May 5, 2014

Short title: TAMD simulations to explore AC conformations

Keywords: enhanced sampling; protein hydratation; adenyl cyclase; Bordetella pertussis; collective variables

The institution at which the work was performed: Institut Pasteur, Paris, France

Contact information for the author responsible for correspondence:

Thérèse E Malliavin

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

E-mail address: terez@pasteur.fr

Temperature-accelerated molecular dynamics gives insights into globular conformations sampled in the free state of the AC catalytic domain

Edithe Selwa[1,5], Tru Huynh[1], Giovanni Ciccotti[2,3], Luca Maragliano[4*] and Thérèse E Malliavin[1*]

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

[2] Department of Physics, University of Roma "La Sapienza", Italy

[3] School of Physics, University College Dublin, Ireland

[4] Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia,

Genoa, Italy

[5] Université Pierre et Marie Curie, Cellule Pasteur UPMC, rue du Dr Roux, 75015 Paris, France

* These authors contributed equally to this work

Key words: *Bordetella pertussis*, adenyl cyclase, molecular dynamics, temperature-accelerated molecular dynamics, water organization

Abstract

The catalytic domain of the adenyl cyclase (AC) toxin from *Bordetella pertussis* is activated by interaction with calmodulin (CaM), resulting in cAMP overproduction in the infected cell. In the X-ray crystallographic structure of the complex between AC and the C terminal lobe of calmodulin, the toxin displays a markedly elongated shape. As for the structure of the isolated protein, experimental results support the hypothesis that more globular conformations are sampled, but information at atomic resolution is still lacking. Here we employ temperature-accelerated molecular dynamics (TAMD) simulations to generate putative all-atom models of globular conformations sampled by CaM-free AC. As collective variables we use centers of mass coordinates of groups of residues selected from the analysis of standard MD simulations. Results show that TAMD allows extended conformational sampling and generates AC conformations that are more globular than in the complexed state. These structures are then refined via energy minimization and further unrestrained MD simulations to optimize inter-domain packing interactions, thus resulting in the identification of a set of hydrogen bonds present in the globular conformations.

INTRODUCTION

Bordetella pertussis is the bacterial pathogen responsible for pertussis, or whooping cough, a respiratory disease that despite the presence of a vaccine has made a dramatic comeback in the recent years [1, 2]. The infection is developed through the following molecular mechanism. *B. pertussis* inhibits the host immune system by the action of its adenyl cyclase, CyaA. The toxic protein enters the attacked cell in an inactivated form, gets activated by association with endogenous calmodulin (CaM), and triggers overproduction of adenosine monophosphate (cAMP), which in high concentration perturbs the cell signaling system making its immune response inefficient.

Given its pivotal role in the infection pathway, an atomic-detailed knowledge of CyaA structure and interactions can have great impact in developing inhibitors to fight *B. pertussis* [3]. Indeed, in the case of the structurally related adenyl cyclase from *Bacillus anthracis*, EF, a detailed understanding of the CaM dependent activation conformational transition helped in developing a new family of inhibitors, the thiophen ureido acids [4].

Currently, a high-resolution X-ray crystallographic structure of CyaA catalytic domain (AC) in complex with the C-terminal lobe of CaM (C-CaM) is available [5] (Figure 1). The structure shows that AC comprises three separate domains, named CA, CB and SA (respectively green, orange and purple in Figure 1). The SA domain includes the α-helices H, F, G and H'. In the complex, C-CaM docks on the H helix, making also contacts with the CA domain. The CaM-associated AC conformation is remarkably elongated, with the SA helices roughly parallel to an imaginary

axis joining the geometrical centers of CA and CB domains. To date, no experimentally determined structure is available for the isolated, CaM-free, AC protein. Indeed, unbound AC domains tend to aggregate, thus preventing crystallization and a straightforward analysis of SAXS curves (A. Chenal, personal communication). Extensive biophysical analyses on CyaA activation, however, provided the following information [6]: (i) the CaM bound and unbound AC conformations show a very similar secondary structure content, (ii) the unbound conformation displays a more globular shape than the bound state, with hydrodynamic parameters similar to that of a sphere, (iii) the transition to a more globular shape might be caused by a docking of the α helix H against the CA domain, with an angular displacement of the helix axis that could reach 90°. On the other hand, a study of the fluorescence of Trp-242, located in the helix H, has revealed [7] in the isolated AC a large internal mobility of this residue, which seems to indicate that the isolated AC undergoes a conformational equilibrium between several protein conformations.

Recently, standard molecular dynamics (MD) simulations of the isolated AC were performed [8], using as starting structure its conformation in complex with C-CaM. Results show that once separated from CaM the AC structure becomes less elongated by moving SA and CB with respect to CA. However, the MD time-scale of few tens of nanoseconds did not allow to reach a conformation where SA is significantly close to CA, thus suggesting that the conformational transition occurs in the hundreds of ns or μs time-scale at least.

During the last fifteen years, the importance of functional conformational changes in proteins has been highlighted in several cases [9–16]. Although very important for

protein function, these changes can be extremely difficult to characterize at atomic resolution, as the biophysical methods most used for protein structure determination most often require conformational stability. Computational methods as MD simulations offer a reliable alternative to obtain insight on transient conformations, but, unfortunately, the time-scales sampled by standard MD simulations are still much shorter than those of the biologically relevant conformational transitions, because a system spends most of its time being trapped in metastable basins by (free) energy barriers. Indeed, even the longest trajectories obtained using purposely engineered computing clusters [17] sample only a relatively small number of reactive events.

Numerous *enhanced sampling* techniques have been developed which aim to directly simulate protein conformational changes [18–32]. Temperature Accelerated Molecular Dynamics (TAMD) [33] is a more recently proposed method that circumvents some of the limitations of the algorithms mentioned above, in particular the restriction to a small number of collective variables. In TAMD, a set of extra variables are introduced, coupled to the original system's coordinates via collective variables (CVs). The original and new variables are then evolved concurrently in condition of effective adiabatic separation and at two different temperatures, the physical temperature for the original system and a higher, artificial temperature for the new variables. In this way, the system navigates the free energy landscape associated to the extra variables overcoming barriers that are even much higher than the energy at the physical temperature. TAMD can be employed to reconstruct the free energy landscape from direct sampling via re-weighting [33, 34], or using a mean-force interpolation method [35]. TAMD and its extensions have already been

applied to a variety of rare events sampling studies [36–43] and have proved to be particularly useful in protein conformational searches [44–50].

In the present work, we use TAMD to explore the conformational space of the CaM-free CyaA catalytic domain (AC), in search of globular metastable conformations. To define the collective variables, we propose an approach based on: (i) the analysis of previously performed standard MD simulations of the isolated protein [8] and (ii) the selection of sets of residues which can be considered to move as rigid bodies. Our results show that TAMD is able to induce an enhanced sampling of isolated AC conformational space, generating conformations that are more globular than the CaM-complexed one. A refinement procedure is applied to the TAMD derived structures to optimize packing interactions between domains, providing AC conformations which are stable along further unrestrained MD simulations. The final optimized structures are remarkably more globular, with the α-helix H establishing long-range interactions with the region CA, as the hydrogen bonds between the residue pairs ASP-360,ASP-359/ARG-246, which provide potential candidates for cross-linking experiments.

To our knowledge, the structures presented here are the first available all-atom models of globular conformations sampled by the isolated AC CyaA domain. The present work also addresses the question of finding a strategy to reach basins of local energy minima in the case where one end-point of the transition corresponds to a highly dynamical state not defined by a unique conformation.

1 MATERIALS AND METHODS

Molecular Dynamics simulations

The starting point of the simulations was the same as the one used in [8], ie. the AC PDB structure 1YRT in which the residues 226-232 invisible in the X-ray crystallographic structure [5] were built using Modeller9v4 [51]. The force field CHARMM22 including the cross-term (CMAP) correction [52, 53] was used and Na^+ counterions were added to neutralize the system. The system was then hydrated in a box of TIP3P [54] water molecules using a cutoff of 12 Å and periodic boundary conditions.

The MD and TAMD trajectories were performed using NAMD 2.7b2 [55]. A cutoff of 12 Å and a switching distance of 10 Å were defined for non-bonded interactions calculations, while long-range electrostatic interactions were calculated with the Particle Mesh Ewald (PME) protocol [56]. The simulations were realized in the NPT ensemble. The simulations were performed at temperature 300 K and pressure 1 atm. Temperature was regulated according to a Langevin thermostat [57], and the pressure was regulated with the Langevin piston Nose-Hoover method [58, 59]. The SHAKE algorithm [60, 61] was used to keep rigid all covalent bonds involving hydrogens, enabling a time step of 2 fs. Atomic coordinates were saved every picosecond.

At the beginning of each trajectory, the system was first minimized for 1,000 steps, then heated up gradually from 0 K to 300 K in 30,000 integration steps. Finally, the system was equilibrated for 50,000 steps. If not otherwise stated, the residue numbering in the AC domain is those of the crystallographic structure 1YRT

[5]. Set-up for the MD simulations is summarized in Table S1.

TAMD simulations

Let us indicate with $(\boldsymbol{x}, \boldsymbol{z})$ the set of physical (i.e. Cartesian) and extra variables, and with $\boldsymbol{\theta}(\boldsymbol{x})$ a set of collective variables, functions of the Cartesian ones. Assuming Langevin dynamics for the physical system, TAMD equations of motion can be written as:

$$\begin{cases} M\ddot{\boldsymbol{x}} = -\gamma\dot{\boldsymbol{x}} - \nabla_x V(\boldsymbol{x}) - \kappa \sum_{\alpha=1}^{N} (\theta_\alpha(\boldsymbol{x}) - z_\alpha)\nabla_x \theta_\alpha(\boldsymbol{x}) \\ \qquad\qquad\qquad\qquad\qquad\qquad + \sqrt{2M\gamma\beta^{-1}}\,\boldsymbol{\eta}^x(t) \\ \bar{\gamma}\dot{\boldsymbol{z}} = \kappa(\boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{z}) + \sqrt{2\bar{\gamma}\bar{\beta}^{-1}}\,\boldsymbol{\eta}^z(t) \end{cases} \tag{1}$$

where M is the mass matrix, $V(\boldsymbol{x})$ is the empirical classical potential of the system, $\boldsymbol{\eta}^{x,z}(t)$ are white noises (i.e. Gaussian processes with mean 0 and covariance $< \eta^p_\alpha(t)\eta^p_{\alpha'}(t') >= \delta_{\alpha\alpha'}\delta(t-t')$, with $p = \boldsymbol{x}, \boldsymbol{z}$), $\kappa > 0$ is the so-called spring force constant, $\gamma, \bar{\gamma} > 0$ are friction coefficients of the Langevin thermostats, $\beta^{-1} = k_B T$, $\bar{\beta}^{-1} = k_B \bar{T}$ with k_B the Boltzmann constant and $T, \bar{T}$ temperatures.

Equation (1) describes the motion of $\boldsymbol{x}$ and $\boldsymbol{z}$ under the extended potential

$$U_\kappa(\boldsymbol{x}, \boldsymbol{z}) = V(\boldsymbol{x}) + \tfrac{1}{2}\kappa \left\| \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{z} \right\|^2 . \tag{2}$$

It was shown in [33] that by adjusting the parameter κ so that $\boldsymbol{z}(t) \approx \boldsymbol{\theta}(\boldsymbol{x}(t))$ and the friction coefficient $\bar{\gamma}$ so that the $\boldsymbol{z}$ move slower than $\boldsymbol{x}$, one can generate a trajectory $\boldsymbol{z}(t)$ in z-space which effectively moves at the artificial temperature $\bar{T}$ on the free energy hyper-surface $F(\boldsymbol{z})$ defined at the physical temperature T. Hence, by construction, the limiting equation for $\boldsymbol{z}(t)$ in Eq. 1 samples the distribution $e^{-\bar{\beta}F(z)}$.

Then, using $\bar{T} > T$ in (1) accelerates the exploration of the free energy landscape by the $\boldsymbol{z}(t)$ trajectory, as energy barriers can be crossed more easily.

In the present work, six TAMD trajectories of 40ns each were performed, using three different temperatures and the Cartesian coordinates of three or four centers of mass as collective variables. The starting point of all TAMD trajectories was the AC conformation obtained after recording 1 ns of standard MD with NAMD. We indicate these sets of variables as (CM_1^a, CM_2^a, CM_3^a) and (CM_1^b, CM_2^b, CM_3^b, CM_4^b) and describe them in details in Results. In each TAMD simulation, the parameters of the MD Langevin thermostat are such that: $\gamma^{-1} = 0.5$ps, and $\beta^{-1} = 0.6$kcal/mol, corresponding to a temperature of 298 K. The TAMD algorithm was implemented in NAMD via a tcl script for calculating the evolution of collective variables. The restraint constant force κ was 100 kcal/(mol.Å^2). The value for the artificial friction on the $\boldsymbol{z}$ variables can be determined following the principle that the separation of time scales between $\boldsymbol{x}$ and $\boldsymbol{z}$ must be such that the $\boldsymbol{x}$ have time to equilibrate before the $\boldsymbol{z}$ move substantially. In practice, we proceeded as suggested in [44], i.e. we ran standard MD trajectories with the collective variables restrained at $\theta(\boldsymbol{x}) = \boldsymbol{z}$ fixed, and monitored the mean force estimators $G_j(N)$ defined for each collective variable j as:

$$G_j(N) = \frac{\kappa}{N} \sum_{i=1}^{N} [\theta_j(x(t_i)) - z_j] \tag{3}$$

where $\theta_j(x(t_i))$ is the instantaneous value at the time t_i of the collective variable. The time required for $G_j(N)$ to reach a plateau (Figure S1) allows one to extract the characteristic time of relaxation of the Cartesian variables to a fixed value of the variables $\boldsymbol{z}$, and hence an estimate for the time-scales separation $\bar{\gamma}/\gamma$. For the 9

coordinates corresponding to the coordinates of (CM_1^a, CM_2^a, CM_3^a) (Figure S1a) and for the 12 variables corresponding to the coordinates of (CM_1^b, CM_2^b, CM_3^b, CM_4^b) (Figure S1b), the estimator (3) converges in 5000 simulation time-steps. As the simulations time-steps are of 0.002 ps, a friction $\bar{\gamma}$ of 50 ps^{-1} is enough to allow system relaxation.

In order to explore the AC conformational space at different levels, TAMD trajectories were run using values of 4, 6 and 20 $kcal.mol^{-1}$ for the artificial thermal energy $\bar{\beta}^{-1}$ of the Langevin thermostat attached to the collective variables. These thermal energies correspond to artificial temperatures $\bar{T}$ of 1986, 2980 and 9933 K. We stress again here that although we employ high temperatures to evolve the z variables, we also use high friction on them to ensure that the x variables have the time to locally equilibrate to new values of z before these move substantially. This aspect makes TAMD quite different from a standard MD run at high temperature: the use of high temperature values does not alter the folded structure of the protein.

The set-up of the TAMD simulations is summarized in Table S1. Additional analysis methods for the conformational sampling and for the detection of rigid clusters of AC residues, are described in the Supplementary Materials.

RESULTS

Here, we describe the procedure we followed to obtain equilibrated conformations of the isolated AC which are in qualitative agreement with the limited available experimental data. A scheme of our strategy is represented in Figure S2. Starting from the extended AC structure observed in the complex with calmodulin, we chose a set of collective variables and performed TAMD simulations which allowed us to enhance sampling of the isolated protein. Conformers that showed the largest drift from the starting condition where then selected and submitted to an optimization procedure based on energy minimization and refinement by unrestrained MD simulations. In this way, we finally obtained a set of equilibrated, globular isolated AC conformations.

Definition of collective variables

The first step of the present study, i.e. the choice of the collective variables, was performed in the following way. First, using the Lyman-Zuckerman approach [62] we analyzed the standard MD trajectory of the isolated AC that was previously [8] performed using AMBER [63] and starting from the AC conformation in the C-CaM complexed crystallographic structure 1YRT [5]. Using a cutoff value of 3 Å we could identify seven AC representative conformations, four of which are displayed in Figure 2a. As expected from the limited sampling capabilities of a standard MD simulations, the trajectory is trapped in a free-energy basin surrounding the local minimum of the crystallographic conformation. The protocol PiSQRD [64] (described in the Supplementary Materials), was then used to break down the protein

into rigid groups of residues, in order to account for 80% of the possible internal mobility of each representative AC conformation. The PiSQRD analyses were applied using the Web interface `pisqrd.escience-lab.org` [65].

The groups produced by the PiSQRD protocol are considered as moving together as rigid bodies, and their centers of mass coordinates are thus good candidates for collective variables to describe the isolated AC conformational change. Among the groups detected by PiSQRD, only the residues common to the definitions obtained for each representative conformation were kept to define the centers of mass. In this way, we avoid bias introduced by outlier representative conformations, which may arise from transiently observed mobility in standard MD simulations.

From the PiSQRD analyses, it was thus possible to describe the conformational variability of AC using only 3 or 4 groups of residues, whose centers of mass we indicate as (CM_1^a, CM_2^a, CM_3^a) and (CM_1^b, CM_2^b, CM_3^b, CM_4^b) (Figure 2 and Table I). In both sets, the first group (CM_1^a or CM_1^b) is defined quite similarly (Table I) using residues from the region CB. The second group (CM_2^a or CM_2^b) contains in both cases residues from the region CA. CM_2^b, however, does not include the residues 325-364, corresponding to the C terminal region (in cyan in Figure 2) and the helix α16. These residues are thus considered to move independently when four rigid domains are employed. This seems to be an accurate description since it agrees with the large reorganization undergone by this region along the MD trajectories [8]. Finally, the SA region contains one (CM_3^a) or two (CM_3^b, CM_4^b) groups. The group CM_3^a is defined using residues from the SA region that are close to CA: residues 245-248 from helix H, helix G and residues 207, 208 at the C terminal part of helix

F. Otherwise, CM_3^b and CM_4^b are defined at the two extremes of SA, one closest to SA and the other including the loop 226-232. We note that this loop, reconstructed by homology modeling as it is not present in the crystallographic structure 1YRT, displayed a large degree of flexibility in standard MD simulations [8].

Summarizing, from the analysis of standard MD simulations, two sets of collective variables were chosen to be used in TAMD, corresponding respectively to the Cartesian coordinates of three and four centers of mass, i.e. 9 and 12 variables. The definitions of centers of mass located in CA and SA are quite variable depending whether three or four centers of mass are defined. This variations agree with the larger conformational fluctuations observed previously [8] for CA and SA regions along standard MD simulations.

TAMD induced conformational change

TAMD trajectories of 40 ns were run using as collective variables the centers of mass coordinates (CM_1^a, CM_2^a, CM_3^a) and (CM_1^b, CM_2^b, CM_3^b, CM_4^b) described in the previous section. As effective temperatures, $\bar{\beta}^{-1}$ values of 4, 6 and 20 kcal.mol^{-1} were chosen. In the following, we refer to the TAMD trajectories using the names: TAMD-**I**COM-**J**, where **I** = 3 or 4 indicates the number of collective variables and **J** = 4, 6, 20 kcal.mol^{-1} the $\bar{\beta}^{-1}$ values.

During the TAMD trajectories, the RMSD of the AC domain (Figure 3) increased from 2 to 8 Å using the collective variables (CM_1^a, CM_2^a, CM_3^a) and from 2 to 12 Å using the collective variables (CM_1^b, CM_2^b, CM_3^b, CM_4^b). The fact that larger RMSD values are obtained with four centers of mass shows that they work

better in pushing the system away from the C-CaM complexed minimum. We also point out that the maximum RMSDs observed in TAMD trajectories with 3 and 4 collective variables are respectively the double or the triple of the one observed in the same length standard MD simulation (black curve in Figure 3). The smallest values of gyration radius of the AC domain (Figure S3) are observed for the TAMD simulations recorded using 4 centers of mass, a $\bar{\beta}^{-1}$ value of 20 kcal/mol at the same time interval (25-40 ns) where the largest RMSD values are observed.

The amounts of α helix and β strands were calculated for one frame out of every ten along the MD and TAMD trajectories using the software STRIDE [66]. All profiles for α-helices and β-strands superimpose nicely (Figure S4), which proves that during TAMD trajectories the protein secondary structure content is not altered, even when high effective temperatures were employed. This stability also agrees with the small variation of $\approx 3.7 \pm 1\%$ experimentally observed [6] in α-helical content. Overall, the percentages of secondary structures, α-helices and β-strands, (Table S2) observed along TAMD and MD trajectories, are all around 30 and 15%, which is in qualitative agreement with the experimental percentages of $\approx$ 24% α-helices and $\approx$ 19% β-sheets [6] observed by CD.

Representative AC conformations were obtained using the Lyman-Zuckerman algorithm [62] with a cutoff $d_c = 2.7$ Å on the different TAMD trajectories (Figures 4b-d) as well as on the MD trajectory (Figure 4a). The superposition of the conformations on the CA region reveals that the relative orientations of the SA, CA and CB vary much more in T4C-06 (Figure 4c) and T4C-20 (Figure 4d) trajectories than in a standard MD trajectory run along the same time interval (Figure 4a).

Furthermore, the internal conformational change of CA, CB and SA domains was analyzed along T4C-4, T4C-6, T4C-20 (Figure S5). For the three values of $\bar{\beta}^{-1}$, the global drift (black curve in Figure S5) is much larger than for each single region CA, CB or SA, which shows that the relative displacements of the regions describe most of the total drift.

The regions CA and CB, located around the catalytic site, also move with respect to each other during T4C trajectories (Figure S6). Several sub-regions with conformational RMSD mostly smaller than 1.5 Å were detected: the N terminal region (residues 16-26), the regions CA_1 (residues 256-272), CA_2 (residues 7-35, 273-341) and CB_1 (residues 58-186) and the α helix AB (residues 44-54). During TAMD trajectories, the large variations of the distances between these regions are the sign of a major reorganization of the interface: the α helix AB moves apart from the N terminal region (Figure S6a) and the regions CB_1 and CA_1 move also apart from each other (Figure S6b). The regions CB_1 and CA_2 move closer to each other in T4C-06 and apart from each other in T4C-20 (Figure S6c). As the CA/CB interface surrounds the catalytic site, we speculate that the motion just described may cause a major destabilization of the site, consistent with the observed loss of AC activity in the absence of CaM.

Summarizing, with respect to standard MD, TAMD significantly increased the conformational changes in isolated AC. The regions undergoing the largest displacement are SA and CA. The architecture of the CA/CB interface is strongly reorganized, which agrees with the loss of enzymatic activity in isolated AC.

Analysis of protein-solvent and inter-domain interactions

As described in the previous section, during the trajectory T4C-20, the AC domain displays a large conformational change, where the region SA shows large variations of its orientation with respect to CA. In some of the observed conformations the two domains are close to each other. Nevertheless, in our TAMD trajectories we did not observe long-lasting, stable closed conformations of the protein. In order to investigate this issue we decided to first explore possible interactions of the protein with the surrounding solvent. To this purpose, we searched all simulated trajectories for water molecules forming hydrogen bonds with protein atoms (Figure 5). Bridging water molecules were selected as the ones forming at least two hydrogen bonds with protein donor or acceptor groups. In the standard MD trajectory (Figure 5, black curve), as well as in most of the TAMD trajectories, the number of water bridges is more or less constant and oscillates around 110, whereas after the first 20 ns of the T4C-20 and T3C-20 trajectory (Figure 5: green curve), the number of water bridges drops. The largest drop of about 90 is observed for T4C-20. By contrast, in T3C-4 and T4C-4, no transition is observed in the number of water bridges (Figure 5: blue curves).

In Figure 6, we report the presence of water bridges in matrices, where a point is added at the position (i,j) if a water bridge was observed between residues i and j. Cumulative populations of water bridges are plotted for the 29-30 ns interval of standard MD (Figure 6a) and from different intervals of the trajectory T4C-20 (Figure 6b-d). The 9-10 ns interval of T4C-20 (Figure 6b) displays a contact map similar to that observed for standard MD (Figure 6a), and interval 29-30 ns of T4C-

20 (Figure 6c) displays the minimum number of water bridges. In the interval 39-40 ns of T4C-20 (Figure 6d), a larger number of water bridges is observed, but the system did not come back to the state observed during the interval 9-10 ns. The reduction of the number of water bridges is proportionally more drastic between residues close in the sequence than between residues far apart in the sequence (long-range waters). Some long-range water molecules, highlighted with circles on Figure 6c, are still visible in the contact map of the interval 29-30 ns of T4C-20, connecting residues GLY-299, GLU-301, GLN-302, ASN-304, GLU-308, ALA-309, ASP-310 (catalytic loop), and residues 41-44, ARG-327 located on two sides of the catalytic site. The presence of these water molecules is certainly an obstacle to the enzymatic reaction.

In order to probe whether conformations sampled during T4C-20 could evolve spontaneously toward the basin of globular AC structure, three conformations were extracted at 36, 39 and 40 ns, corresponding to extrema of conformational RMSD (Figure 3b, green curve). As noted before, the smallest values of gyration radius values were observed (Figure S3b, green curve) for conformers showing the largest RMSD. Because however the fluctuations of gyration radius are quite large, we decided to pick the AC conformations based on the RMSD values which display a more regular trend.

Classical molecular dynamics simulations CONT-36, CONT-39 and CONT-40 were started from these conformations, and showed after 1 ns, a rapid drift up to 6 Å back to a completely elongated conformation similar to the one observed in complex with calmodulin. This result might indicate that during T4C-20 the system

did not have time to fully equilibrate degrees of freedom orthogonal to the collective variables. For example, solvent-protein interactions and interactions between the different domains that are likely to be relevant for stabilization of the globular conformation might require local equilibration not achieved during the TAMD runs. The disappearance of water bridges provides a mechanistic explanation for the backward drift. Indeed, water bridges could mediate long-range interactions between protein residues located in the regions SA and CA, and could thus induce a stable packing of SA on CA.

Stabilization of calmodulin-free AC conformations

In this section we describe the procedure we followed to optimize packing interactions at the newly formed interface between the SA and CA regions, thus stabilizing the globular conformation. We performed successive runs of energy minimization in presence of distance restraints, starting from the AC conformation extracted at 36 ns from TAMD trajectories. The minimizations were performed in the absence of solvent in order to avoid the trapping of water molecules at the SA/CA interface, which would prevent the establishment of strong interaction between the two domains. A high dielectric constant was used to mimic solvation condition.

The first set of distance restraints was established considering the water bridges observed between SA and CA regions along the TAMD trajectories. In the successive minimization runs, the list of distance restraints is each time increased by adding the restraints corresponding to the new hydrogen bonds established during the previous minimization run. Hence, all restraints added during the second and

third minimization runs are not arbitrary chosen and correspond to hydrogen bonds established during the previous minimization cycle.

Each minimization was performed for 100,000 steps under harmonic restraints applied to force the establishment of the selected hydrogen bonds: the harmonic restraints were applied using the command extraBonds of NAMD. The minimizations were performed *in vacuum* with a dielectric constant equal to 80 and in the absence of Particle Mesh Ewald (PME) protocol [56] or boundary conditions. For each hydrogen bond, the distance between the protein donor and acceptor atoms is restrained. For each observed water bridge, we restrained the distance between the protein acceptor atom establishing a hydrogen bond with the water oxygen, and the protein donor atom establishing a hydrogen bond with the water hydrogen. The distance restraints were applied with lower and upper limits of 1.2 and 2 Å, and a force constant of 30 kcal/(mol.Å^2).

From the simulation T4C-20 we extracted the conformation corresponding to 36 ns and we submitted it to three runs of *in vacuo* energy minimization. The first minimization was performed with two restraints only (Table S3) applied between pairs of atoms of residues ASN-200, LYS-34 and ASN-35 (Figure S7A) involved in the most stable water bridges observed along T4C-20.

We then analyzed the hydrogen bonds established during the first minimization run, focusing on pairs of residues that are separated by more than 50 other residues along the sequence. We found that 17 hydrogen bonds were established between the region SA and the regions CA and C terminal in more than 70% of the minimization steps. These hydrogen bonds involved atoms from residues ARG-342/ARG-253

(1 hydrogen bond), ASP-353/ARG-240 (6 hydrogen bonds), LEU-356/ARG-253 (2 hydrogen bonds), GLY-357/ARG-253 (3 hydrogen bonds), GLY-357/THR-249 (1 hydrogen bond), GLY-357/ARG-244 (4 hydrogen bonds) (Figure S7B). Since some of these bonds occur at different steps of the minimization, they are not all compatible with each other. For this reason, we divided them into 9 different lists that we applied separately (Table S4). These were added to the two restraints used during the first minimization cycle, thus producing 9 different minimized conformations, labeled from 21 to 29.

During the third minimization cycle, the different sets of hydrogen bonds established for more than 70% of the minimization steps, were added (Table S5) to each of the eight previous restraint lists. These restraints involve residues shown in Figure S7C. The 9 minimized conformations obtained after the third minimization cycle, are labeled from 31 to 39.

Different relative orientations are observed for the α helix H, in the 9 minimized conformations obtained after the third minimization cycles (Figure 7). Interestingly, this helix contains TRP-242 (drawn in licorice) which was shown by fluorescence spectroscopy [7] to display large internal mobility in free AC. A model of AC in which the α helix H would fluctuate between several orientations would explain such a mobility.

The stability of the 9 conformations was then checked by resolvating them in a water box, as described in the subsection "Molecular Dynamics Simulations" of the Materials and Methods, and submitting each one to 20-ns standard unrestrained MD simulations. Among these simulations, the trajectories MIN-32 and MIN-36,

initiated from the minimized conformations 32 and 36, display a plateau in RMSD at about 3.5 Å (Figure 8a), corresponding to almost superimposable. initial and final conformations of AC (Figure 8c). The RMSD deviations from the crystal structure (Figure 8b) are larger than 8.0 Å, which is much larger than the conformational drift of 5 Å observed for the isolated AC along the trajectory MD (Figure 3, black curve). The minimization procedure allowed thus to obtain two more globular conformations which were stable along 20 ns of unrestrained molecular dynamics simulation. The last conformations of the two trajectories are given as supplementary material.

In order to further validate the AC conformations sampled during the equilibration trajectories MIN-32 and MIN-36, the global shape of the protein was analyzed in order to check its agreement with the experimental observations [6]. First, the average value of the gyration radius has been calculated on MIN and on MD trajectories (Table S6). In MIN trajectories, the average gyration radii were of 23.4 Å, in qualitative agreement with the experimentally measured anhydrous radius of 23 Å. It should be also noticed that these values are smaller then the gyration value of 24.9 ± 0.3 measured on the trajectory MD.

The ratio between the inertia moments of the AC system was calculated on the protein surrounded by water molecules located at less than 20 Å to a protein atom, in order to take into account the water molecules dragged by the diffusion of the protein or influenced by being in the vicinity of the protein. This cutoff of 20 Å was chosen as the distance for which the motions of water molecules were experimentally shown [67] to be influenced by the protein movement. Using this approach, the calculation of the inertia moments showed, in agreement with Ref. [6]

that the system is prolate, with a mean ratio between the second and third inertia moment of 1.06 along MD and MIN trajectories. Second, the average ratio a/b between the first and the second inertia moment is 1.6 for MD and 1.4 for the MIN trajectories. The MIN trajectories sample thus conformations closer to the spherical shape than the MD trajectory.

In order to get a qualitative overview of the origin of stabilization of compact conformations, we calculated (Table S7): (i) the different components of the protein energy, (ii) the different components of the interaction energy between the protein and the solvent. Significant differences are observed between MIN-32/MIN-36 and MD for the electrostatic Coulomb energy and the total energy of the protein, which are more negative on MIN-32 and MIN-36. The more negative Coulomb energy agrees with the establishment of stabilizing hydrogen bonds within the protein. Concerning the interaction energy of the protein with the solvent, the opposite trend is observed: the Coulomb interaction energy is more negative for MD trajectory than for MIN trajectories, which supports the formation of protein-solvent electrostatic interactions in MD trajectory, arising from the disruption of protein-protein interactions observed in MIN trajectories.

Interactions stabilizing the globular conformation of isolated AC

We now turn to the analysis of hydrogen bonds connecting protein atoms of CA/C terminal and SA regions during the standard MD trajectories MIN-32 and MIN-34 (Table II). For comparison,we also analyze the standard MD trajectories CONT-36,

CONT-39, CONT-40 (Table S8), which drifted back to the elongated conformation. As a first, general result, the total number of inter-domain hydrogen bond is larger for MIN than for CONT trajectories, which agrees with the more globular shape of the protein structure in MIN simulations.

In particular, in CONT trajectories stable hydrogen bonds are only observed between GLU-276 and ARG-258 and between GLY-341 and LEU-198 (Table S8). It should be noted that in the PDB crystallographic structure 1YRT, the GLU-276 and ARG-258 sidechains are only about 7 Å from each other, and GLY-341 and LEU-198 already establish a hydrogen bond.

On the contrary, in MIN trajectories, in addition to hydrogen bonds between GLU-276 or VAL-271 and ARG-258 and between GLY-341 and LEU-198, hydrogen bonds are observed more than 80% of the time between ASP-360 or ASP-359 and ARG-246 (Table II). The hydrogen bonds between ASP-359 and ARG-246 were already present in the restraints used for minimizing the conformations (Table S3). It is worthwhile to point out that in the crystal structure of the C-CaM complex, ASP-360/ASP-359 and ARG-246 are located on two distinct CaM/AC interaction surfaces, more than 18 Å apart. We thus speculate that the hydrogen bonds between ASP-360/ASP-359 and ARG-246 are relevant for the stability of the globular isolated protein conformations and could be thus tested in cross-linking experiments to lock CyaA in an always inactive conformation.

DISCUSSION

The aim of the present work was to provide conformations of the CaM-free AC domain of the adenyl cyclase from *Bordetella pertussis*. To this purpose, we employed TAMD to enhance sampling of isolated AC conformational space, using as collective variables the Cartesian coordinates of the centers of mass of 3 and 4 groups of residues. The limited experimental information available on the unbound state of AC, resulting from hydrodynamic measurements [6], supports a model where the protein conformation is more globular than in the complex with CaM. Following these guidelines, a set of extremely drifted conformations sampled during TAMD trajectories was selected and submitted to a refinement procedure whose goal was to optimize packing interactions between the regions CA and SA that had moved close to each other during the accelerated trajectory. The reliability of obtained conformations was checked by verifying their stability along standard unrestrained MD simulations.

This finally yielded reasonable, although still tentative, predictions for a sample of protein conformations populating the free-energy basin of the calmodulin-free state of the protein. These predictions permit to infer inter-domain contacts responsible for the stability of the globular conformations.

First of all, the present study confirms the efficiency of TAMD in enhancing exploration of the conformational space of a protein. Indeed, the comparison of representative conformations extracted from the standard MD trajectory and from TAMD reveals the wider conformational change occurring in the latter. This enhancement is especially remarkable since similar simulation lengths were used for

recording both types of trajectories. It is also important to stress that due to the lack of structural information on the isolated state, the collective variables were defined from knowledge of the bound state only. Nevertheless, TAMD was able to push the system out of the local minimum energy basin sampled by standard MD. Notably, we observed that using four centers of mass coordinates as collective variables worked better in pushing the system away from the C-CaM complex minimum. Since the differences between the two sets of centers of mass are located in the CA and SA regions, this result points to the importance of an accurate description of these regions to understand the conformational transition. In the time-span here considered, however, TAMD did not generate a complete transition to the energy basin of unbound AC. This result might indicate that degrees of freedom orthogonal to the collective variables have a relevant role in stabilizing the isolated domain. In particular, the stabilization of the globular conformation most likely requires local rearrangement of sidechains to optimize the newly established domain-domain interface, a fact which had no time to occur during the TAMD runs. Here we were able to identify some of these interactions, namely hydrogen bonds between the SA and CA domains. In future studies, it would be worthwhile to use variables describing these interactions in free energy calculations. Also, the results here presented suggest which residues to mutate in cross-linking experiments to probe the shape of the CaM-free protein conformation *in vitro* and to possibly lock it in a state incapable of binding CaM, thus being constitutively inactive.

The conformational changes of AC observed during TAMD trajectories allow us here to draw hypotheses about the inactivation mechanism of the protein. First, as

extrapolated from previous MD simulations [8], the conformational change associated with the variations of AC activity are localized at the interface between CA and CaM. Second, it provides additional detailed information about the deformation of the catalytic site exerted by the movement of CB with respect to CA.

We have analyzed in detail the role of protein-solvent interactions during the conformational change. First, in TAMD trajectories stable water bridges are formed between the catalytic loop and other regions of the protein, which we speculate might perturb the catalytic activity. Second, a dramatic decrease of water bridges is associated with the largest conformational change of the protein. These results suggest that a substantial modification of protein-solvent interactions is correlated with AC conformational transition. In particular, the large decrease in number of water bridges during the protein transition to a more globular conformation is reminiscent of the large solvent density fluctuations observed at the de-wetting transition of a bio-molecule [68].

The kind of transition explored here, between a well-defined bound state and a much more dynamical unbound state, is quite often encountered in biology. From a computational stand-point, the lack of precision in the definition of the unbound state makes more difficult the definition of appropriate collective variables, as well as the definition of the target region of the conformational transition. The present work thus provides an important illustration of how to address such class of problems.

Furthermore, the predicted packing geometry for the globular state of AC can provide valuable information about possible interactions between AC and thiophen ureido acid inhibitors [4]. Indeed, an extensive bioinformatics analysis of the *Bacil-*

lus anthracis adenyl cyclase predicted that these inhibitors bind to a pocket called SABC. A quite similar pocket is observed in AC, which supports the hypothesis of efficiency of the same inhibitors on AC. Among the residues lining the pocket are LYS-34, ASN-200 and PHE-275, which are involved in some of the intra-protein hydrogen bonds in the globular AC conformations discussed in this work. We therefore speculate that thiophen ureido acids inhibitors could act on AC by stabilizing its globular conformation via interactions with these residues.

Notable advantages of TAMD are that i) it is an untargeted exploration of the collective variables space, and ii) it can be used with multiple collective variables (12 were used here), thus making it suitable to tackle a problem like the one reported in this work where a single collective variable is not easy to identify *a priori* and some of the structural quantities such as the gyration radius are likely to fluctuate between different values.

To conclude, although TAMD was not by itself sufficient to yield a stable isolated protein conformation, it remarkably generated the spontaneous large-scale movement of the SA helices towards the CA domain which was not observed in standard MD simulations at the timescale recorded here, thus providing a good quality guess for the unbound protein model. Starting from this guess, a refined conformation was obtained via local adjustment of the newly formed inter-domain surface. This work hence provides important information about interactions among domains responsible for isolated AC structure stabilization which in the future will be further exploited: (i) computationally, by free energy calculations using collective variables which properly describe the motion of the relevant domains, (ii) experimentally, by

realizing cross-linking experiments to check the effective proximities of the residues.

CONCLUSIONS

Upon CaM unbinding, the catalytic domain of the adenyl cyclase toxin from *Bordetella pertussis* undergoes a transition from an elongated conformation, which was previously experimentally characterized, to a more globular conformation only qualitatively described from experiments. In this work, we showed that TAMD trajectories started from the protein configuration in the bound state allowed a more extensive sampling of the conformational space than standard MD trajectories. Out of the enhanced sampling simulations we selected extremely drifted protein conformations and refined them by energy minimization and further standard simulations to obtain a set of metastable globular conformations. Based on the analysis of protein-protein and protein-solvent interactions, we were able to identify a series of connected residues in the metastable protein conformations which constitute the first available prediction of inter-residue proximities observed in the CaM-free *B. pertussis* CyaA catalytic domain.

ACKNOWLEDGMENTS

ES thanks the Ministère de l'Enseignement Supérieur for PhD support. TM and TH acknowledge CNRS and Institut Pasteur for funding. GC acknowledges financial support from SFI grant no. 08-IN.1-I1869 and from the Istituto Italiano di Tecnologia under the SEED project grant no. 259 SIMBEDD - Advanced Compu-

tational Methods for Biophysics, Drug Design and Energy Research. We thank Dr. Elodie Laine for useful discussions about cyclase inhibitors. Drs. Paulo Ricardo Battista and Albert Y. Lau and Prof. Eric Vanden-Eijnden are acknowledged for fruitful discussions.

References

[1] Walker C. Whooping cough case numbers rise across the UK and US. Nurs Child Young People 2012;**24**:4.

[2] Murphy JF. Pertussis has re-emerged. Ir Med J 2012;**105**:260.

[3] Seifert R, Dove S. Towards selective inhibitors of adenylyl cyclase toxin from Bordetella pertussis. Trends Microbiol 2012;**20**:343–351.

[4] Laine E, Goncalves C, Karst J, Lesnard A, Rault S, Tang WJ, Malliavin TE, Ladant D, Blondel A. Use of allostery to identify inhibitors of calmodulin-induced activation of Bacillus anthracis Edema Factor. Proc Natl Acad Sci USA 2010;**107**:11277–11282.

[5] Guo Q, Shen Y, Lee YS, Gibbs CS, Mrksich M, Tang WJ. Structural basis for the interaction of Bordetella pertussis adenylyl cyclase toxin with calmodulin. EMBO J 2005;**24**:3190–3201.

[6] Karst JC, Pérez AC Sotomayor, Guijarro JI, Raynal B, Chenal A, Ladant D. Calmodulin-induced conformational and hydrodynamic changes in the cat-

alytic domain of Bordetella pertussis adenylate cyclase toxin. Biochemistry 2010;**49**:318–328.

[7] Gallay J, Vincent M, de la Sierra IM Li, Munier-Lehmann H, Renouard M, Sakamoto H, Bârzu O, Gilles AM. Insight into the activation mechanism of *Bordetella pertussis* adenylate cyclase by calmodulin using fluorescence spectroscopy. Eur J Biochem 2004;**271**:821–833.

[8] Selwa E, Laine E, Malliavin TE. 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.

[9] Whitley MJ, Lee AL. Exploring the role of structure and dynamics in the function of chymotrypsin inhibitor 2. Proteins 2011;**79**:916–924.

[10] Doring K, Surrey T, Grünewald S, John E, Jähnig F. Enhanced internal dynamics of a membrane transport protein during substrate translocation. Protein Sci 2000;**9**:2246–2250.

[11] Sacquin-Mora S, Delalande O, Baaden M. Functional modes and residue flexibility control the anisotropic response of guanylate kinase to mechanical stress. Biophys J 2010;**99**:3412–3419.

[12] Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. Nat Chem Biol 2010;**6**:352–358.

[13] Kokkinidis M, Glykos NM, Fadouloglou VE. Protein flexibility and enzymatic catalysis. Adv Protein Chem Struct Biol 2012;**87**:181–218.

[14] Ramanathan A, Agarwal PK. Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. PLoS Biol 2011;**9**:e1001193.

[15] Boekelheide N, Salomón-Ferrer R, Miller TF. Dynamics and dissipation in enzyme catalysis. Proc Natl Acad Sci U S A 2011;**108**:16159–16163.

[16] Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, Benkovic SJ, Wright PE. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science 2011;**332**:234–238.

[17] Klepeis J, Lindorff-Larsen K, Dror R, Shaw D. Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol 2009;**19**:120–127.

[18] Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling and throwing ropes over rough mountain passes, in the dark. Annu Rev Phys Chem 2002;**53**:291–318.

[19] Elber R, Karplus MM. A method for determining reaction paths in large molecules: application to myoglobin. Chem Phys Lett 1987;**139**:375–380.

[20] Mills G, Jönsson H. Quantum and thermal effects in H2 dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. Phys Rev Lett 1994;**72**:1124–1127.

[21] Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. String method in collective variables: Minimum free energy paths and isocommittor surfaces. J Chem Phys 2006;**125**:024106.

[22] Vanden-Eijnden E, Venturoli M. Markovian milestoning with Voronoi tessellations. J Chem Phys 2009;**130**:194101.

[23] Fischer S, Karplus M. Conjugate peak refinement and an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. Chemical Physics Letters 1992;**194**:252–261.

[24] E W, Vanden-Eijnden E. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. Annu Rev Phys Chem 2010;**61**:391–420.

[25] Huber T, Torda AE, van Gunsteren WF. Local elevation: A method for improving the searching properties of molecular dynamics simulation. J. Comput.-Aided Mol. Design 1994;**8**:695–708.

[26] Grubmuller H. Predicting Slow Structural Transitions in Macromolecular Systems - Conformational Flooding. Phys Rev E 1995;**52**:2893–2906.

[27] Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci U S A 2002;**99**:12562–12566.

[28] Tribello GA, Ceriotti M, Parrinello M. A self-learning algorithm for biased molecular dynamics. Proc Natl Acad Sci U S A 2010;**107**:17509–17514.

[29] Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys 2004;**120**:11919–11929.

[30] Bucher D, Grant BJ, Markwick PR, McCammon JA. Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. PLoS Comput Biol 2011;**7**:e1002034.

[31] Schlitter J, Engels M, Kruger P, Jacoby E, Wollmer A. Targeted Molecular-Dynamics Simulation of Conformational Change - Application to the T[-]R Transition in Insulin. Mol Simul 1993;**10**:291–308.

[32] Sotomayor M, Schulten K. Single-molecule experiments in vitro and in silico. Science 2007;**316**:1144–1148.

[33] 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.

[34] Ferrenberg AM Swendsen RH. New Monte Carlo technique for studying phase transitions. Phys. Rev. Lett. 1988;**61**:2635–2638.

[35] Maragliano L., Vanden-Eijnden E. Single-sweep methods for free energy calculations. J. Chem. Phys. 2008;**128**:184110.

[36] 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.

[37] Monteferrante M, Bonella S, Meloni S, Eijnden E Vanden, Ciccotti G. Calculations of free energy barriers for local mechanisms of hydrogen diffusion in alanates. Scientific Modeling & Simulation SMNS 2008;**15**:187–.

[38] Sterpone F, Bonella S, Meloni S. Early Stage of the Dehydrogenation of NaAlH4 by Ab Initio Rare Event Simulations. J Phys Chem C 2012;**116**:19636–19643.

[39] Lapelosa M, Abrams CF. A Computational Study of Water and CO Migration Sites and Channels Inside Myoglobin. J Chem Theory Comput 2013;**9**:1265–1271.

[40] Ciccotti G, Meloni S. Temperature accelerated Monte Carlo (TAMC): a method for sampling the free energy surface of non-analytical collective variables. Phys Chem Chem Phys 2011;**13**:5952–5959.

[41] Abrams CF, Vanden-Eijnden E. On-the-fly free energy parameterization via temperature accelerated molecular dynamics. Chem Phys Lett 2012;**547**:114–119.

[42] Yamamori Y, Kitao A. MuSTAR MD: Multi-scale sampling using temperature accelerated and replica exchange molecular dynamics. J Chem Phys 2013;**139**:145105–145115.

[43] Lucid J, Meloni S, MacKernan D, Spohr E, Ciccotti G. Probing the Structures of Hydrated Nafion in Different Morphologies Using Temperature Accelerated Molecular Dynamics simulations. J Phys Chem C 2012;:DOI: 10.1021/jp309038n.

[44] Abrams CF, Vanden-Eijnden E. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. Proc Natl Acad Sci U S A 2010;**107**:4961–4966.

[45] Vashisth H, Maragliano L, Abrams CF. "DFG-flip" in the insulin receptor kinase is facilitated by a helical intermediate state of the activation loop. Biophys J 2012;**102**:1979–1987.

[46] Vashisth H, Brooks CL. Conformational sampling of maltose-transporter components in Cartesian collective variables is governed by the low-frequency normal modes. J Phys Chem Lett 2012;**3**:3379–3384.

[47] Nygaard R, Zou Y, Dror RD, Mildorf TJ, Arlow DH, Manglik A, Pan AC, Liu CW, Fung JJ, Bokoch MP, Thian FS, Kobilka TS, Shaw DE, Mueller L, Prosser RS, Kobilka BK. The dynamic process of $\beta(2)$-adrenergic receptor activation. Cell 2013;**152**:532–542.

[48] Vashisth H, Abrams CF. All-atom structural models of insulin binding to the insulin receptor in the presence of a tandem hormone-binding element. Proteins 2013;**81**:1017–1030.

[49] Scarpazza DP, Ierardi DJ, Lerer AK, Mackenzie KM, Pan AC, Banka JA, Chow E, Dror RO, Grossman JP, Killebrew D, Moraes MA, Predescu C, Salmon JK, Shaw DE. Extending the Generality of Molecular Dynamics Simulations on a Special-Purpose Machine. Proceedings of the 27th IEEE International Parallel and Distributed Processing Symposium 2013;:933–945.

[50] Vashisth H, Storaska AJ, Neubig RR, Brooks CL. Conformational dynamics of a regulator of G-protein signaling protein reveals a mechanism of allosteric inhibition by a small molecule. ACS Chem. Biol 2013;:accepted.

[51] Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative Protein Structure Modeling using Modeller. Current Protocols in Bioinformatics 2006;**5**:5.6.

[52] MacKerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields and limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comp Chem 2004;**25**:1400–1415.

[53] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;**102**:3586–3616.

[54] Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;**79**:926–935.

[55] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;**26**:1781–1802.

[56] Darden T, York D, Pedersen L. Particle Mesh Ewald and an N.log(N) method for Ewald sums in large systems. J Chem Phys 1993;**98**:3684–90.

[57] Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. San Diego, California: Academic press; 2002.

[58] Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. J Chem Phys 1994;**101**:4177–4189.

[59] Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation and the Langevin piston method. J Chem Phys 1995;**103**:4613–4622.

[60] Ryckaert J.P., Ciccotti G., Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints and Molecular dynamics of n-alkanes. J. Comput. Phys. 1977;**23**:327–341.

[61] Andersen HC. Rattle and A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. J Comp Phys 1983;**52**:24–34.

[62] Lyman E, Zuckerman DM. Ensemble-based convergence analysis of biomolecular trajectories. Biophys J 2006;**91**:164–172.

[63] Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods R. The Amber biomolecular simulation programs. J Computat Chem 2005;**26**:1668–1688.

[64] Aleksiev T, Potestio R, Pontiggia F, Cozzini S, Micheletti C. PiSQRD and a web server for decomposing proteins into quasi-rigid dynamical domains. Bioinformatics 2009;**25**:2743–2744.

[65] Potestio R, Pontiggia F, Micheletti C. Coarse-grained description of protein internal dynamics and an optimal strategy for decomposing proteins in rigid subunits. Biophys J 2009;**96**:4993–5002.

[66] Frishman D, Argos P. Knowledge-Based Protein Secondary Structure Assignment. Proteins: Structure, Function, and Genetics 1995;**23**:566–579.

[67] Born BI, Kim SJ, Ebbinghaus S, Gruebele M, Havenith M. The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin. Faraday Discuss 2009;**141**:161–173.

[68] Patel A, Varilly P, Jamadagni S, Hagan M, Chandler D, Garde S. Sitting at the edge: how biomolecules use hydrophobicity to tune their interactions and function. J Phys Chem B 2012;**116**:2498–2503.

2 Figure Legends

Figure 1

Figure 1: X-ray crystallographic structure of the catalytic domain (AC) of the adenyl cyclase CyaA (PDB entry 1YRT) drawn in cartoons representation. The region SA colored in purple, contains the α-helix H. The loop 226-232 (Hom loop) at the left extremity of SA, colored in salmon, was missing in the X-ray crystallographic structure and was modeled [8] using Modeller9v4 [51]. The other protein regions are colored in green (CA), orange (CB), cyan/blue (C terminal tail: C tail) and yellow (catalytic loop: C loop). The C-CaM lobe, colored in red, interacts with SA and the C terminal tail. Calcium ions are represented by silver beads. The panel a) displays AC interacting with Ca^{2+}-loaded CaM, and the panel b) only AC, viewed from another side.

Figure 2

Figure 2: (a) Four representative conformations extracted from the standard MD trajectory of the isolated AC previously recorded [8] using AMBER [63]. These conformations were determined by the Lyman and Zuckerman algorithm (see in supplementary material the section "Analysis of conformational sampling"). These conformations are labeled with their corresponding frame numbers in the MD trajectory. The three (b) and four (c) centers of mass obtained from the PiSQRD analysis (see in supplementary material the section "Detection of rigid clusters of AC residues") of the representative conformations are represented as transparent spheres and labeled similarly as in Table I. In the displayed conformations, the AC regions are colored in purple (SA), in pink (loop 226-232 modeled by homology), in green (CA), orange (CB), cyan (C terminal tail: C tail) and yellow (catalytic loop: C loop). The Cα atoms defining each center of mass are drawn as spheres and colored in the following way (CM_1^a, CM_1^b: brown, CM_2^a, CM_2^b: blue, CM_3^a, CM_4^b: red, CM_3^b: magenta).

Figure 3

Figure 3: Global conformational drift (Å) calculated on the atoms Cα of the domain AC, along the MD (black curve) and the TAMD trajectories run using as collective variables three (a) and four (b) centers of mass. The curves are colored in blue, red and green for artificial temperatures $\bar{\beta}^{-1}$ of 4, 6 and 20 kcal.mol^{-1}.

Figure 4

Figure 4: Representative conformations extracted using the Lyman-Zuckerman algorithm [62] with a cutoff $d_C = 2.7$ Å from the (a) MD and (b) T4C-4, (c) T4C-6, (d) T4C-20. The protein regions are colored in purple (SA), green (CA), cyan (C terminal), yellow (catalytic loop) and orange (CB). The representative conformations were transformed by rigid roto-translation to get their regions CA superimposed to each other.

Figure 5

Figure 5: Number of water bridges along the MD trajectory (black curves) and the TAMD trajectories using as collective variables four (a) or three (b) centers of mass.

Figure 6

Figure 6: Positions of the water molecules (water bridges) connecting the AC acceptor and donor groups. Each water bridge connecting atoms from the residues i and j is drawn at the position i and j of a matrix: (a) Matrix obtained from summing up water bridges observed in the 29-30 ns interval of trajectory MD. (b-d) Matrix obtained from summing up water bridges observed in the 9-10 (b), 29-30 (c) and 39-40 (d) ns intervals of the trajectory T4C-20. Among the water bridges present in the 29-30 and 39-40 ns intervals, those connecting residues located on the two sides of the catalytic site, are highlighted by circles.

Figure 7

Figure 7: Minimized conformations of AC obtained during the second and the third minimization cycles. The minimized conformations were transformed by rigid roto-translation to get their regions CA superimposed to each other. The residue Trp-242 is drawn in red sticks.

Figure 8

Figure 8: (a) Global conformational drift (Å) from the minimized conformations, calculated on the atoms Cα of the domain AC, along the unrestrained 20 ns molecular dynamics trajectory MIN-32 (black curve) and MIN-36 (red curve). (b) Global conformational drift (Å) from the X-ray crystallographic structure of AC, calculated on the atoms Cα of the domain AC, along the unrestrained 20 ns molecular dynamics trajectory MIN-32 (black curve) and MIN-36 (red curve). (c) Superimposition of initial and final conformations of the trajectories MIN-32 and MIN-36. In the initial conformations, the CA and SA regions are colored respectively in green and purple, whereas they are colored in lime and pink in the final superimposed conformations.

Table I

Three centers of mass (3COM)	Residues definition	Region
CM_1^a	55-59 61-187	CB
CM_2^a	7-43 189-200 249-265 268-292 298-364	CA
CM_3^a	209-244	loop extremity of SA
Four centers of mass (4COM)	**Residues definition**	**Region**
CM_1^b	61-185	CB
CM_2^b	7-34 38-43 188-194 261-310 313-324	CA
CM_3^b	201-205 207 208 245-248	extremity of SA close to CA
CM_4^b	225-231	loop extremity of SA

Table I: Definition of centers of mass used as collective variables in TAMD simulations. The AC residue numbers used for the definition are given along with their localization in AC regions.

Table II

MIN-32 Hydrogen bonds		%
GLY-363-O	GLY-254-HN	39.9
LEU-362-O	ARG-259-Hζ22	60.9
ASP-360-Oδ2	ARG-246-Hϵ	78.1
ASP-360-Oδ1	ARG-246-Hζ22	83.8
ASP-360-Oδ1	ARG-246-Hϵ	45.7
LEU-362-O	ARG-259-Hζ22	60.9
ASP-360-Oδ2	ARG-246-Hϵ	78.1
ASP-360-Oδ2	ARG-246-Hζ22	89.2
ARG-258-O	ARG-348-Hζ11	41.3
THR-273-Oγ1	ARG-258-Hζ11	58.5
GLY-341-O	LEU-198-HN	82.6
GLU-276-Oϵ2	ARG-258-Hζ12	97.6
GLU-276-Oϵ2	ARG-258-Hζ22	99.7
ALA-252-O	THR-273-Hγ1	67.4
MIN-36 Hydrogen bonds		%
LEU-362-O	ARG-259-Hζ22	41.4
ASP-359-Oδ1	ARG-246-Hϵ	75.7
ASP-359-Oδ2	ARG-246-Hζ22	81.5
ASP-359-Oδ2	ARG-246-Hϵ	83.4
ASP-359-Oδ1	ARG-246-Hζ22	86.5
GLY-363-O	ARG-258-Hϵ	87.9
GLY-341-O	LEU-198-HN	91.8
VAL-271-O	ARG-258-HN	97.6
VAL-271-O	ARG-258-Hζ12	99.2

Table II: Presence of hydrogen bonds between SA and CA/C terminal tail during the trajectories MIN-32 and MIN-36. The acceptor and donor atoms of the water bridges and hydrogen bonds are given along with the percentage of their presence along the MD trajectories.

3 Figure

Figure 1

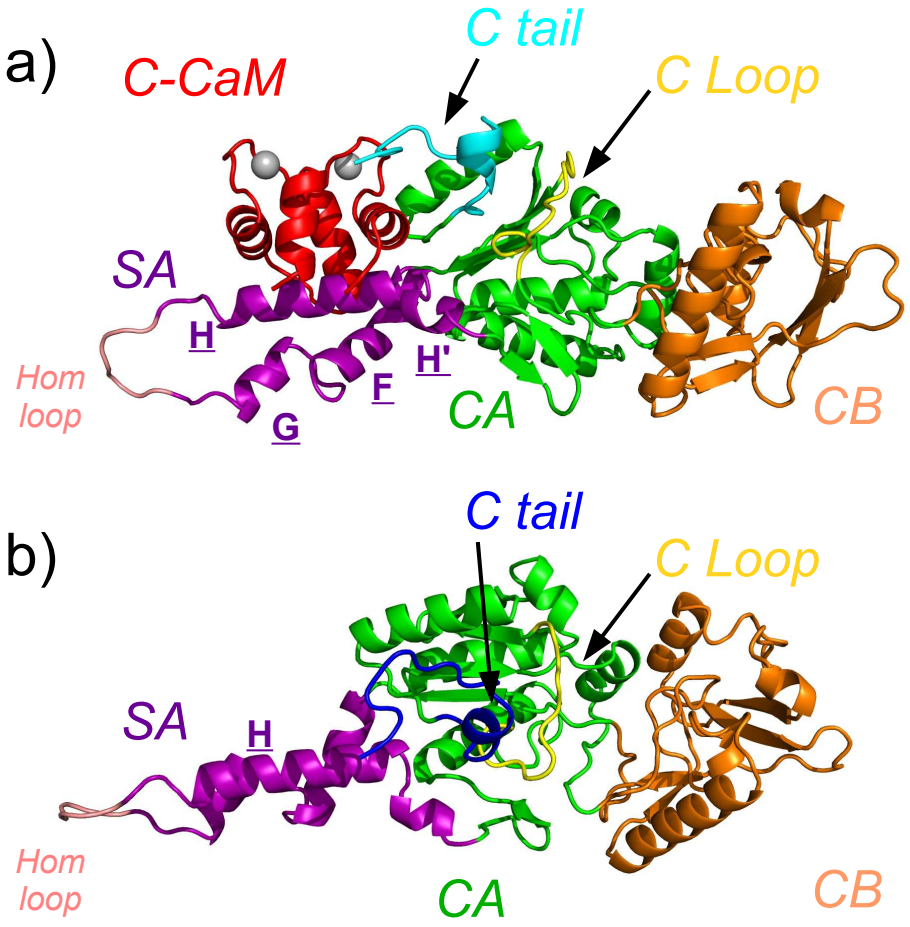


Figure 2

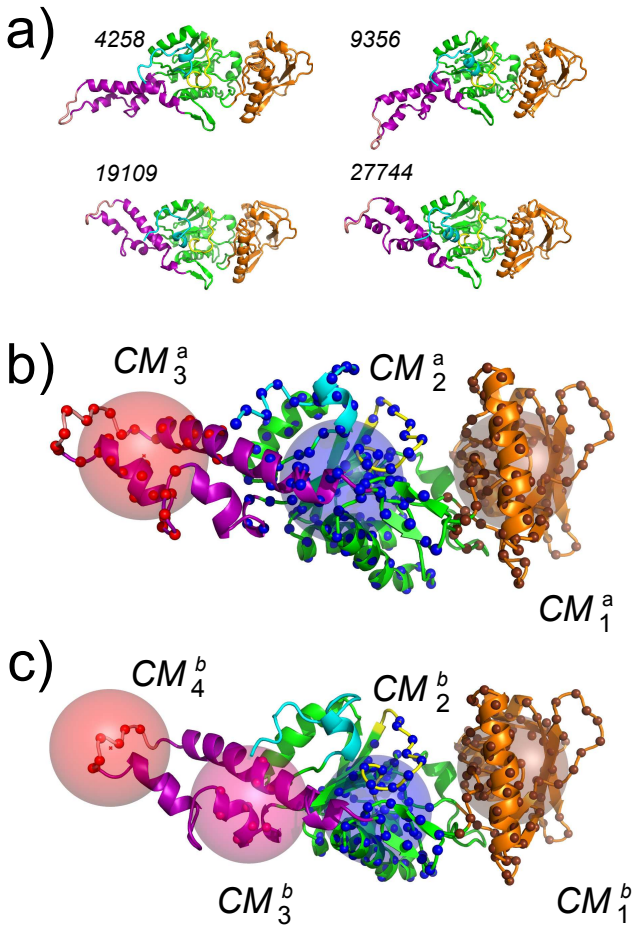


Figure 3

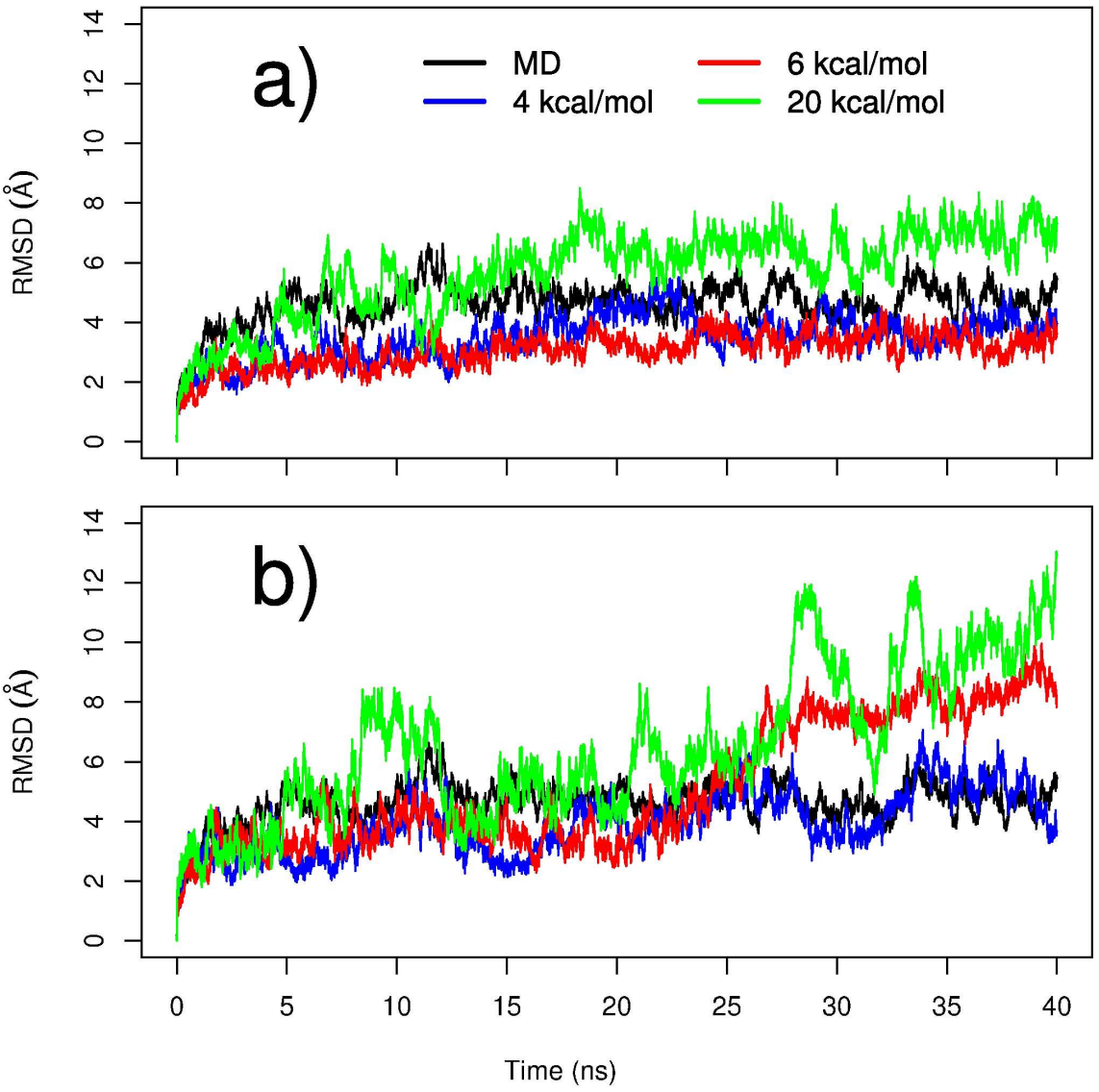

a)
MD
4 kcal/mol
6 kcal/mol
20 kcal/mol
RMSD (Å)
0
2
4
6
8
10
12
14
b)
RMSD (Å)
0
5
10
15
20
25
30
35
40
Time (ns)

Figure 4

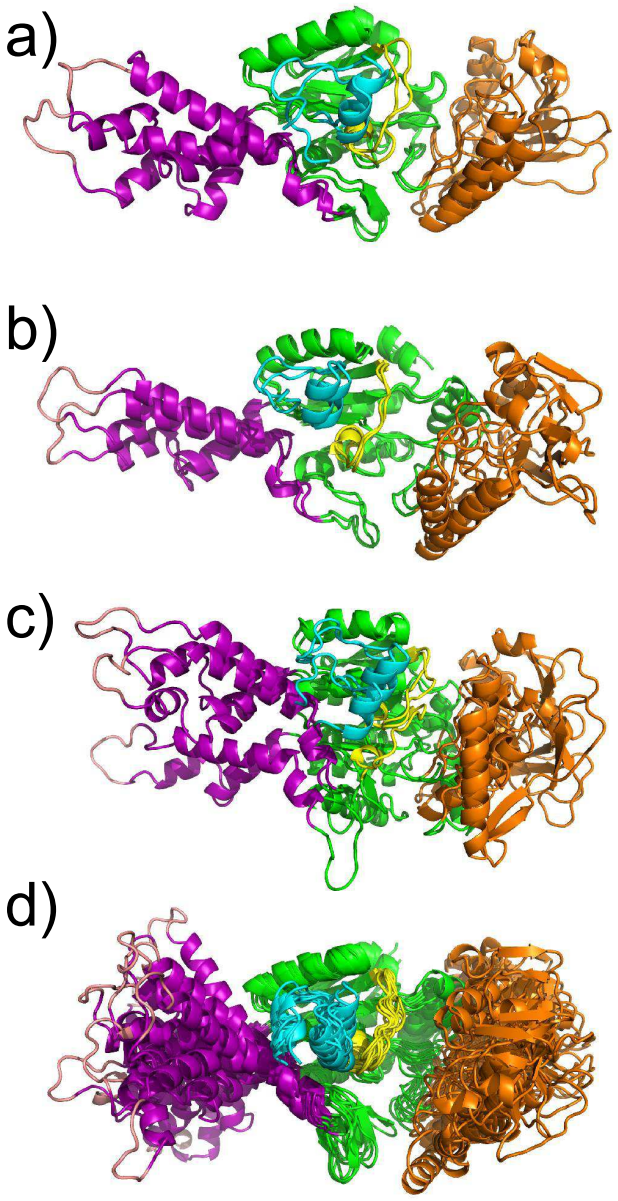


Figure 5

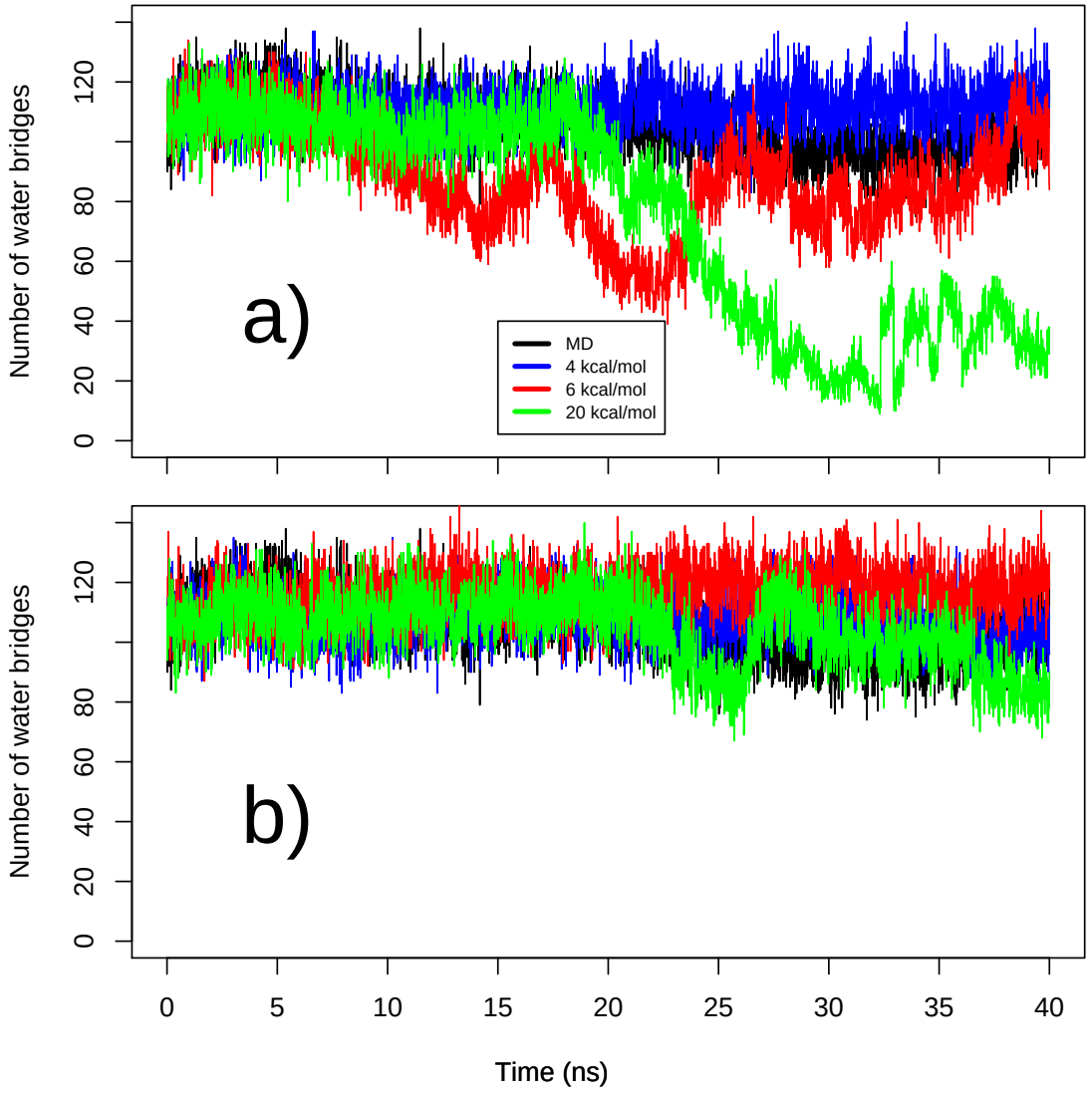

Number of water bridges
0
20
40
60
80
120
a)
MD
4 kcal/mol
6 kcal/mol
20 kcal/mol
Number of water bridges
0
20
40
60
80
120
b)
0
5
10
15
20
25
30
35
40
Time (ns)

Figure 6

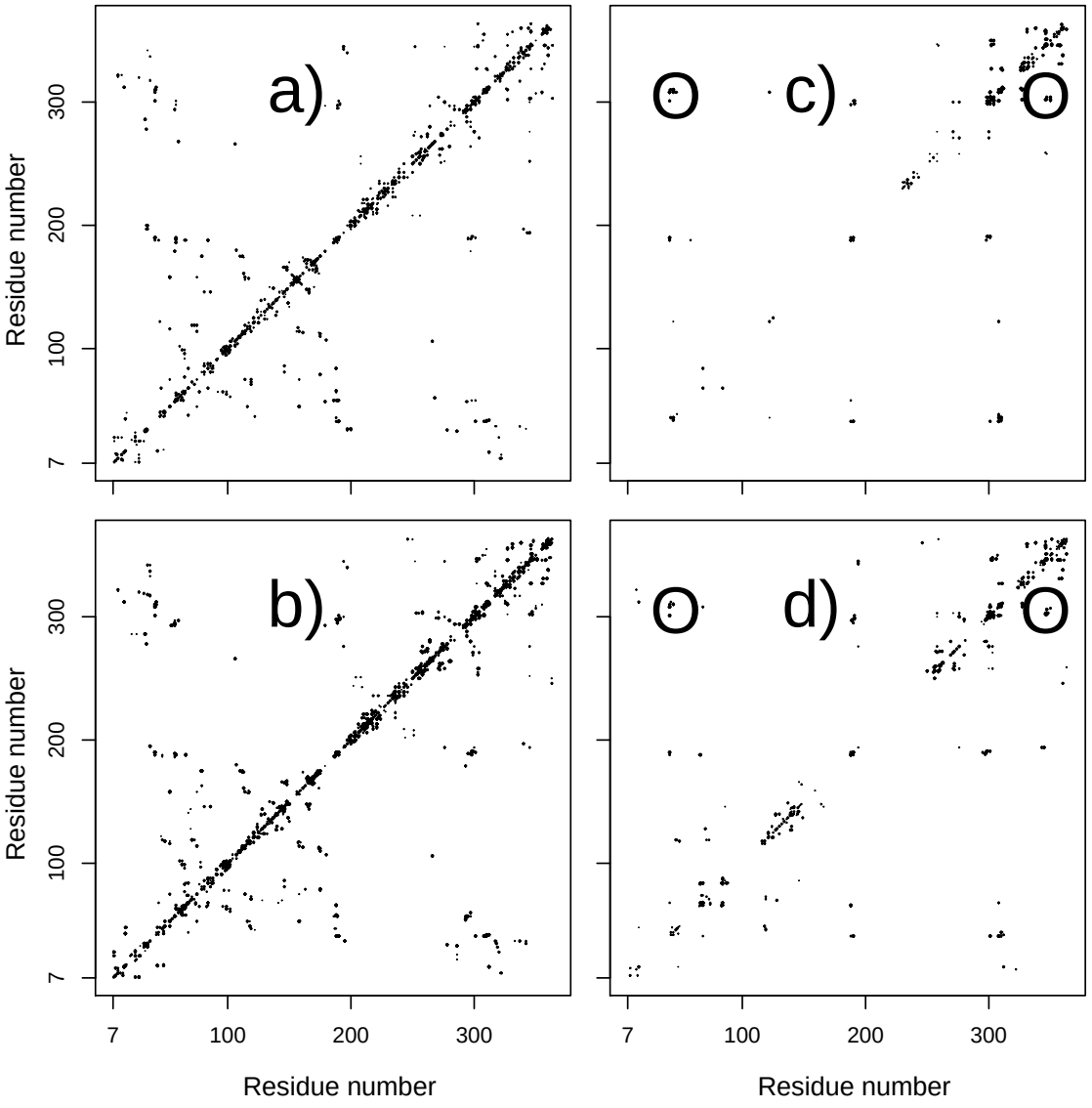


Figure 7

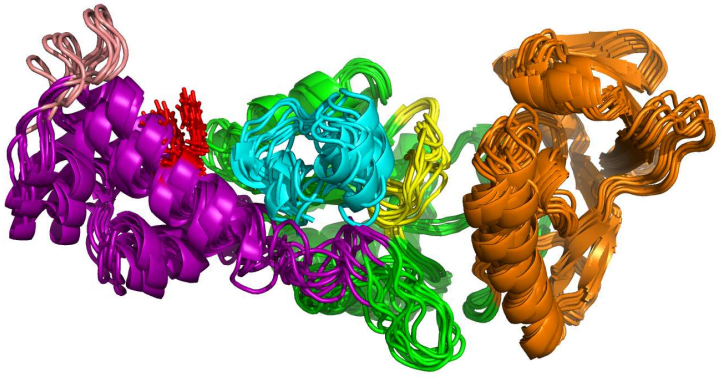
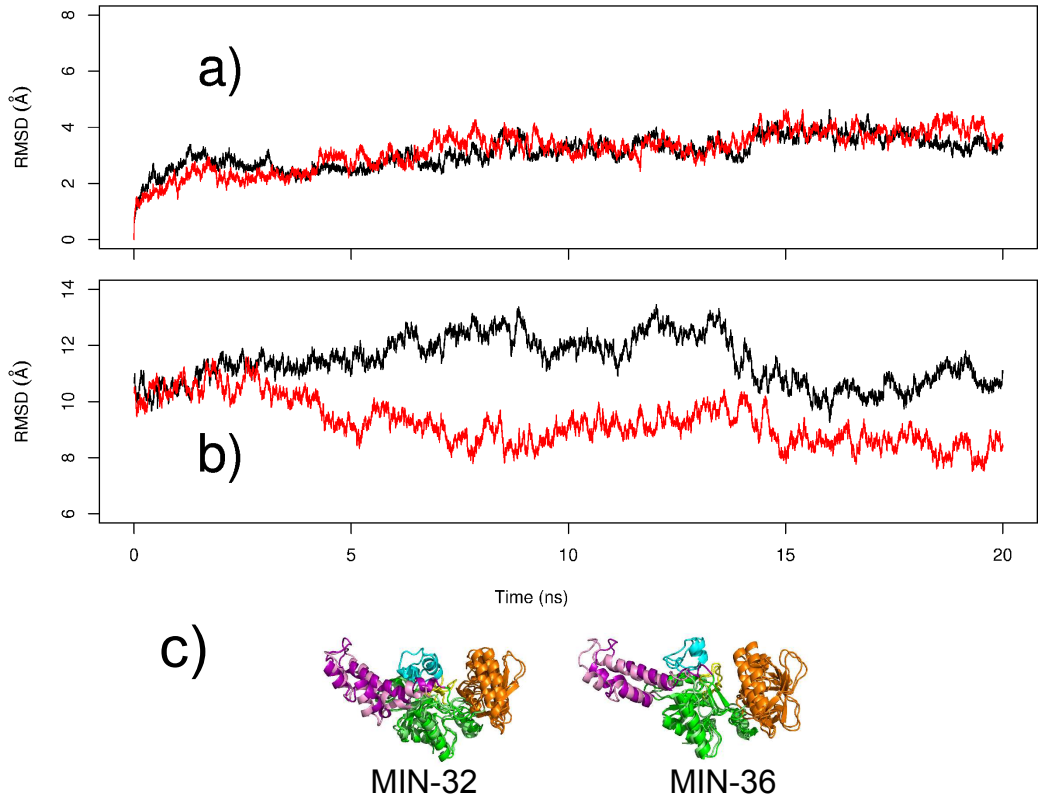


Figure 8


a)
RMSD (Å)
0
2
4
6
8
b)
RMSD (Å)
6
8
10
12
14
0
5
10
15
20
Time (ns)
c)
MIN-32
MIN-36