

Université de Montréal

**Développement d'une méthode de simulation multi-échelle pour l'étude des  
grandes transformations dans les protéines**

par  
Lilianne Dupuis

Département de Biochimie  
Faculté de Médecine

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)  
en Bio-informatique

13 décembre, 2011

© Lilianne Dupuis, 2011.

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée:

**Développement d'une méthode de simulation multi-échelle pour l'étude des  
grandes transformations dans les protéines**

présentée par:

Lilianne Dupuis

a été évaluée par un jury composé des personnes suivantes:

Hervé PHILIPPE,	président-rapporteur
Normand MOUSSEAU,	directeur de recherche
Rémi SAUVÉ,	membre du jury
Guillaume LAMOUREUX,	examineur externe
Michel DESJARDINS,	représentant du doyen de la FES

Thèse acceptée le: .....



## RÉSUMÉ

Les protéines accomplissent leur fonction dans la cellule grâce à leur faculté de changer de forme. Chaque classe de protéines peut se caractériser par une structure spécialisée partagée par ses membres avec un certain degré de variabilité. Tel est le cas des protéines à motifs mains-EF, qui se transforment en liant et déliant l'ion calcium. Ce motif permet à la Troponin C de s'ouvrir et se refermer afin de moduler le mécanisme de contraction des fibres musculaires. Un mécanisme similaire permet à la Calmoduline de gérer l'activité de divers canaux cellulaires.

Les techniques de simulations numériques peuvent aider à comprendre les trajectoires de ces transformations. Le projet principal de cette thèse consistait à développer une méthode informatique multi-échelle permettant de simuler des mouvements complexes à l'intérieur d'une protéine. La représentation multi-échelle développée peut changer et s'adapter en cours de simulation. La méthode, ART holographique, explore l'espace en générant des basculements d'ensembles atomiques, selon des champs de force atomistiques non biaisés indiquant à tout moment comment les ensembles doivent pivoter. La méthode réduit le calcul des fluctuations locales mais conserve une représentation spatiale complète.

La représentation multi-échelle est combinée à une technique de recherche de passages de transition énergétiquement favorables, ART nouveau, qui conduit la trajectoire moléculaire d'étape en étape. Appliquée à plusieurs protéines, dont la Calmodulin et la Troponin C, ART holographique génère des trajectoires de transformation entre des conformations distantes de celles-ci, déjà connues grâce aux techniques de RMN ou de cristallographie.

L'usage d'une représentation spatiale complète tout au long de la simulation favorise le discernement de certains détails des mécanismes. Le rôle, l'ordre d'intervention, ainsi que la coopérativité de certains résidus et structures impliqués dans le mécanisme des paires main-EF ont été explorés plus en détail et un état intermédiaire est proposé.

**Mots clés : Simulation, multi-échelle, protéines, coopérativité, transitions, mouvements d'ensemble, technique d'activation relaxation**

## ABSTRACT

Proteins accomplish their function inside cells by means of conformational changes. Each protein class may be characterized by a specialized structure shared by its members with some variability. EF-hands proteins present a special motif which transforms itself while binding or unbinding the calcium ion. This structure allows Troponin C domains to open and close as it modulates the muscular fibers contraction. A similar mechanism allow Calmodulin to manage the activity of a diversity of protein channels.

Computational techniques may help discover how these transformations occur. The main project of this thesis was the development of a multi-scale computational method for the simulation of complex motions inside a protein. The multi-scale approach is designed to adapt and change all along the simulation. The method, holographic ART, explore conformational space by generating swiveling and rotation of atomic ensembles, leaded by non biased atomistic forcefields. This determines at each step the overall motion, keeping a complete spatial representation, but with minimal local fluctuations computation.

The multi-scale representation is combined with a unbiased open ended algorithm for identifying transitions states, ART nouveau, which guides the molecular trajectory from state to state. Applied to several proteins, the method was able to generate transformation trajectories between distant conformations known from NMR and crystallography techniques.

The use of a complete spatial representation throughout the simulation allows the method to capture atomistic details of each event. The purpose, the intervention order, as well as cooperativity between some residues and sub-structures involved in the EF-hand pair mechanism have been explored more in detail and an intermediate state is proposed.

**Keywords: simulation, multi-scale, protein, cooperativity, transition, ensemble motions, activation relaxation technique**

## TABLE DES MATIÈRES

<b>RÉSUMÉ</b> . . . . .	<b>iii</b>
<b>ABSTRACT</b> . . . . .	<b>iv</b>
<b>TABLE DES MATIÈRES</b> . . . . .	<b>v</b>
<b>LISTE DES TABLEAUX</b> . . . . .	<b>ix</b>
<b>LISTE DES FIGURES</b> . . . . .	<b>x</b>
<b>LISTE DES ANNEXES</b> . . . . .	<b>xiii</b>
<b>LISTE DES SIGLES</b> . . . . .	<b>xiv</b>
<b>DÉDICACE</b> . . . . .	<b>xv</b>
<b>REMERCIEMENTS</b> . . . . .	<b>xvi</b>
<b>CHAPITRE 1 : INTRODUCTION GÉNÉRALE</b> . . . . .	<b>1</b>
<b>CHAPITRE 2 : INTRODUCTION AUX PROTÉINES</b> . . . . .	<b>5</b>
2.1 Hiérarchie de la structure protéique . . . . .	5
2.1.1 Le niveau primaire . . . . .	5
2.1.2 Le niveau secondaire . . . . .	7
2.1.3 Le niveau tertiaire . . . . .	7
2.1.4 L'encodage du repliement . . . . .	11
2.1.5 Le niveau quaternaire . . . . .	13
2.2 La dynamique des protéines . . . . .	14
2.2.1 Études in vitro . . . . .	15
2.2.2 Paysage énergétique . . . . .	17
2.2.3 Jeu entre les niveaux . . . . .	21

2.2.4	Niveaux d'organisation . . . . .	23
2.2.5	Limitations des méthodes expérimentales . . . . .	24
<b>CHAPITRE 3 : MÉTHODES DE SIMULATION . . . . .</b>		<b>26</b>
3.1	Stratégies d'accélération basées sur le potentiel . . . . .	27
3.2	Stratégies d'accélération basées sur l'échantillonnage . . . . .	30
3.3	Hiérarchie et combinaisons de stratégies . . . . .	33
3.3.1	Hiérarchies structurales . . . . .	34
3.3.2	Combinaison de techniques pour l'échantillonnage . . . . .	37
3.4	Problématiques et questions ouvertes . . . . .	38
<b>CHAPITRE 4 : INTRODUCTION AUX PROTÉINES À MAINS EF . . . .</b>		<b>43</b>
4.1	Le motif main-EF . . . . .	44
4.2	Domaines à mains EF . . . . .	45
4.3	Études informatiques . . . . .	48
<b>CHAPITRE 5 : STRATÉGIE ET ORGANISATION . . . . .</b>		<b>50</b>
5.1	Vision stratégique . . . . .	50
5.1.1	Directions de l'espace, coopérativité et étages . . . . .	50
5.1.2	Les blocs et leur utilisation . . . . .	52
5.1.3	Les coordonnées internes . . . . .	52
5.1.4	<i>ART nouveau</i> . . . . .	53
5.1.5	Stratégies développées dans ce projet doctoral . . . . .	54
5.2	Organisation du logiciel . . . . .	55
5.3	Niveau holographique . . . . .	61
<b>CHAPITRE 6 : TRAVAIL AVEC LES POTENTIELS ATOMIQUES . . . .</b>		<b>66</b>
6.1	Intégration des potentiels à la méthode HMM . . . . .	66
6.2	Les potentiels à solvant implicite . . . . .	68
6.3	OPEP . . . . .	69
6.4	<i>EOPEP</i> . . . . .	72

6.5	Travaux d'implantation de CHARMM19/EFF1 . . . . .	75
6.6	Bilan . . . . .	76
<b>CHAPITRE 7 : LA MÉTHODE HOLOGRAPHIQUE . . . . .</b>		<b>77</b>
7.1	Introduction . . . . .	78
7.2	Methodology . . . . .	80
7.2.1	Basic protein motion . . . . .	80
7.2.2	Overview of the holographic multiscale method . . . . .	81
7.2.3	Construction of elastic blocks and projected flexible regions . . . . .	82
7.2.4	Sampling method . . . . .	85
7.2.5	Forcefields . . . . .	86
7.2.6	Description of a typical holographic ART transition event . . . . .	88
7.3	Protein models . . . . .	90
7.4	Results and Discussion . . . . .	91
7.5	Conclusions and perspectives . . . . .	92
<b>CHAPITRE 8 : ÉTUDE DE PROTÉINES À MAINS EF . . . . .</b>		<b>101</b>
8.1	Introduction . . . . .	103
8.2	Methodology . . . . .	105
8.2.1	ART . . . . .	106
8.2.2	Holographic Multiscale Algorithm . . . . .	106
8.2.3	Forcefields . . . . .	107
8.2.4	Sequences studied . . . . .	109
8.2.5	Analysis . . . . .	110
8.3	Results and Discussion . . . . .	111
8.3.1	Simulations on calmodulin NT . . . . .	111
8.3.2	Simulations on troponinC NT . . . . .	128
8.4	Conclusions and perspectives . . . . .	131
8.5	Acknowledgements . . . . .	132

<b>CHAPITRE 9 : DISCUSSION ET CONCLUSION</b>	<b>133</b>
9.1 Bilan pour les Mains EF	135
9.1.1 Travaux futurs pour les mains EF	136
9.1.2 Validations expérimentales	137
9.2 Bilan pour ART holographique	140
9.2.1 Échantillonnage	141
9.2.2 Transformations dans les boucles avec extrémités fixes	143
9.2.3 Développement des potentiels	143
9.3 Perspectives d'application	144
9.3.1 Mutagénèse et modifications post-traductionnelles	145
9.3.2 Diversité moléculaire	146
9.3.3 Complexes moléculaires	147
9.3.4 Repliement	147
9.4 Épilogue	148
<b>BIBLIOGRAPHIE</b>	<b>150</b>

## LISTE DES TABLEAUX

3.I	Buts principaux selon les stratégies de simulation. . . . .	39
4.I	Études de protéines à mains-EF par simulation. . . . .	49
7.I	EOPEP parameters for side-chain atoms. . . . .	88
7.II	Mean RMSD between events : statistics for a few simulation examples.	90
8.I	Calmodulin and troponin C : RMSD statistics for selected trajectories. . . . .	113

## LISTE DES FIGURES

2.1	Formation de la chaîne primaire protéique. . . . .	6
2.2	Formation d'une structure secondaire protéique. . . . .	8
2.3	L'hélice $\alpha$ . . . . .	9
2.4	Formation $\beta$ anti-parallèle en épingle. . . . .	10
2.5	Formation $\beta$ parallèle en hélice. . . . .	10
2.6	Assemblage tertiaire. . . . .	12
2.7	Assemblage quaternaire de quatre instances d'une protéine. . . . .	14
2.8	La troponine C englobant un domaine de la troponin I en présence d'ions calcium. . . . .	14
2.9	Les motifs répétés de l'ankyrine. . . . .	19
2.10	Paysages énergétiques. . . . .	20
3.1	Niveaux des simulations multi-échelle. . . . .	35
4.1	Schéma d'une main EF. . . . .	45
4.2	Protéines à mains-EF. . . . .	46
5.1	Les quatre niveaux de <i>ART holographique</i> . . . . .	57
5.2	Nature des mouvements amples dans une protéine. . . . .	58
5.3	Passage d'une conformation hélice $\alpha$ vers feuillet $\beta$ ou l'inverse. . . . .	60
5.4	Coordonnées relatives. . . . .	61
5.5	Déroulement des simulations avec ART nouveau. . . . .	62
5.6	Tableaux des positions atomiques cartésiennes et des forces ato- miques cartésiennes. . . . .	62
5.7	Translation entre les niveaux de potentiel atomistique et <i>HMM</i> . . . . .	63
5.8	Chevauchement des blocs holographiques. . . . .	64
6.1	Intégration des potentiels atomistiques dans <i>HMM</i> . . . . .	67
6.2	Un trio holographique. . . . .	67
6.3	Modèle <i>OPEP</i> . . . . .	70



6.4	ASP20 dans la calmoduline apo. . . . .	73
7.1	Relative peptidic planes. . . . .	95
7.2	Flexible C $\alpha$ pivots. . . . .	95
7.3	Software organization. . . . .	96
7.4	A typical HMM/ART event during an EOPEP closing simulation of Calmodulin NT. . . . .	96
7.5	Extended protein A with 3 helices aligned compared to native model.	97
7.6	Protein A at end of EOPEP simulation 20, compared with native model. . . . .	97
7.7	Protein G with beta sheet widely opened, compared to native model.	97
7.8	Amino acid sequences of the protein A and protein G domains used in our simulations. . . . .	98
7.9	24 holographic simulations starting from the artificially open mo- del of protein A. . . . .	99
7.10	48 holographic simulations starting from the artificially open mo- del of protein G. . . . .	100
8.1	Sequences for the NT domains of calmodulin and troponin C. . . . .	109
8.2	Simulations of apo calmodulin NT model. . . . .	112
8.3	Closing simulations of calmodulin NT model. . . . .	115
8.4	Closing pathway for simulation eop9h. . . . .	116
8.5	Diagrammatic representation of the calmodulin NT closing me- chanism. . . . .	118
8.6	Evolution of the energy as a function of accepted ART events du- ring eop9h simulation. . . . .	121
8.7	Evolution of angles between helices. . . . .	123
8.8	Comparison between open and close models, calmodulin NT and troponin C NT. . . . .	124
8.9	Extension of the 24 EOPEP calmodulin NT closing simulations. . . . .	126

8.10	Cartoon representation of the PHE19 as it enters and leaves the intermediate pocket. . . . .	127
8.11	Closing simulations of troponin C NT. . . . .	129
9.1	Deux étapes d'une simulation avec ajout du calcium au modèle fermé du domaine NT de la Calmoduline. . . . .	137
9.2	Étapes principales de 24 simulations <i>HMM-EOP</i> pour la fermeture de la calmoduline. . . . .	144
9.3	Exemples de modifications post-traductionnelles. . . . .	146
9.4	Dimères de protéines. . . . .	148

**LISTE DES ANNEXES**

**Annexe I : Curriculum vitae . . . . . xix**

## LISTE DES SIGLES

apo	Conformation native et sans ligands d'une protéine
ART	Activation Relaxation Tehcnique
CHARMM	Chemistry at HaRvard Macromolcular Mechanics
CT	C-terminal
EEF1	Effective Energy Function for Proteins in Solution
EOPEP	Extended version of OPEP with atomistic side-chains
HMM	Méthode holographique multi-échelle
holo	Conformation d'une protéine en présence d'un ligand
NT	N-terminal
OPEP	Optimized Potential for Efficient Structure Prediction
RMN	Résonnance magnétique nucléaire

## Dédicace

À la mémoire de ma mère Jacqueline Lafrance,

À mon père Antonio Dupuis,

À mes frères Marcel et Jacques,

À mes enfants Jean-Marc, Marie-Ève et Sébastien, et à leur père Daniel

## REMERCIEMENTS

Après plusieurs années de pratique comme conceptrice de systèmes informatiques, puis plusieurs années de pratique de l'art parental, je me suis intéressée à la biochimie. Je souhaitais surtout comprendre la mécanique des macro-molécules. J'ai découvert avec plaisir que mon bac en informatique m'offrait un chemin : l'université de Montréal proposait un nouveau programme de maîtrise en bio-informatique.

Mes premiers remerciements vont d'emblée au professeur Normand Mousseau, mon directeur de thèse. Les recherches du professeur Normand Mousseau, en physique des matériaux complexes, correspondaient le plus à ce que je recherchais. L'ouverture de M. Mousseau envers la multidisciplinarité l'incite à accueillir dans son équipe non seulement des physiciens et biophysiciens, mais aussi des chimistes, et des bio-informaticiens comme moi. Je me suis donc embarquée dans une année préparatoire à l'hiver 2004, puis j'ai débuté une maîtrise en bio-informatique, suivie d'un passage direct au doctorat à l'automne 2006. La largeur de vision de Normand Mousseau ne l'empêche pas d'exercer un suivi bien aiguïté de ceux qu'il guide. Il est très disponible, même à l'étranger ou en congé (alors que j'étais en fin de thèse), grâce aux télé-technologies. Je ne me suis pas sentie restreinte dans la réalisation de mes idées. M. Mousseau est à la fois capable de nous laisser assez de liberté pour suivre nos élans créatifs, et de rectifier notre trajectoire quand elle devient inefficace. Il ne compatit pas avec nos faiblesses. Il donne des directives pour les dépasser, car son but est plus important que de seulement enrichir le bilan des accomplissements de l'équipe. Il s'assure que ses étudiants soient autonomes et compétents lorsqu'ils le quittent.

Je dois ajouter qu'en intégrant le groupe de Normand Mousseau, j'ai bénéficié d'une méthode avant-gardiste qu'il a mis au point avec ses collaborateurs, *ART nouveau*. Bien qu'elle soit déjà beaucoup appliquée, je la perçois comme une technique d'avenir qui commence seulement à déployer ses possibilités. Merci aussi à Philippe Derreumaux en France, car j'ai beaucoup apprécié le travail qu'il a accompli en créant le potentiel *OPEP*, qui s'est avéré important dans le cadre de ma recherche. Mes remerciements vont aussi à M. Karplus et al. pour leur oeuvre avec le potentiel CHARMM19 et le modèle de

solvation EEF1, aptes à être utilisés pour de longues trajectoires.

Je remercie l'équipe qui m'a accueillie, ainsi que ceux qui sont venus s'ajouter au fil des ans. Parties à l'étranger, Fedwa El Mellouhi et Rozita Laghaei sont demeurées mes amies. J'ai poursuivi les diners autour de notre table pliante avec Jean-François St-Pierre, Jean-François Joly, Jessica Nasica, Sébastien Côté, Laurent Karim Béland, Ali Kerrache, Peter Brommer et Saïd Bouzakraoui. J'ai vu moins souvent d'autres membres de l'équipe de physique mais bien apprécié leur conversation, dont Wei Chen, Kevin Levasseur-Smith et Jonathan Duquette. Je me souviens aussi de mes coéquipiers de mes débuts en biochimie et bio-informatique, dont Guy Larochelle, Stéphanie Côté, Coralie Lafontant et bien sûr Véronique Lisi, qui m'a bien encouragée lors de mes débuts un peu traumatisants en présentation scientifique.

Chez moi j'ai toujours eu le support et l'attention de mes enfants, Jean-Marc, Marie-Ève et Sébastien. La collaboration avec leur père, Daniel, a beaucoup aidé. Mes parents et mes frères ont beaucoup été là pour moi. Mon frère Jacques est souvent venu m'aider à la maison, et mon père a effectué des travaux, même passé 80 ans.

Un grand merci aux nombreux professeurs de biochimie et de bio-informatique qui m'ont informée et formée durant les années qui ont mené au doctorat et les premières de celui-ci. Une mention spéciale à la professeure Gertraud Burger, qui a bien voulu m'accueillir dans le programme, pour son engagement envers la bio-informatique. Je reculerais aussi jusqu'en 1982, au BAC en informatique à Sherbrooke, où le professeur Jacques Hagel m'a communiqué la pique de l'analyse informatique. Merci aussi à Mmes Marie Pageau, Marie Robichaud et Elaine Meunier pour leur support technique, administratif et moral. Merci encore à Mme Marie Pageau, M. Serguei Chteinberg, M. Normand Mousseau, Mme Pascale Legault, M. James Omichinski, M. Le Van Nguyen et M. Yves Claude, pour leur confiance lors de mes assistanats pour leur cours de biochimie, de bio-informatique ou d'informatique.

Je remercie mon jury pré-doctoral, qui a accepté que je poursuive. M. Normand Mousseau bien sûr, mais aussi les professeurs Rémy Sauvé, Serguei Chteinberg et Hervé Philippe (président-rapporteur). J'ai aussi eu la chance de recevoir les enseignements de MM. Philippe et Chteinberg. Un grand merci à mon jury de thèse : MM. Normand Mous-

seau, Hervé Philippe, Rémi Sauvé, Michel Desjardins et à M. Guillaume Lamoureux de Concordia, examinateur externe.

Les analystes de Calcul Québec ont été d'une aide précieuse pour faciliter l'utilisation des ressources de calcul de haute performance pour les simulations. Merci beaucoup à Jacques Richer pour son aide avec openMP et les infrastructures Latex de la thèse.

Cette recherche a été possible grâce aux financements du programme de bourse de bio-informatique BiT, du NSERC, du FQRNT et la fondation canadienne de chaires de recherche.

Cette aventure multidisciplinaire a nourri mon grand intérêt envers les niveaux d'organisation de la matière et leurs inter-relations. Beaucoup d'efforts concernaient aussi mon développement personnel : apprendre à m'ouvrir aux différents modèles de pensée, à combiner plusieurs disciplines. Le plus difficile était probablement de développer la capacité de structurer, affirmer et communiquer mes idées. Ces années m'ont permis de discerner mes habilités et mon rôle dans ce que je perçois maintenant comme un grand travail d'équipe. Celui-ci a débuté bien avant moi, au milieu du 20<sup>e</sup> siècle. Je me vois comme une bio-informaticienne spécialisée dans la dynamique des structures tridimensionnelles. Mon principal intérêt est le développement de méthodes informatiques permettant d'approfondir les connaissances en biochimie et en biophysique dans ce domaine. J'ai écrit cette thèse en évitant dans la mesure du possible les jargons spécifiques aux informaticiens, aux physiciens ou aux biochimistes. J'ai tenu à écrire pour les informaticiens travaillant en science et les biochimistes en général, et non pas seulement pour ceux qui combinent les deux disciplines. Il y a donc une section qui explique la base de la structure des protéines. Mon souhait est que cette thèse représente une intégration multidisciplinaire fertile. J'espère vraiment qu'elle constitue un passage qui conduit à de nouvelles possibilités.



## CHAPITRE 1

### INTRODUCTION GÉNÉRALE

Les protéines furent d'abord connues comme une classe de grandes molécules de composition semblable et furent nommées du grec "prota" (première importance) parce qu'elles s'avéraient essentielles dans l'alimentation (par Jons Jakob Berzelius en 1838) [270]. Par la suite, on a réussi à en obtenir des cristaux [25, 250].

En 1936, à partir de l'hémoglobine et de l'oxygène, Pauling et Coryell démontrent que la structure d'une protéine peut changer au contact d'un ligand [210]. Dès lors on pouvait imaginer l'importance de comprendre les liens entre la flexibilité d'une protéine, sa capacité à changer de forme, et son aptitude à servir de mécanisme pour l'accomplissement d'une fonction. Des années plus tard, Linus Pauling imagine en pliant une feuille de papier ce que pourrait être l'architecture de l'hélice  $\alpha$ , un des principaux motifs architecturaux des protéines. En 1951, il propose avec Robert B. Corey et H. R. Branson les modèles architecturaux des structures hélices  $\alpha$  et feuillets  $\beta$ , qui s'avèrent prouvés en laboratoire par la suite [209]. On appelle maintenant ces motifs structures secondaires de la protéine. Ils se construisent à partir d'une chaîne primaire d'acides aminés, codée par les gènes. Le milieu du siècle dernier vit aussi l'apparition de techniques permettant une détermination de plus en plus complète de ces séquences [227]. Le début du chapitre 2 donnera un aperçu des niveaux de structures protéiques pour les non-initiés.

Les protéines permettent à la cellule de se structurer et effectuent pour elle une grande diversité de fonctions. Chacune accomplit son action grâce à la flexibilité de certaines régions internes, qui lui permettent de changer de forme. Beaucoup d'entre elles peuvent ainsi accomplir un travail mécanique. Dans le cas des protéines enzymatiques, des changements de conformation permettent de positionner des acides aminés réactifs dont l'action est de couper ou assembler d'autres protéines ou molécules. Les protéines peuvent ainsi s'adapter entre elles, ou se lier à d'autres types de molécules pour former des structures de soutien, des rails de transport, de glissières mécaniques, ou servir de canaux de transfert à travers la surface de la cellule ou de ses organelles. De longues

structures protéiques permettront par exemple la séparation des chromosomes lors de la mitose cellulaire, ou encore assureront la croissance des axones selon les signaux reçus par des canaux. Beaucoup de protéines font office de signal, de clef ou de mécanisme pour modifier l'opération de celles qui forment des canaux ou des glissières. Les protéines excrétées formeront des structures extra-cellulaires tels le fil d'araignée, la soie du ver, les cheveux, des flagelles de toutes sortes, les os et les cartilages à base de protéines collagènes, ou des organes de propulsion telle la flagelle du spermatozoïde.

Pour comprendre le fonctionnement d'une protéine on doit connaître les conformations possibles que chacune peut adopter et les chemins de transition entre ces conformations. Les méthodes expérimentales telles que la RMN et la cristallographie ont fourni ces dernières décennies un grand nombre de modèles tridimensionnels de protéines. La cristallographie [28] cible des protéines moyennes ou volumineuses, d'une certaine stabilité, aptes à former des cristaux. La structure est déterminée à partir de l'étude de la diffraction de rayons X traversant les cristaux. Découverte en 1946 par Felix Bloch et Edouard Mills Purcell, la spectroscopie *RMN* (résonance magnétique nucléaire) cible généralement des protéines plus petites et plus flexibles, en milieu aqueux [111]. Elle utilise un champ magnétique externe pour discerner les relations entre certains types d'atome et déduire la structure de la protéine et sa fluctuation. Pour plusieurs protéines nous avons ainsi accès à leur forme dite native, sans ligand, ainsi qu'à certaines formes stables qu'elles adoptent lorsqu'elles sont en contact avec des ions, protéines ou autres molécules. L'étude de la composition en acides aminés [231] aide à déterminer les domaines mobiles d'une molécule. Dans plusieurs cas, la mutagénèse dirigée [205] et la fluorescence par résonance [242] permettent de connaître quels domaines des protéines se rapprochent au cours de leur travail et interagissent entre eux.

Les chemins de transition sont plus complexes à découvrir. La quête d'une image plus raffinée des mouvements internes a amené le développement des méthodes de simulation informatique. La protéine est complexe et comporte plusieurs niveaux d'organisation qui fonctionnent à différents rythmes vibratoires. La simulation complète d'un mécanisme moléculaire peut devenir très coûteuse en temps de calcul. Beaucoup s'investissent donc dans la création de techniques permettant d'accélérer le calcul de ces

mouvements. La stratégie principale du travail décrit dans cette thèse a été d'utiliser l'organisation hiérarchique des protéines de manière à générer des mouvements de grande envergure, en évitant les longs calculs liés aux fluctuations locales. La représentation multi-échelle a été conçue pour faciliter la génération de mouvements d'ensemble à partir des calculs de champs de forces atomistiques non biaisés. Elle a été appliquée à la simulation de transformations moléculaires comportant des états intermédiaires.

Cette brève introduction est suivie de trois chapitres d'introduction plus élaborés. Le chapitre 2 explique les caractéristiques structurales et dynamiques des protéines utiles au développement de cette thèse. Alors que son début est consacré aux non-initiés du domaine protéique, le chapitre explore certaines propriétés protéiques déterminantes pour le développement de la thèse.

Le chapitre 3 présente un tableau des principales stratégies informatiques permettant leur étude. La complexité des protéines nécessite plusieurs approches. Ce chapitre permet de situer la présente recherche dans cet éventail d'efforts collectifs et de justifier son utilité.

Le chapitre 4 introduit pour sa part les protéines à mains-EF, dont certaines ont fait l'objet d'une application de la méthode développée dans cette thèse. Nous donnons ici un aperçu des études précédentes.

Une description des stratégies de la technique développée (*holographic multi-scale method* ou *HMM*) fait l'objet du chapitre 5, qui explique aussi l'organisation générale du logiciel. Le chapitre 6 est consacré aux travaux liés à l'implantation des champs de forces atomistiques dans *HMM*, en particulier à *EOPEP*, une extension tout atome du potentiel gros grain *OPEP* [171] qu'il s'est avéré utile de développer pour l'étude des protéines choisies. Le premier article sur la méthode multiéchelle holographique (*HMM*) constitue le chapitre 7. On y voit que la méthode a été mise à l'épreuve via des déformations distantes de la conformation native de protéines à hélice  $\alpha$  et feuillet  $\beta$ . Le but était de vérifier l'efficacité de la méthode à utiliser des forces et énergies non-biasées pour générer des transformations conduisant à des configurations similaires aux conformations natives.

Une étude détaillée de la transformation des domaines N-Terminaux des protéines

Calmoduline et Troponine C a été réalisée comme application plus concrète de la méthode. Ces molécules ont été choisies parce que beaucoup d'études expérimentales existent à leur sujet, ce qui nous permettait une évaluation des forces et des faiblesses de la méthode ainsi que des potentiels informatiques utilisés. L'analyse des simulations qui ont généré une trajectoire entre les formes connues expérimentalement a permis de discerner davantage le rôle de certains acides aminés dans un enchainement de mécanismes de nature non harmonique, ainsi que la coopérativité entre certains groupes. Cette étude a été réalisée avec deux potentiels atomiques différents, *EOPEP* et *CHARMM19* [192] joint au modèle de solvation *EEFI* [145]. L'article fait l'objet du chapitre 8.

Finalement, nous discutons des progrès apportés par le projet et des lacunes sur lesquelles travailler, pour conclure avec les applications possibles de la méthode et les nouveaux buts à atteindre.

## CHAPITRE 2

### INTRODUCTION AUX PROTÉINES

Ce travail participe à un vaste effort collectif qui a débuté vers le milieu du siècle dernier avec les découvertes des expérimentateurs et théoriciens des protéines. Nous avons hérité d'eux un éventail de notions sur la structure et la dynamique énergétique de la protéine. Nous allons inventorier ici les connaissances qui seront utilisées pour cette thèse, en nous attardant sur les travaux qui apportent des concepts clefs pour celle-ci. Ce parcours se fait selon un ordonnancement logique menant aux motivations de la présente recherche.

La première section est une base utile aux non-initiés du domaine de la structure des protéines. Les habitués peuvent poursuivre à la section 2.2.

#### 2.1 Hiérarchie de la structure protéique

Pour comprendre comment une protéine bouge, il faut d'abord connaître son organisation, qui s'établit en plusieurs niveaux. Des forces différentes prédominent à chaque niveau.

##### 2.1.1 Le niveau primaire

La protéine est formée par le ribosome [218], qui utilise le code génétique pour guider l'assemblage de la chaîne d'acides aminés qui la compose. Les acides aminés ont en commun un carbone, dit  $C\alpha$ , ainsi qu'un petit groupe d'atomes permettant de former des liens dit *peptidiques* entre ces carbones  $\alpha$ . La jonction peptidique est planaire. Ces jonctions permettent de former une longue chaîne linéaire, bordée tout le long de courtes chaînes latérales. Il existe vingt sortes d'acides aminés et donc vingt sortes de chaînes latérales, qui confèrent à chacun leur spécificité (Voir Fig. 2.1) .

Seules les forces covalentes président à la formation de la chaîne primaire et cet assemblage se produit à l'intérieur du tunnel du ribosome [158]. Celui-ci pourrait aussi

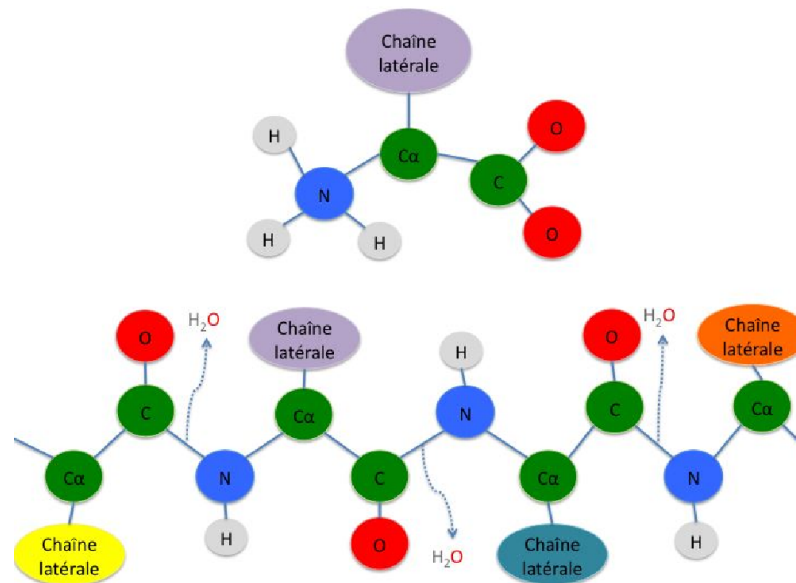


Figure 2.1 – Formation de la chaîne primaire protéique. En haut : Un acide aminé non relié. En bas : une chaîne de quatre acides aminés. Un lien peptidique se forme entre deux acides aminés grâce à la production d’une molécule d’eau. La suite de carbones  $\alpha$  reliés par des liens peptidiques forme la chaîne principale de la protéine. Les protéines diffèrent par leur nombre d’acides aminés et par la diversité des chaînes latérales. La chaîne d’acides aminés constitue la structure primaire de la protéine. À noter : l’extrémité de la chaîne commençant par un azote est dite N-terminale, par opposition à l’autre qui est dite C-terminale. Le ribosome commence à construire la chaîne à partir du N-terminal.

favoriser une certaine tendance tri-dimensionnelle dans la chaîne à sa sortie, en vue de la formation des structures secondaires.

### 2.1.2 Le niveau secondaire

La chaîne principale de la protéine peut se replier sur elle même pour former des structures secondaires. Ces structures sont basées sur des ponts hydrogènes se formant entre un azote d'une jonction peptidique et un oxygène d'une autre jonction peptidique qui se partagent un proton (Fig. 2.2) Les jonctions peptidiques agissent donc comme des carrefours, formant dans un sens les liens covalents entre les acides aminés et dans l'autre sens des ponts hydrogènes avec d'autres jonctions peptidiques. Deux types de structures régulières sont très fréquentes : les hélices  $\alpha$  (Fig. 2.3) et les feuillets  $\beta$  (Fig. 2.4 et 2.5). Les premières se constituent grâce à des ponts hydrogènes s'établissant entre les oxygènes d'une série consécutive de jonctions peptidiques et les azotes de chaque jonction peptidique située quatre acides aminés plus loin. Les feuillets  $\beta$  se constituent lorsque deux ou plusieurs sections de la chaîne principale se retrouvent côte à côte. Des ponts hydrogènes peuvent ainsi se constituer avec une section de chaîne voisine et même deux sections de chaînes situées de part et d'autre. Les sections de chaîne qui se longent ainsi peuvent appartenir à la même protéine ou à différentes protéines. Selon le nombre de lignes parallèles (ou anti-parallèles) le feuillet sera de plus en plus large. Parce qu'il présente une courbure, il peut même se refermer sur lui-même pour former un baril  $\beta$ . Dans une protéine, plusieurs structures secondaires peuvent coexister, reliées entre elles par des boucles moins structurées.

### 2.1.3 Le niveau tertiaire

Les éléments de structures secondaires forment entre eux un assemblage qu'on nomme structure tertiaire. Ce troisième niveau ne provient généralement pas de la chaîne principale, mais plutôt de l'interaction entre les chaînes latérales. Celles-ci sont relativement

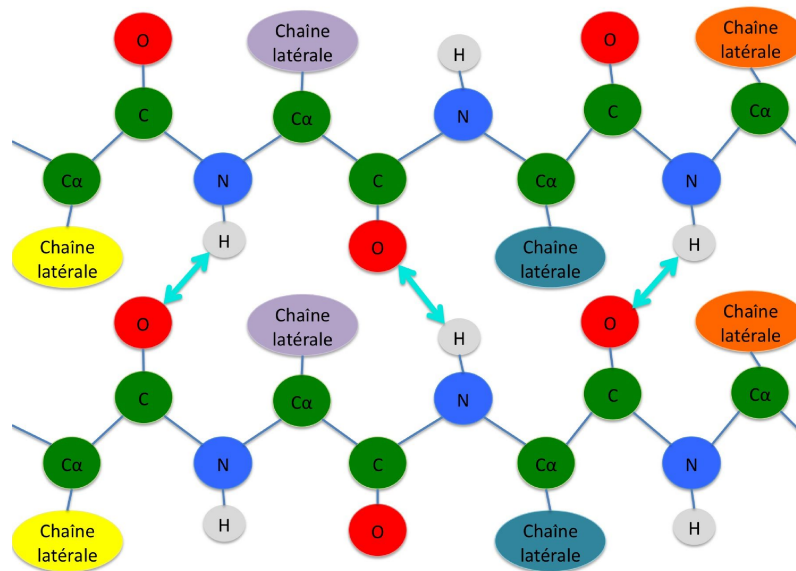


Figure 2.2 – Formation d’une structure secondaire protéique. Schéma de ponts hydrogènes établis entre deux sections parallèles de la chaîne. On voit qu’ici un acide aminé sur deux interagit avec l’autre chaîne, alors que les autres sont orientés à l’opposé. Le présent arrangement caractérise les structures secondaires de type feuillet  $\beta$  aux brins parallèles.



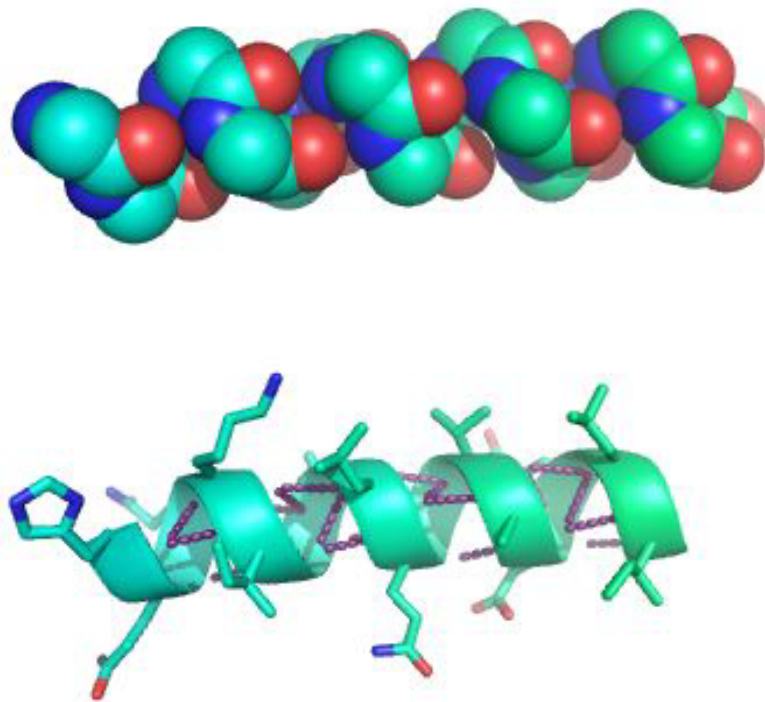


Figure 2.3 – L'hélice  $\alpha$ . En haut : Représentation sans les chaînes latérales. Les oxygènes (en rouge) de toutes les jonctions peptidiques sont impliqués dans des ponts hydrogènes avec l'azote (en bleu) du quatrième suivant, ce qui assure une compaction élevée. Les hydrogènes liés aux azotes ne sont pas montrés ici, mais on peut observer le fort rapprochement des oxygènes et des azotes. En bas : Représentation simplifiée de la même hélice  $\alpha$  à laquelle on a ajouté les chaînes latérales. Ces dernières sont impliquées dans les formations tertiaires. Les traits pourpres à l'intérieur de l'hélice représentent les ponts hydrogènes.

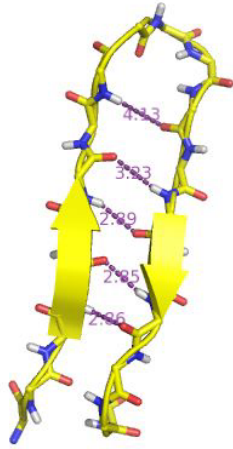


Figure 2.4 – Formation  $\beta$  anti-parallèle en épingle. Représentation simplifiée sans les chaînes latérales. Images générées avec pymol.



Figure 2.5 – Formation  $\beta$  parallèle en hélice. Représentation simplifiée, sans les chaînes latérales. Dans les formations  $\beta$ , un acide aminé sur deux forme un pont hydrogène avec un correspondant dans la section de chaîne qui longe la sienne. Images générées avec pymol.

courtes et très variables. Parmi les vingt sortes de chaînes latérales, certaines sont entièrement ou presque entièrement composées de carbones et d'hydrogènes comme le sont les huiles. En environnement aqueux, elles constituent le principal moteur de repliement tertiaire. Leur contact avec l'eau n'étant pas énergiquement favorable, on les dit hydrophobes. En milieu aqueux, elles entraînent la protéine à se replier de manière à être positionnées à l'intérieur, hors du contact avec l'eau. Les chaînes latérales hydrophobes peuvent plus aisément se positionner à l'extérieur de la protéine lorsque celle-ci est insérée à l'intérieur d'une membrane cellulaire, qui constitue un milieu hydrophobe. En milieu aqueux, ce sont les chaînes latérales polaires ou chargées, dites hydrophiles, qui se trouvent à entourer la protéine, car elles interagissent favorablement avec l'eau. Il ne sera pas nécessaire pour lire cette thèse de connaître en détail les vingt acides aminés. Les quatre aspects des acides aminés à retenir pour la lecture sont :

1. Hydrophobie versus hydrophilie : Les chaînes latérales chargées et polaires sont attirées par l'eau. Le contact entre l'eau et les chaînes latérales majoritairement composées de carbone n'est pas énergétiquement favorable. Elles se retrouvent donc entre elles, hors du contact avec l'eau.
2. Ponts hydrogènes et structures secondaires. Les contacts entre les plans peptidiques sont basés sur l'attraction du proton (l'hydrogène) associé à l'azote d'une jonction par l'oxygène d'autre jonction. La coopérativité entre plusieurs ponts hydrogènes permet de générer des structures régulières et stables.
3. Présence d'une diversité de dimensions et de formes dans les chaînes latérales.
4. Constitution planaire de la jonction peptidique.

#### **2.1.4 L'encodage du repliement**

Les protéines codées par le ribosome représenteraient seulement une fraction des séquences qu'on peut générer aléatoirement [153, 201]. Seules les protéines présentant un ratio adéquat de résidus hydrophobes par rapport aux résidus hydrophiles auraient été retenues par l'évolution. La nécessité d'une certaine compaction [54] et le contact avec l'eau [16] sont les paramètres qui déterminent ce ratio. Le contact avec l'eau est aussi un déterminant du nombre d'acides aminés privilégié pour les domaines : alors qu'un trop petit nombre ne fournit pas suffisamment d'interactions internes pour leur stabilisation, un trop grand nombre augmente les coûts entropiques (liés à la réduction du désordre) [288].

Le repliement en forme de globule est avantageux pour l'énergie de surface [296]. Les protéines à répétition se démarquent pourtant du modèle globulaire en adoptant une forme allongée. L'induction précoce, rapide et stable des structures secondaires empêcherait une réorganisation globale [71].

On observe depuis longtemps que le comportement du repliement serait principalement binaire, dicté par la présence ou l'absence de résidus hydrophobes [194, 213, 303]. Les positions hydrophobes détermineraient le repliement, sans beaucoup d'égard

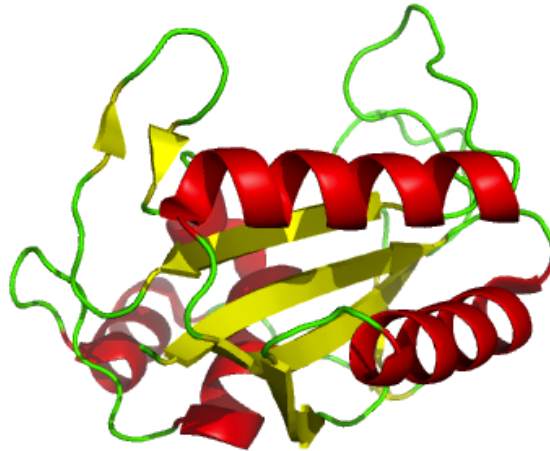


Figure 2.6 – Assemblage tertiaire. Hélices  $\alpha$  (en rouge) et feuillets  $\beta$  (en jaune). Enzyme HPPK, fichier 1Q0N.pdb. Image générée avec pymol.

à l'identité particulière des résidus tant hydrophobes que polaires. Les études par énumération exacte de modèles hydrophobe/polaire (HP) sur réseau font ressortir la pertinence d'un code de repliement quasi binaire basé sur les états hydrophobe et polaire [22, 33, 54]. Ceci s'explique par le fait que les résidus hydrophiles sont aussi "confortables" avec l'eau qu'entre eux. Ils entreraient donc peu en ligne de compte lors du repliement. Par exemple, les myoglobines, qui transportent l'oxygène, ont chez beaucoup d'espèces la même structure précise en deux couches d'hélices. Par contre, la comparaison des séquences indique que les résidus polaires se positionnent majoritairement à l'extérieur et sont hautement interchangeables. Les résidus hydrophobes sont situés à l'intérieur et ont souvent les mêmes positions dans la séquence [213].

Le repliement est à tel point relié à la disposition des résidus hydrophobes dans la séquence, que bien des méthodes de prédictions de structure sont basées sur celle-ci. Certains rythmes d'alternance des résidus hydrophobes conduisent à des structures particulières, comme c'est le cas de la répétition aux positions +3, +4, +3, +4... des résidus hydrophobes le long de la séquence, qui privilégie la formation d'hélices  $\alpha$  coiled-coil [118, 160]. La disposition des résidus hydrophobes dans la séquence permet donc de prédire certains types de structures. Par exemple on utilise une méthode d'alignement

des sections d'une même chaîne protéique pour y détecter la présence d'hélices  $\beta$ . Dans ce cas une suite continue de ponts hydrogènes parallèles dispose la chaîne principale comme une hélice autour d'un cylindre [46] (voir Fig. 2.5).

Selon les expériences de dénaturation (voir section 2.2.1), les protéines globulaires solubles formeraient d'abord un globule fondu [72, 97, 136, 232]. Les contacts hydrophobes formerait alors un réseau, et leur rapprochement favoriserait la formation subséquente des structures secondaires [13, 127]. Cependant, les expériences de dénaturation ne reflètent pas nécessairement les conditions réelles de repliement de sortie ribosome, qui privilégieraient un repliement graduel à partir du N-terminal [69]. Quoiqu'il en soit, les caractéristiques spatiales très diverses des chaînes latérales ont aussi un rôle à jouer, et non pas seulement leur nature hydrophobe ou hydrophile. Les conformations spatiales plus ou moins volumineuse constituent des contraintes négatives empêchant certains chemins. L'entassement spatial n'est pas le moteur du repliement, il est le résultat de la sélection naturelle. Il permet la perpétuation de la protéine en assurant sa stabilité et son bon fonctionnement. Les études de la mécanique de repliement de protéines de structure très semblables, mais dont les séquences diffèrent passablement, confirment cette hypothèse [12]. Le positionnement et le ratio optimal des résidus hydrophobe/hydrophile résulterait aussi de la sélection naturelle des séquences qui favorisent un repliement stable. Les interactions atomiques précises influenceraient peu le repliement.

La recherche d'arrangements spatiaux optimaux est un thème important pour le développement de cette thèse.

### **2.1.5 Le niveau quaternaire**

Malgré son repliement tertiaire, une protéine peut présenter certaines zones hydrophobes à sa surface. Ceci favorise son assemblage avec d'autres protéines identiques, pour former des homo-dimères, homo-trimères ou autres homo-polymères. Si les protéines sont un peu différentes nous obtenons des hétéro-polymères. Les canaux des membranes cellulaires sont généralement un assemblage homo-polymère ou hétéro-polymère [141, 186, 267]. Les complexes moléculaires, ces appareils spécialisés, se forment à partir de protéines très différentes qui se combinent même avec d'autres types

de molécules tels les acides nucléiques. Le ribosome [218], cette usine à protéines, est un assemblage de plusieurs protéines et molécules d'acide ribonucléique.

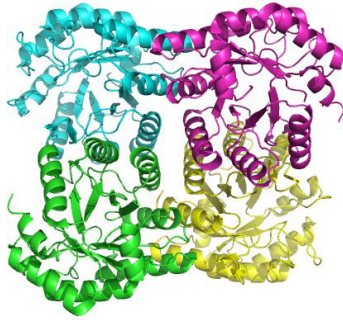


Figure 2.7 – Assemblage quaternaire de quatre instances d'une protéine. Images générées avec pymol.

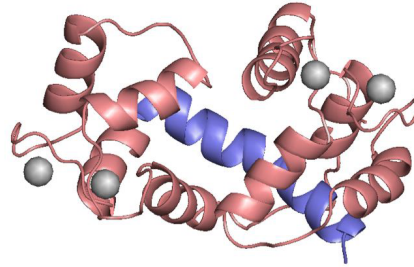


Figure 2.8 – La troponine C englobant un domaine de la troponin I en présence d'ions calcium. Images générées avec pymol.

## 2.2 La dynamique des protéines

Chaque protéine est donc agencée selon une structure précise. Dans cette structure, les éléments de structures secondaires, plutôt stables, sont reliés entre eux par des boucles plus flexibles. Ceci permet une certaine mobilité entre les éléments. Les protéines effectuent des mouvements lorsqu'elles accomplissent leur fonction et adoptent donc plus d'une conformation. Ces aspects dynamiques des protéines ont été étudiés par les physiciens, chimistes, biochimistes, mathématiciens. Les chercheurs s'intéressent à différents aspects de la dynamique des protéines, dans ces aspects généraux, puis pour des détails de plus en plus précis :

1. Flexibilité des différentes zones : localisation, rythme et amplitude des mouvements.
2. Interactions entre les éléments : blocage ou coopérativité.
3. Trajectoire d'un mécanisme, étapes et ordonnancement des événements.

### 2.2.1 Études in vitro

Commençons par la connaissance des zones flexibles et mobiles par rapport aux zones stables. L'analyse de la diffraction des rayons X qui ont traversé les cristaux d'une protéine fournissent de nombreux modèles tri-dimensionnels de celle-ci dans son état stable [28]. La position moyenne de chaque atome est obtenue grâce l'étude de la densité électronique. Mais on obtient aussi une indication de leur niveau d'agitation grâce aux facteurs de *Debye-Waller* (B-factors) [42]. Ceux-ci sont un indice de l'instabilité des différentes zones, ou tendance au mouvement. Ces indications concernent les fluctuations de faible barrière énergétique, mais permettent de cibler les zones plus flexibles de la molécule.

La résonance électromagnétique *RMN* est aussi beaucoup utilisée pour déterminer les structures moléculaires [178]. Elle peut être un peu moins précise que la cristallographie, mais elle est utilisée en solutions aqueuses, condition plus proche des conditions réelles. Elle vise les molécules qui se prêtent moins bien à la cristallographie ou bien apporte des confirmations ou informations complémentaires [24, 120]. La *RMN* utilise des champs électromagnétiques pour déterminer la position relative des atomes grâce à l'examen des influences mutuelles sur les spins nucléaires. Différents techniques axées sur la *RMN* permettent aussi de déterminer les zones en mouvements. Par exemple, les vibrations rapides des liens N-H dans la chaîne principale indiquent les zones où celle-ci a une plus grande liberté [111].

Les informations obtenues sur la flexibilité et le niveau de mouvement n'indiquent généralement pas leur amplitude ou leur trajectoires précises. Cette connaissance est plus difficile à obtenir. Dans certains cas, on peut faire appel à des fluorophores greffés par mutagenèse sur des domaines moléculaires a priori éloignés. Cela permet de détecter si ces domaines entrent en contact durant l'accomplissement d'une fonction [91, 110, 242, 289]. Le *GFP* est un fluorophore très connu basé sur les propriétés de fluorescence du plus volumineux des acides aminés, le tryptophane. Dans le *GFP* le tryptophane est bien protégé, positionné à l'intérieur d'un baril de brins  $\beta$  anti-parallèles de plus de 200 acides aminés. On a étudié la façon dont sa fluorescence est activée par certaines

chaînes latérales chargées ou polaires [37, 274]. Plutôt que de greffer le *GFP* au complet, il est possible de muter un résidu tryptophane à une position stratégique qui générera éventuellement de la fluorescence grâce à un rapprochement spécifique [8, 66, 107]. Ainsi donc avec les techniques de fluorescence, on arrive à connaître le mouvement interne plus complet de certaines protéines.

D'autres techniques sont dédiées à l'étude des forces et énergies impliquées dans les processus internes ou externes de la protéine. Les études par tractions (*single-molecule force spectroscopy* [221, 222]) permettent d'évaluer la force d'attachement de certains domaines moléculaires. De telles études ont été réalisées avec la calmoduline [121, 122] en relation avec ses différentes cibles moléculaires. Dans ces études on réussit à attacher les deux extrémités protéiques, l'une à un matériel fixe, l'autre à un nano-instrument grâce auquel on applique une force de traction.

Les études de dénaturations [86, 187, 224, 229] permettent d'étudier les états avoisinant le repliement. On obtient grâce à elles une courbe du taux de dépliement en fonction de la concentration de dénaturant ou de la température [187]. Plus la pente de dénaturation (ou valeur « m ») est abrupte, plus la coopération entre les résidus est élevée pendant l'étape de transition. Le délai qui précède la courbure est quant à lui un indice du niveau de stabilité de l'état replié. La mutagenèse d'un seul résidu à la fois est utilisée pour étudier la contribution des contacts tertiaires à la stabilité de la forme native. Le calcul de la valeur  $\phi$  indique si la variation de l'énergie de transition provoquée par la mutation est proportionnelle (valeur proche de 1) ou non (valeur 0) à la variation de l'énergie de l'état final :

$$\phi = \frac{\Delta(\Delta G^{t-d})}{\Delta(\Delta G^{r-d})} \quad \text{où t-d = transition-dénaturé, r-d = replié-dénaturé} \quad (2.1)$$

Ceci donne un indice de la participation du résidu muté à la phase de transition. On peut obtenir la répartition de ces valeurs  $\phi$  tout le long de la protéine. Les valeurs  $\phi$  proches de 1 indiquent que l'état de transition est perturbé par la mutation dans la même mesure que l'état final. Cela suggère qu'une certaine organisation de cette région est déjà présente dans l'état de transition. La déstabilisation par des mutants individuels



peut ensuite être comparée à la déstabilisation par deux mutations à la fois. Ceci permet de mieux comprendre les processus de coopérativité. On se sert de cette méthode pour discerner les résidus qui agissent ensemble lors de l'étape de transition principale. Des dénaturants (urée, guanine) sont aussi utilisés pour connaître les changements dans le rythme de dépliement selon les mutations, par rapport à celui de la protéine native. On peut déterminer ainsi les zones (valeur- $m$  élevée) dont les autres zones ont besoin pour se replier.

La dynamique de fonctionnement d'une protéine suscite certaines questions particulières :

1. Les transitions s'effectuent-elles en une seule phase ou existe-t-il des états intermédiaires stables ?
2. Comment les parties de la protéines interagissent-elles entre elles ?
3. Les mouvements de la protéine mature impliquent-ils des transformations de la structure secondaire ou seulement de la structure tertiaire ?

### 2.2.2 Paysage énergétique

La variation de l'énergie potentielle lors des transitions peut être représentée par un paysage énergétique. Celui-ci est une image de l'aisance avec laquelle la protéine effectue son parcours de transformation, ou au contraire, illustre les blocages qui entravent son chemin. Plus la coopérativité est élevée entre ses éléments, plus la protéine est susceptible d'effectuer son parcours en une seule phase. En vue d'appuyer le développement de cette thèse, nous allons maintenant faire ressortir ici que la coopérativité peut être étroitement liée à la présence de multiples interactions concordantes agissant selon une même direction de l'espace.

Les études par dénaturation ont permis d'étudier la balance des forces dans le cas des protéines à motifs à répétition, qui représentent une simplification du cas général et facilitent la compréhension de l'oscillation entre les états d'équilibre. Parmi les cas particulièrement étudiés il y a l'ankyrine [174, 180], *TPR* [71, 166, 174], *CagY* du *Helicobacter pylori* [50] et la spectrine [222]. Les motifs répétés se présentent sous forme de

paires d'hélices  $\alpha$  séparées par des boucles. Dans bien des cas, la structure tertiaire peut être considérée comme quasi unidimensionnelle, avec des contacts hydrophobes solides établis entre chaque structure répétée et une rareté ou absence de contact avec les zones éloignées dans la séquence.

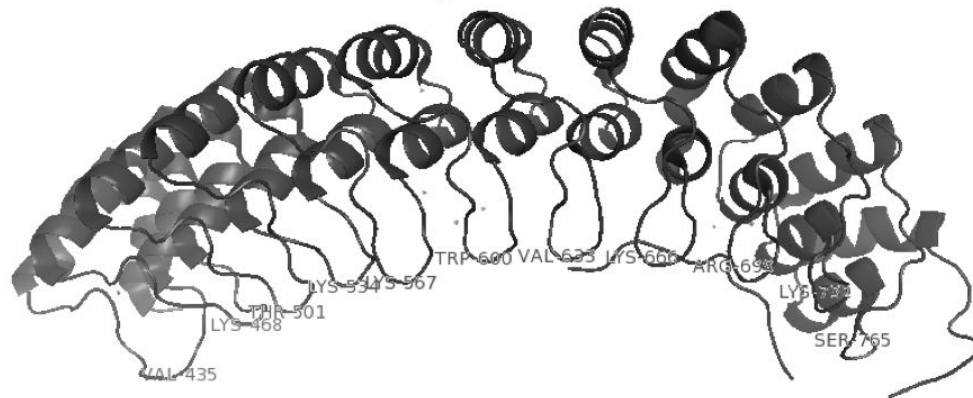
Des études de dénaturation et de mutagenèse ont démontré une forte coopérativité entre les motifs lors de la transition entre l'état déplié et l'état replié. À cause de la forte coopérativité entre les unités de repliement (foldons), le repliement des zones à répétition suit un trajet simple de la forme d'un entonnoir descendant jusqu'au minimum énergétique de l'état replié. Ces propriétés rendent les protéines à répétition aptes à être étudiées selon un modèle mathématique de Ising [146, 305]. Ferreiro et al. [71] ont ainsi étudié les propriétés énergétiques des structures répétées des protéines TPR, en représentant chaque foldon (de 34 d'acides aminés chacun) par un état binaire « Replié » ou « Non replié ». Selon eux de légères modifications ont plus d'effets pour perturber le processus de formation chez les protéines à répétitions que chez les protéines comportant davantage d'interactions tertiaires.

Selon le modèle, chaque foldon  $j$  tend à se déplier à cause du gain d'entropie  $s_j$  (multiplié par la température), mais que ce gain est contrecarré par les gains d'énergie libre des foldons repliés : grâce à leur énergie interne  $\varepsilon_j^i$ , mais aussi grâce à la diminution de l'énergie de surface  $\varepsilon_{j,j+1}^s$  entre chaque paire de foldons accolés [71].

$$H = \sum_{j=1}^N [-Ts_j\delta_{jU} - \varepsilon_j^i\delta_{jF}] - \sum_{i=1}^{N-1} \varepsilon_{j,j+1}^s\delta_{jF}\delta_{(j+1)F} \quad (2.2)$$

où U=unfolded F=folded et  $\delta_j$  vaut 1 si l'état est présent, 0 sinon.

Selon la seconde partie de l'équation, lorsqu'un élément  $j$  se retrouve dans un état replié ( $\delta_{jF}=1$ ) un fort effet de coopérativité entraînera le repliement des éléments  $j-1$  et  $j+1$ . Le système est donc binaire entre deux états, bien que les unités binaires soient des motifs d'environ 30 ou 40 résidus. Le facteur entropique  $s_j$  est fonction de la concentration de dénaturant et de la sensibilité au dénaturant. La protéine se dépliera si la concentration de dénaturant ou la température augmente. L'augmentation de l'énergie de surface entre les foldons a pour effet de rehausser la stabilité (une plus grande quantité de déna-



Ankyrin 1N11.pdb

Figure 2.9 – Les motifs répétés de l’ankyrine. L’architecture de l’ankyrine se présente sous la forme de la suite répétée d’un même motif hélice-boucle-hélice. Fichier 1N11.pdb, avec mes annotations. Image générée avec pymol.

turant est requise pour amorcer le dépliement), mais aussi la coopérativité (la rapidité du dépliement). En conditions normales la forte coopérativité entre les éléments induit un repliement dont le comportement au niveau énergétique prend la forme d’un entonnoir (voir Fig. 2.10(b)).

L’empilement linéaire des spires dans les molécules à motifs répétés n’est pas sans rappeler l’empilement linéaire des spires dans une hélice  $\alpha$  bien que ces architectures se situent à des niveaux d’organisation moléculaire différents. Une transition significative survient lorsque plusieurs éléments susceptibles de collaborer ensemble se retrouvent rapprochés de façon à pouvoir interagir dans une même direction de l’espace. Nous verrons que la technique développée dans cette thèse permet la génération de basculements d’ensembles dans l’espace, et permet ainsi de rechercher ces rapprochements favorables. Le parcours énergétique des protéines globulaires est plus complexe (voir Fig 2.10(b)), présentant des rugosités, des pauses dans le parcours. Ces blocages peuvent parfois correspondre à des obstacles stériques qui empêchent certains rapprochements de se faire.

Depuis longtemps on cherche à savoir si les protéines en général, ou telle protéine

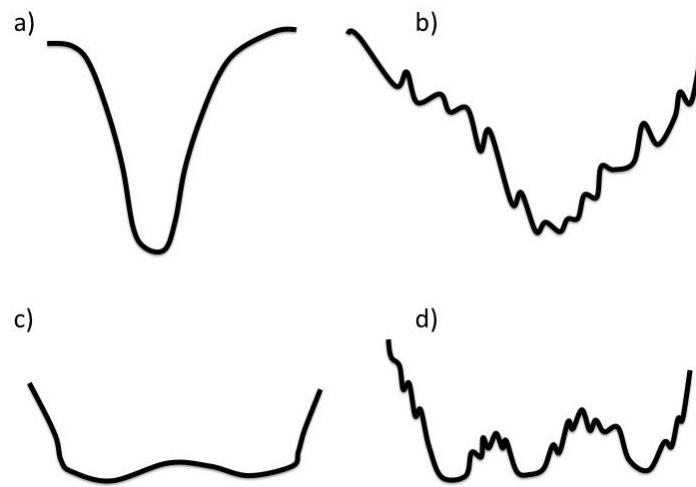


Figure 2.10 – Paysages énergétiques. (a) Paysage énergétique du repliement d'une protéine à répétition : entonnoir. Deux états seulement : déplié et natif. (b) Paysage énergétique du repliement d'une protéine globulaire : états intermédiaires, plusieurs phases. (c) Paysage énergétique d'une protéine native à deux états simples. (d) Paysage énergétique d'une d'une protéine native plus complexe : états intermédiaires entre des états plus stables.

en particulier, atteignent leur état natif en une seule phase de transition ou en plusieurs étapes. Les expériences de dénaturation/renaturation des protéines globulaires montrent que les petites protéines se replient en une seule étape, alors qu'un processus en plusieurs étapes est nécessaire pour les plus grosses [12]. Plusieurs chemins appartenant à un ensemble d'états de transition [55] sont empruntés au cours d'étapes de réorganisation pour retrouver l'état natif [203]. Les éléments de structures n'apparaîtraient pas de façon corrélée et des pauses seraient présentes dans les trajets énergétiques avant que le tout ait convergé vers l'état natif. Onuchic et al. suggèrent via des théories statistiques que les frustrations provoquant ces pauses sont minimales et que nous avons plutôt affaire à un entonnoir rugueux [201, 202].

Déplaçons maintenant cette réflexion vers les cas où la protéine a atteint son état natif et ne se déplace plus qu'entre des états énergétiques avoisinants, ou dont la différence énergétique relative entre les états peut être modifiée par les conditions environnementales ou des clefs appelées ligands [121]. Dans le cas d'une molécule à deux états, les paramètres environnementaux peuvent faire en sorte que les 2 états présentent un niveau énergétique assez semblable, avec une faible barrière énergétique entre l'état fermé (replié) et l'état ouvert (déplié). Nous n'avons donc plus un terrain énergétique en forme d'entonnoir, mais bien une colline douce entre deux bassins (Fig. 2.10(c)). Pour les molécules plus complexes, un terrain plus rugueux, voire plus élevé, séparent deux bassins ou plus (Fig. 2.10(d)). Les conditions environnementales et les ligands peuvent changer la répartition des bassins et des sous-bassins (états intermédiaires).

### 2.2.3 Jeu entre les niveaux

La protéine est organisée en plusieurs niveaux et on pourrait supposer que l'état natif atteint, seul son dernier niveau, ou structure tertiaire, oscille d'un état stable à l'autre. Mais d'abord, le niveau tertiaire s'est-il vraiment formé avant le niveau secondaire, tel que leur nom le laisserait supposer ?

Les premiers modèles préconisaient qu'une hélice  $\alpha$  pourrait se former spontanément sans subir l'influence des autres parties de la chaîne. Un paramètre de simulation favorisant le démarrage d'une première boucle d'hélice était nécessaire, les boucles sui-

vantes pouvant être induites selon un modèle de Ising [305]. Dans la foulée de cette idée, on a examiné la possibilité que les structures secondaires soient formées avant la structure tertiaire, et se rassemblent pour conduire à la structure tertiaire [74]. Cette idée est appuyée par les observations qui indiquent que la tendance au repliement est une tendance locale [14]. Cependant, selon les études de protéines à répétitions, la formation d'hélice unique serait impossible [71]. Les nouveaux modèles préconisent que la formation des structures secondaires est facilitée par des interactions tertiaires inductrices [217]. L'élongation de ces structures régulières obéit aussi aux lois de la nucléation [27, 43, 179, 190], selon lesquelles la formation d'un premier noyau diminue ensuite l'énergie nécessaire pour l'agglomération des éléments suivants. De plus, certaines protéines ne forment leurs structures secondaires qu'au moment de leur assemblage en dimères [151].

La prédiction des hélices  $\alpha$  est principalement basée sur la recherche de résidus hydrophobes situés à intervalles réguliers de façon à se retrouver sur un même côté de l'hélice  $\alpha$  [118, 160]. La force hydrophobe serait suscitée par le contact tertiaire entre les régions de la chaîne et favoriserait alors la formation d'hélices selon un processus d'optimisation de l'intercalation des résidus hydrophobes. Ces interactions tertiaires inductrices s'effectuent chez les protéines à répétitions entre des sections de chaîne très voisines, alors que des régions éloignées dans la chaîne interagissent ensemble chez les protéines plus complexes.

Lorsque les protéines atteignent leur forme native, leurs structures secondaires ont une grande stabilité. Le jeu de leurs mouvements se situerait alors davantage au niveau de la structure tertiaire. Comme nous le verrons au chapitre 8, l'ouverture et la fermeture des domaines en main-EF de la Troponin C est un exemple de mécanisme dans lequel les structures secondaires (quatre hélices  $\alpha$ ) sont préservées [155].

D'autres mécanismes sont au contraire basés sur l'apparition et la disparition de structures secondaires. On observe par exemple la disparition et réapparition d'un secteur d'hélice lors du large mouvement de boucle de la tyrosine kinase entre son état passif et son état actif [293]. La Troponine C et la Calmoduline ont une forme très similaire : deux domaines (N-terminal et C-terminal) reliés par un long lien flexible qui peut

prendre la forme d'une longue hélice  $\alpha$  [11, 77, 137]. Cette longue hélice se plie en deux lorsque les deux lobes interagissent avec des ligands, nous donnant un autre exemple d'instabilité de structures secondaires. La présence de glycines, qui n'ont pas de chaîne latérale, favorise une fluctuation dans les structures secondaires. On retrouve un autre exemple d'instabilité des structures secondaires au niveau d'une structure en feuillet  $\beta$  qui réunit les deux motifs mains-EF d'un domaine. Cette structure est déstabilisée en présence de l'ion calcium [75, 161, 238, 294].

Des états transitoires d'une structure secondaire à l'autre existent aussi. Une transition d'hélice  $\alpha$  vers feuillet  $\beta$  survient chez les protéines qui forment des fibres de type amyloïde, comme dans le cas de la protéine prion. Ollesch et al. ont utilisé la technique *time-resolved Fourier transform infrared spectroscopy* pour étudier la transition de  $\alpha$  vers  $\beta$  induite par la réduction d'un lien disulfide et menant à la forme nocive de la protéine [198]. La configuration anormale est davantage structurée que la conformation régulière. Elle agirait comme un moule qui abaisse la barrière énergétique entre les deux conformations et favorise l'agglutination [43].

#### 2.2.4 Niveaux d'organisation

Les protéines comportent en elles-mêmes plusieurs niveaux d'organisation et font elles-même partie d'organisations dites quaternaires, mais cela ne s'arrête pas là. Dans une cellule, la complexité de tissus protéiques de grande envergure s'établit grâce à plusieurs autres niveaux, dont les agencements peuvent être très semblables pour plusieurs tissus [1]. Selon Ackbarow et Buehler, chaque niveau fonctionne selon une échelle de temps différente, permettant de séparer la réactivité par niveau et assurant la robustesse des tissus qui structurent la cellule ou forment les cartilages, les os ou les cheveux. La longue hélice  $\alpha$  de la vimentine s'enroulera autour d'une autre permettant de former la base de filaments intermédiaires, qui eux forment la base de fibres, puis de tissus plus complexes. Le tout premier niveau de tels tissus, la séquence de la protéine, contient des ingrédients dit "silencieux" qui assurent l'universalité et la souplesse des niveaux plus élevés mais sont sans effet sur l'organisation du niveau le plus global. Plus précisément, l'élasticité des niveaux intermédiaires est assurée par l'élasticité et l'uniformité des hé-

lices  $\alpha$  du niveau le plus bas. Mais la séquence primaire contient aussi des éléments plus variables, dont le rôle n'intervient qu'après la construction de plusieurs niveaux d'organisation, pour créer des jonctions spécifiques aux divers types de réseaux du niveau cellulaire. Ils assurent par exemple la différence marquée entre un cheveu et le filet de laminine enveloppant le noyau cellulaire, bien qu'ils soient construits à partir des mêmes types de niveaux intermédiaires [1].

Le comportement d'un niveau supérieur ne peut donc être simplement défini par le comportement moyen de ses sous-éléments. Car certains acides aminés présents au niveau le plus primaire ne deviennent actifs qu'après l'établissement de plusieurs couches. C'est aussi le cas des centres réactifs, dont le positionnement s'établit avec une grande précision consécutivement au repliement de la protéine. Créé à partir d'éléments précis (et conservés entre les espèces [299]) de la séquence primaire, le centre réactif ne participe pas à l'hierarchisation de la protéine, mais interagira avec des processus de niveau cellulaire. En ce qui concerne les tissus complexes, l'émergence de la diversité structurale se produit à partir d'éléments marginaux qui étaient d'abord inutiles dans le cadre du fonctionnement des couches inférieures et intermédiaires, et dont le rôle n'apparaît qu'en bout de ligne.

### **2.2.5 Limitations des méthodes expérimentales**

Les méthodes d'investigation *in vitro* sont limitées entre autre par la capacité de résolution des instruments. À l'échelle atomique, on peut difficilement obtenir directement l'image d'un processus en mouvement. On peut obtenir des indices, mais ceux-ci peuvent aussi être faussés par le processus expérimental. Les indices de mobilité (*B-factors*) de la cristallographie deviennent trop proches de la marge d'erreur [219] pour les plus protéines de plus grande envergure. Les facteurs de *Debye-Waller* en surfaces exposées sont moins fiables car ils peuvent être faussés par des contacts non naturels entre les protéines du cristal [64, 65]. La spectroscopie par résonance électromagnétique (*RMN*) intervient dans des conditions aqueuses plus proches de la réalité biologique mais la taille des protéines étudiées est plus limitée qu'en cristallographie. Néanmoins certains progrès techniques permettent d'identifier des rythmes lents dans de grosses



molécules [111, 283].

La détermination de la flexibilité permet de connaître les possibilités de mouvement. Les trajets effectifs sont plus difficiles à déterminer expérimentalement, car les techniques d'investigation sont invasives et affectent les processus. Les fluorophores sont greffés par mutagenèse à des résidus de nature précise. Ils sont volumineux et il faut choisir l'endroit où les greffer pour ne pas trop affecter le mouvement interne. Ils sont cependant excellents pour étudier les migrations au sein ou en dehors des cellules. Utilisé seul, le tryptophane est moins invasif pour détecter les interactions internes, mais il demeure tout de même le plus volumineux des acides aminés.

Les méthodes de dénaturation cherchent à évaluer les énergies de transitions. Elles se basent sur la mesure de signaux indiquant le dépliement ou repliement, tels que la fluorescence, par rapport à la quantité de dénaturant (température ou produit). Ils cherchent ainsi à déterminer le paysage énergétique entre les états déplié et replié. Mais comme Smart l'explique, cela suppose que l'agent dénaturant agit uniquement sur l'énergétique de dépliement. Mais à cause de la violence des agents dénaturants, la courbe du dépliement cacherait une ligne de base représentant la transmission d'une partie de l'énergie à des processus autres que le dépliement (par exemple vibration des chaînes latérales, effet sur le fluorophore) [240]. Il y a donc des efforts de développement pour des techniques de transfert d'énergie moins invasives qui permettent de distinguer les deux processus [53, 99, 240].

## CHAPITRE 3

### MÉTHODES DE SIMULATION

Nous voyons ici comment l'informatique permet d'approfondir certaines questions apportées par les expérimentateurs. Nous traçons un portrait des moyens informatiques développés pour étudier la flexibilité, les mouvements et les mécanismes des protéines, dans cette grande quête pour comprendre leur fonction. Cela permettra de situer les innovations de cette thèse par rapport à l'ensemble des initiatives informatiques.

De plus en plus de modèles informatiques tridimensionnels sont obtenus grâce à la cristallographie et à la *RMN*. Les théories et les expérimentations ont fourni les connaissances en énergétique qui permettent de modéliser informatiquement les forces et les énergies qui gouvernent les interactions atomiques. Le but des méthodes informatiques est d'étudier les zones laissées dans l'ombre par les méthodes expérimentales, ou de prolonger et approfondir ce qui a été découvert par les expérimentateurs et théoriciens.

Les simulations informatiques visent à déterminer les mouvements à l'intérieur d'une protéine à partir des informations données par les fonctions d'un potentiel ou champs de force. Le paysage énergétique informatique est déterminé par le potentiel. Celui-ci est utilisé par la méthode d'échantillonnage (la plus ancienne étant la dynamique moléculaire [150, 173]), pour faciliter le choix d'une direction dans l'espace des conformations et découvrir des formes voisines accessibles. En améliorant les méthodes d'échantillonnage on cherche à inventorier efficacement et exhaustivement les conformations possibles, et idéalement la probabilité de chacune.

Des calculs très complets des interactions atomiques requièrent l'utilisation de potentiels de mécanique quantique, qui évaluent la densité électronique autour des noyaux atomiques. Ceci n'est possible que pour l'étude de phénomènes atomiques dont l'échelle de temps est très courte, et pour des molécules ou zones moléculaires de dimension restreinte, pour des études très spécifiques [147, 211]. Pour accéder à des temps moléculaires plus élevés ou à des structures de plus grande dimension, on doit faire des compromis et utiliser différentes stratégies d'approximation pour les champs de force.

On peut aussi élaborer des techniques d'échantillonnage accélérées en plus ou au lieu de chercher à réduire les calculs du potentiel. Les prochaines sections présentent les stratégies d'accélération en trois volets : stratégies basées sur le potentiel, stratégies basées sur l'échantillonnage, combinaison de stratégies. Nous verrons que certaines méthodes se combinent avantageusement.

### 3.1 Stratégies d'accélération basées sur le potentiel

Le paysage énergétique parcouru est défini par le champs de force utilisé. La simplification du champs de force permet de diminuer la rugosité de cet espace et de faciliter son exploration. Les stratégies consistent d'abord à réduire la quantité de calcul, puis à éliminer des atomes, ou à les regrouper en ensembles de plus en plus grands. Voici selon cet ordre un aperçu de stratégies courantes :

#### 1. Potentiels quantiques

Les potentiels qui évaluent les forces atomiques au niveau quantique se situent à l'extrémité la plus coûteuse en temps CPU. On les utilise pour étudier l'interaction de résidus avec un ion par exemple [147, 302].

#### 2. Mécanique moléculaire

Ensuite vient la mécanique moléculaire qui ne tient plus compte des électrons. Elle évalue les liens covalents et leur angles comme des ressorts simples, la position optimale des angles diédraux (entre quatre atomes consécutifs) et les interactions non liées, hydrophobiques et ioniques. Tous les atomes sont encore ici représentés.

#### 3. Élimination d'atomes

Toute une catégorie d'atomes peut faire partie d'un ensemble intangible, si ceux-ci ne sont pas représentés mais que leur effet est inclus dans les relations entre les atomes visés par l'étude. Les potentiels à solvant implicite ont pour but d'éviter les calculs liés aux très nombreuses molécules de solvant (souvent l'eau) avec lesquelles la protéine réagit. Le solvant aqueux a pour effet de diminuer l'impact

des liaisons entre les chaînes polaires et chargées, car elles sont aussi confortables dans l'eau que liées entre elles. Les interactions entre elles n'interviennent qu'en milieu fermé. Les éléments courts se retrouveront les premiers en milieu fermé et c'est ainsi que les ponts hydrogènes de la chaîne principale induiront mutuellement leur enclenchement pour former les structures secondaires. Les potentiels à solvant implicite évaluent ces situations selon deux stratégies : étude statistique des contacts, comme on le voit par exemple pour le potentiel *OPEP* [52, 171], ou évaluation de l'absence de l'eau, selon un calcul de la *SASA* (*solvent accessible surface area*). Ces dernières intègrent aussi des paramètres statistiques liés aux types d'atomes. Des approximations de la *SASA*, calculables en des temps CPU accessibles, et permettant la dérivation de l'énergie (pour le calcul des forces) ont été développées. Des méthodes telles *MSEED* [212] étudient les intersections entre des sphères autour des atomes pour trouver la surface accessible [79, 244]. Des approches telles *EEFI* [145] et *POPS* [32] utilisent une fonction incluant la distance entre les atomes, leur dimension et la dimension de la molécule de solvant pour évaluer la probabilité d'inclusion d'une molécule de solvant [101, 280]. Des atomes peuvent non seulement être éliminés au niveau du solvant, mais devenir implicites dans la protéine. Par exemple les atomes d'hydrogène reliés aux carbones sont souvent omis et inclus dans les carbones correspondants [192]. Ces carbones spéciaux annoncent déjà l'idée du regroupement d'atomes car ils ont une dimension et un comportement différents des atomes de carbone simples.

#### 4. Regroupement d'atomes

Le champs de force utilise typiquement le temps CPU selon le carré du nombre d'éléments. Pour réduire le nombre d'interactions à calculer, on rassemble des éléments en un seul ensemble simplifié. Une bille permet de représenter plusieurs atomes, augmentant la durée moléculaire couverte par les simulations informatiques. Avec les modèle de  $G\ddot{o}$ , une chaîne protéique peut être représentée par ses atomes de carbone  $\alpha$  seulement [71, 105, 148, 265]. Par exemple, les modèles de  $G\ddot{o}$  utilisés par Ferrerio et Walczak pour étudier les protéines à répétitions ex-

pliquent les causes de baisse de la coopérativité constatée lors des expériences impliquant la mutation d'un simple résidu impliqué dans l'énergie de contact entre les foldons [71] : l'énergie de surface ainsi altérée dans une zone provoque une scission qui interrompt le processus coopératif de tout le système. Ferreiro et Walczak indiquent qu'un minimum de trois foldons consécutifs doivent être intacts pour qu'on puisse obtenir un effet coopératif. L'effet d'entonnoir énergétique proviendrait selon eux d'un mécanisme de nucléation où les premières structures formées recruteraient les boucles avoisinantes pour les intégrer dans un patron régulier.

On peut aussi avoir une certaine variété de billes au lieu de seulement un seul type [128, 157]. D'autres types de billes sont utilisées dans les simulations d'acides nucléiques [257] ou de phospholipides membranaires [234].

La définition d'une bille implique une définition des champs de forces qui modulent son comportement. On classe les modèles gros grains selon l'approche utilisée pour construire champ de force [48, 264] : 1) Constitutifs (biaisés selon un modèle de protéine particulier) [148], 2) Basés sur des connaissances (empiriques, statistiques) [51], 3) Corrélés sur les forces atomistiques moyennes [40, 112, 247, 258] en vue d'une modélisation élastique.

Notons aussi que certaines approches cherchent à réduire l'espace conformationnel d'une protéine en considérant principalement les angles diédraux, seuls susceptibles d'engendrer des changements de conformation significatifs [85, 90, 126].

## 5. Réseau rigide

Alors que les approches par réseaux élastiques (Voir section 3.2) se basent sur l'étude des fréquences moléculaires, les approches par réseaux rigides se basent sur la densité des interactions atomiques. Celles-ci conviennent aux molécules hétérogènes, où la densité des interactions fortes entre les atomes peut varier beaucoup selon les régions de la molécule. La méthode *FIRST* utilise un algorithme appelé *Pebble Game* [113, 114]. Les atomes d'une protéine sont représentés par les sommets d'un graphe et les interactions fortes entre les atomes par des barres entre les sommets. On déplace des contraintes le long de ces barres, jusqu'à ce que

les zones rigides soient identifiées. Par la suite, lors de simulations de type Monte Carlo, chaque zone rigide bougera collectivement à l'aide de légers déplacements aléatoires (Méthode *FRODA*) [281], méthode qui s'applique aussi aux ARN [80].

## 6. Corde élastique

Si la molécule s'y prête, l'approximation peut englober tous les atomes en une représentation d'unique ressort. Une macro-molécule ou une structure moléculaire de type longiligne sera modélisée comme un unique ressort ou une suite de ressorts que l'on peut compresser et étirer [21, 184, 214]. Dans la section 3.3.1 nous voyons que les tissus biologiques sont souvent des assemblages de plusieurs niveaux de molécules longilignes élastiques. Ces études d'élasticité longilignes ont leur contrepartie bidimensionnelle dans l'étude de la multi-échelle des membranes lipidiques [109].

La diminution du nombre d'éléments à calculer permet de réduire le temps de calcul selon un facteur appromaxivement en  $N^2$ . Si la définition du comportement du solvant est implicitement contenue dans celui des atomes de la protéine, on évite par exemple le calcul lié aux molécules d'eau, souvent très nombreuses. Ces approximations lissent l'espace des conformations à visiter. Mais plus on simplifie l'espace des conformations plus il peut devenir distordu. Il risque de fournir des chemins moins fiables ou approximatifs et dépourvus de certains détails importants.

## 3.2 Stratégies d'accélération basées sur l'échantillonnage

Cette section explique comment simplifier le parcours effectué par les simulations. Avec les différents calcul de forces, le temps CPU tend à augmenter selon le carré du nombre d'éléments. Mais la recherche des conformations possibles risque pour sa part de faire croître le temps CPU exponentiellement. Ce problème devient bien plus considérable que celui du champ de force lorsqu'on augmente la taille de la molécule étudiée. Ce problème est réduit si l'on utilise des champs de force de plus en plus précis, qui évitent l'errance dans des trajets non réalistes. Ceci entre en contradiction avec la stra-

tégie de simplification des champs de force. Un autre problème provient du risque de longue perte de temps en errance à l'intérieur de puits d'énergie locaux. Les méthodes d'échantillonnage accéléré visent donc à rechercher les mouvements d'une certaine envergure, évitant les variations locales de conformation.

On pourrait classer ces méthodes suivant la façon dont elles suivent de près ou survole la courbure du terrain énergétique. À une extrémité nous avons la dynamique moléculaire qui suit de près les courbures du terrain. Ici on quitte un creux du terrain en espérant qu'une route s'y tracera au fur et à mesure et nous dirigera au bon endroit. À l'autre extrémité, nous avons les méthodes d'interpolation, qui utilisent la connaissance des conformations source et destination et tracent un parcours entre les deux. La destination s'atteint facilement, mais le parcours peut prendre des raccourcis irréalistes. Et entre les deux stratégies, il y a des méthodes qui cherchent à éviter les minimums locaux et à faire des déplacements d'une certaine envergure.

### 1. Dynamique moléculaire

Utilisée depuis plusieurs décennies[150, 173], la dynamique moléculaire demeure la méthode d'échantillonnage la plus connue et la plus courante [20, 36, 55, 60, 123, 138, 164, 165, 189, 193, 216, 236, 291]. Elle utilise le calcul des forces pour faire varier la vitesse et la direction de chaque atome selon la formule  $F = m a$ . Les changements de positions modifient les forces et les forces influencent les positions. L'ensemble atomique fera beaucoup de va et vient dans l'espace local, mais selon les lois des probabilités et grâce à l'énergie cinétique, il empruntera éventuellement une voie de transition vers une conformation plus distante.

### 2. Analyse des modes normaux

Le puits harmonique (terrain énergétique en forme de bol) dans lequel une molécule effectue un va-et-vient correspond à des valeurs propres du système. Or, les molécules présentent des fréquences plus lentes, qui correspondent à des fluctuations de plus grande envergure. Pour les simuler, on peut construire un réseau élastique ( $EN$ ) en reliant chaque atome à d'autres par des contraintes élastiques déterminées par examen des interrelations observées dans les dynamiques

moléculaires [260]. Mais de plus en plus on préfère extrapoler ces mouvements harmoniques de grande ampleur en utilisant l'*étude des modes normaux* (*Normal mode analysis* ou *NMA*) [30, 115, 149, 260]. Le *NMA* constitue une méthode d'extrapolation à partir des caractéristiques structurales d'une conformation stable de la molécule. Ses calculs se basent sur l'analyse matricielle des dérivées de l'énergie [88, 134]. Les grands mouvements d'allers-retours peuvent être modélisés par des forces élastiques correspondant à leur amplitude et à leur fréquence. Certains atomes plus fortement reliés se déplacent ensemble. La conservation d'une représentation moléculaire complète limite toutefois l'ampleur des molécules étudiables. La méthode *RTB* (*Rotation Translation Blocks*) combine l'approche gros grain à la *NMA* en définissant des blocs d'atomes [3, 252, 253] et permet l'étude de cibles plus imposantes. Les déplacements de basse fréquence dépendraient de l'organisation générale de la molécule et non pas des détails locaux [159]. Les approches gros-grain sont donc très utilisées en combinaison avec le *NMA* [65, 102, 104, 266, 292, 300].

### 3. Recherche de points de selle

Au lieu de demeurer dans un bassin harmonique quelle que soit son ampleur, on peut vouloir trouver un passage pour le quitter et trouver un bassin avoisinant. Bien que certaines protéines présentent des mécanismes où les rythmes vibratoires lents s'effectuent sans obstacles, il y a les protéines et autres molécules dont certains éléments sont bloqués dans l'une ou l'autre de leur position harmonique par d'autres éléments [119, 162]. Ces cas présentent des barrières à surmonter lors du passage d'un état à l'autre. Ces mécanismes peuvent être étudiés par les méthodes de points de selle [17, 31, 163, 167, 170]. La forme de la selle de cheval évoque la forme d'un col entre deux montagnes lisses. L'environnement d'un minimum local est exploré afin de découvrir de tels passages énergétiquement favorables vers un minimum local avoisinant. Ces méthodes suivent la courbure des forces dans l'environnement d'un minimum local. Elles ont d'abord été utilisées dans l'étude des verres amorphes [18, 181], mais aussi dans l'étude des agrégations



amyloïdes [139, 171, 298] et du repliement des protéines [246].

#### 4. Interpolations

L'interpolation d'un chemin nécessite que les conformations de départ et d'arrivée soient connues. Son but est d'éliminer les errances aux abords ou loin du chemin de transition. L'interpolation intègre au potentiel la connaissance des contacts qui devront être formés dans la conformation de destination. Cette intégration se fait selon un pourcentage graduel au fur et à mesure des itérations de la simulation [125]. L'interpolation est très utilisée avec les modèles de Gō pour comprendre la dynamique générale d'un mécanisme. L'étude des chemins d'ouverture et fermeture de protéines à mains-EF a bénéficié de cette méthode [265, 266]. L'interpolation peut aussi être utilisée avec des modèles atomistiques [15].

### 3.3 Hiérarchie et combinaisons de stratégies

La complexité des protéines, leur hétérogénéité structurale, rend difficile l'utilisation d'une approche accélérée unique pour l'étude de leurs mécanismes. Chaque stratégie a ses cibles privilégiées et ses limites. Leur choix dépend de ce que l'on cherche à étudier. Les méthodes peuvent être combinées, cibler différentes zones de la molécule (ou de l'assemblage de molécules) ou différentes étapes du chemin représentant le processus étudié.

On distingue plusieurs niveaux de complexité dans une protéine, et aussi plusieurs échelles de temps dans son mécanisme. Alors que le temps est géré par la méthode d'échantillonnage, les caractéristiques spatiales sont gérées par le champs de force. Nous avons donc en partant déjà deux niveaux : celui de la méthode d'échantillonnage et celui du modèle et de son champ de force. Une dynamique moléculaire qui utilise un champ de force, comporte déjà deux niveaux : celui du parcours qui génère les conformations consécutives, et celui de la structure ou du modèle, traité par le champs de force.

Le niveau parcours est découpé en étapes et le niveau structure peut être découpé selon différents niveaux de définition. La méthode *Froda* [281] utilise deux niveaux de définition : les blocs qui sont définis grâce à la méthode *FIRST* [113], et le niveau ato-

mique pour les zones flexibles. Elle gère les blocs rigides comme des ensembles d'éléments, alors que les zones flexibles sont gérées selon les atomes individuels. Le niveau structurel a donc ici deux niveaux de définition.

Les méthodes de recherche de passage de transition [31, 167] sont un exemple de découpage du niveau parcours car elles génèrent des étapes. Le niveau parcours dessine la trajectoire au fil des acceptations et des rejets d'événements. Ces étapes sont définies par la technique de convergence et la relaxation. Nous avons donc deux échelles de définition dans le temps : celle du parcours complet et celle de chaque événement du parcours. Nous verrons dans la section 3.3.2 que l'on peut utiliser des méthodes d'interpolation pour mieux définir les états intermédiaires. Dans une simulation informatique, nous pouvons donc fonctionner selon une hiérarchie de quatre niveaux, tel que l'exprime la figure 3.1.

La prochaine section explique les deux principales façons de combiner deux niveaux de définition ou plus de la structure étudiée. La section suivante relate les approches couramment utilisées concernant le discernement des événements dans le temps.

### 3.3.1 Hiérarchies structurales

À l'instar de méthodes telles la RMN et la cristallographie (*B-factors*), les méthodes numériques démontrent que de grandes régions d'une molécule sont plutôt rigides, stables, ou seulement sujettes à des fluctuations locales alors que d'autres sont dédiées aux mouvements importants. Nous pouvons tirer parti de ces différents niveaux de mobilité dans la protéine et combiner plusieurs niveaux de définitions. Ces modèles gros grains où cohabitent plus d'un niveau de résolution sont dits parallèles [264] ou concurrents [285] par opposition aux modèles sériels ou séquentiels, qui les utilisent plutôt consécutivement l'un à l'autre. En d'autres termes, nous pouvons soit fonctionner avec différentes couches de définition que l'on adapte entre elles (fonctionnement parallèle ou en concurrent), ou en effectuant la projection d'un niveau de définition sur un autre (fonctionnement sériel ou consécutif).

#### 1. Couches

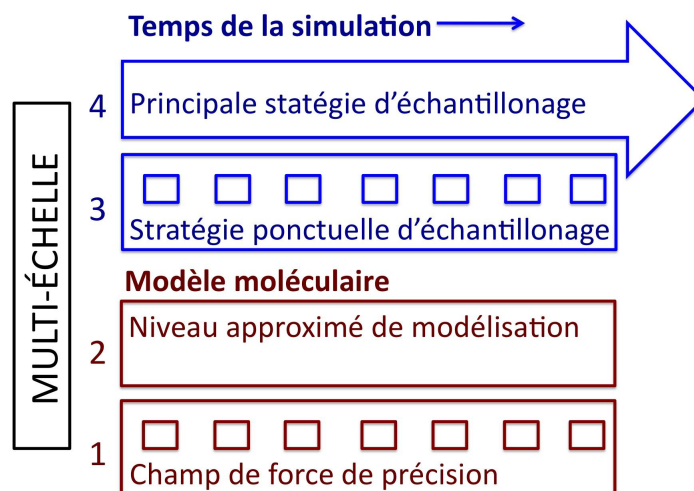


Figure 3.1 – Niveaux des simulations multi-échelle. Nous représentons ici les approches multi-échelles selon deux échelles de temps et deux échelles de structure. Temps : Niveau 1 = Parcours, Niveau 2=Événement. Structure : Niveau 3=Ensemble, 4=Élément).

Il est possible de définir plusieurs degrés de résolution. Ceci s'applique en par exemple aux systèmes où des calculs quantiques sont nécessaires pour des parties ciblées d'un système. Les méthodes QM/MM intègrent des calculs quantiques (*QM*) aux calculs de mécanique moléculaire (*MM*) [23, 185, 211, 248], deux types de potentiel que nous avons introduit à la section 3.1. Les chaînes latérales d'une molécule, dont les mouvements sont limités dans l'espace relativement à la chaîne principale, peuvent être l'objet d'une approximation, alors que tous les atomes de la chaîne principale sont conservés [51, 171], ce qui permet entre autres d'étudier la formation des fibres amyloïdes [176, 228, 279]. L'approche du *NMA* peut utiliser une hiérarchie d'approximation gros-grain [56]. Avec la méthode *FRODA*, les blocs rigides déterminés par la méthode *FIRST* cohabitent avec des zones flexibles, ce qui constitue aussi deux niveaux de définition [90]. Les blocs de *FIRST* sont aussi utilisés dans l'approche combinée *RTB* et *NMA* pour l'étude de protéines globulaires [3], augmentant l'efficacité par rapport aux approches *NMA* de base.

La définition de couches convient aussi aux ensembles de molécules libres, qui

peuvent passer d'une zone à l'autre. Ainsi, dans l'approche de Matej et Delle, une molécule d'eau sera une simple sphère, un triangle ou trio d'atomes, selon la zone qu'elle traverse. Cela permet de concentrer les calculs complexes dans des zones stratégiques, dans le cas d'un solvant autour d'une macro-molécule par exemple [216].

Selon Woodcock et al., la cohabitation de deux niveaux peut être soustractive ou additive. Avec les modèles additifs il faut définir une zone d'interface entre les deux représentations. Les modèles soustractifs vont calculer la zone à haute définition en mode haute précision, calculer tout l'ensemble en mode approximation, puis soustraire ces derniers calculs pour la zone précise [285].

Des structures longilignes régulières s'organisent en plusieurs niveaux dans les tissus biologiques. Le modèle de la corde élastique peut avoir plusieurs niveaux de définition et permettre entre autres de modéliser la structure de la chromatine [233, 272, 276]. On étudie ainsi l'interaction de filaments protéiques pour différents types de fibres : l'actine, qui constituent les fibres musculaires si on monte de plusieurs niveaux [269], ou la kératine, dont plusieurs niveaux d'agencement construisent un cheveu [5]. À chaque niveau de représentation, on effectue un montage de fibres du niveau précédent, qui est réduit à une simple formule de ressort représentant le comportement statistique moyen de ses éléments. Diverses formules d'élasticité peuvent être établies pour définir le comportement en étirement, recourbement ou torsion [142].

## 2. Projection

Ici la molécule est totalement décrite selon les deux niveaux de résolution et on effectue la traduction d'une résolution à l'autre. Des modèles élastiques sont dérivés d'informations obtenues lors de simulations atomistiques [40, 112]. Les simulations atomistiques d'une protéine dans une membrane de phospholipides fournissent des moyennes de fluctuations des coordonnées internes qui sont ensuite utilisées en simulation *NMA* [247].

L'approche peut aussi être inversée. On étudie la formation des fibres de type prion

par une telle approche : les fibres sont formées lors de simulations utilisant le modèle OPEP [51, 171] (qui est lui-même à deux niveaux de résolution concurrents), puis on génère des représentations *tout atome* à partir des résultats obtenus, pour vérifier la stabilité et raffiner les détails [189].

### 3.3.2 Combinaison de techniques pour l'échantillonnage

Le potentiel peut donc être défini en plusieurs niveaux. La méthode d'échantillonnage peut aussi combiner plusieurs techniques. On vise à obtenir un meilleur échantillonnage des conformations possibles, ou à mieux définir les étapes du parcours.

Le raffinement des étapes de transition est utilisé dans le cadre des méthodes de traction et d'interpolation. On évalue un point de selle entre deux points bien connus, ou encore on cherche à explorer une zone imprécise du parcours. Lorsque l'on connaît, parfois approximativement, la conformation de deux minimums locaux voisins, on peut utiliser une méthode d'interpolation entre les deux conformations et ainsi évaluer la montée d'énergie nécessaire (énergie libre d'activation) pour ce passage [192, 245], entre autre avec la technique *Umbrella sampling* [19] (*échantillonnage parapluie*). Certains calculs, qui utilisent la hauteur relative de l'énergie du point de selle par rapport à celle de chacun des deux minimums, peuvent fournir une estimation statistique sur les rythmes de passage [19, 206]. Dans les cas complexes il est plus difficile de savoir si ces conformations représentent véritablement des étapes sur le chemin. Un chemin d'interpolation est souvent établi approximativement mais peut bénéficier de travaux de raffinement grâce à des *essaimages* ponctuels de simulations [207]. Selon Roux et al., une trop grande dispersion de celles-ci indique la nécessité de repositionner le chemin. On appelle *recuit d'une transition* (*reaction path annealing*) certaines techniques d'activation qui permettent l'étude d'un présumé point du parcours et conduisent à une meilleure définition du chemin [59].

En ce qui concerne les dynamiques moléculaires, on peut chercher à rendre l'échantillonnage plus exhaustif. Avec l'approche par échange de répliques, des simulations sont menées à différentes températures. À une température donnée, on obtiendra une certaine diversité dans les énergies des conformations générées. Elles seront en moyenne élevées en proportion de la température choisie. Il se présentera des situations où des énergies

plus basses seront atteintes, correspondant à des énergies atteignables par les simulations roulant à une température un peu plus basse. La simulation peut alors se poursuivre selon cette température plus basse. Ceci permet un échantillonnage plus exhaustif des conformations [36, 139, 230]. Cette méthode s'utilise aussi selon une variante hamiltonienne (qui travaille au niveau des énergies d'attraction et de répulsion entre les atomes non-liés) [138, 287].

Lorsqu'on a pu faire un assez grand nombre de simulations pour atteindre un niveau statistique, on peut classifier les conformations et établir la probabilité de tel ou tel type de conformation ou d'assemblage [45, 189]. Pour les ensembles atomiques qui ont un nombre limité de conformations possibles, il peut être intéressant de répertorier et de mémoriser chacun de ceux-ci, ainsi que les transitions entre les états [175]. Ceci peut permettre d'établir le ou les chemins optimaux, la probabilité de chacun, et discerner les événements rares.

### **3.4 Problématiques et questions ouvertes**

Beaucoup d'approches et de combinaisons d'approches sont donc disponibles pour aborder la complexité des macro-molécules. Mais quels sont les manques encore à combler ?

La fluctuation dans les minimums locaux représente le principal problème de perte de temps CPU. Dans la réalité moléculaire, c'est grâce à cette fluctuation que le passage de transition vers un changement plus important sera trouvé et se produira. Les puits harmoniques permettent d'établir des rythmes moléculaires différents à chaque échelle d'organisation, ce qui est nécessaire au fonctionnement moléculaire. Mais au niveau des simulations informatiques, on veut éviter de perdre du temps dans les fluctuations locales et trouver le chemin hors des puits harmoniques. Trois autres problèmes sont à surmonter : on doit pouvoir faire la distinction entre les conformations permises par la flexibilité et les mouvements effectivement réalisés par la molécule. On doit faciliter la découverte des mouvements d'ensemble. On veut aussi échantillonner le mieux possible l'espace des conformations. Le tableau 3.I donne un aperçu de la recherche de cinq

objectifs selon les différentes stratégies. Dans ce tableau les méthodes sont classées dans l'ordre où elles ont été abordées au cours du chapitre. Il se termine avec l'approche *HMM* (Méthode holographique multi-échelle) développée dans cette thèse. Un *OUI* majuscule indique qu'il s'agit d'un objectif principal pour la méthode. Un plus petit *oui* indique que la méthode vise une amélioration à ce niveau.

Stratégie	Réduire fluctuations locales	Franchir les barrières	Mouvem. collectifs	Augmenter réalisme mouvements	Echantill. exhaustif
Gros grains	oui		OUI		
Dynamiques moléculaires		oui		OUI	
Extrapolations (NMA)	OUI		OUI	oui	
Recherche des points de selle	OUI	OUI		OUI	
Tractions	OUI		oui		
Intrapolations	OUI		oui		
Couches	OUI		OUI	oui	
Projections	OUI		OUI	OUI	
Échantillonnage parapluie	OUI	OUI	oui	oui	
Essaimage				OUI	
Recuit		OUI			OUI
Échange de répliques		OUI			OUI
Mémorisation des états	OUI				OUI
HMM	OUI		OUI	OUI	

Tableau 3.I – Buts principaux selon les stratégies de simulation. Les OUI en lettres majuscules indiquent les buts principaux recherchés par les méthodes en question. Les oui en lettres minuscules indiquent ce que la méthode peut aussi favoriser, mais dans une mesure moindre. HMM est la méthode qui est développée dans cette thèse.

Voici une description des colonnes du tableau 3.I

1. **Réduction des fluctuations locales.** On veut éviter de voyager indéfiniment dans

un puits harmonique.

2. **Franchissement des barrières.** On veut découvrir et/ou étudier les passages de transitions.
3. **Mouvements collectifs.** On veut découvrir et réaliser des mouvements coopératifs.
4. **Réalisme des mouvements.** On veut éviter les parcours erronés. On veut éviter de confondre flexibilité et mouvement effectif.
5. **Échantillonnage exhaustif.** On veut connaître toutes les conformations possibles, leur probabilité respective et les chemins entre elles.

Il apparaît clair que les mécanismes protéiques se produisent à différentes échelles de leur architecture. Certaines études s'intéressent aux mouvements amples et se prêtent bien aux approximations qui passent sous silence les détails mécaniques plus fins. Par contre certains mécanismes se présentent sous forme d'un enchaînement séquentiel où la survenue d'un événement spécifique dépend des conditions mises en place par le mécanisme précédent. Nous le constatons lors du montage des niveaux des grandes structures protéiques, alors que certains résidus ne jouent un rôle que rendu au dernier niveau du montage multidimensionnel, assurant la forme de l'architecture globale (voir section 2.2.4), ou lorsqu'un site ne devient actif qu'après le montage précis des niveaux de la structure protéique. Le mécanisme d'une protéine peut présenter plus d'une seule phase, alors que des éléments ponctuels entrent en jeu en cours de route pour changer l'état général et réorienter son déroulement.

Les stratégies d'accélération sont soumises au risque de manquer de discernement car elles déplacent beaucoup d'éléments à la fois, ou effectuent de grands mouvements sans discernement des détails. Les méthodes d'interpolation intègrent de plus en plus de techniques visant à découvrir et étudier les transitions qui impliquent des barrières d'énergies. Les méthodes d'interpolation ou d'activation de point de selle, tels les *échantillonnages parapluie (umbrella sampling)*, *recuits*, *essaimages*, permettent de simuler



une transition, en autant qu'on ait la chance de connaître les conformations voisines séparées par ce passage, sinon elles permettent de l'évaluer et de l'affiner.

Les modes du *NMA* démontrent la flexibilité de la protéine, mais on comprend de plus en plus qu'on doit faire la différence entre flexibilité et mouvements effectifs. Les modes demandent une étude plus complexe pour savoir combien d'entre eux sont nécessaires pour représenter les mouvements corrélés de la protéine (qui sont davantage en rapport avec le mécanisme réel de la protéine) car les modes harmoniques peuvent être contradictoires entre eux [286]. Dans certaines situations l'un des modes harmoniques se retrouve bloqué dans l'un de ses états par d'autres éléments [162]. De plus, les fluctuations harmoniques, même lentes, se font aux alentours du bassin d'attraction d'une configuration à l'équilibre, et ne sont pas aussi valables passé une certaine distance de coupure [259]. La protéine est susceptible d'être attirée par d'autres bassins d'attraction [225]. Des efforts vont dans le sens d'élargir la méthode *NMA* pour intégrer ces divergences aux cas "réguliers" [119, 225].

Les méthodes de passage de transition [17, 31, 167] permettent de simuler des chemins complexes sans biais sur le trajet ou la destination. Utilisées avec des représentations atomistiques, elles améliorent dans une bonne mesure la rapidité de l'échantillonnage des conformations environnantes. Pour augmenter la capacité d'atteindre des conformations plus éloignées, on peut aussi les utiliser avec des approximations d'échelle. Mais pour les protéines aux transformations plus complexes le principal effet de longs déplacements est de faire ressortir l'erreur présente dans l'approximation du champ de force. Comme nous le verrons au chapitre 8, les déplacements d'une certaine envergure sont guidés par des contraintes spatiales qui réduisent le nombre de chemins possibles. Ces contraintes risquent d'être mise inadéquatement sous silence par les approximations.

Les changements de conformation sont susceptibles de réorganiser la définition des zones rigides de la molécule [168]. Une reformulation de la définition du haut niveau est nécessaire si la nouvelle conformation implique une réorganisation des ponts hydrogènes, dont dépend particulièrement la définition des zones rigides. Lorsqu'un changement de configuration modifie le profil de rigidité d'une molécule, il en redéfinit par conséquent l'élasticité et la flexibilité. Ce problème se présente aussi si nous voulons

étudier l'effet des entrées en contact d'une protéine avec les autres molécules participant à sa fonction. Plusieurs ligands ont justement comme rôle de changer le profil de rigidité/flexibilité de la molécule.

Les dynamiques moléculaires *tout atome* standard permettent de conserver à tout moment une mouvance dans la répartition des zones rigides et flexibles et elles déplacent les éléments entre eux selon un calcul réaliste des forces et énergies. Nous avons avantage à préserver le mieux possible ces deux conditions au fil des simulations multi-échelle pour assurer leur réalisme. Si l'on se fie aux promesses du tableau 3.I au sujet de la méthode *HMM*, et aux capacités des méthodes de recherche de points de selle (voir même tableau), on peut espérer que la combinaison de *HMM* et de la technique de recherche de point de selle *ART nouveau* [17, 163, 167, 174] permette de découvrir des mouvements d'ensemble significatifs et réalistes. Les chapitres 5 et 7 décrivent *HMM* (Méthode multi-échelle holographique).

## CHAPITRE 4

### INTRODUCTION AUX PROTÉINES À MAINS EF

Les tests de la méthode ont d'abord été effectués sur des modèles de protéines artificiellement étendus, à 15 ou 20 Å du modèle natif correspondant. Ces travaux ont permis d'évaluer la capacité de *HMM*, combinée à *ART nouveau*, à générer des mouvements d'envergure. Ils sont décrits au chapitre 7. Mais le but de la méthode est l'étude de mécanismes réels de protéines dans l'accomplissement de leur fonction. Il nous fallait cibler une classe de protéines accessibles à une toute première application de *HMM*, en tenant compte des capacités et limites actuelles de la méthode. Pour valider la capacité de la méthode à générer des trajectoires entre conformations distantes, il nous fallait aussi avoir sous la main deux modèles distants pour chaque protéine étudiée, provenant de la *RMN* ou de la *cristallographie*. De préférence, le mécanisme de ces protéines devait avoir fait l'objet de beaucoup d'études expérimentales. En résumé, les protéines recherchées devaient répondre aux critères suivants :

1. Dimensions petites à moyennes.
2. Présence de plusieurs structures secondaires relativement stables.
3. Boucles flexibles non contraintes.
4. Transformations non-harmonique, de bonne envergure, difficiles à étudier par les autres méthodes informatiques.
5. Disponibilité de modèles distants pour la même protéine, provenant de la *RMN* ou de la *cristallographie*.
6. Plusieurs études in vitro disponibles concernant le mécanisme.

Nos essais préliminaires ont démontré que certaines protéines à mains EF répondaient à tous ces critères et se prêtaient bien à nos simulations holographiques. Nous

pouvons démarrer les simulations à partir de l'un des modèles de la protéine, dont nous modifions d'abord les conditions pour favoriser une transformation en direction d'un autre modèle. Beaucoup de domaines mains EF adoptent une conformation ouverte en présence de calcium, tandis que l'absence de calcium favorise la fermeture du domaine. Nous pouvons donc soit démarrer des simulations à partir de la conformation ouverte connue, en retirant le calcium, ou démarrer à partir de la conformation fermée à laquelle nous ajoutons un modèle de l'ion calcium.

Les protéines à mains EF ont bénéficié de plusieurs études expérimentales pour mieux connaître leur interaction avec les diverses protéines cibles. Elles sont présentes dans toutes les cellules eucariotes et servent de second messenger pour un grand nombre de processus déclenchés par l'arrivée d'ions  $\text{Ca}^{2+}$  [235, 294, 295]. La présence de calcium favorise l'ouverture de leur cœur hydrophobe [39, 63, 108], elles lient alors des protéines cibles. Elles modulent ainsi des processus telle la signalisation [70, 91, 197], la plasticité cérébrale [7, 96, 195], la transmission nerveuse [172], la contraction musculaire [49, 68, 241], la vision [6, 256] et l'endocytose [98, 251]. Certains domaines à mains-EF, tel celui de la calbindine, se transforment peu en présence de calcium et sont plutôt réputés servir de tampon pour son emmagasinage [108].

#### **4.1 Le motif main-EF**

Le motif main-EF est formé d'une hélice reliée à une autre par une boucle flexible, généralement capable de lier l'ion  $\text{Ca}^{2+}$ . Une boucle typique de liaison au calcium comporte 12 résidus [87] dont certaines positions sont très conservées [304]. Différentes affinités sont possibles, faisant en sorte que certains domaines lient le calcium de manière endémique, alors que d'autres le lient seulement en présence d'un surplus de calcium dans la cellule. Les subtilités dans la composition des boucles permettent donc un étalonnage fin des divers mécanismes selon le taux de calcium. La figure 4.1 explique le modèle de référence de la boucle liant le calcium.

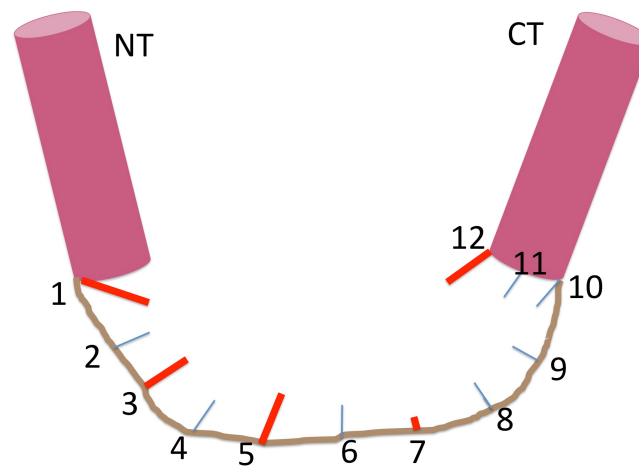


Figure 4.1 – Schéma d'une main EF. Les 12 positions types d'une boucle de liaison au calcium, sises entre deux hélices  $\alpha$ . Les positions 1, 3, 5 et 12 sont souvent des résidus porteurs d'oxygène(s) permettant de lier le calcium. Le calcium peut être aussi coordonné par l'oxygène de la chaîne principale en position 7.

## 4.2 Domaines à mains EF

La majorité des motifs à mains EF fonctionnent par paire, la seconde hélice du premier motif étant reliée par une boucle à la première hélice du deuxième motif. Les expériences de Grabarek ont démontré que les deux mains-EF d'un domaine sont aussi liées entre elles par une structure en feuillet  $\beta$  qui joue un rôle important lors du mécanisme d'ouverture [94]. Les deux boucles de liaison au calcium se longent dans leur partie centrale et leur chaîne principale forme entre elles des ponts hydrogènes. La figure 8.5 du chapitre 8 illustre la consolidation du feuillet suite au départ du calcium. Selon Grabarek, l'arrivée du calcium restructurerait de manière importante cette région, qui serait à l'origine de la tendance au domaine de s'ouvrir ou de se refermer [95]. Meyer, Mabu-chi et Grabarek ont exploré par mutagenèse l'importance de la phenylalanine précédant immédiatement la première boucle de liaison au calcium du domaine C-terminal. Ce résidu est situé hors du coeur hydrophobe dans la conformation fermée et stabiliserait celle-ci. Il serait aussi important pour la conformation ouverte, pour assurer l'affinité aux ligands [177].

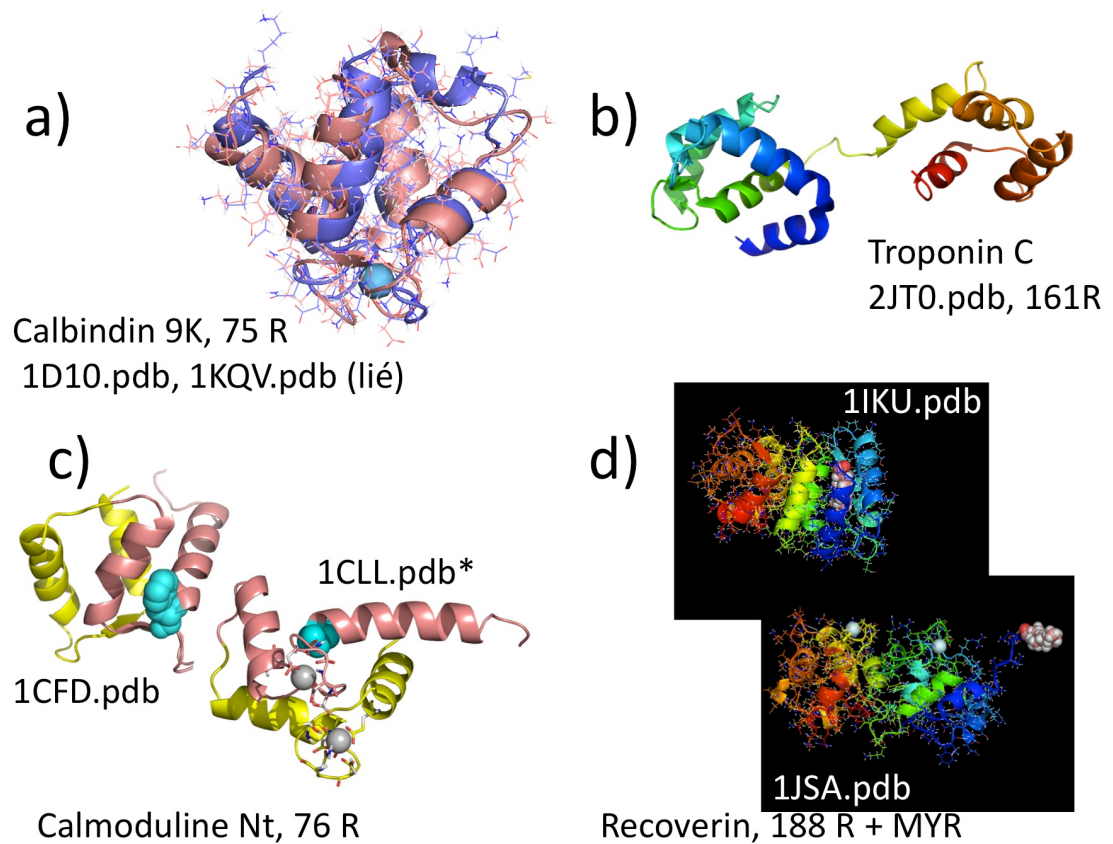


Figure 4.2 – Protéines à mains-EF. (a) Calbindine 9K : En bleu, liée au calcium. En pêche, non liée. Formée d'une paire de motif main-EF. (b) Troponine C apo : domaine NT à gauche, domaine CT à droite. (c) Domaine NT de la calmoduline : apo à gauche, lié à deux calcium à droite. La première paire EF est représentée en pêche et la deuxième en jaune. (d) La recoverine avec sa terminaison "myristoyl" externalisée lorsque la protéine est liée au calcium (1JSA.pdb) mais internalisée en absence de calcium (1IKU.pdb). Images générées avec pymol.

Chaque domaine est donc formé de quatre hélices  $\alpha$  pour la majorité des protéines EF. La figure 4.2 (a) illustre la calbindin 9k, qui comporte effectivement deux mains-EF en un seul domaine. L'image montre une superposition de la conformation liée au calcium avec la conformation non liée au calcium. Sous l'action du calcium les calbindines effectuent une transformation locale, sans ouverture ample de leur conformation [191]. Celles-ci sont particulièrement présentes dans les astrocytes et modulent entre autre l'assemblage des micro-tubules [195, 204].

Bien des protéines à mains EF comportent deux domaines, dits N-terminal et C-terminal, reliés par une zone flexible [11, 237, 275] qui prend dans certains cas la conformation d'une longue hélice. Tel est le cas de la calmoduline et de la troponine C, qui adoptent alors la forme d'une haltère. La figure 4.2 (b) montre la troponine C lorsque cette zone de liaison entre les domaines NT et CT est déstructurée. Alors que le domaine CT de la Troponine C est toujours lié à la troponine I, sa cible, le domaine NT lie la troponine I seulement en présence d'un surplus de calcium dans la cellule. Le domaine ouvre et accapare alors la troponine I, qui empêchait la contraction musculaire. L'arrivée du calcium provoque donc indirectement la contraction musculaire [34, 68, 82, 154, 155, 239, 243].

La figure 4.2 (c) montre un seul domaine de la calmoduline, le domaine N-terminal, dans sa version apo et fermée à gauche, dans sa version ouverte et liée au calcium à droite. Les deux hélices composant chaque motif EF-hand prennent en présence du calcium une position perpendiculaire. La première main EF est représentée en couleur pêche et la deuxième en jaune. La calmoduline est ubiquitaire dans les cellules eucariotes. Elle est présente dans de nombreux processus qui doivent être activés ou désactivés par l'arrivée des ions  $\text{Ca}^{2+}$  dans la cellule [235, 294, 295]. En ouvrant sous l'influence du calcium un domaine à mains-EF, elle se liera par exemple à un domaine de canal cellulaire, qui changera alors son activité.

La recoverine est un autre exemple de protéine à mains EF qui effectue une transformation d'envergure avec l'arrivée ou le départ du calcium. Tout comme pour la troponine C et la calmoduline, la recoverine comporte deux domaines à mains-EF, pour un total de quatre boucles de liaison au calcium. La figure 4.2 (d) montre des changements majeurs

entre les conformations liées et non liées. La recoverine est impliquée dans le processus de la vision. En présence de calcium, la recoverine extériorise un appendice "myristoyl" dont le rôle est d'inhiber une protéine influencée par la lumière, la rhombopsine. En absence de calcium, la recoverine enferme sa terminaison "myristoyl" [6, 256].

### 4.3 Études informatiques

Afin d'explorer plus à fond certaines propriétés, les protéines à mains EF ont aussi été simulées par ordinateur. C'est particulièrement le cas pour la calmoduline. La table 4.I liste différents types d'études ciblant des protéines à mains EF. Les dynamiques moléculaires tout-atome non biaisées ont permis d'explorer le comportement du lien central entre les domaines NT et CT de la calmoduline et de la troponine, ainsi que le positionnement relatif de ces domaines et l'affinité des domaines ouverts à plusieurs protéines cibles. Elles n'ont cependant pas permis de simuler une trajectoire complète entre les formes fermées et ouvertes d'un domaine à mains-EF.

Les études des modes normaux s'appliqueraient mal à des protéines telle la calmoduline, parce qu'avec la complexité de leur trajectoire, ces modes changeraient en cours de route [254]. Les dynamiques biaisées utilisant le potentiel de Gō fournissent pour leur part des trajectoires générales, mais limitées dans les détails à cause de la représentation de chaque acide aminé par une seule sphère.

Les simulations par traction du calcium dans l'étude de Zhang et al [301], en tout atome, confirment le rôle de la transformation du feuillet  $\beta$  dans la trajectoire de transformation du domaine CT de la calmoduline, tel que décrit dans les études de Grabarek. Nous avons pour notre part étudié en tout atome et sans biais, les domaines NT de la calmoduline et de la troponine C. Cette étude est publiée dans l'article décrit au chapitre 8. Nos simulations avec *HMM* explorent plus à fond le rôle du feuillet  $\beta$  pour la trajectoire de fermeture, ainsi que les étapes d'interaction entre certains acides aminés clef, qui ont des rôles coopératifs, ou agissent comme contraintes stériques et comme voie de passage pour guider la trajectoire.



Tableau 4.I – Études de protéines à mains-EF par simulation.

Méthode utilisée	Portée de l'étude
Dynamiques moléculaires tout atome	Déplacement relatif domaines NT et CT, lien central - protéine Mlc1p avec ligand 1Q4 [84] - TPNc [137], avec ligands [100, 268] - CAM [60, 77, 137], avec ligands [140, 290] Calmoduline NT lié au calcium : fluctuation [61, 271] CAM boucle(s) de liaison et ions [129, 147] Calmoduline NT lié au calcium : reconnaissance/liaison de différents ligands [76] Calbindine 9K : couplage entre les boucles de liaison [169]
Simulations biaisées - potentiel de $G\bar{o}$	Trajectoire générale, ouverture, fermeture domaine [265, 266, 300]
- traction du calcium	Fermeture domaine CT de la calmoduline : transitions dans le feuillet $\beta$ [301]
Normal Mode Analysis	Peu applicable à la calmoduline [134, 254]

## CHAPITRE 5

### STRATÉGIE ET ORGANISATION

Nous développons ici le cadre théorique et organisationnel qui sous-tend les réalisations décrites aux chapitres 7 et 8 (articles). Dans un premier temps nous examinons trois notions issues des travaux précédents : le rôle de la coopérativité et des directions de l'espace dans les niveaux d'organisation, la modélisation gros grain par blocs rigides et les représentations par coordonnées internes. À partir de celles-ci, nous traçons les lignes stratégiques de la nouvelle méthode multi-échelle, puis nous décrivons l'organisation logique qui permet son utilisation sur des modèles protéiques. La technique développée est cependant décrite plus en détail dans le chapitre 7, et les travaux relatifs aux champs de force sont expliqués au chapitre 6.

#### 5.1 Vision stratégique

Dans un premier temps, voici un survol indiquant l'importance des jeux de directions et de coopérativité dans l'espace tri-dimensionnel, autant pour l'agencement des domaines de protéines entre eux, que pour l'agencement des protéines entre elles. Ceci constitue la principale notion à la base de la méthode développée, qui mise sur une stratégie de déplacements d'envergure, avec adaptation fine des sous-éléments, le tout selon des champs de force réaliste. Les volets de cette stratégie sont ensuite expliqués.

##### 5.1.1 Directions de l'espace, coopérativité et étages

Commençons par examiner la situation dans une optique large. La hiérarchisation intérieure à une protéine implique consécutivement les trois directions de l'espace. Il en va ensuite de même pour son montage en fibres, filaments ou complexes. Alors que les acides aminés s'assemblent, la structure se déploie dans une direction de l'espace. Lorsqu'un pont hydrogène se forme, la structure se déploie dans deuxième direction de l'espace. Mais cela peut se défaire facilement. Mais avec plusieurs ponts hydrogènes

parallèles collaborant dans la même direction, on diminue la possibilité de défaire ce deuxième étage : par exemple avec une spire d'hélice  $\alpha$  [2] ou avec deux brins  $\beta$  qui se longent. Cela veut dire que la probabilité d'aller vers le désordre est diminuée car le système s'est fermé sur lui-même, protégé par une barrière énergétique. On a accédé à un creux dans le paysage énergétique.

Avec les probabilités cinétiques, les éléments désorganisés iront tôt ou tard se mouler le long de la structure stable et celle-ci s'allongera. Puis une troisième direction de l'espace peut entrer en jeu. Les feuilletts  $\beta$  se collent grâce aux forces hydrophobes et forment des fibres telles celles des vers à soie [103]. La coopérativité des nombreuses interactions parallèles hydrophobes constituent une grande barrière énergétique qui rend ces fibres très difficiles à défaire. Les hélices  $\alpha$  aussi, en se longant, se consolident en largeur (donc selon la troisième direction de l'espace), qui donne davantage de solidité à la précédente direction de l'espace. Par exemple les hélices  $\alpha$  de tropomyosines (un des nombreux constituants des fibres musculaires) se longent et s'entrelacent, formant des *coiled-coil* [106].

Les assemblages de structures secondaires en protéines globulaires impliquent aussi ainsi la troisième dimension de l'espace. Si on va plus loin, les microtubules offrent un exemple de montage impliquant consécutivement cinq changements de direction : trois directions de l'espace sont consécutivement sollicitées comme pour le montage de la tubuline qui est une protéine globulaire. Ensuite une suite de globules sont assemblés en filaments dans une direction de l'espace. Puis treize filaments se longent, formant une courbure dans la direction transversale, qui se ferme en un tube [92], flexible et bien plus solide. Dans ces montages, les niveaux additionnels s'établissant dans une direction spatiale différente protègent les sous-niveaux contre les incidents aléatoires. La flexibilité des niveaux inférieurs détermine aussi le comportement des niveaux supérieurs, car la résistance du niveau supérieur dépend de la résistance des niveaux inférieurs. Selon les études de Ackbarow et al, les hélices  $\alpha$  des vimentines se déplient pour absorber l'énergie du poids déformant excessif que subissent les étages supérieurs de ces structures de filaments parcourant la cellule [1].

Chaque étage d'assemblages moléculaires contient un certain nombre de protéines

(ou autres molécules) en un agencement régulier (tissus élastiques et fibres) ou irrégulier (complexes moléculaires). Ceci rappelle les agencements réguliers et irréguliers internes à une protéine. Avec ses assemblages de structures  $\alpha$ , de feuillets  $\beta$  et de boucles diverses, une protéine est souvent plus complexe qu'une fibre, si l'on considère les sous-éléments de la fibre comme de simples blocs.

### 5.1.2 Les blocs et leur utilisation

Un bloc est une région, dans une molécule ou un assemblage moléculaire, que l'on considère rigide par rapport aux autres parties plus fluctuantes. Jacob et Thorpe étudient les zones rigides à l'intérieur des protéines à partir du réseau des interactions fortes reliant des sommets représentant les atomes. Leur algorithme *FIRST* [113, 114] discerne les régions flexibles des régions rigides en déplaçant d'un sommet à l'autre des jetons représentant les degrés de liberté (*pebble game*). Ces blocs peuvent ensuite être utilisés en simulation selon la méthode *FRODA*. Cette technique effectue de légers déplacements aléatoires de ces régions en bloc, tout en simulant les zones flexibles en tout atome. Elle est utilisée pour interpoler la trajectoire possible entre deux conformations [78]. Gohlke et al. appliquent la technique *FRODA* en intégrant les blocs comme éléments dans un réseau élastique [90] ou utilisent les blocs dans une approche *NMA* [4]. Ceci permet d'étudier des mouvements larges tel ceux inhérents à la protéine liant la maltodextrine, la protéine barnase, ainsi que l'isomérase disulphide [117] ou de se concentrer sur les parties flexibles de structures aussi complexes que le tunnel du ribosome [80, 81].

Tama, Sanejouand et Brooks et al. combinent aussi des blocs à l'approche *NMA*, via un modèle gros grain de blocs (*RTB*) de six acides aminés [252, 253], pour l'étude de protéines de plusieurs centaines d'acides aminés (HIV-1 protease, Triglyceride lipase, myosine) ou l'étude du ribosome [252, 253].

### 5.1.3 Les coordonnées internes

L'approche *NMA* basée sur les blocs peut utiliser des modèles exprimés en coordonnées internes. Celles-ci représentent les éléments les uns relativement aux autres plutôt

que selon une base absolue. Par exemple Chu et Voth représentent chaque monomère du filament d'actine par quatre sphères gérées relativement les unes aux autres [40]. Ceci établit deux niveaux de simulation, les monomères étant les pièces du réseau élastique supérieur. L'utilisation de coordonnées internes permet une certaine indépendance entre les niveaux de simulation : les monomères sont déplacés comme un seul élément en mode NMA, mais les quatre sphères composant chaque monomère sont aussi déplacées de manière interne entre elle.

En dynamique moléculaire par traction, les coordonnées internes permettent une représentation simplifiée utilisant les angles entre les atomes de la chaîne peptidique [223]. Des molécules en tout atome sont simulés avec la technique de passage de transition *ARTIST* [297] qui utilise avec succès les coordonnées internes du programme *LIGAND* [143, 144].

#### 5.1.4 *ART nouveau*

La méthode *ART* représente la première implantation d'échantillonnage de conformations par passages d'états de transition [17, 18, 167, 181, 182]. D'abord utilisée pour l'étude des verres amorphes, *ART* a permis d'étudier le repliement [246, 277] et l'agglutination des protéines [176, 228]. *ART* fonctionne en quatre étapes :

1. Activation d'un ou de groupes d'atomes choisis aléatoirement, dans une direction aléatoire.
2. Convergence vers une valeur propre négative du système, suivie jusqu'à ce que les forces soient près de zéro. Ceci définit un point de selle. Nous poussons alors légèrement plus loin, pour que la relaxation qui suivra ne retourne pas dans le bassin précédent.
3. Relaxation vers le nouveau bassin.
4. Acceptation selon un critère de température de Métropolis (équation 5.1) .

$$E - E_{ref} < -(0.002 \cdot T \cdot \log(\text{random}[0 - 1.0])) \quad (5.1)$$

### 5.1.5 Stratégies développées dans ce projet doctoral

Le projet principal de la thèse a consisté à développer une approche multi-échelle, appelée *HMM* ou *Holographic Multiscale Method*, utilisée en conjonction avec la méthode d'échantillonnage *ART nouveau*. Le terme *holographique* est expliqué à la section 5.2. La méthode *HMM* n'a pas été développée à partir de méthodes multi-échelle existantes. Elle inclut cependant certains principes connus : celui des blocs et celui des coordonnées internes, mais avec des différences importantes. En ce qui concerne les blocs, les modèles multi-échelle existants ne tiennent pas compte du changement de la répartition des zones rigides au cours du fonctionnement moléculaire. Les blocs rigides conservent les mêmes définitions qu'au début des simulations. Il en est de même dans les modèles élastiques, bien que le filament d'actine de Chu et Voth utilise des blocs avec certaines fluctuations internes [40]. Nous avons élargi la définition de bloc de deux façons. Un bloc peut basculer et se mouvoir par rapport à l'ensemble dont il fait partie. Au lieu d'être déplacé aléatoirement ou selon un champ de force prédéfini, un bloc aurait avantage à être déplacé selon un champ de force réaliste qui s'adapte au fur et à mesure du déplacement. Dans un contexte de transformation complexe, la précision des interactions devrait être augmentée et non pas être diminuée. On doit pouvoir suivre l'évolution des forces et énergies au cours du mouvement. Deuxièmement, les blocs devraient pouvoir se transformer de l'intérieur, et la définition de leurs frontières devrait changer et s'adapter en cours de simulation. *HMM* réalise des déplacements d'ensemble mouvant plutôt que des mouvements de blocs rigides. La transformation intérieure est facilitée par l'utilisation de coordonnées internes.

La stratégie du projet est de rechercher au fur et à mesure de la simulation, des pivots et des directions de l'espace permettant le basculement des parties de ce niveau supérieur, que ce soit pour le former ou le déformer. La détection de lieux et de directions précises permettra d'enclencher ou de défaire les forces de coopérativité entre deux zones de l'assemblage. Les caractéristiques souhaitées sont de cinq catégories :

1. **Mouvements d'ensemble.** Génération de mouvement d'ensemble plutôt que des fluctuations locales. De grandes zones de la protéine basculent par rapport à une

sélection de pivots.

2. **Passage de barrières.** Capacité de simuler des trajectoires comportant des barrières énergétiques : simulations conduites par une méthode de recherche de transitions énergétiquement favorables (*ART nouveau*).
3. **Mutli-échelle adaptatif.** Définition dynamique des blocs de haut niveau, qui peuvent se redéfinir au cours d'un mouvement, selon l'évolution des forces atomiques.
4. **Précision de niveau atomistique.** Utilisation de champs de forces atomistiques et non-biaisés, avec définition implicite du solvant. Déplacement des zones de haut niveau par basculement et rotation selon les indications du champ de force atomistique. Les blocs de haut niveau conservent une définition atomique précise. Les forces qui leur sont appliquées sont la projection des forces du potentiel atomique. La relaxation est en mode tout-atome après chaque événement, pour parfaire l'adaptation des formes. On peut utiliser avec la méthode différents potentiels : atomiques, gros grain, ou une combinaison des deux.
5. **Analyse des interactions.** Capacité de consulter la contribution énergétique de chaque résidu et de chaque atome pour chaque conformation générée. Les graphiques produits permettent de mieux discerner les enchaînements de cause à effet, ainsi que les phénomènes de coopérativité.

## 5.2 Organisation du logiciel

Le modèle multi-échelle développé *HMM* est utilisé dans le cadre de simulations avec la technique de recherche d'états de transition *ART nouveau* [17, 167]. Les deux méthodes forment ensemble la méthode *ART holographique*, qui permet l'étude des possibilités de transformations spatiales des molécules, mais sans notion de température ni d'échelle de temps moléculaire. La méthode se prête à l'utilisation de potentiels en solvant implicite. *ART nouveau* assure les niveaux 4-Parcours et 3-Évènement tel que le montre la figure 5.1. Ce tableau est la version ART-holographique du tableau plus

général 3.1 du chapitre 3. Le niveau trajectoire est formé des conformations énergétiquement avantageuses choisies par *ART nouveau* selon un critère de température Métropolis (équation 5.1). *HMM* réalise le niveau 2-Ensembles du tableau 5.1 grâce aux mouvements d'ensemble qu'elle permet de générer. Les champs de force atomiques réalisent le niveau 1-Éléments.

Les événements de transition alternent avec les relaxations en tout atome. Selon la classification de Tozzini [264], ces deux niveaux de représentations sont utilisés de manière sérielle ou séquentielle plutôt que parallèle ou concurrente : les blocs se répartissent sur la molécule au complet et leurs forces sont reconstituées à partir de celles du modèle atomistique. L'alternance entre les deux niveaux se fait au rythme de l'alternance entre transition en mode multi-échelle et relaxation tout atome. Mais cela est aussi effectué à chaque instant d'une transition, car à chaque pas de celle-ci les forces de basculement des blocs sont réévaluées à partir des nouvelles forces atomiques.

Avant de poursuivre, nous devons définir sur quelle base nous construisons les blocs de haut niveau. Notre optique n'est pas de les utiliser comme des régions stables, mais comme des régions qui peuvent basculer en un seul ensemble. Or, tout changement d'importance au sein d'une protéine survient par rotation autour des angles diédraux  $\phi$  et  $\psi$  [85] situés de part et d'autre des  $C\alpha$  de la chaîne principale. La figure 5.2 montre que seuls les changements dans les angles de torsion de la chaîne principale sont susceptibles de générer des transformations importantes. Prenant racine aux carbones  $\alpha$ , les chaînes latérales n'effectuent pour leur part que des mouvements locaux. La figure 5.3 illustre le passage d'un acide aminé d'une conformation à l'autre, par basculement des plans peptiques de part et d'autre de son  $C\alpha$ . Ces basculements sont la conséquence de changements dans les angles de torsion  $\phi$  (angle diédral entre les atomes N, $C\alpha$ ,C,N) et  $\psi$  (angle diédral entre les atomes C,N, $C\alpha$ ,C). L'angle diédral  $\omega$  ( $C\alpha$ ,C,N, $C\alpha$ ) situé à l'intérieur de la liaison peptique rend celle-ci planaire et rigide. Nous générerons donc nos mouvements en utilisant les rotations et les basculements des plans peptiques les uns par rapport aux autres. La figure 5.4 montre que chaque plan peptique peut être représenté par une base orthogonale. La base orthogonale peut être relative à celle du plan précédent. Les carbones  $\alpha$  situés entre chaque jonction peptique constituent alors



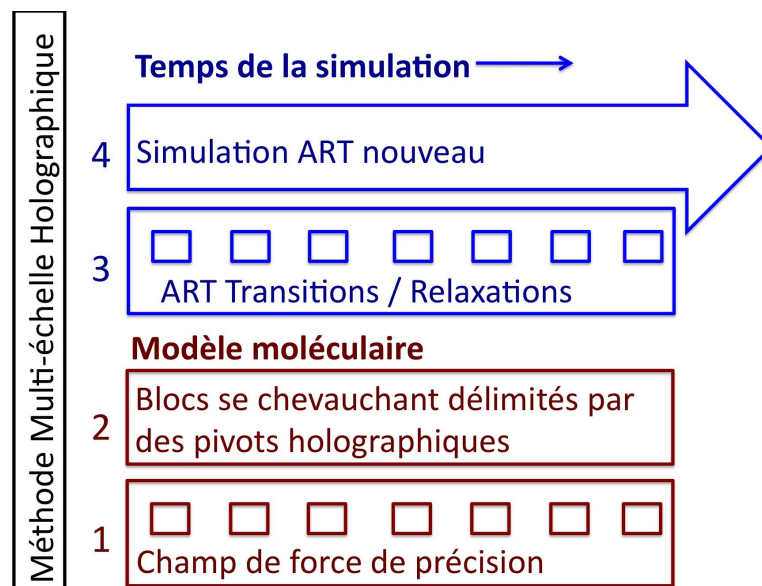


Figure 5.1 – Les quatre niveaux de *ART holographique*. *HMM* telle qu'utilisée en combinaison avec la méthode *ART nouveau* [17, 167]. *ART nouveau* génère des trajectoires (niveau 4) en choisissant des événements (niveau 3). Les événements sont issus de déplacements d'ensemble permis par *HMM* (niveau 2), et d'ajustements atomiques plus élémentaires (niveau 1).

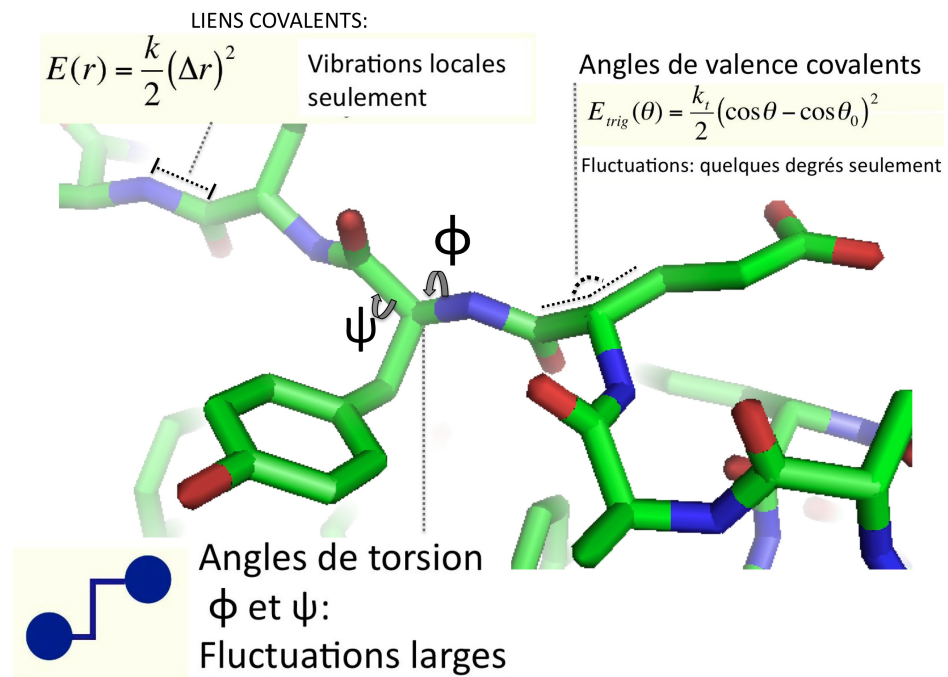


Figure 5.2 – Nature des mouvements amples dans une protéine. Les changements d'envergure se produisent seulement via des changements dans les angles de torsion entre quatre atomes consécutifs de la chaîne principale.

les uniques pivots qui sont nécessaires à la méthode *HMM*.

Les définitions de blocs se superposent : on obtient ainsi une définition du comportement dynamique et adaptatif du haut niveau plutôt qu'une définition rigide. Le modèle projette les forces atomiques sur les multiples définitions de blocs, pour déterminer leur influence sur leur rotation et leur basculement, de façon à mouvoir de grandes zones en un seul ensemble. Les blocs contiennent en eux-mêmes les pivots de d'autres blocs et peuvent donc se transformer en cours de route selon les changements induits dans les forces atomiques. Le modèle est dit *holographique* parce que la tendance de la molécule complète est projetée sur chaque  $C\alpha$  défini comme pivot. Ceci rappelle les plaques photographiques qui génèrent les hologrammes. Chaque point de ces plaques comporte une image complète de l'image tridimensionnelle, mais vue sous un angle différent des autres.

Jusqu'à 90% de des atomes d'une protéine forment naturellement des zones ordonnées [14], majoritairement des hélices  $\alpha$  et les feuillets  $\beta$ . Nous pouvons exploiter la présence de zones stables en définissant des  $C\alpha$  comme pivots seulement dans les extrémités de celles-ci et dans les boucles qui les relient. Ces zones seront tout de même relaxées en tout-atome après chaque transition.

Parce que cette formulation multi-échelle est dynamique, elle présente l'avantage de pouvoir être mise à l'épreuve à partir d'une certaine désorganisation structurale de la protéine native. Nous avons donc prévu des tests posant certains défis, qui consistent à démarrer les simulations à partir de formes partiellement dépliées par rapport à la forme native. Ils sont décrits au chapitre 7.

Le logiciel *HMM* a été construit pour simuler la transformation à l'intérieur d'une protéine, mais aussi en prévision de simulations sur des montages de molécules. La méthode d'échantillonnage *ART nouveau* peut manipuler des vecteurs de positions et de forces sans avoir besoin de connaître leur nature holographique ou cartésienne (figure 5.5). Les étapes d'activation, de transition et de relaxation de ART font toutes appel à la même fonction de calcul des forces, à laquelle *ART* fournit l'état courant des positions afin d'obtenir l'état courant des forces et énergies pour les besoins de ses algorithmes de convergence. Le type et niveau d'architecture impliqué est caché dans la

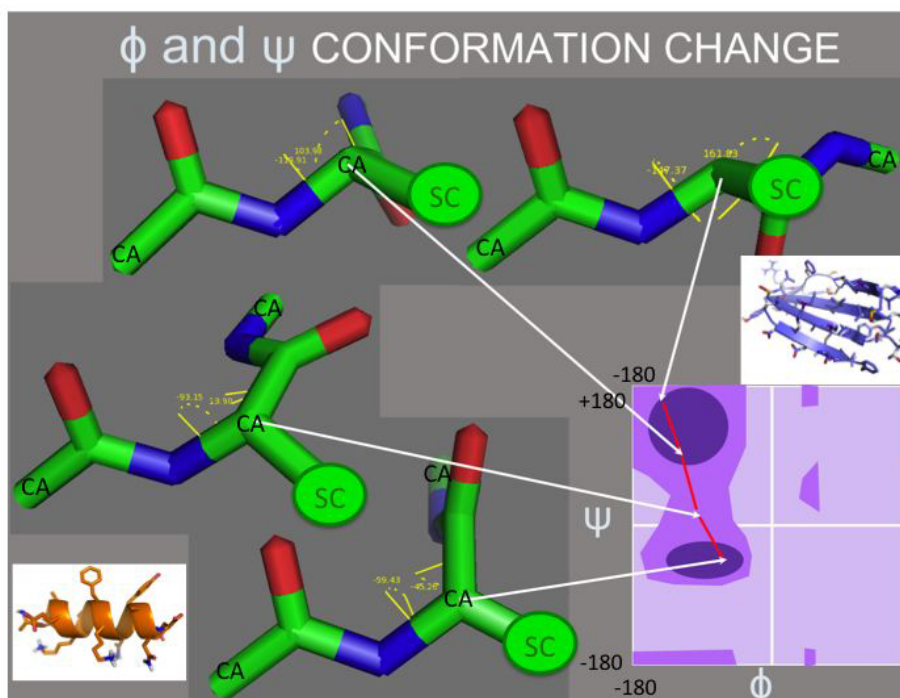


Figure 5.3 – Passage d’une conformation hélice  $\alpha$  vers feuillet  $\beta$  ou l’inverse. Le diagramme de Ramachandran illustre les angles  $\phi$  et  $\psi$  typiques de ces conformations. Le changement correspond à la réorientation de chaque plan peptidique par rapport au précédent. Le carbone  $\alpha$  sert de pivot entre trois ensembles : la chaîne principale qui le précède, la chaîne principale qui le suit, et la chaîne latérale (ici représentée par le sigle SC).

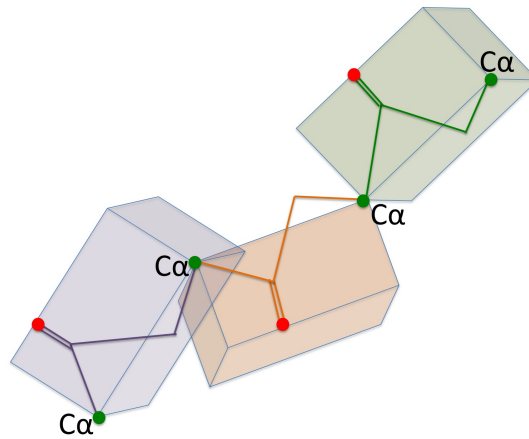


Figure 5.4 – Coordonnées relatives. Un plan peptidique peut être représenté selon une base orthogonale relative à la base orthogonale du plan peptidique précédent. Cela permet de fonctionner en coordonnées internes.

fonction de calcul des forces et est transparent pour *ART*.

La phase de relaxation de *ART* utilise directement les tableaux atomiques cartésiens des positions et des forces, tels qu'on les voit dans l'exemple de la figure 5.6. Par contre, les étapes d'activation et de transition utilisent les tableaux des positions et des forces selon l'architecture holographique.

### 5.3 Niveau holographique

L'architecture holographique de haut niveau est construite au moment de l'initialisation à partir des positions atomiques. La technique d'échantillonnage *ART nouveau* appelle à répétition la fonction de calcul des forces holographiques, et juge ainsi de la façon de faire évoluer la position relative des blocs holographiques. La fonction de calcul des forces utilise les positions pour déterminer les forces. Elle manipule une structure qui dissimule la complexité du système étudié. Pour une protéine transformée à l'échelle holographique, la première étape sera de reconstruire la représentation atomistique à partir des positions des blocs en coordonnées sphériques. Les forces atomiques sont alors obtenues par appel au potentiel atomistique. Nous avons implanté deux potentiels différents

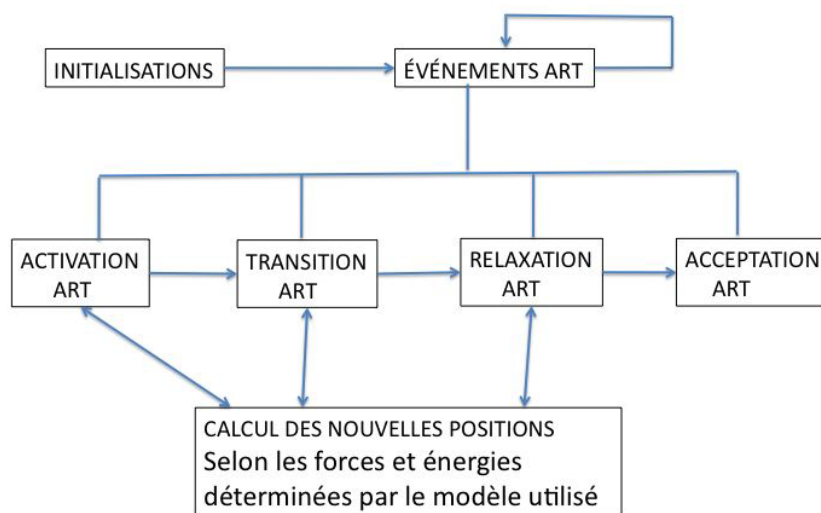


Figure 5.5 – Déroulement des simulations avec ART nouveau.

ATOM	1			2			3			4		
Coord	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
position	13.0	-4.1	-3.0	11.7	-3.7	-3.6	10.6	-4.0	-2.6	9.6	-4.7	-2.9
force	0.031	-0.022	0.009	0.04	-0.007	-0.05	0.008	0.001	0.07	-0.04	-0.02	0.01

Figure 5.6 – Tableaux des positions atomiques cartésiennes et des forces atomiques cartésiennes.

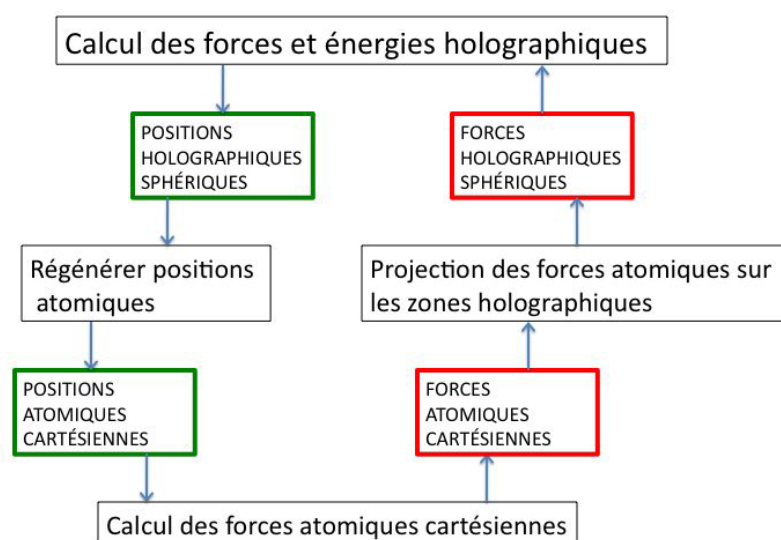


Figure 5.7 – Translation entre les niveaux de potentiel atomistique et *HMM*. La représentation holographique utilise le potentiel atomistique par pour calculer les forces holographiques.

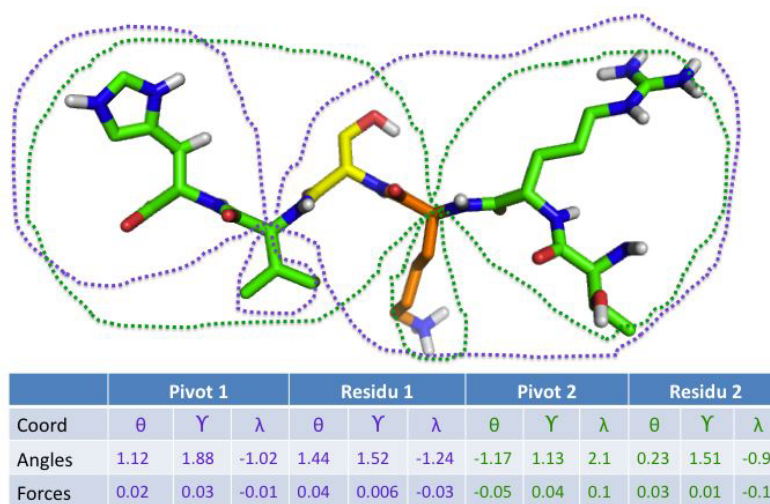


Figure 5.8 – Chevauchement des blocs holographiques. Représentation de deux découpages holographiques basé chacun sur un pivot déterminé. Tableau des positions des blocs holographiques associées à chaque pivot et tableau des forces sphériques correspondantes.



dans la méthode : *EOPEP* et *CHARMM19* conjoint au modèle de solvation *EEF1*. Nous décrivons ces potentiels au chapitre suivant. À partir des forces atomiques, la méthode *HMM* reconstitue les forces sphériques pour chaque découpage holographique du haut niveau, et les fournit ainsi à *ART* (voir figure 5.7). Les détails de ces transformations sont décrits au chapitre 7.

Il a est à noter qu'ainsi nous ne cherchons aucunement à réduire les calculs des forces atomiques. Non seulement nous effectuons les calculs atomiques au complet, mais nous ajoutons les calculs qui permettent de projeter les forces atomiques de différentes façons, une pour chaque définition de bloc, qui se chevauchent les unes les autres. La figure 5.8 donne un exemple pour seulement deux pivots  $C\alpha$ . À chaque pivot nous avons la première partie de la molécule, la chaîne latérale et la seconde partie de la molécule. Nous avons donc trois angles pour définir le comportement de la seconde partie par rapport à la première et trois angles pour définir le comportement de la chaîne latérale par rapport à la première partie. Le vecteur des forces indique la tendance au basculement et rotation de ces angles.

Nous allons descendre à une plus petite échelle avec le chapitre 6, qui décrit l'implantation des champs de force atomistiques dans la méthode.

## CHAPITRE 6

### TRAVAIL AVEC LES POTENTIELS ATOMIQUES

La *méthode multi-échelle holographique* peut être utilisée en conjonction avec des potentiels gros grains ou tout-atome. Elle a d'abord intégré le potentiel gros grain à solvant implicite *OPEP* [171]. Celui-ci définit la chaîne protéique principale en tout atome et les chaînes latérales par une sphère aux propriétés spécifiques à chaque acide aminé. L'un des travaux de cette thèse a consisté à développer une extension atomistique, *EOPEP* à partir du potentiel *OPEP*. Les atomes additionnels des chaînes latérales peuvent être ajoutés pour certaines de celles-ci, ou pour la totalité de celles-ci. Cela permet de définir une plus grande précision où cela est nécessaire. L'implantation dans la méthode du potentiel à solvant implicite *CHARMM19* [192] a été réalisée, ainsi que celle du modèle de solvation *EFF1* [145] qui s'est avéré un complément important de *CHARMM19*. *CHARMM19* peut être utilisé indépendamment ou en conjonction avec *EFF1*.

#### 6.1 Intégration des potentiels à la méthode HMM

L'utilisation de potentiels à solvant implicite dans *HMM* nécessite une implantation particulière de ceux-ci. Il n'est pas suffisant pour *HMM* de connaître la valeur et la direction des forces pour chaque atome. Les différents potentiels sont donc intégrés à la méthode selon une infrastructure générale utile à celle-ci. La série verticale de rectangles gris dans la partie gauche de la figure 6.1 représente ce que l'on demande typiquement à un potentiel atomique. Les rectangles de couleurs à gauche représentent des calculs additionnels utiles à nos représentations multi-échelles.

Nous avons besoin de faire la distinction entre les forces internes aux blocs et les forces relationnelles entre ceux-ci. De plus, ceux-ci se superposent. La figure 6.2 montre la ségrégation des forces internes par rapport aux forces relationnelles pour un pivot situé à l'acide aminé numéro K. Pour *HMM*, la somme des forces sur un atome doit être

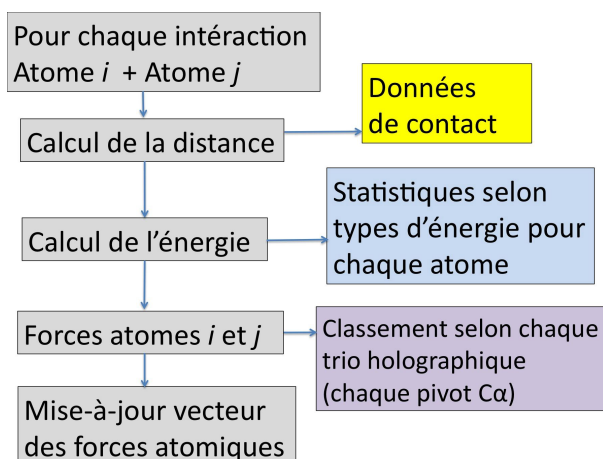


Figure 6.1 – Intégration des potentiels atomistiques dans *HMM*. À gauche en gris : calculs types pour la génération du vecteur des forces par atome. À droite, calculs additionnels : rectangle lilas : répartition des forces selon les trios holographiques associés à chaque pivot  $C\alpha$  ; rectangle bleu : statistiques d'énergie pour chaque atome, pour les besoins d'analyse des enchaînements d'événements et de la coopérativité ; rectangle jaune : mise-à-jour d'informations de contact (facultatif, utilisé pour les assemblages de molécules).

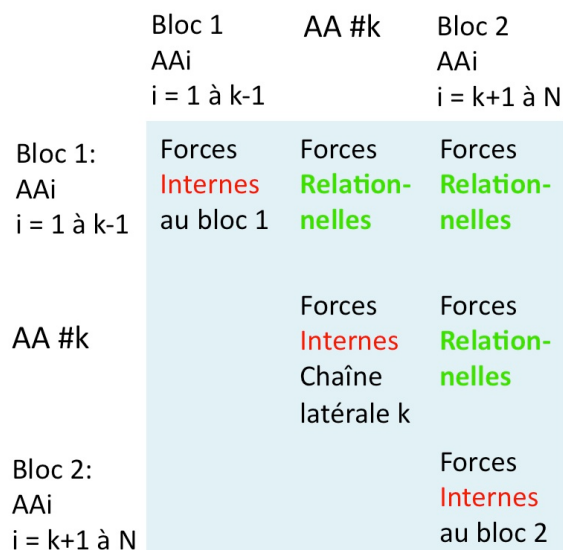


Figure 6.2 – Un trio holographique. Discrimination entre les forces internes et les forces relationnelles pour les besoins des découpages holographiques. Ici un seul découpage est représenté, basé sur un pivot  $C\alpha$  de l'acide aminé numéro K.

discriminée selon sa provenance, et cela pour chacun des découpages holographiques de haut niveau. Ainsi pour un découpage holographique de la molécule en trois parties au pivot  $C\alpha$  numéro  $K$  (partie précédente, partie suivante, chaîne latérale), nous voulons omettre pour chaque atome  $i$  les forces provenant des atomes du bloc où il est situé. Nous voulons seulement obtenir les forces qui sont en provenance des deux autres parties. Et ce discernement des forces doit être fait pour chacun des découpages holographiques, c'est-à-dire pour chaque pivot  $C\alpha$ .

Afin de mieux analyser les phénomènes de cause à effet et de coopérativité, des fichiers spéciaux peuvent être générés avec la compilation des énergies par résidu, et par atome. Ceci a permis de construire les graphiques d'évolution des énergies par résidu que l'on peut voir dans l'article présenté au chapitre 8. On y représente la variation de l'énergie concernant des résidus stratégiques. Nous avons utilisé l'énergie globale, mais nous pouvons aussi faire le discernement selon le type d'énergie et par atome. Au moment des calculs inter-atomiques, nous pouvons aussi compiler des informations concernant la surface utile aux méthodes de plus haute échelle. Cette compilation est facultative et peut être activée par un paramètre de configuration.

Nous allons consacrer la suite du chapitre aux explications concernant l'implantation des potentiels tout atome et le développement de *EOPEP*.

## 6.2 Les potentiels à solvant implicite

Pour étudier les capacités de la méthode *HMM* à générer des changements de longue portée, nous l'avons utilisée avec des potentiels à solvant implicite décrivant l'occupation spatiale de manière assez précise. Le choix de deux potentiels tout atome basés sur des stratégies différentes permet de mieux discerner comment les performances et les limitations de la méthode sont reliées à la méthode *HMM* elle-même ou au potentiel atomistique utilisé. Le modèle *EEF1* [145], associé aux paramètres du champs de force *CHARMM19* [192], établit ses calculs sur la capacité d'inclure ou d'exclure l'eau dans le volume séparant chaque paire d'atomes donnée, ainsi que selon le type de ces atomes. Pour sa part le potentiel gros grain *OPEP* [171] intègre l'effet de l'eau dans les attrac-

tions et répulsions moléculaires, grâce à des études statistiques. À partir d'*OPEP*, nous avons développé le potentiel tout atome *EOPEP* de manière à ce qu'il utilise le même nombre d'atomes que *CHARMM19/EEF1*, soit tous les atomes lourds (carbones, oxygènes, azotes, soufres) et tous les hydrogènes polaires (associés aux azotes et à certains oxygènes), tandis que les hydrogènes associés aux carbones sont implicites.

Ce chapitre décrit l'élaboration des chaînes latérales ayant permis de créer *EOPEP* à partir d'*OPEP*. Mais tout d'abord débutons par un résumé sur le potentiel gros grain *OPEP*.

### 6.3 OPEP

Développé par Philippe Derreumaux [171], le potentiel *OPEP* définit tous les atomes de la chaîne principale et regroupe ceux de chaque chaîne latérale en une seule bille aux propriétés caractéristiques. L'énergie totale est la somme des énergie des interactions covalentes (locales), de celles des interactions non liées et de celles des interactions des ponts hydrogènes (équation 6.1) .

$$E = +E_{local} + E_{nonbonded} + E_{H-bridge} \quad (6.1)$$

Les énergies locales comprennent les énergies de liaison, celles des angles de valence et diédraux, des diédraux impropres, ainsi qu'un terme favorisant les conformations fréquentes  $\phi$  et  $\psi$  des diagrammes de Ramachandran :

$$E_{local} = w_b \sum_{liens} k_b (r - r_{eq})^2 + w_a \sum_{angles} k_a (\alpha - \alpha_{eq})^2 + w_{\Omega} \sum_{imp-tor} k_{\Omega} (\Omega - \Omega_{eq})^2 + E_{\phi,\psi} \quad (6.2)$$

où

$$E_{\phi,\psi} = w_{\phi,\psi} \left( \sum_{\phi} K_{\phi\psi} (\phi - \phi_0)^2 + \sum_{\psi} K_{\phi\psi} (\psi - \psi_0)^2 \right) \quad (6.3)$$

À part le fait qu'il s'agit d'un potentiel à solvant implicite, deux caractéristiques sont

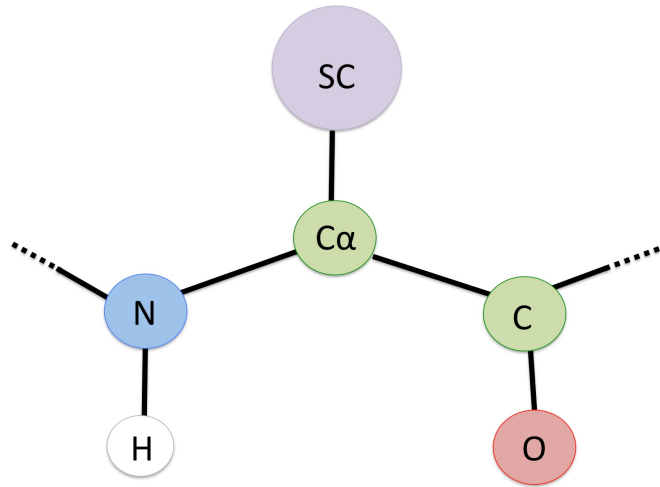


Figure 6.3 – Modèle *OPEP*. Modélisation des acides aminés en six atomes dans *OPEP* [51]. SC représente la chaîne latérale.

principalement à la base de l'originalité et des capacités d'*OPEP*. La première concerne la formation des structures secondaires, la seconde la formation des structures tertiaires. *OPEP* calcule l'énergie des ponts hydrogènes ( $E_{HB1}$ ), ainsi que la coopérativité entre les ponts hydrogènes ( $E_{HB2}$ ) (équations 6.4 à 6.10).

$$E_{HB} = E_{HB1-4} + E_{HB1>4} + E_{HB2\alpha} + E_{HB2\beta} \quad (6.4)$$

Pour calculer les ponts hydrogènes entre les plans peptidiques de la chaîne principale, *OPEP* utilise une fonction de type Lennard-Jones 10-12 [133] (équations 6.5 à 6.7).

$$E_{HB1-4} = w_{hb1-4} \sum_{bonds} \epsilon_{hb1-4} \mu(r_{ij}) v(\alpha_{ij}) \quad (6.5)$$

$$E_{HB1>4} = w_{hb1>4} \sum_{bonds} \epsilon_{hb1>4} \mu(r_{ij}) v(\alpha_{ij}) \quad (6.6)$$

où

$$\mu(r_{ij}) = 5\left(\frac{\sigma}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma}{r_{ij}}\right)^{10} \quad (6.7)$$

$$v(\alpha_{ij}) = \cos^2 \alpha_{ij} \text{ si } \alpha_{ij} > 90^0, 0 \text{ sinon} \quad (6.8)$$

L'équation 6.7 concerne la distance entre l'hydrogène et l'atome donneur. Alors que le nombre 12 réfère à l'exposant du terme répulsif, le nombre 10 réfère à l'exposant du terme attractif. L'équation 6.8 tient compte de l'angle entre les atomes N-H et l'oxygène. Le calcul de coopérativité des ponts hydrogènes concerne la formation des hélices  $\alpha$  et des feuillets  $\beta$  (équations 6.9 et 6.10) :

$$E_{HB2\alpha} = \sum_{bonds} \varepsilon_{\alpha}^{coop} \exp(-\frac{1}{2}(r_{ij} - \sigma)^2) \exp((-\frac{1}{2}(r_{kl} - \sigma)^2)\Delta(ijkl)) \quad (6.9)$$

$$E_{HB2\beta} = \sum_{bonds} \varepsilon_{\beta}^{coop} \exp(-\frac{1}{2}(r_{ij} - \sigma)^2) \exp(-\frac{1}{2}(r_{kl} - \sigma)^2)\Delta'(ijkl) \quad (6.10)$$

Cette stratégie est basée sur la constatation que la prolongation d'une structure secondaire (en hélice  $\alpha$  ou en feuillet  $\beta$ ) est plus facile que son initiation. Lors de nos études de la calmoduline, des simulations ont été générées avec et sans ces calculs de coopérativité. La présence du calcul de coopérativité s'est avérée indispensable pour produire la consolidation du feuillet  $\beta$  nécessaire au retour vers la forme apo.

La deuxième caractéristique importante d'*OPEP* est le calcul des attractions et répulsions entre les chaînes latérales. Alors que tous les atomes de la chaîne principale sont représentés, chaque chaîne latérale est représentée par une sphère d'une dimension qui la caractérise et située à une distance spécifique du carbone  $\alpha$  qui en est la racine (voir figure 6.3). Pour chaque couple différent impliquant deux chaînes latérales, une attraction/répulsion de type Lennard-Jones 6-12 [131] est définie (Équation 6.11). Alors que le terme en exposant 12 est attractif, le terme en exposant 6 est répulsif. Le sigle *H* exprime un calibrage issu des statistiques sur de nombreuses protéines. Des équations de ce type sont aussi utilisées pour les interactions non liées impliquant les autres type d'atomes.

$$E_{wdw} = \varepsilon_{ij} \left( \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^0}{r_{ij}} \right)^6 \right) H(\varepsilon_{ij}) - \left( \frac{r_{ij}^0}{r_{ij}} \right)^6 H(-\varepsilon_{ij}) \quad (6.11)$$

## 6.4 EOPEP

Pour le travail d'étude de la transformation du domaine NT de la calmoduline (Voir chapitre 8), des tests préliminaires avaient été effectués en utilisant le potentiel *OPEP* [171] gros grains. Les premières simulations concernaient la stabilité de la conformation apo. Elles ont fait ressortir que cette protéine requérait une précision accrue pour deux aspects de la représentation atomique. Le plus important concernait la définition des ponts hydrogènes entre les chaînes latérales munies d'oxygènes et les azotes de la chaîne principale. König et Boresch ont démontré que pour les petites chaînes latérales polaires, la conformation locale de la chaîne principale était davantage déterminante que les statistiques d'affinité entre les chaînes latérales [132]. Ceci s'est effectivement avéré un facteur très important entraînant le choix de modéliser dans *EOPEP* les ponts hydrogènes entre les chaînes latérales et la chaîne principale. Les interactions polaires n'entrent souvent en jeu qu'après la formation de la structure tertiaire. Alors elles peuvent avoir des rôles importants : établir des zones de contact avec les ligands et avoir alors un effet important sur la structure générale. Certains contacts polaires ont un rôle important en bordure des structures secondaires (*capping*) ainsi qu'à l'intérieur des boucles flexibles et stabilisent le jeu de celles-ci. Ce qui pourrait paraître un détail fait une grande différence aussitôt que l'on génère des mouvements d'envergure. Dans le cas de la calmoduline, une définition plus complète de plusieurs sérines, thréonines, aspartates et glutamates fermant les bouts des hélices s'est avérée indispensable afin de stabiliser leur mouvement par rapport à la structure en feuillet  $\beta$  servant d'échafaudage à la base du jeu entre les quatre hélices du domaine NT. Dans le cas de la calmoduline, le résidu ASP20 forme dans la conformation apo plusieurs ponts hydrogènes avec la chaîne principale de la première boucle et agit ainsi comme un épaulement. Ces interactions pont-H, et plusieurs autres, sont indispensables pour stabiliser le mouvement relatif des hélices (Fig. 6.4).

Le deuxième développement de précision qui s'est avéré nécessaire concernait la représentation spatiale plus complète des chaînes latérales. Car au sein du noyau hydrophobe du domaine NT de la Calmoduline, non moins de cinq phénylalanines, cinq méthionines et nombre d'autres chaînes latérales hydrophobes, adaptent leur forme entre



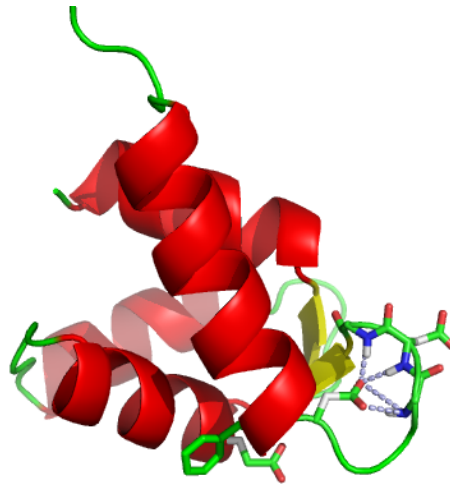


Figure 6.4 – ASP20 dans la calmoduline apo. Dans la conformation apo, le premier résidu de la boucle 1 de la calmoduline, ASP20 forme plusieurs ponts hydrogènes avec les azotes de la chaîne principale.

elles de manière précise.

Pour *EOPEP*, nous avons déployé les chaînes latérales en définissant les atomes de carbone, d'oxygène et d'azote, ainsi que l'hydrogène de nature polaire. Ceux-ci ont aussi été ajoutés aux azotes et oxygènes concernés. Les carbones ont été définis avec leur hydrogènes implicites. Pour les dimensions, nous avons utilisé les paramètres de type Van der Waal décrites pour le potentiel CHARMM19 [192]. Lors du déploiement des chaînes latérales en tout atome, nous perdons l'information statistique d'*OPEP*, qui décrit les affinités entre les différentes chaînes. Par contre, les énergies d'attraction entre les types d'atomes sont encore décrites dans *EOPEP* avec la même stratégie de type Lennard-Jones 6-12 de *OPEP* (équation 6.11), à l'exception des ponts hydrogènes. Ceci permet éventuellement de modéliser une protéine avec deux niveaux de définition *OPEP* et *EOPEP*, selon le besoin de précision. Pour les simulations décrites dans cette thèse nous avons cependant utilisé partout la définition tout atome.

Bien qu'employant les mêmes formules Lennard-Jones 10-12 (équations 6.7-6.8),

la force des ponts hydrogènes impliquant des atomes de chaînes latérales a été réduite par rapport à celles des ponts hydrogènes spécifiques à la chaîne principale. Lorsque le receveur est un oxygène d'un groupe carboxyle (aspartate, glutamate), nous conservons la pleine force si l'interaction est par rapport aux azotes de la chaîne principale. Une interaction envers un azote ou un oxygène donneur d'hydrogène d'une chaîne latérale entraîne cependant une diminution d'un facteur vingt. Une diminution additionnelle d'un facteur dix est appliquée si l'interaction implique un receveur de chaîne latérale non carboxyle. Ceci reproduit empiriquement, mais assez bien la tendance plus faible des chaînes latérales à former des ponts hydrogènes, étant donnée leur immersion dans le solvant aqueux. Leur formation se produit seulement lorsque les forces hydrophobes ont forcé les rapprochements des chaînes impliquées. La calmoduline et la protéine A ont servi pour les tests d'étalonnage. Cet étalonnage empirique visait aussi l'ajustement des forces hydrophobes, principalement entre les carbones des chaînes latérales, ainsi que celles des autres interactions entre oxygènes, carbones, azotes, soufres. Il s'est avéré préférable que les interactions entre les différents types de carbone (ceux non fortement polarisés par un lien covalent avec un atome chargé) adoptent une courbe Lennard-Jones 6-12 uniforme entre eux. Une contribution de  $0.05(\text{Kcal/mol})^{-2}$  par atome de carbone (non lié à un atome chargé) a été utilisée pour la profondeur du puits énergétique. Ceci favorise l'ajustement des formes entre elles, alors que des variations basées sur des statistiques entraînait moins bonne adaptation entre les formes. Car selon le comportement hydrophobe, la molécule d'eau doit être présente ou absente et non pas partiellement présente. Les statistiques selon les types de carbone sont le reflet du comportement moyen et conviennent aux gros grains, qui eux ont perdu l'information sur la forme. Pour les atomes hydrophobes, utiliser les statistiques pour calibrer les tendances revient un peu à se servir de la conséquence de l'adaptation des formes comme si elle était la cause, et peut induire un certain biais.

Le comportement des atomes de soufre de la méthionine a été directement calqué sur celui de la bille méthionine de *OPEP*. Les interactions des atomes polaires ne concernant pas la formation des ponts hydrogènes ont été modélisées sur la base des celles de certaines chaînes gros grain *OPEP*. Le comportement des atomes d'oxygène et d'azote est

calqué sur celui de la chaîne gros grain aspartate de *OPEP*, mais diminué d'un facteur de deux. Le comportement des hydrogènes est diminué d'un facteur de vingt par rapport à celui de l'arginine. Ces ajustements ont été établis suite à des tests de longue durée évaluant la stabilité d'un ensemble de protéines, dont la protéine A native et le modèle apo de domaine N-terminal de la calmoduline. La stabilité devait résister sur plusieurs jours aux les activations de bonne envergure générées par *ART holographique*.

## 6.5 Travaux d'implantation de CHARMM19/EFF1

Le potentiel *CHARMM19* peut être sélectionné dans le fichier d'initialisation, avec ou sans le module de solvation implicite *EFF1*. Les différents termes de calcul des forces propres à ces modèles sont invoqués dans la même infrastructure générale qui sert pour *OPEP* ou *EOPEP*. L'implantation de *CHARMM19* peut différer un peu du modèle officiel au niveau du calcul des forces de torsions. La méthode en fait le parcours exhaustif, mais selon le système utilisé pour *OPEP* et *EOPEP*. Les calculs de torsions sont calibrés pour s'ajuster selon les potentiels. Les articles présentés aux chapitres 7 et 8 montrent des simulations effectuées avec *CHARMM19* en association avec *EFF1*. L'équation 6.12 représente l'énergie de solvation selon le modèle *EFF1*.

$$\Delta G^{sly} = \sum_i \Delta G_i^{ref} - \sum_{j>i} \left( \frac{2\Delta G_i^{free}}{4\pi\sqrt{\pi}\lambda_i r_{ij}^2} \exp(-x_{ij}^2) V_j + \frac{2\Delta G_j^{free}}{4\pi\sqrt{\pi}\lambda_j r_{ij}^2} \exp(-x_{ji}^2) V_i \right) \quad (6.12)$$

où pour chaque type d'atome,  $V$  est son volume type de solvation, et  $\Delta G^{free}$  son énergie libre de solvation.  $\lambda$  est un paramètre qui vaut 3.5 Å, sauf pour les groupes ioniques neutralisés où il vaut 6.0 Å. Ces groupes correspondent aux chaînes latérales chargées et aux terminaisons NT et CT, pour lesquels la somme des charges s'annule dans le potentiel.

Des simulations ont aussi été réalisées avec *CHARMM19* sans *EFF1*. De manière générale on constate alors une plus grande difficulté d'effectuer les mouvements requis, la cause étant une attraction excessive entre les chaînes latérales polaires ou chargées.

Par contre la définition des structures secondaires est parfois mieux assurée, à cause d'une attraction plus efficace au sein des ponts hydrogène de la chaîne principale.

## 6.6 Bilan

Bien que le potentiel *EOPEP* en soit à ses premières utilisations, il a réussi à guider des simulations vers des conformations proches de cibles éloignées de la conformation de départ (voir chapitres 7 et 8). Il en est de même pour *CHARMM19*, que nous avons utilisé seulement selon ses caractéristiques de base, à l'exception de l'appui important apporté par le modèle de solvation *EEFI*. Nous pouvons conclure qu'il sera de plus en plus possible de simuler des trajectoires complexes en utilisant des potentiels *ab initio*. Ces potentiels ont été implantés dans le but de vérifier l'efficacité de *HMM*. Mais la méthode *HMM* est en mesure d'accepter des potentiels de différents niveaux de complexité. L'ajout de termes additionnels est aisément paramétrisable dans le fichier d'initialisation. Les simulations avec *HMM* constituent des tests poussés pour l'évaluation de la capacité des potentiels ou d'un aspect spécifique de représenter les états transitoires, et non pas seulement les états stables.

## CHAPITRE 7

### LA MÉTHODE HOLOGRAPHIQUE

Le développement de la méthode multi-échelle holographique a débuté le jour où M. Mousseau m'a demandé de modifier le modèle de la protéine A pour déplacer une hélice  $\alpha$  hors de sa position normale. Il fallait bien sûr que je crée une méthode qui la guiderait vers sa place normale. Les défis se sont ensuite intensifiés, ce qui m'a conduit à concevoir un moyen de faire basculer plusieurs larges régions en même temps. La représentation de ces régions en coordonnées internes s'est avérée fructueuse, à condition que les forces d'ensemble soient ajustées au fur et à mesure selon les indications de champs de force atomistiques.

Ma contribution à cet article en tant qu'étudiante a consisté à concevoir et développer la méthode holographique multi-échelle. J'ai aussi développé *EOPEP*, une version étendue du potentiel gros grain *OPEP*, en déployant les chaînes latérales en tout atome et en calibrant leurs forces atomistiques. J'ai réalisé l'implantation du champs de force *CHARMM19* et du modèle de solvation *EEFI* dans la méthode. J'ai effectué des simulations à partir de modèles déployés de domaines des protéines A et G, que j'ai analysées, et j'ai écrit la première version de l'article.

L'article a été publié au "Journal of physics : conference series" en février 2012.

## **Holographic multiscale method used with non-biased atomistic forcefields for simulation of large transformations in protein**

L Dupuis<sup>1</sup> and N Mousseau<sup>2</sup>

<sup>1</sup> Département de Biochimie et Centre Robert-Cedergren, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7

<sup>2</sup> Département de Physique, Regroupement Québécois sur les Matériaux de Pointe (RQMP) et Centre Robert-Cedergren, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7

### **ABSTRACT**

**We present a multiscale approach for simulating protein flexibility. The originality of our method is its ability to perform dynamic multiscaling based on continuous revaluation of overlapping areas. The holographic multiscale method overcomes the limitations of motions determined by predefined and fixed high-level descriptions and allows the reproduction of residue-specific impact on large scale motion. The method is tested with two different non-biased all atom implicit solvent forcefields. These show stretched proteins A and G, with maintained secondary structure, folding back near native states in a small number of transition events, demonstrating the advantages of this multiscale approach.**

### **7.1 Introduction**

The success of macromolecular computer simulations rests on describing appropriately two crucial aspects of these complex systems : the interactions between particles, provided by forcefields of various descriptive levels, and the physical geometrical realization of the structure, which can be extracted from experimental data, such as X-ray

diffraction or NMR, or constructed by methods for sampling the system's configurational space. In spite of ever growing computational resources, it is still necessary to find the simplest level of description and sampling that will provide the best balance for answering the problem we are interested in.

This right balance is particularly difficult to find when looking at large-scale protein motions, that are often controlled by atomic-scale details but occur on time scale beyond current's general sampling methods. To overcome these opposite requirements, it is necessary to consider multiscaling approaches that can focus on crucial aspects of both of interactions and motions.

Traditional molecular dynamics, for example, describes exactly, within the forcefield limits, motion at all scales, from thermal vibrations to docking and folding. However, because no bias is provided to favor large motion, the large majority of the computing efforts go to describing thermal oscillations, providing little new understanding. Monte Carlo approaches benefit from considerably more freedom in selecting the appropriate level of motion [73, 196, 281]. However, since these do not generally take forces into account, the size of moves is typically limited by collisions and steric constraints. These are avoided by methods such as Normal Mode Analysis (NMA) [30, 115, 149, 260], a method that identifies directly the softest directions of change by first mapping the protein as an elastic network to compute the lowest-frequency modes of the dynamical matrix, a remarkably good approximation. As low frequencies only depend on the general shape of the molecule [159], NMA can be used with simple coarse-grained level representations [3, 253, 266, 300].

There exists many problems and phenomena, however, where large scale motion of a molecule is affected by mutation of single residues [41, 43, 177, 263] and that cannot be described by coarse-grained harmonic state analysis [162]. For such movements, it is essential to adopt a multilevel representation that adapts to the role of the various molecular structures and incorporates non-harmonic motions. The latter can be achieved with activated methods that propose a non-biased approach for the study of complex transformations affecting the general shape of proteins. These approaches perform continuous evaluations of force curvatures in order to discover energetically feasible transition pas-

sages, providing a natural way to leave local minima and identify physically-possible pathways [17, 31, 167, 246]. Focusing on the atomic scale, however, these methods rapidly become dominated by small, local rearrangements that contribute very little to overall configurational changes.

Here, we propose a multiscale algorithm that allows the activated method to focus on large scale motion while retaining the capacity to adapt this representation to atomistic constraints and motion. The holographic multiscale method (HMM) introduces a dynamical definition of the various representation scale that increases considerably the versatility of this approach compared with standard multiscale methods [3, 113, 168, 253, 281]. More precisely, the HMM combines the activation-relaxation technique (ART nouveau), an activated method that has been used extensively for protein folding and aggregation, both in cartesian [182, 228, 246, 277] and in internal coordinates [298], with an adaptive structure-dependent multiscale representation that focuses on large cooperative ensemble motions while allowing, at every step, atomistic relaxation that fully incorporates the role of specific residues and interactions on large scale dynamics.

The structure of this paper is as follows. We begin with a brief overview of multilevel characteristic of proteins, on which the holographic view is based. We then present the details of the holographic multiscale method, including the multiscale projection into reduced internal coordinates. We briefly discuss our in-house implicit-solvent reduced-representation extended OPEP forcefield and then present a few test applications that demonstrate the strength of the algorithm.

## **7.2 Methodology**

### **7.2.1 Basic protein motion**

A protein is a chain of a few tens to many thousands of amino acids characterized by a common part composed of a short sequence of a nitrogen and a carbon atom positioned on each side of a central carbon ( $C\alpha$ ) and by a lateral chain, or residue, of varying length and composition, rooted at the  $C\alpha$ . Amino acids are linked by their common part through a series of covalent bonds called peptidic links, which form the



protein's main-chain or backbone. The presence of covalent bonds restricts effectively the possible motion of each amino acid with respect to its neighbors to the torsion angles situated on both sides of the  $C\alpha$ . With these degrees of freedom, the main-chain can adopt a number of conformations that determine the protein secondary structures —  $\alpha$ -helices,  $\beta$ -sheets, loops, random coils, etc. — which are further packed into a three-dimensional organization called tertiary structure. This well-defined organization is guided by delicate balance of backbone–backbone, side-chain–side-chain, backbone–side-chain and protein-solvent interactions, controlled by hydrogen bonds, hydrophobic, polar and charge groups.

Large structural changes in the protein necessarily involve large modifications of the main-chain torsion angles,  $\phi$  and  $\psi$ , located, respectively, around the N- $C\alpha$  bond, and the  $C\alpha$ -C bond. For example, significant changes in these angles are required to go from  $\phi \approx -55$  and  $\psi \approx -45$  degrees for  $\alpha$ -helices to  $\phi \approx -130$  to  $-160$  and  $\psi \approx 130$  to  $160$  degrees for  $\beta$ -sheets. Even though these angles largely characterize the protein spatial organization, other angles and bond-lengths will adapt slightly to every configuration in order to minimize steric effects and optimize stability. The holographic multiscale method combines these two levels of relaxation. Following other large scale motion techniques [40], the HMM focuses on these angles to generate the activation. Its originality comes from the fact that it allows all degrees of freedom to adapt during the relaxation phase.

### 7.2.2 Overview of the holographic multiscale method

To focus on large scale cooperative motion, the holographic multiscale method proposes a multilevel approach that starts with the protein in a local energy minimum and goes as follows :

1. The protein conformation is described on the basis of a list of flexible  $C\alpha$  pivots. It is at this moment that a multiscale description is introduced, by the identification of flexible (or *active*) regions, typically associated with loops and random-coil conformations, and rigid regions, those with stable secondary structures. At each

active  $C\alpha$  pivot the protein is split into three blocks : the chain fragment preceding the pivot, that following it, and the lateral chain. Each pivot possesses six degrees of freedom associated with the rotation of the last two sections with respect to the first one and to each other. After each displacement step, the all-atom Cartesian space representation of the molecule is regenerated in order to compute the atomistic forces which are projected back onto each overlapping block definition delimited by the various freed pivots (Fig. 7.3). This generates a smoothed configurational subspace, eliminating small, but non-diffusive, barriers.

2. A first activated event is generated, using ART nouveau, on this reduced configurational subspace, leading to a first-order saddle point associated with large-scale cooperative motion.
3. From this saddle-point associated with a transition state, the protein is brought into a local-energy minimum using damped molecular dynamics applied to the all-atom Cartesian representation with all degrees of freedom allowed to move.

These three steps represent an ART event, going from one local energy minimum to another, passing through an activated state and following a physically-possible pathway, that allows an efficient sampling of the configurational energy landscape. The next sections describe in more details each of these steps.

### 7.2.3 Construction of elastic blocks and projected flexible regions

As mentioned in the previous section, the crucial step in HMM is the construction of the multiscale representation. For this, we first construct a spherical representation of positions and forces projected onto the peptidic plane angular degrees of freedom that determine the rotation of the  $\phi$  and  $\psi$  angles. For each of these planes, we compute an orthogonal basis, using the  $C\alpha$  to  $C\alpha$  axis direction as first axis,

$$\vec{a}_1 = \frac{\vec{r}_{C\alpha 2} - \vec{r}_{C\alpha 1}}{\|\vec{r}_{C\alpha 2} - \vec{r}_{C\alpha 1}\|}, \quad (7.1)$$

and the perpendicular direction toward the main-chain C-bound oxygen as second axis,

$$\vec{a}_2 = \frac{\vec{r}_{oxy-C\alpha_1} - (\vec{r}_{oxy-C\alpha_1} \times \vec{a}_1) \times \vec{a}_1}{\|\vec{r}_{oxy-C\alpha_1} - (\vec{r}_{oxy-C\alpha_1} \times \vec{a}_1) \times \vec{a}_1\|}. \quad (7.2)$$

The third axis is deduced from a right hand rule

$$\vec{a}_3 = \frac{\vec{a}_1 \times \vec{a}_2}{\|\vec{a}_1 \times \vec{a}_2\|}. \quad (7.3)$$

This forms the absolute base A for each peptidic plane  $\delta$ . We also need to create its relative base  $B_\delta$ , which is orthogonal :

$$B_\delta = (A_{\delta-1} \cdot A_\delta^T)^T \text{ giving } \vec{b}_1, \vec{b}_2, \vec{b}_3 \quad (7.4)$$

This relative representation allows us to evaluate the angular position (swiveling) of the following area relatively to the preceding, using two spherical angles : the longitude  $\gamma$  and the latitude  $\lambda$  of the  $C\alpha-C\alpha$  axis ( $b_1$ ), relatively to the preceding plane, defined as :

$$\lambda = -\arccos \frac{b_{1,2}}{\|\vec{b}_1\|} \quad (7.5)$$

and

$$\gamma = \frac{b_{1,3}}{|b_{1,3}|} \arccos \frac{b_{1,1}}{\|\vec{b}_1\| \times |\sin \lambda|}. \quad (7.6)$$

Because the blocks are three-dimensional, a third angle represents the rotation  $\theta$  of the plane around its own  $C\alpha-C\alpha$  axis ( $b_1$ ). As this axis also swivels in space, the evaluation of its rotation requires a comparison with defined referential which we take to be the  $x$  axis of the preceding base. We first determine the axis of rotation  $c$  between axis  $b_1$  and the  $x$  axis :

$$\vec{c} = \vec{x} \otimes \vec{b}_1 = (0, b_{1,3}, -b_{1,2}). \quad (7.7)$$

We then determine the rotation matrix  $C$  around this  $c$  axis for the angular course needed between axis  $b_1$  and  $x$  axis. We obtain this angle from the arc-cosinus of the scalar product between axis  $b_1$  and  $x$  axis. This rotation matrix  $C$  is then used to apply a

theoretical swivelling of the current matrix  $B$  toward the  $x$  axis of the referential previous block, giving the matrix  $S$  :

$$S = (C \cdot B^T)^T. \quad (7.8)$$

The first axis of matrix  $S$  coincide with the  $x$  axis. The second axis of matrix  $S$  is used to determine the current rotation angle of the block, using its angular position relative to the  $y$  axis :

$$\theta = \frac{s_{2,3} - s_{2,1}}{|s_{2,3} - s_{2,1}|} \arccos \frac{s_{2,2}}{\|\vec{s}_2\|} \quad (7.9)$$

The modification of these three angles corresponds to a holographic swiveling/rotation around the peptidic junction  $\delta-1, \delta$ . With this representation, the entire protein following a  $C\alpha$  pivot is allowed to swivel and rotate relatively to the preceding part. This representation uses 3 angles instead of the 2 torsion  $\phi$  and  $\psi$  angles. This means that some variation in the valence angles is allowed here, in addition of the easier dihedral rotations. The valence angles around the  $C\alpha$  may be stressed at the beginning of the activation, but the convergence method will soon relaxed them and conducted the motions to rotate according to  $\phi$  and  $\psi$  dihedral angles. The position of the lateral chain is also expressed in spherical representation relatively to the preceding peptidic plane, including a rotation around its own  $C\alpha-C\beta$  axis.

Atomic positions must be retrieved at each step of the simulation in order to compute the forces. We first get back axis  $b_1$  of the relative base  $B$  :

$$\vec{b}_1 = (\sin \lambda \times \cos \gamma, \cos \lambda, \sin \lambda \times \sin \gamma). \quad (7.10)$$

The  $b_2$  axis is obtained using the new  $\theta$  angle, reversing the steps of equations 7.7,7.8,7.9. From matrix  $B$  we get back the absolute base  $A$  :

$$A_\delta = (A_{\delta-1}^T, B_\delta^T)^T. \quad (7.11)$$

Then the new positions are computed from the initial positions, using the initial absolute base  $A$  and the current absolute base  $A'$  as transformation matrices.

$$\vec{r}_i^{\lambda'} = (A'^T \cdot A)^T \cdot \vec{r}_i^{\lambda}, \quad (7.12)$$

where  $i$  is the atomic number and  $\vec{r}_i$  is a three-dimensional cartesian position vector.

The rotation forces around the different axes are in correspondence with angles around these axes. The force contribution of all the atoms of the molecule is evaluated for each pair of holographic areas. The lateral chain positioned at the concerned  $C\alpha$  pivot is also designed as a block that can rotate relatively to them. The contribution of each atom for the block rotation around a particular axis is inversely proportional to its radius distance to the axis of rotation, and the part of the force that is perpendicular to both the axis and the radius between the axis and the atom.

$$F_{a_l} = \frac{\sum(\vec{f} \times (\vec{r} - (\vec{r} \cdot a_l) \cdot a_l))}{\sum \|\vec{r}\|} \quad (7.13)$$

where  $l = 1, 2, 3$ .

Using the current array of relative angular positions of each overlapping block pair, and the newly computed corresponding arrays of spherical block forces, ART nouveau gives new arrays of relative angular positions. At each step this new array is used to compute back cartesian atomistic positions and then the cartesian atomistic forces from forcefield. Those are used by HMM to update the spherical holographic block forces used by ART. Fig. 7.3 summarizes this circular process.

## 7.2.4 Sampling method

We use the activation-relaxation technique, ART nouveau [167, 182], to generate energetically favorable transition events from one conformation to the other. Since the method has been reviewed recently [163, 170], we present here only the general scheme as applied into HMM. Starting from a configuration in a local minimum, a random direction of deformation is selected among the available degrees of freedom (pivots). The configuration is slowly deformed. At each step, the lowest eigenvalues of the Hessian matrix are computed using a Lanczos scheme. This random deformation is continued

until the lowest eigenvalue becomes negative, indicating the presence of a nearby first-order saddle point. The configuration is then pushed, in the reduced holographic projection, along the eigenvector corresponding to this eigenvalue while the energy is relaxed in the perpendicular hyperplane. This activation phase stops when the total force becomes close to zero and the configuration has converged onto the transition state. During this first phase, motion can take place around all free pivots ; the rest of the protein is moved as blocks. For the relaxation phase, the configuration is then nudged over the saddle point and the system is relaxed into a new minimum using, now, the full set of degrees of freedom in cartesian coordinates. The new configuration is then accepted or rejected using a Metropolis criterion.

The advantage of this approach is that, by using the system's forces to direct the activation, collisions are naturally avoided, allowing large collective motions. An example of an ART nouveau step can be seen in Fig. 7.4 and is discussed in Section 7.2.6.

### 7.2.5 Forcefields

The holographic multiscale method can be used with any non-biased implicit solvation atomistic force field. We present here results with two different force fields : CHARMM19 with the solvation model EEF1 as well as the extended optimized-potential for extended proteins (EOPEP) forcefield, an extended version of the OPEP coarse grain forcefield.

For better integration with HMM, we developed an in-house version of CHARMM19, based on Ref. [192]. Following Ponder's approach with Tinker [215], we use our own torsion angle description, based on OPEP's [171], but adapting the prefactors to those of CHARMM19. We combined this potential with the EEF1 solvation model volume exclusion terms [145]. For the latter, we applied ionic neutralization of the side-chains, the N-terminal and the C-terminal, as specified in EEF1.

To distinguish between forcefield and methodological limitations, we also developed our own extended version of the OPEP forcefield. The original OPEP is an implicit solvent reduced-representation forcefield that includes all heavy main-chain atoms as well as a single bead for the lateral chains, with statistically-derived interactions.

Among its original features is a four-body interaction term to describe the cooperativity observed in the H-bond formation for stabilizing secondary structure. The quality of this potential has been well-tested in protein folding and protein aggregation simulations for chains of 70 residues or less [139, 171, 246, 298]. While this potential performs very well for secondary and tertiary structures, it fails to describe precisely more flexible regions where long residues often dominate. EOPEP inherits OPEP's ability to stabilize secondary structure using a cooperativity term for the evaluation of hydrogen bridges as well as provide side-chain hydrogen-bridge. In addition to these terms, EOPEP develops fully the lateral chains to ensure better packing for large residues.

More precisely, EOPEP represents all side-chain's carbon, oxygen, nitrogen, sulfur, and polar hydrogen (linked to O or N) atoms. The various carbon configurations are defined with implicit hydrogens. For these atoms, we adopt the van der Waals atomic dimensions of CHARMM19 [192]. Following OPEP, attractive and repulsive interactions between side-chain atoms are described using a screened Lennard Jones 6-12 term, except for the side-chain hydrogen bridges, for which we use OPEP's Lennard Jones 10-12 term used in the N-H-O interactions. This choice ensures that EOPEP can be used in conjunction with OPEP, providing a multilevel description inside a single protein and decreasing significantly the computational costs while adding the appropriate degree of complexity where needed. Because of their small size, however, proteins studied here are described fully by EOPEP.

The value for the extended set of parameters is shown in Table 7.I. To avoid overstabilizing side-chain-side-chain interactions, the hydrogen bridge strength for lateral chain is considerably reduced compared to the main-chain H-bridge. The hydrogen bridge receptor of carboxyl groups (aspartate, glutamate) conserves main-chain interaction strength, while other oxygen receptor see their contribution decreased by a factor ten, and the contribution of side-chain hydrogen bridge donor (nitrogens or some oxygens) by a factor of 20. These choices reproduce the aqueous immersion effect on them, as such contacts are generally only established consecutively to tertiary structure for-

Tableau 7.I – EOPEP parameters for side-chain atoms. These parameters are used in addition to the covalent bonding and main-chain main-chain interactions already defined in OPEP. EOPEP parameters were adjusted according to the procedure described in the text.

Interaction	Implementation	Partial contribution
Side-chain Hydrogen Bridge		
Carboxyl oxygen	3 terms LJ10-12	As OPEP Main-chain O
Other oxygen receptor	3 terms LJ10-12	0.10 vs Main-chain O
Nitrogen and oxygen donor	3 terms LJ10-12	0.20 vs Main-chain N
Side-chain non polar Carbons	2 terms LJ6-12	$0.05(\text{Kcal/mol})^2$
Sulfur	2 terms LJ6-12	As OPEP MET
Non HB interaction from polar atoms		
Oxygen and Nitrogen	2 terms LJ6-12	0.5 of OPEP ASP
Hydrogen	2 terms LJ6-12	0.05 of OPEP ARG

mation, which are directed by hydrophobicity [12]. Hydrophobic side-chain interactions also follow OPEP and are described by a uniform Lennard Jones 6-12 function, with a  $0.05 (\text{kcal/mol})^2$  well-depth contribution for any type of non-polar side chain carbon atom. Sulfur atoms, for their part, are directly copied from the OPEP methionin behavior.

The EOPEP forcefield was calibrated through extensive stability tests of the native state of protein A as well as calmodulin C and N terminals using long holographic ART simulations. EOPEP was subsequently used unchanged in folding simulations of these three sequences as well as Troponin C NT and Protein G.

### 7.2.6 Description of a typical holographic ART transition event

A typical HMM-ART event takes a few minutes on a single CPU. Fig. 7.4 reports details for a typical accepted event during a closing simulation of the open Calmodulin NT model, which undergoes transformation in its binding loops when deprived of its calcium ions [161, 265]. For this simulation, we defined as pivots residues 18 to 31, which include the first binding loop, and residues 54 to 67, which include the second binding loop. The event was started from a energy minimum at  $-372.9 \text{ kcal/mol}$ . The ART nouveau steps are identified on the plot and correspond to the exit from the harmonic basin, the convergence to the saddle point, the push over the saddle and the relaxation to a new



minimum. Looking at panels (a) and (b), we note that the energy goes up rapidly as the protein is deformed, reaching 1450 kcal/mol, at the exit of the harmonic basin in a random direction (step 3). Since the convergence to the saddle point is accompanied by a relaxation of the  $N-1$  other excited degrees of freedom, the energy falls rapidly during this next phase but the saddle point is still found at 528.6 kcal/mol above the initial minimum (step 17). This unphysically large energy is due to the fact that only angles around pivots are allowed to move, creating considerable strain in the other degrees of freedom, such as bond lengths and various angles. Indeed, most of this strain is released during the first steps of the relaxation to the final minimum, which takes place in Cartesian space, on all coordinates, and the final structure (step 37) is at 0.4 Å away and 4 kcal/mol (with a total energy of -376.7 kcal/mol) below the corresponding initial state values.

Panel (c) of Fig. 7.4 reports the root mean square average of  $\phi$  and  $\psi$  Ramachandran angles within each binding loop,

$$\theta_{\text{RMS}} = \sum_i \sqrt{(\Delta\phi_i)^2 + (\Delta\psi_i)^2} \quad (7.14)$$

where  $i$  runs on the active residues of loop 1 or 2.

The initial activation takes place mostly in the first binding loop, with relatively little deformation in the second loop. Interestingly, most of the angular displacement in this phase takes place during the exit of the harmonic basin and only small angular adjustments are observed during the convergence to the saddle point. Relaxation allows the full protein to react to this transformation and we see the second loop adjusting significantly during relaxation, creating a large cooperative event that changes considerably the protein conformation.

Table 7.II reports the averaged RMSD between initial and final states of events for various simulations. Events can be very large, typically between 2.0 and 4.0 Å. The largest moves are unlikely to lead to low energy structures, however, and at low Metropolis temperature, most moves tend to be rejected. Statistics for accepted events at 300 K show that most lead to displacements are around 1 Å, which is still considerable.

Tableau 7.II – Mean RMSD between events : statistics for a few simulation examples.

Protein model	Forcefield	Sim	mean RMSD		mean RMSD	
			at saddle all events	at min all events	at saddle accepted only	at min accepted only
Protein A	CHARMM19	5	2.29	2.39	1.27	1.28
Protein A	CHARMM19	16	2.93	2.96	0.96	0.77
Protein A	EOPEP	1	4.13	4.38	1.11	1.03
Protein A	EOPEP	20	3.10	3.25	1.16	1.09
Protein G	CHARMM19	2	2.24	2.03	1.26	0.98
Protein G	CHARMM19	11	2.07	1.94	1.12	0.95
Protein G	EOPEP	2	2.74	2.66	1.25	0.83
Protein G	EOPEP	14	1.93	1.52	1.19	0.69

### 7.3 Protein models

We tested here the ability of the method to perform long distance transformations. To do that, we first stretched two proteins while preserving intact their secondary structure. Folding simulations for protein A domain of protein A (1BDC pdb file), a 60-residue three- $\alpha$ -helix bundle, was started from a configuration with the three helices aligned and a 16.6 Å RMSD with respect to the native configuration (Fig. 7.5). A 56-residues fragment of protein G model (pdb file 1GB1), composed of a  $\alpha$ -helix and a four-stranded  $\beta$ -sheet, was also artificially opened to provide a starting conformation at 19.6 Å RMSD away from the native conformation (Fig. 7.7).

Figure 7.2 shows that it is possible to consider some part of the molecule as stable regions and avoid activations in these areas. Figure 7.8 reports the sequence of the protein A and protein G domains used and the  $C\alpha$  pivots that were allowed for activation. At each activation we performed a random choice of two trios of consecutive C-alpha pivots among the ones underlined in Figure 7.8. We activate at each of these chosen pivots in random directions for the three peptidic plane angles of the main chain and the for the three rotating block angles of the side chain.

## 7.4 Results and Discussion

We performed 24 folding simulations starting from the open version of protein A model, each with 150 to 300 accepted events for CHARMM19 and 300 to 500 for EOPEP at a Metropolis temperature of 300 K (except eop16a, which is discussed below). The evolution of the best runs, both in terms of energy and RMSD are reported in Fig. 7.9. EOPEP simulations 2 and 20 bring the model under 3.5 Å in respectively 3 hrs and 5.5 hrs CPU time for 50 to 75 events. Simulation 20 reaches 2.8 Å at the lowest-energy state of all simulation after 1.6 day CPU time. Convergence is slightly less for CHARMM19/EEF1, where the best simulation (che21a) comes to 3.8 Å with the native state.

We obtain similar results with protein G. Fig. 7.10 reports a selection of the 48 folding simulations starting from the stretched protein G model, each running for 200 to 400 accepted events at 300 K. The best RMSD descent is obtained with the CHARMM19/EEF1 simulation number 33, reaching 3.7 Å with the native model. This simulation also displays the lowest energy of all 48 runs using the CHARMM19/EEF1 forcefield. Several EOPEP simulations fall below 5 Å RMSD within 7 to 14 hours CPU time.

From HMM point of view, these bests results are excellent as HMM simulations tests of the stability for these two proteins show that the folding runs end up near the effective native state of each potential. Extensive HMM sampling simulations at 300 K starting from the native model for both protein A and protein show a stabilization near 2.5 and 1.7 Å-RMSD for protein A under EOPEP and CHARMM/EEF1 and 2.5 and 3.0 Å-RMSD, respectively, for protein G (data not shown), comparable to the best results of HMM folding.

The fact that only a fraction of all simulations reach the forcefield minima is due to the constant temperature scheme used here. If we simply heat up a structure trapped in a relatively high energy or RMSD basin and allow it to leave this metastable state, it is easy to bring the sequence in more stable state. This is the case for protein A EOPEP simulation number 16 which was trapped in a basin at around 10 Å RMSD after 250 steps. Raising the Metropolis temperature to 4000 K for about 150 events, the run ex-

plores configurations up to 17 Å RMSD away from the crystalline state, but rapidly folds to 4.6 Å RMSD after the temperature is brought back to 300 K. A similar but shorter test performed at the end of simulation che45g (protein G with CHARMM19/EEF1) enable the protein to relax from 7 to 6 Å RMSD from the native state.

Comparing the RMSD and energy evolutions in Fig. 7.9 and 7.10, we note that the energy drop typically follows the RMSD change, reflecting the existence of periods of local adjustment for minimizing the energy after a large configurational change. While this occurs for the four simulation sets presented in Fig. 7.9 and 7.10, the phenomenon is particularly obvious for protein G CHARMM19/EEF1 simulations presented in Fig. 7.10 (a) and (b) : while both energy and RMSD follow a similar drop in value, for each of the 6 presented simulations, the energy relaxation occur systematically tens or more events after a RMSD abrupt slope.

## 7.5 Conclusions and perspectives

This work presents the holographic multiscale method for generating efficiently large range protein motions. Large RMSD and energy steps are generated using the ART nouveau sampling method combined with a multiscale angular representation. In particular, by coupling block motion for the activation to full-scale atomic relaxation, HMM takes full advantages of the secondary structures while allowing it to change in response to large motion. This step is crucial to couple the role of side-chains, for example, with the stabilization of tertiary structure.

Results presented here on two model systems demonstrate the method's efficiency for generating large reorganization along well-controlled pathways. Clearly, however, a description of the energy landscape at the atomistic level increases the competition between local packing and block motion. This makes it difficult to accept events that get away from local energy basins without destabilizing the whole structure. The use of simulated annealing or other relaxation techniques would greatly help here, and further study will focus on identifying the best algorithm for constructing the proper Metropolis temperature sequence, which could involve an additional set of activated moves on side

chains, after each global move, to optimize the local structure before accepting or rejecting the event. The crucial result presented here, however, is that HMM can generate large motions that are not necessarily observed by normal mode analysis but that can nevertheless take place without uncontrolled steric clashes.

In addition to the model systems presented here, HMM has also been applied with success to characterize the folding of two EF-hands proteins, Calmodulin and Troponin C, upon removal of Ca ions. This work, to be published somewhere else, shows that HMM can be used to identify the specific events, atomic interactions and cooperative moves responsible for large scale changes [58].

HMM suffers nevertheless from two limitations. First, the method is not very effective for constrained loops as it requires being able to move a part of a protein with respect to the other, something that cannot easily be done while ensure a closing constraint. An appropriate selection of the initial deformation in ART nouveau step should be able to lift this inefficiency but more work remain to be done to establish a universal approach.

More complex, however, is the limitation on the forcefield. Because HMM works effectively at zero K, i.e., without thermal vibrations, it must be used with implicit solvent schemes, which are less precise than fully explicit models. The inherent flexibility of implicit-solvent model are obvious, especially with the coupling of the coarse-grained OPEP forcefield with the all-atom EOPEP. More work, however, must be done to ensure that these models work well away from the native state but this is beyond the development of a method such as HMM.

Overall, HMM provides an efficient and effective tool for characterizing the multiscale cooperativity associated with folding and large-scale motion. By providing physically possible pathways, HMM generates ensemble motion that are likely to be close to how real proteins work as long as forcefield are sufficiently reliable.

## **Acknowledgements**

This work was funded in part by the BiT bioinformatic scholarship program, the NSERC, FQRNT and the Canada Research Chair Foundation. Calculations were done

using resources from Calcul Québec.

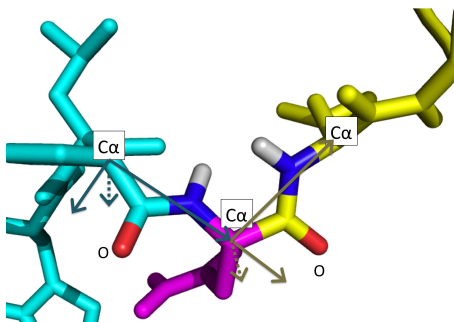


Figure 7.1 – A  $C\alpha$  is present at the root of the residue represented in magenta. An orthogonal basis is defined for the two peptidic planes on each side of this pivot, using the two ending  $C\alpha$  of the plane and the perpendicular direction within the plane, on the oxygen side. We obtain the absolute bases matrices  $A_{\delta-1}$  and  $A_{\delta}$ . Relative matrix  $B_{\delta}$  is then defined as the representation of matrix  $A_{\delta}$  in the orthogonal basis of matrix  $A_{\delta-1}$ . By projecting the cartesian-defined forces on these bases, we obtain spherical rotating/swivelling forces of the whole cyan area relatively to the whole yellow area. The side-chain represented here in magenta is also evaluated relatively to the two large areas, and also simulated as a rotating/swiveling block.

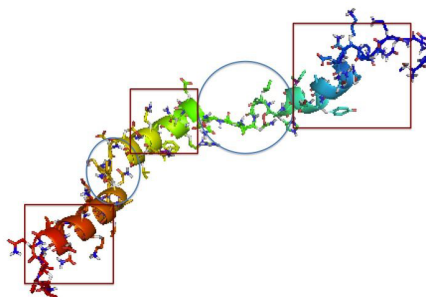


Figure 7.2 – All  $C\alpha$ 's of a protein may be defined as pivots. For multiscale simulations, we consider rather some part of the molecule (squares) as structurally stable and define pivots only on  $C\alpha$ 's in the flexible zones (circles). Protein A, for example, is defined in the following way :  $\alpha$ -helices are considered stable zones and all  $C\alpha$ 's in the loops and at the end of  $\alpha$ -helices parts bordering them are fully flexible.

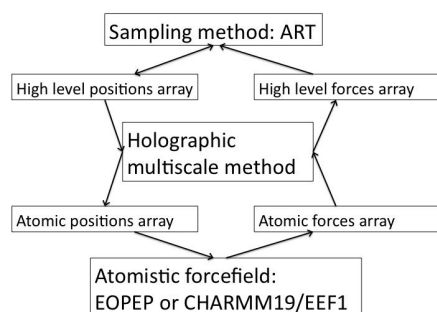


Figure 7.3 – Software organization. The activation and saddle passage steps of ART use the high-level pivot and blocks representation. After each iteration, ART returns new relative angular positions for the various blocks. These are translated back into Cartesian coordinates and the respective force is computed and then projected onto the reduced space corresponding to the high-level description.

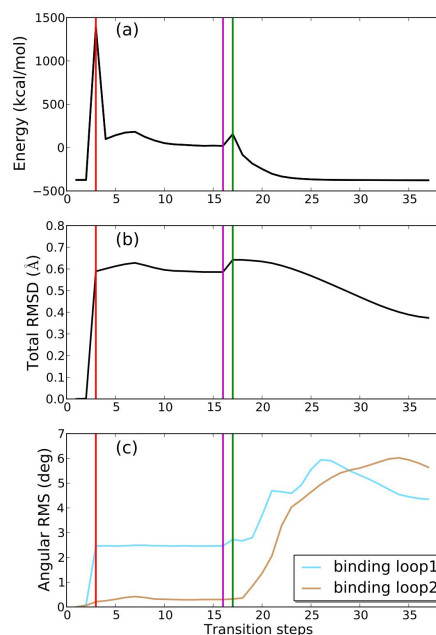


Figure 7.4 – A typical HMM/ART event during an EOPEP closing simulation of Calmodulin NT. (a) Evolution of the total energy as a function of iteration number. (b) Global  $C\alpha$  RMSD measured from iteration zero. (c) Root-mean square angular displacement computed over  $\phi$  and  $\psi$ , as measured from iteration zero and averaged over the two binding loops. The open CAM protein model is taken from pdb 1CLL, with residues 1-3 added, totalizing 76 residues. Vertical lines delineate the various phases of an activation/relaxation event. From left to right : push outside of the harmonic basin, convergence to the saddle point, push over the saddle, relaxation into the final minimum.



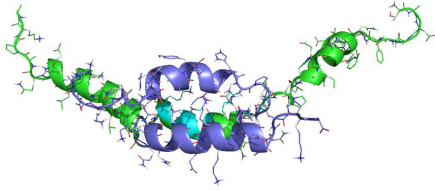


Figure 7.5 – Extended protein A with 3 helices aligned compared to native model. In cyan : the central helix 2. In blue : helices 1 and 3 folded toward helix 2 in the native form. In green : helices 1 and 3 extended away from helix 2, at the start of our simulations.

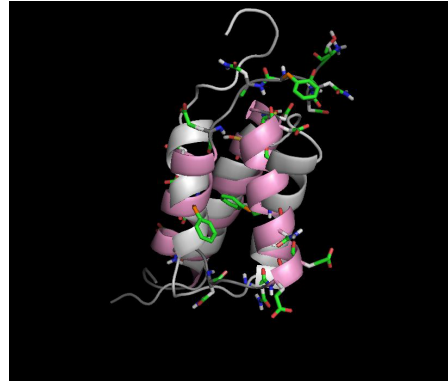


Figure 7.6 – Pink : protein A at end of EOPEP simulation 20. White : the native protein A model.

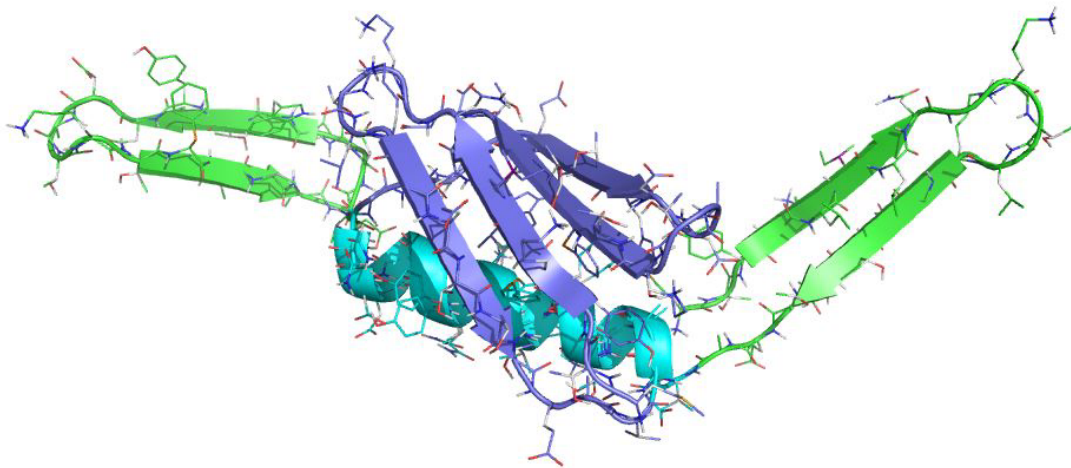


Figure 7.7 – Protein G with beta sheet widely opened, compared to native model. In cyan : the central helix. In blue : the beta sheet formation folded along the central helix in the native form. In green : in our simulations starting model, the beta sheet formation is split away at each end of the central helix

### protein A

TADNKFNKEQQNAFYEILHLPNLNEEQRNG

FIQSLKDPSQSANLLAEAKKLNDAQAPKA

### protein G

MTYKLILNGKTLKGETTTEAVDAAATAEKV

FKQYANDNGVDGEWTYDDATKTFTVTE

Figure 7.8 – Amino acid sequences of the protein A and protein G domains used in our simulations.  $\alpha$ -helices are represented in green,  $\beta$ -strands in orange and the loops in light purple. In the present study, activations and saddle passages were allowed for the  $C\alpha$  pivots underlined in this figure. Relaxations were performed in all-atom mode on the entire sequences.

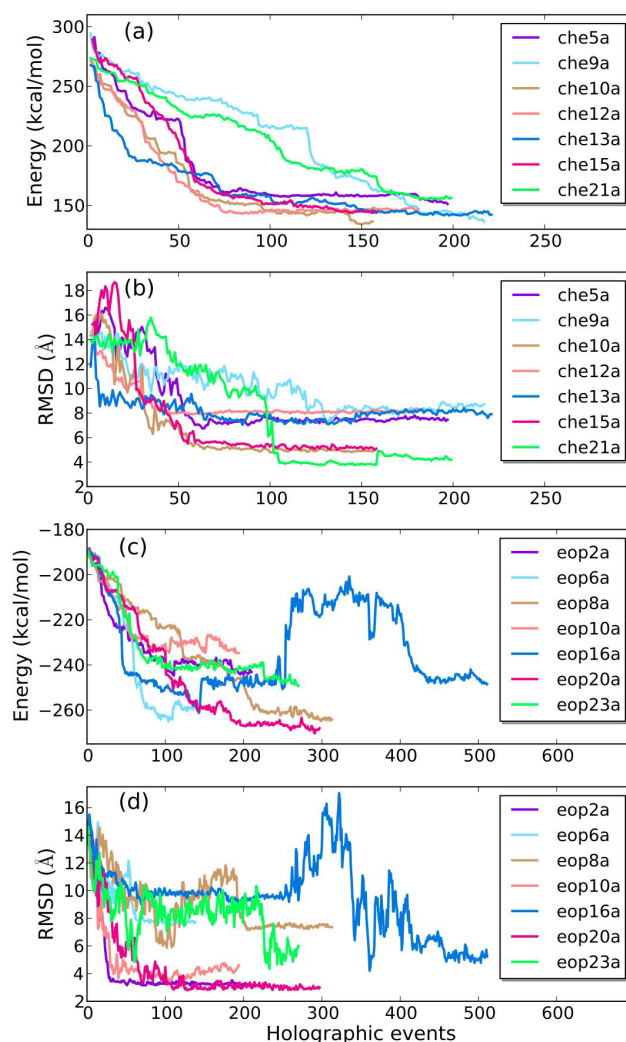


Figure 7.9 – 24 holographic simulations are performed starting from the artificially open model of protein A (three  $\alpha$ -helices aligned), using both CHARMM19/EEF1 and EOPEP. From each set we show the four trajectories visiting the lowest-energy structures and the four trajectories visiting structures with the lowest RMSD with respect to the native conformation (only simulation ch5a meets both criteria). (a) Energy and (b) RMSD evolution as a function of the number of accepted ARTn events using CHARMM1/EEF1. (c) and (d), same but with EOPEP forcefield. The RMSD is measured relative to the C $\alpha$  of the native form, excluding the unstable ending loops (residues 1-11 and 55-60). All simulations are run at 300 K, except eop16a, which is run at a Metropolis temperature of 4000 K from event 250 to a few events past 400 and then returned at 300 K.

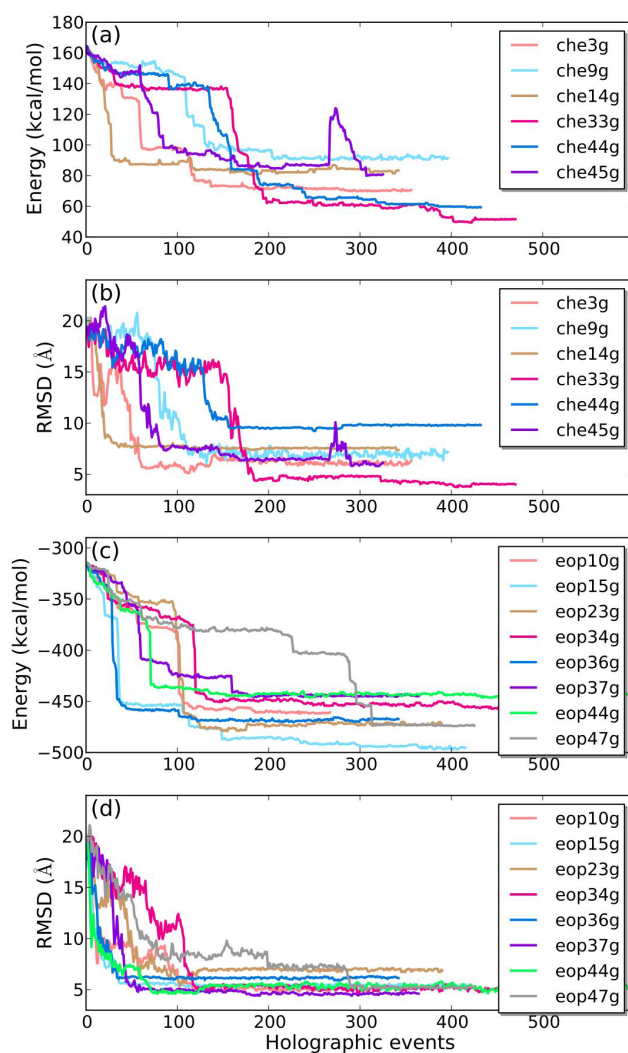


Figure 7.10 – 48 holographic simulations starting from the artificially open model of protein G. From each set we show the four trajectories visiting the lowest-energy structures and the four trajectories visiting structures with the lowest RMSD with respect to the native conformation (simulations che3g and che33g meet both criteria). (a) Energy and (b) RMSD evolution as a function of the number of accepted ARTn events using CHARMM1/EEF1. (c) and (d), same but with EOPEP forcefield RMSD is relative to all the  $C\alpha$  of the native form. All simulations are run at 300 K, except che45g, which is run at a Metropolis temperature of 4000 K for a few events around event 270 and then returned at 300 K.

## CHAPITRE 8

### ÉTUDE DE PROTÉINES À MAINS EF

Nous avons voulu voir comment la méthode *HMM* pouvait être appliquée à l'étude d'un mécanisme réel. Or, les domaines à mains EF de la molécule calmoduline (décrits au chapitre 4 et dans le présent chapitre) repositionnent leurs hélices  $\alpha$  de façon marquée lors de l'accomplissement de sa fonction. De plus, grâce à la cristallographie et à la *RMN*, nous disposons des deux conformations type pour cette protéine : ouverte ou fermée. En absence de calcium, la conformation ouverte a tendance à revenir vers la forme fermée. Nous pouvons donc générer des simulations à partir de la conformation ouverte et tester ainsi la capacité de la méthode à conduire vers la forme fermée. Pour mieux valider la méthode, nous avons utilisé une seconde protéine à mains EF : la troponine C. Celle-ci est de forme similaire à la calmoduline, avec environ 70% d'identité de séquence. Cela s'est avéré intéressant, car nous avons pu relier une différence dans le mécanisme à des différences précises dans la séquence.

Ma contribution d'étudiante a consisté à préparer et générer les simulations, à développer certains outils d'analyse, à analyser les simulations et à produire la version initiale de l'article. Les outils d'analyse requis concernaient le suivi de l'évolution des angles entre les hélices, ainsi que le suivi de l'évolution des énergies en distinguant la contribution de chaque résidu. Cela a permis d'observer les liens de cause à effet, ainsi que les phénomènes de coopérativité.

L'article a été publié au "Journal of Chemical Physics" en janvier 2012 et la figure 8.10(d) a fait la page couverture.

## **Understanding the EF-hand folding pathway using non-biased interatomic potentials**

L Dupuis<sup>1</sup> and N Mousseau<sup>2</sup>

<sup>1</sup>Département de Biochimie, Centre Robert-Cedergren and GEPRON, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7

<sup>2</sup> Département de physique, Centre Robert-Cedergren and GEPRON, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7

### **ABSTRACT**

**The EF-hand superfamily of proteins is characterized by the presence of calcium binding helix-loop-helix structures. Many of these proteins undergo considerable motion responsible for a wide range of properties upon binding but the exact mechanism at the root of this motion is not fully understood. Here, we use an unbiased accelerated multiscale simulation scheme, coupled with two force fields — CHARMM-EEF1 and the extended OPEP — to explore in details the closing pathway, from the unbound holo state to the closed apo state, of two EF-hand proteins, the calmodulin and troponin C N-terminal modules. Based on a number of closing simulations for these two sequences, we show that the EF-hand  $\beta$ -scaffold, identified as crucial by Grabarek for the EF-hand opening driven by calcium binding, is also important in closing the EF-hand. We also show the crucial importance of the phenylalanine situated at the end of first EF-hand helix, and identify an intermediate state modulating its behavior, providing a detailed picture of the closing mechanism for these two representatives of EF-hand proteins.**

Keywords : calmodulin, troponin, simulation, multiscale, protein folding, cooperativity

## 8.1 Introduction

The EF-hand motif plays a crucial role in eukaryotic cellular signaling. Protein members of this large family are characterized by a helix-loop-helix structure. Upon binding  $\text{Ca}^{2+}$  ions these motifs can act as simple calcium buffers as well as undergo large motion associated with a number of processes [10, 87, 294]. Among these EF-hand proteins, calmodulin has probably attracted the most attention. Using calcium as trigger token, this small 148 residues protein opens its hydrophobic cores [161] and binds to hundred of different protein targets [235, 294, 295], participating into a wide range of processes such as signal transduction [70, 91, 197], endocytosis [98, 251], muscular contraction [241], cell cycle regulation and mitosis [208, 262, 273], genes regulation [199, 200] and neural plasticity [7, 96]. calmodulin adopts an overall dumbbell shape, with a N-terminal module made of two EF-hands motifs at one end (helix pairs A-B and C-D), linked with a long and highly flexible  $\alpha$ -helix to a C-terminal module with a similar two EF-hands shape. Both ends may attach to different sites of a protein channel or grip around the same protein domain [9, 200]. The CT domain is considered to be more rigid, binding structurally to its target, while the NT domain, which has less affinity to calcium, is considered the regulatory domain [301].

Each EF-hand is composed of two typical helix- $\text{Ca}^{2+}$  chelation loop-helix units. It is this pair of turn-loop structures that provides the rich conformational space needed to accommodate the large scope of protein targets. To ensure its  $\text{Ca}^{2+}$  binding role, the canonical loop sequence in EF-hands is mostly preserved with a number of crucial residues at fixed positions [87]. While the role of each residue in binding this element has been extensively studied [87], their respective motion on the pathway connecting the holo to apo conformations is much less understood.

The full structure of the apo and the calcium-bound states of CaM is well known [35, 135]. It has served as a basis for characterizing the overall kinetics of the CaM EF-hands. This work is simplified somewhat by the fact that NT and CT domains are mostly independent from each other, so that it is appropriate to consider their properties separately [75, 261]. A number of experiments confirm the large conformational space sampled

by the unbound sequence. NMR relaxation experiments have shown that the kinetics is rapid for calcium-free CaM, suggesting important conformational changes over a 10-ns time scale [261]. However, the flexibility of the NT domain is not necessarily sufficient to span the full apo-holo conformational space [29]. In this case, the addition of calcium could force both NT and CT domains to stay open in absence of protein target, facilitating binding [39, 75, 130]. Other works suggest rather that calmodulin's inherent flexibility is sufficient to populate the closed form and the open form in presence of calcium and that the calcium loaded open form would only be stabilized by protein target presence [271].

Much less is known, however, about the exact pathway for going from one state to the other and numerous computational studies have attempted to answer part of the question. Confirming experimental observations, molecular dynamical simulations with local constraints have shown that the bound state of the CT domain displays large thermal fluctuations [61]. A number of studies have also focused directly on the apo to (unbound) holo pathway. Because of the difficulty associated with representing accurately both states, all simulations impose a bias in the force field to guarantee convergence to experimentally-derived states. Steered molecular dynamics, for example, was used to force  $\text{Ca}^{2+}$  binding and dissociation from various loops in the CT domain [129, 301]. Zhang *et al.*, for their part, used a double-Gō forcefield to study the unbound NT domain [300]. A similar approach was followed by Chen and Hummer, this time on the CT domain [38], allowing them to identify three phases along the pathway — apo-folded, apo-unfolded and holo-like structures — as well as by Tripathi and Portman who study in details the structural changes taking place during the transformation and find a significant rigidification of the loops in the NT domain as the protein moves from apo to holo conformations [265, 266]. These results generally support the experimentally-derived two-step Grabarek model for the  $\text{Ca}^{2+}$  binding mechanisms centered on a structure named  $\text{EF}\beta$ -scaffold, associated with the residues forming the  $\beta$ -sheet in the apo conformation, which defines the ion's position and allows the repositioning of the helices to ensure binding [93, 94].

In this study, we revisit this question with an unbiased approach that allows us to



characterize finely the pathway from the holo to the apo form. Following previous simulations, we first focus on the unbound two EF-hand motifs of the N-terminal module of calmodulin which has been extensively characterized. For comparison, we also simulate the holo to apo pathway for the EF-hand motif of the troponin C N-terminal, for which little numerical work exists. For this, we apply a multiscale scheme that samples the phase space efficiently while providing a clear description of intermediate states. This method is combined with two potentials — CHARMM19 [192] with EEF1 solvation model [145] and an extended version of the OPEP forcefield [171] (EOPEP).

While the details of the closing pathway differ slightly between the two sequences, our results confirm the EF-hand  $\beta$ -scaffold as a conformation switch, as proposed by Grabarek et Meyer [94, 177]. While Grabarek principally explored the conformational change towards the open protein form upon ion binding, our study focus on the closing pathway upon ion unbinding. We observe that the reverse pathway is initiated by the reverse reorganization of the  $\beta$ -scaffold. Our simulations also underline the importance of the phenylalanine situated at the end of first EF-hand helix of calmodulin and troponin C, providing a more complete picture of the closing mechanism for these representatives of this important class of proteins.

## 8.2 Methodology

The numerical study of closing pathways in flexible proteins is challenging and most simulations of EF-hand dynamics have used G $\bar{o}$ -like biased potentials to ensure successful transitions back and forth between the two states [38, 265, 266, 300]. Such a bias accelerates the effective dynamics allowing the determination of a free energy surface while ensuring that both terminal states visited correspond to the experimentally-derived structure. By removing possible metastable states, however, the bias can also affect the closing pathway in uncontrolled ways. To complement these works, we follow here an unbiased approach using a multiscale activated algorithm based on the activation-relaxation technique (ART nouveau) [17, 167] that has been applied successfully to pro-

tein folding [246, 277], protein aggregation [228, 278] and protein flexibility [298]. The holographic multiscale approach, which is presented in details in Ref. [57], allows to concentrate the conformational sampling on the flexible angles, which increases significantly the efficiency of the method. For completeness, we summarize the algorithm below.

This sampling method is coupled to two all-atom potentials with implicit solvent : CHARMM19 [192] with EEF1 solvation model [145] and an extended version of the OPEP model [171]. We describe briefly below the various elements of our approach.

### 8.2.1 ART

Sampling of the closing trajectory is performed using the Activation Relaxation Technique (ART nouveau) [17, 167, 182]. Starting from a conformation in a local minimum, the protein is deformed in a random direction according to the prescription described in the next subsection until a direction of negative curvature appears on the energy surface. The conformation is then pushed along this direction while the force is minimized in the perpendicular hyperplane, bringing the system onto a first-order saddle point corresponding to a local transition state. From there, the conformation is brought into a new local minimum. This set of conformations — initial minimum, saddle point, final minimum — forms an event. The full move is accepted using a Metropolis criterion based on the energy difference between the final and initial minima and a temperature of 300 K.

This method has been successfully applied to characterize the folding of a 16-residue  $\beta$ -hairpin [277] and the 60-residue protein A [246], as well as the aggregation of amyloid peptides [228, 279].

### 8.2.2 Holographic Multiscale Algorithm

The critical part for the efficiency of ART nouveau lies in the initial deformation. Since protein motions are based on angular deformations, it is much more efficient to work within internal coordinates [143, 144, 297]. Even in this reduced coordinate space,

a dominant fraction of possible moves involves breaking the secondary structure and lead to high-energy structures that would be rejected with the ART nouveau algorithm. Since non-local motions necessarily involve rotation of the  $\psi$  and  $\phi$  dihedral angles on each side of the main-chain  $C\alpha$ 's, our strategy is to constrain the initial deformation to loop  $\psi$  and  $\phi$  angles, with  $\alpha$ -helices able to react to the deformation, but not initiate it.

In the holographic multiscale approach, each main-chain movable angle is treated as a pivot around which the protein, separated as two blocks, and the lateral chain can rotate and swivel. As a consequence, the definition of flexible and rigid regions is dynamical : the set of splitting points change during one transition event, avoiding large block collisions, and increasing the flexibility of the approach. At each step during the activation phase, forces are evaluated in real space and projected on each pivot, allowing a local elastic response to the large scale motion generated by the move [57].

For the N-terminal domains of calmodulin and troponin C, we allow all  $C\alpha$  on the residues forming both calcium loops as well as those bordering them to act as eventual pivots. During the activation step, we randomly choose a few of them as initial pivots, with random rotating directions. These deformations, which do not have to respect biological rules, are corrected naturally during activation as all angles are allowed to adapt as the system converges to the transition state. To remove any bias imposed by the presence of rigid secondary-structure regions, the relaxation part of each event takes place in real-space representation using a damped-MD algorithm and applied to all degrees of freedom.

### 8.2.3 Forcefields

Until now, most EF-hand closing simulations have used biased potentials. We focus rather on two non-biased implicit-solvent force fields coupled with our sampling technique : our in-house implementation of CHARMM19 [192] with EEF1 solvation model [145] and an extended potential based on the OPEP model [171].

The implicit water solvent CHARMM19 potential is designed to be used in conjunction with different solvation models. We implement our own version of CHARMM19, as described in Ref. [192]. Following the approach of Ponder with Tinker [215], we use

our own torsion angle description, based on OPEP's [171], but adapting the prefactors to those of CHARMM19. This potential is combined with the EEF1 solvation model volume exclusion terms [145]. We applied ionic neutralization of the side chains, the N-terminal and the C-terminal, as specified in EEF1.

We also use an extended version of OPEP potential [171], generated by developing C, O, N and H atoms of the side chains. The coarse grain OPEP forcefield is based on a statistical approach for the evaluation of the side chains interactions, which are represented as unique spheres of different size, and introduces a cooperativity for main chain hydrogen bonds. It has proven very reliable for a large number of sequences and proteins of about 70 residues or less, including amyloids [139, 246, 298]. We find, however, that it is less appropriate for highly flexible sequence such as CaM and TnC, where a detailed description of the side-chain packing is necessary to ensure the stability of the native state. In order to maintain the OPEP's advantages, we develop here a new all-atom version of the potential, the extended OPEP maintaining a simple Lennard Jones 6-12 energy to compute the interaction between the different types of carbon, oxygen and nitrogen. As for the CHARMM19 forcefield, all polar hydrogens are explicit in EOPEP. Hydrogen bridges are also computed between side chain/side chain and main chain/side chain in addition to main chain/main chain, using the traditional 10-12 force term of OPEP. Main chain hydrogen bridge cooperativity is still evaluated as in OPEP. The dimensions adopted for the different side chain atom kinds (including the different kind of carbon implicitly attached to different number of hydrogens) have been taken from the CHARMM19 van der Waals parameters. Parameters are adjusted to generate a stable structure for the NT domain of calmodulin and protein A, but not for troponin C.

An additional advantage of EOPEP is that it is possible to couple it with the coarse-grained OPEP to provide a multilevel description inside a single protein, decreasing significantly the computational costs while adding the appropriate degree of complexity where needed. Because of their small size, however, both proteins studied here are described fully by EOPEP.

### 8.2.4 Sequences studied

We focus on two protein domains : the N-terminal module of calmodulin and the N-terminal module of the muscular troponin C (Fig 8.1). Each domain consists of two EF-hands whose cooperative binding to two  $\text{Ca}^{2+}$  exposes a hydrophobic core. Each EF-hand is defined by an  $\alpha$ -helix-loop- $\alpha$ -helix sequence. This motif is found in several other calcium binding proteins, with some variations. In calmodulin, each of the EF-hand motif central loop contains 12 residues with one glutamate and several aspartates, whose carboxyl group oxygen atoms that bind to a  $\text{Ca}^{2+}$  ion [147, 152]. Binding to calcium forces a significant rearrangement of the two helices as they move from a rather parallel orientation in the apo form to a mostly perpendicular orientation [161]. In the closed form, the two calcium binding loops are interrelated by the formation of an interloop short  $\beta$ -sheet stabilized by three hydrogen bonds [135] and identified as the EF-hand  $\beta$ -scaffold by Grabarek [93, 94].

The 76-residues NT module CaM apo model used is based on the NMR resolved structures of Ref. [135] (PDB : 1CFD) while the calcium bound NT model comes from the crystallographic characterization of Ref. [35] (PDB : 1CLL). We added the 3 missing first residues to the later model, and used residues of the NT part, up to the 76th. Both NT configurations are illustrated in Fig. 8.4 panels (f) and (a) respectively.

The troponin C also presents a dumbbell shape with two E-hands motifs at each end. This protein is dedicated to the activation of actin and myosin muscle microfibers components, by blocking the inhibitor, troponin I [68]. It is known that the CT module of

```

CaM NT
EFhand1      ADQLTEEQIAEFKEAFSLFDKDGGTITTKELGTVMRSL
EFhand2      GQNPTEAELQDMINEVDADGNGTIDFEFLTMMARKM
TpC NT
EFhand1      ASMTDQQAEARAF LSEEMIAEFKAAFDMFDADGGGDIKSTELGTVMRMLG
EFhand2      QNPTKEELDAIIEEVDEDGSGTIDFEFLVMMVRQMKEDA

```

Figure 8.1 – Sequences for the NT domains of calmodulin and troponin C. Calcium binding residues are in red ;  $\alpha$ -helices are in green and the EF-hand  $\beta$ -scaffold residues are underlined. The residues at loop position 7 bind using the main chain oxygen.

troponin C is structurally linked to troponin I, while the action of NT module is triggered by the calcium ion [67]. The NT domain is composed of 5 helices, as helices 2-5 compose the 2 EF-hand motifs (helix pairs A-B and C-D), while the first helix is denoted “N”. Not counting helix N, the N-terminal domain has 70 percent identity with the calmodulin NT. Each troponin binding loops also contains 12 residues, but with a sequence that differs slightly from that of calmodulin. The first binding loop of the NT has lower affinity to calcium [82], presenting a glycine at position 5 instead of an aspartate. For the simulations, we used the entire 90 residues of model 01 presented in PDB 1TNQ [83] without the  $\text{Ca}^{2+}$  ions. We compared the configurations generated by the simulations with each of the 40 closed models included in PDB 1TNP [83].

### 8.2.5 Analysis

For all sequences RMSD is measured on  $\text{C}\alpha$ 's only. For calmodulin, RMSD includes residues 3 to 76, leaving aside residues 1-3, which are disordered and highly flexible. For troponin C, we omit the unstable extremities — 1-15 and 86-90 — from the RMSD calculations. RMSD are measured with respect to the non-relaxed original experimentally-derived PDB files. Angles between helices are computed using the principal axis method. The inertia tensor, which includes all main chain and the  $\text{C}\beta$  atoms on each helix, is diagonalized and the principal axis is used as reference. Our analysis also includes individual amino acid energy difference, underlying cooperativity between specific groups. Energy contribution by amino acid is evaluated by adding contribution of every main chain and side chain atom composing this amino acid. Individual atom energy statistic comes from the different forcefield terms where two, three and four-body contributions are split evenly between atoms involved.

## 8.3 Results and Discussion

### 8.3.1 Simulations on calmodulin NT

#### 8.3.1.1 Stability of the native apo conformation

We first apply CHARMM19-EEF1 and EOPEP to the native unbound closed model of CaMnt to assess its stability and the size of the apo basin under each potential. For both forcefields, we perform a local unconstrained minimization from the experimental apo form at  $T=0$  followed by eight independent ART simulations with a 300 K Metropolis temperature and a minimum of 100 accepted events. This is a stringent test as displacement per event can reach 1 Å RMSD.

Both simulation sets lead to stable apo structures, with unmodified  $\beta$ -sheet and  $\alpha$ -helices. With CHARMM19-EEF1, the CaMnt apo form stabilises within a region of about 1.5 to 2.5 Å RMSD away from the experimentally-derived structure, with the lowest-energy structures fluctuating between 2.0 and 2.5 Å (Fig. 8.2 (a) and (b)). Using EOPEP, the calmodulin NT apo form evolves to a region of about 2.5 Å (Fig. 8.2 (c) and (d)). Trajectories visiting the lowest-energy structures tend to remain slightly closer to the experimental-derived structure, at about 2.3 Å. In all cases, the displacement away from the experimental data is due to fluctuations in the tertiary structure, with unchanged secondary structure.

#### 8.3.1.2 Folding from the unbound holo conformation

Having established the stability zone of the apo conformation, we now turn to characterizing the closing pathway. For each forcefield, 24 independent simulations with a minimum of 100 accepted events are launched starting from the open model of CaMnt with the calcium ions removed, at 5 Å RMSD of the apo structure, and without imposed bias, with a 300 K Metropolis temperature.

Table 8.I reports the averaged RMSD between initial and final states of events for various holographic ART simulations on the open CaMnt model. The generated events were typically between 1.2 and 1.7 Å. The largest moves are unlikely to lead to low

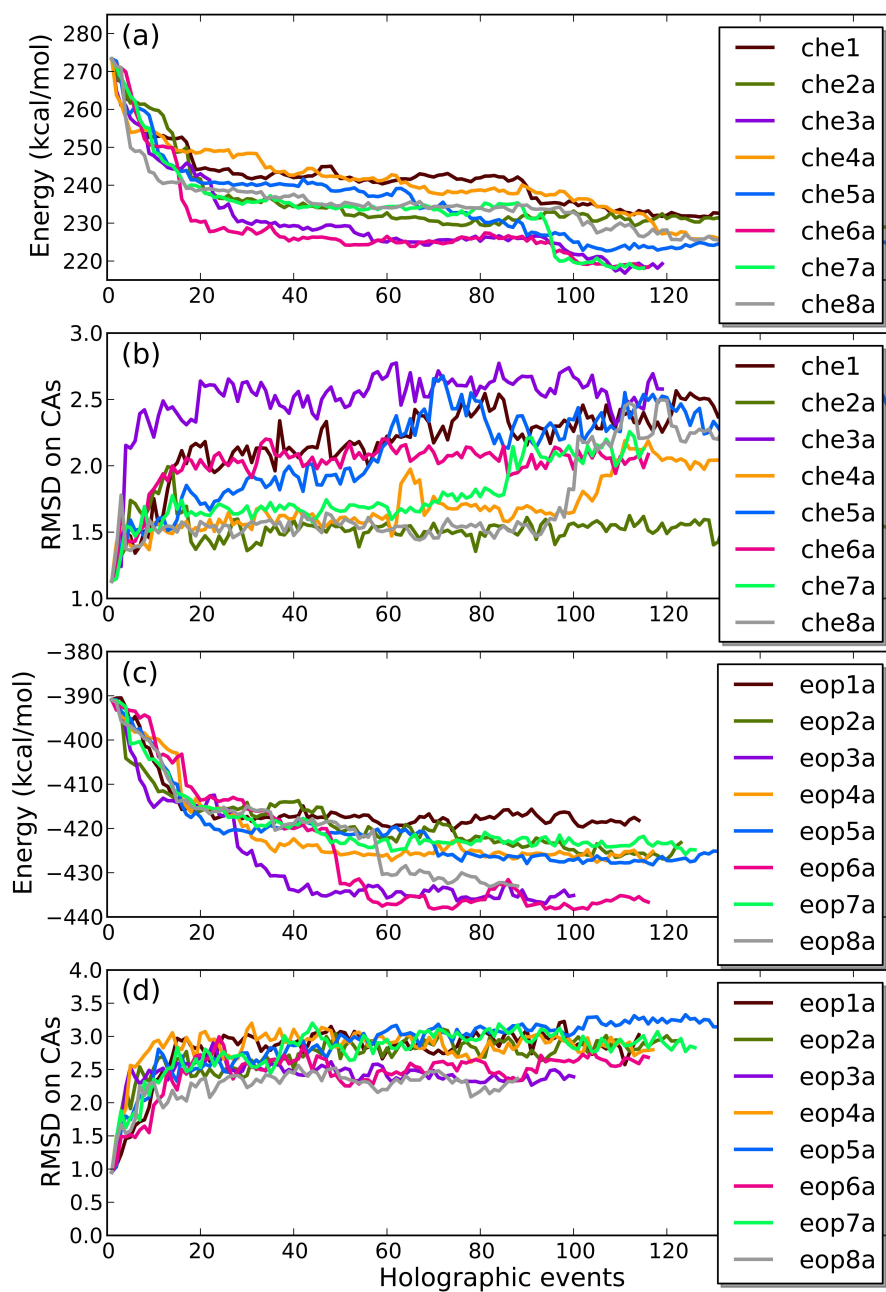


Figure 8.2 – Energy and RMSD measured from the experimentally-derived closed calmodulin NT structure as a function of accepted ART-event for eight independent simulation started from the experimental apo conformation. (a) and (b) Energy and RMSD results, respectively, for CHARMM19-EEF1 ; (c) and (d) same for EOPEP. RMSD are computed on  $C\alpha$ , excluding unstable residues 1-3.



energy structures, however, and at low Metropolis temperature, most moves tend to be rejected. Statistics for accepted events at 300 K show that most lead to displacements are around 0.6 Å.

Tableau 8.I – RMSD statistics for selected trajectories. Averaged displacement measured from the initial minimum to the saddle point and the final minimum for various simulations on both the calmodulin and troponin C using OPEP char CHARMM/EEF1 potentials. Statistics are given both for the full sets of events and those accepted only.

Model	Forcefield	Sim	Averaged RMSD in (Å)			
			saddle all events	final min all events	saddle accep.	final min accep.
CAM	CHARMM	14	1.53	1.24	0.73	0.44
CAM	CHARMM	18	1.77	1.64	0.79	0.61
CAM	EOPEP	5	1.78	1.70	0.81	0.62
CAM	EOPEP	9	1.73	1.61	0.83	0.65
TPC	CHARMM	11	1.24	1.09	0.72	0.53
TPC	CHARMM	17	1.35	1.15	0.79	0.58
TPC	EOPEP	16	1.88	1.84	0.83	0.71
TPC	EOPEP	21	1.69	1.63	0.87	0.71

Using this protocol with CHARMM19-EEF1, only one pathway out of 24 reaches the apo state, passing within a minimum distance of 2.2 Å and stabilizing at 2.4 Å. Most other runs visit higher-energy structures at 3.5 to 4.0 Å RMSD from the apo state. For legibility, we show respectively in Fig. 8.3 the configurational energy (panel (a)) as well as the RMSD (panel (b)) for the four simulations leading to the lowest energy as well as the four with the smallest RMSD with respect to the target CaMnt apo form, for a total of six curves (trajectories che12h and che14h belong to both groups). We see that, for the six trajectories, the energy goes down systematically, reaching a plateau during the runs in all but one case (simulation che12h). RMSD fluctuations are large except for simulation che14h, that reaches the apo state within 20 events.

Folding is easier with EOPEP, and ten out of the 24 runs performed with this force-field sample within 2.2 Å of the experimentally-derived closed structure, reaching a distance identical to the closed-state basin in the previous section. As with CHARMM19-

EEF1, Figure 8.3 (c) and (d) shows the energy and RMSD as a function of accepted even for the four trajectories with lowest energy and RMSD. Again, the minimum RMSD, 1.9 Å to 2.3 Å, is obtained very quickly, typically within less than 30 events. Not all of these apo-like structures are stable, however, and, in some cases, the trajectory evolves away from this conformation as the energy goes down. Here again, an energy plateau is generally attained within 80 to 90 events.

Before looking into the successful closing trajectories, we first summarize the critical closing steps extracted from our simulations, represented in Fig. 8.4 :

1. Increase in the distance between the side chain oxygen atoms of the calcium binding loops.
2. Consolidation/elongation of the  $\beta$ -sheet linking the two calcium loops : a bond is formed between THR29 and GLY61.
3. Repositioning of PHE19 at the outside of the NT module, allowing a tighter closure of helices B-C toward helices A and D. During this process, the PHE19 must pass through a a three-residue pocket, LEU-MET-VAL, positioned at the bottom of helix B and where it can be trapped as a metastable state before being ejected from the hydrophobic core.
4. General increase of hydrophobic contacts as each EF-hand helices pair falls into a stable parallel conformation.

Of the 24 EOPEP simulations, 5 fold into the apo state, all following the above mechanisms. For clarity, we focus on the simulation eop9h, which flows near the experimentally-derived closed apo form but also adopts all the important structural elements of the apo state, including the  $\beta$  strand, the exit of PHE19, and the closing of the hydrophobic core.

### 8.3.1.3 Oxygen carrier binding residues

Simulations start from the open model with  $\text{Ca}^{2+}$  removed. The calcium binding residues are located at positions 1, 3, 5, 7, and 12 of each of the two binding loops. At position 12, we have a glutamate, GLU31, for the first EF-hand, positioned at the beginning

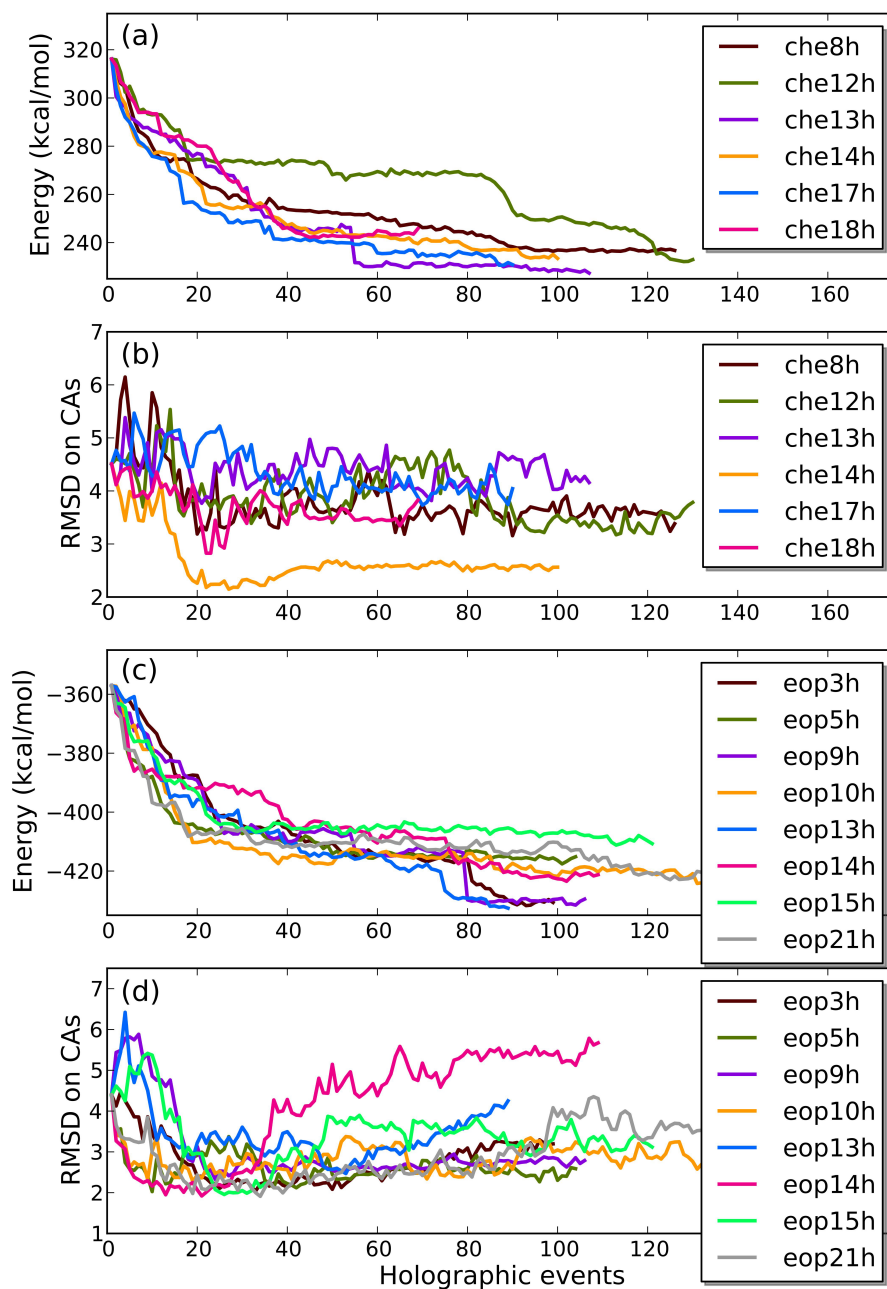


Figure 8.3 – Energy and RMSD measured from the experimentally-derived closed CAMnt structure as a function of accepted ART-event for independent simulations started from the experimentally-derived open conformation with Ca ion removed. From a set of 24 trajectories, we present the four simulations with the lowest-energy and RMSD structures, respectively. Panels (a) and (b) show the evolution of the energy and RMSD for CHARMM-EEF1 simulations, respectively; panels (c) and (d) present results for EOPEP simulations.

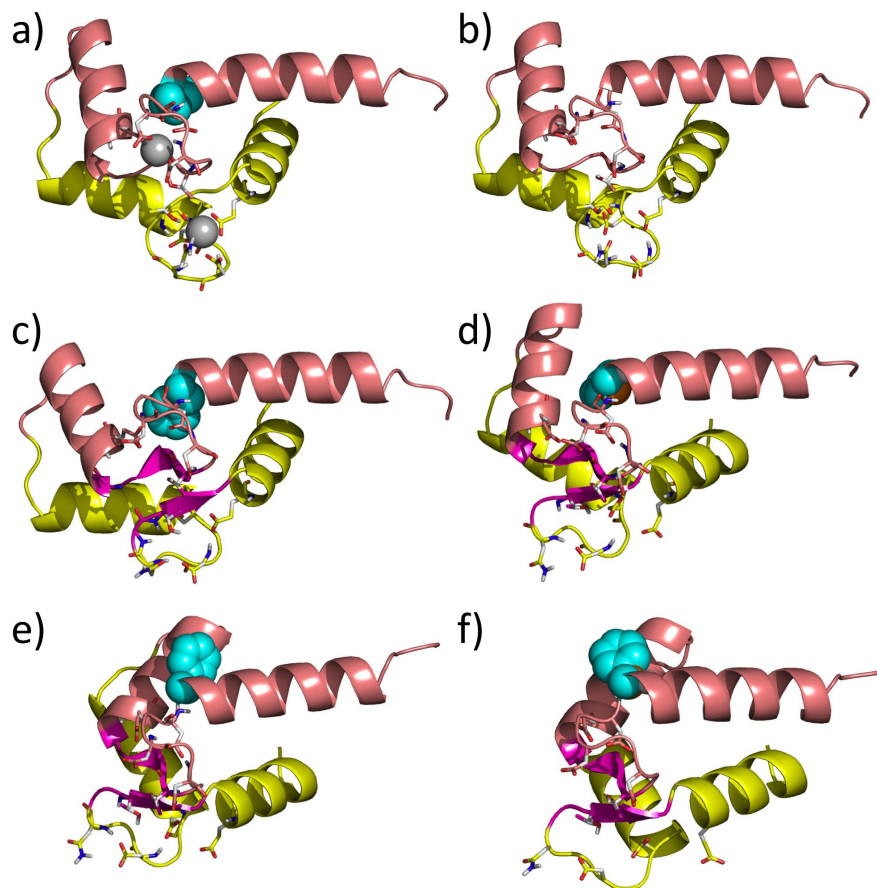


Figure 8.4 – Closing pathway for simulation eop9h. Each EF-hand pair goes from a rather perpendicular arrangement in the holo state toward a more parallel state in the apo form. (a) 1CLL.pdb model up to residue 76, residues 1-3 added. (b) After a first minimization without  $\text{Ca}^{2+}$ , we observe an increase in the distance between oxygen carrier residues of the loop. (c) Event 3 : consolidation of the  $\beta$ -sheet (magenta) linking the 2 calcium binding loops. A hydrogen bridge is formed between THR29 and ILE61. (d) Event 16 : Increase in the number of hydrophobe interactions, the core begin to close, but the presence of PHE19 (cyan) prevent a better packing. (e) Event 19 : PHE19 gets out of the hydrophobic core, allowing a better closing of the core. (f) Event 24 : The helices pack together near the apo form. The end helices (A and D) are held relatively fixed in these views, in order to show the cooperative motion of helices B and C. For ease of viewing, helices A-B (EF-hand 1) are colored in peach and helices C-D (EF-hand2) in yellow.

of the EF-hand exiting helix. In the presence of  $\text{Ca}^{2+}$  ions, GLU31 maintains the bottom of helix B near the calcium loop. As the calcium is removed, it leaves behind several negatively charged oxygen within a close distance in the loop area, a very unfavorable cohabitation [249]. Not surprisingly, this unstable conformation relaxes instantaneously in our simulations; this step has already taken place by event number one. Fig. 8.6 (a) shows, however, that not all side chains on the binding loop reach equilibrium at once and a few residues, such as GLU20, GLU56 and GLU58, continue their reorganization. This allows, for example, GLU20 to form several hydrogen bridges with the main chain nitrogen of the loop, acting as an elbow that stabilizes the motion between helices A and B in the apo conformation.

#### 8.3.1.4 $\beta$ -sheet scaffold consolidation

After a few events, a number of EOPEP simulations show the formation of the  $\beta$ -sheet scaffold between the binding loops. This event generally takes place early in simulation (event 22 of eop3h, event 11 of eop5h and eop13h, event 3 of eop9h, event 10 of eop10h, event 4 of eop15h, event 9 of eop21h), except for simulation eop14h (event 36). Fig. 8.6 (b), which shows the four residues involved in this short  $\beta$ -sheet, illustrates the precocity of this event in simulation 9, with a concomitant and abrupt energy drop for THR29 and ILE61. In parallel, we see that ILE27 and ILE63 consolidate their double hydrogen bridge. These results support the EF-hand  $\beta$ -scaffold model of Grabarek [93, 94] which strongly correlates the ILE27 Phi and Psi conformation changes with the angular transition of helices A and B between the CaM closed and open conformations. Interestingly, these first reorganization steps often bring the structure away from the final closed form before converging towards it, as can be seen in Fig. 8.3 for EOPEP simulations 9, 13 and 15. Among the 10 EOPEP simulations closing within 2.25 Å of the apo state, only simulation eop14h manages to close the protein before zipping the  $\beta$ -sheet. This near native-state structure is unstable, however, and further relaxation pushes the simulation away from the apo state as can be seen in Fig. 8.3. This mispacking correlates with a delayed  $\beta$ -sheet zipping that occurs at event 36.

None of the CHARMM-EEF1 simulations manages to close the  $\beta$ -sheet scaffold,

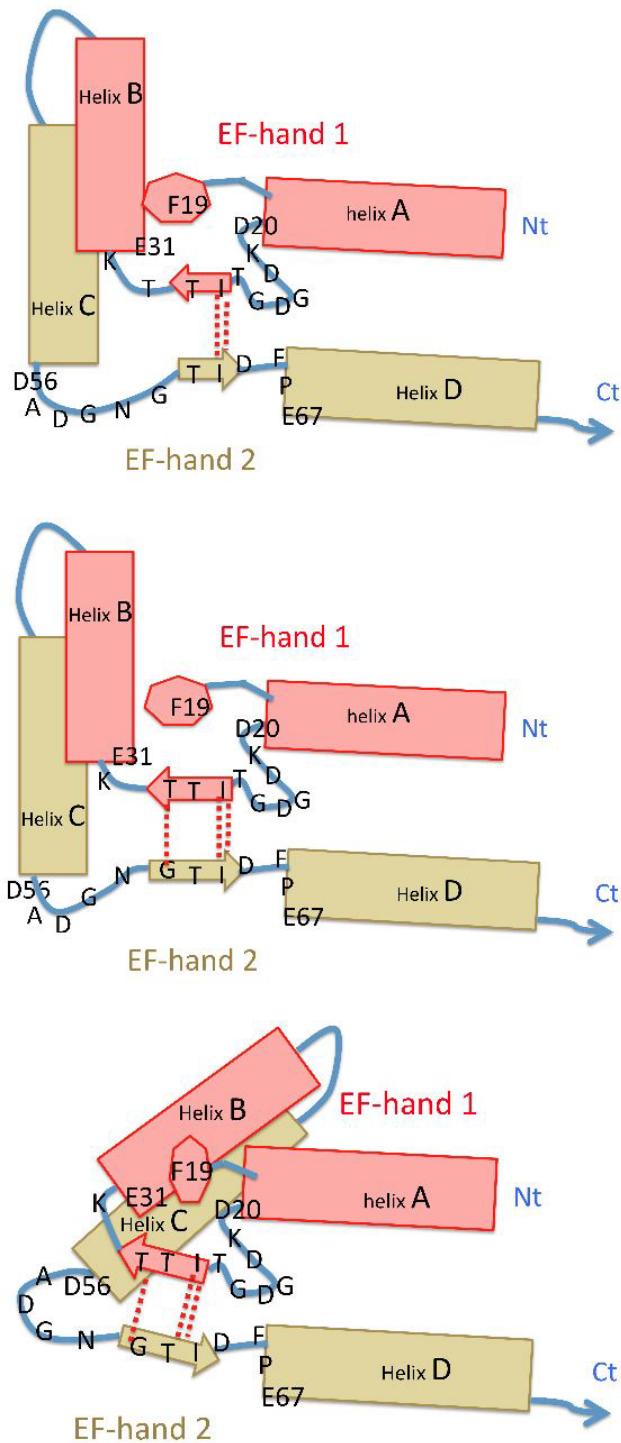


Figure 8.5 – Diagrammatic representation of the calmodulin NT closing mechanism. Top : initial open state. Middle : Consolidation of the  $\beta$ -sheet, widening out the space between helices, freeing PHE19 latch. Bottom : Closing of the helices in parallel configuration, PHE19 getting outside the hydrophobic core.

which explains why only one trajectory out of 24 reaches the apo conformational basin. However, we observe the  $\beta$ -sheet zipping with CHARMM19 without solvation terms (data not shown), thus enhancing the strength of main chain hydrogen bridge. As expected, the inappropriate attractions from charged side chains resulting from the absence of a solvent prevent the molecule from performing a proper closing.

In addition, we also tested for 24 simulations on calmodulin NT using EOPEP without using the main chain cooperativity term. None of these were able to close the  $\beta$ -sheet scaffold, nor complete the closing transformation. Clearly, as pointed out by Grabarek, the EF-hand  $\beta$ -scaffold is a crucial step on the pathway connecting holo and apo. This transition is schematized by the first two panels of the diagram in Fig. 8.5.

### 8.3.1.5 Exit of phenylalanine 19

$\beta$ -sheet zipping is generally followed by the formation of hydrophobic contacts between helices. Examination of Fig. 8.6(c) indicates that these contacts appear in two phases during simulation eop9h.

First, as ILE27 and ILE63 adopt a  $\beta$ -conformation, a few contacts are initiated, involving PHE16, VAL35 and VAL55 (events 1-10, Fig. 8.6-c). Second, around event 19, we observe a more general energy relaxation for hydrophobic core residues. To improve packing, two residues, PHE12 and PHE19, move away from the core. PHE19 escapes the hydrophobic core in only four simulations. From this and other simulations, it seems that an early occurrence helps to further stabilize the structure in the apo conformational basin (event 19 of eop9h, event 15 of eop10h, as well as events 10-14 of che14h). While the RMSD of several other simulations can also reach values below 2.5 Å, their final states do not display the correct position for the PHE19, which should be outside the hydrophobic core (see Figs. 8.4 (e) and (f) and 8.8 (a)). This suggests that both the  $\beta$ -sheet formation and the PHE19 escape must occur in a specific order during closing to ensure that the protein reaches the correct apo state. The PHE19 has a structurally analog in the CT lobe, PHE 92. In both domains, this phenylalanine is located at the end of the first EF-hand helix, just before the first aspartate of the first binding loop. Experimental mutation studies of this position in the CT module revealed an instability of the closed

CT module form when PHE is mutated to ALA. This instability is associated with a decreased accessibility of several side chains of the hydrophobic core in the calcium bound open form [177]. From our simulations, PHE19 appears to play a role in the open architecture, behaving as a lock of the EF-hand into the open form (see top panel of Fig. 8.5 and both blue models in Fig. 8.8 with this position pointed by cyan arrows). Thus, the absence of the bulky phenylalanine at this strategic position may compromise the full opening of the hydrophobic core.

Simulations eop9h, eop10h, eop11h and eop23h represent the ideal folding scenario as all structural details associated with the apo state are in place by the end of the simulation. It is instructive to compare with trajectory eop21h, for example, that visits low RMSD structures ( $2.0 \text{ \AA}$ ) and manages a late expulsion of PHE19 towards the solvent at event 98. The order in which this step takes place is associated with an incorrect packing of the hydrophobic core that prevents full stabilization into the native state. Even though PHE19 is now outside of the hydrophobic core, the conformation does not correspond as well to the native apo state and it moves away from it after event 98 (Fig. 8.3). With further simulations, however, as discussed in Subsection 8.3.1.8, the protein reaches a stable apo state similar to the of the four other fully closed simulations.

Two more sets of 24 simulations showed a similar success rate for reaching the apo from the holo state with, respectively 4 and 5 successful PHE19 exit, after about 100 accepted events (data not shown), suggesting that the results obtained here are typical. As shown in Subsection 8.3.1.8, extending the simulation time or raising the temperature can increase significantly this rate of success.

### 8.3.1.6 Closure of the hydrophobic core

The evolution of a selection of 14 hydrophobic residues is shown in Fig. 8.6 (c). Simulation eop9h reaches its lowest RMSD,  $2.3 \text{ \AA}$ , at event 24 but an overall energy relaxation for the hydrophobic core extends clearly up to events 85. Starting from the open state, the NT module undergoes a motion that first moves away from the closed model before converging back towards the apo basin. It then continues to relax through a series of smaller adjustments as the hydrophobic residues react to each other. The non-



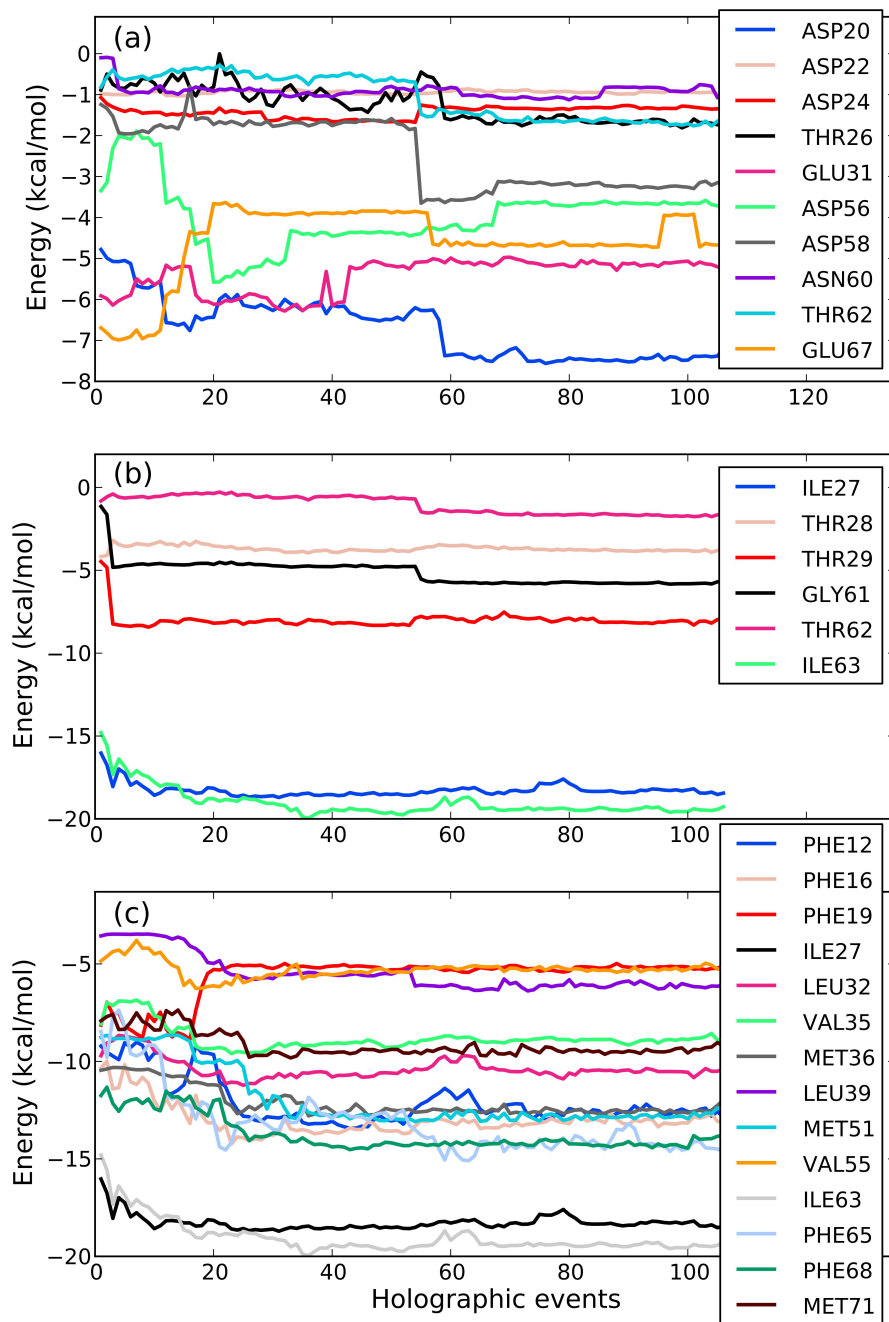


Figure 8.6 – Evolution of the energy as a function of accepted ART events during eop9h simulation. (a) Evolution of the calcium binding residues. Following calcium removal, the zero temperature minimization already causes a significant energy drop before the first event. (b) Evolution of the six residues participating in the formation of the short  $\beta$ -sheet associated with the closed form. (c) Evolution of the CaMnt 14 hydrophobic residues.

linear relaxation is clearly seen at events 17 and 18, when PHE19 leaves the hydrophobic core for an energetically uncomfortable position that allows several other hydrophobic residues to adopt low-energy conformation, as illustrated by the third panel of Fig. 8.5. PHE12 initially accompanies PHE19's extraction from the hydrophobic core, but while PHE19 remains in its outside position, PHE12 later inserts back into the core.

### **8.3.1.7 EF-hand helices angular evolution**

It is possible to characterize the closing trajectories by following the relative angle between the four helices forming the EF-hand [10, 294]. Fig. 8.7 shows the evolution of various angles for simulations eop9h and eop14h. These graphics underline the cooperativity between the two EF-hand pairs. For the first simulation, both pairs A-B and C-D move rapidly, within 20 and 45 events, respectively, from the open to the closed values, and oscillate around their experimental value as the protein adjusts the finer structural details. In some cases, eop14h, for example (Fig. 8.7, bottom), we observe an overshoot motion in the first helix pair that pushes the protein away from the native state.

### **8.3.1.8 Higher temperature simulations : characterization of the pathways to the apo state from metastable minima**

Even though the low temperature of the first simulation set allows us to clearly identify the various steps needed for bringing calmodulin into its apo state, it leaves a large fraction of the runs into metastable, non-native conformations. To assess the validity of the model presented in the previous paragraphs, it is necessary to see how the protein manages to escape from these metastable states and find its way to the close conformation.

For this purpose, we extended the simulations on the 24 EOPEP runs at two Metropolis temperatures : 300 K, similar to the initial run, and 900 K, to allow the protein to escape local minima. Even though 900 K might appear large, this temperature is fictive as thermal fluctuations are not taken into account here and the protein still moves from local-energy minimum to local-energy minimum [246]. For both temperatures, 1500

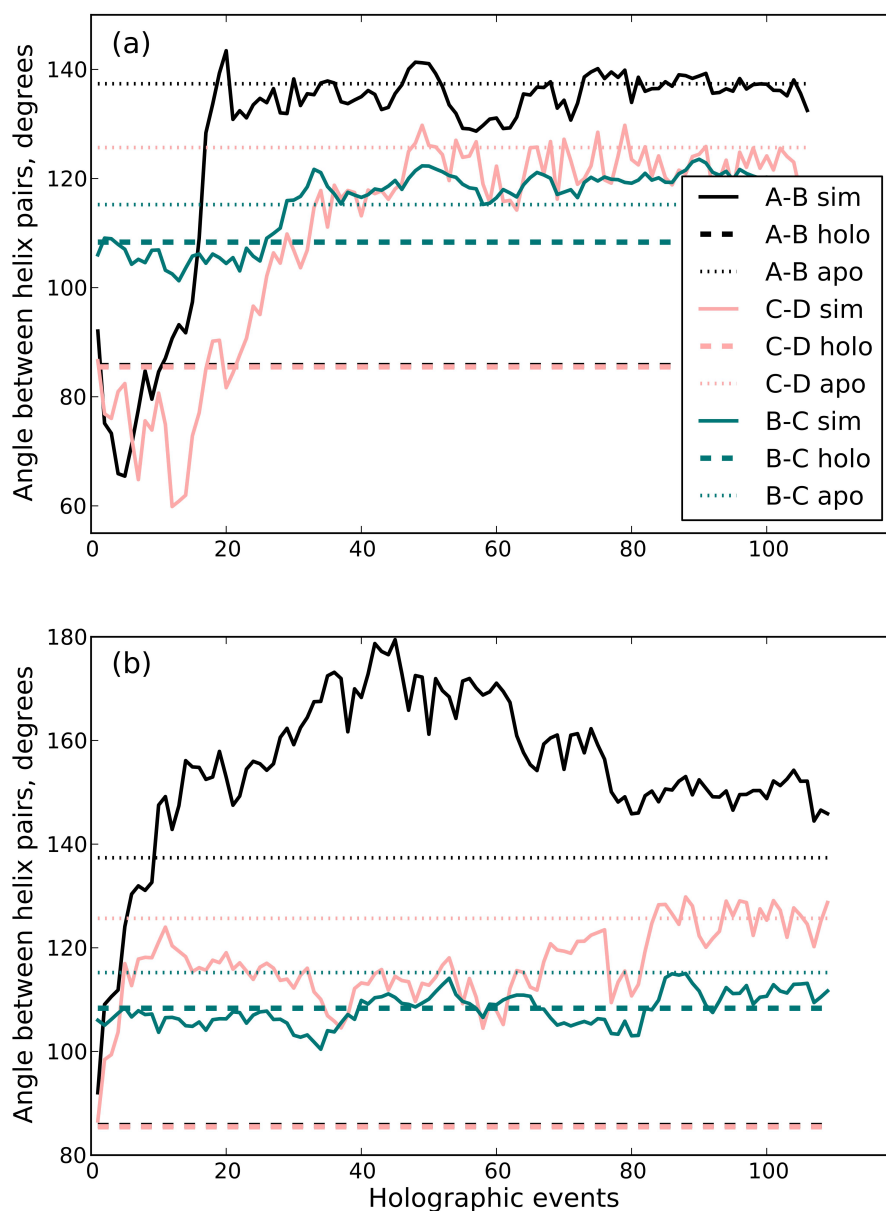


Figure 8.7 – Top : Evolution of the angles between various helix pairs as a function of accepted ART events for simulation eop9h. Dashed (dotted) lines indicate the value for the experimentally-derived open (closed) form. Bottom : Evolution of the angles between various pairs of helices as a function of accepted ART events for simulation eop14h. Dashed (dotted) lines indicate the value for the experimentally-derived open (closed) form.

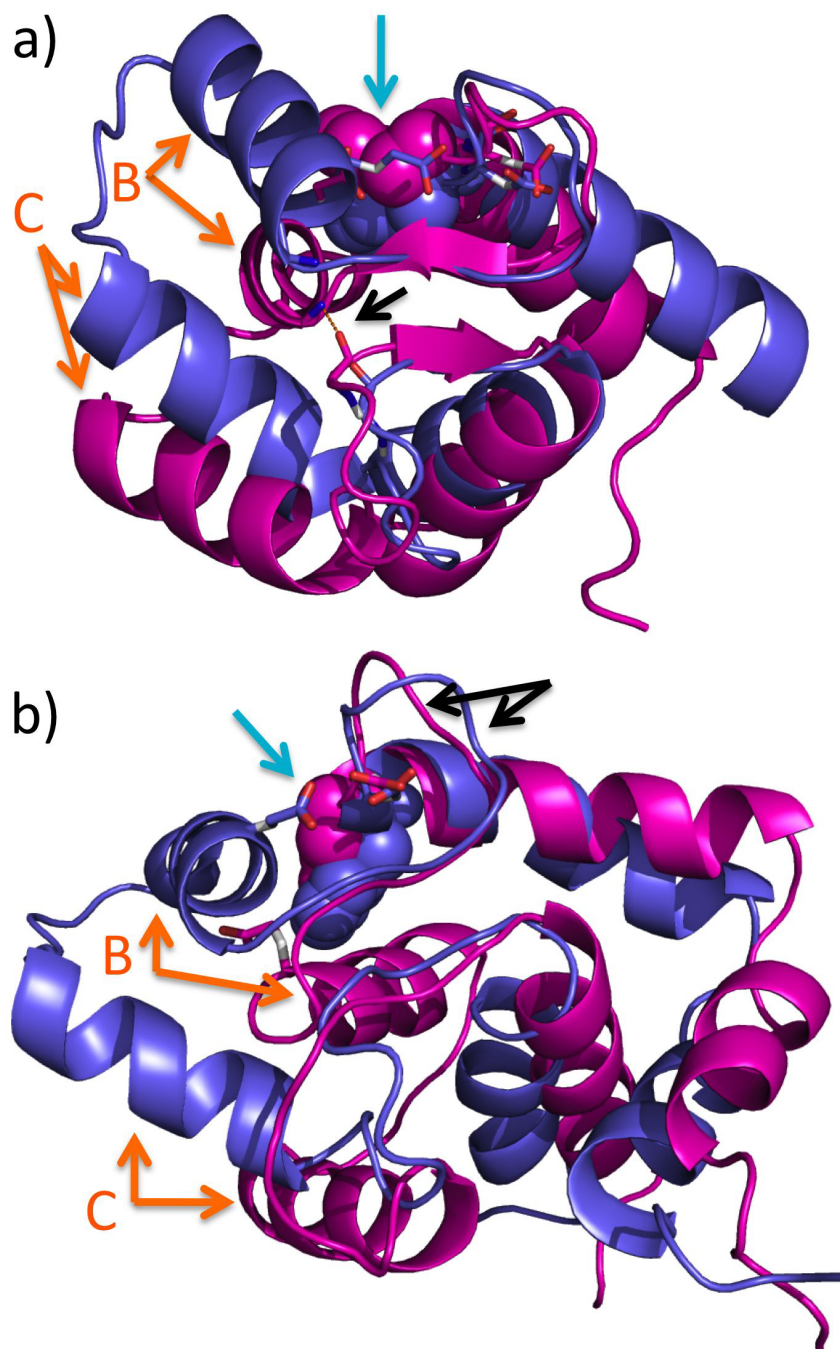


Figure 8.8 – Comparison between open and close models for both (a) CAMnt and (b) TpCnT. Open conformations are shown in blue and closed structure in magenta. The models are aligned on the stable part of their beta sheets. The orange arrows point out the variation of helices B and C relatively to their respective EF-hand partners A and D. Black arrows indicates the most variable location within or nearby the  $\beta$ -scaffold, not the same for both types of molecule. Cyan arrows point the pre-loop 1 phenylalanine, that undergoes an important relocalization relatively to helix B.

events are attempted.

At 300 K, this leads to a further 2 runs (numbers 5 and 12) that manage to evict PHE19. At 900 K, closing rate is even larger, will 11 runs showing an end conformation with the PHE19 expelled and a conformation compatible with the apo state of calmodulin, and a further 6 runs with the PHE19 in an intermediate state on the pathway to ejection. Figure 8.9 shows the evolution of the total energy and RMSD as measured from the experimental state for these 24 runs. For clarity, these are separated into four subgroups : the five runs already in apo state at the onset of the 900 K simulations, the 6 runs that expel PHE19 at 900 K and converged into the closed state, the 6 runs with PHE19 stabilized in an intermediate state and the 7 that do not fold.

As discussed above, the intermediate state can be found on the pathway of each of the 11 simulations that reach a closed state either in the original 300 K or the new prolonged runs (Fig. 8.9(a) and (b)). In this intermediate position, PHE19 of helix A is inserted within a hydrophobic pocket localized at the beginning of helix B. This pocket is formed of LEU 32, MET 36 and VAL 35 (see Fig. 8.10). This position corresponds to a metastable state favorable to a further PHE19 escape. The importance of stabilizing PHE19 is confirmed by the fact that only runs that do not position PHE19 on the outside or in an intermediate position are rapidly destabilized with an RMSD that reaches between 5 and 9 Å (Fig. 8.9(d) and Fig. 8.10(f)). Those seven simulations all show PHE19 inside the hydrophobic core and away from the pocket, correlating with an unstable behavior.

Even if the 900 K conformations are somewhat disturbed compared to those at 300 K, close states with externally positioned PHE19 are stable and never bring this residue back inside the hydrophobic core. This suggests that this movement in solution should only take place after the disruption of the  $\beta$ -scaffold by an incoming Ca ion.

Extended simulations at 300 K and 900 K were also performed with CHARMM19-EEF1. Due to the observed difficulty for forming an elongated  $\beta$ -scaffold associated with this forcefield, no new simulation manages to expel PHE19. Che14t, however, the only run reaching the apo state remained very stable at 2.5Å RMSD during the 900 K run. At 900 K, only simulation 19 manages to reach the intermediate state with PHE19 in the LEU-MET-VAL pocket. This locks the protein in a conformation near the apo state,

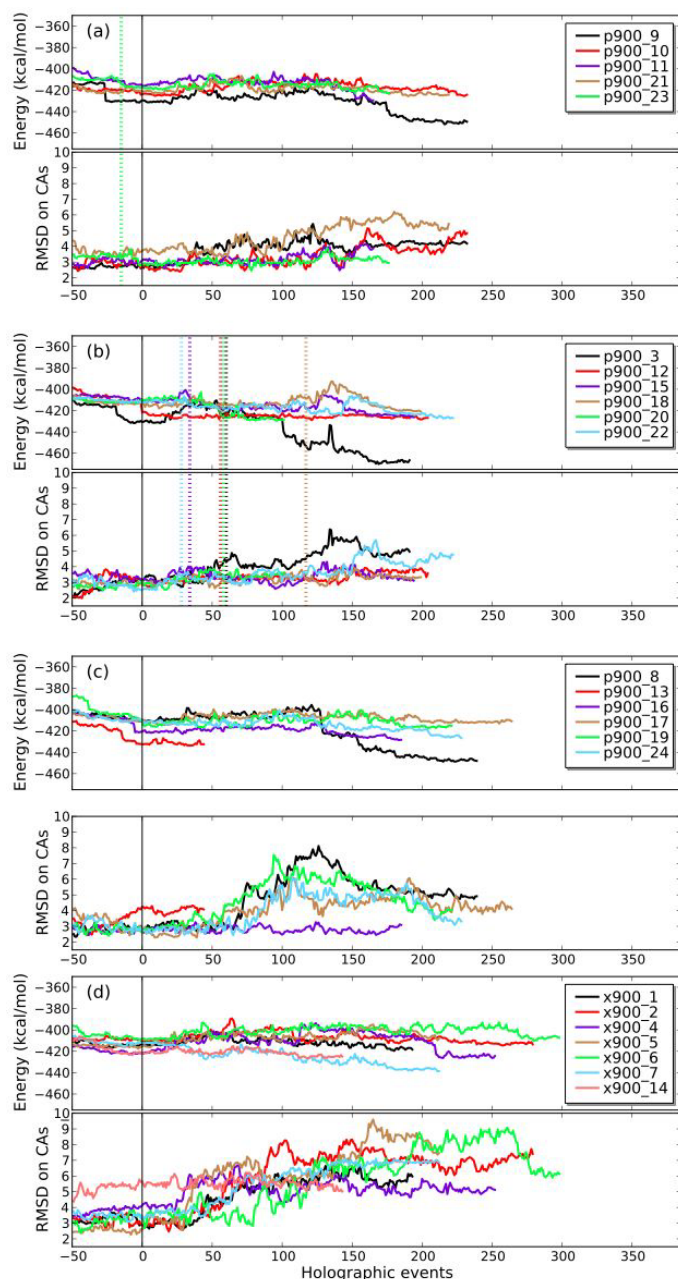


Figure 8.9 – Extension of the 24 EOPEP calmodulin NT closing simulations, using a 900 K acceptance Metropolis temperature. The vertical continuous black lines represent the beginning of the runs at this temperature following the last 50 events accepted at 300 K. The energy (top) and RMSD from the experimentally-derived apo state (bottom) are shown for the : (a) 5 simulations that had already succeeded PHE19 escape at metropolis 300 K. (b) 6 additional simulations succeeding PHE19 escape ; (c) 6 simulations with PHE19 in an intermediate state. (d) 7 simulations with destabilizing trajectory. The dashed vertical lines in (a) and (b) indicate the escape event of PHE19. The line is not shown for simulations 9, 10, 11 and 21 because this event occurred before the beginning of this graph.



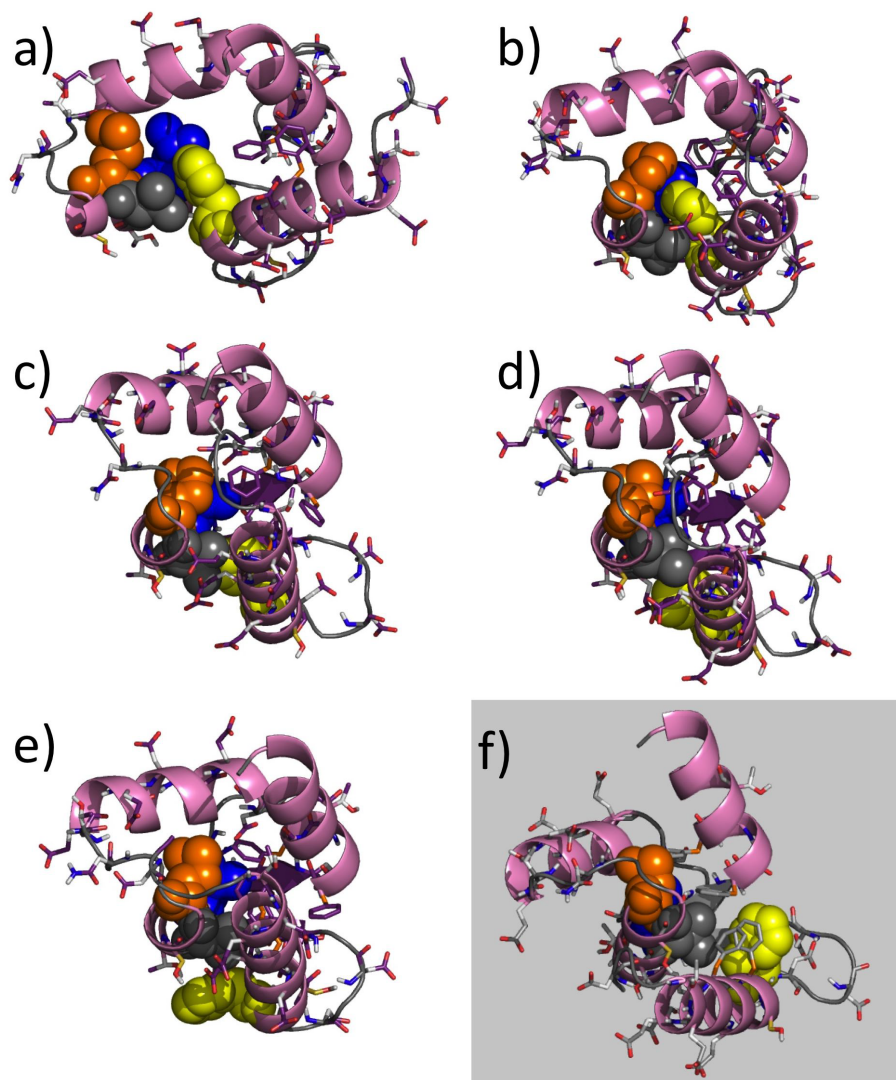


Figure 8.10 – Cartoon representation of the PHE19 as it enters and leaves the intermediate pocket. PHE19 of helix A and the tree pocket residues of helix B are represented as spheres : PHE19 in yellow, VAL35 in grey, MET36 in orange and LEU32 in blue. (a) The open calmodulin NT model. (b) While closing, the LEU-MET-VAL pocket receives PHE19. (c) and (d) Valine 35 swivels and PHE19 finds its way toward (e) exit. (f) Typical unstable structures for simulation at 900 K where the PHE19 is neither trapped in the intermediate state nor ejected outside the hydrophobic core. We see the  $\beta$ -sheet in dark grey at the back of the molecule. This structure appears in (c,d,e) in dark purple.

at 2.5–3.0 Å RMSD. This observation confirms both the importance of the  $\beta$ -scaffold and the PHE19 for stabilizing the close state.

### 8.3.2 Simulations on troponinC NT

We also performed CHARMM19-EEF1 and EOPEP simulations on the five helices of the 90-residues muscular troponin C N-terminal module. We used the first model included in the NMR PDB file 1TNQ [83] as starting point for our stability studies. We proceed following the same protocol as for CaM with ART simulations started from the apo state to assess the sequence's stability with both potentials, with the RMSD computed from the best fit of 1TNP PDB file, i.e. the 30th structure for CHARMM-EEF1 and the 3rd for EOPEP. In both cases, the apo state is stable and trajectories remain between 1.9 and 2.9 Å RMSD with CHARMM-EEF1 and between 2.3 and 3.3 Å RMSD for EOPEP $\alpha$ .

Twenty-four 100-accepted-event simulations are then launched from the unbound open structure of skeletal TnC NT with both potentials. We used the unbound first model from the PDB file 1TNQ [83] as the initial state. Table 8.I also shows the averaged displacements for TpCnT computed for a number of trajectories. Here, generated events were typically between 1.1 and 1.8 Å with, again, significantly smaller displacements for accepted events at 300 K.

Fig. 8.11 shows the four trajectories visiting, respectively, the lowest-energy structure and the smallest RMSD for CHARMM-EEF1 (top) and EOPEP (bottom). Several CHARMM19-EEF1 simulations (Fig. 8.11(a)) succeed in converging near 2.5 Å RMSD from the apo state, simulation che18t reaching 2.5 Å. As with CaM, the structure also collapses rapidly with EOPEP and the trajectories visit conformations between 1.9 and 2.3 Å from the experimentally-derived apo state within the first 20 to 30 events (EOPEP $\alpha$  simulations 6, 13 and 23, Fig. 8.11), converging to a distance between 2.5 and 3.0 Å.

Folding of troponin C follows basically the same main steps as calmodulin NT, with some nuances in the  $\beta$ -scaffold behavior. The first TpCnT EF-hand motif is preceded by a phenylalanine (here at position 29) that plays a function similar to PHE19 and PHE92 in calmodulin. Within the 24 EOPEP $\alpha$  simulations, PHE29 exit was realized



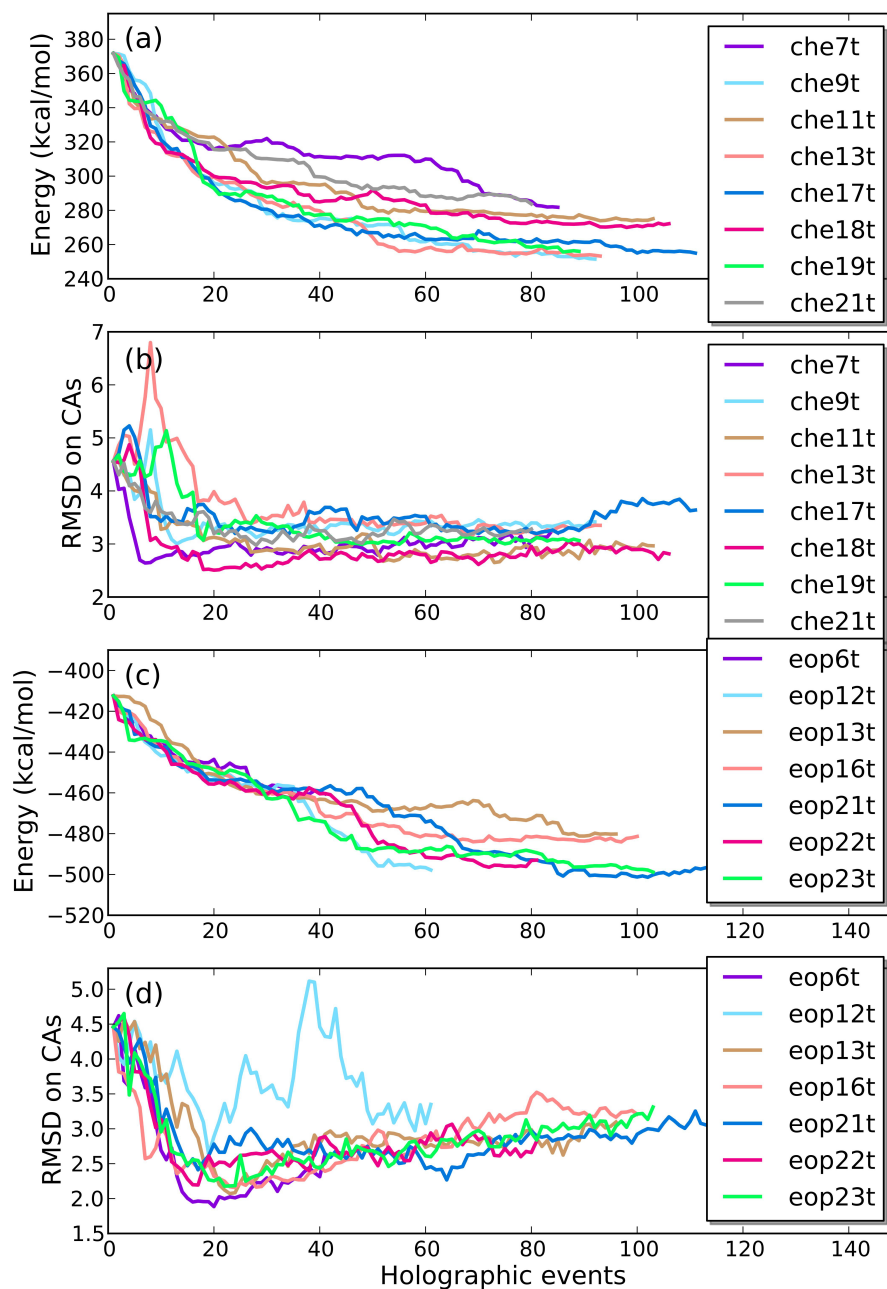


Figure 8.11 – Energy and RMSD measured from the experimentally-derived closed TPCnt structure as a function of accepted ART-event for independent simulations started from the experimentally-derived open conformation with Ca ion removed. From a set of 24 trajectories, we present the four simulations with the lowest-energy and RMSD structures, respectively. Panels (a) and (b) show the evolution of the energy and RMSD for CHARMM-EEF1 simulations, respectively. RMSD is relative to model 30 of PDB 1TNP; panels (c) and (d) present results for EOPEP simulations, with RMSD relative to model 03 of PDB 1TNP. Simulation eop22h reaches both energy and RMSD criteria.

in 8 simulations, including eop13t (events 23-31), eop21t (events 104-110) and eop22t (events 13-15), represented in Fig. 8.11(c)(d). With CHARMM19-EEF1, PHE29 exit is partially achieved by simulations che17t (events 92-93) and che18t (events 78-80). With this potential, many other simulations also show a relatively low RMSD with the apo form even though PHE19 is not expelled from the hydrophobic core.

As with calmodulin, previous to PHE29 exit, we observe modifications within the troponin C NT  $\beta$ -scaffold that sets up the structure for its eviction. As compared with CaM, the region forming the  $\beta$ -scaffold has almost the same conformation in the TpC holo and apo forms, both models showing a single hydrogen bridge, at ILE37-ILE74. However as calcium is removed, two additional hydrogen bridges are formed during the first energy minimization : ILE37-ILE73 and GLY35-PHE75. The  $\beta$ -sheet consolidates therefore from loop 1 position 8–loop 2 position 8 toward helices A and D (loop 1 position 6–loop 2 position 10) while the CaMnt  $\beta$ -sheet, which also starts from the ILE at loop 1 and loop 2 in position 8, zips in the reverse direction, towards helices B and C, involving loop 1 at position 10 and loop 2 at position 6.

In the successful EOPEP PHE29 exit simulations, the  $\beta$ -scaffold conformation change is later followed by a reconfiguration of the 3-glycine area bordering the  $\beta$ -sheet (GLY33-34-35 at loop 1, positions 4 to 6). This zone provides a higher flexibility for TpC, and in most simulations its restructuring closely precedes the exit of PHE29. This could explain why pre-loop 1 phenylalanine exit was realized in a significantly higher proportion of folding simulations for TpCnt than for CaMnt.

The LEU-MET-VAL pocket also plays a role within the troponin C Nt (residues 42,46,45), even though, in EOPEP simulations, PHE29 is less deeply engaged within it as compared with its calmodulin Nt analog. In simulations eop6t and eop16t presented Fig. 8.11(a) and (b), we observe that PHE29 is on its way to exit the pocket, positioned just under valine 45. In simulations eop12t and eop23t, PHE29 is still within the LEU-MET-VAL pocket, in the intermediated state.

In both CHARMM-EEF1 simulations that nearly succeed in expelling PHE (che17t and che18t of Fig. 8.11(a) and (b)), PHE29 is outside the LEU-MET-VAL pocket, but lies between VAL45 and helix A. In the other simulations, PHE29 stays strongly inside

the LEU-MET-VAT pocket. We observe considerable instability for the  $\beta$ -scaffold with respect to the closed form model at the THR39 ending in the troponin C Nt CHARMM-EEF1 simulations. This situation may prevent the LEU-MET-VAL pocket of helix B to gain distance from PHE29 of helix A. This points again at the importance of the  $\beta$ -scaffold formation, even though this structure is less defined in troponin C NT than in calmodulin NT.

## 8.4 Conclusions and perspectives

We simulated the folding pathway, from  $\text{Ca}^{2+}$ -unbound open holo conformations to closed apo states, of two well-studied EF-hand : the NT domains of calmodulin and troponin C. Trajectories are generated using an activated algorithm, ART nouveau, along with the multi-scale holographic method, which lead to an unbiased sampling of the conformation space with a focus on the flexible regions. This allows us to complement recent numerical work using Gō-like biased forcefield approaches.

Our results are compatible with the EF-hand  $\beta$ -sheet scaffold model proposed by Grabarek which explains the apo to holo conformation change [93, 94]. Here we provide a complete, four-stage, picture of the reverse motion, or folding from holo toward apo conformation : (1) the removal of the  $\text{Ca}^{2+}$  atom, forces the side-chain oxygen atom to repel each other ; (2) upon release of the oxygen carrier residues, the  $\beta$ -sheet scaffold linking the 2 binding loops is able to consolidate, widening out the loops area ; (3) this allows helix B swivel towards the inner side of PHE19 (CaMnt) or PHE29 (TpCnt), expelling the residue outside the hydrophobic core after a variable time spent for the PHE19 or PHY29 into an intermediate state ; (4) this motion leads to a tighter packing of helices B and C toward helices A and D, optimizing and stabilizing the closing of the hydrophobic core.

There is much interest in the reverse transition, where  $\text{Ca}^{2+}$  ions bind to the two loops. This transition may be favored by the large apo fluctuations shown by experimentalists. But the oxygen carrier residues repulsion is likely to prevent the tightening of the binding loops and, consequently, a significant occurrence of the typical open conforma-

tion. Thus, the  $\beta$ -sheet scaffold may act as a switch for both conformational direction changes. The closed form is stabilized by the  $\beta$ -sheet consolidation/elongation tendency along with the hydrophobic attractions between helices. The open form is a consequence of the  $\beta$ -sheet disruption by oxygenated side chains attraction toward calcium, leading to spatial constraints locking open the first EF-hand pair. This description of the  $\beta$ -sheet scaffold conformational switch activated by calcium is also supported by the inability of  $Mg^{2+}$  to bring up the open conformation neither for calmodulin or troponin C [95].

The expulsion of the phenylalanine at the end of helix A appears to be modulated by the presence of a LEU-MET-VAL pocket at the beginning of helix B. The valine acts as a transitional door by which the phenylalanine goes out. The  $\beta$ -scaffold elongation favors a slight rotation of helix B (Fig. 8.10(b)), which allow the full expulsion of PHE19 or PHE29. Due to the high stability of this state, calcium binding should be necessary to induce the reverse tendency and thus favor reentrance of the phenylalanine.

The holographic approach used in this work has proven able to compute a class of ensemble motions, suitable for the calmodulin and troponin C NT conformations. In the future, we intend to apply this method to larger systems where the multiscaling approach is essential. Here, it allows us to identify a very detailed picture of the atomic scale steps crucial for the large conformational transition observed in EF-hands.

## 8.5 Acknowledgements

This work was funded in part by the BiT bioinformatic scholarship program, FQRNT, NSERC and the Canada Research Chair Foundation. Calculations were done using resources from Calcul Québec.

## CHAPITRE 9

### DISCUSSION ET CONCLUSION

Au chapitre 3 nous avons vu comment de nombreux travaux ont permis de discerner les lignes importantes du comportement des protéines au travers de leur apparente complexité. Les réseaux élastiques et le *NMA* ont aidé à faire le discernement entre les échelles de temps des mécanismes moléculaires [30, 104, 259]. Le regroupement d'atomes a permis de simuler des processus autrement inaccessibles [51, 89, 113, 128, 148]. L'étude des protéines en ciblant leurs échelles de temps plus globales est une étape d'importance majeure pour l'avancement de nos connaissances.

À partir de là, des stratégies visant la conquête des interactions plus complexes ont pu être mises en place. Les efforts plus récents axés sur les passages de transition [31, 167] ou l'analyse de zones spécifiques du parcours [19, 206, 207, 245] permettent d'aborder la rugosité des paysages énergétiques. Alors que la tendance était à la simplification, à partir des bases ainsi acquises, une tendance à intégrer des informations plus détaillées se dessine. Par exemple, les techniques de *NMA* ont d'abord dénudé le modèle protéique, faisant ressortir l'importance de la conformation par rapport aux spécificités en ce qui concerne les mouvements harmoniques [159]. Mais là aussi on cherche maintenant à remettre des informations jugées pertinentes. Par exemple, Sanejouand préconise d'utiliser le modèle de solvation *EEF1* dans la modélisation *NMA* [225].

Dans ce courant stratégique, les travaux de la présente thèse ont consisté à placer les échelles de temps dans l'ordre hiérarchique dit "*top-down*" au lieu de "*bottom-up*". Nous avons fait l'inverse du processus traditionnel qui consiste à parcourir le temps moléculaire en cumulant des processus locaux de fréquences élevées, jusqu'à ce que la probabilité des fluctuations permettent l'évènement de plus basse fréquence. L'accent a été mis sur les mouvements d'envergure que peuvent amener les basculements possibles, mais en conservant la capacité d'éléments spécifiques d'influencer le cours de plus gros ensembles. Nous l'avons vu au chapitre 8 avec l'exemple du déplacement de la phénylalanine 19 de la calmoduline, qui permet une fermeture plus stable du domaine.

Nous avons assuré la finesse de la représentation spatiale à tout moment des passages de transition, avec en plus une adaptation fine lors de la relaxation après chaque transition.

Plus nous avons de niveaux moléculaires à étudier, plus les rapports de temps entre les échelles génèrent des temps de simulation impossibles. La méthode *HMM* ne dépend pas de la survenue d'un événement rare à un niveau pour permettre un basculement à un autre niveau. Elle génère les différents basculements possibles entre les parties, qu'ils soient fréquents ou non. C'est alors le rôle du potentiel atomistique utilisé d'évaluer l'avantage énergétique de chaque transformation proposée. La méthode *HMM* s'est avérée efficace dans la génération de conformations de plus en plus compactes et basses en énergie.

Cependant les grands mouvements ne s'effectuent pas nécessairement à fréquence lente. Nous devons distinguer entre les vibrations lentes des protéines, qui sont liées aux fluctuations permises par la structure, et les mécanismes coopératifs, qui sont rapides. Les processus de repliement en deux phases principales, une rapide et une longue, sont observés dans les protéines [47, 72, 97, 306]. Dans la première phase qui est rapide, la constitution du globule fondu implique des interactions distantes. Tandis que la deuxième phase est lente, avec les molécules d'eau emprisonnées qui se dégagent et les structures fines qui s'ajustent.

Les simulations de fermeture de domaines rapportées aux chapitres 7 et 8 (Fig. 7.9, 7.10, 8.3 et 8.11) mettent en évidence la capacité de la méthode *HMM* de simuler deux phases : dans les premières heures des simulations nous observons d'abord un changement rapide dans les structures, suivi d'un plateau. Les ajustements sont alors plus locaux, mais l'énergie continue de descendre. Nous pouvons considérer la première phase comme davantage cinétique, alors que différents mouvements sont possibles. Il devient ensuite plus difficile de générer des mouvements amples conduisant à des évaluations de moindre énergie lorsque l'on parvient à un certain entassement moléculaire. Les basculements des chaînes latérales interviennent alors pour affiner les ajustements locaux.

## 9.1 Bilan pour les Mains EF

En ce qui concernent le mécanisme des mains EF, la méthode a permis d'explorer certains phénomènes du trajet de fermeture. Pour la troponine C comme pour la calmoduline, les deux boucles liant le calcium se longent en leur centre et se lient par des ponts hydrogènes. Cette zone forme un feuillet  $\beta$  plus ou moins bien défini selon la présence ou l'absence de l'ion calcium. L'attraction des oxygènes des chaînes latérales et de la chaîne principale par le calcium tend à perturber ce feuillet. Au contraire, les ponts hydrogènes de ce feuillet se consolident en l'absence de calcium. Cette zone initierait donc autant le mécanisme d'ouverture, selon Grabarek [93, 94], que le mécanisme de fermeture du domaine, comme le démontrent nos simulations (chapitre 8). Dans le cas du mécanisme de fermeture, la restructuration du feuillet  $\beta$  entraîne un évasement de la zone des boucles, facilité par la répulsion entre les chaînes oxygénées, qui n'est plus contrebalancée par l'ion positif. Si la contraction de la base des mains EF favorise l'ouverture du domaine, son évasement permet au domaine de se refermer.

Le mouvement de fermeture amène la phénylalanine terminant l'hélice A du domaine NT (phe19 dans le cas de la calmoduline, phe29 pour la troponine C) à l'intérieur d'une poche hydrophobe (formée par leu32-val35-met36 dans le cas de la calmoduline, par leu42-val45-met46 dans le cas de la troponine C). Cet état intermédiaire est présent sur le chemin de toutes les simulations qui ont complété la transformation (voir section 8.3.1.8). Pour évaluer la complétion, nous avons utilisé le RMSD (avec les  $C\alpha$  de la conformation apo) mais aussi considéré comme critères obligatoires la consolidation du feuillet  $\beta$  ainsi que la position de phe19/29 à l'extérieur du domaine. L'ensemble des simulations démontrent la nécessité de la formation  $\beta$  avant que la sortie de la phénylalanine ne puisse survenir. Après un passage plus ou moins prolongé dans la poche hydrophobe LMV, la phénylalanine effectue sa sortie en transitant par la valine. Le pivotement de leur chaîne latérale respective favorise cette transition.

Nous avons constaté la capacité de la méthode à produire un chemin complet pour la fermeture du module NT de la calmoduline, avec 11/24 simulations, et 6 simulations sur le même chemin mais non complétées, car dans l'état intermédiaire avec la phé-

nylalanine 19 sise dans la pochette leu-met-val. Nos simulations ne présentent pas la transition inverse où la phénylalanine réussirait à revenir dans la poche intermédiaire. Ceci suggère la nécessité d'une déstructuration du feuillet  $\beta$  pour permettre le retour de la phénylalanine dans le coeur hydrophobe, et donc d'une induction par l'ion calcium.

En plus de permettre la proposition d'un enchaînement d'événements pour le mécanisme de fermeture, les simulations mettent en relief la coopérativité de certains éléments. La figure 8.6 (c) montre que la fermeture du domaine N-terminal de la calmoduline implique la coopérativité entre plusieurs résidus. Nous avons généré des fichiers contenant les statistiques énergétiques de chaque résidu au fur et à mesure des simulations. En particulier, nous observons que le positionnement extérieur de phe19 va de pair avec une montée en énergie de ce résidu, permettant une baisse de l'énergie de plusieurs résidus du coeur hydrophobe.

Les protéines troponine C et calmoduline sont de proches cousines dans l'arbre phylogénétique [124]. Leur domaine NT comporte cependant 30% de différence de séquence et il est intéressant de voir si cela a une répercussion au niveau du mécanisme. Nous avons observé en fait une plus grande facilité pour la sortie de la phe29 de TPC NT que pour la sortie de phe19 de CAM NT. L'observation des films des simulations suggère que cela est imputable à une plus grande flexibilité de la bouche 1 avec 3 glycines consécutives en position 3-4-5 pour la troponine C par rapport à gly-asp-gly pour la calmoduline (voir Fig. 8.1). La flexibilité de cette zone libère un peu plus d'espace pour le passage de phe29.

### 9.1.1 Travaux futurs pour les mains EF

Une prochaine étape serait de simuler les mains-EF en relation avec leurs protéines cibles. Cela nécessite d'étendre la méthode holographique à la simulation de complexes. Cela requiert aussi la maîtrise des ions, dans ce cas-ci le calcium. Un modèle de l'ion  $\text{Ca}^{2+}$  avait été développé dans le cadre du développement du potentiel *EOPEP*. Ce modèle de calcium avait été ajouté au modèle fermé et certaines simulations présentaient le chemin suivant : resserrement des boucles et déstructuration du feuillet bêta, retour de phe19 à l'intérieur du coeur hydrophobe en transitant par val35, leur chaîne latérale



basculant l'une par rapport à l'autre. La figure 9.1 montre un événement illustrant ce passage, qui mène à une plus ample ouverture du domaine. Le modèle doit être perfectionné car beaucoup de simulations ne menaient pas à l'ouverture.

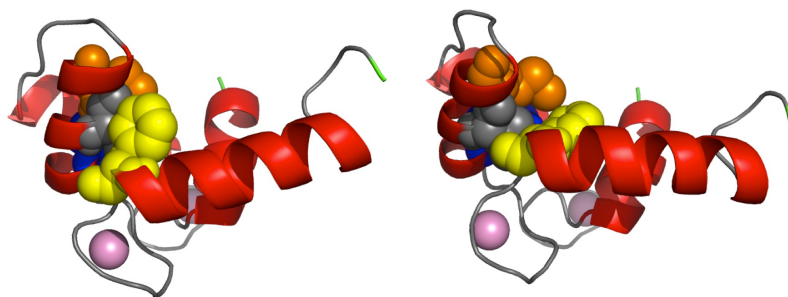


Figure 9.1 – Étapes 174 (à gauche) et 175 (à droite) d'une simulation avec ajout du calcium au modèle fermé du domaine NT de la Calmoduline. En présence du modèle calcium, le feuillet  $\beta$  est perturbé et la zone des boucles s'est rétrécie. À l'étape 174, le domaine est partiellement ouvert et phe19 est encore à l'extérieur du coeur hydrophobe. À l'étape suivante, phe19 est entré dans le coeur hydrophobe et le domaine a ouvert davantage. Dans les deux figures nous pouvons observer la valine, de couleur grise, qui agit comme porte. Seulement certaines chaînes latérales sont illustrées ici pour plus de clarté, et l'ion calcium est représenté en rose.

### 9.1.2 Validations expérimentales

Nos simulations informatiques appuient l'idée que les domaines mains-EF étudiés suivent des trajectoires précises, guidés par des composants aux positions stratégiques, au lieu suivre une diversité de chemins. Pour valider nos résultats informatiques nous les avons comparés avec les observations expérimentales. Nous espérons aussi apporter de précisions supplémentaires. On peut ensuite s'interroger dans quelle mesure certains détails des simulations peuvent être à leur tour vérifiés expérimentalement. Selon un point de vue le rôle des simulations est d'apporter des précisions inaccessibles au niveau expérimental. Mais de telles précisions sont elles utiles si elles ne peuvent pas guider la recherche expérimentale vers de nouvelles avenues ? Dans cette optique nous devons examiner les implications expérimentales des hypothèses amenées par les simulations.

Peut-on capturer expérimentalement un détail précis observé dans les simulations ? Peut-on observer expérimentalement les effets à plus grandes échelles d'une modification précise qui seraient prédits en simulation ?

Pourrions nous par exemple vérifier par mutagenèse l'importance de la consolidation du feuillet  $\beta$  dans le mécanisme de la calmoduline ainsi que celle de la val35 comme porte de passage pour phe19 ? Dans ces deux cas, il serait indiqué de générer des simulations des états mutés afin d'avoir une image préalable de ce qui pourrait subvenir expérimentalement. Il serait difficile de vérifier expérimentalement la présence d'un état intermédiaire dans la trajectoire de fermeture car il faudrait partir de l'état ouvert. Avec l'expression *in vitro* nous obtenons les molécules natives et donc les domaines NT et CT fermés. Une mutation dans le but de stabiliser l'état intermédiaire proposé pour la trajectoire de fermeture risque d'empêcher d'atteindre l'état ouvert préalable à cette trajectoire. En ce qui concerne val35, la prédiction serait que sa mutation vers une chaîne plus encombrante stabiliserait l'état intermédiaire de la trajectoire de fermeture mais aussi l'état natif fermé, empêchant l'entrée de phe19 dans le coeur hydrophobe. À l'inverse, une mutation de val35 pour une chaîne plus légère pourrait désorganiser l'état fermé ou l'état ouvert, mais d'une manière difficilement prévisible. Des simulations avec cette mutation pourraient cependant donner des indices.

Selon notre analyse des simulations de fermeture du domaine NT de la calmoduline l'élongation du feuillet  $\beta$  permet l'évasement des boucles et rend possible la sortie de phe19. Encore ici, nous devons partir de l'état ouvert pour le vérifier. Par contre, nous pourrions partir de l'état fermé mais privilégier l'état ouvert en affaiblissant le pont-hydrogène thr29N-gly61O qui consolide le feuillet  $\beta$ . Nous pouvons présumer que le domaine s'ouvrirait à l'addition de l'ion calcium, mais aurait de la difficulté à revenir à la conformation ouverte. Une première vérification concernerait la qualité de l'ouverture du domaine. Une activation normale de l'enzyme cible est indicatrice d'une position correcte de phe19 à l'intérieur du coeur hydrophobe [255]. Selon nos simulations, au retrait du calcium, l'incapacité d'allonger correctement le feuillet  $\beta$  devrait entraver la sortie de phe19 hors du coeur hydrophobe. Pour empêcher cette élongation, les résidus thr29 et/ou gly61 pourrait être mutés pour des résidus plus encombrants entravant leur

contact au niveau de la chaîne principale.

En résumé, si l'on considère que les calmodulines exprimées expérimentalement présentent leurs domaines mains-EF initialement fermés, une mutation de val35 en un résidu plus encombrant telle la phénylalanine stabiliserait l'état fermé du domaine NT. Par contre une mutation de gly61 en un résidu plus encombrant, telle la tyrosine, permettrait éventuellement l'ouverture du domaine NT en présence de calcium, et stabiliserait peut-être l'état ouvert par la suite. Pour en faire la vérification, nous devons donc pouvoir tester la qualité de l'état ouvert et de l'état fermé du domaine NT de la calmoduline. La vérification de l'état ouvert pourrait être réalisée par des tests d'efficacité de liaison à une enzyme. Le blocage dans l'état fermé pourrait être vérifié par des tests d'ellipticité (à l'aide du dichroïsme circulaire) [177, 188] comparés entre les domaines sauvages fermés et ouverts et mutant G61T.

Tan, Mabuchi et Grabarek ont observé que l'inhibition d'un domaine ou de l'autre de la calmoduline était suffisante pour entraver l'efficacité de l'enzyme phosphodiestérase [255]. Pour tester l'efficacité de la liaison à cet enzyme, on peut donc cloner la calmoduline complète et la sur-exprimer. Les tests d'efficacité de l'enzyme nous permettrait ensuite d'évaluer la qualité de l'état ouvert du domaine NT de la calmoduline.

Liste des démarches expérimentales :

1. Obtenir l'ADN complémentaire contenant la séquence entière de la calmoduline (foie humain ou de grenouille par exemple).
2. Mutagénèse. En plus du type sauvage, obtenir le mutant G61Y pour empêcher l'élongation du feuillet  $\beta$  : changer le codon pour TAC de la tyrosine. Obtenir aussi le mutant V35F pour entraver l'entrée de phe19 dans le coeur hydrophobe : changer le codon pour TTT de la phénylalanine.
3. Clonage. Couper les séquences cDNA sauvage et mutées avec les enzymes de restrictions KpnI et PstI. Sur-exprimer avec la réaction en chaîne par polymérase [183]. Vérifier les trois chaînes par séquençage avec didésoxynucléotide [226]. Pour chacune des trois variantes, insérer dans un vecteur d'expression tel PKK233-2 et intégrer à une culture bactérienne *Escherichia coli* [255].

4. Purification des trois variantes de la protéine (sauvage V35F et G61T).
5. Il faut s'assurer que les mutants se soient formés selon une structure comparable au type sauvage. On pourrait utiliser le dichroïsme circulaire pour comparer les conformations [177, 188].
6. Vérification de l'exposition au solvant.
7. Vérification de l'ouverture du domaine. Pour le type sauvage comme pour les deux mutants, tester l'efficacité de l'enzyme associée au fur et à mesure que l'ion calcium est ajouté. Pour les deux mutants on compare la courbe d'efficacité de l'enzyme avec celle du type sauvage. Nous escomptons que la qualité de l'ouverture et de la liaison du mutant G61Y demeure assez bonne. Nous escomptons que l'ouverture soit compromise pour le mutant V35F.
8. Vérification du blocage de l'état ouvert. On effectue ensuite un retrait graduel du calcium pour le mutant G61Y et pour le type sauvage afin de comparer leur diminution de liaison à l'enzyme. Nous escomptons que G61Y demeure plus ou moins ouvert, la sortie de phe19 étant rendue plus difficile.
9. Vérification de l'affinité au calcium. Pour bien discerner l'effet des mutations sur la liaison au calcium de l'effet des mutations sur l'ouverture du domaine, on doit aussi évaluer la courbe d'affinité au calcium.

## 9.2 Bilan pour ART holographique

Les succès obtenus proviennent des stratégies suivantes :

1. Génération de mouvements d'ensemble qui précèdent les ajustements locaux.
2. Utilisation de la précision spatiale pour des mouvements de grande amplitude.
3. Pour se rendre plus facilement à bon port, en plus des calculs atomistiques détaillés, on ajoute des calculs holographiques pour guider les mouvements d'ensemble.

4. Adaptabilité de la représentation multi-échelle, au fur et à mesure de l'évolution des forces.
5. Capacité d'évaluer l'effet au niveau de la trajectoire de faibles différences de séquence entre certaines macro-molécules. Ceci pourrait aussi s'appliquer à l'évaluation des mutations et des modifications post-traductionnelles.

La méthode est à ses débuts et certains problèmes sont observés :

1. Dépendance aux champs de force, qui peuvent guider vers des conformations éloignées de la conformation cible connue grâce aux méthodes expérimentales.
2. Incapacité de générer des transformations dans les boucles intercalées entre deux zones contraintes spatialement.
3. Les algorithmes ont besoin d'être perfectionnés pour assurer la sortie des états de stagnation dans les minimums locaux lorsqu'on a atteint une condition compacte.
4. Les simulations n'ont pas pour l'instant été générées en assez grand nombre pour que l'on puisse évaluer la probabilité des chemins observés.

### 9.2.1 Échantillonnage

Après un certain nombre d'événements apportant une restructuration rapide, nous observons que certaines simulations mènent à des conformations différentes du modèle cible principal. Bien sûr, il y a toujours la question de savoir si les trajectoires d'une protéine mènent à des conformations diverses d'énergie similaire à la protéine native ou plutôt si les différents chemins conduisent vers la même conformation finale [47]. Mais pour le bien des processus moléculaires l'optimisation de l'énergie devrait favoriser la conformation la plus courante. Les résultats finaux de plus basse énergie sont dépendants de la justesse des champs de force utilisés.

Le déroulement des simulations des mains-EF semble très lié à la réussite de la fusion initiale. Une tendance aux séjours dans des conformations alternatives est illustrée par les différents plateaux des graphiques RMSD. Aux alentours de l'événement 240, la

simulation eop16a (figure 8.3 (c) et (d)) a été menée à plus haute température de Métropolis, provoquant la réouverture des domaines, puis une meilleure fermeture après retour à une température d'acceptation de 300 K vers l'événement 405.

Pour améliorer l'échantillonnage, deux points sont à considérer. Les chemins possibles sont considérablement réduits si nous utilisons un potentiel atomistique au lieu d'un potentiel gros grain. Par rapport au potentiel gros grain OPEP d'origine, le développement de EOPEP a considérablement réduit la survenue des conformations alternatives. Ceci a été observé au fur et à mesure que furent ajoutées des interactions importantes pour les mains EF : ponts hydrogènes entre chaînes latérales et chaîne principale, représentation spatiale affinée facilitant l'imbrication complexe propre aux coeurs hydrophobes des mains EF, discernement entre les forces attractives et répulsives au sein des résidus, tels la glutamine et l'asparagine par exemple. L'utilisation de champs de forces de plus en plus précis peut améliorer les simulations de certains cas de transformation complexe.

Deuxièmement nous pouvons aussi réduire le problème des trappes liées à la forme du paysage énergétique. Les conformations alternatives, qu'elles soient réalistes ou non, peuvent présenter des zones de descente énergétique menant à des minima locaux d'où il peut être plus difficile de remonter. Des ajustements fins peuvent en outre se produire, conduisant à une évaluation énergétique moindre. Le mouvement de sortie qui suit alors, n'ayant pas bénéficié de ces ajustements fins, aura une évaluation plus élevée et sera rejeté. Lors de l'étude de la fermeture de la calmoduline, les caractérisations supplémentaires à plus haute énergie d'acceptation (900K) étaient donc appropriées pour évaluer l'habilité de la méthode à résoudre ces situations. Un algorithme détectant une stagnation dans un plateau serait souhaitable pour réactiver la simulation.

Il serait aussi possible de construire une image des états possibles et des chemins entre ces états. La méthode permet la génération de fichiers .REL (positions relatives des plans peptidiques et des blocs des chaînes latérales) pour chaque événement. Ces données prennent peu d'espace et pourraient être gardées en mémoire. Elles seraient classées hiérarchiquement, permettant de retrouver rapidement le plus proche voisin. Pour retrouver rapidement la conformation la plus proche, la hiérarchie de comparaison

serait la suivante : comparaison avec les positions liées aux pivots centraux seulement, comparaison avec les autres pivots, puis en considérant aussi la position des blocs latéraux. La figure 9.2 montre un exemple plus simple de classification, construite d'après les états principaux tirés des simulations de la fermeture de la calmoduline (chapitre 8). L'étude de ces états avait permis de déterminer le chemin pris par les simulations qui avaient réussi à atteindre les critères importants de la configuration cible.

On peut aussi s'inspirer des techniques d'échange de répliques pour diversifier l'échantillonnage. Des simulations peuvent être conduites selon une plus haute température de Métropolis. Il se présente alors des cas aléatoires où l'énergie est relativement basse, fournissant un minimum local qui n'aurait peut-être pas été échantillonné autrement.

### **9.2.2 Transformations dans les boucles avec extrémités fixes**

La génération de transformations entre zones contraintes reste un point majeur à améliorer. Une compréhension plus étendue des différents types de mouvements d'ensemble susceptibles de survenir dans une protéine est nécessaire. On doit inventorier les différents types de processus coopératifs possibles et développer des représentations à partir desquelles on peut les générer facilement. Ces transformations coopératives ne sont pas seulement des stratégies informatiques visant à détecter les mouvements d'envergure possibles. Des mouvements coordonnés instantanés surviennent au sein des macromolécules. Le déplacement d'un simple élément dans un plan de l'espace 3D peut signifier le basculement d'une large région dans un autre plan de cet espace. Les transformations peuvent aussi consister en des basculements internes à une zone, en des étirements ou en des contractions. Ces derniers types de transformation peuvent s'effectuer entre zones contraintes et doivent donc être implantés dans la méthode.

### **9.2.3 Développement des potentiels**

La convergence des simulations *ART nouveau* - *HMM* vers la cible recherchée est dépendante de la justesse des potentiels atomiques non-biaisés utilisés. La possibilité de générer des trajectoires non biaisées, alors que l'on connaît la destination souhai-

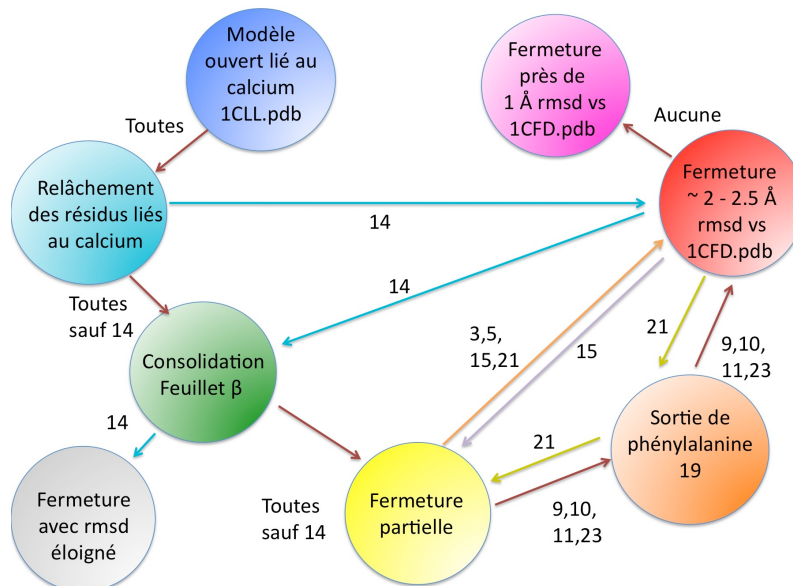


Figure 9.2 – Étapes principales de 24 simulations *HMM-EOP* pour la fermeture de la calmoduline. L'enchaînement optimal des événements serait celui des simulations 9, 10, 11 et 21 car la sortie de phe19 permettrait la stabilité de la conformation fermée.

table, s'avère un moyen de tester de manière sévère les potentiels atomistiques. Toute faiblesse d'une représentation des forces et énergies conduit rapidement à des conformations non réalistes, alors même que la simulation multi-échelle génère des descentes énergétiques efficaces, mais selon un paysage énergétique erroné. Il n'est plus question ici de calibrer un potentiel en utilisant des déformations relativement locales d'un éventail de molécules. Cette stratégie ne permet que d'étalonner les représentations des états moléculaires stables. Nous en sommes ici à générer des simulations permettant de discerner en quelques heures les capacités et incapacités d'un potentiel à bien représenter non seulement les états moléculaires stables, mais aussi les états transitoires entre des états éloignés.

### 9.3 Perspectives d'application

Si l'on classe la méthode selon son domaine d'expertise, nous dirions d'abord qu'elle vise les phénomènes non-harmoniques, qui comportent des états intermédiaires, en par-



ticulier à cause de son utilisation en conjonction avec *ART nouveau*. La méthode s'applique plus particulièrement aux changements d'envergure, bien qu'elle permette de discerner le rôle d'éléments de niveau atomistique. Elle comporte un avantage sur les approximations gros grain habituelles, car elle effectue les déplacements d'ensemble selon des champs de force précis et actualisés à tout moment, au lieu d'appliquer aux ensembles des forces prédéterminées et non adaptatives. Mais la méthode est à ses débuts et son utilisation est encore limitée. Pour une utilisation plus universelle, il faut corriger son incapacité à générer des transformations entre zones contraintes.

L'application de la méthode *HMM* aux domaines des protéines calmoduline, tropoline C, protéine A et protéine G démontre qu'il est possible de rejoindre des conformations distantes avec des simulations non biaisées. Pour ces protéines, la méthode a permis de discerner des processus non-harmoniques, ainsi que coopératifs, analysés grâce aux statistiques de chaque résidu (voir figure 8.6).

### 9.3.1 Mutagenèse et modifications post-traductionnelles

Les premiers essais sur la calmoduline, qui se sont faits en absence de la définition de ponts hydrogènes entre les chaînes latérales et la chaîne principale, ont démontré l'importance des résidus oxygénés non seulement pour la coordination du calcium, mais aussi pour la stabilité de la conformation non liée (voir section 6.4). Ceci met en relief le potentiel de la méthode pour tester différentes hypothèses de mutagenèse. Qu'advierait-il par exemple si l'on mutait *in silico* phe19 de la calmoduline en une simple alanine ? Nous pourrions peut-être reproduire les grandes lignes des constatations faites par Meyer, Mabuchi et Grabarek au sujet de cette mutation [177]. Il peut être avantageux d'effectuer des tests informatiques préliminaires avec *ART* holographique avant de s'engager dans la genèse de protéines mutantes. Si l'on est limité dans le nombre de mutations à générer et évaluer, les trajectoires obtenues en simulation pourrait aider à ordonner les mutations selon leur intérêt.

Les modifications post-traductionnelles jouent un grand rôle dans le fonctionnement des protéines et des acides nucléiques. La phosphorylation des résidus sérine, thréonine ou tyrosine permet de désactiver ou d'activer diverses protéines [44]. La glycolisa-

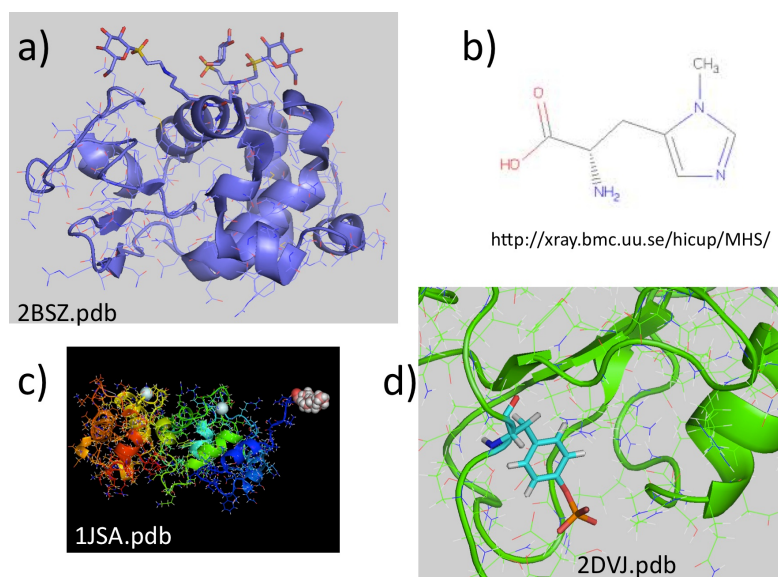


Figure 9.3 – Exemples de modifications post-traductionnelles. a) Glycolisation de deux résidus lysine. b) Méthylation de l’histidine. c) myristoylation du C-terminal de la recoverin. d) Phosphorylation d’un résidu tyrosine. Images générées avec pymol.

tion permet l’adhésion entre les cellules [282, 284]. Les histones qui contrôlent l’ADN sont aussi régulées par des modifications post-traductionnelles, ainsi que l’ADN lui-même [62]. La figure 9.3 donne quelques exemples de modifications post-traductionnelles. Comme c’est le cas pour la mutagenèse, ces modifications sont simples à modéliser *in silico* et pourrait donner lieu à des simulations comparatives.

### 9.3.2 Diversité moléculaire

La maîtrise des différentes sortes de molécules n’est qu’une question d’ajout de définitions structurales (atomes, liens, angles de valence et de torsions) et d’ajouts de types d’atomes aux potentiels. En ce qui concerne leur manipulation au niveau holographique, il convient de ne sélectionner que les pivots les plus flexibles pour définir des basculements de plans relatifs. Le cas le plus difficile serait celui des ADN et ARN. Leur chaîne comporte beaucoup plus de pivots flexibles que celles des protéines. Par contre les bases latérales ont une certaine rigidité et pivotent autour du C1 du ribose.

### 9.3.3 Complexes moléculaires

Le travail de doctorat s'est concentré sur le fonctionnement interne de protéines individuelles, mais l'idée est réutilisable pour d'autres échelles d'organisation moléculaire. La représentation en coordonnées sphériques est flexible et peut aussi bien accommoder un assemblage de molécules que l'on déplace comme des blocs rotatifs, qu'un assemblage des parties d'une seule protéine. Le méthode *HMM* permet de faire basculer les régions internes à une molécule comme des blocs dans l'espace. Ces habilités peuvent être appliquées aux déplacements de protéines dans un assemblage (fibres et complexes moléculaires) car chaque molécule peut être représentée comme un bloc mu par les forces atomistiques projetées sur ses axes de rotation.

Une première étape accessible serait d'abord l'application de la méthode à des études de dimérisation. Levy, Wolynes et Onuchic ont étudié trois sortes de dimérisation : obligatoire, induite et non-obligatoire [151]. Lors de la dimérisation non-obligatoire, deux molécules s'assemblent ultérieurement à leur repliement individuel. Le repliement des dimères obligatoires survient suite à l'association des deux molécules. Pour le troisième cas de figure, l'une des protéines a déjà atteint son repliement natif et induit le repliement de la seconde (voir Fig. 9.4). Les capacités de ART holographique seraient à évaluer pour la simulation de ces trois scénarios. En parallèle de la transformation individuelle des protéines, on pourrait aussi faire basculer celles-ci comme des blocs dans l'espace.

Des ajustements aux potentiels à solvant implicite, ou l'utilisation de potentiel à solvants explicites, permettrait la simulation de protéines membranaires. Elles sont souvent composées d'hélices  $\alpha$  dont certaines basculent par rapport aux autres [26, 116, 156, 220], d'où l'intérêt de les étudier avec ART holographique.

### 9.3.4 Repliement

La variété des protéines va de pair avec une diversité des scénarios de repliement. Comme nous l'avons vu lors des tests des protéines A et G ouvertes, ART holographique peut effectuer le repliement dans le cas où les structures secondaires sont déjà formées. Mais bien souvent les structures secondaires sont induites consécutivement à

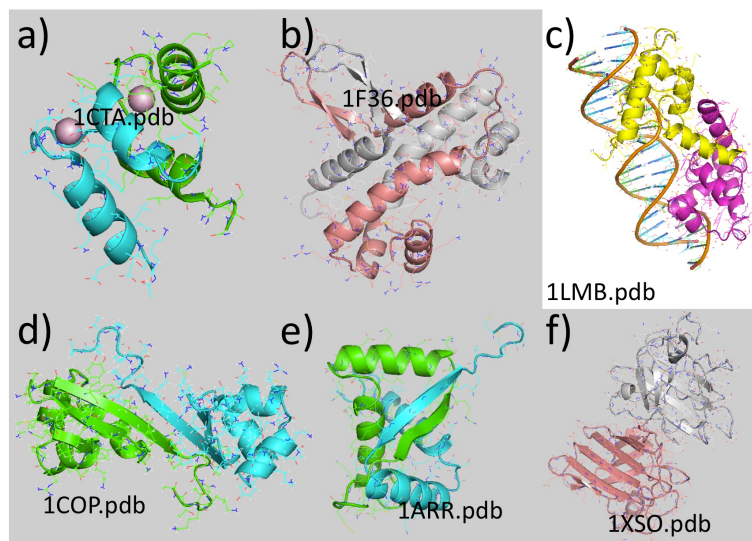


Figure 9.4 – Dimères de protéines. a) b) e) Dimères obligatoires. c) d) f) Dimères non obligatoires. (Images créées avec Pymol).

une organisation pré-tertiaire survenant lors d'une fusion préliminaire [13, 127]. Il est donc nécessaire de parfaire la méthode pour la simulation des transformations locales.

## 9.4 Épilogue

Des variations importantes peuvent apparaître dans les mécanismes moléculaires ou dans l'architecture des complexes moléculaires consécutivement à des variations locales plus ou moins subtiles telles que les mutations, les modifications post-traductionnelles et l'effet de différents ligands. Les progrès en rapidité et en précision des techniques numériques peuvent permettre d'affiner la compréhension de ces différences.

La poursuite du développement de la méthode *HMM* va de pair avec celle des modules informatiques associés à celle-ci : méthodes d'échantillonnage et champs de forces atomiques. Parce qu'en phase de développement, *ART holographique* doit d'abord être appliqué à des protéines accompagnées de solides informations expérimentales. Notre système permet ainsi de discerner plusieurs détails des trajectoires des mécanismes. Ceci peut conduire au développement de champs de forces de plus en plus aptes à reproduire

des transformations complexes connues. Ce n'est que lorsque les résultats de la méthode seront reconnus valides dans un bon nombre de situations que l'on pourra aborder l'exploration de mécanismes difficilement accessibles par les voies expérimentales. Les trajectoires hypothétiques devront ensuite être validées par de nouvelles expériences *in vitro*. Car les simulations informatiques s'inscrivent dans la partie théorique du cycle traditionnel de la méthode scientifique : données - hypothèses - expérimentations basées sur les hypothèses - nouvelles données...

## BIBLIOGRAPHIE

- [1] T. Ackbarow and M. J. Buehler. Strength and robustness of protein materials. *Encyc. of Nanoscience and Nanotechnology*, 23 :349–387, 2011.
- [2] T. Ackbarow, X. Chen, S. Keten, and M. J. Buehler. Hierarchies, multiple energy barriers, and robustness govern the fracture mechanics of alpha-helical and beta-sheet protein domains. *Proceedings of the National Academy of Sciences*, 104(42) :16410–16415, 2007.
- [3] A. Ahmed and H. Gohlke. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins : Structure, Function, and Bioinformatics*, 63(4) :1038–1051, 2006.
- [4] A. Ahmed, F. Rippmann, G. Barnickel, and H. Gohlke. A normal mode-based geometric simulation approach for exploring biologically relevant conformational transitions in proteins. *Journal of Chemical Information and Modeling*, 51(7) :1604–1622, 2011.
- [5] R. L. C. Akkermans and P. B. Warren. Multiscale modelling of human hair. *Philosophical Transactions : Mathematical, Physical and Engineering Sciences*, 362(1821) :pp. 1783–1793, 2004.
- [6] J. B. Ames, R. Ishima, T. Tanaka, J. I. Gordon, L. Stryer, and M. Ikura. Molecular mechanics of calcium-myristoyl switches. *Nature*, 389 :198–202, 1997.
- [7] R. Andersen, Y. Li, M. Resseguie, and J. E. Brenman. Calcium/calmodulin-dependent protein kinase ii alters structural plasticity and cytoskeletal dynamics in drosophila. *The journal of Neuroscience*, 25(39) :8878–8888, 2005.
- [8] B. Antonny, P. Chardin, M. Roux, and M. Chabre. Gtp hydrolysis mechanisms in ras p21 and in the ras-gap complex studied by fluorescence measurements on tryptophan mutants. *Biochemistry*, 30(34) :8287–8295, 1991.

- [9] H. Asmara, E. Minobe, Z. A. Saud, and M. Kameyama. Interactions of calmodulin with the multiple binding sites of *cav1.2*  $Ca^{2+}$  channels. *J Pharmacol Sci*, 112 :397–404, 2010.
- [10] E. Babini, I. Bertini, F. Capozzi, C. Luchinat, A. Quattrone, and M. Turano. Principal component analysis of the conformational freedom within the ef-hand superfamily. *Journal of Proteome Research*, 4(6) :1961–1971, 2005.
- [11] A. Babu, J. Gulati, and al. Critical minimum length of the central helix in troponin c for the  $Ca^{2+}$  switch in muscular contraction. *The Journal of Biological Chemistry*, 268(26) :19232–19238, 1993.
- [12] D. Baker. A surprising simplicity to protein folding. *Nature*, 405 :39–43, 2000.
- [13] R. L. Baldwin. Making a network of hydrophobic clusters. *Science*, 295(5560) :1657–1658, 2002.
- [14] R. L. Baldwin and G. D. Rose. Is protein folding hierarchic ? i. local structure and peptide folding. *Trends in Biochemical Sciences*, 24 :26–33, 1999.
- [15] N. K. Banavali and B. Roux. The n-terminal end of the catalytic domain of src kinase hck is a conformational switch implicated in long-range allosteric regulation. *Structure*, 13(11) :1715 – 1723, 2005.
- [16] M. A. A. Barbosa and A. F. Pereira de Araújo. Relevance of structural segregation and chain compaction for the thermodynamics of folding of a hydrophobic protein model. *Phys. Rev. E*, 67(5) :051919, May 2003.
- [17] G. T. Barkema and N. Mousseau. Event-based relaxation of continuous disordered systems. *Phys. Rev. Lett.*, 77(21) :4358–4361, 1996.
- [18] G. T. Barkema and N. Mousseau. Identification of relaxation and diffusion mechanisms in amorphous silicon. *Phys. Rev. Lett.*, 81 :1865–1868, 1998.

- [19] C. Bartels and M. Karplus. Multidimensional adaptive umbrella sampling : Applications to main chain and side chain peptide conformations. *Journal of Computational Chemistry*, 18(12) :1450–1462, 1997.
- [20] T. Bastug and S. Kuyucak. Molecular dynamics simulations of calcium binding gramicin a. *Chemical Physics Letters*, 424 :82–85, 2006.
- [21] C. J. Benham. An elastic model of the large-scale structure of duplex dna. *Biopolymers*, 18(3) :609–623, 1979.
- [22] B. Berger and T. Leighton. Protein folding in the hydrophobic-hydrophilic (hp) model is np-complete. *Journal of Computational Biology*, 5(1) :27–40, 1998.
- [23] J. Bergès, G. Rickards, A. Rauk, and C. Houée-Levin. Qm/mm study of electron addition on protein disulfide bonds. *Chemical Physics Letters*, 421(1-3) :63 – 67, 2006.
- [24] M. Berjanskii and D. S. Wishart. Nmr : prediction of protein flexibility. *Nature Publishing Group*, 1 :683–688, 2006.
- [25] J. D. Bernal. Biographical memoirs of fellows of the royal society. *The Royal Society*, 9 :1–35, 1963.
- [26] M. Betanzos, C.-S. Chiang, H. R. Guy, and S. Sukharev. A large iris-like expansion of a mechanosensitive channel protein induced by membrane tension. *Nature Structural Biology*, 9 :704–710, 2002.
- [27] K. Binder and D. Stauffer. Statistical theory of nucleation, condensation and coagulation. *Advances in Physics*, 25(4) :343–396, 1977.
- [28] T. L. Blundell and L. Johnson. Protein crystallography. *Molecular Biology Series*, 1976.
- [29] V. Borsi, C. Luchinat, and G. Parigi. Global and local mobility of apocalmodulin monitored through fast-field cycling relaxometry. *Biophysical Journal*, 97 :1765–1771, 2009.



- [30] B. Brooks and M. Karplus. Harmonic dynamics of proteins : normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences*, 80(21) :6571–6575, 1983.
- [31] J. M. Carr, S. A. Trygubenko, and D. J. Wales. Finding pathways between distant local minima. *The Journal of Chemical Physics*, 122 :234903, 2005.
- [32] L. Cavallo, J. Kleinjung, and F. Fraternali. Pops : a fast algorithm for solvent accessible surface areas at atomic and residu level. *Nucleic Acids Research*, 31(13) :3364–3366, 2003.
- [33] H. S. Chan and K. A. Dill. Transition states and folding dynamics of proteins and heteropolymers. *The Journal of Chemical Physics*, 100(12) :9238–9257, 1994.
- [34] M. Chandra, K. B. Campbel, and al. Interaction between myosin heavy chain and troponin isoforms modulate cardiac myofiber contractile dynamics. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, 293 :1595–1607, 2007.
- [35] R. Chattopadhyaya, W. E. Meador, A. R. Means, and F. A. Quioco. Calmodulin structure refined at 1.7 a resolution. *J. Mol. Biol.*, 228 :1177, 1992.
- [36] Y. Chebaro, X. Dong, R. Laghaei, P. Derreumaux, and N. Mousseau. Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. *J. Phys. Chem.*, 113 :267–274, 2009.
- [37] Y. Chen and M. D. Barkley. Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, 37(28) :9976–9982, 1998.
- [38] Y. G. Chen and G. Hummer. Slow conformational dynamics and unfolding of the calmodulin c-terminal domain. *J. Am. Chem. Soc.*, 129(9) :2414–2415, 2007.
- [39] J. J. Chou, S. Li, C. B. Klee, and A. Bax. Solution structure of ca<sup>2+</sup>-calmodulin reveals flexible hand-like properties of its domains. *Nature structural biology*, 8(11) :990–997, 2001.

- [40] J.-W. Chu and G. A. Voth. Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophysical Journal*, 90 :1572–1582, 2006.
- [41] M. W. Clarkson and A. L. Lee. Long-range dynamic effects of point mutations propagate through side chains in the serine protease inhibitor eglin c†. *Biochemistry*, 43(39) :12448–12458, 2004. PMID : 15449934.
- [42] G. M. Clore and C. D. Schwieters. Concordance of residual dipolar couplings, backbone order parameters and crystallographic b-factors for a small [alpha]/[beta] protein : A unified picture of high probability, fast atomic motions in proteins. *Journal of Molecular Biology*, 355(5) :879 – 886, 2006.
- [43] F. E. Cohen and S. B. Prusiner. Pathologic conformations of prion proteins. *Annual Review of Biochemistry*, 67 :793–819, 1998.
- [44] P. Cohen. Review lecture : Protein phosphorylation and hormone action. *Proceeding of the Royal Society*, 234 :113–144, 1988.
- [45] S. Côté, P. Derreumaux, and N. Mousseau. Distinct morphologies for amyloid beta monomer : Abeta(1-40), ab(1-42) and ab(1-40)(d23n). *Journal of Chemical Theory and Computation*, 2011.
- [46] L. Cowen, P. Bradley, M. Menke, J. King, and B. Berger. Predicting the beta-helix fold from protein sequence data. *Journal of Computational Biology*, 9(2) :261–276, 2002.
- [47] L. Cruzeiro-Hansson and P. A. S. Silva. Protein folding : thermodynamic versus kinetic control. *Journal of Biological Physics*, 27 :S6–S8, 2001.
- [48] C. Czaplewski, A. Liwo, M. Makowski, S. Ołdziej, and H. A. Scheraga. Coarse-grained models of proteins : Theory and applications. pages 35–83, 2011.
- [49] J. Davis and S. B. Tikunova. Ca<sup>2+</sup> exchange with troponin c and cardiac muscle dynamics. *Cardiovascular Research*, 77 :619–626, 2008.

- [50] R. M. Delahay, G. D. Balkwill, K. A. Bunting, W. Edwards, J. C. Atherton, and M. S. Searle. The highly repetitive region of the helicobacter pylori cagY protein comprises tandem arrays of an [alpha]-helical repeat module. *Journal of Molecular Biology*, 377(3) :956 – 971, 2008.
- [51] P. Derreumaux. Generating ensemble averages for small proteins from extended conformations by monte carlo simulations. *Phys. Rev. Lett.*, 85(1) :206–209, Jul 2000.
- [52] P. Derreumaux and N. Mousseau. Coarse-grained protein molecular dynamics simulations. 126(2) :025101, 2007.
- [53] A. Dhar, K. Girdhar, D. Singh, H. Gelman, S. Ebbinghaus, and M. Gruebele. Protein stability and folding kinetics in the nucleus and endoplasmic reticulum of eucaryotic cells. *Biophysical journal*, 101 :421 – 430, 2011.
- [54] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan. Principles of protein folding—a perspective from simple exact models. *Protein Science*, 4 :561–602, 1995.
- [55] F. Ding, N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophysical journal*, 83 :3525–3532, 2002.
- [56] P. Doruker, R. L. Jernigan, and I. Bahar. Dynamics of large proteins through hierarchical levels of coarse-grained structures. *Journal of Computational Chemistry*, 23(1) :119–127, 2002.
- [57] L. Dupuis and N. Mousseau. Holographic multiscale method used with non-biased atomistic forcefields for simulation of large transformations in protein. *Journal of Physics : Conference Series*, 341 :012015, 2012.
- [58] L. Dupuis and N. Mousseau. Understanding the ef-hand closing pathway using non-biased interatomic potentials. *The Journal of Chemical Physics*, 136 :035101, 2012.

- [59] P. Eastman, N. Grønbech-Jensen, and S. Doniach. Simulation of protein folding by reaction path annealing. *114(8) :3823–3841*, 2001.
- [60] M. G. Elad Project and al. A molecular dynamics study of the effect of  $ca^{2+}$  removal on calmodulin structure. *Biophysical Journal*, 90 :3842–3850, 2006.
- [61] J. Elezgaray, G. Marcou, and Y. H. Sanejouand. Exploring the natural conformation changes of the c-terminal domain of calmodulin. *PHYSICAL REVIEW*, 66 :0319081–0319088, 2002.
- [62] M. Esteller. Cancer epigenomics : Dna methylomes and histone-modification maps. *Nat Rev Genet*, 8 :286–298, 2007.
- [63] J. Evanans, G. Carlstrom, and al. Nmr studies of the e140q mutant of the carboxy-terminal domain of calmodulin reveal global conformational exchange in the  $ca^{2+}$ -saturated state. *NMR Studies of the E140Q Mutant of the Carboxy-Terminal Domain of Calmodulin Reveal Global Conformational Exchange in the CA2+-Saturated State*, 36 :3448–3457, 1997.
- [64] E. Eyal, S. Gerzon, V. Potapov, M. Edelman, and V. Sobolev. The limit of accuracy of protein modeling : Influence of crystal packing on protein structure. *Journal of Molecular Biology*, 351(2) :431 – 442, 2005.
- [65] E. Eyal, L.-W. Yang, and I. Bahar. Anisotropic network model : systematic evaluation and a new web interface. *Bioinformatics*, 22(21) :2619–2627, 2006.
- [66] M. Fandrich, V. Forge, K. Buder, M. Kittler, C. M. Dobson, and S. Diekmann. Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. *Proceedings of the National Academy of Sciences*, 100(26) :15463–15468, 2003.
- [67] C. Farah, C. Miyamoto, C. Ramos, A. da Silva, R. Quaggio, K. Fujimoi, L. Smillie, and F. Reinach. Structural and regulatory functions of the nh<sub>2</sub>- and cooh-terminal regions of skeletal muscle troponin i. *The Journal of Biological Chemistry*, 269(7) :5230–5240, 1994.

- [68] C. Farah and F. Reinach. The troponin complex and regulation of muscle contraction. *The FASEB journal*, 9 :755–767, 1995.
- [69] D. V. Fedyukina and S. Cavagnero. Protein folding at the exit tunnel. *Annu. Rev. Biophys.*, 40 :337–359, 2011.
- [70] M. D. Feldkamp, S. E. O’Donnell, L. Yu, and M. A. Shea. Allosteric effects of the antipsychotic drug trifluoperazine on the energetics of calcium binding by calmodulin. *Proteins : Structure, Function, and Bioinformatics*, 78(10) :2265–2282, 2010.
- [71] D. U. Ferreira, A. M. Walczak, A. K. Elizabeth, and P. G. Wolynes. The energy landscapes of repeat-containing proteins : Topology, cooperativity, and the folding funnels of one-dimensional architectures. *PLoS Computational Biology*, 4 :1–13, 2008.
- [72] A. R. Fersht. Optimization of rates of protein folding : the nucleation-condensation mechanism and its implications. *Proceedings of the National Academy of Sciences*, 92(24) :10869–10873, 1995.
- [73] K. A. Fichtorn and W. H. Weinberg. Theoretical foundations of dynamical monte carlo simulations. *The Journal of Chemical Physics*, 95(2) :1090–1096, 1991.
- [74] A. V. Finkelstein and O. B. Ptitsyn. A theory of protein molecule self-organization : Iv. helical and irregular local structures of unfolded protein chains. *Journal of Molecular Biology*, 103(1) :15 – 24, 1976.
- [75] B. E. Finn, J. Evenas, T. Drakenberg, J. P. Waltho, E. Thulin, and S. Forsén. Calcium-induced structural changes and domain autonomy in calmodulin. *Nature structural biology*, 2(9) :777, 1995.
- [76] G. a. Fiorin, P. M., and al. Using metadynamics to understand the mechanism of calmodulin/target recognition at atomic detail. *Biophysical Journal*, 91 :2768–2777, 20056.

- [77] R. R. Fiorin, G. and. Biekofsky, A. Pastor, and P. Carloni. Unwinding the helical linker of calcium-loaded calmodulin : A molecular dynamics study. *PROTEINS : Structure, Function, and Bioinformatics*, 61 :829–839, 2005.
- [78] S. Flores, N. Echols, D. Milburn, B. Hespeneide, K. Keating, J. Lu, S. Wells, E. Z. Yu, M. Thorpe, and M. Gerstein. The database of macromolecular motions : new features added at the decade mark. *Nucleic Acids Research*, 34(suppl 1) :D296–D301, 2006.
- [79] R. Fraczkiewicz and W. Braun. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *Journal of Computational Chemistry*, 19(3) :319–333, 1998.
- [80] S. Fulle and H. Gohlke. Determining rna flexibility by graph theory : Ribosomal exit tunnel as a case study. *From Computational Biophysics to Systems Biology*, 40 :213–216, 2008.
- [81] S. Fulle and H. Gohlke. Statics of the ribosomal exit tunnel : Implications for co-translational peptide folding, elongation regulation, and antibiotics binding. *Journal of Molecular Biology*, 387(2) :502 – 517, 2009.
- [82] S. M. Gagné, L. Tsuda, S. and. Spyropoulos, L. E. Kay, and B. D. Sykes. Backbone and methyl dynamics of the regulatory domain of troponin c : Anisotropic rotational diffusion and contribution of conformational entropy to calcium affinity. *J. Mol. Biol.*, 278 :667–686, 1998.
- [83] S. M. Gagné, S. Tsuda, M. X. Li, L. B. Smillie, and B. D. Sykes. Structures of the troponin c regulatory domains in the apo and calcium-saturated states. *Nature structural biology*, 2(9) :784–789, 1995.
- [84] A. Ganoth, E. Nachliel, R. Friedman, and M. Gutman. Molecular dynamics study of a calmodulin-like protein with a iq peptide : Spontaneous refolding of the protein around the peptide. *PROTEINS : Structure, Function, and Bioinformatics*, 64 :133–146, 2006.

- [85] A. E. García. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 68(17) :2696–2699, Apr 1992.
- [86] G. Gibrat, G. C. T., and al. Biophysical study of thermal denaturation of apo-calmodulin : Dynamics of native and unfolded states. *Biophysical Journal*, 95 :5247–5256, 2008.
- [87] J. L. Gifford, M. P. Walsh, and H. J. Vogel. Structure and metal-ion binding properties of the  $ca^{2+}$ -binding helix-loop-helix ef-hand motifs. *Biochem, J.*, 405 :199–221, 2007.
- [88] N. Go, T. Noguti, and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proceeding of the National Academy of Sciences of the U.S.A.*, 80 :3696–3700, 1983.
- [89] N. Go, T. Noguti, and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proceedings of the National Academy of Sciences*, 80(12) :3696–3700, 1983.
- [90] H. Gohlke and M. F. Thorpe. A natural coarse graining for simulating large biomolecular motion. *Biophysical Journal*, 91 :2115–2120, 2006.
- [91] J. T. Gonçalves and W. Stühmer. Calmodulin interaction with heag1 visualized by fret microscopy. *PLoS ONE*, 5(5) :e10873, 2010.
- [92] I. Goze and U. Z. Littauer. Microtubule protein : Tubulin. *Scandinavian Journal of Immunology*, 15 :299–316, 1982.
- [93] Z. Grabarek. Structure of a trapped intermediate of calmodulin : Calcium regulation of ef-hand proteins from a new perspective. *J. Mol. Biology*, 346 :1351–1366, 2005.
- [94] Z. Grabarek. Structural basis for diversity of the ef-hand calcium-binding proteins. *J. Mol. Biol.*, 359 :509–525, 2006.

- [95] Z. Grabarek. Insights into modulation of calcium signaling by magnesium in calmodulin, troponin c and related ef-hand proteins. *Biochimica et Biophysica Acta*, 1813 :913–921, 2011.
- [96] H. M. Gradin, U. Marklund, N. Larsson, T. A. Chatila, and M. Gullberg. Regulation of microtubule dynamics by  $ca^{+2}$ /calmodulin-depedent kinase iv/gr-dependent phosphorylation of oncoprotein 18. *Molecular and Cellular Biology*, 17(6) :3459–3467, 1997.
- [97] M. M. Gromiha and S. Selvaraj. Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins : application of long-range order to folding rate prediction. *Journal of Molecular Biology*, 310(1) :27 – 32, 2001.
- [98] H. Grotsch, J. P. Giblin, F.-Z. Idrissi, I.-M. Fernandez-Golbano, J. R. Collette, T. M. Newpher, V. Robles, S. K. Lemmon, and M.-I. Geli. Calmodulin dissociation regulates myo5 recruitment and function at endocytic sites. *The EMBO journal*, 29 :2899–2914, 2010.
- [99] M. Gruebele, G. Scott, Y. Chemla, and M. Gruebele. Better folding thermodynamics from kinetics. *J. Chem. Phys.*, 2011.
- [100] J. Gulati, H. Weinstein, and al. Functional role of arginin-11 in the n-terminal helix of skeletal troponin c : Combined mutagenesis and molecular dynamics investigation. *Biochemistry*, 34 :7348–7355, 1995.
- [101] O. Guvench and C. L. Brooks III. Efficient approximate all-atom solvent accessible surface area method parameterized for folded and denatured protein conformations. *J Comput Chem*, 25 :1005–1014, 2003.
- [102] T. Haliloglu, I. Bahar, and B. Erman. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.*, 79(16) :3090–3093, Oct 1997.



- [103] C. Y. Hayashi, N. H. Shipley, and R. V. Lewis. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *International Journal of Biological Macromolecules*, 24(2-3) :271 – 275, 1999.
- [104] K. Hinsen. Analysis of domain motions by approximate normal mode calculations. *Proteins : Structure, Function, and Bioinformatics*, 33(3) :417–429, 1998.
- [105] T. X. Hoang and M. Cieplak. Molecular dynamics of folding of secondary structures in go-type models of proteins. *The Journal of Chemical Physics*, 112(15) :6851–6862, 2000.
- [106] R. S. Hodges, J. Sodek, L. B. Smillie, and L. Jurasek. Tropomyosin : Amino acid sequence and coiled-coil structure. *CSH symposia*, 37 :299–310, 1973.
- [107] A. Holmgren. Tryptophan fluorescence study of conformational transitions of the oxidized and reduced form of thioredoxin. *Journal of Biological Chemistry*, 247(7) :1992–1998, 1972.
- [108] M. Ikura. Calcium binding and conformational response in ef-hand proteins. *Trends in Biochemical Sciences*, 21(1) :14 – 17, 1996.
- [109] A. Imparato, J. C. Shillcock, and R. Lipowsky. Shape fluctuations and elastic properties of two-component bilayer membranes. *EPL (Europhysics Letters)*, 69(4) :650, 2005.
- [110] Y. Ishii, T. Yoshida, T. Funatsu, T. Wazawa, and T. Yanagida. *Chemical Physics*, 247(1) :163 – 173, 1999.
- [111] R. Ishima and D. A. Torchia. Protein dynamics from nmr. *at Struct Mol Biol*, 7 :704–743, 2000.
- [112] S. Izvekov and G. A. Voth. Modeling real dynamics in the coarse-grained representation of condensed phase systems. 125(15) :151101, 2006.

- [113] D. J. Jacobs, A. J. Rader, L. A. Kuhn, and M. F. Thorpe. Protein flexibility predictions using graph theory. *Proteins : Structure, Function, and Bioinformatics*, 44(2) :150–165, 2001.
- [114] D. J. Jacobs and M. F. Thorpe. Generic rigidity percolation : The pebble game. *Phys. Rev. Lett.*, 75(22) :4051–4054, 1995.
- [115] D. Janežič and B. R. Brooks. Harmonic analysis of large systems. ii. comparison of different protein models. *Journal of Computational Chemistry*, 16(12) :1543–1553, 1995.
- [116] Y. Jiang, A. Lee, J. Chen, M. Cadene, B. T. Chait, and R. MacKinnon. The open pore conformation of potassium channels. *Nature*, 417 :523–526, 2002.
- [117] J. E. Jimenez-Roldan, S. A. Wells, R. B. Freedman, and R. A. Roemer. Integration of first, froda and nmm in a coarse grained method to study protein disulphide isomerase conformational change. *Journal of Physics : Conference Series*, 286(1) :012002, 2011.
- [118] D. T. Jones. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 292(2) :195 – 202, 1999.
- [119] B. Juanico, Y.-H. Sanejouand, F. Piazza, and P. De Los Rios. Discrete breathers in nonlinear network models of proteins. *Phys. Rev. Lett.*, 99(23) :238104, Dec 2007.
- [120] O. Julien, S. B. D., and al. Nmr studies of the dynamics of a bifunctional rhodamine probe attached to troponin c. *J. AM. CHEM. SOC.*, 130 :2602–2609, 2008.
- [121] J. Junker, Z. F., and M. Rief. Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science*, 323 :633–637, 2009.
- [122] J. Junker and M. Rief. Evidence for a broad transition-state ensemble in calmodulin folding from single-molecule force spectroscopy. *Protein Folding*, 3 :3306–3309, 2010.

- [123] M. Karplus. Molecular dynamics simulations of biomolecules. *Accounts of Chemical Research*, 35(6) :321–323, 2002.
- [124] H. Kawasaki, S. Nakayama, and R. H. Kretsinger. Classification and evolution of ef-hand proteins. *BioMetals*, 11 :277–295, 1998.
- [125] M. K. Kim, R. L. Jernigan, and G. S. Chirikjian. Efficient generation of feasible pathways for protein conformational transitions. *Biophysical Journal*, 83(3) :1620 – 1630, 2002.
- [126] A. Kitao and N. Gō. Conformational dynamics of polypeptides and proteins in the dihedral angle space and in the cartesian coordinate space : Normal mode analysis of deca-alanine. *Journal of Computational Chemistry*, 12(3) :359–368, 1991.
- [127] J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson, and H. Schwalbe. Long-range interactions within a nonnative protein. *Science*, 295(5560) :1719–1722, 2002.
- [128] D. K. Klimov and D. Thirumalai. Mechanisms and kinetics of beta-hairpin formation. *Proceedings of the National Academy of Sciences*, 97(6) :2544–2549, 2000.
- [129] C. Kobayashi and S. Takada. Protein grabs a ligand by extending anchor residues : Molecular simulation for ca<sup>2+</sup> binding to calmodulin loop. *Biophysical Journal*, 90 :3043–3051, 2006.
- [130] Y. Komeiji, Y. Yutaka Ueno, and M. Uebayasi. Molecular dynamics simulations revealed ca<sup>2+</sup> -dependent conformational change of calmodulin. *FEBS Letters*, 521 :133–139, 2002.
- [131] C. L. Kong. Combining rules for intermolecular potential parameters. ii. rules for the lennard-jones (12-6) potential and the morse potential. *J Chem Phys*, 59(5) :2464–2467, 1973.

- [132] G. König and S. Boresch. Hydration free energies of amino acids : Why side chain analog data are not enough. *J. Phys. Chem.*, 113 :8967–8974, 2009.
- [133] T. Kortemme, A. V. Morozov, and D. Baker. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *Journal of Molecular Biology*, 326(4) :1239 – 1259, 2003.
- [134] W. G. Krebs, V. Alexandrov, C. A. Wilson, N. Echols, H. Yu, and M. Gerstein. Normal mode analysis of macromolecular motions in a database framework : Developing mode concentration as a useful classifying statistic. *Proteins : Structure, Function, and Bioinformatics*, 48(4) :682–695, 2002.
- [135] H. Kuboniwa, T. N., G. S., H. Ren, C. B. Klee, and A. Bax. Solution structure of calcium-free calmodulin. *Nature Structural Biology*, 2 :768–776, 1995.
- [136] K. Kuwajima. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins : Structure, Function, and Bioinformatics*, 6(2), 1989.
- [137] M. E. L., P.-A. J.-L., and W. H. Structural dynamics of calmodulin and troponin c. *Protein Engineering*, 3 :625–637, 1991.
- [138] R. Laghaei and N. Mousseau. Spontaneous formation of polyglutamine nanotubes with molecular dynamics simulations. *J Chem Phys*, 132(16) :165102, 2010.
- [139] R. Laghaei, N. Mousseau, and G. Wei. Effect of the disulfide bond on the monomeric structure of human amylin studied by combined hamiltonian and temperature replica exchange molecular dynamics simulations. *J. Phys. Chem.*, 114 :7071–7077, 2010.
- [140] E. Laine, T. E. Malliavin, and al. The conformational plasticity of calmodulin upon calcium complexation gives a model of its interaction with the oedema factor of *Bacillus anthracis*. *Proteins*, 71 :1813–1829, 2009.

- [141] D. Langosch, L. Thomas, and H. Betz. Conserved quaternary structure of ligand-gated ion channels : the postsynaptic glycine receptor is a pentamer. *Proceedings of the National Academy of Sciences*, 85(19) :7394–7398, 1988.
- [142] J. Langowski and D. W. Heermann. Computational modeling of the chromatin fiber. *Seminars in Cell and Developmental Biology*, 18(5) :659 – 667, 2007. Membrane Lipid Microdomains : Roles in Signalling and Disease and 3D Chromatin ; Structure Inside the Cell Nucleus.
- [143] R. Lavery, I. Parker, and J. Kendrick. A general approach to the optimization of the conformation of ring molecules with an application to valinomycin. *J Biomol Struct Dyn*, 4 :443–462, 1986.
- [144] Z. Lavery, K. Zakrzewska, and H. Sklenar. Jumna (junction minimisation of nucleic acids). *Computer Physics communication*, 91 :135–158, 1995.
- [145] T. Lazaridis and M. Karplus. Effective energy function for proteins in solution. *PROTEINS : Structure, Function and Genetics*, 35 :133–152, 1999.
- [146] T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. ii. lattice gas and ising model. *Phys. Rev.*, 87(3) :410–419, Aug 1952.
- [147] M. Lepsik and M. J. Field. Binding of calcium and other metal ions to the ef-hand loops of calmodulin studied by quantum chemical calculations and molecular dynamics simulations. *The Journal of Physical Chemistry B*, 111(33) :10012–10022, 2007.
- [148] C. Levinthal. Are there pathways for protein folding? *Journal of Medical Physics*, 65(1) :44–45, 1968.
- [149] M. Levitt, C. Sander, and P. S. Stern. Protein normal-mode dynamics : Trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of Molecular Biology*, 181(3) :423 – 447, 1985.

- [150] M. Levitt and A. Warshel. Computer simulation of protein folding. *Nature*, 253 :694–698, 1975.
- [151] Y. Levy, P. G. Wolynes, and J. N. Onuchic. Protein topology determines binding mechanism. *Proceedings of the National Academy of Sciences of USA*, 101(2) :511–516, 2004.
- [152] A. Lewit-Bentley and S. Réty. Ef-hand calcium-binding proteins. *Current Opinion in Structural Biology*, 10 :637–643, 2000.
- [153] H. Li, R. Helling, C. Tang, and N. Wingreen. Emergence of preferred structures in a simple model of protein folding. *Science*, 273(5275) :666–669, 1996.
- [154] M. X. Li, S. B. D., and al. Kinetic studies of calcium and cardiac troponin i peptide binding to human cardiac troponin c using nmr spectroscopy. *Eur. Biophys J.*, 31 :245–256, 2002.
- [155] D. A. Lindhout and S. B. D. Structure and dynamics of the c-domain of human cardiac troponin c in complex with the inhibitory region of human cardiac troponin i. *The Journal of Biological Chemistry*, 278(29) :27024–27034, 2003.
- [156] J. Liu and S. A. Siegelbaum. Change of pore helix conformational state upon opening of cyclic nucleotide-gated channels. *Neuron*, 28 :899–909, 2000.
- [157] A. Liwo, M. Khalili, and H. A. Scheraga. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7) :2362–2367, 2005.
- [158] J. Lu and C. Deutsch. Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol*, 12 :1123–1129, 2005.
- [159] M. Lu and J. Ma. The role of shape in determining molecular motions. *Biophysical Journal*, 89(4) :2395 – 2401, 2005.

- [160] A. Lupas, M. V. Dyke, and J. Stock. Predicting coiled coils from protein sequences. *Science*, 252(5009) :pp. 1162–1164, 1991.
- [161] T. T. M. Zhang and M. Ikura. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nature structural biology*, 2(9) :758–767, 1995.
- [162] J. Ma. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, 13(3) :373 – 380, 2005.
- [163] E. Machado-Charry, D. Caliste, L. Genovese, N. Mousseau, and P. Pochet. Optimized energy landscape exploration using the ab initio based art-nouveau. *J. Chem Phys.*, 2011.
- [164] A. D. MacKerell and N. K. Banavali. All-atom empirical force field for nucleic acids : Ii. application to molecular dynamics simulations of dna and rna in solution. *Journal of Computational Chemistry*, 21(2) :105–120, 2000.
- [165] J. A. D. MacKerell, M. Karplus, and al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *American Chemical Society*, 102 :3586–3616, 1998.
- [166] E. R. G. Main, Y. Xiong, M. J. Cocco, L. D’Andrea, and L. Regan. Design of stable [alpha]-helical arrays from an idealized tpr motif. *Structure*, 11(5) :497 – 508, 2003.
- [167] R. Malek and N. Mousseau. Dynamics of lennard-jones clusters : A characterization fo the activation-relaxtion technique. *Phys. Rev.*, 62 :7723–7728, 2000.
- [168] T. Mamonova, B. Hespeneide, R. Straub, M. F. Thorpe, and M. Kurnikova. Protein flexibility using constraints from molecular dynamics simulations. *Physical Biology*, 2(4) :S137, 2005.

- [169] S. Marchand and B. Roux. Molecular dynamics study of calbindin d9k in the apo and singly and doubly calcium-loaded states. *Proteins : Structure, Function, and Bioinformatics*, 33(2) :265–284, 1998.
- [170] M.-C. Marinica, F. Willaime, and N. Mousseau. Energy landscape of small clusters of self-interstitial dumbbells in iron. *Phys. Rev. B*, 83(9) :094119, Mar 2011.
- [171] J. Maupetit, P. Tuffery, and P. Derreumaux. A coarse-grained protein force field for folding and structure prediction. *Proteins*, 69 :394–408, 2007.
- [172] H. V. McCue, L. P. Haynes, and R. D. Burgoyne. The diversity of calcium sensor proteins in the regulation of neuronal function. *CHS perspectives*, 2 :1, 2010.
- [173] J. A. McGammon and M. Karplus. Dynamics of folded proteins. *Nature*, 267(5612) :585–590, 1977.
- [174] C. C. Mello and D. Barrick. An experimentally determined protein folding energy landscape. *Proceedings of the National Academy of Sciences of the United States of America*, 101(39) :14102–14107, 2004.
- [175] F. E. Mellouhi, L. J. Lewis, and N. Mousseau. The kinetic activation-relaxation technique : A powerful off-lattice on-the-fly kinetic monte carlo algorithm. *Phys. Rev.*, 78 :153202, 2008.
- [176] A. Melquiond, G. Boucher, N. Mousseau, and P. Derreumaux. Following the aggregation of amyloid-forming peptides by computer simulations. *The Journal of Chemical Physics*, 122 :1245–1255, 2005.
- [177] D. F. Meyer, Y. Mabuchi, and Z. Grabarek. The role of phe-92 in the ca<sup>2+</sup> induced conformational transition in the c-terminal domain of calmodulin. *The Journal of Biological Chemistry*, 271(9) :11284–11290, 1996.
- [178] G. T. Montelione, D. Zheng, Y. J. Huang, K. C. Gunsalus, and T. Szyperski. Protein nmr spectroscopy in structural genomics. *nature structural biology*, pages 982–985, 2000.



- [179] D. Moroni, T. Wolde, P. Rein, and P. G. Bolhuis. Interplay between structure and size in a critical crystal nucleus. *Physical Review L.*, 94(23) :235703, 2005.
- [180] L. K. Mosavi, T. J. Cammett, D. C. Desrosiers, and Z. yu Peng. The ankyrin repeat as molecular architecture for protein recognition. *Protein Science*, 13(6) :1435–1448, 2004.
- [181] N. Mousseau and G. T. Barkema. Activated mechanisms in amorphous silicon : An activation-relaxation-technique study. *Phys. Rev.*, B 61 :1898–1906, 2000.
- [182] N. Mousseau, P. Derreumaux, G. T. Barkema, and R. Malek. Sampling activated mechanisms in proteins with the activation-relaxation technique. *Mol. Graph. and Modeling*, 19 :78–86, 2001.
- [183] K. B. Mullis and F. A. Faloona. Specific synthesis of dna in vitro via a polymerase-catalyzed chain reaction. *Academic Press*, 155 :335–350, 1987.
- [184] M. G. Munteanu, K. Vlahovicek, S. Parthasarathy, I. Simon, and S. Pongor. Rod models of dna : sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends in Biochemical Sciences*, 23(9) :341 – 347, 1998.
- [185] R. B. Murphy, D. M. Philipp, and R. A. Friesner. A mixed quantum mechanics/molecular mechanics (qm/mm) method for large-scale modeling of chemistry in protein environments. *Journal of Computational Chemistry*, 21(16) :1442–1457, 2000.
- [186] L. S. Musil and D. A. Goodenough. Multisubunit assembly of an integral plasma membrane channel protein, gap junction connexin43, occurs after exit from the er. *Cell*, 74(6) :1065 – 1077, 1993.
- [187] J. K. Myers, C. N. Pace, and J. M. Scholtz. Denaturant m values and heat capacity changes : relation to changes in accessible surface areas of protein unfolding. *Protein Sci.*, 4(10) :2138–2148, 1995.

- [188] K. Nakanishi, N. Berova, and R. Woody. Circular dichroism : principles and applications. *VCH*, pages 1–570, 1994.
- [189] J. Nasica-Labouze, M. Meli, P. Derreumaux, G. Colombo, and N. Mousseau. A multiscale approach to characterize the early aggregation steps of the amyloid-forming peptide gnnqqny from the yeast prion sup-35. *PLoS Comput Biol*, 7(5) :e1002051, 05 2011.
- [190] J. Nasica-Labouze and N. Mousseau. Kinetics of amyloid growth. *Book chapter*, Ed. Philippe Derreumaux.
- [191] M. R. Nelson and W. J. Chazin. Structures of ef-hand ca 2+-binding proteins : Diversity in the organization, packing and response to ca 2+ binding. *BioMetals*, 11 :297–318, 1998.
- [192] E. Neria, S. Fischer, and M. Karplus. Simulation of activation free energies in molecular systems. *Journal of Chemical Physics*, 105(5) :1902–1921, 1996.
- [193] H. D. Nguyen and C. K. Hall. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46) :16180–16185, 2004.
- [194] A. Nicholls, K. A. Sharp, and B. Honig. Protein folding and association : Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins : Structure, Function, and Bioinformatics*, 11(4) :281–296, 1991.
- [195] H. Nishiyama, T. Knöpfel, S. Endo, and S. Itohara. Glial protein s100b modulates long-term neuronal synaptic plasticity. *Proceedings of the National Academy of Sciences*, 99 :4037–4042, 2002.
- [196] N. Noguti, T. Go. Efficient monte carlo method for simulation of fluctuating conformations of native proteins. *Biopolymers*, 3 :527–546, 1985.
- [197] D. J. OConnell, M. C. Bauer, J. OBrien, W. M. Johnson, C. A. Divizio, S. L. OKane, T. Berggrd, A. Merino, K. S. Okerfeldt, S. Linse, and D. J. Cahill. Inte-

- grated protein array screening and high throughput validation of 70 novel neural calmodulin-binding proteins. *Molecular and Cellular Proteomics*, 9 :1118–1132, 2010.
- [198] J. Ollesch, E. Künnemann, R. Glockshuber, and K. Gerwert. Prion protein alpha-to-beta transition monitored by time-resolved fourier transform infrared spectroscopy. *Appl. Spectroscopy*, 61(10) :1025–1031, 2007.
- [199] J. Onions, S. Hermann, and T. Grundstrom. Basic helix-loop-helix protein sequences determining differential inhibition by calmodulin and s-100 proteins. *Journal of Biological Chemistry*, 272(38) :23930–23937, 1997.
- [200] J. Onions, S. Hermann, and T. Grundstrom. A novel type of calmodulin interaction in the inhibition of basic helix-loop-helix transcription factor. *Biochemistry*, 39 :4366–4374, 2000.
- [201] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding : The energy landscape perspective. *Annual Review of Physical Chemistry*, 48(1) :545–600, 1997.
- [202] J. N. Onuchic, H. Nymeyer, A. E. García, J. Chahine, and N. D. Socci. The energy landscape theory of protein folding : Insights into folding mechanisms and scenarios. 53 :87 – 130, IN3–IN4, 131–152, 2000.
- [203] J. N. Onuchic and P. G. Wolynes. Theory of protein folding. *Current Opinion in Structural Biology*, 14(1) :70 – 75, 2004.
- [204] T. Ostendorp, E. Leclerc, A. Galichet, M. Koch, N. Demling, B. Weigle, C. W. Heizmann, P. M. H. Kroneck, , and G. Fritz. Structural and functional insights into rage activation by multimeric s100b. *EMBO J*, 26 :3868–3878, 2007.
- [205] D. E. Otzen and A. R. Fersht. Analysis of protein–protein interactions by mutagenesis : direct versus indirect effects. *Protein Engineering*, 12(1) :41–45, 1999.

- [206] A. C. Pan and B. Roux. Building markov state models along pathways to determine free energies and rates of transitions. 129(6) :064107, 2008.
- [207] A. C. Pan, D. Sezer, and B. Roux. Finding transition pathways using the string method with swarms of trajectories. *The Journal of Physical Chemistry B*, 112(11) :3432–3440, 2008.
- [208] G. Papoff, N. Trivieri, R. Crielesi, F. Ruberti, S. Marsilio, and G. Ruberti. Fadd-calmodulin interaction : A novel player in cell cycle regulation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 3 :898–911, 2010.
- [209] L. Pauling, R. B. Corey, and H. R. Branson. Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci*, 37(4) :205–211, 1951.
- [210] L. Pauling and C. D. Coryell. The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proc Natl Acad Sci U S A*, 22(4) :210–216, 1936.
- [211] M. D. Peraro, K. Spiegel, G. Lamoureux, M. D. Vivo, W. F. DeGrado, and M. L. Klein. Modeling the charge distribution at metal sites in proteins for molecular dynamics simulations. *Journal of Structural Biology*, 157(3) :444 – 453, 2007.
- [212] G. Perrot, S. H. A., and al. Mseed : A program for the rapid analytical determination of accessible surface areas and their derivatives. *Journal of Computational Chemistry*, 13(1) :1–11, 1992.
- [213] M. F. Perutz, J. C. Kendrew, and H. C. Watson. Structure and function of haemoglobin : II. some relations between polypeptide chain configuration and amino acid sequence. *Journal of Molecular Biology*, 13(3) :669 – 678, 1965.
- [214] F. Pichierri and A. Sarai. Elastic properties of the poly-glycine [alpha]-helix from periodic scf-lmo calculations. *Chemical Physics Letters*, 322(6) :536 – 542, 2000.
- [215] Ponder. <http://dasher.wustl.edu/tinker>, 2010.

- [216] M. Praprotnik, L. D. Site, and K. Kremer. A macromolecule in a solvent : Adaptive resolution molecular dynamics simulation. *The Journal of Chemical Physics*, 126(13) :134902, 2007.
- [217] O. B. Ptitsyn. Molten globule and protein folding. *Academic Press*, 47 :83–229, 1995.
- [218] G. R. and N. H. F. Ribosomes and translation. *Annual review of biochemistry*, 66 :38, 1997.
- [219] A. A. Rashin, A. H. L. Rashin, and R. L. Jernigan. Protein flexibility : coordinate uncertainties and interpretation of structural differences. *Acta Crystallogr D Biol Crystallogr.*, 65 :1140–1161, 2009.
- [220] A. Raybaud, Y. Dodier, P. Bissonnette, M. Simoes, D. G. Bichet, R. Sauvé, and L. Parent. The role of the gx9gx3g motif in the gating of high voltage-activated ca<sup>2+</sup> channels. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 281 :39424–39436, 2006.
- [221] M. Rief, F. Oesterhelt, B. Heymann, and H. E. Gaub. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science*, 275(5304) :1295–1297, 1997.
- [222] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub. Single molecule force spectroscopy of spectrin repeats : low unfolding forces in helix bundles. *Journal of Molecular Biology*, 286(2) :553 – 561, 1999.
- [223] R. Rohs, C. Etchebest, and R. Lavery. Unraveling proteins : A molecular mechanics study. *Biophysical Journal*, 76(5) :2760 – 2768, 1999.
- [224] C. A. Royer. Revisiting volume changes in pressure-induced protein unfolding. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1595(1-2) :201 – 209, 2002.

- [225] Y.-H. Sanejouand. Elastic network models : Theoretical and empirical foundations. 2011.
- [226] F. Sanger, S. Nicklen, and A. R. Coulson. Dna sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12) :5463–5467, 1977.
- [227] F. Sanger and H. Tuppy. The amino-acid sequence in the phenylalanyl chain of insulin. 2. the investigation of peptides from enzymic hydrolysates. *Biochemical Journal*, 49(4) :481–490, 1951.
- [228] S. Santini, G. Wei, N. Mousseau, and P. Derreumaux. Pathway complexity of alzheimer’s b-amyloid ab16-22 peptide assembly. *Structure*, 12(7) :1245–1255, 2004.
- [229] M. M. Santoro and D. W. Bolen. Unfolding free energy changes determined by the linear extrapolation method. 1. unfolding of phenylmethanesulfonyl .alpha.-chymotrypsin using different denaturants. *Biochemistry*, 27(21) :8063–8068, 1988.
- [230] M. Sayadi, S. Tanizaki, and M. Feig. Effect of membrane thickness on conformational sampling of phospholamban from computer simulations. *Biophysical Journal*, 98(5) :805 – 814, 2010.
- [231] A. Schlessinger and B. Rost. Protein flexibility and rigidity predicted from sequence. *Proteins : Structure, Function, and Bioinformatics*, 61(1) :115–126, 2005.
- [232] G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas’, and R. I. Gilmanshin. Study of the “molten globule” intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*, 31(1) :119–128, 1991.
- [233] S. Sharma, F. Ding, and N. V. Dokholyan. Multiscale modeling of nucleosome dynamics. *Biophysical Journal*, 92(5) :1457 – 1470, 2007.

- [234] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein. A coarse grain model for phospholipid simulations. *The Journal of Physical Chemistry B*, 105(19) :4464–4470, 2001.
- [235] X. Shen, C. A. Valencia, J. Szostak, B. Dong, and R. Liu. Scanning the human proteome for calmodulin-binding proteins. *PNAS*, 102(17) :5969–5974, 2005.
- [236] C. M. Shepherd and H. J. Vogel. A molecular dynamics study of  $Ca^{2+}$ -calmodulin : Evidence of interdomain coupling and structural collapse on the nanosecond timescale. *Biophysical Journal*, 87 :780–791, 2004.
- [237] B. D. Slaughter, J. R. Unruh, M. W. Allen, R. J. Bieber Urbauer, and C. K. Johnson. Conformational substates of calmodulin revealed by single-pair fluorescence resonance energy transfer : influence of solution conditions and oxidative modification. *Biochemistry*, 44(10) :3694–36707, 2005.
- [238] C. M. Slupsky, S. B. D., and al. Solution secondary structure of calcium-saturated troponin c monomer determined by multidimensional heteronuclear nmr spectroscopy. *Protein Science*, 4 :1279–1290, 1995.
- [239] C. M. Slupsky and B. D. Sykes. Nmr solution structure of calcium-saturated skeletal muscle troponin c. *Biochemistry*, 34 :15953–15964, 1995.
- [240] A. G. Smart. Kinetic experiments shed light on protein-folding thermodynamics. *Physics Today*, pages 18–20, Aug 2011.
- [241] G. Smith and M. MacQuaide. Swallowing a spider to catch a fly : Ca-calmodulin dynamics in adult cardiomyocytes. *Journal of Molecular and Cellular Cardiology*, 41 :416–417, 2006.
- [242] B. Somogyi, Z. Lakos, Á. Szarka, and M. Nyitrai. Protein flexibility as revealed by fluorescence resonance energy transfer : an extension of the method for systems with multiple labels. *Journal of Photochemistry and Photobiology B : Biology*, 59 :26–32, 2000.

- [243] L. Spyropoulos, S. B. D., and al. Temperature dependence of dynamics and thermodynamics of the regulatory domain of human cardiac troponin c. *Biochemistry*, 40 :12541–12551, 2001.
- [244] S. Sridharan, A. Nicholls, and K. A. Sharp. A rapid method for calculating derivatives of solvent accessible surface areas of molecules. *Journal of Computational Chemistry*, 16(8) :1038–1044, 1994.
- [245] J.-F. St-Pierre, M. Karttunen, N. Mousseau, T. Róg, and A. Bunker. Use of umbrella sampling to calculate the entrance/exit pathway for z-pro-proline inhibitor in prolyl oligopeptidase. *Journal of Chemical Theory and Computation*, 7(6) :1583–1594, 2011.
- [246] J.-F. St-Pierre, N. Mousseau, and P. Derreumaux. The complex folding pathways of protein a suggest a multiple-funnelled energy landscape. *J. Chem. Phys.*, 128 :0145101–1, 2008.
- [247] P. J. Stansfeld and M. S. P. Sansom. From coarse grained to atomistic : A serial multiscale approach to membrane protein simulations. *Journal of Chemical Theory and Computation*, 7(4) :1157–1166, 2011.
- [248] M. Strajbl, G. Hong, and A. Warchel. Ab initio qm/mm simulation with proper sampling : First principle calculations of the free energy of the autodissociation of water in aqueous solution. *The journal of Physical Chemistry B*, 106(51) :13333–13343, 2002.
- [249] S. Strobl, C. Fernandez-Catalan, M. Braun, R. Huber, H. Masumoto, K. Nakagawa, A. Irie, H. Sorimachi, G. Bourenkova, H. Bartunik, K. Suzuki, and W. Bode. The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium. *Proceedings of the National Academy of Sciences*, 97(2) :588–592, 2000.
- [250] J. B. Sumner. The isolation and crystallization of the enzyme urease. *Journal of Biological Chemistry*, 69(2) :435–441, 1926.



- [251] T. Sun, X.-S. Wu, J. Xu, B. D. McNeil, Z. P. Pang, W. Yang, L. Bai, S. Qadri, J. D. Molkentin, D. T. Yue, and L.-G. Wu. The role of calcium/calmodulin-activated calcineurin in rapid and slow endocytosis at central synapses. *The Journal of Neuroscience*, 30(35) :11838–11847, 2010.
- [252] F. Tama and C. L. Brooks. Symmetry, form, and shape : Guiding principles for robustness in macromolecular machines. *Annual Review of Biophysics and Biomolecular Structure*, 35(1) :115–133, 2006.
- [253] F. Tama, F. X. Gadea, O. Marques, and Y.-H. Sanejouand. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins : Structure, Function, and Bioinformatics*, 41(1) :1–7, 2000.
- [254] F. Tama and Y.-H. Sanejouand. Conformational change of proteins arising from normal mode calculations. *Protein Engineering*, 14(1) :1–6, 2001.
- [255] R.-Y. Tan, Y. Mabuchi, and Z. Grabarek. Blocking the  $Ca^{2+}$ -induced conformational transitions in calmodulin with disulfide bonds. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 271(13) :7479–7483, 1996.
- [256] T. Tanaka, J. B. Amest, T. S. Harvey, L. Stryer, and M. Ikura. Sequestration of the membrane-targeting myristoyl group of recoverin in the calcium-free state. *Nature*, 376 :444–447, 1995.
- [257] H. L. Tepper and G. A. Voth. A coarse-grained model for double-helix molecules in solution : Spontaneous helix formation and equilibrium properties. *The Journal of Chemical Physics*, 122(12) :124906, 2005.
- [258] I. F. Thorpe, J. Zhou, and G. A. Voth. Peptide folding using multiscale coarse-grained models. *The Journal of Physical Chemistry B*, 112(41) :13079–13090, 2008.
- [259] M. Tirion. Large amplitude elastic motion in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77 :1905–1908, 1996.

- [260] M. M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77(9) :1905–1908, Aug 1996.
- [261] N. Tjandra, H. Kuboniwa, H. Ren, and A. Bax. Rotational dynamics of calcium-free calmodulin studied by <sup>15</sup>n-nmr relaxation measurements. *Eur. J. Biochem.*, 230 :1014–1024, 1995.
- [262] K. Torok, M. Wilding, L. Groigno, R. Patel, and M. Whitaker. Imaging the spatial dynamics of calmodulin activation during mitosis. *Current Biology*, 8 :692–699, 1998.
- [263] A. Tousignant and J. N. Pelletier. Protein motions promote catalysis. *Chemistry and Biology*, 11 :1037–1042, 2004.
- [264] V. Tozzini. Multiscale modeling of proteins. *Accounts of Chemical Research*, 43(2) :220–230, 2010.
- [265] S. Tripathi and J. J. Portman. Inherent flexibility and protein function : The open/closed conformational transition in the n-terminal domain of calmodulin. *The Journal of Chemical Physics*, 128 :205104, 2008.
- [266] S. Tripathi and J. J. Portman. Inherent flexibility determines the transition mechanisms of the ef-hands of calmodulin. *PNAS*, 106 :2104–2109, 2009.
- [267] V. M. Unger, N. M. Kumar, N. B. Gilula, and M. Yeager. Three-dimensional structure of a recombinant gap junction membrane channel. *Science*, 283 :1176–1180, 1999.
- [268] J. F. Varughese and Y. Chalovich J. M., Li. Molecular dynamics studies of troponin (tni-tnt-tnc) complexes : Insight into the regulation of muscle contraction. *Journal of Biomolecular Structure and Dynamics*, 28(2) :159–173, 2010.
- [269] M. Viceconti, D. Testi, F. Taddei, S. Martelli, G. J. Clapworthy, and S. V. S. Jan. Biomechanics modeling of the musculoskeletal apparatus : Status and key issues. *Proceedings of the IEEE*, 94(4) :725–739, april 2006.

- [270] H. B. Vickery. The origin of the world protein. *Yale J Biol Med.*, 22(5) :387–393, 1950.
- [271] D. Vigil, S. C. Gallagher, J. Trehwella, and A. E. Garcia. Functional dynamics of the hydrophobic cleft in the n-domain of calmodulin. *Biophysical Journal*, 80 :2082–2092, 2001.
- [272] E. Villa, A. Balaeff, L. Mahadevan, and K. Schulten. Multiscale method for simulating protein-dna complexes. *Multiscale Modeling and Simulation*, 2(4) :527–553, 2004.
- [273] M. V. Vinogradova, V. S. Reddy, A. S. N. Reddy, E. P. Sablin, and R. J. Fletterick. Crystal structure of kinesin regulated by  $ca^{2+}$ -calmodulin. *The journal of Biological Chemistry*, 279 :23504–23509, 2004.
- [274] J. T. Vivian and P. R. Callis. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal*, 80(5) :2093 – 2109, 2001.
- [275] C.-K. Wang, R. Liao, and H. C. Cheung. Rotational dynamics of skeletal muscle troponin c. *The Journal of Biological Chemistry*, 268 :14671–14677, 1993.
- [276] G. Wedemann and J. Langowski. Computer simulation of the 30-nanometer chromatin fiber. *Biophysical Journal*, 82(6) :2847 – 2859, 2002.
- [277] G. Wei, P. Derreumaux, and N. Mousseau. Sampling the complex energy landscape of a simple b-hairpin. *The Journal of Chemical Physics*, 119(13) :6403–6406, 2004.
- [278] G. Wei, N. Mousseau, and P. Derreumaux. Sampling the self-assembly pathways of kffe hexamers. *Biophysical Journal*, 87(6) :3648–3658, 2004.
- [279] G. Wei, W. Song, P. Derreumaux, and N. Mousseau. Self-assembly of amyloid-forming peptides using molecular dynamics simulations and the opep coarse-grained force field. *Frontiers in Biosciences*, 13 :5681–5692, 2008.

- [280] J. Weiser, P. S. Shenkin, and W. C. Still. Approximate solvent-accessible surface areas from tetrahedrally directed neighbor density. *Biopolymers*, 50 :373–380, 1999.
- [281] S. Wells, S. Menor, B. Hespeneide, and M. F. Thorpe. Constrained geometric simulation of constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.*, 2 :S126–S127, 2005.
- [282] H. Welzl and O. Stork. Cell adhesion molecules : Key players in memory consolidation ? *New in physiological sciences*, 18 :147–150, 2003.
- [283] G. Wider and K. Wüthrich. Nmr spectroscopy of large molecules and multimolecular assemblies in solution. *Current Opinion in Structural Biology*, 9(5) :594 – 601, 1999.
- [284] J. M. Withka, D. F. Wyss, A. A. R., E. L. Reinherz, and M. A. Recny. Structure of the glycosylated adhesion domain of human t lymphocyte glycoprotein cd2. *Structure*, 1(1) :69–81, 1993.
- [285] H. L. Woodcock, B. T. Miller, M. Hodoscek, A. Okur, J. D. Larkin, J. W. Ponder, and B. R. Brooks. Mscale : A general utility for multiscale modeling. *Journal of Chemical Theory and Computation*, 7(4) :1208–1219, 2011.
- [286] A. W. V. Wynsberghe and Q. Cui. Interpreting correlated motions using normal mode analysis. *Structure*, 14(11) :1647 – 1653, 2006.
- [287] C. Xu, J. Wang, and H. Liu. A hamiltonian replica exchange approach and its application to the study of side-chain type and neighbor effects on peptide backbone conformations. *Journal of Chemical Theory and Computation*, 4(8) :1348–1359, 2008.
- [288] D. Xu and R. Nussinov. Favorable domain size in proteins. *Folding and Design*, 3(1) :11 – 17, 1998.

- [289] Q. Xu, W. J. Brecht, K. H. Weisgraber, R. W. Mahley, and Y. Huang. Apolipoprotein e4 domain interaction occurs in living neuronal cells as determined by fluorescence resonance energy transfer. *279(24)* :25511–25516, 2004.
- [290] C. Yang, J. G. S., and K. K. Structure, dynamics and interaction with kinase targets : computer simulations of calmodulin. *Biophysica and Biophysica Acta*, 1697 :289–300, 2004.
- [291] L. Yang, C.-h. Tan, M.-J. Hsieh, J. Wang, Y. Duan, P. Cieplak, J. Caldwell, P. A. Kollman, and R. Luo. New-generation amber united-atom force field. *The Journal of Physical Chemistry B*, 110(26) :13166–13176, 2006.
- [292] L.-W. Yang, X. Liu, C. J. Jursa, M. Holliman, A. J. Rader, H. A. Karimi, and I. Bahar. ignm : a database of protein functional motions based on gaussian network model. *Bioinformatics*, 21(13) :2978–2987, 2005.
- [293] S. Yang and B. Roux. Src kinase conformational activation : Thermodynamics, pathways, and mechanisms. *PLoS Comput Biol*, 4(3) :e1000047, 03 2008.
- [294] K. L. Yap, J. B. Ames, M. B. Swindells, and M. Ikura. Diversity of conformational states and changes within the ef-hand protein superfamily. *PROTEINS : Structure, Function, and Genetics*, 37 :499–507, 1999.
- [295] K. L. Yap, J. Kim, K. Truong, M. Sherman, T. Yuan, and M. Ikura. Calmodulin target database. *Journal of structural and functional genomics*, 1 :8–14, 2000.
- [296] M. yi Shen, F. P. Davis, and A. Sali. The optimal size of a globular protein domain : A simple sphere-packing model. *Chemical Physics Letters*, 405(1-3) :224 – 228, 2005.
- [297] M. Yun, R. Lavery, M. Mousseau, K. Zakrzewska, and P. Derreumaux. Artists : An activated method in internal coordinate space for sampling protein energy landscapes. *PROTEINS : Structure, Function and Genetics*, 63 :967–975, 2006.

- [298] M.-R. Yun, N. Mousseau, and P. Derreumaux. Sampling small-scale and large-scale conformational changes in proteins and molecular complexes. *The Journal of Chemical Physics*, 126 :105101–105110, 2007.
- [299] B. Zhang, L. Rychlewski, K. Pawlowski, J. S. Fetrow, J. Skolnick, and A. Godzik. From fold predictions to function predictions : Automation of functional site conservation analysis for functional genome predictions. *Protein Science*, 8(5) :1104–1115, 1999.
- [300] Y. Zhang, D. Jasnow, and Z. M. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *PNAS*, 104(46) :18043–18048, 2008.
- [301] Y. Zhang, H. Tan, Y. Lu, Z. Jia, and G. Chen. Ca<sup>2+</sup> dissociation from the c-terminal ef-hand pair in calmodulin : A steered molecular dynamics study. *FEBS Letters*, 582 :1355–1361, 2008.
- [302] W. Zhong, J. P. Gallivan, Y. Zhang, L. Li, H. A. Lester, and D. A. Dougherty. From ab initio quantum mechanics to molecular neurobiology : A cation-pi binding site in the nicotinic receptor. *Proceedings of the National Academy of Sciences*, 95(21) :12088–12093, 1998.
- [303] R. Zhou, X. Huang, C. J. Margulis, and B. J. Berne. Hydrophobic collapse in multidomain protein folding. *Science*, 305(5690) :1605–1609, 2004.
- [304] Y. Zhou, W. Yang, M. Kirberger, H.-W. Lee, G. Ayalasomayajula, and J. J. Yang. Prediction of ef-hand calcium-binding proteins and analysis of bacterial ef-hand proteins. *Proteins : Structure, Function, and Bioinformatics*, 65 :643–655, 2006.
- [305] B. H. Zimm and J. K. Bragg. Theory of the phase transition between helix and random coil in polypeptide chains. *The Journal of Chemical Physics*, 31(2) :526–535, 1959.
- [306] R. Zwanzig. Two-state models of protein folding kinetics. *PNAS*, 94(1) :148–150, 1997.

## **Annexe I**

### **Curriculum vitae**

#### **Expériences d'assistanat d'enseignement (2006-2010) :**

Cours bin1001 - Introduction aux problèmes et méthodes de recherche en bio-informatique.

Cours bcm2003 - Application d'outils bio-informatiques.

Cours bin6003 – Architecture des polymères biologiques

Cours bcm2501 – Macromolécules biologiques

Cours ift1969 – Introduction à l'informatique scientifique

Cours ift1810 – Introduction à l'informatique

#### **Présentations orales :**

**15/11/2006** - Simulations multi-échelle de la flexibilité des protéines

Colloque Bio-Informatique Robert Cedergren

Colloque national organisé par l'université de Montréal

**15/02/2007** – Multiscale simulations of protein flexibility

Workshop on Dynamics under constraint II

Bellairs Research Institute of McGill University, Holetown, Barbados

Colloque international organisé par l'université McGill

**17/11/2007** - Multiscale simulations of protein flexibility (Holographic)

Colloque Bio-Informatique Robert Cedergren

Colloque national organisé par l'université de Montréal

**4/12/2007** – Pivots and flexibility in proteins

Présentation WEB (Netmeeting) à partir de l'université de Montréal

Sur le site « Flexweb – Analysis of Flexibility in Biomolecules and Networks »

Arizona State University (<http://flexweb.asu.edu/netmeetings/>)

**16/6/2011** – Overlapping High-level protein representations combined with atomistic non-biased forcefields for simulations

HPCS, UQAM, Montréal

**Présentations affiches :**

**15/5/2006** – Simulations of protein flexibility

Auteurs : Lilianne Dupuis, Normand Mousseau, Philippe Derreumaux

Rigidity, Flexibility, and Motion in Biomolecules workshop

Fiesta Inn Resort, Tempe, Arizona

Colloque international organisé par l'université d'Arizona

**3/8/2006** – Simulations of protein flexibility

Auteurs : Lilianne Dupuis, Normand Mousseau, Philippe Derreumaux

École d'été Réseau Québécois Étudiant des Matériaux de Pointe

Hotel Cheribourg, Magog

Colloque provincial organisé par le RQMP (Réseau québécois des matériaux de pointe)

**14/5/2007** - Finding pathways at different reality levels in simulations of Protein Flexibility

Auteurs : Lilianne Dupuis, Normand Mousseau, Philippe Derreumaux

Réunion annuelle du RQMP (Réseau québécois des matériaux de pointe)

Centre Mont-Royal, Montréal

**9/11/2007** - Multiscale simulations of protein flexibility

Auteurs : Lilianne Dupuis, Normand Mousseau

Colloque Bio-Informatique Robert Cedergren

Université de Montréal

Colloque national organisé par l'université de Montréal

**21-22/5/2008** - A holographic multi-scale approach for simulation of protein flexibility

Auteurs : Lilianne Dupuis, Normand Mousseau

GEPROM – Symposium annuel 2008

Les bases moléculaires et génétiques des arythmies cardiaques

Colloque international organisé par l'université de Montréal

**9/3/2011** - Multi-scale simulations of EF-hands proteins transformation with non-biased all atom potentials



Biophysical Society 55th Annual Meeting

Baltimore, USA

**Expériences de travail :**

**Hiver et automne 1983, automne 1984 et été 1985 :**

Stages d'étude universitaire :

- Ministère des transports, Montréal.
- Centre d'insémination artificielle de St Hyacinthe.
- Videoway Communication, Montréal (2 stages).

**1986-1987** : 1<sup>er</sup> emploi : Analyste programmeur chez Futurtex, Montréal.

Conception et réalisation d'un système reliant les pages de fournisseurs de contenus destinés à la câblodistribution. Langage C. Système DOS sur PC. Conception et programmation. Supervision de stagiaires. Rédaction de manuels.

**1987-1988** : Analyste programmeur chez Systèmes informatiques Phillips, Ville St-Laurent.

Conception d'un logiciel d'édition d'image à intégrer à un environnement multifenêtres.

Langage C. Conception orientée objet. Système Unix sur PC 386.

Conception. Supervision de stagiaires. Rédaction de manuels en anglais.

**1988-1993** : Analyste programmeur pour Videoway Communication, Montréal.

Conception et réalisation de systèmes clients-serveurs. Bases de données, télécommunications, Câblodistribution. Langage C. Ordinateurs Sun. Système Unix. Oracle. Environnement multi-tâches Utilisation de logiciels de conception modulaire (niveaux fonctionnel et organique).

**1994-1995** : Chargée de projet pour Videoway Communication, Montréal.

Mêmes environnements que ci haut. Développement de systèmes de diffusion sur le câble.

Différents niveaux de télécommunication. Logiciels de gestion de projets.

Supervision de 3 informaticiens. Répartition des tâches et intégration.

**1996-2003** : Retrait parental