



HAL
open science

Analyzing in silico the relationship between the activation of the edema factor and its interaction with calmodulin

Irène Pitard, Damien Monet, Pierre Louis Goossens, Arnaud Blondel, Thérèse Malliavin

► To cite this version:

Irène Pitard, Damien Monet, Pierre Louis Goossens, Arnaud Blondel, Thérèse Malliavin. Analyzing in silico the relationship between the activation of the edema factor and its interaction with calmodulin. In press. pasteur-02999615v1

HAL Id: pasteur-02999615

<https://pasteur.hal.science/pasteur-02999615v1>

Preprint submitted on 25 Nov 2020 (v1), last revised 4 Dec 2020 (v2)

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

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

Copyright

November 25, 2020

Analyzing *in silico* the relationship between the activation of the edema factor and its interaction with calmodulin

Irène Pitard^{1,2,3}, Damien Monet^{1,2,3}, Pierre L. Goossens⁴, Arnaud Blondel^{1,2} and Thérèse E
Malliavin^{1,2}

¹ Institut Pasteur and CNRS UMR 3528, rue du Dr Roux, Unité de Bioinformatique Structurale,
75015 Paris, France

² Institut Pasteur and CNRS USR 3756, rue du Dr Roux, Centre de Bioinformatique, Biostatistique
et Biologie Intégrative, 75015 Paris, France

³ Ecole Doctorale Université Paris Sorbonne, 75005 Paris, France

⁴ Institut Pasteur, rue du Dr Roux, Unité Yersinia, 75015 Paris, France

Key words: protein-protein interaction; *Bacillus anthracis*; virulence factor;
cavity detection; allostery

Corresponding author: Thérèse E Malliavin, Institut Pasteur and CNRS UMR 3528, Unité de Bioinformatique Structurale, 25 rue du Dr Roux, 75015 Paris, France E-mail address: therese.malliavin@pasteur.fr

Short title: Interaction adefovir/edema factor

Abstract

Molecular dynamics (MD) simulations have been recorded on the complex between the edema factor (EF) of *Bacillus anthracis* and calmodulin (CaM), starting from a structure with the orthosteric inhibitor adefovir bound in the EF catalytic site. The starting structure has been destabilized by alternately suppressing different co-factors, such as adefovir ligand or ions, revealing several long-distance correlations between the conformation of CaM, the geometry of the CaM/EF interface, the enzymatic site and the overall organization of the complex. An allosteric communication between CaM/EF interface and the EF catalytic site, highlighted by these correlations, was confirmed by several bioinformatics approaches from the literature. A network of hydrogen bonds and stacking interactions extending from the helix V of CaM, and the residues of the switches A, B and C, and connecting to catalytic site residues, is a plausible candidate for the mediation of allosteric communication. The greatest variability in volume between the different MD conditions was also found for cavities present at the EF/CaM interface and in the EF catalytic site. The similarity between the predictions from literature and the volume variability might introduce the volume variability as new descriptor of allostery.

Contribution to the Field Statement

In the field of drug discovery, medium affinity of ligands or internal flexibility of the targets makes difficult to determine precisely the binding sites. Indeed, as classical structural biology uses a signal averaged over time or space, the signal-to-noise ratio is rapidly degraded in the event of internal dynamics. However, an interesting aspect of systems with a internal mobility or a large conformational transition, is the great variability of the cavities present on the molecular surface, leading to the appearance of cryptic and allosteric pockets which expand the possibilities of ligand binding. But, due to the internal mobility, these pockets have more chances to be transiently populated and are also more difficult to observe in structural biology. *In silico* molecular dynamics approaches are thus methods of choice for investigating the druggability of protein or protein complexes cavities in case of large internal mobility. By studying here the case of a fairly flexible protein-protein interaction, we found that a procedure basing on molecular dynamics simulations in presence of various co-factors and detecting the large variability in the volumes of cavities tracked along the trajectories, makes it possible to propose allosteric regions similar to those obtained by current approaches from the literature.

Introduction

As the interactions between proteins are essential in all biological processes, the modulation of these interactions with small ligands is a promising direction [1–8]. During the last decade, virtual screening has experienced a turning point where interest has widened from protein active sites to cryptic sites [9, 10] 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 [11, 12], by mixed-solvent MD simulations [13, 14] or by biased MD simulations [15, 16]. 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 [17–20] is particularly suitable in the case of inhibition of protein-protein interactions, because the conformational changes in which allosteric communication plays an important role when establishing the interaction, may create or modify cavities [21].

Allostery, discovered in the early days of molecular biology [22, 23] has since then, evolved from a model of discrete protein conformations to a more continuous description of the free energy landscape of proteins [21, 24, 25]. 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 [26–28]. The evolution of allostery description has allowed the emergence of so-called allosteric ligands [21, 29], for which the binding to the protein has a long-range influence on the protein conformation via residue networks. In this regard, a pioneering approach has been based on a detailed mechanistic simulation of functional motions [30]. The 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 [31]. 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 [32–35]. In particular, it was shown [36] that most of the known allosteric site motions show high correlations with corresponding orthosteric site motions, whereas other surface cavities did not. Many of the prediction methods are based on a description of protein structures as an elastic network in which each residue is replaced by the atom $C\alpha$ and the structure deformations are modelled through a set of springs between these atoms [32, 37]. The structure-based statistical mechanical model of allostery (SBSMMA) [37] has been introduced which permits to calculate the free energy variation due to allosteric communication.

In the present work, we investigate the use of molecular dynamics (MD) simulations and bioinformatics approaches to analyze protein-protein interactions, 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 [38]. The EF/CaM complex has been extensively studied by structural biology and biophysical techniques [38–47] as well as by molecular modeling [30, 48–53]. This complex (Figure 1A) 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 wide range of relative re-orientations of N terminal (N-CaM) and C terminal (C-CaM) lobes [54, 55], 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 organization (Figure 1A). Strikingly, in switch C two strings β and a connection loop present in the structure of isolated EF are converted to a α helix in the structure of the EF/CaM complex.

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 [56–61]. Among them, the ligand adefovir [57] was found by X-ray crystallography to bind in the catalytic site in the presence of an 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 [62]. Thiophen ureoacid ligands have been discovered following virtual screening on the pocket SABC, formed by residues from the three switches A, B, and C [30]. Since it is believed that they do not bind to the enzymatic site of EF, compounds 10506-2A and thiophene ureoacids must by definition bind to an allosteric site.

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 [57], 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 analysis of the trajectories made it possible to detect a network of hydrogen bonds and stacking interactions, connecting the EF catalytic site and the EF/CaM interface and showing a strong destabilization when the co-factors are removed from the EF/CaM structure. In addition, cavities

present in the EF/CaM complex have been tracked along all MD trajectories, and cavity volume variability has been proposed as a method of detecting allosteric pockets. The results obtained by this approach on EF/CaM have been confirmed by analyzes [33, 37]. Therefore, the network of hydrogen bonds and stacking interactions connecting the EF catalytic site and the EF/CaM interface could be considered a plausible candidate for the allosteric communication path within the complex structure.

Materials and Methods

Preparation of the systems for MD simulations

The X-ray crystallographic complex of EF with the inhibitor adefovir [57] (PDB entry: 1PK0, Figure 1A) served as the starting point of the MD trajectories. The protein chain was analyzed using Molprobit [63] (molprobit.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) [64, 65]. The chains A and D of the structure 1PK0 were neutralized using potassium ions and solvated with water molecules (Table I). The loop (residues 675-695) located between helices L and M [44] in the helical domain is disordered and not visible in the initial 1PK0 structure. This missing loop was added to the structure using the CHARMM GUI interface. The force field CHARMM36 [66–68] and the TIP3P water

model [69] 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 [70]. Six different systems have been prepared with different molecular compositions, starting from the structure 1PK0 then by removing various co-factors: ions Mg^{2+} , Ca^{2+} and adefovir (Table I).

Recording MD trajectories

The MD trajectories were recorded using NAMD 2.13 [71] (www.ks.uiuc.edu/Research/namd/). The simulations were performed in the NPT ensemble. 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 [72]. The RATTLE algorithm [73, 74] 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 integration steps. Then, the system was equilibrated for 50,000 steps. For each of six various conditions (Table I), 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 .

Analysis of MD trajectories

The root-mean-square deviations (RMSD, Å) of atomic coordinates, their root-mean-square fluctuations (RMSF, Å), as well as distance and angle analysis between atoms along the

recorded trajectories were performed using cpptraj [75]. Angles between CaM α helix axes were calculated using python scripts based on the python MDAnalysis library [76, 77], 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 an 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 trajectory were calculated using a python script based on the python MDAnalysis library [76, 77] and the software FreeSASA [78]. 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 are defined according to [79]. 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. The accessible surface was calculated using the Lee-Richards algorithm [80] with a probe radius of 1.4 Å.

Cavities of the EF/CaM complex were detected using the software *mkgridXf* [81]. The cavities are delimited between two surfaces: the inner and out surfaces determined by rolling probes of radii 1.4 and 8 Å on the protein atoms. In addition, allosteric pockets were predicted using the Web server of the PARS approach [32] (bioinf.uab.cat/cgi-bin/pars-cgi/pars.pl).

Results

The removal of co-factors destabilizes the organization of the EF/CaM complex

The evolution of coordinate root-mean-square deviations (RMSD, Å) 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 and 6 Å (Figure 2). The reproducibility of the replicas of a given trajectory has been further analyzed by realizing a clustering on each replica. This clustering was performed using a self-organizing map approach, described in the Supplementary Material. RMSD coordinates were calculated between the representative conformations extracted from clustering by comparing them two by two (Figure S1). The average RMSD values between replicas of the same trajectory (yellow cells) are larger than the average RMSD values within each replicas (grey cells), but are mostly smaller than the average RMSD between different trajectories. Interestingly, EF_ade_Mg_CaM_Ca, which corresponds to the full system, displays the smallest RMSD of 3.9 Å between the replicas.

The coordinate root-mean-square fluctuations (RMSF, Å) (Figure 3) display similar profiles for all trajectories. Unsurprisingly, a certain variability is observed for the peak of fluctuations on the loop (residues 675-695) missing in the initial X-ray crystallographic structure and modelled when preparing the system for molecular dynamics simulation, as described in the Materials and Methods. Globally, the N-CaM region fluctuates more than the C-CaM region, loaded with Ca^{2+} ions.

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 replicas. The trajectory EF_CaM_Ca, corresponding to the active toxin ready to interact with adefovir, has an internal mobility similar to that of EF_ade_Mg_CaM_Ca. The similarity of internal mobility for these two EF/CaM complexes mirrors 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 defined as the angle between the axes of helices IV and V (Figure 4). 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 vary in the 18-21 Å range (Figure 4,x-axis). Removal of Magnesium (EF_ade_CaM_Ca) or of Calcium (EF_ade_Mg_CaM) ions in the presence of adefovir induces an increase in the range of variations. On the other hand, the removal of the ligand makes it possible to maintain a narrow distribution of values, but shifts the radius of gyration to 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 4, 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. But, the systems EF_ade_CaM_Ca and EF_ade_Mg_CaM, in which the adefovir is present and only one ion type is conserved show large drifts of angle towards smaller angles down to 100° or larger angles up to 140°. Furthermore, the 2D contour plot describing the joint probability distribution of the gyration

radius and the bending angle ((Figure 4) reveals a strong correlation between the variations of these two parameters for systems EF_ade_CaM_Ca and EF_ade_Mg_CaM.

The drifts in the bending angle of the central α helix is a sign of a destabilization of the EF/CaM complex. Indeed, CaM displays in the EF/CaM complex a relatively unusual extended conformation [82]. It was highlighted in [48] using normal mode analysis of different trajectories recorded on EF/CaM that an exact fit of the extended conformation of CaM to the EF structure is required for the complex stability. Indeed, in the EF/CaM complex, the extended structure of CaM is used to push away the helical domain from the remaining part of EF. The variation of bending angle in CaM central helix has a direct influence on the extension of the conformation of CaM and therefore on its adjustment to the activated EF. On the other hand, the variation of gyration radius of the whole complex corresponds to a major perturbation and perturbs the catalytic site and catalytic activity of EF. Thus the correlation observed between the gyration radius and the bending angle ((Figure 4) proves that the fit of extended CaM to the complex EF/CaM has an influence on the EF function.

Overall, the removal of ions induces perturbations in the EF/CaM complex, which displays important conformational changes highlighted by variations of the gyration radius. These perturbations are visible through the atomic fluctuations as well as the correlated variations of gyration radius and CaM central α helix bending. The bending of the central helix α of CaM, due to the local environment, is related to the radius of gyration, describing the general shape of the complex, thus highlighting a long-distance effect.

CaM conformation inside the EF/CaM complex conserves features of the isolated CaM

The CaM conformation will be analyzed and compared to the literature [83] information in order to assess the fitting of CaM to the interaction with EF. CaM is an extremely flexible protein [54, 55], its conformation being strongly modulated by the loading of Calcium ions [84–86]. This allows CaM to bind various target proteins and peptides involved in the signaling processes [87].

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 S1). Side chain atoms O δ from residues D93, D95 and residues D131, D133 show weaker coordination than the other coordinated residues.

It is well known from the literature that the presence of Calcium ions has a strong influence on the conformation of the isolated CaM. The angles between the α helices of EF hands in CaM increase, as well as the accessible surfaces of hydrophobic patches, upon Calcium loading [82, 84, 85]. In addition, in the absence of Calcium ions, the central α helix is more disordered [86, 88]. These variations of CaM conformation in the presence of Calcium ions allows a better interaction of CaM with peptides involved in Calcium signaling [83, 87].

Using as definition of the hydrophobic patches the residues previously listed in Materials and Methods, the surfaces of hydrophobic patches of N-CaM and C-CaM (Figure 5) display quite different trends. The C-CaM patch, corresponding to EF-hands loaded with ions Ca^{2+} , displays profiles mostly concentrated around 200 \AA^2 , with some replica displaying few jumps up to 900 \AA^2 . By contrast, the N-CaM patch show much more diversity spanning a range of $200\text{-}600 \text{ \AA}^2$. Similar trends have been observed for accessible surfaces of methionines in previous simulations of the literature [79], with a cumulative exposed surface of N-CaM Met of about 26 \AA^2 in apo-form and 88 \AA^2 with Calcium, whereas the C-CaM methionines displayed cumulative exposed surfaces of 45 \AA^2 in apo-form and 124 \AA^2 with Calcium (Table 4 of [79]). The paradoxical behavior of the C-CaM patch which is blocked at a smaller accessible surface value than the N-CaM patch can be explained by the presence of the C_A region of EF which blocks the motions of C-CaM, as C-CaM is inserted between helical domain and C_A . At the contrary, one side of N-CaM interacts with the helical domain, whereas the opposite side of N-CaM is free, which allows more mobility of the N-CaM patch. The variations of hydrophobic patches in the complex EF/CaM are not similar to those described for the isolated CaM. Indeed, in the isolated CaM, the variations of hydrophobic patches favor the CaM interaction with signaling peptides, but this aspect is not relevant in the case of EF/CaM interaction.

In isolated CaM, EF hands show a trend to open when CaM is loaded with calcium [89, 90]. Similarly, in previous MD simulations of the EF/CaM complex [48], 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 the present simulations (Figure 6). In the present work, the EF-hands 3 and 4, located in C-CaM loaded with Ca^{2+} ions, display quite stable values

around respectively 80 and 90°, corresponding to open conformations. The EF-hand 4 fluctuates slightly more than the EF-hand 3, specially 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 display the largest variations in the range 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), displays equilibrium between closed and semi-open conformations. These EF-hands are located in the N-CaM lobe, which is not loaded with Calcium ions and interacts also in a much less intricate way with EF as only one side of N-CaM interacts with EF helical domain. For these two reasons, the angles between the two α helices are much more variable between the different conditions of simulations as well as between replicas of a given condition. In addition, the EF-hand 2 shows greater variability than the EF-hand 1, because it is farther away from the helical domain.

The overview of the CaM conformations reveals that the EF hands of N-CaM, which are not loaded with calcium ions, show much more heterogeneity in conformations, and populate conformations corresponding to closed and semi-open EF hand configurations. Otherwise, the overall conformations of CaM as well as the coordination of calcium are not strikingly modified among the trajectories recorded here.

A network of amino-acid interactions connect the EF catalytic site with CaM

The EF activity has been investigated according to the accessible surface of the catalytic site, 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. This accessible surface varies in the 400-1200 Å² range (Figure 7, left). This range of values agrees with the average catalytic surfaces of EF previously observed in MD studies [48], as: 928 Å² in the complex with 2 Calcium-loaded CaM, 866 Å² in the complex with the 4 Calcium-loaded CaM and 501 Å² in the complex with the apo CaM. Noticeably, in the presence of adefovir (black, yellow, magenta and green curves) the surfaces are smaller around 400-800 Å². This corresponds to a more closed catalytic site, in agreement with the inhibitory effect of adefovir. Removal of ion Mg²⁺ or removal of ions Ca²⁺ and Mg²⁺ in the presence of adefovir (magenta and yellow curves) induces a certain shift towards larger values, but the largest shifts towards the 800-1200 Å² range, is observed if adefovir is removed (cyan and blue curves). The trajectories EF_CaM_Ca (blue curves) corresponding to the activated EF display the most open catalytic site, which supports the use of accessible surface as an estimator of the EF catalytic activity.

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

The SABC pocket, previously used for the virtual screening having conducted to the discovery of thiophen ureoacids [30], is formed by residues, A496, P499, I538, E539, P542,

S544, S550, W552, Q553, T579, Q581, L625, Y626, Y627, N629, and N709, belonging to the three switches A, B, and C [44]. The accessible surface of the SABC pocket (Figure 7, right) varies in a much smaller range of 200-800 Å² than the catalytic pocket. Although the two pockets are defined by a similar number of residues, large differences are nonetheless observed for pocket SABC, sign of significant reorganizations in this region, resulting in some cases in the disappearance of a large accessible surface .

Initial analysis of the X-ray crystallographic structure 1PK0 [57] has revealed a network of hydrogen bonds connecting adefovir and residues from catalytic site with EF residues at the EF/CaM interface and residues from CaM (Figure 1B). This network displays hydrogen bond and stacking interactions which have been monitored along trajectories (Table S2). The contacts involve 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 1B). Interestingly, in the X-ray structures of EF [44], these switches undergo 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 Mg²⁺ are strongly reduced. In the initial X-ray crystallographic structure (PDB entry: 1PK0), the ion Yb³⁺ 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 Mg²⁺ and Nε2-H577 is stable along the trajectories EF_ade_CaM_Ca and EF_ade_Mg_CaM (Table S2) 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 interaction significantly present along all trajectories (Table S2). 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 replicas of EF_ade_CaM (D). The adefovir hydrogen bonds conserved with protein mainly involve protons of the amine groups 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 artefactual 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.

Analysis of the EF / CaM interface, as well as the interactions within the catalytic site, highlight the influence of the variation in the composition of the system on the accessible surface of the catalytic site, impacting the enzymatic activity. In addition, a network of interactions between EF and CaM residues detected in the initial X-ray crystallographic structure, undergo strong destabilization when co-factors are removed: this network could be considered as a plausible communication path for allosteric correlation between CaM and the EF catalytic site.

Analysis of cavities deformation to detect allosteric pockets

In this section, we investigate the possibility to use the cavity tracking implemented in the software *mkgridXf* [81] along EF MD trajectories in order 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 [32, 91] or described by a normal mode perturbation [92]. Also, a recent analysis of a large set of protein structures containing ligands showed [93] 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 [32, 37]. Allosterity has been thus repeatedly associated to larger local or global deformability. Beside, cavity tracking of *mkgridXf* [81] along MD trajectories made possible to correlate deformations of individual cavities to the principal components of the proteins global motions [94]. The association of the observation made by Desdouits et al [94] with the literature approaches conducted us to investigate in the present work on the reliability of analyzing the cavity deformation to predict allosteric sites.

A systematic analysis and tracking of the cavities present in the complex was performed along trajectories using *mkgridXf* [81, 94]. The cavities were determined by rolling probes as described in the Materials and Methods. Each cavity was tracked along MD trajectories using a description based on a consensus list of protein atoms delineating the cavities [81].

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 120-200ns 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 8). For most of the cavities, the volumes are smaller than 100\AA^3 for all trajectories. Few of them (indexes #5, #64, #83, #97, #130, #136, #140) display larger volumes up to 2500\AA^3 as well as large variations among the various trajectory conditions. The consensus residues

defining each of these cavities have been determined using a cutoff of 2.5 on the residue score and are listed in Table S3.

For both replicas, the cavity #140, 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), and the central α -helix (#64). The EF regions are the helical region (#5, #64, #83, #97, #130) and the region C^A (#136). Two cavities located at the interface between the EF-hands and the region C^A (#97 and #136) display smaller volumes. According to the rationale exposed at the beginning of the section and which we explore on the EF / CaM complex, the “variable cavities” cited above and located at the interface between CaM and different EF regions are allosteric cavity candidates.

In order to probe this proposition, we compared the approach proposed here to current approaches in the literature. The initial conformation of the EF/CaM complex was thus processed with different web servers able to predict druggable or allosteric pockets (Figure 9). The servers PARS [32], DeepSite [95], FTMap [11], POCASA [96], and CavityPlus [33] were used. Globally, the *mkgridXf* cavities 140 and 64, located respectively in the catalytic site and at the EF/CaM interface, correspond to predicted druggable pockets. The cavity 130, located in the helical domain, is present in some of the predicted sets of druggable pockets. Then, the server CavityPlus allows to identify the location of allosteric sites, using the CorrSite method [36]. Using the catalytic site as the orthosteric site, the corresponding allosteric pockets predicted by CavityPlus correspond to the *mkgridXf* cavities 64, 130 and 83, located at the EF/CaM interface and in the helical domain of EF. To summarize, the set of *mkgridXf* cavities with large volume variation contain druggable pockets, and three of

the *mkgridXf* cavities are predicted to be allosteric pockets with respect to the catalytic site. This last result supports the existence of an allosteric communication between the EF/CaM interface and the catalytic site.

As described in the introduction, a theoretical frame for detection of allostery has been developed during the last five years: the structure-based statistical mechanical model of allostery (SBSMMA) [37], based on a description of the protein as an elastic network between $C\alpha$. The effect of ligand is implicitly modelled through an additional energetic term added to the harmonic energy of the elastic network. The initial EF/CaM complex conformation has been processed on the AlloSigMA server (allosigma.bii.a-star.edu.sg/home) [34] implementing the SBSMMA model. The $\Delta h_i^{(m\uparrow)}$ energy values, describing the allosteric communication between protein residues have been plotted for all residues involved in the *mkgridXf* cavities 5, 64, 83, 97, 130, 136 and 140 (Figure 10). The probe residues sample the cavity residues (filled triangles) and, for all cavities except 140, positive $\Delta h_i^{(m\uparrow)}$ values describing a free energy change are observed for several residues located in the catalytic site (filled rectangle). The AlloSigMA results thus agree with an allosteric communication between *mkgridXf* cavities 5, 64, 83, 97, 130, 136 and the catalytic site.

The allosteric communication demonstrated by the variability of cavity volumes as well as by the energetics of the elastic network, is also supported by the variation of the network of interactions described in the previous section between the catalytic site and the helix V of CaM. This network is a plausible candidate for a communication path within the EF/CaM structure.

The cavities present in the EF/CaM complex were tracked along MD trajectories recorded with various perturbation conditions related to functional aspects of EF activity. In that way,

the EF/CaM interface has been pointed out as a region containing pockets allosteric with respect to the orthosteric catalytic site. The targeting of these pockets by virtual screening has higher chances to conduct to the discovery of allosteric ligands.

Discussion

In the present work, MD trajectories were recorded from the 1PK0 crystallographic structure of the EF/CaM complex, in the presence of various sets of co-factors: ions Ca^{2+} and Mg^{2+} and ligand adefovir.

The main finding from the comparison of trajectories is that the removal of ions has a strong effect on the conformations of the complex, at local and global levels. 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 interactions even in the presence of ion Mg^{2+} . This distant influence of ions agrees with the existence of a network of hydrogen bonds and stacking interactions which connects the catalytic pocket with the EF/CaM interface and which is destabilized if co-factors are removed. Noticeably a similar network of hydrogen bonds has been observed and validated using MD and mutagenesis [97] 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 [57], 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 have enforced the binding of adefovir inhibitor to the catalytic site or induced the specific conformation of adefovir in the site. Previous computational analyses [49, 98] already highlighted the artefactual character of some ions observed in the catalytic site of various EF X-ray crystallographic structures.

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 the EF hands also show greater variations in N-CaM than in C-CaM, which could be related to a weaker interaction between EF and N-CaM.

The tracking of *mkgridXf* cavities in the EF/CaM complex along MD trajectories revealed large variations of cavity volumes in two regions: (i) the catalytic site and (ii) the interface between EF and CaM. The analysis of initial EF/CaM complex structure in the frame of the model SBSMMA [37] has shown that the residues located in the *mkgridXf* cavities at the interface between EF and CaM, display allosteric communication with residues in the catalytic site. The procedure followed in the present manuscript has several points in common with the structure-based statistical mechanical model of allostery (SBSMMA). Indeed, SBSMMA states that the allostery can be quantitatively demonstrated using a model of elastic network which is deformed when the protein undergoes transition between unperturbed and perturbed states. In the present work, the unperturbed state is the MD trajectories recorded on the EF/CaM complex in presence of all co-factors, and the perturbed states are the MD trajectories recorded when one or more co-factors have been removed. In the frame of SBSMMA, the deformable cavities located at the interface between CaM and EF qualify them as being related in an allosteric way to the catalytic site and thus to the catalytic

function of EF. Consequently, ligands designed to bind such cavities could have an allosteric effect on the catalytic activity of EF. In that respect, one should note that an inhibitor of EF, the compound 10506-2A, has been claimed to bind to the helical region [62].

During the last decade, many approaches have been developed for detecting pockets susceptible to bind allosteric ligands, and allosteric paths through protein structures [32, 91, 92, 99–106]. 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 mode or elastic network analysis. In the present analysis, we decided to focus on the protein cavities. The relationship found here between the variability of *mkgridXf* cavity volumes and protein long distance communication is not surprising since such correlation has been observed previously [94] between *mkgridXf* cavities deformation and protein functional motions.

The proposition of using the volume variability of cavities for detection of allosteric communication has been only used here on the EF/CaM system. It is obvious that the application on a unique system does not constitute a general proof of concept. Nevertheless, 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 [32, 91]. Indeed, by contrast with the network where only one atom (generally $C\alpha$) 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 quadratic energy surface of the elastic network model. Finally, perturbing the system by removing co-factors, as ions, highly involved in the EF function, makes the observation of

protein deformation more specifically related to the inhibition of EF function. However, all these improvements are accessible at the cost of a very intense computational effort.

Acknowledgments

Irène Pitard thanks DGA for PhD financial support (grant 2017033). Authors thank CNRS and Institut Pasteur for recurrent funding. This work was also supported by the INCEPTION project ANR-16-CONV-0005.

References

- [1] Morelli X, Bourgeas R, Roche P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol* **15** (2011) 475–481.
- [2] 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** (2012) 20–33.
- [3] 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** (2014) 1291–1307.
- [4] Huang SY. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. *Drug Discov. Today* **19** (2014) 1081–1096.

- [5] Aguirre C, Cala O, Krimm I. Overview of Probing Protein-Ligand Interactions Using NMR. *Curr Protoc Protein Sci* **81** (2015) 1–17.
- [6] Kuenemann MA, Sperandio O, Labbé CM, Lagorce D, Miteva MA, Villoutreix BO. In silico design of low molecular weight protein-protein interaction inhibitors: Overall concept and recent advances. *Prog. Biophys. Mol. Biol.* **119** (2015) 20–32.
- [7] Fischer ES, Park E, Eck MJ, Thomä NH. SPLINTS: small-molecule protein ligand interface stabilizers. *Curr. Opin. Struct. Biol.* **37** (2016) 115–122.
- [8] Shin WH, Christoffer CW, Kihara D. In silico structure-based approaches to discover protein-protein interaction-targeting drugs. *Methods* **131** (2017) 22–32.
- [9] Vajda S, Beglov D, Wakefield AE, Egbert M, Whitty A. Cryptic binding sites on proteins: definition, detection, and druggability. *Curr Opin Chem Biol* **44** (2018) 1–8.
- [10] Beglov D, Hall DR, Wakefield AE, Luo L, Allen KN, Kozakov D, et al. Exploring the structural origins of cryptic sites on proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115** (2018) E3416–E3425.
- [11] Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **10** (2015) 733–755.
- [12] Cimermancic P, Weinkam P, Rettenmaier TJ, Bichmann L, Keedy DA, Woldeyes RA, et al. CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J. Mol. Biol.* **428** (2016) 709–719.

- [13] Ghanakota P, DasGupta D, Carlson HA. Free Energies and Entropies of Binding Sites Identified by MixMD Cosolvent Simulations. *J Chem Inf Model* **59** (2019) 2035–2045.
- [14] Martinez-Rosell G, Lovera S, Sands ZA, De Fabritiis G. PlayMolecule CrypticScout: Predicting Protein Cryptic Sites Using Mixed-Solvent Molecular Simulations. *J Chem Inf Model* **60** (2020) 2314–2324.
- [15] Comitani F, Gervasio FL. Exploring Cryptic Pockets Formation in Targets of Pharmaceutical Interest with SWISH. *J Chem Theory Comput* **14** (2018) 3321–3331.
- [16] Sun Z, Wakefield AE, Kolossvary I, Beglov D, Vajda S. Structure-Based Analysis of Cryptic-Site Opening. *Structure* **28** (2020) 223–235.
- [17] Tschammer N. Allosteric Modulators of the Class A G Protein Coupled Receptors. *Adv. Exp. Med. Biol.* **917** (2016) 185–207.
- [18] Deredge DJ, Huang W, Hui C, Matsumura H, Yue Z, Moënne-Loccoz P, 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** (2017) 3421–3426.
- [19] Feng C, Roy A, Post CB. Entropic allostery dominates the phosphorylation-dependent regulation of Syk tyrosine kinase release from immunoreceptor tyrosine-based activation motifs. *Protein Sci.* **27** (2018) 1780–1796.
- [20] Ni D, Liu N, Sheng C. Allosteric Modulators of Protein-Protein Interactions (PPIs). *Adv. Exp. Med. Biol.* **1163** (2019) 313–334.

- [21] Goodey NM, Benkovic SJ. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **4** (2008) 474–482.
- [22] Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12** (1965) 88–118.
- [23] Koshland DE, Némethy G, Filmer D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5** (1966) 365–385.
- [24] Liu J, Nussinov R. Allosteric: An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput. Biol.* **12** (2016) e1004966.
- [25] Wodak SJ, Paci E, Dokholyan NV, Berezovsky IN, Horovitz A, Li J, et al. Allosteric in Its Many Disguises: From Theory to Applications. *Structure* **27** (2019) 566–578.
- [26] van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat. Methods* **10** (2013) 896–902.
- [27] Gur M, Madura JD, Bahar I. Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase. *Biophys. J.* **105** (2013) 1643–1652.
- [28] Raman AS, White KI, Ranganathan R. Origins of Allosteric and Evolvability in Proteins: A Case Study. *Cell* **166** (2016) 468–480.
- [29] Nussinov R, Tsai CJ. Allosteric in disease and in drug discovery. *Cell* **153** (2013) 293–305.

- [30] Laine E, Goncalves C, Karst J, Lesnard A, Rault S, Tang W, et al. Use of allostery to identify inhibitors of calmodulin- induced activation of *Bacillus anthracis* Edema Factor. *Proc Natl Acad Sci USA* **107** (2010) 11277–11282.
- [31] J Zhang and R Nussinov, editor. *Protein Allostery in Drug Discovery*. Advances in Experimental Medicine and Biology (Springer) (2019).
- [32] Panjkovich A, Daura X. PARS: a web server for the prediction of Protein Allosteric and Regulatory Sites. *Bioinformatics* **30** (2014) 1314–1315.
- [33] Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, et al. CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Res.* **46** (2018) W374–W379.
- [34] Tan ZW, Guarnera E, Tee WV, Berezovsky IN. AlloSigMA 2: paving the way to designing allosteric effectors and to exploring allosteric effects of mutations. *Nucleic Acids Res.* **48** (2020) W116–W124.
- [35] Ghode A, Gross L, Tee W, Guarnera E, Berezovsky I, Biondi R, et al. Synergistic allostery in multiligand-protein interactions. *Biophys. J.* (2020) in press.
- [36] Ma X, Meng H, Lai L. Motions of Allosteric and Orthosteric Ligand-Binding Sites in Proteins are Highly Correlated. *J Chem Inf Model* **56** (2016) 1725–1733.
- [37] Guarnera E, Berezovsky IN. Toward Comprehensive Allosteric Control over Protein Activity. *Structure* **27** (2019) 866–878.

- [38] 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* **278** (2003) 29261–29266.
- [39] Shen Y, Lee Y, Soelaiman S, Bergson P, Lu D, Chen A, et al. Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylyl cyclase exotoxins. *EMBO J* **21** (2002) 6721–6732.
- [40] Shen Y, Lee Y, Soelaiman S, Bergson P, Lu D, Chen A, et al. Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylyl cyclase exotoxins. *EMBO J* **21** (2002) 6721–6732.
- [41] 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* **283** (2008) 23836–23845.
- [42] Drum C, Yan S, Sarac R, Mabuchi Y, Beckingham K, Bohm A, 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** (2000) 36334–36340.
- [43] Drum C, Shen Y, Rice P, Bohm A, Tang W. Crystallization and preliminary X-ray study of the edema factor exotoxin adenylyl cyclase domain from *Bacillus anthracis* in the presence of its activator, calmodulin. *Acta Crystallogr D Biol Crystallogr* **57** (2001) 1881–1884.
- [44] Drum C, Yan S, Bard J, Shen Y, Lu D, Soelaiman S, et al. Structural basis for the

- activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature* **415** (2002) 396–402.
- [45] Shen Y, Guo Q, Zhukovskaya NL, Drum CL, Bohm A, Tang WJ. Structure of anthrax edema factor-calmodulin-adenosine 5'-(alpha,beta-methylene)-triphosphate complex reveals an alternative mode of ATP binding to the catalytic site. *Biochem. Biophys. Res. Commun.* **317** (2004) 309–314.
- [46] Guo Q, Shen Y, Zhukovskaya NL, Florian J, Tang WJ. Structural and kinetic analyses of the interaction of anthrax adenylyl cyclase toxin with reaction products cAMP and pyrophosphate. *J. Biol. Chem.* **279** (2004) 29427–29435.
- [47] Shen Y, Zhukovskaya NL, Guo Q, Florian J, Tang WJ. Calcium-independent calmodulin binding and two-metal-ion catalytic mechanism of anthrax edema factor. *EMBO J.* **24** (2005) 929–941.
- [48] 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* **71** (2008) 1813–1829.
- [49] 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* **77** (2009) 971–983.
- [50] Laine E, Blondel A, Malliavin T. Dynamics and energetics: a consensus analysis of the impact of calcium on EF-CaM protein complex. *Biophys J* **96** (2009) 1249–1263.

- [51] Laine E, Blondel A, Malliavin T. Dynamics and energetics: a consensus analysis of the impact of calcium on EF-CaM protein complex. *Biophys J* **96** (2009) 1249–1263.
- [52] 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* **99** (2010) 2264–2272.
- [53] Laine E, Ladant LMD, 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** (2012) 580–604.
- [54] Bertini I, Del Bianco C, Gelis I, Katsaros N, Luchinat C, Parigi G, et al. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc. Natl. Acad. Sci. U.S.A.* **101** (2004) 6841–6846.
- [55] Anthis NJ, Doucleff M, Clore GM. 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** (2011) 18966–18974.
- [56] Soelaiman S, Wei BQ, Bergson P, Lee YS, Shen Y, Mrksich M, et al. Structure-based inhibitor discovery against adenylyl cyclase toxins from pathogenic bacteria that cause anthrax and whooping cough. *J. Biol. Chem.* **278** (2003) 25990–25997.
- [57] Shen Y, Zhukovskaya NL, Zimmer MI, Soelaiman S, Bergson P, Wang CR, et al. Selective inhibition of anthrax edema factor by adefovir, a drug for chronic hepatitis B virus infection. *Proc. Natl. Acad. Sci. U.S.A.* **101** (2004) 3242–3247.

- [58] Chen Z, Moayeri M, Zhao H, Crown D, Leppla SH, Purcell RH. 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** (2009) 13487–13492.
- [59] Geduhn J, Dove S, Shen Y, Tang WJ, 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.* **336** (2011) 104–115.
- [60] Taha H, Dove S, Geduhn J, Konig B, Shen Y, Tang WJ, 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 Schmiedebergs Arch. Pharmacol.* **385** (2012) 57–68.
- [61] Taha HM, Schmidt J, Gottle M, Suryanarayana S, Shen Y, Tang WJ, 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** (2009) 693–703.
- [62] Lee YS, Bergson P, He WS, Mrksich M, Tang WJ. Discovery of a small molecule that inhibits the interaction of anthrax edema factor with its cellular activator, calmodulin. *Chem. Biol.* **11** (2004) 1139–1146.
- [63] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66** (2010) 12–21.

- [64] Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, 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** (2016) 405–413.
- [65] Jo S, Cheng X, Lee J, Kim S, Park SJ, Patel DS, et al. CHARMM-GUI 10 years for biomolecular modeling and simulation. *J Comput Chem* **38** (2017) 1114–1124.
- [66] MacKerell A, Bashford D, Bellott M, Dunbrack R, Evanseck J, Field M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102** (1998) 3586–3616.
- [67] MacKerell A, Feig M, Brooks C. Extending the treatment of backbone energetics in protein force fields and limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comp Chem* **25** (2004) 1400–1415.
- [68] Best R, Zhu X, Shim J, Lopes P, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi1 and chi2 dihedral angles. *J Chem Theor Comput* **8** (2012) 3257–3273.
- [69] Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79** (1983) 926–935.
- [70] Kim S, Lee J, Jo S, Brooks CL, Lee HS, Im W. CHARMM-GUI ligand reader and

- modeler for CHARMM force field generation of small molecules. *J Comput Chem* **38** (2017) 1879–1886.
- [71] Phillips J, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem* **26** (2005) 1781–1802.
- [72] Darden T, York D, Pedersen L. Particle Mesh Ewald and an N.log(N) method for Ewald sums in large systems. *J Chem Phys* **98** (1993) 3684–90.
- [73] Ryckaert J, Ciccotti G, Berendsen H. Numerical integration of the cartesian equations of motion of a system with constraints and Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23** (1977) 327–341.
- [74] Andersen H. Rattle: a "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J Comp Phys* **52** (1983) 24–34.
- [75] Roe D, TE Cheatham I. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9** (2013) 3084–3095.
- [76] Michaud-Agrawal N, Denning E, Woolf T, Beckstein O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem* **32** (2011) 2319–2327.
- [77] Gowers R, Linke M, Barnoud J, Reddy T, Melo M, Seyler S, et al. MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference, Austin, TX, 2016* **32** (2016) 102–109.
- [78] Mitternacht S. FreeSASA: An open source C library for solvent accessible surface area calculation. *F1000Research* **5** (2016) 189.

- [79] Yang C, Jas G, Kuczera K. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *Biochimica and Biophysica Acta* **1697** (2004) 289–300.
- [80] Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55** (1971) 379–400.
- [81] 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** (2019) 3506–3518.
- [82] Yamniuk AP, Vogel HJ. Calmodulin’s flexibility allows for promiscuity in its interactions with target proteins and peptides. *Mol. Biotechnol.* **27** (2004) 33–57.
- [83] Crivici A, Ikura M. Molecular and structural basis of target recognition by calmodulin. *Annu Rev Biophys Biomol Struct* **24** (1995) 85–116.
- [84] Zhang M, Tanaka T, Ikura M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat Struct Biol* **2** (1995) 758–767.
- [85] Finn B, Evenas J, Drakenberg T, Waltho J, Thulin E, Forsen S. Calcium-induced structural changes and domain autonomy in calmodulin. *Nat Struct Biol* **2** (1995) 777–783.
- [86] Komeiji Y, Ueno Y, Uebayasi M. Molecular dynamics simulations revealed Ca(2+)-dependent conformational change of Calmodulin. *FEBS Lett.* **521** (2002) 133–139.
- [87] Ikura M, Clore GM, Gronenborn AM, Zhu G, Klee CB, Bax A. Solution structure of

- a calmodulin-target peptide complex by multidimensional NMR. *Science* **256** (1992) 632–638.
- [88] Barbato G, Ikura M, Kay LE, Pastor RW, Bax A. Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry* **31** (1992) 5269–5278.
- [89] Zhang M, Tanaka T, Ikura M. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat. Struct. Biol.* **2** (1995) 758–767.
- [90] Finn BE, Evenäs J, Drakenberg T, Waltho JP, Thulin E, Forsén S. Calcium-induced structural changes and domain autonomy in calmodulin. *Nat. Struct. Biol.* **2** (1995) 777–783.
- [91] Guarnera E, Berezovsky IN. Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. *PLoS Comput. Biol.* **12** (2016) e1004678.
- [92] Greener JG, Sternberg MJ. AlloPred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinformatics* **16** (2015) 335.
- [93] Alfayate A, Rodriguez Caceres C, Gomes Dos Santos H, Bastolla U. Predicted dynamical couplings of protein residues characterize catalysis, transport and allostery. *Bioinformatics* **35** (2019) 4971–4978.
- [94] Desdouits N, Nilges M, Blondel A. Principal Component Analysis reveals correlation of cavities evolution and functional motions in proteins. *J. Mol. Graph. Model.* **55** (2015) 13–24.

- [95] Jiménez J, Doerr S, Martínez-Rosell G, Rose AS, De Fabritiis G. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics* **33** (2017) 3036–3042.
- [96] Yu J, Zhou Y, Tanaka I, Yao M. Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics* **26** (2010) 46–52.
- [97] Selwa E, Davi M, Chenal A, Sotomayor-Pérez A, Ladant D, Malliavin T. Allosteric activation of *Bordetella pertussis* adenylyl cyclase by calmodulin: molecular dynamics and mutagenesis studies. *J Biol Chem* **289** (2014) 21131–21141.
- [98] 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.
- [99] Daily MD, Gray JJ. Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. *PLoS Comput. Biol.* **5** (2009) e1000293.
- [100] Mitternacht S, Berezovsky IN. Binding leverage as a molecular basis for allosteric regulation. *PLoS Comput. Biol.* **7** (2011) e1002148.
- [101] Bowman GR, Geissler PL. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U.S.A.* **109** (2012) 11681–11686.
- [102] Clarke D, Sethi A, Li S, Kumar S, Chang RWF, Chen J, et al. Identifying Allosteric Hotspots with Dynamics: Application to Inter- and Intra-species Conservation. *Structure* **24** (2016) 826–837.

- [103] Song K, Liu X, Huang W, Lu S, Shen Q, Zhang L, et al. Improved Method for the Identification and Validation of Allosteric Sites. *J Chem Inf Model* **57** (2017) 2358–2363.
- [104] Pflieger C, Mingos A, Boehm M, McClendon CL, Torella R, Gohlke H. Ensemble and Rigidity Theory-Based Perturbation Approach To Analyze Dynamic Allostery. *J Chem Theory Comput* **13** (2017) 6343–6357.
- [105] Huang M, Song K, Liu X, Lu S, Shen Q, Wang R, et al. AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses. *Nucleic Acids Res.* **46** (2018) W451–W458.
- [106] Abrusan G, Marsh JA. Ligand-Binding-Site Structure Shapes Allosteric Signal Transduction and the Evolution of Allostery in Protein Complexes. *Mol. Biol. Evol.* **36** (2019) 1711–1727.

System	solute	number of TIP3P waters	neutralizing K ⁺ ions	total number of atoms
EF_ade_Mg_CaM_Ca	EF, CaM, Mg ²⁺ , 2 Ca ²⁺ , adefovir	68219	9	215183
EF_ade_CaM_Ca	EF, CaM, 2 Ca ²⁺ , adefovir	68220	11	215187
EF_ade_Mg_CaM	EF, CaM, Mg ²⁺ , adefovir	69773	13	219847
EF_ade_CaM	EF, CaM, adefovir	69795	15	219914
EF_CaM_Ca	EF, CaM, 2 Ca ²⁺	69849	11	220034
EF_CaM	EF, CaM	69785	15	219844

Table I: Systems composition.

Figure 1

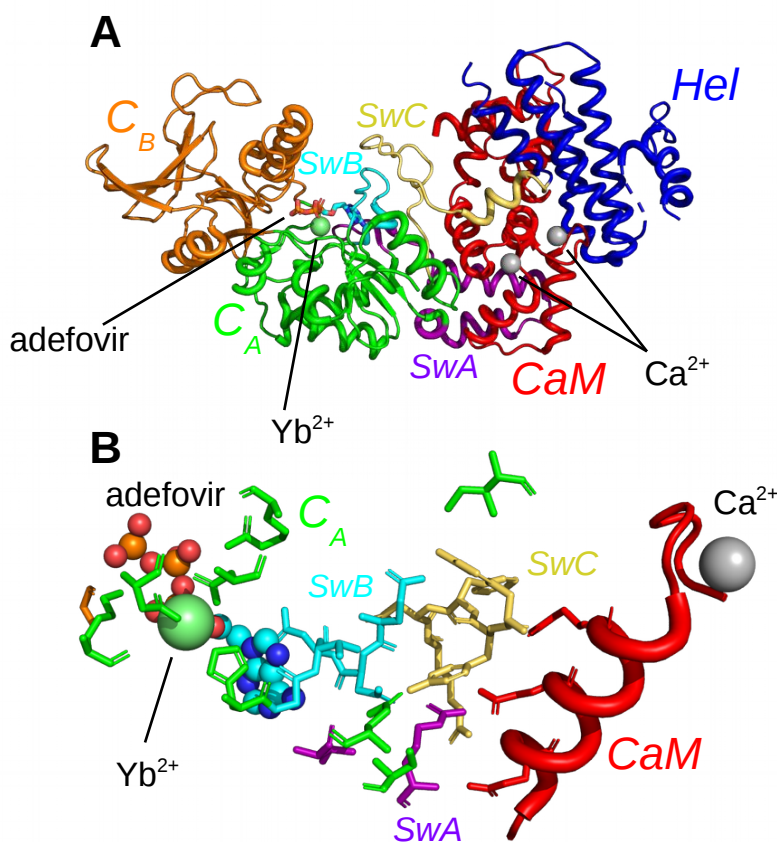


Figure 1: A. X-ray crystallographic structure (PDB entry: 1PK0) of the complex EF/CaM with the ions Ca^{2+} colored in gray and the ion Yb^{2+} colored in lime. The ligand adefovir is drawn in ball-and-sticks, the regions C_A (residues 292-349 and 490-622), C_B (residues 350-489) and helical (residues 660-800, labeled "Hel" on the figure). 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. The helix V of CaM is colored in salmon.

B. Network of residues in the X-ray crystallographic structure (PDB entry: 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 than in A.

Figure 2

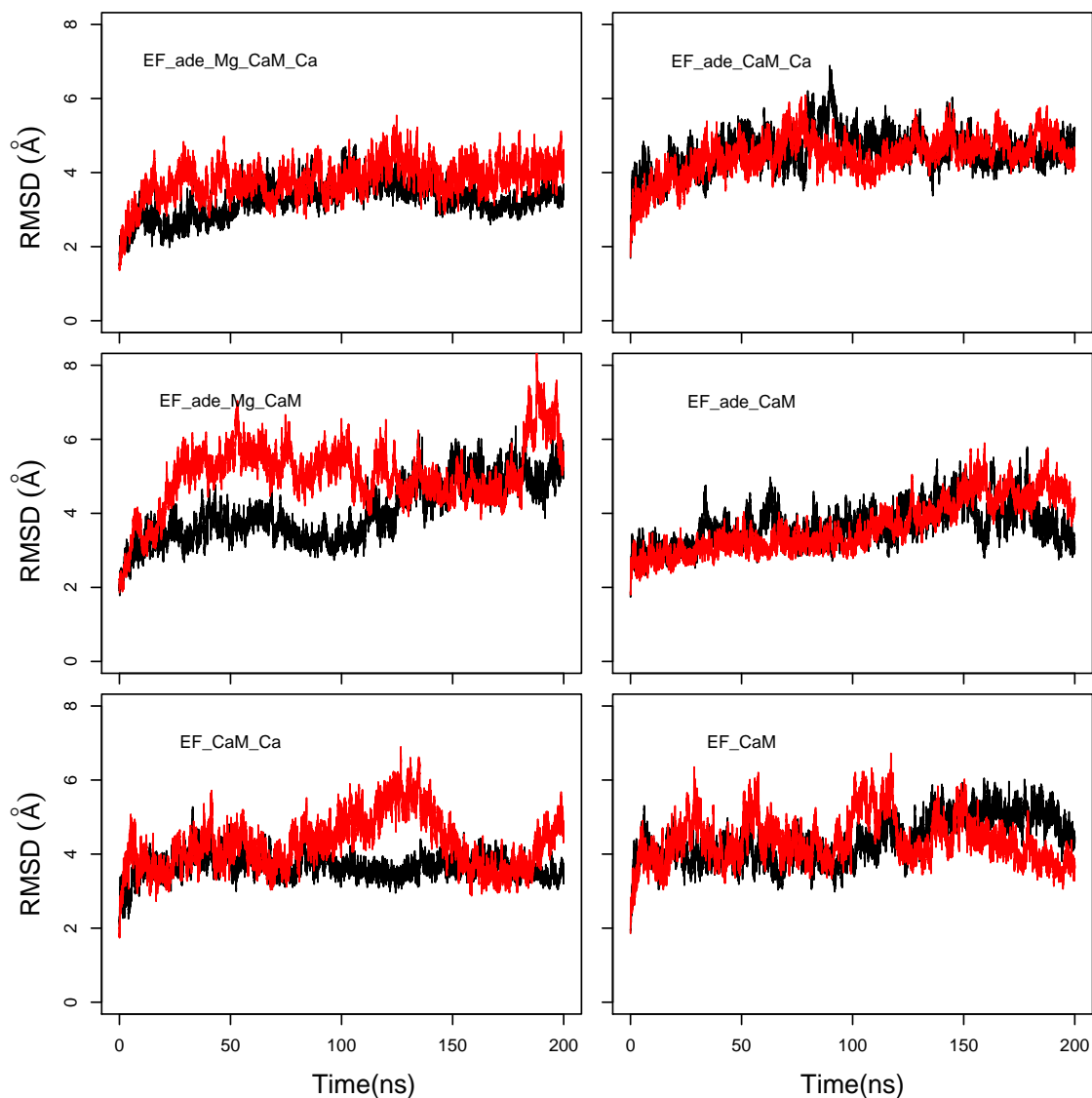


Figure 2: Coordinate root-mean-square deviations (RMSD: Å) of the backbone heavy atoms of EF and CaM with respect to the PDB structure 1PK0, calculated along all trajectories. For each trajectory, the black and red curves correspond to the two replicas R1 and R2.

Figure 3

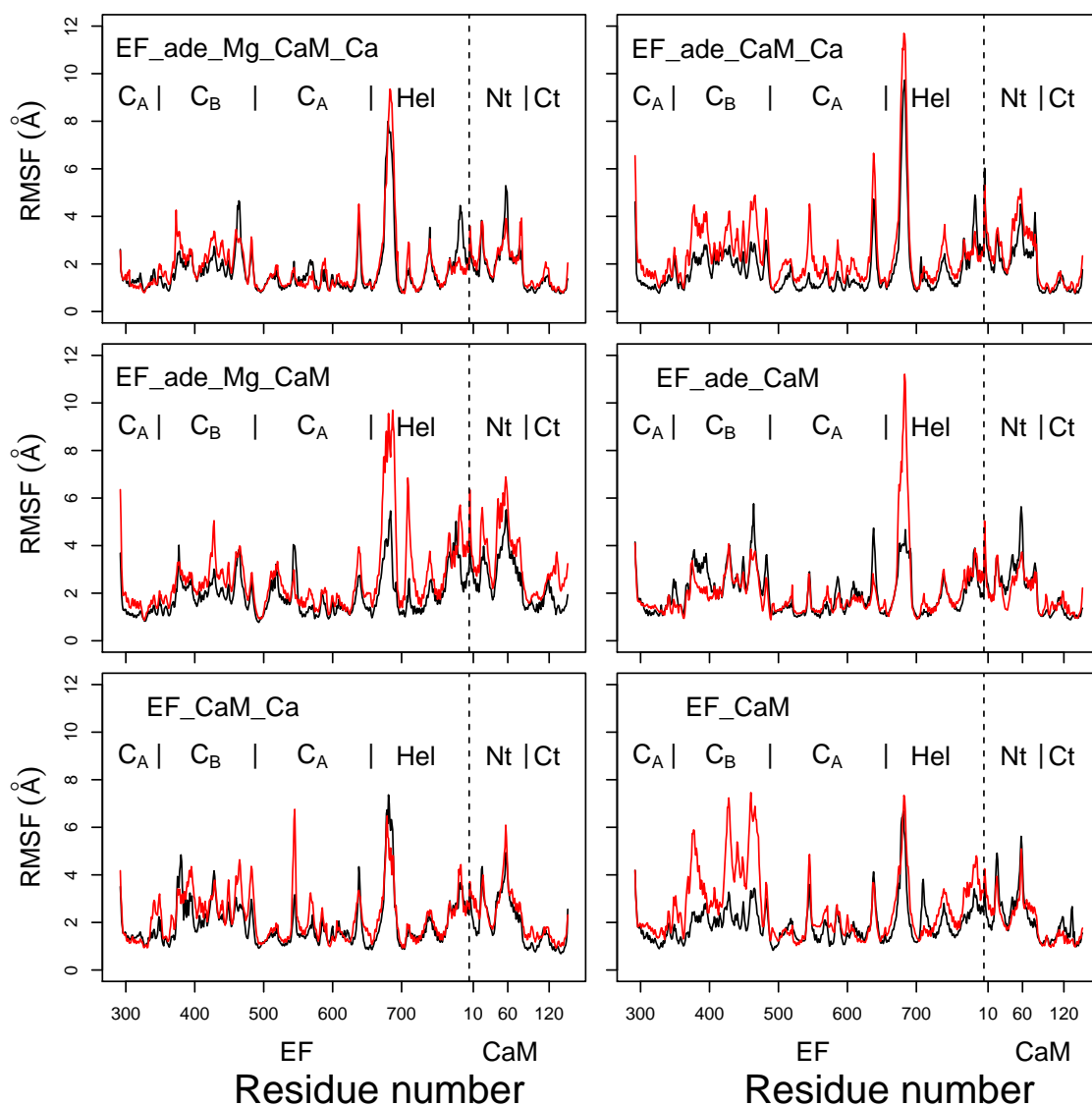


Figure 3: Coordinate root-mean-square fluctuations (RMSF: Å) of the backbone heavy atoms of the complex EF/CaM, each frame being superimposed on the chains A and D of the PDB entry 1PK0. For each trajectory, the black and red curves correspond to replicas R1 and R2. The N terminal residue of CaM is indicated 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.

Figure 4

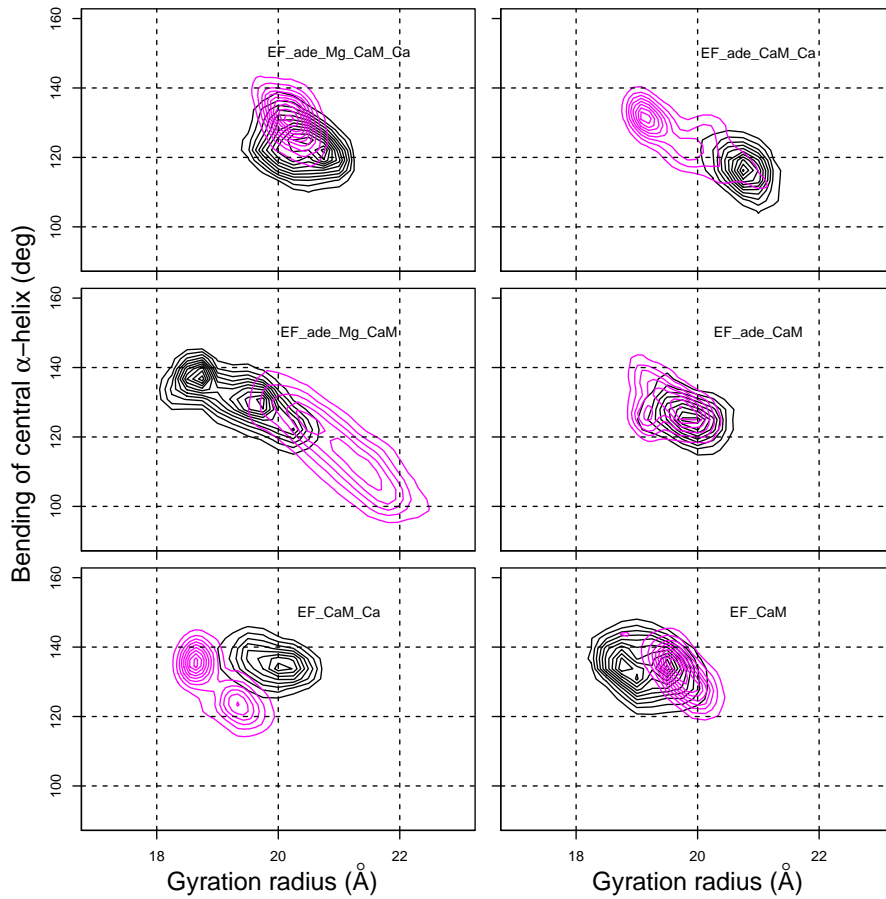


Figure 4: Contour plots describing the variation of the gyration radius (\AA) with respect to the bending angle of the CaM central α -helix (deg) determined as the angles between axes of α helices IV (residues 66-73) and V (residues 83-92). **The contour lines describe the joint probability distribution of these two parameters.** For each trajectory, the black and magenta contours correspond to replicas R1 and R2.

Figure 5

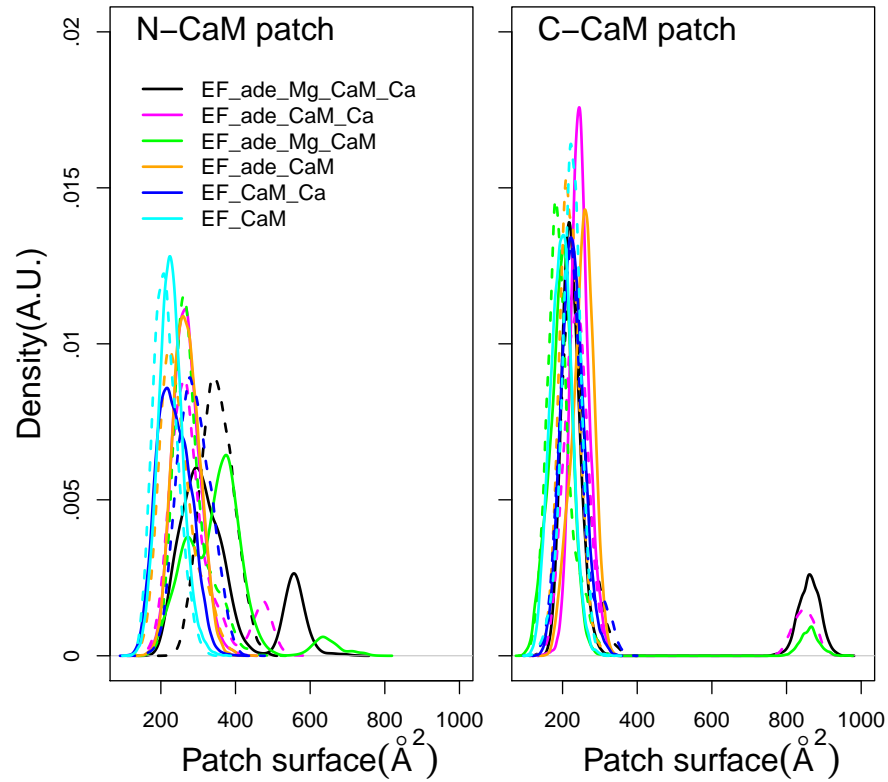


Figure 5: Distribution of accessible surface (\AA^2) of the hydrophobic patch of N-CaM and C-CaM. The N-CaM patch is defined by the residues A10, F12, A15, L18, F19, L32, M36, L39, M51, V55, M71, M72, M76. The C-CaM patch is defined by the residues I85, A88, V91, F92, L105, M109, L112, L116, M124, F141, M144, M145, A147. For each trajectory, the plain and dashed curves correspond to replicas R1 and R2.

Figure 6

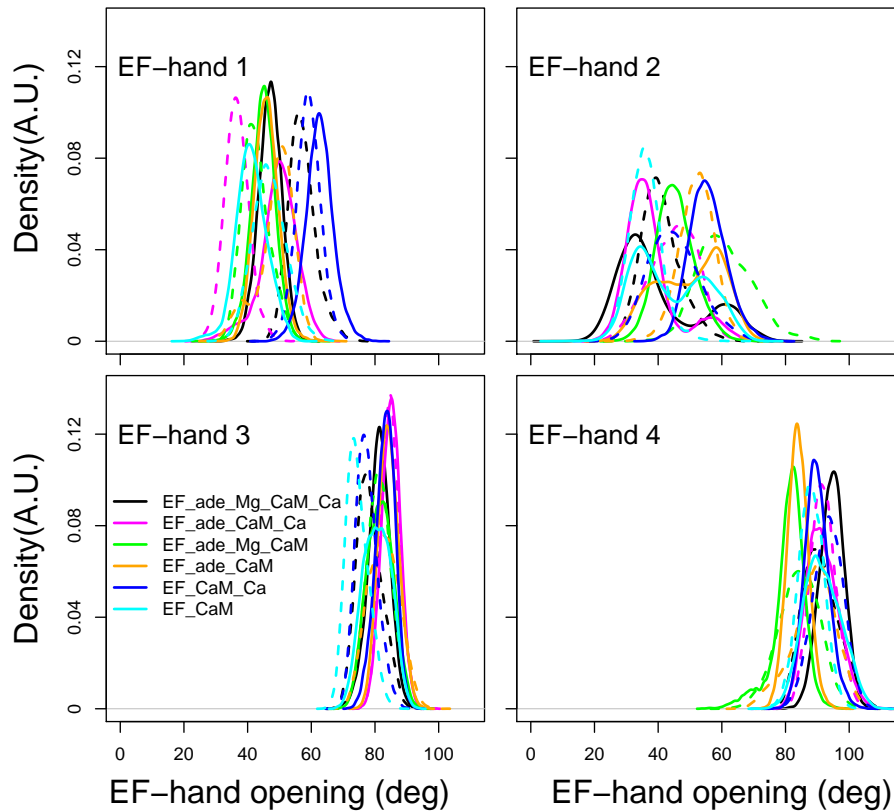


Figure 6: Angle of the EF-hands (deg) calculated between axes of α helices including residues 8-19 (helix I), 31-37 (helix II), 46-53 (helix III), 66-73 (helix IV), 83-92 (helix V), 103-110 (helix VI), 119-127 (helix VII), 139-145 (helix VIII). The angle of EF-hand 1 is the angle between helices I and 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.

Figure 7

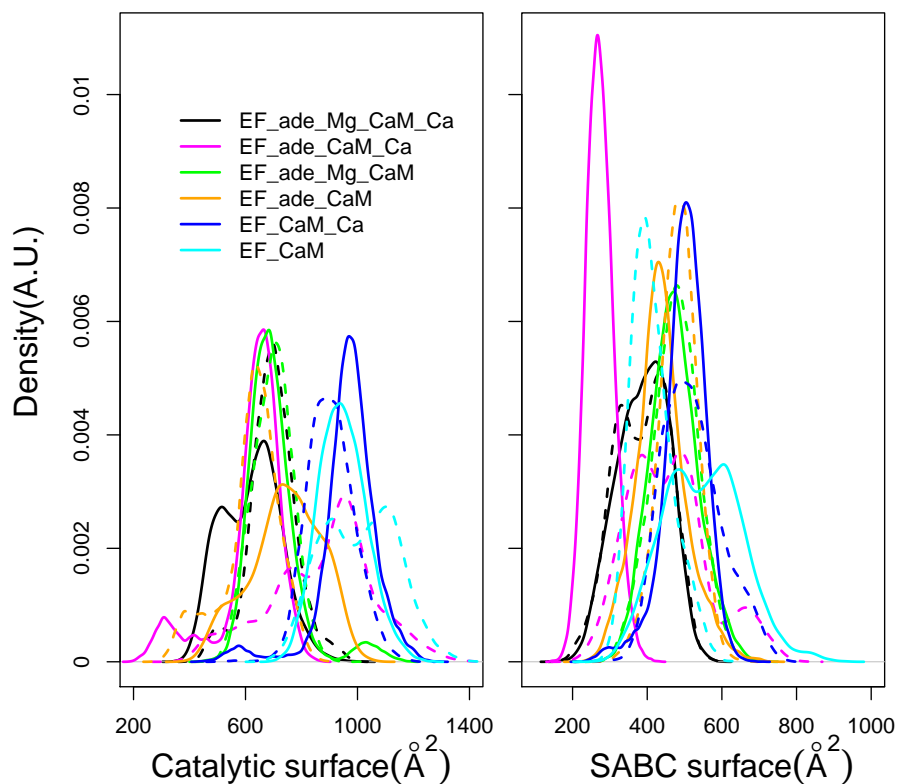


Figure 7: Left. Distribution of accessible catalytic surface (\AA^2) of EF, calculated as the sum of the accessible surfaces of residues H351, K353, S354, K372, R329, K346, L348, D491, D493, H577, G578, T579, D582, N583, E588, F586, T548 of EF. Right. Distribution of accessible surface (\AA^2) of the pocket SABC, 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.

Figure 8

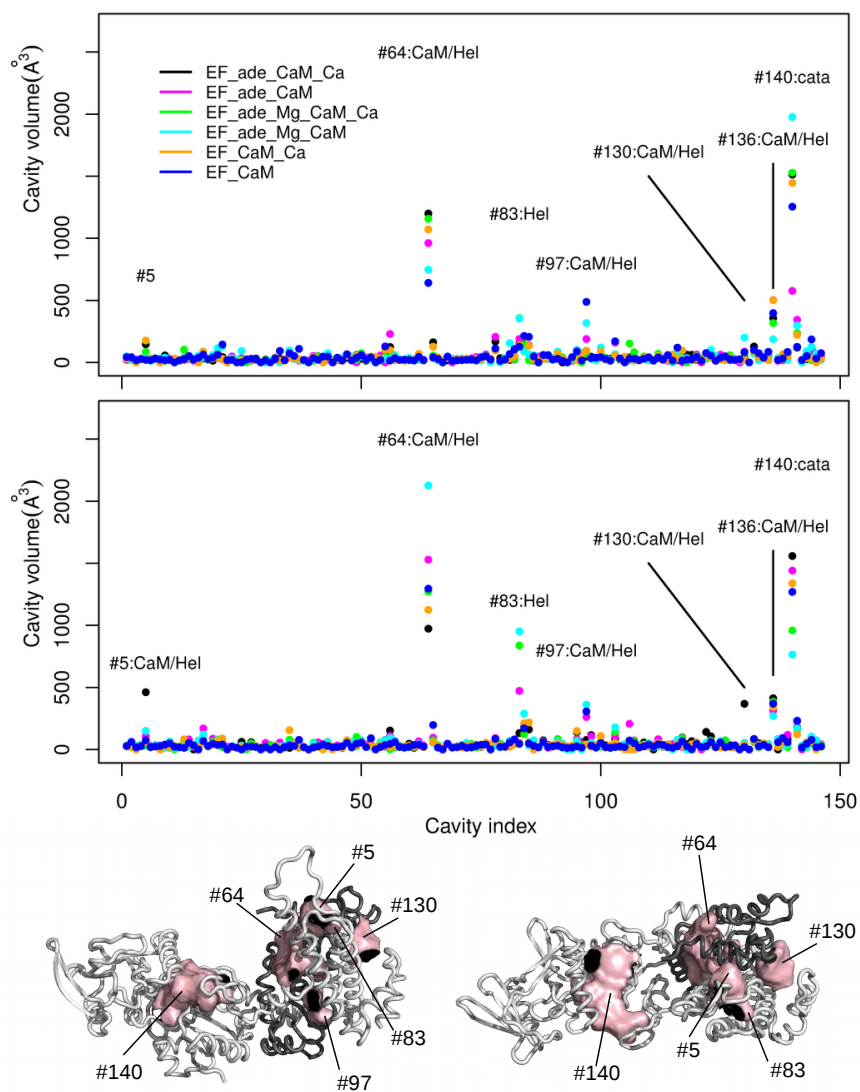


Figure 8: Top. Averaged volumes (\AA^3) of the cavities detected by *mkgridXf* plotted along the cavity index. The two plots correspond to the two replicas series of the trajectories. The points are colored according to the trajectory on which the volume was averaged: EF_ade_CaM_Ca (black), EF_ade_CaM (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\AA^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. CaM is colored in dark grey.

Figure 9

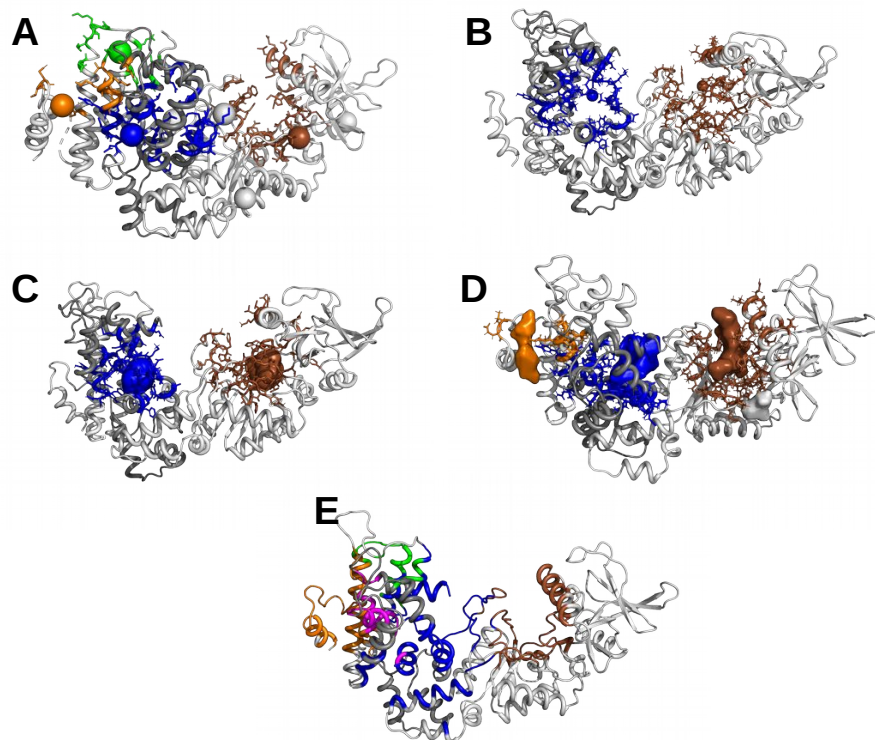


Figure 9: Prediction of pockets druggability or allosterity on the initial EF/CaM structure 1PK0 using several methods: A. PARS (bioinf.uab.cat/cgi-bin/pars-cgi/pars.pl), prediction of the druggable pockets, three of them correspond to *mkgridXf* cavities: 64 (blue), 130 (orange) and 140 (brown). B. Deepsite (www.playmolecule.com/deepsite), prediction of two druggable pockets, displaying scores of 0.999 and corresponding to *mkgridXf* cavities: 64 (blue) and 140 (brown). C. FTMap (ftsites.bu.edu), prediction of two druggable pockets, corresponding to *mkgridXf* cavities: 64 (blue) and 140 (brown). D. POCASA (altair.sci.hokudai.ac.jp/g6/Research/POCASA_e.html), prediction of five druggable pockets, two of them corresponding to the *mkgridXf* cavity 140 (brown), and two others to the *mkgridXf* cavities: 64 (blue) and 130 (orange). E. CavityPlus (www.pkumdl.cn:8000/cavityplus/index.php), prediction of the allosteric pockets influenced by the orthosteric pocket located in the catalytic site (brown). Four allosteric pockets are predicted with the corresponding scores: Cavity1 (3.36), Cavity5 (1.50), Cavity10 (0.75), Cavity4 (0.70). These cavities correspond to the *mkgridXf* cavities: 64 (blue), 130 (orange/magenta), 83 (green).

Figure 10

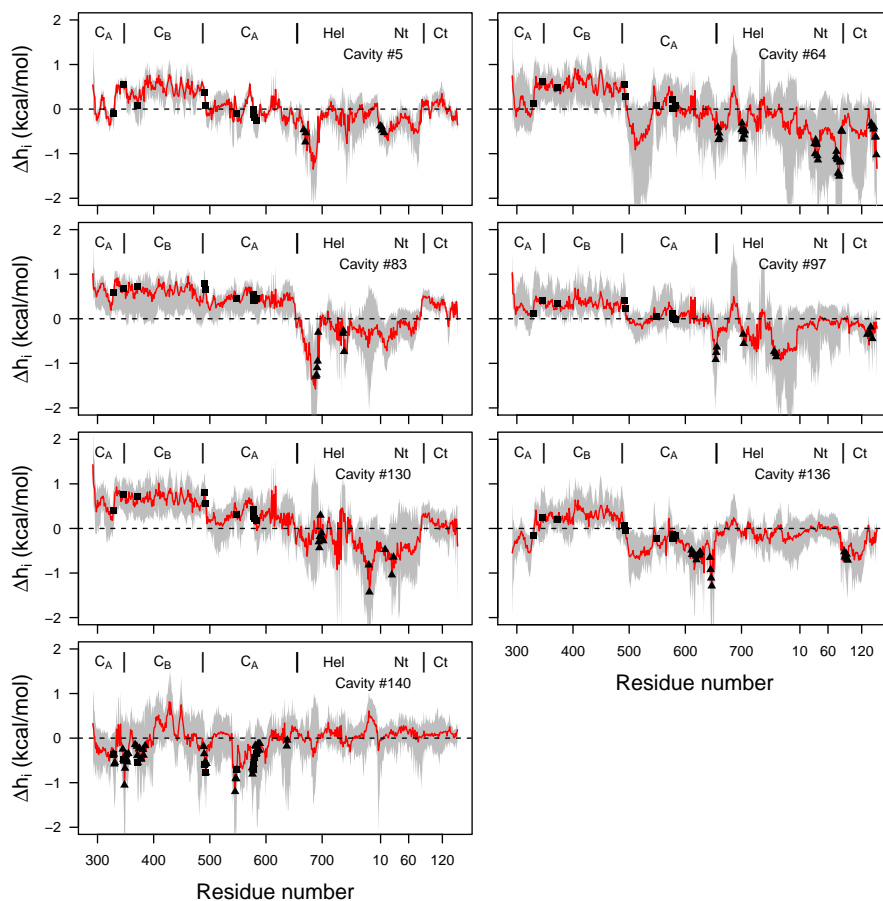


Figure 10: Plot of free energy $\Delta h_i^{(m\uparrow)}$ (kcal/mol) profile, where the probe residue m samples the residues defining each of the selected cavities (Table S3), and i samples all residues of the complex (x-axis). The average free energy profile is plotted in red with range of values drawn in grey. The residues defining the cavity (Table S3) are shown with filled triangles, and the residues 329,346,372,577,491, 493,548,578,579,583 of the catalytic site are shown with filled rectangles. The dashed line corresponds to the zero energy value.