New Insights into Protein (Un)Folding Dynamics

ARTICLE *in* JOURNAL OF PHYSICAL CHEMISTRY LETTERS · MARCH 2015

Impact Factor: 7.46 · DOI: 10.1021/acs.jpclett.5b00055

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

New Insights into Protein (Un)Folding Dynamics

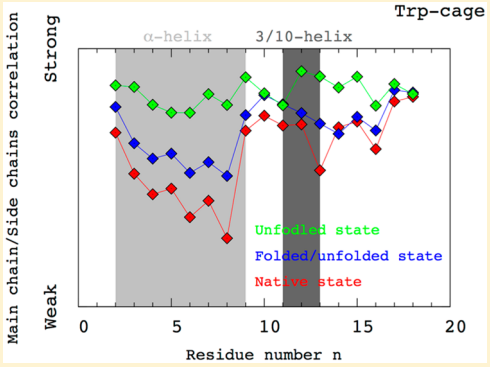
Yoann Cote,[†] Gia G. Maisuradze,*[,‡] Patrice Delarue,[†] Harold A. Scheraga,*[,‡] and Patrick Senet*[,†,‡]

[†]Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne 9 Av., Alain Savary, BP 47 870, F-21078 Dijon Cedex, France

[‡]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States

Supporting Information

ABSTRACT: A fundamental open problem in biophysics is how the folded structure of the main chain (MC) of a protein is determined by the physics of the interactions between the side chains (SCs). All-atom molecular dynamics simulations of a model protein (Trp-cage) revealed that strong correlations between the motions of the SCs and the MC occur transiently at 380 K in unfolded segments of the protein and during the simulations of the whole amino-acid sequence at 450 K. The high correlation between the SC and MC fluctuations is a fundamental property of the unfolded state and is also relevant to unstructured proteins as intrinsically disordered proteins (IDPs), for which new reaction coordinates are introduced. The presented findings may open a new door as to how functions of IDPs are related to conformations, which play a crucial role in neurodegenerative diseases.



To understand protein folding, one needs to understand how the motions of the side chains (SCs) are propagated to the main chain (MC) and vice versa. The motions of the nonpolar SCs are one of the (hydrophobic) forces driving the folding of a protein: these motions orient the MC locally.[1] Similarly, the formation of hydrogen bonds between the amide N−H bonds and the carbonyl C═O groups of the MC affects local motions of the SCs. The molecular dynamics (MD) simulations of a model protein (Trp-cage), reported here, demonstrate that the locally unfolded regions of a protein favor strong coupling between the structural fluctuations of the SCs and of the MC. In the unfolded state, the SCs and the MC are strongly coupled to each other, unlike the SCs and MC of residues in secondary structures in the native state. A high correlation between the motions of the SCs and the MC can guide folding of the protein. This property may be important in intrinsically disordered proteins (IDPs). IDPs lack stable structure and are highly flexible; they play crucial roles in the cell and are related to several human diseases. Many IDPs undergo binding-induced folding, which, according to present findings, should be facilitated by the strong correlation between the SCs and the MC in the disordered regions of the IDPs.

To shed light on the local coupling between the SCs and the MC in protein folding, we analyzed the fluctuations of the coarse-grained dihedral angles (CGDAs) γ_n built on four successive C^α atoms (MC) and the CGDAs δ_n built on two successive C^α atoms and their respective C^β atoms (SCs) of the Trp-cage during folding/unfolding events by using all-atom MD simulations in explicit solvent (water). The CGDAs are a subset of coordinates monitored in all-atom MD simulations to characterize large protein structural fluctuations and protein folding.[2,3] It should be noted that, for Gly residues, a pseudo-C^β atom is defined as the position of its side-chain H atom to define the CGDA δ_n.[4]

The ultrafast folder Trp-cage[5] was chosen because it has been studied extensively by MD simulations and experiments and because unfolding events can be reproduced by unbiased all-atom MD simulations in water within a reasonable computational time. The Trp-cage is a 20-residue protein designed to understand protein-folding mechanisms.[5] It consists of one α-helix and one (3/10)-helix. It is a very fast folder that folds spontaneously and cooperatively into its native structure in 4 μs.[6] Because of its small size and its fast kinetics, the Trp-cage has become a typical protein for MD simulations of protein folding in implicit[7−9] or explicit[10−12] solvent.

Two reference MD trajectories (with different initial conditions) of the Trp-cage in the native state were produced over a 400 ns time scale at 300 K. The time evolution of the correlation between the fluctuations of the MC and of the SCs of the Trp-cage during unbiased unfolding/folding MD trajectories is investigated at 380 K on the same time scale for two MD trajectories with different initial conditions. An artificially high temperature (450 K) was also chosen for one simulation of a completely unfolded state of the Trp-cage on the same time scale. The initial structure of the MD runs is the model 1 of PDB 1L2Y unless otherwise specified. The dynamics of the MC and the SCs around each residue n were monitored by recording the CGDA γ_n and δ_n, respectively, as in our previous work.[4] Large structural fluctuations (unfolding events at 380 K) were monitored by recording the

Received: January 10, 2015
Accepted: March 9, 2015

DOI: 10.1021/acs.jpclett.5b00055
J. Phys. Chem. Lett. 2015, 6, 1082−1086

root-mean-square-deviation (rmsd) between the C^{α} positions at a given time at their corresponding positions in the initial structure. A new quantity, named the steric parameter, s_n, has been defined to characterize these fluctuations. It measures the distance of the nth SC relative to the rest of the protein. A low value of s_n means that the SC_n is far from the rest of the protein. On the contrary, a high value of s_n means that the SC_n occurs at short distances from the rest of the protein. Near the Pro residues, the variation of s_n has to be interpreted with caution because the SC is covalently linked to the backbone; hence a variation of s_n corresponds to a motion of the Pro ring relative to the backbone rather than a real motion outward or toward the protein. All MD calculations were performed at several constant temperatures and pressure (1 bar) by using the GROMACS package[13] and the Amber99SB force field[14] with the TIP3P force field[15] for the water. Results of one representative MD trajectory at each temperature are described here. Similar results found for other MD trajectories and methodological details are provided in the Supporting Information.

In agreement with our previous results for the VA3 protein in its native state,[4] the Pearson correlation coefficient, R,[16] between the time series $\gamma_n(t)$ and $\delta_n(t)$, averaged over time at 300 K (native state of the Trp-cage), varies strongly along the amino-acid sequence (Figure 1 A). It is low for CGDAs built only on residues located inside secondary structures (n = 3−8) and large for CGDAs located at the boundaries of the α helix (n = 2, 9), of the 3/10 helix (n = 11 and 13), and in the flexible parts: the loop (n = 10) and the C-terminus of the Trp-cage (n = 14−20). The correlation coefficient R at n = 12 is high despite the fact that it is located in a helix because of the very short length and instability[17] of the 3/10 helix.

Increase in temperature has a significant impact on the time-average correlation coefficient R, as shown in Figure 1 A. Except for n = 14, the average correlation coefficient is higher at 380 K than at 300 K, and its value is more uniform along the amino-acid sequence at 380 K than at 300 K. (The minimum and maximum values of R are R_{min} = 0.3 and R_{max} = 0.8 at 300 K and R_{min} = 0.5 and R_{max} = 0.8 at 380 K.) At 380 K, the CGDAs visited both folded states and unfolded states, as shown by the time evolution of the rmsd between the initial structure and each snapshot of the MD runs extracted every picosecond (Figure 2). There are large fluctuations (>0.4 nm) at 380 K relative to the mean (black solid line Figure 2) between 130 and 200 ns and after 350 ns for which the protein is transiently unfolded. (See the insets of Figure 2.) These large structural changes are the main cause of an increase in the average correlation coefficient, as demonstrated in Figure 1B. Indeed, the values of R averaged at 380 K over the time period for which the Trp-cage is folded (Figure 1B) are close to the same values in the native state at 300 K (red curve in Figure 1A), in particular, in the α-helix segment, whereas the values of R averaged at 380 K over the time period for which the Trp-cage is unfolded (Figure 1B) are higher than the averaged values of R at 380 K (blue curves in Figure 1A,B). At 450 K, the time average correlation coefficient R between the SCs and the MC is nearly equal for all residues: it varies between 0.75 and 0.9 (Figure 1A).

Calculation of the free-energy profile as a function of the rmsd (Figure S1 in the SI) shows that the most probable structures of the Trp-cage have an rmsd of 0.67 nm at 450 K and are thus unfolded. Calculation of the average of s_n over the entire duration of the trajectory at 450 K (Figure S2 in the SI)

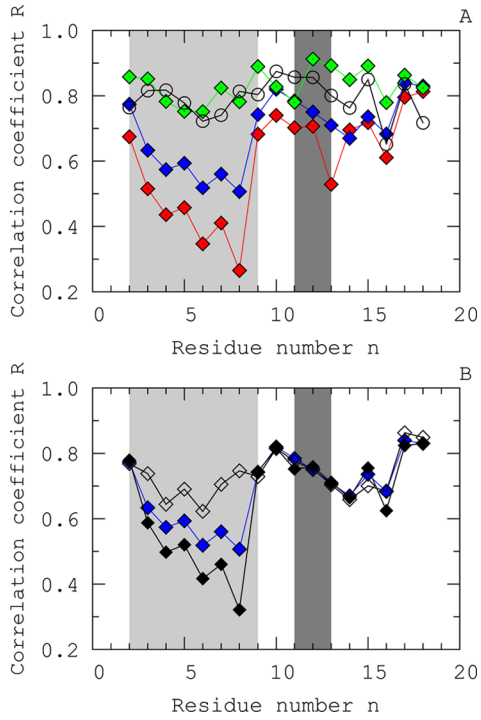


Figure 1. Variation along the amino-sequence of the Pearson correlation coefficient R computed between the time series $\gamma_n(t)$ and $\delta_n(t)$ for the MD trajectories of the Trp-cage. The α-helix and the 3/10-helix are indicated by light-gray and dark-gray stripes, respectively. (A) Average of R over the entire duration of the trajectory at 300 (red diamonds), 380 (blue diamonds), and 450 K (green diamonds), starting with an initial *native* structure. Values of R for an initial *unfolded* structure simulated at 300 K during 1 ns are shown by empty circles. (See the text.) (B) Average of R computed at 380 K over the entire duration of the trajectory (blue diamonds, over only the snapshots for which the structure is folded (rmsd < 0.21 nm) (black diamonds) and over only the snapshots for which the structure is unfolded (rmsd > 0.67 nm) (empty diamonds). The cutoff values are the average rmsd at 300 K (0.21 nm) and the average rmsd at 450 K (0.67 nm).

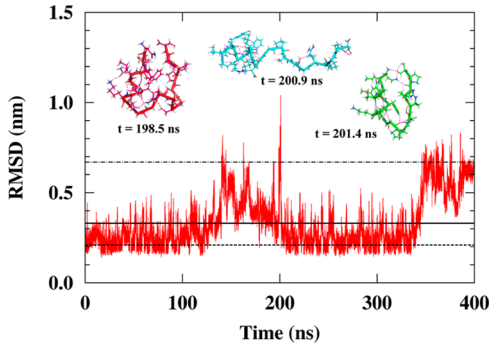


Figure 2. Time evolution of the rmsd of the C^{α} atoms at every picosecond relative to the initial native structure computed at 380 K. The average values of the rmsd calculated over the entire duration of the trajectories at 380 (horizontal black solid line), 300 (lower horizontal dashed line), and 450 K (upper horizontal dotted-dashed line) are shown for comparison.

shows that s_n decreases for all residues (except at n = 8 and 13) compared with its values in the native structure.

To test whether these structural changes (indicated by s_n) are responsible for the large values of R, we selected one of the

DOI: 10.1021/acs.jpclett.5b00055
J. Phys. Chem. Lett. 2015, 6, 1082−1086

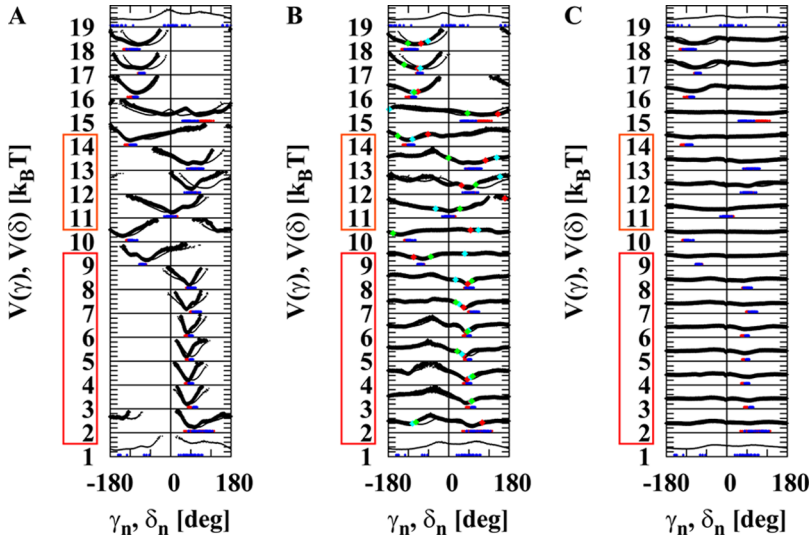


Figure 3. Effective FEPs $V(\gamma_n)$ (thick black lines) and $V(\delta_n)$ (thin black lines) along the amino-acid sequence of the Trp-cage (γ_n, n = 2−18; δ_n, n = 1−19) computed as in ref 18. FEPs computed from a 400 ns MD trajectory performed at 300 (A), 380 (B), and 450 K (C). All FEPs are compared with the NMR-derived structural data computed from the 1L2Y PDB files (blue diamonds, γ_n; red diamonds, δ_n). Residues 2−8 form an α-helix, and residues 11−14 form a 3/10-helix. The red, blue, and green labels on the FEPs in panel B correspond to the values of the CGDAs at 198.5, 200.9, and 201.4 ns, respectively (typical transient unfolding event shown in the insets of Figure 2).

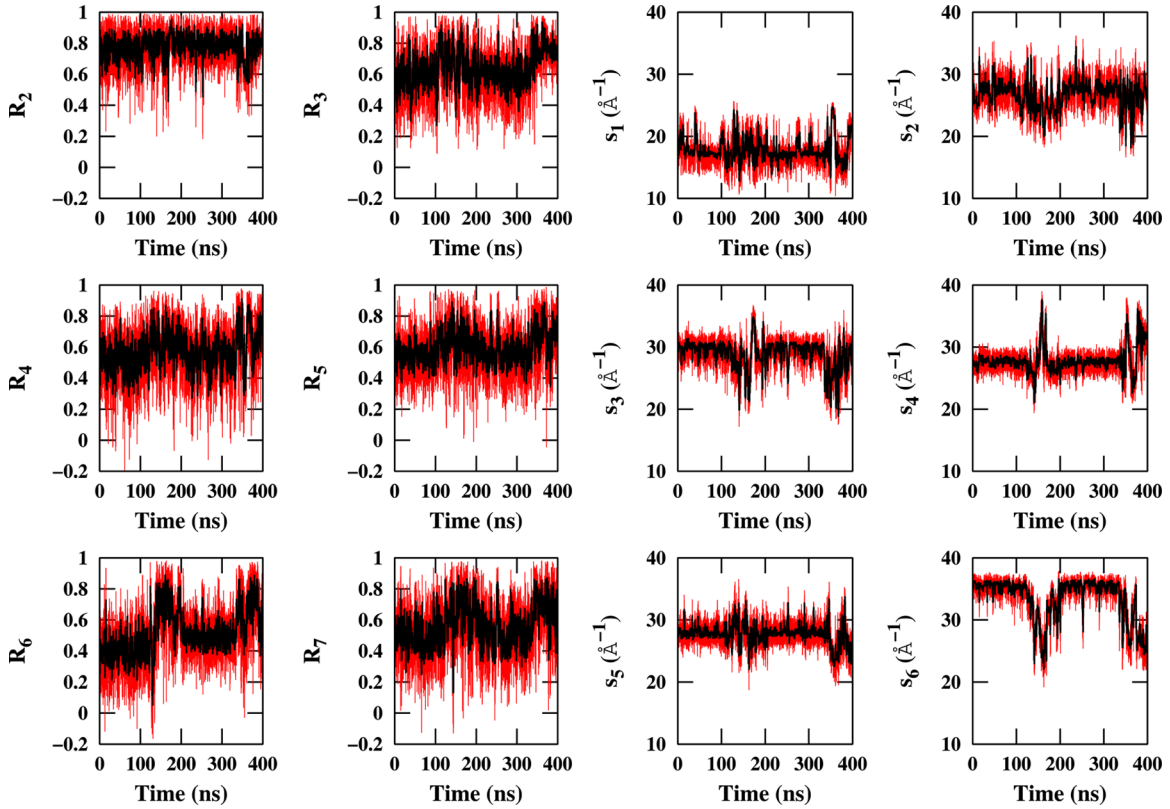


Figure 4. Correlation factor, R, and steric parameter, s_n, as a function of time at 380 K for selected CGDAs. In each panel, the red curve represents a calculation on a sliding window of 100 ps, and the black curves are a mobile average computed over 500 ps.

most probable extended structures of the Trp-cage (rmsd = 0.67 nm) as the initial structure of a new trajectory of 20 ns at 300 K. The values of R averaged over the first nanosecond of this trajectory (before the Trp-cage has time to fold to a non-native compact structure (empty circles in Figure 1 A)) are very close to those computed for the entire trajectory at 450 K (green diamonds in Figure 1 A). The values of R averaged over longer time (10 and 20 ns) remain around 0.7. (See Figure S3 in the SI.) Thus, the conformations of the folded/unfolded Trp-cage are encompassed in the variation of the correlation coefficient along the sequence at both 380 (folding/unfolding) and 450 K (unfolded). These variations at the two temperatures reflect the particular role of each CGDA in the folding/unfolding events.

DOI: 10.1021/acs.jpclett.5b00055
J. Phys. Chem. Lett. 2015, 6, 1082−1086

To relate the variation of R to the local constraints on the SCs and to reveal the disordered segments of the Trp-cage, the free-energy profiles (FEPs) of the CGDAs were computed from their probability density in the MD trajectories.[18] The FEPs $V(\gamma_n)$ and $V(\delta_n)$ are shown in Figure 3 at 300 (panel A) and 380 K (panel B), where they are compared with the values of the CGDAs γ_n and δ_n computed from the NMR-derived models.[5] Good agreement is observed between the NMR-derived values of γ_n and of δ_n and the minima of the FEPs in the native state (Figures 3 A). As expected from Figure 2, at 380 K, the protein explores regions in which the α-helix (residues 2−8) and the 3/10 helix (residues 11−14) are unfolded (Figure 3 B) (points located far from the minima of the FEPs).

At 380 K, the SCs (n = 4−8, 13, 16), for which R increases drastically compared with its value at 300 K (Figure 1 A), are those that are less constrained at 380 K than at 300 K; that is, they have wider FEPs at 380 K than at 300 K (compare Figure 3B with Figure 3A). The SCs at n = 14 and 15 have fewer constraints in the native state; that is, they have wide FEPs at 300 K (Figure 3 A), and their correlation coefficients barely change with an increase in temperature, as shown in Figure 1 A. The CGDAs built on Pro residues at the C-terminus of the Trp-cage (n = 17 and 18) are highly correlated (Figure 1 A) because these SCs are covalently linked to the MC. In summary, at 380 K the release of constraints on the MC, that is, the flatness of the FEPs at 380 K (Figure 3 B), increases the time average correlation between the motions of the SCs and of the backbone compared with its value at 300 K. (See Figure 1 A.) The FEPs of all CGDAs (except n = 16−19 built on Pro residues) are flat at 450 K (Figure 3C), indicating that the protein is unfolded at this temperature, as was expected from the FEP as a function of the rmsd (Figure S1 in the SI) and from the average values of s_n (Figure S2 in the SI). The loss of the secondary structures at 450 K releases the constraints imposed by the hydrogen bonds in the MC in the native state, which explains the significant increase in the correlation between the SCs and MC at 450 K compared with 300 K (Figure 1 A). It should be noted that R_{17} and R_{18} have similar values between 300 and 450 K (Figure 1 A) because they are built on rigid Pro residues.

Because the Trp-cage folds and unfolds as a function time, one expects a time dependence of the correlation coefficient, R. Indeed, the correlation coefficient averaged over all of the CGDAs as a function of time at 380 K increases transiently when the secondary structures are lost (Figure S4 in the SI). To identify the residues responsible for the transient unfolding at 380 K (Figure 2), we computed the steric parameter, s_n, and the correlation coefficient, R, as a function of time for selected CGDAs in Figure 4 and for all CGDAs in Figures S5 and S6 in the SI. The large structural fluctuations of the Trp-cage at 380 K (Figure 2) coincide with the large motion of SC_6 away from the protein and consequently with an increase of in the correlation coefficient at these times because SC_6 is more flexible: s_6 decreases between 130 and 200 ns and after 350 ns and the mobile average of R_6 increases up to 0.8 at these times. (See s_6 and R_6 (black curve) in Figure 4.) The motion of Trp_6 during these periods of time influences all of the other SCs at the same times; namely, significant variations of s_n and R are found for n = 2−5 (Figure 4), n = 13, 14, 16, and 19 (Figures S5 and S6 in the SI). It should be noted that the orientation and the rotameric state of Trp_6 SC play a crucial role in the folding of the Trp-cage.[19−21] Fluctuations of the Pro SCs at the C-terminus were also observed independent of the motion of Trp6. For example, s_{17} increases at t = 20 ns when s_{18} does not vary and s_{19} decreases (Figure S6 in the SI).

In summary, a high correlation between SC and MC fluctuations is observed during protein unfolding events (at 380 K) and in the unfolded (unstructured) state (at 450 K) of the protein. An unstructured very flexible state is one of the main properties of IDPs.[22] The characterization and the structural definition of IDPs is an active field of research with potential medical applications.[23] Some IDPs, such as α-synuclein,[24] are indeed involved in neurodegenerative diseases. Because of their flexibility, it is possible to extend the present results to IDPs, and we may expect large values of the correlation factor R for most of the residues of the protein. In fact, R (Figures 1 and 4) and s_n (Figure 4, Figures S2 and S6 in the SI) might be good descriptors of the dynamics of IDPs and more generally of unfolded proteins, and a high correlation between the SCs and the MC may facilitate the binding-induced folding.

This research was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB10-19767), and the PARI NANO2BIO (Région Bourgogne) and was conducted by using the HPC resources from DSI-CCUB (Université de Bourgogne).

■ ASSOCIATED CONTENT

Supporting Information

Results of additional trajectories and methodological details. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*G.G.M.: E-mail: gm56@cornell.edu.
*H.A.S.: E-mail: has5@cornell.edu.
*P.S.: E-mail: psenet@u-bourgogne.fr.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Matheson, R. R.; Scheraga, H. A. A Method for Predicting Nucleation Sites for Protein Folding Based on Hydrophobic Contacts. *Macromolecules* **1978**, *11*, 819−829.
(2) Liwo, A.; Khalili, M.; Scheraga, H. A. Ab Initio Simulations of Protein-folding Pathways by Molecular Dynamics with the United-Residue Model of Polypeptide Chains. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362−2367.
(3) Korkut, A.; Hendrickson, W. A. A Force Field for Virtual Atom Molecular Mechanics of Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15667−15672.
(4) Cote, Y.; Senet, P.; Delarue, P.; Maisuradze, G. G.; Scheraga, H. A. Anomalous Diffusion and Dynamical Correlation Between the Side Chains and the Main Chain of Proteins in their Native State. *Proc. Natl. Acad. of Sci. U.S.A.* **2012**, *109*, 10346−10351.
(5) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.
(6) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.
(7) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.
(8) Pitera, J. W.; Swope, W. Understanding Folding and Design: Replica-exchange Simulations of "Trp-cage" Miniproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587−7592.
(9) Schug, A.; Herges, T.; Verma, A.; Lee, K. H.; Wenzel, W. Comparison of Stochastic Optimization Methods for All-Atom

DOI: 10.1021/acs.jpclett.5b00055
J. Phys. Chem. Lett. 2015, 6, 1082−1086

Folding of the Trp-Cage Protein. *ChemPhysChem* **2005**, *6*, 2640−2646.
(10) Zhou, R. Trp-cage: Folding Free Energy Landscape in Explicit Water. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280−13285.
(11) Paschek, D.; Nymeyer, H.; Garcia, A. E. Replica Exchange Simulation of Reversible Folding/unfolding of the Trp-cage Miniprotein in Explicit Solvent: On the Structure and Possible Role of Internal Water. *J. Struct. Biol.* **2007**, *157*, 524−533.
(12) Paschek, D.; Hempel, S.; Garcia, A. E. Computing the Stability Diagram of the Trp-cage Miniprotein. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17754−17759.
(13) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. J. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(14) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712−725.
(15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(16) Pearson, K. Mathematical Contributions to the Theory of Evolution. III Regression, Heredity and Panmixia. *Philos. Trans. R. Soc., A* **1896**, *187*, 253−318.
(17) Nemethy, G.; Phillips, D. C.; Leach, S. J.; Scheraga, H. A. A Second Right-handed Helical Structure with the Parameters of the Pauling-Corey a-helix. *Nature* **1967**, *214*, 363−365.
(18) Senet, P.; Maisuradze, G. G.; Foulie, C.; Delarue, P.; Scheraga, H. A. How Main-chains of Proteins Explore the Free-Energy Landscape in Native States. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19708−19713.
(19) Day, R.; Paschek, D.; Garcia, A. E. Microsecond Simulations of the Folding/Unfolding Thermodynamics of the Trp-Cage Miniprotein. *Proteins* **2010**, *78*, 1889−1899.
(20) Chowdhury, S.; Lee, M. C.; Duan, Y. Characterizing the Rate-Limiting Step of Trp-Cage Folding by All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. B* **2004**, *108*, 13855−13865.
(21) Kannan, S.; Zacharias, M. Role of Tryptophan Side Chain Dynamics on the Trp-Cage Mini-Protein Folding Studied by Molecular Dynamics Simulations. *PLoS One* **2014**, *9*, e88383.
(22) Dyson, H. J.; Wright, P. E. Coupling of Folding and Binding for Unstructured Proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54−60.
(23) Salmon, L.; et al. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132*, 8407−8418.
(24) Eliezer, D. In *Parkinson's Disease*; Nass, R., Przedborski, S., Eds.; Elsevier, Inc.: Boston, 2008; pp 575−595.

DOI: 10.1021/acs.jpclett.5b00055
J. Phys. Chem. Lett. 2015, 6, 1082−1086