JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Finding Conformational Transition Pathways from Discrete Molecular Dynamics Simulations

Pedro Sfriso,[†] Agusti Emperador,[†] Laura Orellana,[†] Adam Hospital,[†,‡] Josep Lluis Gelpí,[†,§,∥] and Modesto Orozco*[,†,‡,∥]

[†]Joint IRB-BSC Program in Computational Biology, Institute of Research in Biomedicine, Josep Samitier 1-5, Barcelona, 08028, Spain

[‡]Structural Bioinformatics Node, Instituto Nacional De Bioinformática, Institute of Research in Biomedicine, Josep Samitier 1-5, Barcelona, 08028, Spain

[§]Computational Bioinformatics Node, Instituto Nacional De Bioinformática, Barcelona Supercomputing Center, Jordi Girona 29, Barcelona, 08034, Spain

[∥]Departament de Bioquímica, Facultat de Biologia, Universtitat de Barcelona, Avgda Diagonal 647, Barcelona, 08028, Spain.

Supporting Information

ABSTRACT: We present a new method for estimating pathways for conformational transitions in macromolecules from the use of discrete molecular dynamics and biasing techniques based on a combination of essential dynamics and Maxwell−Demon sampling techniques. The method can work with high efficiency at different levels of resolution, including the atomistic one, and can help to define initial pathways for further exploration by means of more accurate atomistic molecular dynamics simulations. The method is implemented in a freely available Web-based application accessible at http://mmb.irbbarcelona.org/MDdMD.

INTRODUCTION

Proteins are dynamic entities, whose conformations change in response to a variety of external stimuli, such as temperature, solvent composition, presence of ligands, and electric or mechanical fields.[1] There is now an overwhelming amount of evidence showing that protein function is directly related to protein flexibility,[2−4] and it is clear that evolution has made an effort to not only optimize protein structure but also to design flexibility patterns optimal for function.[5−10] Furthermore, it seems that evolution has used very often the intrinsic flexibility patterns of ancestor proteins to create new macromolecules,[2,11] in a conservative strategy to maintain fold and flexibility. Databases are full of examples where the same protein is found in different conformations due to the presence (or absence) of different ligands.[8,12,13] However, there are very few examples of experimental characterization of protein conformational transitions, since dynamic high-resolution techniques are still in their infancy,[10,14−16] which implies that most of the knowledge from conformational transitions in proteins is derived from molecular simulation techniques.[1,17−34]

Atomistic molecular dynamics (MD) is probably the most accurate and universal simulation technique for the study of protein flexibility. MD is based on a rigorous theoretical formalism and uses physical potentials (the force field) that have been refined and optimized for decades.[35,36] Unfortunately, practical use of MD is limited by the gap between the transition and the currently accessible simulation times, which precludes the use of direct-unbiased MD approaches to study large conformational changes. As a response to this problem, a variety of techniques have been developed to force the sampling along the direction of the transition.[37,38] These biasing techniques can provide encouraging results[38−45] but are very expensive computationally and can lead to incorrect results if the transition coordinate or the restraint protocols are not well tuned. In this complex scenario, morphing coarse-grained (CG) models have been gaining importance as an inexpensive alternative to obtain first guesses of the transition paths.[5−9,13,21,25,46−63]

Within the morphing CG paradigm, the protein is represented at the C_α level ignoring side chains. Transitions are simulated using different approaches, the simplest ones are based on interpolation schemes between original and final conformations using either Cartesian or internal coordinates.[64−69] A more physical variant of the morphing CG method is based on the assumption that evolution has created proteins precoded to perform biologically relevant transitions, which means that biologically relevant conformational transitions are likely to happen along soft modes,[13,53,70−73] i.e., the easiest deformation modes of proteins. These morphing methods require a definition of the protein Hamiltonian, which is often obtained by means of the elastic network models (ENM), where the *minimal frustration* principle[74] is assumed, and accordingly the perturbation energy associated with deformations of protein conformations from known experimental structure follow a harmonic behavior:

$$E = \sum_{i,j} \delta_{ij} K_{ij} (R_{ij} - R_{ij}^0)^2 \quad (1)$$

where i and j are residues, δ_{ij} is a delta function equal to 1 when i and j are at less than a given distance and 0 otherwise, K is a spring constant (linear or distance dependent), R_{ij} stands for inter-residue distance, and the superscript 0 refers to the value of R_{ij} in the reference structure.

Received: June 14, 2012
Published: August 23, 2012

The preferred deformation modes, i.e., those along which the energetic cost of deforming a protein is minimum, are obtained by diagonalization of the Hessian matrix associated with the Hamiltonian outlined in eq 1 (normal-mode analysis; NMA). Methods based on the EN-NMA approach provided then a guess of the transition by activating the movements along low frequency modes overlapping with the transition vector. Transitions obtained by animating natural deformation modes of proteins are more realistic than those obtained by simple interpolation schemes.[70] Unfortunately, EN-NMA morphing approaches also present some shortcomings; a very clear one is the corruption in the covalent structure of the protein related to large displacements along a limited number of modes. To partially alleviate these problems, different authors[51,70] have recalled the principle of minimal frustration also along the transition, and accordingly, lower modes are recomputed for intermediate structures obtained along the transition. Other authors[52,53] have developed EN-NMA in the internal coordinate space. Unfortunately, none of these elegant approaches is useful when transition is not coded in the first essential deformation modes of the protein, or when it implies side chain movements ignored in a C_α representation.

In this paper, we present a new method to trace large conformational transitions based on a very fast discrete molecular dynamics (dMD) algorithm. Plausible trajectories are obtained by following ballistic equations of motions are biased toward the target structure by means of a Maxwell−Demon engine (see below), which incorporates information of essential deformation movements of the protein. The method can work with any level of resolution (including all-atoms and hybrid levels), is very efficient computationally, and displays very good performance in a large variety of test systems.

METHODOLOGICAL APPROACH

Basic Discrete Molecular Dynamics Algorithm. The basic dMD formalism[62,75−77] assumes that particles move in the ballistic regime (constant velocity) until a particle−particle collision occurs. In dMD, the potential energy is defined with stepwise discontinuous functions of the particle−particle distance instead of continuous functions used in standard molecular dynamics. In the absence of any collision, the particles move linearly with constant velocity. The position of a given particle at the time of the next collision is

$$\vec{r}_i(t + t_c) = \vec{r}_i(t) + \vec{v}_i t_c \tag{2}$$

where $\vec{r}_i$ and $\vec{v}_i$ stand for positions and velocities and t_c is the minimum among the collision times t_{ij} between each pair of particles i and j:

$$t_{ij} = \frac{-b_{ij} \pm \sqrt{b_{ij}^2 - v_{ij}^2(r_{ij}^2 - d^2)}}{v_{ij}^2} \tag{3}$$

where r_{ij} is the square modulus of $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$, v_{ij} is the square modulus of $\vec{v}_{ij} = \vec{v}_j - \vec{v}_i$, $b_{ij} = \vec{r}_{ij}\cdot\vec{v}_{ij}$, and d is the distance corresponding to the wall of the square well.

When two particles collide, there is an elastic transfer of linear momentum into the direction of the vector $\vec{r}_{ij}$:

$$m_i\vec{v}_i = m_i\vec{v}_i' + \Delta\vec{p} \tag{4}$$

where the prime denotes the variables after the collision. The new velocities after collision are obtained by applying conservation rules:

$$m_i v_i + m_j v_j = m_i v_i' + m_j v_j' \tag{5}$$

$$\frac{1}{2}m_i v_i^2 + \frac{1}{2}m_j v_j^2 = \frac{1}{2}m_i v_i'^2 + \frac{1}{2}m_j v_j'^2 + \Delta V \tag{6}$$

where ΔV stands for the depth of the square well defining the interatomic potential.

The transferred momentum can be easily determined from

$$\Delta p = \frac{m_i m_j}{m_i + m_j}\left\{\sqrt{(v_j - v_i)^2 - 2\frac{m_i + m_j}{m_i m_j}\Delta V} - (v_j - v_i)\right\} \tag{7}$$

Note that the two particles can go out of the well as long as

$$\Delta V < \frac{m_1 m_2}{2(m_1 + m_2)}(v_j - v_i)^2 \tag{8}$$

Otherwise, if the particles remain in the well, eq 7 reduces to

$$\Delta p = \frac{m_i m_j}{m_i + m_j}\{\sqrt{(v_j - v_i)^2} - (v_j - v_i)\} \tag{9}$$

which, taking the negative solution of the root, leads to

$$\Delta p = \frac{2m_i m_j}{m_i + m_j}(v_i - v_j) \tag{10}$$

In summary, in dMD no forces should then be calculated; the equations of motion should be integrated on the femtosecond scale, yielding then to very efficient simulations.[58,60,78−80] Previous works in our group have shown how dMD is able to reproduce well equilibrium dynamics of proteins as determined by explicit-solvent atomistic MD simulations[58,60,79] and can be used to perform robust minimizations of protein−protein complexes (manuscript in preparation). Other authors demonstrated the sampling capability of dMD folding small proteins.[76,81] Also, encouraging results from dMD have been reported in the analysis of many aspects of protein and nucleic acids dynamics,[61−63,77,82−85] macromolecular aggregation,[56,86−88] and macro- and supramolecular transitions.[89]

Force Field Description. dMD is based on the use of simple or multiple-step square potentials to describe physicochemical interactions. Our dMD interaction potentials include an implicit solvation term (derived from Lazaridis−Karplus functions[90]) and van der Waals and electrostatic terms. In the aim of increasing the speed of the simulations, we have chosen our simplest version of dMD:[59] in the case of attractive interactions, two-step potentials that define square wells are used, and a soft barrier in the case of repulsive interactions. As in standard dMD calculations, infinite wells were used to maintain all bonds and angles in the protein near equilibrium values, preventing then distortions of the chemical backbone. In this implementation of the method, well-defined secondary structure elements were enforced during the transition by defining very deep square potentials between hydrogen bonded groups. These secondary-structure constraints are automatically released in cases where initial and final secondary structures do not match.

Biasing Techniques. The core of our morphing procedure is a biasing algorithm, which enhances dMD sampling in the direction of the transition and also, if possible, along the essential deformation modes (Figure 1). The first is quantified by simple metrics, which in principle are applied only to the C_α's:

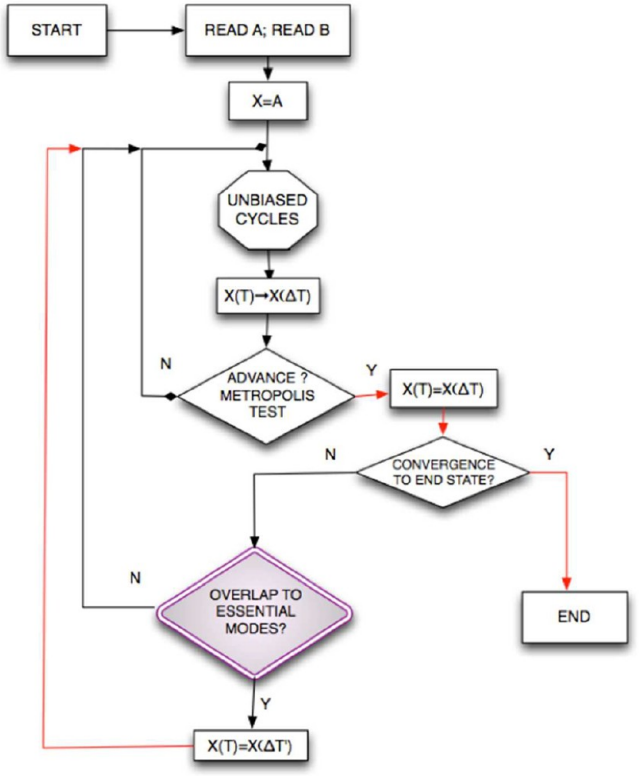

Figure 1. Flowchart of the basic MDdMD algorithm. Red lines indicate a definitive temporal advance toward the target structure. NMA biasing criteria (purple) only apply under certain conditions (see Figure 2).

$$\Gamma = \sum_{i=1}^{N} \| \vec{r}_{i,\mathrm{B}} - \vec{r}_{i,\mathrm{X}} \| \omega(i) \tag{11}$$

where N is the total number of residues, B is the target structure, X is the sampled conformation (X = A for the original conformation), and $\omega(i)$ is a weighting function defined as

$$\omega(i) = \begin{cases} 0 & \text{if } \| \vec{r}_{i,\mathrm{B}} - \vec{r}_{i,\mathrm{X}} \| < r_{\mathrm{cut}} \\ \| \vec{r}_{i,\mathrm{B}} - \vec{r}_{i,\mathrm{X}} \| & \text{if } \| \vec{r}_{i,\mathrm{B}} - \vec{r}_{i,\mathrm{X}} \| > r_{\mathrm{cut}} \end{cases} \tag{12}$$

where r_{cut} is an estimate of oscillation around equilibrium structures generated by the thermal noise (1.5 Å from the study of our MODEL database[91]).

We define the transition vector ($\Delta\vec{R}$) as that connecting the sampled structure to the target one ($\vec{R}_B - \vec{R}_X$) and the essential transition vector defined from the combination of eigenvectors that better reconstructs the transition:

$$\vec{V}_{\mathrm{NM}} = \sum_{j=1}^{m} \alpha_j \vec{v}_j \tag{13}$$

where $\vec{v}j$ stands for the eigenvectors obtained from normal-mode analysis (see below), and α_j is the normalized overlap between the selected mode (j) and the transition vector. In order to reduce the noise, the sum extends for the m modes with $\alpha_j > 0.15$ pertaining to the group of 10 lower frequency ones (the essential deformation space).

Following a Maxwell−Demon approach, the bias toward the target structure is not introduced by an energy penalty, but using informational criteria.[39,92] Thus, after at a certain simulation step (t), the progress variable (Γ) is computed (eq 11) and compared with that obtained in a previous accepted movement ($t - \Delta t$). Following the Metropolis Monte Carlo procedure, the simulation segment ($t - \Delta t$)→(t) is preaccepted or not based on the probability p_t:

$$p_t = \begin{cases} 1 & \text{if } \Gamma_t < \Gamma_{t-1} \\ \exp\left[-\frac{1}{2} \left(\frac{\Gamma_t - \Gamma_{t-1}}{\beta \mathrm{RMSD}_{(X,t,B)}} \right)^2 \right] & \text{if } \Gamma_t > \Gamma_{t-1} \end{cases} \tag{14}$$

where β is dynamically adjusted to guarantee a user-input acceptation rate (recommended value around 40%), and the time frame (Δt) is typically 100 RTU (see eq 15; in our experience, 1 RTU corresponds to around 10−20 ps of standard protein dynamics).

$$\mathrm{RTU} = \frac{0.15 \text{ Total Collisions}}{\#\text{residues}} \text{ at } T = 300 \text{ K} \tag{15}$$

The protocol outlined above is very efficient for driving transitions (without aberrant contacts or distortions of chemical structure), but it does not guarantee that such a transition follows essential deformation movements, which in some cases might bias the transition to biologically unlikely paths. To guarantee that, if possible, transition uses the default conformational flexibility of the protein, we compute the overlap between the essential transition vector ($\vec{V}_{\mathrm{NM}}$) and the transition vector ($\Delta\vec{R}$), taking first $\vec{V}_{\mathrm{NM}}$ defined from the eigenvectors of the original structure (A). We found that, if the overlap is above 0.6, we can consider that essential deformation space contains information on the transition and proceed as described in the following paragraph. When the overlap is smaller than this cutoff, we consider that the essential deformation space does not contain useful information to improve our definition of the transition pathway, and transition is fully guided by the dMD force field and the Maxwell−Demon.

Assuming that there is a good overlap between the transition vector and the essential transition vector defined from eigenvectors of the original structure (X = A), we incorporate additional conditions into the preaccepted configuration as defined by probability function:

$$p_{\mathrm{NM}} = \begin{cases} 1 & \text{if } \tau > T \\ \exp\left[-\frac{(T - \tau)^2}{\beta_{NM}} \right] & \text{if } \tau < T \end{cases} \tag{16}$$

where the index τ is the normalized projection between two structures separated by a significant period of time ($\Delta t'$) in the range 20−100 RTU ($\Delta t'$ is automatically adjusted depending on protein size) and T is by default 0.6. The β_{NM} factor is set to 0.1 to guarantee a smooth evolution of the probability function in the range 0.15−0.6.

As noted by others (see the Introduction), when the structure moves from the starting conformation (i.e., X ≠ A), the principle of minimal frustration is not granted, and accordingly eigenvectors computed for A lose predictive power. Thus, when our algorithm detects that the transition vector between target structure and current transition structure at $t + \Delta t'$ ($X_{t+\Delta t'}$) does not overlap with the essential transition vector determined from the eigenvectors for the starting structure (X = A), it assumes that original essential deformation space is no longer informative. At this point, if the transition structure ($X_{t+\Delta t'}$) is close to the final conformation (X = B; based on a spherical cutoff

adjusted to the size of the protein, typically around 3.5 Å), we use the eigenvectors of B to compute the essential transition vector. Otherwise, we recomputed the eigenvectors assuming that structure $X_{t+\Delta t'}$ is a minimum and applied the same protocol outlined before (see Figures 1 and 2).

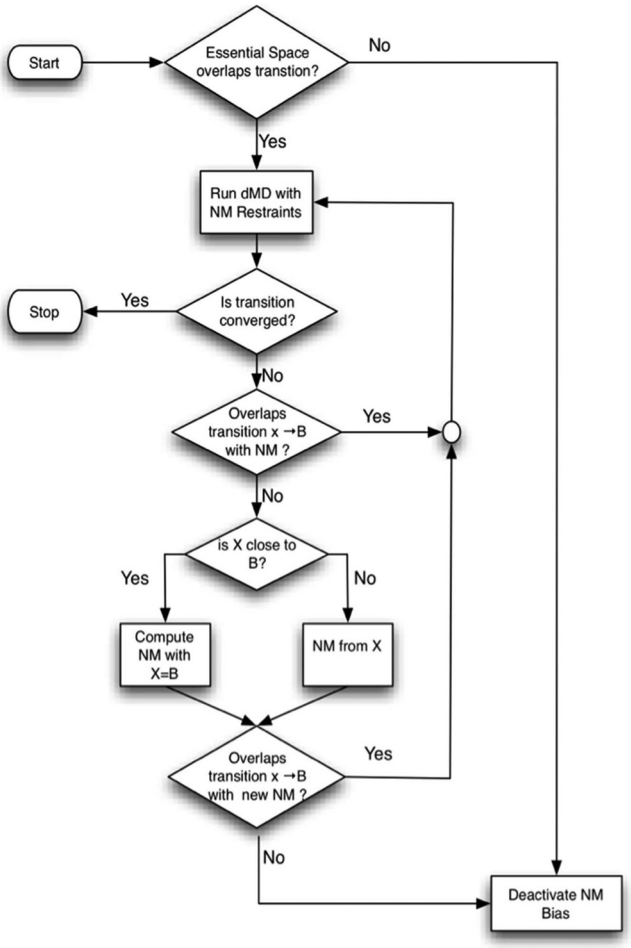

Figure 2. Detail on the implementation of the NMA bias based on the initial and current overlap between transition and essential deformation space.

In a small number of cases, especially when transition is close to the target structure, the essential deformation movements are not very useful for driving the transition. In this case, eq 16 increases dramatically the rejection rate, making final convergence slower. The algorithm recognizes automatically this situation, inactivating from this point of the introduction of information on essential modes as a guide for the transition (Figure 2).

Computing Normal Modes. The essential deformation space of a given structure X is computed using our version of the elastic network model. It is based on a Go-like harmonic potential with differential sequential and spatial cutoff functions and a Kovac's distance dependent force constant adjusted to reproduce essential dynamics obtained from atomistic molecular dynamics simulations (see refs 13 and 80 for details of the method). It is worth noting that there is a very good agreement between the type of flexibility predicted by this EN-NMA model and that obtained by dMD simulations,[59,60] which guarantees the physical consistency of the present hybrid model. The EN-NMA routine is incorporated in our code for recomputing eigenvectors when required.

Convergence into Target Basin. It is not trivial to determine when a transition has reached the target structure, since protein structures are continuously moving by thermal noise (in average 1−2 Å based on atomistic MD simulations). Bearing this in mind, we have decided not to attempt to reach a very low RMSD to the target conformation. In fact, even for a perfect force field, it is unrealistic to obtain a zero RMSD structure when a protein structure is naturally oscillating at room temperature.[22]

Thus, we adopt here a convergence criterion based on the slope of RMSD to target structure respect to time. This criterion prevents our method from obtaining stressed structures that would be only an artificial minimum as a result of the strong bias necessary to achieve close to zero RMSD values, but the user should be aware that a further refinement using explicit solvent atomistic MD simulation might be necessary. Thus, in our procedure, below a user-provided RMSD cutoff (recommended values around 2 Å for a 200 residues protein to 3 Å for a larger proteins around 600−700 residues), we conclude that the system has reached the target equilibrium state (near target experimental structure) when in the last 20−40 Reduced Simulated Time Units (RTU; see eq 15) of dynamics there is a negligible change in RMSD.

RESULTS AND DISCUSSION

Transition Characterization. We tested the ability of our method to obtain reasonable estimates of pathways for conformational transitions by exploring a large database of cases where there are at least two clearly different structures in the protein data bank (PDB[93]). The validation data set (see Table 1) contains 47 protein pairs ranging from small (around 100) to very large (around 1000) proteins. In all cases, trajectories have been followed in both directions, yielding a total of 94 transitions, all of them followed at the all-heavy-atoms level of resolution. Conformational changes in the flexible part (obtained by computing RMSD after manual alignment of rigid part) vary from small (less than 3 Å) to very large (more than 20 Å), which means that we are trying to reproduce not only trivially small transitions but also massive conformational changes (see Figure 3), which are more challenging for transition pathway detectors. Analysis of the overlap between the transition vector and the lowest eigenvectors of the equilibrium conformations shows in general a reasonable overlap between the transition and the essential deformation pattern of proteins (see Table 1 and Figure 3), confirming that often biologically relevant conformational changes follow the essential deformation modes of proteins.[3,69] However, 68 of the 94 transitions display overlaps below 75%, and 34 of them below 50% (Figure 3), meaning that a significant number of transitions cannot be simply explained by the essential deformation pattern of proteins. Finally, it is worth noting that the reversibility in the transitions is not always granted, since overlaps between essential deformation and transition vectors are significantly different when considered in the A→B and B→A directions for a non-negligible number of cases (see Table 1).

A detailed analysis of the entire data set reveals that 44% of transitions correspond to conformational changes between the open and closed forms of the protein, 39% to induced binding to other macromolecules, 60% to the binding of small ligands or cofactors, and one requires post-translational modifications (see Table S1). The database contains no example of trivial

Table 1. List of Proteins Considered in the Validation of the Method[a]

structure pair	# residues	RMSD (Å)	overlap	ΔRg/Å	difficulty
1yzy*/2ahm*	71	7	0.39/0.45	0.16	++
1szv*/1vet*	91	5.24	0.23/0.26	1.05	++
1l5e/1l5b	101	6.7	0.81/0.83	0.31	−−
1wrp/3wrp	108	2.48	0.49/0.46	0.61	+−
1xfr*/2fjy*	123	5.84	0.30/0.32	0.95	++
1e7xA/1dzsB	129	3.4	0.07/0.10	0.90	+−
1cfd/1cfc	148	5.43	0.94/0.92	1.74	−−
1h2d*/1oc3*	158	8.6	0.27/0.27	1.53	++
2gja/1rfl	162	8.71	0.20/0.33	2.04	+−
1r3e*/1ze1*	169	5.94	0.29/0.33	0.22	++
1ybj*/1dk0*	173	5.64	0.24/0.22	0.06	++
1aje*/1ees*	174	6.82	0.39/0.56	1.81	++
1cbuB/1c9kB	180	3.55	0.38/0.52	0.32	+−
1ex6/1ex7	186	3.64	0.88/0.79	0.83	−−
1s2h*/1go4*	190	4.9	0.48/0.56	0.28	+−
1bcc/2bcc	196	7.45	0.53/0.54	0.52	+−
2rh5/2rgx	202	5.99	0.88/0.62	2.51	+−
4ake/1ake	214	7.19	0.88/0.62	3.08	−−
1ggaA/1wdnA	220	5.34	0.86/0.61	1.56	−−
2lao/1lst	238	4.81	0.90/0.59	1.53	−−
3pjr*/1qhh*	261	9.38	0.47/0.61	0.33	+−
1urp/2dri	271	4.24	0.92/0.88	1.01	−−
1ram/1leiA	273	3.38	0.94/0.90	1.46	−−
5at1/8atc	310	2.59	0.66/0.45	0.50	+−
1 cmkA/1 cmkB	317	3.62	0.94/0.72	1.27	−−
3dap/1dap	320	4.35	0.90/0.74	1.30	−−
1eyk/1nuz	327	4.54	0.56/0.28	1.00	+−
1bp5/1a8e	329	6.81	0.86/0.67	1.98	−−
1jqj/2pol	366	1.99	0.48/0.38	0.46	+−
1omp/1anf	370	3.91	0.86/0.86	0.90	−−
8adh/6adh	374	1.27	0.24/0.31	0.16	+−
9aat/1ama	401	1.67	0.35/0.44	0.27	+−
1ux5/1y64	411	10.33	0.84/0.61	3.75	++
1qf5/1hoo	431	3.03	0.32/0.56	0.65	+−
1yyo/1yyw	438	17.96	0.14/0.40	2.37	++
1bnc/1dv2	452	4.51	0.83/0.79	1.56	−−
1rkm/2rkm	517	3.24	0.92/0.66	0.58	−−
1sx4/1oel	524	12.61	0.77/0.76	3.87	−−
1hp1/1hpu	525	9.93	0.52/0.53	0.34	++
2hmi/3hvt	556	3.45	0.59/0.60	0.61	+−
1i7d/1d6m	620	3.65	0.61/0.48	0.08	+−
8ohm/1cu1	645	4.62	0.77/0.71	0.29	+−
1lfg/1lfh	691	6.54	0.76/0.85	1.08	−−
1qvi/1kk8	837	27.61	0.72/0.78	1.70	+−
1q9x/1q9y	899	5.91	0.81/0.46	0.57	+−
1ih7/1ig9	903	6.47	0.77/0.48	0.68	+−
1su4/1iwo	994	13.93	0.71/0.56	0.92	+−

[a]For each transition, we quoted the starting and final structures (in PDB code), the size (in number of residues) of the protein, the root mean square between the two structures, the change in radius of gyration, the overlap between the transition vector, the expected difficulty of the transition (see main text), and the essential deformation space of the reference proteins (first 10 modes, overlap is computed for each pair (a/b) in both directions: a→b, first number and b→a, second number).

conformational changes, but it is clear that the degree of complexity of finding a reasonable transition path is not homogeneous across the data set. Thus, we classified transition

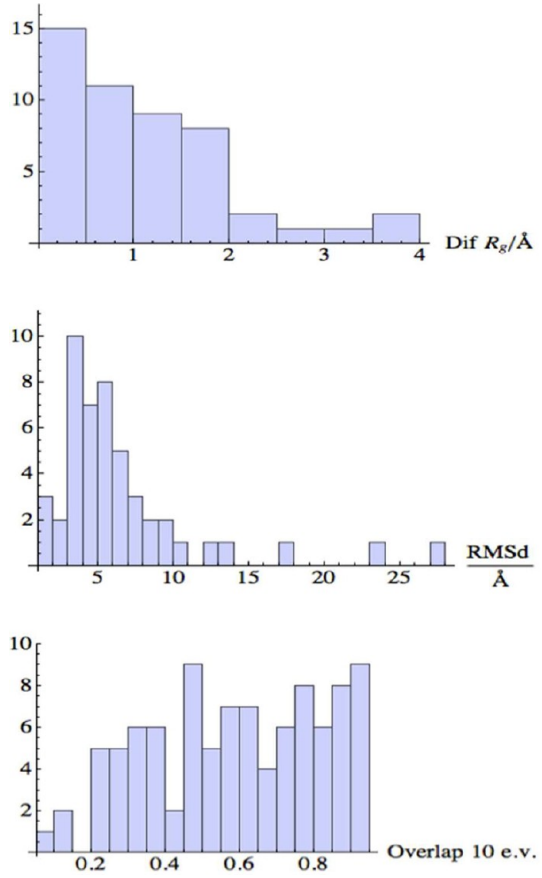

Figure 3. Distribution of difference in radius of gyration (top), between starting and final structures, RMSD between both stats (middle), and overlap of initial essential deformation space with transition (bottom) for all the transitions considered.

between all protein pairs in three categories (simple (−−), difficult (−+), and very difficult (++)) based on three simple descriptors: (i) RMSD between the pairs, (ii) the maximum overlap between the transition vector and essential deformation modes, and (iii) the expected reversibility. On the basis of this, our extended database contains 33% simple transitions, 46% difficult transitions, and 22% very difficult transitions (see Table S1).

Global Performance of the Method. We managed to reach the final structural basin for 91 of the 94 transitions (see Table S2), without any *ad hoc* adjustment of the method for difficult cases. Looking in detail to the final structures, we did not find local errors, like too large bonds or unrealistic angles or dihedrals, and only in three cases did we detect some problems in local geometry (see Table S2) corresponding to cases where experimental transition implies a disruption of elements of the secondary structure, which were not captured by our default setup. In summary, our method has a maximum failure ratio of only 6% when exploring a large database of transitions, some of them very difficult to trace (see Table 1). From 94% of the successful cases, we can detect a few cases where the transition is too small and there is overlap between the starting and final basins, making it difficult to guarantee the success of the transition. In these cases (six cases from the 96 considered; see Table S2), direct atomistic studies based on umbrella sampling, targeted MD, and metadynamics of alternative biasing techniques seem a more sensible election than coarse-grained approaches.

dx.doi.org/10.1021/ct300494q | J. Chem. Theory Comput. 2012, 8, 4707−4718

It is worth noting that essential deformation as defined by normal-mode analysis helps to find a reasonable pathway in around 70% of the cases; even in a significant number of cases (36%), the reference structure used to determine the essential deformation space was changed in the course of transition (see Methodological Approach). The method managed to obtain physically plausible transitions (see Figure 4) in cases of dramatic changes (see for example 1qvi←→1kk8 or 1sx4←→1oel transitions in Table S2), in very large systems (see for example 1su4←→1iwo or 1ih7←→1ig9), and also in cases (like 1ybj←→1dk0 and 1h2d←→1oc3) where there is a very poor overlap between essential deformation space and the transition vector (see Figure 4 and Table S2). Using the reduced simulation trajectory time (RTU relative to size and RMSD of the transition), we can evaluate the real difficulty of our procedure to reach the target structure. As shown in Table S3 and Figure S1, the method finds transition paths very quickly in most cases. In general, the cases (Figure S1) where pathways are difficult to find correspond to transitions labeled as "very difficult" in Table 1, typically cases where essential deformation space is quite orthogonal to the transition, and where there is a large ratio of rejection by the Maxwell−Demon algorithm. Translation of reduced simulation time to wall-clock time is difficult, but for most (60%) of the transitions outlined here, the algorithm finishes in less than 1−2 h on a standard laptop computer. Transition pathways obtained with MDdMD are consistent with all the structural parameters considered in our physical based force field, though we expect them to be a valuable initial estimation. However, we should stress that the first guess of transition path may still not be the kinetically optimal one, so further refinements using much more rigorous (and computationally expensive) methods might be needed.

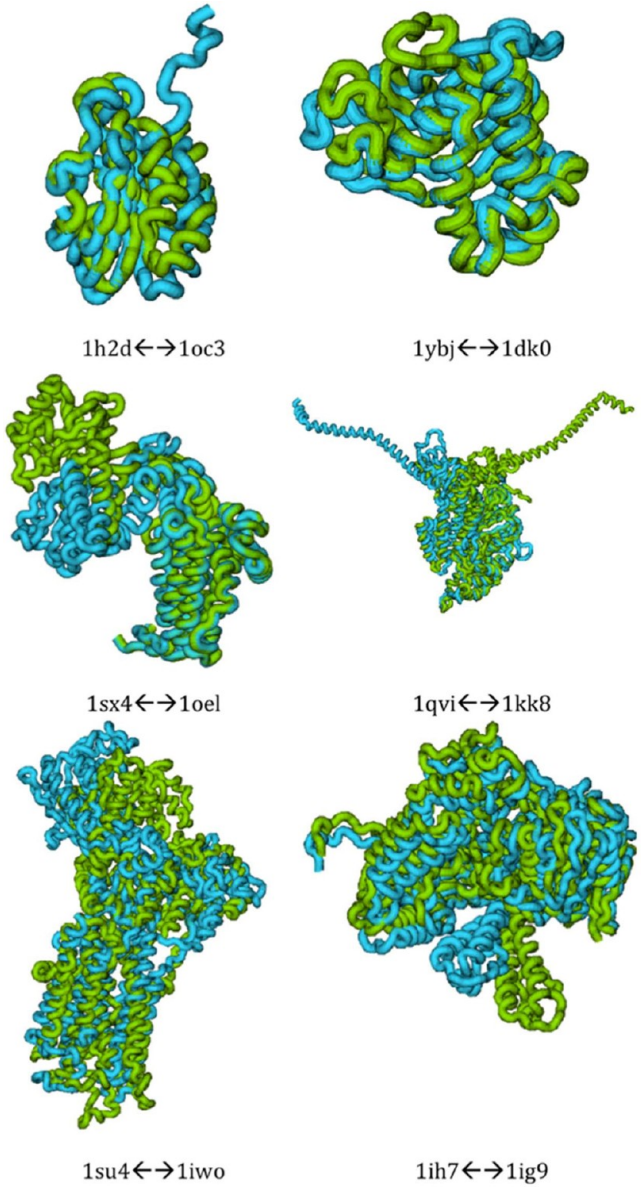

Figure 4. Structural ribbon superposition of both experimental ends of some nontrivial conformational transitions explored.

Reversibility. Experimentally, there are many transitions which are not equally easy in both directions (A→B and B→A). This is clearly visible in cases where the overlap between the transition vector A←→B and the essential deformation spaces of A and B are very different (see Table 1), warning about reversibility problems in our simulations. Fortunately, even the reduced simulation times in the A→B and B→A directions can be quite different for some pairs of proteins (see Table S2); there is only one case (adenosylcobinamide kinase, 1cbu/1c9k) where our method shows irreversibility. Very interestingly, the analysis of the RMSD bidimensional plots (see examples in Figure 5) indicates that for a given point in the transition paths, structures sampled in the A→B and B→A directions are quite similar, suggesting microscopic reversibility in the transition. The same is clear by looking at the evolution Maxwell−Demon acceptance rate and the RMSD (to target) along the transition (see examples in Figure 5). In summary, our method, which is not based on interpolation or on the use of geometrical energy restraints, is able to find with reduced computational effort feasible and (macro- and microscopically) reversible transitions between distant conformations.

Local Geometrical Quality. One of the main problems of interpolation techniques and NMA-based morphing techniques relies on the lack of quality of the local geometry, which might display unrealistic bond lengths or angles, or even steric clashes. Our procedure, which is based on a very simple, but still reliable, physical potential, eliminates completely these problems. No violation of chemical connectivity or steric clashes are found during transitions. PROSA profiles[94] reveal that not only final structures but also generated intermediate conformations fulfill the standard requirements of a folded protein (see examples in Figure 6). Violations of Ramachandran's maps are quite small along the transitions (Figure 7), and the pattern of native contacts predicted by our method matches in general that found experimentally (Figure 7). It is very encouraging that even the current version of the method uses only the deviation from the target structure of the C_α's as a decision variable for the Maxwell−Demon selection procedure; side chains are typically well positioned (see examples in Figure 7). In any case, it is worth it to note that any error in side chain positioning can be easily corrected in our algorithm by just including information-biased torsion parameters in the dMD potential function.

Sampling of Transition Intermediates. The procedure outlined here allows a fast determination of transition paths, which drives transition between two known conformations keeping all moment geometries that do not violate the chemical structure and that in general can be explained from the pattern of easiest deformations of the reference structures. In the lack of experimental dynamic data, there is, however, no guarantee on

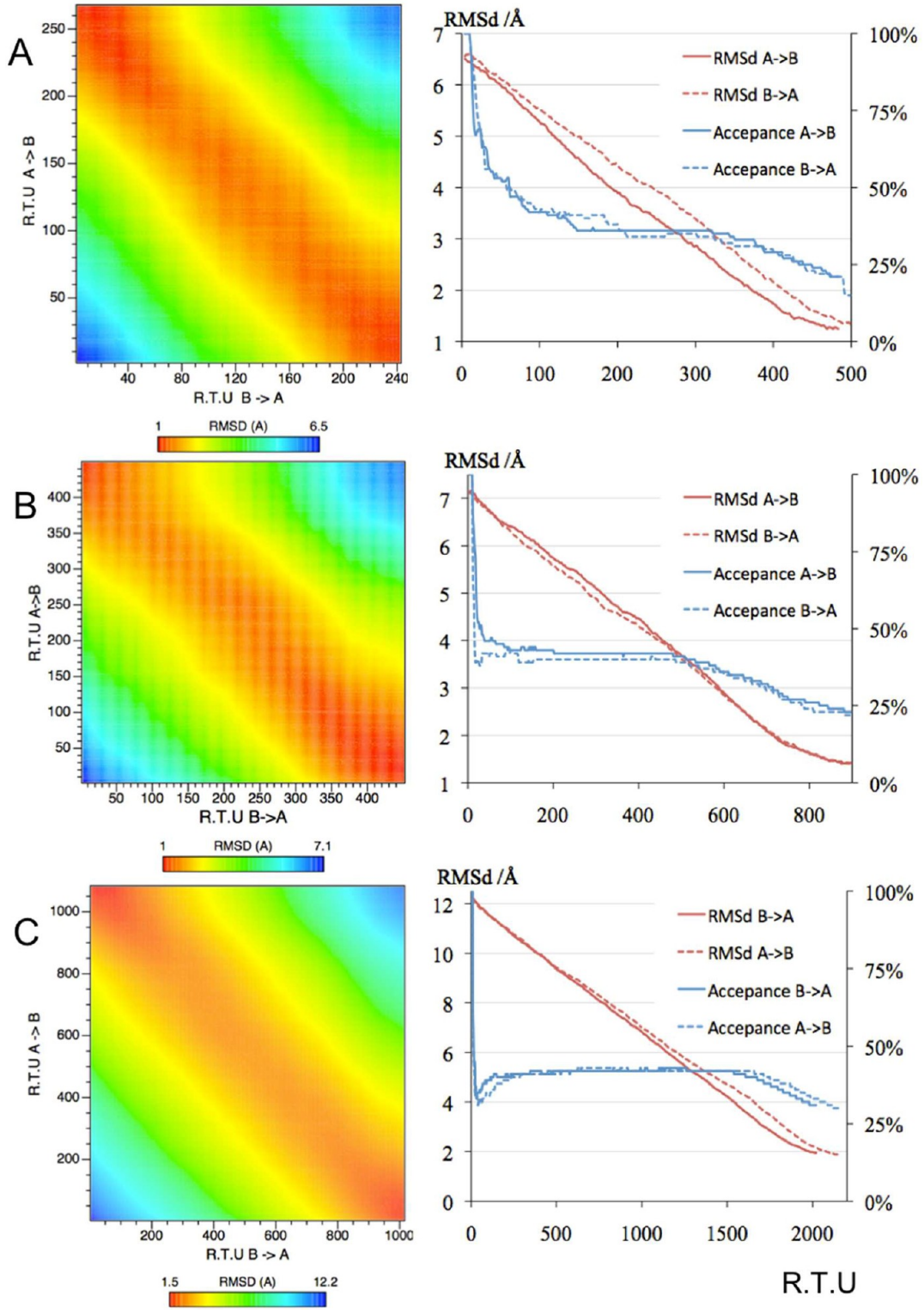

Figure 5. Examples (1l5e ←→ 1l5b, top; 1ake←→4ake, middle; and 1sx4←→1oel, bottom) of transitions obtained by the MDdMD procedure. On the left panel, bidimensional RMSD plots are presented to show the reversibility of the transitions (nearly diagonal distributions). On the right-hand side, Maxwell−Demon acceptance and RMSD profiles along transitions are presented to confirm the reversibility (noted as near-superposition of the curves).

the goodness of the proposed path. Fortunately, for a few cases, PDB reports intermediate structures, i.e., conformations of the protein that are expected to be in the path of transition between two more distant protein conformers (see Figure 6). It must be stressed that there is no direct evidence that the experimental intermediate is a real intermediate for the transition, but it is a reasonable assumption that a good transition path should approach at some point the putative intermediate structure. Analysis of the corresponding transition (Figure 8, Figure S2) illustrates how our transition trajectory typically passes close to the suggested experimental intermediate sampling in all of the case structures, with clear folded-like properties (see Figure 6)

dx.doi.org/10.1021/ct300494q | J. Chem. Theory Comput. 2012, 8, 4707−4718

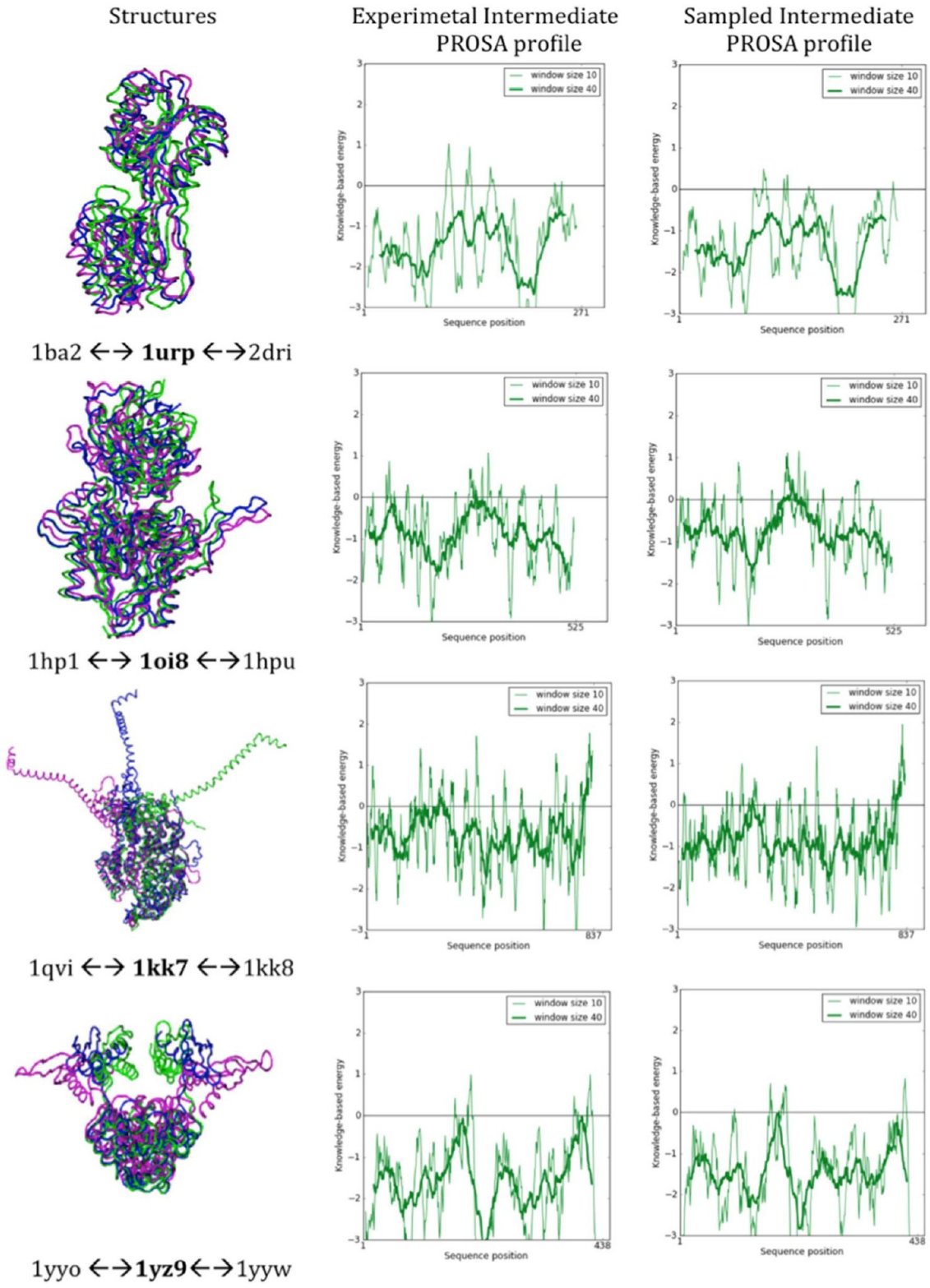

Figure 6. Study of the similarity between experimental and MDdMD transition intermediates. Left: detail of the transitions (with the intermediate displayed in blue) considered here. Right: PROSA profiles for the experimental and MDdMD sampled intermediate (note that no bias was introduced in the simulations to approach MDdMD sampling to experimental intermediate).

and non-negligible RMSDs between the sampled structure and the assumed "intermediate" crystal structure, which are in some cases clearly linked to mobile movements.

Transition Perturbation. The procedure outlined here is very flexible, allowing the introduction of any external effect into the calculation. This allows us to determine, for example, how a given transition might be facilitated or disturbed by the presence of external fields or molecules. Figure 9 illustrates the power of the method to characterize the disturbing effect of the ligand on the close→open transition of adenylate kinase (1ake→4ake), which

dx.doi.org/10.1021/ct300494q | J. Chem. Theory Comput. 2012, 8, 4707−4718

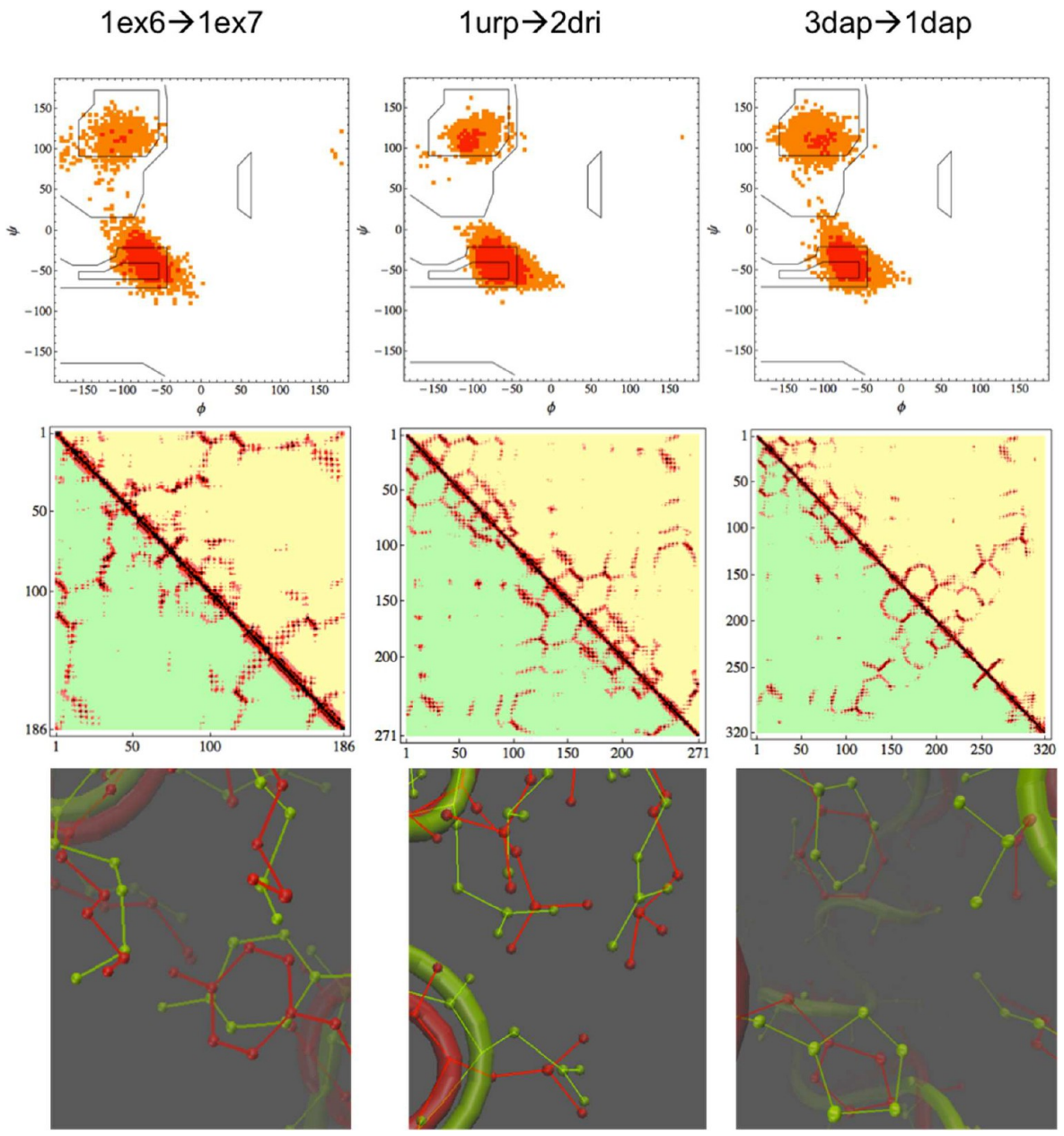

Figure 7. Detail of the local quality of sampled structures for three random cases. Top: Ramachandran maps for all sampled structures (red show higher populated states). Middle: C_β based contact maps for the sampled (above diagonal) and experimental (below diagonal) end state. Bottom: Some structural details of the side chains in the experimental vs final sampled structures.

is characterized by a dramatic decrease in the velocity of the algorithm to advance toward the target structure and the much larger value of RMSD to the target value obtained in the simulations.

MDdMD Application. MDdMD can be run through a Web-based interface (http://mmb.irbbarcelona.org/MDdMD). The user can upload both input and target structures or fetch them from the PDB. Since only Cα from the target structure is used, there is no need for both proteins to be coincident, so mutated structures or close homologues could also be simulated. A reduced set of parameters could also be defined: simulation temperature, acceptance ratio, and the desired final RMSD. Alternatively, most relevant trajectories for the already available simulations (see Table S2) could be examined. After an initial check for the consistency of parameters and structures, simulations are launched. MDdMD simulations are executed on our Web-applications cluster, under an SGE batch queue engine. The simulation can be followed interactively on an intermediate Web page that includes information about the current stage of the transition: an RMSD/acceptance rate plot showing the current acceptance rate and the RMSD between the simulated and the target structure; the superposition between initial, current, and target structures can be visualized through a JMol applet (http://jmol.sourceforge.net). From the intermediate screen, the simulation can be stopped by the user at any time, obtaining the accumulated trajectory as a final result. This helps to avoid nonproductive simulations, or finish simulations that have already reached their target. In any case, the user is informed through an email message of the completion of the simulation, and a link to the final results is provided.

The resulting conformational transition trajectory can be visualized in JMol or downloaded, either in PDB or Amber's

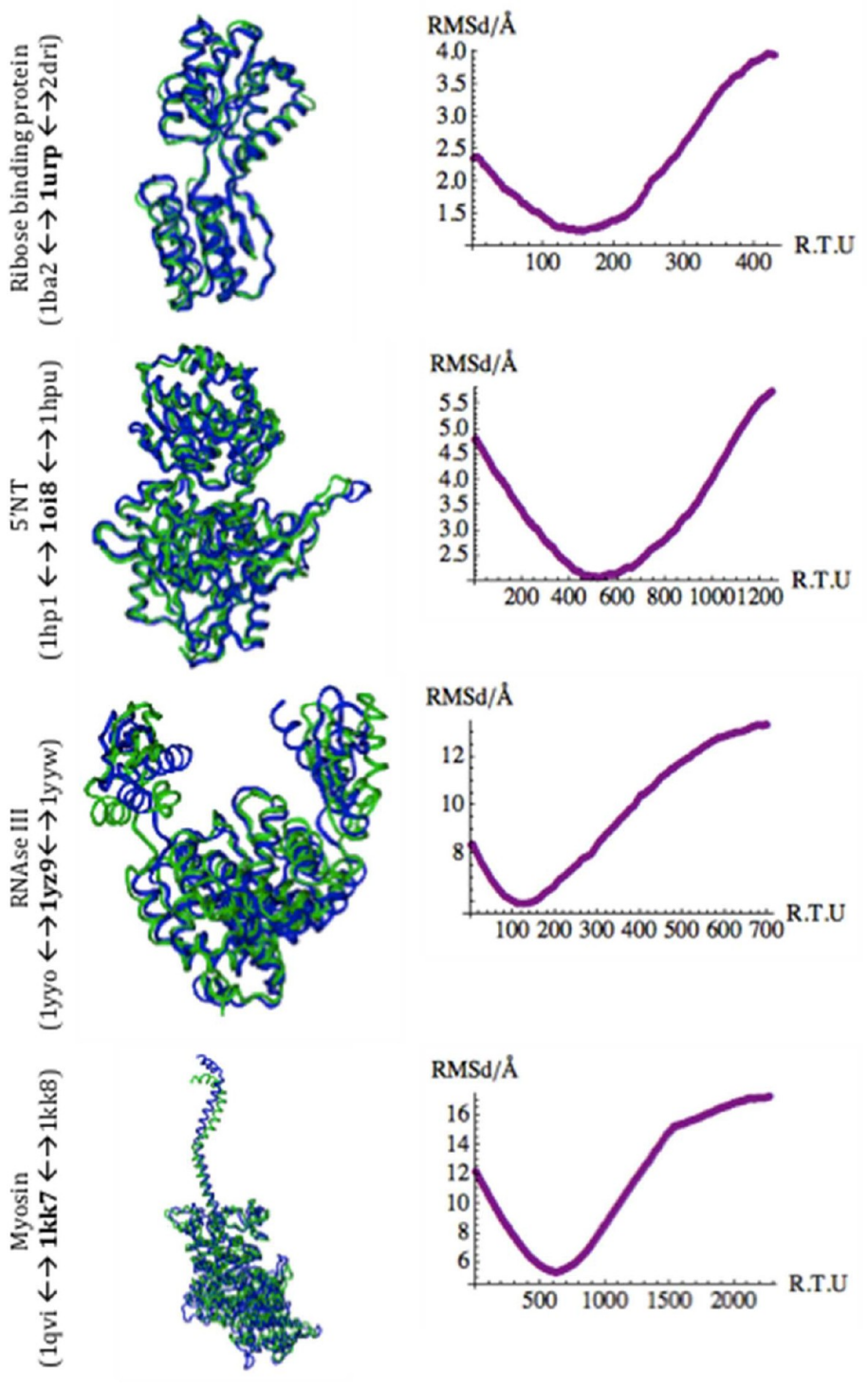

Figure 8. Evolution of the RMSD (to the experimental intermediate) along the simulation (green) together with the experimental intermediate (blue).

ASCII CRD formats, that make them compatible with most MD analysis software.

■ CONCLUSIONS

We present an extension of the discrete molecular dynamics algorithm to trace conformational transitions. The method can work at any level of resolution (including the all-heavy-atoms one explored here) and drives physically meaningful transitions toward the target structure using a Maxwell−Demon procedure enriched by introducing information on the essential deformation pattern of proteins. The method is extremely flexible, allowing the introduction of any desired experimental constraints or the simulation of perturbing elements in the transition. Testing of the method in an extended set of transitions (nearly 100) reveals a success rate around 94%, including some very difficult transitions, involving large movements that do not align with the essential deformability pattern of proteins. The intrinsic speed of dMD makes the technique very efficient computationally and competitive with less physical approaches.

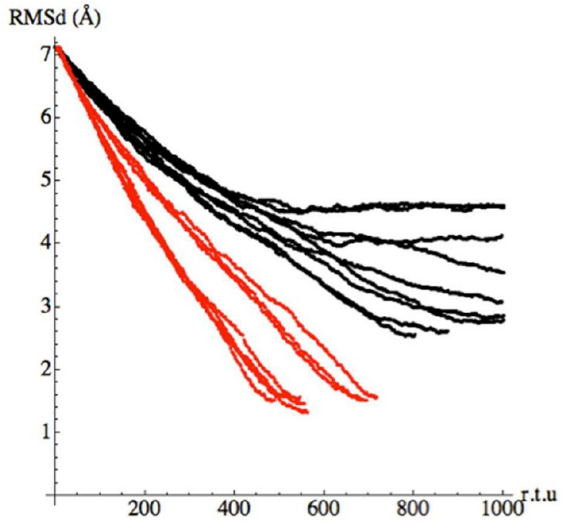

Figure 9. RMSD (to target structure) evolution along simulation time in the 1ake→4ake transition in the absence (red lines) and presence (black lines) of a ligand in the binding site of bis(adenosine)5′-pentaphosphate. The disturbing effect of the ligand is clearly visible in the failure of black lines to reach the target structure. Multiple lines correspond to individual trajectories performed to verify the robustness of the results.

■ ASSOCIATED CONTENT

Supporting Information

Detailed data of studied conformational transitions. The material is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: modesto.orozco@irbbarcelona.org.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science and Innovation (BIO2009-10964 and Consolider E-Science), European Research Council (ERC Advanced Grant), Scalalife European Project, and the Fundación Marcelino Botín. P.S is an IRB-la Caixa predoctoral fellow. M.O. is an ICREA Academia investigator.

■ REFERENCES

(1) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hübner, C. G.; Kern, D. *Nature* **2007**, *450*, 838−844.
(2) Velazquez-Muriel, J. A.; Rueda, M.; Cuesta, I.; Pascual-Montano, A.; Orozco, M.; Carazo, J.-M. *BMC Struct. Biol.* **2009**, *9*, 6.
(3) Bakan, A.; Bahar, I. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14349−14354.
(4) Yang, L.; Song, G.; Jernigan, R. L. *Biophys. J.* **2007**, *93*, 920−929.
(5) Bahar, I.; Chennubhotla, C.; Tobi, D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633−640.
(6) Tobi, D.; Bahar, I. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 18908−18913.
(7) Eyal, E.; Dutta, A.; Bahar, I. *WIREs Comput. Mol. Sci.* **2011**, *1*, 426−439.
(8) Dobbins, S. E.; Lesk, V. I.; Sternberg, M. J. E. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10390−10395.
(9) Gerstein, M.; Krebs, W. *Nucleic Acids Res.* **1998**, *26*, 4280−4290.
(10) Falke, J. J. *Science* **2002**, *295*, 1480−1481.
(11) Leo-Macias, A.; Lopez-Romero, P.; Lupyan, D.; Zerbino, D.; Ortíz, A. R. *Biophys. J.* **2005**, *88*, 1291−1299.
(12) Stein, A.; Rueda, M.; Panjkovich, A.; Orozco, M.; Aloy, P. *Structure* **2011**, *19*, 881−889.
(13) Orellana, L.; Rueda, M.; Ferrer-Costa, C.; Lopez-Blanco, J. R.; Chacón, P.; Orozco, M. *J. Chem. Theory Comput.* **2010**, *6*, 2910−2923.
(14) Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128−132.
(15) Ban, D.; Funk, M.; Gulich, R.; Egger, D.; Sabo, T. M.; Walter, K. F. A.; Fenwick, R. B.; Giller, K.; Pichierri, F.; de Groot, B. L.; Lange, O. F.; Grubmüller, H.; Salvatella, X.; Wolf, M.; Loidl, A.; Kree, R.; Becker, S.; Lakomek, N.-A.; Lee, D.; Lunkenheimer, P.; Griesinger, C. *Angew. Chem., Int. Ed.* **2011**, *50*, 11437−11440.
(16) Fenwick, R. B.; Esteban-Martin, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. *J. Am. Chem. Soc.* **2011**, *133*, 10336−10339.
(17) Kubitzki, M. B.; de Groot, B. L. *Structure* **2008**, *16*, 1175−1182.
(18) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807−822.
(19) Shimamura, T.; Weyand, S.; Beckstein, O.; Rutherford, N. G.; Hadden, J. M.; Sharples, D.; Sansom, M. S. P.; Iwata, S.; Henderson, P. J. F.; Cameron, A. D. *Science* **2010**, *328*, 470−473.
(20) Paci, E.; Lindorff-Larsen, K.; Dobson, C. M.; Karplus, M.; Vendruscolo, M. *J. Mol. Biol.* **2005**, *352*, 495−500.
(21) Orozco, M.; Orellana, L.; Hospital, A.; Naganathan, A.; Emperador, A.; Carrillo, O.; Gelpí, J. *Advances in Protein Chemistry and Structural Biology*; Christov, C., Ed.; Burlington Academic Press: Burlington, MA, 2011; Vol. *85*, pp 183−215.
(22) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpí, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 796−801.
(23) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6679−6685.
(24) Amadei, A.; Linssen, A.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412−425.
(25) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.
(26) Juraszek, J.; Bolhuis, P. G. *Biophys. J.* **2008**, *95*, 4246−4257.
(27) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. *Adv. Chem. Phys.* **2002**, 1−78.
(28) Juraszek, J.; Vreede, J.; Bolhuis, P. G. *Chem. Phys.* **2012**, *396*, 30−44.
(29) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(30) Bolhuis, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12129−12134.
(31) Wales, D. J. *Int. Rev. Phys. Chem.* **2006**, *25*, 237−282.
(32) Khalili, M.; Wales, D. J. *J. Phys. Chem. B* **2008**, *112*, 2456−2465.
(33) Wales, D. J.; Bogdan, T. V. *J. Phys. Chem. B* **2006**, *110*, 20765−20776.
(34) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *119*, 9947−9955.
(35) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585−590.
(36) Brooks, C.; Karplus, M.; Pettitt, M. *Adv. Chem. Phys.* **1988**, *71*, 35−58.
(37) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148−154.
(38) Beckstein, O.; Denning, E. J.; Perilla, J. R.; Woolf, T. B. *J. Mol. Biol.* **2009**, *394*, 160−176.
(39) Perilla, J. R.; Beckstein, O.; Denning, E. J.; Woolf, T. B. *J. Comput. Chem.* **2010**, *32*, 196−209.
(40) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(41) Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826−843.
(42) Schlitter, J. *J. Mol. Graph.* **1995**, *12*, 84−89.
(43) Krüger, P.; Verheyden, S.; Declerck, P. J.; Engelborghs, Y. *Protein Sci.* **2001**, *10*, 798−808.
(44) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(45) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Science* **2002**, *296*, 1832−1835.

(46) Bahar, I.; Rader, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.
(47) Derreumaux, P.; Mousseau, N. *J. Chem. Phys.* **2007**, *126*, 025101.
(48) Tirion, M. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.
(49) Devane, R.; Shinoda, W.; Moore, P. B.; Klein, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 2115−2124.
(50) Kim, M. K.; Chirikjian, G. S.; Jernigan, R. L. *J. Mol. Graphics Modell.* **2002**, *21*, 151−160.
(51) Kim, M. K.; Jernigan, R. L.; Chirikjian, G. S. *Biophys. J.* **2002**, *83*, 1620−1630.
(52) Mendez, R.; Bastolla, U. *Phys. Rev. Lett.* **2010**, *104*, 228103.
(53) Lopez-Blanco, J. R.; Garzón, J. I.; Chacón, P. *Bioinformatics* **2011**, *27*, 2843−2850.
(54) Rueda, M.; Chacón, P.; Orozco, M. *Structure* **2007**, *15*, 565−575.
(55) Ding, F.; Dokholyan, N. V. *Trends Biotechnol.* **2005**, *23*, 450−455.
(56) Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Biophys. J.* **2002**, *83*, 3525−3532.
(57) Shirvanyants, D.; Ding, F.; Tsao, D.; Ramachandran, S.; Dokholyan, N. V. *J. Phys. Chem. B* **2012**, *116*, 8375−8382.
(58) Emperador, A.; Meyer, T.; Orozco, M. *J. Chem. Theory Comput.* **2008**, *4*, 2001−2010.
(59) Emperador, A.; Meyer, T.; Orozco, M. *Proteins: Struct., Funct., Bionf.* **2010**, *78*, 83−94.
(60) Emperador, A.; Carrillo, O.; Rueda, M.; Orozco, M. *Biophys. J.* **2008**, *95*, 2127−2138.
(61) Ding, F.; Buldyrev, S.; Dokholyan, N. *Biophys. J.* **2005**, *88*, 147−155.
(62) Zhou, Y.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 14429−14432.
(63) Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. *RNA* **2008**, *14*, 1164−1173.
(64) Krebs, W. G.; Gerstein, M. B. *Nucleic Acids Res.* **2000**, *28*, 1665−1675.
(65) Ye, Y.; Godzik, A. *Nucleic Acids Res.* **2004**, *32*, W582−W585.
(66) Flores, S.; Echols, N.; Milburn, D.; Hespenheide, B.; Keating, K.; Lu, J.; Wells, S.; Yu, E. Z.; Thorpe, M.; Gerstein, M. *Nucleic Acids Res.* **2006**, *34*, D296−D301.
(67) Lindahl, E.; Azuara, C.; Koehl, P.; Delarue, M. *Nucleic Acids Res.* **2006**, *34*, W52−W56.
(68) Franklin, J.; Koehl, P.; Doniach, S.; Delarue, M. *Nucleic Acids Res.* **2007**, *35*, W477−W482.
(69) Weiss, D. R.; Levitt, M. *J. Mol. Biol.* **2009**, *385*, 665−674.
(70) Yang, Z.; Májek, P.; Bahar, I. *PLoS Comput. Biol.* **2009**, *5*, e1000360.
(71) Lezon, T. R.; Sali, A.; Bahar, I. *PLoS Comput. Biol.* **2009**, *5*, e1000496.
(72) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. *Annu. Rev. Biophys.* **2010**, *39*, 23−42.
(73) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. *Chem. Rev.* **2010**, *110*, 1463−1497.
(74) Bryngelson, J.; Onuchic, J.; Socci, N.; Wolynes, P. *Proteins: Struct., Funct., Bionf.* **1995**, *21*, 167−195.
(75) Proctor, E. A.; Ding, F.; Dokholyan, N. *WIREs Comput. Mol. Sci.* **2011**, *1*, 80−92.
(76) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010−1018.
(77) Urbanc, B.; Borreguero, J.; Cruz, L.; Stanley, H. *Methods Enzymol.* **2006**, *412*, 314−338.
(78) Smith, S.; Hall, C.; Freeman, B. *J. Comput. Phys.* **1997**, *134*, 16−30.
(79) Taketomi, H.; Ueda, Y.; Gō, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 445−459.
(80) Camps, J.; Carrillo, O.; Emperador, A.; Orellana, L.; Hospital, A.; Rueda, M.; Cicin-Sain, D.; D'Abramo, M.; Gelpí, J. L.; Orozco, M. *Bioinformatics* **2009**, *25*, 1709−1710.
(81) Nguyen, H. D.; Hall, C. K. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16180−16185.
(82) Gherghe, C. M.; Leonard, C. W.; Ding, F.; Dokholyan, N. V.; Weeks, K. M. *J. Am. Chem. Soc.* **2009**, *131*, 2541−2546.
(83) Hajdin, C. E.; Ding, F.; Dokholyan, N. V.; Weeks, K. M. *RNA* **2010**, *16*, 1340−1349.
(84) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Folding Des.* **1998**, *3*, 577−587.
(85) Ding, F.; Lavender, C. A.; Weeks, K. M.; Dokholyan, N. V. *Nat. Methods* **2012**, 603−608.
(86) Peng, S.; Ding, F.; Urbanc, B.; Buldyrev, S. V.; Cruz, L.; Stanley, H. E.; Dokholyan, N. V. *Phys. Rev. E* **2004**, *69*, 041908.
(87) Urbanc, B.; Betnel, M.; Cruz, L.; Bitan, G.; Teplow, D. *J. Am. Chem. Soc.* **2010**, *132*, 4266−4280.
(88) Urbanc, B.; Cruz, L.; Yun, S.; Buldyrev, S. V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17345−17350.
(89) Ding, F.; Borreguero, J.; Buldyrev, S. V.; Stanley, H.; Dokholyan, N. *Proteins: Struct., Funct., Bionf.* **2003**, *53*, 220−228.
(90) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Bionf.* **1999**, *35*, 133−152.
(91) Meyer, T.; D'Abramo, M.; Hospital, A.; Rueda, M.; Ferrer-Costa, C.; Pérez, A.; Carrillo, O.; Camps, J.; Fenollosa, C.; Repchevsky, D.; Gelpí, J. L.; Orozco, M. *Structure* **2010**, *18*, 1399−1409.
(92) Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2004**, *87*, 800−811.
(93) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(94) Wiederstein, M.; Sippl, M. J. *Nucleic Acids Res.* **2007**, *35*, W407−W410.