JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Temperature Accelerated Molecular Dynamics with Soft-Ratcheting Criterion Orients Enhanced Sampling by Low-Resolution Information

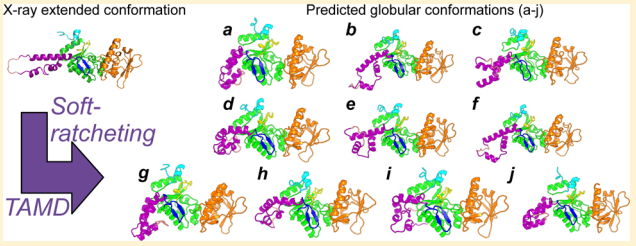
Isidro Cortes-Ciriano,[†] Guillaume Bouvier,[†] Michael Nilges,[†] Luca Maragliano,[‡] and Thérèse E. Malliavin*[,†]

[†]Unité de Bioinformatique Structurale, CNRS UMR 3528, Structural Biology and Chemistry Department, Institut Pasteur, 25-28, rue Dr. Roux, 75 724 Paris, France

[‡]Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genoa, Italy

Supporting Information

ABSTRACT: Many proteins exhibit an equilibrium between multiple conformations, some of them being characterized only by low-resolution information. Visiting all conformations is a demanding task for computational techniques performing enhanced but unfocused exploration of collective variable (CV) space. Otherwise, pulling a structure toward a target condition biases the exploration in a way difficult to assess. To address this problem, we introduce here the soft-ratcheting temperature-accelerated molecular dynamics (sr-TAMD), where the exploration of CV space by TAMD is coupled to a soft-ratcheting algorithm that filters the evolving CV values according to a predefined criterion. Any low resolution or even qualitative information can be used to orient the exploration. We validate this technique by exploring the conformational space of the inactive state of the catalytic domain of the adenyl cyclase AC from *Bordetella pertussis*. The domain AC gets activated by association with calmodulin (CaM), and the available crystal structure shows that in the complex the protein has an elongated shape. High-resolution data are not available for the inactive, CaM-free protein state, but hydrodynamic measurements have shown that the inactive AC displays a more globular conformation. Here, using as CVs several geometric centers, we use sr-TAMD to enhance CV space sampling while filtering for CV values that correspond to centers moving close to each other, and we thus rapidly visit regions of conformational space that correspond to globular structures. The set of conformations sampled using sr-TAMD provides the most extensive description of the inactive state of AC up to now, consistent with available experimental information.



INTRODUCTION

In the first days of structural biology, biomolecules were considered as rigid molecular objects, due in particular to the predominant role played by X-ray crystallography.[1,2] Nevertheless, during the three last decades, the vision of protein structure evolved and the importance of internal dynamics, of excited states, and of conformational transitions is now widely recognized. This view is illustrated, for example, by the hypothesis of prion structural transition,[3] by the discovery of the functional importance of intrinsically disordered proteins,[4] by the observation of conformational transitions in protein−ligand, protein−protein, and protein−nucleic acid interaction,[5−7] and by the detection of folding intermediates.[8]

Many computational methods have been proposed in order to explore protein conformational transitions. Performing extensive molecular dynamics (MD) simulations[9,10] underwent recently a large development.[11−15] Exploring conformational transitions can also be realized by biasing the coordinate evolution, which has been realized by a large variety of approaches, such as adaptive biasing force (ABF),[16,17] steered molecular dynamics (SMD),[18] and targeted molecular dynamics (TMD).[19] However, these methods require precise knowledge of the two extremities of the transition. Other methods such as metadynamics,[20,21] accelerated molecular dynamics (aMD),[22] and nonequilibrium candidate Monte Carlo[23] can perform exploration of the conformational space without having knowledge about the final state.

In temperature accelerated molecular dynamics (TAMD), proposed by Maragliano and Vanden-Eijnden,[24] a set of additional variables are introduced and coupled to the original system via collective variables (CVs). The original and new variables are then evolved concurrently but on two different time scales[25] and at different temperatures, the one for the extra variables being higher. In this way, the system follows the high temperature diffusion of the CVs over their free energy landscape defined at the physical temperature, thus being able to overcome barriers that are higher than the physical energy. Efficient, adiabaticity based, control of multitemperature systems was introduced long ago in MD,[26,27] and TAMD borrows from these and other approaches such as adiabatic free energy dynamics (AFED).[28,29] So far, TAMD has been successfully employed in several rare events studies, in particular protein conformational search.[30−40]

Received: February 16, 2015
Published: May 20, 2015

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454

In the present work, we propose to combine TAMD with a soft-ratcheting criterion (SR),[41] which uses a predefined variable to check whether sampling is progressing toward a target condition. More in detail, at each step, we use TAMD to propose new values of the CVs and we accept them if they progress toward the defined destination. If they lead away from the target, we accept them with a probability (see below), which gets smaller if the distance from the target is larger. The SR algorithm has already shown to give meaningful insight on transition mechanisms,[42−45] even when exploiting only low-resolution or qualitative information for the progress variable. Most importantly, it was shown to produce less biased transition paths than other approaches such as targeted MD.[41] The advantage of pairing it with TAMD is that we monitor progress in CV space, which implies that in the limit of adiabatic separation between the original and CV dynamics we are acting on the exploration of the *true* free energy surface. Notably, the adiabatic condition in TAMD is unaffected by the SR criterion and can be determined as illustrated previously.[29]

The behavior of sr-TAMD, TAMD, steered molecular dynamics (SMD), and molecular dynamics (MD) has been analyzed for the simple transition of alanine dipeptide (see the Supporting Information), and sr-TAMD displays faster transitions than the other approaches. The efficiency of the sr-TAMD approach is demonstrated on the adenyl cyclase (CyaA) domain (AC) from *Bordetella pertussis*, the agent of whooping cough.[46,47] CyaA is activated in the host immune system by the action of calmodulin, triggering overproduction of adenosine monophosphate (cAMP) which perturbs the cell's immune response. A high-resolution X-ray crystallographic structure of the CyaA catalytic domain in complex with the C-terminal lobe of CaM (C-CaM)[48] (Figure 1) shows an elongated conformation of CaM-associated AC. On the other hand, as aggregation of unbound AC domains prevents crystallization (A. Chenal and D. Ladant, personal communication), no experimentally determined structure is available for the isolated, CaM-free, AC protein. Extensive biophysical analyses, however, showed[49] that CaM bound and unbound AC conformations display a very similar secondary structure content, and that the unbound conformation is more globular than the bound state.

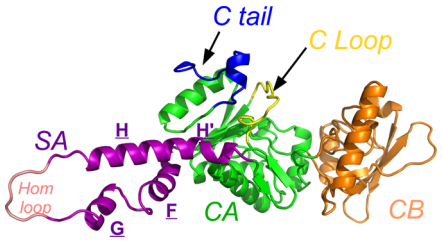


Figure 1. Conformation of the catalytic domain (AC) of the adenyl cyclase CyaA observed in the X-ray crystallographic structure (PDB entry 1YRT) of the AC/C-CaM complex, drawn in cartoon representation. AC comprises three separate domains, named CA, CB, and SA (respectively green, orange, and purple). The SA domain includes the α-helices H, F, G, and H′. The region SA colored in purple contains the α-helices F, G, and 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[56] using Modeller9v4.[70] 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 sr-TAMD approach is shown here to improve significantly the exploration of the conformational space of the isolated AC domain, in particular by inducing a more pronounced decrease of the protein gyration radius than a previous study[39] by standard TAMD. The stability (or long-time metastability) of the AC structures obtained by sr-TAMD is demonstrated via standard MD simulations. The final structures are analyzed to investigate the CaM-free conformation of the protein. These structures provide a picture of the CyaA inactive state which agrees with the available experimental knowledge.

1. SR-TAMD EXTENDS CONFORMATIONAL SPACE EXPLORATION

Starting from the AC crystallographic structure[48] and after having simulated a 40 ns MD trajectory, three sr-TAMD trajectories (Table S1, Supporting Information) were recorded using an artificial energy $\overline{\beta}^{-1}$ of 20 kcal/mol and the geometric centers CM_1^b, CM_2^b, CM_3^b, and CM_4^b [39] as collective variables. The scaling coefficient ρ of the soft-ratcheting criterion (eq 2 and Figure S1A, Supporting Information) determines how often the proposed target values of collective variables are accepted. ρ values greater than 1 Å did not permit the compaction of AC, whereas the ρ value of 0.5 Å induced a disruption of its secondary structure due to a drastic compaction of the protein. The percentage of rejected target values of collective values (Figure S1B, Supporting Information) is about 70% for ρ = 1 Å, and this value was thus chosen for the present study.

The AC conformations obtained at the end of sr-TAMD trajectories were extracted from the water box, resolvated, and submitted to a procedure of minimization, thermalization, and equilibration and then to unrestrained MD simulations (named CONT in Table S1, Supporting Information). This resolvatation procedure has the advantage to more strongly probe the stability of the globular conformation, with respect to the procedure starting a simulation from the solvated system at the end of the sr-TAMD trajectories. The system characteristics are given in Table S2 (Supporting Information). Fifteen trajectories were recorded starting from nine representative globular conformations sampled in each of the three sr-TAMD trajectories.

The previously recorded[39] MD and TAMD trajectories sample similar R_g values in the range 24.5−26.5 Å, whereas the current sr-TAMD samples a distinct range of R_g values: 22.5−24 Å (Figure 2). Furthermore, the continuing MD trajectories (Table S1, Supporting Information) starting from sr-TAMD conformations sample the same gyration values as the sr-TAMD trajectories, which supports the stability of globular conformations extracted from sr-TAMD.

The root-mean-square deviation (RMSD: Å) of α carbon coordinates along the sr-TAMD trajectories (Figure S2C, Supporting Information) reveals a variation of 8 Å during 2 ns, which is stronger and faster than the drifts smaller than 4 Å observed along the previously recorded 2 ns of TAMD trajectories (Figure 3 in ref 39). Furthermore, the R_g values decrease down to 22 Å (Figure S2A, Supporting Information). The variation of the radius of gyration of more than 3 Å from 25 to 22 Å is quite remarkable and shows that the sr-TAMD permits one to explore the regions of low gyration radius in the protein conformational space. Indeed, along previously recorded TAMD trajectories,[39] a value of 23.4 Å was observed for globular conformations.

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454

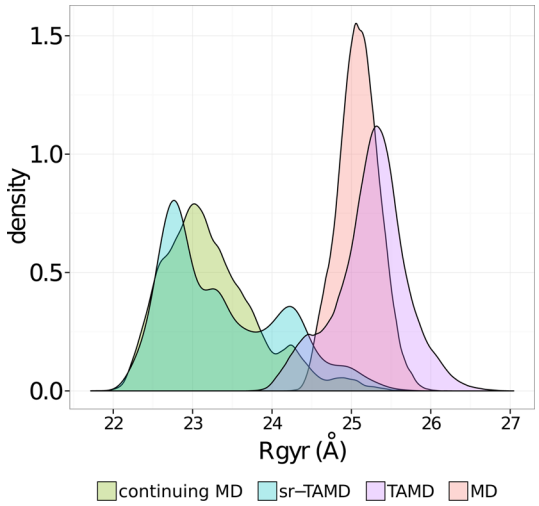


Figure 2. Distribution of gyration radius values sampled during MD (red), TAMD (blue) trajectories recorded in ref 39 and sr-TAMD (green), continuing MD (yellow) trajectories recorded in the present work. sr-TAMD enables the sampling of compact AC conformations displaying R_g values in the 22−24 Å range.

The drift of α carbon coordinates (Figure S2D, Supporting Information) for the 15 MD continuing trajectories shows that most of them display a conformational drift smaller than 5 Å. Only one trajectory comes back to the radius of gyration value of 25 Å observed in the extended conformation, and most of the trajectories display a value smaller than 24 Å (Figure S2B, Supporting Information). Thus, the continuing MD trajectories display metastable AC conformations that are markedly different from the elongated one.

The differences in the content of secondary structure between sr-TAMD, continuing MD, and the extended crystallographic conformation are mostly clustered in the α helices of the SA domain (Figure S3A,B, Supporting Information), and show that the isolated AC has a lesser α helix content in the SA domain. This agrees with the observations experimentally made by circular dichroism (CD).[49]

In conclusion, using the soft-ratcheting criterion together with TAMD permitted a broader exploration of the AC conformational space with respect to the previous TAMD calculation.[39] Furthermore, the sr-TAMD allowed us to obtain globular, metastable conformations as demonstrated by subsequent standard MD trajectories.

2. FEATURES OF THE INACTIVE STATE CONFORMATIONS

Although the sr-TAMD conformations are generated according to a biased probability, we find it instructive to inspect them. The 272 000 conformations sampled along the sr-TAMD and continuing MD trajectories (Table S1, Supporting Information), complemented with the extended conformation of AC in the X-ray crystallographic structure 1YRT,[48] were analyzed by the self-organizing map (SOM) approach,[50] described in more detail in the Supporting Information. This SOM approach is an unsupervised machine learning method which permits the clustering of a set of conformations without having to define a reference conformation, nor a superimposition procedure for the conformations.[51−53] The output of SOMs is represented as a U-matrix (Figure 3A), where each pixel is colored according

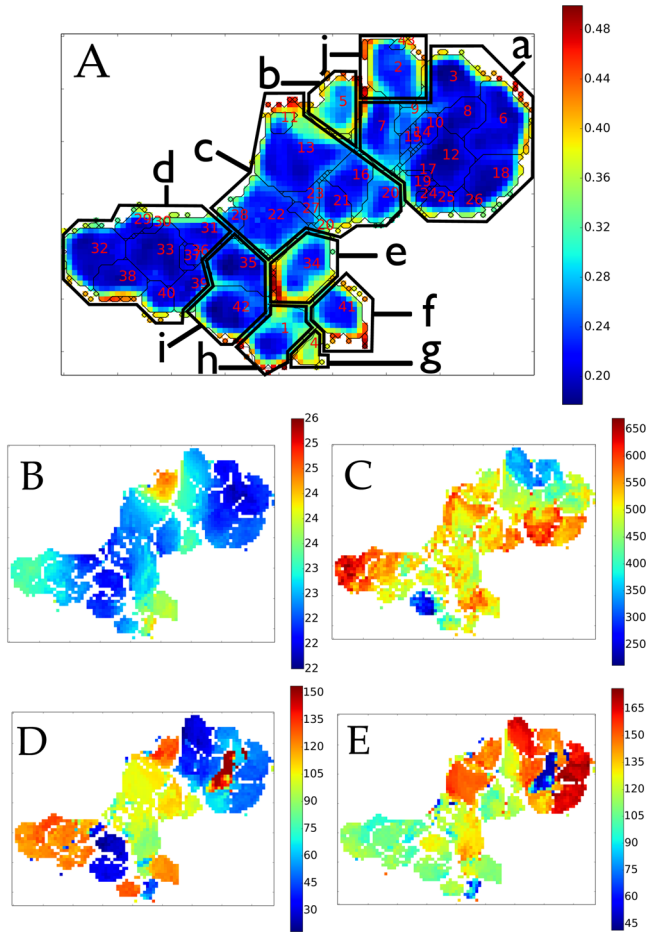


Figure 3. (A) U-matrix of the SOM analysis of 272 000 AC conformations sampled in sr-TAMD and MD continuing trajectories (Table S1, Supporting Information). The cluster labels (from 1 to 43) are plotted in magenta. The local similarity scale (right of the plot) between the conformations is expressed in Å. Projections on the U-matrix of the radius of gyration (B: Å), of the catalytic accessible surface (C: Å^2), and of the angle values between α-helices H and F (D: deg) and α-helices H and G (E: deg). In panels B−E, the neurons to which no value is projected as they contain no protein conformation are drawn in white.

to the local similarity between protein conformations (eq S4, Supporting Information). The region with the smallest values (colored in blue in Figure 3A) thus corresponds to the most homogeneous clusters of conformations, whereas larger values (colored in orange-red in Figure 3A) appear in the heterogeneous regions present between these clusters. Using a flooding approach (described in the Supporting Information), 43 clusters of conformations were determined from the U-matrix (Figure 3A). Using a barrier of local similarity equal to 0.36 Å in the U-matrix, 10 groups, labeled from *a* to *j* (Figure 3A and Table S3, Supporting Information), were determined and will be used in the following as representative descriptions of the conformations sampled by isolated AC.

A conformation was randomly extracted from each group (Figure 4). This set of representative conformations displays large reorganization of the SA, C terminal tail, and part of CA, arising from the changes in relative orientations of α helices of these regions. The SA reorganization supports the experimental observation[49] of a large CD variation in this region. The β hairpin (residues 259−273), colored in blue in Figure 4, is in

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454

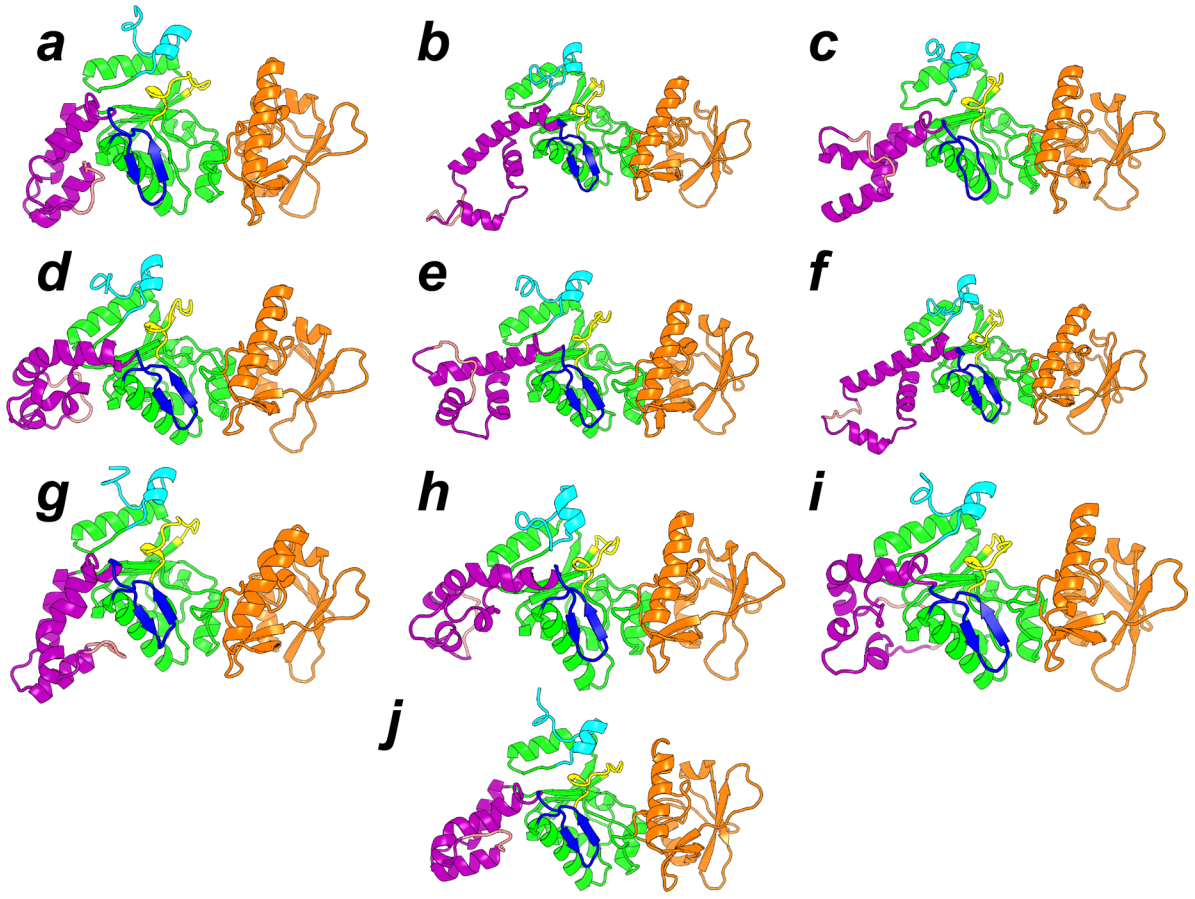


Figure 4. Representative conformations extracted from the 10 clusters detected from the SOM analysis (Table S3, Supporting Information). The conformation labels correspond to the labels written in Figure 3A. Each conformation is drawn in cartoons, with the region CA in green, the region CB in orange, and the region SA in magenta. The catalytic loop is colored in yellow, and the C terminal loop, in cyan. The β-hairpin region located in CA (residues 259−273) is colored in blue.

most of the conformations less accessible to the solvent, due to the vicinity of the region SA. This reduced accessibility of the β hairpin in the inactive state of AC agrees with its important role in the interaction between AC and the N-terminal lobe of CaM (N-CaM).[54]

The sr-TAMD trajectories srT4C-β20-A and srT4C-β20-B sample distinct groups of conformations *a* and *d*, whereas the conformations sampled in the sr-TAMD srT4C-β20-C are equivalently shared between *c* and *g* (Table S4, Supporting Information). The representative conformations of the groups *a*, *d*, and *g* correspond to a similar packing of the α helices of SA on the region CA (Figure 4). The continuing MD trajectories may visit several groups of conformations, but most of them sample the same group as that sampled during the corresponding sr-TAMD trajectory (Table S4, Supporting Information). This agrees with the sampling acceleration of TAMD with respect to MD, already observed in the analysis of the AChBP apo/holo transition.[40] Few exceptions to this rule are nevertheless observed: (i) the sampling of the group *j* in one MD trajectory continuing from srT4C-β20-A, (ii) the sampling of groups *h* and *i* in two MD trajectories continuing from srT4C-β20-B, and (iii) the sampling of groups *b*, *e*, and *f* in three MD trajectories continuing from srT4C-β20-C. The new groups sampled along MD trajectories (*j*, *h*, *c*, *b*, *e*, *f*) are always neighbors of the groups sampled in sr-TAMD trajectories (*a*, *d*, *i*, *g*) on the SOM U-matrix (Figure 3A).

Several geometric parameters have been projected on the U-matrix (Figure 3A): the radius of gyration of AC (Figure 3B), the catalytic accessible surface calculated with MSMS 2.5.3[55] on the catalytic residues K58, H63, K84, D188, D190, H298, N304, and F306 (Figure 3C), the angle between the axes of α helices H and G (Figure 3D), and the angle between the axes of α helices H and F (Figure 3E).

The radius of gyration R_g displays values in the 22−26 Å range, the maximum of 26 Å being only observed for the group *b* containing the extended conformation of AC observed in the X-ray crystallographic structure 1YRT.[48] In contrast, the smallest values of 22 Å are observed in groups *a*, *c*, and *d*, which are populated by conformations sampled during the three sr-TAMD trajectories, and in group *i*, sampled during a MD continuing from srT4C-β20-B (Table S4, Supporting Information). R_g values around 23 Å are observed for groups *f* and *g* and transiently in groups *a*, *c*, and *d*.

The catalytic accessible surface (Figure 3B) varies from 250 up to 650 Å^2. The upper range of values (colors yellow to red), in the range 500−650 Å^2, agrees with the values observed along the MD simulations previously recorded[56] on the AC/C-CaM complexes, in which AC is active. This agrees with the basal activity observed for AC in the absence of CaM.[57] Much smaller values in the 250−400 Å^2 range (colors blue to cyan), observed in all representative groups, correspond to reduced enzymatic activity for which the entrance of the substrates into the active site is obstructed.

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454

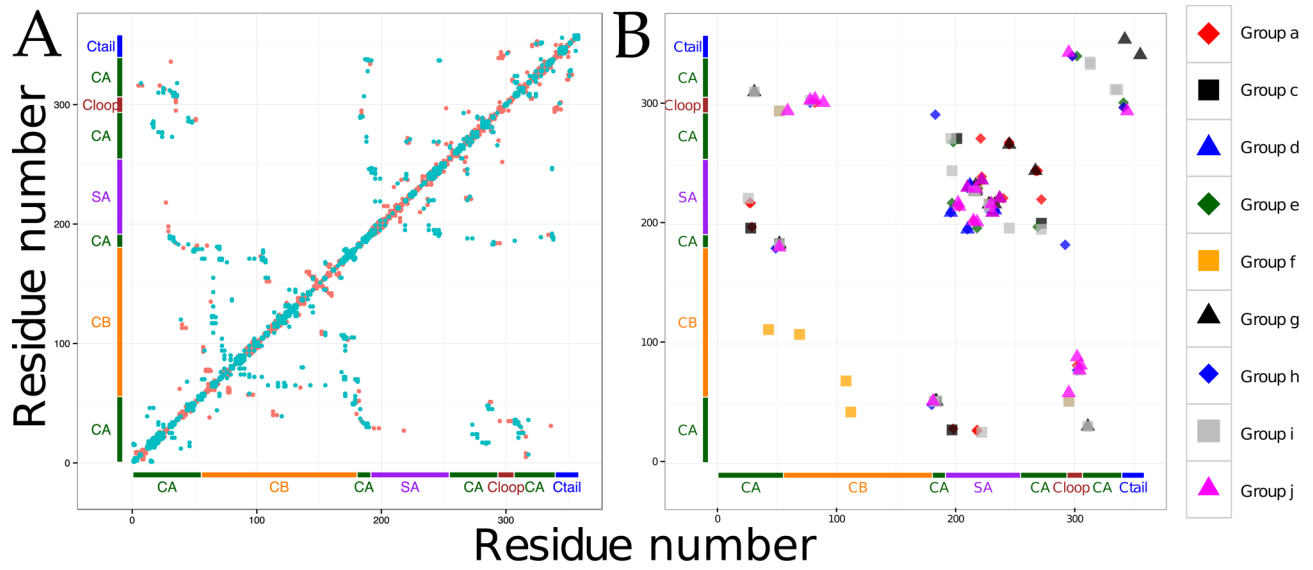


Figure 5. Analysis of interactions within the AC conformations. (A) Contact map representing interactions between residues in the representative conformation of group *a* (Table S3, Supporting Information, and Figure 4). Each point is drawn between residues for which the distance between geometric centers is smaller than 7 Å. The points are colored in red if the two residues are polar and charged and in blue if the two residues are hydrophobic (ALA, LEU, VAL, ILE, PRO, PHE, MET, GLY). (B) Contact map of the hydrogen bonds observed for more than 40% of TAMD and continuing MD trajectories and being not present in the extended crystallographic structure of AC.

Each group determined from the SOM clusters (Figure 3A) displays quite heterogeneous values for the catalytic accessible surfaces in contrast with the much more homogeneous values observed for the radius of gyration. Quite different enzymatic activities can thus correspond to a similar global shape of AC.

The angles between the axes of α helices H and F (Figure 3D) and between the axes of α helices H and G (Figure 3E) vary, respectively, in the intervals 20−150 and 45−170°, which span a large part of the possible angle values. The group *a* displays heterogeneous angle values, and the angle values between H and F and between H and G show mirror images, which agrees with an inverse orientation of the α helices F and G with respect to the α helix H observed in the representative conformations (Figure 4). Also, in a region located in the middle of group *a*, both angles are shifted by about 100°, which corresponds to a cooperative reorientation of the α helices F and G with respect to H. These variations illustrate the extreme plasticity of the region SA and agrees with the large fluorescence anisotropy of Trp-242 experimentally observed[58] for AC in the inactive state.

Overall, the AC conformations sampled along the sr-TAMD and the continuing MD trajectories display variable packing of the α helices in SA, and a large range of values for the catalytic accessible surface. These large variations prove the efficiency of sr-TAMD for providing a large exploration of the conformational space, while ensuring that the radius of gyration is kept in a lower range of values.

3. STABILITY OF THE AC CONFORMATIONS IN THE INACTIVE STATE

Contact maps between centers of mass of residue side chains closer than 7 Å were determined on the representative conformation of group *a* (Figure 5A). The contacts between hydrophobic (blue) and hydrophilic (red) residues are mostly superimposed. Similar contact maps are observed for all groups of representative conformations. This equivalence between polar and hydrophobic interactions justifies *a posteriori* the optimization approach, based on hydrogen bond formation, which was previously taken to get more globular conformations of AC.[39] In the following, we are thus focusing on a more detailed analysis of the hydrogen bonds in the representative conformations.

The contact map displaying the hydrogen bonds between protein residues present in more than 40% of the sr-TAMD and continuing MD trajectories and not present in the extended crystallographic conformation of AC (Figure 5B) reveals the appearance of a set of hydrogens inside the SA region. Other contacts appear between the C-loop and CA in groups *c*, *d*, *e*, *j*, and *h*, between the C-loop and CB for groups *j*, *i*, *h*, *f*, and *c*, and between CA and CB for groups *f*, *g*, *h*, and *j*. Interestingly, groups *i* and *j* contain conformations with a small catalytic accessible surface. The presence of the contacts involving CA, CB, and the C-loop could thus play a role in the closing of the catalytic site. The C-tail, which was shown[59] to be implicated in the AC activation, establishes contacts with CA and the C-loop.

The residues involved in the hydrogen bonds can be manually regrouped in eight sets (Table S5, Supporting Information). Three sets (1, 4, 5) are present in more than five groups, whereas the other sets are only present in a minority of groups: this reflects the diversity of conformations sampled by AC in the inactive site. The sets 1−3 are located at the CA−CB interface, at different regions of the catalytic site. The general deformation of the catalytic site induced by hydrogen bond formation as well as the involvement of residues GLU-301, GLU-308, ASP-310, and GLU-311 from the catalytic loop in these sets agrees with the decrease of AC enzymatic activity observed for isolated AC. The fourth set of residues is located on the catalytic loop and the C tail. Unsurprisingly, ASN-347 is present in this set of residues: the modification of these residues to ALA, performed simultaneously with other modifications, was shown[59] to have an influence on the interaction of AC with calmodulin. The fifth set of residues is located in SA and appears in all groups of conformations *a*−*j*. Sets 6 and 7 involve residues from CA and from α helices H

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454

and F. The last set of hydrogen bonds is formed between residues from the extremity of SA and residues from the N terminal part of AC (CA region), which are quite apart in the extended X-ray structure of AC (Figure 1).

The energies of interaction between the regions SA and CA and between the regions CA and CB were calculated using the VMD package namdenergy[60] and projected on the U-matrix (Figure S4, Supporting Information). The total energies measured between SA and CA (Figure S4B, Supporting Information) span less negative values than the ones measured between CA and CB (Figure S4A, Supporting Information), in agreement with the variability of SA positions with respect to CA. Groups *i* and *a* display the most negative interaction values between SA and CA, in agreement with the representative conformations where groups *i* and *a* display the SA region best packed on CA (Figure 4). Group *i* is the only group displaying simultaneously large attractive interactions for CA−CB and SA−CA. Overall, the interaction energies between CA and CB and between SA and CA display variability in agreement with the conformational equilibrium undergone by AC in the inactive state.

To summarize, the conformations sampled by AC are stabilized by polar as well as hydrophobic contacts. The analysis of hydrogen bonds permits one to determine the residues essential for the formation of the inactive AC state.

4. DISCUSSION

In the present work, we introduced a new method, the soft-ratcheting temperature accelerated molecular dynamics (sr-TAMD), which is a modification of the temperature-accelerated molecular dynamics (TAMD) algorithm, where the choice of the target values for collective variables is biased using a soft-ratcheting criterion.[41] The sr-TAMD is a new enhanced sampling method, with improved capacities for exploring the conformational space of biomolecules. This soft-ratcheting procedure was inspired by the Metropolis criterion,[61] and ensures that, if the application of the criterion conducts to a dead end in the conformational space, backward steps could be taken, in order to overcome local energy barriers. The application of the soft-ratcheting procedure does not contradict the previous hypotheses of adiabatic separation present in the TAMD algorithm, which insures that the system explores its real free energy surface (eq 2 in the Supporting Information): the target values of collective variables are chosen by soft-ratcheting among the values proposed by the TAMD algorithm.

The addition of the soft-ratcheting criterion presents an additional advantage with respect to TAMD: it permits a quite qualitative definition of the features searched for in the conformation of the biomolecular system. For example, in the present case, we only used the information that the AC shape should become more globular. This qualitative information was obtained from the hydrodynamic measurements performed on isolated AC[49] and did not require any information on the atomic coordinates of the target conformation. Other qualitative properties could be the relative displacements of various regions of a biomolecular system, or the validation of experimental parameters such as electronic microscopy envelopes. This qualitative criterion can be adapted to a large variety of situations, as (i) biomolecular domains coming closer together or moving apart, (ii) dissociation of protein−protein or protein−ligand complexes, (iii) shape deformation of biomolecules, and (iv) diffusion of ions or ligands on protein surfaces.

The sr-TAMD bears important differences to driven approaches as steered molecular dynamics (SMD)[18] or target molecular dynamics (TMD):[19] first, the ratchet-like probability filter perturbs the dynamics less than the above methods, and second, it affects the evolution of the CVs, i.e., dynamics on the free energy surface provided that the adiabatic condition in TAMD is satisfied. It is also important to note that, while the original SR algorithm is typically used within the framework of dynamic importance sampling (DIMS), where a reweighting is performed at each step to obtain the unbiased probability of the generated trajectory, here we suggest to use sr-TAMD only for a quick and minimally biased exploration of the free energy space without reweighting, and to validate the obtained path and structures afterward with more accurate approaches. As an example, we anticipate that sr-TAMD can be an efficient tool to obtain in short time high-quality initial paths for techniques such as the string method[24] or transition path sampling.[62] Combinations with recent, more refined importance sampling methods can also be envisioned.[63]

For many biomolecular systems undergoing functional conformational transitions, we only know one end point from high-resolution structural data. Moreover, often the transition involves an ensemble of conformations, as for example the unfolded state of a protein,[64−66] or multimerization/precipitation.[67] The approach suggested here may find many applications in studying biomolecular conformational transitions, and thus represents a milestone for the use of enhanced sampling methods for characterizing structurally undefined conformational states.

The efficiency of sr-TAMD has been validated on the simple conformational transition of the di-Ala peptide. Then, the application of the proposed approach sr-TAMD on the adenyl cyclase AC from *Bordetella pertussis* permitted the most extensive exploration available up to now of the inactive state of this protein. The analysis of the AC conformations sampled along sr-TAMD and continuing MD trajectories reveals a larger variability in the SA region and in parts of the CA region. The SA flexibility is produced by large changes in the relative orientations of α helices H, F, and G. The catalytic accessible surface varies also a lot along the conformations, being smaller or similar to the catalytic accessible surface of the active AC. This supports the residual enzymatic activity observed in isolated AC.[57]

The analysis of interactions in the conformations sampled in the AC inactive state showed that the hydrogen bonds play an important role in the conformation architecture and permitted us to identify pairs of residues whose interaction is essential to observe these conformations. The predicted representative conformations can be used in the future to (i) fit SAXS curves, (ii) predict residues cross-linking which would trap AC into its inactive state,[68] and (iii) propose possible transition paths from the active to inactive states of AC, which can be starting points for the calculation of free energy profiles.

5. CONCLUSIONS

We introduced soft-ratcheting TAMD (sr-TAMD) which uses a soft-ratcheting criterion[41] for filtering the values of the collective variables proposed by steps of TAMD.[24] The exploration of the free energy surface performed by TAMD is thus filtered on the basis of progress toward a target, which can be defined on the basis of qualitative information only. We validated the efficiency of this approach on the transition of the alanine dipeptide and on the inactive state of the adenyl cyclase

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454

AC from *Bordetella pertussis*. For AC, the exploration of regions of low radii of gyration was considerably enhanced with respect to the exploration performed with TAMD alone.[39] In this way, an extensive description of the conformations sampled by AC in an inactive state was provided.

6. EXPERIMENTAL SECTION

A general description of the TAMD approach is given in the Supporting Information.

We used as collective variables (CVs) the Cartesian coordinates of the same geometric centers that have been previously defined,[39] and we indicate them as CM_1^b, CM_2^b, CM_3^b, and CM_4^b. A soft-ratcheting criterion[41] was added to the equations describing the TAMD algorithm (eq S1, Supporting Information), in order to orient the trajectory by accepting or rejecting the proposed CV target values CM_{1new}^b, CM_{2new}^b, CM_{3new}^b, and CM_{4new}^b in the following way. First, the geometric center $G(CM_b^1, CM_2^b, CM_3^b, CM_4^b)$ of the CV current values was calculated as the average values of the x, y, and z coordinates of the centers CM_i^b. Also, the geometric center $G_{new}(CM_{1new}^b, CM_{2new}^b, CM_{3new}^b, CM_{4new}^b)$ is calculated in the same way from the coordinates of the proposed CM_{inew}^b target values for the CVs. Each pair of CM_i^b, CM_{inew}^b of geometric centers is then analyzed by comparing the distances $d(CM_i^b, G)$ and $d(CM_{inew}^b, G_{new})$. If each distance $d(CM_{inew}^b, G_{new})$ is smaller than the corresponding distance $d(CM_i^b, G)$, the candidate target values are accepted as current targets for collective variables. Otherwise, a random number r between 0 and 1 is generated and the target values CM_{inew}^b are accepted if r is smaller than the following factor F

$$F = f_1 f_2 f_3 f_4 \tag{1}$$

where each acceptance probability f_i is expressed (Figure S1A, Supporting Information) as

$$f_i = \exp(-(d(CM_i^b, G) - d(CM_{inew}^b, G_{new}))^2/\rho^2) \tag{2}$$

and ρ is the scaling coefficient of the soft-ratcheting criterion. This soft-ratcheting criterion has been introduced in the tcl script used to implement the TAMD algorithm in NAMD 2.7b2.[69] All details about setup of the systems and simulation runs are in the Supporting Information.

■ ASSOCIATED CONTENT

Supporting Information

A description (i) of the procedure for the initialization of simulations, (ii) of the theoretical background of the TAMD approach, (iii) of the algorithm of self-organizing maps, used to cluster the sampled AC conformations, (iv) of backward transition from a globular to an extended conformation of AC, (v) of the comparison of TAMD, sr-TAMD, MD, and SMD on the alanine dipeptide, and (vi) the ten representative conformations of the Figure 4 in PDB format. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00153.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: therese.malliavin@pasteur.fr.

Funding

I.C.-C. thanks the Paris-Pasteur International PhD Programme for funding. T.E.M. thanks CNRS and Institut Pasteur for funding. M.N. thanks the European Union (FP7-IDEAS-ERC 294809) for funding.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Jonathan Weare and Eric Vanden-Eijnden are acknowledged for fruitful discussions.

■ REFERENCES

(1) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181*, 662−666.

(2) Blake, C.; Koenig, D.; Mair, G.; North, A.; Phillips, D.; Sarma, V. Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthe sis at 2 Angstrom Resolution. *Nature* **1965**, *206*, 757−761.

(3) Harrison, R.; Sharpe, P.; Singh, Y.; Fairlie, D. Amyloid peptides and proteins in review. *Rev. Physiol. Biochem. Pharmacol.* **2007**, *159*, 1−77.

(4) Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **2012**, *37*, 509−516.

(5) Drum, C.; Yan, S.; Bard, J.; Shen, Y.; Lu, D.; Soelaiman, S.; Grabarek, Z.; Bohm, A.; Tang, W. Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature* **2002**, *415*, 396−402.

(6) Allain, F.; Gubser, C.; Howe, P.; Nagai, K.; Neuhaus, D.; Varani, G. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature* **1996**, *380*, 646−650.

(7) Russo, A.; Tong, L.; Lee, J.; Jeffrey, P.; Pavletich, N. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* **1998**, *395*, 237−243.

(8) Korzhnev, D.; Religa, T.; Kay, L. Transiently populated intermediate functions as a branching point of the FF domain folding pathway. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17777−17782.

(9) Duan, Y.; Wang, L.; Kollman, P. The early stage of folding of villin headpiece subdomain observed in a 200-ns fully solvated molecular dynamics simulation. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9897−9902.

(10) Liu, Y.; Strümpfer, J.; Freddolino, P.; Gruebele, M.; Schulten, K. Structural Characterization of -Repressor Folding from All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2012**, *3*, 1117−1123.

(11) Jensen, M.; Jogini, V.; Borhani, D.; Leffler, A.; Dror, R.; Shaw, D. Mechanism of voltage gating in potassium channels. *Science* **2012**, *336*, 229−233.

(12) Piana, S.; Lindorff-Larsen, K.; Shaw, D. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5915−5920.

(13) Piana, S.; Lindorff-Larsen, K.; Shaw, D. Atomistic description of the folding of a dimeric protein. *J. Phys. Chem. B* **2013**, *117*, 12935−12942.

(14) Salomon-Ferrer, R.; Goetz, A.; Poole, D.; Grand, S. L.; Walker, R. Routine microsecond molecular dynamics simulations with AMBER - Part II: Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.

(15) Stone, J.; McGreevy, R.; Isralewitz, B.; Schulten, K. GPU-accelerated analysis and visualization of large structures solved by molecular dynamics flexible fitting. *Faraday Discuss.* **2014**, *169*, 265−283.

(16) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008**, *128*, 144120.

(17) Dickson, B.; Legoll, F.; Lelièvre, T.; Stoltz, G.; Fleurat-Lessard, P. Free energy calculations: an efficient adaptive biasing potential method. *J. Phys. Chem. B* **2010**, *114*, 5823−5830.

(18) Wriggers, W.; Schulten, K. Investigating a back door mechanism of actin phosphate release by steered molecular dynamics. *Proteins* **1999**, *35*, 262−273.

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454

(19) Schlitter, J.; Engels, M.; Krüger, P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graphics* **1994**, *12*, 84−89.
(20) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(21) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(22) Markwick, P.; McCammon, J. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20053−20065.
(23) Nilmeier, J.; Crooks, G.; Minh, D.; Chodera, J. Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E1009−1018.
(24) 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.
(25) Vanden-Eijnden, E. Numerical Techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.* **2003**, *1*, 385−391.
(26) Sprik, M. Computer simulation of the dynamics of induced polarization fluctuations in water. *J. Phys. Chem.* **1991**, *95*, 2283−2291.
(27) Blöchl, P.; Parrinello, M. Adiabaticity in first-principles molecular dynamics. *Phys. Rev. B* **1992**, *45*, 9413.
(28) Rosso, L.; Minary, P.; Zhu, Z.; Tuckerman, M. On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *J. Chem. Phys.* **2002**, *116*, 4389−4403.
(29) Abrams, J.; Tuckerman, M. Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations. *J. Phys. Chem. B* **2008**, *112*, 15742−15757.
(30) 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.
(31) Abrams, C.; 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.
(32) Vashisth, H.; Maragliano, L.; Abrams, C. "DFG-flip" in the insulin receptor kinase is facilitated by a helical intermediate state of the activation loop. *Biophys. J.* **2012**, *102*, 1979−1987.
(33) Vashisth, H.; Brooks, C. 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.
(34) Nygaard, R.; Zou, Y.; Dror, R.; Mildorf, T.; Arlow, D.; Manglik, A.; Pan, A.; Liu, C.; Fung, J.; Bokoch, M.; Thian, F.; Kobilka, T.; Shaw, D.; Mueller, L.; Prosser, R.; Kobilka, B. The dynamic process of $\beta(2)$-adrenergic receptor activation. *Cell* **2013**, *152*, 532−542.
(35) Lapelosa, M.; Abrams, C. A Computational Study of Water and CO Migration Sites and Channels Inside Myoglobin. *J. Chem. Theory Comput.* **2013**, *9*, 1265−1271.
(36) Vashisth, H.; Abrams, C. 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.
(37) Scarpazza, D.; Ierardi, D.; Lerer, A.; Mackenzie, K.; Pan, A.; Banka, J.; Chow, E.; Dror, R.; Grossman, J.; Killebrew, D.; Moraes, M.; Predescu, C.; Salmon, J.; Shaw, D. 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.
(38) Vashisth, H.; Storaska, A.; Neubig, R.; Brooks, C. Conformational dynamics of a regulator of G-protein signaling protein reveals a mechanism of allosteric inhibition by a small molecule. *ACS Chem. Biol.* **2013**, *8*, 2778−2784.
(39) Selwa, E.; Huynh, T.; Ciccotti, G.; Maragliano, L.; Malliavin, T. Temperature-accelerated molecular dynamics gives insights into globular conformations sampled in the free state of the AC catalytic domain. *Proteins* **2014**, *82*, 2483−2496.
(40) Hosseini-Naveh, M.; Malliavin, T.; Maragliano, L.; Cottone, G.; Ciccotti, G. Conformational changes in acetylcholine binding protein investigated by temperature accelerated molecular dynamics. *PLoS One* **2014**, *9*, e88555.
(41) Perilla, J.; Beckstein, O.; Denning, E.; Woolf, T. Computing ensembles of transitions from stable states: Dynamic importance sampling. *J. Comput. Chem.* **2011**, *2*, 196−209.
(42) Shimamura, T.; Weyand, S.; Beckstein, O.; Rutherford, N.; Hadden, J.; Sharples, D.; Sansom, M.; Iwata, S.; Henderson, P.; Cameron, A. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **2010**, *328*, 470−473.
(43) Shimamura, T.; Weyand, S.; Beckstein, O.; Rutherford, N.; Hadden, J.; Sharples, D.; Sansom, M.; Iwata, S.; Henderson, P.; Cameron, A. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **1992**, *328*, 470−473.
(44) Perilla, J.; Leahy, D.; Woolf, T. Molecular dynamics simulations of transitions for ECD epidermal growth factor receptors show key differences between human and drosophila forms of the receptors. *Proteins* **2013**, *81*, 1113−1126.
(45) Perilla, J.; Leahy, D.; Woolf, T. Molecular dynamics simulations of transitions for ECD epidermal growth factor receptors show key differences between human and drosophila forms of the receptors. *Proteins* **2013**, *81*, 1113−1126.
(46) Walker, C. Whooping cough case numbers rise across the UK and US. *Nurs. Child. Young People* **2012**, *24*, 4.
(47) Murphy, J. Pertussis has re-emerged. *Ir. Med. J.* **2012**, *105*, 260.
(48) Guo, Q.; Shen, Y.; Lee, Y.; Gibbs, C.; Mrksich, M.; Tang, W. Structural basis for the interaction of Bordetella pertussis adenylyl cyclase toxin with calmodulin. *EMBO J.* **2005**, *24*, 3190−3201.
(49) Karst, J.; Pérez, A. S.; Guijarro, J.; Raynal, B.; Chenal, A.; Ladant, D. Calmodulin-induced conformational and hydrodynamic changes in the catalytic domain of Bordetella pertussis adenylate cyclase toxin. *Biochemistry* **2010**, *49*, 318−328.
(50) Bouvier, G.; Desdouits, N.; Ferber, M.; Blondel, A.; Nilges, M. An automatic tool to analyze and cluster macromolecular conformations based on Self-Organizing Maps. *Bioinformatics* **2015**, *31*, 1490−1492.
(51) Miri, L.; Bouvier, G.; Kettani, A.; Mikou, A.; Wakrim, L.; Nilges, M.; Malliavin, T. Stabilization of the integrase-DNA complex by Mg2+ ions and prediction of key residues for binding HIV-1 integrase inhibitors. *Proteins* **2014**, *82*, 466−478.
(52) Bouvier, G.; Duclert-Savatier, M.; Desdouits, N.; Meziane-Cherif, D.; Blondel, A.; Courvalin, P.; Nilges, M.; Malliavin, T. Functional motions modulating VanA ligand binding unraveled by self-organizing maps. *J. Chem. Inf. Model.* **2014**, *54*, 289−301.
(53) Spill, Y.; Bouvier, G.; Nilges, M. A convective replica-exchange method for sampling new energy basins. *J. Comput. Chem.* **2013**, *34*, 132−140.
(54) Springer, T.; Goebel, E.; Hariraju, D.; Finley, N. Mutation in the β-hairpin of the Bordetella pertussis adenylate cyclase toxin modulates N-lobe conformation in calmodulin. *Biochem. Biophys. Res. Commun.* **2014**, *453*, 43−48.
(55) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Fast and robust computation of molecular surfaces. In *Proc. 11th ACM Symp. Comput. Geom.*, Vancouver, June 5−7, 1995, ACM: New York, 1995; pp C6−C7.
(56) Selwa, E.; Laine, E.; Malliavin, T. Differential role of Calmodulin and Calcium ions in the stabilization of the catalytic domain of adenyl cyclase CyaA from *Bordetella pertussis*. *Proteins* **2012**, *80*, 1028−1040.
(57) Wolff, J.; Cook, G.; Goldhammer, A.; Berkowitz, S. Calmodulin activates prokaryotic adenylate cyclase. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 3841−3844.
(58) Gallay, J.; Vincent, M.; de la Sierra, I. L.; Munier-Lehmann, H.; Renouard, M.; Sakamoto, H.; Bârzu, O.; Gilles, A. Insight into the activation mechanism of Bordetella pertussis adenylate cyclase by calmodulin using fluorescence spectroscopy. *Eur. J. Biochem.* **2004**, *271*, 821−833.
(59) Selwa, E.; Chenal, A.; Sotomayor-Pérez, A.; Ladant, D.; Malliavin, T. Allosteric Activation of Bordetella pertussis Adenylyl

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454

Cyclase by Calmodulin: Molecular Dynamics and Mutagenesis Studies. *J. Biol. Chem.* **2014**, *289*, 21131−21141.
(60) Hsin, J.; Arkhipov, A.; Yin, Y.; Stone, J.; Schulten, K. Using VMD: an introductory tutorial *Curr. Protoc. Bioinf.* **2008**, Chapter 5, Unit 5.7.
(61) Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(62) Bolhuis, P.; Chandler, D.; Dellago, C.; Geissler, P. Transition path sampling and throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(63) Guttenberg, N.; Dinner, A.; Weare, J. Steered transition path sampling. *J. Chem. Phys.* **2012**, *136*, 234103.
(64) Liu, Z.; Huang, Y. Advantages of proteins being disordered. *Protein Sci.* **2014**, *23*, 539−550.
(65) Uversky, V. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci.* **2013**, *22*, 693−724.
(66) Dyson, H.; Wright, P. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54−60.
(67) Malinovska, L.; Kroschwald, S.; Alberti, S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim. Biophys. Acta* **2013**, *1834*, 918−931.
(68) Prevost, M.; Sauguet, L.; Nury, H.; Renterghem, C. V.; Huon, C.; Poitevin, F.; Baaden, M.; Delarue, M.; Corringer, P. A locally closed conformation of a bacterial pentameric proton-gated ion channel. *Nat. Struct. Mol. Biol.* **2012**, *19*, 642−649.
(69) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(70) Eswar, N.; Webb, B.; Marti-Renom, M.; Madhusudhan, M.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling using Modeller. *Curr. Protoc. Bioinf.* **2006**, *26*, 5.6.

DOI: 10.1021/acs.jctc.5b00153
J. Chem. Theory Comput. 2015, 11, 3446−3454