

Exploration of conformational transition pathways from coarse-grained simulations

Pedro Sfriso^{1,2}, Adam Hospital^{1,2,3}, Agustí Emperador^{1,2} and Modesto Orozco^{1,2,3,4,*}¹Institute for Research in Biomedicine (IRB Barcelona), ²Joint IRB-BSC Program in Computational Biology, Baldiri Reixac 10, Barcelona 08028, Spain, ³National Institute of Bioinformatics and ⁴Department of Biochemistry and Molecular Biology, University of Barcelona, Av. Diagonal 647, Barcelona 08028, Spain

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: A new algorithm to trace conformational transitions in proteins is presented. The method uses discrete molecular dynamics as engine to sample protein conformational space. A multiple minima Go-like potential energy function is used in combination with several enhancing sampling strategies, such as metadynamics, Maxwell Demon molecular dynamics and essential dynamics. The method, which shows an unprecedented computational efficiency, is able to trace a wide range of known experimental transitions. Contrary to simpler methods our strategy does not introduce distortions in the chemical structure of the protein and is able to reproduce well complex non-linear conformational transitions. The method, called GOdMD, can easily introduce additional restraints to the transition (presence of ligand, known intermediate, known maintained contacts,...) and is freely distributed to the community through the Spanish National Bioinformatics Institute (<http://mmb.irbbarcelona.org/GOdMD>).

Availability: Freely available on the web at <http://mmb.irbbarcelona.org/GOdMD>.

Contact: modesto.orozco@irbbarcelona.org or modesto@mmb.pcb.ub.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on February 26, 2013; revised on April 29, 2013; accepted on May 30, 2013

1 INTRODUCTION

Many biological functions of proteins such as mechanic work, signal transduction or enzymatic activity are modulated by a key property of them: flexibility (Henzler-Wildman *et al.*, 2007; Micheletti, 2013; Velazquez-Muriel *et al.*, 2009). Flexibility is a property that has been refined and maintained by evolution (Falke, 2002; Micheletti, 2013) and that, in turn has been also exploited by evolution to generate new proteins in a conservative mechanism, which guarantees the maintenance of the structural scaffold as well as the relevant deformation pattern (Leo-Macias *et al.*, 2005; Stein *et al.*, 2011). Structural databases show (Gerstein and Krebs, 1998) increasing number of proteins having alternative structures depending on external factors (such as crystallization conditions, posttransductional chemical modifications, presence of ligands, changes in solvent

environment, etc). This probes the existence of dramatic conformational transitions in proteins, but giving no information on how such transitions happen.

Recent refinements of experimental techniques have provided direct evidence on the mechanisms of conformational transitions for some model proteins (Ban *et al.*, 2011; Eisenmesser *et al.*, 2002; Fenwick *et al.*, 2011; Kern and Zuiderweg, 2003; Lindorff-Larsen *et al.*, 2005). However, we are still far from the point that all conformational transitions could be described by means of experimental methods. This situation forces the use of simulation techniques, which has been largely refined in the past years (Best *et al.*, 2005; Bolhuis *et al.*, 2002; Karplus and Kuriyan, 2005; Kubitzki and de Groot, 2008; Maragakis and Karplus, 2005; Miyashita, 2003; Okazaki *et al.*, 2006), providing information of increasing quality for a large number of conformational transitions in proteins (Sfriso *et al.*, 2012; Stein *et al.*, 2011).

Molecular dynamics (MD), using atomistic force-fields and explicit representation of solvent (McCammon *et al.*, 1977), is probably the most accurate theoretical technique for reproducing protein flexibility. Recent computational approaches have made possible the simulation of thousands of proteins (Meyer *et al.*, 2010) and the derivation of up to millisecond trajectories for proteins (Dror *et al.*, 2012; Lindorff-Larsen *et al.*, 2011). Unfortunately, most conformational transitions are still far from the capabilities of plain atomistic MD, forcing the use of biasing schemes (Beckstein *et al.*, 2009; Das *et al.*, 2006; Laio and Parrinello, 2002; Leone *et al.*, 2010; Liphardt *et al.*, 2002; Schlitter, 1994), designed to maximize the sampling along a given variable that is believed to capture conformational transition motions. Biased-MD protocols are extremely powerful, but they are expensive computationally, require expertise from the user and can lead to incorrect results when the transition coordinate is not well defined.

Coarse-grained (CG) models (Bahar and Rader, 2005; Dobbins *et al.*, 2008; Lopez-Blanco *et al.*, 2011; Mendez and Bastolla, 2010; Orozco *et al.*, 2011; Tozzini, 2005; Whitford *et al.*, 2007) are inexpensive, but still accurate alternatives to atomistic MD simulations, which have gained significant popularity in recent years. Unfortunately, the use of CG models require the assumption of a certain loss of detail in the simulation, for example, explicit solvent is ignored, which prevent the study of specific water-protein interactions and side chains are either ignored or dramatically simplified (Mendez and Bastolla, 2010; Marrink *et al.*, 2007), which raises problems

*To whom correspondence should be addressed.

to reproduce ligand-target interactions. CG models have been calibrated against experimental data, such as B-factors, structural variability in databases or atomistic MD, and despite their simplicity, provide often results of surprising quality with limited computer resources and reduced expertise from the user.

CG-methods have been largely used to trace conformational transitions in proteins. The first implementations followed mainly interpolation schemes between original and final conformations (Delarue and Sanejouand, 2002; Flores *et al.*, 2006; Franklin *et al.*, 2007; Kim *et al.*, 2002; Krebs and Gerstein, 2000; Lindahl *et al.*, 2006; Weiss and Levitt, 2009). Interpolation protocols are fast and guarantee the completion of the transition, but often at the expense of unrealistic intermediate conformations, which are not good starting points for refinement through more accurate atomistic simulations. Alternative CG-transition methods have been developed under the assumption that biologically relevant transitions should follow the easiest deformation movements of the proteins (Bahar *et al.*, 2010; Lezon *et al.*, 2009; Mendez and Bastolla, 2010; Yang *et al.*, 2009) defined as the softest deformation modes obtained by diagonalization of the Hessian matrix derived from an elastic network model (ENM) Hamiltonian:

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

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 K_{ij} in the reference structure.

Morphing methods based on the use of essential deformation modes provide more reasonable approaches than linear interpolation schemes, but are far from being optimal since (i) the covalent structure of the protein can be damaged when large displacements along a limited number of Cartesian eigenvectors are made, (ii) a non-negligible number of transitions do not align with the intrinsic deformation pattern of proteins (Stein *et al.*, 2011), and even those cases where the alignment between the essential deformation space and the transition vector is good, displacement along the essential space of one of the proteins never allows a full transition between the two conformational states. Recent advances in the field, such as the use of multiple reference structures to define a transition-dependent essential deformation space (Sfriso *et al.*, 2012; Yang *et al.*, 2009) or the derivation of the essential deformation modes in the dihedral space (Lopez-Blanco *et al.*, 2011; Mendez and Bastolla, 2010), have alleviated but not solved these problems.

In this article, we present a new approach to obtain ultra-fast, but accurate, conformational transition pathways in proteins out of physics-based molecular mechanics simulations. By combining an efficient sampling technique [discrete molecular dynamics (dMD; Orozco *et al.*, 2011; Proctor *et al.*, 2011)] with a novel multiple-well Go-like scheme. Sampling is enhanced with biasing techniques such as metadynamics (Laio and Parrinello, 2002) and Maxwell Demon MD (Rueda *et al.*, 2004; Sfriso *et al.*, 2012). The method was tested in a large battery of transitions (near 50 pairs of structures) obtaining, with reduced

computational cost, reasonable pathways in all cases, including in cases of extreme difficulty, where transition requires, for example, partial unfolding of the protein. Owing to the physical nature of the method, sampled intermediate structures maintain covalent structure, and no steric clashes are allowed. In cases where experimental intermediates are characterized, the method finds transition paths passing close to them. Surprisingly, the method outperforms our previous more detailed algorithm, Maxwell Demon discrete molecular dynamics procedure (MDdMD; Sfriso *et al.*, 2012), and is also competitive with respect to alternative methods in the literature (Supplementary Figs S1–S3). Our novel approach is flexible, allowing the introduction of any perturbation. Furthermore, the user can easily bias trajectories to guarantee that known experimental intermediates are sampled.

2 METHODS

In this approach we have applied the dMD method [see (Emperador *et al.*, 2008b; Proctor *et al.*, 2011) for details]. In dMD, particles move in the ballistic regime, with constant velocity until a collision occurs. Collisions occur at the particle–particle distances corresponding to a discontinuity in the dMD potential. The velocities of the particles after the collision are obtained using the rules of conservation of momentum and energy. This allows avoiding the integration of Newton's equations of motion, speeding up the simulations as compared with usual MD. Discrete MD has successfully been applied to protein folding (Ding *et al.*, 2005; Zhou and Karplus, 1997), macromolecular dynamics (Emperador *et al.*, 2008a, 2010), RNA structural predictions (Ding *et al.*, 2008, 2012; Gherghe *et al.*, 2009) and protein aggregation (Ding and Dokholyan, 2008; Urbanc *et al.*, 2004, 2010). Recently dMD has been also used to robust energy minimization of protein–protein complexes (Emperador *et al.*, 2013) and protein–ligand interactions (Proctor *et al.*, 2012).

2.1 Force-field representation

We adopted a CG representation of the protein, where only C α s were explicitly represented. We used a single square well with infinite walls to define the bonded interactions between consecutive beads. The non-bonded interactions were described by a multiple well Go-like model (Taketomi *et al.*, 1975; Ueda *et al.*, 1978) depending on the experimental inter-particle distance. Non-bonded terms of the Hamiltonian were classified in two categories based on whether or not the inter-particle distance is different between the initial (A) and final (B) conformations. If they are, we define a double well potential (Fig. 1) centered in the respective experimental values: \bar{r}_A and \bar{r}_B . When the distances are similar in both conformations (i.e. $\bar{r}_A \sim \bar{r}_B$), we used a single, but wider, well (50% larger) centered at $\frac{1}{2}(\bar{r}_A + \bar{r}_B)$. To avoid over-restraining the system with physically irrelevant interactions, the potential energy was defined only for those particles within a cutoff ($\bar{r}_A, \bar{r}_B > \text{radii of gyration}/2$). It is worth noting that the height of the wells for all non-bonded interactions is finite (0.5 kcal/mol), allowing the particles to escape from the wells if required.

This Go-like scheme is convenient to complete transition paths [$>85\%$ of the root mean square deviation (RMSd) difference between end structures) because target minimum acts as attractor. This situation enables us to use softer biasing schemes and to better recover native contacts (see Supplementary Materials for details).

2.2 Accelerating the transition

Within the Hamiltonian definition above, the A→B transition occurs through a movement of particles jumping from wells centered at \bar{r}_A to those centered at \bar{r}_B . Unfortunately, as a side effect of Go-like potentials,

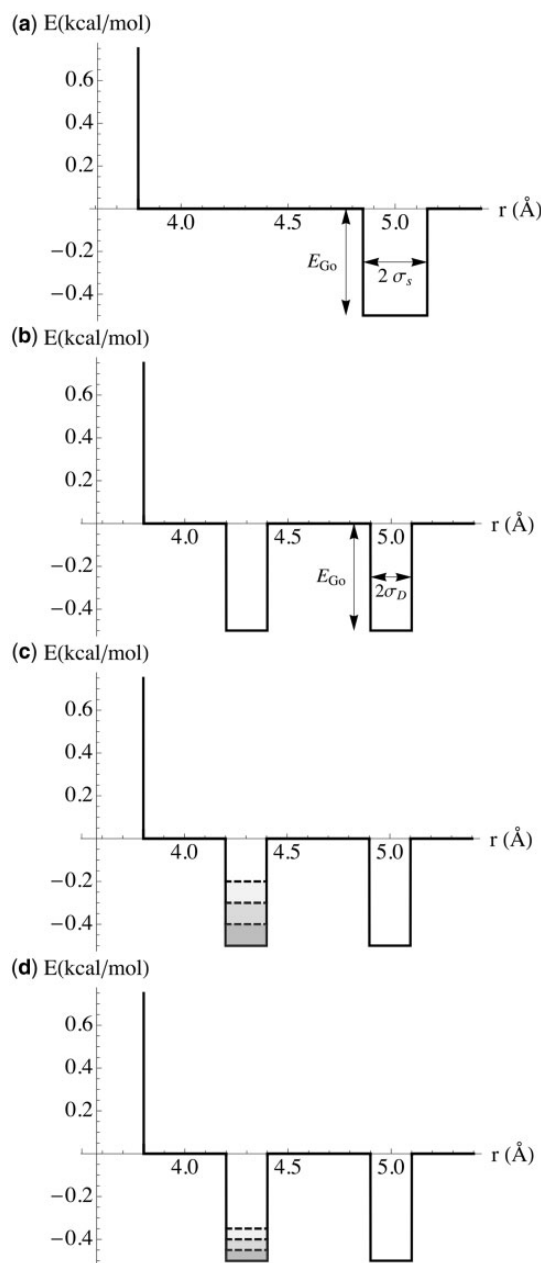


Fig. 1. Interaction potentials used in this study. (a) single well corresponding to a particle–particle non-bonded pair that does not change their distance during transition ($E_{GO} = 0.5$ kcal/mol, $\sigma_s = 0.15$ Å). (b) Double well representing two states of particle radial distance that corresponds to two known structures ($E_{GO} = 0.5$ kcal/mol, $\sigma_D = 0.10$ Å). (c) Example of how discrete metadynamics increases the potential energy of a pair of particles whose movement overlaps with the essential deformation space of the initial state and (d) the same situation but when the overlap is much lower

spontaneous hopping from one minimum to another is difficult, making unlikely to find a spontaneous transition. We decided then to bias the trajectory to guarantee enrichment of sampling along the transition path. For this purpose, we implemented here a complex biasing scheme designed to favor the transition, but avoiding the use of restrains that

would guide the trajectory along arbitrary (and probably unrealistic paths). The first level of biasing is designed to move the protein far from the starting conformation, while the second level is designed to enrich biasing to approach the protein to the final conformation. For the interested reader, a flowchart of the algorithm is presented in Supplementary Figure S4.

2.2.1 Escaping from the initial minima We have implemented a discrete version of the metadynamics method (Laio and Parrinello, 2002) to guarantee that the system leaves the original minima. Metadynamics penalizes visits to the original well by raising gradually the energy of the well (i.e. by filling it; Fig. 1), which increases the chances of the system to escape from the attraction of the Go-potentials to the starting structure. As a consequence, as trajectory progresses sampling of the original conformation is less and less probable and system departs from the original conformation. Metadynamics is efficient as a method to escape from a minimum, but such a divergence happens in a random way, which can yield to unrealistic deformations when applied in an unsupervised fashion. To solve this problem, we have coupled metadynamics to an ENM (Emperador *et al.*, 2008a; Orellana *et al.*, 2010) in such a way that wells associated to inter-particle distances showing large changes along the first five essential deformation modes were filled faster than those associated to inter-particle distances that are not coupled to the essential deformation modes of the protein (Fig. 1). This strategy guarantees the exploration of alternative ways to escape from the original structure minima, while increasing the possibility to sample preferred pathways as defined by the essential deformation modes (first five are considered by default, but results are largely invariant in the range 3–10 modes (data not shown). Note that the ENM analysis is done only with the initial structure irrespectively of which the target structure is, avoiding the definition of too linear pathways. Note also that the full dMD simulations are done, which means that all the degrees of freedom (and not only the five preferred ones) are sampled.

2.2.2 Moving toward the final conformation The ENM-metadynamics procedure outlined above guarantees that the protein moves apart from the original structure sampling preferentially essential deformation modes. However, there is no guarantee that such movements will approach the protein to the final conformation. To guide the trajectory toward the final structure, without introducing arbitrary energy restraints, we have implemented a Maxwell Demon biasing algorithm (Perilla *et al.*, 2010; Rueda *et al.*, 2004; Sfriso *et al.*, 2012) with a control magnitude (Γ) defined as follows:

$$\Gamma = \sum_{i=1}^N \omega(i) \|r_{i,B} - r_{i,X}\| \quad (2)$$

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 an optimized weight function dependent on the inter-particle distance and the size of protein (Supplementary Fig. S5; results are robust to ω values in the range 15–25 Å are used).

Following the MDdMD procedure the bias toward the target structure is not introduced by an energy penalty, but using a less interfering informational criteria (Perilla *et al.*, 2010). The scheme is simple and efficient, after a certain number of dMD simulation steps (time t) the value of the progress variable (Γ ; equation 2) is compared with that obtained in a previous accepted movement ($t-\Delta t$). The simulation fragment $t-\Delta t \rightarrow t$ is then accepted or rejected based on a simple Metropolis test (equation 3):

$$p_t = \begin{cases} 1 & \text{if } \Gamma_t < \Gamma_{t-\Delta t} \\ \exp\left[-\left(\frac{\Gamma_t - \Gamma_{t-\Delta t}}{\beta \text{RMSd}_{(X;t,B)}}\right)^2\right] & \text{if } \Gamma_t > \Gamma_{t-\Delta t} \end{cases} \quad (3)$$

where p is the acceptance probability, $\text{RMSd}_{(X;t,B)}$ is RMSd between structure sampled at time $t(X)$ and target structure, β is dynamically

adjusted to guarantee an acceptance rate of 70%, and the time frame (Δt) is typically 100 time steps. These large acceptance rates, combined with a weight function with a maximum at ~ 15 Å enables the system to explore no so direct pathways that could lead to stressed structures and provides the best local geometries (Supplementary Figs S6 and S7). The presence of the Go-potentials of the target structure in the Hamiltonian avoids the need to increase β when trajectory approaches the final target conformation.

3 RESULTS

To evaluate the power of GOMD, we apply it to trace transitions between known equilibrium conformations of proteins. After analyzing protein data bank (PDB), we define a set of 94 transitions corresponding to 47 proteins showing two distinct conformations. This large benchmark set was extracted from a previous work in our group (Sfriso *et al.*, 2012) and spans a wide range of proteins from 100 to 1000 residues, showing different shapes and secondary structure composition. The database contains no trivial conformational changes and some of the conformational transitions represent dramatic geometrical alterations in the structure of the protein, including refolding in some cases. Many of the transitions are coupled to the binding to small ligands (60%) and/or macromolecules (39%), which increase the difficulties to trace reasonable conformational transition pathways. Further analysis of the data set reveals that a significant fraction of the transitions (44%) correspond to open/close ones, which can a priori generate hysteresis problems. Finally, 22% of the transitions do not align well [overlap (accumulated dot product) between top five normal modes and conformational transition is below 0.25] with the first five essential deformation modes detected from EN-normal mode analysis (NMA), i.e. that they will be very difficult to represent by simply activating movements along intrinsic deformation modes. Looking at different criteria (Sfriso *et al.*, 2012) we consider 62 of the 94 transitions as difficult or very difficult to follow, which means that we are validating our method with the most exigent transition dataset available.

GOMD has been able to find reasonable pathways for all studied transitions, even those requiring large refolding processes. The algorithm can provide multiple transition trajectories, defining a scenario of multiple pathways, which fits better into the transition funnel theory (Dill and Chan, 1997; Portella and Orozco, 2010). Furthermore, no violations of covalent distances, nor steric clashes or chemical meaningless conformations are sampled during transitions, which are always smooth, without the presence of apparent discontinuities or hysteresis effects (Supplementary Fig. S8). PROSA (Wiederstein and Sippl, 2007) calculations performed in random conformations sampled during the different transition yields to native-like profiles (Supplementary Figs S9 and S10), indicating that we are not sampling unrealistic conformations along the transition. The method is extremely efficient, >60% of trajectories are finished in <2 min wall-clock time in a laptop computer (2.4 GHz Intel Core 2 Duo) and even the most difficult transitions are finished in <30 min in the same computer. We are working in a parallel algorithm to explore larger systems with similar efficiency.

3.1 Sampling known intermediate conformations

There are a few cases in PDB of distinct conformations for a protein, where there is in addition to the start and end conformation a third structure, which is intermediate (at least in terms of RMSd) between the other two. Following Weiss and Levitt (2009) we can suggest that in general the third structure can be near a preferred passing point on the transition between the extreme conformations (i.e. we can consider an 'intermediate' in the transition). Thus, for the five cases where this putative intermediate is available, we determined transition paths between the two more diverse conformations. In all cases (Fig. 2) we found smooth and reasonable transitions, which go in the direction of the putative intermediate, even for the most difficult cases, where simpler methods have serious difficulties. In small systems (Ribose Binding Protein, 5'-NT), the transition passes through the putative intermediate. In larger systems, small RMSd deviations to the putative intermediate are much more difficult to obtain, partly because of noise introduced by the presence of

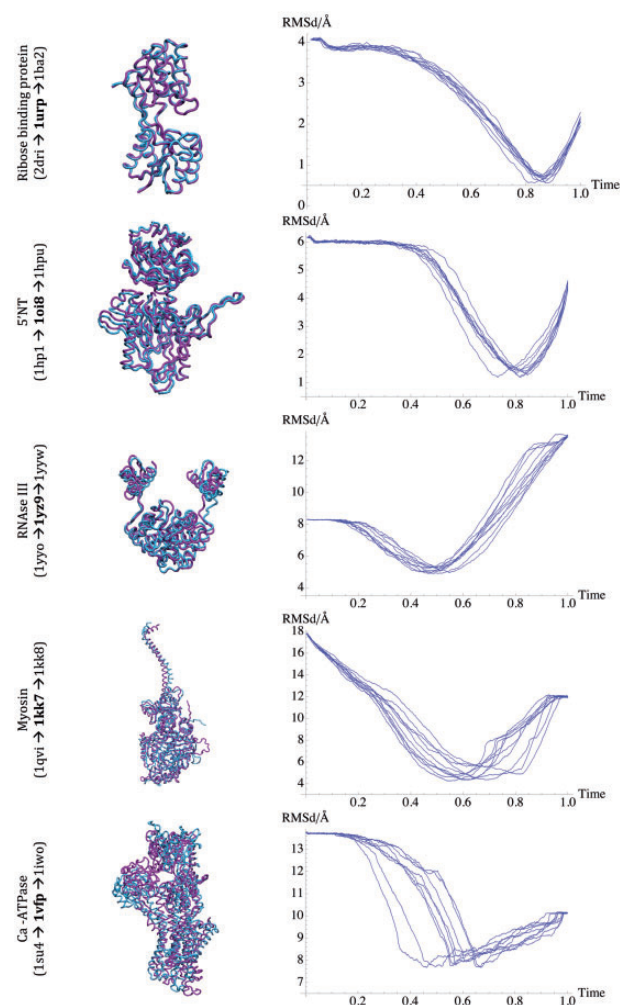


Fig. 2. Structural superposition of experimental intermediate structure and sampled conformations obtained along transitions. In the right-hand side column we display the RMSd profile (taking intermediate as reference) obtained in 10 independent transition pathways

long connecting loops, and partly because the intrinsic difficulty of tracing large conformational transitions. Although some large RMSd values are obtained (due mostly to stochastic loop movements), the potentials used in GOdMD capture better intermediate states than other state-of-the-art morphing methods (Supplementary Fig. S3). However, it is clear that the simulation drives spontaneously the transition toward conformational regions close to the putative intermediate. PROSA profiles and TMScore (Zhang and Skolnick, 2004) of both sampled and putative experimental intermediates are similar (Supplementary Figs S9 and S10 and Supplementary Table S2 for details), supporting the quality of the transition path.

3.2 Non-linearity of the transitions

One of the main caveats of biasing methods based on forcing a regular reduction of RMSd to the target structure (as most morphing methods) is that they force a linear transition path, that might be unrealistic (Fig. 3), and that does not provide information on the bottlenecks in the transition. Our method is able to detect non-linear transition pathways, and provides information on where are the bottlenecks, i.e. those points where the transition seems nearly stopped for long periods, and which correspond to regions of high rejection rate in the Maxwell Demon (Supplementary Fig. S11). Interestingly, the method is also able to disconnect local (measured as % of native contacts) to global transitions (measured by the RMSd to the target), showing how for some transitions local rearrangements happen first, while, on the contrary, for other transitions global

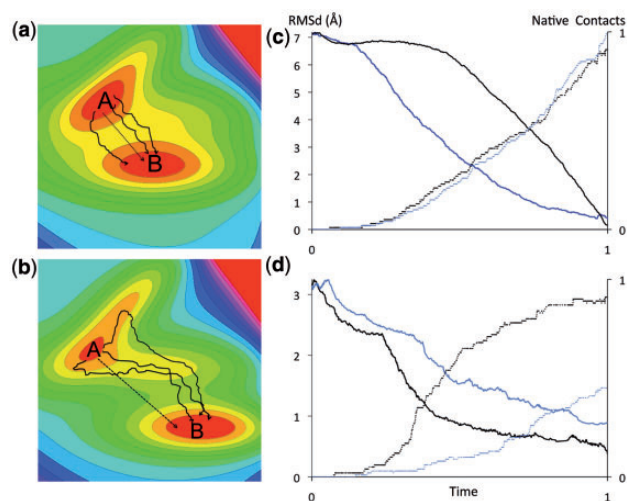


Fig. 3. (a) Energy surface where a linear conformational transition is expected to be realistic (x and y correspond to two general conformational variables, and color code means the energy order: red<yellow<green<blue). (b) Energy surface where transition does not follow a linear pathway. (c) An example of a real transition (4ake → lake) that behaves linearly, and where pathway obtained by our current procedure (black line) and a linear one (blue; obtained by forcing low acceptance rates and ignoring ENM analysis). (d) As in panel (c), but for a non-linear transition 1c9kB → 1cubB, where normal (black) and forced linear transitions (blue) are different, the latter being incorrect, because little target-native contacts are recovered (despite the low RMSd to target)

movement is before local conformational rearrangements (Figs 3 and 4).

It is worth noting that, the physical nature of the method guarantees that the guiding engine (NMA-bias metadynamics and a Maxwell Demon here) finds conformational pathways that never explore regions of unrealistically large energy. This guarantees that along all the transitions we sample protein-like conformations (Fig. 4), avoiding sampling always linear pathways that can lead to strong local distortions.

3.3 Introducing experimental information

One of the main advantages of our method is its flexibility, which makes very easy to introduce the effect of external perturbations in the transition, or to bias the trajectory by experimental information.

To show the capabilities of our algorithm, we analysed the effect of ligand binding in a transition, and also illustrate how the existence of experimentally validated intermediate can be used to bias the transition towards preferred pathways. For the first purpose we studied three additional systems for which ligand binding is known to displace conformational equilibrium: D-Allose binding protein (1gud/1rpj), L-Leucine Binding Protein (1usg/1usi) and Osmo-protection protein (1sw5/1sw2). After running 10 simulations (five in the absence and five in the presence of the ligand), we observed that those trajectories containing ligand were less efficient to reach the final state (Supplementary Figs S12 and S13 for details). These are just ‘prove of concept’ calculations, but provide encouraging evidence that our simple and fast algorithm can capture qualitatively the effect of the ligand in altering conformational pathways. To explore the possibility to bias GOdMD transitions by using experimental information on intermediate we have repeated the study of the very difficult ISU4 → IIWO transition (Weiss and Levitt, 2009) assuming that the intermediate structure IVFP corresponds to a necessary pass-point in the transition. So, we decoupled our sampling methods to find the transition pathways that better accommodate this experimental information.

With the same Hamiltonian defined by the end states (A, B; see Methodological Approach), the Maxwell Demon MD is now first referred to the experimental intermediate structure (IVFP) and then switched to the target structure (when sampling of intermediate structure is converged). With this protocol, we selected the path closer to the putative intermediate state I. Figure 5 shows the obtained results. A clear improvement on the sampling of I is observed, showing that while the potential

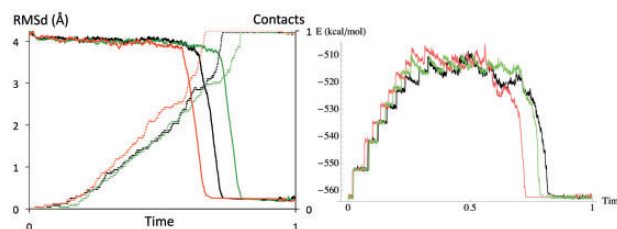


Fig. 4. Typical RMSd to target and potential energy profiles. Energy is presented in relative units. Different (three for simplicity) trajectories lead to similar but not identical results

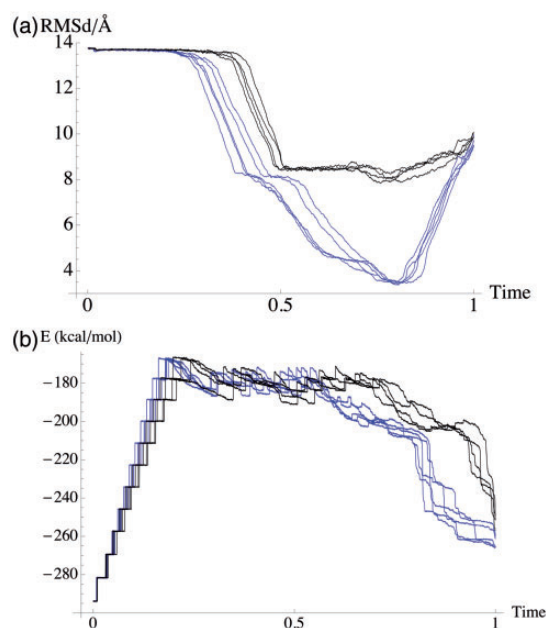


Fig. 5. ISU4 → IIWO transition including (blue) or not (black) information on the existence of an intermediate IVFP Transition is repeated five times to guarantee reproducibility. (a) Profiles of RMSd with respect to experimental intermediate structure. (b) Potential energy evolution along the trajectory

energy surface is defined considering only A and B, the transition derived is fully compatible with the presence of the putative intermediate I.

Note that the new bias introduced in the trajectory by the experimental restraint does not impact dramatically on the energy, and it even improves energy relaxation toward the final state (Fig. 5).

4 CONCLUSION

We present GODMD, a new method to trace conformational transition at the CG level. The method uses the fast CG dMD algorithm coupled to two biasing methods: (i) metadynamics adapted to follow proteins easiest deformation pattern, and (ii) a Metropolis-based Maxwell Demon algorithm. The method is fast and, contrary to many other approaches, always finds a smooth transition path when tested in a large dataset of transitions. Obtained trajectories maintain the covalent structure of the protein avoiding also unfavourable contacts, sampling intermediate structures with ‘protein-like’ properties, a clear advantage with respect to usual morphing schemes. The method is extremely flexible and can be easily adapted to introduce the effect of external perturbations in the transition or to bias the transition using experimental information. Limitations of the method are likely to be coupled to its CG nature that precludes the derivation of atomistic information on the transition. The approach is then expected to be especially powerful to obtain rough transition pathways that will be further refined by atomistic MD-based algorithms. A computer program

implementing our method is publicly available as a web server at mmb.irbbarcelona.org/GODMD.

Funding: Ministerio de Economía y Competitividad, Spain, the European Research Council (ERC-Advanced Grant (BIO2012-32868 to M.O.); the Instituto Nacional de Bioinformática (INB); the Consolider E-Science Project (to M.O.); Framework VII Scalalife Project (to M.O.); Fundación Marcelino Botín (to M.O.). P.S. is a ‘La Caixa fellow’ and M.O. is an ICREA Academia Researcher.

Conflict of Interest: none declared.

REFERENCES

- Bahar, I. and Rader, A.J. (2005) Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.*, **15**, 586–592.
- Bahar, I. *et al.* (2010) Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.*, **39**, 23–42.
- Ban, D. *et al.* (2011) Kinetics of conformational sampling in ubiquitin. *Angew. Chem. Int. Ed. Engl.*, **50**, 11437–11440.
- Beckstein, O. *et al.* (2009) Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open ↔ closed transitions. *J. Mol. Biol.*, **394**, 160–176.
- Best, R.B. *et al.* (2005) Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure*, **13**, 1755–1763.
- Bolhuis, P.G. *et al.* (2002) Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, **53**, 291–318.
- Das, P. *et al.* (2006) Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc. Natl. Acad. Sci. USA*, **103**, 9885–9890.
- Delarue, M. and Sanejouand, Y.H. (2002) Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J. Mol. Biol.*, **320**, 1011–1024.
- Dill, K.A. and Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.*, **4**, 10–19.
- Ding, F. and Dokholyan, N.V. (2008) Dynamical roles of metal ions and the disulfide bond in Cu, Zn superoxide dismutase folding and aggregation. *Proc. Natl. Acad. Sci. USA*, **105**, 19696–19701.
- Ding, F. *et al.* (2005) Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. *Biophys. J.*, **88**, 147–155.
- Ding, F. *et al.* (2008) Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA*, **14**, 1164–1173.
- Ding, F. *et al.* (2012) Three-dimensional RNA structure refinement by hydroxyl radical probing. *Nat. Methods*, **9**, 603–608.
- Dobbins, S.E. *et al.* (2008) Insights into protein flexibility: the relationship between normal modes and conformational change upon protein-protein docking. *Proc. Natl. Acad. Sci. USA*, **105**, 10390–10395.
- Dror, R.O. *et al.* (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.*, **41**, 429–452.
- Eisenmesser, E.Z. *et al.* (2002) Enzyme dynamics during catalysis. *Science*, **295**, 1520–1523.
- Emperador, A. *et al.* (2008a) Exploring the suitability of coarse-grained techniques for the representation of protein dynamics. *Biophys. J.*, **95**, 2127–2138.
- Emperador, A. *et al.* (2008b) United-atom discrete molecular dynamics of proteins using physics-based potentials. *J. Chem. Theory Comput.*, **4**, 2001–2010.
- Emperador, A. *et al.* (2010) Protein flexibility from discrete molecular dynamics simulations using quasi-physical potentials. *Proteins*, **78**, 83–94.
- Emperador, A. *et al.* (2013) Efficient relaxation of protein-protein interfaces by discrete molecular dynamics simulations. *J. Chem. Theory Comput.*, **9**, 1222–1229.
- Falke, J.J. (2002) Enzymology. a moving story. *Science*, **295**, 1480–1481.
- Fenwick, R.B. *et al.* (2011) Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J. Am. Chem. Soc.*, **133**, 10336–10339.
- Flores, S. *et al.* (2006) The database of macromolecular motions: new features added at the decade mark. *Nucleic Acids Res.*, **34**, D296–D301.

- Franklin, J. et al. (2007) MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape. *Nucleic Acids Res.*, **35**, W477–W482.
- Gerstein, M. and Krebs, W. (1998) A database of macromolecular motions. *Nucleic Acids Res.*, **26**, 4280–4290.
- Gherghe, C.M. et al. (2009) Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics. *J. Am. Chem. Soc.*, **131**, 2541–2546.
- Henzler-Wildman, K.A. et al. (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature*, **450**, 838–844.
- Karplus, M. and Kuriyan, J. (2005) Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA*, **102**, 6679–6685.
- Kern, D. and Zuiderweg, E.R. (2003) The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.*, **13**, 748–757.
- Kim, M.K. et al. (2002) Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.*, **83**, 1620–1630.
- Krebs, W.G. and Gerstein, M.B. (2000) The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res.*, **28**, 1665–1675.
- Kubitzki, M.B. and de Groot, B.L. (2008) The atomistic mechanism of conformational transition in adenylate kinase: a TEE-REX molecular dynamics study. *Structure*, **16**, 1175–1182.
- Laio, A. and Parrinello, M. (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, **99**, 12562–12566.
- Leo-Macias, A. et al. (2005) An analysis of core deformations in protein superfamilies. *Biophys. J.*, **88**, 1291–1299.
- Leone, V. et al. (2010) Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.*, **20**, 148–154.
- Lezon, T.R. et al. (2009) Global Motions of the Nuclear Pore Complex: Insights from Elastic Network Models. *PLoS Comput. Biol.*, **5**, e1000496.
- Lindahl, E. et al. (2006) NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.*, **34**, W52–W56.
- Lindorff-Larsen, K. et al. (2005) Simultaneous determination of protein structure and dynamics. *Nature*, **433**, 128–132.
- Lindorff-Larsen, K. et al. (2011) How Fast-Folding Proteins Fold. *Science*, **334**, 517–520.
- Liphardt, J. et al. (2002) Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, **296**, 1832–1835.
- Lopez-Blanco, J.R. et al. (2011) iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics*, **27**, 2843–2850.
- Maragakis, P. and Karplus, M. (2005) Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J. Mol. Biol.*, **352**, 807–822.
- Marrink, S. et al. (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, **111**, 7812–7824.
- McCammon, J.A. et al. (1977) Dynamics of folded proteins. *Nature*, **267**, 585–590.
- Mendez, R. and Bastolla, U. (2010) Torsional network model: normal modes in torsion angle space better correlate with conformation changes in proteins. *Phys. Rev. Lett.*, **104**, 228103.
- Meyer, T. et al. (2010) MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics trajectories. *Structure*, **18**, 1399–1409.
- Micheletti, C. (2013) Comparing proteins by their internal dynamics: Exploring structure-function relationships beyond static structural alignments. *Phys. Life Rev.*, **10**, 1–26.
- Miyashita, O. (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. USA*, **100**, 12570–12575.
- Okazaki, K.I. et al. (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA*, **103**, 11844–11849.
- Orellana, L. et al. (2010) Approaching elastic network models to molecular dynamics flexibility. *J. Chem. Theory Comput.*, **6**, 2910–2923.
- Orozco, M. et al. (2011) Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. *Adv. Protein Chem. Struct. Biol.*, **85**, 183–215.
- Perilla, J.R. et al. (2010) Computing ensembles of transitions from stable states: dynamic importance sampling. *J. Comput. Chem.*, **32**, 196–209.
- Portella, G. and Orozco, M. (2010) Multiple routes to characterize the folding of a small DNA hairpin. *Angew. Chem. Int. Ed. Engl.*, **49**, 7673–7676.
- Proctor, E.A. et al. (2011) Discrete molecular dynamics. *WIREs Comput. Mol. Sci.*, **1**, 80–92.
- Proctor, E.A. et al. (2012) Discrete molecular dynamics distinguishes native-like binding poses from decoys in difficult targets. *Biophys. J.*, **102**, 144–151.
- Rueda, M. et al. (2004) Exploring the counterion atmosphere around DNA: what can be learned from molecular dynamics simulations? *Biophys. J.*, **87**, 800–811.
- Schlitter, J. (1994) Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graph.*, **12**, 84–89.
- Sfriso, P. et al. (2012) Finding conformational transition pathways from discrete molecular dynamics simulations. *J. Chem. Theory Comput.*, **8**, 4707–4718.
- Stein, A. et al. (2011) A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks. *Structure*, **19**, 881–889.
- Taketomi, H. et al. (1975) Studies on protein folding, unfolding and fluctuations by computer simulation. *Int. J. Pept. Protein Res.*, **7**, 445–459.
- Tozzini, V. (2005) Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.*, **15**, 144–150.
- Ueda, Y. et al. (1978) Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers*, **17**, 1531–1548.
- Urbanc, B. et al. (2004) In silico study of amyloid beta-protein folding and oligomerization. *Proc. Natl. Acad. Sci. USA*, **101**, 17345–17350.
- Urbanc, B. et al. (2010) Elucidation of amyloid β -protein oligomerization mechanisms: discrete molecular dynamics study. *J. Am. Chem. Soc.*, **132**, 4266–4280.
- Velazquez-Muriel, J.A. et al. (2009) Comparison of molecular dynamics and superfamily spaces of protein domain deformation. *BMC Struct. Biol.*, **9**, 6.
- Weiss, D.R. and Levitt, M. (2009) Can morphing methods predict intermediate structures? *J. Mol. Biol.*, **385**, 665–674.
- Whitford, P.C. et al. (2007) Conformational transitions of adenylate kinase: switching by cracking. *J. Mol. Biol.*, **366**, 1661–1671.
- Wiederstein, M. and Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, **35**, W407–W410.
- Yang, Z. et al. (2009) Allosteric transitions of supramolecular systems explored by network models: application to chaperonin GroEL. *PLoS Comput. Biol.*, **5**, e1000360.
- Zhang, Y. and Skolnick, J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins*, **57**, 702–710.
- Zhou, Y. and Karplus, M. (1997) Folding thermodynamics of a model three-helix-bundle protein. *Proc. Natl. Acad. Sci. USA*, **94**, 14429–14432.