Analysis of the mechanism of action of inhibitors targeting the allosteric activation of edema factor of Bacillus anthracis

Irène Pitard

Sorbonne Université

Direction Général de l'Armement / Institut Pasteur

École doctorale complexité du vivant

Bioinformatique structurale/ Dynamique fonctionnelle, UMR3528

Analyse du mécanisme d'action d'inhibiteurs ciblant l'activation allostérique du facteur œdématogène de *Bacillus anthracis*

Par Irène Pitard

Thèse de doctorat de Biochimie et Biologie Structurale

Dirigée par Thérèse Malliavin et Pierre L. Goossens

Présentée et soutenue publiquement le 23 Octobre 2020

Devant un jury composé de:

Président du jury: Prof. Olivier Lequin, Sorbonne Université

Rapporteur: Dr. Sophie Zinn-Justin, CEA Paris-Saclay

Rapporteur: Prof. Dr. med. Roland Seifert, Medizinische Hochschule Hannover

Examinateur: M.C.F. Elodie Laine, Sorbonne Université

Examinateur: Dr. Fabrice Biot, Institut de Recherche Biomédicale des Armées

Directrice de thèse: Dr. Thérèse Malliavin, Institut Pasteur

Co-directeur de thèse: Dr. Pierre L. Goossens, Institut Pasteur

Acknowledgments

I deeply thank the members of the jury, who agreed to judge my thesis work in particularity Prof. Olivier Lequin for accepting to be the chairmanship of this presented work. To Dr. Sophie Zinn-Justin and Prof. Dr. Roland Seifert, I address my thanks for the honor they do me by agreeing to judge this work.

I want to express my thanks to Dr. Michael Nilges for welcoming me into the Structural Bioinformatic Unit of at Institut Pasteur and for encouraging me to present my work all around the world.

I would particularly like to thank Dr. Thérèse Malliavin who directed this thesis. She was kind enough to accept me while I was abroad doing my engineer internship at the Massachusetts Institut of Technology (Cambridge, USA). I want to thank you for your availability and your trust. I would like to thanks Dr. Pierre Goossens for being my co-director. Thanks to both of you to give me the opportunity to discover a new field for me: Structural Biology. The thesis project was a challenging wonderful experience.

A special acknowledgement is for Dr. Iñaki Guijarro for proving nonstop input, encouragement and guidance during this Ph.D. work.

I would like to deeply express my appreciation to Dr. Daniel Ladant, Dr. Ahmed Haouz, Dr. Jacques Bellalou and Dr. Stéphane Pêtres, who, during lively discussions, have always taken a critical and constructive look at this work. I am very grateful to you for kindly allowing me to follow experiments on your platforms and your laboratories. Thank you for your faith, your continuous support and constant feedback.

I also extend my thanks to Marilyne Davi, Catherine Simenel, Mireille Nowakowski, Dr. Ariel Mechaly, Dr. Rémy Lemeur, Christophe Thomas, Dr. Guillaume Bouvier, Dr. Massimiliano Bonomi and Dr. Sébastien Brier for their invaluable technical assistance and their support. It was a great pleasure to work with you.

I would like to sincerely thanks Dr. Peggy Suzanne and Dr. Patrick Dallemagne from Centre d'Etudes et de Recherches sur le Médicament de Normandie, without whom, the molecules would only have remained virtual.

I would like to thanks my DGA tutor, Dr. Fabrice Biot, to follow my work and your availability for my thesis comity and the discussions. I profoundly want to thank all the members of my thesis comity Prof. Catherine Vénien-Bryan and Dr. Alexandre de Brevern to encourage me during these three years.

I thank the Structural Bioinformatics Unit and all the team and platforms who I worked with, for their sympathy and for always listening to me. I would like to address many thanks to Maya Um for her support and her logistic help and Dr. Tru Huynh and Dr. Bruno Vitorge for technical help. I would like to thank Dr. Muriel Delepierre for her priceless advice. I would like to thanks Dr. Arnaud Blondel for the long-appreciated discussions. I also sincerely acknowledge Dr. Nadia Izadi-Pruneyre and Dr. Alexandre Chenal for believing in my abilities and for the valuable comments.

There are not strong enough words to express my gratitude to everyone that contributes to this Ph.D. work on both professional and personal level.

Thanks also to everyone that I did not mention, your support was evident to me.

Résumé court

Le facteur œdémateux (EF), une toxine majeure de *Bacillus anthracis*, est activé par la calmoduline de l'hôte (CaM) pour produire des concentrations supra physiologiques d'AMP cyclique (AMPc) conduisant à une perturbation des voies de signalisation. L'interaction EF-CaM induit des changements conformationnels dans une région switch allostérique de EF conduisant à la formation du site catalytique fonctionnel. Des études antérieures *in silico* ciblant cette région switch, complétées par des données expérimentales, ont montré que les uréidoacides thiophènes (TUA) inhibent l'activité enzymatique de EF. Cependant, les connaissances sur le site de liaison et sur l'interaction étaient manquantes. Nous présentons ici une étude de l'interaction du TUA-diCl, le composé le plus actif, avec les protéines EF, CaM et le complexe EF-CaM à l'aide d'essais biochimiques couplés à des méthodes biophysiques et de modélisations moléculaires. Le TUA-diCl interagit avec EF isolé, le complexe EF-CaM et de manière inattendue avec CaM. L'étude du site de liaison entre le composé TUA-diCl et la protéine CaM par RMN indique que le composé se lie aux patchs hydrophobes de CaM qui deviennent accessibles lorsque la CaM est complexée par les ions calciums. Ceci entraîne un compactage de la structure de CaM et des changements de la dynamique interne de la protéine. Les données enzymatiques, de fluorescence et de RMN montrent que l'inhibition d'EF est due à l'interaction du composé sur EF et ne dépend pas de la présence de CaM. Des expériences de compétition entre le TUA-diCl et l'inhibiteur du site catalytique EF 2'-MANT-3'-dATP, indiquent que TUA-diCl est un inhibiteur allostérique de EF. Les expériences HDX-MS ont révélé que le TUA-diCl se lierait au domaine hélicoïdal de EF, une région critique pour l'insertion de CaM. De plus, des approches *in silico* ont mis en évidence plusieurs sites de liaison possible dans le domaine hélicoïdal. Par conséquent, TUA-diCl représente une nouvelle classe d'inhibiteurs de EF avec un mécanisme d'action allostérique et ouvrant la voie vers la conception de molécules thérapeutiques innovantes.

Mots clés : Anthrax, *Bacillus cereus*, facteur œdémateux, calmoduline, inhibition allostérique, biologie structurale, RMN, dynamique moléculaire, rayon X, HDX-MS

Abstract

Edema factor (EF), a major *Bacillus anthracis* toxin, is activated by host calmodulin (CaM) to produce supraphysiological concentrations of cyclic AMP (cAMP) thus perturbing intracellular signaling. The EF-CaM interaction induces conformational changes in an allosteric switch region of EF that lead to the formation of the catalytic site. Previous *in silico* studies targeting this switch region, complemented with experimental data, showed that thiophen ureidoacids (TUA) inhibit the enzyme catalytic activity. However, knowledge of the binding site and inhibition mode of TUA compounds are still lacking. Here, we characterize the interaction of the most active TUA compound (TUA-diCl) with EF, CaM and EF-CaM using biochemical assays coupled to biophysical methods and molecular modeling. We show that TUA-diCl interacts with EF, EF-CaM and unexpectedly with CaM. Mapping of the binding site by NMR, showed that TUA-diCl binds to the exposed hydrophobic patches of calcium loaded CaM, causing the compaction and changes in internal dynamics of the protein. Importantly, enzymatic, fluorescence and NMR data show that EF inhibition is due to the interaction of the compound with EF and is CaM-independent. Furthermore, competition experiments between TUA-diCl and the EF catalytic-site inhibitor 2'-MANT-3'-dATP, indicate that TUA-diCl is an allosteric inhibitor of EF. HDX-MS identifies a putative binding site of TUA-diCl on the helical domain of EF, a critical region for CaM insertion. Several possible binding pockets in the helical domain are analyzed *in silico*. TUA-diCl represents a new class of EF inhibitors with an allosteric mechanism, opening the way towards the design of innovative therapeutic compounds.

Keywords: Anthrax, *Bacillus cereus*, Edema factor, calmodulin, allosteric inhibition, structural biology, NMR, molecular dynamics, X-ray, HDX-MS

Résumé long

Introduction

Mon travail de thèse a été consacré à l'étude de l'un des principaux facteurs de virulence de *Bacillus anthracis*, l'agent causal de l'anthrax ou maladie du charbon. *B. anthracis* est une bactérie Gram positive produisant trois facteurs de virulence : une capsule et deux toxines. Les toxines résultent de l'assemblage de l'antigène protecteur (PA) avec le facteur œdémateux (EF) ou le facteur létal (LF). PA s'oligomérise puis se fixe aux récepteurs cellulaires de la toxine de l'anthrax de l'hôte. Le complexe est ensuite internalisé par endocytose. Puis, les enzymes EF and LF sont transloquées dans le cytosol par la formation du pore PA. EF est activée par la calmoduline (CaM) intracellulaire et catalyse de manière non régulée la formation d'AMP cyclique conduisant à des perturbations des voies de signalisations cellulaires de l'hôte. Par ailleurs, EF semble être la principale cause responsable des formes cutanées et systémiques de l'anthrax.

Les approches thérapeutiques développées contre la maladie de charbon ciblent les différentes étapes d'intoxication : l'entrée des protéines dans les cellules hôtes, l'activation de la protéine ou de l'activité catalytique. Dans le travail présenté ici, la stratégie utilisée pour inhiber EF est le développement de ligands allostériques.

Au cours d'une précédente thèse financée par la DGA, des inhibiteurs TUA (composés de type « thiophène ureio-acide » ont été identifiés *in silico* et validés expérimentalement en ciblant une poche allostérique sur EF nommée SABC. Le composé le plus actif est le 5- (3 ', 4'-dichlorophényl) -3 - [(N-phénylamino)acide carbonylamino] -thiophen-2-carboxylique (TUA-diCl). Cependant, aucune donnée expérimentale concernant leurs sites de liaison sur la protéine EF et leurs mécanismes d'inhibition n'étaient disponibles. La finalité de ce projet de thèse était donc de mieux comprendre le mode d'inhibition et de déterminer le ou les sites d'interaction des inhibiteurs TUA par des méthodes expérimentales et des approches *in silico*.

Chapitre I. Propriétés biophysiques et biochimiques du système d'étude

La découverte de nouveaux médicaments peut être menée par criblage à haut débit de librairies de molécules chimiques. Au cours de ce processus les molécules trouvées peuvent avoir tendance à s'agréger. En s'auto-associant ces molécules peuvent interagir avec de nombreux partenaires au sein de la cellule et de manière non-spécifique. Le composé TUA-diCl a été trouvé par criblage virtuel sur une poche allostérique détectée en modélisant le chemin de transition entre la forme inactive et la forme active de EF. Il est important de connaître ses propriétés intrinsèques afin de mieux appréhender son mode d'action.

Nous avons d'abord étudié les propriétés hydrophobes de la molécule TUA-diCl. En effet, nous avons montré que le TUA-diCl avait tendance à s'auto-agréger. Les cycles thiophène et dichloro-aryle étaient les plus impactés. Ce dernier pouvait être séquestré dans des micelles de détergent réduisant probablement son effet inhibiteur et sa disponibilité pour interagir avec le complexe EF-CaM. Bien qu'une petite agrégation de ce composé soit clairement un inconvénient pour cibler spécifiquement une protéine, nous avons mis au point des conditions dans lesquelles le composé ne s'associe pas de manière significative. Par ailleurs, nous avons constaté que les ions calcium nécessaires à l'activation de l'EF par CaM étaient défavorables à l'agrégation et surtout, nous avons montré que le composé n'inhibait pas de manière non spécifique EF à travers ses agrégats. Trouver des conditions dans lesquelles le composé ne s'agrège pas et inhibe EF ouvre la voie à la compréhension de son mode d'inhibition avec EF.

La deuxième partie du système d'étude concerne le matériel biologique. *B. anthracis* et *B. cereus* appartiennent à la famille *B. cereus* mais seul *B. anthracis* est une souche pathogène de classe II. Pour des questions de sécurité, l'étude suivante se concentre sur l'adénylate cyclase produite par des souches pathogènes de *B. cereus.* La protéine EF_3 correspondant au domaine catalytique responsable de l'activité d'adénylate cyclase (résidus 291-800) a été produite en grande quantité et son protocole de production a été optimisé. La production de la protéine EF_3 a été ensuite transposée à grande échelle (fermenteurs de 4 L). Le milieu et l'induction de l'expression ont été adaptés. Ceci permet d'obtenir un large stock homogène de biomasse et de le purifier en utilisant les étapes d'affinité au nickel et de filtration sur gel en routine, utiles pour les essais de cristallogenèse. La protéine CaM a été également produite et purifiée. Les protéines EF et CaM ont été ensuite analysées par différentes techniques biophysiques afin de contrôler leur taille, leur stabilité et leur homogénéité. Des tests enzymatiques ont été effectués

afin de vérifier si les enzymes étaient fonctionnelles, c'est-à-dire capable de catalyser la formation d'AMPc à partir de l'ATP. L'activité de EF_3 mesurée (k_{cat}) en présence de 5% (v/v) de DMSO est comprise entre 1000 et 2000 molécules de AMPc produites par seconde comme attendue.

Chapitre II. Interactions moléculaires entre l'inhibiteur TUA-diCl et le facteur œdèmateux, ainsi que son activateur calmoduline

A ce jour, il existe peu d'inhibiteurs allostériques de EF. Le mécanisme du composé TUA pensé à l'origine était de verrouiller EF dans son état inactif et empêcher ainsi l'insertion de CaM dans EF. Cibler l'association des deux protéines EF et CaM peut permettre de trouver des inhibiteurs allostériques de l'interaction protéine-protéine (IPPI). Trouver des inhibiteurs des IPPI est difficile mais offre de nombreuses possibilités d'approches thérapeutiques. Jusqu'à présent, un seul composé, le 10506-2A, s'est révélé agir comme IPPI envers EF-CaM.

Nous avons démontré par les méthodes de biochimie reposant sur les tests d'activité que le mécanisme d'inhibition du TUA-diCl envers EF n'est pas CaM dépendant. En effet, TUA-diCl est capable d'inhiber l'activité d'adénylate cyclase en absence de calcium. De plus, l'augmentation de la concentration de CaM n'influe pas sur inhibition du TUA-diCl.

Les expériences RMN basées sur l'observation du ligand (STD et NOE transférés) mettant en compétition un ligand du site catalytique, 2'-MANT-3'dATP et le TUA-diCl ont montré que les TUA ne se lient pas au niveau du site actif de EF_3. Les TUA sont donc des inhibiteurs allostériques.

Nous avons aussi établi par STD que les composés TUA interagissent avec EF_3, le complexe EF_3-CaM et, de façon inattendue, avec CaM. Nous avons confirmé que TUA-diCl est plus affin que TUA-Cl pour les trois systèmes et nous avons décidé de centrer notre étude sur le composé le plus actif. Enfin, nous avons établi que tous les protons des molécules TUA sont impliqués dans l'interaction avec EF_3, EF_3-CaM et CaM. Afin de connaître la structure tertiaire du ligand TUA-diCl interagissant avec EF ou EF-CaM, nous avons utilisé en parallèle l'amarrage moléculaire assisté par ordinateur comme outil pour prédire les possibles conformations du ligand TUA-diCl au sein de EF et EF-CaM. Nous avons ensuite utilisé la cartographie des épitopes fournie par les données STD pour filtrer les résultats de l'amarrage.

Les régions d'interaction suggérées sont le domaine hélicoïdal de EF et les parties N- et C-terminales de CaM.

Le site de liaison du TUA-diCl a été étudié par trois techniques: la fluorescence, HDX-MS et la cristallographie. Des essais de cristallogenèse ont été réalisés sur la protéine EF_3 et le complexe EF_3-CaM en présence ou non d'inhibiteurs. Des cristaux de EF_3-CaM ont été obtenus avec une diffraction de 4.75 Å et de 4.3 Å. La structure a été ensuite résolue. La structure du complexe EF-CaM est similaire à celles déjà publiées, le complexe est donc bien fonctionnel. Malheureusement, jusqu'à ce jour aucun cristal en présence du TUA-diCl n'a été détecté. Par fluorescence, nous avons montré que TUA-diCl est plus affin pour EF_3 que pour CaM et qu'il n'y aurait qu'un site de liaison sur EF_3. De plus, les expériences de HDX-MS sur EF_3 en présence de la molécule ont révélé une région où l'on observe une réduction d'accessibilité au solvant. Cette région se trouve dans le domaine hélicoïdal (résidus 704-729), et contient seulement un résidu (N709) de la poche SABC. L'ensemble de ces résultats préliminaires sembleraient indiquer que la molécule TUA-diCl se lie seulement sur un site de EF et que ce site est différent de celui initialement prévu pour inactiver la protéine.

Chapitre III. Caractérisation du mécanisme d'interaction entre le TUA-diCl et CaM

Dans la littérature, il est courant que de petites molécules conçues ou criblées pour inhiber des protéines activées par CaM, se lient également de manière inattendue à cette dernière. Nous nous sommes intéressées au mécanisme d'interaction avec CaM pour comprendre le mécanisme global d'inhibition du composé TUA-diCl.

Afin de déterminer le site d'interaction de TUA-diCl sur la calmoduline et de déterminer la constante d'affinité de l'interaction, nous avons effectué des expériences de perturbation des déplacements chimiques de CaM lors de la liaison de TUA-diCl. Pour ce faire, nous avons suivi les variations de déplacement chimique des groupements amides de CaM sur des spectres de corrélation ^{1}H-^{15}N (HSQC) lors de l'ajout de TUA-diCl, à l'aide d'une protéine CaM doublement marquée $^{15}N/^{13}C$ (Giotto Biotech). Nous avons au préalable attribué à l'aide de méthodes classiques les signaux de la chaîne principale de CaM dans les conditions expérimentales utilisées. L'interaction du TUA-diCl avec CaM en présence de calcium a été suivie à 37 °C sur un spectromètre Bruker avec une fréquence de résonance ^{1}H de 600 MHz. Nous avons montré que TUA-diCl se lie à des régions riches en résidus hydrophobes et plus

particulièrement les résidus suivants : F19, D20, I27, N53, V55, A57, I63, T70, M72, A73, R74, M76, Y77, R106, M109, N111, G113, K115, V121, I130, V142, Q143, M145, T146, A147 et K148. Le composé TUA-diCl est plus affin pour le lobe N-ter de que pour le lobe C-ter de CaM. Par ailleurs, en absence de calcium (apo-CaM), l'interaction avec TUA-diCl est abolie. L'interaction est calcium dépendante, ce qui a également été montré par les expériences de RMN proton monodimensionnelles.

La dynamique de la CaM a été étudiée en absence et en présence du ligand en suivant la relaxation du noyau ^{15}N pour déterminer le temps de réorientation de la molécule. Les mouvements des domaines C-ter et N-ter de CaM en absence de TUA-diCl semblent indépendants. Lors de l'ajout du ligand, les mouvements des deux domaines semblent être moins indépendants et la structure semble plus fermée.

Des études *in silico* ont été également menées afin d'identifier les interactions entre le ligand et la CaM.

Chapitre IV. Analyse *in silico* de la relation entre l'activation du facteur œdèmateux et son interaction avec la calmoduline

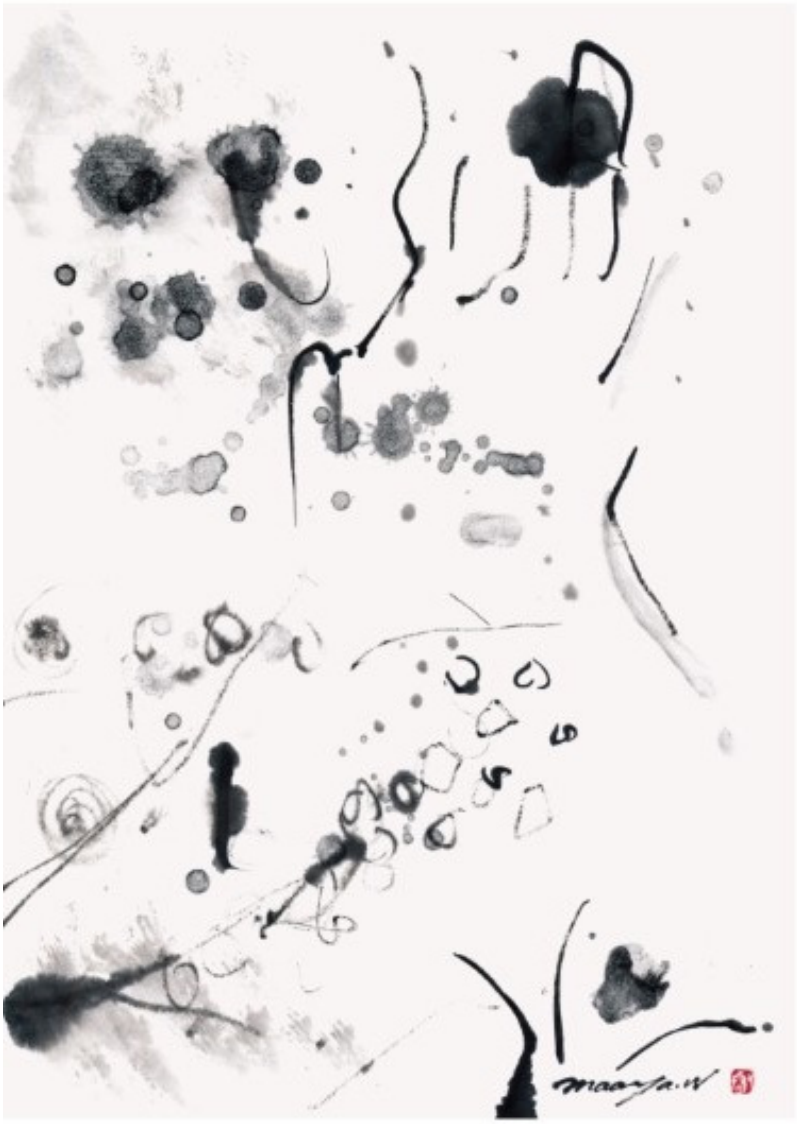
Dans le chapitre IV, nous avons combiné l'utilisation de simulations de dynamiques moléculaires (MD) avec des approches bio-informatiques pour analyser les interactions protéine-protéine. Le complexe EF-CaM a été largement étudié par la biologie structurale et les techniques biophysiques ainsi que par la modélisation moléculaire. Ce complexe représente un très bon système d'étude où la mobilité des deux partenaires est importante pour l'interaction et l'activation de la toxine. La CaM possède une plasticité remarquable. En effet, la CaM est piégée dans une conformation étendue dans le complexe EF-CaM et l'insertion de la CaM au sein de EF induit un large changement conformationnel de EF permettant ainsi son activation. L'approche dans ce chapitre est de détecter les poches allostériques permettant d'inhiber l'activité d'EF. Le système EF-CaM en interaction avec un ligand orthostérique l'adéfovir a été déstabilisé en retirant alternativement les ions Mg^{2+}, Ca^{2+} ou le ligand adefovir. Les poches les plus affectées par la déstabilisation étaient le site catalytique et l'interface EF-CaM. En ciblant l'interface EF-CaM, on pourrait envisager d'inhiber EF de manière allostérique. En parallèle, des simulations du ligand en interaction avec les trois systèmes ont été lancées. Ces simulations ont montré que l'interface EF-CaM a été déstabilisée par le TUA-diCl. En combinant des

données expérimentales et des études *in silico,* le site de liaison et le mécanisme d'action peuvent être mis en évidence, offrant une nouvelle stratégie pour explorer l'interaction moléculaire et trouver une nouvelle voie pour inhiber spécifiquement une cible.

Ce travail a été réalisé d'octobre 2017 à octobre 2020, au sein de l'Unité de Bioinformatique Structurale (M. Nilges) à l'Institut Pasteur, sous la direction du Dr. T. Malliavin et Dr. P.L. Goossens, en collaboration avec les différentes plateformes présentes à l'Institut Pasteur.

La vie n'est facile pour aucun de nous.
Mais quoi, il faut avoir de la persévérance,
et surtout de la confiance en soi.
Il faut croire que l'on est doué pour quelque chose, et que,
Cette chose, il faut l'atteindre coûte que coûte.

Marie Curie



Interprétation par le calligraphe Maaya Wakasugi de photos de microscope de *Bacillus anthracis* de Robert Koch.

Table of Contents

Acknowledgments 2

Résumé court 4

Abstract 5

Résumé long 6

Table of Contents 14

Index of Figures 16

Index of Tables 18

Abbreviations 19

Foreground 21

A. Introduction 23

A.1 Anthrax bioterrorism risk 23

A.1.1 The virulence factors of anthrax 24

A.1.2 Intoxication process by LT and ET 24

A.2 Edema Factor (EF) from *B. anthracis*, a bacterial adenylyl cyclase toxin 25

A.2.1 Mechanisms of entry of the toxins into eukaryotic cells 26

A.2.1.1 PA binding to cell-receptor 27

A.2.1.2 PA prepore to pore formation 28

A.2.1.3 Translocation of EF and LF across the membrane 30

A.2.1.4 Differences in the translocation process of EF and LF 33

A.2.2 Adenylyl cyclase catalytic site 34

A.2.2.1 Mammalian adenylyl cyclase 34

A.2.2.2 Catalytic mechanism of EF 39

A.2.2.3 Comparison of the catalytic mechanisms of EF, CyaA and ExoY 44

A.2.2.4 Activation of EF by CaM 46

A.2.2.5 Comparison of the activation modes of EF and CyaA by CaM 48

A.3 Identification of ligands targeting EF 50

A.3.1 Inhibitors of EF catalytic site 52

A.3.1.1 Ethyl 5-aminopyrazolo[1,5-α]quinazoline-3-carboxylate 52

A.3.1.2 Adefovir and PMEApp 53

A.3.1.3 3-[(9-oxo-9H-fluorene-1-carbonyl)-amino]-benzoic acid 54

A.3.1.4 P-site inhibitors 54

A.3.1.5 (M)ANT-nucleotides 55

A.3.2 EF Allosteric inhibitors 57

A.3.2.1 4-[4-(4-Nitrophenyl)-thiazolylamino]-benzene-sulfonamide 58

A.3.2.2 CaM inhibitors 59

A.3.2.3 Thiophen ureidoacids 59

B. Aims of the project 62

C. Materials and Methods 63

C.1 Thiophen ureidoacids 63

C.2 Protein expression and purification 64

C.2.1 Plasmids and cloning vectors 64

C.2.2 Protein expression and purification 65

C.2.3 Protein analysis 68

C.3 Enzymatic assays 68

C.4 NMR 70

C.5 Fluorescence 72

C.6 HDX-MS experiments ... 73
C.7 X-ray crystallography ... 74
C.8 Molecular docking ... 76
C.9 *In silico* studies of the EF-CaM complex ... 77

D. Experimental results & discussions ... 81

Chapter I. Biophysical and biochemical properties of the system ... 81

D.I.1 Biophysical properties of TUA compounds ... 81

CONTEXT ... 81

RESULTS ... 81

DISCUSSION ... 87

D.I.2 Biochemical and biophysical analysis of the catalytic activity of EF_3 ... 90

RESULTS ... 90
D.I.2.1 Optimization of EF_3 expression ... 90
D.I.2.2 EF_3 purification ... 94
D.I.2.3 Quality control ... 95
D.I.2.4 Buffer optimization ... 96
D.I.2.5 CaM purification ... 97
D.I.2.6 EF_3-CaM complex formation ... 97
D.I.2.7 Adenylyl cyclase activity of EF_3 ... 99

Chapter II. Molecular interactions between TUA-diCl, edema factor and its host activator calmodulin ... 101

INTRODUCTION ... 101

RESULTS ... 102
D.II.1 TUA-diCl inhibition of adenylyl cyclase activity ... 102
D.II.2 Interaction of TUA-diCl with the EF_3, CaM and EF_3 -CaM ... 104
D.II.3 Binding affinities of TUA-diCl for EF_3 and CaM ... 108
D.II.4 Interaction of TUA-diCl with the AC384 and AC384-CaM complex ... 109
D.II.5 Is TUA-diCl an allosteric inhibitor? ... 110
D.II.6 Structural analysis of the complex TUA-diCl-EF_3 complex by HDX-MS ... 116
D.II.7 Structural investigation on the complex EF_3-CaM interplay with TUA-diCl ... 120
D.II.8 Molecular docking of TUA-diCl with EF_3 and the complex EF_3-CaM ... 125

DISCUSSION ... 129

Chapter III. Interaction of TUA-diCl with CaM ... 133

INTRODUCTION ... 133

RESULTS ... 134
D.III.1 The interaction of TUA-diCl with CaM is calcium-dependent ... 134
D.III.2 Binding site of TUA-diCl ... 136
D.III.3 Influence of TUA-diCl binding on the dynamics of holo-CaM ... 139
D.III.4 Binding mechanism monitored by line shape analysis ... 143
D.III.5 Molecular docking and dynamics simulations studies ... 144
D.III.5.1 Molecular docking filtered using CSPs ... 145
D.III.5.2 Molecular docking using CSPs as restraints ... 145
D.III.5.3 Molecular docking on the N- and C-CaM lobes ... 148

DISCUSSION ... 153

Chapter IV: In silico analyzing of the relationship between the activation of the Edema Factor and its Interaction with Calmodulin ... 157

INTRODUCTION ... 157

RESULTS ... 159

D.IV.1 The removal of co-factors destabilizes the architecture of the EF-CaM complex 159
D.IV.2 CaM conformation in the EF-CaM complex conserves features of the isolated CaM 163
D.IV.3 A network of amino-acid interactions connects the EF catalytic site with CaM 167
D.IV.4 Analysis of cavities deformation to detect allosteric pockets 172
D.IV.5 TUA-diCl interacts with the EF-CaM interface 176
D.IV.1 TUA-diCl destabilizes CaM within the complex EF-CaM 178
D.IV.2 TUA-diCl destabilizes the network interaction connecting EF to CaM 180

DISCUSSION 185

E. Conclusion 188

F. Appendix 192

VIII. Bibliography 210

Index of Figures

Figure 1- Anthrax toxins mode of action. 26
Figure 2- PA_{83} monomer binding to CMG2. 28
Figure 3- Monomeric PA_{63} in prepore and pore conformations. 29
Figure 4- Clamps reorganization during pore formation. 30
Figure 5- Allosteric mechanism of EF-LF translocation. 32
Figure 6- Structural changes of EF during the intoxication process. 33
Figure 7- Crystal structure of the catalytic core 5C1.2C2 in complex with Gsα. FKS and ATP. 35
Figure 8- Activation of mACs by $G_{s\alpha}$. 36
Figure 9- Crystal structures of the catalytic site of mAC. 37
Figure 10- MANT-GTP inhibitor in the AC catalytic site. 38
Figure 11- Catalytic site of mACs. 39
Figure 12- Structural changes during EF activation. 40
Figure 13- Catalytic site of EF. 41
Figure 14- The two binding sites of ATP in the catalytic site. 42
Figure 15- Catalytic site of EF. 43
Figure 16- Two active metal ions in the catalytic reaction. 44
Figure 17- Structural comparison of EF, CyaA and ExoY. 46
Figure 18- Scheme of the mechanism of activation of EF by CaM. 48
Figure 19- Interaction of CaM with CyaA and EF. 49
Figure 20- Scheme of the inhibition mechanism of Fabs toward of B. anthracis toxins from Farcasanu et al[123]. 51
Figure 21- Ethyl 5-aminopyrazolo[1,5-α]quinazoline-3-carboxylate 53
Figure 22- Adefovir compounds 53
Figure 23- DC5: 3-[(9-oxo-9H-fluorene-1-carbonyl)-amino]-benzoic acid. 54
Figure 24- the P-site inhibitor 2'-d-3'-AMP 55
Figure 25- (M)ANT structures inhibitors. 56
Figure 26- 4-(4-(4-amino-7H-pyrrolo[2,3-d]pyrimidin-7-yl)piperidine-1-carbonyl)benzenesulfonyl fluoride trifluoroacetate. 57
Figure 27- 10506-2A and nitro-10506-2A. 59
Figure 28- Calmidazolium chloride. 59
Figure 29- SABC cavity in EF inactive conformation 60
Figure 30- 3,4-dichloro-5-phenyl-3-(3-phenylureido)thiophene-2-carboxylic acid. 61
Figure 31- Chemical synthesis of TUA compounds 63
Figure 32- Plasmid map: 65
Figure 33- The TUA-diCl compound. 82
Figure 34- ^1H spectra of TUA-diCl. 82
Figure 35- STD spectra of TUA-diCl at varying concentrations 85
Figure 36- Effect of the Tween 20 on the aggregation and on the inhibition power of TUA-diCl. 86

Figure 37- TUA-diCl bioavailability radar obtained with the interface swissADME[249] 87
Figure 38- The chemical groups of the TUA-ligand that can interact with different partners. 88
Figure 39- Chemical structures of the 10506-2A on the left and on the TUA-diCl on the right. 89
Figure 40- Location of the different amino acids between B. anthracis and B. cereus. 90
Figure 41- Strategy for protein overexpression. 91
Figure 42- SDS-PAGE evaluation of the different growth conditions EF_3 overexpression 93
Figure 43- Growth profile of EF_3 expressing cells in a 4 L bioreactor. 94
Figure 44- Size exclusion chromatography of purified EF_3 protein elution. 94
Figure 45- Quality control of EF_3 95
Figure 46- NanoDSF results. 96
Figure 47- CaM sample. 97
Figure 48- SEC elution of the three species. 98
Figure 49- 1D NMR spectra of the three species, ... 98
Figure 50- Stability of the EF_3-CaM complex monitored by SEC. 99
Figure 51- EF_3 enzymatic activity assays as a function of CaM. 100
Figure 52- TUA-diCl inhibition of adenylyl cyclase activity of EF_3 (A) and AC384 (B). 102
Figure 53- Activation of EF_3 with different concentrations of CaM in the presence of TUA-diCl. 103
Figure 54- TUA-diCl interaction with CaM, EF_3 and EF_3 -CaM by WaterLOGSY. 104
Figure 55- TUA-diCl interaction with CaM, EF_3 and EF_3 -CaM by followed by STD at 25°C. 105
Figure 56- The order of the addition of CaM does not affect the NMR spectra and the inhibition of TUA-diCl. 107
Figure 57- Binding isotherms of TUA-diCl with EF_3 and CaM monitored by fluorescence. 109
Figure 58- TUA-diCl interaction with AC384 and AC384-CaM assessed by STD. 110
Figure 59- Michaelis Menten kinetics of ATP of EF_3 in the presence of TUA-diCl. 111
Figure 60- Binding of 2'-MANT-3'-dATP and TUA-diCl to EF_3-CaM assessed by STD. 112
Figure 61- Transferred NOE spectrum of TUA-diCl in the presence of EF_3. 114
Figure 62- NOESY spectrum of TUA-diCl and 2'-MANT-3'-dATP in the presence of EF_3-CaM. 115
Figure 63- Peptide maps of EF_3 118
Figure 64- Effects of TUA-diCl binding on the solvent accessibility of EF_3. 119
Figure 65- Tiny crystal of the protein EF_3. 120
Figure 66- X-ray crystallographic structures of the complex EF_3-CaM. 125
Figure 67- Superimposition of the X-ray structure of the published and the herein obtained EF_3-CaM X-ray structures 125
Figure 68- Computation of theoretical STD ratios of the docking poses. 126
Figure 69- Cartoon representation of the three lowest RMSE poses and cavities for EF_3 and EF_3-CaM. 128
Figure 70- Possible inhibition mechanisms of EF_3-CaM by the TUA-diCl. 131
Figure 71- Calcium-dependent interaction between TUA-diCl and CaM. 135
Figure 72- Interaction between TUA-diCl and holo-CaM. 137
Figure 73- Peptide maps of CaM. 138
Figure 74- Effects of TUA-diCl binding on the solvent accessibility of CaM. 139
Figure 75- Effect of TUA-diCl on the internal dynamics of holo-CaM. 142
Figure 76- Line shape fit for residue M145. 144
Figure 78- Clustering of the docking poses obtained with HADDOCK based on the CSPs data. 146
Figure 79- Cluster analysis with the RMSE of the calculated and experimental STD values 147
Figure 80- Comparison between the best docked structures using the CSP information as restraints or to filter data. 147
Figure 81- Starting points of the MD simulations for the complex CaM-TUA-diCl obtained by minimizing the RMSE between the STD_{th} and STD_{exp} values. 149
Figure 82- Best extracted pose of TUA-diCl along the MD trajectory matching with the experimental CSPs data. 151
Figure 83- Comparison of TUA-diCl with some holo-CaM inhibitors. 154
Figure 84- Possible mechanisms of the TUA-diCl mode of binding to CaM. Both CaM domain plays an important role in the activation 156
Figure 85- Root-mean-square deviations (RMSD) of the backbone heavy atoms of EF and CaM with respect to the PDB structure 1PK0, ... 160
Figure 86-Coordinate root-mean-square fluctuations (RMSF) of the backbone heavy atoms of the complex EF-CaM 161
Figure 87- Contour plots describing the variation of the gyration radius with respect to the bending angle of the CaM central α-helix 163
Figure 88- Distribution of accessible surface of the hydrophobic patch of N-CaM and C-CaM. 165
Figure 89- Angle of the EF-hands 166

Figure 90- A. X-ray crystallographic structure (PDB ID 1PK0) of the complex EF-CaM..... 167
Figure 91- Distribution of accessible catalytic surface and the SABC pocket..... 168
Figure 92- Averaged volumes of the cavities detected by mkgridXf plotted along the cavity index 174
Figure 94- RMSD and RMSF..... 177
Figure 95- Starting and ending poses of the MD simulations..... 178
Figure 96- CaM features within the complex EF-CaM in presence of TUA-diCl..... 180
Figure 97- Distribution of accessible catalytic surface and the SABC pocket..... 182
Figure 98- Contacts between the TUA-diCl and the complex EF-CaM..... 183
Figure 99- Contacts between the TUA-diCl and EF..... 184

Index of Tables

Table 1- Systems for EF-CaM interaction studies 79
Table 2- Systems for TUA-diCl interaction studies with EF and EF-CaM..... 80
Table 3- Culture growth monitored by OD600 measurements..... 92
Table 4- Values of the STD_{exp} ratio of individual ligand protons in STD experiments. 106
Table 5- Kinetic parameters of enzymatic activity of EF_3-CaM with and without TUA-diCl. 111
Table 6- Crystallization conditions for the different systems. 122
Table 7- Data collection and preliminary refinement statistics for the EF_3-CaM complex. 124
Table 8- RMSE between the calculated and experimental STD ratios..... 127
Table 9- Effect of TUA-diCl binding on the tumbling correlation time of CaM. 140
Table 10- Results of the molecular docking on two cavities using STD experimental data. 148
Table 11- Residues of CaM involved in the interaction with CaM inhibitors and TUA-diCl residues with high CSPs..... 152
Table 12- Average distances (Å) between calcium ions (labeled Cal1 and Cal2) located in CaM EF-hands 3 and 4 and atoms of C-CaM residues..... 164
Table 13- Relevant interactions observed along MD trajectories connecting atoms from EF, CaM and adefovir (ade), as well as ions Ca^{2+} and Mg^{2+}..... 171
Table 14- The percentage of formation of the EF-CaM hydrogen network in presence of TUA-diCl..... 181

Abbreviations

Å, Angstroms

AC, Catalytic domain of CyaA

mAC, Mammalian Adenylyl cyclase

ACD, Adenylyl cyclase domain

ANT, Anthraniloyl-

ANTXR, Anthrax toxin receptors

AMP, Adenosine monophosphate

ATP, Adenosine tri-phosphate

cAMP, Cyclic adenosine monophosphate

CaM, Calmodulin

CMZ, Calmidazolium chloride

CSP, Chemical shift perturbation

CyaA, Adenylate cyclase toxin of *Bordetella pertussis*

DMSO, dimethyl sulfoxide

EDTA, Ethylenediaminetetraacetic acid

EGTA, Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid

EF, Edema factor

EF_3, Adenylyl cyclase domain of edema factor (residue 292-800)

ET, Edema toxin

ExoY, Exoenzyme Y

FKS, Forskolin

GF, Gel filtration

HD, helical domain

HDX-MS, Hydrogen/deuterium exchange mass spectrometry

HSQC/HMQC, Heteronuclear single/multiple quantum coherence/correlation

iPPIs, Inhibitors of protein-protein interactions

PPIs, Protein-protein interactions

LF, Lethal factor

LT, Lethal toxin

mAb, Monoclonal antibody

MANT, (N-methyl)-anthraniloyl

MAPK, Mitogen-activated protein kinase

MD, Molecular dynamics

MIDAS, Metal divalent ion-dependent adhesion site

NanoDSF, Nano differential scanning fluorimetry

NMR, Nuclear magnetic resonance

NOE, Nuclear Overhauser effect

trNOE, transferred Nuclear Overhauser effect

PA, Protective antigen

PABD, Protective antigen binding domain

PDB, Protein data bank

PKA, cAMP-dependent protein kinase

PMEA, 9-(2-phosphonylmethoxyethyl)adenine)

RMSD, Root mean square deviation

RMSE, Root mean square error

RMSF, Root mean square fluctuation

SDS-PAGE, Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SA, SB, SC, Switch A, Switch B, Switch C

SEC, size exclusion

SOFAST, band-selective optimized flip-angle short-transient experiment

STD-NMR, Saturation transfer difference

TFP, Trifluoperazine

TRACT, TROSY for rotational correlation times

TUA, Thiophen ureidoacids

vWA, von-Willebrand factor

WaterLOGSY, Water ligand observed via gradient spectroscopy

Foreground

My Ph.D. work was devoted to the study of one of the major virulence factors of *Bacillus anthracis,* the causative agent of anthrax. *B. anthracis* is a Gram-positive carrying three virulence factors: one poly-γ-D glutamyl capsule and two toxins. The toxins consist of the assembly of the protective antigen (PA) with the edema factor (EF) or lethal factor (LF). EF, the object of this work, is an adenylate cyclase enzyme that catalyzes the conversion of ATP into the second messenger cAMP (3', 5'-cyclic adenosine monophosphate). EF is responsible of increasing cAMP concentration in host cells altering water homeostasis and provoking edema. The enzyme participates in the cutaneous and systemic forms of anthrax. Upon PA binding to the anthrax toxin host cell receptors on innate immune system cells, PA oligomerizes and forms a complex with EF (or LF), the complex is endocytosed and after acidification of the endosome is converted to a pore, which allows EF (or LF) to translocate into the host cell cytosol. Once in the cytosol, EF is activated by host calmodulin (CaM) and displays its adenylyl cyclase activity leading to supraphysiological levels of cAMP that disrupt the host signaling pathways.

In order to investigate a biological threat and find inhibitors, we have to better understand the molecular targets of the pathogen, their functions and mechanisms of action. Combining several approaches such as molecular modeling and experimental structural biology techniques can unravel the function of molecular machineries. X-ray crystallography, Nuclear Magnetic Resonance (NMR), fluorescence and cryo-Electron Microscopy (cryo-EM) coupled to biochemistry and other biophysical techniques have revealed key features of the translocation mechanism, the activation of the proteins and the catalytic activity (some examples are highlighted in the revue in appendix). The intoxication mechanism is now better understood paving the way to find toxin inhibitors. Therapeutics approaches can target the different steps of intoxication: the entry of the proteins in the host cells, the activation of the protein or the catalytic activity. Orthosteric or allosteric strategies can be considered to inhibit the function of the toxins.

During a previous Ph.D. thesis in the laboratory, funded by the DGA, an original way of inhibition targeting the activation of EF by CaM was found. The strategy was to discover *in silico* an allosteric site present in the inactive conformation of EF that disappears upon CaM

binding. Interaction with CaM induces the folding of a disordered switch in EF that results in the appearance of the catalytic site and the subsequent activation of the enzyme. By targeting this allosteric switch region, *in silico* studies complemented with experimental data, showed that thiophen ureidoacids (TUA) inhibit cAMP production. The objectives of the present thesis are to give some insights on the inhibition site and mode of action of the TUA compounds. Indeed, up to now, no structural data concerning the binding site or mechanism of inhibition are available.

A. Introduction

A.1 Anthrax bioterrorism risk

Infectious agents can be classified in three priority categories A, B and C depending on their spreading capacity, the associated mortality rate, as well as on the public health preparedness and on the induced public fear[1,2]. In 2018, *Bacillus anthracis* (anthrax) was listed in category A of the priority pathogens together with *Clostridium botulinum* (botulism), *Yersinia pestis* (plague), *Variola major* (smallpox), *Francisella tularensis* (tularemia), Lassa, Machupo, Dengue, Ebola and Marburg viruses[1,2]. Anthrax, a threatful infectious disease that mainly affects herbivores but can be transmitted to humans, is caused by the gram-positive *Bacillus anthracis*, a member of the *Bacillus cereus* group. *B. anthracis* is encountered in two forms: the dormant and resistant spores and the vegetative bacilli. Spores are present in the soil and can survive for a long period (~ 100 years). Spores can survive in drastic environments, being resistant to heat, ultraviolet or ionizing radiation, pressure and chemical agents[3]. The spores are the infectious morphotype, which can germinate within the host and release virulence factors leading to the host death. The microscopic spores, can be easily added to sprays, food or water to be used as a biological weapon[4]. Anthrax has been indeed used in several circumstances[5]: (i) during the first world war by the German army to infect the livestock of the Allied Nations, (ii) during the second world war, by the Japanese on the Chinese prisoners in Manchuria and in 11 Chinese cities, (iii) by UK military in 1942–1943 by testing a spore delivery system leading to the death of all island's sheep, (iv) accidentally by releasing anthrax in the environment during an accident in 1979 in a biological weapon center in Sverdlovsk, URSS, (v) by the Iraqi on the Kurds, (vi) by the Aum Shinrikyo group in Tokyo and (vii) recently in 2001, after attacks on the World Trade Center. In the last case, 22 persons got infected via letters containing powder with *B. anthracis* spores. Although a convention prohibiting biological weapons was signed in 1975 by almost all world countries except Chad, South Soudan, Namibia, Eritrea, Comoros and Israel[6,7], this type of weapon has been strongly suspected to be used thereafter.

Current biodefense research strategies against anthrax are focusing on understanding the biology and the response of the human immune system towards the infectious agent, on developing and improving drugs, vaccines and quick tools for diagnosis, as well as on establishing resources and biosafety for laboratories[8].

A.1.1 The virulence factors of anthrax

Bacillus anthracis produces virulence factors such as a poly-γ-D-glutamic acid capsule and a toxin made up of three different proteins: Protective Antigen (PA, 83 kDa), Edema Factor (EF, 89 kDa) and Lethal Factor (LF, 90 kDa), which are individually nontoxic. The tripartite anthrax toxin is an AB toxin, composed of an A component displaying an enzymatic activity inside the host cytosol and a B component involved in cell binding. The EF and LF proteins constitute the A moieties, and PA the B one[9]. The AB toxin family includes also cholera, diphtheria, botulism, tetanus and pertussis toxins[9]. The intoxication process is due to the pairwise association of LF and PA (LT: Lethal Toxin) and EF and PA (ET: Edema Toxin). EF and LF follow the same mechanism to enter host cells by establishing an interaction between their N-terminal domain and PA but have different enzymatic activities and targets. While LF is a zinc protease that specifically cleaves the majority of mitogen-activated protein kinase (MAPK) kinase proteins[10,11] causing the deregulation of cellular signaling pathways, EF is an adenylate cyclase, which upon activation by calmodulin (CaM) leads to supraphysiologic levels of the second messenger 3'-5' cyclic adenosine monophosphate (cAMP)[12,13]. In this thesis, we focus on EF activation and activity.

A.1.2 Intoxication process by LT and ET

After either cutaneous, gastrointestinal or inhalational anthrax infections, the virulence factors are released into the host after germination of the spores. In the early stages of infection, ET causes edema. The first step of bacterial invasion is the germination of the spores within the host immune cells after phagocytosis[14], germination can also occur extracellularly[15]. LT and ET display their activity in particular in the host innate immune cells, namely neutrophils, macrophages, and dendritic cells (DCs)[16,]. The two toxins LT and ET indeed cooperate to promote bacterial invasion freezing the host immune response as well as the human neutrophil NADPH oxidase activity[16]. ET toxicity is due to its adenylate cyclase activity that allows the enzyme to overproduce second messengers [mainly cAMP, but also cyclic cytidine monophosphate (cCMP) and cyclic uridine monophosphate (cUMP)] involved in differentiation, proliferation and apoptosis of the immune cells[17]. The cAMP overproduction by ET stimulates the transcription factor involved in the expression of exchange protein activated by cAMP (Epac) and cAMP-dependent protein kinase A (PKA). PKA then triggers the activation of cAMP-responsive element binding (CREB)[18,19]. ET also promotes the

activation of cell surface anthrax toxin receptors in DC and in the monocyte cells that migrate into the lymph node[20,21,22], where the production of the toxin would be increased[23]. LT and ET inhibit immune functions such as phagocytosis and chemotaxis. LT and ET inhibit the expression of pro-inflammatory cytokines such as TNF-α, IL-10, IL-12p70 and suppress interferon expression[16,24]. The immune signaling pathways are disrupted hampering the recruitment of immune cells and causing their apoptosis. The maturation and activation of DCs by ET alter the adaptive immune system and thus contribute to the bacterial dissemination[16]. Besides, the activation, chemotaxis and proliferation of T cells are also affected by LT and ET. Depending on the polarization of the T cells, the immune response is not the same for the different subtypes of lymphocytes Th1, Th17 and Th2. Th1 activate the macrophage, Th17 stimulate the local inflammation by cell recruitment, whereas Th2 stimulate the production of antibodies and B lymphocyte[24,25]. ET induces the polarization of T cell to Th2 and Th17 rather than to Th1[24,25]. At the same time, LT would be involved in the suppression of lymphocyte B proliferation and antibody production[26]. However, the role of ET in B cells is not yet understood. Noninvasive imaging technologies show how can ET induce apoptosis of lymphoid cells and play a major role in the early stage of the inhalational infection[27].

ET and LT disrupt the immune system of the host leading to dramatic consequences for the host. To fight against bioterrorism threats, finding new treatments against anthrax is crucial. Understanding the function and the mechanism of *B. anthracis* proteins at a molecular level is important to this end. The structures of *B. anthracis* proteins at atomic resolution can give insights into the function and mode of action of the proteins.

A.2 Edema Factor (EF) from *B. anthracis,* a bacterial adenylyl cyclase toxin

For many years, structural information available on *B. anthracis* toxins were only separated structures of EF[28,29], PA[30,31,32,33] and LF[34], determined by X-ray crystallography. During the last years, with the development of cryo-EM, several multi-component structures appeared, giving new structural insights on the different steps of the toxin translocation process through the host membrane. In this introduction chapter, we describe the different structures highlighting these recently obtained multicomponent structures to shed light on the mechanism of *B. anthracis* toxins.

A.2.1 Mechanisms of entry of the toxins into eukaryotic cells

Once the three exotoxins (PA, LF, EF) are released into the blood, the PA monomers bind to anthrax toxin receptors (ANTXR). Two different anthrax toxin receptors are known: ANTXR1, the tumor endothelial marker 8 (TEM8/ATR)[35] and the ANTXR2, the capillary morphogenesis protein 2 (CMG2)[36]. More recently, the low-density lipoprotein receptor-related protein 6 (LRP6), a cell surface protein, has been found to be an ANTXR co-receptor[37]. The monomer PA is cleaved by a host furin membrane protease, producing PA_{20} and PA_{63} and PA_{63} monomers spontaneously assemble into heptamers (70%)[38,30,39,40] and octamers (30%)[31,41,42]. PA oligomers are known to assemble either (i) in plasma[43] or (ii) at the cell-surface (Figure 1). Up to three (four) molecules of either EF and/or LF interact with a PA_{63} heptamer[44] (octamer[31,41,42]). The complex is then endocytosed through an endosomal acidic compartment. The acidic pH of endosomal vesicles causes conformational changes of the toxins and concomitant transmembrane pore formation and toxins released in the cytosol.

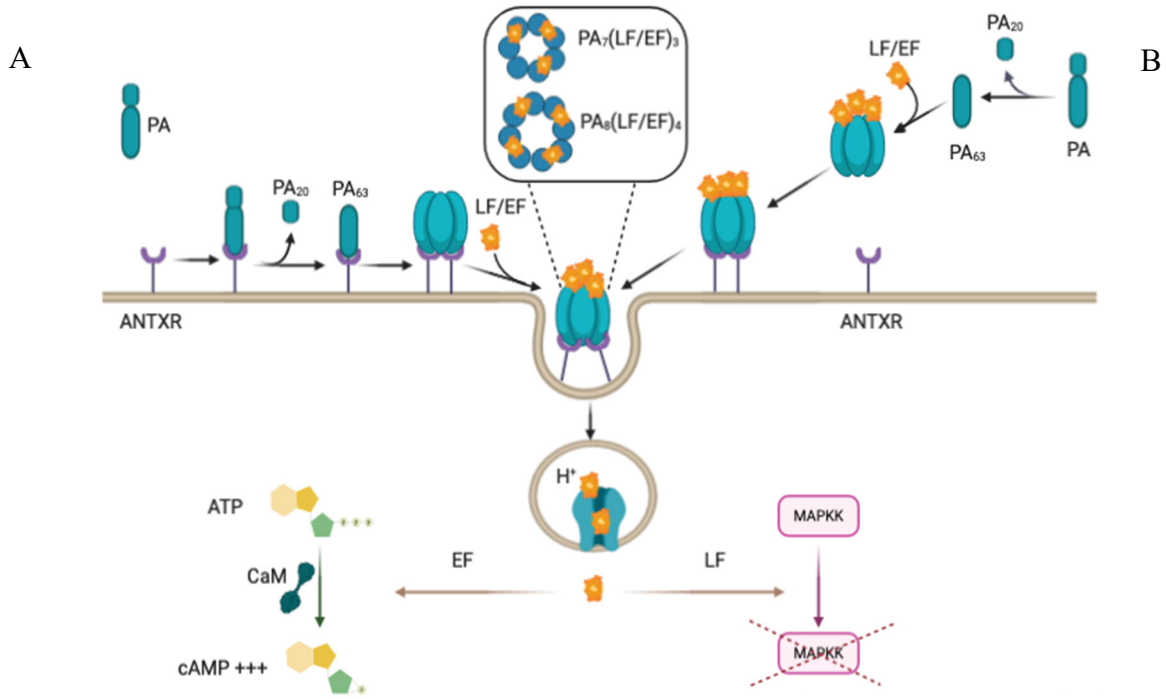


Figure 1- Anthrax toxins mode of action. *The initial steps of the anthrax toxins translocation into the host cell take place either at the membrane surface (A) or in plasma (B). PA (blue) is cleaved by furin, a membrane protease, producing PA_{20} and PA_{63}. While PA_{20} is released in the extracellular matrix, PA_{63} agglomerates at the surface of the host membrane (ANTXR, violet) or in the blood either as heptamers or octamers. EF or LF (in yellow) bind to PA_{63} oligomers, the ET or LT complexes are then internalized into the host cell through endosomal vesicles. The acidic conditions of endosomal vesicles trigger conformational changes leading to pore formation and EF/LF translocation through the vesicles'*

membrane. Once in the cytosol, EF rises up the level of cAMP favoring water efflux and thereby edema, whereas LF cleaves MAPKK, generating a series of events conducting to the host cell death.

This and all thesis scheme figures were created with BioRender.com.

A.2.1.1 PA binding to cell-receptor

PA is a cell binding protein mediating the translocation of EF and LF through the host membrane. The first X-ray crystallographic structure of the 83 kDa PA determined in 1997[32], revealed four domains (Figure 2). Domain 1 (blue in Figure 2) consists of the N-terminal fragment PA_{20} after furin cleavage and Domain 1' (cyan in Figure 2), which corresponds to the N-terminal region of PA_{63}, constitutes the binding site of EF and LF and is involved in the oligomerization of PA_{63} subunits. Domain 2 (hot pink in Figure 2) is composed of a β barrel core and a large flexible loop ($2\beta_2$-$2\beta_3$ loop, purple in Figure 3) between the β-strands $2\beta_2$ and $2\beta_3$. Under acidic pH, this loop folds into two β-strands and inserts into the membrane, leading to the formation of the β-barrel channel of the pore. Domain 2, as well as Domain 3 is also involved the oligomerization of PA_{63} subunits[45]. Domain 4 (dark red in Figure 2) is a receptor-binding domain located at the carboxy-terminal of PA_{63}[32].

The two anthrax toxin receptors (ANTXR) described at the beginning of section 2.1, are constituted of a N-terminal extracellular von-Willebrand factor (vWA), an also extracellular immunoglobin-like domain and a single transmembrane region with a cytosolic domain located at the C-terminal. The vWA domains of the two ANTXR share ~ 60% sequence similarity[36]. PA binds to the extracellular vWA domain of ANTXR. The vWA domain includes a metal divalent ion-dependent adhesion site (MIDAS). Two X-ray crystallographic structures[33,46] had been reported for the complex between PA and the VWA domain of CMG2 (colored in coral in Figure 2). These structures elucidate the role of the MIDAS[33,30,46,47]. The MIDAS undergoes a large structural change between a closed to an open form and is essential for PA-receptor interaction[48,47]. The residue D683 of PA coordinating the Mn^{2+} cation in the MIDAS of the ANTXR receptors is crucial for PA-ANTXR binding. This structure also underlines that the $2\beta_3$-$2\beta_4$ loop (in dark olive green in Figure 2) located in domain 2, in addition, to domain 4 of PA_{63} both interact with the CMG2. The interaction between $2\beta_3$-$2\beta_4$ loop and vWA is sensitive to pH[49].

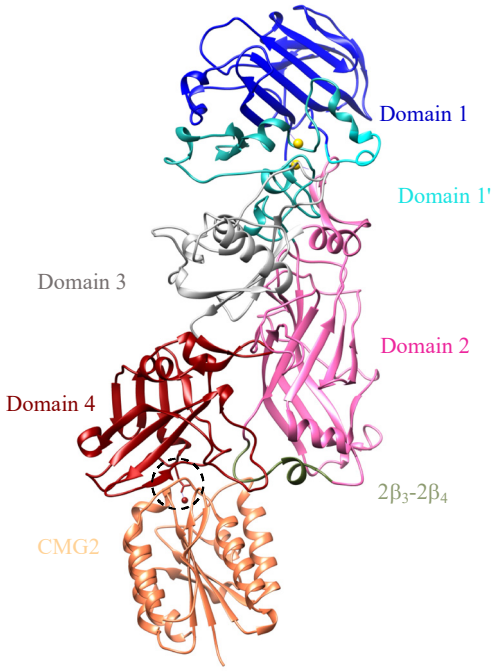


Figure 2- PA_{83} monomer binding to CMG2. *Cartoon representation of the structure of PA bound to the von-Willebrand factor of CMG2 (PDB ID 1T6B). Domain 1 is divided into two parts: the PA_{20} domain after furin cleavage (blue, residues 1-167), and Domain 1' (cyan, residues 168-259). Domain 2 (residues 250-487) is shown in hot pink. The $2\beta_3$-$2\beta_4$ loop (dark olive green) located in Domain 2 interacts with CMG2 colored in coral. Domain 3 (residues 488-596) is colored in grey and Domain 4 (residues 597-735) in dark red. Residue D683 (sticks), interacts with CMG2 and the metal located in the MIDAS of the vWA domain of CMG2. The Mn^{2+} ion at the interface of PA-CMG2 is colored in red.*

This and all thesis structure figures were obtained with Chimera[50] and the missing residues were reconstructed using the CHARMM GUI interface[51] (ww.charmm-gui.org).

A.2.1.2 PA prepore to pore formation

A recently solved high resolution cryo-EM structure provided important mechanistic information on the structural changes involved in the conversion of the prepore into the pore[52] (Figure 3). The pore is formed upon acidification in the endosome. Under acidic conditions, the PA_{63} prepore and ANTXR dissociate[53,54]. The comparison of the structures of the prepore PA_7 (Figure 3 A) and of the pore PA_7 (Figure 3 B) clearly shows that Domain 4 (dark red) of PA_{63} moves apart from Domain 2 (hot pink) of PA_{63} and that the orientation of the $2\beta_{10}$–$2\beta_{11}$ loop (forest green) located in Domain 2 of PA_{63} is different. These structural changes allow the Domain 2 $2\beta_2$-$2\beta_3$ loop (purple) to rotate and bend. The β-strands $2\beta_3$, $2\beta_2$ and the $2\beta_2$-$2\beta_3$ loop form the channel of the PA complex and the $2\beta_2$-$2\beta_3$ loop anchors on the endosomal membrane[52]. The only difference between heptameric and octameric complexes is the conformation of the $2\beta_{10}$–$2\beta_{11}$ loop.

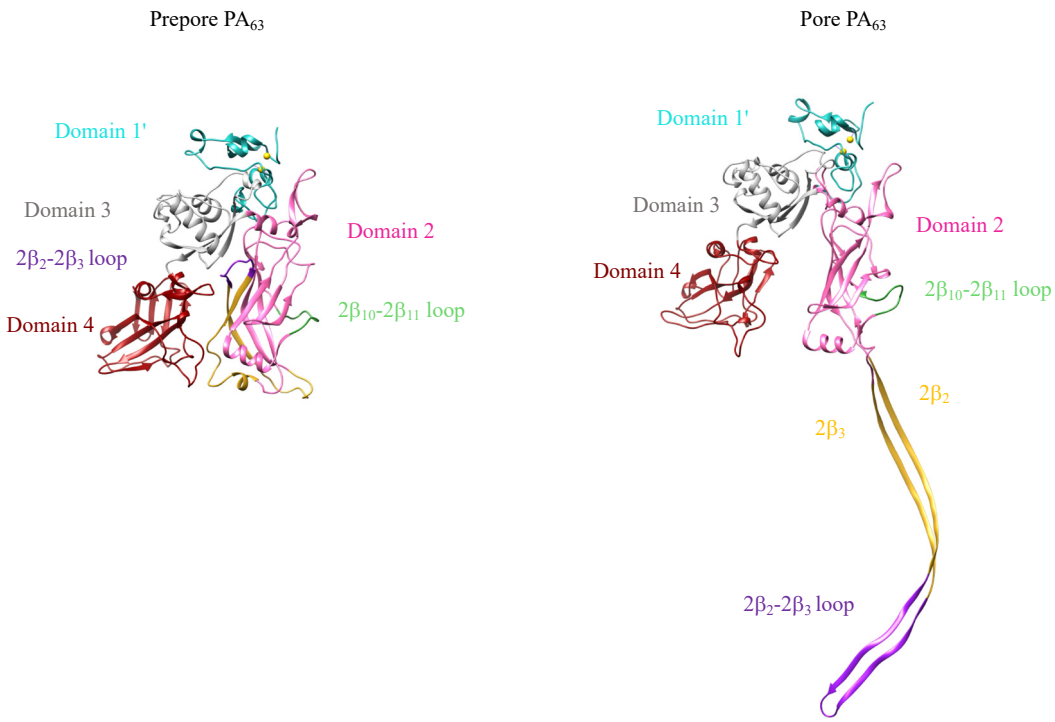


Figure 3- Monomeric PA_{63} in prepore and pore conformations. *The PA domains are colored following the same code as in Figure 2. In the left side of the figure, the structure of PA_{63} monomer forming the prepore is represented (PDB ID 1TZO) and in the right side, the structure of PA_{63} monomer in the prepore (PDB ID 3J9C). During the formation of the pore under acidification of the endosome, the loop $2\beta_{10}$-$2\beta_{11}$ (in forest green) is flipped and the β strands $2\beta_2$ and $2\beta_3$ (in gold) and $2\beta_2$-$2\beta_3$ loop (in purple) bend and assemble to participate in the formation of a β barrel inside the endosomal membrane.*

The pore is composed of three key clamps that are essential for the translocation process: the α (in cyan in Figure 4), the ϕ (green) and the charge (orange) clamps. The α clamp is located at the top interface of two PA units[42]. A salt bridge between residues K397 from one PA Domain 2 and D426 from the adjacent PA Domain 2 joins two PA units together. The loops containing these residues, ie the 397 loop (pink) above residue F427 and 426 loop (navy) below residue F427, bring together residues F427 from different PA subunits to form the ϕ clamp at the entrance of the pore channel[55,56]. The ϕ clamp is a ring constituted of seven (for a heptameric complex) F427 residues[57] from the $2\beta_{10}$–$2\beta_{11}$ loops of the PA subunits[52]. In the prepore the diameter of the ϕ clamp is 30 Å and in the pore it is 6 Å. The proton pH gradient between the endosome and the cytosol is maintained by the ϕ clamp. The ϕ and α clamps are dynamic clamps that communicate through a long-distance allosteric network that includes the 397 loop (pink). The allosteric network is regulated by the proton pH gradient[58,55,56]. The presence of the ϕ clamp coupled with a proton pH gradient (ΔpH) are required for the translocation of the unfolded EF and LF proteins [57,59,60]. The charge clamp (orange), located inside the β-barrel channel, is a

highly negatively charge surface composed of aspartate (D276) and glutamate (E343, E335) residues[61].

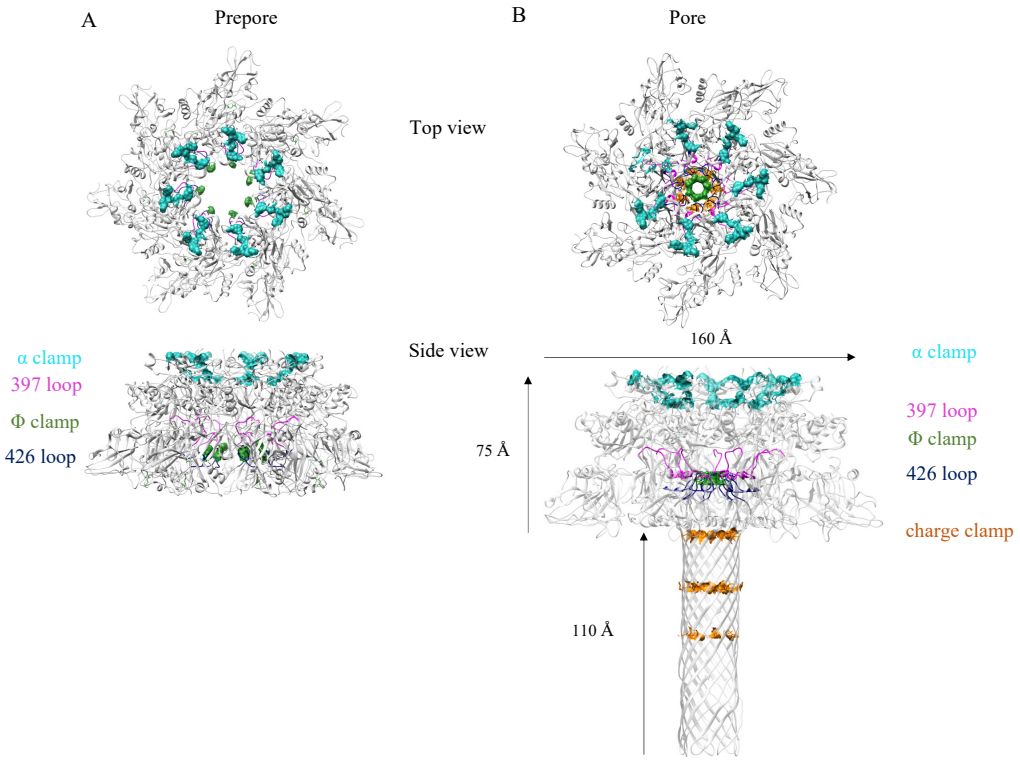


Figure 4- Clamps reorganization during pore formation. *A. Top and side views of the structure of the PA prepore (A, PDB ID 1TZO) and pore (B, PDB ID 6UZB) complexes. The α clamp is colored in cyan, the ϕ clamp in forest green and the charge clamp in orange. The structural rearrangements of the 397 (magenta) and 426 (navy) loops upon pore formation, bring close residue F427 of each PA subunit to form the narrow ϕ clamp of the pore. Indeed, the diameter of the ϕ clamp goes from 30 Å in the prepore to 6 Å in the pore. The charge clamp in the channel is composed of residues D276, E343 and E335.*

A.2.1.3 Translocation of EF and LF across the membrane

The structural features of the membrane translocation of EF and LF have been revealed by X-ray crystallographic[42] and cryo-EM[52,62] structures. Both toxins display similar translocation mechanisms. Once the PA_{63} prepore is formed, EF or LF binds to PA. The structures of octameric $[PA_8(LF_N)_4]$ and heptameric $[PA_7(LF_N)_3]$ PA in complex with the N-terminal region of LF (LF_N) showed no significant structural differences, suggesting a common translocation mechanism for both types of oligomer. As shown in Figure 5, EF and LF bind to the same site on PA_{63} oligomers via a homologous protective antigen-binding domain (PABD) located at the N-terminal extremity of EF (EF_N) and LF (LF_N)[34]. EF_N and LF_N bind at the

interface of two PA_{63} subunits, several sites in EF_N and LF_N allowing for several binding events[42]. The C-terminal of EF_N and LF_N binds to the hydrophobic interface of the PA_{63} prepore. This interaction engages the residues V223, L226 and Y227 of EF_N or the residues V232, L235, H229, Y223, L188 and Y236 of LF_N with residues F202, P205, I297 and I210 of PA_{63}[42,62]. Moreover, salt bridges between residues D187 and D184 of LF_N and residues K213 and K214 of PA_{63} are formed[42]. The α-helix located in the N-terminal of EF/LF binds non-specifically to the α-clamp of PA_{63}. The residues L33 to K35 of EF_N or I43 to K45 of LF_N interact with residues L203 to P205 of the α-clamp site of PA_{63}[62]. The α clamp stabilizes and orients the unfolded polypeptide (EF_N or LF_N) intermediate into the channel[42]. The different orientations of F464 in PA with LF_N and EF_N show that the α clamp is adaptable to bind non-specifically a large range of helices[62]. As mentioned before, acidification of the endosome induces conformational changes leading to the prepore to pore transition and the translocation of the toxins. Under acidic conditions, the acidic residues of the polypeptide are protonated and thereby the intramolecular hydrogen bonds and the salt bridges within the C-terminal of the polypeptide are destabilized. The α-helix of the polypeptide moves down to the ϕ clamp and the α clamp interacts with the polypeptide, triggering allosteric changes on the ϕ clamp[58,63]. The polypeptide passes through the tight ϕ clamp gate where the polypeptide unfolds. The unfolded polypeptide is driven into the charge clamp. A proton gradient ΔpH inside the channel connecting the endosomal compartment and the cytosol constitutes the driving force that drives the unfolded polypeptide into the charge clamp[60]. The membrane potential (Δψ) also contributes toxin translocation[59,42]. The deprotonation of the acidic residues of the polypeptide in the β barrel, helps the polypeptide to translocate through the channel. The polypeptide refolds in its secondary structure inside the channel and reaches the cytosol where it adopts again its tertiary structure.

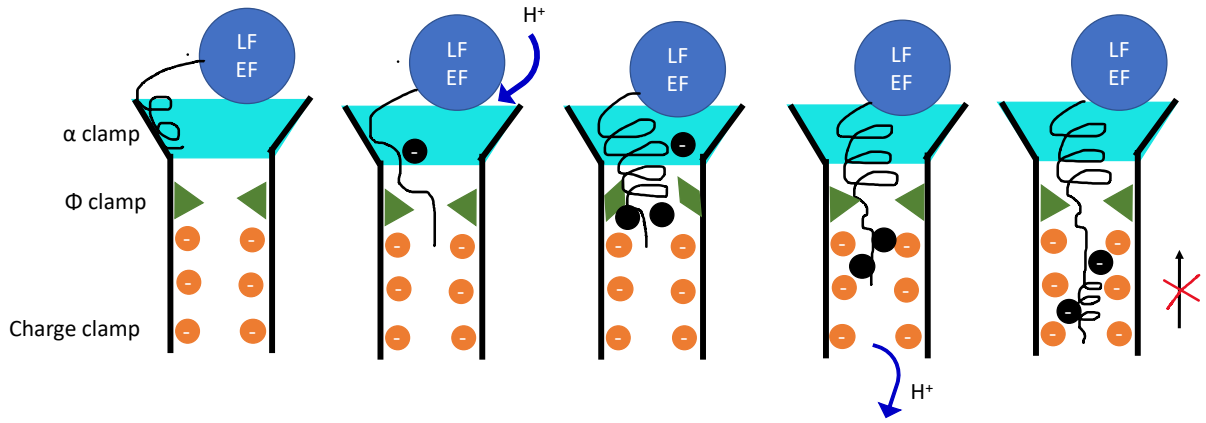


Figure 5- Allosteric mechanism of EF-LF translocation. *In the first step, acidic conditions destabilize the C-terminal region of PABD in the translocated polypeptide (LF or EF) and partially protonate the polypeptide chain (black wire). The N terminal α helix of the polypeptide binds to the α clamp (colored in cyan) and moves to the ϕ clamp region (colored in green) using favorable interactions through the allosteric network previously described. The conformational equilibrium of the ϕ clamp (colored in orange) between a dilated and a tight ring helps to unfold the polypeptide. The extended polypeptide then moves to the charge clamp, where the acidic residues of the polypeptide are deprotonated. A partial refolding of the polypeptide within the channel prevents it to retrograde and applies a force favoring the translocation. The cycle of α clamp binding and toxin unfolding is repeated until the full polypeptide transits to the cytosol.*

In 2020, a cryo-EM structure of $PA_7(LF_N)_3$ suggested that the binding of the first LF molecule influences the binding of the two next LF molecules. The second LF bound, was found to be in an intermediate state driven by the first LF binding event[63]. During the binding of three LF molecules to the prepore complex, two LF chains would be in an intermediate state and one LF chain in the open state described in the $PA_8(LF_N)_4$ X-ray structure[42].

Despite major advances in the understanding of translocation, several features remain unclear, such as the order of events and the translocation mechanism for β-sheet regions of the polypeptide. β-sheet would use another mechanism such as the extended-chain Brownian ratchet to translocated through the PA pore. First, the negatively-charged residues of the peptide would be neutralized by the proton in the α clamp. Then the peptide would translocate via a Brownian motion within the ϕ clamp to the charged clamp where the peptide is deprotonated and is trapped on the *trans* side of the channel[64].

A.2.1.4 Differences in the translocation process of EF and LF

Despite displaying homologous PADB regions, the first step of the translocation process differs for LF and EF. In contrast to LF, EF undergoes a large rearrangement in order to be translocated and then activated by the host calmodulin. Indeed, in the cryo-EM structure of the pore PA_{63} in complex with EF, the PABD and the helical domain (HD) of EF interact together (Figure 6B). This interaction was unexpected as in the crystal structure of EF with its cell activator, calmodulin (CaM), the PABD interacts with the adenylyl cyclase domain (ACD) of the toxin (Figure 6A)[65]. While in the crystal structure of EF-CaM, residues 20-30 of PABD are disordered[65], these residues are involved in an α-helix in the prepore complex and interact with the α-clamp. These observations indicate that the PABD domain would detach from the ACD domain upon CaM binding[65]. Furthermore, the cryo-EM structure of the PA/EF complex highlights the multi-functional character of the helical domain of EF, which connects the ACD and PABD domains and interacts with PA[62]. The ACD and the HD domains are consequently reorganized upon PA complex binding.

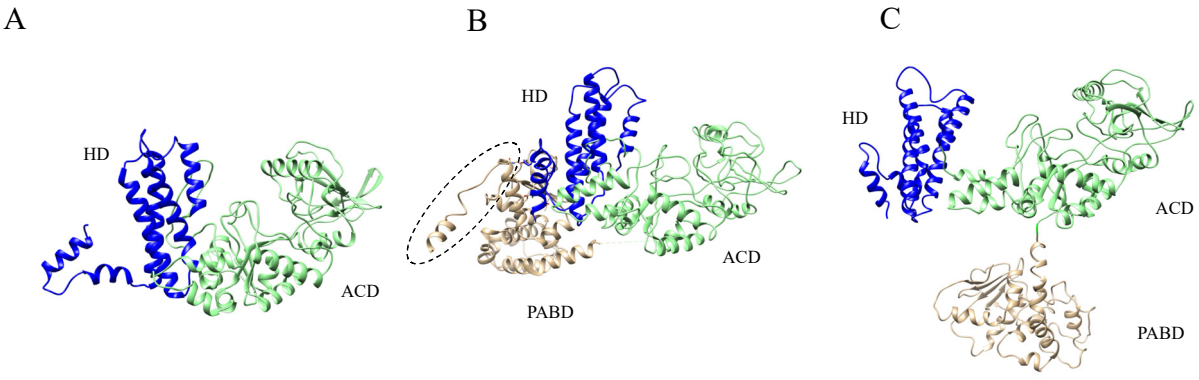


Figure 6- Structural changes of EF during the intoxication process.*X-ray crystallographic structure (PDB ID 1K8T) of inactive EF without the PABD domain highlighting the HD (blue) and ACD (light green) domains. B. Cryo-EM structure (PDB ID 6UZB) of EF bound to a PA_{63} oligomer showing the PABD domain (tan) interacting with the helical domain (HD). PA_{63} binding triggers the folding of the N-terminal part of PABD (dashed circle), which is unfolded in the active form of EF. C. X-ray crystallographic structure (PDB ID 1XFV) of active EF with the PABD domain (tan) connected to the adenylyl cyclase domain (ACD) by a flexible linker (light green). While HD shows the same orientation in the inactive (panel A) and pore-associated conformations (panel B), it swings 70 degrees away from ACD upon CaM binding (panel C).*

Once LF and EF reach the cytosol, these enzymes can display their respective activities. In the following section, we describe the catalytic and activation mechanisms of EF once it is in the cytosol.

A.2.2 Adenylyl cyclase catalytic site

Leppla demonstrated that EF is a bacterial adenylyl cyclase[12] activated by calmodulin[13]. The common function of adenylyl cyclases is to catalyze the conversion of the adenosine 3,5-triphosphate (ATP) into cyclic adenosine 3',5'-mono-phosphate (cAMP) with pyrophosphate (PPi) as a by-product[66]. Adenylyl cyclases are divided into six classes. In particular, class II and class III correspond to bacterial and mammalian adenylyl cyclases, respectively[67]. Mammalian cells require cAMP produce for numerous intracellular processes. Pathogenic bacteria take advantage of these processes to impair the host defense mechanism by overproducing cAMP in innate immune cells and thus deregulating the cAMP-associated processes. The mechanisms of activation and catalytic activity were first clarified for mammalian adenylyl cyclases, linking structural and functional information. Describing the first general catalysis mechanism established for mammalian adenylyl cyclases, is useful to understand the mechanism of the bacterial cyclases and to highlight their differences.

A.2.2.1 Mammalian adenylyl cyclase

The mammalian adenylyl cyclases (mACs) are classified into nine membrane-bound isoforms (mAC1-9) and a soluble isoform. These nine membrane-bound isoforms associated enzymes are regulated by the G-protein ($G_{\alpha s}$), P-site inhibitors, and depending on their class, by other G-protein subunits, forskolin (FKS), calcium, calcium/calmodulin (CaM) and protein kinases[66]. The soluble adenylyl cyclase is activated by bicarbonate and calcium[67].

Based on different X-ray crystallographic structures[68,69,70] of mACs, their first catalytic mechanism was proposed in 1997. The catalytic core is composed of two cytoplasmic (~ 40 kDa) functional domains named C1 and C2 that share a similar structure. The heterodimer complex C1/C2 is also common to guanylyl cyclase[71]. A prototypical mAC catalytic core composed of a mAC type 5 domain C1 subunit A (5C1, in forest green in Figure 7) and mAC type 2 domain C2 subunit A (2C2, in orange red in Figure 7) was crystallized with the activators forskolin (FKS, in blue in Figure 7) and guanosine 5'-O-(3-thiotri-phosphate)–activated $G_{s\alpha}$ ($G_{s\alpha}$.$GTP_{\gamma S}$)[68]. The structure, solved in the presence of a P-site inhibitor and PPi, demonstrated that two sites are present at the large C1/C2 interface: an ATP binding and a second site for one molecule of FKS and $G_{s\alpha}$.$GTP_{\gamma S}$ (Figure 7)[68]. A second structure[69,72] mimicking the heterodimer C1/C2 catalytic core with two mAC type 2 domains C2 (2C2) showed the presence of two

binding sites for FKS in the homodimer, in agreement with an ATP binding site and an additional binding site[69].

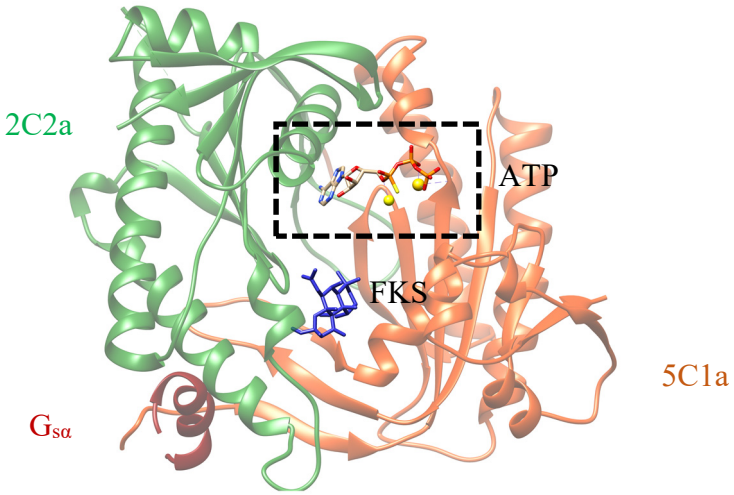


Figure 7- Crystal structure of the catalytic core 5C1.2C2 in complex with Gsα. FKS and ATP. *The prototype catalytic core of mACs is composed of the domain 5C1 (orange red) and 2C2 (forest green) (PDB ID 1CJK). mAC is activated by $G_{s\alpha}$ and especially by interaction with the switch II region of $G_{s\alpha}$ (dark red). The catalytic site shows two different sites: one for ATP (tan sticks) and FKS (blue sticks). The yellow spheres represent the Mg^{2+} and Mn^{2+} ions.*

The activator ($G_{s\alpha}$) interacts with C2 through a hydrophobic bridge. By comparing the crystal structure of 5C1.2C2.FS.$G_{s\alpha}$.$GTP_{\gamma S}$ and 2C2.FS, small structural changes were noticed, modifying the relative orientation of C1/C2 leading to the organization of the catalytic site (Figure 8)[68,72].

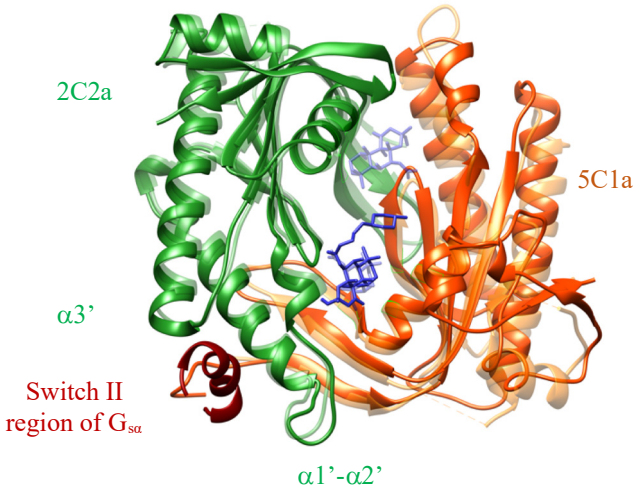


Figure 8- Activation of mACs by $G_{s\alpha}$. *Superimposition of the structures of 2C2.FS (PDB ID 1AB8, light green and red orange) and 5C1.2C2.FS.$G_{s\alpha}$.$GTP_{\gamma S}$ (PDB ID 1AZS), following the same color code as Figure 7. The activator $G_{s\alpha}$ (dark red) makes hydrophobic and polar contacts with the catalytic core 5C1.2C2. The reorientation of 5C1 and 2C2 is due to the insertion of $G_{s\alpha}$ switch II between the 2C2 α1'-α2' loop (black circle) and the α3' helix.*

P-site inhibitors are purine ring-containing non-competitive inhibitors of adenylyl cyclase activators such as Mn^{2+}, FKS and $G_{s\alpha}$ but competitors of cAMP[73]. The structures of adenylyl cyclase in the presence of the P-site inhibitors 2'-d-3'-AMP (2′-deoxyadenosine 3′-monophosphate) and 2′,5′-dd-3′-ATP (2′,5′-dideoxy-adenosine 3′-triphosphate) were instrumental to shed light on the binding mode of ATP[72]. These compounds bind to the ATP binding site only in the presence of PPi and a divalent ion (Mn^{2+}, Mg^{2+}, Zn^{2+}). The crystal structures in the presence or the absence of the latter ATP-like compounds revealed an open (5C1.2C2.FS.$G_{s\alpha}$.$GTP_{\gamma S}$) and a closed (5C1.2C2.FS.$G_{s\alpha}$.$GTP_{\gamma S}$.2'-d-3'-AMP.PPi) conformation of the catalytic site of mACs. Upon 2'-d-3'-AMP binding, the β_2-β_3, α_1-α_2 and α_3-β_4 loops of 5C1 and the β_7-β_8 loop of 2C2 move closer to the substrate changing the catalytic site to an active closed conformation. The following residues are required in the 2C2 domain for the activity: R1029 stabilizes α-phosphate intermediate, N1025 forms a hydrogen bond with a water molecule and the nitrogen N9 of ATP, D1018 and K938 form hydrogen bonds with nitrogen N6 and N1 of ATP, respectively. In the 5C1 domain: D396 and D440 interact with 3'-ribosyl OH of ATP and with the metal ion[68] The metal ion deprotonates the 3'-OH from ATP which then attacks the α-phosphate from ATP and releases first cAMP and then PPi[74]. In 1998, the catalytic mechanism of mAC was elucidated. Two divalent metal ions observed in the structure of the enzyme (hereby named A and B) stabilize the transition state of the enzymatic

conversion of ATP into cAMP. Metal A either induces a direct attack or activates a water molecule to deprotonate the 3'OH-hydroxyl group of ATP. The second metal B promotes a stabilization by binding to the ß and γ phosphate groups[70,75] (Figure 9B). The 3'endo conformation of ATP favors the nucleophilic attack of its 3'-OH on the α-phosphate[76].

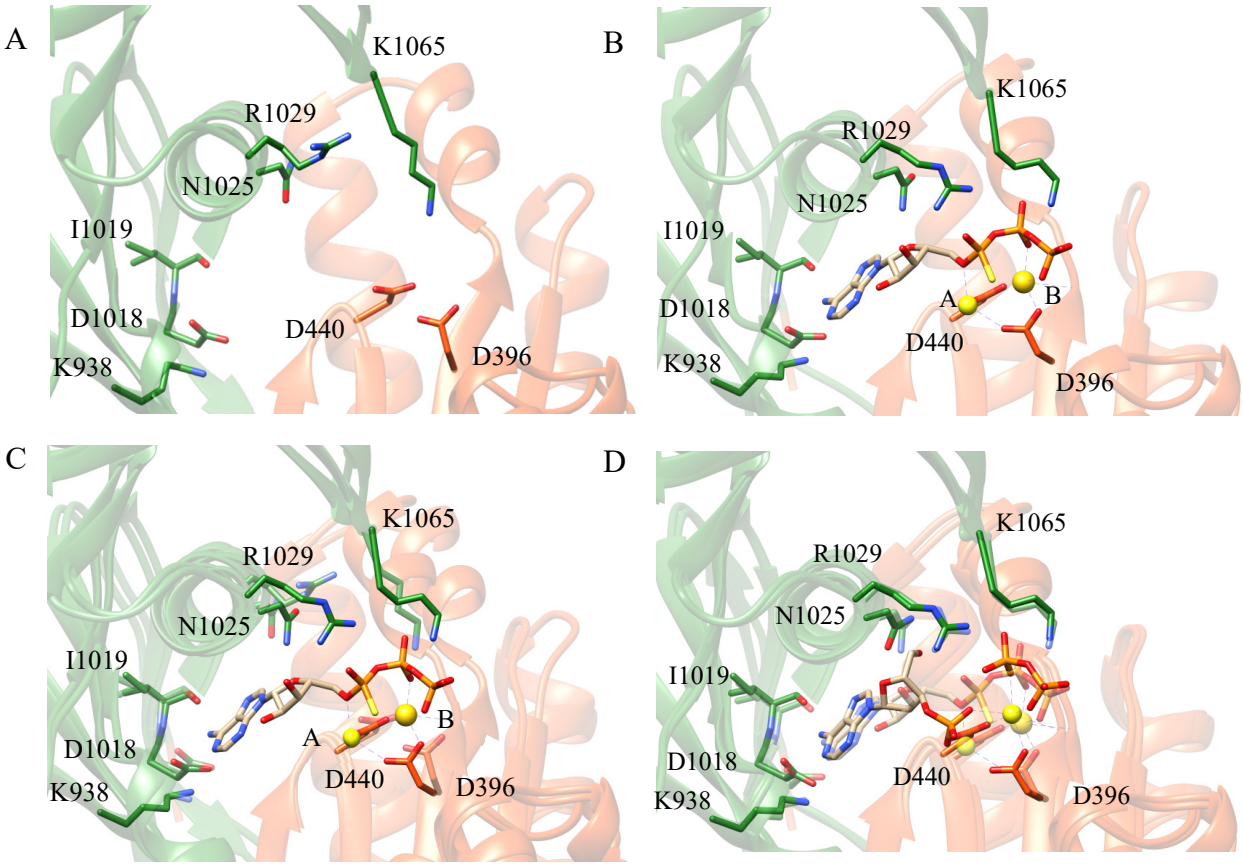


Figure 9- Crystal structures of the catalytic site of mAC. *A. Structure of the open state 5C1.2C2 (PDB ID 1AZS) B. Structure of the closed state (PDB ID 1CJK) with two metal ions (A, Mg^{2+}, B, Mn^{2+}, in yellow) and a non-hydrolysable analog of ATP ($ATP_{\gamma S}$-Rp) colored by atom (nitrogen in blue, oxygen in red, phosphorus in orange, carbon in tan). C. Superimposition of structures shown in A (lighter colors) and B, indicating that the key residues (side-chain in sticks representation colored by atom) of the catalytic site are reoriented upon ATP and metal ions binding, especially R1029, K1065, N1025, D440 and D396. D. Superimposition of the closed state shown in B (lighter colors) and the crystal structure of cAMP and PPi in the catalytic core (PDB ID 1CS4). The phosphate chain of cAMP is 1 Å away from the one of $ATP_{\gamma S}$-Rp, whereas PPi is superimposed with the phosphate chain of $ATP_{\gamma S}$-Rp.*

In 2006, an ATP inhibitor, 2'(3')-O-(N-methylanthraniloyl)-guanosine5'-triphosphate (MANT-GTP), was crystallized in the same position as the P-site inhibitor in the mAC catalytic site[77] (Figure 10). The MANT compounds are fluorescent probes for ATP binding sites and mAC inhibitors. MANT-GTP interacts with two metal ions. The comparison of the bound structures of ATP and MANT-GTP revealed that the base rings of the two molecules are

oriented differently. The interaction with MANT-GTP prevents a closed state of the catalytic site and blocks the conformation in an intermediate state leading to an inactivation of mAC. Residue R1029 is locked in a more open state and the orientation of N1025 is different, highlighting the importance of these two residues for the enzymatic activity.

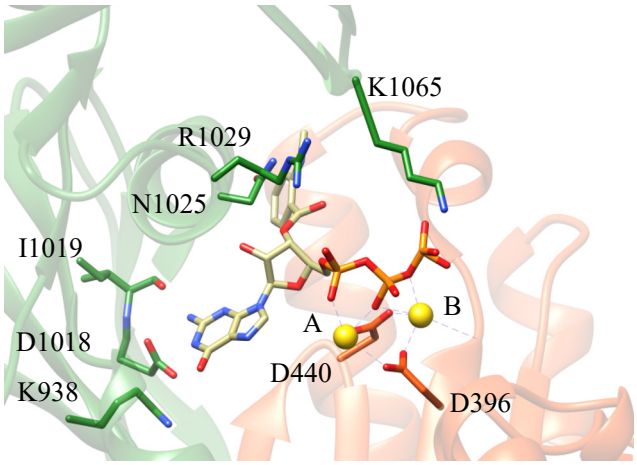


Figure 10- MANT-GTP inhibitor in the AC catalytic site. *The inhibitor MANT-GTP (shown in sticks colored by atom as in Figure 9) occupies the ATP binding site of mAC. The additional steric bulk group in position 2'/3' of ATP, induces the rotation of residues R1029 and N1025 and blocks the catalytic activity of mAC. The two metal ions Mn^{2+} (gold spheres) are coordinated by the triphosphate chain of MANT-GTP and residues D440 and D396 of mAC.*

Metal ions play an important role in the activity of mACs. Depending on the metal ions used, the catalytic activity of mACs can be inhibited, providing insights of the importance of the location and size of the metal ions. mACs are activated by Mn^{2+} and inhibited by high concentrations of Ca^{2+} and Zn^{2+} ions. The presence of Ca^{2+} has been shown to affect the binding of metal A/B ions and to inhibit mAC (Figure 11A)[78]. As can be observed in Figure 11A, Ca^{2+} ions stabilize the enzyme in a more open and inactive form. The larger van der Waals radius of Ca^{2+} ions disturbs the position of protein residues and hence the protein-substrate interactions in the catalytic site. Hydrogen bonds and van der Waals interactions essential for the catalysis of the cyclization of ATP are thus broken for residues R1029, N1025, D1018, I019, K065 (Figure 11B). The Ca^{2+} ions inhibit mAC by displacing metal A and interacting with β, γ phosphates and metal B. In such conformation, the smaller Mg^{2+} ion cannot bind anymore. Moreover, the second Ca^{2+} ion cannot participate in the nucleophilic attack of the group hydroxyl 3'OH of ATP, further blocking the catalysis. The steric effect of Ca^{2+} in the A metal position is thus at the origin of the inhibitory effect of Ca^{2+} ions on the catalytic reaction.

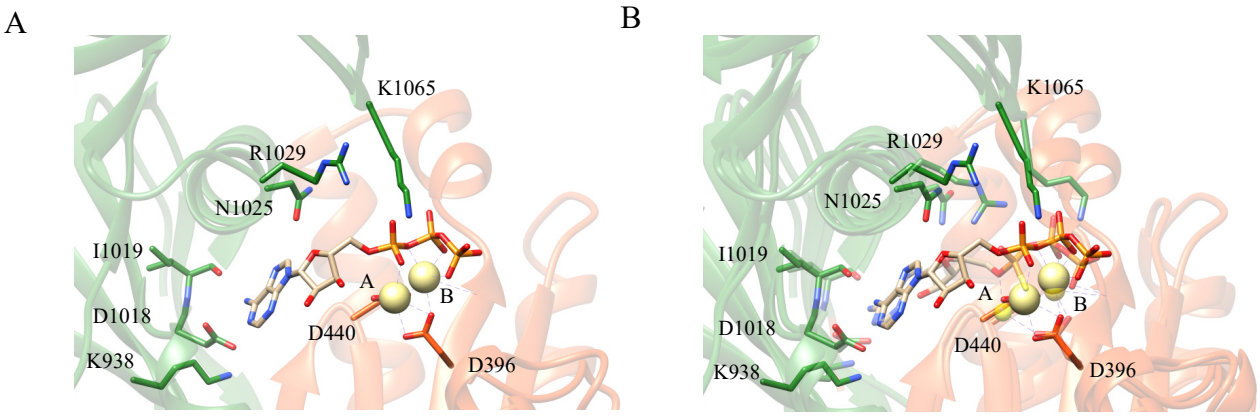


Figure 11- Catalytic site of mACs. *A. Crystal structure of mAC and ATP (tan sticks), in the presence of Ca^{2+} ions (gold spheres) (PDB ID 3C16) B. Superimposition of ATP in presence of Ca^{2+} ions (A) with $ATP_{\gamma S}$-Rp with two metal ions (A, Mg^{2+}, B, Mn^{2+}) (PDB ID 1CJK, lighter colors). The structural superimposition highlights the displacement of the residues K1065, R1029, N1025, that explains the inactivation of the adenylyl cyclase by calcium.*

The X-ray crystallographic structures in combination with biochemical approaches were crucial to unravel the molecular mechanism of action of mAC. The orientation of C1 towards C2 is the crucial step to activate mACs, generating a structural rearrangement of the catalytic site. Nevertheless, Tang and coworkers have emphasized that that EF C1/C2 display a relatively low sequence similarity with mammalian ACs[67].

A.2.2.2 Catalytic mechanism of EF

Like mACs, EF converts ATP into cAMP and the by-product PPi. EF is a very efficient enzyme that can convert ~ 1000-2000 molecules of cAMP per second[79,80]. This rate is higher than those of mACs[79,80]. To understand the catalytic mechanism of EF, one had to wait until 2001 when the first X-ray crystallographic structure was solved[28]. The structure of isolated EF corresponding to the inactive state of the enzyme is different from the structure of the active enzyme in complex with the EF activator calmodulin (EF-CaM)[28]. EF is composed of an N-terminal domain interacting with PA (PABD) and of a C-terminal domain (EF_3) divided into two domains C_A, C_B and a helical domain (HD) (Figure 12). The catalytic core is located at the interface of the C_A (in green in Figure 12) and C_B domains (in orange in Figure 12). Two loops in C_A are essential for the catalytic activity: switch A (SA, in purple in Figure 12) and B (SB, in cyan in Figure 12). A third loop, switch C (SC, in pink in Figure 12), connects the catalytic core to the helical domain. The helical domain (in blue in Figure 12) undergoes a significant displacement upon interaction with CaM (in red in Figure 12) that gets inserted between the

helical and C_A domains[28]. Upon CaM binding, SA and SC reorganize to stabilize SB and form the active catalytic site. The catalytic loop SB is highly flexible and dynamic, allowing the enzyme to efficiently produce cAMP molecules (> 1000 s^{-1}). The transition from an inactive towards an active catalytic site is different in EF and in mACs. Indeed, in contrast to mammalian ACs, the relative orientations of C_A and C_B domains are not modified during the activation by CaM. The mechanism of activation of EF by CaM is complex involving significant structural changes for EF as well as for CaM.

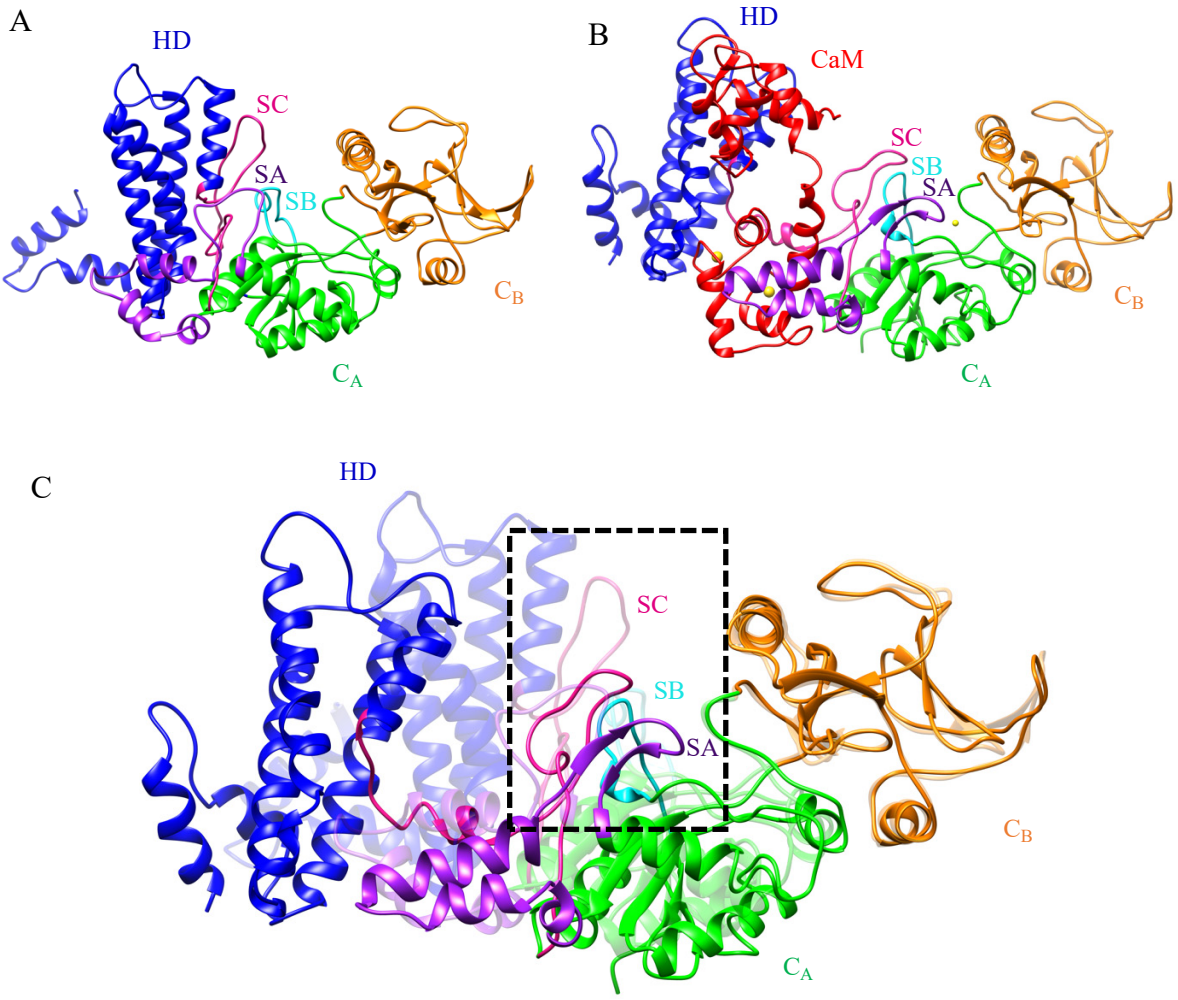


Figure 12- Structural changes during EF activation. *A. X-ray crystallographic structures of (A) inactive EF (PBD ID 1K8T) and (B) active EF in complex with CaM (PDB ID 1K93). and (C) superposition of the inactive (lighter colors) and active forms of EF displayed in A and B. The C_A and C_B domains of EF forming the catalytic core are represented in green and orange, respectively and the helical domain of EF is shown in blue. CaM (red) is inserted between the HD and C_A domains, causing the rotation and distancing of HD from C_A. The insertion of CaM also leads to a structural reorganization of EF switches A (purple), B (cyan) and C (pink). Switches C and A interact with CaM and stabilize the catalytic loop switch B.*

Another structure of EF-CaM[29], bound to the non-hydrolysable native substrate 3'dATP revealed that the substrate analog is coordinated to a single metal Yb^{3+} ion. EF residues D491 and D493 interact with the Yb^{3+} ion and with the α and β phosphates of 3'dATP (Figure 13). Residues R329, K346, K353, and K372 interact with the triphosphate chain of 3'dATP and N583 with the O4' oxygen of the ribose of 3'dATP. In contrast to mACs, EF uses a threonine residue (T548) instead of an aspartic residue to interact with the nitrogen N6 of the ribose of 3'dATP. Another specificity of EF is that H351 was proposed to be a catalytic base that deprotonates the oxygen of the 3' OH group at the difference with mACs for which the role of catalytic base is held by a metal ion. In agreement with this, other kinetic studies with EF mutants showed that residue H351 was essential for the catalytic activity[81].

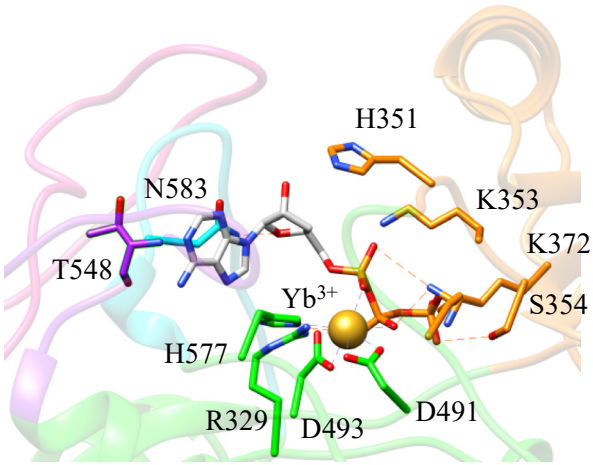


Figure 13- Catalytic site of EF. *ATP binding in the catalytic site in presence of Yb^{3+} (PDB ID 1K90). Domains are colored following the same code as in Figure 12, 3'-dATP is colored by atoms (grey for carbon, red for oxygen, blue for nitrogen and orange for phosphorous).*

A minor binding site of ATP was found in the crystal structure including adenosine 5′-(α,β-methylene)-triphosphate (AMPCPP)[82]. The binding modes of AMPCPP and ATP differ and display different orientations of the adenine and ribose cycles (Figure 14). The adenine cycle interacts with EF residues K382 and F586, instead of residues T548 and T579 as previously observed for ATP (Figure 14). The rotation of the adenine and ribose rings increases the distance between the 3'OH of AMPCPP and H351 and N583. In this conformation, the 3'OH of AMPCPP cannot attack the α-phosphorus of AMPCPP because the two groups are too far from each other. This alternative binding site offers new possibilities for developing bacterial adenylyl cyclase inhibitors.

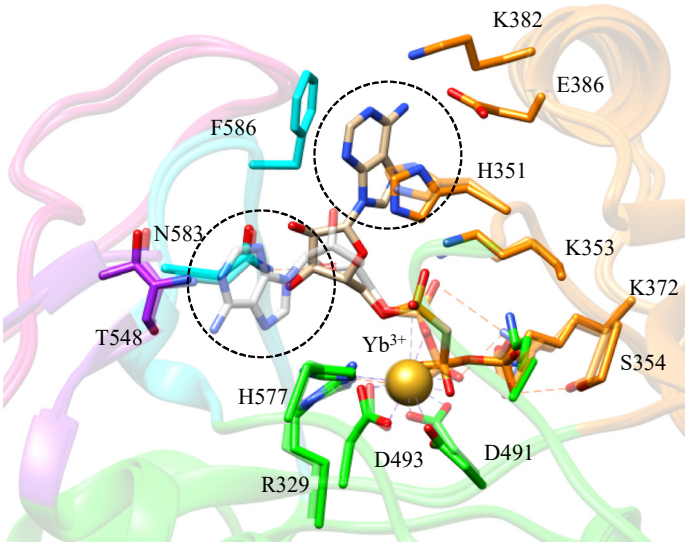


Figure 14- The two binding sites of ATP in the catalytic site. *Comparison of the classic ATP binding site (PDB ID 1K90, lighter colors) and the minor ATP binding site (PDB ID 1S26) observed for AMPCPP. In the alternative binding site, the base ring (indicated by the dashed circle) is rotated by 75 degrees and interacts with residues K382 and E586.*

As the catalytic mechanisms of mACs and DNA polymerase involved two metal ions, it was suggested that EF would also adopt a two-metal-ion mechanism. The crystal structure[83] of EF solved in the presence of the two products of the catalytic reaction, cAMP and PPi (Figure 15), would be in favor of this two-metal-ion process as in contrast to the previous structure solved with 3'dATP that showed only one metal ion in the catalytic site, the structure with cAMP and PPi displays two metal ions. The latter structure shows that the nucleophilic attack on ATP 3'-OH is a crucial step and suggests that the transition state would implicate a bipyramidal penta-coordinated phosphorus by an associative or concerted mechanism[83]. Both products cAMP and PPi are ~ 4.3 - 5.1 Å away from each other, metal ion A is coordinated to H577, D493 and 3'O of cAMP, and metal ion B to D491 and the phosphate chains of cAMP and PPi. PPi also interacts with K346 and K372[83] (Figure 15 B). The adenine ring of cAMP and ATP are in different orientations due to the rotation of the N-glycosidic bond[83]. The ribose adopts two different conformations for ATP (3'endo, *anti*-conformation) and cAMP (2'exo, *syn*-conformation)[83]. EF residues T548, G578, T579, D582 and N583 establish van der Waals contacts with the adenine ring of cAMP[83]. The N6 of cAMP interacts with T548 and the O4' of cAMP with N583[83]. The release order of the products is different from that of the mACs because, in the case of EF, PPi is released first[83].

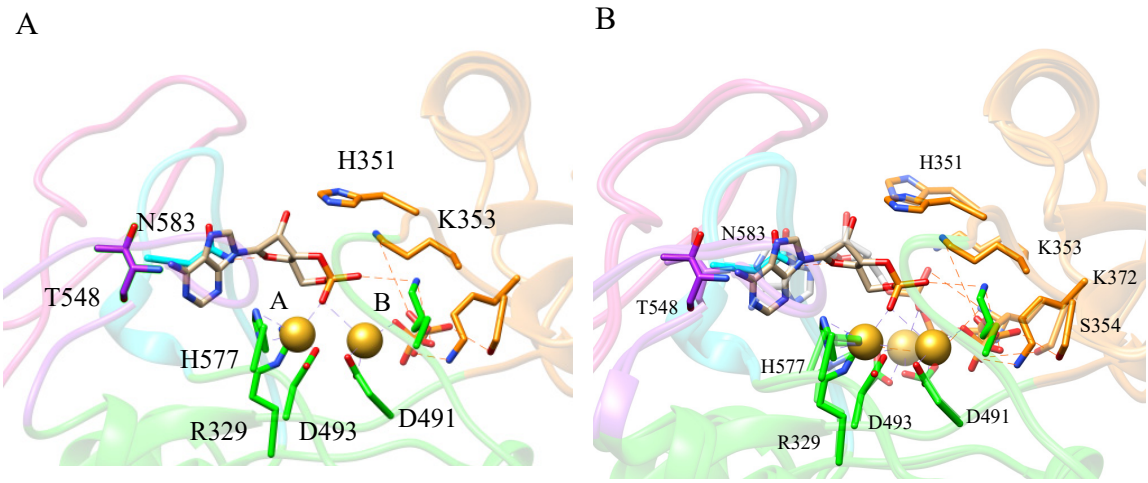


Figure 15- Catalytic site of EF. *A. The catalysis products cAMP and PPi (sticks) in presence of two Yb^{3+} ions are shown as spheres (PDB ID 1SK6). B. Superposition of EF-CaM-3'dATP (see Figure 13) and the structure shown in A. The base ring is oriented differently in both structures and PPi is farther from cAMP (ca. 4.3-5.1 Å) in the two metal ions structure. The domains are shown following the same colors as in Figure 12. PPi atoms interact with S354, K372 and K346. The adenine ring of cAMP is rotated 160 degrees relative to the adenine ring of 3'-dATP.*

In 2005, a new structure helped to establish that the two-metal-ions catalytic mechanism is possible and to redefine the role of H351[65]. In the X-ray structure of EF-CaM-3'dATP, two Mg^{2+} are present (Figure 16A). The metal ion A is coordinated to D491, D493 and His 577 and is close to the 3′-OH of ATP, which facilitates the deprotonation of ATP 3'-OH. Aspartic residue 493 and the 3'-dATP α oxygen of the triphosphate chain, both interact with metal ion B[65]. In the EF-CaM-cAMP crystal structure, residue H351 (first hypothesized to play a direct base catalysis role) is located 6 Å away from the 3'-OH group of cAMP, indicating that this residue cannot be a catalytic base[65]. This structure further reveals that residue N583 of EF is essential for positioning the 3'-OH group of the ligand for its nucleophile attack. The role of H351 is either to deprotonate a water molecule that then would attack the 3'OH or to stabilize an OH^{--} ion close to the 3'-OH group[65]. The combined action of metal ion A and H351 is essential for the nucleophilic attack.

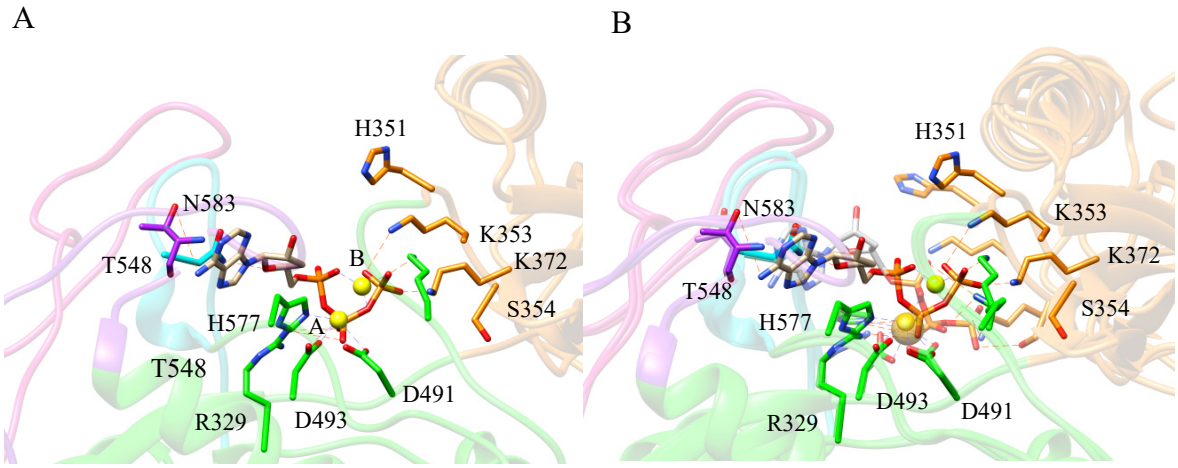


Figure 16- Two active metal ions in the catalytic reaction. *A. Two-metal-ion structure (PDB ID 1XFV). The two Mg^{2+} yellow spheres involved in the catalytic reaction of ATP are shown as spheres and 3'dATP is drawn as sticks. B. Superposition of a two-metal-ion (shown in A) and of a one metal ion structure (PDB ID 1K90, see Figure 13A). The rotation and position of residues are more impacted in the C_B domain (orange) than in the C_A domain (green) by the number of metal ions in the catalytic site.*

A molecular modeling (MD) approach was used to study the one and two metal ion catalytic mechanisms[84]. In the presence of a single ion, ATP was strongly bound suggesting that it would be the first step of the catalysis. An additional basic residue present in EF would help in the hypothesis of a single ion. The phosphate tail of ATP was more extended for the nucleophilic attack. For a two-metal-ion mechanism, both products cAMP and PPi were tightly bound and the nucleophilic attack was favorable. A two-metal-ion process would be the final step in order to stabilize the transition state. Moreover, a high Mg^{2+} ion concentration inactivates EF and the X-ray crystallographic structure with two Mg^{2+} were obtained in this condition. Nevertheless, it is still not clear whether the catalytic mechanism involves one or two ions or both. From MD simulations, a one ion process would be preferred[84,85]. The dissociation of PPi requires more force than that of cAMP as it has also been shown experimentally[74]. The protonation state of PPi would affect the dragging forces, the ionic interactions and hence the dissociation of the product[85,86].

A.2.2.3 Comparison of the catalytic mechanisms of EF, CyaA and ExoY

Like EF, the adenylyl cyclase toxins CyaA from *Bordetella pertussis* and ExoY from *Pseudomonas aeruginosa* are type II ACs and as such are activated upon interaction with

specific cofactors. CyaA and EF are activated by CaM, whereas ExoY requires filamentous actin (F-actin)[87]. Their modes of host cell intoxication are different: while CyaA is directly translocated across the membrane[80,88], ExoY is translocated via a type III secretion system[89] and EF via endocytosis. The sequences of the N-terminal region of EF and CyaA ACD domains show 34% identities[90]. This region displays 29% identity for EF and ExoY[90]. Furthermore, the three proteins share a similar ATP-binding motif suggesting a common catalytic mechanism[65,87,91]. Nevertheless, ExoY exhibits a preference for guanosine 3′,5′-cyclic monophosphate (cGMP) and uridine 3′,5′-cyclic monophosphate (cUMP) rather than for cAMP and nucleoside 3′,5′-cyclic monophosphate (cNMP), which are the preferred substrates of EF and CyaA[92]. The global structure of ExoY, CyaA and EF present similarities (Figure 17)[93]. In the three adenylyl cyclases, one asparagine residue (N583 for EF, N304 for CyaA and N297 for ExoY) is important for substrate binding and speeds-up the catalytic reaction[94]. The catalytic mechanism of CyaA involves a two-metal ion mechanism, with residues D188, D190 and H298 coordinating the two metal ions, R37, Y58 Y65 and Y84 interacting with the phosphate chain of ATP and N304 with the ribose 3'O of ATP. The catalytic site of CyaA is larger than the EF one and accommodates a third metal ion located far from ATP and is believed to maintain structurally the catalytic site of CyaA[91]. In the case of ExoY, it is not possible to establish if catalysis requires one or two metal ions due to the unavailability of ACD structures in complex with its substrate[93]. The residues playing the role of H351 for EF are H63 for CyaA and Y86 for ExoY. Noteworthy, the structural differences of the C-terminal region of ACD domains evidence the different modes of activation of the three enzymes (Figure 17).

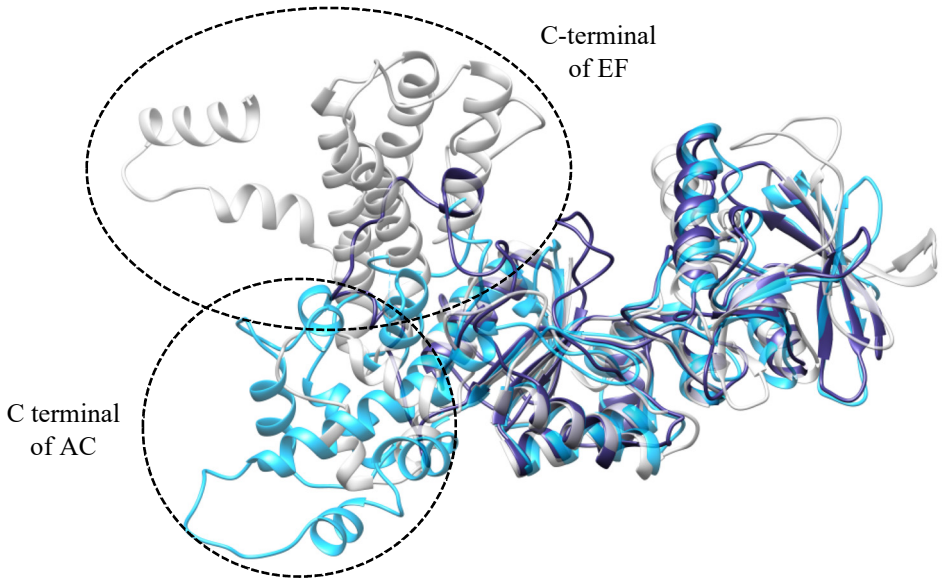


Figure 17- Structural comparison of EF, CyaA and ExoY. *EF (PDB ID 1K8T) is represented in light grey, CyaA (PDB ID 2COL) in deep sky blue and ExoY (PDB ID 5XNW) in dark slate blue. The three C-terminal regions of the ACD are different, reflecting a unique activation mechanism for each protein (dashed circle).*

A.2.2.4 Activation of EF by CaM

EF is activated by CaM, a small ubiquitous protein expressed in all eukaryotic cells. CaM binds Ca^{2+} ions and is involved in a Ca^{2+} dependent manner, in numerous cellular processes such as cell signaling, signal transduction and regulation of cytoskeletal elements. X-ray crystallographic structures of Ca^{2+}-CaM show that CaM is composed of two globular domains separated by an interconnecting helix[95,96]. In the Ca^{2+}-CaM structure, two Ca^{2+} are bound to each lobe and the loading of Ca^{2+} ions induces the appearance of hydrophobic patches[95] in which mostly methionine (M36, M51, M71, M72, M76, M109, M124, M144 and M145) and phenylalanine (F19, F68, F92 and F141) residues are implicated[95]. The binding sites of Ca^{2+} are helix-loop-helix motifs, called EF-hands. The N-terminal lobe (N-CaM) is composed of EF-hands EF-1 (helices A and B) and EF-2 (helices C and D) and the C-terminal lobe (C-CaM) of EF-hands EF-3 (helices E and F) and EF-4 (helices G and H). Solution Nuclear Magnetic Resonance (NMR) studies of CaM revealed a very flexible linker between the two lobes implying that the first determined X-ray structure was not representative of CaM in solution[97]. The structure of apo-CaM, without Ca^{2+} ions, was also solved by NMR[98]. Holo-CaM, loaded with four Ca^{2+} ions, displays a more compact shape than apo-CaM. In apo-CaM, N-CaM is in a closed conformation with small angles between the helices, whereas C-CaM is

in a semi-open conformation with larger angles between the helices[99]. This observation suggests that the C-terminal lobe has a higher affinity for Ca^{2+} ions than the N-terminal lobe[100], because in holo-CaM, both C-CaM and N-CaM adopt an open conformation. Binding of Ca^{2+} ions increases the angles (close to 90 degrees) and the distances between the helices within a given EF-hand[98], resulting in a larger accessibility of the hydrophobic patches described above and thereby permitting CaM to interact with a wide range of signaling peptides[101,102,103,104]. The two lobes are engaged in the interaction with the target peptides and form a hydrophobic tunnel[98]. The peptides form an amphiphilic α-helix, which is the canonical recognition motif for binding to CaM[101,105]. Nevertheless, different binding modes involving an extended conformation of CaM have been observed for various interactions, such as the interactions with ionic pumps and channels[106,107,108] and the EF-CaM complex[90].

Hydrophobic and hydrophilic contacts are observed between EF and CaM. These involve residues in EF switch A (L501, K505, K525, Q526), in the switch C linker region (R540, R630, D647), which is disordered in the EF inactive state, and in the HD domain (A661, S668, K671, R691). N-CaM interacts with the helical domain and C-CaM with the catalytic core and helical domains.

In the EF-CaM complex, CaM is loaded with only two Ca^{2+} ions binding to the C-terminal lobe[90]. The two CaM lobes are essential for EF activation[90] and protect the C-terminal of EF against digestion by proteases[90]. The mechanism of activation of EF by CaM was clarified by using NMR[109]. ^{15}N- and ^{13}C-labeled CaM was titrated with Ca^{2+} ions and in the absence or presence of EF in order to probe the dynamics of the interaction. The N-terminal lobe of apo-CaM is able to bind EF, whereas the C-terminal lobe of CaM requires to be calcium-loaded to interact with EF, in agreement with the X-ray structure of EF-CaM[28]. Moreover, the level of Ca^{2+} ions affects the EF catalytic activity, with low levels being required for EF activation by CaM and high levels inhibiting the catalysis[24]. This calcium dependence may be (at least partially) rationalized by the higher calcium affinity of the C-lobe and the requirement of an open and closed conformation of the C- and N-lobes, respectively.

The stability and dynamics of EF-CaM have been investigated by MD using Ca^{2+} as the variable parameter[111]. The Ca^{2+} ion level affects the energetic contributions of CaM, EF and EF-CaM[24,112]. Three levels of Ca^{2+} were characterized: 4-Ca^{2+}, 2-Ca^{2+} and 0-Ca^{2+}. In the 4-Ca^{2+} and 0-Ca^{2+} simulations, the N-CaM was partially open and the bending of the linker of CaM was affected. The removal of Ca^{2+} affected both EF and EF-CaM stability, whereas in the 4-Ca^{2+} simulation, the stability of EF was not impacted. In the 2-Ca^{2+} simulation the two Ca^{2+} ions were located in C-CaM and the hydrogen bonds between N-CaM and HD of EF observed

in the structure were mainly present but absent in the case of the 0-Ca^{2+} and 4-Ca^{2+} simulations. The most stable conformation was found for the 2-Ca^{2+} simulation. In this simulation, the N-CaM was blocked in an intermediate conformation that depended on a network of residues: the interface region between EF (helical domain, SA) and CaM and the catalytic region (C_A, C_B) of EF. Without calcium (0-Ca^{2+}), no communication through this network was observed, while for 4-Ca^{2+}, the network was destabilized. These results emphasize that the mechanism of binding is different between the two CaM lobes. The C-terminal lobe interacts with EF in a $^{Ca2+}$ dependent way whereas this is not the case for the N-terminal lobe at low/moderate calcium concentrations.

The mechanistic model proposed by Ulmer and coworkers[109] is depicted in Figure 18. N-CaM first interacts with the helical domain of EF and the 2Ca^{2+} present in the N-CaM are expulsed. This interaction thus locks N-CaM in a closed conformation, and finally, holo C-CaM rotates and inserts between the C_A and helical domains in an open conformation[17,79,111].

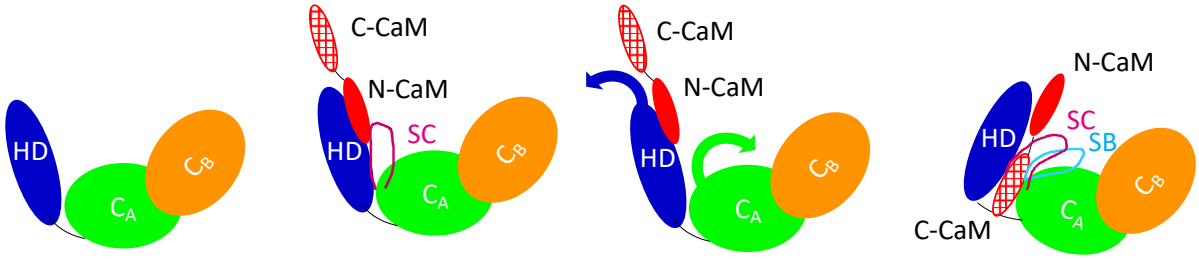


Figure 18- Scheme of the mechanism of activation of EF by CaM. *The EF domains are represented in green (C_A), orange (C_B) and blue (HD) and the CaM lobes are displayed in hatched-red (C-CaM) and red (N-CaM). Binding of N-CaM to HD triggers large structural rearrangements in HD and switch C (pink) leading to the insertion of the C-CaM lobe between HD and C_A. The catalytic loop SB (cyan line) is then reorganized and stabilized by SC.*

A.2.2.5 Comparison of the activation modes of EF and CyaA by CaM

As CyaA and EF are homologous bacterial adenylate cyclase toxins activated by the same host activator, CaM, in this section, we compare their activation processes.

CaM binds to the C-terminal region of both proteins ACDs[91] but establishes different interactions with each protein and the nature of the critical residues for the interaction are different (Figure 19). The catalytic domain of CyaA is hereafter named AC. C-CaM (in red in Figure 19) binds to EF switch A (SA, in purple in Figure 19) and to CyaA F, and H' helices (in purple in Figure 19). The essential residues are W242 and E526 (in black sticks in Figure 19) for the CyA and EF interactions with CaM, respectively[91]. Furthermore, while for the activation

of EF both the N- and C-terminal lobes of CaM are needed, CyaA can be activated by only one lobe[110]. Importantly, for both proteins, CaM binding stabilizes a loop responsible for their activity: switch B (H577-E592, in cyan in Figure 19) for EF and T300-K312 in CyaA (in cyan in Figure 19). Only scarce structural data for the N-CaM/CyaA interaction are available. First, a β hairpin region (residues 259–273) of CyaA seems to interact with N-CaM[113]. As no structural data of CaM with the full CyaA ACD is available, the mechanism of activation is still unclear. Of note, the affinity for CaM is higher for CyaA than for EF[110].

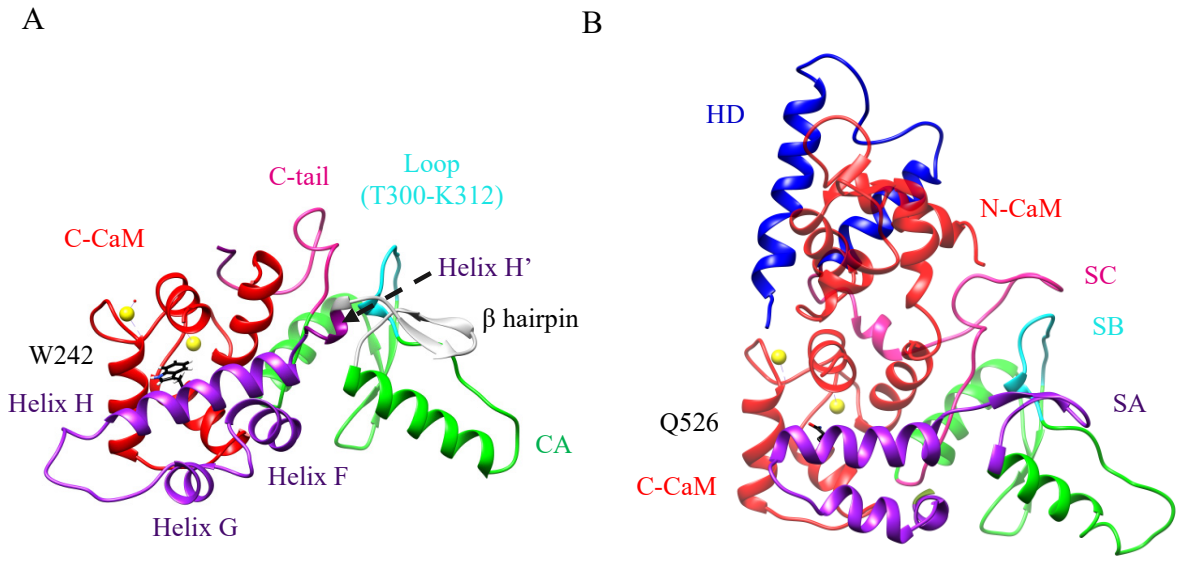


Figure 19- Interaction of CaM with CyaA and EF. *CyaA (A) and EF (B) interaction with CaM. CaM is shown in red, and the calcium ions are displayed as yellow spheres. In A, C-CaM interacts with the C-tail (pink), the H/H' and F α-helices of AC (purple) and N-CaM interacts with a ß hairpin (grey). These contacts with CaM trigger the stabilization of the catalytic loop (cyan) located in the CyaA CA domain (green). In B, N-CaM interacts with the HD of EF (blue) and C-CaM with switches SA (in purple) and SC (pink) of the CA domain (green). Insertion of CaM stabilizes the SB loop (cyan) involved in catalysis. Switch A of EF and H/H' and F α-helices of CyaA ACD seem to make similar contacts with C-CaM. EF SC switch and CyaA C- interact with the catalytic loops of EF and CyaA, respectively.*

Bioinformatic studies confirmed the energetic difference between EF/CaM and AC/C-CaM by showing differences in the superimposition of the influence diagrams of EF/CaM and AC/C-CaM due to the movement of HD of EF by the insertion of CaM[114,115]. The HD is not present in the structure of CyaA. In the structure of C-CaM/CyaA, the α helix H of AC is in direct interaction with CaM. The surface of interaction between adenylyl cyclase domain and CaM is larger for EF/C-CaM than for AC/C-CaM. The interaction between C-CaM and CyaA would be established through the hydrophobic patch of C-CaM[116,117]. In MD simulations, the suppression of Ca^{2+} affects the hydrogen bonds network from C-CaM to the AC catalytic site

and disturbs in that way the function of AC. Mutations of these residues show the perturbation of the catalytic activity of AC. And also, the SA loop not visible in the X-ray crystallographic structure of CyaA, is highly flexible during molecular dynamics simulations recorded in the absence of Ca^{2+} ions. Synchrotron radiation circular dichroism (SR-CD), hydrogen/deuterium exchange mass spectrometry (HDX-MS) and small angle X-ray scattering (SAXS) indicate that a 75 residues loop of AC undergoes a disorder to order transition upon CaM binding[118]. The SA domain of EF is much shorter than the one of CyaA. For bacterial AC toxins, the activation takes place through the stabilization of several catalytic loops, whereas for mAC the juxtaposition of C1 and C2 is responsible for the activation. As the catalytic site is different in bacterial AC toxins and mACs, it is relevant to develop specific inhibitors of the bacterial AC catalytic site.

The use and improvement of experimental structural biology techniques coupled to biochemistry and in interplay with computational approaches such as molecular dynamics simulation, have enabled to better understand the mode of action of AC toxins and enabled studies to find new ways to target EF and anthrax

A.3 Identification of ligands targeting EF

In order to discover and develop inhibitors of EF, three strategies have been explored[119]: (i) to block the entry of toxins into host cells, (ii) to prevent the activation of EF by host CaM and (iii) to inhibit EF enzymatic activity.

Blockade of toxin entry strategies targeting PA was followed to develop preventive and therapeutic approaches. Vaccines and antibodies as well as antibiotics were developed but produced side-effects due to high reactogenicity and antibiotic resistance, respectively[120,121]. These treatments are administered to high-risk population and in the early stage of infection. The problems encountered when applying this strategy, prompt for the need to find an efficient and rapid way to block anthrax infection and toxicity.

More recently, investigations aim to neutralize EF with monoclonal antibodies (mAb) by hijacking the protein in the blood[122,123]. For this purpose, Chen and coworkers developed a phage display library of antibody fragment antigen-binding regions (Fab) clones against EF[122]. The library was created from lymphocytes elicited from chimpanzees immunized against EF. The EF-binding selected Fab fragments were then converted to humanized monoclonal IgG

antibodies[122]. Among the four developed antibodies that specifically target EF, EF13D is the most promising[122]. EF13D inhibits the production of cAMP *in vitro* and *in vivo* by ET by interacting with the helical domain of EF, thus preventing the binding of CaM. EF13D is able to separate CaM from EF, because it binds to EF with a 50- to 130-fold higher affinity to CaM than EF (Figure 20)[122].

Based on this immunologic strategy, a phage-display synthetic Fab library was recently developed and two other antibody Fabs towards EF, named A4 and B7, were identified[123]. As these Fabs establish only a monovalent interaction with EF, a C-terminal linker was added to A4 and B7 to obtain bivalent interactions[123]. The antibodies B7 and EF13D are both competitors of CaM binding to EF. A4 and B7 bind to different sites on EF. Although B7 displays more potency than A4, using a mixture of A4/B7 could be a good alternative for inhibiting EF, because B7/A4 Fab mixtures were very effective at blocking the entry of EF into anthrax-sensitive human COS cells. Although EF13D mAbs and B7 or B7/A4 Fabs are able to inhibit the catalytic activity of EF *in vitro*, in vivo they probably exert their neutralizing activity by blocking the entry of EF into cells before endosomal engagement[123].

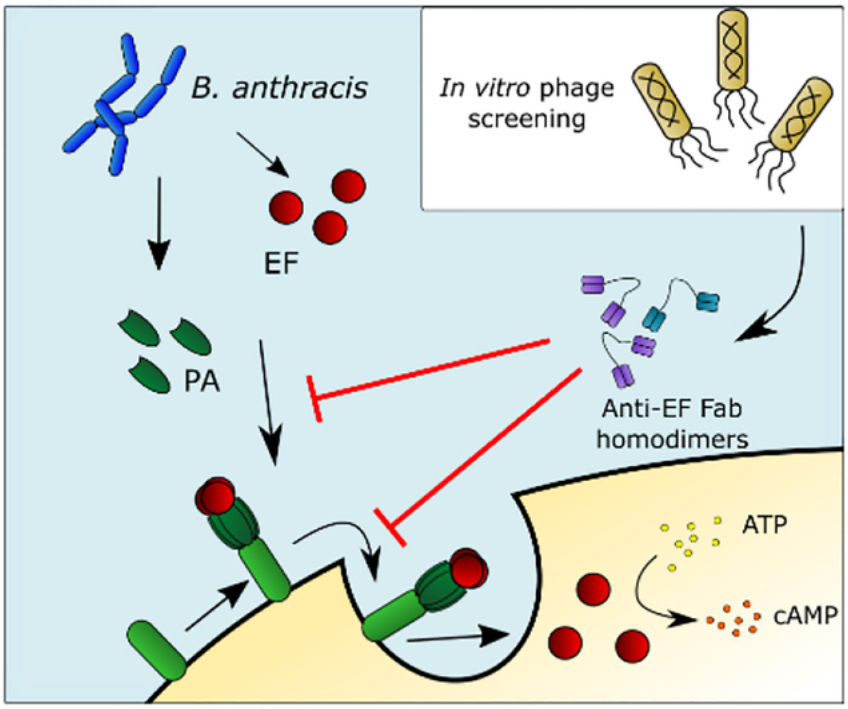


Figure 20- Scheme of the inhibition mechanism of Fabs toward B. anthracis toxins from Farcasanu et al[123]. *Monoclonal antibody Fab fragments hijack EF and thus prevent its endocytosis and its activation by CaM in the cytosol.*

In the following section 3.1, we present the results obtained in the literature following the strategies of inhibition of EF targeting it in the cytosol either by preventing its activation by host CaM or directly inhibiting its enzymatic activity. Both strategies have been applied by targeting orthosteric or allosteric sites. An orthosteric site is a functional site where the endogenous substrate or ligand binds. For an enzyme, the orthosteric site is the catalytic site. An allosteric site is by definition a site other than the site binding the endogenous ligand. Targeting the allosteric site with a ligand can trigger long distance inactivation of enzyme function. We first present, the inhibitors binding to the catalytic site of EF and then we focus on the allosteric inhibitors of EF.

A.3.1 Inhibitors of EF catalytic site

A.3.1.1 Ethyl 5-aminopyrazolo[1,5-α]quinazoline-3-carboxylate

Along with having solved several EF-CaM X-ray crystallographic structures, the team of W. J. Tang was the first to report inhibitors of EF by using an *in silico* structure-based method[124]. Their strategy to inactive EF was to target the catalytic site of EF. As the catalytic site of EF is different from that of human mACs, they were expecting to obtain inhibitors specific of EF. About 200,000 commercially available molecules were docked into the catalytic site and the 19 top-scoring compounds were selected for *in vitro* inhibition assays and to assess their effect on the morphology of Y1 cells. A family of quinazoline compounds showed specific inhibition of EF and CyaA. Among them, ethyl 5-aminopyrazolo[1,5-α]quinazoline-3-carboxylate (Figure 21) was the most promising one ($K_i = 20$ μM)[124]. Docking of this compound on the EF-CaM-3'-dATP structure revealed that the three cycles of the molecule overlay with the adenosine group of 3'-dATP and, like 3'-dATP, interact with residues T548 and T579. This ATP competitor inhibits the morphological change in Y1 cells and consequently the intoxication process but displays poor selectivity because it inhibits mACs I, II and V[124]. By modifying the ethyl ester group of the molecule (dashed circle in Figure 21), with a secondary amide, the selectivity for EF and the solubility of the compound were improved[124].

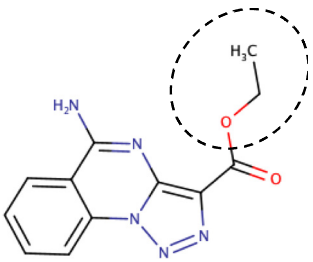
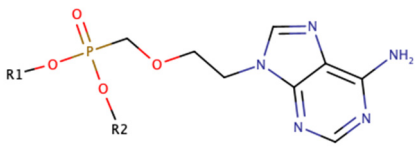


Figure 21- Ethyl 5-aminopyrazolo[1,5-α]quinazoline-3-carboxylate.

A.3.1.2 Adefovir and PMEApp

Adefovir dipivoxil (bis-POM-PMEA, Figure 22), an active, clinically approved viral drug against hepatitis B[125,126], revealed to be also active on EF[127]. Adefovir dipivoxil (PMEA) is metabolized by cellular kinases *in situ* into the active adefovir diphosphate, PMEApp. PMEApp inhibits very efficiently the *in vitro* production of cAMP (K_i = 27 nM). The prodrug Adefovir dipivoxil prevents the intoxication by ET by reversing the down regulation induced by EF on the production of cytokines[128] and reduces anthrax-related hypotension and lethality in mice[129,27]. This non-cyclizable ATP analog competes for the same site with the endogenous substrate ATP but displays a higher affinity for EF than ATP (K_m = 168-194 µM) due to supplementary van der Waals interactions and hydrogen bonds with N583 and H577 with EF, while keeping favorable negative charges. PMEApp revealed also to be active on the catalytic site of CyaA (K_i = 25 nM). However, PMEApp has shown an inhibition of mACs only 10 to 500 fold weaker than for EF[127,130], leading to potent side effects linked to kidney[131,132]. Recently, other derivatives have been tested in order to lower the cytotoxicity and to improve the half-life of the molecules in plasma[133].



PMEA: R1 = R2 = H
bis-POM-PMEA: R1 = R2 = $CH_2OC(O)C(CH_3)_3$
PMEApp: R1 = PP, R2 = H

Figure 22- Adefovir compounds.

A.3.1.3 3-[(9-oxo-9H-fluorene-1-carbonyl)-amino]-benzoic acid

A virtual screening based on the assembly of fragments to build 3D pharmacophores was used to find EF active site inhibitors[134]. Although also displaying good scores for mACs, 19 compounds with the highest Autodock[135] docking score towards EF were selected for further cell assays. Some compounds were able to reduce the production of cAMP by ET in RAW 264.7 cells[136], among them, three molecules had IC_{50} values lower than 10 µM. Compounds were not tested on purified EF or on mACs. ETEC-infected mice, which produce an adenylyl cyclase toxin that has a high degree of identity with EF[137], were treated with 3-[(9-oxo-9H-fluorene-1-carbonyl)-amino]-benzoic acid (DC5, Figure 23). The treatment prevented diarrhea and intestinal edema[136,138]. Furthermore, the high mutagenic toxicity risk of the compound predicted by the OSIRIS program[139] was reduced by modifying the benzoic acid moiety, while keeping the affinity for EF[140]. For example, the N-(3-carbamimidoylphenyl)-9-oxo-9H-fluorene-1-carboxamide showed the same affinity than DC5 but a lower cytoxicity. In addition, this compound offers large possibilities of modulation in order to increase its affinity[140].

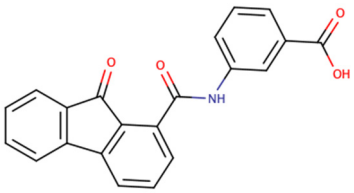


Figure 23- DC5: 3-[(9-oxo-9H-fluorene-1-carbonyl)-amino]-benzoic acid.

A.3.1.4 P-site inhibitors

Nucleotide-based EF inhibitors have also been explored. The first idea to inhibit EF was to use compounds already known to target mACs, like the P-site inhibitors already described in section A.2.2.1. P-site inhibitors are purine nucleotides with a phosphate or polyphosphate at the 3'-O-ribosyl position. These P-site inhibitors were effective on mACs but not on EF, confirming that the catalytic site and activity of EF and mACs are different. For example, 2′-d-3′-AMP (Figure 24) displays a higher affinity for mACs than for EF[141,142]. The 2′-d-3′-AMP is a competitive inhibitor of ATP for EF and a non-competitive inhibitor of ATP for mACs.

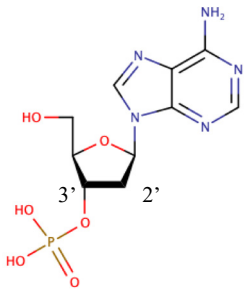


Figure 24- the P-site inhibitor 2'-d-3'-AMP.

A.3.1.5 (M)ANT-nucleotides

The addition of an anthraniloyl (ANT) or a N-methylanthraniloyl (MANT) group at the 3'-O-ribosyl position of ATP was found to be important for the inhibition of EF. The 2'-d-3'ANT-ATP was the first reported compound of the (M)ANT family inhibiting the active site of EF in the presence of Mg^{2+} (K_i = 10 μM)[143]. Subsequently, different compounds of this type have been developed and tested. By modifying the chemical properties on the 2' or 3'-O-ribosyl position on the native substrate ATP, the selectivity and affinity towards EF were considerably improved[144]. Moreover, 2′(3′)-O-ribosyl-substituted MANT-nucleotides are able to monitor the conformational changes in the catalytic site of mACs and EF due to their intrinsic fluorescence. EF and mACs show different sensitivities to pyrimidine-based MANT compounds, further emphasizing the differences between the catalytic sites of these enzymes[145]. Indeed, on the one hand, MANT-GTP (described previously) has a lower potency inhibition toward EF (μM range inhibition) than mACs (nM range inhibition), and on the other hand, unlike mACs, EF is sensitive to the pyrimidine base cytosine[146], with MANT-CTP (K_i = 100 nM, Figure 25 A), together with the adenine-based propyl-ANT-ATP (K_i = 80 nM, Figure 25 B) being one of the most effective inhibitors of EF[143]. The Bis-MANT molecules are less selective than mono-MANT due to a steric hindrance[147]. The most potent inhibitor of EF catalytic site known to date is 3′-MANT-2′-dATP (Figure 25, right). The hydrophobic interaction established by the N-methyl group of this compound is crucial for its high affinity (K_i = 10 nM in presence of Mn^{2+})[148,149]. This compound shows a ~ 5 - 50-fold higher affinity for EF than for mACs.

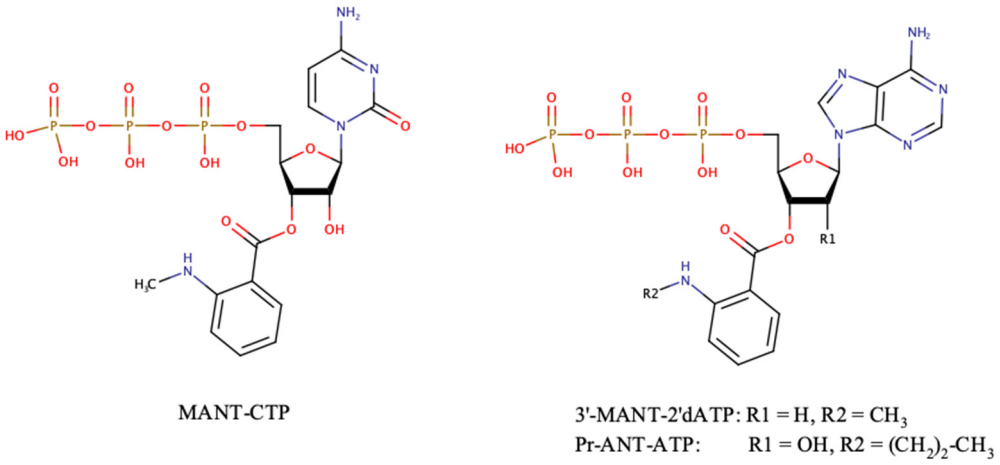


Figure 25- (M)ANT structures inhibitors.

A.3.4.5 5'-p-fluorosulfonylbenzoyl 5'-adenosine

A novel nucleotide-based family of compounds was designed recently[150] from observing the EF amino acid residues mediating the interaction between 3'-dATP–and EF-CaM. Three lysine residues were identified to interact with the phosphate chain in the structure of the complex. The key idea to obtain a molecule that could interact with the lysine residues was to introduce an electrophile group in the molecule. To this end, the authors[150] chose a fluorosulfonyl group. The compound 5′-*p*-fluorosulfonylbenzoyl adenosine (FSBA) irreversibly inactivated EF by establishing a covalent bond. FSBA was then adapted to the target, by modulating the nucleotide, the electrophilic group and the linker. Bicycles such as pyrimidine demonstrated to be important for the inhibition. The 4-(4-(4-amino-7H-pyrrolo[2,3-d]pyrimidin-7-yl)piperidine-1-carbonyl)benzenesulfonyl fluoride trifluoroacetate (Figure 26) is the most active (EC_{50} = 12.0 nM, K_i = 0.3 μM) compound of this series. The electrophile group needs to be further investigated as it shows low metabolic stability. Furthermore, selectivity and affinity have still to be improved.

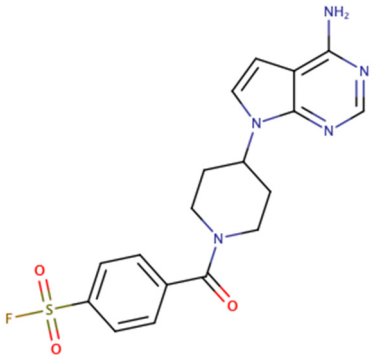


Figure 26- 4-(4-(4-amino-7H-pyrrolo[2,3-d]pyrimidin-7-yl)piperidine-1-carbonyl)benzenesulfonyl fluoride trifluoroacetate.

A.3.2 EF Allosteric inhibitors

In some cases, it is not suitable to target the catalytic site, because of specificity problems. Indeed, many enzymes with similar functions can have analogous active sites. For example, GPCRs, kinases and ion channels ATPases share a highly conserved/ATP site[151,152]. Therefore, targeting the orthosteric site of ATPases leads to side effects and drug resistance. In the case of EF, although possible, targeting the catalytic site makes it difficult to obtain compounds with very high potency and selectivity for EF relative to mACs. Considering that mACs and EF are structurally very different proteins outside of the catalytic domain, another rational approach for drug development is the design of ligands that target unique binding sites present in EF but not in mACs. Finding allosteric inhibitors is a different way to understand the mechanism of action of EF at a molecular level and to develop new therapeutic approaches.

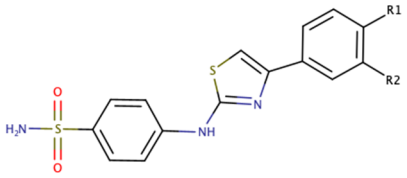
Allostery is an essential biological process, in which the perturbation of a binding site induces conformational[153,154,155,156,157] and/or dynamical changes of a distal functional site and modulates the affinity and activity of the macromolecule[158,159,160]. Initially proposed by the seminal work of Monod, Wyman and Changeux[153,154], the concept of allostery has been studied for many years along three main axes: thermodynamic equilibrium, free energy landscape and structure[161,162,163,154]. Allostery can be considered as an intrinsic property of all proteins[164]. It is nowadays generally admitted that protein function is achieved by an interplay between structural and dynamical properties of the system[165,166,167,168]. Therefore, protein function is usually modulated by the surrounding environment (binding to another protein, pH, ionic strength, temperature or covalent modification)[156,169,170,171]. These events can alter the internal dynamics of biomolecules with or without inducing large conformational changes[172,173]. As the direct experimental observation of allosteric signaling is difficult, computational methods[174]

have been used to study and predict the communication between the active and allosteric sites through networks of amino acid and the impact on protein function[175,176]. Theoretical predictions can then be probed experimentally by investigating how the protein function is affected by the mutation of those residues. The identification of allosteric sites is crucial to understand the mechanisms of action of the target protein, molecular interactions and to find new modulation strategies for drug discovery[169,177,178]. Such modulators can selectively increase or decrease the activity of the target. Allosteric drugs can be specific and avoid off-target side effects[179,180]. However, it is not trivial to determine whether a particular region of a protein can be an allosteric site and can affect the function of the target upon ligand binding. Once an allosteric site is identified, the challenge is to find molecules that specifically bind to that site.

Here below, we present some reported inhibitors exhibiting allosteric modulation of the activation of EF by CaM.

A.3.2.1 4-[4-(4-Nitrophenyl)-thiazolylamino]-benzene-sulfonamide

The aim of this study is to discover a new class of inhibitors targeting the protein-protein interactions (iPPIs), thus would prevent the association of EF with CaM. A chemical library of 10 000 compounds using a Y1 cell-assay approach was screened. Twenty-four ligand candidates were identified and tested by Surface Plasmon Resonance[181]. In this test, CaM was immobilized and the association between EF and CaM followed. The molecule 4-[4-(3,4-dichlorophenyl)-thiazolylamino]-benzenesulfonamide (10506-2A, Figure 27) showed to be an iPPI preventing (i) EF-CaM association (IC_{50} = 10-15 μM) and (ii) the morphological change of Y1 cells. The binding site was identified by fluorescence and photolabeling. Compound 10506-2A binds at the interface between the active site and the helical domain of EF. However, 10506-2A was cytotoxic. In order to decrease the cytoxicity, the 3,4 di-chloro group was replaced by a nitro group to produce the compound nitro-10506-2A (Figure 27). Unexpectedly, due to the hydrophobicity of the two molecules, nitro-10506-2A and 10506-2A were also able to bind CaM although they displayed a higher affinity for EF than for CaM. Noteworthy, in the literature, it is relatively common for small hydrophobic molecules designed or screened to inhibit proteins activated by CaM, to also unexpectedly bind to CaM[182,183].



10506-2A: R1 = R2 = Cl
nitro-10506-2A: R1 = NO_2, R2 = H

Figure 27- 10506-2A and nitro-10506-2A.

A.3.2.2 CaM inhibitors

As an alternative to inactivate EF, it is possible to target its cellular activator, CaM. In that sense, several CaM-inhibitors were also tested. The CaM inhibitors CMZ (calmidazolium chloride)[184], TFP (trifluoperazine)[185] and W-7 [N-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide hydrochloride][186] exhibited inhibition of proteins activated by CaM. CMZ has a higher affinity for holo-CaM (nanomolar range) than TFP and W-7 (micromolar range)[187]. Among the three compounds, CMZ (Figure 28) is the most potent inhibitor of EF and CyaA[188,189,190]. Its inhibition is through a CaM-independent allosteric mechanism (20% inhibition with 10 µM of the compound for EF). CMZ cannot be considered as an inhibitor due to selectivity (it also inhibits mAC1 and mAC8)[188], potency and toxicity[191] problems. A general drawback of the development of CaM inhibitors is that other proteins regulated by CaM can be unintentionally perturbed, resulting in problems of selectivity and concomitant toxicity in host cells.

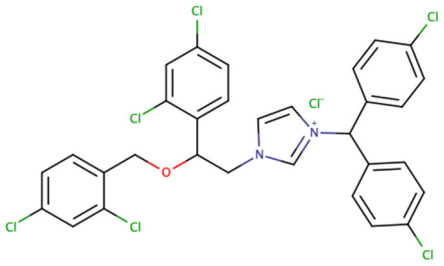


Figure 28- Calmidazolium chloride.

A.3.2.3 Thiophen ureidoacids

During the Ph.D. thesis of Elodie Laine in the Structural Bioinformatics Unit (2006-2009), funded by the DGA, TUA inhibitors (thiophene ureido-acid compounds) were identified

in silico by modeling the conformational transition between the active and inactive forms of EF and by targeting pockets displaying large variations along the transition path[119,192]. In particular, a cavity called SABC involving residues from switches A, B and C was selected as it exists in the inactive form of EF and early disappears along the path towards the active form (Figure 29). In addition, as described previously, the three switches are crucial for EF activation and catalytic activity. The SABC pocket thus constitutes an interesting site to block EF in an intermediate inactive conformation and thus prevent its activation by CaM. As the SABC pocket is also observed in CyaA, targeting SABC can be an alternative way to inhibit specifically the bacterial adenylyl cyclase toxins EF and CyaA.

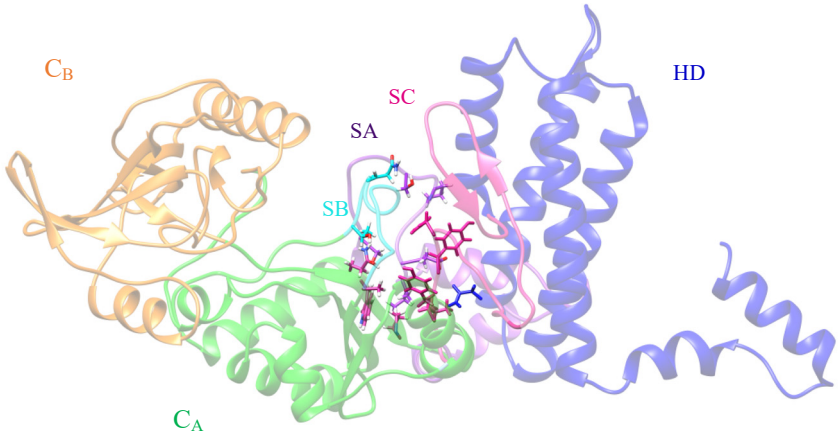


Figure 29- SABC cavity in EF inactive conformation. *For clarity, the representation of EF is flipped horizontally. The target pocket SABC is located at the interface of the SA (hot pink), SB (cyan) and SC (violet) switches. The sidechains of the SABC cavity residues are shown as sticks (A496, P499, I538, E539, P542, S544, S550, W552, Q553, T579, Q581, L625, Y626, Y627, N629, N709). The pocket is located between the catalytic core and the helical domain where CaM inserts upon interaction.*

A virtual screening on the SABC pocket was performed with 28,000 compounds of the National Chemical Library[193] and 18 compounds were chosen and validated experimentally. Among them, six belonged to the TUA family[194,195]. The 3,4-dichloro-5-phenyl-3-(3-phenylureido)thiophene-2-carboxylic acid (Figure 30) was the most active one (IC_{50} = 2-3 µM for EF) and revealed to be active also on CyaA[192].

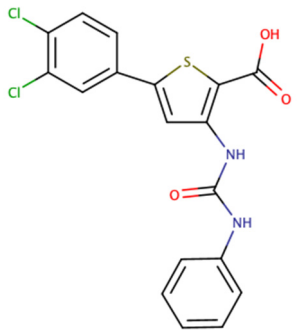


Figure 30- 3,4-dichloro-5-phenyl-3-(3-phenylureido)thiophene-2-carboxylic acid.

However, information on the mechanism of inhibition and binding site of TUA compounds is lacking. In addition, this family of compounds needed to be tested on other adenylate cyclases, and the cytoxicity and activity on cells needed to be assessed. The purpose of this thesis project is therefore to better understand the inhibition mode of TUA compounds and to experimentally determine the interaction site(s) of TUA inhibitors.

B. Aims of the project

This work builds-up on the previous work done by Elodie Laine[192]. The aim is to determine the binding site or sites of TUA inhibitors by experimental methods and by molecular modeling approaches and to shed light on their mechanism of inhibition. Acquiring structural data on the binding site of the inhibitor and study *in silico* the interplay between edema factor inhibition and activation is instrumental to unravel the mode of action of these TUA molecules. Here, we present a study of the interaction of TUA-diCl, with the protein EF from *Bacillus cereus* from the pathogenic G9241 *Bacillus cereus* strain, a very close homolog of the *B. anthracis* protein.

Chapter I is devoted to the analysis of biochemical and biophysical properties of TUA compounds and the biochemical and analytical characterization of the protein EF.

In Chapter II, we characterize the interactions of TUA compounds with EF, the complex EF-CaM and, unexpectedly, with CaM using adenylyl cyclase enzymatic activity assays, ligand-observed NMR experiments including STD (saturation transfer difference), Waterlogsy and trNOESY (transferred nuclear Overhauser Effect spectroscopy), hydrogen/deuterium exchange mass spectrometry (HDX-MS), crystallogenesis and crystallography studies. Moreover, we assess competition experiments for investigating on the inhibition mechanism of the TUA-diCl toward EF.

In Chapter III, we investigate in detail the interaction between the TUA compounds and CaM by NMR using ^{15}N and ^{13}C labeled CaM. This interaction is calcium dependent, which indicates that these hydrophobic compounds bind to hydrophobic surfaces of CaM which are formed in the presence of calcium between the N and C lobes of CaM.

In Chapter IV, we present an *in silico* study of the interplay between EF inhibition and activation. The EF-CaM complex was studied by molecular modeling in order to detect allosteric pockets. For this, several simulations were recorded in various conditions in which the complex was destabilized by alternately removing different cofactors (ions and ligand). After concatenation of all trajectories, all protein cavities were systematically detected and tracked. The cavities displaying the largest volume variations along the various conditions were located either in the catalytic site or at the EF-CaM interface. By targeting the EF-CaM interface, one could thus imagine inhibiting EF allosterically. In addition, we performed molecular dynamic simulations on the complex EF-CaM in the presence of the TUA-diCl. This MD simulation would help understanding if the TUA-diCl destabilized the complex EF-CaM.

C. Materials and Methods

C.1 Thiophen ureidoacids

The thiophen ureidoacids (TUA) adenylate cyclase inhibitors used in this work, TUA-diCl and TUA-Cl, were synthesized by P. Suzanne (CERMN, Centre d'Etudes et de Recherche sur le Médicament de Normandie). The compounds were obtained through a four-step process as schematized in Figure 31 and described earlier[192].

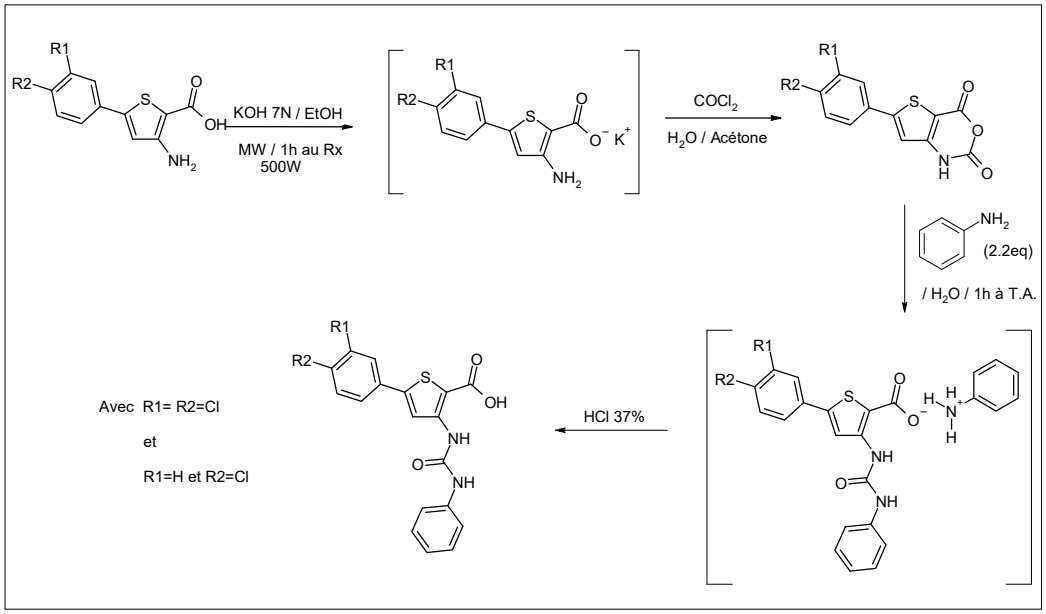


Figure 31- Chemical synthesis of TUA compounds

TUA-diCl and TUA-Cl were obtained as dry green or yellow powders, respectively. Stock solutions (20 mM concentration) were prepared by dissolving the compounds in deuterated dimethyl sulfoxide (DMSO-d_6), aliquoted and kept frozen at - 20°C until use.

The purity and integrity of the compounds were assessed by ^{1}H and ^{13}C NMR. Resonances of both compounds were assigned in DMSO-d_6 and in the buffer used for NMR experiments (buffer A: 20 mM Tris-d11, pH 7.5, 100 mM NaCl, 0.2 mM $CaCl_2$, 5% DMSO-d_6). In aqueous buffer, mono and dichloro TUA compounds showed a tendency to auto-associate. Therefore, signals were assigned at different concentrations. NMR experiments were performed on 800 (Avance NEO) or 600 MHz (AVANCE III HD) Bruker spectrometers (see

section C.4 for details on the spectrometers). Resonances were assigned with standard mono dimensional (1D) ^{1}H, 1D ^{13}C, DEPTQ, bidimensional (2D) homonuclear ^{1}H–^{1}H COSY,^{1}H–^{1}H NOESY (mixing time of 500 ms) and heteronuclear ^{1}H–^{13}C HSQC and HMBC. The TUA-diCl was also assigned in buffer A. Here below, are given the resonances in DMSO-d_6.

5-(4-chlorophenyl)-3-(phenylcarbamoylamino)thiophene-2-carboxylic acid (TUA-Cl). Yellow solid, mp. 208 °C, 60%.
^{1}H NMR (600 MHz, DMSO-d_6) δ 6.98 (t, *J* = 7.4 Hz, 1H), 7.34 – 7.24 (m, 2H), 7.52 (ddd, *J* = 9.3, 7.5, 5.1 Hz, 4H), 7.75 – 7.67 (m, 2H), 9.95 (s, 1H), 8.27 (s, 1H), 10.20 (s, 1H).
^{13}C NMR (600 MHz, DMSO-d_6) δ 117.1, 118.9, 119.0, 122.5, 127.7, 129.2, 129.7, 132.7, 133.6, 140.2, 145.5, 145.6, 151.8, 165.1.

5-(3,4-dichlorophenyl)-3-(phenylcarbamoylamino)thiophene-2-carboxylic acid (TUA-diCl). Green solid, mp. 219–220 °C. 63%.
^{1}H NMR (600 MHz, DMSO-d_6) δ = 7.01 (t, J= 7.4 Hz, 1 H), 7.30 (dd, J= 8.0, 7.4 Hz, 2H), 7.53 (d, J= 8.0 Hz, 2 H), 7.64–7.68 (dd, J= 8.5, 1.5 Hz, 1 H), 7.72 (d, J= 8.5 Hz, 1 H), 7.95 (d, J= 1.5 Hz, 1 H), 8.37 (s, 1 H), 9.64 (s, 1 H), 10.02 (s, 1 H), 13.37 (br s, 1 H, COOH).
^{13}C NMR (600 MHz, DMSO-d_6) δ = 109.7, 119.0, 120.0, 122.8, 126.4, 127.8, 129.3, 131.9, 132.01, 132.6, 133.9, 139.9, 144.2, 145.5, 151.8, 165.1.

C.2 Protein expression and purification

C.2.1 Plasmids and cloning vectors

Plasmids and cloning vectors of Bacillus cereus. Plasmids using the pET vector (Novagen) coding for *B. cereus* EF_3 were designed from the vector used by WJ Tang (pProEX-H6-EF)[196] for *B. anthracis* EF_3. New restriction enzyme cleavage sites had been added to the pProEX-H6-EF vector, NcoI & NheI at the N-ter end and KpnI at the C-ter end. The EF_3 gene from *B. cereus* G9241 strain (a gift from P. L. Goossens) was cloned using the HiFI / Gibson method into the vector pETM11 between the NotI and KpnI sites. The vector pETM11-EF2-Short codes for a fusion protein consisting of a N-terminal Glu residue, a hexahistidine tag (6-His tag), a TEV cleavage site and residues 291-798 of the catalytic domain of EF_3. The plasmid also codes for a kanamycin resistance gene and a T7 promoter. This plasmid was designed to obtain EF_3 for

crystallogenesis and *in vitro* experiments (Figure 32). The EF_3 protein coded by pETM11-EF2-Short lacks the flexible PA interaction domain, which could make difficult crystallogenesis and has not been included to obtain *B. anthracis* EF_3 structures.

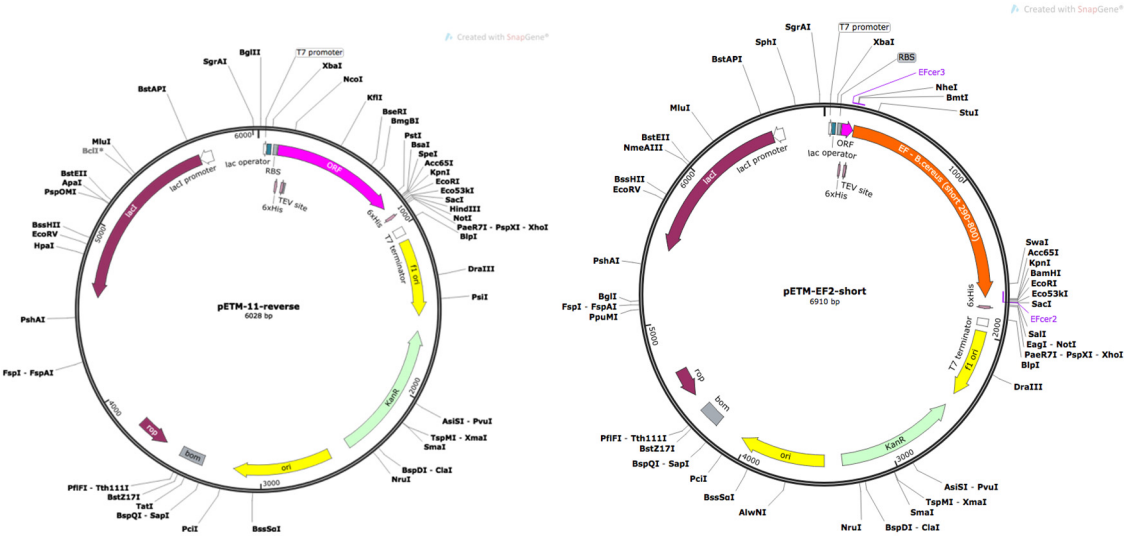


Figure 32- Plasmid map: *pETM11vector (left) and pETM11-EF2-Short plasmid (right). The EF_3 gene is represented in orange, the replication origin (ori, f1 ori) in yellow, the kanamycin resistance gene in green and the lacI promoter in purple.*

C.2.2 Protein expression and purification

Expression of EF_3 and CaM was performed in collaboration with the Production and Purification of Recombinant Proteins Technological Platform and the team of Daniel Ladant (Biochemistry of Macromolecular Interactions Unit, BIM) of the Institut Pasteur.

Expression of EF_3. *B. cereus* EF_3 was expressed using the plasmid pETM11-EF2-Short described in the previous section and BL21(DE3) Star™ CodonPlus *Escherichia coli* (*E. coli*) strains. Expression was done either in flasks or, after optimization in bioreactors. Cells were heat-shock transformed: 5 µL of 10 ng/µL plasmid miniprep were mixed gently, incubated for 45 min on ice, placed at 42°C for 30 seconds and then immediately immersed in ice for 2 min. Once the thermal shock was finished, 500 µL of Luria Bertani (LB) 1X medium were added, cells were incubated for 45 s at 37 °C and 200 rpm and then 0.2 mL of the bacterial suspension were spread on LB Agar petri dishes containing the antibiotics kanamycin (Kan, 50 µg/mL) and chloramphenicol (Cm, 30 µg/mL). Plates were incubated overnight at 37 °C. For expression, a preculture of 20 mL of LB media supplemented with Kan and Cm was incubated

at 37 °C with shaking at 200 rpm overnight. The cells were then further grown at 37 °C in Nzytech, an auto-inducible, media (with 50 µg/mL Kan) starting from on OD_{600} ~ 0.2 at 37 °C to an OD_{600} ~ 4 (corresponding to approximately 4 h of incubation), and then the temperature was lowered to 16°C for 16 h postinduction incubation. After expression, cells (~ 15 g/L of culture) were collected by centrifugation 17 500 g, 8 °C for 15 min and stored at -80 °C.

Once the conditions were optimized for 1 L cultures of EF_3, a scale up was performed to express the protein in 4 L bioreactors. Scale up was done using the rich medium HDM containing 0.1% glucose and the antibiotics Kan/Cm. Cells were grown for ~ 4 h at 37 °C, expression was induced with 1 mM IPTG after reaching a DO600 of 22. The temperature was then switched to 16 °C and cells were incubated for 16 h. The final DO600 was ~ 43 and the obtained biomass was 200 g. Cells were collected by centrifugation at 17 500 g and 8 °C for 15 min and stored at -80 °C.

Expression of CaM. Human calmodulin was also expressed in 4 L bioreactors. CaM was overproduced in *E. coli* BLR cells carrying the heat-inducible plasmid pDLTCaM41[117]. *E. coli* was grown at 30 °C in LB medium containing 100 µg/mL ampicillin to an OD_{600} of 0.6–0.8. To induce CaM expression, the growth temperature was shifted to 42 °C and cells were incubated for 3 h. The cells (120 g) were collected by centrifugation at 17 500 g and 8°C for 15 min and stored at -80°C.

Purification of EF_3. Frozen *E. coli* cell pellets were thawed and homogenized in a Waring Commercial Blender using a D-Cell and lysed at 4 °C and 1 kbar in 20 mM Tris-HCl pH 8.6, 500 mM NaCl, with an EDTA-free protease inhibitor cocktail (Roche Diagnosis) and few microliters of benzonase. The benzonase was used to remove nucleic acids, reduce the viscosity of the lysate and prevent cell clumping. The lysate was centrifuged for 1 h at 47 800 g and 8 °C. The recombinant 6-His-tagged EF_3 protein was present in the supernatant. The strategy of purification was based on immobilized metal affinity chromatography (IMAC) with nickel nitrilotriacetic acid (Ni-NTA) matrices. The supernatant of the *E. coli* lysate was applied directly onto a Protino Ni-NTA 5 mL column previously equilibrated with 20 mM Tris-HCl pH 8.6, 500 mM NaCl, 20 mM imidazole. The column was washed with 5 column volumes (CV) of the equilibration buffer with a flow rate of 1 mL / min. The 6-His-tagged EF_3 protein was retained on the column and was eluted with a gradient of imidazole from 20 to 200 mM. The 6-His tag was cleaved overnight at 4 °C in 20 mM Tris-HCl, pH 8.6, 500 mM NaCl by using a ratio 1:100 (w:w) ratio of hexahistdine tagged Tobacco Etch Virus (TEV) protease:EF_3. EF_3

was applied again onto the Ni-NTA column and directly eluted with the proteolysis buffer and concentrated using 30 kDa-cutoff Amicon tubes into a final volume of 5 mL. The last purification step was a size exclusion chromatography (SEC) on a Superdex-200 hiLoad 16/600x column (20 mM Tris-HCl, pH 8.6, 50 mM NaCl) at 4 °C with a flow of 0.5 mL/min. Finally, EF_3 was then concentrated with a 30 kDa-cutoff Amicon concentrator (> 5 mg/mL) and stored at -80 °C. The purity of the EF_3 preparation at the different steps was monitored by SDS-PAGE analysis.

Purification of CaM. CaM expression frozen cell pellets were thawed, homogenized and lysed like for EF_3 (see above) using a different buffer: 20 mM HEPES pH 7.5, 50 mM NaCl, supplemented with complete protease inhibitor cocktail (Roche) and few microliters of benzonase. The strategy of CaM purification was based on differential precipitation of proteins. The first step of purification was a precipitation in 50% ammonium sulfate (CaM remains soluble) followed by a glacial acetic acid precipitation in 50% ammonium sulfate (CaM does precipitate). After centrifugation (20 min at 17 200 g at 4 °C) the CaM pellet was resuspended in 20 mM HEPES pH 7.5, 1 mM EDTA and loaded on a phenyl Sepharose chromatography column for a hydrophobic interaction chromatography (HIC). In the absence of calcium, CaM (apo-CaM) does not expose hydrophobic patches and thus does not bind to phenyl Sepharose while some minor contaminants do. The unbound fraction was supplemented by 5 mM $CaCl_2$ and a second HIC chromatography was performed in the presence of 2 mM $CaCl_2$ to charge CaM with Ca^{2+} ions and form holo-CaM, which presents hydrophobic patches formed by interactions of the N- and C-terminal lobes upon calcium binding. In the presence of calcium, CaM binds to the phenyl Sepharose column. CaM was eluted with 20 mM HEPES pH 7.5 supplemented with 2 mM EDTA to hijack calcium ions. An EDTA gradient (0 to 2 mM) was further added to the eluted protein sample, which was further purified at 4°C by SEC on a Superdex-75 hi-load 26/60 column equilibrated in 20 mM HEPES pH 7.4, 50 mM NaCl (0.5 mL/min flow). Finally, CaM was then concentrated with a 3 kDa-cutoff Amicon concentrator (> 5 mg/mL) and stored at -80 °C. A SDS-PAGE was performed at each step to control the purification.

Purification of the complex EF_3-CaM. For crystallogenesis the EF_3-CaM complex was isolated by SEC on a S-200 column in 20 mM Tris-HCl pH 8.6, 20 mM NaCl, 2 mM $CaCl_2$ at 4°C with a flow rate of 0.5 mL/min. The sample was equilibrated for 10 minutes before loading it to the column using a molar ratio of 2:3:6 (EF_3:CaM:Ca^{2+}) as described[65].

C.2.3 Protein analysis

Analysis of the purified proteins was performed in collaboration with the Molecular Biophysics (PFBMI) and the Biological NMR platforms (PFBioNMR) of the Institut Pasteur.

Mass Spectrometry. The integrity and identity of the proteins was assessed by matrix-adsorbed time of flight mass spectrometry (MALDI TOF) on an Ultrafex TOF/TOF (Bruker) equipment used in linear and positive mode. For analysis, 1 µL of protein was added to 1 µL of saturated sinapinic acid (10 mg/mL) prepared in 50% acetonitrile and 0.1% of trifluoroacetic acid.

Dynamic light scattering. Protein samples were analyzed by dynamic light scattering (DLS) to control homogeneity, detect aggregates and determine the size of proteins. Analyses were performed on a DynaPro Plate Reader II (Wyatt) instrument, with 20 acquisitions of 10 cycles of 10 sec acquisitions. DLS was also used for buffer optimization on samples kept 24 h at 4 °C. Twenty conditions were tested with 5 different pH values (6, 7, 7.5, 8, 8.5) and four salt NaCl concentrations (50, 150, 300, 500 mM).

NanoDSF. The buffer optimization was also performed by NanoDSF (nano differential scanning fluorimetry) on a Prometheus nt48 (Nanotemper) apparatus. The principle of the technique is to analyze the thermal stability of proteins by monitoring their intrinsic fluorescence signal. The melting temperature (Tm) for each buffer condition characterized by DLS was determined in duplicate by NanoDSF. The temperature range used for the melting curves was from 20 °C to 95 °C with a 1 °C/min temperature gradient.

C.3 Enzymatic assays

General Principe. The enzymatic activity and the inhibitory potential of the compounds were tested by a non-radioactive adenylyl cyclase (AC) activity assay. EF_3 in the presence of host holo-CaM, its activator, AC catalyzes the following reaction:

$$ATP \xrightarrow{Mg^{2+}} cAMP + PPi + H^+$$

In the assay, ATP is converted to cAMP in a first step, and in a second step, cAMP is separated from ATP by Al_2O_3 at pH 7.5: while deprotonated ATP (ATP^{4-}) is retained on the

Al_2O_3 powder, $cAMP^-$ is not. ATP can later be detected by light absorption read at 260 and 340 nm, and the amount of cAMP can be deduced.

Enzymatic Assay. *B. cereus* EF_3 was diluted into the dilution buffer (10 mM Tris-HCl pH 8.0, 0.001% (v/v) Tween 20) to 1 nM and 40 µL of diluted EF_3 were added (0.4 nM final concentration) to 50 µL of AC Assay buffer (100 mM Tris-HCl, pH = 8.0, 15 mM $MgCl_2$, 0.2 mM $CaCl_2$, 0.2-1 mg BSA, containing various concentrations of CaM – from 0.02 nM to 2 µM – or without CaM, and either no DMSO or 10% DMSO). The mixture was pre-incubated for 5 minutes at 30 °C. The enzymatic reaction was initiated by addition of ATP (10 µL of a 20 mM stock solution prepared in H_2O, pH 7.0) at 30 °C under agitation and carried out for 10 min at 30 °C under agitation. The enzymatic reaction was stopped by addition of Al_2O_3 powder (about 0.2 mL dry powder). Then 0.9 mL of 20 mM HEPES, pH = 7.5, 0.1 M NaCl were added and the mixture was well mixed and left for few min (or up to few hrs). The tube was centrifuged 3-5 min at 16 000 g. The supernatant (2 x 0.3 mL), containing the unbound cAMP, was transferred to a UV- transparent microplate. The Optical Densities at 260 and 340 nm (OD_{260}, OD_{340}) were recorded with a microplate reader. The quantity of cAMP was determined from the absorbance at 260 nm using an extinction coefficient for cAMP of 15.4 mM (the absorbance at 340 nm is used as a control for measuring background light scattering due to occasional Al_2O_3 powder carryover). The enzymatic activity (k_{cat}) is expressed in mol of cAMP produced per sec per mol of the enzyme, corresponding to k_{cat} of the enzyme in these conditions. The enzymatic activity assay was measured at different final concentrations of CaM (i.e. 1 µM, 0.1 µM, 10 nM, 1 nM, 0.1 nM). It was also tested in the absence of free calcium by adding an excess of EGTA (2 mM final concentrations) to the AC Assay buffer.

Inhibition assay. The compounds (5 µL from a corresponding 5× concentrated solutions in 20 mM HEPES, pH 7.5, NaCl 0.1 M, 10% v/v DMSO) were added to the indicated final concentrations (i.e. 100, 60, 40, 30, 20, 10, 6, 3 and 1µM) to EF_3 and incubated for 15 min at 30 °C. Then CaM (10 µL from a 200 nM stock solution in 10 mM Tris-HCl pH 8.0) was added to a final concentration of 20 nM and the mixtures were incubated for 5 min at 30 °C and the enzymatic activity was triggered and measured as described above. Controls without inhibitor were performed. The data were performed in duplicate. The same procedure was followed for the interaction studies with AC384 (gently provided by D. Ladant team). AC384 corresponds to the catalytic domain of CyaA (residues 1-384).

C.4 NMR

Samples. For experiments with unlabeled proteins, samples were buffer exchanged on PD10 columns (Minitrap G25 – 0.97x2.8 cm, GE Healthcare) to study proteins in buffer 20 mM Tris-d11 (98% deuterated Tris, Eurisotop) pH 7.5, 100 mM NaCl, 0.2 mM $CaCl_2$, 5% DMSO-d_6. The stock of TUA ligands was dissolved at 20 mM in 100% DMSO-d_6. Ligand observed interaction studies were performed with a 1:100 protein (0.5 µM): ligand (50 µM) ratio. The ratio chosen for the interaction study was 100-fold excess of ligand (50 µM) as compared to the proteins (0.5 µM) in presence of 5% of DMSO-d_6. For the competition experiments, another ligand used was the 2'-MANT-3'-dATP (Jena Bioscience). The complex was formed mixing 0.5 µM of EF_3 with 1 µM of CaM, 500 µM of 2'-MANT-3'-dATP and 50 µM of TUA-diCl were then added in the presence of 15 mM of magnesium and 5% of DMSO-d_6. For protein observed experiments with CaM, lyophilized doubly labeled $^{15}N/^{13}C$ CaM (ordered from Giotto Biotech) was dissolved in 20 mM HEPES, 100 mM NaCl, 2 mM $CaCl_2$, pH 7, 5%D_2O. For holo-CaM resonance assignment, the protein concentration was 258 µM. For the interaction study, 110 µM of $^{15}N/^{13}C$ CaM were titrated with TUA-diCl (10 to 245 µM) with a final DMSO-d_6 concentration of 6%. All NMR experiments were performed in 3 mm NMR tubes.

Data acquisition and analysis. NMR spectra were collected at the Institut Pasteur Biological NMR Technological core facility on Bruker spectrometers operating at a proton frequency of 800.6 MHz (Avance NEO) or 599.4 MHz (Avance III HD) both equipped with a cryogenically cooled triple resonance $^1H\{^{13}C/^{15}N\}$ TCI probe. The magnetic field was locked either with D_2O or DMSO-d_6. The ^1H chemical shifts were referenced to external DSS (dimethyl silapentane sulfonate). Data were recorded with Topspin 3.5.7 or 4.05 (Bruker) and processed and analyzed with Topspin 4.05, NMRPipe[197] and CCPNMR analysis 2.4.2[198].

Experiments with unlabeled proteins (ligand observed) were performed at 25°C, and those with $^{15}N/^{13}C$ CaM (protein observed) at 37°C. To characterize the interaction of TUA compounds with the proteins we performed saturation transfer difference (STD)[199,200], water-ligand observed via gradient spectroscopy[201] (WaterLOGSY), and nuclear Overhauser effect spectroscopy[202] (NOESY) experiments. These ligand-observed techniques rely on the differential magnetization transfer or relaxation properties of macromolecules and small free compounds and on the exchange between the bound and free forms of the ligand during the experiment. The ligand is in excess relative to the unlabeled macromolecule. The effect of saturation (STD, WaterLOGSY, NOESY) or transverse relaxation (T2) of the macromolecule

is transferred to the bound ligand and then upon complex dissociation at equilibrium to the observed free ligand resonances. The conditions (protein:ligand ratio and concentration) were optimized taking into account the auto-association of the free ligand and finding conditions in which auto-association did not contribute to signal build-up: ratio 1:100 and 50 µM ligand concentration.

STD experiments (1024 scans) were acquired with on- and off-resonance frequencies at 0 and 40 ppm, respectively, a 0.5 ppm saturation bandwidth and a saturation time of 2 seconds. Saturation was achieved using trains of Eburp2[203] pulses. The STD value (STD_{exp})was defined as the ratio of the areas of each signal in the difference (A_{STD}) and reference (A_0) spectra:

$$\mathrm{STD_{exp}} = A_{STD}/A_0$$

WaterLOGSY experiments (1024 scans) were recorded with a 2 s inversion time for the water resonance. ^{1}H-^{1}H NOESY experiments were acquired with 80 scans, and a mixing time of 600 ms. Except in WaterLOGSY experiments, the water signal was suppressed using excitation sculpting with gradient sequences[204].

To follow the interaction of CaM with TUA-diCl, the chemical shifts and relaxation properties of CaM were characterized. Hence assignment of the backbone and CB resonances of holo-CaM were required. For the assignment of holo-CaM signals, longitudinal-relaxation enhanced BEST (band-selective excitation short transient) versions of standard HNCO, HNcaCO, HNCA, NHcoCA, HNCACB, CBCAcoNH 3D experiments included in the NMRLib were used[205]. Amide protons were excited at 9 ppm with a 4-ppm excitation bandwidth. The same experiments were acquired to assign the resonances of holo-CaM (99 µM) in the presence of TUA-diCl (245 µM). The titration of TUA-diCl in presence of CaM was followed by ^{1}H-^{15}N-SOFAST-HMQC[206] (band-selective optimized flip-angle short-transient heteronuclear multiple quanta correlation) and HSQC[207] (heteronuclear single quantum correlation) experiments. We used the following equation to calculate the chemical shift perturbation with $\Delta\delta_H$ and $\Delta\delta_N$ corresponding to the chemical shift of ^{1}H and ^{15}N respectively.

$$\Delta\delta = \sqrt{{\Delta\delta_H}^2 + (0.159 \times \Delta\delta_N)^2}$$

Binding data were analyzed with TITAN[208] to extract kinetic and thermodynamic data from the evolution of the chemical shifts and line shapes of CaM amide signals upon titration of CaM with TUA-diCl. The ^{1}H-^{15}N SOFAST data was used as input for the analysis.

The dynamics of holo-CaM in the presence or absence of TUA-diCl was assessed by amide ^{15}N 1D TRACT and 2D T1, T2 ^{1}H-^{15}N NOE relaxation experiments. The tumbling correlation time (τ_c) was estimated from ^{15}N relaxation using the BEST version of the TRACT sequence (TROSY for rotational correlation times[209]) to assess the compactness of different states of CaM (bound and free CaM in this instance). The TRACT experiment allows one to determine the τ_c for large proteins and/or exchanging systems in which the transversal relaxation is very fast. T1, T2 and ^{1}H-^{15}N NOE experiments were performed using the pulse sequences described by Lewis Kay and co-workers[210]. TRACT, T1 and T2 data were fit to mono-exponential decays to obtain the corresponding relaxation times.

C.5 Fluorescence

The interaction of TUA-diCl with EF_3 and CaM was monitored by means of the intrinsic fluorescence of TUA-diCl. Fluorescence spectra were recorded at room temperature using a TECAN Infinite 200 Pro instrument with the Tecan i-control software with samples in 384-well flat bottom polystyrene plates (Corning) that contained 20 µL solution per well. Emission spectra were recorded from 380 to 750 nm with excitation at 350 nm and 20 (emission) and 5-10 nm (excitation) bandwidths. For the interaction study of TUA-diCl with EF_3 and CaM, the concentration of TUA-diCl was fixed at 10 µM and the protein concentration was varied between 60 nM and 100 µM in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 2 mM $CaCl_2$, 5% (v/v) DMSO. For the self-auto association studies of TUA-diCl, the concentration of the compound varied from 10 µM to 400 µM with different concentrations of Tween 20 (0%, 0.005%, 0.05%) in 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.2 mM $CaCl_2$, 5% (v/v) DMSO.

C.6 HDX-MS experiments

Sample Preparation. All labeling reactions were performed at room temperature in 20 mM deuterated HEPES buffer, 150 mM NaCl, 2 mM $CaCl_2$, pD 7.4, 2.5% DMSO (labeling buffer) unless specified. Prior to initiating HDX-MS, the quality of each recombinant protein was assessed by intact mass measurement (data not shown). Two experimental conditions were analyzed. In the first instance, EF_3 (15 µL at 5.6 µM in 20 mM HEPES, 150 mM NaCl, 2 mM $CaCl_2$, pH 7.4) was incubated alone and in the presence of a 100-fold molar excess TUA-diCl (62.5 µM in 20 µL) using a final 2.5% (v/v) DMSO concentration. In parallel, CaM (15 µL at 6.7 µM in 20 mM HEPES, 150 mM NaCl, 2 mM $CaCl_2$, pH 7.4) was incubated in both the presence and the absence of an 87-fold molar excess TUA-diCl (62.5 µM in 20 µL) in 2.5% DMSO. The excess of TUA-diCl ensures that all complexes remain stable over the entire labeling reaction. After 1 h equilibration at room temperature, continuous labeling was initiated by dilution (7-fold, final D_2O/H_2O ratio of 85.7:14.3%) using labeling buffer supplemented or not with 62.5 µM TUA-diCl. Aliquots of 20 µL (12-14 pmol of proteins) were withdrawn at each labeling time point and quenched upon mixing with 40 µL of an ice cold quench solution of 3% formic acid, 4 M urea to decrease the pH to 2.5 (final D_2O/H_2O ratio = 28.6:71.4%). Quenched samples were immediately snap frozen in liquid nitrogen and stored at -80°C. Undeuterated samples were obtained following the same experimental procedure. One unique sample was prepared per protein and condition (preliminary results).

Data Acquisition. Quenched samples were thawed and immediately injected onto a HDX manager connected to two nanoACQUITY UPLC M-Class pumps (Waters Corporation, Milford, MA). The temperature of the HDX manager was maintained at 0°C to minimize back exchange. 50 µL of each labeled sample (i.e., 10 pmol of EF_3; 12 pmol of CaM) were digested using an in-house packed column (2.0 x 20 mm, 63 µL bed volume) of immobilized pig pepsin agarose beads (Thermo Scientific, Rockford, IL) for 2 min at 20°C. Peptides were directly trapped and desalted onto a C18 Trap column (VanGuard BEH 1.7 µm, 2.1 x 5 mm, Waters Corporation, Milford, MA) at a flow rate of 100 µL/min (0.15% formic acid) and separated by a 8 min linear gradient of 5-30% acetonitrile followed by a short 2 min increase from 30% to 40% of acetonitrile at 40 µL/min using an ACQUITY UPLC BEH C18 analytical column (1.7 µm, 1 × 100 mm, Waters Corporation, Milford, MA). After each run, the pepsin column was manually cleaned with two consecutive injections of 1% formic acid, 5% acetonitrile, 1.5 M

guanidinium chloride, pH 1.7. Blank injections were performed between each run to confirm the absence of carry-over.

Mass spectra were acquired in resolution and positive ion-mode (*m/z* 50-2000) on a Synapt G2-Si HDMS mass spectrometer (Waters Corporation, Milford, MA) equipped with a standard ESI source and lock-mass correction. Peptic peptides were identified in undeuterated samples by data independent acquisition (MS^E) using the same chromatographic conditions than for the deuterated samples. Four distinct MS^E trap collision energy ramps were employed to optimize the efficiency of the fragmentation: 10-30V (low), 15-35V (medium), 20-45V (high), and 10-45V (mixed mode).

Data Analysis. The initial peptide maps were generated by database searching in ProteinLynX Global server 3.0 (Waters corporation, Milford, MA) using the following processing and workflow parameters: low and elevated intensity thresholds set to 250.0 and 100.0 counts; intensity threshold sets to 750.0 counts; automatic peptide and fragment tolerance; non-specific primary digest reagent; false discovery rate sets to 4%. Each fragmentation spectrum was manually inspected for assignment confirmation. The peptide maps were further refined in DynamX 3.0 (Waters corporation, Milford, MA) using the following filters: minimum intensity = 5000; minimum products per amino acid = 0.20; minimum sum intensity for products: 1000; minimum score = 7.0; maximum MH+ error (ppm) = 10.

DynamX 3.0 was used to extract the centroid masses of all peptides selected for HDX-MS. One unique charge state was used per peptide and no back-exchange correction was performed. HDX-MS results are reported as relative deuterium exchange level expressed in fractional exchange. Fractional exchange data were calculated by dividing the experimental uptake value by the theoretically maximum number of exchangeable backbone amide hydrogens that could be replaced into each peptide in 85.7% excess deuterium.

C.7 X-ray crystallography

Crystallization. Nano sitting drop high throughput crystallization screenings were performed using a Mosquito robot (TTPlabtech) in the crystallography core facility of Institut Pasteur[211]. Each crystallization plate was stored and imaged using a RockImager 1000 (Formulatrix) either 4 or 18°C. The frequencies of the images are (i) point zero, (ii) daily in the first week, (iii) every three days during the rest of the first month, (iv) once per week the second month, (v) twice a month, and then (vi) once a month. The images were visualized and scored via Rock Maker

Web server. For crystal optimization, the design of the matrices was set up using the software CryMon (Protein BioSolutions Inc) and the crystallization solutions were generated with the MatrixMaker robot (Protein BioSolutions Inc). The kits used for initial screening are: (i) Crystal screen 1&2 from Hampton, (ii) Wizard 1&2 from Emerald Biosystems, (iii) Structure screen 1&2 from Molecular Dimensions limited, (iv) JBS1 to JBS4 from Jena Bioscience (v) JBS5 to JBS8 from Jena Bioscience, (vi) SaltRX from Hampton and (vi) PEGion 1&2 from Hampton and in some screening (vii) Index from Hampton. 27 combinations of proteins samples (EF_3, H_6 EF_3, EF_3-CaM, H_6- EF_3-CaM, CaM) in complex with different molecules (3'-dATP from Sigma-Aldrich, TUA-diCl) at different concentrations (1 mM to 674 mM) had been tested using the seven or eight kits at two temperatures (4 to 18 degrees) which represent about 17 544 conditions of crystallization. For each screening, a new batch of purified protein has been used. The complex EF_3-CaM was purified by size exclusion chromatography using a superdex S-200 column with a buffer composed of 20 mM Tris-HCl, pH 8.6, 20 mM NaCl, 2 mM $CaCl_2$ at 4°C. The molar ratio used was 2:3:6 (EF_3:CaM:Ca^{2+}) as described and incubated for 10 min before injection in the column[64]. After the crystallization screening, the samples were frozen and stored at -80 °C. Crystal manual optimization was performed in Linbro plates with the hanging-drop method by mixing 2 µl of protein with 2 µl of reservoir solution.

Data collection and structure determination. For data collection, the crystals were flash-cooled in liquid nitrogen using the condition of crystallization supplemented with 30% (v/v) glycerol as a cryoprotectant. The X-ray diffraction data were collected on beamlines PROXIMA-1 or PROXIMA-2A (Synchrotron SOLEIL, St Aubin, France) and processed with autoPROC (Global Phasing Ltd.). The crystal structures of the CaM in complex with EF_3 was solved by the molecular replacement technique using the previously reported EF_3-CaM structure (PDB ID 1PK0)[127] as search models with program Phaser[212]. Final models of the complexes were obtained through interactive cycles of manual model building with Coot[213] and reciprocal space refinement with Buster[214]. X-ray data collection and model refinement statistics are summarized in Table 7. Figures showing the crystallographic model were generated and rendered with Pymol (Schrodinger, LLC) and/or Chimera[50].

C.8 Molecular docking

Molecular docking. TUA-diCl was docked onto the X-ray crystallographic structures of EF (PDB ID 1K8T), CaM (PDB ID 6M7H) and EF-CaM with the PABD domain (residues 1-291) truncated (PDB ID 1XFY) using AutodockVina 1.1.2[135]. The exhaustiveness of AutodockVina was set to 20 and the random seed to 42. Docking of TUA-diCl to CaM (PDB ID 6M7H) using NMR chemical shift perturbation (CSP) data as constraints was performed with the HADDOCK2.2 web interface[215] (https://www.bonvinlab.org/software/haddock2.2). Only residues with CSP values > 0.15 were included in the active residue list entry.

Cavities detection. Cavities were detected with the software *mkgridXf*, developed in-house by the team of Arnaud Blondel[216]. *MkgridXf* is designed to identify cavities by detecting void at the surface of a protein structure on a three-dimensional orthogonal grid using a small spherical probe with a radius set to 1.4 Å (r_{in}). The procedure was repeated with a large sphere of 8.0 Å (r_{ou}) for EF and EF-CaM and 6.0 Å for CaM to identify the external volume accessible to the solvent which is excluded from the cavities volume. For CaM, the values of parameters r_{ou} were set to the value of 6.0 Å obtain two cavities, one for N-CaM and another for C-CaM.
Allosteric pockets were predicted using the Web server of the PARS approach[217] (bioinf.uab.cat/cgi-bin/ pars-cgi/pars.pl).

Analysis of the molecular docking with STD data. The method used was developed to compare the theoretical and experimental STD values was developed in the Structural Bioinformatic unit by G. Bouvier[218]. The theoretical STD (STD_{th}) is defined as the buriedness (b_i) of the consider hydrogen atom of the ligand (i). The buriedness was defined as the number of hydrogens of the protein (n_i) in a sphere of 6 Å around ligand hydrogen i, divided by the total number of protein hydrogens in the binding pocket (n_j).

$$b_i = \frac{n_i}{\sum_{j}^{N} n_j}$$

The comparison with the experimental STD value STD_{exp} (previously described) was done using the root mean square error (RMSE).

$$RMSE = \sqrt{\frac{\sum(STD_{exp} - b_i)^2}{n}}$$

Analysis of the Molecular docking with the CSP. Complexes of CaM with TUA-diCl are clustered using the self-organizing map (SOM) approach[219]. The conformations of the complex between CaM and TUA-diCl are first aligned and the Cartesian coordinates of the small molecule are used as input for the SOM. The procedure of the SOM clustering was followed as described[218].

C.9 *In silico* studies of the EF-CaM complex

Molecular dynamics (MD) simulations for analyzing the relationship between the activation of the edema factor and its interaction with calmodulin. The X-ray crystallographic complex of EF with the inhibitor adefovir[127] (PDB ID 1PK0, Figure 90) served as the starting point of the MD trajectories. The protein chain was analyzed using Molprobity[220] (molprobity.biochem.duke.edu), in order to add hydrogen atoms and to select the sidechains orientations optimizing the network of hydrogen bonds. The ion Yb^{3+} present in the catalytic site, was replaced by a physiologically compatible ion Mg^{2+}. The files to perform MD simulations were prepared with the CHARMM-GUI interface (www.charmm-gui.org)[221,222]. The chains A and D of the structure 1PK0 were neutralized using potassium ions and solvated with water molecules (Table 1). The force field CHARMM36[223–225] and the TIP3P water model[226] were used to model the physical interactions. The parameters for ligand adefovir were obtained using the CHARMM-GUI interface (www.charmm-gui.org). with the Ligand Reader & Modeler tool[227]. Six different systems were prepared with different molecular compositions, starting from the structure 1PK0 in which the C-CaM lobe is loaded with two ions Ca^{2+} by removing various co-factors. The MD trajectories were carried out using NAMD 2.13[228] (www.ks.uiuc.edu/Research/namd/). The simulations were realized in the NPT ensemble, at a temperature of 300 K and a pressure of 1 atm. Temperature was regulated according to a Langevin thermostat[229], and the pressure was regulated with the Langevin piston Nose-Hoover method[230,231]. A cutoff of 12 Å and a switching distance of 10 Å were defined for non-bonded interactions, while long-range electrostatic interactions were calculated with the Particle Mesh Ewald (PME) protocol. The RATTLE algorithm[232,233] was used to keep all covalent bonds involving hydrogens rigid, enabling a time step of 2 fs. Atomic coordinates were saved every 10 picoseconds. At the beginning of each trajectory, the system was first minimized for 10,000

steps, then heated up gradually from 0 K to 300 K in 300,000 steps. Then, the system was equilibrated for 50,000 steps. For each of six different conditions (Table 1), two independent trajectories of 200 ns were recorded (named the replicas R1 and R2) and corresponding to a total simulation duration of 2.4 μs.

MD simulations of the complex TUA-diCl with EF, CaM and EF-CaM. The systems studied with MD simulations (Table 2) were prepared using CHARMM-GUI[51]. The starting points of the MD were built from the atomic coordinates of the following PDB entries (EF-CaM without the PABD domain 1XFY, EF from 1K8T, CaM from 6M7H) used above for the TUA-diCl docking. The ligand coordinates obtained from the docking pose with the lowest RMSE between STD_{exp} and STD_{th} were added to the protein system. All systems were simulated at atomistic resolution using CHARMM36[225] and the CHARMM General Force Field[234] for the protein and the ligand, respectively. The initial models were enclosed in a periodic cell and solvated with TIP3P water molecules[226]. Potassium and chloride ions were added to ensure charge neutrality at a physiological concentration of 0.15 M. The starting models were first energy-minimized using the steepest-descent algorithm and then equilibrated at 300K and 1 atm for 1 ns, using positional restraints on all the protein and ligand heavy atoms. Temperature and pressure during the NPT equilibration were ensured by the Bussi-Donadio-Parrinello thermostat[235] and the Berendsen barostat[236]. A time step of 2 fs was used together with LINCS constraints[237] on heavy atoms chemically bonded with hydrogens. Van der Waals and Coulomb interactions were cut off at 1.2 nm. A force-switch was applied to Van der Waals forces starting at 1.0 nm. Particle-mesh Ewald[238] was used for long-range electrostatic interactions. Production simulations were carried out in the canonical (NVT) ensemble at 300K, enforced by the Bussi-Donadio-Parrinello thermostat[235]. All simulations were performed using GROMACS 2019.6[239].

Analysis of MD trajectories. The root-mean-square deviations (RMSD) of atomic coordinates the root-mean-square fluctuations (RMSF), as well as distance and angle analysis between atoms along the trajectories were performed using cpptraj[240]. Angles between CaM α helix axes were calculated using the MDAnalysis library[241,242], the helices being defined as CaM regions including residues 8-19 (helix I), 31-37 (helix II), 46-53 (helix III), 66-73 (helix IV), 83-92 (helix V), 103-110 (helix VI), 119-127 (helix VII), 139-145 (helix VIII). The axis of a helix

spanning residues n to p is defined as a segment connecting the geometric centers of the atoms $C\alpha^{n}$ and $C\alpha^{n+2}$ and of the atoms $C\alpha^{p}$ and $C\alpha^{p+2}$.

The solvent accessible surfaces of residues along the trajectories were calculated using the python MDAnalysis library[241,242] and the software FreeSASA[243]. The EF catalytic site surface was defined as the sum of solvent accessible surfaces of EF residues H351, K353, S354, K372, R329, K346, L348, D491, D493, H577, G578, T579, D582, N583, E588, F586, and T548. The surface of the SABC pocket was defined as the sum of the solvent accessible surfaces of EF residues A496, P499, I538, E539, P542, S544, S550, W552, Q553, T579, Q581, L625, Y626, Y627, N629, and N709. The surface of hydrophobic patches is defined according to Yang *et al.*[116]. The N-CaM patch is formed by the N-CaM residues A10, F12, A15, L18, F19, L32, M36, L39, M51, V55, M71, M72, and M76. The C-CaM patch is formed by the C-CaM residues I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, and A147.

ID	System	Proteins | ligand | crystallographic ions | solution ions	Cell type	# water molecules	# K^+	Length [ns]
A	EF_ade_Mg_CaM_Ca	EF, CaM | adefovir | Mg^{2+}, 2 Ca^{2+}	cubic	68219	9	200x2
B	EF_ade_Mg_CaM	EF, CaM | adefovir | Mg^{2+}	cubic	69773	13	200x2
C	EF_ade_CaM_Ca	EF, CaM | adefovir | 2 Ca^{2+}	cubic	68220	11	200x2
D	EF_ade_CaM	EF, CaM | adefovir	cubic	69795	15	200x2
E	EF_CaM_Ca	EF, CaM | 2 Ca^{2+}	cubic	69849	11	200x2
F	EF_CaM	EF, CaM	cubic	69785	15	200x2

Table 1- Systems for EF-CaM interaction studies

ID	System	Proteins | ligand | crystallographic ions | solution ions	Cavity number	Cell type	# water molecules	# K/Cl ions	Length [ns]
1	EF	EF | TUA-diCl | SO_4^{2-}	4	triclinic	26562	82/81	200
2	EF	EF | TUA-diCl | SO_4^{2-}	18	triclinic	26675	83/82	200
4	EF-CaM	EF, CaM | TUA-diCl | 2 Ca^{2+}, Mg^{2+}	17	triclinic	26396	90/83	200
5	EF-CaM	EF, CaM | TUA-diCl | 2 Ca^{2+}, Mg^{2+}	36	triclinic	26399	90/83	200
6	EF-CaM	EF, CaM | TUA-diCl | 2 Ca^{2+}, Mg^{2+}	14	triclinic	26405	90/83	200
CaM-1	CaM	CaM | TUA-diCl | 4 Ca^{2+}	1	dodecahedron	6562	36/20	1066
CaM-2	CaM	CaM | TUA-diCl | 4 Ca^{2+}	2	dodecahedron	6555	36/20	1055

Table 2- Systems for TUA-diCl interaction studies with EF and EF-CaM

D. Experimental results & discussions

Chapter I. Biophysical and biochemical properties of the system

D.I.1 Biophysical properties of TUA compounds

CONTEXT

In the drug discovery field, high throughput screening is essential to find a hit for a specific target. During this process, hits can be found to be promiscuous molecules[244]. Most often, these molecules are hydrophobic, self-aggregate and interact with many other targets causing target aggregation[244]. Indeed, hydrophobic molecules can be challenging because these tend to aggregate and these aggregates can interact nonspecifically with the surface of the protein targets[244,245,246]. It is thus important to fully characterize the properties of the ligand in order to better apprehend its mode of action. In this section, we describe the characterization of the hit used in this work, which was found by virtual screening on an allosteric pocket identified by modeling the transition path between the inactive and active forms of the edema factor protein. The most active compound was the 5-(3',4'-dichlorophenyl) -3-[(N-phenylamino) carbonylamino]-thiophen-2-carboxylic acid (TUA-diCl, Figure 33).

RESULTS

As TUA-diCl is a hydrophobic molecule containing three aromatic rings and only one charge and was poorly soluble in H_2O, we characterized the molecule before assessing protein interactions. The NMR ^1H chemical shifts of the TUA-diCl signals are presented in Figure 33. The auto-association of the molecule can be monitored by NMR following the ^1H chemical shift variations and by fluorescence following the intrinsic fluorescence of the compound as a function of concentration. For these experiments, TUA-diCl in DMSO-d_6 was diluted into the buffer keeping the final percentage of DMSO-d_6 constant, at 5% (v/v), to help the solubilization of the molecule.

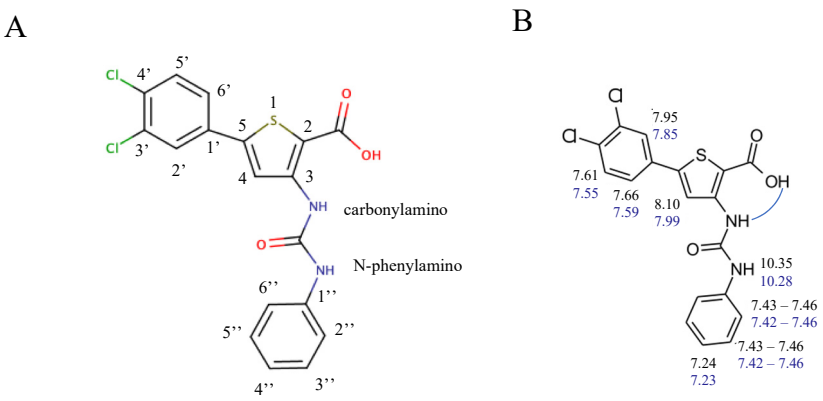


Figure 33- The TUA-diCl compound. *A. The TUA-diCl molecule and atom numbering are represented This numbering is used throughout the manuscript. B. The ^{1}H chemical shifts of TUA-diCl at 20 µM (black) and 300 µM (blue) are shown.*

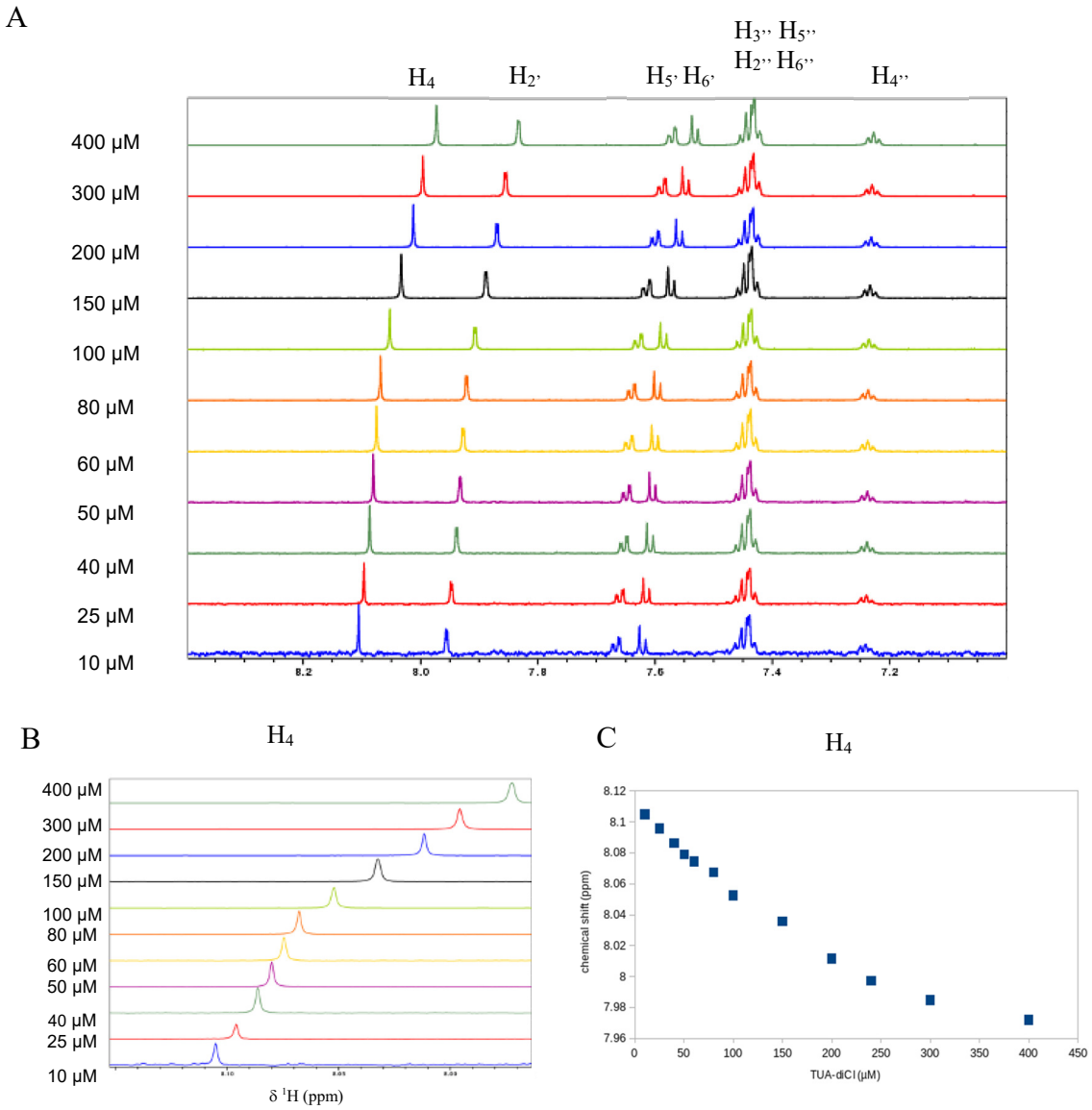


Figure 34- ^{1}H spectra of TUA-diCl. *A. ^{1}H spectra depending on TUA-diCl concentration, starting from 10 µM and ending at 400 µM. In the spectra only the proton belonging to the 3 rings are shown, the N-*

phenylamine was excluded. B. ^{1}H zoom spectra on the proton H_4 (located in thiophen moiety) depending on the addition of TUA-diCl molecules. C. Graphic of the chemical shifts of H_4 depending on TUA-diCl concentrations.

The ^{1}H chemical shifts of many TUA-diCl signals vary upon successive additions of the compound, indicating that the molecule self-assembles in a fast-exchange regime on the chemical shift time scale (Figure 34). The chemical shift of the proton of the thiophen and di-chloride cycles and the N-phenylamine proton (not shown) were the most impacted. In contrast, the chemical shift of the phenyl ring was not perturbed, indicating that auto-association mainly involves the thiophen and di-chloride rings. Even though the number of aggregation (number of molecules in the oligomeric state or states) is not known, an apparent measure of the half transition constant of aggregates can be determined from the chemical shift variations for each proton. In Figure 34, we present the chemical shift changes of proton H_4, which belongs to the thiophen moiety. The apparent half transition constant of TUA-diCl aggregates is ~ 100-120 μM.

Additionally, aggregation can be followed by the saturation transfer difference (STD) NMR technique. STD is a method used for ligand screening and mapping the atoms of the ligand in interaction with a protein. Two spectra are recorded: one with a specific saturation on the protein far from the signals of the ligand (saturation spectrum) and the second one with off-resonance saturation (reference spectrum), very far from the signals of the protein and the ligand. The saturation on large molecules like proteins is very efficient and rapidly diffuses within the protein and to the ligand protons in close contact with the protein. The saturation of the bound ligand is transferred to the free ligand, which is in excess relative to the protein. The difference of the two spectra (saturation – reference) displays only the signals of the ligand in close contact with the protein and the STD_{exp} value (difference spectrum area of the signal normalized relative to the corresponding reference spectrum signal) is a measure of the relative distances of the ligand hydrogens to the protein hydrogen atoms. No STD signal indicates no interaction. If a small compound aggregates into high molecular weight species, the signals within the assembly will be very broad and invisible in the NMR spectra. Upon saturation (at 0 ppm in this instance) in a region with no visible ligand signals but with invisible resonances of the ligand within the aggregates in exchange with the visible ligand, the bound ligand will be saturated and its saturation will be transferred to the soluble ligand upon dissociation. Under these conditions, the ligand will show signals in the STD difference spectrum.

At concentrations close to or higher than the apparent aggregation, the STD spectra of TUA-diCl with saturation at 0 ppm (and reference at -40 ppm), ie > 7 ppm away from the closest TUA-diCl signal, shows signals indicating that indeed, the NMR visible species are in equilibria with high molecular weight species.

Addition of detergent has been proposed to solve the aggregation problem and is regularly used in screening to avoid selecting aggregator compounds, as the inhibition power of promiscuous compounds is sensitive to detergent[247]. We thus tested the effect of the non-ionic detergent Tween 20 on the aggregation and activity of TUA-diCl by STD, fluorescence and enzymatic assays, using concentrations below and above the critical micellar concentration (CMC). We also tested the effect of calcium by STD, as the activation of EF by CaM is calcium-dependent. As can be observed in Figure 35, STD data show that Tween 20 at 0.05% (> CMC) displaces the aggregation of TUA-diCl to higher concentrations and at 0.005% (< CMC) shows only a modest effect. With regards to calcium, the same atoms of TUA-diCl seem to be involved in self-assembly. However, like Tween 20, calcium ions displace the aggregation equilibrium, requiring higher TUA-diCl concentrations to observe a STD signal in the presence of calcium (Figure 35A, B). This suggests that calcium has an adverse effect on the stacking of TUA-diCl molecules.

Due to its conjugated unsaturation, TUA-diCl is fluorescent, with maxima excitation and emission wavelengths of 350 and 504 nm (at low concentrations), respectively. Like its NMR spectra, the fluorescence spectra (normalized intensity with respect to concentration, shape and maximum emission wavelength) of TUA-diCl are concentration-dependent, and can thus be used to monitor its aggregation. At high concentrations of the molecule, as generally observed for hydrophobic interactions, a blue shift (higher intensity at 480/530 nm ratios) and a signal increase are observed (Figure 36). A low Tween 20 concentration (0.01%) produces a minor displacement of aggregation towards higher TUA-diCl concentrations, but at 0.05%, Tween 20 micelles efficiently avoid TUA-diCl aggregation even at the highest TUA-diCl concentration (400 µM) tested. NMR and fluorescence data indicate that Tween 20 micelles are able to shift the equilibrium between soluble and aggregated species of TUA-diCl, most likely by sequestering TUA-diCl in micelles.

Tween 20 also shows a negative effect on both, the enzymatic activity of EF and its inhibition by TUA-diCl and the effect is more important with Tween 20 micelles than with near

CMC concentrations (Figure 36). Addition of Tween 20 at a concentration in the CMC range (0.01%) or at 0.05%, leads to a decrease of k_{cat}. In the absence of Tween, addition of TUA-diCl at 50 µM, a concentration at which the compound does not aggregate as stated by NMR (1D, STD) and fluorescence, the catalytic activity of EF is reduced to ~ 7%, (no Tween 20) or ~ 13% at low Tween 20 concentration, whereas in the presence of Tween micelles, its activity is only reduced to ~ 70%. It can be deduced that detergent micelles capture TUA-diCl molecules, leading to lower availability of the compound to interact with the protein.

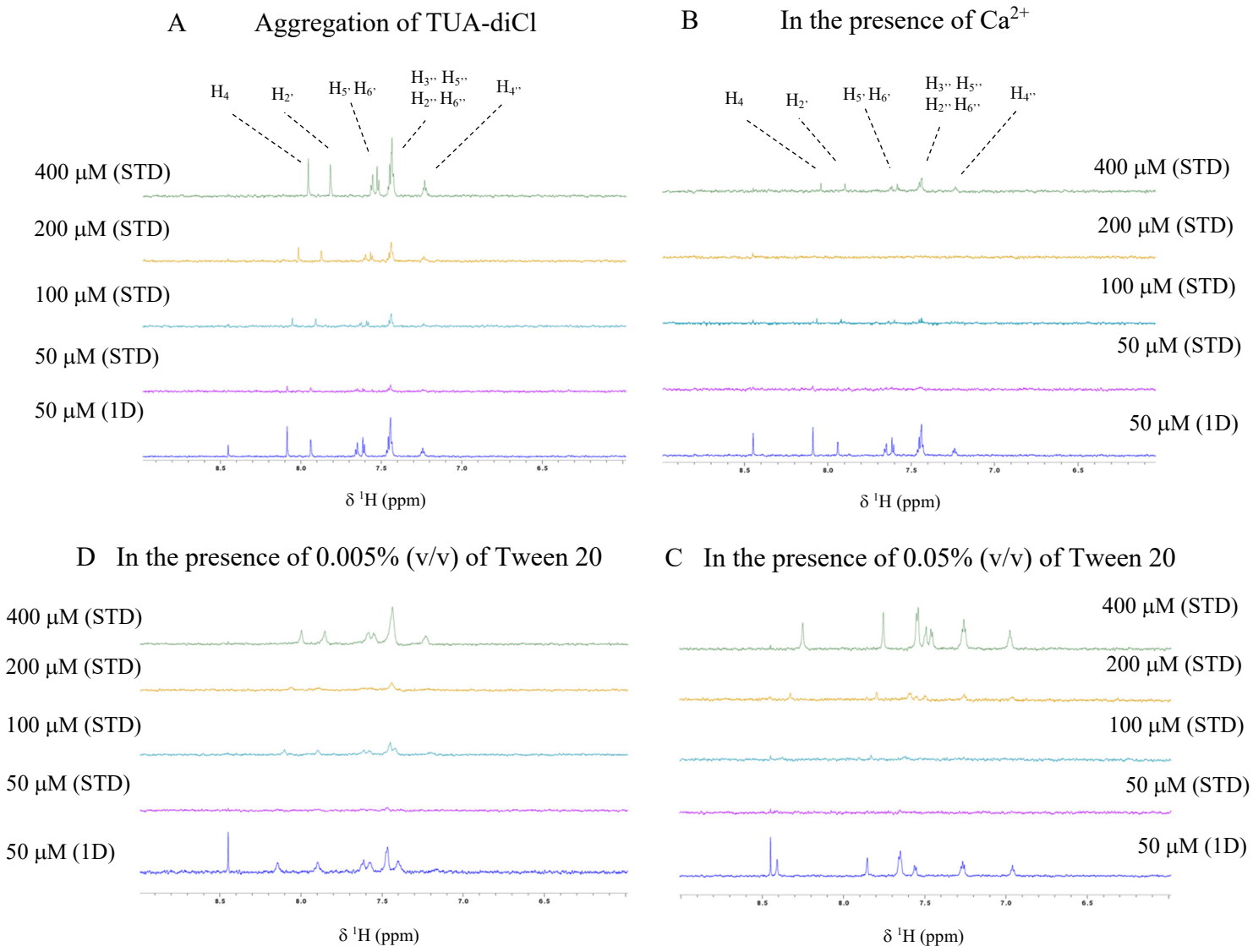


Figure 35- STD spectra of TUA-diCl at varying concentrations *(A) in the presence of 0.2 mM $CaCl_2$ (B) or in the presence of Tween 20 at 0.005% (C) or 0.05% (D) without added calcium.*

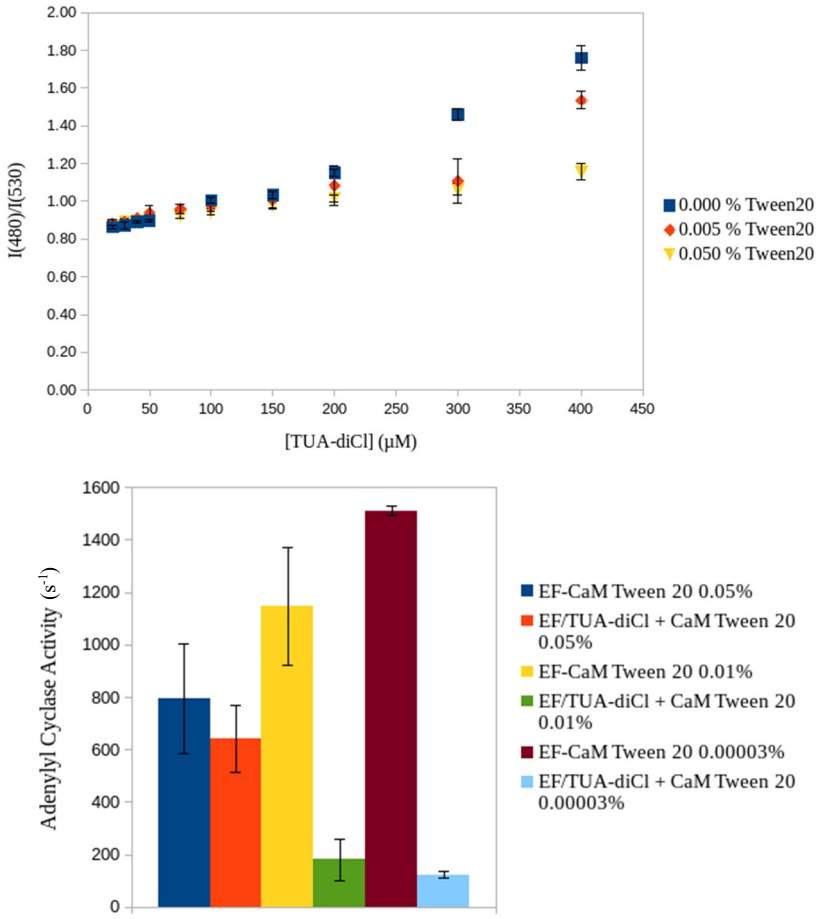


Figure 36- Effect of the Tween 20 on the aggregation and on the inhibition power of TUA-diCl. *(Top) TUA-diCl aggregation monitored by fluorescence (ratio of intensities at 480 and 530 nm) at varying Tween 20 concentrations (0.00, 0.005 and 0.050%). (Bottom) Effect of Tween 20 on the inhibition of EF adenylate cyclase activity. Enzymatic assays were performed as described in the Materials and Methods section C.3, using a fixed concentration of TUA-diCl (50 µM) and 3 Tween 20 concentrations (0.00003, 0.01 and 0.05%). Fluorescence and activity tests were performed in the presence of 0.2 mM $CaCl_2$.*

It is generally admitted that drugs should comply with the so-called Lipinski rules[248]. According to the physicochemical properties predicted using swissADME[249] (Figure 37), TUA-diCl respects three out of the four Lipinski rules: its molecular weight is 405.30 g/mol (< 500 Da), the number of H-bond acceptors is 2 (< 10) and of H-bond donors is 3 (< 5), but the $logP_0$ is 4.17 higher than the limit (> 4.15). TUA-diCl possesses too many unsaturations, not enough soluble groups and the lipophilicity is too high to be predicted as a good drug-likeness molecule. The TUA-diCl has two structural fragments which are typically found in drugs, a urea group and an acid carboxylic[250].

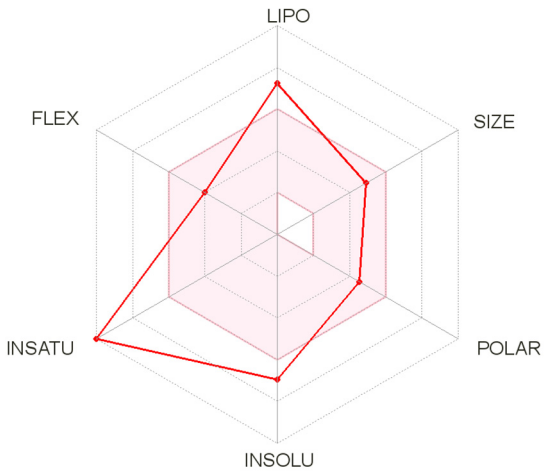


Figure 37- TUA-diCl bioavailability radar obtained with the interface swissADME[249]. *The lipophilicity is 7.16, the molecular weight 405.30 g/mol, the topological polar surface is 89.60 Å^2, log S is -6.95, the saturation is 0.00, the rotatable bonds are 6.*

DISCUSSION

We have shown that TUA-diCl tends to self-aggregate into high molecular weight species (invisible by NMR in the conditions we used) through the thiophen and dichloro-aryl rings and can be sequestered by detergent micelles that reduce its inhibitory effect most likely by reducing the compound availability to interact with the EF-CaM complex. While small compound aggregation is clearly a disadvantage to specifically target a protein, we found conditions in which the compound does not auto associate-significantly (50 µM and lower concentrations), found that calcium ions that are required for EF activation by CaM, are unfavorable for aggregation and importantly, we showed that the compound does not unspecifically inhibit EF through its aggregates. Finding conditions in which the compound does not aggregate and inhibits EF paves the way to understand its mode of EF inhibition, which is the purpose of my doctoral work.

TUA-diCl has several chemical groups that can potentially interact with a protein. The electronegative groups /atoms such as chlorides and thiophen can act as cationic pincers. In that sense, the observed inhibitory effect of Ca^{2+} on TUA-diCl aggregation suggests that the compound could indeed be a cationic pincer (and brings into mind a possible mechanism for the mechanism of Ca^{2+} inhibition of aggregation, i.e. interaction of Ca^{2+} could unfavorably change the stacking of the thiophen and dichloro-aryl rings). TUA-diCl can potentially interact

with proteins via hydrophobic forces including π–π stacking of its three aromatic groups, hydrogen bonds with the carbonyl of the carboxylic acid and the urea moieties, π-cation interactions or ionic interactions of its acidic groups (Figure 38)[251–254]. All these groups offer wide possibilities of interaction. Thiophen and its derivatives have indeed been used in numerous therapeutic approaches and demonstrated antimicrobial, anti-inflammatory, anticancer, anticonvulsant, antidepressant, analgesic antimalarial, antiviral, antihypertensive and antioxidant activities[255,256]. However, the thiophen combined with a urea group can present disadvantages such as bioavability issues and hepatic toxicity[257].

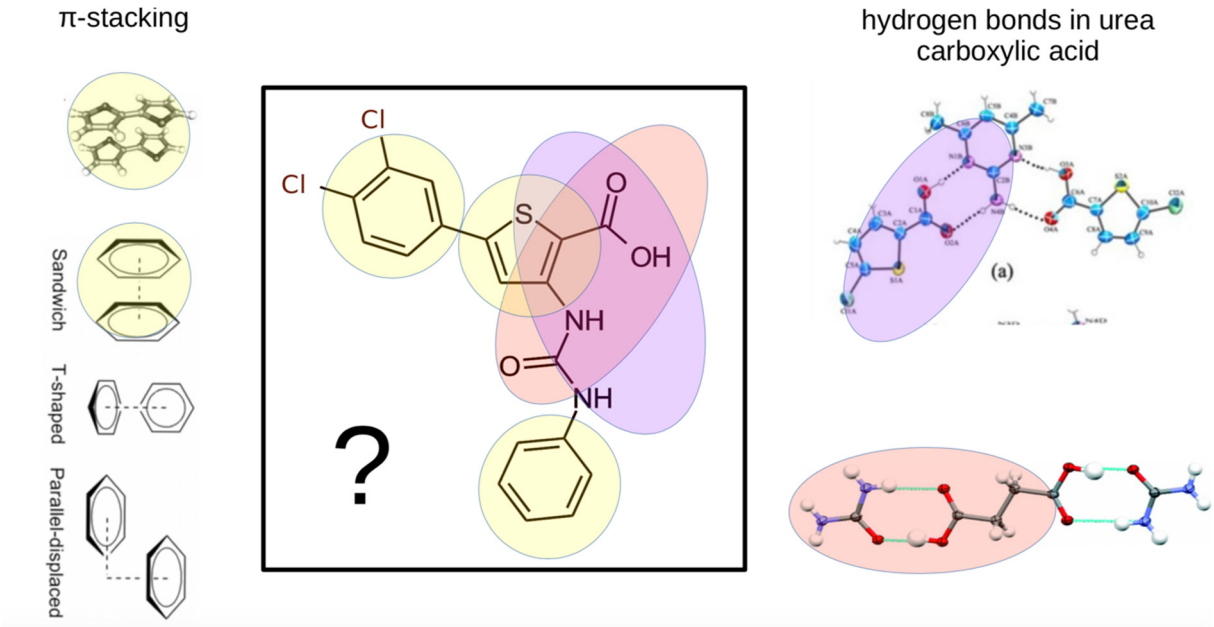


Figure 38- The chemical groups of the TUA-ligand that can interact with different partners. *π–stacking is showed in yellow, the hydrogen bonding groups of urea and carboxylic acids moieties are represented in orange and purple.*

TUA-diCl has been used in two different studies. In the first one, which is at the origin of this work, TUA-diCl was found *in silico* targeting an allosteric pocket in EF and AC. TUA-diCl was tested *in vitro* and showed inhibition of CaM activated EF and AC[192]. In a second unrelated study, this compound was also found *in silico* when targeting the switch region of the RNA polymerase (RNAP)[258]. TUA-diCl and other TUA compounds were active *in vitro* and bound the RNAP switch region as shown by competition STD and NOESY experiments[259]. In both cases, the ligand targeted hydrophobic cavities and the di-chloro compounds were found to be the most active within the family, indicating that the di-chloro group plays an important role in inhibition. In the case of EF, another compound named 10506-2A was found to bind in

an allosteric site[181]. Both, TUA-diCl and 10506-2A are composed of three hydrophobic cycles. The benzenesulfonamide and phenylthiazole groups are essential for the EF inhibition activity of 10506-2A[181], suggesting that in addition to the di-chloro phenyl cycle, the thiophen group coupled with the phenyl group would participate in the inhibitory effect of TUA-diCl (Figure 39).

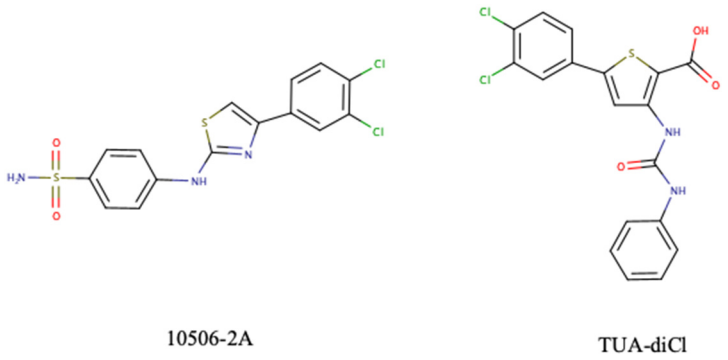


Figure 39- Chemical structures of the 10506-2A on the left and on the TUA-diCl on the right.

Although TUA-diCl will definitively not be a drug against EF, it is a chemically interesting starting point to find a potential hit against EF. Moreover, it is predicted to bind to an allosteric site. If this hypothesis were confirmed, TUA-diCl would reveal an allosteric site avoiding specificity issues due to the presence of ATP binding active sites in many human and bacteria proteins. Knowing the binding site and obtaining the structure of the complex would be crucial to modify the TUA-diCl moieties to increase its inhibitory effect prior to be pharmacological optimized. Several techniques exist to optimize molecules such as building pharmacophore model[260,261] and recently with the development of deep learning[262], solubility and biodisponibility can be increased and cytoxicity reduced, which is critical for the therapeutic approaches of this project.

Our goal was to understand the mechanism of EF inhibition and determine whether it binds to an allosteric site and identify its binding site on EF as a first step to design better inhibitors with better physicochemical and pharmacophore properties.

D.I.2 Biochemical and biophysical analysis of the catalytic activity of EF_3

B. anthracis and *B. cereus* belong to the *B. cereus* family but only *B. anthracis* is a pathogenic class II strain. For security issues, the following study was focusing on adenylate cyclase produced by pathogenic strains of *B. cereus*. Indeed, these strains, isolated either in the USA (strain G9241) or in Africa from great apes (CA or CI strains), produced an enzyme EF_3 with the same activity as that of EF_3 isolated from *B. anthracis*. These two strains from *B. anthracis* and *B. cereus* had three or four different amino acids. They were located away from the enzymatic site and thus do not affect the enzymatic activity (Figure 40). In addition, products derived from *B. cereus* strains are not affected by micro-organism and highly pathogenic toxin (MOT) regulations.

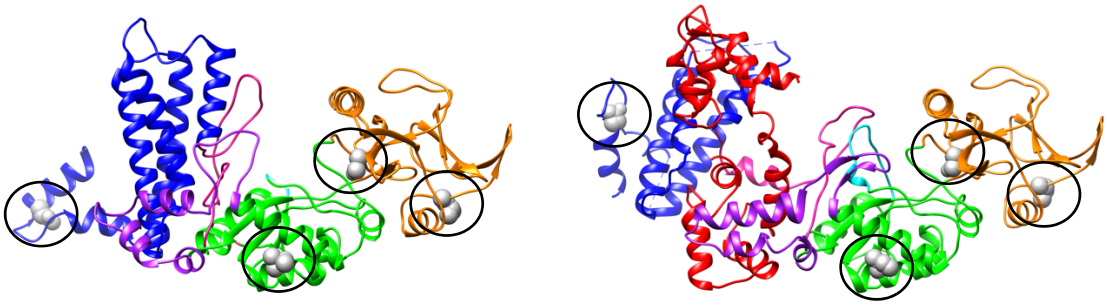


Figure 40- Location of the different amino acids between B. anthracis and B. cereus. *The X-ray structure of EF_3 (A, PDB ID 1K8T) and EF-CaM (B, PDB ID 1K93) with the four different amino acids. The four amino acids (black circle) are located at the solvent surface of the C_A and the helical domain of EF_3 far away from the catalytic site and the interaction site with CaM.*

RESULTS

D.I.2.1 Optimization of EF_3 expression

The recombinant *B. cereus* EF_3 protein expressed here comprises the adenylyl cyclase domain without the protective antigen binding domain (i.e. corresponding to residues 291 to 798 of EF). We chose to work with this species for the crystallogenesis and crystallization study. Most of the X-ray crystallographic structures of EF found in the protein data bank are without PABD as it is a flexible domain that could interfere with the crystallogenesis of the protein.

The aim of the optimization of protein production was to obtain a high yield and soluble protein. The following parameters were examined: host bacterial strains, temperature and expression time. In total, four expression conditions were investigated (Figure 41). For each strain (BL213 pDia17 or BL21 DE3 Star^{TM} CodonPlus) expression during 8 h at 37 °C or 4 h at 37 °C and then 16 h expression at 16 °C.

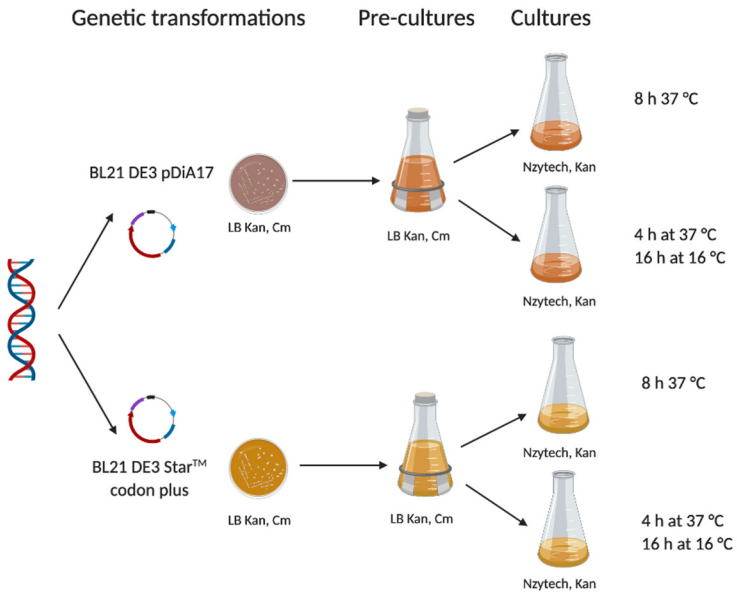


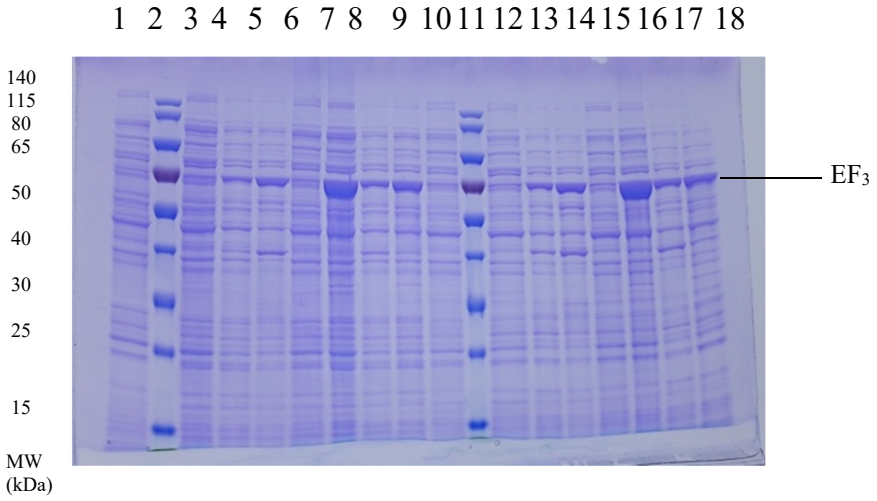
Figure 41- Strategy for protein overexpression. *The scheme represents the different conditions tested. Expression was auto-induced after 4 hours of growth at 37°C when the cultures reached an OD600 of 3.5-4.*

The two strains used for the expression of EF_3 were *E. coli* BL21 DE3 derivatives that integrate on their chromosome a copy of the gene coding for the polymerase of phage T7 that is under the control of a Lac promoter. This T7 polymerase will activate the transcription of the gene coding for EF_3 that is placed downstream of the T7 promoter on plasmid ETM11-EF2-Short. The strategy for overexpressing EF_3 was to use 2 competent strains of *E. coli*: BL21 DE3 pDiA17 and BL21 DE3 Star^{TM} CodonPlus. The BL21 DE3 Star^{TM} CodonPlus promotes the stability of mRNA transcripts and expresses tRNAs that recognize rare codons in E. coli and can thus increase the expression levels of certain proteins. The plasmid pDiA17 codes for the LacIq repressor in order to better repress the expression of the T7 polymerase before induction (and thus reduce the basal expression of the EF_3 protein). The preculture was performed in Luria-Bertani (LB) for the four conditions. Bacterial strains of *E. coli* were grown in NZY auto

induction LB medium (Nzytech) in presence of the kanamycin antibiotic (50 µg/mL). Isopropyl β-D-1-thiogalactopyranoside (IPTG) is not added for Nzytech culture. By using an autoinduction media, the induction of the T7 polymerase will be progressive and high cell densities will be obtained. The T7 polymerase then triggers high expression of the plasmid born EF_3 gene. The culture conditions used are *E. coli* BL21 DE3 Star™ CodonPlus 4h at 37°C and 16°C and *E. coli* BL21 DE3 pDia17 4 h at 37 °C and 16 °C. The cultures were performed in presence of 50 µg/mL of kanamycin. The absorbance at 600 nm was recorded at the beginning, before induction (4 h) and at the end of the culture (Table 3). The different conditions were tested on 100 mL cultures.

	***E. coli* cells**	**OD600 at 0 h (initial)**	**OD600 at 4 h (before induction)**	**$OD600_1$ at 8 h (End of the culture)**	**$OD600_2$ at 16 h (End of the culture)**
Condition 1: 8 h à 37 °C	pDIA17	0.336	3.98	6.48	
	Star™ CodonPlus	0.300	3.56	6.84	
Condition 2: 4 h at 37 °C and 16 h at 16 °C	pDIA17	0.328	4.14		5.52
	Star™ CodonPlus	0.289	3.82		5.12

Table 3- Culture growth monitored by OD600 measurements



condition 1 pDia17					condition 2 pDia17				condition 1 Star™ CodonPlus					condition 2 Star™ CodonPlus			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
S		S	I		S		I		S		S	I		S		I	
4 h	marker	8 h	4 h	8 h	4 h	16 h	4 h	16 h	4 h	marker	8 h	4 h	8 h	4 h	16 h	4 h	16 h

Figure 42- SDS-PAGE evaluation of the different growth conditions EF_3 overexpression *For each condition S represents the soluble fraction of the lysate and I the insoluble fraction before induction (4 h) and at the end of the culture (8 h condition 1 or 16 h condition 2). The gel was stained by Coomassie blue. EF_3 corresponds to the major band of about 65 kDa (according to molecular size markers run in lanes 2 and 11).*

The final biomass is higher in condition 1 ($OD600_1 > OD600_2$, Table 3). Extracts of the cells grown at the different stages were analyzed by SDS-PAGE which was performed to establish which conditions led to the best over expression of EF_3 (Figure 42).

EF_3 was well expressed in all conditions. Yet, temperature conditions influenced expression and the folding of the protein. In the 1st condition (8 h at 37 °C), the protein was insoluble whereas in the 2nd condition (4 h at 37 °C and 16 h at 16 °C) both proteins were soluble at the end of the culture. Condition 2 was found to be the most favorable. By decreasing the temperature, the folding of the protein and the yield of the relative fraction soluble form of EF_3 was favored. The *E. coli* strains behaved similarly in terms of production of soluble protein. We hence chose to work with *E. coli* Star™ CodonPlus grown 4 h at 37 °C and 16 h postinduction at 16 °C. After culture, cells (15 g per L) were collected by centrifugation and stored at - 80 °C.

A scale-up of the production of EF_3 was performed using a 4 L bioreactor at the Production and Purification of Recombinant Proteins Technological Platform. The conditions were similar to those used in the 1 L flask but the media and induction were adapted for a 4 L batch. We used a high-density medium in the presence of 1% (v/v) glucose, 50 µg/mL kanamycin and 30 µg/mL chloramphenicol. Bacteria were inoculated at an OD_{600} = 0.75, the culture was induced by adding 1 mM of IPTG after 4 h (OD_{600} = 22, Figure 43) and culture was continued at 16 °C for 16 h. The advantage of the fermenter was that the level of oxygen was kept constant and not limiting leading to much higher biomass (200 g, 50 g per L). It was important for the following project to have a uniform production batch in order to obtain high level of proteins in a reproducible manner.

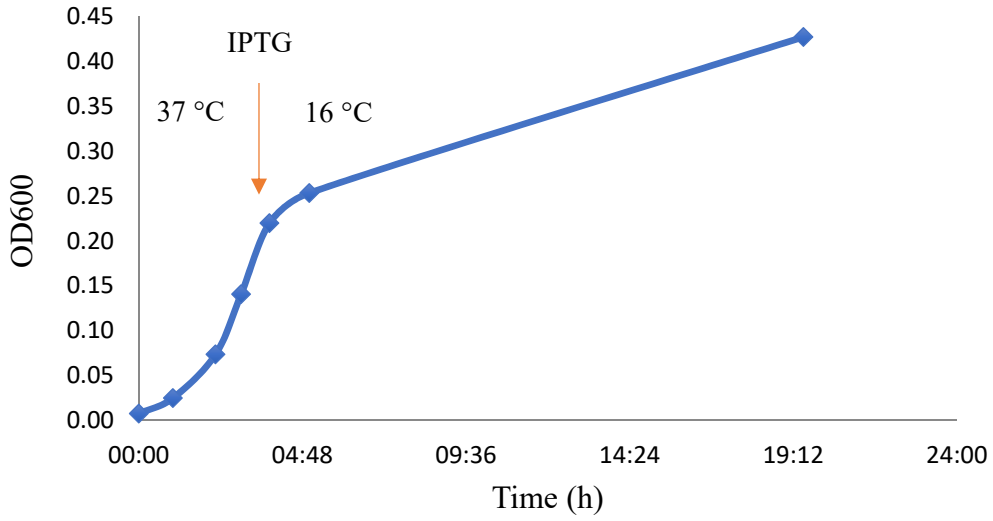


Figure 43- Growth profile of EF_3 expressing cells in a 4 L bioreactor. *The OD600 curve over time is shown in blue.*

D.I.2.2 EF_3 purification

After thawing the lysed cell pellet, EF_3 was purified following a three-days procedure. The purification of EF_3 was facilitated by using the 6-His tag at the N-terminus of EF_3. EF_3 was purified as described in Materials and Methods section C.2. The strategy of purification is based on immobilized metal affinity chromatography (IMAC) with nickel nitrilotriacetic acid (Ni-NTA) matrices as the 6-His-tag located in EF_3 has a high affinity for nickel. The last step of purification is size exclusion chromatography. The elution profile of the last step, the standard size exclusion chromatogram is represented in Figure 44.

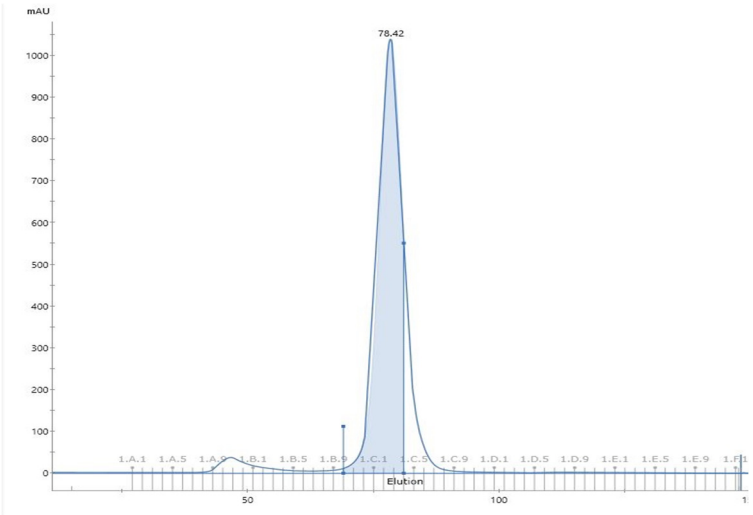


Figure 44- Size exclusion chromatography of purified EF_3 protein elution. *The blue peak corresponds to the selected EF_3 fractions.*

D.I.2.3 Quality control

The protein sample was further checked by different techniques such as SDS-PAGE, mass spectrometry, DLS, UV spectra (Figure 45) and by activity tests. The concentration of the protein has been assessed by UV spectra at 280 nm in a quartz cuvette. The average amount of purified proteins obtained for 15 g of cell pellet was about 20 mg. EF_3 appeared rather uniform by DLS with a hydrodynamic radius of 2.17 nm, although a negligible amount of high aggregates could be detected.

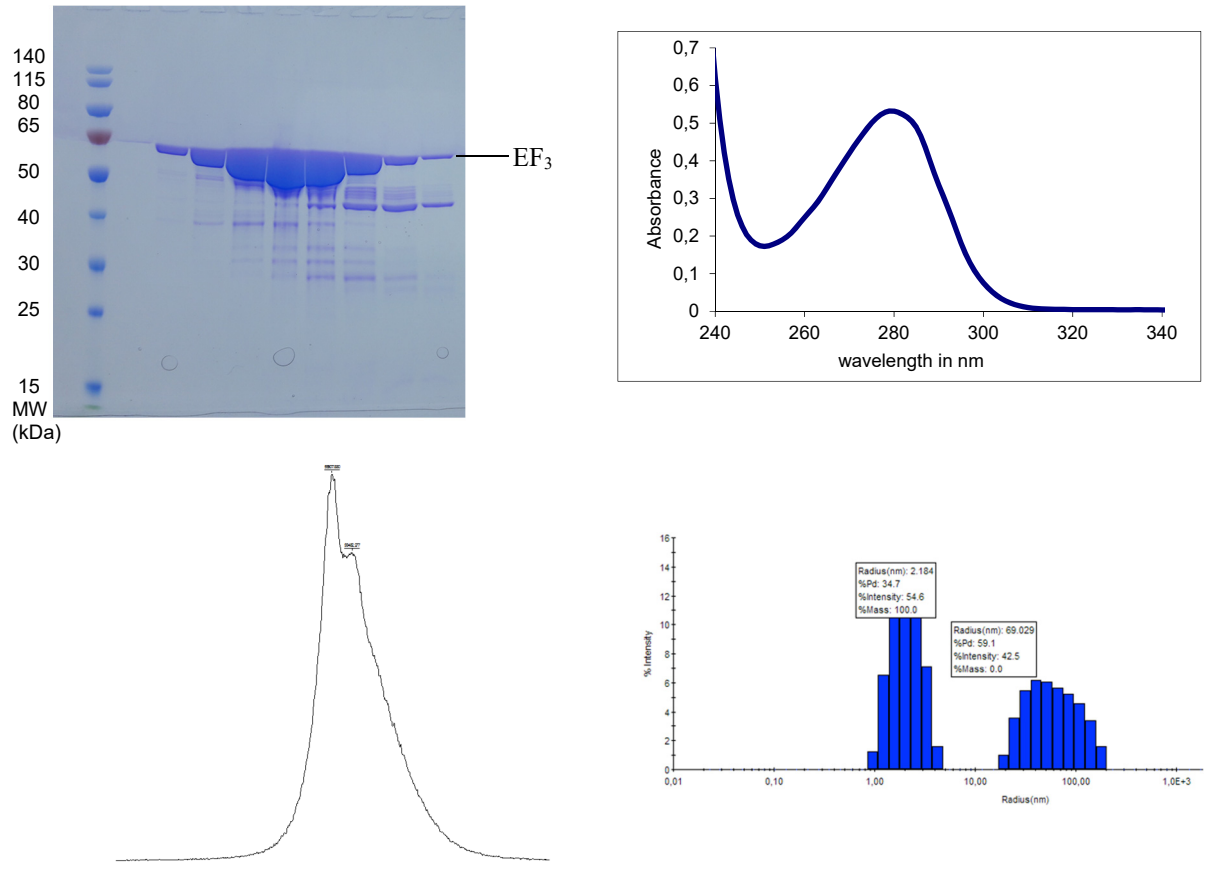


Figure 45- Quality control of EF_3 *(20 mM Tris-HCl, 7.5, 100 mM NaCl) A. SDS-PAGE of the EF_3 gel filtration chromatography B. UV spectrum C. MALDI-TOF and D. DLS*

Two pics are present on the MALDI-TOF, meaning that the protein was present in two forms. The difference between these two pics are 178 Da, and likely due to alpha-N-6-phosphogluconoylation which is not surprising for a 6-His tagged protein purified using *E. coli* cells[263]. The sample is validated and further used for X-ray crystallography, NMR, inhibition

activity assays, fluorescence purpose. A buffer screening was performed due to crystallization problems.

D.I.2.4 Buffer optimization

Different buffers were tested using NanoDSF and DSL methods to study different stability conditions in order to attempt to solve crystallization issues. Twenty conditions were tested. It was important to study the behavior of the protein in different buffers in order to improve the purification protocol, the crystallogenesis conditions and storage of the protein. Five different pH (6.0, 7.0, 7.5, 8.0, 8.5) and four concentrations of NaCl (50, 150, 300, 500 mM). The best condition deduced from these tests was 20 mM Tris-HCl pH = 8.5, 500 mM NaCl (Figure 46). In this condition, the protein has a low polydispersity (> 20%). Conditions with high salinity (> 300 mM) show the best Tm. The condition 20 mM Tris-HCl pH = 8.5, 500 mM NaCl was among the best one. This condition was implemented in the purification protocol of EF.

Capillary	Sample ID	ON (Ratio) ↗	ON (Scattering) ↗	IP #1 (Ratio) ↗
7	20 mM HEPES pH 7 300 mM NaCl	34.9°C		41.9°C
8	20 mM HEPES pH 7 500 mM NaCl	35.3°C		41.9°C
20	20 mM TRIS pH 8.5 500 mM NaCl	33.3°C		41.7°C
11	20 mM HEPES pH 7.5 300 mM NaCl	36.4°C		41.6°C
6	20 mM HEPES pH 7 150 mM NaCl	36.3°C		41.3°C
5	20 mM HEPES pH 7 50 mM NaCl	32.4°C		41.2°C
12	20 mM HEPES pH 7.5 500 mM NaCl	34.3°C		41.1°C
15	20 mM TRIS pH 8 300 mM NaCl	33.8°C		40.9°C
14	20 mM TRIS pH 8 150 mM NaCl	35.8°C	42.4°C	40.9°C
18	20 mM TRIS pH 8.5 150 mM NaCl	34.4°C	41.7°C	40.8°C
9	20 mM HEPES pH 7.5 50 mM NaCl	36.3°C	38.8°C	40.7°C
13	20 mM TRIS pH 8 50 mM NaCl	36.6°C	39.4°C	40.6°C
17	20 mM TRIS pH 8.5 50 mM NaCl	36.1°C	38.5°C	40.5°C
19	20 mM TRIS pH 8.5 300 mM NaCl	24.5°C		40.5°C
10	20 mM HEPES pH 7.5 150 mM NaCl	28.0°C		40.0°C
3	20 mM MES pH 6 300 mM NaCl	35.7°C		39.9°C
4	20 mM MES pH 6 500 mM NaCl	35.2°C		39.5°C
16	20 mM TRIS pH 8 500 mM NaCl	30.1°C		39.0°C
1	20 mM MES pH 6 50 mM NaCl	34.1°C	32.2°C	37.8°C
2	20 mM MES pH 6 150 mM NaCl	31.6°C		37.3°C

Figure 46- NanoDSF results. *The table contains the twenty conditions, with the different buffers represented in the second column and the Tm of the protein in the last column.*

D.I.2.5 CaM purification

Calmodulin was purified according to the procedure described in Materials and Methods section C.2. The elution profile of the final step of size exclusion (SEC) is shown in Figure 47. The mass of the eluted protein was checked by mass spectrometry and corresponded to the expected mass for CaM.

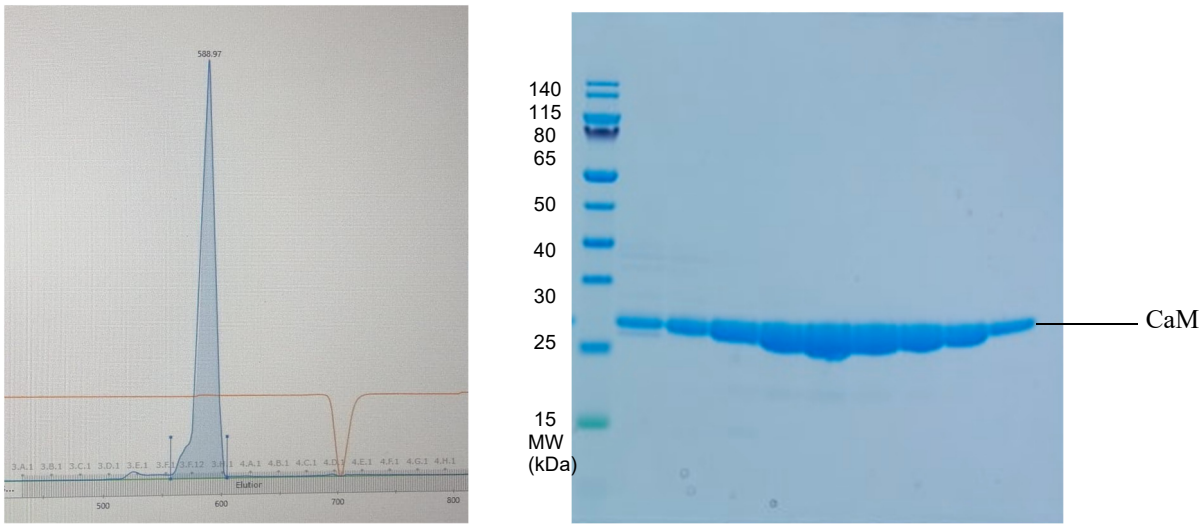


Figure 47- CaM sample. *A. Gel filtration (GF) chromatogram B. SDS-PAGE of the SEC step of CaM purification. The CaM band (16 kDa) is visible between the 30 and 25 kDa markers.*

D.I.2.6 EF_3-CaM complex formation

The complex was formed in the presence of Ca^{2+} by adding two molar equivalents of EF_3 to three molar equivalents of CaM[65]. The EF_3, CaM and their complex have different elution times in the SEC, allowing the isolation of the EF_3-CaM complex and showing that a stable complex is formed (Figure 48). The complex His_6-EF_3-CaM was obtained using the same conditions as for EF_3-CaM. The formation of the complex was also identified by ^1H 1D spectra NMR (Figure 49). The ^1H 1D spectrum of EF_3-CaM was not the resulting addition of the ^1H 1D spectra of EF_3 and CaM highlighting the formation of the complex EF_3-CaM.

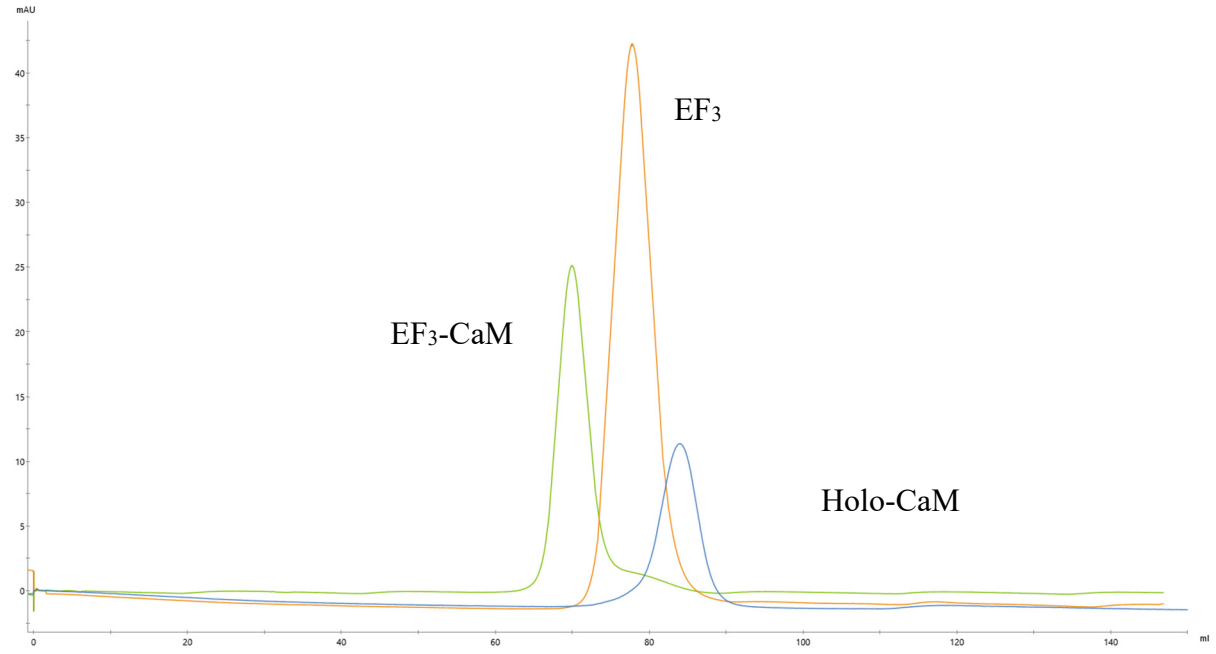


Figure 48- SEC elution of the three species. *The EF_3-CaM complex (green) eluted on a S-200 at ~ 70 mL, EF_3 (orange) had an elution volume of ~ 80 mL and CaM (blue) eluted at ~ 85 mL.*

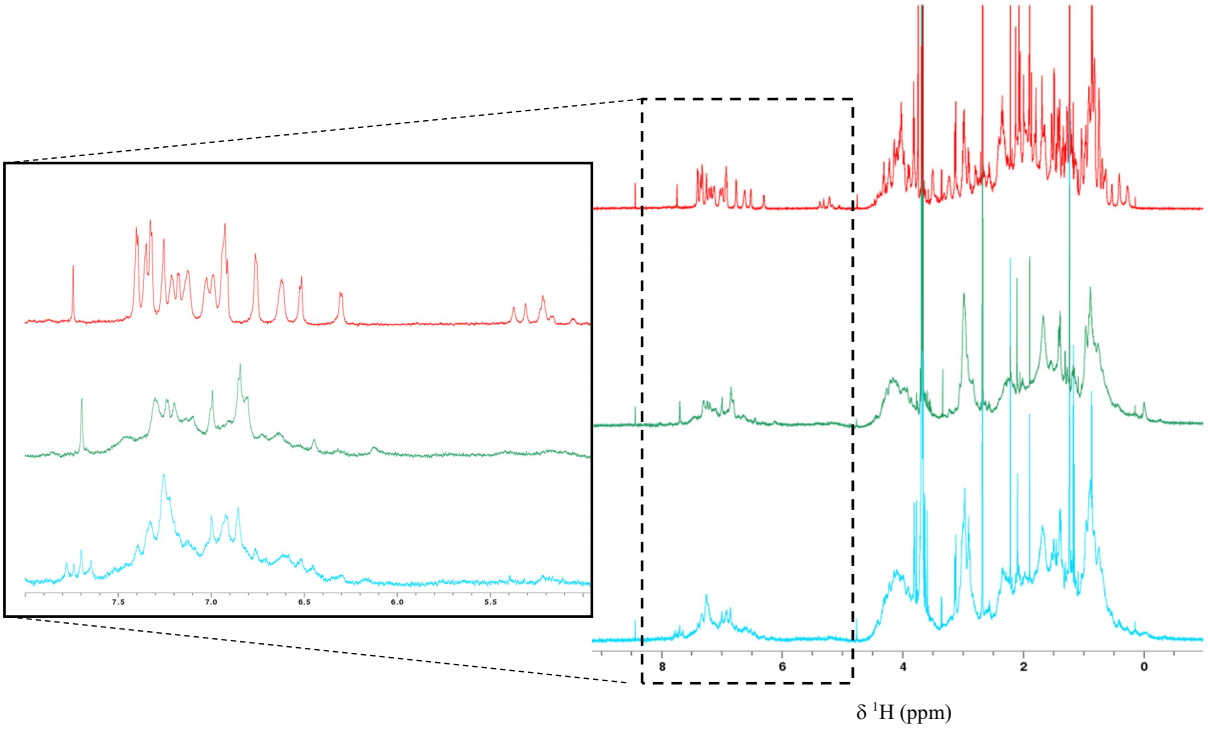


Figure 49- 1D NMR spectra of the three species, *CaM (red), EF_3 (green) and the complex EF_3-CaM (cyan). The concentration of protein used was 20 µM.*

The overtime stabilities of EF_3, CaM and the complex were tested using NMR, enzymatic activity and SEC. Importantly, as revealed by SEC, the complex was still stable after one week at 4 °C (Figure 50), rendering possible crystal formation. NMR ^1H 1D spectra of EF_3, CaM and EF_3-CaM also showed that the proteins were stable for several days.

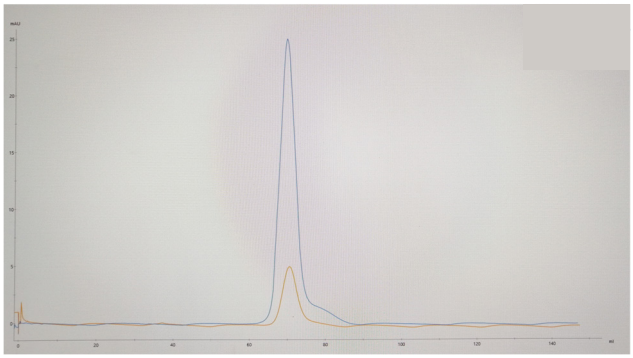


Figure 50- Stability of the EF_3-CaM complex monitored by SEC. *In blue gel filtration chromatograms of the EF_3-CaM complex a time 0 (blue trace) and after seven days at 4 °C (orange trace). In the second injection (orange trace) less protein was loaded onto the SEC.*

D.I.2.7 Adenylyl cyclase activity of EF_3

The enzymatic activity of *B. cereus* EF_3 was determined in the presence of different CaM concentrations as described in the Materials and Methods C.3. The catalytic rate (k_{cat}) of EF_3 at saturating CaM concentrations (ie above 100 nM) was around ~ 1000-2000 cAMP molecules produced per second per EF_3 molecules, in good agreement with prior results from the literature or the *B. anthracis* enzyme[83]. As TUA-diCl is hydrophobic, it was conserved in DMSO and biophysical experiments were performed with 5% (v/v) DMSO to increase its solubility, the catalytic activity of EF_3 was tested in the absence or in the presence of DMSO (Figure 51). We found that DMSO (5%) did not affect the enzymatic catalytic: in both cases ~ 1 200 molecules of cAMP were converted per second. The adenylyl cyclase activity of EF_3 was half-maximal at about 5 nM of CaM, in good agreement with prior studies[110] both in the presence or in the absence of DMSO. As the interaction of EF_3-CaM is strongly calcium-dependent, we investigated the effect of Ca^{2+} chelating EGTA on the EF_3 activity and found that EF_3 requires calmodulin binding to be active.

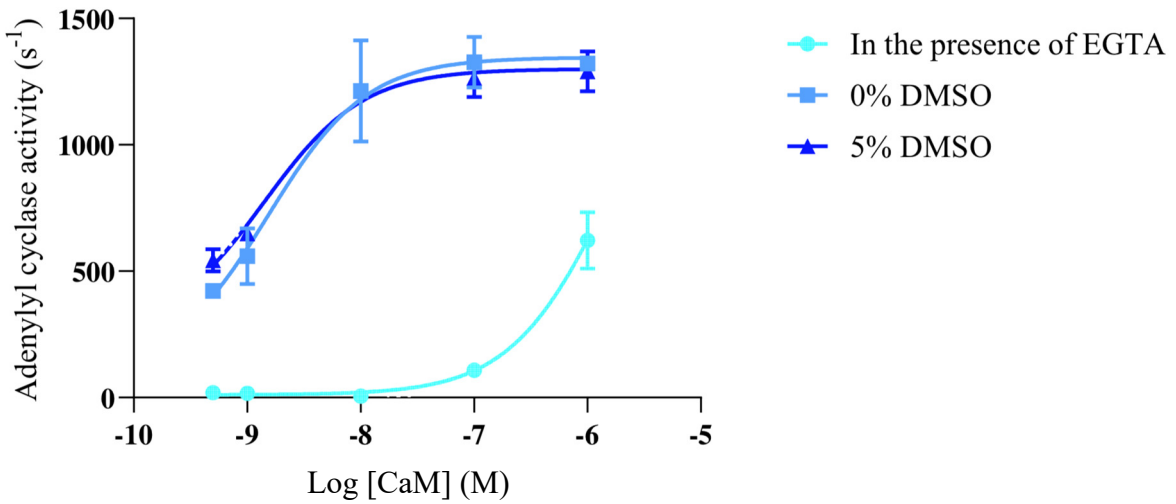


Figure 51- EF_3 enzymatic activity assays as a function of CaM. *The adenylyl cyclase activity (expressed as cAMP produced per sec per enzyme molecule) of EF_3 (0.4 nM) was measured at the indicated CaM concentrations in the presence of 5% (v/v) DMSO (square) or without DMSO (diamonds) in the presence of 0.1 mM $CaCl_2$, or the presence of 2 mM EGTA without DMSO (triangles). EF_3 and CaM were mixed for 5 min at 30 °C under agitation, the reaction was started by adding ATP (2 mM) to the solution and stopped after 10 min. The cAMP produced was measured by recording the absorbance at 260 nm and 340 nm as described in Materials and Methods section C.3.*

Chapter II. Molecular interactions between TUA-diCl, edema factor and its host activator calmodulin

INTRODUCTION

Previous studies in the lab focused on identifying potential allosteric inhibitors of the edema factor[192]. An in homemade method was developed and used to target an allosteric pocket found in adenylate cyclase EF and AC. For EF, the method was based on modeling the transition path between the inactive state of the enzymes without CaM and the close active state in interaction with CaM. In the case of EF, the insertion of CaM between the helical domain and the catalytic core of EF induces the formation of the active site. The pocket was identified by detecting cavities present in the inactive state of EF that disappear along the transition path to the active state. The chosen pocket was located at the interface of switches A, B and C of EF, called SABC. Then, a virtual targeting SABC screening was carried out. The hit compounds that belong to the ureidothiophene-carboxylic acids (TUA) family, were further tested *in vitro* on an enzymatic activity assay revealing that TUA-diCl exhibited the lowest IC_{50} (2 μM)[192]. The TUA-compound mechanism was though to lock EF in an inactive open conformation thus preventing the insertion of CaM in EF. By targeting an allosteric pocket close to the CaM binding site, the association of CaM and EF would be disrupted and TUA compounds could function like inhibitors of protein-protein interaction (IPPIs). Interestingly, in a screening of EF inhibitors, only one compound named (10506-2A) and belonging to the TUA family was found to act as an IPPIs of EF-CaM[181]. Here, to understand the mode of action of TUA-diCl, we characterize its interaction with EF and its activator CaM using enzyme kinetics studies, NMR ligand-observed techniques, HDX-MS and molecular docking. Although EF is the main focus of this work, we also performed some experiments on *Bordetella pertussis* CyaA AC domain.

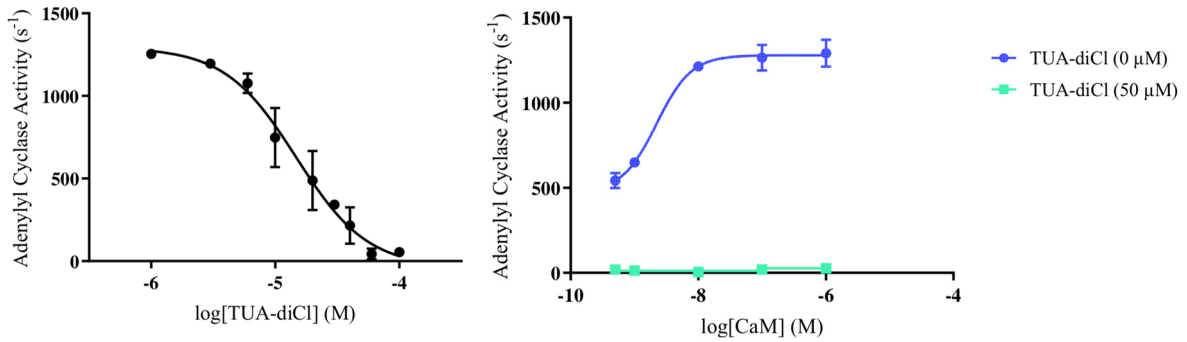
RESULTS

D.II.1 TUA-diCl inhibition of adenylyl cyclase activity

The inhibition effect of TUA-diCl was investigated on both EF_3 and AC384 (residues 1-384 of CyaA) which corresponds to the catalytic domain of CyaA, the adenylyl cyclase toxin secreted by *Bordetella pertussis* (kindly provided by D. Ladant team, Institut Pasteur).

The *in vitro* adenylyl cyclase inhibition assay was performed with 20 nM of CaM, which was determined to be the lowest concentration of the activator for maximum EF_3 and AC384 activity, while preventing a high concentration of free holo-CaM, which is known to interact with hydrophobic small molecules (Figure 52). TUA-diCl was pre-incubated with EF_3 or CyaA and then CaM was added to the solution. In the assay, TUA-diCl showed a ~ 2-fold lower IC_{50} for EF_3 15 µM ± 1 than for AC384 37 µM ± 1.

A EF_3



B AC384

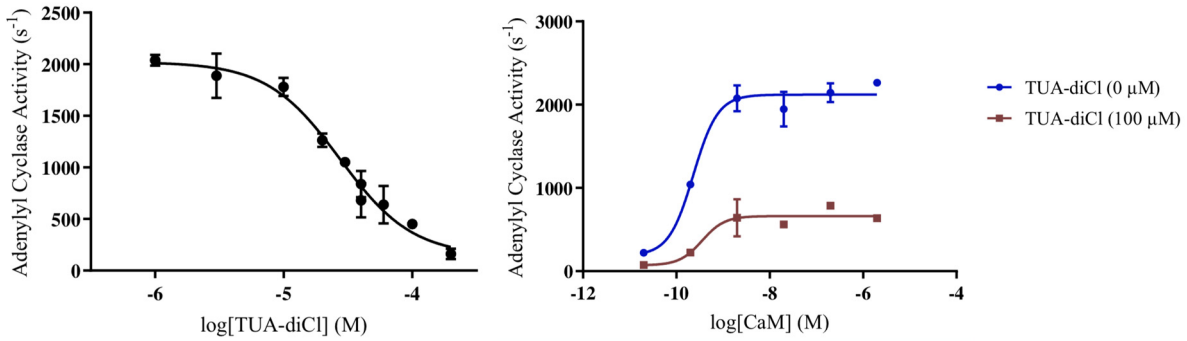


Figure 52- TUA-diCl inhibition of adenylyl cyclase activity of EF_3 (A) and AC384 (B). *The dose-response curve (left) and the inhibition at fixed TUA-diCl and varying CaM concentrations (right) are shown. TUA-diCl inhibits EF_3 and AC384. Increasing the concentration of CaM did not affect the inhibition of the compound.*

Although TUA-diCl was selected *in silico* to target directly EF_3 and prevent its activation, the compound is hydrophobic and could potentially interact with holo-CaM and thereby inhibit EF_3 or AC384, as observed for some CaM-dependent adenylate cyclases[188].

In order to establish if TUA-diCl exerts its inhibitory effect on EF_3/AC384 by binding to CaM, the concentration of the inhibitor was fixed to a concentration showing a high inhibition (50 µM for EF_3 and 100 µM for AC384 Figure 52 on the right panel) while the concentration of the activator CaM was varied from 5 nM to 1 µM. At 50 µM of TUA-diCl, its inhibition of the compound on EF_3 was not displaced by increasing CaM concentrations. Indeed, above the minimum CaM concentration for maximum activity, adding CaM in the presence of TUA-diCl did not increase the enzymatic activity. We observed the same result with AC384 at 100 µM of TUA-diCl. Hence, at a high concentration of TUA-diCl, the inhibition does not depend on the concentration of CaM (IC_{50} = 12 ± 1 µM). To further test if EF_3 inhibition is CaM-dependent, we compared the dose-response curves of TUA-diCl at three different CaM concentrations (20 nM, 100 nM and 1 µM) (Figure 53). The concentration of CaM did not effectively influence the inhibitory effect of TUA-diCl, suggesting that TUA-diCl is not a CaM-dependent adenylyl-cyclase inhibitor.

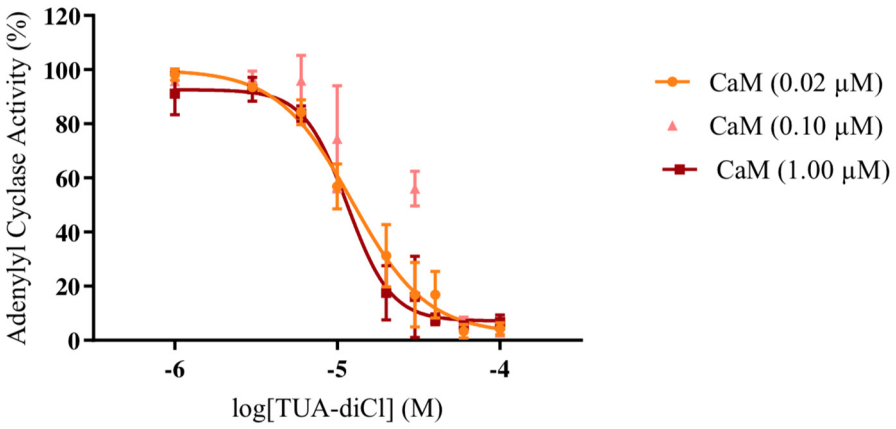


Figure 53- Activation of EF_3 with different concentrations of CaM in the presence of TUA-diCl. *The adenylyl cyclase activity was recorded with three different concentrations of CaM (0.02, 0.1 and 1 µM) in the presence of varying TUA-diCl concentrations. The adenylyl cyclase activity is represented in percentage.*

D.II.2 Interaction of TUA-diCl with the EF_3, CaM and EF_3 -CaM

The interaction of TUA-diCl with the different proteins was investigated by ligand-observed NMR techniques (WaterLOGSY, STD) to better characterize its mode of action.

Like the STD experiment presented in Chapter I, the WaterLOGSY experiment is based on the transfer of saturation information between the bound and free forms of a ligand. The technique is based on the observation that most X-ray structures of protein-ligand complexes show resident water mediating interactions between the ligand and the protein and on the fact that resident water will tumble more slowly that hydration and bulk water, behaving like a 'big molecule'. The sign of saturation of the resident water will be inversed relative to the sign of bulk and hydration water and this will be transmitted to the free ligand that was in contact with resident water within the complex. In practice, in the WaterLOGSY experiment, water is selectively saturated. If a small molecule binds to the protein, its signals will be of the opposite sign with respect to the signal of other small molecules in the sample. Of note, the sign of the signals of protons that directly exchange with water (NH, NH2…), will show the same sign as the ligand signals. As shown in Figure 54, the signals of TUA-diCl are positive in the WaterLOGSY spectra of TUA-diCl in the presence of EF_3, EF_3 -CaM and, surprisingly of holo-CaM, in contrast to the signals of DMSO at 2.67 ppm, Tris at 3.65 ppm, an impurity at 8.43 ppm and the signals of non-exchangeable protons of isolated TUA-diCl. Hence, TUA-diCl binds to the three proteins.

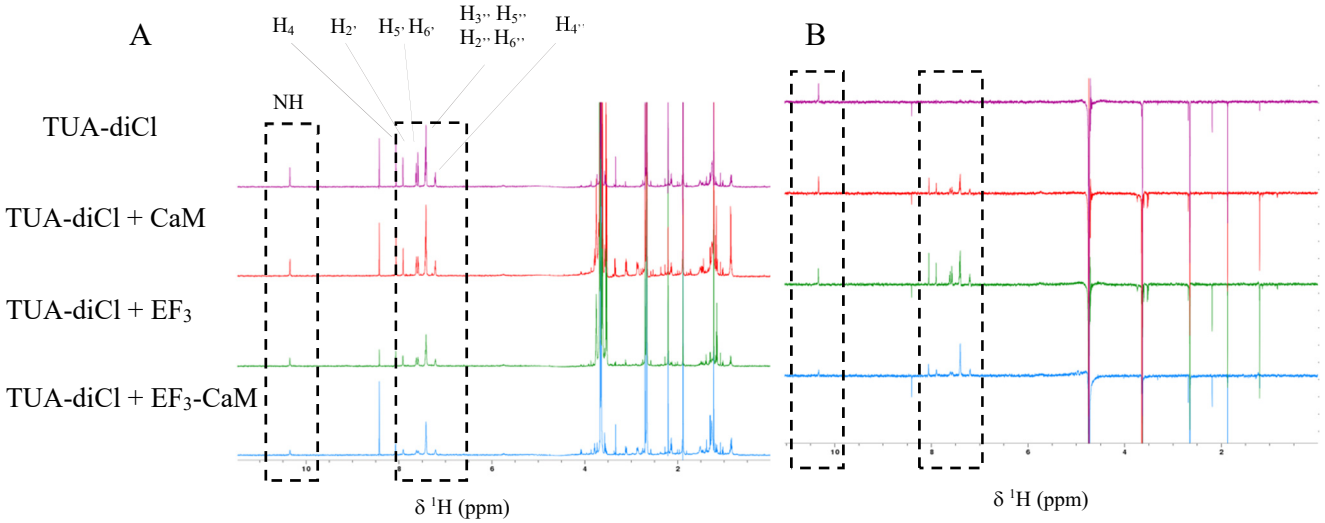


Figure 54- TUA-diCl interaction with CaM, EF_3 and EF_3 -CaM by WaterLOGSY. *^{1}H reference NMR spectra are shown on the panel A and WaterLOGSY experiments on the panel B. Protein-ligand ratios were 1:100. From top to bottom, as indicated by the labels: TUA-diCl (purple), TUA-diCl in the*

presence of CaM (red), EF_3 (green) and EF_3 -CaM (blue). Please note that the phenylamine NH signal at 10.35 ppm is positive and intense in the Waterlogsy spectrum of TUA-diCl in (A), because it directly exchanges with water. Signals of residually protonated Tris (Tris-d_{11} deuterated at 98%) at 3.65 ppm, DMSO (DMSO-d_6 deuterated at 99.97%) at 2.67 ppm, an impurity at 8.43 ppm and other buffer impurities are negative in all WaterLOGSY spectra.

We further characterized the interactions of TUA-diCl by STD. As mentioned in Chapter I, the presence of ligand signals in a STD spectrum indicates an interaction between the ligand and the protein and the signals observed indicate which atoms of the ligand are in close contact with the protein. Like in WaterLOGSY experiments, the STD spectra of EF_3, EF_3-CaM and CaM indicate that TUA-diCl is able to interact with the three proteins (Figure 55). As expected, no significant signal is observed in the control STD spectrum of TUA-diCl obtained at 50 µM, a concentration at which the molecule is essentially monomeric. Moreover, intense signals are observed for the 3 cycles of TUA-diCl for the three proteins and a weak signal for the NH of the N-phenylamino group, indicating that mainly the three rings are involved in the interaction with EF_3, CaM and their complex. A quantitative analysis of the STD data, suggests that the dichloro and thiophen rings establish somewhat stronger contacts with EF_3, CaM and EF_3 -CaM than the phenyl ring (Table 4). The unexpected interaction of TUA-diCl with CaM will be described in Chapter III.

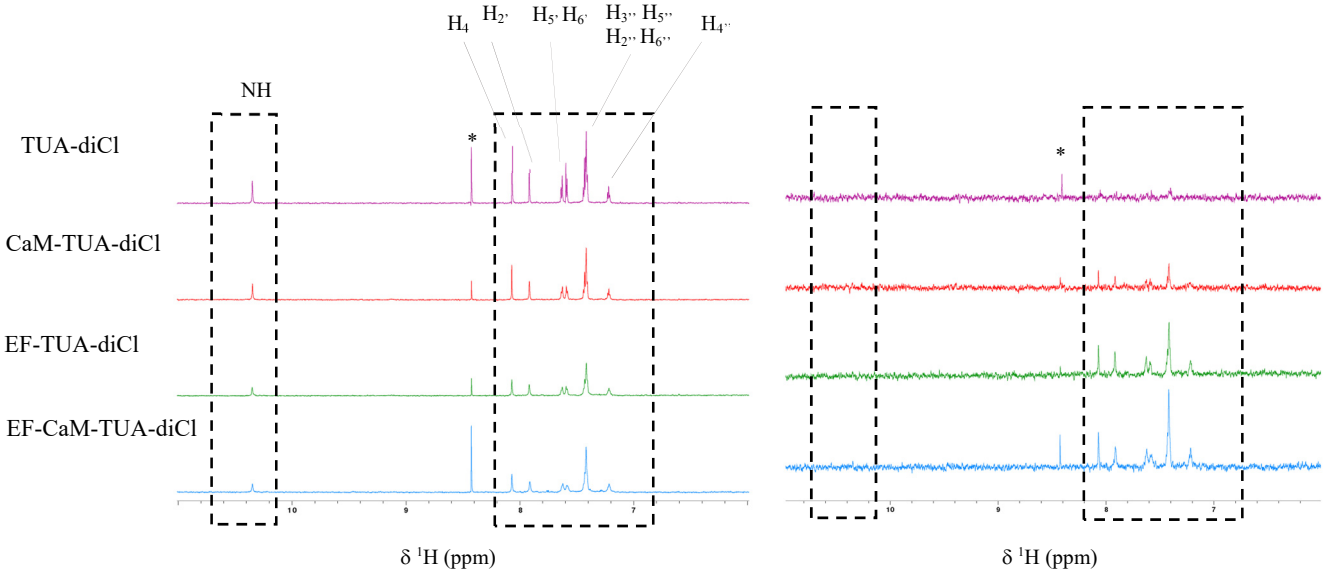


Figure 55- TUA-diCl interaction with CaM, EF_3 and EF_3 -CaM by followed by STD at 25°C. *The corresponding off-resonance control and the corresponding STD ^{1}H NMR spectra are shown on the left and right, respectively. From top to bottom, as indicated by the labels: TUA-diCl (purple), TUA-diCl in the presence of CaM (red), EF_3 (green) and EF_3 -CaM (blue). The concentration of TUA-diCl was 50*

μM. For spectra with proteins a 100:1 ligand:protein ratio was used. The signal at 8.43 ppm labeled with an asterisk) arises from an impurity.

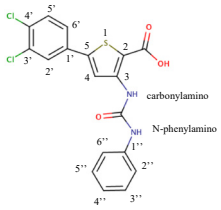
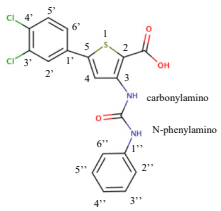
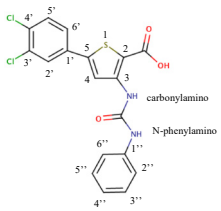
Proteins	*Protons*	*STD_{exp} (normalized)*
EF_3 	NH (phenylamino)	0.13
	H_4	1.00
	$H_{2'}$	0.94
	$H_{5'}$	0.96
	$H_{6'}$	0.77
	$H_{2''}$, $H_{3''}$, $H_{6''}$, $H_{5''}$	0.72
	$H_{4''}$	0.73
EF_3 -CaM 	NH (phenylamino)	0.35
	H_4	1.00
	$H_{2'}$	0.98
	$H_{5'}$	0.99
	$H_{6'}$	0.78
	$H_{2''}$, $H_{3''}$, $H_{6''}$, $H_{5''}$	0.83
	$H_{4''}$	0.95
CaM 	NH (phenylamino)	0.15
	H_4	0.58
	$H_{2'}$	0.79
	$H_{5'}$	0.77
	$H_{6'}$	0.73
	$H_{2''}$, $H_{3''}$, $H_{6''}$, $H_{5''}$	0.55
	$H_{4''}$	1

Table 4- Values of the STD_{exp} ratio of individual ligand protons in STD experiments. *The STD_{exp} ratio is the integral (surface) of a signal in the STD difference spectrum relative to the corresponding integral in the off-resonance reference spectrum. In the table, the STD_{exp} values have been normalized relative to the highest STD_{exp} value.*

It was previously observed that for a monochloride version of TUA-diCl (TUA-Cl) with a weaker inhibitory effect, the order of addition of the compound had an impact on its inhibitory capacity[192]. We thus checked whether adding TUA-diCl to the preincubated EF_3-CaM complex, or to EF_3 and after incubating adding CaM, had an effect on the NMR spectra of the mixture and on the inhibition of the adenylate cyclase activity of EF_3. The order of addition of CaM did neither affect the STD spectrum of the resulting mixture nor significantly

influenced the inhibition of EF_3 enzymatic activity (Figure 56). Indeed after preincubating the complex and then adding the compound, the STD spectra were effectively identical (Figure 56C) and TUA-diCl exhibited an IC_{50} of 12 ± 1 µM (Figure 56A) compared to the value of 15 ± 1 µM obtained when CaM was added to the EF_3 -TUA-diCl complex. We also tested the compound TUA-Cl and while we did not observe differences in the STD spectra, we observed, as previously demonstrated by Laine and coworker[192], that the order of addition of CaM had an effect. Indeed, the inhibitory effect of TUA-Cl was much weaker when added to the preformed complex (53% at 50 µM of TUA-Cl) than when added to EF_3 (6% at 50 µM TUA-Cl). As the EF_3-CaM dissociation rate must be very slow (K_d in the 5 - 20 nM range[110]), these results suggest that TUA-Cl is not able to perturb EF_3 in the CaM-bound active conformation.

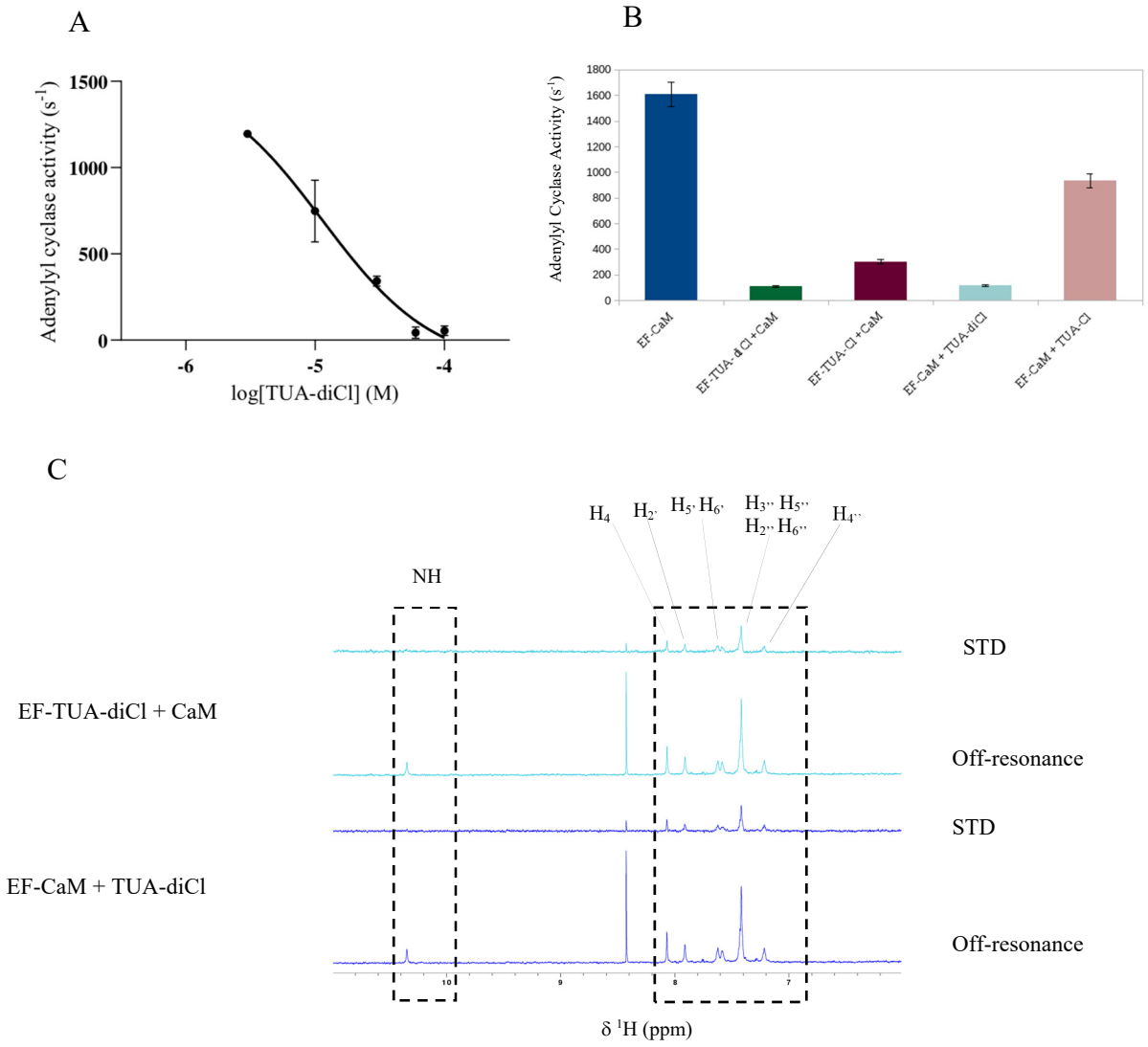


Figure 56- The order of the addition of CaM does not affect the NMR spectra and the inhibition of TUA-diCl. *A. Adenylyl cyclase activity of EF_3 upon addition of TUA-diCl to the pre-incubated EF_3 (0.4 nM) and CaM (1 µM) mixture for 10 min. B. The concentration of TUA-diCl was set to 50 µM. The adenylyl cyclase activity of EF_3-CaM complex is represented in blue. The binary EF_3-CaM or EF_3-TUA-*

diCl mixtures (as indicated) were incubated for 10 min after which TUA-diCl (forest green) or CaM (light blue) was added and the ternary mixtures equilibrated for 10 min prior to recording the adenylyl cyclase activity. The same procedure was applied for the TUA-Cl compound, the EF_3-CaM were pre-incubated and TUA-Cl was added (pink) and the EF_3-TUA-Cl were pre-incubated and CaM was added (red). C. Reference and STD spectra of the ternary mixtures obtained with the addition order indicated in the figure labels. In C, binary mixtures were also incubated for 10 min before adding the third partner.

D.II.3 Binding affinities of TUA-diCl for EF_3 and CaM

We showed that TUA-diCl is fluorescent and used this property to follow its interaction with its different partners and establish its relative affinities to EF_3 and CaM in preliminary experiments. We first defined a wavelength to excite TUA-diCl but not the protein (350 nm) and followed the fluorescence signal at 504 nm to monitor the behavior of the small molecule. The experiments were done at a low and fixed TUA-diCl concentration (10 µM) to avoid auto-association problems. TUA-diCl is hydrophobic and its interactions with EF_3 and CaM are predicted to be mainly hydrophobic. Normally, upon binding into a hydrophobic environment a blue shift accompanied of an enhancement of the fluorescence intensity are observed. However, we observed a quenching of the TUA-diCl fluorescence signal after diluting the compound from concentrated DMSO solutions into the buffer containing EF_3 or CaM (5% final DMSO). The binding isotherms after normalization to 100 are shown in Figure 57. It is clear from the curves that TUA-diCl shows a much higher affinity for EF_3 than for CaM. The apparent K_d valued determined by fitting the curves considering a simple equilibrium with a 1:1 stoichiometry are 0.7 ± 0.4 µM for EF_3 and 93 ± 35 µM for CaM. Although the data for EF_3 at the plateau are noisy, the shape of the EF_3 fit curve displays two straight lines, indicating that (i) the 10 µM TUA-diCl concentration used is much higher than the K_d and that under these conditions, one can determine the stoichiometry of the binding (ii) but the estimates of the dissociation constant are not reliable. As the slope change is observed at 10 µM EF_3 concentration, these data suggest that one molecule of TUA-diCl would bind to one molecule of EF_3. These preliminary data need to be tested further but it indicates an apparent $K_d < 1$ µM for EF_3 and ~ 90 µM for CaM. The EF_3-CaM binding curve is very similar to the EF_3 one (not shown), indicating that under the conditions used, the fluorescence of bound TUA-diCl is dominated by its interaction with EF_3 in agreement with the relative EF_3 and CaM affinities.

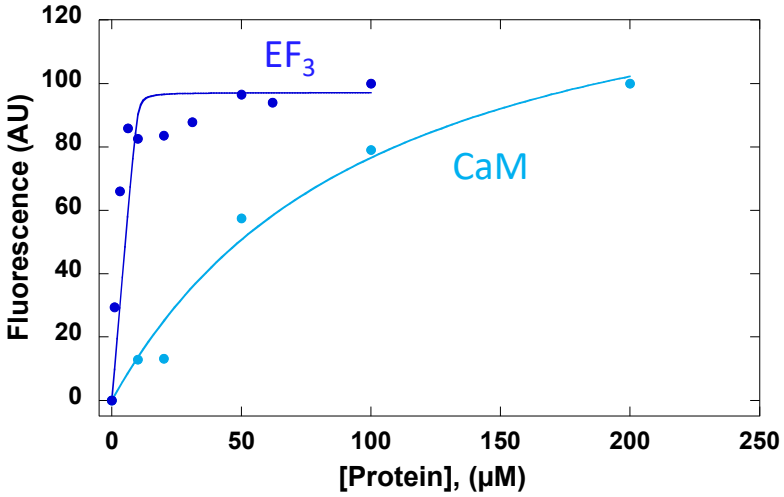


Figure 57- Binding isotherms of TUA-diCl with EF_3 and CaM monitored by fluorescence*. The concentration of TUA-diCl was fixed to 10 µM. The signal is followed at 504 nm after excitation at 350 nm. Binding data was normalized to 100. Curves are the fits of the data to calculate the K_D for a simple two-states equilibrium and a 1:1 stoichiometry.*

D.II.4 Interaction of TUA-diCl with the AC384 and AC384-CaM complex

The interaction of TUA-diCl with CyaA in the presence or in the absence of CaM was also analyzed by STD. The ratio for ligand:protein was the same as used for the interaction of EF_3, namely 100:1. STD spectra indicate that TUA-diCl interacts with AC384 and the complex AC384-CaM and, like in the case of EF_3, the three rings of the compound are involved in the interaction (Figure 58).

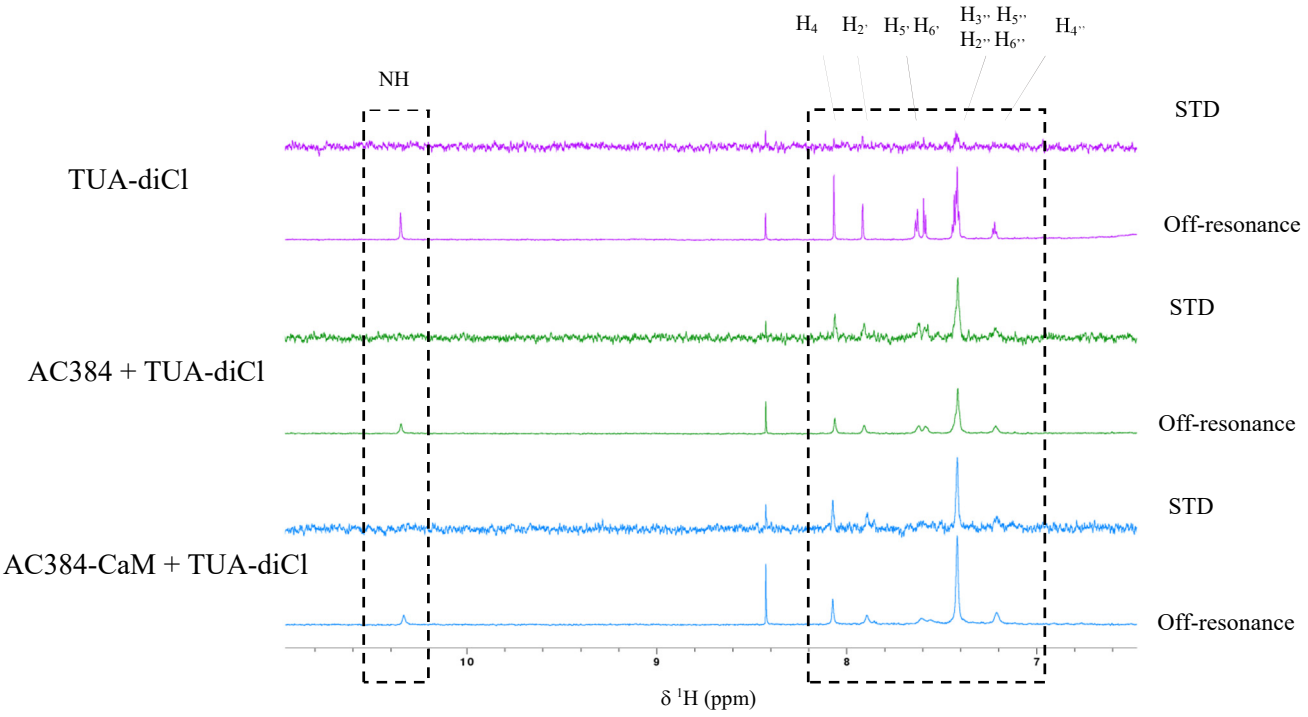
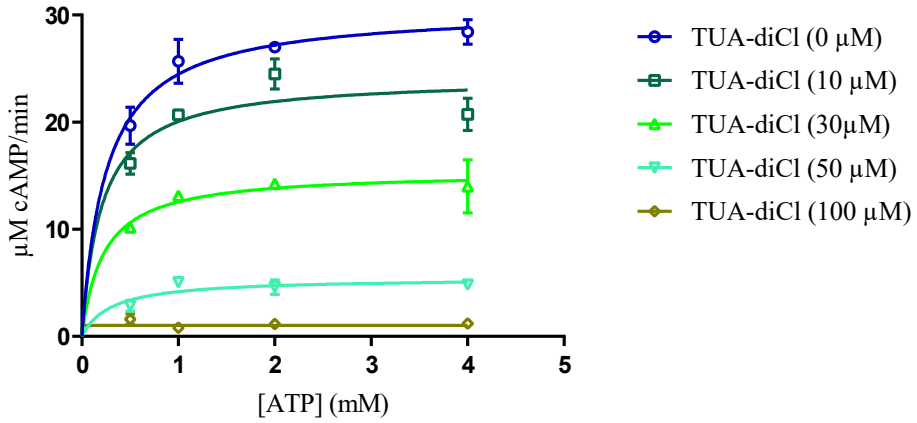


Figure 58- TUA-diCl interaction with AC384 and AC384-CaM assessed by STD. *The ligand:protein ratio used in the experiments is 100:1. The labels indicate the compounds/proteins present in the corresponding off-resonance and STD difference spectra. For these experiments, the concentrations were: AC384 0.5 µM, CaM 1 µM and TUA-diCl 50 µM. The STD spectra in the presence of proteins display signals corresponding to the protons of the three cycles of TUA-diCl and no signal for the NH of the N-phenylamino group, similar to what was observed for the TUA-diCl–EF_3 interaction. The control STD spectrum of TUA-diCl alone does not show any significant signal.*

D.II.5 Is TUA-diCl an allosteric inhibitor?

TUA-diCl was found *in silico* by targeting an allosteric pocket present in EF_3, but we had no experimental evidence of the binding site to support that the compound is an allosteric inhibitor. To test if TUA-diCl is an allosteric inhibitor, we performed competition enzymatic assays, NMR NOESY and STD experiments and characterized the interaction of EF_3-TUA diCl by HDX-MS. Competition enzymatic assays between the natural EF_3 substrate (ATP) and TUA-diCl with four different concentrations of each compound were performed (Figure 59, Table 5). The Lineweaver-Burk representation of the data at 0 and 30 µM TUA-diCl of the data (Figure 59B) clearly shows that the Km is the same under both conditions (same intercept on the abscissa) but the Vmax is different (different intercept on ordinates axis and different slope). Indeed, the Michaelis constant Km of the reaction was ~ 250 µM with or without TUA-diCl but the Vmax is slower in the presence of the inhibitor illustrating that TUA-diCl is a non-competitive inhibitor and supporting the hypothesis that TUA-diCl is an allosteric inhibitor.

A



B

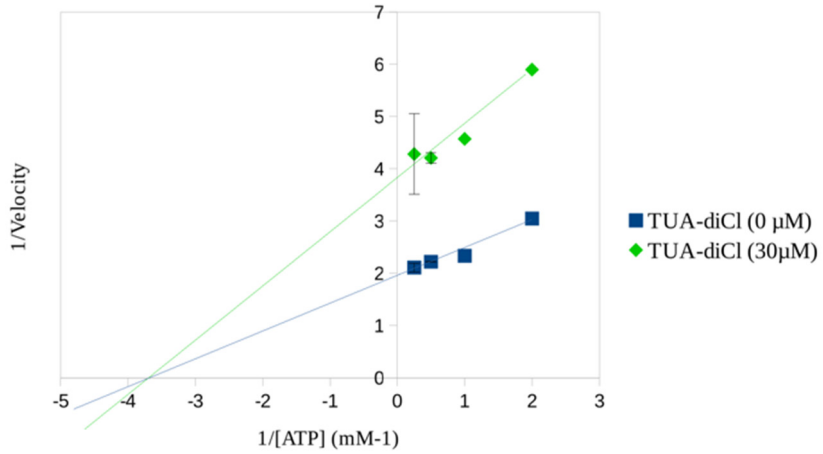


Figure 59- Michaelis Menten kinetics of ATP of EF_3 in the presence of TUA-diCl. *The final protein concentrations are 0.4 nM for EF_3 and 1 µM for CaM; the concentration of ATP varied from 0.5 to 4 mM. TUA-diCl is incubated 10 min with EF_3 and then CaM is added and the ternary mixture is incubated 5 min before starting the reaction by adding ATP. The reaction is stopped after 10 or 15 min for high concentrations of TUA-diCl (50 µM and 100 µM). The Michaelis Menten curves are shown in A and the corresponding Lineweaver–Burk plots for 0 and 30 µM TUA-diCl are displayed in B. The ordinate intercepts (1/Vmax) and the slopes of the curves (Km/Vmax) with the inhibitor (green) and without it (blue) are different, while the intercept on the abscissa axis (1/Km) coincide, suggesting that TUA-diCl is a non-competitive ATP inhibitor that doesn't bind on the active site of EF_3.*

	EF_3-CaM	*EF_3-CaM + TUA-diCl (30 µM)*
k_{cat} (s^{-1})	*1200*	*600*
Vmax (cAMPµM/sec)	*0.518*	*0.261*
Km (µM)	*0.250*	*0.250*

Table 5- Kinetic parameters of enzymatic activity of EF_3-CaM with and without TUA-diCl.

In addition to kinetics experiments, the binding site of TUA-diCl on EF_3-CaM was studied by NMR using STD and transferred NOE (trNOE) experiments. To establish whether

TUA-diCl binds to the catalytic site, we used 2'-MANT-3'-dATP, an ATP analog known to bind in the catalytic site of EF_3 (inhibition constant K_i of ~ 5 µM in the presence of Mg^{2+})[144,264]. We performed STD and trNOE experiments adding Mg^{2+} ions (15 mM $MgCl_2$) to the buffer to ensure that 2'-MANT-3'-dATP binds to the EF_3-CaM complex. While control STD experiments with the isolated compounds show very weak or no signals in the difference spectrum, experiments with the binary mixtures TUA-diCl or 2'-MANT-3'-dATP with EF_3-CaM do display STD signals (Figure 60). Thus, the binding of both compounds can be detected by STD. Under the conditions used (10-fold excess of 2'-MANT-3'-dATP relative to TUA-diCl), the STD spectrum of the ternary mixture (TUA-diCl, 2'-MANT-3'-dATP, EF_3-CaM), shows all the signals of both compounds detected in the corresponding binary mixtures with EF_3-CAM. In other words, both compounds can bind to EF_3-CaM.

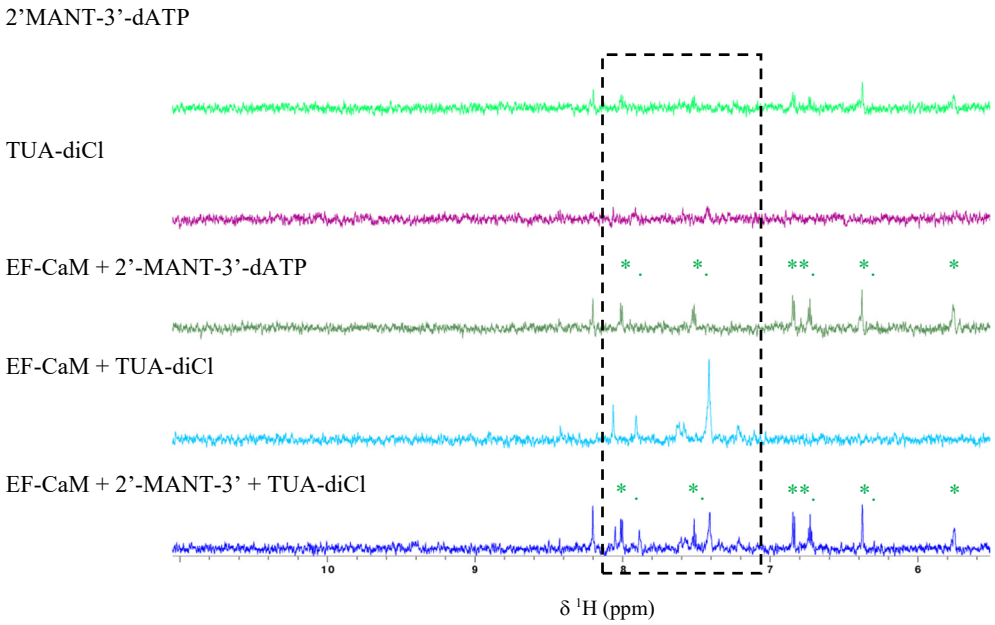


Figure 60- Binding of 2'-MANT-3'-dATP and TUA-diCl to EF_3-CaM assessed by STD. *The labels indicate the compounds/proteins present in the corresponding STD difference spectra. For these experiments, the concentrations (following the order of addition and incubation) were: EF_3 0.5 µM CaM 1 µM, 2'-MANT-3'-dATP 500 µM, TUA-diCl 50 µM. Experiments were performed in the presence of magnesium (15 mM $MgCl_2$) to ensure 2'-MANT-3'-dATP binding. Note that 2'-MANT-3'-dATP was added at 10-fold higher concentration than TUA-diCl. The signals of the TUA-diCl are delimited with a dashed line rectangle and the signals of 2'-MANT-3'-dATP are indicated with in green stars.*

Although STD experiments show that TUA-diCl and the ATP analog can bind to EF_3-CaM when mixed together, these experiments do not provide site specific information. We thus used trNOE experiments to study the binding of both compounds. trNOE experiments are

regular 2D ^{1}H-^{1}H NOESYs performed in a system with a large excess of free ligand that exchanges with a minor form bound to a macromolecule. Under appropriate conditions, these experiments can give distance information of the bound conformation of the ligand read on the signals of the free form. In order to observe transferred NOEs, the free and bound forms of the ligand have to be in fast exchange and significantly exchange during the mixing time (typically ranging from 30 to 800 ms) of the experiment. The NOE intensity and sign depend on tumbling correlation time (τ_c) and hence on the size of the macromolecule, as well as on the spectrometer frequency (ω): small molecules ($\omega\tau_c << 1$) have weak positive or close to zero (positive or negative) NOEs ($\omega\tau_c \approx 1$), whereas macromolecules ($\omega\tau_c >> 1$) show strong negative NOEs. Under appropriate conditions (exchange rate, protein:ligand ratios, the relative strength of bound and free NOEs, binding site occupancy), the NOEs observed on the free form will be dominated by the NOE of the bound form and will be strong and negative. In contrast, the NOE cross-peaks will be positive for the ligand in the absence of the macromolecule.

The NOE spectrum of TUA-diCl in the presence of EF_3 indicates that (i) as expected TUA-diCl binds to EF_3 (negative NOEs with the same sign as the diagonal in the presence of EF_3, positive NOEs in its absence), (ii) the conditions used are appropriate to observe trNOEs (Figure 61). The same NOESY spectrum can be observed for TUA-diCl in the presence of the EF_3-CaM complex. In the control NOESY spectrum of TUA-diCl alone, no negative NOE exists (not shown).

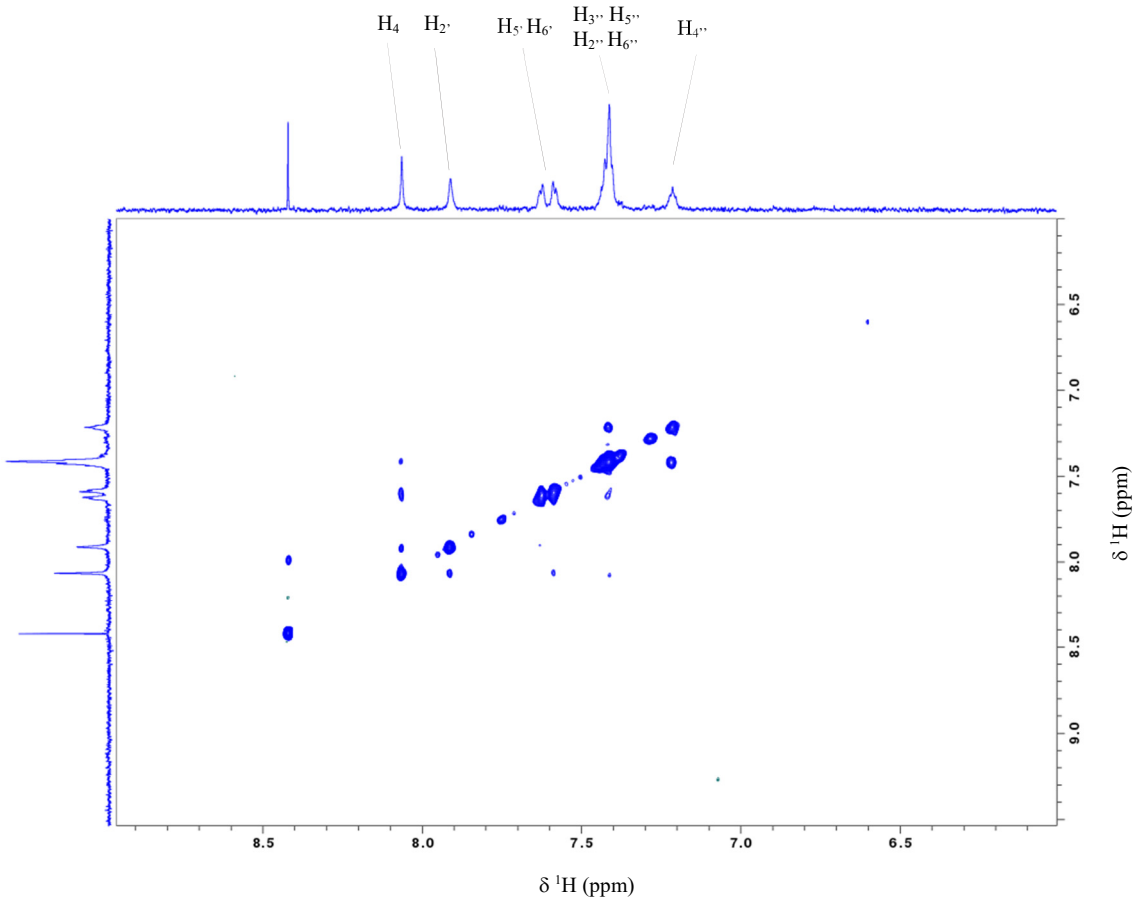


Figure 61- Transferred NOE spectrum of TUA-diCl in the presence of EF_3. *The TUA-diCl NOE cross-peaks are negative (same sign as the diagonal), which shows that TUA-diCl binds to EF_3. The spectrum was acquired with a 600 ms mixing time using 0.5 µM EF_3 and 50 µM TUA-diCl.*

If two ligands bind to the same site of a macromolecule, NOEs between the bound forms of the ligands will be visible in a trNOE spectrum. These NOEs, which do not exist between the free forms of the ligands can build-up via two different mechanisms. In the first mechanism called ILOE (inter-ligand NOE), if the ligands are simultaneously present on the same site (ternary complex) and close to each other (< 5 Å), a NOE can exist between them[265]. The second possibility consists on protein mediated NOEs that are transferred between competitive ligands (do not bind simultaneously) to the protein via the same protons of the protein[266]. Hence, trNOE spectra can be used to investigate whether two ligands bind to the same site.

We first used TUA-Cl with TUA-diCl to check if under the conditions used, it was possible to observe interligand NOEs. The two molecules only differ by Cl/H atom on the TUA-diCl di-chloro ring so they must bind to the same site in a competitive manner. By adding the two TUA compounds to EF_3-CaM or EF_3, we identified a negative intermolecular cross-peak between TUA-diCl and TUA-Cl (not shown) indicating that protein-mediated interligand NOEs can indeed be detected. We then recorded the trNOE spectrum of the ternary mixture 2'-MANT-3'-dATP, TUA-diCl, EF_3-CaM (Figure 62). The trNOE spectrum of the two molecules

in the presence of EF_3-CaM (with 15 mM Mg^{2+} in the buffer) shows intramolecular negative NOEs of TUA-diCl and 2'-MANT-3'-dATP, indicating that both compounds bind to the EF_3-CaM complex. Importantly, no intermolecular NOE exists between both compounds, further supporting that TUA-diCl is an allosteric inhibitor.

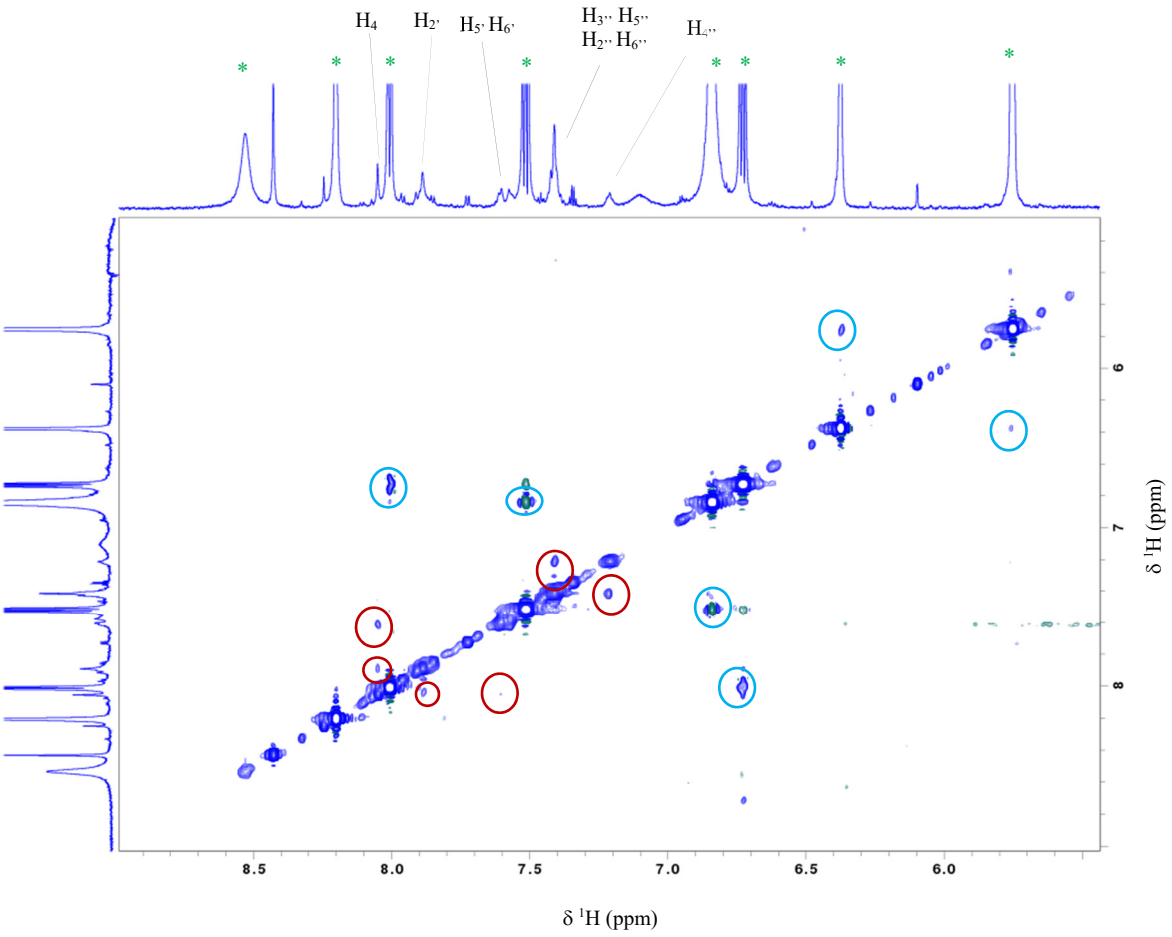


Figure 62- NOESY spectrum of TUA-diCl and 2'-MANT-3'-dATP in the presence of EF_3-CaM. *TUA-diCl (50 µM, grey line with labels) and 2'-MANT-3'-dATP (500 µM, green stars) EF_3-CaM ratios were 100:1and 1000:1, respectively. The spectrum was acquired with a 600 ms mixing time. Negative NOEs are shown in blue and positive NOEs in green. TUA-diCl and 2'-MANT-3'-dATP intramolecular NOEs are circled in red and cyan, respectively. The signals of 2'-MANT-3'-dATP in the 1D projections are trimmed for visualization purposes Both ligands show intramolecular negative NOEs but no intermolecular NOE.*

D.II.6 Structural analysis of the complex TUA-diCl-EF_3 complex by HDX-MS

We next performed HDX-MS experiments in collaboration with S. Brier (Biological NMR technological Core Facilities, Institut Pasteur) to identify the binding site of TUA-diCl on EF.

Hydrogen deuterium exchange followed by mass spectrometry is a powerful technique that provides information on conformation, dynamics, protein folding and binding sites (epitope mappings). It is based on the exchange of amide hydrogens with the solvent hydrogens or deuterons (H_2O, D_2O), the rate of which depends mainly on the temperature, the pH, the sequence of the protein, the burial of amide groups and the formation of hydrogen bonds, as well as on the sequence on the protein. In this technique, the protein is diluted in a solution with D_2O to start the exchange, samples are taken at different times of exchange. The exchange is stopped (by lowering the pH and the temperature) and at the same time an acidic protease is rapidly used to generate peptides (usually pepsin) and then the samples are frozen. The incorporation of deuterium is then readout by mass spectrometry after different times of exchange. Briefly, to map the binding site of a compound on a protein, after establishing suitable conditions, the exchange profiles of the protein with and without the ligand are compared. Upon binding, solvent accessible amides of the protein will be buried and show lower deuterium incorporation rates, leading to the identification of the binding site. Importantly, HDX exchange will not only report on the binding site, but also on conformational changes or differences in the dynamics of the free and bound protein. The resolution of HDX-MS will depend on many factors (sequence coverage by the proteolytic peptides, size of the peptides…) but is usually of the order of six or more residues.

To investigate the effects of TUA-diCl binding on EF_3, the quench and pepsin conditions were first optimized to generate a peptide map with high sequence coverage. A total of 71 unique peptides covering 99.0% of the EF_3 sequence were selected and used for HDX-MS (Figure 63). The uptake profile generated with the isolated EF_3 protein is typical of a well folded protein. Dynamic events (i.e., increase of deuterium uptake with time), indicative of the presence of secondary structural elements, were observed throughout the entire protein including the C_A domain (residues 294-349 and 490-622 encompassing Switch A (502-501) and Switch B (579-591)), the C_B domain (residues 350-489), switch C (residues 623-659), and the C terminal helical domain (residues 660-800) (Figure 26 A). Only few specific peptides of

EF_3, such as peptide 626-641 in Switch C, and peptides 679-688, 764-773, 782-790 and 783-790 in the helical domain display no dynamic HDX behavior and correspond to loop between structural elements.

TUA-diCl binding induces massive changes in solvent accessibility throughout EF. The formation of the complex is associated with large increases in deuterium uptake within the different domains of the protein compared to the isolated protein (Figure 64A). More specifically, the main changes occurred in the catalytic site (segments 325-378 and 477-494, C_A domain) and within regions of the C_A, Switch C, and the helical domain directly involved in CaM binding (regions 502-523 and 615-703) (Figure 64A, C). Interestingly, the CaM binding sites present the C_A (including Switch A and C) and the helical domain lose their dynamic HDX behavior in the presence of TUA-diCl and reach their maximal uptake values after 10 sec labeling only.

The difference in relative fractional uptake between the free- and TUA-diCl-bound states provides a more quantitative and visual assessment of the differences between states (Figure 64B). A negative uptake difference indicates an increase in solvent accessibility, while a positive value is indicative of a TUA-diCl-induced protective effect on the exchangeable amide hydrogens. The greatest uptake increases imposed by TUA-diCl were detected within the catalytic site and the CaM-binding regions. These modifications indicate that TUA-diCl affects the dynamics and/or the structure of two critical regions important for the activity of the toxin. Interestingly, a decrease in solvent accessibility compatible with a masking effect was also detected in the helical domain. The solvent accessibility of peptides 704-723, 704-728 and 704-729 was reduced by ~ 10% in the complex compared to the free state (Figure 64B). Residues 704-729 encompass the C-terminal and the N-terminal regions of helices M and N, respectively, and the loop connecting both helices (Figure 63C). In addition, the identified region contains one residue (N709) involved in the formation of the allosteric SABC pocket that was used to identify TUA-diCl by in silico screening (Figure 64C). It is important to note that the TUA-diCl-induced masking effect was restricted to the region 704-729 containing one residue of the SABC pocket.

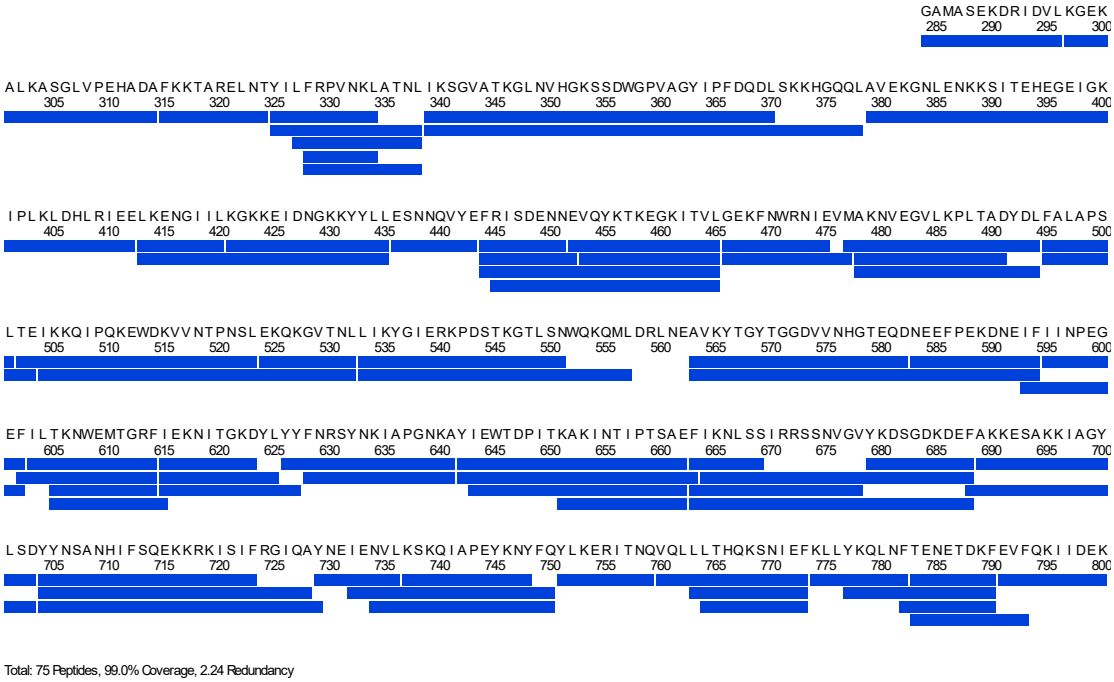


Figure 63- Peptide maps of EF_3 *generated after 2 min pepsin digestion at 20 °C under quench conditions. Each blue bar corresponds to a unique peptide identified by data independent MS/MS acquisition.*

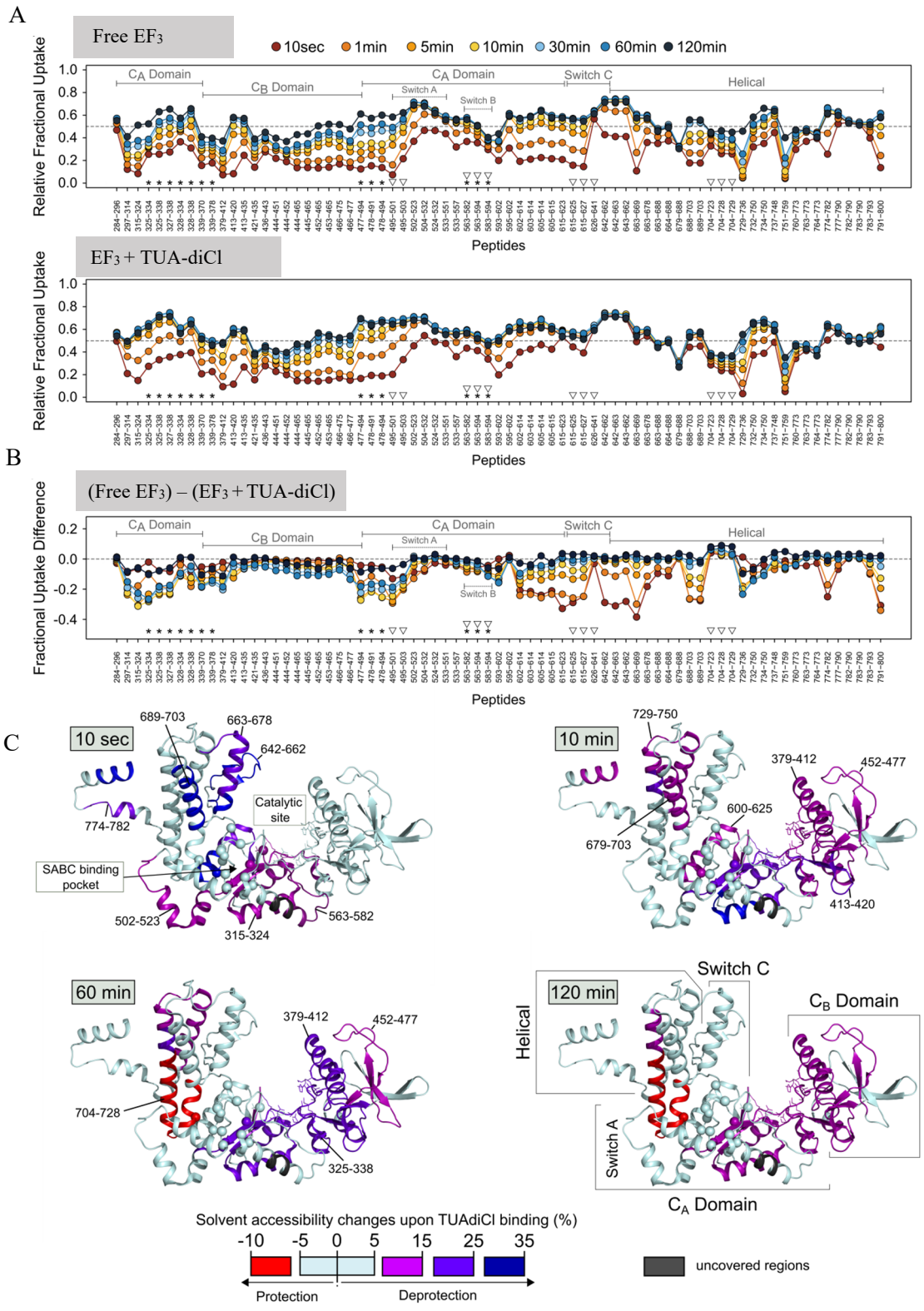


Figure 64- Effects of TUA-diCl binding on the solvent accessibility of EF3. *A. Relative fractional uptake profiles were monitored for EF_3 alone (free) and in the presence of 100X molar excess TUA-diCl. All labeling experiments were performed at room temperature using 85.7% excess D_2O and 2.5% final DMSO. The relative fractional exchange data calculated at each time point and for each condition was plotted as a function of peptide position. Each dot corresponds to the value extracted from one unique replicate (preliminary results). The location of the C_A domain, switch A, C_B domain, switch C,*

and helical domain is indicated. Peptides containing residues involved in the catalytic site or forming the SABC binding pocket are marked with a black asterisk or an empty triangle, respectively. B. Fractional uptake difference plot showing the differences in uptake calculated between the free and the TUA-diCl bound states at each time point and for each peptide. A positive value indicates a TUA-diCl-induced protective effect, while a negative value is indicative of increased solvent accessibility within the complex. C. The fractional uptake differences values calculated after 10 sec, 10 min, 60 min and 120 min labeling were mapped onto the crystal structure of EF_3 in the free state (PDB ID 1K8T). Spheres indicate the SABC pocket. The binding of TUA-diCl dramatically increases the solvent accessibility of elements located in the C_A and the helical domain. The only region showing a decrease in accessibility locates in the helical domain and includes residues N709 from the SABC pocket (peptide 704-728).

D.II.7 Structural investigation on the complex EF_3-CaM interplay with TUA-diCl

Aiming to determine the binding site of the TUA compound. in the EF_3-CaM complex, I sought to use X-ray crystallography approaches. I used protocols and equipment at the crystallography facility of the Institut Pasteur to generate crystals of my protein samples. Several approaches were tested, which included varying the temperature, the concentration of proteins and crystallization kits used, representing 17 544 conditions[211]. Tiny crystals of EF_3 were grown in 0.1 M HEPES, pH 7.5, 1.4 M sodium citrate (EF_3 20 mg/mL). The formation of EF_3 crystal was highly favored by sodium citrate conditions (Figure 65). No crystal was observed for the co-crystallization of EF_3 with TUA-diCl compound. Tiny crystals of the protein with its His-tag, H_6-EF_3, were also grown in the same conditions used for EF_3, suggesting that the presence of the tag doesn't affect the crystallization of EF_3. These crystals were manually reproduced in order to increase their sizes, by varying the crystallization drop volumes and different ratio of protein : mother liquor.

A B

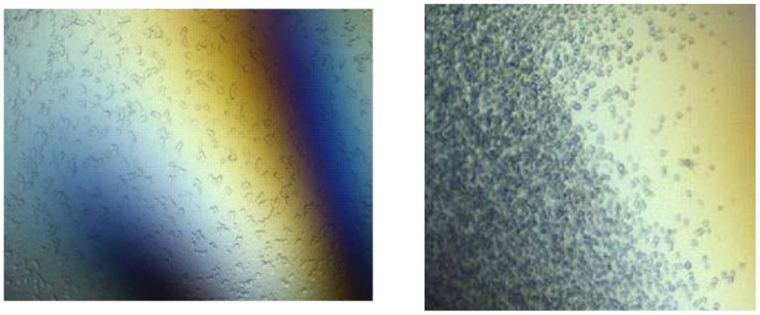
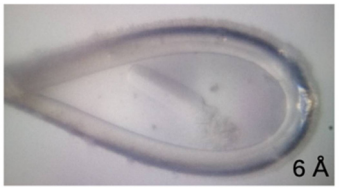
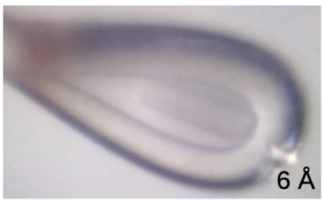
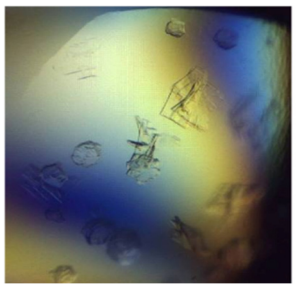
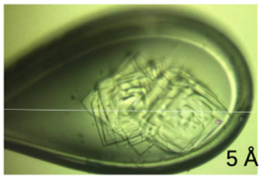


Figure 65- Tiny crystal of the protein EF_3. *The conditions are A. 4 M sodium formate and B. 0.1 M HEPES, pH 7.5, 1.4 M sodium citrate.*

The second strategy was to crystalize the EF_3-CaM complex in the presence or the absence of 3'-dATP, a non-hydrolysable analog of its endogenous substrate and/or TUA-diCl compound. As the inhibition was realized in the presence of CaM and that TUA-diCl interacted with EF_3-CaM and CaM. The complexes EF_3-CaM and H_6 EF_3-CaM were isolated by SEC S-200 as described previously. Different crystals were grown and fished in the conditions displayed in Table 6.

Samples	**Crystalization conditions**	**T (°C)**	**Crystals and diffraction resolution**
EF_3-CaM 21.2 mg/mL, 3'-dATP 3mM, TUA-diCl 1mM **A structure**	0.2 M Ammonium sulfate, 0.1 M Sodium citrate, pH 5.6, 15% (w/v) PEG 4 k	18 °C	 
EF_3-CaM, 21.2 mg/mL	0.2 M Ammonium sulfate, 0.1 M Sodium citrate, pH 5.6, 15% (w/v) PEG 4 k	18 °C	 
H_6- EF_3-CaM, 25.7 mg/mL	0.2 M Ammonium sulfate, 0.1 M Sodium citrate, pH 5.6, 15% (w/v) PEG 4 k	18 °C	 > 6 Å
EF_3-CaM, 15.6 mg/mL	0.2 M Magnesium acetate, 10% (w/v) PEG 8 k	4 °C	 

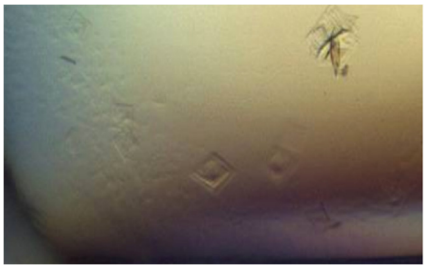
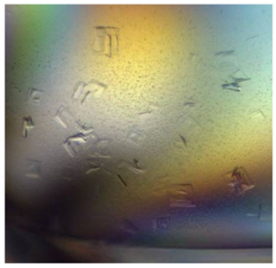
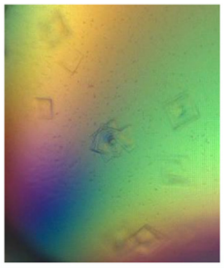
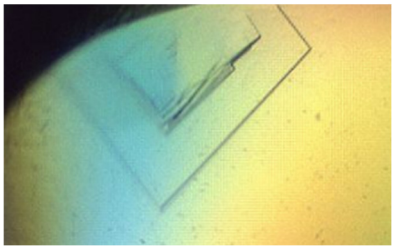
EF_3-CaM, 15.6 mg/mL	0.1 M HEPES, pH 7.5, 0.2 M calcium acetate, 10% (w/v) PEG 8 k, 40% (v/v) GBL	18 °C	 > 6 Å
EF_3-CaM, 15.6 mg/mL	0.1 M HEPES, pH 7.5, calcium acetate 0.2 M, 10% (w/v) PEG 8 k, 0.1 M calcium chloride	18 °C	 > 6 Å
EF_3-CaM, 15.6 mg/mL	0.1 M HEPES, pH 7.5, 0.2 M calcium acetate, 10% (w/v) PEG 8 k, 30% (v/v) DMSO	18 °C	 > 6 Å
EF_3-CaM, 15.6 mg/mL **B structure**	0.1 M HEPES, pH 7.5, 0.2 M calcium acetate, 10% (w/v) PEG 8 k, 0.1 M spermine HCl_4	18 °C	 5 Å

Table 6- Crystallization conditions for the different systems. *For the crystals indicated as A and B, X-ray diffraction data are presented in Table 7.*

In Table 6, we can observe the different shapes and sizes of the crystals. Most of the crystals obtained were stacked together, increasing the difficulty to solve a possible structure. A microfocus (5 µm) beam at the Proxima 2 beamline (Synchrotron SOLEIL) was employed to expose on a spot where the crystal was unique and the thickness sufficient. The maximum X-ray diffraction resolution oscillated between 5 and 20 Å. The complex of His tagged EF_3 associated with CaM was also tested and crystals appeared in the same condition as for EF_3-CaM but no resolution improvement was seen. In order to improve the growing condition of EF_3-CaM crystal and the packing of the crystals, optimization tests were performed. For each

screening crystals grew using 0.2 M ammonium sulfate, 0.1 M sodium citrate, pH 5.6, 15% (w/v) PEG 4 k. Therefore, these components were added and the pH (4.6 to 9.3), the concentration of the precipitant (PEG 4 k, 7.5% to 15% w/v) and the salt (0.1 M to 0.5 M) were varied. The nature of the precipitant and salt were adapted using glycerol, ethylene glycol, PEG 400, calcium acetate, magnesium chloride, ammonium tartrate dibasic. Another option was to design an additive matrix based on a second promising condition (0.1 M HEPES, pH 7.5, 10% (w/v) PEG 8 k, 0.2 M calcium acetate) by using a commercial additive screen from Hampton. We observed that alcoholic solvents helped the crystallization. Other encouraging crystals were found to grow in DMSO conditions. So, adding the small molecule dissolved in DMSO to the complex EF_3-CaM should not interfere with the crystallogenesis of the complex.

After many attempts, recently two structures (named A and B) of the complex EF_3-CaM were solved and refined at 4.27 Å (Table 6A) and 4.75 Å (Table 6B) and by A. Mechaly (Institut Pasteur) (Table 7; Figure 67). In these X-ray crystallographic structures, the complex EF_3-CaM was formed (Figure 66). In the X-ray crystallographic structure B, the complex forms a V-shaped dimer, whereas in the A structure the complex is a trimer as observed in the previous published X-ray structures[28,29,82,110,127]. A notable difference was the space group. The space group of the published structure of the complex is I_{222} [28,29,82,110,127]. Due to the low resolution, the comparison at the atomic resolution between the already obtained and the published structures is difficult (Figure 67). Unfortunately, despite the formation of crystals in the presence of the TUA-diCl and the low resolution, of the diffraction data, no density that might correspond to the ligand was observed. During these three years of my Ph.D. work, none of the many attempts to co-crystallize EF_3-CaM in complex with TUA-diCl was unsuccessful. Efforts will be pursued to obtain high quality diffracting to solve the structure of the EF_3-CaM complex bound to the small molecule by soaking and/or seeding methods.

	A structure EF_3-CaM (21.2 mg/mL) 3'dATP (3mM) TUA-diCl (1mM)	**B structure** EF_3-CaM (15.6 mg/mL)
Beamline	PROXIMA-2	PROXIMA-1
Space Group	P 4_2 2_1 2	P 1 2_1 1
a, b, c (Å)	243.77 243.77 122.85	66.48 138.32 124.30
α, b, γ (º)	90.00 90.00 90.00	90.00 100.63 90.00
Resolution (Å)	49.83 / 4.27 (4.31 / 4.27)	122.17 / 4.75 (5.81 / 4.75)
Completeness (%)	99.6	52.3
R-Value Free	0.277	0.336
R-Value Work	0.230	0.273
CC (work)	0.910	0.774
CC (free)	0.866	0.705
CC (1/2)	0.985	0.992
Rmerge	1.175	0.164
Multiplicity	53.6	4.3
Mean((I)/sd(I))	5.1	5.0
Total number of observations	1400559	25083
Total number unique	26118	5893
Reflections used in refinement	26070	5893
Wilson B-factor	125.27	219.04
Protein atoms	15036	10024
RMS (bonds)	0.008	0.004
RMS (angles)	1.04	0.63

Table 7- Data collection and preliminary refinement statistics for the EF_3-CaM complex.

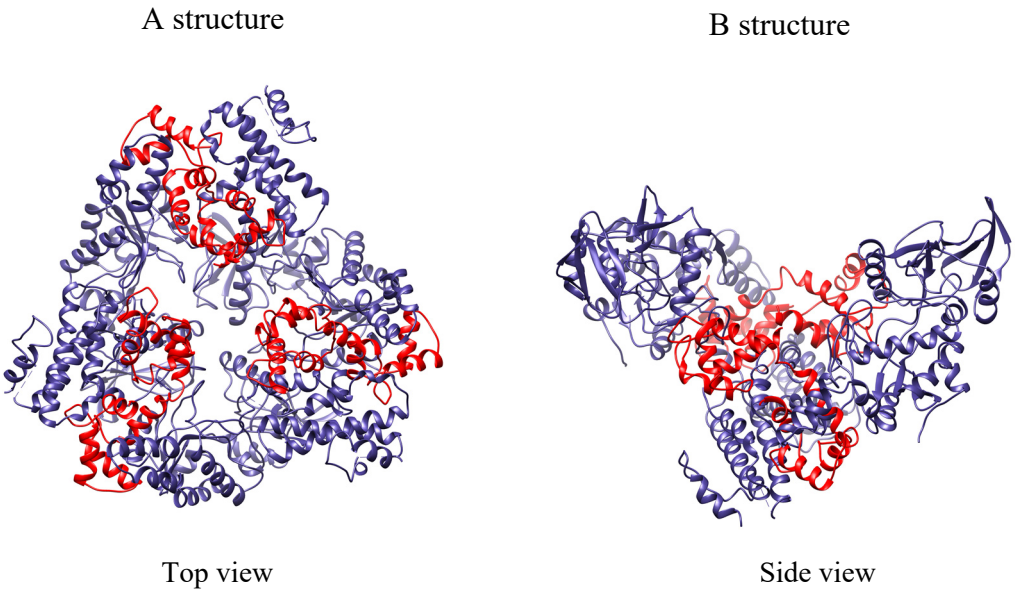


Figure 66- X-ray crystallographic structures of the complex EF_3-CaM. *The protein EF_3 is represented in dark slate blue and the CaM in red. The A structure (left) is a trimer of the complex EF_3-CaM whereas the B structure is a V-shaped dimer (right). Different views of the dimer and trimer and displayed.*

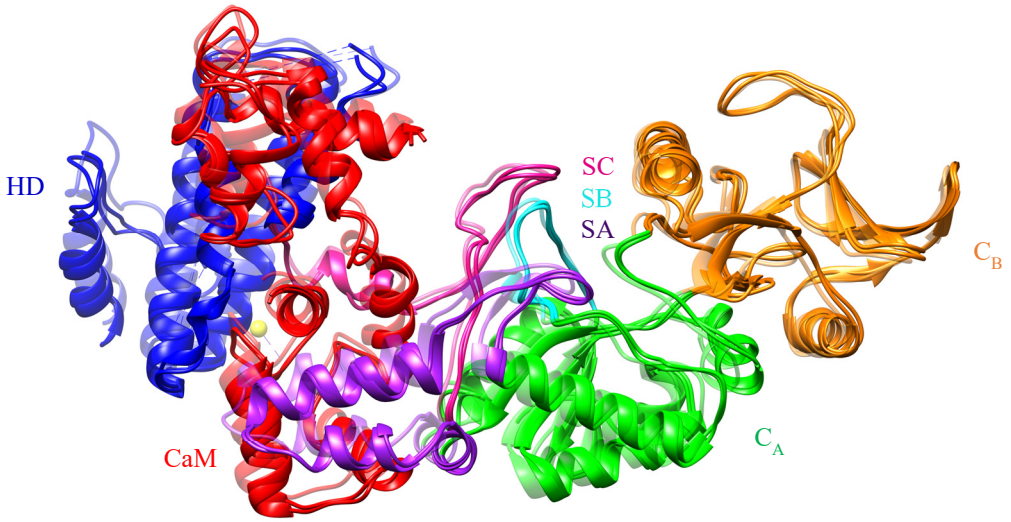


Figure 67- Superimposition of the X-ray structure of the published and the herein obtained EF_3-CaM X-ray structures. *In transparency, the X-ray structure of EF_3-CaM (1K93, colored by domain) is represented. The trimer A and dimer B structures that we obtained superimpose well with the B. anthracis 1K93 structure albeit the different resolutions of the published (2.95 Å) and the A (4.27 Å) and B (4.75 Å) structures.*

D.II.8 Molecular docking of TUA-diCl with EF_3 and the complex EF_3-CaM

Molecular docking can predict binding sites and conformations of a ligand bound to a macromolecule. The docking software optimize the interaction protein-ligand conformation by minimizing a scoring function that evaluates the energy of the complex. Due to the difficulty

of predicting an interaction energy from a three-dimensional structure, the docking algorithms produce an ensemble of ligand conformations bound to the protein. Here, we used docking as a tool to predict a hypothetical ligand-protein three-dimensional structure. Here, instead of only relying on a docking score function to classify the predicted poses, we used the STD experimental data as filters to help in the *in silico* prediction of the binding site of TUA-diCl on EF. We thus, performed molecular docking calculations using the experimental STD_{exp} ratios (Table 4), which provide quantitative information on ligand protons in close contact with the protein, as well as the NOESY data, which excludes the catalytic site as a possible binding site, to filter and validate the predicted binding site(s). By calculating the minimum distance between each ligand hydrogen and the surrounding protein ones, a theoretical epitope mapping is obtained for each pose and compared with the experimental data. Nevertheless, several factors can complicate this approach, such as baseline distortions in the experimental spectra or the approximations used in the docking calculation. We used *mkgridXf*[216] to calculate the cavities and Autodock Vina to dock TUA-diCl on EF_3 and EF_3-CaM. The ligand hydrogens in closer contact with protein hydrogens can be established by a high STD_{exp} ratio. The relative distances to the protein hydrogens were established from the experimental STD_{exp} values (Table 4) after normalization with respect to the highest STD_{exp} value. The theoretical STD (STD_{th}) is calculated by counting the protein hydrogens within a sphere of 6 Å of the considered ligand one divided by the total number of hydrogens of the binding pocket (Figure 68).

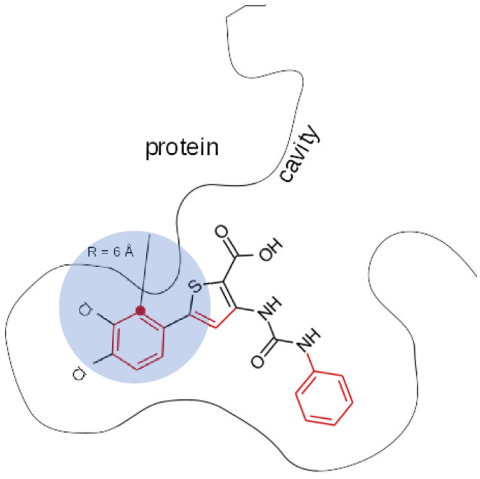


Figure 68- Computation of theoretical STD ratios of the docking poses. *The cavity is schemed as a black line and the molecule is buried within the cavity. The hydrogens close to the protein are shown in red. The blue circle represents an example of how the STD_{th} is calculated for a particular hydrogen proton of the ligand (red point).*

The starting PDB structures used for the docking calculations were 1K8T for EF and 1XFY for EF_3-CaM (without the PABD domain). The missing regions were reconstructed using the CHARMM-GUI[51] interface. Twenty-five cavities were found in EF_3 and 42 cavities in the complex EF_3-CaM using *mkgridXf*[216]. The ligand TUA-diCl was docked into twenty poses within each of these cavities. Once the cavities and the poses were defined, we used the experimental data to filter the docking poses. The best poses were obtained by minimizing the root mean square error (RMSE) between the theoretical and experimental STD values. The three best poses/cavities (lowest RMSE) for each protein are presented in Table 8 and in Figure 69. In the best poses with EF_3, TUA-diCl was located in the insertion region of CaM in EF_3, whereas with the EF_3-CaM complex, the compound interacted with both EF_3 and CaM. The helical domain is involved in the best poses for EF_3 and its complex, in agreement with experimental data, namely (i) the poses fit with STD data showing that the three cycles of the molecule TUA-diCl are in interaction with EF_3, (ii) TUA-diCl does not bind to the catalytic site as evidenced by enzymatic competition and trNOE experiments and (iii) the only region that becomes protected from the solvent upon binding of TUA-diCl to EF_3 (residues 704-728) is located in the helical domain and matches with pose ID 3 (residues A708, Q714, K717 and R718). However, the presence of these charged residues invites further studies to fully understand the interaction of the ligand with its pocket. In chapter IV, we combine the experimental data on the two systems with the docking data and molecular dynamic (MD) simulations, to better explore the mechanism of TUA-diCl binding and its binding site. Interestingly the helical domain plays an important role in the activation of EF and seems to be the target of TUA-diCl, suggesting that it is a promising target to develop allosteric inhibitors of the EF_3 toxin activation.

Systems	ID	Cavity number	RMSE	Energy (kcal/mol)	Volume ($Å^3$)	Position
EF_3	1	4	0.18	- 7.40	2080.58	SABC loop
	2	20	0.19	- 8.20	245.02	Helical domain
	3	18	0.19	- 7.90	47.92	Helical domain
EF_3-CaM	4	17	0.15	- 7.50	3539.58	Helical domain
	5	36	0.16	- 7.30	146.47	Helical domain
	6	14	0.16	- 6.40	96.97	Helical domain

Table 8- RMSE between the calculated and experimental STD ratios.

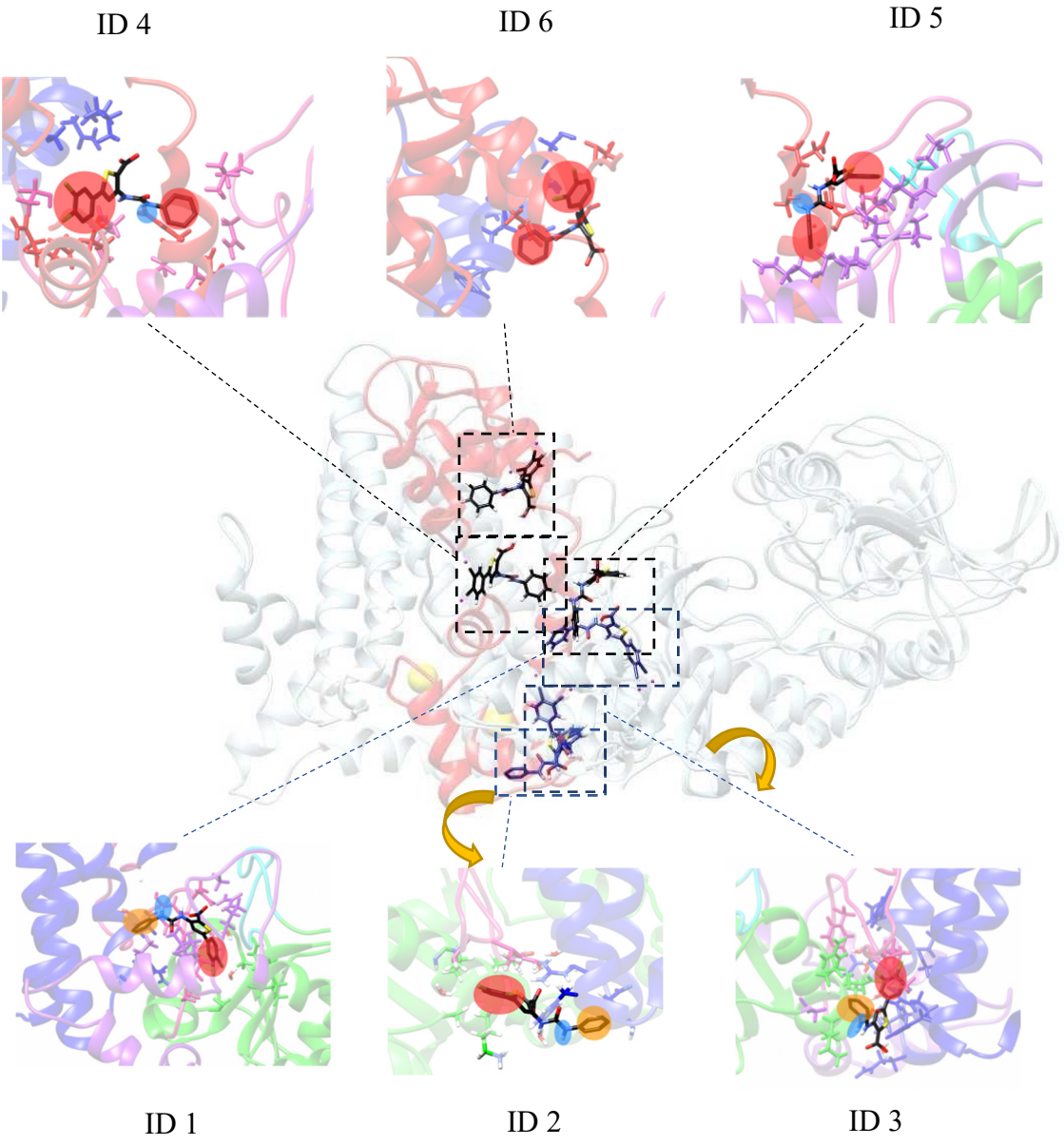


Figure 69- Cartoon representation of the three lowest RMSE poses and cavities for EF_3 and EF_3-CaM. *The structure of the EF_3-CaM complex is displayed as a transparent ribbon, with EF_3 in white, CaM in red, calcium atoms in yellow together with the three best poses of TUA-diCl on EF_3 (blue) and EF_3-CaM (black). IDs 1 to 3 (bottom) are the best poses for EF_3 and IDs 4 to 6 for EF_3-CaM. The hydrogens of TUA-diCl at close, medium or relatively long distances with the respect to the protein as assessed by STD_{exp} values (Table 4) are highlighted with red, orange and blue ellipses, respectively. For ID 2 and 3 the protein was flipped vertically to facilitate visualization. The binding regions are indicated in Table 8.*

DISCUSSION

The ligand TUA-diCl, discovered *in silico* by targeting an allosteric site on the edema factor toxin[192], has been extensively studied by various structural biology and biochemical approaches. The interactions of TUA-diCl with isolated EF_3 as well as with the EF_3-CaM complex have indeed been investigated by a large variety of NMR approaches such as STD, WaterLOGSY and trNOE, as well as by HDX-MS, crystallogenesis and crystallography, fluorescence and enzymatic assays. The effect of the compound on *Bordetella pertussis* adenylate (domain AC3884) toxin activity has also been characterized.

The *in vitro* adenylyl cyclase inhibition assays with the *B. cereus* and *Bordetella pertussis* adenylate toxins showed that TUA-diCl displays IC_{50} values of 15 ± 1 µM for the EF_3-CaM and 36 ± 1 µM for the AC384-CaM complexes. AC and EF share similar enzymatic activities and global structural organization and are activated by calcium loaded CaM but display different features in their activation. One could wonder whether the interaction between the ligand and the adenylate cyclases takes place in an indirect or direct way. In addition, we unexpectedly found by STD, WaterLOGSY and 1D ^1H NMR (not shown), that TUA-diCl directly binds to CaM, probably due to the ligand hydrophobic properties. However, several experimental evidences support that the inhibition is due to a direct interaction with EF_3. Indeed, we showed (i) by STD and WaterLOGSY experiments that the compound binds to EF_3, (ii) by fluorescence that TUA-diCl has a much higher affinity for EF_3 than for CaM and (iii) enzymatic assays indicated that varying the CaM concentration did not affect the IC_{50} constant. These results suggest that the inhibition mechanism is CaM-independent. The interaction of TUA-diCl with CaM will be described in more detail in Chapter III.

Hydrophobic compounds like TUA-diCl that tend to form aggregates can inhibit non-specifically different proteins by causing their aggregation. To avoid this problem, we worked at low TUA-diCl concentrations in which we previously showed that the compound did not aggregate (see Chapter I). Moreover, the IC_{50} and the apparent K_d of the interaction with EF_3 (< 1 µM) determined by fluorescence corresponds to concentrations in which TUA-diCl molecules do not stack together, and preliminary fluorescence results indicate that one molecule of TUA-diCl binds to one molecule of EF_3, strongly arguing for the specificity of the interaction with the EF and CyaA virulence factor adenylate cyclase domains.

We followed several research axes to determine if TUA-diCl was indeed an allosteric inhibitor and to characterize its binding site on EF_3. STD and trNOE competition experiments

with TUA-diCl and the ATP analog 2'-MANT-3'-dATP that binds to the catalytic site, indicated that both molecules were able to bind to the EF_3-CaM complex. In addition, no trNOE interaction cross-peak was observed between both molecules in the trNOE spectra, suggesting that both molecules bind to different sites on the EF_3-CaM complex. Competition enzymatic assays with different ATP and TUA-diCl concentrations further indicated that TUA-diCl does not influence the Michaelis-Menten constant K_m of the ATP conversion into cAMP, further demonstrating that TUA-diCl is a non-competitive inhibitor of ATP and hence binds outside the catalytic site.

As revealed by HDX-MS experiments, TUA-diCl binding has a dramatic effect on the solvent accessibility of several regions of EF_3, denoting a general destabilization of the protein. The catalytic site, the switch B, and C regions and the helical domain that interacts with calmodulin are the most affected regions. Conversely, the only region that becomes protected upon TUA-diCl binding and could putatively be attributed to the TUA-diCl binding site (residues 704-728), is a helix-loop-helix motif between helices M and N in the helical domain and includes one residue of the SABC pocket (N709) that was the target cavity used to identify TUA-diCl *in silico*. The HDX-MS observations agree well with NMR and enzymatic activity competition data to establish that TUA-diCl is an allosteric inhibitor and places the presumed binding region close but distinct to the SABC pocket. Interestingly, the 704-728 region matches with one of the best cavities identified by our docking procedure (ID 3) using experimental STD ratios of the compound, and this region is close to the region involved in CaM binding. It should be mentioned that a previously described inhibitor of EF_3, the ligand 10506-2A, displays a similar IC_{50} and shares common chemical groups with TUA-diCl, such as three aryl rings with a thiophen linking a phenyl and an amine. Interestingly, cross-linking studies showed that the compound binds to the EF switch C and helical regions (residues 631-690 and 740-777)[180] but the putative binding site in the former study and this work differs.

On the ligand side, STD experiments revealed that the three cycles of TUA-diCl interact with EF_3 and EF_3-CaM, whereas the linker between the thiophen and phenylamine cycles seems to be less important. Furthermore, the di-chloride ring seems to play an important role as a substitution of a chlorine by a hydrogen (TUA-Cl) leads to a weaker affinity for EF_3 (STD data not shown). Of note, for TUA-Cl, we confirmed here the previous observation[192] that adding CaM after or before the ligand has an effect on EF inhibition: TUA-Cl is unable to inhibit the preformed complex EF-CaM, whereas TUA-diCl inhibits it.

In parallel to the characterization described above, we made many attempts to crystallize EF_3 and EF_3-CaM with TUA-diCl to obtain the structure of the ligand and to reveal the

conformational changes undergone upon ligand binding. We solved two X-ray structures, both corresponding to the active state of EF_3, which along with activity assays indicated that the protein samples prepared during this work were in the native state. Unfortunately, no extra density that could correspond to the TUA-diCl inhibitor was detected in the structure. The difficulties encountered for crystallization can be *a posteriori* explained by the drastic effect of TUA-diCl observed by HDX-MS and the self-aggregation of the compound at high concentrations. Indeed, HDX-MS revealed that EF_3 is destabilized upon TUA-diCl binding, and we showed that at the high concentrations of TUA-diCl required for crystallization (up to 1 mM) the molecule self-aggregates. To my knowledge, no crystal structure of a complex with an allosteric inhibitor has been reported so far for any pathogenic adenylyl cyclase.

More generally, different mechanisms of inhibition by TUA-diCl can be considered: (i) TUA-diCl could indirectly inhibit the activity of EF by binding to CaM; (ii) TUA-diCl could prevent CaM binding by locking EF in an inactive conformation; (iii) TUA-diCl might bind to the EF-CaM complex and allosterically disturb the catalytic site or compete with CaM for EF binding; (iv) TUA-diCl could dissociate the pre-formed EF-CaM complex (Figure 70).

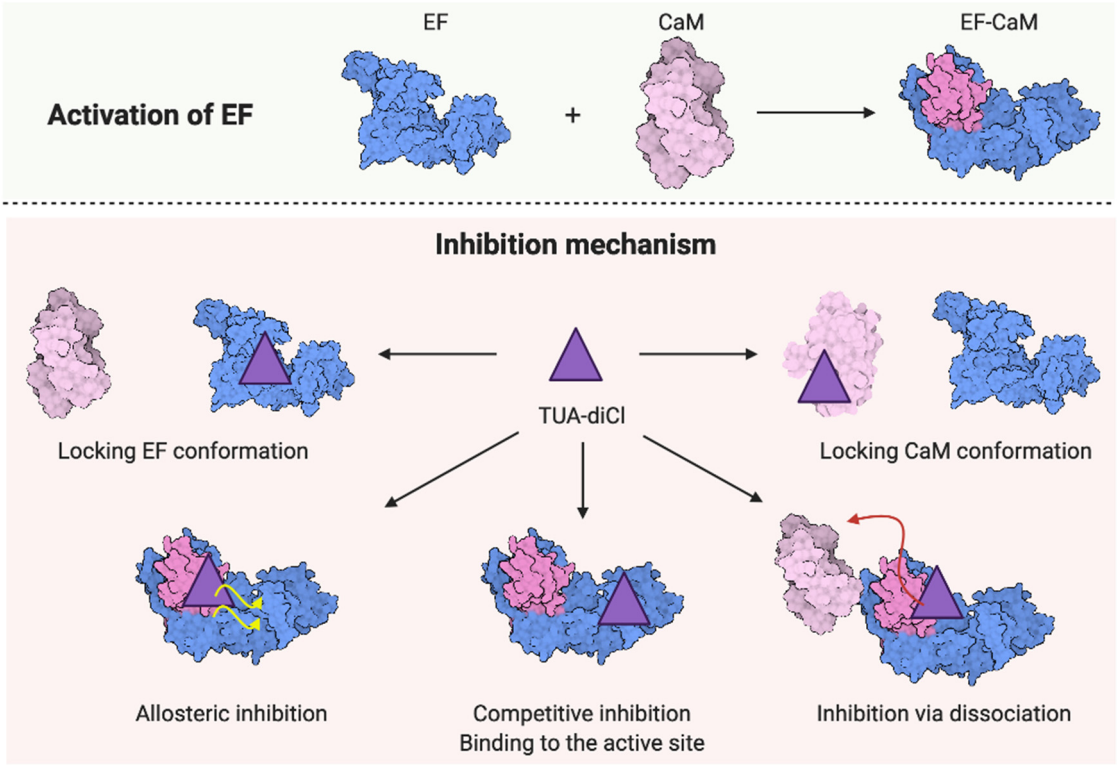


Figure 70- Possible inhibition mechanisms of EF_3-CaM by the TUA-diCl. *(top) Activation of EF_3 by CaM binding and (bottom) five possible inhibition scenarios. CaM is represented in pink, EF in blue, and the compound is shown as a purple triangle.*

Our experimental data allows us to distinguish between these scenarios and clearly shows that TUA-diCl acts by binding to an allosteric site that could lock EF in an inactive conformation either free or in complex with CaM. The fact that CaM shows a higher affinity for EF-CaM (K_d in the tenths of nanomolar range) than TUA-diCl (K_d < 1 µM, IC_{50} = 13 ± 2 µM) and that the order of addition of CaM to EF (EF_3-CaM + TUA-diCl or EF_3-TUA-diCl + CaM) does not affect its inhibition activity, suggests that TUA-diCl can bind to the EF-CaM complex and does not directly compete with CaM. In agreement with HDX-MS data, the compound could then destabilize EF, thereby disrupting the catalytic site, and possibly causing the dissociation of CaM.

The experimental data and *in silico* docking information allow us to highlight plausible binding sites in the helical domain of EF and particularly the region 704-728 which is the only region that becomes less accessible to the solvent upon TUA-diCl binding. In agreement with the virtual screening approach previously made in the laboratory to identify TUA-diCl, the compound is an allosteric EF_3 inhibitor. However, the binding site of the molecule seems to be different from the one previously imagined. Further structural studies need to be performed to characterize the binding site and from there, the interactions with the target site to improve the properties of the molecule.

Chapter III. Interaction of TUA-diCl with CaM

INTRODUCTION

Calmodulin is a calcium-binding protein essential for the proper functioning of the eukaryotic cells. This ubiquitous protein interacts with numerous target peptides and adopts different conformations depending on its interaction with the target and on the concentration of Ca^{2+}. CaM is composed of two globular regions, the C-terminal and the N-terminal lobes, hereafter named in the following N-CaM and C-CaM, which are connected by a flexible linker[95,96,98]. The various conformations sampled by the central linker from an extended α–helix[106,107,108] to an unwound conformation[267] confers a high conformational plasticity to CaM. In each lobe, the loading of Ca^{2+} ions induces an exposition to the solvent of a series of residues forming a so-called hydrophobic patch. The N-CaM patch is composed of the residues A10, F12, A15, L18, F19, L32, M36, L39, M51, V55, M71, M72, M76, and the C-CaM patch of residues I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, A147[95,102–104]. In the presence of its target peptides or small compounds, CaM most often wraps the peptide or the small inhibitors using the exposed hydrophobic patches to establish the interaction, thus adopting a compact globular conformation. Compounds like calmidazolium chloride (CMZ), Tripfuoperazine (TFP) and W7 are well-established CaM inhibitors[189,]. The X-ray crystallography structure of CaM with TFP revealed that one, two or four molecules of TFP bind depending on the stoichiometry used[268,269,185,270]. In the case of a 1:1 CaM-TFP complex, the TFP is located in the C-CaM lobe[268], whereas for the 1:4 CaM-TFP, two molecules bind to the N-CaM (TFP-2) and C-CaM (TFP-1) and two others (TFP-3 and TFP-4) are located at the interface of the two lobes[185]. NMR studies of the interaction between CaM and W7 revealed that two molecules (W7-1 and W7-2) are trapped in the two hydrophobic patches of C-CaM and N-CaM without interacting which each other[271]. At the contrary, for the structure of CaM interaction with two N-(3,3-diphenylpropyl)-N′-[1-R-(3,4-bis-butoxyphenyl)ethyl]-propylene-diamine (AAA) molecules , the AAA molecules are located in the N-CaM and the C-CaM and interact with each other, extending the hydrophobic core-like region within CaM[272]. Another feature showing the diversity of the CaM interaction and its plasticity is that CaM is able to bind simultaneously two chemically different ligands such as TFP and KAR-2 or vinblastine and KAR-2[273]. A recent X-ray crystallography of CaM bound to KN93 showed a configuration similar to those previously observed for the four TFP ligands.

In this structure, two KN93 molecules bind to the two CaM lobes (KN93-1 to N-CaM and KN93-2 to C-CaM) with a third molecule (KN93-3) bound to the interdomain via a cooperative mechanism[183]. The ligands TFP, W7, AAA and KN93 would lock the hydrophobic patches of CaM preventing them to recognize and interact with their natural target[185,272,274]. The mechanism of action of these molecules would be to compete for CaM, thus preventing its interaction with its target. However, it is common that small hydrophobic molecules bind to CaM in an unwanted way thus provoking side effects[185,188,274,275,276]. As CaM is involved in numerous cell processes, targeting CaM offers new therapeutic opportunities for different cancers, inflammatory diseases and neurodegenerative diseases[274,276,277,278,279].

Here, we characterize the interaction between the ligand TUA-diCl with CaM combining NMR with an *in silico* approach. This combination of experimental and *in silico* data provides some structural and mechanism insight into the CaM-TUA-diCl complex.

RESULTS

As described in Chapter II, we observed by STD and WaterLOGSY NMR experiments that TUA-diCl is able to interact with CaM. In order to determine the interaction site of TUA-diCl on calmodulin, the affinity constant of the interaction, the effect of TUA-diCl on the dynamics of CaM and to establish if the interaction is calcium-dependent, we worked with $^{15}N/^{13}C$ doubly labeled CaM. We first assigned the ^{1}H, ^{13}C and ^{15}N backbone and CB chemical shifts of calcium loaded CaM in our experimental conditions (pH 7.0, 37 °C) using conventional methods described in the Materials and Methods section C.4.

D.III.1 The interaction of TUA-diCl with CaM is calcium-dependent

Backbone amide ^{1}H and ^{15}N chemical shifts are very sensitive to the chemical environment and constitute excellent probes to map binding sites, detect conformational modifications and changes in dynamics upon binding for each residue of a protein. To evaluate the effect of calcium in the interaction of TUA-diCl with CaM, we compared the ^{1}H-^{15}N correlation spectra (SOFAST-HMQC) of holo-CaM (2 mM $CaCl_2$ in the buffer) and apo-CaM (buffer without calcium supplemented with 3 mM EGTA) (Figure 71). Upon addition of TUA-diCl to holo-CaM, most of the ^{1}H-^{15}N peaks shift, indicating that, as observed by STD, WaterLOGSY (Chapter II) and ^{1}H 1D experiments (not shown), the compound binds to CaM,

probably changing the conformation/dynamics of the protein. In contrast, TUA-diCl does not show any significant effect on the apo-CaM ^{1}H-^{15}N correlation spectra, demonstrating that in the absence of calcium the interaction is abolished. Hence, calcium is required for the interaction, which suggests that TUA-diCl binds to the exposed hydrophobic patches that are formed when the N- and C-CaM lobes are calcium loaded. TUA-diCl was initially reported not to affect CaM binding to the CyaA-derived H-helix peptide (peptide H) at a concentration able to inhibit EF_3-CaM[192]. Based on this observation from fluorescence anisotropy experiments, it was anticipated that EF_3 inhibition by TUA-diCl was specific as this compound did not affect CaM under the conditions used[192]. However, here we showed using NMR that TUA-diCl is able to interact with CaM and most likely changes its conformation. The lack of effect of CaM could be attributed to the high-affinity of peptide H for CaM (K_d ~ 10 nM for C-CaM and ~ 300 nM for N-CaM, Alexandre Chenal, Institut Pasteur, personal communication) compared to the affinity of TUA-diCl and the concentrations used (1µM CaM, 1 µM peptide H, 10 µM TUA-diCl).

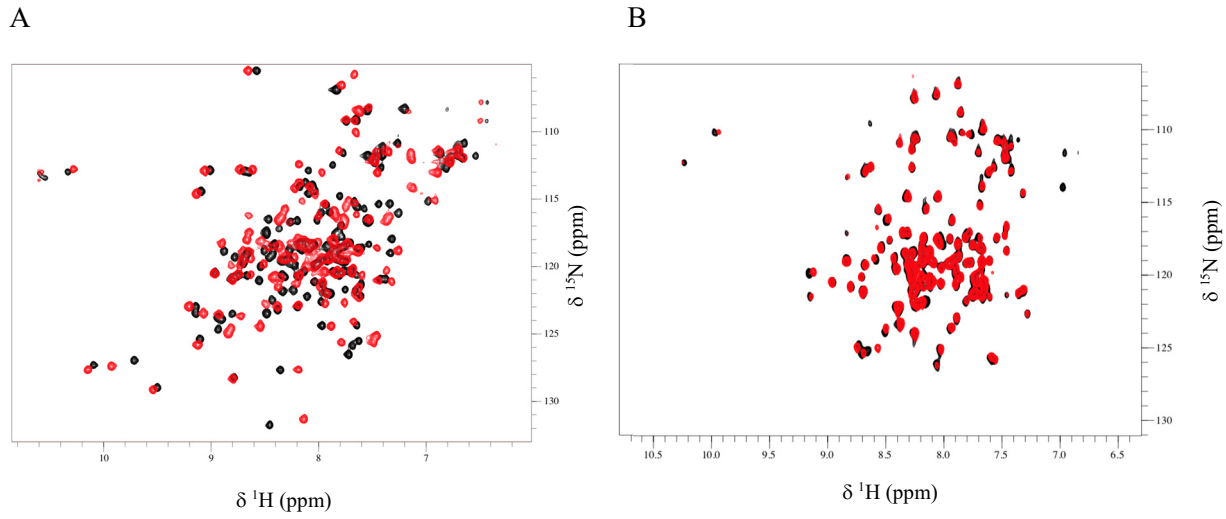


Figure 71- Calcium-dependent interaction between TUA-diCl and CaM. *^{1}H-^{15}N SOFAST spectra of holo-CaM (A) and apo-CaM (B) (60 µM) in the presence (red, 200 µM) or absence (black) of TUA-diCl (200 µM). Spectra were recorded at 37°C. For the holo-CaM samples, the buffer was supplemented with 2 mM $CaCl_2$ and for the apo-CaM samples with 3 mM sodium-EGTA.*

D.III.2 Binding site of TUA-diCl

The binding site was investigated using NMR and HDX-MS. To map the binding site by NMR, we performed chemical shift perturbation (CSP) experiments of CaM with TUA-diCl. TUA-diCl was titrated from 0 to a final concentration of 250 µM and a SOFAST-HMQC experiment was acquired for each TUA-diCl concentration. Upon addition of the ligand, changes of the amide ^{1}H and ^{15}N chemical shifts and/or changes in intensity were observed for many residues indicating that their local environment is perturbed by the ligand (Figure 72). We observed two different exchange regimes for the signals that were perturbed: for some amide signals like I130 (Figure 72B), a gradual variation of chemical shift without significant changes in intensity indicated a fast exchange regime, while for others like F19, V55, A57 and T70, the signals broadened, eventually disappeared and appeared back again at high TUA-diCl concentrations (see A57 signal Figure 72B). This later behavior was indicative of an intermediate to fast exchange regime on the chemical shift time scale. To completely assign the resonances of CaM at the highest TUA-diCl concentration used (245 µM), we followed the signals variations upon titration and confirmed the assignment of the backbone and CB resonances using triple-resonance 3D experiments. The most affected residues were: F19, D20, I27, N53, V55, A57, I63, T70, M72, A73, R74, M76, Y77, R106, M109, N111, G113, K115, V121, I130, V142, Q143, M145, T146, A147 and K148. Of those, S38, M71 and V108 were exchange-broadened and the peaks at the highest TUA-diCl concentrations could not be assigned. The most affected residues are in the N- and C- lobes, as well as in the linker region (M76) of CaM, but most of the highly impacted residues are in the N-CaM lobe (Figure 72C).

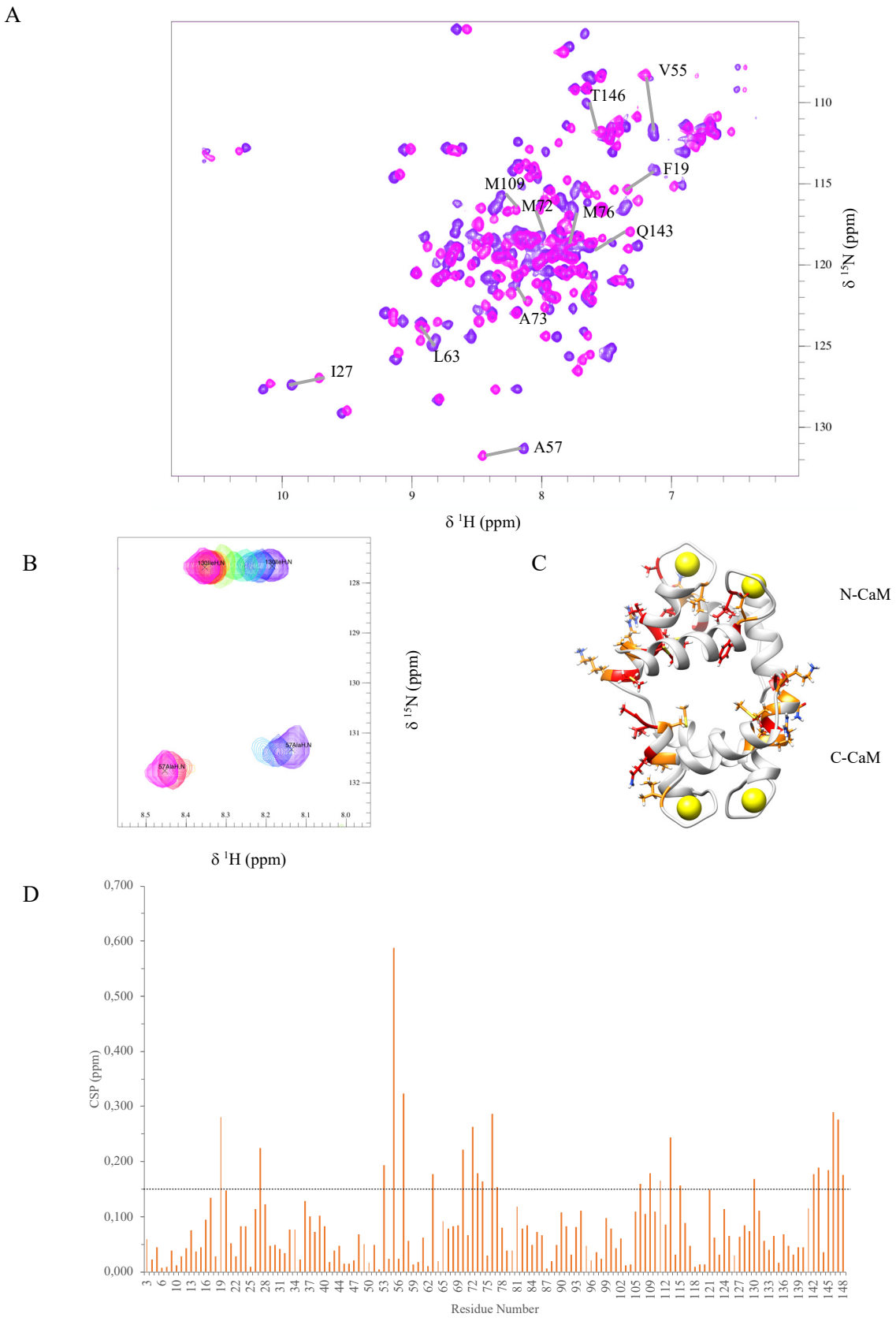


Figure 72- Interaction between TUA-diCl and holo-CaM. *A.* ^{1}H *-* ^{15}N *HSQC spectra of the first and last points of the titration of holo-CaM (~ 100 µM) at 0 (purple) and 245 µM (pink) TUA-diCl concentration. B. Examples of the chemical shift and intensity variations for a signal in fast (I130) and intermediate/fast (A57) regime. C. Residues with highly perturbed amide resonances upon TUA-diCl binding are*

highlighted on the structure of holo-CaM (PDB ID 6M7H) in red (CSP > 0.20) and orange (0.15 < CSP ≤ 0.20). Calcium ions are shown as yellow spheres. D. Residual CSP values calculated from the chemical shifts of the first to the last point of the titration of TUA-diCl. The dashed line at 0.04 ppm corresponds to the threshold CSP value chosen to highlight the most affected residues. CSP values were calculated from the ^{1}H and ^{15}N chemical shift variations ΔH and ΔN using the formula: $CSP = \sqrt{\Delta H^2 + (0.159 \times \Delta N)^2}$.

HDX-MS was used to gain additional structural information on the effects of TUA-diCl-binding to CaM. Pepsin digestion of CaM yielded 39 peptides covering 97.3% of the protein sequence (Figure 73). As observed with EF_3, the HDX-MS pattern of the isolated protein is typical of a well folded protein. The N- and C-CaM regions exhibit dynamic HDX patterns, indicative of the presence of secondary structural elements. However, the region linking the N-CaM and C-CaM lobes (peptides 73-84 and 73-85, Figure 74A) is fully exchanged after 10 sec labeling and displays no dynamic HDX activity. Consequently, these regions are solvent exposed and behaves as a flexible linker between the N- and C-CaM lobes.

Binding of TUA-diCl induces a pronounced reduction in solvent accessibility throughout the protein, with both N- and C-CaM lobes equally affected (Figure 74B). The massive reductions in uptake imposed by TUA-diCl strongly suggest that the inhibitor reshapes calmodulin from its conformation with independent N- and C-lobes into a more compact structure. Similar HDX results were obtained with the myosin light-chain kinase peptide, which is known to impose a compact globular structure to CaM upon binding[118].

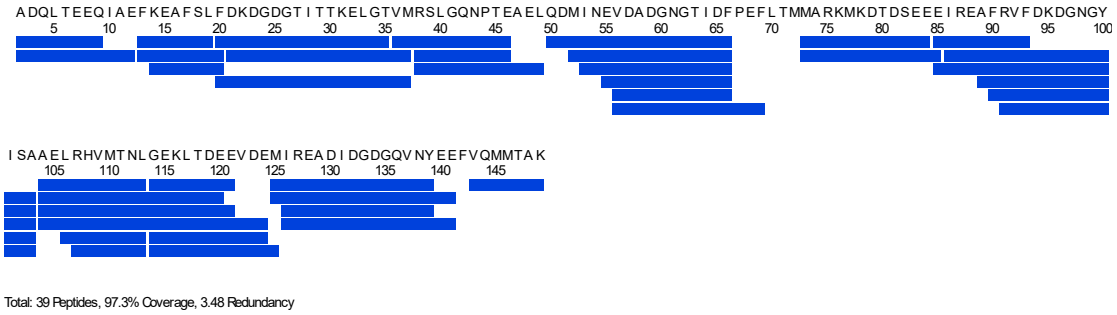


Figure 73- Peptide maps of CaM. *Maps were generated after 2 min pepsin digestion at 20°C under quench conditions. Each blue bar corresponds to a unique peptide identified by independent MS/MS acquisition.*

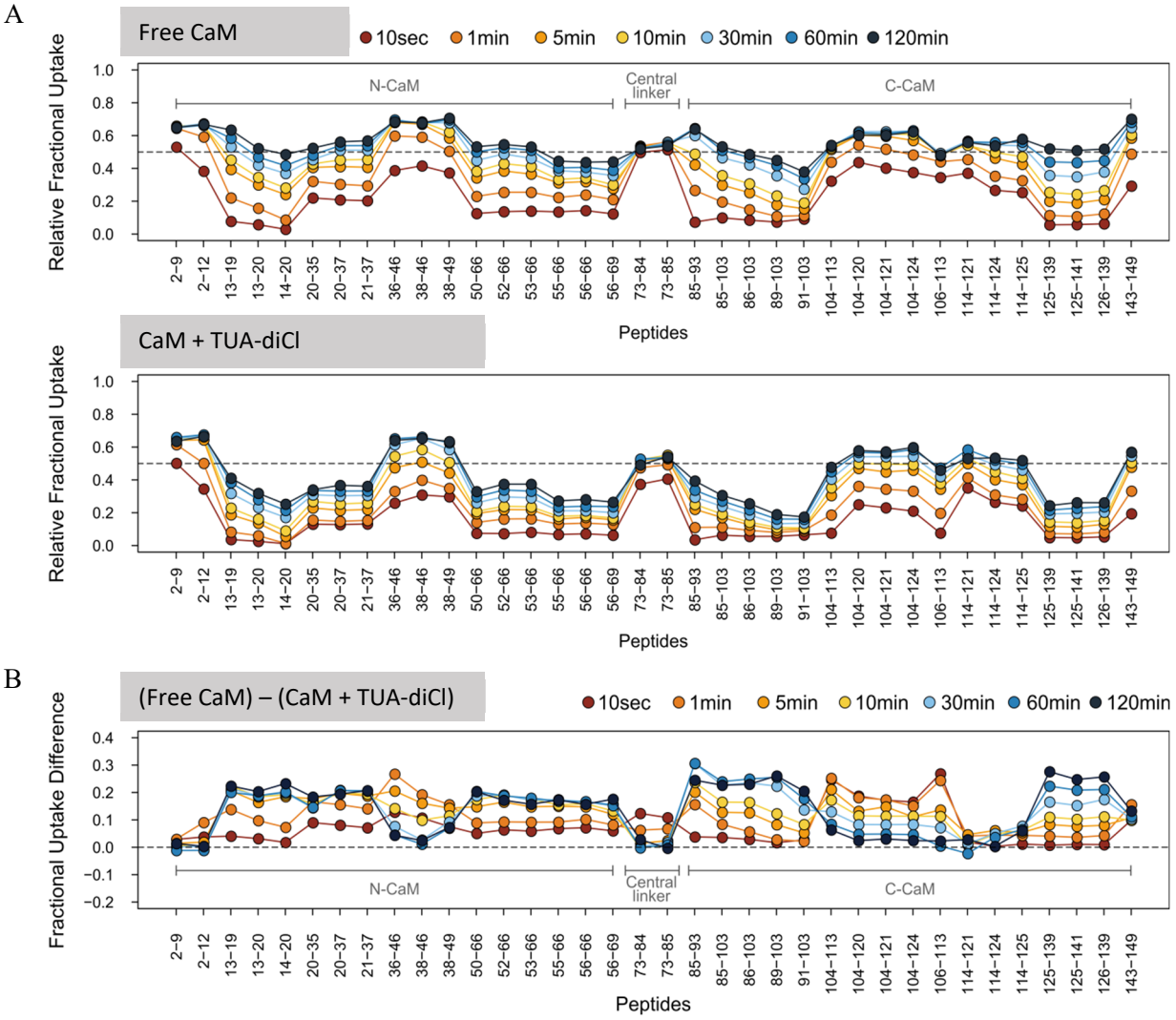


Figure 74- Effects of TUA-diCl binding on the solvent accessibility of CaM. *A. Relative fractional uptake profiles monitored for CaM alone (free) and in the presence of 87X molar excess of TUA-diCl. All labeling experiments were performed at room temperature using 85.7% excess D_2O and 2.5% final DMSO. The relative fractional exchange data calculated at each time point and for each condition was plotted as a function of peptide position. Each dot corresponds to the value extracted from one unique replicate (preliminary results). The position of the N-CaM, C-CaM and the central linker is indicated. B) Fractional uptake difference plot showing the differences in uptake calculated between the free- and the TUA-diCl-bound states at each time point and for each peptide. TUA-diCl binding to CaM dramatically reduces the solvent accessibility of the entire protein, with both N- and C-CaM lobes affected.*

D.III.3 Influence of TUA-diCl binding on the dynamics of holo-CaM

HDX data indicated that TUA-diCl reduces the solvent accessibility of the entire CaM, most likely stabilizing it in a compact globular conformation. NMR CSP data is in accordance

with this hypothesis as chemical shifts can vary because of conformational changes or different dynamics in addition to direct binding effects, and many residues from both lobes and the linker region sense the interaction. To establish whether a change in dynamics/conformation could be associated to the observed CSP upon TUA-diCl binding, we studied the dynamics of CaM in the presence and absence of TUA-diCl by means of ^{15}N relaxation experiments.

The N- and C-lobes of CaM are linked by a flexible region. As a consequence, both lobes show somewhat independent motions. If TUA-diCl interacts with both lobes and brings them together to form a more compact protein, this should be observable by variations of the tumbling correlation (τ_c) time upon binding. Using NMR, the correlation time of a protein is typically determined from the ratios of the longitudinal (R_1) and transversal (R_2) ^{15}N relaxation rates. However, in the presence of TUA-diCl, the transverse relaxation rate of most residues increases due to the exchange between the bound and free conformations on the millisecond-microsecond time range. In other words, the R_2 rate has a significant contribution from exchange hampering its use to calculate the τ_c. We thus used the 1D version of the TRACT sequence[209] to determine the tumbling correlation time of CaM in the presence or absence of TUA-diCl. This method allows one to calculate the τ_c for systems with high R_2 relaxation rates and is not sensitive to chemical exchange phenomena. In the absence of TUA-diCl, at 37°C, CaM shows a τ_c of 3.7 ± 0.2 ns (Table 9). This value can be compared to 5.5 ns, which is the theoretical tumbling correlation time of a spherical protein with CaM molecular weight and amino acid composition at 37°C using the Stokes-Einstein relation. Hence, this low τ_c value CaM is indicative of fast tumbling and of independent motions of the N- and C-CaM lobes as previously observed[97]. Upon addition of TUA-diCl (245 µM) to CaM (99 µM), the correlation time of CaM increases to 4.8 ± 0.3 ns. Because under the conditions used, CaM is not saturated with TUA-diCl (see below), the correlation time of the CaM-TUA-diCl complex will be even higher. Thus, CaM becomes more globular and compact, in agreement with the solvent accessibility data obtained by HDX-MS.

	CaM (248 µM)	CaM+TUA-diCl (99:245µM)
τ_c (ns)	3.7 ± 0.2	4.8 ± 0.3

Table 9- Effect of TUA-diCl binding on the tumbling correlation time of CaM.

The compaction of CaM upon interaction with TUA-diCl should have an impact on the internal dynamics of the protein. We thus characterized the influence of ligand binding on the ^{15}N T1 and T2 relaxation times and the ^{1}H-^{15}N heteronuclear NOE (Figure 75). Free calcium loaded CaM shows similar internal T1, T2 and ^{1}H-^{15}N NOEs for residues in the N- (1-76) and C- lobes (82-148), with the exception of residues 115-116 located in a dynamic loop between two α-helices in C-CaM and the N- (1-4) and C-termini (146-148) of the protein. Because the tumbling correlation time depends on the T1/T2 ratio, we can conclude that both lobes show the same tumbling rate. The very dynamic behavior of the linker region (77-81) can be established by the high T2 and T1 times and the low ^{1}H-^{15}N NOEs. The internal dynamics of holo-CaM corresponds well to those described in the literature[97] with a central flexible linker[280,281] and two folded lobes with independent and similar dynamics.

TUA-diCl binding has a profound effect on the dynamics of CaM evidenced on the T2 and ^{1}H-^{15}N NOE parameters. All the residues show a lower T2 in the presence of TUA-diCl, with mean and standard deviations of 126 ± 35 and 44 ± 18 ms in the absence or presence of TUA-diCl, respectively. In addition, the T2 values become very variable between residues. This drop and variability of the transverse relaxation times can be attributed to the exchange between the free and bound conformations in the ms - μs time range that has a substantial contribution to the transverse relaxation rate. The exchange contribution for a two-state binding depends on the exchange rate (k_{ex}) between the free and bound forms ($k_{ex} = k_{off} + k_{on}$ [TUA-diCl], the difference of the resonance frequencies between both forms and the magnetic field. For a single binding event (same k_{ex}), the relative exchange contribution depends on the difference of chemical shifts of the bound and free forms and, along with the CSP values, can be used to map the binding site along with the changes in dynamics due to conformational changes. Interestingly, residues (73-82) in the highly flexible linker region of free CaM show very low T2 values, most likely indicating that this region is involved in the interaction, although a conformational exchange between different bound forms of CaM on the ms-μs time range cannot be excluded. Similar conclusions can be drawn for the region 110-113, which also shows high CSP values. The global drop in T2 values is associated to a global increase in ^{1}H-^{15}N NOE values (free: 0.67 ± 0.16, with TUA-diCl: 0.72 ± 0.18), with only some residues in both lobes showing significantly lower values in the presence of TUA-diCl (39, 55, 100, 102, 110, 113, 121, 130 and 141). This increase can be assigned to a global rigidification of calmodulin in the presence of the ligand, as the heteronuclear NOE values are associated to the amplitude of motion on the ns-ps time scale, with higher NOEs denoting low amplitude motions and vice versa. The two regions in which fast motions get more restricted in the CaM bound form are

(73-76) in the vicinity of the linker (77-81) and (87-88) in the C-CaM lobe. Thus, the linker N-terminal region and its vicinity show lower amplitude fast motions typical of folded regions in globular proteins. Although NOE data could not be determined for residues 77, 82-86, the NOEs for residues 79 and 80 denote high amplitude motions as in free CaM.

In summary, TUA-diCl influences the internal dynamics of CaM, with a global restriction of rapid motions in the ns-ps time scale, very fast ^{15}N transverse relaxation rates due to conformational exchange on the ms-µs time scale and a compaction of the protein evidenced by a lower tumbling time. The flexible linker residues 79 and 80, conserve their dynamic nature in complex with TUA-diCl but residues (73-76) show typical NOE values of ordered residues in globular proteins.

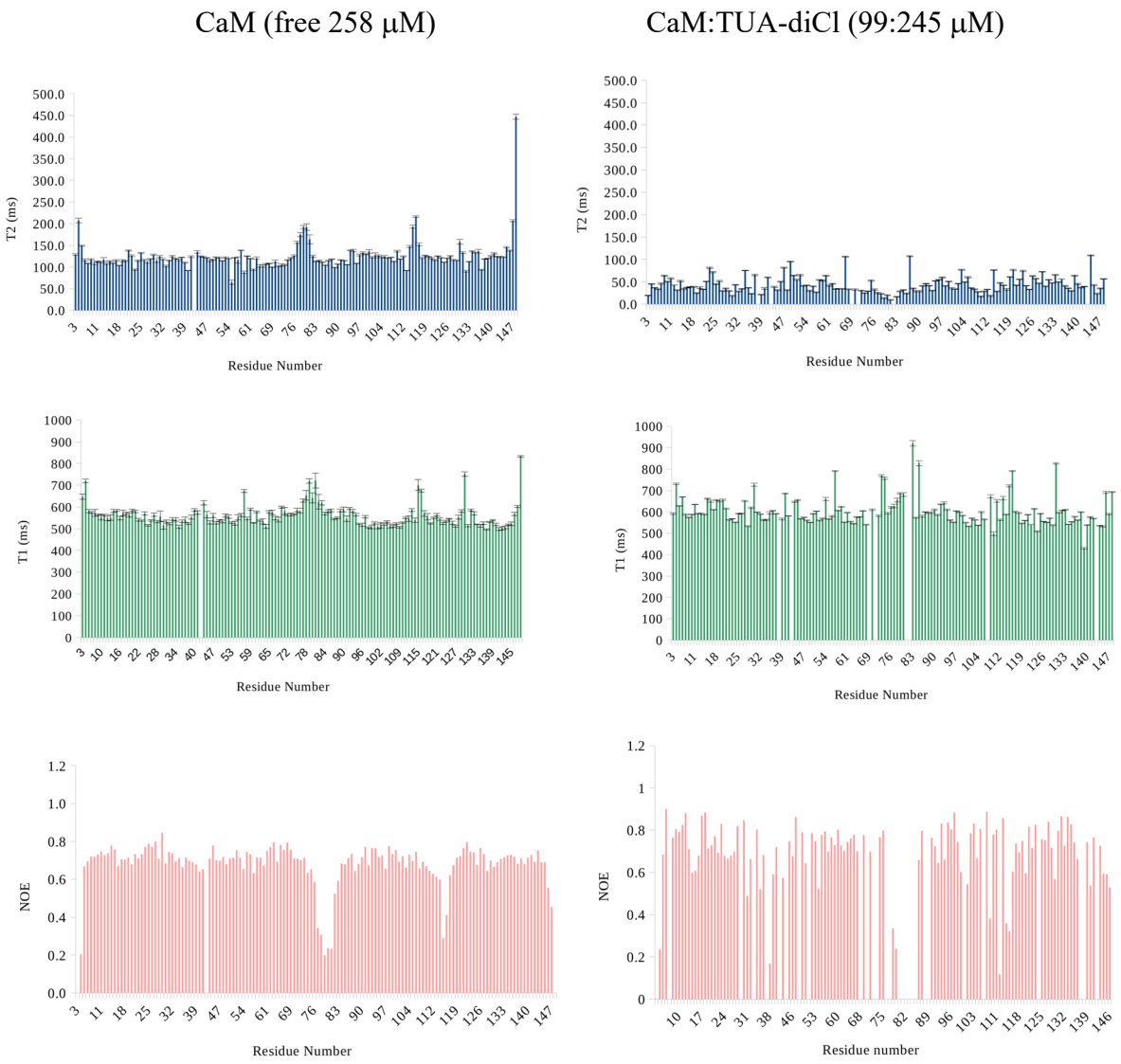


Figure 75- Effect of TUA-diCl on the internal dynamics of holo-CaM. *The ^{15}N relaxation T2 (top), T1 (middle) times and heteronuclear ^{1}H-^{15}N NOE (bottom) of holo-CaM without (left) or with TUA-diCl*

(right) are plotted as a function of the residue number. Data were acquired at 37 °C and 600 MHz using CaM concentrations of 248 µM (free) or 99 µM when in the presence of 245 µM TUA-diCl. It should be noted that the heteronuclear NOE data is very noisy due to exchange broadening of the signals (low T2 values) and the low concentration of calmodulin. Consequently, it could not be determined for several residues.

D.III.4 Binding mechanism monitored by line shape analysis

The evolution of the CaM amide signals (chemical shift and line shapes) caused by increasing concentrations of TUA-diCl was monitored to obtain the dissociation constant and the exchange rates between the free and bound forms. To this end, we used the software TITAN[208] (titration analysis). In contrast to other line shape analysis software that fit single traces to the relaxation equations in the presence of exchange, Titan uses a quantum mechanics algorithm to follow the evolution of the magnetization during the experiment (HSQC, HMQC…) and performs a global fit of the line shapes and positions of 2D ^{1}H-^{15}N peaks of selected residues for all ligand concentrations. As a first approach, we considered the simplest mechanism, a two-state exchange with a 1:1 stoichiometry, and fitted the signals of 22 residues (in 9 spectra) distributed in the structure of CaM. Using a two-state model, the equilibrium dissociation constant is 21 ± 2 µM and the dissociation rate k_{off} is 727 ± 39 s^{-1}. Consequently, the association rate constant k_{on} is 3.5 ± 0.2 10^{7} $M^{-1}.s^{-1}$. A two-state model produces a reasonable fit that can be considered as a first approach to obtain approximate values of the thermodynamic and kinetic parameters of the equilibrium (Figure 76). However, the intensities of the peaks during the transition are not always well-described by the fit to a two-state model and 1:1 stoichiometry, indicating that the mechanism of the TUA-diCl interaction with CaM is more complex. Further work fitting to more complex models must be performed to obtain a better description of the binding mechanism and its associated parameters.

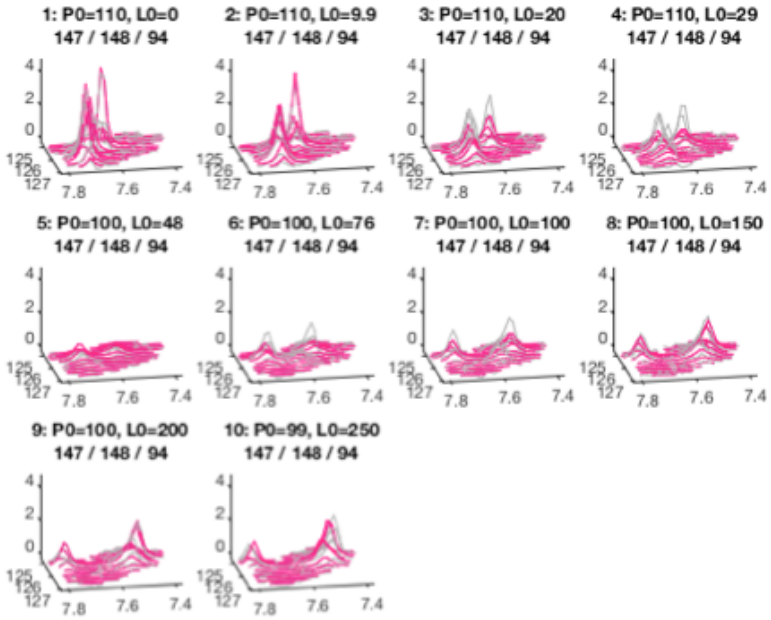


Figure 76- Line shape fit for residue M145. *^{1}H (abscissa) and ^{15}N (ordinate) evolution of the signal of residue M145 at different protein (Po) and TUA-diCl (Lo) µM concentrations. The fit (magenta) and the experimental (grey) traces are displayed. The peaks of residue M145 are in a cluster of signals with residues 94,147 and 148 indicated in the individual plot legends. The image and the fit were produced with the software Titan.*

D.III.5 Molecular docking and dynamics simulations studies

Two docking strategies were employed in an attempt to characterize the binding mechanism of TUA-diCl to CaM. The first strategy was to dock the ligand on the two lobe regions of the CaM structure and then to filter the possible solutions using two experimental pieces of information derived from NMR: (i) the CSP values of CaM and, (ii) STD_{exp} ratios obtained for TUA-diCl. Docking using experimental data (CSP) to filter the solutions was performed with AutodockVina[135], whereas docking using CSP data as restraints was done on the HADDOCK2.2[215] webserver (https://www.bonvinlab.org/software/haddock2.2). STD data was used either to filter the solutions obtained with CSP data as constraints or to directly filter the solutions with CSP. Only one TUA-diCl molecule was docked onto the CaM structure. In agreement with the tumbling correlation time, internal dynamics, the CSP and HDX-MS data, we chose a compact structure of CaM for all docking procedures and MD simulations (see below), namely the structure of CaM in interaction with the small hydrophobic compound KN93 (PDB ID 6M7H[183]). For all docking procedures, only the residues with the highest CSP values were used to filter the solutions or to constrain the calculations (F19, V55, M72, M76, G113, Q143, M145, T146, A147).

D.III.5.1 Molecular docking filtered using CSPs

TUA-diCl was initially placed in the single central cavity found by *mkgridXf*[216] software using a high r_{ou} value (8 Å). Twenty molecules of TUA-diCl were then docked using AutodockVina[135] on this cavity. Here, we choose to use principally the CSPs data over the STD data. The main reason for this choice is that, while from the STD data we know that most of the hydrogens of the ligand are involved in the interaction. CSP provides precious information on the protein residues that may be located in the binding site. The minimum average distance between the CSPs and the atoms of the ligand was thus used to filter the data and select the best matching conformations (Figure 77). In these models, TUA-diCl was located in the center of CaM and residues from the N- and C-CaM lobes as well as close to the linker region were engaged in the interaction with the ligand.

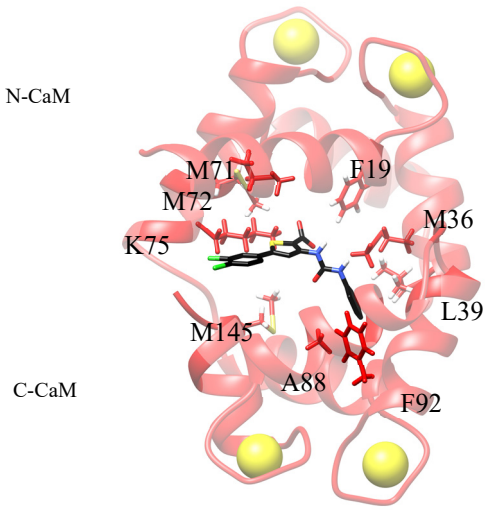


Figure 77- Best docked position of the ligand filtered by the experimental CSP values*. The CaM residues interacting with the TUA-diCl (black) are indicated and shown in red sticks colored by heteroatoms. The calcium ions are displayed as yellow spheres.*

D.III.5.2 Molecular docking using CSPs as restraints

The docking calculations with HADDOCK2.2[215] were done generating 200 structures representing different poses of one molecule of TUA-diCl within CaM in agreement with the CSP values. During the docking procedure the protein (CaM) and the ligand (TUA-diCl) were kept flexible, which permits an enhanced exploration of the mutual interactions. The 200

structures were then clustered using the self-organizing map (SOM)[219] approach into four clusters, representing the variability of molecular conformations (Figure 78).

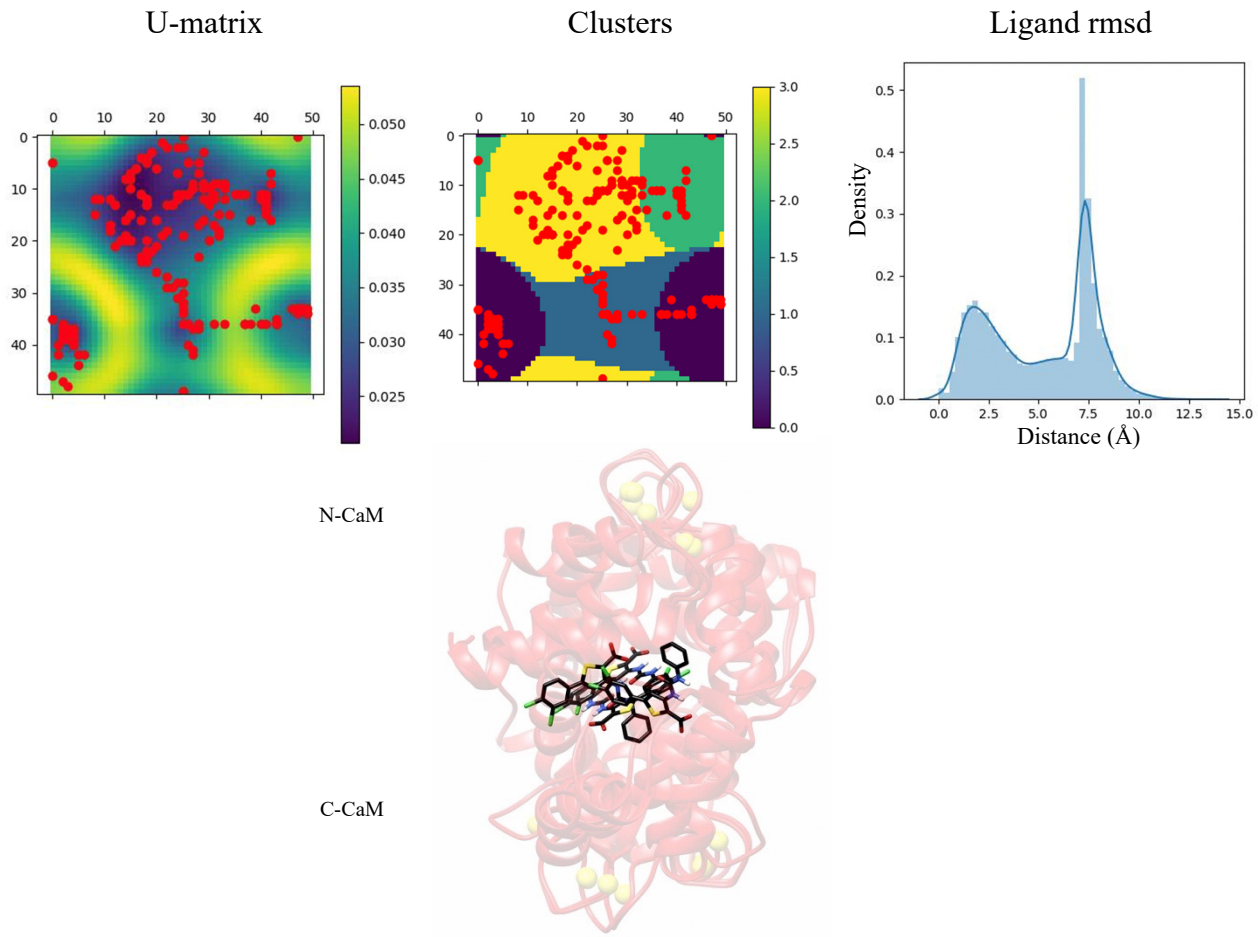


Figure 78- Clustering of the docking poses obtained with HADDOCK based on the CSPs data. *(Top) From the left to the right, the self-organizing map U-matrix and the clusters are showed and the pairwise distance between the position of the ligand among the 200 conformations. (Bottom) Superimposition of the representative conformations of TUA-diCl (in black) and CaM (in red) obtained from the four clusters. The red dots on the U-matrix plot and on the cluster map represent each CaM-TUA-diCl conformation.*

The theoretical STD_{th} values were calculated for each representative conformation of each cluster and the comparison between the STD_{th} and the STD_{exp} was carried out as described in Chapter II. The poses with the lowest RMSE of each cluster were then compared (Figure 79). One orientation of the molecule seems to better match the STD data. In this orientation, the residues of CaM in close contact with TUA-diCl are A15, M72, Y75, D80, M145. The methionines involved are located in both CaM lobes, indicating that the compound can interact simultaneously with the two lobes.

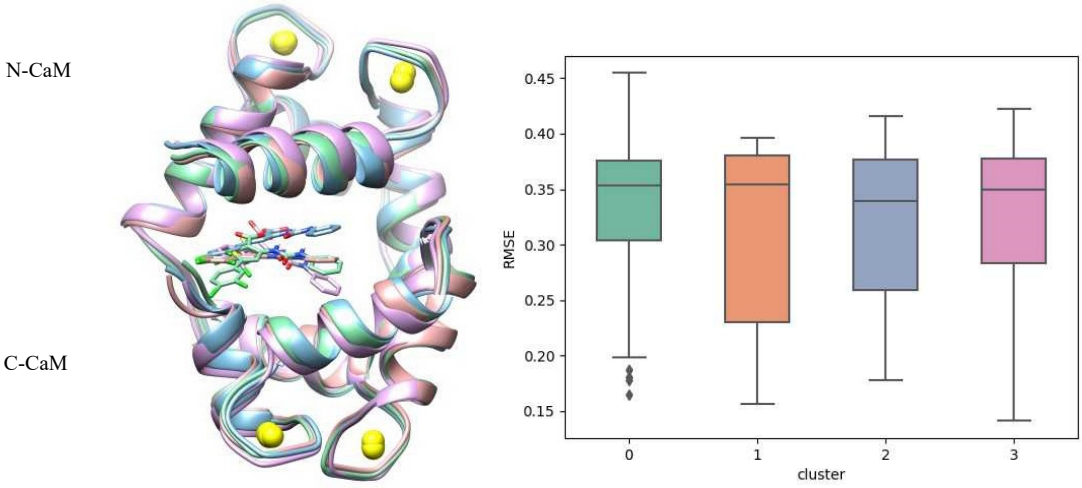


Figure 79- Cluster analysis with the RMSE of the calculated and experimental STD values. *(Left) Superimposition of the representative conformations for each of the clusters obtained by the self-organizing maps approach. The protein CaM represented in carton and the ligand TUA-diCl represented in licorice, are colored according to the cluster. (Right) Boxplot of the RMSE values of* STD_{th} *and* STD_{exp} *plotted for each of the four clusters. The same cluster colors are used for the left and right panel. Outliers are shown as diamonds.*

By superimposing the protein backbone of the best conformations obtained from each procedure, we noticed that the ligand was almost overlaid, with the di-Cl ring interacting with the C-CaM lobe terminal residues (M144, M145 and M76 vicinity of the linker). Although the residues in close contact with TUA-diCl do not exactly match the residues with high CSP used as ambiguous constraints or filters, except for A15, all the residues highlighted by the models are highly impacted (high CSP and/or low T2 values) and close to the residues used as input for the calculations. Overall, that the two dockings performed with different software and strategies using experimental data as constraints or to filter solutions converge and reasonably satisfy the experimental data, suggests that one molecule could bind to both lobes in a closed CaM structure.

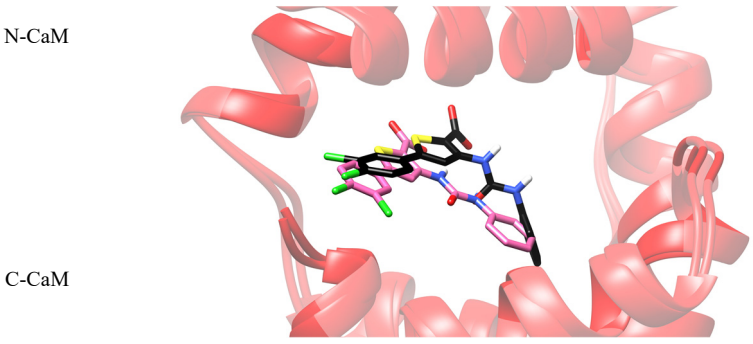


Figure 80- Comparison between the best docked structures using the CSP information as restraints or to filter data. *Superimposition of the best poses obtained using HADDOCK (pink) and AutodockVina*

(black) using CSP values as constraints or as filters, respectively. The CaM residues with high CSP values are colored in a darker red.

D.III.5.3 Molecular docking on the N- and C-CaM lobes

Previous calculations were done based on a single TUA-diCl molecule placed in a single cavity found with *mkgridXf*[216] using the r_{ou} parameter set to 8 Å. In order to investigate the stability of the ligand within the CaM lobes and investigate if the experimental data could be explained by the interaction of two TUA-diCl molecules with one CaM molecule, we used *mkgridXf*[216] with a lower r_{ou} value (6 Å) to establish cavities in the different lobes and performed MD simulations. Under these conditions, one cavity could be defined in each of the CaM lobes. Of note, some residues of a different lobe can participate in the cavity and interactions of the other lobe; in particular, in a closed CaM structure, the C-terminus of the protein is close to C-terminal moiety of the α-helix preceding the linker and the central flexible linker (residues 72, 75, 76). Twenty molecules of TUA-diCl were docked into the two cavities using STD data and Autodock Vina as explained before and then a molecular dynamics simulation of 1 μs was launched for the lowest STD RMSE poses of each cavity. The five lowest RMSE model statistics are displayed in Table 10 and the two highest scoring poses and the RMSE distributions plotted as boxplots are shown in Figure 81. The lowest RMSE values are found for the docking poses on C-CaM (cavity number 2), but the average RMSE value is lower for the docking poses on N-CaM (cavity number 1) (top of Figure 81). KN93 is a small hydrophobic molecule with three aromatic cycles and a size comparable to TUA-diCl. In the crystal structure of CaM with KN93, that we use here for docking and MD simulations, three molecules of KN93 are present[183]. As evidenced in Figure 81C, D the starting points for the MD simulations with TUA-diCl overlap with the molecules KN93-1 and KN93-2.

Cavity number	*RMSE*	*Energy (kcal.mol^{-1})*	*Volume (Å^3)*	*Position*
2	0.14	- 7.20	94.95	C-CaM
2	0.16	- 6.50	94.95	C-CaM
1	0.17	- 7.60	132.07	N-CaM
1	0.17	- 7.20	132.07	N-CaM
1	0.17	- 6.50	132.07	N-CaM

Table 10- Results of the molecular docking on two cavities using STD experimental data.

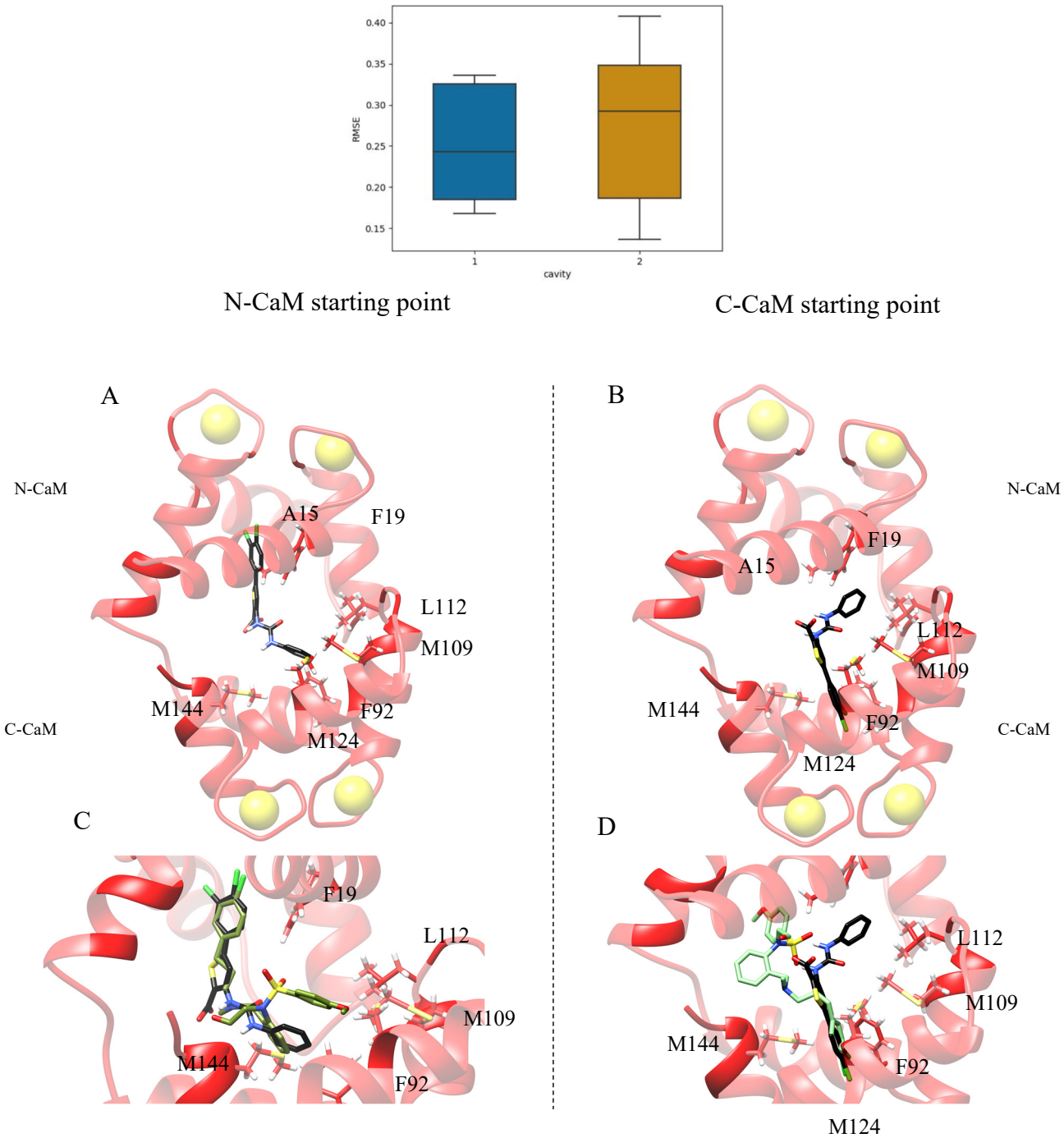


Figure 81- Starting points of the MD simulations for the complex CaM-TUA-diCl obtained by minimizing the RMSE between the STD_{th} and STD_{exp} values. *(Top) Box plot of the RMSE between the calculated and experimental STDs for each docking pose on each cavity. The boxplot is colored by cavities, with N-CaM represented in blue and C-CaM in orange. (A, B) Starting structures for MD with TUA-diCl (black sticks) in N-CaM (A) and (B) C-CaM cavities. (C, D) Superimpositions of the starting points of the MD simulations on each cavity with the KN93-1 (C) and KN93-2 (D) molecules in the X-ray crystallographic structure 6M7H used for docking and MD simulations. KN93 molecules are shown in different greens. The residues with high CSP values are highlighted in deep red; calcium ions are represented as yellow sphere.*

In the best poses obtained using experimental CSP and STD information, the CaM residues involved in the interaction with TUA-diCl for the N-CaM cavity were: F19, I27, L39, I63, F68, M71, Y75, F92, M145 and for the C-CaM cavity: A15, F19, F92, M109, L112, M124, M144. Most of these residues displayed significantly high CSP and/or low T2 values, suggesting that two (or more) TUA-diCl molecules could bind to CaM, with preferential interactions on each lobe. Moreover, the N- and C-CaM TUA-diCl molecules can coexist in a single structure. Interestingly, the di-Cl ring of TUA-diCl superimposes with an aryl ring of the KN93-1 and KN93-2 molecules buried in the different CaM lobes, suggesting that the orientation of this ring could be very important. Another interesting feature of this docking is that the negatively charged acetate group of TUA-diCl in the C-CaM cavity points towards K115 and G113 (both residues show high CSPs) suggesting that an ionic interaction that neutralizes buried charges could be important to stabilize the complex in closed conformation. Together, these results suggest that 2 molecules of TUA-diCl could bind in N- and C-CaM cavities similar to those occupied by KN93.

During the MD simulations the TUA-diCl ligand was highly flexible. Interestingly, the molecule docked in the N-CaM lobe migrated from an extremity to the central cavity of CaM and ended up in the C-CaM region, whereas the molecule docked in the C-CaM cavity stayed within the C-CaM lobe. These results suggest that the interactions of the ligand within the C-CaM lobe are more stable and/or that the energetic cost of opening the C-CaM cavity to allow TUA-diCl to migrate out of it is higher. We extracted 200 frames along the trajectories in which the ligand and the CaM residues with higher CSP values (F19, V55, I63, M72, M76, G113, M145, A146, T147) showed the lower distances. To this end, the minimum distances between the heavy atoms of the ligand and the protein were calculated using PLUMED[282]. The simulations recorded from the two starting points were different and displayed different interaction patterns. Eight of the CaM residues with a high CSP match with the ligand positions along the MD trajectory starting from the TUA-diCl pose on N-CaM, whereas only five CaM residues with a high CSP match with the ligand positions along the MD trajectory starting from the TUA-diCl pose on C-CaM. Although the MD trajectory starting from the ligand pose in the N-CaM cavity fits better the experimental data and the binding of one molecule of TUA-diCl could explain the quasi totally of the observed higher CSPs, the ligand is not stable at this position (Figure 82).

By comparing the CaM residues affected by the addition of TUA-diCl with the residues involved in the interaction with the CaM ligand KN93 (Table 11), seven residues interact both with TUA-diCl and KN93-1 and six in the case of KN93-2. In addition, the X-ray

crystallographic structure of the CaM-KN93 complex was superimposed with the MD frame in which TUA-diCl is close to the eight residues with a high CSPs. The position of TUA-diCl is parallel to KN93-1 and the di-Cl ring points toward the N-CaM lobe.

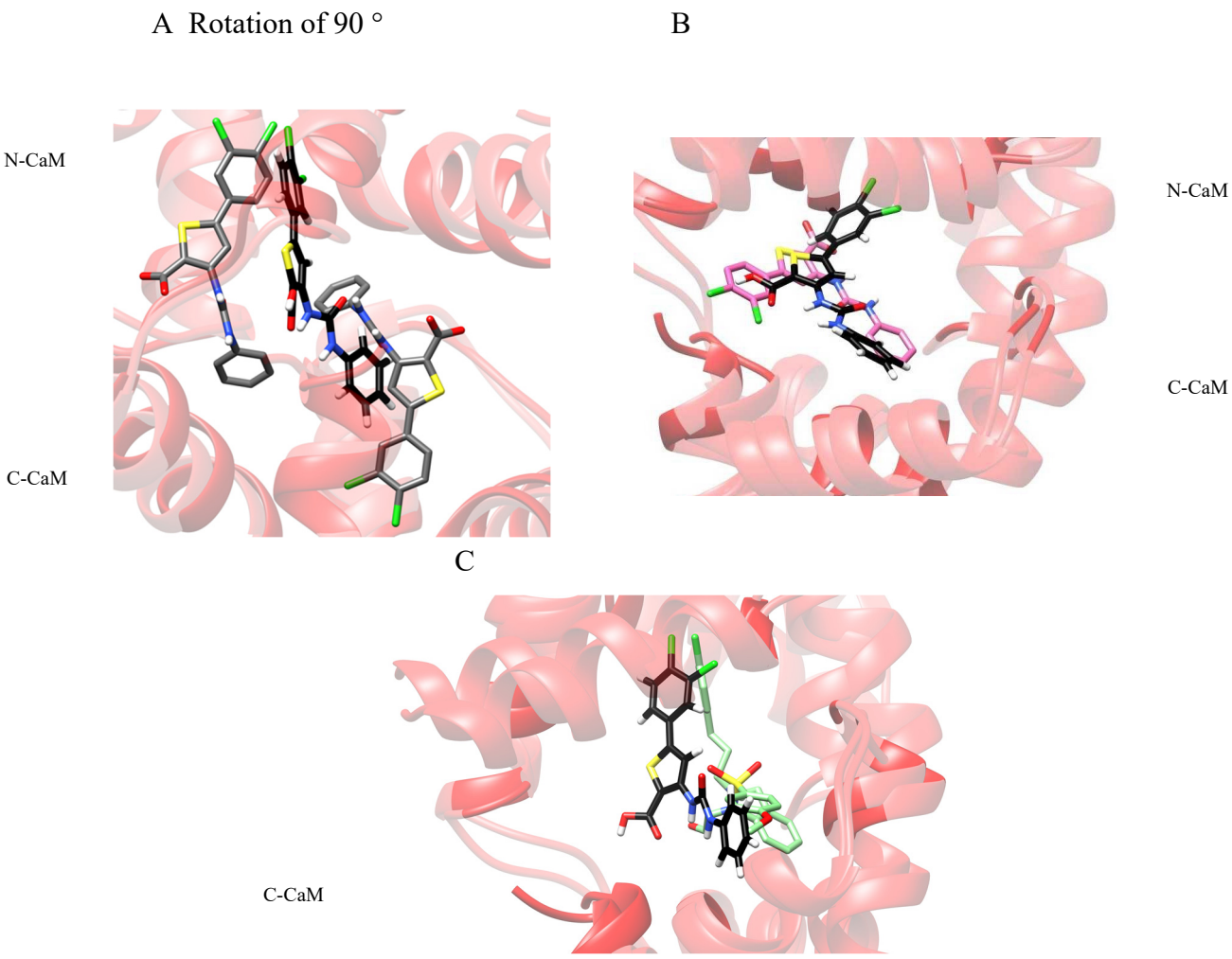


Figure 82- Best extracted pose of TUA-diCl along the MD trajectory matching with the experimental CSPs data. *A. Comparison between the positions of the TUA-diCl ligand starting point of the MD simulation (grey) and the extracted pose using the CSP data (black). CaM is in red transparency and the residues with high CSPs are highlighted in red. B. The pose of TUA-diCl refined with the CSP and STD data (pink) is superimposed with the pose of the ligand extracted from the simulation in agreement with the CSP data (black). C. KN93-2 (green) is close to the position of TUA-diCl (black) obtained during the MD simulations.*

Residue number CaM	TFP-1	TFP-2	TFP-3	TFP-4	W7-1	W7-2	KN93-1	KN93-2	KN93-3	TUA-diCl
	C-CaM	Interdomain	Interdomain	N-CaM	N-CaM	C-CaM	N-CaM	C-CaM	interdomain	
	X-ray (PDB ID 1LIN)[185]				NMR (PDB ID 1MUX)[271]		X-ray (PDB ID 6M7H)[183]			
8			X							
12			X					X		
11		X	X							
14		X								
15		X	X							
18		X							X	
19				X	X		X	X		X
27				X	X		X			X
32				X	X		X			
36					X		X			
39							X			X (CSP = 0.102)
51				X	X		X			X
52				X	X					
54				X						
55				X	X					X
63				X	X		X			X
68				X	X		X	X		
71				X	X		X	X		
72			X		X			X		X
74			X							
76								X		X
92	X		X			X	X	X		
100	X					X		X		
105	X					X	X			
109		X				X	X		X	X
112							X			
114		X								
115		X								X
116		X								
120		X								
124	X	X				X	X	X	X	X (CSP = 0.113)
125	X					X		X		
127	X									
128	X					X				
136	X					X				
141	X					X		X		
144	X		X			X			X	
145			X			X	X	X		X
147			X							X

Table 11- Residues of CaM involved in the interaction with CaM inhibitors and TUA-diCl residues with high CSPs.

DISCUSSION

The interaction of TUA-diCl with CaM was studied by heteronuclear and homonuclear NMR and HDX-MS experiments. Using the ligand epitope mapping from experimental STD ratios and/or the chemical shift perturbations measured on the protein by NMR, docking positions of TUA-diCl were predicted by in silico docking approaches. The stability of the predicted ligand poses was probed by recording MD trajectories.

We showed that the interaction of TUA-diCl with CaM is calcium-dependent, and thus requires the presence of Ca^{2+} loaded CaM. It is known that holo-CaM shows exposed hydrophobic patches that can wrap around segments of residues in a helical conformation or interact with small hydrophobic molecules such as TUA-diCl. Binding of TUA-diCl leads to the compaction of CaM as reflected by the decrease in solvent accessibility monitored by HDX-MS, the reduction in the tumbling rate, and the overall reduction of fast motions of CaM in the ps-ns time scale. Binding of TUA-diCl has also a significant impact in the environment of many CaM amide groups as established by the CSP analysis, indicating that the N- and C-CaM lobes, together with residues in the vicinity of the flexible linker are involved in the interaction with the ligand. Moreover, most of the residues with high CSP values are in the hydrophobic patches of CaM that have been observed to interact with other small hydrophobic molecules, and the STD epitope mapping indicates that the three aromatic cycles of TUA-diCl are in close contact with CaM. Together these results indicate that TUA-diCl interacts with exposed hydrophobic CaM patches, bridging together the N- and C-CaM lobes in a compact globular structure of CaM. Within the complex, residues 79-80 in the linker still display high amplitude motions, showing similar fast dynamics to those of flexible loops within globular proteins.

In order to establish a mechanism of interaction and the binding site(s), it is required to know its stoichiometry. However, the low affinity of the system ($K_d \geq 20$ µM as established for a 1:1 two-state binding mechanism) and the aggregation of TUA-diCl at concentrations above 50 µM (which can change the nature of the interactions) hampered the direct determination of the stoichiometry in this work. We thus resorted to in silico docking, MD simulations and to the comparison with the interaction of CaM with small hydrophobic molecules.

Like for TUA-diCl, interactions between aromatic rings of small hydrophobic CaM ligands and the hydrophobic patches have been observed for CaM binders such as W7, TFP,

KN93 or CMZ (Figure 83). These molecules display chemical features that are also encountered in TUA-diCl: at least two aryl groups, a charge at neutral pH (positive for W7, KN93, and CMZ, negative for TUA-diCl). These CaM ligands interact with several methionines located in the patches. The high CSP values of M72, M76 and M145, strongly suggest that these residues are also involved in the interaction with TUA-diCl.

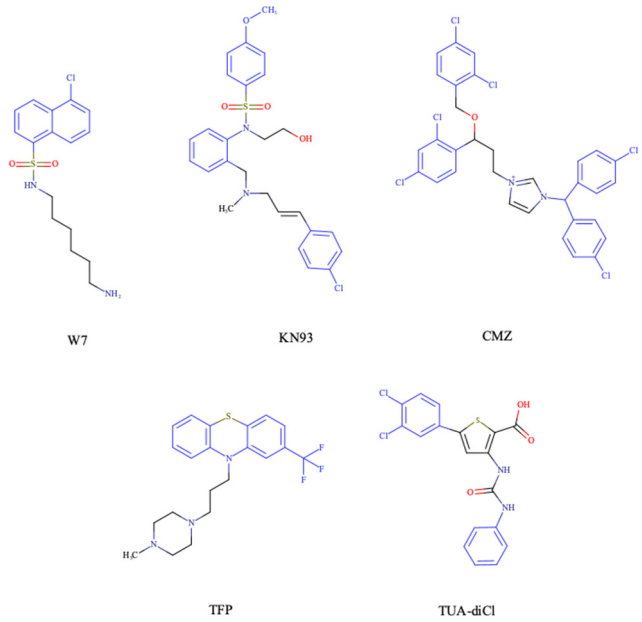


Figure 83- Comparison of TUA-diCl with some holo-CaM inhibitors. *All the molecules show two or more aryl groups (in blue). W7 and KN9 have a sulfonamide moiety and TUA-diCl a thiophen moiety.*

The comparison of the CaM residues with high CSP values in the presence of TUA-diCl with the residues that interact with these molecules in the corresponding structures, reveals CaM residues in common for TUA-diCl with W7-1, TFP-4 (4 residues) and with the two KN93 molecules bound to each of the CaM lobes: KN93-1 (7 residues), KN93-2 (6 residues) (Table 11). Thus, *a posteriori*, the structure of KN93 in complex with CaM that we chose seems to be the best available option for docking and MD simulations. Depending on the TUA-diCl-CaM stoichiometry (1 or 2 to 1), two ligand molecules could be bound to CaM interacting each with cavities in the N or C lobes, or one ligand would bind between N-CaM and C-CaM (Figure 84). The closed conformation of CaM could cope with one or two additional TUA-diCl molecules.

Most of the known interactions of CaM with small ligands of the size of TUA-diCl or even bigger, display more than one ligand involved in the complex, suggesting that more than one ligand might be necessary to keep CaM in a closed compact conformation. Although

dockings on a single cavity converged into very similar structures that can account for several of the experimental observations (mainly CSPs and STDs), docking on two cavities (a single ligand at a time) agrees better with the CSP data. This suggests that two molecules, rather than one, could be involved in the interaction with CaM. Moreover, the coincidence of the residues with significantly high CSP values and the residues that interact with KN93 molecules 1 and 2 is striking, TUA-diCl can occupy cavities in both lobes and the, docked TUA-diCl molecules superimpose well with the KN93 molecules, further supporting that two TUA-diCl molecules bind to CaM and suggesting that very similar cavities are used by the ligand in both cases.

In the case of KN93, three molecules are involved in the interaction. Two molecules, presumably KN93-1 and KN93-2, bind to CaM with similar affinities interacting with both lobes and without interacting with each other. In a cooperative mechanism, a third molecule of KN93 binds with a much lower affinity[183] to form a hydrophobic tunnel. In the case of TUA-diCl, no sign of cooperativity was observed in the binding data obtained by NMR and fluorescence, but both methods lead to different Kd with a ~ 4x fold difference. We cannot statue at this stage, whether this difference is significant and both methods report on different binding events or not. Irrespective of the number of TUA-diCl molecules that bind to CaM, the NMR, HDX-MS and in silico data indicate that TUA-diCl stabilizes a closed compact conformation of CaM, that can explain how this compound would inhibit the activation of EF. Indeed, in the structure of the EF-CaM complex, the N-CaM and C-CaM lobes are extended and interact with the exposed hydrophobic patches with different regions of EF. Therefore, a closed conformation cannot interact with EF to activate it (Figure 84).

Importantly, given that the affinity of TUA-diCl is much higher for EF than for CaM, and that we showed in Chapter II that EF inhibition is not CaM dependent, the direct interaction of TUA-diCl with CaM should not play a significant role in the observed EF inhibition. Nevertheless, to further develop a CaM inhibitor, much attention should be paid to avoid unwanted interactions with calmodulin

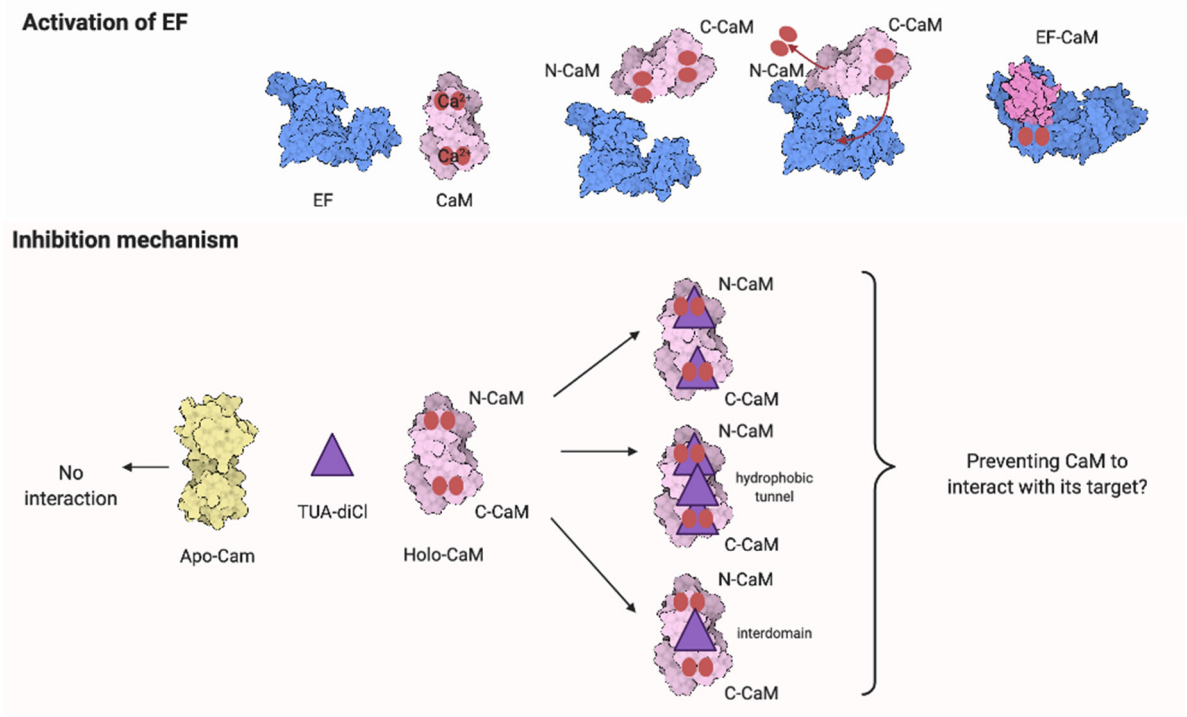


Figure 84- Possible mechanisms of the TUA-diCl mode of binding to CaM. Both CaM domain plays an important role in the activation. *(Top) Activation of EF by CaM. The interaction between the N-CaM lobe and EF is required for the insertion of the C-terminal of CaM within the helical domain and the catalytic core of EF. N-CaM loses its two calcium ions (red spheres) when bound to EF. TUA-diCl can potentially disrupt the activation of EF by stabilizing a closed compact conformation that would hamper the interaction of the N-CaM lobe with EF and the insertion of the C-CaM lobe in the helical domain of EF. One molecule in a central cavity, two molecules each one in a CaM lobe, or three or more molecules forming a hydrophobic core between the lobes could be implicated in the interaction.*

Chapter IV. *In silico* analyzing of the relationship between the activation of the edema factor and its interaction with calmodulin

INTRODUCTION

As the interactions between proteins are essential in all biological processes, the modulation of these interactions with small ligands is a promising direction[283–290]. During the last decade, virtual screening has experienced a turning point where interest has widened from protein active sites to cryptic sites[291,292] not visible in the isolated protein conformation but formed upon ligand binding. These cryptic sites have been proposed to be detected by analysis of protein structures[293,294], by mixed-solvent MD simulations[295,296] or by biased MD simulations[297,298]. Inhibition of protein-protein interactions, search for new cryptic sites and targeting protein function in an allosteric way are three closely related goals. Targeting allostery[299–302] is particularly appropriate in the case of inhibition of protein-protein interactions, as the induced conformation changes in which allosteric communication plays a strong role during the establishment of interaction, can create or modify cavities[303].

Allostery, discovered in the early days of molecular biology[153,155] has since then, evolved from a model of discrete protein conformations to a more continuous description of the free energy landscape of proteins[162,163,303]. Consequently, the allostery is nowadays quite often described as being closely connected to the variations of equilibrium between protein conformations. Networks of protein residues have been pointed out to be involved in the transmission of these equilibrium variations[304–306]. The evolution of allostery description has allowed the emergence of so-called allosteric ligands[169,303], for which the binding to the protein has a long-range influence on the protein conformation via residue networks. These allosteric ligands and the associated allosteric pockets attract a lot of attention, because they expand the range of pockets and ligands that could be exploited in effector design by virtual screening studies[307]. In addition, the diversity of allosteric phenomena can be utilized to limit the emergence of resistance mutations in target proteins, making carefully selected allosteric ligands more robust to the appearance of resistance in pathogens.

As a result, numerous methods have been developed for allosteric pocket detection in protein structures. Many of these methods are based on a description of protein structures as an elastic network in which each residue is replaced by one atom (generally the Cα) and the structural deformations are modeled through a set of springs between this atoms[217,308]. Ligands are systematically docked on each protein pocket. The overall structure deformation due to the ligand docking is then detected. The pockets for which docking produces a deformation larger than a given threshold are predicted as allosteric pockets. Other approaches are based on a detailed mechanistic simulation of the functional motions[192].

In the present chapter, we investigate the use of molecular dynamics (MD) simulations and bioinformatics approaches to analyze protein-protein interactions, the variability of cavities and allostery for one example of protein-protein interaction corresponding to the activation of a virulence factor. The Edema Factor (EF) of Bacillus anthracis is activated in the cytoplasm of the host cell by interacting with the ubiquitous protein calmodulin (CaM). This interaction depends on the level of Ca^{2+} loaded by CaM, the C terminal lobe of CaM (C-CaM) displaying the highest affinity for ions Ca^{2+} during the interaction with EF[109]. The EF-CaM complex has been extensively studied by structural biology and biophysical techniques as well as by molecular modeling[192,84,111,112,115,119]. This complex (Figure 90A) represents a very good example of an interaction with induced conformational selection for both partners. Indeed, free CaM in solution displays a very heterogeneous set of conformations, with a wide range of relative re-orientations of N terminal (N-CaM) and C terminal (C-CaM) lobes[310,311], whereas CaM in complex with EF is blocked in an extended conformation. Similarly, the inactive state and the activated state of EF display largely different conformations. The helical region (residues 660-800) is moved apart from the C_A (residues 292-349 and 490-622) region to allow CaM insertion. A large conformational reorganization of switches A (residues 502-551, purple), B (residues 578-591, cyan) and C (residues 630-659, yellow) also takes place and the catalytic site is reshaped in its active architecture (Figure 90A). Indeed, switch B is disordered and not visible in electronic density in the absence of CaM. Strikingly, in switch C, two β-strands and a connecting loop are converted to an α helix in the EF-CaM complex structure.

In the literature, both orthosteric and allosteric ligands have been proposed to inhibit EF activity. Several orthosteric inhibitors, binding to the catalytic site, have been discovered[122,124,127,143,147,312]. Among them, the ligand adefovir[127] was found by X-ray crystallography to bind in the catalytic site in the presence of a Yb^{3+} ion coordinated by adefovir

as well as by protein residues. On the other hand, the compound 10506-2A has been shown to be an IPPI (inhibitor of protein-protein interaction) and to bind close to the EF helical regions[181]. Thiophen ureidoacid (TUA) ligands have been discovered following a virtual screening on the pocket SABC, formed by residues from the three switches A, B, and C[192]. The TUA compounds was designed to prevent the activation of EF by CaM, as this SABC pocket is present in the closed state of EF and early disappearing in the open state. The TUA-diCl is found to be the most active one. As they do not bind to the EF enzymatic site, compounds 10506-2A and thiophen ureidoacids should thus by definition act through allostery.

Here, we propose the following approach to detect protein regions which should be targeted using an allosteric approach to inhibit the activity of EF. Starting from the X-ray crystallographic structure of EF-CaM complex bound to the orthosteric inhibitor adefovir[127], we destabilized it by removing alternatively several co-factors: the ion Mg^{2+} present in the catalytic site, the ions Ca^{2+} loaded by CaM or the ligand adefovir. The corresponding trajectories were analyzed to detect the long-distance communications associated with the structure destabilization. By analogy with the methods used for allosteric pockets prediction[217,308], the cavities in the EF-CaM complex were tracked along the MD trajectories, and the ones displaying the largest shape variations were labeled as allosteric. All these analyses point out the interface EF-CaM as a good candidate to discover allosteric inhibitors of the EF enzymatic activity. This interface is indeed destabilized by TUA-diCl an allosteric inhibitor. By combining experimental data and *in silico* studies the binding site and mechanism of action can be highlighted, offering new strategies to explore the molecular interaction and find new pathways to inhibit specifically a target.

RESULTS

D.IV.1 The removal of co-factors destabilizes the architecture of the EF-CaM complex

The evolution of the root-mean-square deviations (RMSD) with respect to the initial conformations of chains A (EF) and D (CaM) in the structure 1PK0 calculated along the trajectories displays quite similar trends, with a plateau attained in most of the cases after 50 ns

and located between 4 Å and 6 Å (Figure 85). The highest values are observed for trajectories EF_ade_Mg_CaM and EF-CaM.

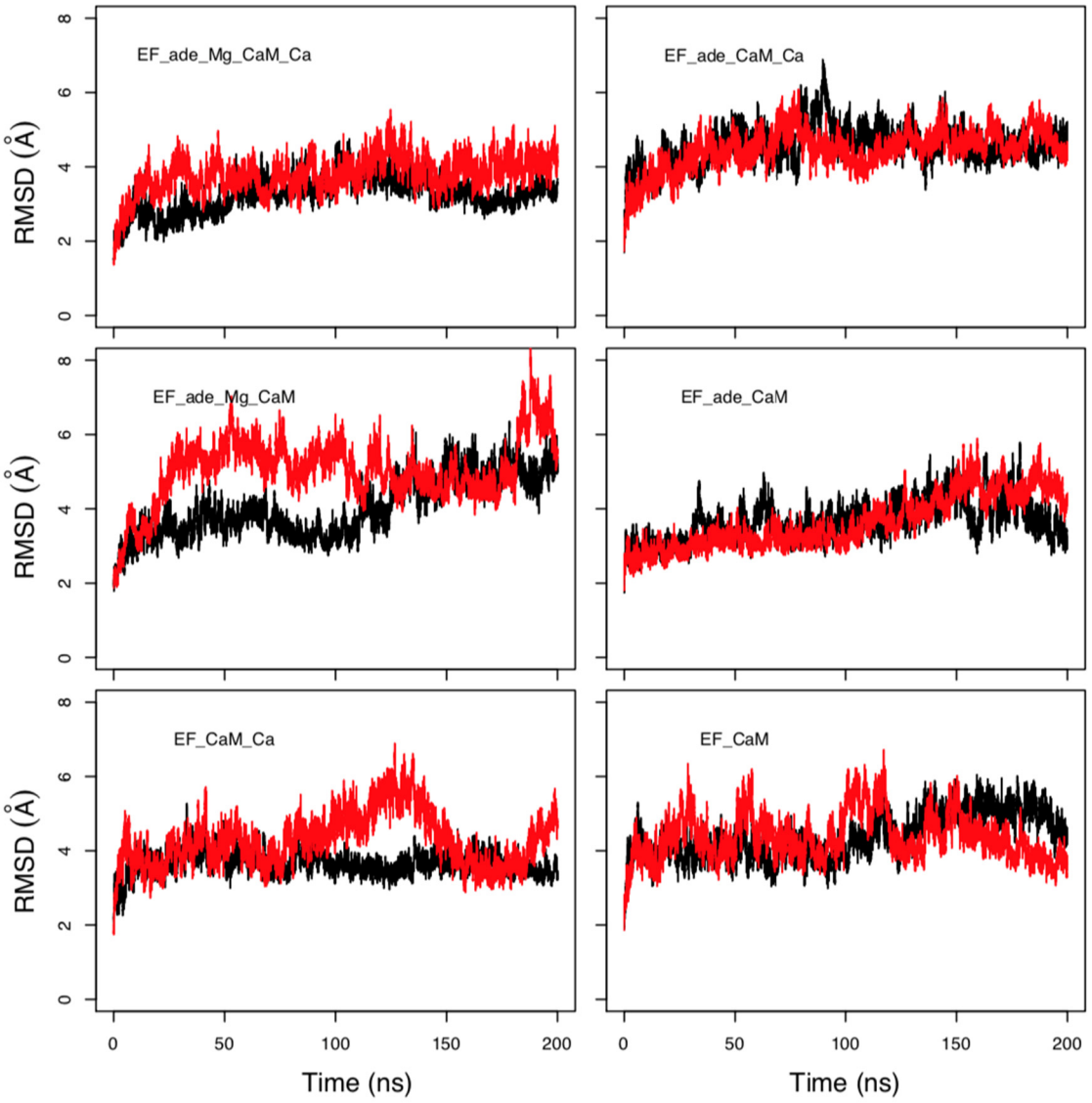


Figure 85- Root-mean-square deviations (RMSD) of the backbone heavy atoms of EF and CaM with respect to the PDB structure 1PK0, *calculated along all trajectories in Table 1. For each trajectory, the black and red curves correspond to the two replicas R1 and R2.*

The coordinate root-mean-square fluctuations (RMSF) (Figure 86) display similar profiles for all trajectories. A peak of fluctuations is observed for the loop (residues 675-695) located between helices L and M[29] in the helical domain close to the calcic loop of the EF-hand 1 of N-CaM. This loop was disordered and thus not visible in the initial 1PK0 structure. Globally, the N-CaM region fluctuates more than the C-CaM region, loaded with Ca^{2+} ions.

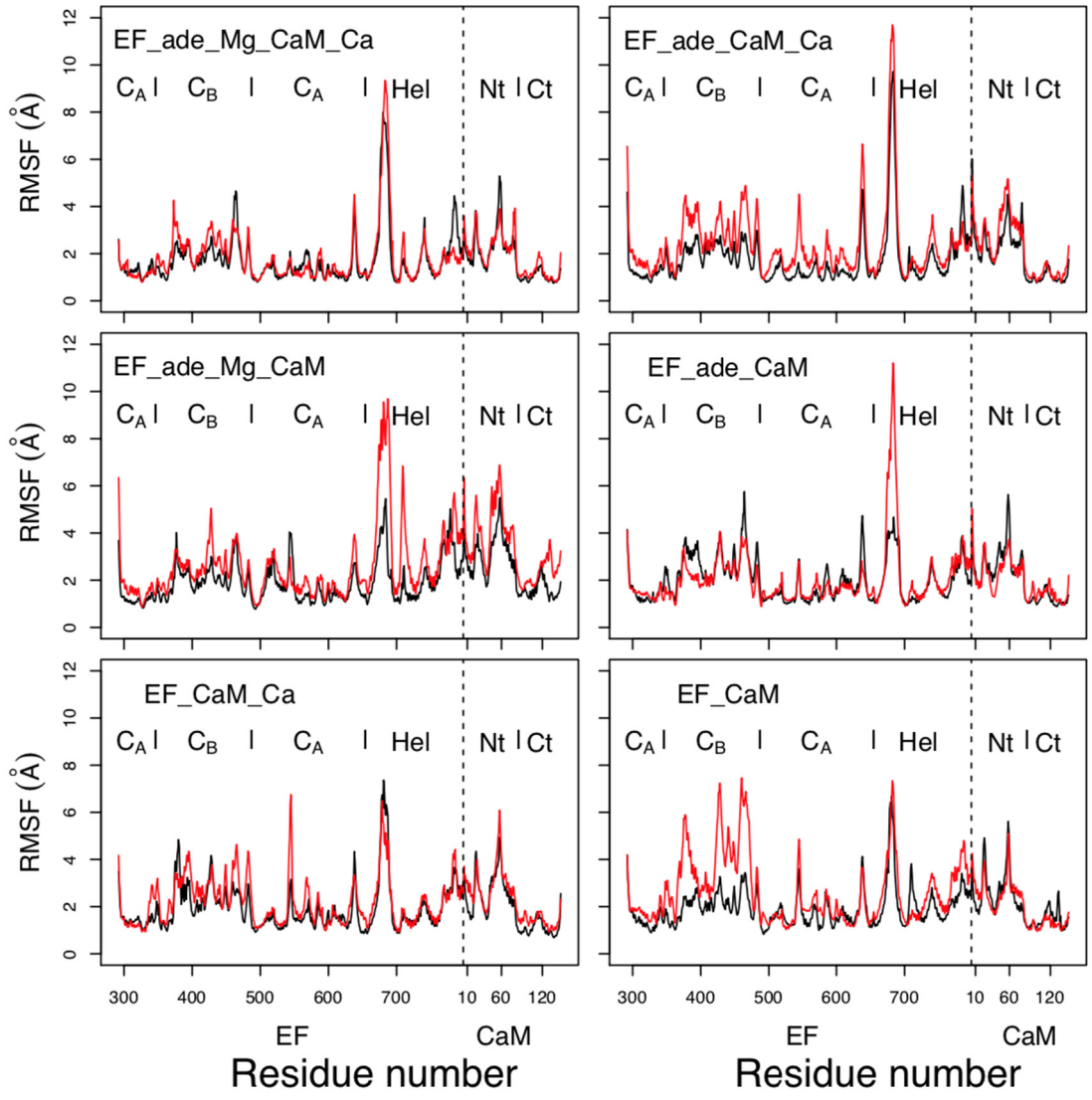


Figure 86-Coordinate root-mean-square fluctuations (RMSF) of the backbone heavy atoms of the complex EF-CaM, *each frame being superimposed on the chains A and D of the PDB ID 1PK0. For each trajectory, the black and red curves correspond to the two replicas. The N terminal residue of CaM is indicated by a vertical dashed line, the EF and CaM regions are labeled on the top of each plot. The N-CaM and C-CaM regions are labeled Nt and Ct for sake of clarity.*

Interestingly, the removal of ions does not increase much the mobility provided that the ligand adefovir is still present (trajectory EF_ade_CaM). By contrast, the removal of one type of ion in the presence of the ligand (trajectories EF_ade_CaM_Ca and EF ade Mg CaM) or of the ions and the ligand (trajectory EF-CaM) increases the internal mobility of EF and produces a shift between the two replicas. The trajectory EF_CaM_Ca, corresponding to the active toxin ready to interact with adefovir, has an internal mobility similar to EF_ade_Mg_CaM_Ca. The similarity of internal mobility for these two EF-CaM complexes are in agreement with their functional correspondence.

The global shape of the EF-CaM complex was analyzed by monitoring the gyration radius of the complex as well as the bending angle of the central α helix of CaM (Figure 87). The gyration radius mainly samples values in the 19-22 Å range, with the exception of EF_ade_Mg_CaM in which the gyration radius varies in the 18-21 Å range (Figure 87, x-axis). The removal of magnesium (EF_ade_CaM_Ca) or calcium (EF_ade_Mg_CaM) ions in the presence of adefovir induces an increase in the range of variations. By contrast, removal of the ligand (trajectories EF CaM) allows to keep a narrow distribution of values, but produces a shift in the gyration radius toward smaller values, around 19 Å. Thus, both ligand and ions have an influence on the complex expansion.

The bending of the central α helix of CaM, monitored as the angle between axes of α helices IV and V (Figure 87, y-axis), is relatively stable around 130° for the full system EF_ade_Mg_CaM_Ca as well as for EF_CaM_Ca and EF CaM. However, the systems EF_ade_CaM and EF ade Mg CaM, in which the adefovir is present and the calcium ions are removed, show large drifts of angle towards smaller angles down to 100° or larger angles up to 140°. Furthermore, the 2D contour plot of the gyration radius in function of the bending angle reveals a strong correlation between their variations for systems EF_ade_CaM_Ca and EF_ade_Mg_CaM (Figure 87).

The drifts in the bending angle of the central α helix is a sign of a destabilization of the EF-CaM complex. Indeed, the analysis of global CaM conformation along MD simulations[111] revealed that calmodulin acts as a spring that maintains EF in an open conformation.

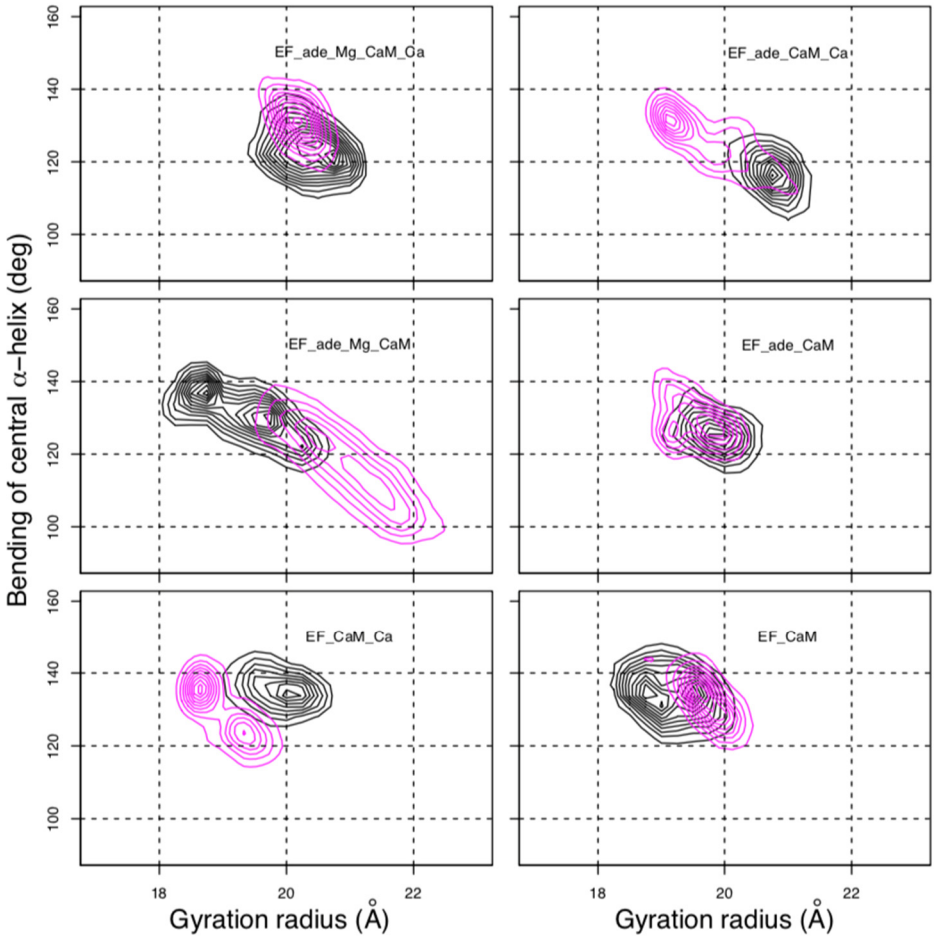


Figure 87- Contour plots describing the variation of the gyration radius with respect to the bending angle of the CaM central α-helix *determined as the angles between axes of α helices IV (residues 66-73) and V (residues 83-92). The contour line describes the joint probability distribution of these two parameters. For each trajectory, the black and magenta contours correspond to replicas R1 and R2.*

Overall, the removal of ions induces perturbations in the EF-CaM complex. These perturbations are visible in the atomic fluctuations as well as in the correlated variations of gyration radius and CaM central α helix bending. Hence, the bending of CaM central α helix, which describes a local movement, is connected to the gyration radius, which instead describes the overall shape of the complex. This connection highlights a long-distance effect.

D.IV.2 CaM conformation in the EF-CaM complex conserves features of the isolated CaM

In the previous section, the removal of calcium ions bound to CaM has resulted in the destabilization of the EF-CaM complex. CaM is an extremely flexible protein[310,311], its conformation being strongly modulated by the loading of calcium ions[98,313,314]. This allows

CaM to bind various target proteins involved in the signaling processes[102]. In the present section, the CaM conformation was analyzed and compared to the literature[315] according to various structural parameters in order to assess the fitting of CaM to the interaction with EF.

In CaM, the EF-hand domains are helix-loop-helix motifs responsible for calcium binding. The calcium ions bound to the C-CaM lobe are coordinated by the carbonyl oxygen of residue Y99 and side-chain carbonyl groups of residues D93, D95 and E104 in the EF-hand 3, and by the carbonyl oxygen of residue Q135 and side-chain carbonyl groups of residues D131, D133 and E140 in the EF-hand 4. The average coordination distances reveal that the ions keep similar coordination geometry along all trajectories (Table 12). Side-chain atoms Oδ from residues D93, D95 and residues D131, D133 showed weaker coordination than the other coordinated residues.

Calcium	CaM atom	EF_ade_Mg_CaM_Ca	EF_ade_CaM_Ca	EF_CaM_Ca
Cal1	O-Y^{99}	2.3 ± 0.1 2.3 ± 0.1	2.3 ± 0.1 2.3 ± 0.1	2.3 ± 0.1 2.3 ± 0.1
Cal1	Oδ1-D^{93}	3.2 ± 0.84 2.1 ± 0.05	2.1 ± 0.05 2.1 ± 0.05	2.1 ± 0.1 2.1 ± 0.05
Cal1	Oδ2-D^{93}	3.8 ± 0.2 2.7 ± 0.8	3.8 ± 0.2 3.8 ± 0.2	3.6 ± 0.2 3.6 ± 0.2
Cal1	Oδ1-D^{95}	3.7 ± 0.2 3.8 ± 0.2	3.7 ± 0.2 3.7 ± 0.2	2.2 ± 0.1 2.2 ± 0.1
Cal1	Oδ2-D^{95}	2.1 ± 0.05 2.1 ± 0.05	2.1 ± 0.05 2.1 ± 0.05	3.6 ± 0.1 3.6 ± 0.1
Cal1	Oϵ1-E^{104}	2.2 ± 0.1 2.3 ± 0.1	2.2 ± 0.1 2.2 ± 0.1	2.2 ± 0.1 2.2 ± 0.1
Cal1	Oϵ2-E^{104}	2.2 ± 0.1 2.2 ± 0.1	2.2 ± 0.1 2.2 ± 0.1	2.3 ± 0.1 2.3 ± 0.1
Cal2	O-Q^{135}	2.3 ± 0.1 2.3 ± 0.1	2.4 ± 0.1 2.4 ± 0.1	2.3 ± 0.1 2.3 ± 0.1
Cal2	Oδ1-D^{131}	2.1 ± 0.05 2.1 ± 0.05	2.1 ± 0.05 2.1 ± 0.05	2.1 ± 0.05 2.1 ± 0.05
Cal2	Oδ2-D^{131}	3.7 ± 0.2 3.6 ± 0.2	3.7 ± 0.2 3.7 ± 0.2	3.8 ± 0.2 3.8 ± 0.2
Cal2	Oδ1-D^{133}	2.2 ± 0.1 2.2 ± 0.1	2.2 ± 0.1 2.2 ± 0.1	2.1 ± 0.1 2.1 ± 0.05
Cal2	Oδ2-D^{133}	3.6 ± 0.1 3.6 ± 0.1	3.6 ± 0.1 3.6 ± 0.1	3.8 ± 0.2 3.8 ± 0.2
Cal2	Oϵ1-E^{140}	2.2 ± 0.1 2.2 ± 0.1	2.2 ± 0.1 2.2 ± 0.1	2.2 ± 0.1 2.2 ± 0.1
Cal2	Oϵ2-E^{140}	2.3 ± 0.1 2.3 ± 0.1	2.3 ± 0.1 2.3 ± 0.1	2.3 ± 0.1 2.2 ± 0.1

Table 12- Average distances (Å) between calcium ions (labeled Cal1 and Cal2) located in CaM EF-hands 3 and 4 and atoms of C-CaM residues. *The two lines of each table cell correspond to the two replicas recorded for each trajectory.*

It is well known from the literature that the presence of calcium ions has a strong influence on the conformation of the isolated CaM[98,313]. The angles between the α helices of

EF-hands in CaM increase, as well as the accessible surfaces of hydrophobic patches, upon calcium loading[98,313]. In addition, in the absence of calcium ions, the central α helix is more disordered[97,314]. These variations allow a better interaction of CaM with peptides involved in calcium signaling[102,315]. Using the definition of the hydrophobic patches the residues previously listed in Materials and Methods section C.9, the surfaces of hydrophobic patches of N- CaM and C-CaM (Figure 88) display quite different trends. The C-CaM patch, which corresponds to EF-hands loaded with ions Ca^{2+}, displays consistent profiles around 200 $Å^2$, with some replica displaying few jumps up to 900 $Å^2$. By contrast, the N-CaM patch shows much more diversity spanning a range of 200-600 $Å^2$. Similar trends have been observed for accessible surfaces of methionines in previous simulations[116], with a cumulative exposed surface of N-CaM Met of about 26 $Å^2$ in apo-form and 88 $Å^2$ with calcium, whereas the C-CaM methionines displayed cumulative exposed surfaces of 45 $Å^2$ in apo-form and 124 $Å^2$ with calcium (Table 4 in [116]).

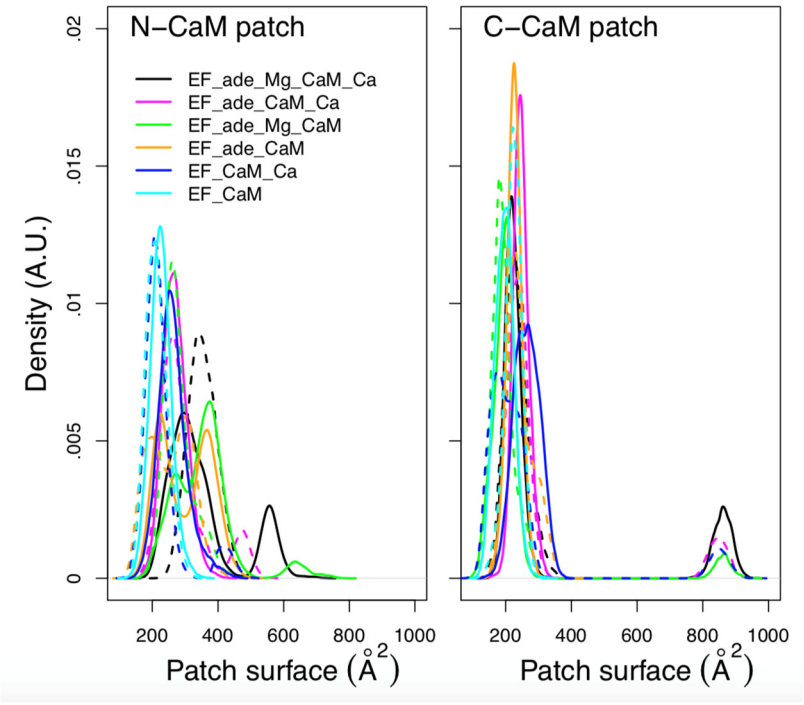


Figure 88- Distribution of accessible surface of the hydrophobic patch of N-CaM and C-CaM. *The N-CaM patch is defined by the residues A10, F12, A15, L18, F19, L32, M36, L39, M51, V55, M71, M72, M76. The C-CaM patch is defined by the residues I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, A147. For each trajectory, the plain and dashed curves correspond to replicas R1 and R2.*

In isolated CaM, the α helices in EF-hands display a tendency to move apart when CaM is loaded with calcium[98,313]. Similarly, in previous MD simulations of the EF-CaM complex[111], the C-CaM EF-hands, loaded with Ca^{2+}, display more open α helices than the N-CaM EF-hands.

Such behavior is also observed in our simulations (Figure 89). In C-CaM, the EF-hands 3 and 4, loaded with Ca^{2+} ions, display quite stable values around respectively 80 and 90°, corresponding to open conformations of EF-hands. The EF-hand 4 fluctuates slightly more than the EF-hand 3, especially for some trajectories in which calcium or Magnesium ions are absent (trajectories EF_ade_CaM and EF ade Mg CaM). By contrast, the angles of EF-hands 1 and 2, located in N-CaM, display much wider variations among trajectories. The angle of EF-hand 1 is located in the 40-80° range for most of the trajectories, corresponding to conformations of the hand oscillating between closed and semi-open configurations. The angle of EF-hand 2 displays the largest variations in the range of 20-80°. For EF_ade_Mg_CaM trajectories (green curves), the EF-hand 2 explores open conformations with angles larger than 80° and, for one replica of EF_ade_Mg_CaM_Ca (black curves), EF_ade_CaM (yellow curves) and EF_CaM (cyan curves), it displays equilibrium between closed and semi-open conformations. The removal of calcium ions from C-CaM thus induces a shift of the of N-CaM hands towards more open states.

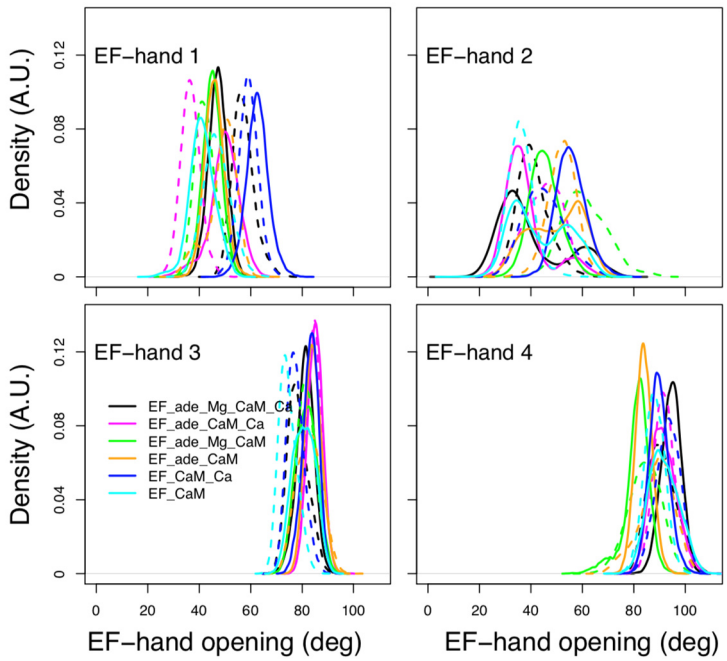


Figure 89- Angle of the EF-hands *(deg) calculated between axes of α helices including residues 8-19 (helix I), 31-37 (helix II), 46-53 (helix III), 66-73 (helix IV), 83-92 (helix V), 103-110 (helix VI), 119-127 (helix VII), 139-145 (helix VIII). The angle of EF-hand 1 is the angle between helices I and II. The angle of EF-hand 2 is the angle between helices III and IV. The angle of EF-hand 3 is the angle between helices V and VI. The angle of EF-hand 4 is the angle between helices VII and VIII.*

Overall, EF-hands in N-CaM, which are not loaded with calcium ions, display much more conformational heterogeneity and populate states corresponding to closed as well as to semi-open EF-hand. The global conformation of CaM as well as the calcium coordination is not strikingly modified among the trajectories recorded here.

D.IV.3 A network of amino-acid interactions connects the EF catalytic site with CaM

The edema factor activity has been investigated according to various conformational features: (i) the accessible surface of the catalytic site, which reports its functional state, (ii) the interface between CaM and EF which is directly related to the EF activation. Initial analysis of the X-ray crystallographic structure 1PK0 revealed a network of hydrogen bonds connecting adefovir and residues from the catalytic site with EF residues at the EF-CaM interface and residues from CaM (Figure 90, B).

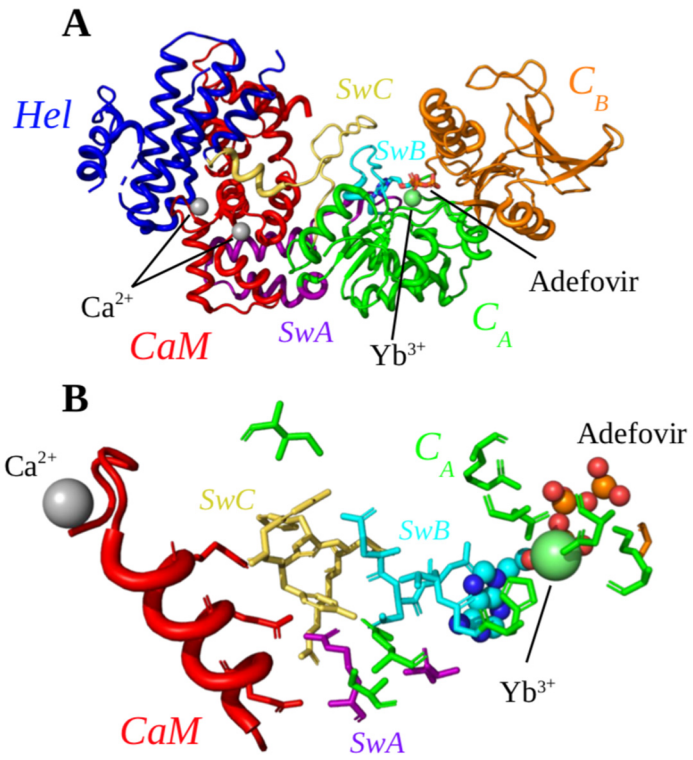


Figure 90- A. X-ray crystallographic structure (PDB ID 1PK0) of the complex EF-CaM *with the ions Ca^{2+} colored in gray and the ion Yb^{3+} colored in lime. The ligand adefovir is drawn in sticks, the regions C_A (residues 292349 and 490622), C_B (residues 350-489) and helical (residues 660-800) of EF are colored in green, orange and blue and CaM is colored in red. The switches A (residues 502-551), B*

(residues 578-591) and C (residues 630-659), labeled SwA, SwB and SwC, are colored in purple, cyan and yellow. B. The network of residues in the X-ray crystallographic structure (PDB ID 1PK0) of the complex EF-CaM connecting residues of the catalytic site to residues of the α helix V of CaM. The ligand adefovir is drawn in spheres. The helix V is drawn in carton. The residues of CaM, region C_A and of switches A, B and C are in the same colors as in A.

The accessible surface of the catalytic site of EF (Figure 91, left), calculated as the sum of solvent accessible surfaces of residues H351, K353, S354, K372, R329, K346, L348, D491, D493, H577, G578, T579, D582, N583, E588, F586, and T548, varies in the 400-1200 Å^2 range. This range of values agrees with the average catalytic surfaces previously observed in MD studies[111], as 928 Å^2 for the 2 calcium-loaded CaM, 866 Å^2 for the 4 calcium-loaded CaM and 501 Å^2 for the apo-CaM. Noticeably, in the presence of adefovir (black, yellow, magenta and green curves) the surfaces are smaller around 400-800 Å^2. This corresponds to a more closed catalytic site, in agreement with the inhibitory effect of adefovir. Removal of ion Mg^{2+} or removal of ions Ca^{2+} and Mg^{2+} in the presence of adefovir (magenta and yellow curves) induces a certain shift towards larger values, but the largest shifts towards the 800-1200 Å^2 range, is observed if adefovir is removed (cyan and blue curves).

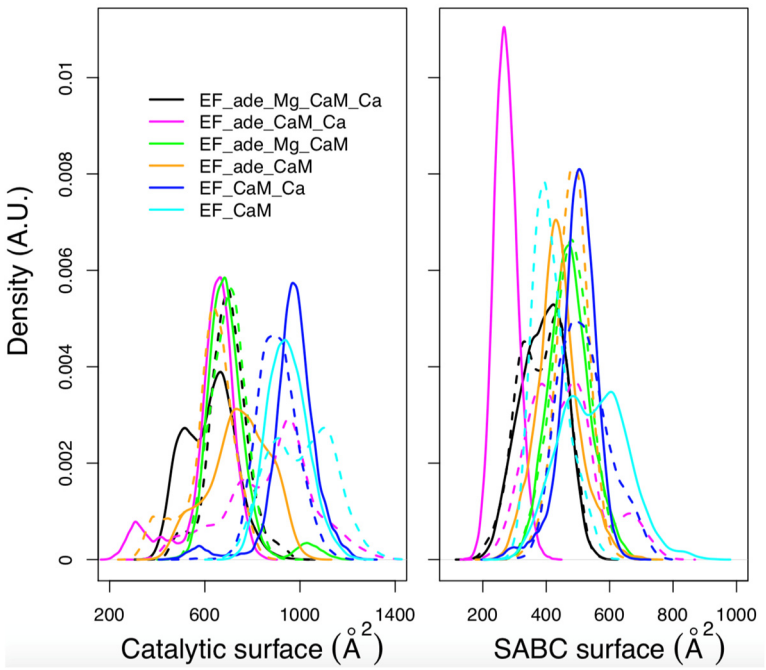


Figure 91- Distribution of accessible catalytic surface and the SABC pocket. *Left. Distribution of accessible catalytic surface (Å^2) of EF, calculated as the sum of the accessible surfaces of residues H351, K353, S354, K372, R329, K346, L348, D491, D493, H577, G578, T579, D582, N583, E588, F586, T548 of EF. Right. Distribution of accessible surface (Å^2) of the SABC cavity, calculated as the*

sum of the accessible surfaces of residues A496, P499, I538, E539, P542, S544, S550, W552, Q553, T579, Q581, L625, Y626, Y627, N629, and N709 of EF. For each trajectory, the plain and dashed curves correspond to replicas R1 and R2.

The interface between CaM and EF has been analyzed in the two following aspects: (i) the accessible surface of the SABC pocket and (ii) the variation of interactions along the residue network spanning from the catalytic site to CaM.

The SABC pocket, previously used for the virtual screening that leads to the discovery of thiophen ureidoacids[192], is formed by residues, A496, P499, I538, E539, P542, S544, S550, W552, Q553, T579, Q581, L625, Y626, Y627, N629, and N709, which belong to the three switches A, B, and C[29]. The accessible surface of the SABC pocket (Figure 91, right) varies in a much smaller range of 200-800 $Å^2$ than the catalytic pocket. Although the two pockets are defined by a similar number of residues, large differences are nonetheless observed for pocket SABC. This is a sign of significant reorganizations in this region, which results in some cases in the disappearance of a large accessible surface.

A network of interacting residues initially observed in the 1PK0 structure[127] (Figure 90B) displays hydrogen bond and stacking interactions which have been monitored along trajectories (Table 13). The network of contacts involves the residues T519, T548, Q553, G578, D582, N583 starting from the catalytic site and expanding to CaM. These residues are located in the switches A (residues 502-551), B (residues 578-591) and C (residues 630-659) (Figure 90B). Interestingly, in the X-ray structures of EF[29], this switch undergoes a major reorganization between the inactive state and the active state of EF.

Overall, the proportion of formed interactions strongly decreases as soon as ions are removed from the system. In particular, the interactions involving ion Mg^{2+} are strongly reduced. In the initial X-ray crystallographic structure (PDB ID 1PK0), the ion Yb^{3+} is penta-coordinated by three atoms from adefovir (O1-EMA, P3-EMA, O2-EMA) and two atoms from EF (Nε2-H577, O-Y492). In the MD trajectories, only the contact between Mg^{2+} and Nε2-H577 is stable along the trajectories EF_ade_CaM_Ca and EF_ade_Mg_CaM (Table 13) and only three contacts are still present in the trajectory EF_ade_Mg_CaM_Ca: the ones with O2-ade and Nε2-H577 at a significant level and the one with O-Y492 at a negligible frequency. This

loss of contacts could be due to the reduction of the charge and of the van der Waals radius between ions Yb^{3+} and Mg^{2+}.

Among the interactions between adefovir and protein present in the structure 1PK0, only three (hydrogen bonds ade-H6/O-T548, ade-H7/O-T579 and stacking ade/N583) are still present along the trajectory EF_ade_Mg_CaM_Ca. Among them, the stacking between the indole part of adefovir and the aliphatic part of N583 is the only one significantly present along all trajectories (Table 13). Nevertheless, as soon as the ions are removed from the system, the interaction frequency is reduced in one replica of EF_ade_Mg_CaM (B), EF_ade_CaM_Ca (C) and in all replica of EF_ade_CaM (D). The adefovir hydrogen bonds conserved with protein mainly involve protons of the NH group on the indole part. The contacts involving the phosphate group P1 in the structure 1PK0 are completely lost along all trajectories. This destabilization of the adefovir/protein contact and Magnesium contacts as soon as the ion Yb^{3+} is replaced by the ion Mg^{2+} could support an artifactual character of the structure 1PK0, in which the stability of adefovir in the catalytic site was obtained by the presence of the non-biological ion Yb^{3+}.

The hydrogen bond and stacking interactions involving EF and CaM residues are reduced between trajectories EF_ade_Mg_CaM_Ca (A) and EF_ade_Mg_CaM (B) and between trajectories EF_ade_CaM_Ca (C) and EF_ade_CaM (D), when the ions Ca^{2+} are removed from the system. Overall, the removal of Ca^{2+} ions has more influence to reduce the contact stability as the removal of Mg^{2+}. Indeed, six interactions decrease under a formation percentage of 10% if Ca^{2+} ions are removed, whereas no interaction decreases below this percentage in the absence of Mg^{2+}. Interestingly, this influence is visible for hydrogen bonds established between the side-chain guanidino group of R630 from EF and the side-chain carboxyl groups of residues E84 and E87 from CaM, for the hydrogen bond between the sidechains of R540 from EF and the carboxyl groups of E87 from CaM, and also for the stacking between F628 from EF and R90 from CaM. Noticeably, since the CaM residues E84, E87 and R90 are located at the C terminal part of the α helix V, just before the EF-hand 3 in C-CaM, the presence or absence of Ca^{2+} ions in this lobe has a direct effect on the interaction EF-CaM.

interaction type	A	B	C	D	E	F
Ion Mg^{2+}						
MG/O2-ade	100/32	-	0.0/0.0	-	-	-
MG/O-Y492	3.5/3.5	-	0.0/0.0	-	-	-
MG/Nϵ2-H577	100/100	-	100/100	-	-	-
adefovir/protein	A	B	C	D	E	F
H14-ade/Oδ-D493	0.7/2.8	0/0	0/0	0/0	-	-
P1-ade/Hζ-K346	0/0	0/0	0/0	0/0	-	-
O5-ade/sidechain-R329	1.3/0	2.1/0	4.6/0.5	0.1/4.4	-	-
H13-ade/sidechain-R329	0/8.9	0/0	0/0	0/0	-	-
H11-ade/Oδ-D491	1.5/0	0/0	0/0	0/0	-	-
H6-ade/O-T548	92/96	0/0	0/0	0/0	-	-
H6-ade/HN-T548	26/5	0/0	0/0	0/0	-	-
H7-ade/O-G578	12/4	0/0	0/0	0/0	-	-
H7-ade/O-T579	79/95	0/0	0/0	0/0	-	-
N3-ade/O-T548	0/0	0/0	0/0	0/0	-	-
Stacking ade/N583	92/99	50/96	88/23	2.9/23	-	-
EF/EF	A	B	C	D	E	F
Hγ1-T579/Oδ-D582	99/100	45/94	100/34	51/81	83/11	23/83
Oϵ1-Q553/HN-T579	35/52	26/19	75/56	17/0	0.7/6	34/0
K346-Hζ1/G352-O	88/92	99/94	0/0	34/60	6/1.4	0.9/0.2
Q553-Hϵ/D579-Oδ	60/97	36/44	100/22	21/49	4/3.3	12/15
Q553-Hϵ/S550-Oγ	17/3.4	2/2.5	58/11	1.4/0.5	8/8	3.7/6
Stacking R540/Y627	58/17	0.2/1.7	91/23	9/0.1	13/12	2.2/69
H577-Hδ1/G578-O	93/88	94/93	28/11	38/33	41/24	29/25
N581-Oϵ1/N629-HN	99/99	95/99	92/98	90/99	93/80	85/99
D582-Oδ/Y627-HH	77/86	27/19	91/22	37/52	63/30	53/46
N583-O/N629-Hδ	36/29	82/39	28/58	35/68	38/90	27/58
Stacking I615/Y626	70/89	60/49	80/72	63/76	75/73	79/59
Stacking Y626/I619	0/0	0/0	0/0	0/0	0/0	0/0
EF/CaM	A	B	C	D	E	F
R630-sidechain/E87-Oϵ	97/73	24/8	69/59	60/16	53/19	37/100
R630-sidechain/E84-Oϵ	69/29	21/2.4	40/34	20/33	7/4	2.5/98
sidechain-R540/Oϵ-E87	68/19	3.4/25	92/48	4.5/0	21/0.5	1.8/70
Stacking F628/R90	49/13	2.3/7	52/32	7/11	8/5	8/86

Table 13- Relevant interactions observed along MD trajectories connecting atoms from EF, CaM and adefovir (ade), as well as ions Ca^{2+} and Mg^{2+}. *The selected interactions were picked up in the initial X-ray crystallographic structure 1PK0. The percentage of formation corresponds to the percentage of frames for which the distance is smaller than 2.5 Å in the case of hydrogen bonds and smaller than 4.5 Å in the case of stacking interactions. The trajectory names given in the table correspond to A=EF_ade_Mg_CaM_Ca, B=EF ade Mg CaM, C=EF_ade_CaM_Ca, D=EF ade CaM, E=EF_CaM Ca, F=EF CaM.*

The analysis of the interface EF-CaM, as well as the interactions within the catalytic site, highlight the influence that the variation of system composition has on the EF propensity to form interactions and on the accessible surface of its catalytic site, impacting its ability to perform its enzymatic activity. In addition, a network of interactions between EF and CaM

residues supports the existence of an allosteric correlation between CaM and the EF catalytic site.

D.IV.4 Analysis of cavities deformation to detect allosteric pockets

In this section, we investigate the possibility to use MD simulations with the cavity tracking implemented in the software *mkgridXf*[216] to predict allosteric sites in the EF-CaM complex. This approach is motivated by the following reasons. As already quoted in the introduction, several approaches for prediction of allosteric sites are based on a measure of the deformation of the protein described by an elastic network under the docking of ligands[217,316] or described by a normal mode perturbation[317]. Also, a recent analysis of a large set of protein structures containing ligands showed[318] that the binding sites of allosteric ligands display larger deformations. Similarly, several bioinformatics approaches predict allosteric pockets as the ones on which ligand binding induces the largest variations in protein structures[217,308]. Allostery has been thus repeatedly associated with larger local or global deformability. Besides, cavity tracking of *mkgridXf*[216] along MD trajectories made it possible to correlate deformations of individual cavities to the principal components of the protein global motions[319]. The association of the observation made by Desdouits *et al.*[319] with the literature approaches conducted us to propose the analysis of cavities deformation obtained via cavity tracking as a possible new way for the prediction of allosteric sites. A systematic analysis and tracking of the cavities present in the complex were performed along trajectories using *mkgridXf*[216,319]. The cavities were determined by rolling probes as described in Materials and Methods section C.9. Each cavity was tracked along MD trajectories using a description based on a consensus list of protein atoms delineating the cavities[216].

From each trajectory, only proteins EF and CaM have been kept, water molecules, ions and adefovir being removed. One frame every 40 was kept over the time interval 120-200 ns of the trajectories EF_ade_CaM_Ca, EF ade CaM, EF_ade_Mg_CaM_Ca, EF ade Mg CaM, EF_CaM_Ca and EF CaM, and then concatenated. The two trajectory replicas series were then analyzed independently with *mkgridXf* in order to probe the reproducibility of the cavity analyses.

The deformations of protein cavities have been monitored through the variations of their volumes. The volumes of each cavity averaged along each trajectory are plotted as points along

the cavity index, the points being colored according to the trajectory (Figure 92). For most of the cavities, the volumes are smaller than 100 $Å^3$ for all trajectories. Few of them (indexes #5, #64, #83, #97, #130, #136, #140) display larger volumes up to 2500 $Å^3$ as well as large variations among the various trajectory conditions.

For both replicas, the cavity #140, which is located inside the catalytic site, is among the most variable cavities. Other very variable cavities are located at the interface between different CaM and EF regions. The CaM regions are the EF-hand 1 (#5, #130), the EF-hand 3 (#136), the EF-hand 4 (#97), the central α-helix (#64). The EF regions are the helical region (#5, #64, #83, #97, #130) and the region C_A (#136). Interestingly, the cavity located at the interface between the EF-hand 3 and the region C_A (#136) displays small volumes. According to the rational exposed at the beginning of the present section, the "variable cavities" quoted above and located at the interface between CaM and different EF regions can be considered as allosteric cavities.

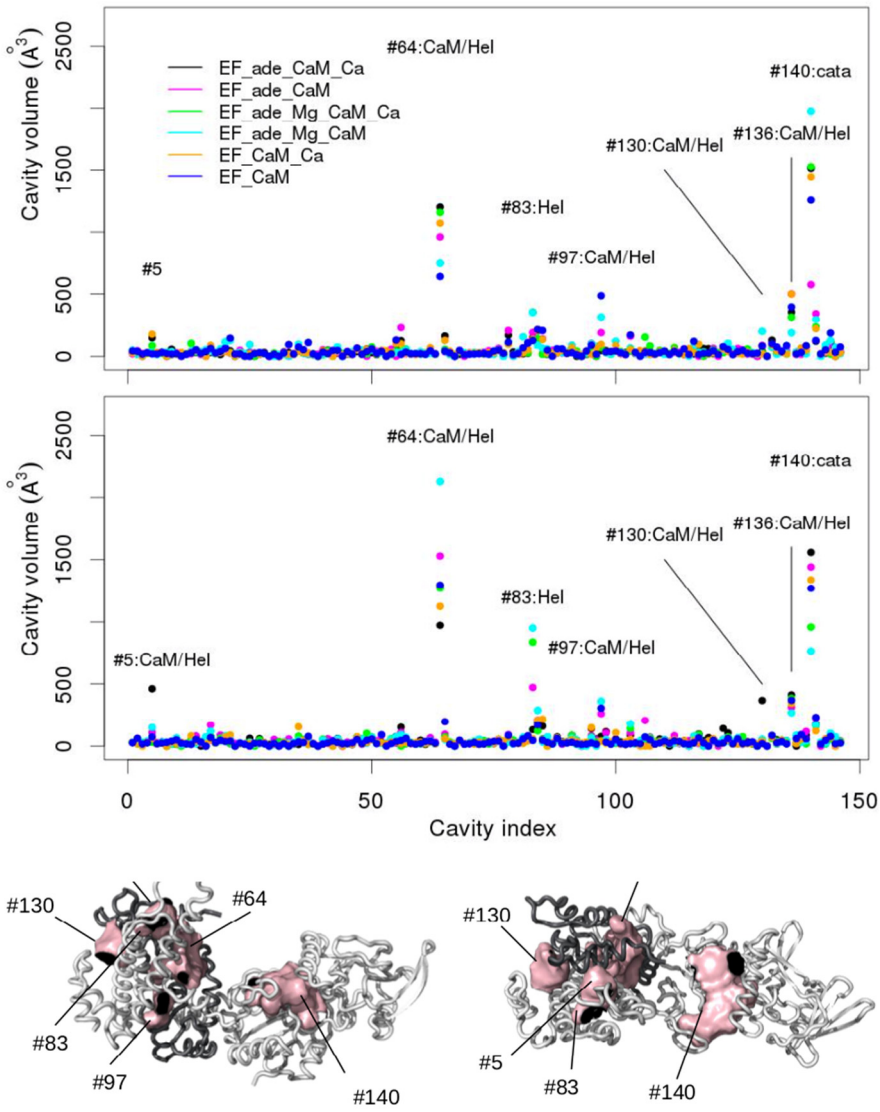


Figure 92- Averaged volumes of the cavities detected by mkgridXf plotted along the cavity index *(top). The two plots correspond to the two replicas series of the trajectories. The points are colored according to the trajectory on which the volume was averaged: EF_ade_CaM_Ca (black), EF_ade_CaM (magenta), EF_ade_Mg_CaM_Ca (green), EF_ade_Mg_CaM (cyan), EF_CaM_Ca (orange) and EF_CaM (blue). The cavities, for which at least one volume larger than 250 Å^3 has been observed, are labeled with the cavity number and annotated according to the cavity location: Hel=helical domain, cata=catalytic site, Hel/CaM=interaction interface between helical domain and CaM. (Bottom) Opposite views of the complex EF-CaM with CaM colored in dark grey. The cavities labeled on the plots are drawn in surfaces and colored in pink.*

In order to compare our approach to other approaches proposed in the literature to predict allosteric cavities, the three conformations of the EF-CaM complex extracted from the X-ray crystallographic structure 1PK0 were analyzed using the software PARS[217]. Before performing this analysis, all non-protein molecules have been removed from the structure. The centers of predicted sites (Figure 93) are located in the catalytic site and at the interface of CaM

and EF, in regions similar to the ones selected by detecting cavities along MD trajectories with *mkgridXf*. However, few other predicted pockets, which are marked by an asterisk in Figure 93, are observed in the region CB, which is far from the EF-CaM interface. The approach PARS predicts most of the allosteric cavities in regions similar to those predicted by *mkgridXf*. The additional predicted cavities may be due to the unique protein conformation analyzed by PARS, whereas the pockets detected by *mkgridXf* are obtained from the analysis of the MD trajectories.

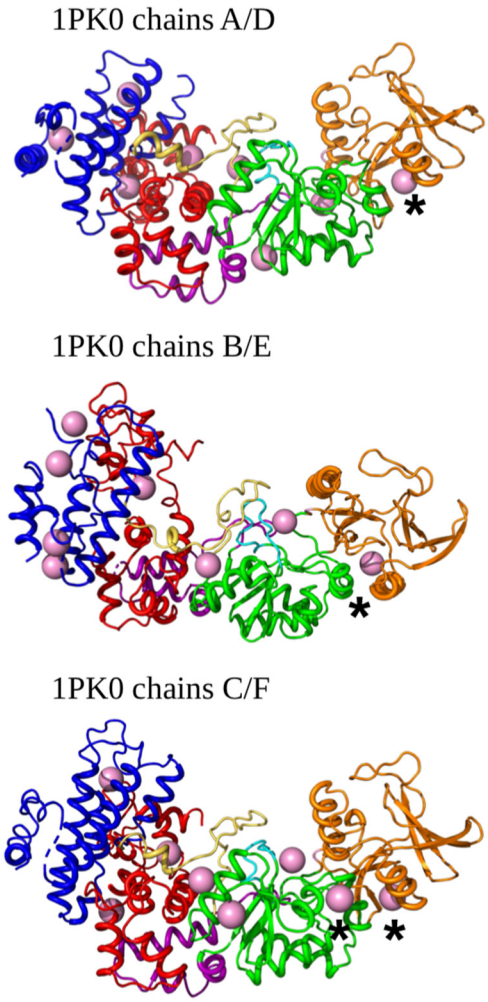


Figure 93- Allosteric sites predicted using PARS[217]*. The color code for protein structure is the same as in* ***Figure*** *90. The centers of predicted pockets are indicated using pink spheres. The ones not located in the catalytic site or at the EF-CaM interface are labeled with an asterisk.*

The cavities present in the EF-CaM complex were tracked along MD trajectories performed in various conditions related to functional aspects of EF activity. In this way, the EF-CaM interface has been pointed out as a region containing allosteric pockets. The targeting

of these pockets by virtual screening has higher chances to conduct to the discovery of allosteric ligands.

D.IV.5 TUA-diCl interacts with the EF-CaM interface

The ligand TUA-diCl was discovered by *in silico* targeting of the allosteric SABC pocket, but there are, at this time, no experimental evidences on the binding site of the molecule. In Chapter II, we discussed the possible binding mode of action and the binding site of the TUA-diCl. Another way of further investigating the mechanism of the small molecule is to study its interaction and the protein using MD simulations. Trajectories of 200 ns were recorded on the complex of the ligand TUA- diCl in interaction with the proteins (Materials and Methods section C.9, Table 2). The starting points of the simulation, which was chosen based on the consistency with the available experimental data, were defined in Chapter II. As the EF cavities 18 and 20 are very close, we selected the cavity 18 (ID 3) for the starting conformation for the MD simulations and cavity 4 for EF (ID 1) and the cavities 17 (ID 4), 36 (ID 5) and 14 (ID 6) for EF-CaM (Table 2). The intramolecular variations with respect to the initial structures of the systems are shown in Figure 94: the RMSD varies between 3 Å and 5 Å for the EF-CaM systems, while the RMSD values for EF, especially for ID 3, are higher. The RMSD of the ligand, which was computed after aligning the backbone atoms of the protein, was calculated along the trajectories to monitor the variation and the flexibility of the ligand (Figure 94, B and F). During the five simulations, the final positions of the ligand TUA-diCl are different from their initial positions (Figure 95), but the location of the ligand is most of the time close to its initial position. Different conformations have been explored by the ligand.

The RMSF profile for ID 1 is similar to ID 3 (Figure 94, D) and ID 4 to ID 5 and 6 (Figure 94, C). A fluctuation peak is observed for ID 1 and ID 3 in the helical domain of EF traducing the flexibility of this domain in presence of the ligand. On the contrary, for the complex EF-CaM in interaction with the ligand (ID 4, 5 and 6), the helical domain in these cases is less flexible. Interestingly, a second fluctuation peak is observed for ID 4 in the region of the SC loop. The SC loop is located close to the helical domain and this loop plays a critical role in the activation of EF by CaM. This loop stabilizes the interaction between EF and CaM.

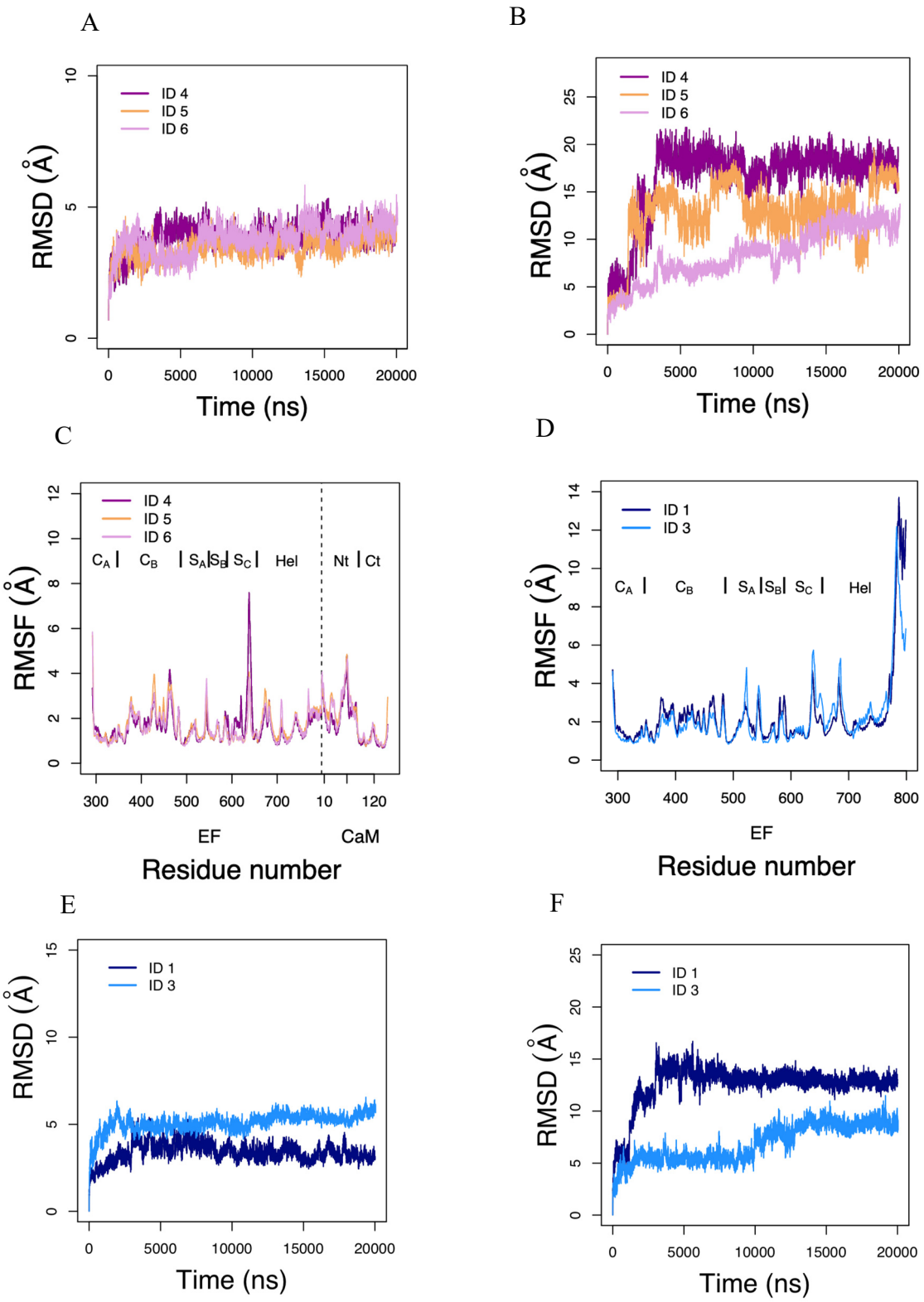


Figure 94- RMSD and RMSF *of the backbone heavy atoms of EF (E) and EF-CaM (A) with respect to the respective initial PDB and RMSD of the ligand TUA-diCl (B, F) with respect to the backbone of the protein EF (residue 291 to 780) (D) and EF-CaM (residue 292 to 799 for EF and 3 to 148 for CaM) (C), calculated along all trajectories. The different starting points of the ligand for dynamic simulations of TUA-diCl-EF-CaM are colored in the gradient of pink and in a blue gradient for TUA-diCl-EF. The ID 1 is colored in navy, ID 3 in dodger blue, ID 4 in dark magenta, ID 5 in sandy brown and ID 6 in plum.*

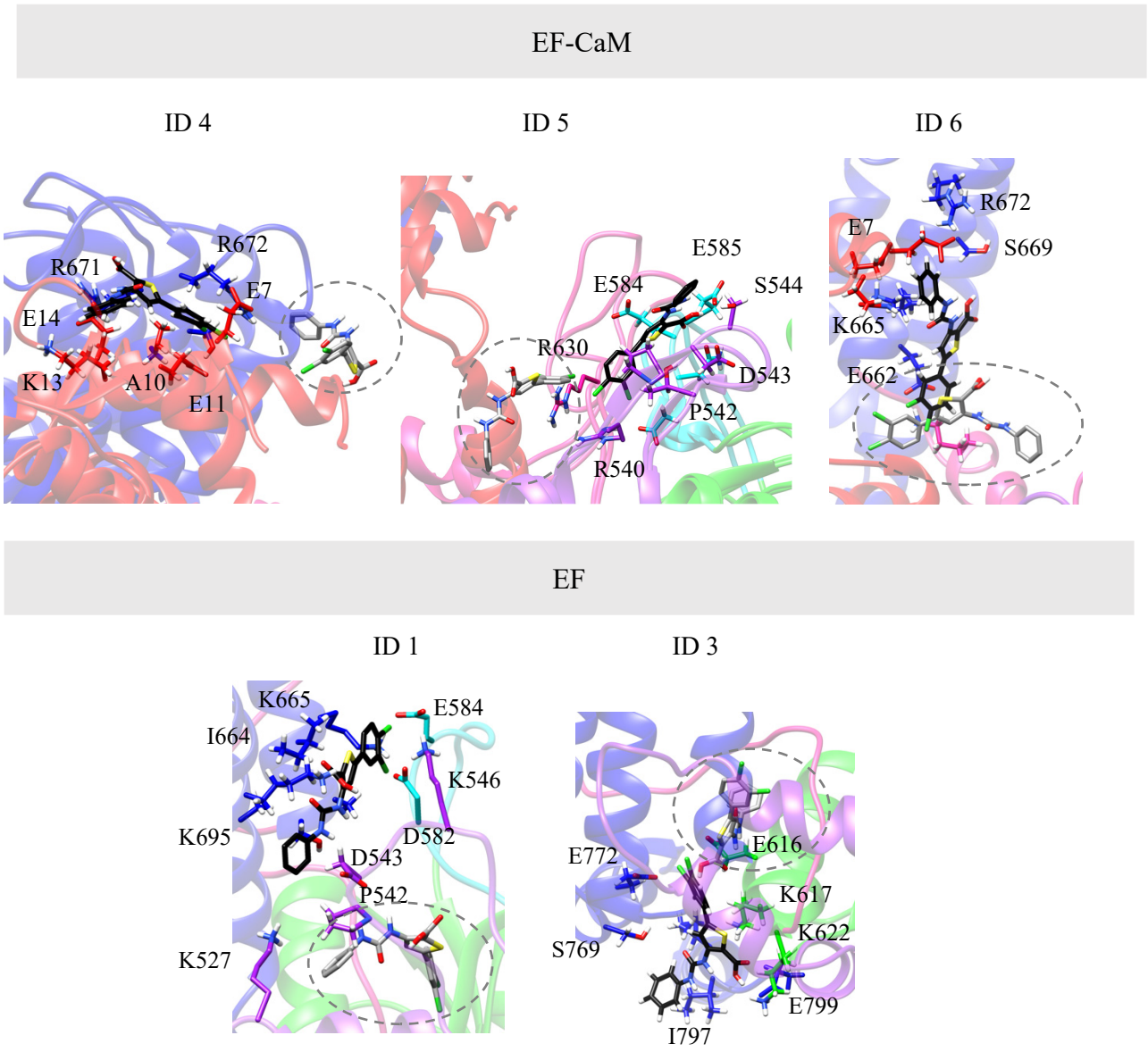
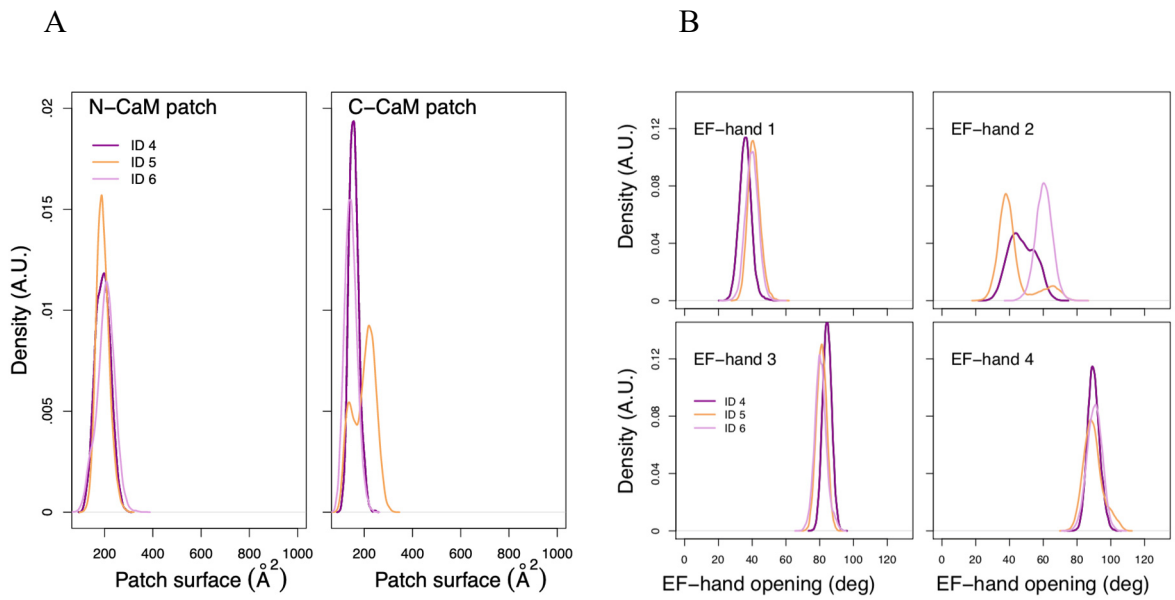


Figure 95- Starting and ending poses of the MD simulations. *The starting poses of the ligand TUA-diCl are marked with a dashed grey circle, whereas the ending poses are colored in black. These ending poses were found to be present between 50% and 75% of the time during MD simulations.*

D.IV.1 TUA-diCl destabilizes CaM within the complex EF-CaM

The conformation of CaM within the complex EF-CaM and the interaction between EF and CaM were assessed to understand whether the ligand is able to disrupt the interaction between the two proteins. The isolated CaM within the complex has particular features as previously described. In the three simulations, the C-CaM without the ligand is around 220 Å^2 and the N-CaM fluctuates between 200 Å^2 and 600 Å^2. The hydrophobic surface of the patches N-CaM and C-CaM in presence of the small molecule are both reduced (Figure 96). The N-CaM patch for the three simulations is superimposed (200 Å^2). The C-CaM profile is different

for the simulation ID 4, 6 and ID 5. For ID 5, two surfaces are observed: the former measures around 100 $Å^2$, which corresponds to the C-CaM patch of ID 4, 6 and the latter measures 220 $Å^2$, as in the case of the complex without the ligand. Differences were observed in the EF-hands of CaM with and without the ligand. As the presence of calcium is affecting the different angles on the EF-hands of CaM, we can determine the conformation of CaM within the complex in presence of the ligand. The EF-hands 3 and 4 in presence of the small molecule located to the C-CaM are in an open-conformation (around 80 to 90°, Figure 96) as in the holo EF-CaM complex. On the contrary, EF-hand 2 varies a lot as similarly described in the previous section. The EF-hand 2 is more affected by the position of the ligand than the other hands. The angle of EF-hand 2 adopts a more closed-conformation in ID 5 (around 40°) and a semi-opened conformation (around 60°) in ID 6, while it oscillates between the two conformations for ID 4 (around 40 to 60°). By comparing with the EF-hand opening angle of the N-CaM with and without the ligand, the angle in presence of the ligand displays a more closed-conformation. The N-CaM plays a different behavior in presence of the ligand (Figure 96).



C

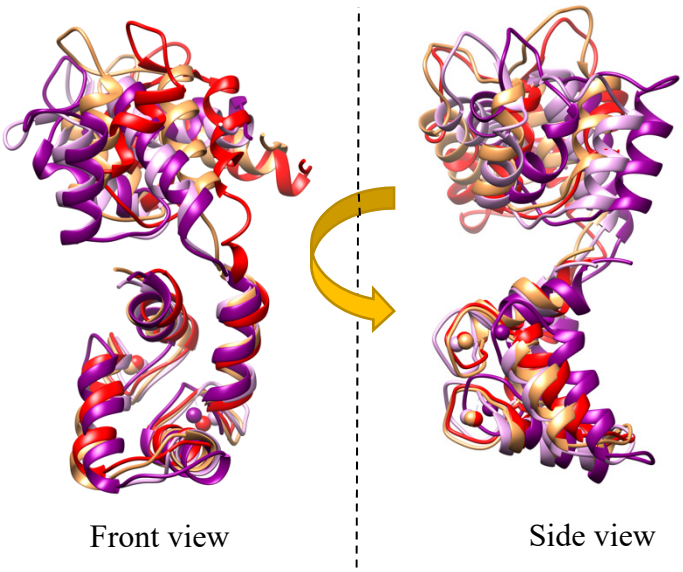


Figure 96- CaM features within the complex EF-CaM in presence of TUA-diCl. *In panel A, the distribution surface of the hydrophobic patches of CaM with the different starting points of TUA-diCl is represented, while in panel B the distribution of the EF-hands of CaM. The panel C illustrates the superimposition of one representative conformation of CaM with the two calcium during the three MD simulations (ID 4 in dark magenta, ID 5 in sandy brown and ID 6 in plum) and the conformation of CaM in the X-ray structure 1XFY (in red). The N-terminal of CaM is highly impacted by the presence of the TUA-diCl and the two Ca^{2+} within the EF-hand of C-CaM are displaced.*

D.IV.2 TUA-diCl destabilizes the network interaction connecting EF to CaM

The previous described hydrogen networks between EF and CaM are destabilized in the presence with the small molecule in the ID 4 and ID 6 simulation. The percentage of interactions between EF and CaM previously analyzed (Table 14) is globally decreasing in presence of the ligand TUA-diCl, as we have observed when the ions were removed (Table 13). The loss of connection is due to the interaction of TUA-diCl with the N-CaM, C-CaM, the helical domain and the loop region of EF. ID 6 is interacting with the residues 633-634, 642-669 of EF and 8, 82-86, 138-139 of CaM and ID 5 with the residues 538-542, 582-584, 630 of EF and 84 of CaM. The ligand in ID 4 interacts strongly with residues 665-672 in the helical domain and the residues 7-14 of the N-CaM. ID 4 and ID 5 by interacting with the key residues 630 of EF and 84 of CaM destabilize the allosteric network of EF and CaM. Destabilizing the hydrogen network between EF and CaM by a small molecule would destabilize the complex in an allosteric way.

EF/EF	ID 4	ID 5	ID 6
Hγ1-T579/Oδ-D582	74.95	26.94	55.50
O1-Q553/HN-T579	0.14	9.08	15.92
K346-Hζ1/G352-O	95.04	71.39	89.19
Q553-H/D582-Oδ	3.69	27.83	11.89
Q553-H/S550-Oγ	2.54	1.04	7.05
Stacking R540/Y627	13.66	94.86	21.16
H577-Hδ1/G578-O	70.56	69.62	72.92
N581-O1/N629-HN	87.06	97.5	98.75
D582-Oδ/Y627-HH	62.40	52.89	100.00
N583-O/N629-Hδ	27.21	31.55	68.43
Stacking I615/Y626	16.93	24.32	39.92
Stacking Y626/I619	2.67	0.01	0.00
EF/CaM	ID 4	ID 5	ID 6
R630-sidechain/E87-O	30.60	53.96	32.11
R630-sidechain/E84-O	6.50	47.30	24.20
sidechain-R540/O-E87	4.20	45.00	13.10
Stacking F628/R90	9.66	81.24	27.32

Table 14- The percentage of formation of the EF-CaM hydrogen network in presence of TUA-diCl

The surface of the catalytic site of ID 4, 5 and 6 is around 920 $Å^2$ (Figure 97) similar to the complex loaded with two calcium without any ligand of the catalytic site. Moreover, the accessible surface of the catalytic site pocket in the MD with EF is higher than in the complex (1200 $Å^2$), as the catalytic site is an open-conformation. The SABC pocket is found to be in a different conformation. The accessible surface of the cavity SABC in the case of ID 5 is reduced to 360 $Å^2$ and increased in the case of the ID 4 to 500 $Å^2$ and ID 6 to 550 $Å^2$. For EF the accessible surface of SABC is higher for ID 3 around 950 $Å^2$ and around 600 $Å^2$ for ID 1. The surface of the cavity is increasing when the ligand interacts only with the loop SC (ID 3 and 6) and decreasing when the ligand interacts with SA, SB and SC (ID 1 and 5).

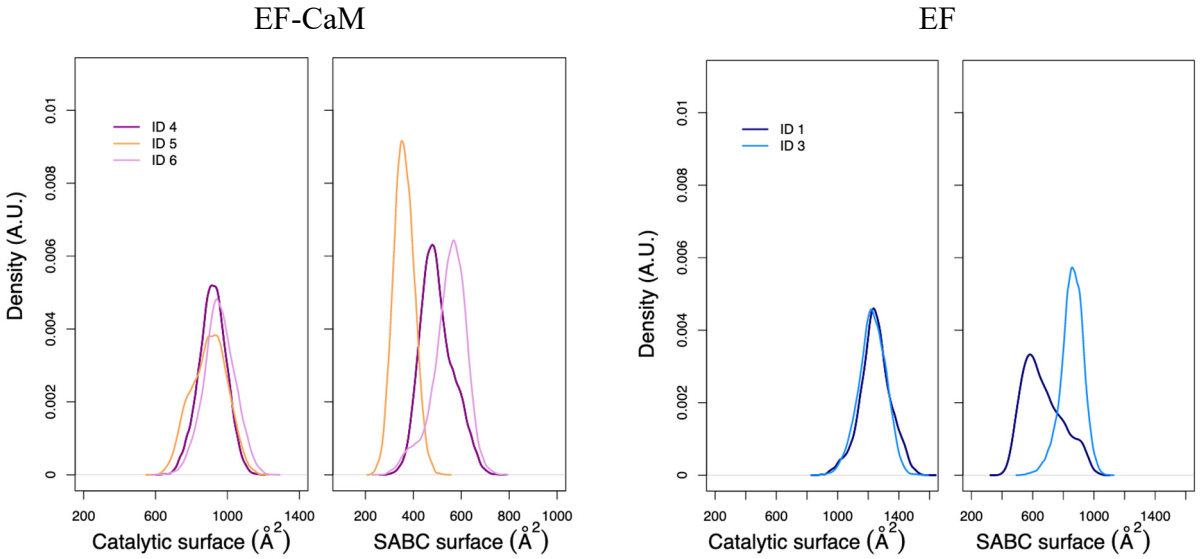


Figure 97- Distribution of accessible catalytic surface and the SABC pocket *(Å²) of EF, calculated as mentioned previously. On the left, the two surfaces are calculated on EF for the TUA-diCl-EF-CaM system and on the right for the TUA-diCl-EF system.*

The STD_{th} ratios were calculated along the trajectories as explained in Materials and Methods section C.8. The positions of the ligand with a RMSE < 0.2 and present over 70% of the corresponding frames were extracted and presented in

Figure 98 for the complex EF-CaM and in Figure 99 for EF. TUA-diCl in the simulation ID 4 was interacting with the residue E7, A10 and E11 of the N-CaM. TUA-diCl interacted with the C-CaM in the simulation ID 6 with the residues R86, R90, Y138, E139 and in the simulation ID 4 with the residues S81 and E84. The pocket of ID 5 was the closest among the ID 4, 5 and 6 to the SABC pocket sharing the residues I538, E539, P542 of EF. By projecting the residues of the interaction of TUA-diCl with the complex EF-CaM on the X-ray structure of EF alone, we observed that all of the helical domain was covered by the three molecules (ID 4, 5, 6, Figure 98). We did the same analysis for the TUA-diCl in interaction with EF and we highlighted the residues of EF in interaction with TUA-diCl on the X-ray structure of the complex EF-CaM. Interestingly, the residues forming the pocket of ID 1 and 3 in EF were split by the insertion of CaM on each side of CaM. Once the complex EF-CaM was formed, theses pockets disappeared. The residue E539, P542, Q581, Y627 and N629 of SABC were involved in the TUA-diCl interaction with EF (ID 1, Figure 99). By analyzing the two closest pockets to SABC, we noticed that the ID 5 (TUA-diCl-EF-CaM) and ID 1 (TUA-diCl-EF) shared two residues in common the residue E539 and P542 which were present in SABC.

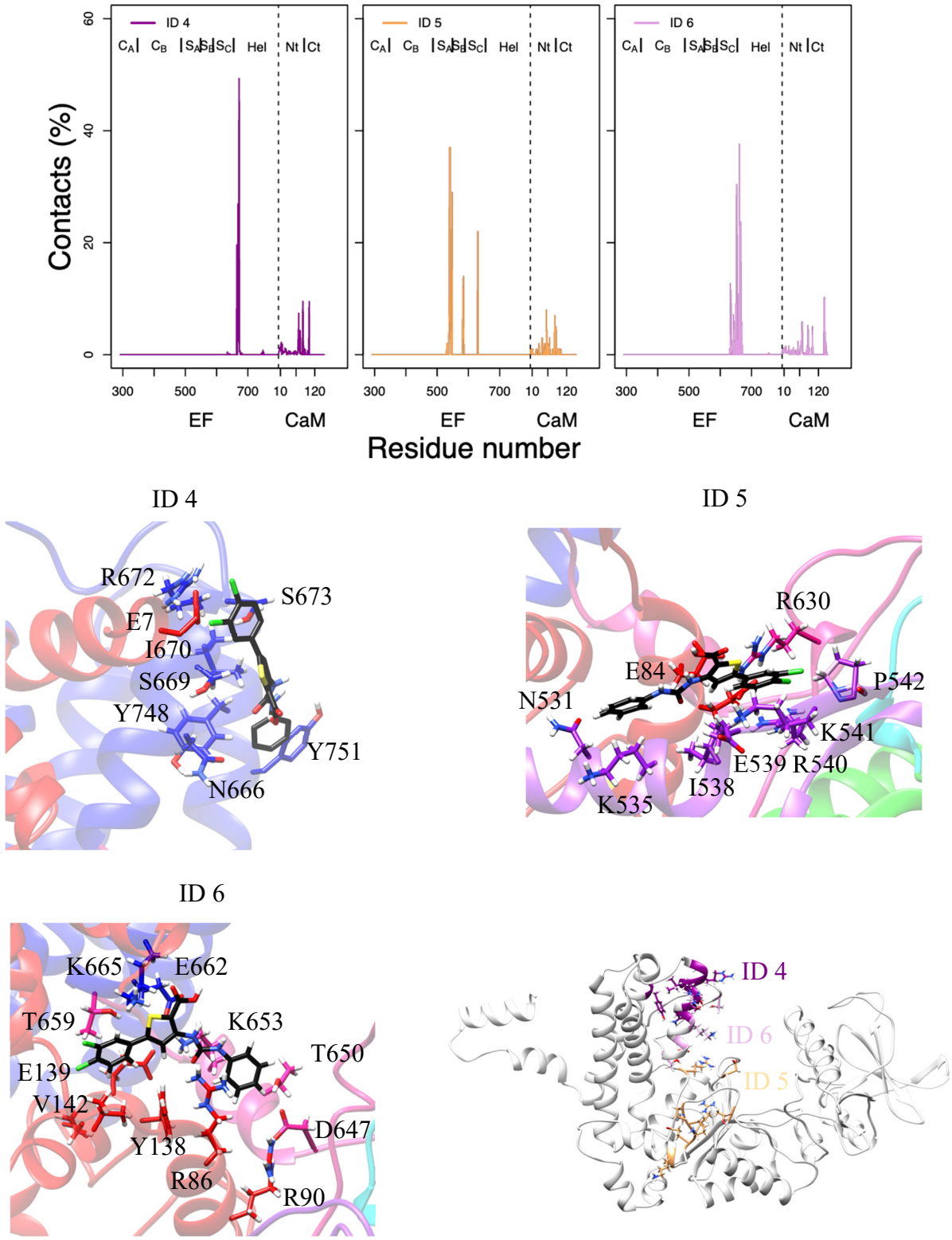


Figure 98- Contacts between the TUA-diCl and the complex EF-CaM. *The contacts of 4 Å between the TUA-diCl and EF-CaM were calculated along the trajectory. On the top of the figure, the percentages of formation of contacts are plotted in function of the residues number of EF-CaM along the trajectory. Below the contact map, the extracted poses of TUA-diCl that best agree with the* STD_{exp} *are represented. The CaM is represented in red, calcium in yellow, the helical domain of EF in blue, the* C_A *domain in green, the loop switch A in purple, the switch B in cyan and the switch C in deep pink.*

The residues in contact with the TUA-diCl (black) are represented in stick colored by domain. The EF-CaM residues in interaction with the TUA-diCl are plotted in the X-ray structure in the closed conformation of EF, in dark magenta for ID 4, in sandy brown for ID 5 and in plum for ID 6.

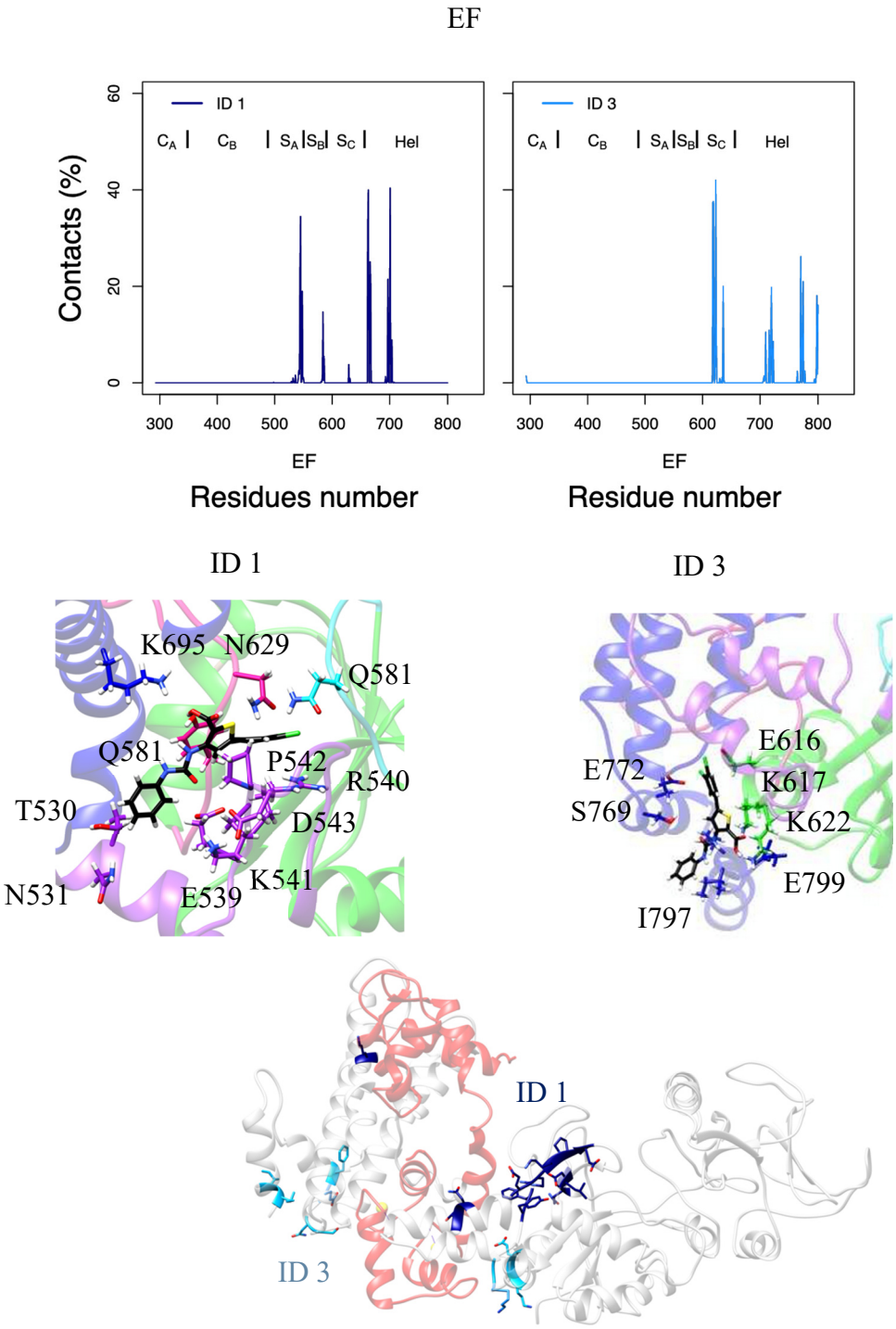


Figure 99- Contacts between the TUA-diCl and EF. *The description is the same as the previous figure. The helical domain of EF is colored in blue the CA domain is colored in green, the SA in purple, the SB in cyan and switch C is colored in deep pink. At the bottom of the figure, the residues in interaction with TUA-diCl of the closed conformation of EF are represented on the X-ray structure of the complex EF-CaM (white-red), in navy for ID 1, in dodger blue for ID 3.*

DISCUSSION

In the present work, MD trajectories starting from the X-ray crystallographic structure 1PK0 of EF-CaM complex bound to the inhibitor adefovir, have been performed in different conditions: in the presence and absence of co-factors, ions and ligand.

The most obvious, yet striking observation arising from the comparison of the trajectories is that the absence of ions has a strong effect on the complex dynamics, not only locally but also at long distance. Indeed, the removal of Mg^{2+} ion destabilizes the interactions between EF and adefovir, but also affects contacts between EF and CaM. Similarly, the removal of Ca^{2+} ions destabilizes the interaction of EF with CaM, but also the geometry of the catalytic site and the EF/adefovir interaction even in the presence of ion Mg^{2+}. This distant influence of ions applies through a network of hydrogen bonds and stacking interactions connecting the catalytic pocket with the EF-CaM interface. An allosteric influence of the EF-CaM complex to the catalytic pocket is thus highlighted. Noticeably a similar network of hydrogen bonds has been observed and validated using MD and mutagenesis[320] in the adenylyl cyclase (AC) toxin from *Bordetella pertussis*.

Another observation is that the replacement of the ion Yb^{3+}, observed in the initial X-ray crystallographic structure 1PK0[127], by the more biologically relevant Mg^{2+} ion induces a destabilization of numerous contacts between adefovir, ion and residues of the catalytic pocket. Consequently, the establishment of interactions due to the presence of Yb^{3+} ion could strengthen the binding of adefovir inhibitor to the catalytic site or induce the specific conformation of adefovir in the site. Previous computational analyses already highlighted the artifactual character of some ions observed in the catalytic site of various EF X-ray crystallographic structures[85,84].

The analysis of CaM conformations in the EF-CaM revealed that the removal of ions Ca^{2+} induces an unfitting of the conformation of CaM to its position in the complex as it is visible by the CaM central helix bending. The angles of EF-hands display also larger variations in N-CaM than in C-CaM, which might be related to a weakening of the EF/N- CaM interaction.

The tracking of cavities in the EF-CaM complex conformations extracted from the MD trajectories revealed large variations of cavity volumes in two regions: (i) the catalytic site and (ii) the interface between EF and CaM. The large deformability of cavities located at the interface between CaM and EF qualifies them as being related in an allosteric way to the catalytic site geometry and thus to the catalytic function of EF. Consequently, ligands designed

to bind such pockets could have an allosteric effect on the catalytic activity of EF. In that respect, one should note that a previously discovered inhibitor of EF, the compound 10506-2A, has been proposed to bind to the helical region[181].

The TUA-diCl compound is an example of ligand, which was demonstrated in Chapter II to inhibit the EF enzymatic activity in an allosteric way. The *in silico* docking approaches as well as the HDX-MS measurements described in Chapter II pointed out the EF helical domain as one of the preferred binding regions for TUA-diCl. Starting from the complexes between TUA-diCl and EF or EF/CaM obtained in Chapter II by *in silico* docking approaches, MD trajectories have been recorded. Although the positions of TUA-diCl are not very stable along these trajectories, we also observed that CaM, loop Switch C along with the hydrogen network with EF and CaM are destabilized, which in turn perturb at long-distance the catalytic site. In particular, the lobe N-CaM is displaced in the presence of TUA-diCl. In one of the trajectories (ID 1), some residues of the pocket in which TUA-diCl is bound, are part of the SABC pocket, which was initially targeted at the time of the discovery of TUA inhibitors. In another trajectory (ID 3), the ligand position is consistent with EF regions pointed by HDX-MS experiments. These two regions could be examined to improve the prediction of the TUA-diCl position on the EF or EF/CaM surface.

During the last decade, many approaches have been developed for detecting pockets susceptible to bind allosteric ligands, and allosteric paths through protein structures[217,316,317,321–327]. These approaches are mostly based on a graph description of protein structures. The graphs are then analyzed either from the point of view of protein rigidity and graph theory, or from a more physical point of view of normal modes or elastic network analysis. Many approaches require the presence of ligands in order to calculate their influence on the protein structure and thus to evaluate the allosteric effect of the ligand/cavity. In our analysis, we decided to focus on protein cavities: this approach avoids docking of ligands and is it thus independent from the choice of a specific ligand library. The relationship found here between the variability of *mkgridXf* cavity volumes and protein long distance communication is not surprising since such correlation has previously been observed[319] between *mkgridXf* cavities deformation and protein functional motions.

The use of cavities tracking along MD trajectories presents some advantages with respect to methods based on the modeling of protein via an elastic interaction network[217,316]. Indeed, by contrast with the network where only one atom (generally Cα) per residues is included in the calculation, the cavity calculation and tracking take information about all atoms into account, as well as their mutual interactions and their interaction with the solvent and co-

factors. Moreover, the model for internal dynamics of the complex is more realistic than the energy model of the elastic networks. Finally, the use of perturbation in ion binding highly involved in the EF function makes the observation of protein deformability more reliable as compared with the other approaches used in literature. However, all these improvements come at a higher computational cost.

E. Conclusion

The development of structural biology techniques during the last decades has revolutionized our understanding of cellular processes by providing a view at increasing resolution. The structural information that nowadays can be obtained has opened new avenues for modulating biological processes, in particular on the front of the rational design of drugs.

Among the possible strategies to modify the behavior of biologically relevant protein complexes with ligands, targeting the catalytic site of enzymes has attracted a lot of interest, as it permits to directly influence biological reactions. Nevertheless, this approach presents some disadvantages, as chemical inhibitors might display chemical similarity with the natural endogenous substrate of the target, and thus interfere in deleterious ways with the biological system, for example by inducing secondary effects. When targeting toxins with enzymatic activity, the possibilities of developing inhibitors are often hampered by the fact that the catalytic site can be highly conserved across species for enzymes sharing the same substrate and activity. The often-observed rigidity of the catalytic site can also difficult the development of inhibitors.

Consequently, targeting protein regions others than the catalytic sites becomes a valuable alternative, in particular for regions undergoing internal mobility for which the wealth of druggable pocket is enlarged, with the appearance of cryptic and allosteric pockets. One should notice that in this case, the conformational variability of the targeted site can reduce the entropic barrier of protein-ligand interaction compared to a rigid catalytic site. This reduction of the entropic barrier allows interactions with a smaller enthalpic component, which are easier to design. However, the affinity of the ligand can also be reduced.

The possible reduction of affinity mentioned above will have dramatic consequences on the experimental study of ligand-protein interactions. Indeed, this type of study strongly benefits from the assistance of experimental structural biology approaches. However, for most of these approaches, the acquisition of good quality data requires an average in time or in space of the experimental observables. In the case of partial disorder, averaging of the data over conformational ensembles might result in attenuation of the signal and consequently in reducing the available information. Thus, structural studies often face additional difficulties in this context.

This thesis originates from a molecular modeling and virtual screening study published in 2010 on the edema factor toxin (EF) of *B. anthracis*. After entering the host cell, EF is activated by interaction with the ubiquitous protein calmodulin (CaM) and acts then as a virulence factor by displaying an adenylyl cyclase enzymatic activity. This *in silico* study, which specifically targeted a pocket outside the catalytic site with the aim of proposing an allosteric inhibitor, conducted to the discovery of a series of compounds of the thiophen ureidoacids family (TUA) that inhibited the enzymatic activity of EF.

In this thesis, the mode of action of the TUA compounds on EF has been systematically investigated using a wide variety of biochemical and structural biology approaches including biochemical assays, NMR, crystallography, HDX-MS and molecular modeling. One major result of this work is that the main hypotheses presented in the initial study of 2010 were verified. The TUA compounds are indeed able to directly interact with EF and with the complex EF-CaM using an allosteric site. HDX-MS revealed that a region at the interface between the two partners of the complex (residues 704-728) could be the binding region of the ligand.

A surprising observation of this work is that the hydrophobic TUA ligand that we characterized (TUA-diCl) is able to interact with isolated CaM by binding to the hydrophobic patches that become accessible when CaM is loaded with calcium ions. Although, this interaction keeps CaM in a compact state that cannot activate EF, we showed that TUA-diCl inhibits the enzyme by a direct interaction with it. The observation that an inhibitor of a CaM dependent enzyme can actually bind to CaM is interesting as it could reveal a general trend of ligands supposedly inhibiting the function of proteins partners of CaM in the cell. This phenomenon was indeed observed for KN93, an inhibitor of the CaMKII kinase[183,275].

The interaction of EF with its activator CaM is very strong, with a dissociation constant of the order of tenths of nanomolar (depending on the conditions). At a first glance, it is thus surprising that a small ligand with a much lower affinity can inhibit the activity of EF without binding to the catalytic site. Our results show that TUA-diCl can bind to EF in the presence and absence of CaM, and by altering the EF conformation, hamper the activation by CaM and the formation of the catalytic site. The preliminary HDX analysis was pivotal to understand this phenomenon, as it revealed that TUA-diCl dramatically increases the solvent accessibility of EF, in particular of the helical and catalytic regions and hence destabilizes these regions of the protein.

This research does not stop here. In the near future, the HDX-MS data should be repeated to confirm the initial observations and to characterize the effect of TUA-diCl on the EF-CaM complex. If the observed unordering of EF were confirmed, techniques such as SAXS, coupled to techniques to measure the diffusion coefficients (analytical ultracentrifugation, diffusion ordered NMR spectroscopy [DOSY], …) should be used to analyze the conformational changes and possible dissociation of the EF-CaM complex produced by TUA-diCl.

Although the preliminary HDX-MS results are not encouraging in this sense, one could persevere in the attempts to obtain crystals of EF-CaM in presence of the ligand and detect its electronic density using soaking techniques. The system could also be studied by cryo-EM, both micro-ED or single-particle. By cryo-EM, in the absence of major conformational changes, the subtraction of the maps of the EF-CaM complex (~ 95 kDa, in the presence of the PABD domain of EF), with and without the ligand could lead to the localization of the binding site even if atomistic models were not achieved. The location of the binding site could then be confirmed by mutagenesis with the help of *in silico* approaches.

An interesting possibility to better understand the mechanism of action of the ligand and to localize the binding site is to use chemical shift perturbation methods on methyl labeled EF. This method has the advantage that it does not require the structure of the complex with the ligand. It is possible to ^{13}C label the protonated methyl groups of specific residues (Ile, Val, Leu, Met should be the most interesting) on a deuterated background of the protein. This labeling strategy allows one to obtain high quality ^{1}H-^{13}C correlation spectra of methyl groups for high molecular weight proteins like EF. A simple comparison of the spectra with and without ligand should identify the signals of the CH_3 groups involved in the interaction in the absence of major conformational changes or inform on important conformational changes as initially observed by HDX-MS. If no major conformational changes were detected (high numbers of affected CH_3 resonances), then the assignment of the CH_3 resonances could be done based on NOESY spectra and the structure of EF and allow one to identify the binding site. It should be noted that the assignment can be difficult and expensive.

Several TUA ligands with various substitutions inhibit EF-CaM. In addition, the compound 10506-2A contains chemical features analogous to those of TUA ligands. These chemical functions could be included in a QSAR analysis in order to propose new compounds

with a better affinity to EF or EF-CaM, while reducing/avoiding the interaction with CaM. Combining a virtual screening approach with these structural data could lead to discover new hits and further test them *in vitro*.

The original approach developed by A. Blondel in the laboratory lead to the discovery of a family of allosteric inhibitors of EF, which can be very valuable probes to understand the activation of the enzyme. Several molecular details of the mechanism of action of these compounds were highlighted in this Ph.D. work and revealed an allosteric region in the helical domain of EF that seems to be very important for the enzyme. This compound should also be useful to study the activation of EF *in cellulo* and importantly, it will be used to test if it can prevent the entry of EF into cells. Indeed, EF internalization requires its interaction with the protein PA, and this interaction is mediated by the PABD and helical domains of EF. It will be very interesting to see if TUA-diCl can inhibit EF internalization because this would completely abolish the toxic activity of EF and would lead to a novel strategy to inhibit the toxic activity of the enzyme by targeting the helical domain.

F. Appendix

During the PhD, I had the opportunity to write a review on the interplay between experimental evidences with molecular modeling in order to elucidate (i) the entry of bacterial toxins such as diphtheria, botulinic and tetanus and (ii) virulence factors of adenylyl cyclases EF, AC and ExoY into host cells.

toxins

MDPI

Review

Structural Biology and Molecular Modeling to Analyze the Entry of Bacterial Toxins and Virulence Factors into Host Cells

Irène Pitard [1,2,3] and Thérèse E Malliavin [1,2,*]

1 Unité de Bioinformatique Structurale, Institut Pasteur and CNRS UMR3528, 75015 Paris, France; irene.pitard@pasteur.fr
2 Centre de Bioinformatique, Biostatistique et Biologie Intégrative, Institut Pasteur and CNRS USR3756, 75015 Paris, France
3 Sorbonne Université, Collège Doctoral, Ecole Doctorale Complexité du Vivant, 75005 Paris, France
* Correspondence: therese.malliavin@pasteur.fr

Received: 28 May 2019; Accepted: 18 June 2019; Published: 24 June 2019

Abstract: Understanding the functions and mechanisms of biological systems is an outstanding challenge. One way to overcome it is to combine together several approaches such as molecular modeling and experimental structural biology techniques. Indeed, the interplay between structural and dynamical properties of the system is crucial to unravel the function of molecular machinery's. In this review, we focus on how molecular simulations along with structural information can aid in interpreting biological data. Here, we examine two different cases: (i) the endosomal translocation toxins (diphtheria, tetanus, botulinum toxins) and (ii) the activation of adenylyl cyclase inside the cytoplasm (edema factor, CyA, ExoY).

Keywords: anthrax; botulinium toxin; diphtheria toxin; bordetella pertussis; adenylyl cyclase; molecular modeling; enhanced sampling; drug design; calmodulin

Key Contribution: The review describes application of techniques of modeling and structural biology study to analyze the entry of bacterial toxins and virulence factors into host cells. This permits to put in evidence common features in structure and dynamics for diphtheria, botulinic and tetanus toxins on one side, and for adenylyl cyclases EF and AC on the other side.

1. Introduction

Structural biology evolved from the determination of the structure of small metalloproteins [1] that are moderately flexible around a stable conformation up, carried out in the late 50s, to the current studies of intrinsically disordered proteins [2]. Such development has been accelerated by the expansion of biophysical techniques. The combined uses of these methods provides an increasing amount of information on structure and dynamics of biomolecular systems. One major result of this development is the recognition that structure, dynamics and conformational transitions are intricately connected in all biomolecular systems, and they are also closely related to the cellular and physiological processes in which the biomolecules are involved. In addition, due to intrinsic gaps in experimental studies, the parallel use of molecular modeling and structural biology approaches is essential to understand protein function, in particular in the case of conformational flexibility. Indeed, transient phenomena as mobility or conformational transitions are difficult to experimentally quantify, as most experimental techniques used in structural biology only gain sufficient sensitivity by time and/or space averaging of the signal.

Molecular modeling approaches face other issues. One long-lasting problem is the size of the conformational space that must be explored, which is still a challenge despite the development of high-power computing and of enhanced sampling techniques [3–5]. Another problem is the limitations induced by the physical model of inter-atomic interactions or *force-field* [6,7]. Nevertheless, in numerous examples, the use of molecular modeling along with structural biology or biophysics approaches allowed a deeper investigation of the system [8–10].

In the present review, we present how the cooperative use of in silico and experimental structural biology approaches improve our knowledge of specific steps of the insertion of bacterial toxins and virulence factors into the host cells. We focus the review on two events, widely recognized as essential in the toxicology field: (i) the translocation of diphtheria, botulinic and tetanus toxins through the membrane of endosome vesicles, (ii) the activation of adenylyl cyclases in the cytoplasm of host cells. The choice of these biological topics was motivated by the existence of several X-ray crystallographic structures of these toxins, as well as by the presence of large conformational transitions investigated by molecular modeling studies.

2. Bacterial Toxins

2.1. Diphtheria Toxin

Diphtheria toxin (DT) was recognized as a major disease-causing agent at the end of 19th century, when filtered bacterial culture was shown to reproduce the disease in animal models. This 58 kDa protein includes three domains: the catalytic (C) domain (residues 1–193), the middle translocation (T) domain (residues 205–378) and the receptor-binding (R) domain (residues 386–535) (Figure 1). The monomeric DT [11] is a Y-shaped molecule, formed by two fragments A and B connected by a disulfide bridge. The fragment A corresponds to the domain C, which contains α and β secondary structures, colored in orange or in olive green in Figure 1. The fragment B contains: (i) the domain T, formed by 9 α helices labeled TH1 to TH9, and colored in magenta or in purple in Figure 1; (ii) the domain R, a flattened β barrel, colored in green or in turquoise in Figure 1 and displaying an overall topology similar to those of the immunoglobulin (Ig) domains [11].

DT is endocytosed into the cells by binding its fragment B to the heparin-binding epidermal growth factor-like precursor (HB-EGF). Under acidic conditions of the late endosome, the domain C is cleaved and translocated into the cytosol. Once in the cytoplasm the C domain then catalyzes the transfer of the ADP-ribose moiety of NAD onto the elongation factor-2 (EF-2), rendering EF-2 inactive and consequently causing protein synthesis to stop in the host cell.

Several structures of DT [11–15] revealed either monomeric or dimeric forms, each containing its own arrangement of the three domains. The monomeric DT displays a packed arrangement of the three domains, whereas the dimeric DT is formed by swapping of the R domains from two monomeric DT [13] in extended conformation (Figure 1A). Dimerization of DT can be induced by freezing the protein in mixed phosphate buffers, which are known to decrease in pH from 7.0 to 3.6 during freezing. This observation led Carroll et al. [16] to propose that the decrease in pH causes dimerization and, on the basis of this study, it has also been proposed [13] that lowering the pH can convert the monomeric DT into an open form. The pH effect is supported by the existence of numerous charged and polar residues present at the interface between R and C domains. Indeed, the three salt bridges stabilizing the interface at neutral pH will be disrupted at low pH leading to the transition toward the extended monomer structure [13]. Within the T domain, the hydrophobic helices TH8 and TH9 are sandwiched by two layers of amphiphilic helices, TH1 to TH4 and TH5 to TH7. The loops between TH5 and TH6-7 and between TH8 and TH9, which contain acidic residues D290, D292, D295 and E349, D352, respectively, are called the dagger tips [12] (Figure 1A). At acidic pH, these acidic residues can be protonated, thus changing their solubility preference from water to membrane. This hypothesis is confirmed by an X-ray crystallographic structure of DT at low pH [17] (Figure 1B), which reveals an unfolding of the TH2, TH3 and TH4 α-helices from the domain T

which exposes an hydrophobic surface that includes the TH5 and TH8 α-helices, and the loop region connecting the TH8 and TH9 α-helices.

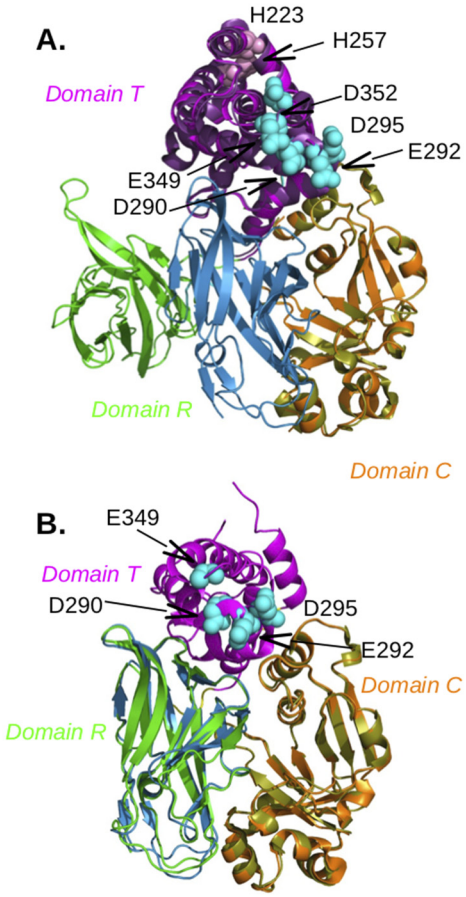


Figure 1. Diphtheria toxin (DT) structures determined with closed conformation (PDB entry: 1F0L [18]), with open conformation (PDB entry: 1DDT [12]), at acidic pH (PDB entry: 4OW6 [17]). The structures are drawn in cartoon. (**A**) Superimposition of the structures 1DDT and 1F0L on the domain C, The domains are colored in the following way: domain C (orange for 1DDT, olive green for 1F0L), domain T (magenta for 1DDT, violet for 1F0L), domain R (green for 1DDT, turquoise for 1F0L). The moving apart of the green domain T in the open conformation of 1F0L is visible. The residues D290, E292, D295, E349, D352, drawn with spheres and colored in cyan, are located in the dagger tip. The residues H223 and H257 are drawn as spheres and are colored in pink. (**B**) Superimposition of the structures 4OW6 and 1F0L on the domain C, The domains are colored in the following way: domain C (orange for 4OW6, olive green for 1F0L), domain T (magenta for 4OW6, not shown for sake of clarity for 1F0L), domain R (green for 4OW6, turquoise for 1F0L). The residues D290, E292, D295, D352, drawn with spheres and colored in cyan, are located in the dagger tip. The region containing residues H223 and H257 is unfolded in 4OW6 because of the acidic pH conditions.

The membrane translocation and the effect of acidic pH were further studied by a large set of biophysical methods: fluorescence [19–23], specular neutron reflectometry and solid-state

nuclear magnetic resonance (NMR) spectroscopy [24], substituted-cysteine accessibility [25], X-ray crystallography [17], hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) [26,27], site-directed spin labeling [28], site-directed mutagenesis [29,30], hydrophobic photo-labeling [31], and introduction of disulfide bridges [32]. These studies provide a very contrasting overview of the conformational transition of DT at low pH and of its translocation through the membrane.

Using fluorescence techniques, several pieces of information were obtained on the behavior of the T domain within DT: (i) the hydrophobic helices TH5-TH9 tend to be more exposed to aqueous solution in the isolated T domain [20]; (ii) the dagger tip connecting TH5 to TH6-7 is stable on the cis (insertion) side of the membrane in the isolated T domain, whereas it moves between the cis and trans sides of the membrane if the domain C is present [20]; (iii) the TH6-TH7 segment displays [25] constriction in the channel possibly formed by the T domain in the membrane; (iv) the T domain acts as a chaperone for the C domain translocation at acidic pH [21]; (v) the mutations E362Q and E349D/D352N, located in the region TH8–TH9 (named as the dagger tip), cause changes to how T inserts into the membrane and the E362Q mutation induces insertion of TH8-TH9 in the membrane at neutral pH [22]; (vi) the protonation of the residue H257, located in the T domain, acts as a pH-sensitive switch that triggers conformational change, resulting in T insertion into the membrane, and the neighboring residue H223 was suggested as a modulator (safety-latch) of H257 protonation [29]; (vii) the H223Q mutant, which lacks the latch, displays unfolding at less acidic conditions (pH < 7.5) with respect to the pH < 6.5 required for WT protein unfolding [30].

The effect of pH was also studied on the isolated T domain. Using hydrogen/deuterium exchange experiments coupled to mass spectrometry [26], the structure of the T domain isolated at neutral pH displays high protection for α helix TH5, which is assigned to the formation of dimers. However, at acidic pH the molten globule state displays high protection within the helical hairpin TH8-TH9, which was assigned to self-assembly of this hydrophobic part of the domain, due to oligomerization. Nevertheless, pH-dependent HDX [27] kinetic analysis reveals a different picture in which transition occurs from the native state (W-state) to a membrane-competent state (W+-state), with exposure to the solvent of the hairpin TH8-9.

Neutron reflectivity experiments [24] provided direct evidence that the T domain's C-terminal helices penetrate deeply into the hydrocarbon core, while its N-terminal helices penetrate the polar head-group region. This penetration is performed via a pH-dependent, two-step process: (a) a destabilization of the native state through the protonation of histidine side chains (pH 7 to 6) followed by the interaction between the solvent-exposed hydrophobic surfaces of the molten-globule state and the membrane (pH 7 to 6); (b) the reorganization of the T domain conformation leading to a membrane-inserted state (pH 6 to 4). Cysteine-scanning mutagenesis performed [28] on 28 consecutive sites comprising the TH8 helix and the TH8-TH9 inter-helical loop, reveals upon membrane binding at pH 4.6, a major structural reorganization of the domain, with most of TH8 retaining its helicity and the TH8-TH9 loop converted into a new helical structure. Among double mutant cysteines, all mutants forming internal disulfide bridges within the catalytic domain were translocated less efficiently [32].

Several results have been reported for the oligomeric states of DT and of the domain DT in the membrane. It was hypothesized [31,33] that DT exists as an oligomer in membranes. A concerted approach including fluorescence quenching with molecular dynamics (MD) simulations [23] proposes an open-channel model. It was also proposed that there is no single conformation in the transmembrane state, but, rather, a collection of states with different folds and topologies [34]. It was also observed that in the channel formed by an isolated T domain in the membrane contains only one T domain molecule [35]. Besides, it was established [36,37] that two conformations of the T domain can exist in the membrane with helices TH5-9 close to the membrane surface (P state) or more deeply inserted (TM state). The loop between TH8 and TH9, corresponding to one dagger tip, was exposed to the trans side of the bilayer, while other solvent-exposed residues in the helices TH5-9 region are located near the cis surface [19].

To investigate the extensive but contradictory experimental information on the T and DT interaction with the membrane, several molecular modeling studies have been realized. Flores et al. [38] developed an original accelerated MD (DISEI-aMD) approach that biased the electrostatic interactions between atom pairs of the solute: the corresponding MD trajectories showed that, in agreement to circular dichroism and fluorescence experiments, the protonation of histidines triggers partial unfolding of N-terminal TH helices, exposing hydrophobic sites while retaining a global compact structure. However, the two independent recorded trajectories of 6.8 and 9.5 µs showed differences in the conformational changes observed for these TH helices. Furthermore, coarse-grained MD simulations of T domain in the presence of lipids [39] revealed highly varied membrane insertion modes of the T domain that depended on the trajectory replicas.

The initial models of the protein conformations and protein-membrane association [39] were further used as starting points of microsecond MD trajectories [40], in which the membrane-bound conformations stayed mainly close to the initial conformation, but inserted deeply in the membrane. Extensive equilibrium MD simulations of the domain T at low pH with a combined length of over 8 µs [41] demonstrate that histidine protonation results in substantial molecular rearrangements characterized by the unfolding of helices TH1 and TH2 and the loss of close contacts between the C- and N-terminal segments. During the 6-µs simulation at low pH, the kinking of helix TH1, initiated by rotation of backbone dihedral angles of K216, decreases the α-helical content on the C-terminus of TH1 and was followed by unfolding of its N-terminus. The structural changes include exposure of the TH8-9 hairpin containing the dagger tip, in preparation for its subsequent transmembrane insertion. In addition, pKa of histidines were calculated by Thermodynamic Integration and validated by fluorescence and fluorescence resonance energy transfer (FRET) experiments [41]. Thermodynamic Integration indicates diverse roles for the different histidines in the conformational switching of the T domain triggered by the pH: H223 is a likely candidate for early protonation, but H257 has the highest free energy of protonation, in relation with a large perturbation of the native structure.

The overall view of DT translocation through the endosome membrane is that this process follows a very complex mechanism. The numerous experiments performed on this system have provided some insights, but several contradicting models have been proposed. MD simulations provide a conformational view with respect to the experimental observations, highlighting the great heterogeneity of conformation observed for DT in the different environments studied. The role of some His residues was clarified, but, due to the enormous complexity of the conformational landscape of DT during the translocation, up to now only a superficial exploration was possible, despite the large computational time invested.

2.2. Botulinium and Tetanus Toxins

The botulinum toxins (BoNTs) and the tetanus toxin (TeNT) are the most potent toxins known, they target the nervous system and are consequently named neurotoxins [42]. The BoNTs are produced by a variety of anaerobic spore–forming Clostridial species categorized as *Clostridium botulinum*, and cause botulism mostly through food poisoning. Among the BoNTs, various types (from A to G) exist, from which the serotypes A, B and E are lethal. The typical symptom of botulism is flaccid paralysis, which is the inability to contract skeletal muscles, inducing impaired vision, followed by paralysis of facial muscles and ultimately respiratory failure due to diaphragm paralysis. The TeNT is produced by *Clostridium tetani* and causes the disease tetanus, which is drastically different from botulism and is characterized by periodic hypercontraction of skeletal muscles called spastic paralysis.

In agreement with the different symptoms associated to the poisoning, BoNT and TeNT target neuromuscular junction but with distinct destinations, as the catalytic domain of BoNTs is released into the cytosol of motor neurons, whereas the catalytic domain of TeNT is finally released into the cytosol of inhibitory neurons. Both toxin families enter into the synaptic recycling vesicle system, through binding to gangliosides (a class of glycosphingolipids) and to synaptic vesicle membrane proteins. Once the neurotoxin is embarked into a synaptic vesicle, a change of pH toward acidic values

induces a conformational change of the toxin, and consequently the translocation of the catalytic domain through the vesicle membrane to the synapse cytosol. This catalytic domain is then separated from BoNT or TeNT and expelled from the vesicle by the cleavage of a very conserved disulfide bridge. The catalytic domain, a zinc-dependent protease, cleaves various SNARE proteins depending on the type of the toxin, thus generating the toxic effect.

BoNTs and TeNT target different neurons, and medium identity percentage in the 30–40% range, is observed for the primary sequences among TeNT and the different types of BoNTs. Nevertheless, all structures determined for these toxins share quite similar organization. At the time of toxin synthesis in *Clostridium botulinum* and *Clostridium tetani*, a unique chain is present. Depending on the toxin types, this chain is cleaved in two peptidic chains, either during secretion from the bacterial pathogen or later within the host. The mature toxins are thus formed by two protein chains (light chain: Lc and heavy chain: Hc) covalently connected through a disulfide bridge. All structures (Figure 2) are organized as a butterfly, the central domain (the butterfly abdomen), named H_N, is α helical and corresponds to the translocation domain, helping the catalytic domain to pass the vesicule membrane. One butterfly wing is formed by the light chain previously described, and corresponds to the catalytic domain, a zinc-dependent protease named LC. The other wing (Hc), covalently connected to the H_N domain, contains the N-terminal (H_{CN}) and C-terminal (H_{CC}) receptor-binding regions. These regions, displaying closed β structures similar to immunoglobulin fold, establish the previously described interactions of BoNT with synaptic vesicle membrane proteins and gangliosides.

Several approaches are under development for preventing the lethality of BoNTs. The AntiBotABE program has started the development of an oligoclonal antibody cocktail [43–45]. Virtual screening of BoNTs target Lc producing numerous compounds preventing the cleavage of SNARE proteins by binding to the active site of the zinc metalloprotease [46–53]. But, by contrast and thanks to their characteristic paralysis effects, BoNTs are also used to treat an increasing number of medical disorders [42] including human neuromuscular disorders characterized by involuntary muscle contractions (strabismus, blepharospasm, and hemifacial spasm), as well as cosmetic applications.

In addition of displaying similar structure organization, X-ray crystallographic structures revealed relative similar positions of domains among BoNTs and between BoNTs and TeNT (Figure 2). Several BoNT types (BoNT/A1 [54], BoNT/B1 [55]) are in an open conformation, displaying a flat position of the wings, as for a naturalized butterfly (Figure 2A). Whereas, the BoNT/E [56] and TeNT [57] structures display two distinct closed conformations, in which both wings are folded together, as if the butterfly was perched upon a flower (Figure 2B), with relative different positions of the wings between the two toxins. Another intermediate conformation of BoNT is observed for progenitor toxin complexes (PTC) of BoNT/A1 [58], and BoNT/E [59], which protects the toxin and facilitates its absorption in the gastrointestinal tract. In all structures, a very long loop, the belt, belonging to the chain Hc, grips the catalytic domain LC. Two Ca^{2+}, described as essential [60] for the translocation, are bound within the translocation domain of BoNT/B1 [60]. Interestingly, in the BoNT/E structure [56], a Na^{+} ion is located at a position equivalent to that of a Ca^{2+}.

Another important aspect of the conformational landscape of BoNTs is the influence of the pH. As described above, acidic pH in synaptic vesicles induces a conformational change of BoNTs. As pH drops under 6, variation of protonation in histidine residues could play a major role in this change. Nevertheless, X-ray crystallographic structures of BoNT/B, determined at various pH values [60], do not display any variation of conformation. But, decrease of pH was also shown [61] to induce a decrease in ellipticity of the CD spectra. Small angle X-ray scattering (SAXS) measurements on TeNT revealed a conformational change as a function of pH [57]. The formation of the BoNT/E PTC complex is similarly favored by pH variations and networks of acidic/basic residues were put in evidence at the interface between BoNT/E and the non-toxic-non-hemagglutinin protein (NTNHE) [59].

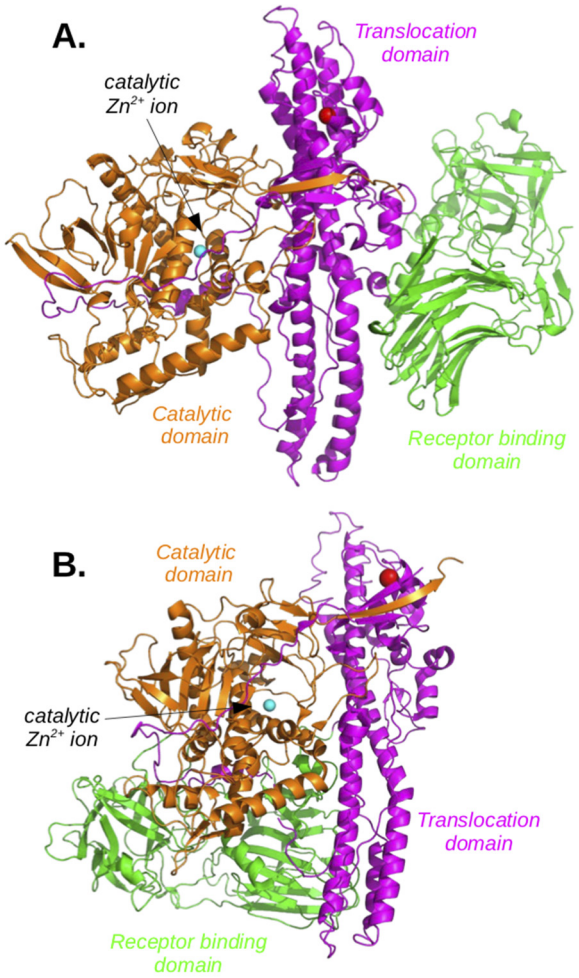


Figure 2. BoNT structures in open (**A**) and closed (**B**) conformations, corresponding to PDB entries: 3BTA (BoNT/A1 [54]) and 3FFZ (BoNT/E [56]). The protein chains are drawn in cartoon, with the translocation domain colored in magenta, the receptor binding domain colored in green, and the catalytic domain colored in orange. The catalytic ion Zn^{2+} is drawn as a sphere and colored in cyan and an ion Ca^{2+} (**A**) or Na^{+} (**B**) located at equivalent positions in the two structures is colored in blue.

The conformational landscape of BoNTs presents several interesting aspects. First, it is a fascinating example of large conformational variations in close relationship with important physiological processes. Second, shifts in pH play an important role in inducing structural changes, and it is challenging to properly model such effects. Third, understanding the mechanism of action of BoNTs will open the way to educated engineering of these toxins. Despite this, only one MD simulation study [62] have been performed [62] on an BoNT/A uncleaved chain.

DT and BoNT/TeNT toxins have numerous similar features in common. These toxins are composed of two polypeptidic chains attached by a disulfide bridge. Their structures display packed/extended and open/closed conformations, as well as non toxic dimers stabilized through a network of hydrogen

bonds between acidic residues. The R domain of DT displays an immunoglobulin fold which is also encountered in the domains HCC and HNN of BoNTs and TeNT. The translocation domain in both proteins is formed from α helices and the catalytic domains display an α/β fold.

However, these two protein families display also strong differences: (i) their size, about 500 residues for DT, about 1300 residues for BoNT and TeNT; (ii) a quite intricate topology for BoNT/TeNT with a belt region extending from the HN domain to grip the catalytic LC domain, whereas the topology of DT is based on more short-range interactions. The BoNT/E structure [56] displays disordered regions, whereas a more limited disorder is observed in any DT structure [17].

3. Adenylyl Cyclase Virulence Factors

The adenylyl cyclase toxins are present in several pathogens. Up to now, three of these toxins have been studied at the molecular level: the edema factor (EF) from *Bacillus anthracis*, CyaA from *Bordetella pertussis* and ExoY from *Pseudomonas aeruginosa* [63,64]. X-ray crystallographic structures have been determined for EF [65], CyaA [66] and ExoY [67]. EF and CyaA are activated as adenylyl cyclase by interaction with the ubiquitous protein calmodulin (CaM), present in the cytoplasm of the host cell. The activated adenylyl cyclase triggers overproduction of cyclic adenosine monophosphate (cAMP), which in high concentration perturbs the cell signaling system making its immune response inefficient. ExoY is activated by interaction with filamentous actin [64]. Results obtained by molecular modeling on the catalytic domains of CyaA and EF have been reviewed in details in Refs. [68,69].

3.1. Plasticity of the Adenylyl Cyclase Interaction with Calmodulin

CyaA revealed itself as a very good prototype for developing various biotechnological applications [70,71] which then drove the interest for its structure [66]. On the other hand, the interest in edema factor is based upon defence against its potential to cause harm in anthrax-based bio-terrorism [72]. The availability of high-resolution X-ray crystallographic structures for free EF, EF complexed to CaM and the catalytic domain AC of CyaA complexed to the C terminal lobe of CaM (C-CaM) (Figure 3) permits to study the adenylyl cyclase dynamical properties and to relate them to the toxin function. For the AC/CaM complex, the N-terminal lobe of CaM has been predicted to bind AC at different positions [73]. The enzymatic reaction of anthrax adenylyl cyclase has also been examined through the determination of various X-ray crystallographic structures including the reactant [74] and the reaction products [75] and showing EF uses two-metal–ion catalysis [76–78].

The interactions of AC and EF with calmodulin are different, although they share similar global features (Figure 3). The barrier of activation of AC is smaller than the one of EF, as the affinity of AC for calmodulin is about 0.2 nM [79], whereas it is 20 nM for EF [66,80]. This difference is also visible in an analysis of energetic influences. This coarse-grained analysis model [81,82], obtained by dividing the two partners of interaction into regions, analyzes the energetic influences between these regions. The influence diagrams obtained for the EF/CaM and AC/C-CaM complexes are qualitatively different as many more influences and a more intricate pattern are observed for the EF/CaM than for the AC/C-CaM diagram. This topological feature is not surprising, as the interaction between EF and CaM required to move apart the EF helical domain and the CA domain in order to insert CaM (Figure 3A,B). In contrast, the α helix H of AC, to which C-CaM is in direct interaction (Figure 3C), is more accessible. Consequently, the surface of interaction between adenylyl cyclase and CaM is larger for EF/C-CaM than for AC/C-CaM.

Observation of AC/CaM interaction in the X-ray crystallographic structure [66] (Figure 3C) suggests that this interaction mostly arises from the α helix H of AC. Indeed, mutations of CaM methionines, belonging to the CaM hydrophobic patch [83], strongly decrease the AC affinity for CaM [84]. Furthermore, the α helix H can be considered as a structural anchor, because the peptide spanning the sequence of helix H folds as an α-helix in solution [85,86].

However, available MD trajectories allowed to enlarge this point of view. Indeed, the removal of Ca^{2+} from the complex AC/C-CaM induced during the course of MD simulations [82] the breaking of

hydrogen bonds involving the AC residues D360, R338 and N347 located in the C-terminal extremity of AC, the AC residues Q302, E301 and N304 located in the catalytic loop, and the C-CaM residue R90 located in an α helix of the EF hand 3. The hydrogen bonds connecting these residues form a network from C-CaM to the AC catalytic site and their breaking is thus directly related to the AC function. The importance of these residues was confirmed by mutagenesis studies [87]: mutation to alanines induces a decrease of the affinity of AC for CaM, which agrees with the model of network described above and with a more complex interaction interface than the one initially guessed. Independent experimental studies of CaM [88] reveal that the interaction of calmodulin with AC increases the apparent Ca^{2+} binding in C-CaM.

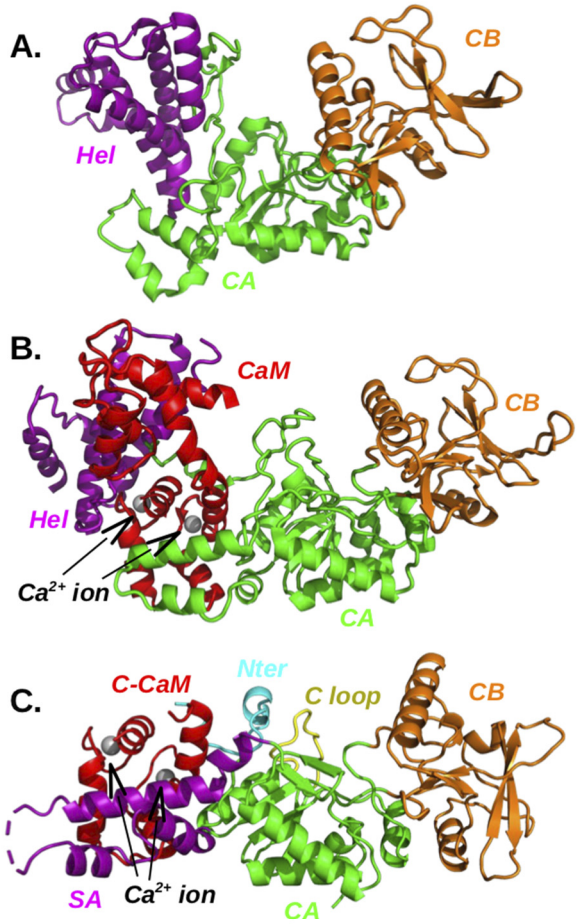


Figure 3. Adenylyl cyclase structures: (**A**) isolated EF (PDB entry: 1K8T [65]), (**B**) EF/CaM complex (PDB entry:1K90 [65]), (**C**) AC/C-CaM complex (PDB entry: 1YRT [66]). The full calmodulin (CaM) and the C terminal lobe of calmodulin (C-CaM) are colored in red with ions Ca^{2+} in silver. The domains CA and CB are colored in green and orange. The helical domain (EF) and the SA domain (AC) are colored in magenta. In the AC/C-CaM complex (C), the N terminal tail (Nter) of AC is colored in cyan and the catalytic loop (C loop) is colored in yellow.

In the complex AC/C-CaM, the loop containing residues 226–232, located at the extremity of SA region which is not visible in the crystal structure, but has been reconstructed using Modeller [89], displays large internal mobility along the MD trajectories [82]. The other parts of the SA region (Figure 3C) display larger mobility if Ca^{2+} are removed, and even more in the free protein AC. The SA mobility is induced by large variations of relative orientations between α helices. Using a combination of SAXS, HDX-MS, and synchrotron radiation circular dichroism (SR-CD), it was shown [90] that, in the absence of CaM, AC exhibits significant structural disorder, and that a 75-residue-long stretch within AC undergoes a disorder-to-order transition upon CaM binding.

In the structures of AC, the C terminal lobe of CaM is complexed with two Ca^{2+}. By contrast, several levels of complexation are present in EF/CaM complexes. An NMR study showed that the calcic loops of the C terminal lobe of CaM display the best affinity for Ca^{2+} ions [91]. For MD trajectories recorded on EF/CaM complexes using various levels of complexation by Ca^{2+} [92], CaM displays tendency to be more or less elongated and is the best fitted to the interaction with EF, in the case of two Ca^{2+} bound to C-CaM. The MD simulations of AC/CaM [82] and EF/CaM [81,92,93], revealed that unlike observations on isolated calmodulin [94,95], the open conformation of the helix-turn-helix motifs (the so-called EF hand) is kept in the presence as well as in the absence of Ca^{2+}.

Along the course of MD simulations [82], free AC conformations show a general tendency to become less elongated, by compacting the protein's extremities, the regions SA and CB. Nevertheless, the conformational drift of AC corresponds to oscillations around the X-ray crystallographic structure [66] and no conformational transition to a new basin was observed for the protein [82]. Enhanced sampling approaches, as temperature accelerated MD (TAMD) [96,97] and sr-TAMD [98], were used to further explore the conformational landscape of AC. The sr-TAMD approach allowed to obtain series of AC conformations [98] displaying a significantly less elongated shape than the starting X-ray crystallographic structure, as well as the conformations sampled in the previous TAMD. This decrease in elongation is obtained by a large reorganization of the α helices in the SA domain. The β hairpin (residues 259–273) is in most conformations less accessible to the solvent than in the AC/CaM complex, in agreement with its important role in the interaction with the N-terminal lobe of CaM (N-CaM) [99].

3.2. Searching Inhibitors for Adenylyl Cyclases

Whooping cough, caused by *Bordetella pertussis*, remains predominantly sensitive to antibiotic treatment, yet resistant strains are evolving [100,101] making the search for toxin inhibitors a sensible precautionary measure [100,102–106]. Search for inhibitors of edema factor have been also performed [102,107–114] and are reviewed in [113]. Some compounds are active against both CyaA and EF [115,116]. Monoclonal antibodies were raised against EF [117].

A virtual screening study of the adenylyl cyclase EF from *Bacillus anthracis* led [118] to the discovery of EF inhibitors, belonging to the series of thiophen ureidoacids. This study targeted the pocket SABC, located between the catalytic loop and the CaM binding site, which displays a large variation of shape between the active and inactive states of EF. Surprisingly, these EF inhibitors also displayed activity against AC [118], which was discovered by chance when AC was used as a control to avoid the detection of promiscuous ligands. The similarity between the inhibition of EF and AC is supported by the sequence alignment of proteins of the multifunctional-autoprocessing repeats-in-toxin (MARTX) toxin family [119], including EF, AC and ExoY. Although the sequence similarity is not very high, the alignment coupled to the possible similarity of the inhibitors interaction in EF and AC can be exploited to push forward the structural, biophysical and biochemical knowledge on the proteins MARTX.

4. Conclusions

The examples described here illustrate the importance of the combined use of molecular modeling and experimental structural biology approaches. Indeed, from the initially determined structures of

EF and AC, it would have been impossible to guess that the same family of compounds, the thiophen ureidoacids, could inhibit the enzymatic activity of both toxins. The use of molecular modeling can thus make possible to rationalize structural properties. Similarly, the comparison of the DT and BoNT/TeNT structures enabled detecting common physico-chemical properties, such as the open/closed conformation transition or the partial unfolding at low pH. Therefore, it could be expected that the description of the DT already obtained is qualitatively transferable to toxins BoNT and TeNT.

Author Contributions: Writing—review and editing, I.P. and T.E.M.; conceptualization, T.E.M.; supervision, T.E.M.

Funding: This research was funded by DGA support 2017033 and the Ecole Doctorale Complexité du Vivant (ED515).

Acknowledgments: We thank CNRS and Institut Pasteur for funding. Thérèse E Malliavin thanks Christine Rasetti-Escargueil, Michel-Robert Popoff and Emmanuel Lemichez for fruitful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kendrew, J.; Bodo, G.; Dintzis, H.; Parrish, R.; Wyckoff, H.; Phillips, D. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **1958**, *181*, 662–666. [CrossRef]
2. Dunker, A.K.; Babu, M.M.; Barbar, E.; Blackledge, M.; Bondos, S.E.; Dosztanyi, Z.; Dyson, H.J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; et al. What's in a name? Why these proteins are intrinsically disordered: Why these proteins are intrinsically disordered. *Intrinsically Disord Proteins* **2013**, *1*, e24157. [CrossRef]
3. Bernardi, R.; Melo, M.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* **2015**, *1850*, 872–877. [CrossRef]
4. Fujisaki, H.; Moritsugu, K.; Matsunaga, Y.; Morishita, T.; Maragliano, L. Extended Phase-Space Methods for Enhanced Sampling in Molecular Simulations: A Review. *Front. Bioeng. Biotechnol.* **2015**, *3*, 125. [CrossRef]
5. Miao, Y.; McCammon, J. Unconstrained Enhanced Sampling for Free Energy Calculations of Biomolecules: A Review. *Mol. Simul.* **2016**, *42*, 1046–1055. [CrossRef] [PubMed]
6. Wu, H.; Jiang, F.; Wu, Y. Significantly Improved Protein Folding Thermodynamics Using a Dispersion-Corrected Water Model and a New Residue-Specific Force Field. *J. Phys. Chem. Lett.* **2017**, *8*, 3199–3205. [CrossRef]
7. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958. [CrossRef]
8. Ma, B.; Nussinov, R. Release factors eRF1 and RF2: A universal mechanism controls the large conformational changes. *J. Biol. Chem.* **2004**, *279*, 53875–53885. [CrossRef] [PubMed]
9. Zachariae, U.; Grubmuller, H. A highly strained nuclear conformation of the exportin Cse1p revealed by molecular dynamics simulations. *Structure* **2006**, *14*, 1469–1478. [CrossRef] [PubMed]
10. Lee, E.H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K. Discovery through the computational microscope. *Structure* **2009**, *17*, 1295–1306. [CrossRef] [PubMed]
11. Choe, S.; Bennett, M.J.; Fujii, G.; Curmi, P.M.; Kantardjieff, K.A.; Collier, R.J.; Eisenberg, D. The crystal structure of diphtheria toxin. *Nature* **1992**, *357*, 216–222. [CrossRef] [PubMed]
12. Bennett, M.J.; Eisenberg, D. Refined structure of monomeric diphtheria toxin at 2.3 A resolution. *Protein Sci.* **1994**, *3*, 1464–1475. [CrossRef] [PubMed]
13. Bennett, M.J.; Choe, S.; Eisenberg, D. Domain swapping: Entangling alliances between proteins. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3127–3131. [CrossRef] [PubMed]
14. Bell, C.E.; Eisenberg, D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry* **1996**, *35*, 1137–1149. [CrossRef] [PubMed]
15. Bell, C.E.; Eisenberg, D. Crystal structure of nucleotide-free diphtheria toxin. *Biochemistry* **1997**, *36*, 481–488. [CrossRef] [PubMed]
16. Carroll, S.F.; Barbieri, J.T.; Collier, R.J. Dimeric form of diphtheria toxin: Purification and characterization. *Biochemistry* **1986**, *25*, 2425–2430. [CrossRef] [PubMed]
17. Leka, O.; Vallese, F.; Pirazzini, M.; Berto, P.; Montecucco, C.; Zanotti, G. Diphtheria toxin conformational switching at acidic pH. *FEBS J.* **2014**, *281*, 2115–2122. [CrossRef] [PubMed]

18. Steere, B. *Characterization of High-Order Oligomerization and Energetics in Diphteria Toxin*; University of California: Los Angeles, CA, USA, 2001.
19. Rosconi, M.P.; Zhao, G.; London, E. Analyzing topography of membrane-inserted diphtheria toxin T domain using BODIPY-streptavidin: At low pH, helices 8 and 9 form a transmembrane hairpin but helices 5-7 form stable nonclassical inserted segments on the cis side of the bilayer. *Biochemistry* **2004**, *43*, 9127–9139. [CrossRef] [PubMed]
20. Wang, J.; London, E. The membrane topography of the diphtheria toxin T domain linked to the a chain reveals a transient transmembrane hairpin and potential translocation mechanisms. *Biochemistry* **2009**, *48*, 10446–10456. [CrossRef]
21. Chassaing, A.; Pichard, S.; Araye-Guet, A.; Barbier, J.; Forge, V.; Gillet, D. Solution and membrane-bound chaperone activity of the diphtheria toxin translocation domain towards the catalytic domain. *FEBS J.* **2011**, *278*, 4516–4525. [CrossRef]
22. Ghalak, C.; Rodnin, M.; Vargas-Uribe, M.; McCluskey, A.; Flores-Canales, J.; Kurnikova, M.; Ladokhin, A. Role of Acidic Residues in Helices TH8–TH9 in Membrane Interactions of the Diphtheria Toxin T Domain. *Toxins* **2015**, *7*, 1303–1323.
23. Kyrychenko, A.; Lim, N.; Vasquez-Montes, V.; Rodnin, M.; Freites, J.; Nguyen, L.; Tobias, D.; Mobley, D.; Ladokhin, A. Refining Protein Penetration into the Lipid Bilayer Using Fluorescence Quenching and Molecular Dynamics Simulations: The Case of Diphtheria Toxin Translocation Domain. *J. Membrane Biol.* **2018**, *251*, 379–391. [CrossRef] [PubMed]
24. Chenal, A.; Prongidi-Fix, L.; Perier, A.; Aisenbrey, C.; Vernier, G.; Lambotte, S.; Haertlein, M.; Dauvergne, M.T.; Fragneto, G.; Bechinger, B.; et al. Deciphering membrane insertion of the diphtheria toxin T domain by specular neutron reflectometry and solid-state NMR spectroscopy. *J. Mol. Biol.* **2009**, *391*, 872–883. [CrossRef] [PubMed]
25. Kienker, P.; Wu, Z.; Finkelstein, A. Mapping the membrane topography of the TH6–TH7 segment of the diphtheria toxin T-domain channel. *J. Gen. Physiol.* **2015**, *145*, 107–125. [CrossRef] [PubMed]
26. Man, P.; Montagner, C.; Vitrac, H.; Kavan, D.; Pichard, S.; Gillet, D.; Forest, E.; Forge, V. Accessibility changes within diphtheria toxin T domain when in the functional molten globule state, as determined using hydrogen/deuterium exchange measurements. *FEBS J.* **2010**, *277*, 653–662. [CrossRef] [PubMed]
27. Li, J.; Rodnin, M.; Ladokhin, A.; Gross, M. Hydrogen-Deuterium Exchange and Mass Spectrometry Reveal the pH-Dependent Conformational Changes of Diphtheria Toxin T Domain. *Biochemistry* **2014**, *53*, 6849–6856. [CrossRef] [PubMed]
28. Oh, K.; Zhan, H.; Cui, C.; Altenbach, C.; Hubbell, W.; Collier, R. Conformation of the Diphtheria Toxin T Domain in Membranes: A Site-Directed Spin-Labeling Study of the TH8 Helix and TL5 Loop. *Biochemistry* **1999**, *38*, 10336–10343. [CrossRef] [PubMed]
29. Rodnin, M.V.; Kyrychenko, A.; Kienker, P.; Sharma, O.; Posokhov, Y.O.; Collier, R.J.; Finkelstein, A.; Ladokhin, A.S. Conformational switching of the diphtheria toxin T domain. *J. Mol. Biol.* **2010**, *402*, 1–7. [CrossRef] [PubMed]
30. Rodnin, M.V.; Li, J.; Gross, M.L.; Ladokhin, A.S. The pH-Dependent Trigger in Diphtheria Toxin T Domain Comes with a Safety Latch. *Biophys. J.* **2016**, *111*, 1946–1953. [CrossRef]
31. D'Silva, P.R.; Lala, A.K. Organization of diphtheria toxin in membranes. A hydrophobic photolabeling study. *J. Biol. Chem.* **2000**, *275*, 11771–11777. [CrossRef]
32. Falnes, P.O.; Choe, S.; Madshus, I.H.; Wilson, B.A.; Olsnes, S. Inhibition of membrane translocation of diphtheria toxin A-fragment by internal disulfide bridges. *J. Biol. Chem.* **1994**, *269*, 8402–8407. [PubMed]
33. Sharpe, J.C.; London, E. Diphtheria Toxin Forms Pores of Different Sizes Depending on Its Concentration in Membranes: Probable Relationship to Oligomerization. *J. Membr. Biol.* **1999**, *171*, 209–221. [CrossRef] [PubMed]
34. Ladokhin, A. pH-Triggered Conformational Switching along the Membrane Insertion Pathway of the Diphtheria Toxin T-Domain. *Toxins* **2013**, *5*, 1362–1380. [CrossRef] [PubMed]
35. Gordon, M.; Finkelstein, A. The number of subunits comprising the channel formed by the T domain of diphtheria toxin. *J. Gen. Physiol.* **2001**, *118*, 471–480. [CrossRef] [PubMed]
36. Wang, Y.; Malenbaum, S.E.; Kachel, K.; Zhan, H.; Collier, R.J.; London, E. Identification of shallow and deep membrane-penetrating forms of diphtheria toxin T domain that are regulated by protein concentration and bilayer width. *J. Biol. Chem.* **1997**, *272*, 25091–25098. [CrossRef] [PubMed]

37. Rosconi, M.P.; London, E. Topography of helices 5-7 in membrane-inserted diphtheria toxin T domain: Identification and insertion boundaries of two hydrophobic sequences that do not form a stable transmembrane hairpin. *J. Biol. Chem.* **2002**, *277*, 16517–16527. [CrossRef] [PubMed]
38. Flores-Canales, J.; Vargas-Uribe, M.; Ladokhin, A.; Kurnikova, M. Membrane Association of the Diphtheria Toxin Translocation Domain Studied by Coarse-Grained Simulations and Experiment. *J. Membrane Biol.* **2015**, *248*, 529–543. [CrossRef] [PubMed]
39. Flores-Canales, J.; Kurnikova, M. Targeting Electrostatic Interactions in Accelerated Molecular Dynamics with Application to Protein Partial Unfolding. *J. Chem. Theory Comput.* **2015**, *11*, 2550–2559. [CrossRef]
40. Flores-Canales, J.; Kurnikova, M. Microsecond Simulations of the Diphtheria Toxin Translocation Domain in Association with Anionic Lipid Bilayers. *J. Phys. Chem. B* **2015**, *119*, 12074–12085. [CrossRef]
41. Kurnikov, I.; Kyrychenko, A.; Flores-Canales, J.; Rodnin, M.; Simakov, N.; Vargas-Uribe, M.; Posokhov, Y.; Kurnikova, M.; Ladokhin, A. pH-Triggered Conformational Switching of the Diphtheria Toxin T-Domain: The Roles of N-Terminal Histidines. *J. Mol. Biol.* **2013**, *425*, 2752–2764. [CrossRef]
42. Dong, M.; Masuyer, G.; Stenmark, P. Botulinum and Tetanus Neurotoxins. *Annu. Rev. Biochem.* **2019**, in press. [CrossRef] [PubMed]
43. Miethe, S.; Mazuet, C.; Liu, Y.; Tierney, R.; Rasetti-Escargueil, C.; Avril, A.; Frenzel, A.; Thullier, P.; Pelat, T.; Urbain, R.; et al. Development of Germline-Humanized Antibodies Neutralizing Botulinum Neurotoxin A and B. *PLoS ONE* **2016**, *11*, e0161446. [CrossRef] [PubMed]
44. Derman, Y.; Selby, K.; Miethe, S.; Frenzel, A.; Liu, Y.; Rasetti-Escargueil, C.; Avril, A.; Pelat, T.; Urbain, R.; Fontayne, A.; et al. Neutralization of Botulinum Neurotoxin Type E by a Humanized Antibody. *Toxins (Basel)* **2016**, *8*, 257. [CrossRef] [PubMed]
45. Rasetti-Escargueil, C.; Avril, A.; Miethe, S.; Mazuet, C.; Derman, Y.; Selby, K.; Thullier, P.; Pelat, T.; Urbain, R.; Fontayne, A.; et al. The European AntibotABE Framework Program and Its Update: Development of Innovative Botulinum Antibodies. *Toxins (Basel)* **2017**, *9*, 309. [CrossRef] [PubMed]
46. Zuniga, J.E.; Schmidt, J.J.; Fenn, T.; Burnett, J.C.; Arac, D.; Gussio, R.; Stafford, R.G.; Badie, S.S.; Bavari, S.; Brunger, A.T. A potent peptidomimetic inhibitor of botulinum neurotoxin serotype A has a very different conformation than SNAP-25 substrate. *Structure* **2008**, *16*, 1588–1597. [CrossRef] [PubMed]
47. Hu, X.; Legler, P.M.; Southall, N.; Maloney, D.J.; Simeonov, A.; Jadhav, A. Structural insight into exosite binding and discovery of novel exosite inhibitors of botulinum neurotoxin serotype A through in silico screening. *J. Comput. Aided Mol. Des.* **2014**, *28*, 765–778. [CrossRef] [PubMed]
48. Teng, Y.H.; Berger, W.T.; Nesbitt, N.M.; Kumar, K.; Balius, T.E.; Rizzo, R.C.; Tonge, P.J.; Ojima, I.; Swaminathan, S. Computer-aided identification, synthesis, and biological evaluation of novel inhibitors for botulinum neurotoxin serotype A. *Bioorg. Med. Chem.* **2015**, *23*, 5489–5495. [CrossRef]
49. Guo, J.; Wang, J.; Gao, S.; Ji, B.; Waichi Chan, E.; Chen, S. Substrate-based inhibitors exhibiting excellent protective and therapeutic effects against Botulinum Neurotoxin A intoxication. *Sci. Rep.* **2015**, *5*, 16981. [CrossRef]
50. Kumaran, D.; Adler, M.; Levit, M.; Krebs, M.; Sweeney, R.; Swaminathan, S. Interactions of a potent cyclic peptide inhibitor with the light chain of botulinum neurotoxin A: Insights from X-ray crystallography. *Bioorg. Med. Chem.* **2015**, *23*, 7264–7273. [CrossRef]
51. Kumar, G.; Agarwal, R.; Swaminathan, S. Small molecule non-peptide inhibitors of botulinum neurotoxin serotype E: Structure-activity relationship and a pharmacophore model. *Bioorg. Med. Chem.* **2016**, *24*, 3978–3985. [CrossRef]
52. Minnow, Y.V.; Goldberg, R.; Tummalapalli, S.R.; Rotella, D.P.; Goodey, N.M. Mechanism of inhibition of botulinum neurotoxin type A light chain by two quinolinol compounds. *Arch. Biochem. Biophys.* **2017**, *618*, 15–22. [CrossRef] [PubMed]
53. Jacobson, A.R.; Adler, M.; Silvaggi, N.R.; Allen, K.N.; Smith, G.M.; Fredenburg, R.A.; Stein, R.L.; Park, J.B.; Feng, X.; Shoemaker, C.B.; et al. Small molecule metalloprotease inhibitor with in vitro, ex vivo and in vivo efficacy against botulinum neurotoxin serotype A. *Toxicon* **2017**, *137*, 36–47. [CrossRef] [PubMed]
54. Lacy, D.B.; Tepp, W.; Cohen, A.C.; DasGupta, B.R.; Stevens, R.C. Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat. Struct. Biol.* **1998**, *5*, 898–902. [CrossRef] [PubMed]
55. Swaminathan, S.; Eswaramoorthy, S. Structural analysis of the catalytic and binding sites of Clostridium botulinum neurotoxin B. *Nat. Struct. Biol.* **2000**, *7*, 693–699. [CrossRef] [PubMed]

56. Kumaran, D.; Eswaramoorthy, S.; Furey, W.; Navaza, J.; Sax, M.; Swaminathan, S. Domain organization in Clostridium botulinum neurotoxin type E is unique: Its implication in faster translocation. *J. Mol. Biol.* **2009**, *386*, 233–245. [CrossRef] [PubMed]
57. Masuyer, G.; Conrad, J.; Stenmark, P. The structure of the tetanus toxin reveals pH-mediated domain dynamics. *EMBO Rep.* **2017**, *18*, 1306–1317. [CrossRef]
58. Gu, S.; Rumpel, S.; Zhou, J.; Strotmeier, J.; Bigalke, H.; Perry, K.; Shoemaker, C.B.; Rummel, A.; Jin, R. Botulinum neurotoxin is shielded by NTNHA in an interlocked complex. *Science* **2012**, *335*, 977–981. [CrossRef]
59. Eswaramoorthy, S.; Sun, J.; Li, H.; Singh, B.R.; Swaminathan, S. Molecular Assembly of Clostridium botulinum progenitor M complex of type E. *Sci Rep* **2015**, *5*, 17795. [CrossRef]
60. Eswaramoorthy, S.; Kumaran, D.; Keller, J.; Swaminathan, S. Role of metals in the biological activity of Clostridium botulinum neurotoxins. *Biochemistry* **2004**, *43*, 2209–2216. [CrossRef]
61. Chellappan, G.; Kumar, R.; Santos, E.; Goyal, D.; Cai, S.; Singh, B.R. Structural and functional analysis of botulinum neurotoxin subunits for pH-dependent membrane channel formation and translocation. *Biochim. Biophys. Acta* **2015**, *1854*, 1510–1516. [CrossRef]
62. Chen, X.; Deng, Y. Long-time molecular dynamics simulations of botulinum biotoxin type-A at different pH values and temperatures. *J. Mol. Model.* **2007**, *13*, 559–572. [CrossRef] [PubMed]
63. Tang, W.; Guo, Q. The adenylyl cyclase activity of anthrax edema factor. *Mol. Asoects Med.* **2009**, *30*, 423–430. [CrossRef] [PubMed]
64. Belyy, A.; Mechold, U.; Renault, L.; Ladant, D. ExoY, an actin-activated nucleotidyl cyclase toxin from P. aeruginosa: A minireview. *Toxicon* **2018**, *149*, 65–71. [CrossRef] [PubMed]
65. Drum, C.; Yan, S.; Bard, J.; Shen, Y.; Lu, D.; Soelaiman, S.; Grabarek, Z.; Bohm, A.; Tang, W. Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature* **2002**, *415*, 396–402. [CrossRef] [PubMed]
66. Guo, Q.; Shen, Y.; Lee, Y.; Gibbs, C.; Mrksich, M.; Tang, W. Structural basis for the interaction of Bordetella pertussis adenylyl cyclase toxin with calmodulin. *EMBO J.* **2005**, *24*, 3190–3201. [CrossRef] [PubMed]
67. Khanppnavar, B.; Datta, S. Crystal structure and substrate specificity of ExoY, a unique T3SS mediated secreted nucleotidyl cyclase toxin from Pseudomonas aeruginosa. *Biochim. Biophys. Acta Gen. Subj.* **2018**, *1862*, 2090–2103. [CrossRef]
68. Malliavin, T.E. Molecular Modeling of the Catalytic Domain of CyaA Deepened the Knowledge of Its Functional Dynamics. *Toxins (Basel)* **2017**, *9*, 199. [CrossRef]
69. Laine, E.; Martinez, L.; Ladant, D.; Malliavin, T.; Blondel, A. Molecular motions as a drug target: Mechanistic simulations of anthrax toxin edema factor function led to the discovery of novel allosteric inhibitors. *Toxins (Basel)* **2012**, *4*, 580–604. [CrossRef]
70. Ladant, D.; Ullmann, A. Bordetella pertussis adenylate cyclase: A toxin with multiple talents. *Trends Microbiol.* **1999**, *7*, 172–176. [CrossRef]
71. Karimova, G.; Pidoux, J.; Ullmann, A.; Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5752 –5756. [CrossRef]
72. Narayanan, N.; Lacy, C.R.; Cruz, J.E.; Nahass, M.; Karp, J.; Barone, J.A.; Hermes-DeSantis, E.R. Disaster Preparedness: Biological Threats and Treatment Options. *Pharmacotherapy* **2018**, *38*, 217–234. [CrossRef] [PubMed]
73. Guo, Q.; Jureller, J.; Warren, J.; Solomaha, E.; Florián, J.; Tang, W. Protein-protein docking and analysis reveal that two homologous bacterial adenylyl cyclase toxins interact with calmodulin differently. *J. Biol. Chem.* **2008**, *283*, 23836–23845. [CrossRef] [PubMed]
74. Shen, Y.; Guo, Q.; Zhukovskaya, N.L.; Drum, C.L.; Bohm, A.; Tang, W.J. 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. [CrossRef] [PubMed]
75. Guo, Q.; Shen, Y.; Zhukovskaya, N.L.; Florian, J.; Tang, W.J. Structural and kinetic analyses of the interaction of anthrax adenylyl cyclase toxin with reaction products cAMP and pyrophosphate. *J. Biol. Chem.* **2004**, *279*, 29427–29435. [CrossRef] [PubMed]
76. Shen, Y.; Zhukovskaya, N.L.; Guo, Q.; Florian, J.; Tang, W.J. Calcium-independent calmodulin binding and two-metal-ion catalytic mechanism of anthrax edema factor. *EMBO J.* **2005**, *24*, 929–941. [CrossRef] [PubMed]

77. Martínez, L.; Laine, E.; Malliavin, T.; 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. [CrossRef]
78. Martínez, L.; Malliavin, T.; Blondel, A. Mechanism of reactant and product dissociation from the Anthrax Edema Factor: A Locally Enhanced Sampling and Steered Molecular Dynamics Study. *Proteins* **2011**, in press.
79. Shen, Y.; Lee, Y.; Soelaiman, S.; Bergson, P.; Lu, D.; Chen, A.; Beckingham, K.; Grabarek, Z.; Mrksich, M.; Tang, W. Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylyl cyclase exotoxins. *EMBO J.* **2002**, *21*, 6721–6732. [CrossRef]
80. 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. [CrossRef]
81. Laine, E.; Martínez, L.; Blondel, A.; Malliavin, T. 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. [CrossRef]
82. Selwa, E.; Laine, E.; Malliavin, T. Differential role of Calmodulin and Calcium ions in the stabilization of the catalytic domain of adenyl cyclase CyaA from *Bordetella pertussis*. *Proteins* **2012**, *80*, 1028–1040. [CrossRef] [PubMed]
83. Yang, C.; Jas, G.; Kuczera, K. Structure, dynamics and interaction with kinase targets: Computer simulations of calmodulin. *Biochim. Biophys. Acta* **2004**, *1697*, 289–300. [CrossRef] [PubMed]
84. Vougier, 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. [CrossRef] [PubMed]
85. Prêcheur, B.; Siffert, O.; Bârzu, O.; Craescu, C. 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*, 67–72. [CrossRef] [PubMed]
86. Craescu, C.; Bouhss, A.; Mispelter, J.; Diesis, E.; Popescu, A.; Chiriac, M.; Bârzu, O. Calmodulin binding of a peptide derived from the regulatory domain of Bordetella pertussis adenylate cyclase. *J. Biol. Chem.* **1995**, *270*, 7088–7096. [CrossRef] [PubMed]
87. Selwa, E.; Huynh, T.; Ciccotti, G.; Maragliano, L.; Malliavin, T. Temperature-accelerated molecular dynamics gives insights into globular conformations sampled in the free state of the AC catalytic domain. *Proteins* **2014**, *82*, 2483–2496. [CrossRef]
88. Springer, T.; Emerson, C.; Johns, C.; Finley, N. Interaction with adenylate cyclase toxin from Bordetella pertussis affects the metal binding properties of calmodulin. *FEBS Open Bio* **2017**, *7*, 25–34. [CrossRef]
89. Webb, B.; Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **2014**, *1137*, 1–15.
90. O'Brien, D.P.; Durand, D.; Voegele, A.; Hourdel, V.; Davi, M.; Chamot-Rooke, J.; Vachette, P.; Brier, S.; Ladant, D.; Chenal, A. Calmodulin fishing with a structurally disordered bait triggers CyaA catalysis. *PLoS Biol.* **2017**, *15*, e2004486. [CrossRef]
91. Ulmer, T.; Soelaiman, S.; Li, S.; Klee, C.; Tang, W.; Bax, A. Calcium dependence of the interaction between calmodulin and anthrax edema factor. *J. Biol. Chem.* **2003**, *278*, 29261–29266. [CrossRef]
92. Laine, E.; Yoneda, J.; Blondel, A.; Malliavin, T. 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. [CrossRef] [PubMed]
93. Laine, E.; Blondel, A.; Malliavin, T. Dynamics and energetics: A consensus analysis of the impact of calcium on EF-CaM protein complex. *Biophys. J.* **2009**, *96*, 1249–1263. [CrossRef] [PubMed]
94. 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. [CrossRef] [PubMed]
95. Finn, B.; Evenas, J.; Drakenberg, T.; Waltho, J.; Thulin, E.; Forsen, S. Calcium-induced structural changes and domain autonomy in calmodulin. *Nat. Struct. Biol.* **1995**, *2*, 777–783. [CrossRef] [PubMed]
96. Maragliano, L.; 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. [CrossRef]
97. 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. [CrossRef]

98. Cortes-Ciriano, I.; Bouvier, G.; Nilges, M.; Maragliano, L.; Malliavin, T. Temperature Accelerated Molecular Dynamics with Soft-Ratcheting Criterion Orients Enhanced Sampling by Low-Resolution Information. *J. Chem. Theory Comput.* **2015**, *11*, 3446–3454. [CrossRef]
99. Springer, T.; Goebel, E.; Hariraju, D.; Finley, N. Mutation in the β-hairpin of the Bordetella pertussis adenylate cyclase toxin modulates N-lobe conformation in calmodulin. *Biochem. Biophys. Res. Commun.* **2014**, *453*, 43–48. [CrossRef]
100. Seifert, R.; Dove, S. Towards selective inhibitors of adenylyl cyclase toxin from Bordetella pertussis. *Trends Microbiol.* **2012**, *20*, 343–351. [CrossRef]
101. Yang, Y.; Yao, K.; Ma, X.; Shi, W.; Yuan, L.; Yang, Y. Susceptibility to Erythromycin and Virulence-Related Genotype Changes in China (1970–2014). *PLoS ONE* **2015**, *10*, e0138941. [CrossRef]
102. Soelaiman, S.; Wei, B.Q.; Bergson, P.; Lee, Y.S.; Shen, Y.; Mrksich, M.; Shoichet, B.K.; Tang, W.J. Structure-based inhibitor discovery against adenylyl cyclase toxins from pathogenic bacteria that cause anthrax and whooping cough. *J. Biol. Chem.* **2003**, *278*, 25990–25997. [CrossRef]
103. Gottle, M.; Dove, S.; Steindel, P.; Shen, Y.; Tang, W.J.; Geduhn, J.; Konig, B.; Seifert, R. Molecular analysis of the interaction of Bordetella pertussis adenylyl cyclase with fluorescent nucleotides. *Mol. Pharmacol.* **2007**, *72*, 526–535. [CrossRef]
104. Taha, H.M.; Schmidt, J.; Gottle, M.; Suryanarayana, S.; Shen, Y.; Tang, W.J.; Gille, A.; Geduhn, J.; Konig, B.; Dove, S.; et al. Molecular analysis of the interaction of anthrax adenylyl cyclase toxin, edema factor, with 2′(3′)-O-(N-(methyl)anthraniloyl)-substituted purine and pyrimidine nucleotides. *Mol. Pharmacol.* **2009**, *75*, 693–703. [CrossRef]
105. Geduhn, J.; Dove, S.; Shen, Y.; Tang, W.J.; Konig, B.; Seifert, R. Bis-halogen-anthraniloyl-substituted nucleoside 5′-triphosphates as potent and selective inhibitors of Bordetella pertussis adenylyl cyclase toxin. *J. Pharmacol. Exp. Ther.* **2011**, *336*, 104–115. [CrossRef]
106. Taha, H.; Dove, S.; Geduhn, J.; Konig, B.; Shen, Y.; Tang, W.J.; Seifert, R. Inhibition of the adenylyl cyclase toxin, edema factor, from Bacillus anthracis by a series of 18 mono- and bis-(M)ANT-substituted nucleoside 5′-triphosphates. *Naunyn Schmiedebergs Arch. Pharmacol.* **2012**, *385*, 57–68. [CrossRef]
107. Gille, A.; Lushington, G.H.; Mou, T.C.; Doughty, M.B.; Johnson, R.A.; Seifert, R. Differential inhibition of adenylyl cyclase isoforms and soluble guanylyl cyclase by purine and pyrimidine nucleotides. *J. Biol. Chem.* **2004**, *279*, 19955–19969. [CrossRef] [PubMed]
108. Lee, Y.S.; Bergson, P.; He, W.S.; Mrksich, M.; Tang, W.J. Discovery of a small molecule that inhibits the interaction of anthrax edema factor with its cellular activator, calmodulin. *Chem. Biol.* **2004**, *11*, 1139–1146. [CrossRef] [PubMed]
109. Shen, Y.; Zhukovskaya, N.L.; Zimmer, M.I.; Soelaiman, S.; Bergson, P.; Wang, C.R.; Gibbs, C.S.; Tang, W.J. Selective inhibition of anthrax edema factor by adefovir, a drug for chronic hepatitis B virus infection. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3242–3247. [CrossRef]
110. Suryanarayana, S.; Wang, J.L.; Richter, M.; Shen, Y.; Tang, W.J.; Lushington, G.H.; Seifert, R. Distinct interactions of 2′- and 3′-O-(N-methyl)anthraniloyl-isomers of ATP and GTP with the adenylyl cyclase toxin of Bacillus anthracis, edema factor. *Biochem. Pharmacol.* **2009**, *78*, 224–230. [CrossRef] [PubMed]
111. Schein, C.H.; Chen, D.; Ma, L.; Kanalas, J.J.; Gao, J.; Jimenez, M.E.; Sower, L.E.; Walter, M.A.; Gilbertson, S.R.; Peterson, J.W. Pharmacophore selection and redesign of non-nucleotide inhibitors of anthrax edema factor. *Toxins (Basel)* **2012**, *4*, 1288–1300. [CrossRef] [PubMed]
112. Chen, D.; Ma, L.; Kanalas, J.J.; Gao, J.; Pawlik, J.; Jimenez, M.E.; Walter, M.A.; Peterson, J.W.; Gilbertson, S.R.; Schein, C.H. Structure-based redesign of an edema toxin inhibitor. *Bioorg. Med. Chem.* **2012**, *20*, 368–376. [CrossRef] [PubMed]
113. Seifert, R.; Dove, S. Inhibitors of Bacillus anthracis edema factor. *Pharmacol. Ther.* **2013**, *140*, 200–212. [CrossRef] [PubMed]
114. Jiao, G.S.; Kim, S.; Moayeri, M.; Thai, A.; Cregar-Hernandez, L.; McKasson, L.; O'Malley, S.; Leppla, S.H.; Johnson, A.T. Small molecule inhibitors of anthrax edema factor. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 134–139. [CrossRef] [PubMed]
115. Lubker, C.; Seifert, R. Effects of 39 Compounds on Calmodulin-Regulated Adenylyl Cyclases AC1 and Bacillus anthracis Edema Factor. *PLoS ONE* **2015**, *10*, e0124017. [CrossRef] [PubMed]

116. Česnek, M.; Skácel, J.; Jansa, P.; Dračinský, M.; Šmîdková, M.; Mertliková-Kaiserová, H.; Soto-Velasquez, M.P.; Watts, V.J.; Janeba, Z. Nucleobase Modified Adefovir (PMEA) Analogues as Potent and Selective Inhibitors of Adenylate Cyclases from Bordetella pertussis and Bacillus anthracis. *ChemMedChem* **2018**, *13*, 1779–1796. [CrossRef] [PubMed]
117. Chen, Z.; Moayeri, M.; Zhao, H.; Crown, D.; Leppla, S.H.; Purcell, R.H. Potent neutralization of anthrax edema toxin by a humanized monoclonal antibody that competes with calmodulin for edema factor binding. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13487–13492. [CrossRef] [PubMed]
118. Laine, E.; Goncalves, C.; Karst, J.; Lesnard, A.; Rault, S.; Tang, W.; Malliavin, T.; 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. [CrossRef] [PubMed]
119. Belyy, A.; Raoux-Barbot, D.; Saveanu, C.; Namane, A.; Ogryzko, V.; Worpenberg, L.; David, V.; Henriot, V.; Fellous, S.; Merrifield, C.; et al. Actin activates Pseudomonas aeruginosa ExoY nucleotidyl cyclase toxin and ExoY-like effector domains from MARTX toxins. *Nat. Commun.* **2016**, *7*, 135–182. [CrossRef]

VIII. Bibliography

1. NIAID Emerging Infectious Diseases/ Pathogens | NIH: National Institute of Allergy and Infectious Diseases. https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens.
2. CDC | Bioterrorism Agents/Diseases (by category) | Emergency Preparedness & Response. https://emergency.cdc.gov/agent/agentlist-category.asp (2019).
3. Mock, M. & Fouet, A. Anthrax. *Annual Review of Microbiology* **55**, 647–671 (2001).
4. Inglesby, T. V. *et al.* Anthrax as a Biological Weapon: Medical and Public Health Management. *JAMA* **281**, 1735–1745 (1999).
5. BBC NEWS | Americas | Using anthrax as a weapon. http://news.bbc.co.uk/2/hi/americas/1604621.stm.
6. Where global solutions are shaped for you | Disarmament | The Biological Weapons Convention. https://www.unog.ch/80256EE600585943/(httpPages)/04FBBDD6315AC720C1257180004B1B2F?OpenDocument.
7. Armes biologiques – UNODA. https://www.un.org/disarmament/fr/amd/armes-biologiques/.
8. NIAID Strategic Plan for Biodefense Research 2007 Update. 10 (2007).
9. Barth, H., Aktories, K., Popoff, M. R. & Stiles, B. G. Binary bacterial toxins: biochemistry, biology, and applications of common Clostridium and Bacillus proteins. *Microbiol. Mol. Biol. Rev.* **68**, 373–402, table of contents (2004).
10. Vitale, G. *et al.* Anthrax Lethal Factor Cleaves the N-Terminus of MAPKKs and Induces Tyrosine/Threonine Phosphorylation of MAPKs in Cultured Macrophages. *Biochemical and Biophysical Research Communications* **248**, 706–711 (1998).
11. Duesbery, N. S. *et al.* Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280**, 734–737 (1998).
12. Leppla, S. H. Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. *PNAS* **79**, 3162–3166 (1982).
13. Labruyere, E. *et al.* Characterization of ATP and calmodulin-binding properties of a truncated form of Bacillus anthracis adenylate cyclase. *Biochemistry* **29**, 4922–4928 (1990).
14. Guidi-Rontani, C., Levy, M., Ohayon, H. & Mock, M. Fate of germinated Bacillus anthracis spores in primary murine macrophages. *Molecular Microbiology* **42**, 931–938 (2001).
15. Corre, J.-P. *et al.* In Vivo Germination of Bacillus anthracis Spores During Murine Cutaneous Infection. *J Infect Dis* **207**, 450–457 (2013).
16. Tournier, J.-N., Rossi Paccani, S., Quesnel-Hellmann, A. & Baldari, C. T. Anthrax toxins: A weapon to systematically dismantle the host immune defenses. *Molecular Aspects of Medicine* **30**, 456–466 (2009).
17. Wolter, S., Dittmar, F. & Seifert, R. cCMP and cUMP in Apoptosis: Concepts and Methods. in *Non-canonical Cyclic Nucleotides* (ed. Seifert, R.) 25–47 (Springer International Publishing, 2017). doi:10.1007/164_2016_5007.

18. Park, J. M. *et al.* Signaling Pathways and Genes that Inhibit Pathogen-Induced Macrophage Apoptosis— CREB and NF-κB as Key Regulators. *Immunity* **23**, 319–329 (2005).
19. Moayeri, M., Leppla, S. H., Vrentas, C., Pomerantsev, A. P. & Liu, S. Anthrax Pathogenesis. *Annu. Rev. Microbiol.* **69**, 185–208 (2015).
20. Larabee, J. L. *et al.* Glycogen Synthase Kinase 3 Activation Is Important for Anthrax Edema Toxin-Induced Dendritic Cell Maturation and Anthrax Toxin Receptor 2 Expression in Macrophages ▿ . *Infect Immun* **79**, 3302–3308 (2011).
21. Maldonado-Arocho, F. J. & Bradley, K. A. Anthrax Edema Toxin Induces Maturation of Dendritic Cells and Enhances Chemotaxis towards Macrophage Inflammatory Protein 3β. *Infection and Immunity* **77**, 2036–2042 (2009).
22. Kim, C. *et al.* Antiinflammatory cAMP signaling and cell migration genes co-opted by the anthrax bacillus. *Proceedings of the National Academy of Sciences* **105**, 6150–6155 (2008).
23. Glomski, I. J., Piris-Gimenez, A., Huerre, M., Mock, M. & Goossens, P. L. Primary Involvement of Pharynx and Peyer's Patch in Inhalational and Intestinal Anthrax. *PLoS Pathog* **3**, (2007).
24. Paccani, S. R. *et al.* Anthrax toxins inhibit immune cell chemotaxis by perturbing chemokine receptor signalling. *Cellular Microbiology* **9**, 924–929 (2007).
25. Rossi Paccani, S. *et al.* The adenylate cyclase toxin of Bacillus anthracis is a potent promoter of TH17 cell development. *Journal of Allergy and Clinical Immunology* **127**, 1635–1637 (2011).
26. Fang, H., Xu, L., Chen, T. Y., Cyr, J. M. & Frucht, D. M. Anthrax Lethal Toxin Has Direct and Potent Inhibitory Effects on B Cell Proliferation and Immunoglobulin Production. *The Journal of Immunology* **176**, 6155–6161 (2006).
27. Dumetz, F. *et al.* Noninvasive Imaging Technologies Reveal Edema Toxin as a Key Virulence Factor in Anthrax. *The American Journal of Pathology* **178**, 2523–2535 (2011).
28. Drum, C. L., Shen, Y., Rice, P. A., Bohm, A. & Tang, W.-J. 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 Cryst D* **57**, 1881–1884 (2001).
29. Drum, C. L. *et al.* Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature* **415**, 396–402 (2002).
30. Lacy, D. B., Wigelsworth, D. J., Melnyk, R. A., Harrison, S. C. & Collier, R. J. Structure of heptameric protective antigen bound to an anthrax toxin receptor: A role for receptor in pH-dependent pore formation. *PNAS* **101**, 13147–13151 (2004).
31. Kintzer, A. F. *et al.* The protective antigen component of anthrax toxin forms functional octameric complexes. *J. Mol. Biol.* **392**, 614–629 (2009).
32. Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. Crystal structure of the anthrax toxin protective antigen. *Nature* **385**, 833–838 (1997).
33. Lacy, D. B., Wigelsworth, D. J., Scobie, H. M., Young, J. A. T. & Collier, R. J. Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: An anthrax toxin receptor. *PNAS* **101**, 6367–6372 (2004).
34. Pannifer, A. D. *et al.* Crystal structure of the anthrax lethal factor. *Nature* **414**, 229–233 (2001).

35. Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. Identification of the cellular receptor for anthrax toxin. *Nature* **414**, 225–229 (2001).
36. Scobie, H. M., Rainey, G. J. A., Bradley, K. A. & Young, J. A. T. Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *PNAS* **100**, 5170–5174 (2003).
37. Wei, W., Lu, Q., Chaudry, G. J., Leppla, S. H. & Cohen, S. N. The LDL Receptor-Related Protein LRP6 Mediates Internalization and Lethality of Anthrax Toxin. *Cell* **124**, 1141–1154 (2006).
38. Milne, J. C., Furlong, D., Hanna, P. C., Wall, J. S. & Collier, R. J. Anthrax protective antigen forms oligomers during intoxication of mammalian cells. *J. Biol. Chem.* **269**, 20607–20612 (1994).
39. Katayama, H. *et al.* GroEL as a molecular scaffold for structural analysis of the anthrax toxin pore. *Nat Struct Mol Biol* **15**, 754–760 (2008).
40. Katayama, H. *et al.* Three-dimensional structure of the anthrax toxin pore inserted into lipid nanodiscs and lipid vesicles. *PNAS* **107**, 3453–3457 (2010).
41. Kintzer, A. F., Sterling, H. J., Tang, I. I., Williams, E. R. & Krantz, B. A. Anthrax Toxin Receptor Drives Protective Antigen Oligomerization and Stabilizes the Heptameric and Octameric Oligomer by a Similar Mechanism. *PLoS One* **5**, (2010).
42. Feld, G. K. *et al.* Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers. *Nat Struct Mol Biol* **17**, 1383–1390 (2010).
43. Kintzer, A. F. *et al.* Role of the Protective Antigen Octamer in the Molecular Mechanism of Anthrax Lethal Toxin Stabilization in Plasma. *Journal of Molecular Biology* **399**, 741–758 (2010).
44. Mogridge, J., Cunningham, K. & Collier, R. J. Stoichiometry of Anthrax Toxin Complexes. *Biochemistry* **41**, 1079–1082 (2002).
45. Mogridge, J., Mourez, M. & Collier, R. J. Involvement of Domain 3 in Oligomerization by the Protective Antigen Moiety of Anthrax Toxin. *J Bacteriol* **183**, 2111–2116 (2001).
46. Santelli, E., Bankston, L. A., Leppla, S. H. & Liddington, R. C. Crystal structure of a complex between anthrax toxin and its host cell receptor. *Nature* **430**, 905–908 (2004).
47. Fu, S. *et al.* The Structure of Tumor Endothelial Marker 8 (TEM8) Extracellular Domain and Implications for Its Receptor Function for Recognizing Anthrax Toxin. *PLoS One* **5**, (2010).
48. Shimaoka, M., Takagi, J. & Springer, T. A. Conformational Regulation of Integrin Structure and Function. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 485–516 (2002).
49. Rajapaksha, M. *et al.* pH effects on binding between the anthrax protective antigen and the host cellular receptor CMG2. *Protein Sci.* **21**, 1467–1480 (2012).
50. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).
51. S, J., T, K., Vg, I. & W, I. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859–1865 (2008).
52. Jiang, J., Pentelute, B. L., Collier, R. J. & Zhou, Z. H. Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature* **521**, 545–549 (2015).
53. Rainey, G. J. A. *et al.* Receptor-specific requirements for anthrax toxin delivery into cells. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13278–13283 (2005).

54. Williams, A. S., Lovell, S., Anbanandam, A., El-Chami, R. & Bann, J. G. Domain 4 of the anthrax protective antigen maintains structure and binding to the host receptor CMG2 at low pH. *Protein Sci* **18**, 2277–2286 (2009).
55. Melnyk, R. A. & Collier, R. J. A loop network within the anthrax toxin pore positions the phenylalanine clamp in an active conformation. *PNAS* **103**, 9802–9807 (2006).
56. Brown, M. J., Thoren, K. L. & Krantz, B. A. Role of the α clamp in the protein translocation mechanism of anthrax toxin. *J Mol Biol* **427**, 3340–3349 (2015).
57. Krantz, B. A. A Phenylalanine Clamp Catalyzes Protein Translocation Through the Anthrax Toxin Pore. *Science* **309**, 777–781 (2005).
58. Das, D. & Krantz, B. A. Peptide- and proton-driven allosteric clamps catalyze anthrax toxin translocation across membranes. *Proc Natl Acad Sci USA* **113**, 9611–9616 (2016).
59. Thoren, K. L., Worden, E. J., Yassif, J. M. & Krantz, B. A. Lethal factor unfolding is the most force-dependent step of anthrax toxin translocation. *PNAS* **106**, 21555–21560 (2009).
60. Krantz, B. A., Finkelstein, A. & Collier, R. J. Protein Translocation through the Anthrax Toxin Transmembrane Pore is Driven by a Proton Gradient. *Journal of Molecular Biology* **355**, 968–979 (2006).
61. Wynia-Smith, S. L., Brown, M. J., Chirichella, G., Kemalyan, G. & Krantz, B. A. Electrostatic Ratchet in the Protective Antigen Channel Promotes Anthrax Toxin Translocation. *J. Biol. Chem.* **287**, 43753–43764 (2012).
62. Hardenbrook, N. J. *et al.* Atomic structures of anthrax toxin protective antigen channels bound to partially unfolded lethal and edema factors. *Nat Commun* **11**, 1–10 (2020).
63. Antoni, C. *et al.* Cryo-EM structure of the fully-loaded asymmetric anthrax lethal toxin in its heptameric pre-pore state. *PLOS Pathogens* **16**, e1008530 (2020).
64. Das, D. & Krantz, B. A. Secondary Structure Preferences of the Anthrax Toxin Protective Antigen Translocase. *Journal of Molecular Biology* **429**, 753–762 (2017).
65. Shen, Y., Zhukovskaya, N. L., Guo, Q., Florián, J. & Tang, W.-J. Calcium-independent calmodulin binding and two-metal–ion catalytic mechanism of anthrax edema factor. *The EMBO Journal* **24**, 929–941 (2005).
66. Sunahara, R. K., Dessauer, C. W. & Gilman, A. G. Complexity and Diversity of Mammalian Adenylyl Cyclases. *Annu. Rev. Pharmacol. Toxicol.* **36**, 461–480 (1996).
67. Tang, W.-J. & Hurley, J. H. Catalytic Mechanism and Regulation of Mammalian Adenylyl Cyclases. *Mol Pharmacol* **54**, 231–240 (1998).
68. Tesmer, J. J. Crystal Structure of the Catalytic Domains of Adenylyl Cyclase in a Complex with Gs·GTPS. *Science* **278**, 1907–1916 (1997).
69. Zhang, G. *et al.* Characterization and crystallization of a minimal catalytic core domain from mammalian type II adenylyl cyclase. *Protein Science* **6**, 903–908 (1997).
70. Tesmer, J. J. & Sprang, S. R. The structure, catalytic mechanism and regulation of adenylyl cyclase. *Curr. Opin. Struct. Biol.* **8**, 713–719 (1998).
71. Sunahara, R. K. *et al.* Exchange of substrate and inhibitor specificities between adenylyl and guanylyl cyclases. *J. Biol. Chem.* **273**, 16332–16338 (1998).
72. Zhang, G., Liu, Y., Ruoho, A. E. & Hurley, J. H. Structure of the adenylyl cyclase catalytic core. *Nature* **386**, 247–253 (1997).

73. Florio, V. A. & Ross, E. M. Regulation of the catalytic component of adenylate cyclase. Potentiative interaction of stimulatory ligands and 2',5'-dideoxyadenosine. *Mol Pharmacol* **24**, 195–202 (1983).
74. Dessauer, C. W. & Gilman, A. G. The catalytic mechanism of mammalian adenylyl cyclase. Equilibrium binding and kinetic analysis of P-site inhibition. *J. Biol. Chem.* **272**, 27787–27795 (1997).
75. Mitterauer, T., Hohenegger, M., Tang, W.-J., Nanoff, C. & Freissmuth, M. The C2 Catalytic Domain of Adenylyl Cyclase Contains the Second Metal Ion (Mn^{2+}) Binding Site †. *Biochemistry* **37**, 16183–16191 (1998).
76. Tesmer, J. J. G. *et al.* Molecular Basis for P-Site Inhibition of Adenylyl Cyclase,. *Biochemistry* **39**, 14464–14471 (2000).
77. Mou, T.-C., Gille, A., Fancy, D. A., Seifert, R. & Sprang, S. R. Structural Basis for the Inhibition of Mammalian Membrane Adenylyl Cyclase by 2 ′(3′)- *O* -(*N* -Methylanthraniloyl)-guanosine 5 ′-Triphosphate. *J. Biol. Chem.* **280**, 7253–7261 (2005).
78. Mou, T.-C., Masada, N., Cooper, D. M. F. & Sprang, S. R. Structural Basis for Inhibition of Mammalian Adenylyl Cyclase by Calcium. *Biochemistry* **48**, 3387–3397 (2009).
79. Tang, W.-J. & Guo, Q. The Adenylyl Cyclase Activity of Anthrax Edema Factor. *Mol Aspects Med* **30**, 423–430 (2009).
80. Ladant, D. & Ullmann, A. Bordetella pertussis adenylate cyclase: a toxin with multiple talents. *Trends in Microbiology* **7**, 172–176 (1999).
81. Gupta, M., Alam, S. & Bhatnagar, R. Kinetic characterization and ligand binding studies of His351 mutants of Bacillus anthracis adenylate cyclase. *Archives of Biochemistry and Biophysics* **446**, 28–34 (2006).
82. Shen, Y. *et al.* 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.* **317**, 309–314 (2004).
83. Guo, Q., Shen, Y., Zhukovskaya, N. L., Florián, J. & Tang, W.-J. Structural and Kinetic Analyses of the Interaction of Anthrax Adenylyl Cyclase Toxin with Reaction Products cAMP and Pyrophosphate. *J. Biol. Chem.* **279**, 29427–29435 (2004).
84. Martínez, L., Laine, E., Malliavin, T. E., Nilges, M. & Blondel, A. ATP conformations and ion binding modes in the active site of anthrax edema factor: A computational analysis. *Proteins: Structure, Function, and Bioinformatics* **77**, 971–983 (2009).
85. Martínez, L., Malliavin, T. E. & Blondel, A. Mechanism of reactant and product dissociation from the anthrax edema factor: A locally enhanced sampling and steered molecular dynamics study. *Proteins: Structure, Function, and Bioinformatics* **79**, 1649–1661 (2011).
86. Chen, D. *et al.* Accounting for ligand-bound metal ions in docking small molecules on adenylyl cyclase toxins. *Proteins: Structure, Function, and Bioinformatics* **67**, 593–605 (2007).
87. Belyy, A. *et al.* Actin activates Pseudomonas aeruginosa ExoY nucleotidyl cyclase toxin and ExoY-like effector domains from MARTX toxins. *Nat Commun* **7**, 1–14 (2016).
88. Voegele, A. *et al.* Translocation and calmodulin-activation of the adenylate cyclase toxin (CyaA) of Bordetella pertussis. *Pathog Dis* **76**, (2018).

89. Hauser, A. R. The Type III Secretion System of Pseudomonas aeruginosa: Infection by Injection. *Nat Rev Microbiol* **7**, 654–665 (2009).
90. Drum, C. L. *et al.* An Extended Conformation of Calmodulin Induces Interactions between the Structural Domains of Adenylyl Cyclase from Bacillus anthracis to Promote Catalysis. *J. Biol. Chem.* **275**, 36334–36340 (2000).
91. Guo, Q. *et al.* Structural basis for the interaction of Bordetella pertussis adenylyl cyclase toxin with calmodulin. *The EMBO Journal* **24**, 3190–3201 (2005).
92. Beckert, U. *et al.* ExoY from Pseudomonas aeruginosa is a nucleotidyl cyclase with preference for cGMP and cUMP formation. *Biochemical and Biophysical Research Communications* **450**, 870–874 (2014).
93. Khanppnavar, B. & Datta, S. Crystal structure and substrate specificity of ExoY, a unique T3SS mediated secreted nucleotidyl cyclase toxin from Pseudomonas aeruginosa. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1862**, 2090–2103 (2018).
94. Belyy, A., Mechold, U., Renault, L. & Ladant, D. ExoY, an actin-activated nucleotidyl cyclase toxin from P. aeruginosa: A minireview. *Toxicon* **149**, 65–71 (2018).
95. Babu, Y. S., Bugg, C. E. & Cook, W. J. Structure of calmodulin refined at 2.2 A resolution. *J. Mol. Biol.* **204**, 191–204 (1988).
96. Babu, Y. S. *et al.* Three-dimensional structure of calmodulin. *Nature* **315**, 37–40 (1985).
97. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry* **31**, 5269–5278 (1992).
98. Zhang, M., Tanaka, T. & Ikura, M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat Struct Mol Biol* **2**, 758–767 (1995).
99. Swindells, M. B. & Ikura, M. Pre-formation of the semi-open conformation by the apo-calmodulin C-terminal domain and implications binding IQ-motifs. *Nat. Struct. Biol.* **3**, 501–504 (1996).
100. Minowa, O. & Yagi, K. Calcium Binding to Tryptic Fragments of Calmodulin1. *The Journal of Biochemistry* **96**, 1175–1182 (1984).
101. O'Neil, K. T. & DeGrado, W. F. How calmodulin binds its targets: sequence independent recognition of amphiphilic α-helices. *Trends in Biochemical Sciences* **15**, 59–64 (1990).
102. Ikura, M. *et al.* Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256**, 632–638 (1992).
103. Meador, W., Means, A. & Quiocho, F. Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. *Science* **257**, 1251–1255 (1992).
104. Meador, W., Means, A. & Quiocho, F. Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science* **262**, 1718–1721 (1993).
105. Cox, J. A., Comte, M., Fitton, J. E. & DeGrado, W. F. The interaction of calmodulin with amphiphilic peptides. *J. Biol. Chem.* **260**, 2527–2534 (1985).
106. Schumacher, M. A., Rivard, A. F., Bächinger, H. P. & Adelman, J. P. Structure of the gating domain of a Ca 2+ -activated K + channel complexed with Ca 2+ /calmodulin. *Nature* **410**, 1120–1124 (2001).
107. Osawa, M. *et al.* A novel target recognition revealed by calmodulin in complex with Ca 2+ -calmodulin-dependent kinase kinase. *Nat Struct Mol Biol* **6**, 819–824 (1999).

108. Elshorst, B. *et al.* NMR Solution Structure of a Complex of Calmodulin with a Binding Peptide of the Ca2+ Pump,. *Biochemistry* **38**, 12320–12332 (1999).
109. Ulmer, T. S. *et al.* Calcium Dependence of the Interaction between Calmodulin and Anthrax Edema Factor. *J. Biol. Chem.* **278**, 29261–29266 (2003).
110. Shen, Y. *et al.* Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylyl cyclase exotoxins. *The EMBO Journal* **21**, 6721–6732 (2002).
111. Laine, E., Yoneda, J. D., Blondel, A. & Malliavin, T. E. The conformational plasticity of calmodulin upon calcium complexation gives a model of its interaction with the oedema factor of Bacillus anthracis. *Proteins* **71**, 1813–1829 (2008).
112. Laine, E., Blondel, A. & Malliavin, T. E. Dynamics and Energetics: A Consensus Analysis of the Impact of Calcium on EF-CaM Protein Complex. *Biophys J* **96**, 1249–1263 (2009).
113. Springer, T. I., Goebel, E., Hariraju, D. & Finley, N. L. Mutation in the β-hairpin of the Bordetella pertussis adenylate cyclase toxin modulates N-lobe conformation in calmodulin. *Biochem. Biophys. Res. Commun.* **453**, 43–48 (2014).
114. Selwa, E., Laine, E. & Malliavin, T. E. Differential role of calmodulin and calcium ions in the stabilization of the catalytic domain of adenyl cyclase CyaA from Bordetella pertussis. *Proteins* **80**, 1028–1040 (2012).
115. Laine, E., Martínez, L., Blondel, A. & Malliavin, T. E. Activation of the edema factor of Bacillus anthracis by calmodulin: evidence of an interplay between the EF-calmodulin interaction and calcium binding. *Biophys. J.* **99**, 2264–2272 (2010).
116. Yang, C., Jas, G. S. & Kuczera, K. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *Biochim. Biophys. Acta* **1697**, 289–300 (2004).
117. Vougier, S. *et al.* Essential role of methionine residues in calmodulin binding to Bordetella pertussis adenylate cyclase, as probed by selective oxidation and repair by the peptide methionine sulfoxide reductases. *J. Biol. Chem.* **279**, 30210–30218 (2004).
118. O'Brien, D. P. *et al.* Calmodulin fishing with a structurally disordered bait triggers CyaA catalysis. *PLoS Biol.* **15**, e2004486 (2017).
119. Laine, E., Martínez, L., Ladant, D., Malliavin, T. & Blondel, A. Molecular motions as a drug target: mechanistic simulations of anthrax toxin edema factor function led to the discovery of novel allosteric inhibitors. *Toxins (Basel)* **4**, 580–604 (2012).
120. Brossier, F., Levy, M. & Mock, M. Anthrax Spores Make an Essential Contribution to Vaccine Efficacy. *Infect Immun* **70**, 661–664 (2002).
121. Bower, W. A. *et al.* Use of Anthrax Vaccine in the United States: Recommendations of the Advisory Committee on Immunization Practices, 2019. *MMWR Recomm Rep* **68**, 1–14 (2019).
122. Chen, Z. *et al.* Potent neutralization of anthrax edema toxin by a humanized monoclonal antibody that competes with calmodulin for edema factor binding. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13487–13492 (2009).
123. Farcasanu, M. *et al.* Rapid discovery and characterization of synthetic neutralizing antibodies against anthrax edema toxin. *Biochemistry* **58**, 2996–3004 (2019).
124. Soelaiman, S. *et al.* Structure-based Inhibitor Discovery against Adenylyl Cyclase Toxins from Pathogenic Bacteria That Cause Anthrax and Whooping Cough. *J. Biol. Chem.* **278**, 25990–25997 (2003).

125. Hadziyannis, S. J. *et al.* Adefovir Dipivoxil for the Treatment of Hepatitis B e Antigen–Negative Chronic Hepatitis B. *New England Journal of Medicine* **348**, 800–807 (2003).
126. Marcellin, P. *et al.* Adefovir Dipivoxil for the Treatment of Hepatitis B e Antigen–Positive Chronic Hepatitis B. *New England Journal of Medicine* **348**, 808–816 (2003).
127. Shen, Y. *et al.* Selective inhibition of anthrax edema factor by adefovir, a drug for chronic hepatitis B virus infection. *PNAS* **101**, 3242–3247 (2004).
128. Tournier, J.-N. *et al.* Anthrax Edema Toxin Cooperates with Lethal Toxin to Impair Cytokine Secretion during Infection of Dendritic Cells. *J Immunol* **174**, 4934–4941 (2005).
129. Suffredini, D. A. *et al.* Shock and lethality with anthrax edema toxin in rats are associated with reduced arterial responsiveness to phenylephrine and are reversed with adefovir. *American Journal of Physiology-Heart and Circulatory Physiology* **313**, H946–H958 (2017).
130. Shoshani, I., Laux, W. H. G., Périgaud, C., Gosselin, G. & Johnson, R. A. Inhibition of Adenylyl Cyclase by Acyclic Nucleoside Phosphonate Antiviral Agents. *J. Biol. Chem.* **274**, 34742–34744 (1999).
131. Lin, C., Xu, C., Teng, A., Yeh, L.-T. & Peterson, J. Pharmacokinetics of Pradefovir and PMEA in Healthy Volunteers After Oral Dosing of Pradefovir. *The Journal of Clinical Pharmacology* **45**, 1250–1258 (2005).
132. Erion, M. D. *et al.* Liver-Targeted Drug Delivery Using HepDirect Prodrugs. *J Pharmacol Exp Ther* **312**, 554–560 (2005).
133. Česnek, M. *et al.* Nucleobase Modified Adefovir (PMEA) Analogues as Potent and Selective Inhibitors of Adenylate Cyclases from Bordetella pertussis and Bacillus anthracis. *ChemMedChem* **13**, 1779–1796 (2018).
134. Chen, D. *et al.* Novel inhibitors of anthrax edema factor. *Bioorganic & Medicinal Chemistry* **16**, 7225–7233 (2008).
135. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem* **31**, 455–461 (2010).
136. Schein, C. H. *et al.* Pharmacophore Selection and Redesign of Non-nucleotide Inhibitors of Anthrax Edema Factor. *Toxins* **4**, 1288–1300 (2012).
137. Lasaro, M. A., Mathias-Santos, C., Rodrigues, J. F. & Ferreira, L. C. S. Functional and immunological characterization of a natural polymorphic variant of a heat-labile toxin (LT-I) produced by enterotoxigenic *Escherichia coli* (ETEC). *FEMS Immunol Med Microbiol* **55**, 93–99 (2009).
138. Moen, S. T. *et al.* Testing the Efficacy and Toxicity of Adenylyl Cyclase Inhibitors against Enteric Pathogens Using In Vitro and In Vivo Models of Infection. *Infection and Immunity* **78**, 1740–1749 (2010).
139. Chemistry Tools. https://www.organic-chemistry.org/prog/.
140. Chen, D. *et al.* Structure-based redesign of an edema toxin inhibitor. *Bioorganic & Medicinal Chemistry* **20**, 368–376 (2012).
141. Johnson, R. A. & Shoshani, I. Inhibition of Bordetella pertussis and Bacillus anthracis adenylyl cyclases by polyadenylate and 'P'-site agonists. *J. Biol. Chem.* **265**, 19035–19039 (1990).

142. Dessauer, C. W., Tesmer, J. J. G., Sprang, S. R. & Gilman, A. G. The interactions of adenylate cyclases with P-site inhibitors. *Trends in Pharmacological Sciences* **20**, 205–210 (1999).
143. Taha, H. M. *et al.* Molecular Analysis of the Interaction of Anthrax Adenylyl Cyclase Toxin, Edema Factor, with 2′(3′)-O-(N-(methyl)anthraniloyl)-Substituted Purine and Pyrimidine Nucleotides. *Mol Pharmacol* **75**, 693–703 (2009).
144. Seifert, R. & Dove, S. Inhibitors of Bacillus anthracis edema factor. *Pharmacology & Therapeutics* **140**, 200–212 (2013).
145. Seifert, R., Lushington, G. H., Mou, T.-C., Gille, A. & Sprang, S. R. Inhibitors of membranous adenylyl cyclases. *Trends Pharmacol. Sci.* **33**, 64–78 (2012).
146. Göttle, M. *et al.* Cytidylyl and Uridylyl Cyclase Activity of Bacillus anthracis Edema Factor and Bordetella pertussis CyaA. *Biochemistry* **49**, 5494–5503 (2010).
147. Taha, H. *et al.* Inhibition of the adenylyl cyclase toxin, edema factor, from Bacillus anthracis by a series of 18 mono- and bis-(M)ANT-substituted nucleoside 5′-triphosphates. *Naunyn-Schmiedeberg's Arch Pharmacol* **385**, 57–68 (2012).
148. Gille, A. *et al.* Differential Inhibition of Adenylyl Cyclase Isoforms and Soluble Guanylyl Cyclase by Purine and Pyrimidine Nucleotides. *J. Biol. Chem.* **279**, 19955–19969 (2004).
149. Hübner, M., Dizayee, S., Matthes, J., Seifert, R. & Herzig, S. Effect of MANT-nucleotides on L-type calcium currents in murine cardiomyocytes. *Naunyn-Schmiedeberg's Arch Pharmacol* **383**, 573–583 (2011).
150. Jiao, G.-S. *et al.* Small molecule inhibitors of anthrax edema factor. *Bioorg. Med. Chem. Lett.* **28**, 134–139 (2018).
151. Wenthur, C. J., Gentry, P. R., Mathews, T. P. & Lindsley, C. W. Drugs for Allosteric Sites on Receptors. *Annu Rev Pharmacol Toxicol* **54**, 165–184 (2014).
152. Zheng, J. *et al.* Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* **32**, 2154–2161 (1993).
153. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* **12**, 88–118 (1965).
154. Changeux, J.-P. The concept of allosteric modulation: an overview. *Drug Discovery Today: Technologies* **10**, e223–e228 (2013).
155. Koshland, D. E., Némethy, G. & Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits*. *Biochemistry* **5**, 365–385 (1966).
156. Tsai, C.-J., Ma, B. & Nussinov, R. Folding and binding cascades: Shifts in energy landscapes. *Proceedings of the National Academy of Sciences* **96**, 9970–9972 (1999).
157. Kumar, S., Ma, B., Tsai, C.-J., Sinha, N. & Nussinov, R. Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Science* **9**, 10–19 (2000).
158. Clarkson, M. W., Gilmore, S. A., Edgell, M. H. & Lee, A. L. Dynamic coupling and allosteric behavior in a nonallosteric protein. *Biochemistry* **45**, 7693–7699 (2006).
159. Guo, J. & Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **116**, 6503–6515 (2016).
160. Kern, D. & Zuiderweg, E. R. The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology* **13**, 748–757 (2003).

161. Tsai, C.-J. & Nussinov, R. A Unified View of "How Allostery Works". *PLoS Comput Biol* **10**, (2014).
162. Liu, J. & Nussinov, R. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput. Biol.* **12**, e1004966 (2016).
163. Wodak, S. J. *et al.* Allostery in Its Many Disguises: From Theory to Applications. *Structure* **27**, 566–578 (2019).
164. Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins* **57**, 433–443 (2004).
165. Frauenfelder, H. & McMahon, B. Dynamics and function of proteins: The search for general concepts. *PNAS* **95**, 4795–4797 (1998).
166. Berendsen, H. J. & Hayward, S. Collective protein dynamics in relation to function. *Current Opinion in Structural Biology* **10**, 165–169 (2000).
167. Frauenfelder, H., Sligar, S. & Wolynes, P. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
168. Zhuravlev, P. I. & Papoian, G. A. Functional versus folding landscapes: the same yet different. *Curr. Opin. Struct. Biol.* **20**, 16–22 (2010).
169. Nussinov, R. & Tsai, C.-J. Allostery in Disease and in Drug Discovery. *Cell* **153**, 293–305 (2013).
170. Hilser, V. J., Wrabl, J. O. & Motlagh, H. N. Structural and energetic basis of allostery. *Annu Rev Biophys* **41**, 585–609 (2012).
171. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).
172. Nussinov, R. & Tsai, C.-J. Allostery without a conformational change? Revisiting the paradigm. *Current Opinion in Structural Biology* **30**, 17–24 (2015).
173. Cooper, A. & Dryden, D. T. F. Allostery without conformational change: A plausible model. *Eur Biophys J* **11**, 103–109 (1984).
174. Daura, X. Advances in the Computational Identification of Allosteric Sites and Pathways in Proteins. in *Protein Allostery in Drug Discovery* (eds. Zhang, J. & Nussinov, R.) 141–169 (Springer, 2019). doi:10.1007/978-981-13-8719-7_7.
175. Zhang, W., Xie, J. & Lai, L. Correlation Between Allosteric and Orthosteric Sites. in *Protein Allostery in Drug Discovery* (eds. Zhang, J. & Nussinov, R.) 89–105 (Springer, 2019). doi:10.1007/978-981-13-8719-7_5.
176. Süel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **10**, 59–69 (2003).
177. Gorman, S. D., D'Amico, R. N., Winston, D. S. & Boehr, D. D. Engineering Allostery into Proteins. in *Protein Allostery in Drug Discovery* (eds. Zhang, J. & Nussinov, R.) 359–384 (Springer, 2019). doi:10.1007/978-981-13-8719-7_15.
178. Cheng, X. & Jiang, H. Allostery in Drug Development. in *Protein Allostery in Drug Discovery* (eds. Zhang, J. & Nussinov, R.) 1–23 (Springer, 2019). doi:10.1007/978-981-13-8719-7_1.
179. Congreve, M., Oswald, C. & Marshall, F. H. Applying Structure-Based Drug Design Approaches to Allosteric Modulators of GPCRs. *Trends in Pharmacological Sciences* **38**, 837–847 (2017).

180. Christopoulos, A. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat Rev Drug Discov* **1**, 198–210 (2002).
181. Lee, Y.-S., Bergson, P., He, W. S., Mrksich, M. & Tang, W.-J. Discovery of a Small Molecule that Inhibits the Interaction of Anthrax Edema Factor with Its Cellular Activator, Calmodulin. *Chemistry & Biology* **11**, 1139–1146 (2004).
182. Tokumitsu, H. *et al.* KN-62, 1-[N,O-bis(5-isoquinolinesulfonyl)-N-methyl-L-tyrosyl]-4-phenylpiperazi ne, a specific inhibitor of Ca2+/calmodulin-dependent protein kinase II. *J. Biol. Chem.* **265**, 4315–4320 (1990).
183. Johnson, C. N. *et al.* The CaMKII inhibitor KN93-calmodulin interaction and implications for calmodulin tuning of NaV1.5 and RyR2 function. *Cell Calcium* **82**, 102063 (2019).
184. Gietzen, K., Wüthrich, A. & Bader, H. R 24571: A new powerful inhibitor of red blood cell Ca++-transport ATPase and of calmodulin-regulated functions. *Biochemical and Biophysical Research Communications* **101**, 418–425 (1981).
185. Vandonselaar, M., Hickie, R. A., Quail, W. & Delbaere, L. T. J. Trifluoperazine-induced conformational change in Ca 2+ -calmodulin. *Nat Struct Mol Biol* **1**, 795–801 (1994).
186. Caulfield, M. P. *et al.* The naphthalenesulphonamide calmodulin antagonist W7 and its 5-iodo-1-C8 analogue inhibit potassium and calcium currents in NG108-15 neuroblastoma × glioma cells in a manner possibly unrelated to their antagonism of calmodulin. *Neuroscience Letters* **125**, 57–61 (1991).
187. Johnson, J. D. & Wittenauer, L. A. A fluorescent calmodulin that reports the binding of hydrophobic inhibitory ligands. *Biochem J* **211**, 473–479 (1983).
188. Lübker, C. & Seifert, R. Effects of 39 Compounds on Calmodulin-Regulated Adenylyl Cyclases AC1 and Bacillus anthracis Edema Factor. *PLoS One* **10**, (2015).
189. Ahlijanian, M. K. & Cooper, D. M. Antagonism of calmodulin-stimulated adenylate cyclase by trifluoperazine, calmidazolium and W-7 in rat cerebellar membranes. *J Pharmacol Exp Ther* **241**, 407–414 (1987).
190. Schuler, D. *et al.* Interactions of Bordetella pertussis adenylyl cyclase toxin CyaA with calmodulin mutants and calmodulin antagonists: Comparison with membranous adenylyl cyclase I. *Biochemical Pharmacology* **83**, 839–848 (2012).
191. Hayes, M. P., Soto-Velasquez, M., Fowler, C. A., Watts, V. J. & Roman, D. L. Identification of FDA-Approved Small Molecules Capable of Disrupting the Calmodulin–Adenylyl Cyclase 8 Interaction through Direct Binding to Calmodulin. *ACS Chem. Neurosci.* **9**, 346–357 (2018).
192. Laine, E. *et al.* Use of allostery to identify inhibitors of calmodulin-induced activation of Bacillus anthracis edema factor. *PNAS* **107**, 11277–11282 (2010).
193. Chimiothèque Nationale. http://chimiotheque-nationale.cn.cnrs.fr/?lang=fr.
194. Le Foulon, F.-X., Braud, E., Fabis, F., Lancelot, J.-C. & Rault, S. Synthesis and combinatorial approach of the reactivity of 6- and 7-arylthieno[3,2-d][1,3]oxazine-2,4-diones. *Tetrahedron* **59**, 10051–10057 (2003).
195. Le Foulon, F.-X., Braud, E., Fabis, F., Lancelot, J.-C. & Rault, S. Solution-Phase Parallel Synthesis of a 1140-Member Ureidothiophene Carboxylic Acid Library. *J. Comb. Chem.* **7**, 253–257 (2005).

196. Drum, C. L. *et al.* An Extended Conformation of Calmodulin Induces Interactions between the Structural Domains of Adenylyl Cyclase from *Bacillus anthracis* to Promote Catalysis. *J. Biol. Chem.* **275**, 36334–36340 (2000).
197. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
198. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
199. Mayer, M. & Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **38**, 1784–1788 (1999).
200. Mayer, M. & Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **123**, 6108–6117 (2001).
201. Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M. & Stockman, B. WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J. Biomol. NMR* **21**, 349–359 (2001).
202. Thrippleton, M. J. & Keeler, J. Elimination of Zero-Quantum Interference in Two-Dimensional NMR Spectra. *Angewandte Chemie International Edition* **42**, 3938–3941 (2003).
203. Geen, H. & Freeman, R. Band-selective radiofrequency pulses. *Journal of Magnetic Resonance (1969)* **93**, 93–141 (1991).
204. Hwang, T. L. & Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* **112**, 275–279 (1995).
205. Favier, A. & Brutscher, B. NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J. Biomol. NMR* **73**, 199–211 (2019).
206. Schanda, P. & Brutscher, B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J. Am. Chem. Soc.* **127**, 8014–8015 (2005).
207. Kay, L., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665 (1992).
208. Waudby, C. A., Ramos, A., Cabrita, L. D. & Christodoulou, J. Two-Dimensional NMR Lineshape Analysis. *Scientific Reports* **6**, 24826 (2016).
209. Lee, D., Hilty, C., Wider, G. & Wüthrich, K. Effective rotational correlation times of proteins from NMR relaxation interference. *Journal of Magnetic Resonance* **178**, 72–76 (2006).
210. Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
211. Weber, P. *et al.* High-Throughput Crystallization Pipeline at the Crystallography Core Facility of the Institut Pasteur. *Molecules* **24**, (2019).
212. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Cryst* **40**, 658–674 (2007).
213. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst D* **60**, 2126–2132 (2004).

214. Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., & Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O. BUSTER v.2.11.2. *Cambridge, United Kingdom: Global Phasing Ltd.* (2011).
215. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **428**, 720–725 (2016).
216. Monet, D., Desdouits, N., Nilges, M. & Blondel, A. mkgridXf: Consistent Identification of Plausible Binding Sites Despite the Elusive Nature of Cavities and Grooves in Protein Dynamics. *J. Chem. Inf. Model.* **59**, 3506–3518 (2019).
217. Panjkovich, A. & Daura, X. PARS: a web server for the prediction of Protein Allosteric and Regulatory Sites. *Bioinformatics* **30**, 1314–1315 (2014).
218. Bouvier, G. *et al.* Target Engagement and Binding Mode of an Antituberculosis Drug to Its Bacterial Target Deciphered in Whole Living Cells by NMR. *Biochemistry* **58**, 526–533 (2019).
219. Kohonen, T. *Self-Organizing Maps*. (Springer-Verlag, 2001). doi:10.1007/978-3-642-56927-2.
220. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
221. Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).
222. Jo, S. *et al.* CHARMM-GUI 10 Years for Biomolecular Modeling and Simulation. *J Comput Chem* **38**, 1114–1124 (2017).
223. MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
224. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* **25**, 1400–1415 (2004).
225. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. *J Chem Theory Comput* **8**, 3257–3273 (2012).
226. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
227. Kim, S. *et al.* CHARMM-GUI Ligand Reader & Modeler for CHARMM Force Field Generation of Small Molecules. *J Comput Chem* **38**, 1879–1886 (2017).
228. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802 (2005).
229. Frenkel, D. & Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*. (Elsevier, 2001).
230. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).

231. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
232. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327–341 (1977).
233. Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* **52**, 24–34 (1983).
234. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671–690 (2010).
235. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
236. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
237. Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
238. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
239. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
240. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
241. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**, 2319–2327 (2011).
242. Gowers, R. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. in 98–105 (2016). doi:10.25080/Majora-629e541a-00e.
243. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Res* **5**, 189 (2016).
244. Feng, B. Y., Shelat, A., Doman, T. N., Guy, R. K. & Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nature Chemical Biology* **1**, 146–148 (2005).
245. Ryan, A. J., Gray, N. M., Lowe, P. N. & Chung, C. Effect of Detergent on "Promiscuous" Inhibitors. *J. Med. Chem.* **46**, 3448–3451 (2003).
246. Boulton, S. *et al.* Mechanisms of Specific versus Nonspecific Interactions of Aggregation-Prone Inhibitors and Attenuators. *J. Med. Chem.* **62**, 5063–5079 (2019).
247. Zega, A. NMR Methods for Identification of False Positives in Biochemical Screens. *J. Med. Chem.* **60**, 9437–9447 (2017).
248. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was

originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. *Advanced Drug Delivery Reviews* **46**, 3–26 (2001).
249. Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **7**, 42717 (2017).
250. Muegge, I., Heald, S. L. & Brittelli, D. Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* **44**, 1841–1846 (2001).
251. Jones, A. O. F., Leech, C. K., McIntyre, G. J., Wilson, C. C. & Thomas, L. H. Engineering short, strong hydrogen bonds in urea di-carboxylic acid complexes. *CrystEngComm* **16**, 8177–8184 (2014).
252. Cysewski, P., Przybylek, M., Ziółkowska, D. & Mroczyńska, K. Exploring the cocrystallization potential of urea and benzamide. *Journal of Molecular Modeling* **22**, 103 (2016).
253. Jennifer, S. J. & Muthiah, P. T. Design of co-crystals/salts of some Nitrogenous bases and some derivatives of thiophene carboxylic acids through a combination of hydrogen and halogen bonds. *Chem Cent J* **8**, 20 (2014).
254. Scherlis, D. A. & Marzari, N. π-Stacking in Thiophene Oligomers as the Driving Force for Electroactive Materials and Devices. *J. Am. Chem. Soc.* **127**, 3207–3212 (2005).
255. Archna, Pathania, S. & Chawla, P. A. Thiophene-based derivatives as anticancer agents: An overview on decade's work. *Bioorganic Chemistry* **101**, 104026 (2020).
256. Shah, R. & Verma, P. K. Therapeutic importance of synthetic thiophene. *Chem Cent J* **12**, (2018).
257. Gramec, D., Peterlin Mašič, L. & Sollner Dolenc, M. Bioactivation Potential of Thiophene-Containing Drugs. *Chem. Res. Toxicol.* **27**, 1344–1358 (2014).
258. Sahner, J. H., Groh, M., Negri, M., Haupenthal, J. & Hartmann, R. W. Novel small molecule inhibitors targeting the "switch region" of bacterial RNAP: Structure-based optimization of a virtual screening hit. *European Journal of Medicinal Chemistry* **65**, 223–231 (2013).
259. Fruth, M. *et al.* Binding Mode Characterization of Novel RNA Polymerase Inhibitors Using a Combined Biochemical and NMR Approach. *ACS Chem. Biol.* **9**, 2656–2663 (2014).
260. Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today* **15**, 444–450 (2010).
261. Dror, O., Shulman-Peleg, A. & Wolfson, R. N. and H. J. Predicting Molecular Interactions in silico: I. A Guide to Pharmacophore Identification and its Applications to Drug Design. *Current Medicinal Chemistry* vol. 11 71–90 https://www.eurekaselect.com/63022/article (2003).
262. Zhou, Z., Kearnes, S., Li, L., Zare, R. N. & Riley, P. Optimization of Molecules via Deep Reinforcement Learning. *Scientific Reports* **9**, 10752 (2019).
263. Geoghegan, K. F. *et al.* Spontaneous alpha-N-6-phosphogluconoylation of a 'His tag' in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **267**, 169–184 (1999).

264. Suryanarayana, S. *et al.* Distinct interactions of 2′- and 3′-O-(N-methyl)anthraniloyl-isomers of ATP and GTP with the adenylyl cyclase toxin of Bacillus anthracis, edema factor. *Biochemical Pharmacology* **78**, 224–230 (2009).
265. Li, D., DeRose, E. F. & London, R. E. The inter-ligand Overhauser effect: A powerful new NMR approach for mapping structural relationships of macromolecular ligands. *J Biomol NMR* **15**, 71–76 (1999).
266. Sánchez-Pedregal, V. M. *et al.* The INPHARMA Method: Protein-Mediated Interligand NOEs for Pharmacophore Mapping. *Angewandte Chemie International Edition* **44**, 4172–4175 (2005).
267. Fallon, J. L. & Quiocho, F. A. A Closed Compact Structure of Native Ca2+-Calmodulin. *Structure* **11**, 1303–1307 (2003).
268. Cook, W. J., Walter, L. J. & Walter, M. R. Drug Binding by Calmodulin: Crystal Structure of a Calmodulin-Trifluoperazine Complex. 7.
269. Bg, V. *et al.* Simultaneous binding of drugs with different chemical structures to Ca2+-calmodulin: crystallographic and spectroscopic studies. *Biochemistry* **37**, 15300–15310 (1998).
270. Feldkamp, M. D., O'Donnell, S. E., Yu, L. & Shea, M. A. Allosteric Effects of the Anti-Psychotic Drug Trifluoperazine on the Energetics of Calcium Binding by Calmodulin. *Proteins* **78**, 2265–2282 (2010).
271. Osawa, M. *et al.* Solution structure of Calmodulin-W-7 complex: the basis of diversity in molecular recognition11Edited by P. E. Wright. *Journal of Molecular Biology* **276**, 165–176 (1998).
272. Harmat, V. *et al.* A new potent calmodulin antagonist with arylalkylamine structure: crystallographic, spectroscopic and functional studies11Edited by R. Huber. *Journal of Molecular Biology* **297**, 747–755 (2000).
273. Horváth, I. *et al.* The Structure of the Complex of Calmodulin with KAR-2: A NOVEL MODE OF BINDING EXPLAINS THE UNIQUE PHARMACOLOGY OF THE DRUG. *J. Biol. Chem.* **280**, 8266–8274 (2005).
274. Weiss, B., Prozialeck, W. C. & Wallace, T. L. Interaction of drugs with calmodulin: Biochemical, pharmacological and clinical implications. *Biochemical Pharmacology* **31**, 2217–2226 (1982).
275. Wong, M. H. *et al.* The KN-93 Molecule Inhibits Calcium/Calmodulin-Dependent Protein Kinase II (CaMKII) Activity by Binding to Ca2+/CaM. *Journal of Molecular Biology* **431**, 1440–1459 (2019).
276. Audran, E. *et al.* A general framework to characterize inhibitors of calmodulin: Use of calmodulin inhibitors to study the interaction between calmodulin and its calmodulin binding domains. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1833**, 1720–1731 (2013).
277. Braunewell, K.-H. The darker side of Ca2+ signaling by neuronal Ca2+-sensor proteins: from Alzheimer's disease to cancer. *Trends in Pharmacological Sciences* **26**, 345–351 (2005).
278. Shim, J. S., Lee, J., Kim, K. N. & Kwon, H. J. Development of a new Ca2+/calmodulin antagonist and its anti-proliferative activity against colorectal cancer cells. *Biochemical and Biophysical Research Communications* **359**, 747–751 (2007).

279. Zaichick, S. V., McGrath, K. M. & Caraveo, G. The role of Ca2+ signaling in Parkinson's disease. *Dis Model Mech* **10**, 519–535 (2017).
280. Kuboniwa, H. *et al.* Solution structure of calcium-free calmodulin. *Nat Struct Mol Biol* **2**, 768–776 (1995).
281. Wriggers, W., Mehler, E., Pitici, F., Weinstein, H. & Schulten, K. Structure and Dynamics of Calmodulin in Solution. *Biophysical Journal* **74**, 1622–1639 (1998).
282. Bonomi, M. *et al.* PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **180**, 1961–1972 (2009).
283. Morelli, X., Bourgeas, R. & Roche, P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol* **15**, 475–481 (2011).
284. Tuffery, P. & Derreumaux, P. Flexibility and binding affinity in protein–ligand, protein–protein and multi-component protein interactions: limitations of current computational approaches. *J R Soc Interface* **9**, 20–33 (2012).
285. Zhang, X., Betzi, S., Morelli, X. & Roche, P. Focused chemical libraries--design and enrichment: an example of protein-protein interaction chemical space. *Future Med Chem* **6**, 1291–1307 (2014).
286. Huang, S.-Y. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. *Drug Discov. Today* **19**, 1081–1096 (2014).
287. Aguirre, C., Cala, O. & Krimm, I. Overview of Probing Protein-Ligand Interactions Using NMR. *Curr Protoc Protein Sci* **81**, 17.18.1-17.18.24 (2015).
288. Kuenemann, M. A. *et al.* In silico design of low molecular weight protein–protein interaction inhibitors: Overall concept and recent advances. *Progress in Biophysics and Molecular Biology* **119**, 20–32 (2015).
289. Fischer, E. S., Park, E., Eck, M. J. & Thomä, N. H. SPLINTS: Small-molecule protein ligand interface stabilizers. *Curr Opin Struct Biol* **37**, 115–122 (2016).
290. Shin, W.-H., Christoffer, C. W. & Kihara, D. In silico structure-based approaches to discover protein-protein interaction-targeting drugs. *Methods* **131**, 22–32 (2017).
291. Vajda, S., Beglov, D., Wakefield, A. E., Egbert, M. & Whitty, A. Cryptic binding sites on proteins: definition, detection, and druggability. *Curr Opin Chem Biol* **44**, 1–8 (2018).
292. Beglov, D. *et al.* Exploring the structural origins of cryptic sites on proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3416–E3425 (2018).
293. Kozakov, D. *et al.* The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **10**, 733–755 (2015).
294. Cimermancic, P. *et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J. Mol. Biol.* **428**, 709–719 (2016).
295. Ghanakota, P., DasGupta, D. & Carlson, H. A. Free Energies and Entropies of Binding Sites Identified by MixMD Cosolvent Simulations. *J Chem Inf Model* **59**, 2035–2045 (2019).
296. Martinez-Rosell, G., Lovera, S., Sands, Z. A. & De Fabritiis, G. PlayMolecule CrypticScout: Predicting Protein Cryptic Sites Using Mixed-Solvent Molecular Simulations. *J. Chem. Inf. Model.* **60**, 2314–2324 (2020).

297. Comitani, F. & Gervasio, F. L. Exploring Cryptic Pockets Formation in Targets of Pharmaceutical Interest with SWISH. *J Chem Theory Comput* **14**, 3321–3331 (2018).
298. Sun, Z., Wakefield, A. E., Kolossvary, I., Beglov, D. & Vajda, S. Structure-Based Analysis of Cryptic-Site Opening. *Structure* **28**, 223-235.e2 (2020).
299. Tschammer, N. Allosteric Modulators of the Class A G Protein Coupled Receptors. *Adv. Exp. Med. Biol.* **917**, 185–207 (2016).
300. Deredge, D. J. *et al.* Ligand-induced allostery in the interaction of the Pseudomonas aeruginosa heme binding protein with heme oxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3421–3426 (2017).
301. Feng, C., Roy, A. & Post, C. B. Entropic allostery dominates the phosphorylation-dependent regulation of Syk tyrosine kinase release from immunoreceptor tyrosine-based activation motifs. *Protein Sci.* **27**, 1780–1796 (2018).
302. D, N., N, L. & C, S. Allosteric Modulators of Protein-Protein Interactions (PPIs). *Adv Exp Med Biol* **1163**, 313–334 (2019).
303. Goodey, N. M. & Benkovic, S. J. Allosteric regulation and catalysis emerge via a common route. *Nature Chemical Biology* **4**, 474–482 (2008).
304. van den Bedem, H., Bhabha, G., Yang, K., Wright, P. E. & Fraser, J. S. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat. Methods* **10**, 896–902 (2013).
305. Gur, M., Madura, J. D. & Bahar, I. Global Transitions of Proteins Explored by a Multiscale Hybrid Methodology: Application to Adenylate Kinase. *Biophys J* **105**, 1643–1652 (2013).
306. Raman, A. S., White, K. I. & Ranganathan, R. Origins of Allostery and Evolvability in Proteins: A Case Study. *Cell* **166**, 468–480 (2016).
307. Zhang, J. & Nussinov, R. *Protein Allostery in Drug Discovery*. (Springer Singapore, 2019). doi:10.1007/978-981-13-8719-7.
308. Guarnera, E. & Berezovsky, I. N. Toward Comprehensive Allosteric Control over Protein Activity. *Structure* **27**, 866-878.e1 (2019).
309. Guo, Q. *et al.* Protein-Protein Docking and Analysis Reveal That Two Homologous Bacterial Adenylyl Cyclase Toxins Interact with Calmodulin Differently. *J. Biol. Chem.* **283**, 23836–23845 (2008).
310. Bertini, I. *et al.* Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *PNAS* **101**, 6841–6846 (2004).
311. Anthis, N. J., Doucleff, M. & Clore, G. M. Transient, sparsely populated compact states of apo and calcium-loaded calmodulin probed by paramagnetic relaxation enhancement: interplay of conformational selection and induced fit. *J. Am. Chem. Soc.* **133**, 18966–18974 (2011).
312. Geduhn, J. *et al.* Bis-halogen-anthraniloyl-substituted nucleoside 5'-triphosphates as potent and selective inhibitors of Bordetella pertussis adenylyl cyclase toxin. *J. Pharmacol. Exp. Ther.* **336**, 104–115 (2011).
313. Finn, B. E. *et al.* Calcium-induced structural changes and domain autonomy in calmodulin. *Nat Struct Mol Biol* **2**, 777–783 (1995).
314. Komeiji, Y., Ueno, Y. & Uebayasi, M. Molecular dynamics simulations revealed Ca2+-dependent conformational change of Calmodulin. *FEBS Letters* **521**, 133–139 (2002).

315. Crivici, A. & Ikura, M. Molecular and Structural Basis of Target Recognition by Calmodulin. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 85–116 (1995).
316. Guarnera, E. & Berezovsky, I. N. Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. *PLOS Computational Biology* **12**, e1004678 (2016).
317. Greener, J. G. & Sternberg, M. J. E. AlloPred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinformatics* **16**, 335 (2015).
318. Alfayate, A., Rodriguez Caceres, C., Gomes Dos Santos, H. & Bastolla, U. Predicted dynamical couplings of protein residues characterize catalysis, transport and allostery. *Bioinformatics* **35**, 4971–4978 (2019).
319. Desdouits, N., Nilges, M. & Blondel, A. Principal Component Analysis reveals correlation of cavities evolution and functional motions in proteins. *J. Mol. Graph. Model.* **55**, 13–24 (2015).
320. Selwa, E. *et al.* Allosteric activation of Bordetella pertussis adenylyl cyclase by calmodulin: molecular dynamics and mutagenesis studies. *J. Biol. Chem.* **289**, 21131–21141 (2014).
321. Daily, M. D. & Gray, J. J. Allosteric Communication Occurs via Networks of Tertiary and Quaternary Motions in Proteins. *PLoS Comput Biol* **5**, (2009).
322. Bowman, G. R. & Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *PNAS* **109**, 11681–11686 (2012).
323. Clarke, D. *et al.* Identifying Allosteric Hotspots with Dynamics: Application to Inter- and Intra-species Conservation. *Structure* **24**, 826–837 (2016).
324. Song, K. *et al.* Improved Method for the Identification and Validation of Allosteric Sites. *J Chem Inf Model* **57**, 2358–2363 (2017).
325. Pfleger, C. *et al.* Ensemble- and Rigidity Theory-Based Perturbation Approach To Analyze Dynamic Allostery. *J. Chem. Theory Comput.* **13**, 6343–6357 (2017).
326. Huang, M. *et al.* AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses. *Nucleic Acids Res* **46**, W451–W458 (2018).
327. Abrusán, G. & Marsh, J. A. Ligand-Binding-Site Structure Shapes Allosteric Signal Transduction and the Evolution of Allostery in Protein Complexes. *Mol Biol Evol* **36**, 1711–1727 (2019).