure représentation du solvant disponible actuellement en modélisation moléculaire. Elle permet en effet de pouvoir réaliser des simulations stables de dynamique moléculaire à l'échelle de plusieurs dizaines de nanosecondes. De plus, la représentation au niveau atomique de ces molécules rend possible l'observation de la formation de liaisons hydrogène entre le soluté et le solvant. Ainsi, l'utilisation d'un solvant explicite est indispensable lorsque l'eau est en interaction directe avec la biomolécule et que son rôle n'est pas limité à ses effets électrostatiques.

Il existe différents modèles de molécules d'eau en modélisation moléculaire, qui diffèrent par le nombre de sites (de 3 à 5 sites pour représenter la répartition des charges dans la molécule), les différentes charges partielles de ces sites, ou la géométrie de la molécule (longueur des liaisons $O-H$ et angle de valence \widehat{HOH}). Pour l'étude des biomolécules, les modèles les plus utilisés sont les modèles TIP3P (Jorgensen *et al.*, 1983), SPC (Berendsen *et al.*, 1981) ou SPC/E (Berendsen *et al.*, 1987), qui représentent tous trois la molécule d'eau avec trois sites (l'atome d'oxygène et les deux atomes d'hydrogène). Le tableau F.1 présente quelques paramètres qui différencient ces modèles d'eau.

Dans nos études, nous avons utilisé le modèle d'eau TIP3P, car les paramètres des champs de force disponibles dans le logiciel AMBER ont été optimisés pour l'utilisation de ce modèle. De plus, comme dans la plupart des simulations, la conformation des molécules d'eau est fixée. Dans le logiciel AMBER, cette rigidité est assurée par deux liaisons fixes $O - H$ de 0,9572 Å et une liaison fictive $H - H$ de 1,514 Å.

Modèle	r_{OH} (Å)	q_O (e)	q_H (e)	\widehat{HOH} (°)
SPC	1,0	-0,82	0,41	109,47
SPC/E	1,0	-0,8476	0,4238	109,47
TIP3P	0,9572	-0,834	0,417	104,52

TABLE F.1 – Valeurs de quelques paramètres pour les modèles d'eau SPC, SPC/E et TIP3P. r_{OH} est la longueur de la liaison $O - H$, q_O et q_H sont les charges respectivement des atomes d'oxygène et d'hydrogène données en unité atomique ($e = 1,6.10^{-19} C$).

En dépit de ses nombreux avantages, la représentation des molécules d'eau de manière explicite présente certains inconvénients. Le nombre de molécules à ajouter pour solvater une protéine peut en effet être très important. Par exemple, dans le cas du dimère *BE* de la cadhérine étudié dans le chapitre 6, nous avons dû ajouter 24632 molécules d'eau soit 73896 atomes pour une protéine de 6600 atomes environ. Les interactions entre molécules du solvant sont donc très coûteuses en temps de calcul, d'autant plus qu'elles ne sont pas l'objet de nos simulations. Ceci rend de plus très difficile l'obtention de grandeurs thermodynamiques qui nécessitent des temps de simulation très longs pour pouvoir calculer des moyennes significatives.

F.2 Solvant implicite

L'utilisation de modèles de solvant implicite est une alternative de plus en plus crédible à l'utilisation de solvants explicites. Dans ces modèles, le soluté est considéré comme une cavité de faible constante diélectrique ϵ_i au sein du solvant représenté comme un milieu continu de constante diélectrique ϵ_s , élevée. Ils permettent notamment de dépasser certaines limitations des modèles de solvant explicite en étant plus avantageux en termes de temps de calcul et en permettant l'obtention de grandeurs thermodynamiques. Au cours de notre étude, nous avons exploité ces deux propriétés en utilisant deux approches différentes : le modèle Poisson-Boltzmann et le modèle Born généralisé.

F.2.1 Modèle Poisson-Boltzmann

Dans un milieu continu de constante diélectrique ϵ , le potentiel électrostatique ϕ obéit à l'équation de Poisson :

$$\nabla \cdot [\epsilon(r)\nabla\phi(r)] = -4\pi\rho(r)$$

où ρ représente la densité de charges. Il est possible de modéliser la présence d'ions mobiles en solution et, dans le cas d'un sel monovalent, on obtient l'équation de Poisson-Boltzmann :

$$\nabla \cdot [\epsilon(r)\nabla\phi(r)] - \kappa'(r)\sinh[\phi(r)] = -4\pi\rho(r)$$

où κ' est relié à la constante de Debye-Hückel κ par la relation :

$$\kappa'^2 = \epsilon\kappa^2 = \frac{8\pi N_A e^2 I}{1000 k_B T}$$

I est la force ionique de la solution, N_A le nombre d'Avogadro, k_B la constante de Boltzmann, T la température et e la charge élémentaire. On obtient l'équation de Poisson-Boltzmann linéarisée en ne conservant que le premier terme du développement de la fonction \sinh en série de Taylor :

$$\nabla \cdot [\epsilon(r)\nabla\phi(r)] - \kappa'(r)\phi(r) = -4\pi\rho(r)$$

Cette équation ne peut être résolue de façon analytique que pour des cas très simples (en modélisant le soluté par des sphères ou des ellipsoïdes par exemple). Des méthodes numériques ont donc été développées pour résoudre cette équation pour les cas généraux, dont la plus utilisée est celle dite des différences finies (FDPB pour "*Finite Difference Poisson-Boltzmann*"). Cette méthode consiste à superposer au système un maillage cubique régulier : à chaque point de la grille ainsi définie, on fixe une valeur pour chacun des paramètres de l'équation de Poisson-Boltzmann (une densité de charge, une constante diélectrique et une valeur de force ionique), et le potentiel électrostatique est calculé en chacun de ces points par une méthode itérative. Une fois le potentiel défini en tout point, l'enthalpie libre de solvatation est donnée par :

$$\Delta G_{PB} = \frac{1}{2} \sum_{i=1}^N q_i (\phi_i - \phi_i^0)$$

où ϕ_i et ϕ_i^0 représentent respectivement le potentiel électrostatique solution de l'équation de Poisson-Boltzmann pour une constante diélectrique ϵ_s élevée (78,5 pour l'eau) et pour une constante diélectrique de 1 (pour le vide), et q_i la charge de l'atome i .

La résolution de l'équation de Poisson-Boltzmann représente aujourd'hui une référence dans le calcul des enthalpies libres de solvation. C'est pourquoi nous l'avons utilisé dans le calcul des enthalpies libres de complexation. En revanche, son utilisation dans des simulations nécessitant le calcul de dérivées (minimisation d'énergie ou dynamique moléculaire) est délicate du fait des discontinuités inhérentes à l'utilisation de calculs sur grilles. Ainsi, pour ce type de simulations, d'autres modèles de solvant implicite dérivables ont été développés comme le modèle de Born généralisé.

F.2.2 Modèle Born généralisé

Dans le cadre de nos minimisations d'énergie sous contrainte (voir chapitre 3), nous avons utilisé le modèle de Born généralisé (aussi noté GB pour "*Generalized Born*"). Ce modèle est basé sur la formule de Born qui donne l'enthalpie libre de solvation d'un ion de charge q et de rayon a par la formule :

$$\Delta G_{Born} = -\frac{q^2}{2a} \left(1 - \frac{1}{\epsilon_s}\right)$$

Pour une molécule constituée de N atomes de charge q_1, \dots, q_N et de rayons a_1, \dots, a_N , et dans le cas où la distance r_{ij} entre deux atomes i et j est suffisamment grande devant le rayon des atomes, l'extension de la formule de Born donne :

$$\Delta G_{solv} = -\sum_{i=1}^N \frac{q_i^2}{2a_i} \left(1 - \frac{1}{\epsilon_s}\right) - \frac{1}{2} \sum_{i=1}^N \sum_{i \neq j}^N \frac{q_i q_j}{r_{ij}} \left(1 - \frac{1}{\epsilon_s}\right)$$

La première somme de cette expression représente la solvation propre de chacune des N charges d'après la formule de Born, et la seconde double somme représente l'écrantage des interactions coulombiennes lors de l'immersion de la molécule dans le solvant.

Le modèle GB propose une approximation de l'enthalpie libre de solvation pour une molécule sous la forme :

$$\Delta G_{GB} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon_s}\right) \sum_{i,j} \frac{q_i q_j}{f_{GB}}$$

f_{GB} est une fonction qui doit répondre à deux critères. Elle doit se réduire, pour $i = j$, au terme de solvation propre avec un rayon de Born effectif α_i , et pour de grandes distances r_{ij} à l'écrantage des interactions coulombiennes, permettant ainsi de retrouver la formule précédente.

L'expression de f_{GB} utilisée dans nos simulations est :

$$f_{GB} = \sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp\left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j}\right)}$$

Le rayon de Born effectif est calculé à l'aide de la méthode de Hawkins, Cramer et Truhlar (Hawkins *et al.*, 1995 ; Hawkins *et al.*, 1996) :

$$\alpha_i^{-1} = \rho_i^{-1} - \sum_{j \neq i} g(\vec{r}_i, \vec{r}_j, \rho_i, \rho_j)$$

où \vec{r}_i est le vecteur position de l'atome i et ρ_i son rayon de Born intrinsèque. En notant r_{ij} la distance entre les atomes i et j , la fonction g est donnée par :

$$g(\vec{r}_i, \vec{r}_j, \rho_i, \rho_j) = \frac{1}{2} \sum_j \left[\frac{1}{L_{ij}} - \frac{1}{U_{ij}} + \frac{r_{ij}}{4} \left(\frac{1}{U_{ij}^2} - \frac{1}{L_{ij}^2} \right) + \frac{1}{2r_{ij}} \ln \left(\frac{L_{ij}}{U_{ij}} \right) + \frac{\rho_j^2}{4r_{ij}} \left(\frac{1}{L_{ij}^2} - \frac{1}{U_{ij}^2} \right) \right]$$

avec :

$$\begin{aligned} L_{ij} &= 1 & \text{si } (r_{ij} + \rho_j) \leq \rho_i \\ L_{ij} &= \rho_i & \text{si } (r_{ij} - \rho_j) \leq \rho_i < (r_{ij} + \rho_j) \\ L_{ij} &= r_{ij} - \rho_j & \text{si } \rho_j \leq (r_{ij} - \rho_j) \end{aligned}$$

et :

$$\begin{aligned} U_{ij} &= 1 & \text{si } (r_{ij} + \rho_j) \leq \rho_i \\ U_{ij} &= r_{ij} - \rho_j & \text{si } \rho_i < r_{ij} + \rho_j \end{aligned}$$

Il est possible d'incorporer un terme de Debye-Hückel pour rendre compte d'une concentration saline non nulle. L'enthalpie libre de solvation est alors donnée par :

$$\Delta G_{GB} = -\frac{1}{2} \left(1 - \frac{e^{-\kappa f_{GB}}}{\epsilon_s} \right) \sum_{i,j} \frac{q_i q_j}{f_{GB}}$$

où κ est la constante de Debye-Hückel.

La méthode de Born généralisé a été utilisée pour réaliser des simulations de dynamique moléculaire stables de protéines ou d'ADN sur plusieurs nanosecondes (Tsui et Case, 2000 ; Calimet *et al.*, 2001). Elle permet de diminuer le temps de simulation (pour une protéine globulaire d'environ 2000 atomes, le gain de temps de calcul s'élève à 30 % environ par rapport à une simulation en solvant explicite) mais également d'accélérer les changements structuraux du fait de l'absence de frictions avec les molécules de solvant (Tsui et Case, 2000).

Annexe G

Calculs d'enthalpies libres de complexation par la méthode MM-PBSA

L'enthalpie libre est une des grandeurs thermodynamiques les plus importantes, mais elle est difficile à obtenir lorsqu'on s'intéresse aux macromolécules flexibles, dont la surface d'énergie potentielle possède de nombreux minima reliés entre eux par de faibles barrières énergétiques.

Au cours de nos travaux sur les cadhérines, nous avons voulu comparer les enthalpies libres de formation de différents dimères de EC1-2. Pour cela, nous avons utilisé la méthode MM-PBSA, qui permet d'estimer les enthalpies libres de complexation à partir de trajectoires de dynamique moléculaire. Cette méthode a été employée pour la première fois sous sa forme actuelle à la fin des années 1990 (Srinivasan *et al.*, 1998) et a depuis été utilisée avec succès pour déterminer les enthalpies libres de formation de complexes protéine/ligand, protéine/protéine, protéine/ADN, ou protéine/ARN (Kollman *et al.*, 2000).

G.1 Principe de la méthode

La méthode MM-PBSA est fondée sur le fait que l'enthalpie libre est une fonction d'état et que sa variation dépend des états initiaux et finaux et non du chemin. Ceci peut s'écrire, dans le cas de la formation d'un complexe, sous la forme :

$$\Delta G = G^{comp} - (G^1 + G^2)$$

où G^{comp} représente l'enthalpie libre du complexe et G^1 et G^2 les enthalpies libres de chacun des monomères.

En réalité, l'enthalpie libre absolue G d'une espèce en solution n'est pas une grandeur mesurable, et on utilise le cycle thermodynamique décrit sur la figure G.1.

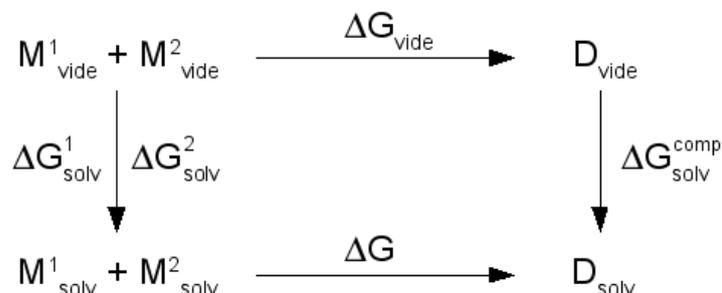


FIGURE G.1 – Cycle thermodynamique utilisé pour calculer l'enthalpie libre de formation ΔG du complexe D.

L'enthalpie libre de formation du complexe s'écrit donc :

$$\Delta G = \Delta G_{vide} + \Delta G_{solv}^{comp} - \Delta G_{solv}^1 - \Delta G_{solv}^2$$

En notant que $\Delta G_{vide} = G_{vide}^{comp} - (G_{vide}^1 + G_{vide}^2)$, on obtient :

$$\Delta G = \Delta G^{comp} - (\Delta G^1 + \Delta G^2)$$

avec $\Delta G^x = G_{vide}^x + \Delta G_{solv}^x$.

L'approche MM-PBSA consiste à calculer chacun de ces trois termes à partir d'un ensemble de structures issues de trajectoires¹ de dynamique moléculaire ou de Monte Carlo (en retirant les molécules d'eau et les ions éventuellement présents). L'enthalpie libre de complexation s'écrit alors :

$$\Delta G = \langle \Delta G^{comp}(i) \rangle_i - (\langle \Delta G^1(i) \rangle_i + \langle \Delta G^2(i) \rangle_i)$$

où chacun des termes du membre droit de l'équation représente une moyenne sur l'ensemble des structures issues des dynamiques. Les termes $\Delta G^x(i)$ sont estimés par la somme des énergies du champ de force (H_{MM}), d'un terme de solvation (ΔG_{solv}) et d'un terme entropique ($-TS$) :

$$\Delta G^x(i) = H_{MM}^x(i) + \Delta G_{solv}^x(i) - TS^x(i)$$

1. Ces ensembles de structures peuvent être extraites de plusieurs trajectoires (une pour le complexe et une pour chacun des monomères "libres") ou bien extraites d'une unique trajectoire du complexe, cette dernière procédure négligeant les changements conformationnels induits par la complexation.

G.2 Terme enthalpique

La contribution enthalpique correspond aux énergies du champ de force :

$$H_{MM} = H_{int} + H_{vdW} + H_{elec}$$

H_{int} correspond à l'ensemble des énergies liées du champ de force (élongation de liaisons, torsions angulaires), H_{vdW} aux interactions de van der Waals et H_{elec} aux énergies coulombiennes. L'expression de ces différents termes est détaillée dans l'annexe B.2. Dans le calcul de ces contributions, aucune troncature de potentiel n'est réalisée et il est possible d'utiliser un champ de force différent de celui utilisé pour l'obtention de la trajectoire en dynamique moléculaire.

G.3 Terme de solvation

Le terme de solvation se décompose en une contribution polaire et une contribution non polaire :

$$\Delta G_{solv} = \Delta G_{polaire} + \Delta G_{non\ polaire}$$

La contribution polaire est calculée à l'aide d'un modèle de solvation implicite (PB ou GB, voir annexe F). Dans notre cas, nous avons utilisé le modèle de Poisson-Boltzmann, qui sert actuellement de référence aux calculs d'enthalpies libres de solvation, implémenté dans le programme *pbsa* (Luo *et al.*, 2002) de la suite AMBER.

La contribution non polaire provient de la création d'une cavité de faible constante diélectrique dans le solvant et des interactions de van der Waals entre les molécules de solvant et le soluté. Elle ne dépend pas du caractère polaire ou chargé de la molécule et est modélisée par un terme dépendant de la surface accessible au solvant (*SAS*) du soluté :

$$\Delta G_{non\ polaire} = \gamma SAS + b$$

Les valeurs de *SAS* sont obtenues par la méthode LCPO (Weiser *et al.*, 1999). Les valeurs des paramètres γ et b sont généralement prises de façon à reproduire les enthalpies libres de solvation de petits alcanes. Dans nos calculs, nous avons fixé γ à $0,005 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ et b à 0 kcal.mol^{-1} .

G.4 Terme entropique

Le terme entropique est décomposé en trois termes calculés à partir de formules de thermodynamique statistique, en appliquant un modèle de gaz parfait pour le soluté. Les trois termes représentent les entropies de translation, de rotation et de vibration :

$$\begin{aligned}
 S_{trans} &= 2,5 R - R \ln \left(\frac{\Lambda^3 P}{k_B T} \right) \\
 S_{rot} &= 1,5 R + R \ln \left[\pi^{\frac{1}{2}} \left(\frac{8\pi^2 k_B T}{h^2} \right)^{\frac{3}{2}} \det(A)^{\frac{1}{2}} \right] \\
 S_{vib} &= \sum_{i=1}^M \left[\frac{\frac{h\nu_i}{k_B T}}{\exp\left(-\frac{h\nu_i}{k_B T}\right) - 1} - \ln \left(1 - \exp\left(-\frac{h\nu_i}{k_B T}\right) \right) \right]
 \end{aligned}$$

k_B est la constante de Boltzmann, P la pression (fixée à 1 bar), R la constante des gaz parfaits, T la température, A le tenseur d'inertie de la molécule, et ν_i la fréquence de vibration du i^e mode de vibration. Λ est la longueur d'onde de de Broglie donnée par :

$$\Lambda = \frac{h}{\sqrt{2\pi M k_B T}}$$

où M représente la masse totale de la molécule. Les fréquences des modes de vibration sont obtenues par analyse des modes normaux (voir annexe E). Ces modes normaux sont calculés sur un sous-ensemble réduit de structures issues de la dynamique moléculaire, car le calcul des modes normaux constitue la partie la plus coûteuse en termes de temps de calcul. Il est en théorie également possible d'obtenir ces fréquences à partir du calcul des modes quasi-harmoniques. Cependant, cette méthode souffre d'un problème de convergence et les valeurs obtenues par ce biais semblent surestimées (Gohlke et Case, 2004).

G.5 Limitations

La méthode MM-PBSA, si elle permet d'obtenir des estimations d'enthalpies libres de complexation, pâtit de plusieurs limitations :

1. Dans cette méthode, l'enthalpie libre de complexation est calculée comme la différence de grands termes (fréquemment plusieurs milliers de kcal.mol⁻¹) avec des variations importantes pour obtenir une valeur assez faible (de l'ordre de quelques kcal.mol⁻¹), avec des incertitudes importantes.
2. Les valeurs de ΔG calculées sont très sensibles au choix des paramètres, notamment les charges et rayons atomiques utilisés pour calculer les termes de solvation.

Bibliographie

- Aken, E. V., Wever, O. D., da Rocha, A. S. C. et Mareel, M. (2001) Defective E-cadherin/catenin complexes in human cancer. *Virchows Arch.*, **439** (6), 725–751.
- Alattia, J. R., Ames, J. B., Porumb, T., Tong, K. I., Heng, Y. M., Ottensmeyer, P., Kay, C. M. et Ikura, M. (1997) Lateral self-assembly of E-cadherin directed by cooperative calcium binding. *FEBS Lett.*, **417** (3), 405–408.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. et Watson, J. (1994) *Molecular biology of the cell, Third Edition*. Garland Publishing.
- Allen, M. et Tildesley, D. (1987) *Computer simulations of liquids*. Clarendon Press, Oxford.
- Allouche, D., Parello, J. et Sanejouand, Y. H. (1999) Ca²⁺/Mg²⁺ exchange in parvalbumin and other EF-hand proteins. A theoretical study. *J. Mol. Biol.*, **285** (2), 857–873.
- Anfinsen, C. B. et Scheraga, H. A. (1975) Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.*, **29**, 205–300.
- Babu, C. S., Dudev, T., Casareno, R., Cowan, J. A. et Lim, C. (2003) A combined experimental and theoretical study of divalent metal ion selectivity and function in proteins : application to E. coli ribonuclease H1. *J. Am. Chem. Soc.*, **125** (31), 9318–9328.
- Bahadur, R. P., Chakrabarti, P., Rodier, F. et Janin, J. (2003) Dissecting subunit interfaces in homodimeric proteins. *Proteins*, **53** (3), 708–719.
- Bahadur, R. P., Chakrabarti, P., Rodier, F. et Janin, J. (2004) A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.*, **336** (4), 943–955.
- Bahar, I., Atilgan, A. R. et Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding & Design*, **2** (3), 173–181.
- Bahar, I. et Jernigan, R. L. (1999) Cooperative fluctuations and subunit communication in tryptophan synthase. *Biochemistry*, **38** (12), 3478–3490.
- Bairoch, A. (2000) The ENZYME database in 2000. *J. Phys. Chem.*, **28** (1), 304–305.

- Banavali, N. K. et MacKerell, A. D. (2002) Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.*, **319** (1), 141–160.
- Bartlett, G. J., Porter, C. T., Borkakoti, N. et Thornton, J. M. (2002) Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.*, **324** (1), 105–121.
- Bastard, K., Thureau, A., Lavery, R. et Prévost, C. (2003) Docking macromolecules with flexible segments. *J. Comput. Chem.*, **24** (15), 1910–1920.
- Bayas, M., Schulten, K. et Leckband, D. (2004) Force dissociation of the strand dimer interface between C-cadherin ectodomains. *Mech. Chem. Biosys.*, **1** (2), 101–111.
- Bayas, M. V., Leung, A., Evans, E. et Leckband, D. (2006) Lifetime measurements reveal kinetic differences between homophilic cadherin bonds. *Biophys. J.*, **90** (4), 1385–1395.
- Bayas, M. V., Schulten, K. et Leckband, D. (2003) Forced detachment of the CD2-CD58 complex. *Biophys. J.*, **84** (4), 2223–2233.
- Bayly, C. I., Cieplak, P., Cornell, W. D. et Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges : the resp model. *J. Phys. Chem.*, **97** (40), 10269–10280.
- Berendsen, H., Postma, J., van Gunsteren, W. et Hermans, J. (1981) *Intermolecular forces*. D. Reidel Publishing Company.
- Berendsen, H. J. C., Grigera, J. R. et Straatsma, T. P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.*, **91** (24), 6269–6271.
- Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. et Haak, J. R. (1984) Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, **81** (8), 3684–3690.
- Björkman, A. J., Binnie, R. A., Zhang, H., Cole, L. B., Hermodson, M. A. et Mowbray, S. L. (1994) Probing protein-protein interactions. The ribose-binding protein in bacterial transport and chemotaxis. *J. Biol. Chem.*, **269** (48), 30206–30211.
- Björkman, A. J. et Mowbray, S. L. (1998) Multiple open forms of ribose-binding protein trace the path of its conformational change. *J. Mol. Biol.*, **279** (3), 651–664.
- Blaszczyk, J., Li, Y., Yan, H. et Ji, X. (2001) Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes. *J. Mol. Biol.*, **307** (1), 247–257.
- Boggon, T. J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B. M. et Shapiro, L. (2002) C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science*, **296** (5571), 1308–1313.

-
- Brieher, W. M., Yap, A. S. et Gumbiner, B. M. (1996) Lateral dimerization is required for the homophilic binding activity of C-cadherin. *J. Cell Biol.*, **135** (2), 487–496.
- Brink, J., Ludtke, S. J., Kong, Y., Wakil, S. J., Ma, J. et Chiu, W. (2004) Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure*, **12** (2), 185–191.
- Bryant, Z., Pande, V. S. et Rokhsar, D. S. (2000) Mechanical unfolding of a beta-hairpin using molecular dynamics. *Biophys. J.*, **78** (2), 584–589.
- Bryant, Z., Stone, M. D., Gore, J., Smith, S. B., Cozzarelli, N. R. et Bustamante, C. (2003) Structural transitions and elasticity from torque measurements on dna. *Nature*, **424** (6946), 338–341.
- Bustamante, C., Bryant, Z. et Smith, S. B. (2003) Ten years of tension : single-molecule DNA mechanics. *Nature*, **421** (6921), 423–427.
- Calimet, N., Schaefer, M. et Simonson, T. (2001) Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins*, **45** (2), 144–158.
- Carloni, P., Rothlisberger, U. et Parrinello, M. (2002) The role and perspective of ab initio molecular dynamics in the study of biological systems. *Acc. Chem. Res.*, **35** (6), 455–464.
- Case, D., Darden, T., Cheatham III, T., Simmerling, C., Wang, J., Duke, R., Luo, R., Merz, K., Wang, B., Pearlman, D., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J., Ross, W. et Kollman, P. (2004) *AMBER8*. University of California San Francisco, CA.
- Case, D., Pearlman, D., Caldwell, J., Cheatham III, T., Wang, J., Ross, W., Simmerling, C., Darden, T., Merz, K., Stanton, R., Cheng, A., Vincent, J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R., Duan, Y., Pitner, J., Massova, I., Seibel, G., Singh, U., Weimer, P. et Kollman, P. (2002) *AMBER 7*. University of California San Francisco, CA.
- Chakrabarti, P. et Janin, J. (2002) Dissecting protein-protein recognition sites. *Proteins*, **47** (3), 334–343.
- Chakravarty, S. et Varadarajan, R. (2002) Elucidation of factors responsible for enhanced thermal stability of proteins : a structural genomics based study. *Biochemistry*, **41** (25), 8152–8161.
- Chappuis-Flament, S., Wong, E., Hicks, L. D., Kay, C. M. et Gumbiner, B. M. (2001) Multiple cadherin extracellular repeats mediate homophilic binding and adhesion. *J. Cell Biol.*, **154** (1), 231–243.

- Chatfield, D. C., Szabo, A. et Brooks, B. R. (1998) Molecular dynamics of staphylococcal nuclease : comparison of simulation with n-15 and c-13 nmr relaxation data. *J. Am. Chem. Soc.*, **120** (21), 5301–5311.
- Chattopadhyaya, R., Meador, W. E., Means, A. R. et Quioco, F. A. (1992) Calmodulin structure refined at 1.7 Å resolution. *J. Mol. Biol.*, **228** (4), 1177–1192.
- Cheatham, T. E., Miller, J. L., Fox, T., Darden, T. A. et Kollman, P. A. (1995) Molecular-dynamics simulations on solvated biomolecular systems. The Particle Mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.*, **117** (14), 4193–4194.
- Chen, C. P., Posy, S., Ben-Shaul, A., Shapiro, L. et Honig, B. H. (2005) Specificity of cell-cell adhesion by classical cadherins : critical role for low-affinity dimerization through beta-strand swapping. *Proc. Natl. Acad. Sci. U. S. A.*, **102** (24), 8531–8536.
- Chen, J., Lu, Z., Sakon, J. et Stites, W. E. (2000) Increasing the thermostability of staphylococcal nuclease : implications for the origin of protein thermostability. *J. Mol. Biol.*, **303** (2), 125–130.
- Chen, J. et Stites, W. E. (2001) Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry*, **40** (50), 15280–15289.
- Chitaev, N. A. et Troyanovsky, S. M. (1998) Adhesive but not lateral E-cadherin complexes require calcium and catenins for their formation. *J. Cell Biol.*, **142** (3), 837–846.
- Chothia, C. et Lesk, A. M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **5** (4), 823–826.
- Christofori, G. (2003) Changing neighbours, changing behaviour : cell adhesion molecule-mediated signalling during tumour progression. *EMBO J.*, **22** (10), 2318–2323.
- Cieplak, P., Cornell, W. D., Bayly, C. et Kollman, P. A. (1995) Application of the multi-molecule and multiconformational resp methodology to biopolymerscharge derivation for dna, rna, and proteins. *J. Comput. Chem.*, **16** (11), 1357–1377.
- Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J. L., Chatenay, D. et Caron, F. (1996) DNA : an extensible molecule. *Science*, **271** (5250), 792–794.
- Cooper, G. et Hausman, R. (2006) *The cell : a molecular approach, Fourth Edition*. Sinauer Associates, Inc.
- Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J. et Blundell, T. L. (1990) X-ray analyses of aspartic proteinases. II. Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 Å resolution. *J. Mol. Biol.*, **214** (1), 199–222.

- Cornell, W. D., Cieplak, P., Bayly, C. I. et Kollman, P. A. (1993) Application of resp charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. *J. Am. Chem. Soc.*, **115** (21), 9620–9631.
- Darden, T., York, D. et Pedersen, L. (1993) Particle Mesh Ewald : an n.log(n) method for Ewald sums in large systems. *J. Chem. Phys.*, **98** (12), 10089–10092.
- Deshpande, N., Address, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z., Green, R. K., Flippen-Anderson, J. L., Westbrook, J., Berman, H. M. et Bourne, P. E. (2005) The RCSB Protein Data Bank : a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.*, **33** (Database issue), D233–D237.
- Diamond, R. (1990) On the use of normal-modes in thermal parameter refinement theory and application to the bovine pancreatic trypsin-inhibitor. *Acta Crystallogr., Sect. A : Found. Crystallogr.*, **46**, 425–435.
- Dinner, A. R., Blackburn, G. M. et Karplus, M. (2001) Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature*, **413** (6857), 752–755.
- Dohoney, K. M. et Gelles, J. (2001) Chi-sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules. *Nature*, **409** (6818), 370–374.
- Duguay, D., Foty, R. A. et Steinberg, M. S. (2003) Cadherin-mediated cell adhesion and tissue segregation : qualitative and quantitative determinants. *Dev. Biol.*, **253** (2), 309–323.
- Duong, T. H. et Zakrzewska, K. (1997) Influence of drug binding on DNA flexibility : a normal mode analysis. *J. Biomol. Struct. Dyn.*, **14** (6), 691–701.
- Durand, P., Trinquier, G. et Sanejouand, Y. H. (1994) A new approach for determining low-frequency normal-modes in macromolecules. *Biopolymers*, **34** (6), 759–771.
- Echols, N., Milburn, D. et Gerstein, M. (2003) MolMovDB : analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.*, **31** (1), 478–482.
- Fernández-Recio, J., Totrov, M. et Abagyan, R. (2002) Soft protein-protein docking in internal coordinates. *Protein Sci.*, **11** (2), 280–291.
- Flatters, D., Zakrzewska, K. et Lavery, R. (1997) Internal coordinate modeling of dna : force field comparisons. *J. Comput. Chem.*, **18** (8), 1043–1055.
- Flores, S., Echols, N., Milburn, D., Hespeneide, B., Keating, K., Lu, J., Wells, S., Yu, E. Z., Thorpe, M. et Gerstein, M. (2006) The Database of Macromolecular Motions : new features added at the decade mark. *Nucleic Acids Res.*, **34** (Database issue), D296–D301.

- Florin, E. L., Moy, V. T. et Gaub, H. E. (1994) Adhesion forces between individual ligand-receptor pairs. *Science*, **264** (5157), 415–417.
- Foty, R. A. et Steinberg, M. S. (2005) The differential adhesion hypothesis : a direct evaluation. *Dev. Biol.*, **278** (1), 255–263.
- Frank, M. et Kemler, R. (2002) Protocadherins. *Curr. Opin. Cell Biol.*, **14** (5), 557–562.
- Frauenfelder, H., Parak, F. et Young, R. D. (1988) Conformational substates in proteins. *Ann. Rev. Biophys. Biophys. Chem.*, **17**, 451–479.
- Frenkel, D. et Smit, B. (2002) *Understanding molecular simulations : from algorithms to applications, Second edition*. Academic press.
- Gerstein, M. et Krebs, W. (1998) A database of macromolecular motions. *Nucleic Acids Res.*, **26** (18), 4280–4290.
- Gerstein, M., Lesk, A. M. et Chothia, C. (1994) Structural mechanisms for domain movements in proteins. *Biochemistry*, **33** (22), 6739–6749.
- Getz, G., Vendruscolo, M., Sachs, D. et Domany, E. (2002) Automated assignment of SCOP and CATH protein structure classifications from FSSP scores. *Proteins*, **46** (4), 405–415.
- Gilquin, B., Guilbert, C. et Perahia, D. (2000) Unfolding of hen egg lysozyme by molecular dynamics simulations at 300K : insight into the role of the interdomain interface. *Proteins*, **41** (1), 58–74.
- Giudice, E., Varnai, P. et Lavery, R. (2001) Energetic and conformational aspects of a : t base-pair opening within the dna double helix. *ChemPhysChem*, **2** (11), 673–+.
- Gohlke, H. et Case, D. A. (2004) Converging free energy estimates : MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.*, **25** (2), 238–250.
- Grottesi, A., Ceruso, M.-A., Colosimo, A. et Nola, A. D. (2002) Molecular dynamics study of a hyperthermophilic and a mesophilic rubredoxin. *Proteins*, **46** (3), 287–294.
- Grubmüller, H., Heymann, B. et Tavan, P. (1996) Ligand binding : molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, **271** (5251), 997–999.
- Gumbiner, B. M. (1996) Cell adhesion : the molecular basis of tissue architecture and morphogenesis. *Cell*, **84** (3), 345–357.
- Haliloglu, T., Bahar, I. et Erman, B. (1997) Gaussian dynamics of folded proteins. *Phys. Rev. Lett.*, **79** (16), 3090–3093.
- Halle, B. (2002) Flexibility and packing in proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **99** (3), 1274–1279.

-
- Handschuh, G., Lubber, B., Hutzler, P., Höfler, H. et Becker, K. F. (2001) Single amino acid substitutions in conserved extracellular domains of E-cadherin differ in their functional consequences. *J. Mol. Biol.*, **314** (3), 445–454.
- Haney, P. J., Stees, M. et Konisky, J. (1999) Analysis of thermal stabilizing interactions in mesophilic and thermophilic adenylate kinases from the genus *Methanococcus*. *J. Biol. Chem.*, **274** (40), 28453–28458.
- Harrison, O. J., Corps, E. M., Berge, T. et Kilshaw, P. J. (2005) The mechanism of cell adhesion by classical cadherins : the role of domain 1. *J. Cell Sci.*, **118** (Pt 4), 711–721.
- Hawkins, G. D., Cramer, C. J. et Truhlar, D. G. (1995) Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.*, **246** (1-2), 122–129.
- Hawkins, G. D., Cramer, C. J. et Truhlar, D. G. (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, **100** (51), 19824–19839.
- Hayward, S., Kitao, A. et Berendsen, H. J. (1997) Model-free methods of analyzing domain motions in proteins from simulation : a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins*, **27** (3), 425–437.
- Hayward, S., Kitao, A. et Go, N. (1994) Harmonic and anharmonic aspects in the dynamics of BPTI : a normal mode analysis and principal component analysis. *Protein Sci.*, **3** (6), 936–943.
- Hayward, S., Kitao, A. et Go, N. (1995) Harmonicity and anharmonicity in protein dynamics : a normal mode analysis and principal component analysis. *Proteins*, **23** (2), 177–186.
- He, W., Cowin, P. et Stokes, D. L. (2003) Untangling desmosomal knots with electron tomography. *Science*, **302** (5642), 109–113.
- Hegyí, H. et Gerstein, M. (1999) The relationship between protein structure and function : a comprehensive survey with application to the yeast genome. *J. Mol. Biol.*, **288** (1), 147–164.
- Hinsen, K. (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins*, **33** (3), 417–429.
- Hinsen, K., Thomas, A. et Field, M. J. (1999) Analysis of domain motions in large proteins. *Proteins*, **34** (3), 369–382.
- Hirano, S., Mihara, K., Yamazaki, Y., Kamikubo, H., Imamoto, Y. et Kataoka, M. (2002) Role of C-terminal region of staphylococcal nuclease for foldability, stability, and activity. *Proteins*, **49** (2), 255–265.

- Holm, L. et Sander, C. (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.*, **26** (1), 316–319.
- Humphrey, W., Dalke, A. et Schulten, K. (1996) VMD : visual molecular dynamics. *J. Mol. Graph.*, **14** (1), 33–38.
- Hunenberger, P. H., Mark, A. E. et Vangunsteren, W. F. (1995) Computational approaches to study protein unfolding - hen egg-white lysozyme as a case-study. *Proteins*, **21** (3), 196–213.
- Häussinger, D., Ahrens, T., Aberle, T., Engel, J., Stetefeld, J. et Grzesiek, S. (2004) Proteolytic E-cadherin activation followed by solution NMR and X-ray crystallography. *EMBO J.*, **23** (8), 1699–1708.
- Häussinger, D., Ahrens, T., Sass, H.-J., Pertz, O., Engel, J. et Grzesiek, S. (2002) Calcium-dependent homoassociation of E-cadherin by NMR spectroscopy : changes in mobility, conformation and mapping of contact regions. *J. Mol. Biol.*, **324** (4), 823–839.
- Idiris, A., Alam, M. T. et Ikai, A. (2000) Spring mechanics of alpha-helical polypeptide. *Protein Eng.*, **13** (11), 763–770.
- Isin, B., Doruker, P. et Bahar, I. (2002) Functional motions of influenza virus hemagglutinin : a structure-based analytical approach. *Biophys. J.*, **82** (2), 569–581.
- Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D. et Schulten, K. (2001a) Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.*, **19** (1), 13–25.
- Isralewitz, B., Gao, M. et Schulten, K. (2001b) Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.*, **11** (2), 224–230.
- Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. et Schulten, K. (1997) Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, **72** (4), 1568–1581.
- Jain, A. K., Murty, M. N. et Flynn, P. J. (1999) Data clustering : a review. *ACM Computing Surveys*, **31** (3), 264–323.
- Janin, J. et Chothia, C. (1985) Domains in proteins : definitions, location, and structural principles. *Methods Enzymol.*, **115**, 420–430.
- Jones, S. et Thornton, J. M. (2004) Searching for functional sites in protein structures. *Curr. Opin. Chem. Biol.*, **8** (1), 3–7.
- Jorgensen, W. (1981) Transferable interatomic potential functions for water, alcohols and ethers. *J. Am. Chem. Soc.*, **103**, 335–340.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. et Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79** (2), 926–935.

- Keepers, J., Kollman, P. A. et James, T. L. (1984) Molecular mechanical studies of base-pair opening in d(CGCGC) :d(GCGCG), dG5 . dC5, d(TATAT) :d(ATATA), and dA5 . dT5 in the B and Z forms of DNA. *Biopolymers*, **23** (11 Pt 2), 2499–2511.
- Keskin, O. (2002) Comparison of full-atomic and coarse-grained models to examine the molecular fluctuations of c-AMP dependent protein kinase. *J. Biomol. Struct. Dyn.*, **20** (3), 333–345.
- Keskin, O., Jernigan, R. L. et Bahar, I. (2000) Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.*, **78** (4), 2093–2106.
- Kitao, A., Hayward, S. et Go, N. (1994) Comparison of normal-mode analyses on a small globular protein in dihedral angle space and cartesian coordinate space. *Biophys. Chem.*, **52** (2), 107–114.
- Koch, A. W., Manzur, K. L. et Shan, W. (2004) Structure-based models of cadherin-mediated cell adhesion : the evolution continues. *Cell. Mol. Life Sci.*, **61** (15), 1884–1895.
- Koch, A. W., Pokutta, S., Lustig, A. et Engel, J. (1997) Calcium binding and homoassociation of E-cadherin domains. *Biochemistry*, **36** (25), 7697–7705.
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A. et Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules : combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **33** (12), 889–897.
- Laskowski, R. A., Luscombe, N. M., Swindells, M. B. et Thornton, J. M. (1996) Protein clefts in molecular recognition and function. *Protein Sci.*, **5** (12), 2438–2452.
- Laskowski, R. A., Watson, J. D. et Thornton, J. M. (2003) From protein structure to biochemical function? *J. Struct. Funct. Genomics*, **4** (2-3), 167–177.
- Laur, O. Y., Klingelhöfer, J., Troyanovsky, R. B. et Troyanovsky, S. M. (2002) Both the dimerization and immunochemical properties of E-cadherin EC1 domain depend on Trp(156) residue. *Arch. Biochem. Biophys.*, **400** (1), 141–147.
- Lavery, R., Lebrun, A., Allemand, J. F., Bensimon, D. et Croquette, V. (2002) Structure and mechanics of single biomolecules : experiment and simulation. *J. Phys.-Condens. Mat.*, **14** (14), R383–R414.
- Lavery, R., Pullman, A. et Pullman, B. (1981) Steric accessibility of reactive centers in B-DNA. *Int. J. Quantum Chem.*, **20** (1), 49–62.
- Lavery, R., Sklenar, H., Zakrzewska, K. et Pullman, B. (1986) The flexibility of the nucleic acids : (II). The calculation of internal energy and applications to mononucleotide repeat DNA. *J. Biomol. Struct. Dyn.*, **3** (5), 989–1014.

- Lavery, R., Zakrzewska, K. et Sklenar, H. (1995) JUMNA (JUnction Minimization of Nucleic-Acids). *Comput. Phys. Commun.*, **91** (1-3), 135–158.
- Leach, A. (2001) *Molecular modelling - Principles and applications, Second edition*. Prentice Hall.
- Lebrun, A. et Lavery, R. (1996) Modelling extreme stretching of DNA. *Nucleic Acids Res.*, **24** (12), 2260–2267.
- Lebrun, A. et Lavery, R. (1998) Modeling the mechanics of a DNA oligomer. *J. Biomol. Struct. Dyn.*, **16** (3), 593–604.
- Leckband, D. et Sivasankar, S. (2000) Mechanism of homophilic cadherin adhesion. *Curr Opin Cell Biol*, **12** (5), 587–592.
- Lesk, A. M. et Chothia, C. (1984) Mechanisms of domain closure in proteins. *J. Mol. Biol.*, **174** (1), 175–191.
- Li, Z. et Scheraga, H. A. (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U. S. A.*, **84** (19), 6611–6615.
- Linke, W. A., Kulke, M., Li, H., Fujita-Becker, S., Neagoe, C., Manstein, D. J., Gautel, M. et Fernandez, J. M. (2002) PEVK domain of titin : an entropic spring with actin-binding properties. *J. Struct. Biol.*, **137** (1-2), 194–205.
- Lipari, G. et Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules .1. theory and range of validity. *J. Am. Chem. Soc.*, **104** (17), 4546–4559.
- Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. et Bustamante, C. (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science*, **292** (5517), 733–737.
- Lu, H., Isralewitz, B., Krammer, A., Vogel, V. et Schulten, K. (1998) Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.*, **75** (2), 662–671.
- Lu, H., Krammer, A., Isralewitz, B., Vogel, V. et Schulten, K. (2000) Computer modeling of force-induced titin domain unfolding. *Adv. Exp. Med. Biol.*, **481**, 143–60 ; discussion 161–2.
- Lu, H. et Schulten, K. (1999) Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins*, **35** (4), 453–463.
- Lu, H. et Schulten, K. (2000) The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.*, **79** (1), 51–65.
- Luo, R., David, L. et Gilson, M. K. (2002) Accelerated poisson-boltzmann calculations for static and dynamic systems. *J. Comput. Chem.*, **23** (13), 1244–1253.

- Ma, J., Sigler, P. B., Xu, Z. et Karplus, M. (2000) A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.*, **302** (2), 303–313.
- Makagiansar, I. T., Nguyen, P. D., Ikesue, A., Kuczera, K., Dentler, W., Urbauer, J. L., Galeva, N., Alterman, M. et Siahaan, T. J. (2002) Disulfide bond formation promotes the cis- and trans-dimerization of the E-cadherin-derived first repeat. *J. Biol. Chem.*, **277** (18), 16002–16010.
- Marquart, M., Walter, J., Deisenhofer, J., Bode, W. et Huber, R. (1983) The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr., Sect. B : Struct. Sci.*, **39** (AUG), 480–490.
- Marques, O. et Sanejouand, Y. H. (1995) Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins*, **23** (4), 557–560.
- Martin, A. C., Orengo, C. A., Hutchinson, E. G., Jones, S., Karmirantzou, M., Laskowski, R. A., Mitchell, J. B., Taroni, C. et Thornton, J. M. (1998) Protein folds and functions. *Structure*, **6** (7), 875–884.
- Masugata, K., Ikai, A. et Okazaki, S. (2002) Molecular dynamics study of mechanical extension of polyalanine by afm cantilever. *Appl. Surf. Sci.*, **188** (3-4), 372–376.
- Mazur, A. K. (2002) DNA dynamics in a water drop without counterions. *J. Am. Chem. Soc.*, **124** (49), 14707–14715.
- McCammon, J. et Harvey, S. (1987) *Dynamics of proteins and nucleic acids*. Cambridge University Press.
- Moy, V. T., Florin, E. L. et Gaub, H. E. (1994) Intermolecular forces and energies between ligands and receptors. *Science*, **266** (5183), 257–259.
- Murzin, A. G., Brenner, S. E., Hubbard, T. et Chothia, C. (1995) SCOP : a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247** (4), 536–540.
- Nagafuchi, A. et Takeichi, M. (1988) Cell binding function of E-cadherin is regulated by the cytoplasmic domain. *EMBO J.*, **7** (12), 3679–3684.
- Nagano, N., Orengo, C. A. et Thornton, J. M. (2002) One fold with many functions : the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.*, **321** (5), 741–765.
- Nagar, B., Overduin, M., Ikura, M. et Rini, J. M. (1996) Structural basis of calcium-induced E-cadherin rigidification and dimerization. *Nature*, **380** (6572), 360–364.
- Najmudin, S., Nalini, V., Driessen, H. P., Slingsby, C., Blundell, T. L., Moss, D. S. et Lindley, P. F. (1993) Structure of the bovine eye lens protein gammaB(gammaII)-

- crystallin at 1.47 Å. *Acta Crystallogr., Sect. D : Biol. Crystallogr.*, **49** (Pt 2), 223–233.
- Navizet, I., Cailliez, F. et Lavery, R. (2004a) Probing protein mechanics : residue-level properties and their use in defining domains. *Biophys. J.*, **87** (3), 1426–1435.
- Navizet, I., Lavery, R. et Jernigan, R. L. (2004b) Myosin flexibility : structural domains and collective vibrations. *Proteins*, **54** (3), 384–393.
- Neri, M., Anselmi, C., Cascella, M., Maritan, A. et Carloni, P. (2005) Coarse-grained model of proteins incorporating atomistic detail of the active site. *Phys. Rev. Lett.*, **95** (21), 218102.
- Nicholls, P. (2000) Introduction : the biology of the water molecule. *Cell. Mol. Life Sci.*, **57** (7), 987–992.
- Niessen, C. M. et Gumbiner, B. M. (2002) Cadherin-mediated cell sorting not determined by binding or adhesion specificity. *J. Cell Biol.*, **156** (2), 389–399.
- Noguti, T. et Go, N. (1985) Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. *Biopolymers*, **24** (3), 527–546.
- Nose, A., Tsuji, K. et Takeichi, M. (1990) Localization of specificity determining sites in cadherin cell adhesion molecules. *Cell*, **61** (1), 147–155.
- Oh, B. H., Pandit, J., Kang, C. H., Nikaido, K., Gokcen, S., Ames, G. F. L. et Kim, S. H. (1993) 3-Dimensional structures of the periplasmic Lysine Arginine Ornithine-binding protein with and without a ligand. *J. Biol. Chem.*, **268** (15), 11348–11355.
- Orengo, C. A., Pearl, F. M. G. et Thornton, J. M. (2003) The CATH domain structure database. *Methods Biochem. Anal.*, **44**, 249–271.
- Overduin, M., Harvey, T. S., Bagby, S., Tong, K. I., Yau, P., Takeichi, M. et Ikura, M. (1995) Solution structure of the epithelial cadherin domain responsible for selective cell adhesion. *Science*, **267** (5196), 386–389.
- Ozawa, M., Engel, J. et Kemler, R. (1990) Single amino acid substitutions in one Ca²⁺ binding site of uvomorulin abolish the adhesive function. *Cell*, **63** (5), 1033–1038.
- Ozawa, M. et Kemler, R. (1990) Correct proteolytic cleavage is required for the cell adhesive function of uvomorulin. *J. Cell Biol.*, **111** (4), 1645–1650.
- Patel, S. D., Chen, C. P., Bahna, F., Honig, B. et Shapiro, L. (2003) Cadherin-mediated cell-cell adhesion : sticking together as a family. *Curr. Opin. Struct. Biol.*, **13** (6), 690–698.
- Patel, S. D., Ciatto, C., Chen, C. P., Bahna, F., Rajebhosale, M., Arkus, N., Schieren, I., Jessell, T. M., Honig, B., Price, S. R. et Shapiro, L. (2006) Type II cadherin

- ectodomain structures : implications for classical cadherin specificity. *Cell*, **124** (6), 1255–1268.
- Pearl, F., Todd, A., Sillitoe, I., Dibley, M., Redfern, O., Lewis, T., Bennett, C., Marsden, R., Grant, A., Lee, D., Akpor, A., Maibaum, M., Harrison, A., Dallman, T., Reeves, G., Diboun, I., Addou, S., Lise, S., Johnston, C., Sillero, A., Thornton, J. et Orengo, C. (2005) The CATH domain structure database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. *Nucleic Acids Res.*, **33** (Database issue), D247–D251.
- Perahia, D., Levy, R. M. et Karplus, M. (1990) Motions of an alpha-helical polypeptide-comparison of molecular and harmonic dynamics. *Biopolymers*, **29** (4-5), 645–677.
- Perret, E., Leung, A., Feracci, H. et Evans, E. (2004) Trans-bonded pairs of E-cadherin exhibit a remarkable hierarchy of mechanical strengths. *Proc. Natl. Acad. Sci. U. S. A.*, **101** (47), 16472–16477.
- Perret, E., Leung, A., Morel, A., Feracci, H. et Nassoy, P. (2002) Versatile decoration of glass surfaces to probe individual protein-protein interactions and cellular adhesion. *Langmuir*, **18** (3), 846–854.
- Pertz, O., Bozic, D., Koch, A. W., Fauser, C., Brancaccio, A. et Engel, J. (1999) A new crystal structure, Ca²⁺ dependence and mutational analysis reveal molecular details of E-cadherin homoassociation. *EMBO J.*, **18** (7), 1738–1747.
- Pokutta, S., Herrenknecht, K., Kemler, R. et Engel, J. (1994) Conformational changes of the recombinant extracellular domain of E-cadherin upon calcium binding. *Eur. J. Biochem.*, **223** (3), 1019–1026.
- Ravindranathan, K. P., Gallicchio, E. et Levy, R. M. (2005) Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *J. Mol. Biol.*, **353** (1), 196–210.
- Rohs, R., Etchebest, C. et Lavery, R. (1999) Unraveling proteins : a molecular mechanics study. *Biophys. J.*, **76** (5), 2760–2768.
- Rosenberg, J. M., Seeman, N. C., Day, R. O. et Rich, A. (1976) RNA double helices generated from crystal structures of double helical dinucleoside phosphates. *Biochem. Biophys. Res. Commun.*, **69** (4), 979–987.
- Ryckaert, J. P., Ciccotti, G. et Berendsen, H. J. C. (1977) Numerical-integration of cartesian equations of motion of a system with constraints molecular-dynamics of N-alkanes. *J. Comput. Phys.*, **23** (3), 327–341.
- Sacquin-Mora, S. et Lavery, R. (2006) Investigating the local flexibility of functional residues in hemoproteins. *Biophys. J.*, **90** (8), 2706–2717.

- Schwesinger, F., Ros, R., Strunz, T., Anselmetti, D., Güntherodt, H. J., Honegger, A., Jeremutus, L., Tiefenauer, L. et Pluckthun, A. (2000) Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. *Proc. Natl. Acad. Sci. U. S. A.*, **97** (18), 9972–9977.
- Schwieters, C. D. et Clore, G. M. (2001) Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J. Magn. Reson.*, **152** (2), 288–302.
- Shan, W., Yagita, Y., Wang, Z., Koch, A., Svenningsen, A. F., Gruzglin, E., Pedraza, L. et Colman, D. R. (2004) The minimal essential unit for cadherin-mediated intercellular adhesion comprises extracellular domains 1 and 2. *J. Biol. Chem.*, **279** (53), 55914–55923.
- Shan, W. S., Tanaka, H., Phillips, G. R., Arndt, K., Yoshida, M., Colman, D. R. et Shapiro, L. (2000) Functional cis-heterodimers of N- and R-cadherins. *J. Cell Biol.*, **148** (3), 579–590.
- Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Grübel, G., Legrand, J. F., Als-Nielsen, J., Colman, D. R. et Hendrickson, W. A. (1995) Structural basis of cell-cell adhesion by cadherins. *Nature*, **374** (6520), 327–337.
- Smith, D. A., Brockwell, D. J., Zinober, R. C., Blake, A. W., Beddard, G. S., Olmsted, P. D. et Radford, S. E. (2003) Unfolding dynamics of proteins under applied force. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, **361** (1805), 713–728.
- Smith, J. C. (1991) Protein dynamics comparison of simulations with inelastic neutron-scattering experiments. *Quart. Rev. Biophys.*, **24** (3), 227–291.
- Smith, S. B., Cui, Y. et Bustamante, C. (1996) Overstretching B-DNA : the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, **271** (5250), 795–799.
- Sotomayor, M., Corey, D. P. et Schulten, K. (2005) In search of the hair-cell gating spring elastic properties of ankyrin and cadherin repeats. *Structure*, **13** (4), 669–682.
- Srinivasan, J., Miller, J., Kollman, P. A. et Case, D. A. (1998) Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.*, **16** (3), 671–682.
- Swindells, M. B. (1995) A procedure for the automatic determination of hydrophobic cores in protein structures. *Protein Sci.*, **4** (1), 93–102.

-
- Tajkhorshid, E., Aksimentiev, A., Balabin, I., Gao, M., Isralewitz, B., Phillips, J. C., Zhu, F. et Schulten, K. (2003) Large scale simulation of protein mechanics and function. *Adv. Protein Chem.*, **66**, 195–247.
- Takeichi, M. (1990) Cadherins : a molecular family important in selective cell-cell adhesion. *Annu. Rev. Biochem.*, **59**, 237–252.
- Tama, F., Gadea, F. X., Marques, O. et Sanejouand, Y. H. (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins*, **41** (1), 1–7.
- Tama, F. et Sanejouand, Y. H. (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng.*, **14** (1), 1–6.
- Tamura, K., Shan, W. S., Hendrickson, W. A., Colman, D. R. et Shapiro, L. (1998) Structure-function analysis of cell adhesion by neural N-cadherin. *Neuron*, **20** (6), 1153–1163.
- Teague, S. J. (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.*, **2** (7), 527–541.
- Thomas, A., Hinsen, K., Field, M. J. et Perahia, D. (1999) Tertiary and quaternary conformational changes in aspartate transcarbamylase : a normal mode study. *Proteins*, **34** (1), 96–112.
- Thornton, J. M., Todd, A. E., Milburn, D., Borkakoti, N. et Orengo, C. A. (2000) From structure to function : approaches and limitations. *Nat. Struct. Biol.*, **7 Suppl**, 991–994.
- Tirion, M. M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, **77** (9), 1905–1908.
- Tomschy, A., Fauser, C., Landwehr, R. et Engel, J. (1996) Homophilic adhesion of E-cadherin occurs by a co-operative two-step interaction of N-terminal domains. *EMBO J.*, **15** (14), 3507–3514.
- Tsui, V. et Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers*, **56** (4), 275–291.
- Vieille, C., Burdette, D. S. et Zeikus, J. G. (1996) Thermozyms. *Biotechnol. Ann. Rev.*, **2**, 1–83.
- Villa, E., Balaeff, A. et Schulten, K. (2005) Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proc. Natl. Acad. Sci. U. S. A.*, **102** (19), 6783–6788.
- Voet, D. et Voet, J. (1998) *Biochimie*. De Boeck Université.

- Vriend, G. (1990) WHAT IF : a molecular modeling and drug design program. *J. Mol. Graph.*, **8** (1), 52–56.
- Várnai, P. et Zakrzewska, K. (2004) DNA and its counterions : a molecular dynamics study. *Nucleic Acids Res.*, **32** (14), 4269–4280.
- Wallace, A. C., Borkakoti, N. et Thornton, J. M. (1997) TESS : a geometric hashing algorithm for deriving 3D coordinate templates for searching structural databases. application to enzyme active sites. *Protein Sci.*, **6** (11), 2308–2323.
- Wallace, A. C., Laskowski, R. A. et Thornton, J. M. (1996) Derivation of 3D coordinate templates for searching structural databases : application to Ser-His-Asp catalytic triads in the serine proteinases and lipases. *Protein Sci.*, **5** (6), 1001–1013.
- Wang, J. M., Cieplak, P. et Kollman, P. A. (2000) How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21** (12), 1049–1074.
- Warshel, A. et Levitt, M. (1976) Theoretical studies of enzymic reactions : dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **103** (2), 227–249.
- Weiser, J., Shenkin, P. S. et Still, W. C. (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps (lcpo). *J. Comput. Chem.*, **20** (2), 217–230.
- Wriggers, W. et Schulten, K. (1997) Protein domain movements : detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. *Proteins*, **29** (1), 1–14.
- Yang, J., Dokurno, P., Tonks, N. K. et Barford, D. (2001) Crystal structure of the M-fragment of alpha-catenin : implications for modulation of cell adhesion. *EMBO J.*, **20** (14), 3645–3656.
- Yang, L. W. et Bahar, I. (2005) Coupling between catalytic site and collective dynamics : a requirement for mechanochemical activity of enzymes. *Structure*, **13** (6), 893–904.
- Yap, A. S., Briehner, W. M., Pruschy, M. et Gumbiner, B. M. (1997) Lateral clustering of the adhesive ectodomain : a fundamental determinant of cadherin function. *Curr. Biol.*, **7** (5), 308–315.
- York, D. M., Darden, T. A. et Pedersen, L. G. (1993) The effect of long-range electrostatic interactions in simulations of macromolecular crystals. A comparison of the Ewald and truncated list methods. *J. Chem. Phys.*, **99** (10), 8345–8348.
- Yuan, Z., Zhao, J. et Wang, Z. X. (2003) Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng.*, **16** (2), 109–114.

- Zacharias, M. (2003) Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.*, **12** (6), 1271–1282.
- Zhu, B., Chappuis-Flament, S., Wong, E., Jensen, I. E., Gumbiner, B. M. et Leckband, D. (2003) Functional analysis of the structural basis of homophilic cadherin adhesion. *Biophys. J.*, **84** (6), 4033–4042.
- Závodszky, P., Kardos, J., Svingor et Petsko, G. A. (1998) Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **95** (13), 7406–7411.
- Åqvist, J. (1990) ion-water interaction potentials derived from free-energy perturbation simulations. *J. Phys. Chem.*, **94** (21), 8021–8024.