

RESEARCH ARTICLE

Hierarchical Conformational Analysis of Native Lysozyme Based on Sub-Millisecond Molecular Dynamics Simulations

Kai Wang¹, Shiyang Long¹, Pu Tian*^{1,2}

1 School of Life Sciences, Jilin University, Changchun, China, **2** MOE Key Laboratory of Molecular Enzymology and Engineering, Jilin University, Changchun, China

* tianpu@jlu.edu.cn



OPEN ACCESS

Citation: Wang K, Long S, Tian P (2015) Hierarchical Conformational Analysis of Native Lysozyme Based on Sub-Millisecond Molecular Dynamics Simulations. PLoS ONE 10(6): e0129846. doi:10.1371/journal.pone.0129846

Academic Editor: Xuhui Huang, Hong Kong University of Science and Technology, HONG KONG

Received: March 12, 2015

Accepted: May 12, 2015

Published: June 9, 2015

Copyright: © 2015 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are available within the paper and its Supporting Information files.

Funding: This research was partially funded by a start-up fund from Jilin University and by 481 National Natural Science Foundation of China under grant number 31270758. 482 Computational resources were partially supported by High Performance Computing 483 Center of Jilin University, China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Hierarchical organization of free energy landscape (FEL) for native globular proteins has been widely accepted by the biophysics community. However, FEL of native proteins is usually projected onto one or a few dimensions. Here we generated collectively 0.2 milli-second molecular dynamics simulation trajectories in explicit solvent for hen egg white lysozyme (HEWL), and carried out detailed conformational analysis based on backbone torsional degrees of freedom (DOF). Our results demonstrated that at micro-second and coarser temporal resolutions, FEL of HEWL exhibits hub-like topology with crystal structures occupying the dominant structural ensemble that serves as the hub of conformational transitions. However, at 100ns and finer temporal resolutions, conformational substates of HEWL exhibit network-like topology, crystal structures are associated with kinetic traps that are important but not dominant ensembles. Backbone torsional state transitions on time scales ranging from nanoseconds to beyond microseconds were found to be associated with various types of molecular interactions. Even at nanoseconds temporal resolution, the number of conformational substates that are of statistical significance is quite limited. These observations suggest that detailed analysis of conformational substates at multiple temporal resolutions is both important and feasible. Transition state ensembles among various conformational substates at microsecond temporal resolution were observed to be considerably disordered. Life times of these transition state ensembles are found to be nearly independent of the time scales of the participating torsional DOFs.

Introduction

During last two decades, great progress has been made in the study of major conformational transitions for some functionally important proteins [1–10]. Such transitions usually exhibit spatial displacement of some structural elements up to a few nanometers, often involve collective motions of many residues and occur on time scales ranging from microseconds to milliseconds and beyond. Computationally, elastic network model (ENM) based methodologies have

Competing Interests: The authors have declared that no competing interests exist.

been successfully applied to explain functional relevance of such large scale conformational change [11, 12]. Unfortunately, the complexity of protein conformational space does not stop here. For a given protein, each major conformation is compatible with many conformational substates (CSs) that have different combinations of side chain and backbone torsional states. Transitions among these CSs manifest themselves as rotation of backbone dihedrals ϕ , ψ and side chains on time scales ranging from sub-nanoseconds to micro-seconds and beyond. These motions are qualitatively termed as flexibility, which is widely acknowledged to be important or sometimes critical in various molecular interactions [13]. However, due to the astronomically large number of possible CS, not much effort has been invested in studying CS transitions, distributions and organization in conformational space. A series of seminal experimental studies on the rebinding dynamics of CO by myoglobin provided strong support for hierarchical organization of free energy landscape (FEL) in native proteins [14, 15]. However, multi-dimensional native FEL of protein is only projected onto one measurable dimension in these experiments, the topological organization of CS at various temporal resolutions and corresponding detailed structural information of the whole protein is not available.

Native (functional) protein FEL has been analyzed in detail by utilizing atomistic molecular dynamics (MD) simulation trajectories that are a few hundred nanoseconds to microseconds-long [16]. Major methodologies are various forms of principal component analysis (PCA) that are based upon either cartesian or internal coordinates (mainly dihedral angles). Hierarchical organization of CS is supported by these studies on the FEL of various proteins in the vicinity of starting crystal structures. However, in addition to limitation of low dimensions, time scale is quite limited and conclusions made may not be readily extended to the global organization of native FEL. The majority of experimental and computational studies of FEL for native proteins up to date are designed to explain how given known functions/properties are related to the underlying FEL. Reduction of dimension (such as projection of FEL onto certain experimentally measurable quantities [14] or various forms of PCA [16]) is an effective means. With increasing environmental challenges (e.g. nanomaterials and new drugs) faced by biological systems, our concern is expanded from interactions among endogenous biological molecular partners to interactions between interested biomolecules and large number of chemical entities. Therefore, to transform the role of conformational analysis of proteins from explaining known to predicting a wide spectrum of molecular interactions, it might potentially be helpful to study global native FEL of proteins in as high dimensional space that we may understand as possible. Despite great progresses that have been achieved by the docking community [17, 18], it is well acknowledged that obtaining accurate structures of complexes from structures of their comprising proteins is a difficult task. A sufficiently complete backbone native structural ensemble was demonstrated to improve docking significantly [19]. Proper treatment of CS (or flexibility) has become a major bottleneck in improving prediction power of docking (both protein-ligand and protein-protein) methodologies [13, 20]. Therefore, a deeper fundamental understanding of the distribution and organization of CS is highly desired.

Great breakthroughs have been achieved in understanding folding mechanisms of small proteins by millisecond MD simulations [9, 21–24]. Despite acknowledged uncertainties [24–27], many MD simulation folding studies [24, 28] demonstrated that modern molecular mechanical force fields are reasonably accurate to distinguish the native ensemble from many unfolded ensembles. Inspired by successful folding studies with MD simulations [24], we posed the following questions: i) how are CS organized within conformational space of native proteins at various temporal resolutions (or free energy hierarchies); ii) how does FEL of native proteins generated by molecular mechanical force fields compare with available experimental data and; iii) among astronomically large numbers of possible CS, what fractions are of realistic significance in determining physiochemical properties and biological functions of a given

protein. These questions are difficult to tackle in a general sense. We chose to use HEWL as a model protein and CHARMM22 (with CMAP) as a typical set of force field parameters. This particular case is likely to serve as a representative for many other globular proteins. We generated collectively 0.2ms MD trajectories for HEWL in explicit solvents (water and NaCl). Based on distributions of backbone dihedrals, 50,000,000 snapshots were clustered at temporal resolutions ranging from $\sim ns$ to $\sim 20\mu s$. It is found that at μs and coarser temporal resolutions, native FEL exhibit a hub-like topology with the dominating ensemble, which harbors nearly all crystal structures, serving as the hub connecting smaller structural ensembles among which mutual transitions happen occasionally. At 100ns and finer temporal resolutions, however, HEWL clusters form network-like organization, and crystal structures are associated with significant ensembles that form kinetic traps. The number of backbone CS that is of statistical significance is found to be rather limited even at ns temporal resolution, revealing strong conformational correlations among backbone dihedrals. Comparison with available crystal structures indicates that backbone torsional state transitions involved in various molecular interactions span time scales ranging from ns to $\sim 10\mu s$.

Results

Sets of HEWL CS ensembles at multiple temporal resolutions

To explore the native conformational space of HEWL, we generated 2000 100 – ns MD trajectories and saved 50,000,000 structures (with 4ps intervals). To verify that our MD trajectories have achieved sufficient sampling for the native HEWL conformational space, we performed clustering on various subsets of trajectories. The high level agreement regarding the relative importance of significant clusters indicate that our sampling of the HEWL conformational space is reasonably sufficient (Fig. A in [S1 File](#)). Distributions were plotted for each ϕ and ψ that exhibit multiple-peaks (Table A and Fig. B in [S1 File](#)) under given temporal resolutions ([Table 1](#)), and were utilized to assign all the saved structures to different clusters and transition state ensembles. Briefly, 127 backbone dihedrals with two-major-peak distributions were defined as two-state torsional DOFs, and were divided into 5 different temporal resolutions according to the number of observed transitions between the two torsional state of each selected torsion. Under a given time resolution (say T_2), two snapshots belong to the same cluster if and only if for each of the torsional DOF that has the same or coarser time resolutions (i.e. two-state torsional DOFs at T_0, T_1 and T_2), they share the same torsional state (see [Methods](#) for details). The number of clusters obtained is listed for five specified temporal resolutions (T_0, T_1, T_2, T_3 and T_4 , see [Table 1](#)). At the temporal resolution T_4 ($\sim ns$), the statistical weight (W) of snapshots in each cluster (used interchangeably with CS hereafter) is plotted in [Fig 1a](#) and the percentage of snapshots in the largest n clusters (W_{sum}) were plotted as a function of n in [Fig 1b](#)

Table 1. Five temporal resolutions utilized for hierarchical backbone conformational analysis of HEWL. Corresponding number of conformational substates is provided in the right column.

Temporal resolution	N_{Trans}	$200(\mu s)/N_{Trans}$	N_{CS}
T_0	$0 < N_{Trans} \leq 10$	20 μs to 200 μs	10
T_1	$10 < N_{Trans} \leq 100$	2 μs to 20 μs	35
T_2	$100 < N_{Trans} \leq 1000$	200ns to 2 μs	434
T_3	$1000 < N_{Trans} \leq 10000$	20ns to 200ns	12967
T_4	$10000 < N_{Trans} \leq 100000$	2ns to 20ns	40356

doi:10.1371/journal.pone.0129846.t001

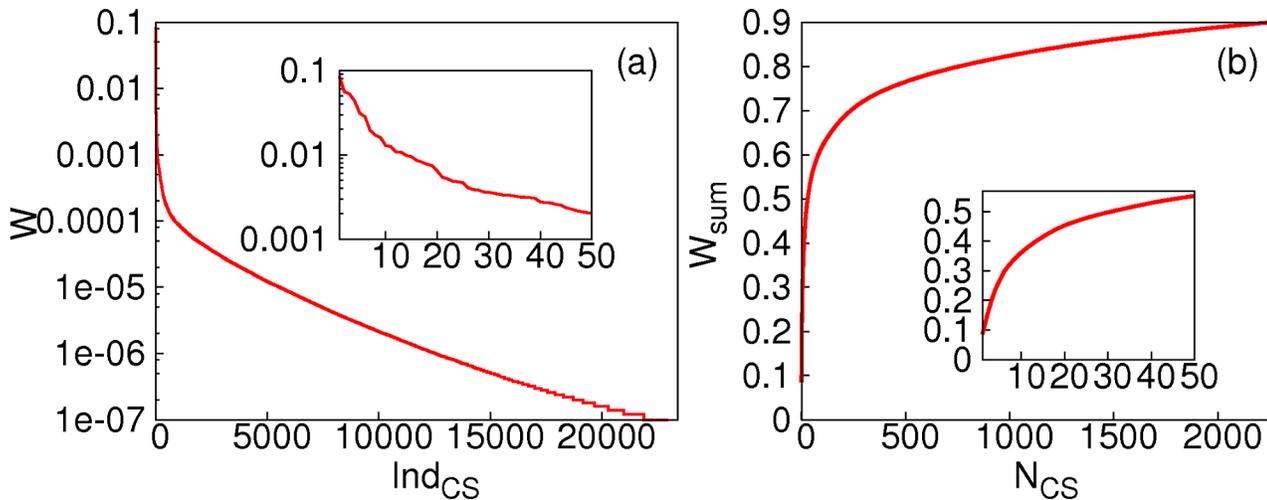


Fig 1. a) Statistical weight (W) of the largest 23400 (out of 40356) CSs at time resolution T_4 , the horizontal axis are indices for CS (Ind_{CS}). Inset: W for the 50 largest CSs. b) Collective statistical weight (W_{sum}) of the N largest CSs (N_{CS}) at time resolution T_4 . Inset: magnification for small N_{CS} .

doi:10.1371/journal.pone.0129846.g001

(See Fig. C in [S1 File](#) for similar plots at other temporal resolutions). Apparently, the number of CS observed in our simulations is quite limited, and the number of CS that have statistical significance is much less (with only 868 CSs has a weight larger than 0.01% at the temporal resolution T_4). These observations suggest that quantitative analysis of significant CS at temporal resolutions as fine as ns is potentially feasible, at least for backbone conformational space. It is noted that the number of clusters depends highly on the temporal resolution used to perform clustering, and is much smaller for coarser temporal resolutions ([Table 1](#)). Statistics at temporal resolution T_0 is quite limited, and our analysis and discussion hereafter focus on temporal resolutions T_1 through T_4 .

The hierarchical organization of obtained HEWL CS in backbone conformational space is plotted in [Fig 2\(a-d\)](#). At temporal resolution T_1 , a hub-like topology is observed and the dominating ensemble serves as the hub of conformational transitions. At finer temporal resolutions (T_3 and T_4), CSs of HEWL exhibit network-like organization, and both the average number of neighboring clusters (that have direct transitions to and from a given cluster) and its variation increases ([Fig 2](#)). In [Fig 2](#), Approximately half of the lines represent more than a dozen direct transitions between clusters, and a significant fraction (20% to 30% dependent upon temporal resolutions) of lines represent hundreds or more inter-cluster transitions. This fact further suggest that our sampling is reasonably sufficient.

Comparison with wild type crystal structure ensemble

To examine the experimental relevance of our obtained CSs, 120 crystal structures of wild type HEWL are clustered using the same criteria established for MD snapshots. As shown in [Fig 2](#) and [Table B](#) in [S1 File](#), At temporal resolutions T_1 and T_2 , nearly all examined crystal structures are located in the dominating structural ensemble from our simulations. However, at T_3 and finer temporal resolutions ($\leq 200ns$), crystal structures are scattered in more and more clusters that are significant but not dominant anymore. This observation might indicate the deficiency of the utilized force fields in differentiating CS on such fine temporal resolutions. Dominance

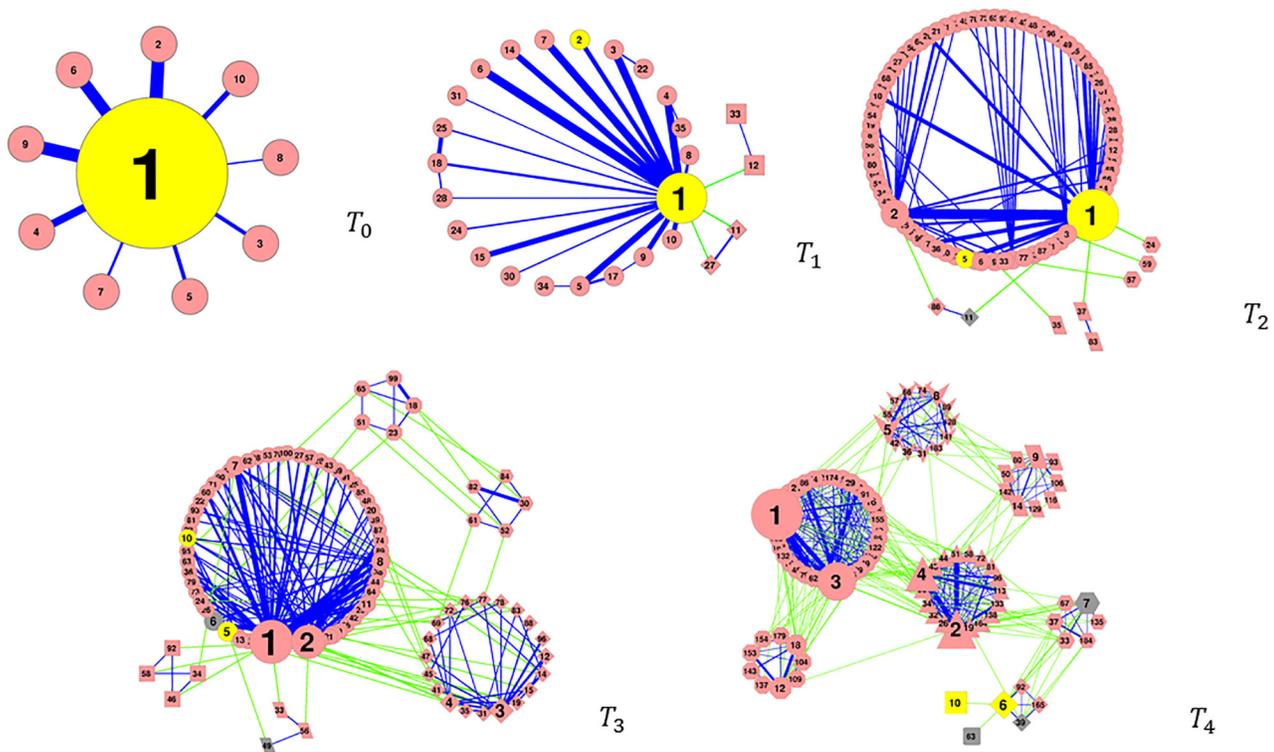


Fig 2. CS organization at temporal resolutions T_0 , T_1 , T_2 , T_3 and T_4 . Each line (edge) between the two defining nodes (representing CSs) indicates occurrence of direct transitions between these two CSs. Sizes of nodes are indicative of CS's statistical weight and thickness of lines are indicative of the number of direct transitions between these two CSs (but not strictly to the proportion). Blue lines represent transitions between clusters that merge to the same cluster on the preceding coarser temporal resolution, while green lines represent transitions between clusters that belong to different clusters on the preceding coarser time resolution. Yellow nodes harbor HEWL crystal structures in both single and complex forms, gray nodes harbor HEWL crystal structures in complexes with other proteins, and the remaining nodes in pink represent CSs that do not harbor any crystal structures. Only significant CSs are shown.

doi:10.1371/journal.pone.0129846.g002

by a structural ensemble that does not correspond to crystal structure(s) has been attributed to inaccuracy of force fields [9]. Nevertheless, we need to note that the simulation condition is quite different from crystallization conditions of any utilized crystal structures. Finally a given crystallization process is not guaranteed to capture the most populous solution conformational substate at arbitrarily given temporal resolutions. Therefore, the exact reason for mismatch between dominant simulated ensembles and crystal structural ensemble (CSE) at fine temporal resolutions is not clear. It is likely that all three factors contribute simultaneously. As shown in Fig 3a, At temporal resolution T_4 , clusters harboring CSE are observed to have significantly less inter-cluster connections when compared with other clusters that have similar statistical weight (i.e. neighbors on the X axis, inter-cluster connections normalized by statistical weight are shown in Fig 3b). It is noted that such “kinetic trap” property is only observed on relatively fine temporal resolutions (T_3 and T_4). At coarse temporal resolutions (T_1 and T_2), each major CSE seems to be “contaminated” by many non-CSE solution states on finer temporal resolutions (Fig. D in S1 File).

We analyzed distributions of crystal structures among the clusters established from MD trajectories at time resolutions T_1 through T_4 (see supporting text for details). Briefly, for backbone torsional transitions associated with various molecular interactions that were captured by crystal structures, their corresponding time scales were found to vary from $\sim ns$ to multiple μs .

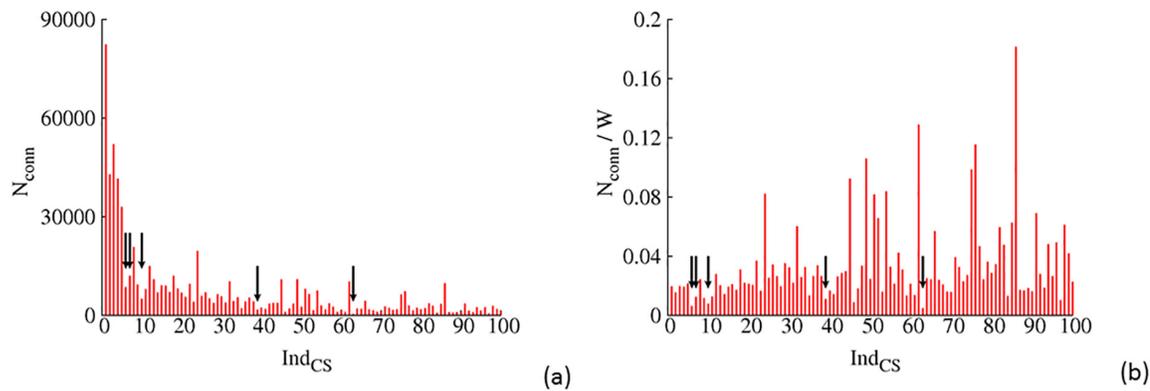


Fig 3. Extent of inter-cluster transitions for the largest 100 CSs at temporal resolution T_4 , CSs that harboring crystal structures are indicated by arrows. a) The total number of transitions (N_{conn}), b) upon normalization with the corresponding statistical weight (N_{conn}/W).

doi:10.1371/journal.pone.0129846.g003

Functionally relevant substrate-bound and inhibitor-bound structures, while sharing large hydrophobic solvent accessible surface area (hSASA) (Fig 4), locate in different clusters at temporal resolutions T_3 and T_4 . Site for change of backbone torsional states associated with molecular interactions do not necessarily co-localize with corresponding interaction residues. This observation is in agreement with the concept that all dynamic proteins are potentially allosteric [29]. All HEWL crystal structures that are in complexes with antibodies and a few other small proteins locate in significant clusters obtained from our simulations of free form HEWL.

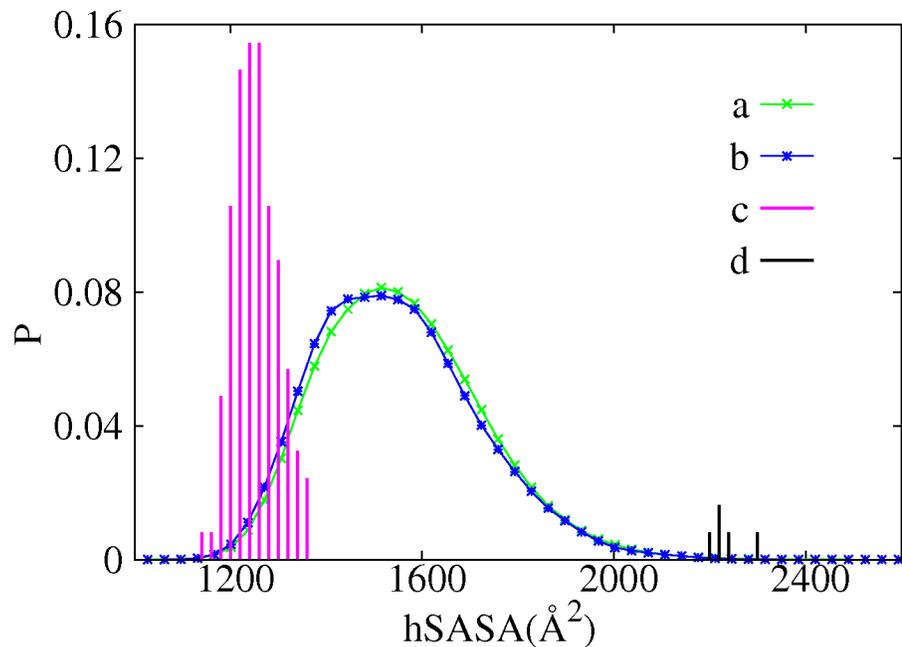


Fig 4. Probability distribution of hSASA for the crystal structure ensemble and MD simulation ensembles. Green line a): from trajectories that originate from crystal structures without sugar substrates/inhibitor bound; Blue line b): from trajectories that originate from crystal structures with sugar substrates/inhibitor bound; Purple impulses c): from crystal structure ensemble without sugar substrates/inhibitor bound; Black impulses d): from crystal structures with sugar substrates/inhibitor bound.

doi:10.1371/journal.pone.0129846.g004

Therefore, the conformational selection mechanism [30] seems to dominate for interactions of HEWL and its protein partners.

Comparison with mutant crystal structure ensemble

There are many crystal structures available in PDB for a wide variety of HEWL mutants (see Table C in [S1 File](#)). We were interested in examining how point mutations impact selection of potential CSE by crystallization. To this end, we performed the same clustering procedure with 29 mutant structures and listed the results in Table C in [S1 File](#). It is found that mutants crystal structures are located in the same clusters as the wild type ones under similar crystallization conditions. As shown in Fig. E in [S1 File](#), analyzed mutations are distributed across the whole protein. Therefore, at least for this set of mutants, single (and multiple) point mutations do not significantly influence the net result of CSE selection during crystallization. This observation suggests that for HEWL, crystallization conditions are more important than point mutations in selecting CSE. Therefore, caution has to be applied in explaining any structural change observed in mutant crystal structures under different crystallization conditions. In crystallization attempts, adjusting solution condition is a more widely utilized strategy than point mutations. Such practice is consistent with our current observation. The rationale might be that solution conditions directly impact many surface residues that are more relevant in crystal contacts, and consequently has a larger probability to result in significant change of interaction networks that determine selection of CSE in target protein molecules. This observation is likely to be true for other globular proteins.

Physical properties of major CS and transition state ensembles

To examine structural differences within each and among different clusters. We calculated the pair-wise root mean squared deviation (pWRMSD) based on backbone atoms, and the results are shown in [Fig 5a](#). At temporal resolution T_4 , inter-cluster structural differences as measured by pWRMSD are on average slightly larger than that within each cluster, as one would intuitively expect. However, distributions for inter-cluster and intra-cluster pWRMSD are both single-peaked and largely overlap. A few other molecular scale physical quantities, including the number of hydrogen bonds (nHB), the number of native contacts (nNC), radius of gyration (R_g), and $hSASA$ were calculated for major CSs at multiple temporal resolutions. As shown in [Fig 5\(b\)–5\(d\)](#) and [Fig. F in S1 File](#), CSs occupied by crystal structures exhibit significantly larger nHB , nNC and smaller $hSASA$ than other ones, while sharing similar average R_g .

We also calculated these quantities for transition state ensembles (between CSs). As shown by [Fig 6](#), At temporal resolution T_2 , transition state ensembles at various temporal resolutions exhibits larger $hSASA$, larger R_g , smaller nHB and nNC . Similar but less significant observations were made at other temporal resolutions (see [Fig. G in S1 File](#)). These observations suggest that transition structural states corresponding to change of backbone torsional states are more disordered and expanded. Disorder within native state of well-structured protein is not an established concept. Due to their small statistical significance in equilibrium conformational distributions, these transition state ensembles have trivial impact on thermodynamic properties. For the same reason, direct experimental comparison is extremely difficult at present. However, large-amplitude backbone dihedral rotations are important in protein conformational dynamics. Therefore, further investigation in this aspect is necessary.

Life time of transition state ensembles at multiple temporal resolutions

Life time for TSEs of protein folding have been investigated extensively by experimental, theoretical and computational studies and found to vary from multiple ps to hundreds of ns [31]–

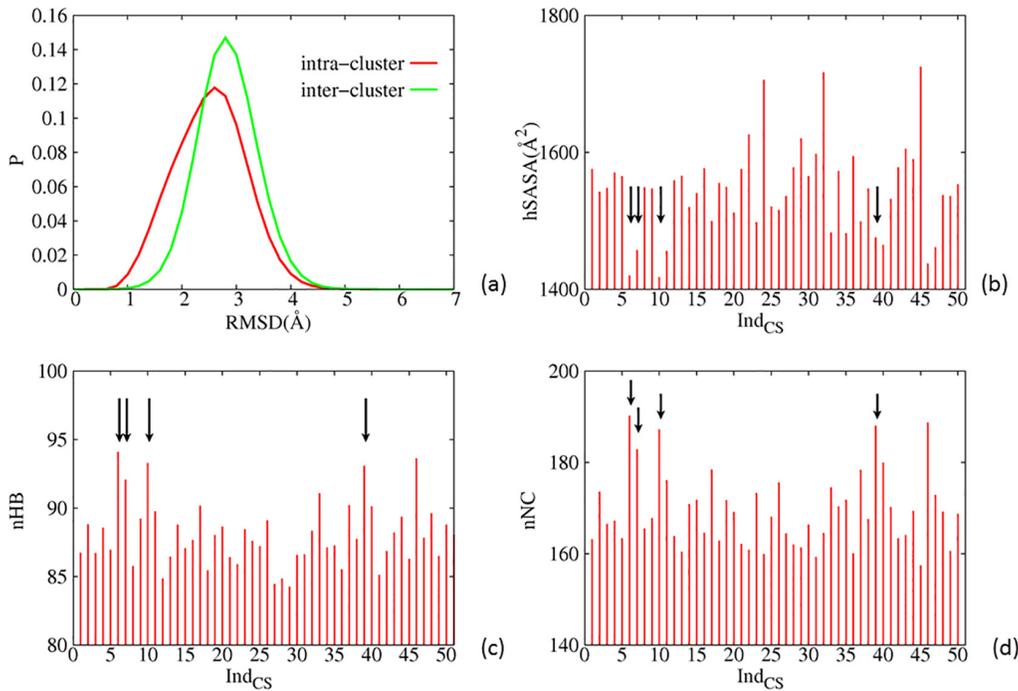


Fig 5. A few physical properties of major clusters at temporal resolution T_4 . a) Distributions of intra-cluster (red) and inter-cluster (green) pwRMSD. b) $hSASA$, c) nHB and d) nNC of the 50 largest clusters. Arrows in b), c) and d) indicate clusters harboring crystal structures.

doi:10.1371/journal.pone.0129846.g005

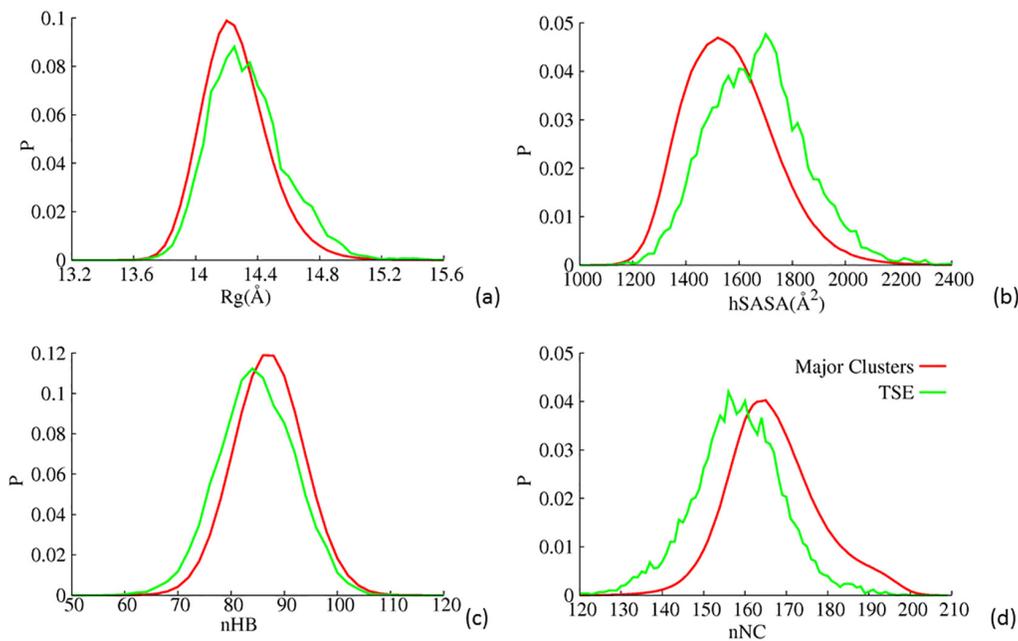


Fig 6. Distributions of a) R_g , b) $hSASA$, c) nHB and d) nNC for major clusters (the 50 largest ones) and TSEs at temporal resolution T_2 .

doi:10.1371/journal.pone.0129846.g006

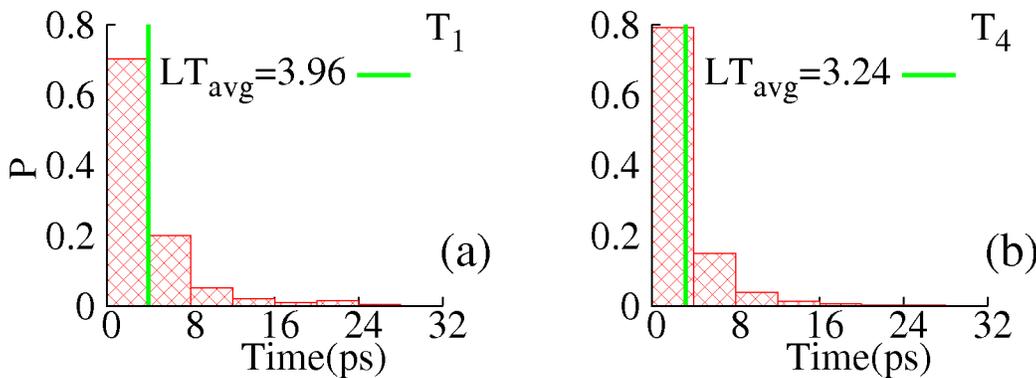


Fig 7. Distributions of TSE life times at temporal resolution a) T_1 and b) T_4 . The green impulses indicate estimated averages (LT_{avg}).

doi:10.1371/journal.pone.0129846.g007

[34]. No quantitative characterization of life time for TSE of backbone torsional transitions in native protein is available to the best of our knowledge, however. We calculated averages and distributions of life times for TSEs at various temporal resolutions and plotted the results in Fig 7. It is interesting that both distributions and averages of torsional transition times are similar for all backbone torsions that were used in clustering, regardless of the average waiting times, which are used to define time scales for corresponding torsional DOFs and range over five orders of magnitude. As cluster splitting backbone torsions at various temporal resolutions are distributed across the whole primary sequence (Table A in S1 File), this observation implies that life time of backbone torsional transition state is insensitive to local environment. While specific life times for transition state of backbone torsional DOFs might vary for different force fields, its observed invariance across primary sequences implies insensitivity to local environment, which is likely to be a general feature of proteins.

Side chain torsional degrees of freedom

Side chain torsional DOFs, together with backbone ones, are generally termed “soft” DOFs and have been demonstrated to be the major source of protein configurational entropy [35]. We calculated distributions of χ_1 for all rotatable side chains that have at least one heavy atom beyond C_β . It is found that each χ_1 significantly populate all three major rotameric states (Fig. H in S1 File). Waiting time for these interconversions ranging from a few nanoseconds to microseconds and beyond (See Fig. I in S1 File). Addition of χ_1 s in the clustering results in a dramatic increase of cluster number at each temporal resolution, and N_{CS} amounts to 7382009 at temporal resolution T_4 . Further addition of χ_2 increases N_{CS} to 38437034 at T_4 . These observations reflect significantly less correlations among side chains when compared with backbone dihedrals as calculated by us (Fig. J in S1 File) and reported elsewhere [36], and confirm the conclusion from an earlier study [37] that globular proteins have solid like backbone DOFs and liquid like side chain DOFs. These side chain based clusters are not helpful in understanding experimental observables and function of proteins since each single cluster has negligible statistical weight. Therefore, we did not perform further analysis on them. This observation by no means negates the importance of side chain flipping in function of some proteins (e.g. open/close of a channel) [38, 39], where it is the marginal probability of a specific one or a few side chain torsional state (s), not the probability of a unique side chain torsional state combination (which is always

negligible in statistical weight), that accounts for the corresponding structural, dynamical or functional importance.

Discussion

Selection of clustering criteria and methodology

In folding and design studies, RMSD from a specific native structure (usually a crystal structure) or pWRMSD is widely utilized for clustering and a number of accelerated methods are developed for this purpose [40–42]. This is quite effective as structural differences between native ensemble and various unfolded ensembles are significant. Another type of widely used quantities are principal components (PC) obtained from diagonalizing the covariance matrix [43]. Principal component analysis (PCA) has been shown to be effective in reducing dimension of protein FEL as first few PCs usually dominate the whole FEL that is explored by simulations on the order of hundreds of nanoseconds [16]. This is not necessarily true for very large number of snapshots exhibiting strong structural diversity. When PCA is applied to snapshots from single or a few 100–*ns* HEWL trajectories, in agreement with many previous studies, the first few PCs dominate the whole FEL. However, when snapshots from hundreds of 100–*ns* trajectories were used, strong structural diversity renders the first few PCs much less dominant (Fig. K in S1 File). It is important to note that PCA analysis is based on the structural covariance matrix, which does not have any time scale related information. The dominating eigenvectors corresponding to collective motions that are of the largest spatial magnitude, not necessarily the longest time scale. Therefore, both RMSD and PCs (from PCA) are not in line with our major goal of analyzing the explored conformational space at multiple temporal resolutions.

Markov state models (MSM) is a powerful technique with explicit temporal resolution consideration, and has been successfully applied to analyze major functionally important metastable states for many proteins [44]. There have been significant improvement of kinetic network models, such as the super-level-set hierarchical clustering [45] and Hierarchical Nyström methods [46] for multiple resolution analysis of bimolecular dynamics, since earlier days of MSM [47]. We initially chose to use distributions of backbone dihedrals for clustering based on the well-acknowledged importance of these torsional DOFs for protein conformational transitions, the deterministic clustering results, and explicit consideration of temporal resolution. When combined with the radix sort algorithm [48] and bit-encoded torsional states, the clustering process may be accomplished with trivial computational cost. The above discussed results indicate that clustering based on backbone torsional states divide crystal structure ensemble into physically meaningful sub-clusters. Additionally, each backbone torsional DOF is associated with a specific residue, potentially rendering easy integration of conformational analysis with sequence analysis and mutation experiments. Explicit knowledge of time scales and distributions of individual backbone torsions in specific environment may be utilized by machine learning methods for prediction of CS distribution and dynamics. Therefore, given its low computational cost, conceptual simplicity and practical utility, backbone dihedral distribution based clustering is a useful approach for conformational analysis of proteins. Apparently, MSM is a more theoretically advanced methods for multiple-resolution analysis of MD trajectory sets. However, specifying number of clusters for different hierarchies, which we do not have much information *a priori*, becomes a challenge. Additionally, the clustering itself become computationally intensive for large number of snapshots as in our case. It would be interesting to have a detailed comparison between our backbone dihedral based clustering and MSM analysis in the future.

The significance of MD simulation results and the crystal structural ensemble

While admitting that each current force field parameter set has its own limitations, it is certainly true that they describe very realistic models of proteins. The significance of simple lattice model [49–51] and Go model [52, 53] studies can not be over emphasized in developing protein folding theories. Recent atomistic simulations of the folding of many small globular proteins [24], while greatly enriched original simple models, proved rather than negated the significance of their major conclusions. At microsecond temporal resolution (T_2), dozens of transitions are observed between the dominating cluster that harboring nearly all crystal structures and other small clusters, suggesting that the force fields is effective in locating the major native structural states. Therefore we believe that our results, despite the fact that it does not strictly reflecting the global FEL of HEWL under physiological conditions, give a highly realistic model, and the conclusions regarding the topology and organization of CS are likely to be representative for many globular proteins.

Crystal structures of proteins, despite their being obtained from non-physiological conditions, have beautifully rationalized numerous critical molecular biological processes (e.g. central dogma related processes, energy production, signal transduction). This is consistent with our remarkable observation that 114 out of 120 crystal structures, which were obtained under widely different conditions, are located within a central $2 \mu\text{s}$ bottom region of the native HEWL FEL explored by 0.2ms MD simulations. Five out of the remaining six crystal structures are obtained under extreme conditions such as low hydration and high sodium nitrate concentration (see supporting information for details). Our simulation results, together with the large crystal structural ensemble, suggest that HEWL is likely to have only one narrow major native FEL well (within a few microseconds) that harbors structural states to bind substrates and interact with other proteins. Since force fields are always defective to some extent, very long trajectories might get lost in unrealistic conformational states, running multiple moderately long trajectories in the vicinity of high resolution experimental structures may alleviate this artifact to some degree. Therefore, to sample conformational space of proteins with similar type of native FEL in the vicinity of a crystal structure, utilizing multiple independent microseconds-long MD trajectories might be a better strategy than running a single milli-second or longer MD trajectory.

Conclusions and the prospects

In this study we performed a hierarchical conformational analysis for HEWL, a typical globular protein with sufficient structural complexity, at multiple temporal resolutions ranging from ns to μs . We observed hub-like topology at microsecond and coarser temporal resolutions, and increasingly diversified structural states and more connectivity among CS at increasingly finer temporal resolutions. Various molecular interactions captured in CSE were found to associate with CS transitions covering time scales from nanoseconds to microseconds. CSE for HEWL are found to be associated with kinetic-trap clusters at $\sim 10\text{--}100ns$ temporal resolution. The number of CS that have non-negligible statistical weight is quite limited, even at ns temporal resolution. These observations suggest that to study CS of native globular proteins at temporal resolutions ranging from ns to μs is both important and potentially feasible for prediction of molecular interactions. However, our study is limited to one specific protein. The apparent immediate questions needs to be answered are: i) How is distribution and organization of CS related to specific protein size, folds and sequences; ii) What are the roles of hierarchical CS in protein-protein and protein-ligand interactions. We are working on a few more proteins with

different sizes and folds to further increase our understanding in these directions. We hope that this work will stimulate more investigations in this line.

Methods

Molecular dynamics simulations

MD simulation systems were setup and equilibrated with VMD [54] and NAMD [55]. Production runs were performed using ACEMD [56]. 120 crystal structures of HEWL, including 44 in protein-protein-interaction complexes and 76 in free form (or with bound small molecules), were taken from PDB (www.pdb.org). Each structure is solvated with 6575 TIP3P water molecules and neutralized with 5 Na^+ and 13 Cl^- ions, resulting in a system with 21703 atoms. The 120 HEWL simulation systems were first equilibrated for 200ps in NVT ensemble and for 1ns in NPT ensemble, both harmonic restraint on protein heavy atoms. After that, restraints were released and 3ns of NPT run were performed to obtain a proper volume (a cubic box with 59.582 Å sides) that is used in the following NVT production runs. Starting configurations for production NVT runs were selected from the last 1ns of the previous NPT run with the criteria of having right box size (within 0.01Å). CHARMM22 (with CMAP) force fields were utilized with a 9Å non-bonded cutoff distance in production runs. The simulation time step is 2fs for equilibration with NAMD and 4fs with hydrogen repartition for production run with ACEMD. Electrostatics were treated by PME with grid size being approximately 1Å. 20 independent trajectories were started from each equilibrated PDB structure by using 20 different random number seeds for initial atomic velocity assignment, resulting in 2400 initial trajectories. Trajectories stopped due to machine failure (mainly writing error due to full disks, GPU memory error and operating system failing) were discarded. 2000 100 – ns trajectories, with collectively 50,000,000 snapshots, were utilized for final analysis.

Clustering of structural and transition state ensembles at various temporal resolutions

Backbone dihedral distributions and torsional states. Distribution of each backbone dihedral angle (ϕ , ψ) is constructed by using histograms with bin size being 1°. To perform clustering, local minima are found for each backbone dihedral distribution. Whenever the angle distance between two local minima of a given dihedral is smaller than 60, the minimum with smaller probability is taken as an effective. After this initial filtering, a backbone dihedral is selected as a potential clustering dihedral if two or more effective local minima exist. Each region between two neighboring local minima is defined as a torsional state for that torsional DOF. As dihedral angles are cyclic variables, the number of state is equal to the number of local minima utilized for splitting corresponding dihedrals. We found that 127 backbone dihedrals (out of the total of 256) exhibit two-peak distributions and are defined as two-state torsional DOFs that were used in clustering.

Time scales. The time scale for each specific two-state backbone dihedral is defined by the average waiting time between two transition events regardless of its specific directions and routes (For circular two-state system with state A , B and minima MIN_A , MIN_B , there are two possible directions and four possible routes, $A \rightarrow MIN_A \rightarrow B$, $A \rightarrow MIN_B \rightarrow B$, $B \rightarrow MIN_A \rightarrow A$, $B \rightarrow MIN_B \rightarrow A$). Specifically, within our collective trajectory time of 200μs, if N transitions (N_{trans}) happened between the two states of an given dihedral, then the average waiting time is $t_w = \frac{200}{N_{trans}} \mu s$, which is used to define time scale of the corresponding torsional DOF.

Clustering at five temporal resolutions. To establish backbone based hierarchical clustering, we first divided all 127 two-state backbone dihedrals into five groups corresponding to five

different temporal resolutions T_0 ($20\mu s < t_w < 200\mu s$), T_1 ($2\mu s < t_w < 20\mu s$), T_2 ($200ns < t_w < 2\mu s$), T_3 ($20ns < t_w < 200ns$) and T_4 ($2ns < t_w < 20ns$). Two structures are assigned to the same cluster only if for each clustering participating torsional DOF, both structures are in the same torsional state. When performing clustering at a given temporal resolution, two-state backbone dihedrals with shorter time scales are treated as single-state DOF and are excluded. For example, when clustering at temporal resolution T_2 , we use backbone dihedrals that have time scales within the range of T_0 , T_1 and T_2 , but ignore those with time scales within the range of T_3 and T_4 . Due to limited statistics, we did not perform detailed analysis at temporal resolution T_0 . Backbone torsional DOFs utilized for each temporal resolution is listed in Table A in [S1 File](#).

Transition state ensembles. First we define transition state ensemble for effective local minimum of all two-state backbone torsional DOFs (ϕ and ψ) as follows. At temporal resolutions T_3 and T_4 , a 5 degree region (2.5 degree to each side of an effective local minimum) is defined as the transition state. At temporal resolutions T_0 , T_1 and T_2 , for a given local minimum of a specific torsional distribution, number of snapshots in 2.5 degree bins to each side of the minimum are assigned to $N_{+1}, N_{+2}, N_{+3}, \dots$ and $N_{-1}, N_{-2}, N_{-3}, \dots$. If $N_{\pm(i+2)} - N_{\pm(i+1)} \gg N_{\pm(i+1)} - N_{\pm i}$ ($i = 1, 2, 3, \dots$), then bins $1, \dots, i + 1$ are defined as the transition state region of the corresponding local minimum. Due to finite sampling, limited connectivity on each level of hierarchical conformational space, and correlations between (among) different backbone torsional DOF, local minima in distributions of each two-state torsional DOF do not necessarily correspond to the transition ensembles between different clusters. Only when transition state of a specific DOF coincide with transition between two clusters at a given temporal resolution, it is taken as transition state of corresponding clusters and snapshots within that state are counted as transition state structures.

Principal component analysis. Dihedral principal component analysis (dPCA) is utilized in this study. To tackle the circular property of dihedral angles, we used the following trigonometric functions to represent backbone dihedrals (ϕ and ψ) [57]. $q_{2i-1} = \cos \phi_i$, $q_{2i} = \sin \psi_i$, with i running through the number of residues. Subsequently, covariance matrices $\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle$ are constructed for selected (or all) trajectories and diagonalized to obtain principal components. We also performed PCA analysis based on cartesian coordinates for protein backbone atoms.

Calculation of various physical properties

Radius of gyration. Radius of gyration for each snapshot is calculated as $r_g = \sqrt{\frac{\sum_{i=1}^{n_{atom}} m_i r_i^2}{\sum_{i=1}^{n_{atom}} m_i}}$, with m_i and r_i being the mass and the distance to the molecular center of mass for atom i respectively, n_{atom} is the total number of atoms in HEWL.

Native contacts. A residue contact is defined for two non-sequential residues with C_α distance being smaller than 6.5\AA . Residue contacts that are shared by two thirds (80) or more of the 120 crystal structures utilized in this study are defined as native contacts.

Hydrophobic solvent accessible surface area (hSASA). A 1.4\AA -diameter sphere is used to probe HEWL surface of hydrophobic residues (as defined by VMD) with VMD [54].

Hydrogen bonds. Existence of a hydrogen bond is defined by a distance between a donor and an acceptor being smaller than 3.5\AA (cutoff distances ranging from 3.0\AA to 3.5\AA gives similar ordering among different clusters when the number of hydrogen bonds is compared), and by the corresponding “donor \rightarrow H \rightarrow acceptor” bend angle being larger than 130 .

Supporting Information

S1 File.

(PDF)

Acknowledgments

This research was partially funded by a start-up fund from Jilin University and by National Natural Science Foundation of China under grant number 31270758. Computational resources were partially supported by High Performance Computing Center of Jilin University, China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: PT. Performed the experiments: KW. Analyzed the data: KW SL. Wrote the paper: PT.

References

1. Boehr DD, McElheny D, Dyson HJ, Wright PE. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*. 2006; 313:1638–42. PMID: [16973882](#)
2. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*. 2007; 449:1078–1082. doi: [10.1038/nature06232](#) PMID: [17960247](#)
3. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, et al. Intrinsic motions along an enzymatic reaction trajectory. *nature*. 2007 12; 450:838–44. doi: [10.1038/nature06410](#) PMID: [18026086](#)
4. Biehl R, Hoffmann B, Monkenbusch M, Falus P, Préost S, Merkel R, et al. Direct Observation of Correlated Interdomain Motion in Alcohol Dehydrogenase. *Phys Rev Lett*. 2008 Sep; 101:138102. doi: [10.1103/PhysRevLett.101.138102](#) PMID: [18851497](#)
5. Hodges C, Bintu L, Lubkowska L, Kashlev M, Bustamante C. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science*. 2009; 325:626–8. doi: [10.1126/science.1172926](#) PMID: [19644123](#)
6. Dong M, Husale S, Sahin O. Determination of protein structural flexibility by microsecond force spectroscopy. *Nat Nanotechnol*. 2009; 4:514–7. doi: [10.1038/nnano.2009.156](#) PMID: [19662014](#)
7. Vreede J, Juraszek J, Bolhuis PG. Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein. *Proc Natl Acad Sci USA*. 2010; 107:2391–6. doi: [10.1073/pnas.0908754107](#)
8. Inoue R, Biehl R, Rosenkranz T, Fitter J, Monkenbusch M, Radulescu a, et al. Large domain fluctuations on 50-ns timescale enable catalytic activity in phosphoglycerate kinase. *Biophysical journal*. 2010 Oct; 99(7):2309–17. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3042550&tool=pmcentrez&rendertype=abstract>. doi: [10.1016/j.bpj.2010.08.017](#) PMID: [20923666](#)
9. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-level characterization of the structural dynamics of proteins. *Science (New York, NY)*. 2010 Oct; 330(6002):341–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20947758>. doi: [10.1126/science.1187409](#)
10. Wang Y, Tang C, Wang E, Wang J. Exploration of multi-state conformational dynamics and underlying global functional landscape of maltose binding protein. *PLoS computational biology*. 2012 Jan; 8(4): e1002471. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3330084&tool=pmcentrez&rendertype=abstract>. doi: [10.1371/journal.pcbi.1002471](#) PMID: [22532792](#)
11. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett*. 1996; 77:1905–08. doi: [10.1103/PhysRevLett.77.1905](#) PMID: [10063201](#)
12. Gur M, Zomot E, Bahar I. Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions. *The Journal of chemical physics*. 2013 Sep; 139(12):121912. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3739829&tool=pmcentrez&rendertype=abstract>. doi: [10.1063/1.4816375](#) PMID: [24089724](#)
13. Sinko W, Lindert S, McCammon JA. Accounting for Receptor Flexibility and Enhanced Sampling Methods in Computer-Aided Drug Design. *Chemical Biology & Drug Design*. 2013; 81(1):41–49. doi: [10.1111/cbdd.12051](#)

14. Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC. Dynamics of ligand binding to myoglobin. *Biochemistry*. 1975 Dec; 14(24):5355–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1191643>. doi: [10.1021/bi00695a021](https://doi.org/10.1021/bi00695a021) PMID: [1191643](https://pubmed.ncbi.nlm.nih.gov/1191643/)
15. Frauenfelder H, Sligar S, Wolynes P. The energy landscapes and motions of proteins. *Science*. 1991; 254(5038):1598–1603. Available from: <http://www.sciencemag.org/content/254/5038/1598.abstract>. doi: [10.1126/science.1749933](https://doi.org/10.1126/science.1749933) PMID: [1749933](https://pubmed.ncbi.nlm.nih.gov/1749933/)
16. Zhuravlev PI, Papoian Ga. Functional versus folding landscapes: the same yet different. *Current opinion in structural biology*. 2010 Feb; 20(1):16–22. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20102791>. doi: [10.1016/j.sbi.2009.12.010](https://doi.org/10.1016/j.sbi.2009.12.010) PMID: [20102791](https://pubmed.ncbi.nlm.nih.gov/20102791/)
17. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. *Proteins: Structure, Function, and Bioinformatics*. 2010; 78(15):3073–3084. doi: [10.1002/prot.22818](https://doi.org/10.1002/prot.22818)
18. Gallicchio E, Levy RM. Prediction of SAMPL3 host-guest affinities with the binding energy distribution analysis method (BEDAM). *Journal of computer-aided molecular design*. 2012 May;(5):505–516. doi: [10.1007/s10822-012-9552-3](https://doi.org/10.1007/s10822-012-9552-3) PMID: [22354755](https://pubmed.ncbi.nlm.nih.gov/22354755/)
19. Pons C, Fenwick RB, Esteban-Martín S, Salvatella X, Fernandez-Rrecio J. Validated Conformational Ensembles Are Key for the Successful Prediction of Protein Complexes. *Journal of Chemical Theory and Computation*. 2013; 9(3):1830–1837. doi: [10.1021/ct300990h](https://doi.org/10.1021/ct300990h)
20. Lexa KW, Carlson HA. Protein flexibility in docking and surface mapping. *Quarterly Reviews of Biophysics*. 2012 8; 45:301–343. doi: [10.1017/S0033583512000066](https://doi.org/10.1017/S0033583512000066) PMID: [22569329](https://pubmed.ncbi.nlm.nih.gov/22569329/)
21. Lindorff-larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science*. 2011; 334:517–520. doi: [10.1126/science.1208351](https://doi.org/10.1126/science.1208351) PMID: [22034434](https://pubmed.ncbi.nlm.nih.gov/22034434/)
22. Lane TJ, Bowman GR, Beauchamp K, Voelz Va, Pande VS. Markov state model reveals folding and functional dynamics in ultra-long MD trajectories. *Journal of the American Chemical Society*. 2011 Nov; 133(45):18413–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3227799&tool=pmcentrez&rendertype=abstract>. doi: [10.1021/ja207470h](https://doi.org/10.1021/ja207470h) PMID: [21988563](https://pubmed.ncbi.nlm.nih.gov/21988563/)
23. Dickson A, Brooks CL. Native states of fast-folding proteins are kinetic traps. *Journal of the American Chemical Society*. 2013 Mar; 135(12):4729–34. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23458553>. doi: [10.1021/ja311077u](https://doi.org/10.1021/ja311077u) PMID: [23458553](https://pubmed.ncbi.nlm.nih.gov/23458553/)
24. Lane TJ, Shukla D, Beauchamp Ka, Pande VS. To milliseconds and beyond: challenges in the simulation of protein folding. *Current opinion in structural biology*. 2013 Feb; 23(1):58–65. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23237705>. doi: [10.1016/j.sbi.2012.11.002](https://doi.org/10.1016/j.sbi.2012.11.002) PMID: [23237705](https://pubmed.ncbi.nlm.nih.gov/23237705/)
25. Sosnick TR, Hinshaw JR. *Biochemistry*. How proteins fold. *Science (New York, NY)*. 2011 Oct; 334(6055):464–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22034424>. doi: [10.1126/science.1214018](https://doi.org/10.1126/science.1214018)
26. Chung HS, McHale K, Louis JM, Eaton Wa. Single-molecule fluorescence experiments determine protein folding transition path times. *Science (New York, NY)*. 2012 Feb; 335(6071):981–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22363011>. doi: [10.1126/science.1215768](https://doi.org/10.1126/science.1215768)
27. Vymětal J, Vondrášek J. Critical Assessment of Current Force Fields. Short Peptide Test Case. *Journal of Chemical Theory and Computation*. 2013; 9(1):441–451. doi: [10.1021/ct300794a](https://doi.org/10.1021/ct300794a)
28. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. Systematic validation of protein force fields against experimental data. *PLoS one*. 2012 Jan; 7(2):e32131. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3285199&tool=pmcentrez&rendertype=abstract>. doi: [10.1371/journal.pone.0032131](https://doi.org/10.1371/journal.pone.0032131) PMID: [22384157](https://pubmed.ncbi.nlm.nih.gov/22384157/)
29. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function, and Bioinformatics*. 2004; 57(3):433–443. Available from: <http://dx.doi.org/10.1002/prot.20232>.
30. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences*. 2010; 35(10):539–546. Available from: <http://www.sciencedirect.com/science/article/pii/S0968000410000885>. doi: [10.1016/j.tibs.2010.04.009](https://doi.org/10.1016/j.tibs.2010.04.009) PMID: [20541943](https://pubmed.ncbi.nlm.nih.gov/20541943/)
31. Bolhuis PG, Chandler D, Dellago C, Geissler PL. TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annual Review of Physical Chemistry*. 2002; 53(1):291–318. doi: [10.1146/annurev.physchem.53.082301.113146](https://doi.org/10.1146/annurev.physchem.53.082301.113146) PMID: [11972010](https://pubmed.ncbi.nlm.nih.gov/11972010/)
32. Best RB, Hummer G. Diffusive Model of Protein Folding Dynamics with Kramers Turnover in Rate. *Phys Rev Lett*. 2006 Jun; 96:228104. doi: [10.1103/PhysRevLett.96.228104](https://doi.org/10.1103/PhysRevLett.96.228104) PMID: [16803349](https://pubmed.ncbi.nlm.nih.gov/16803349/)
33. Chung HS, Louis JM, Eaton WA. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proceedings of the National Academy of Sciences*. 2009; 106(29):11837–11844. Available from: <http://www.pnas.org/content/106/29/11837.abstract>. doi: [10.1073/pnas.0901178106](https://doi.org/10.1073/pnas.0901178106)

34. Best RB, Mittal J. Microscopic events in -hairpin folding from alternative unfolded ensembles. *Proceedings of the National Academy of Sciences*. 2011; 108(27):11087–11092. Available from: <http://www.pnas.org/content/108/27/11087.abstract>. doi: [10.1073/pnas.1016685108](https://doi.org/10.1073/pnas.1016685108)
35. Li DW, Brüschweiler R. In silico relationship between configurational entropy and soft degrees of freedom in proteins and peptides. *Phys Rev Lett*. 2009 Mar; 102:118108. doi: [10.1103/PhysRevLett.102.118108](https://doi.org/10.1103/PhysRevLett.102.118108) PMID: [19392246](https://pubmed.ncbi.nlm.nih.gov/19392246/)
36. Fenwick RB, Esteban-Martín S, Richter B, Lee D, Walter KFA, Milovanovic D, et al. Weak Long-Range Correlated Motions in a Surface Patch of Ubiquitin Involved in Molecular Recognition. *Journal of the American Chemical Society*. 2011; 133(27):10336–10339. doi: [10.1021/ja200461n](https://doi.org/10.1021/ja200461n) PMID: [21634390](https://pubmed.ncbi.nlm.nih.gov/21634390/)
37. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. *Nature*. 2005; 433:128–132. Available from: http://www.nature.com/nature/journal/v433/n7022/supinfo/nature03199_S1.html. doi: [10.1038/nature03199](https://doi.org/10.1038/nature03199) PMID: [15650731](https://pubmed.ncbi.nlm.nih.gov/15650731/)
38. Winn PJ, Lüdemann SK, Gauges R, Lounnas V, Wade RC. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proceedings of the National Academy of Sciences*. 2002; 99(8):5361–5366. Available from: <http://www.pnas.org/content/99/8/5361.abstract>. doi: [10.1073/pnas.082522999](https://doi.org/10.1073/pnas.082522999)
39. Gora A, Brezovsky J, Damborsky J. Gates of Enzymes. *Chemical Reviews*. 2013; 113(8):5871–5923. doi: [10.1021/cr300384w](https://doi.org/10.1021/cr300384w) PMID: [23617803](https://pubmed.ncbi.nlm.nih.gov/23617803/)
40. Hildebrandt AK, Dietzen M, Lengauer T, Lenhof HP, Althaus E, Hildebrandt A. Efficient computation of root mean square deviations under rigid transformations. *Journal of Computational Chemistry*. 2014; 35(10):765–771. doi: [10.1002/jcc.23513](https://doi.org/10.1002/jcc.23513)
41. Magis C, Di Tommaso P, Notredame C. T-RMSD: a web server for automated fine-grained protein structural classification. *Nucleic Acids Research*. 2013; 41(W1):W358–W362. Available from: <http://nar.oxfordjournals.org/content/41/W1/W358.abstract>. doi: [10.1093/nar/gkt383](https://doi.org/10.1093/nar/gkt383) PMID: [23716642](https://pubmed.ncbi.nlm.nih.gov/23716642/)
42. Hung LH, Guerquin M, Samudrala R. GPU-Q-J, a fast method for calculating root mean square deviation (RMSD) after optimal superposition. *BMC Research Notes*. 2011; 4(1):97. doi: [10.1186/1756-0500-4-97](https://doi.org/10.1186/1756-0500-4-97) PMID: [21453553](https://pubmed.ncbi.nlm.nih.gov/21453553/)
43. Zhou X, Chou J, Wong S. Protein structure similarity from principle component correlation analysis. *BMC Bioinformatics*. 2006; 7(1):40. doi: [10.1186/1471-2105-7-40](https://doi.org/10.1186/1471-2105-7-40) PMID: [16436213](https://pubmed.ncbi.nlm.nih.gov/16436213/)
44. Chodera JD, Noé F. Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology*. 2014; 25(0):135–144. Theory and simulation / Macromolecular machines. Available from: <http://www.sciencedirect.com/science/article/pii/S0959440X14000426>. doi: [10.1016/j.sbi.2014.04.002](https://doi.org/10.1016/j.sbi.2014.04.002) PMID: [24836551](https://pubmed.ncbi.nlm.nih.gov/24836551/)
45. Huang X, Yao Y, Bowman GR, Sun J, Guibas LJ, Carlsson G, et al. Constructing multi-resolution markov state models (msms) to elucidate RNA hairpin folding mechanisms. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*. 2010;p. 228–239. PMID: [19908375](https://pubmed.ncbi.nlm.nih.gov/19908375/)
46. Yao Y, Cui RZ, Bowman GR, Silva DA, Sun J, Huang X. Hierarchical Nystrom methods for constructing Markov state models for conformational dynamics. *The Journal of Chemical Physics*. 2013; 138(17):- . doi: [10.1063/1.4802007](https://doi.org/10.1063/1.4802007)
47. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *The Journal of Chemical Physics*. 2007; 126(15):- . doi: [10.1063/1.2714538](https://doi.org/10.1063/1.2714538) PMID: [17461665](https://pubmed.ncbi.nlm.nih.gov/17461665/)
48. Cormen TH, Leiserson CE, Rivest RL, Stein C. *Introduction to Algorithms*. 3rd ed. MIT Press and McGraw-Hill; 2009.
49. Sbreveali A, Shakhnovich E, Karplus M. How does a protein fold? *nature*. 1995 5; 369:248–251. Available from: <http://dx.doi.org/10.1038/369248a0>.
50. Mirny L, Shakhnovich E. PROTEIN FOLDING THEORY: From Lattice to All-Atom Models. *Annual Review of Biophysics and Biomolecular Structure*. 2001; 30(1):361–396. doi: [10.1146/annurev.biophys.30.1.361](https://doi.org/10.1146/annurev.biophys.30.1.361) PMID: [11340064](https://pubmed.ncbi.nlm.nih.gov/11340064/)
51. Bonneau R, Baker D. AB INITIO PROTEIN STRUCTURE PREDICTION: Progress and Prospects. *Annual Review of Biophysics and Biomolecular Structure*. 2001; 30(1):173–189. doi: [10.1146/annurev.biophys.30.1.173](https://doi.org/10.1146/annurev.biophys.30.1.173) PMID: [11340057](https://pubmed.ncbi.nlm.nih.gov/11340057/)
52. Nobuhiro GŌ, Abe H. Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers*. 1981; 20(5):991–1011. doi: [10.1002/bip.1981.360200511](https://doi.org/10.1002/bip.1981.360200511)
53. Abe H, Nobuhiro GŌ. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. *Biopolymers*. 1981; 20(5):1013–1031. doi: [10.1002/bip.1981.360200512](https://doi.org/10.1002/bip.1981.360200512) PMID: [7225529](https://pubmed.ncbi.nlm.nih.gov/7225529/)

54. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996; 14(1):33–38. doi: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: [8744570](https://pubmed.ncbi.nlm.nih.gov/8744570/)
55. Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K. Accelerating molecular modeling applications with graphics processors. *J Comput Chem*. 2007; 28:2618–2640. doi: [10.1002/jcc.20829](https://doi.org/10.1002/jcc.20829) PMID: [17894371](https://pubmed.ncbi.nlm.nih.gov/17894371/)
56. Harvey MJ, Giupponi G, Gabritiis GD. ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput*. 2009; 5:1632–9. doi: [10.1021/ct9000685](https://doi.org/10.1021/ct9000685)
57. Mu Y, Nguyen PH, Stock G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins: Structure, Function, and Bioinformatics*. 2005; 58(1):45–52. doi: [10.1002/prot.20310](https://doi.org/10.1002/prot.20310)