



**HAL**  
open science

# Analyse de l'activation du facteur oedémateux de Bacillus anthracis par la calmoduline, en vue de la recherche d'inhibiteurs

Elodie Laine

► **To cite this version:**

Elodie Laine. Analyse de l'activation du facteur oedémateux de Bacillus anthracis par la calmoduline, en vue de la recherche d'inhibiteurs. Médicaments. Université Pierre et Marie Curie - Paris VI, 2009. Français. NNT: . tel-00483602

**HAL Id: tel-00483602**

**<https://theses.hal.science/tel-00483602>**

Submitted on 14 May 2010

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

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

**THESE DE DOCTORAT DE  
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité : **Sciences de la Vie**  
Ecole doctorale : **iVIV**

Présentée par

**Elodie Laine**

Pour obtenir le grade de

**DOCTEUR de L'UNIVERSITE PIERRE ET MARIE CURIE**

Sujet de la thèse :

**Analyse de l'activation du facteur oedémateux de *Bacillus anthracis*  
par la calmoduline, en vue de la recherche d'inhibiteurs**

Soutenue le 2 octobre 2009 devant le jury composé de :

Thérèse Malliavin	Directrice de thèse
Arnaud Blondel	Co-directeur de thèse
Massimo Marchi	Rapporteur
Bruno Villoutreix	Rapporteur
Claudine Mayer	Examinatrice



La virulence de la bactérie Gram<sup>+</sup> *Bacillus anthracis*, responsable de la maladie du charbon, est due à la présence d'une capsule et deux toxines. Chaque toxine résulte de l'assemblage de l'antigène protecteur (PA) avec l'un des deux facteurs, létal (LF) ou oedémateux (EF), dans le cytoplasme de la cellule hôte. EF est une adénylyl cyclase, qui transforme l'ATP en AMPc de manière incontrôlée, provoquant des dérèglements cellulaires. Elle est activée par la calmoduline (CaM), impliquée dans de nombreuses voies de signalisation du calcium. Des structures cristallographiques et une étude par *RMN* ont montré que la stabilité du complexe EF-CaM dépend du niveau de calcium fixé à CaM.

Des simulations de dynamique moléculaire du complexe, avec 0, 2 ou 4 ions calcium, ont permis de caractériser l'effet du calcium sur la plasticité conformationnelle des deux partenaires et de proposer un modèle de l'interaction EF-CaM. L'analyse conjointe des corrélations dynamiques et des influences énergétiques a fait émerger le concept de connexité du réseau de résidus comme critère de stabilité. La large transition conformationnelle induite chez EF par la fixation de CaM a été décrite, grâce à la détermination d'un chemin de réaction plausible, par modélisation moléculaire. Les conformations intermédiaires obtenues ont servi à guider la recherche rationnelle d'inhibiteurs de la toxine EF, dans le cadre d'une approche combinant méthodes computationnelles et expérimentales. Une stratégie innovante, impliquant le criblage virtuel d'une poche allostérique plutôt que du site catalytique de l'enzyme, a identifié six molécules actives, inhibant totalement l'activité de EF à des concentrations de 10-100  $\mu\text{M}$ .

**Mots clés :** *Bacillus anthracis*, facteur oedémateux, calmoduline, calcium, dynamique moléculaire, détermination de chemin de réaction, recherche rationnelle d'inhibiteurs, allostérie.

Ce travail a été réalisé d'octobre 2006 à octobre 2009, au sein de l'Unité de Bioinformatique Structurale (M. Nilges), Institut Pasteur, 25-28, rue du Dr. Roux, F-75015 Paris, sous la direction du Dr. Thérèse Malliavin et du Dr. Arnaud Blondel.

**Activation of the edema factor of *Bacillus anthracis* by calmodulin probed by molecular dynamics simulations : a new way to find inhibitors.**

The virulence of the Gram<sup>+</sup> bacterium *Bacillus anthracis*, responsible for anthrax disease, is caused by a capsule and two toxins. Each toxin is formed from the assembly of the protective antigen (PA) and either the lethal factor (LF) or the edema factor (EF) in the cytoplasm of host cells. EF is an adenylate cyclase, that produces cAMP from ATP in an uncontrolled fashion, provoking severe cellular dysfunction. EF is activated by calmodulin (CaM), involved in many calcium signaling pathways. Crystallographic structures and an *NMR* study showed that the level of calcium bound to CaM influences the stability of the EF-CaM complex.

Molecular dynamics simulations of the complex, with 0, 2 or 4 calcium ions, enabled to characterize the effect of calcium on the conformational plasticity of each partner and to propose a model for the EF-CaM interaction. The joint analysis of dynamical correlations and energetic influences raised the concept of residue network connectedness as a stability criterion. The large conformational transition undergone by EF upon CaM binding was described through the determination of a plausible reaction path. The obtained intermediate conformations were further used to guide the rational search for inhibitors of the EF toxin, in an approach combining computational and experimental methods. An innovative strategy involving the virtual screening of an allosteric pocket instead of the catalytic site of the enzyme, identified six active compounds able to fully inhibit EF activity at concentrations of 10-100  $\mu\text{M}$ .

**Key words :** *Bacillus anthracis*, edema factor, calmodulin, calcium, molecular dynamics, reaction path determination, rational search for inhibitors, allostery.

# Remerciements

Je voudrais tout d'abord remercier Michael Nilges, qui m'a accueillie au sein de son laboratoire, et m'a ainsi permis de travailler dans un environnement scientifique de très haute qualité. Je suis très reconnaissante envers Thérèse Malliavin, qui m'a guidée et m'a fait confiance tout au long de ces trois années. Ses qualités scientifiques, son enthousiasme sans faille et sa disponibilité ont joué un rôle déterminant dans l'accomplissement de cette thèse. Je remercie vivement Arnaud Blondel pour son engagement dans la co-direction de ce travail. Ses conseils avisés ont été d'une grande aide aux moments cruciaux de ma réflexion.

Les résultats présentés ici sont le fruit de multiples collaborations, notamment avec l'unité Biochimie des Interactions Macromoléculaires de l'Institut Pasteur, dirigée par Daniel Ladant. Merci à Christophe Goncalves et Johanna Karst pour leurs contributions. Je remercie également Alexandre Chenal, dont les indications et la documentation m'ont servi à rédiger ce manuscrit. Je tiens à exprimer ma gratitude envers Wei-Jen Tang, de l'Université de Chicago, qui a fourni la protéine EF. Merci à Aurélien Lesnard et Sylvain Rault, du Centre d'Etudes et de Recherches sur le Médicament de Normandie, pour nous avoir donné accès à leur chimiothèque. Enfin, je sais gré à Leandro Martínez et Julliane Yoneda de leur implication.

Mon travail quotidien à l'Institut Pasteur a été grandement facilité par Tru Huyn et Renée Communal. Je voudrais les remercier pour leur aide face aux difficultés informatiques et administratives. Je suis très heureuse d'avoir passé ces trois années au sein de l'équipe BIS et je remercie tous ses membres pour leur accueil chaleureux et nos discussions, en particulier Nathalie, David, Aymeric, Olivier, Claire et Benjamin.

Je remercie Elisabeth Labruyère-Dadaglio, pour nos entretiens réguliers, ainsi que toutes les personnes à l'Institut Pasteur qui m'ont permis de mener cette thèse dans des conditions de travail excellentes.

Je remercie les Pr. Jean Delettré et Germain Trugnan, directeurs successifs de l'école doctorale iVIV, pour leur suivi et pour m'avoir permis de bénéficier de formations parallèles.

Je suis reconnaissante envers les membres du jury de cette thèse, Massimo Marchi, Bruno Villoutreix et Claudine Mayer, pour l'intérêt qu'ils portent à mon travail.

Mes dernières pensées vont à ma famille, mes parents, mon frère et Simon, pour leur amour et leur soutien indéfectible. Merci à Pierre, Romain, Fanny, Nicolas, Karoll et Clio, qui m'ont accompagnée durant ces trois années.

*Cette thèse a bénéficié d'un financement de la Délégation Générale pour l'Armement (DGA/MRIS) du ministère de la Défense, géré par le Centre National de la Recherche Scientifique (CNRS).*



*Ce qui est simple est toujours faux.*

*Ce qui ne l'est pas est inutilisable.*

Paul Valéry



# Table des matières

<b>Abbreviations</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1 Contexte Biologique . . . . .	1
1.1 La maladie du charbon . . . . .	1
1.2 Risque bio-terroriste . . . . .	2
1.3 Facteurs de virulence, facteur oedémateux . . . . .	4
2 Aspects structuraux et problématiques . . . . .	6
2.1 La protéine EF . . . . .	6
2.2 La calmoduline et le calcium . . . . .	8
2.3 Le complexe EF-CaM . . . . .	9
3 Plan de l'étude . . . . .	11
<b>2 Approches méthodologiques</b>	<b>13</b>
1 La dynamique moléculaire . . . . .	13
1.1 Principe de la dynamique classique . . . . .	13
1.2 Champs de force . . . . .	15
1.3 Estimation de l'énergie libre : méthode <i>MMPBSA</i> . . . . .	17
1.4 Échantillonnage et transition conformationnelle . . . . .	19
2 Le criblage virtuel . . . . .	22
2.1 Principe . . . . .	22
2.2 Définition du site actif . . . . .	24
2.3 Algorithmes de <i>docking</i> . . . . .	25
FlexX . . . . .	25
ICM . . . . .	26
2.4 Fonctions de <i>scoring</i> . . . . .	27
FlexX . . . . .	28
ICM . . . . .	29
2.5 Défis et perspectives . . . . .	29
<b>3 Stabilité de l'Interaction EF-CAM</b>	<b>31</b>
1 Introduction . . . . .	31
2 Matériels et Méthodes . . . . .	32
2.1 Préparation des structures initiales . . . . .	32
2.2 Simulations de dynamique moléculaire . . . . .	32
2.3 Analyse des trajectoires . . . . .	34

2.4	Simulations LES . . . . .	35
3	Résultats et discussion . . . . .	35
3.1	Stabilité du complexe EF-CaM . . . . .	35
	Analyse de convergence . . . . .	36
	Dynamique Essentielle . . . . .	38
3.2	Plasticité conformationnelle des partenaires EF et CaM . . . . .	40
	Fluctuations atomiques . . . . .	40
	Caractéristiques géométriques de CaM . . . . .	41
	Conformation de EF et contacts avec CaM . . . . .	43
3.3	Inter-dépendance entre Ca <sup>2+</sup> /N-CaM et N-CaM/EF. . . . .	46
	Changements conformationnels de CaM et coordination des ions calcium . . . . .	46
	Modulation de l'affinité de N-CaM pour le calcium . . . . .	49
4	Conclusion . . . . .	53
<b>4</b>	<b>Aspects Dynamiques et Énergétiques du Complexe EF-CAM</b>	<b>55</b>
1	Introduction . . . . .	55
2	Matériels et Méthodes . . . . .	56
2.1	Modèle de réseau élastique . . . . .	56
2.2	Matrices de corrélations . . . . .	57
2.3	Analyse locale des traits caractéristiques . . . . .	58
2.4	Cartes de dépendances énergétiques . . . . .	59
3	Résultats . . . . .	61
3.1	Modèle de réseau élastique . . . . .	61
3.2	Analyse locale des traits caractéristiques . . . . .	62
3.3	Corrélations généralisées . . . . .	66
3.4	Cartes de dépendances énergétiques . . . . .	69
4	Discussion . . . . .	74
5	Conclusion . . . . .	75
<b>5</b>	<b>Recherche d'Inhibiteurs de EF</b>	<b>77</b>
1	Introduction . . . . .	77
2	Matériels et Méthodes . . . . .	78
2.1	Simulations et chemin de transition . . . . .	78
	Structures . . . . .	78
	Simulations <i>DM</i> . . . . .	78
	Chemin de transition conformationnelle . . . . .	78
2.2	Méthodes <i>In silico</i> de recherche d'inhibiteurs . . . . .	80
	Identification des poches . . . . .	80
	Préparation du récepteur et du ligand . . . . .	80
	Procédure de <i>docking</i> . . . . .	80
	Analyse bioinformatique des inhibiteurs . . . . .	81
2.3	Méthodes expérimentales . . . . .	82
	Ligands testés . . . . .	82
	Essai biochimique des composés . . . . .	83
	Essai enzymatique de l'adénylate cyclase . . . . .	83
	Test par anisotropie de fluorescence à l'état stationnaire . . . . .	84

3	Résultats . . . . .	84
3.1	Chemin de transition conformationnelle . . . . .	84
3.2	Criblage virtuel et sélection des ligands . . . . .	87
3.3	Caractérisation <i>in vitro</i> des inhibiteurs putatifs . . . . .	90
3.4	Analyse bioinformatique des inhibiteurs identifiés . . . . .	92
4	Discussion . . . . .	95
4.1	La détermination d'un chemin de transition est essentielle . . . . .	95
4.2	La poche SABC est stable au regard de la dynamique interne de EF . . . . .	96
4.3	La comparaison des structures de CyaA et de EF suggère un mode de liaison similaire sur les deux enzymes. . . . .	98
4.4	Les touches validées ont une action allostérique sur l'interaction EF-CaM . . . . .	100
5	Conclusion . . . . .	103
<b>6</b>	<b>Conclusion générale et perspectives</b>	<b>105</b>
	<b>Bibliographie</b>	<b>109</b>
	<b>Annexe</b>	<b>123</b>
A.1	Tableau des scores de <i>docking</i> des molécules sélectionnées <i>in silico</i> . . . . .	123
A.2	Publication décrivant la plasticité conformationnelle des partenaires de l'interaction EF-CaM . . . . .	125
A.3	Publication décrivant l'effet électrostatique du calcium sur le réseau de résidus du complexe EF-CaM . . . . .	143

# ABBREVIATIONS

<b>2D</b>	en deux dimensions
<b>3D</b>	en trois dimensions
<b>AC</b>	adénylate cyclase
<b>ACE2</b>	<i>analytic continuum electrostatics potential</i>
<b>ACP</b>	analyse en composantes principales
<b>ADN</b>	acide désoxyribonucléique
<b>AMP<sub>c</sub></b>	adénosine monophosphate cyclique
<b>ARN</b>	acide ribonucléique
<b>ATP</b>	adénosine triphosphate
<b>CaM</b>	calmoduline
<b>CPR</b>	<i>conjugate peak refinement</i>
<b>DM</b>	dynamique moléculaire
<b>EDM</b>	<i>energetic dependency maps</i>
<b>EF</b>	facteur oedémateux
<b>ENM</b>	<i>elastic network model</i>
<b>FFT</b>	<i>fast Fourier transform</i>
<b>GC</b>	<i>generalized correlations</i>
<b>HTS</b>	<i>high throughput screening</i>
<b>LES</b>	<i>locally enhanced sampling simulations</i>
<b>LF</b>	facteur létal
<b>LFA</b>	<i>local feature analysis</i>
<b>MC</b>	Monte Carlo
<b>MD</b>	<i>metadynamics</i>
<b>MMPBSA</b>	<i>molecular mechanics Poisson-Boltzmann surface area</i>
<b>PA</b>	antigène protecteur
<b>PDB</b>	<i>Protein Data Bank</i>
<b>Pi</b>	phosphate inorganique
<b>PME</b>	<i>particle mesh Ewald</i>
<b>PPase</b>	pyrophosphatase
<b>PPi</b>	pyrophosphate inorganique
<b>QSAR</b>	<i>quantitative structure-activity relationship</i>
<b>RE</b>	<i>replica exchange</i>
<b>RMN</b>	résonance magnétique nucléaire
<b>RMSD</b>	<i>root mean square deviation</i>
<b>RX</b>	cristallographie aux rayons X
<b>SASA</b>	<i>solvent accessible surface area</i>
<b>SAXS</b>	<i>small angle X-ray scattering</i>
<b>SMD</b>	<i>steered molecular dynamics simulations</i>
<b>SIFt</b>	<i>structural interaction fingerprint</i>
<b>VDA</b>	<i>virtual dihedral angle</i>

# Introduction

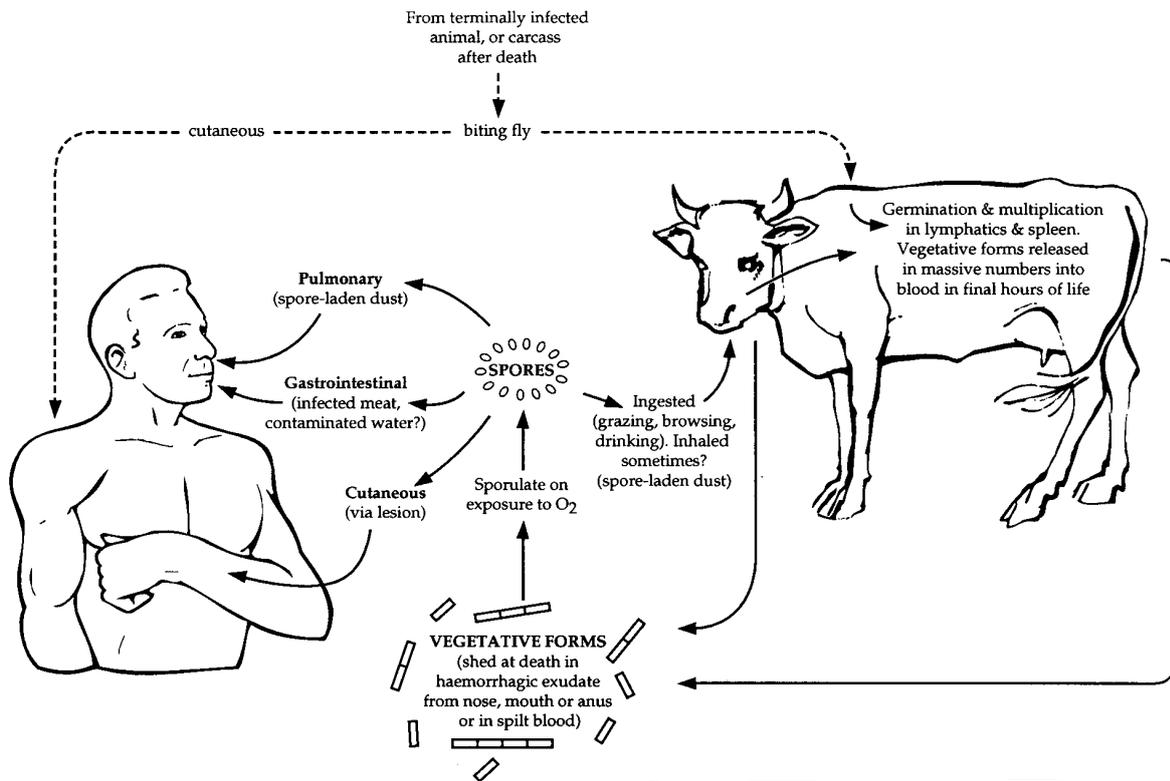
## 1 Contexte Biologique

### 1.1 La maladie du charbon

Le charbon ("anthrax" en anglais) est une maladie infectieuse aiguë, qui touche principalement les herbivores mais peut aussi affecter l'Homme [122]. Connue depuis l'antiquité, elle doit son nom au sang noir, poisseux et incoagulable des animaux morts contaminés. Le médecin Casimir Davaine établit l'étiologie de la maladie en 1863, mettant ainsi pour la première fois en évidence l'origine bactérienne d'une maladie animale et humaine [68; 67]. L'étude du charbon par Louis Pasteur permit de confirmer le rôle des germes dans les maladies et de populariser le concept de la vaccination, notamment à travers une expérience publique sur une cinquantaine de moutons, à Pouilly-le-Fort en 1881 [71]. Il faudra attendre 1955, pour qu'une toxine du charbon soit mise en évidence [232].

L'agent infectieux de la maladie est la bactérie sporulante à Gram positif *Bacillus anthracis* [189], qui appartient au groupe *Bacillus cereus* des Bacillaceae. Comme les autres bactéries du genre *Bacillus*, *B. anthracis* est observée sous deux formes : la spore, forme résistante, et le bacille, forme végétative. Le bacille est immobile (absence de flagelle), mesure 1.0-1.5  $\mu\text{m}$  de diamètre sur 3-5  $\mu\text{m}$  de longueur et forme des chaînes à l'aspect "tiges de bambou". La spore est la forme infectante, capable de germer dans les tissus d'un hôte mammifère.

Il existe trois voies de contamination chez l'Homme [76] (voir cycle infectieux Fig. 1) : (i) le charbon cutané, à travers une plaie, est la forme la plus fréquente (90-95 % des cas) et se caractérise par l'apparition d'une papule rouge, qui évolue en une vésicule purigineuse, puis se transforme en escarre noirâtre ; (ii) le charbon intestinal, causé par l'ingestion d'aliments contaminés, provoque une inflammation aiguë du tube digestif ; (iii) le charbon pulmonaire, suite à l'inhalation de spores, se traduit par un syndrome grippal, suivi par une aggravation brutale avec d'importants troubles respiratoires, toux, maux de tête, douleurs musculaires, qui évoluent vers un choc septique. Cette dernière forme est quasiment toujours mortelle en l'absence de traitement, avec une dose létale 50 % de l'ordre de  $10^4$  spores. A ce jour, aucun cas de transmission inter-humaine n'a été recensé.



**FIG. 1: Cycle infectieux de l'anthrax.** La spore joue un rôle central dans le cycle, même si l'infection peut aussi être acquise via l'absorption des formes végétatives lorsque, par exemple, les Hommes ou les carnivores mangent la viande d'un animal mort du charbon ou lorsque les mouches transmettent la maladie. Ce schéma est extrait du rapport de l'*Organisation Mondiale pour la Santé* [252].

La maladie du charbon a pratiquement disparu en France, mais elle reste endémique ou sporadique dans les pays d'Afrique subsaharienne et d'Asie, en Europe du Sud, en Amérique Centrale et du Sud et dans certaines régions d'Australie. L'antibiothérapie entraîne en général une guérison rapide, à condition qu'elle soit administrée très tôt (avant l'apparition des premiers symptômes, ou juste après). La ciprofloxacine et la doxycycline sont utilisées pour la thérapie et la prophylaxie antibiotique de la maladie. S'il existe des vaccins pour animaux efficaces, les vaccins à usage humain sont moins performants et sont réservés aux groupes à hauts risques [45]. Conçus dans les années 1960, ils sont très réactogènes [95], doivent être administrés en six injections [129], et protègent insuffisamment contre le charbon pulmonaire. Récemment, l'*Unité Toxines et Pathogénie Bactérienne*, à l'Institut Pasteur, responsable du *Centre National de Référence du Charbon*, a mis au point une composition vaccinale efficace, susceptible d'être utilisée chez l'Homme [39; 98].

## 1.2 Risque bio-terroriste

Les spores de la bactérie *Bacillus anthracis* se propagent facilement dans l'air et dans l'eau et sont très résistantes à la chaleur, à la dessiccation, aux rayonnements ultra-violets et aux

désinfectants [275; 235]. Elle peuvent survivre plusieurs décennies sous la terre et contaminer le bétail broutant à la faveur de conditions plus favorables. Ces qualités font de *B. anthracis* un agent infectieux potentiellement dangereux pour la santé publique, s'il est disséminé de façon malveillante. De fait, il est considéré comme un agent biologique représentant un risque bio-terroriste, de catégorie A selon la classification proposée par les *Centers for Disease Control and Prevention* (CDC) et le *National Institute of Allergy and Infectious Diseases* (NIAD) américains [216; 129]. Un certain nombre de micro-organismes sont ainsi classés en fonction de critères, tels que le taux de morbidité, de mortalité, la capacité de production, de dissémination... Les agents de catégorie A constituent les risques potentiels les plus élevés (voir classification Table I).

Maladie	Impact sur la santé publique		Potentiel de dissémination		Perception publique <sup>d</sup>	Préparation spéciale <sup>e</sup>	Catégorie
	Morbidité	Mortalité	P-D <sup>b</sup>	P - P <sup>c</sup>			
Coqueluche	+	++	+	++++	+++	+++	A
Anthrax	++	+++	+++	0	+++	+++	A
Peste <sup>f</sup>	++	+++	++	++	++	+++	A
Botulisme	++	+++	++	0	++	+++	A
Tularémie	++	++	++	0	+	+++	A
FHV <sup>g</sup>	++	+++	+	+	+++	++	A
EV <sup>h</sup>	++	+	+	0	++	++	B
Fièvre Q	+	+	++	0	+	++	B
Brucellose	+	+	++	0	+	++	B
Morve	++	+++	++	0	0	++	B
Mélioïdose	+	+	++	0	0	++	B
Psittacose	+	+	++	0	0	+	B
Ricine	++	++	++	0	0	++	B
Typhus	+	+	++	0	0	+	B
Choléra <sup>i</sup>	+	+	++	+/-	+++	+	B
Shigellose <sup>i</sup>	+	+	++	+	+	+	B

<sup>a</sup> Les agents ont été rangés de la menace la plus sérieuse (++++) à la moins sérieuse (0).

<sup>b</sup> Potentiel de production et dissémination en quantités susceptibles d'affecter une large population, basé sur la disponibilité, les exigences de bio-sécurité BSL, le mode d'infection le plus efficace, et la stabilité environnementale.

<sup>c</sup> Transmission de personne à personne.

<sup>d</sup> Niveau de perception publique, évalué à partir du nombre de fois où un agent ou une maladie est apparu dans les médias.

<sup>e</sup> Niveau de préparation spéciale en terme de santé publique, évalué en considérant les besoins en stockage de médicaments, en surveillance et en prévention, ainsi qu'en capacités de diagnostic.

<sup>f</sup> Peste pneumonique.

<sup>g</sup> Fièvres hémorragiques virales dues aux Filovirus (Ebola, Marburg) ou aux Arenavirus (par ex. Lassa, Machupo).

<sup>h</sup> Encéphalites virales.

<sup>i</sup> Exemples de maladies transmises par la nourriture ou l'eau.

**TAB. I: Critères et poids<sup>a</sup> utilisés pour évaluer les agents constituant une menace biologique potentielle.** Ce tableau est extrait du rapport du CDC [216].

Le bacille du charbon a été rendu tristement célèbre lors d'une série d'attentats à la lettre piégée en 2001, aux États-Unis. A cette occasion, parmi les 22 personnes qui développèrent la maladie, 11 furent contaminées par la forme pulmonaire, avec un taux de mortalité de 45% [20]. Ces événements ont fait prendre conscience de la menace bio-terroriste et de la nécessité d'une préparation spécifique des États pour gérer les crises liées aux maladies infectieuses, émergentes ou ré-émergentes [205].

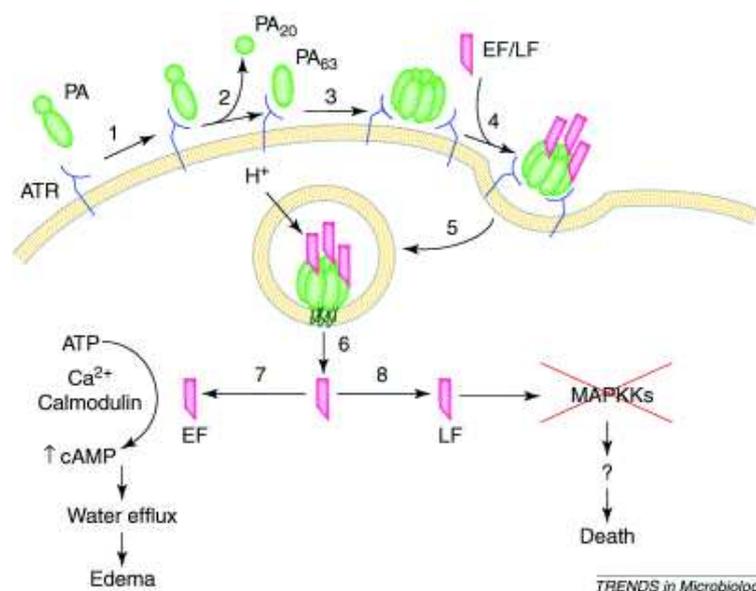
Auparavant, *B. anthracis* avait été militarisée à plusieurs reprises [129] : (i) par les Japonais contre les Chinois, au sein des unités 100 et 731 entre 1932 et 1945 [115], (ii) par le Royaume-Uni qui testa le charbon sur l'île écossaise de Gruinard pendant la Seconde Guerre Mondiale [175; 176], (iii) par l'armée irakienne dans sa lutte contre les rebelles kurdes en 1987-8 [291]. Par ailleurs, un accident provoqua la libération d'un nuage d'anthrax d'une usine d'armes biologiques à Sverdlovsk, en URSS, en 1979.

La *Convention sur l'interdiction des armes biologiques* (CABT), ratifiée en 1975, est le premier traité multilatéral de désarmement à bannir une catégorie entière d'armes. Elle oblige notamment les 171 États parties contractantes à ne pas mettre au point, fabriquer, stocker, acquérir ni conserver des agents micro-biologiques ou autres agents biologiques, comme le bacille du charbon, en quantités qui ne sont pas destinées à des fins prophylactiques, de protection ou à d'autres fins pacifiques.

### 1.3 Facteurs de virulence, facteur oedémateux

La maladie du charbon survient après germination des spores à l'intérieur des macrophages et dissémination des bacilles dans l'organisme [189]. *B. anthracis* produit un nombre de facteurs de virulence, critiques pour l'établissement de l'infection et la pathogénèse [58]. Parmi les principaux facteurs, la capsule, formée d'un polymère d'acide D-glutamique, permet à la bactérie d'échapper à la phagocytose [78]. Sa synthèse est régulée par le plasmide pXO2 (60 MDa). La bactérie sécrète aussi deux toxines de type A-B, codées par des gènes du plasmide pXO1 (110 MDa). L'antigène protecteur (PA) joue le rôle de facteur d'adhésion cellulaire (sous-unité B). PA<sub>83</sub> se fixe sur un récepteur ATR de la cellule hôte, où il est clivé par des protéases cellulaires, libérant ainsi son fragment N-terminal PA<sub>20</sub>. Le fragment actif PA<sub>63</sub> s'oligomérisse en heptamères, qui chargent chacun jusqu'à trois sous-unités A pour les transloquer dans le cytosol, par endocytose. La sous-unité A peut être le facteur létal (LF), une métallo-protéase [267], ou le facteur oedémateux (EF), une adénylate cyclase (AC) [63]. La description historique de la pathogénicité de *B. anthracis* identifiait LF comme responsable de la létalité associée au charbon systémique, et EF comme responsable de la formation d'oedème [189; 58]. Cependant, des observations récentes [188; 102; 88] suggèrent qu'il existe une coopération entre les deux toxines [40]. Le fait qu'elles soient co-régulées et qu'elles ciblent des types cellulaires similaires pourrait impliquer qu'elles détériorent la fonction cellulaire et sapent les défenses immunitaires de l'hôte en synergie [25; 250].

**FIG. 2: Modèle du mode d'action de la toxine du charbon.** PA se fixe sur un récepteur membranaire ubiquitaire (ATR) (1), puis est clivé par une protéase membranaire (2). Tandis que la sous-unité PA<sub>20</sub> (20 kDa) est relarguée dans le milieu extra-cellulaire, PA<sub>63</sub> (63 kDa) s'oligomérisé pour former un heptamère (3). EF et/ou LF se fixe(nt) sur l'heptamère (4), puis la toxine assemblée est endocytée (5) et acheminée vers les endosomes. Là, le pH bas déclenche un changement conformationnel dans l'heptamère, promouvant la formation d'un pore sélecteur de cation et la translocation des entités enzymatiques à travers la membrane (6). L'adénylate cyclase EF provoque une augmentation de la concentration d'AMPc, résultant en un afflux d'eau et un oedème (7). La métallo-protéase LF clive les kinases de kinases activées par les agents mitogènes (MAPKKs). Ces actions initient une série d'événements qui conduisent à la mort de l'hôte (8). Ce schéma est extrait de la revue de Mourez et al. de 2001 [194].



De nombreuses bactéries pathogènes sécrètent des toxines, dont l'action consiste à augmenter la quantité d'AMP cyclique intracellulaire [11]. Le facteur oedémateux EF, qui nous intéresse ici, est activé par la calmoduline (CaM) en tant qu'adénylate cyclase (AC) dans le cytoplasme de la cellule hôte. Il transforme l'ATP, réserve d'énergie principale de la cellule, en AMPc de manière incontrôlée [170]. L'accumulation d'AMPc provoque des dérèglements cellulaires, à travers une série de mécanismes qui restent peu connus. On sait toutefois que, dans les macrophages, l'AMPc inhibe l'activité du facteur de transcription NF- $\kappa$ B, qui normalement stimule la production de plusieurs facteurs inflammatoires [198; 18; 9; 61]. L'action de LF sur les kinases de kinases MAPKK coupe également des voies de signalisation qui convergent vers l'activation de NF- $\kappa$ B, ce qui suggère une route par laquelle les deux toxines collaborent pour saper les défenses immunitaires de l'hôte et promouvoir la mort des macrophages. Par ailleurs, Raymond et al. [211] ont récemment montré que EF réprime l'expression par les macrophages de la phospholipase sPLA2-IIA, un composant majeur de la réponse immune ayant des propriétés bactéricides.

Beaucoup d'efforts ont été investis dans le développement de vaccins à base de PA [143; 104] et dans la recherche d'inhibiteurs de LF [64; 251; 137] ou ciblant le pore PA [106]. Cependant, les déterminants moléculaires à l'origine de la pathogénicité de la bactérie ne sont pas complètement caractérisés et beaucoup de questions demeurent quant au mode d'action des toxines de *B. anthracis*. Les récentes études mettant en évidence le rôle important de EF, et révélant une synergie d'action entre les deux facteurs oedémateux et létal, ouvrent la voie à de nouvelles approches pharmacologiques pour traiter le charbon et font de EF une cible potentielle intéressante.

L'étude de l'activité enzymatique de EF [109; 110; 135] est allée de pair avec la recherche d'inhibiteurs compétitifs de l'ATP : Soelaiman et al. [233] ont découvert un inhibiteur de l'activité de EF par criblage *in vitro* d'une banque de petites molécules ( $K(i) = 20 \mu M$ ); Shen et al. [227] ont révélé que l'adéfovir, utilisé dans le traitement contre l'hépatite B, est aussi actif sur EF ( $K(i) = 27 nM$ ). Une petite molécule inhibitrice de l'interaction protéique EF-CaM a également été identifiée parmi une banque de 100 000 composés [168]. Enfin, récemment, une approche de recherche de médicament basée sur la structure a permis de découvrir quatre nouveaux inhibiteurs, avec des  $IC_{50}$  de l'ordre de quelques  $\mu M$  [50].

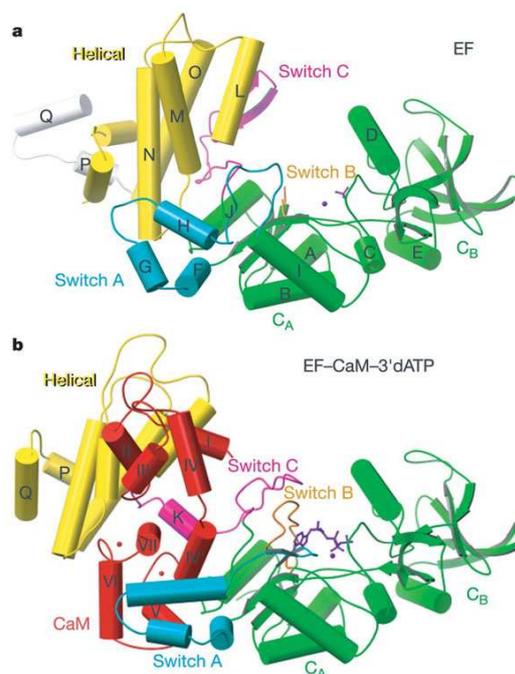
## 2 Aspects structuraux et problématiques

### 2.1 La protéine EF

La protéine EF se compose d'un domaine N-terminal de fixation à PA (PABD, résidus 64-290) et d'une région C-terminale catalytique (ACD, résidus 291-800) [77; 228]. PABD est homologue de séquence et de structure au domaine équivalent chez LF [155]. Le coeur catalytique diffère des adénylates cyclases (AC) mammifères, et présente une homologie avec CyaA de *Bordetella pertussis* et ExoY de *Pseudomonas aeruginosa*. Néanmoins, le domaine hélicoïdal Hel (résidus 660-800) de ACD, serré contre le coeur catalytique (résidus 294-622) dans l'état inactif, est unique à EF.

EF est très efficace pour catalyser la production d'AMPc, à un taux de  $1-2 \text{ m.s}^{-1}$  largement plus élevé que les AC mammifères [242; 226], mais seulement lorsqu'il est lié à la calmoduline (CaM). Il est probable qu'une telle activation *in situ* constitue une protection de la bactérie contre sa propre toxine. Le choix de la calmoduline peut s'expliquer de plusieurs manières [170] : (i) CaM est très abondante dans le cytoplasme de toutes les cellules supérieures et a été hautement conservée au cours de l'évolution ; (ii) elle est le principal médiateur de la signalisation cellulaire du calcium. Les premières structures, disponibles dans la *RCSB Protein Data Bank* (PDB) [31], de EF-ACD seul (entrée PDB : 1K8T) et EF-ACD en complexe avec la calmoduline (entrée PDB : 1K93) ont été résolues par cristallographie aux rayons X (*RX*) en 2002, par Drum et al [77] (voir structures de EF, Fig. 3). Lors de l'interaction avec la calmoduline, EF subit une large transition conformationnelle qui aboutit à la formation du site catalytique. Le domaine Hel effectue une translation de  $15 \text{ \AA}$  et une rotation de  $30^\circ$  par rapport au coeur catalytique, ouvrant un espace dans lequel CaM va s'insérer.

**FIG. 3: Structures secondaires de EF.** (a) EF libre ; (b) le complexe EF-CaM-3'-deoxy-ATP. CaM (en rouge) est maintenue d'un côté par une projection en forme de doigt de CA (en bleu), et de l'autre côté par la boucle SC (en magenta, dénotée *Switch C*) et le domaine Hel (en jaune, dénoté *Helical*). Les boucles SA (dénotée *Switch A*) et SB (dénotée *Switch B*) sont en cyan et en orange, respectivement. L'analogue 3'-deoxy-ATP (en violet) et l'ion métallique (en violet) sont fixés entre les domaines CA et CB (en vert), qui forment le site actif. Dans la structure de EF libre, les deux dernières hélices HP et HQ (résidus 771-799, en blanc), qui ne font partie ni de l'interface avec CaM ni du site actif, sont échangées entre deux molécules de EF. Cet échange n'a probablement pas d'importance physiologique, EF étant un monomère d'après ses propriétés en gel de filtration. Cette figure est extraite de la publication des structures 1K8T et 1K93 de EF [77].



La réaction catalytique de EF, médiée par l'attaque de l'oxygène 3' du phosphate-( $\alpha$ ), est semblable à la réaction de cyclisation des AC mammifères, ou à la réaction de transfert nucléotidyle des polymérases de l'ADN et de l'ARN [237; 243; 109]. Pour promouvoir cette réaction, EF doit dans un premier temps fixer l'ATP, puis faciliter la déprotonation de l'hydroxyle OH en 3', stabiliser l'intermédiaire phosphorane penta-coordonné, et enfin libérer les produits AMPc et pyrophosphate (PPi) [228]. Le mécanisme enzymatique de la transformation d'ATP en AMPc par EF n'est à ce jour pas clarifié. En 2005, une structure cristallographique du complexe EF-CaM (entrée PDB : 1XFV), résolue par Shen et al. [228], a permis de proposer un mécanisme catalytique faisant intervenir deux ions métalliques  $Mg^{2+}$  : le premier agirait de concert avec l'histidine H351 pour faciliter la déprotonation de l'hydroxyle OH en 3' ; le second favoriserait la rupture de la liaison entre les phosphates ( $\alpha$ ) et ( $\beta$ ). Ce mécanisme de catalyse à deux ions métalliques est également employé par les AC mammifères. Cependant, d'autres structures disponibles dans la PDB présentent des configurations du site catalytique différentes, quant à la conformation du substrat (3'-deoxy-ATP, dans les structures) ou des produits (AMPc et PPi), ou quant aux modes de liaison des ions  $Mg^{2+}$  impliqués dans la catalyse. Notamment, certaines structures ne comportent qu'un seul ion lanthanide  $Yb^{3+}$  (utilisé pour la cristallisation) dans le site catalytique [77; 109]. Aussi, une observation détaillée de la structure 1XFV montre que : (i) la conformation du 3'-deoxy-ATP et les positions des ions  $Mg^{2+}$  dans le site diffèrent de celles observées chez les AC mammifères ; (ii) les conditions de cristallisation impliquent une très grande concentration de  $Mg^{2+}$  (200 mM  $MgCl_2$ ), qui serait suffisante pour inhiber l'activité de EF-CaM ; (iii) la carte de densité électronique autour des ions  $Mg^{2+}$  est relativement ambiguë. Toutes ces informations semblent contredire, ou au moins jettent le doute sur l'hypothèse d'un mécanisme commun aux

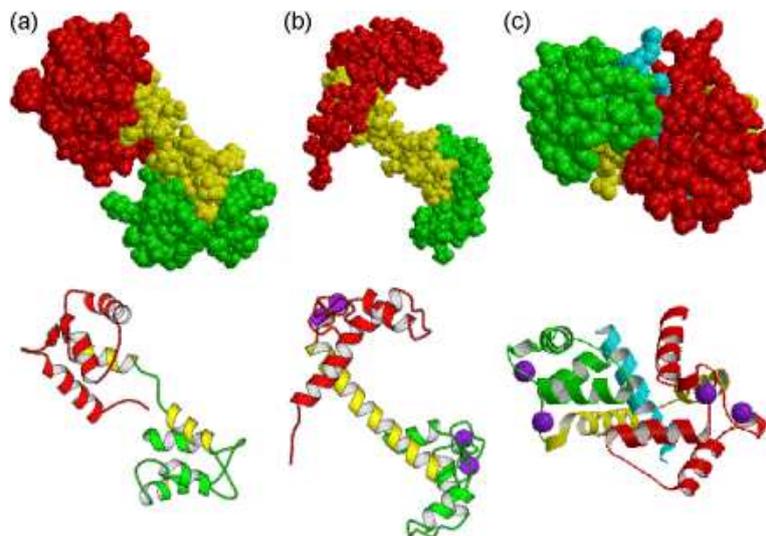
AC et à EF [180].

## 2.2 La calmoduline et le calcium

La calmoduline (CaM) est une petite protéine flexible, ubiquitaire des cellules eucaryotes. Senseur du calcium [186; 268; 258; 55], elle est le membre le plus représentatif de la super-famille des protéines « EF-hand ». Elle joue un rôle crucial dans de nombreuses voies de signalisation du calcium et est impliquée dans la survie, la croissance, la prolifération et la différenciation cellulaire. En réponse à différents stimuli extra-cellulaires, les niveaux de calcium s'élèvent, d'une concentration de base de  $10^{-7}$  M jusqu'à  $6.10^{-6}$  M. En fixant plus ou moins d'ions  $\text{Ca}^{2+}$ , CaM est capable de réguler plus de trente cibles protéiques diverses [256; 62; 276]. Sa plasticité remarquable en fait un bon sujet pour étudier les implications de la flexibilité structurale des protéines [166; 202; 51; 93]. Par ailleurs, elle a joué un rôle historique dans le développement des méthodes par résonance magnétique nucléaire (*RMN*) multi-dimensionnelles [128; 126; 127; 28; 96].

Les 148 résidus de CaM sont arrangés en deux domaines globulaires, N-terminal (N-CaM, résidus 5-64) et C-terminal (C-CaM, résidus 93-147), reliés par un « linker » central flexible [22]. Chaque domaine consiste en une paire de sites de fixation au calcium en hélice-boucle-hélice, dits motifs « EF-hand ». C-CaM a une affinité plus grande pour le calcium que N-CaM. La fixation coopérative d'ions  $\text{Ca}^{2+}$  sur chacun des domaines induit un changement conformationnel, faisant passer les motifs « EF-hand » d'une conformation fermée à ouverte. Cette ouverture provoque l'exposition au solvant de patches hydrophobes, qui participent à la reconnaissance moléculaire des cibles de CaM [284; 130].

**FIG. 4: Représentation schématique des trois formes structurales de la calmoduline.** (a) Structure de apo-CaM (entrée PDB : 1CFD) [154]; (b) Structure de Ca-CaM (entrée PDB : 1CLL) [47]; (c) Structure du complexe entre CaM et le peptide smMLCK (entrée PDB : 1CDL) [184]. Chaque structure est dessinée de manière schématique, en représentation « cartoon ». N-CaM est en rouge, C-CaM est en vert, le « linker » central est en jaune, les ions calcium sont en violet, le peptide smMLCK est en cyan. Cette figure est extraite de l'étude de CaM par simulations de dynamique moléculaire, réalisée par Yang et al. [284].

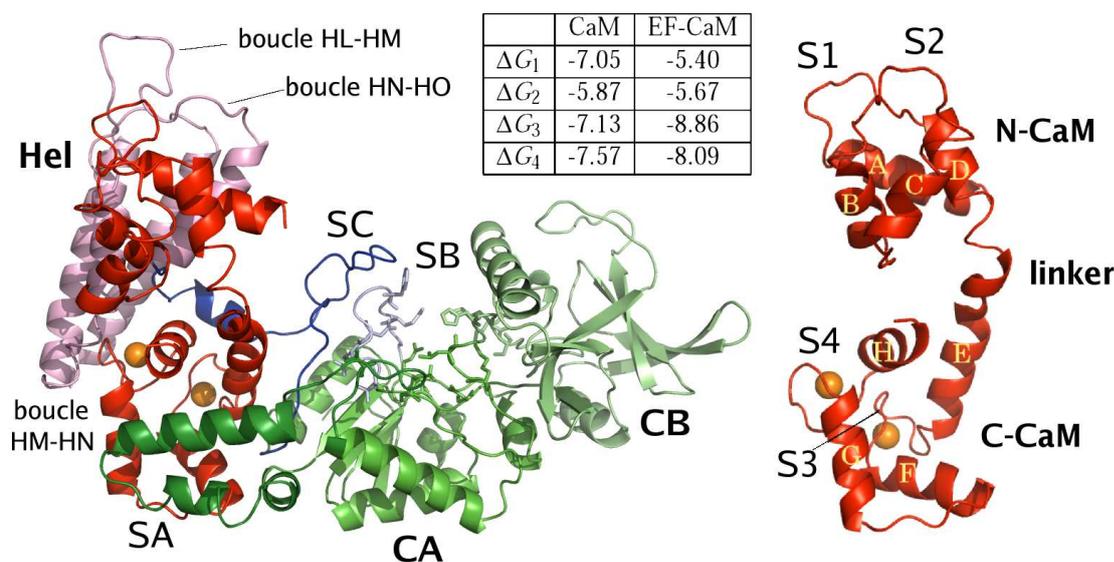


Les premières structures cristallographiques résolues de la calmoduline mammifère montraient la protéine libre en forme de cloche, avec ses deux domaines ouverts liés par une hélice droite [21; 47], et une conformation compacte pour la protéine en complexe avec des peptides cibles [184] (voir

structures de CaM, Fig. 4). Cependant, des simulations de dynamique moléculaire [282; 229; 87], des expériences par *RMN* [56; 33] et une récente étude à l'échelle de la molécule unique [136] démontrent que la CaM libre est en fait très flexible et qu'il existe probablement plusieurs états conformationnels en solution. De manière similaire, des structures récentes [281; 83] ont mis en évidence une variété de sous-états en environnement cristallin.

Généralement, quand CaM interagit avec une protéine cible, les deux lobes sont en contact, formant un tunnel hydrophobe dans lequel est engouffré le peptide de fixation de la cible [125; 184; 185; 286; 86; 130]. Le motif de reconnaissance canonique a été décrit comme une hélice  $\alpha$  amphiphatique [214; 24]. Toutefois, des structures récentes [81; 222; 46; 285; 134] ont révélé des modes d'interaction CaM-cibles inattendus [120; 264]. Parmi ses structures, celle du complexe EF-CaM révèle CaM en conformation étendue, partageant avec EF une surface d'interaction enfouie particulièrement large de  $5\,900\text{ \AA}^2$  [77].

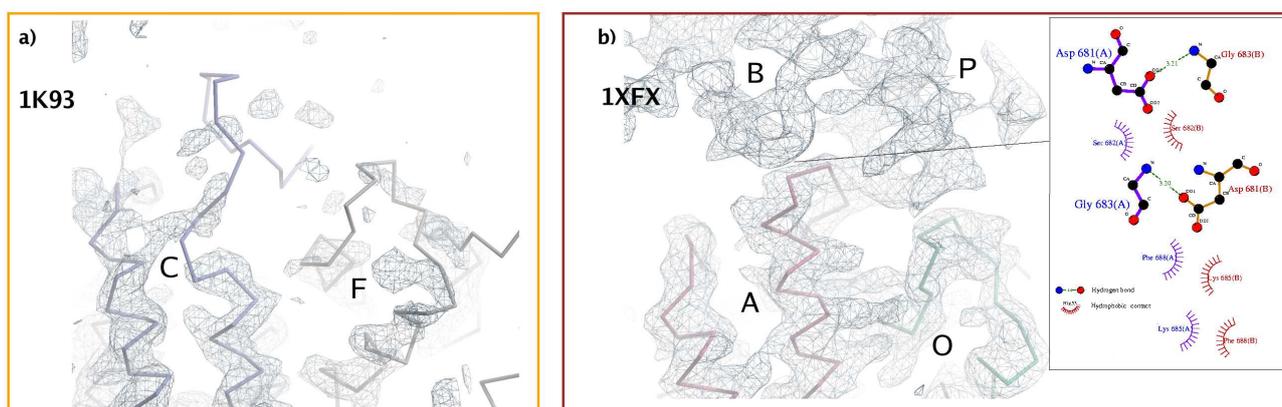
### 2.3 Le complexe EF-CaM



**FIG. 5: Structure du complexe EF-CaM.** La structure cristallographique du complexe EF-CaM [77] est montrée en représentation « cartoon ». La calmoduline (*en rouge*) est chargée de deux ions calcium (*sphères en orange*) sur son domaine C-terminal. (a) Le domaine Hel (résidus 660-767) de EF (*en rose*) est composé des hélices HL, HM, HN et HO. CA est représenté en vert (résidus 294-349 et 490-622), avec la boucle SA en vert foncé (résidus 502-551) et la boucle SB en bleu clair (résidus 579-591), et CB est représenté en lime (résidus 350-489). A l'interface entre CA et CB, les résidus catalytiques sont indiqués en représentation « licorice ». La boucle SC est représentée en bleu (résidus 623-659). (b) CaM est composée de huit hélices : HA, HB, HC dans N-CaM, HD et HE dans le « linker » central et HF, HG, HH dans C-CaM. Les boucles de liaison au calcium sont localisées entre les hélices HA et HB dans le site 1 (S1), les hélices HC et HD dans le site 2 (S2), les hélices HE et HF dans le site 3 (S3) et les hélices HG et HH dans le site 4 (S4). S3 et S4 sont chargés d'ions  $\text{Ca}^{2+}$  (*sphères en orange*). Le tableau en insert donne en  $\text{kcal.mol}^{-1}$  les affinités des quatre sites de CaM pour le calcium, calculées à partir des données de Ulmer et al. [255]. La visualisation a été réalisée avec Pymol [72].

La première structure de EF-CaM (entrée PDB : 1K93) [77] contient la région catalytique ACD de EF, en complexe avec CaM chargée de deux ions  $\text{Ca}^{2+}$  sur son domaine C-terminal (Fig. 5). La calmoduline est insérée entre le domaine Hel (résidus 660-800) et le coeur catalytique (résidus 294-622), enserrée par la boucle SC (résidus 623-659). Les résidus catalytiques sont situés à l'interface des domaines CA (résidus 294-349, 490-622) et CB (résidus 350-489). Dans le mécanisme de l'interaction décrit par Drum et al. [77], c'est le domaine N-terminal (N-CaM) qui initie le contact avec le domaine Hel de EF. Le domaine C-CaM établit ensuite des interactions avec la boucle SA (résidus 502-551), située à l'interface entre le domaine Hel et le coeur catalytique dans la forme inactive de EF. S'ensuivent des réarrangements de la boucle SA et de la boucle SC (résidus 630-659), celle-ci venant stabiliser la boucle SB (résidus 578-591), partie intégrante du site catalytique dans le complexe.

D'autres structures, résolues en 2005 par Shen et al. [228], contiennent EF intact, incluant son domaine PABD, situé à plus de 40 Å de la zone d'interface avec EF. CaM est alors chargée d'un nombre variable d'ions calcium : deux  $\text{Ca}^{2+}$  (entrée PDB : 1XFY), trois  $\text{Ca}^{2+}$  (entrée PDB : 1XFZ) ou quatre  $\text{Ca}^{2+}$  (entrée PDB : 1AFX). L'influence de cette complexation au calcium, sur la stabilité et même la formation de EF-CaM, a été mise en évidence par une étude *RMN* de CaM, libre et liée à EF [255]. Le complexe chargé de deux ions  $\text{Ca}^{2+}$ , fixés sur C-CaM, apparaît comme la forme prépondérante, et la présence de EF modifie l'affinité de CaM pour le calcium. Typiquement, l'affinité des sites 3 et 4 de C-CaM augmente, alors que celle des sites 1 et 2 de N-CaM diminue (voir insert, Fig. 5). Shen et al. [228] ont proposé une explication atomique de l'affinité réduite de N-CaM à partir de l'analyse structurale de EF-CaM. Il est intéressant de noter, toutefois, que malgré cette influence du calcium, les structures cristallographiques du complexe sont globalement similaires, l'architecture globale de EF-CaM étant conservée quel que soit le niveau de calcium.



**FIG. 6: Boucle HL-HM du domaine Hel de EF (résidus 675-684) en interaction avec N-CaM. (a-b) Cartes de densité électronique contourées à  $1.10 \sigma$  des chaînes C et F de 1K93 (a), et des chaînes A et O de 1AFX, ainsi que des chaînes B et P des molécules symétriques (b). La figure en insert dans (b), calculée avec LIGPLOT [271], montre les liaisons hydrogènes et les interactions hydrophobes établies entre les boucles HL et HM des chaînes A et B. La visualisation des cartes de densité électronique a été réalisée avec COOT [82].**

Les différences entre structures se situent surtout au niveau local, par exemple s’agissant du squelette de la région de contact entre N-CaM et le domaine Hel de EF (résidus 672-695, 737-745 et 784-793) [228]. La boucle HL-HM (résidus 675-684) (Fig. 5) adopte une conformation déformée dans la structure 1K93 (chaîne C), illustrée par la formation d’un lacet protubérant entre les résidus 681 et 691 et par la perte de structure secondaire de l’hélice HM (résidus 688 à 695) (Fig. 6a). Au contraire, dans 1XFX, la partie N-terminale de l’hélice HM est bien formée (Fig. 6b). Ces différences reflètent une carte de densité électronique pauvre dans cette région de 1K93 (Fig. 6a), tandis que la densité électronique est nettement améliorée dans 1XFX, 1XFY et 1XFZ (Fig. 6b). De plus, les facteurs B correspondants sont particulièrement élevés dans 1K93 (données non montrées), dénotant une flexibilité plus grande que dans 1XFX. Les environnements cristallins des deux structures sont également différents : la boucle HL-HM de 1XFX est en contact avec la même boucle de la molécule symétriquement liée (Fig. 6b), de telle sorte que la conformation de la boucle est étroitement maintenue. Les boucles HL-HM des deux molécules symétriques (chaînes A et B) sont orientées perpendiculairement, autorisant la formation d’un motif spécifique de liaisons hydrogènes et d’interactions hydrophobes (voir insert, Fig. 6b) qui sert de chapeau stabilisateur à l’hélice HM [19].

### 3 Plan de l’étude

Les interactions entre bio-molécules, au sein de la cellule, modulent la transmission de signaux essentiels à la survie, la croissance, la différenciation, la prolifération. A l’intérieur même des bio-molécules, l’information se propage par des changements conformationnels et des mouvements concertés, qui sous-tendent la fonction biologique.

L’espace conformationnel exploré par les protéines est décrit par un paysage énergétique. En solution, plusieurs sous-états, associés à des bassins d’énergie, sont à l’équilibre, et les taux d’inter-conversion entre eux sont déterminés par les barrières énergétiques qui les séparent. Selon le modèle Monod-Wyman-Changeux [192], lors d’une interaction moléculaire, l’équilibre entre sous-états est déplacé, de sorte que la conformation favorisée par la fixation du ligand devient plus peuplée et adopte une énergie plus basse. Il y a donc influence réciproque entre les interactions protéiques et le paysage énergétique des bio-molécules impliquées. En modifiant le paysage énergétique, enthalpiquement à travers des changements conformationnels, ou entropiquement via la modulation de la flexibilité, on déplace l’équilibre des populations de sous-états. De telles modifications peuvent ainsi rendre la conformation active défavorable et empêcher indirectement l’inter-conversion induite par la fixation du substrat [266; 167].

Le complexe EF-CaM, englobant environ six cents résidus, est un bon exemple où la mobilité structurale joue un rôle crucial dans l’établissement de la fonction enzymatique, parce que : (i) la calmoduline ubiquitaire, à la plasticité remarquable, est piégée en conformation étendue par EF dans le complexe ; (ii) une large transition conformationnelle de la toxine EF est nécessaire à son

activation ; (iii) le niveau de calcium influence la stabilité du complexe, sans pourtant affecter son architecture globale.

Cette thèse comporte trois volets, dédiés à l'étude du comportement dynamique du complexe actif EF-CaM et à la recherche d'inhibiteurs de la toxine. Dans un premier temps, l'influence du calcium sur la stabilité du complexe, attestée expérimentalement, a été sondée à travers la caractérisation de la plasticité conformationnelle des partenaires de l'interaction, par dynamique moléculaire. Une attention particulière a été portée aux mouvements globaux de la calmoduline, et à sa capacité à fixer le calcium dans le complexe, afin de mieux comprendre les mécanismes par lesquels la flexibilité de cette protéine est associée à la régulation spécifique de multiples cibles. Dans un deuxième temps, l'effet électrostatique des ions calcium a été modélisé sur le réseau de résidus du complexe, pris cette fois dans son ensemble et non plus considéré du point de vue de chaque partenaire. Les aspects dynamiques et énergétiques de la propagation du signal calcium, au niveau local et global, ont été caractérisés et reliés dans un cadre commun de compréhension. Enfin, la mobilité structurale de la protéine EF a été décrite par des méthodes de modélisation moléculaire, afin de sonder sa relation avec la fonction enzymatique. Une approche originale a été élaborée et appliquée pour la recherche d'inhibiteurs, visant à cibler non pas le site enzymatique lui-même, mais plutôt la transition conformationnelle conduisant à l'activation de la toxine. La validation de l'approche a été menée à travers des tests expérimentaux, révélant plusieurs inhibiteurs capables d'empêcher l'interaction entre EF et CaM.

# Approches méthodologiques

## 1 La dynamique moléculaire

La dynamique moléculaire (*DM*) consiste en la modélisation, par le calcul informatique, de l'évolution d'un système de particules au cours du temps. Conçue dans les années 1950 dans le cadre de la physique théorique [12], cette technique est maintenant largement appliquée en science des matériaux et pour l'étude des biomolécules. Les simulations *DM* sont un moyen de sonder les relations structure-dynamique-fonction, en explorant l'espace conformationnel des biomolécules dans le vide, l'eau, une membrane lipidique... La génération d'ensembles statistiques de conformations permet d'accéder aux propriétés physico-chimiques et grandeurs thermodynamiques caractéristiques des protéines. La dynamique moléculaire est aussi utilisée pour déterminer des structures à partir de données de biologie structurale (cristallographie aux rayons X *RX*, résonance magnétique nucléaire *RMN*, microscopie électronique *EM*, diffusion des rayons X aux petits angles *SAXS*...).

### 1.1 Principe de la dynamique classique

Le comportement des particules atomiques obéit aux lois de la mécanique quantique relativiste. Le principe de la dynamique moléculaire classique consiste à décrire le mouvement de ces particules à partir des lois de Newton, de la mécanique classique [259; 165], en admettant un certain nombre d'approximations. L'énergie potentielle  $U$  du système étudié est représentée par un champ de force empirique (voir section 1.2). A énergie totale  $E$  constante, l'évolution des positions  $\vec{r}_i$  des atomes est déterminée par l'équation du mouvement de Newton :

$$\vec{a}_i = \frac{d^2\vec{r}_i}{dt^2} = \frac{\vec{F}_i}{m_i}, \quad (1)$$

$$\vec{F}_i = -\text{grad}_i U, \quad (2)$$

où  $\vec{F}_i$  est la force totale appliquée sur la particule de masse  $m_i$ .

A chaque pas de temps  $t$ , la position et la vitesse de la particule  $i$  sont interpolées à partir de son accélération instantanée  $\vec{a}_i$  et de ses position et vitesse au temps  $(t - \delta t)$ . L'accélération

$\vec{a}_i$  est déterminée grâce à la force  $\vec{F}_i$ , supposée constante dans l'intervalle de temps  $\delta t$ . Dans le logiciel AMBER 8 [41], avec lequel ont été effectuées les simulations présentées ici, l'équation du mouvement est résolue avec l'algorithme du « saut-de-mouton » (*leap-frog* [119]), variante de l'algorithme de Verlet [261; 262] :

$$\vec{v}(t + \frac{1}{2}\delta t) = \vec{v}(t - \frac{1}{2}\delta t) + \delta t \cdot \vec{a}(t), \quad (3)$$

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \cdot \vec{v}(t + \frac{1}{2}\delta t). \quad (4)$$

Les vitesses initiales sont générées de manière aléatoire selon une distribution de Maxwell-Boltzmann. L'intervalle de temps  $\delta t$  doit être suffisamment petit par rapport aux périodes de mouvement des atomes [165]. Sachant que les liaisons impliquant des hydrogènes ont des périodes de vibration de l'ordre de 10 fs, un pas d'intégration  $\delta t = 1$  fs semble approprié.

Typiquement, un protocole de simulation dans le logiciel AMBER 8 se déroule de la manière suivante :

1. **Construction du système.** Les coordonnées atomiques sont obtenues à partir de structures tridimensionnelles, résolues par cristallographie ou par *RMN* et généralement déposées dans la *RCSB Protein Data Bank* [31] (PDB). Les atomes manquants, notamment les hydrogènes, sont éventuellement modélisés. Des contre-ions sont ajoutés pour neutraliser le système, en utilisant une grille de potentiel de Coulomb dans un rayon de 4 Å autour de la biomolécule. Le système est ensuite placé au centre d'une boîte de molécules de solvant pré-équilibré. Le volume du soluté est « découpé » dans cette boîte en retirant les molécules de solvant chevauchantes, ie dont au moins un atome est situé à moins de 1 Å du soluté.
2. **Minimisation de l'énergie.** La configuration utilisée pour initier la simulation doit être « énergétiquement réaliste », au risque d'obtenir des résultats aberrants. Ainsi, afin de réduire des forces initiales trop importantes, le système est relaxé lors d'une étape de minimisation d'énergie. Parmi les différentes méthodes mathématiques de minimisation, AMBER 8 utilise une combinaison de deux méthodes de gradient, la méthode de la plus grande pente [60] (*steepest descent*) et la méthode du gradient conjugué [90] (*conjugate gradient*). A chaque cycle de minimisation, la configuration du système est modifiée selon une direction déterminée par le gradient de l'énergie potentielle. Dans un premier temps, la méthode de la plus grande pente permet de tendre rapidement vers un minimum d'énergie local, puis la méthode du gradient conjugué sert à affiner la minimisation pour converger vers un minimum local.
3. **Thermalisation - Équilibration - Production.** L'introduction de la biomolécule dans la boîte de solvant et la définition des bords de celle-ci induisent des lacunes de van der Waals. Pour rétablir une densité optimale, une étape d'équilibration est nécessaire. Le système est d'abord chauffé à la température souhaitée, en maintenant la pression constante. Plusieurs cycles se succèdent ensuite, à volume constant puis à pression constante, ce qui permet

d'ajuster le volume et de corriger la densité de la boîte. A l'issue de cette étape d'équilibration, la trajectoire de la simulation *DM* proprement dite peut enfin être produite.

## 1.2 Champs de force

Dans le cadre de simulations *DM* en coordonnées cartésiennes, l'hamiltonien  $H$  du système étudié comporte un terme d'énergie cinétique  $E_c = \sum_{i=1}^N \frac{1}{2} m_i \vec{v}_i^2$  et un terme d'énergie potentielle  $U$ , c'est-à-dire le champ de force ou la fonction d'interaction. Ce dernier décrit l'énergie d'interaction entre les particules du système et dépend d'un certain nombre de paramètres empiriques. Le champ de force utilisé dans AMBER 8 [41] est exprimé comme la somme de cinq types d'interactions :

$$U = \sum_{\text{liaisons}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dièdres}} \sum_n \frac{1}{2} V_n [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \epsilon_{ij} \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] \quad (5)$$

Les trois premiers termes correspondent respectivement aux énergies d'élongation des liaisons covalentes, de déformation angulaire et de déformation des angles de torsion (dièdres) entre atomes liés. Les deux derniers termes représentent les contributions électrostatiques et de van der Waals entre atomes non liés, c'est-à-dire entre atomes d'une même molécule séparés par plus de deux liaisons, ou entre atomes de molécules différentes.

**Interactions entre atomes liés.** Les énergies d'élongation des liaisons covalentes et de déformation angulaire sont exprimées par des potentiels harmoniques.  $r$  et  $\theta$  sont la longueur de liaison et l'angle instantanés, en Å et °,  $r_{eq}$  et  $\theta_{eq}$  sont les valeurs de référence,  $K_r$  et  $K_\theta$  sont les constantes de force associées en kcal.mol<sup>-1</sup>.Å<sup>-2</sup>. L'énergie de déformation des angles dièdres est décrite par une série de Fourier et dépend de l'angle dièdre instantané  $\phi$  en °, d'un terme d'énergie  $V_n$  en kcal.mol<sup>-1</sup>, d'un terme de multiplicité  $n$  et des facteurs de phase  $\gamma$  pour lesquels l'énergie de torsion est minimale.

**Interactions électrostatiques.** Les interactions électrostatiques entre deux atomes  $i$  et  $j$  sont évaluées avec la loi de Coulomb :

$$U_{el} = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (6)$$

où  $q_i$  et  $q_j$  sont les charges des atomes  $i$  et  $j$ ,  $r_{ij}$  est la distance entre les deux atomes en Å, et  $\epsilon_0$  est la constante électrique (ou permittivité diélectrique du vide) qui vaut  $8,85418782 \times 10^{-12} F.m^{-1}$ .

**Interactions de van der Waals.** Les interactions de van der Waals sont décrites par un potentiel de Lennard-Jones en 6-12 :

$$U_{vdw} = \sum_{i < j} \varepsilon_{ij} \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] \quad (7)$$

où  $r_{ij}$  est la distance entre les atomes  $i$  et  $j$  en Å,  $R_{ij}^*$  et  $\varepsilon_{ij}$  sont les paramètres de Lennard-Jones. Le premier terme en  $r_{ij}^{-12}$  correspond à la répulsion entre deux atomes dû au principe d'exclusion de Pauli et le second terme en  $r_{ij}^{-6}$  représente la dispersion attractive de London.  $\varepsilon_{ij}$  correspond à la profondeur du puits de potentiel à la distance de séparation  $R_{ij}^*$  (voir Fig. 1). Les valeurs de  $\varepsilon_{ij}$  et  $R_{ij}^*$  sont obtenues à partir des règles de Lorentz-Berthelot :

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \quad (8)$$

$$R_{ij}^* = \frac{R_{ii}^* + R_{jj}^*}{2} \quad (9)$$

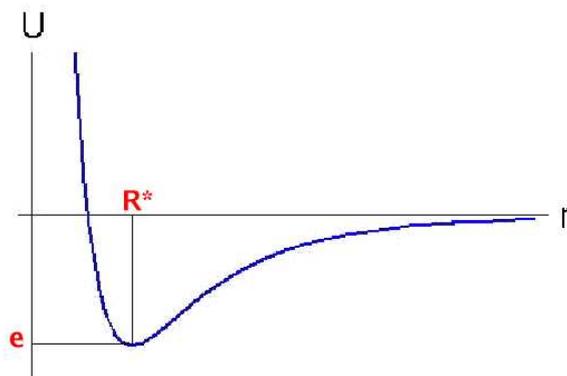


FIG. 1: Schéma général des potentiels de Lennard-Jones

**Interactions à longue distance.** Les systèmes simulés comportent généralement plusieurs milliers d'atomes, ce qui rend le traitement explicite de toutes les interactions très long. Ainsi, des méthodes d'approximation sont utilisées pour évaluer les interactions à longue distance. Les interactions de van der Waals, qui diminuent en  $\frac{1}{r^6}$ , sont généralement négligées au delà d'un certain rayon de coupure. Par contre, les interactions coulombiennes varient de manière inversement proportionnelle à la distance  $r_{ij}$ , et ne peuvent donc être omises. Dans AMBER, elles sont évaluées par le protocole du maillage particulier d'Ewald [65] (*particle mesh Ewald* ou *PME*). Cet algorithme est basé sur la méthode de la sommation d'Ewald, qui se propose de calculer de manière efficace les interactions (électrostatiques) à longue distance entre des particules et toutes leurs images périodiques. Le système est considéré comme composé d'une unité centrale reproduite en une infinité d'images (conditions limites périodiques 3D). Chaque particule qui sort de la boîte primitive (ou d'une boîte image) par un côté est remplacée par une particule entrant par le côté opposé. Les molécules situées près des bords de la boîte sont entourées par des molécules

provenant d'autres boîtes. Le nombre de particules simulées est ainsi réduit et l'on s'affranchit des effets de bords. La somme d'Ewald est composée de trois termes :

$$U_{ewald} = U^R + U^I + U^0 \quad (10)$$

$U^R$  est la somme des interactions à courte distance dans l'espace direct réel (à l'intérieur de l'unité primitive),  $U^I$  est la somme des interactions à longue distance dans l'espace réciproque imaginaire (les unités images) et  $U^0$  est un terme correctif constant. Dans l'algorithme *PME*, le premier terme  $U^R$  est évalué directement dans un rayon de coupure donné alors que le second terme  $U^I$  est approximé par une transformée de Fourier rapide (*fast Fourier transform* ou *FFT*) sur une grille où les charges sont interpolées en chaque point. La complexité du calcul est ainsi réduite à  $o(N(\log N))$ ,  $N$  étant le nombre de particules du système.

**Liaisons X-H.** Dans le cadre de notre étude, les vibrations haute fréquence des liaisons de type X-H, impliquant des hydrogènes, ont une influence très limitée sur le comportement global du système et ne nous intéressent pas dans le détail. Il est possible de figer ces vibrations en « gelant » les distances X-H à leurs valeurs d'équilibre, avec la méthode SHAKE [217]. Le pas d'intégration peut ainsi être augmenté à 2 fs. La contrainte de distance appliquée est la suivante :

$$r_{ij}^2(t) - d_{ij}^2 = 0 \quad (11)$$

où  $r_{ij}$  est la distance entre les atomes  $i$  et  $j$  et  $d_{ij}$  est la distance de référence. En pratique, un écart  $\varepsilon$  est toléré, de sorte que l'inégalité suivante est remplie :

$$\frac{r_{ij}^2(t) - d_{ij}^2}{d_{ij}^2} < \varepsilon \quad (12)$$

où  $\varepsilon$  vaut typiquement  $10^{-5}$ .

### 1.3 Estimation de l'énergie libre : méthode *MMPBSA*

Il existe plusieurs méthodes de thermodynamique statistique pour calculer la différence d'énergie libre entre deux états d'un même système moléculaire ou entre deux systèmes moléculaires différents. Dans l'approche *MMPBSA* [236; 151; 273; 213; 272] (*molecular mechanics Poisson-Boltzmann surface area*), l'énergie libre de liaison d'un ligand à un récepteur est exprimée comme la différence entre l'énergie libre du complexe et les énergies libres des ligand et récepteur isolés :

$$\Delta\Delta G_{liaison} = \Delta G_{complexe} - [\Delta G_{récepteur} + \Delta G_{ligand}] \quad (13)$$

Les énergies libres des différents composants du système sont moyennées sur un ensemble de conformations obtenues par simulation de dynamique moléculaire. Les conformations du complexe,

du ligand et du récepteur sont en général issues d'une même trajectoire, pour éviter l'introduction d'erreurs dues à plusieurs échantillonnages  $DM$  incomplets.

L'énergie libre  $\Delta G$  est exprimée comme une somme de contributions :

$$\Delta G = \Delta G_{MM} + \Delta G_{solv}^p + \Delta G_{solv}^{np} - T\Delta S \quad (14)$$

L'énergie de mécanique moléculaire  $\Delta G_{MM}$  regroupe l'énergie interne  $E_{int}$ , les interactions électrostatiques coulombiennes  $E_{ele}$  et les interactions de van der Waals  $E_{vdw}$ . L'énergie de solvation se décompose en une contribution polaire  $\Delta G_{solv}^p$ , calculée en résolvant l'équation de Poisson-Boltzmann [274; 148], et en une contribution non polaire  $\Delta G_{solv}^{np}$ , fonction de la surface accessible au solvant [59]. L'entropie conformationnelle est estimée à partir d'une analyse en modes normaux, mais elle est souvent omise à cause de temps de calculs longs et d'erreurs importantes dans son estimation.

Pour évaluer les interactions soluté-solvant et solvant-solvant, la méthode *MMPBSA* utilise une représentation implicite du solvant, dans laquelle les molécules, d'eau en l'occurrence, sont traitées comme un continuum non structuré.

**Interactions polaires de solvation** Les interactions polaires sont évaluées selon la théorie électrostatique de Poisson-Boltzmann (PB) par une approximation de champ moyen [225; 70]. Le soluté est représenté par un corps diélectrique dont la forme est définie par les coordonnées atomiques et les rayons des cavités atomiques. Il est composé d'un ensemble de charges ponctuelles placées au centre de chaque atome qui génèrent un champ électrostatique. Le potentiel électrostatique  $\phi_j$  à la charge atomique  $j$  est déterminé en résolvant l'équation de Poisson-Boltzmann :

$$\Delta\varepsilon(\vec{r})\Delta\phi(\vec{r}) = -4\pi p(\vec{r}) - 4\pi \sum_i z_i \exp(-z_i\phi(\vec{r})/k_B T) \quad (15)$$

où  $\varepsilon(\vec{r})$  est la constante diélectrique,  $\phi(\vec{r})$  est le potentiel électrostatique,  $p(\vec{r})$  est la charge du soluté,  $z_i$  est la charge de l'ion de type  $i$ ,  $k_B$  est la constante de Boltzmann et  $T$  est la température.

Dans le module *MMPBSA* d'AMBER 8, la forme linéarisée de l'équation PB est utilisée. La linéarisation est effectuée par la méthode des différences finies [69], dans laquelle : (i) les charges atomiques sont cartographiées sur une grille de points de différences finies ; (ii) des conditions périodiques/non-périodiques sont assignées ; (iii) un modèle diélectrique est appliqué pour définir la frontière entre le solvant (diélectrique fort) et l'intérieur du soluté (diélectrique faible).

**Interaction non polaires de solvation** Les interactions non polaires sont typiquement évaluées par un terme proportionnel à la surface accessible au solvant (*SASA*) :

$$\Delta G_{solv}^{np} = \gamma SASA + c \quad (16)$$

Dans AMBER 8 , une méthode alternative est également proposée (Tan & Luo, en préparation), dans laquelle la contribution non polaire est composée d'un terme de dispersion (attractif) et d'un terme de cavité (répulsif) :

$$\Delta G_{solv}^{np} = \Delta G_{disp} + \Delta G_{cav} \quad (17)$$

$\Delta G_{disp}$  est calculé par une intégration numérique sur la surface accessible au solvant, qui rend compte des interactions attractives de solvation avec une distance de coupure infinie.  $\Delta G_{cav}$  est proportionnel à la SASA, comme décrit plus haut.

## 1.4 Échantillonnage et transition conformationnelle

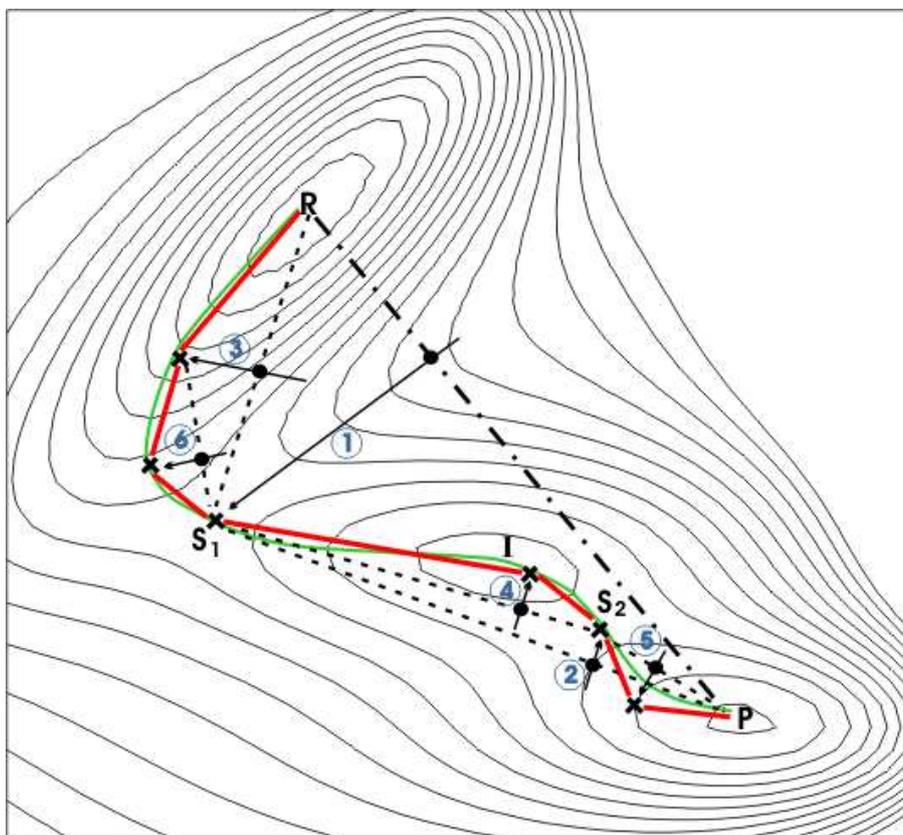
Au cours des trente dernières années, le développement des ressources informatiques – avec récemment le projet BlueGene, ainsi que l'amélioration des algorithmes de calculs et des champs de force, a permis d'allonger les temps de simulation et d'étudier des systèmes de plus en plus gros [140; 149]. Il est maintenant possible de modéliser le comportement de protéines en solution aqueuse pendant plusieurs centaines de nanosecondes. Cependant, la plupart des mouvements et interactions nécessaires aux fonctions biologiques dans la cellule se produisent sur une échelle de temps allant de la microseconde à la minute. Au niveau moléculaire, les réarrangements conformationnels induits par ces mouvements et interactions sont associés à des barrières énergétiques importantes, que les simulations *DM* ne permettent généralement pas de traverser. En effet, les champs de force empiriques produisent des paysages énergétiques faits de multiples minima locaux, dans lesquels les simulations sont facilement piégées.

Plusieurs méthodes offrent la possibilité d'améliorer ou d'accélérer l'échantillonnage des conformations et d'accéder ainsi à des événements rares, tels que la dissociation d'un ligand ou l'agrégation de polypeptides. Parmi les récents développements, on peut citer la méthode d'échange de répliques [114; 239] (*replica exchange* ou *RE*), dans laquelle plusieurs copies d'un même système sont simulées en parallèle, par dynamique moléculaire. A chaque réplique est attribuée une valeur différente d'une variable indépendante, par ex. la température. Les valeurs sont échangées entre répliques voisines à des intervalles de temps donnés, selon un critère de Metropolis [187].

Des simulations permettent un échantillonnage accru au niveau local [80] (*locally enhanced sampling* ou *LES*), dans lesquelles est dupliquée une partie du système, comme un ligand ou une région flexible, en de multiples copies qui créent un champ moyen environnant. Les copies n'interagissent pas entre elles et leurs interactions avec le reste du système sont moyennées en fonction du nombre de copies. Le potentiel est par conséquent lissé et les barrières énergétiques abaissées, tandis que les copies sont libres de s'éloigner les unes des autres [215]. Dans le cas d'un complexe protéine-ligand, l'abaissement des barrières énergétiques (qui doivent être surmontées pour observer la dissociation du ligand) augmente la constante de dissociation du ligand de manière exponentielle [254].

Dans la méthode de dynamique moléculaire orientée [105; 133; 132] (*steered molecular dynamics* ou *SMD*), une force extérieure est appliquée sur une partie du système, pour guider

celui-ci dans une direction donnée. Cette force est une fonction du temps, par exemple un ressort harmonique ou un ressort dont la constante de raideur augmente linéairement avec le temps. Typiquement, ces simulations sont utilisées pour modéliser la dissociation d'un ligand. Elles peuvent aussi servir de point de départ pour la résolution de chemins de transition entre deux états, inactif et actif, d'une protéine. Ainsi, la méthode d'affinement par pic conjugué [89] (*Conjugate peak refinement* ou *CPR*), développé par Fischer & Karplus, ajoute un nombre suffisant de conformations intermédiaires à cette conjecture initiale pour déterminer un chemin d'énergie minimum, adiabatique. Les conformations sont optimisées de telle sorte qu'elles se rapprochent de « vallées » énergétiques et que les conformères de plus haute énergie coïncident avec des points sels. Le chemin produit constitue un « film » du processus de transition, même si chaque atome a la possibilité de bouger de manière indépendante. Il rend compte uniquement des mouvements essentiels à la transition. L'algorithme *CPR* a été utilisé dans cette étude, pour caractériser la transition conformationnelle de EF (voir Chapitre 5).



**FIG. 2: Description schématique de l'algorithme *CPR*.** Les lignes de contour montrent le paysage énergétique 2D d'une molécule hypothétique (avec des minima à R, I, P). Un chemin initial est produit, ici la simple ligne d'interpolation (-.-.) entre les états du réactant (R) et du produit (P) (les résultats ne changent pas si la transition est considérée comme R->P, ou P->R). Chaque cycle *CPR* (numérotés en bleu) ajoute automatiquement une conformation intermédiaire (.) à la région de plus haute énergie (le 'pic') et l'optimise par minimisation conjuguée (->) de telle sorte qu'elle se rapproche du fond de la vallée énergétique (x). Ce nouveau point est ensuite incorporé dans le chemin par deux segments adjacents d'interpolation (- - -). La répétition de cette procédure heuristique conduit à un chemin final (en rouge) qui épouse la forme de la surface d'énergie et passe à travers les points sels exacts (S1, S2), proche du chemin d'énergie minimum (en vert). Ce schéma est extrait de la publication de Fischer & Karplus [89].

Enfin, la méthode méta-dynamique (*metadynamics* ou *MD*) [160] utilise un potentiel d'énergie additionnel pour biaiser l'exploration de l'espace conformationnel. Ce potentiel est une somme de gaussiennes centrées sur les valeurs de variables collectives choisies. Les variables collectives incluent idéalement tous les modes pertinents pour décrire le système étudié et difficiles à échantillonner lors de simulations *DM* classiques. Le potentiel additionnel dépend du temps et guide le système hors des minima d'énergie pour le forcer à explorer des régions non encore visitées de l'espace conformationnel.

En biologie, les fonctions importantes de la cellule nécessitent le concours de larges agrégats macro-moléculaires, de l'ordre de plusieurs dizaines de nanomètres. Un nombre croissant de ces agrégats sont étudiés expérimentalement par des techniques de faible à moyenne résolution telles que la cryo-microscopie électronique [27], la diffusion des rayons X aux petits angles [197] ou la spectrométrie de masse [224]. L'utilisation de représentations simplifiées, dites gros-grains, intégrant une multitude de degrés de libertés dans un ensemble réduit, s'inscrit dans ce contexte. Dans ces modèles, les biomolécules ne sont plus décrites à l'échelle atomique, mais plus grossièrement par des grains ou des perles, qui représentent les résidus, ou les domaines comme des corps rigides à six degrés de liberté, trois de translation et trois de rotation. Le niveau de granularité résulte d'un compromis entre l'exactitude et la transférabilité du modèle. Dans le modèle en réseau élastique (*elastic network model* ou *ENM*) [244; 23], la protéine est représentée par un réseau de ressorts harmoniques qui connectent entre eux les atomes interagissants ou les résidus proches. L'analyse en mode normaux (*normal mode analysis* ou *NMA*) de ce réseau permet d'obtenir les modes collectifs intrinsèques de la protéine, qui sont les modes de plus basses fréquences. Ils décrivent les mouvements globaux, supposément fonctionnels, accessibles à proximité de l'état natif, considéré comme l'état d'énergie minimale. L'hypothèse sous-jacente de ce modèle est que les changements structuraux sont déterminés par la topologie des contacts inter-résidus de l'état natif. Malgré l'extrême simplicité de cette approche, l'analyse en mode normaux gros-grains parvient à prédire correctement les fluctuations thermiques mesurées expérimentalement. Elle est notamment utilisée pour élucider les mouvements induits par la formation d'un complexe et ainsi proposer des conformations cibles pour l'arrimage de ligands, pour affiner des structures à basse résolution [73], ou encore pour explorer des chemins de transition conformationnelle entre deux états donnés [74].

D'autres modèles gros-grains existent, notamment ceux inspirés du modèle de Go [253], et des champs de force sont spécialement développés pour modéliser de larges systèmes, comme les bicouches lipidiques ou les agrégats de protéines. Les paramètres physico-chimiques sont déterminés soit à partir de simulations tout-atomes (*bottom-up*), soit de manière empirique (*top-down*).

## 2 Le criblage virtuel

L'élaboration d'un nouveau médicament est un processus coûteux, long et risqué : près d'un milliard d'euros et 15 ans en moyenne sont nécessaires à la mise sur le marché d'une seule molécule, sur douze molécules en phase I d'essais cliniques [44]. Au cours des trente dernières années, l'activité de recherche et développement (R&D) au sein de l'industrie pharmaceutique a connu des changements majeurs. L'essor de la génomique fonctionnelle avec le projet Génome Humain dans les années 1990-2000 a mis en évidence un nombre toujours croissant de cibles thérapeutiques potentielles. En parallèle, le développement des méthodes de biologie structurale (cristallographie aux rayons X, puis *RMN*) a permis de faire le lien entre les phénomènes biologiques et les interactions physico-chimiques sous-jacentes. Face à cet ensemble pléthorique de nouvelles cibles et structures, les industries pharmaceutiques ont entrepris la synthèse et le criblage à haut-débit de larges bibliothèques combinatoires de composés chimiques. Cependant, le taux de réussite de ces approches systématiques est relativement faible. C'est pourquoi des approches rationnelles sont aujourd'hui privilégiées, qui combinent méthodes expérimentales et computationnelles. Dans ce contexte, le criblage *in silico* de molécules est de plus en plus utilisé pour identifier de potentiels candidats médicaments et prédire leurs modes de liaison et affinités [145; 212].

Notamment, les approches basées sur la structure (*structure-based*) ont prouvé leur utilité dans la découverte et le développement d'un certain nombre de molécules actives [147; 290; 200]. Elles se sont récemment popularisées, grâce (i) à l'augmentation des ressources computationnelles [17; 111; 231], dont les autres techniques *in silico* ont également bénéficié, (ii) à l'amélioration des méthodes et des modèles, s'agissant notamment d'échantillonner l'espace conformationnel [223; 241; 181; 7], (iii) à la croissance des banques de structures tri-dimensionnelles de biomolécules [32; 230].

### 2.1 Principe

De manière générale, le criblage virtuel s'apparente au criblage réel à haut débit (*high throughput screening* ou *HTS*). Plutôt que de tester *in vitro* l'activité potentielle de 100 000 à 2 000 000 molécules sur une cible, le criblage virtuel a pour objectif de prédire *in silico* la capacité d'un grand nombre de molécules à se fixer sur cette cible. Cette méthode computationnelle ne nécessite pas la synthèse chimique de toutes les molécules criblées, ce qui constitue un avantage non négligeable en terme de coûts. La qualité des résultats du criblage dépend de l'efficacité des algorithmes à discriminer les agents actifs des composés inactifs. Les molécules prédites comme actives, les « touches », sont destinées, après validation expérimentale, à servir de points de départ pour les chimistes médicinaux.

A l'origine d'un projet de conception rationnelle de nouveau médicament, deux cas de figure peuvent se présenter : soit la structure tridimensionnelle de la cible thérapeutique est connue, soit la cible thérapeutique n'a pas été identifiée (par exemple les récepteurs couplés aux protéines G).

Le choix de la cible lui-même est une tâche difficile, qui dépend de beaucoup d'aspects expérimentaux externes qui ne seront pas discutés ici, comme par exemple la disponibilité d'un essai pour des tests expérimentaux ou l'importance du rôle de la cible dans les processus biologiques *in vivo*. Dans le cas où la structure de la cible n'est pas disponible, les méthodes basées sur la connaissance du ligand (*ligand-based drug design*) sont privilégiées, comme par exemple, les techniques de relation quantitative structure-activité (*quantitative structure-activity relationship* ou *QSAR*), qui permettent de corréler l'activité des molécules avec leurs propriétés physico-chimiques. Nous nous intéressons ici au deuxième cas, lorsque la structure de la protéine cible est connue (*structure-based drug design*). Plusieurs protocoles [207; 8; 97; 260; 193], basés sur le criblage virtuel de banques de ligands sur des structures 3D cibles, se déclinent en quatre étapes :

**1. Construction du modèle structural du récepteur et pré-sélection de composés.**

Les coordonnées atomiques 3D de la cible sont obtenues à partir d'une structure cristallographique de bonne résolution (meilleure que 3 Å). Les atomes manquants sont éventuellement modélisés et la structure minimisée. Les coordonnées atomiques – bi-dimensionnelles – des composés sont obtenues à partir de banques de molécules. Les structures 3D sont générées par des logiciels comme CORINA [218], Concord et Confort [1] ou Converter [2]. Il peut aussi être nécessaire d'attribuer des charges partielles, de gérer les différents états de protonation et la stéréochimie des centres chiraux. La banque est souvent filtrée, pour ne retenir que des molécules dont les propriétés chimiques s'apparentent à celles de médicaments ou de composés actifs connus. Les protocoles de filtrage sont des variantes de la règle de cinq de Lipinski [172], qui définit des critères physico-chimiques descripteurs de la bio-disponibilité orale (poids moléculaire, coefficient de partition, nombre de donneurs/accepteurs d'hydrogènes, d'atomes lourds, de cycles...etc).

**2. Arrimage (*docking*) et évaluation de l'affinité des ligands (*scoring*).** Deux points clés de la procédure de criblage virtuel sont la qualité de l'échantillonnage conformationnel et la capacité de la fonction de *scoring* à discriminer les vrais inhibiteurs des composés inactifs. Plusieurs méthodes permettent un échantillonnage étendu de l'espace conformationnel du ligand, par : (i) construction incrémentale [207], (ii) simulations de Monte Carlo (*MC*) dans l'espace des angles de torsion [8], (iii) recherche systématique complétée par des simulations *MC* [97], (iv) algorithme génétique [260; 193]. La flexibilité conformationnelle de la cible est aussi explorée à l'aide de méthodes spécifiques [182; 26; 246]. L'arrimage de chaque ligand produit un ensemble de modes de liaison ou « poses », qui sont évaluées par une fonction de *scoring*. Plusieurs études sur la qualité des fonctions de *scoring* ont montré qu'aucune fonction n'est universelle [190; 144].

**3. Validation expérimentale d'un ensemble de molécules sélectionnées.** A partir de leur score d'arrimage, ou affinité prédite, les molécules criblées *in silico* sont classées, et un petit ensemble de « touches » est choisi pour des tests *in vitro* ou *in vivo*. Ces tests expérimentaux permettent de valider l'approche *in silico* et de caractériser le facteur d'enrichissement, ie la proportion de molécules réellement actives sur la part de la banque qui a

été sélectionnée.

## 2.2 Définition du site actif

La construction du modèle structural du récepteur est une étape cruciale du criblage virtuel. Les structures  $RX$ , d'où sont tirées les coordonnées des récepteurs, présentent plusieurs limitations inhérentes aux méthodes et modèles utilisés pour les résoudre. Par exemple, la conformation piégée en environnement cristallin n'est pas toujours représentative de l'état de la protéine en solution. De plus, même une bonne résolution, comprise entre 1.5 et 2.5 Å, ne permet pas de définir l'état de protonation des résidus. Enfin, et surtout, la fixation d'un ligand à la surface d'une protéine est un processus dynamique, au cours duquel la protéine et le ligand subissent des réarrangements conformationnels, parfois importants, dont l'état cristallin ne rend pas compte.

Au delà du choix de la structure, un autre aspect important concerne le choix du site de liaison potentiel ciblé à la surface biomoléculaire. La définition précise du site ne peut être évitée, étant donné que la plupart des logiciels se basent sur les interactions établies au niveau atomique entre le ligand et la biomolécule pour guider l'arrimage. Ce point est très sensible pour le succès du criblage, un mauvais choix de poche pouvant complètement biaiser le calcul et ainsi exclure toute chance de prédire des touches qui seront validées expérimentalement.

Quand elles sont disponibles, les structures de complexes substrat-récepteur co-cristallisés permettent de localiser précisément le site actif. Cependant, pour de nombreuses protéines, la fonction biologique n'est pas caractérisée au niveau moléculaire, ou simplement inconnue [279]. Des outils d'alignement de séquences ou de structures peuvent alors être utilisés pour localiser, voire modéliser par homologie, le site actif [53; 152; 49]. On peut aussi envisager de bloquer l'activité d'une protéine en ciblant une région différente de son site actif, auquel cas le site de liaison (allostérique) n'est pas défini *a priori*. Un certain nombre d'algorithmes sont dédiés au problème de la prédiction de sites de liaison, certains basés sur des considérations géométriques, d'autres en utilisant des calculs d'énergie. Les méthodes géométriques ont pour but d'identifier des dépressions à la surface protéique ou des volumes fermés par des atomes saillants [201; 118; 169; 35; 121; 99]. Elles sont généralement rapides et faciles à mettre en oeuvre mais ne permettent pas de distinguer une simple cavité d'un site de liaison fonctionnel. C'est pourquoi, des méthodes développées récemment introduisent des descripteurs physico-chimiques (degré d'enfouissement, exposition au solvant...) pour mieux comparer et classifier les sites de liaisons [195; 277; 141; 112]. D'autres algorithmes sont basés sur des calculs d'énergie d'interaction, pour positionner des fragments moléculaires ou des ligands-sondes à la surface de la protéine [163; 57; 161].

L'algorithme PocketFinder [15; 14], utilisé dans cette étude, définit une grille de potentiel (de Lennard-Jones, en 6-12) pour cartographier les interactions de van der Waals entre le récepteur et une sonde carbone de 1.7 Å. La carte de potentiel est ensuite lissée pour mettre en exergue les régions où le potentiel de van der Waals est bas, de manière continue, sur un espace significatif.

Le lissage s'effectue par moyennage itératif à chaque noeud  $P_{ijk}^n$  de la grille :

$$P_{ijk}^{n+1} = \frac{1}{2}[P_{ijk}^n + \frac{1}{6}(P_{(i-1)jk}^n + P_{(i+1)jk}^n + P_{i(j-1)k}^n + P_{i(j+1)k}^n + P_{ij(k-1)}^n + P_{ij(k+1)}^n)], \quad (18)$$

ce qui permet d'approximer la convolution :

$$P(\vec{r}) = \int f(\vec{p}, \vec{r}) P^0(\vec{p}) d\vec{p}, \quad (19)$$

où  $f(\vec{p}, \vec{r}) = \exp(-(\frac{\vec{p}-\vec{r}}{\lambda})^2)$ , avec le paramètre de lissage  $\lambda = \sqrt{\frac{2}{3}N_{iter}}$  fixé à 2.6 Å. Puis, la carte est contourée pour délimiter les volumes des enveloppes prédites. Le niveau de contour  $P_{cont}$  est évalué à partir de la moyenne  $\mu_{P_{ijk}}$  et de l'écart-type  $\sigma_{P_{ijk}}$  de tous les noeuds  $P_{ijk}$  de la grille :

$$P_{cont} = \mu_{P_{ijk}} - \tau\sigma_{P_{ijk}}, \quad (20)$$

avec le paramètre de seuil  $\tau = 4.6$ . Enfin, les enveloppes sont classées en fonction de leurs volumes : celles avec un volume inférieur à 100 Å sont éliminées. Les poches détectées sont aussi classées selon leur taille et leur forme, pour prédire leur capacité à interagir avec des molécules « drug-like ».

## 2.3 Algorithmes de *docking*

Il existe plus de soixante algorithmes dédiés au problème de la prédiction des position et orientation, la « pose », d'un ligand dans un site de liaison protéique. Aucun n'est à ce jour capable de pleinement rendre compte de la réalité biophysique du processus de *docking*, si bien que la qualité des prédictions obtenues avec un algorithme particulier dépend encore largement du système biologique lui-même [190].

Le placement du ligand dans le site actif est réalisé : (i) de manière rigide, en utilisant un ou plusieurs conformères pour tenir compte de la flexibilité du ligand, (ii) en re-construisant le ligand de manière incrémentale dans le site actif, (iii) par exploration stochastique de l'espace conformationnel (algorithmes génétiques, simulations de *Monte Carlo*). Dans notre étude, nous avons utilisé les logiciels FlexX [206] et ICM [6], dont les algorithmes de *docking* sont détaillés ici.

### FlexX

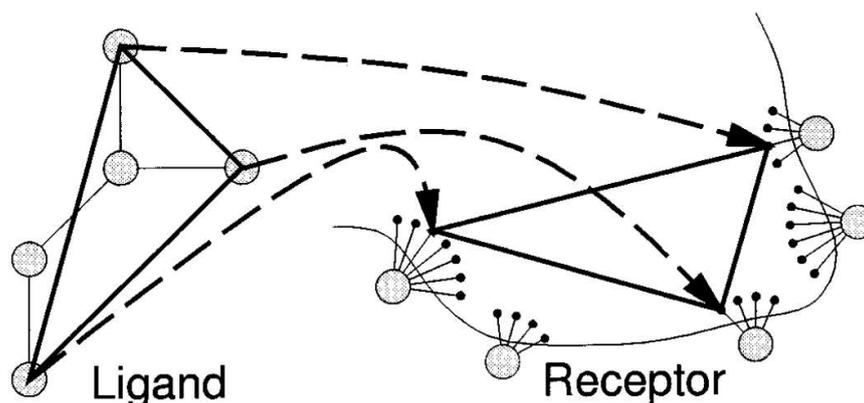
FlexX [206] procède par reconstruction incrémentale du ligand dans le site de liaison, en cherchant à maximiser les interactions ligand-récepteur [208] :

1. Sélection d'un fragment de base du ligand (susceptible d'établir des interactions directionnelles avec le récepteur)
2. Placement du fragment de base indépendamment du reste du ligand, en faisant correspondre des centres d'interaction du ligand et du site de la protéine récepteur (voir Fig. 3)

3. Construction du complexe récepteur-ligand, en partant d'un ensemble de placements favorables du fragment de base et en utilisant un arbre de recherche pour trouver les feuilles qui contiennent les placements les plus favorables énergétiquement.

Les interactions moléculaires sont représentées par des surfaces sphériques. A chaque groupe interagissant sont associés un type et une géométrie (centre, rayon et surface). Les différents types d'interaction décrits sont les liaisons hydrogènes et les interactions hydrophobes entre cycles aromatiques et/ou groupements méthyles et/ou amides.

**FIG. 3: Algorithme de placement de FlexX** : il s'agit de faire correspondre trois centres d'interaction (*sphères grises*) du ligand avec trois points discrets d'interaction du site actif (*points noirs*), pour définir une unique transformation du ligand dans le site actif. Ce schéma est extrait de la publication de Rarey et al. [208].



## ICM

La méthode ICM [8; 6] utilise les coordonnées internes pour décrire les conformations et positions relatives des biomolécules. Les positions et les angles de torsion des atomes du ligand ainsi que les angles de torsion des chaînes latérales du site de liaison sont considérés comme variables alors que le reste des degrés de liberté de la protéine sont fixés. Le processus de *docking* effectue une exploration stochastique de l'espace conformationnel par un algorithme de *Monte Carlo* [8; 247] :

1. Modification aléatoire du placement du ligand par mouvement pseudo-brownien ou rotation des liaisons internes
2. Optimization locale de l'énergie (en utilisant les paramètres ECEPP3 avec une constante diélectrique distance-dépendante  $\epsilon = 4r$  [196])
3. Evaluation de l'énergie de solvatation et de la contribution entropique des chaînes latérales
4. Application du critère de Metropolis [187] à une température donnée et retour à l'étape 1

Le critère de Metropolis permet d'accepter des conformations de plus haute énergie en utilisant une dépendance à la température. Ainsi, si l'énergie de la nouvelle conformation,  $E_{n+1}$ , est plus basse que l'énergie avant modification,  $E_n$ , la nouvelle conformation est acceptée. Sinon, le critère de Metropolis s'applique [187] :

$$\begin{cases} \text{si } \frac{\exp(-E_n/RT)}{\exp(-E_{n+1}/RT)} > z, & \text{la conformation est acceptée} \\ \text{sinon,} & \text{elle est rejetée} \end{cases}$$

avec  $z$  un nombre compris entre 0 et 1.

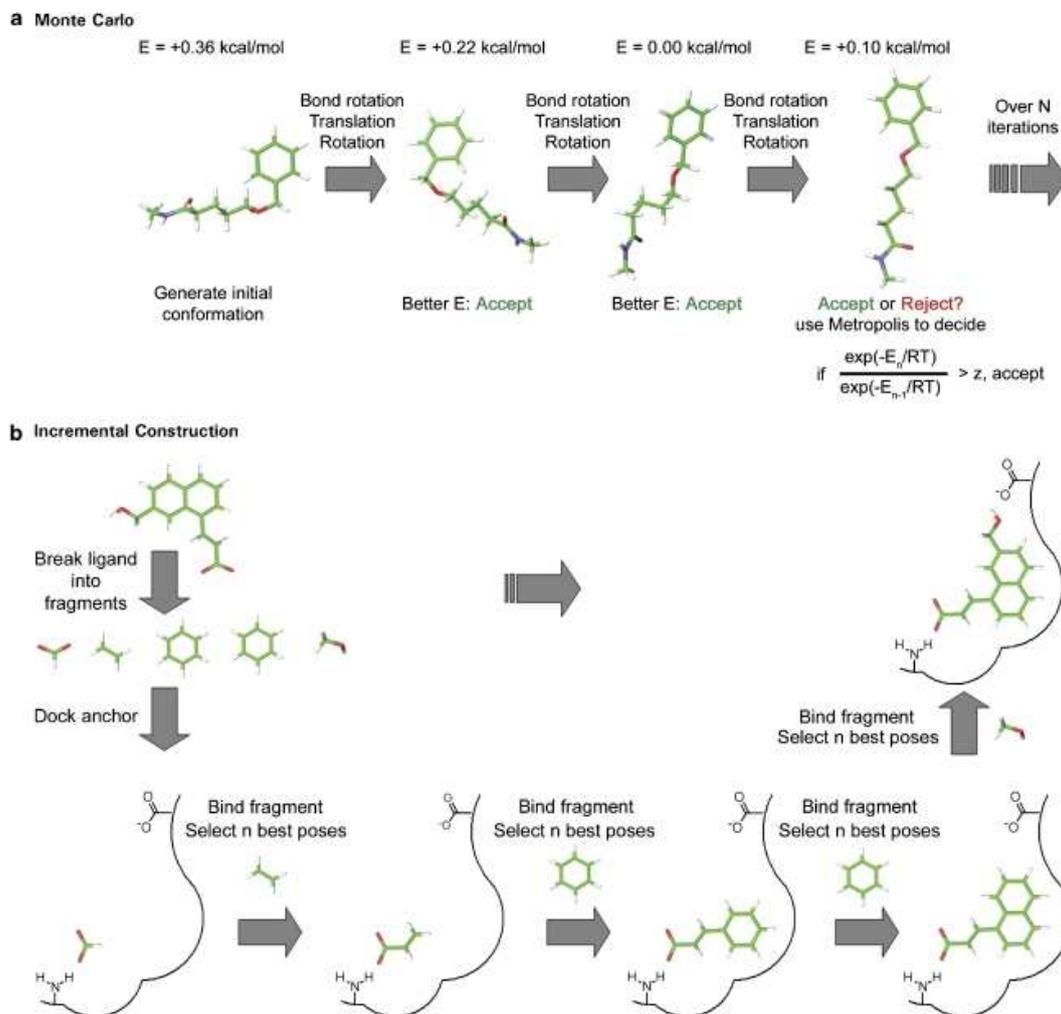


FIG. 4: Représentation schématique d'un algorithme génétique (a) et d'une recherche *Monte Carlo* (b), dans le contexte du *docking* d'un ligand flexible. Ce schéma est extrait de la revue de Moitessier et al. [190]

## 2.4 Fonctions de *scoring*

Dans chacun des algorithmes de *docking*, la recherche de placements du ligand dans le site de liaison s'accompagne de l'évaluation énergétique de ces placements. Les fonctions de score estiment l'affinité des ligands par une expression simplifiée de l'énergie de liaison  $\Delta G = \Delta H - T\Delta S$ . Cette estimation permet à la fois de guider le *docking* et de prédire l'affinité de liaison : généralement, la fonction utilisée pour évaluer la pose finale (optimale) du ligand est plus élaborée que la fonction utilisée pour guider le *docking*. La plupart des fonctions de scores présentent des limitations du fait de leur simplicité [190], et la difficulté à décrire complètement l'interaction entre un récepteur et un ligand avec un minimum de ressources se traduit souvent par une corrélation faible entre les scores prédits et les affinités mesurées. Ainsi, les fonctions de *scoring* ne prennent pas (ou mal) en

compte les aspects énergétiques de la désolvatation et l'entropie conformationnelle, de sorte que les résultats de criblage sont pollués par un grand nombre de faux positifs.

On distingue les fonctions de *scoring* (i) empiriques, (ii) celles basées sur un champ de force et (iii) celles basées sur la connaissance (des complexes protéine-ligand). Les fonctions empiriques, développées à partir du travail de Böhm [34], proposent d'évaluer l'énergie de liaison comme une somme simple de contributions (liaisons hydrogènes, liaisons ioniques, effets hydrophobes, rigidification des liaisons tournantes). Les fonctions basées sur un champ de force comprennent des termes d'interactions entre atomes non liés (électrostatiques et de van der Waals). Enfin, les fonctions basées sur la connaissance utilisent des potentiels de paires d'atomes, définis à partir des distributions statistiques des paires de types d'atomes rencontrées dans les structures cristallographiques de complexes protéine-ligand. Nous détaillons ici les fonctions de *scoring* de FlexX et ICM.

## FlexX

Dans FlexX, l'énergie de liaison est estimée par une fonction de *scoring* similaire à celle de Böhm [34] :

$$\begin{aligned} \Delta G = & \Delta G_0 + \Delta G_{rot} * N_{rot} + \Delta G_{hb} \sum_{\text{liais. H neutres}} f(\Delta R, \Delta \alpha) + \Delta G_{io} \sum_{\text{int. ioniques}} f(\Delta R, \Delta \alpha) \\ & + \Delta G_{aro} \sum_{\text{int. arom.}} f(\Delta R, \Delta \alpha) + \Delta G_{lipo} \sum_{\text{cont. lipo.}} f \star (\Delta R, \Delta \alpha) \end{aligned} \quad (21)$$

$N_{rot}$  est le nombre de liaisons tournantes qui sont immobilisées dans le complexe. La fonction de calibration  $f(\Delta R, \Delta \alpha)$ , qui permet de pénaliser les déviations par rapport à la géométrie idéale, et les paramètres ajustables  $\Delta G_0$ ,  $\Delta G_{rot}$ ,  $\Delta G_{hb}$  et  $\Delta G_{io}$ , proviennent directement du travail de Böhm. Un paramètre additionnel  $\Delta G_{aro} = -0.7 \text{ kJ.mol}^{-1}$  permet de prendre en compte les interactions impliquant des groupes aromatiques. La dernière contribution ( $\Delta G_{lipo}$ ) est une modification de l'énergie de contact lipophile de Böhm. Elle est sommée sur tous les contacts entre atomes appariés. La fonction  $f \star (\Delta R, \Delta \alpha)$  prend en compte les contacts ayant une distance plus ou moins idéale et pénalise les contacts trop proches. Elle est exprimée comme :

$$f \star (\Delta R, \Delta \alpha) = \begin{cases} 0, & \text{si } \Delta R > 0.6 \\ 1 - \frac{\Delta R - 0.2}{0.4}, & \text{si } 0.2 < \Delta R \leq 0.6 \\ 1, & \text{si } -0.2 < \Delta R \leq 0.2 \\ 1 - \frac{-\Delta R - 0.2}{0.4}, & \text{si } -0.6 < \Delta R \leq -0.2 \\ \frac{\Delta R + 0.6}{0.2}, & \text{si } \Delta R \leq -0.6 \end{cases}$$

$\Delta R = R - R_0$  étant la déviation de la distance entre les centres des atomes par rapport à la distance idéale, prise comme la somme des rayons de van der Waals augmentés chacun de  $0.3 \text{ \AA}$ .  $\Delta \alpha$  est la déviation de l'angle de la liaison hydrogène  $\angle_{N/O-H..O/N}$  par rapport à sa valeur idéale

de 180°.

## ICM

La fonction de *scoring* utilisée dans ICM, basée sur les champs de force d'AMBER et de ECEPP/3 [196], est exprimée comme :

$$E_{score} = \Delta E_{IntFF} + T\Delta S_{Tor} + \alpha_1 \Delta E_{HBond} + \alpha_2 \Delta E_{HBDesol} + \alpha_3 \Delta E_{SolEl} + \alpha_4 \Delta E_{HPhob} + \alpha_5 Q_{Size} \quad (22)$$

où  $\Delta E_{IntFF}$  inclut l'énergie moléculaire mécanique interne du ligand et l'interaction de van der Waals entre le ligand et le récepteur.  $T\Delta S_{Tor}$  est la perte d'entropie conformationnelle du ligand, proportionnelle au nombre d'angles de torsion libres immobilisés dans le complexe.  $\Delta E_{HBond}$  est le terme de liaison hydrogène.  $\Delta E_{HBDesol}$  est un terme qui rend compte de la rupture de liaisons hydrogènes avec le solvant.  $\Delta E_{SolEl}$  est l'énergie de désolvatation électrostatique, calculée en résolvant l'équation de Poisson-Boltzmann en utilisant la méthode des éléments finis [248].  $\Delta E_{HPhob}$  est le gain d'énergie libre hydrophobe, proportionnel à la surface accessible au solvant des atomes du récepteur et du ligand qui sont enfouis dans le complexe, avec un paramètre de tension de surface fixé à 0.012 kcal.Å<sup>-2</sup>.  $Q_{Size}$  est un terme de correction qui pénalise les gros ligands, pour éviter le biais introduit par l'augmentation du nombre d'interactions avec l'augmentation de la taille du ligand. Les poids  $\alpha_i$  ( $i \in 1, 2, \dots, 5$ ) ont été optimisés sur un ensemble divers de complexes.

## 2.5 Défis et perspectives

Le succès des approches de criblage virtuel reste encore largement conditionné par le choix du système. Chaque algorithme de *docking* et chaque fonction de *scoring* présente des imperfections et des biais différents, si bien qu'aucun n'est universel. Le choix des outils et des options utilisés se fait au cas par cas et l'inclusion d'un maximum de connaissance du système biologique étudié est bien souvent nécessaire pour avoir de bonnes chances de succès. Un des défis majeurs, pour améliorer la transférabilité des méthodes d'un système à l'autre, est la modélisation de la flexibilité du récepteur et le traitement du solvant, en particulier des molécules d'eau présentes dans le site de liaison. Des méthodes spécifiques permettent aujourd'hui d'accorder une certaine liberté aux chaînes latérales du site de liaison de la cible, mais la fixation d'un ligand peut être associée à des réarrangements conformationnels substantiels, difficiles à décrire. Les fonctions de *scoring* comportent beaucoup d'approximations et rendent mal compte des aspects énergétiques de la solvation et de l'entropie conformationnelle. La tendance actuelle, pour pallier les manques de chaque fonction particulière, consiste à regrouper plusieurs termes de différentes fonctions pour obtenir des fonctions consensus plus prédictives [84]. Les méthodes actuelles résultent d'un nécessaire compromis entre la rapidité de calcul et la qualité du modèle utilisé. Avec le développement toujours croissant des ressources informatiques, on peut espérer que de plus en plus d'aspects biophysiques des interactions protéine-ligand pourront être décrits avec justesse.



# Stabilité de l'Interaction EF-CaM

## 1 Introduction

EF-CaM donne un exemple de complexe protéique où la calmoduline, senseur ubiquitaire du calcium, est recrutée par un pathogène pour l'activation de sa toxine. CaM a été largement étudiée, par des méthodes expérimentales et computationnelles. Néanmoins, les mécanismes qui gouvernent son interaction avec le calcium et les déterminants de son affinité pour diverses (une grande variété de) cibles restent peu connus [55]. Ainsi, le travail présenté ici se focalise principalement sur les mouvements internes de CaM, lors de son interaction avec EF.

Les données expérimentales disponibles sur EF-CaM révèlent une influence des ions  $\text{Ca}^{2+}$  fixés à CaM sur la stabilité, voire la formation, du complexe. Ainsi, dans le premier volet de notre étude, nous avons cherché à mettre en évidence cette influence, par des simulations de dynamique moléculaire de EF-CaM, avec différents niveaux de calcium. Nous avons choisi d'analyser de manière détaillée, à la fois le comportement dynamique des deux partenaires de l'interaction EF-CaM, et la liaison des ions  $\text{Ca}^{2+}$  à CaM, au sein du complexe.

Trois simulations *DM* de 15 ns ont été réalisées à partir d'une structure expérimentale où CaM est chargée de deux ions  $\text{Ca}^{2+}$ . Cette structure a été perturbée, tantôt en retirant, tantôt en ajoutant deux ions  $\text{Ca}^{2+}$ . Les résultats de ces perturbations peuvent être comparés, entre eux, et avec les données expérimentales. La simulation du complexe perturbé, auquel on a ajouté deux ions  $\text{Ca}^{2+}$ , est également comparée à une simulation réalisée à partir d'une structure *RX* existante, contenant quatre ions  $\text{Ca}^{2+}$ . Les différences entre les structures initiales se répercutent dans les simulations, et sont révélées, à l'échelle atomique, à travers les changements conformationnels de CaM et la coordination du calcium. Enfin, des simulations permettant un échantillonnage accru au niveau local apportent une indication qualitative de la modulation de l'affinité de N-CaM pour le calcium.

Il est utile de signaler l'existence de simulations *DM* précédentes de EF-CaM, effectuées par Zhao & al [289]. Cependant, les auteurs n'avaient considéré que le complexe EF-(2Ca-CaM), le temps de simulation était de seulement 350 ps, et un rayon de coupure doux avait été utilisé pour les interactions de longue portée, au lieu du protocole du maillage particulier d'Ewald.

## 2 Matériels et Méthodes

### 2.1 Préparation des structures initiales

Le complexe EF-CaM a été simulé à partir de deux structures cristallographiques différentes. Un premier jeu de coordonnées a été obtenu à partir du fichier PDB 1K93 [77], dans lequel deux ions  $\text{Ca}^{2+}$  sont fixés sur C-CaM. Les chaînes C (EF) et F (CaM) ont été sélectionnées. Les résidus terminaux 768-798 de EF, dont certains désordonnés, ont été retirés. Trois complexes ont été construits à partir de ces coordonnées : (i) 1k93-2Ca, avec deux ions calcium, a été obtenu directement ; (ii) 1k93-0Ca, sans calcium, a été généré en retirant les deux ions  $\text{Ca}^{2+}$  ; (iii) 1k93-4Ca, avec quatre ions calcium, a été généré en ajoutant deux ions  $\text{Ca}^{2+}$  à N-CaM. Pour ce faire, le domaine N-terminal d'une structure récente de CaM libre, chargée de calcium (entrée PDB : 1OOJ) [240], a été superposé au domaine N-terminal de CaM liée à EF.

Un second jeu de coordonnées a été obtenu à partir du fichier PDB 1XFX [228], dans lequel quatre ions  $\text{Ca}^{2+}$  sont fixés sur la calmoduline. Les chaînes C (EF) et Q (CaM) ont été sélectionnées, en omettant les résidus 3-4 de CaM. Ces coordonnées ont fourni la structure initiale du complexe 1xfx-4Ca. La chaîne Q du fichier PDB 1XFX a également été isolée pour générer la structure initiale de CaM seule, libérée de EF : CaM-1xfx-4Ca.

	1k93-0Ca	1k93-2Ca	1k93-4Ca	1xfx-4Ca	CaM-1xfx-4Ca
Nombre de contre-ions	13	9	5	5	5
Dimensions de la boîte d'eau	124 x 82 x 99 Å <sup>3</sup>			108 x 104 x 91 Å <sup>3</sup>	58 x 64 x 76 Å <sup>3</sup>
Nombre de molécules d'eau	24 632	24 642	24 652	30 928	8 459
Nombre total d'atomes	83 851	83 879	83 907	103 291	27 601

TABLE I: Caractéristiques des systèmes simulés.

Les atomes d'hydrogènes ont été ajoutés avec le module LEAP d'AMBER 8.0 [41], en utilisant les paramètres parm99. Le système a été neutralisé avec des contre-ions  $\text{Na}^+$ , puis hydraté par une boîte d'eau, avec un rayon de coupure de 9 Å. Le potentiel TIP3P [138] a été utilisé pour modéliser les molécules d'eau et les frontières de la boîte ont été traitées avec les conditions limites périodiques 3D. Les caractéristiques des systèmes simulés sont détaillées dans la Table I. Les paramètres de Lennard-Jones des ions  $\text{Ca}^{2+}$  sont un rayon de van der Waals  $R = 1.7131$  Å et une profondeur de puits  $\epsilon = 0.459789$  kcal.mol<sup>-1</sup> [16]. Ils ont déjà permis de reproduire la géométrie de coordination du calcium observée expérimentalement, dans une simulation de CaM seule [87].

### 2.2 Simulations de dynamique moléculaire

Les trajectoires de dynamique moléculaire (*DM*) ont été calculées avec AMBER 8.0 [41]. Un rayon de coupure de 9 Å a été appliqué pour les interactions de Lennard-Jones, et les interactions longue-portée ont été calculées à l'aide du protocole du maillage particulier d'Ewald (*particle*

*mesh Ewald* ou *PME* [65]). Le système est couplé au thermostat de Berendsen [30] pour réguler la température et la pression.

Les simulations ont été initiées par une succession de minimisations sous contraintes. Le système a été chauffé pendant 10 ps jusqu'à 300 K à volume constant, tout en contraignant les atomes du complexe (ou de CaM seule, dans le cas CaM-1xfx-4Ca) et les positions des ions  $\text{Ca}^{2+}$  avec une constante de force de  $25 \text{ kcal.mol}^{-1}.\text{Å}^2$ . L'étape d'équilibration s'est déroulée de la façon suivante : un round *DM* de 5 ps, à volume constant ; quatre rounds *DM* de 2.5 ps, à pression constante, en réduisant les contraintes de position de  $25 \text{ kcal.mol}^{-1}.\text{Å}^2$  à  $5 \text{ kcal.mol}^{-1}.\text{Å}^2$  ; enfin, un dernier round de 10 ps, avec une contrainte de  $2.5 \text{ kcal.mol}^{-1}.\text{Å}^2$ .

Pour le chauffage et l'équilibration du complexe 1k93-4Ca, aucune contrainte n'a d'abord été appliquée sur les ions  $\text{Ca}^{2+}$  nouvellement ajoutés, ni sur les boucles de coordination (résidus 20-31 et 56-67). Cependant, à l'issue de l'étape d'équilibration, les deux cations N-terminaux avaient largement dévié de leurs positions initiales (Table II). Une seconde équilibration a donc été menée pendant 50 ps, en appliquant des contraintes de distance de  $10 \text{ kcal.mol}^{-1}.\text{Å}^2$  entre le premier ion  $\text{Ca}^{2+}$  et les oxygènes O $\delta$ 1 de Asp-20, O $\delta$ 2 de Asp-22, O $\delta$ 1 de Asp-24, O de Thr-26 et entre le second  $\text{Ca}^{2+}$  et les oxygènes O $\delta$ 1 de Asp-56, O $\delta$ 1 de Asp-58, O $\delta$ 1 de Asn-60, O de Thr-62. La plupart des atomes des boucles de coordination sont ainsi revenus à des distances du cation comparables à celles observées dans la structure *RX* 100J [240] de  $\text{Ca}^{2+}$ -CaM (Table II). Toutefois, les atomes O $\epsilon$ 1 et O $\epsilon$ 2 de Glu-31, O $\epsilon$ 1 et O $\epsilon$ 2 de Glu-67, sont plus éloignées, à 4.8-8.5 Å, comme observé dans la structure *RX* 1XFX [228] du complexe.

S1	Ca <sup>2+</sup> -CaM (100J)	1k93-4Ca 1 <sup>ere</sup> équil.	1k93-4Ca 2 <sup>eme</sup> équil.	S2	Ca <sup>2+</sup> -CaM (100J)	1k93-4Ca 1 <sup>ere</sup> équil.	1k93-4Ca 2 <sup>eme</sup> équil.
D20.O $\delta$ 1*	2.5	4.3	2.7	D56.O $\delta$ 1*	2.7	2.6	2.7
D20.O $\delta$ 2	4.2	2.6	2.6	D56.O $\delta$ 2	4.3	4.7	2.6
D22.O $\delta$ 1	3.5	2.6	2.6	D58.O $\delta$ 1*	2.8	9.9	2.6
D22.O $\delta$ 2*	3.0	4.2	2.6	D58.O $\delta$ 2	3.9	9.0	4.1
D24.O $\delta$ 1*	2.4	6.2	2.5	N60.O $\delta$ 1*	2.6	8.7	3.1
D24.O $\delta$ 2	4.2	7.9	4.4				
T26.O*	2.6	2.8	2.6	T62.O*	2.6	9.9	2.8
E31.O $\epsilon$ 1	2.7	6.6	8.5	E67.O $\epsilon$ 1	2.7	11.3	5.2
E31.O $\epsilon$ 2	2.5	5.3	7.3	E67.O $\epsilon$ 2	2.9	10.8	4.8

**TAB. II: Distances de coordination du calcium dans les sites S1 et S2.** Les distances (en Å) entre les ions  $\text{Ca}^{2+}$  et les atomes qui les coordonnent sont comparées, entre : (i) la structure 100J de  $\text{Ca}^{2+}$ -CaM [240] utilisée pour obtenir le complexe 1k93-4Ca, (ii) la conformation issue de la première étape d'équilibration de 1k93-4Ca (sans contrainte de distances), (iii) la conformation issue de la seconde étape d'équilibration, au cours de laquelle des contraintes de distances ont été appliquées entre les atomes marqués d'une astérisque \* et les positions des ions  $\text{Ca}^{2+}$ . O $\epsilon$ 1 et O $\epsilon$ 2 de E31, O $\epsilon$ 1 et O $\epsilon$ 2 de E67 restent plus éloignés des ions  $\text{Ca}^{2+}$ , après la seconde étape d'équilibration, que dans la structure 100J de  $\text{Ca}^{2+}$ -CaM libre. Ils sont aussi localisés loin des ions  $\text{Ca}^{2+}$  dans la structure cristallographique 1XFX de EF-(4Ca-CaM) [228], à des distances de 4.0 Å, 4.4 Å, 6.7 Å et 4.6 Å respectivement.

Les simulations *DM* ont été réalisées pendant 15 ns, à une température de 300 K et une pression de 1 atm. L'algorithme SHAKE [217] a été utilisé pour « geler » les liaisons covalentes

impliquant des hydrogènes à leurs valeurs d'équilibre, ce qui a permis un pas d'intégration de 2 fs. Les coordonnées atomiques ont été enregistrées toutes les ps, fournissant un ensemble de 15 000 conformations par trajectoire.

## 2.3 Analyse des trajectoires

Afin d'évaluer le degré de convergence des simulations *DM*, nous avons utilisé une approche récemment développée par Lyman & Zuckerman [174]. Elle consiste à générer, pour chaque trajectoire, un ensemble de structures représentatives de « référence », puis de classer toutes les conformations dans des groupes de « référence ».

Un rayon de coupure  $r$  est d'abord fixé de manière arbitraire. Une conformation est choisie au hasard dans la trajectoire : c'est la première structure de « référence ». Toutes les conformations présentant une déviation (*root mean square deviation* ou *RMSD*) plus petite que  $r$  par rapport à cette structure de « référence » sont retirées de la trajectoire. Une seconde structure de « référence » est ensuite tirée au hasard parmi les conformations restantes, et de nouveau toutes les conformations présentant un *RMSD* plus petit que  $r$  par rapport à cette structure de « référence » sont retirées. L'opération est répétée, jusqu'à épuisement des conformations. On obtient ainsi un ensemble de structures de « référence » différentes d'au moins  $r$  Å les unes des autres.

La trajectoire est ensuite divisée en deux moitiés. Les conformations issues de chaque moitié sont associées à la structure de « référence » la plus proche, en terme de *RMSD*, constituant ainsi des groupes de « référence ». Si la simulation a convergé, alors chaque groupe de « référence » est peuplé à parts égales avec des conformations issues de la première et de la deuxième moitié de la trajectoire. En tout, cinq exécutions de l'algorithme d'analyse de convergence ont été effectuées pour chaque trajectoire.

Pour comparer cette nouvelle méthode à une mesure plus traditionnelle, nous avons calculé la fonction d'auto-corrélation sur les positions des atomes  $C(\alpha)$  :

$$C(t) = \frac{N!}{2!(N-2)!} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \langle v_{ij}(t_0).v_{ij}(t_0+t) \rangle \quad (1)$$

où  $v_{ij}$  est le vecteur de position entre les  $C(\alpha)$   $i$  et  $j$ , et  $N$  est le nombre de  $C(\alpha)$ . La courbe obtenue peut être décomposée en fonctions exponentielles. Pour chaque simulation,  $C(t)$  a été ajustée sur un modèle exponentiel  $f(t) = c_1 e^{-t/\tau_1} + c_2 e^{-t/\tau_2} + c_3 e^{-t/\tau_3}$  [29], en utilisant Xmgrace [278].

Un angle dièdre virtuel (*Virtual Dihedral Angle* ou *VDA*) a été calculé entre les centres de masse des quatre motifs « EF-hand » de CaM. Il décrit la rotation relative des lobes N-CaM et C-CaM l'un par rapport à l'autre. Traditionnellement, le *VDA* est défini comme l'angle pseudo-dièdre formé par les quatre ions  $\text{Ca}^{2+}$  fixés à CaM [199], mais cette définition est pertinente seulement si l'hélice centrale est droite [203].

Pour évaluer le fléchissement de l'hélice centrale, l'angle entre les axes des segments hélicoïdaux 64-72 et 86-94 a été calculé. Les axes sont définis par les points successifs situés au milieu des atomes (i)  $N_i$  et  $N_{i+2}$ , (ii)  $C(\alpha)_i$  et  $C(\alpha)_{i+2}$ , (iii)  $C_i$  et  $C_{i+2}$ , avec  $i$  l'indice du résidu dans la séquence. L'ouverture des motifs « EF-hand » a été évalué de manière similaire, en calculant les angles formés entre les hélices HA (résidus 8-19) et HB (résidus 31-37) de S1, les hélices HC (résidus 46-53) et HD (résidus 64-73) de S2, les hélices HE (résidus 83-92) et HF (résidus 103-110) de S3, et les hélices HG (résidus 119-127) et HH (résidus 139-145) de S4.

L'analyse en composantes principales (*principal component analysis* ou *ACP*) des mouvements de EF et de CaM a été effectuée avec le module PTRAJ d'AMBER 8.0 [41] et 9.0 [43]. La structure secondaire a été déterminée avec le logiciel DSSP [139]. Les liaisons hydrogènes et interactions hydrophobes ont été identifiées avec le programme HBPLUS 2.25 [183] dans le paquet LIGPLOT 4.4.2 [271]. Le logiciel R 2.4.1 [204] a servi pour les graphiques d'analyse.

## 2.4 Simulations LES

Le principe des simulations permettant un échantillonnage accru au niveau local (*locally enhanced sampling* ou *LES*) est exposé au Chapitre 2, paragraphe 1.1.4. Elles ont été utilisées ici pour étudier la dissociation du calcium du domaine N-CaM. Les simulations ont été réalisées par Leandro Martínez, avec CHARMM version 33b2 [38]. Les conformations issues des étapes de thermalisation et d'équilibration de 1k93-4Ca et 1xfx-4Ca ont été prises comme points de départ. L'eau et les ions ont été retirés et plusieurs copies de  $\text{Ca}^{2+}$  ont été insérées. Pour chacun des deux sites S1 et S2 et chacun des deux complexes 1k93-4Ca et 1xfx-4Ca, 13 trajectoires de 200 ps ont été calculées avec 1, 2, 4, 8, 16, 18, 20, 22, 24, 26, 28, 30 et 32 copies de  $\text{Ca}^{2+}$ . Un ion  $\text{Ca}^{2+}$  est considéré comme dissocié s'il est à plus de 8 Å du centre de masse des résidus Asp-20, Asp-22, Asp-24, Thr-26, Glu-31 dans S1 ou Asp-56, Asp-58, Asn-60, Thr-62, Glu-67 dans le S2 à la fin des trajectoires. Pour mimer les effets du solvant, une fonction diélectrique sigmoïdale a été utilisée [164], avec  $\epsilon_{eel}$  proportionnel à 1.4 fois la distance inter-atomique et une fonction « switch » opérant entre 7 et 8 Å (Martínez et al., soumis).

# 3 Résultats et discussion

## 3.1 Stabilité du complexe EF-CaM

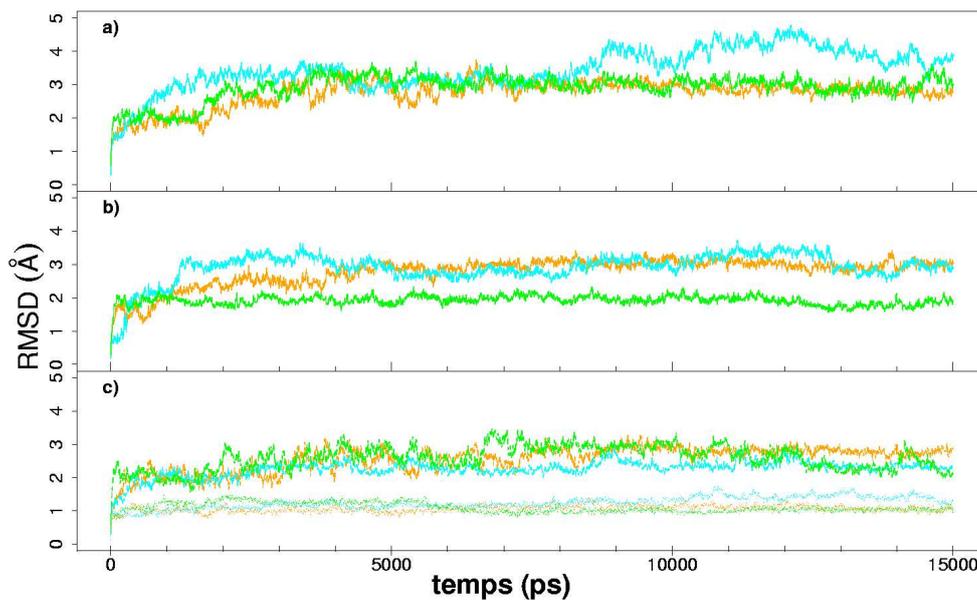
L'influence du niveau de calcium sur la stabilité du complexe a été étudiée, en se focalisant sur les trajectoires *DM* 1k93-0Ca, 1k93-2Ca et 1k93-4Ca, obtenues à partir d'une structure cristallographique commune.

### Analyse de convergence

La variation conformationnelle a été quantifiée à l'aide de la déviation standard (*RMSD*) des positions atomiques de CaM par rapport à la conformation initiale (Fig. 1). Les déviations des squelettes de CaM dans 1k93-2Ca et 1k93-4Ca se stabilisent autour de 2.7 Å (Fig. 1a), alors que la déviation dans 1k93-0Ca varie entre 3.5 et 5 Å jusqu'à la fin de la simulation. Ces valeurs sont petites par rapport au *RMSD* de 15 Å de CaM libre [229; 87], car l'amplitude des mouvements de la protéine est réduite par la présence de EF. Les déviations calculées pour EF sont de 3.7 Å (1k93-0Ca), 2.9 Å (1k93-2Ca) et 4.1 Å (1k93-4Ca).

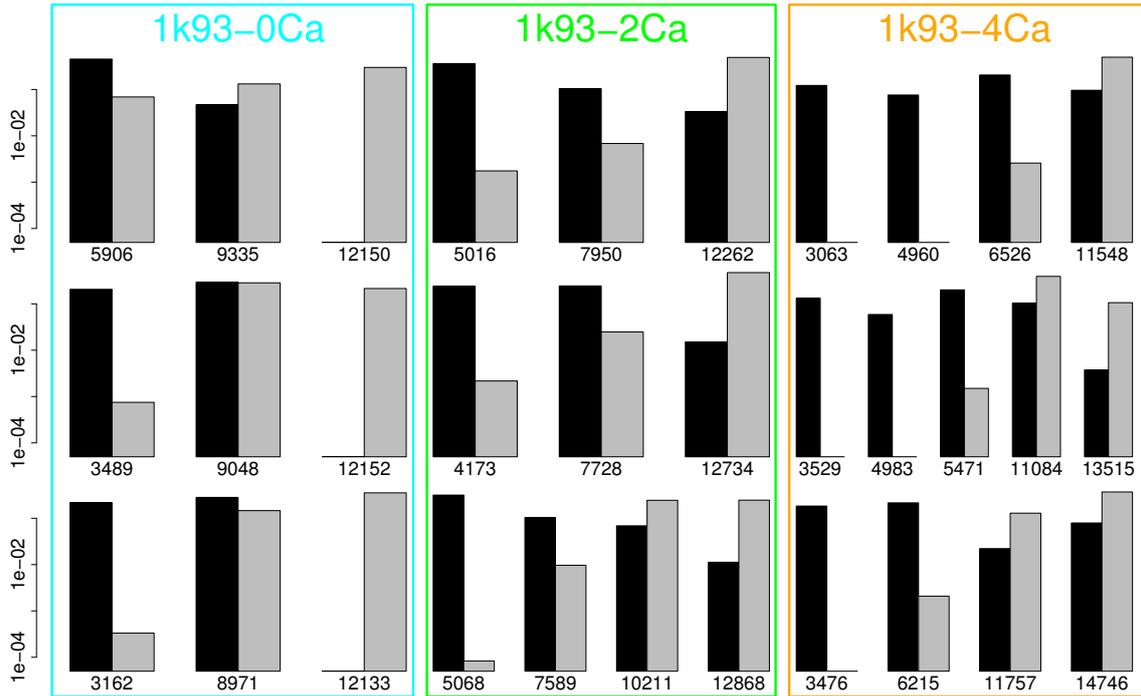
C-CaM (résidus 93-147) est très stable, autour de 1-1.5 Å, alors que le *RMSD* de N-CaM (résidus 5-64) atteint 3 Å (Fig. 1c). Les profils de *RMSD* de la protéine entière et de chaque domaine sont similaires, indiquant que les mouvements collectifs contribuent peu aux fluctuations de CaM en complexe avec EF, ce qui contraste avec CaM libre [229; 87].

Il apparaît sur ces profils *RMSD* que 3 ns sont nécessaires à la relaxation du complexe par rapport à la structure cristallographique initiale. Ainsi, dans la suite, la plupart des analyses ne prennent en compte que les 12 dernières ns des trajectoires.



**FIG. 1: *RMSD* du squelette de la calmoduline.** La déviation est calculée par rapport à la structure *RX* initiale (1K93) [77] : (a) *RMSD* total, (b) *RMSD* du linker central, en superposant les résidus 65-92, (c) *RMSD* des lobes N-CaM (*en haut, ligne en tirets*) et C-CaM (*en bas, ligne en pointillés*), en superposant les résidus 5-64 et 93-147 respectivement. 1k93-0Ca est en cyan, 1k93-2Ca en vert et 1k93-4Ca en orange.

La stabilité relative des complexes a été caractérisée à travers une analyse de convergence des trajectoires [174]. À chaque exécution de l'algorithme, des conformations différentes sont tirées au hasard dans les trajectoires *DM*. Ainsi, afin de s'assurer de la robustesse de la méthode, plusieurs analyses ont été menées, pour chaque trajectoire. Les résultats obtenus sont similaires (Fig. 2).



**FIG. 2: Analyse de convergence** : histogrammes de peuplement des groupes de référence pour (a) 1k93-0Ca, (b) 1k93-2Ca et (c) 1k93-4Ca. Les conformations de la première moitié de la trajectoire sont en noir et les conformations de la deuxième moitié sont en gris. Chaque ligne correspond à un choix aléatoire différent. Les numéros des structures de référence sont indiqués sur l'axe des abscisses et correspondent à leurs positions en ps le long de la trajectoire.

Un rayon de coupure fixé à 2.5 Å a permis d'extraire trois structures de référence de la trajectoire 1k93-0Ca (Fig. 2a), trois à quatre structures de 1k93-2Ca (Fig. 2b), et quatre à cinq structures de 1k93-4Ca (Fig. 2c). Tous les groupes de référence de 1k93-2Ca sont peuplés par des conformations issues des deux moitiés de la trajectoire (Fig. 2b), ce qui indique un bon degré de convergence. Les histogrammes de 1k93-0Ca dénotent aussi un bon degré de convergence, avec toutefois un ensemble de nouvelles conformations vers la fin de la simulation, autour de 12 ns (Fig. 2a). Pour 1k93-4Ca, la tendance à converger est moins nette : les conformations issues de la première moitié de la simulation sont sur-représentées dans les structures de référence (Fig. 2c). L'ajustement des fonctions d'auto-corrélation sur des courbes exponentielles confirme que la trajectoire 1k93-4Ca converge plus lentement que les trajectoires 1k93-0Ca et 1k93-2Ca (Table III).

**TAB. III: Coefficients  $c_1$  et  $\tau_1$  des courbes exponentielles**, qui s'ajustent le mieux sur les fonctions d'auto-corrélation de 1k93-0Ca, 1k93-2Ca et 1k93-4Ca. Les coefficients de corrélation  $R^2$  sont de 0.98-0.99.

	1k93-0Ca	1k93-2Ca	1k93-4Ca
$c_1$	0.37	0.39	0.27
$\tau_1$	45.9	59.8	139.1

## Dynamique Essentielle

La plasticité conformationnelle des partenaires de l'interaction EF-CaM a été caractérisée par leur dynamique essentielle [13]. Les trois premiers modes *ACP* identifiés chez CaM contribuent à hauteur de 61%, 58% et 63% des fluctuations totales du squelette de la protéine dans 1k93-0Ca, 1k93-2Ca et 1k93-4Ca. Comme attendu, ces contributions sont plus faibles que celles calculées pour CaM libre, de 88% et 71% en forme apo et chargée de calcium [284], révélant des mouvements moins collectifs de CaM en complexe avec EF. Dans 1k93-0Ca et 1k93-4Ca, les valeurs propres diminuent brusquement, alors que le profil de 1k93-2Ca est plus progressif (Fig. 3a). Ceci suggère que les mouvements de CaM sont encore moins collectifs dans 1k93-2Ca que dans les deux autres simulations : CaM fluctue autour d'une position stable dans 1k93-2Ca, alors qu'elle évolue vers de nouvelles conformations dans 1k93-0Ca et 1k93-4Ca.

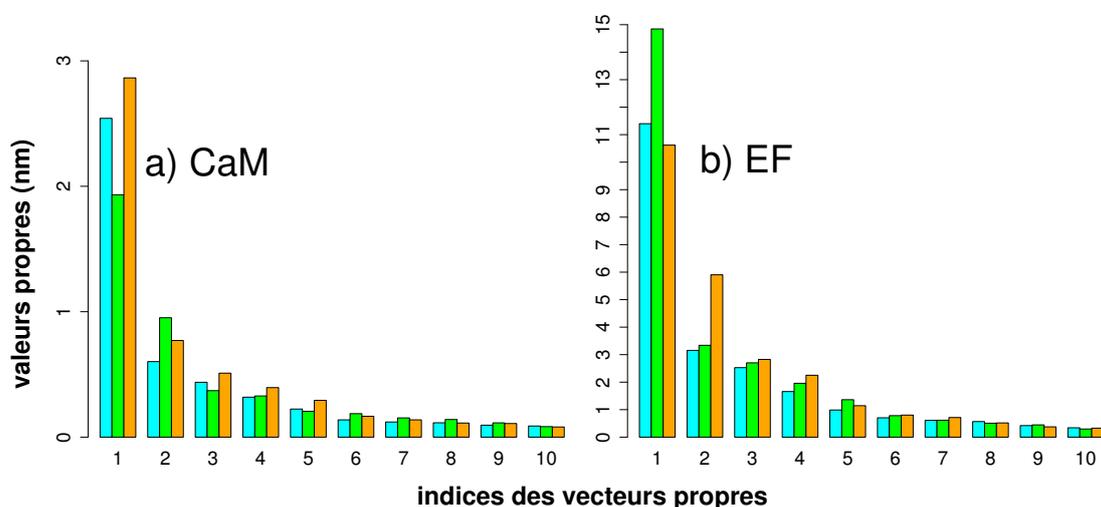


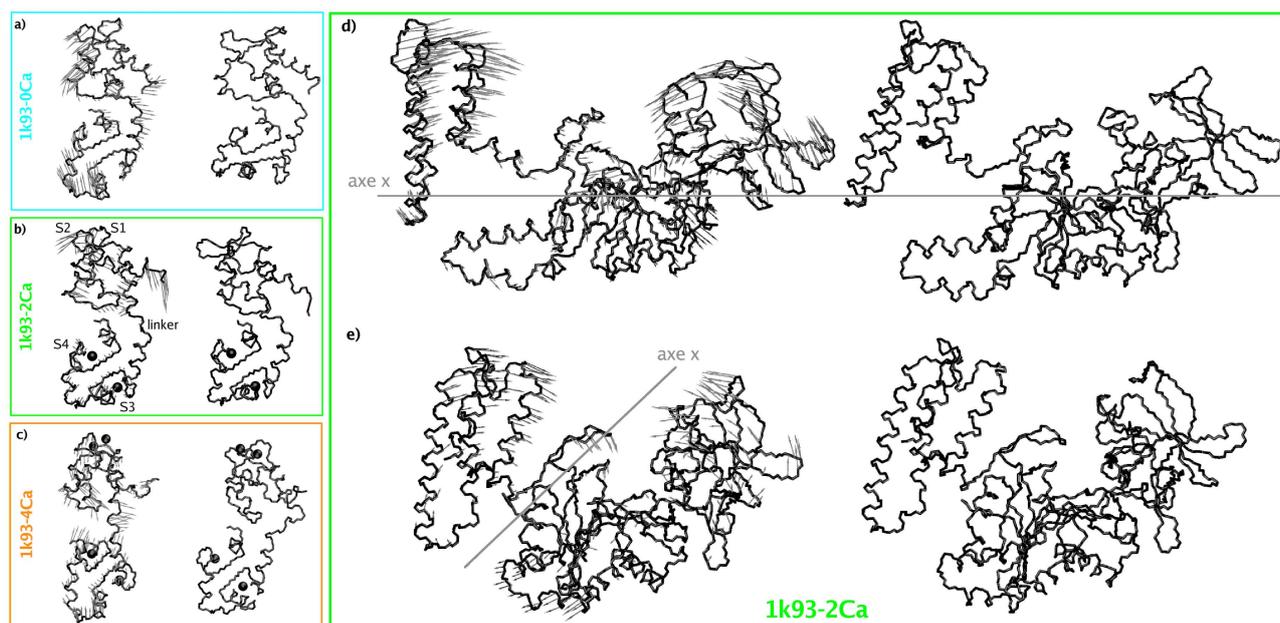
FIG. 3: Valeurs propres de la dynamique essentielle : (a) CaM, (b) EF. 1k93-0Ca est en cyan, 1k93-2Ca en vert et 1k93-4Ca en orange.

Les deux conformations extrêmes adoptées par CaM dans les simulations, le long du premier mode, sont représentées sur la Figure 4. Dans les trois complexes, on observe des déformations de N-CaM, particulièrement de la queue N-terminale dans 1k93-2Ca (Fig. 4b). Dans 1k93-0Ca (Fig. 4a) et 1k93-4Ca (Fig. 4c), on observe également un mouvement des deux lobes l'un vers l'autre (*flèches vers le haut et le bas*), tandis que le linker central est repoussé (*flèches vers la droite*). Ce mouvement de domaine, particulièrement prononcé dans 1k93-0Ca, pourrait expliquer l'augmentation du *RMSD* à la fin de la simulation (cf. Fig. 1a).

Dans l'analyse *ACP* de EF (Fig. 3b), la première valeur propre dans 1k93-2Ca est plus grande que les premières valeurs propres dans les deux autres complexes (Fig. 3b), alors que c'était le contraire chez CaM (Fig. 3a). Dans 1k93-2Ca, les valeurs propres de EF diminuent rapidement, et 19 modes sont suffisants pour décrire 90% des fluctuations totales de la protéine. Dans 1k93-4Ca et 1k93-0Ca, 21 et 24 modes sont nécessaires, signe que les mouvements de EF sont plus diffus

dans ces deux complexes que dans 1k93-2Ca. Là encore, c'est le contraire qui était observé chez CaM (Fig. 3a).

Dans 1k93-2Ca, les deux premiers modes décrivent respectivement une compaction (Fig. 4d) et une torsion (Fig. 4e). Ainsi, le domaine Hel (résidus 672-695) et le domaine CB (résidus 350-489) se rapprochent l'un de l'autre selon le premier mode, tandis qu'ils tournent l'un par rapport à l'autre selon le second mode. Les deux premiers modes dans 1k93-4Ca sont similaires. Par contre, dans 1k93-0Ca, les deux premiers modes de EF décrivent des déformations des régions les plus flexibles : le site catalytique, les boucles du domaine Hel et la boucle SA.



**FIG. 4: Conformations extrêmes selon les modes ACP.** Pour chaque protéine, la structure « initiale » est à gauche et la structure « finale », obtenue par déformation du squelette le long du mode considéré, est à droite. Les flèches grises représentent les composantes du mode sur chaque atome. (a-c) Conformations extrêmes de CaM le long du premier mode de sa dynamique essentielle dans : (a) dans 1k93-0Ca, le premier mode décrit des déformations de S2 (résidus 45-59) dans N-CaM, de la boucle HF-HG (résidus 108-119) dans C-CaM, et du linker central (résidus 74-80); (b) dans 1k93-2Ca, le premier mode décrit des déformations de la queue N-terminale (résidus 5-9) et de S2 (résidus 44-60) dans N-CaM; (c) dans 1k93-4Ca, le premier mode décrit des déformations de S1 (résidus 21-40) dans N-CaM, du linker central (résidus 74-79), et localement dans C-CaM (résidus 113-115 et 146-147). (d-e) Conformations extrêmes de EF dans 1k93-2Ca le long des (d) premier et (e) second modes de sa dynamique essentielle. La visualisation a été réalisée sous VMD [123], avec le module IED [191]

Au regard des résultats symétriques obtenus pour les deux partenaires, l'interprétation suivante peut être formulée : lorsque le complexe est stable,  $2\text{Ca}^{2+}$ -CaM établit une large surface de contact avec EF et sa mobilité est réduite ; la déstabilisation de l'interaction EF-CaM par l'ajout ou le retrait de calcium permet à CaM de fluctuer de manière plus collective et plus indépendante vis-à-vis de EF. EF adopte un comportement diamétralement opposé : dans le complexe avec deux ions  $\text{Ca}^{2+}$ , il a la liberté de fluctuer autour d'une position d'équilibre, à travers des mouvements globaux du domaine hélicoïdal et du coeur catalytique ; dans les formes déstabilisées, 1k93-0Ca et 1k93-4Ca, la mobilité de EF est contrainte et ses mouvements sont plus locaux, centrés sur des régions flexibles.

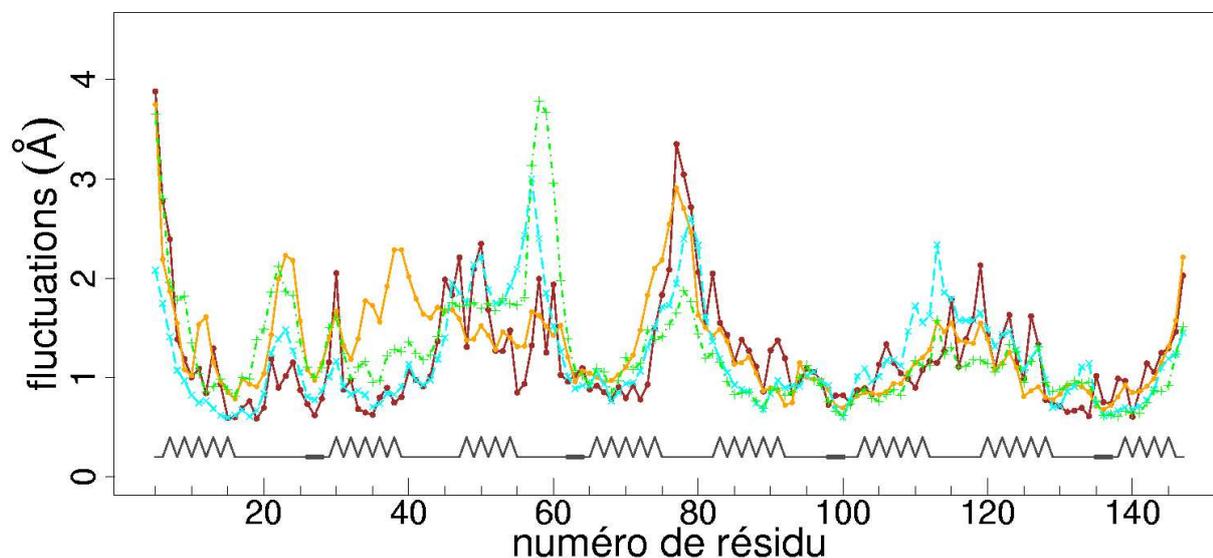
La trajectoire 1k93-2Ca présente les caractéristiques les plus stables en termes de convergence et de dynamique essentielle, en accord avec les résultats expérimentaux [255].

### 3.2 Plasticité conformationnelle des partenaires EF et CaM

La plasticité conformationnelle des partenaires de l'interaction EF-CaM a été caractérisée avec différents niveaux de calcium, en utilisant les trajectoires *DM* 1k93-0Ca, 1k93-2Ca et 1k93-4Ca.

#### Fluctuations atomiques

Les fluctuations atomiques par résidu de CaM et de EF ont été calculées. Les variations de fluctuations chez EF sont petites et ne sont donc pas décrites ici. Les fluctuations atomiques de CaM (Fig. 5) montrent que N-CaM (résidus 5-64) est plus flexible que C-CaM (résidus 93-147), en accord avec son plus grand *RMSD* (Fig. 1c) et avec le fait qu'il interagit plus faiblement avec EF [77; 228]. La fluctuation moyenne globale de C-CaM est d'environ 1 Å quel que soit le niveau de calcium. Par contre, les fluctuations moyennes globales de N-CaM et du linker central augmentent avec le niveau de calcium : les valeurs de fluctuations du linker sont 0.2 Å supérieures dans 1k93-4Ca, alors que les valeurs de fluctuations pour N-CaM sont respectivement 1.15, 1.35 et 1.41 Å dans 1k93-0Ca, 1k93-2Ca et 1k93-4Ca.



**FIG. 5: Fluctuations atomiques par résidu.** Les fluctuations atomiques du squelette de CaM sont calculées le long des trajectoires 1k93-0Ca, 1k93-2Ca, 1k93-4Ca, et aussi 1xfx-4Ca. 1k93-0Ca est en cyan, 1k93-2Ca en vert, 1k93-4Ca en orange et 1xfx-4Ca en marron. La structure secondaire de CaM est également indiquée au dessus de l'axe des abscisses.

Des études par *RMN* de CaM ont récemment montré que la dynamique interne de la protéine est affectée par la fixation de peptides cibles, et ont révélé l'importance de l'entropie conformationnelle dans la reconnaissance de ces peptides [94]. Le rôle crucial des changements dans la dynamique

des chaînes latérales a été mis en évidence dans les complexes smMLCKp-CaM and CaMKK $\alpha$ p-CaM [178; 93]. Pour vérifier si de tels changements pouvaient être observés dans CaM liée à EF, les variations de fluctuations suite à l'ajout ou au retrait de calcium ont été analysées le long des simulations (données non montrées). Globalement, les chaînes latérales et le squelette présentent des profils similaires et les changements dus au niveau de calcium sont de faible amplitude : la dynamique des chaînes latérales de CaM n'est pas particulièrement importante dans l'interaction EF-CaM. Néanmoins, l'augmentation de la mobilité de N-CaM dans 1k93-4Ca et de C-CaM dans 1k93-0Ca pourrait indiquer une augmentation de l'entropie conformationnelle, pour compenser la perte de complémentarité entre CaM et EF.

### Caractéristiques géométriques de CaM

Le rayon de giration de CaM varie entre 18.6 et 19.6 Å le long des simulations, avec une déviation standard de 0.3 Å (Table IV). Ces valeurs révèlent une conformation intermédiaire entre la structure cristallographique étendue de Ca<sup>2+</sup>-CaM (21.9 Å) [47] et la structure compacte de CaM liée au peptide smMLCK (16.1 Å) [184]. Dans 1k93-4Ca, le rayon est légèrement supérieur à ceux dans 1k93-2Ca et 1k93-0Ca (Table IV), dénotant une extension relative de la structure de la protéine.

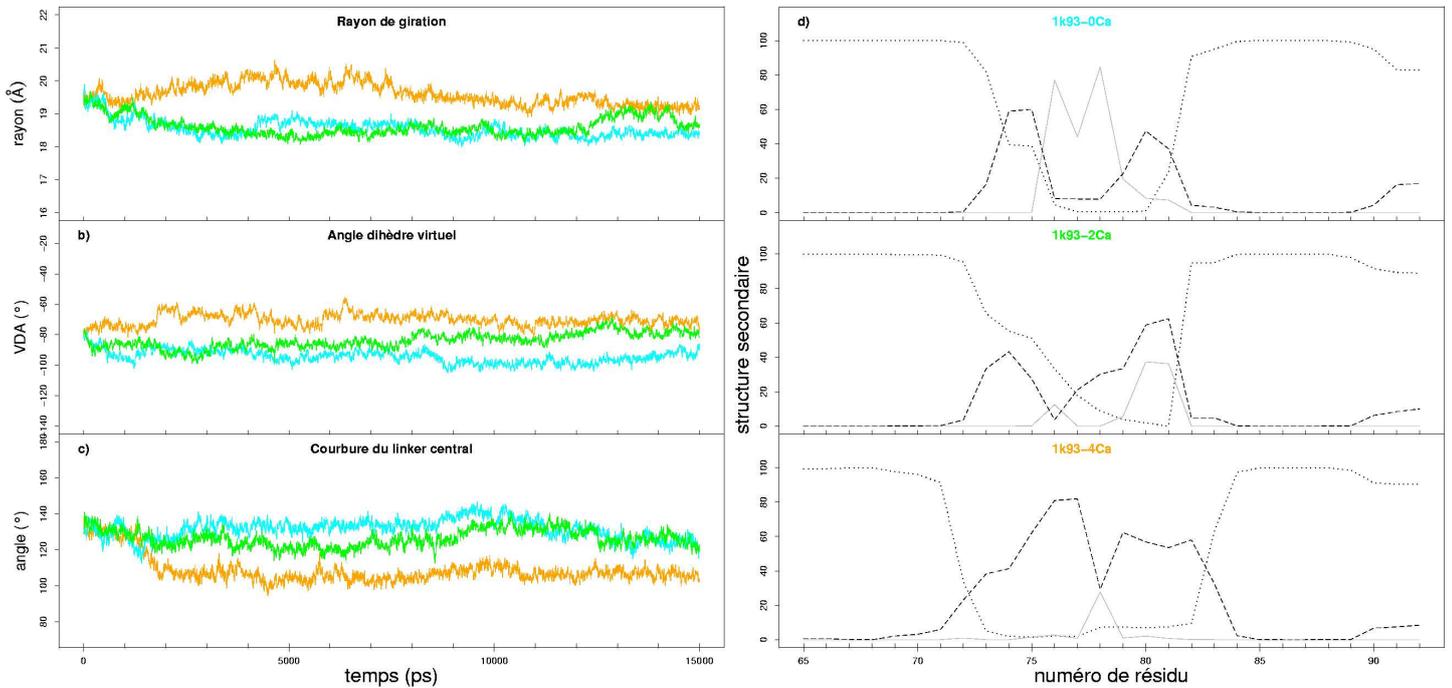
		Rayon de giration (Å)	VDA (°)	Courbure du linker (°)
Structure <i>RX</i>	1K93 (2Ca)	19.5	-80.5	124
Trajectoires <i>DM</i>	1k93-0Ca	18.5 ± 0.2	-95.2 ± 3.8	132 ± 5
	1k93-2Ca	18.5 ± 0.2	-82.7 ± 4.6	126 ± 5
	1k93-4Ca	19.6 ± 0.4	-69.8 ± 3.7	106 ± 4

**TAB. IV: Caractéristiques géométriques de CaM.** Le rayon de giration, l'angle dièdre virtuel et le fléchissement du linker central sont calculés sur la structure *RX* initiale 1K93 [77] d'une part, et moyennés le long des trajectoires *DM* 1k93-0Ca, 1k93-2Ca et 1k93-4Ca d'autre part.

L'angle dièdre virtuel (*VDA*) formé par les centres de masse des motifs « EF-hand » caractérise l'orientation relative des lobes (Fig. 6*b*). CaM adopte un angle aigu ou droit (entre -100 et -60°) pendant les simulations, comme dans d'autres complexes formés entre CaM et ses cibles [184], alors que CaM libre présente un angle aigu [47; 282; 229; 87]. Dans 1k93-0Ca et 1k93-4Ca, le *VDA* vaut environ -94° et -70° respectivement, alors qu'il fluctue entre ces deux valeurs dans 1k93-2Ca, autour de -84°. Ainsi, le retrait (ou l'ajout) de calcium stabilise l'orientation des lobes vers une plus grande (ou plus petite) valeur de *VDA*.

Le comportement du linker central est décrit à travers sa structure secondaire (Fig. 6*d*) et son degré de fléchissement (Fig. 6*c*). Le temps de présence de la structure secondaire hélicoïdale est le plus important dans 1k93-2Ca, à 74.6% du temps total de simulation (Fig. 6*d*), en accord avec la stabilité du *RMSD* (Fig. 1). La partie médiane du linker central subit une perte de structure hélicoïdale entre les résidus 73 et 82 dans 1k93-0Ca, 71 et 84 dans 1k93-2Ca, 69 et 84 dans

1k93-4Ca. Ces régions alternent entre des structures de type hélice  $\alpha$ , coude (« turn ») et pelote statistique (« random coil »), dans des proportions différentes selon le niveau de calcium. Le retrait de deux ions  $\text{Ca}^{2+}$  de C-CaM provoque le dépliement partiel du linker central (résidus 76-78) dans 1k93-0Ca, tandis que l'ajout de deux ions  $\text{Ca}^{2+}$  dans N-CaM provoque une transition partielle en coude (résidus 73-82) dans 1k93-4Ca.



**FIG. 6:** Évolution des caractéristiques géométriques de CaM le long des trajectoires *DM*. (a) Rayon de giration, (b) VDA, et (c) fléchissement du linker central de CaM, le long des simulations : 1k93-0Ca est en cyan, 1k93-2Ca en vert et 1k93-4Ca en orange. (d) Structure secondaire du linker central, moyennée le long des simulations. Le pourcentage de temps de simulation passé par chaque résidu dans chaque type de structure secondaire est reporté sur le graphe. Les lignes noire avec tirets courts, noire avec tirets longs et grise continue représentent les types hélice  $\alpha$ , coude (« turn ») et pelote statistique (« random coil »).

Le degré de fléchissement du linker est évalué par l'angle entre les hélices HD et HE (Fig. 6c). Alors que l'angle de courbure dans 1k93-0Ca et 1k93-2Ca est stable autour de  $132^\circ$  et  $126^\circ$ , il vaut environ  $109^\circ$  dans 1k93-4Ca. Ce fléchissement plus prononcé pourrait être le résultat de la conversion de la structure hélicoïdale en coude, les résidus centraux formant une charnière. De précédentes simulations *DM* de  $\text{Ca}^{2+}$ -CaM libre montrent aussi une perte de structure hélicoïdale dans le linker central [282; 283; 229], associée à un fléchissement. De plus, dans la simulation de 20 ns de Shepherd et Vogel [229], le dépliement partiel des résidus 75-81 participe au fléchissement du linker et à une compaction de CaM. Ici, nous n'observons toutefois pas de compaction, car le fléchissement n'est pas suffisant pour contre-balancer l'expansion de la protéine due à l'ajout d'ions  $\text{Ca}^{2+}$  dans N-CaM de 1k93-4Ca.

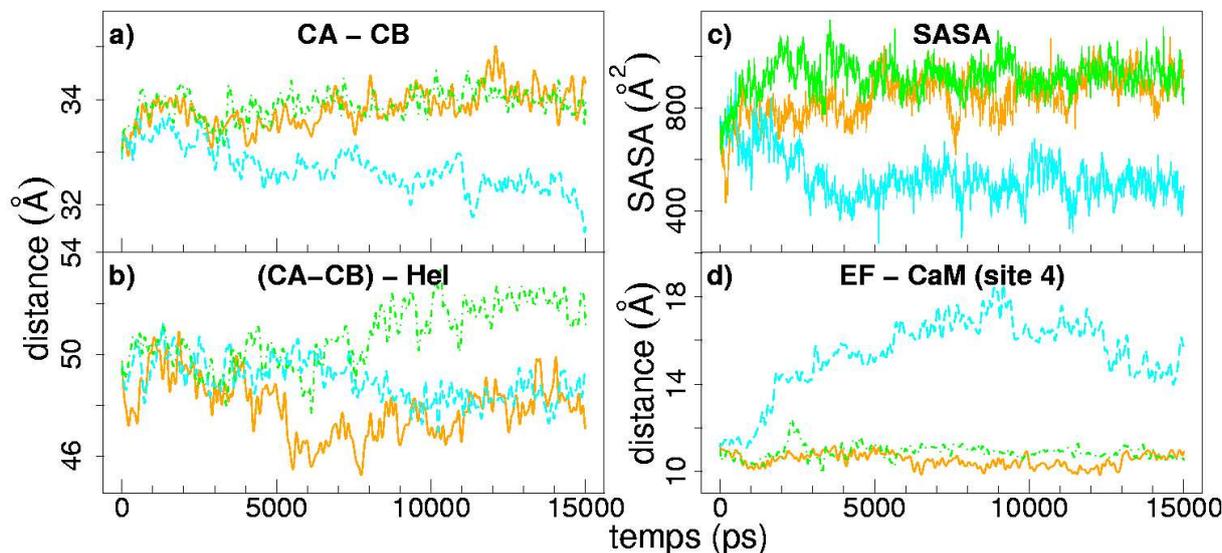
Il a été observé dans la littérature [229; 87] que le fléchissement ou dépliement du linker central est stabilisé de manière lâche par des liaisons hydrogènes, formées entre les résidus centraux et le

N-terminus. Bien qu'il manque les résidus 1 à 4 dans la séquence de CaM liée à EF, nous avons trouvé des interactions stabilisantes similaires dans 1k93-0Ca et 1k93-4Ca, mais pas dans 1k93-2Ca. Ce n'est pas surprenant, au regard de la plus grande flexibilité de la queue N-terminale dans 1k93-2Ca (Fig. 5a). Dans 1k93-0Ca, les liaisons hydrogènes T5.N-T77.O, T5.N-T79.O, T5.H $\gamma$ 1-T77.O, T5.O $\gamma$ 1-T79.H $\gamma$ 1 semblent stabiliser le dépliement du linker. Dans 1k93-4Ca, la liaison hydrogène Q8.O $\epsilon$ 1-T79.H $\gamma$ 1, formée dans l'intervalle 8-15 ns, et les contacts hydrophobes entre le N-terminus et les résidus du linker participent au fléchissement de celui-ci.

L'analyse du rayon de giration, de l'angle dièdre virtuel, de la structure secondaire et de la conformation du linker, et des contacts électrostatiques et hydrophobes, montre que les conformations de CaM dans 1k93-0Ca et 1k93-2Ca sont globalement proches, tandis que celle adoptée dans 1k93-4Ca est significativement différente.

### Conformation de EF et contacts avec CaM

L'analyse de la conformation de EF dans le complexe EF-CaM révèle un comportement calcium-dépendant, différent de celui de CaM. L'ajout de deux ions Ca<sup>2+</sup> a peu d'effet sur la conformation de EF, tandis que le retrait de calcium affecte la protéine entière, et plus particulièrement la géométrie du site catalytique. En effet, plusieurs distances caractéristiques de EF évoluent de manière similaire le long des simulations 1k93-2Ca et 1k93-4Ca (Fig. 7 et Table V).



**FIG. 7:** Évolution des caractéristiques géométriques de EF le long des trajectoires *DM* : (a) distance entre les centres de masse de CA (résidus 292-349 et 490-622) et CB (résidus 350-489), (b) distance entre les centres de masse du coeur catalytique (CA et CB) et le domaine Hel (résidus 660-767), (c) *SASA* du site catalytique (H351, K353, S354, K372, R329, K346, L348, D491, D493, H577, G578, T579, D582, N583, E588, F586, T548), (d) distance entre le centre de masse de la queue C-terminale de EF (résidus 758-767) et le centre de masse des atomes de coordination du calcium dans S4 de CaM. 1k93-0Ca est en cyan (tirets longs), 1k93-2Ca en vert (tirets pointillés) et 1k93-4Ca en orange (trait plein).

Par contre, les profils de 1k93-0Ca se démarquent, notamment la distance entre les domaines

CA et CB diminue le long de la trajectoire (Fig. 7a). L'ouverture du site catalytique, mesurée par la distance entre les centres de masse des résidus H577, G578, T579, D582, N583, E588, F586 de SB et H351, K353, S354, K372 de CB, est également significativement plus petite dans 1k93-0Ca que dans 1k93-2Ca et 1k93-4Ca (Table V).

Les distances triangulaires entre les résidus R329, K353, K346, sont également indicatrices de l'ouverture du site catalytique dans les trajectoires *DM* (Table V) : Dans 1k93-0Ca, R329.C $\gamma$  est plus proche de K353.N $\zeta$  et de K346.N $\zeta$  que dans 1k93-2Ca et 1k93-4Ca. Le calcul de la surface accessible au solvant (*SASA*) dans le site catalytique (Fig. 7c et Table V) renforce l'hypothèse d'un resserrement notable du site dans 1k93-0Ca. Les valeurs de *SASA* calculées dans 1k93-2Ca et 1k93-4Ca sont plus grandes, dénotant même un gain d'exposition par rapport à la structure *RX* initiale.

		SB-CB (Å)	R329.C $\gamma$ K353.N $\zeta$ (Å)	R329.C $\gamma$ - K346.N $\zeta$ (Å)	<i>SASA</i> (Å <sup>2</sup> )
EF libre struct. <i>RX</i>	1K8T (EF)	-	11.3	12.2	786
EF-CaM struct. <i>RX</i>	1K93 (2Ca)	13.6	7.5	9.6	703
EF-CaM traj. <i>DM</i>	1k93-0Ca	11.3 $\pm$ 0.5	6.4 $\pm$ 0.8	11.0 $\pm$ 1.0	501 $\pm$ 58
	1k93-2Ca	14.4 $\pm$ 1.1	11.1 $\pm$ 1.8	13.1 $\pm$ 1.2	928 $\pm$ 55
	1k93-4Ca	13.7 $\pm$ 0.9	11.4 $\pm$ 1.9	13.7 $\pm$ 1.6	866 $\pm$ 70

**TAB. V: Caractéristiques géométriques de EF.** Les distances caractéristiques de EF et la *SASA* sont d'une part calculées dans les structures *RX* de EF seul 1K8T [77] et du complexe 1K93 [77], d'autre part moyennées le long des simulations *DM* 1k93-0Ca, 1k93-2Ca et 1k93-4Ca. Les résidus pris en compte dans le calcul de la *SASA* du site catalytique sont H351, K353, S354, K372, R329, K346, L348, D491, D493, H577, G578, T579, D582, N583, E588, F586 et T548 [77]. La distance SB-CB est calculée entre les centres de masse des résidus H577, G578, T579, D582, N583, E588 et F586 de SB et des résidus H351, K353, S354 et K372 de CB. Les résidus R329, K346, K353 ont été choisis car ils forment un réseau de liaisons hydrogènes dans les structures cristallographiques du complexe [228].

Dans le but de décrire les contacts entre EF et CaM dans le complexe, les interactions hydrophobes et les liaisons hydrogènes établies pendant au moins 50% du temps de simulation ont été analysées (Table VI). 40 résidus de C-CaM, en moyenne, sont impliqués dans des contacts, alors que seulement 5 résidus de N-CaM interagissent avec EF (Table VI). Dans 1k93-2Ca, les liaisons hydrogènes constituent la majorité des contacts N-CaM/EF, tandis qu'ils sont minoritaires dans 1k93-0Ca et 1k93-4Ca. Cette inversion est en accord avec la variabilité conformationnelle de N-CaM observée dans les simulations (voir fluctuations Fig. 5). Le rapport liaisons hydrogènes sur contacts hydrophobes peut donc être relié à la spécificité de l'interaction EF-CaM.

Par ailleurs, la Figure 7d révèle que le retrait de calcium de C-CaM dans 1k93-0Ca autorise la queue C-terminale de EF à s'éloigner de S4 de CaM, jusqu'à 18 Å. Au contraire, la distance reste stable, autour de 10.5 Å, dans 1k93-2Ca et 1k93-4Ca. Dans les structures cristallographiques de EF-CaM, avec deux et quatre ions Ca<sup>2+</sup> [77; 228], un pont salin clé est formé entre Asp-647 de SC de EF et Arg-90 du linker de CaM. Deux ponts salins entre O $\delta$ 2 de Asp-647 et les N $\epsilon$  et N $\eta$ 2 de Arg-90 sont stables le long des trajectoires 1k93-0Ca et 1k93-2Ca. Dans 1k93-4Ca, les chaînes

**TAB. VI: Nombre de résidus de CaM impliqués dans des contacts avec EF.** Les ratios entre le nombre de résidus impliqués dans des liaisons hydrogènes et le nombre de résidus impliqués dans des interactions hydrophobes, pendant au moins 50% du temps de simulation, sont reportés. Les résidus sont regroupés par blocs :  $G_{HA}$  comprend l'hélice HA de S1 (résidus 5-22),  $G_{HB}$  comprend l'hélice HB de S1 (résidus 23-42),  $G_{C-CaM}$  comprend les hélices HE et HH (résidus 80-147).

	1k93-0Ca	1k93-2Ca	1k93-4Ca
$G_{HA}$	4/6	4/3	3/6
$G_{HB}$	2/4	1/1	0/2
$G_{C-CaM}$	12/28	15/29	16/28

latérales des deux résidus s'éloignent l'un de l'autre, mais les deux ponts salins sont re-formés vers 12 ns jusqu'à la fin de la trajectoire.

Les caractéristiques géométriques de CaM et de EF mesurées le long de la trajectoire 1k93-2Ca concordent avec les valeurs expérimentales calculées sur la structure  $RX$  initiale [77]. L'évolution de ces caractéristiques dans les trajectoires du complexe, perturbé par l'ajout ou le retrait de calcium, permet (i) d'élaborer une représentation plus dynamique des complexes 1k93-2Ca et 1k93-4Ca existants [77; 228], (ii) de donner des raisons de l'instabilité expérimentale de 1k93-0Ca [255].

La caractéristique dynamique la plus évidente est le rôle joué par le mouvement de fléchissement du linker central de CaM. En effet, les fluctuations de CaM dans le complexe sont pour la plupart le résultat de déformations dans N-CaM et du linker. Ainsi, dans 1k93-4Ca, la fixation de calcium provoque l'expansion de la protéine et un fléchissement notoire du linker, stabilisé par des contacts avec la queue N-terminale de CaM. Ce mouvement est important pour l'ajustement de la taille de CaM à la distance Hel-CB de EF. A travers cette variation d'ajustement, CaM joue le rôle d'un ressort qui maintient EF en conformation ouverte, induisant l'activation du site catalytique adényl cyclase, le niveau de complexation de deux  $Ca^{2+}$  permettant le meilleur ajustement de CaM à la conformation active ouverte de EF.

L'effet du retrait de tous les ions calcium du complexe donne des informations sur le système 1k93-0Ca, pour lequel aucune structure n'a été résolue. Ulmer et al. [255] ont proposé un modèle où apo-CaM n'interagit avec EF qu'à travers son lobe N-terminal. En effet, on observe que l'interaction entre EF et le domaine C-terminal de CaM est affaibli dans 1k93-0Ca.

Cependant, les analyses du rayon de giration, de l'angle dièdre virtuel, de l'ouverture des motifs « EF-hand » et de l'exposition des patches hydrophobes montrent que les conformations de CaM dans 1k93-0Ca et 1k93-2Ca sont proches, alors que celle dans 1k93-4Ca se démarque. Ceci suggère qu'une fois C-CaM inséré et maintenu entre le domaine hélicoïdal et le cœur catalytique de EF, la stabilité du complexe n'est pas affectée de manière dramatique par le retrait de calcium. Une barrière énergétique s'oppose à la formation de EF-(0Ca-CaM), mais il existe aussi une barrière énergétique qui s'oppose à la dissociation du complexe suite au retrait des ions  $Ca^{2+}$  C-terminaux. Les résultats présentés ici sont principalement influencés par cette seconde barrière.

De plus, les réactions à la perturbation du niveau de calcium suggèrent un mécanisme pour la déstabilisation du complexe EF-(0Ca-CaM). En effet, notre étude montre que l'ajout d'ions

$\text{Ca}^{2+}$  dans N-CaM augmente la flexibilité du lobe, alors que le retrait d'ions  $\text{Ca}^{2+}$  de C-CaM provoque peu de changements. C-CaM est profondément inséré entre le coeur catalytique de EF et le domaine Hel, de sorte que la tension induite par le retrait de calcium ne peut pas être relaxée dans le domaine lui-même. Par conséquent, elle se propage à EF, expliquant les déformations de EF observées dans 1k93-0Ca : rapprochement des domaines de EF et effondrement du site catalytique.

### 3.3 Inter-dépendance entre $\text{Ca}^{2+}$ /N-CaM et N-CaM/EF.

L'influence mutuelle entre les interactions  $\text{Ca}^{2+}$ /N-CaM et N-CaM/EF a été caractérisée sur les simulations 1k93-4Ca et 1xfx-4Ca, où CaM est chargée de quatre ions  $\text{Ca}^{2+}$ , et qui ont été obtenues à partir de deux structures *RX* différentes.

#### Changements conformationnels de CaM et coordination des ions calcium

Dans CaM libre, la fixation de calcium provoque l'ouverture des motifs « EF-hand » [286; 86; 229] : les angles inter-hélices passent de  $42\text{-}52^\circ$  chez apo-CaM à  $70\text{-}95^\circ$  chez  $\text{Ca}^{2+}$ -CaM [21; 47]. Cette ouverture s'accompagne de l'exposition au solvant de patches hydrophobes, impliqués dans l'interaction avec les peptides cibles [284].

Dans les structures *RX* et dans les simulations *DM* du complexe, les deux motifs « EF-hand » C-terminaux (S3 et S4) sont ouverts, à  $71.0\text{-}105.6^\circ$ , quel que soit le niveau de calcium (Table VII). Quant aux motifs N-terminaux (S1 et S2), ils sont fermés dans les structures 1K93 et 1XFX, tandis que leur degré d'ouverture varie dans les simulations (Table VII). S1 garde une conformation fermée quel que soit le niveau de calcium. S2 adopte une conformation ouverte suite à l'ajout de calcium dans 1k93-4Ca, alors que son angle est stable autour de  $55.3^\circ$  dans 1xfx-4Ca. Globalement, S2 présente une plus grande variabilité que les autres sites, comme Shepherd et Vogel [229] l'ont montré dans leur simulation de  $\text{Ca}^{2+}$ -CaM libre.

		Site 1 ( $^\circ$ )	Site 2 ( $^\circ$ )	Site 3 ( $^\circ$ )	Site 4 ( $^\circ$ )	<i>SASA</i> ( $\text{\AA}^2$ )	
						N-CaM	C-CaM
CaM libre struct. <i>RX</i>	1CFD (0Ca)	39.4	51.6	48.7	46.5	406	490
	1CLL (2Ca)	85.8	93.2	70.4	86.1	694	784
EF-CaM struct. <i>RX</i>	1K93 (2Ca)	$45.5 \pm 0.2$	$56.3 \pm 0.0$	$78.1 \pm 0.1$	$88.3 \pm 0.8$	641	784
	1XFX (4Ca)	$45.8 \pm 0.0$	$58.8 \pm 0.2$	$76.4 \pm 0.0$	$83.0 \pm 0.0$	468	733
EF-CaM traj. <i>DM</i>	1k93-0Ca	$33.4 \pm 3.9$	$62.6 \pm 8.5$	$79.8 \pm 3.8$	$94.8 \pm 4.3$	$443 \pm 45$	$810 \pm 38$
	1k93-2Ca	$41.2 \pm 2.3$	$60.2 \pm 7.3$	$74.4 \pm 3.4$	$98.3 \pm 3.8$	$476 \pm 68$	$835 \pm 32$
	1k93-4Ca	$43.7 \pm 7.2$	$75.7 \pm 7.7$	$80.4 \pm 2.7$	$101.0 \pm 4.6$	$505 \pm 42$	$840 \pm 31$
	1xfx-4Ca	$36.7 \pm 4.1$	$55.3 \pm 4.5$	$79.3 \pm 2.9$	$95.4 \pm 3.8$	$457 \pm 44$	$742 \pm 32$

**TAB. VII: Ouverture des motifs « EF-hand » de CaM.** Les angles inter-hélicoïdaux des sites de fixation au calcium S1, S2, S3 et S4, ainsi que l'exposition des patches hydrophobes dans N-CaM et C-CaM sont calculés sur les structures cristallographiques 1CFD de apo-CaM [154], 1CLL de  $\text{Ca}^{2+}$ -CaM [47], 1K93 [77] et 1XFX [228] de EF-CaM, et son moyennés le long des simulations 1k93-0Ca, 1k93-2Ca, 1k93-4Ca et 1xfx-4Ca. L'exposition des patches hydrophobes est évaluée en sommant la *SASA* des résidus concernés/impliqués, selon la définition de Yang et al. [284].

L'exposition des patches hydrophobes, définis par Yang et al. [284], a été évaluée en calculant leur *SASA* (Table VII). Les valeurs d'exposition de C-CaM dans EF-CaM sont similaires à celles dans Ca<sup>2+</sup>-CaM libre [47] (entrée PDB : 1CLL). Par contre, les valeurs de N-CaM dans EF-CaM sont proches de celles d'apo-CaM libre [154] (entrée PDB : 1CFD), même si un gain d'exposition de 12% est observé pour ce lobe, suite à l'ajout de calcium dans 1k93-4Ca.

L'ouverture de S2 dans 1k93-4Ca, induite par la fixation d'un ion Ca<sup>2+</sup>, est accompagnée d'une rigidification de la partie N-terminale de sa boucle calcique (résidus 54-60) par rapport à 1k93-2Ca (Fig. 5, section 3.2). Comme par un effet de contre-balancement, la région située en amont (résidus 29-47), comprenant HB et la partie N-terminale de la boucle HB-HC, devient plus flexible. Inversement dans 1xfx-4Ca, HC et la partie N-terminale de la boucle calcique (résidus 54-60) sont globalement plus flexibles, tandis que la région en amont (résidus 29-44) est plus rigide, à un niveau similaire à 1k93-0Ca. Les fluctuations de S1 (résidus 20-31) augmentent dans 1k93-4Ca par rapport à 1k93-2Ca, alors qu'elles sont inférieures à 1.5 Å dans 1xfx-4Ca, à l'exception notable de Lys-30.

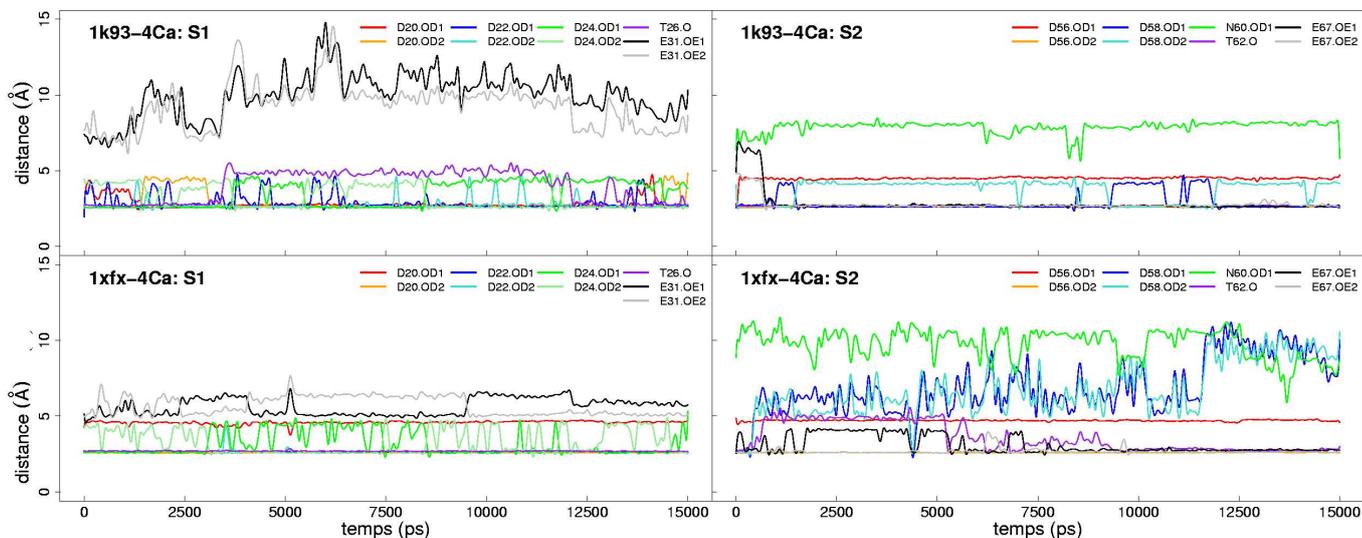
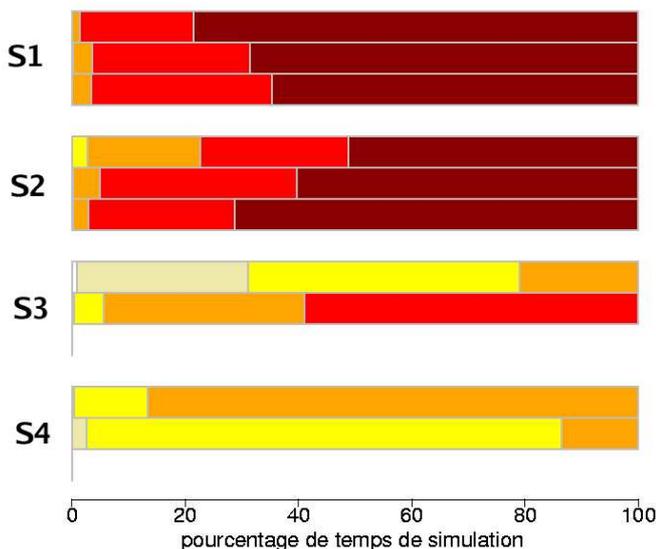
Chaque ion calcium est coordonné par les résidus conservés Asp<sup>1</sup>, Asp<sup>3</sup>, (Asp/Asn)<sup>5</sup>, X<sup>7</sup> et Glu<sup>12</sup>, en positions 1, 3, 5, 7 et 12 dans la séquence de la boucle calcique. La coordination complète du calcium est atteinte lorsque les six atomes d'oxygène suivants sont à 2.6 Å du cation : un des deux oxygènes O $\delta$  de Asp<sup>1</sup>, un des deux oxygènes O $\delta$  de Asp<sup>3</sup>, un oxygène O $\delta$  de (Asp/Asn)<sup>5</sup>, l'oxygène du squelette du résidu en position 7, et les deux oxygènes O $\epsilon$  de Glu<sup>12</sup>. Les ions Ca<sup>2+</sup> C-terminaux (S3 et S4) bénéficient d'une coordination complète ou quasi-complète le long des simulations 1k93-2Ca, 1k93-4Ca et 1xfx-4Ca (Fig. 8a). Ceci démontre que la fixation du calcium dans ces sites est stable est qu'elle n'est pas affectée par la présence de calcium dans N-CaM. Par contre, les ions Ca<sup>2+</sup> N-terminaux (S1 et S2) sont seulement partiellement coordonnés dans 1k93-4Ca et 1xfx-4Ca. Dans 1k93-4Ca, l'ion Ca<sup>2+</sup> fixé à S1 est coordonné par trois oxygènes pendant plus de 80% du temps de simulation (Fig. 8a). L'attachement de l'ion Ca<sup>2+</sup> dans S2 est plus fort, avec cinq oxygènes pendant plus de 50% du temps, en accord avec la rigidification de la boucle calcique (Fig. 5). Dans 1xfx-4Ca, l'ion Ca<sup>2+</sup> fixé à S1 est coordonné par quatre oxygènes pendant plus de 80% du temps de simulation, tandis que l'ion Ca<sup>2+</sup> fixé à S2 est coordonné par trois oxygènes pendant 50% du temps, et par deux ou quatre oxygènes pendant 30% ou 20% du temps (Fig. 8a). La coordination du calcium dans S2 est donc plus faible dans 1xfx-4Ca que dans 1k93-4Ca. C'est la tendance inverse qui est observée pour S1.

Dans S1 (Fig. 8b et 8d), il manque les deux oxygènes O $\epsilon$  de Glu-31 dans la sphère de coordination de l'ion Ca<sup>2+</sup>, le long des deux trajectoires 1k93-4Ca et 1xfx-4Ca. Glu-31 est en position 12 dans la boucle calcique et sera donc dénoté Glu<sup>12</sup>-31 à partir de maintenant. Une notation similaire est adoptée pour les autres résidus coordonnant le calcium. La distance entre les oxygènes de Glu<sup>12</sup>-31 (*courbes en noir et gris*) et l'ion calcium est d'environ 10 Å dans 1k93-4Ca (Fig. 8b), largement plus grande que celle de 6 Å dans 1xfx-4Ca (Fig. 8d). De plus, l'oxygène du squelette de Thr<sup>7</sup>-26 a tendance à s'éloigner à 5 Å de Ca<sup>2+</sup> dans 1k93-4Ca (Fig. 8b : *courbe en violet*),

tandis qu'il reste dans la sphère de coordination de  $\text{Ca}^{2+}$  dans 1xfx-4Ca, à 2.6 Å (Fig. 8d : *courbe en violet, dans la ligne de base*).

**FIG. 8: Coordination des ions calcium.**

(a) Nombre d'oxygènes qui coordonnent l'ion  $\text{Ca}^{2+}$  dans S1, S2, S3, S4 de 1k93-2Ca, 1k93-4Ca et 1xfx-4Ca, de bas en haut : le code couleur de blanc à rouge foncé correspond à des modes de coordination allant de un à six oxygènes. (b-e) Distances de coordination dans les sites de fixation du calcium : (b) S1 de 1k93-4Ca, (c) S2 de 1k93-4Ca, (d) S1 de 1xfx-4Ca, (e) S2 de 1xfx-4Ca. Dans S1, les distances sont calculées entre le cation et O $\delta$ 1 de Asp<sup>1-20</sup> (rouge), O $\delta$ 2 de Asp<sup>1-20</sup> (orange), O $\delta$ 1 de Asp<sup>3-22</sup> (bleu), O $\delta$ 2 de Asp<sup>3-22</sup> (cyan), O $\delta$ 1 de Asp<sup>5-24</sup> (vert foncé), O $\delta$ 2 de Asp<sup>5-24</sup> (vert clair), O de Thr<sup>7-26</sup> (violet), O $\epsilon$ 1 de Glu<sup>12-31</sup> (noir), et O $\epsilon$ 2 de Glu<sup>12-31</sup> (gris). Dans S2, les distances sont calculées entre le cation et O $\delta$ 1 de Asp<sup>1-56</sup> (rouge), O $\delta$ 2 de Asp<sup>1-56</sup> (orange), O $\delta$ 1 de Asp<sup>3-58</sup> (bleu), O $\delta$ 2 de Asp<sup>3-58</sup> (cyan), O $\delta$ 1 de Asn<sup>5-60</sup> (vert foncé), O de Thr<sup>7-62</sup> (violet), O $\epsilon$ 1 de Glu<sup>12-67</sup> (noir), et O $\epsilon$ 2 de Glu<sup>12-67</sup> (gris). La distance optimale de coordination est de 2.6 Å.



Dans S2 (Fig. 8c et 8e), les deux oxygènes O $\epsilon$  de Glu<sup>12-67</sup> sont positionnés à plus de 4 Å de l'ion  $\text{Ca}^{2+}$  dans les deux structures initiales de 1k93-4Ca et 1xfx-4Ca. Cependant, ils entrent rapidement dans la sphère de coordination à 2.6 Å dans les simulations (*courbes en noir et gris*). Ils restent à cette distance tout le long de la trajectoire 1k93-4Ca (Fig. 8d) et coordonnent le cation pendant toute la deuxième moitié de la trajectoire 1xfx-4Ca (Fig. 8e). Par contre, O $\delta$ 1 d'Asn<sup>5-60</sup> (*courbe en vert foncé*) est à plus de 7 Å du  $\text{Ca}^{2+}$  dans les deux trajectoires. La différence la plus flagrante est observée pour Asp<sup>3-58</sup> (*courbes en cyan et bleu*) : les deux oxygènes O $\delta$  échangent leurs positions le long de la trajectoire 1k93-4Ca (Fig. 8c), tandis qu'ils sortent rapidement de la sphère de coordination dans 1xfx-4Ca (Fig. 8e).

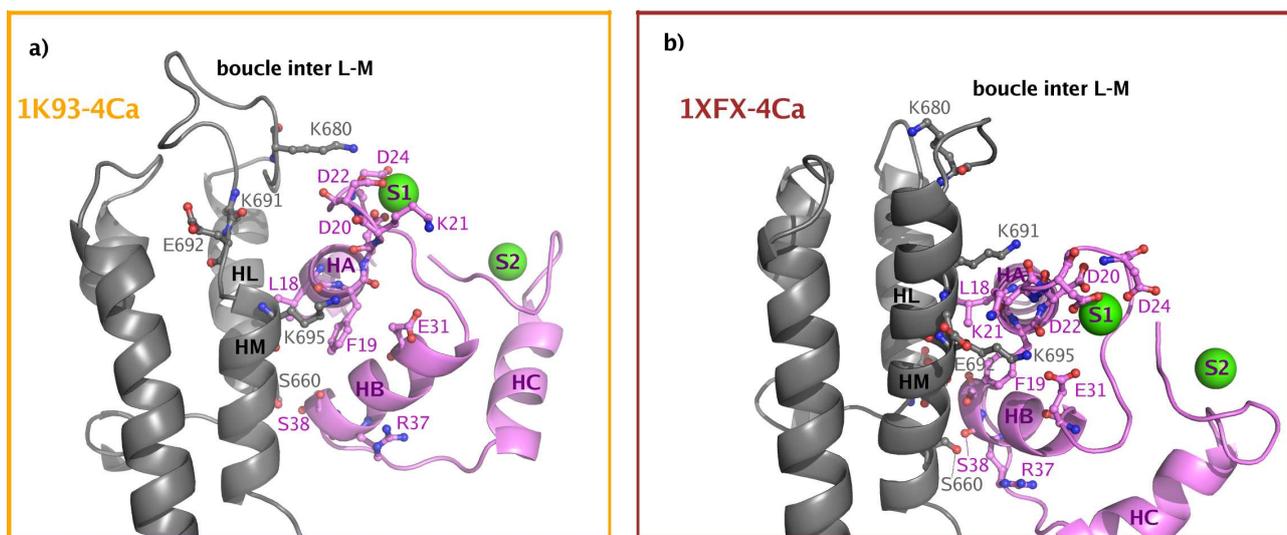
Comparé à 1k93-4Ca, 1xfx-4Ca présente une plus forte coordination du calcium dans S1, mais une coordination plus instable dans S2. De tels résultats sont en accord avec, d'une part la rigidité de S1, et d'autre part la flexibilité de S2, accrues dans 1xfx-4Ca par rapport à 1k93-4Ca.

Les deux simulations 1k93-4Ca et 1xfx-4Ca révèlent une coordination incomplète des ions  $\text{Ca}^{2+}$  fixés à N-CaM, en accord avec l'affinité réduite de ce domaine pour le calcium. L'ajout de deux ions  $\text{Ca}^{2+}$  provoque l'ouverture partielle du second motif « EF-hand » dans 1k93-4Ca, alors que les deux motifs sont bloqués en conformation fermée dans 1xfx-4Ca. S2 n'est pas directement contraint par EF, ce qui l'autorise à une plus grande plasticité que S1, de telle sorte que Glu<sup>12-67</sup> entre durablement dans la sphère de coordination dans les deux simulations. Au contraire, Glu<sup>12-31</sup> ne peut atteindre la sphère de coordination du  $\text{Ca}^{2+}$  dans S1, soit à cause de la très grande flexibilité des résidus 29-44, induite par la fixation de calcium sur S2 dans 1k93-4Ca (Fig. 5a), soit à cause d'interactions électrostatiques avec EF dans 1xfx-4Ca (Fig. 10b).

Afin de tester la capacité de  $4\text{Ca}^{2+}$ -CaM à s'ouvrir une fois libérée de EF, la protéine a été simulée pendant 15 ns, EF ayant été retiré de la structure *RX* 1XFX [228]. La déviation moyenne est de 7.8 Å le long de cette trajectoire, et la simulation ne converge pas vers l'ouverture de N-CaM. Les angles inter-hélicoïdaux sont plutôt stables autour de  $40 \pm 3^\circ$  dans S1,  $47 \pm 5^\circ$  dans S2,  $84 \pm 10^\circ$  dans S3 et  $76 \pm 3^\circ$  dans S4. La SASA des patches hydrophobes vaut environ 400 Å<sup>2</sup> dans N-CaM et 800 Å<sup>2</sup> dans C-CaM. Une analyse de la dynamique essentielle a permis d'observer une ré-orientation des lobes l'un par rapport à l'autre, conduisant à un VDA de  $-105.6 \pm 11.4^\circ$ . Il apparaît donc plus facile pour  $4\text{Ca}^{2+}$ -CaM libérée de EF de ré-orienter ses lobes plutôt que d'ouvrir son domaine N-terminal.  $4\text{Ca}^{2+}$ -CaM garde la conformation intermédiaire adoptée dans le complexe EF-( $4\text{Ca}$ -CaM), avec les Glu-12 hors des sphères de coordination du calcium. Cette conformation intermédiaire est similaire à la structure de CaM piégée par un pont disulfure, résolue par Grabarek [103].

### Modulation de l'affinité de N-CaM pour le calcium

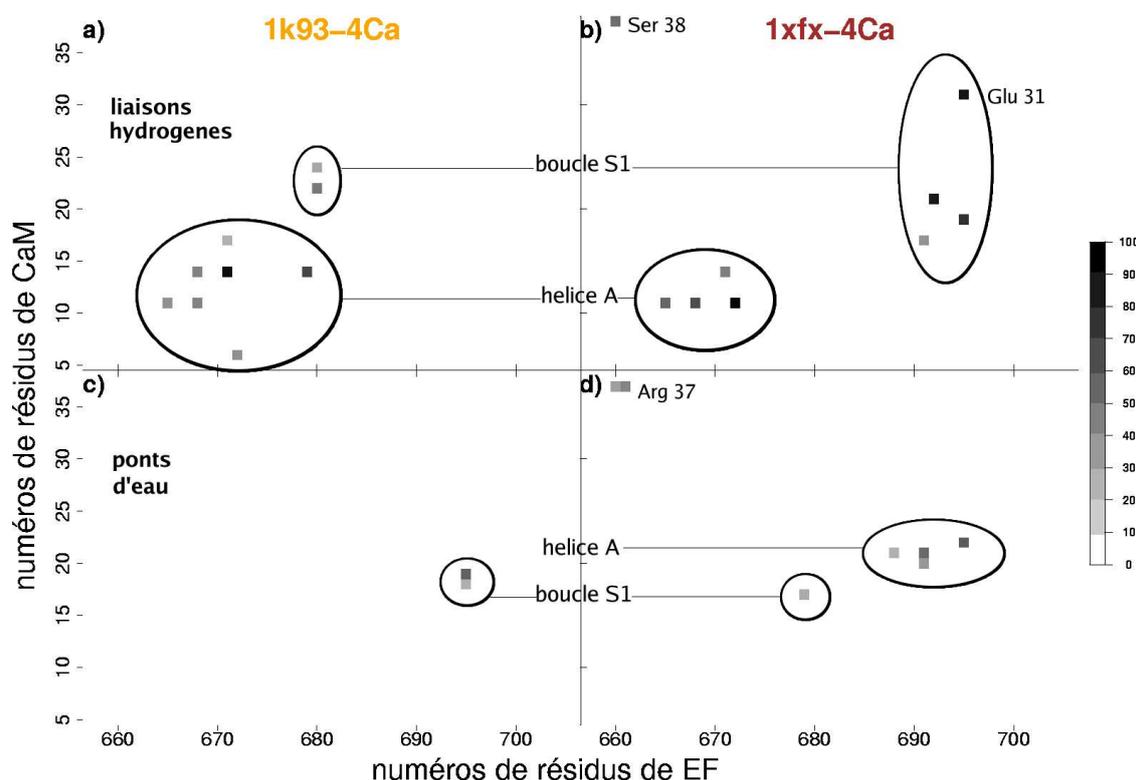
Comme décrit dans le Chapitre 1 d'Introduction, la boucle HL-HM de EF (résidus 672-695), en contact direct avec le premier motif « EF-hand » de N-CaM (résidus 5-38), adopte des conformations différentes dans 1K93 [77] et dans 1XFX, 1XFY, 1XFZ [228]. Il en résulte des réseaux d'interactions différents (liaisons hydrogènes, ponts salins et ponts d'eau) entre cette région de EF (résidus 657-706) et N-CaM, dans les structures *RX* et le long des trajectoires 1k93-4Ca et 1xfx-4Ca (Fig. 9 et Fig. 10). Glu-6, Glu-11, Glu-14 et Ser-17 de l'hélice HA de CaM établissent des contacts directs avec EF (résidus 665-691) dans les deux trajectoires. Cependant, les contacts diffèrent au niveau de la boucle calcique de S1 (résidus 20-28) et de l'hélice HB (résidus 29-38).



**FIG. 9: Boucle HL-HM du domaine Hel de EF (résidus 675-684) en interaction avec N-CaM.** Représentation « cartoon » de la boucle HL-HM de EF (en gris) et de N-CaM (en rose) dans les conformations obtenues au bout de 8 ns des simulations (c) 1k93-4Ca et (d) 1xfx-4Ca. Les hélices HL et HM de EF sont marquées en noir, les hélices HA, HB, HC et S1, S2 de N-CaM sont marquées en violet. Les résidus Ser-660, Lys-680, Lys-691, Glu-692 et Lys-695 de EF (noms en gris) et les résidus Leu-18, Phe-19, Asp-20, Lys-21, Asp-22, Asp-24, Glu-31, Arg-37 et Ser-38 de CaM (noms en magenta) sont mis en avant avec la représentation « licorice », et les ions calcium sont représentés comme des sphères vertes. La visualisation a été réalisée avec Pymol [72].

En effet, la forme protubérante de la boucle HL-HM de EF (résidus 675-684) et le déroulement partiel de l'hélice HM (résidus 685-695) qui en résulte, affectent la position et l'orientation des chaînes latérales des résidus de cette région (Fig. 9 *c-d*). Par exemple, Lys-680 forme des liaisons hydrogènes avec Asp-22 (42% du temps de simulation) et Asp-24 (19%) dans 1k93-4Ca, tandis qu'il n'est pas orienté vers CaM dans 1xfx-4Ca. De manière similaire, Lys-695 établit des ponts d'eau avec Leu-18 (15%) et Phe-19 (52%) dans 1k93-4Ca (Fig. 10 *c*), alors qu'il interagit à travers des liaisons hydrogènes avec Phe-19 (76%) et Glu-31 (89%) (Fig. 10 *b*) et via un pont d'eau avec Asp-22 (44%) dans 1xfx-4Ca (Fig. 10 *d*). De plus, dans 1xfx-4Ca, Lys-691 et Glu-692 établissent des ponts d'eau avec Asp-20 (27%) et Lys-21 (50%) et une liaison hydrogène avec Lys-21 (88%) (Fig. 10 *b* et 10 *d*).

De telles interactions ont un impact sur la conformation de la boucle de S1 et sur l'orientation globale du premier motif « EF-hand » de CaM par rapport à EF. Les résidus Asp-22 et Asp-24 de CaM interagissent avec Lys-680 de EF dans 1k93-4Ca, tandis que les résidus Asp-20, Lys-21, Asp-22 et Glu-31 sont en contact avec Lys-691, Glu-692 et Lys-695 dans 1xfx-4Ca. En conséquence, l'hélice sortante HB de S1 est positionnée plus près des hélices HL et HM de EF dans 1xfx-4Ca que dans 1k93-4Ca (Fig. 9 *c* et 9 *d*), ce qui favorise la formation de liaisons hydrogènes et de ponts d'eau entre les résidus Arg-37, Ser-38 de CaM et Ser-660, Ala-661 de EF (Fig. 10 *b* et 10 *d*). De cette façon, HB de CaM est attachée à EF par ses deux extrémités N-terminale (Glu-31) et C-terminale (Arg-37, Ser-38) dans 1xfx-4Ca, tandis qu'elle n'est pas contrainte par EF dans 1k93-4Ca.



**FIG. 10: Réseau d'interaction entre le domaine Hel de EF (résidus 657-706) et N-CaM (résidus 5-38) :** (a,b) liaisons hydrogènes et (c,d) ponts d'eau, formés le long des trajectoires (a,c) 1k93-4Ca et (b,d) 1xfx-4Ca sont matérialisés par des carrés. L'échelle des couleurs pour le pourcentage de temps de présence va de gris clair à noir. Seules les interactions présentes plus de 15% du temps de simulation sont reportées. Les étiquettes sur l'axe vertical central indiquent la localisation de l'hélice A et de la boucle calcique de S1 dans la séquence de N-CaM. Les cercles associés mettent en évidence les interactions entre ces régions de N-CaM et EF.

Globalement, le premier motif « EF-hand » de CaM est plus fermement attaché à EF dans 1xfx-4Ca que dans 1k93-4Ca : dans 1xfx-4Ca d'une part, le réseau d'interactions avec EF s'étend tout le long de la séquence du motif « EF-hand » (résidus 6, 11, 14, 17, 19, 20, 21, 22, 31, 37 et 38) ; dans 1k93-4Ca d'autre part, des interactions sont établies seulement à travers HA (résidus 6, 11, 14, 19) et à travers la partie N-terminale de la boucle calcique (résidus 22 et 24). Pour les deux configurations étudiées, l'existence d'un tel réseau, composé de résidus cruciaux pour la fixation du calcium, suggère une explication atomique pour l'affinité réduite de N-CaM liée à EF pour le calcium.

Afin de mieux appréhender le processus de dissociation du calcium de N-CaM lié à EF, des simulations *LES* du complexe EF-CaM ont été réalisées, en partant des structures équilibrées de 1k93-4Ca et 1xfx-4Ca (Table VIII). Les ions  $\text{Ca}^{2+}$  liés à S1 (LES-1k93-(1), LES-1xfx-(1)) ou S2 (LES-1k93-(2), LES-1xfx-(2)) ont été répliqués indépendamment 1, 2, 4, 8, 16, 18, 20, 22, 24, 26, 28, 30 ou 32 fois pour augmenter les probabilités de dissociation. Pour les deux systèmes 1k93-4Ca et 1xfx-4Ca, les premiers stages de la dissociation sont observés à partir de 16 copies pour S1 et 20 copies pour S2, indiquant que les ions sont plus fermement attachés à S2 (Table VIII). 18% des ions sont complètement dissociés avec 22 copies dans LES-1k93-(1), alors que des événements

de dissociation complète n'apparaissent qu'à partir de 26 copies dans LES-1k93-(2). Même au delà de 26 copies, les probabilités de dissociation sont plus grandes dans LES-1k93-(1) que dans LES-1k93-(2). Ce meilleur attachement du  $\text{Ca}^{2+}$  dans S2 par rapport à S1 dans 1k93-4Ca est cohérent avec sa meilleure coordination (cf paragraphe 3.3).

Le calcium se dissocie également plus facilement dans LES-1xfx-(1) (24 copies) que dans LES-1xfx-(2) (26 copies). Cependant, les probabilités de dissociation sont similaires, atteignant 100% à la fin des deux simulations à 32 copies. Les probabilités globales de dissociation suggèrent que l'affinité de S2 pour le calcium est diminuée dans 1xfx-4Ca comparé à 1k93-4Ca, tandis que l'affinité de S1 pour le calcium est légèrement augmentée. Par conséquent, le rapport des probabilité de dissociation entre les deux motifs « EF-hand » de N-CaM s'accordent avec les nombres de coordination calculés dans la section précédente 3.3 (Fig. 8a).

Nombres de copies	LES-1k93-(1)	LES-1k93-(2)	LES-1xfx-(1)	LES-1xfx-(2)
1	N	N	N	N
2	N	N	N	N
4	N	N	N	N
8	N	N	N	N
16	S	N	S	N
18	S	N	S	N
20	S	S	S	S
22	4D	S	S	S
24	7D	S	6D	S
26	17D	1D	16D	10D
28	17D	4D	9D	6D
30	29D	8D	22D	29D
32	32D	19D	32D	32D

**TAB. VIII: Simulations *LES* de la dissociation du calcium dans le complexe EF-CaM.** LES-1k93-(1) et LES-1k93-(2) sont les simulations de la dissociation du calcium des sites S1 et S2 de 1k93-4Ca respectivement. LES-1xfx-(1) et LES-1xfx-(2) sont les simulations de la dissociation du calcium des sites S1 et S2 de 1xfx-4Ca respectivement. Le cation est dupliqué en 1, 2, 4, 8, 16, 18, 20, 22, 24, 26, 28, 30 ou 32 copies et les événements de dissociation sont observés pendant 200 ps : (N) pas de dissociation, (S) dissociation amorcée – certaines copies peuvent même sortir et revenir, (D) nombre de copies dissociées à la fin de la simulation.

Les simulations *LES*, aussi bien que l'analyse du réseau d'interaction EF/N-CaM et de la coordination des ions  $\text{Ca}^{2+}$ , permettent d'expliquer la perte d'affinité de N-CaM pour le calcium dans le complexe. Du fait de la contrainte appliquée par EF sur HA de N-CaM, à travers un réseau de contacts électrostatiques, S1 atteint seulement la première étape du processus de fixation du calcium, sans le concours du glutamate en position 12. En plus de cette action directe sur S1, EF parvient à moduler l'affinité de S2 pour le calcium, de manière indirecte. Dans 1k93-4Ca, la plus grande plasticité de S2 autorise un meilleur attachement du calcium dans ce site. Dans 1xfx-4Ca, l'établissement d'un réseau d'interactions plus dense entre S1 et EF permet de bloquer les mouvements de HB, attachée à EF par ses deux extrémités. La tension introduite par la présence d'un ion  $\text{Ca}^{2+}$  dans S2 ne peut être relaxée à travers S1, rompant la coopérativité entre les deux sites du lobe, et générant une instabilité dans S2.

## 4 Conclusion

Dans ce premier chapitre de résultats, nous avons focalisé notre attention sur le comportement dynamique des partenaires EF et CaM, et nous avons évalué l'impact de l'ajout ou du retrait d'ions calcium sur la stabilité du complexe. Des simulations de 15 ns de EF-CaM, correspondant à différents niveaux de calcium, ont été réalisées et ont pu être validées par une analyse de convergence et par des comparaisons avec les données expérimentales. Ces simulations ont été effectuées avec des conditions plus réalistes (limites périodiques, *PME*) et dans un temps largement plus long que de précédentes simulations de 350 ps de EF-(2Ca-CaM) [289].

L'analyse des mouvements et des caractéristiques dynamiques de CaM, et de EF, a clairement mis en évidence la stabilité de la forme du complexe EF-(2Ca-CaM). La conformation adoptée par CaM dans cette forme est en effet la mieux adaptée à EF. Un modèle de l'interaction EF-CaM peut ainsi être proposé, selon lequel CaM joue le rôle d'un ressort, qui maintient le site catalytique de EF en conformation ouverte active. L'interaction avec EF requiert que C-CaM soit en conformation ouverte, chargé de calcium, et bloque N-CaM en conformation fermée. La fixation de calcium sur ce lobe reste possible, mais elle demeure incomplète, N-CaM étant piégé dans une conformation intermédiaire.

Les descriptions détaillées du réseau de contacts électrostatiques entre EF et N-CaM et de la coordination du calcium dans les boucles calciques ont permis d'expliquer, à l'échelle atomique, l'affinité réduite de N-CaM pour le calcium dans EF-CaM. La comparaison de deux simulations du complexe EF-(4Ca-CaM), l'une à partir d'une structure cristallographique perturbée, et l'autre à partir d'une structure existante déjà chargée de quatre ions  $\text{Ca}^{2+}$ , a révélé une influence mutuelle entre les interactions EF-CaM et  $\text{Ca}^{2+}$ -CaM. Des simulations *LES*, bien que réalisées en solvant implicite, ont confirmé qualitativement une modulation par EF de l'affinité de N-CaM pour le calcium, à travers l'application d'une contrainte mécanique.

La présence de EF n'affecte pas seulement l'affinité de CaM pour le calcium, mais également la coopérativité entre les deux lobes à travers le linker central. En effet, l'ajout d'ions  $\text{Ca}^{2+}$  sur N-CaM provoque des fluctuations dans N-CaM lui-même et dans le linker central. C-CaM, quant à lui, est profondément inséré entre le coeur catalytique et le domaine hélicoïdal de EF, qui lui laisse très peu de mobilité. Ainsi, la tension introduite par le retrait de calcium de ce lobe ne peut être relaxée et provoque des changements dans les fluctuations de N-CaM et de EF, notamment au niveau du coeur catalytique.

Au delà de la caractérisation de l'interaction EF-CaM qu'il permet, ce travail apporte une contribution à la connaissance globale de la remarquable flexibilité de CaM et à la compréhension de son interaction avec le calcium. Une approche plus globale de la structure et de la dynamique du complexe EF-CaM est cependant nécessaire pour appréhender les effets du calcium à travers son réseau de résidus.



# Aspects Dynamiques et Énergétiques du Complexe EF-CaM

## 1 Introduction

La fonction des protéines et des complexes protéiques est modulée par des changements conformationnels et des mouvements concertés, qui résultent de la transmission d'information à travers leur réseau de résidus [54]. Diverses méthodes servent à explorer la dynamique collective des protéines, pour décrire l'importance de cette dynamique pour la fonction biologique. L'analyse en composantes principales [13; 10] permet d'identifier les modes essentiels d'un système, de plus grande amplitude et impliquant le plus grand nombre d'atomes, à partir d'une simulation *DM* par exemple [238; 234; 153]. Une alternative aux simulations est l'analyse en modes normaux [37; 100], qui se propose d'identifier les mouvements lents intrinsèques des protéines, encodés dans leur structure native. L'utilisation d'un potentiel harmonique simple et de la seule topologie de la structure native, dans le modèle de réseau élastique [244; 23], a permis de prédire avec une précision étonnante le comportement global de très larges systèmes [54; 131]. Si ces méthodes sont largement utilisées pour décrire les mouvements lents ou collectifs des complexes protéiques, la relation entre corrélations dynamiques et influences énergétiques au sein des ces systèmes a cependant été peu étudiée.

Le complexe EF-CaM se prête particulièrement bien à une telle étude, dans la mesure où le niveau de calcium affecte la stabilité du complexe, sans pour autant modifier son architecture globale (Cf. Chapitre 1 d'Introduction, section 2.3). De plus, on a vu, au Chapitre 3, que la perturbation du niveau de calcium induit des changements conformationnels locaux (de N-CaM et du site catalytique de EF) et modifie la dynamique essentielle des partenaires. Qu'en est-il de l'effet électrostatique des ions  $\text{Ca}^{2+}$  sur l'ensemble du réseau de résidus du complexe ? La mise en évidence de cet effet permet-elle d'inférer un modèle de la propagation du signal calcium à travers le réseau ?

Dans ce deuxième volet de notre étude, nous proposons de sonder cette propagation par une analyse double dynamique et énergétique. Nous avons choisi d'utiliser plusieurs méthodes, certaines

récemment développées, pour appréhender les aspects dynamiques et énergétiques de EF-CaM, sous différents angles et à différentes échelles. Le modèle de réseau élastique a d'abord été appliqué pour identifier les mouvements intrinsèques du complexe. Comme attendu, ce modèle ne permet pas de caractériser les changements dus au calcium. Une analyse en composantes principales des mouvements de EF-CaM a été effectuée sur les simulations *DM* décrites au chapitre précédent, et une technique statistique a été appliquée, afin d'exploiter les résultats pour décrire la dynamique locale du réseau de résidus et d'identifier des domaines dynamiques locaux. L'effet du calcium a enfin été caractérisé par des approches basées sur les corrélations dynamiques d'une part et sur les énergies libres de liaison d'autre part. L'introduction de nouveaux concepts pour décrire l'architecture du complexe donne une vision consensuelle des résultats des deux méthodes. Une interprétation commune révèle que le retrait de calcium rompt le couplage entre la région catalytique et l'interface EF-CaM, alors que l'ajout de calcium déstabilise EF au sein du complexe.

## 2 Matériels et Méthodes

Les détails concernant la préparation des structures initiales et la production des trajectoires *DM* sont données dans le Chapitre 3.

### 2.1 Modèle de réseau élastique

Dans la théorie du réseau élastique (*elastic network model* ou *ENM*) [244; 23], la structure du complexe est modélisée par un réseau de  $N = 619$  noeuds, chacun représentant un  $C(\alpha)$  donné. Deux noeuds  $i$  et  $j$  sont liés par un ressort de constante  $\gamma_*$ , si la distance qui les sépare est inférieure à un certain rayon de coupure  $r_c$ . Le modèle peut être exprimé par une matrice de Kirchhoff  $\Gamma$ , définie comme suit [23] :

$$\Gamma_{ij} = \begin{cases} -\gamma_* & \text{si } i \neq j \text{ et } R_{ij} \leq r_c, \\ 0 & \text{si } i \neq j \text{ et } R_{ij} > r_c, \\ -\sum_{i \neq j} \Gamma_{ij} & \text{si } i = j. \end{cases} \quad (1)$$

Les facteurs  $B$  expérimentaux et les fluctuations *DM* ont été prédits par la méthode *ENM*, selon les équations suivantes [23] :

$$\langle \Delta R_k^2 \rangle = [\Gamma^{-1}]_{kk} \quad (2)$$

$$B_k = \frac{8\pi^2}{3} \langle \Delta R_k^2 \rangle \quad (3)$$

où  $\langle \Delta R_k^2 \rangle$  et  $B_k$  sont la fluctuation moyenne au carré et le facteur de température du résidu  $k$ . L'inverse de la matrice de Kirchhoff  $\Gamma^{-1}$  est la matrice de covariance prédite. En normalisant

$\Gamma^{-1}$ , on obtient une matrice de corrélation prédite  $C$  :

$$c_{ij} = \frac{[\Gamma^{-1}]_{ij}}{(\sqrt{[\Gamma^{-1}]_{ii}}\sqrt{[\Gamma^{-1}]_{jj}})} \quad (4)$$

Les structures cristallographiques du complexe EF-CaM, 1K93 [77], 1XFX, 1XFY et 1XFZ [228] – avec respectivement 2, 3 et 4  $\text{Ca}^{2+}$ , ont été processées par *ENM* pour prédire les facteurs  $B$  expérimentaux. Les fluctuations *DM* ont été d’une part calculées sur les 12 dernières ns de chaque simulation, et d’autre part prédites par *ENM*, en utilisant des conformations représentatives de chaque trajectoire. Ces conformations représentatives sont les structures de référence identifiées avec la méthode d’analyse de convergence de Lyman et al. [174] (Cf. Chapitre 3, section 3.1), et correspondent aux conformations *DM* prises à : 5906 ps, 9335 ps et 12150 ps de 1k93-0Ca, 5016 ps, 7950 ps et 12262 ps de 1k93-2Ca, 3529 ps, 4983 ps, 5471 ps, 11084 ps et 13515 ps de 1k93-4Ca. Un rayon de coupure  $r_c$  de 9 Å et une constante de raideur  $\gamma_\star = 0.05$  permettent d’obtenir le meilleur ajustement des valeurs prédites sur les facteurs  $B$  expérimentaux.

## 2.2 Matrices de corrélations

Les fluctuations atomiques, les matrices de corrélations de Pearson et les matrices de corrélations généralisées ont été évaluées pour les  $N = 619$   $C(\alpha)$  du complexe EF-CaM, en considérant les 12 dernières ns de chaque simulation *DM*. Les fluctuations atomiques ( $N$ ) et les matrices de corrélations de Pearson ( $N \times N$ ) ont été calculées avec le module PTRAJ d’AMBER 9 [43]. Les corrélations ont été quantifiées par le coefficient de Pearson  $r$  :

$$r[\vec{x}_i, \vec{x}_j] = \frac{\langle \vec{x}_i \cdot \vec{x}_j \rangle}{\sqrt{(\langle \vec{x}_i \rangle^2 \langle \vec{x}_j \rangle^2)}} \quad (5)$$

où  $\vec{x}_i$  et  $\vec{x}_j$  sont les vecteurs de fluctuation des positions des résidus  $i$  et  $j$ .

Les corrélations généralisées (*generalized correlations* ou *GC*), récemment développées par Lange & Grubmüller [162] et basées sur une métrique différente, ont été calculées avec l’outil `g_correlation`, intégré au logiciel GROMACS [257]. Par analogie avec le coefficient de Pearson, le coefficient de corrélation généralisée  $r_{MI}$  est défini comme suit [162] :

$$r_{MI}[\vec{x}_i, \vec{x}_j] = (1 - \exp -2I[\vec{x}_i, \vec{x}_j]/3)^{-1/2} \quad (6)$$

où  $I[\vec{x}_i, \vec{x}_j]$  est l’information mutuelle entre les résidus  $i$  et  $j$  [162] :

$$I[\vec{x}_i, \vec{x}_j] = H[\vec{x}_i] + H[\vec{x}_j] - H[\vec{x}_i, \vec{x}_j], \quad (7)$$

estimée à partir de la mesure du contenu d’information  $H[\vec{x}] = -\int p(\vec{x}) \ln p(\vec{x}) d\vec{x}$ , avec  $p(\vec{x})$  la distribution de probabilité du vecteur de fluctuation  $\vec{x}$ .

	1k93-0Ca	1k93-2Ca	1k93-4Ca
$h \geq 5.17$	$0.44 \pm 0.10$ (2)	$0.45 \pm 0.10$ (2)	$0.46 \pm 0.10$ (1)
	$0.60 \pm 0.08$ (1)	$0.60 \pm 0.07$ (2)	$0.60 \pm 0.08$ (2)
$5 \geq h < 5.17$	$0.44 \pm 0.10$ (2)	$0.45 \pm 0.10$ (2)	$0.46 \pm 0.10$ (1)
	$0.60 \pm 0.08$ (1)	$0.60 \pm 0.07$ (1)	$0.56 \pm 0.08$ (3) $0.71 \pm 0.07$ (2)
$h < 5$	$0.46 \pm 0.10$ (3)	$0.46 \pm 0.10$ (3)	$0.46 \pm 0.10$ (1)
	$0.46 \pm 0.16$ (2)	$0.47 \pm 0.22$ (2)	$0.56 \pm 0.08$ (3)
	$0.60 \pm 0.08$ (1)	$0.60 \pm 0.07$ (1)	$0.71 \pm 0.07$ (2)

**TAB. I: Classification hiérarchique.** Sont reportées les corrélations moyennes des groupes de résidus, identifiés à partir des matrices de corrélations généralisées de 1k93-0Ca, 1k93-2Ca et 1k93-4Ca, pour différentes hauteurs de coupure  $h$ . Avec  $h_{opt} = 5$ , deux, deux et trois groupes sont respectivement identifiés, qui présentent des valeurs de corrélations moyennes significativement différentes. Les numéros des groupes sont indiqués entre parenthèses.

Les résidus ont été classés à partir de leurs corrélations généralisées, en utilisant le programme R [204]. Des matrices de dissimilarité ont d'abord été déterminées en calculant les distances euclidiennes entre les lignes de chaque matrice de corrélations. Une classification hiérarchique a ensuite été appliquée sur les matrices de dissimilarité, avec la méthode « complete linkage » [101]. Les arbres ainsi générés ont été coupés à une certaine hauteur  $h$  pour identifier des groupes. La valeur optimale  $h_{opt} = 5$  permet d'obtenir un nombre de groupes minimal et suffisant pour couvrir toute la gamme de corrélations généralisées (Table I). La même classification a été appliquée sur les matrices de corrélations de Pearson, mais aucune hauteur de coupure n'a permis d'obtenir des groupes avec des corrélations moyennes significativement différentes.

### 2.3 Analyse locale des traits caractéristiques

L'analyse locale des traits caractéristiques (*local feature analysis* ou *LFA*), adaptée par Zhang & Wrighers [288] à l'étude des protéines, permet d'identifier des groupes de résidus, dont les fluctuations décrivent les traits caractéristiques les plus marquants de la dynamique locale. Une analyse en composantes principales (*ACP*) des mouvements des  $C(\alpha)$  du complexe dans chaque trajectoire  $DM$  a d'abord été réalisée, avec le module PTRAJ d'AMBER 9 [43]. Parmi les  $3N$  vecteurs propres  $\Psi_r$ , associés aux valeurs propres  $\lambda_r$ ,  $r = 1, 2, \dots, 3N$ , les quinze premiers se sont révélés suffisants pour expliquer 81 à 84 % des fluctuations totales. C'est pourquoi ces vecteurs ont été retenus pour appliquer le formalisme *LFA* [288]. En bref, ce formalisme opère une transformation de la matrice de corrélations, de telle sorte que les corrélations hors-diagonale sont réduites, et permet ainsi d'identifier quinze degrés de liberté « graines », représentant quinze résidus « graines ».

A partir des  $n = 15$  modes *ACP* globaux, on peut définir  $3N$  fonctions de sortie *LFA*  $O(i)$  [288] :

$$O(i) = \sum_{r=1}^n \frac{A_r}{\sqrt{\lambda_r}} \Psi_r(i) \quad (8)$$

où  $A_r$  est la projection des fluctuations atomiques  $\Delta x_i$  sur le vecteur propre  $\Psi_r$  :  $A_r = \sum_{i=1}^{3N} \Psi_r(i) \Delta x_i$ .

Les corrélations entre les fonctions de sortie *LFA* sont résiduelles, données par [288] :

$$\langle O(i)O(j) \rangle = \sum_{r=1}^n \Psi_r(i)\Psi_r(j) \equiv P(i, j), \quad (9)$$

de telle sorte qu'à la limite  $n \rightarrow 3N$ , les sorties  $O(i)$  sont complètement décorréelées dans l'espace :  $P(i, j) \rightarrow \delta(i, j)$ .

L'algorithme de « clairsemage » décrit par Zhang & Wriggers [288] propose d'approximer l'ensemble des  $3N$  fonctions de sortie  $O(i)$  par seulement quinze fonctions de sortie  $O(i_m)_{i_m \in \mathcal{M}}$ , reconstruits à partir d'un ensemble  $\mathcal{M}$  de quinze degrés de liberté. A chaque itération de l'algorithme, appliqué avec le programme R [204], les  $O(i)$  ont été reconstruits en utilisant l'ensemble courant  $\mathcal{M}^{(m)}$  de  $m$  degrés de liberté, et les erreurs de reconstruction  $O^{err}(i)$  ont été évaluées. Parmi les  $3N$  degrés de liberté, l'indice présentant la plus grande erreur de reconstruction  $O^{err}(i_{m+1})$  a été choisi comme le  $(m + 1)$ ème indice graine dans  $\mathcal{M}$ , à la condition que le  $C(\alpha)$  auquel correspond cet indice n'ait pas déjà été choisi. Les indices graines sont ajoutés à  $\mathcal{M}$  jusqu'à un nombre de quinze  $C(\alpha)$  graines. Les résidus graines ainsi identifiés servent à construire une représentation topographique locale du complexe EF-CaM, où chaque résidu graine et ses résidus voisins corrélés définissent un domaine dynamique local.

La corrélation résiduelle  $P(h, k)$  entre le résidu graine  $h$  et n'importe quel autre résidu  $k$  a été évalué comme suit [288] :

$$P(h, k) = \sum_{d=1}^3 \sum_{r=1}^{n=15} \Psi_r(h_d)\Psi_r(k_d), \quad (10)$$

où  $d$  est l'indice de coordonnée (x,y,z) et  $\Psi_r$  est le vecteur propre *ACP* associé à la valeur propre  $\lambda_r$ .

## 2.4 Cartes de dépendances énergétiques

L'algorithme décrit par Hamacher et al. dans une étude de l'assemblage de la sous-unité ribosomale [113] a été utilisé pour identifier les influences énergétiques au sein du complexe EF-CaM. L'algorithme est basé sur l'affirmation suivante : '*Si le retrait de la protéine Z change l'ordre relatif de X-plus-fortement-lié-que-Y à Y-plus-fortement-lié-que-X, alors l'ordre d'assemblage probabiliste est également changé*' [113]. Dans le cas de EF-CaM, on considère des pseudo-domaines au lieu de protéines. Ainsi, la méthode des cartes de dépendances énergétiques (*energetic dependency maps* ou *EDM*) permet d'inférer un schéma de dépendance de stabilité de pseudo-domaines, plutôt qu'un ordre d'assemblage probabiliste de protéines.

Une nouvelle définition des domaines ou pseudo-domaines du complexe EF-CaM a été dérivée des résultats *LFA* (Table V). L'énergie libre de liaison d'un pseudo-domaine  $D$  au complexe  $C$  a été évaluée sur des conformations extraites toutes les 10 ps des 12 dernières ns de la trajectoire

$DM$  de  $C$ . L'énergie libre de liaison  $\Delta\Delta G[D, C]$  est exprimée comme :

$$\Delta\Delta G[D, C] = \Delta G[C] - (\Delta G[C \setminus D] + \Delta G[D]), \quad (11)$$

où  $C \setminus D$  symbolise le complexe  $C$ , duquel le domaine  $D$  a été retiré. Dans chaque complexe, les énergies libres de liaison de chaque pseudo-domaine ont été calculées et rangées, de façon à obtenir un classement de référence, avec le rang 1 correspondant à la meilleure énergie.

Dans le but de déterminer l'influence d'un pseudo-domaine donné  $D^*$  sur les autres pseudo-domaines,  $D^*$  a été retiré du complexe  $C$  et les énergies libres de liaison de chacun des pseudo-domaines restants  $D$ , au complexe tronqué  $C \setminus D^*$ , ont été calculées comme suit :

$$\Delta\Delta G[D, C \setminus D^*] = \Delta G[C \setminus D^*] - (\Delta G[C \setminus (D + D^*)] + \Delta G[D]) \quad (12)$$

Ces énergies libres de liaison ont ensuite été rangées et le classement obtenu a été comparé au classement de référence. Lorsque le rang d'un pseudo-domaine  $D$  sur  $C \setminus D^*$  s'est trouvé moins bon que son rang de référence, une influence énergétique a été attribuée de  $D^*$  sur  $D$ . L'intensité de l'influence est équivalente au nombre de rangs perdus par  $D$ . Un total de  $7 + (6 * 7) = 49$  énergies libres de liaison ont été calculées et les influences énergétiques ont été matérialisées sur des cartes de dépendances.

Les énergies libres de liaison ont été évaluées avec la méthode *MMPBSA* [236], décrite dans le Chapitre 2, section 1.3. L'espacement de la grille a été fixé à 0.5 Å et des constantes diélectriques de 1 et 80, pour l'intérieur et l'extérieur du soluté, ont été choisies. La limite diélectrique a été définie en utilisant une sphère d'eau de rayon 1.6 Å sur la surface atomique, avec les rayons optimisés par Tan & Luo, basés sur les types d'atomes et les charges d'AMBER. La contribution non polaire  $\Delta G_{sol}^{np}$  a été calculée comme la somme d'un terme répulsif  $\Delta G_{cavite}$ , proportionnel à la *SASA*, et un terme attractif  $\Delta G_{dispersion}$ , calculé par une approche d'intégration de surface (Tan & Luo, en préparation, cité dans [43]) :  $\Delta G_{sol}^{np} = \Delta G_{dispersion} + \Delta G_{cavite} = \Delta G_{dispersion} + 0.04356 * SASA - 1.008$ . Du fait de variances significatives dans le calcul de l'entropie du soluté  $S_{solute}$  avec la méthode *MMPBSA* [48; 236], ce terme n'a pas été inclus dans la valeur de l'énergie libre de liaison totale.

Les valeurs moyennes de la plupart des énergies libres de liaison se sont trouvées très proches les unes de autres, de sorte que les déviations standards ne permettaient pas de les ranger correctement. Des tests non-paramétriques de Kolmogorov-Smirnov [179] ont donc été effectués sur les séries temporelles des énergies, avec le programme R [204]. L'énergie  $\Delta\Delta G_1$  était considérée comme effectivement plus grande que  $\Delta\Delta G_2$  lorsque le test de Kolmogorov-Smirnov donnait une  $p$ -value inférieure ou égale à 0.05.

## 3 Résultats

### 3.1 Modèle de réseau élastique

Le modèle de réseau élastique (*ENM*) [244; 23] a été appliqué aux structures cristallographiques du complexe EF-CaM [228; 77]. Les facteurs  $B$  expérimentaux ont ainsi pu être prédits avec des coefficients de régression  $R^2$  supérieurs à 0.95 (Table II). Les fluctuations atomiques et les matrices de corrélations ont également été prédites pour un ensemble de conformations représentatives, extraites des trajectoires 1k93-0Ca (à 5906, 9335 et 12150 ps), 1k93-2Ca (à 5016, 7950 et 12262 ps) et 1k93-4Ca (à 3529, 4983, 5471, 11084 et 13515 ps), identifiées par l'analyse de convergence au Chapitre 3, section 3.1. Les fluctuations prédites par la méthode *ENM* ont pu être ajustées sur les fluctuations *DM* avec des coefficients  $R^2$  entre 0.90 et 0.93 (Table II). On peut noter que les matrices de corrélations *ENM*, calculées à partir des structures *RX* ou des conformations *DM*, sont très similaires (Table IV). De ce fait, et malgré des coefficients  $R^2$  élevés, la méthode *ENM* est incapable de révéler des changements dans la dynamique interne, entre les différents niveaux de calcium.

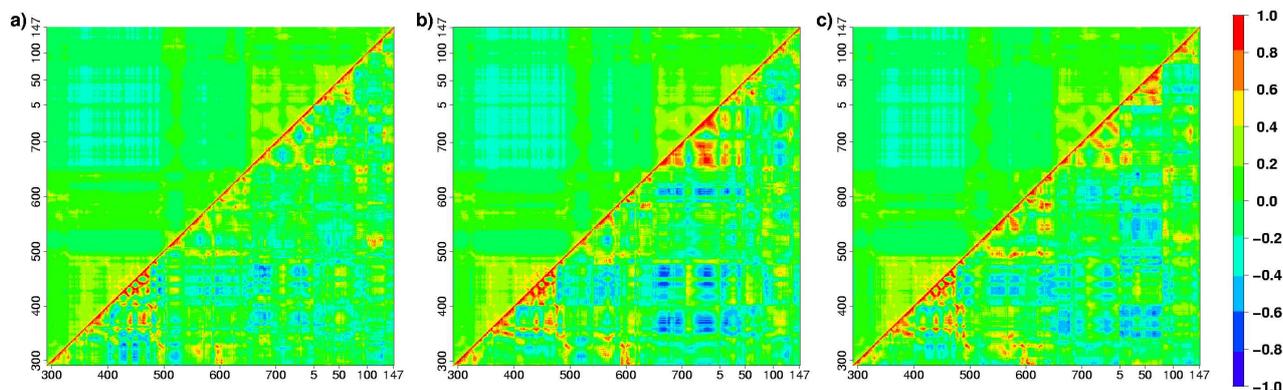
		$a$	$R^2$
Structure <i>RX</i>	1K93 (2Ca)	1.06	0.97
Structure <i>RX</i>	1XFY (2Ca)	0.97	0.98
Structure <i>RX</i>	1XFZ (3Ca)	0.99	0.98
Structure <i>RX</i>	1AFX (4Ca)	0.99	0.99
1k93-0Ca (DM)	5906 ps	0.95	0.90
	9335 ps	0.95	0.90
	12150 ps	0.94	0.90
1k93-2Ca (DM)	5016 ps	0.98	0.93
	7950 ps	0.97	0.93
	12262 ps	0.98	0.92
1k93-4Ca (DM)	3529 ps	1.00	0.91
	4983 ps	0.99	0.92
	5471 ps	0.99	0.92
	11084 ps	1.01	0.91
	13515 ps	1.00	0.91

**TAB. II: Coefficients de pente  $a$  et de régression  $R^2$  pour l'ajustement linéaire des fluctuations prédites par *ENM*.** Les fluctuations sont prédites à partir de représentations *ENM* de conformations représentatives extraites des simulations *DM* 1k93-0Ca, 1k93-2Ca et 1k93-4Ca. L'ajustement est effectué sur les facteurs  $B$  expérimentaux des structures *RX* 1K93 [77] et 1AFX, 1XFY, 1XFZ [228], et sur les fluctuations calculées le long des trajectoires *DM* 1k93-0Ca, 1k93-2Ca et 1k93-4Ca. Les positions des conformations le long des simulations *DM* sont données en picosecondes.

Les matrices de corrélations *ENM* ont été comparées aux matrices de corrélations de Pearson calculées le long des simulations *DM* 1k93-0Ca, 1k93-2Ca et 1k93-4Ca (Fig. 1). Les corrélations *ENM* sont globalement plus petites que les corrélations de Pearson : 98 % des valeurs *ENM* sont comprises entre -0.4 et 0.4, alors que les corrélations de Pearson balaient l'intervalle entre -1 et

+1. De plus, les motifs de corrélation de Pearson sont plus fins que ceux prédits par la méthode *ENM* (Fig. 1), illustrant le fait que cette dernière décrit des mouvements collectifs, provenant principalement de la forme moléculaire intrinsèque du complexe.

Deux blocs de résidus anti-corrélés sont mis en évidence dans les matrices *ENM* (Fig. 1, *en haut à gauche*) : (i) la région catalytique, formée par les domaines CA et CB (résidus 292-501), (ii) la région d'interface entre les deux partenaires, formée par le domaine Hel de EF (résidus 648-767) et N-CaM (résidus 5-76). Ces résultats sont en accord avec l'analyse en composantes principales des mouvements de EF présentée dans le Chapitre 3, section 3.1, qui révélait une compaction et une torsion entre le coeur catalytique et le domaine Hel. La méthode *ENM* est donc capable de prédire des corrélations à longue distance, bien qu'elle soit basée sur des contraintes géométriques locales. Les matrices de corrélations de Pearson révèlent des corrélations positives et négatives similaires (Fig. 1, *en bas à droite*). L'anti-corrélation entre les domaines CB et Hel de EF est particulièrement visible dans 1k93-2Ca (Fig. 1b).



**FIG. 1: Comparaison des matrices *ENM* et des matrices de Pearson.** Les matrices de corrélations *ENM* prédites sont en haut à gauche et les matrices de corrélations de Pearson sont en bas à droite. Les matrices de corrélation *ENM* ont été prédites pour des conformations représentatives extraites des simulations *DM* de 1k93-0Ca, 1k93-2Ca et 1k93-4Ca, prises à : (a) 9335 ps pour 1k93-0Ca, (b) 7950 ps pour 1k93-2Ca, (c) 11084 ps pour 1k93-4Ca. Les matrices de corrélation de Pearson ont été calculées sur les trajectoires *DM* de 1k93-0Ca, 1k93-2Ca et 1k93-4Ca. Les numéros de résidus de EF (292-767) et de CaM (5-147) sont indiqués sur les axes x et y.

La méthode *ENM* permet d'extraire des mouvements globaux du complexe. Elle n'est toutefois pas suffisamment précise pour discriminer les différents niveaux de calcium. Afin de traiter cette question, d'autres méthodes ont donc été utilisées : (i) l'analyse locale des traits caractéristiques (*LFA*) [288], (ii) le coefficient de corrélation généralisée (*GC*) [162].

### 3.2 Analyse locale des traits caractéristiques

Dans le but de caractériser plus finement la dynamique locale de EF-CaM, une analyse locale des traits caractéristiques [288] (*local feature analysis* ou *LFA*) a été effectuée sur les trajectoires *DM*. Une *ACP* a été appliquée sur les mouvements des  $C(\alpha)$ , et les quinze premiers vecteurs propres ont été retenus pour la *LFA* [288]. L'algorithme *LFA* de « clairsemage » [288] a permis d'identifier

quinze résidus graines. Comme dans le papier original de Zhang & Wriggers [288], ces résidus graines sont pour la plupart localisés dans les régions flexibles de chaque complexe (Table III) : les queues N- et C-terminales de EF et de CaM, la région catalytique et les boucles SA et SC de EF, les sites de fixation au calcium et le linker central de CaM. La grande flexibilité des résidus graines Arg-292 et Gln-767 de EF et Glu-6 de CaM est associée à des mouvements indépendants, une caractéristique inhérente des parties terminales des protéines. Comme attendu, ces résidus présentent des profils de corrélations résiduelles (Eq. 2.3) plats, contenant peu d'information.

		1k93-0Ca	1k93-2Ca	1k93-4Ca
<b>Queues</b>		D292 <sup>EF</sup> , N767 <sup>EF</sup>	R292 <sup>EF</sup> , E6 <sup>CaM</sup>	E6 <sup>CaM</sup>
<b>EF</b>	<b>CB</b>	T392, K461	N376, G429, E449, L465	V380, G429, N450, E482
	<b>SA/SC</b>		T519 <sup>SA</sup> , G638 <sup>SC</sup>	T545 <sup>SA</sup>
	<b>Hel</b>	S673, D684, E687, P743, Y745	N675, S682, F688, P743	S674, D684, D686
<b>CaM</b>	<b>N-CaM</b>	D24, A57, N60	A57, G61	D24, S38, A46
	<b>linker</b>	D78		R74
	<b>C-CaM</b>	E120, M124	D129	E104, D118

**TAB. III: Résidus graines LFA.** Les résidus graines sont localisés (i) dans les queues N- et C-terminales de EF et de CaM, (ii) dans le domaine CB, les boucles SA et SC et le domaine Hel de EF, (iii) dans les domaines N- et C-terminaux et le linker central de CaM. L'appartenance des résidus dans la structure est indiquée en exposant quand c'est nécessaire.

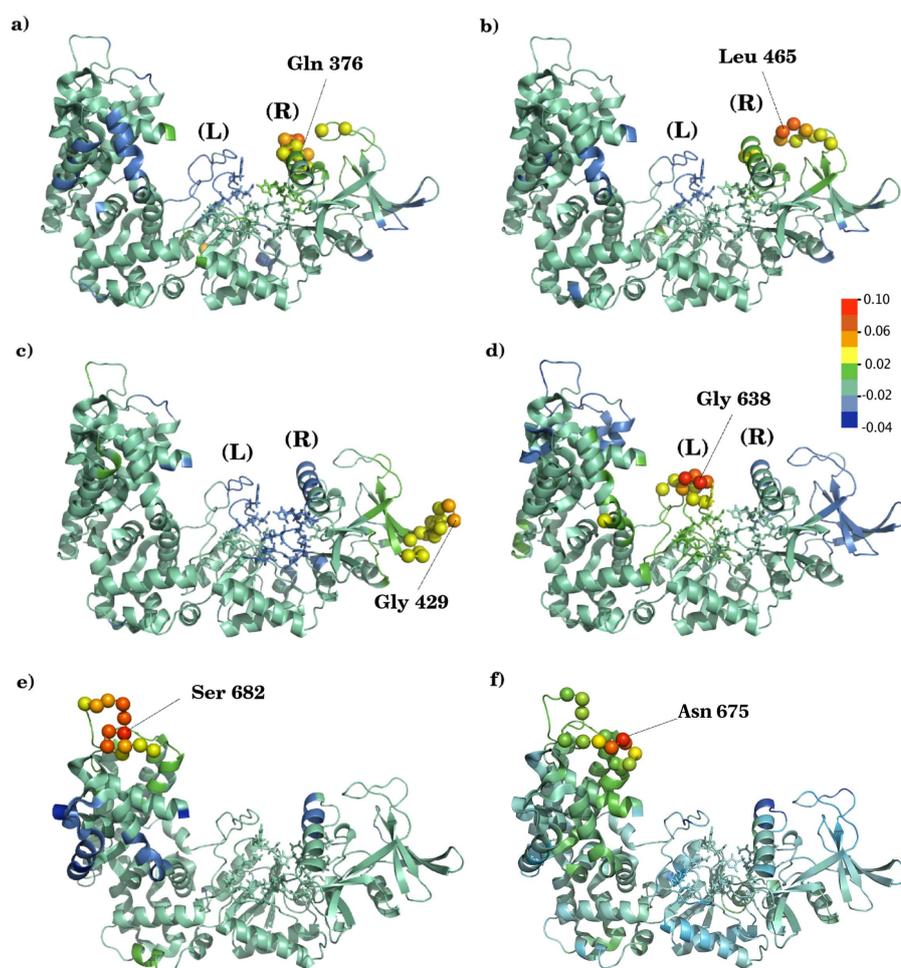
Plus intéressants, sont les résidus graines associés à un groupe de résidus localement corrélés, représentant des domaines dynamiques locaux. Dans chaque complexe, trois groupes sont localisés dans des régions importantes pour l'activité catalytique de EF, l'interaction EF-CaM ou l'interaction Ca<sup>2+</sup>-CaM. Ils sont décrits ici en détail, en se focalisant principalement sur la trajectoire 1k93-2Ca.

Les résidus graines identifiés dans EF de 1k93-2Ca, ainsi que leurs groupes de résidus associés ( $P \geq 0.03$ ), sont dépeints Fig. 2. Les résidus catalytiques sont indiqués en représentation « lico-riche » et la structure entière est colorée en fonction du profil de corrélation résiduelle du résidu graine considéré. Les cinq résidus graines Leu-465, Gln-376, Gly-429, Glu-449 (non montré dans la Fig. 2) et Gly-638, sont localisés dans la région catalytique (Table III, Fig. 2*a-d*). Comme indiqué Fig. 2*a* et 2*b*, les deux groupes formés autour de Gln-376 et Leu-465 sont sur le bord droit (R) du site catalytique et partagent les résidus 376, 379, 465, 467. Les deux résidus graines présentent des corrélations résiduelles négatives (jusqu'à -0.03, *en bleu*) avec les résidus du bord gauche (L). Réciproquement, le groupe formé autour de Gly-638 (Fig. 2*d*), qui inclut des résidus des boucles SB et SC, dans le bord gauche (L), présente des anti-corrélations résiduelles (Eq. 2.3), jusqu'à -0.03, avec les résidus du bord droit (R).

Par conséquent, la dynamique locale de la région catalytique est polarisée autour de deux spots anti-corrélés : Gly-638 de la boucle SC, représentatif du côté gauche et Gln-376 de CB, intimement lié à Leu-465, représentatif du côté droit. Cette polarisation, qui met en évidence le mouvement d'ouverture/fermeture du site, n'est pas observée dans 1k93-0Ca, où le nombre de résidus graines dans la région catalytique est réduit (Table III), en accord avec l'effondrement

observé dans la simulation *DM* au Chapitre 3.

Plus loin du site catalytique de EF, Gly-429 est anti-corrélé (jusqu'à -0.02) avec la plupart des résidus catalytiques et en périphérie proche (Fig. 2c). Glu-449 présente un profil de corrélation similaire. Cette région est également identifiable dans 1k93-4Ca, par les résidus graines Gly-429 et Asn-450 (Table III).

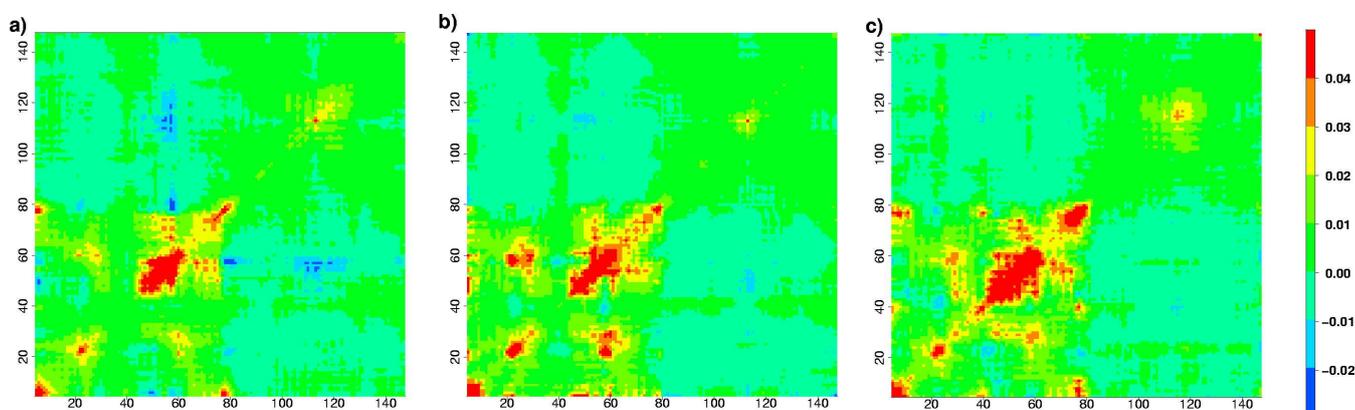


**FIG. 2: Résidus graines *LFA* de EF dans 1k93-2Ca.** Les  $C(\alpha)$  des résidus graines localisés dans la région catalytique et dans le domaine Hel de 1k93-2Ca sont représentés sous forme de sphères, ainsi que leurs groupes de résidus associés (corrélation résiduelle  $\geq 0.03$ ) : (a) Gln-376, (b) Leu-465, (c) Gly-429, (d) Gly-638, (e) Ser-682, (f) Phe-688. Les résidus catalytiques sont indiqués en représentation « licorice ». (L) et (R) marquent le côté gauche et le côté droit du site catalytique. L'échelle de couleur donne les profils de corrélations résiduelles des résidus graines considérés. La visualisation a été réalisée avec Pymol [72].

Dans 1k93-2Ca, des résidus graines ont aussi été identifiés dans des régions de EF en contact avec CaM (Table III) : (i) Thr-519 dans la boucle SA, en contact avec C-CaM, (ii) Asn-675, Ser-682, Phe-688 et Pro-743 dans la boucle HL-HM de Hel, en contact avec le premier motif « EF-hand » de N-CaM (Fig. 5). Ser-682 (Fig. 2d) et Phe-675 (Fig. 2e) présentent tous deux des corrélations résiduelles positives (jusqu'à 0.02) avec la boucle HN-HO (résidus 738-745, Fig. 5)). De plus, Ser-682 et Phe-688 (dont les profils de corrélation sont similaires à celui de Ser-682)

présentent des corrélations résiduelles négatives (jusqu'à  $-0.03$ ) avec l'hélice HC et le linker central de CaM. Il est intéressant de noter que des résidus graines ont été identifiés dans les boucles SA et SC et dans la boucle HL-HM de Hel dans 1k93-2Ca, étant donné que ces régions sont d'une importance particulière pour l'interaction EF-CaM [228; 77]. Par ailleurs, le retrait ou l'ajout de calcium, tout en affectant la stabilité du complexe, conduit à des caractéristiques dynamiques locales assez différentes : aucun résidu graine n'est trouvé dans la boucle SA de 1k93-0Ca ni dans la boucle SC de 1k93-0Ca et de 1k93-4Ca ; le nombre de résidus graines dans les boucles HL-HM et HN-HO de Hel passe à cinq dans 1k93-0Ca et à trois dans 1k93-4Ca.

Des résidus graines ont également été identifiés dans CaM (Table III) et leurs corrélations résiduelles sont largement influencées par le niveau de calcium (Fig. 3). Dans 1k93-2Ca, deux groupes de résidus sont définis dans la boucle calcique de S2, autour de Ala-57 et Gly-61. Dans la matrice de corrélations résiduelles (Fig. 3b), le second motif « EF-hand » de CaM apparaît comme un bloc, qui s'étend du résidu 43 jusqu'au résidu 80 du linker central. Un pic de corrélation est aussi visible dans S1 (résidus 20-31) et deux pics hors diagonaux démontrent l'inter-corrélation entre les deux boucles calciques (résidus 20-31 et 56-67). Aucune corrélation n'est observée dans C-CaM, si ce n'est un pic étroit dans la boucle HF-HG entre les deux motifs « EF-hand » (résidus 108-118). Asp-129 a été identifié comme résidu graine, mais ses mouvements sont plutôt indépendants car il présente un profil de corrélation très plat, avec des valeurs absolues en dessous de 0.01 (non montré).



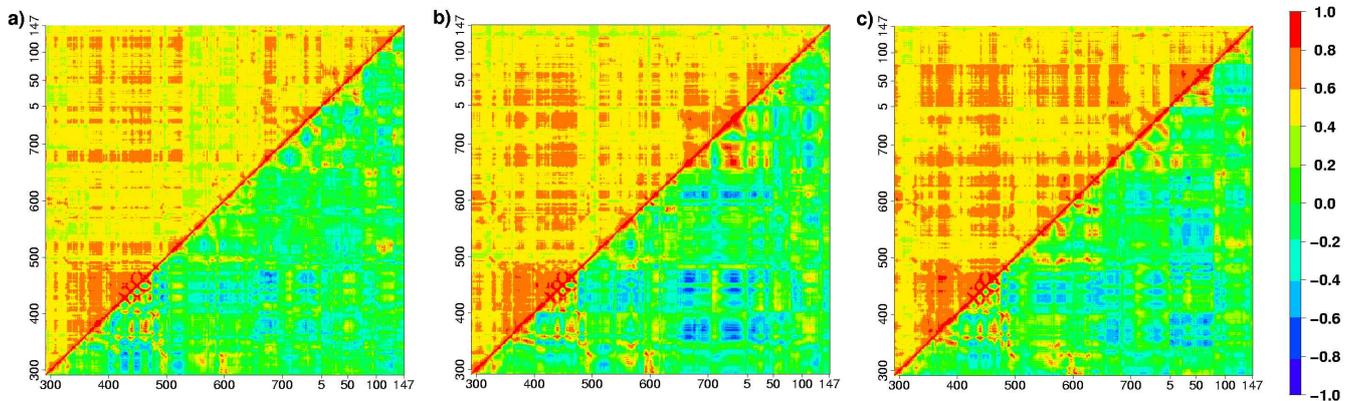
**FIG. 3: Matrices de corrélations résiduelles *LFA* de CaM :** (a) 1k93-0Ca, (b) 1k93-2Ca et (c) 1k93-4Ca. Les numéros de résidus de CaM (5-147) sont indiqués sur les axes x et y.

Le retrait de calcium de S3 et S4 dans 1k93-0Ca réduit la corrélation entre S1 et S2, tandis qu'elle induit un nouveau pic de corrélation entre le résidu graine Asp-78 et les résidus N-terminaux (Fig. 3a), qui est relié au dépliement du linker central pendant la simulation. De plus, les résidus 54-58 de S2 sont anti-corrélés avec Asp-78 et l'hélice HG (résidus 108-129) de C-CaM (Fig. 3a). L'ajout de calcium dans N-CaM (S1 et S2) induit une expansion du bloc de corrélation du second motif « EF-hand » vers l'hélice HB (résidus 27-41) (Fig. 3c), tandis que des résidus graines sont trouvés plus haut dans la séquence de S2 (Ser-38 et Ala-45). Les deux pics hors diagonaux du

premier motif « EF-hand » de 1k93-2Ca (Fig. 3b) deviennent moins intenses, tandis que des pics d'anti-corrélation apparaissent entre la boucle calcique de S1 et la boucle HB-HC (résidus 39-44). Comme dans 1k93-0Ca, le linker central (résidu graine Arg-74) est corrélé à la queue N-terminale, en accord avec le pliage observé pendant la simulation 1k93-4Ca au Chapitre 3.

Les conclusions suivantes peuvent être tirées concernant l'effet du retrait ou de l'ajout de calcium sur le réseau de résidus de CaM. L'impact direct du retrait de calcium est observé dans la boucle HF-HG (Fig 5b) de C-CaM dans 1k93-0Ca. De manière plus indirecte, les mouvements de la boucle HF-HG et du linker central deviennent anti-corrélés avec ceux de HC de N-CaM. Ceci révèle une propagation du signal calcium. D'autre part, la matrice de corrélation de 1k93-4Ca met en évidence la propagation des changements conformationnels induits par le calcium à travers S2 jusqu'à l'hélice HB de S1. L'information est transmise entre les deux motifs « EF-hand » à travers des mouvements anti-corrélés de la boucle calcique et de la boucle HB-HC. Globalement, la méthode *LFA* permet de mettre en lumière des régions, autour desquelles la dynamique de EF est organisée, et de comprendre la propagation du signal calcium le long de la structure de CaM.

### 3.3 Corrélations généralisées



**FIG. 4: Comparaison des matrices *GC* et des matrices de Pearson.** Les matrices *GC* sont en haut à gauche et les matrices de Pearson en bas à droite. Elles ont été calculées à partir des trajectoires (a) 1k93-0Ca, (b) 1k93-2Ca, (c) 1k93-4Ca. Les numéros de résidus de EF (292-767) et de CaM (5-147) sont reportés sur les axes x et y.

Les mouvements globaux de EF-CaM ont été analysés avec des calculs de corrélations généralisées (*GC*) [162] sur les trajectoires *DM* des trois complexes. Basée sur la théorie de l'information mutuelle, cette mesure de corrélation se veut plus précise et moins biaisée que le coefficient de corrélation de Pearson, traditionnellement utilisé [162]. En effet, 50% des corrélations de Pearson (Fig. 4, *en bas à droite*) sont inférieures à 0.2 en valeur absolue (*en vert*). Au contraire, les corrélations généralisées (Fig. 4, *en haut à gauche*), atteignent des valeurs intenses ( $> 0.8$ ) et relient un nombre plus grand de résidus, distants dans la séquence protéique. De plus, les matrices de

corrélations de Pearson de 1k93-0Ca, 1k93-2Ca et 1k93-4Ca partagent entre 65 et 73% de similarité, tandis que les matrices  $GC$  partagent 45 à 68% de similarité (Table IV), et sont donc plus sensibles au niveau de calcium.

	1k93-0Ca	1k93-2Ca	1k93-4Ca
1k93-4Ca	0.45	0.49	1
1k93-2Ca	0.68	1	0.65
1k93-0Ca	1	0.73	0.71

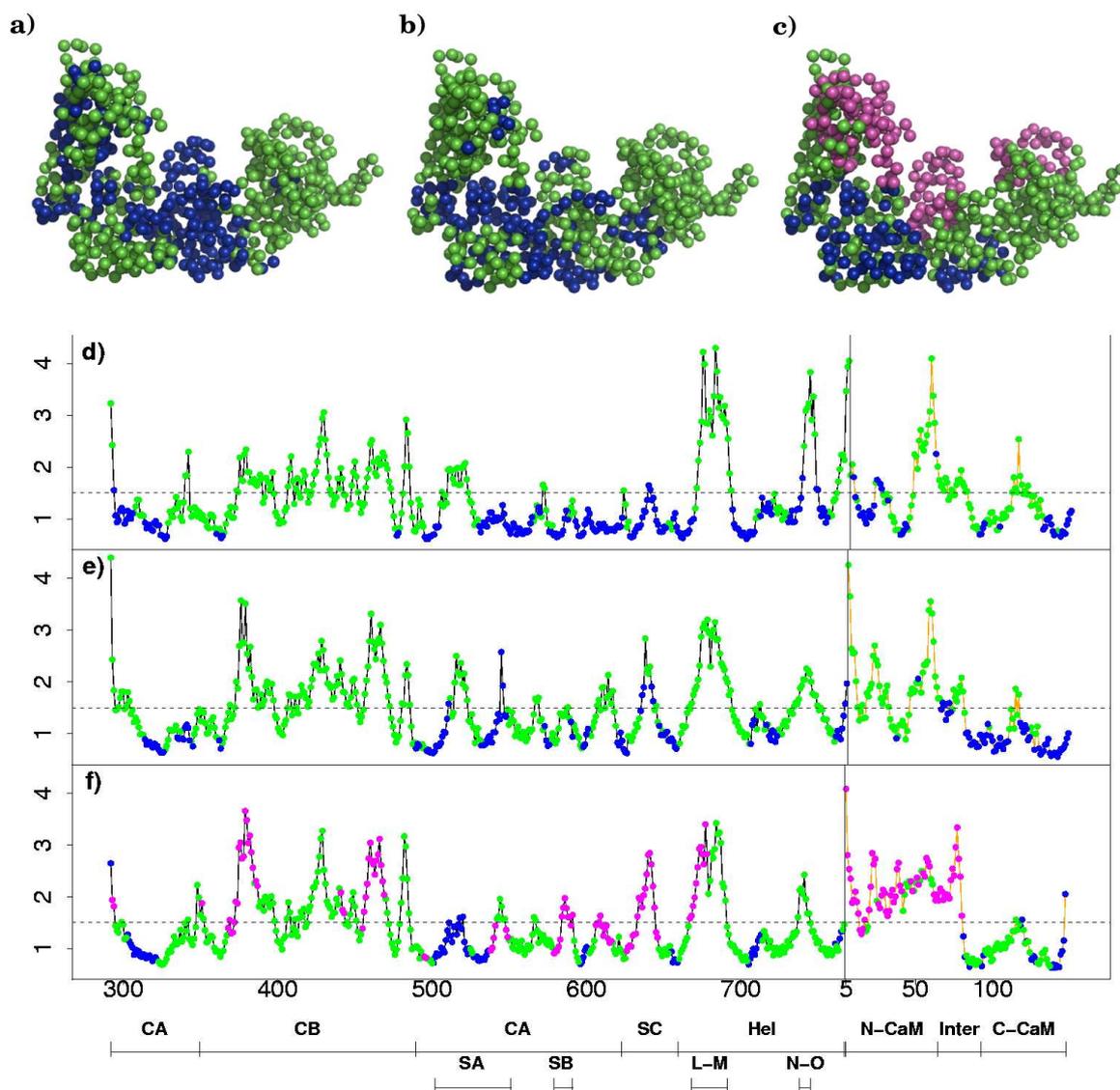
**TAB. IV: Similarités entre matrices de corrélations** : les similarités reportées sont celles des matrices de corrélations généralisées (*en haut à gauche*) et des matrices de corrélations de Pearson (*en bas à droite*), entre les différents niveaux de calcium. Le degré de similarité  $S$  entre deux matrices est évalué comme leur produit scalaire normalisé :  $S = \frac{\sum_{i,j} m_{ij} n_{ij}}{\sqrt{(\sum_{i,j} m_{ij}^2)(\sum_{i,j} n_{ij}^2)}}$ .

Une classification hiérarchique a été appliquée sur les matrices  $GC$  des trois complexes (Fig. 4). Les résidus ont été regroupés en fonction des similarités entre leurs profils de corrélation. Ainsi, les résidus membres du même groupe partagent une même corrélation globale avec le reste du réseau de résidus et présentent des corrélations similaires entre eux. La hauteur de coupure pour la classification a été déterminée comme décrit dans la section 2. Matériels et Méthodes. En utilisant cette approche, trois groupes ont été identifiés (Fig. 5 et Table I) : G1 (*en bleu*) regroupe les résidus non corrélés, avec une corrélation moyenne inférieure à 0.5, G2 (*en vert*) regroupe les résidus faiblement corrélés, avec des corrélations comprises entre 0.5 et 0.6, et G3 (seulement pour 1k93-4Ca, *en rouge*) regroupe des résidus fortement corrélés, avec une corrélation moyenne de 0.7.

Dans 1k93-2Ca, les résidus sont classés en deux groupes. Une large majorité de 446 résidus sont faiblement corrélés, dans G2 (Fig. 5b, *en vert*). La plupart des 173 résidus restants, dans G1 (*en bleu*), présentent de petites fluctuations  $DM$  (Fig. 5e). Cependant, les résidus de EF suivants : 510, 544-545 dans la boucle SA, 636, 641-642 dans la boucle SC, 766-767 dans la queue C-terminale ; et les résidus 50, 66-67, 70 dans S2 de CaM, présentent des fluctuations supérieures à 1.5 Å. Par conséquent, leurs mouvements sont relativement indépendants du reste du réseau de résidus.

Dans 1k93-0Ca, un plus grand nombre de 245 résidus peuplent G1 (*en bleu*). 149 d'entre eux forment une large région intermédiaire, localisée dans le domaine CA et la boucle SC, entre le coeur catalytique de EF et l'interface EF-CaM (Fig. 5a) et présentent des fluctuations  $DM$  uniformément plus petites que dans 1k93-2Ca (Fig. 5d). Au contraire, certains résidus de C-CaM, qui n'étaient pas corrélés (G1) dans 1k93-2Ca, deviennent faiblement corrélés (G2) dans 1k93-0Ca, tandis que leurs fluctuations augmentent. De plus, le retrait de calcium provoque le partitionnement des domaines Hel (résidus 660-767) et N-CaM (résidus 5-76) en quatre sous-ensembles de résidus flexibles et corrélés, séparés par des résidus plus rigides et non corrélés (Fig. 5d) : (i) la boucle HL-HM (résidus 668-691), (ii) la boucle HN-HO (résidus 738-745), (iii) la queue C-terminale de EF (résidus 754-767) et la queue N-terminale de CaM (résidus 5-6), (iv) le second motif « EF-hand » ainsi que le linker central (résidus 42-87). Ce résultat est en accord

avec la concentration de résidus graines observée dans les boucles HL-HM et HN-HO de EF, au paragraphe 3.2, avec la méthode *LFA*.



**FIG. 5: Classification hiérarchique des résidus en fonction de leurs corrélations généralisées.** Les groupes de résidus sont définis à partir des matrices  $GC$  de (a,d) 1k93-0Ca, (b,e) 1k93-2Ca et (c,f) 1k93-4Ca. (a-c) Les  $C(\alpha)$  sont représentés comme des sphères, colorées en fonction de leur groupe : les résidus non corrélés ( $< 0.5$ ) en G1 sont bleus, les résidus légèrement corrélés (0.5-0.6) en G2 sont verts, et les résidus fortement corrélés (0.7) en G3 sont rouges. (d-f) Fluctuations atomiques des  $C(\alpha)$  de EF (trait noir) et de CaM (trait orange), calculées le long des 12 dernières ns de chaque trajectoire *DM*. Les points sont colorés selon le code couleur des groupes. La valeur de 1.5 Å est indiquée par un trait en pointillés sur chaque graphe. Les positions dans la séquence du complexe des différents domaines CA, CB, Hel, N-CaM, le linker central (*Inter*) et C-CaM, de même que celles des boucles SA, SB, SC, HL-HM (*L-M*) et HN-HO (*N-O*), sont indiquées en bas.

Dans la matrice  $GC$  de 1k93-4Ca (Fig. 4c), il a été possible de classer les résidus dans trois groupes. G3 (*en rouge*) inclut des résidus fortement corrélés, localisés dans trois régions distinctes du complexe (Fig. 5c) : (i) dans CB, (ii) dans SB, SC ou CA, (iii) dans HL de Hel et N-CaM.

Ceci suggère que le changement conformationnel induit dans N-CaM par la fixation de calcium, a des résonances jusque dans la région catalytique de EF. En opposition à ces régions fortement corrélées, un groupe de résidus non corrélés (*en bleu*) est formé autour de C-CaM, comprenant les résidus 303-322 et 596-601 de CA, 501-537 de SA, 655-659 de SC et 705-713 de Hel.

L'analyse des matrices  $GC$  permet de dépeindre des modèles dynamiques du complexe EF-CaM, sensibles au niveau de calcium. Le modèle du complexe 1k93-2Ca montre une certaine homogénéité, la plupart des résidus étant faiblement corrélés, exception faite de régions localisées. Par contre, les modèles de corrélation des deux autres complexes illustrent la déstabilisation de l'interaction EF-CaM, mise en évidence dans le Chapitre 3. Par exemple, 1k93-0Ca est divisé en deux blocs corrélés, séparés par une région de résidus non corrélés. Le premier bloc correspond à CB et le second bloc est fragmenté en plusieurs groupes flexibles et corrélés, délimités par des résidus non corrélés et relativement rigides. Ainsi, le retrait de calcium rompt le couplage entre l'interface EF-CaM et la région catalytique, ce qui est en accord avec l'instabilité du complexe EF-CaM, expérimentalement observée en l'absence de calcium [255]. D'un autre côté, l'ajout de calcium dans 1k93-4Ca induit des corrélations élevées entre N-CaM et le coeur catalytique de EF. La tension introduite dans N-CaM, mise en évidence dans le paragraphe 3.2, est transmise à travers le complexe, même à de longues distances. Globalement, cette analyse montre que des corrélations dynamiques modérées sont indicatrices d'un complexe stable.

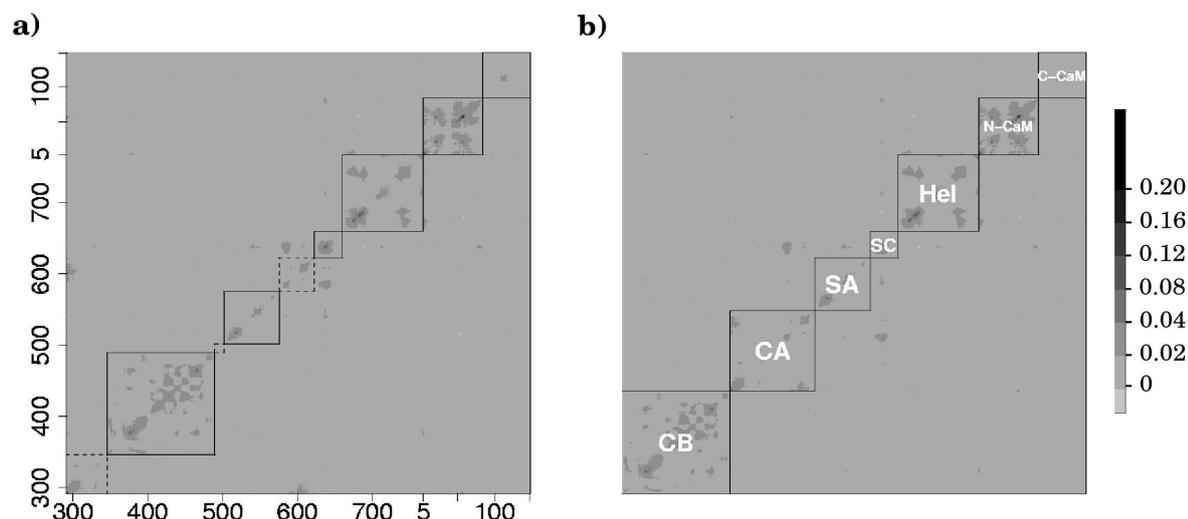
### 3.4 Cartes de dépendances énergétiques

	Définition de biologie structurale [77]		Définition <i>LFA</i>	
EF	CA	294-349 ; 490-622	CA	292-345 ; 490-501 ; 576-622
	SA	502-551	SA	502-575
	SB	578-591		
	CB	350-489	CB	346-489
	SC	630-659	SC	623-659
	Hel	660-767	Hel	660-767
CaM	N-CaM	5-64	N-CaM	5-84
	linker	65-92		
	C-CaM	93-147	C-CaM	85-147

**TAB. V: Définition des domaines ou pseudo-domaines du complexe EF-CaM :** numéros des résidus délimitant (i) les domaines de EF-CaM dans la définition de biologie structurale, selon Drum et al. [77], (ii) les pseudo-domaines de EF-CaM dans la nouvelle définition avec la méthode *LFA*. Les résidus impliqués dans le site catalytique de EF sont Arg-329 (L), Lys-346 (R), Leu-348 (R), His-351 (R), Lys-353 (R), Ser-354 (R), Lys-372 (R), Asp-491, Asp-493 (L), Thr-548 (L), His-577 (L), Thr-579 (L), Asp-582 (L), Asn-583 (L), Phe-586 (L), Glu-588 (L), et sont localisés soit sur le côté gauche (L), soit sur le côté droit (R) du site.

L'effet du calcium sur les influences énergétiques au sein de EF-CaM a été étudié en réalisant des calculs *MMPBSA* d'énergie libre sur les trajectoires *DM*. L'approche proposée par Hamacher et al. [113] et décrite avec plus de détails dans la section 2. Matériels et Méthodes, consiste, dans un premier temps, à calculer et à ranger les énergies libres de liaison de chaque domaine  $X$  au reste

du complexe. On obtient ainsi un classement de référence. Dans un deuxième temps, on détermine l'influence du retrait de chacun des autres domaines  $Y$  sur ce classement de référence. Si le rang de  $X$  est moins bon quand  $Y$  est retiré, alors on considère que  $Y$  a une influence énergétique sur la stabilité de  $X$  au sein du complexe. Les influences énergétiques sont matérialisées par des flèches sur des cartes de dépendances (Fig. 7).



**FIG. 6: Matrice de corrélations résiduelles *LFA*.** Elle a été évaluée à partir des quinze premiers vecteurs *ACP*. Les pseudo-domaines sont délimités par des traits pleins, excepté CA, sur la gauche, pour lequel des traits pointillés sont utilisés. **b** Les lignes et les colonnes ont été permutes de manière à grouper les résidus appartenant à CA dans le second bloc. Les numéros de résidus de EF (292-767) et CaM (5-147) sont indiqués sur les axes x et y.

Sept pseudo-domaines ont été définis à l'intérieur de EF-CaM, avec les résultats de l'approche *LFA*. Ces définitions impliquent des ajustements à la définition de biologie structurale proposée par Drum et al. [77] (Table V). La matrice de corrélations résiduelles *LFA* de 1k93-2Ca est reportée Fig. 6a, avec les nouvelles délimitations des sept pseudo-domaines. Un premier ajustement a été effectué aux limites de CA et CB, définis par les résidus 292-349, 491-522, et les résidus 350-489 respectivement selon Drum et al. [77]. Dans cette définition de biologie structurale (Table V), la plupart des résidus localisés sur le côté droit du site catalytique appartient à CB et la plupart des résidus localisés sur le côté gauche appartient à CA. Les seules exceptions sont observées pour Lys-346 et Leu-348, à la fois sur le côté droit et dans CA. Considérant la polarisation du site, mise en évidence par l'approche *LFA*, CB a été re-défini entre les résidus 346 et 489. De plus, SA a été extrait de CA, les résidus 502-575 formant un motif de corrélation bien défini, non corrélé avec le reste de CA (Fig. 6a). Les résidus 623-629 ont été ajoutés à la définition de SC (résidus 623-659) et la définition originale de Hel a été conservée (résidus 660-767). Dans CaM, le linker central et N-CaM sont associés dans un même motif de corrélation (Fig. 6a). N-CaM a donc été re-défini des résidus 5 à 84 et C-CaM des résidus 85 à 147. Après permutation des colonnes et des lignes, la matrice de corrélation *LFA* (Fig. 6b) ne présente plus de corrélations inter-domaines, excepté

entre SC et SB (résidus 576-591) de CA.

	$\Delta\Delta E_{ele}$	$\Delta\Delta E_{vdw}$	$\Delta\Delta E_{int}$	$\Delta\Delta G_{sol}^p$	$\Delta\Delta G_{sol}^{np}$	$\Delta\Delta G$
CA	$-552 \pm 87$	$-314 \pm 11$	$52 \pm 4$	$800 \pm 85$	$11 \pm 3$	$-3 \pm 22$
CB	$-407 \pm 60$	$-82 \pm 5$	$24 \pm 3$	$497 \pm 58$	$14 \pm 2$	$44 \pm 13$
SA	$-1410 \pm 91$	$-241 \pm 9$	$20 \pm 3$	$1526 \pm 84$	$17 \pm 3$	$-89 \pm 24$
SC	$-881 \pm 45$	$-241 \pm 9$	$19 \pm 3$	$1049 \pm 43$	$2 \pm 2$	$-51 \pm 20$
Hel	$-1911 \pm 78$	$-161 \pm 9$	$10 \pm 2$	$2035 \pm 75$	$21 \pm 2$	$-6 \pm 17$
N-CaM	$-1387 \pm 86$	$-101 \pm 11$	$10 \pm 2$	$1445 \pm 82$	$25 \pm 2$	$-9 \pm 18$
C-CaM	$-705 \pm 87$	$-251 \pm 12$	$10 \pm 2$	$830 \pm 77$	$16 \pm 3$	$-101 \pm 29$

**TAB. VI: Énergies libres de liaison de référence des sept pseudo-domaines de 1k93-2Ca.** Les différentes contributions à l'énergie libre de liaison totale  $\Delta\Delta G$  sont donnés en  $\text{kcal.mol}^{-1}$  : l'énergie d'interaction de Coulomb ( $\Delta\Delta E_{ele}$ ), l'énergie de van der Waals ( $\Delta\Delta E_{vdw}$ ), l'énergie interne ( $\Delta\Delta E_{int}$ ), l'énergie de solvation polaire ( $\Delta\Delta G_{sol}^p$ ) et l'énergie de solvation non-polaire ( $\Delta\Delta G_{sol}^{np}$ ).

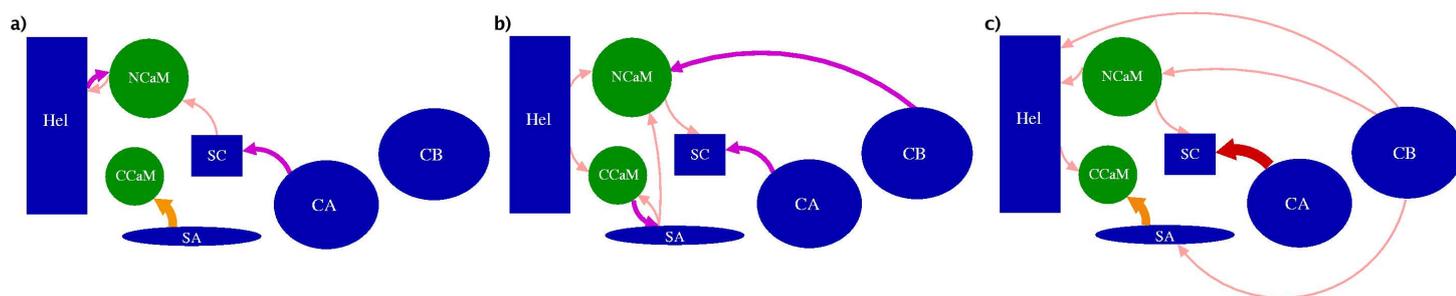
	$\Delta\Delta\Delta E_{ele}$	$\Delta\Delta\Delta E_{vdw}$	$\Delta\Delta\Delta E_{int}$	$\Delta\Delta\Delta G_{sol}^p$	$\Delta\Delta\Delta G_{sol}^{np}$	$\Delta\Delta\Delta G$
<b>influences de courte portée</b>						
CA $\rightarrow$ SC <sup>0</sup>	$430 \pm 64$	$105 \pm 6$	$-8 \pm 2$	$-478 \pm 60$	$3 \pm 2$	$52 \pm 14$
CA $\rightarrow$ SC <sup>2</sup>	$408 \pm 38$	$107 \pm 7$	$-9 \pm 2$	$-464 \pm 37$	$7 \pm 2$	$49 \pm 14$
CA $\rightarrow$ SC <sup>4</sup>	$407 \pm 34$	$116 \pm 8$	$-9 \pm 2$	$-457 \pm 34$	$6 \pm 2$	$63 \pm 15$
SA $\rightarrow$ C-CaM <sup>0</sup>	$803 \pm 47$	$82 \pm 6$	$0 \pm 0$	$-807 \pm 45$	$2 \pm 2$	$80 \pm 10$
SA $\leftrightarrow$ C-CaM <sup>2</sup>	$612 \pm 47$	$79 \pm 8$	$0 \pm 0$	$-603 \pm 46$	$2 \pm 2$	$90 \pm 14$
SA $\rightarrow$ C-CaM <sup>4</sup>	$565 \pm 37$	$77 \pm 5$	$0 \pm 0$	$-585 \pm 33$	$0 \pm 1$	$56 \pm 9$
Hel $\leftrightarrow$ N-CaM <sup>0</sup>	$1715 \pm 82$	$84 \pm 8$	$0 \pm 0$	$-1738 \pm 89$	$-7 \pm 2$	$54 \pm 17$
Hel $\rightarrow$ N-CaM <sup>2</sup>	$1411 \pm 58$	$49 \pm 8$	$0 \pm 0$	$-1450 \pm 60$	$-11 \pm 2$	$-1 \pm 14$
N-CaM $\rightarrow$ Hel <sup>4</sup>	$1128 \pm 76$	$64 \pm 8$	$0 \pm 0$	$-1159 \pm 78$	$-6 \pm 2$	$28 \pm 11$
Hel $\rightarrow$ C-CaM <sup>2</sup>	$666 \pm 39$	$55 \pm 5$	$0 \pm 0$	$-689 \pm 35$	$-6 \pm 1$	$25 \pm 15$
Hel $\rightarrow$ C-CaM <sup>4</sup>	$768 \pm 29$	$66 \pm 5$	$0 \pm 0$	$-807 \pm 32$	$-8 \pm 1$	$19 \pm 13$
SC $\rightarrow$ N-CaM <sup>0</sup>	$552 \pm 29$	$3 \pm 2$	$0 \pm 0$	$-531 \pm 28$	$0 \pm 1$	$25 \pm 6$
N-CaM $\rightarrow$ SC <sup>2</sup>	$473 \pm 27$	$4 \pm 2$	$0 \pm 0$	$-451 \pm 27$	$0 \pm 1$	$26 \pm 5$
N-CaM $\rightarrow$ SC <sup>4</sup>	$442 \pm 34$	$6 \pm 3$	$0 \pm 0$	$-417 \pm 31$	$-3 \pm 1$	$29 \pm 11$
SA $\rightarrow$ N-CaM <sup>2</sup>	$692 \pm 58$	$12 \pm 2$	$0 \pm 0$	$-682 \pm 56$	$-4 \pm 1$	$18 \pm 7$
<b>influences de longue portée</b>						
CB $\rightarrow$ N-CaM <sup>2</sup>	$84 \pm 13$	$0 \pm 0$	$0 \pm 0$	$-80 \pm 13$	$0 \pm 0$	$4 \pm 2$
CB $\rightarrow$ N-CaM <sup>4</sup>	$106 \pm 14$	$0 \pm 0$	$0 \pm 0$	$-105 \pm 13$	$0 \pm 0$	$1 \pm 1$
CB $\rightarrow$ SA <sup>4</sup>	$22 \pm 23$	$11 \pm 4$	$0 \pm 0$	$-34 \pm 22$	$-3 \pm 1$	$-5 \pm 7$
CB $\rightarrow$ Hel <sup>4</sup>	$-55 \pm 6$	$0 \pm 0$	$0 \pm 0$	$54 \pm 7$	$0 \pm 0$	$-1 \pm 1$

**TAB. VII: Différences d'énergie libre de liaison et de chacun des termes d'énergie.** Chaque différence correspond à une influence énergétique identifiée dans 1k93-0Ca<sup>(0)</sup>, 1k93-2Ca<sup>(2)</sup> et 1k93-4Ca<sup>(4)</sup>, donnés en  $\text{kcal.mol}^{-1}$  :  $\Delta\Delta\Delta G[D* \leftrightarrow D] = \Delta\Delta G[D, C \setminus D*] - \Delta\Delta G[D, C] = \Delta\Delta G[D*, C \setminus D] - \Delta\Delta G[D*, C]$ . La flèche de  $\Delta\Delta\Delta G[D* \rightarrow D]$  indique que le retrait de  $D*$  déprécie le rang de  $D$ , et réciproquement. La flèche de  $\Delta\Delta\Delta G[D* \leftrightarrow D]$  indique que le retrait de  $D$  déprécie également le rang de  $D*$  (voir Fig. 7). La quantité  $\Delta\Delta\Delta G[D* \leftrightarrow D]$  est symétrique par rapport à  $D$  et  $D*$ .

Les énergies libres de liaison ont été calculées sur chaque trajectoire  $DM$  avec la méthode *MMPBSA* [236]. Les contributions des différents termes d'énergie à l'énergie libre de liaison totale  $\Delta\Delta G$  sont donnés pour 1k93-2Ca (Table VI). Ils sont du même ordre de grandeur pour les trois complexes. Les termes les plus larges sont les énergies d'interaction de Coulomb  $\Delta\Delta E_{ele}$  et les pénalités de désolvation polaire  $\Delta\Delta G_{sol}^p$ . Ils présentent des valeurs absolues dans l'intervalle 400-2000  $\text{kcal.mol}^{-1}$  et se compensent l'un l'autre. Les autres termes sont beaucoup plus faibles :

les énergies de van der Waals sont comprises entre  $-80$  et  $-320$  kcal.mol $^{-1}$ , les termes de solvation non polaire et les énergies internes valent entre  $0$  et  $50$  kcal.mol $^{-1}$ .

Comme montré sur la Fig. 7, les résultats *EDM* sont sensibles au niveau de calcium et révèlent à la fois des influences de courte portée et de longue portée. Les influences de courte portée sont associées à des variations défavorables d'énergies libres de liaison  $\Delta\Delta G$  (Table VII). Ainsi, les différences d'énergie libre positives  $\Delta\Delta\Delta G[\text{CA} \rightarrow \text{SC}]$ ,  $\Delta\Delta\Delta G[\text{SA} \leftrightarrow \text{C-CaM}]$ ,  $\Delta\Delta\Delta G[\text{Hel} \leftrightarrow \text{N-CaM}]$  (excepté dans 1k93-2Ca) et  $\Delta\Delta\Delta G[\text{Hel} \rightarrow \text{C-CaM}]$  sont dues aux pénalités de van der Waals. Par contre, la pénalité électrostatique de Coulomb est prépondérante dans la différence positive  $\Delta\Delta\Delta G[\text{SC} \leftrightarrow \text{N-CaM}]$ . La différence  $\Delta\Delta\Delta G[\text{SA} \rightarrow \text{N-CaM}]$  résulte d'effets mixtes électrostatiques et de van der Waals. Par conséquent, les influences de courte portée correspondent à des interactions/contacts hydrophobes ou électrostatiques directs. Au contraire, les influences de longue portée de CB sur SA, Hel et N-CaM ne sont pas associées à de larges variations d'énergie (Table VII). En effet, les contributions de van der Waals et non polaires sont plutôt faibles, tandis que les contributions électrostatiques et polaires se compensent l'une l'autre, ce qui n'est pas surprenant puisque CB est à plus de  $30$  Å de SA, Hel ou N-CaM. Les pertes de rangs pour ces domaines proviennent en fait d'un changement favorable dans l'énergie de liaison de CA quand CB est retiré, de  $36$  et  $65$  kcal.mol $^{-1}$  dans 1k93-2Ca et 1k93-4Ca respectivement.



**FIG. 7: Cartes de dépendances énergétiques :** (a) 1k93-0Ca, (b) 1k93-2Ca et (c) 1k93-4Ca. Le complexe EF-CaM est représenté schématiquement par ses sept pseudo-domaines, avec CaM en vert (pseudo-domaines N-CaM et C-CaM) et EF en bleu (pseudo-domaines CA, CB, SA, SC, Hel). Les flèches représentent les influences énergétiques, dont la largeur dépend du nombre de rangs perdus dans la comparaison des énergies libres de liaison : l'intensité vaut 1 en rose, 2 en mauve, 3 en orange et 4 en rouge. Les cartes ont été dessinées avec Xfig [3].

Hamacher et al. [113] ont utilisé la méthode *EDM* pour prédire l'ordre d'assemblage de la petite sous-unité ribosomale. Ici, au lieu de considérer l'assemblage de protéines au sein d'une supra-structure, nous avons considéré l'assemblage de pseudo-domaines, certains liés covalamment, au sein d'un complexe protéine-protéine. Du fait de ce choix, et parce que l'établissement de l'interaction EF-CaM induit des réarrangements conformationnels qui ne sont pas échantillonnés dans nos simulations, l'ordre d'assemblage que nous obtenons (Fig. 7) ne décrit pas le mécanisme réel de la formation du complexe EF-CaM. Au lieu de cela, il donne une description du schéma de dépendance de stabilité entre les différentes parties du complexe, comme dans une architecture en château de cartes : plus un pseudo-domaine est influencé par les autres, plus il occupe une place basse dans la hiérarchie des stabilités relatives au sein du complexe. Ainsi, nous proposons que

les pseudo-domaines non soumis à influence définissent les points de nucléation de l'architecture du complexe.

De cette façon, CA, CB et Hel sont les points de nucléation de 1k93-2Ca (Fig. 7b). Les influences de Hel et SA sur C-CaM correspondent à des interactions directes à l'interface EF-CaM. De plus, l'influence de C-CaM sur SA peut mettre en évidence le changement conformationnel que SA subit lorsqu'il interagit avec C-CaM [77]. N-CaM est sujet aux influences de SA, CB et Hel, ce qui signifie qu'il joue un rôle moins important que C-CaM dans l'interaction avec EF, en accord avec les expériences [77; 228]. Enfin, les influences de N-CaM et CA sur SC illustrent le changement conformationnel de SC lors de l'interaction avec N-CaM, qui conduit à la stabilisation de SB, et ainsi du site catalytique [77].

Dans 1k93-0Ca (Fig. 7a), l'ensemble de nucléation (CA, CB et SA) est déplacé vers la région catalytique de EF, révélant un affaiblissement de l'interaction EF-CaM. Aucune influence de longue portée n'est observée, de telle sorte que le couplage est rompu entre la région d'interface EF-CaM et la région catalytique. Cette rupture est en accord avec l'effondrement du site catalytique observé dans cette simulation au Chapitre 3 et avec les corrélations généralisées, au paragraphe 3.3 (Fig. 5a). L'inter-dépendance de SA et C-CaM est convertie en une influence plus forte de SA sur C-CaM. Hel n'est plus lié à C-CaM : le couplage entre les deux pseudo-domaines de CaM est rompu.

Dans 1k93-4Ca (Fig. 7c), une réduction de l'ensemble de nucléation (CA, CB) et de nouvelles influences de longue portée de CB sur Hel et SA, au sein même de EF, sont observées. Elles suggèrent une architecture en château de cartes plus contrainte et plus instable, et révèlent une déstabilisation des domaines de EF en contact avec CaM. Une telle déstabilisation peut être reliée au changement conformationnel induit par l'ajout de calcium dans N-CaM observé au Chapitre 3, qui trouve une résonance jusque dans la région catalytique de EF (cf. Fig. 5c). On peut noter que l'influence de CA sur SC est renforcée. De plus, l'influence de Hel sur C-CaM est maintenue. L'inter-dépendance de SA et C-CaM est convertie en une influence plus forte de SA sur C-CaM, comme observé dans 1k93-0Ca.

Les résultats *EDM* fournissent un modèle pour la dépendance énergétique relative des domaines à l'intérieur du complexe et varient selon le niveau de calcium. La hiérarchie des stabilités relatives des domaines met en lumière le mécanisme de l'interaction EF-CaM décrit expérimentalement [77; 228]. En effet, dans 1k93-2Ca, les influences énergétiques sont centrées autour de l'interface EF-CaM. Le retrait ou l'ajout de calcium produit différents effets sur les cartes de dépendances, conduisant toutes deux à une déstabilisation. D'un côté, dans 1k93-0Ca, les couplages sont rompus au sein du complexe, particulièrement entre le coeur catalytique et le domaine Hel de EF. De l'autre côté, dans 1k93-4Ca, le nombre de points de nucléation diminue et de nouvelles influences au sein de EF révèlent des contraintes qui déstabilisent l'architecture du complexe.

## 4 Discussion

Le complexe EF-CaM résulte de l'assemblage de deux protéines, elles-même composées de domaines globulaires attachés par un lien flexible : les domaines N- et C-terminaux reliés par le linker central dans CaM, le domaine hélicoïdal et le coeur catalytique reliés par la boucle SC dans EF. Le réseau de résidus est organisé en deux blocs : (i) la région d'interface, où CaM et EF sont en contact, composée de Hel et SA de EF et des deux domaines de CaM, (ii) la région catalytique, où le site catalytique réside, composée des domaines CA et CB. SC joue un rôle important dans le maintien de l'activité catalytique du complexe, car il interagit avec CaM et stabilise SB dans le site catalytique.

Deux approches différentes ont été utilisées pour sonder la relation entre aspects dynamiques et énergétiques de EF-CaM, l'une basée sur les corrélations dynamiques, en utilisant les corrélations généralisées, et l'autre basée sur les influences énergétiques, en utilisant les calculs d'énergie libre. De manière intéressante, les résultats des deux méthodes sont en accord, et l'on peut donc en faire une interprétation commune dans le contexte de l'effet du calcium sur le réseau de résidus de EF-CaM. Les deux méthodes dépeignent une image homogène et équilibrée de l'architecture de 1k93-2Ca. La stabilité de 1k93-2Ca est associée à des corrélations modérées à travers le réseau de résidus entier et à des influences énergétiques centrées autour de l'interface EF-CaM. La déstabilisation du complexe, par le retrait ou l'ajout de calcium, est aussi révélée par les deux méthodes : dans 1k93-0Ca, les corrélations manquantes dans les matrices  $GC$ , aussi bien que les influences énergétiques manquantes dans les  $EDM$  suggèrent une rupture de couplage entre les différentes parties du réseau de résidus, tandis que dans 1k93-4Ca, de fortes corrélations dans les matrices  $GC$  et de nouvelles influences énergétiques entre les domaines de EF mettent en évidence un déséquilibre dans le réseau de résidus.

Au vu de ces résultats, nous pouvons formuler l'hypothèse que la stabilité du complexe dépend de l'organisation de son réseau de résidus et, particulièrement, de la capacité de chaque noeud de ce réseau à communiquer avec tous les autres, à travers des mouvements corrélés ou des contributions à une architecture en château de cartes. Ce résultat est en accord avec Bahar et al. [54; 245], qui ont établi une relation formelle entre la dynamique des protéines et la transmission d'information à travers leurs structures [54], ainsi qu'avec le processus d'association protéine-protéine [245]. Nous proposons de dénoter cette propriété la connexité du réseau de résidus. Comme décrit précédemment, la connexité du réseau de résidus de EF-CaM est affectée par le niveau de complexation au calcium. D'une part, la rupture de couplage et le déséquilibre des points de nucléation dans 1k93-0Ca révèlent un manque de communication à travers le réseau. D'autre part, les fortes corrélations entre N-CaM et le coeur catalytique de EF et les influences à longue portée au sein de EF dans 1k93-4Ca correspondent à une connexité étendue du réseau. A cause de cette connexité étendue, l'architecture en château de cartes du complexe est déstabilisée. Les modèles présentés ici sont donc en accord qualitatif avec la stabilité relative de l'interaction EF-CaM selon le niveau de calcium [255].

Néanmoins, la question demeure, de savoir comment les corrélations dynamiques et les contributions énergétiques peuvent être reliées de manière formelle. D'une part, les influences *EDM* sont basées sur des changements relatifs – et pas nécessairement absolus, d'énergies libres de liaison, et révèlent des dépendances énergétiques entre les domaines, dans le contexte de la stabilité ou l'intégrité du complexe. D'autre part, elles reflètent dans une certaine mesure les covariances dynamiques, à travers le support des domaines entre eux, par analogie avec le principe d'action-réaction dans la mécanique classique.

## 5 Conclusion

Dans le chapitre précédent, nous avons caractérisé la plasticité conformationnelle des partenaires de l'interaction EF-CaM. Ici, nous avons cherché à étudier l'effet des ions  $\text{Ca}^{2+}$  sur l'organisation du réseau de résidus du complexe pris dans son ensemble. Pour cela, nous avons abordé le problème sous différents angles, à la fois géométrique et énergétique, à la fois local et global. Nous avons pu donner une description qualitative de la propagation du signal calcium à travers le complexe, en s'intéressant aux aspects dynamique locaux. L'émergence du concept de connexité du réseau de résidus comme critère de stabilité a permis d'unifier nos modèles dynamiques et énergétiques du complexe.



# Recherche d’Inhibiteurs de EF

## 1 Introduction

La recherche et le développement de nouveaux inhibiteurs par des approches qui ciblent les interactions entre biomolécules est une voie prometteuse pour moduler les processus cellulaires de manière rationnelle [265]. En effet, les interactions physiques qui contrôlent la machinerie cellulaire sont régulées de manière allostérique à travers la modulation d’équilibres conformationnels [220; 249]. Dans le contexte d’une interaction protéine-protéine établie à travers un déplacement de l’équilibre conformationnel des deux partenaires, il devrait être possible d’empêcher cette interaction en ciblant une poche à la surface d’une des deux biomolécules, qui n’apparaîtrait que dans un sous-ensemble de conformations à l’équilibre. Cette approche est séduisante car elle permet d’élargir le champ des cibles potentielles et peut être un moyen de prévenir les phénomènes de résistance.

L’identification d’un site susceptible de jouer un rôle allostérique nécessite une description plausible des mécanismes biophysiques de l’interaction, qui peut être obtenue par la détermination d’un chemin de réaction entre les conformations inactive et active de la biomolécule cible. Dans ce troisième volet, nous présentons une méthode de détermination de chemin par modélisation moléculaire, permettant de définir une poche allostérique à la surface de EF, afin d’identifier de nouveaux inhibiteurs de la toxine.

La plupart des approches utilisées jusqu’ici pour la recherche d’inhibiteurs de EF [233; 227; 168; 135; 50] se sont focalisées sur le site enzymatique, se basant sur des tests *in vitro* ou *in vivo*. Un seul inhibiteur de l’interaction EF-CaM a pu être identifié parmi 10 000 molécules. Récemment, une stratégie *in silico*, faisant usage d’un pharmacophore 3D construit à partir de fragments, a été appliquée avec succès [50]. Ici, nous abordons le problème sous un angle nouveau, en combinant méthodes expérimentales et computationnelles.

L’étude approfondie de la dynamique du complexe EF-CaM et de l’influence du calcium sur sa stabilité a mis en évidence la plasticité conformationnelle remarquable des partenaires de l’interaction EF et CaM. En l’occurrence, la protéine EF subit une large transition lors de la formation du complexe. Un protocole original, basé sur l’utilisation de l’algorithme d’affinement par pic conjugué [89] (*conjugate peak refinement* ou *CPR*), a permis de caractériser cette mobilité structurale.

L'obtention de conformations intermédiaires plausibles a servi à l'élaboration d'une stratégie de sélection de touches potentielles par criblage virtuel. Parmi les 18 touches prédites *in silico*, six inhibiteurs ont été validés expérimentalement, ce qui correspond à un niveau d'enrichissement d'environ 30%. L'un d'entre eux supprime l'activité de EF avec une affinité meilleure que 10  $\mu\text{M}$ .

## 2 Matériels et Méthodes

### 2.1 Simulations et chemin de transition

#### Structures

Quatre structures cristallographiques de EF ont été utilisées, en complexe avec CaM (entrées PDB : 1K93, 1K90, 1XFX) ou libre (entrée PDB : 1K8T) [77; 228]. La conformation de CyaA analysée ici correspond à la structure cristallographique de CyaA-(C-CaM) (entrée PDB : 1YRT) [108].

#### Simulations *DM*

Les simulations *DM* ont été effectuées avec les modules sander et PMEMD d'AMBER 8 et 10 [41; 42]. Les conditions de simulation du complexe EF-CaM chargé de zéro (1k93-0Ca), deux (1k93-2Ca) ou quatre (1k93-4Ca) ions  $\text{Ca}^{2+}$  sont décrites dans le Chapitre 3, section Matériels et Méthodes. Une simulation additionnelle de 15 ns de EF libre en conformation fermée a été réalisée, avec les mêmes conditions, si ce n'est que la température a été régulée par un thermostat de Langevin [173], avec une fréquence de collision de 2  $\text{ps}^{-1}$ .

#### Chemin de transition conformationnelle

La protéine EF a été simulée seule, sans ions ni co-facteurs, en utilisant le champ de force parm19 dans CHARMM version 29 [38]. La représentation du solvant est implicite, selon deux niveaux d'approximation. Le niveau le plus simple consiste en une fonction diélectrique distance-dépendante de forme sigmoïdale [164], combinée à une fonction de déplacement de force prenant effet entre 7 et 8 Å. Dans la seconde phase d'affinement du chemin (voir plus loin), un modèle de solvation plus sophistiqué a été élaboré, basé sur un potentiel analytique de l'électrostatique de continuum (*analytic continuum electrostatics potential* ou *ACE2*) [221]. La constante diélectrique a été fixée à 1 à l'intérieur de la protéine, et 80 à l'extérieur. Une fonction de déplacement de force prend effet entre 10 et 12 Å pour compléter le modèle *ACE2*, et une fonction de transfert est appliquée entre 7 et 8 Å, pour modéliser l'interaction de van der Waals.

La structure 1K8T de EF seule (forme fermée) et la chaîne C de la structure 1K93 du complexe EF-CaM (forme ouverte) [77] ont été choisies comme points initial et final de la transition conformationnelle. La boucle SB (résidus 289-299) de EF étant désordonnée dans la structure

1K8T, elle a été modélisée par simple translation de ses coordonnées depuis la structure 1K93. Une minimisation aux extrémités de cette boucle SB modélisée a permis de finaliser son insertion dans le point initial de la transition. Les coordonnées internes des points initial et final ont ensuite été comparées, pour détecter d'éventuels retournements de chaînes latérales et exclure des mouvements rotatifs. Les orientations des chaînes latérales ont été manuellement rendues similaires dans les deux états de la protéine, à condition que l'opération ne rompsse pas de liaisons hydrogènes importantes.

L'étape initiale de la détermination du chemin de transition consiste en deux simulations *DM* orientées (*steered molecular dynamics* ou *SMD*), chacune partant de chaque point extrême. La force de guidage a été évaluée à partir du *RMSD* entre les deux points extrêmes, avec une constante de force de  $0.5 \text{ kcal.mol}^{-1}.\text{Å}^{-1}$ . Les simulations ont été effectuées à une température de 300 K, en utilisant un thermostat de Langevin [85], avec un coefficient d'amortissement de  $100 \text{ ps}^{-1}$ . Un pas de temps de 1 fs a été choisi, pour une durée totale de 100 ps, chaque simulation fournissant un ensemble de 500 conformations.

Les ensembles de conformations échantillonnées le long des deux simulations *SMD* représentent le premier jet du chemin de transition. L'affinement de celui-ci se fait de manière itérative, en alternant deux types de processus : (i) une réduction du nombre de conformations, (ii) un échantillonnage entre des paires de conformations. L'objectif de ces processus est de définir le chemin de transition par une série de conformations liant les points initial et final existants, dans le contexte du champ de force utilisé. Pour être considéré comme un chemin de transition, la série de conformations doit vérifier la condition suivante : l'énergie de chaque conformation de la série, ou de toute conformation obtenue par interpolation linéaire entre deux conformations de la série, est inférieure à un seuil donné. La réduction du seuil guide l'affinement du chemin.

Le processus (i) (réduction du nombre de conformations) consiste à découper le chemin courant de manière itérative. Pour chaque sous-chemin allant de la conformation  $n_k$  au point final de la transition (forme ouverte), on cherche l'indice le plus grand  $k'$  tel que le chemin allant de  $n_k$  à  $n_{k'}$  répond à l'exigence du seuil d'énergie. La conformation  $n_{k'}$  devient alors la conformation qui suit  $n_k$  dans le chemin réduit.

Le processus (ii) (échantillonnage entre des paires de conformations) utilise l'algorithme d'affinement par pic conjugué [89] (*conjugate peak refinement* ou *CPR*) entre des paires de conformations  $n_k$  et  $n_{k'}$  respectant plusieurs critères : (a) le nombre de conformères entre les conformations  $n_k$  et  $n_{k'}$  est en dessous d'un certain seuil  $\nu$ , (b) le *RMSD* entre les coordonnées des conformations  $n_k$  et  $n_{k'}$  est plus petit qu'un rayon de coupure  $\delta$ , (c) le rapport du *RMSD* sur la distance curvilinéaire  $D_{curv}$  est inférieur à un certain ratio  $\epsilon$ .  $D_{curv}$  est défini comme :

$$D_{curv} = \sum_{s=k}^{k'} RMSD(X(n_s), X(n_{s+1})) \quad (1)$$

où  $X(n_s)$  représente les coordonnées du  $n_s^{eme}$  conformère, et  $RMSD(X(n_s), X(n_{s+1}))$  est le *RMSD*

entre les coordonnées des conformères successifs  $n_s$  et  $n_{s+1}$  dans le chemin.

Les critères (a) et (b) sont utilisés de façon à limiter la difficulté que constitue une recherche *CPR*, alors que le critère (c) a pour vocation de limiter les possibles détours faits par le chemin de transition dans le paysage conformationnel. Ces critères ont été appliqués manuellement, en fonction des résultats obtenus dans les précédents cycles d'affinement. Lors de l'exécution de l'algorithme *CPR*, les mouvements des atomes sont contraints dans un rayon de 0.5 Å.

Chaque cycle d'affinement a été composé d'un processus de réduction (i) et d'un processus de génération de conformères (ii), et a nécessité environ 300 heures CPU sur un processeur à 2 666 Mhz. Après 45 cycles d'affinement, le modèle électrostatique simple, par fonction diélectrique sigmoïdale, a été abandonné au profit du modèle *ACE2*, dans le but d'éviter la formation de vides à l'intérieur de la protéine. A l'issue de l'affinement, le chemin retenu pour le criblage virtuel était composé de 80 conformations, il avait une longueur curvilinéaire totale de 80 Å et son seuil de réduction d'énergie était de  $-29\ 000\ \text{kcal.mol}^{-1}$ .

## 2.2 Méthodes *In silico* de recherche d'inhibiteurs

### Identification des poches

Les poches ont été identifiées avec le module PocketFinder [14] du logiciel ICM [6], qui propose d'identifier des sites de fixation « druggable » à la surface des protéines. L'algorithme est décrit dans le Chapitre 2, section 2.2. Une recherche systématique de poches a été effectuée à la surface des structures *RX* de EF, 1K93 et 1K8T [77], des 80 conformations intermédiaires du chemin de transition déterminées comme décrit plus haut, et des conformations représentatives extraites des trajectoires *DM* du complexe EF-CaM (voir Chapitre 3, section Matériels et Méthodes) et de EF libre.

### Préparation du récepteur et du ligand

La protonation des histidines pour le récepteur a été assignée selon les recommandations de WHATIF [270]. Les structures 2D des 30 976 composés ont été téléchargées en format SDF depuis la *Chimiothèque Nationale* [4], composée de molécules très diverses. Le logiciel CORINA [219; 218] a servi à générer les structures 3D des composés, à leur ajouter des hydrogènes et à les convertir en format MOL2. 2 647 composés ont été retirés de la banque, suite à des échecs de construction 3D ou dans le cas de fragments multiples.

### Procédure de *docking*

Pour arrimer les ligands, les logiciels FlexX [206] et ICM [6] ont été utilisés. Comme expliqué au Chapitre 2, section 2.3, l'algorithme de *docking* de FlexX [207; 206] procède par re-construction incrémentale du ligand à l'intérieur du site potentiel de liaison, alors que l'algorithme de *docking* de ICM [7; 181; 6] consiste en une simulation de *Monte Carlo* (*MC*), suivie par une minimisation d'énergie locale et une sélection par critère de Metropolis [247; 8]. Pour FlexX, les paramètres par

défaut du récepteur ont été utilisés. Pour ICM, le facteur de rigueur a été fixé à 10, et 5 exécutions indépendantes de l'algorithme *MC* ont été réalisées pour chaque composé.

### Analyse bioinformatique des inhibiteurs

Les degrés de similarité de forme entre les composés ont été évalués avec le logiciel ROCS [5], basé sur l'idée que des molécules ont des formes similaires lorsque leurs volumes se recouvrent bien. L'algorithme de ROCS utilise des fonctions gaussiennes pour représenter le volume moléculaire et maximiser le recouvrement entre deux molécules. Le degré de recouvrement entre deux volumes moléculaires est évalué à partir de leur coefficient de Tanimoto [280]. Le score obtenu ici est dit *combo*, car il est égal à la somme du coefficient de Tanimoto et d'un score qui rend compte du degré de similarité chimique entre les deux molécules. Ces deux termes sont compris entre 0 et 1. Le degré de similarité chimique est évalué en utilisant un champ de force dit *de couleur*, qui définit six types de centres chimiques (donneur de liaison H, accepteur de liaison H, hydrophobe, anion, cation et cycle). Le champ de force par défaut, ImplicitMillsDean, a été utilisé, qui inclut un modèle simple de pKa.

Les composés ont été groupés en appliquant un algorithme de classification hiérarchique à la matrice de similarité. Les groupes sont formés par fusions successives, en utilisant le critère de lien complet (*complete linkage*), c'est-à-dire que la distance entre deux groupes est définie comme la distance maximale entre ses membres [116; 101]. Le logiciel R [124] a été utilisé pour la construction, la classification et la visualisation de la matrice.

L'analyse par empreintes d'interactions structurales [75; 177; 36] (*structural interaction fingerprints* ou *SIFt*) consiste en la caractérisation et la classification des poses prédites par le *docking* d'un ligand, basées sur les interactions établies avec la protéine ciblée. Chaque pose est représentée par un vecteur binaire *SIFt*, dans lequel la valeur 1 ou 0 de chaque composant indique la présence ou l'absence d'interaction(s) spécifique(s) entre le ligand et un résidu du site de liaison de la protéine. De façon à réduire le nombre de poses sujettes à l'analyse *SIFt*, une classification hiérarchique des poses de chaque ligand a été effectuée avec FlexX, en utilisant le critère de lien complet (*complete linkage*) avec un rayon de coupure de 3 Å [208; 79].

Les interactions ligand-protéine ont été extraites et classées par FlexX en 10 types : {donneur de liaison H}/{accepteur de liaison H}, {accepteur de liaison H}/{donneur de liaison H}, {centre phényl}/{cycle phényl}, {centre phényl}/{phényl-CH<sub>3</sub>}, {centre phényl}/{amide}, {cycle phényl}/{centre phényl}, {phényl-CH<sub>3</sub>}/{centre phényl}, {amide}/{centre phényl}, contact avec la chaîne principale de la protéine, contact avec la chaîne latérale de la protéine. Les vecteurs binaires de 10 bits de long, représentant le profil d'interaction de chacun des 16 résidus définissant la poche criblée, ont été ensuite concaténés pour obtenir un *SIFt* de 160 bits de long.

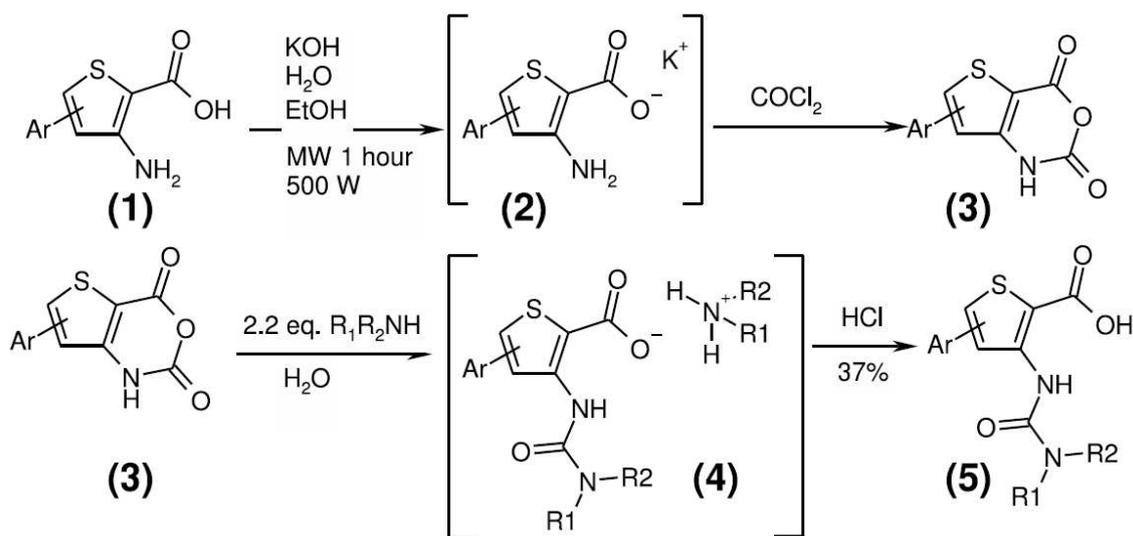
Le degré de similarité entre deux *SIFts* *A* et *B* a été évalué à partir de leur coefficient de Tanimoto  $T_c$  [280] :

$$T_c(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

où  $|A \cap B|$  est le nombre de bits ON communs aux deux A et B,  $|A \cup B|$  est le nombre de bits ON présents dans l'un des deux A ou B. Les empreintes d'interactions ont été regroupées à travers une classification hiérarchique [116]. La construction, la classification et la visualisation des *SIFts* ont été réalisées avec le logiciel R [124].

## 2.3 Méthodes expérimentales

### Ligands testés



**FIG. 1: Synthèse de la librairie des uréido-acides du thiophène.** Les amino-esters initiaux (1) ont été préparés selon la méthode de Kirsch pour les amino-esters substitués-4 [146; 209; 210] ou la méthode de Arnold-Vilsmeier-Haack pour les amino-esters substitués-5 [171; 117]. Leur hydrolyse alcaline sous des conditions de chauffage micro-ondes conduit à des intermédiaires amino-carboxylates non isolés (2). L'ajout de phosgène permet d'isoler les anhydrides (3), qui, placés dans une suspension d'eau avec 2.2 équivalents d'amine primaire ou secondaire, réagissent pour former des uréido-carboxylates de thiophène d'ammonium solubles (4). L'acidification ultérieure du mélange réactionnel permet de récupérer l'uréido-acide en solide précipité après filtration (5). L'excès d'amine est éliminé par simple lavage à l'eau, sous forme d'hydrochloride soluble.

La série d'inhibiteurs identifiés est issue de la chimiothèque du *Centre d'Etudes et de Recherche sur le Médicament de Normandie* (CERMN). Le CERMN a développé un programme de synthèse parallèle [91; 92] dont l'objectif est de produire de nouveaux dérivés du thiophène ayant des propriétés pharmacologiques potentielles. Dans ce cadre, une librairie de 1 140 uréido-acides du thiophène a été synthétisée (Fig. 1) par réaction d'un ensemble de 60 amines primaires ou secondaires avec 19 thiéno[3,2-d]- ou thiéno[2,3-d][1,3]oxazine-2,4-diones. Les composés ont été obtenus par une simple stratégie combinatoire en phase de solution sur une échelle de 200-400 mg. Les rendements dépassent les 70% et le pourcentage de pureté est supérieur à 80%.

### Essai biochimique des composés

Les composés ont été dissous dans du diméthylsulfoxyde (DMSO), à une concentration de stockage de 10 mM. La conservation est effectuée à -20 °C. Les protéines EF, CyaA et CaM ont été purifiées comme décrit dans de précédentes études [226; 269; 142]. La pyrophosphatase inorganique (PPase) de *E. coli* a été achetée à Sigma (ref I5907) et le kit Pi-ALS à Innova Biosciences (Innova Biosciences Ltd, Babraham Hall, Cambridge UK).

### Essai enzymatique de l'adénylate cyclase

Afin de caractériser le potentiel inhibiteur des différents composés envers les toxines bactériennes adénylates cyclases (AC), un test colorimétrique de mesure d'activité adénylate cyclase, en format micro-plaque, a été conçu. Les AC convertissent l'ATP en AMPc, relarguant du pyrophosphate inorganique (PPi). La molécule PPi peut ensuite être clivée par la PPase (par ex. de *E. coli*, ref Sigma I5907) en deux ions phosphates (Pi). Ceux-ci sont détectables par colorimétrie, grâce au changement d'absorbance de la teinte vert malachite – lue à 595 nm, en présence de complexes phospho-molybdates.

Les tests ont été effectués comme suit : EF de *Bacillus anthracis* a été dilué dans la solution tampon des AC (20 mM de Hepes-Na [pH 7.5], 15 mM de MgCl<sub>2</sub>, 0.2 mM de CaCl<sub>2</sub> et 0.01% de Tween-20 (v/v)), complétée par 4 unités/ml de PPase (1 unité de PPase hydrolyse 1  $\mu$ mol de PPi par minute), jusqu'à une concentration finale de 0.2-0.5 nM; 25  $\mu$ L ont été distribués dans les 96 puits d'une plaque de microtitration. Les composés (10  $\mu$ L, provenant d'une solution concentrée 5 x dans un essai tampon) ont été ajoutés jusqu'aux concentrations finales indiquées (i.e. 100, 10 ou 1  $\mu$ M), pour 20 minutes d'incubation à 22 °C. Puis, CaM (5  $\mu$ L, provenant d'une solution de stockage à 20  $\mu$ M dans un essai tampon) a été ajoutée jusqu'à une concentration finale de 2  $\mu$ M, et les mélanges ont ensuite été incubés pendant 10 minutes à 30 °C. Le déclenchement de la réaction enzymatique s'est fait par ajout d'ATP (10  $\mu$ L pour une concentration finale de 2 mM) et la micro-plaque a été incubée à 30 °C, avec agitation. Dix échantillons de 1  $\mu$ L ont ensuite été prélevés de chaque puits, à différents temps d'incubation (de 3 à 15 minutes) et transférés dans une seconde plaque de microtitration contenant, dans chaque puits, 100  $\mu$ L d'un mélange de Pi-ALS – 10  $\mu$ L de H<sub>2</sub>O, 80  $\mu$ L de ré-agent Pi ColorLock ALS et 10  $\mu$ L de solution Accelerator, le ré-agent et la solution étant fournis dans le kit Pi-ALS. Les conditions acides de ce mélange Pi-ALS ont permis de stopper immédiatement la réaction enzymatique. Après 5 minutes d'incubation à température ambiante, 10  $\mu$ L de solution stabilisatrice (du kit Pi-ALS) ont été ajoutés pour éviter la rupture non-enzymatique du substrat phosphorylé, en conditions acides (selon les instructions du kit). Après 30-60 minutes d'incubation supplémentaires à température ambiante, les densités optiques à 595 nm (OD<sub>595</sub>) ont été enregistrées avec un lecteur de micro-plaques (Tecan, Lyon, France). Dans certaines expériences (voir la section Résultats), CaM a été mélangée à EF *avant* l'ajout des composés.

Une courbe standard a été réalisée en parallèle, en ajoutant des concentrations connues de Pi dans le mélange Pi-ALS. Elle a servi à convertir les valeurs OD<sub>595</sub> en moles de PPI produites. L'activité enzymatique a été calculée à partir de la vitesse initiale de synthèse de PPI (dans les conditions décrites plus haut, l'accumulation de PPI était linéaire dans le temps). En absence de composé, l'activité spécifique de EF s'est trouvée comprise entre 1 et 2 mmol de PPI par minute et par mg de protéine (100% d'activité). Lorsque CaM était omise dans le mélange du test, l'activité de synthèse de PPI était en dessous du seuil de détection, i.e. < 1 μmol de PPI par minute et par mg de protéine (0% d'activité). Les tests avec CyaA de *Bordetella pertussis* ont été réalisés de manière similaire. La concentration finale de CyaA était de 0.1-0.2 nM. L'activité spécifique de CyaA était de 2.5 mmol de PPI par minute et par mg de protéine en présence de CaM et en-dessous du seuil de détection, i.e. < 1 μmol PPI par minute et par mg de protéine en l'absence de CaM.

### Test par anisotropie de fluorescence à l'état stationnaire

Le peptide de 22 résidus P<sub>233-254</sub>, correspondant à la séquence principale de reconnaissance de CaM chez CyaA [158], et qui englobe les résidus 233-254 (LDRERIDLLWKIARAGARSAVG, avec N<sub>ter</sub> acétylé), a été synthétisé par Genosphere Biotechnologies (Paris, France). La fixation calcium-dépendante du peptide P<sub>233-254</sub> à CaM a été suivie en mesurant l'anisotropie de fluorescence à l'état stationnaire de son unique résidu Tryptophane (CaM n'a pas de Tryptophane). Le test a été mené avec un spectro-fluorimètre FP-6200 (Jasco, Japan) dans un porteur de cellule à thermostat intégré (effet Peltier), en utilisant une cellule de quartz de longueur de chemin 1 cm (101.QS de Hellma). La longueur d'onde d'excitation a été fixée à 295 nm et les spectres d'émission ont été enregistrés à 25 °C, de 300 à 400 nm. Des bandes passantes de 5 et 10 nm ont été respectivement utilisées pour les rayons d'excitation et d'émission. L'anisotropie du résidu Trp du peptide P<sub>233-254</sub> (2 μM) a été mesurée avec et sans CaM (2 μM) et/ou G2 (10 μM) dans une solution tampon A (20 mM de Hepes [pH 7.4] et 150 mM de NaCl), complétée avec 0.2 mM de CaCl<sub>2</sub> ou 2 mM d'EGTA. Les expériences d'anisotropie de fluorescence à l'état stationnaire ont été réalisées avec un faisceau polarisé verticalement (V) et horizontalement (H), en utilisant des polariseurs FDP-223 aux ouvertures d'excitation (x) et d'émission (m). Les valeurs d'anisotropie  $r$ , mesurées à une longueur d'onde de  $345 \pm 1$  nm, ont été calculées comme suit :

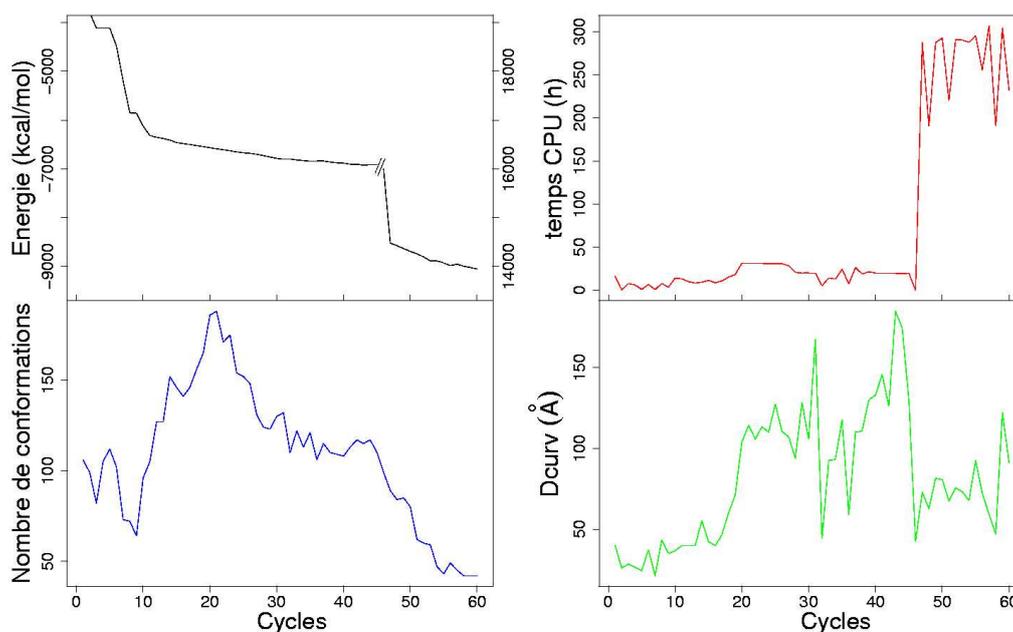
$$r = \frac{V_x V_m - \frac{H_x V_m}{H_x H_m} V_x H_m}{V_x V_m + 2 \frac{H_x V_m}{H_x H_m} V_x H_m}. \quad (3)$$

## 3 Résultats

### 3.1 Chemin de transition conformationnelle

Un chemin de transition adiabatique entre les conformations inactive et active de EF a été calculé par Arnaud Blondel. L'évolution de l'affinement du chemin de transition de EF est décrite

par les paramètres suivants (Fig. 2) : le temps CPU requis pour un cycle d'affinement, l'énergie du système, le nombre de conformères dans le chemin et la longueur curvilinéaire (voir Eq. 1 dans la section Matériels et Méthodes) du chemin. Les profils de temps CPU (*courbe rouge*) et d'énergie (*courbe noire*) montrent une transition autour du cycle n°45, quand le modèle de solvant passe de la fonction diélectrique sigmoïdale au potentiel *ACE2* [221].

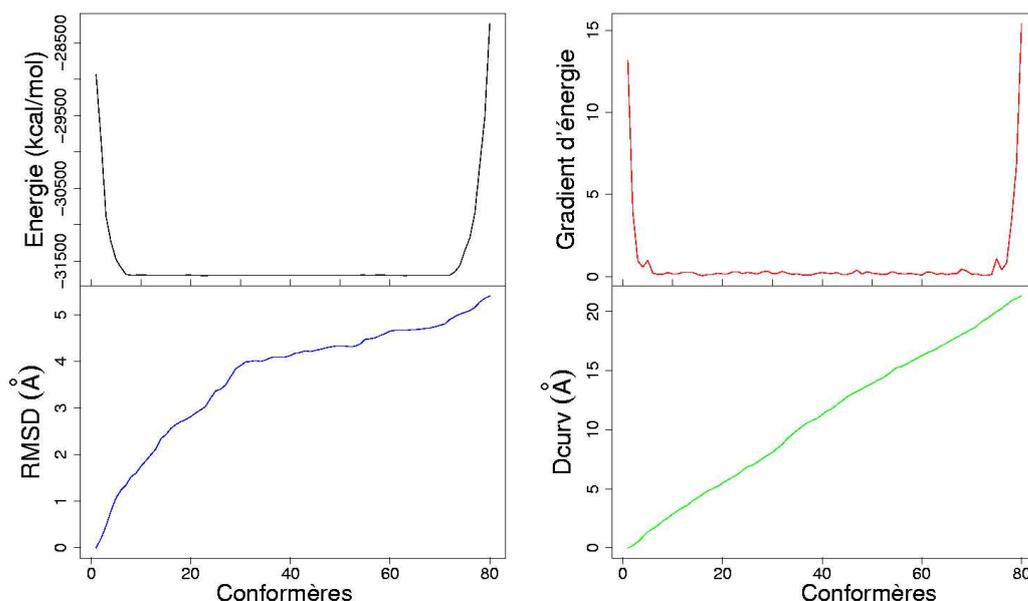


**FIG. 2: Historique de l'affinement du chemin de transition.** L'énergie maximum en kcal/mol (*courbe noire, en haut à gauche*), le temps CPU en heures (*courbe rouge, en haut à droite*), le nombre de conformères après réduction (*courbe verte, en bas à gauche*) et la longueur curvilinéaire  $D_{curv}$  en Å avant réduction (*courbe bleue, en bas à droite*) sont donnés pour un cycle d'affinement. Les échelles à gauche et à droite du graphe de l'énergie s'appliquent respectivement aux portions gauche et droite de la courbe. Ce choix de deux échelles est fait pour éviter que les variations de la courbe ne soient écrasées à cause de l'écart entre les valeurs correspondant au modèle de solvant le plus simple (de 1 à 45) et au modèle de solvant *ACE2* (46 à 60) respectivement.

L'énergie (*courbe noire*) présente une diminution systématique au cours des cycles d'affinement, et ce pour chacun des niveaux d'approximation du solvant. Le nombre de conformères dans le chemin (*courbe verte*) varie au cours des 40 premiers cycles, pour ensuite diminuer et se stabiliser autour d'une valeur minimum. La longueur curvilinéaire  $D_{curv}$  tend également à se stabiliser à partir du cycle 46, dans une moindre mesure cependant. A l'image de cette stabilisation, le chemin semble donc converger.

Les variations des caractéristiques suivantes ont été enregistrées le long du chemin final, obtenu à l'issue du dernier cycle d'affinement (Fig. 3) : l'énergie, le gradient d'énergie, le *RMSD* entre les coordonnées des conformations successives et du point initial, la longueur curvilinéaire. Comme l'indiquent les profils plats de l'énergie et du gradient d'énergie, l'ensemble des conformations obtenues dans ce chemin final ne présente aucun clash stérique. Cependant, les conformères de la

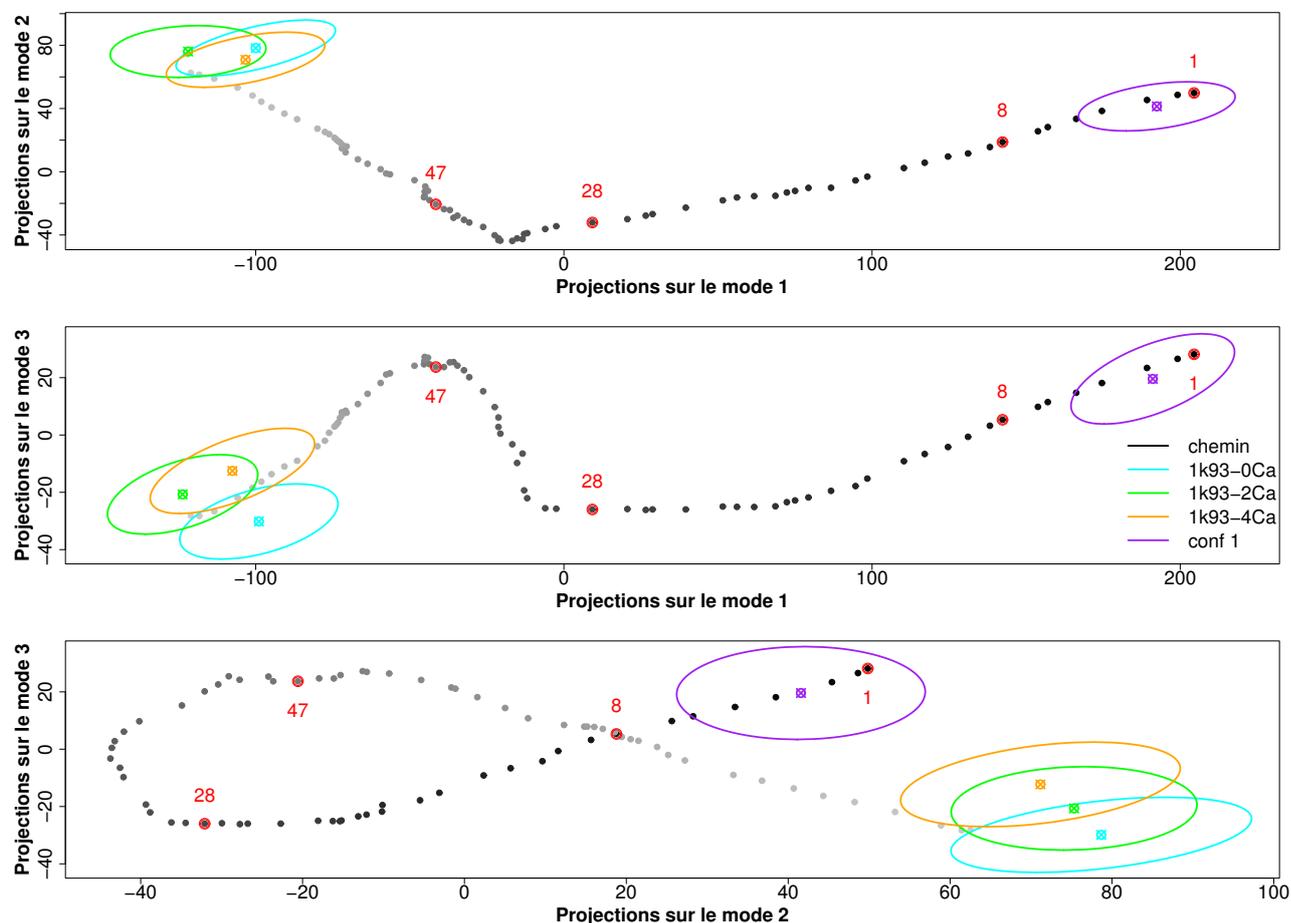
transition ont des caractéristiques énergétiques différentes de celles observées dans les structures  $RX$ , de sorte que la valeur d'énergie augmente beaucoup aux extrémités du chemin. Ceci peut venir de la nature des méthodes d'affinement du chemin, qui, basées sur la minimisation de l'enthalpie, ne prennent pas en compte l'entropie du système. En l'absence de compensation entropie/enthalpie, l'optimisation du chemin conduit à des valeurs d'énergie excessivement faibles.



**FIG. 3: Caractéristiques du chemin de transition final.** L'énergie en kcal/mol (courbe noire, en haut à gauche), le gradient d'énergie (courbe rouge en haut à droite), le  $RMSD$  par rapport au point initial en Å (courbe bleue, en bas à gauche) et la longueur curvilinéaire  $D_{curv}$  en Å (courbe verte, en bas à droite) sont donnés.

Une analyse en composantes principales ( $ACP$ ) a été effectuée sur les coordonnées atomiques des 80 conformères composant le chemin final. Les conformations du chemin, ainsi que les conformations échantillonnées dans les simulations  $DM$  de EF et de EF-CaM, ont été projetées sur les trois premiers modes  $ACP$  du chemin (Fig. 4). Les projections des trajectoires  $DM$  sont très regroupées dans l'espace des modes  $ACP$ , à proximité du point final du chemin de transition. Ceci prouve que l'information acquise grâce au chemin n'aurait pas pu être obtenue à partir des trajectoires à l'équilibre. En particulier, les conformations *conf 8*, *conf 28* et *conf 47*, utilisées comme contrôles négatifs pour filtrer les composés criblés (voir plus bas), sont localisées dans des régions de l'espace non accessibles par les conformations  $DM$ . Les projections de la trajectoire de EF libre (*en violet*), quant à elles, sont situées à proximité du point initial du chemin. On peut noter que le centre de projection de la trajectoire 1k93-2Ca (*en vert*) est le plus proche du point final du chemin dans les trois panels : les ellipses de projection des trajectoires 1k93-0Ca (*en cyan*) et 1k93-4Ca (*en orange*) dévient dans le même sens selon le premier mode, ou en sens opposé selon le troisième mode.

Il apparaît donc que le chemin produit ici constitue une description plausible de la transition conformationnelle entre les formes fermée et ouverte de EF. Il est aussi cohérent avec l'espace conformationnel échantillonné par les simulations *DM* à l'équilibre.

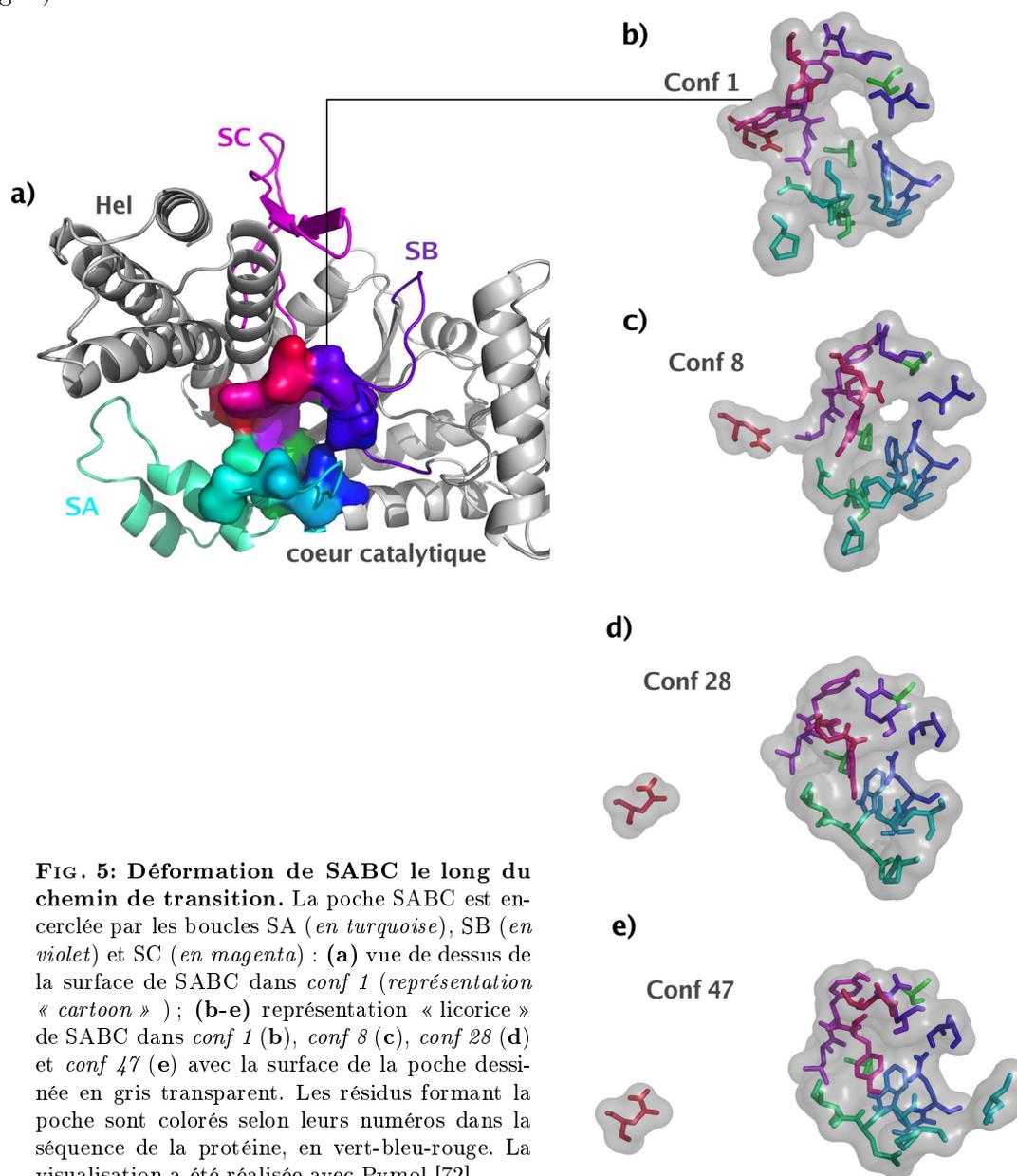


**FIG. 4: Projections sur les trois premiers modes ACP du chemin de transition.** Les projections des conformations successives du chemin, depuis le point initial jusqu'au point final, sont colorées de gris clair à noir. Les conformères *conf 8*, *conf 28* et *conf 47* du chemin sont indiqués en rouge. Les projections des simulations *DM* du complexe EF-CaM chargé de zéro, deux, ou quatre ions  $\text{Ca}^{2+}$  et de EF libre sont colorées en cyan, vert, orange et violet, respectivement. Les projections des simulations *DM* sont représentées par une ellipsoïde dont le centre correspond à la projection de la conformation *DM* moyenne (*carré plein*), et dont les dimensions indiquent la dispersion des projections selon les deux modes considérés.

### 3.2 Criblage virtuel et sélection des ligands

Une recherche systématique de poches a été effectuée à la surface de EF, le long du chemin de transition, avec le module PocketFinder d'ICM [14]. Dix sites de liaison potentiels ont été identifiés, avec des volumes allant de 100 à 860 Å<sup>3</sup>, parmi lesquels le site catalytique. A la surface de *conf 1*, outre la poche catalytique de plus grande taille, une cavité de 463 Å<sup>3</sup> a été détectée à l'interface des boucles SA (résidus 502-551), SB (résidus 578-591) et SC (résidus 630-659). Cette poche, également visible dans la structure 1K8T, est dénotée SABC. Les trois boucles SA, SB et

SC jouent un rôle critique dans l'activation de EF, soit à travers des contacts directs avec CaM, soit à travers la stabilisation du site catalytique [77; 228]. De ce fait, elles subissent de larges réarrangements conformationnels durant la transition, ce qui modifie la forme de la poche SABC (Fig. 5).

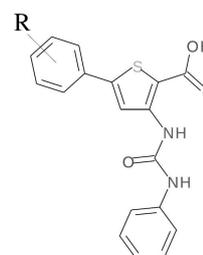


Tandis que la cavité de la poche est bien formée dans les premières étapes du chemin de transition (Fig. 5b, *conf 1*), elle disparaît progressivement en avançant dans la transition (Fig. 5c, 5d et 5e), à cause de : (i) l'espace croissant entre le domaine Hel (résidus 660-767) et la région catalytique (résidus 294-622), (ii) les réarrangements du squelette des boucles SA et SC, (iii) la réorientation des chaînes latérales des résidus en contact avec CaM dans le complexe. Ces observations permettent de formuler l'hypothèse qu'une petite molécule capable de se fixer dans la cavité de la poche, et ainsi de maintenir sa forme, pourrait piéger EF en conformation fermée,

empêchant de fait son activation par CaM. Par ailleurs, SABC n'est pas particulièrement étroite et présente un bon équilibre entre les résidus chargés ou polaires et les résidus non polaires. Ces pré-dispositions en font un bon site de liaison candidat pour un *docking* avec des logiciels comme FlexX [206]. SABC a donc été retenue pour le criblage virtuel de la *Chimiothèque Nationale*, dans le but de sélectionner des inhibiteurs potentiels de l'activité de EF.

Le criblage virtuel de la *Chimiothèque Nationale* [4], comprenant alors 28 329 molécules, a été réalisé avec FlexX [206], en ciblant la poche SABC, définie par les résidus : A496, P499, I538, E539, P542, S544, S550, W552, Q553, T579, Q581, L625, Y626, Y627, N629, N709. La conformation initiale *conf 1* du chemin de transition a été utilisée comme conformation cible. Un premier élagage de la banque de molécules a été effectué, en retirant les composés ayant des scores moins bons qu'un seuil  $\theta$ , calculé comme la moyenne  $\mu$  des meilleurs scores de tous les composés moins la déviation standard  $\sigma$  :  $\theta = \mu - \sigma = -12.6 - 4.7 = -17.3$ . Parmi les 3 649 composés retenus, les meilleurs 1% ont été sélectionnés comme touches potentielles (voir Table I et Annexe A.1).

**TAB. I: Touches sélectionnées par criblage virtuel et validées expérimentalement.** La formule chimique commune des six composés est donnée au dessus du tableau, à droite. Dans le tableau, les codes, les identifiants, les noms, la composition du radical R, les scores de *docking* et les taux d'inhibition de l'activité de EF à 100  $\mu$ M sont reportés. Les meilleurs scores obtenus avec FlexX sont donnés pour les conformations *conf 1*, *conf 8*, *conf 28*, *conf 47* du chemin et la structure 1K8T. Pour la conformation initiale cible *conf 1*, le score FlexX moyen avec déviation standard est également donné (deuxième ligne), ainsi que l'intervalle de scores obtenus avec ICM (troisième ligne). Les scores de *docking* des composés sélectionnés initialement, mais finalement non retenus pour les tests enzymatiques, sont reportés en Annexe A.1



	F3	F4	G2	G3	G4	G5
CEsr- <sup>a</sup>	3341000	3342000	4451000	4452000	4454000	4456000
AB- <sup>b</sup>	00003633	00003634	00004720	00004721	00004723	00004725
Nom	TUAF	TUAOCH <sub>3</sub>	TUAdiCl	TUABr	TUA	TUACl
R	4-F	4-OCH <sub>3</sub>	3,4-diCl	4-Br	-	4-Cl
<i>conf 1</i> <sup>FlexX</sup>	-27.66	-27.09	-27.31	-28.12	-27.51	-27.86
	-22.94 ± 2.17	-24.62 ± 1.76	-23.89 ± 2.06	-22.73 ± 2.51	-22.78 ± 2.25	-22.88 ± 2.29
<i>conf 1</i> <sup>ICM</sup>	[-11.46 ; -23.53]	[-19.85 ; -20.68]	[-16.30 ; -18.89]	[-20.60 ; -20.92]	[-19.00 ; -19.45]	[-19.39 ; -20.53]
1K8T <sup>FlexX</sup>	-30.74	-29.43	-31.16	-30.74	-30.74	-30.74
<i>conf 8</i> <sup>FlexX</sup>	-19.25	-17.85	-19.38	-19.25	-20.53	-19.25
<i>conf 28</i> <sup>FlexX</sup>	-14.45	-13.44	-12.90	-12.62	-13.28	-12.24
<i>conf 47</i> <sup>FlexX</sup>	-15.82	-14.76	-16.05	-15.82	-15.89	-15.82
Inhibition	102 ± 2	96 ± 7	98 ± 6	101 ± 1	96 ± 7	98 ± 3

<sup>a</sup> Identifiant dans la chimiothèque du CERMN.

<sup>b</sup> Identifiant dans la *Chimiothèque Nationale* (septembre 2007).

Les conformations intermédiaires *conf 8*, *conf 28* et *conf 47* (Fig. 5b-d), tirées du chemin de transition, ont ensuite été utilisées comme contrôles négatifs pour filtrer les faux positifs. Les composés capables de se lier aux états déformés de la poche SABC, observés à la surface de ces conformations, aussi bien qu'à la poche bien formée trouvée dans la conformation initiale *conf*

1, ont été considérés comme faux positifs. En effet, la stratégie adoptée consistait à chercher des ligands capables de maintenir la forme initiale de la poche. Les touches potentielles ont été retenues si et seulement si leur score moyen minoré de leur déviation standard sur 20 poses de *docking* à *conf 1* était meilleur que le meilleur score de *docking* à chacune des conformations *conf 8*, *conf 28* ou *conf 47*. De cette façon, trois composés ont été éliminés de l'ensemble (voir Table I et Annexe A.1). Un second filtre a été appliqué à travers le *re-docking* des ligands sélectionnés, avec le logiciel ICM [6], qui utilise une représentation plus précise de la physique du système basée sur un champ de force de mécanique moléculaire [247; 8]. Pour chaque ligand considéré, l'intervalle des scores ICM s'est trouvée assez éloigné du score FlexX (voir Table I et Annexe A.1). Les composés ayant des scores ICM positifs ou supérieurs à -10 kcal/mol ont été retirés. Globalement, 28 touches potentielles ont finalement été sélectionnées et 18 ont pu être testées *in vitro*.

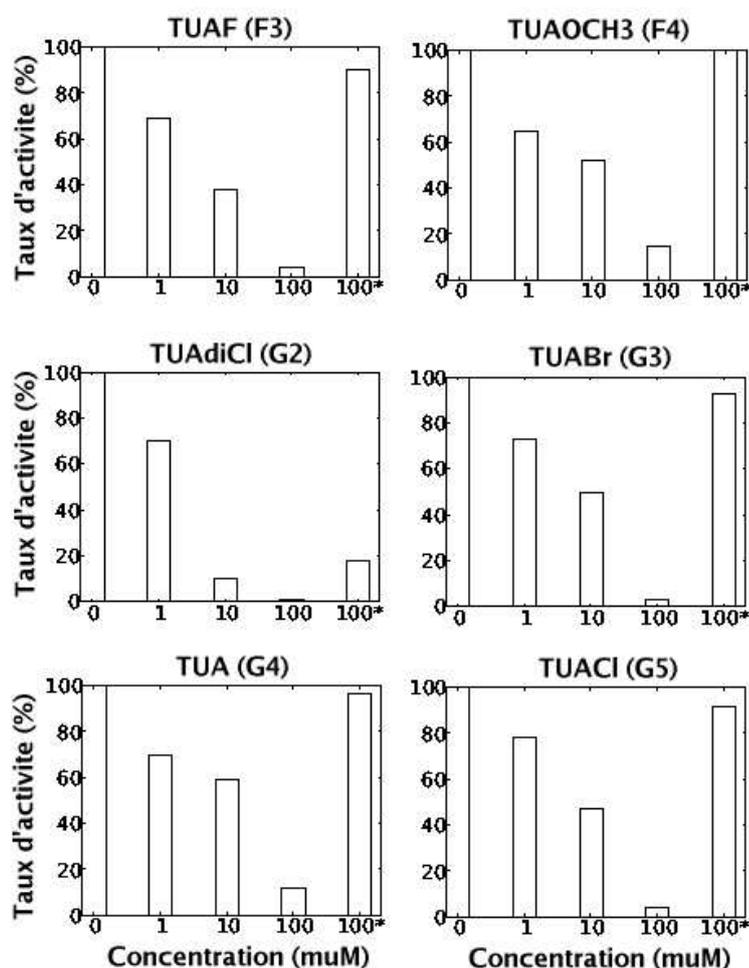
### 3.3 Caractérisation *in vitro* des inhibiteurs putatifs

Le potentiel d'inhibition des 18 composés sélectionnés et disponibles, envers l'activité enzymatique de EF, a été caractérisé expérimentalement, par Christophe Goncalves, en collaboration avec l'unité Biochimie des Interactions Macromoléculaires (Daniel Ladant, Institut Pasteur). Lors d'une première série de tests, EF (0.5 nM) a été pré-incubé avec 100  $\mu$ M du composé considéré, puis CaM a été ajoutée (2  $\mu$ M). Après un temps additionnel de 10 minutes d'incubation, pour permettre la formation éventuelle du complexe actif EF-CaM, l'activité enzymatique a été déterminée à l'aide d'un test colorimétrique permettant de mesurer la quantité de pyrophosphate (PPi), produit re-largué lors de la conversion d'ATP en AMPc. Comme montré Figure 6 et Table I, six composés inhibent totalement l'activité de EF dans ces conditions, tous les six partageant la même unité uréido-acide du thiophène (TUA). Cependant, testés dans de plus faibles concentrations, les composés voient leur pouvoir inhibiteur diminuer : à 10  $\mu$ M, seul le composés TUAdiCl inhibe l'activité de EF à plus de 80%, avec un  $IC_{50}$  global de 2-3  $\mu$ M. De fait, TUAdiCl se révèle être un inhibiteur puissant de EF, tandis que les autres uréido-acides du thiophène sont moins efficaces.

Dans une seconde série de tests, EF a été pré-incubé avec CaM *avant* l'ajout des ligands sélectionnés. Du fait de l'affinité de CaM pour EF [226; 108], aucun pouvoir inhibiteur sur le complexe actif pré-formé n'était attendu, de la part de molécules sélectionnées pour leur capacité à empêcher l'association entre les deux protéines. Si cinq des six touches sont en effet incapables d'inhiber l'activité de EF dans ces conditions (Fig. 6), le composé TUAdiCl provoque quant à lui une diminution drastique de l'activité enzymatique du complexe EF-CaM.

Ces résultats expérimentaux valident donc une série de molécules comme inhibiteurs de l'activation de EF par CaM, parmi les ligands sélectionnés *in silico* pour cibler la poche SABC. Le membre le plus puissant de cette famille de molécules, TUAdiCl, est même capable d'inhiber EF lié à CaM, ce qui suggère un mode de liaison multiple ou différent de celui d'un simple antagoniste de la fixation de CaM.

**FIG. 6:** Taux d'activité de EF observés expérimentalement, sous l'effet des uréido-acides du thiophène sélectionnés. Pour chaque composé, les barres représentent : (1) le contrôle, (2-4) les pourcentages d'activité de EF, à des concentrations de composés de 1, 10 et 100  $\mu\text{M}$ , dans le cas d'une pré-incubation *avant* ajout de CaM, (5) le pourcentage d'activité de EF, à une concentration de composés de 100  $\mu\text{M}$ , dans le cas où ceux-ci sont ajoutés *après* formation du complexe EF-CaM (100\*).



Pour explorer plus avant la sélectivité de ces six molécules actives, une troisième série de tests a été réalisée afin de mesurer leur pouvoir inhibiteur sur la toxine adénylate cyclase CyaA de *Bordetella pertussis*, l'agent responsable de la coqueluche. CyaA est aussi activée par CaM et partage 25% d'identité de séquence avec EF. De plus, la structure de son site catalytique lié à CaM révèle un haut degré d'homologie structurale avec le complexe EF-CaM, bien que leurs surfaces de contact respectives avec CaM soient distinctes [108; 107].

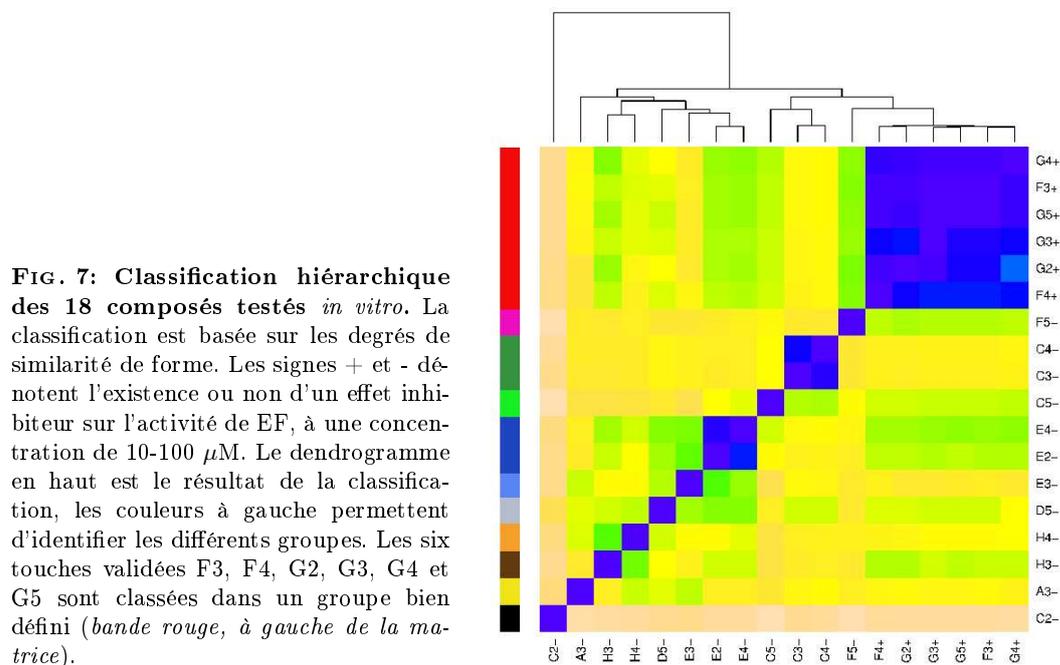
Les six molécules inhibent l'activité de CyaA à 100  $\mu\text{M}$  (non montré). Toutefois, à une concentration 10 fois plus faible, aucune inhibition significative n'a été mise en évidence, dénotant une capacité limitée de ces composés à agir comme antagonistes de la fixation de CaM sur CyaA.

Un certain nombre de petites molécules hydrophobes sont capables de bloquer l'interaction de CaM avec des enzymes cibles [263]. Ainsi, l'hypothèse d'une fixation directe des inhibiteurs uréido-acides du thiophène à CaM, plutôt qu'à EF, a été testée par Johanna Karst, de l'unité Biochimie des Interactions Macromoléculaires (Daniel Ladant, Institut Pasteur). Pour cela, la capacité de liaison de CaM à un peptide de 22 acides aminés, P<sub>233-254</sub>, correspondant à la séquence principale de liaison de CyaA à CaM [158], a été mesurée en présence du plus puissant inhibiteur TUAdiCl.

La liaison a été suivie grâce aux variations d’anisotropie de fluorescence  $r$  de l’unique Tryptophane du peptide P<sub>233–254</sub> (CaM n’a pas de Tryptophane). La valeur d’anisotropie du peptide libre (2  $\mu\text{M}$ ) vaut 0.03 (en présence de calcium), et augmente à 0.09 lors de l’ajout d’une concentration équimolaire de CaM. Cette augmentation révèle la fixation du peptide à CaM. Ces valeurs ne varient pas en présence de 10  $\mu\text{M}$  de TUAdiCl ( $r = 0.027$  et  $0.095$ , sans et avec CaM). Ceci indique que TUAdiCl, à une concentration suffisante pour inhiber quasi-complètement l’activité de EF, n’a pas d’effet sur la capacité d’association de CaM à un peptide spécifique de la liaison CyaA-CaM. Le pouvoir inhibiteur de TUAdiCl envers l’activité de EF résulte donc d’une interaction directe avec l’enzyme EF, et non pas d’une inhibition de son activateur CaM.

Faisant suite à la détermination de la structure de EF [77], plusieurs composés ont été identifiés expérimentalement comme inhibiteurs de l’activité de EF [233; 227; 168; 50], dont les  $\text{IC}_{50}$  se situent dans l’intervalle 0.027-100  $\mu\text{M}$ . Le composé TUAdiCl apparaît donc comme un inhibiteur au moins aussi efficace, avec un  $\text{IC}_{50}$  de quelques  $\mu\text{M}$ . La comparaison des formules chimiques montre que les composés décrits ici représentent une nouvelle famille chimique de molécules inhibitrices de EF. On peut remarquer que, dans tous les cas reportés, sauf un [50], les ligands actifs sur EF le sont également sur CyaA, comme c’est le cas ici.

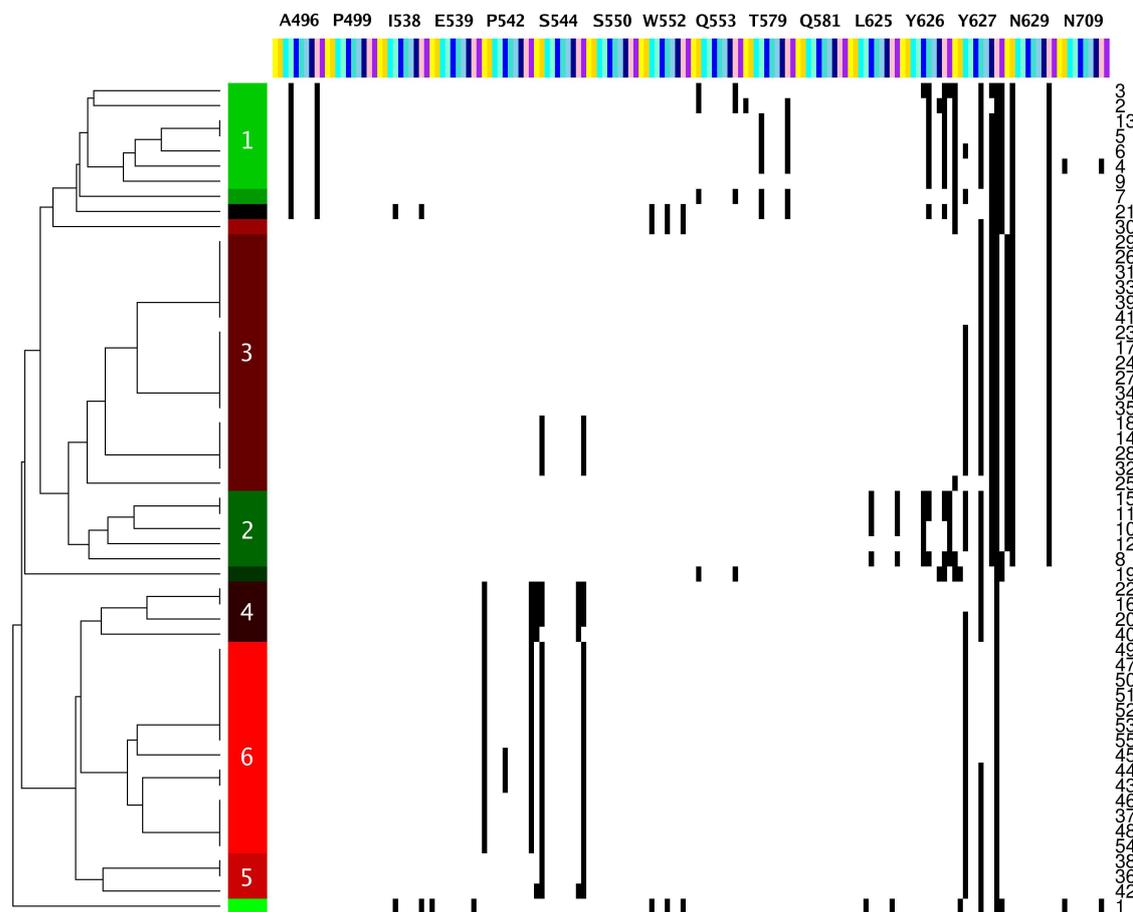
### 3.4 Analyse bioinformatique des inhibiteurs identifiés



**FIG. 7: Classification hiérarchique des 18 composés testés *in vitro*.** La classification est basée sur les degrés de similarité de forme. Les signes + et - dénotent l’existence ou non d’un effet inhibiteur sur l’activité de EF, à une concentration de 10-100  $\mu\text{M}$ . Le dendrogramme en haut est le résultat de la classification, les couleurs à gauche permettent d’identifier les différents groupes. Les six touches validées F3, F4, G2, G3, G4 et G5 sont classées dans un groupe bien défini (*bande rouge, à gauche de la matrice*).

Les tests expérimentaux ont permis de valider la stratégie *in silico* de sélections des ligands. Afin de mieux comprendre le mode d’action des six inhibiteurs, une analyse bioinformatique des composés et de leurs modes de liaison prédits a été menée. Les degrés de similarité de formes des ligands sélectionnés ont été évalués avec le logiciel ROCS [5], et les matrices obtenues ont

été soumises à un algorithme de classification hiérarchique (Fig. 7). Comme énoncé plus haut, les six inhibiteurs partagent un échafaudage chimique commun et ne varient entre eux que par la nature d'un substituant – majoritairement halogène, attaché à un groupe phényle (Table I). Les autres composés testés sont majoritairement isolés et leurs formes diffèrent des inhibiteurs validés, à l'exception peut-être du composé F5, de sorte qu'aucune tendance globale ne se dessine.

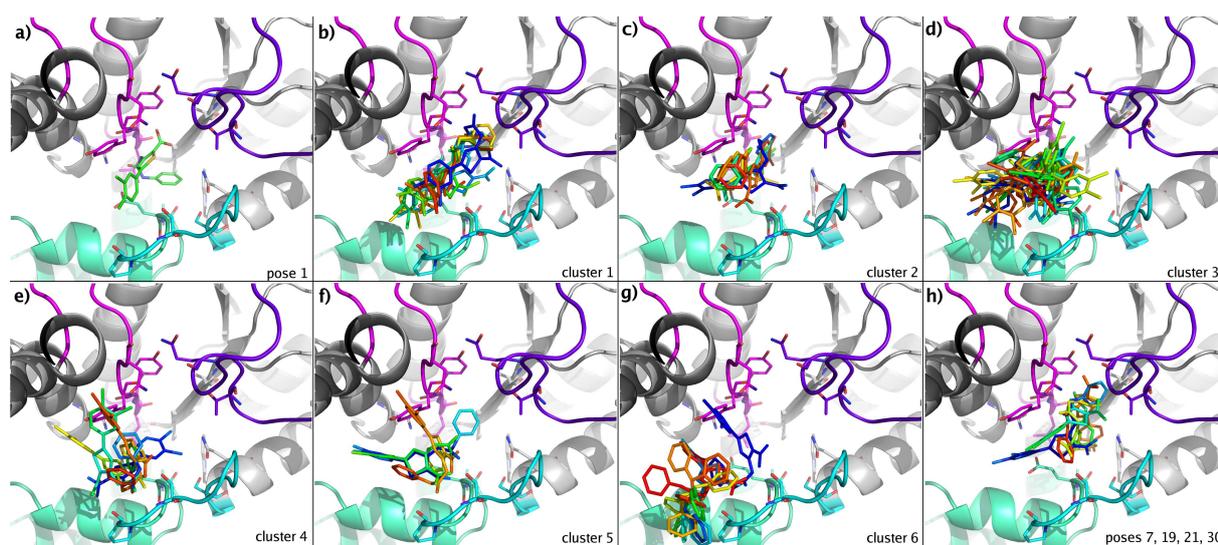


**FIG. 8: Classification hiérarchique des poses de docking de TUAdiCl (G2).** La classification est basée sur les *SIFts* des 55 poses de TUAdiCl pour *conf 1*. Chaque *SIFt* est représenté par une ligne dans la matrice, avec le rang de la pose correspondante inscrit sur la droite. L'existence d'un type d'interaction entre TUAdiCl et un résidu de la poche SABC – parmi A496, P499, I538, E539, P542, S544, S550, W552, Q553, T579, Q581, L625, Y626, Y627, N629 et N709 – est indiquée par un point noir. Au sein de chaque profil de résidu, chaque résidu étant indiqué en haut de la carte, on compte 10 bits différents, représentant 10 types d'interactions ligand-résidu différents : {donneur de liaison H}/{accepteur de liaison H} (*en jaune*), {accepteur de liaison H}/{donneur de liaison H} (*en doré*), {centre phényle}/{cycle phényle} (*en cyan*), {centre phényle}/{phényle-CH<sub>3</sub>} (*en marine*), {centre phényle}/{amide} (*en bleu*), {cycle phényle}/{centre phényle} (*en turquoise*), {phényle-CH<sub>3</sub>}/{centre phényle} (*en bleu ciel*), {amide}/{centre phényle} (*en bleu foncé*), contact avec la chaîne principale de la protéine (*en violet*), contact avec la chaîne latérale de la protéine (*en rose*). Le dendrogramme résultant de la classification hiérarchique est montré sur la gauche et le panel de couleurs associé indique les groupes et les singletons identifiés. Le code couleur de vert-noir-rouge est fonction du rang de pose moyen pour chaque groupe de poses.

Les poses prédites par *docking* des six touches ont été caractérisées en détails à travers une analyse des empreintes d'interactions structurales (*SIFt*) [75], décrite dans la section Matériels

et Méthodes. Globalement, les matrices *SIFT* obtenues pour les six touches sont similaires et les interactions clés avec EF restent les mêmes. Par conséquent, seuls les résultats *SIFT* pour le plus puissant inhibiteur TUAdiCl (G2) sont décrits ici (Fig. 8 et 9). Une classification hiérarchique des poses de *docking*, basée sur leur *SIFT*, dévoile six groupes et cinq singletons. La meilleure pose est isolée dans un singleton (Fig. 8 et Fig. 9a), en cohérence avec la dégradation brutale des scores de  $-27.31$  à  $-23.65$  entre la première et la seconde poses. Les cinq autres touches présentent aussi cette caractéristique (non montré).

Dans une large majorité de poses, TUAdiCl interagit avec la chaîne latérale de Tyr-627, principalement à travers des contacts hydrophobes (Fig. 8). On peut aussi noter la présence de liaisons hydrogènes avec la chaîne latérale de Asn-629 et, dans une moindre mesure, des interactions hydrophobes avec Tyr-626 et Leu-625. Comme mentionné précédemment, Leu-625, Tyr-626, Tyr-627 et Asn-629 de la boucle SC, sont en contact avec CaM dans le complexe EF-CaM. Ainsi, ils participent à la distortion de la poche SABC le long du chemin de transition, à travers la ré-orientation de leur chaînes latérales. Par conséquent, l'existence de ces interactions suggère que TUAdiCl, de même que les cinq autres touches, est capable d'empêcher directement la fixation de CaM à EF.



**FIG. 9:** Poses de *docking* de G2 classées à travers une analyse *SIFT* : (a) pose 1, (b) groupe 1, (c) groupe 2, (d) groupe 3, (e) groupe 4, (f) groupe 5, (g) groupe 6, (h) superposition des poses singletons 7, 19, 21 et 30. Les résidus définissant la poche SABC sont dessinés en bâtons et colorés selon leurs positions dans la structure de la protéine : SA (en turquoise), SB (en violet) et SC (en magenta). Le ligand TUAdiCl est aussi dessiné en bâtons et en couleurs multiples.

TUAdiCl établit aussi des contacts hydrophobes avec les chaînes latérales de Ala-496 et Thr-579 dans le groupe 1, comprenant les 2<sup>ème</sup> à 7<sup>ème</sup> meilleures poses, et dans les poses 7 et 21. Ces deux résidus se situent à l'opposé de Leu-625, Tyr-626, Tyr-627 et Asn-629 dans la poche, à des distances  $C\alpha-C\alpha$  de 8-12 Å. En établissant des contacts avec les deux côtés de la poche dans le groupe 1, TUAdiCl apparaît donc capable de se placer de façon à épouser et maintenir la forme de la cavité (Fig. 9b).

Le groupe 1 et tous les singletons (poses 1, 7, 19, 21 et 30) établissent aussi des interactions

avec les chaînes latérales de Trp-552 et Gln-553, également situés à l'opposé des résidus de SC dans la poche. Les poses 1, 7, 19, 21 et 30 peuvent donc être vues comme des modes de liaison alternatifs, où TUAdiCl est encore capable de maintenir la forme de la poche (Fig. 9a et 9h).

Les poses des groupes restants ne sont clairement pas aussi bien positionnées à l'intérieur de la poche (Fig. 9c-g). Dans les groupes 2 et 3 (Fig. 9c-d), TUAdiCl paraît seulement attaché à SC, avec une diversité de poses particulièrement élevée dans le groupe 3. Dans les groupes 4-6 (Fig. 9e-g), les groupes 5 et 6 étant composés principalement des poses les moins bonnes, TUAdiCl est partiellement à l'extérieur de la poche et établit des interactions avec Pro-542 et Ser-544 de SA (Fig. 8). De telles interactions ne semblent donc pas déterminantes pour une fixation efficace du ligand, dans la perspective de maintenir la forme de la poche. De plus, elles ne sont pas observées dans les matrices *SIFt* des cinq autres touches.

Globalement, l'analyse par *SIFts* des poses prédites par FlexX suggère un mode de liaison commun des six inhibiteurs à la poche SABC. Des interactions clés se retrouvent dans une large majorité de poses, qui impliquent les résidus de SC en contact avec CaM dans le complexe EF-CaM. L'établissement d'interaction avec des résidus situés sur le côté opposé de la poche dans un certain nombre de poses parmi les meilleures soutient l'hypothèse que les six inhibiteurs sont capables de maintenir la forme de la poche.

## 4 Discussion

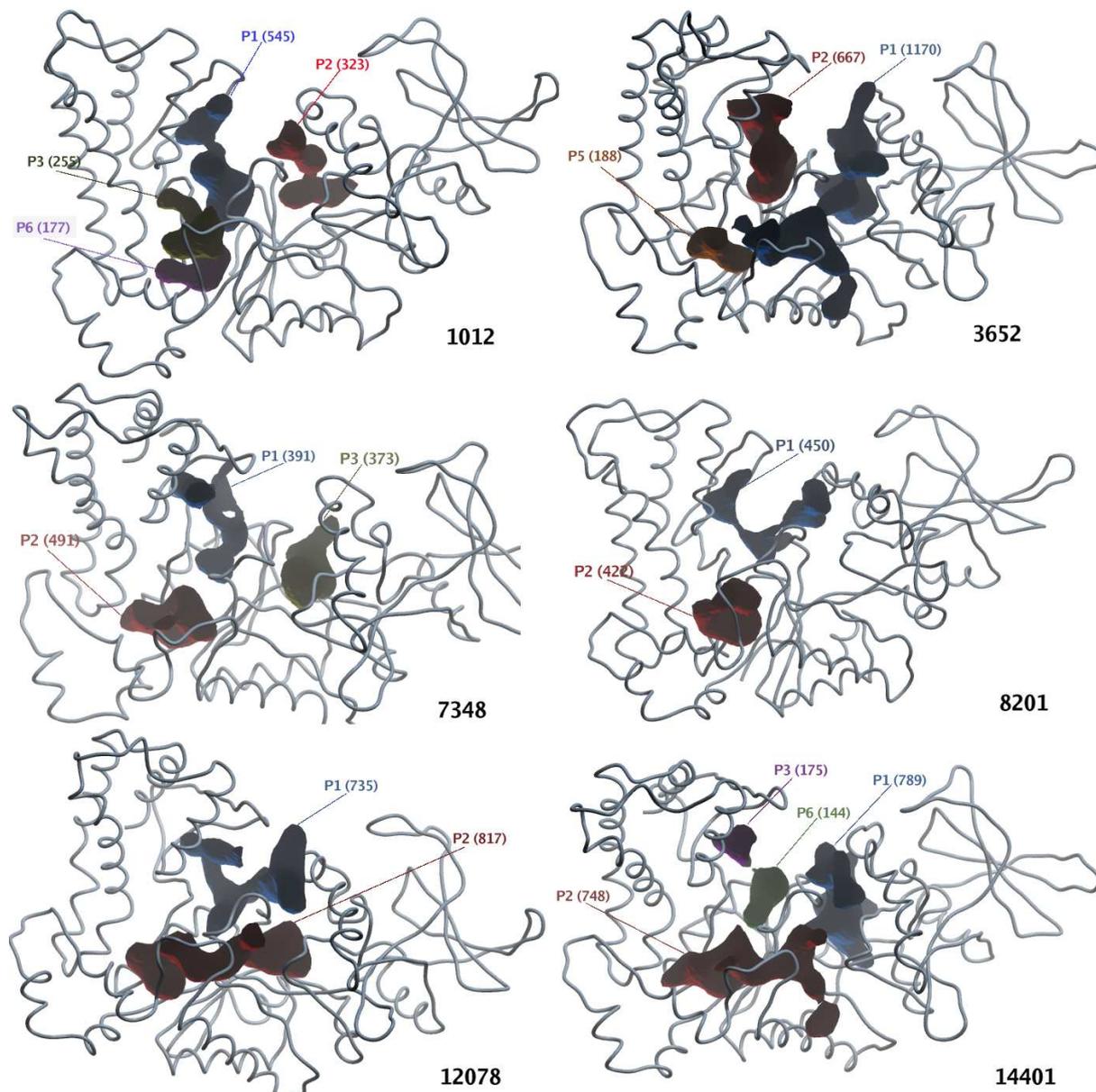
### 4.1 La détermination d'un chemin de transition est essentielle

L'originalité de l'approche développée ici tient dans la détermination par modélisation moléculaire d'un chemin de transition conformationnelle plausible et de son utilisation dans un contexte de recherche d'inhibiteurs. Les conformations intermédiaires obtenues, grâce à une application extensive de l'algorithme *CPR*, sont énergétiquement plausibles et s'inscrivent dans un chemin minimal adiabatique. Elle fournissent des informations quant aux aspects structuraux de la transition conformationnelle de EF, qui peuvent être exploitées, en l'occurrence pour l'identification d'un site de liaison potentiel et la sélection de ligands.

Bien que le criblage virtuel ait principalement ciblé des conformations issues de la première moitié du chemin, la détermination d'un chemin complet était essentielle pour le choix de la poche. En effet, il était nécessaire de suivre l'évolution de la surface biomoléculaire de la cible EF pour analyser l'évolution des poches, et en particulier de la poche SABC. Cibler une poche formée uniquement dans des conformations proches de la conformation fermée initiale est plus efficace que cibler une poche formée dans les dernières étapes du chemin, où le processus de fixation de CaM est facilité.

## 4.2 La poche SABC est stable au regard de la dynamique interne de EF

Comme décrit dans la section Matériels et Méthodes, la boucle SB manquante dans la structure *RX* de EF fermé (entrée PDB : 1K8T) [77] a été modélisée de façon à obtenir le point initial du chemin de transition. La conformation de SB a simplement été prise de la structure *RX* de EF-CaM (entrée PDB : 1K93) [77].



**FIG. 10:** Poches détectées sur les conformations représentatives extraites des simulations *DM* de **EF libre**. La position de chaque conformation dans la trajectoire est indiquée en bas à droite (ps). Les poches sont dénommées P1 à P6, et leurs volumes sont donnés en  $\text{\AA}^3$ .

Cette modélisation a évidemment un impact important sur la forme de la cavité de SABC, le

long du chemin de transition, et n'est bien sûr pas neutre au regard de la définition de la poche. Pour voir si des conformations alternatives de SB influenceraient la poche SABC, une simulation *DM* de 15 ns de la conformation initiale *conf 1* du chemin de transition a été réalisée. Le *RMSD* de la structure se stabilise au bout de 1 ns autour de 3 Å. Le *RMSD* des coordonnées du squelette de la boucle SB varie autour de 1.78 Å, avec une déviation standard de 0.24 Å et une valeur maximale de 2.42 Å. Les fluctuations ne dépassent pas 2.2 Å pour ces résidus et la structure secondaire de la boucle est quasi-complètement stable au bout de 5 ns (non montré).

La méthode de Lyman & Zuckerman [174] (voir Chapitre 3) a été appliquée pour extraire six conformations représentatives de la trajectoire, avec un rayon de coupure de 2 Å. Une recherche de poches a été menée à la surface de ces conformations, prises à 1012 ps, 3652 ps, 7348 ps, 8201 ps, 12078 ps et 14401 ps. Les poches localisées dans la même région que la poche SABC ou à l'interface entre la poche SABC et la poche catalytique sont montrées Figure 10. Dans toutes les conformations, au moins une poche est détectée qui contient plus de la moitié des seize résidus de SABC (Table II).

Conformation (ps)	Numéro de poche	Volume (Å <sup>3</sup> )	Résidus en commun avec SABC
1012	p1	545	(3/23 = 13%) A496, Y626, N629
	p2	323	(0/9 = 0%)
	p3	256	(7/13 = 54%) E539, P542, S544, Q553, Q581, Y627, N629
	p6	177	(11/14 = 79%) A496, P499, I538, E539, W552, Q553, T579, Q581, L625, Y627, N709
3652	p1	1170	(4/33 = 12%) P542, S544, Q553, Y627
	p2	667	(3/25 = 12%) Q553, Y627, N629
	p5	188	(8/13 = 62%) P499, I538, E539, Q553, Q581, L625, Y627, N709
7348	p1	391	(4/25 = 16%) A496, Q581, Y627, N629
	p2	491	(11/21 = 52%) P499, E539, P542, W552, Q553, T579, Q581, L625, Y627, N629, N709
	p3	373	(0/16 = 0%)
8201	p1	450	(0/26 = 0%)
	p2	422	(11/17 = 65%) A496, P499, I538, P542, W552, Q553, T579, L625, Y627, N709
12078	p1	735	(2/27 = 7%) Q581, N629
	p2	817	(8/29 = 28%) P499, I538, P542, W552, Q553, L625, Y627, N709
14401	p1	789	(0/26 = 0%)
	p2	748	(9/28 = 32%) P499, I538, P542, W552, T579, Q581, L625, Y627, N709
	p3	175	(0/10 = 0%)
	p6	144	(0/10 = 0%)

**TAB. II: Poches détectées à proximité de SABC.** Les poches sont détectées à la surface de conformations représentatives extraites à 1012 ps, 3652 ps, 7348 ps, 8201 ps, 12078 ps et 14401 ps de la trajectoire *DM* de *conf1*. Le nom de la poche, donné par le module PocketFinder d'ICM, son volume, le nombre relatif et le nom des résidus en commun avec SABC sont reportés.

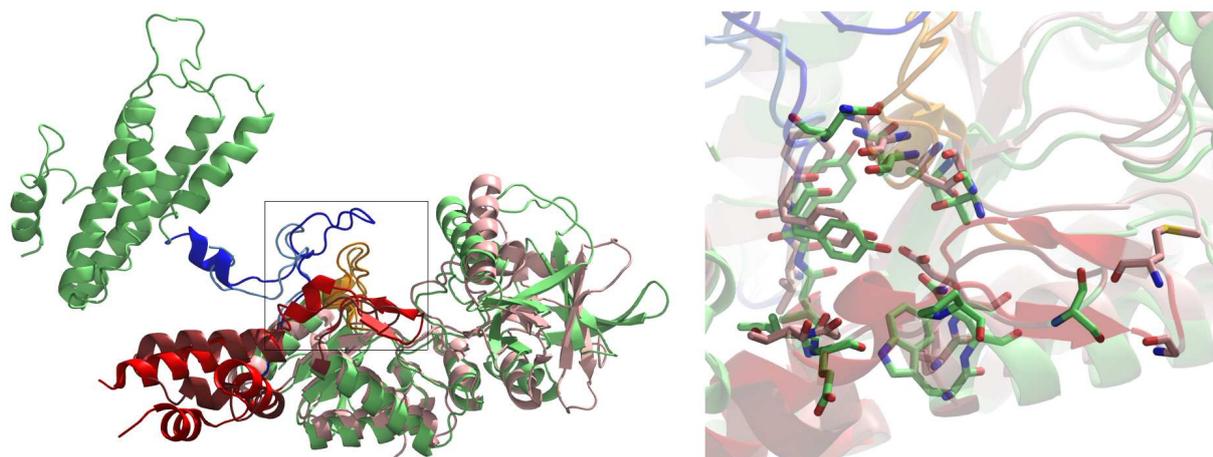
Les conformations extraites du premier tiers de la trajectoire, à 1012 ps et 3652 ps, présentent de petites poches de 177 Å<sup>3</sup> et 188 Å<sup>3</sup>, composées à plus de 60% de résidus de SABC. Les conformations extraites du second tiers, à 7348 ps et 8201 ps, présentent des poches de 491 Å<sup>3</sup> et 422 Å<sup>3</sup>, dont les formes ressemblent à la poche SABC et qui sont composées des résidus de SABC pour plus de 50%. Enfin, les conformations du dernier tiers, à 12078 ps et 14401 ps, présentent de larges poches de 817 Å<sup>3</sup> et 748 Å<sup>3</sup>, s'étendant de l'interface entre SA, SB et SC jusqu'à la poche catalytique. Ces larges poches sont encore définies par des résidus de SABC à hauteur de 30%. Trois types de configurations sont donc balayées le long de la simulation *DM* de *conf 1* et

SABC, telle qu'elle a été définie pour le criblage virtuel, apparaît comme une poche consensuelle enlaçant les résidus susceptibles de former des cavités à l'interface de SA, SB et SC dans EF fermé. La présence de tous les résidus de SABC est justifiée dans la définition de la poche, car tous se retrouvent dans au moins une poche de chaque conformation représentative.

### 4.3 La comparaison des structures de CyaA et de EF suggère un mode de liaison similaire sur les deux enzymes.

Les six touches sélectionnées *in silico* inhibent l'activité catalytique de CyaA, lorsqu'elles sont ajoutées avant CaM, à une concentration de 100  $\mu\text{M}$ . Ceci suggère qu'elles sont capables de piéger CyaA dans une conformation inactive, et ainsi d'empêcher la formation du complexe actif CyaA-CaM. Aucune structure du coeur catalytique de CyaA libre n'est actuellement disponible. Par conséquent, l'existence d'une poche similaire à SABC de EF à la surface de CyaA ne peut être affirmée. Cependant, l'alignement des structures des deux enzymes en complexe avec CaM fournit des éléments de réponse quant à l'effet des inhibiteurs.

L'alignement entre CyaA et EF, proposé par Guo et al. [108], montre que les hélices HF-HG-HH-HH', la boucle T300-K312 et la queue C-terminale de CyaA correspondent aux boucles SA, SB et SC de EF, respectivement. Ces régions de CyaA sont susceptibles de jouer des rôles équivalents à leurs homologues chez EF, dans l'interaction avec CaM et dans la stabilisation du site catalytique [107; 108]. De plus, ces régions correspondantes de CyaA et de EF adoptent des conformations similaires dans les complexes CyaA-(C-CaM) [108] et EF-CaM [77] (Fig. 11), suggérant qu'elles pourraient subir des réarrangements conformationnels substantiels similaires lors de la fixation de CaM.



**FIG. 11: Superposition de EF et CyaA.** Les structures de EF (1K93) et de CyaA (1YRT), en complexe avec CaM, sont représentées en lime et rose respectivement. La boucle SA est soulignée en rouge foncé et rouge, SB en marron et orange, SC en bleu et cyan, pour EF et CyaA respectivement. Sur la droite, un zoom de la région SABC est présenté, et les résidus de EF définissant SABC sont dessinés en bâtons.

A la surface de CyaA liée à C-CaM (1YRT), une petite cavité est détectée à l'interface des hélices HF-HG-HH-HH', de la boucle T300-K312 et de la queue C-terminale. Toutefois, les résidus

qui la composent ne correspondent pas directement aux résidus de la poche SABC de EF, dans l'alignement des deux protéines : le *RMSD* entre les coordonnées des résidus correspondants de EF et CyaA vaut 7.55 Å. La surface de CyaA, en conformation liée à C-CaM, ne présente donc pas de cavité équivalente à SABC de EF. On peut néanmoins constater que neuf résidus de la poche SABC de EF possèdent des homologues identiques ou similaires chez CyaA, sur un total de seize, d'après l'alignement des structures des deux enzymes (Table III). Considérant que la structure de CyaA seule n'est pas assez stable pour être cristallisée, on peut émettre l'hypothèse que les hélices HF-HG-HH-HH', la boucle T300-K312 et la queue C-terminale sont très flexibles, et donc susceptibles d'adopter, en solution, des conformations proches de EF libre, qui seraient favorisées par la fixation des inhibiteurs identifiés.

**TAB. III: Résidus de la poche SABC de EF et leurs correspondants chez CyaA.** L'alignement des séquences des deux protéines est tiré de [108]. Une étoile  $\star$  indique des résidus identiques, un triangle  $\triangle$  des résidus similaires et un cercle  $\circ$  des résidus partageant une forme commune.

EF		CyaA	
pré-SA	A496 $\star$	A193 $\star$	pré-F
pré-SA	P499 $\star$	P196 $\star$	pré-F
SA	I538	G254	pré-F
SA	E539	T255	épingle B
SA	P542	G265	épingle B
SA	S544	M267	épingle B
CA	S550 $\triangle$	T273 $\triangle$	CA
CA	W552 $\triangle$	F275 $\triangle$	CA
CA	Q553 $\circ$	E276 $\circ$	CA
SB	T579 $\star$	T300 $\star$	boucle
SB	Q581 $\star$	Q302 $\star$	boucle
SC	L625 $\triangle$	V343 $\triangle$	queue C-ter
SC	Y626 $\circ$	F344 $\circ$	queue C-ter
SC	Y627 $\star$	Y345 $\star$	queue C-ter
SC	N629 $\star$	N347 $\star$	queue C-ter
Hel	N709		

En l'absence d'informations structurales précises, il n'est pas aisé de déterminer l'action des ligands à un niveau atomique. Cependant, on peut se servir de la comparaison entre les structures de CyaA et EF pour élucider le site de fixation préférentiel des ligands. Le Tableau IV donne les détails de toutes les poches détectées à la surface de *conf 1* de EF, ainsi que leur degré de similarité avec les groupes de résidus correspondants dans l'alignement avec CyaA. Outre les deux premières poches, à savoir la poche catalytique et SABC, qui présentent des hauts degrés de similarité, les rapports de similarité sont plutôt faibles, excepté pour la petite poche n°6 avec deux tiers des résidus conservés. La poche n°6 est située à l'interface des boucles SB et SC, séparée de la poche SABC seulement par la chaîne latérale de Gln-581. Une telle délimitation est une conséquence de la modélisation de SB dans *conf 1*. Autoriser une certaine flexibilité à cette boucle

peut mener à la fusion des deux poches, comme on l'a vu précédemment dans la trajectoire *DM* de *conf 1*. Le fait que, à l'exception de SABC et du site catalytique, seule une petite cavité trouvée dans le voisinage de celle-là présente une forte similarité avec CyaA, renforce l'hypothèse que les inhibiteurs potentiels se fixent sur la région ciblée, à l'interface de SA, SB et SC.

numéro de poche	volume ( $\text{\AA}^3$ )	nombre de résidus	nombre d'homologues chez CyaA			
			total	conservés	similaires	de formes similaires
1 (catal)	871	23	23	17	2	1
2 (SABC)	463	16	15	6	3	2
3	223	12	8	1	1	0
4	210	12	12	1	1	1
5	183	10	4	2	0	0
6	168	9	8	6	0	0
7	156	12	12	1	0	1
8	135	9	3	0	0	0
9	123	10	0	0	0	0
10	110	8	8	3	0	0
11	103	11	9	1	0	0

**TAB. IV: Description des poches détectées à la surface de *conf 1* de EF, et comparaison avec CyaA.** Pour chaque poche détectée à la surface de *conf 1* de EF, le numéro, le volume, le nombre de résidus, le nombre total d'homologues chez CyaA (absence de délétion/insertion dans l'alignement), le nombre d'homologues identiques, similaires ou de formes similaires, sont reportés. Des résidus sont considérés comme similaires s'ils appartiennent à la même famille physico-chimique. Ils ont seulement des formes similaires lorsqu'ils partagent les même contraintes stériques.

#### 4.4 Les touches validées ont une action allostérique sur l'interaction EF-CaM

On a vu que TUAdiCl est capable d'inhiber le complexe EF-CaM pré-formé. Ce résultat étonnant implique un mode de liaison alternatif ou multiple. En effet, la poche SABC a justement été choisie en raison de sa présence limitée le long du chemin de transition. Elle n'est visible à la surface d'aucune des structures cristallographiques de EF-CaM [228; 77] ni des dernières conformations du chemin. Une recherche de poches menée à la surface des structures *RX* permet cependant de détecter une petite cavité (221  $\text{\AA}$ ), que l'on qualifiera de résiduelle, à l'interface de SA, SB et SC dans la structure 1XFX, qui a un résidu en commun avec la définition de SABC (non montré).

Pour analyser l'influence de la dynamique interne de la protéine sur les cavités à sa surface, une recherche systématique de poches a été effectuée sur des conformations représentatives extraites des trajectoires *DM* de 15 ns du complexe EF-CaM, 1k93-0Ca, 1k93-2Ca et 1k93-4Ca. Des poches résiduelles à l'interface des trois boucles ont été trouvées à la surface de presque toutes les conformations, avec des volumes plus larges, de 104 à 326  $\text{\AA}^3$  (Table V) et un nombre croissant de résidus en commun avec la définition de SABC. La conformation 9335 de 1k93-0Ca présente même une poche résiduelle d'un volume supérieur à 300  $\text{\AA}^3$ , qui partage le plus grand nombre (6, 50%) de résidus en commun avec la poche SABC (Table V). Elle est notamment composée des

résidus Leu-625 et Tyr-627, dont l'importance pour la liaison de CaM et dans l'analyse *SIFT* a été mise en évidence, mais aussi de Ile-538, Trp-552, Gln-553 et Gln-581.

Conformation (ps)		Volume ( $\text{\AA}^3$ )	Résidus en commun avec SABC
0Ca	5906	192	(3/7 = 43%) I538,Q581,Y627
	9335	326	(6/12 = 50%) I538,W552,Q553,Q581,L625,Y627
	12150	202	(3/10 = 30%) E539,Q581,N629
2Ca	5016	104	(1/7 = 14%) Q581
	7950	-	-
	12262	144	(0/6 = 0%)
4Ca	3529	257	(0/10 = 0%)
	4983	126	(2/7 = 29%) Q290,N338
	5471	191	(2/8 = 25%) E248,Q290
	11084	-	-
	13515	-	-

**TAB. V: Poches identifiées à l'interface de SA, SB et SC dans le complexe EF-CaM.** Les poches sont détectées à l'interface de SA, SB et SC à la surface de EF dans des conformations représentatives extraites des simulations *DM* du complexe, 1k93-0Ca, 1k93-2Ca et 1k93-4Ca. Le volume de chaque poche, le nombre relatif et le nom des résidus en commun avec SABC sont reportés.

Comme observé dans le Chapitre 3, le complexe EF-CaM sans calcium, obtenu en retirant les deux ions  $\text{Ca}^{2+}$  de la forme prépondérante du complexe en solution, est déstabilisé dans les trajectoires *DM* : CaM adopte une conformation moins adaptée à EF et l'effondrement du site catalytique de EF est observé. Cet effondrement est caractérisé par une réduction notable des dimensions du site et de sa *SASA*. Les valeurs de *SASA*, calculées pour les conformations représentatives des trois trajectoires de EF-CaM, sont reportées dans le tableau VI. Les valeurs des conformations issues de 1k93-2Ca et 1k93-4Ca sont proches de celles calculées dans les structures *RX* 1K90 [77] et 1XFV [228], avec une tendance à la hausse, alors que les valeurs de 1k93-0Ca présentent une diminution de 40%. Des distances caractéristiques de la fixation de l'ATP dans les structures *RX* ont également été calculées. Elles ont été définies à partir de l'analyse des interactions entre l'ATP et les résidus du site catalytique et sont indicatrices de la longueur occupée par l'ATP. On peut noter que globalement les valeurs de la distance K346.N $\zeta$ -K372.N $\zeta$  sont plus grandes dans 1k93-0Ca, comparées aux autres trajectoires et aux structures *RX*, tandis que les valeurs de la distance R329.C $\zeta$ -S354.O $\gamma$  sont plus petites. Cette tendance est encore plus marquée pour la conformation 9335 de 1k93-0Ca.

Par ailleurs, les six ligands validés ont été arrimés avec FlexX à la poche résiduelle de chaque conformation, tandis que l'ATP a été arrimé au site catalytique. Globalement, les scores de ces touches sont dans le même intervalle que les scores obtenus durant l'arrimage sur la poche SABC dans *conf 1* (non montré). Au contraire, les scores de *docking* de l'ATP sur les conformations *DM* sont moins bons que ceux obtenus pour la structure *RX* 1K93 [77], utilisée comme point de départ des trajectoires *DM*. En particulier, le site catalytique de la conformation 9335, extraite de la trajectoire à zéro  $\text{Ca}^{2+}$  fixe l'ATP avec le plus mauvais score -19.78.

		<i>SASA</i> catal. (Å <sup>2</sup> )	Dist. catal. (Å)	
			K346.Nζ - K372.Nζ	R329.Cζ - S354.Oγ
Struct. <i>RX</i>	1K90	714	7.5 ± 0.3	12.1 ± 0.1
	1XFV	892	7.0 ± 0.0	13.0 ± 0.0
Traj. <i>DM</i>	5906 <sup>0</sup>	475	13.4	10.0
	9335 <sup>0</sup>	504	14.5	9.6
	12150 <sup>0</sup>	422	12.1	11.9
	5016 <sup>2</sup>	919	10.5	12.3
	7950 <sup>2</sup>	916	11.7	12.9
	12262 <sup>2</sup>	947	10.9	14.9
	3529 <sup>4</sup>	791	7.5	13.0
	4983 <sup>4</sup>	886	8.0	12.8
	5471 <sup>4</sup>	845	11.6	13.6
	11084 <sup>4</sup>	824	12.3	14.7
13515 <sup>4</sup>	910	8.9	14.8	

**TAB. VI: Caractéristiques du site catalytique dans le complexe EF-CaM.** Les distances et *SASA* caractéristiques sont calculées sur les structures *RX*, 1K90 [77] et 1XFV [228], et sur les conformations représentatives extraites des trajectoires *DM* du complexe. Le nombre d'ions Ca<sup>2+</sup> est indiqué en exposant : <sup>0</sup> renvoie à la trajectoire 1k93-0Ca, <sup>2</sup> renvoie à 1k93-2Ca et <sup>4</sup> renvoie à 1k93-4Ca.

La conformation 9335 de 1k93-0Ca donne donc un exemple où la présence d'une poche raisonnablement large à l'interface de SA, SB et SC est associée à l'effondrement et à la distortion du site catalytique. Il est clair que la déstabilisation observée pour 1k93-0Ca est cohérente avec l'impossibilité expérimentale [255] d'observer la formation du complexe EF-CaM sans calcium. Néanmoins, on peut émettre l'hypothèse que la capacité de TUAdiCl à se fixer sur EF-CaM, et ainsi à inhiber l'activité de EF, pourrait s'expliquer par l'existence de conformations similaires à la conformation 9335, soit à l'équilibre en solution, ou obtenues par ajustement induit (*induced fit*) lors de l'interaction avec le composé. L'influence mutuelle mise ici en évidence entre l'espace délimité par les boucles SA, SB et SC, et la poche catalytique, suggère bien un effet allostérique de la liaison des ligands validés à EF.

Les tests enzymatiques menés ici ont permis de valider l'approche *in silico*, et en particulier la capacité des logiciels de *docking* FlexX et ICM à discriminer des composés actifs parmi une banque de molécule. Les scores prédits sont cependant différents d'un logiciel à l'autre et n'ont pu être clairement corrélés aux affinités mesurées expérimentalement. Un certain nombre de questions demeurent quant au mode d'action des inhibiteurs. L'alignement des deux toxines EF et CyaA révèle des correspondances et apporte des arguments en faveur de la réalité de l'interaction ligands-SABC. L'analyse des simulations *DM* du complexe permet également d'apporter des éléments de réponse, en mettant en évidence une influence mutuelle entre la poche catalytique et la cavité à l'interface de SA, SB et SC, s'agissant du volume et de la configuration, et donc, on peut le supposer, de la capacité de liaison du substrat ou des inhibiteurs. Un telle influence apporte des arguments pour un possible effet allostérique induit sur le site catalytique, à travers la fixation d'un petit composé à la poche SABC. Cependant, elle a été révélée dans une forme perturbée

du complexe, qui ne correspond à aucune structure expérimentale. Les conformations tirées de la trajectoire *DM* fournissent donc seulement des indications sur la plasticité conformationnelle de la protéine en solution.

## 5 Conclusion

Une approche originale, combinant méthodes computationnelles et expérimentales, où l'activité enzymatique est inhibée via le ciblage d'une interaction entre biomolécules, a été élaborée. Une telle approche, qui implique le criblage d'une poche allostérique plutôt que du site actif de l'enzyme EF, a permis d'identifier une nouvelle famille de molécules inhibitrices de l'activité de la toxine EF, dont le membre le plus puissant a une affinité au moins aussi bonne que les inhibiteurs connus. Le choix de la poche ciblée a été rendu possible par la détermination d'un chemin de transition, résultat de calculs de modélisation moléculaire. Ce chemin n'existe cependant que dans le contexte d'un champ de force particulier, qui comporte des imperfections et approximations, et une forme de validation expérimentale serait souhaitable. Dans ce sens, des expériences de fluorescence [168] ou par *RMN* pourraient attester l'existence de certaines de ces conformations, en suivant des distances ou liaisons hydrogènes transitoires.

En l'absence de données structurales précises, il est difficile de déterminer le mode d'action des inhibiteurs à l'échelle atomique. La résolution d'une structure co-cristallisée permettrait de localiser le site de fixation des inhibiteurs et de véritablement valider l'approche. De plus, une étude plus approfondie de la dynamique de la région catalytique de CyaA, qui reste peu connue, pourrait être envisagée, avec la possibilité d'un transfert de la connaissance acquise sur EF et des méthodologies développées ici.

Ce travail ouvre une nouvelle voie pour la recherche d'inhibiteurs par criblage virtuel, étendant la diversité des cibles biologiques potentielles et la variabilité des inhibiteurs potentiels. En effet, le choix d'une poche allostérique rend possible le ciblage de protéines dont le site actif n'est pas accessible, et peut aussi être une solution pour pallier les phénomènes de résistance.



# Conclusion générale et perspectives

Cette étude de l'activation du facteur oedémateux EF de *B. anthracis* par la calmoduline s'inscrit dans le cadre plus général des relations structure-dynamique-fonction. La plupart des fonctions biologiques sont régies par des interactions biophysiques entre biomolécules, qui sont par nature dynamiques. Dans cette perspective, nous nous sommes attachés à décrire les influences à l'oeuvre dans le complexe EF-CaM, comme préliminaire à la recherche rationnelle d'inhibiteurs de la toxine EF. L'analyse de simulations de dynamique moléculaire (*DM*) du complexe a permis de caractériser la plasticité conformationnelle des deux partenaires et de proposer un modèle de l'interaction, dans lequel CaM agit comme un ressort qui maintient EF en conformation ouverte active. Les observations issues des trajectoires *DM* ont pu être reliées aux données expérimentales disponibles dans la littérature. Par ailleurs, l'influence mutuelle entre les interactions EF-CaM et  $\text{Ca}^{2+}$ -CaM a pu être mise en évidence et caractérisée qualitativement, à travers des simulations *LES* et des calculs *MMPBSA*.

L'effet électrostatique des ions  $\text{Ca}^{2+}$  sur le réseau de résidus du complexe EF-CaM a également été modélisé, à travers une analyse à la fois dynamique et énergétique. L'interprétation conjointe des motifs de corrélations dynamiques et des influences énergétiques a fait émerger le concept de connexité du réseau comme critère de stabilité. L'utilisation d'un panel de méthodes récentes a permis d'obtenir à la fois des informations sur le comportement dynamique local et une image globale de l'architecture du complexe. Des développements méthodologiques sont maintenant envisagés, s'agissant d'affiner le modèle de réseau élastique, ou de définir de nouvelles normes pour l'analyse des traits caractéristiques, dans le but d'identifier des *routes* par lesquelles l'information se propage à travers le réseau de résidus.

La mobilité structurale de la toxine EF, illustrée par la large transition conformationnelle lors de l'interaction avec CaM, a été décrite à l'aide de méthodes de dynamique moléculaire. Les conformations intermédiaires obtenues, composant un chemin d'énergie minimum plausible, ont servi de point de départ pour la recherche rationnelle d'inhibiteurs de EF. Au lieu de cibler le site catalytique de l'enzyme, nous avons choisi de cibler l'interaction entre EF et CaM, à travers le criblage virtuel d'une poche à la surface de EF. Cette poche est présente dans les conformations proches de la forme inactive de EF, tandis que sa cavité disparaît le long du chemin. Nous avons ainsi identifié six molécules inhibitrices de l'activité de EF, avec des affinité de quelques  $\mu\text{M}$ .

D'autres stratégies auraient pu ou pourraient être envisagées, par exemple telles que les molécules sélectionnées soient capables d'activer la toxine en absence de CaM, provoquant des dommages au sein de la bactérie même.

Le travail présenté ici est largement basé sur des simulations de dynamique moléculaire de quinze nanosecondes. Le long des trajectoires, nous avons pu détecter des changements conformationnels et mesurer des caractéristiques géométriques. A travers le traitement de l'ensemble des conformations, nous avons également été en mesure de décrire qualitativement des propriétés dynamiques et énergétiques. Toutefois, quinze nanosecondes ne permettent qu'une exploration limitée de l'espace conformationnel autour de la structure cristallographique initiale, et les événements observés le long des trajectoires restent particuliers, parmi une multitude de réalisations possibles. L'obtention de données quantitatives, s'agissant par exemple de l'affinité de la calmoduline pour le calcium, aurait nécessité l'utilisation de méthodes plus sophistiquées et plus demandeuses en ressources et temps de calcul, qui dépassent le cadre de cette étude.

Un des intérêts forts présentés par la protéine EF, outre son implication dans la pathogénicité du bacille du charbon, est sa mobilité structurale remarquable. La résolution d'un chemin de transition a fourni des conformations intermédiaires, localisées dans différentes régions du paysage énergétique de l'enzyme. Il n'est cependant pas trivial de valider un tel chemin adiabatique par des méthodes expérimentales. Le caractère unique de ce chemin s'inscrit dans le contexte d'un champ de force particulier. En solution, cette unicité est remise en cause. De plus, le fossé d'énergie observé entre les structures cristallographiques et les conformations intermédiaires pose question. Malgré cela, les résultats positifs obtenus lors des tests enzymatiques des ligands sélectionnés prouvent l'utilité et la pertinence d'une telle modélisation.

Les expériences réalisées *in vitro* ont apporté une forme de validation de notre approche *in silico*. Toutefois, si le protocole utilisé a permis d'identifier des molécules actives parmi un ensemble de composés inactifs, les scores calculés par les logiciels de *docking* ne sont pas corrélés avec les affinités mesurées expérimentalement. Cet état de fait illustre les limites des fonctions de *scoring*. Dans l'optique d'une optimisation éventuelle des composés, des expériences additionnelles sont nécessaires pour s'assurer de la spécificité des inhibiteurs et de la localisation du site de liaison (par cristallographie aux rayons X ou spectroscopie de fluorescence). Des calculs d'énergie libre de liaison permettraient également de caractériser les déterminants de l'affinité de chacun des six inhibiteurs.

Comme on l'a vu dans l'introduction, *Bacillus anthracis* constitue un risque bioterroriste sérieux. Etant donné le rôle majeur que jouent les toxines de la bactérie dans la pathogénèse du charbon, l'identification d'inhibiteurs spécifiques de PA, LF et EF est une voie importante pour le

développement de traitements efficaces. La plupart des études menées jusqu'ici se sont focalisées sur le facteur létal LF. Ici, grâce à une approche originale, nous avons identifié une nouvelle famille de composés inhibiteurs de l'activité de EF, présentant des affinités au moins aussi bonnes que les inhibiteurs existants [233; 227; 168; 50].

La calmoduline est un senseur du calcium qui régule de nombreuses fonctions vitales pour la cellule. Si son comportement dynamique a déjà été largement décrit [282; 283; 229; 284; 87; 203; 150; 287], son interaction avec EF n'avait jusque là pas été étudiée de manière extensive par des outils de modélisation moléculaire. De précédentes simulations avaient été réalisées en des temps beaucoup plus courts et dans des conditions différentes [289]. Les simulations présentées ici apportent donc une meilleure compréhension de la modulation du rôle de senseur de calcium de la calmoduline par un pathogène.

Les tests enzymatiques ont été réalisés en collaboration avec l'unité de Biochimie des Interactions Macromoléculaires (Daniel Ladant, Institut Pasteur), qui possède une large expertise sur la toxine CyaA de *Bordetella pertussis* [157; 156; 159; 66; 52]. Comme évoqué dans le dernier chapitre, CyaA est activée par CaM et partage 25% d'identité de séquence avec EF. Les modes d'interaction avec la calmoduline diffèrent, du fait notamment que CyaA ne possède pas de domaine hélicoïdal. Ainsi, elle présente une activité basale et le lobe C-terminal de CaM suffit à l'activer. La structure du domaine catalytique de l'enzyme non liée à CaM et les mécanismes détaillés de l'interaction Cya-CaM restent à élucider. La connaissance acquise ici sur la toxine EF pourrait être transférée à l'étude de CyaA. On a vu notamment que les inhibiteurs de l'activité de EF sont aussi efficaces sur CyaA. Des simulations *DM* de CyaA libre et en complexe avec CaM sont envisagées dans un premier temps, pour caractériser la dynamique de la protéine. L'expertise de l'unité de Daniel Ladant permettrait un dialogue intense entre méthodes computationnelles et expérimentales dans le cadre d'un tel projet.

Les résultats de cette thèse confirment l'utilité des approches de bioinformatique structurale pour caractériser la biophysique des interactions entre biomolécules et pour la recherche (amont) de nouveaux médicaments. Les simulations de dynamique moléculaire sont une méthode établie pour sonder le comportement des systèmes biomoléculaires sur des échelles de temps et d'espace difficilement accessibles expérimentalement. Avec le développement massif des ressources informatiques et de la parallélisation des calculs, il est aujourd'hui possible de simuler fidèlement des systèmes de plusieurs dizaines de milliers d'atomes en des temps de l'ordre de la microseconde. Par ailleurs, les approches *in silico* de recherche de nouveaux médicaments, comme la chemoinformatique, présentent un avantage majeur en terme de rapidité et de coûts vis-à-vis des criblages *in vitro* à haut-débit. Ainsi, allant de pair avec les efforts de séquençage, notamment du génome humain, la tendance actuelle est à la rationalisation et la systématisation de l'utilisation des outils de bioinformatique structurale et de chemoinformatique pour la recherche pharmaceutique. Un

compromis nécessaire reste à trouver entre leur capacité à traiter rapidement de gros volumes de données et la qualité des modèles sous-jacents, dont dépend le pouvoir prédictif de ces outils.

La façon d'envisager le comportement des systèmes biomoléculaires a récemment évolué. Ainsi, au-delà de l'action individuelle d'effecteurs, tels que les enzymes, c'est la régulation allostérique des interactions entre biomolécules qui détermine les fonctions biologiques. Ce qui a été tenté ici, et qui constitue une voie de développement pour la recherche de nouveaux traitements, est la modulation directe des interactions biomoléculaires par des petites molécules, plutôt que la voie traditionnelle qui consiste à cibler le site catalytique d'une enzyme. Par ailleurs, la vision originelle univoque et statique des structures protéiques a évolué vers un nouveau paradigme selon lequel les protéines existent à l'état natif sous plusieurs conformations fonctionnellement actives, qui inter-convertissent en quelques millisecondes, voire plus. La caractérisation et la description de ces conformations constituent un défi futur pour la modélisation moléculaire.

# BIBLIOGRAPHIE

- [1] <http://www.tripos.com>.
- [2] <http://www.accelrys.com>.
- [3] <http://www.xfig.org/>.
- [4] <http://chimiotheque-nationale.enscm.fr>.
- [5] *ROCS, version 2.2, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com, 2006.*
- [6] R Abagyan. *ICM Manual v.3.1*. Molsoft LLC, La Jolla, CA, 2005.
- [7] R.A. Abagyan and A.K. Mazur. New methodology for computer-aided modelling of biomolecular structure and dynamics. 2. local deformations and cycles. *J. Biomol. Struct. Dyn.*, 6 :833–845, 1989.
- [8] R.A. Abagyan, M.M. Totrov, and D.A Kuznetsov. ICM : A new method for protein modeling and design : Applications to docking and structure prediction from the distorted native conformation. *J. Comp. Chem.*, 15 :488–506, 1994.
- [9] L. Abrami, N. Reig, and F. G. van der Goot. Anthrax toxin : the long and winding road that leads to the kill. *Trends Microbiol.*, 13 :72–78, 2005.
- [10] R. Abseher and M. Nilges. Efficient sampling in collective coordinate space. *Proteins*, 39 :82–88, 2000.
- [11] N. Ahuja, P. Kumar, and R. Bhatnagar. The adenylate cyclase toxins. *Crit. Rev. Microbiol.*, 30 :187–196, 2004.
- [12] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27 :1208–1209, 1957.
- [13] A. Amadei, A.B. Linssen, and H.J. Berendsen. Essential dynamics of proteins. *Proteins*, 17 :412–425, 1993.
- [14] J. An, M. Totrov, and R. Abagyan. Comprehensive identification of druggable protein ligand binding sites. *Genome Inform.*, 15 :31–41, 2004.
- [15] J. An, M. Totrov, and R. Abagyan. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol. Cell Proteomics*, 4 :752–761, 2005.
- [16] J. Aqvist. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.*, 94 :8021–8024, 1990.
- [17] I.T. Arkin, H. Xu, M.O. Jensen, E. Arbely, E.R. Bennett, K.J. Bowers, E. Chow, R.O. Dror, M.P. Eastwood, R. Flitman-Tene, B.A Gregersen, J.L. Klepeis, I. Kolossváry, Y. Shan, and D.E. Shaw. Mechanism of Na<sup>+</sup>/H<sup>+</sup> Antiporting. *Science*, 317 :799–803, 2007.
- [18] P. Ascenzi, P. Visca, G. Ippolito, A. Spallarossa, M. Bolognesi, and C. Montecucco. Anthrax toxin : a tripartite lethal combination. *FEBS Lett.*, 531 :384–388, 2002.
- [19] R. Aurora, T.P. Creamer, R. Srinivasan, and G.D. Rose. Local interactions in protein folding : lessons from the  $\alpha$ -helix. *J. Biol. Chem.*, 272 :1413–1416, 1997.
- [20] No authors listed. Update : Investigation of bioterrorism-related anthrax–Connecticut, 2001. *MMWR Morb. Mortal. Wkly. Rep.*, 50 :1077–1079, 2001.
- [21] Y.S. Babu, C.E. Bugg, and W.J. Cook. Structure of calmodulin refined at 2.2 Å resolution. *J. Mol. Biol.*, 204 :191–204, 1988.
- [22] Y.S. Babu, J.S. Sack, T.J. Greenhough, C.E. Bugg, A.R. Means, and W.J. Cook. 3-Dimensional structure of calmodulin. *Nature.*, 315 :37, 1985.
- [23] I. Bahar, A.R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.*, 2 :173–181, 1997.
- [24] M. Bähler and A. Rhoads. Calmodulin signaling via the IQ motif. *FEBS Lett.*, 513 :107–113, 2002.
- [25] C. T. Baldari, F. Tonello, S. R. Paccani, and C. Montecucco. Anthrax toxins : A paradigm of bacterial immune suppression. *Trends Immunol.*, 27 :434–440, 2006.

- [26] K Bastard, C Prévost, and M Zacharias. Accounting for loop flexibility during protein-protein docking. *Proteins*, 62 :956–969, 2006.
- [27] W. Baumeister and A. C. Steven. Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem. Sci.*, 25 :624–631, 2000.
- [28] A. Bax and M. Ikura. An efficient 3D NMR technique for correlating the proton and  $^{15}\text{N}$  backbone amide resonances with the alpha-carbon of the preceding residue in uniformly  $^{15}\text{N}/^{13}\text{C}$  enriched proteins. *J. Biomol. NMR*, 1 :99–104, 1991.
- [29] C. Baysal and A.R. Atilgan. Relaxation kinetics and the glassiness of native proteins : coupling of timescales. *Biophys. J.*, 3 :1570–1576, 2005.
- [30] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81 :3684–3690, 1984.
- [31] H. M. Berman, T. N. Bhat, P. E. Bourne, Z. Feng, G. Gilliland, H. Weissig, and J. Westbrook. The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.*, 7 Suppl :957–959, 2000.
- [32] H.M. Berman. The Protein Data Bank : a historical perspective. *Acta Cryst A*, 64 :88–95, 2008.
- [33] I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchina, G. Parigi, M. Peana, A. Provenzani, and M.A. Zoroddu. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc. Natl. Acad. Sci. USA*, 101 :6841–6846, 2004.
- [34] H.J. Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Design*, 8 :243–256, 1994.
- [35] G. P. Brady and P. F. Stouten. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.*, 14 :383–401, 2000.
- [36] S.C. Brewerton. The use of protein-ligand interaction fingerprints in docking. *Curr. Opin. Drug. Discov. Devel.*, 11 :356–364, 2008.
- [37] B. Brooks and M. Karplus. Harmonic dynamics of proteins : normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA.*, 80 :6571–6575, 1983.
- [38] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4 :187–217, 1983.
- [39] F. Brossier, M. Levy, and M. Mock. Anthrax spores make an essential contribution to vaccine efficacy. *Infect. Immun.*, 70 :661–664, 2002.
- [40] J. C. Burnett, E. A. Henchal, A. L. Schmaljohn, and S. Bavari. The evolving field of biodefence : therapeutic developments and diagnostics. *Nat Rev Drug Discov*, 4 :281–297, 2005.
- [41] D.A. Case, T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, and P.A. Kollman. *AMBER 8*. University of California, San Francisco, 2004.
- [42] DA Case, TA Darden, TE Cheatham, CL III Simmerling, J Wang, RE Duke, R Luo, M Crowley, RC Walker, W Zhang, KM Merz, B Wang, S Hayik, A Roitberg, G Seabra, I Kolossváry, KF Wong, F Paesani, J Vanicek, X Wu, SR Brozell, T Steinbrecher, H Gohlke, L Yang, C Tan, J Mongan, V Hornak, G Cui, DH Mathews, MG Seetin, C Sagui, V Babin, and PA Kollman. *AMBER 10*. University of California, San Francisco, 2008.
- [43] D.A. Case, T.A. Darden, T.E. Cheatham, C.L. III Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman. *AMBER 9*. University of California, San Francisco, 2006.
- [44] C. T. Caskey. The drug development crisis : efficiency and safety. *Annu. Rev. Med.*, 58 :1–16, 2007.

- [45] CDC. Use of anthrax vaccine in response to terrorism : supplemental recommendations of the Advisory Committee on Immunization Practices. *JAMA*, 288 :2681–2682, 2002.
- [46] V. Chandran, E.J. Stollar, K. Lindorff-Larsen, J.F. Harper, W.J. Chazin, C.M. Dobson, B.F. Luisi, and J. Christodoulou. Structure of the regulatory apparatus of a calcium-dependent protein kinase (CDPK) : a novel mode of calmodulin-target recognition. *J. Mol. Biol.*, 357 :400–410, 2006.
- [47] R. Chattopadhyaya, W.E. Meador, A.R. Means, and F.A. Quiocho. Calmodulin structure refined at 1.7 Å resolution. *J. Mol. Biol.*, 228 :1177–1192, 1992.
- [48] T.E. 3rd Cheatham, J. Srinivasan, D.A. Case, and Kollman P.A. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dyn.*, 16 :265–80, 1998.
- [49] V. Chelliah, L. Chen, T. L. Blundell, and S. C. Lovell. Distinguishing structural and functional restraints in evolution in order to identify interaction sites. *J. Mol. Biol.*, 342 :1487–1504, 2004.
- [50] D. Chen, M. Misra, L. Sower, J. W. Peterson, G. E. Kellogg, and C. H. Schein. Novel inhibitors of anthrax edema factor. *Bioorg. Med. Chem.*, 16 :7225–7233, 2008.
- [51] K. Chen, J. Ruan, and L.A. Kurgan. Prediction of three dimensional structure of calmodulin. *Protein J.*, 25 :57–70, 2006.
- [52] A. Chenal, J. I. Guijarro, B. Raynal, M. Delepierre, and D. Ladant. RTX calcium binding motifs are intrinsically disordered in the absence of calcium : implication for protein secretion. *J. Biol. Chem.*, 284 :1781–1789, 2009.
- [53] G. Cheng, B. Qian, R. Samudrala, and D. Baker. Improvement in protein functional site prediction by distinguishing structural and functional constraints on protein family evolution using computational design. *Nucleic Acids Res.*, 33 :5861–5867, 2005.
- [54] C. Chennubhotla and I. Bahar. Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput. Biol.*, 3 :1716–1726, 2007.
- [55] D. Chin and A.R. Means. Calmodulin : a prototypical calcium sensor. *Trends Cell Biol*, 10 :322–328, 2000.
- [56] 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. *Nat. Struct. Biol.*, 8 :990–997, 2001.
- [57] R. G. Coleman and K. A. Sharp. Travel depth, a new shape descriptor for macromolecules : application to ligand binding. *J. Mol. Biol.*, 362 :441–458, 2006.
- [58] R. J. Collier and J. A. Young. Anthrax toxin. *Annu. Rev. Cell Dev. Biol.*, 19 :45–70, 2003.
- [59] M. L. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221 :709–713, 1983.
- [60] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math. Soc.*, 49 :1 – 23, 1943.
- [61] M. A. Crawford, C. V. Aylott, R. W. Bourdeau, and G. M. Bokoch. *Bacillus anthracis* toxins inhibit human neutrophil NADPH oxidase activity. *J. Immunol.*, 176 :7557–7565, 2006.
- [62] A. Crivici and M. Ikura. Molecular and structural basis of target recognition by calmodulin. *Annu. Rev. Biophys. Biomol. Struct.*, 24 :85–116, 1995.
- [63] F. Dal Molin, F. Tonello, D. Ladant, I. Zornetta, I. Zamparo, G. Di Benedetto, M. Zaccolo, and C. Montecucco. Cell entry and cAMP imaging of anthrax edema toxin. *EMBO J.*, 25 :5405–5413, 2006.
- [64] G. A. Dalkas, A. Papakyriakou, A. Vlamis-Gardikas, and G. A. Spyroulias. Low molecular weight inhibitors of the protease anthrax lethal factor. *Mini Rev. Med. Chem.*, 8 :290–306, 2008.
- [65] T Darden, D York, and L. Pedersen. Particle mesh Ewald : An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98 :10089–10092, 1993.
- [66] N. Dautin, G. Karimova, and D. Ladant. Bordetella pertussis adenylate cyclase toxin : a versatile screening tool. *Toxicon*, 40 :1383–1387, 2002.

- [67] C. Davaine. Nouvelles recherches sur les infusoires du sang dans la maladie connue sous le nom de sang de rate. *Comptes rendus de l'Académie des sciences*, 57 :351–353, 386–387, 1863.
- [68] C. Davaine. Recherches sur les infusoires du sang dans la maladie connue sous le nom de sang de rate. *Comptes rendus de l'Académie des sciences*, 57 :220–223, 1863.
- [69] M.E. Davis and J.A. McCammon. Solving the finite difference linearized Poisson-Boltzmann equation : A comparison of relaxation and conjugate gradient methods. *J. Comput. Chem.*, 10 :386 – 391, 1989.
- [70] M.E. Davis and J.A. McCammon. Calculating Electrostatic Forces from Grid-Calculated Potentials. *J. Comput. Chem.*, 11 :401–409, 1990.
- [71] Patrice Debré. *Louis Pasteur*, chapter p. 420-426. Flammarion, 1994 (tirage 1997).
- [72] W. L. Delano. The pymol molecular graphics system : <http://www.pymol.org>, 2002.
- [73] M. Delarue and P. Dumas. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci. U.S.A.*, 101 :6957–6962, 2004.
- [74] M. Delarue and Y. H. Sanejouand. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases : the elastic network model. *J. Mol. Biol.*, 320 :1011–1024, 2002.
- [75] Z. Deng, C. Chuaqui, and J. Singh. Structural interaction fingerprint (SIFt) : a novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.*, 47 :337–344, 2004.
- [76] T. C. Dixon, M. Meselson, J. Guillemin, and P. C. Hanna. Anthrax. *N. Engl. J. Med.*, 341 :815–826, 1999.
- [77] C.L. Drum, S.Z. Yan, J. Bard, Y.Q. Shen, D. Lu, S. Soelaiman, Z. Grabarek, A. Bohm, and W.J. Tang. Structural basis for the activation of anthrax adenyl cyclase exotoxin by calmodulin. *Nature*, 415 :396–402, 2002.
- [78] M. Drysdale, S. Heninger, J. Hutt, Y. Chen, C. R. Lyons, and T. M. Koehler. Capsule synthesis by *Bacillus anthracis* is required for dissemination in murine inhalation anthrax. *EMBO J.*, 24 :221–227, 2005.
- [79] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, Inc., New York, USA, 1973.
- [80] R. Elber and M. Karplus. Enhanced sampling in molecular dynamics : Use of the Time-Dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.*, 112 :9161–9175, 1990.
- [81] B. Elshorst, M. Hennig, H. Forsterling, A. Diener, M. Maurer, P. Schulte, H. Schwalbe, C. Griesinger, J. Krebs, H. Schmid, T. Vorherr, and E. Carafoli. NMR solution structure of a complex of calmodulin with a binding peptide of the Ca<sup>2+</sup> pump. *Biochemistry*, 38 :12320–12332, 1999.
- [82] P. Emsley and K. Cowtan. Coot : model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography.*, 60 :2126–2132, 2004.
- [83] J.L. Fallon and F.A. Quijcho. A closed compact structure of native Ca<sup>2+</sup>-calmodulin. *Structure*, 11 :1303–1307, 2003.
- [84] M. Feher. Consensus scoring for protein-ligand interactions. *Drug Discov. Today*, 11 :421–428, 2006.
- [85] S.E. Feller, Y. Zhang, R.W. Pastor, and B.R. Brooks. Constant pressure molecular dynamics simulation : The Langevin piston method. *J. Chem. Phys.*, 103 :4613–4621, 1995.
- [86] B.E. Finn, J. Evenas, T. Drakenberg, J.P. Waltho, E. Thulin, and S. Forsen. Calcium-induced structural changes and domain autonomy in calmodulin. *Nat. Struct. Biol.*, 2 :777–783, 1995.
- [87] G. Fiorin, R.R. Biekofsky, A. Pastore, and P. Carloni. Unwinding the helical linker of calcium-loaded calmodulin : a molecular dynamics study. *Proteins*, 61 :829–839, 2005.
- [88] A. M. Firoved, G. F. Miller, M. Moayeri, R. Kakkar, Y. Shen, J. F. Wiggins, E. M. McNally, W. J. Tang, and S. H. Leppla. *Bacillus anthracis* edema toxin causes extensive tissue lesions and rapid lethality in mice. *Am. J. Pathol.*, 167 :1309–1320, 2005.
- [89] S. Fischer and M. Karplus. Conjugate peak refinement : an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chemical Physics Letters*, 194 :252–261, 1992.

- [90] R. Fletcher and C.M. Reeves. Function minimisation by conjugate gradients. *The computer journal*, 1964.
- [91] F.X. Le Foulon, E. Braud, F. Fabis, J.C. Lancelot, and S. Rault. Synthesis and combinatorial approach of the reactivity of 6-and 7-arylthieno[3,2-d][1,3]oxazine-2,4-diones. *Tetrahedron*, 59 :10051–10057, 2003.
- [92] F.X. Le Foulon, E. Braud, F. Fabis, J.C. Lancelot, and S. Rault. Solution-phase parallel synthesis of a 1140-member ureidothiophene carboxylic acid library. *J. Comb. Chem.*, 7 :253–257, 2005.
- [93] K.K. Frederick, J.K. Kranz, and A.J. Wand. Characterization of the backbone and side chain dynamics of the CaM-CaMKII $\alpha$  complex reveals microscopic contributions to protein conformational entropy. *Biochemistry*, 45 :9841–9848, 2006.
- [94] K.K. Frederick, M.S. Marlow, K.G. Valentine, and A.J. Wand. Conformational entropy in molecular recognition by proteins. *Nature.*, 448 :325–329, 2007.
- [95] A. M. Friedlander, P. R. Pittman, and G. W. Parker. Anthrax vaccine : evidence for safety and efficacy against inhalational anthrax. *JAMA*, 282 :2104–2106, 1999.
- [96] M.S. Friedrichs, L. Mueller, and M. Wittekind. An automated procedure for the assignment of protein  $^1\text{HN}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}\alpha$ ,  $^1\text{H}\alpha$ ,  $^{13}\text{C}\beta$  and  $^1\text{H}\beta$  resonances. *J. Biomol. NMR*, 4 :703–726, 1994.
- [97] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, and P.S. Shenkin. Glide : a new Approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*, 47 :1739–1749, 2004.
- [98] Y. P. Gauthier, J. N. Tournier, J. C. Paucod, J. P. Corre, M. Mock, P. L. Goossens, and D. R. Vidal. Efficacy of a vaccine based on protective antigen and killed spores against experimental inhalational anthrax. *Infect. Immun.*, 77 :1197–1207, 2009.
- [99] F. Glaser, R. J. Morris, R. J. Najmanovich, R. A. Laskowski, and J. M. Thornton. A method for localizing ligand binding pockets in protein structures. *Proteins*, 62 :479–488, 2006.
- [100] N. Go, T. Noguti, and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA.*, 80 :3696–3700, 1983.
- [101] A.D. Gordon. *Classification*. Boca Raton, 2nd edition edition, 1999.
- [102] Y. Gozes, M. Moayeri, J. F. Wiggins, and S. H. Leppla. Anthrax lethal toxin induces ketotifen-sensitive intradermal vascular leakage in certain inbred mice. *Infect. Immun.*, 74 :1266–1272, 2006.
- [103] Z. Grabarek. Structure of a trapped intermediate of calmodulin : calcium regulation of EF-hand proteins from a new perspective. *J. Mol. Biol.*, 346 :31351–66, 2005.
- [104] J. D. Grabenstein. Vaccines : countering anthrax : vaccines and immunoglobulins. *Clin. Infect. Dis.*, 46 :129–136, 2008.
- [105] H. Grubmüller, B. Heymann, and P. Tavan. Ligand binding : molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271 :997–999, 1996.
- [106] K. Gujraty, S. Sadacharan, M. Frost, V. Poon, R. S. Kane, and J. Mogridge. Functional characterization of peptide-based anthrax toxin inhibitors. *Mol. Pharm.*, 2 :367–372, 2005.
- [107] Q. Guo, J. E. Jureller, J. T. Warren, E. Solomaha, J. Florián, and W. J. Tang. Protein-protein docking and analysis reveal that two homologous bacterial adenylyl cyclase toxins interact with calmodulin differently. *J. Biol. Chem.*, 283 :23836–23845, 2008.
- [108] Q. Guo, Y. Shen, Y.S. Lee, C.S. Gibbs, M. Mrksich, and W.J. Tang. Structural basis for the interaction of Bordetella pertussis adenylyl cyclase toxin with calmodulin. *EMBO J.*, 24 :3190–3201, 2005.
- [109] Q. Guo, Y. Shen, N. L. Zhukovskaya, J. Florián, and W. J. Tang. Structural and kinetic analyses of the interaction of anthrax adenylyl cyclase toxin with reaction products cAMP and pyrophosphate. *J. Biol. Chem.*, 279 :29427–29435, 2004.
- [110] M. Gupta, S. Alam, and R. Bhatnagar. Kinetic characterization and ligand binding studies of His351 mutants of *Bacillus anthracis* adenylyl cyclase. *Arch. Biochem. Biophys.*, 446 :28–34, 2006.

- [111] C. Hadley. Biologists think bigger. *EMBO Rep.*, 5 :236–238, 2004.
- [112] T. A. Halgren. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model*, 49 :377–389, 2009.
- [113] K. Hamacher, J. Trylska, and J.A. McCammon. Dependency map of proteins in the small ribosomal subunit. *PLoS Comput. Biol.*, 2 :e10, 2006.
- [114] U.H.E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.*, 281 :140, 1997.
- [115] S. Harris. *Factories of Death : Japanese Biological Warfare, 1932-45 and the American Cover-Up*. Routledge, 1994.
- [116] J.A. Hartigan. *Clustering Algorithms*. Inc. New York, NY, USA, 1975.
- [117] H. Hartmann and J. Liebscher. *J. Synthesis*, 275, 1984.
- [118] M. Hendlich, F. Rippmann, and G. Barnickel. LIGSITE : automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.*, 15 :359–363, 1997.
- [119] R. W. Hockney. vol. 9. In *Methods in Computational Physics*, pages 135 – 211. Academic Press, New York, 1970.
- [120] K.P. Hoeflich and M. Ikura. Calmodulin in action : diversity in target recognition and activation mechanisms. *Cell*, 108 :739–742, 2002.
- [121] B. Huang and M. Schroeder. LIGSITEcsc : predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.*, 6 :19, 2006.
- [122] M. J. Hudson, W. Beyer, R. Böhm, A. Fasanella, G. Garofolo, R. Golinski, P. L. Goossens, U. Hahn, B. Hallis, A. King, M. Mock, C. Montecucco, A. Ozin, F. Tonello, and S. H. Kaufmann. Bacillus anthracis : balancing innocent research with dual-use potential. *Int. J. Med. Microbiol.*, 298 :345–364, 2008.
- [123] W. Humphrey, A. Dalke, and K. Schulten. VMD : visual molecular dynamics. *J. Mol. Graph.*, 14 :33–38,27–28, 1996.
- [124] R. Ihaka and R. Gentleman. R : A Language for Data Analysis and Graphics. *J Comp Graph Statist*, 5 :299–314, 1996.
- [125] M. Ikura, G.M. Clore, A.M. Gronenborn, G. Zhu, C.B. Klee, and A. Bax. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science.*, 256 :632–638, 1992.
- [126] M. Ikura, L.E. Kay, and A. Bax. A novel approach for sequential assignment of  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  spectra of proteins : heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, 29 :4659–4667, 1990.
- [127] M. Ikura, L.E. Kay, and A. Bax. Improved three-dimensional  $^1\text{H}$ - $^{13}\text{C}$ - $^1\text{H}$  correlation spectroscopy of a  $^{13}\text{C}$ -labeled protein using constant-time evolution. *J. Biomol. NMR*, 1 :299–304, 1991.
- [128] M. Ikura, D. Marion, L.E. Kay, H. Shih, M. Krinks, C.B. Klee, and Bax A. Heteronuclear 3D NMR and isotopic labeling of calmodulin. Towards the complete assignment of the  $^1\text{H}$  NMR spectrum. *Biochem. Pharmacol.*, 40 :153–160, 1990.
- [129] T. V. Inglesby, T. O’Toole, D. A. Henderson, J. G. Bartlett, M. S. Ascher, E. Eitzen, A. M. Friedlander, J. Gerberding, J. Hauer, J. Hughes, J. McDade, M. T. Osterholm, G. Parker, T. M. Perl, P. K. Russell, and K. Tonat. Anthrax as a biological weapon, 2002 : updated recommendations for management. *JAMA*, 287 :2236–2252, 2002.
- [130] H. Ishida, K. Takahashi, K. Nakashima, Y. Kumaki, M. Nakata, K. Hikichi, and M. Yazawa. Solution structures of the N-terminal domain of yeast calmodulin :  $\text{Ca}^{2+}$ -dependent conformational change and its functional implication. *Biochemistry*, 39 :13660–13668, 2000.
- [131] B. Isin, K. Schulten, E. Tajkhorshid, and I. Bahar. Mechanism of Signal Propagation upon Retinal Isomerization : Insights from Molecular Dynamics Simulations of Rhodopsin Restrained by Normal Modes. *Biophys. J.*, in press, 2008.
- [132] B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.*, 11 :224–230, 2001.

- [133] B. Isralewitz, S. Izrailev, and K. Schulten. Binding pathway of retinal to bacterio-opsin : a prediction by molecular dynamics simulations. *Biophys. J.*, 73 :2972–2979, 1997.
- [134] Y. Izumi, H. Watanabe, N. Watanabe, A. Aoyama, Y. Jinbo, and N. Hayashi. Solution X-ray scattering reveals a novel structure of calmodulin complexed with a binding domain peptide from the HIV-1 matrix protein p17. *Biochemistry.*, 47 :7158–7166, 2008.
- [135] M. E. Jimenez, K. Bush, J. Pawlik, L. Sower, J. W. Peterson, and S. R. Gilbertson. Synthesis and screening of small molecule inhibitors of anthrax edema factor. *Bioorg. Med. Chem. Lett.*, 18 :4215–4218, 2008.
- [136] C.K. Johnson. Calmodulin, conformational states, and calcium signaling. A single-molecule perspective. *Biochemistry.*, 45 :14233–14246, 2006.
- [137] S. L. Johnson, L. H. Chen, E. Barile, A. Emdadi, M. Sabet, H. Yuan, J. Wei, D. Guiney, and M. Pellicchia. Structure-activity relationship studies of a novel series of anthrax lethal factor inhibitors. *Bioorg. Med. Chem.*, 17 :3352–3368, 2009.
- [138] W.L. Jorgensen. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.*, 103 :335–340, 1981.
- [139] W Kabsch and C. Sander. Dictionary of protein secondary structure : Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 12 :2577–2637, 1983.
- [140] K. Kadau, J. L. Barber, T. C. Germann, and B. J. Alder. Scaling of atomistic fluid dynamics simulations. *Phys Rev E Stat Nonlin Soft Matter Phys*, 78 :045301, 2008.
- [141] Y. Kalidas and N. Chandra. PocketDepth : a new depth based algorithm for identification of ligand binding sites in proteins. *J. Struct. Biol.*, 161 :31–42, 2008.
- [142] G. Karimova, C. Fayolle, S. Gmira, A. Ullmann, C. Leclerc, and D. Ladant. Charge-dependent translocation of Bordetella pertussis adenylate cyclase toxin into eukaryotic cells : implication for the in vivo delivery of CD8(+) T cell epitopes into antigen-presenting cells. *Proc. Natl. Acad. Sci. U.S.A.*, 95 :12532–12537, 1998.
- [143] W. A. Keitel. Recombinant protective antigen 102 (rPA102) : profile of a second-generation anthrax vaccine. *Expert Rev Vaccines*, 5 :417–430, 2006.
- [144] E. Kellenberger, N. Foata, and D. Rognan. Ranking targets in structure-based virtual screening of three-dimensional protein libraries : methods and problems. *J. Chem. Inf. Model*, 48 :1014–1025, 2008.
- [145] J. Kirchmair, S. Distinto, D. Schuster, G. Spitzer, T. Langer, and G. Wolber. Enhancing drug discovery through in silico screening : strategies to increase true positives retrieval rates. *Curr. Med. Chem.*, 15 :2040–2053, 2008.
- [146] G. Kirsch, D. Caignant, and P.J. Caignant. *Heterocycl. Chem.*, 19 :443, 1982.
- [147] D.B. Kitchen, H. Decornez, J.R. Furr, and J. Bajorath. Docking and scoring in virtual screening for drug discovery : methods and applications. *Nat Rev Drug Discov*, 3 :935–349, 2004.
- [148] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase : effects of ionic strength and amino-acid modification. *Proteins*, 1 :47–59, 1986.
- [149] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.*, 19 :120–127, 2009.
- [150] C. Kobayashi and S. Takada. Protein grabs a ligand by extending anchor residues : molecular simulation for Ca<sup>2+</sup> binding to calmodulin loop. *Biophys. J.*, 90 :3043–3051, 2006.
- [151] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham. Calculating structures and free energies of complex molecules : combining molecular mechanics and continuum models. *Acc. Chem. Res.*, 33 :889–897, 2000.
- [152] D. M. Kristensen, R. M. Ward, A. M. Lisewski, S. Erdin, B. Y. Chen, V. Y. Fofanov, M. Kimmel, L. E. Kavrakli, and O. Lichtarge. Prediction of enzyme function based on 3D templates of evolutionarily important amino acids. *BMC Bioinformatics*, 9 :17, 2008.

- [153] M.B. Kubitzki and B.L. de Groot. Molecular dynamics simulations using temperature-enhanced essential dynamics replica exchange. *Biophys. J.*, 92 :4262–4270, 2007.
- [154] H. Kuboniwa, N. Tjandra, S. Grzesiek, H. Ren, C.B. Klee, and A. Bax. Solution structure of calcium-free calmodulin. *Nat. Struct. Biol.*, 2 :768–776, 1995.
- [155] D. B. Lacy, M. Mourez, A. Fouassier, and R. J. Collier. Mapping the anthrax protective antigen binding site on the lethal and edema factors. *J. Biol. Chem.*, 277 :3006–3010, 2002.
- [156] D. Ladant. Interaction of Bordetella pertussis adenylate cyclase with calmodulin. Identification of two separated calmodulin-binding domains. *J. Biol. Chem.*, 263 :2612–2618, 1988.
- [157] D. Ladant, C. Brezin, J. M. Alonso, I. Crenon, and N. Guiso. Purification and immunological characterization of bacterial and secreted adenylate cyclase of Bordetella pertussis. *Ann Sclavo Collana Monogr*, 3 :295–298, 1986.
- [158] D. Ladant, S. Michelson, R. Sarfati, A. M. Gilles, R. Predeleanu, and O. Bârzu. Characterization of the calmodulin-binding and of the catalytic domains of Bordetella pertussis adenylate cyclase. *J. Biol. Chem.*, 264 :4015–4020, 1989.
- [159] D. Ladant and A. Ullmann. Bordetella pertussis adenylate cyclase : a toxin with multiple talents. *Trends Microbiol.*, 7 :172–176, 1999.
- [160] A. Laio, A. Rodriguez-Forteza, F. L. Gervasio, M. Ceccarelli, and M. Parrinello. Assessing the accuracy of metadynamics. *J. Phys. Chem. B*, 109 :6714–6721, 2005.
- [161] M. R. Landon, D. R. Lancia, J. Yu, S. C. Thiel, and S. Vajda. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J. Med. Chem.*, 50 :1231–1240, 2007.
- [162] O.F. Lange and H. Grubmüller. Generalized correlation for biomolecular dynamics. *Proteins.*, 62 :1053–1061, 2006.
- [163] A. T. Laurie and R. M. Jackson. Q-SiteFinder : an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*, 21 :1908–1916, 2005.
- [164] R. Lavery, H. Sklenar, K. Zakrzewska, and B. Pullman. The flexibility of the nucleic acids : (II). The calculation of internal energy and applications to mononucleotide repeat DNA. *J. Biomol. Struct. Dyn.*, 3 :989–1014, 1986.
- [165] A.R. Leach. *Molecular Modelling : Principles and Applications* . Addison Wesley Publishing Company, 2001.
- [166] A. L. Lee, K. A. Sharp, J. K. Kranz, X. J. Song, and A. J. Wand. Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry*, 41 :13814–13825, 2002.
- [167] G. M. Lee and C. S. Craik. Trapping moving targets with small molecules. *Science*, 324 :213–215, 2009.
- [168] Y. S. Lee, P. Bergson, W. S. He, M. Mrksich, and W. J. Tang. Discovery of a small molecule that inhibits the interaction of anthrax edema factor with its cellular activator, calmodulin. *Chem. Biol.*, 11 :1139–1146, 2004.
- [169] J. Liang, H. Edelsbrunner, and C. Woodward. Anatomy of protein pockets and cavities : measurement of binding site geometry and implications for ligand design. *Protein Sci.*, 7 :1884–1897, 1998.
- [170] R. C. Liddington. Anthrax : a molecular full nelson. *Nature*, 415 :373–374, 2002.
- [171] J. Liebscher, B. Neumann, and H.J. Hartmann. *J. Prakt. Chem.*, 325 :915, 1983.
- [172] C.A. Lipinski, F. Lombardo, B.W. Dominy, and P.J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.*, 23 :3–25, 1997.
- [173] R.J. Loncharich, B.R. Brooks, and R.W. Pastor. Langevin dynamics of peptides : the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers*, 32 :523–535, 1992.
- [174] E. Lyman and D.M. Zuckerman. Ensemble-based convergence analysis of biomolecular trajectories. *Biophys. J.*, 91 :164–172, 2006.

- [175] R. J. Manchee, M. G. Broster, J. Melling, R. M. Henstridge, and A. J. Stagg. Bacillus anthracis on Gruinard Island. *Nature*, 294 :254–255, 1981.
- [176] R. J. Manchee, M. G. Broster, A. J. Stagg, and S. E. Hibbs. Formaldehyde Solution Effectively Inactivates Spores of *Bacillus anthracis* on the Scottish Island of Gruinard. *Appl. Environ. Microbiol.*, 60 :4167–4171, 1994.
- [177] G. Marcou and D. Rognan. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model*, 47 :195–207, 2007.
- [178] M.S. Marlow and A.J. Wand. Conformational dynamics of calmodulin in complex with the calmodulin-dependent kinase kinase alpha calmodulin-binding domain. *Biochemistry*, 45 :8732–8741, 2006.
- [179] G. Marsaglia, Tsang W.W., and J. Wang. Evaluating Kolmogorov’s distribution. *J. Stat. Soft.*, 8 :18, 2003.
- [180] L. Martínez, E. Laine, T. E. Malliavin, M. Nilges, and A. Blondel. ATP conformations and ion binding modes in the active site of anthrax edema factor : A computational analysis. *Proteins*, Jul 2009.
- [181] A.K. Mazur and R.A. Abagyan. New methodology for computer-aided modelling of biomolecular structure and dynamics. 1. non-cyclic structures. *J. Biomol. Struct. Dyn.*, 6 :815–832, 1989.
- [182] J.A. McCammon. Target flexibility in molecular recognition. *Biochim. Biophys. Acta.*, 1754 :221–224, 2005.
- [183] IK McDonald and JM. Thornton. Satisfying hydrogen bonding potential in proteins. *JMB*, 238 :777–793, 1994.
- [184] W.E. Meador, A.R. Means, and F.A. Quijcho. Target enzyme recognition by calmodulin : 2.4 Å structure of a calmodulin-peptide complex. *Science*, 257 :1251–1255, 1992.
- [185] W.E. Meador, A.R. Means, and F.A. Quijcho. Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science.*, 262 :1718–1721, 1993.
- [186] A.R. Means. Molecular mechanisms of action of calmodulin. *Recent Prog. Horm. Res.*, 44 :223–262, 1988.
- [187] N. Metropolis, A. Rosebluth, M. Rosebluth, and A. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21 :1087–1092, 1953.
- [188] M. Moayeri, D. Haines, H. A. Young, and S. H. Leppla. *Bacillus anthracis* lethal toxin induces TNF-alpha-independent hypoxia-mediated toxicity in mice. *J. Clin. Invest.*, 112 :670–682, Sep 2003.
- [189] M. Mock and A. Fouet. Anthrax. *Annu. Rev. Microbiol.*, 55 :647–671, 2001.
- [190] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C. R. Corbeil. Towards the development of universal, fast and highly accurate docking/scoring methods : a long way to go. *Br. J. Pharmacol.*, 153 Suppl 1 :7–26, 2008.
- [191] John Mongan. Interactive essential dynamics. *J. Comput. Aided Mol. Design*, 18 :433–436, 2004.
- [192] J. Monod, J. Wyman, and J. P. Changeux. ON THE NATURE OF ALLOSTERIC TRANSITIONS : A PLAUSIBLE MODEL. *J. Mol. Biol.*, 12 :88–118, 1965.
- [193] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, and A.J. Olson. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 19 :1639–1662, 1998.
- [194] M. Mourez, D. B. Lacy, K. Cunningham, R. Legmann, B. R. Sellman, J. Mogridge, and R. J. Collier. 2001 : a year of major advances in anthrax toxin research. *Trends Microbiol.*, 10 :287–293, 2002.
- [195] M. Nayal and B. Honig. On the nature of cavities on protein surfaces : application to the identification of drug-binding sites. *Proteins*, 63 :892–906, 2006.
- [196] G. Nemethy, K.D. Gibson, K. A. Palmer, C.N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H.A. Scheraga. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.*, 96 :6472 – 6484, 1992.

- [197] C. Neylon. Small angle neutron and X-ray scattering in structural biology : recent examples from the literature. *Eur. Biophys. J.*, 37 :531–541, 2008.
- [198] J. O’Brien, A. Friedlander, T. Dreier, J. Ezzell, and S. Leppla. Effects of anthrax toxin components on human neutrophils. *Infect. Immun.*, 47 :306–310, 1985.
- [199] J.L. Pascual-Ahuir, E.L. Mehler, and H. Weinstein. Calmodulin structure and function : Implication of arginine in the compaction related to ligand binding. *Molecular Engineering*, 1 :231–247, 1991.
- [200] I. Pauli, L.F. Timmers, R.A. Caceres, M.B. Soares, and W.F. de Azevedo. In silico and in vitro : identifying new drugs. *Curr. Drug Targets*, 9 :1054–1061, 2008.
- [201] K. P. Peters, J. Fauck, and C. Frömmel. The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J. Mol. Biol.*, 256 :201–213, 1996.
- [202] P. Petrone and V.S. Pande. Can conformational change be described by only a few normal modes? *Biophys. J.*, 90 :1583–1593, 2006.
- [203] E. Project, R. Friedman, E. Nachliel, and M. Gutman. A molecular dynamics study of the effect of  $\text{Ca}^{2+}$  removal on calmodulin structure. *Biophys. J.*, 90 :3842–3850, 2006.
- [204] R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
- [205] E. Raoult. Bioterrorisme : Rapport de Mission du Professeur Raoult. Technical report, Ministère de la Santé, de la Famille et des Personnes Handicapées, Ministère délégué à la Recherche et aux Nouvelles Technologies, 2003.
- [206] M. Rarey. FlexX release 2.1 User Guide. Technical report, BioSolveIt, 2007.
- [207] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, 261 :470–489, 1996.
- [208] M. Rarey, S. Wefing, and T. Lengauer. Placement of medium-sized molecular fragments into active sites of proteins. *J. Comput. Aided Mol. Design*, 10 :41–54, 1996.
- [209] S. Rault, J.C. Lancelot, B. Letois, M. Robba, and Y. Labat. *French Patent*, 9,203,732, 1992.
- [210] S. Rault, J.C. Lancelot, B. Letois, M. Robba, and Y. Labat. *Chem. Abstr.*, 123 :11841, 1995.
- [211] B. Raymond, D. Leduc, L. Ravaux, R. Le Goffic, T. Candela, M. Raymondjean, P. L. Goossens, and L. Touqui. Edema toxin impairs anthracidal phospholipase A2 expression by alveolar macrophages. *PLoS Pathog.*, 3 :e187, 2007.
- [212] U. Rester. From virtuality to reality - Virtual screening in lead discovery and lead optimization : a medicinal chemistry perspective. *Curr Opin Drug Discov Devel*, 11 :559–568, 2008.
- [213] C. M. Reyes and P. A. Kollman. Structure and thermodynamics of RNA-protein binding : using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.*, 297 :1145–1158, 2000.
- [214] A.R. Rhoads and F. Friedberg. Sequence motifs for calmodulin recognition. *FASEB J.*, 11 :331–340, 1997.
- [215] A. Roitberg and R. Elber. Modeling side chains in peptides and proteins : Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *J. Chem. Phys.*, 85 :9277, 1991.
- [216] L. D. Rotz, A. S. Khan, S. R. Lillibridge, S. M. Ostroff, and J. M. Hughes. Public health assessment of potential biological terrorism agents. *Emerging Infect. Dis.*, 8 :225–230, 2002.
- [217] J.-P. Ryckaert, G. Ciccotti, and H.J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints : Molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23 :327–341, 1977.
- [218] J. Sadowski and J. Gasteiger. From atoms and bonds to three-dimensional atomic coordinates : automatic model builders. *Chem. Rev.*, 1993.
- [219] J. Sadowski, J. Gasteiger, and G. Klebe. Comparison of automatic three-dimensional model builders using 639 x-ray structures. *J. Chem. Inf. Comput. Sci.*, 1994.

- [220] Y. Savir and T. Flusty. Conformational proofreading : the impact of conformational changes on the specificity of molecular recognition. *PLoS ONE*, 2 :e468, 2007.
- [221] M. Schaefer, C. Bartels, and M. Karplus. Solution conformations and thermodynamics of structured peptides : molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.*, 284 :835–848, 1998.
- [222] M.A. Schumacher, A.F. Rivard, H.P. Bachinger, and J.P. Adelman. Structure of the gating domain of a  $\text{Ca}^{2+}$ -activated  $\text{K}^+$  channel complexed with  $\text{Ca}^{2+}$ /calmodulin. *Nature*, 410 :1120–1124, 2001.
- [223] E.I. Shakhnovich. Protein design : a perspective from simple tractable models. *Fold Des*, 3 :R45–58, 1998.
- [224] M. Sharon and C. V. Robinson. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.*, 76 :167–193, 2007.
- [225] K. A. Sharp and B. Honig. Electrostatic interactions in macromolecules : theory and applications. *Annu. Rev. Biophys. Biophys. Chem.*, 19 :301–332, 1990.
- [226] Y. Shen, Y.S. Lee, S. Soelaiman, P. Bergson, D. Lu, A. Chen, K. Beckingham, Z. Grabarek, M. Mrksich, and W.J. Tang. Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylyl cyclase exotoxins. *EMBO J*, 21 :6721–6732, 2002.
- [227] Y. Shen, N. L. Zhukovskaya, M. I. Zimmer, S. Soelaiman, P. Bergson, C. R. Wang, C. S. Gibbs, and W. J. Tang. Selective inhibition of anthrax edema factor by adefovir, a drug for chronic hepatitis B virus infection. *Proc. Natl. Acad. Sci. U.S.A.*, 101 :3242–3247, 2004.
- [228] Y. Shen, N.L. Zhukovskaya, Q. Guo, J. Florián, and W.J. Tang. Calcium-independent calmodulin binding and two-metal-ion catalytic mechanism of anthrax edema factor. *EMBO J*, 24 :929–941, 2005.
- [229] C.M. Shepherd and H.J. Vogel. A molecular dynamics study of  $\text{Ca}^{2+}$ -calmodulin : evidence of interdomain coupling and structural collapse on the nanosecond timescale. *Biophys. J.*, 87 :780–791, 2004.
- [230] J.M. Shin and D.H. Cho. PDB-Ligand : a ligand database based on PDB for the automated and customized classification of ligand-binding structures. *Nuc. Acids Res.*, 33 :D238–D241, 2005.
- [231] M.R. Shirts and V.S. Pande. Screen savers of the world, unite! *Science*, 290 :1903–1904, 2000.
- [232] J. Keppie et J. L. Stanley Smith, H. The chemical basis of the virulence of *Bacillus anthracis*. V. The specific toxin produced by *B. anthracis* in vivo. *British journal of experimental pathology*, 36 :460–472, 1955.
- [233] S. Soelaiman, B. Q. Wei, P. Bergson, Y. S. Lee, Y. Shen, M. Mrksich, B. K. Shoichet, and W. J. Tang. Structure-based inhibitor discovery against adenylyl cyclase toxins from pathogenic bacteria that cause anthrax and whooping cough. *J. Biol. Chem.*, 278 :25990–25997, 2003.
- [234] V. Spiwok, P. Lipovová, and B. Králová. Metadynamics in essential coordinates : free energy simulation of conformational changes. *J. Phys. Chem.*, 111 :3073–3076, 2007.
- [235] E. A. Spotts Whitney, M. E. Beatty, T. H. Taylor, R. Weyant, J. Sobel, M. J. Arduino, and D. A. Ashford. Inactivation of *Bacillus anthracis* spores. *Emerging Infect. Dis.*, 9 :623–627, 2003.
- [236] J. Srinivasan, T.E. Cheatham, J.P. Cieplak, P.A. Kollman, and D.A. Case. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Am. Chem. Soc.*, 120 :9401–9409, 1998.
- [237] T. A. Steitz. A mechanism for all polymerases. *Nature*, 391 :231–232, 1998.
- [238] M. Stepanova. Dynamics of essential collective motions in proteins : theory. *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys.*, 76 :051918, 2007.
- [239] Y. Sugita and Y. Okamoto. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.*, 314 :141, 1999.
- [240] J. Symersky, G. Lin, S. Li, S. Qiu, M. Carson, N. Schormann, and M. Luo. Structural genomics of *caenorhabditis elegans* : crystal structure of calmodulin. *Proteins*, 53 :947–949, 2003.
- [241] K. Tai. Conformational sampling for the impatient. *Biophys. Chem.*, 107 :213–220, 2004.

- [242] W. J. Tang, J. Krupinski, and A. G. Gilman. Expression and characterization of calmodulin-activated (type I) adenylylcyclase. *J. Biol. Chem.*, 266 :8595–8603, 1991.
- [243] J. J. Tesmer, R. K. Sunahara, R. A. Johnson, G. Gosselin, A. G. Gilman, and S. R. Sprang. Two-metal-Ion catalysis in adenylyl cyclase. *Science*, 285 :756–760, 1999.
- [244] M.M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77 :1905–1908, 1996.
- [245] D. Tobi and I. Bahar. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA.*, 102 :18908–18913, 2005.
- [246] M Totrov and A Abagyan. Flexible ligand docking to multiple receptor conformations : a practical alternative. *Curr Opin Struct Biol*, 18 :178–184, 2008.
- [247] M. Totrov and R. Abagyan. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins*, Suppl 1 :215–220, 1997.
- [248] M. Totrov and R. Abagyan. Rapid boundary element solvation electrostatics calculations in folding simulations : successful folding of a 23-residue peptide. *Biopolymers*, 60 :124–133, 2001.
- [249] C.J. Tsai, A. del Sol, and R. Nussinov. Allostery : absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.*, 378 :1–11, 2008.
- [250] B. E. Turk. Manipulation of host signalling pathways by anthrax toxins. *Biochem. J.*, 402 :405–417, 2007.
- [251] B. E. Turk. Discovery and development of anthrax lethal factor metalloproteinase inhibitors. *Curr Pharm Biotechnol*, 9 :24–33, 2008.
- [252] P. C. B. Turnbull. Guidelines for the surveillance and control of anthrax in humans and animals (publication no. WHO/EMC/ZDI./98.6). Technical report, Department of Communicable Diseases Surveillance and Response, World Health Organization, Geneva, Switzerland, 1998.
- [253] Y. Ueda, H. Taketomi, and N. Go. Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers*, 17 :1531 – 1548, 1978.
- [254] A. Ulitsky and R. Elber. The thermal equilibrium aspects of the time dependent Hartree and the locally enhanced sampling approximations : formal properties, a correction, and computational examples for rare gas clusters. *J. Chem. Phys.*, 98 :3380–3388, 1993.
- [255] T.S. Ulmer, S. Soelaiman, S. Li, C.B. Klee, W.J. Tang, and A. Bax. Calcium dependence of the interaction between calmodulin and anthrax edema factor. *J. Biol. Chem.*, 278 :29261–29266, 2003.
- [256] N.V. Valeyev, D.G. Bates, P. Heslop-Harrison, I. Postlethwaite, and N.V. Kotov. Elucidating the mechanisms of cooperative calcium-calmodulin interactions : a structural systems biology approach. *BMC Syst. Biol.*, 2 :48, 2008.
- [257] D van Der Spoel, E Lindahl, B Hess, G Groenhof, AE Mark, and Berendsen HJ. GROMACS : fast, flexible, and free. *J. Comput. Chem.*, 26 :1701–1718, 2005.
- [258] L. Van Eldik and D. Watterson. *Calmodulin and Signal Transduction*. Academic Press., San Diego, 1998.
- [259] W.F. van Gunsteren and H. J. C. Berendsen. Computer Simulation of Molecular Dynamics : Methodology, Applications, and Perspectives in Chemistry. *Ang. Chem. Inter. Ed. in En.*, 29 :992 – 1023, 1990.
- [260] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, and R.D. Taylor. Improved Protein-Ligand Docking Using GOLD. *Proteins*, 52 :609–623, 2003.
- [261] L. Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159 :98 – 103, 1967.
- [262] L. Verlet. Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions. *Phys. Rev.*, 165 :201 – 214, 1968.
- [263] B.G. Vertessy, V. Harmat, Z. Böcskei, G. Náray-Szabó, F. Orosz, and J. Ovádi J. Simultaneous binding of drugs with different chemical structures to Ca<sup>2+</sup>-calmodulin : crystallographic and spectroscopic studies. *Biochemistry*, 37 :15300–15310, 1998.

- [264] S.W. Vetter and E. Leclerc. Novel aspects of calmodulin target recognition and activation. *Eur. J. Biochem.*, 270 :404–414, 2003.
- [265] B. O. Villoutreix, K. Bastard, O. Sperandio, R. Fahraeus, J. L. Poyet, F. Calvo, B. Déprez, and M. A. Miteva. In silico-in vitro screening of protein-protein interactions : towards the next generation of therapeutics. *Curr Pharm Biotechnol*, 9 :103–122, Apr 2008.
- [266] V. J. Vinson. Proteins in motion. Introduction. *Science*, 324 :197, 2009.
- [267] G. Vitale, L. Bernardi, G. Napolitani, M. Mock, and C. Montecucco. Susceptibility of mitogen-activated protein kinase kinase family members to proteolysis by anthrax lethal factor. *Biochem. J.*, 352 Pt 3 :739–745, 2000.
- [268] H.J. Vogel. The Merck Frosst Award Lecture 1994. Calmodulin : a versatile calcium mediator protein. *Biochem. Cell. Biol.*, 72 :357–376, 1994.
- [269] S. Vouquier, J. Mary, N. Dautin, J. Vinh, B. Friguet, and D. Ladant. Essential role of methionine residues in calmodulin binding to Bordetella pertussis adenylate cyclase, as probed by selective oxidation and repair by the peptide methionine sulfoxide reductases. *J. Biol. Chem.*, 279 :30210–30218, 2004.
- [270] G. Vriend. WHAT IF : a molecular modeling and drug design program. *J. Mol. Graph.*, 8 :52–56, 1990.
- [271] AC Wallace, RA Laskowski, and JM. Thornton. LIGPLOT : A program to generate schematic diagrams of protein-ligand interactions. *Prot Eng*, 8 :127–134, 1995.
- [272] J. Wang, P. Morin, W. Wang, and P. A. Kollman. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.*, 123 :5221–5230, 2001.
- [273] W. Wang and P. A. Kollman. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.*, 303 :567–582, 2000.
- [274] J. Warwicker and H. C. Watson. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J. Mol. Biol.*, 157 :671–679, 1982.
- [275] A. Watson and D. Keir. Information on which to base assessments of risk from environments contaminated with anthrax spores. *Epidemiol. Infect.*, 113 :479–490, 1994.
- [276] H. Weinstein and E.L. Mehler. Ca<sup>2+</sup>-binding structural dynamics in functions of calmodulin. *Annu. Rev. Physiol.*, 56 :213–236, 1994.
- [277] M. Weisel, E. Proschak, and G. Schneider. PocketPicker : analysis of ligand binding-sites with shape descriptors. *Chem. Cent. J.*, 1 :7, 2007.
- [278] Weizmann Institute of Science, Israel. *Grace User’s Guide, Version 5.1*, 2007.
- [279] J. C. Whisstock and A. M. Lesk. Prediction of protein function from protein sequence and structure. *Q. Rev. Biophys.*, 36 :307–340, 2003.
- [280] P Willett, J.M. Barnard, and G.M. Downs. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.*, 38 :983–996, 1998.
- [281] M.A. Wilson and A.T. Brunger. The 1.0 Å crystal structure of Ca(2+)-bound calmodulin : an analysis of disorder and implications for functionally relevant plasticity. *J. Mol. Biol.*, 301 :1237–1256, 2000.
- [282] W. Wriggers, E. Mehler, F. Pitici, H. Weinstein, and K. Schulten. Structure and dynamics of calmodulin in solution. *Biophys. J.*, 74 :1622–1639, 1998.
- [283] C. Yang, G.S. Jas, and K. Kuczera. Structure and dynamics of calcium-activated calmodulin in solution. *J. Biomol. Struct. Dyn.*, 19 :247–271, 2001.
- [284] C. Yang, G.S. Jas, and K. Kuczera. Structure, dynamics and interaction with kinase targets : computer simulations of calmodulin. *Biochim. Biophys. Acta.*, 1697 :289–300, 2004.
- [285] Q. Ye, H. Wang, J. Zheng, Q. Wei, and Z. Jia. The complex structure of calmodulin bound to a calcineurin peptide. *Proteins.*, 73 :19–27, 2008.

- [286] M. Zhang, T. Tanaka, and M. Ikura. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat. Struct. Biol.*, 2 :758–767, 1995.
- [287] Y. Zhang, H. Tan, Y. Lu, Z. Jia, and G. Chen.  $\text{Ca}^{2+}$  dissociation from the C-terminal EF-hand pair in calmodulin : a steered molecular dynamics study. *FEBS Lett.*, 582 :1355–1361, 2008.
- [288] Z. Zhang and W. Wriggers. Local feature analysis : a statistical theory for reproducible essential dynamics of large macromolecules. *Proteins.*, 64 :391–403, 2006.
- [289] J. Zhao, S.A. Roy, and D.J. Nelson. MD simulations of anthrax edema factor : calmodulin complexes with mutations in the edema factor "switch a" region and docking of 3'-deoxy ATP into the adenyl cyclase active site of wild-type and mutant edema factor variants. *J. Biomol. Struct. Dyn.*, 21 :159–170, 2003.
- [290] Z. Zhou, A.K. Felts, R.A. Friesner, and R.M. Levy. Comparative performance of several flexible docking programs and scoring functions : enrichment studies for a diverse set of pharmaceutically relevant targets. *J. Chem. Inf. Model*, 47 :1599–1608, 2007.
- [291] R. A. Zilinskas. Iraq's biological weapons. The past as future? *JAMA*, 278 :418–424, 1997.

# ANNEXE

## A.1 Tableau des scores de *docking* des molécules sélectionnées *in silico*

Dans le tableau suivant, sont reportés les scores de *docking* des meilleurs 1% de ligands, parmi les 3649 composés retenus, sur la poche SABC. Les meilleurs scores obtenus avec FlexX sont donnés pour la conformation initiale du chemin de transition *conf 1*, la structure cristallographique de EF fermé 1K8T [77] et les conformations intermédiaires *conf 8*, *conf 28*, *conf 47* du chemin de transition. Le score moyen obtenu avec FlexX pour *conf 1* est évalué en prenant les vingt meilleures poses de *docking* de chaque composé. L'intervalle de score donné dans la colonne *conf 1*<sup>ICM</sup> reflète les valeurs extrêmes des meilleurs scores obtenus avec ICM, lors de cinq exécutions de l'algorithme de *docking* sur *conf 1*. Les symboles ● et ★ indiquent respectivement les composés éliminés à cause de leurs mauvais scores ICM ou de leur mauvaise performance envers les contrôles négatifs *conf 8*, *conf 28*, *conf 47*. Le symbole ◇ indique les composés sélectionnés, testés et validés *in vitro* comme inhibiteurs de l'activité de EF.

statut	ID ( AB-000)	conf 1 <sup>Flex</sup> X	1K8T <sup>Flex</sup> X	conf 1 <sup>ICM</sup>	conf 1 <sup>Flex</sup> X (mean)	conf 8 <sup>Flex</sup> X	conf 28 <sup>Flex</sup> X	conf 47 <sup>Flex</sup> X
non reçu	08808	-37.71	-32.69	-22.36	-33.42 ± 1.95	-22.33	-15.66	-15.88
testé	02221	-34.94	-30.45	-27.25	-32.41 ± 1.31	-19.88	-16.85	-13.95
retiré •	08809	-34.06	-32.63	[-2.59;-6.18]	-33.22 ± 0.31	-26.79	-24.13	-19.58
testé	01436	-33.24	-30.20	[-11.86;-14.68]	-24.79 ± 2.95	-19.14	-12.84	-14.81
non reçu	08828	-33.12	-28.34	[-18.01;-24.83]	-31.25 ± 0.68	-24.75	-16.85	-14.70
non reçu	08813	-32.36	-32.51	[-12.54;-23.95]	-27.70 ± 2.15	-22.88	-15.94	-14.66
retiré ★	28121	-30.57	-34.93	-11.98	-26.32 ± 1.32	-26.88	-16.97	-19.84
retiré ★	16699	-30.04	-33.73	[-6.40;-14.15]	-24.99 ± 2.88	-26.29	-13.35	-16.21
testé	14688	-29.57	-25.52	-13.88	-25.69 ± 1.05	-14.00	-3.05	-3.96
retiré •	31018	-29.55	-32.08	+4.05	-28.21 ± 0.58	-23.16	-19.30	-12.56
testé	02227	-29.44	-27.81	[-23.30;-26.14]	-26.20 ± 1.95	-18.81	-14.33	-16.10
non reçu	25669	-28.85	-29.13	[-15.17;-28.01]	-24.48 ± 1.17	-18.81	-12.46	-11.05
testé	18538	-28.71	-30.21	[-18.05;-19.80]	-25.14 ± 1.25	-21.41	-13.52	-13.13
testé	01990	-28.54	-28.06	[-19.82;-28.72]	-24.46 ± 1.37	-20.62	-12.77	-12.81
testé	05497	-28.40	-32.60	[-19.17;-22.60]	-25.12 ± 1.03	-20.93	-17.75	-15.64
testé	03735	-28.36	-23.52	[-11.88;-17.18]	-22.01 ± 2.52	-9.51	-3.58	-6.89
retiré •	28288	-28.31	-23.72	+5.49	-24.96 ± 1.12	-13.49	-12.28	-4.77
testé ◇	04721	-28.12	-30.74	[-20.60;-20.92]	-22.73 ± 2.51	-19.25	-12.62	-15.82
non reçu	22813	-27.93	-25.92	[-20.28;-21.45]	-24.82 ± 1.64	-17.47	-7.22	-10.51
testé ◇	04725	-27.86	-30.74	[-19.39;-20.53]	-22.88 ± 2.29	-19.25	-12.24	-15.82
testé	05597	-27.82	-33.33	[-18.52;-18.94]	-22.10 ± 1.59	-19.69	-13.84	-14.52
non reçu	21350	-27.67	-18.36	[-18.41;-18.63]	-18.97 ± 3.72	-16.42	-8.01	-7.81
testé ◇	03633	-27.66	-30.74	[-11.46;-23.53]	-22.94 ± 2.17	-19.25	-14.45	-15.82
testé	13697	-27.63	-28.79	[-17.77;-22.75]	-25.54 ± 0.71	-23.20	-14.55	-8.29
non reçu	08589	-27.62	-26.68	[-18.58;-18.82]	-23.66 ± 1.66	-16.28	-12.99	-12.17
testé	13694	-27.53	-27.51	[-21.66;-23.71]	-26.10 ± 0.45	-20.53	-14.17	-6.10
testé ◇	04723	-27.51	-30.74	[-19.00;-19.45]	-22.78 ± 2.25	-20.53	-13.28	-15.89
non reçu	30016	-27.33	-28.28	[-10.77;-11.01]	-22.33 ± 1.66	-17.00	-16.17	-12.20
testé ◇	04720	-27.31	-31.16	[-16.30;-18.89]	-23.89 ± 2.06	-19.38	-12.90	-16.05
retiré •	22934	-27.16	-19.26	[-8.09;-11.02]	-20.30 ± 2.61	-10.54	-6.72	-7.26
non reçu	28047	-27.16	-35.03	[-17.29;-27.46]	-25.70 ± 0.66	-24.21	-16.68	-19.41
retiré ★ •	26943	-27.14	-39.96	cannot dock	-25.47 ± 0.89	-34.03	-2.71	-7.99
non reçu	26363	-27.12	-20.33	[-22.14;-24.48]	-20.27 ± 2.50	-6.07	-2.48	-5.36
testé ◇	03634	-27.09	-29.43	[-19.85;-20.68]	-24.62 ± 1.76	-17.85	-13.44	-14.76
not receive	22795	-27.07	-35.36	[-16.03;-24.44]	-22.55 ± 1.34	-21.20	-13.42	-10.12
testé	02222	-27.07	-21.44	[-10.75;-19.06]	-23.24 ± 1.50	-13.70	-10.21	-7.00

## A.2 Publication décrivant la plasticité conformationnelle des partenaires de l'interaction EF-CaM

# The conformational plasticity of calmodulin upon calcium complexation gives a model of its interaction with the oedema factor of *Bacillus anthracis*

Elodie Laine,<sup>1</sup> Julliane D. Yoneda,<sup>2</sup> Arnaud Blondel,<sup>1</sup> and Thérèse E. Malliavin<sup>1\*</sup>

<sup>1</sup>Unité de Bioinformatique Structurale, CNRS URA 2185 Institut Pasteur, 25–28, rue du Dr. Roux, Paris 75 724, France

<sup>2</sup>Programa de Pós-graduação em Química Orgânica, Universidade Federal Fluminense, Outeiro de S. Joao Batista s/n, 24020150, Niteroi, RJ, Brazil

## ABSTRACT

We analyzed the conformational plasticity of calmodulin (CaM) when it is bound to the oedema factor (EF) of *Bacillus anthracis* and its response to calcium complexation with molecular dynamics (MD) simulations. The EF–CaM complex was simulated during 15 ns for three different levels of calcium bound to CaM. They were respectively no calcium ion (EF–(Apo–CaM)), two calcium ions bound to the C-terminal domain of CaM (EF–(2Ca–CaM)), and four calcium ions bound to CaM (EF–(4Ca–CaM)). Calculations were performed using AMBER package. The analysis of the MD simulations illustrates how CaM forces EF in an open conformation to form the adenylyl cyclase enzymatic site, especially with the two calcium form of CaM, best suited to fit the open conformation of EF. By contrast, CaM encounters bending and unwinding of its flexible interlinker in EF–(Apo–CaM) and EF–(4Ca–CaM). Calcium binding to one domain of CaM affects the other one, showing a transmission of information along the protein structure. The analysis of the CaM domains conformation along the simulations brings an atomistic and dynamic explanation for the instability of these complexes. Indeed the EF-hand helices of the N-terminal domain tend to open upon calcium binding (EF–(4Ca–CaM)), although the domain is locked by EF. By contrast, the C-terminal domain is strongly locked in the open conformation by EF, and the removal of calcium induces a collapse of EF catalytic site (EF–(Apo–CaM)).

Proteins 2008; 71:1813–1829.  
© 2008 Wiley-Liss, Inc.

**Key words:** calmodulin; *Bacillus anthracis*; oedema factor; molecular dynamics; calcium; solvent accessible surface; essential dynamics; NMR.

## INTRODUCTION

Structure flexibility plays a key role in target recognition for a large variety of proteins, in particular, for the ubiquitous protein calmodulin (CaM)<sup>1</sup> which participates in calcium signalling pathways that regulate many crucial biological functions. CaM actually played a historical role in the development of multidimensional NMR methods.<sup>2–6</sup> It is also a good model for understanding the implications of protein flexibility<sup>7–10</sup> and the thermodynamics of protein–protein interactions.<sup>10–14</sup>

CaM is composed of 148 residues and consists of two globular domains (lobes) connected by a flexible linker. Each domain can bind up to two calcium ions. The first crystallographic structures of mammalian CaM displayed a dumbbell-like shape for free CaM<sup>15,16</sup> and a compact conformation for CaM in complex with target peptides.<sup>17</sup> However, molecular dynamics (MD) simulations,<sup>18–20</sup> NMR experiments,<sup>21,22</sup> and a recent single-molecule study<sup>23</sup> revealed that free CaM is actually very flexible and there probably exist several conformational substrates in solution. Similarly, recent crystal structures<sup>24,25</sup> highlighted a variety of substates in the crystalline environment. The relationship between calcium ion complexation and biological activity was recently reviewed for the whole EF-hand protein family.<sup>26,27</sup>

Mostly, when CaM interacts with a target the lobes contact each other, forming a hydrophobic channel that engulfs the CaM-binding domain.<sup>28–31</sup> But the oedema factor (EF)–CaM complex,<sup>32</sup> which we discuss here, and other novel structures<sup>33,34</sup> have revealed an unexpected extended conformation of CaM.<sup>35,36</sup> EF is secreted by the Gram-positive spore-forming bacterium *Bacillus anthracis*, the etiological agent of anthrax. Activated as an adenylyl cyclase by CaM it produces cAMP from ATP in an uncontrolled fashion, and to this

Grant sponsors: French Ministry of Defence (DGA/MRIS), The National Scientific Research Centre (CNRS), The Pasteur Institute.

\*Correspondence to: Thérèse E. Malliavin, Institut Pasteur, 25–28, rue du Dr. Roux, Paris 75 724, France. E-mail: terez@pasteur.fr

Received 13 June 2007; Revised 18 September 2007; Accepted 4 October 2007

Published online 3 January 2008 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.21862

**Table I**  
Preparation Details of the MD Simulations

	EF-(Apo-CaM)	EF-(2Ca-CaM)	EF-(4Ca-CaM)	4Ca-CaM
Number of counterions	13	9	5	5
Water box dimensions		$123.9 \times 81.9 \times 99.2 \text{ \AA}^3$		$57.5 \times 64.0 \times 76.1 \text{ \AA}^3$
Number of water molecules	24,632	24,642	24,652	8,459
Total number of atoms	83,851	83,879	83,907	27,601

extent impairs the host immune defences.<sup>37–40</sup> Inhibiting the formation of the EF-CaM complex<sup>41</sup> is thus an important target to fight against anthrax toxicity.

The first X-ray structure of the EF-CaM complex displayed CaM in an extended conformation, loaded with two calcium ions and making extensive contacts with EF.<sup>32</sup> Since then Shen *et al.*<sup>42</sup> have solved several crystal structures of the complex, with different levels of calcium. These structures, together with a recent NMR study of free CaM and EF-bound CaM,<sup>43</sup> suggest that the level of calcium bound to CaM influences the stability and even the formation of the EF-CaM complex.

This work focuses mainly on CaM within the EF-CaM complex, as the internal motions of this protein deserve particular attention,<sup>44</sup> and the behavior of EF is presented in relation with CaM. The general purpose is to explore the mechanics of the EF-CaM interaction by perturbing the EF-(2Ca-CaM) complex, where the C-terminal lobe (sites 3 and 4) is loaded with two  $\text{Ca}^{2+}$ . Indeed it is experimentally deduced<sup>43</sup> that 2  $\text{Ca}^{2+}$ -CaM has the highest affinity for EF, the complex with 4  $\text{Ca}^{2+}$ -CaM being less stable and the complex with Apo-CaM not formed. CaM response to the perturbation of EF-(2Ca-CaM) is studied to relate this behavior to the knowledge on free CaM,<sup>18–20,45</sup> and a mechanical model of the EF-CaM interaction is proposed.

We report three 15-ns MD simulations of EF-CaM with different calcium complexations: EF-(Apo-CaM) with no calcium ion, EF-(2Ca-CaM) with two calcium ions bound to the C-terminal domain of CaM, and EF-(4Ca-CaM) with four calcium ions bound to CaM. It is worth mentioning that a previous MD study of the complex was performed by Zhao *et al.*<sup>46</sup> However the authors only considered the EF-(2Ca-CaM) complex, simulations were performed during a very short time period of 350 ps and a smooth cutoff was used for long-range interactions instead of the PME protocol.

The report is organized along two main axes. On the one hand, the trajectories are analyzed to validate them according to literature knowledge about the EF-CaM complex stability, CaM dynamics, and calcium affinity. On the other hand, the system perturbation induced by adding or removing a calcium ion is described and related to a model of the EF-CaM interaction.

## MATERIALS AND METHODS

### Preparation of initial coordinate files

Coordinates of the EF-CaM complex were obtained from the PDB entry 1K93.<sup>32</sup> Chains C (EF) and F (CaM) were selected, as they have the smallest number of missing residues. Residues 768 to 798, which comprised some disordered regions, were deleted. Coordinates of the released 4Ca-CaM were obtained from calmodulin X-ray structure in the EF-(4Ca-CaM) complex<sup>42</sup> (PDB id 1XFX: chain Q). Hydrogens were added using the LEaP module from AMBER 8.0<sup>47</sup> with the parm99 parameter set.  $\text{Na}^+$  counterions were added to neutralize the system. The complex was then hydrated by a box of water molecules using a cutoff of 9 Å along the three axis, in order to discard any water molecule located further than 9 Å from any solute molecule. The TIP3P potential for water molecules<sup>48</sup> and periodic boundary conditions were used. Details are summed up in Table I.

Complexes with different levels of calcium were generated in the following way: the 1K93 structure produced the complexation with 2  $\text{Ca}^{2+}$ , the EF-(Apo-CaM) complex was generated by removing the cations from 1K93, and the EF-(4Ca-CaM) complex was generated by adding 2  $\text{Ca}^{2+}$  in the N-terminal calcium loops of CaM. The N-terminal domain of a recent free  $\text{Ca}^{2+}$ -loaded CaM structure<sup>49</sup> (PDB id: 1OOJ) was superimposed on the N-terminal domain of EF-bound CaM to position the calcium ions in the coordinating loops. Ions were Coulomb/Lennard-Jones parametrized (Van der Waals radius  $R = 1.7131 \text{ \AA}$ , well depth  $\epsilon = 0.459789 \text{ kcal mol}^{-1}$ ).<sup>50</sup>

### Molecular dynamics simulations

MD trajectories were calculated using AMBER 8.0<sup>47</sup> on an IBM cluster of Power 4 processors (IDRIS and CINES super computing centres). A cutoff of 9 Å was used for Lennard-Jones interactions, and long-range electrostatic interactions were calculated with the Particle Mesh Ewald (PME) protocol.<sup>51</sup> Temperature and pressure were regulated according to the Berendsen thermostat.<sup>52</sup>

Simulations were initiated by some rounds of semi-restrained and then unrestrained minimization of the entire system. Heating of the system was performed dur-

**Table II**

Comparison of Calcium Coordination Distances (Å) Calculated Between the N-Terminal Calcium Ions and the Corresponding Coordinating Atoms

	Free Ca <sup>2+</sup> -CaM structure 100J <sup>a</sup>	EF-(4Ca-CaM) 1st equilibration <sup>b</sup>	EF-(4Ca-CaM) 2nd equilibration <sup>c</sup>
<b>Site 1</b>			
D20.Oδ1*	2.5	4.3	2.7
D20.Oδ2	4.2	2.6	2.6
D22.Oδ1	3.5	2.6	2.6
D22.Oδ2*	3.0	4.2	2.6
D24.Oδ1*	2.4	6.2	2.5
D24.Oδ2	4.2	7.9	4.4
T26.O*	2.6	2.8	2.6
E31.Oε1	2.7	6.6	8.5
E31.Oε2	2.5	5.3	7.3
<b>Site 2</b>			
D56.Oδ1*	2.7	2.6	2.7
D56.Oδ2	4.3	4.7	2.6
D58.Oδ1*	2.8	9.9	2.6
D58.Oδ2	3.9	9.0	4.1
N60.Oδ1*	2.6	8.7	3.1
T62.O*	2.6	9.9	2.8
E67.Oε1	2.7	11.3	5.2
E67.Oε2	2.9	10.8	4.8

During the second equilibration step, distance restraints were applied in each site between the flagged (\*) atoms and the calcium position. Oε1 and Oε2 of E31, Oε1 and Oε2 of E67, which show larger distances than in free Ca<sup>2+</sup>-CaM (pdb entry 100J)<sup>49</sup> even after the second equilibration step, are also found far from the calcium ions in the crystallographic structure of EF-(4Ca-CaM) (pdb entry 1XFX),<sup>42</sup> at distances of 4.0 Å, 4.4 Å, 6.7 Å, and 4.6 Å.

<sup>a</sup>In the crystallographic structure of free Ca<sup>2+</sup>-CaM (pdb entry 100J).<sup>49</sup>

<sup>b</sup>In the EF-(4Ca-CaM) model, after the first equilibration step (without distance restraints).

<sup>c</sup>After the second equilibration step.

ing 10 ps up to 300 K at constant volume, while restraining the atoms of the complex (or calmodulin resp.) and the calcium positions with a force constant of 25 kcal mol<sup>-1</sup> Å<sup>2</sup>. The equilibration process was then performed in the following way: 1 MD round of 5 ps at constant volume and 4 MD rounds of 2.5 ps at constant pressure were run while reducing the position restraints from 25 kcal mol<sup>-1</sup> Å<sup>2</sup> to 5 kcal mol<sup>-1</sup> Å<sup>2</sup>; eventually a last MD round of 10 ps was performed with a restraint of 2.5 kcal mol<sup>-1</sup> Å<sup>2</sup>.

For heating and equilibration of the EF-(4Ca-CaM) complex, no restraints were applied on the newly added calcium ions—neither on the CaM EF-hands coordination loops (residues 18–30 and 52–66), so that the system could better accommodate them. At the end of the equilibration process however, both N-terminal calcium ions were found far away from their initial positions (Table II). Hence a second equilibration was performed during 50 ps. Position restraints of 2.5 kcal mol<sup>-1</sup> Å<sup>2</sup> were again applied on the whole system, except for calcium positions and the corresponding coordination loops. In addition, some specific distance restraints of 10 kcal mol<sup>-1</sup> Å<sup>2</sup> were applied between the first Ca<sup>2+</sup> and the oxygens Oδ1 of Asp 20, Oδ2 of Asp 22, Oδ1 of Asp 24, O of Thr 26 and between the second Ca<sup>2+</sup> and the oxygens Oδ1 of

Asp 56, Oδ1 of Asp 58, Oδ1 of Asn 60, O of Thr 62. As a result, most of the atoms in the coordination loops get as close from the calcium ions as in the X-ray structure 100J<sup>49</sup> (Table II). However, atoms Oε1 and Oε2 of Glu 31, Oε1 and Oε2 of Glu 67 are found farther away from the calcium ions (4.8–8.5 Å), as it is also the case in the crystal structure 1XFX of EF-(4Ca-CaM).<sup>42</sup>

MD simulations were performed at temperature of 300 K and pressure of 1 atm during 15 ns. The SHAKE algorithm<sup>53</sup> was used to keep rigid all covalent bond involving hydrogens, enabling a time step of 2 fs. Atom coordinates were saved every ps (15,000 snapshots).

### Analysis of the trajectories

To assess the convergence of the simulations we used the ensemble-based approach recently developed by Lyman and Zuckerman.<sup>54</sup> For each trajectory, it consists of generating a set of representative or “reference” structures and then clustering all the 15,000 conformations in corresponding “reference” groups.

An arbitrary cutoff  $r$  is first defined. A conformation is randomly picked up among the 15,000 conformations of the trajectory: this is the first “reference” structure. All conformations that show a RMSD from this “reference” structure smaller than  $r$  are removed from the trajectory. A second “reference” structure is picked up among the remaining conformations and again all conformations that show a RMSD smaller than  $r$  are removed. The operation is repeated until there is no remaining conformation. Finally, we obtain a set of “reference” structures that deviate from each other by at least  $r$  Å.

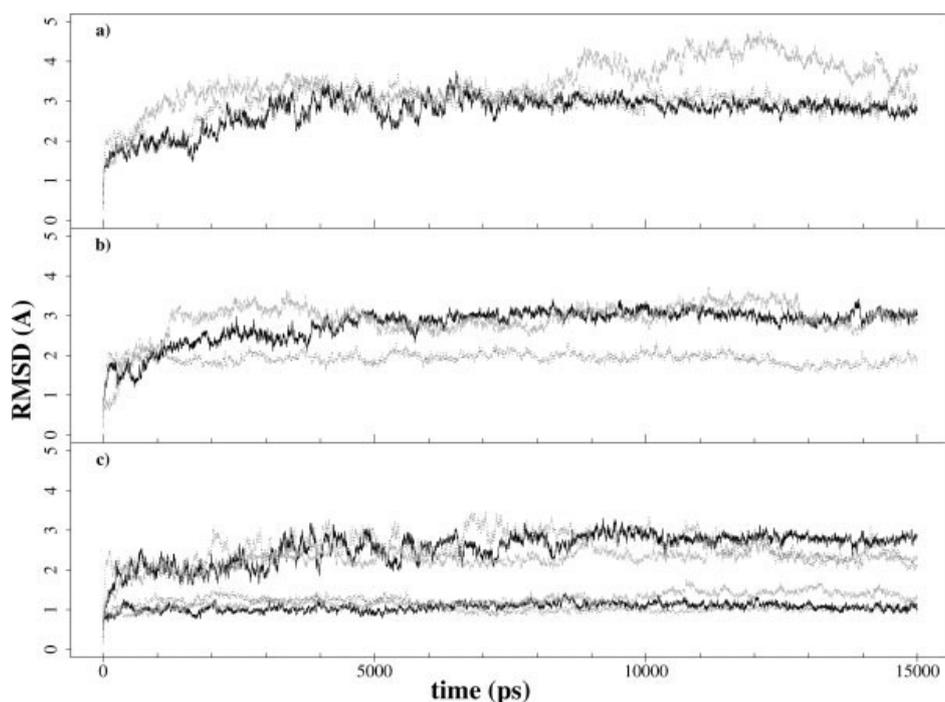
The trajectory is then split in two halves and conformations from each half are associated to their closest “reference” structure in terms of RMSD, constituting “reference” groups. If the simulation has converged, each “reference” structure must be represented by an equivalent number of conformations from both halves of the trajectory.

To compare this new method to a more traditional measurement, we computed the average time auto-correlation function on Cα atom positions:

$$C(t) = \frac{N!}{2!(N-2)!} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \langle v_{ij}(t_0) \cdot v_{ij}(t_0 + t) \rangle \quad (1)$$

where  $v_{ij}$  is the position vector between Cα atoms  $i$  and  $j$  and  $N$  is the number of Cα atoms. The obtained curve can be decomposed in exponential functions. For each simulation,  $C(t)$  was fitted to an exponential model function  $f(t) = c_1 e^{-t/\tau_1} + c_2 e^{-t/\tau_2} + c_3 e^{-t/\tau_3}$ <sup>55</sup> using XmGrace.<sup>56</sup>

A virtual dihedral angle (VDA), which describes the relative rotation of the lobes to each other, was calculated between the centres of mass of the four CaM EF-hands. Traditionally the VDA is taken as the pseudo-dihedral



**Figure 1**

Backbone RMSD of calmodulin with respect to the initial crystal structure (1K93)<sup>32</sup> (a) Total RMSD. (b) RMSD of the interlinker region using fit to residues 65–92. (c) RMSD of the N-terminal (up) and C-terminal domain (bottom) using independent fits to residues 5–64 and 92–147, respectively. In all plots, the EF-(Apo-CaM) trajectory is shown in grey (long dotted line), the EF-(2Ca-CaM) and EF-(4Ca-CaM) trajectories are shown in black (short dotted line and solid line, respectively).

angle formed by the four calcium ions bound to CaM.<sup>57</sup> But, since this angle definition is relevant only for a straight central helix,<sup>45</sup> we used EF-hand centres of mass instead of calcium positions here.

The bending of the interlinker was evaluated by calculating the angle formed between the axes of  $\alpha$ -helices IV (residues 64–72) and V (residues 86–94). Axes were defined with the successive middle points between atoms (i) N( $i$ ) and N( $i+2$ ), (ii) C $\alpha$ ( $i$ ), and C $\alpha$ ( $i+2$ ), (iii) C( $i$ ) and C( $i+2$ ),  $i$  being the residue index in the sequence. The EF-hand opening was evaluated in a similar way, from the angles between helices I (residues 8–19) and II (residues 31–37) in site 1, helices III (residues 46–53) and IV (residues 66–73) in site 2, helices V (residues 83–92) and VI (residues 103–110) in site 3, and helices VII (residues 119–127) and VIII (residues 139–145) in site 4.

Unless otherwise stated, trajectories were analyzed using the PTRAJ module of AMBER 8.0.<sup>47</sup> Determination of the secondary structure was performed by DSSP.<sup>58</sup> Electrostatic binding energies  $\Delta E_{\text{ele}}$  of calcium ions were calculated with the MMPBSA module of AMBER 8.0.<sup>47</sup> SASA calculations were performed by MSMS<sup>59</sup> with a probe sphere radius of 1.5 Å. Interactions were identified with the HBPLUS 2.25 program<sup>60</sup> in the LIGPLOT 4.4.2 package.<sup>61</sup> VMD<sup>62</sup> was used for

visualization and the analysis graphics were realised using the R package.<sup>63</sup>

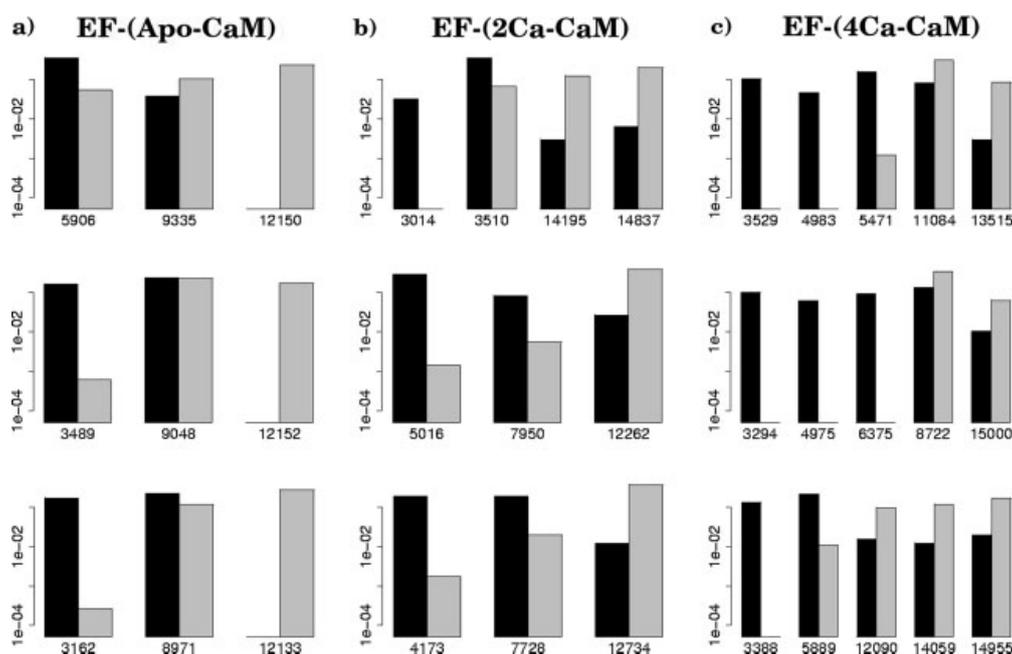
## RESULTS

To analyze CaM behavior in the EF-CaM complex, the following aspects were studied: (i) the relative stability of the different calcium levels; (ii) the influence of calcium level on calmodulin domain coupling and interaction with EF; (iii) the mechanics of the interaction between CaM and EF. In this work, we refer to the most recent MD simulations of free Apo-CaM<sup>64</sup> and Ca<sup>2+</sup>-loaded CaM<sup>19,20,65</sup> in order to compare the conformational changes of EF-bound CaM to those of free CaM.

### Relative stability of the EF-CaM complex for different levels of calcium

The relative stability of the different complexes was first investigated through an analysis of the convergence of the trajectories.

The conformational drift along the trajectories was evaluated by calculating the root mean square deviation (RMSD) of CaM coordinates from the initial coordinates (Fig. 1). Backbone RMS deviations stabilize at around



**Figure 2**

Convergence analysis on the (a) EF-(Apo-CaM), (b) EF-(2Ca-CaM), and (c) EF-(4Ca-CaM) trajectories. Each line corresponds to an analysis for a given choice of the seed. The reference structures numbers are indicated in x-coordinates and correspond to their positions in ps along the trajectory. For each reference group are indicated the proportions of members in the first half (black bars) and in the second half (grey bars) of the trajectory.

2.7 Å for EF-(2Ca-CaM) and EF-(4Ca-CaM) [Fig. 1(a)], whereas deviations for EF-(Apo-CaM) vary up to the end of the simulation in the 3.5–5 Å range. These values are small compared to RMSD values of free CaM, up to 15 Å,<sup>19,20</sup> because the amplitude of movement is reduced by EF. The RMS deviations calculated for EF (data not shown) are 3.7 Å (EF-(Apo-CaM)), 2.9 Å (EF-(2Ca-CaM)), and 4.1 Å (EF-(4Ca-CaM)): the smallest value is observed for the most experimentally stable complex.

Looking at Figure 1(c) we observe that the C-terminal domain (residues 93–147) is very stable with average RMSD between 1 and 1.5 Å whereas the RMSD is greater in the N-terminal domain (residues 5–64) up to 3 Å. Besides the RMSD of the whole protein and that of the domains are similar, indicating that, in complex with EF, global motions do not have the important contribution to the fluctuations that was found for free CaM.<sup>19,20</sup> We deduced from RMSD profiles [Fig. 1(a)] that 3 ns are necessary for the complex to relax from the initial crystallographic structure. These were thus removed for the convergence analysis and the latter study of CaM and EF.

To assess the convergence of the simulations in the last 12 ns we used the approach developed recently by Lyman and Zuckerman.<sup>54</sup> The robustness of the method was tested with different seeds for the conformation random

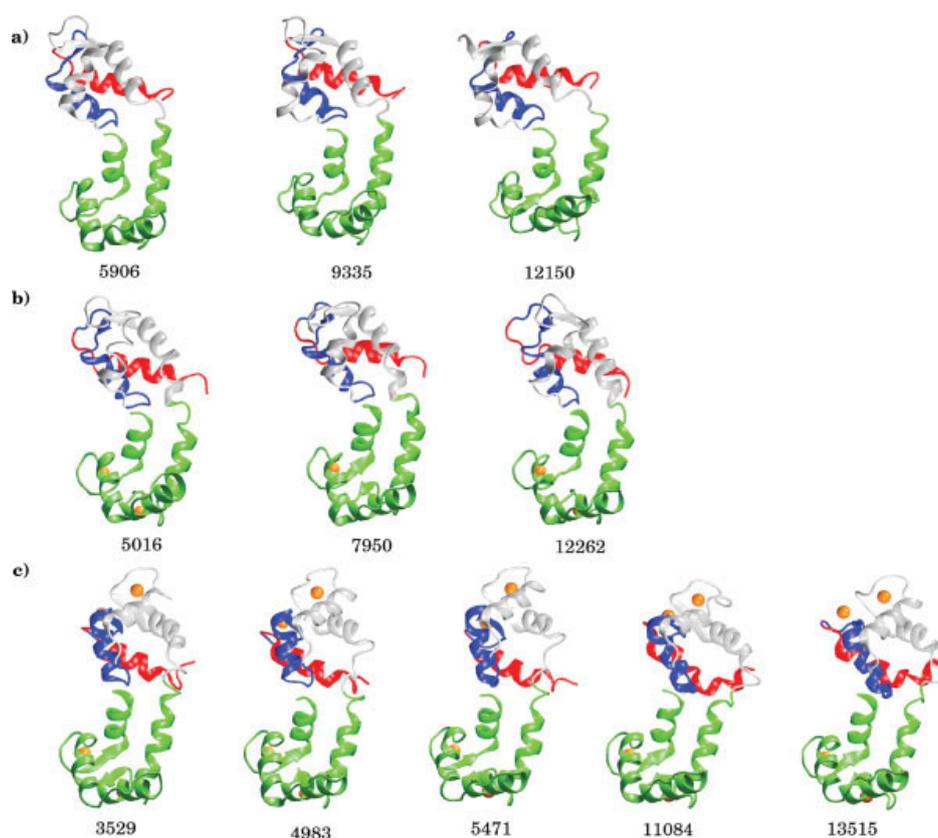
picking up. In effect, results were similar for all seeds and we display here three analyses (Fig. 2).

Structural diversity proved to be limited in the trajectories as a cutoff of 3.5 Å was not sufficient to obtain more than one reference structure for EF-(2Ca-CaM) or EF-(Apo-CaM). Thus, a 2.5 Å cutoff was chosen, which enables to define three (or four) reference structures from EF-(Apo-CaM) [Fig. 2(a)] and EF-(2Ca-CaM) [Fig. 2(b)] and five reference structures from EF-(4Ca-CaM) [Fig. 2(c)].

EF-(2Ca-CaM) gives a good convergence criteria, since most of the reference structures are represented in both halves of the trajectory [Figs. 2(b) and 3(b)]. EF-(Apo-CaM) also, but to a lesser extent since the complex seems to explore an ensemble of new conformations near the end of the trajectory (12 ns) [Figs. 2(a) and 3(a)]. Concerning EF-(4Ca-CaM), conformations from the first half of the simulation are more represented than conformations from the second half [Figs. 2(c) and 3(c)].

The fitting of the average time auto-correlation functions onto exponential curves confirms that the EF-(4Ca-CaM) trajectory converges more slowly than the EF-(Apo-CaM) and EF-(2Ca-CaM) trajectories (Table III).

EF-(2Ca-CaM) is the only trajectory displaying at the same time one of the smallest drifts and a good convergence: this is in agreement with the stability experimen-

**Figure 3**

Comparison of reference conformations from the (a) EF-(Apo-CaM), (b) EF-(2Ca-CaM), and (c) EF-(4Ca-CaM) trajectories. Calcium ions are colored in orange. Helix I of site 1 (residues 5–22), helix II of site 1 (residues 23–42), and helices V to VIII (residues 80–147) are shown in red, blue, and green, respectively. They correspond to the blocks defined in Table IX.

tally observed for the EF-CaM complex with 2  $\text{Ca}^{2+}$  ions.

The relative stability of the different complexes was then analyzed through the essential dynamics (ED)<sup>66</sup> of the two partners.

The three first modes calculated for calmodulin only contribute 61%, 58%, and 63% to the total CaM backbone mean-square fluctuations in EF-(Apo-CaM), EF-

(2Ca-CaM), and EF-(4Ca-CaM), respectively. These are smaller than the 88% and 71% contributions calculated by Yang *et al.*<sup>31</sup> for apo and calcium-loaded forms, respectively, on MD simulations of free CaM. As expected, CaM in complex with EF undergoes less collective motions than free CaM. Eigenvalues rapidly drop in EF-(Apo-CaM) and EF-(4Ca-CaM), while EF-(2Ca-CaM) profile is smoother [Fig. 4(a)]. This suggests that motions are even less collective in EF-(2Ca-CaM) than in the two other forms: while CaM fluctuates around a stable position in EF-(2Ca-CaM), the protein evolves toward new conformations in EF-(Apo-CaM) and EF-(4Ca-CaM).

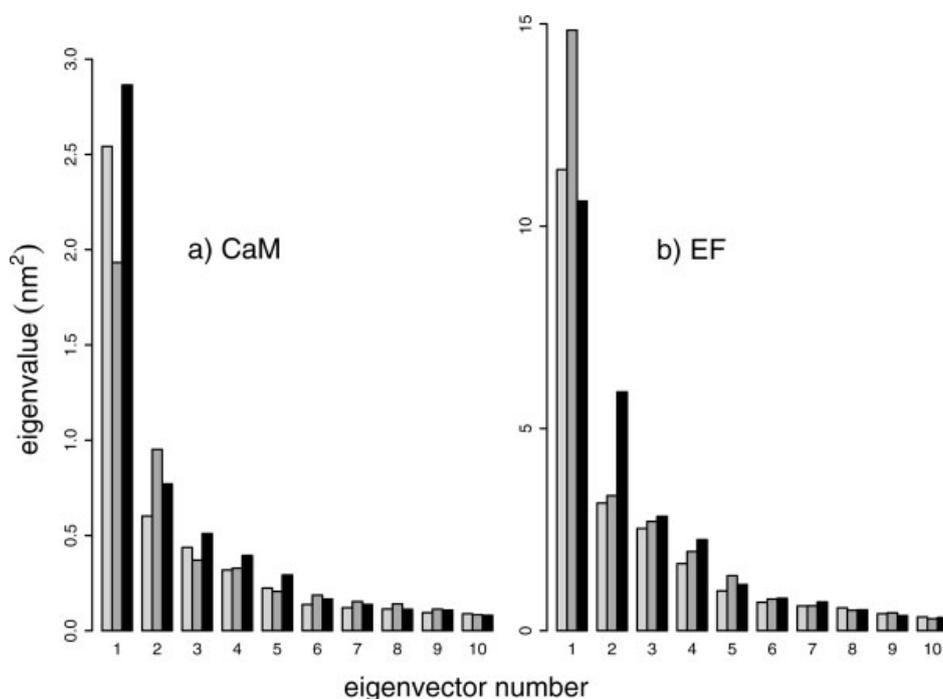
The two extreme backbone conformations of calmodulin along the first mode are displayed on Figure 5, together with the components of the mode on each atom. Deformations within the N-terminal domain are observed in the three complexes and movement of the N-tail is particularly visible in EF-(2Ca-CaM). Deformations of the interlinker region are observed in EF-(4Ca-

**Table III**

Coefficients  $c_1$  and  $\tau_1$  of the Best Fitted Exponential Curves to the Average Time Autocorrelation Functions of EF-(Apo-CaM), EF-(2Ca-CaM), and EF-(4Ca-CaM)

	EF-(Apo-CaM)	EF-(2Ca-CaM)	EF-(4Ca-CaM)
$c_1$	0.37	0.39	0.27
$\tau_1$	45.9	59.8	139.1

Correlations between the empirical functions and the fitted curves are within the 0.98–0.99 range.

**Figure 4**

Eigenvalues obtained from the ED analysis performed on CaM (left) and EF (right), between 3 and 15 ns: EF-(Apo-CaM) is in lightgray, EF-(2Ca-CaM) in darkgray, and EF-(4Ca-CaM) in black.

CaM) and to a lesser extent in EF-(Apo-CaM). In EF-(Apo-CaM), both lobes tend to get closer to each other (up and down arrows) while the interlinker is pushed away (right arrows): this motion could explain the increase in global RMSD at the end of the simulation [see Fig. 1(a)].

Hence, the analysis of the collective motions of CaM reveals larger ones for EF-(Apo-CaM) and EF-(4Ca-CaM), which are experimentally unstable complexes. It is interesting to note that, in these two trajectories, the first mode involves a deformation of the interlinker domain.

In the ED analysis<sup>66</sup> of EF (Fig. 4), the first eigenvalue for EF-(2Ca-CaM) is the largest one [Fig. 4(b)], eigenvalues then rapidly drop and 19 modes are sufficient to describe 90% of the total fluctuations. On the other hand, 21 modes are necessary to explain 90% of the total fluctuations in EF-(4Ca-CaM) and three more are needed in EF-(Apo-CaM). Thus, motions appear to be more diffuse in these two complexes than in EF-(2Ca-CaM).

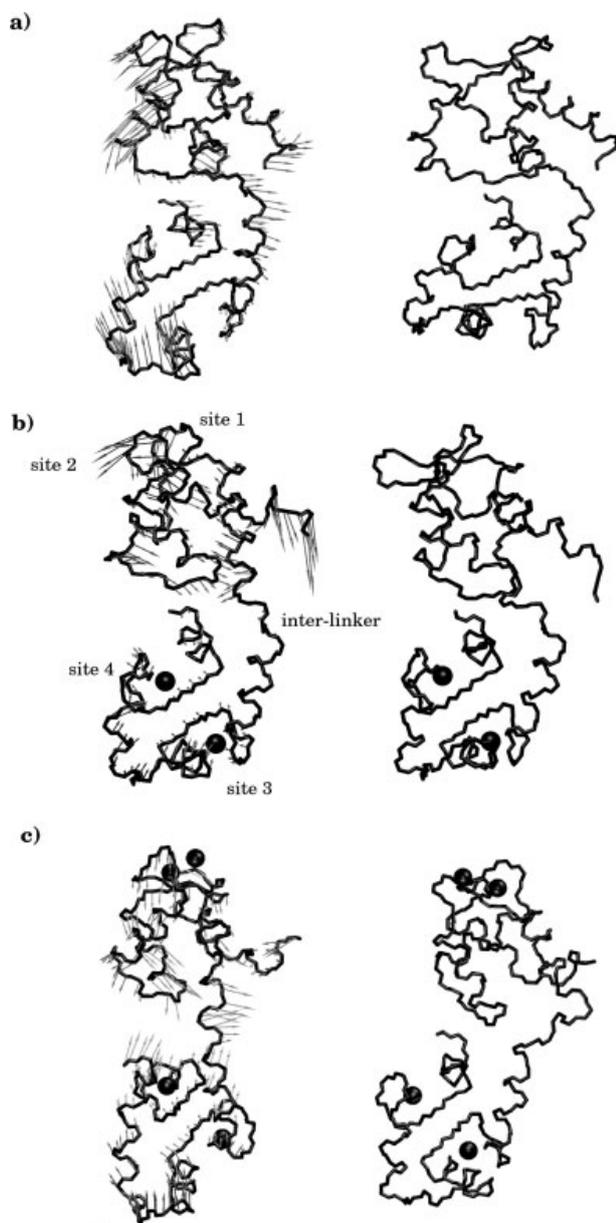
In EF-(2Ca-CaM) and in EF-(4Ca-CaM), the two first modes describe respectively a compaction [Fig. 6(a)] and a torsion [Fig. 6(b)]. Indeed, the helical domain and the CB domain move closer to each other along the  $x$  axis in the first mode, while they rotate relatively to each other around the  $x$  axis in the second mode. By

contrast, the two first modes of EF in EF-(Apo-CaM) describe deformations taking place in the most flexible regions, namely the catalytic site, loops in the helical domain and switch A.

These results suggest that EF motions are more collective in EF-(2Ca-CaM) than in the other complexes, which is the contrary of what was observed for CaM. The following interpretation can then be drawn: when the complex is stable,  $2\text{Ca}^{2+}$ -CaM makes extensive contacts with EF and its mobility is reduced; destabilization of the EF-CaM interaction upon calcium binding or removal allows CaM to move independently from EF, in a more collective way. EF adopts the opposite behavior: as the complex is stable with two  $\text{Ca}^{2+}$ , EF has the freedom to fluctuate around an equilibrium position through global motions of the helical domain and the catalytic core; while the complex is destabilized (EF-(Apo-CaM) and EF-(4Ca-CaM)), EF mobility appears constrained and it tends to display more local motions inside its flexible regions.

#### Calmodulin flexibility, calcium affinity, and domain coupling in calmodulin

Atomic fluctuations per residue were calculated for CaM [Fig. 7(a,b)] as well as for EF [Fig. 7(c)] with the



**Figure 5**

Extreme conformations obtained by deforming the mean backbone conformation of CaM in (a) EF-(Apo-CaM), (b) EF-(2Ca-CaM), and (c) EF-(4Ca-CaM) along the first mode. On the left is represented the “starting” structure with gray arrows displaying the components of the first mode on each atom. On the right is represented the “final” structure obtained by moving along the first mode. (a) In EF-(Apo-CaM), the first mode describes deformations of site 2 (residues 45–59) within the N-terminal domain, of the region between helices VI and VII (residues 108–119) within the C-terminal domain, and of the interlinker (residues 74–80). (b) In EF-(2Ca-CaM), the first mode describes deformations of the N-terminal tail (residues 5–9) and site 2 (residues 44–60) within the N-terminal domain. (c) In EF-(4Ca-CaM), the first mode describes deformations of site 1 (residues 21–40) in the N-terminal domain, of the interlinker (residues 74–79), and locally within the C-terminal domain (residues 113–115 and 146–147).

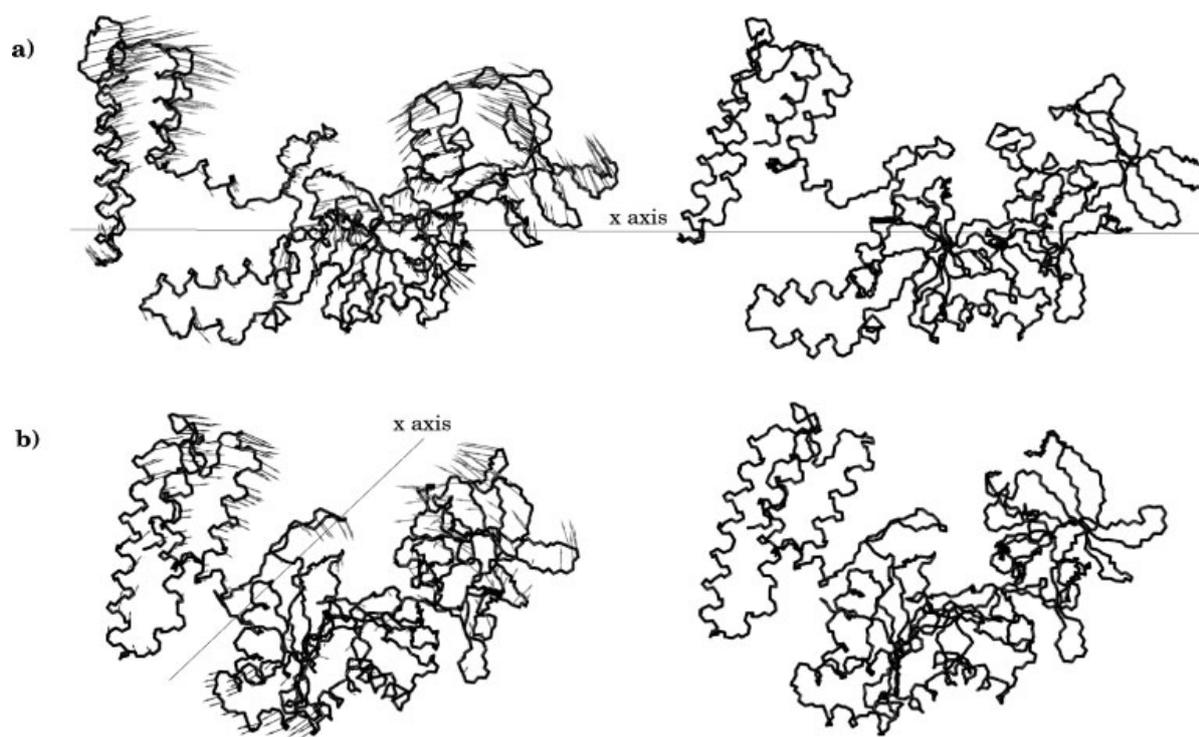
different levels of calcium. Changes were small for EF and will not be further described. In CaM, the N-terminal lobe (residues 5–64) interacts weakly with EF<sup>32,42</sup> and is flexible, in agreement with its larger RMSD [Fig. 1(c)]. Site 2 rigidifies after binding a calcium ion in EF-(4Ca-CaM) while the region upstream (residues 29–47) becomes flexible as a counterbalancing effect. By contrast site 1 coordination loop (residues 16–29) slightly gets more flexible when the level of calcium increases. The interlinker region also shows some flexibility, especially for EF-(4Ca-CaM). Although the C-terminal lobe (residues 93–147) makes extensive contacts with EF, the loop connecting the two EF-hands (residues 111–118) displays atomic fluctuations above 1.5 Å, and these fluctuations increase in the apo form.

NMR studies of CaM recently revealed that CaM internal dynamics is affected by peptide binding and highlighted the importance of conformational entropy in target recognition.<sup>67</sup> The crucial role of side-chains dynamics changes was illustrated in the smMLCKp-CaM and CaMKK $\alpha$ p-CaM complexes.<sup>10,13</sup> To decide whether such changes can be found in EF-bound CaM, fluctuations upon calcium binding or removal were analyzed along the simulations [Fig. 7(b)]. Overall side-chains and backbone fluctuations show similar profiles and changes according to calcium level are of small magnitude: CaM side-chains dynamics is not especially important in the EF-CaM interaction. Nevertheless, the increasing mobility of the N-terminal domain in EF-(4Ca-CaM) and of the C-terminal domain in EF-(Apo-CaM) could indicate an conformational entropic compensation for the loss of affinity of CaM for EF.

In free-CaM, calcium binding provokes the opening of the helix-loop-helix EF-hands.<sup>19,29,30</sup> Interhelices angles increase from 42–52° in Apo-CaM<sup>29</sup> to 70–95° in Ca<sup>2+</sup>-CaM.<sup>15,16</sup> This exposes an ensemble of hydrophobic residues to solvent, which can then interact with target peptides.<sup>31</sup>

In the simulations of EF-CaM, both C-terminal EF-hands (sites 3 and 4) are open irrespective to the level of calcium, with interhelices angles ranging from 71.0 to 105.6° (Table IV). On the other hand, conformations of the N-terminal EF-hands (sites 1 and 2) vary upon calcium binding or removal (Table IV). Overall site 2 shows a greater variability than the other sites, as Shepherd and Vogel<sup>19</sup> showed in their 20-ns simulation of free Ca<sup>2+</sup>-CaM. It fluctuates between a closed and a semi-open conformation in EF-(Apo-CaM) and EF-(2Ca-CaM), and slightly opens upon calcium binding in EF-(4Ca-CaM), switching between a semi-open and an open conformation.

The exposition of hydrophobic patches (defined by Yang *et al.*<sup>31</sup>) was evaluated (Table V) by calculating their solvent accessible surface area (SASA). In agreement with the X-ray structures 1K93, 1XFX, and 1XFY, SASA



**Figure 6**

Extreme conformations obtained by deforming the mean backbone conformation of EF in EF-(2Ca-CaM) along (a) the first mode and (b) the second mode. On the left is represented the “starting” structure with gray arrows displaying the components of the first mode on each atom. On the right is represented the “final” structure obtained by moving along the first mode.

values in the C-terminal domain are close to those of free  $\text{Ca}^{2+}$ -loaded CaM<sup>16</sup> (PDB entry: 1CLL), whereas SASA values in the N-terminal domain are close to those of free Apo-CaM<sup>68</sup> (PDB entry: 1CFD), suggesting that the formation of the EF-CaM complex selects the apo N-terminal and the  $\text{Ca}^{2+}$ -loaded C-terminal conformations, leading as expected to EF-(2Ca-CaM) as the most stable complex. Upon calcium binding, however, the N-terminal domain shows a significant gain in exposure of 12%, consistent with the opening of site 2.

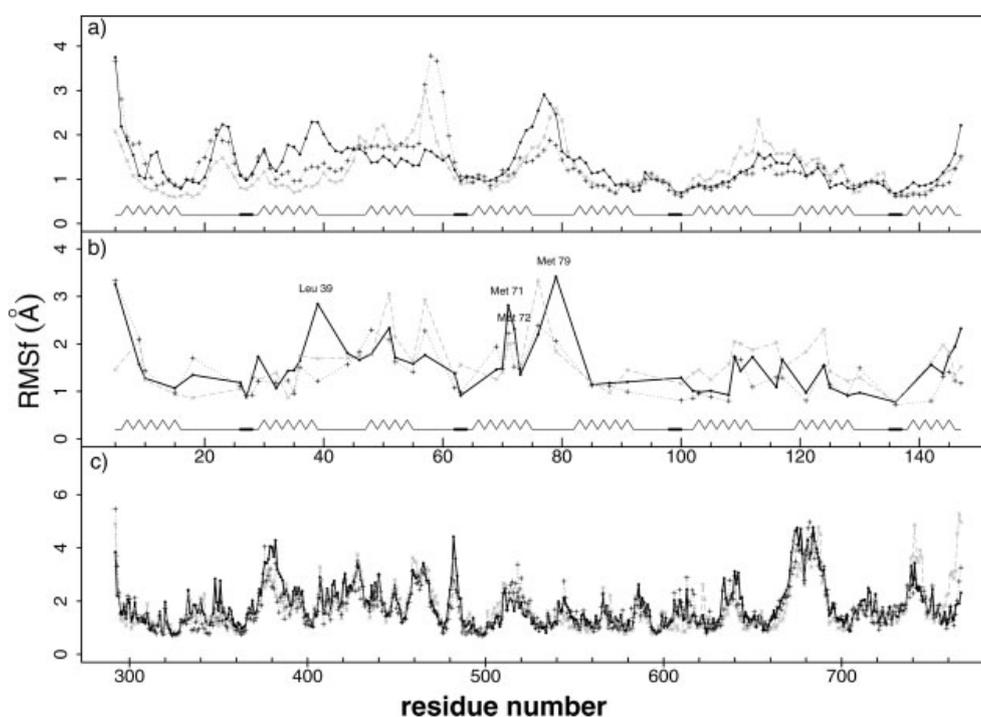
Each calcium is coordinated by residues 1, 3, 5, 7, and 12 in the sequence of the interhelical loop of each EF-hand. The coordination and its electrostatic energies were analyzed (Fig. 8 and Table VI). The electrostatic binding energy  $\Delta E_{\text{ele}}$  was calculated from the Coulombic interaction of calcium with the coordination loop, and is certainly a crude approximation of the experimental binding affinity. But,  $\Delta E_{\text{ele}}$  gives an overall estimation of calcium coordination and allows therefore a sensitive comparison of the different sites.

Figure 8(a) displays the percentage of the simulation time spent with a given number of oxygens coordinat-

ing the  $\text{Ca}^{2+}$  ion in each EF-hand. C-terminal calcium ions (sites 3 and 4) are fully coordinated during 71% and 65% of EF-(2Ca-CaM) simulation time and are always coordinated by at least four oxygens. This proves the stability of the sites. Electrostatic binding energy (Table VI) of site 4 is significantly higher than that of site 3, in agreement with Ulmer *et al.*<sup>43</sup> finding on affinity.

Coordination percentages [Fig. 8(a)] and electrostatic binding energies (Table VI) in EF-(4Ca-CaM) do not significantly change for sites 3 and 4, suggesting that the stability of these sites is not altered by calcium binding on sites 1 and 2. This confirms the hypothesis of Ulmer *et al.*<sup>43</sup> that binding calcium on N-CaM does not affect C-CaM affinity for calcium.

By contrast, sites 1 and 2 in the N terminal domain are incompletely coordinated. Most of the time (84%), site 1 makes only 3 coordinations out of 6. Coordination in site 2 is slightly higher, involving four or five oxygens during 35% and 59% of the time respectively. This is consistent with the stiffness observed in site 2 but not in site 1 upon calcium binding [Fig. 7(a)], and with the



**Figure 7**

Fluctuations per residue considering the last 12 ns of the EF-CaM simulations: (a,b) all-atoms and methyl-bearing side-chains fluctuations of calmodulin—the secondary structure is displayed at the bottom, (c) EF atomic fluctuations. The EF-(Apo-CaM) trajectory is shown in grey (long dotted line), the EF-(2Ca-CaM) and EF-(4Ca-CaM) trajectories are shown in black (short dotted line and solid line, respectively).

more favorable electrostatic binding energy of  $\text{Ca}^{2+}$  in site 2 than in site 1 (Table VI).

Missing coordinating ligands can be identified in Figures 8(b,c). Glu 31 (black and grey) and Thr 26 (purple) are observed too far to participate in the coordination of site 1 [Fig. 8(b)], as described by Shen *et al.*<sup>42</sup> on the crystallographic structure of EF-(4Ca-CaM). In site 2 [Fig. 8(c)], Asn 60 (dark green) stays away from the  $\text{Ca}^{2+}$ , as in Shen *et al.*<sup>42</sup> crystallographic structure of EF-(4Ca-CaM). However, these authors found that Glu 67 did not coordinate  $\text{Ca}^{2+}$  in site 2, while both side-

chains carbonyl oxygens of Glu 67 are close to the calcium ion in the simulations, around  $2.7 \pm 0.2 \text{ \AA}$  [Fig. 8(c)]. The affinity of N-CaM for calcium<sup>42,43</sup> is reduced by EF locking the domain in a closed conformation and preventing both sites 1 and 2 to fully bind calcium. This is in total agreement with SASA and EF-hand opening.

To test whether  $4\text{Ca}^{2+}$ -CaM could pursue toward full opening of the EF-hands when released from EF, the protein was simulated during 15 ns after removal of EF from the EF-(4Ca-CaM) crystallographic structure<sup>42</sup> (PDB entry: 1XFX). Mean RMS deviation on the 3–15 ns inter-

**Table IV**

Average Interhelical Angles in Calcium sites 1, 2, 3, and 4

	Site 1 (°)	Site 2 (°)	Site 3 (°)	Site 4 (°)
EF-(2Ca-CaM) 1K93	45.5 ± 0.2	56.3 ± 0.0	78.1 ± 0.1	88.3 ± 0.8
EF-(2Ca-CaM) 1XFY	45.3 ± 0.0	53.8 ± 0.0	75.4 ± 0.0	83.2 ± 0.0
EF-(4Ca-CaM) 1XFX	45.8 ± 0.0	58.8 ± 0.2	76.4 ± 0.0	83.0 ± 0.0
EF-(Apo-CaM) MD	33.4 ± 3.9	62.6 ± 8.5	79.8 ± 3.8	94.8 ± 4.3
EF-(2Ca-CaM) MD	41.2 ± 2.3	60.2 ± 7.3	74.4 ± 3.4	98.3 ± 3.8
EF-(4Ca-CaM) MD	43.7 ± 7.2	75.7 ± 7.7	80.4 ± 2.7	101.0 ± 4.6

These angles describe the opening of EF hands.

**Table V**Solvent Accessible Surface Area (SASA, Å<sup>2</sup>) Calculated on the Hydrophobic Patches of the N-Terminal and C-Terminal Domains of Calmodulin

	No calcium		Two calcium ions		Four calcium ions	
	N-term	C-term	N-term	C-term	N-term	C-term
Free CaM (X-ray) <sup>16,68</sup>	406	490	—	—	694	784
EF-bound CaM (X-ray) <sup>32</sup>	—	—	641	784	—	—
EF-bound CaM (X-ray) <sup>42</sup>	—	—	468	733	527	766
EF-bound CaM (MD)	443 ± 45	810 ± 38	476 ± 68	835 ± 32	505 ± 42	840 ± 31

Values are calculated on (i) free CaM and on (ii) CaM in complex with EF. Free CaM values are calculated on the 1CDL X-Ray structure of free Apo-CaM<sup>68</sup> and on the 1CLL X-Ray structure of free Ca<sup>2+</sup>-loaded CaM.<sup>16</sup> EF-bound CaM values are calculated on the X-Ray structures of the EF-CaM complex – the 1K93 structure of EF-(2Ca-CaM),<sup>32</sup> and the 1XFY and 1XFX structures of EF-(2Ca-CaM) and EF-(4Ca-CaM),<sup>42</sup> and on the EF-(Apo-CaM), EF-(2Ca-CaM) and EF-(4Ca-CaM) MD trajectories.

val is 7.8 Å. Contrary to what was expected, the simulation did not converge toward the opening of the N-terminal domain. Interhelical angles were rather stable around 40° ± 3° in site 1, 47° ± 5° in site 2, 84° ± 10° in site 3, and 76° ± 3° in site 4. SASA of the hydrophobic patches was about 400 Å<sup>2</sup> in N-CaM and 800 Å<sup>2</sup> in C-CaM. An essential dynamics (ED) analysis allowed to observe a reorientation of the lobes relative to each other (Fig. 9). It appears thus easier for the released 4Ca-CaM to reorientate its lobes than to open its N-terminal domain. This observation agrees with the structure of the free 4Ca<sup>2+</sup>-CaM recently solved by Grabarek,<sup>71</sup> where N-CaM is artificially locked in a closed conformation by a disulfide bond, revealing a two-step calcium binding mechanism. The structure of 4Ca-CaM in the EF-(4Ca-CaM) complex is also intermediate between the closed and open conformations, lacking the final Glu 12 ligand in the N-terminal coordination loops. This state corre-

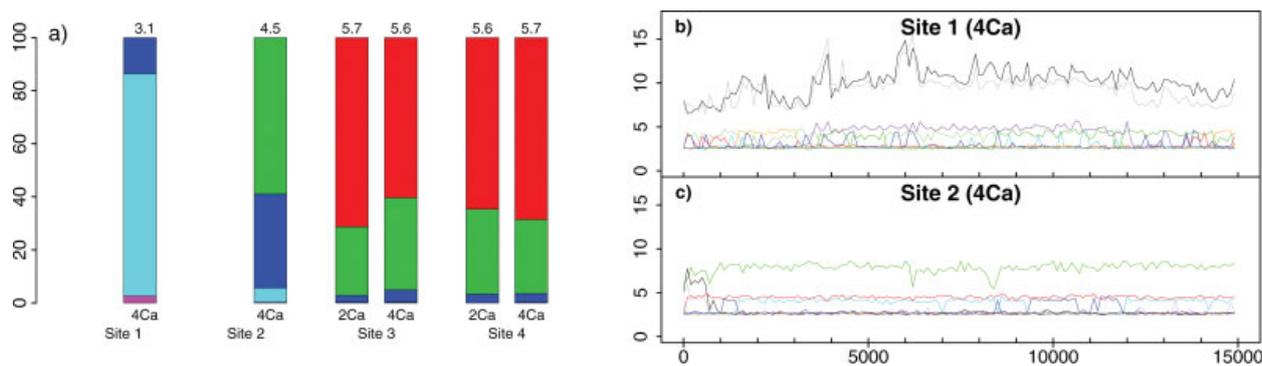
sponds to the first step of dissociation observed in a steered MD.<sup>72</sup>

### Mechanical model of the CaM-EF interaction

A mechanical model of the CaM-EF interaction was derived from the analyses of (i) calmodulin conformation and (ii) the EF-CaM interaction.

The average CaM radii of gyration [Fig. 10(a)] during the simulations range from 18.6 to 19.6 Å with standard deviations of 0.3 Å. These values reveal an intermediate conformation between the extended conformation of Ca<sup>2+</sup>-CaM in crystal (21.9 Å)<sup>16</sup> and the compact conformation of smMLCK peptide-bound CaM (16.1 Å).<sup>17</sup>

In EF-(Apo-CaM) and EF-(2Ca-CaM), CaM tends to become more compact during the trajectories than in the crystalline environment<sup>32,42</sup> (Table VII). In EF-(4Ca-CaM) however, CaM average radius of gyration is slightly

**Figure 8**

Calcium coordination in (b,c) sites 1 and 2 of EF-(4Ca-CaM), (a) sites 1–4 of EF-(2Ca-CaM) and EF-(4Ca-CaM). (a) Each colour stands for a level of calcium coordination: 2 oxygens in purple, 3 oxygens in light blue, 4 oxygens in dark blue, 5 oxygens in green, and full coordination with 6 oxygens in red. At the top is written the mean coordination level. Oxygens are considered to coordinate the calcium ion if they are located at less than 2.8 Å from it, consistent with typical calcium coordination distances found in the literature.<sup>69,70</sup> (b,c) Coordination distances in (b) site 1 and (c) site 2 along the EF-(4Ca-CaM) trajectory. In site 1, coordination distances are calculated between the calcium ion and D20.Oδ1 (red), D20.Oδ2 (orange), D22.Oδ1 (dark blue), D22.Oδ2 (light blue), D24.Oδ1 (dark green), D24.Oδ2 (light green), T26.O (purple), E31.Oε1 (black), and E31.Oε2 (grey). In site 2, coordination distances are calculated between the calcium ion and D56.Oδ1 (red), D56.Oδ2 (orange), D58.Oδ1 (dark blue), D58.Oδ2 (light blue), N60.Oδ1 (dark green), T62.O (purple), E67.Oε1 (black), and E67.Oε2 (grey).

**Table VI**Electrostatic Binding Energies  $\Delta E_{ele}$  of Calcium Ions in Each Site of EF-(2Ca-CaM) and EF-(4Ca-CaM), in kcal mol<sup>-1</sup>

	Site 1	Site 2	Site 3	Site 4
EF-(2Ca-CaM)	—	—	-1245 ± 102	-1444 ± 102
EF-(4Ca-CaM)	-1160 ± 100	-1294 ± 100	-1265 ± 100	-1475 ± 100

larger, of  $19.6 \pm 0.4$  Å. Indeed, as described in a previous section, the N-terminal domain slightly opens upon calcium binding in EF-(4Ca-CaM).

The VDA formed by the EF-hands centres of mass characterizes the relative orientation of the lobes [Fig. 10(b)]. CaM adopts an acute or right VDA (between  $-100$  and  $-60^\circ$ ) during the simulations, as in other target-CaM complexes,<sup>17</sup> whereas free CaM displays an obtuse VDA.<sup>16,18–20</sup> For EF-(4Ca-CaM) and EF-(Apo-CaM), the VDA was around  $-70^\circ$  and  $-94^\circ$ , respectively, while it fluctuates between the two other profiles around  $-84^\circ$  for EF-(2Ca-CaM). Thus, binding (respectively removing) two calcium ions stabilizes the orientation of CaM lobes, toward a smaller (respectively larger) VDA value.

To characterize the behavior of the interlinker region, its secondary structure was determined using DSSP<sup>58</sup> and averaged over the whole simulations [Fig. 10(d)]. The percentages of simulation time spent in helical secondary structure, hydrogen bonded turn, or random coil are given for each residue. EF-(2Ca-CaM) displays the highest average helical content (74.6%), consistent with the stability of the RMSD (Fig. 1).

The central region of the interlinker lacks of helical secondary structure: the nonhelical structures extends from residues 73 to 82 in EF-(Apo-CaM), 71 to 84 in EF-(2Ca-CaM), and 69 to 84 in EF-(4Ca-CaM). These residues switch from helix to turn or random coil in different proportions according to the level of calcium. Removing two calcium ions from the C-terminal domain provokes a partial unfolding of the interlinker (residues 76–78) in EF-(Apo-CaM) whereas binding two calcium ions on the N-terminal domain provokes a partial transition to turn (residues 73–82) in EF-(4Ca-CaM).

We evaluated the bending of the interlinker as the angle between helices IV and V [Fig. 10(c)]. While the angle of EF-(Apo-CaM) and EF-(2Ca-CaM) stays stable around  $132^\circ$  and  $126^\circ$ , it stabilizes around  $109^\circ$  for EF-(4Ca-CaM). This more pronounced bending might be the result of the conversion of the helical secondary structures to turns, with a hinge formed by the central residues.

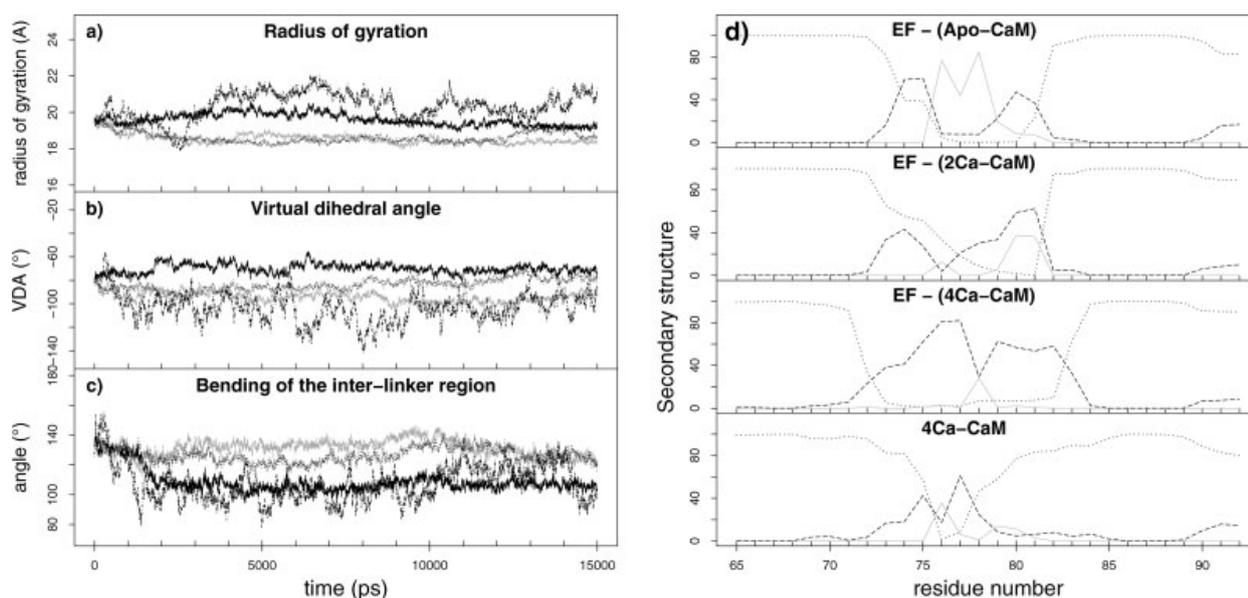
Previous MD simulations of free Ca<sup>2+</sup>-loaded CaM also showed a loss of helical structure in the interlinker<sup>18,19,65</sup> associated with the bending. Moreover, Shepherd and Vogel<sup>19</sup> found in their 20-ns simulation

that partial unfolding of residues 75–81 participated in the bending of the interlinker and in the associated compaction event of CaM. We do not observe such a compaction as the bending is not sufficient to counter-balance the expansion of the N-terminal domain in EF-(4Ca-CaM).

It was observed in the literature<sup>19,20</sup> that the bending or unwinding of the interlinker region is loosely stabilized by some hydrogen bonds formed between central residues and the N-terminus. Although CaM sequence in the EF-CaM complex lacks residues 1–4, we found similar stabilizing interactions in EF-(Apo-CaM) as well as in EF-(4Ca-CaM), but not in EF-(2Ca-CaM), consistent with the greater flexibility of CaM N-tail in this complex [Fig. 7(a)]. Indeed, in EF-(Apo-CaM), hydrogen bonds T5.N-T77.O, T5.N-T79.O, T5.H $\gamma$ 1-T77.O, T5.O $\gamma$ 1-T79.H $\gamma$ 1 appear to stabilize the unwinding of the inter-

**Figure 9**

First mode of the essential dynamics analysis of the simulation of 4Ca-CaM released from EF. Gray arrows display the components of the mode on each backbone atom.

**Figure 10**

(a–c) Radius of gyration (a), VDA (b), and bending of the interlinker (c) of calmodulin along the simulations. The EF-(Apo-CaM) trajectory is shown in gray (long dotted line), the EF-(2Ca-CaM) and EF-(4Ca-CaM) trajectories are shown in black (dotted and thick line, respectively), the 4Ca-CaM trajectory is shown in black (dotted and long dashed line). (d) Secondary structure averaged over the whole simulations. The percentage of simulation time spent by each residue in each secondary structure type is given. The black short dotted, black long dotted, and grey solid lines represent the helical, turn, and random coil types, respectively.

linker. In EF-(4Ca-CaM), the Q8.Oε1-T79.Hγ1 hydrogen bond, formed in the 8–15 ns interval, and hydrophobic interactions between N-terminus and interlinker residues participate in the bending of the interlinker.

Therefore, analysis of the radius of gyration, of the VDA, of the bending of the interlinker, and the of interactions within CaM show that CaM conformations in EF-(Apo-CaM) and EF-(2Ca-CaM) are similar whereas that in EF-(4Ca-CaM) is significantly different. One should notice that EF-(2Ca-CaM) and EF-(4Ca-CaM) conformations in crystal were quite similar, and that MD trajectories allow further exploration of the conformational space.

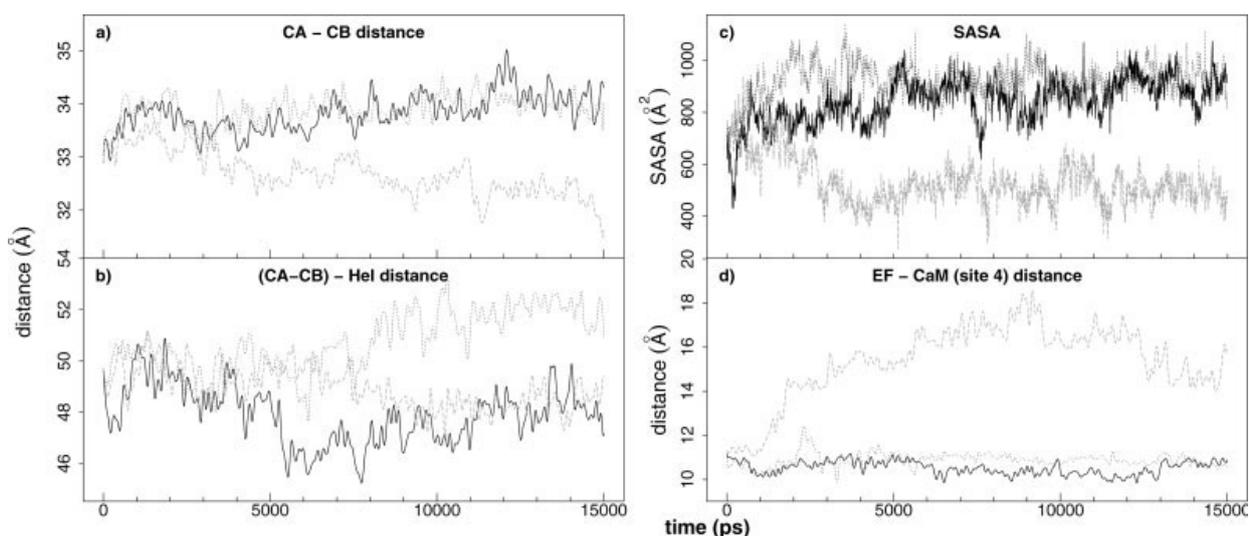
The analysis of EF conformation in the EF-CaM complex reveals a calcium-dependent behavior different from

that of CaM. Adding two calcium ions to EF-(2Ca-CaM) has little effect on EF conformation, whereas removing two calcium ions affects the whole protein, and especially the geometry of the catalytic site. As showed in Figure 11(a,b), the time evolution of some characteristic distances between the centres of mass of EF domains are similar for EF-(2Ca-CaM) and EF-(4Ca-CaM) but are often significantly different for EF-(Apo-CaM). Opening of the EF catalytic site was monitored through the mean distance between the centres of mass of residues from Switch B (H577, G578, T579, D582, N583, E588, F586) and residues from CB (H351, K353, S354, K372) (Table VIII). Indeed, this distance has a similar evolution, in the 13–15 Å range, for EF-(2Ca-CaM) and EF-(4Ca-CaM), while it decreases slightly, around 11.3 Å for EF-(Apo-CaM). N329, K353,

**Table VII**

Conformational Parameters of CaM (Radius of Gyration, Virtual Dihedral Angle (VDA), and Bending of the Interlinker Region) Calculated on EF-CaM X-ray Structures<sup>32,42</sup> and on MD Trajectories

		Radius of gyration (Å)	VDA (°)	Bending (°)
X-ray PDB	EF-(2Ca-CaM)(1K93) <sup>32</sup>	19.5	-80.5	124
	EF-(2Ca-CaM)(1XFY) <sup>42</sup>	19.2	-76.9	125
	EF-(4Ca-CaM)(1XFX) <sup>42</sup>	19.4	-77.3	120
MD trajectories	EF-(Apo-CaM)	18.5 ± 0.2	-95.2 ± 3.8	132 ± 5
	EF-(2Ca-CaM)	18.5 ± 0.2	-82.7 ± 4.6	126 ± 5
	EF-(4Ca-CaM)	19.6 ± 0.4	-69.8 ± 3.7	106 ± 4

**Figure 11**

Characteristic features of EF conformations along the trajectories: (a) distance between the centers of mass of CA (residues 292 to 349 and 490 to 622) and CB (residues 350 to 489), (b) distance between the centers of mass of the catalytic core (grouping CA and CB) and Hel (residues 660 to 767), (c) SASA of the catalytic site (H351, K353, S354, K372, R329, K346, L348, D491, D493, H577, G578, T579, D582, N583, E588, F586, T548), (d) distance between the center of mass of EF C-tail (residues 758 to 767) and the center of mass of the coordinating ligands in site 4 of CaM.

K346 establish hydrogen bonds with the ligand in Shen *et al.*<sup>42</sup> and the triangle distances between these residues (Table VIII) is thus also indicative of the opening of EF catalytic site during the trajectories. In EF-(Apo-CaM) N329.C $\gamma$  is closer to K353.N $\zeta$  and to K346.N $\zeta$ , than in the other two trajectories. Calculations of the accessible surface area (SASA) in the catalytic site also support [Fig. 11(c), Table VIII] the hypothesis of the active site collapse in EF-(Apo-CaM). The SASA values calculated from the simulations of EF-(2Ca-CaM) and EF-(4Ca-CaM) are slightly larger, but in qualitative agreement with the X-ray structures 1XFX and 1XFY of the complexes.

To characterize contacts between EF and CaM in the complex, hydrophobic interactions and hydrogen bonds established for at least 50% of the simulation time were analyzed (Table IX). The C-terminal domain of calmodulin makes extensive contacts with EF: an average of 40 C-terminal residues are involved in either hydrophobic interactions or hydrogen bonds, whereas there are only about five N-terminal residues interacting with EF (Table IX). In the N-terminal domain, the hydrogen bonds form the majority of contacts for EF-(2Ca-CaM), whereas they are a minority for EF-(Apo-CaM) and EF-(4Ca-CaM). The changes observed in the N-terminal

**Table VIII**

Average Characteristic Distances and Solvent Accessible Surface (SASA) Within EF Catalytic Site Including the Residues (H351, K353, S354, K372, R329, K346, L348, D491, D493, H577, G578, T579, D582, N583, E588, F586, T548)

	SB-CB (Å)	N329.C $\gamma$ K353.N $\zeta$ (Å)	N329.C $\gamma$ -K346.N $\zeta$ (Å)	SASA (Å <sup>2</sup> )
Apo-EF 1K8T <sup>32</sup>	—	11.3	12.2	786
EF-(2Ca-CaM) 1K93 <sup>32</sup>	13.6	7.5	9.6	703
EF-(2Ca-CaM) 1XFV <sup>42</sup>	15.1	11.1	10.1	833
EF-(4Ca-CaM) 1XFX <sup>42</sup>	15.2	11.2	10.3	827
EF-(Apo-CaM) MD	11.3 $\pm$ 0.5	6.4 $\pm$ 0.8	11.0 $\pm$ 1.0	501 $\pm$ 58
EF-(2Ca-CaM) MD	14.4 $\pm$ 1.1	11.1 $\pm$ 1.8	13.1 $\pm$ 1.2	928 $\pm$ 55
EF-(4Ca-CaM) MD	13.7 $\pm$ 0.9	11.4 $\pm$ 1.9	13.7 $\pm$ 1.6	866 $\pm$ 70

The SB-CB distance is calculated between the centers-of-mass of residues H577, G578, T579, D582, N583, E588, and F586 from SB and residues H351, K353, S354, and K372 from CB. Residues N329, K346, K353 were chosen as they establish in the X-ray structure<sup>42</sup> a network of hydrogen bonds.

**Table IX**

Number of CaM Residues that Are Involved in Hydrogen Bonds with EF/  
Number of CaM Residues that Are Involved in Hydrophobic Interactions with  
EF (for at least 50% of the Simulation Time)

	EF-(Apo-CaM)	EF-(2Ca-CaM)	EF-(4Ca-CaM)
Helix I	4/6	4/3	3/6
Helix II	2/4	1/1	0/2
C-term	12/28	15/29	16/28

Residues are grouped by blocks: Helix I comprises helix I of site 1 (residues 5–22), Helix II comprises helix II of site 1 (residues 23–42), C-term comprises helices V to VIII (residues 80–147).

domain agree with the conformational variability of the reference structures (Fig. 3) in the convergence analysis. The balance between hydrogen bonds and hydrophobic contacts may thus be related to the specificity of the EF-CaM interaction.

In the crystallographic structures of EF-CaM with 2 and 4  $\text{Ca}^{2+}$ ,<sup>32,42</sup> a key salt bridge was formed between Asp 647 from the switch C of EF and Arg 90 from the interlinker of CaM. Two salt bonds between O $\delta$ 2 of Asp 647 and the N $\epsilon$  and N $\eta$ 2 of Arg 90 are stable along the EF-(2Ca-CaM) as well as the EF-(Apo-CaM) trajectories. In EF-(4Ca-CaM), side-chains of the two residues rapidly moved away from each other, but both salt bonds are formed again around 12 ns up to the end of the trajectory.

Figure 11(d) reveals that the removal of calcium from C-CaM in EF-(Apo-CaM) allows EF C-tail to move away from site 4 (11.1 Å at  $t = 0$  and around 16.0 Å in the 3–15 ns interval). On the contrary, the distance stays stable, around 10.5 Å for EF-(2Ca-CaM) and EF-(4Ca-CaM).

## CONCLUSION

In this study, we focused on the conformational plasticity of CaM in the EF-CaM complex and we evaluated the impact of calcium binding or removal on the stability of the complex. We performed three 15-ns MD simulations of the EF-CaM complex corresponding to three different levels of calcium. Such simulations exhibit a much longer time period and more accurate conditions (periodic boundaries, PME) than the early 350-ps simulations of EF-(2Ca-CaM).<sup>46</sup> Values of radius of gyration, VDA, EF-hands opening, and hydrophobic patches calculated on the EF-(2Ca-CaM) and EF-(4Ca-CaM) trajectories were found in good agreement with measurements on the corresponding experimental X-ray structures.<sup>32,42</sup> The EF-(2Ca-CaM) trajectory displays the most stable features (convergence, calcium coordination, essential dynamics), in agreement with experiments.

The trajectories presented here have been validated through their comparison with X-ray crystallographic structures. They can consequently be used: (i) to draw a

more dynamical picture of the EF-(2Ca-CaM) and EF-(4Ca-CaM) complexes, (ii) to give reasons for the experimental instability of the EF-(Apo-CaM) complex.

The most striking dynamics feature is the role played by the bending motion of CaM interlinker. Indeed, fluctuations of EF-bound CaM are mostly the result of deformations inside the N-terminal domain and bending of the interlinker region. This motion is important for the fitting of the size of CaM to the distance between EF helical and CB domains. In EF-(4Ca-CaM), calcium binding provokes the expansion of the N-terminal domain and the bending of the interlinker, from 126 down to 109°. The exposure of hydrophobic patches promotes contacts between the interlinker region and the N-tail of CaM, which stabilizes the bending of the interlinker.

Through the bending of the interlinker, CaM plays the role of a spring which maintains EF in an open conformation, inducing the activation of the adenylyl cyclase catalytic site. The two  $\text{Ca}^{2+}$  complexation level gives the best fit to EF open conformation: indeed, EF-(Apo-CaM) and EF-(4Ca-CaM) both display a deformation of the interlinker.

The N- and C-terminal lobes of CaM are locked by EF in configurations similar to closed-like and open-like conformations, whatever the level of calcium complexation. Nevertheless, limited opening of the N-terminal domain of CaM was observed in EF-(4Ca-CaM) and illustrates the general opening mechanism of CaM upon calcium binding. Site 1 only reaches the first step of calcium binding and its calcium coordination loop lacks the final Glu 12 ligand. Consequently N-CaM is trapped in an intermediate closed conformation, from which it cannot escape.

The presence of EF not only disturbs CaM affinity for calcium but also the coupling mechanism between the two CaM lobes through the interlinker region. EF allows coupling between CaM lobes, but only in one direction, that is from C-CaM to N-CaM. Indeed, binding two calcium ions on N-CaM provokes only changes in the fluctuations of both the interlinker and the N-terminal domain itself. On the other hand, removing two calcium ions from the C-CaM provokes changes in the fluctuations of the C-terminal domain itself and also in the N-terminal domain. This asymmetry in coupling is also observed in the variation of hydrophobic patches. The N-terminal hydrophobic patches of CaM gain almost the same surface (29 Å<sup>2</sup>) upon calcium binding than the one they loose (33 Å<sup>2</sup>) in EF-(Apo-CaM). On the other hand, the C-terminal patches gain about 5 Å<sup>2</sup> in EF-(4Ca-CaM) and loose about 25 Å<sup>2</sup> upon calcium removal. This might be a consequence of a stronger interaction between EF and C-CaM than with N-CaM.

The removal of all calciums was also studied in the present work, which gives information about the EF-(Apo-CaM) system, for which no structure has been solved to date. Ulmer *et al.*<sup>43</sup> proposed a model where

Apo-CaM only interacts with EF through its N-terminal domain. Indeed, we observed that the interaction between EF and the C-terminal domain of CaM is weakened in EF-(Apo-CaM).

However, our analyses of radius of gyration, VDA, opening of the sites, and exposure of hydrophobic patches showed that the conformation of CaM in EF-(Apo-CaM) is closer to that of EF-(2Ca-CaM) than that of EF-(4Ca-CaM). This suggests that once the C-terminal domain is inserted and blocked between the helical domain and the catalytic core of EF, the stability of the complex is not dramatically affected upon calcium removal. An energetic barrier opposes the formation of EF-(Apo-CaM) but there is also a symmetric energetic barrier opposing the dissociation of the complex upon removal of the C-terminal calcium ions. The results presented are mainly influenced by the second barrier.

Besides, reactions to calcium binding or removal suggest a mechanism for the destabilization of the EF-(Apo-CaM) complex. Indeed, our study suggests that the binding of calcium on N-CaM tends to open the domain, but this opening is restrained by EF helical domain, which keeps the catalytic site open and maintains its contacts with CaM site 1. On the other hand, C-CaM is deeply inserted between the catalytic core and the helical domain of EF, and cannot relax. The removal of  $\text{Ca}^{2+}$  ions from C-CaM will induce a tension toward EF, and thus explains the deformations observed in EF bound to Apo-CaM, which bring the EF domains closer and induce a collapse of the catalytic site.

The present work brings a contribution to the global knowledge of CaM conformational plasticity and describes its molecular interaction with a pathological agent. By giving clues to a better understanding of the mechanisms played in the EF-CaM complex, it enables to better characterize the CaM molecular features responsible for its regulation of many targets. It also permits to define what makes the specificity of CaM recruitment by a toxin, compared to the recognition and activation of target proteins in cellular signalling pathways.

## ACKNOWLEDGMENTS

We thank the French Ministry of Defence (DGA/MRIS), the National Scientific Research Centre (CNRS) and the Pasteur Institute for fundings and support. We also acknowledge the French National Calculations Centres IDRIS (Orsay) and CINES (Montpellier) for computational resources. JY acknowledges the FAPERJ (Brazil) for an undergraduate fellowship.

## REFERENCES

- Ikura M, Ames JB. Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: two ways to promote multifunctionality. *Proc Natl Acad Sci USA* 2006;103:1159–1164.
- Ikura M, Marion D, Kay LE, Shih H, Krinks M, Klee CB, Bax A. Heteronuclear 3D NMR and isotopic labeling of calmodulin. Towards the complete assignment of the  $^1\text{H}$  NMR spectrum. *Biochem Pharmacol* 1990;40:153–160.
- Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 1990;29:4659–4667.
- Ikura M, Kay LE, Bax A. Improved three-dimensional  $^1\text{H}$ - $^{13}\text{C}$ - $^1\text{H}$  correlation spectroscopy of a  $^{13}\text{C}$ -labeled protein using constant-time evolution. *J Biomol NMR* 1991;1:299–304.
- Bax A, Ikura M. An efficient 3D NMR technique for correlating the proton and  $^{15}\text{N}$  backbone amide resonances with the alpha-carbon of the preceding residue in uniformly  $^{15}\text{N}/^{13}\text{C}$  enriched proteins. *J Biomol NMR* 1991;1:99–104.
- Friedrichs MS, Mueller L, Wittekind M. An automated procedure for the assignment of protein  $^1\text{HN}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}\alpha$ ,  $^1\text{H}\alpha$ ,  $^{13}\text{C}\beta$  and  $^1\text{H}\beta$  resonances. *J Biomol NMR* 1994;4:703–726.
- Lee AL, Sharp KA, Kranz JK, Songn XJ, Wandn AJ. Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry* 2002;41:13814–13825.
- Petrone P, Pande VS. Can conformational change be described by only a few normal modes? *Biophys J* 2006;90:1583–1593.
- Chen K, Ruan J, Kurgan LA. Prediction of three dimensional structure of calmodulin. *Protein J* 2006;25:57–70.
- Frederick KK, Kranz JK, Wand AJ. Characterization of the backbone and side chain dynamics of the CaM-CaMKIIp complex reveals microscopic contributions to protein conformational entropy. *Biochemistry* 2006;45:9841–9848.
- Qin Z, Wertz SL, Jacob J, Savino Y, Cafiso DS. Defining protein-protein interactions using site-directed spin-labeling: the binding of protein kinase C substrates to calmodulin. *Biochemistry* 1996;35:13272–13276.
- Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat Struct Biol* 2000;7:72–77.
- Marlow MS, Wand AJ. Conformational dynamics of calmodulin in complex with the calmodulin-dependent kinase kinase  $\alpha$  calmodulin-binding domain. *Biochemistry* 2006;45:8732–8741.
- Ishida H, Vogel HJ. Protein-peptide interaction studies demonstrate the versatility of calmodulin target protein binding. *Protein Pept Lett* 2006;13:455–465.
- Babu YS, Bugg CE, Cook WJ. Structure of calmodulin refined at 2.2 Å resolution. *J Mol Biol* 1988;204:191–204.
- Chattopadhyaya R, Meador WE, Means AR, Quijcho FA. Calmodulin structure refined at 1.7 Å resolution. *J Mol Biol* 1992;228:1177–1192.
- Meador WE, Means AR, Quijcho FA. Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science* 1992;257:1251–1255.
- Wriggers W, Mehler E, Pitici F, Weinstein H, Schulten K. Structure and dynamics of calmodulin in solution. *Biophys J* 1998;74:1622–1639.
- Shepherd CM, Vogel HJ. A molecular dynamics study of  $\text{Ca}^{2+}$ -calmodulin: evidence of interdomain coupling and structural collapse on the nanosecond timescale. *Biophys J* 2004;87:780–791.
- Fiorin G, Biekofsky RR, Pastore A, Carloni P. Unwinding the helical linker of calcium-loaded calmodulin: a molecular dynamics study. *Proteins* 2005;61:829–839.
- Chou JJ, Li S, Klee CB, Bax A. Solution structure of  $\text{Ca}^{2+}$ -calmodulin reveals flexible hand-like properties of its domains. *Nat Struct Biol* 2001;8:990–997.
- Bertini I, Del Bianco C, Gelis I, Katsaros N, Luchina C, Parigi G, Peana M, Provenzani A, Zoroddu MA. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc Natl Acad Sci USA* 2004;101:6841–6846.
- Johnson CK. Calmodulin, conformational states, and calcium signaling. A single-molecule perspective. *Biochemistry* 2006;45:14233–14246.

24. Wilson MA, Brunger AT. The 1.0 Å crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. *J Mol Biol* 2000;301:1237–1256.
25. Fallon JL, Quijcho FA. A closed compact structure of native Ca<sup>2+</sup>-calmodulin. *Structure* 2003;11:1303–1307.
26. Gifford JL, Walsh MP, Vogel HJ. Structures and metal-ion-binding properties of the Ca<sup>2+</sup>-binding helix-loop-helix EF-hand motifs. *Biochem J* 2007;405:199–221.
27. Capozzi F, Casadei F, Luchinat C. EF-hand protein dynamics and evolution of calcium signal transduction: an NMR view. *J Biol Inorg Chem* 2006;11:949–962.
28. Ishida H, Takahashi K, Nakashima K, Kumaki Y, Nakata M, Hikichi K, Yazawa M. Solution structures of the N-terminal domain of yeast calmodulin: Ca<sup>2+</sup>-dependent conformational change and its functional implication. *Biochemistry* 2000;39:13660–13668.
29. Zhang M, Tanaka T, Ikura M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat Struct Biol* 1995;2:758–767.
30. Finn BE, Evenas J, Drakenberg T, Waltho JP, Thulin E, Forsen S. Calcium-induced structural changes and domain autonomy in calmodulin. *Nat Struct Biol* 1995;2:777–783.
31. Yang C, Jas GS, Kuczera K. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *Biochim Biophys Acta* 2004;1697:289–300.
32. Drum CL, Yan SZ, Bard J, Shen YQ, Lu D, Soelaiman S, Grabarek Z, Bohm A, Tang WJ. Structural basis for the activation of anthrax adenyl cyclase exotoxin by calmodulin. *Nature* 2002;415:396–402.
33. Elshorst B, Hennig M, Forsterling H, Diener A, Maurer M, Schulte P, Schwalbe H, Griesinger C, Krebs J, Schmid H, Vorherr T, Carafoli E. NMR solution structure of a complex of calmodulin with a binding peptide of the Ca<sup>2+</sup> pump. *Biochemistry* 1999;38:12320–12332.
34. Schumacher MA, Rivard AF, Bachinger HP, Adelman JP. Structure of the gating domain of a Ca<sup>2+</sup>-activated K<sup>+</sup> channel complexed with Ca<sup>2+</sup>/calmodulin. *Nature* 2001;410:1120–1124.
35. Hoeflich KP, Ikura M. Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell* 2002;108:739–742.
36. Vetter SW, Leclerc E. Novel aspects of calmodulin target recognition and activation. *Eur J Biochem* 2003;270:404–414.
37. Liddington RC. Anthrax: a molecular full nelson. *Nature* 2002;415:373–374.
38. Ascenzi P, Viscia P, Ippolitob G, Spallarossac A, Bolognesic M, Montecuccod C. Anthrax toxin: a tripartite lethal combination. *FEBS Lett* 2002;531:384–388.
39. Collier RJ, Young JA. Anthrax toxin. *Annu Rev Cell Dev Biol* 2003;19:45–70.
40. Abrami L, Reig N, van der Goot FG. Anthrax toxin: the long and winding road that leads to the kill. *Trends Microbiol* 2005;13:72–78.
41. Lee YS, Bergson P, He WS, Mrksich M, Tang WJ. Discovery of a small molecule that inhibits the interaction of anthrax edema factor with its cellular activator, calmodulin. *Chem Biol* 2004;8:1139–1146.
42. Shen Y, Zhukovskaya NL, Guo Q, Florian J, Tang WJ. Calcium-independent calmodulin binding and two-metal-ion catalytic mechanism of anthrax edema factor. *EMBO J* 2005;24:929–941.
43. Ulmer TS, Soelaiman S, Li S, Klee CB, Tang WJ, Bax A. Calcium dependence of the interaction between calmodulin and anthrax edema factor. *J Biol Chem* 2003;278:29261–29266.
44. Chin D, Means AR. Calmodulin: a prototypical calcium sensor. *Trends Cell Biol* 2000;10:322–328.
45. Project E, Friedman R, Nachliel E, Gutman M. A molecular dynamics study of the effect of Ca<sup>2+</sup> removal on calmodulin structure. *Biophys J* 2006;90:3842–3850.
46. Zhao J, Roy SA, Nelson DJ. MD simulations of anthrax edema factor: calmodulin complexes with mutations in the edema factor “switch a” region and docking of 3'-deoxy ATP into the adenyl cyclase active site of wild-type and mutant edema factor variants. *J Biomol Struct Dyn* 2003;21:159–170.
47. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA. AMBER 8. San Francisco: University of California; 2004.
48. Jorgensen WL. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J Am Chem Soc* 1981;103:335–340.
49. Symersky J, Lin G, Li S, Qiu S, Carson M, Schormann N, Luo M. Structural genomics of *Caenorhabditis elegans*: crystal structure of calmodulin. *Proteins* 2003;53:947–949.
50. Aqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. *J Phys Chem* 1990;94:8021–8024.
51. Darden T, York D, Pedersen L. Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–10092.
52. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–3690.
53. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 1977;23:327–341.
54. Lyman E, Zuckerman DM. Ensemble-based convergence analysis of biomolecular trajectories. *Biophys J* 2006;91:164–172.
55. Baysal C, Atilgan AR. Relaxation kinetics and the glassiness of native proteins: coupling of timescales. *Biophys J* 2005;3:1570–1576.
56. Grace User's Guide, Version 5.1. Israel: Weizmann Institute of Science; 2007.
57. Pascual-Ahuir JL, Mehler EL, Weinstein H. Calmodulin structure and function: Implication of arginine in the compaction related to ligand binding. *Mol Eng* 1991;1:231–247.
58. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;12:2577–2637.
59. Sanner MF, Olson AJ, Spehner JC. Fast and robust computation of molecular surfaces. In: Proc 11th ACM Symp Comp Geom C6–C7; June 5–7, 1995, Vancouver, British Columbia, Canada.
60. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. *JMB* 1994;238:777–793.
61. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Prot Eng* 1995;8:127–134.
62. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;14:33–38,27–28.
63. Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comp Graph Stat* 1996;5:299–314.
64. Yang C, Kuczera K. Molecular dynamics simulations of calcium-free calmodulin in solution. *J Biomol Struct Dyn* 2002;19:801–819.
65. Yang C, Jas GS, Kuczera K. Structure and dynamics of calcium-activated calmodulin in solution. *J Biomol Struct Dyn* 2001;19:247–271.
66. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. *Proteins* 1993;17:412–425.
67. Frederick KK, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. *Nature* 2007;448:325–329.
68. Kuboniwa H, Tjandra N, Grzesiek S, Ren H, Klee CB, Bax A. Solution structure of calcium-free calmodulin. *Nat Struct Biol* 1995;2:768–776.
69. Einspahr H, Bugg CE. Calcium and its role in biology. New York: M. Dekker; 1984.
70. Cates MS, Teodoro ML, Phillips GN Jr. Molecular mechanisms of calcium and magnesium binding to parvalbumin. *Biophys J* 2002;82:1133–1146.
71. Grabarek Z. Structure of a trapped intermediate of calmodulin: calcium regulation of ehand proteins from a new perspective. *J Mol Biol*. 2005;346:31351–31366.
72. Kobayashi C, Takada S. Protein grabs a ligand by extending anchor residues: molecular simulation for Ca<sup>2+</sup> binding to calmodulin loop. *Biophys J* 2006;90:3043–3051.

### A.3 Publication décrivant l'effet électrostatique du calcium sur le réseau de résidus du complexe EF-CaM

# Dynamics and Energetics: A Consensus Analysis of the Impact of Calcium on EF-CaM Protein Complex

Elodie Laine, Arnaud Blondel, and Thérèse E. Malliavin\*

Unité de Bioinformatique Structurale, CNRS URA 2185, Département de Biologie, Structurale et Chimie, Institut Pasteur, Paris, France

**ABSTRACT** We have studied the relationship between dynamical correlations and energetic contributions in an attempt to model the transmission of information inside protein-protein complexes. The complex formed between the edema factor (EF) of *Bacillus anthracis* and calmodulin (CaM) was taken as an example, as the formation and stability of the complex depend on the calcium complexation level. The effect of calcium through EF-CaM residue network has been investigated with various approaches: 1), the elastic network model; 2), the local feature analysis; 3), the generalized correlations; and 4), the energetic dependency maps (EDMs), on 15-ns molecular dynamics simulations of the complex loaded with 0, 2, or 4  $\text{Ca}^{2+}$  ions. The elastic network model correctly describes the basic architecture of the complex but is poorly sensitive to the level of calcium compared to the other methods. The local feature analysis allows us to characterize the local dynamics of the complex and the propagation of the calcium signal through CaM. The analyses of global dynamics and energetics—through generalized correlations and EDMs—provide a comprehensive picture of EF-CaM architecture and can be unified by using the concept of residue network connectedness. A medium connectedness, defined as the ability of each residue to communicate with all remaining parts of the complex, is observed for the  $2\text{Ca}^{2+}$  level, which was experimentally identified as the most stable form of EF-CaM. The hierarchy of relative stabilities given by the EDMs sheds a new light on the EF-CaM interaction mechanism described experimentally and supports an organization of the complex architecture centered around nucleation points.

## INTRODUCTION

Functional changes in protein-protein complexes are underlain by conformational changes and concerted motions arising from information transmission through their residue networks (1). In recent years, various methods have been used to explore the collective dynamics of proteins and study their relevance for biological function, among them principal component analysis (PCA) (2–6) and normal mode analysis (1,7–10). However, the relationship between dynamical correlations and energetic contributions has not been studied extensively in protein complexes. The complex formed between the edema factor (EF) of *Bacillus anthracis* and calmodulin (CaM) provides a good test case for the existing methods as the complex stability depends on the number of calciums bound to CaM (2, 3, or 4  $\text{Ca}^{2+}$ ) while the main architecture of the complex remains practically unchanged. Indeed the x-ray structures with three or four calcium ions display no significant change of the complex global architecture. Furthermore, the analysis of molecular dynamics (MD) simulations performed by Laine et al. (11) showed that the perturbation of the EF-CaM complex by adding and removing  $\text{Ca}^{2+}$  ions only changed the local conformations of some parts of the system, and not its overall architecture. It is thus a challenge to understand the propagation of the electrostatic effects of  $\text{Ca}^{2+}$  ions through the structure of EF-CaM.

EF-CaM gives an example of recruitment of an endogenous protein by a pathogen to activate its toxins (12). Indeed,

upon binding of calmodulin, a ubiquitous actor of many calcium-signaling pathways (13–18), the toxin EF of *B. anthracis* undergoes a large conformational transition that leads to the formation and activation of its adenylyl cyclase catalytic site, which transforms ATP into cAMP. The accumulation of cAMP causes cellular dysfunction and allows EF to impair host immune defenses (19).

In the crystallographic structures of the complex (20,21) (Fig. 1), CaM displays an extended conformation—which is unusual in CaM-target complexes, and is loaded with various numbers of calcium ions:  $2\text{Ca}^{2+}$  (Fig. 1 a),  $3\text{Ca}^{2+}$ , and  $4\text{Ca}^{2+}$ . Calmodulin (in red) is inserted between the catalytic core of EF and its helical domain Hel (residues 660–767, in pink). EF catalytic core is composed of domains CA (residues 292–349 and 490–622, in green) and CB (residues 350–489, in lime). The catalytic residues (in licorice) lie at the CA-CB interface. The switch A (domain CA, residues 402–551, in dark green) is in direct contact with CaM. Domain Hel is composed of four helices L, M, N, and O separated by three interhelical loops: inter L-M loop (residues 668–691), inter M-N loop (residues 706–713), and inter N-O loop (residues 738–745). A loop named switch C (residues 623–659, in blue) connects Hel to the catalytic core of EF.

CaM is composed of two globular domains linked by a central flexible helix (residues 65–92) (Fig. 1 b): the N-terminal domain N-CaM (residues 5–64) and the C-terminal domain C-CaM (residues 93–147). Each domain consists in a pair of calcium-binding EF-hands (22,23) which opens upon calcium binding, exposing hydrophobic patches (24–27). However, the affinity and cooperativity of CaM calcium binding sites are affected by the presence of EF

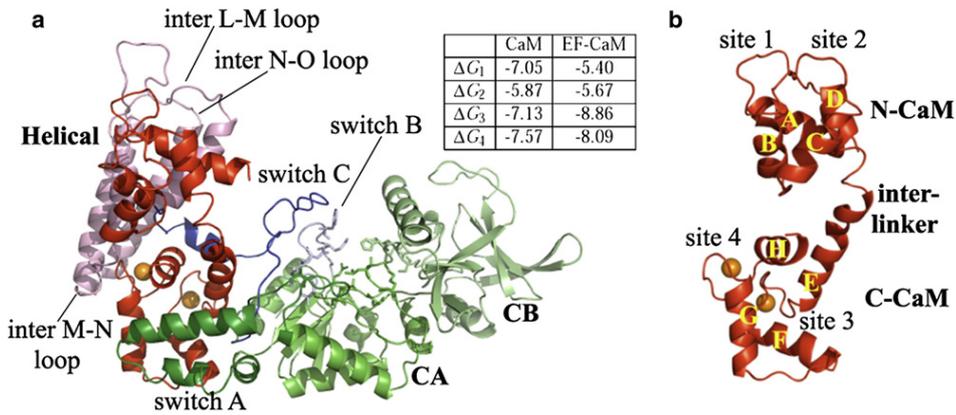
Submitted July 16, 2008, and accepted for publication October 21, 2008.

\*Correspondence: terez@pasteur.fr

Editor: Nathan Andrew Baker.

© 2009 by the Biophysical Society  
0006-3495/09/02/1249/15 \$2.00

doi: 10.1016/j.bpj.2008.10.055



623–659). (b) CaM is composed of eight helices: A, B, and C in N-CaM; D and E in the interlinker; and F, G, and H in C-CaM. Calcium-binding loops are located between helices A and B in site 1, helices C and D in site 2, helices E and F in site 3, and helices G and H in site 4. Sites 3 and 4 are loaded with  $\text{Ca}^{2+}$  ions (spheres in orange). (Inset) Table gives the affinities of the four sites of CaM for calcium calculated from Ulmer et al. (28), in kcal/mol.

(28). In particular, EF locks N-CaM in a closed conformation, reducing its affinity for calcium (see table inset in Fig. 1), whereas it binds C-CaM in an open conformation, loaded with calcium (11). Accordingly, the predominant form of the complex is EF-(2Ca-CaM) (28), in which two  $\text{Ca}^{2+}$  ions are bound to the C-terminal domain of CaM.

To study the stability of EF-CaM in solution, we previously performed three 15-ns molecular dynamics (MD) simulations of the complex (11), with different levels of calcium: EF-(0Ca-CaM), EF-(2Ca-CaM), and EF-(4Ca-CaM). The analysis of these simulations helped characterize the conformational plasticity of each partner and led us to propose a model of the EF-CaM interaction in which CaM acts as a spring that maintains EF in an open active conformation.

In this work, we investigated the effect of calcium addition/removal through the structure of the whole complex. The signal propagation was studied on a simplified representation of EF-CaM, in which only  $\text{C}(\alpha)$  were retained. The elastic network model was used first to study the natural collective dynamics of the complex and compare them to that of the MD dynamics. The effect of calcium was then characterized with approaches based on dynamical correlations on the one hand and on energetic influences on the other hand. The introduction of new concepts to describe the complex architecture gave a consensual view of the results from the two methods. These results reveal that removing calcium breaks the coupling of EF-CaM residue network, whereas calcium binding destabilizes EF within the complex.

## MATERIALS AND METHODS

### Molecular dynamics simulations

Details concerning the preparation of initial coordinate files and the production of molecular dynamics trajectories were given previously (11). Coordinates of the EF-CaM complex with two  $\text{Ca}^{2+}$  were obtained from the PDB

entry 1K93 (20). The EF-(0Ca-CaM) complex was generated by removing the cations and the EF-(4Ca-CaM) complex was generated by adding two  $\text{Ca}^{2+}$  in the N-terminal calcium loops of CaM. The Coulomb/Lennard-Jones parameters of the  $\text{Ca}^{2+}$  ions were taken from the parm99 parameter set of AMBER 8 (29), as a van der Waals radius  $R = 1.7131 \text{ \AA}$  and a well-depth  $\epsilon = 0.459789 \text{ kcal/mol}$  (30). These calcium parameters do not include any modeling of polarization. Nevertheless, this parameterization proves satisfactory, as in the MD simulations performed on isolated calmodulin (31) and on the EF-CaM complex (11), the systems exhibited a behavior and especially a calcium coordination geometry in agreement with the experimental observations.

The three complexes were simulated with AMBER 8 (29) during 15 ns in explicit solvent (>80,000 atoms). The MD trajectories were processed with the PTRAJ module of AMBER 9 (32). A convergence study (11) showed that the system took 3 ns to relax. The MD simulations were analyzed by Laine et al. (11) and the conformational drifts were qualitatively described. In EF-(0Ca-CaM) and EF-(4Ca-CaM), the two lobes of CaM slightly rotate with respect to each other, whereas in EF-(2Ca-CaM) the drift of CaM is very limited. Besides EF encounters collapse of its enzymatic site in EF-(0Ca-CaM) and N-CaM tends to expand in EF-(4Ca-CaM). Consequently the last 12 nanoseconds of each simulation were used for the analyzes. PYMOL (33) was used for visualization.

### Elastic network model

In the elastic network theory (9,34), the complex structure is modeled as a network of  $N$  nodes, each representative of a given residue  $v_i$ . Two nodes  $i$  and  $j$  are linked by a spring of constant  $\gamma^*$  if they are closer than a chosen cutoff distance  $r_c$ . The model can be expressed by a Kirchoff matrix  $\Gamma$ , defined as (34)

$$\Gamma_{ij} = \begin{cases} -\gamma^* & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{i \neq j} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (1)$$

A  $\text{C}(\alpha)$  representation of the EF-CaM complex was used to construct the elastic network model (ENM). Both experimental B-factors and MD fluctuations were predicted from the ENM according to (34)

$$\langle \Delta R_k^2 \rangle = [\Gamma^{-1}]_{kk}, \quad (2)$$

$$B_k = 8\pi^2 \langle \Delta R_k^2 \rangle / 3, \quad (3)$$

where  $(\Delta R_k^2)$  and  $B_k$  are the mean-square fluctuation and the temperature factor of residue  $k$ . The inverse of the Kirchoff matrix  $\Gamma^{-1}$  is the predicted covariance matrix. Predicted correlation matrices  $C$  were thus calculated by normalizing  $\Gamma^{-1}$ :

$$C_{ij} = \frac{[\Gamma^{-1}]_{ij}}{\left(\sqrt{[\Gamma^{-1}]_{ii}}\sqrt{[\Gamma^{-1}]_{jj}}\right)}. \quad (4)$$

The crystallographic structures 1K93 (20), and 1XFX, 1XFY, and 1XFZ (21), respectively loaded with 2, 3, or 4Ca<sup>2+</sup>, of the EF-CaM complex, were processed to predict the experimental B-factors. The MD fluctuations were calculated on the 12 last ns of each simulation on the one hand and were predicted using representative conformations from each simulation on the other hand. These representative conformations were identified previously (11) through a convergence analysis of the trajectories and correspond to MD snapshots taken at: 5906 ps, 9335 ps, and 12,150 ps of the EF-(0Ca-CaM) simulation, 5016 ps, 7950 ps, and 12,262 ps of the EF-(2Ca-CaM) simulation, 3529 ps, 4983 ps, 5471 ps, 11,084 ps, and 13,515 ps of the EF-(4Ca-CaM) simulation. The cutoff  $r_c$  was fixed at 9 Å and  $\gamma^*$  was set to 0.05 as these values permitted the best fit of predicted values on experimental B-factors.

### Pearson and generalized correlation matrices

Atomic fluctuations, Pearson correlation matrices and generalized correlation matrices were calculated for the  $N = 619$   $C(\alpha)$  of the EF-CaM complex, considering the 12 last ns of each MD trajectory. Atomic fluctuations and Pearson correlation matrices ( $N \times N$ ) were computed with the PTRAJ module of AMBER 9 (32). Correlations were quantified according to the Pearson coefficient,

$$r[\vec{x}_i, \vec{x}_j] = \frac{\langle \vec{x}_i \cdot \vec{x}_j \rangle}{\sqrt{(\langle \vec{x}_i \rangle^2 \langle \vec{x}_j \rangle^2)}}, \quad (5)$$

where  $\vec{x}_i$  and  $\vec{x}_j$  are the positional fluctuation vectors of residues  $i$  and  $j$ , respectively.

Generalized correlations recently developed by Lange and Grubmüller (35) and which are based on a different metric were computed by the software `g_correlation`, part of the GROMACS framework (36). By analogy to the Pearson coefficient, the generalized correlation coefficient is defined as (35)

$$r_M[\vec{x}_i, \vec{x}_j] = \frac{1}{\sqrt{1 - \exp(-2I[\vec{x}_i, \vec{x}_j]/3)}}, \quad (6)$$

where  $I[\vec{x}_i, \vec{x}_j]$  is the mutual information between residues  $i$  and  $j$  (35)

$$I[\vec{x}_i, \vec{x}_j] = H[\vec{x}_i] + H[\vec{x}_j] - H[\vec{x}_i, \vec{x}_j], \quad (7)$$

estimated from the measure of information content  $H[\vec{x}] = -\int p(\vec{x}) \ln p(\vec{x}) d\vec{x}$ , where  $p(\vec{x})$  is the probability distribution of the fluctuation vector  $\vec{x}$ .

Hierarchical clustering analysis was performed on the generalized correlation (GC) matrices using the statistics program R (37). Dissimilarity matrices were first determined by computing the Euclidean distances between the rows of each correlation matrix. Hierarchical clustering was performed on the dissimilarity matrices using the complete linkage method (38). The trees produced by the clustering process were cut at a given height  $h_{\text{opt}} = 5$  to identify clusters. The optimal value  $h_{\text{opt}}$  of the cutting height was determined to obtain the minimum number of clusters sufficient to span the range of generalized correlations (Table 1). The same clustering analysis was applied also on the Pearson correlation matrices but no optimal cutting height could be determined as the clusters always gave the same mean correlation.

**TABLE 1 Hierarchical clustering analysis: mean correlation values are given for clusters identified from the generalized correlation matrices of EF-(0Ca-CaM), EF-(2Ca-CaM), and EF-(4Ca-CaM) with different cut heights  $h$**

	EF-(0Ca-CaM)	EF-(2Ca-CaM)	EF-(4Ca-CaM)
$5.17 \leq h < 7.1$	$0.44 \pm 0.10$	$0.45 \pm 0.10$	$0.46 \pm 0.10$
	$0.60 \pm 0.08$	$0.60 \pm 0.07$	$0.60 \pm 0.08$
	$0.44 \pm 0.10$	$0.45 \pm 0.10$	$0.46 \pm 0.10$
$5 \leq h < 5.17$	$0.60 \pm 0.08$	$0.60 \pm 0.07$	$0.56 \pm 0.08$
	$0.46 \pm 0.10$	$0.46 \pm 0.10$	$0.46 \pm 0.10$
$4.4 \leq h < 5$	$0.46 \pm 0.16$	$0.47 \pm 0.22$	$0.56 \pm 0.08$
	$0.60 \pm 0.08$	$0.60 \pm 0.07$	$0.71 \pm 0.07$

For  $h_{\text{opt}} = 5$  two, two, and three clusters are identified for EF-(0Ca-CaM), EF-(2Ca-CaM), and EF-(4Ca-CaM), respectively, which display significantly different mean correlation values.

The comparison of correlation matrices were performed using the statistics program R (37). The similarity  $S$  between two correlation matrices  $M$  and  $N$  is calculated as their normalized dot product:

$$S = \frac{\sum_{i,j} m_{ij} n_{ij}}{\sqrt{\left(\sum_{i,j} m_{ij}^2\right) \left(\sum_{i,j} n_{ij}^2\right)}}$$

### Local feature analysis

In the local feature analysis (LFA) developed by Zhang and Wrighers (39), residue clusters representative of the most striking features of local dynamics are determined. A principal component analysis (PCA) was first performed on each MD trajectory with the PTRAJ module of AMBER 9 (32). Among the  $3N$  eigenvectors  $\Psi_r$  associated with eigenvalues  $\lambda_r$ ,  $r = 1, 2, \dots, 3N$ , the first 15 ones were sufficient to describe 81–84% of the total fluctuations. These vectors were consequently used to apply the LFA formalism (39) on the  $C(\alpha)$  motions of the complex. In brief, this formalism projects the correlation matrix in such a way that it reduces off-block diagonal correlations and identifies 15 seed degrees of freedom and hence 15 seed residues.

From the  $n = 15$  global PCA modes one can define  $3N$  local LFA output functions  $O(i)$  with minimum correlation (39),

$$O(i) = \sum_{r=1}^n \frac{A_r}{\sqrt{\lambda_r}} \Psi_r(i), \quad (8)$$

where  $A_r$  is the projection of atomic fluctuations  $\Delta x_i$  onto eigenvector  $\Psi_r$ :  $A_r = \sum_{i=1}^{3N} \Psi_r(i) \Delta x_i$ . The residual correlations between LFA outputs are given by (39)

$$\langle O(i)O(j) \rangle = \sum_{r=1}^n \Psi_r(i) \Psi_r(j) \equiv P(i, j), \quad (9)$$

so that in the limit  $n \rightarrow 3N$  they are completely decorrelated by space:  $P(i, j) \rightarrow \delta(i, j)$ .

Rather than computing all  $3N$  outputs  $O(i)$ , the sparsification algorithm described by Zhang and Wrighers (39) was applied using the statistics program R (37) to approximate the entire  $3N$  outputs  $O(i)$  with only a small subset of 15 outputs  $O(i_m)_{i_m \in M}$ , where  $M$  is the ensemble of 15 seed degrees of freedom chosen to reconstruct the outputs  $O(i)$ . At each iteration of the sparsification algorithm, the outputs  $O(i)$  were reconstructed given the current set  $M^{(m)}$  of  $m$  degrees of freedom and the reconstruction mean square errors  $O^{\text{err}}(i)$  were evaluated. Out of the  $3N$  available degrees of freedom, the seed index displaying the maximum reconstruction error  $O^{\text{err}}(i_{m+1})$  and not being already picked up was chosen as the  $(m + 1)^{\text{th}}$  index into  $M$ . Seed indexes were added to  $M$  until 15 indexes were chosen, standing for 15 seed  $C(\alpha)$  atoms or residues. Consequently, the obtained 15 seed residues are representative of the most striking features of local dynamics. They permit us to construct a local topographic representation of the EF-CaM

complex, as each seed residue and its neighboring correlated residues (residue cluster) define a local dynamic domain.

The residual correlation  $P(h, k)$  between the seed residue  $h$  and any other residue  $k$  was evaluated as (39)

$$P(h, k) = \sum_{d=1}^3 \sum_{r=1}^{n=15} \Psi_r(h_d) \Psi_r(k_d), \quad (10)$$

where  $d$  is the  $(x, y, z)$  coordinate index and  $\Psi_r$  is the PCA eigenvector associated with eigenvalue  $\lambda_r$ .

## Dependency maps

The algorithm described by Hamacher et al. in a study of the ribosomal subunit assembly (40) was used to analyze energetic dependencies inside the complex. The algorithm is based on the assumption that: "If the removal of protein Z changes the relative order from X-more-strongly-bound-than-Y to Y-more-strongly-bound-than-X, then the probabilistic assembly order is also changed" (40). This assumption was adapted to the EF-CaM case, in which pseudodomains are considered instead of proteins, and in which the stability dependency scheme of pseudodomains is considered instead of the probabilistic assembly order of proteins.

A new definition of the domains (pseudodomains) of the EF-CaM complex was derived from the LFA (Table 2). The binding free energy of a pseudodomain  $D$  to the complex  $C$  was evaluated on snapshots extracted every 10 ps from the 12 last ns of the MD trajectory of  $C$ . The binding free energy  $\Delta\Delta G[D, C]$  is expressed as

$$\Delta\Delta G[D, C] = \Delta G[C] - (\Delta G[C \setminus D] + \Delta G[D]), \quad (11)$$

where  $C \setminus D$  stands for complex  $C$  from which domain  $D$  was removed. Within each complex, binding free energies were calculated for all pseudodomains and reference ranks were attributed accordingly, with rank 1 corresponding to the best binding energy.

To determine the influence of a given pseudodomain  $D^*$  on the other pseudodomains,  $D^*$  was removed from the complex  $C$  and the binding free energy of each one of the remaining pseudodomains  $D$  to the truncated complex  $C \setminus D^*$  was calculated as follows:

$$\Delta\Delta G[D, C \setminus D^*] = \Delta G[C \setminus D^*] - (\Delta G[C \setminus (D + D^*)] + \Delta G[D]). \quad (12)$$

The free energies of binding on  $C \setminus D^*$  were then ranked accordingly. When the rank of a pseudodomain  $D$  on  $C \setminus D^*$  was worse than its reference rank, an energetic influence was attributed to  $D^*$  on  $D$ . The influence intensity is the number of ranks that  $D$  lost in  $C \setminus D^*$  context as compared to  $C$ . A total of  $7 + (6^*7) = 49$  binding free energy time series were calculated and energetic influences were materialized on dependency maps.

Binding free energies were evaluated by MMPBSA method (41) using the MMPBSA module of AMBER 9 (32), as

$$\Delta G = \Delta E_{MM} + \Delta G_{solv} - TS_{solute}, \quad (13)$$

where  $\Delta E_{MM}$  is the molecular mechanics contribution expressed as the sum of the electrostatics ( $\Delta E_{elec}$ ), van der Waals ( $\Delta E_{vdw}$ ), and internal ( $\Delta E_{int}$ ) contributions, and  $\Delta G_{solv}$  is the solvation free energy contribution to binding expressed as the sum of polar ( $\Delta G_{sol}^{polar}$ ) and nonpolar ( $\Delta G_{sol}^{nonpolar}$ ) solvation free energies. The polar contribution  $\Delta G_{sol}^{polar}$  was calculated by solving the Poisson-Boltzmann (PB) equation with the PBSA module of AMBER 9 (32). The grid spacing was set to 0.5 Å and dielectric constants of 1 and 80 were used for the interior and exterior of the solute, respectively. The dielectric boundary was defined using a 1.6 Å probe water on the atomic surface. The nonpolar contribution  $\Delta G_{sol}^{nonpolar}$  was calculated as the sum of a repulsive (cavity) term  $\Delta G_{cavity}$  proportional to the solvent-accessible surface area

**TABLE 2 Residues numbers delimiting 1), the EF-CaM domains in the structural biology definition by Drum et al. (20); and 2), the EF-CaM pseudodomains in the new definition with the LFA method**

	Structural biology definition (20)		LFA definition	
EF	CA	294–349 and 490–622	CA	292–345 and 490–501 and 576–622
	Switch A	502–551	SA	502–575
	Switch B	578–591		
	CB	350–489	CB	346–489
	Switch C	630–659	SC	623–659
	Hel	660–767	Hel	660–767
CaM	N-CaM	5–64	N-CaM	5–84
	Interlinker	65–92		
	C-CaM	93–147	C-CaM	85–147

Residues involved in the catalytic site of EF are Arg-329 (L), Lys-346 (R), Leu-348 (R), His-351 (R), Lys-353 (R), Ser-354 (R), Lys-372 (R), Asp-491, Asp-493 (L), Thr-548 (L), His-577 (L), Thr-579 (L), Asp-582 (L), Asn-583 (L), Phe-586 (L), and Glu-588 (L) and belong either to the left side (L) or to the right side (R) of the site.

(SASA) and an attractive (dispersion) term  $\Delta G_{disp}$  computed by surface-integration approach (Tan and Luo, in preparation quoted in (32)):  $\Delta G_{sol}^{nonpolar} = \Delta G_{disp} + \Delta G_{cavity} = \Delta G_{disp} + 0.04356 * SASA - 1.008$ . Because of significant variances in computing solute entropy  $S_{solute}$  using the MMPBSA approach, this term was not included in the total binding free energy value.

For most binding free energies, mean values were very close and standard deviations did not enable us to properly rank them. Nonparametric Kolmogorov-Smirnov tests (42) were therefore performed on the energy estimations time series with the statistics program R (37). The binding free energy of  $D$  was actually considered worse than that of  $D^*$  when the Kolmogorov-Smirnov test gave a  $p$ -value  $\leq 0.05$ .

## RESULTS

A previous analysis of the 15-ns MD simulations of the three complexes EF-(0Ca-CaM), EF-(2Ca-CaM), and EF-(4Ca-CaM) led to a model for EF-CaM interaction and to the description of the conformational changes associated with the binding or removal of two  $Ca^{2+}$  ions (11). In particular, binding calcium to N-CaM was shown to provoke a limited opening of the second EF-hand and both binding and removal of calcium reduced the dependence of CaM on EF while CaM interlinker underwent bending or unwinding. In this work, a scheme relating dynamics and energetics in the frame of simplified models is proposed to analyze the EF-CaM complex. Hence, the EF-CaM calcium dependence and the propagation of the calcium binding or removal signal through the residue network of the complex have been investigated. A full panel of methods for analyzing the complex dynamics was used to determine the type of dynamics analysis that agrees the most with the energetic information. The different methods used are based on: 1), elastic network theory (34); 2), local feature analysis theory (39); 3), mutual information theory (35); and 4), energetic dependency maps (40).

### Elastic network model

The elastic network model (ENM) (9,34) was applied to the crystallographic structures of the EF-CaM complex (21,20).

**TABLE 3** Slope  $a$  and coefficient of regression  $R^2$  for the linear fitting of the ENM predicted fluctuations on the experimental B-factors (x-ray) (20,21) and on the MD fluctuations (MD)

		$a$	$R^2$
1K93(20) (x-ray)		1.06	0.97
1XFX(21) (x-ray)		0.99	0.99
1XFY(21) (x-ray)		0.97	0.98
1XFZ(21) (x-ray)		0.99	0.98
EF-(0Ca-CaM) (MD)	5906 ps	0.95	0.90
	9335 ps	0.95	0.90
	12,150 ps	0.94	0.90
EF-(2Ca-CaM) (MD)	5016 ps	0.98	0.93
	7950 ps	0.97	0.93
	12,262 ps	0.98	0.92
EF-(4Ca-CaM) (MD)	3529 ps	1.00	0.91
	4983 ps	0.99	0.92
	5471 ps	0.99	0.92
	11,084 ps	1.01	0.91
	13,515 ps	1.00	0.91

The MD fluctuations were predicted by the ENM for representative conformations extracted from the MD simulations. The snapshots' positions in the trajectories of EF-(0Ca-CaM), EF-(2Ca-CaM), and EF-(4Ca-CaM) are given in picoseconds.

The predicted B-factors were found in agreement with the experimental values, as already demonstrated by Bahar et al. (34), with coefficients of regression  $R^2$  above 0.95 (Table 3). Fluctuations and correlation matrices were also predicted for a set of representative conformations from the MD trajectories of EF-(0Ca-CaM) (at 5906, 9335, and 12,150 ps), EF-(2Ca-CaM) (at 5016, 7950, and 12,262 ps), and EF-(4Ca-CaM) (at 3529, 4983, 5471, 11,084, and 13,515 ps), previously identified (11) through a convergence analysis. ENM fluctuations fit MD fluctuations with slightly lower  $R^2$  coefficients, in the 0.90–0.93 range (Table 3). Noticeably ENM correlation matrices from either crystallographic structures or MD conformations were very similar (Table 4). Consequently, despite high  $R^2$  coefficients with experimental values, the ENM is unable to highlight changes in internal dynamics between the different levels of calcium.

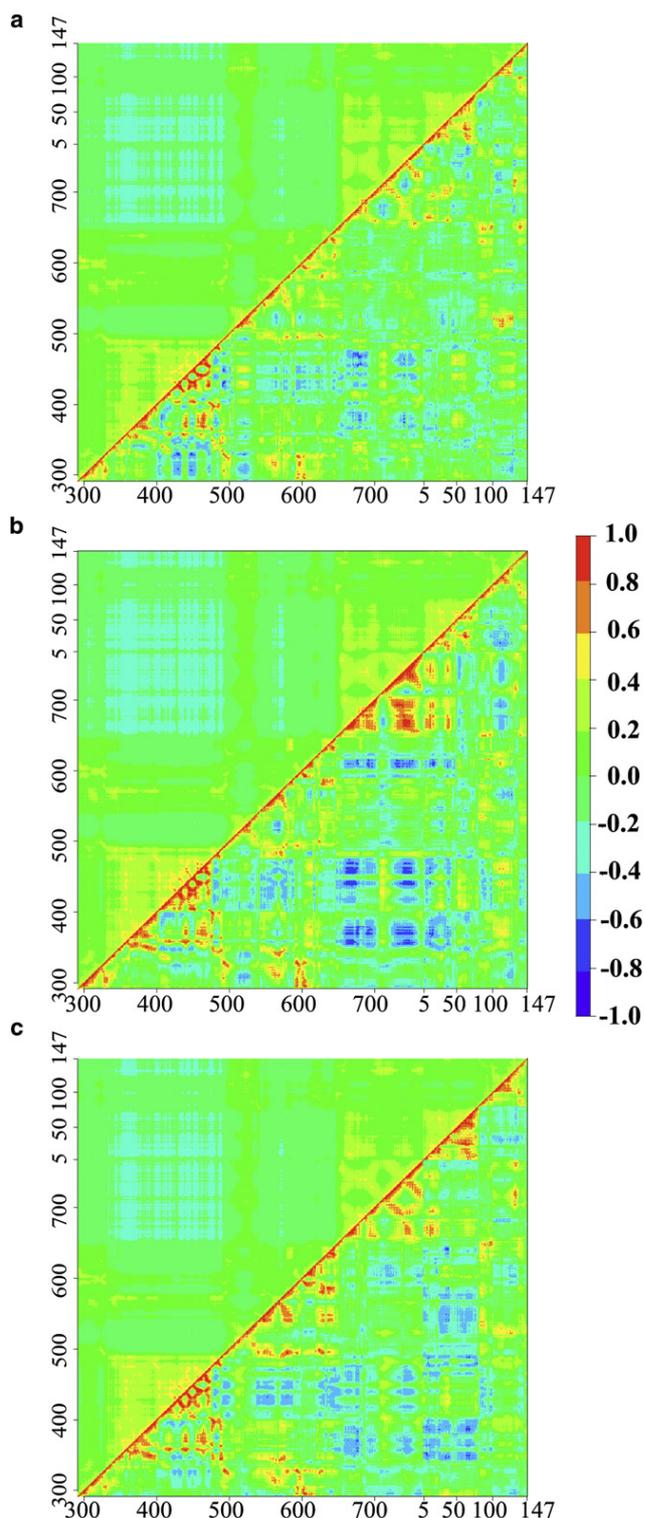
The ENM correlation matrices for EF-(0Ca-CaM), EF-(2Ca-CaM), and EF-(4Ca-CaM) were compared to the Pearson correlation matrices calculated from the MD trajectories (Fig. 2). ENM correlations (*upper left*) are overall smaller than Pearson correlations (*lower right*), since ~98% of the

**TABLE 4** Similarity shared by the generalized correlation matrices (*upper left*) and the Pearson (superscript  $P$ ) and the elastic network (superscript  $EN$ ) correlation matrices (*lower right*) between the different levels of calcium

	EF-(0Ca-CaM)	EF-(2Ca-CaM)	EF-(4Ca-CaM)
EF-(4Ca-CaM)	0.45	0.49	1
EF-(2Ca-CaM)	0.68	1	$0.65^P / 0.98^{EN}$
EF-(0Ca-CaM)	1	$0.73^P / 0.98^{EN}$	$0.71^P / 0.98^{EN}$

The degree of similarity  $S$  between two matrices is evaluated as their normalized scalar product:  $S = \frac{\sum_{i,j} m_{ij} n_{ij}}{\sqrt{(\sum_{i,j} m_{ij}^2)(\sum_{i,j} n_{ij}^2)}}$ .

$$S = \frac{\sum_{i,j} m_{ij} n_{ij}}{\sqrt{(\sum_{i,j} m_{ij}^2)(\sum_{i,j} n_{ij}^2)}}$$



**FIGURE 2** Correlation matrices predicted by the ENM (*upper left*) and Pearson correlation matrices (*lower right*). ENM correlation matrices were predicted for representative conformations extracted from the MD simulations, taken at: (a) 9335 ps for EF-(0Ca-CaM), (b) 7950 ps for EF-(2Ca-CaM), and (c) 11,084 ps for EF-(4Ca-CaM). Pearson correlation matrices were computed on the last 12 ns of each MD simulation. Residue numbers of EF (292–767) and CaM (5–147) are indicated on  $x$  and  $y$  axes.

**TABLE 5** LFA seed residues

		EF-(0Ca-CaM)	EF-(2Ca-CaM)	EF-(4Ca-CaM)
Tails		Arg-292, Gln-767 (EF)	Arg-292 (EF), Glu-6 (CaM)	Glu-6 (CaM)
EF	Domain CB	Thr-392, Lys-461	Gln-376, Gly-429, Glu-449, Leu-465	Val-380, Gly-429, Asn-450, Glu-482
	Switches A and C		Thr-519 (SA), Gly-638 (SC)	Thr-545 (SA)
	Domain Hel	Ser-673, Asp-684, Glu-687, Pro-743, Tyr-745	Asn-675, Ser-682, Phe-688, Pro-743	Ser-674, Asp-684, Asp-686
CaM	N-CaM	Asp-24, Ala-57, Asn-60	Ala-57, Gly-61	Asp-24, Ser-38, Ala-46
	Interlinker	Asp-78		Arg-74
	C-CaM	Gly-120, Met-124	Asp-129	Glu-104, Asp-118

Seed residues are located 1), in the N- and C-tails of EF and of CaM; 2), in domain CB, switches A and C and domain Hel of EF; and 3), in the N- and C-terminal domains and the interlinker of CaM. Residues locations in the structure are indicated in parenthesis when necessary: switch A is denoted SA and switch C is denoted SC.

ENM values are within the range  $-0.4$  to  $0.4$  whereas the Pearson correlations span the whole interval between  $-1$  and  $1$ . This is not a consequence of a smaller number of significantly contributing modes in the Pearson matrices spectra, which would make artificially high correlations (data not shown). Furthermore, the Pearson correlation patterns are much finer than those of ENM (Fig. 2), illustrating the fact that ENM describes collective motions, arising mainly from the intrinsic molecular shape of the complex.

The ENM correlation matrices unravel two anti-correlated residue blocks within EF-CaM (Fig. 2, upper left): 1), the catalytic region, formed roughly by domains CA and CB (residues 292–501); and 2), the EF-CaM interface region formed roughly by Hel of EF (residues 648–767) and N-CaM (residues 5–76). These results are in agreement with the principal component analysis of EF motions (11), which revealed a compaction and a torsion between the catalytic core and the helical domain of EF as the two first modes of EF-(2Ca-CaM) dynamics. This shows that ENM is able to predict long-range correlations to some extent, although it is based on local geometrical constraints only. Pearson correlation matrices reveal similar positive and negative correlations (Fig. 2, lower right): for instance, the anti-correlation between domains CB and Hel of EF is particularly visible in EF-(2Ca-CaM) (Fig. 2 b, lower right).

Although ENM could extract the global motions of the complex, it was not acute enough to discriminate between the various levels of calcium. To tackle that question, other methods were used: 1), post-processing of the Pearson covariance matrices by the local feature analysis theory (39); and 2), the generalized correlation coefficient (35).

### Local feature analysis

To better study the local dynamics of EF-CaM, a local feature analysis (LFA) (39) was performed on the MD trajectories. Principal component analysis (PCA) was applied on the  $C(\alpha)$  motions and then, the first 15 eigenvectors were retained for the LFA (39). The LFA sparsification algorithm (39) selected 15 seed degrees of freedom standing for 15 seed residues. As in the original article of Zhang and Wriggers (39), seed residues are mostly allocated to flexible

regions within each complex (Table 5): the N- and C-tails of EF and CaM, the catalytic region and the switches A and C of EF, the calcium binding sites and the interlinker of CaM. The high flexibility of the seed residues located in the N- and C-tails (Arg-292 and Gln-767 in EF, Glu-6 in CaM) is associated with independent motions, an inherent feature of the terminal parts of a protein. Consequently these residues display flat correlation profiles, bearing little information (data not shown). More interestingly, the other seed residues are associated with locally correlated residues that form clusters representing local dynamic domains. In the three complexes, these clusters are found in regions important for the EF catalytic activity, the EF-CaM interaction, or the  $\text{Ca}^{2+}$ -CaM interaction. They are described here in more details, focusing mainly on the EF-(2Ca-CaM) trajectory.

Seed residues found in EF of EF-(2Ca-CaM), together with their associated clusters (residual correlation greater or equal to 0.03), are displayed in Fig. 3. The catalytic residues are highlighted in licorice and the whole structure is colored according to the residual correlation profile of the seed residue. The five seed residues Leu-465, Gln-376, Gly-429, Glu-449 (not shown in Fig. 3), and Gly-638 are located in the catalytic region (Table 5, Fig. 3, a–d). As shown in Fig. 3, a and b, the two clusters formed around Gln-376 and Leu-465 lie on the right edge (R) of the catalytic site and share residues 376, 379, 465, and 467 in common. Both seed residues display negative residual correlations (down to  $-0.03$ , in blue) with residues located in the left edge (L) of the catalytic site. Reciprocally the cluster formed around Gly-638 (Fig. 3 d), which includes residues from switches B and C in the left edge (L), displays anti-correlations (down to  $-0.03$ ) with residues located in the right edge (R). As a consequence, the local dynamics in the catalytic region are polarized around two anti-correlated spots: Gly-638 from switch C representative of the left side and Gln-376 from CB, closely related to Leu-465, representative of the right side. This polarization, which highlights the opening/closing movement of the site, cannot be seen in EF-(0Ca-CaM), where the number of seed residues in the catalytic region is reduced (Table 5), consistent with the collapse observed in the MD simulation (11).

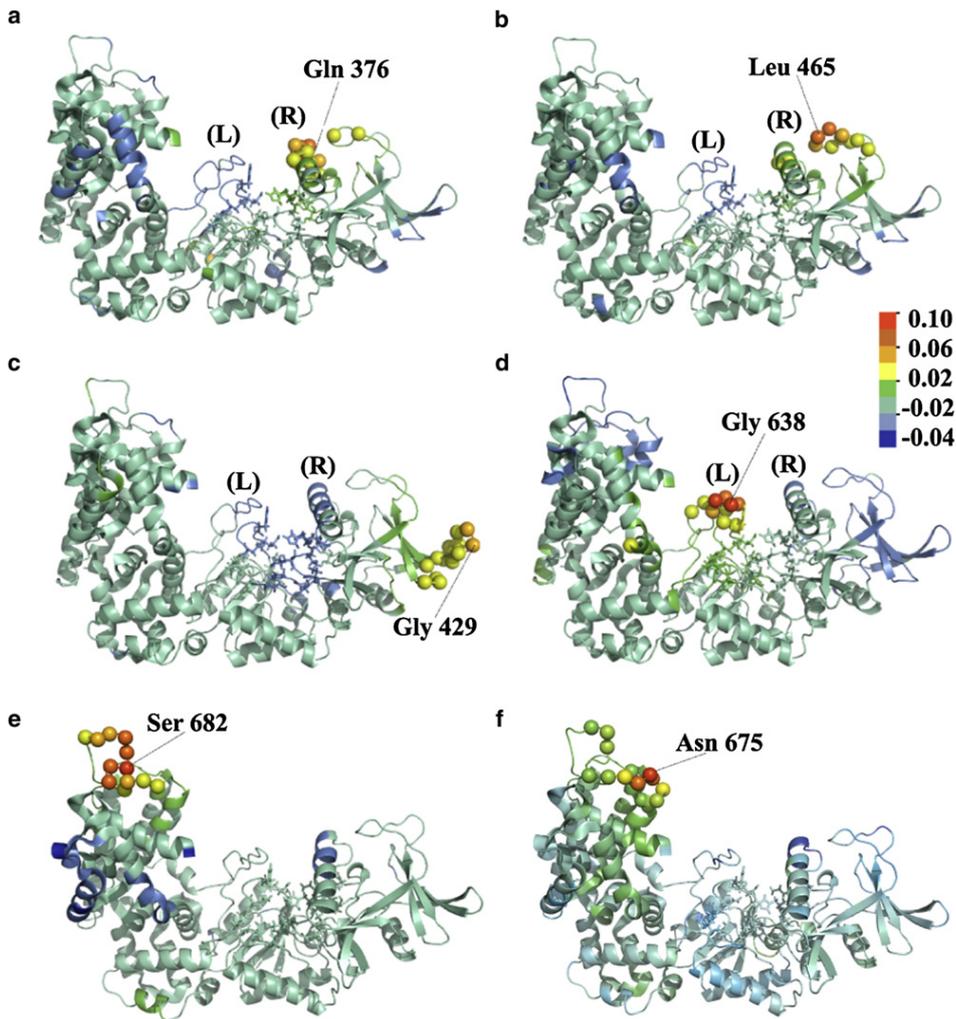


FIGURE 3  $C(\alpha)$  atoms of the LFA seed residues located in the catalytic region and in the helical domain of EF-(2Ca-CaM), together with their associated clusters (residual correlation greater or equal to 0.03), represented in spheres: (a) Gln-376, (b) Leu-465, (c) Gly-429, (d) Gly-638, (e) Ser-682, and (f) Asn-675. Catalytic residues are displayed in licorice. The terms (L) and (R) stands for the left side and right side of the catalytic site, respectively. Colors follow the residual correlation profile with respect to each considered seed residue, from red down to blue.

Further away from EF catalytic site, Gly-429 is anti-correlated (down to  $-0.02$ ) with most of the catalytic residues and residues at both edges of the catalytic site (Fig. 3 c). Glu-449 displays a similar correlation profile and is not shown. This region is also highlighted in EF-(4Ca-CaM) by seed residues Gly-429 and Asn-450 (Table 5).

In EF-(2Ca-CaM), seed residues were also selected in regions of EF that are contacting CaM (Table 5): 1), Thr-519 within switch A, in contact with C-CaM; and 2), Asn-675, Ser-682, Phe-688, and Pro-743 within the inter L-M loop of Hel, in contact with the first EF-hand of N-CaM (Fig. 1). Both Ser-682 (Fig. 3 e) and Asn-675 (Fig. 3 e) display positive residual correlations (up to 0.02) with the inter N-O loop (residues 738–745, Fig. 1). Additionally, Ser-682 and Phe-688 (whose correlation profile is similar to that of Ser-682, data not shown) display negative residual correlations (down to  $-0.03$ ) with the entering helix C of site 2 and the interlinker of CaM. It is interesting that seed residues were found in switch A, switch C, and the inter L-M loop of Hel in EF-(2Ca-CaM), as these regions are of particular importance for the interaction between EF and CaM

(20,21). Furthermore, removing or adding calcium, while affecting the stability of the complex, leads to quite different local dynamics features: no seed residue is found in switches A and C of EF-(0Ca-CaM), neither in switch C of EF-(4Ca-CaM); the number of seed residues within the inter L-M and N-O loops of Hel is increased up to 5 in EF-(0Ca-CaM) and decreased down to 3 in EF-(4Ca-CaM).

Seed residues are also located in CaM (Table 5) and their residual correlations are largely influenced by the level of calcium (Fig. 4). In EF-(2Ca-CaM), two clusters are defined in the calcium-binding loop of site 2 around Ala-57 and Gly-61. On the residual correlation matrix (Fig. 4 b), the second EF-hand of CaM appears as a block extending from residue 43 toward residue 80 of the interlinker. A correlation peak is also located in site 1 (residues 20–31) and two off-diagonal peaks demonstrate the intercorrelation between the two calcium binding loops (residues 20–31 and 56–67). No correlation is observed in the C-terminal domain, except for a narrowed peak located in the inter F-G loop between the two EF-hands (residues 108–118). Asp-129 was selected as a seed residue, but its motions are rather independent as it

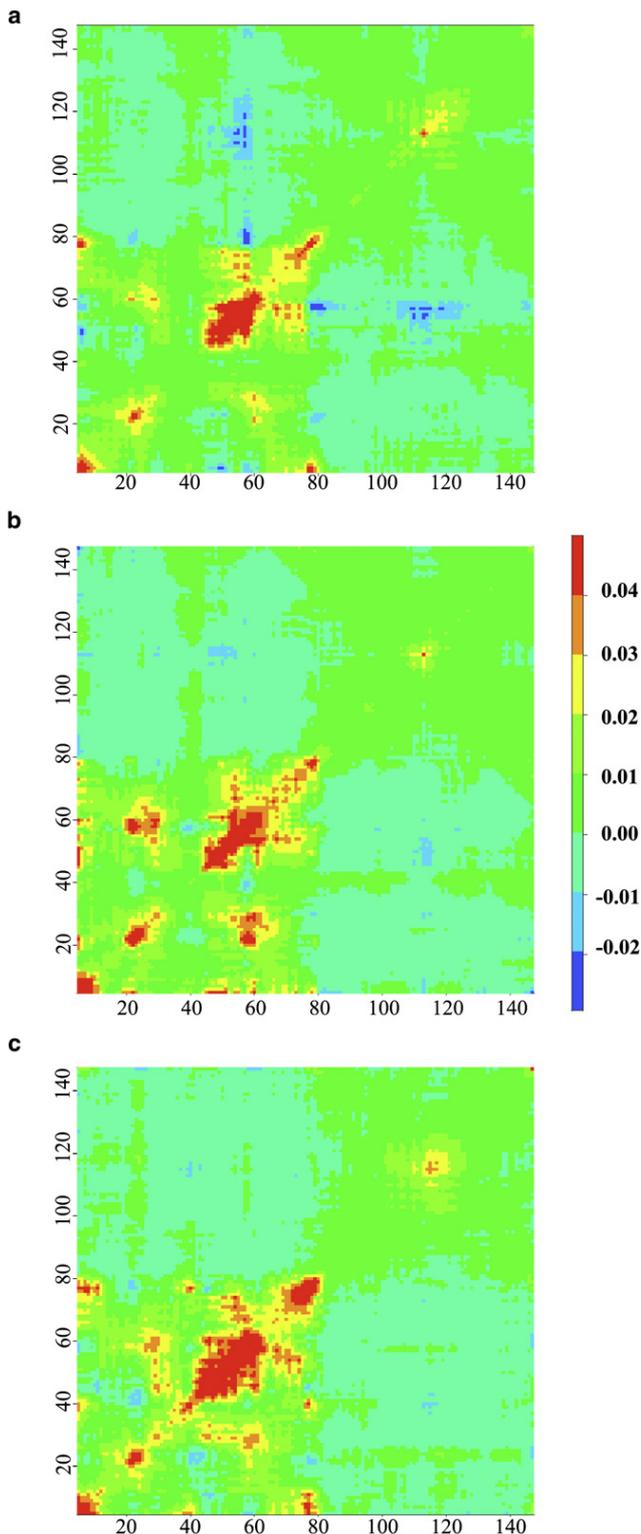


FIGURE 4 LFA residual correlation matrices of CaM within (a) EF-(0Ca-CaM), (b) EF-(2Ca-CaM), and (c) EF-(4Ca-CaM). Residues numbers of CaM (5–147) are indicated on *x* and *y* axes.

displays a very flat correlation profile with absolute values  $<0.01$ .

Removing calcium from sites 3 and 4 in EF-(0Ca-CaM) reduces the correlation between site 1 and site 2 while inducing a new correlation peak between seed residue Asp-78 and the N-tail residues (Fig. 4 *a*), which is related to the unwinding of the interlinker during the simulation (11). Furthermore, residues 54–58 in site 2 are anti-correlated with Asp-78 and helix G (residues 108–129) of C-CaM (Fig. 4 *a*). Calcium binding in N-CaM (sites 1 and 2) induces an expansion of the correlation block of the second EF-hand toward helix B (residues 27–41) (Fig. 4 *c*) while seed residues are located earlier in the sequence of site 2 (Ser-38 and Ala-45). The two off-diagonal peaks located in the first EF-hand of EF-(2Ca-CaM) residual correlation matrix (Fig. 4 *b*) become less intense while anti-correlation peaks arise between the calcium binding loop of site 1 and the inter B-C loop (residues 39–44). As in EF-(0Ca-CaM), the interlinker (seed residue Arg-74) is correlated with the N-tail, consistent with the bending observed during the EF-(4Ca-CaM) simulation (11).

The following conclusions can be drawn regarding the effect of calcium binding/removal on the CaM residue network. The direct impact of removing calcium is observed in the interhelical F-G loop (Fig. 1 *b*) of C-CaM in EF-(0Ca-CaM). Additionally, motions of the F-G loop and of the interlinker become anti-correlated with that of helix C of N-CaM, which thus reveals the propagation of calcium information. On the other hand, the EF-(4Ca-CaM) correlation matrix highlights the propagation of calcium-induced conformational changes through site 2 toward helix B of site 1. Information is transmitted between the two EF-hands through anti-correlated motions of the calcium-binding loop and of the interhelical B-C loop. Overall, the LFA enables us to highlight local dynamic regions around which EF dynamics are organized and helps us to understand the propagation of calcium information along the CaM structure.

### Generalized correlations

Global motions of EF-CaM were analyzed with generalized correlations calculations (35) on the MD trajectories of the three complexes. Based on mutual information theory, this correlation measure is intended to be more accurate and less biased than the commonly used Pearson correlation coefficient (35). Indeed, the Pearson measure (Fig. 5, *lower right*) displays  $>50\%$  of absolute correlations smaller than 0.2 (in *green*), which is not the case for the generalized correlations (Fig. 5, *upper left*). The GC matrices display a larger number of intense values ( $>0.8$ ) that relate larger numbers of residues far away in the proteins sequences (Fig. 5, *lower right*). In addition, the Pearson correlation matrices of EF-(0Ca-CaM), EF-(2Ca-CaM), and EF-(4Ca-CaM) share 65–73% similarity, while the generalized correlation matrices share 45–68% similarity (Table 4), and thus are more sensitive to the level of calcium.

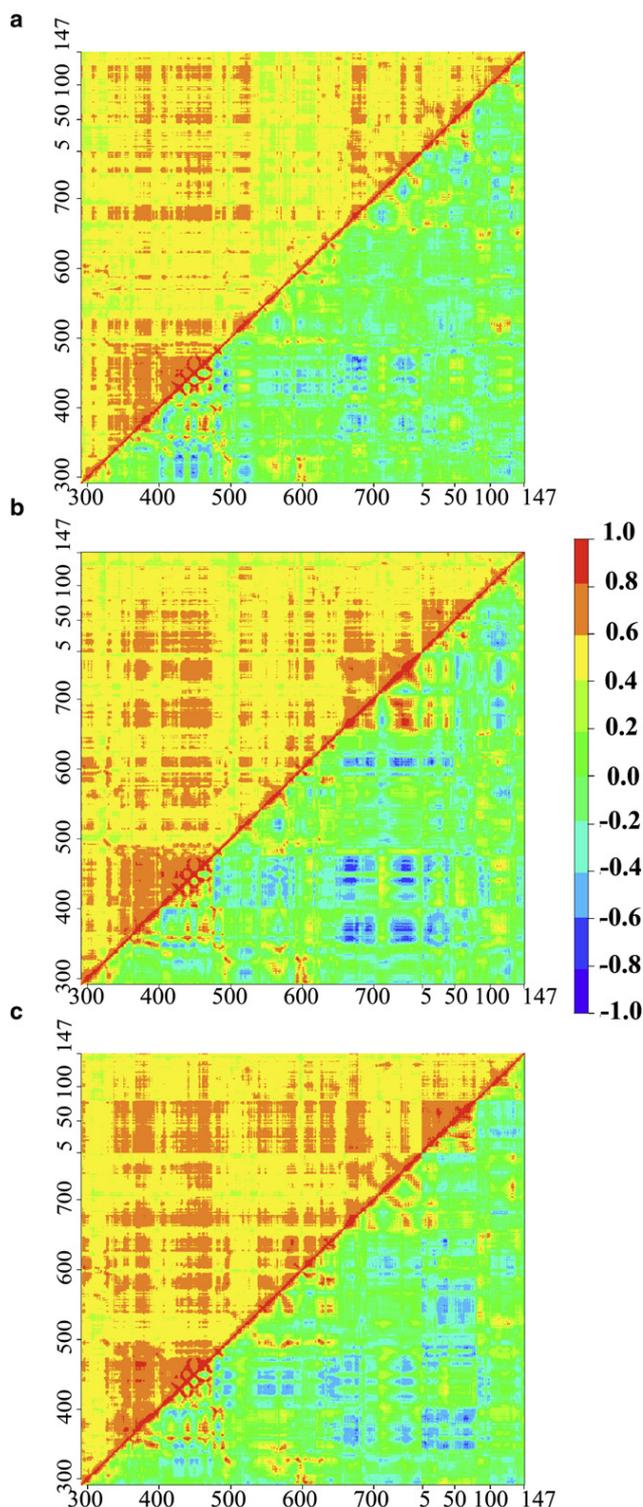


FIGURE 5 Generalized correlation matrices (*upper left*) and Pearson correlation matrices (*lower right*) for (a) EF-(0Ca-CaM), (b) EF-(2Ca-CaM), and (c) EF-(4Ca-CaM). Residues numbers of EF (292–767) and CaM (5–147) are indicated on *x* and *y* axes.

A hierarchical clustering analysis was applied to the GC matrices of the three complexes (Fig. 5). Residues were grouped based on the similarity of their correlation profiles, given that each row of a correlation matrix represents the correlation profile of a particular residue. Consequently, residues, being members of the same cluster, have the same global correlation to the remaining part of the residue network and display similar correlations to each other. The cutting height of the clustering was determined as described in [Materials and Methods](#). Using this approach, three clusters were identified (Fig. 6 and Table 2): cluster 1 (in *blue*) is composed of uncorrelated residues with mean correlation below 0.5; cluster 2 (in *green*) is composed of slightly correlated residues with correlations in the 0.5–0.6 range; and cluster 3 (only for EF-(4Ca-CaM), in *red*) is composed of strongly correlated residues with a mean correlation value of 0.7.

In EF-(2Ca-CaM), residues are grouped in two clusters. A large majority, 446 residues, display intermediate correlations in cluster 2 (Fig. 6 *b*, in *green*). Most of the 173 remaining uncorrelated residues of cluster 1 (in *blue*) have small fluctuations in the MD trajectory (Fig. 6 *e*). However, residues 510, 544–545 in switch A of EF, residues 636, 641–642 in switch C of EF, residues 766–767 in the C-tail of EF, and residues 50, 66–67, and 70 in site 2 of CaM display fluctuations above 1.5 Å. Consequently, the motions of these residues are relatively independent from the rest of the protein residue network.

In EF-(0Ca-CaM), a larger number of 245 residues populate cluster 1 (in *blue*). One-hundred-and-forty-nine of them form a large intermediate region located in CA and switch C between the catalytic core of EF and the EF-CaM interface (Fig. 6 *a*) and display MD fluctuations uniformly smaller than in EF-(2Ca-CaM) (Fig. 6 *d*). In contrast, some residues within C-CaM which were uncorrelated (cluster 1) in EF-(2Ca-CaM), become slightly correlated (cluster 2) in EF-(0Ca-CaM) while their fluctuations get larger. Furthermore, removing calcium provokes the splitting of Hel (residues 660–767) and N-CaM (residue 5–76) in four subsets of correlated flexible residues separated by uncorrelated more rigid ones (Fig. 6 *d*): 1), the inter L-M loop (residues 668–691); 2), the inter N-O loop (residues 738–745); 3) the C-tail of EF (residues 754–767) and the N-tail of CaM (residues 5–6); and 4), the second EF-hand together with the interlinker (residues 42–87). This finding is in agreement with the previously observed concentration of LFA seed residues in the L-M and N-O loops of EF.

In the GC matrix of EF-(4Ca-CaM) (Fig. 5 *c*), it was possible to group the residues into three clusters. Cluster 3 (in *red*) includes highly correlated residues located in three distinct regions of the complex (Fig. 6 *c*): 1), in CB; 2), in switch B, switch C, or CA; and 3), in helix L of Hel and N-CaM. This suggests that the conformational changes occurring in N-CaM upon calcium binding echo up to the catalytic region of EF. As opposed to these highly correlated regions, a group of uncorrelated residues (in *blue*) is formed

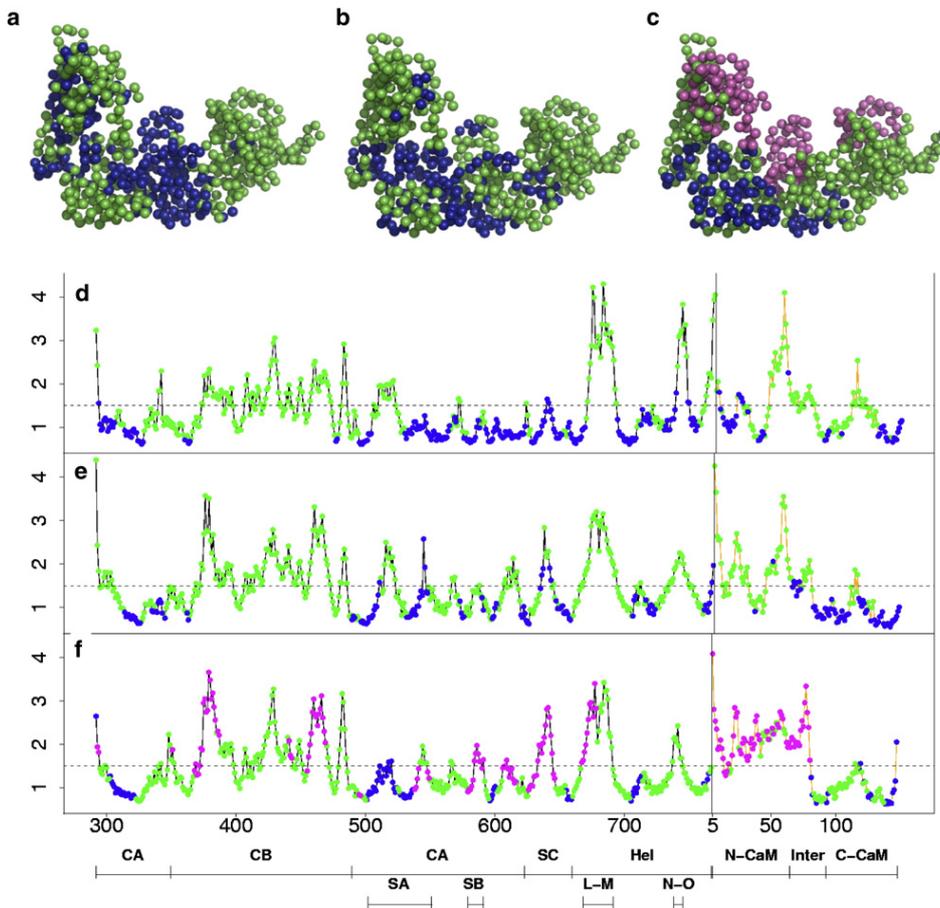


FIGURE 6 Clusters defined from the generalized correlation matrices and MD atomic fluctuations of (a and d) EF-(0Ca-CaM), (b and e) EF-(2Ca-CaM), and (c and f) EF-(4Ca-CaM). (a–c)  $C(\alpha)$  atoms are represented as spheres and colored with respect to their cluster: cluster 1 of uncorrelated residues ( $<0.5$ ) is displayed in blue, cluster 2 of slightly correlated residues (0.5–0.6) is displayed in green, and cluster 3 of highly correlated residues (0.7) is displayed in red. (d–f)  $C(\alpha)$  atomic fluctuations of EF (black line) and of CaM (orange line), calculated on the last 12 nanoseconds of each MD trajectory. Points are colored according to the clusters color-code. The value of 1.5 Å is marked with a dotted line on each plot. The locations of the different domains CA, CB, Hel, N-CaM, the interlinker (Inter), and C-CaM of the complex, as well as those of switches A (SA), B (SB), and C (SC), and of the interhelical L-M and N-O loops along the sequence, are indicated at the bottom.

around C-CaM, comprising residues 303–322 and 596–601 of CA, residues 501–537 of switch A, residues 655–659 of switch C, and residues 705–713 of Hel.

The analysis of the GC matrices enables us to draw correlation pictures of the EF-CaM complex that are sensitive to the level of calcium. The EF-(2Ca-CaM) picture displays homogeneity as most of the residues are slightly correlated except for localized regions. By contrast, correlation pictures of the two other complexes illustrate the destabilization of the EF-CaM interaction (11). For instance, EF-(0Ca-CaM) is divided in two correlated blocks separated by a region of uncorrelated residues formed by CA and switch C. The first block corresponds to CB and the second block is fragmented in several flexible correlated groups delimited by uncorrelated rather rigid residues. Therefore, calcium removal breaks the coupling between EF-CaM interface and the catalytic region, which agrees with the instability of the EF-CaM complex experimentally observed in absence of calcium (28). On the other hand, adding calcium in EF-(4Ca-CaM) induces high correlations between N-CaM and the catalytic core of EF. This suggests that calcium binding induces constraint as observed previously (11), and the information is transmitted throughout the complex, even at long distances from N-CaM. Overall, this analysis supports that moderate dynamical correlations are indicative of a stable complex.

## Dependency maps

The effect of calcium on EF-CaM energetics was investigated by performing MMPBSA free energy calculations on the MD trajectories. The approach proposed by Hamacher et al. (40) and described in more details in [Materials and Methods](#) consists in calculating and ranking the binding free energy of each domain  $X$  on the rest of the complex and determining the influence of the removal of each other domain  $Y$  on the ranking. If the rank of  $X$  is lower when  $Y$  is removed, then  $Y$  is considered to have an energetic influence on the stability of  $X$  within the complex. Energetic influences are materialized by arrows on dependency maps (see Fig. 8).

Seven pseudodomains were defined inside EF-CaM with the results obtained with the LFA approach. These definitions implied some adjustments to the structural biology definition presented by Drum et al. (20) (Table 2). The matrix of EF-(2Ca-CaM) LFA residual correlations is displayed in Fig. 7 a, with our new pseudodomains delimitations. A first adjustment was made for CA and CB, defined as residues 292–349, 490–522, and residues 350–499, respectively, by Drum et al. (20). As shown in the structural biology definition (Table 2), most of the residues located in the right side of the catalytic site belong to CB and most of the

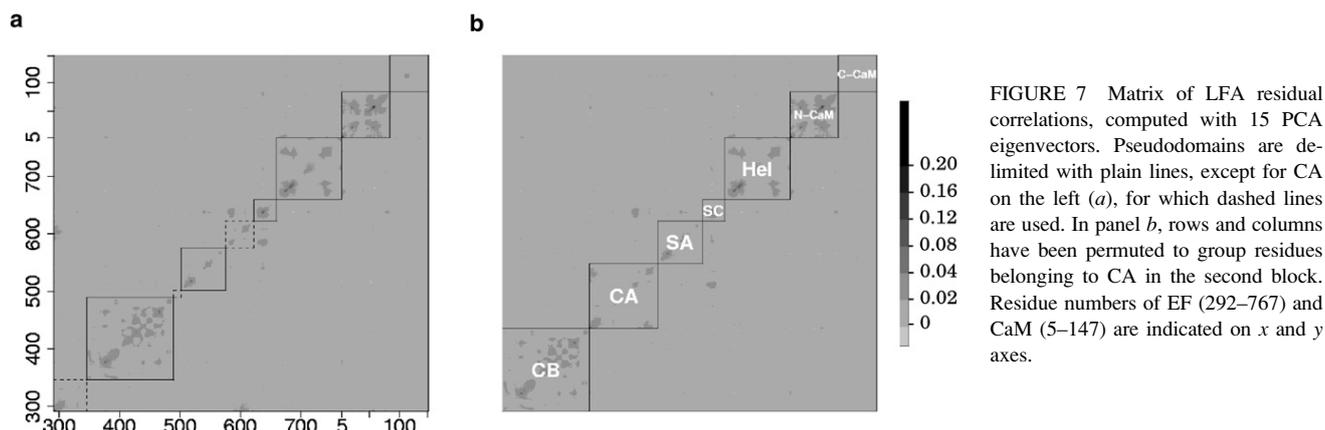


FIGURE 7 Matrix of LFA residual correlations, computed with 15 PCA eigenvectors. Pseudodomains are delimited with plain lines, except for CA on the left (*a*), for which dashed lines are used. In panel *b*, rows and columns have been permuted to group residues belonging to CA in the second block. Residue numbers of EF (292–767) and CaM (5–147) are indicated on *x* and *y* axes.

residues located in the left side belong to CA. The only exceptions are observed for Lys-346 and Leu-348, which are located in the right side of the catalytic site and belong to CA. Considering the polarization of the catalytic site shown by the LFA, we thus redefined CB from residues 346 to 489. We also extracted switch A from CA, as residues 502–575 form a well-defined correlation pattern and are not correlated with the remaining part of CA (Fig. 7 *a*). We added residues 623–629 to the definition of switch C (residues 623–659) and we kept the original definition of Hel (residues 660–767). Since the LFA correlation pattern of CaM associates the interlinker and N-CaM (Fig. 7 *a*), we redefined N-CaM from residues 5 to 84 and C-CaM from residues 85 to 147. After columns and rows permutations, the LFA correlation matrix (Fig. 7 *b*) displays no interdomain correlations except between switch C and switch B (residues 576–591) within CA.

Binding free energies were calculated on each MD trajectory with MMPBSA method (41). The contributions of the different energy terms to the total binding free energy  $\Delta\Delta G$  are given for EF-(2Ca-CaM) (Table 6). They are of the same order for the three complexes (data not shown). The largest terms are the Coulomb interaction energies  $\Delta\Delta E_{\text{ele}}$  and the polar desolvation penalties  $\Delta\Delta G_{\text{sol}}^{\text{polar}}$ , with absolute values in the 400–2000 kcal/mol range that compensate each other. The other terms are much smaller: van der Waals energies are between  $-80$  and  $-320$  kcal/mol, and both nonpolar solvation terms and internal energies are between 0 and 50 kcal/mol.

As visible on Fig. 8, energetic dependency maps (EDMs) are sensitive to the level of calcium and reveal both short-range and long-range influences. Short-range influences appear to be associated with unfavorable changes in binding free energies  $\Delta\Delta G$  (Table 7). Indeed, the positive free energy differences  $\Delta\Delta\Delta G[\text{CA} \rightarrow \text{SC}]$ ,  $\Delta\Delta\Delta G[\text{SA} \leftrightarrow \text{C-CaM}]$ ,  $\Delta\Delta\Delta G[\text{Hel} \leftrightarrow \text{N-CaM}]$  (except in EF-(2Ca-CaM)), and  $\Delta\Delta\Delta G[\text{Hel} \rightarrow \text{C-CaM}]$  are due to van der Waals penalties. By contrast, the Coulomb electrostatic penalty is preponderant in the positive  $\Delta\Delta\Delta G[\text{SC} \leftrightarrow \text{N-CaM}]$  and the difference  $\Delta\Delta\Delta G[\text{SA} \rightarrow \text{N-CaM}]$  results from mixed Coulomb electrostatic and van der Waals effects. Consequently, short-range energetic influences correspond to direct hydrophobic or electrostatic interactions/contacts. However, long-range influences of CB on SA, Hel, and N-CaM are not associated with large free energy changes (Table 7). Indeed, van der Waals and nonpolar contributions are rather small while electrostatic and polar contributions compensate each other, which is not surprising since CB is  $>30$  Å away from SA, Hel, or N-CaM. The losses of ranks for these domains are actually promoted by a favorable change in the binding energy of CA to the complex when CB is removed, of 36 and 65 kcal/mol in EF-(2Ca-CaM) and EF-(4Ca-CaM), respectively.

Hamacher et al. (40) used EDMs to predict the assembly ordering of the small ribosomal subunit. Here, instead of considering the assembly of proteins within a suprastructure, we considered the assembly of pseudodomains, some of which covalently bonded, within a protein-protein complex.

TABLE 6 Reference binding free energies of the seven pseudodomains of EF-(2Ca-CaM)

	$\Delta\Delta E_{\text{ele}}$	$\Delta\Delta E_{\text{vdw}}$	$\Delta\Delta E_{\text{int}}$	$\Delta\Delta G_{\text{sol}}^{\text{polar}}$	$\Delta\Delta G_{\text{sol}}^{\text{nonpolar}}$	$\Delta\Delta G$
CA	$-552 \pm 87$	$-314 \pm 11$	$52 \pm 4$	$800 \pm 85$	$11 \pm 3$	$-3 \pm 22$
CB	$-407 \pm 60$	$-82 \pm 5$	$24 \pm 3$	$497 \pm 58$	$14 \pm 2$	$44 \pm 13$
SA	$-1410 \pm 91$	$-241 \pm 9$	$20 \pm 3$	$1526 \pm 84$	$17 \pm 3$	$-89 \pm 24$
SC	$-881 \pm 45$	$-241 \pm 9$	$19 \pm 3$	$1049 \pm 43$	$2 \pm 2$	$-51 \pm 20$
Hel	$-1911 \pm 78$	$-161 \pm 9$	$10 \pm 2$	$2035 \pm 75$	$21 \pm 2$	$-6 \pm 17$
N-CaM	$-1387 \pm 86$	$-101 \pm 11$	$10 \pm 2$	$1445 \pm 82$	$25 \pm 2$	$-9 \pm 18$
C-CaM	$-705 \pm 87$	$-251 \pm 12$	$10 \pm 2$	$830 \pm 77$	$16 \pm 3$	$-101 \pm 29$

The different contributions to the total free energy of binding  $\Delta\Delta G$  are given in kcal/mol as the Coulomb interaction energy ( $\Delta\Delta E_{\text{ele}}$ ), the van der Waals energy ( $\Delta\Delta E_{\text{vdw}}$ ), the internal energy ( $\Delta\Delta E_{\text{int}}$ ), the polar solvation energy ( $\Delta\Delta G_{\text{sol}}^{\text{polar}}$ ), and the nonpolar solvation energy ( $\Delta\Delta G_{\text{sol}}^{\text{nonpolar}}$ ).

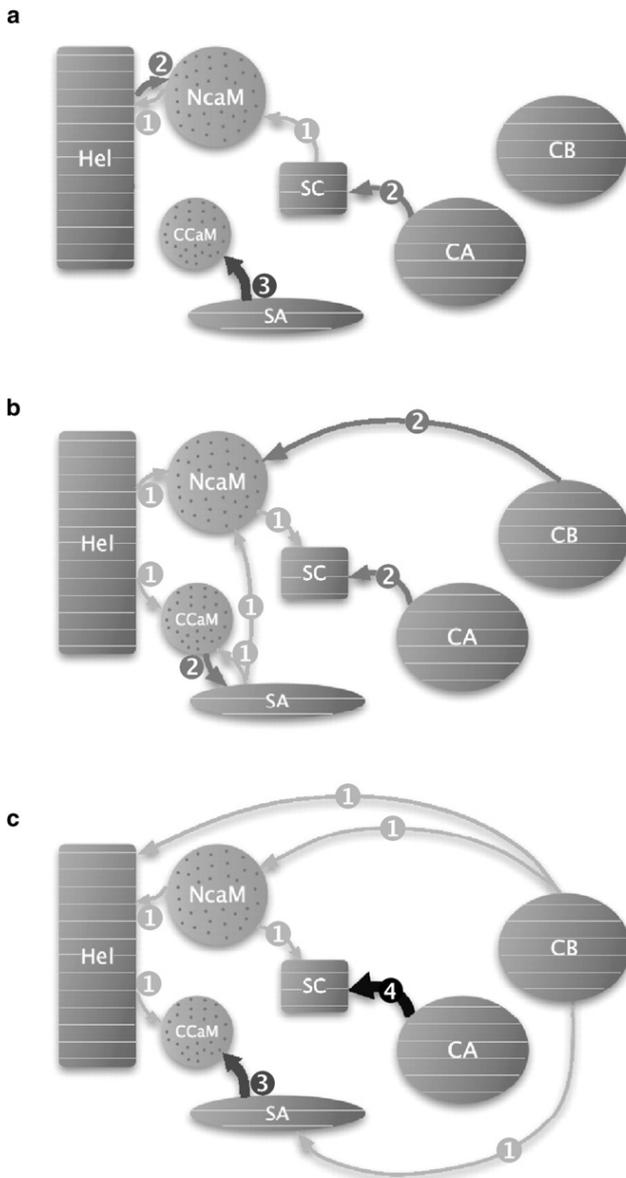


FIGURE 8 Dependency maps of (a) EF-(0Ca-CaM), (b) EF-(2Ca-CaM), and (c) EF-(4Ca-CaM). The EF-CaM complex is schematically represented by its seven pseudodomains, with CaM (pseudodomains N-CaM and C-CaM) with dots and EF (pseudodomains CA, CB, SA, SC, Hel) with horizontal lines. The arrows represent the energetic influences: the larger the greater intensity or number of lost ranks in the comparison of free binding energies (light shading is 1, medium shading is 2, dark shading is 3, and solid is 4).

Because of this choice, and because the establishment of the EF-CaM interaction induces conformational rearrangements not sampled in our simulations, the assembly ordering we get from the maps (Fig. 8) does not describe the actual mechanism of the EF-CaM complex formation. Instead, it gives a description of the stability dependence scheme between the different parts of the complex, as in a house-of-cards architecture: the more a pseudodomain is influenced by others, the lower the place it occupies in the hierarchy of the relative

stabilities inside the complex. Accordingly, we propose that pseudodomains not subjected to any influence consist in nucleation points for the architecture of the complex.

This way, CA, CB, and Hel are the nucleation points of EF-(2Ca-CaM) (Fig. 8 b). The influences of both Hel and switch A on C-CaM correspond to direct interactions at the EF-CaM interface. Additionally, the influence of C-CaM on switch A may highlight the conformational change switch A undergoes when interacting with C-CaM (20). N-CaM is subject to the influences of switch A, CB, and Hel, which means that it plays a less important role than C-CaM in the interaction with EF, in agreement with experiments (20,21). Finally, the influences of N-CaM and CA on switch C illustrate the conformational change that switch C undergoes while interacting with N-CaM and which leads to the stabilization of the catalytic site by ordering switch B of CA (20).

In EF-(0Ca-CaM) (Fig. 8 a), the nucleation ensemble (CA, CB, and switch A) is displaced toward EF catalytic region, which reveals that EF-CaM interaction is weakened. No long-range influence is observed, so that the coupling is broken between the EF-CaM interface region and the catalytic region. This breakage is consistent with the collapse of the catalytic site observed during this simulation (11) and with the generalized correlations (Fig. 6 a). The interdependence of switch A and C-CaM is converted into a stronger influence of switch A on C-CaM. Hel is no longer linked with C-CaM: the coupling between the two CaM domains is broken.

In EF-(4Ca-CaM) (Fig. 8 c), a reduction of the nucleation ensemble (CA, CB) and new long-range influences within the EF protein, from CB to both Hel and switch A, are observed. They suggest a more constrained and unstable house-of-cards architecture and reveal a destabilization of the EF domains that contact CaM. Such a destabilization can be related to the calcium-induced conformational change of N-CaM observed during this simulation (11), which find echoes up to the catalytic region of EF according to the generalized correlations (Fig. 6 c). Noticeably, the influence of CA on switch C is reinforced. In addition, the influence of Hel on C-CaM is maintained. The interdependence of switch A and C-CaM is converted into a stronger influence of switch A on C-CaM, as observed in EF-(0Ca-CaM).

The EDMs give a model for the relative energetic dependence of the domains inside the complex and vary according to the level of calcium. The hierarchy in the domains relative stabilities sheds a new light on the EF-CaM interaction mechanism described experimentally (20,21). Indeed, in EF-(2Ca-CaM), energetic influences are centered around the EF-CaM interface. Removing or adding calcium produces different effects on the dependency maps, both leading to a destabilization. On the one hand, in EF-(0Ca-CaM), couplings within the complex are broken, especially between the catalytic core and the helical domain of EF. On the other hand, in EF-(4Ca-CaM), the number of

**TABLE 7** Differences in total free binding energies and in each energy term corresponding to the energetic influences identified in EF-(0Ca-CaM) (<sup>0</sup>), EF-(2Ca-CaM) (<sup>2</sup>), and EF-(4Ca-CaM) (<sup>4</sup>), given in kcal/mol below, in the legend

	$\Delta\Delta\Delta E_{\text{ele}}$	$\Delta\Delta\Delta E_{\text{vdw}}$	$\Delta\Delta\Delta E_{\text{int}}$	$\Delta\Delta\Delta G_{\text{sol}}^{\text{polar}}$	$\Delta\Delta\Delta G_{\text{sol}}^{\text{nonpolar}}$	$\Delta\Delta\Delta G$
Short-range influences						
CA → SC ( <sup>0</sup> )	430 ± 64	105 ± 6	-8 ± 2	-478 ± 60	3 ± 2	52 ± 14
CA → SC ( <sup>2</sup> )	408 ± 38	107 ± 7	-9 ± 2	-464 ± 37	7 ± 2	49 ± 14
CA → SC ( <sup>4</sup> )	407 ± 34	116 ± 8	-9 ± 2	-457 ± 34	6 ± 2	63 ± 15
SA → C-CaM ( <sup>0</sup> )	803 ± 47	82 ± 6	0 ± 0	-807 ± 45	2 ± 2	80 ± 10
SA ↔ C-CaM ( <sup>2</sup> )	612 ± 47	79 ± 8	0 ± 0	-603 ± 46	2 ± 2	90 ± 14
SA → C-CaM ( <sup>4</sup> )	565 ± 37	77 ± 5	0 ± 0	-585 ± 33	0 ± 1	56 ± 9
Hel ↔ N-CaM ( <sup>0</sup> )	1715 ± 82	84 ± 8	0 ± 0	-1738 ± 89	-7 ± 2	54 ± 17
Hel → N-CaM ( <sup>2</sup> )	1411 ± 58	49 ± 8	0 ± 0	-1450 ± 60	-11 ± 2	-1 ± 14
N-CaM → Hel ( <sup>4</sup> )	1128 ± 76	64 ± 8	0 ± 0	-1159 ± 78	-6 ± 2	28 ± 11
Hel → C-CaM ( <sup>2</sup> )	666 ± 39	55 ± 5	0 ± 0	-689 ± 35	-6 ± 1	25 ± 15
Hel → C-CaM ( <sup>4</sup> )	768 ± 29	66 ± 5	0 ± 0	-807 ± 32	-8 ± 1	19 ± 13
SC → N-CaM ( <sup>0</sup> )	552 ± 29	3 ± 2	0 ± 0	-531 ± 28	0 ± 1	25 ± 6
N-CaM → SC ( <sup>2</sup> )	473 ± 27	4 ± 2	0 ± 0	-451 ± 27	0 ± 1	26 ± 5
N-CaM → SC ( <sup>4</sup> )	442 ± 34	6 ± 3	0 ± 0	-417 ± 31	-3 ± 1	29 ± 11
SA → N-CaM ( <sup>2</sup> )	692 ± 58	12 ± 2	0 ± 0	-682 ± 56	-4 ± 1	18 ± 7
Long-range influences						
CB → N-CaM ( <sup>2</sup> )	84 ± 13	0 ± 0	0 ± 0	-80 ± 13	0 ± 0	4 ± 2
CB → N-CaM ( <sup>4</sup> )	106 ± 14	0 ± 0	0 ± 0	-105 ± 13	0 ± 0	1 ± 1
CB → SA ( <sup>4</sup> )	22 ± 23	11 ± 4	0 ± 0	-34 ± 22	-3 ± 1	-5 ± 7
CB → Hel ( <sup>4</sup> )	-55 ± 6	0 ± 0	0 ± 0	54 ± 7	0 ± 0	-1 ± 1

$\Delta\Delta\Delta G[D^* \leftrightarrow D] = \Delta\Delta G[D, C \setminus D^*] - \Delta\Delta G[D, C] = \Delta\Delta G[D^*, C \setminus D] - \Delta\Delta G[D^*, C]$ . The arrow in  $\Delta\Delta\Delta G[D^* \rightarrow D]$  indicates that removal of  $D^*$  lowers the rank of  $D$ , and reciprocally. The arrow in  $\Delta\Delta\Delta G[D^* \leftrightarrow D]$  indicates that removal of  $D$  also lowers the rank of  $D^*$  (see Fig. 8). The quantity  $\Delta\Delta\Delta G[D^* \leftrightarrow D]$  is symmetric with respect to  $D$  and  $D^*$ .

nucleation points is decreased and new influences within EF reveal constraints that destabilize the complex architecture.

## DISCUSSION

EF-CaM is a good example of a protein-protein complex in which each partner displays remarkable conformational plasticity (11,20). The complex results from the assembly of two proteins composed of globular domains attached by a flexible linker: the N- and C-terminal domains linked by the central helix in CaM, the helical domain, and the catalytic core linked by switch C in EF. The residue network is organized into two blocks: 1), the interface region where CaM and EF contact each other, composed of Hel and switch A of EF and both domains of CaM; and 2), the catalytic region where the catalytic site lies, composed of domains CA and CB. Switch C plays an important role in maintaining the catalytic activity of the complex, as it both interacts with CaM and stabilizes switch B within the catalytic site.

We used here two different approaches to characterize the effect of  $\text{Ca}^{2+}$  on the global dynamics and energetics of EF-CaM, one based on dynamical correlations using generalized correlations and the other based on energetic influences using free energy calculations. Interestingly, we find good agreements between the results of the two methods, which can thus be interpreted in relation to each other in the framework of EF-CaM residue network. Both methods display an homogeneous and equilibrated picture of EF-(2Ca-CaM) architecture. Indeed, the stability of EF-(2Ca-CaM) is associated with moderate correlations over the whole residue network in the

GC matrix, and with balanced nucleation points and energetic influences essentially between EF and CaM on the EDMs. Destabilization of the complex by either binding or removal of calcium ions is also revealed by both methods: in EF-(0Ca-CaM), missing correlations in the GC matrix as well as missing energetic influences in the EDMs suggest a breakage of the couplings between the different parts of the residue network while in EF-(4Ca-CaM) high correlations in the GC matrix and new energetic influences between the domains of EF highlight an imbalance in the residue network.

Consequently, the observations from analysis of GCs and EDMs support the hypothesis that the stability of the complex depends upon the organization of its residue network and particularly upon the ability of each node of the network to communicate with all remaining parts of the complex through slightly correlated motions or balanced contributions to a house-of-card architecture. This finding is in agreement with Chennubhotla and Bahar (1) and Tobi and Bahar (43), who established the relationship among protein dynamics, information transmission (1), and protein-protein association process (43). We propose to denote this property as the residue network connectedness. As previously described, the connectedness of EF-CaM residue network is affected by the level of calcium complexation. On the one hand, the coupling breakage and the imbalance of the nucleation region in EF-(0Ca-CaM) reveal a lack of communication throughout the complex residue network. On the other hand, the high correlations between N-CaM and EF catalytic core and the long-range influences inside EF in EF-(4Ca-CaM) correspond to an extended connectedness

of the residue network. Because of this extended connectedness, the house-of-card architecture of the complex is destabilized. Both the lack of communication observed for EF-(0Ca-CaM) and the imbalance contributions observed for EF-(4Ca-CaM) are the signs of the complex destabilization according to the model presented here. They are thus in qualitative agreement with the relative stability of the EF-CaM interaction with different calcium levels (28).

The question remains, however: how can dynamical correlations and energetic contributions be formally related? On the one hand, the EDMs are based on relative but not necessarily absolute changes in free energies of binding, and they reveal energetic dependencies between domains within the framework of the complex stability or integrity. On the other hand, the EDMs reflect covariance dynamics to some extent through the support of domains to each other. This support is described by the action-reaction principle in classical mechanics, as the opposite directions of the action and reaction forces.

## CONCLUSION

We used various approaches, some recently developed, to analyze the geometrical and energetic bases of the effect of calcium removal or addition on the stability and connectedness of the EF-CaM residue network. While the elastic network model gave a picture of the natural collective motions of EF-CaM, the LFA sparsification algorithm, generalized correlations and dependency maps gave information relative to the calcium level. The LFA theory highlighted local dynamic regions that may be particularly sensitive and helped us to understand the propagation of the calcium information through CaM. Generalized correlations gave comprehensive pictures of the global dynamics of the complex. Finally, dependency maps revealed the energetic influences at play between the different domains of the complex. Interestingly, information based on the generalized correlations and that based on the energetics were in good agreement and could be related to give a better understanding of the calcium-dependent interaction mechanisms. This relationship was established through the concept of residue network connectedness, indicative of the complex stability.

The edema toxin of *B. anthracis* has recently been shown to impair the host immune system by inhibiting the expression of anthracidal phospholipase A2 by macrophages (44). Adenylyl cyclase EF thus appears as an interesting biological target for fighting anthrax. Perspectives for this work are the modeling of the conformational transition associated with the formation of the EF-CaM complex, to describe the mechanisms of the EF activation. Specific regions involved in the internal dynamics of the complex could be targeted to disturb the stability of the catalytic site and/or block the activation of EF.

We thank Julliane Yoneda for setting up the molecular dynamics simulations. The suggestion of the word “connectedness” by a reviewer is acknowledged.

This work was supported by the French Ministry of Defense (DGA/MRIS), the National Scientific Research Center (CNRS), and the Pasteur Institute. We also acknowledge the French National Calculations Centers IDRIS (Orsay) and CINES (Montpellier) for computational resources.

## REFERENCES

- Chennubhotla, C., and I. Bahar. 2007. Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput. Biol.* 3:1716–1726.
- Amadei, A., A. Linssen, and H. Berendsen. 1993. Essential dynamics of proteins. *Proteins*. 17:412–425.
- Abseher, R., and M. Nilges. 2000. Efficient sampling in collective coordinate space. *Proteins*. 39:82–88.
- Stepanova, M. 2007. Dynamics of essential collective motions in proteins: theory. *Phys. Rev. E*. 76:051918.
- Spiwok, V., P. Lipovová, and B. Králová. 2007. Metadynamics in essential coordinates: free energy simulation of conformational changes. *J. Phys. Chem. B*. 111:3073–3076.
- Kubitzki, M., and B. de Groot. 2007. Molecular dynamics simulations using temperature-enhanced essential dynamics replica exchange. *Biophys. J.* 92:4262–4270.
- Brooks, B., and M. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*. 80:6571–6575.
- Gō, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA*. 80:3696–3700.
- Tirion, M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
- Inin, B., K. Schulten, E. Tajkhorshid, and I. Bahar. 2008. Mechanism of signal propagation upon retinal isomerization: insights from molecular dynamics simulations of rhodopsin restrained by normal modes. *Biophys. J.* 95:789–803.
- Laine, E., J. Yoneda, A. Blondel, and T. Malliavin. 2008. The conformational plasticity of calmodulin upon calcium complexation gives a model of its interaction with the edema factor of *Bacillus anthracis*. *Proteins*. 71:1813–1829.
- Liddington, R. 2002. Anthrax: a molecular full Nelson. *Nature*. 415:373–374.
- Vogel, H. 1994. The Merck Frost Award Lecture 1994. Calmodulin: a versatile calcium mediator protein. *Biochem. Cell Biol.* 72:357–376.
- Chin, D., and A. Means. 2000. Calmodulin: a prototypical calcium sensor. *Trends Cell Biol.* 10:322–328.
- Bahler, M., and A. Rhoads. 2002. Calmodulin signaling via the IQ motif. *FEBS Lett.* 513:107–113.
- Hoeflich, K., and M. Ikura. 2002. Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell*. 108:739–742.
- Vetter, S., and E. Leclerc. 2003. Novel aspects of calmodulin target recognition and activation. *Eur. J. Biochem.* 270:404–414.
- Ikura, M., and J. Ames. 2006. Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: two ways to promote multifunctionality. *Proc. Natl. Acad. Sci. USA*. 103:1159–1164.
- Ascenzi, P., P. Visca, G. Ippolito, A. Spallarossa, M. Bolognesi, et al. 2002. Anthrax toxin: a tripartite lethal combination. *FEBS Lett.* 531:384–388.
- Drum, C., S. Yan, J. Bard, Y. Shen, D. Lu, et al. 2002. Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature*. 415:396–402.
- Shen, Y., N. Zhukovskaya, Q. Guo, J. Florian, and W. Tang. 2005. Calcium-independent calmodulin binding and two-metal-ion catalytic mechanism of anthrax edema factor. *EMBO J.* 24:929–941.
- Capozzi, F., F. Casadei, and C. Luchinat. 2006. EF-hand protein dynamics and evolution of calcium signal transduction: an NMR view. *J. Biol. Inorg. Chem.* 11:949–962.

23. Gifford, J., M. Walsh, and H. Vogel. 2007. Structures and metal-ion-binding properties of the  $\text{Ca}^{2+}$ -binding helix-loop-helix EF-hand motifs. *Biochem. J.* 405:199–221.
24. Finn, B., J. Evenas, T. Drakenberg, J. Waltho, E. Thulin, et al. 1995. Calcium-induced structural changes and domain autonomy in calmodulin. *Nat. Struct. Biol.* 2:777–783.
25. Zhang, M., T. Tanaka, and M. Ikura. 1995. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat. Struct. Biol.* 2:758–767.
26. Ishida, H., K. Takahashi, K. Nakashima, Y. Kumaki, M. Nakata, et al. 2000. Solution structures of the N-terminal domain of yeast calmodulin:  $\text{Ca}^{2+}$ -dependent conformational change and its functional implication. *Biochemistry*. 39:13660–13668.
27. Yang, C., G. Jas, and K. Kuczera. 2004. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *Biochim. Biophys. Acta.* 1697:289–300.
28. Ulmer, T., S. Soelaiman, S. Li, C. Klee, W. Tang, et al. 2003. Calcium dependence of the interaction between calmodulin and anthrax edema factor. *J. Biol. Chem.* 278:29261–29266.
29. Case, D., T. Darden, T. Cheatham, C. Simmerling, J. Wang, et al. 2004. AMBER 8. University of California, San Francisco.
30. Aqvist, J. 1990. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* 94:8021–8024.
31. Fiorin, G., R. Biekofsky, A. Pastore, and P. Carloni. 2005. Unwinding the helical linker of calcium-loaded calmodulin: a molecular dynamics study. *Proteins*. 61:829–839.
32. Case, D., T. Darden, T. Cheatham, C. I. Simmerling, J. Wang, et al. 2006. AMBER 9. University of California, San Francisco.
33. Delano, W.L. 2002. The PyMOL Molecular Graphics System. <http://www.pymol.org>.
34. Bahar, I., A. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* 2:173–181.
35. Lange, O., and H. Grubmüller. 2006. Generalized correlation for biomolecular dynamics. *Proteins*. 62:1053–1061.
36. van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, et al. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
37. R Development Core Team. 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>, ISBN 3-900051-07-0.
38. Gordon, A. 1999. Classification, 2nd Ed.. Chapman & Hall/CRC, Boca Raton, FL.
39. Zhang, Z., and W. Wriggers. 2006. Local feature analysis: a statistical theory for reproducible essential dynamics of large macromolecules. *Proteins*. 64:391–403.
40. Hamacher, K., J. Trylska, and J. McCammon. 2006. Dependency map of proteins in the small ribosomal subunit. *PLoS Comput. Biol.* 2:e10.
41. Srinivasan, J., T. Cheatham, J. Cieplak, P. Kollman, and D. Case. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate—DNA helices. *J. Am. Chem. Soc.* 120:9401–9409.
42. Marsaglia, G., W. W. Tsang, and J. Wang. 2003. Evaluating Kolmogorov's distribution. *J. Stat. Software.* 8:1–4.
43. Tobi, D., and I. Bahar. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA.* 102:18908–18913.
44. Raymond, B., D. Leduc, L. Ravaux, R. Goffic, T. Candela, et al. 2007. Edema toxin impairs anthracidal phospholipase A2 expression by alveolar macrophages. *PLoS Pathog.* 3:e187.