



**HAL**  
open science

# Analyse du contrôle allostérique et prédiction de structure pour une toxine de pathogène : l'apport des simulations de dynamique moléculaire

Edithe Selwa

► **To cite this version:**

Edithe Selwa. Analyse du contrôle allostérique et prédiction de structure pour une toxine de pathogène : l'apport des simulations de dynamique moléculaire. Biochimie, Biologie Moléculaire. Université Pierre et Marie Curie - Paris VI, 2012. Français. NNT : 2012PAO66681 . tel-00836270

**HAL Id: tel-00836270**

**<https://theses.hal.science/tel-00836270>**

Submitted on 20 Jun 2013

**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

**Edithe Selwa**

Pour obtenir le grade de

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

Sujet de la thèse :

**Analyse du contrôle allostérique et prédiction de structure**  
**pour une toxine de pathogène:**  
**l'apport des simulations de dynamique moléculaire**

Soutenue le 25 Septembre 2012 devant le jury composé de :

Jean Cognet	Président
Paul Fleurat-Lessard	Rapporteur
Charles Robert	Rapporteur
Thérèse Malliavin	Directeur de thèse
Daniel Ladant	Examinateur



La protéine CyaA est un facteur de virulence majeur de *Bordetella pertussis*, impliqué dans la maladie de la coqueluche. Le domaine catalytique AC de CyaA est directement transféré dans la cellule hôte eucaryote, où il est activé comme adénylcyclase par la calmoduline, une protéine ubiquitaire et sensible aux ions calcium. Ainsi, AC transforme l'ATP en AMPc de manière incontrôlée, ce qui conduit à des dérèglements cellulaires.

Seule la structure de AC complexé à la calmoduline chargée d'ions calcium avait été résolue par cristallographie aux rayons X. À partir de cette structure, des simulations de dynamique moléculaire de AC libre, et en complexe avec la calmoduline nous ont permis de caractériser l'effet de la calmoduline et des ions calcium sur la plasticité conformationnelle du complexe. Les tendances conformationnelles de AC libre ont aussi été étudiées. L'analyse conjointe des influences énergétiques et des liaisons hydrogènes a révélé un réseau d'interactions entre AC et la calmoduline, dans lequel trois résidus clés, susceptibles de jouer un rôle allostérique sur l'activité de l'adénylcyclase ont été modifiés par mutagenèse dirigée. Ces tests expérimentaux ont conduit à la mise en évidence d'une région allostérique qui assure la communication de l'information de transition conformationnelle entre le site de fixation de la calmoduline et le site catalytique. Une exploration conformationnelle plus approfondie de AC à l'état non-lié a été entreprise par une méthode innovante de dynamique accélérée par la température (*TAMD*). Elle nous a conduit à la prédiction de conformations échantillonnées de AC dans son état libre. Ces prédictions pourraient être utilisées à l'avenir pour stabiliser l'état libre et faciliter l'étude expérimentale de sa structure.

**Mots clés :** *Bordetella pertussis*, calmoduline, calcium, dynamique moléculaire, allostérie, dynamique accélérée par la température, *TAMD*.

Ce travail a été réalisé d'octobre 2009 à octobre 2012, 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 au sein de l'Unité de Biochimie des Interactions Macromoléculaires, sous la direction de Daniel Ladant. Les travaux sur la *TAMD* ont été réalisés en collaboration avec Luca Maragliano (ITT, Gênes) et Giovanni Ciccotti (La Sapienza, Rome et UCD, Dublin).



## **Analysis of the allosteric control and prediction of a structure for a pathogenic toxin : the contribution of the molecular dynamics simulations**

The protein CyaA is an important virulence factor of *Bordetella pertussis* involved in the whooping cough disease. The AC domain of CyaA is activated in eukaryotic host cells as an adenylylase by the ubiquitous and calcium sensitive protein calmodulin (CaM). This activation transforms the ATP into cAMP in an uncontrolled way, leading to severe cellular dysfunction.

Only the structure of AC complexed with calcium-loaded calmodulin has been resolved by X-ray crystallography. Starting from this structure, molecular dynamics simulations of AC free and in complex with calmodulin enabled to characterize the effect of the calmodulin and calcium ions on the conformational plasticity of the complex. Conformational tendencies of free AC were also studied. The joint analysis of energetic influences and hydrogen bonds revealed a network of interactions between AC and calmodulin, in which three key residues, that may play an allosteric role on the catalytic activity of the adenylylase, were modified by site-directed mutagenesis. These experimental tests have led to the identification of an allosteric region which assures the communication of the conformational transition information between the binding site of the calmodulin and the catalytic site. A more thorough conformational exploration of the unbound AC state was undertaken by an innovative temperature accelerated molecular dynamics (*TAMD*) method. It led to a prediction of some sampled conformations of AC in its unbound state. These predictions could be used in the future to stabilize the free state and to facilitate the experimental study of its structure.

**Key words :** *Bordetella pertussis*, calmodulin, calcium, molecular dynamics, allostery, temperature accelerated molecular dynamics, TAMD.

# Remerciements

Je voudrais tout d'abord remercier Michael Nilges, directeur de l'unité de Bioinformatique Structurale et du département de Biologie Structurale et Chimie, pour m'avoir offert l'opportunité d'intégrer son laboratoire et de travailler au sein d'une équipe scientifique de très haute qualité. Je tiens également à remercier Thérèse Malliavin qui m'a accordée sa confiance et guidée tout au long de ces trois années de thèse. Je suis également très reconnaissante envers Elodie Laine, dont les travaux de thèse ont inspiré une grande partie de mes travaux.

Les résultats présentés ici sont issus de nombreuses collaborations, notamment avec l'unité de Biochimie des interactions macromoléculaires, dirigée par Daniel Ladant envers qui j'exprime ma grande reconnaissance pour m'avoir permis de réaliser des travaux expérimentaux au sein de son laboratoire en toute confiance. Un grand merci à Marilynne Davi, pour m'avoir supervisée en biologie moléculaire et pour sa gentillesse. Je remercie également Alexandre Chenal pour sa contribution dans la réalisation expérimentale d'une partie de ma thèse, ainsi que pour me donner la possibilité de participer à la réalisation de son projet. Merci également à Johanna Spiaczka-Karst et Ana Cristina Sotomayor Pérez pour leur contribution, ainsi qu'à tous les membres de l'équipe BIM pour l'accueil chaleureux qu'ils m'ont réservé. Je tiens également à exprimer ma gratitude envers Luca Maragliano (ITT, Gênes) et Giovanni Ciccotti (La Sapienza, Rome et UCD, Dublin) pour leur collaboration et leur grande aide sur la TAMD.

A l'équipe BIS, Arnaud Blondel, Nathalie Duclert, et tous les autres membres, je vous suis reconnaissante pour les échanges et les regards critiques qu'ils ont apportés sur mes travaux. Merci à Tru Huynh, pour toutes les solutions apportées face aux difficultés et aux problèmes informatiques rencontrés. Un grand merci à Renée Communal, qui a énormément apporté à la vie au sein du laboratoire, et pour toute l'aide administrative qu'elle m'a apportée. Je remercie tous les membres, anciens membres et amis de BIS, en particulier Fabien, Nathan, Yannick, Guillaume, Paulo et Claire, pour leur soutien quotidien et leur bonne humeur qui ont beaucoup apporté aux conditions de travail au sein de l'équipe. Je suis très heureuse d'avoir passé ces trois années en votre compagnie.

Je remercie Elisabeth Labruyère-Dadaglio de l'Institut Pasteur pour les entretiens dans le cadre du tutorat, et Germain Trugnan, directeur de l'école doctorale iViv, pour son suivi régulier et les formations dont il m'a permis de bénéficier. J'adresse ma reconnaissance aux membres du jury de cette thèse, Paul Fleurat-Lessard, Charles Robert et Jean Cognet pour avoir accepté de juger cette thèse.

Pour terminer, j'aimerais remercier ma maman, pour tout l'amour et tout le soutien qu'elle me porte, et mon papa, qui veille sur moi de là-haut et que j'espère rendre fier. Merci à tous mes amis fidèles, pour m'avoir accompagnée au cours de ces trois années.



*Un problème livré à lui-même se dessèche ou pourrit.  
Mais fertilisez un problème à l'aide d'une solution  
et vous allez en faire éclore des dizaines.*

Norman Frederik Simpson



# Table des matières

	<b>Abbreviations</b>	<b>xiii</b>
<b>I</b>	<b>Interaction entre biomolécules et dynamique des complexes</b>	<b>1</b>
1	Interactions entre biomolécules . . . . .	1
1.1	Différents type d'interaction : protéine-protéine, protéine-ADN, protéine-sucre, protéine-ligand . . . . .	1
1.2	Cibles de choix pour influencer sur les mécanismes biologiques . . . . .	6
2	Méthode d'étude des complexes . . . . .	7
2.1	Méthodes de détermination des structures de complexes . . . . .	7
	La cristallographie aux rayons X . . . . .	8
	La Résonance Magnétique Nucléaire (RMN) . . . . .	9
	Couplage dipolaire résiduel (CDR) . . . . .	12
2.2	Mouvements induits sur les structures de protéines par la formation d'un complexe . . . . .	12
3	Importance du désordre en biologie . . . . .	14
3.1	Importance fonctionnelle du désordre . . . . .	14
3.2	Rôle de la dynamique dans la formation de complexes . . . . .	15
	Les trois modèles d'interactions protéine-protéine . . . . .	16
	Les profils énergétiques de ces trois modèles . . . . .	17
	Les avantages . . . . .	18
<b>II</b>	<b>Les toxines : un modèle d'interaction protéine-protéine</b>	<b>21</b>
1	Le rôle des toxines dans l'infection par les bactéries pathogènes . . . . .	21
1.1	Panorama des toxines de pathogènes . . . . .	21
1.2	Toxines de pathogènes augmentant la production d'AMPc . . . . .	25
	Rôle de l'AMPc dans la transduction de signaux . . . . .	25
	Rôle de l'AMPc chez <i>E.coli</i> . . . . .	28
	Les toxines entraînant une surproduction d'AMPc . . . . .	29
2	La toxine adénylcyclase CyaA . . . . .	32
2.1	La coqueluche . . . . .	32
	Epidémiologie . . . . .	32
	Le germe pathogène et la maladie . . . . .	32
	Mécanisme d'infection et symptômes . . . . .	33
	Traitements et vaccination . . . . .	34
2.2	L'adénylcyclase CyaA : Biogenèse, Structure et mécanisme d'invasion . . . . .	35
	Biogenèse et Organisation structurale de CyaA . . . . .	35

	Mécanisme d'invasion des cellules eucaryotes . . . . .	37
2.3	CyaA, un outil biotechnologique . . . . .	38
2.4	Connaissances structurales et cinétiques . . . . .	41
	La calmoduline et le calcium . . . . .	42
	Le complexe AC/CaM . . . . .	43
	Comparaison avec le complexe EF/CAM . . . . .	45
2.5	Carte des mutants de AC . . . . .	48
3	Plan de l'étude . . . . .	51
<b>III Présentation de quelques méthodes d'exploration de l'espace conformationnel des protéines par des simulations de DM</b>		
1	La dynamique moléculaire . . . . .	53
1.1	Principe de la dynamique moléculaire classique . . . . .	53
1.2	Champ de force . . . . .	54
1.3	Exemple de protocole de simulation d'un système . . . . .	57
2	La dynamique moléculaire avancée . . . . .	59
2.1	Méthodes basées sur la mécanique moléculaire . . . . .	60
	Le réseau élastique . . . . .	60
	Affinement par pic conjugué ( <i>Conjugate Peak Refinement</i> ou <i>CPR</i> ) . . . . .	60
2.2	Méthodes fondées sur la dynamique moléculaire . . . . .	61
	Dynamique moléculaire ciblée ou <i>Targeted Molecular Dynamics (TMD)</i> . . . . .	61
	Dynamique moléculaire dirigée ou <i>Steered Molecular Dynamics (SMD)</i> . . . . .	62
	Echange de répliques (ou <i>Replica Exchange (RE)</i> ) . . . . .	62
	Adaptive Tempering . . . . .	62
	Dynamique moléculaire accélérée ( <i>Accelerated Molecular Dynamics (AMD)</i> ) . . . . .	63
	Dynamique accélérée par la température ( <i>Temperature Accelerated Dynamics (TAD)</i> ) . . . . .	63
	La méta-dynamique ( <i>Metadynamics</i> ) . . . . .	64
	Dynamique moléculaire accélérée par la température ou <i>Temperature accelerated molecular dynamics (TAMD)</i> . . . . .	65
<b>IV Rôle de la calmoduline et des ions <math>\text{Ca}^{2+}</math> dans la stabilisation du domaine catalytique de l'adénylcyclase CyaA de <i>B. pertussis</i></b>		
1	Introduction . . . . .	67
2	Matériels et Méthodes . . . . .	68
2.1	Préparation des systèmes simulés . . . . .	68
2.2	Simulations de Dynamique Moléculaire . . . . .	70
2.3	Convergence des trajectoires . . . . .	70
2.4	Carte de dépendances énergétiques . . . . .	71
3	Résultats . . . . .	73
3.1	Dérives conformationnelles . . . . .	73
3.2	Tendances conformationnelles de AC et activité catalytique . . . . .	78
3.3	Conformation de C-CaM et interaction avec AC . . . . .	83
3.4	Influences énergétiques . . . . .	85
4	Discussion . . . . .	90
5	Conclusion . . . . .	91

<b>V</b>	<b>Influence de mutations introduites dans AC sur son affinité pour la calmoduline</b>	<b>93</b>
1	Introduction . . . . .	93
2	Matériels et méthodes . . . . .	95
2.1	Simulations de Dynamique Moléculaire . . . . .	95
2.2	Construction des plasmides . . . . .	97
2.3	Expression et purification des variants de AC . . . . .	98
2.4	Détermination de l'activité enzymatique adénylyclase . . . . .	99
3	Résultats . . . . .	100
3.1	L'interface C-CaM/AC dans les structures cristallographiques . . . . .	100
3.2	Analyse de l'interaction C-CaM/AC au cours des trajectoires de DM . . . . .	102
3.3	Les fluctuations de C-CaM dans les trajectoires de DM . . . . .	107
3.4	Réseau d'interactions au sein de AC au cours des trajectoires de DM . . . . .	109
3.5	Caractérisation expérimentale des variants de AC modifiés . . . . .	111
4	Discussion . . . . .	114
5	Conclusion . . . . .	116
<b>VI</b>	<b>Exploration de la transition conformationnelle de l'état lié à l'état non-lié de AC</b>	<b>117</b>
1	Introduction . . . . .	117
2	Matériels et méthodes . . . . .	120
2.1	Simulations de Dynamique Moléculaire . . . . .	120
2.2	Simulations TAMD . . . . .	120
2.3	Echantillonnage des conformations . . . . .	122
3	Résultats . . . . .	123
3.1	Choix des variables collectives . . . . .	123
3.2	Exploration de l'espace conformationnel de AC . . . . .	125
3.3	Variation de la structure du solvant . . . . .	127
3.4	Prédiction de conformations de AC stables dans le bassin d'énergie de AC isolé . . . . .	130
4	Discussion . . . . .	139
5	Conclusion . . . . .	140
<b>VII</b>	<b>Conclusion générale et perspectives</b>	<b>143</b>
	<b>Bibliographie</b>	<b>147</b>
	<b>Annexes</b>	<b>160</b>
A.1	Tableau des pourcentages de formation des liaisons hydrogène établies entre C-CaM et AC. . . . .	160
A.2	Tableau des pourcentages de formation des liaisons hydrogène entre les résidus situés dans la boucle L $\alpha$ 16Cter reliant l'hélice $\alpha$ 16 et la région Cterminale. . . . .	162
A.3	Séquence ADN et traduction en acides aminés du variant ACm3A . . . . .	164
A.4	Séquence ADN et traduction en acides aminés du variant ACm2A . . . . .	165
A.5	Séquence ADN et traduction en acides aminés du variant ACm1A . . . . .	166
A.6	Publication décrivant le rôle de la calmoduline et des ions calcium dans la stabilisation du domaine catalytique de l' adényl-cyclase CyaA de <i>Bordetella pertussis</i> . . . . .	167





# ABBREVIATIONS

<b>3D</b>	en trois dimensions
<b>AC</b>	adénylcyclase (Domaine catalytique de CyaA)
<b>ACP</b>	analyse en Composantes Principales
<b>ADN</b>	acide désoxyribonucléique
<b>AMD</b>	<i>accelerated molecular dynamics</i>
<b>AMPc</b>	adénosine monophosphate cyclique
<b>APC</b>	cellule présentatrice d'antigène
<b>ARN</b>	acide ribonucléique
<b>ATP</b>	adénosine triphosphate
<b>AT</b>	<i>anthrax toxin</i>
<b>CaM</b>	calmoduline
<b>C-CaM</b>	lobe C-CaM de la calmoduline
<b>CD</b>	cellule dendritique
<b>CDK</b>	kinase dépendante des cyclines
<b>CDKI</b>	inhibiteur de kinase dépendante des cyclines
<b>CDR</b>	couplage dipolaire résiduel
<b>CMH</b>	complexe majeur d'histocompatibilité
<b>COM</b>	<i>Center of mass</i>
<b>CPR</b>	<i>conjugate peak refinement</i>
<b>CREB</b>	<i>cAMP response element binding protein</i>
<b>CREM</b>	<i>cAMP-responsive-element-binding modulator</i>
<b>CT</b>	toxine cholérique
<b>CTL</b>	lymphocytes T Cytotoxiques
<b>DM</b>	dynamique moléculaire
<b>DT</b>	Toxine diphtérique
<b>EDM</b>	<i>energetic dependency maps</i>
<b>EF</b>	Facteur œdémateux
<b>ENM</b>	<i>elastic network model</i>
<b>FFT</b>	<i>fast Fourier transform</i>
<b>FRET</b>	<i>fluorescence resonance energy transfer</i>
<b>GDP</b>	guanosine diphosphate
<b>GTP</b>	guanosine triphosphate
<b>LF</b>	Facteur Létal
<b>MC</b>	Monte Carlo
<b>MD</b>	<i>metadynamics</i>
<b>MMPBSA</b>	<i>molecular mechanics Poisson-Boltzmann surface area</i>
<b>NMA</b>	<i>Normal mode analysis</i>
<b>OMS</b>	organisation mondiale pour la santé
<b>PA</b>	antigène protecteur
<b>PDB</b>	<i>Protein Data Bank</i>
<b>Pi</b>	phosphate inorganique
<b>PID</b>	protéine intrinsèquement désordonnée
<b>PKA</b>	protéine kinase A
<b>PKR</b>	protéine kinase R
<b>PME</b>	<i>particle mesh Ewald</i>
<b>PPiase</b>	pyrophosphatase inorganique
<b>PPi</b>	pyrophosphate inorganique

<b>RE</b>	<i>replica exchange</i>
<b>RMN</b>	résonance magnétique nucléaire
<b>RMSD</b>	<i>root mean square deviation</i>
<b>RMSF</b>	<i>root mean square fluctuation</i>
<b>RX</b>	crystallographie 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>TAC</b>	<i>toxine adénylcyclase</i>
<b>TAD</b>	<i>Temperature Accelerated Dynamics</i>
<b>TAMD</b>	<i>Temperature Accelerated Molecular Dynamics</i>
<b>TMD</b>	<i>Targeted molecular dynamics</i>

# Chapitre I

## Interaction entre biomolécules et dynamique des complexes

---

L'objet de la présente thèse est de montrer, sur l'exemple de la toxine adénylcyclase de *Bordetella pertussis*, que la bioinformatique donne des informations pour explorer les modes d'interaction des biomolécules. Les interactions entre biomacromolécules sont à la base de la plupart des phénomènes biologiques. Depuis 50 ans, le développement de la biologie structurale a permis de commencer à relier ces interactions globales à des interactions physico-chimiques au niveau atomique. Une meilleure compréhension de ces interactions est indispensable pour espérer un jour pouvoir contrôler leur établissement.

Dans ce travail de thèse, nous nous sommes intéressés au développement de méthodes pour prédire et analyser les changements conformationnels subis par une protéine lors de l'établissement d'une interaction. Nous avons choisi comme exemple d'application le domaine catalytique de l'adénylcyclase CyaA de *Bordetella pertussis*. En effet, c'est un bon exemple d'interaction dans laquelle les deux partenaires, subissent de grands réarrangements conformationnels. De plus, c'est un système qui a été largement étudié depuis 20 ans par notre collaborateur Daniel Ladant (Unité Biochimie des Interactions macromoléculaires, Pasteur), ce qui nous a donné accès à la possibilité de tester expérimentalement des hypothèses issues de l'analyse bioinformatique.

## 1 Interactions entre biomolécules

### 1.1 Différents type d'interaction : protéine-protéine, protéine-ADN, protéine-sucre, protéine-ligand

Les protéines représentent l'échafaudage de la vie. Elles assurent des fonctions cruciales dans tous les processus biologiques essentiels, comme les fonctions de transports ou de stockage d'autres molécules, la transduction du signal, un support mécanique pour la défense immunitaire, ou encore le contrôle de la croissance et de la différenciation cellulaire. Certaines sont des enzymes spécifiques qui catalysent des réactions chimiques importantes dans les systèmes biologiques.

Leurs propriétés fonctionnelles sont déterminées par leur structure tridimensionnelle, mais c'est surtout leur capacité à interagir avec d'autres biomolécules et à former des complexes qui est responsable de leur rôle dominant en biochimie.

Il existe différents types d'interaction entre une protéine et une autre biomolécule :

#### – Interaction protéine-protéine

La plupart des protéines s'assemblent pour former un complexe pour accomplir leur fonction. Une étude expérimentale récente a montré pour la première fois que chez la levure, 80% des protéines purifiées (soit environ 2000) étaient sous forme de complexe [1]. Ceci illustre l'importance des interactions protéine-protéine dans la cellule. Actuellement, les principes qui régissent ces interactions ne sont que partiellement et assez mal compris. Il existe des bases de données recensant toutes les interactions protéine-protéine, telles que STRING (<http://string-db.org/>), BioGRID (<http://thebiogrid.org/>) ou encore Brookhaven Protein Databank ([http://www-jmg.ch.cam.ac.uk/cil/SGTL/Brookhaven.html](http://www.jmg.ch.cam.ac.uk/cil/SGTL/Brookhaven.html)).

Selon Jones (1996) [2], les complexes se divisent en quatre catégories : (i) les homodimères, avec deux sous-unités identiques, (ii) les complexes enzymes-inhibiteurs, (iii) les complexes antigène-protéine, (iv) tous les autres hétérocomplexes incluant des complexes permanents et des complexes entre monomères indépendants non-permanents.

Il a été décrit que les complexes permanents (ceux qui ne se dissocient pas) possèdent des interfaces protéine-protéine qui sont plus compactes et moins planaires et avec moins de liaisons hydrogène que les autres complexes.

L'étude de la proportion de chaque acide aminé impliqué dans la surface d'interaction a révélé que les résidus hydrophobes sont principalement impliqués dans les interactions entre homodimères. En ce qui concerne les hétérodimères, la faible proportion de ces résidus hydrophobes est équilibrée par une proportion plus importante de résidus polaires. Cette différence peut être expliquée par les rôles que jouent ces deux types de complexes. En effet, les homodimères fonctionnent rarement en tant que monomères, c'est pourquoi leurs surfaces hydrophobes sont enfouies en permanence dans le complexe. Beaucoup d'hétérocomplexes sont souvent dissociés en monomères en solution et assurent une fonction biologique dans cet état. Leurs interfaces ne peuvent donc pas être aussi hydrophobes que celles des homodimères, car une exposition trop importante des régions hydrophobes serait énergétiquement défavorable. Les hétérocomplexes qui ne se dissocient pas présentent un taux de liaisons hydrogène équivalent à celui des homodimères.

Le nombre de segments peptidiques interagissant à l'interface est également important. Les résidus d'interface séparés par au moins cinq résidus appartiennent à un segment différent. En général, le nombre de segments interagissant à l'interface de deux protéines varie entre 1 et 11, cependant, ce nombre est compris entre deux ou cinq pour les complexes enzyme-inhibiteur, lorsque l'inhibiteur est un peptide. Les inhibiteurs peptidiques ont pour but de mimer un long segment de chaîne polypeptidique, dans la conformation requise pour le clivage par une enzyme, et de ce fait, elles présentent toutes une boucle canonique qui domine l'interaction. A l'inverse, les autres interfaces sont très segmentées. Les structures secondaires impliquées sont souvent

des combinaisons entre les hélices  $\alpha$ , les brins et les boucles, même si parfois un seul type est impliqué. Les interfaces impliquant les feuillettes  $\beta$  se divisent en trois catégories, (i) celles qui interagissent en étendant des feuillettes à travers des liaisons hydrogène classiques, (ii) dans lesquelles les feuillettes s'empilent au-dessus d'un autre, (iii) celles où les structures sont mixtes, ni étendues, ni empilées.

#### – Interaction protéine-ADN/ARN

Des fonctions importantes de l'ADN dépendent d'interactions avec les protéines. Ces interactions peuvent être spécifiques ou non à une séquence d'ADN unique.

Parmi les interactions non-spécifiques, on retrouve les histones qui sont des protéines permettant l'enroulement de l'ADN pour le rendre plus compact. Chez les procaryotes, ces histones sont remplacés par une multitude de protéines. Ces interactions non-spécifiques sont formées par des résidus basiques des histones, qui forment des liaisons ioniques avec le groupement phosphate du squelette de l'ADN. C'est pourquoi elles sont indépendantes de la séquence. D'autres protéines peuvent également se lier de manière non-spécifique à un ADN courbé pour permettre le sur-enroulement de celui-ci afin de former le chromosome. Parmi les interactions protéine-ADN non-spécifiques de la séquence, il existe un groupe de protéines qui se lient spécifiquement à un simple brin de l'ADN pour le répliquer. Les polymérases par exemple sont particulièrement importantes pour la réplication et la transcription de l'ADN. La protéine de réplication A est également impliquée dans la réplication de l'ADN, mais aussi dans la recombinaison et la réparation de celui-ci.

A l'inverse, certaines protéines se lient à une séquence spécifique particulière de l'ADN. Cela concerne surtout les facteurs de transcription, qui sont des protéines régulant la transcription de l'ADN en ARNm en l'activant ou l'inhibant. Des modifications d'un type de facteur de transcription peuvent engendrer des conséquences sur des milliers de gènes. Par conséquent, ces protéines sont souvent les cibles des processus de transduction de signal qui contrôlent les réponses aux changements environnementaux.

Quatre types d'interactions potentielles peuvent exister entre protéines et ADN. (i) Des ponts salins formés entre les phosphates et les chaînes latérales d'acides aminés chargés positivement (N de la Lys, guanidinium de l'arginine, histidine protonée). (ii) Des liaisons hydrogène pouvant s'établir entre les phosphates, sucres et bases des acides nucléiques et les liaisons peptidiques ou les chaînes latérales des acides aminés hydrophiles. (iii) Des interactions d'empilement (*dits "stackings"*) impliquant les chaînes latérales des acides aminés aromatiques (Trp, Tyr, Phe, His) et les bases. (iv) Des interactions hydrophobes entre les bases des acides nucléiques et les chaînes latérales des acides aminés non polaires.

Les protéines décrivent des motifs de fixation aux acides nucléiques. Il y a par exemple le motif "Hélice-tour-hélice" portés par des protéines qui sont en général sous forme dimérique avec des sites de fixation palindromiques sur l'ADN. La première hélice sert de verrou conformationnel qui permet l'orientation correcte de l'hélice de reconnaissance, qui doit se coucher à l'intérieur du grand sillon de l'ADN. Il y a également le motif en brins  $\beta$  qui s'étend dans le grand ou

le petit sillon pour y établir de nombreux contacts. Il est souvent accompagné d'une hélice  $\alpha$  qui apporte la spécificité de la reconnaissance de l'ADN. Les petits domaines en doigts à Zinc interagissent également avec l'ADN et l'ARN. Ils requièrent la coordination de plusieurs atomes de zinc pour stabiliser leur repliement. Leurs structures sont constituées de deux feuilletts  $\beta$  antiparallèles et d'une hélice  $\alpha$ . Ce motif leur permet d'interagir dans les grands sillons, avec quatre nucléotides. Le motif "Leucine Zipper" est encore un autre motif particulier, qui fait intervenir la répétition de leucines tous les sept résidus pour les positionner du même côté d'une longue hélice amphipatique, ce qui la stabilise par rapport au corps plus ou moins globulaire de la protéine. Cette hélice interagit avec une autre hélice du même genre pour former le zip. Ce n'est pas un motif qui se fixe à l'ADN, mais plutôt un motif qui permet le positionnement correct d'une région de la protéine se fixant à l'ADN.

#### – Interaction protéine-sucre

Les lectines sont définies comme des protéines d'origine non-immune qui se lient spécifiquement et de façon réversible aux sucres et ne montrent aucune activité enzymatique pour ces substrats [3]. Elles se retrouvent dans toutes les classes d'organismes (virus, bactéries, plantes, insectes, animaux) et sont capables d'agglutiner les cellules et de les glycoconjuguer. Ces molécules possèdent une capacité unique de "lire" l'information biologique qui est codifiée dans la structure tridimensionnelle des sucres. Elles représentent en fait les récepteurs spécifiques pour les interactions protéine-sucre qui jouent un rôle clé dans une multitude de processus de reconnaissance moléculaire et de signalisation cellulaire. Situées à la surface des bactéries, des virus, ou des parasites intestinaux, elles reconnaissent les glycoconjugués présents sur la surface des cellules épithéliales et facilitent donc les processus de colonisation et d'infection. Les lectines animales ont des rôles très importants dans la croissance et le développement des organismes supérieurs. Dans les lectines spécifiques pour les monosaccharides, les sites de liaison sont généralement des dépressions peu profondes sur la surface de la protéine, tandis que dans les lectines spécifiques aux oligosaccharides, les sites de liaison sont plus profonds et montrent une complémentarité pour le ligand semblable à celle qui existe entre les enzymes-substrats. Les interactions lectine-ligand se caractérisent par des liaisons hydrogène fortes formées par les chaînes latérales des résidus chargés. Les groupements NH ou CO de la chaîne principale participent à la reconnaissance. Parfois, les interactions hydrophobes jouent aussi un rôle dans la force d'interaction. Les sucres sont des molécules très polaires mais la disposition des groupements OH peut créer des zones hydrophobes qui peuvent donner lieu à des interactions avec les résidus aromatiques. Ce type d'interaction est appelé "*stacking interaction*" [4]. Certaines lectines montrent une présence exceptionnelle de calcium dans le site de liaison pour interagir directement avec le sucre.

Chez les glycoprotéines, les domaines carbohydrates ont des propriétés hydrophiles et visqueuses qui sont importantes pour leur fonction. Les domaines carbohydrates sont liés de manière covalente aux polypeptides *via* les chaînes latérales de Thr, Ser ou Asn. Mais dans certaines glycoprotéines, comme par exemple les ribonucléases du pancréas, ces domaines ne sont ni avantageux, ni désavantageux, mais reflètent le fait qu'un enzyme non-spécifique liant le sucre reconnaît un

motif structural d'une protéine en tant que substrat. Les complexes carbohydate-protéine artificiels sont plus résistants à la température, aux dénaturants et aux protéases, ce qui fait d'eux des candidats potentiels pour en faire des agents thérapeutiques.

Dans un autre exemple, les sucres jouent un rôle important dans les interactions spécifiques entre les surfaces des cellules et les macromolécules solubles. La reconnaissance intercellulaire, comme la formation des tissus issus de différents types cellulaires, est également fondée sur la diversité des glycoprotéines.

#### – Interaction protéine-lipide

Les protéines pénètrent les interfaces des phases aqueuses et apolaires. Elles permettent de transporter les lipides dans les solutions aqueuses, comme par exemple le cholestérol dans le sang, et elles prennent part aux activités membranaires [5]. Dans les membranes, les lipides forment des barrières et des compartiments biologiques. Les rôles de transduction du signal, de transports spécifiquement membranaires ou de transduction d'énergie sont quand à eux assurés par les protéines. Les lipoprotéines sont difficiles à cristalliser, les séquences en acides aminés indiquent que la partie intra-membranaire des protéines serait composée d'un long segment de résidus apolaires. Un exemple est donné par protéine membranaire de *Halobacterium halobium*. La sous-unité de cette protéine consiste essentiellement en sept hélices  $\alpha$  qui traversent la membrane. Un autre exemple bien connu est la pompe à calcium dans le réticulum sarcoplasmique des muscles qui contient une membrane très spécialisée. En effet, sa seule fonction semble être le relargage et l'accumulation d'ions calcium. Les lipides constituent une barrière tandis que la protéine, sous forme de cylindre, permet le passage de ces ions.

#### – Interaction protéine-ligand

Les ligands sont des petites molécules qui peuvent se lier de manière réversible à une molécule et former un complexe avec elle. Il en existe de plusieurs types différents : substrats, cofacteurs, ions métalliques, activateurs, inhibiteurs, neurotransmetteurs, nucléotides. Ils peuvent se lier à des protéines comme des enzymes, ou des récepteurs tels les récepteurs nicotiques ou les récepteurs impliqués dans la transduction du signal. Les protéines sont des cibles thérapeutiques, car on peut synthétiser des ligands qui prennent la place des ligands naturels pour modifier les fonctions de la protéine cible. Par exemple, la fixation d'un ligand sur un site allostérique modifie le site actif de la protéine. C'est pourquoi les interactions protéine-ligand sont particulièrement intéressantes à étudier dans le cadre de l'amarrage ("*docking*") de ligands pour le développement de médicaments. De plus, les informations sur le mode d'amarrage d'un ligand permettent de comprendre structurellement son mode d'action et d'envisager son amélioration.

Les ligands interagissent avec les protéines à travers différents types d'interactions :

- les liaisons hydrogène : plus fortes que les van der Waals, moins fortes que ionique ou covalent,
- les interactions ioniques : elles ont lieu entre des groupes chimiques ioniques ou interactions électrostatiques charge-charge. Elles peuvent également former des "ponts salins", qui sont



des interactions entre des résidus positifs (Arg, Lys) et négatifs (Asp, Glu).

- les interactions apolaires : elles ont lieu entre groupes non chargés. Les interactions sont de type dipôle permanent–dipôle induit (induction, Debye) ou dipôle induit–dipôle induit (dispersion, London). Elles concernent aussi les interactions répulsives à courte distance qui sont responsables de la condition de complémentarité stérique).
- les autres : Cations- $\pi$ , complexation de métal, les interactions hydrophiles (interaction polaire favorable, contribution du solvant défavorable), interactions hydrophobes (interaction apolaire favorable, contribution du solvant favorable).

Les systèmes immunoglobines-ligand sont un modèle pour les interactions protéine-ligand [5].

## 1.2 Cibles de choix pour influencer sur les mécanismes biologiques

Les interactions protéine-protéine sont les unités fonctionnelles de la régulation des processus biologiques. C'est pourquoi elles sont reconnues comme étant une classe importante de cibles thérapeutiques [6]. Des études récentes sur le protéome humain suggèrent qu'il existerait environ 650 000 interactions dans l'interactome humain [7]. Beaucoup d'interactions protéine-protéine impliquées dans la signalisation cellulaire sont reliées à des maladies spécifiques, comme le cancer, les maladies immunitaires ou neurodégénératives, et les maladies infectieuses. De ce fait, les interactions protéine-protéine représentent une nouvelle voie de recherche de molécules bioactives agissant au niveau de leurs interfaces. Les applications de recherche d'inhibiteurs ciblant les interfaces entre les protéines sont en expansion.

Ainsi, pour traiter le cancer, ce sont surtout les interactions entre des molécules cruciales impliquées dans l'apoptose qui sont étudiées, afin de prévenir une croissance cellulaire anormale. Pour les pathologies du système immunitaire, par exemple les maladies autoimmunes, ce sont les cytokines et leur récepteurs qui sont étudiés afin de trouver des modulateurs de la prolifération des cellules T et des réponses immunitaires. Dans les maladies neurodégénératives, les interactions protéine-protéine de l'apoptose sont ciblées pour prévenir la mort des neurones [6]. Et pour les maladies infectieuses, les interactions entre les virus (ou les bactéries) et les protéines de l'hôte peuvent être également ciblées.

D'autre part, les interfaces protéine-protéine sont structurellement très diversifiées, ce qui peut aboutir à une régulation plus spécifique des protéines ciblées, permettant ainsi de diminuer les effets secondaires et d'augmenter l'efficacité thérapeutique. En plus des interfaces d'interaction, on peut également cibler les sites allostériques. De plus, cette méthode est également applicable aux interactions protéine-lipide ou protéine-acide nucléique. Ces approches peuvent conduire à de nouvelles opportunités pour la découverte de nouveaux traitements dans le futur.

## 2 Méthode d'étude des complexes

### 2.1 Méthodes de détermination des structures de complexes

La structure tridimensionnelle d'une protéine permet d'obtenir des informations sur la fonction biologique d'une protéine. Elle permet d'avoir accès à de nombreuses propriétés comme la forme du repliement de la protéine, la forme de la surface, les propriétés électrostatiques, l'agencement du site catalytique, la liaison protéine/ligand, la localisation de résidus mutés ou conservés, ou encore l'association de sous-unités.

Il existe de nombreuses méthodes de résolution de structure, à différentes échelles. Les techniques de microscopie optique par exemple, ne disposent pas de la résolution suffisante pour distinguer les détails des molécules au niveau atomique. La cryomicroscopie électronique permet de reconstituer la structure de complexes macromoléculaires à basse résolution, en réalisant au préalable une congélation "flash" de la protéine en solution, par plongement rapide dans l'azote liquide. Un cliché de microscopie électronique à transmission permet ensuite de visualiser un grand nombre d'images de la protéine d'intérêt dans diverses orientations, qui seront ensuite analysées par des outils informatiques. Parmi les méthodes à basse résolution, il y a aussi la diffusion des rayons X aux petits angles (DXPA) (dits en anglais "Small Angle X-Ray Scattering (SAXS)") ou la diffusion des neutrons aux petits angles ("Small Angle Neutron Scattering (SANS)") qui sont deux techniques permettant d'obtenir des informations sur la taille et la forme d'objets contenus dans une solution, comme par exemple l'état d'oligomérisation et la forme des macromolécules biologiques. Le principe est simple, une solution concentrée de la macromolécule est placée dans un faisceau parallèle de rayons X ou de neutrons. La protéine diffuse le rayonnement incident. L'intensité du signal diffusé en fonction de l'angle par rapport au faisceau incident est analysé sur toutes les orientations de la molécule, donnant ainsi des informations sur sa taille et sa forme. En fonction de l'échantillon, ces trois techniques peuvent atteindre une précision de 1 à 2 nm de résolution.

Cependant, pour comprendre la dynamique d'une structure protéique, et réaliser une analyse des mouvements moléculaires, une détermination de la structure 3D à l'échelle "atomique" est indispensable. Les données qui en résultent sont souvent utilisées dans les projets de développement de médicaments, car elles permettent de comprendre le mode d'action d'un médicament sur sa cible d'un point de vue moléculaire et structural. La prédiction des médicaments se fait par des méthodes expérimentales d'analyse structurale et des approches computationnelles à partir de la structure. Les deux techniques les plus utilisées sont la cristallographie aux rayons X et la résonance magnétique nucléaire, qui sont des méthodes physiques indirectes. Les coordonnées atomiques des structures tridimensionnelles ainsi résolues sont disponibles sur la *RCSB Protein Data Bank* (PDB) [8].

## La cristallographie aux rayons X

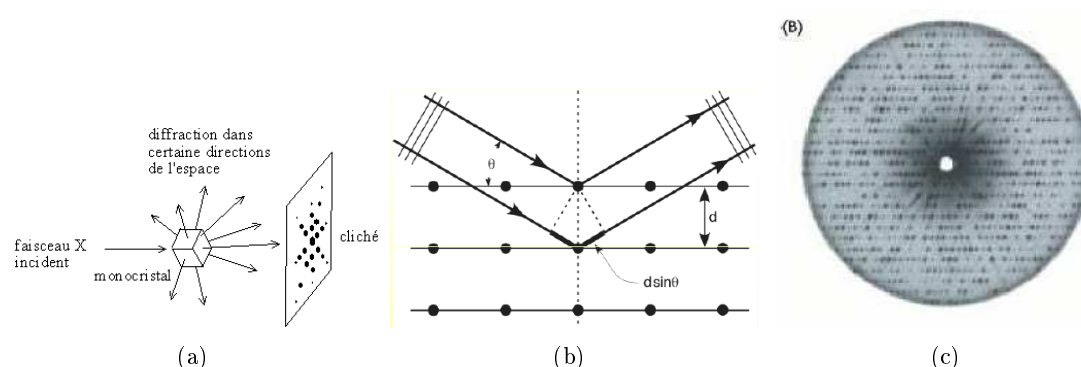
La cristallographie aux rayons X est la méthode la plus ancienne et toujours la plus utilisée [9].

L'application de la cristallographie à l'étude structurale des macromolécules, en particulier des protéines, doit sa naissance aux études de Max Perutz et John Kendrew qui ont obtenu conjointement en 1962 le prix Nobel de chimie pour leur travail sur la détermination de la structure tridimensionnelle de l'hémoglobine et de la myoglobine. Cette méthode est basée sur l'étude de la diffraction du rayonnement électromagnétique des rayons X par des cristaux de la molécule d'intérêt, car les longueurs d'onde (comprises entre 0,01 nm et 10 nm) sont de l'ordre des distances qui séparent les plans atomiques des réseaux cristallins. On obtient ainsi une carte de diffraction correspondant à une image virtuelle de la protéine, qui est ensuite traduite par un procédé mathématique en une image réelle.

La première étape consiste donc à obtenir des cristaux de la protéine d'intérêt, en ajoutant du sulfate d'ammonium ou un autre sel à la solution concentrée de la protéine en question, pour réduire sa solubilité. Une cristallisation lente favorise la formation d'un réseau cristallin hautement ordonné, à l'inverse de précipités amorphes. Cette étape est souvent limitante, car les conditions physico-chimiques (pH, tampon, ions...) favorables sont variables d'une molécule à l'autre, et parce que les méthodes de détermination des conditions sont très empiriques. De plus, le degré de pureté de la molécule est également un facteur déterminant pour la résolution de sa structure.

Lorsque que le cristal est obtenu avec une taille satisfaisante (quelques dizaines à quelques centaines de microns), la deuxième étape consiste à le "bombarder" de rayons X monochromatiques, produits par une source adaptée (par exemple un synchrotron). Les rayons sont diffractés par les électrons des molécules ordonnées dans le cristal, et sont détectés par un film photographique qui laisse alors apparaître des tâches noircies (figures I.1(a) et I.1(c)) en fonction de leur intensité, ou par un détecteur électronique. Les ondes issues de l'atome interfèrent, et le film photographique qui les reçoit ne laisse apparaître de tâches que si ces ondes sont en phase, car les ondes en opposition de phase s'annulent. L'amplitude de l'onde diffractée par un atome est proportionnelle au nombre d'électrons qu'il contient. Ainsi, un atome de carbone diffracte donc six fois plus "intensément" qu'un atome d'hydrogène. Le fait que les molécules à étudier forment un cristal permet d'amplifier le signal de diffraction, car les rayons ont plus de chance d'être en phase, contrairement à des molécules en solution où les signaux de diffraction se mélangent. Une illustration d'une carte de diffraction est donnée sur la figure I.1(c). Des images du motif de diffraction sont enregistrées pour différentes orientations du cristal. L'espacement des tâches renseigne sur la maille du cristal, c'est-à-dire les dimensions de l'élément de volume qui est régulièrement répété dans l'espace pour donner naissance au cristal.

L'étape suivante est de reconstruire une carte de la densité électronique en 3D de la protéine à partir des intensités observées, grâce à la transformée de Fourier. Pour chaque tâche, cette opération mathématique produit une onde de densité électronique, dont l'amplitude est proportionnelle à la racine carrée de l'intensité observée de la tâche. D'autre part, la phase de chaque



**FIG. 1.1: Diffraction aux rayons X** (a) : (b) : Les rayons X sont diffractés selon la loi de Bragg (Prix Nobel 1915) :  $\lambda = 2d_{h,k,l} \sin\theta$  (c) : Diagramme de diffraction de la myoglobine. Photo extraite de Stryer, 5ème édition.

onde détermine si elle renforce d'autres ondes ou si elle annule les ondes contribuées par d'autres tâches. Ces phases peuvent être déduites de modèles de diffraction très bien connus et produits par des marqueurs atomes lourds, comme l'uranium ou le mercure sur des sites spécifiques de la protéine. La carte de densité électronique est représentée par des couches parallèles empilées les unes sur les autres. Sur chaque couche est représentée la distribution de la densité électronique par des lignes de contour, comme les lignes des cartes topographiques qui représentent l'altitude. La résolution de la carte de densité est déterminée par le nombre d'intensités diffusées utilisées dans la synthèse de Fourier, mais dépend beaucoup de la qualité du cristal. Plus la valeur de la résolution est petite, et meilleure elle est.

La dernière étape consiste à interpréter la carte de densité électronique, en positionnant au mieux les atomes de la protéine dans celle-ci, avec l'aide de la séquence en amino-acides de la protéine, et de modèles de carte de densité spécifique à chaque acide aminé. La structure peut ensuite être raffinée par différents cycles d'analyses, de minimisation d'énergie (calcul d'énergie) et de reconstruction pour arriver à un modèle plausible.

La cristallographie aux rayons X présente le grand avantage de ne pas connaître de limite liée à la taille des objets moléculaires étudiés. Par contre, pour pouvoir utiliser cette technique, il faut absolument disposer de cristaux en quantité et qualité suffisante, ce qui constitue en général l'étape limitante des études par cristallographie aux rayons X.

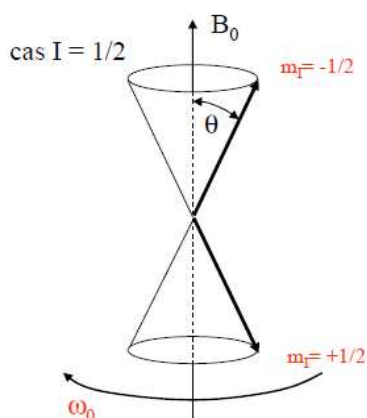
## La Résonance Magnétique Nucléaire (RMN)

Le phénomène de résonance magnétique nucléaire a été découvert en 1938 par Isidor Isaac Rabi, qui reçu à cette occasion un prix Nobel de physique en 1944. La spectrométrie RMN est une méthode qui utilise les propriétés magnétiques des noyaux des atomes pour obtenir des informations géométriques (distances, angles...) sur la molécule. A partir de ces informations géométriques, il est ensuite possible de reconstruire une structure 3D.

Les noyaux atomiques sont caractérisés par un spin nucléaire  $I$  et un moment magnétique

nucléaire  $\mu$  qui sont tous deux quantifiés et équivalents à de petits aimants :  $\mu = \gamma I$ , avec  $\gamma$  le rapport gyromagnétique caractéristique de chaque isotope. Le nombre de spin  $I$  détermine le nombre  $N$  d'orientations (de positions) que peut prendre le noyau au sein d'un champ magnétique uniforme. Les valeurs de  $I$  peuvent être prévues empiriquement à partir du nombre  $p$  de protons et du nombre  $n$  de neutrons constituant le noyau. Si  $p$  et  $n$  sont pairs, alors  $I=0$  (exemple :  $^{12}\text{C}$ ,  $^{16}\text{O}$ ) ; si  $(p+n)$  pairs et  $p$  impair, alors  $I$  est un entier (exemple :  $^2\text{H}$ ,  $^{14}\text{N}$ ) pour lesquels  $I=1$  ; si  $(p+n)$  est impair, alors  $I=1/2$  entier (exemple : les isotopes  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{31}\text{P}$ ).

En l'absence de champ magnétique extérieur, tous les noyaux, quel que soit leur moment magnétique, ont la même énergie. En revanche, si on applique un champ magnétique extérieur statique  $\vec{B}_0$  très intense (typiquement 10 à 20 tesla), ces spins s'alignent avec le champ et définissent  $(2I + 1)$  niveaux d'énergie, (soit  $(2I + 1)$  orientations possibles), équidistants de  $\omega_0 = \gamma B_0$  (figure I.2), avec  $\omega_0$  la fréquence angulaire de Larmor des rotations des spins autour de l'axe de  $\vec{B}_0$ .

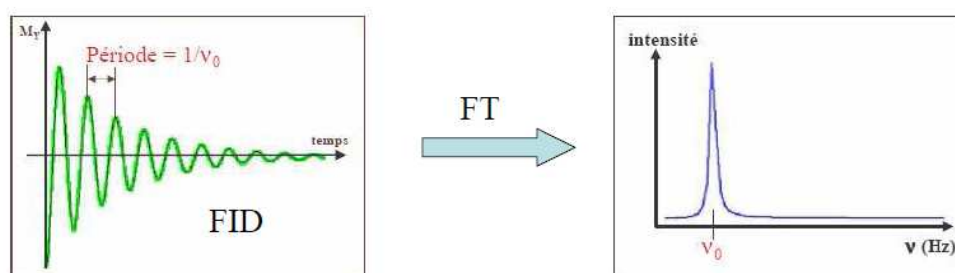


**FIG. I.2:** Rotation d'un spin autour du champ magnétique  $\vec{B}_0$ . Le paramètre  $m_I$  représente les deux états du spin  $I$  dans le cas  $I = 1/2$ .  $\theta$  l'angle décrit par le spin et le vecteur du champ magnétique  $B_0$  autour duquel il tourne.  $\omega_0$  la fréquence angulaire de Larmor.

De ce fait, si  $I=0$ , il n'y a pas d'interaction possible entre les spins, et les noyaux correspondants ne peuvent pas être utilisés en RMN. Si  $I = 1/2$ , deux états de spin sont alors possibles, donc deux moments magnétiques. C'est pourquoi en spectroscopie RMN on utilise principalement des isotopes de spin  $1/2$ .

Lorsque l'énergie d'un rayonnement électromagnétique correspond à la différence d'énergie entre deux niveaux d'énergie, celui-ci peut être absorbé par le noyau atomique. Le retour à l'équilibre est appelé relaxation magnétique longitudinale. L'énergie  $E$  d'une onde électromagnétique est proportionnelle à sa fréquence  $\nu$  :  $E = h\nu$  où  $h$  est la constante de Planck. La résonance RMN est donc caractérisée par cette fréquence exprimée en Hertz, appelée fréquence de Larmor (ou fréquence de résonance) et notée  $\nu_0$ , qui ne dépend que du champ appliqué  $B_0$  et de la constante gyromagnétique  $\gamma$  :  $\nu_0 = \gamma B_0$ .

En spectroscopie RMN, on place l'échantillon dans un champ  $B_0$  constant dans l'espace, et stationnaire, de façon à aligner les moments magnétiques de spin dans ce champs. Puis, on applique un rayonnement magnétique radiofréquence notée  $B_1$ , orienté perpendiculairement au plan de  $B_0$ , avec une fréquence angulaire  $\omega$  égale à la résonance de Larmor  $\omega_0$  pour perturber le spin. Le signal mesuré durant la relaxation du système (qui décroît exponentiellement) est appelé un FID (*Free Induction Decay*), qui se présente sous forme d'une sinusoïde, ou de superpositions de sinusoïdes, selon le nombre de résonances présentes dans le spectre RMN. Une transformée de Fourier de ce signal permet la visualisation d'un spectre RMN en fréquence. On obtient ainsi un pic par proton (figure I.3).



**FIG. I.3: Mesure du signal RMN** A gauche : Mesure du signal FID pendant la relaxation. A droite : Après avoir appliqué la transformée de Fourier.

En réalité, on en observe plusieurs. En effet, la même année, Proctor et Dickinson découvrent indépendamment le phénomène de déplacement chimique, découverte fondamentale pour l'essor des applications de la RMN en chimie organique. Quand une molécule est placée dans le champ magnétique  $B_0$ , des courants d'électrons sont induits dans les diverses orbitales moléculaires, induisant ainsi des moments magnétiques qui s'opposent proportionnellement à l'action du champ magnétique  $B_0$ . Le champ réellement ressenti par les noyaux n'est donc pas exactement le champ appliqué, mais un champ réduit. Cet effet est nommé le diamagnétisme. D'autre part, d'autres effets peuvent modifier le champ local, comme par exemple le paramagnétisme électronique, lorsque les électrons sont délocalisés (noyaux aromatiques, liaisons chimiques fortement polarisées,...) ou non appariés, ce qui va à l'inverse augmenter le champ ressenti par les noyaux.

L'environnement électronique de chaque noyau est alors caractérisé par une constante d'écran  $\sigma$  qui contient des contributions des deux effets dia- et paramagnétiques. Le champ local ressenti correspond alors à  $B_{Local} = B_0(1 - \sigma)$ . La fréquence de résonance de chaque noyau dépend donc directement de la constante d'écran :  $\nu_L = \nu_0(1 - \sigma)$ . La différence de fréquence  $\delta$  induite par les différents environnements électroniques est appelée "déplacement chimique" et correspond à la relation suivante :

$$\delta = ((\nu_L - \nu_{ref})/\nu_0) * 10^6 \quad (I.1)$$

la référence choisie dépendant de l'isotope. Les gammes de déplacement chimiques sont très diverses suivant les isotopes étudiés. Le déplacement chimique donne une idée du groupement chimique. Il est invariable avec le champ magnétique. Il existe une table des déplacements chimiques des protons dans les protéines pour faciliter l'attribution des spectres.

La Résonance Magnétique Nucléaire (RMN) permet l'étude d'échantillons en solution, ce qui donne accès à l'analyse de phénomènes comme les échanges conformationnels ou les interactions inter-moléculaires de faible affinité. Mais il faut noter une limitation de la RMN : la taille des objets moléculaires étudiés. En effet, à partir de 150-200 résidus, l'étude par RMN devient très complexe et nécessite des protocoles très lourds pour la préparation des échantillons.

### Couplage dipolaire résiduel (CDR)

En spectroscopie RMN, la résonance des spins peut également être modifiée par des interactions spin-spin qui se divisent en deux catégories : les interactions dipolaires magnétiques, qui peuvent être directes à travers l'espace, on appelle alors ce phénomène "couplages dipolaires directs", ou les interactions créées par l'intermédiaire des électrons situés dans les orbitales moléculaires de liaison, on parle alors de "couplages dipolaires indirects ou scalaires". Le couplage scalaire est produit par les effets physiques suivants. L'interaction de contact de Fermi polarise le moment magnétique des électrons dans le sens opposé au moment magnétique du noyau. Dans une liaison chimique, les spins sont couplés grâce aux électrons de liaison qui transmettent cette polarisation de proche en proche. Ce couplage est dit indirect car il résiste aux réorientations moléculaires. Il est indépendant du champ magnétique  $B_0$  et de la fréquence de résonance du noyau. Il représente un excellent indicateur de la présence de liaisons chimiques. Prenons par exemple deux noyaux  $k$  et  $j$ . Selon l'orientation de son moment magnétique, le noyau  $k$  va induire un petit champ qui va s'ajouter ou se soustraire au champ appliqué, ce qui va augmenter ou diminuer le champ nécessaire pour atteindre la résonance de  $j$ . Il est également appelé "couplage J" en référence à la variable  $J_{ij}$  qui en définit l'intensité.

L'intensité du couplage direct dépend de la distance  $r_{ij}$  entre deux noyaux  $i$  et  $j$  en interaction et de leurs rapports gyromagnétiques  $\gamma_i$  et  $\gamma_j$ . Il dépend de l'angle formé entre le vecteur inter-nucléaire  $\vec{r}_{ij}$  et le champ appliqué  $B_0$ . Ce type de couplage est moyenné dans les liquides où les molécules prennent toutes les orientations possibles, et n'est donc pas mesurable. Mais en milieu anisotrope ou orienté, on peut mesurer le couplage dipolaire résiduel. La protéine ou le complexe protéique étudié est solubilisé dans un milieu qui orientent les objets moléculaires individuels tous dans la même direction. Les constantes de couplage dipolaire résiduelle peuvent alors être mesurées, et donnent des informations sur l'orientation des liaisons N-H par rapport au champ magnétique  $B_0$ . Elles fournissent aussi des contraintes sur l'orientation relative de partenaires au sein d'un complexe.

## 2.2 Mouvements induits sur les structures de protéines par la formation d'un complexe

La plupart des protéines nécessitent un réarrangement conformationnel afin de pouvoir interagir avec leur ligands ou leur substrat. Ce réarrangement se fait par la formation d'un complexe avec une autre protéine. C'est le cas par exemple des complexes enzymatiques, où la liaison d'un activateur permet à l'enzyme d'acquérir une forte spécificité d'interaction avec son substrat. Les kinases par exemple, qui utilisent l'ATP pour phosphoryler les protéines sur leurs résidus sérine/thréonine ou tyrosine situés sur une boucle dite "d'activation", comme c'est le cas pour la protéine kinase DYR (DYRK) [10], sont activées par leur propre phosphorylation, ou par l'interaction avec d'autres protéines ou ligands.

Par exemple, la liaison du peptide Ste5 sur la kinase Fus3 modulerait allostériquement son

activité en coordonnant une orientation optimale entre les lobes N-terminal et C-terminal de la kinase [10], distaux du site actif. L'effet et le mécanisme d'activation sont comparables à l'effet de l'extension C-terminale des protéine kinases A (PKA), qui stabilise leur état actif.

Les protéines kinases R (PKR) doivent former un dimère pour être activé [10]. L'un des deux domaines kinase met en interaction une région comportant l'hélice C dans le lobe N de l'autre domaine kinase. Elles possèdent également un couplage allostérique entre la liaison du substrat sur une autre hélice et la conformation de la boucle d'activation.

Les protéines kinases dépendantes des cyclines (CDKs) sont importantes dans le contrôle du cycle cellulaire. Elles sont activées et inactivées à différents moments du cycle [11]. Les CDKs sont partiellement activées par leurs sous-unités de régulation cycline, qui sont synthétisées périodiquement et dégradées au cours du cycle cellulaire. Le complexe cycline/CDK est entièrement activé par la phosphorylation d'un résidu thréonine dans la boucle T par la kinase activatrice des CDK (*CDK-activating kinase CAK*). Les CDKs sont régulées par des inhibiteurs (CDKIs) qui interrompent le cycle cellulaire, en réponse à divers signaux. L'un d'entre eux,  $p19^{INK4d}$  est spécifique des Cdk4 et Cdk6, et les active en présence ou en absence des cyclines D dont elles sont dépendantes. La structure de Cdk6 est composée de deux domaines, séparés par le site catalytique, et est très similaire à celle de Cdk2. La cycline interagit avec une hélice du domaine N-terminal, dont la conformation change pour activer la kinase.  $p19^{INK4d}$  se lie au niveau de l'interface des deux domaines, au niveau d'un seul côté du site catalytique. Des interactions proches entre  $p19^{INK4d}$  et le "linker" de Cdk6, qui lie les deux domaines, semblent être déterminantes pour la spécificité. La liaison de  $p19^{INK4d}$  induit un mouvement considérable du domaine N-terminal de la kinase, relativement au domaine C-terminal, inhibant ainsi la liaison productive avec l'ATP, et les mouvements de l'hélice (qui interagit avec la cycline) vers sa conformation active. Il est possible que les réarrangements conformationnels permettent un mouvement de cette hélice vers sa conformation active avec la liaison de la cycline, mais la liaison productive de l'ATP serait peu probable.

Dans les kinases dépendantes de la calmoduline, la région de régulation bloque le site actif. La liaison de la CaM relève cette auto-inhibition, révélant ainsi un domaine kinase actif. Ainsi la phosphorylation de la boucle d'activation n'est pas nécessaire. Par exemple, la liaison de la CaM met en évidence un résidu Thr majeur pour l'auto-phosphorylation de CaMK-II, qui se situe au niveau de la région de régulation.

Les adénylates kinases sont des kinases d'AMP qui catalysent la réaction  $Mg^{2+} ATP + AMP \rightleftharpoons Mg^{2+} + ADP + ADP$  [12]. Elles maintiennent le niveau des nucléotides à l'équilibre, facilitant ainsi l'utilisation de l'ATP et leur reconstitution. L'enzyme doit être spécifique, dans le but d'éviter des réactions de transfert du radical phosphoryl involontaires, dont l'hydrolyse. Cette spécificité est procurée par deux domaines qui se referment fermement avec la liaison des substrats. Cependant, la forte liaison du substrat conduit l'enzyme dans un puits énergétique qui ralentit la catalyse. Comme le transfert des radicaux phosphoryl entre les phosphates est énergétiquement neutre, il n'y a donc pas de différence énergétique au cours de la réaction qui permettrait à l'enzyme de sortir de ce puits. Müller [12] suggère donc un modèle selon lequel l'énergie serait contrebalancée par la



diminution de l'énergie de liaison interne, qui a lieu par la mobilisation d'une chaîne polypeptidique dans une région isolée du centre actif.

La maturation des pré-ARNm eucaryotes en ARNm matures fait intervenir des ribonucléoprotéines [13]. Dans un premier temps, la reconnaissance intermoléculaire spécifique entre ces protéines et le pré-ARNm requiert l'interaction de boucles variables du domaine ribonucléoprotéique avec les régions hélicoïdales de l'ARN. Puis, dans un deuxième temps, la formation du complexe ordonne la boucle simple-brin flexible de l'ARN contre la surface en feuillet  $\beta$  de la protéine, et réorganise la région carboxy-terminale de celle-ci afin de maximiser la complémentarité de la surface et la reconnaissance du groupe fonctionnel. Le rôle de la structure de l'ARN et la plasticité illustré par ce mécanisme explique comment une forte liaison ( $K_d \approx 10^{-11}$  M) et une spécificité importante peuvent être acquises avec une surface intermoléculaire relativement petite ( $2,3 \text{ \AA}^2$ ). Les contacts intermoléculaires se distinguent en deux catégories. La première implique une région de potentiel électrostatique positif. Ces résidus participent dans l'interaction de type "corps rigide" entre la protéine et les nucléotides de l'ARN, pour établir la formation du complexe. La deuxième catégorie de contacts se produit par ajustement induit "*induced-fit*", qui est un modèle d'interaction dans lequel la complémentarité entre les deux partenaires est imparfaite (voir section 3.2) et dont la liaison induit une réorganisation conformationnelle de l'un des deux. Cette catégorie implique les résidus à la surface du feuillet  $\beta$  et de l'hélice  $\beta 4$  boucle C, dont les contacts ne peuvent avoir lieu sans la réorganisation de l'hélice C.

## 3 Importance du désordre en biologie

### 3.1 Importance fonctionnelle du désordre

Les protéines assurent l'échafaudage de la vie, et assurent une grande variété des réactions catalytiques essentielles. De ce fait, il a longtemps été supposé qu'elles doivent être structurellement stables et repliées correctement pour assurer leur fonction. Cependant, les progrès réalisés dans le séquençage des génomes a permis de réfuter cette hypothèse. En effet, le génome humain complet, publié en 2003 [14; 15] est significativement trop petit par rapport au code nécessaire pour exprimer une telle variété de protéines. De plus, une proportion significative des séquences protéiques qui seraient codées par le génome ne formeraient pas de structure tridimensionnelle stable. C'est pourquoi le génome doit inclure un code pour une multitude de formes alternatives de protéines, dont la plupart pourrait être partiellement ou complètement désordonnées (destructurées) ou pas totalement structurées dans leurs états fonctionnels [16]. D'autre part, en parallèle mais indépendamment, des études expérimentales ont permis de découvrir des exemples de protéines et domaines importants qui n'étaient pas totalement structurés ou qui étaient complètement désordonnés en solution, mais qui demeuraient parfaitement fonctionnels. Les raisons pour lesquelles un domaine est désordonné varient en fonction des systèmes dans lequel ce genre de domaine se trouve. Certaines protéines sont déstructurées en solution en l'absence de

leur partenaire physiologique, mais se structureraient en leur présence, formant ainsi un complexe stable pouvant être purifié et caractérisé [17; 18; 19].

Le concept de “protéines intrinsèquement désordonnées” [20] a émergé dans plusieurs domaines indépendamment et simultanément. La terminologie dans chaque domaine est donc un peu différente. Ainsi, les protéines intrinsèquement désordonnées sont également désignées sous le nom de protéines destructurées, nativement non repliées, nativement désordonnées, ou encore “rhéomorphiques“. Les protéines désordonnées sont souvent des protéines régulatrices de la transcription, traduction, du cycle cellulaire, de la transduction du signal, de la phosphorylation d’autres protéines, et tous les processus impliqués dans le contrôle physiologique. Certaines d’entre elles sont impliquées dans des maladies génétiques, ou dans des maladies non héréditaires. Dans le cas de maladies génétiques, des mutations dans les régions structurées des protéines vitales peuvent être létales, alors que des mutations dans les régions désordonnées sont moins conséquentes.

Diverses raisons possibles peuvent expliquer l’importance du désordre intrinsèque d’une protéine (ou d’une partie de protéine). Par exemple, les interactions entre un nombre de facteurs de transcription et l’ADN constituent une classe d’interactions où des régions destructurées contribuent à l’augmentation de l’affinité séquence-spécifique entre les partenaires [21]. Ce phénomène a également été récemment observé dans les interactions ARN-protéines [22]. L’augmentation de la spécificité est un phénomène important pour les complexes qui doivent se former suite à une transduction de signal, ils doivent également être capables de se dissocier une fois leur fonction accomplie. D’autre part, beaucoup de fonctions centrales de la cellule, comme la transcription de gènes, nécessitent l’assemblage de complexes multi-protéiques. L’un des rôles majeurs d’un coactivateur comme le CBP est de connecter plusieurs membres variés du complexe multi-protéique, en interagissant avec de nombreuses protéines différentes à la fois, à différents endroits de l’ADN. Pour permettre cette connexion multi-protéique, le CBP possède donc un nombre de séquences de “linker” assez longs, pour permettre une architecture différente de ces sites. Certaines protéines désordonnées jouent un rôle crucial dans plusieurs maladies, notamment le cancer, qui ouvrirait une nouvelle vision de concevoir des médicaments [23].

## 3.2 Rôle de la dynamique dans la formation de complexes

La vision classique des protéines a longtemps été basée exclusivement sur les protéines globulaires, et leur nature dynamique était largement ignorée. Cette vision était surtout due au fait que la cristallographie aux rayons X était la seule méthode disponible à l’époque (ou majoritairement utilisée) pour résoudre les structures tridimensionnelles des protéines. En effet, cette approche ne permettait de résoudre que des conformations rigides et uniques, ce qui laissait penser que chaque protéine adopte une conformation unique et permanente dans des conditions physiolo-

giques. Cependant, les avancées technologiques dès 1980 ont permis l'émergence d'une vision plus dynamique aux protéines et leurs interactions. Le développement de la RMN et d'autres méthodes spectroscopiques ont permis de détecter des conformations autres que les conformations fortement peuplées, ce qui a conduit à la découverte de la vraie diversité des protéines en solution. L'évolution en matière de techniques génétiques associés à de nouveaux projets génomiques, a également contribué à la conclusion que de nombreuses protéines et régions protéiques ne peuvent former de structure stable lorsqu'elles sont isolées. Ce groupe de protéines, appelées Protéines Intrinsèquement Désordonnées (PID), ne peut être caractérisé par des structures uniques bien définies, mais existe dans un ensemble de conformations rapidement interconvertibles.

### Les trois modèles d'interactions protéine-protéine

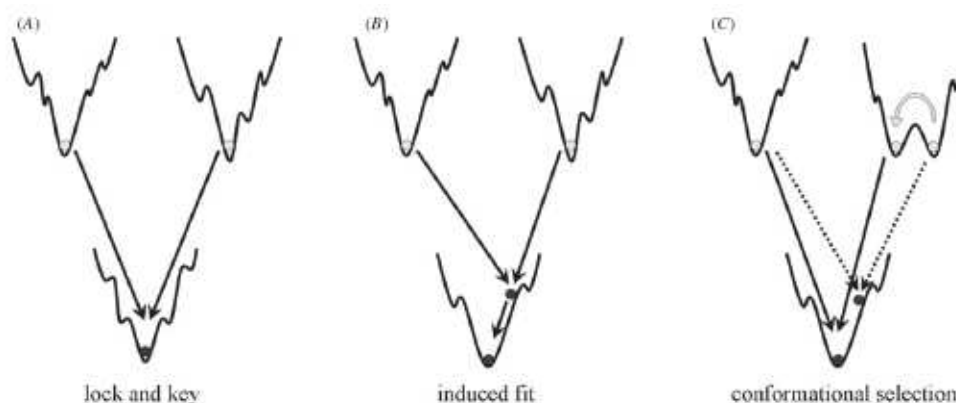
A l'origine, des modèles variés ont été développés pour décrire les interactions enzyme-substrat. Désormais, ils sont également applicables de manière plus générale à des interactions protéine-protéine génériques. Il existe trois modèles principaux expliquant l'interaction entre deux protéines à l'état libre conduisant à la formation d'un complexe. Le premier modèle, le plus classique, est appelé le modèle "*clé serrure*" (Fischer, 1890). Il supporte l'idée que les deux partenaires doivent être complémentaires chimiquement et géométriquement au niveau de leurs interfaces d'interaction, sans induire quelconque changement conformationnel dans la forme libre et la forme liée des partenaires. Il s'applique surtout aux protéines globulaires. L'exemple de la trypsine et BPTI illustre bien ce modèle, car la structure des deux partenaires à l'état libre est bien définie et reste inchangée à l'état lié. Cependant, beaucoup de protéines ne conservent pas la même conformation dans les deux états. Le deuxième modèle, nommé le mécanisme par "ajustement induit" (dit en anglais "*induced-fit*") (Koshland, 1958 [24]), suggère une complémentarité imparfaite entre les deux partenaires. Et donc, lors de la liaison, la structure de l'un (ou des deux) change. Néanmoins, ces deux modèles supposent une conformation stable.

Le troisième modèle propose une autre explication en ce qui concerne les différences conformationnelles entre les états non-liés et liés. Le concept de "*sélection conformationnelle*" dit que l'un des deux, ou même les deux partenaires adoptent de multiples conformations à basse énergie à l'état non-lié. La différence entre les énergies libres de ces états détermine la distribution de la population de ces conformations. L'interaction entre les partenaires perturbe cet équilibre. La conformation du complexe adoptée ayant l'énergie la plus basse est différente de la structure dominante à l'état non lié, et correspond à l'une des conformations alternatives d'énergie plus haute.

Pour illustrer ces mécanismes, on peut représenter l'espace conformationnel par le paysage énergétique des protéines, qui se traduit par une représentation où l'énergie de chacune d'entre elles est une fonction des degrés de liberté, comme par exemple les angles dièdres le long de la chaîne carbonée.

### Les profils énergétiques de ces trois modèles

Les profils énergétiques des trois modèles cités précédemment sont représentés sur la figure I.4.



**FIG. I.4: Représentation schématique des profils énergétiques des trois modèles d'interaction entre les protéines globulaires.** (A) Le modèle “clé-serrure”, (B) le modèle “ajustement induit”, (C) le modèle “sélection conformationnelle”. L'énergie du système est dessinée selon une coordonnée unique de l'espace conformationnel. Les états initiaux et finaux des protéines sont représentés par des points blancs et noirs respectivement. Les flèches pleines décrivent les chemins de liaison, tandis que les flèches en pointillés montrent les chemins de liaison avec des énergies défavorables. Cette figure est extraite de Meszaros, 2011 [25].

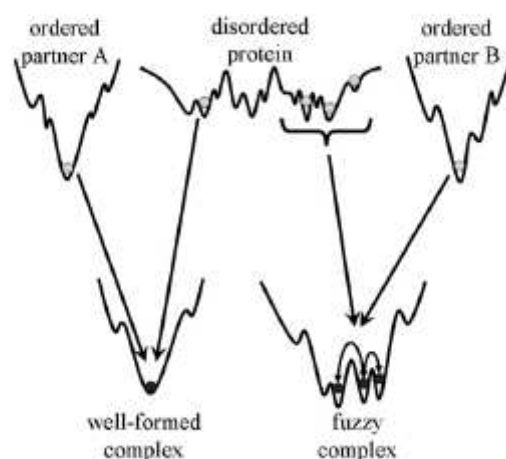
Le paysage énergétique des protéines globulaires bien repliées est comparé à un entonnoir, où il y a un grand nombre de conformations ayant une énergie haute (donc défavorable), et beaucoup moins de conformations à basse énergie. Le point le plus bas (le minimum global) représente l'unique structure native. Dans le modèle “clé-serrure”, les deux partenaires ont un minimum global bien défini à l'état non-lié, chacun ayant sa propre fonction énergétique correspondant à la conformation native. L'interaction entre ces deux partenaires conduit à une fonction d'énergie combinée qui a elle-même son propre minimum global unique où chaque partenaire conserve la conformation qu'il avait à l'état libre. Le modèle “induced-fit” est similaire, à la différence que le complexe formé ne correspond pas à un minimum énergétique. Le complexe va, à partir de cet état, atteindre le minimum global en réarrangeant la conformation de l'un ou des deux partenaires. Dans le concept de la “Sélection conformationnelle”, au moins un des partenaires a deux, voir plusieurs minima très prononcés séparés par une barrière énergétique. Lorsqu'ils interagissent ensemble, la conformation correspondant à l'un de ces minima est sélectionnée par le partenaire.

L'importance de ce dernier modèle repose sur le fait qu'il peut prendre en compte l'hétérogénéité des protéines. Grâce au développement des expériences RMN de relaxation dispersion [26], des conformations de haute énergie peu peuplées ont pu être détectées, ce qui a mis en évidence l'importance de ces conformations lors de la liaison. La généralisation de ce modèle peut être appliquée aux protéines intrinsèquement désordonnées. En effet, les PIDs décrivent un profil énergétique bien différent des protéines globulaires. Tout d'abord, la différence entre l'état complètement déstructuré des PIDs et l'énergie globale minimum est bien moins grande que pour les protéines bien repliées. De même, la différence entre le minimum local et le minimum global

est petite. Et pour finir, ces minima locaux ne sont pas séparés par de grandes barrières énergétiques. Par conséquent, les états proches équivalents peuvent être facilement atteints. Les PIDs peuvent contenir des éléments structuraux préformés, comme les hélices  $\alpha$ . Ces éléments peuvent être visualisés par peu de vallées plus profondes, mais l'apparence générale des profils énergétique ne change pas. Les PIDs forment ainsi une classe assez hétérogène et adoptent un équilibre dynamique avec un continuum de conformations interchangeables rapidement.

Les segments ordonnés des PIDs peuvent également adopter une structure bien définie unique lorsqu'elles forment un complexe. Le profil énergétique du complexe qu'elles forment peut donc aussi ressembler à un entonnoir avec un minimum global unique. Cependant, ces protéines échantillonnent un grand nombre de conformations différentes à l'état non-lié précédant l'interaction. Donc, la conformation finale est choisie à partir d'un ensemble de conformations au lieu d'un nombre limité de conformations (figure I.5).

**FIG. I.5:** *Représentation schématique des profils énergétiques des interactions entre une protéine globulaire et une protéine intrinsèquement désordonnée (PID).* L'énergie du système est dessinée contre une coordonnée unique de l'espace conformationnel. Les états initiaux et finaux des protéines sont représentés par des points blancs et noirs respectivement. La protéine globulaire décrit un profil énergétique en forme d'entonnoir tandis que la PID décrit un profil plus plat et rude. Le complexe formé peut soit devenir complètement ordonné et représenté comme un entonnoir, soit retenir une certaine flexibilité conduisant à un complexe dit "flou" en anglais "fuzzy". Cette figure est extraite de Meszaros (2011) [25].



Les segments désordonnés ne deviennent pas toujours complètement ordonnés lors de la liaison, mais peuvent garder leur nature dynamique même dans une partie du complexe. C'est ce qu'on appelle les complexes "flous" dit "fuzzy" en anglais qui sont des complexes en partie désordonnés (figure I.5). La conformation adoptée dans le complexe peut refléter la préférence conformationnelle à l'état non-lié, mais elle peut aussi être influencée par le partenaire. En général, la flexibilité intrinsèque des segments désordonnés leur permet de lier plusieurs partenaires, et les différents niveaux de flexibilité présents dans ces molécules et leurs complexes leur permet d'assurer leurs fonctions dans différents cas.

### Les avantages

#### – thermodynamiques :

Il a été décrit que le désordre intrinsèque du partenaire permet de découpler la spécificité de l'interaction de son affinité. Cette propriété confère un avantage aux interactions des PIDs car les interactions hautement spécifiques d'affinité modérée sont souvent nécessaires dans les mécanismes de régulation et de transduction de signaux, où les interactions doivent pouvoir être rompues rapidement et efficacement [20; 27].

D'autre part, en conséquence des interconversions rapides d'une multitude de conformations, les PIDs ne présentent pas de charges discrètes pour leur partenaire, mais plutôt une charge électrostatique moyenne qui reflète la charge nette du segment désordonné [28]. Les régions désordonnées sont des sites de phosphorylation prédominants [29], pouvant être reliés à la régulation des interactions à travers l'effet "polyélectrostatique" [28]. D'autres modifications post-traductionnelles qui modifient la charge nette d'une protéine et qui peuvent aussi avoir des effets polyélectrostatiques sont l'acétylation des résidus lysine et la modification avec l'ubiquitine ou la protéine SUMO. Grâce à cet effet électrostatique à longue portée, une protéine désordonnée contenant de multiples motifs de liaison peut interagir avec un récepteur structuré ne possédant qu'un seul site de liaison [28]. La flexibilité des régions désordonnées joue aussi un rôle très important pour assurer la fonction biologique. En effet, la flexibilité de ces régions leur permet de faire des réarrangements conformationnels favorables, de manière à privilégier la distribution des charges selon le type d'interaction formée avec le ligand. De plus, le désordre augmentant la plasticité et la malléabilité des protéines, une même séquence protéique peut interagir avec différents partenaires.

– **stériques :**

Le désordre conformationnel permet la liaison entre une PID avec une protéine repliée qui est difficilement accessible d'un point de vue stérique. Les PIDs peuvent se lier sur plusieurs surfaces d'un partenaire, l'enveloppant et augmentant ainsi la surface enfouie [30]. Les régions ordonnées peuvent toujours interagir avec les "linkers" désordonnés, facilitant ainsi un certain nombre de possibilités pour la régulation.

– **par rapport à l'évolution :**

Les eucaryotes représentent le domaine possédant le taux de PIDs le plus élevé. Or des études précédentes ont montré que les régions désordonnées sont celles qui présentent les taux de mutations les plus élevés, sans que cela n'affecte ni la stabilité ni la fonction des PIDs. Seule l'accumulation de quelques mutations engendre ces conséquences. Ces deux informations combinées ensemble expliqueraient comment les eucaryotes auraient pu évoluer rapidement pour devenir des organismes pluricellulaires complexes. De plus, la taille de la molécule augmentant avec la complexité, il faudrait que les protéines structurées soient considérablement assez larges pour interagir. Le désordre, lui, aide à limiter la taille des protéines, ainsi que l'encombrement cellulaire [31].

– **par rapport à l'activité catalytique :**

Les fluctuations atomiques avec un intervalle de temps rapide (ps-ns) peuvent faciliter les mouvements plus lents ( $\mu$ s-ms) qui ont une amplitude plus élevée et qui sont catalytiquement plus productifs [32]. La fréquence de ces mouvements concertés s'avère donc être limitante pour la réaction catalytique.



## Chapitre II

# Les toxines : un modèle d'interaction protéine-protéine

---

## 1 Le rôle des toxines dans l'infection par les bactéries pathogènes

### 1.1 Panorama des toxines de pathogènes

Les toxines bactériennes sont des substances solubles qui altèrent le métabolisme normal des cellules de l'hôte. La production d'une toxine est spécifique d'une espèce bactérienne qui produit une pathologie qui lui est associée. Les **exotoxines** sont des protéines produites par les bactéries et sécrétées dans le milieu environnant. Les **endotoxines** sont des lipopolysaccharides de la membrane externe de certaines bactéries Gram négatif, libérées lors de la lyse cellulaire [33].

Les toxines ne sont pas indispensables à la croissance des bactéries, mais peuvent participer à leur survie en détruisant les leucocytes par exemple, et à leur diffusion, en dégradant les protéines de la matrice du tissu conjonctif.

Elles représentent les substances biologiques les plus puissantes contre l'être humain. En effet, elles agissent à des concentrations extrêmement faibles ( $\sim 100$  ng / personne pour la toxine tétanique ou botulinique).

Les toxines sont de spécificité variable, et agissent de plusieurs manières, sans forcément détruire directement les cellules de l'hôte [33]. Certaines agissent à la surface des cellules, d'autres au niveau intracellulaire.

#### 1. Les toxines agissant depuis le milieu extracellulaire de la cellule cible.

- les toxines qui aident les bactéries à diffuser dans les tissus :



Ce sont des enzymes de dégradation sécrétées par les bactéries, qui agissent comme facteurs de diffusion en facilitant la dispersion des micro-organismes infectants. Il y a par exemple l'hyaluronidase qui détruit la base du tissu conjonctif, la DNase qui fluidifie le pus rendu visqueux par l'ADN libéré des leucocytes morts ou encore la streptokinase qui détruit les barrières de fibrine se trouvant sur le chemin des streptocoques envahisseurs.

– **les toxines qui lysent les cellules :**

Elles agissent habituellement comme des lipases ou en s'insérant dans la membrane pour former des pores. Un exemple de lipase est donné par la lécithinase, qui lyse les cellules sans discrimination car son substrat est ubiquitaire dans les membranes des mammifères. Plusieurs toxines présentant ces caractéristiques sont appelées hémolysines, car elles sont capables de lyser les hématies. Cependant, elles sont également capable de lyser des leucocytes. Ainsi, les micro-organismes éliminent les défenses de l'hôte, en créant un milieu anaérobie, nécrotique, et riche en nutriments dans lequel ils peuvent se développer. L'autre classe perturbe la perméabilité de la membrane en y formant des pores protéiques non sélectifs pour l'eau et les ions. Ainsi, l'eau pénètre dans le cytoplasme entraînant ainsi la lyse de la cellule. Même à faible concentration, les fonctions cellulaires peuvent être altérées par de petites perturbations dans la membrane qui provoquent la fuite des ions, indispensables à la survie de la cellule. Ces mécanismes peuvent donc détruire efficacement les phagocytes, qui sont la première ligne de défense de l'hôte.

## 2. Les toxines à cible intracellulaire (de type AB)

La plupart des toxines à cible intracellulaire sont synthétisées sous la forme de précurseur inactifs qui doivent être activés pour devenir toxiques. La toxine de la diphtérie (de *Corynebacterium diphtheriae*), du choléra (de *Vibrio cholerae*), du tétanos (de *Clostridium tetani*), et des shigelles (de *Shigella dysenteriae*) en font partie. Le précurseur inactif est généralement composé de deux parties, l'une impliquée dans la liaison à la membrane de la cellule hôte (on appelle cette sous-unité "B" pour "Binding"), l'autre responsable de l'activité toxique (sous-unité "A" pour "Active"). Ces toxines sont appelées toxines AB. Le plus souvent, la sous-unité B se lie à un récepteur membranaire de la cellule hôte et la toxine est ensuite endocytée dans une vésicule membranaire (endosome) dans la cellule. L'acidification de l'endosome induit une séparation des sous-unités A et B. La sous-unité B reste dans l'endosome et sera recyclée à la surface. La sous-unité A est transportée vers des organelles spécifiques avant d'être libérée dans le cytosol. Cependant, l'entrée de la toxine dans la cellule cible peut également avoir lieu par une translocation directe. Dans ce cas, la liaison de la sous-unité B sur le récepteur induit la formation d'un pore dans la membrane, au travers duquel la sous-unité A est transférée dans le cytoplasme. La sous-unité A de la protéine peut ensuite être activée par clivage protéolytique et réduction des ponts disulfure pour devenir toxique.

Un exemple le plus classique du mécanisme d'entrée d'une toxine dans la cellule hôte, est celui de la toxine diphtérique (DT), dont la bactérie *Corynebacterium diphtheriae* est responsable d'une maladie respiratoire sévère. Le mécanisme est illustré sur la figure II.1.

Un clivage protéolytique de DT, de la famille des furines, a lieu à la surface de la cellule, et génère les deux sous-unités DTA et DTB reliées par un pont disulfure [34]. DTA constitue la sous-unité "A", et DTB la sous-unité "B". DTB est constitué de deux domaines : DTT qui facilite la translocation à travers la membrane endosomale, et DTR qui se fixe au récepteur membranaire de la cellule cible. DT est ensuite endocyté. L'acidification de l'endosome provoque un changement conformationnel majeur qui permet aux hélices  $\alpha$  de DTT de pénétrer dans la membrane de l'endosome (endomembrane) et de former un canal permettant la translocation de DTA vers le cytosol. Dans sa forme insérée dans la membrane, DTT peut aussi jouer le rôle de chaperon en interagissant partiellement avec DTA, facilitant ainsi son dépliement pendant sa translocation. DTA passe ensuite dans le cytosol *via* des facteurs cytosoliques, où il est restructuré vers sa forme active, puis relargué grâce à la réduction du pont disulfure.

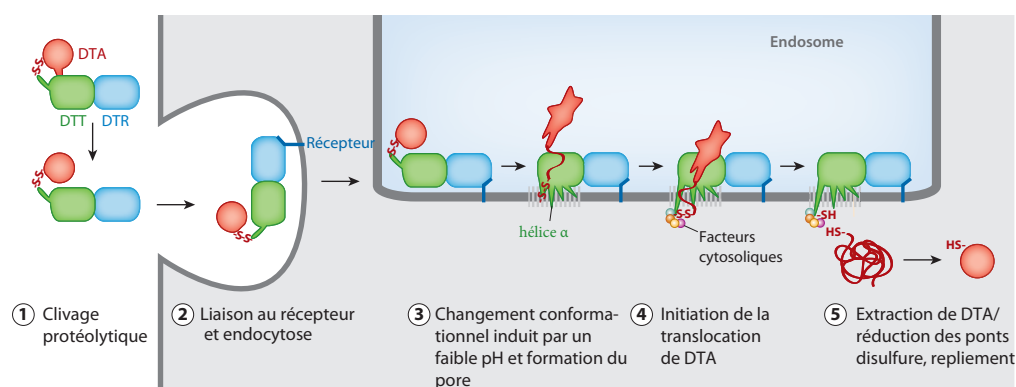


FIG. II.1: *Désassemblage de la toxine diphtérique au niveau de la membrane plasmique et dans les endosomes.* Adaptation de la figure 4b de la revue de Takamasa Inoue [34].

Voici quelques exemples de toxines à cible intracellulaire :

#### – les toxines qui bloquent la synthèse protéique :

La toxine diphtérique (de *Corynebacterium diphtheriae*), dont le mécanisme d'entrée dans la cellule hôte a été décrit précédemment, est une ADP ribosyl-transférase activement dirigée contre le facteur d'élongation 2 (ou EF2), que l'on trouve dans les protéines eucaryotes. EF2 catalyse l'hydrolyse du guanosine triphosphate (GTP) nécessaire au mouvement des ribosomes sur l'ARNm, permettant ainsi la synthèse protéique. Par une réaction d'ADP-ribosylation de EF2, la toxine diphtérique inhibe ainsi la synthèse protéique [33]. L'exotoxine A sécrétée par *P. aeruginosa* agit également de la même façon.

La toxine shiga de *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei* ou encore *Shigella boydii*, bloque la synthèse protéique en agissant comme une N-glycosidase. C'est une toxine de type AB<sub>5</sub> (avec 5 sous-unités B) dont la sous-unité A est responsable du clivage d'une

liaison N-glycosidique de l'adénine à la position nucléotidique 4324 de l'ARN ribosomique 28S de la sous-unité ribosomale 60S, altérant ainsi la synthèse protéique [35].

– **les toxines qui interfèrent sur la transduction du signal :**

Parmi les toxines qui ne détruisent pas la cellule, il en existe qui fonctionnent en élevant la concentration d'AMP cyclique (AMPc), qui joue un rôle essentiel dans la transduction du signal. C'est le cas de la toxine cholérique (de *Vibrio cholerae*), dont le tissu cible est l'épithélium de l'intestin grêle. La sous-unité B a une affinité spécifique pour la muqueuse épithéliale intestinale. La sous-unité A agit par ADP-ribosylation de la GTPase, qui est impliquée dans la régulation de la synthèse d'AMP cyclique, altérant ainsi la production de ce dernier. Un autre exemple est celui de la toxine CyaA de *Bordetella pertussis*, sur laquelle porte la présente thèse. La toxine, qui est une adénylcyclase, est composée de plusieurs domaines, dont un qui permet sa liaison à un récepteur d'une cellule phagocytaire, et d'autres qui lui permettent d'être directement transloquée dans la cellule hôte, où elle sera activée comme adénylcyclase par la calmoduline, protéine ubiquitaire dans les cellules eucaryotes. Elle catalyse alors l'hydrolyse de l'ATP en AMPc et PPI de manière incontrôlée. L'AMPc est produit en grandes quantités, ce qui altère le système immunitaire de l'hôte.

– **les toxines qui bloquent la transmission synaptique :**

Les plus létales d'entre elles sont les neurotoxines tétanique et botulinique, produites par *Clostridium tetani* et *Clostridium botulinum*, qui fonctionnent de façon intracellulaire. Les neurotoxines botuliniques et tétanique sont synthétisées sous forme d'une seule chaîne protéique de 150 kDa environ. Ces protéines ne possèdent pas de séquence signal et leur sécrétion intervient par un processus encore mal défini d'exfoliation des parois bactériennes ou par lyse de ces parois. L'activation des neurotoxines fait intervenir une protéolyse qui clive la protéine précurseur en deux fragments A et B. Les deux chaînes restent réunies par un pont disulfure.

La toxine tétanique est produite dans la plaie d'inoculation et diffuse jusqu'aux extrémités terminales des motoneurones. Le fragment B, activé par protéolyse et par réduction des ponts disulfures, se lie aux récepteurs gangliosidiques des neurones. Après internalisation, il migre le long des axones vers la moelle épinière et le tronc cérébral par un mécanisme de transport rétrograde. À ce niveau, il se fixe au niveau des terminaisons présynaptiques et bloque la libération de neurotransmetteurs inhibiteurs, en l'occurrence la glycine et surtout l'acide gamma amino-butyrique (GABA). La diminution de l'inhibition résulte en une augmentation de l'activité des neurones moteurs et provoque les spasmes musculaires et les contractions musculaires rigides caractéristiques du tétanos.

La toxine botulinique, elle, est produite dans la nourriture contaminée. Elle transite par le tube digestif, et est activée par les protéases intestinales. Elle traverse ensuite la muqueuse intestinale et est véhiculée par le sang ou la lymphe jusqu'aux extrémités

axonales amyéliniques des motoneurones, où elle est internalisée. Contrairement à la toxine tétanique, la toxine botulinique reste à l'extrémité présynaptique des motoneurones et agit sur les terminaisons nerveuses périphériques. La toxine botulinique provoque ainsi un blocage présynaptique de la libération d'acétylcholine aux jonctions neuromusculaires. L'interruption de la stimulation nerveuse entraîne une relaxation irréversible des muscles, conduisant à l'arrêt respiratoire.

## 1.2 Toxines de pathogènes augmentant la production d'AMPc

La communication entre les cellules se fait grâce à des molécules diffusant dans les milieux extracellulaires. Elles sont détectées par des protéines transmembranaires (récepteurs) qui sont activés par des messagers moléculaires secondaires intracellulaires. Ces messagers induisent à leur tour une réponse cellulaire à travers une cascade de réactions moléculaires à partir de la face intracellulaire de la membrane plasmique. Ce processus est appelé **transduction du signal**. Les messagers moléculaires sont impliqués dans de nombreux processus tels que l'activation de protéines, la synthèse protéique, la division, la différenciation et la mort cellulaire (apoptose), le métabolisme énergétique de l'organisme, l'homéostasie, la croissance de l'individu, le fonctionnement des organes etc.

L'adénosine monophosphate cyclique (AMPc) est un second messager moléculaire qui joue un rôle important dans la signalisation cellulaire, mais aussi dans la régulation de l'expression de gène. Il est le produit de l'hydrolyse de l'ATP par une adénylcyclase présente dans les membranes plasmiques. Cette réaction libère du pyrophosphate qui est immédiatement dégradé par une pyrophosphatase, rendant la réaction irréversible (figure II.2).

### Rôle de l'AMPc dans la transduction de signaux

L'interaction spécifique entre une molécule de communication intercellulaire et un de ses récepteurs spécifiques entraîne des réactions moléculaires en cascade à partir de la face intracellulaire de la membrane plasmique : c'est le processus de transduction du signal qui initie la réponse de la cellule à la stimulation extracellulaire portée par le messager intercellulaire. Les récepteurs transmembranaires qui sont impliqués dans la transduction du signal sont dits métabotropes. Ils assurent la transmission du signal extracellulaire au travers de la membrane puis sa traduction en un nouveau messager intracellulaire, l'AMPc, interprétable par la cellule en terme de réponse physiologique. Il existe principalement deux types de récepteurs métabotropiques : les récepteurs tyrosine-kinases, qui activent la transduction du signal par phosphorylation d'une protéine cible, et les récepteurs couplés aux protéines G, qui activent des effecteurs enzymatiques, comme par exemple l'adénylcyclase qui catalyse l'hydrolyse de l'ATP en AMPc. Ces récepteurs sont aussi connus sous le nom de récepteurs à sept domaines transmembranaires. Un exemple est illustré sur la figure II.2.

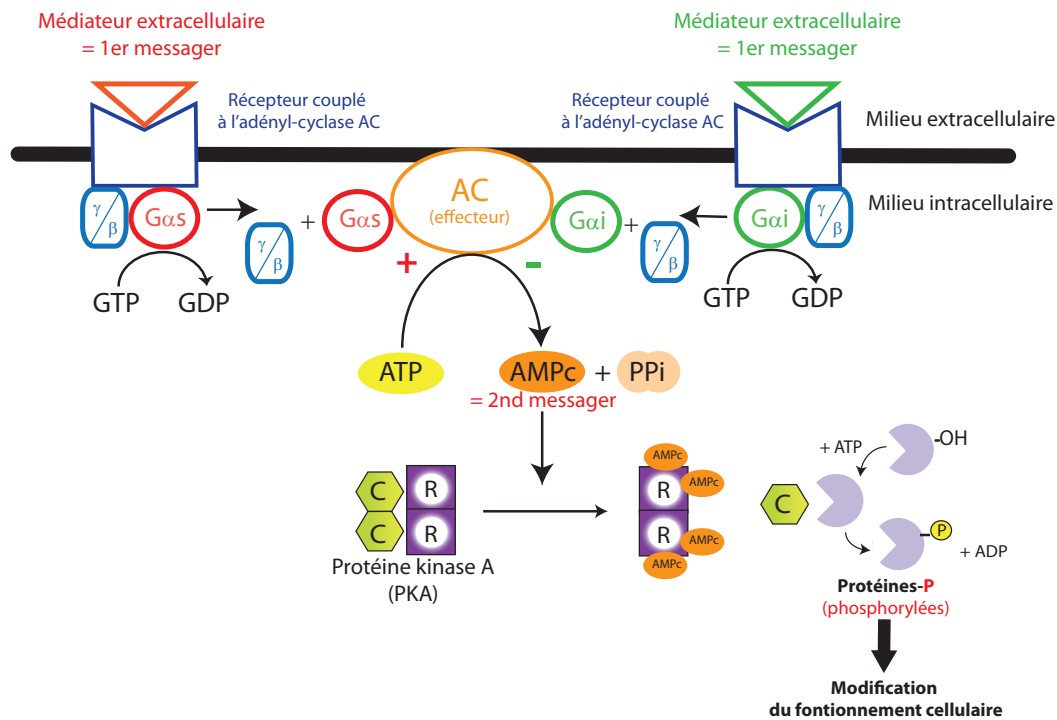


FIG. II.2: Régulation et rôle de l'AMPc

Les protéines G appartiennent au vaste groupe des GTPases. Chaque protéine G est un complexe de sous-unités  $\alpha$ ,  $\beta$  et  $\gamma$ , qui est couplée à un récepteur II.2. Elle permet de transmettre le message apporté par le premier messenger (par exemple une hormone activatrice) à un enzyme effecteur, ici l'adényl cyclase. La sous-unité  $\alpha$  est liée à une molécule de Guanosine Diphosphate (GDP) ou de Guanosine Triphosphate (GTP). Le transfert d'information apporté par le premier messenger extracellulaire se fait par la fixation de ce dernier sur le récepteur, ce qui va changer la conformation de la sous-unité  $G\alpha_s$ , lui permettant ainsi de libérer sa molécule GDP et de lier une molécule de GTP (Guanosine Triphosphate). La sous-unité  $G\alpha_s$  se dissocie alors du dimère  $\beta\gamma$  et va activer l'adényl cyclase (d'où le nom  $G\alpha_s$ , s pour "stimulateur"). Cependant ce message peut aussi être inhibiteur de l'adényl cyclase si un médiateur extracellulaire inhibiteur se fixe sur un récepteur couplé à la sous-unité  $G\alpha_i$  (i pour "inhibiteur"). L'hydrolyse assez lente du GTP en GDP désactive la protéine G et arrête le couplage récepteur/effecteur, évitant ainsi une action trop prolongée du système de transduction.

L'adényl cyclase activée catalyse l'hydrolyse de l'ATP en AMPc, second messenger de la transduction du signal qui va à son tour activer une protéine-kinase A (PKA). Celle-ci est composée de deux sous-unités régulatrices R et deux sous-unités catalytiques C. En se liant sur les sous-unités R, l'AMPc libère les sous-unités C, qui vont catalyser la phosphorylation d'enzymes spécifiques de nombreuses voies métaboliques, comme par exemple la glycolyse, lipolyse, glycogénolyse ou gluconéogenèse (figure II.3) au niveau cytosolique.

La PKA activée peut aussi avoir pour cible des récepteurs qui forment les canaux ioniques

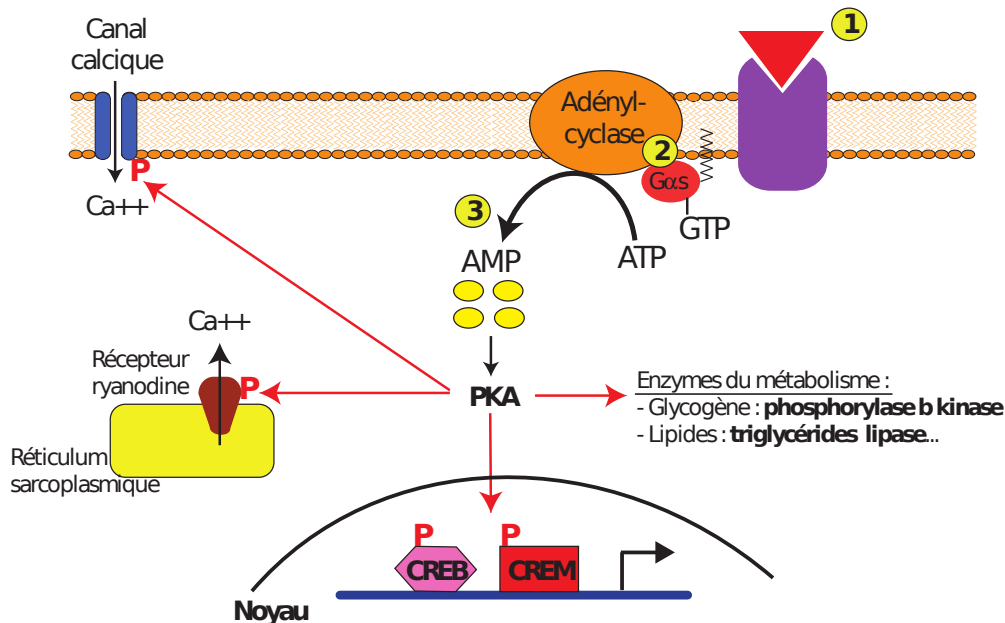


FIG. II.3: Cycle du PKA

calciques insérés dans les membranes du réticulum endoplasmique ou du réticulum sarcoplasmique (appelé récepteur de la ryanodine). Ces derniers sont exprimés dans les muscles squelettiques, mais également dans le myocarde et le cerveau. La phosphorylation du récepteur de la ryanodine active ce dernier, ce qui entraîne une sortie des ions  $\text{Ca}^{2+}$  vers le cytoplasme. La concentration de  $\text{Ca}^{2+}$  qui en résulte dans le réticulum est alors suffisamment faible pour induire une fermeture du canal (figure II.3), et donc la contraction musculaire. Au niveau nucléaire, les facteurs transcriptionnels CREB (“*cAMP response element binding protein*”) et CREM (“*cAMP-responsive-element-binding modulator*”) sont également les cibles de PKA. Ils modulent la transcription de nombreux gènes. Pour ne citer qu’un exemple la phosphorylation de CREB permet la progression du cycle cellulaire de la phase G1 en S en stimulant la transcription des cyclines spécifiques D et E. En effet, les mécanismes de la régulation du cycle cellulaire reposent essentiellement sur deux structures protéiques complémentaires appelées Cdk (Cycline-dependent kinase) et cycline. À chaque phase du cycle cellulaire correspondent un ou plusieurs complexes Cdk-cycline. Au cours de la phase G1, ce sont les cyclines spécifiques D et E qui sont transcrites. L’activation de Cdk4 par la cycline D et l’activation de Cdk2 par la cycline E vont ensuite permettre la phosphorylation de protéines jouant un rôle dans la progression du cycle cellulaire.

### Rôle de l'AMPc chez *E.coli*

Chez la bactérie *E.coli*, les gènes impliqués dans un processus métabolique sont organisés sous forme d'opéron, c'est-à-dire sous forme groupés et organisés en une seule unité de transcription. La bactérie trouve sa source de carbone dans le catabolisme des sucres. Si le glucose est la source "préférée", le lactose peut également être consommé par la bactérie et métabolisé en galactose et glucose. Un schéma de la régulation positive de la transcription des gènes de l'opéron lactose est proposé sur la figure II.4.

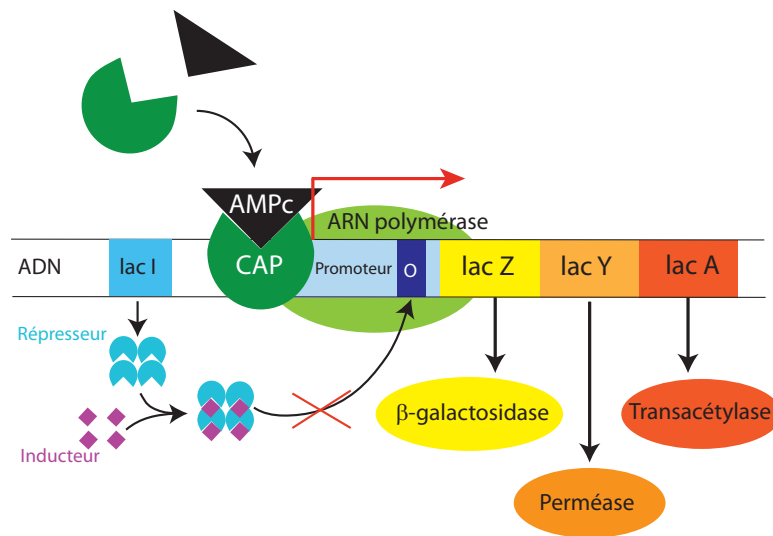


FIG. II.4: Opéron lactose de *E.coli* : régulation positive de la transcription des gènes

Dans l'opéron lactose on trouve trois gènes, dont deux sont indispensables à la dégradation du lactose. Le gène lacZ code pour la  $\beta$ -galactosidase, qui hydrolyse le lactose en galactose et glucose. Le gène lacY code pour la lactose perméase qui permet l'entrée du lactose dans la bactérie. Enfin, le gène lacA coderait pour une  $\beta$ -thiogalactoside acétyltransférase permettant à la cellule d'utiliser les thiogalactosides [36]. En amont de l'opéron, on trouve le gène lacI, qui code pour une protéine régulatrice appelée répresseur. Celle-ci inhibe la transcription des gènes de l'opéron en se fixant sur le site opérateur O. En présence de lactose et de glucose, le répresseur est inactivé par l'allolactose, un isomère du lactose, libérant ainsi l'opérateur. Les gènes de l'opéron peuvent donc être transcrits grâce à l'ARN polymérase.

Cependant, tant que le glucose est présent, la bactérie va le métaboliser préférentiellement, et n'a donc pas besoin de synthétiser les enzymes de l'opéron lactose. Ceci implique donc l'existence d'un autre mécanisme de régulation que l'on appelle la répression catabolique. Ce n'est que lorsque la concentration en glucose diminue que le métabolisme du lactose devient nécessaire. Ce signal de carence est traduit par une augmentation du taux d'AMPc, qui forme alors un complexe avec la protéine CAP (pour Catabolite gene Activator Protein). Ce complexe se lie à l'ADN en amont du site de fixation de l'ARN-polymérase. L'interaction CAP-AMPc agit comme un inducteur qui renforce l'affinité de l'ARN-polymérase pour le promoteur de l'opéron.

Cette régulation positive permet d'augmenter d'un facteur 50 la transcription de l'opéron lactose. En présence de glucose, il n'y a pas de complexe CAP-AMPc formé, ce qui réduit la transcription à un niveau très faible.

### Les toxines entraînant une surproduction d'AMPc

#### – La toxine cholérique :

La toxine cholérique (CT) de *Vibrio cholerae* et l'entérotoxine thermolabile (LT) de *Escherichia coli* sont à l'origine des syndromes diarrhéiques respectivement provoqués par le vibrion cholérique et par certaines souches de *E. coli* entérotoxino-gènes. Ces deux toxines sont constituées d'une sous-unité A catalytique associée à 5 sous-unités B en anneau capables d'interagir avec les récepteurs cellulaires. Dans le cas de la CT, la sous-unité B (appelée CTB) se lie aux récepteurs gangliosidiques des cellules épithéliales de l'intestin. Contrairement au mécanisme classique d'entrée dans la cellule (comme celui de la toxine diphtérique), les sous-unités A (appelée CTA) et B (CTB) ne sont pas désassemblées à la surface de la cellule. Le complexe toxine-récepteur est ensuite endocyté et transporté vers la machinerie de sécrétion du réticulum endoplasmique, qui en temps normal, "rétro-transloque" les protéines mal repliées vers le cytosol pour qu'elles y soient dégradées. CTB se lie sur un complexe de rétrotranslocation présent sur la membrane du réticulum endoplasmique, composé de deux protéines Derlin1 (ou Derlin2) et Hrd1. C'est à ce niveau que les sous-unités CTA et CTB sont séparées. En effet, CT est réduit par une réductase, générant ainsi la sous-unité CTA1. La protéine disulfide isomérase (PDI), en se fixant sur le complexe Derlin1-Hrd1, va déplier CTA1 par une réaction d'oxydo-réduction. La sous-unité CTA1 dépliée peut alors être transloquée vers le cytosol au travers d'un canal formé par le complexe Derlin1-Hrd1. Se repliant rapidement dans le cytosol, elle évite ainsi la dégradation par le protéasome. Ensuite, en ajoutant un ADP ribose sur la protéine G $\alpha$ s, impliquée dans la transduction du signal, CTA1 bloque l'activité GTPase de celle-ci, ce qui maintient donc G $\alpha$ s en permanence sous sa forme active. L'activité adénosine diphosphate ribosylase entraîne donc l'activation constitutive de l'adényl cyclase membranaire. Cette protéine effectrice induit une augmentation de la concentration en adénosine monophosphate cyclique (AMPc) intracellulaire. Ce second messenger activera, à son tour, une protéine kinase A qui viendra phosphoryler les canaux ioniques, modifiant ainsi leur activité. Il s'ensuit une sécrétion importante de chlorure de sodium et, par osmose, d'eau en direction de la lumière intestinale, entraînant ainsi la diarrhée.

D'autre part, certains micro-organismes apportent leur propre adényl cyclase, qu'ils transloquent dans le cytosol de celle-ci, ce qui a pour conséquence une surproduction d'AMPc [37]. En effet, ces toxines adényl cyclases bactériennes sont dotées d'un coeur catalytique qui diffère des AC mammifères, et qui leur confère un taux de production d'AMPc largement plus élevé que celui des AC mammifères [38; 39] lorsqu'elles sont activées dans la cellule hôte eucaryote. En général, elles sont activées par la calmoduline, qui est une protéine eucaryote ubiquitaire sensible aux ions calcium, et extrêmement abondante dans le cytoplasme de toutes les cellules supérieures. La calmoduline est hautement conservée au cours de l'évolution et elle est le principal médiateur



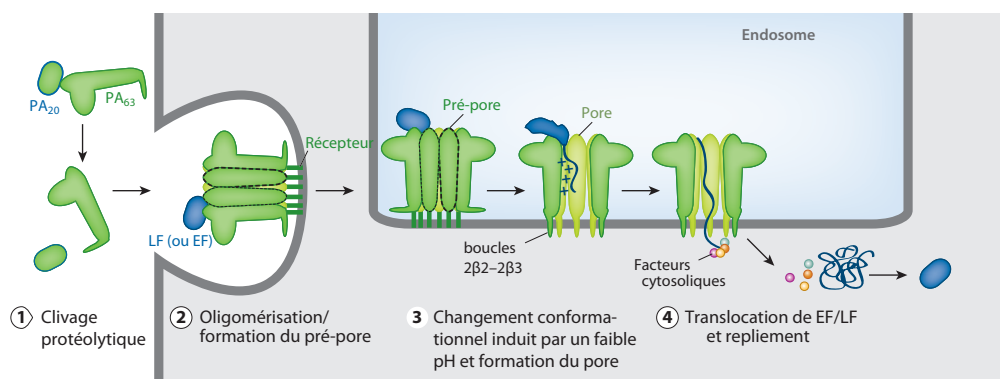
de la signalisation cellulaire du calcium [40]. Actuellement, seulement trois toxines bactériennes ont été clairement identifiées comme adénylcyclase capables d'intoxiquer les cellules eucaryotes :

– **L'adénylcyclase CyaA :**

L'adénylcyclase CyaA, également connue sous le nom de ACT, AC-Hly et toxine AC [41], est produite par *Bordetella pertussis*, bactérie responsable de la maladie de la coqueluche ainsi que par d'autres bactéries apparentées du même genre, *Bordetella parapertussis*, et *Bordetella bronchiseptica*, qui sont des pathogènes d'animaux. La toxine CyaA cible les cellules du système immunitaire (macrophages, leucocytes...). Cette exotoxine possède un mécanisme d'invasion des cellules eucaryotes unique. En effet, c'est une protéine bifonctionnelle dotée d'un domaine C-terminal hémolytique lui permettant d'être "directement" internalisée dans la cellule hôte, et d'un domaine N-terminal catalytique qui possède l'activité d'adénylcyclase dépendante de la calmoduline. L'activation de l'adénylcyclase par cette dernière, permet la catalyse de la synthèse d'AMPc de manière incontrôlée, ce qui conduit à l'altération du système immunitaire [42; 43; 44]. Les propriétés de cet agent pathogène et son mécanisme d'invasion des cellules sont expliqués en détail dans la section suivante.

– **Le facteur œdémateux de l'anthrax (EF) :**

Le facteur œdémateux (EF), produit par *B. anthracis*, est l'un des facteurs responsables de la maladie du charbon, autrement dit l'anthrax. Il cible non seulement les cellules du système immunitaire (macrophages, les neutrophiles, cellules T, les monocytes) mais également les cellules endothéliales et les plaquettes [45]. Homologue à CyaA d'un point de vue structural, ces deux protéines ne partagent que 25% d'identité de séquence. Son mécanisme d'invasion de la cellule hôte est illustré sur la figure II.5.



**FIG. II.5: Mécanisme d'entrée de la toxine de l'anthrax (AT) au niveau de la membrane plasmique et dans les endosomes.** Adaptation de la figure 4a de la revue de Takamasa Inoue [34].

La toxine de l'anthrax (AT) est composée de trois sous-unités : le facteur œdémateux (EF), le facteur létal (LF) et un antigène protecteur (PA). Au travers de plusieurs étapes, PA permet la translocation de EF et LF dans le cytoplasme de la manière suivante. Tout d'abord, la forme

native de 83 kDa de PA (PA<sub>83</sub>), se lie au récepteur de la toxine de l'anthrax. Ensuite, PA<sub>80</sub> est clivé en un fragment de 20 kDa (PA<sub>20</sub>) et de 63 kDa (PA<sub>63</sub>) par des endoprotéases membranaires de la famille des furines. Enfin, PA<sub>20</sub> se dissocie pour permettre à PA<sub>63</sub> d'oligomériser et former un heptamère (ou octamère) appelé pré-pore. La surface du pré-pore heptamérique ou octamérique, devenue accessible après le relargage de PA<sub>20</sub>, peut ainsi lier trois à quatre molécules LF et/ou EF respectivement. Ces complexes ainsi assemblés sont ensuite endocytés par la cellule, dans un compartiment acide. Le faible pH de l'endosome induit ensuite un grand changement conformationnel du pré-pore PA<sub>63</sub> en pore, permettant ainsi l'insertion de la boucle 2β2-2β3 de PA<sub>63</sub> dans la membrane de l'endosome (endomembrane). La liaison au récepteur membranaire de l'endosome affecte également la formation du pore en diminuant le seuil de conversion de pré-pore en pore d'une unité de pH. Un faible pH déclenche aussi un dépliement de EF et LF, permettant ainsi leur translocation à travers le pore grâce à un gradient de pH. Pour terminer, des facteurs cytosoliques tirent LF et EF dans le cytosol, où ces dernières se replient pour restaurer leur activité catalytique. Le facteur létal LF agit comme une protéase dépendante du zinc qui coupe la partie N-terminale des MAP kinases ("*mitogen-activated protein kinase kinases (MAPKK)*"), ce qui inhibe ces kinases en les empêchant de se lier efficacement à leurs substrats. La transduction du signal est ainsi altérée, conduisant ainsi à l'apoptose. Le facteur œdémateux (EF) agit comme une adénylcyclase, activée par la calmoduline dépendante des ions Ca<sup>2+</sup>, qui augmente la concentration intracellulaire en AMPc, perturbant ainsi la transduction du signal, ce qui conduit également à la mort cellulaire.

Une comparaison entre les propriétés et la structure de EF et de CyaA est décrite dans la section 2.4.

#### – ExoY :

Produite par *P. aeruginosa*, un pathogène humain souvent responsable d'infections nosocomiales. Les pathologies qu'elle engendre sont diverses (infection de l'oeil, des plaies, des brûlures, des urines, des poumons, des méningites etc) mais sévères lorsque la défense immunitaire normale de l'hôte est altérée. La virulence du pathogène est en partie due à l'expression des gènes du régulon exoenzyme S, importants pour la régulation, la sécrétion et l'intoxication des cellules eucaryotes [46; 47; 48]. ExoY est l'un des effecteurs exprimés par ce régulon [47], qui est transloqué dans le cytoplasme des cellules eucaryotes *via* un mécanisme de sécrétion de type III. Ce type de sécrétion est une organisation complexe et multi-protéique qui joue un rôle clé dans le processus infectieux, permettant l'injection direct d'effecteurs spécifiques bactériens dans le cytosol des cellules eucaryotes. Ainsi, l'intoxication est limitée au site de colonisation, contrairement à EF et CyaA qui peuvent exercer leur effet à distance du site d'infection. Homologue à CyaA et EF au niveau de deux régions, dont la région catalytique, ExoY possède une activité adénylcyclase qui conduit à une surproduction de l'AMPc intracellulaire de l'hôte. Son activité basale est proche de celle de CyaA [47]. Cependant, à la différence de CyaA et EF, elle n'est pas activée par la calmoduline, mais par un facteur eucaryotique distinct de cette dernière et encore inconnu à ce jour.

## 2 La toxine adénylcyclase CyaA

### 2.1 La coqueluche

#### Epidémiologie

La coqueluche est une maladie respiratoire très contagieuse qui se développe par épidémie. La première description clinique authentique de la maladie est celle de Guillaume de Baillou en 1578 au cours d'une épidémie à Paris, faite sous le nom de *Tussis quintina* [49; 50]. Au siècle suivant, Willis et Sydenham décrivent le syndrome coquelucheux sous le nom de pertussis (étymologiquement : toux sévère). De nombreuses épidémies furent décrites au XVIII<sup>ème</sup> et XIX<sup>ème</sup> siècle. Mais ce n'est qu'en 1900 que l'agent principal, la bactérie *Bordetella pertussis*, fût découvert par Jules Bordet et Octave Gengou, qui ne l'ont isolée qu'en 1906 après la mise au point du premier milieu de culture à base d'extrait de pommes de terre [51].

Depuis 1980, il existe une recrudescence de la coqueluche chez les adolescents et les adultes n'étant pas vaccinés. En France, entre 1999 et 2000, la coqueluche est la première cause de décès par infection bactérienne communautaire chez les nourrissons ayant entre 10 jours de vie et 2 mois. Aux États-Unis en 2005, 12 000 cas de coqueluche ont été diagnostiqués, soit 6 fois plus qu'en 1980. De plus, au Canada, la proportion de cas de coqueluche chez les adolescents (de plus de 15 ans) et les adultes est passée de 9,6 % en 1995 à 31,3 % en 2004. Cette augmentation peut être sans doute attribuée au déclin de l'immunité chez les adolescents et les adultes.

La coqueluche est actuellement l'une des dix causes les plus mortelles dues à des maladies infectieuses, et a causé en 2008 16 millions de cas dans le monde dont 95% dans les pays en développement, et 195 000 morts infantiles [52]. D'après Crowcroft & Pebody [53], la coqueluche est l'une des cinq causes majeures des morts chez les jeunes enfants.

#### Le germe pathogène et la maladie

Le germe pathogène responsable de la coqueluche appartient au genre *Bordetella* qui inclu 9 espèces dont les trois plus étudiés sont *Bordetella pertussis* (coqueluche), *Bordetella parapertussis* (paracoqueluche, pathologie analogue à la coqueluche), et *Bordetella bronchiseptica*. Le séquençage d'un ancêtre commun aux trois espèces et des études comparatives de phylogénie ont permis de supposer que *B.pertussis* et *B.parapertussis* auraient dérivé indépendamment à partir de différents clones de *B.bronchiseptica* [54; 55; 56].

*B. pertussis* est un pathogène strictement humain qui n'a pas de réservoir animal ou environnemental connu à ce jour. Ce coccobacille à Gram négatif encapsulé, est de petite taille (0.8  $\mu\text{m}$  sur 0.4  $\mu\text{m}$ ), et ne produit pas de spores. Il est extrêmement labile et ne peut survivre en-dehors de son hôte [57]. De ce fait, il a une facilité à se transmettre de manière très efficace d'un hôte à un autre par voie aérienne, d'autant plus qu'il présente une affinité exclusive pour les muqueuses des voies respiratoires de l'homme. Cependant, à la phase d'état de la maladie, il n'y a plus de bacilles viables dans les sécrétions : les signes de la maladie sont dus aux différents facteurs de virulence

suivants : a) les adhésines : l'hémagglutinine filamenteuse (FHA), pertactine, les fimbriae de type 2 et 3 ; b) les toxines : toxine pertussique (PTX), la toxine adénylcyclase (TAC), la cytotoxine trachéale (TCT), le lipooligosaccharide et l'endotoxine coquelucheuse [58]. La FHA, la pertactine et les fimbriae facilitent la fixation aux cellules hôtes cibles, et la PTX, TCT, et TAC permettent à la bactérie de détruire le revêtement épithélial et d'échapper au système immunitaire de l'hôte.

Dans les régions où les enfants n'ont pas été vaccinés (pays en développement), la transmission se fait d'enfant à enfant. A l'inverse, dans les pays où les enfants sont vaccinés depuis des décennies (pays industrialisés) la transmission se fait d'adulte à enfant par le biais de gouttelettes respiratoires de personnes infectées [58]. Longtemps considérée comme une maladie de la petite enfance, elle peut être sévère pour l'homme à tout âge, mais elle est particulièrement dramatique, voire mortelle pour les nourrissons de moins de six mois et les personnes à risque comme les femmes enceintes ou les personnes âgées. La coqueluche ne confère pas une immunité à vie et peut être contractée plusieurs fois.

### Mécanisme d'infection et symptômes

Elle se caractérise par trois phases [59] :

1) une phase d'incubation de 9 à 10 jours environ sans aucun symptôme, suivi d'une rhinorrhée atypique de deux semaines environ. Pendant cette période, *B.pertussis* est capable de s'attacher à l'épithélium de la trachée via diverses molécules adhésives (les adhésines). Les mouvements des cils vibratiles des cellules trachéales sont alors bloqués, permettant ainsi la multiplication des bactéries dans les voies respiratoires et une accumulation de mucus et de débris cellulaires. Les symptômes catarrhaux présentés à ce stade, notamment une toux, ne sont pas spécifiques de la maladie, ce qui rend le diagnostique difficile.

2) une phase paroxystique qui se caractérise par des quintes de toux pouvant se terminer par le classique chant du coq. Dans les cas typiques, la toux est particulièrement paroxysmique la nuit et s'accompagne souvent de vomissements (elle évoque souvent le chant du coq). Ces symptômes peuvent être dus à une difficulté à expulser le mucus des bronches et de la trachée, mais il n'y a actuellement aucune explication possible à ces toux paroxismales caractéristiques. Cette phase peut durer plusieurs mois. Chez le nourrisson, la coqueluche peut ne provoquer qu'une apnée et une cyanose, tandis que chez l'adolescent et l'adulte, une toux persistante, non caractéristique, peut être la seule manifestation de la maladie.

3) une phase de convalescence qui peut durer de une à plusieurs semaines et pendant laquelle la toux diminue graduellement.

La coqueluche peut entraîner chez les jeunes enfants des complications sérieuses comme la pneumonie, les affections neurologiques (crises convulsives, encéphalites) ou des hernies. Chez le nourrisson, la coqueluche peut être très grave voire mortelle. Les cas de coqueluche sont cliniquement identifiables chez l'enfant âgé de 1 à 5 ans, les cas graves et les décès sont surtout observés au cours des premières semaines ou des premiers mois de la vie. Chez les grands enfants, les adolescents et les adultes, en fonction de leur restant d'immunité, la coqueluche passe souvent inaperçue. C'est pourquoi il est important de confirmer la maladie par un diagnostique biologique

(culture de germe, identification du matériel génétique par PCR) pour arrêter la transmission qui est très facile car la maladie est très contagieuse.

### Traitements et vaccination

Lorsque la maladie s'est déclarée, un traitement par antibiothérapie est alors nécessaire. L'érythromycine est l'antibiotique le plus souvent utilisé, car il élimine la présence de la bactérie dans les sécrétions, ce qui diminue ainsi les risques de contamination. En général, les antibiotiques sont efficaces durant le temps d'incubation, lorsque la maladie n'a pas encore été diagnostiquée.

Depuis les années 1940, la vaccination constitue un progrès remarquable dans la prévention de cette infection. Si la baisse de la mortalité due à la coqueluche a été observée avant l'arrivée de la vaccination, c'est la mise en place de programmes de vaccination généralisée qui a permis de réduire considérablement la mortalité.

L'Organisation Mondiale pour la Santé (OMS) estime qu'en 2008 [58], la vaccination contre la coqueluche a permis d'éviter environ 687 000 décès dans le monde. La vaccination à grande échelle pratiquée dans les années 1950 à 1960 a conduit à une réduction spectaculaire (>90%) de l'incidence de la coqueluche et de la mortalité dans les pays industrialisés. Cependant, une étude dans ces pays montre que cette protection décline au bout de 4 ans, et ce pendant 8 ans [60]. On sait peu de choses sur la durée de l'efficacité du vaccin dans les pays en développement. De plus en plus de cas de coqueluche sont décelés chez les grands enfants, les adolescents et les adultes, qui deviennent alors des sources importantes de contamination pour les nourrissons.

Il existe deux types de vaccins anticoquelucheux : les vaccins anticoquelucheux à germes entiers, préparés à partir de *B. pertussis* tuées par chauffage et traitées au formol, et les vaccins acellulaires préparés à partir de composants bactériens sélectionnés et hautement purifiés.

Le vaccin anticoquelucheux à germes entiers est souvent associé à des anatoxines diphtériques et tétaniques (DTP) ou à d'autres vaccins administrés systématiquement au cours de la petite enfance (vaccins anti-*Haemophilus influenzae* type b (Hib), anti-hépatite B et antipoliomyélitique inactivé (VPI)). Ainsi, la réponse immunitaire est dirigée contre toute une série d'antigènes bactériens. Ce vaccin présente une bonne efficacité mais qui a tendance à diminuer progressivement d'après des études faites en Pologne et au Royaume-Uni.

Les vaccins acellulaires contiennent un ou plusieurs constituants antigéniques suivants : PTX, HAF, pertactine et fimbriae de types 2 et 3. Le rôle de chacun d'entre eux dans la protection n'est pas clairement établi, mais aucune altération importante de l'efficacité des vaccins n'a été documentée, quels que soient les changements survenus au cours du temps dans le patrimoine génétique des souches de *B. pertussis* circulant. Des études ont permis de montrer que les vaccins de ce type, à plusieurs constituants, sont plus efficaces que ceux qui n'en ont que deux ou un seul. Ils peuvent également être associés à d'autres vaccins comme pour les vaccins à germes entiers.

L'efficacité des vaccins anticoquelucheux entiers ou acellulaires varie en fonction de la souche

bactérienne responsable de la coqueluche. Toutefois, les meilleurs vaccins acellulaires sont plus efficaces que les moins bons vaccins entiers, mais peuvent être moins efficaces que les meilleurs vaccins entiers, pour prévenir la coqueluche.

Des données épidémiologiques indiquent que l'efficacité des traitements actuels pourraient s'estomper du fait d'une dérive antigénique et de la sélection des clones les moins sensibles aux vaccins ne se sont pas matérialisés. Mais rien n'indique qu'il y ait une émergence d'une résistance aux antimicrobiens chez ce germe.

## 2.2 L'adénylcyclase CyaA : Biogenèse, Structure et mécanisme d'invasion

La toxine CyaA (toxine adénylcyclase (TAC)) est l'un des facteurs de virulence majeure de *Bordetella pertussis*. Elle joue un rôle essentiel dans les premiers stades de colonisation du système respiratoire par *B. pertussis* [61; 62; 63]. Elle peut envahir de nombreux types cellulaires, comme les érythrocytes par exemple, plus connues sous le nom d'hématies, ou encore les myocytes (cellules des fibres musculaires), mais surtout les cellules du système immunitaire inné. Khelef et al [64] ont montré que *B. pertussis* induisait directement l'apoptose des macrophages alvéolaires *ex vivo*. Plus récemment, Harvill et al.[65] ont montré grâce à un modèle murin d'infection, que les neutrophiles et les macrophages sont, *in vivo*, les principales cibles de la toxine CyaA de *B. bronchiseptica*, une Bordetelle responsable d'infection chez l'animal. La toxine CyaA représente ainsi un mécanisme essentiel de défense contre les premiers effecteurs de la réponse immunitaire innée de l'hôte [66].

*B. pertussis* sécrète la protéine CyaA sous forme d'un polypeptide de 1706 kDa composé des trois domaines RD, H et AC. Dans un premier temps, son domaine RD (nommé R sur la figure II.6) se lie au récepteur de la cellule hôte. Puis dans un deuxième temps, son mécanisme unique d'entrée dans la cellule lui permet de perméabiliser la membrane plasmique grâce à son domaine "hémolysine" (H) et de transférer directement son domaine catalytique (AC) exclusivement, dans le cytosol de la cellule hôte (figure II.6). Enfin, AC est activé par la calmoduline qui est une protéine ubiquitaire et sensible aux ions calcium [67], engendrant ainsi une surproduction d'AMPc conduisant à la mort cellulaire.

### Biogenèse et Organisation structurale de CyaA

La toxine CyaA est codée par le gène *cyaA* dont l'expression est régulée par les signaux environnementaux [42; 68]. CyaA est une protéine de 1706 acides aminés, synthétisée sous forme d'un précurseur inactif, proCyaA, qui est converti en toxine active, capable d'intoxiquer les cellules-cibles, par acylation sélective de deux résidus lysine, la Lys860 et la Lys983 [69]. Cette modification post-traductionnelle est réalisée par une acyltransférase spécifique, CyaC, qui catalyse le transfert d'un groupe acyl de l'acyl-ACP ("*Acyl-Carrier Protein*") sur les groupes  $\epsilon$ -aminé des deux lysines de proCyaA. Le polypeptide acylé est ensuite sécrété à travers l'enveloppe bactérienne par une machinerie de sécrétion (de type I) constituée des protéines CyaB, CyaD et CyaE [70]. Les gènes

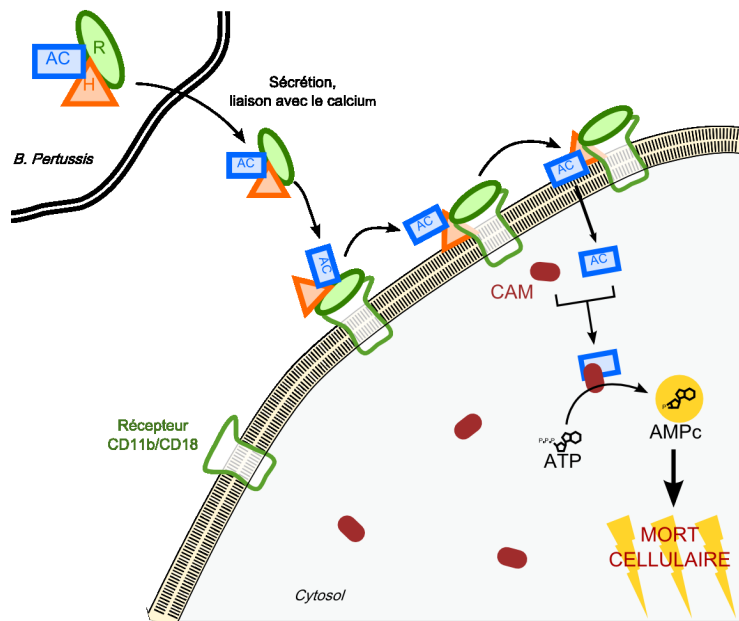


FIG. II.6: Schéma du mécanisme d'invasion de l'adénylcyclase *CyaA* dans la cellule hôte.

codant pour ces protéines sont organisés sur le chromosome de *B. pertussis* en opéron avec le gène de structure *cyaA*.

*CyaA* est une protéine bi-fonctionnelle car elle possède une activité adénylcyclase et une activité hémolytique (figure II.7). Le domaine catalytique (AC) est localisé dans les 400 résidus amino-terminaux tandis que la région responsable de l'activité hémolytique (domaine hémolysine) est localisée dans les 1300 résidus carboxy-terminaux.

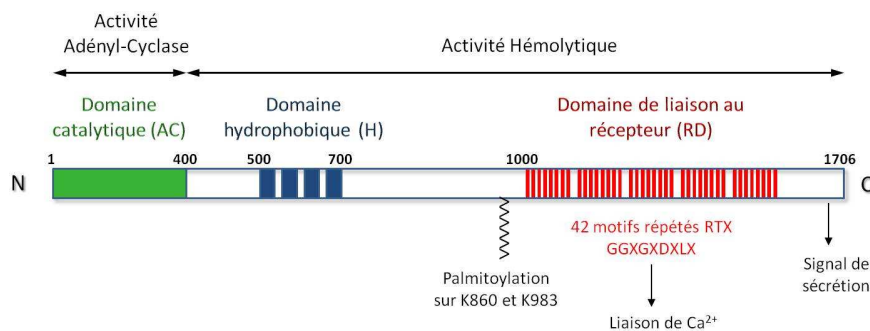


FIG. II.7: Organisation structurale de *CyaA* en trois domaines. Adaptée de Ladant et Ullmann [42]

Le domaine hémolysine possède une faible activité hémolytique intrinsèque, qui résulte de sa capacité à former dans des membranes biologiques, des canaux ioniques sélectifs pour les cations [71; 72; 73]. Ce domaine est par ailleurs essentiel pour la liaison de *CyaA* avec les cellules eucaryotes cibles ainsi que pour la translocation du domaine catalytique à travers la membrane plasmique des cellules-cibles. Plusieurs régions peuvent être identifiées dans ce domaine "hémolysine" :

- le domaine hydrophobe (H) entre les résidus 500 à 720. Des segments hydrophobes peuvent

former des hélices  $\alpha$  transmembranaires [42; 74]. Les délétions internes à cette région abolissent à la fois la cytotoxicité et l'activité hémolytique de CyaA [71];

- une région qui contient deux lysines acylées par CyaC;
- le domaine de liaison au récepteur (RD) qui comprend 30 à 40 copies d'un motif répété de neuf résidus dont le prototype est GGXGXDXLX (X représente n'importe quel acide aminé). Ce motif RTX est un motif structural caractéristique de la famille des toxines bactériennes. La liaison de calcium sur ces motifs RTX provoque des changements de conformation notables de CyaA [75; 71] qui sont, vraisemblablement, directement impliqués dans l'interaction avec la membrane et la translocation du domaine catalytique;
- l'extrémité C-terminale qui contient le signal de sécrétion reconnu par un appareil de sécrétion (de type I) constitué des protéines CyaB, CyaD, CyaE.

Le domaine catalytique possède quant à lui une activité enzymatique très élevée ( $k_{cat} = 2000 \text{ s}^{-1}$ ) en présence de CaM, qui se lie avec une forte affinité à CyaA ( $K_D < 0.1 \text{ nM}$ ) [43; 76].

### Mécanisme d'invasion des cellules eucaryotes

Le mécanisme d'entrée de CyaA dans les cellules eucaryotes, lui, est unique [42; 77]. Classiquement, les toxines entrent dans la cellule hôte par endocytose. Le domaine catalytique AC lui, est internalisé par une translocation directe dans le cytosol de la cellule hôte par un processus connu sous le nom d'internalisation ou intoxication [42; 78; 79]. Plusieurs résultats expérimentaux confortent ce modèle de translocation directe. Par exemple, une augmentation d'AMPc a pu être détectée dans les cellules cibles quelques secondes après l'ajout de la toxine CyaA, ce qui montre qu'il s'agit bien d'un processus de translocation et non d'endocytose, et qu'il est très rapide. Si il s'agissait d'une endocytose, l'augmentation d'AMPc n'aurait pas été détectée aussi vite. D'autre part, il a été démontré également que CyaA ne dépendait pas de l'acidification de vésicules endocytiques. De plus, CyaA peut intoxiquer des cellules qui n'ont pas de trafic membranaire comme les érythrocytes de Mammifères.

Le processus d'entrée se déroule schématiquement en trois étapes suivantes :

1. CyaA se lie sélectivement au récepteur, l'intégrine CD11b/CD18 ( $\alpha_M\beta_2$ ), par son domaine de liaison au récepteur (RD) en présence de calcium sur les motifs RTX, et avec une forte affinité [66]. L'intégrine CD11b/CD18 est naturellement exprimée par certains leucocytes, notamment les neutrophiles, les macrophages, les cellules dendritiques et les cellules NK [80]. De ce fait, ces cellules phagocytiques sont les principales cibles de la toxine CyaA lors de l'infection.
2. Ensuite, les segments hydrophobes de la protéine s'insèrent spontanément dans la membrane plasmique, de manière irréversible et à faible concentration de calcium [78]. La liaison de CyaA avec ces cellules requiert l'intégrité de la partie C-terminale de CyaA (résidus 400 à 1706) ainsi que l'acylation des lysines K860 et K983 [69]. Par ailleurs, il a été démontré que l'insertion de la toxine dans la membrane des érythrocytes conduit à la formation de pores perméables aux cations [71; 72; 78]. Cependant, des mutants dans lesquels l'activité



hémolytique a été abolie (incapables de perméabiliser la membrane par la formation de pores) sont toujours capables de transloquer le domaine catalytique AC. Cette constatation suggère que la formation de pore n'est pas nécessaire pour permettre l'entrée de AC [81]. De plus, il a été démontré [72] que l'activité hémolytique de CyaA permettait la perméabilisation de la membrane en formant des pores cationiques, mais que le diamètre de ces derniers est trop petit pour permettre à AC, même totalement déplié, de passer au travers. Les mécanismes moléculaires exacts par lesquels CyaA pénètre dans les cellules cibles restent à ce jour méconnus.

3. Le domaine catalytique (AC) de CyaA est transloqué à travers la membrane plasmique. Cette translocation est un processus qui dépend de la température, du potentiel transmembranaire des cellules-cibles [82], et de la concentration en calcium. AC requiert d'une forte concentration extracellulaire d'ions  $\text{Ca}^{2+}$  de l'ordre du mM (au moins supérieure à 0.2 mM) [83; 84; 78]. Le domaine RD est probablement directement impliqué dans ce mécanisme [75; 85], étant donné que cette région possède les principaux sites qui ont une faible affinité de liaison pour le calcium, et qui subissent des changements conformationnels dus à la liaison du calcium [85; 86]. La température joue également un rôle non négligeable. Alors que l'insertion de CyaA dans la membrane peut se dérouler entre 4 et 37°C, la translocation, elle, n'a lieu qu'à une température d'environ 20°C [78; 87]. La source d'énergie nécessaire au processus de translocation est à ce jour inconnue, mais il a été démontré [82] qu'un gradient électrochimique à travers la membrane plasmique pourrait en être à l'origine dans les myocytes cardiaques. D'autre part, il a été montré [88] que AC pouvait se déstructurer partiellement même à des températures physiologiques, ce qui favoriserait son insertion dans la membrane.

Cependant, des études [89; 90] ont montré que CyaA peut également être endocytée dans les cellules exprimant le récepteur CD11b/CD18. La découverte du mécanisme de translocation directe de AC dans le cytosol de la cellule hôte a enrichi l'imagination des chercheurs, et leur a permis de détourner les propriétés de CyaA pour en faire un outil biotechnologique, dont les aspects sont décrits dans la section suivante.

### 2.3 CyaA, un outil biotechnologique

De nombreuses applications biologiques ont été développées à partir des propriétés biologiques de CyaA durant ces deux dernières décennies. Par exemple, CyaA peut être utilisée comme système vecteur d'antigène pour le développement de vaccins anticancéreux, mais aussi en tant que méthode de criblage efficace pour cibler les protéines, et en tant que système double-hybride pour étudier les interactions protéine-protéine.

#### – Le système vecteur d'antigène :

Les lymphocytes T Cytotoxiques (CTL) représentent le mécanisme de défense majeure en réponse à une variété d'infections virales, bactériennes ou parasitaires. Ces cellules éliminent les

cellules infectées en reconnaissant un peptide endogène de très petite taille, appelé antigène, et associé à des molécules du Complexe Majeur d'Histocompatibilité (CMH) de classe I à la surface de la cellule. Il existe également des épitopes associés à des CMH de classe II qui stimulent des lymphocytes T-CD4<sup>+</sup> auxiliaires, importants dans la stimulation de réponses T cytotoxiques efficaces.

De nombreux antigènes de tumeur sont reconnus par les lymphocytes T CD4<sup>+</sup> (CD4) ou T CD8<sup>+</sup> (CTL), aussi bien des tumeurs murines qu'humaines [91]. CyaA ayant la propriété de délivrer son domaine catalytique à travers la membrane cytoplasmique des cellules eucaryotes, l'idée est donc d'utiliser CyaA comme transporteur d'antigènes d'intérêt (un épitope spécifique) dans le cytoplasme des cellules présentatrices d'antigènes (APCs) afin d'induire des réponses immunitaires en stimulant les CTL ou les CD4.

Le principe consiste à greffer un antigène (épitope spécifique) dans un site permissif au sein du domaine catalytique AC de CyaA, de manière à ne pas modifier sa structure et/ou son activité catalytique [92] (fig II.8). Par exemple, Fayolle et al [93], ont introduit génétiquement un épitope T CD8<sup>+</sup> du virus VIH ou du virus de la chorioméningite lymphocytaire (LCMV) dans un même site permissif du domaine catalytique, pour fabriquer des toxines recombinantes capables de stimuler chez des souris de fortes réponses T cytotoxiques (CTL) spécifiques de l'épitope inséré (figure II.8 gauche). Ces travaux ont démontré pour la première fois qu'une toxine recombinante portant des épitopes exogènes peut induire *in vivo* des réponses CTL spécifiques. Dans une autre étude [94], un épitope CD8<sup>+</sup> OVA (LCMV) a également induit de fortes réponses CTL, capables de lyser des cellules tumorales exprimant l'antigène ovalbumine. En outre, l'utilisation de CyaA recombinante confère également une immunité antitumorale *thérapeutique* efficace [94].

La sélectivité de CyaA pour les cellules dendritiques *in vivo* peut aussi être exploitée pour délivrer des épitopes associés au CMH de classe II, et stimuler ainsi les lymphocytes T-CD4<sup>+</sup> auxiliaires. En utilisant une CyaA recombinante portant à la fois un épitope de classe I, comme OVA, et un épitope de classe II (MalE), Schlecht et al [90] ont montré que ces antigènes pouvaient soit être transférés dans le cytoplasme pour rejoindre la voie de présentation des épitopes de classe I (figure II.8 gauche), soit être endocytés pour rejoindre la voie de présentation des épitopes de classe II (figure II.8 droite).

Les résultats accumulés durant les dix dernières années ont pleinement confirmé le potentiel de la molécule CyaA en vaccinologie et des phases d'essais cliniques d'immunothérapie anticancéreuse sont actuellement en cours. Par exemple, la protéine CyaA recombinante a déjà été expérimentée en tant que vaccin anticancer via un épitope spécifique de mélanomes, et même en temps que vaccin thérapeutique contre les papillomavirus humains [96].

#### – **Système rapporteur de translocation de protéine :**

Le domaine catalytique AC de CyaA est fortement activé par la calmoduline dans les cellules eucaryotes exclusivement. De ce fait, il est possible d'utiliser AC comme un rapporteur sélectif d'une translocation de protéine d'une bactérie à des cellules eucaryotes. La majorité des bactéries délivrent des protéines dans les cellules hôtes dans le but d'exercer leur fonction pathologique.

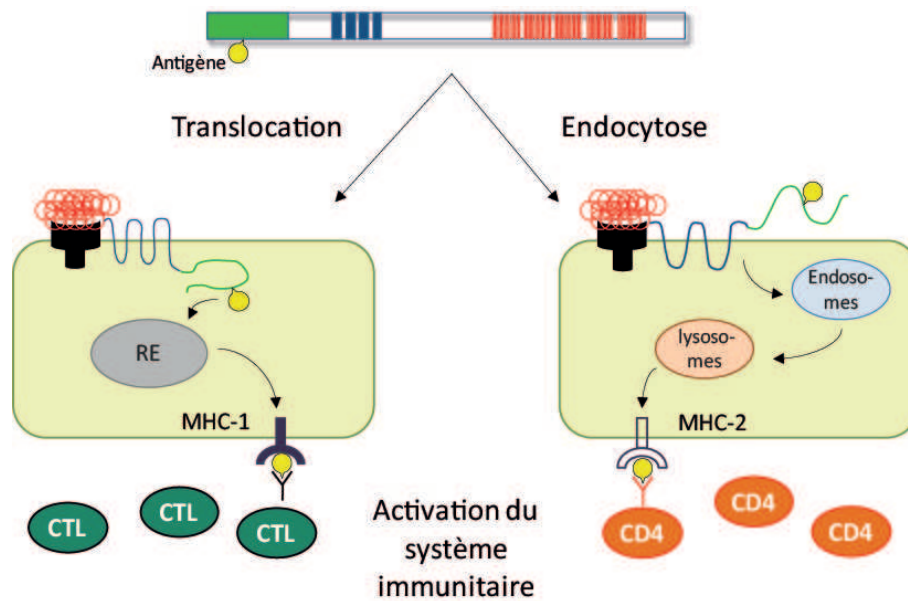


FIG. II.8: *CyaA*, système vecteur d'antigène : Une protéine CyaA recombinante portant un antigène génétiquement inséré dans le domaine catalytique, se fixe de manière sélective sur le récepteur intégrine CD11b/CD18 des cellules dendritiques (CD). Après translocation à travers la membrane plasmique des CD (gauche), l'antigène est dégradé en peptides par le protéasome dans le réticulum endoplasmique (RE), qui peuvent alors s'engager dans la voie de présentation cytoplasmique des épitopes associés aux CMH-I, pour stimuler les lymphocytes T cytotoxiques CD8<sup>+</sup> (CTL). Alternativement, la protéine CyaA recombinante est endocytée *via* le récepteur CD11b/CD18 (droite). Après protéolyse dans les endosomes/lysosomes, les peptides générés, et notamment les épitopes CD4<sup>+</sup> dérivés de l'antigène, peuvent s'associer aux CMH de classe II, pour stimuler des lymphocytes T auxiliaires (Th ou CD4). Adaptation d'après Ladant, 2006[95].

Parmi de nombreux systèmes de sécrétion, la machine de sécrétion de type III est utilisée par beaucoup de pathogènes pour injecter leur facteurs de virulence dans le cytosol de la cellule hôte, *via* un contact direct entre la surface de la bactérie et la surface de la cellule cible.

Cornelis et ses collaborateurs ont été les premiers à utiliser ce système AC-rapporteur [97]. Ils ont construit des protéines fusionnées avec Yops (*Yersinia outer membrane protein*) et le domaine catalytique AC : les enzymes hybrides Yop<sup>+</sup>-AC n'ont pas montré d'activité de AC dans *Yersinia*, à cause de l'absence de calmoduline. Une fois transféré dans le cytoplasme des cellules eucaryotes par le mécanisme de sécrétion de type III, l'enzyme hybride a été activé par la calmoduline de l'hôte et a synthétisé de l'AMPc, qui a pu être mesuré. L'accumulation d'AMPc dans les cellules cibles reflète donc directement la translocation Yop<sup>+</sup>-AC<sup>+</sup>.

#### – Le système double-hybride :

A partir du domaine catalytique AC de CyaA, Daniel Ladant et ses collaborateurs [98; 42] ont conçu un système double hybride bactérien qui permet de cribler et de sélectionner des interactions entre deux protéines. Ce système est basé sur la complémentarité entre deux fragments de AC : T25 et T18 [99]. Exprimés indépendamment dans *E. coli*, ces deux domaines sont incapables d'interagir ensemble, et ne peuvent donc pas reconstituer l'enzyme active [98]. A l'inverse, lorsque les fragments T25 et T18 sont fusionnés avec les protéines X et Y (figure II.9) qui peuvent inter-

agir ensemble, l'association des protéines hybrides T25-X et T18-Y permet une complémentarité fonctionnelle entre T25 et T18, et donc une restauration de l'activité de l'enzyme AC [42; 100].

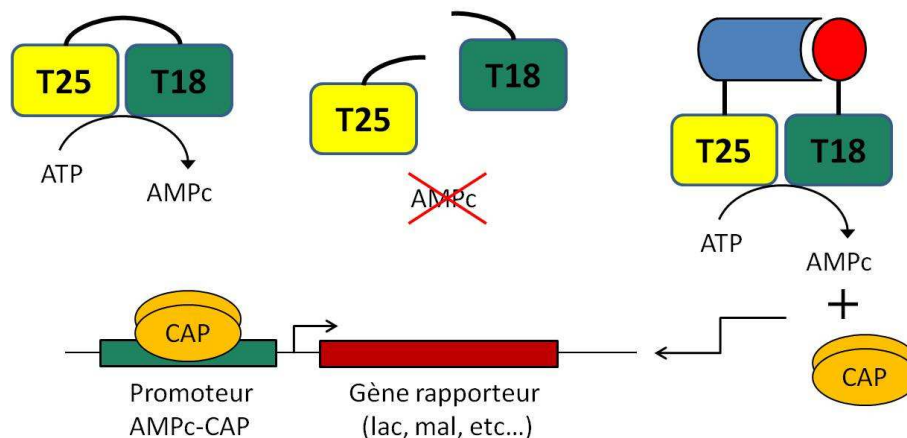


FIG. II.9: CyaA, système double-hybride, d'après Ladant [42].

Cette activation est aisément détectable par la synthèse d'AMPc basale en l'absence de calmoduline. L'AMPc est une molécule de signalisation capable de se lier au facteur de transcription catabolique CAP. Lorsque l'AMPc est synthétisé, il se lie à CAP et le complexe peut contrôler l'expression d'un grand nombre de gènes, appelés gènes cataboliques comme les opérons *lac* et *mal*, impliqués dans le catabolisme du lactose et du maltose. Par conséquent, les bactéries démontrent la capacité d'utiliser comme seule source de carbone, le lactose et le maltose, et peuvent facilement être distinguables sur un milieu sélectif.

## 2.4 Connaissances structurales et cinétiques

La structure du domaine catalytique AC de CyaA isolé, c'est-à-dire non liée à la calmoduline, demeure inconnue à ce jour. En effet, il n'y a pas de conditions de cristallisation publiées pour le domaine AC isolé. En effet, une légère agrégation du domaine AC est observée en solution et perturbe les mesures de SAXS (Alexandre Chenal, communication personnelle), et cette tendance à l'agrégation peut rendre difficile l'obtention de conditions de cristallisation pour AC. Par ailleurs, AC est trop grande pour être étudiée par RMN. Seule la structure du complexe AC/CaM a été résolue par cristallographie aux rayons X, mais uniquement avec le lobe C-CaM de la calmoduline chargé en ions  $\text{Ca}^{2+}$ .

Des expériences de transfert d'énergie par résonance en fluorescence (FRET) [101], des mesures de "cross-linking", des mutations et des amarrages protéine-protéine ("protein-protein docking") ont été réalisées par Guo et collaborateurs [102] afin de prédire la position du domaine N-terminal dans le complexe AC/CaM.

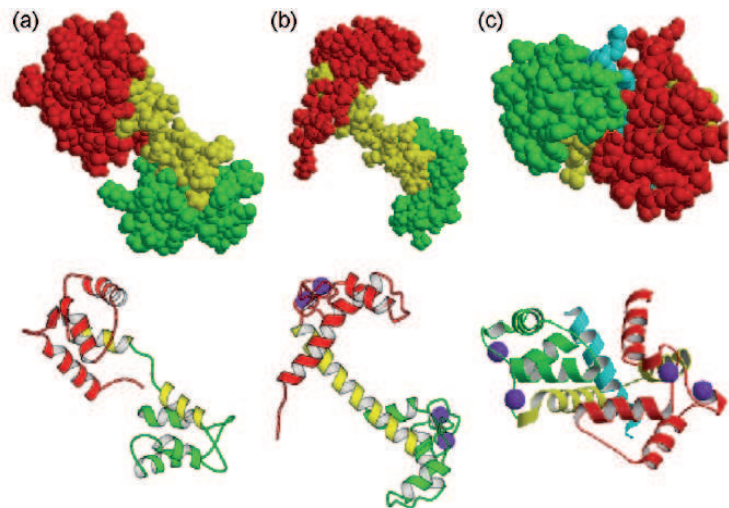
Le coeur catalytique de CyaA diffère des adénylcyclases des mammifères, et présente une homologie avec le facteur œdémateux (EF) de *Bacillus anthracis*.

## La calmoduline et le calcium

La calmoduline (CaM) est une protéine ubiquitaire des cellules eucaryotes. Senseur des ions calcium [103; 104; 105; 106], elle joue un rôle crucial dans de nombreuses voies de signalisation du calcium, mais aussi dans la survie, la croissance, la prolifération et la différenciation cellulaire.

La calmoduline, constituée de 148 résidus, est divisée en 2 domaines globulaires : N-terminal (Nter, résidus 5-64) et C-terminal (Cter, résidus 93-147), reliés par un “linker“ central flexible [107] (figure II.10 a)). CaM est le membre le plus représentatif de la super-famille des protéines à main-EF. Chaque domaine est composé d'une paire de sites de fixation au calcium de la forme hélice-boucle-hélice, dit motif "main-EF". Le lobe C-CaM présente une affinité plus élevée pour le calcium que le lobe N-CaM. La fixation d'ions  $\text{Ca}^{2+}$  sur chacun des deux domaines induit une transition conformationnelle d'un état fermé (inactif) de ces mains EF vers un état ouvert (actif) où des régions hydrophobes sont exposés au solvant, ce qui permet une reconnaissance moléculaire des cibles de CaM [108].

**FIG. II.10: Représentation schématique des trois conformations structurales de la calmoduline** (a) Structure de l'apo-CaM, entrée PDB : 1CFD [109]. (b) Structure de Ca-CaM, entrée PDB : 1CCL [110]. (c) Structure du complexe CaM/peptide smMLCK, entrée PDB : 1CDL [111]. Chaque structure est représentée en mode van der Waals (au-dessus) et ruban (en-dessous). Les domaines N-terminal, C-terminal sont colorés en rouge et vert respectivement. Le linker central est coloré en jaune, les ions calcium en violet, et le peptide smMLCK en cyan. Cette figure est extraite de Yang [108].



Sa plasticité remarquable fait d'elle un bon sujet d'étude des implications de la flexibilité structurale des protéines [112; 113; 114; 115].

La première structure de la calmoduline, réalisée par cristallographie aux rayons X [116], a montré une structure étendue en forme “d’haltère”. Les domaines N et C terminal, contenant chacun deux mains EF, sont les poids de l’haltère, l’hélice  $\alpha$  centrale formant la tige (Figure II.10b).

Des structures suivantes, déterminées par cristallographie aux rayons X [111] et par RMN [117], ont montré qu’en interaction avec un peptide cible, la calmoduline adoptait une structure compacte, où les deux lobes N et C terminal viennent se placer de part et d’autre du peptide, l’hélice central étant cassée au milieu pour permettre la pliure (Figure II.10 c). La structure en solution adopte une forme où l’hélice  $\alpha$  centrale est en partie désordonnée ce qui permet des mouvements considérables des deux lobes [118; 110; 119; 120].

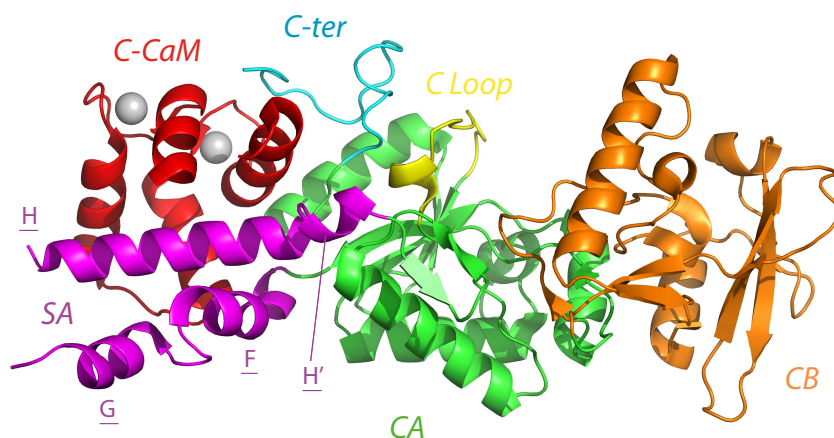
Par ailleurs, des études structurales ont montré que l’activation de la CaM par la fixation

des  $\text{Ca}^{2+}$  est accompagnée par l'ouverture des mains EF de chaque lobe [121]. Cette ouverture permet d'exposer des régions hydrophobes qui sont alors disponibles pour interagir avec des peptides cibles repliés sous forme d'hélice  $\alpha$ . Ainsi, en combinant la flexibilité de l'hélice centrale avec les déformations des régions hydrophobes, CaM peut interagir avec une dizaine de variétés de fragments peptidiques (soit plus de 100 protéines) avec une forte affinité malgré la variabilité de séquence de ces peptides [122; 108; 123; 124]. La dynamique conformationnelle joue un rôle prédominant à la fois dans la liaison du  $\text{Ca}^{2+}$  et dans la reconnaissance de la cible qui illustre l'importance de la plasticité structurale dans tous les aspects de la fonction de CaM.

Cependant, la résolution plus récente de structures de complexes impliquant la calmoduline [125; 126; 127], a révélée de nouveaux modes d'interactions inattendus entre la CaM et sa cible. Parmi ces structures, celle de la CaM complexée avec le facteur œdémateux (EF) [128], un homologue de CyaA, a révélé une conformation inattendue de la CaM, qui se présente alors sous sa forme étendue. D'autres structures [129; 130] décrivent également cette tendance.

### Le complexe AC/CaM

Quatre structures du complexe entre le domaine catalytique AC et CaM ont été obtenues par cristallographie à rayons X par Guo (2005) [131]. Elles sont disponibles dans la "RCSB Protein Data Bank" (PDB) [8] sous les entrées PDB 1YRT, 1YRU, 1ZOT et 2COL. La structure 1YRT, qui a été résolue avec une résolution de 2.10 Å, est représentée sur la figure II.11.



**FIG. II.11:** Structure cristallographique par rayons X du domaine catalytique (AC) de l'adényl cyclase CyaA (code PDB 1YRT) représentée en ruban. Les hélices F, G, H et H' colorées en violet forment le domaine SA. La boucle 226-232 reliant les hélices G et H à l'extrémité gauche de SA, n'est pas observée sur la densité électronique. Les autres régions de la protéine sont colorées en vert (CA), orange (CB), cyan (la queue C-terminale) et jaune (boucle catalytique : C loop). Le lobe C-CaM, coloré en rouge, interagit avec SA et la queue C-terminale. Les ions calcium sont représentés par des sphères grises. La visualisation a été faite avec pymol.

Cette structure contient le domaine catalytique AC (résidus 7 à 364) de CyaA en complexe avec le lobe C-CaM de la calmoduline (résidus 79 à 147), chargé de deux ions  $\text{Ca}^{2+}$ . La calmoduline

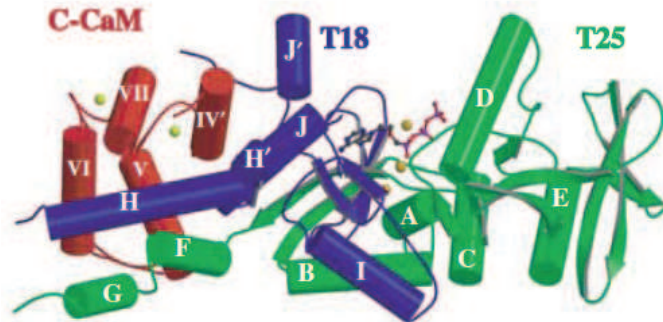
est insérée entre la région Switch A (SA) (résidus 198 à 260) et la région CA (résidus 1-55;181-191;255-358) qui inclut la région C-terminale (C-ter) (résidus 340-358). Les résidus 226 à 232 formant une boucle à l'extrémité du switch A (résidus 198 à 260) n'ont pas pu être observés sur la carte de la densité électronique. Le site catalytique se situe à l'interface de la région CA (résidus 1-55;181-191;255-358), et de la région CB (résidus 56 à 180). La boucle catalytique (C Loop) (résidus 294-306) est comprise dans la région CA. [131].

Les régions de contact de AC avec la calmoduline incluent trois segments : l'hélice  $\alpha$  F (H197–R206), l'hélice  $\alpha$  H (R235–R246), l'hélice  $\alpha$  H0 avec la boucle la précédant (R250–R259), ainsi que l'extrémité C-terminal (E346–R360). Ces observations structurales sont en accord avec l'observation que le segment de P196 à M267 contribue à environ 90% de l'énergie de liaison à la calmoduline [132; 131].

Les trois autres structures sont pratiquement identiques à celle-ci. La structure de l'entrée PDB 1YRU, a été résolue avec une résolution de 2.50 Å, dans les mêmes conditions mais en excès de chlorure de calcium (concentration 1mM). Pour obtenir un meilleur aperçu de la liaison de l'ATP et de la catalyse par CyaA, les structures 1ZOT et 2COL ont été résolues avec une résolution de 2.20 Å en présence d'ions  $\text{Ca}^{2+}$ , d'ions  $\text{Mg}^{2+}$  et d'adéfovir diphosphate pour la première, et d'ions  $\text{Ca}^{2+}$ , d'ions  $\text{Mg}^{2+}$  et de pyrophosphate (PPi) pour la deuxième. L'adéfovir diphosphate est un métabolite actif d'un médicament antiviral, l'adéfovir dipivoxil. Il mime l'ATP et inhibe de manière compétitive la catalyse de CyaA avec une forte affinité (25nM  $K_i$ ) [133].

L'architecture de AC peut également être représenté comme l'assemblage de deux domaines, T18 (fragment C-terminal de 18 KDa) et T25 (fragment N-terminal de 25 KDa) [134] (figure II.12). Les deux fragments peuvent être clivés par une trypsine et tous deux interagissent avec la C-CaM, bien que T18 établit plus de contacts (figure II.12). Cette constatation est en accord avec le fait que T18 est le domaine primaire de liaison de la CaM [134]. Les deux domaines sont indispensables pour la catalyse.

**FIG. II.12:** *Structure cristallographique par rayons X du domaine catalytique AC de CyaA en présence d'adéfovir diphosphate.* Entrée PDB : 1ZOT. Les domaines T25 et T18 sont colorés en vert et violet respectivement. C-CaM est représentée en rouge. Les ions calcium sont colorés en jaune et les ions magnésium en orange. Les atomes de carbone, d'oxygène, de nitrogène et de phosphore de l'adéfovir diphosphate sont respectivement colorés en gris, rouge, bleu, et violet. Figure extraite de Guo, 2005 [131].



Les structures de AC sont représentatives de l'activité catalytique de AC car il a été démontré [39] que la C-CaM est capable d'activer complètement le domaine AC. AC peut également être complètement activé par N-CaM, mais les concentrations nécessaires de cette dernière ne sont pas physiologiques. De plus, chaque lobe séparé a une affinité réduite pour AC, comparé à la

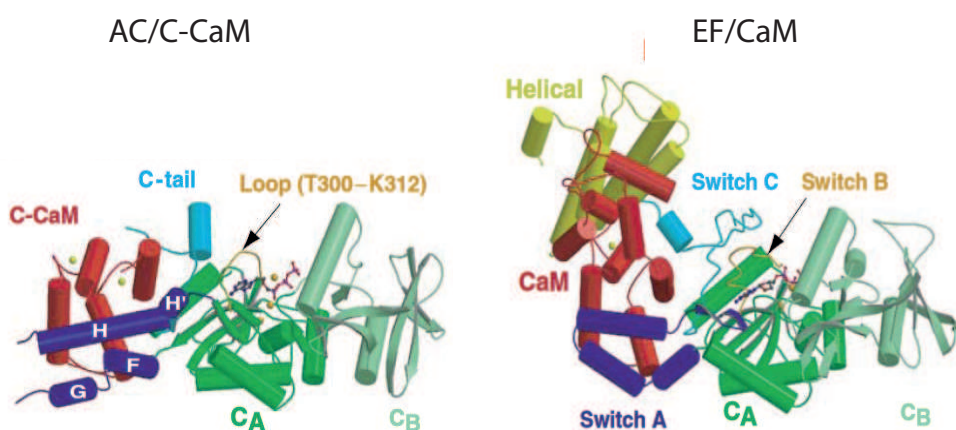
CaM intacte.

#### – Interaction C-CaM/AC

Les surfaces de AC chargées positivement permettent de lier la C-CaM, protéine acide avec un point isoélectrique (pI) de 4.1. Quatre régions de AC sont impliquées dans cette liaison : l'hélice F (H197-R206), H (R235-R246), H'(R250-R259) et la queue C-terminale (E346-R360). Ceci est en accord avec les résultats indiquant que le segment P196-M267 contribue à environ 90% de l'énergie de liaison à CaM [132].

Quelques résidus clés de l'interaction C-CaM/AC ont été identifiés. Parmi eux, le tryptophane 242 situé sur l'hélice H de AC joue un rôle important dans l'activation de AC par C-CaM, et quatre méthionines (M109, M124, M144, M145) de C-CaM sont importantes pour l'affinité de C-CaM pour AC [135; 136]. En-dehors de l'hélice H, trois autres régions de AC interviennent également. Elles comprennent les ponts salins formés entre les résidus R250, R258 et R259 sur l'hélice H', et les résidus E84 et E87 de C-CaM. Elles comprennent également l'interaction des résidus de la queue C-terminale de AC, E346, R348, D360, avec E87 et R90 de C-CaM. De plus, l'hélice F établit aussi des contacts avec C-CaM. H197 de AC forme une liaison hydrogène avec E114 de C-CaM, et R206 de AC avec la chaîne principale de G113, qui forme aussi un pont d'eau avec E114.

#### Comparaison avec le complexe EF/CaM



**FIG. II.13: Comparaison entre AC/C-CaM et EF/CaM** Structure secondaire de AC/CaM (à droite) et de EF/CaM (à gauche). Les domaines CA et CB des deux complexes sont colorés en vert et vert-clair respectivement. C-CaM et CaM sont colorés en rouge. Le domaine hélicoïdal de EF est représenté en jaune. Les trois régions switch, switch A, switch B et switch C de EF sont colorés en violet, orange et cyan respectivement. Les régions hélices F-H', loop T300-K312 et la queue C-terminale de AC, qui correspondent aux trois switches de EF sont colorés en violet, orange et cyan respectivement. Les ions calcium sont colorés en jaune et les ions magnésium en orange. Les atomes de carbone, d'oxygène, de nitrogène et de phosphore de la molécule adéfovire diphosphate (AC) et 3'dATP (EF) sont colorés en gris, rouge, bleu et violet respectivement. Figure extraite de Guo et al., 2005 [131].

Le facteur œdémateux (EF) est l'exotoxine de *B. anthracis* structurellement homologue à AC de CyaA, mais ne partageant que 25% d'identité de séquence. Elle est également activée par la



calmoduline pour catalyser l'hydrolyse de l'ATP en AMPc et pyrophosphate. Elle est divisée en plusieurs domaines similaires à ceux de AC (figure II.13) qui comprend les domaines CA et CB, dont l'interface constitue le site catalytique, et les domaines hélicoïdaux [128; 137]. Dans AC, les structures des domaines CA (résidus 1-61 et 187-363) et CB (résidus 62-186) sont proches de celles des domaines correspondants de EF avec des écarts quadratiques moyens respectifs de 1.17 et 1.08 Å. Cependant la distribution des charges à la surface de ces domaines diffère entre AC et EF.

Les switches C, B, et A du facteur œdémateux correspondent respectivement à (figure II.13) la queue C-terminale (cyan), la boucle catalytique (jaune) et les hélices F et H de SA (violet) [131; 128]. Comme ces switches subissent des changements conformationnels importants pendant l'activation du facteur œdémateux, on peut se demander si leurs homologues dans CyaA subissent également des réarrangements substantiels pendant l'activation du domaine AC.

#### – Comparaison entre les interactions C-CaM/AC et CaM/EF

Deux régions de liaison à C-CaM, les hélices F et H de CyaA, sont positionnées de manière similaire aux hélices de la région switch A de EF. Cependant, les deux autres régions de CyaA qui font des contacts avec CaM sont différentes de celles de EF. La queue C-terminale de CyaA se différencie d'un point de vue structural du switch C correspondant de EF. CyaA possède une hélice de plus, l'hélice H', qui interagit également avec la C-CaM, et qui est remplacée par un "domaine" hélicoïdal se liant à l'extrémité N-CaM [128; 137]. Les comparaisons de séquence révèlent que les résidus liant la C-CaM chez CyaA ne sont pas conservés avec les résidus correspondant chez EF. Les interactions clés avec la CaM divergent entre les deux protéines. Par exemple, le résidu W242 de AC joue un rôle vital dans l'interaction avec la C-CaM, tandis que le résidu correspondant dans EF, la glutamine 526 n'a qu'un rôle mineur [128].

CyaA a une affinité environ 100 fois plus forte pour la CaM que EF, bien que sa surface de contact avec la C-CaM ( $3395 \text{ \AA}^2$ ) soit plus petite que celle de EF ( $3914 \text{ \AA}^2$ ) [39]. De plus, CyaA ne possède pas le domaine hélicoïdal correspondant de EF, qui est un domaine interagissant avec la CaM et vital pour l'activation de EF par celle-ci. Néanmoins, AC possède une complémentarité de surface avec C-CaM beaucoup plus significative que EF, ce qui explique sa forte affinité pour la C-CaM.

Contrairement à EF/CaM, la calmoduline entière n'a pas besoin de calcium pour activer AC (shen 2002 [39]). N-CaM et C-CaM peuvent activer complètement AC, bien que l'affinité mesurées par des valeurs de EC50, diminue. C-CaM peut activer AC de manière optimale dans des concentrations de calcium physiologiques ( $0.03\text{-}0.1 \mu\text{M}$ ). L'activation de AC par N-CaM demande une concentration élevée de calcium ( $0.1\text{-}1 \mu\text{M}$ ). L'activation par N-CaM dépend de la présence de calcium, mais pas celle par C-CaM.

#### – Propriétés catalytiques

CyaA et EF ont des propriétés catalytiques similaires. Leur affinité est équivalente pour l'ATP ( $0.2\text{-}1 \text{ mM}$ ) et les produits de la réaction ( $0.6\text{-}4 \text{ mM}$ ) (Guo et al, 2004)[138]. Elles partagent

également une constante catalytique  $k_{cat}$  de 1000-5000  $s^{-1}$  et une efficacité catalytique d'environ  $10^7 M^{-1}s^{-1}$ . Les produits de la réaction sont relargués aléatoirement, bien qu'il y ait une préférence pour le relargage de PPI en premier lieu [138].

Les deux adénylcyclases catalysent la réaction réversible  $ATP \rightleftharpoons cAMP + PPI + H^+$ . Pour AC, le taux d'avancement de la réaction inverse, synthétisant l'ATP à partir de l'AMPc et du PPI, est plus faible que le taux d'avancement de la réaction directe (l'hydrolyse de l'ATP) d'un facteur 13 [138]. A l'inverse, pour EF, ces mêmes réactions directes et inverses partagent un taux d'avancement similaire.

La structure de AC/CaM en complexe avec l'adéfovir diphosphate montre comment ce dernier se lie au site catalytique. L'adéfovir diphosphate forme un réseau de liaisons hydrogène pour lier les résidus cruciaux pour la catalyse, qui se situent à l'interface des domaines CA et CB. L'interaction de AC avec la base adénine et les triphosphates de l'adéfovir diphosphate est similaire à l'interaction de EF avec le 3'deoxy ATP, un analogue de l'ATP non cyclisable. Trois ions magnésium sont observés. Deux d'entre eux se trouvent à des endroits similaires à ceux présents dans le complexe EF/CaM/3'deoxy ATP, c'est-à-dire dans le site actif de EF. L'un est coordonné par deux résidus aspartate (D188 et D190) et une histidine (H298), l'autre par des phosphates  $\alpha$  et  $\beta$  des moitiés triphosphates. La catalyse est supposée utiliser deux ions Magnesium [137] bien que des simulations de dynamique moléculaire [139; 140] ont montré qu'un mécanisme à un seul ion pourrait être plausible.

Comparé à EF/CaM, la région CB de AC a une rotation de  $23^\circ$  par rapport à la région CA. De ce fait, les résidus catalytiques dans CB (résidus H63 et K84) se sont éloignés d'environ 2 Å de ceux dans CA (résidus K58, D188, D190, H298, N304, F306). La poche catalytique est donc plus ouverte dans AC que dans EF.

La structure de AC/CaM en complexe avec le pyrophosphate (PPI) révèle que la liaison du PPI dans le site actif est similaire à celle du diphosphate terminal de l'adéfovir diphosphate. Le résidu N304 de CA, qui est essentiel pour la liaison d'ions métalliques catalytiques et le positionnement du ribose, est pratiquement identique au résidu correspondant (N583) dans EF [137]. De même, les résidus de CB qui sont cruciaux pour la liaison du triphosphate de l'ATP (R37, K58, K65, K84) et la déprotonation de 3'OH (H63) sont bien alignés avec les résidus correspondants dans EF.

#### – Mécanisme d'activation catalytique par la CaM

Bien que les résidus de AC lient la CaM diffèrent considérablement de ceux de EF, le mécanisme d'activation reste semblable. Chez EF, les switches A, B et C subissent des transitions conformationnelles significatives [128]. Ces régions correspondent aux régions switch A, alias SA, la boucle catalytique (T300-K312) et la queue C-terminale de AC. L'étape clé de l'activation de EF par CaM est la stabilisation de la boucle catalytique, alias le switch B [128]. En effet, il a été déterminé que le résidu catalytique clé du switch B impliqué dans ce mécanisme est l'asparagine 583, car il permet de garder en place le ribose durant la catalyse. Pour AC, Guo et al. ont démontré que lorsque le résidu correspondant, l'asparagine 304, est muté en alanine, l'efficacité catalytique est diminuée d'un facteur 72, mais qu'en revanche son affinité pour la CaM

demeure inchangée. D'autre part, dans EF, la CaM induit l'interaction entre la boucle catalytique et l'ATP, grâce à la phénylalanine 586, qui correspond chez CyaA au résidu F306. La modification du résidu correspondant chez AC en alanine conduit à une protéine qui possède une efficacité catalytique et une affinité pour CaM équivalente à la protéine sauvage. L'alignement structural des deux protéines montre que ces deux résidus sont positionnés de la même façon. Ces deux informations réunies permettent de confirmer l'hypothèse selon laquelle CaM permet la stabilisation de la boucle catalytique pour l'avancement de la catalyse.

L'interaction entre la boucle catalytique (switch B) et le switch C est une étape clé dans l'activation catalytique de EF par CaM [128]. La région correspondante dans AC (queue C-terminale) ne partage pas d'homologie de structure avec le switch C de EF, mais elle le mime dans les contacts essentiels formés entre C-CaM et la boucle catalytique de AC. Le mouvement concerté de la queue C-terminale et des hélices F et H' de AC lorsque la CaM se lie, joue un rôle clé dans la stabilisation de la boucle T300-K312 pour l'activation catalytique.

L'activation catalytique de EF par CaM commence par les interactions coordonnées du domaine hélical de EF avec la N-CaM, et celui du switch A avec C-CaM [128; 141]. Les interactions des hélices F, H et H' de AC avec C-CaM miment l'interaction du switch A de EF avec la CaM. Tandis que N-CaM peut activer à lui seul AC, il n'y a pas de segment dans AC qui pourrait mimer le domaine hélical de EF [39]. Le rôle de ce domaine hélical est de bloquer EF dans un état inactivé [39], ce qui explique pourquoi EF ne possède pas d'activité basale, contrairement à AC.

## 2.5 Carte des mutants de AC

L'adénylcyase AC de CyaA est divisée en deux segments, le segment N-terminal T25 (résidus 1-224) et le segment C-terminal T18 (résidus 225-399) [134]. Les deux segments sont impliqués dans la liaison avec la CaM mais T18 est le segment qui forme le plus de contacts. Le site catalytique se situe entre T18 et T25 [138]. Chaque segment porte un résidu tryptophane, W242 étant situé dans le segment T18 et W69 dans le segment T25.

Des études précédentes ont montré que le "cœur" du site de liaison de AC à la CaM est situé autour du résidu tryptophane 242 [44; 142; 143; 144]. Ainsi, les premières expériences de mutagenèse dirigée se concentraient sur le résidu W242.

### – Mutations de résidus dans la région de contact avec la CaM.

Glaser et al. (1989) [135] ont construit les quatre mutants W242V, W242R, W242G, W242D. En effet, W242 se situe au milieu de l'hélice  $\alpha$  H, et les résidus aspartate et glycine ont tendance à déstructurer les hélices, contrairement aux résidus tryptophane et valine. De plus, Cox et al (1985) [145] avaient déjà suggéré qu'une hélice  $\alpha$  est l'élément d'interaction avec la CaM essentiel des adénylcyases bactériennes et d'autres protéines interagissant avec elle, chose qui a été confirmée par de nombreuses études des interactions peptide-CaM [111; 117]. Et la séquence autour du résidu W242 a des caractéristiques communes aux peptides liant la calmoduline, comme par exemple la présence de résidus basiques et hydrophobes formant une

hélice  $\alpha$  amphiphilique [146; 135]. Glaser et al. ont démontré que ces quatre mutants expriment des activités considérablement différentes à une concentration de CaM qui est saturante pour l'enzyme sauvage (10 nM). Le mutant W242V est le moins affecté, comme attendu, il conserve 90% de l'activité maximale du WT. Les mutants W242D et W242G sont les plus affectés dans leur activation par la CaM et dans leur activité. Le mutant W242D, le plus affecté, ne présente que 6.6% de l'activité normale, et requiert 350 nM de CaM pour être activé contre 0.20 nM pour le "wild-type". Le mutant W242G a une sensibilité pour la CaM réduite d'un facteur 1000. En revanche, l'affinité pour l'ATP n'a changé pour aucun des quatre mutants.

Guo et al. (2005) [131] ont construit également le mutant W242G afin d'étudier les effets des lobes N-CaM et C-CaM séparément sur l'activité de AC. Il en résulte que le mutant a une sensibilité réduite aussi bien pour la C-CaM que pour la N-CaM, ce qui suggère que CyaA peut interagir aussi bien avec N- ou C-CaM pour former des complexes hétérogènes. D'autre part, le mutant W242G a une activation par la CaM drastiquement réduite [135].

Le mutant W242V fut également étudié par Gilles et al. [142] et Bouhss et al [132].

Guo et al. ont pu définir, grâce à la structure cristallographique du complexe AC/C-CaM qu'ils ont résolue, d'autres régions étant en contact avec la C-CaM : l'hélice F (H197–R206), l'hélice H (R235–R246) et l'hélice H' avec la boucle qui la précède (R250–R259), la queue C-terminale (E346–R360). Ces régions sont en accord avec le fait que le segment P196 à M267 contribue à environ 90% de l'énergie de liaison à la CaM [132].

Ces auteurs ont donc construit des mutants indépendants qui contiennent des mutations au niveau de résidus compris dans ces régions. Ainsi, ils ont démontré que le mutant R206A (hélice F) adopte un comportement comparable à celui de AC sauvage, et que les trois autres mutants W242G (hélice H), R258A/R259A (Hélice H') et E346A/R348A (C-ter) ont une activation par la CaM réduite en présence de calcium. Par conséquent, leur efficacité catalytique est également réduite ( $k_{cat}/K_m$ ), ainsi que leur affinité pour la CaM, bien que cela soit moins apparent à des concentrations de  $Ca^{2+}$  élevées.

#### – Mutations dans le site catalytique.

Avant la détermination de structure de AC, il avait été remarqué [44] que le domaine T25 isolé présentait une activité enzymatique faible mais détectable, ce qui indiquait qu'il est porteur du site catalytique. Très tôt, des études de mutagenèse dirigée ont donc ciblé des résidus de cette région.

Les premiers mutants de cette région, K58Q et K65Q, ont été construits par Glaser et al. (1989) [135], car le site de liaison à l'ATP était supposé être localisé autour de ces résidus. L'activité catalytique de ces deux mutants a été drastiquement réduite pour des conditions de CaM saturante, comparé à l'activité de AC sauvage. En effet, K65Q (la plus active des deux) a une activité de 1.4  $\mu\text{mol}/\text{min.mg}$  de protéine contre 2540  $\mu\text{mol}/\text{min.mg}$  de protéines pour AC sauvage. Les propriétés de leur liaison à la CaM n'ont pas été modifiées.

T18 est le segment majeur de liaison à la calmoduline, mais les deux fragments sont nécessaires pour restaurer une activation par la CaM et une activité catalytique normale [134; 143]. Bouhss et al. (1993) [132] ont démontré que l'intersection entre les segments T18 et T25 d'une longueur de 72 acides aminés, est responsable de 90% de l'énergie de liaison du complexe AC/CaM, grâce aux mutants [132] : R235E, L241E, K243E, W242V (tous localisés dans l'hélice H), R221Q, R223Q, R224E (tous situés dans l'hélice G), R164E, R165E, K166E, W69V, P184W (CB). De plus, ces auteurs ont démontré que la catalyse et la liaison de la CaM sont dramatiquement altérées, et que le résidu W69 est important pour la stabilisation de la structure tertiaire de la protéine *via* des interactions de Van der Waals.

Bouhss et al. ont également construit le mutant W69F (CB) en 1996 [147], qui ne conserve que 40 % de l'activité de AC sauvage, et qui perd 9 % d'énergie de liaison à CaM.

D'autre part, sur les neuf résidus histidine localisées au niveau du site catalytique de *Bordetella pertussis*, trois (H63, H106 et H298) sont conservés dans l'adénylcyclase de *Bordetella anthracis*. Une série de mutations sur deux de ces trois histidines a été réalisée par Munier et al. (1992) [148] : H63Q, H63R, H63E, H63V, H106R, H106Q, H106N. Ces modifications de AC ont conduit à une diminution de l'activité catalytique, montrant ainsi l'importance de ces deux histidines pour une activité catalytique efficace. Cependant, la mutation du résidu H63 se révèle la plus critique, avec une perte d'activité d'un facteur 2000 à 3000, tandis que l'affinité pour la CaM n'est pas altérée.

#### – Mutations de résidus de AC susceptibles de lier N-CaM

Récemment, Guo et al. ont prédit la position du lobe N-CaM par des expériences d'amarrage protéine-protéine ("*Docking*") et de simulations de dynamique moléculaire, et ont testé leurs hypothèses par des expériences de mutagenèse dirigée, de "pontage" ("*cross-linking*") et de transfert d'énergie entre molécules fluorescentes (FRET) [102].

Leur modèle les conduisit à supposer que les résidus acides de N-CaM forment des ponts salins avec les résidus basiques de AC R223, R224 et R235. D'après ce modèle, la mutation de ces trois résidus en alanine devrait affecter l'activation de AC par la CaM, mais pas par la C-CaM. Les tests furent négatifs, suggérant que N-CaM ne se lie pas à l'hélice G.

Une interaction entre l'épingle à cheveux  $\beta$  derrière l'hélice H a également été envisagée. Les mutants E256A, Q260A, R262A ont une sensibilité d'activation par la CaM réduite mais pas par la C-CaM, ce qui suggère que N-CaM se lie bien à cette région.

D'autre part, le domaine hélicoïdal C-terminal du facteur œdémateux (*Bordetella anthracis*) fait des contacts extensifs avec N-CaM et cette interaction sert à ancrer la CaM pour induire un changement conformationnel de cette toxine. Guo et al. ont émis l'hypothèse que le domaine C-terminal de AC (résidus 361 à 373) pourrait lier N-CaM. Deux mutants ayant la région 361-366 pour l'un et 367-372 pour l'autre tronquée ont été construits. La délétion de la région 361-366 montre une réduction de 50% de la sensibilité d'activation de l'activité catalytique de CyaA

par la CaM, contrairement à la région 367-372. En revanche, cette délétion n'a pas affecté la sensibilité d'activation par la C-CaM, qui est, comme attendue, environ 30 fois plus faible que CaM.

Dans d'autres modèles, l'épingle à cheveux  $\beta$  (résidus 255-264) ainsi que les résidus Q260 et R262 et/ou E256 interagissent avec N-CaM [102]. Des mutations au niveau des résidus de N-CaM qui interagissent avec ces régions confirme cette observation. Mais pour évaluer davantage ce modèle et estimer la proximité des résidus appariés, des expériences de "cross-linking" ont été réalisées en mutant les résidus A257, Q260, T263 de AC et Q41 et D50 de N-CaM en cystéine, dans l'optique de former des ponts disulfures. Cette expérience est rendue possible du fait que ni AC ni CaM ne possède de cystéine. Un pont disulfure a été observé entre les résidus Q260C de AC et D50C de CaM.

Cette même expérience reproduite en mutant les résidus A94 et A225 situés à l'extrémité opposée de CyaA et les résidus Q41 et T26 de CaM en cystéine n'a eu qu'un effet minimal sur l'activation de CyaA par la CaM.

### 3 Plan de l'étude

Les interactions entre bio-molécules régulent la plupart des mécanismes biologiques au sein de la cellule, tels la transduction du signal, la croissance, la différenciation et bien d'autres encore, comme il l'a été décrit au cours de ce premier chapitre. Beaucoup d'interactions protéine-protéine impliquées dans la signalisation cellulaire sont associées à des maladies spécifiques, dont les maladies immunitaires ou infectieuses. Dans le cadre des recherches thérapeutiques, les interactions protéine-protéine constituent de nouvelles cibles privilégiées pour la recherche de composés bioactifs agissant à leur interface, ou au niveau de sites allostériques. L'étude de ces interactions représente donc un enjeu majeur.

L'activation d'une protéine par son partenaire implique des changements conformationnels importants pour propager cette information. La flexibilité d'une protéine joue également un rôle essentiel dans son activation enzymatique. Une étude précédente sur le facteur oedémateux EF de *Bacillus anthracis* [149], responsable de la maladie de l'anthrax, a été menée afin de comprendre comment sa grande mobilité structurale et son interaction avec la calmoduline, protéine d'une flexibilité remarquable, joue un rôle crucial dans l'établissement de sa fonction enzymatique. Ces travaux ont orienté le choix de l'objet d'étude de cette thèse sur l'étude du complexe AC/CaM, qui possède 25% d'identité de séquence avec le complexe EF/CaM, mais qui est structuralement homologue à ce dernier. Cependant, les interactions sont différentes, à cause de l'absence d'un domaine hélicoïdal dans AC/CaM. D'autre part, CyaA est une protéine étudiée depuis plus de trente ans et bien connue des expérimentalistes de l'Institut. Cette thèse étant majoritairement portée sur des simulations de dynamique moléculaire, il est donc plus aisé de comparer les

résultats avec ceux de la littérature déjà très vaste, ainsi que de les valider par de nouvelles expériences au sein de l'Institut.

Cette thèse comporte trois parties. La première partie consiste en l'étude de la dynamique interne du domaine catalytique AC de l'adénylcyclase, et de la stabilisation de celui-ci par le lobe C-terminal de la calmoduline (C-CaM). L'influence de ce dernier, ainsi que des ions calcium sur la stabilité du complexe a été analysée à travers l'étude de la plasticité des deux partenaires, réalisée par dynamique moléculaire. Dans ce même chapitre, les aspects énergétiques de ces influences sur l'interaction AC-CaM ont également été caractérisés, et a permis l'identification d'un réseau d'interactions entre ces deux partenaires. Ainsi, des résidus clés ont été mis en évidence, pouvant être mis en relation avec la fonction enzymatique. Le rôle que jouent ces résidus clés sur le contrôle allostérique de l'activation de AC a été analysé dans le chapitre V, par des simulations de dynamique moléculaire classique incluant des mutations au niveau de ces résidus, mais également par des expériences de mutagenèse dirigée, que j'ai réalisées dans l'unité de Biochimie des interactions macromoléculaires de l'Institut Pasteur, dirigée par Daniel Ladant. Ces expériences ont permis de mettre en évidence une région intermédiaire clé, formée par ces résidus, qui assure la transmission de l'information de transition conformationnelle entre C-CaM et le site catalytique de AC. La dernière partie de cette thèse est consacrée à la prédiction d'une structure du domaine AC isolé se rapprochant des descriptions expérimentales connues. En effet, aucune structure tridimensionnelle à haute résolution de AC n'a été résolue, mais des mesures hydrodynamiques attestent qu'il se présente sous une forme compacte. Les simulations de dynamique moléculaire classique réalisées dans le premier chapitre n'ont pas permis d'observer une telle transition. Une approche récente et originale, appelée "Dynamique moléculaire accélérée par la température" (*Temperature Accelerated Molecular Dynamic (TAMD)*) [150], a donc été utilisée pour accélérer la dynamique moléculaire du domaine AC, dans le but de franchir les barrières énergétiques qui séparent les différents états de AC, et prédire ainsi une structure stable de AC à l'état non-lié.

## Chapitre III

# Présentation de quelques méthodes d'exploration de l'espace conformationnel des protéines par des simulations de dynamique moléculaire

---

## 1 La dynamique moléculaire

La dynamique moléculaire consiste à étudier, par le calcul numérique, l'évolution d'une molécule au cours du temps. Conçue à l'origine dans le cadre de la physique théorique, elle s'est maintenant étendue aux domaines des sciences des matériaux et des biomolécules. Elle permet ainsi d'explorer l'espace conformationnel d'une protéine à partir de sa structure, dont les coordonnées ont été résolues par des méthodes de biologie structurale (cristallographie aux rayons X, résonance magnétique nucléaire (RMN)), dans le but de comprendre la relation structure-dynamique-fonction, ou de déterminer des conformations inaccessibles par ces méthodes.

### 1.1 Principe de la dynamique moléculaire classique

Elle consiste à étudier la trajectoire d'une molécule en appliquant les lois de la mécanique classique newtonienne, c'est-à-dire à simuler les mouvements atomiques au cours du temps, en admettant un certain nombre d'approximations. Chaque atome  $i$  de la molécule est considéré comme une masse ponctuelle  $m_i$  dont le mouvement est déterminé par l'ensemble des forces  $\vec{F}_i$  exercées sur lui par les autres atomes en fonction du temps  $t$ . 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 suivante :

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



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

où  $\vec{a}_i$  est l'accélération de l'atome  $i$ .

Pour obtenir la trajectoire de chaque atome en fonction du temps, ces équations du mouvement sont intégrées en général par l'algorithme de Verlet [151], qui consiste à calculer à chaque pas de temps  $t$  :

- l'accélération instantanée  $\vec{a}_i$  de chaque atome  $i$  à partir des forces  $\vec{F}_i$  s'exerçant sur chacun d'entre eux, et de leur masse atomique  $m_i$ .
- la vitesse atomique  $\vec{v}_i$  à partir de l'accélération instantanée  $\vec{a}_i$ .
- la position atomique  $\vec{r}_i$  à partir de la vitesse atomique  $\vec{v}_i$ .

L'intégration de ces équations se fait en subdivisant la trajectoire en une série d'états discrets séparés par des intervalles de temps très courts dont la longueur définit le pas d'intégration  $\delta t$ . Dans le logiciel AMBER [152] avec lequel une partie des simulations présentées ici a été effectuée, l'équation du mouvement est résolue avec l'algorithme du "saut-de-mouton" (*leap-frog* [153]), une variante de l'algorithme de Verlet :

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

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

Les vitesses initiales sont générées de manière aléatoire selon la distribution de Maxwell-Boltzmann. L'intervalle de temps  $\delta t$  doit être plus petit que l'échelle de temps (1 fs) associée aux mouvements de plus haute fréquence du système, c'est-à-dire les vibrations des liaisons chimiques dans lesquelles un hydrogène participe. La stabilité numérique du système est ainsi assurée. Des algorithmes tels que SHAKE [154] ou RATTLE [154; 155] peuvent être appliqués pour utiliser un pas de temps  $\delta t$  plus grand (environ 2 fs). Ces algorithmes consistent à "geler" les vibrations de haute fréquence des liaisons de type X-H, qui ont, dans le cadre de notre étude, une influence très limitée sur le comportement global du système.

## 1.2 Champ de force

Le champ de force en modélisation moléculaire est l'ensemble des formes fonctionnelles et des paramètres utilisés pour calculer l'énergie potentielle du système dans un logiciel donné.

L'énergie totale du système a deux composantes qui sont l'énergie cinétique  $E_c = \sum_{i=1}^N \frac{1}{2} m_i v_i^2$  et l'énergie potentielle  $U$ . Cette dernière est calculée en fonction de la position exacte des atomes.

Elle est exprimée comme la somme de cinq types d'interactions :

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

Les trois premiers termes sont des termes dits “liés”, car ils correspondent respectivement aux énergies de compression ou d'élongation de la liaison covalente, de déformation angulaire et de déformation des angles de torsion (dièdres) entre atomes liés de manière covalente. Les deux derniers termes sont dits “non-liés” car ils correspondent aux contributions électrostatiques et de van der Waals entre atomes séparés de plus de deux liaisons.

Chacun de ces termes comporte des constantes qui sont définies par un champ de force empirique adapté en fonction du système étudié (ADN, protéine ...). Ils sont codés dans les programmes standards de modélisation moléculaire, tel que CHARMM [156], AMBER [152], NAMD [157] ou GROMACS [158]. Ces derniers diffèrent essentiellement au niveau de la paramétrisation des valeurs  $K_r$ ,  $r_{eq}$ ,  $\theta_{eq}$  et des autres paramètres cités dans l'équation.

#### Interactions entre atomes liés :

- Énergie de compression ou d'élongation des liaisons : elle est exprimée par un potentiel harmonique, où  $r$  est la longueur exprimée en Å entre une paire d'atomes liés. Lorsque cette longueur dévie de sa valeur de référence, (ou valeur d'équilibre),  $r_{eq}$ , cette énergie est donc augmentée.  $r_{eq}$  est une constante du champ de force choisi.  $K_r$  est une constante de force associée en kcal.mol<sup>-1</sup>.Å<sup>-2</sup>.
- Énergie de déformation des angles : ce terme est une somme des énergies d'élongation de chaque angle de valence  $\theta$  de la molécule, exprimées également par un potentiel harmonique par rapport à la valeur de référence  $\theta_{eq}$ .  $K_\theta$  est une constante de force associée en kcal.mol<sup>-1</sup>.deg<sup>-2</sup>. Les angles sont exprimés en °.
- Énergie de déformation des angles dièdres : elle est exprimée par un potentiel de torsion qui modélise la variation du changement d'énergie selon la liaison rotative autour d'une liaison. Elle est décrite par une série de Fourier et dépend de l'angle dièdre instantané  $\phi$  en degrés, d'un terme d'énergie  $V_m$  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 entre atomes non-liés :

- Interactions électrostatiques : La somme des interactions électrostatiques entre deux atomes

$i$  et  $j$  est évaluée par la loi de Coulomb :

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

où  $q_i$  et  $q_j$  sont les charges partielles portées par les atomes  $i$  et  $j$ ,  $r_{ij}$  la distance en Å entre ces atomes, et  $\epsilon_0$  la constante diélectrique (ou perméabilité diélectrique du vide) qui est égale à  $8,85418782 \times 10^{-12} \text{F.m}^{-1}$ . Cette énergie diminue lentement avec l'augmentation de la distance entre atomes ( $\sim 1/r$ ).

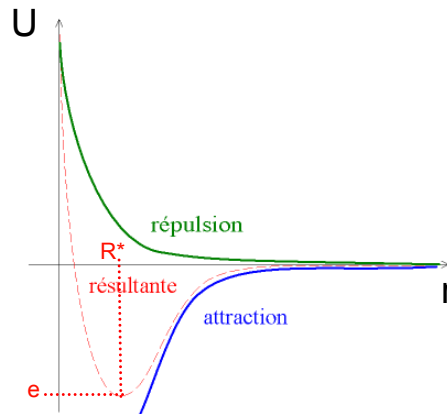
- Interactions de van der Waals : elles sont décrites par un potentiel de Lennard-Jones en 6-12 :

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

où  $r_{ij}$  est la distance entre les atomes  $i$  et  $j$  en Å  $R_{ij}^*$  et  $\epsilon_{ij}$  sont les paramètres de Lennard-Jones. Le paramètre en  $r_{ij}^{-12}$  correspond à la répulsion entre deux atomes dû au principe d'exclusion de Pauli (courbe verte de la figure III.1). Le terme en  $r_{ij}^{-6}$  représente la dispersion attractive de London (courbe bleue de la figure III.1).  $\epsilon_{ij}$  (notée  $e$  sur la figure III.1) correspond à la profondeur du puits de potentiel à la distance  $R_{ij}^*$ . Ces deux derniers paramètres sont estimés par la règle de Lorentz-Berthelot :

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad (\text{III.8})$$

$$R_{ij}^* = \frac{1}{2}(R_{ii}^* + R_{jj}^*) \quad (\text{III.9})$$



**FIG. III.1: Représentation graphique du potentiel de Lennard-Jones en 6-12.** Le paramètre  $U$  représente l'énergie associée aux interactions de van der Waals, et  $r$  la distance entre deux atomes  $i$  et  $j$ .

### Interactions à longue distance :

Les systèmes sont simulés dans une boîte, qui impose donc des limites spatiales. Afin de s'affranchir des effets de bords, on modélise ces derniers dans une boîte à bords périodiques, c'est-à-dire qu'on

modélise des systèmes de taille finie délimités par des boîtes et répliqués (virtuellement) à l'infini. Ainsi, chaque atome interagit avec les autres atomes de la boîte, mais aussi avec ses images virtuelles. Cependant, les systèmes comportent des milliers d'atomes, ce qui rend le traitement de toutes les interactions très long. Ainsi, des méthodes d'approximation sont utilisées pour évaluer les interactions à longue distance :

- image minimale : chaque atome “voit” au plus une seule image de tous les autres atomes présents dans le système. L'énergie d'interaction est calculée avec l'image de l'atome la plus proche, qu'elle soit à l'intérieur de la boîte ou dans une boîte voisine.
- rayon de coupure ou *cut-off* : il est associé à l'image minimale. Au-delà d'une certaine distance, les interactions entre paires d'atomes sont considérées comme nulle. De manière générale, lors de l'emploi de conditions aux limites périodiques, la valeur du cut-off ne doit pas dépasser la moitié de la taille minimale de la boîte.
- Somme d'Ewald : Les interactions de van der Waals diminuent en  $\frac{1}{r^6}$ , et peuvent donc être négligées au-delà d'un certain rayon de coupure. À l'inverse, les interactions coulombiennes varient de manière inversement proportionnelle à la distance  $r_{ij}$ , et ne peuvent donc être omises. Elles sont donc souvent évaluées par le protocole du maillage particulier d'Ewald [159]. Fondé sur la somme d'Ewald, cet algorithme calcule de manière efficace les interactions (électrostatiques) à longue distance entre les atomes et toutes leurs images périodiques. La somme d'Ewald est composée de trois termes :

$$U_{Ewald} = U^R + U^I + U^0 \quad (\text{III.10})$$

Le premier terme  $U^R$  est la somme des interactions à courte distance dans l'espace direct réel,  $U^I$  est la somme des interactions à longue distance dans l'espace réciproque imaginaire (images de la boîte), et  $U^0$  est un terme correctif constant. Dans l'algorithme PME,  $U^R$  est évalué directement dans un rayon de coupure donné et  $U^I$  est approximé par une transformée de Fourier rapide sur une grille où les charges sont interpolées en chaque point.

### 1.3 Exemple de protocole de simulation d'un système

Le résultat d'une dynamique moléculaire dépend des conditions initiales, basées sur différents paramètres mais aussi sur la conformation de départ et le jeu de vitesses initiales, qui suit une distribution de Maxwell-Boltzmann obtenue à partir d'un générateur de nombre aléatoire.

Classiquement, la mise au point d'une simulation se déroule de la manière suivante :

1. **Construction du système** : Les coordonnées atomiques initiales sont obtenues à partir des structures tridimensionnelles à haute résolution, résolues par cristallographie aux rayons X ou par RMN. Elles sont disponibles sur la *RCSB Protein Data Bank* (PDB) [8]. Les atomes manquants, classiquement les hydrogènes, sont modélisés. Des contre-ions sont ajoutés pour neutraliser les systèmes chargés, 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 périodique avec un solvant explicite ou implicite pré-équilibré. Dans le cas d'un solvant explicite, les molécules d'eau sont modélisées généralement avec le modèle TIP3P [160]. Le volume du soluté est "découpé" dans cette boîte en retirant les molécules de solvant chevauchantes, dont au moins un atome est situé à moins de 1 Å du soluté. Dans le cas de solvant implicite, les molécules d'eau ne sont pas modélisées, mais leur effet est mimé en changeant la constante diélectrique  $\epsilon_{ij}$  du milieu par celle de l'eau, dans l'expression du potentiel électrostatique.

- 2. Minimisation de l'énergie :** Le plus souvent, les points de départ des simulations de dynamique moléculaire sont des structures de biomolécules obtenues par RMN ou par cristallographie aux rayons X. Les longueurs et angles de liaisons de ces structures ne sont pas forcément compatibles avec les valeurs définies dans le champ de force. Par ailleurs, suite à des erreurs ou au manque de données expérimentales, les structures peuvent présenter des défauts : superposition des atomes, coordonnées atomiques manquantes... Ces défauts et ce manque d'adéquation au champ de force induit des valeurs trop grandes de l'énergie potentielle et de son gradient. Les forces appliquées sur le système sont alors trop grandes, ce qui donne des accélérations trop élevées, et fait bouger les atomes de manière violente, ce qui déstabilise la simulation numérique. Pour obtenir un point de départ "énergétiquement réaliste" et énergétiquement compatible avec le champ de force, la première étape d'une simulation consiste à relaxer le système par une étape de minimisation d'énergie, qui va détecter un minimum énergétique local. Il existe plusieurs algorithmes de minimisation. Les méthodes d'ordre 0 utilisent simplement des évaluations de la fonction d'énergie. Les méthodes d'ordre 1 utilisent la dérivée première de la fonction  $f$  (gradient). La méthode de la plus grande pente (*Steepest Descent*) et celle du gradient conjugué en font partie. Enfin, les méthodes d'ordre 2 utilisent le gradient et le Hessien (dérivée seconde) de  $f$  (Newton-Raphson, Quasi-Newton, Newton-tronqué...). Plus l'ordre est grand et plus l'algorithme est précis, mais plus il consomme de ressources (CPU, mémoire). En général, une combinaison de la méthode de la plus grande pente (*Steepest Descent*) et du Gradient conjugué est utilisée. Dans un premier temps, la méthode de la plus grande pente permet de tendre plus rapidement vers un minimum d'énergie local. Puis la méthode du gradient conjugué sert à affiner la minimisation pour une meilleure convergence vers ce minimum local.
- 3. Thermalisation - Equilibration - Production :** Pour rétablir une densité optimale après l'introduction de la protéine dans la boîte de solvant, une étape d'équilibration est nécessaire. Le système est donc d'abord chauffé à la température souhaitée avec des contraintes harmoniques sur les positions atomiques du soluté. Le système est ensuite équilibré à volume constant puis à pression constante pour ajuster le volume et corriger la densité de la boîte. Les contraintes harmoniques sont relâchées progressivement. À la fin de cette étape, le système est prêt à être simulé dans un ensemble thermodynamique choisi, et produire ainsi une trajectoire de DM.

Il existe quatre ensembles thermodynamiques différents. L'ensemble micro-canonique, dit

NVE, est réalisé à nombre d'atomes  $N$ , volume  $V$  et énergie totale  $E$  du système constants. L'ensemble canonique NVT est réalisé dans les mêmes conditions, mais avec une température  $T$  constante et une énergie totale du système variable. L'ensemble isotherme-isobare, plus connu sous le nom NPT, est réalisé à pression  $P$  et température  $T$  constantes. Enfin, dans l'ensemble grand-canonique, ou  $\mu$ VT, le potentiel chimique  $\mu$ , le volume et la température sont constants, laissant ainsi l'énergie et le nombre d'atomes fluctuer au cours de la simulation.

Dans les dynamiques moléculaires de système protéique, un ensemble très utilisé est l'ensemble NPT. La température  $T$  est reliée à l'énergie cinétique  $E_c$  :

$$E_c = \sum_{i=1}^N \frac{1}{2} m_i v_i^2 = \frac{k_b T}{2} (3N - N_c) \quad (\text{III.11})$$

avec  $N$  le nombre d'atomes,  $(3N - N_c)$  le nombre de degré de liberté,  $m_i$  la masse atomique,  $k_b$  la constante de Boltzmann. La température peut être contrôlée en modifiant les vitesses  $v_i$  des atomes et molécules du système. Par exemple, les vitesses peuvent être ajustées de manière à avoir  $T = T_{ref}$  à chaque pas. L'algorithme de Berendsen [161] par exemple, ajuste les vitesses à un taux proportionnel à la différence de température.

La pression est reliée au volume du système et aux interactions entre atomes :  $PV = Nk_b T$ . Elle peut donc être contrôlée en faisant varier le volume du système et en ajustant la position des atomes les uns par rapport aux autres.

## 2 La dynamique moléculaire avancée

Comme il a été décrit dans la section précédente, le potentiel utilisé pour une simulation de dynamique moléculaire requiert l'utilisation d'un très petit pas d'intégration, de l'ordre de la femtoseconde. De ce fait, le temps réel de simulation est relativement limité et constitue un obstacle majeur dans l'étude de changements conformationnels. En effet, la plupart des fonctions biologiques dans la cellule se produisent sur une échelle de temps allant de la microseconde à la milliseconde. Or les ressources informatiques, bien qu'elles se soient énormément développées ces dernières années, ne permettent de simuler des gros systèmes que sur quelques centaines de nanosecondes. D'autre part, les réarrangements conformationnels induits par ces mouvements et interactions sont associés à des barrières énergétiques importantes, qui ne peuvent généralement pas être traversées par des simulations de DM classique.

Pour palier ce problème, de nombreuses approches ont été développées. Elles se divisent principalement en deux classes, l'une étant fondée sur la mécanique moléculaire, l'autre sur la dynamique moléculaire. Quelques unes d'entre elles sont présentées dans les sections suivantes.

## 2.1 Méthodes basées sur la mécanique moléculaire

### Le réseau élastique

Dans le modèle en réseau élastique, ou *elastic network model (ENM)* [162; 163], la protéine est représentée par un réseau de ressorts harmoniques qui connectent entre eux les atomes interagissants (en général les carbones  $\alpha$  ou les résidus proches. L'analyse en modes normaux (*normal mode analysis (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 accessibles à proximité de l'état natif. Le calcul des modes normaux se fait dans le vide, en trois étapes : (i) minimisation de l'énergie potentielle, fonction complexe des  $3N$  coordonnées avec  $N$  le nombre d'atomes, (ii) calcul de la matrice "Hessienne" (dérivée seconde de l'énergie potentielle), (iii) détermination des valeurs propres et vecteurs propres de la matrice Hessienne par diagonalisation. Il existe  $3N$  vecteurs propres et chacun d'entre eux correspond à un mode spécifique, les six premiers de valeur propre nulle correspondant aux mouvements de translation et de rotation.

Les modes normaux peuvent également être utilisés pour l'analyse des mouvements. Le Hessien est alors calculé indépendamment des masses, donc dans l'espace des positions, ce qui correspond à une analyse quasi-harmonique, elle-même équivalente à une analyse en composantes principales (ACP). Cela revient ainsi à étudier l'amplitude de déplacement. La matrice Hessienne est calculée de cette façon :

$$H_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j} \quad (\text{III.12})$$

où  $V$  est le potentiel calculé à partir du champ de force, et  $q_i$  le déplacement de la  $i^{\text{me}}$  coordonnée par rapport au minimum d'énergie.

À partir de ces vecteurs propres, on peut estimer la flexibilité de la protéine à partir du calcul des fluctuations atomiques RMSFs (*Root-Mean-Square Fluctuations*) :

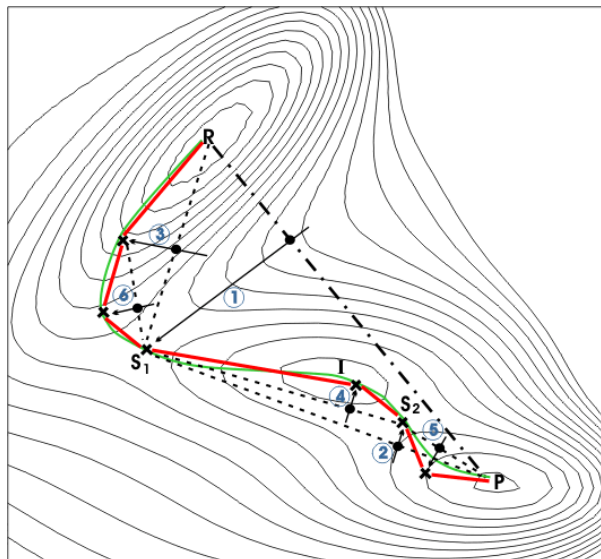
$$RMSF_i = k_B T \sqrt{\sum_j^{\text{modes}} \frac{L_{ij}^2}{\lambda_j^2}} \quad (\text{III.13})$$

où  $i$  est le  $i^{\text{eme}}$  degré de liberté,  $k_B$  est la constante de Boltzmann,  $T$  la température et  $L_{ij}$  le  $i^{\text{eme}}$  composant du  $j^{\text{eme}}$  vecteur propre de raideur (et non de fréquence)  $\lambda_j$ .

### Affinement par pic conjugué (*Conjugate Peak Refinement* ou *CPR*)

L'affinement par pic conjugué [164] (*Conjugate Peak Refinement* ou *CPR*), développée par Fischer et Karplus, est l'une des méthodes heuristiques la plus largement utilisée pour calculer un chemin de transition entre deux états, inactif et actif, d'une protéine. Cette méthode ajoute des conformations intermédiaires entre ces deux états en nombre suffisant. Le but est de trouver un chemin de transition d'énergie minimum, et adiabatique. Les conformations correspondant aux maxima d'énergie sont donc optimisées de façon à ce qu'elles convergent vers des points selle. Le

résultat de cette méthode CPR dépend donc crucialement du chemin initial donné en entrée, et ne fournit qu'un seul chemin parmi de nombreux autres possibles. L'algorithme CPR est présenté sur la figure III.2.



**FIG. III.2: Représentation schématique de l'algorithme CPR.** Les lignes de contour représentent le paysage énergétique en deux dimensions d'une molécule hypothétique, avec des minima au niveau des points R, I, et P. Une première conjecture est faite sur le chemin entre les points R (Réactif) et P (Produit), représentée ici par l'interpolation linéaire (- . - . -) entre ces deux points. Chaque cycle CPR, numéroté en bleu, ajoute automatiquement un nouveau point intermédiaire (●) au niveau de la région de plus haute énergie (le "pic") et l'optimise par une minimisation des gradients conjugués (->) de manière à ce qu'il se place à proximité de la vallée d'énergie (X). Ce nouveau point est ensuite incorporé dans le chemin par les deux nouveaux segments d'interpolation (- - -). La répétition de cette procédure heuristique conduit à la génération d'un chemin final (—) qui suit la vallée de la surface d'énergie et passe exactement par deux points selles (S1 et S2), proche du chemin d'énergie minimum (—).

## 2.2 Méthodes fondées sur la dynamique moléculaire

### Dynamique moléculaire ciblée ou *Targeted Molecular Dynamics (TMD)*

La dynamique moléculaire ciblée ou *Targeted Molecular Dynamics*[165] est utilisée pour déterminer le chemin de transition d'une conformation initiale vers une conformation finale, toutes deux connues, c'est-à-dire dont on dispose des coordonnées atomiques. Cette méthode utilise une contrainte harmonique qui permet de réduire, à chaque pas de temps, l'écart quadratique moyen (*Root mean square deviation (RMSD)*) d'une structure mouvante vers une structure cible. La force appliquée sur chaque atome est donnée par le gradient de potentiel suivant :

$$U_{TMD} = \frac{1}{2} \frac{k}{N} [RMSD(t) - RMSD^*(t)]^2 \quad (\text{III.14})$$

où  $RMSD(t)$  est le RMSD instantané entre les coordonnées atomiques actualisées et les coordonnées atomiques cibles, et  $RMSD^*(t)$  évolue linéairement à partir du RMSD initial au premier pas de TMD, vers le RMSD final au dernier pas de TMD.  $k$  est une constante de force



et  $N$  le nombre d'atomes cibles.

### Dynamique moléculaire dirigée ou *Steered Molecular Dynamics (SMD)*

La dynamique moléculaire dirigée [166] ou *Steered Molecular Dynamics (SMD)* est une extension de la première. Comme pour la TMD, la structure finale est requise afin de guider le système dans une direction donnée. Une contrainte harmonique dotée d'une constante de force  $k$  est appliquée sur le centre de masse des atomes marqués, qui est forcé à évoluer avec une vitesse  $v$  constante dans la direction contrainte  $\vec{n}$ . La méthode SMD repose donc sur le potentiel appliqué au système suivant :

$$U(\vec{r}_1, \vec{r}_2, \dots, t) = \frac{1}{2}k \left[ vt - (\vec{R}(t) - \vec{R}_0) \cdot \vec{n} \right]^2. \quad (\text{III.15})$$

Ici,  $t \equiv N_{ts}dt$  où  $N_{ts}$  est le nombre de pas écoulé au cours de la simulation, et  $dt$  l'échelle du pas de temps en femtosecondes.  $\vec{R}(t)$  représente le centre de masse instantané des atomes marqués et  $\vec{R}_0$  est le centre de masse initial. Le vecteur  $\vec{n}$  est normalisé avant d'être utilisé.

Ces deux méthodes présentées ici consistent à explorer l'espace conformationnel entre deux états connus, et ne peuvent donc être appliquées si la conformation de l'état final est inconnue. D'autres méthodes, également basées sur la dynamique moléculaire, ont été développées pour une exploration accélérée de l'espace conformationnel, indépendamment d'un état final. Quelques unes d'entre elles sont présentées dans les points suivants.

### Echange de répliques (ou *Replica Exchange (RE)*)

Dans la méthode d'échange de réplique [167; 168] (ou *Replica Exchange (RE)*), le système est initialement dupliqué en plusieurs copies. A chaque réplique est attribuée une température différente, dans le but de traverser la barrière énergétique. Ces copies sont ensuite simulées en parallèle par dynamique moléculaire. Les valeurs de température sont échangées entre répliques voisines à des intervalles de temps donnés, selon un critère de Metropolis [169].

### Adaptive Tempering

Cette méthode récemment développée par Zhang (2010) [170] est proche de la méthode d'échange des répliques à copie unique pour l'actualisation dynamique de la température au cours de la simulation. La température  $T$  est une variable aléatoire dans l'intervalle  $[Tmin, Tmax]$  gouverné par l'équation :

$$dE/dT = E - E(T) - 1/T + \text{sqrt}(2)T\xi, \quad (\text{III.16})$$

où  $\xi$  est un bruit Gaussien. Lorsque l'énergie potentielle pour une structure donnée est inférieure à l'énergie moyenne, la température est diminuée. A l'inverse, lorsque l'énergie courante est plus élevée que l'énergie moyenne, la température est alors augmentée. Ainsi, on accède à un échantillonnage conformationnel plus rapide pour trouver les conformations d'énergie minimale.

### Dynamique moléculaire accélérée (*Accelerated Molecular Dynamics (AMD)*)

La dynamique moléculaire accélérée [171] (ou *Accelerated Molecular Dynamics (AMD)*), développée récemment par McCammon, améliore l'exploration de l'espace conformationnel du système en réduisant les barrières énergétiques séparant les différents états du système. Cette méthode modifie le paysage énergétique du potentiel en augmentant les "puits" des bassins énergétiques qui sont en-dessous d'un certain seuil, tout en gardant ceux qui sont au-dessus de celui-ci intacts. Ainsi, les barrières énergétiques séparant des bassins énergétiques adjacents sont réduits, permettant ainsi un échantillonnage de l'espace conformationnel difficilement accessible par dynamique moléculaire classique.

Dans la forme originale de la méthode AMD, lorsque l'énergie potentielle du système tombe en-dessous d'un seuil d'énergie,  $E$ , un potentiel "accélérateur" est ajouté, de telle façon que le potentiel modifié,  $V^*(\mathbf{r})$ , soit relié au potentiel originel  $V(\mathbf{r})$ , *via*

$$V^*(\mathbf{r}) = V(\mathbf{r}) + \Delta V(\mathbf{r}), \quad (\text{III.17})$$

où  $\Delta V(\mathbf{r})$  est le potentiel accélérateur,

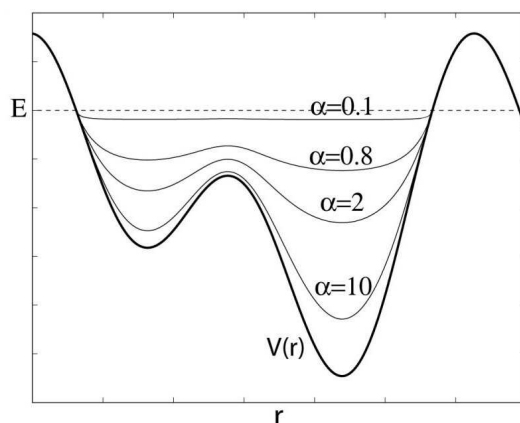
$$\Delta V(\mathbf{r}) = \begin{cases} 0 & V(\mathbf{r}) \geq E \\ \frac{(E-V(\mathbf{r}))^2}{\alpha+E-V(\mathbf{r})} & V(\mathbf{r}) < E. \end{cases} \quad (\text{III.18})$$

Comme le montre la figure III.3, l'énergie seuil  $E$  contrôle la portion du potentiel de surface affectée par "l'accélérateur", tandis que le facteur accélérateur  $\alpha$  détermine la forme du potentiel modifié.

L'énergie "accélétratrice" peut être appliquée sur le potentiel des angles dièdres, sur le potentiel total ou sur les deux simultanément.

### Dynamique accélérée par la température (*Temperature Accelerated Dynamics (TAD)*)

L'idée de la dynamique accélérée par la température [172] (ou *Temperature Accelerated Dynamic (TAD)*) est d'accélérer les transitions en augmentant la température, tout en filtrant les transitions qui n'auraient pas dû se produire à température originale. La trajectoire est libre d'explorer par elle-même des chemins de transitions pour sortir du bassin énergétique, aucune information n'est requise sur la nature du mécanisme réactionnel. Dans chaque bassin, le système



**FIG. III.3: Représentation schématique de la méthode AMD.** Le potentiel original est indiqué par une ligne en pointillés. Lorsqu'il descend en-dessous de l'énergie seuil  $E$  (ligne en gras), un potentiel "accélérateur" est ajouté. Les profils d'énergie modifiés (lignes fines) ont des barrières énergétiques plus basses séparant les bassins énergétiques adjacents.

évolue à haute température  $T_{high}$  (tandis que la température d'intérêt est plus basse  $T_{low}$ ). À chaque fois qu'une transition en-dehors du bassin est détectée, le point selle de cette transition est calculé, puis la trajectoire est ramenée dans le bassin pour y être prolongée. Cette procédure répétitive génère une liste de chemins et de temps "d'échappement" pour le système à haute température. La probabilité de distribution du temps requis pour le premier échappement pour chaque mécanisme est une exponentielle. La théorie harmonique de la transition des états donne une dépendance Arrhénius du taux sur la température, ne dépendant que de la hauteur de la barrière statique, il est alors possible d'extrapoler chaque temps d'échappement observé à  $T_{high}$  pour obtenir le temps d'échappement correspondant à  $T_{low}$ , qui est obtenu correctement à partir de la distribution exponentielle à  $T_{low}$ . Le temps pour chaque événement observée à  $T_{high}$  extrapolé à  $T_{low}$  est donné par l'équation suivante :

$$T_{low} = T_{high} \exp^{E_a(\beta_{low} - \beta_{high})} \quad (\text{III.19})$$

### La méta-dynamique (*Metadynamics*)

La méthode méta-dynamique [173] ("*metadynamics (MD)*") utilise un potentiel d'énergie additionnel pour biaiser l'exploration de l'espace conformationnel et accélérer les événements rares. Ce potentiel est une somme de Gaussiennes centrées sur les valeurs de variables collectives choisies qui décrivent le système. Une variable collective est une fonction des coordonnées atomiques du système, dont la variation est utilisée pour décrire la transition conformationnelle que l'on veut étudier.

Une nouvelle Gaussienne est ajoutée à chaque intervalle de temps  $t_G$ , le potentiel biaisé au

temps  $t$  est donné par :

$$V_G(S(x), t) = \sum_{t'=t_G, 2t_G, 3t_G, \dots} \omega \exp\left(-\frac{(S(x) - s_{t'})^2}{2\delta s^2}\right) \quad (\text{III.20})$$

où  $\omega$  et  $\delta s$  sont la hauteur et la largeur des Gaussiennes et  $s_t = S(x(t))$  est la valeur de la variable collective au temps  $t$ . Avec le temps, ce potentiel remplit les minima de la surface d'énergie libre et guide le système en-dehors de ces minima. De cette manière, la somme des Gaussiennes encourage le système à explorer tout le paysage énergétique, sans revenir à l'étape précédente, et produit l'image négative de l'énergie libre. À ce niveau, l'énergie libre devient une constante en fonction des variables collectives.

La méthode de méta-dynamique a l'avantage de ne pas avoir besoin d'une estimation initiale du paysage énergétique à explorer. Cependant, le choix des variables collectives est une étape cruciale qui n'est pas triviale.

### Dynamique moléculaire accélérée par la température ou *Temperature accelerated molecular dynamics (TAMD)*

La dynamique moléculaire accélérée par la température [150] (ou *Temperature Accelerated Molecular Dynamics (TAMD)*) est différente de la dynamique accélérée à cause de l'introduction d'une température supplémentaire comme dans la méthode TAD énoncée précédemment, et se rapproche plutôt de la méta-dynamique et de la SMD. En effet, cette méthode fait évoluer les valeurs cibles des variables collectives à chaque pas de temps, en utilisant le thermostat de Langevin, qui fonctionne à une température plus élevée que la température de la trajectoire de dynamique moléculaire. La TAMD présente plusieurs avantages : (i) les variables collectives sont forcées d'évoluer sur une surface d'énergie libre non biaisée du système, (ii) la simulation peut se faire en modèle tout-atome et en solvant explicite, (iii) le choix du type et du nombre de variables collectives est très flexible (des centres de masse, distances entre atomes ou centres de masse, ou même des angles entre deux centres de masse), (iv) contrairement à la SMD, la connaissance de la structure finale n'est pas requise.

Les équations de la TAMD sont les suivantes :

$$m_i \ddot{x}_i = -\frac{\partial V(x)}{\partial x_i} - k \sum_{j=1}^m (\theta_j(x) - z_j) \frac{\partial \theta_j(x)}{\partial x_i} - m_i \gamma \dot{x}_i + \sqrt{2\beta^{-1} \gamma m_i} \eta_i^x(i) \quad (\text{III.21})$$

$$\bar{\gamma} \bar{m}_j \dot{z}_j = k(\theta_j(x) - z_j) + \sqrt{2\bar{\beta}^{-1} \bar{\gamma} \bar{m}_j} \eta_j^z(i) \quad (\text{III.22})$$

où  $\theta(x) = (\theta_1(x), \dots, \theta_m(x))$  sont les variables collectives, qui sont fonction des coordonnées cartésiennes des atomes, et  $z = (z_1, \dots, z_m)$  sont des valeurs particulières de ces variables.  $V(x)$  est l'énergie potentielle du champ de force considéré,  $k$  est la constante de force.  $\gamma$  est le coefficient de friction de Langevin.  $\eta$  est un puits Gaussien de moyenne nulle.  $\beta = \frac{1}{k_B T}$  où  $k_B$  est la constante

de Boltzmann et  $T$  la température du système.  $\bar{\beta} = \frac{1}{k_B T}$  l'inverse de la température artificielle  $\bar{T}$ . Le jeu d'équations ci-dessus décrit le mouvement de  $x(t)$  et de  $\theta(t)$  dans le potentiel étendu :

$$U_k(x, \theta) = V(x) + \frac{k}{2} \sum_{j=1}^m (\theta_j(x) - z_j)^2 \quad (\text{III.23})$$

Il a été montré dans [150] que si : (i)  $\bar{\gamma}$  est choisi suffisamment grand pour garantir que  $z(t)$  évolue lentement par rapport aux variables  $x(t)$ , (ii)  $k$  est suffisamment grand pour que  $z_j(t) \approx \theta_j(x(t))$  à chaque pas de temps, alors la force agissant sur  $z$  est approximativement égale à l'opposé du gradient d'énergie libre  $F$  à la température  $T$  :

$$\bar{\gamma} \bar{m}_j \dot{z}_j = -\frac{\partial F(z)}{\partial z_j} + \sqrt{2\bar{\beta}^{-1} \bar{\gamma} \eta_j^z(i)} \quad (\text{III.24})$$

La TAMD est apparentée à la SMD (“*Steered Molecular Dynamics*”) à cause du terme de contrainte harmonique sur les valeurs cibles  $z_j$  des variables collectives  $\theta_j$ . Mais, les valeurs cibles des variables collectives ne sont pas imposées de l'extérieur, mais évoluent en étant attachées à un thermostat de Langevin. Ce thermostat est caractérisé par une température supérieure à celle du système, ce qui permet d'accélérer l'évolution des valeurs cibles des variables collectives, et donc l'exploration de l'espace conformationnel.

Dans le cadre de notre étude, c'est cette dernière méthode d'accélération de la dynamique moléculaire qui a été utilisée pour échantillonner d'avantage l'espace conformationnel du domaine AC de la protéine CyaA de *Bordetella pertussis*. En effet, la structure de l'état inactif demeure inconnue à ce jour, et les simulations de dynamique moléculaire classique n'ont pas permis de franchir la barrière énergétique séparant l'état non-actif (dont la structure a été résolue par cristallographie aux rayons X), d'un autre bassin énergétique.

## Chapitre IV

# Rôle de la calmoduline et des ions $\text{Ca}^{2+}$ dans la stabilisation du domaine catalytique de l'adénylcyclase CyaA de *Bordetella pertussis*

---

## 1 Introduction

Les adénylcyclases (ACs) EF (Facteur œdémateux) de *Bacillus anthracis* et CyaA de *Bordetella pertussis* jouent un rôle essentiel dans l'invasion des cellules hôtes, en surexprimant de l'AMPc dans ces cellules de manière incontrôlée, ce qui altère les processus de signaux cellulaires à travers de mécanismes variés [174; 175; 176; 177; 178; 179; 180; 181].

Les domaines catalytiques du facteur œdémateux et de CyaA sont activés comme adénylcyclase par l'interaction avec la calmoduline (CaM), protéine ubiquitaire sensible aux ions  $\text{Ca}^{2+}$ . CaM comprend deux lobes connectés par un linker flexible. Chaque lobe de CaM contient quatre hélices  $\alpha$ , arrangées par paire appelée "mains EF".

Pour le facteur œdémateux comme pour le domaine catalytique AC de CyaA, l'interaction avec CaM induit une transition conformationnelle importante. La transition du facteur œdémateux a été décryptée par la détermination des structures cristallographiques aux rayons X de l'enzyme libre et couplée avec la CaM [128]. A l'inverse, les structures cristallographiques du domaine AC n'ont été résolues qu'en complexe avec le lobe C-terminal de calmoduline (C-CaM), chargé en ions  $\text{Ca}^{2+}$  [131]. Ces structures de AC sont représentatives de l'activité catalytique de AC car il a été démontré [39] que la C-CaM est capable d'activer complètement le domaine AC. Dans la structure cristallographique du complexe AC/C-CaM (figure IV.1), le domaine AC inclut trois régions principales, à savoir la région CA (en vert) au milieu de la structure, la région CB (en orange) et la région SA (en violet) aux deux extrémités. Les séquences du facteur œdémateux et de AC ont montré [131] que la queue C-terminale (cyan), la boucle catalytique (jaune) et les hélices F et H de SA (violet) correspondent aux switchs C, B, et A du facteur œdémateux [128].

Comme ces switches subissent des changements conformationnels importants pendant l'activation du facteur œdémateux, on peut se demander si leurs homologues dans CyaA subissent également des réarrangements substantiels pendant l'activation du domaine AC.

L'information structurale obtenue sur le facteur œdémateux par Wei-Jen Tang et ses collaborateurs [182; 128; 138; 183; 137], a ouvert la voie à une caractérisation minutieuse des caractéristiques physico-chimiques de l'activité de AC en utilisant des approches de modélisation moléculaire [139; 140]. De plus, la prédiction d'un chemin de transition conformationnelle plausible, fondé sur la connaissance des structures du facteur œdémateux libre et en complexe avec la CaM, a permis de proposer des inhibiteurs de son activation, qui ont été validés expérimentalement par la suite [184]. Le rôle des ions  $\text{Ca}^{2+}$  dans l'interaction facteur œdémateux/CaM a également été étudié [185; 186; 149].

Le manque d'information structurale disponible sur le domaine AC isolé rend l'analyse de l'activation par modélisation moléculaire moins simple que pour le facteur œdémateux. Cependant, le site de liaison du lobe N-terminal de la CaM a été prédit [102], par des modélisations moléculaires et des expériences de cross-linkings. Par ailleurs, des mesures d'ultracentrifugation et de diffusion de la lumière ont montré [77] que le domaine AC isolé en solution est hydraté et sphérique.

Pour obtenir des informations sur le rôle de C-CaM et des ions  $\text{Ca}^{2+}$  dans l'activation du domaine AC, nous avons réalisé des simulations de dynamique moléculaire (DM) des trois systèmes : AC-(2Ca-C-CaM) à partir de la structure PDB 1YRT [131], AC-(0Ca-C-CaM) où les ions  $\text{Ca}^{2+}$  ont été retirés de la structure, et AC-isolé dans lequel la C-CaM et les ions calcium ont également été retirés de la structure 1YRT. La suppression des ions  $\text{Ca}^{2+}$  et de C-CaM a induit une large dérive conformationnelle de AC, conduisant potentiellement à une protéine de forme moins allongée en rapprochant les régions SA et CB (localisées aux deux extrémités de la protéine) l'une de l'autre.

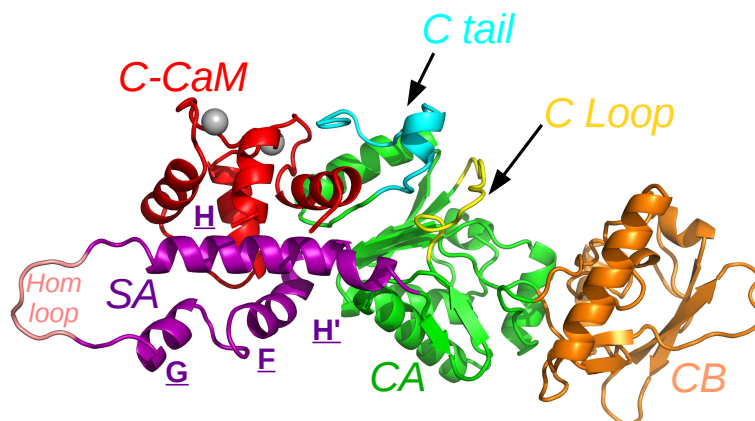
Une analyse des influences énergétiques dans le complexe, réalisée en utilisant une approche précédemment proposée sur le complexe facteur œdémateux/CaM [186], a révélé une influence simple de C-CaM sur la région CA dans les systèmes AC-(0Ca-C-CaM) et AC-(2CA-C-CaM). Ce schéma d'influence simple est en accord avec la forte affinité de AC pour la CaM, et corrèle avec la formation d'un réseau de liaisons hydrogène connectant C-CaM, la queue C-terminale, CA et des résidus de la boucle catalytique.

## 2 Matériels et Méthodes

### 2.1 Préparation des systèmes simulés

Les trois systèmes AC-(2Ca-C-CaM), AC-(0Ca-C-CaM) et AC-isolé ont été construits de la façon suivante. Les coordonnées de AC-(2Ca-C-CaM) ont été extraites à partir du fichier pdb 1YRT [131], et les coordonnées de AC-(0Ca-C-CaM) ont été obtenues en retirant les coordonnées des ions  $\text{Ca}^{2+}$  du même fichier pdb. Le système AC-isolé a été généré de la même façon, en retirant

les coordonnées des ions et de C-CaM. Les résidus 226 à 232 formant la boucle à l'extrémité de la région SA n'ont pas pu être observés dans la structure cristallographique. Ils ont donc été modélisés avec le protocole `build_loop_fixed.py` de Modeller9v4 [187]. La boucle Hom loop ainsi construite est colorée en saumon sur la Figure IV.1.



**FIG. IV.1:** Structure cristallographique par rayons X du domaine catalytique (AC) de l'adénylcyclase *CyaA* (code PDB 1YRT) représentée en ruban. Les hélices F, G, H et H' colorées en violet forment le domaine SA. La boucle 226-232 (Hom loop) à l'extrémité gauche de SA et modélisée avec Modeller9v4[187], est colorée en saumon. Les autres régions de la protéine sont colorées en vert (CA), orange (CB), cyan (la queue C-terminale) et jaune (boucle catalytique : C loop). Le lobe C-CaM, coloré en rouge, interagit avec SA et la queue C-terminale. Les ions calcium sont représentés par des sphères grises. La visualisation a été faite avec pymol.

Les atomes d'hydrogène ont été ajoutés avec le module LEaP de AMBER 9 [152] avec le champ de force FF99SB [188] et des contre-ions  $Na^+$  ont été également ajoutés pour neutraliser le système. Une boîte de molécules d'eau de type TIP3P [189] a été construite en utilisant un rayon de coupure de 10 Å avec des conditions périodiques. Les détails sont résumés dans le tableau I.

Trajectoire	AC-(2Ca-C-CaM)	AC-(0Ca-C-CaM)	AC-isolé
Nombre de contre-ions $Na^+$	14	18	4
Dimensions de la boîte d'eau	92.6 x 106.9 x 103.3 Å <sup>3</sup>	91.3 x 106.9 x 106.4 Å <sup>3</sup>	85.2 x 106.9 x 103.3 Å <sup>3</sup>
Nombre de molécules d'eau	26455	27056	24681
Nombre total d'atomes	85825	87630	79433

**TAB. I:** Détails des systèmes moléculaires simulés dans les trajectoires de dynamique moléculaire (DM).

Les ions calcium ont été paramétrés avec les paramètres de Coulomb et Lennard-Jones suivants [190] : rayon de van der Waals  $R = 1.7131$  Å, profondeur du puits  $\epsilon = 0.459789$  kcal.mol<sup>-1</sup>. Trois trajectoires supplémentaires AC-(2Ca-C-CaM)-2nd, AC-(0Ca-C-CaM)-2nd et AC-isolé-2nd ont été réalisées dans les mêmes conditions pour vérifier la reproductibilité des résultats.



## 2.2 Simulations de Dynamique Moléculaire

Les trajectoires ont été calculées sur 30 ns en utilisant le module PMEMD de AMBER 9.0 [191], sur des clusters de 4 à 8 processeurs. Un rayon de coupure de 10 Å a été défini pour les calculs des interactions de Lennard-Jones, tandis que les interactions électrostatiques à longue portée ont été calculées par la méthode "Particle Mesh Ewald" (PME) [159]. La température et la pression ont été régulées avec le thermostat de Berendsen [161], à une température de 298 K et une pression de 1 atm. L'algorithme SHAKE [154] a permis de garder les liaisons covalentes impliquant les hydrogènes rigides et ainsi d'utiliser un pas d'intégration de 2 fs. Les coordonnées atomiques ont été sauvegardées toutes les picosecondes (30 000 conformations).

Les trajectoires de DM ont été initialisées par une minimisation, répétée avec 10 cycles de "Steepest Descent", suivis par 490 cycles de Gradient Conjugué. Durant ces minimisations, les contraintes harmoniques sur les positions atomiques du soluté ont été progressivement relâchées avec une constante de force allant de 100 à 0 kcal/(mol.Å<sup>2</sup>). Ensuite, les systèmes ont été thermalisés de 0 à 298 K pendant 20 ps, en appliquant des contraintes harmoniques sur les positions atomiques du soluté avec une constante de force de 25 kcal/(mol.Å<sup>2</sup>). Pour terminer, l'équilibration a été lancée avec un cycle de 5 ps à volume constant, suivi de 4 cycles de 2.5 ps à pression constante, un cycle de 10 ps et un dernier de 60 ps. Des forces harmoniques restrictives ont été appliquées sur les coordonnées atomiques des solutés durant les cinq premiers cycles, puis la constante de force a été relâchée de 25 à 0 kcal/(mol.Å<sup>2</sup>) sur tous les atomes.

## 2.3 Convergence des trajectoires

La convergence des trajectoires a été analysée par une approche développée par Lyman et Zuckerman [192]. Elle consiste à générer un jeu de structures représentatives, dites "de référence" et de classer ensuite les conformations de la trajectoire dans les groupes de référence correspondants, en fonction du cut-off  $d_c$  fixé arbitrairement et représentant une valeur de RMSD. Dans un premier temps, une conformation est choisie aléatoirement dans la trajectoire. Elle est appelée "la structure de référence". Ensuite, toutes les conformations pour lesquelles le RMSD calculé par rapport à la structure de référence est plus petit que le cutoff  $d_c$  sont classées dans ce groupe et retirées de la trajectoire. Une seconde structure de référence est alors choisie aléatoirement parmi les conformations restantes, et comme précédemment, toutes les conformations pour lesquelles le RMSD calculé par rapport à cette seconde structure de référence est inférieur à  $d_c$ , sont classées dans ce second groupe, et retirées de la trajectoire. Cet algorithme est répété jusqu'à ce que toutes les conformations de la trajectoire soient classées. Ainsi, les groupes de structures diffèrent les uns des autres d'au moins  $d_c$  Å.

Dans un deuxième temps, les structures de référence sont utilisées pour vérifier la dérive conformationnelle des trajectoires. Chaque trajectoire est coupée en deux moitiés, et les conformations de chaque moitié sont classées dans le groupe de référence le plus proche en terme de RMSD. D'après cette méthode, si la trajectoire a atteint un état stationnaire, chaque groupe de référence

devrait être représenté par un nombre équivalent de conformations provenant des deux moitiés de la trajectoire. Comme la dérive conformationnelle au cours de la première nanoseconde est assez grande et peut ainsi biaiser l'analyse, cette nanoseconde a donc été retirée avant de lancer l'analyse.

## 2.4 Carte de dépendances énergétiques

L'approche des influences énergétiques a pour but de comparer le classement des énergies d'interaction entre les régions d'une protéine, en fonction de leur environnement. Ces influences ont été analysées pour chaque trajectoire de DM avec l'algorithme introduit par Hamacher et ses collaborateurs [193] et précédemment utilisé sur le complexe facteur œdémateux/CaM [186]. Les influences énergétiques sont distinctes des énergies d'interaction entre les régions, car une région peut gagner ou perdre un rang même si l'énergie de liaison varie peu. Le changement dans le classement des énergies donne une information sur l'hierarchie des régions dans l'assemblage du complexe, selon le principe général formulé par Hamacher et ses collaborateurs [193] : "Si la suppression d'une protéine Z change l'ordre relatif de X-plus-fortement-lié-que-Y à Y-plus-fortement-lié-que-X, alors l'ordre de l'assemblage probable est aussi changé, et nous attribuons une influence de la protéine Z sur la protéine X."

Les systèmes moléculaires ont été divisés en quatre régions : CA (incluant la queue C-terminale et la boucle catalytique), CB et SA pour AC, et C-CaM. L'énergie d'interaction a été calculée comme l'énergie MMPBSA de liaison  $\Delta G$  [194] avec AMBER 9 [191], sur les conformations extraites toutes les 10 ps au cours des 27 dernières nanosecondes des trajectoires DM :

$$\Delta G = \Delta E_{MM} + \Delta G_{solv} \quad (IV.1)$$

où  $\Delta E_{MM}$  représente la contribution mécanique, exprimée comme la somme des contributions électrostatiques ( $\Delta E_{ele}$ ), de van der Waals ( $\Delta E_{vdw}$ ), et internes ( $\Delta E_{int}$ ).  $\Delta G_{solv}$  est la contribution de l'énergie libre de solvation exprimée comme la somme des énergies libres de solvation polaires ( $\Delta G_{sol}^{polaire}$ ) et non-polaires ( $\Delta G_{sol}^{nonpolaire}$ ). Le module pbsa de AMBER9 a permis de calculer la contribution polaire  $\Delta G_{sol}^{polaire}$  en résolvant l'équation de Poisson-Boltzmann (PB) sur une grille de pas 0.5 Å et des constantes diélectriques de 1 et 80 pour l'intérieur et l'extérieur du soluté. La frontière diélectrique a été définie en faisant rouler une sonde d'eau de 1.4 Å sur la surface atomique. La contribution non-polaire  $\Delta G_{sol}^{nonpolaire}$  a été calculée comme la somme d'un terme répulsif (cavité)  $\Delta G_{cavite}$  proportionnel à la surface accessible au solvant (SASA) et d'un terme attractif (dispersion)  $\Delta G_{disp}$  calculé par l'approche d'intégration de surface [195].

$$\Delta G_{sol}^{nonpolaire} = \Delta G_{disp} + \Delta G_{cavite} \quad (IV.2)$$

Les influences énergétiques ont été évaluées par l'algorithme suivant. Dans un premier temps, l'énergie MMPBSA de liaison de chaque région  $D$  au complexe  $C$  est calculé ainsi :

$$\Delta\Delta G[D, C] = \Delta G[C] - (\Delta G[C \setminus D] + \Delta G[D]) \quad (\text{IV.3})$$

Les énergies  $\Delta\Delta G[D, C]$  sont ensuite classées de la plus grande à la plus petite, donnant ainsi un rang de référence pour chaque région.

Dans un deuxième temps, pour évaluer l'influence énergétique d'une région  $D^*$  sur les autres, la région  $D^*$  est retirée du complexe, et l'énergie MMPBSA de liaison de chaque région  $D$  sur le complexe tronqué  $C \setminus D^*$  est calculée :

$$\Delta\Delta G[D, C \setminus D^*] = \Delta G[C \setminus D^*] - (\Delta G[C \setminus (D + D^*)] + \Delta G[D]) \quad (\text{IV.4})$$

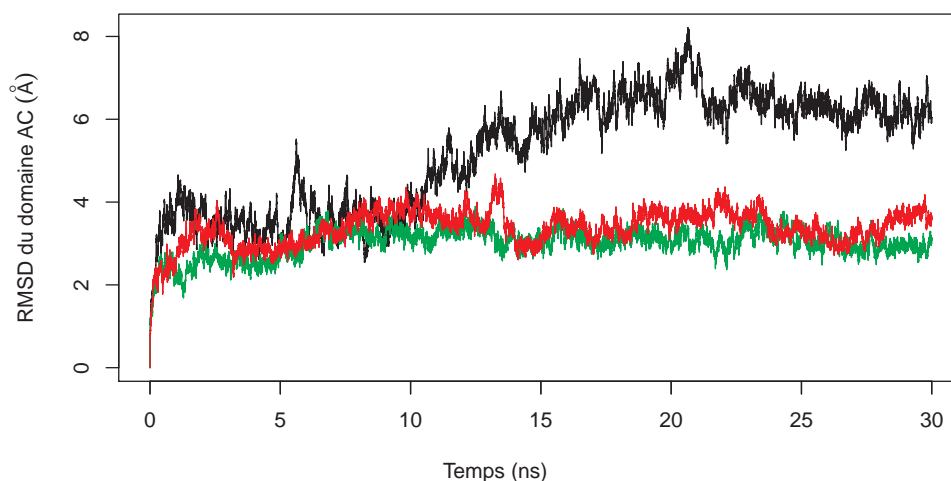
Les énergies calculées sont ensuite classées et vérifiées statistiquement comme précédemment et les listes  $\Delta\Delta G[D, C]$  et  $\Delta\Delta G[D, C \setminus D^*]$  sont comparées. Si une région  $D'$  a perdu un rang par rapport à son rang de référence, alors  $D^*$  a une influence énergétique sur  $D'$ . L'intensité de l'influence correspond au nombre de rangs perdus par  $D'$ .

Le classement des régions est vérifié par rapport aux déviations standard. Si la différence des valeurs moyennes des énergies est plus grande que la somme des déviations standard, le classement produit par les valeurs moyennes des énergies est supposé être fiable. Par ailleurs, le classement a aussi été testé statistiquement par un test non-paramétrique Mann-Whitney [196] réalisé sur les estimations des énergies au cours du temps, avec le logiciel R [197] (commande *wilcox.test*). Les deux hypothèses alternatives  $\Delta\Delta G(D) > \Delta\Delta G(D')$  et  $\Delta\Delta G(D') > \Delta\Delta G(D)$  sont testées, et celui ayant une p-valeur proche de 1 est choisi. Les influences énergétiques ont été calculées à l'aide du script `mm_pbsa.pl` du package AMBER 9 [152]. L'analyse d'une trajectoire de 30 ns requiert une semaine de temps CPU sur un processeur de 2.6 GHz.

## 3 Résultats

### 3.1 Dérives conformationnelles

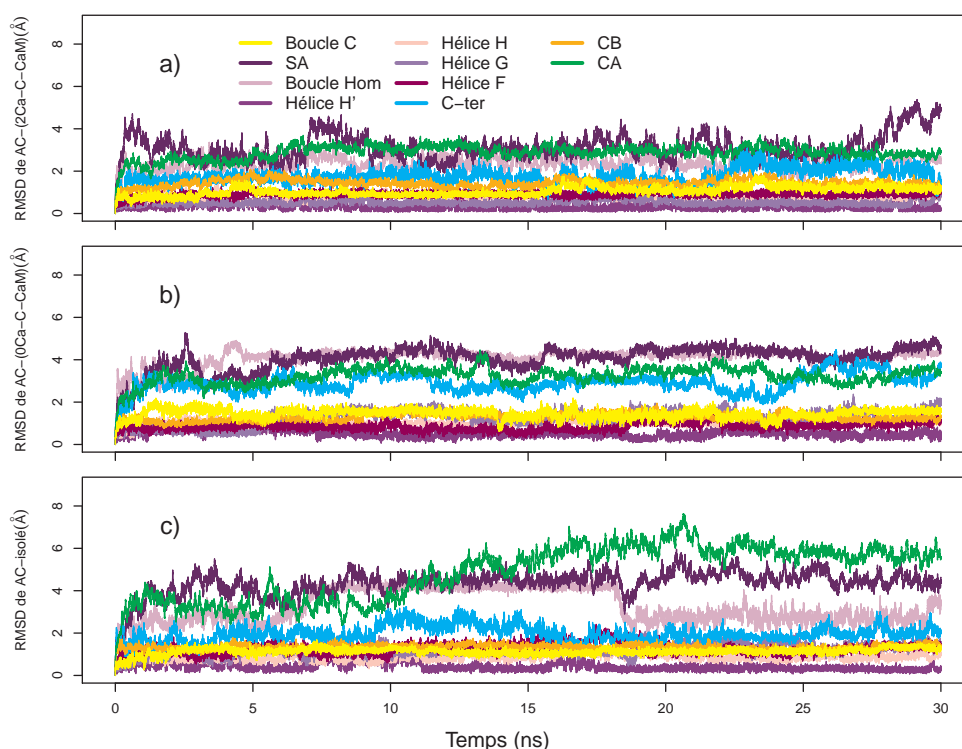
La dérive conformationnelle du domaine AC décrit un plateau autour de 3 Å en présence de C-CaM (Fig. IV.2)(trajectoires AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM)) et atteint 6 Å en absence de C-CaM (trajectoire AC-isolé). L'interaction avec la calmoduline semble donc être un élément essentiel pour la stabilisation de AC, la présence d'ions  $Ca^{2+}$  ne jouant qu'un rôle marginal. Ces observations sont en accord avec la faible influence de la concentration en ions  $Ca^{2+}$  sur l'activité catalytique [39].



**FIG. IV.2: RMSD sur les carbones  $\alpha$  des trois systèmes.** La dérivation est calculée par rapport à la première conformation des trajectoires de : AC-isolé (en noir), AC-(0Ca-C-CaM) (en rouge), AC-(2Ca-C-CaM) (en vert)

Le RMSD des coordonnées atomiques de C-CaM est semblable dans les deux trajectoires AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM), et varie autour de 1.2 Å. Cette faible différence correspond à une réorganisation assez limitée du lobe. D'autre part, d'autres simulations de dynamique moléculaire ont également montré de faibles variations du RMSD en présence de  $Ca^{2+}$  (0.8-2 Å) [198; 199; 108; 200] et en leur absence (2-3 Å) [201; 202; 203]. La C-CaM est donc stabilisée par son interaction avec AC dans le complexe AC/C-CaM.

Les dérives conformationnelles ont été analysées en détail en calculant les RMSDs sur les coordonnées des carbones  $\alpha$  des différentes régions de AC suivantes, en les superposant indépendamment les unes sur les autres (Fig IV.3) : CA, CB, la queue C-terminale (C-ter), SA, la boucle catalytique (Boucle C), les hélices alpha H, H', G, F, et la boucle 226-232 (Boucle Hom) qui est non observable dans la structure cristallographique 1YRT et qui a donc été modélisée par homologie. Les couleurs respectives des courbes des RMSD sont définies dans la légende de la figure IV.3 a).



**FIG. IV.3: Dérives conformationnelles des différentes régions du domaine AC, estimées par le calcul du RMSD des coordonnées atomiques des carbones  $\alpha$ .** La conformation de référence est la première de chaque trajectoire. (a) Trajectoire AC-(2Ca-C-CaM), (b) trajectoire AC-(0Ca-C-CaM) et (c) trajectoire AC isolé. Les courbes sont colorées en fonction de la région selon la légende de la figure (a). Les régions de AC sont définies de cette façon : CA comprend les résidus 7-61, 187-197, 261-299 et 313-345, la boucle catalytique comprend les résidus 300-312, la queue C-terminale contient les résidus 346-364, CB est composée des résidus 62-186 et SA des résidus 198-260. Les hélices H, F, G, H' correspondent aux résidus 234-253, 198-210, 214-223, 256-260.

En absence de C-CaM ou des ions  $\text{Ca}^{2+}$  (Figs IV.3 b, c), la boucle Hom (rose), CA (vert) et SA (violet) sont principalement responsables de l'augmentation du RMSD. En effet, la Boucle Hom est stable autour de 2 Å pour AC-(2Ca-C-CaM), 4 Å pour AC-(0Ca-C-CaM), et varie entre 2 et 4 Å dans AC-isolé. La dérive conformationnelle de CA (vert) présente une légère augmentation dans AC-(0Ca-C-CaM), et une plus importante, jusqu'à 6 Å, dans AC-isolé. Cependant, il est à noter que dans l'intervalle entre 12 et 30 ns de la trajectoire de AC-isolé (Fig IV.3 c), le profil du RMSD de CA est similaire à celui de AC (Fig IV.2 courbe noire). La dérive globale de AC est donc principalement due à la dérive conformationnelle de CA. Le RMSD de SA (violet) passe de 3 Å dans AC-(2Ca-C-CaM) à 4 Å dans AC-(0Ca-C-CaM), mais garde un profil similaire dans AC-(0Ca-C-CaM) et AC-isolé. La grande dérive de SA est en accord avec la correspondance de cette région avec le switch A de l'adénylcyase EF [131].

La boucle catalytique (jaune), CB (orange) et les hélices alpha H, H', G, F (représentées dans un dégradé de violet), présentent des RMSD inférieurs à 2 Å dans toutes les trajectoires. La faible variation observée pour CB démontre que cette région reste rigide, et ceci est en accord avec sa position qui est la plus éloignée du site de liaison de la CaM. Les petites dérives observées pour

les hélices alpha H, H', G et F indiquent que ces hélices sont structurellement stables, ce qui est confirmé par une analyse indépendante des structures secondaires. Les dérives observées pour SA viennent donc des variations de l'orientation entre les hélices H, F, G et H'. En effet, de grandes déviations standards sont observées pour les angles formés par les hélices G-F et H-H' dans AC-(0Ca-C-CaM) et AC-isolé (Table II). Par conséquent, la variabilité des orientations relatives des hélices combinée au RMSD de la boucle Hom décrite précédemment produisent cette forte dérive conformationnelle de SA.

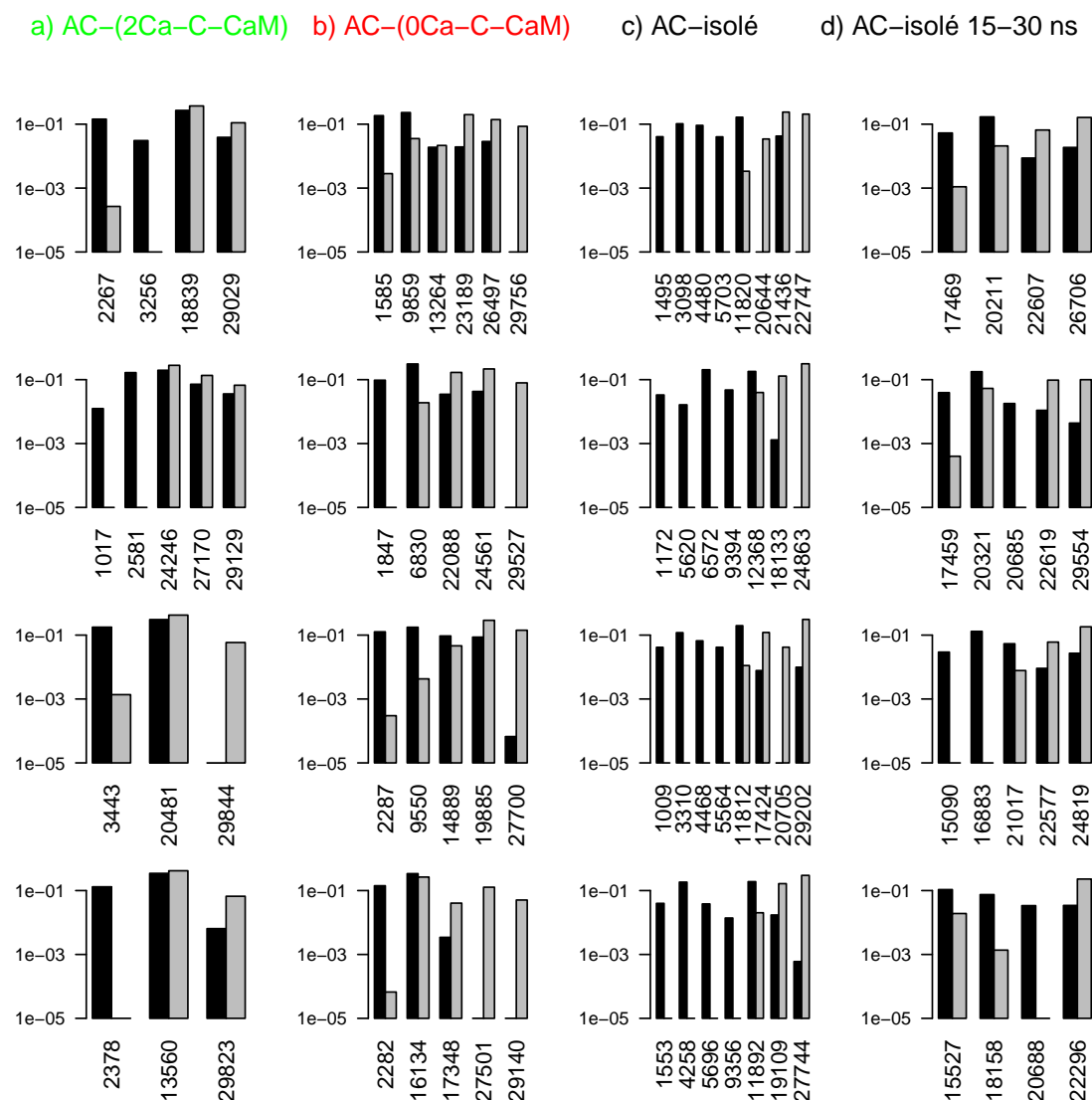
Trajectoire	AC-(2Ca-C-CaM)	AC-(0Ca-C-CaM)	AC-isolé
Angle H-G (°)	45.1 ± 7.1	23.0 ± 8.9	34.1 ± 6.4
Angle G-F (°)	160.4 ± 9.1	148.5 ± 8.0	133.8 ± 13.1
Angle H-F (°)	54.1 ± 4.0	52.0 ± 2.8	65.6 ± 6.7
Angle H-H' (°)	39.3 ± 3.2	72.7 ± 13.7	31.3 ± 13.0

**TAB. II:** Angles (°) entre les axes des hélices  $\alpha$  H (résidus 234-253), G (résidus 214-223), F (résidus 198-210) et H' (résidus 256-260). Les axes ont été définis en prenant les points successifs se situant au milieu des segments formés par les atomes (i) N( $i$ ) et N( $i+2$ ), (ii) C- $\alpha$ ( $i$ ) et C- $\alpha$ ( $i+2$ ), (iii) C( $i$ ) et C( $i+2$ ),  $i$   $i$  étant l'index du résidu dans la séquence.

Les dérives conformationnelles calculées sur les trois trajectoires de Dynamique Moléculaire supplémentaires, non représentées, AC-(2Ca-C-CaM)-2nd, AC-(0Ca-C-CaM)-2nd et AC-isolé-2nd, montrent des résultats similaires à ceux décrits précédemment. Cependant les dérives sont un peu plus importantes pour : (i) SA dans la trajectoire du complexe AC/CaM en présence d'ions  $Ca^{2+}$ , et (ii) la queue C-terminale dans la trajectoire du complexe AC/CaM en absence d'ions  $Ca^{2+}$ . La boucle qui a été reconstruite par homologie est moins mobile dans la trajectoire de AC isolée. Néanmoins, les mobilités relatives des différentes régions sont préservées, ce qui prouve la reproductibilité des observations faites sur les trajectoires de DM.

La convergence des trajectoires a également été analysée avec la méthode de Lyman et Zuckerman [192], décrite dans les Matériels et Méthodes. Des groupes de conformations représentatives ont été extraits des trajectoires en appliquant des valeurs de cut-off  $d_c$  pour le RMSD de 2.0, 2.5, 2.7 et 3.0 Å. L'état stationnaire des trajectoires a été vérifié en regardant si les différents groupes étaient peuplés de la même façon dans les deux moitiés de la trajectoire. La figure IV.4 montre les résultats obtenus en utilisant quatre graines différentes pour le choix aléatoire de la conformation. Toutes les trajectoires montrent des variations au niveau du nombre de groupes générés, tout comme au niveau des populations relatives au sein des groupes dans les deux moitiés d'une trajectoire. Cette variabilité peut être due à la difficulté du choix de la conformation de référence pour chaque groupe, du fait du manque de sensibilité du RMSD dans une protéine non globulaire. La trajectoire AC-isolé (Fig IV.4 c) décrit des tendances distinctes de celles observées dans les trajectoires AC-(2Ca-C-CaM) (Fig IV.4 a) et AC-(0Ca-C-CaM) (Fig IV.4 b), du fait que de nombreux groupes sont peuplés par des conformations provenant uniquement de la première moitié de la trajectoire. Ceci indique que dans AC-isolé, le soluté a quitté son bassin conformationnel initial. L'analyse de la convergence entre 15 et 30 ns de AC-isolé (Fig IV.4 d) montre

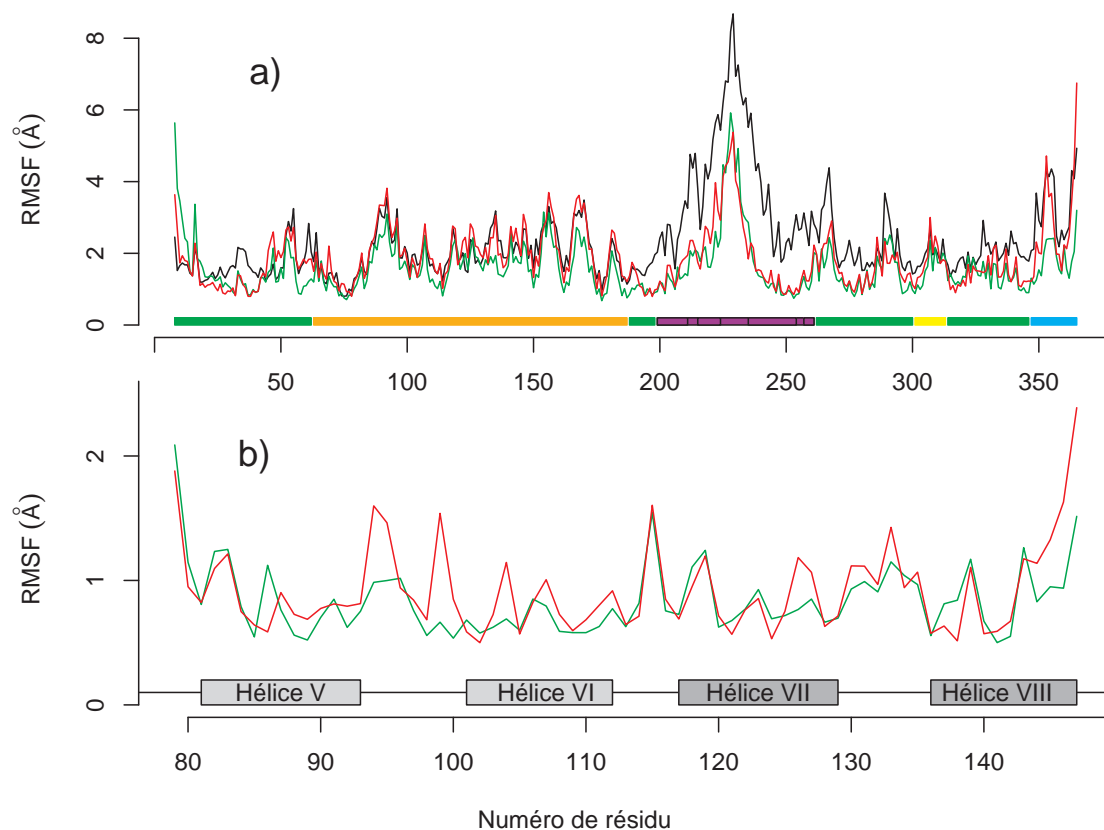
un plus petit nombre de groupes générés, et une stabilisation de la population de ces groupes comparé à la trajectoire entière, ce qui est le signe que la simulation de DM a atteint ses limites dans l'exploration de l'espace conformationnel de AC.



**FIG. IV.4:** Analyse de la convergence des 2 à 30 ns de (a) EF-(2Ca-C-CaM), (b) EF-(0Ca-C-CaM) et (c) AC-isolé et sur (d) l'intervalle 15-30 ns de AC-isolé. Chaque ligne correspond à une analyse pour un rayon de coupure donné. Les numéros des structures de référence sont indiqués en abscisse et correspondent à leurs positions en ps au cours de la trajectoire. Pour chaque groupe de référence, la population des conformations appartenant à la première moitié de la trajectoire sont reportées avec des barres noires, et celle appartenant à la seconde moitié avec des barres grises.

Les fluctuations atomiques du domaine AC montre leur plus grand pic de fluctuation dans la boucle 226-232 construite pas homologie (Fig IV.5 a). Cette boucle est également à l'origine des plus grandes variations de fluctuations entre les différentes trajectoires de DM, avec une augmentation maximale de 2 Å. De plus, des différences dans les profils de fluctuation de AC-

isolé et les autres trajectoires se distinguent pour les résidus 180 à 357, correspondant à SA, à l'extrémité C-terminale, à la boucle catalytique et la partie de CA qui les relie. Ces différences sont en accord avec les grandes dérives conformationnelles observées pour SA et CA dans AC-isolé (Fig IV.3 c). Pour C-CaM (Fig IV.5 b), des augmentations d'une faible flexibilité sont notables lorsque les ions  $Ca^{2+}$  sont retirés du complexe, dans la partie C-terminale et dans la boucle de liaison au  $Ca^{2+}$  de la main EF 3, entre les hélices V et VI.



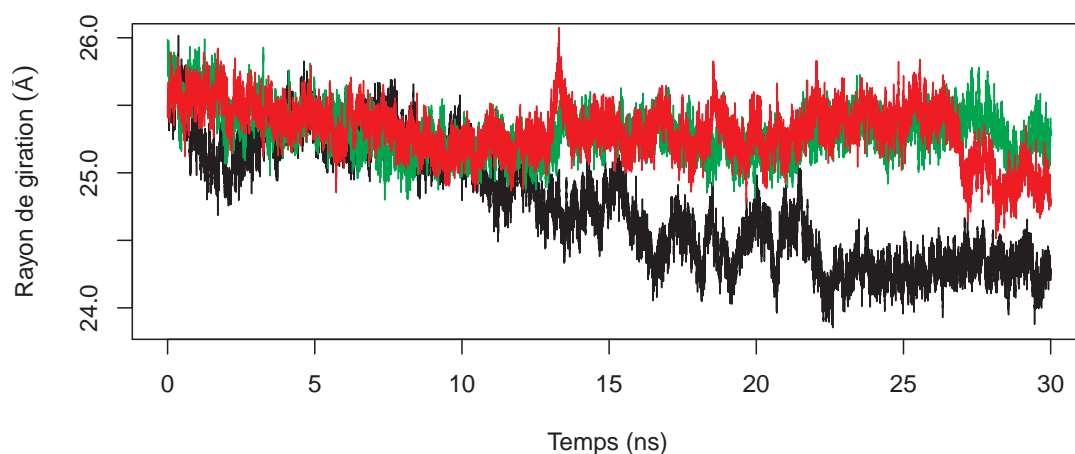
**FIG. IV.5:** Fluctuations atomiques par résidu calculées sur (a) le domaine AC et (b) le lobe C-CaM. Pour chaque figure, les courbes sont représentées en vert pour AC-(2Ca-C-CaM), en rouge pour AC-(0Ca-C-CaM) et en noir pour AC-isolated. En-dessous de la figure (a), les régions de AC sont illustrées le long de la séquence : CA (vert), CB (orange), SA (violet), la queue C-terminale (cyan), la boucle catalytique (jaune). Pour C-CaM (b), les hélices  $\alpha$  sont représentées dans des boîtes.

Les différentes analyses conformationnelles et de convergence décrites précédemment conduisent à une description des variations structurales des systèmes étudiés. Le domaine AC isolé dérive de manière plus importante que lorsqu'il est complexé avec la C-CaM. Les plus importantes variations sont principalement localisées dans les régions SA et CA.



### 3.2 Tendances conformationnelles de AC et activité catalytique

Aucune structure à haute résolution de AC isolée n'est disponible. Cependant, des mesures hydrodynamiques [77] ont détecté une forme sphérique de ce domaine dans les trois états, en solution. Dans notre étude, nous analysons la forme générale du soluté en calculant le rayon de giration défini comme l'écart quadratique moyen des distances entre le centre géométrique du soluté et tous ses atomes. Le rayon de giration (Fig IV.6) est similaire et stable entre 26.5 et 27.0 Å pour les trajectoires AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM), et décroît de 25.5 à 24.2 Å pour la trajectoire AC-isolé. Les variations entre les rayons de AC-isolé et les autres trajectoires sont petites, mais révèlent néanmoins une tendance significative du soluté à diminuer de taille. De plus, les moments d'inertie calculés au cours de la trajectoire AC-isolé (Fig IV.7) montrent une décroissance des moments d'inertie les plus grands, démontrant ainsi que la protéine a tendance à être moins allongée.



**FIG. IV.6:** Variation du rayon de giration du domaine AC au cours des trajectoires AC-(2Ca-C-CaM) (vert), AC-(0Ca-C-CaM) (rouge) and AC-isolé (noir).

Une analyse en composantes principales (ACP) de la matrice de covariance des dynamiques donne accès à des informations additionnelles sur les tendances conformationnelles du soluté. Différents profils de vecteurs propres sont obtenus (Fig IV.8) entre AC-isolé et les autres trajectoires. Pour AC-isolé, seule la première valeur propre est dominante et 10 vecteurs propres suffisent pour expliquer 90% des mouvements internes de la protéine, tandis que pour AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM), ce sont les 2 et les 3 premières valeurs propres respectives qui sont dominantes et il faut 30 et 19 vecteurs propres respectivement pour expliquer 90% des mouvements internes des complexes. Le domaine AC décrit ainsi plus de mouvements collectifs.

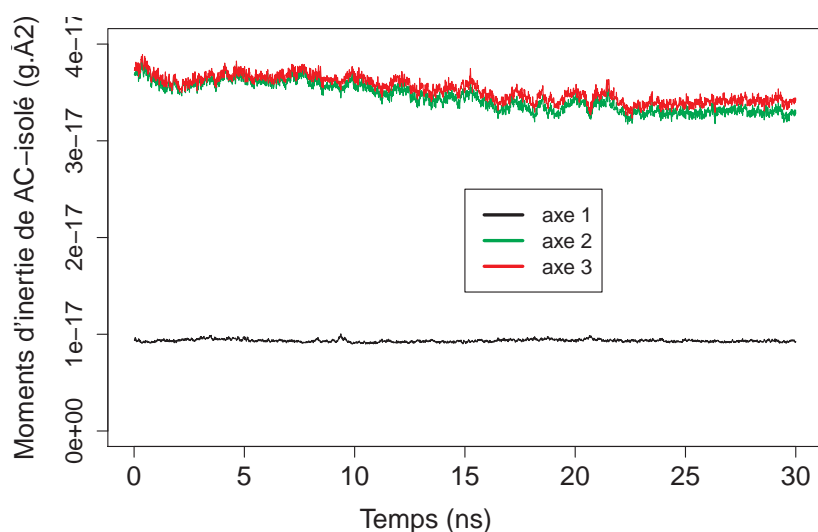


FIG. IV.7: *Variations des moments d'inertie au cours de la trajectoire AC-isolé.* Les moments sont calculés par rapport aux axes 1, 2 et 3, correspondant aux première, deuxième et troisième plus grandes valeurs propres du tenseur d'inertie, calculé sur chaque structure de la trajectoire. 1/10 des conformations enregistrées ont été utilisées pour le calcul.

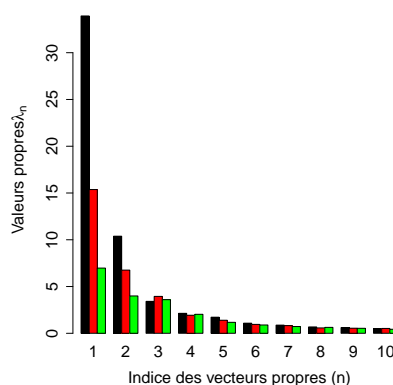
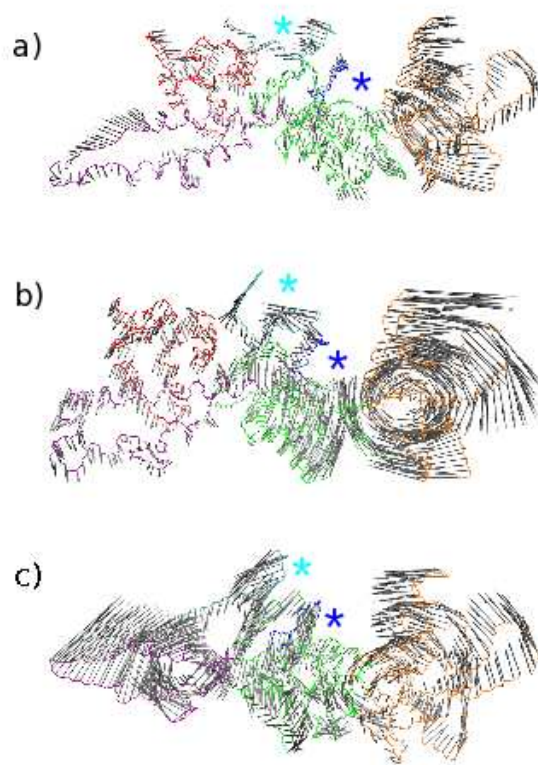


FIG. IV.8: *Valeurs propres de la dynamique essentielle :* Trajectoire AC-(2Ca-C-CaM) en vert, AC-(0Ca-C-CaM) en rouge et AC-isolé en noir.

La projection du premier mode sur la structure du soluté montre des amplitudes plus larges pour les sous-domaines SA et CB (Fig IV.9), et démontre une tendance de ces régions à tourner autour d'un axe perpendiculaire au plan de la figure. Ce mouvement est le plus prononcé pour AC-isolé et soutient l'idée d'une déformation assez large de AC conduisant à une forme sphérique en rapprochant les régions SA et CB l'une vers l'autre, conformément aux résultats des mesures hydrodynamiques [77]. En outre, dans AC-(2Ca-C-CaM) (fig IV.9 a), CB décrit des mouvements plus importants que le reste de AC, tandis que dans AC-(0Ca-C-CaM) (fig IV.9 b), l'amplitude des mouvements de CB augmente au même titre que ceux de CA. Dans AC-isolé (fig IV.9 c), un large déplacement est observé pour SA, tandis que les régions CA et CB voient leurs amplitudes décroître relativement. La projection sur le premier mode observée ici pour CB avec sa petite

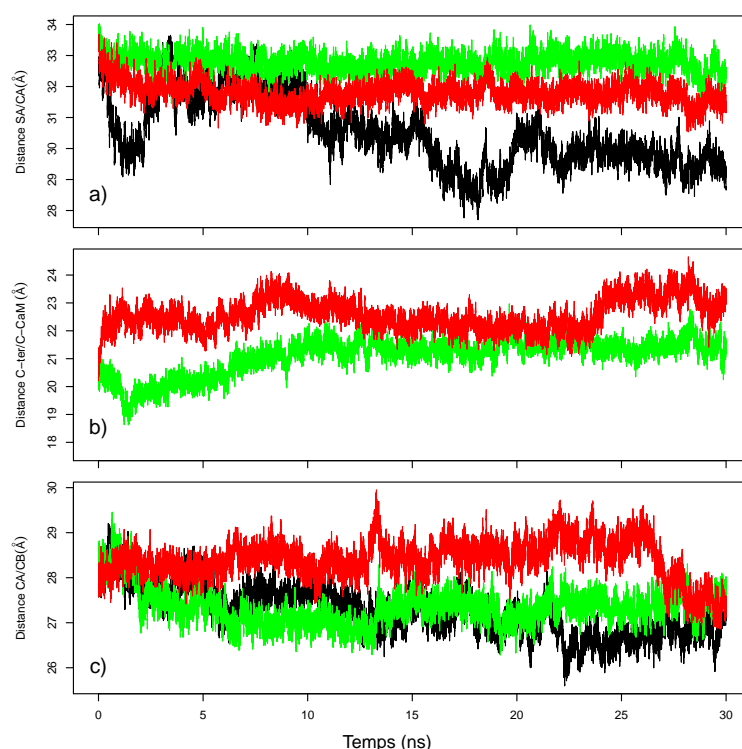
dérive conformationnelle (figs IV.2 et IV.3) prouvent que cette région bouge en tant que corps rigide.



**FIG. IV.9:** *Projections du premier mode sur les structures de :* a) AC-(2Ca-C-CaM), b) AC-(0Ca-C-CaM) et c) AC-isolé.

Les tendances conformationnelles des complexes ont ensuite été comparées à leur architecture statique, obtenue à partir des distances entre les centres de masse des différentes régions (Fig IV.10).

En présence de C-CaM, les centres de masse de CA et SA restent à une distance stable de 32 à 33 Å (fig IV.10 a), tandis que dans AC-isolé, la distance CA/SA oscille et décroît d'environ 4 Å. Cette grande variation de la distance est en accord avec les mouvements observés dans AC-isolé par l'analyse en composantes principales de la matrice de covariance (fig IV.9 c). La distance entre les centres de masse de la partie C-terminale et de C-CaM augmente d'environ 2 Å lorsque les ions  $\text{Ca}^{2+}$  sont retirés de C-CaM (fig IV.10 b). Cette augmentation est en accord avec la dérive conformationnelle plus grande de l'extrémité C-terminale observée pour AC-(0Ca-C-CaM) comparée à AC-(2Ca-C-CaM) (fig IV.3 a et b). Au cours de la trajectoire AC-(2Ca-C-CaM), la distance entre les centres de masse de CA et CB (fig IV.10 c) demeure stable autour de 27-28 Å. Elle augmente dans AC-(0Ca-C-CaM) et diminue légèrement dans AC-isolé. Les positions relatives de CA et de CB sont donc stables dans les systèmes étudiés.



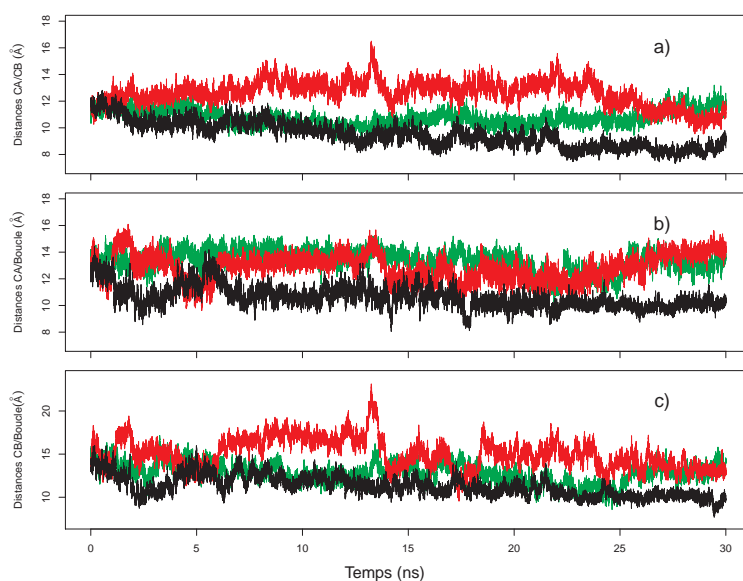
**FIG. IV.10:** Variations des distances entre les centres de masse de : (a) SA et CA, (b) C-CaM et C-ter, (c) CA et CB. Les courbes correspondent à : AC-(2Ca-C-CaM) en vert, AC-(0Ca-C-CaM) en rouge et AC-isolé en noir.

Pour établir un lien entre les tendances conformationnelles de AC et son activité catalytique potentielle, une analyse de la déformation du site catalytique a été faite en calculant les distances entre les résidus catalytiques suivants [131] : K58, H63, K84, D188, D190, H298, N304, et F306. Trois centres de masse,  $CM_{CA}$ ,  $CM_{CB}$  et  $CM_{CLoop}$  sont respectivement calculés à partir de : K58, D188, D190 et H298 situés dans le sous-domaine CA, H63 et K84 situés dans CB, N304 et F306 situés dans la boucle catalytique.

Lorsque les deux ions  $Ca^{2+}$  sont retirés (trajectoire AC-(0Ca-C-CaM)), CB s'éloigne légèrement de CA et la boucle C, comme le montrent les calculs des distances qui augmentent entre  $CM_{CA}$  et  $CM_{CB}$  (fig IV.11 a) et entre  $CM_{CLoop}$  et  $CM_{CB}$  (Fig IV.11 c), alors que la distance entre  $CM_{CA}$  et  $CM_{CLoop}$  reste constante IV.11(b).

Dans AC-isolé, toutes les distances décroissent, en particulier la distance entre  $CM_{CA}$  et  $CM_{CB}$  qui présente une diminution de 4 Å. Ces variations dans la géométrie du site catalytique peuvent être reliées à l'efficacité catalytique à travers la surface catalytique accessible au solvant. En effet, une surface catalytique petite restreint l'accès au site catalytique, tandis qu'une surface accessible plus grande facilite la liaison du substrat.

Dans les trajectoires de dynamique moléculaire analysées ici, la surface catalytique oscille entre une valeur moyenne de 527 Å<sup>2</sup> pour AC-(2Ca-C-CaM), une valeur maximale de 614 Å<sup>2</sup> pour AC-(0Ca-C-CaM), et une valeur minimale de 460 Å<sup>2</sup> pour AC-isolé (table III).



**FIG. IV.11:** Variations de la géométrie du site catalytique décrites par les distances entre les centres de masse : (a)  $CM_{CA}$  et  $CM_{CB}$ , (b)  $CM_{CA}$  et  $CM_{Loop}$ , (c)  $CM_{CB}$  et  $CM_{Loop}$ . Les centres de masse,  $CM_{CA}$ ,  $CM_{CB}$  et  $CM_{Loop}$  sont respectivement calculés à partir de : K58, D188, D190 et H298 situés dans CA, H63 et K84 situés dans CB, N304 et F306 situés dans la boucle catalytique. Les courbes correspondent à AC-(2Ca-C-CaM) en vert, AC-(0Ca-C-CaM) en rouge et AC-isolé en noir.

trajectoire	AC-(2Ca-C-CaM)	AC-(0Ca-C-CaM)	AC-isolé
Surface catalytique ( $\text{Å}^2$ )	$527.3 \pm 54.4$	$614.7 \pm 50.2$	$460.5 \pm 91.7$
main EF angle 3 ( $^\circ$ )	$88 \pm 2.5$	$92 \pm 2.8$	-
main EF angle 4 ( $^\circ$ )	$98.7 \pm 3.2$	$98.3 \pm 4.1$	-
Surface accessible de pièce hydrophobe ( $\text{Å}^2$ )	$764.0 \pm 24.5$	$785.3 \pm 32.1$	-
Distance main EF 3 $\alpha$ hélice H ( $\text{Å}$ )	$11.5 \pm 0.4$	$10.5 \pm 0.3$	-
Distance main EF 4 $\alpha$ hélice H ( $\text{Å}$ )	$11.6 \pm 0.3$	$12.1 \pm 0.3$	-

**TAB. III:** Analyse de la structure du site catalytique de l'adényl cyclase et de C-CaM au cours des trajectoires AC-(2Ca-C-CaM), AC-(0Ca-C-CaM) et AC-isolé. La surface catalytique est la surface accessible au solvant ( $\text{Å}^2$ ) calculée avec MSMS 2.5.3 [204] sur les résidus catalytiques K58, H63, K84, D188, D190, H298, N304 and F306 définis dans la Figure 5c de Ref. [131]. La structure de C-CaM est caractérisée par : (i) la surface accessible au solvant ( $\text{Å}^2$ ) des résidus I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, A147, définissant la partie hydrophobe [108], (ii) les angles moyens entre les axes des hélices  $\alpha$  V et VI (résidus 82-92 et 102-111) et des hélices  $\alpha$  VII et VIII (résidus 118-128 et 138-146), définissant l'ouverture des mains EF 3 et 4, (iii) les distances minimales ( $\text{Å}$ ) entre les centres de masse des mains EF et l'axe de l'hélice H.

La géométrie de la poche de AC-(0Ca-C-CaM) est donc compatible avec une efficacité catalytique similaire à celle de AC-(2Ca-C-CaM), conformément aux observations expérimentales [39] montrant l'absence d'influence des ions  $\text{Ca}^{2+}$  sur l'activité catalytique de AC. Comme attendu, la surface la plus ouverte dans AC-(0Ca-C-CaM) correspond à la plus grande distance observée

entre les centres de masse de CA et CB (fig IV.10 c). A l'inverse, dans AC-isolé, la géométrie (fig IV.11) tout comme la surface accessible au solvant du site catalytique (table III) révèlent une fermeture du site, gênant ainsi la liaison du substrat.

Il est à noter que pour le facteur œdémateux (EF) de *B. anthracis*, [185], la délétion des ions  $Ca^{2+}$  a un effet opposé par rapport à ce qui est observé pour AC/C-CaM, dont la taille du site catalytique passe de 928 à 501 Å<sup>2</sup>. Cette fermeture du site catalytique dans le facteur œdémateux EF est en accord avec la réduction de son activité catalytique pour des concentrations en  $Ca^{2+}$  décroissantes [39; 205].

L'analyse des tendances conformationnelles globales et de la géométrie du site catalytique de AC permet de relier la dérive conformationnelle importante décrite par AC-isolé avec la fermeture du site catalytique. A l'inverse, l'absence ou la présence des ions  $Ca^{2+}$  n'a qu'un effet limité sur l'architecture globale du complexe et de l'efficacité catalytique.

### 3.3 Conformation de C-CaM et interaction avec AC

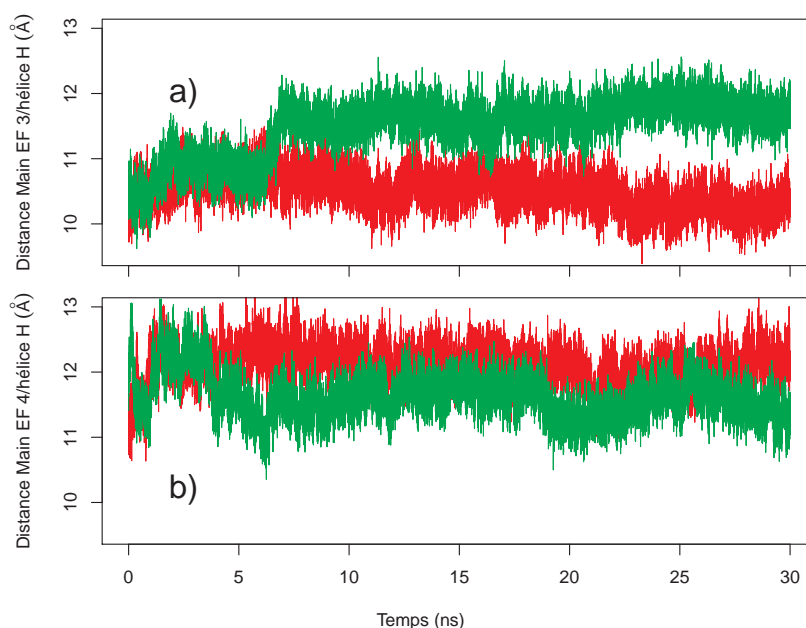
Comme évoqué dans la section "Dérives conformationnelles", la conformation de C-CaM décrit seulement peu de variations au cours des simulations. Plusieurs paramètres ont été choisis (table III pour obtenir une image plus détaillée de cette variation : l'ouverture des mains EF [121; 206], la surface accessible au solvant des parties hydrophobes [108], et la coordination des ions  $Ca^{2+}$ . L'ouverture des mains EF 3 et 4 de C-CaM a été déterminée à partir des angles 3 et 4 entre les axes des hélices  $\alpha$  V et VI, et entre les axes des hélices  $\alpha$  VII et VIII.

Dans AC-(2Ca-C-CaM), les angles 3 et 4 sont stables autour de 88 et 99° respectivement, ces valeurs étant à peine plus grandes que celles observées dans la structure cristallographique 1YRT (87 et 85°). Dans AC-(0Ca-C-CaM), la main EF 3 a tendance à s'ouvrir avec un angle d'environ 92°. Cette ouverture est en accord avec les importantes fluctuations observées dans la boucle  $Ca^{2+}$  de la main EF 3 (fig IV.5 b), mais reste paradoxal par rapport à la fermeture des mains EF normalement observée dans la CaM isolée en l'absence d'ions  $Ca^{2+}$  [121; 206]. Dans notre cas, la fermeture est gênée par l'interaction de la C-CaM avec AC, tout comme il a été observé dans les simulations de dynamique moléculaire réalisées sur le facteur œdémateux EF de *B. anthracis* [185].

D'après la littérature [108], les pièces hydrophobes de C-CaM (I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, A147) devraient être plus accessibles si les mains EF sont plus ouvertes. C'est en effet le cas dans nos simulations, qui présentent une légère augmentation de la surface accessible au solvant des parties hydrophobes (table III) passant de 764 Å<sup>2</sup> dans AC-(2Ca-C-CaM) à 785 Å<sup>2</sup> dans AC-(0Ca-C-CaM), ce qui est en accord avec l'ouverture de l'angle 3 décrit précédemment. Cette augmentation de la surface de la partie hydrophobe indique une interaction plus rapprochée entre C-CaM et AC. Il est à noter que la partie hydrophobe contient également les méthionines 109, 124, 144 et 145, dont l'oxydation réduit considérablement l'affinité de CaM pour AC [136]. L'oxydation des méthionines peut ainsi gêner la variation de la surface des parties et induire par conséquent une rupture de la liaison de C-CaM avec AC. L'augmentation de

la surface de la partie est opposée à la réduction de la surface observée dans le complexe EF/CaM [185], lorsque les ions  $\text{Ca}^{2+}$  sont retirés (Table V de [185]). Ainsi, pour le facteur œdémateux tout comme pour AC, la surface de la partie hydrophobe décrit des variations qui sont en accord avec les variations du site catalytique décrites dans la sous-section "Tendances conformationnelles de AC et activité catalytique" : une plus grande surface de la partie hydrophobe correspond à un site catalytique plus ouvert, et une plus petite surface à un site catalytique plus fermé.

Les mains EF 3 et 4 de C-CaM interagissent de façon parallèle avec l'hélice  $\alpha$  H (figure IV.1), la main EF 3 étant plus proche de la région CA et la main EF 4 plus proche de l'extrémité de SA. Parmi ces deux mains EF, la main EF 4 semble avoir une interaction plus forte avec l'hélice. En effet, comme souligné dans la sous-section "Dérives conformationnelles", la délétion des ions  $\text{Ca}^{2+}$  induit une augmentation de la flexibilité dans la partie C-terminale de C-CaM (fig IV.5 b) qui est incluse dans la main EF 4. De plus, dans AC-(2Ca-C-CaM), la distance de la main EF 4 à l'axe de l'hélice  $\alpha$  H oscille entre 10 et 12 Å (fig IV.12 b) au cours de la trajectoire, tandis que la distance impliquant la main EF 3 montre une transition nette autour de 7 ns (fig IV.12 a). Cette transition est produite par une petite rotation de l'hélice  $\alpha$  H de SA.



**FIG. IV.12:** Variations des distances minimales (Å) entre le centre de masse de C-CaM (a) mains EF 3 ou (b) main EF 4 et l'axe de l'hélice  $\alpha$  H de SA. Les courbes correspondent à AC-(2Ca-C-CaM) en vert et AC-(0Ca-C-CaM) en rouge.

La coordination des ions  $\text{Ca}^{2+}$  dans le lobe C-CaM (table IV) est similaire aux observations précédemment faites dans la structure 100J de  $\text{Ca}^{2+}$ -CaM isolée (Table II de [185]).

La coordination des ions calcium dans la main EF 3 implique les chaînes latérales de D93, D95, N97, et E104 et le carbonyle de Y99. La coordination dans la main EF 4 implique les chaînes

Main EF	Atome	distance moyenne (Å)	Minimum (Å)	Maximum (Å)
3	Asp-93 O $\delta$ 1/O $\delta$ 2	$2.6 \pm 0.1$	2.3	3.1
3	Asp-95 O $\delta$ 1/O $\delta$ 2	$2.6 \pm 0.1$	2.3	3.0
3	Asn-97 O $\delta$ 1	$2.7 \pm 0.1$	2.4	3.5
3	Tyr-99 O	$2.6 \pm 0.1$	2.4	3.3
3	Glu-104 O $\epsilon$ 1/O $\epsilon$ 2	$2.6 \pm 0.1$	2.4	2.9
4	Asp-129 O $\delta$ 2	$2.6 \pm 0.1$	2.3	3.2
4	Asp-131 O $\delta$ 1/O $\delta$ 2	$2.6 \pm 0.1$	2.3	3.0
4	Asp-133 O $\delta$ 1/O $\delta$ 2	$3.1 \pm 1.1$	2.3	7.4
4	Asp-133 O	$2.9 \pm 0.5$	2.4	6.1
4	Gln-135 O	$2.7 \pm 0.1$	2.4	3.5

**TAB. IV:** Coordination des ions  $\text{Ca}^{2+}$  dans AC-(2Ca-C-CaM). Les atomes de la boucle calcium sont détectés à une distance plus proche que 2.8 Å. Pour les distances impliquant les couples d'atomes O $\delta$ 1/O $\delta$ 2 et O $\epsilon$ 1/O $\epsilon$ 2, seule la plus petite des deux distances est relevée. Pour chaque interaction de coordination, la distance moyenne et la déviation standard sont indiquées, ainsi que les valeurs minimales et maximales des distances au cours de la trajectoire.

latérales de D129, D131, et D133 et les carbonyles de D133 et D135. Les liaisons hydrogène formées par les chaînes latérales de D131 et D133 sont plus faibles que les autres, ce qui indique une coordination plus faible des ions  $\text{Ca}^{2+}$  par la main EF 4, et peut être mis en parallèle avec l'interaction plus importante entre la main EF 4 et l'hélice  $\alpha$  H (fig IV.10 b).

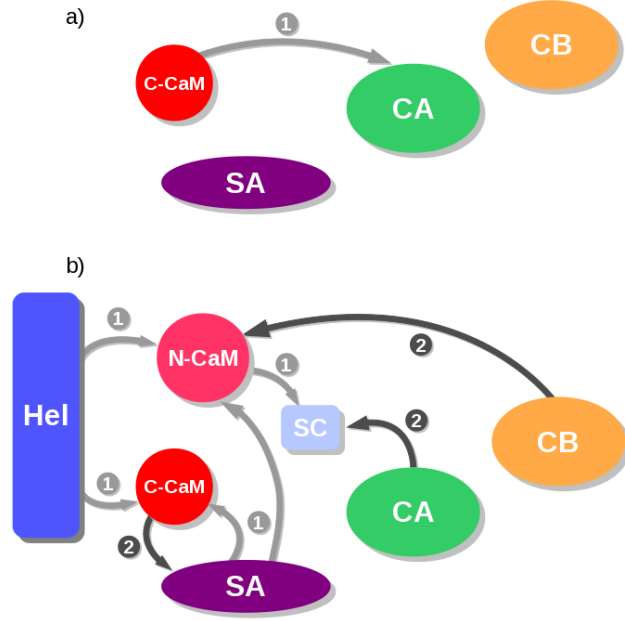
La C-CaM ne montre pas beaucoup de variabilité conformationnelle, contrairement aux observations faites précédemment sur la structure du complexe entre CaM et le facteur œdémateux de *B. anthracis* [185; 149]. La mobilité majeure observée sur C-CaM est l'ouverture de la main EF 3 lorsque les ions  $\text{Ca}^{2+}$  sont retirés. Ce mouvement est contradictoire par rapport au comportement habituel de la CaM en solution, et a tendance à prouver que l'interaction avec AC rend les sous-états conformationnels de C-CaM différents de ceux observés pour la CaM en solution. Dans les petites variations conformationnelles du lobe C-CaM, les mains EF 3 et 4 ne partagent pas les mêmes caractéristiques. La main EF 3, plus proche de la queue C-terminale et de la région CA, lie les ions  $\text{Ca}^{2+}$  plus solidement, et est plus sensible à la délétion des  $\text{Ca}^{2+}$ .

### 3.4 Influences énergétiques

Les influences énergétiques entre les régions SA, CA, CB et C-CaM ont été calculées à partir des différentes trajectoires comme décrit dans le Matériel et Méthodes. Aucune influence énergétique n'a été identifiée au cours de la trajectoire AC-isolé, tandis qu'une seule influence de C-CaM sur CA a été observée dans AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM) (fig IV.13 a).

Des variations similaires dans l'énergie MMPBSA ont été observées pour AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM). Dans le complexe entier *C*, les régions SA et C-CaM ont des énergies d'interaction  $\Delta\Delta G[D, C]$  de référence favorables, alors que CA et CB présentent des énergies d'interaction défavorables. Néanmoins, CA est classé à une meilleure position que CB, comme le montrent les





**FIG. IV.13: Schéma des influences énergétiques** pour (a) AC-(2Ca-C-CaM), AC-(0Ca-C-CaM) et (b) la simulation de DM précédemment réalisée [186] sur le complexe EF/CaM en présence des ions  $\text{Ca}^{2+}$ . Les régions du complexe sont représentées par des figures géométriques. Les flèches indiquent l'influence énergétique d'un domaine sur l'autre, l'intensité de l'influence étant marquée par le nombre de rangs perdus par le domaine influencé.

tests non-paramétriques de Mann-Whitney [196] (Tables V et VI). Dans le complexe  $C \setminus C\text{-CaM}$ , toutes les régions affichent des énergies d'interaction  $\Delta\Delta G[D, C \setminus C\text{-CaM}]$  défavorables.

(a) C	$\Delta\Delta E_{ele}$	$\Delta\Delta E_{vdw}$	$\Delta\Delta E_{int}$	$\Delta\Delta G_{sol}^{polar}$	$\Delta\Delta G_{sol}^{nonpolar}$	$\Delta\Delta G[D, C]$
CB (4)	$-215 \pm 55$	$-77 \pm 6$	$24 \pm 3$	$310 \pm 53$	$14 \pm 2$	$56 \pm 17$
CA (3)	$-51 \pm 59$	$-191 \pm 8$	$47 \pm 4$	$213 \pm 54$	$28 \pm 3$	$46 \pm 22$
SA (2)	$-1801 \pm 47$	$-144 \pm 7$	$23 \pm 3$	$1816 \pm 42$	$6 \pm 2$	$-100 \pm 15$
C-CaM (1)	$-892 \pm 55$	$-149 \pm 10$	$0 \pm 0$	$888 \pm 45$	$12 \pm 2$	$-141 \pm 17$
(b) $C \setminus C\text{-CaM}$	$\Delta\Delta E_{ele}$	$\Delta\Delta E_{vdw}$	$\Delta\Delta E_{int}$	$\Delta\Delta G_{sol}^{polar}$	$\Delta\Delta G_{sol}^{nonpolar}$	$\Delta\Delta G$
CA (3)	$-680 \pm 59$	$-132 \pm 7$	$47 \pm 4$	$828 \pm 56$	$18 \pm 2$	$82 \pm 18$
CB (2)	$-241 \pm 53$	$-77 \pm 6$	$24 \pm 3$	$337 \pm 51$	$14 \pm 2$	$57 \pm 17$
SA (1)	$-428 \pm 15$	$-55 \pm 3$	$23 \pm 3$	$481 \pm 14$	$6 \pm 1$	$27 \pm 7$

**TAB. V: Analyse des influences énergétiques de AC-(2Ca-C-CaM) avec la suppression de C-CaM.** (a) Les énergies libres de liaison de référence  $\Delta\Delta G[C, D]$  des quatre régions C-CaM, SA, CA et CB sont exprimées en kcal/mol avec leur contributions : 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 ( $\Delta\Delta G_{sol}^{polar}$ ) et l'énergie de solvation non-polaire ( $\Delta\Delta G_{sol}^{nonpolar}$ ). Le rang de chaque région dans la liste des énergies d'interaction est inscrit entre parenthèses avec le nom de la région. Le rang entre CB et CA a été vérifié par le test non-paramétrique de Mann-Whitney [196], réalisé sur les estimations de l'énergie au cours du temps, avec le logiciel R [197] (commande *wilcox.test*). Une p-value de  $2.2 \cdot 10^{-16}$  a été obtenue avec l'hypothèse alternative  $\Delta\Delta G(CA) < \Delta\Delta G(CB)$ , ce qui attribue un rang 3 pour CA et un rang 4 pour CB. (b) Energies d'interaction perturbées  $\Delta\Delta G[D, C \setminus C\text{-CaM}]$  et contributions correspondantes obtenues pour C-CaM, SA, CA et CB sur le système tronqué  $C \setminus C\text{-CaM}$ . Le rang entre CB et CA a été vérifié par le même test non-paramétrique Mann-Whitney [196]. Une p-value de  $2.2 \cdot 10^{-16}$  a été obtenue avec l'hypothèse alternative  $\Delta\Delta G(CA) > \Delta\Delta G(CB)$ , ce qui attribue le rang 2 pour CB et le rang 3 pour CA et donc une perte de 1 pour le rang de CA.

(a) C	$\Delta\Delta E_{ele}$	$\Delta\Delta E_{vdw}$	$\Delta\Delta E_{int}$	$\Delta\Delta G_{sol}^{polar}$	$\Delta\Delta G_{sol}^{nonpolar}$	$\Delta\Delta G[D, C]$
CB (4)	$-131 \pm 50$	$-79 \pm 6$	$24 \pm 3$	$210 \pm 43$	$12 \pm 2$	$36 \pm 19$
CA (3)	$304 \pm 81$	$-177 \pm 11$	$46 \pm 4$	$-189 \pm 70$	$28 \pm 3$	$12 \pm 20$
SA (2)	$-2045 \pm 86$	$-163 \pm 10$	$22 \pm 3$	$2074 \pm 74$	$10 \pm 2$	$-102 \pm 21$
C-CaM (1)	$-842 \pm 67$	$-153 \pm 9$	$0 \pm 0$	$827 \pm 65$	$14 \pm 2$	$-154 \pm 22$
(b) C\C-CaM	$\Delta\Delta E_{ele}$	$\Delta\Delta E_{vdw}$	$\Delta\Delta E_{int}$	$\Delta\Delta G_{sol}^{polar}$	$\Delta\Delta G_{sol}^{nonpolar}$	$\Delta\Delta G$
CA (3)	$-625 \pm 35$	$-133 \pm 7$	$46 \pm 4$	$747 \pm 35$	$18 \pm 3$	$54 \pm 17$
CB (2)	$-186 \pm 53$	$-79 \pm 6$	$24 \pm 3$	$266 \pm 36$	$12 \pm 2$	$37 \pm 18$
SA (1)	$-444 \pm 36$	$-54 \pm 4$	$22 \pm 3$	$487 \pm 30$	$7 \pm 1$	$20 \pm 10$

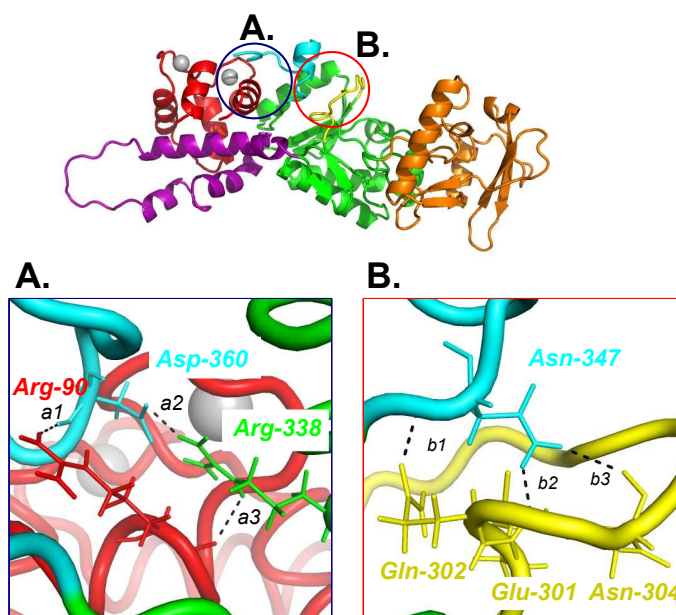
**TAB. VI: Analyse des influences énergétiques de AC-(0Ca-C-CaM) lorsque la C-CaM est retirée.** (a) Les énergies libres de liaison de référence  $\Delta\Delta G[C, D]$  des quatre régions C-CaM, SA, CA et CB sont exprimées en kcal/mol en fonction de leurs contributions : en énergie d'interaction Coulomb ( $\Delta\Delta E_{ele}$ ), énergie de van der Waals ( $\Delta\Delta E_{vdw}$ ), l'énergie interne ( $\Delta\Delta E_{int}$ ), l'énergie de solvatation polaire ( $\Delta\Delta G_{sol}^{polaire}$ ) et l'énergie de solvatation non-polaire ( $\Delta\Delta G_{sol}^{nonpolaire}$ ). Le rang de chaque région dans la liste des énergies d'interaction est écrit entre parenthèses. Le rang entre CB et CA a été vérifié par un test non-paramétrique Mann-Whitney [196] réalisé sur les estimations d'énergie au cours du temps avec le logiciel R [197] (commande *wilcox.test*). Une p-valeur de  $2.2 \cdot 10^{-16}$  a été obtenu avec l'hypothèse alternative  $\Delta\Delta G(CA) < \Delta\Delta G(CB)$ , ce qui attribue un rang 3 pour CA et un rang 4 pour CB. (b) Energies d'interaction perturbées  $\Delta\Delta G[D, C \setminus C-CaM]$  et les contributions correspondantes obtenues pour C-CaM, SA, CA et CB sur le système tronqué C\C-CaM. Les rangs entre CB et CA et entre SA et CB ont été vérifié par le même test non-paramétrique Mann-Whitney [196] que précédemment. Des p-valeurs de  $2.2 \cdot 10^{-16}$  ont été obtenues avec l'hypothèse alternative  $\Delta\Delta G(CA) > \Delta\Delta G(CB)$  et  $\Delta\Delta G(CB) > \Delta\Delta G(SA)$ , attribuant ainsi le rang 1 pour SA, 2 pour CB et 3 pour CA et une perte de 1 pour le rang de CA.

Les différentes composantes de l'énergie d'interaction de CB sont similaires dans C et C\C-CaM, tandis que les valeurs absolues des composantes de SA diminuent, l'énergie de solvatation étant divisée par 4. En dépit de la réduction des composantes de l'énergie, SA conserve le meilleur rang dans la liste des interactions, alors que les valeurs absolues de l'énergie de solvatation polaire  $\Delta\Delta G_{sol}^{polar}$  et de l'énergie électrostatique  $\Delta\Delta G_{ele}$  de CA sont augmentées. C'est ce qui est responsable de la perte d'un rang de CA (Tables V et VI). La réaction des différentes régions par rapport à la suppression des ions  $Ca^{2+}$  est en accord avec leurs positions relatives par rapport à C-CaM : CB se situe du côté opposé dans le complexe, tandis que CA et SA forment des interactions directes avec C-CaM.

L'unique influence observée dans AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM) ainsi que les plus grandes énergies de liaison  $\Delta\Delta G_{sol}^{polar}$  et  $\Delta\Delta E_{ele}$  soutient le rôle essentiel des ions  $Ca^{2+}$  et de C-CaM dans l'activation complète de AC. La perte d'un rang de CA est due dans les deux cas à l'importante variation des énergies  $\Delta\Delta G_{sol}^{polar}$  et  $\Delta\Delta G_{ele}$ , mais  $\Delta\Delta G_{ele}$  varie plus dans le cas de AC-(0Ca-C-CaM) (Tables V et VI). Cet effet peut être mis en parallèle avec la rupture des liaisons hydrogène entre les résidues de la queue C-terminale (N347, D360), CA (R338), la boucle catalytique (E301, Q302, N304), et C-CaM (R90) (figure IV.14).

Ces liaisons hydrogène sont présentes dans AC-(2Ca-C-CaM), encore partiellement présentes dans AC-(0Ca-C-CaM), et principalement rompues dans AC-isolé (Table VII).

La rupture de ces liaisons hydrogène est en accord avec l'augmentation de la distance entre les centres de masse de la queue C-terminale et de la C-CaM (Fig IV.10 b). Les liaisons hydrogène



**FIG. IV.14:** Vue d'ensemble de la structure du complexe AC-C-CaM (1YRT) représenté en cartoon avec les zooms A et B à l'interface entre C-CaM, la queue C-terminale et la boucle catalytique. Le lobe C-CaM est coloré en rouge, la queue C-terminale en cyan, la boucle catalytique en jaune, le reste de CA en vert, et les ions  $\text{Ca}^{2+}$  sont représentés en sphères grises. Les résidus impliqués dans les liaisons hydrogène dans AC-(2Ca-C-CaM) et pour lesquels les liaisons hydrogène sont rompues dans AC-(0Ca-C-CaM) et AC sont représentés en bâtons et colorés en fonction de la région du complexe à laquelle ils appartiennent. Les liaisons hydrogène correspondantes sont notées  $a1$ ,  $a2$ ,  $a3$  pour l'interaction C-CaM/queue C-terminale/CA et  $b1$ ,  $b2$ ,  $b3$  pour l'interaction queue C-terminale/boucle catalytique.

Liaison hydrogène	Partenaires	AC-(2Ca-C-CaM) (%)	AC-(0Ca-C-CaM) (%)	AC-isolé (%)
$a1$	D360-O/R90-H $\eta$ 21	97.9	76.1	-
$a2$	R338-H $\eta$ 12/D360-O $\delta$ 2	44.9	40.8	32.0
$a3$	R338-H $\epsilon$ /R90-O	84.2	13.8	-
$b1$	Q302-O $\epsilon$ 1/N347-H	92.3	0.1	81.5
$b2$	E301-O/N347-H $\delta$ 22	98.8	0.0	21.8
$b3$	N304-O/N347-H $\delta$ 21	98.9	3.8	0.9

**TAB. VII:** Pourcentages de formation des liaisons hydrogène connectant C-CaM, la région CA, la queue C-terminale, et la boucle catalytique. Les noms correspondent aux noms de la Figure IV.14, les partenaires de liaison hydrogène et le pourcentage de temps au cours des trajectoires, lorsque la distance est inférieure à 2.5 Å

formées entre R90 de C-CaM et R326, N336 de AC impliquant tous les groupes carboxyles et aminés de ces chaînes latérales rappellent les ponts salins observés entre K525 du facteur œdémateux et E114 de CaM [128].

L'absence d'influences énergétiques lorsque la CaM et les ions  $\text{Ca}^{2+}$  sont retirés, est en accord avec la déstabilisation expérimentalement observée de CyaA isolée. En effet, la protéine isolée

semble avoir des difficultés à cristalliser et il a été montré qu'elle s'agrège partiellement en solution (Alexandre Chenal, communication personnelle). Les contributions enthalpiques, considérées dans la présente analyse, devraient jouer un rôle plus limité dans les influences énergétiques.

Le schéma des influences énergétiques a été précédemment interprété [186] comme donnant une description d'un schéma dépendant de la stabilité entre les différentes parties du complexe. Comme dans l'architecture d'un château de cartes : plus une région est influencée par les autres, plus elle occupe une faible place dans la hiérarchie des stabilités relatives au sein du complexe. Dans ce cadre, l'unique influence de C-CaM sur CA observée pour AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM) démontre que CA occupe la plus petite place dans la hiérarchie des stabilités des régions de AC/C-CaM, et est stabilisé par C-CaM. La région CA est donc le meilleur candidat pour des modifications ayant pour but de moduler l'interaction C-CaM/AC en présence et en absence d'ions  $Ca^{2+}$ . Pour AC-isolé, aucune influence énergétique n'est observée, et par conséquent aucune stabilité hiérarchique. Le diagramme correspondant des influences énergétiques précédemment obtenu pour le complexe facteur œdémateux/CaM (Fig IV.13 b) [186], était bien plus complexe que celui obtenu pour le complexe AC/C-CaM (Fig IV.13 a). Ceci peut être relié à une barrière énergétique plus grande pour l'interaction EF/CaM par rapport à l'interaction AC/CaM, cette barrière induisant une hiérarchie des stabilités des régions. En effet, l'affinité du facteur œdémateux pour la CaM est d'environ 20 nM [39], alors que l'affinité de AC pour la CaM est d'environ 0.2 nM [131; 135].

## 4 Discussion

Le domaine catalytique AC de CyaA de *Bordetella pertussis* a été simulé en complexe avec la C-CaM chargée d'ions  $\text{Ca}^{2+}$  et avec l'apo-C-CaM, mais également en tant que protéine isolée. Le choix de simuler le complexe avec seulement une portion de la calmoduline est justifié par le fait que seule l'interaction de la C-CaM avec AC a pu être résolue par cristallographie [131], et par des mesures expérimentales prouvant que la C-CaM suffit pour activer complètement AC [39].

L'étude du domaine AC par des simulations de dynamique moléculaire a été initialement motivée par le fait que des informations sur la structure de la protéine isolée à l'équilibre et en solution demeurent parcellaires. En effet, il n'y a à ce jour aucune structure cristallographique de AC isolé qui ait été résolue. Seules des mesures hydrodynamiques [77] révèlent que la protéine isolée est plus sphérique et hydratée que le complexe AC/CaM.

L'objectif de cette étude est extrêmement ambitieux, du fait que les deux partenaires subissent des changements conformationnels majeurs lorsque leur interaction est rompue. Par conséquent, une trajectoire de DM enregistrée dans un intervalle de 30 ns a peu de chances d'atteindre l'équilibre. En effet, l'apo-C-CaM ne se dissocie pas du domaine AC, et ce dernier ne parvient pas, en l'absence de C-CaM, à atteindre une forme sphérique comparable aux mesures hydrodynamiques [77].

Cependant, la durée relativement courte des simulations n'empêche pas une analyse pertinente de la réaction de AC par rapport aux différentes perturbations suivantes : la suppression des ions  $\text{Ca}^{2+}$  ou du lobe C-terminal de la calmoduline (C-CaM) chargée d'ions  $\text{Ca}^{2+}$ . En effet, nous cherchons ici à caractériser les tendances conformationnelles relatives du domaine AC, et non ses dérives conformationnelles absolues. C'est pourquoi des dynamiques moléculaires de quelques dizaines de nanosecondes dans trois conditions différentes suffisent pour en extraire des informations importantes. De plus, ces simulations ont été répétées une seconde fois avec une graine différente et aboutissent à des résultats similaires.

L'étude présentée ici a permis d'obtenir des informations plus précises sur ce point, comme les caractéristiques structurales de l'activation de AC. Dans un premier temps, les préférences conformationnelles décrites par AC au cours des trajectoires sont qualitativement en accord avec les observations faites par Karst et ses collaborateurs [77], étant donné que l'évolution globale de la structure tend vers une forme plus globulaire. Cependant, à cause de la durée assez limitée de l'intervalle exploré au cours des simulations, seul le début de la transition vers une forme sphérique et hydratée a pu être échantillonné. L'analyse de la convergence par l'approche de Lyman et Zuckerman [192] faite sur les 15 à 30 ns de AC-isolé, a montré une tendance de ce dernier à rester dans un seul bassin conformationnel. Ceci signifie que la simulation de dynamique moléculaire a atteint sa limite dans l'exploration de l'espace conformationnel de AC, et que des méthodes plus adaptées à une exploration de l'espace conformationnel plus étendue devraient être envisagées pour surmonter les barrières conformationnelles.

Dans un deuxième temps, les simulations de DM ont apporté des informations précises sur le rôle relatif des ions  $\text{Ca}^{2+}$  et de C-CaM dans l'activation de l'adénylcyclase. D'après les informations

expérimentales antérieures aux simulations, ni les ions  $Ca^{2+}$ , ni même la CaM entière ne sont nécessaires pour l'activation [39]. A partir de l'analyse des simulations, nous pouvons déduire que l'interaction C-CaM/AC décrit deux aspects différents. Le premier aspect est que l'interaction de C-CaM avec la région SA est essentiellement basée sur un effet stérique, rapprochant SA de l'autre extrémité de AC. En effet, en l'absence de C-CaM, le déplacement de SA induit une déstabilisation importante de CA qui ferme le site catalytique. Le second aspect est que les ions  $Ca^{2+}$  n'ont qu'un faible effet direct sur le site catalytique, étant donné que le site s'ouvre légèrement en l'absence des ions, ce qui est en accord avec l'observation que AC peut toujours être activé en l'absence de Calcium [39]. Cependant, conformément aux observations expérimentales [39], les ions  $Ca^{2+}$  facilitent l'interaction entre C-CaM et AC. L'influence de C-CaM et des ions  $Ca^{2+}$  est médiée par les interactions entre C-CaM et la queue C-terminale, comme le montrent : (i) le diagramme des influences énergétiques, (ii) la corrélation des liaisons hydrogène dans C-CaM, la queue C-terminale et CA, avec la présence de C-CaM et des ions  $Ca^{2+}$ , (iii) l'augmentation des dérivées conformationnelles de la queue C-terminale et de la boucle C en l'absence des ions, (iv) les distances entre les centres de masse de C-CaM et de la queue C-terminale. L'importance de la queue C-terminale et de la boucle catalytique correspond avec l'équivalence établie par Guo et ses collaborateurs [131], entre ces régions et les switch C et B qui subissent des grands changements conformationnels durant l'activation du facteur œdémateux [128].

Le domaine AC de l'adénylcyclase de *Bordetella pertussis* partage quelques caractéristiques communes avec le facteur œdémateux de *Bacillus anthracis*, en particulier ce qui concerne son activation par une interaction avec la calmoduline. Cependant, la détermination des deux structures de AC [131] et du facteur œdémateux [128] a révélé une différente organisation structurale, en accord avec le faible pourcentage d'identité de séquence de 25 % entre AC et le facteur œdémateux. L'analyse des dynamiques moléculaires du domaine AC en présence et en absence de C-CaM et des ions  $Ca^{2+}$  confirme que les deux adénylcyclases se comportent différemment. Tout d'abord, la suppression des ions  $Ca^{2+}$  ouvre le site catalytique, alors qu'elle le ferme pour le facteur œdémateux. De plus, dans la structure du facteur œdémateux, le switch C, équivalent à la queue C-terminale de AC, est lié covalamment au domaine hélicale d'environ 150 résidus, qui a pour rôle de bloquer le facteur œdémateux dans un état inactif [39]. Dans AC, la queue C-terminale est plus libre de ses mouvements. Ceci explique la barrière d'activation plus petite de AC par rapport au facteur œdémateux. L'interaction CaM/C-ter est donc un point important de l'activation, comme l'indiquent les influences énergétiques et le réseau de liaisons hydrogène.

## 5 Conclusion

Le domaine catalytique AC de l'adénylcyclase CyaA de *Bordetella pertussis* a été simulé dans différents environnements d'interaction. L'interaction entre AC et la calmoduline présentent deux aspects majeurs. En premier, la calmoduline bloque la région SA de façon à garder la géométrie du site catalytique dans une forme activée, en stabilisant la région CA. En second, les ions  $Ca^{2+}$

jouent un rôle indirect dans l'activation, stabilisant l'interaction C-CaM/AC en induisant la formation d'un réseau de liaisons hydrogène partant de la C-CaM jusqu'à la boucle catalytique à travers la queue C-terminale. Cette description est qualitativement en accord avec les informations expérimentales cinétiques [39] et hydrodynamiques [77], précédemment rapportées sur ce système, et met en évidence les différents rôles de CaM et des ions  $\text{Ca}^{2+}$  dans l'activation du domaine AC.

## Chapitre V

# Influence de mutations introduites dans AC sur son affinité pour la calmoduline

---

## 1 Introduction

Les interactions intermoléculaires jouent un rôle fondamental en biologie. Les interactions ADN-protéine et ARN-protéine régulent l'expression des gènes, la réparation, la recombinaison ou encore la réplication, et les interactions protéine-protéine sont impliquées dans la signalisation cellulaire, dans le transport et le stockage, dans la phosphorylation protéique, ainsi que dans les réactions enzymatiques.

Depuis quelques années, l'étude des interactions protéine-protéine a été utilisée pour analyser les propriétés globales des cellules et des bactéries [207], dans le cadre d'une approche de biologie des systèmes. Les interactions protéine-protéine jouent également un rôle important dans la virulence des pathogènes [208; 209; 210], et pourraient être une cible de choix pour développer de nouvelles approches pour la lutte contre les effets pathogènes de ces organismes [211; 184; 212].

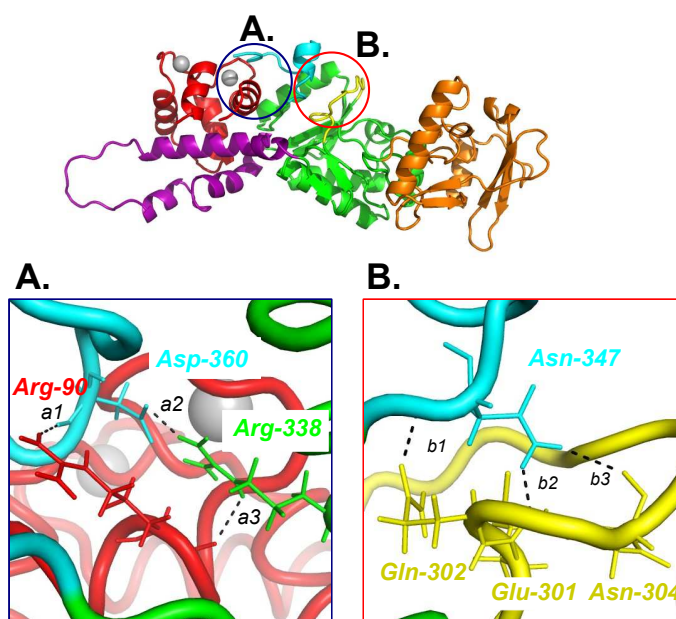
L'ensemble des interactions protéine-protéine étudiées au niveau moléculaire global [213; 214] constitue ce qu'on appelle l'interactome. Dans cet interactome, la plasticité des interactions biomoléculaires permettent à une seule protéine d'interagir avec plusieurs autres protéines. Ainsi, les interactions protéine-protéine ne sont donc pas déterminées par une interaction spécifique entre deux atomes, mais par un réseau d'interaction à l'interface protéine-protéine qui peut se réorganiser si les deux partenaires de l'interaction sont modifiés. Par conséquent, il est difficile de mettre en évidence des résidus particuliers dont les modifications induisent une perte significative de l'affinité entre deux partenaires.

Au sein d'une protéine, la détection de réseau d'interactions des résidus responsables du transfert du signal conformationnel, a suscité beaucoup d'intérêt au cours de cette dernière décennie [215; 216; 217; 218]. Des méthodes, basées sur des alignements de séquences protéiques apparentées [219; 220], ou sur l'analyse de la covariance des déplacements chimiques de RMN [221], ont permis de détecter des réseaux au sein de nombreuses protéines. Certaines analyses ont ouvert la voie à une ingénierie des protéines [222]. Cependant, les méthodes d'alignement de séquences de



protéines utilisées pour détecter l'évolution moléculaire, peuvent être améliorées [223].

Dans le chapitre IV, l'interaction entre le domaine catalytique AC de *Bordetella pertussis* et la calmoduline a été analysée par des simulations de dynamique moléculaire. L'étude des influences énergétiques au sein du complexe AC/C-CaM a révélé une influence unique de C-CaM sur CA (qui inclut la région C-terminale et la boucle catalytique), démontrant ainsi que la région CA est stabilisée par C-CaM, et qu'elle est donc la meilleure candidate pour des modifications ayant pour but de moduler l'interaction C-CaM/AC. L'analyse des liaisons hydrogène formées par des résidus de CA dans les complexes AC/C-CaM uniquement a mis en évidence un groupe de liaisons hydrogène formées par trois résidus, ARG338, ASN347 et ASP360. ARG338 et ASN360 sont situés dans les régions AC et C-terminale respectivement, et forment des liaisons hydrogène entre eux mais aussi avec le résidu ARG90 de C-CaM (figure V.1). ASN347, qui est situé dans la région C-terminale, interagit avec la boucle catalytique. Ces résidus sont donc susceptibles de jouer un rôle dans l'activation de l'adénylcyclase, et dans son interaction avec la CaM.



**FIG. V.1:** Vue d'ensemble de la structure du complexe AC-C-CaM (1YRT) représenté en ruban avec les zooms A et B à l'interface entre C-CaM, la queue C-terminale et la boucle catalytique. Le lobe C-CaM est coloré en rouge, la queue C-terminale en cyan, la boucle catalytique en jaune, le reste de CA en vert, et les ions  $\text{Ca}^{2+}$  sont représentés en sphères grises. Les résidus impliqués dans les liaisons hydrogène dans AC-(2Ca-C-CaM) et pour lesquels les liaisons hydrogène sont rompues dans AC-(0Ca-C-CaM) et AC sont représentés en bâtons et colorés en fonction de la région du complexe à laquelle ils appartiennent. Les liaisons hydrogène correspondantes sont notées  $a1$ ,  $a2$ ,  $a3$  pour l'interaction C-CaM/queue C-terminale/CA et  $b1$ ,  $b2$ ,  $b3$  pour l'interaction extrémité C-terminale/boucle catalytique.

Dans le présent chapitre, ces prédictions ont été vérifiées en modifiant ces trois résidus en alanine par mutagenèse dirigée. Trois variants de AC ont été ainsi construits et produits dans *E. coli*. Le premier, ACm3A, comporte ces trois mutations simultanément. Le second, ACm2A,

ne contient que les deux mutations ARG338ALA et ASP360ALA. Le dernier, ACm1A, ne porte que la mutation ASN347ALA. L'effet des différentes combinaisons de mutations sur l'activité catalytique des variants et sur leur affinité pour la calmoduline a ensuite été caractérisé par des tests enzymatiques. Des simulations de dynamique moléculaire ont été aussi réalisées sur les différents variants ACm1A, ACm2A et ACm3A complexés à C-CaM. L'analyse comparative des interactions qui ont lieu au niveau de l'interface C-CaM/AC entre les différents variants est en accord avec un modèle de réseau de résidus influant sur l'interaction C-CaM/AC.

## 2 Matériels et méthodes

### 2.1 Simulations de Dynamique Moléculaire

Les trois systèmes, AC-(2Ca-C-CaM), AC-(0Ca-C-CaM) et AC-isolé, simulés avec AC sauvage, ont déjà été décrits dans le chapitre IV. Trois autres systèmes comprenant les protéines modifiées ACm1A, ACm2A et ACm3A complexés à C-CaM, ont été construits de la manière suivante. La dernière conformation de la simulation AC-(2Ca-C-CaM) a été extraite et les chaînes latérales des résidus mutés ont été remplacées par les chaînes latérales du résidu alanine, en utilisant le module LEaP de AMBER 10 [152]. Les topologies des protéines modifiées ont été générées avec le champ de force FF99SB [188] et des contre-ions  $\text{Na}^+$  ont été ajoutés pour neutraliser le système. Les trois systèmes ont ensuite été placés dans une boîte d'eau de type TIP3P [189] avec un rayon de coupure de 10 Å dans des conditions périodiques. Les ions calcium ont été paramétrés avec les paramètres de Coulomb et Lennard-Jones suivants [190] : rayon de van der Waals  $R = 1.7131$  Å, profondeur du puits  $\epsilon = 0.459789$  kcal.mol<sup>-1</sup>. Les détails pour la construction des systèmes AC-(2Ca-C-CaM) and AC-(0Ca-C-CaM) ainsi que pour ACm1A-(2Ca-C-CaM), ACm2A-(2Ca-C-CaM), ACm3A-(2Ca-C-CaM) sont indiqués dans la Table I.

Les trajectoires de DM de 50 ns ont été produites avec le module PMEMD de AMBER 9.0 [191], sur des clusters de 4 et 8 processeurs. Un rayon de coupure de 10 Å a été défini pour les calculs des interactions de Lennard-Jones, tandis que les interactions électrostatiques à longue portée ont été calculées par la méthode "Particle Mesh Ewald" (PME) [159]. La température et la pression ont été régulées avec le thermostat de Berendsen [161], à une température de 298 K et une pression de 1 atm. L'algorithme SHAKE [154] a permis de garder les liaisons covalentes impliquant les hydrogènes rigides et ainsi d'utiliser un pas d'intégration de 2 fs. Les coordonnées atomiques ont été sauvegardées toutes les deux picosecondes (25 000 conformations).

Les trajectoires de DM ont été initialisées par une minimisation, répétée avec 10 cycles de "Steepest Descent", suivis par 490 cycles de Gradient Conjugué. Durant ces minimisations, les contraintes harmoniques sur les positions atomiques du soluté ont été progressivement relâchées avec une constante de force allant de 100 à 0 kcal/(mol.Å<sup>2</sup>). Ensuite, les systèmes ont été thermalisés de 0 à 298 K pendant 20 ps, en appliquant des contraintes harmoniques sur les positions atomiques du soluté avec une constante de force de 25 kcal/(mol.Å<sup>2</sup>). Pour terminer,

Trajectoire	AC-(2Ca-C-CaM)	AC-(0Ca-C-CaM)
Nombre de contre-ions Na <sup>+</sup>	14	18
Dimensions de la boîte d'eau	92.6 x 106.9 x 103.3 Å <sup>3</sup>	91.3 x 106.9 x 106.4 Å <sup>3</sup>
Nombre de molécules d'eau	26455	27056
Nombre total d'atomes	85825	87630
Temps simulé (ns)	30	30

Trajectoire	ACm1A-(2Ca-C-CaM)	ACm2A-(2Ca-C-CaM)	ACm3A-(2Ca-C-CaM)
Mutations	ASN347ALA	ARG338ALA ASP360ALA	ARG338ALA ASN347ALA ASP360ALA
Nombre de contre-ions Na <sup>+</sup>	14	14	14
Dimensions de la boîte d'eau	112.5 x 85.8 x 114.7 Å <sup>3</sup>	112.5 x 85.8 x 114.7 Å <sup>3</sup>	112.5 x 85.8 x 114.7 Å <sup>3</sup>
Nombre de molécules d'eau	29003	29008	29008
Nombre total d'atomes	93465	93468	93464
Temps simulé (ns)	50	50	50

TABLE I: Détails des systèmes moléculaires simulés par dynamique moléculaire (DM).

l'équilibration a été lancée avec un cycle de 5 ps à volume constant, suivi de 4 cycles de 2.5 ps à pression constante, un cycle de 10 ps et un dernier de 60 ps. Des contraintes harmoniques ont été appliquées sur les coordonnées atomiques des solutés durant les cinq premiers cycles, puis la constante de force a été relâchée de 25 à 0 kcal/(mol.Å<sup>2</sup>).

La formation des liaisons hydrogène entre les groupes donneurs et accepteur de la protéine a été suivie en sélectionnant les paires d'atomes proches de moins de 3 Å toutes les 10 conformations. La sélection a été implémentée avec la librairie python Biskit [224]. Dans la suite de ce chapitre, les résidus de C-CaM sont désignés par leur numéro dans la séquence de C-CaM présente dans la structure cristallographique 1YRT [131], suivi de la lettre 'a'. Par ailleurs, dans certaines analyses, l'extrémité C-terminale est considérée indépendamment de la région CA, contrairement au choix fait dans le chapitre "Rôle de la calmoduline et des ions Ca<sup>2+</sup> dans la stabilisation du domaine catalytique de l'adénylcyclase CyaA de *Bordetella pertussis*".

## 2.2 Construction des plasmides

Le plasmide pTRAC384GK utilisé pour l'expression de AC sauvage (correspondant aux 384 premiers codons de CyaA, suivi par deux résidus Gly et Lys) est décrit dans Vouquier et al (2004) [136]. Les plasmides pKTRACm2A et pKTRACm3A, utilisés pour exprimer ACm2A et ACm3A

respectivement ont été construits à partir du plasmide pKTRAC, un vecteur d'expression pour les protéines CyaC et CyaA de *Bordetella pertussis*. Ce plasmide a été obtenu à partir du plasmide pTRACG (Gmira et al, 2001 [225]) en remplaçant le gène de résistance à l'ampicilline (Amp) par le gène de résistance à la kanamycine (Kan) du plasmide pKT25 (Karimova et al, 2001 [226]). Dans pKTRAC, les gènes *cyaC* et *cyaA* sont exprimés sous le contrôle du promoteur Pr du phage  $\lambda$ . Le plasmide a une origine de répllication ColE1 et exprime également un répresseur thermosensible  $\lambda$ cI<sup>857</sup> qui réprime fortement la transcription au niveau du promoteur Pr du phage  $\lambda$  à une température inférieure à 32°C. L'expression des protéines CyaC et CyaA est déclenchée en augmentant la température à 37-42°C.

Le plasmide pKTRACm3A a été construit en sous-clonant entre les sites uniques *AgeI* et *BamHI* de pKTRAC, un gène synthétique (synthétisé par GeneArt, Life Technologies SAS, France) codant les résidus 320 à 384 de CyaA, suivi par les deux résidus Gly et Lys, dans lequel les trois codons, ARG338, ASN347 et ASP360 ont été changés en codons Ala. La séquence du gène ACm3A obtenu est donnée en annexe A.3.

Le plasmide pKTRACm2A a été construit en sous-clonant, entre les sites uniques *SnaBI* et *BsiWI* du plasmide pKTRACm3A, un oligonucléotide synthétique double-brins. Ce dernier résulte de l'hybridation des deux oligonucléotides 5'-GTTTTCTACGAAAACCGCGC-3' et 5'-GTACGCGCGGTTTTCGTAGAAAAC-3' (Eurofins MWG GmbH, Ebersberg, Germany), qui restore le codon sauvage ASN347 (il est à noter que *SnaBI* a été aboli avec l'insertion du nucléotide (Annexes A.4). Les gènes CyaA' modifiés dans pKTRACm2A et pKTRACm3A ont été vérifiés par un séquençage d'ADN (Eurofins MWG GmbH, Ebersberg, Germany).

Le plasmide pKT7AC contient un gène synthétique (synthétisé par GeneArt, Life Technologies SAS, France), codant pour les 384 premiers codons de CyaA, optimisés pour une forte expression dans *E.coli* et suivi par les deux résidus Gly et Lys. Ce gène synthétique AC est cloné sous le contrôle du promoteur T7 et d'une séquence synthétique RBS dans le vecteur pMK-RQ (GeneArt, Life Technologies SAS, France), un plasmide avec une origine de répllication ColE1 et un gène de résistance à la kanamycine (Kan).

Le plasmide pKT7ACm1A, utilisé pour l'expression de ACm1A, a été construit en sous-clonant entre les sites uniques *SnaBI* et *BsiWI* du plasmide pKT7AC, un oligonucléotide double-brins synthétique, obtenus par hybridation des deux oligonucléotides 5'-GTGTTCTACGAAGCTCGCGC-3' et 5'-GTACGCGCGAGCTTCGTAGAACAC-3' (Eurofins MWG GmbH, Ebersberg, Germany), modifiant ainsi le codon ASN347 en alanine (le site *SnaBI* est supprimé avec l'insertion de l'oligonucléotide, Annexes A.5). Le gène CyaA' modifié dans pKT7ACm1A a été vérifié par séquençage d'ADN (Eurofins MWG GmbH, Ebersberg, Germany).

### 2.3 Expression et purification des variants de AC

Les protéines AC sauvage, ACm2A et ACm3A ont été exprimées dans la souche *Escherichia coli* BLR (Novagen), transformées indépendamment avec les plasmides pTRAC384GK, pKTRACm2A ou pKTRACm3A respectivement, et mises en culture à 30°C, en milieu LB [227]

contenant 100  $\mu\text{g}/\text{ml}$  d'ampiciline (pour le plasmide pTRAC384GK) ou 50  $\mu\text{g}/\text{ml}$  de kanamycine (pour les plasmides pKTRACm2A et pKTRACm3A), jusqu'à une densité optique (600 nm) de 0.6–0.8. Puis, l'expression des protéines ACs a été déclenchée en augmentant la température à 42°C. Après 2h30 de croissance à 42°C, les cellules ont été collectées par centrifugation (20 min, 10000 g, 4°C), et les culots de cellules ont été congelés à -20°C.

ACm1A a été exprimée dans les souches KRX de *E. coli* (Promega, Madison, USA), transformées avec le plasmide pKT7ACm1A, et mises en culture à 37°C, en milieu LB contenant 50  $\mu\text{g}/\text{ml}$  de kanamycine, jusqu'à une densité optique (600 nm) de 0.6–0.8. Puis, l'expression de la protéine ACm1A a été déclenchée par l'addition de 0.1% de rhamnose (pour induire l'expression de l'ARN polymérase T7 codée dans le chromosome). Après 180 min de culture à 37°C, les cellules ont été collectées par centrifugation comme décrit précédemment.

Les culots de cellules ont été ensuite resuspendus dans un tampon Hepes-Na (20 mM, pH 7.5), et lysés par sonication à 4°C. La suspension soniquée a été centrifugée 20 min à 12000 g et 4°C. Le surnageant a été retiré, et le culot contenant les protéines AC agrégées en corps d'inclusion, a été resuspendu dans de l'urée 8 M, Hepes-Na, 20 mM pH 7.5, et mis en agitation toute une nuit à 4°C. Puis, après 20 min de centrifugation à 15000 g à 4°C, le surnageant ("extrait urée"), contenant les variants AC solubilisés, a été collecté.

Les protéines AC sauvage et AC modifiées ont été purifiées selon un protocole standard [136], en deux étapes successives de chromatographie sur colonnes de DEAE-Sephadex (20mL) échangeuses d'anion. L'extrait urée est d'abord chargé une première fois sur une colonne de DEAE-Sephadex équilibrée avec de l'urée 8M, Hepes-Na 20 mM, pH 7.5. Dans ces conditions, les protéines ne se fixent pas sur la résine, et se retrouvent essentiellement dans la fraction non retenue (FT). Cette dernière a été diluée cinq fois dans l'Hepes-Na, 20 mM pH 7.5, et passé une seconde fois sur une nouvelle colonne DEAE-Sephadex équilibrée dans de l'Hepes-Na, 20 mM pH 7.5. Les protéines sont maintenant retenues sur la résine et, après des lavages extensifs en tampon Hepes-Na, 20 mM pH 7.5, les protéines ont pu être éluées sous forme soluble avec l'Hepes-Na 20 mM, contenant du NaCl (100 à 200 mM). Les protéines ont ensuite été concentrées par ultrafiltration, en centrifugeant les protéines dans un tube contenant une membrane imperméable aux protéines > 10kDa (Millipore, Billerica, MA).

La pureté des protéines a été analysée par électrophorèse en gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) et est dans l'ensemble supérieure à 95 %. Le contenu protéique a été déterminé à partir du spectre d'absorption en utilisant un coefficient d'extinction à 278 nm de 28000  $\text{M}^{-1}.\text{cm}^{-1}$ . L'absence de contamination significative en acides nucléiques a été évaluée en mesurant le rapport entre la densité optique à 278 nm sur la densité optique à 260 nm. Pour l'ensemble des protéines modifiées, le rapport est supérieur à 1,5.

La calmoduline (CaM) a été purifiée comme décrit précédemment par Vouhier [136].

## 2.4 Détermination de l'activité enzymatique adénylyclase

L'activité adénylyclase a été déterminée avec un test colorimétrique récemment décrit (Laine et al. 2010 [184]) adapté au format microplaque. AC hydrolyse l'ATP en AMPc et en pyrophosphate inorganique (PPi) qui peut être ensuite clivé par la pyrophosphatase inorganique en deux molécules de phosphate (Pi). Ces dernières peuvent être détectées par colorimétrie lorsqu'elles sont complexées avec le phosphomolybdate vert de malachite qui absorbe à 595 nm. La quantité de Pi produite a été déterminée avec le réactif Pi ColorLock <sup>TM</sup> ALS (Innova Biosciences Cambridge, UK).

Les tests ont été réalisés dans des microplaques de 96 puits, à 30°C dans un volume final de 50  $\mu$ L contenant un mélange réactionnel de Tris-HCl, 50 mM, pH = 8.0, MgCl<sub>2</sub> 10 mM, CaCl<sub>2</sub> 0.1 mM, sérum albumine bovine 0.5 mg/ml, pyrophosphatase inorganique de *E.coli* à 2 Unités/mL (Sigma-Aldrich ; la quantité de pyrophosphatase ajoutée dans le milieu réactionnel a été déterminée de façon à ce qu'elle soit suffisante pour une conversion immédiate et complète du PPi en Pi), 0 à 1  $\mu$ M de CaM, et 0.1 à 1 nM (selon leur activité spécifique) de protéines AC modifiées (diluées à partir de solutions stock mesurées dans du Tris-HCl 10 mM, 0.2 % Tween 20, pH = 8.0). Les mélanges ont été pré-incubés pendant 10 min à 30 °C sous agitation.

Les réactions enzymatiques ont été initiées par l'addition d'ATP à une concentration finale de 2 mM, et les microplaques ont été incubées à 30 °C sous agitation. À différents temps d'incubation (de 2 à 10 min), 10  $\mu$ L de mélange réactionnel de chaque puit ont été transférés dans une seconde microplaque contenant dans chaque puit, 100  $\mu$ L de mélange Pi-ALS, contenant 20  $\mu$ L d' H<sub>2</sub>O et 80  $\mu$ L de réactif Pi ColorLock ALS (procuré par le kit Pi-ALS de Innova Biosciences). La réaction enzymatique est immédiatement arrêtée par les conditions acides du mélange de Pi-ALS. Après 10-13 min d'incubation à température ambiante, 10  $\mu$ L de solution "stabilisante" (procurée par le kit Pi-ALS) ont été ajoutés pour abolir l'hydrolyse non-enzymatique de l'ATP dans les conditions acides (selon les instructions du kit). Après 30-60 min d'incubation supplémentaire à température ambiante, les densités optiques à 595 nm (DO<sub>595</sub>) ont été enregistrées avec un lecteur de microplaques (Tecan, Lyon, France).

Une courbe standard a été calculée en ajoutant des concentrations connues de Pi dans le mélange Pi-ALS pour convertir des valeurs de DO<sub>595</sub> en moles de PPi produits. L'activité enzymatique a été calculée à partir de la vitesse initiale de la synthèse de PPi. Dans les conditions décrites ci-dessus, l'accumulation de PPi était linéaire par rapport au temps durant les six premières minutes au minimum.

## 3 Résultats

### 3.1 L'interface C-CaM/AC dans les structures cristallographiques

Dans cette section, nous avons voulu tout d'abord vérifier si les interactions entre AC et CaM déterminées dans le chapitre IV sont conservées dans les quatres structures cristallographiques

du complexe AC/C-CaM [131] publiées avec les entrées PDB 1YRT, 1YRU, 1ZOT et 2COL. Pour cela, nous avons analysé les liaisons hydrogène dont la distance entre les atomes donneurs et accepteurs est inférieure à 3 Å (tableau II).

Vingt et une liaisons hydrogène établies entre C-CaM et AC ont été détectées. Parmi elles, 10 liaisons relient C-CaM avec SA, 4 relient C-CaM avec CA, et 7 relient C-CaM avec la queue C-terminale. Les liaisons hydrogène s'établissant dans les quatre structures sont formées exclusivement entre C-CaM et le domaine SA, ce qui indique que les interactions à l'interface entre C-CaM et SA sont mieux conservées que les autres régions de l'interface C-CaM/AC.

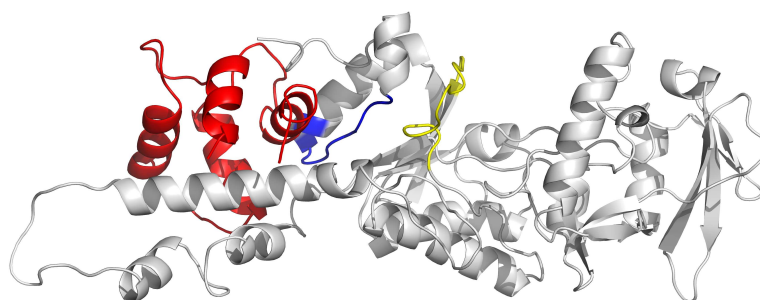
D'autre part, le seul pont d'eau observé entre C-CaM et AC dans toutes les structures cristallographiques relie les résidus GLY-113a-O et ARG-206-NH2 dans SA (tableau III).

Parmi les résidus impliqués dans les liaisons hydrogène et les ponts d'eau au niveau de l'interface C-CaM/AC, le résidu ARG338 est observé dans trois des quatre structures, dont deux fois en interaction avec le résidu ARG90a (C-CaM). De la même manière, le résidu ASP360 est observé dans deux structures en interaction avec ARG90a. Parmi les résidus de AC interagissant avec la C-CaM au travers des ponts d'eau, ARG338 est observée dans trois structures PDB sur quatre et est connecté à ARG90a. Les résidus modifiés sont donc souvent impliqués dans des liaisons hydrogène ou des ponts d'eau à l'interface C-CaM/AC. Mais des variations de partenaires sont observées.

Les résidus compris entre ARG338 et ASN347, forment un morceau d'hélice  $\alpha$  et une boucle étendue : nous appellerons dans la suite  $L\alpha 16Cter$ , puisqu'elle recouvre une partie de la dernière hélice  $\alpha 16$  et la partie C-terminale de AC (figure V.2). La majorité des résidus de  $L\alpha 16Cter$  forme des liaisons hydrogène dans trois structures cristallographiques sur quatre (Table IV), contrairement à ce qui est observé à l'interface C-CaM/AC (table II), ce qui montre que cette boucle est un point d'ancrage solide de la partie C-terminale de AC. De plus, les résidus de  $L\alpha 16Cter$  forment moins de ponts (table V) d'eau que ceux de l'interface C-CaM/AC (Table III).

Liaisons Hydrogènes C-CaM/AC	Pourcentage de formation	Entrées PDB	Région de AC
ARG-90a-O/ARG-338-NH1	25.0	1YRU	CA
ASP-93a-O/ARG-338-NH1	25.0	1YRT	CA
ASP-80a-OD2/ARG-246-NH2	25.0	1ZOT	SA
GLU-127a-OE1/ARG-235-NH2	25.0	1YRT	SA
GLU-82a-OE1/ALA-364-N	25.0	1YRT	Cter
GLU-87a-OE1/ARG-348-NH2	25.0	1YRT	Cter
ARG-86a-NE/LEU-362-O	25.0	1ZOT	Cter
ARG-86a-NH1/LEU-362-O	25.0	2COL	Cter
ASP-80a-OD2/ARG-250-NE	50.0	1YRT,1YRU	CA
GLU-114a-OE2/VAL-215-N	50.0	1ZOT,2COL	SA
GLU-120a-OE1/ARG-237-NH1	50.0	1YRT,1ZOT	SA
GLU-87a-OE2/ARG-258-NH2	50.0	1YRU,2COL	SA
GLU-87a-OE2/ARG-348-NE	50.0	1YRT,1ZOT	Cter
GLY-113a-O/ARG-206-NH2	50.0	1YRU,1ZOT	SA
LYS-94a-NZ/GLY-339-O	50.0	1ZOT,2COL	CA
ARG-90a-NE/ASP-360-OD1	75.0	1YRT,1ZOT,2COL	Cter
ARG-90a-NH2/ASP-360-OD2	75.0	1YRT,1ZOT,2COL	Cter
MET-124a-O/TRP-242-NE1	75.0	1YRT,1ZOT,2COL	SA
GLU-84a-OE1/ARG-259-NH2	100.0	1YRT,1YRU,1ZOT,2COL	SA
GLU-84a-OE2/ARG-259-NE	100.0	1YRT,1YRU,1ZOT,2COL	SA
GLU-87a-OE1/ARG-258-NH2	100.0	1YRT,1YRU,1ZOT,2COL	SA

**TAB. II:** Liaisons hydrogènes observées pour les entrées PDB 1YRT, 1YRU, 1ZOT et 2COL, correspondant aux structures cristallographiques aux rayons X du complexe AC/CaM. Les numéros des résidus de C-CaM sont suivis de la lettre 'a'. Les régions de AC contenant les résidus impliqués sont : CA, Cter et SA. CA comprend les résidus 7-61, 187-197, 261-299 et 313-345, l'extrémité C-terminale contient les résidus 346-364, SA comprend les résidus 198-260.



**FIG. V.2:** Structure du complexe AC/CaM. La protéine est représentée en mode "cartoon". La C-CaM est représentée en rouge, la boucle catalytique C-loop (résidus 300 à 312) en jaune, et la boucle L $\alpha$ 16Cter (résidus ARG338 à ASN347) en bleu.



Pont d'eau reliant AC et C-CaM	Pourcentage de formation	Entrée PDB	Région de AC
MET-145a-SD/ARG-246-O	25.0	1ZOT	SA
ASN-97a-N/ASP-359-O	25.0	1YRU	Cter
ASP-80a-OD2/ARG-250-NE	25.0	1YRU	CA
ASP-80a-OD2/THR-255-OG1	25.0	1YRT	SA
ARG-90a-NH1/GLU-346-OE2	25.0	1YRU	Cter
ARG-90a-NH2/GLU-346-OE2	25.0	1YRT	Cter
GLU-114a-OE1/ARG-237-NH2	25.0	2COL	SA
GLU-114a-OE2/ARG-206-NE	25.0	1YRU	SA
GLU-83a-OE1/ARG-348-NH1	25.0	2COL	Cter
GLU-84a-OE1/ARG-259-NH2	25.0	1YRT	SA
GLU-87a-OE1/THR-255-OG1	25.0	1ZOT	SA
GLU-87a-OE2/GLU-346-N	25.0	1YRT	Cter
ASP-95a-N/GLY-335-O	25.0	2COL	CA
GLU-84a-O/THR-255-OG1	50.0	1ZOT,2COL	SA
GLU-87a-O/TYR-345-N	50.0	1ZOT,2COL	CA
LEU-112a-O/ARG-206-NE	50.0	1YRT,2COL	SA
ARG-90a-O/ARG-338-NH1	75.0	1YRT,1ZOT,2COL	CA
GLY-113a-O/ARG-206-NH2	100.0	1YRT,1YRU,1ZOT,2COL	SA

**TAB. III:** Ponts d'eau observés pour les entrées PDB 1YRT, 1YRU, 1ZOT et 2COL, correspondant aux structures cristallographiques aux rayons X du complexe AC/C-CaM. Les numéros des résidus de C-CaM sont suivis de la lettre 'a'. Les régions de AC contenant les résidus impliqués sont : CA, Cter et SA. Les régions de AC contenant les résidus impliqués sont : CA, Cter et SA. CA comprend les résidus 7-61, 187-197, 261-299 et 313-345, l'extrémité C-terminale contient les résidus 346-364, SA comprend les résidus 198-260.

Liaisons Hydrogènes C-CaM/AC	Pourcentage de formation	Entrées PDB	Région AC
GLY-341-O/LEU-198-N	50.0	1YRT,1YRU	Cter/SA
TYR-342-OH/ASN-35-ND2	100.0	1YRT,1YRU,1ZOT,2COL	Cter/CA
VAL-343-N/PRO-196-O	100.0	1YRT,1YRU,1ZOT,2COL	Cter/SA
TYR-345-OH/GLU-276-OE2	100.0	1YRT,1YRU,1ZOT,2COL	Cter/CA
TYR-345-OH/ASN-303-ND2	100.0	1YRT,1YRU,1ZOT,2COL	Cter/C-loop
GLU-346-OE1/LEU-357-N	75.0	1YRT,1YRU,1ZOT	Cter/Cter
ASN-347-ND2/ASN-304-O	75.0	1YRT,1ZOT,2COL	Cter/C-loop
ASN-347-ND2/GLU-301-O	100.0	1YRT,1YRU,1ZOT,2COL	Cter/C-loop

**TAB. IV:** Liaisons hydrogènes observées pour les résidus de la boucle L $\alpha$ 16Cter dans les entrées PDB 1YRT, 1YRU, 1ZOT et 2COL.

### 3.2 Analyse de l'interaction C-CaM/AC au cours des trajectoires de DM

Dans cette section, trois simulations de dynamique moléculaire indépendantes ont été réalisées sur le complexe formé entre la C-CaM et AC modifié de trois façons différentes. Dans le premier système, ACm3A-(2Ca-C-CaM), trois résidus de AC, ARG338, ASN347 et ASP360 (figure V.1 A et B) ont été modifiés en alanine. Le second, ACm2A-(2Ca-C-CaM), ne contient que les deux

Résidus liés par un pont d'eau	Pourcentage de formation	Code PDB	Région de AC
GLU-340-N/GLN-336-O	25.0	1YRT	Cter/CA
TYR-342-OH/ASN-35-OD1	50.0	1YRT,1YRU	Cter/CA
TYR-342-N/ARG-338-O	75.0	1YRT,1ZOT,2COL	Cter/CA
TYR-345-N/ILE-194-O	75.0	1YRT,1ZOT,2COL	Cter/SA

**TAB. V:** Ponts d'eau observés pour les résidus de la boucle  $L\alpha 16C_{ter}$  dans les entrées PDB 1YRT, 1YRU, 1ZOT et 2COL.

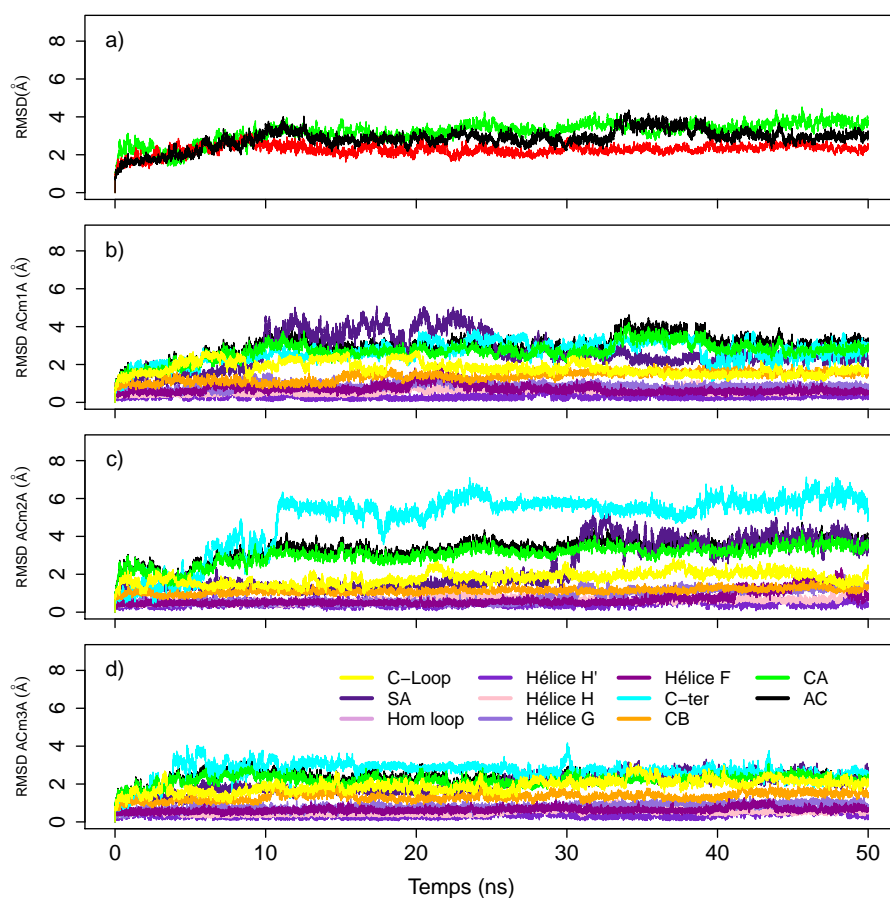
modifications ARG338ALA et ASP360ALA (figure V.1 A). Le dernier, ACm1A-(2Ca-C-CaM), ne porte que la mutation ASN347ALA figure(V.1B). La simulation de ces variants a pour objectif de vérifier leur stabilité, et d'analyser les variations d'interaction au niveau de l'interface C-CaM/AC.

La dérive conformationnelle du complexe AC/C-CaM a été calculée pour les trajectoires ACm1A-(2Ca-C-CaM), ACm2A-(2Ca-C-CaM) et ACm3A-(2Ca-C-CaM) (Figure IV.2a). Elle est globalement similaire pour les trois protéines modifiées, et atteint rapidement un plateau de 3 Å. De plus, le plateau du RMSD est similaire à celui obtenu pour les trajectoires de DM sur le complexe AC/C-CaM dont AC sauvage (Chapitre IV, simulations AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM)) et qui restaient stables au cours de la simulation. Ces résultats nous permettent de dire que les mutations ne déstabilisent pas la structure du complexe déjà formé, au moins aux échelles de temps considérées dans les simulations.

La dérive conformationnelle a également été calculée sur chaque région de AC des trois trajectoires des protéines modifiées (Figure V.3b-d). Pour ACm1A-(2Ca-C-CaM) (Figure V.3b) et ACm3A-(2Ca-C-CaM) (Figure V.3d), la dérive conformationnelle de chaque région est inférieure à 4 Å, ce qui indique une stabilité de ces régions dans le complexe. Les plus grandes valeurs de RMSD sont observées pour toutes les régions sauf CB. Cette observation est en accord avec le fait que les mutations introduites sont localisées dans le voisinage des régions CA, SA, C-loop et C-ter. Le grand RMSD observé pour SA est analogue à ce qui avait été observé dans la dynamique AC-(2Ca-C-CaM) (Chapitre IV, Figure IV.2).

Pour ACm2A-(2Ca-C-CaM) (Figure V.3c), le RMSD de la région CA, qui contient l'extrémité C-terminale et la boucle catalytique, est légèrement plus élevé que dans ACm1A et ACm3A avec un plateau à 4 Å. De plus, le RMSD obtenu pour l'extrémité C-terminale est beaucoup plus élevé avec un plateau à 6 Å. Par contre, dans les trois trajectoires des protéines modifiées, le RMSD de la boucle catalytique reste stable autour de 2 Å. Une observation plus précise de la figure V.3c montre que les dérives conformationnelles de CA et de l'extrémité C-terminale sont parallèles, et que l'extrémité C-terminale est donc majoritairement responsable de la perturbation de CA. L'une des deux mutations introduites dans ACm2A est située au niveau de l'interface C-CaM/C-ter. L'extrémité C-terminale subit donc une réorganisation conformationnelle due à l'introduction de cette mutation, mais cette réorganisation reste limitée à cette région et ne se propage pas au reste de la protéine.

L'analyse en composantes principales (ACP) de la matrice de covariance des fluctuations



**FIG. V.3:** Dérives conformationnelles estimées par le calcul du RMSD des coordonnées atomiques des carbones  $\alpha$ . La conformation de référence est la première de chaque trajectoire. (a) Dérives calculées sur le domaine AC des trajectoires ACm1A-(2Ca-C-CaM) (courbe noire), ACm2A-(2Ca-C-CaM) (courbe verte) et ACm3A-(2Ca-C-CaM) (courbe rouge). (b-d) Dérives calculées sur les différentes régions du domaine AC pour les trajectoires ACm1A-(2Ca-C-CaM) (b) ACm2A-(2Ca-C-CaM) (c) et ACm3A-(2Ca-C-CaM) (d). Les courbes sont colorées en fonction de la région selon la légende de la figure (d). Les régions de AC sont définies de cette façon : CA comprend les résidus 7-61, 187-197, 261-299 and 313-345 for CA, la boucle catalytique comprend les résidus 300-312, la queue C-terminale contient les résidus 346-364, CB est composée des résidus 62-186 et SA des résidus 198-260. Les hélices H, F, G, H' correspondent aux résidus 234-253, 198-210, 214-223, 256-260.

atomiques (Figure V.4) montre des caractéristiques similaires à celles observées sur les trajectoires du complexe où AC est non muté. En effet, les extrémités SA et CB sont celles qui bougent le plus avec une tendance à se rapprocher l'une de l'autre, tandis que CA et C-CaM sont les moins mobiles. Les trois mutations produisent des projections similaires du premier vecteur propre, la seule différence étant un mouvement de la queue C-terminale plus large dans ACm2A, ce qui est en accord avec la dérive conformationnelle de cette région décrite dans le paragraphe précédent.

Dans le but d'obtenir des informations plus précises sur l'interaction entre C-CaM et AC modifié, les liaisons hydrogène établies entre ces deux protéines ont été étudiées. Une analyse globale a d'abord été faite sur le nombre de liaisons hydrogène établies entre C-CaM et les différentes régions de AC modifié dans plus de 50% de la trajectoire (Tableau VI). Parmi les trois trajectoires des variants de AC, ACm3A-(2Ca-C-CaM) est celui qui décrit le moins de liaisons hydrogène

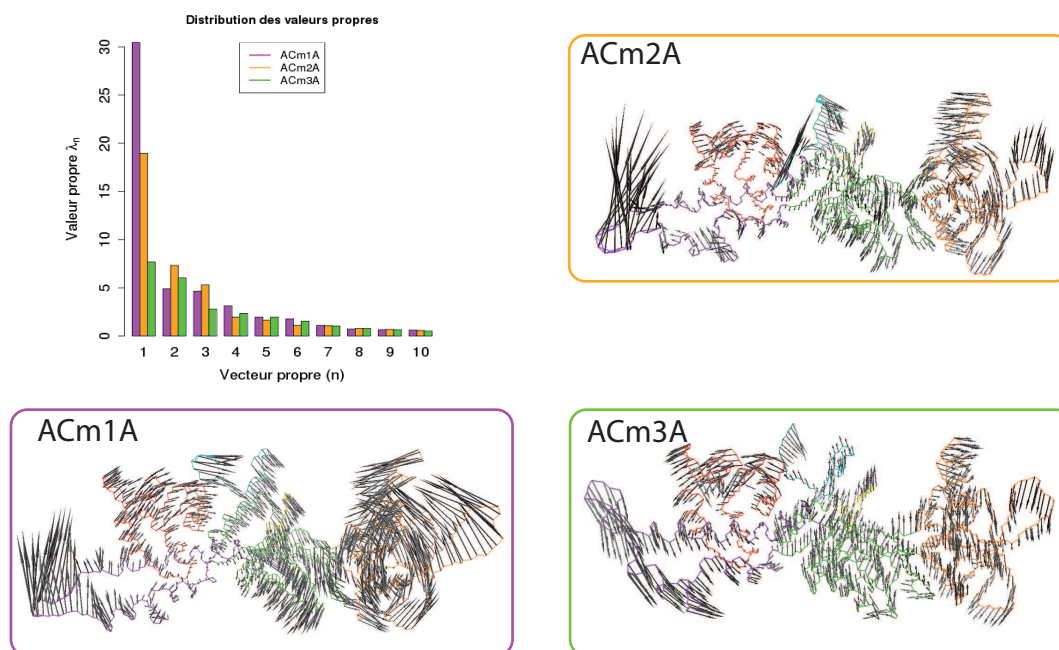


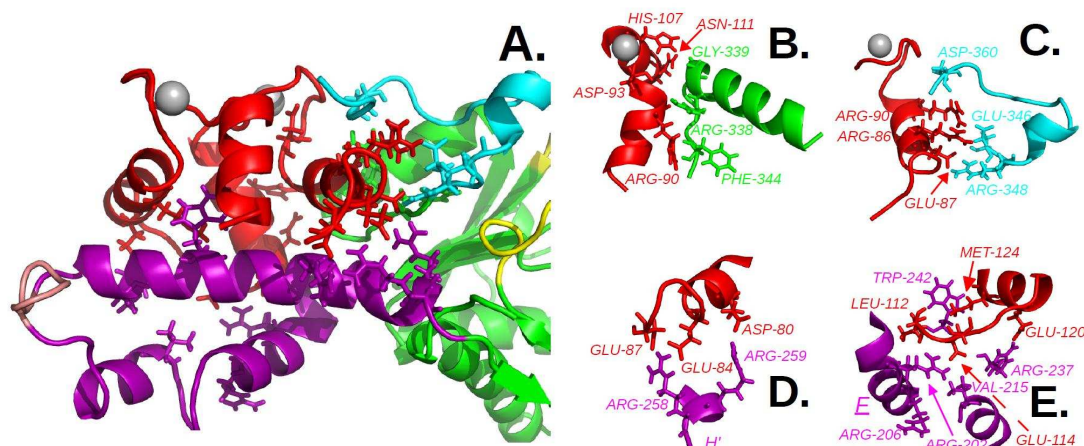
FIG. V.4: *Analyse en Composantes Principales de la matrice de covariance de la dynamique interne des mutants calculée sur les trajectoires ACm1A-(2Ca-C-CaM), ACm2A-(2Ca-C-CaM) et ACm3A-(2Ca-C-CaM)*. (a) Distribution des valeurs propres colorées en violet pour ACm1A-(2Ca-C-CaM), en orange pour ACm2A-(2Ca-C-CaM) et en vert pour ACm3A-(2Ca-C-CaM). Projection sur le premier mode du système ACm1A-(2Ca-C-CaM) (b), du système ACm2A-(2Ca-C-CaM) (c), du système ACm3A-(2Ca-C-CaM) (d). Les régions du complexe sont colorées en rouge (C-CaM), violet (SA), vert (CA), orange (CB), cyan (C-ter) et jaune (C-loop).

entre C-CaM et l'extrémité C-terminale, ACm1A-(2Ca-C-CaM) étant celui qui en montre le plus. Cependant, dans toutes les trajectoires sauvages (AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM)) produites en présence ou en absence d'ions  $\text{Ca}^{2+}$ , plus de huit liaisons hydrogène sont observées entre C-CaM et l'extrémité C-terminale, ce qui représente au moins le double de ce qui a été observé pour ACm3A-(2Ca-C-CaM).

Trajectoire	C-CaM/Cter	C-CaM/CA	C-CaM/SA
ACm1A-(2Ca-C-CaM)	14	4	23
ACm2A-(2Ca-C-CaM)	8	3	23
ACm3A-(2Ca-C-CaM)	4	2	23
ACWT-(2Ca-C-CaM)	8	4	19
ACWT-(0Ca-C-CaM)	9	2	22
ACWTbis-(2Ca-C-CaM)	8	3	15
ACWTbis-(0Ca-C-CaM)	10	5	16

TAB. VI: *Nombre de liaisons hydrogène présentes entre C-CaM et AC à plus de 50 % du temps des trajectoires*. Les liaisons hydrogène sont classées selon la région où les résidus de AC sont situés : C-CaM/Cter représente les liaisons hydrogène entre C-CaM et l'extrémité C-terminale, C-CaM/CA représente les liaisons hydrogène entre C-CaM et la région CA exclusivement (sans l'extrémité C-terminale), C-CaM/SA représente les liaisons hydrogène entre C-CaM et SA. CA comprend les résidus 7-61, 187-197, 261-299 and 313-345, l'extrémité C-terminale contient les résidus 346-364, SA comprend les résidus 198-260.

Les liaisons hydrogène entre C-CaM et AC ont ensuite été analysées plus en détail en sélectionnant les liaisons hydrogène établies à plus de 80% du temps pour au moins une des simulations (Annexe A.1). En regardant plus en détail les liaisons hydrogène formées entre C-CaM et AC, quelques groupes de résidus ont été extraits (Figure V.5).



**FIG. V.5:** (A). *Vue générale de l'interface entre C-CaM et AC.* Les résidus impliqués dans une liaison hydrogène formée dans plus de 80% d'une trajectoire donnée, sont représentés en bâtons. (B-E). Zooms sur différentes parties de l'interface, avec les mêmes résidus représentés en bâtons. (B) Interface C-CaM/CA. (C) Interface C-CaM/Cter. (D) Interface C-CaM/SA1. (E) Interface C-CaM/SA2. Pour (D) et (E), les noms des hélices sont soulignés. Les résidus ARG-237 et TRP-242 appartiennent à l'hélice H. SA1 comprend l'hélice H' de SA, tandis que SA2 contient des fragments des hélices H, F et G.

Le premier groupe (Figure V.5B) inclut les résidus ARG-90a, ASP-93a, HIS-107a, ASN-111a des deux côtés de la première main EF de C-CaM, et les résidus ARG-338, GLU-339 et PHE-344 de CA. Le résidu ARG-338 a été muté en alanine dans ACm2A et ACm3A. Ce groupe montre également une variation dans les liaisons hydrogène établies parmi les trajectoires de AC sauvage (WT) et des variants de AC, ainsi qu'entre AC sauvage et les variants (Annexe A.2).

Le second groupe (Figure V.5C), inclus les résidus ARG-86a, GLU-87a et ARG-90a de la première hélice  $\alpha$  de C-CaM qui interagissent avec les résidus ASP-360, GLU-346 et ARG-348 de l'extrémité C-terminale. Le résidu ARG-360 a été modifié en ALA dans ACm2A et ACm3A. Le nombre de liaisons hydrogène varie de manière assez importante entre les différentes trajectoires. Le nombre de liaisons hydrogène présentes dans plus de 80% de la trajectoire est compris entre 5 et 7 sur l'ensemble des trajectoires, à l'exception de ACm3A. Dans ACm3A, seule la liaison hydrogène ARG-90a-2HH2/GLU-346-OE1 est présente à plus de 80% du temps.

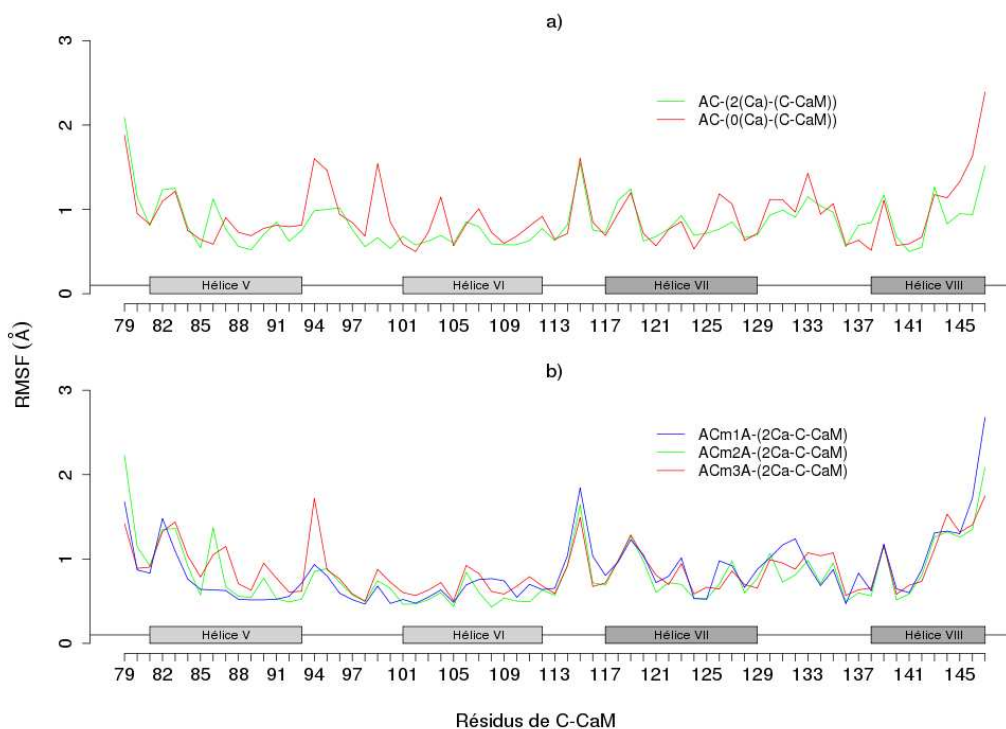
Le troisième groupe (Figure V.5D) contient les résidus ASP-80a, GLU-84a, GLU-87a et l'extrémité N-terminale de C-CaM, et les résidus ARG-258, ARG-259 de l'hélice  $\alpha$  H' de la région SA. Les liaisons hydrogène entre GLU-84a et ARG-259 sont présentes dans toutes les trajectoires, et ancrent l'extrémité N-terminale de C-CaM à l'hélice  $\alpha$  H'.

Le quatrième groupe (Figure V.5E) relie les résidus LEU-112a, GLU-114a, GLU-120a et MET-

124a situés sur la boucle qui connecte les deux mains EF de C-CaM, aux résidus ARG-202, ARG-206 (hélice  $\alpha$  F de SA), VAL-215 (hélice  $\alpha$  G de SA), et ARG-237, TRP-242 (hélice  $\alpha$  H de SA). La liaison hydrogène MET-124a-O/TRP-242-HE1 est présente dans toutes les trajectoires avec des pourcentages de formation toujours supérieurs à 99%. Cette stabilité de la liaison hydrogène est en accord avec les observations faites par Vouquier [136] que la mutation de Met-124a dans CaM détériore fortement son interaction avec AC.

### 3.3 Les fluctuations de C-CaM dans les trajectoires de DM

Le comportement de C-CaM a été suivi au cours des trajectoires ACm1A-(2Ca-C-CaM), ACm2A-(2Ca-C-CaM) et ACm3A-(2Ca-C-CaM) et comparé aux observations faites dans le chapitre IV sur le domaine AC sauvage. Les fluctuations ont été calculées (Figure V.6) au cours de chacune de ces trajectoires.



**FIG. V.6:** *Fluctuations par résidu de C-CaM pour les trajectoires des complexes AC/CaM sauvages et des complexes formés entre C-CaM et les variants de AC.* Le nom des trajectoires est donné dans la légende. (a) Les trajectoires des complexes sauvages, (b) Les trajectoires comprenant ACs modifiés complexés avec C-CaM.

ACm3A-(2Ca-C-CaM) (Figure V.6b : courbe rouge) a un comportement assez particulier par rapport aux autres trajectoires. En effet, des fluctuations plus importantes que pour les autres mutations sont observées au niveau de la partie C-terminale de l'hélice V et de la boucle de la première main EF. Cette région de C-CaM a été précédemment décrite comme interagissant avec la région CA (Figure V.5B) et la partie C-terminale (Figure V.5C). Elle est en contact avec la

boucle L $\alpha$ 16Cter définie sur la Figure V.2. Les fluctuations plus importantes de cette région de C-CaM dans cette trajectoire concordent avec la rupture de la plupart des liaisons hydrogène présentes entre C-CaM et CA (Figure V.5B).

La coordination des ions calcium par les résidus de C-CaM est stable au cours des trajectoires des protéines modifiées (Table VII), à l'exception de quelques interactions dans la seconde main EF. En effet, la liaison hydrogène entre le calcium et Asp-133O $\delta$ 1/O $\delta$ 2 est perturbée dans ACm2A-(2Ca-C-CaM) et ACm3A-(2Ca-C-CaM). La liaison hydrogène entre le calcium et ASP-133-O est perturbé dans ACm1A-(2Ca-C-CaM) et ACm2A-(2Ca-C-CaM), et plus faiblement dans ACm3A-(2Ca-C-CaM). Ces perturbations de Asp-133 O $\delta$ 1/O $\delta$ 2 et ASP-133-O ont également été observées dans la trajectoire de AC-(2Ca-C-CaM) (Chapitre IV, table IV). Les modifications introduites ont donc tendance à augmenter les instabilités déjà observées pour le complexe sauvage.

Main EF	Atome	ACm1A-(2Ca-C-CaM)	ACm2A-(2Ca-C-CaM)	ACm3A-(2Ca-C-CaM)
3	Asp-93 O $\delta$ 1/O $\delta$ 2	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1
3	Asp-95 O $\delta$ 1/O $\delta$ 2	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1
3	Asn-97 O $\delta$ 1	2.7 $\pm$ 0.1	2.7 $\pm$ 0.1	2.7 $\pm$ 0.1
3	Tyr-99 O	2.7 $\pm$ 0.1	2.7 $\pm$ 0.1	2.7 $\pm$ 0.1
3	Glu-104 O $\epsilon$ 1/O $\epsilon$ 2	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1
4	Asp-129 O $\delta$ 1/O $\delta$ 2	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1
4	Asp-131 O $\delta$ 1/O $\delta$ 2	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1	2.6 $\pm$ 0.1
4	Asp-133 O $\delta$ 1/O $\delta$ 2	2.6 $\pm$ 0.1	4.8 $\pm$ 1.5	3.8 $\pm$ 1.5
4	Asp-133 O	6.0 $\pm$ 0.7	2.9 $\pm$ 0.6	2.7 $\pm$ 0.2
4	Gln-135 O	2.7 $\pm$ 0.1	2.7 $\pm$ 0.1	2.7 $\pm$ 0.1

**TAB. VII: Coordination des ions Ca<sup>2+</sup> dans les trajectoires des variants de AC complexés avec C-CaM.** Les atomes des boucles liant le calcium proches d'au plus 2.8 Å de ce dernier sont détectés. Pour les distances impliquant les couples d'atomes O $\delta$ 1/O $\delta$ 2 et O $\epsilon$ 1/O $\epsilon$ 2, la plus petite des deux distances est suivie. Pour chaque interaction de coordination, la distance moyenne et l'écart-type sont indiqués.

L'ouverture des mains EF (Table VIII) est légèrement plus petite pour la main EF 3 que pour la main EF 4, tout comme dans les trajectoires des complexes avec AC sauvage (Chapitre IV, table III). Les mutations induisent de légères variations de ces angles. De la même façon, les valeurs des distances entre les mains EF et l'hélice  $\alpha$  H ainsi que les surfaces de patches hydrophobes de C-CaM (Table VIII) sont du même ordre que pour les trajectoires sauvages, et ne changent pas beaucoup lorsque les mutations sont introduites. Les régions hydrophobes (Table VIII) décrivent des valeurs qui sont du même ordre que celles observées pour les trajectoires sauvages en présence et en absence d'ions calcium (chapitre IV, table III), et elles ne varient pas beaucoup non plus avec l'introduction des mutations. Les modifications introduites dans AC vont donc avoir un effet limité sur la structure de C-CaM, ce qui est en accord avec le fait que, quelque soit la mutation, l'interaction entre AC et C-CaM est toujours maintenue.

Trajectoire	ACm1A-(2Ca-C-CaM)	ACm2A-(2Ca-C-CaM)	ACm3A-(2Ca-C-CaM)
Surface catalytique ( $\text{\AA}^2$ )	$544.8 \pm 51.7$	$641.5 \pm 45.0$	$517.2 \pm 68.8$
Angle de la main EF 3 ( $^\circ$ )	$89.9 \pm 2.1$	$86.1 \pm 2.3$	$90.1 \pm 3.2$
Angle de la main EF 4 ( $^\circ$ )	$95.6 \pm 5.1$	$99.4 \pm 3.8$	$96.1 \pm 4.9$
Surface Accessible du patch hydrophobe ( $\text{\AA}^2$ )	$774.7 \pm 33.2$	$779.1 \pm 38.5$	$767.9 \pm 29.2$
Distance entre la main EF 3 et l'hélice $\alpha$ H ( $\text{\AA}$ )	$10.8 \pm 0.3$	$11.8 \pm 0.2$	$10.8 \pm 0.5$
Distance entre la main EF 4 et l'hélice $\alpha$ H ( $\text{\AA}$ )	$11.7 \pm 0.3$	$11.6 \pm 0.3$	$11.4 \pm 0.3$

**TAB. VIII:** *Analyse du site catalytique adénylcyclase et de la structure de C-CaM au cours des trajectoires ACm1A-(2Ca-C-CaM), ACm2A-(2Ca-C-CaM) et ACm3A-(2Ca-C-CaM).* La surface catalytique est la surface accessible au solvant ( $\text{\AA}^2$ ) calculée avec 2.5.3 [204] sur les résidus catalytiques K58, H63, K84, D188, D190, H298, N304 et F306 définis dans la figure 5c de [131]. La structure de C-CaM est caractérisée par : (i) la surface accessible au solvant ( $\text{\AA}^2$ ) des résidus I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, A147, définissant le patch hydrophobe [108], (ii) les angles moyens entre les axes des hélices  $\alpha$  V et VI (résidus 82-92 et 102-111) et des hélices  $\alpha$  VII et VIII (résidus 118-128 et 138-146), définissant l'ouverture des mains EF 3 et 4, (iii) les distances minimales ( $\text{\AA}$ ) entre les centres de masse des mains EF et l'axe de l'hélice H.

### 3.4 Réseau d'interactions au sein de AC au cours des trajectoires de DM

Des résidus situés sur la boucle L $\alpha$ 16Cter définie sur la figure V.2, sont également connectés par un jeu de liaisons hydrogène. Les liaisons hydrogène entre les résidus de la boucle et d'autres résidus de AC ont été suivies au cours des trajectoires de DM calculées sur les domaines AC modifiés et sauvage (Annexes A.2 et figure V.7). Ces liaisons hydrogène ont été sélectionnées parmi celles ayant un niveau de formation supérieur à 80% dans au moins une des trajectoires. Les résidus de L $\alpha$ 16Cter impliqués dans une liaison hydrogène sont les suivants, 338, 340, 341, 342, 343, 345, 346 et 347, les résidus 338 et 347 faisant partie des résidus qui ont été modifiés. Ceci montre que la boucle conserve une certaine rigidité interne quelques soient les mutations introduites.

Un jeu de liaisons hydrogène implique les résidus ASN-347 situé dans la queue C-terminale, et les résidus GLU-301, GLN-302, ASN-304 situés dans la boucle catalytique. Ces liaisons hydrogène sont majoritairement rompues lorsque ASN-347 est modifié, à l'exception de la liaison GLN-302-OE1/ALA-347-H qui est formée à 99.8 % du temps dans ACm3A-(2Ca-C-CaM), et absente dans les autres trajectoires. Ainsi, même dans le cas où ASN-347 est muté, la boucle catalytique a tendance à être maintenue à travers l'établissement d'une liaison hydrogène avec le résidu modifié ALA-347. Dans les trajectoires de AC sauvage, les liaisons hydrogène entre ASN-347 et la boucle catalytique sont majoritairement formées quelque soit la présence d'ions calcium ou de C-CaM.

Quelques liaisons hydrogène à plus longue portée connectent les résidus GLU-340, TYR-342, VAL-343 et TYR-345 avec les résidus HIE-197, ASN-35, PRO-196 et GLU-276. Parmi elles,



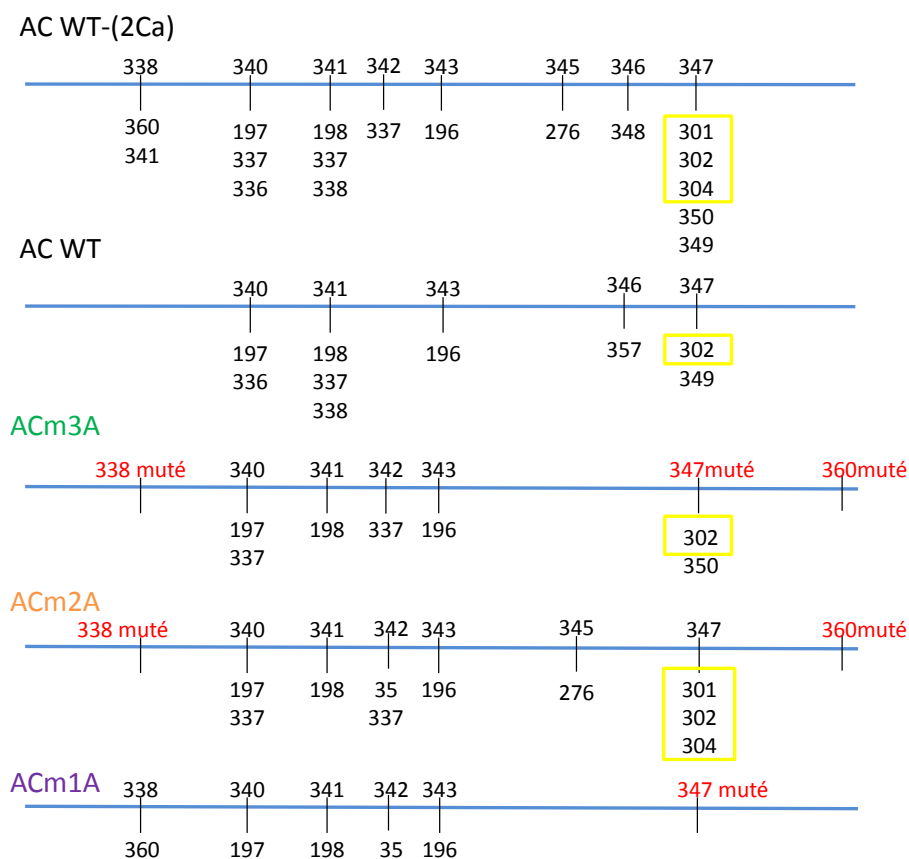


FIG. V.7: Schéma des liaisons hydrogène établies pour plus de 80% du temps entre les résidus de la boucle  $L\alpha 16Cter$  et des résidus de AC. Les résidus de la boucle  $L\alpha 16Cter$  sont indiqués au-dessus du segment bleu. Les résidus de AC impliqués dans des liaisons hydrogène avec les résidus de  $L\alpha 16Cter$  sont indiqués au-dessous du segment. Le résidu 360 est aussi indiqué pour les trajectoires où il a été modifié. Les résidus encadrés en jaune correspondent aux résidus de la boucle catalytique.

les liaisons GLU-340-O/HIE-197-HE2 et VAL-343-H/PRO-196-O sont majoritairement formées dans toutes les trajectoires. TYR-342-HH/ASN-35-OD1 est formé dans ACm1A-(2Ca-C-CaM) et ACm2A-(2Ca-C-CaM), et rompue dans les autres trajectoires. Les liaisons TYR-345-HH/GLU-276-OE2 et TYR-345-HH/GLU-276-OE1 sont rompues dans ACm1A-(2Ca-C-CaM) et majoritairement formées dans les autres trajectoires.

En présence d'ions calcium dans la trajectoires AC-(2Ca-C-CaM), un jeu de liaisons hydrogène est formé entre les résidus ASP-360 et ARG-338. Les pourcentages de formation diminuent lorsque les ions  $Ca^{2+}$  ou la C-CaM sont retirés, mais ne sont pas très affectés par la mutation de ASN-347 seul (ACm1A). Ces résidus ne forment pas de liaison hydrogène lorsqu'ils sont mutés dans ACm2A et ACm3A.

La boucle  $L\alpha 16Cter$  est connectée par des liaisons hydrogène, d'un côté à la boucle catalytique, et de l'autre côté à C-CaM. Deux résidus modifiés sont situés aux deux extrémités de

L $\alpha$ 16Cter. L'un d'eux, le résidu ASN-347 est situé à l'extrémité proche de la boucle catalytique. L'autre, le résidu ARG-338 est situé à l'extrémité proche de C-CaM. Ces deux résidus sont couplés mécaniquement par la chaîne polypeptidique sur laquelle ils sont séparés par 11 résidus seulement. Les mutations dans ACm2A et ACm1A réduisent le réseau d'interaction à C-CaM et à la boucle catalytique respectivement. Cependant, la triple mutation est requise pour obtenir une déstabilisation des liaisons hydrogène. Ceci suggère que le variant ACm3A aurait une affinité plus réduite pour la calmoduline. Néanmoins, les liaisons hydrogène connectant les résidus au sein de la boucle C-terminale, restent stables, ce qui indique que la boucle bouge comme un corps rigide.

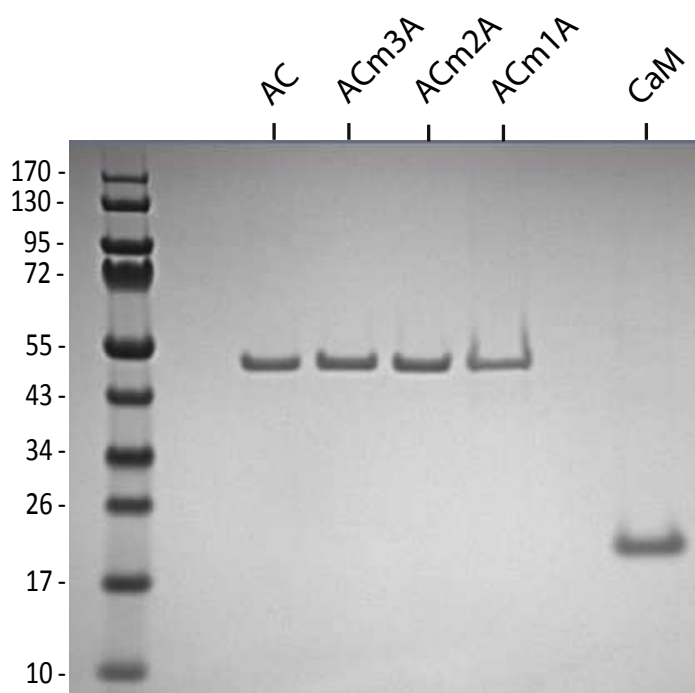
La surface catalytique a été mesurée en calculant la surface accessible au solvant avec le logiciel MSMS 2.5.3 [204] sur les résidus catalytiques K58, H63, K84, D188, D190, H298, N304 et F306, tels qu'ils sont définis dans la figure 5c de Guo [131]. Les résultats sont reportés dans le tableau VIII. Ainsi, on a obtenu des valeurs de 544.8, 641.5 et 517.2 Å<sup>2</sup> pour les trajectoires ACm1A-(2Ca-C-CaM), ACm2A-(2Ca-C-CaM) et ACm3A-(2Ca-C-CaM). Ce dernier a la plus petite valeur de surface accessible au solvant. Dans les trajectoires AC-(2Ca-C-CaM), AC-(0Ca-C-CaM) et AC-isolé (chapitre IV), on avait obtenu des valeurs de 527, 614 et 460 Å<sup>2</sup> respectivement (Chapitre IV table III). Les valeurs de 544.8 et 641.5 Å<sup>2</sup>, obtenues pour ACm1A-(2Ca-C-CaM), ACm2A-(2Ca-C-CaM) sont supérieures à la surface de 527 Å<sup>2</sup> obtenue pour un complexe AC/C-CaM fonctionnel, tandis que la valeur de 517.2 Å<sup>2</sup>, obtenue pour ACm3A-(2Ca-C-CaM), est légèrement inférieure. Or, dans le chapitre IV, on a montré que l'augmentation de la surface accessible au solvant entre AC-(2Ca-C-CaM) et AC-(0Ca-C-CaM) ne semblait pas avoir d'influence sur l'activité, tandis qu'une diminution de cette surface entre AC-(2Ca-C-CaM) et AC-isolé semblait concorder avec une perte d'activité. Donc, de ce point de vue, on peut s'attendre à ce que le variant ACm3A, dont la surface accessible au solvant est légèrement inférieure à celle de AC-(2Ca-C-CaM), ait une activité plus faible, tandis que les activités enzymatiques de ACm1A et ACm2A seraient analogues à l'activité de la toxine sauvage.

### 3.5 Caractérisation expérimentale des variants de AC modifiés

Pour caractériser l'effet des mutations précédemment décrites sur l'activité adényl cyclase et sur l'affinité pour la CaM, les trois variants de AC, ACm3A (mutations ARG338ALA, ASN347ALA, ASP360ALA), ACm2A (mutations ARG338ALA et ASP360ALA) et ACm1A (mutations ASN347ALA) ont été construits, produits et comparés à AC sauvage, qui comprend les résidus 1-384 de CyaA.

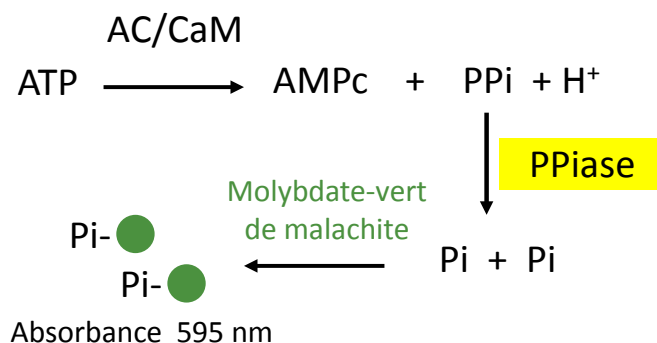
Ces protéines ont été exprimées chez *E. coli* et purifiées à homogénéité (cf Matériels et méthodes), comme le montre l'analyse par gel SDS-PAGE (figure V.8) qui présente les variants de AC purifiés, ainsi que la préparation de CaM, utilisés dans les expériences ci-dessous.

Les propriétés cinétiques des variants de AC ont ensuite été analysées par des tests enzymatiques décrits dans la section Matériels et méthodes. L'adényl cyclase AC hydrolyse l'ATP en AMPc et Pyrophosphate inorganique (PPi), qui est à son tour hydrolysé en phosphate



**FIG. V.8:** Analyse sur gel SDS-PAGE des variants de AC purifiés, AC wild-type et CaM. AC, ACm3A, ACm2A, ACm1A ont été séparés sur un gel SDS-polyacrylamide de 4-12%. Après migration, le gel a été coloré avec du Bleu de Coomassie.

inorganique (Pi) par la pyrophosphatase inorganique (PPiase) ajouté au mélange réactionnel. Le molybdate vert de malachite présent dans le mélange forme un complexe avec les phosphates inorganiques (Pi) qui absorbe alors à 595 nm.

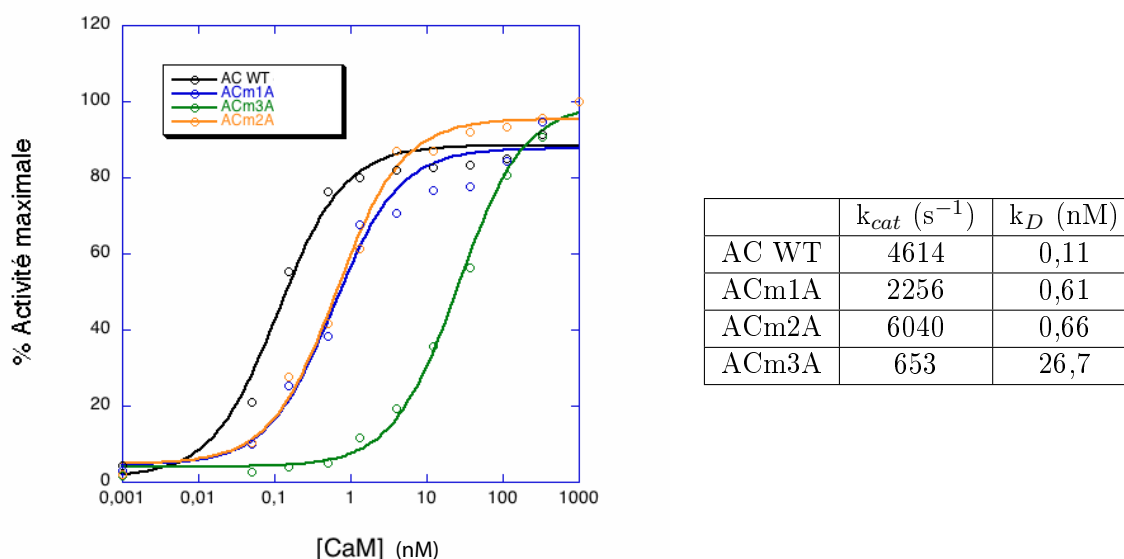


Pour mesurer la quantité d'AMPc produite, on suit donc l'apparition de phosphate inorganique (Pi). Ainsi, l'activité catalytique de AC est déterminée par la constante catalytique  $k_{cat}$ , qui est définie par la formule suivante :

$$k_{cat} = \frac{[AMPc]}{[AC]_{*t}} = \frac{1}{2} \frac{[Pi]}{[AC]_{*t}}$$

L'absorbance d'une solution de concentrations définies en Pi est mesurée en parallèle, pour convertir les valeurs de DO<sub>595</sub> en moles de P<sub>i</sub> produits.

Pour déterminer l'affinité des variants de AC pour CaM, l'activité enzymatique de chaque variant a été mesurée à différentes concentrations de CaM variant de 0 à 1 μM (figure V.9). Sur la figure V.9, l'activité enzymatique est exprimée pour chaque variant, en pourcentage d'activité maximale, déterminée à 1 μM de CaM.



**FIG. V.9: Activation des variants de AC par la CaM.** A gauche : L'activité adénylate-cyclase de AC sauvage (noté "WT") purifié et des mutants ACm3A, ACm2A et ACm1A a été mesurée en présence des concentrations de CaM indiquées sur la figure, et exprimée pour chaque variant en pourcentage d'activité maximale, mesurée en présence de 1 μM CaM. A droite : Tableau récapitulatif des activités catalytiques maximales des protéines, exprimées par la constante catalytique  $k_{cat}$ , et leur affinité pour la CaM, exprimée par la constante de dissociation ( $k_D$ ), qui correspond ici à la constante de demi-activation  $k_{1/2}$ .

L'activité enzymatique maximale de AC sauvage (notée AC WT sur la figure V.9) correspond à un  $k_{cat}$  de 4614 s<sup>-1</sup>. Les variants ACm1A et ACm3A, dans lesquels le résidu ASN347 qui interagit avec la boucle catalytique dans AC sauvage a été modifié en ALA, ont une activité catalytique réduite de moitié pour ACm1A ( $k_{cat} \simeq 2256$  s<sup>-1</sup>), et pour ACm3A une activité ne représentant que 14% de l'activité de AC sauvage ( $k_{cat} \simeq 653$  s<sup>-1</sup>). La présence dans ACm3A des deux mutations supplémentaires, ARG338ALA et ASP360ALA, qui à l'origine interagissent entre eux et avec le résidu ARG90 de la calmoduline, perturbe donc davantage l'activité enzymatique. Par contre, le variant ACm2A, dans lequel seuls ces deux derniers résidus ARG338ALA et ASP360ALA ont été modifiés, ne semble pas présenter d'altération au niveau de son activité. La perte d'activité importante observée pour ACm3A peut donc être due au fait que la boucle catalytique n'est plus stabilisée par le résidu ASN347 qui est modifié dans ces variants.

L'affinité de AC pour CaM est définie par sa constante de dissociation  $k_D$ , qui correspond ici à la constante de demi-activation  $k_{1/2}$ , c'est-à-dire à la concentration de CaM nécessaire pour activer AC à 50% de son activité maximale. Ainsi, AC sauvage possède une forte affinité pour CaM, avec un  $k_D \simeq 0.11$  nM. Les variants ACm1A et ACm2A présente une petite perte d'affinité pour la calmoduline ( $k_D \simeq 0.61$  s<sup>-1</sup> et 0.66 s<sup>-1</sup> respectivement). A l'inverse, le variant ACm3A, qui comporte les trois modifications simultanées, montre une perte d'affinité considérable ( $k_D \simeq 26.7$  nM) environ 250 fois supérieure à l'affinité de AC sauvage.

Les groupes de modifications (ARG338ALA et ASP360ALA) et (ASN347ALA) pris indépendamment ne semblent donc pas avoir beaucoup d'effet sur l'affinité de liaison avec la calmoduline. En revanche, l'affinité est très perturbée lorsque ces trois modifications sont introduites simultanément. Bien que la modification ASN347ALA altère également l'activité enzymatique de l'adénylcyclase, le cumul des trois modifications amplifie cette altération.

## 4 Discussion

Les trajectoires de dynamique moléculaire réalisées dans le chapitre IV, ont permis de mettre en évidence trois résidus, ARG338, ASN347 et ASP360, qui pourraient influencer l'interaction de AC avec la C-CaM. Dans ce chapitre, trois protéines modifiées ont été construites *in silico* et *in vitro* avec les modifications suivantes : la protéine modifiée ACm1A porte la mutation ASN347ALA, la protéine ACm2A porte les deux mutations ARG338ALA, ASP360ALA, et la protéine ACm3A porte les trois mutations ARG338ALA, ASN347ALA et ASP360ALA.

Dans un premier temps, chacun de ces variants de AC a été simulé en complexe avec C-CaM par dynamique moléculaire. Ainsi, l'analyse des trajectoires de DM réalisées sur ACm1A-(2Ca-C-CaM), ACm2A-(2Ca-C-CaM), ACm3A-(2Ca-C-CaM) et la comparaison avec les trajectoires sauvages du chapitre IV a révélé une réduction de l'interaction entre AC et C-CaM dans la trajectoire de ACm3A-(2Ca-C-CaM).

La majorité des liaisons hydrogène souvent observées dans les structures cristallographiques sont également formées dans plus de 80% des trajectoires de DM. L'analyse des trajectoires de DM a montré une forte variabilité dans l'établissement des liaisons hydrogène au sein du complexe. Cette variation est observée entre les différentes protéines, et parmi les trajectoires répétées sur la protéine sauvage. Néanmoins, cette variabilité ne peut pas être entièrement attribuée à la modélisation du système, étant donné que des variations similaires sont observables sur les quatre structures cristallographiques du complexe AC/C-CaM. De plus, les liaisons hydrogène les plus stables au niveau de l'interface C-CaM/AC dans les structures cristallographiques sont aussi celles qui présentent le plus grand pourcentage de formation au cours des trajectoires sauvages. Les liaisons hydrogène des résidus de la boucle L $\alpha$ 16Cter sont également observées avec un pourcentage de formation assez élevé au cours des trajectoires de DM, en particulier les liaisons hydrogène VAL-343-N/PRO-196-O et celles formées entre ASN347 et les résidus de la boucle catalytique.

Les résidus modifiés sont également impliqués dans plusieurs liaisons hydrogène et ponts d'eau dans les structures cristallographiques. Pendant la préparation des simulations de dynamique moléculaire, l'interaction avec l'eau a été modifiée par la procédure de solvatation. Mais l'importance des interactions établies par ces résidus est maintenue dans les trajectoires de dynamique moléculaire par le pourcentage de formation élevé de ces interactions au cours du temps.

Les différents variants de AC ont ensuite été caractérisés expérimentalement. L'affinité de AC sauvage et des variants ACm1A et ACm2A pour CaM est similaire, tandis que celle de ACm3A a été réduite d'un facteur 250 environ. L'activité de ACm3A a été réduite à environ 15 % de l'activité de AC sauvage, et celle de ACm1A d'environ 50 %, ce qui est sans doute dû au fait que le résidu N347 interagit directement avec la boucle catalytique. Toute perturbation de l'interaction entre N347 et la boucle catalytique induit donc une perturbation de l'activité enzymatique. Cependant, la présence des deux mutations ARG338ALA et ASP360ALA supplémentaires qui interagissent avec la calmoduline, perturbe davantage l'activité adénylcyclase.

L'analyse des trajectoires sauvages et des mutants conduite en parallèle avec la caractérisation expérimentale des protéines modifiées montre que la perte d'affinité de ACm3A est due à la perte d'interactions entre C-CaM et CA. Cette déstabilisation est observée seulement pour ACm3A. ACm2A contenant deux mutations à l'interface C-CaM/AC et ACm1A contenant une mutation unique à l'interface Cter/boucle catalytique, ne montrent pas une telle perturbation. Ces observations montrent que ni la seule mutation ASN347ALA, proche de la boucle catalytique, ni les deux mutations ARG338 et ASP360 ne sont suffisantes pour déstabiliser l'interface AC/CaM. En effet, les modifications des deux résidus situés à l'interface C-CaM/AC sont compensées par la formation d'autres liaisons hydrogène qui la stabilisent. En revanche, la mutation des trois résidus induit coopérativement la perte d'affinité pour la CaM, ainsi qu'une réduction importante de l'activité catalytique (15 % de l'activité sauvage). La chaîne peptidique de la boucle L $\alpha$ 16Cter, qui contient les résidus ASN347 et ARG338 à ses deux extrémités, constitue donc un moyen de transmission mécanique de l'information pour l'induction des effets conformationnels.

Dans les travaux de Guo et ses collaborateurs, décrivant la structure cristallographique du complexe AC/C-CaM [131], le double mutant portant les mutations GLU346ALA et ARG348ALA dans la région C-terminale de AC a été construit et caractérisé par ses paramètres cinétiques et son interaction avec la calmoduline. Ce double mutant GLU346ALA/ARG348ALA encadre le résidu ASN347 qui a été modifié dans ce chapitre. Il présente une activité réduite pour des concentrations de calcium libre de 1 et 10  $\mu$ M (pour une concentration de  $\text{Ca}^{2+}$  de 10  $\mu$ M,  $k_{cat} \simeq 1900 \text{ s}^{-1}$  pour AC sauvage,  $k_{cat} \simeq 70 \text{ s}^{-1}$  pour le double mutant GLU346ALA/ARG348ALA). De plus, son affinité pour la CaM est également réduite d'un facteur 204. D'après les résultats des simulations de DM, les liaisons hydrogène dans lesquelles les résidus GLU346 et ARG348 sont impliqués sont majoritairement rompues dans les trajectoires sauvages lorsque les ions calcium ou la C-CaM sont retirés. Dans les trajectoires des mutants, les liaisons hydrogène établies par GLU346

et ARG348 avec les résidus de la C-CaM montrent également un pourcentage de formation de liaisons hydrogène réduit. La sensibilité de ces résidus vis-à-vis des mutations décrites ici révèle qu'ils sont mécaniquement corrélés, et soutient le modèle dans lequel la boucle L $\alpha$ 16Cter est utilisée comme un moyen de transmission des tendances conformationnelles.

## 5 Conclusion

Les effets des modifications introduites dans les protéines ACm1A (ASN347ALA), ACm2A (ARG338ALA, ASP360ALA) et ACm3A (ARG338ALA, ASN347ALA, ASP360ALA) sur l'interaction entre le domaine AC de *Bordetella pertussis* et la calmoduline ont été analysés en mesurant expérimentalement l'affinité de AC pour la CaM, et en réalisant des simulations de dynamique moléculaire sur AC modifié, en complexe avec la C-CaM. ACm1A et ACm2A affichent une affinité pour la CaM similaire à celle observée pour AC sauvage, tandis que ACm3A montre une affinité très réduite. L'analyse des simulations de DM suggère que seul ACm3A montre une déstabilisation de l'interface C-CaM/AC. Cette déstabilisation est transmise de la région proche de la boucle catalytique à l'interface C-CaM/AC à travers la boucle L $\alpha$ 16Cter.

# Chapitre VI

## Exploration de la transition conformationnelle de l'état lié à l'état non-lié de AC

---

### 1 Introduction

Durant les quinze dernières années, l'importance des fluctuations dans la fonction et les interactions d'une protéine a été soulignée dans plusieurs études [228; 229; 230; 231; 232; 233; 234; 235; 236]. Bien que ces fluctuations soient très importantes pour assurer la fonction d'une protéine, il est très difficile de les caractériser à une résolution atomique. En effet, les méthodes biophysiques les plus utilisées pour la détermination d'une structure protéique (RMN et cristallographie aux rayons X) requièrent une certaine stabilité conformationnelle de la protéine.

Pour la cristallographie aux rayons X, cette stabilité est même requise pour la formation de cristaux. Les études par RMN se font sur des échantillons en solution dans laquelle l'équilibre conformationnel est possible. Néanmoins, dans le cas d'échanges conformationnels vastes, les spectres RMN deviennent trop surchargés pour pouvoir être analysés, ou le signal RMN peut aussi disparaître.

Les conformations intermédiaires d'une protéine échantillonnées entre les échanges conformationnels sont très importantes pour caractériser une meilleure compréhension de la fonction protéique, mais d'un autre côté, elles sont également inaccessibles par les méthodes classiques de biologie structurale. C'est pourquoi les méthodes biophysiques computationnelles offrent une alternative fiable pour avoir un aperçu sur les conformations intermédiaires dont la structure ne peut être caractérisée autrement. Cependant, il est à noter que ce problème computationnel est extrêmement complexe.

Jusqu'à présent, de nombreuses méthodes ont été utilisées pour obtenir des informations sur les conformations transitoires des protéines. Une première classe de méthodes, basée sur la mécanique moléculaire, a été initiée par l'affinement par pic conjugué [164] ("*Conjugate peak refinement*" ou *CPR*). Le réseau élastique [237; 238; 239] donne des informations qualitatives sur les transi-



tions conformationnelles. Une méthode récemment proposée [184] a l'avantage de déterminer des chemins de transitions finement raffinés, obtenus par une utilisation itérative et extensive du raffinement de pic conjugué, suivis par le classement des conformations intermédiaires obtenues, afin d'obtenir le chemin le plus direct possible dans l'espace conformationnel de la protéine. Cette classe de méthodes requiert la connaissance des deux conformations extrêmes de la transition. Elles sont habituellement appliquées avec un modèle de solvation implicite, ce qui peut être un facteur limitant, étant donné que le comportement d'un solvant explicite peut influencer la conformation de la protéine.

Une seconde classe de méthodes est fondée sur les équations de dynamique moléculaire, qui se sont avérées efficaces dans certains cas [240; 203; 241]. D'autres méthodes d'exploration de l'espace conformationnel s'appuient sur une déformation de l'énergie potentielle du système, dans le but de favoriser la sortie d'un bassin initial de l'hyper-surface conformationnelle. Ces méthodes ont été proposées il y a déjà quelques années [242]. Elles ont pris récemment un nouveau départ, avec le développement de la métadynamique [243; 244] et la dynamique moléculaire accélérée [245; 246].

Un autre aspect de l'exploration de l'espace conformationnel d'une protéine est l'importance que l'aspect probabilistique des transitions observées soit pris en compte. En effet, la variation conformationnelle de la protéine doit être prise en compte dans le cadre de la mécanique statistique, ce qui signifie que tous les chemins sont en principe possibles, mais que leur probabilité est modulée par l'énergie de transition dont ils ont besoin. Cet aspect est sous-jacent à toutes les méthodes décrites précédemment, qui sont basées sur des simulations de dynamique moléculaire. Il a été exploré et formalisé dans le développement des méthodes d'échantillonnage des chemins de transition [247], et pour la détermination des étapes clés d'une transition [248]. Le formalisme mathématique soulignant l'échantillonnage de chemin de transition a été décrit par Vanden-Eijnden (2010) [249].

Basée sur ce formalisme, la dynamique moléculaire accélérée par la température (dit en anglais "Temperature Accelerated Molecular Dynamic (TAMD)"), a été proposée [150] dans le but d'explorer l'espace conformationnel des protéines, en fonction des variables collectives choisies auparavant. Cette méthode fait évoluer les valeurs cibles des variables collectives en utilisant le thermostat de Langevin, qui fonctionne à une température plus élevée que la température de la trajectoire de dynamique moléculaire. La TAMD présente l'avantage de forcer les variables collectives à évoluer sur une surface d'énergie libre non biaisée du système.

Au cours des années précédentes, la TAMD a démontré son efficacité sur deux problèmes biologiques différents. Elle a tout d'abord été utilisée pour explorer les chemins des déplacements des CO autour de la myoglobine, et pour construire une carte des barrières d'énergie libre rencontrées par les CO lors leur entrée et sortie de la protéine [250]. Puis, dans un deuxième temps, la TAMD a été appliquée pour décrire la transition conformationnelle de deux protéines : GroEL et HIV-1 gp120 [251]. Des variables collectives ont été choisies à partir des structures cristallographiques des deux extrémités de la transition. Puis une TAMD a été produite sur 40 ns

à partir d'une des deux extrémités, afin d'induire une transition conformationnelle de la protéine, de façon à ce qu'elle se rapproche d'une valeur d'au moins 3 Å de l'autre extrémité du chemin.

Dans ce chapitre, notre objectif est d'obtenir des informations sur la transition conformationnelle du domaine catalytique AC de l'adénylcyclase CyaA, à partir de la structure cristallographique de AC complexé avec le lobe C-CaM de la calmoduline [131], la structure de AC isolé à haute résolution étant inconnue. Néanmoins, des analyses biophysiques approfondies [77] apportent les informations sur AC isolée suivantes : (i) la quantité de structure secondaire ne varie pas (voire très peu) entre les états liés et non-liés, (ii) l'état non-lié décrit une forme bien plus compacte que dans l'état lié, avec des paramètres hydrodynamiques proches de ceux d'une sphère. De plus, AC isolé a tendance à s'agréger, ce qui empêche jusqu'à présent sa cristallisation, et rend les analyses des courbes de SAXS impossibles (A. Chenal, communication personnelle).

On recherche des informations sur l'état non-lié de cette protéine. Cet état correspond très probablement à une série de conformations en échange conformationnel les unes avec les autres. Le cas où une protéine est stabilisée par les interactions d'un partenaire et où elle présente une structure instable en l'absence de ce dernier, est fréquemment rencontré dans les systèmes biologiques. Dans le cas présent, (i) l'état instable n'est pas précisément défini par une seule conformation, mais plutôt par une série de conformations qui s'échangent, (ii) les chemins de transition conformationnelle sont également mal définis à cause du manque d'information sur la structure d'une extrémité de la transition. Le problème présent est donc plus complexe et difficile à résoudre que les cas précédemment étudiés [251], mais même des aperçus modestes de l'état dynamique de la protéine pourrait procurer des informations suffisamment riches pour être analysées et testées par les biochimistes.

Dans cette étude, le comportement dynamique de AC dans les simulations de dynamique moléculaire classique [252] a été analysé afin de déterminer un jeu de variables collectives requises pour décrire la dérive de AC de l'état lié à l'état non-lié. Les trajectoires de TAMD produites avec ces variables collectives ont permis un échantillonnage de l'espace conformationnel plus vaste, et une dérive plus importante vers une forme plus compacte de l'état non-lié. Cependant, les conformations qui ont extrêmement dérivé au cours de la TAMD ne restent pas stables au cours d'une DM classique non restreinte à la suite de la TAMD. Une recherche heuristique supplémentaire a dû être réalisée dans le but d'obtenir des conformations stables de AC au cours de trajectoires de DM. Ces conformations explorent un bassin d'énergie minimum local correspondant à l'état non-lié de AC.

Ces travaux soulèvent également la question d'une bonne stratégie pour atteindre les bassins énergétiques correspondant à un minimum d'énergie local, dans le cas où un point final de la transition correspond à un état très dynamique, non défini par une structure à haute résolution.

## 2 Matériels et méthodes

### 2.1 Simulations de Dynamique Moléculaire

Les trajectoires de DM et de TAMD ont été produites avec NAMD 2.7b2 [157] sur des clusters de 4 et 8 processeurs. Un rayon de coupure de 12 Å et une distance de "switch" de 10 Å ont été définis pour les calculs des interactions non-liées. Les interactions électrostatiques à longue portée ont été calculées avec le protocole Particle Mesh Ewald (PME) [159]. Les simulations ont été réalisées dans l'ensemble NPT. La température a été réglée à 300 K avec le thermostat de Langevin [253], et la pression à 1 atm avec la méthode de Nose-Hoover qui utilise le piston de Langevin [254; 255]. L'algorithme RATTLE [154; 155] a été appliqué pour garder les liaisons covalentes impliquant les hydrogènes rigides, permettant ainsi l'utilisation d'un pas de temps de 2 fs. Les coordonnées atomiques ont été enregistrées toutes les picosecondes. Le champ de force CHARMM22 a été appliqué avec la correction CMAP [256; 257]. L'eau a été simulée avec le modèle TIP3P [258].

Les trajectoires ont été initiées avec une minimisation de 1000 pas, puis thermalisées de 0 K à 300 K avec un pas d'intégration de 30000 pas, et enfin, équilibrées sur 50000 pas. Les paramètres principaux des simulations sont donnés dans la table I.

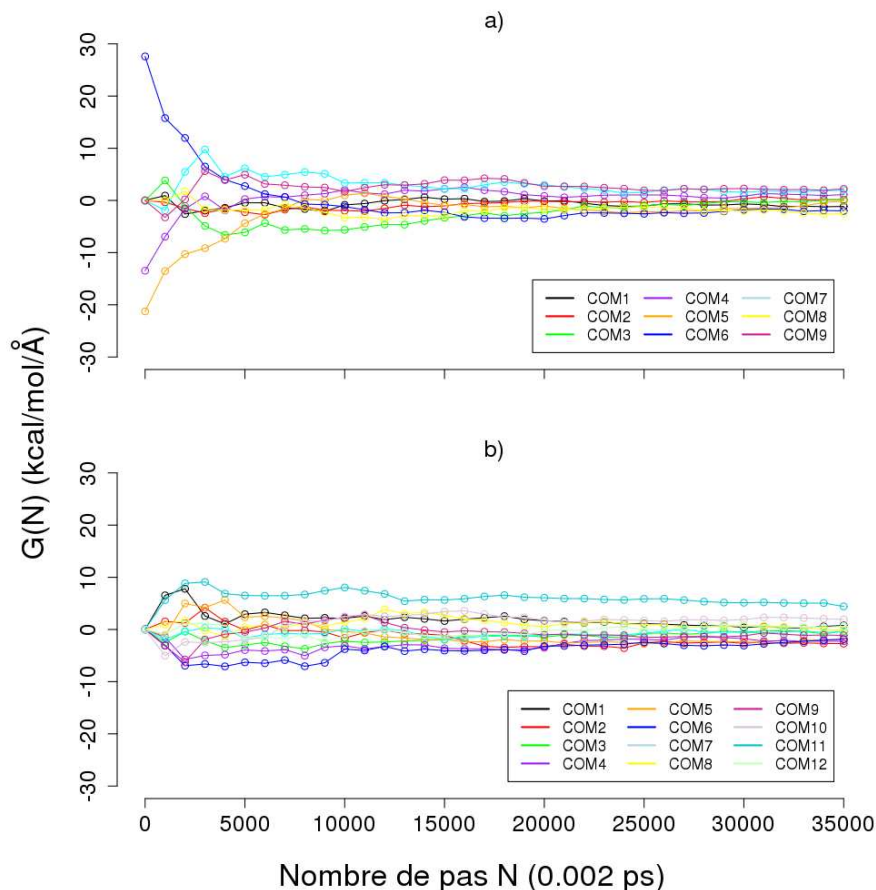
### 2.2 Simulations TAMD

Les principes des simulations de dynamique accélérée par la température ont été décrits dans le chapitre III. Les simulations TAMD ont été précédées par 1 ns de DM classique réalisée dans les mêmes conditions que les simulations de DM du chapitre IV. Trois trajectoires de TAMD, chacune sur 40 ns, ont ensuite été produites dans les mêmes conditions que les DM mais en utilisant trois ou quatre centres de masse comme variables collectives :  $(CM_1^a, CM_2^a, CM_3^a)$  et  $(CM_1^b, CM_2^b, CM_3^b, CM_4^b)$  qui sont décrits dans la partie Résultats et la table II. Pour chaque simulation de TAMD, les paramètres du thermostat de Langevin de DM sont :  $\gamma^{-1} = 0.5$  ps, et  $\beta^{-1} = 0.6$  kcal/mol, correspondant à une température de 298 K.

L'algorithme de TAMD a été implémenté dans NAMD en réalisant une simulation de DM couplée à un script tcl pour le calcul de l'évolution des variables collectives. Celles-ci sont contraintes par une constante de force  $\kappa$  de 100 kcal/(mol.Å<sup>2</sup>). Le thermostat de Langevin artificiel, relié aux variables collectives, a été paramétré de la manière suivante. Le coefficient de friction artificiel  $\bar{\gamma}$  a été déterminé en produisant 70 ps de trajectoire TAMD dite "froide", dans laquelle on ne fait pas évoluer les variables collectives. Puis, à partir de ces trajectoires de TAMD "froides", le paramètre  $G_j(N)$  a été calculé pour chaque variable collective  $j$  comme la somme sur  $N$  pas de temps, comme il l'a été proposé par Abrams (2010) [251], :

$$G_j(N) = \frac{\kappa}{N} \sum_{i=1}^N [\theta_j^*(x(t_i)) - \theta_j] \quad (\text{VI.1})$$

où  $\theta_j$  est la valeur cible et  $\theta_j^*(x(t_i))$  la valeur instantanée au temps  $t_i$  pour les variables collectives. La variation de  $G_j(N)$  le long du pas de temps de la trajectoire (Figure VI.1) permet l'extraction du temps de relaxation caractéristique des variables collectives.



**FIG. VI.1:** Moyenne mobile  $G(N)$  calculée au cours du temps de la simulation de DM dite "froide".  $G(N)$  est calculé sur (a) les 9 variables collectives correspondant aux trois centres de masse, (b) les 12 variables collectives correspondant aux coordonnées Cartésiennes des quatre centres de masse  $CM_1^b, CM_2^b, CM_3^b, CM_4^b$ .

Pour les neuf variables collectives correspondant aux coordonnées de  $(CM_1^a, CM_2^a, CM_3^a)$  (Figure VI.2a) et table II) et pour les douze variables collectives correspondant aux coordonnées de  $(CM_1^b, CM_2^b, CM_3^b, CM_4^b)$  (Figure VI.2b) et Table II), les paramètres  $G_j(N)$  retournent à la valeur à l'équilibre avant 5000 pas de simulation. Comme le pas d'intégration des simulations est de 0.002 ps, un coefficient de friction  $\bar{\gamma}$  de  $50 \text{ ps}^{-1}$  est assez long pour permettre au système de se relaxer.

Dans le but d'explorer l'espace conformationnel de AC à différents niveaux, les trajectoires de TAMD sont produites en utilisant trois valeurs différentes de l'énergie thermique artificielle  $\beta^{-1}$  du thermostat de Langevin artificiel reliés aux variables collectives. Les valeurs de 4, 6 et 20  $\text{kcal.mol}^{-1}$  ont été appliquées, correspondant ainsi aux températures  $\bar{T}$  de 2012, 3019 et 10064 K. Ces principaux paramètres variables ont donné leur nom aux simulations dans la table I.

Nom	Point de départ	Algorithme	Variables Collectives	$\beta^{-1}$ (kcal.mol <sup>-1</sup> )	Temps simulé (ns)
MD	1YRT	DM	-	-	40
MD-CONT-36	TAMD 36 ns	DM	-	-	10
MD-CONT-39	TAMD 39 ns	DM	-	-	10
MD-CONT-40	TAMD 40 ns	DM	-	-	10
MD-MIN-25	Conformation 25 minimisée <sup>a</sup>	DM	-	-	20
MD-MIN-28	Conformation 28 minimisée <sup>a</sup>	DM	-	-	20
MD-MIN-32	Conformation 32 minimisée <sup>a</sup>	DM	-	-	20
MD-MIN-34	Conformation 34 minimisée <sup>a</sup>	DM	-	-	20
TAMD-3COM-4	1YRT	TAMD	CM <sub>1</sub> <sup>a</sup> CM <sub>2</sub> <sup>a</sup> CM <sub>3</sub> <sup>a</sup>	4	40
TAMD-3COM-6	1YRT	TAMD	CM <sub>1</sub> <sup>a</sup> CM <sub>2</sub> <sup>a</sup> CM <sub>3</sub> <sup>a</sup>	6	40
TAMD-3COM-20	1YRT	TAMD	CM <sub>1</sub> <sup>a</sup> CM <sub>2</sub> <sup>a</sup> CM <sub>3</sub> <sup>a</sup>	20	40
TAMD-4COM-4	1YRT	TAMD	CM <sub>1</sub> <sup>b</sup> CM <sub>2</sub> <sup>b</sup> CM <sub>3</sub> <sup>b</sup> CM <sub>4</sub> <sup>b</sup>	4	40
TAMD-4COM-6	1YRT	TAMD	CM <sub>1</sub> <sup>b</sup> CM <sub>2</sub> <sup>b</sup> CM <sub>3</sub> <sup>b</sup> CM <sub>4</sub> <sup>b</sup>	6	40
TAMD-4COM-20	1YRT	TAMD	CM <sub>1</sub> <sup>b</sup> CM <sub>2</sub> <sup>b</sup> CM <sub>3</sub> <sup>b</sup> CM <sub>4</sub> <sup>b</sup>	20	40

**TAB. I: Paramètres des trajectoires.** Les conformations minimisées ont été obtenues à partir de la conformation extraite à 36 ns de la TAMD-4COM-20, qui a été minimisée itérativement.

Trois centres de masse	
CM <sub>3</sub> <sup>a</sup> (COM1)	209-244
CM <sub>2</sub> <sup>a</sup> (COM2)	7-43 189-200 249-265 268-292 298-364
CM <sub>1</sub> <sup>a</sup> (COM3)	55-59 61-187
Quatre centres de masse	
CM <sub>4</sub> <sup>b</sup> (COM1)	225-231
CM <sub>3</sub> <sup>b</sup> (COM2)	201-205 207 208 245-248
CM <sub>2</sub> <sup>b</sup> (COM3)	7-34 38-43 188-194 261-310 313-324
CM <sub>1</sub> <sup>b</sup> (COM4)	61-185

**TAB. II: Définition des centres de masse utilisés comme variables collectives dans les simulations de TAMD.** Les numéros sont les numéros de résidus.

## 2.3 Echantillonnage des conformations

La convergence des trajectoires ainsi que l'échantillonnage de leur population conformationnelle a été analysée avec l'approche développée par Lyman et Zuckerman [192] décrit au chapitre IV.

## 3 Résultats

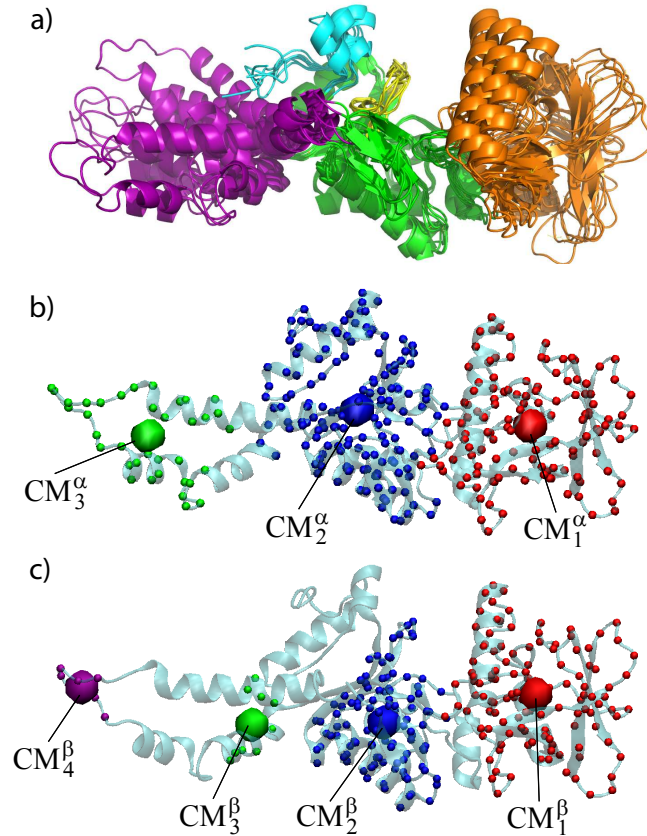
### 3.1 Choix des variables collectives

Le premier paramètre principal de la TAMD à définir est le jeu de variables collectives, suivi par le paramètre  $\bar{\gamma}$  qui est déduit à partir d'une analyse directe du système. Le troisième paramètre,  $\bar{\beta}^{-1}$ , est ensuite choisi pour contrôler la température artificielle à laquelle les variables collectives doivent évoluer. Dans le cas présent, seul le point initial de la transition conformationnelle est connu. Il s'agit de la structure de AC complexée avec la calmoduline chargée en ions calcium [131]. Seules des informations qualitatives de l'état final de cette transition sont disponibles grâce à des mesures hydrodynamiques [77], qui décrivent AC isolée comme étant de forme plus globulaire. De plus, les difficultés rencontrées pour la cristallisation de AC affirme l'existence d'un équilibre conformationnel de AC entre plusieurs conformations. C'est pourquoi le choix des variables collectives décrivant la transition de AC d'un état complexé à un état libre, ne peut être que qualitativement orienté par les informations disponibles. Cependant, ce choix est essentiel, car les résultats obtenus par la TAMD vont dépendre de lui.

Les variables collectives ont donc été choisies de la manière suivante, à partir d'une dynamique moléculaire classique qui a été effectuée sur le système AC isolé décrit dans le chapitre IV.

La trajectoire a été analysée par l'approche de Lyman et Zuckerman [192]. Sept conformations représentatives ont ainsi été extraites (chapitre IV) avec une valeur de coupure de 3 Å. On suppose qu'elles correspondent à un bassin conformationnel de forme harmonique entourant le minimum local défini par la structure de AC déterminée par cristallographie aux rayons X (entrée PDB : 1YRT [131]). Le protocole PiSQRD [259] a ensuite été utilisée sur ces sept conformations pour déterminer les domaines dont les mouvements décrivent la variabilité de celles-ci. Le programme est disponible sur l'interface Web : <http://pisqrd.escience-lab.org> [260]. PiSQRD découpe la protéine en groupes de résidus dont les fluctuations sont bien corrélées. La matrice de covariance de chaque conformation représentative a été modélisée par un réseau élastique fondé sur le modèle  $\beta$ -Gaussian [261], dans lequel chaque amino-acide est représenté par deux centroïdes pour la chaîne peptidique et la chaîne latérale. Les dix premiers modes de la matrice de covariance sont retenus. La décomposition en domaines rigides est déterminée de façon à retenir au moins 80% de la mobilité interne de la protéine.

Le protocole PiSQRD a ainsi produit un découpage en trois domaines rigides sur chaque conformation représentative. Les résidus impliqués dans chaque domaine et retrouvés dans les sept conformations ont été retenus pour définir les centres de masse. De cette façon, on évite le biais qui peut être introduit par la mobilité observée seulement dans quelques parties de la simulation de dynamique moléculaire. Il est également possible de spécifier au logiciel le nombre total de domaines souhaité. Ainsi, nous avons également généré un découpage en quatre domaines pour les mêmes structures représentatives, et calculé les centres de masse de la même façon que précédemment.



**FIG. VI.2:** (a) Les sept conformations représentatives de la trajectoire de DM classique de AC isolé [252], extraites par l'algorithme de Lyman et Zuckerman [192]. Les régions protéiques sont colorées en violet (SA incluant la boucle 226-232 modélisée par homologie), en vert (CA), orange (CB), cyan (queue C-terminale : Cter) et jaune (boucle catalytique : boucle C). Les trois (b) et quatre (c) centres de masse obtenus à partir de l'analyse de PiSQRD des conformations représentatives, sont indiqués de la même façon que dans la table II. Dans (b), les atomes  $C\alpha$  qui définissent les différents centres de masse sont colorés en vert ( $COM1 : CM_3^a$ ), bleu ( $COM2 : CM_2^a$ ) et rouge ( $COM3 : CM_1^a$ ). Dans (c) les atomes  $C\alpha$  qui définissent les différents centres de masse sont colorés en violet ( $COM1 : CM_4^b$ ), vert ( $COM2 : CM_3^b$ ), bleu ( $COM3 : CM_2^b$ ) et rouge ( $COM4 : CM_1^b$ ).

Ainsi, à partir des analyses avec PiSQRD, on a pu décrire la variabilité conformationnelle de AC à travers 3 ou 4 centres de masse :  $(CM_1^a, CM_2^a, CM_3^a)$  et  $(CM_1^b, CM_2^b, CM_3^b, CM_4^b)$  (Figure VI.2).

Les résidus qui définissent les différents centres de masse sont détaillés dans la table II. Dans les deux cas, un centre de masse ( $CM_1^a$  ou  $CM_1^b$ ) est obtenu à partir des résidus formant la région CB avec des définitions assez similaires. Les centres de masse  $CM_2^a$  et  $CM_2^b$  sont localisés dans la région CA.

Si on demande au logiciel PiSQRD de définir 3 centres de masse,  $CM_2^a$  inclut les résidus 325-364, qui correspondent à la région C-terminale et une hélice  $\alpha$  supplémentaire. Ces résidus ne sont pas inclus dans la définition de  $CM_2^b$ . Ils subissent une large réorganisation durant les simulations de DM réalisées sur AC-isolé [252], et ils peuvent donc être considérés comme un groupe qui bouge indépendamment du reste de la structure protéique.

Dans la région SA, un ( $CM_3^a$ ) ou deux ( $CM_3^b, CM_4^b$ ) centres de masse peuvent être obtenus en fonction du nombre de domaines rigides (3 ou 4) définis dans PiSQRD. Le centre de masse  $CM_3^a$  est défini à partir d'une région de SA, plus proche de CA : la moitié de l'hélice H, l'hélice G et les résidus 209, 210 de la partie C-terminale de l'hélice F. Par ailleurs, les deux centres de masse  $CM_3^b$  et  $CM_4^b$  sont déterminés aux deux extrémités de SA, l'un plus proche de SA et l'autre incluant la boucle 226-232. Cette boucle a été reconstruite par homologie car elle n'était pas présente

dans la structure 1YRT. Elle montre un comportement très flexible dans les simulations de DM précédentes (chapitre IV)[252].

### 3.2 Exploration de l'espace conformationnel de AC

Les trajectoires de TAMD ont été simulées sur 40 ns avec les centres de masse ( $CM_1^a$ ,  $CM_2^a$ ,  $CM_3^a$ ) et ( $CM_1^b$ ,  $CM_2^b$ ,  $CM_3^b$ ,  $CM_4^b$ ) choisis précédemment, et des températures  $\bar{\beta}^{-1}$  de 4, 6 et 20 kcal.mol<sup>-1</sup>. Le choix de ces valeurs est justifié par le fait qu'elles sont également employées dans l'exploration conformationnelle des transitions de la sous-unité GRoEL et HIV-1 gp120 [251] qui sont de même taille que AC. En effet, la sous-unité GRoEL de *E Coli* et HIV-1 gp120 contiennent respectivement 547 et 379 résidus tandis que AC est constitué de 358 résidus. De ce fait, des barrières énergétiques similaires entre ces trois protéines peuvent être attendues.

Au cours des trajectoires de TAMD, la dérive conformationnelle du domaine AC (figure VI.3) augmente de 2 à 8 Å lorsqu'on utilise les variables collectives ( $CM_1^a$ ,  $CM_2^a$ ,  $CM_3^a$ ) et de 2 à 12 Å avec les variables collectives ( $CM_1^b$ ,  $CM_2^b$ ,  $CM_3^b$ ,  $CM_4^b$ ).

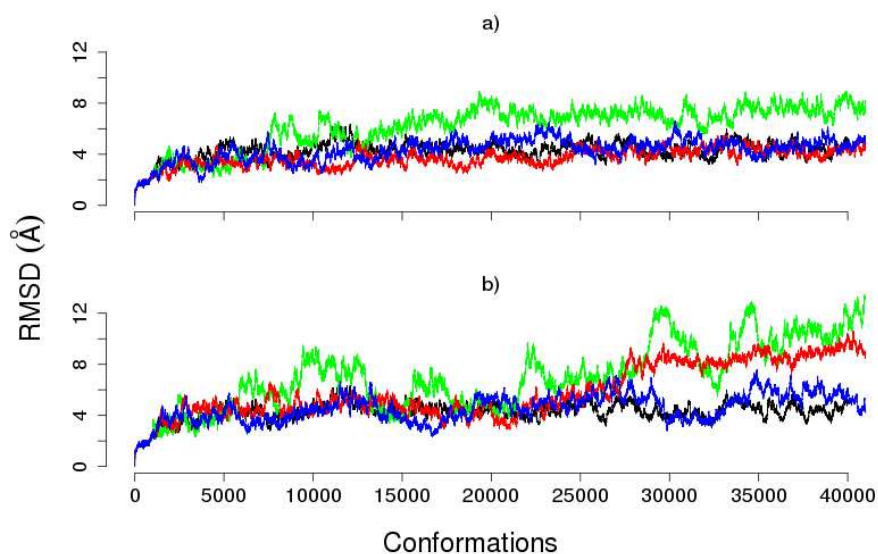
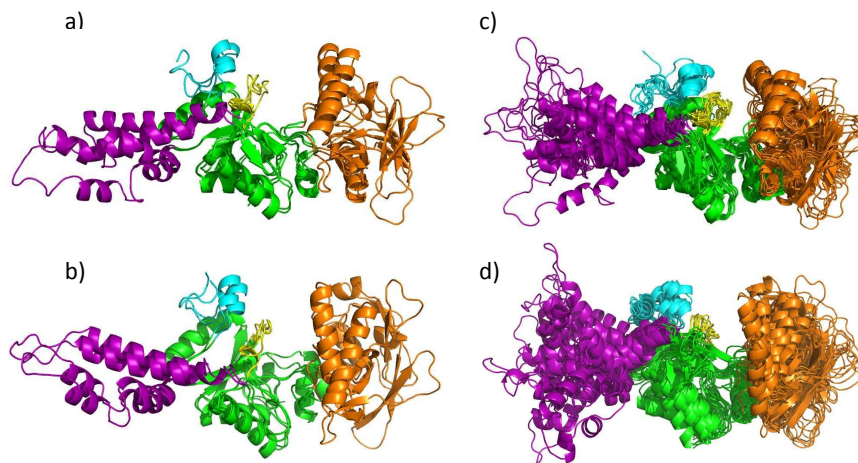


FIG. VI.3: *Dérives conformationnelles (Å) calculées sur les atomes Ca du domaine AC, au cours de la trajectoire de DM et des trajectoires de TAMD produites avec trois (a) ou quatre (b) centres de masse.* Les courbes sont colorées en bleu, rouge et vert pour les énergies thermiques artificielles  $\bar{\beta}^{-1}$  de 4, 6 et 20 kcal.mol<sup>-1</sup>. La courbe calculée pour la DM classique est colorée en noir.

Des valeurs de RMSD plus grandes sont obtenues lorsqu'on utilise quatre centres de masse, ce qui montre que ces dernières sont plus efficaces pour déplacer le système en-dehors du bassin énergétique correspondant à la conformation de AC complexé avec la C-CaM. Les dérives conformationnelles maximales observées pour la TAMD sont deux à trois fois plus grandes que celle observée pour la simulation de DM produite avec NAMD (courbe noire de la figure VI.3).



Les conformations représentatives de chaque trajectoire ont été extraites avec la méthode de Lyman et Zuckerman [192], pour un rayon de coupure de 5 Å. La superposition de celles-ci sur la région CA (figure VI.4 b-d) révèle que les orientations relatives de SA, CA et CB varient de manière beaucoup plus importante dans les trajectoires de TAMD que dans la trajectoire de DM classique (figure VI.4 a)).



**FIG. VI.4:** Conformations représentatives extraites des trajectoires de (a) DM et (b-d) TAMD produites avec les quatre centres de masse pour variables collectives. Les trajectoires de TAMD ont été simulées avec les énergies thermiques artificielles  $\beta^{-1}$  de 4 (b), 6 (c), et 20 (d) kcal.mol<sup>-1</sup>. Les régions de la protéine sont colorées en violet (SA), vert (CA), cyan (C terminal), jaune (boucle catalytique) et orange (CB). Les conformations ont été superposées sur la région CA.

Les dérives conformationnelles des régions CA, CB et SA ont été analysées sur les TAMD produites avec quatre centres de masse pour variables collectives (figure VI.5). Pour les trois valeurs de  $\bar{\beta}^{-1}$ , la dérive conformationnelle globale est bien plus élevée que pour les régions CA, CB et SA considérées individuellement. La région SA dérive le plus, suivie par la région CA puis CB. La faible dérive de CB peut venir d'un effet de rigidification induit par l'utilisation d'un centre de masse unique pour CB. La région SA, pour laquelle deux centres de masse ont été introduits, ne dérive qu'un peu plus. Et la dérive conformationnelle de la région CA, dans lequel seul un centre de masse a été défini, atteint la dérive de SA à la fin des trajectoires. Cependant, dans la région CA, les résidus 325-364, qui correspondent à la région C-terminale et une hélice  $\alpha$ , ne sont pas inclus dans les résidus à partir desquels le centre de masse a été calculé. L'accélération de la dérive, induite par la TAMD, ne dépend pas uniquement du nombre de centres de masse, mais également de la stabilité intrinsèque relative des différentes régions. En effet, les résidus 325-364 incluant la région C terminal et une hélice supplémentaire, montrent une réorganisation importante dans la MD classique décrite dans le chapitre IV. Ces mêmes résidus induisent également une dérive de CA même s'ils ne sont pas inclus dans la définition de centres de masse (figure VI.5).

Les régions CA et CB, localisées autour du site catalytique, subissent également des réarrange-

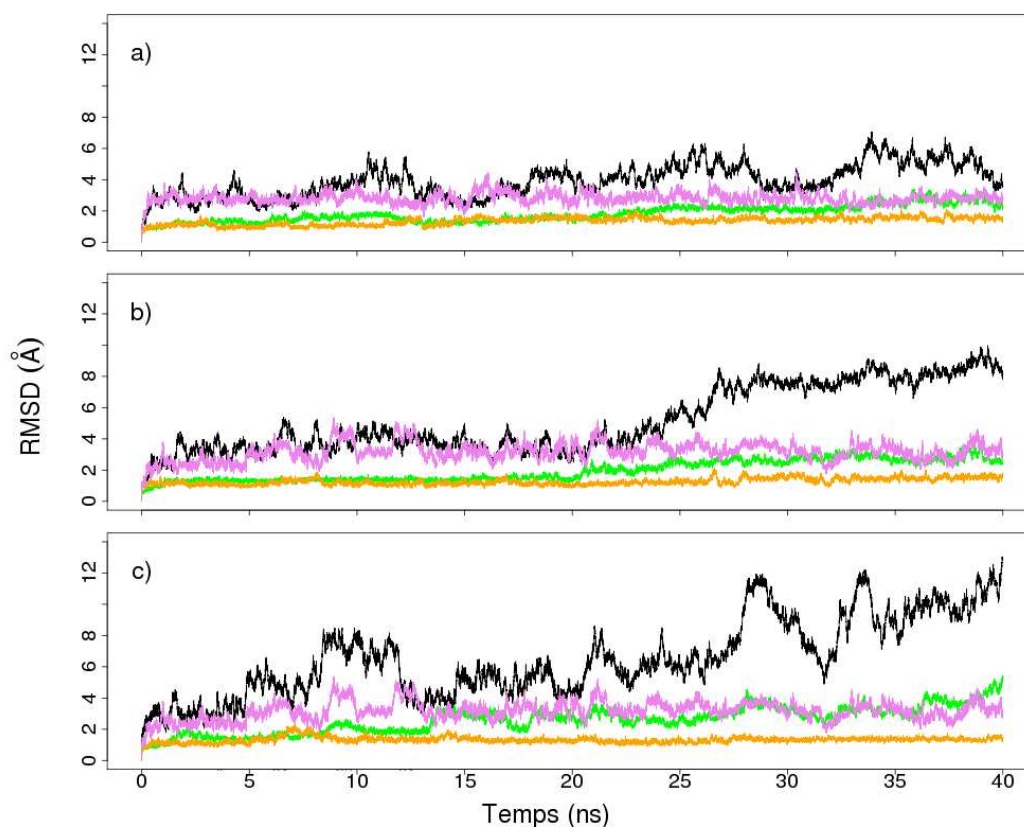
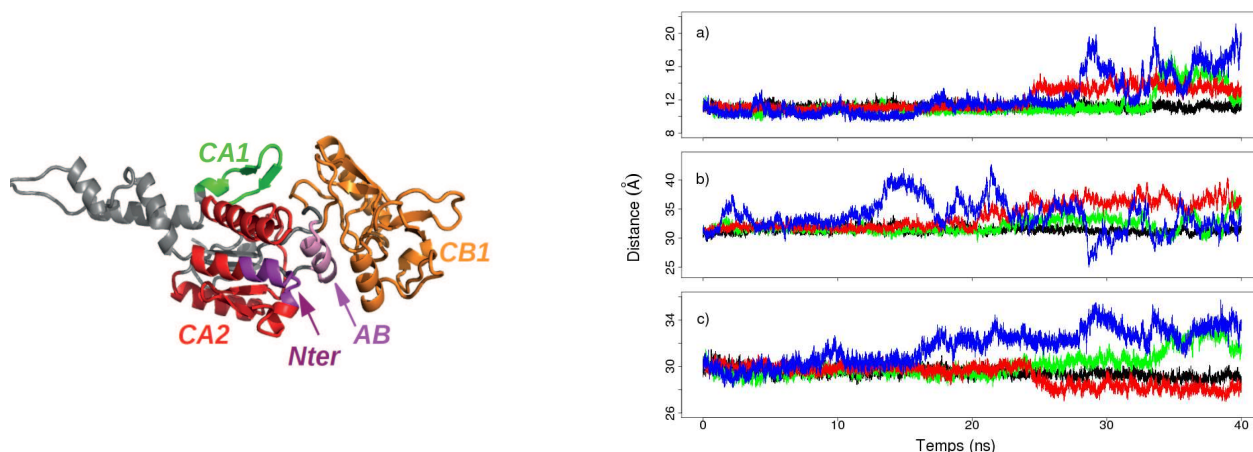


FIG. VI.5: *Dérives conformationnelles (Å) calculées sur les atomes  $C\alpha$  des différentes régions du domaine AC des trajectoires de TAMD réalisées avec quatre centres de masse comme variables collectives. Les valeurs de  $\beta^{-1}$  sont de (a) 4, (b) 6 et (c) 20 kcal.mol<sup>-1</sup>. Les courbes sont colorées en rose (SA), vert (CA), orange (CB) et noire (ensemble de la protéine).*

ments conformationnels au cours des trajectoires de TAMD (Figure VI.6). Quelques sous-régions ont été détectées, qui affichent des dérives conformationnelles pour la plupart inférieures à 1.5 Å : la région N-terminale (résidus 16-26), la région CA1 (résidus 256-272), CA2 (résidus 7-35, 273-341) et CB1 (résidus 58-186) et l'hélice  $\alpha$  AB (résidus 44-54) Figure VI.6a. Certaines distances entre ces régions varient différemment entre la trajectoire de DM classique et les trajectoires de TAMD. L'hélice  $\alpha$  AB bouge par rapport à la région N-terminal (Figure VI.6a). Les régions CB1 et CA1 bougent l'une par rapport à l'autre (Figure VI.6b). Les régions CB1 et CA2 se rapprochent l'une de l'autre dans TAMD-4COM-06, et s'éloignent dans TAMD-4COM-20 (Figure VI.6c). Comme l'hélice  $\alpha$  AB est située sous le site catalytique, son mouvement est donc un signe d'une déstabilisation majeure du site, ce qui est en accord avec la perte d'activité de AC en l'absence de CaM.

### 3.3 Variation de la structure du solvant

Comme décrit précédemment, au cours de la trajectoire de TAMD-4COM-20, le domaine AC affiche une grande dérive conformationnelle, où la région SA montre d'importantes variations dans



**FIG. VI.6: Variation de l'architecture protéique dans les régions CA et CB.** Les distances sont calculées entre les centres de masse des atomes carbone  $\alpha$ . Les courbes représentées sont colorées en noir (DM classique), en bleu (TAMD-4COM-04), rouge (TAMD-4COM-06) et vert (TAMD-4COM-20). a) Distance entre l'hélice  $\alpha$  AB (résidus 44-54) et la région N-terminale (résidus 16-26). b) Distance entre les régions CB1 (résidus 58-186) et CA1 (résidus 256-272). c) Distance entre les régions CB1 (résidus 58-186) et CA2 (résidus 7-35, 273-341). La structure du domaine AC est représentée avec les régions CB1 colorées en orange, l'hélice  $\alpha$  AB colorée en rose, la région CA1 colorée en vert, la région CA2 colorée en rouge et la région N-terminale colorée en violet.

son orientation par rapport à CA. Ces orientations oscillent entre des positions de SA proches de CA, et d'autres positions de SA plus éloignées (Figure VI.4d). Il semblerait donc que la conformation de la protéine explore une région d'entropie conformationnelle grande et qu'elle ne soit pas spécialement attirée par un bassin d'énergie plus faible. Dans le but de prédire des conformations stables adoptées par AC isolée, il faudrait tout d'abord stabiliser la conformation de la protéine autour d'un minimum énergétique local.

Afin de sélectionner les conformations échantillonnées durant TAMD-4COM-20 correspondant à un bassin d'énergie de minimum local, trois conformations ont été extraites à 36, 39 et 40 ns. Elles correspondent aux extrêmes de la dérive conformationnelle calculée (Figure VI.3b). Des simulations de dynamique moléculaire classique MD-CONT-36, MD-CONT-39 et MD-CONT-40 ont été réalisées à partir de ces conformations. Celles-ci montrent une dérive assez rapide vers une forme complètement allongée du domaine AC (Figure VI.7). Cette conformation étendue est assez proche de celle observée dans le complexe avec la calmoduline.

Afin d'examiner l'entropie conformationnelle des configurations enregistrées pendant la simulation TAMD-4COM-20, les molécules d'eau formant au moins deux liaisons hydrogène avec les atomes de la protéine, aussi bien des groupes donneur ou accepteur, ont été analysées. On appelle ce type de liaison des "ponts d'eau". Le nombre de ponts d'eau formés a été calculé sur une conformation sur 10, le long des trajectoires TAMD et comparé au nombre de ponts observés au cours de la simulation de DM classique (Figure VI.8). Dans la trajectoire de DM, le nombre de ponts d'eau est plus ou moins constant et oscille autour de 110, tandis qu'après 20 ns de TAMD-4COM-20 et TAMD-3COM-20 (Figure VI.8, courbes vertes) le nombre de ponts d'eau diminue. Cette diminution est la plus importante pour TAMD-4COM-20 où le nombre de ponts d'eau chute jusqu'à environ 30. Il semblerait que le fait d'appliquer un thermostat d'une tempé-

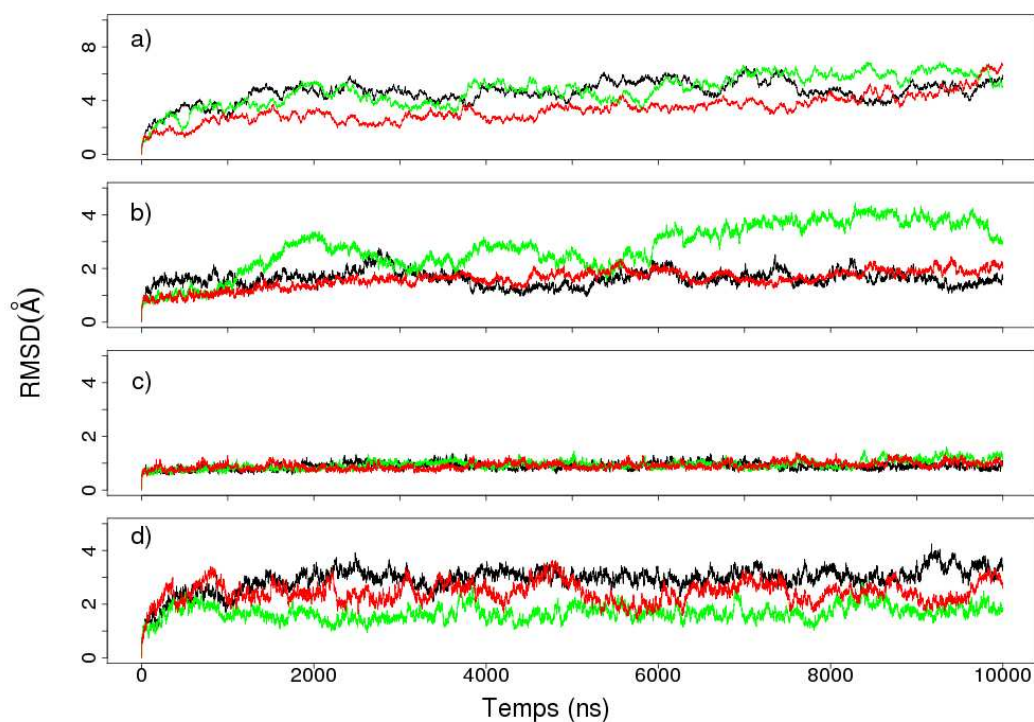


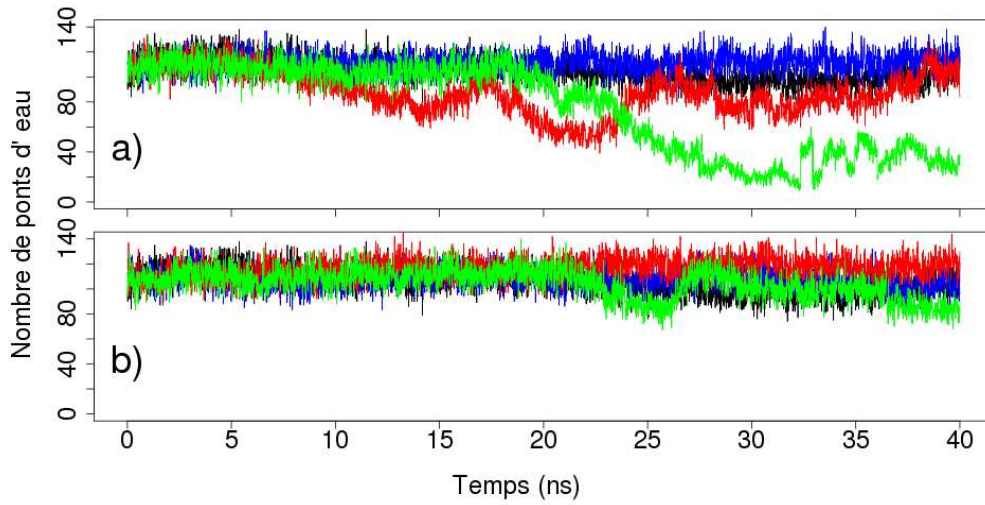
FIG. VI.7: Dérive conformationnelle (Å) du domaine AC au cours d'une DM non-restrainte à partir des conformations extraites à 36, 39 et 40 ns de la trajectoire de TAMD produite à  $\beta^{-1}$  de 20 kcal.mol<sup>-1</sup>. Les courbes sont colorées en noire (36 ns), vert (40 ns) et rouge (39 ns). Le RMSD est calculé en superposant les conformations sur : (a) toute la structure, (b) la région CA, (c) la région CB et (d) la région SA. La structure de référence est la première conformation de la trajectoire.

rature plus élevée durant un certain temps aux centres de masse induise une baisse dramatique de la viscosité du solvant. Ceci peut être induit par la dispersion dans le solvant de l'énergie introduite dans le système par les thermostats de température plus élevée. Dans TAMD-3COM-04 et TAMD-4COM-04, aucune transition n'a été observée (Figure VI.8 courbes bleues).

La disparition de ponts d'eau expliquerait le retour à un état plus allongé de AC, qui est observé dans les trajectoires de DM classique lancées à partir de points extrêmes de la TAMD-4COM-20. En effet, les ponts d'eau pourraient servir d'intermédiaire pour créer des interactions à longue portée entre des résidus situés dans les régions SA et CA, et donc induire un repliement stable de SA sur CA.

Les positions des ponts d'eau dans les structures ont été analysées en calculant des cartes de contacts (Figure VI.9) sur lesquelles un point est dessiné à la position  $(i, j)$  si un pont d'eau est observé entre les résidus  $i$  et  $j$ . La population cumulée des ponts d'eau est représentée à partir des intervalles 29-30 ns de la DM classique (Figure VI.9a), et des intervalles 9-10 ns (Figure VI.9b), 29-30 ns (Figure VI.9c) et 39-40 ns (Figure VI.9d) de la trajectoire TAMD-4COM-20. L'intervalle 9-10 ns de TAMD-4COM-20 (Figure VI.9b) affiche une carte de contact similaire à celle de l'intervalle 29-30 ns de la DM classique (Figure VI.9a).

L'intervalle 29-30 ns de TAMD-4COM-20 (Figure VI.9c) contient le plus petit nombre de



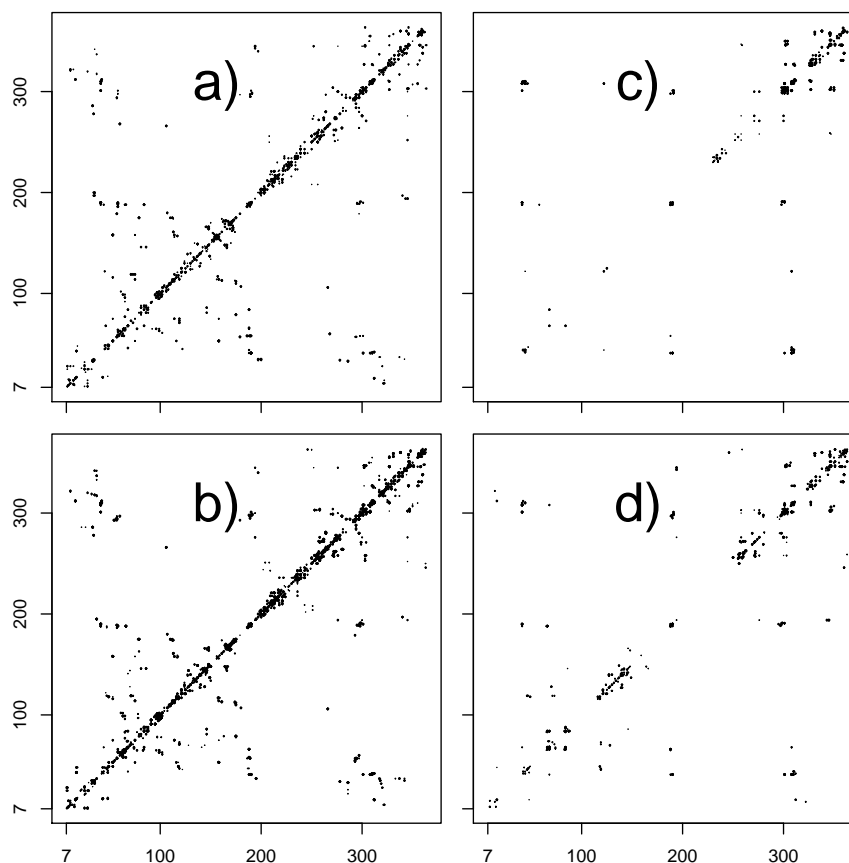
**FIG. VI.8:** *Nombre de ponts d'eau détectés au cours des trajectoires* de (a) DM (courbe noire), TAMd-4COM-04 (courbe bleue), TAMd-4COM-06 (courbe rouge) et TAMd-4COM-20 (courbe verte), (b) DM (courbe noire), TAMd-3COM-04 (courbe bleue), TAMd-3COM-06 (courbe rouge) et TAMd-3COM-20 (courbe verte).

ponts d'eau (Figure VI.8a, courbe verte). La carte de contact est presque vide, à l'exception de quelques molécules d'eau qui connectent les résidus 299, 301, 302, 304, 308, 309, 310 de la boucle catalytique, et 41/44, 327 situés d'un côté et de l'autre de la boucle catalytique. Ces molécules d'eau sont les médiateurs des interactions à longue portée entre les résidus lointains dans la séquence, et leur présence est certainement un obstacle à la réaction enzymatique.

Dans l'intervalle 39-40 ns de TAMd-4COM-20 (Figure VI.9d), un plus grand nombre de ponts d'eau apparaît que dans l'intervalle 29-30 ns, mais le système n'est pas revenu à l'état observé durant l'intervalle 29-30 ns de la DM classique. D'autre part, les ponts d'eau impliqués dans les interactions à longue portée sont toujours présents entre la boucle catalytique et les autres régions de la protéine, il est probable que ces ponts d'eau jouent un rôle important au niveau des interactions impliquant la boucle catalytique.

### 3.4 Prédiction de conformations de AC stables dans le bassin d'énergie de AC isolé

Comme on l'a vu dans la section précédente, les conformations extraites de la trajectoire TAMd-4COM-20 n'étaient pas stables au cours des simulations de DM classique. Afin de stabiliser ces conformations, nous proposons une approche heuristique dans laquelle les conformations de AC extraites de cette trajectoire sont minimisées pour rapprocher les régions SA et CA. Une conformation obtenue à la 36<sup>ème</sup> ns de cette trajectoire a été extraite, et soumise à une série de minimisation où des contraintes de distance ont été ajoutées étape par étape entre les résidus en fonction de leurs distances. Des minimisations de 100000 pas ont été réalisées avec NAMD avec la commande "extraBonds".



**FIG. VI.9: Cartes de contacts décrivant les ponts d'eau.** Chaque point sur la carte de contact aux coordonnées  $i$  et  $j$  correspond à l'observation d'une molécule d'eau formant un pont avec les atomes des résidus  $i$  et  $j$ ; (a) Carte de contact obtenue à partir de la somme des ponts d'eau observés dans l'intervalle 29-30 ns de la trajectoire de DM. (b-d) Cartes de contact obtenues à partir de la somme des ponts d'eau observés dans les intervalles de temps 9-10 ns (b), 29-30 ns (c) et 39-40 ns (d) de la trajectoire TAMD-4COM-20. Les régions sont encadrées de la manière suivante : CA en vert, CB en orange, SA en violet, la région C-terminale en cyan, la boucle catalytique en jaune.

Chaque contrainte de distance est appliquée avec un intervalle de distance 1.2-2 Å et une constante de force de 30 kcal/(mol.Å<sup>2</sup>). Le but de la minimisation avec ces contraintes de distance est de contraindre étape par étape le rapprochement des régions SA et CA. Les contraintes sont choisies en détectant deux ponts d'eau, puis des liaisons hydrogène formées entre les groupes donneur et accepteur de ces deux régions.

Le premier cycle de minimisation a été réalisé avec deux contraintes appliquées entre ASN-200-Oδ1 et ASN-35-Hδ22 et entre LYS-34-O et ASN-200-Hδ21. Ces paires d'atomes ont été sélectionnées car deux ponts d'eau étaient établis entre ces atomes dans 7.8 et 6.2 % de la trajectoire TAMD-4COM-20. Ces ponts d'eau sont les plus stables comparés aux autres ponts formés entre SA et CA (Figure VI.10a).

Les liaisons hydrogène formées au cours du premier cycle de minimisation, entre les résidus éloignés d'au moins 5 résidus dans la séquence ont été analysées. Parmi les 80 liaisons hydrogène formées entre les régions SA, CA et C-terminal, 21 impliquent les résidus ALA-257, ARG-258,



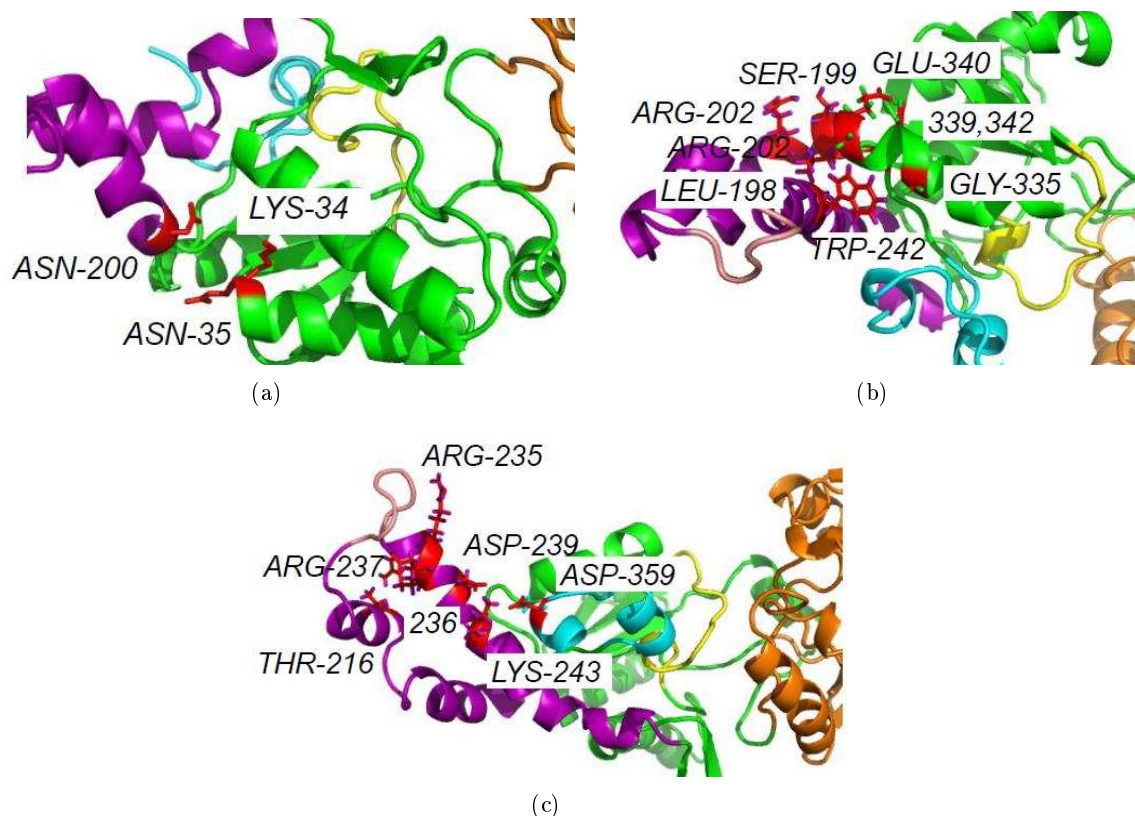


FIG. VI.10: *Position des résidus impliqués dans les jeux de contraintes pour les premiers (a), second (b) et troisième (c) cycle de minimisations.*

ARG-259, GLN-260, THR-273, PHE-275 et GLU-276, situés dans la même région que les résidus ASN-200, LYS-34 et ASN-35 pour lesquels des contraintes de distances ont déjà été appliquées. On constate également la présence de 8 liaisons hydrogène, qui impliquent les résidus ASN-200, LYS-34 et ASN-35, qui étaient déjà contraints dans le cycle de minimisation précédent. Ces 29 liaisons hydrogène n'ont donc pas été prises en compte comme contraintes additionnelles, afin d'explorer les contraintes qui ne sont pas redondantes avec celles déjà appliquées.

D'autre part, 29 liaisons hydrogène impliquent les résidus ASP-359, ASP-360, GLY-361 et GLY-363, situés à l'extrémité C-terminale du domaine AC. Cette partie C-terminal étant assez flexible, ces liaisons hydrogène n'ont pas été considérées comme de bons candidats pour induire une forme plus compacte de AC et stable au cours d'une trajectoire de dynamique moléculaire non-contrainte.

Les 10 liaisons hydrogène qui restaient après avoir retiré celles précédemment décrites, impliquent les résidus LEU-198, SER-199, ARG-202, TRP-242, GLY-335, ARG-338, GLY-339, GLU-340, GLY-341, TYR-342. Parmi elles, trois liaisons hydrogène, ARG-338-O/TRP-242-HE1, GLY-341-O/LEU-198-HN et GLY-341-O/SER-199-HN n'ont pas été prises en compte, car elles sont redondantes avec les autres liaisons hydrogène, et afin de réduire le nombre de listes de contraintes possibles. Comme les 7 liaisons hydrogène restantes ne sont pas compatibles entre elles, elles ont été séparées en 8 listes de contraintes (Table III et Figure VI.10b). Elles ont été rajoutées aux

deux contraintes utilisées au cours du premier cycle, et huit minimisations indépendantes ont été réalisées en utilisant ces jeux de contraintes.

Atome donneur	Atome accepteur	Présence pendant la minimisation (%)
TYR 342 O	LEU 198 HN	3.0
GLY 335 O	TRP 242 H $\epsilon$ 1	58.0
GLU 340 O	ARG 202 HH12	84.0
TYR 342 O	LEU 198 HN	3.0
GLY 335 O	TRP 242 H $\epsilon$ 1	58.0
GLU 340 O	SER 199 HN	65.0
TYR 342 O	LEU 198 HN	3.0
GLY 335 O	TRP 242 H $\epsilon$ 1	58.0
GLU 340 O	SER 199 H $\gamma$ 1	87.0
TYR 342 O	LEU 198 HN	3.0
GLY 335 O	TRP 242 H $\epsilon$ 1	58.0
GLU 340 O	ARG 202 HH11	67.0
TYR 342 O	LEU 198 HN	3.0
GLY 339 O	TRP 242 H $\epsilon$ 1	85.0
GLU 340 O	ARG 202 HH12	84.0
TYR 342 O	LEU 198 HN	3.0
GLY 339 O	TRP 242 H $\epsilon$ 1	85.0
GLU 340 O	SER 199 HN	65.0
TYR 342 O	LEU 198 HN	3.0
GLY 339 O	TRP 242 H $\epsilon$ 1	85.0
GLU 340 O	SER 199 H $\gamma$ 1	87.0
TYR 342 O	LEU 198 HN	3.0
GLY 339 O	TRP 242 H $\epsilon$ 1	85.0
GLU 340 O	ARG 202 HH11	67.0

TAB. III: Liste des contraintes ajoutées pour le 2<sup>me</sup> cycle de minimisation.

Durant le troisième cycle de minimisation, six contraintes provenant de liaisons hydrogène formées au cours du second cycle de minimisation, ont été ajoutées pour chacun des huit jeux de contraintes utilisés au cours du deuxième cycle (Table IV).

Atome donneur	Atome accepteur	Présence pendant la minimisation (%)
THR 216 O $\gamma$ 1	ARG 237 HH12	100.0
GLU 236 O $\epsilon$ 1	ARG 232 HH11	100.0
GLU 236 O $\epsilon$ 2	ARG 223 HH21	100.0
ASP 359 O $\delta$ 2	LYS 243 H $\zeta$ 1	95.0
ASP 239 O $\delta$ 1	LYS 242 H $\zeta$ 2	-
ARG 235 HH21	GLU 332 O $\epsilon$ 2	-

TAB. IV: Liste des contraintes ajoutées pour le 3<sup>me</sup> cycle de minimisation.

Trois contraintes impliquent THR-216, GLU-236, ARG-237 et ARG-223, des résidus situés dans les hélices  $\alpha$  G et H de la région SA (Figure VI.10c). Comme ces liaisons hydrogène étaient déjà formées à 100% pendant le deuxième cycle de minimisation, leur implication tend à renforcer



le repliement existant de SA. Une contrainte supplémentaire a été appliquée entre ASP-359 et LYS-243 afin d'améliorer le repliement de la queue C-terminale. Deux contraintes ont été appliquées entre ASP-239/LYS-243 et ARG-235/GLU-332 afin de rapprocher l'extrémité SA de la région CA.

Les seize conformations finales minimisées obtenues après le deuxième et troisième cycle de minimisation sont illustrées dans la figure VI.11. La liste complète des contraintes est décrite dans la Table V. Pendant le troisième cycle de minimisation, une pliure apparaît dans l'hélice  $\alpha$  H, au niveau du résidu TRP-242 (Figure VI.11). Or, il a été démontré par spectrométrie de fluorescence [262] que le résidu TRP-242 a une grande mobilité interne. En solution, la pliure de l' $\alpha$ -hélice H pourrait être variable au cours du temps, ce qui serait compatible avec une mobilité interne du résidu TRP-242.

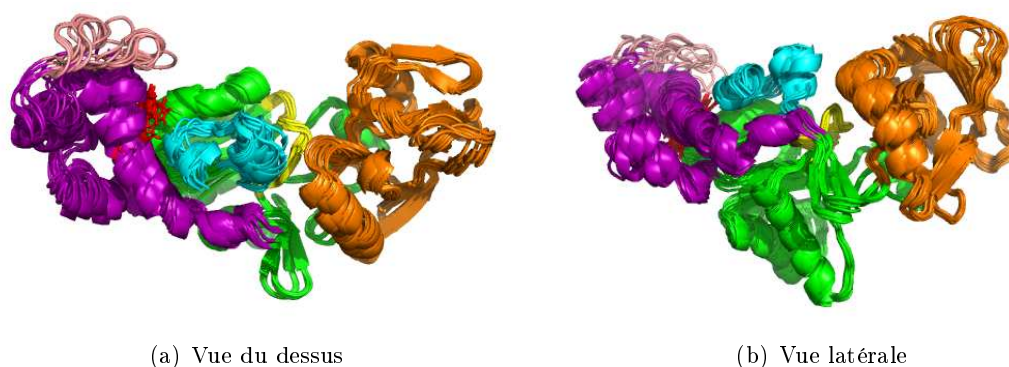


FIG. VI.11: *Conformations minimisées de AC obtenues après le deuxième et troisième cycle de minimisation.* Les conformations sont superposées sur la région CA. Trp-242 est représenté en baton rouge.

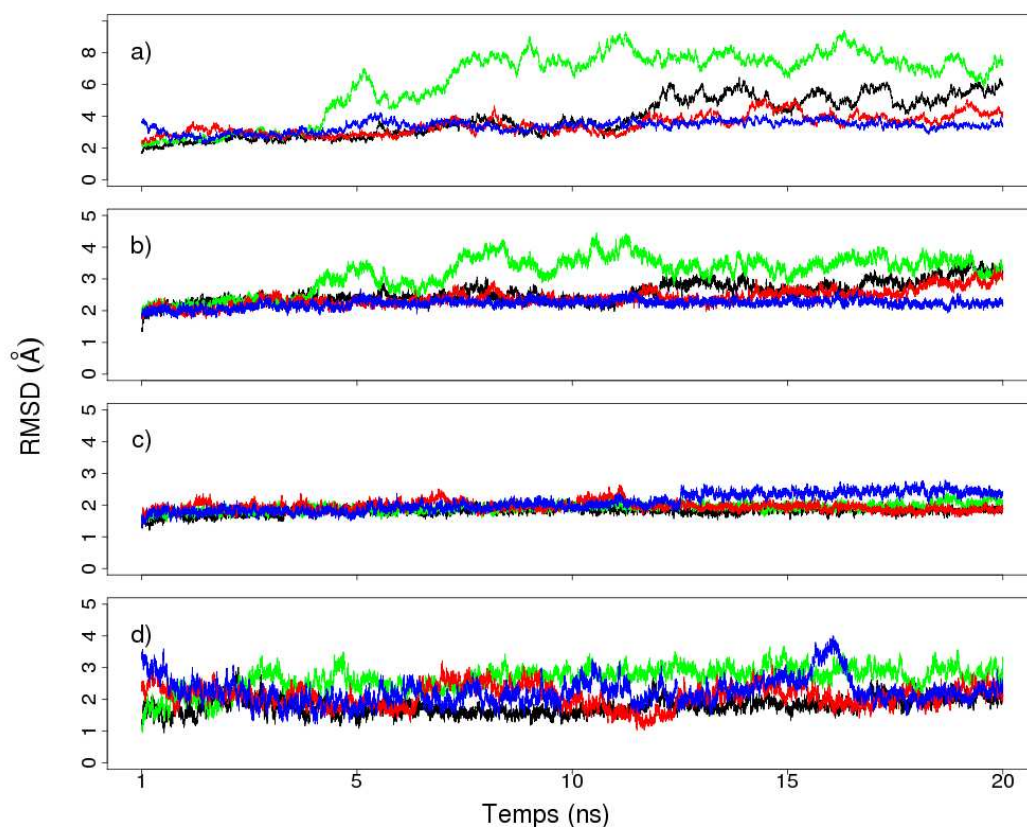
La qualité des 16 conformations obtenues au cours de la seconde et de la dernière minimisation a été vérifiée avec le logiciel Molprobit [263]. Les conformations 25, 28, 32 et 34 sont celles qui ont le moins de résidus situés à l'extérieur des régions permises du diagramme de Ramachandran. Les conformations 25, 28, 32 et 34 ont été sélectionnées et comprenaient entre 6 et 9 résidus avec des angles  $\Phi$  et  $\Psi$  aberrants. Elles ont donc été sélectionnées pour tester leur stabilité avec des simulations de 20 ns de Dynamique Moléculaire non-contraintes : MD-MIN-25, MD-MIN-28, MD-MIN-32 et MD-MIN-34.

Les dérivées liées à ces conformations ont été calculées par rapport à la conformation minimisée 25 (Figure VI.12). Sur les quatre conformations, deux restent proches de la conformation minimisée 25, avec un RMSD des coordonnées inférieur à 4 Å tout au long de la trajectoire de 20 ns. Nous avons donc réussi à obtenir des conformations compactes de AC qui sont relativement stables et qui donnent une description du bassin d'énergie occupé par la protéine AC isolée en solution.

Les ponts d'eau et les liaisons hydrogène formés entre SA et CA dans MD-CONT-36, MD-CONT-39 et MD-CONT-40 (Table VI) sont moins nombreux que ceux formés dans MD-MIN-25, MD-MIN-28, MD-MIN-32 et MD-MIN-34 (Table VII). Par ailleurs, les pourcentages de présence des ponts d'eau sont aussi inférieures pour MD-CONT-36, MD-CONT-39 et MD-CONT-40. Cette observation est en faveur d'une plus grande stabilité des conformations dans MD-MIN-25, MD-

MIN-25			
Atome donneur	Atome accepteur	Origine de la contrainte	% Présence
ASN 200 O $\delta$ 11	ASN 35 H $\delta$ 22	pont	7.8
LYS 34 O	ASN 200 H $\delta$ 21	pont	6.2
TYR 342 O	LEU 198 HN	liaison hydrogène	3.0
GLY 339 O	TRP 242 H $\epsilon$ 1	liaison hydrogène	85.0
GLU 340 O	ARG 202 HH12	liaison hydrogène	84.0
MIN-28			
Atome donneur	Atome accepteur	Origine de la contrainte	% Présence
ASN 200 O $\delta$ 11	ASN 35 H $\delta$ 22	pont	7.8
LYS 34 O	ASN 200 H $\delta$ 21	pont	6.2
TYR 342 O	LEU 198 HN	liaison hydrogène	3.0
GLY 339 O	TRP 242 H $\epsilon$ 1	liaison hydrogène	85.0
GLU 340 O	ARG 202 HH11	liaison hydrogène	67.0
MIN-32			
Atome donneur	Atome accepteur	Origine de la contrainte	% Présence
ASN 200 O $\delta$ 11	ASN 35 H $\delta$ 22	pont	7.8
LYS 34 O	ASN 200 H $\delta$ 21	pont	6.2
TYR 342 O	LEU 198 HN	liaison hydrogène	3.0
GLY 335 O	TRP 242 H $\epsilon$ 1	liaison hydrogène	58.0
GLU 340 O	SER 199 HN	liaison hydrogène	65.0
THR 216 O $\gamma$ 1	ARG 237 HH12	liaison hydrogène	100.0
GLU 236 O $\epsilon$ 1	ARG 223 HH11	liaison hydrogène	100.0
GLU 236 O $\epsilon$ 2	ARG 223 HH21	liaison hydrogène	100.0
ASP 359 O $\delta$ 2	LYS 243 H $\zeta$ 1	liaison hydrogène	100.0
ASP 239 O $\delta$ 1	LYS 243 H $\zeta$ 2	-	-
ARG 235 HH21	GLU 332 O $\epsilon$ 2	-	-
MIN-34			
Atome donneur	Atome accepteur	Origine de la contrainte	% Présence
ASN 200 O $\delta$ 11	ASN 35 H $\delta$ 22	pont	7.8
LYS 34 O	ASN 200 H $\delta$ 21	pont	6.2
TYR 342 O	LEU 198 HN	liaison hydrogène	3.0
GLY 335 O	TRP 242 H $\epsilon$ 1	liaison hydrogène	58.0
GLU 340 O	ARG 202 HH11	liaison hydrogène	67.0
THR 216 O $\gamma$ 1	ARG 237 HH12	liaison hydrogène	100.0
GLU 236 O $\epsilon$ 1	ARG 223 HH11	liaison hydrogène	100.0
GLU 236 O $\epsilon$ 2	ARG 223 HH21	liaison hydrogène	100.0
ASP 359 O $\delta$ 2	LYS 243 H $\zeta$ 1	liaison hydrogène	100.0
ASP 239 O $\delta$ 1	LYS 243 H $\zeta$ 2	-	-
ARG 235 HH21	GLU 332 O $\epsilon$ 2	-	-

TAB. V: Liste totale des contraintes utilisées pour la minimisation des conformations de AC 25, 28, 32 et 34.



**FIG. VI.12:** *Dérive conformationnelle (Å) des trajectoires de 20 ns de Dynamique Moléculaire non-restrictes, produites à partir des quatre conformations minimisées du domaine AC.* Les courbes sont colorées en noire (25), vert (28), rouge (32) et bleue (34). Le RMSD est calculé en superposant les conformations sur : (a) la structure entière, (b) la région CA, (c) la région CB et (d) la région SA. La structure de référence est la conformation minimisée 25.

MIN-28, MD-MIN-32 et MD-MIN-34. Par ailleurs, les liaisons hydrogène observées dans MD-MIN-25, MD-MIN-28, MD-MIN-32 et MD-MIN-34 ne correspondent pas exactement aux contraintes appliquées durant la minimisation, ce qui est le signe d'une certaine réorganisation structurale.

<b>CONT-36</b>					
Ponts d'eau		%	Liaisons hydrogènes		%
ALA-252-O	THR-273-HG1	31.6	GLU-276-OE2	ARG-258-HH11	86.3
VAL-271-O	ARG-258-HH12	30.4	GLU-276-OE2	ARG-258-HH21	88.3
GLU-276-OE1	ARG-258-HH21	35.0	GLU-276-OE1	ARG-258-HH11	88.5
ILE-194-O	ARG-258-HH21	55.6	GLU-276-OE1	ARG-258-HH21	91.3
			GLY-341-O	LEU-198-HN	100.0
<b>CONT-39</b>					
Ponts d'eau		%	Liaisons hydrogènes		%
			GLU-276-OE1	ARG-258-HH11	76.1
			GLU-276-OE2	ARG-258-HH21	90.8
			GLU-276-OE2	ARG-258-HH11	97.2
			GLY-341-O	LEU-198-HN	98.8
			GLU-276-OE1	ARG-258-HH21	99.7
<b>CONT-40</b>					
Ponts d'eau		%	Liaisons hydrogènes		%
ALA-252-O	THR-272-HG1	54.9	GLU-276-OE1	ARG-258-HH11	55.5
VAL-271-O	ARG-258-HH12	35.9	GLU-276-OE2	ARG-258-HH21	93.1
			GLU-276-OE1	ARG-258-HH21	97.5
			GLU-276-OE2	ARG-258-HH11	98.5
			GLY-341-O	LEU-198-HN	99.9

**TAB. VI:** Présence de ponts d'eau et de liaisons hydrogène entre SA et CA/Cter pendant les trajectoires MD-CONT-36, MD-CONT-39 et MD-CONT-40

MIN-25					
Ponts d'eau		%	Liaisons hydrogènes		%
ALA-257-O	ILE-272-HN	74.4	GLU-276-OE1	ARG-258-HH21	53.4
GLN-260-O	ILE-272-HN	76.7	GLU-276-OE2	ARG-258-HH11	89.2
THR-273-OG1	ARG-258-HH22	33.6	GLU-276-OE1	ARG-258-HH11	99.1
GLU-276-OE1	ARG-258-HH21	57.9	GLY-341-O	LEU-198-HN	99.5
ILE-194-O	ARG-258-HH21	68.6	ASP-360-OD1	ARG-246-HE	30.2
ARG-258-O	ARG-348-HN	44.3	ASP-360-OD1	ARG-246-HH22	31.6
GLY-361-O	ARG-259-HH21	34.5	ASP-360-OD2	ARG-246-HH22	35.1
ASP-360-OD2	ARG-246-HH22	35.1	ASP-360-OD2	ARG-246-HE	48.7
ASP-354-OD1	ARG-240-HH22	37.2			
MIN-28					
Ponts d'eau		%	Liaisons hydrogènes		%
ALA-257-O	ILE-272-HN	57.9	GLU-276-OE1	ARG-258-HH21	57.0
GLN-260-O	ILE-272-HN	64.5	GLU-276-OE2	ARG-258-HH21	59.5
GLU-276-OE1	ARG-258-HH21	30.8	GLU-276-OE2	ARG-258-HH11	63.2
TYR-345-OH	ARG-258-HH21	37.2	GLU-276-OE1	ARG-258-HH11	73.4
GLU-276-OE2	ARG-258-HH21	37.7	GLY-341-O	LEU-198-HN	100.0
ARG-258-O	ARG-348-HN	32.5	ASP-360-OD2	ARG-246-HE	33.6
			ASP-360-OD1	ARG-246-HE	34.3
			ASP-360-OD2	ARG-246-HH22	34.9
			ASP-360-OD1	ARG-246-HH22	45.0
MIN-32					
Ponts d'eau		%	Liaisons hydrogènes		%
SER-251-O	PHE-275-HN	40.8	GLU-276-OE1	ARG-258-HH21	33.4
ALA-257-O	ILE-272-HN	46.3	GLU-276-OE2	ARG-258-HH21	61.4
GLN-260-O	ILE-272-HN	48.8	GLU-276-OE1	ARG-258-HH11	64.8
GLY-335-O	TRP-242-HE1	43.1	GLU-276-OE2	ARG-258-HH11	73.5
ARG-258-O	ASN-347-HD22	36.5	GLY-341-O	LEU-198-HN	99.1
ASP-359-OD1	ARG-246-HH11	39.8	ASP-360-OD2	ARG-246-HE	42.9
ASP-359-OD2	ARG-246-HH11	40.3	ASP-360-OD1	ARG-246-HH22	50.4
			ASP-360-OD2	ARG-246-HH22	58.2
			ASP-359-OD2	ARG-246-HH11	70.9
			ASP-359-OD1	ARG-246-HH11	71.7
			ASP-359-OD1	ARG-246-HH21	80.2
			ASP-359-OD2	ARG-246-HH21	83.5
MIN-34					
Ponts d'eau		%	Liaisons hydrogènes		%
ALA-245-O	ARG-338-HE	30.0	GLY-335-O	TRP-242-HE1	30.5
ALA-257-O	ILE-272-HN	73.2	GLU-276-OE2	ARG-258-HH11	35.6
GLN-260-O	ILE-272-HN	79.2	GLU-340-O	ARG-202-HH12	52.3
GLY-335-O	TRP-242-HE1	40.4	GLU-276-OE1	ARG-258-HH21	60.6
GLU-276-OE1	ARG-258-HH21	65.9	GLU-276-OE1	ARG-258-HH11	97.0
ILE-194-O	ARG-258-HH21	85.8	GLY-341-O	LEU-192-HN	99.9
ARG-258-O	ARG-348-HN	45.2	ASP-360-OD1	ARG-246-HE	39.9
ASP-360-OD2	ARG-246-HH22	34.4	ASP-360-OD2	ARG-246-HH22	76.0
			ASP-360-OD2	ARG-246-HE	79.7

**Tab. VII:** Présence de ponts d'eau et de liaisons hydrogène entre SA et CA/Cter pendant les trajectoires MD-MIN-25, MD-MIN-28, MD-MIN-32 et MD-MIN-34.

## 4 Discussion

Le but de ce chapitre consistait à prédire des conformations du domaine AC isolé de l'adénylcyclase de *Bordetella pertussis*. Des jeux de trois ou quatre centres de masse ont été définis comme variables collectives pour l'application de la TAMD. Quelques conformations extrêmes échantillonnées pendant les trajectoires de TAMD ont été analysées de manière plus approfondies afin d'obtenir un jeu de conformations décrivant le bassin d'énergie de minimum local correspondant à l'état non-lié de la protéine.

Avant toute chose, cette étude démontre la puissance de la TAMD dans l'exploration de l'espace conformationnel d'une protéine. En effet, la comparaison entre des conformations représentatives extraites de la trajectoires de DM classique et celles de la trajectoire de TAMD, révèle un échange conformationnel plus large avec la TAMD. Cette amélioration est d'autant plus remarquable du fait que l'on a utilisé la même échelle de temps de simulation pour tous les types de trajectoire.

Les variables collectives ont été définies à partir d'une information procurée par l'état lié. Néanmoins, couplées à la TAMD, elles ont prouvé une efficacité suffisante dans l'amélioration de l'échantillonnage conformationnel de la protéine, sortant du bassin énergétique de minimum local visité par la DM classique. Mais elles n'ont pas permis d'obtenir de transition vers un bassin énergétique correspondant à un état non-lié de AC. D'autres variables plus appropriées, comme par exemple des variables suivant la formation des liaisons hydrogène entre SA et CA, devraient être testées à l'avenir.

Les tendances conformationnelles de AC observées au cours des trajectoires de TAMD donnent des informations sur la façon dont la protéine est inactivée. En effet, le comportement du corps rigide de CB et sa rotation qui l'éloigne par rapport à CA, tranchent avec la grande dérive conformationnelle au sein de CA. Ceci confirme les observations faites précédemment sur les trajectoires de DM classique de AC [252] où l'interface entre CA et C-CaM concentrait les variations associées aux changements de l'activité catalytique de AC.

L'analyse approfondie des trajectoires TAMD soutient l'interférence du comportement du solvant avec les tendances conformationnelles de la protéine. Premièrement, des ponts d'eau stables apparaissent entre la boucle catalytique et d'autres régions de la protéine, perturbant ainsi l'activité catalytique. Par ailleurs, la forte décroissance du nombre de ponts d'eau est associée avec le maximum d'échantillonnage conformationnel de la protéine. Il n'est pas évident de savoir si cette observation est un artéfact de la TAMD ou bien si ce phénomène peut avoir lieu réellement lors de la transition. Néanmoins, cette variation dans le comportement de l'eau joue un rôle dans l'accélération de la transition *in silico*, et la situation pourrait être similaire dans le cas expérimental. En effet, la variation du nombre de ponts d'eau pourrait influencer les couches de solvation plus éloignées que la première couche classique d'hydratation située à 3 Å, en accord avec les observations expérimentales faites par spectroscopie THz [264].

Le type de transition exploré ici, une transition entre un état lié bien défini et un état non-lié plus dynamique est assez souvent rencontré en biologie. Cette étude est applicable au comportement *in vivo* des biomolécules, et ouvre la voie pour d'autres études ultérieures. Dans ce cas, le

manque de précision dans la définition de l'état non-lié rend le choix des variables collectives plus difficile, tout comme la définition de la région cible de la transition conformationnelle.

Le point critique de cette étude est, après avoir été capable de sortir du bassin énergétique associé à l'état lié, d'atteindre un bassin énergétique de l'état non-lié. L'information expérimentale disponible sur l'état non-lié de AC soutient un modèle où l'échange conformationnel de la protéine est limité à des variations mineures autour d'une forme générale plus compacte. En suivant la description hiérarchique largement acceptée des espaces conformationnels des protéines [265; 266], la région cible devrait donc être un minimum énergétique local, défini par une famille de conformations de la protéine. La plupart de ces conformations devrait rester dans ce bassin durant les simulations de dynamique moléculaire classique afin d'être en accord avec la forme compacte homogène observée par l'ultracentrifugation.

En suivant ces directives, nous avons proposé une méthode heuristique pour renforcer l'entrée de la protéine dans le bassin de l'état non-lié. La fiabilité des conformations obtenues a été vérifiée en calculant leur stabilité structurelle au cours de simulations de DM classiques. La variabilité conformationnelle attendue à l'état non-lié hautement dynamique, a été analysée par l'exploration de toutes les possibilités de contraintes dans la formation de liaisons hydrogène. Les réarrangements conformationnels observés au cours des simulations de DM non-restreintes à partir de conformations compactes minimisées, s'accordent avec un bassin d'énergie avec différents minima locaux d'énergies similaires.

La méthode de la TAMD, appliquée sur le système du domaine catalytique isolé de l'adénylcyclase de *Bordetella pertussis* a donné une description qualitative des énergétiques de l'état non-lié de la protéine. Les conformations compactes de la protéine, correspondant aux minima énergétiques locaux du bassin énergétique de l'état non-lié ont été prédites. A l'avenir, les liaisons hydrogène qui induisent les prédictions de ces conformations pourraient être analysées en : (i) définissant d'autres variables collectives qui correspondraient à la formation de ces liaisons hydrogène et en observant la compaction de la conformation au cours des trajectoires de TAMD correspondantes, (ii) en réalisant des expériences de "cross-linking" pour vérifier les proximités suffisantes des résidus reliés par des liaisons hydrogène.

## 5 Conclusion

Le domaine catalytique AC de l'adénylcyclase CyaA de *Bordetella pertussis* décrit une transition d'un état allongé lié à la calmoduline à un état non-lié, qui a été expérimentalement caractérisé comme étant de forme compacte, beaucoup moins étendue, tout en conservant le même degré de structure secondaire. Les dynamiques accélérées par la température ont généré des trajectoires à partir de l'état lié, qui explorent d'avantage l'espace conformationnel que les trajectoires de DM classique. Néanmoins, les conformations extrêmes observées dans les TAMD n'étaient pas stables dans le bassin énergétique de l'état non-lié. Une méthode heuristique, basée sur une analyse des interactions solvant-protéine et protéine-protéine, a permis de sélectionner une série de conforma-

tions qui ont conduit à une estimation des conformations visitées par la protéine dans le bassin énergétique de l'état non-lié.





## Chapitre VII

# Conclusion générale et perspectives

---

Le sujet de cette thèse s'inscrit dans le cadre des interactions protéine-protéine, du contrôle allostérique et de la prédiction de structure. Ce travail de thèse comporte deux objectifs principaux.

Le premier objectif consistait en l'étude des changements conformationnels du domaine catalytique AC de la protéine CyaA de *Bordetella pertussis* lors de son interaction avec le lobe C-terminal de la calmoduline (C-CaM). Pour cela, AC a été simulé par dynamique moléculaire dans trois conditions différentes d'interaction : (i) à l'état non-lié, (ii) à l'état complexé avec C-CaM, (iii) en complexe avec la C-CaM chargée d'ions  $\text{Ca}^{2+}$ .

L'importance de l'interaction de C-CaM dans la stabilisation du complexe a été mise en évidence par diverses analyses. L'analyse des tendances conformationnelles a montré que AC existe sous différents états à l'équilibre, et que l'interaction avec C-CaM le stabilise dans une conformation étendue, favorable à sa fonction enzymatique. L'analyse des mouvements globaux, qui reflètent la plasticité des systèmes, a montré l'importance de C-CaM dans la stabilisation des régions extrêmes SA et CB. La rupture de cette liaison engendre une tendance vers une compaction de AC, en rapprochant les régions SA et CB l'une vers l'autre.

Les influences énergétiques au sein des complexes décrivent un schéma très simple, où seule la C-CaM exerce une influence sur la région CA de AC, montrant ainsi son importance dans la stabilisation de cette région au sein du complexe. Cette observation nous a conduit à l'identification d'un réseau de liaisons hydrogène entre C-CaM et CA qui se rompt avec la délétion de C-CaM et des ions  $\text{Ca}^{2+}$ . Trois résidus clés ont été retenus. Deux d'entre eux, Arg<sup>338</sup> et Asp<sup>360</sup> interagissent non seulement entre eux, mais également avec un résidu de C-CaM, et le troisième, le résidu Asn<sup>347</sup> avec un résidu de la boucle catalytique. L'interaction entre ces groupes de résidus disparaît en l'absence de C-CaM et est perturbée en l'absence d'ions Calcium. On sait par des mesures biochimiques indépendantes que l'absence d'ions Calcium réduit l'affinité de AC pour CaM, et que AC isolé a une activité catalytique quasiment nulle. Le réseau de liaisons hydrogène identifié donne donc un support structural à ces observations biochimiques sur l'affinité AC/CaM et l'activité de AC et permet de formuler l'hypothèse que ce réseau de résidus transmet les perturbations conformationnelles induites par l'interaction avec CaM jusqu'au site catalytique de AC.

L'un des points clés de cette thèse a été l'opportunité d'avoir pu valider cette hypothèse

par des tests *in vitro* des mutants de ces résidus. En effet, l'unité de Biochimie des Interactions Macromoléculaires (Daniel Ladant, Institut Pasteur) possède une expertise sur CyaA depuis une quinzaine d'années. Ainsi, différents mutants de ces résidus ont pu être construits par ingénierie génétique lors d'un stage que j'ai effectué dans cette unité. L'effet de ces mutations sur l'activité enzymatique a pu être mesuré et a validé ainsi les prédictions faites *in vivo* grâce aux simulations de dynamique moléculaire. Parmi les résidus mis en évidence, Arg<sup>338</sup> et Asn<sup>347</sup> se trouvent aux deux extrémités d'une boucle étendue que l'on a désignée sous le nom de L $\alpha$ 16Cter. La boucle L $\alpha$ 16Cter connecte ainsi un site de fixation de la C-CaM avec la boucle catalytique. Elle permet ainsi d'assurer la transmission "allostérique" de l'information conformationnelle induite par la liaison de C-CaM. Les simulations de dynamique moléculaire combinées aux mesures expérimentales, tendent à montrer que L $\alpha$ 16Cter transmet la perturbation structurale induite par CaM au site catalytique.

Le premier modèle de l'allostérie a été proposé dans les années 1960 par Monod Wyman et Changeux [267]. Les protéines allostériques peuvent exister spontanément sous au moins deux états conformationnels interconvertibles à l'équilibre, correspondant à différentes activités fonctionnelles. La liaison d'un ligand stabilise un état conformationnel particulier de la protéine. Dans le cas où cet état correspond à une activité donnée de la protéine, l'effecteur allostérique agit donc comme un régulateur fonctionnel. Le concept de modulateurs allostériques est en train de se mettre en place en pharmacologie, et présente des avantages en terme de spécificité, d'efficacité et de réduction d'effets indésirables. Dans le système qui nous intéresse, il est à noter que des inhibiteurs ont été identifiés pour AC [184] par criblage virtuel, et validés expérimentalement, et que ces inhibiteurs sont supposés interagir de manière allostérique, dans une région située aussi à l'interface CaM/AC, mais distincte de la région qui nous intéresse.

Du point de vue biotechnologique, comme évoqué en introduction, de nombreuses applications biologiques ont été développées à partir des propriétés biologiques de CyaA durant ces deux dernières décennies (système vecteur d'antigène pour le développement de vaccins anti-cancéreux, méthode de criblage efficace pour cibler les protéines, système double-hybride pour l'étude des interactions protéine-protéine). La compréhension du contrôle allostérique de AC pourrait permettre d'envisager une nouvelle approche de biologie synthétique, où l'idée serait d'introduire des ponts disulfure stabilisant les extrémités de la boucle L $\alpha$ 16Cter. La formation ou la rupture de ces ponts disulfures aurait un effet mécanique analogue à la formation de liaison hydrogène et pourrait être contrôlé par la présence d'un réducteur en solution ou à terme d'une autre molécule effectrice. Ainsi, AC étant un puissant catalyseur de l'hydrolyse de l'ATP en AMPc, il deviendrait un moyen biotechnologique intéressant pour la mesure de faible concentration de molécule effectrice à travers la mesure facilement quantifiable d'AMPc produit.

L'autre objectif de la thèse était de prédire une structure stable du domaine catalytique AC de CyaA, étant donné que la structure tridimensionnelle à haute résolution de celui-ci à l'état non-complexé avec la calmoduline n'a pas pu être résolue à ce jour. En effet, AC isolé

a tendance à s'agréger, ce qui rend les courbes de SAXS impossible à analyser (A. Chenal, communication personnelle), et ce qui a probablement empêché jusqu'ici sa cristallisation et l'étude de sa structure par cristallographie aux rayons X. D'autre part, ce domaine est trop gros pour être résolu par RMN, où la taille maximale de la bio-molécule est limitée à 150-200 résidus.

Pour prédire des conformations de AC isolé, une méthode originale et récente (Maragliano et al, 2006 [150]) de dynamique moléculaire accélérée par la température ("*Temperature Accelerated Molecular Dynamics*" ou (*TAMD*)) a été utilisée dans le but d'accélérer l'échantillonnage conformationnel. Cette méthode permet d'étudier le système sans biaiser sa surface d'énergie libre. Par ailleurs, notre expérience nous a montré que des durées de trajectoires analogues à celles d'une dynamique moléculaire classique, étaient suffisantes pour observer des résultats intéressants. Une simulation de dynamique moléculaire classique réalisée sur AC isolé produit un premier échantillonnage de la population conformationnelle au cours de la trajectoire, qui a servi de base pour la définition de variables collectives appropriées. Ensuite, les simulations de *TAMD* ont permis de générer des conformations extrêmes assez éloignées de la conformation de AC en interaction avec C-CaM. Dans ces conformations, la région SA se rapproche fortement de la région CA, montrant une amplification des tendances conformationnelles observées dans la dynamique moléculaire classique. Cependant, ces conformations, bien que leurs structures secondaires soient stables, montrent une grande entropie conformationnelle qui les empêchent de rejoindre un bassin énergétique correspondant à une structure compacte stable. Une méthode heuristique, fondée sur l'analyse du solvant, a permis de sélectionner et de stabiliser certaines d'entre elles dans un bassin énergétique de l'état non-lié.

Toute la problématique des recherches conformationnelles d'un état particulier lorsque sa structure est inconnue repose sur la question "comment peut-on affirmer que la structure prédite correspond bien à celle de l'état recherché?". Néanmoins, il est toujours possible de comparer les résultats *in silico* aux mesures expérimentales *in vitro*, quand elles existent, de l'état actif. Dans notre cas, les mesures biophysiques approfondies et fiables [77], telles la forme sphérique de AC isolé en solution inférée par les paramètres hydrodynamiques, ou encore le faible pourcentage de résidus montrant une variation des structures secondaires, nous permettent seulement de dire que les conformations prédites par les simulations de *TAMD* se "rapprochent de" ou "pourraient correspondre à" l'état inactif de la protéine, ou tout du moins à un état non-lié.

D'autre part, l'identification de liaisons hydrogène stabilisant ces conformations prédites de AC à l'état libre sont un bon point de départ pour des expériences de "pontage" (ou "cross-linking") sur AC, dans le but de recréer ces conformations *in vitro*, et espérer ainsi pouvoir les cristalliser.

Ce travail de bioinformatique structurale a permis d'obtenir des informations complémentaires qui n'étaient pas atteignables par les expériences déjà menées à l'Unité de Biochimie des Interactions Macromoléculaires. Il a aussi permis de proposer aux biochimistes de nouvelles expériences

dont certaines ont déjà été réalisées et ont ouvert de nouvelles perspectives sur le contrôle de l'interaction AC/C-CaM.

# BIBLIOGRAPHIE

1. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 2006;**440**(7084) :631–636.
2. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Nat Acad Sci USA*, 1996;**93**(1) :13–20.
3. Grubhoffer L, Tichá M, Kocourek J. Isolation and properties of a lectin from the seeds of hairy vetch (*Vicia villosa* roth). *Biochem J*, 1981;**195**(3) :623–626.
4. Vyas NK. Atomic features of protein-carbohydrate interactions. *Curr Opin Struct Biol*, 1991;**1**(5) :732–740.
5. Schulz GE, Schirmer RH. Principles of Protein Structure. : Springer; 1979.
6. Jeon YH, Lee JY, Kim S. Chemical modulators working at pharmacological interface of target proteins. *Bioorg & Med Chem*, 2012;**20**(6) :1893–1901.
7. Stumpf MPH, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C. Estimating the size of the human interactome. *Proc Nat Acad Sci USA*, 2008;**105**(19) :6959–6964.
8. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J. The protein data bank and the challenge of structural genomics. *Nat Struct Biol*, 2000;**7 Suppl** :957–959.
9. Drenth J. Principles of Protein X-Ray Crystallography. : Springer-Verlag New York Inc ; 2006.
10. Pellicena P, Kuriyan J. Protein-protein interactions in the allosteric regulation of protein kinases. *Curr Opin Struct Biol*, 2006;**16**(6) :702–709.
11. Brotherton DH, Dhanaraj V, Wick S, Brizuela L, Domaille PJ, Volyanik E, Xu X, Parisini E, Smith BO, Archer SJ, Serrano M, Brenner SL, Blundell TL, Laue ED. Crystal structure of the complex of the cyclin d-dependent kinase cdk6 bound to the cell-cycle inhibitor p19INK4d. *Nature*, 1998;**395**(6699) :244–250.
12. Müller CW, Schlauderer GJ, Reinstein J, Schulz GE. Adenylate kinase motions during catalysis : an energetic counterweight balancing substrate binding. *Structure*, 1996;**4**(2) :147–156.
13. Allain FH-T, Gubser CC, Howe PWA, Nagai K, Neuhaus D, Varani G. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature*, 1996;**380**(6575) :646–650.
14. Collins FS, Morgan M, Patrinos A. The Human Genome Project : Lessons from Large-Scale Biology. *Science*, 2003;**300** :286.
15. Frazier ME, Johnson GM, Thomassen DG, Oliver CE, Patrinos A. The Human Genome Project : Lessons from Large-Scale Biology. *Science*, 2003;**300** :286.
16. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guillot S, Dunker AK. Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 1998; :437–448.
17. Daughdrill GW, Chadsey MS, Karlinsey JE, Hughes KT, Dahlquist FW. The c-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. *Nat Struct Biol*, 1997;**4**(4) :285–291.
18. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and cdk2-bound state : conformational disorder mediates binding diversity. *Proc Nat Acad Sci USA*, 1996;**93**(21) :11504–11509.
19. Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB : a model for activator :coactivator interactions. *Cell*, 1997;**91**(6) :741–752.
20. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cel Biol*, 2005;**6**(3) :197–208.

21. Gearhart MD, Holmbeck SMA, Evans RM, Dyson HJ, Wright PE. Monomeric complex of human orphan estrogen related receptor-2 with DNA : a pseudo-dimer interface mediates extended half-site recognition. *J Mol Biol*, 2003;**327**(4) :819–832.
22. Kucera NJ, Hodsdon ME, Wolin SL. An intrinsically disordered c terminus allows the la protein to assist the biogenesis of diverse noncoding RNA precursors. *Proc Nat Acad Sci USA*, 2011 ;**108**(4) :1308–1313.
23. Cheng Y, LeGall T, Oldfield CJ, Mueller JP, Van YYJ, Romero P, Cortese MS, Uversky VN, Dunker AK. Rational drug design via intrinsically disordered protein. *Trend Biotech*, 2006 ;**24**(10) :435–442.
24. Koshland DE. Application of a theory of enzyme specificity to protein synthesis. *Proc Nat Acad Sci USA*, 1958 ;**44**(2) :98–104.
25. Mészáros B, Simon I, Dosztányi Z. The expanding view of protein–protein interactions : complexes involving intrinsically disordered proteins. *Phys Biol*, 2011 ;**8**(3) :035003.
26. Vallurupalli P, Hansen DF, Kay LE. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc Nat Acad Sci USA*, 2008 ;**105** :11766–11771.
27. Spolar RS, Record, MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science*, 1994 ;**263**(5148) :777–784.
28. Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, Chan HS. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc Nat Acad Sci USA*, 2007 ;**104**(23) :9650–9655.
29. Iakoucheva LM, Radivojac P, Brown CJ, O’Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res*, 2004 ;**32**(3) :1037–1049.
30. Russo AA, Jeffrey PD, Patten AK, Massagué J, Pavletich NP. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature*, 1996 ;**382**(6589) :325–331.
31. Gunasekaran K, Tsai C-J, Kumar S, Zanuy D, Nussinov R. Extended disordered proteins : targeting function with less scaffold. *Trend Biochem Sciences*, 2003 ;**28**(2) :81–85.
32. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 2007 ;**450**(7171) :913–916.
33. Schaechter M, Medoff G, Eisenstein BI. *Microbiologie et pathologie infectieuse*. : De Boeck Supérieur ; 1999.
34. Inoue T, Moore P, Tsai B. How viruses and toxins disassemble to enter host cells. *Ann Rev Microb*, 2011 ;**65**(1) :287–305.
35. Donohue-Rolfe A, Acheson DW, Keusch GT. Shiga toxin : purification, structure, and function. *Rev. Infect. Dis*. 1991 ;**13 Suppl 4** :S293–297.
36. Hediger MA, Johnson DF, Nierlich DP, Zabin I. DNA sequence of the lactose operon : the lacA gene and the transcriptional termination region. *Proc. Natl. Acad. Sci. U.S.A.* 1985 ;**82**(19) :6414–6418.
37. Ahuja N, Kumar P, Bhatnagar R. The adenylate cyclase toxins. *Crit Rev Microb*, 2004 ;**30**(3) :187–196.
38. Tang WJ, Krupinski J, Gilman AG. Expression and characterization of calmodulin-activated (type i) adenylylcyclase. *J Biol Chem*, 1991 ;**266**(13) :8595–8603.
39. Shen Y, Lee YS, Soelaiman S, Bergson P, Lu D, Chen A, Beckingham K, Grabarek Z, Mrksich M, Tang WJ. Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylyl cyclase exotoxins. *EMBO J*, 2002 ;**21** :6721–6732.
40. Cates MS, Teodoro ML, Phillips GN. Molecular mechanisms of calcium and magnesium binding to parvalbumin. *Biophys J*, 2002 ;**82**(3) :1133–1146.
41. Vojtova-Vodolanova J, Basler M, Osicka R, Knapp O, Maier E, Cerny J, Benada O, Benz R, Sebo P. Oligomerization is involved in pore formation by bordetella adenylate cyclase toxin. *The FASEB Journal*, 2009 ;**23**(9) :2831–2843.

42. Ladant D, Ullmann A. *Bordetella pertussis* adenylate cyclase : a toxin with multiple talents. *Trend Microb*, 1999;**7**(4) :172 – 176.
43. Ladant D, Brezin C, Alonso JM, Crenon I, Guiso N. *Bordetella pertussis* adenylate cyclase. purification, characterization, and radioimmunoassay. *J Biol Chem*, 1986;**261**(34) :16264–16269.
44. Ladant D, Michelson S, Sarfati R, Gilles AM, Predeleanu R, Bâzru O. Characterization of the calmodulin-binding and of the catalytic domains of *bordetella pertussis* adenylate cyclase. *J Biol Chem*, 1989;**264**(7) :4015–4020.
45. Moayeri M, Leppla SH. Cellular and systemic effects of anthrax lethal toxin and edema toxin. *Mol Aspect Med*, 2009;**30**(6) :439–455.
46. Frank DW. The exoenzyme s regulon of *pseudomonas aeruginosa*. *Mol Microb*, 1997;**26**(4) :621–629.
47. Yahr TL, Vallis AJ, Hancock MK, Barbieri JT, Frank DW. ExoY, an adenylate cyclase secreted by the *pseudomonas aeruginosa* type III system. *Proc Nat Acad Sci USA*, 1998;**95**(23) :13899–13904.
48. Hritonenko V, Mun JJ, Tam C, Simon NC, Barbieri JT, Evans DJ, Fleiszig SMJ. Adenylate cyclase activity of *pseudomonas aeruginosa* ExoY can mediate bleb-niche formation in epithelial cells and contributes to virulence. *Microb Pathogen*, 2011;**51**(5) :305–312.
49. de Baillou G. *Epidemiorum et ephemeridum* 1969.
50. Musser JM, Hewlett EL, Peppler MS, Selander RK. Genetic diversity and relationships in populations of *bordetella* spp. *J Bacter*, 1986;**166**(1) :230–237.
51. Bordet J, Gengou O. *Le microbe de la coqueluche* ; 1906.
52. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, Jha P, Campbell H, Walker CF, Cibulskis R, Eisele T, Liu L, Mathers C. Global, regional, and national causes of child mortality in 2008 : a systematic analysis. *Lancet*, 2010;**375**(9730) :1969–1987.
53. Crowcroft NS, Pebody RG. Recent developments in pertussis. *Lancet*, 2006;**367**(9526) :1926–1936.
54. van der Zee A, Mooi F, Van Embden J, Musser J. Molecular evolution and host adaptation of *bordetella* spp. : phylogenetic analysis using multilocus enzyme electrophoresis and typing with three insertion sequences. *J Bacter*, 1997;**179**(21) :6609 –6617.
55. Parkhill J, Sebahia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MTG, Churcher CM, Bentley SD, Mungall KL, Cerdeno-Tarraga AM, Temple L, James K, Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A, Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O’Neil S, Ormond D, Price C, Rabinowitsch E, Rutter S, Sanders M, Saunders D, Seeger K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, Whitehead S, Barrell BG, Maskell DJ. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet*, 2003;**35**(1) :32–40.
56. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. Multilocus sequence typing : A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Nat Acad Sci USA*, 1998;**95**(6) :3140–3145.
57. Guiso N. *Bordetella pertussis* and pertussis vaccines. *Clin. Infect. Dis.*, Nov, 2009;**49** :1565–1569.
58. (OMS). Organisation mondiale pour la santé (oms) relevé épidémiologique hebdomadaire 2010.
59. Heininger U, Klich K, Stehr K, Cherry JD. Clinical findings in *Bordetella pertussis* infections : Results of a prospective multicenter surveillance study. *Pediatrics*, 1997;**100**(6) :e10.
60. Wendelboe AM, Van Rie A, Salmaso S, Englund JA. Duration of immunity against pertussis after natural infection or vaccination. *Pediatr Infect Dis J*, 2005;**24**(5 Suppl) :S58–61.
61. Goodwin MS, Weiss AA. Adenylate cyclase toxin is critical for colonization and pertussis toxin is critical for lethal infection by *Bordetella pertussis* in infant mice. *Infect Immun*, 1990;**58**(10) :3445–3447.
62. Khelef N, Sakamoto H, Guiso N. Both adenylate cyclase and hemolytic activities are required by *Bordetella pertussis* to initiate infection. *Microb Pathogen*, 1992;**12**(3) :227–235.



63. Weiss AA, Goodwin MS. Lethal infection by *Bordetella pertussis* mutants in the infant mouse model. *Infect Immun*, 1989;**57**(12) :3757–3764.
64. Khelef N, Zychlinsky A, Guiso N. *Bordetella pertussis* induces apoptosis in macrophages : role of adenylate cyclase-hemolysin. *Infect Immun*, 1993;**61**(10) :4064–4071.
65. Harvill ET, Cotter PA, Yuk MH, Miller JF. Probing the function of bordetella bronchiseptica adenylate cyclase toxin by manipulating host immunity. *Infect Immun*, 1999;**67**(3) :1493–1500.
66. Guermonprez P, Khelef N, Blouin E, Rieu P, Ricciardi-Castagnoli P, Guiso N, Ladant D, Leclerc C. The adenylate cyclase toxin of *Bordetella pertussis* binds to target cells via the  $\alpha M\beta 2$  integrin (Cd11b/Cd18). *J Exp Med*, 2001;**193**(9) :1035–1044.
67. Yáñez M, Gil-Longo J, Campos-Toimil M. Calcium binding proteins. *Adv Exp Med Biol*, 2012;**740** :461–482.
68. Scarlato V, Aricó B, Domenighini M, Rappuoli R. Environmental regulation of virulence factors in bordetella species. *BioEssays* :, *News Rev Mol Cel Develop Biol* 1993;**15**(2) :99–104.
69. El-Azami-El-Idrissi M, Bauche C, Loucka J, Osicka R, Sebo P, Ladant D, Leclerc C. Interaction of *Bordetella pertussis* adenylate cyclase with CD11b/CD18. *J Biol Chem*, 2003;**278**(40) :38514–38521.
70. Glaser P, Sakamoto H, Bellalou J, Ullmann A, Danchin A. Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of *Bordetella pertussis*. *EMBO J*, 1988;**7**(12) :3997–4004.
71. Bellalou J, Sakamoto H, Ladant D, Geoffroy C, Ullmann A. Deletions affecting hemolytic and toxin activities of *Bordetella pertussis* adenylate cyclase. *Infect Immun*, 1990;**58**(10) :3242–3247.
72. Benz R, Maier E, Ladant D, Ullmann A, Sebo P. Adenylate cyclase toxin (CyaA) of *Bordetella pertussis*. evidence for the formation of small ion-permeable channels and comparison with HlyA of *escherichia coli*. *J Biol Chem*, 1994;**269**(44) :27231–27239.
73. Knapp O, Maier E, Polleichtner G, Mašín J, Šebo P, Benz R. Channel formation in model membranes by the adenylate cyclase toxin of *Bordetella pertussis* : Effect of calcium. *Biochemistry*, 2003;**42**(26) :8077–8084.
74. Osičková A, Osička R, Maier E, Benz R, Šebo P. An amphipathic  $\alpha$ -Helix including glutamates 509 and 516 is crucial for membrane translocation of adenylate cyclase toxin and modulates formation and cation selectivity of its membrane channels. *J Biol Chem*, 1999;**274**(53) :37644–37650.
75. Bauche C, Chenal A, Knapp O, Bodenreider C, Benz R, Chaffotte A, Ladant D. Structural and functional characterization of an essential RTX subdomain of *Bordetella pertussis* adenylate cyclase toxin. *J Biol Chem*, 2006;**281**(25) :16914–16926.
76. Wolff J, Cook GH, Goldhammer AR, Berkowitz SA. Calmodulin activates prokaryotic adenylate cyclase. *Proc Nat Acad Sci USA*, 1980;**77**(7) :3841–3844.
77. Karst JC, Sotomayor Pérez AC, Guijarro JI, Raynal B, Chenal A, Ladant D. Calmodulin-induced conformational and hydrodynamic changes in the catalytic domain of bordetella pertussis adenylate cyclase toxin. *Biochemistry* 2010;**49** :318–328.
78. Rogel A, Hanski E. Distinct steps in the penetration of adenylate cyclase toxin of *Bordetella pertussis* into sheep erythrocytes. Translocation of the toxin across the membrane. *J Biol Chem*, Nov, 1992;**267** :22599–22605.
79. Rogel A, Schultz JE, Brownlie RM, Coote JG, Parton R, Hanski E. *Bordetella pertussis* adenylate cyclase : purification and characterization of the toxic form of the enzyme. *EMBO J*, 1989;**8**(9) :2755–2760.
80. Banchereau J, Briere F, Caux C, Davoust J, Lebecque S, Liu Y-J, Pulendran B, Palucka K. Immunobiology of dendritic cells. *Ann Rev Immun*, 2000;**18**(1) :767–811.
81. Osickova A, Masin J, Fayolle C, Krusek J, Basler M, Pospisilova E, Leclerc C, Osicka R, Sebo P. Adenylate cyclase toxin translocates across target cell membrane without forming a pore. *Mol Microb*, 2010;**75**(6) :1550–1562.
82. Otero AS, Yi XB, Gray MC, Szabo G, Hewlett EL. Membrane depolarization prevents cell invasion by *Bordetella pertussis* adenylate cyclase toxin. *J Biol Chem*, Apr, 1995;**270** :9695–9697.

83. Gordon VM, Young WW, Lechler SM, Gray MC, Leppla SH, Hewlett EL. Adenylate cyclase toxins from *Bacillus anthracis* and *Bordetella pertussis*. different processes for interaction with and entry into target cells. *J Biol Chem*, 1989;**264**(25) :14792–14796.
84. Hewlett EL, Gray L, Allietta M, Ehrmann I, Gordon VM, Gray MC. Adenylate cyclase toxin from *Bordetella pertussis*. conformational change associated with toxin activity. *J Biol Chem*, 1991;**266**(26) :17503–17508.
85. Rose T, Sebo P, Bellalou J, Ladant D. Interaction of calcium with *Bordetella pertussis* adenylate cyclase toxin. *J Biol Chem*, 1995;**270**(44) :26370–26376.
86. Chenal A, Guizarro JI, Raynal B, Delepierre M, Ladant D. RTX calcium binding motifs are intrinsically disordered in the absence of calcium. *J Biol Chem*, 2009;**284**(3) :1781–1789.
87. Gray M, Szabo G, Otero AS, Gray L, Hewlett E. Distinct mechanisms for k<sup>+</sup> efflux, intoxication, and hemolysis by *Bordetella pertussis* AC toxin. *J Biol Chem*, 1998;**273**(29) :18260–18267.
88. Karst JC, Barker R, Devi U, Swann M, Davi M, Roser SJ, Ladant D, Chenal A. Identification of a region that assists membrane insertion and translocation of the catalytic domain of *Bordetella pertussis* CyaA toxin. *J Biol Chem*, 2012;.
89. Khelef N, Gounon P, Guiso N. Internalization of *Bordetella pertussis* adenylate cyclase–haemolysin into endocytic vesicles contributes to macrophage cytotoxicity. *Cell Microb*, 2001;**3**(11) :721–730.
90. Schlecht G, Loucka J, Najar H, Sebo P, Leclerc C. Antigen targeting to CD11b allows efficient presentation of CD4<sup>+</sup> and CD8<sup>+</sup> t cell epitopes and in vivo Th1-Polarized t cell priming. *J Immunol*, 2004;**173**(10) :6089 –6097.
91. Boon T, Cerottini JC, van den Eynde B, van der Bruggen P, Van Pel A. Tumor antigens recognized by t lymphocytes. *Ann Rev Immun*, 1994;**12**(1) :337–365.
92. Ladant D, Glaser P, Ullmann A. Insertional mutagenesis of *Bordetella pertussis* adenylate cyclase. *J Biol Chem*, 1992;**267**(4) :2244 –2250.
93. Fayolle C, Sebo P, Ladant D, Ullmann A, Leclerc C. In vivo induction of CTL responses by recombinant adenylate cyclase of *Bordetella pertussis* carrying viral CD8<sup>+</sup> t cell epitopes. *J Immunol*, 1996;**156**(12) :4697 –4706.
94. Fayolle C, Ladant D, Karimova G, Ullmann A, Leclerc C. Therapy of murine tumors with recombinant *Bordetella pertussis* adenylate cyclase carrying a cytotoxic t cell epitope. *J Immunol*, 1999;**162**(7) :4157 –4162.
95. Ladant D. La toxine adénylcyclase de *Bordetella pertussis* : un vecteur d’antigène non-répliatif pour l’immunothérapie. In *Toxines et Cancer*, Goudet-Perrière F, Benoit E, Goyffon M, Marchot P, editors Meeting on Toxinology. Lavoisier Cachan 2006. ISBN : 2-7430-0958-6.
96. Préville X, Ladant D, Timmerman B, Leclerc C. Eradication of established tumors by vaccination with recombinant *Bordetella pertussis* adenylate cyclase carrying the human papillomavirus 16 e7 oncoprotein. *Cancer Research*, 2005;**65**(2) :641 –649.
97. Sory MP, Cornelis GR. Translocation of a hybrid YopE-adenylate cyclase from yersinia enterocolitica into HeLa cells. *Mol Microb*, 1994;**14**(3) :583–594.
98. Karimova G, Pidoux J, Ullmann A, Ladant D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Nat Acad Sci USA*, 1998;**95**(10) :5752 –5756.
99. Karimova G, Ullmann A, Ladant D. *Bordetella pertussis* adenylate cyclase toxin as a tool to analyze molecular interactions in a bacterial two-hybrid system. *Int J Med Microb*, 2000;**290**(4-5) :441–445.
100. Dautin N, Karimova G, Ladant D. *Bordetella pertussis* adenylate cyclase toxin : a versatile screening tool. *Toxicon*, 2002;**40**(10) :1383–1387.
101. Förster T. Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik*, 1948;**437**(1-2) :55–75.
102. Guo Q, Jureller JE, Warren JT, Solomaha E, Florián J, Tang WJ. Protein-protein docking and analysis reveal that two homologous bacterial adenylyl cyclase toxins interact with calmodulin differently. *J Biol Chem* 2008;**283** :23836–23845.

103. Means AR. Molecular mechanisms of action of calmodulin. *Recent Prog. Horm. Res.* 1988;**44** :223–262.
104. Vogel HJ. The Merck Frosst Award Lecture 1994. Calmodulin : a versatile calcium mediator protein. *Biochem Cell Biol*, 1994;**72** :357–376.
105. Van Eldik L, Watterson D. *Calmodulin and Signal Transduction*. San Diego : Academic Press. ; 1998.
106. Chin D, Means AR. Calmodulin : a prototypical calcium sensor. *Trends Cell Biol*, 2000;**10** :322–328.
107. Babu YS, Sack JS, Greenhough TJ, Bugg CE, Means AR, Cook WJ. Three-dimensional structure of calmodulin. *Nature*, 1985;**315**(6014) :37–40.
108. Yang C, Jas GS, Kuczera K. Structure, dynamics and interaction with kinase targets : computer simulations of calmodulin. *Biochim Biophys Acta*, 2004;**1697** :289–300.
109. Kuboniwa H, Tjandra N, Grzesiek S, Ren H, Klee CB, Bax A. Solution structure of calcium-free calmodulin. *Nat Struct Biol*, 1995;**2**(9) :768–776.
110. Chattopadhyaya R, Meador WE, Means AR, Quijcho FA. Calmodulin structure refined at 1.7 Å resolution. *J Mol Biol*, 1992;**228**(4) :1177–1192.
111. 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.
112. Lee AL, Sharp KA, Kranz JK, Song X-J, Wand AJ. Temperature dependence of the internal dynamics of a Calmodulin-Peptide complex. *Biochemistry*, 2002;**41**(46) :13814–13825.
113. Petrone P, Pande VS. Can conformational change be described by only a few normal modes? *Biophys J*, 2006;**90** :1583–1593.
114. Chen K, Ruan J, Kurgan LA. Prediction of three dimensional structure of calmodulin. *Protein J*, 2006;**25** :57–70.
115. Frederick KK, Kranz JK, Wand AJ. Characterization of the backbone and side chain dynamics of the CaM-CaMKII $\alpha$  complex reveals microscopic contributions to protein conformational entropy. *Biochemistry*, 2006;**45** :9841–9848.
116. Taylor DA, Sack JS, Maune JF, Beckingham K, Quijcho FA. Structure of a recombinant calmodulin from *Drosophila melanogaster* refined at 2.2-Å resolution. *J Biol Chem*, 1991;**255** :21375–21380.
117. Ikura M, Clore GM, Gronenborn AM, Zhu G, Klee CB, Bax A. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science*. 1992;**256** :632–638.
118. Osawa M, Tokumitsu H, Swindells MB, Kurihara H, Orita M, Shibamura T, Furuya T, Ikura M. A novel target recognition revealed by calmodulin in complex with Ca<sup>2+</sup>-calmodulin-dependent kinase kinase. *Nat Struct Biol*, 1999;**6**(9) :819–824.
119. Barbato G, Ikura M, Kay LE, Pastor RW, Bax A. Backbone dynamics of calmodulin studied by <sup>15</sup>N relaxation using inverse detected two-dimensional NMR spectroscopy : the central helix is flexible. *Biochemistry*, 1992;**31**(23) :5269–5278.
120. Trehwella J. The solution structures of calmodulin and its complexes with synthetic peptides based on target enzyme binding domains. *Cell Calcium*, 1992;**13**(6-7) :377–390.
121. 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.
122. Meador WE, Means AR, Quijcho FA. Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science*, 1993;**262**(5140) :1718–1721.
123. Clapperton JA, Martin SR, Smerdon SJ, Gamblin SJ, Bayley PM. Structure of the complex of calmodulin with the target sequence of Calmodulin-Dependent protein kinase I : Studies of the kinase activation mechanism. *Biochemistry*, 2002;**41**(50) :14669–14679.
124. Ishida H, Vogel HJ. Protein-peptide interaction studies demonstrate the versatility of calmodulin target protein binding. *Prot Pept Lett*, 2006;**13**(5) :455–465.
125. Elshorst B, Hennig M, Försterling 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**(38) :12320–12332.

126. Schumacher MA, Rivard AF, Blüchinger HP, Adelman JP. Structure of the gating domain of a  $Ca^{2+}$ -activated  $K^{+}$  channel complexed with  $Ca^{2+}$  calmodulin. *Nature*, 2001;**410**(6832) :1120–1124.
127. Juranic N, Atanasova E, Filoteo AG, Macura S, Prendergast FG, Penniston JT, Strehler EE. Calmodulin wraps around its binding domain in the plasma membrane  $Ca^{2+}$  pump anchored by a novel 18-1 motif. *J Biol Chem*, 2010;**285**(6) :4015–4024.
128. 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.
129. Hoefflich KP, Ikura M. Calmodulin in action : Diversity in target recognition and activation mechanisms. *Cell*, 2002 ;**108**(6) :739–742.
130. Vetter SW, Leclerc E. Novel aspects of calmodulin target recognition and activation. *Eur J Biochem*, 2003;**270**(3) :404–414.
131. Guo Q, Shen Y, Lee YS, Gibbs CS, Mrksich M, Tang WJ. Structural basis for the interaction of Bordetella pertussis adenyl cyclase toxin with calmodulin. *Embo J.*, 2005;**24** :3190–3201.
132. Bouhss A, Krin E, Munier H, Gilles AM, Danchin A, Glaser P, Bârzu O. Cooperative phenomena in binding and activation of bordetella pertussis adenylate cyclase by calmodulin. *J Biol Chem*, 1993;**268**(3) :1690–1694.
133. Shen Y, Zhukovskaya NL, Zimmer MI, Soelaiman S, Bergson P, Wang C-R, Gibbs CS, Tang W-J. Selective inhibition of anthrax edema factor by adefovir, a drug for chronic hepatitis b virus infection. *Proc Nat Acad Sci USA*, 2004;**101**(9) :3242–3247.
134. Ladant D. Interaction of bordetella pertussis adenylate cyclase with calmodulin. identification of two separated calmodulin-binding domains. *J Biol Chem*, 1988;**263**(6) :2612–2618.
135. Glaser P, Elmaoglou-Lazaridou A, Krin E, Ladant D, Bârzu O, Danchin A. Identification of residues essential for catalysis and binding of calmodulin in Bordetella pertussis adenylate cyclase by site-directed mutagenesis. *EMBO J*, 1989;**8** :967–972.
136. Vouquier S, Mary J, Dautin N, Vinh J, Friguet B, Ladant D. 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*, 2004;**279** :30210–30218.
137. 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.
138. Guo Q, Shen Y, Zhukovskaya NL, Florian J, Tang WJ. Structural and kinetic analyses of the interaction of anthrax adenyl cyclase toxin with reaction products cAMP and pyrophosphate. *J Biol Chem* 2004;**279** :29427–29435.
139. Martínez L, Laine E, Malliavin TE, Nilges M, Blondel A. Atp conformations and ion binding modes in the active site of anthrax edema factor : A computational analysis. *Proteins* 2009;**77** :971–983.
140. Martínez L, Malliavin TE, Blondel A. Mechanism of reactant and product dissociation from the Anthrax Edema Factor : a Locally Enhanced Sampling and Steered Molecular Dynamics Study. *Proteins* 2011 ;**79** :1649–1461.
141. Ulmer TS, Soelaiman S, Li S, Klee CB, Tang W-J, Bax A. Calcium dependence of the interaction between calmodulin and anthrax edema factor. *J Biol Chem*, 2003;**278**(31) :29261–29266.
142. Gilles AM, Munier H, Rose T, Glaser P, Krin E, Danchin A, Pellecuer C, Bârzu O. Intrinsic fluorescence of a truncated bordetella pertussis adenylate cyclase expressed in escherichia coli. *Biochemistry*, 1990;**29**(35) :8126–8130.
143. Munier H, Gilles A-M, Glaser P, Krin E, Danchin A, Sarfati R, Bârzu O. Isolation and characterization of catalytic and calmodulin-binding domains of bordetella pertussis adenylate cyclase. *Eur J Biochem*, 1991 ;**196**(2) :469–474.
144. Prêcheur B, Siffert O, Bârzu O, Craescu CT. NMR and circular dichroic studies on the solution conformation of a synthetic peptide derived from the calmodulin-binding domain of bordetella pertussis adenylate cyclase. *Eur J Biochem*, 1991 ;**196**(1) :67–72.

145. Cox JA, Comte M, Fitton JE, DeGrado WF. The interaction of calmodulin with amphiphilic peptides. *J Biol Chem*, 1985 ;**260**(4) :2527–2534.
146. O’Neil KT, Wolfe HR, Erickson-Viitanen S, DeGrado WF. Fluorescence properties of calmodulin-binding peptides reflect alpha-helical periodicity. *Science*, 1987 ;**236**(4807) :1454–1456.
147. Bouhss A, Vincent M, Munier H, Gilles AM, Takahashi M, Bârzu O, Danchin A, Gallay J. Conformational transitions within the calmodulin-binding site of bordetella pertussis adenylate cyclase studied by time-resolved fluorescence of trp242 and circular dichroism. *Eur J Biochem*, 1996 ;**237**(3) :619–628.
148. Munier H, Bouhss A, Krin E, Danchin A, Gilles AM, Glaser P, Bârzu O. The role of histidine 63 in the catalytic mechanism of bordetella pertussis adenylate cyclase. *J Biol Chem*, 1992 ;**267**(14) :9816–9820.
149. Laine E, Martínez L, Blondel A, Malliavin TE. Activation of the edema factor of *Bacillus anthracis* by calmodulin : evidence of an interplay between the EF-calmodulin interaction and calcium binding. *Biophys J* 2010 ;**99** :2264–2272.
150. Maragliano L, Ciccotti G, Vanden-Eijnden E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem Phys Lett*, 2006 ;**426** :168–175.
151. Verlet L. Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev*, 1967 ;**159** :98–103.
152. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods R. The Amber biomolecular simulation programs. *J Computat Chem*, 2005 ;**26** :1668–1688.
153. Hockney RW. The potential calculation and some applications. *Meth Comput Phys*, 1970 ;**9** :135–211.
154. 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.
155. Andersen HC. Rattle and A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J Comp Phys*, 1983 ;**52** :24–34.
156. Brooks BR, Brooks III CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caffisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM : The Biomolecular simulation Program. *J Comp Chem*, 2009 ;**30** :1545–1615.
157. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem*, 2005 ;**26** :1781–1802.
158. van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS : fast, flexible, and free. *J Comput Chem*, 2005 ;**26** :1701–1718.
159. Darden T, York D, Pedersen L. Partical Mesh Ewald : an  $N \cdot \log(N)$  method for Ewald sums in large systems. *J Chem Phys*, 1993 ;**98** :3684–90.
160. 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.
161. Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*, 1984 ;**81** :3684–90.
162. Tirion MM. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys Rev Lett*, 1996 ;**77** :1905–1908.
163. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des*, 1997 ;**2** :173–181.
164. Fischer S, Karplus M. Conjugate peak refinement : an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem Phys Lett*, 1992 ;**194** :252–261.

165. Schlitter J, Engels M, Krüger P. Targeted molecular dynamics : a new approach for searching pathways of conformational transitions. *J Mol Graph*, 1994;**12** :84–89.
166. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys J*, 1998;**75** :662–671.
167. Rhee YM, Pande VS. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys J*, 2003;**82** :775–786.
168. Pitera JW, Swope W. Understanding folding and design : replica-exchange simulations of "Trp-cage" miniproteins. *Proc Nat Acad Sci USA*, 2003;**100** :7587–7592.
169. Metropolis N, Rosenbluth A, Rosenbluth M, Teller A. Equation of state calculations by fast computing machines. *J Chem Phys*, 1953;**21** :1087–1092.
170. Zhang C, Ma J. Enhanced sampling and applications in protein folding in explicit solvent. *J Chem Phys*, 2010;**132**(24) :244101.
171. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics : a promising and efficient simulation method for biomolecules. *J Chem Phys*, 2004;**120**(24) :11919–11929.
172. Sørensen MR, Voter AF. Temperature-accelerated dynamics for simulation of infrequent events. *J Chem Phys*, 2000;**112**(21) :9599–9606.
173. Laio A, Rodriguez-Forteza A, Gervasio FL, Ceccarelli M, Parrinello M. Assessing the accuracy of metadynamics. *J Phys Chem B*, 2005;**109**(14) :6714–6721.
174. Gray MC, Hewlett EL. Cell cycle arrest induced by the bacterial adenylate cyclase toxins from *Bacillus anthracis* and *Bordetella pertussis*. *Cell Microbiol*, 2011;**13** :123–134.
175. Göttle M, Dove S, Kees F, Schlossmann J, Geduhn J, König B, Shen Y, Tang WJ, Kaever V, Seifert R. Cytidylyl and uridylyl cyclase activity of bacillus anthracis edema factor and *Bordetella pertussis* CyaA. *Biochemistry*, 2010;**49** :5494–5503.
176. Martín C, Gómez-Bilbao G, Ostolaza H. *Bordetella* adenylate cyclase toxin promotes calcium entry into both CD11b+ and CD11b- cells through cAMP-dependent L-type-like calcium channels. *J Biol Chem*, 2010;**285** :357–364.
177. Paccani SR, Dal Molin F, Benagiano M, Ladant D, D'Elis MM, Montecucco C, Baldari CT. Suppression of T-lymphocyte activation and chemotaxis by the adenylate cyclase toxin of *Bordetella pertussis*. *Infect Immun*, 2008;**76** :2822–2832.
178. Kamanova J, Kofronova O, Masin J, Genth H, Vojtova J, Linhartova I, Benada O, Just I, Sebo P. Adenylate cyclase toxin subverts phagocyte function by RhoA inhibition and unproductive ruffling. *J Immunol*, 2008;**181** :5587–5597.
179. Kim C, Wilcox-Adelman S, Sano Y, Tang WJ, Collier RJ, Park JM. Antiinflammatory cAMP signaling and cell migration genes co-opted by the anthrax bacillus. *Proc Nat Acad Sci USA*, 2008;**105** :6150–6155.
180. Puhar A, Dal Molin F, Horvath S, Ladant D, Montecucco C. Anthrax edema toxin modulates PKA- and CREB-dependent signaling in two phases. *PLoS One*, 2008;**3** :e3564.
181. Raymond B, Leduc D, Ravoux L, Goffic RL, Candela T, Raymondjean M, Goossens PL, Touqui L. Edema Toxin Impairs Anthracidal Phospholipase A2 Expression by Alveolar Macrophages. *PLoS Pathog.*, 2007;**3** :e187.
182. Drum CL, Shen Y, Rice PA, Bohm A, Tang WJ. Crystallization and preliminary X-ray study of the edema factor exotoxin adenylyl cyclase domain from *Bacillus anthracis* in the presence of its activator, calmodulin. *Acta Crystallogr D Biol Crystallogr* 2001;**57** :1881–1884.
183. Shen Y, Guo Q, Zhukovskaya NL, Drum CL, Bohm A, Tang WJ. Structure of anthrax edema factor-calmodulin-adenosine 5'-(alpha,beta-methylene)-triphosphate complex reveals an alternative mode of ATP binding to the catalytic site. *Biochem Biophys Res Commun* 2004;**317** :309–314.
184. Laine E, Goncalves C, Karst J, Lesnard A, Rault S, Tang WJ, Malliavin TE, Ladant D, Blondel A. Use of allostery to identify inhibitors of calmodulin- induced activation of *Bacillus anthracis* Edema Factor. *Proc Nat Acad Sci USA* 2010;**107** :11277–11282.

185. Laine E, Yoneda JD, Blondel A, Malliavin TE. The conformational plasticity of calmodulin upon calcium complexation gives a model of its interaction with the oedema factor of *Bacillus anthracis*. *Proteins* 2008 ;**71** :1813–1829.
186. Laine E, Blondel A, Malliavin TE. Dynamics and energetics : a consensus analysis of the impact of calcium on EF-CaM protein complex. *Biophys J* 2009 ;**96** :1249–1263.
187. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative Protein Structure Modeling using Modeller. *Curr Prot Bioinf*, 2006 ;**5** :5.6.
188. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 2006 ;**65** :712–725.
189. Jorgensen WL. Transferable intermolecular potential functions for water, alcohols and ethers. application to liquid water. *J Am Chem Soc*, 1981 ;**103** :335–340.
190. Aqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. *J Phys Chem*, 1990 ;**94** :8021–8024.
191. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA. AMBER 9. San Francisco : University of California ; 2006.
192. Lyman E, Zuckerman DM. Ensemble-based convergence analysis of biomolecular trajectories. *Biophys J*, 2006 ;**91** :164–172.
193. Hamacher K, Trylska J, McCammon J. Dependency map of proteins in the small ribosomal subunit. *PloS Comput. Biol.*, 2006 ;**2** :e10.
194. Srinivasan J, Cheatham T, Cieplak J, Kollman P, Case D. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J Am Chem Soc*, 1998 ;**120** :9401–9409.
195. Tan C, Tan YH, Luo R. Implicit nonpolar solvent models. *J Phys Chem B*, 2007 ;**111** :12263–12274.
196. Hollander M, Wolfe DA. Nonparametric Statistical Methods. New York John Wiley & Sons ; 1973.
197. R Development Core Team. R : A Language and Environment for Statistical Computing. Vienna, Austria : R Foundation for Statistical Computing ; 2009. ISBN 3-900051-07-0.
198. Wriggers W, Mehler E, Pitici F, Weinstein H, Schulten K. Structure and Dynamics of Calmodulin in Solution. *Biophys J*, 1998 ;**74** :1622–1639.
199. Yang C, Jas GS, Kuczera K. Structure and Dynamics of Calcium-activated Calmodulin in Solution. *J Biomol Struc Dyn*, 2001 ;**19** :247–271.
200. Shepherd CM, Vogel HJ. A Molecular Dynamics Study of Ca<sup>2+</sup>-Calmodulin : Evidence of Interdomain Coupling and Structural Collapse on the Nanosecond Timescale. *Biophys J*, 2004 ;**87** :780–791.
201. Yang C, Kuczera K. Molecular Dynamics Simulations of Calcium-Free Calmodulin in Solution. *J Biomol Struc Dyn*, 2002 ;**19** :801–819.
202. Komeiji Y, Ueno Y, Uebayasi M. Molecular dynamics simulations revealed Ca<sup>2+</sup>-dependent conformational change of Calmodulin. *FEBS Lett*, 2002 ;**521** :133–139.
203. 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.
204. Sanner MF, Olson AJ, Spehner JC. Fast and robust computation of molecular surfaces. In *Proc. 11th ACM Symp. Comp. Geom.* C6–C7 1995.
205. 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.
206. 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.
207. Braun P, Gingras AC. History of protein-protein interactions : From egg-white to complex network. *Proteomics*, 2012 ;**12** :1478–1498.

208. Mendez-Rios J, Uetz P. Global approaches to study protein-protein interactions among viruses and hosts. *Future Microbiol*, 2010;**5** :289–301.
209. Franzosa EA, Xia Y. Structural principles within the human-virus protein-protein interaction network. *Proc Nat Acad Sci USA*, 2011;**108** :10538–10543.
210. Tafforeau L, Rabourdin-Combe C, Lotteau V. Virus-human cell interactomes. *Methods Mol Biol*, 2012;**812** :103–120.
211. Misas-Villamil JC, van der Hoorn RA. Enzyme-inhibitor interactions at the plant-pathogen interface. *Curr Opin Plant Biol*, 2008;**11** :380–388.
212. Villoutreix BO, Labbé CM, Lagorce D, Laconde G, Sperandio O. A leap into the chemical space of Protein-Protein Interaction inhibitors. *Curr Pharm Des*, 2012; :in press.
213. Schwikowski B, Uetz P, Fields S. A network of protein-protein interactions in yeast. *Nat Biotech*, 2000;**18** :1257–1261.
214. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP : the database of interacting proteins. *Nucleic Acids Res*, 2000;**28** :289–291.
215. Kuzu G, Keskin O, Gursoy A, Nussinov R. Constructing structural networks of signaling pathways on the proteome scale. *Curr Opin Struct Biol*, 2012;**22** :367–377.
216. Dubay KH, Bothma JP, Geissler PL. Long-range intra-protein communication can be transmitted by correlated side-chain fluctuations alone. *PLoS Comput Biol*, 2011;**7** :e1002168.
217. del Sol A, Tsai CJ, Ma B, Nussinov R. The origin of allosteric functional modulation : multiple pre-existing pathways. *Structure*, 2009;**17** :1042–1050.
218. Chennubhotla C, Bahar I. Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput Biol*, 2007;**3** :1716–1726.
219. Reynolds KA, McLaughlin RN, Ranganathan R. Hot spots for allosteric regulation on protein surfaces. *Cell*, 2011;**147** :1564–1575.
220. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, 1999;**286** :295–299.
221. Selvaratnam R, Chowdhury S, VanSchouwen B, Melacini G. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc Nat Acad Sci USA*, 2011;**108** :6133–6138.
222. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R. Surface sites for engineering allosteric control in proteins. *Science*, 2008;**322** :438–442.
223. Dickson RJ, Wahl LM, Fernandes AD, Gloor GB. Identifying and seeing beyond multiple sequence alignment errors using intra-molecular protein covariation. *PLoS One*, 2010;**5** :e11082.
224. Grunberg R, Nilges M, Leckner J. Biskit—a software platform for structural bioinformatics. *Bioinformatics*, 2007;**23** :769–770.
225. Gmira S, Karimova G, Ladant D. Characterization of recombinant bordetella pertussis adenylate cyclase toxins carrying passenger proteins. *Res Microb*, 2001;**152**(10) :889–900.
226. Karimova G, Ullmann A, Ladant D. Protein-protein interaction between bacillus stearothermophilus tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. *J Mol Microb Biotech*, 2001;**3**(1) :73–82.
227. Miller JH. *A Short Course In Bacterial Genetics*. Cold Spring Harbor Laboratory Press : Cold Spring Harbor, NY ; 1992.
228. Whitley MJ, Lee AL. Exploring the role of structure and dynamics in the function of chymotrypsin inhibitor 2. *Proteins*, 2011;**79** :916–924.
229. Doring K, Surrey T, Grünewald S, John E, Jähnig F. Enhanced internal dynamics of a membrane transport protein during substrate translocation. *Protein Sci*, 2000;**9** :2246–2250.
230. Sacquin-Mora S, Laforet E, Lavery R. Locating the active sites of enzymes using mechanical properties. *Proteins*, 2007;**67** :350–359.
231. Sacquin-Mora S, Delalande O, Baaden M. Functional modes and residue flexibility control the anisotropic response of guanylate kinase to mechanical stress. *Biophys J*, 2010;**99** :3412–3419.



232. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol*, 2010;**6** :352–358.
233. Kokkinidis M, Glykos NM, Fadouloglou VE. Protein flexibility and enzymatic catalysis. *Adv Protein Chem Struct Biol*, 2012;**87** :181–218.
234. Ramanathan A, Agarwal PK. Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. *PLoS Biol*, 2011;**9** :e1001193.
235. Boekelheide N, Salomón-Ferrer R, Miller TF 3rd. Dynamics and dissipation in enzyme catalysis. *Proc Nat Acad Sci USA*, 2011;**108** :16159–16163.
236. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, Benkovic SJ, Wright PE. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science*, 2011;**332** :234–238.
237. Tama F, Sanejouand YH. Conformational change of proteins arising from normal modes calculations. *Protein Engineering*, 2001;**14** :1–6.
238. Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc Nat Acad Sci USA*, 2004;**101** :6957–6962.
239. Jiang J, Shrivastava IH, Watts SD, Bahar I, Amara SG. Large collective motions regulate the functional properties of glutamate transporter trimers. *Proc Nat Acad Sci USA*, 2011;**108** :15141–15146.
240. Ma B, Nussinov R. Release factors eRF1 and RF2 and a universal mechanism controls the large conformational changes. *J Biol Chem*, 2004;**279** :53875–53885.
241. Zachariae U, Grubmüller H. Importin-beta : structural and dynamic determinants of a molecular spring. *Structure*, 2008;**16** :906–915.
242. van Schaik RC, van Gunsteren WF, Berendsen HJC. Conformational Search by Potential Energy Annealing : Algorithm and Application to Cyclosporin A. *J Comput Aid Mol Design*, 1992;**6** :97–112.
243. Laio A, Parrinello M. Escaping free-energy minima. *Proc Nat Acad Sci USA*, 2002;**99** :12562–12566.
244. Tribello GA, Ceriotti M, Parrinello M. A self-learning algorithm for biased molecular dynamics. *Proc Nat Acad Sci USA*, 2010;**107** :17509–17514.
245. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics and a promising and efficient simulation method for biomolecules. *J Chem Phys*, 2004;**120** :11919–11929.
246. Markwick PR, McCammon JA. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys Chem Chem Phys*, 2011;**13** :20053–20065.
247. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling and throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem*, 2002;**53** :291–318.
248. Faradjian AK, Elber R. Computing time scales from reaction coordinates by milestoning. *J Chem Phys*, 2004;**120** :10880–10889.
249. Weinan E, Vanden-Eijnden E. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu Rev Phys Chem*, 2010;**61** :391–420.
250. Maragliano L, Cottone G, Ciccotti G, Vanden-Eijnden E. Mapping the network of pathways of co diffusion in myoglobin. *J Am Chem Soc*, 2010;**132** :1010–1017.
251. Abrams CF, Vanden-Eijnden E. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc Nat Acad Sci USA*, 2010;**107** :4961–4966.
252. Selwa E, Laine E, Malliavin TE. Differential role of Calmodulin and Calcium ions in the stabilization of the catalytic domain of adenylyl cyclase CyaA from *Bordetella pertussis*. *Proteins*, 2012 ; :1028–1040.
253. Adelman SA, Doll JD. Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - General Formulation for Classical Scattering Off Harmonic Solids. *J Chem Phys*, 1976;**64** :2375–2388.
254. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *J Chem Phys*, 1994;**101** :4177–4189.

255. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation : The Langevin piston method. *J Chem Phys*, 1995 ;**103** :4613–4622.
256. MacKerell Jr AD, Feig M, Brooks III CL. Extending the treatment of backbone energetics in protein force fields and limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comp Chem*, 2004 ;**25** :1400–1415.
257. MacKerell Jr AD, Bashford D, Bellott M, Dunbrack Jr RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher, III WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, 1998 ;**102** :3586–3616.
258. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, 1983 ;**79** :926–935.
259. Aleksiev T, Potestio R, Pontiggia F, Cozzini S, Micheletti C. PiSQRD and a web server for decomposing proteins into quasi-rigid dynamical domains. *Bioinformatics*, 2009 ;**25** :2743–2744.
260. Potestio R, Pontiggia F, Micheletti C. Coarse-grained description of protein internal dynamics and an optimal strategy for decomposing proteins in rigid subunits. *Biophys J*, 2009 ;**96** :4993–5002.
261. Micheletti C, Carloni P, Maritan A. Accurate and efficient description of protein vibrational dynamics and comparing molecular dynamics and Gaussian models. *Proteins*, 2004 ;**55** :635–645.
262. Gallaya J, Vincent M, Li de la Sierra IM, Munier-Lehmann H, Renouard M, Sakamoto H, Bârzu O, Gilles AM. Insight into the activation mechanism of *Bordetella pertussis* adenylate cyclase by calmodulin using fluorescence spectroscopy. *Eur J Biochem*, 2004 ;**271** :821–833.
263. Chen VB, Arendall III WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity and all-atom structure validation for macromolecular crystallography. *Acta Crystal D*, 2010 ;**66** :12–21.
264. Born B, Kim SJ, Ebbinghaus S, Gruebele M, Havenith M. The terahertz dance of water with the proteins and the effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discuss*, 2009 ;**141** :161–173.
265. Krivov SV, Karplus M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc Nat Acad Sci USA*, 2004 ;**101** :14766–14770.
266. Gfeller D, De Los Rios P, Caffisch A, Rao F. Complex network analysis of free-energy landscapes. *Proc Nat Acad Sci USA*, 2007 ;**104** :1817–1822.
267. Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions : a plausible model. *J Mol Biol*, 1965 ;**12** :88–118.

# ANNEXES

A.1 Tableau des pourcentages de formation des liaisons hydrogène établies entre C-CaM et AC.

Paire d'atomes	ACm1A	ACm2A	ACm3A	ACWT	ACWT	ACWTbis	ACWTbis	Région
				-(2Ca)	-(0Ca)	-(2Ca)	-(0Ca)	
GLU-87a-OE1/GLU-346-H	0.0	15.8	0.7	0.1	65.8	0.5	<b>94.6</b>	Cter
ARG-86a-2HH2/GLU-346-OE2	0.4	0.0	0.3	0.3	<b>99.7</b>	3.8	0.0	Cter
ARG-90a-1HH2/ASP-360-OD2	2.0	0.0	0.0	0.0	42.3	0.0	<b>91.7</b>	Cter
GLU-87a-OE2/ARG-348-2HH2	15.0	<b>87.2</b>	44.9	48.6	25.3	0.0	<b>89.6</b>	Cter
ARG-90a-2HH1/GLU-346-OE2	19.4	58.0	58.6	36.9	0.1	<b>95.3</b>	13.2	Cter
ARG-90a-2HH2/GLU-346-OE2	20.4	<b>86.2</b>	65.4	10.8	0.3	<b>80.9</b>	9.8	Cter
GLU-87a-OE2/ARG-348-2HH1	24.4	53.9	47.7	<b>99.4</b>	32.7	0.9	32.7	Cter
GLU-87a-OE1/ARG-348-2HH1	28.9	<b>86.0</b>	16.0	<b>86.3</b>	5.4	1.2	<b>96.1</b>	Cter
ARG-90a-HE/ASP-360-OD2	49.7	0.0	0.0	3.0	65.2	56.7	<b>83.0</b>	Cter
GLU-87a-OE1/ARG-348-2HH2	50.2	<b>88.8</b>	34.1	<b>96.9</b>	3.7	0.9	<b>95.8</b>	Cter
ARG-90a-HE/ASP-360-OD1	56.8	0.0	0.0	4.5	42.4	3.9	<b>89.9</b>	Cter
ARG-90a-HE/ASP-360-O	76.3	0.0	0.0	<b>99.6</b>	26.0	<b>99.9</b>	1.7	Cter
ARG-90a-2HH2/GLU-346-OE1	78.7	<b>91.8</b>	<b>93.9</b>	21.5	<b>98.5</b>	<b>86.0</b>	13.0	Cter
ARG-90a-2HH1/GLU-346-OE1	<b>80.6</b>	71.7	76.1	74.2	<b>99.1</b>	77.4	11.3	Cter
ARG-86a-HE/GLU-346-OE2	<b>81.8</b>	0.7	1.1	7.7	0.0	0.0	0.0	Cter
ARG-86a-HE/GLU-346-OE1	<b>93.4</b>	0.8	5.2	24.7	0.0	0.0	0.0	Cter
ARG-86a-1HH2/GLU-346-OE2	<b>93.6</b>	0.9	4.4	31.5	0.5	0.0	0.0	Cter
ARG-90a-1HH2/ASP-360-O	<b>98.3</b>	0.0	0.0	<b>99.4</b>	<b>85.0</b>	<b>99.8</b>	10.5	Cter
GLU-87a-OE2/GLU-346-H	<b>99.7</b>	7.1	10.6	<b>89.0</b>	35.3	4.1	6.8	Cter
ARG-86a-2HH1/GLU-346-OE1	0.0	0.0	0.0	0.4	<b>99.0</b>	9.0	0.1	Cter
PHE-344-O/ARG-90a-1HH1	0.6	<b>85.7</b>	<b>80.9</b>	0.0	0.0	63.8	0.1	CA
ASP-93a-OD1/ARG-338-2HH2	0.7	0.0	0.0	3.7	63.6	5.0	<b>92.7</b>	CA
ARG-90a-O/ARG-338-1HH2	8.1	0.0	0.0	74.7	39.7	19.9	<b>94.4</b>	CA
GLY-339-O/ASN-111a-1HD2	18.8	<b>90.2</b>	23.7	73.1	0.1	6.7	0.0	CA
ARG-90a-O/ARG-338-HE	66.0	0.0	0.0	<b>94.4</b>	27.5	28.6	<b>93.2</b>	CA
GLY-339-O/HIE-107a-HE2	<b>80.1</b>	<b>92.7</b>	72.4	77.9	17.0	28.6	67.2	CA
ASP-93a-O/ARG-338-HE	<b>96.4</b>	0.0	0.0	20.2	39.5	64.6	31.1	CA
ASP-93a-O/ARG-338-1HH2	<b>99.6</b>	0.0	0.0	27.0	<b>84.6</b>	<b>99.9</b>	<b>99.8</b>	CA
ASP-80a-OD2/ARG-250-HE	47.9	<b>86.6</b>	<b>90.2</b>	76.5	<b>97.0</b>	<b>91.2</b>	40.1	CA
ASP-80a-OD1/ARG-250-1HH2	<b>91.9</b>	<b>91.2</b>	<b>86.8</b>	<b>91.7</b>	<b>97.5</b>	<b>84.0</b>	<b>82.6</b>	CA
ASP-80a-OD1/ARG-250-HE	<b>92.4</b>	<b>81.8</b>	72.6	78.1	59.5	68.5	37.9	CA
ASP-80a-OD2/ARG-250-1HH2	<b>98.2</b>	71.0	<b>93.4</b>	<b>91.8</b>	<b>86.8</b>	<b>82.4</b>	<b>87.1</b>	CA
LEU-112a-O/ARG-202-1HH1	5.8	<b>100.0</b>	<b>100.0</b>	<b>80.3</b>	0.0	0.3	0.0	SA
GLU-120a-OE2/ARG-237-2HH2	49.9	<b>80.2</b>	38.6	0.8	26.6	0.0	31.0	SA
GLU-120a-OE1/ARG-237-2HH1	57.0	37.3	61.9	<b>83.2</b>	<b>80.8</b>	2.3	51.8	SA
GLU-120a-OE1/ARG-237-2HH2	60.6	25.4	67.5	<b>87.0</b>	61.2	2.4	59.9	SA
MET-124a-O/TRP-242-HE1	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>99.9</b>	<b>99.7</b>	<b>100.0</b>	<b>99.9</b>	SA
GLU-114a-OE2/ARG-206-1HH1	0.0	0.0	9.0	9.3	<b>99.1</b>	4.5	27.0	SA
GLU-114a-OE2/VAL-215-H	0.0	0.0	0.0	0.0	<b>97.8</b>	13.9	<b>90.5</b>	SA
GLU-87a-OE1/ARG-258-2HH1	27.2	44.6	63.5	<b>94.6</b>	50.9	77.5	7.2	SA
GLU-87a-OE1/ARG-258-2HH2	<b>97.9</b>	30.8	63.8	<b>98.2</b>	79.9	<b>93.4</b>	4.8	SA
ASP-80a-OD1/ARG-259-2HH2	12.3	61.2	52.9	50.3	<b>83.6</b>	49.9	30.2	SA
ASP-80a-OD2/ARG-259-2HH2	<b>86.9</b>	32.2	35.7	47.4	5.6	21.2	24.3	SA
GLU-84a-OE2/ARG-259-1HH2	<b>95.3</b>	<b>91.2</b>	<b>94.8</b>	<b>96.8</b>	<b>93.2</b>	<b>80.0</b>	<b>80.0</b>	SA
GLU-84a-OE1/ARG-259-1HH2	<b>96.2</b>	<b>90.0</b>	<b>92.1</b>	<b>94.7</b>	<b>94.6</b>	<b>81.4</b>	76.8	SA

TABLE I: *Pourcentage de formation des liaisons hydrogène établies entre C-CaM et AC.* Les liaisons hydrogène ont été sélectionnées parmi celles qui sont formées à plus de 80% du temps dans au moins une des trajectoires. Les pourcentages supérieurs à 80% sont marqués en gras. Les noms des trajectoires ont été modifiés de la façon suivante : le suffixe "WT" désigne les trajectoires "sauvages" du chapitre IV. Par exemple, ACWT correspond à la trajectoire AC-isolé. Le suffixe "bis" correspond aux trajectoires répétées du chapitre IV.

**A.2 Tableau des pourcentages de formation des liaisons hydrogène entre les résidus situés dans la boucle L $\alpha$ 16Cter reliant l'hélice  $\alpha$ 16 et la région Cterminale.**

Paire d' atomes	ACm1A -(2Ca)	ACm2A -(2Ca)	ACm3A -(2Ca)	ACWT -(2Ca)	ACWT -(0Ca)	ACWTbis -(2Ca)	ACWTbis -(0Ca)	ACWT	ACWTbis	Région
ALA-347-H/GLN-302-OE1	7.2	-	<b>99.8</b>	-	-	-	-	-	-	loop-cata/Cter
ASN-347-H/GLN-302-OE1	-	<b>100.0</b>	-	<b>99.5</b>	0.1	73.8	<b>98.8</b>	<b>91.2</b>	<b>90.9</b>	loop-cata/Cter
ASN-347-1HD2/ASN-304-O	-	<b>98.7</b>	-	<b>99.8</b>	5.2	68.5	<b>94.1</b>	2.1	58.3	loop-cata/Cter
ASN-347-2HD2/GLU-301-O	-	<b>98.4</b>	-	<b>99.8</b>	0.0	69.4	<b>92.6</b>	26.1	60.5	loop-cata/Cter
GLU-340-H/GLN-336-O	49.0	71.0	43.4	<b>83.3</b>	75.4	<b>87.1</b>	<b>96.4</b>	<b>82.8</b>	79.9	CA/CA
GLY-341-H/GLN-337-O	45.9	<b>80.9</b>	45.8	<b>86.3</b>	<b>85.6</b>	<b>81.6</b>	<b>93.5</b>	<b>84.1</b>	<b>80.1</b>	CA/CA
GLY-341-H/ARG-338-O	77.8	-	-	<b>95.1</b>	76.9	<b>84.2</b>	<b>92.2</b>	<b>90.6</b>	<b>93.2</b>	CA/CA
GLY-341-H/ALA-338-O	-	<b>96.2</b>	73.3	-	-	-	-	-	-	CA/CA
GLU-340-O/HIE-197-HE2	<b>83.0</b>	<b>99.2</b>	<b>93.3</b>	<b>97.3</b>	<b>88.6</b>	73.7	<b>92.0</b>	<b>89.6</b>	<b>88.5</b>	CA/CA
TYR-342-HH/ASN-35-OD1	<b>83.2</b>	<b>96.9</b>	17.2	0.0	0.5	0.0	0.0	0.0	71.7	CA/CA
VAL-343-H/PRO-196-O	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>99.3</b>	<b>100.0</b>	<b>100.0</b>	<b>99.9</b>	<b>99.9</b>	CA/CA
TYR-345-HH/GLU-276-OE2	12.9	17.4	20.9	57.8	59.0	13.9	<b>96.1</b>	26.3	59.0	CA/CA
TYR-345-HH/GLU-276-OE1	14.2	<b>86.2</b>	66.0	<b>98.8</b>	<b>91.6</b>	14.7	<b>85.8</b>	25.8	32.8	CA/CA
GLU-346-O/ARG-348-1HH1	21.8	<b>87.4</b>	49.2	<b>99.3</b>	7.1	0.0	0.2	0.3	0.1	Cter/Cter
GLU-346-OE1/LEU-357-H	0.0	4.1	0.0	0.0	0.0	9.3	78.3	<b>81.8</b>	78.0	Cter/Cter
ALA-347-O/TYR-350-H	56.6	-	<b>94.6</b>	-	-	-	-	-	-	Cter/Cter
ASN-347-O/TYR-350-H	-	77.8	-	<b>85.3</b>	22.9	<b>93.7</b>	<b>98.1</b>	64.5	76.5	Cter/Cter
ASN-347-O/GLY-351-H	0.0	3.6	0.0	1.4	0.0	<b>91.4</b>	37.6	20.3	45.6	Cter/Cter
ASN-347-OD1/ALA-349-H	-	<b>99.1</b>	-	<b>99.7</b>	<b>93.0</b>	<b>93.3</b>	<b>94.8</b>	<b>97.0</b>	<b>85.3</b>	Cter/Cter
ASP-360-OD1/ARG-338-2HH2	57.2	-	-	<b>82.9</b>	22.1	55.5	0.2	32.4	7.8	Cter/CA
ASP-360-OD2/ARG-338-2HH2	69.6	-	-	<b>82.0</b>	43.7	<b>95.2</b>	2.4	37.0	12.7	Cter/CA
ASP-360-OD1/ARG-338-2HH1	<b>86.0</b>	-	-	<b>89.1</b>	61.0	<b>97.6</b>	6.4	31.6	3.8	Cter/CA
ASP-360-OD2/ARG-338-2HH1	<b>92.1</b>	-	-	<b>96.9</b>	38.8	<b>97.7</b>	7.8	36.6	12.1	Cter/CA
GLY-341-O/LEU-198-H	<b>100.0</b>	<b>99.9</b>	<b>96.8</b>	<b>99.4</b>	<b>99.8</b>	40.7	<b>100.0</b>	<b>99.2</b>	<b>99.3</b>	CA/SA

**TAB. II: Pourcentage de formation des liaisons hydrogène entre les résidus situés dans la boucle  $L\alpha 16$  Cter reliant l'hélice  $\alpha 16$  et la région Cterminale.** Les liaisons hydrogène sélectionnées sont formées à plus de 80% du temps dans au moins une des trajectoires. Les liaisons hydrogène qui ne peuvent pas être formées à cause de la mutation d' un résidu sont indiquées par le symbole '-'. Les noms des trajectoires ont été modifiés de la façon suivante : le suffixe "WT" désigne les trajectoires "sauvages" du chapitre IV. Par exemple, ACWT correspond à la trajectoire AC-isolé. Le suffixe "bis" correspond aux trajectoires répétées du chapitre IV.

## A.3 Séquence ADN et traduction en acides aminés du variant ACm3A

```

1 ATG CAG CAA TCG CAT CAG GCT GGT TAC GCA AAC GCC GCC GAC CGG GAG TCT GGC ATC CCC
  1 M Q Q S H Q A G Y A N A A D R E S G I P

61 GCA GCC GTA CTC GAT GGC ATC AAG GCC GTG GCG AAG GAA AAA AAC GCC ACA TTG ATG TTC
21 A A V L D G I K A V A K E K N A T L M F

121 CGC CTG GTC AAC CCC CAT TCC ACC AGC CTG ATT GCC GAA GGG GTG GCC ACC AAA GGA TTG
41 R L V N P H S T S L I A E G V A T K G L

181 GGC GTG CAC GCC AAG TCG TCC GAT TGG GGG TTG CAG GCG GGC TAC ATT CCC GTC AAC CCG
61 G V H A K S S D W G L Q A G Y I P V N P

241 AAT CTT TCC AAA CTG TTC GGC CGT GCG CCC GAG GTG ATC GCG CGG GCC GAC AAC GAC GTC
81 N L S K L F G R A P E V I A R A D N D V

301 AAC AGC AGC CTG GCG CAT GGC CAT ACC GCG GTC GAC CTG ACG CTG TCG AAA GAG CGG CTT
101 N S S L A H G H T A V D L T L S K E R L

361 GAC TAT CTG CGG CAA GCG GGC CTG GTC ACC GGC ATG GCC GAT GGC GTG GTC GCG AGC AAC
121 D Y L R Q A G L V T G M A D G V V A S N

421 CAC GCA GGC TAC GAG CAG TTC GAG TTT CGC GTG AAG GAA ACC TCG GAC GGG CGC TAT GCC
141 H A G Y E Q F E F R V K E T S D G R Y A

481 GTG CAG TAT CGC CGC AAG GGC GGC GAC GAT TTC GAG GCG GTC AAG GTG ATC GGC AAT GCC
161 V Q Y R R K G G D D F E A V K V I G N A

541 GCC GGT ATT CCA CTG ACG GCG GAT ATC GAC ATG TTC GCC ATT ATG CCG CAT CTG TCC AAC
181 A G I P L T A D I D M F A I M P H L S N

601 TTC CGC GAC TCG GCG CGC AGT TCG GTG ACC AGC GGC GAT TCG GTG ACC GAT TAC CTG GCG
201 F R D S A R S S V T S G D S V T D Y L A

661 CGC ACG CGG CGG GCC GCC AGC GAG GCC ACG GGC GGC CTG GAT CGC GAA CGC ATC GAC TTG
221 R T R R A A S E A T G G L D R E R I D L

721 TTG TGG AAA ATC GCT CGC GCC GGC GCC CGT TCC GCA GTG GGC ACC GAG GCG CGT CGC CAG
241 L W K I A R A G A R S A V G T E A R R Q

781 TTC CGC TAC GAC GGC GAC ATG AAT ATC GGC GTG ATC ACC GAT TTC GAG CTG GAA GTG CGC
261 F R Y D G D M N I G V I T D F E L E V R

841 AAT GCG CTG AAC AGG CGG GCG CAC GCC GTC GGC GCG CAG GAC GTG GTC CAG CAT GGC ACT
281 N A L N R R A H A V G A Q D V V Q H G T

901 GAG CAG AAC AAT CCT TTC CCG GAG GCA GAT GAG AAG ATT TTC GTC GTA TCG GCC ACC GGT
301 E Q N N P F P E A D E K I F V V S A T G

961 GAA TCT CAG ATG CTG ACC CGT GGT CAG CTG AAA GAA TAC ATT GGC CAG CAG GCA GGT GAA
321 E S Q M L T R G Q L K E Y I G Q Q A G E

1021 GGT TAC GTA TTC TAC GAA GCT CGC GCG TAC GGT GTT GCA GGT AAA TCT CTG TTC GAT GCT
341 G Y V F Y E A R A Y G V A G K S L F D A

1081 GGC CTA GGC GCG GCG CCG GGT GTT CCG TCT GGC CGT TCG AAG TTC TCT CCG GAT GTT CTG
361 G L G A A P G V P S G R S K F S P D V L

1141 GAA ACC GTT CCC GGG AAG TAA
381 E T V P G K *

```

FIG. 1: Les codons Arg<sup>338</sup>, Asn<sup>347</sup> et Asp<sup>360</sup> mutés en Ala sont indiqués en rouge. Les sites reconnus par les enzymes de restriction SnaBI et BsiWI sont surlignés en jaune.

## A.4 Séquence ADN et traduction en acides aminés du variant ACm2A

```

1 ATG CAG CAA TCG CAT CAG GCT GGT TAC GCA AAC GCC GCC GAC CGG GAG TCT GGC ATC CCC
1 M Q Q S H Q A G Y A N A A D R E S G I P

61 GCA GCC GTA CTC GAT GGC ATC AAG GCC GTG GCG AAG GAA AAA AAC GCC ACA TTG ATG TTC
21 A A V L D G I K A V A K E K N A T L M F

121 CGC CTG GTC AAC CCC CAT TCC ACC AGC CTG ATT GCC GAA GGG GTG GCC ACC AAA GGA TTG
41 R L V N P H S T S L I A E G V A T K G L

181 GGC GTG CAC GCC AAG TCG TCC GAT TGG GGG TTG CAG GCG GGC TAC ATT CCC GTC AAC CCG
61 G V H A K S S D W G L Q A G Y I P V N P

241 AAT CTT TCC AAA CTG TTC GGC CGT GCG CCC GAG GTG ATC GCG CGG GCC GAC AAC GAC GTC
81 N L S K L F G R A P E V I A R A D N D V

301 AAC AGC AGC CTG GCG CAT GGC CAT ACC GCG GTC GAC CTG ACG CTG TCG AAA GAG CGG CTT
101 N S S L A H G H T A V D L T L S K E R L

361 GAC TAT CTG CGG CAA GCG GGC CTG GTC ACC GGC ATG GCC GAT GGC GTG GTC GCG AGC AAC
121 D Y L R Q A G L V T G M A D G V V A S N

421 CAC GCA GGC TAC GAG CAG TTC GAG TTT CGC GTG AAG GAA ACC TCG GAC GGG CGC TAT GCC
141 H A G Y E Q F E F R V K E T S D G R Y A

481 GTG CAG TAT CGC CGC AAG GGC GGC GAC GAT TTC GAG GCG GTC AAG GTG ATC GGC AAT GCC
161 V Q Y R R K G G D D F E A V K V I G N A

541 GCC GGT ATT CCA CTG ACG GCG GAT ATC GAC ATG TTC GCC ATT ATG CCG CAT CTG TCC AAC
181 A G I P L T A D I D M F A I M P H L S N

601 TTC CGC GAC TCG GCG CGC AGT TCG GTG ACC AGC GGC GAT TCG GTG ACC GAT TAC CTG GCG
201 F R D S A R S S V T S G D S V T D Y L A

661 CGC ACG CGG CGG GCC GCC AGC GAG GCC ACG GGC GGC CTG GAT CGC GAA CGC ATC GAC TTG
221 R T R R A A S E A T G G L D R E R I D L

721 TTG TGG AAA ATC GCT CGC GCC GGC GCC CGT TCC GCA GTG GGC ACC GAG GCG CGT CGC CAG
241 L W K I A R A G A R S A V G T E A R R Q

781 TTC CGC TAC GAC GGC GAC ATG AAT ATC GGC GTG ATC ACC GAT TTC GAG CTG GAA GTG CGC
261 F R Y D G D M N I G V I T D F E L E V R

841 AAT GCG CTG AAC AGG CGG GCG CAC GCC GTC GGC GCG CAG GAC GTG GTC CAG CAT GGC ACT
281 N A L N R R A H A V G A Q D V V Q H G T

901 GAG CAG AAC AAT CCT TTC CCG GAG GCA GAT GAG AAG ATT TTC GTC GTA TCG GCC ACC GGT
301 E Q N N P F P E A D E K I F V V S A T G

961 GAA TCT CAG ATG CTG ACC CGT GGT CAG CTG AAA GAA TAC ATT GGC CAG CAG GCA GGT GAA
321 E S Q M L T R G Q L K E Y I G Q Q A G E

1021 GGT TAC GTT TTC TAC GAA AAC CGC GCG TAC GGT GTT GCA GGT AAA TCT CTG TTC GAT GCT
341 G Y V F Y E N R A Y G V A G K S L F D A

1081 GGC CTA GGC GCG GCG CCG GGT GTT CCG TCT GGC CGT TCG AAG TTC TCT CCG GAT GTT CTG
361 G L G A A P G V P S G R S K F S P D V L

1141 GAA ACC GTT CCC GGG AAG TAA
381 E T V P G K *

```

FIG. 2: Les codons Arg<sup>338</sup> et Asp<sup>360</sup> mutés en Ala sont indiqués en rouge.



## A.5 Séquence ADN et traduction en acides aminés du variant ACm1A

```

1 ATG GGT CAG CAG TCT CAC CAG GCT GGT TAC GCT AAC GCT GCC GAT CGT GAA TCC GGT ATC
1 M G Q Q S H Q A G Y A N A A D R E S G I

61 CCG GCT GCA GTA CTG GAC GGT ATC AAA GCT GTT GCT AAA GAA AAA AAC GCA ACG TTG ATG
21 P A A V L D G I K A V A K E K N A T L M

121 TTC CGC CTG GTA AAC CCG CAC TCT ACC TCA TTA ATT GCT GAA GGC GTT GCT ACT AAA GGC
41 F R L V N P H S T S L I A E G V A T K G

181 CTG GGC GTG CAT GCG AAA TCT TCT GAT TGG GGT CTG CAG GCT GGT TAC ATT CCG GTT AAC
61 L G V H A K S S D W G L Q A G Y I P V N

241 CCG AAC CTG TCT AAA CTG TTC GGC CGC GCA CCG GAA GTT ATC GCA CGT GCA GAC AAC GAC
81 P N L S K L F G R A P E V I A R A D N D

301 GTC AAC TCT TCT CTG GCG CAC GGC CAC ACC GCG GTC GAC CTG ACC CTG TCT AAA GAA CGT
101 V N S S L A H G H T A V D L T L S K E R

361 CTG GAT TAC CTG CGC CAG GCT GGC CTG GTT ACC GGC ATG GCA GAT GGT GTT GTT GCT TCT
121 L D Y L R Q A G L V T G M A D G V V A S

421 AAC CAC GCA GGC TAC GAA CAG TTT GAA TTC CGT GTT AAA GAA ACT AGT GAT GGC CGC TAT
141 N H A G Y E Q F E F R V K E T S D G R Y

481 GCT GTT CAG TAC CGT CGT AAA GGT GGT GAT GAT TTC GAA GCT GTT AAA GTT ATC GGC AAC
161 A V Q Y R R K G G D D F E A V K V I G N

541 GCT GCA GGC ATC CCG CTG ACC GCT GAT ATC GAC ATG TTC GCT ATC ATG CCG CAC TTA AGC
181 A A G I P L T A D I D M F A I M P H L S

601 AAC TTC CGC GAC AGC GCT CGT TCT TCT GTT ACC AGC GGT GAT TCT GTT ACC GAC TAC CTG
201 N F R D S A R S S V T S G D S V T D Y L

661 GCT CGC ACG CGT CGT GCG GCT AGC GAA GCG ACC GGC GGT CTG GAT CGC GAA CGT ATC GAT
221 A R T R R A A S E A T G G L D R E R I D

721 CTG CTG TGG AAA ATT GCG CGT GCC GGC GCA CGC TCT GCT GTT GGT ACC GAA GCG CGC CGT
241 L L W K I A R A G A R S A V G T E A R R

781 CAG TTC CGT TAT GAT GGT GAC ATG AAC ATC GGT GTG ATC ACC GAC TTC GAG CTC GAA GTT
261 Q F R Y D G D M N I G V I T D F E L E V

841 CGT AAC GCT CTG AAC CCG CGT GCA CAC GCG GTT GGT GCG CAG GAC GTT GTA CAG CAC GGC
281 R N A L N R R A H A V G A Q D V V Q H G

901 ACC GAA CAG AAC AAC CCG TTC CCG GAA GCT GAT GAA AAG ATC TTC GTT GTT TCT GCT ACC
301 T E Q N N P F P E A D E K I F V V S A T

961 GGT GAA TCT CAG ATG CTG ACC CGT GGT CAG CTG AAA GAA TAC ATT GGC CAG CAG CGT GGT
321 G E S Q M L T R G Q L K E Y I G Q Q R G

1021 GAA GGT TAC GTG TTC TAC GAA GCT CGC GCG TAC GGT GTT GCA GGT AAA TCT CTG TTC GAT
341 E G Y V F Y E A R A Y G V A G K S L F D

1081 GAT GGC CTA GGC GCG GCG CCG GGT GTT CCG TCT GGC CGT TCG AAG TTC TCT CCG GAT GTT
361 D G L G A A P G V P S G R S K F S P D V

1141 CTG GAA ACC GTC CCG GGT AAA TAA
381 L E T V P G K *

```

**FIG. 3:** Le codon Asn<sup>347</sup> muté en Ala est indiqué en rouge. Un codon Gly a été ajouté juste après le codon d'initiation Met.

**A.6** Publication décrivant le rôle de la calmoduline et des ions calcium dans la stabilisation du domaine catalytique de l'adényl-cyclase CyaA de *Bordetella pertussis*

# Differential role of calmodulin and calcium ions in the stabilization of the catalytic domain of adenylyl cyclase CyaA from *Bordetella pertussis*

Edithe Selwa,<sup>1</sup> Elodie Laine,<sup>2</sup> and Thérèse E. Malliavin<sup>1\*</sup>

<sup>1</sup>Unité de Bioinformatique Structurale, Institut Pasteur and CNRS URA 2185, Structural Biology and Chemistry Department, 25-28, rue du Dr. Roux, 75 724 Paris, France

<sup>2</sup>Laboratoire de Biologie et de Pharmacologie Appliquée Ecole Normale Supérieure de Cachan and CNRS UMR8113, 61, avenue du Président Wilson 94 235 Cachan, France

## ABSTRACT

The catalytic adenylyl cyclase (AC) domain of the protein CyaA from *Bordetella pertussis* is activated by interaction with the C terminal lobe of calmodulin (C-CaM). The AC/C-CaM complex displays an elongated shape, but hydrodynamics measurements on the isolated AC domain allowed to characterize the shape of the protein as spherical. Here, we study by molecular dynamics simulations the complexes between AC and the apo and Ca<sup>2+</sup>-loaded C-CaM, as well as the isolated AC, to characterize the features of AC conformational variability and of AC/C-CaM interaction. The removal of calcium ions from C-CaM increases the AC flexibility, but the removal of C-CaM induces a dramatic drift of the AC conformation. Isolated AC conformations show a general tendency to become less elongated, as the two protein extremities (regions SA and CB) tend to get closer. An analysis of the energetic influences between the C-CaM and the AC regions shows a simple influence scheme, in agreement with the high affinity of AC to CaM. In this scheme, a single influence is observed from C-CaM to the region CA of the AC domain. This influence is correlated to the presence of hydrogen bonds involving residues from C-CaM, and from regions CA, C-terminal tail, and catalytic loop of AC. This study reveals a C-CaM/AC interaction picture where C-CaM stabilizes AC by a steric hindrance on the conformational drift of SA, whereas the Ca<sup>2+</sup> ions allow further stabilization by the establishment of a hydrogen bond network extending from C-CaM to the AC catalytic loop.

Proteins 2011; 00:000–000.  
© 2011 Wiley Periodicals, Inc.

**Key words:** calmodulin; *Bordetella pertussis*; adenylyl cyclase; molecular dynamics; calcium; MMPBSA; essential dynamics.

## INTRODUCTION

The adenylyl cyclases (ACs) Edema Factor from *Bacillus anthracis* and CyaA from *Bordetella pertussis* play an essential role in the host invasion, by overproducing cAMP in host cells, thus disorganizing the cellular signaling processes through a variety of mechanisms.<sup>1–8</sup> The catalytic domains of the Edema Factor and of CyaA are activated as ACs by interaction with the ubiquitous protein calmodulin (CaM). CaM comprises two lobes connected by a flexible linker. Each CaM lobe contains four  $\alpha$  helices, arranged by pairs of helices, named EF-hands.

For both the Edema Factor and the AC domain of CyaA, the interaction with CaM induces a large conformational transition. The Edema Factor transition was deciphered by the determination of X-ray crystallographic structures of the enzyme in free and CaM-bound states.<sup>9</sup> By contrast, crystallographic structures of the AC domain are available only in complex with the Ca<sup>2+</sup>-loaded C terminal lobe of CaM (C-CaM).<sup>10</sup> These AC structures are relevant to the AC activity as C-CaM was shown<sup>11</sup> to be able to fully activate the AC domain. In the X-ray crystallographic structure of AC/C-CaM complex (Fig. 1), the AC domain includes three main regions, CA (green) in the middle of the structure, CB (orange), and SA (purple) at the two extremities. The Edema Factor and AC sequences showed<sup>10</sup> that the C terminal tail (cyan), the catalytic loop (yellow), and helices F and H of SA correspond to the switches C, B, and A of Edema Factor.<sup>9</sup> As these switches undergo large conformational changes during the Edema Factor activation, one may wonder whether their CyaA counterparts also undergo substantial rearrangements during the AC domain activation.

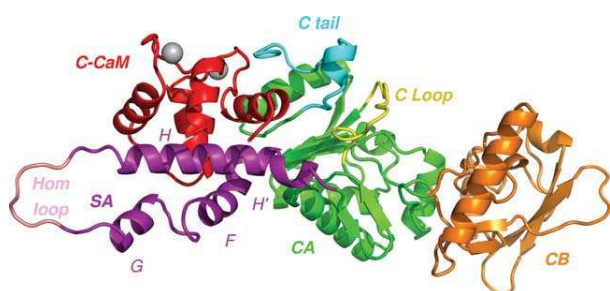
The structural information obtained on Edema Factor by Wei-Jen Tang and his coworkers,<sup>9,12–15</sup> paved the way to a thorough characterization of the physico-chemical features of the AC activity using molecular modeling approaches.<sup>16,17</sup> Furthermore, the prediction of

Additional Supporting Information may be found in the online version of this article.

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

Received 12 June 2011; Revised 4 November 2011; Accepted 14 November 2011

Published online 5 December 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24005



**Figure 1**

X-ray crystallographic structure of the catalytic domain (AC) of the adenylyl cyclase CyaA (PDB entry 1YRT) drawn in cartons. Helices F, G, H, and H' coloured in purple form the region SA. The loop 226–232 (Hom loop) at the left extremity of SA, colored in salmon, was missing in the X-ray crystallographic structure and was modeled using Modeller9v4.<sup>24</sup> The other protein regions are colored in green(CA), orange(CB), cyan(C terminal tail: C tail), and yellow(catalytic loop: C loop). The C-CaM lobe, colored in red, interacts with SA and the C terminal tail. Calcium ions are represented by silver beads.

a plausible conformational transition path, based on the knowledge of the free and CaM-bound Edema Factor structures, allowed to propose inhibitors of the Edema Factor activation, which were subsequently experimentally validated.<sup>18</sup> The role of the calcium ions into the Edema Factor/CaM interaction was also investigated.<sup>19–21</sup>

The lack of structural information available on the isolated AC domain makes the activation analysis by molecular modeling less straightforward than for Edema Factor. Nevertheless, the binding site of the N-terminal lobe of CaM was predicted,<sup>22</sup> based on molecular modeling and cross-linking experiments. Besides, ultracentrifugation and light scattering measurements showed<sup>23</sup> that the free AC domain is hydrated and spherical.

To gain information on the role of C-CaM and  $\text{Ca}^{2+}$  ions in the AC domain activation, we performed molecular dynamics (MD) simulations of three systems: AC-(2Ca–C–CaM) starting from the PDB structure 1YRT,<sup>10</sup> AC-(0Ca–C–CaM) where the  $\text{Ca}^{2+}$  ions were removed from the structure, and AC-isolated where the isolated AC domain was simulated starting from the AC conformation observed in 1YRT. The removal of the  $\text{Ca}^{2+}$  ions and C-CaM induced a large conformational drift of AC, possibly leading to a less elongated protein shape by bringing closer the SA and CB regions located at the protein extremities.

An analysis of the energetic influences in the complex, performed using an approach previously proposed on the Edema Factor/CaM complex,<sup>20</sup> revealed a single influence of C-CaM on the region CA in the AC-(0Ca–C–CaM) and AC-(2Ca–C–CaM) systems. This simple influence scheme is in agreement with the high affinity of AC for CaM, and correlates with the formation of an hydro-

gen bond network connecting C-CaM, the C terminal tail, CA, and catalytic loop residues.

## MATERIALS AND METHODS

### Preparation of simulated systems

The three systems AC-(2Ca–C–CaM), AC-(0Ca–C–CaM), and AC-isolated were built in the following way. Coordinates of AC-(2Ca–C–CaM) were obtained from the PDB entry 1YRT,<sup>10</sup> and coordinates of AC-(0Ca–C–CaM) were produced by removing  $\text{Ca}^{2+}$  ions. The system AC-isolated was generated by removing the ions and C-CaM from 1YRT. As the residues 226–232 forming a loop at the extremity of the SA domain could not be observed in the X-ray crystallographic structure, the loop was inserted using the protocol `build_loop_fixed.py` from Modeller9v4.<sup>24</sup> The modeled loop (Hom loop) is colored in salmon on Figure 1.

Hydrogen atoms were added with the LEaP module from AMBER 9<sup>25</sup> using the force field FF99SB,<sup>26</sup> and  $\text{Na}^+$  counterions were added to neutralize the system. The three systems were then hydrated in a box of TIP3P<sup>27</sup> water molecules using a cutoff of 10 Å and periodic boundary conditions. Details are summed up in Supporting Information Table I. The Calcium Lennard-Jones parameters were a van der Waals radius  $R$  of 1.7131 Å, and a well depth  $\epsilon$  of 0.459789 kcal mol<sup>-1</sup>.<sup>28</sup> Three additional MD trajectories AC-(2Ca–C–CaM)-2nd, AC-(0Ca–C–CaM)-2nd, and AC-isolated-2nd were recorded in the same conditions than the ones described in Supporting Information Table I, to probe the reproducibility of the results.

### MD simulations

The 30 ns MD trajectories were recorded using the Particle mesh Ewald molecular dynamics (PMEMD) module from AMBER 9.0,<sup>29</sup> on clusters of 4 and 8 processors. A cutoff of 10 Å was defined for Lennard-Jones interactions calculations, while long-range electrostatic interactions were calculated with the Particle Mesh Ewald protocol.<sup>30</sup> Temperature and pressure were regulated according to the Berendsen thermostat,<sup>31</sup> and the simulations were performed at temperature 298 K and pressure 1 atm. The SHAKE algorithm<sup>32</sup> was used to keep rigid all covalent bonds involving hydrogens, enabling a time step of 2 fs. Atomic coordinates were saved every picosecond.

The MD trajectories were initiated in the following way. First, minimizations were repeated using 10 steepest descent cycles, followed by 490 conjugated-gradient cycles. During these minimizations, harmonic restraints on the solute atom positions were gradually released with a force constant varying from 100 to 0 kcal/(mol Å<sup>2</sup>). The systems were then thermalized from 0 to 298 K during 20 ps, with harmonic restraints on the solute atom

positions using a force constant of 25 kcal/(mol Å<sup>2</sup>). The equilibration process was then performed with one cycle of 5 ps at constant volume, followed by four cycles of 2.5 ps at constant pressure, one cycle of 10 ps and the last one of 60 ps. Restrained harmonic forces were applied on the solute atom positions during the first five cycles and the force constant was released from 25 to 0 kcal/(mol Å<sup>2</sup>).

### Trajectories convergence

The convergence of the trajectories was analyzed using the approach developed by Lyman and Zuckerman.<sup>33</sup> It consists in generating a set of reference structures, and in classifying the MD conformations in the corresponding reference clusters, according to an arbitrary chosen cut-off  $d_c$  for the coordinate root-mean-square deviation (RMSD). A conformation is randomly picked up from the trajectory: this is the first reference structure. All conformations that show a coordinate RMSD from this reference structure smaller than  $d_c$  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  $d_c$  to the new reference structure are removed. The operation is repeated until there is no remaining conformation, to produce a set of reference structures that deviate from each other by at least  $d_c$  Å.

The reference structures are then used to probe the conformational drift of the trajectories. Each trajectory is split in two halves and conformations from each half are clustered to their closest reference structure in terms of coordinate RMSD. If the trajectory is closer to a stationary state, each conformation cluster should be equally populated by conformations from both halves of the trajectory. As large conformational drift observed during the first nanosecond of each trajectory can bias the convergence analysis, this nanosecond was removed before performing the conformation clustering.

### Energetic influences

The energetic influences approach intends to compare the scale of interaction energies between protein regions, according to their environment. These influences were analyzed along each MD trajectory according to the algorithm introduced by Hamacher et al.<sup>34</sup> and previously used on the Edema Factor/CaM complex.<sup>20</sup> The energetic influences are distinct from interaction energies between regions, as a region can gain or lose rank even with small change of binding energy. The change in energy ranking gives information about the hierarchy of regions in assembling the complex, according to the general principle formulated by Hamacher et al.<sup>34</sup>: “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, and we attribute an influence of protein Z on protein X.”

The molecular systems were divided in four regions: CA (including the C terminal tail and the catalytic loop), CB and SA for AC, and C-CaM. The interaction energy was evaluated as the Molecular Mechanics/Poisson-Boltzmann Surface Area (MMPBSA)<sup>35</sup>  $\Delta G$  binding energy using AMBER 9,<sup>29</sup> on snapshots extracted every 10 ps along the 27 last nanoseconds of the MD trajectories:

$$\Delta G = \Delta E_{MM} + \Delta G_{solv} \quad (1)$$

where  $\Delta E_{MM}$  represents mechanics contribution, expressed as the sum of the electrostatics ( $\Delta E_{ele}$ ), van der Waals ( $\Delta E_{vdw}$ ), and internal ( $\Delta E_{int}$ ) contributions.  $\Delta G_{solv}$  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 pbsa module of AMBER9 allowed to compute the polar contribution  $\Delta G_{sol}^{polar}$  by solving the Poisson-Boltzmann equation with a grid spacing of 0.5 Å and dielectric constants of 1 and 80 for the interior and exterior of the solute. The dielectric boundary was defined by rolling a 1.4 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 and an attractive (dispersion) term  $\Delta G_{disp}$  computed by surface-integration approach.<sup>36</sup>

$$\Delta G_{sol}^{nonpolar} = \Delta G_{disp} + \Delta G_{cavity} \quad (2)$$

The energetic influences were evaluated by the following algorithm. First, MMPBSA binding energy of each region  $D$  to the complex  $C$  is calculated as:

$$\Delta \Delta G[D, C] = \Delta G[C] - (\Delta G[C \setminus D] + \Delta G[D]) \quad (3)$$

The energies  $\Delta \Delta G[D, C]$  are then sorted from the highest to the lowest, producing a reference rank for each region.

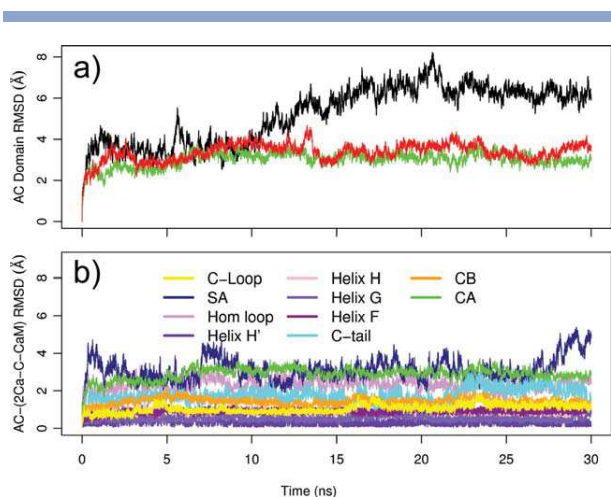
Second, to assess the energetic influence of a region  $D^*$  on the other regions, the region  $D^*$  is removed from the complex, and the MMPBSA binding energy of each region  $D$  on the truncated complex  $C \setminus D^*$  is calculated:

$$\Delta \Delta G[D, C \setminus D^*] = \Delta G[C \setminus D^*] - (\Delta G[C \setminus (D + D^*)] + \Delta G[D]) \quad (4)$$

Resulting energies are then sorted and statistically probed and the  $\Delta \Delta G[D, C]$  and  $\Delta \Delta G[D, C \setminus D^*]$  lists are compared. If a region  $D'$  has lost a rank comparing to its reference rank, so  $D^*$  has an energetic influence on  $D'$ . The influence intensity corresponds to the number of ranks lost by  $D'$ .

The regions ranking is probed according to the standard deviations. If the difference of mean energy values is larger than the sum of the standard deviations, the ranking





**Figure 2**

Conformational drifts, estimated through Carbons  $\alpha$  coordinates RMSD from the starting conformations. (a) Drift calculated on Ca carbons of the AC domain for AC-(2Ca-C-CaM) (green), AC-(0Ca-C-CaM) (red), and AC-isolated (black). (b) Drift calculated on different regions of the AC domain, analyzed for AC-(2Ca-C-CaM). The curves colors according to the analyzed region are given in the legend of (b). The AC regions were defined from the following residue sets: 7–61, 187–197, 261–299, and 313–345 for CA, residues 300–312 for the catalytic loop (C loop), residues 346–364 for the C terminal tail (C tail) residues 62–186 for CB and residues 198–260 for SA. The helices H, F, G, H' correspond to residues 234–253, 198–210, 214–223, 256–260.

produced by the mean energy values is supposed to be reliable. Otherwise, the ranking was further probed by a nonparametric Mann–Whitney test<sup>37</sup> performed on the energy estimations time series using the software R<sup>38</sup> (command `wilcox.test`). Both alternative hypotheses  $\Delta\Delta G(D) > \Delta\Delta G(D')$  and  $\Delta\Delta G(D') > \Delta\Delta G(D)$  are tested, and the one having a  $p$ -value smaller than 0.05 is rejected. The energetic influences were calculated by extensive use of the `mm_pbsa.pl` script of the package AMBER 9.<sup>25</sup> The analysis of a 30 ns trajectory required typically one week of CPU time on 1 processor of a 2.6 GHz workstation.

## RESULTS

### Conformational drifts

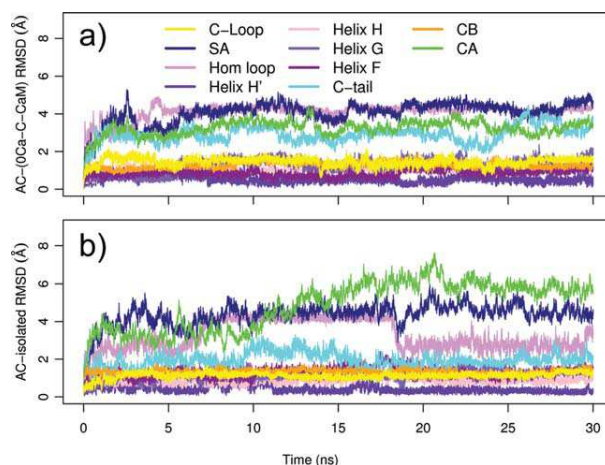
The conformational drift of the AC domain [Fig. 2(a)] displays a plateau around 3 Å in the presence of C-CaM [trajectories AC-(2Ca-C-CaM) and AC-(0Ca-C-CaM)] and jumps to 6 Å in the absence of C-CaM (trajectory AC-isolated). Interaction with the CaM lobe thus seems to be an essential ingredient for stabilizing AC, and the presence of  $\text{Ca}^{2+}$  ions plays only a marginal role in that respect. This is in agreement with the small influence of  $\text{Ca}^{2+}$  concentration to the AC catalytic activity.<sup>11</sup>

The conformational drift of the C-CaM is the same in the two trajectories AC-(2Ca-C-CaM) and AC-(0Ca-C-CaM),

and oscillates around 1.2 Å (data not shown). This small value corresponds to a quite limited reorganization of the lobe. Indeed, RMSDs in the 0.8–2 Å (respectively 2–3 Å) range were previously observed for MD simulations in the presence<sup>39–42</sup> (respectively in the absence<sup>43–45</sup>) of  $\text{Ca}^{2+}$  ions. Therefore, inside the AC/C-CaM complex, C-CaM is stabilized by its interaction with AC.

The conformational drifts were analyzed in more details by considering the  $\alpha$  carbons coordinate RMSD of various AC regions. The RMSDs were calculated by superimposing independently [Figs. 2(b) and 3] the following AC regions: CA, CB, C terminal tail (C tail), SA, catalytic loop (Loop C),  $\alpha$ -helices H, H' G, F, and the loop 226–232 (Hom loop) not visible in the crystallographic structure 1YRT and built by homology modeling. The respective colors of the drift curves are defined in the legend of Figures 2(b) and 3(a).

In the absence of C-CaM or  $\text{Ca}^{2+}$  ions (Fig. 3), Hom loop (pink), CA (green), and SA (dark purple) display the main RMSD increase. Indeed, the Hom loop (pink) is stable around 2 Å for AC-(2Ca-C-CaM), drops to 4 Å in AC-(0Ca-C-CaM), and displays transitions between 4 and 2 Å in AC-isolated, produced by the loop flip. The conformational drift of CA (green) shows a slight increase in AC-(0Ca-C-CaM) and a larger one, up to 6 Å, in AC-isolated. It is worth noting that, in the 12–30 ns range of the trajectory AC-isolated [Fig. 3(b)], the RMSD profile of CA is similar to the RMSD profile for AC [Fig. 2(a), black curve]: the CA drift thus dominates the AC global drift. The SA RMSD (dark purple) drops from around 3 Å in AC-(2Ca-C-CaM) up to 4 Å



**Figure 3**

Conformational drifts, estimated through Carbons  $\alpha$  coordinates RMSD from the starting conformations. Drift calculated on different regions of the AC domain, analyzed for AC-(0Ca-C-CaM) (a) and AC-isolated (b). The curves colors according to the analyzed region are given in the legend of (a). The AC regions were defined as described in the caption in Figure 2.

**Table I**

Angles (°) between the axes of the  $\alpha$  helices H (Residues 234–253), G (Residues 214–223), F (Residues 198–210), and H' (Residues 256–260)

Trajectory (°)	AC-(2Ca–C–CaM)	AC-(0Ca–C–CaM)	AC-isolated
Angle H–G	45.1 ± 7.1	23.0 ± 8.9	34.1 ± 6.4
Angle G–F	160.4 ± 9.1	148.5 ± 8.0	133.8 ± 13.1
Angle H–F	54.1 ± 4.0	52.0 ± 2.8	65.6 ± 6.7
Angle H–H'	39.3 ± 3.2	72.7 ± 13.7	31.3 ± 13.0

Axes were defined by taking the successive middle points between atoms

(a) N(*i*) and N(*i*+2), (b) C- $\alpha$  (*i*) and C- $\alpha$  (*i*+2), (c) C(*i*) and C(*i*+2), *i* being the residue index in the sequence.

in AC-(0Ca–C–CaM), but keeps a similar profile in AC-(0Ca–C–CaM) and AC-isolated. The large SA drift agrees with the correspondence of this region to the switch A of the AC Edema Factor.<sup>10</sup> The C terminal tail (cyan) displays smaller RMSD values than SA, CA, and Hom Loop, but is nevertheless relatively flexible around 2 Å, in agreement with the equivalence of this region to the switch C of the Edema Factor.<sup>10</sup>

The catalytic loop (yellow), CB (orange), and  $\alpha$  helices H, H' G, F (various tones of purple) show RMSD drifts smaller than 2 Å in all trajectories. The small variation observed for CB demonstrates that this region keeps rigid and agrees with its position, the farthest away from the binding site of CaM on AC. The small conformational drifts observed for the  $\alpha$  helices H, H' G, F indicate that these helices are stable, which is confirmed by an independent analysis of the secondary structures (data not shown). Thus, the large drift observed for SA arises from variations in the orientations of helices H, F, G, and H' relative to each other. Indeed, large standard deviations are observed for G–F and H–H' angles in AC-(0Ca–C–CaM) or AC-isolated (Table I). Consequently, the variability in the helices relative orientations together with the large RMSD described earlier for the Hom loop produce the large SA drift.

The conformational drifts obtained for the three additional MD trajectories AC-(2Ca–C–CaM)-2nd, AC-(0Ca–C–CaM)-2nd, and AC-isolated-2nd (Supporting Information Fig. S1) display features quite similar to those described previously. Slightly larger drifts are observed for: (i) SA in the trajectory of the complex AC/C–CaM in the presence of Ca<sup>2+</sup> ions, and of (ii) C terminal tail in the trajectory of the complex AC/C–CaM in the absence of Ca<sup>2+</sup> ions. The loop reconstructed by homology is less mobile in the trajectory of the isolated AC. Nevertheless, the relative mobilities of the different regions are preserved, which proves the reproducibility of the observations made on the MD trajectories.

The trajectories convergence was also analyzed using the method of Lyman and Zuckerman,<sup>33</sup> as described in Materials and Methods. Representative clusters of solute conformations were extracted from the trajectories by using RMSD cut-off values  $d_c$  of 2.0, 2.5, 2.7, and 3.0 Å. The stationary state of the trajectories was probed by

checking whether the clusters were similarly populated in two trajectory halves. The Supporting Information Figure S2 displays the analysis results obtained using four different seeds for the conformation random picking-up. All trajectories display variations in the number of clusters, as well as in the relative populations of the clusters in the two trajectory halves. This variability could arise from a difficulty to choose the right reference clusters, due to the lack of sensitivity of the atomic coordinate RMSD in the context of a non globular protein. The trajectory AC-isolated [Supporting Information Fig. S2(c)] displays trends distinct from the ones observed in AC-(2Ca–C–CaM) [Supporting Information Fig. S1(a)] and AC-(0Ca–C–CaM) [Supporting Information Fig. S2(b)], as many clusters get populated only from the first trajectory half. This indicates that in AC-isolated, the solute has left the initial conformational basin. The convergence analysis of the 15–30 ns of AC-isolated [Supporting Information Fig. S2(d)] shows a smaller number of clusters and a stabilization of the cluster populations with respect to the full AC-isolated trajectory, which are the sign that the MD simulation attained a bottleneck in the exploration of the AC conformational space.

The atomic fluctuations of the AC domain display their largest peak in the loop 226–232 [Supporting Information Fig. S3(a)], built by homology modeling. This loop also displays the largest fluctuations variations between the different MD trajectories, with a maximum increase of 2 Å for the fluctuation peak. Furthermore, differences in the fluctuation profiles of AC-isolated and the other trajectories are mostly visible for the residues 180–357, corresponding to SA, the C terminal tail, the catalytic loop and the portion of CA that connects them. These differences agree with the large conformational drifts observed for SA and CA in AC-isolated [Fig. 3(b)]. For C–CaM [Supporting Information Fig. S3(b)], small flexibility increases are observed when the Ca<sup>2+</sup> ions are removed from the complex, in the C terminal part and in the Ca<sup>2+</sup> loop of the EF hand 3 between helices V and VI [Supporting Information Fig. S2(b)].

The various conformational and convergence analyses conducted here give a consistent description of the structural variations of the studied systems. The isolated AC domain drifts more largely than the AC domain in complex with C–CaM. The largest conformational variations are mainly located in the regions SA and CA.

### Globular tendency of AC and catalytic activity

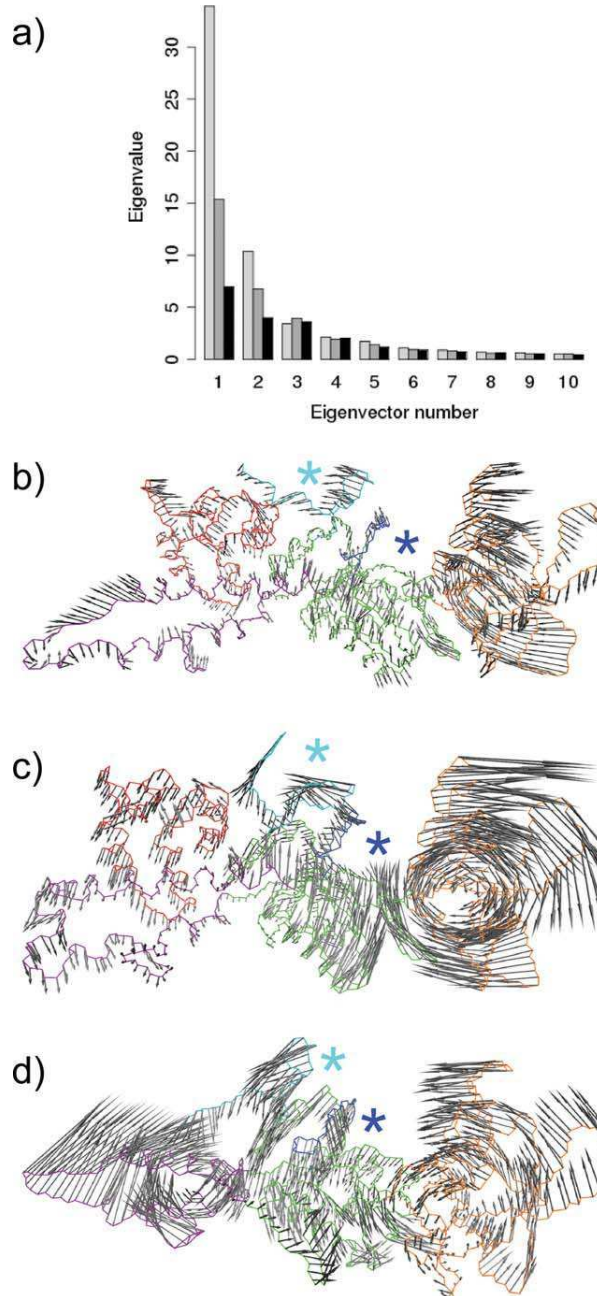
No high resolution structure is available for the isolated AC domain, but hydrodynamic measurements detected<sup>23</sup> a spherical shape of this domain in the free state in solution. Here, we analyzed the overall solute shape by calculating the gyration radius defined as the mean quadratic distance between the solute geometric

center and any solute atom. The gyration radius [Supporting Information Fig. S4] is similar and stable in the 26.5–27.0 Å interval for the trajectories AC-(2Ca–C–CaM) and AC-(0Ca–C–CaM), and drops down from 25.5 to 24.2 Å, in the trajectory AC-isolated. The variations of radii between AC-isolated and the other trajectories are small, but reveals nevertheless a significant tendency of the solute to decrease in size. Furthermore, inertia moments were calculated along the AC-isolated trajectory [Supporting Information Fig. S5] and the largest moments decrease along the trajectory, which demonstrates that the protein displays a tendency to be less elongated.

A principal component analysis of the dynamics covariance matrix gives additional information on the conformational tendencies of the solute: different profiles of eigenvectors are obtained [Fig. 4(a)] for AC-isolated and the other trajectories. For AC-isolated, the first eigenvalue is dominant and 10 eigenvectors explain 90% of the protein internal motion, whereas for AC-(2Ca–C–CaM) and AC-(0Ca–C–CaM), the first 2–3 eigenvalues are dominant, and 30 or 19 eigenvectors explain 90% of the complex internal motion. The free AC thus displays more collective motions.

The first mode projection on the solute structure shows the largest amplitudes for SA and CB [Fig. 4(b–d)], and demonstrates a tendency of these regions to rotate around an axis perpendicular to the plane of the figure. This motion is the most pronounced for AC-isolated, and supports the view of a large AC deformation leading to a spherical shape by bringing SA and CB closer to each other, in agreement with the experimental hydrodynamic measurements.<sup>23</sup> Besides, in AC-(2Ca–C–CaM) [Fig. 4(b)], CB displays much larger motions than the remaining part of AC, whereas in AC-(0Ca–C–CaM) [Fig. 4(c)], the amplitude of CB motions increases while CA also gets to move. In AC-isolated [Fig. 4(d)], a large displacement is observed for SA, whereas the CA and CB motion amplitudes decrease relatively. The large first mode projection observed here for CB along with its small conformational drift (Figs. 2 and 3) prove that this region moves as a rigid body.

The motion tendencies of the complex were compared to the complex static architecture, analyzed from the distances between centers of mass from different regions [Supporting Information Fig. S6]. In the presence of C–CaM, the centers of mass of CA and SA stay at a stable distance, in the 32–33 Å interval [Supporting Information Fig. S6(a)], whereas in AC-isolated, the CA/SA distance oscillates and decreases of about 4 Å. This large distance variation agrees with the motions observed in AC-isolated by principal component analysis of the covariance matrix [Fig. 4(d)]. The distance between the centers of mass of the C terminal tail and of the C–CaM drops of about 2 Å when the Ca<sup>2+</sup> ions are removed from C–CaM [Supporting Information

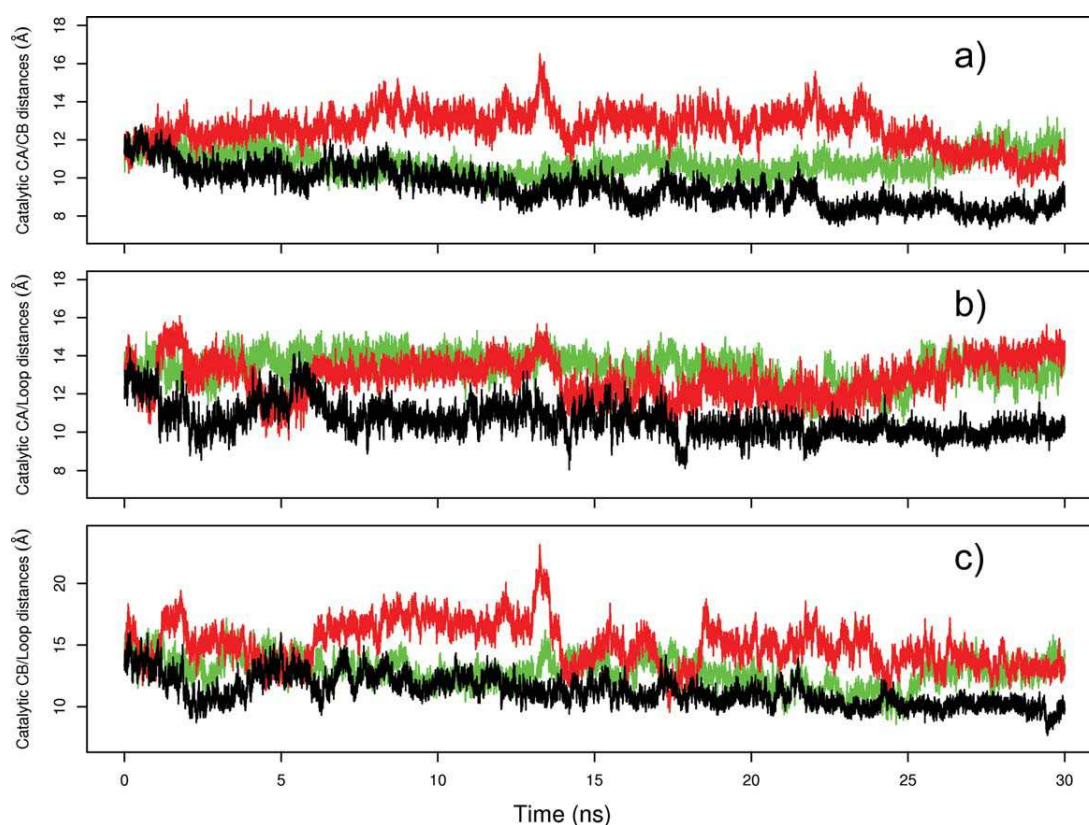


**Figure 4**

Principle component analysis of the internal dynamic covariance matrix of the solute calculated on the trajectories AC-(2Ca–C–CaM), AC-(0Ca–C–CaM), and AC-isolated. (a) Eigenvalues distribution colored as dark [AC-(2Ca–C–CaM)], gray [AC-(0Ca–C–CaM)], and light Gray (AC). Projections of the first mode on the system AC-(2Ca–C–CaM) (b), AC-(0Ca–C–CaM) (c), and AC-isolated (d). The complex regions are colored in red (C–CaM), purple (SA), green (CA), and orange (CB). The C terminal tail and the catalytic loop are colored in cyan and blue, respectively, and highlighted with stars of the same color.

Fig. S6(b)]. This distance increase agrees with the larger motion of the C terminal tail observed in AC-(0Ca–C–CaM) with respect to AC-(2Ca–C–CaM) [Figs. 2(b) and





**Figure 5**

Variations of the catalytic site geometry described by the distances between the centers of mass: (a)  $CM_{CA}$  and  $CM_{CB}$ , (b)  $CM_{CA}$  and  $CM_{CLoop}$ , and (c)  $CM_{CB}$  and  $CM_{CLoop}$ . The centers of mass,  $CM_{CA}$ ,  $CM_{CB}$ , and  $CM_{CLoop}$  are respectively calculated from: K58, D188, D190, and H298 located in CA, H63, and K84 located in CB, N304, and F306 located in the catalytic loop. The curves are plotted in green for AC-(2Ca-C-CaM) in red for AC-(0Ca-C-CaM) and in black AC-isolated.

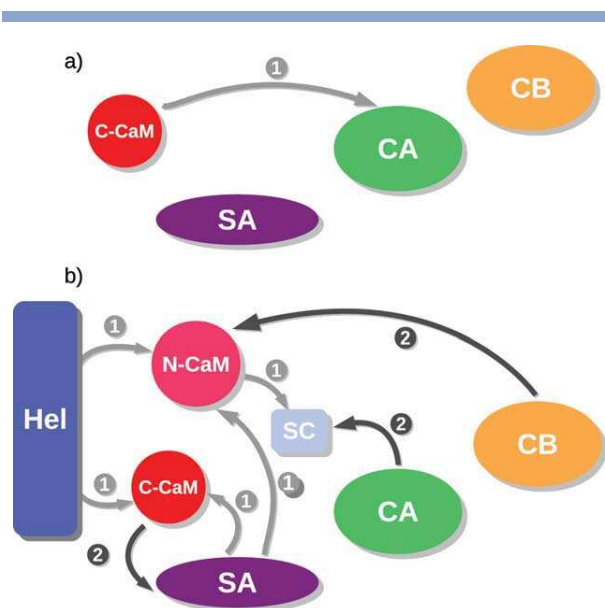
3(a)]. The distance between the centers of mass of CA and CB [Supporting Information Fig. S4(c)] is stable, in the 27–28 Å range, during the trajectory AC-(2Ca-C-CaM), increases in AC-(0Ca-C-CaM), and decreases slightly in AC-isolated. The CA and CB relative positions are thus stable in the studied systems.

To relate the conformational tendency of AC to its possible catalytic activity, the deformation of the catalytic site was analyzed by monitoring distances between the catalytic residues, taken from Figure 5(c) of Ref. 10: K58, H63, K84, D188, D190, H298, N304, and F306. Three centers of mass,  $CM_{CA}$ ,  $CM_{CB}$  and  $CM_{CLoop}$  are, respectively, calculated from: K58, D188, D190, and H298 located in CA, H63, and K84 located in CB, N304, and F306 located in the catalytic loop.

When the Ca<sup>2+</sup> ions are removed [trajectory AC-(0Ca-C-CaM)], CB slightly moves apart from CA and loop C, as the distances  $CM_{CA}/CM_{CB}$  [Fig. 5(a)] and  $CM_{CLoop}/CM_{CB}$  [Fig. 5(c)] increase, while the  $CM_{CA}/CM_{CLoop}$  distance [Fig. 5(b)] keeps constant. In AC-isolated, all distances decrease, the largest drop of 4 Å being observed between  $CM_{CA}$  and  $CM_{CB}$ . These varia-

tions of the catalytic site geometry can be related to the catalytic efficiency through the catalytic accessible surface. Indeed, small accessible surface hinders the access to the catalytic site, whereas large accessible surface facilitates substrate binding. In the MD trajectories analyzed here, the catalytic surface oscillates between medium value (527 Å<sup>2</sup>) for AC-(2Ca-C-CaM), larger value (614 Å<sup>2</sup>) for AC-(0Ca-C-CaM), and smaller value (460 Å<sup>2</sup>) for AC-isolated (Table II). The pocket geometry of AC-(0Ca-C-CaM) is thus compatible with a catalytic efficiency similar to the one of AC-(2Ca-C-CaM), in agreement with the experimental observations<sup>11</sup> showing the lack of Ca<sup>2+</sup> influence on the AC catalytic activity. Unsurprisingly, the more open surface in AC-(0Ca-C-CaM) corresponds to the largest distance between CA and CB centers of mass [Supporting Information Fig. S4(c)]. On the contrary, in AC-isolated, the geometry (Fig. 5) as well as the accessible surface of the catalytic site (Table II) reveal a collapse of the site, hindering substrate binding.

One should notice that for the Edema Factor of *B. anthracis*,<sup>19</sup> the effect of removing Ca<sup>2+</sup> ions is opposite to what is observed for AC/C-CaM, as the catalytic



**Figure 6**

Schemes of energetic influences for (a) AC-(2Ca-C-CaM), AC-(0Ca-C-CaM), and (b) the MD simulation previously recorded<sup>20</sup> on the Edema Factor/CaM complex in the presence of  $\text{Ca}^{2+}$  ions. The complex regions are drawn with geometric figures. The arrows denote the energetic influence of one domain to another, the influence intensity being marked as the number of ranks lost by influenced domain.

surface decreases from 928 to 501  $\text{\AA}^2$ . This collapse of the catalytic site in the Edema Factor is consistent with the reduction of its catalytic activity for decreasing  $\text{Ca}^{2+}$  concentration.<sup>11,46</sup>

The analysis of the global motion tendencies and of the geometry of the catalytic site of AC enables to relate the large conformational drift displayed by AC-isolated with the collapse of the catalytic site. Conversely, the absence or the presence of  $\text{Ca}^{2+}$  ions has only a limited effect on the complex global architecture and catalytic efficiency.

### C-CaM conformation and interaction with AC

As seen in the section “Conformational drift,” the conformation of the C-CaM displays only little variation during the simulations. Several parameters were used (Table II) to obtain a more detailed picture of the variation: opening of the EF-hands,<sup>47,48</sup> accessible surface of the hydrophobic patches,<sup>41</sup> coordination of the  $\text{Ca}^{2+}$  ions. The opening of the EF hands 3 and 4 of C-CaM, was determined from the angles 3 and 4 between the axes of the  $\alpha$  helices V and VI, and between the axes of the  $\alpha$  helices VII and VIII.

In AC-(2Ca-C-CaM), the angles 3 and 4 are stable around 88 and 99°, respectively, these values being only slightly larger than the values of 87 and 85° observed in

the X-ray crystallographic structure 1YRT. In AC-(0Ca-C-CaM), the EF hand 3 displays a tendency to open with an angle of about 92°. This opening agrees with the large fluctuations observed in the  $\text{Ca}^{2+}$  loop of EF hand 3 [Supporting Information Fig. S3(b)], but it is paradoxical with regards to the closure of the EF-hands normally observed in free CaM upon Calcium removal.<sup>47,48</sup> Here, the closure is hindered by the interaction of C-CaM with AC, as it was also observed in the MD simulations previously recorded on the Edema Factor from *B. anthracis*.<sup>19</sup>

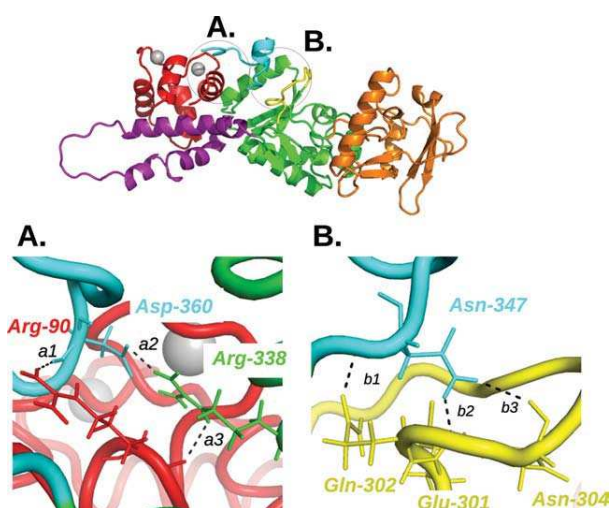
According to the literature,<sup>41</sup> the hydrophobic patch of C-CaM (I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, and A147) should be more accessible if the EF hands are more opened. It is indeed the case here as the accessible surface of hydrophobic patches (Table II) slightly increases from 764 in AC-(2Ca-C-CaM) to 785  $\text{\AA}^2$  in AC-(0Ca-C-CaM), in agreement with the opening of the angle 3 described previously. This patch surface increase is the sign of a closest interaction between C-CaM and AC. One should notice that the hydrophobic patch contains also the methionines 109, 124, 144, and 145, the oxidation of which dramatically decreases the affinity of CaM to AC.<sup>49</sup> The Met oxidation could thus hamper the variation of the patch surface and consequently induce an unbinding of C-CaM from AC. The patch surface increase is opposite to the surface decrease observed in the Edema Factor/CaM complex,<sup>19</sup> when the  $\text{Ca}^{2+}$  ions are removed (Table V of Ref. 19). Thus, for both Edema Factor and AC, the patch surface display variations in agreement with the variations of the catalytic site, described in the subsection “Globular tendency of AC and catalytic activity”: a larger patch surface corresponds to a more open catalytic site and a smaller surface to a collapsed site.

**Table II**

Analysis of the Adenyl Cyclase Catalytic Site and of the C-CaM Structure Along the Trajectories AC-(2Ca-C-CaM), AC-(0Ca-C-CaM), and AC-isolated

Trajectory	AC-(2Ca-C-CaM)	AC-(0Ca-C-CaM)	AC-isolated
Catalytic surface ( $\text{\AA}^2$ )	527.3 ± 54.4	614.7 ± 50.2	460.5 ± 91.7
EF hand angle 3 (°)	88 ± 2.5	92 ± 2.8	–
EF hand angle 4 (°)	98.7 ± 3.2	98.3 ± 4.1	–
Accessible surface of hydrophobic patch ( $\text{\AA}^2$ )	764.0 ± 24.5	785.3 ± 32.1	–
Distance EF hand 3 $\alpha$ helix H ( $\text{\AA}$ )	11.5 ± 0.4	10.5 ± 0.3	–
Distance EF hand 4 $\alpha$ helix H ( $\text{\AA}$ )	11.6 ± 0.3	12.1 ± 0.3	–

The catalytic surface is the solvent accessible surface ( $\text{\AA}^2$ ) calculated with MSMS 2.5.3<sup>54</sup> on the catalytic residues K58, H63, K84, D188, D190, H298, N304, and F306 defined in the Figure 5(c) of Ref. 10. The C-CaM structure is characterized by: (i) the solvent accessible surface ( $\text{\AA}^2$ ) of the residues I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, and A147, defining the hydrophobic patch,<sup>41</sup> (ii) the mean angles between the axes of the  $\alpha$  helices V and VI (residues 82–92 and 102–111) and of the  $\alpha$  helices VII and VIII (residues 118–128 and 138–146), defining the opening of the EF hands 3 and 4, (iii) the minimal distances ( $\text{\AA}$ ) between the centers of mass of the EF hands and the axis of helix H.



**Figure 7**

Overview of the AC-C-CaM complex structure (1YRT) drawn in cartoon with zooms A and B at the interface between C-CaM, C terminal tail, and catalytic loop. The C-CaM lobe is colored in red, the C terminal tail in cyan, the catalytic loop on yellow, the remaining part of CA in green, and the ions Ca<sup>2+</sup> are drawn as silver spheres. The residues involved in hydrogen bonds in AC-(2Ca-C-CaM) and for which the hydrogen bonds are disrupted in AC-(0Ca-C-CaM) and AC-isolated are drawn in sticks and labeled in color according to the complex domain they belong to. The corresponding hydrogen bonds are labeled *a1*, *a2*, *a3* for the interaction C-CaM/C terminal tail/CA and *b1*, *b2*, *b3* for the interaction C terminal tail/catalytic loop.

The EF hands 3 and 4 of C-CaM interact in a parallel way with the  $\alpha$  helix H (Fig. 1), the EF hand 3 being closer to the region CA and the EF hand 4 the closer to the SA extremity. Among the two EF hands, the EF hand 4 seems to have a less tight interaction with the helix. Indeed, as outlined in subsection “Conformational drifts,” the removal of Ca<sup>2+</sup> ions induces an increase of flexibility [Supporting Information Fig. S3(b)] in the C terminal part of C-CaM, included in the EF hand 4. Furthermore, in AC-(2Ca-C-CaM), the distance of EF

hand 4 to the axis of  $\alpha$  helix H oscillates in the 10–12 Å range [Supporting Information Fig. S7(b)] along the trajectory, whereas the distance involving the EF hand 3 displays a sharp transition around 7 ns [Supporting Information Fig. S7(a)]. This transition is produced by a slight rotation of the  $\alpha$  helix H of SA.

The coordination of the Ca<sup>2+</sup> ions in the C-CaM lobe (Table III) is similar to the observations previously made in the structure 1OOJ of free Ca<sup>2+</sup>-CaM (Table II of Ref. 19). The coordination of the calcium in the EF hand 3 involves the sidechains of D93, D95, N97, and E104 and the carbonyl of Y99. The coordination in the EF hand 4 involves the sidechains of D129, D131, and D133 and the carbonyls of D133 and Q135. The hydrogen bonds formed by the D131 and D133 sidechains are weaker than the others, which is the sign of a weaker coordination of Ca<sup>2+</sup> ion by the EF hand 4, and can be put in parallel with the less tight interaction of EF hand 4 with the  $\alpha$  helix H [Supporting Information Fig. S6(b)].

The C-CaM did not display much conformational variability, which is different from the observations previously made on the complex structure between CaM and the Edema Factor from *B. anthracis*.<sup>19,21</sup> The largest mobility observed on C-CaM is the opening of the EF hand 3 when the Ca<sup>2+</sup> ions are removed. This movement disagrees with the usual behavior of CaM in solution, and tends to prove that the interaction with AC makes the conformational sub-states of C-CaM different from those observed for CaM in solution. In the slight conformational variations of the C-CaM lobe, the EF hands 3 and 4 do not share the same features. The EF hand 3, closer to the C terminal tail and the region CA, binds more tightly the Ca<sup>2+</sup> ion, and is more sensitive to the Ca<sup>2+</sup> removal.

### Energetic influences

The energetic influences between the regions SA, CA, CB, and C-CaM were calculated from the different trajectories as described in Materials and Methods. No energetic

**Table III**

Coordination of the Ca<sup>2+</sup> Ions in AC-(2Ca-C-CaM)

EF-hand	Atom	Mean distance (Å)	Minimum (Å)	Maximum (Å)
3	Asp-93 O $\delta$ 1/O $\delta$ 2	2.6 $\pm$ 0.1	2.3	3.1
3	Asp-95 O $\delta$ 1/O $\delta$ 2	2.6 $\pm$ 0.1	2.3	3.0
3	Asn-97 O $\delta$ 1	2.7 $\pm$ 0.1	2.4	3.5
3	Tyr-99 O	2.6 $\pm$ 0.1	2.4	3.3
3	Glu-104 O $\epsilon$ 1/O $\epsilon$ 2	2.6 $\pm$ 0.1	2.4	2.9
4	Asp-129 O $\delta$ 2	2.6 $\pm$ 0.1	2.3	3.2
4	Asp-131 O $\delta$ 1/O $\delta$ 2	2.6 $\pm$ 0.1	2.3	3.0
4	Asn-133 O $\delta$ 1	3.1 $\pm$ 1.1	2.3	7.4
4	Asp-133 O	2.9 $\pm$ 0.5	2.4	6.1
4	Gln-135 O	2.7 $\pm$ 0.1	2.4	3.5

The atoms of the calcium loop closer than 2.8 Å from the ion are detected. For distances involving atom couples O $\delta$ 1/O $\delta$ 2 and O $\epsilon$ 1/O $\epsilon$ 2, the minimum of the two distances is monitored. For each coordination interaction, the mean distance and standard deviation are given, as well as the distance minimum and maximum values along the trajectory.

**Table IV**

Formation Percentages of Hydrogen Bonds Connecting C-CaM, the Region CA, the C Terminal Tail and the Catalytic Loop

Hydrogen bond	Partners	AC-(2Ca-C-CaM) (%)	AC-(0Ca-C-CaM) (%)	AC-isolated (%)
<i>a1</i>	D360-O/R90-H $\eta$ 21	44.9	40.8	–
<i>a2</i>	R338-H $\eta$ 12/D360-O $\delta$ 2	44.9	40.8	32.0
<i>a3</i>	R338-He/R90-O	84.2	13.8	–
<i>b1</i>	Q302-O $\epsilon$ 1/N347-H	92.3	0.1	81.5
<i>b2</i>	E301-O/N347-H $\delta$ 22	98.8	0.0	21.8
<i>b3</i>	N304-O/N347-H $\delta$ 21	98.9	3.8	0.9

The hydrogen bond label drawn in Figure 7, the hydrogen bond partners and the percentage of time along the MD trajectories, when the distance was smaller than 2.5 Å are given.

influence was observed during the trajectories AC-isolated, whereas a single influence of C-CaM on CA was observed in the analysis of AC-(2Ca-C-CaM) [Fig. 6(a)] and AC-(0Ca-C-CaM). Similar variations in MMPBSA energy were observed for AC-(2Ca-C-CaM) and AC-(0Ca-C-CaM). In the full complex *C*, the regions SA and C-CaM have favorable reference interaction energies  $\Delta\Delta G[D, C]$ , whereas CA and CB display unfavorable reference interaction energies. Nevertheless, CA is ranked at a better place than CB, as shown by the non-parametric Mann-Whitney test<sup>37</sup> [Supporting Information Tables S2 and S3]. In the complex AC-CaM, all regions display unfavorable interaction energies  $\Delta\Delta G[D, AC-CaM]$ . The different components of CB interaction energy are similar in *C* and AC-CaM, whereas the absolute values of SA components decrease, the solvation energy being divided by 2. Despite the decrease of energy components, SA keeps the best rank in the interaction list, whereas the absolute values of the polar solvation energy  $\Delta\Delta G_{sol}^{polar}$  and of the electrostatic energy  $\Delta\Delta G_{ele}$  of CA are increased. This is responsible for CA losing one rank (Supporting Information Tables S2 and S3). The reaction of the different regions to the removal of Ca<sup>2+</sup> ions agrees with their relative positions with respect to C-CaM: CB is located at the opposite complex side whereas CA and SA are in direct interaction with C-CaM.

The single influence observed in AC-(2Ca-C-CaM) and AC-(0Ca-C-CaM) along with the much larger  $\Delta\Delta G_{sol}^{polar}$  and  $\Delta\Delta G_{ele}$  of the binding energies supports the essential role of Ca<sup>2+</sup> ions and of C-CaM in the full AC activation. The loose of one CA rank is due in both cases to the large variation of the  $\Delta\Delta G_{sol}^{polar}$  and  $\Delta\Delta G_{ele}$  energies, but  $\Delta\Delta G_{ele}$  vary more in the case of AC-(2Ca-C-CaM) than in the case of AC-(0Ca-C-CaM) (Supporting Information Tables S2 and S3). This effect can be put in parallel with the disruption of hydrogen bonds between residues of the C terminal tail (N347, D360), CA (R338), catalytic loop (E301, Q302, N304), and C-CaM (R90) (Fig. 7). These hydrogen bonds are present in AC-(2Ca-C-CaM), still partially present in AC-(0Ca-C-CaM), and mostly disrupted in AC-isolated (Table IV). The disruption of these hydrogen bonds agrees with the distance increase between the centers of mass of the C terminal tail and of the C-CaM [Supporting Information

Fig. S5(b)]. The hydrogen bonds formed between R90 of C-CaM and R326, N336 of AC involving all the carboxyl and the amide groups of these sidechains (data not shown), is reminiscent of the salt bridge observed between K525 of Edema Factor and E114 of CaM.<sup>9</sup>

The absence of energetic influences when C-CaM and Ca<sup>2+</sup> ions are removed, agrees with the experimentally observed destabilization of the isolated CyaA. Indeed, the isolated protein seems difficult to crystallize and was shown to partially aggregate in solution (Alexandre Chenal, personal communication). The enthalpic contributions, considered in the present analysis, should consequently play a more limited role in the energetic influences.

The energetic influence scheme was previously<sup>20</sup> interpreted as giving a description of the stability dependence scheme between the different parts of the complex, as in a house-of-cards architecture: the more a region is influenced by others, the lower the place it occupies in the hierarchy of the relative stabilities inside the complex. In this frame, the single influence of C-CaM to CA observed for AC-(2Ca-C-CaM) and AC-(0Ca-C-CaM) demonstrates that CA occupies a lower part in the hierarchy of stabilities of the C-CaM/AC regions, and is stabilized by C-CaM. The CA region is thus the best candidate for modifications aiming to modulate the C-CaM/AC interaction in the presence and absence of Ca<sup>2+</sup> ions. For AC-isolated, no energetic influence and consequently no stability hierarchy is observed. The corresponding diagram of energy influences previously obtained on the Edema Factor/CaM complex<sup>20</sup> [Fig. 6(b)] was much more intricate than the one observed on the AC/C-CaM complex [Fig. 6(a)]. This can be related to the higher energy barrier for Edema Factor/CaM interaction with respect to AC/CaM interaction, this barrier inducing a hierarchy of region stabilities. Indeed, the affinity of the Edema Factor for CaM is about 20 nM<sup>11</sup> whereas the affinity of AC to CaM is about 0.2 nM.<sup>10,50</sup>

## DISCUSSION

The catalytic domain of the AC CyaA from *B. pertussis* was simulated in complex with Ca<sup>2+</sup>-loaded and apo C-CaM, and also as an isolated protein. The choice of



considering only a portion of CaM, namely C-CaM, in our simulations was motivated by the fact that the AC/C-CaM interaction was captured by X-ray crystallography<sup>10</sup> and also by experimental evidence showing that C-CaM is able to produce the full AC activation.<sup>11</sup>

The initial motivation for studying the AC domain by MD simulations, was the lack of precise structural information on the conformational equilibrium of this protein isolated in solution. Indeed, as no X-ray crystallographic structure of free AC is available and only hydrodynamics measurements<sup>23</sup> reveal that the isolated protein is more spherical and hydrated than the AC/CaM complex.

The purpose of this study is extremely ambitious, as the two interaction partners undergo major conformational changes, when the interaction is disrupted. Thus, one cannot expect that any MD trajectory recorded on a 10th nanosecond timescale could reach the equilibrium. Indeed, the apo C-CaM does not dissociate from the AC domain, and the isolated AC domain does not reach the spherical shape observed by hydrodynamics measurements.<sup>23</sup>

Nevertheless, the relatively short length of the MD simulations presented here do not hamper their relevance to analyze the reaction of the AC domain according to different perturbations: removal of ions Ca<sup>2+</sup> or removal of C-CaM and of ions Ca<sup>2+</sup>. Indeed, if one is interested in characterizing the relative conformational tendencies and not the absolute conformational drift of the AC domain, MD simulations of a few tens of nanoseconds performed in three different conditions can give valuable information. Moreover, these three simulations were repeated and gave similar results.

The study presented here allowed to obtain more precise information on this point, as well as on the structural features of the AC activation. First, the conformational preferences displayed by AC during the trajectories are in qualitative agreement with the observations of Karst et al.,<sup>23</sup> as the global structure evolution tends to a more globular shape. However, because of the limited time interval explored during the MD simulations, only the beginning of the transition toward a spherical hydrated form can be sampled. The convergence analysis using the Lyman and Zuckerman approach,<sup>33</sup> applied on the 15–30 ns range of AC-isolated, showed a tendency of AC-isolated to stay in one conformational basin. This means that MD simulation reached a bottleneck in the exploration of the AC conformational space, and that methods more adapted to extensive conformational space exploration<sup>51–53</sup> should be used for overcoming conformational barriers.

Second, the MD simulations gave more precise information on the relative role of Ca<sup>2+</sup> ions and C-CaM in the AC activation. The experimental information available prior to the MD simulations, was that neither Ca<sup>2+</sup> ions, nor the full CaM is essential for activation.<sup>11</sup> From the MD simulation analysis, one can deduce that the C-CaM/AC interaction displays two different aspects. The

first aspect is that the interaction of C-CaM with the SA region is essentially based on a steric effect, preventing SA to get closer to the other AC extremity. Indeed, in the absence of C-CaM, the motion of SA induces a large destabilization of CA which closes the catalytic site. The second aspect is that the Ca<sup>2+</sup> ions have only a weak direct effect on the catalytic site, as the site slightly opens in the absence of ions, which agrees with the observation that AC can still be activated in the absence of calcium.<sup>11</sup> However, in agreement with the experimental observations,<sup>11</sup> the Ca<sup>2+</sup> ions facilitate the interaction between C-CaM and AC. The influence of C-CaM and Ca<sup>+</sup> ions is mediated by the interactions between C-CaM and the C terminal tail, as shown by: (i) the energetic influences diagram, (ii) the correlation of hydrogen bonds in C-CaM, C terminal tail, and CA, with the presence of C-CaM and Ca<sup>2+</sup> ions, and (iii) the increased conformational drifts of the C terminal tail and the C loop in the absence of ions, and (iv) the distances between centers of mass of C-CaM and C terminal tail. The importance of the C terminal tail and of the catalytic loop agrees with the equivalence established by Guo et al.,<sup>10</sup> between these regions and the switches C and B which undergo large conformational changes during the Edema Factor activation.<sup>9</sup>

The AC domain of the AC from *B. pertussis* shares several common features with the Edema Factor from *B. anthracis*, in particular concerning its activation via an interaction with CaM. Nevertheless, the determination of the two structures of AC<sup>10</sup> and Edema Factor<sup>9</sup> revealed quite different structural organization, in agreement with the relatively low sequence identity of 25% between AC and Edema Factor. The present analysis of MD of the AC domain in the presence and in the absence of C-CaM and of Ca<sup>2+</sup> ions confirms that the two ACs behave quite differently. First, the removal of Ca<sup>2+</sup> ions opens the catalytic site of AC whereas it closes the catalytic site of Edema Factor. Besides, in the Edema Factor structure, the switch C, equivalent to the C terminal tail of AC, is covalently attached to the helical domain of about 150 residues, which has the role to lock Edema Factor in an inactive state.<sup>11</sup> In AC, the C tail is more free to move away. This explains the smaller barrier of activation of AC with respect to Edema Factor. The CaM/Ctail interaction is thus a key point of the activation and was pointed out by the energetic influences as well as an hydrogen bond network.

## CONCLUSION

The catalytic domain AC of the AC CyaA from *B. pertussis* was simulated in different interaction environments. The interaction between AC and CaM presents two major aspects. First, the CaM blocks the SA region of AC to keep the geometry of the catalytic site in an activated

form by stabilizing the CA region. Second, the  $\text{Ca}^{2+}$  ions have an indirect role in the activation, stabilizing the C-CaM/AC interaction by inducing the formation of an hydrogen bond network spanning from C-CaM to the catalytic loop through the C terminal tail. This description qualitatively agrees with the experimental kinetics<sup>11</sup> and hydrodynamics<sup>23</sup> information previously available on this system, and puts in evidence the differential roles of CaM and  $\text{Ca}^{2+}$  ions in the AC domain activation.

## REFERENCES

- Gray MC, Hewlett EL. Cell cycle arrest induced by the bacterial adenylate cyclase toxins from *Bacillus anthracis* and *Bordetella pertussis*. *Cell Microbiol* 2011;13:123–134.
- Göttle M, Dove S, Kees E, Schlossmann J, Geduhn J, König B, Shen Y, Tang WJ, Kaever V, Seifert R. Cytidylyl and uridylyl cyclase activity of bacillus anthracis edema factor and *Bordetella pertussis* CyaA. *Biochemistry* 2010;49:5494–5503.
- Martin C, Gómez-Bilbao G, Ostolaza H. *Bordetella* adenylate cyclase toxin promotes calcium entry into both CD11b+ and CD11b- cells through cAMP-dependent L-type-like calcium channels. *J Biol Chem* 2010;285:357–364.
- Paccani SR, Dal Molin F, Benagiano M, Ladant D, D'Elios MM, Montecucco C, Baldari CT. Suppression of T-lymphocyte activation and chemotaxis by the adenylate cyclase toxin of *Bordetella pertussis*. *Infect Immun* 2008;76:2822–2832.
- Kamanova J, Kofronova O, Masin J, Genth H, Vojtova J, Linhartova I, Benada O, Just I, Sebo P. Adenylate cyclase toxin subverts phagocyte function by RhoA inhibition and unproductive ruffling. *J Immunol* 2008;181:5587–5597.
- Kim C, Wilcox-Adelman S, Sano Y, Tang WJ, Collier RJ, Park JM. Antiinflammatory cAMP signaling and cell migration genes co-opted by the anthrax bacillus. *Proc Natl Acad Sci USA* 2008;105:6150–6155.
- Puhar A, Dal Molin F, Horvath S, Ladant D, Montecucco C. Anthrax edema toxin modulates PKA- and CREB-dependent signaling in two phases. *PLoS ONE* 2008;3:e3564.
- Raymond B, Leduc D, Ravaux L, Goffic RL, Candela T, Raymond-jean M, Goossens PL, Touqui L. Edema Toxin impairs anthracidal phospholipase A2 expression by alveolar macrophages. *PLoS Pathog* 2007;3:e187.
- 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 adenylate cyclase exotoxin by calmodulin. *Nature* 2002;415:396–402.
- Guo Q, Shen Y, Lee YS, Gibbs CS, Mrksich M, Tang WJ. Structural basis for the interaction of *Bordetella pertussis* adenylate cyclase toxin with calmodulin. *EMBO J* 2005;24:3190–3201.
- Shen Y, Lee YS, Soelaiman S, Bergson P, Lu D, Chen A, Beekingham K, Grabarek Z, Mrksich M, Tang WJ. Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylate cyclase exotoxins. *EMBO J* 2002;21:6721–6732.
- Drum CL, Shen Y, Rice PA, Bohm A, Tang WJ. Crystallization and preliminary X-ray study of the edema factor exotoxin adenylate cyclase domain from *Bacillus anthracis* in the presence of its activator, calmodulin. *Acta Crystallogr D Biol Crystallogr* 2001;57:1881–1884.
- Guo Q, Shen Y, Zhukovskaya NL, Florian J, Tang WJ. Structural and kinetic analyses of the interaction of anthrax adenylate cyclase toxin with reaction products cAMP and pyrophosphate. *J Biol Chem* 2004;279:29427–29435.
- Shen Y, Guo Q, Zhukovskaya NL, Drum CL, Bohm A, Tang WJ. Structure of anthrax edema factor-calmodulin-adenosine 5'-(alpha,-beta-methylene)-triphosphate complex reveals an alternative mode of ATP binding to the catalytic site. *Biochem Biophys Res Commun* 2004;317:309–314.
- 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.
- Martinez L, Laine E, Malliavin TE, Nilges M, Blondel A. ATP conformations and ion binding modes in the active site of anthrax edema factor: a computational analysis. *Proteins* 2009;77:971–983.
- Martinez L, Malliavin TE, Blondel A. Mechanism of reactant and product dissociation from the anthrax edema factor: a locally enhanced sampling and steered molecular dynamics study. *Proteins* 2011;79:1649–1461.
- Laine E, Goncalves C, Karst J, Lesnard A, Rault S, Tang WJ, Malliavin TE, Ladant D, Blondel A. Use of allostery to identify inhibitors of calmodulin-induced activation of *Bacillus anthracis* edema factor. *Proc Natl Acad Sci USA* 2010;107:11277–11282.
- Laine E, Yoneda JD, Blondel A, Malliavin TE. The conformational plasticity of calmodulin upon calcium complexation gives a model of its interaction with the oedema factor of *Bacillus anthracis*. *Proteins* 2008;71:1813–1829.
- Laine E, Blondel A, Malliavin TE. Dynamics and energetics: a consensus analysis of the impact of calcium on EF-CaM protein complex. *Biophys J* 2009;96:1249–1263.
- Laine E, Martínez L, Blondel A, Malliavin TE. Activation of the edema factor of *Bacillus anthracis* by calmodulin: evidence of an interplay between the EF-calmodulin interaction and calcium binding. *Biophys J* 2010;99:2264–2272.
- Guo Q, Jureller JE, Warren JT, Solomaha E, Florián J, Tang WJ. Protein-protein docking and analysis reveal that two homologous bacterial adenylate cyclase toxins interact with calmodulin differently. *J Biol Chem* 2008;283:23836–23845.
- Karst JC, Sotomayor Pérez AC, Guijarro JI, Raynal B, Chenal A, Ladant D. Calmodulin-induced conformational and hydrodynamic changes in the catalytic domain of *Bordetella pertussis* adenylate cyclase toxin. *Biochemistry* 2010;49:318–328.
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* 2006;Chapter 5:Unit 5.6.
- Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods R. The Amber biomolecular simulation programs. *J Comput Chem* 2005;26:1668–1688.
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 2006;65:712–725.
- Jorgensen WL. Transferable intermolecular potential functions for water, alcohols and ethers. application to liquid water. *J Am Chem Soc* 1981;103:335–340.
- Aqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. *J Phys Chem* 1990;94:8021–8024.
- Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA. AMBER 9. San Francisco: University of California; 2006.
- Darden T, York D, Pedersen L. Partial Mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems. *J Chem Phys* 1993;98:3684–3690.
- 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.
- 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.
- Lyman E, Zuckerman DM. Ensemble-based convergence analysis of biomolecular trajectories. *Biophys J* 2006;91:164–172.

34. Hamacher K, Trylska J, McCammon J. Dependency map of proteins in the small ribosomal subunit. *PLoS Comput Biol* 2006;2:e10.
35. Srinivasan J, Cheatham T, Cieplak J, Kollman P, Case D. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J Am Chem Soc* 1998;120:9401–9409.
36. Tan C, Tan YH, Luo R. Implicit nonpolar solvent models. *J Phys Chem B* 2007;111:12263–12274.
37. Hollander M, Wolfe DA. *Nonparametric Statistical Methods*. New York: Wiley; 1973.
38. R Development Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2009.
39. Wriggers W, Mehler E, Pitici F, Weinstein H, Schulten K. Structure and dynamics of calmodulin in solution. *Biophys J* 1998;74:1622–1639.
40. Yang C, Jas GS, Kuczera K. Structure and dynamics of calcium-activated calmodulin in solution. *J Biomol Struct Dyn* 2001;19:247–271.
41. Yang C, Jas GS, Kuczera K. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *Biochim Biophys Acta* 2004;1697:289–300.
42. Shepherd CM, Vogel HJ. A molecular dynamics study of Ca<sup>2+</sup>-calmodulin: evidence of interdomain coupling and structural collapse on the nanosecond timescale. *Biophys J* 2004;87:780–791.
43. Yang C, Kuczera K. Molecular dynamics simulations of calcium-free calmodulin in solution. *J Biomol Struct Dyn* 2002;19:801–819.
44. Komeiji Y, Ueno Y, Uebayasi M. Molecular dynamics simulations revealed Ca(2+)-dependent conformational change of Calmodulin. *FEBS Lett* 2002;521:133–139.
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. 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.
47. 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.
48. 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.
49. Vougie S, Mary J, Dautin N, Vinh J, Friguet B, Ladant D. 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* 2004;279:30210–30218.
50. Glaser P, Elmaoglou-Lazaridou A, Krin E, Ladant D, Bárzu O, Danchin A. Identification of residues essential for catalysis and binding of calmodulin in Bordetella pertussis adenylate cyclase by site-directed mutagenesis. *EMBO J* 1989;8:967–972.
51. Tribello GA, Ceriotti M, Parrinello M. A self-learning algorithm for biased molecular dynamics. *Proc Natl Acad Sci USA* 2010;107:17509–17514.
52. Maragliano L, Cottone G, Ciccotti G, Vanden-Eijnden E. Mapping the network of pathways of co diffusion in myoglobin. *J Am Chem Soc* 2010;132:1010–1017.
53. Abrams CF, Vanden-Eijnden E. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc Natl Acad Sci USA* 2010;107:4961–4966.
54. Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 1996;38:305–320.