Published in final edited form as:
J Phys Chem Lett. 2012 May 3; 3(9): 1117–1123. doi:10.1021/jz300017c.

Structural Characterization of λ-Repressor Folding from All-Atom Molecular Dynamics Simulations

Yanxin Liu[†,‡], **Johan Strümpfer**[†,¶], **Peter L. Freddolino**[†,¶,||], **Martin Gruebele**[*,†,‡,¶,§], and **Klaus Schulten**[*,†,‡,¶]
Beckman Institute, Department of Physics, Center for the Physics of Living Cells, Center for Biophysics and Computational Biology, and Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Abstract

The five-helix bundle λ-repressor fragment is a fast-folding protein. A length of 80 amino acid residues puts it on the large end among all known microsecond folders and its size poses a computational challenge for molecular dynamics (MD) studies. We simulated the folding of a novel λ-repressor fast-folding mutant (λ-HG) in explicit solvent using an all-atom description. By means of a recently developed tempering method, we observed reversible folding and unfolding of λ-repressor in a 10-microsecond trajectory. The folding kinetics was also investigated through a set of MD simulations run at different temperatures that together covered more than 125 microseconds. The protein was seen to fold into a native-like topology at intermediate temperature and a slow-folding pathway was identified. The simulations suggest new experimental observables for better monitoring the folding process, and a novel mutation expected to accelerate λ-repressor folding.

Keywords

enhanced sampling; non-two state process; downhill folding; long time scale simulation; folding pathway

Considerable progress has been made towards understanding protein folding,[1,2] particularly for small fast-folding proteins.[3] The latter usually have under 100 amino acid residues and folding times are less then 500 μs. The small size and short folding times of such proteins make them good candidates for study through molecular dynamics (MD) simulation,[4] permitting the use of all-atom models in explicit solvent.[5–8] The time scale covered by such all-atom MD simulations increased from microsecond[9] to millisecond[10] in the last 15 years. Although computationally expensive, MD simulations provide valuable atom-resolution information on the folding mechanism and transition states.

[*]To whom correspondence should be addressed: mgruebel@illinois.edu; kschulte@ks.uiuc.edu, Phone: 217-244-1604. Fax: 217-244-6078.
[†]Beckman Institute
[‡]Department of Physics, Center for the Physics of Living Cells
[¶]Center for Biophysics and Computational Biology
[§]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[||]Current address: Joint Center for Systems Biology, Columbia University, New York, NY 10032, USA

Supporting Information Available
Detailed description of computational methods, four movies generated from simulation trajectories, and five figures. This material is available free of charge via the Internet at http:/pubs.acs.org/.

In addition to conventional MD simulation, enhanced sampling methods have been widely used in studying protein folding. A well-known example of such methods is replica exchange molecular dynamics (REMD).[11–14] However, the number of replicas needed increases as the total number of degrees of freedom in REMD simulations,[15] making the method impractical for large systems in explicit solvent. Zhang and Ma proposed a single copy tempering method to overcome this limitation of REMD and successfully applied their method to the folding of small model proteins.[16,17] In principle, the single copy tempering method is also suitable for large proteins.

The smallest fast folders (~35 residues) have already served as succesful model systems for both force field development and protein folding studies.[6,18–20] At the large end of the spectrum of fast-folding proteins is the λ-repressor fragment,[21,22] a five-helix bundle protein that consists of about 80 amino acid residues. A large number of mutants exist for the λ-repressor that cover a wide range of stabilities and folding rates.[23] The wild type and mutants of λ-repressor were studied both theoretically[24,25] and computationally using both a Gō-like potential [26,27] and an implicit solvent model.[28] Millisecond time-scale equilibrium MD simulations of λ-repressor in explicit solvent were achieved only recently.[8,29] Bowman et al. constructed folding trajectories for a fast-folding mutant of λ-repressor (D14A) based on Markov state models and suggested a slow folding phase on the time scale of 10 ms.[29] The same mutant was also simulated on a special purpose computer by Lindorff-Larsen et al., in which case an average folding time of 49 μs was observed.[8] Experiments suggest that λ-D14A does not have a ms phase, but other mutants can sample trap states that require ms to recover to the native state.[30]

We report here a series of all-atom MD simulations in explicit solvent on a novel fast-folding mutant of λ-repressor, called λ-HG.[31] The λ-HG contains mutations Y22W, Q32H, A37/49G from the wild type. The full sequence is given in the Supporting Information (SI). These mutations weaken the hydrophobic core but enable an experimental folding time of 15 μs even at its melting temperature T_m = 329 K. This mutant has not been subject to computational study before. The protein folded twice in the 10-μs enhanced sampling simulation described below with smallest C_α -RMSD compared to the crystal structure (PDB code: 3KZ3[23]) of less than 1.7 Å. The results highlight the potential of the enhanced sampling method and the accuracy of the underlying physical model (force field) in studying a relatively large protein with complicated native topology. The subsequent constant temperature simulations reveal the folding kinetics at three different temperatures. A 100-μs MD simulation at intermediate temperature (359 K) partially folded the protein into its native topology and revealed a relatively slow-folding pathway for λ-HG. The latter simulation was made possible through access to the same special purpose computer as used by Lindorff-Larsen et al.[8] Based on the simulations, we propose new experimental observables to identify this slow-folding pathway, as well as a new construct to eliminate the slow-folding pathway and accelerate folding. Depending on the reaction coordinates chosen, the results also suggest that the folding of λ-HG is not a simple two state process as one finds it for many small fast-folding proteins. This suggestion is in agreement with the multi-probe thermal melts measurement by Liu and Gruebele,[31] which shows that λ-HG is not a two-state like kinetic system even at its melting temperature. The multiple state nature of λ-repressor folding requires a more sophisticated analysis of experimental data than was adopted in the past.

Three types of MD simulations on the λ-repressor HG mutant (λ-HG) were carried out and are summarized in Table 1. The first is based on enhanced sampling, in which temperature is a random variable with values chosen between 300 K and 600 K; the second is a conventional equilibrium simulation, starting from the crystal structure; the third is a conventional folding simulation starting from an extended conformation. Equilibrium and

folding simulations were carried out at three temperatures: 329 K, 359 K, and 389 K. The CHARMM27 force field with CMAP corrections[32–34] and TIP3P water model[35] were used for all the simulations reported here.

We performed an enhanced sampling simulation (Sim-ES in Table 1) using NAMD[36] following the protocol suggested by Zhang et al.[17] The 10-μs simulation started from a completely extended conformation to avoid bias towards the native state. The protein folded twice into its native state in the simulation, namely at 0.5 μs and at 4 μs, and was subsequently expelled from the folded state through the applied temperature fluctuations. We consider a single frame conformation as folded if its C_α -RMSD relative to the crystal structure falls into the range defined by the mean ± standard deviation of the native simulation at T = 359 K (Sim-N359 in Table 1). The C_α -RMSD value along the simulation trajectory (Sim-ES in Table 1) is shown in Figure 1. A set of other quantities of interest (such as radius of gyration) are shown in Figure S1 in SI.

Cluster analysis was performed on the enhanced sampling trajectory Sim-ES based on pairwise RMSD. The occupancies of the top 20 clusters are shown in Figure 1 and the representative structures from the each cluster are shown in Figure S2. Two time windows (0 ~ 0.5 μs and 3.5 ~ 4.0 μs), where the folding events occurred, are shown enlarged in Figure S2. The pathway that leads to complete folding of the protein can be followed based on the cluster analysis. Due to the enhanced sampling method used in the current study, the folding pathways observed in Sim-ES may not be the most probable ones. Nevertheless, they represent one of the many physical pathways on the folding landscape. It is worth noting that the folded state is ranked as the most populated cluster without any prior knowledge of the crystal structure. The population distribution is a result of the enhanced sampling algorithm that lower the temperature when the protein is in a low energy state, such as the native state (see Computational Methods in SI). The results suggest a high potential of the enhanced sampling method in force field refinement and protein structure prediction. A representative folded structure of the most populated cluster is shown in Figure 1.

At around t = 0.5 μs the protein remained folded in the native state for a few tens of ns; at around t = 4 μs the protein remained folded in that state for about 500 ns. In either case, the protein is considerably less stable in the native state than expected for a constant temperature simulation. The instability is not surprising, however, as the enhanced sampling method continuously kicks the protein out of any state, including the native one, as explained in Computational Methods in SI. Potential energy and temperature along the enhanced sampling trajectory are shown in Figure S3 in SI. Two time periods around 0.5 μs and 4 μs, i.e., when the protein visits the native state, can be identified as low energy and low temperature regions. Other low energy and low temperature regions in Figure S3 represent stable intermediate states and short-lived kinetic traps along the folding pathways. The key of enhanced sampling is that it allows rapid discovery of the native state with a physical force field in a fully solvated simulation.

In addition to accelerating the search in conformational space, the enhanced sampling method employed here also covers a broad range of temperatures in the simulation, which permits calculation of the temperature dependence of certain structural characteristics given enough sampling. The probability distributions of radius of gyration R_{gyr} and C_α -RMSD values were calculated from simulation Sim-ES and are shown in Figure 2. The probability distribution near T = 305 K reveals several stable states that compete with the native state. The folding of many small, single domain, fast-folding proteins is often found to be a simple two-state process, connecting a basin of rapidly inter-converting unfolded states and a single basin around the folded state. However, this scenario clearly does not apply to the studied

fast-folding λ-repressor mutant that exhibits several basins seen as red islands in Figure 2 when examined through the reaction coordinate of R_{gyr} and C_α -RMSD. High temperature smoothes out the free energy landscape and shifts the equilibrium toward unfolded states; these unfolded states are characterized by large C_α -RMSD values, yet compact structures with small radius of gyration. As discussed previously,[37] the folding kinetics of λ-repressor are probe-dependent. The probability distribution of $D_{W22-H33}$ (distance between sidechain of Trp-22 and sidechain of His-33) and Trp-SASA (solvent accessible surface area of Trp-22), two reaction coordinates that are characteristic of experimental observables, such as tryptophan fluorescence lifetime and tryptophan fluorescence Stokes shift, reveals a two-state like behavior, as shown in Figure S4.

The enhanced sampling simulation Sim-ES permits the observation of the complete folding event, but lacks information on the kinetics of the folding process. To obtain such information, constant temperature simulations are required. We performed such simulations at T = 329, 359, and 389 K. The temperatures are chosen here higher than experimental folding temperatures as it has been shown that the melting temperature of λ-repressor in simulation is higher than the one measured in experiment;[28] previous studies of λ-repressor folding also involved simulations at relatively high temperatures.[8,29] The temperature discrepancy results from the protein force field and water model that utilize parameters which remain unchanged for various temperatures and pressures. The same extended conformation as in the enhanced sampling simulation served as the starting structure. Such extended conformation slows down the folding process compared to observed refolding from thermally and chemically denatured states.[6] However, the TIP3P water model overestimated mobility compensates this shortcome to some degree.[38,39]

The simulation at T = 329 K was carried out for about 14 μs. During an initial collapse of the extended protein, α-helical structure formed rapidly. As shown in Figure 3, α-helical content reached its native level within 2 μs. Subsequently, the protein accumulated more helical structure, even more than formed in the native state. The CHARMM27 force field with CMAP corrections [32–34] used in the current study is known to over-stabilize helical structure,[5,18] which may contribute to this helical overshoot. On the other hand, the helical overshoot may be analogous to the burst phase observed by circular dichroism kinetics measurements of λ-repressor in cryogenic solvent.[40] The helical overshoot state dominates most of the trajectory (see cluster analysis at T = 329 K and the representative structure of cluster # 1 in Figure S5). A set of other quantities of interest (such as radius of gyration) and secondary structure evolution are presented in Figure S6 and Figure S7, respectively. The clusters shown in Figure S5 represent an ensemble of minimum energy compact structures (MECS),[41] which are located near the bottom of the folding energy landscape. The MECS significantly reduce the conformational space that a protein has to sample in order to fold into the native state on a biological time scale. The existence of MECS has been predicted from the heteropolymer lattice model[41] and demonstrated for both ubiquitin[42] and polyglutamine[43] through single molecule force-clamp measurements. Our simulations confirm the existence of MECS even for a fast folding protein like λ-repressor and characterize the structural nature of MECS. The presence of MECS may be a consequence of folding physics principle, or an evolutionary adaptation that allows proteins to fold upon mutation or acquire new functions by making alternative low-energy states available.

At a temperature of T = 359 K, we performed MD simulation Sim-F359 covering 100 μs. After initial collapse of the extended protein to a mostly helical state, the folding trajectory is characterized by a series of interconversions between several transient kinetic intermediate states revealed by cluster analysis as shown in Figure S5. Representative structures from the four most populated clusters are also shown in Figure S5. Helix 1 is the dominant structure element, to which the other four helices add themselves, contributing

also to the main hydrophobic core. The partially folded structure reached at the end of the simulation belongs to cluster # 1, which is constantly visited during the last 60 μs of simulation.

During simulation Sim-F359, helices 1, 2, and 3 assumed their near native conformation as shown in Figure S8. The partially folded structure at the end of the simulation is shown in Figure 4, along with the structure of the native state. Helix 4 and 5 formed individually, but their orientation relative to helix 1 is different from the one in the native state. The folding pathway seen in the simulation differs from what had been proposed before using a variational approach[24,25] or employing implicit solvent replica-exchange molecular dynamics.[28] The previous studies suggested that helix 1 and helix 4 orient correctly relative to the native structure more often than do helices 2 and 3. Other than different methods used in each study, mutation effects also need to be considered. Port-man et al.[24,25] performed their analysis on wild type λ-repressor and Larios et al.[28] studied another fast-folding mutant. We note that both enhanced sampling and constant temperature simulations show helices 1 and 4 to be the most stable structural element, in agreement with the predictions from other studies.[24,28]

The folding pathway observed in our simulation is different from the one suggested previously.[24,25,28] Moreover, the folding observed is slower than measured in experiment,[31] which revealed an activated phase ($k_a = 15\ \mu s^{-1}$) and a molecular phase ($k_m = 2\ \mu s^{-1}$). It is possible that the single trajectory in our simulation follows a variant, slow-folding pathway. The native state of λ-repressor contains a major hydrophobic core formed by helices 1, 2, 3, and 4 with a small hydrophobic patch between the N-terminus of helix 4 and helix 5 (see Figure 4). The residues involved in forming the hydrophobic core and hydrophobic patch are shown in Figure S8. After the hydrophobic core is partially formed by helices 1, 2, and 3, resulting in a near native value for the hydrophobic SASA (Figure S8), the competition between helices 4 and 5 to dock to the major hydrophobic core determines the rate of complete folding. If helix 4 docks to the major hydrophobic core first, helix 5 will likely find the small hydrophobic patch at the N-terminus of helix 4 and complete the folding quickly. On the other hand, if the hydrophobic patch on helix 5 docks to the major hydrophobic core first, the protein will likely form an alternative low-energy metastable intermediate state that features a long escape time, which results in slow folding. The latter is what we observe in our simulation. Helix 5 played an important role in stabilizing the folding intermediate. Although helix 4 is participating in the hydrophobic core, its correct register with helix 1 is prevented by helix 5. This conformation is characterized by a non-native arrangement of helix 4 and 5 compared to the crystal structure (see Figure S8).

In Sim-F359, the protein folded very quickly to the near native state described above, but complete folding was not achieved. The simulation time, even though exceeding the observed folding time (100 μs vs 15 μs), was not long enough to escape the near native intermediate. Factors that may contribute to the long escape time might be a lack of accuracy of the force field and a lack of temperature correspondence between simulation and experiment, i.e., 359 K may not be the optimal folding temperature for λ-HG in the simulation. The TIP3P water model employed here has been shown to affect the dynamics of unfolded peptides[39] and, hence, could also adversely affect the folding. It is also possible that the experiment monitors a faster-equilibrating variable (Trp fluorescence) than the global measurement (RMSD value) used in the simulation, and λ-HG may actually fold slower than suggested by observation. To experimentally reveal the slow-folding pathway seen in the simulation, a new experimental observable needs to be designed that measures the folding rate more accurately and a new construct should be engineered to remove the kinetic trap observed in the simulation and see if the construct, also analyzed through the new observable, folds indeed faster.

To check proximity to the native state for the trajectory at T = 359 K we calculated C_α -RMSD values for the first two helices relative to the crystal structure. The results are shown in Figure 4a. The first two helices are seen to fold into their native conformation and form native contacts with each other several times during the simulation. The first local folding event happened within 18 μs. A stable folded conformation formed at 65 μs and remained stable for more than 30 μs. We note that temperature-jump experiments probing the quenching of Trp-22 fluorescence by His-33 yielded a folding time of 15 μs for the λ-HG mutant;[31] Trp-22 is from helix 1 and His-33 from helix 2, both shown in licorice representation in Figure 4. The early formation of a native-like conformation for helix 1 and helix 2, as observed in the simulation, may yield an experimental signal that looks like one corresponding to formation of the native structure. The distance, D, between Trp-22 and His-33 monitored in the simulation is shown in Figure 4b. Indeed, the native structure value of D is reached in less than 10 μs and is re-established many times during the simulation. The solvent accessible surface area (SASA) of Trp-22 also achieves its native structure value in about 10 μs and remains constant for the last 50 μs.

Using Trp-22 and His-33 as experimental probes is appropriate if the fast-folding process suggested by Larios et al. [28] is taking place, in which case the formation of the native conformation of helix 1 and helix 2 is the rate-limiting step. However, these probes do not distinguish the slow-folding pathway from the fast-folding pathway. A native-like signature will also be observed in the experiment in the early stage of a slow-folding pathway. Our simulations suggest that the last step of slow-folding achieves the correct orientation and packing of helices 1 and 4. We propose that a histidine mutation on helix 4 at position of Leu-69 should be able to resolve the different folding pathways. The side chain distance between Trp-22 and Leu-69 is 6.37 ± 0.75 Å in the native state calculated from Sim-N359, close to the side chain distance between Trp-22 and His-33 (6.33 ±,1.48 Å) calculated from the same simulation, even though the sidechain of Leucine is slightly shorter than that of Histidine. If the quenching occurs between Trp-22 and His-33, it will also occurs between Trp-22 and His-69 once the mutation is made. On the other hand, although the native value of the distance between Trp-22 and His-33 is reached in the near native folding intermediate in Sim-F359 (Figure 4), a small distance between Trp-22 and Leu-69 is only achieved transiently as shown in Figure S9.

The identification of the slow-folding pathway enables one to propose possible fast-folding mutants to be tested in experiment. One way to eliminate the competition between helices 4 and 5 to dock to the major hydrophobic core is simply to delete helix 5 from λ-repressor. We suggest a construct with truncations after Ser-77 from the original λ-HG to remove the hydrophobic patch on helix 5. The resulting 4-helix bundle construct will still have the main hydrophobic core to maintain its stability, but should fold even faster than the λ-HG mutant if our hypothesis is correct. A truncated version of λ-repressor with only helices 1 and 4 left (2-helix bundle) has shown similar stability and folding rate as its parental 5-helix bundle.[44] The same experimental observable (quenching of Trp-22 by His-69) is also suitable for measuring the folding rate of the proposed fast-folding truncated λ-repressor with 4 helices.

A relatively short, 13-μs simulation, Sim-F389, was performed at T = 389 K. A complete folding event was not observed. The simulation confirms that helices 1 and 4 are more stable than the other three helices (see Figure S7 in SI). The protein sampled several low-energy states with native hydrophobic SASA during the trajectory as shown in Figure 3. An analysis of our simulations suggest that λ-repressor folding can be treated as a two-state process when examined through certain reaction coordinates, such as hydrophobic SASA (native vs non-native state). The molecular dynamics simulations reveal, however, structural details of the protein when it visits states with native hydrophobic SASA, showing that SASA value alone does not signify complete folding. Indeed, cluster analysis of our

trajectory revealed that the states with native hydrophobic SASA are highly degenerate. As shown in Figure S5, cluster 4, cluster 1, and cluster 2 are visited at 1 μs, 3 μs, and 8 μs, respectively. We conclude that a two-state model for the folding process of λ-repressor can be an artifact of the chosen probe (or reaction coordinate).

In summary, we combined both enhanced sampling simulation and constant temperature simulation to study the folding of λ-repressor. The enhanced sampling simulations show that the current force field is accurate enough to describe the folding of this five-helix bundle protein made of 80 amino acid residues. The folding of λ-repressor is found not to be a simple two state process. Our study also demonstrates the potential of the enhanced sampling method and long-time MD simulations in revealing protein folding mechanisms, as well as in proposing novel experimental observables and fast-folding mutants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grant P41-RR005969 and R01-GM093318 from the National Institutes of Health and PHY0822613 from the National Science Foundation. The authors acknowledge supercomputer time provided by the National Center for Supercomputing Applications and the Texas Advanced Computing Center via Extreme Science and Engineering Discovery Environment grant (XSEDE) MCA93S028, and the use of the parallel computing resource Taub provided by the Computational Science and Engineering Program at the University of Illinois. Anton computer time was provided by the National Resource for Biomedical Supercomputing and the Pittsburgh Supercomputing Center through grant RC2GM093307 from the National Institutes of Health. The Anton machine had been donated generously by David E. Shaw; we are most grateful for the opportunity to use Anton.[45] We thank Dr. Markus Dittrich and other staff at NRBSC for the support in using the Anton machine and Andrew Blanchard for critical reading of the an early version of the manuscript. Data analysis and figure rendering were done using VMD.[46]

References

1. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of Protein Folding: The Energy Landscape Perspective. Annu Rev Phys Chem. 1997; 48:545–600. [PubMed: 9348663]
2. Thirumalai D, O'Brien EP, Morrison G, Hyeon C. Theoretical Perspectives on Protein Folding. Annu Rev Biochem. 2010; 39:159–183.
3. Kubelka J, Hofrichter J, Eaton WA. The Protein Folding 'Speed Limit'. Curr Opin Struct Biol. 2004; 14:76–88. [PubMed: 15102453]
4. Clementi C. Coarse-Grained Models of Protein Folding: Toy Models or Predictive Tools? Curr Opin Struct Biol. 2007; 18:10–15. [PubMed: 18160277]
5. Freddolino PL, Liu F, Gruebele M, Schulten K. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. Biophys J. 2008; 94:L75–L77. [PubMed: 18339748]
6. Freddolino PL, Schulten K. Common Structural Transitions in Explicit-Solvent Simulations of Villin Headpiece Folding. Biophys J. 2009; 97:2338–2347. [PubMed: 19843466]
7. Freddolino PL, Harrison CB, Liu Y, Schulten K. Challenges in Protein Folding Simulations. Nature Phys. 2010; 6:751–758. [PubMed: 21297873]
8. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How Fast-Folding Proteins Fold. Science. 2011; 334:517–520. [PubMed: 22034434]
9. Duan Y, Kollman P. Pathways to a Protein Folding Intermediate Observed in a 1 Microsecond Simulation in Aqueous Solution. Science. 1998; 282:740–744. [PubMed: 9784131]
10. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, et al. Atomic-level characterization of the structural dynamics of proteins. Science. 2010; 330:341–346. [PubMed: 20947758]
11. Sugita Y, Okamoto Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. Chem Phys Lett. 1999; 314:141–151.

12. Onuchic JN, García AE. Folding a protein in a computer: An Atomic Description of the Folding/Unfolding of Protein A. Proc Natl Acad Sci USA. 2003; 100:13898–13903. [PubMed: 14623983]
13. Rosta E, Hummer G. Error and Efficiency of Replica Exchange Molecular Dynamics Simulations. J Chem Phys. 2009; 131:165102. [PubMed: 19894977]
14. Day R, Paschek D, García AE. Microsecond Simulations of the Folding/Unfolding Thermodynamics of the Trp-Cage Miniprotein. Proteins: Struct, Func Gen. 2010; 78:1889–1899.
15. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. J Chem Phys. 2002; 116:9058–9067.
16. Zhang C, Ma J. Enhanced Sampling in Generalized Ensemble with Large Gap of Sampling Parameter: Case Study in Temperature Space Random Walk. J Chem Phys. 2009; 130:194112. [PubMed: 19466826]
17. Zhang C, Ma J. Enhanced Sampling and Applications in Protein Folding in Explicit Solvent. J Chem Phys. 2010; 132:244101. [PubMed: 20590175]
18. Freddolino PL, Park S, Roux B, Schulten K. Force Field Bias in Protein Folding Simulations. Biophys J. 2009; 96:3772–3780. [PubMed: 19413983]
19. Rajan A, Freddolino PL, Schulten K. Going Beyond Clustering in MD Trajectory Analysis: an Application to Villin Headpiece Folding. PLoS One. 2010; 5:e9890. (12 pages). [PubMed: 20419160]
20. Piana S, Sarkar K, Lindorff-Larsen K, Guo M, Gruebele M, Shaw DE. Computational Design and Experimental Testing of the Fastest-Folding β-Sheet Protein. J Mol Biol. 2011; 405:43–48. [PubMed: 20974152]
21. Huang GS, Oas TG. Submillisecond Folding of Monomericλ Repressor. Proc Natl Acad Sci USA. 1995; 92:6878–6882. [PubMed: 7624336]
22. Yang WY, Gruebele M. Folding at the Speed Limit. Nature. 2003; 423:193–197. [PubMed: 12736690]
23. Liu F, Gao YG, Gruebele M. A Survey of lambda Repressor Fragments from Two-State to Downhill Folding. J Mol Biol. 2010; 397:789–798. [PubMed: 20138892]
24. Portman JJ, Takada S, Wolynes PG. Microscopic Theory of Protein Folding Rates. II. Local Reaction Coordinates and Chain Dynamics. J Chem Phys. 2001; 114:5082–5096.
25. Portman JJ, Takada S, Wolynes PG. Microscopic Theory of Protein Folding Rates. I. Fine Structure of the Free Energy Profile and Folding Routes from a Variational Approach. J Chem Phys. 2001; 114:5069–5081.
26. Pogorelov TV, Luthey-Schulten Z. Variations in the Fast Folding Rates of the λ-Repressor: a Hybrid Molecular Dynamics Study. Biophys J. 2004; 87:207–214. [PubMed: 15240458]
27. Allen LR, Krivov SV, Paci E. Analysis of the Free-Energy Surface of Proteins from Reversible Folding Simulations. PLoS Comput Biol. 2009; 5:e1000428. [PubMed: 19593364]
28. Larios E, Pitera JW, Swope WC, Gruebele M. Correlation of Early Orientational Ordering of Engineered $\lambda_{6–85}$ Structure with Kinetics and Thermodynamics. Chem Phys. 2006; 323:45–53.
29. Bowman GR, Voelz VA, Pande VS. Atomistic Folding Simulations of the Five-Helix Bundle Protein $\lambda_{6–85}$. J Am Chem Soc. 2011; 133:664–667. [PubMed: 21174461]
30. Prigozhin MB, Gruebele M. The Fast and the Slow: Folding and Trapping of $\lambda_{6–85*}$. J Am Chem Soc. 2011; 133:19338–19341. [PubMed: 22066714]
31. Liu F, Gruebele M. Tuning $\lambda_{6–85}$ Towards Downhill Folding at its Melting Temperature. J Mol Biol. 2007; 370:574–584. [PubMed: 17532338]
32. Mackerell AD. Empirical Force Fields for Biological Macromolecules: Overview and Issues. J Comp Chem. 2004; 25:1584–1604. [PubMed: 15264253]
33. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998; 102:3586–3616.
34. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD. Importance of the CMAP Correction to the CHARMM22 Protein Force Field: Dynamics of Hen Lysozyme. Biophys J. 2006; 90:L36–38. [PubMed: 16361340]

35. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys. 1983; 79:926–935.
36. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable Molecular Dynamics with NAMD. J Comp Chem. 2005; 26:1781–1802. [PubMed: 16222654]
37. Ma H, Gruebele M. Kinetics are Probe-Dependent During Downhill Folding of an Engineered λ_{6-85} Protein. Proc Natl Acad Sci USA. 2005; 71:2283–2287. [PubMed: 15699334]
38. Vega C, Abascal JLF, Conde MM, Aragones JL. What Ice Can Teach Us About Water Interactions: a Critical Comparison of the Performance of Different Water Models. Faraday Discuss. 2009; 141:251–276. [PubMed: 19227361]
39. Florová P, Sklenovský P, Banáš P, Otyepka M. Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact. J Chem Theor Comp. 2010; 6:3569–3579.
40. Dumont C, Matsumura Y, Kim SJ, Li J, Kondrashkina E, Kihara H, Gruebele M. Solvent-Tuning the Collapse and Helix Formation Time Scales of λ_{6-85*}. Prot Sci. 2006; 15:2596–2604.
41. Camacho CJ, Thirumalai D. Minimum Energy Compact Structures of Random Sequences of Heteropolymers. Phys Rev Lett. 1993; 71:2505–2508. [PubMed: 10054697]
42. Garcia-Manyes S, Dougan L, Badilla CL, Brujic J, Fernandez J. Direct Observation of an Ensemble of Stable Collapsed States in the Mechanical Folding of Ubiquitin. Proc Natl Acad Sci USA. 2009; 106:10534–10539. [PubMed: 19541635]
43. Dougan L, Li J, Badilla CL, Berne BJ, Fernandeza JM. Single Homopolypeptide Chains Collapse into Mechanically Rigid Conformations. Proc Natl Acad Sci USA. 2009; 106:12605–12610. [PubMed: 19549822]
44. Prigozhin MB, Sarkar K, Law D, Swope WC, Gruebele M, Pitera J. Reducing λ Repressor to the Core. J Phys Chem B. 2011; 115:2090–2096. [PubMed: 21319829]
45. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, et al. Anton, a special-purpose machine for molecular dynamics simulation. Comm ACM. 2008; 51:91–97.
46. Humphrey W, Dalke A, Schulten K. VMD – Visual Molecular Dynamics. J Mol Graphics. 1996; 14:33–38.

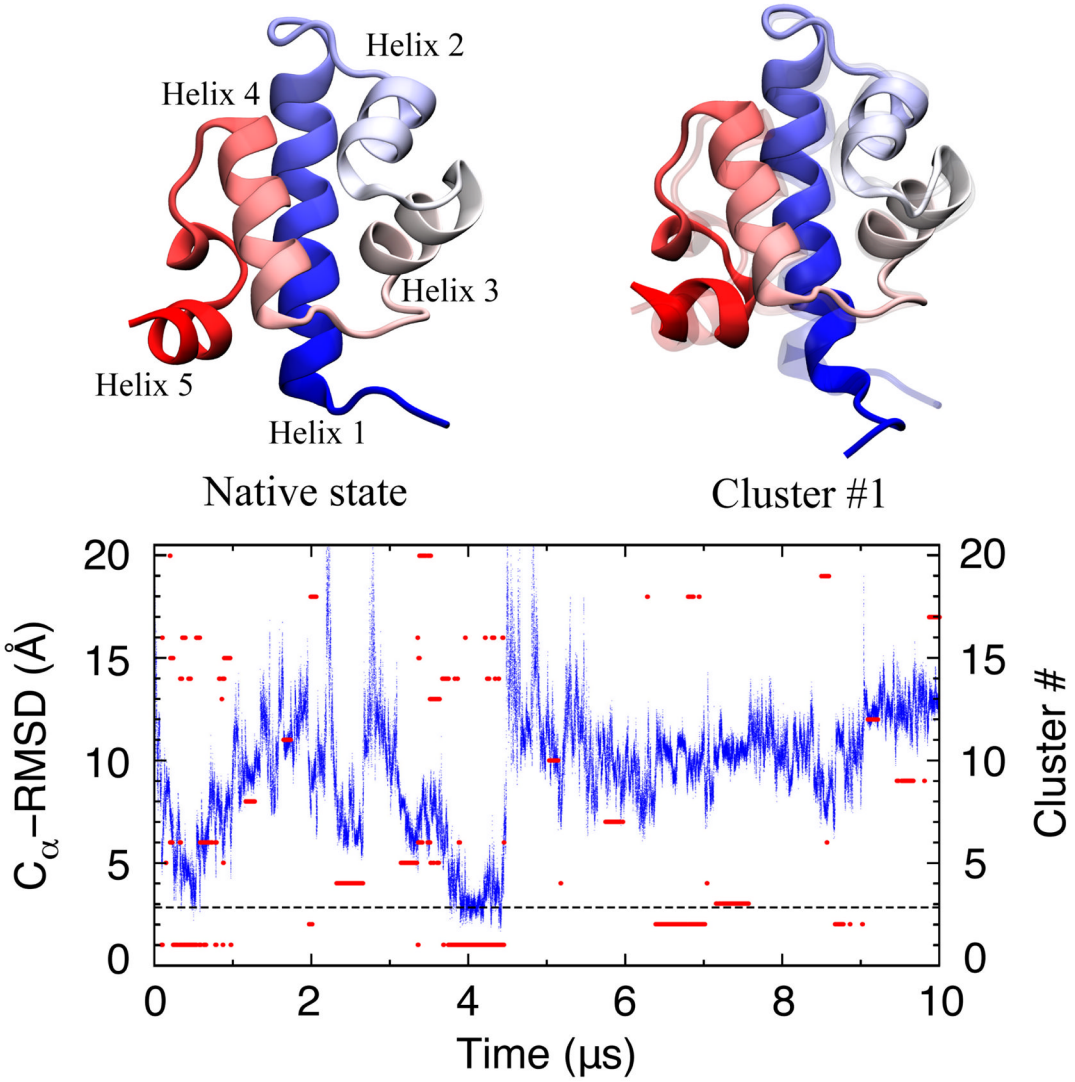

Figure 1.

C_α -RMSD relative to the 3KZ3 [23] crystal structure (shown at top left) calculated from an enhanced sampling simulation (Sim-ES in Table 1). The C_α -RMSD native value is defined by the mean value (dashed line) from the 300-ns equilibrium simulation of the native structure (Sim-N359 in Table 1). The protein folded reversibly into the native value twice in the simulation. Occupancies of the top 20 clusters throughout the trajectory are shown in red. The representative folded structure, identified through the most populated cluster, is shown at top right. The crystal structure is superimposed as a transparent cartoon representation for comparison. Protein coloring runs blue to red from N-terminus to C-terminus.

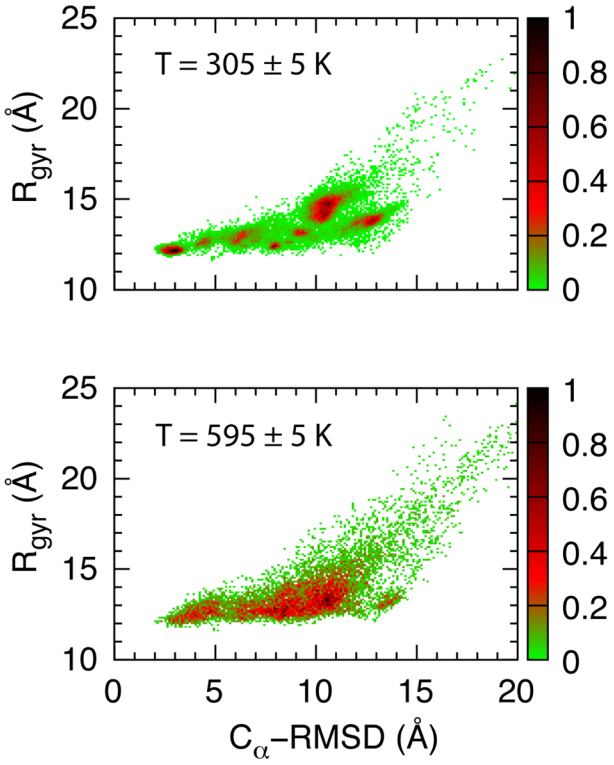

Figure 2.

Temperature dependence of the histogram distribution of radius of gyration R_{gyr} and C_α -RMSD values at T = 305 ± 5 K (top) and at T = 595 ± 5 K (bottom). The histogram distribution was determined from simulation Sim-ES and the associated probability, as shown, is normalized such that the highest value is 1.

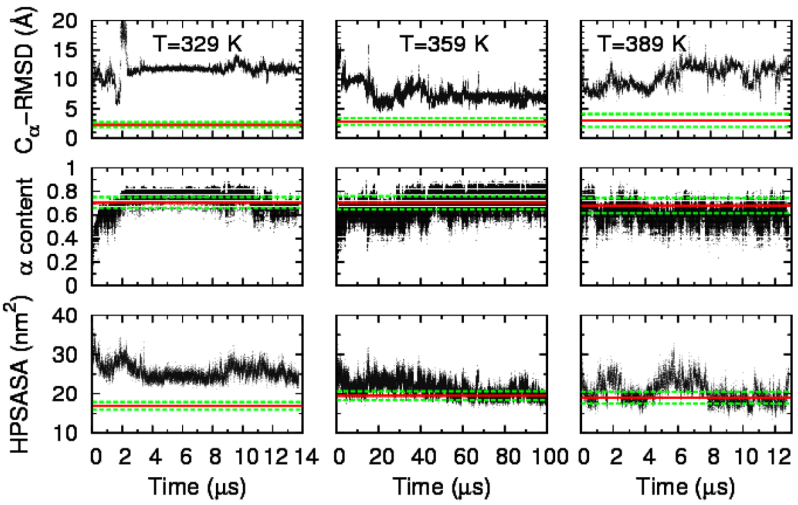

Figure 3.

Characterization of λ-repressor folding trajectory at T = 329 K, T = 359 K, and T = 389 K taken from simulations Sim-F329, Sim-F359, and Sim-F389, respectively. C_α -RMSD values have been calculated relative to the crystal structure 3KZ3;[23] α-content is the fraction of residues that are in α-helical conformation; HPSASA refers to the solvent accessible surface area of hydrophobic groups. The native ranges are defined by the mean value (red solid line) ± standard deviations (green dashed line) from the 300-ns equilibrium simulations of the native structure at the respective temperatures (Sim-N329, Sim-N359, and Sim-N389 in Table 1)

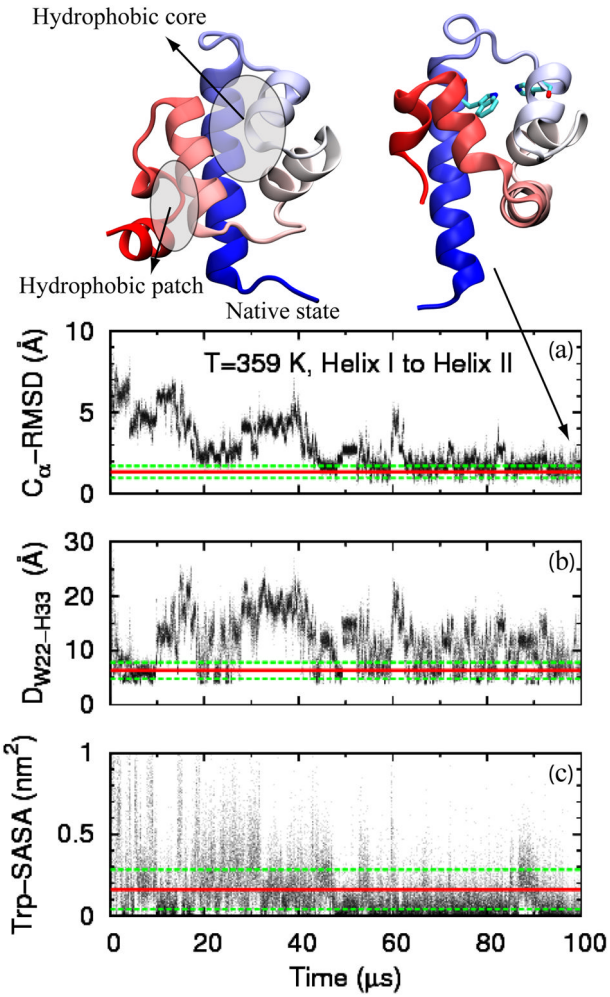

Figure 4.

Observables monitored computationally along the trajectory Sim-F359. (a) C_α -RMSD for the first two helices (from residue Gln-9 to Asp-38) relative to the crystal structure. (b) Distance between sidechain of Trp-22 and sidechain of His-33. (c) Solvent accessible surface area of the sidechain of Trp-22 (Trp-SASA). The crystal structure and the partially folded structure reached in the simulation are provided at the top (same color scheme as in Figure 1). The experimental probe residues tryptophan from helix 1 and histidine from helix 2 are shown in licorice representation.

Table 1

Summary of simulations performed

Name[a]	Init. Conform.	Temp.	Duration	Min./Max. C_α-RMSD	Mean/S.D.[b] C_α-RMSD
Sim-ES	Extended	300–600 K	10.03 μs	1.7/>50.0 Å	*/*
Sim-N329	Crystal	329 K	0.30 μs	0.9/4.2 Å	2.3/0.4 Å
Sim-N359	Crystal	359 K	0.30 μs	1.1/5.3 Å	2.8/0.6 Å
Sim-N389	Crystal	389 K	0.30 μs	1.1/8.0 Å	3.0/1.1 Å
Sim-F329	Extended	329 K	13.73 μs	5.6/>50.0 Å	*/*
Sim-F359	Extended	359 K	100.15 μs	4.3/>50.0 Å	*/*
Sim-F389	Extended	389 K	12.83 μs	5.9/>50.0 Å	*/*

[a] Under the column “Name”, ES denotes an enhanced sampling simulation, N denotes a simulation that started from the native state, i.e., the crystallographic structure, and F denotes a constant temperature folding simulations.

[b] S.D stands for standard deviation; Mean and S.D from native simulations are used as the criteria to judge the degree of folding.