Non-affine displacements encode collective conformational fluctuations in proteins

Dube Dheeraj Prakashchand, Navjeet Ahlawat, Satyabrata Bandyopadhyay,
Surajit Sengupta,* and Jagannath Mondal*

*Tata Institute of Fundamental Research, Center for Interdisciplinary sciences, Hyderabad
500107, India*

E-mail: surajit@tifrh.res.in, +914020203089; jmondal@tifrh.res.in, +914020203091

Abstract

Identifying subtle conformational fluctuations underlying the dynamics of bio macromolecules is crucial for resolving their free energy landscape. We show that a collective variable, originally proposed for crystalline solids, is able to filter out essential macromolecular motions more efficiently than other approaches. While homogenous or ‘affine’ deformations of the biopolymer are trivial, biopolymer conformations are complicated by the occurrence of in-homogenous or ‘non-affine’ displacements of atoms relative to their positions in the native structure. We show that these displacements encode functionally relevant conformations of macromolecule and, in combination with a formalism based upon time-structured independent component analysis, quantitatively resolve the free energy landscape of a number of macromolecules of hierarchical complexity. The kinetics of conformational transitions among the basins can now be mapped within the framework of a Markov state model. The non-affine modes, obtained by projecting out homogenous fluctuations from the local displacements, are found to be responsible for local structural changes required for transitioning between pairs of macro states.

Introduction

The conformational landscape underlying the dynamics of a bio macromolecule is complex and generally composed of rough terrains and shallow valleys. Resolving key basins corresponding to the metastable conformational states of the biomacromolecule in the landscape is a nontrivial task and has remained a topic of contention over the last few decades. A common approach in this direction has been the projection of the biomolecular phase space along judiciously constructed collective variables (CV). A clever choice of CV makes it possible to explore the mechanisms underlying conformational changes. A good CV¹ is one that a) provides dimensionality reduction b) based on dynamics and c) can predict future events. Traditionally intuition-based CVs have not only catered to the needs of deciphering the underlying pathways but also helped in efficiently biasing the MD simulation in enhanced samplings like umbrella sampling and metadynamics, and also building Markov state models (MSM).²⁻¹⁰ However, the heterogeneity associated with individual bio macromolecule has made the choice of CV non-unique, for optimally describing biomolecular conformation. This has resulted in the proposition of diverse sets of CVs over the years for capturing the biomolecular conformational dynamics. Many methodological developments have helped in devising an optimal CV for numerous complex systems. Spectral gap optimization of order parameter (SGOOP)⁶ algorithm proposed by Tiwary and Berne helps in capturing the combination of CVs that record the slowest dynamics of the system. Using a combination of time-structured independent component analysis (TICA) and metadynamics⁵ approach it has been shown that it is plausible to sample the slowest events and additionally drive the system along the slowest modes for enhanced sampling. Parrinello and coworkers,³ have recently developed an approach where initially a metadynamics is carried out using a combination of commonly used CVs and subsequently a variational TICA approach on this sampling reveals slow modes involved in the system. Then further metadynamics is performed using these slow modes as CVs. However, the underlying complexity of the conformational landscape has also necessitated the emergence of machine learning based approaches¹¹⁻¹⁴ for

CV optimisations.

On the other hand, in a crystalline solid, there exists an attractive prescription^{15–18} to resolve its overall displacement from a reference lattice into two displacement subspaces: a) “affine” displacement, which is the trivial homogenous deformations from the reference lattice and b) “non-affine” displacement, which is the remainder of the deformations (see Figure 1). This formalism reveals that the necessary precursors for the initiation of crystal lattice defects such as dislocation dipoles which are provided by the thermally excited non-affine fluctuations.¹⁶ Application of this formalism revealed that rigid solids¹⁹ are thermodynamically metastable even at infinitesimal stress, thereby enabling us to conclude that non-affine fluctuations are crucial in releasing the stress exerted on the yielding crystal. This helps rigid solids under external loading to decay from the metastable state to a stable state through a first order phase-transition. Using the non-affine formalism it is also possible to describe the non-trivial singular deformations like pleats and ripplocations^{20,21} undergone by solids having atoms bound by strong chemical bonds, which disallow any kind of dislocations. Also it has been shown,^{18,22} that colloids can be ordered in any given crystal structure by suppressing the non-affine displacements, with the help of special feedback-control laser traps. Using a similar projection formalism as described in Ganguly et al.,¹⁵ for large bio macromolecules lacking any kind of long-range structural order it has been shown by Dube et al.,²³ that important conformational changes are always accompanied by non-affine fluctuations and that regions in proteins with high susceptibility for non-affine fluctuations correlate with binding hotspots and the magnitude of the spatial correlations reveal allosteric connections in different locations of protein molecules.

As will be detailed below, the current work transitions this idea of ‘non affine displacement’, originally proposed for crystalline solids, to bio-macromolecules. We show that collective variables based upon non affine displacements in a bio-macromolecule, called ‘NAP’ and ‘GNAP’, can capture the crucial structural transitions in bio-macromolecules. In combination with time structured independent component analysis (TICA)^{24,25} and Markov-state

model, GNAP provides a theoretically robust strategy for identifying key conformational motion in protein.

Figure 1: **a.** Illustration of ‘affine’ and ‘non-affine’ deformation in a lattice. **b.** Extension of the concept of non-affinity to protein **c.** The time profile of radius of gyration and GNAP (relative to beta-hairpin conformation) of a 32 bead charged-neutral polymer. The blue beads are positively charged beads while the orange ones are negatively charged. **d.** The representative snapshots corresponding to different GNAPs in the part c

Theory

Non-affine displacements in a bio-macromolecule are derived by filtering out routine homogeneous displacements (known as 'affine deformation') from reference structure. Figure 1a-b schematically introduces us with non-affine displacements as well as the local and global non-affine parameter¹⁵⁻¹⁷ (NAP and GNAP respectively). Figure 1a illustrates affine and non-affine deformation for a crystal. Figure 1b illustrates the formalism in a protein:²³ Consider a neighbourhood Ω_i centered around an atom i . The instantaneous positions of the $j = 1, n_\Omega$ atoms in Ω_i may be compared with the set of fixed reference positions $\mathbf{R}_i$. The reference atomic positions, for example, could represent the native state of the protein. NAP defines the least square error made by trying to define $\mathbf{r}_i$ as a "best fit" affine deformation $\mathcal{D}$ of $\mathbf{R}_i$ viz. $\text{NAP}(i) > 0 = \min_{\mathcal{D}} \sum_{j=1}^{n_\Omega} [\mathbf{r}_j - \mathbf{r}_i - \mathcal{D}(\mathbf{R}_j - \mathbf{R}_i)]^2$. Here the minimisation is over choices of the deformation matrix $\mathcal{D}$.¹⁵ Thus $\text{NAP}(i)$ quantifies the "local structural changes" of i -th particle of the biomolecule. The central quantity of our current work, the global non-affine parameter (GNAP), is the average of NAP over the macromolecule (Figure 1b): $\text{GNAP} = N^{-1} \sum_{i=1}^N \text{NAP}(i)$

One can show^{15,17} that the thermal average of the local non-affine parameter is given by the trace of the displacement co-variance matrix projected onto the non-affine subspace viz. $\langle \text{NAP}(i) \rangle = \text{Tr } \mathbf{P}[\mathbf{C}(i)]\mathbf{P}$. Here $\mathbf{C}(i) = \langle \Delta^T(i)\Delta(i) \rangle$ is the displacement covariance with $\Delta(i)$ consisting of the relative displacements $\mathbf{u}_i - \mathbf{u}_j$ of the atoms from the reference positions arranged as a column vector with $3n_\Omega$ elements. The projection operator $\mathbf{P}$ depends *only* on the reference configuration and is obtained by first defining the $3n_\Omega \times 9$ dimensional block matrix $\mathbf{R}$ with elements $R_{j\alpha,\gamma\gamma'} = \delta_{\alpha\gamma}(R_{j\gamma'} - R_{i\gamma'})$, where the Greek indices correspond to the spatial dimensions and take values 1, 2 or 3. Finally, $\mathbf{P} = \mathbf{P}^2 = 1 - \mathbf{R}(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T$.

While the sum of the eigenvalues of $\mathbf{P}[\mathbf{C}(i)]\mathbf{P}$ determines the thermal average of NAP, the corresponding *eigenvectors* govern the atomic conformations which lead to non-affine displacements. For example, in crystals, these non-affine modes with the largest eigenvalues correspond to precursors to crystal defects^{16,17} and are implicated in plastic events and loss

of rigidity.²⁶ In proteins too, they correspond to local conformational changes which are important for ligand binding and allostery.²³ One may also characterize regions in the protein which are particularly amenable to non-affine displacements using the local NAP susceptibility defined as $\langle (\text{NAP}(i) - \langle \text{NAP}(i) \rangle)^2 \rangle$. In this sense, the NAP susceptibility behaves as any other thermodynamic local response function such as heat capacity or compressibility. We have implemented the computation of GNAP for any macromolecules in popular MD analysis tool PLUMED2.^{27,28}

Simulation method

We have used computational models of three different biomacromolecules in our current work: a) 32-bead model polymer with partial positive and negative charges along different beads, b) C-terminal domain of GB1 β -hairpin (residues 41–56) and c) Bovine pancreatic trypsin inhibitor (BPTI) protein. Two of these systems (polymer and β -hairpin) have been the subject of prior computational studies^{7,29} by our group. (see Figure 2) In the current work, we have employed same simulation protocols as have been done earlier. The polymer mimics bead-in-a-spring model with a total of 32 beads connected by harmonic bonds and angles, the details of which are documented else-where.²⁹ The polymer consists of alternative positive and negative charges on the beads in a periodic fashion. Apart from the electrostatic interaction, the polymer beads additionally interact with each other and the aqueous media via lenard Jones (LJ) interaction. The LJ parameters of the polymer beads have been kept fixed at $\sigma = 0.4$ nm and $\epsilon_p = 1.0$ kJ/mol. The other cross polymer-solution LJ parameters have been determined by geometric combination rule. As in previous case, the solvent has been modelled as SPC/E water model.³⁰ All simulations have been done with a single polymer chain inside a cubic box of volume ≈ 125 nm³ consisting of 3850 water molecules at an average temperature of 300 K and an average pressure of 1 bar using NPT ensemble. The details of the thermostat and barostat in addition to other simulation details can be

Figure 2: Computational models of the three systems studied in the current work: a) 32-bead model polymer, b) C-terminal domain of GB1 β -hairpin (residues 41–56) and c) Bovine pancreatic trypsin inhibitor (BPTI) protein

found in our previous work.²⁹ We have run 200 independent simulations, each 30 ns long, by starting with diverse conformation of the polymer.

The 16-residue long peptide Ace-⁴¹GEWTYDDATKTFTVTE⁵⁶-Nme, as taken out from the full-length GB1 protein NMR structure (PDB: 1GB1³¹) was modelled using AMBER03-STAR^{32,33} forcefield and solvated in a truncated octahedron simulation cell containing 984 TIP3P³⁴ water molecules and electroneutralized. The current work utilizes our previously run 200 short MD trajectories of the β -hairpin, each 100 ns long and spawned from prior replica exchange molecular dynamics (REMD) derived trajectories.⁷ In addition, we have spawned 200 more MD trajectories (and hence a total of 400 trajectories) to improve the statistics of constructed conformational landscape for β -hairpin. The readers are referred to work of Mondal and coworkers⁷ for simulation details.

The simulation trajectories for BPTI correspond to prior work of Shaw and coworkers³⁵ and were generously provided by D. E. Shaw research.

All frameworks of Time structured Independent Component Analysis (TICA) and Markov

state model (MSM) were developed using pyemma.³⁶

Results and Discussion

This communication illustrates that GNAP encodes relevant information concerning conformational folding of a macromolecule and in combination with TICA, can serve as an efficient collective variable. Towards this end, we first consider a model 32-bead charge-neutral polymer with equal number of beads having partial charges of opposite polarity (see representative snapshot in Figure 1c and Figure 2 a)). Recent MD simulation by Mondal and coworkers²⁹ (also see simulation methods) has shown that this polymer can undergo reversible collapsed $\rightleftharpoons$ extended transition in aqueous media and the transition often involves a hairpin-like intermediate as well. We calculate the time profile of GNAP of the polymer from its representative MD simulation trajectories. As depicted in Figure 1c, we find that GNAP captures the dynamic evolution of the polymer with time and responds in unison with conformational fluctuation of the polymer. Specifically, we find that, starting with collapsed conformation as reference, the value of GNAP increases with transition from collapsed to hairpin or extended conformation and vice versa (Figure 1d)

The observation that GNAP of the polymer can dynamically evolve, in sync with its conformational fluctuation, encouraged us to explore the power of GNAP in mapping the free energy landscape of a macromolecule. However, one of the key features of GNAP is that its computation depends on a reference structure. For a conformationally floppy macromolecule, the choice of reference structure may not always be unique, which in turn influences the reconstruction of the free energy landscape. This is demonstrated by the supplementary Figure S1. To circumvent the dependence of GNAP on single reference structure, we here develop a robust protocol by combining techniques of TICA^{24,25} with GNAP. Our proposed work flow is enumerated in Figure 3. The main purpose of this protocol is to optimally combine the GNAPs obtained from multiple reference configurations. Towards this end, we

Figure 3: Scheme representing the computational protocol described in the current work. The trajectories based on adapting sampling is the source of the data points. The data points are first subjected to K-mean clustering. The GNAP calculation requires a reference structure. The output of the K-mean clustering i.e. the representative structures of the clusters are used as the different structures and the GNAP values so obtained act as the raw data for the TICA step. Collecting such TICA components which account for the 90 percent of the kinetics of the system is required to properly build a markov state model.

first perform adaptive molecular dynamics sampling of the macromolecule of interest. The aggregated conformations resulting from the adaptive sampling is subsequently subjected to K-mean clustering.³⁷⁻³⁹ The cluster centres thereby obtained provide us with a set of representative reference structures, each of which is used to compute individual GNAP. These multiple GNAP values, corresponding to as many reference structures, are then optimally combined in a linear fashion within the frame-work of TICA. As has been shown earlier by Pande and Noe,^{24,25} TICA efficiently optimises the coefficients of the linear combination by maximising the auto-correlation of projection. The number of TICA dimensions gets further reduced by choosing only those which conserves 90% of the kinetic-variance.

Figure 4: The 4 state markov model for the 32 bead polymer chain. **a.** Free energy profile along the two underlying slowest mode of relaxation reveal four minima . . **b** the various fluxes connecting the four minima, with the thickness of the arrows representing the relative fluxes. **c.** the conformational macrostate at the four minima. The hairpin structure is the most populous and the collapsed structure is the next most populous structure. Also there a state which is a distorted version of the hairpin very close by the perfect hairpin structure.

When this strategy was applied for the model polymer discussed earlier, a set of 12 reference structures were found to exhaustively cluster the conformational ensembles and hence with these 12 reference structures we calculate a set 12 GNAP values for our adaptive sampling trajectories. In the case of model polymer, we noted that just 3 out of 12 TICA dimensions were sufficient to conserve 90% of the kinetic-variance. Figure 4a projects the

two-dimensional free energy landscape of the model polymer along top two slowest linear combinations of GNAPs, namely TIC_1 and TIC_2 . The free energy surface is obtained after reweighing the probability distribution by the stationary distributions. We find that the optimal linear combinations of GNAP are able to fully resolve the landscape of the model polymer into multiple key conformations (figure 4c), namely extended (state-0, 34.39%), collapsed (state-1, 23.29%), distorted hair-pin (state-2, 7.16%) and symmetric hair pin (state-3, 35.13%), which comes out to be most populated conformation. Reconstructed transition networks (figure 4b) among these GNAP-projected key polymer conformations, as obtained via a Markov state model (MSM) analysis⁴⁰ (over 800 discrete micro states at 250 ps lag time), show diverse array of rapid state-to-state transitions, especially between collapsed and perfect hairpin structure.

Figure 5: **a.** The 5 state markov model for the GB1 protein, mapped on the free energy surface along TIC_1 and TIC_2 (see text for details of TIC_1 and TIC_2) **b.** Transition network among the macro-states (the thickness of the arrows represents the relative fluxes). The free-energy profile recovers that the transitions among the structurally similar conformations 0 , 1 and 2 (these structures lie along the second slowest modes) are somewhat faster compared to those which are structurally quite distinct which requires huge structural transformation **c.** Snapshot of structures corresponding to the minima.

The ability of GNAP in exhaustive quantification of underlying conformational landscape becomes more prominent in hierarchically more complex systems than the model polymer : β -hairpin GB1 peptide and Bovine pancreatic trypsin inhibitor (BPTI) protein (see Figure 2b and 2c). In particular, for GB1 peptide, (see simulation methods and our earlier work⁷ for details of simulation models and methods), we find that (see Figure 5) the combination of GNAP together with the protocol described in Figure 3 brings out the complex free energy landscape along the slowest and second slowest dimensions. Subsequent MSM analysis (based on 1000 discrete micro states at a lag time of 600 ps demonstrates that a 5-state model with state-0 being the symmetrically distorted beta hair-pin (14.85%), state-1, perfect beta hair-pin (49.85%, state-2: asymmetrically distorted beta hair-pin (7.10%), state-3: terminal α -helix (3.67%) and state-4: central α -helix (24.51%) aptly describes the overall landscape (see figure 5). The transition-network shows that flux from state 1 to any other state, especially state 1-to-3 are extremely feeble, implying that the transformation from beta hair-pin to terminal helix is most rare and difficult. This is also backed up by our common understanding that such a transformation requires a huge structural re-arrangement at multiple sites.⁷

An important aspect of the current formalism is that we can investigate the most probable non-affine modes by projecting out homogenous fluctuations from the local displacements. These non-affine modes can potentially infer local structural changes between a pair of macro states. We illustrate this via the supplemental movie S1 which demonstrates the non-affine modes responsible for transition of GB1 from beta hair-pin structure to the central helix structure. The movie also demonstrates a key advantage of GNAP over popular approach of principal component analysis (PCA) (see supplemental movie S2 for PCA):^{41,42} the current approach can explore local changes and the projection on the non-affine subspace ensures that all trivial motions are filtered out and thereby it is possible to retrieve and analyse predominant modes responsible for the conformational changes. Figure 6 quantifies the key difference between PCA and GNAP-analysis for the above-mentioned sampling of GB1-hairpin-transition from β -sheet to central α -helix. In Figure 6a, the spatial distribution

of the normalized NAP-susceptibility^{16,23} across the protein molecule is calculated over the aforementioned sampling. The two particles in red and also circled in red are the centers of high NAP-susceptibility and therefore likely to produce non-affine conformational changes. To explore the difference between PCA and NAP on equal footing, we solve the eigenvalue equation of their individual covariance matrices. The eigen spectra at the two locations, as obtained by current formalism, are shown in Figure 6b where it is clearly seen that the highest eigenmode is completely separated from the rest of the eigen-spectrum for the two high NAP-susceptibility regions. This large separation confirms that the modes associated with the separated eigenvalues are indeed non-random. On the other hand, the PCA-derived eigen spectrum (Figure 6b) shows a relatively lower gap in its associated eigenvalues than those obtained in case of NAP-derived covariance matrix. A comparison of displacement modes at these two locations (Figure 6c) derived from NAP analysis and from PCA analysis shows that the NAP analysis recovers the helical curvatures and the twists that are required to attain the central α -helix begin to appear, while the PCA mode provides a much limited resolution in regard with local mechanism involved in the transition from hair-pin to central alpha helix. It can also be seen from Figure 6d that the net effective displacement vector involved in the transition from β -sheet to central α -helix conformation projected over the most prominent eigenmodes (indicated by the highest eigen values) in the two methods is higher in the case of NAP-analysis than PCA-analysis, which suggests that the modes obtained from the NAP-analysis are able to mimic the underlying dynamics involved in this transition more closely than the PCA analysis.

Finally, going beyond simple polymers and peptides, GNAP is found to aptly describe the conformational phase-space for prototypical protein BPTI (Figure 7). The milli-second long simulations by D.E. Shaw group³⁵ have earlier provided atomic-scale insights into the folding/unfolding events of BPTI. The projection of the trajectory (generously provided by D. E. Shaw research) along the top two TICA-derived slowest GNAP-combination reveals multiple key intermediates of BPTI (Figure 7a). Interestingly, similar projection along more

Figure 6: Comparison between GNAP and PCA analysis for GB1 protein system **a**. Spatial distribution of NAP susceptibility. This analysis provides the hotspots region at indicated regions in protein (the regions, locations 1 and 2, that are red in colour). **b**. The non-affine displacement modes derived from the GNAP analysis better manifest the helical topologies required for going from β -sheet to central α -helix conformation than PCA. **c**. Comparison of eigen-spectrum obtained from NAP analysis at the two locations 1 and 2 with the PCA eigen-spectrum for the entire GB1 system. **d**. The quantification of how well NAP analysis predicts the transition process, compared to the PCA analysis, via the projection of the $\Delta_{transition}$ (defined as the difference between the coordinates of the β -sheet and central α -helix conformations) on the highest eigen value modes obtained in the two methods.

typical CV namely RMSD fails to resolve all the key intermediates (figure 7b) The macro state corresponding to folded native conformation automatically appears as the free energetically most stable state (state 1) in the GNAP reconstructed free energy projection. At this minima, we recover a structure with left-handed conformation of the disulfide bridge between Cys14 and Cys38. Interestingly, this has been previously reported by Shaw and coworkers as the most populated state. On the other hand, state 2 represents conformation with large deviation from native structure, which is the unfolded conformation. Most importantly, the GNAP analysis recovers the well-known "left-handed" conformation of the disulfide bridge between Cys14 and Cys38 in state 1, 2 and 4, as earlier reported by simulations and NMR experiments. Finally, state 3 brings out the 'right handed' conformation^{43,44} (see Grey et al⁴³ and captions of Figure 7 for details.)

Figure 7: Comparison of GNAP with RMSD for BPTI system **a**. The free-energy profile along TIC1 (slowest mode of relaxation) and TIC2 (the second slowest mode of relaxation) constructed out of linear combination of multiple GNAPs corresponding to representative reference structures obtained in the K-mean cluster step **b**. The free-energy profile along TIC1 and TIC2 constructed out of linear combination of RMSDs relative to the same sets of reference representative structures, as used in case of GNAPs. The four structures 1, 2, 3 and 4 have the same reference structures(in grey ghost representation) while the structure from the four minima are shown in cyan with distorted part in green. M (structure 3) is the right-handed disulphide conformation and m_{c14} (structure 1 and 4 (with distortion)) and m_{c38} (structure 2) are the left-handed disulphide conformation, as reported by Grey et al.⁴³ Structure 4 is not captured in the RMSD-TICA free-energy profile.

Conclusion

In summary, the current work demonstrates that the non-affine displacements in bio macro-molecule can act as a good marker for characterising its key conformational fluctuations. Specifically, we formulate a global non-affine parameter (GNAP) by averaging over all local non-affine displacements corresponding to each protein heavy atoms and we show that GNAP in turn can capture the essential collective motion in bio-macromolecules. We also find that GNAP, when used in combination with TICA, judiciously handles the choice of multiple reference states and resolves the underlying free energy landscape into key basins. These are efficiently demonstrated for a series macromolecules of hierarchical complexity (model polymer, peptides and proteins). More over, GNAP is found to provide better resolution of the free energy landscape of a protein BPTI compared to RMSD, a conventional CV regularly used for projecting the biomacromolecular motion. The non-affine modes, obtained by projecting out homogenous fluctuations from the local displacements, are found to be responsible for local structural changes required for transitioning between a pair of macro states. Finally, we show that a local NAP analysis can be used to filter out the essential macromolecular configurations responsible for slowly evolving conformational changes such as that from a β -sheet to α -helix, pinpointing the residues where such a change is likely to initiate. Overall, identification of non-affine fluctuations in protein offers a distinct advantage for understanding the bio-macromolecular dynamics and its quantifier GNAP holds the promise of acting as an optimal CV for projecting the free energy landscape. The availability of an analytical expression for GNAP makes it possible to patch it with any standard MD analysis tool for post-simulation analysis . Along this line, the current work has seen the implementation of GNAP into popular open access MD analysis tool PLUMED.²⁷ This holds future promises of driving a biophysical process along its collective non-affine displacement using enhanced sampling techniques.

Acknowledgement

This work was supported by computing resources obtained from shared facility of TIFR Center for Interdisciplinary Sciences, India. JM would like to acknowledge research intramural research grants obtained from TIFR, India, Ramanujan Fellowship and Early Career Research funds provided by the Department of Science and Technology (DST) of India (ECR/2016/000672). We thank D. E. Shaw research for generous access to their simulation trajectories of Bovine pancreatic trypsin inhibitor.

Supporting Information (SI)

Figure demonstrating the variation of free energy due to different choice of reference structures(PDF). Movies showing non-affine modes and PCA modes for transition from beta-sheet to central helix of GB1 peptide.(mpeg)

References

- (1) McGibbon, R. T.; Husic, B. E.; Pande, V. S. Identification of simple reaction coordinates from complex dynamics. *The Journal of Chemical Physics* **2017**, *146*, 044109.
- (2) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (3) McCarty, J.; Parrinello, M. A variational conformational dynamics approach to the selection of collective variables in metadynamics. *The Journal of Chemical Physics* **2017**, *147*, 204109.
- (4) Yang, Y. I.; Parrinello, M. Refining Collective Coordinates and Improving Free Energy Representation in Variational Enhanced Sampling. *Journal of Chemical Theory and Computation* DOI:10.1021/acs.jctc.8b00231, PMID: 29715017.

- (5) M. Sultan, M.; Pande, V. S. tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *Journal of Chemical Theory and Computation* **2017**, *13*, 2440–2447.
- (6) Tiwary, P.; Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2839–2844.
- (7) Ahalawat, N.; Mondal, J. Assessment and optimization of collective variables for protein conformational landscape: GB1 beta-hairpin as a case study. *The Journal of Chemical Physics* **2018**, *149*, 094101.
- (8) Mendels, D.; Piccini, G.; Parrinello, M. Collective Variables from Local Fluctuations. *The Journal of Physical Chemistry Letters* **2018**, *9*, 2776–2781, PMID: 29733652.
- (9) Yang, Y. I.; Niu, H.; Parrinello, M. Combining Metadynamics and Integrated Tempering Sampling. *The Journal of Physical Chemistry Letters* **2018**, *9*, 6426–6430.
- (10) Capelli, R.; Carloni, P.; Parrinello, M. Exhaustive Search of Ligand Binding Pathways via Volume-Based Metadynamics. *The Journal of Physical Chemistry Letters* **2019**, *10*, 3495–3499.
- (11) Wang, J.; Ferguson, A. L. Nonlinear reconstruction of single-molecule free-energy surfaces from univariate time series. *Phys. Rev. E* **2016**, *93*, 032412.
- (12) Schneider, E.; Dai, L.; Topper, R. Q.; Drechsel-Grau, C.; Tuckerman, M. E. Stochastic Neural Network Approach for Learning High-Dimensional Free Energy Surfaces. *Phys. Rev. Lett.* **2017**, *119*, 150601.
- (13) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nature Communications* **2018**, *9*, 5.
- (14) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted autoencoded variational

- Bayes for enhanced sampling (RAVE). *The Journal of Chemical Physics* **2018**, *149*, 072301.
- (15) Ganguly, S.; Sengupta, S.; Sollich, P.; Rao, M. Nonaffine displacements in crystalline solids in the harmonic limit. *Phys. Rev. E* **2013**, *87*, 042801.
- (16) Ganguly, S.; Sengupta, S.; Sollich, P. Statistics of non-affine defect precursors: tailoring defect densities in colloidal crystals using external fields. *Soft Matter* **2015**, *11*, 4517–4526.
- (17) Popli, P.; Kayal, S.; Sollich, P.; Sengupta, S. Exploring the link between crystal defects and non-affine displacement fluctuations. *Physical Review E* **2019**, *033002*, 1–17.
- (18) Ganguly, S.; Mohanty, P. S.; Schurtenberger, P.; Sengupta, S.; Yethiraj, A. Contrasting the dynamics of elastic and non-elastic deformations across an experimental colloidal Martensitic transition. *Soft Matter* **2017**, *13*, 4689–4697.
- (19) Nath, P.; Ganguly, S.; Horbach, J.; Sollich, P.; Karmakar, S.; Sengupta, S. On the existence of thermodynamically stable rigid solids. *Proceedings of the National Academy of Sciences* **2018**, *115*, E4322–E4329.
- (20) Ganguly, S.; Nath, P.; Horbach, J.; Sollich, P.; Karmakar, S.; Sengupta, S. Equilibrium and dynamic pleating of a crystalline bonded network. *The Journal of Chemical Physics* **2017**, *146*, 124501.
- (21) Ganguly, S.; Das, D.; Horbach, J.; Sollich, P.; Karmakar, S.; Sengupta, S. Plastic deformation of a permanently bonded network: Stress relaxation by pleats. *The Journal of Chemical Physics* **2018**, *149*, 184503.
- (22) Popli, P.; Ganguly, S.; Sengupta, S. Translationally invariant colloidal crystal templates. *Soft Matter* **2018**, *14*, 104–111.

- (23) Dube, D.; Ahalawat, N.; Khandelia, H.; Mondal, J.; Sengupta, S. On identifying collective displacements in apo-proteins that reveal eventual binding pathways. *PLOS Computational Biology* **2019**, *15*, e1006665.
- (24) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *Journal of Chemical Theory and Computation* **2013**, *9*, 2000–2009.
- (25) Perez-Hernandez, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noe, F. Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics* **2013**, *139*, 015102.
- (26) Nath, P.; Ganguly, S.; Horbach, J.; Sollich, P.; Karmakar, S.; Sengupta, S. On the existence of thermodynamically stable rigid solids. *Proceedings of the National Academy of Sciences* **2018**, *115*, E4322–E4329.
- (27) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **2009**, *180*, 1961–1972.
- (28) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185*, 604 – 613.
- (29) Mukherjee, M.; Mondal, J. Osmolyte-Induced Collapse of a Charged Macromolecule. *The Journal of Physical Chemistry B* **2019**, *123*, 4636–4644.
- (30) Berendsen, H.; Grigera, J.; Straatsma, T. P. The Missing Term in Effective Pair Potential. *J.Phys.Chem.* **1987**, *91*, 6269.

- (31) Gronenborn, A.; Filpula, D.; Essig, N.; Achari, A.; Whitlow, M.; Wingfield, P.; Clore, G. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **1991**, *253*, 657–661.
- (32) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point charge force field for molecular mechanics simulations of proteins based on condensed phase quantum mechanical calculations. *Journal of Computational Chemistry* **2003**, *24*, 1999–2012.
- (33) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix Coil Transition of Polypeptides. *The Journal of Physical Chemistry B* **2009**, *113*, 9004–9015.
- (34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.
- (35) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346.
- (36) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noe, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **2015**, *11*, 5525–5542.
- (37) Forgy, E. Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification. *Biometrics* **1965**, *21*, 768–769.
- (38) Wilkin, G. A.; Huang, X. K-Means Clustering Algorithms: Implementation and Com-

- p>
- parison. Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007). 2007; pp 133–136.
- (39) Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics* **1979**, *28*, 100.
- (40) Bowman, G. R.; Pande, V. S.; Noé, F. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Springer Netherlands, 2014.
- (41) Abdi, H.; Williams, L. J. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* **2010**, *2*, 433–459.
- (42) R., B. B.; Dusanka, J.; Martin, K. Harmonic analysis of large systems. I. Methodology. *Journal of Computational Chemistry* *16*, 1522–1542.
- (43) Grey, M. J.; Wang, C.; Palmer, A. G. Disulfide Bond Isomerization in Basic Pancreatic Trypsin Inhibitor: Multisite Chemical Exchange Quantified by CPMG Relaxation Dispersion and Chemical Shift Modeling. *Journal of the American Chemical Society* **2003**, *125*, 14324–14335.
- (44) Otting, G.; Liepinsh, E.; Wuethrich, K. Disulfide bond isomerization in BPTI and BPTI(G36S): An NMR study of correlated mobility in proteins. *Biochemistry* **1993**, *32*, 3571–3582.

TOC Graphic