Published in final edited form as:

J Phys Chem Lett. 2015 March 19; 6(6): 1082–1086. doi:10.1021/acs.jpclett.5b00055.

New Insights into Protein (Un)Folding Dynamics

Yoann Cote¹, Gia G. Maisuradze^{2,§}, Patrice Delarue¹, Harold A. Scheraga^{2,*}, and Patrick Senet^{1,2,†}

¹Laboratoire Interdisciplinaire Carnot de Bourgogne UMR 6303 CNRS-Université de Bourgogne, 9 Av. Alain Savary, BP 47 870, F-21078 Dijon Cedex, France

²Baker Laboratory of Chemistry and Chemical Biology, Cornell University Ithaca, New York 14853

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.

Correspondence to: Gia G. Maisuradze; Harold A. Scheraga; Patrick Senet.

[§]gm56@cornell.edu

^{*}has5@cornell.edu

[†]psenet@u-bourgogne.fr

Supporting Information

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

Keywords

Unfolded state; correlations between the motions of the side chains and main chain; molecular dynamics simulations; steric parameter; Intrinsically Disordered Proteins

In order 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 affect 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, 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 were analyzed 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 in order to define the CGDA δ_n .⁴

The ultrafast folder Trp-cage⁵ 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 in order 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 μ s.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 implicit⁷⁻⁹ or explicit¹⁰⁻¹² 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 root-mean-square-deviation (RMSD) between the C_α positions at a given time at their corresponding positions in the initial structure. To characterize these fluctuations, a new quantity, named the steric parameter s_n , has been defined. It measures the distance of the n^{th} 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 outwards or towards the protein. All MD calculations were performed at several constant temperatures and pressure (1 bar) by using the GROMACS package¹³ and the Amber99SB forcefield¹⁴ with the TIP3P force field¹⁵ 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 1A). It is low for CGDAs built only on residues located inside secondary structures ($n = 3$ to 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 ($n = 14$ to 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 instability¹⁷ of the 3/10 helix.

Increase of temperature has a significant impact on the time average correlation coefficient R as shown in Figure 1A. 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 ps (Figure 2). There are large fluctuations (larger than 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 insets of Figure 2). These large structural changes are the main cause of an increase of 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 and B). At 450 K, the time average correlation coefficient R between the SCs and the MC is nearly equal for all the 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) 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) shows that s_n decreases for all residues (except at $n=8$ and 13) compared to 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 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 ns of this trajectory [before the Trp-cage has time to fold to a non-native compact structure (empty circles in Figure 1A)] are very close to those computed for the entire trajectory at 450 K (green diamonds in Figure 1A). The values of R averaged over longer time (10 and 20 ns) remain around 0.7 (see Figure S3). Thus, the conformations of the folded/unfolded Trp-cage are encompassed in the variation of the correlation coefficient along the sequence both at 380 K (folding/unfolding) and 450 K (unfolded). These variations at the two temperatures reflect the particular role of each CGDA in the folding/unfolding events.

In order 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 K (panel A) and at 380 K (panel B) where they are compared to 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 3A). As expected from Figure 2, at 380 K, the protein explores regions in which the α -helix (residues 2 to 8) and the 3/10 helix (residues 11 to 14) are unfolded (Figure 3B) (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 to its value at 300 K (Figure 1A), are those which are less constrained at 380 K than at 300 K, i.e. they have wider FEPs at 380 K than at 300 K (compare Figure 3B to Figure 3A). The SCs at $n = 14$ and $n = 15$ have fewer constraints in the native state, i.e. they have wide FEPs at 300 K (Figure 3A), and their correlation coefficients barely change with an increase of the temperature as shown in Figure 1A. The CGDAs built on Pro residues at the C-terminus of the Trp-cage ($n = 17$ and 18) are highly correlated (Figure 1A) because these SCs are covalently linked to the MC. In summary, at 380 K the release of constraints on the MC, i.e. the flatness of the FEPs at 380 K (Figure 3B), increases the time average correlation between the motions of the SCs and of the backbone compared to its value at 300 K (see Figure 1A). 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) and from the average values of s_n (Figure S2). 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 of the correlation between the SCs and MC at 450 K compared to 300 K (Figure 1A). It

should be noted that R_{17} and R_{18} have similar values between 300 and 450 K (Figure 1A) 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 the CGDAs as a function of time at 380 K increases transiently when the secondary structures are lost (Figure S4). In order to identify the residues responsible for the transient unfolding at 380 K (Figure 2), the steric parameter s_n and the correlation coefficient R were computed as a function of time for selected CGDAs in Figure 4 and for all CGDAs in Figures S5 and S6. The large structural fluctuations of the Trp-cage at 380 K (Figure 2) coincide with the large motion of SC₆ away from the protein and consequently with an increase of the correlation coefficient at these times because SC₆ 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 Trp₆ during these periods of time influences all the other SCs at the same times, namely, significant variations of s_n and R are found for $n = 2$ to 5 (Figure 4), $n = 13, 14, 16$ and 19 (Figures S5 and S6). It should be noted that the orientation and the rotameric state of Trp₆ 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 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, as α -synuclein²⁴ 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, S2 and S6) 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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).

References

1. Matheson RR, Scheraga HA. A Method for Predicting Nucleation Sites for Protein Folding Based on Hydrophobic Contacts. *Macromolecules*. 1978; 11:819–829.

2. Liwo A, Khalili M, Scheraga HA. Ab Initio Simulations of Protein-folding Pathways by Molecular Dynamics with the United-Residue Model of Polypeptide Chains. *Proc Natl Acad Sci US A*. 2005; 102:2362–2367.
3. Korkut A, Hendrickson WA. A Force Field for Virtual Atom Molecular Mechanics of Proteins. *Proc Natl Acad Sci US A*. 2009; 106:15667–15672.
4. Cote Y, Senet P, Delarue P, Maisuradze GG, Scheraga HA. Anomalous Diffusion and Dynamical Correlation Between the Side Chains and the Main Chain of Proteins in their Native State. *Proc Natl Acad Sci US A*. 2012; 109:10346–10351.
5. Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue Protein. *Nat Struct Biol*. 2002; 9:425–430. [PubMed: 11979279]
6. Qiu L, Pabit SA, Roitberg AE, Hagen SJ. Smaller and Faster:3 The 20-Residue Trp-Cage Protein Folds in 4 μ s. *J Am Chem Soc*. 2002; 124:12952–12953. [PubMed: 12405814]
7. Simmerling C, Strockbine B, Roitberg AE. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J Am Chem Soc*. 2002; 124:11258–11259. [PubMed: 12236726]
8. Pitera JW, Swope W. Understanding Folding and Design: Replica-exchange Simulations of “Trp-cage” Miniproteins. *Proc Natl Acad Sci US A*. 2003; 100:7587–7592.
9. Schug A, Herges T, Verma A, Lee KH, Wenzel W. Comparison of Stochastic Optimization Methods for All-Atom Folding of the Trp-Cage Protein. *Chem Phys Chem*. 2005; 6:2640–2646. [PubMed: 16331731]
10. Zhou R. Trp-cage: Folding Free Energy Landscape in Explicit Water. *Proc Natl Acad Sci US A*. 2003; 100:13280–13285.
11. Paschek D, Nymeyer H, Garcia AE. 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. [PubMed: 17293125]
12. Paschek D, Hempel S, Garcia AE. Computing the Stability Diagram of the Trp-cage Miniprotein. *Proc Natl Acad Sci US A*. 2008; 105:17754–17759.
13. Hess B, Kutzner C, van der Spoel D, Lindahl EJ. 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. [PubMed: 16981200]
15. 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.
16. Pearson K. Mathematical Contributions to the Theory of Evolution. III Regression, Heredity and Panmixia. *Phil Trans R Soc Lond A*. 1896; 187:253–318.
17. Nemethy G, Phillips DC, Leach SJ, Scheraga HA. A Second Right-handed Helical Structure with the Parameters of the Pauling-Corey α -helix. *Nature*. 1967; 214:363–365. [PubMed: 6040611]
18. Senet P, Maisuradze GG, Foulie C, Delarue P, Scheraga HA. How Main-chains of Proteins Explore the Free-energy Landscape in Native States. *Proc Natl Acad Sci US A*. 2008; 105:19708–19713.
19. Day R, Paschek D, Garcia AE. Microsecond Simulations of the Folding/unfolding Thermodynamics of the Trp-cage Miniprotein. *Proteins*. 2010; 78:1889–1899. [PubMed: 20408169]
20. Chowdhury S, Lee MC, 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. [PubMed: 24563686]
22. Dyson HJ, Wright PE. Coupling of Folding and Binding for Unstructured Proteins. *Curr Opin Struct Biol*. 2002; 12:54–60. [PubMed: 11839490]
23. Salmon L, et al. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J Am Chem Soc*. 2010; 132:8407–8418. [PubMed: 20499903]
24. Eliezer, D. Parkinson's Disease. Nass, R.; Przedborski, S., editors. Elsevier Inc; 2008. p. 575-595.

FIG. 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 grey and dark grey stripes, respectively. A) Average of R over the entire duration of the trajectory at 300K (red diamonds), at 380 K (blue diamonds), and at 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 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 cut-off values are the average RMSD at 300 K (0.21 nm) and the average RMSD at 450 K (0.67 nm).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIG. 2.

Time evolution of the RMSD of the C_{α} atoms at every ps, 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 K (horizontal black solid line), at 300 K (lower horizontal dashed line) and at 450 K (upper horizontal dotted-dashed line) are shown for comparison.

FIG. 3.

Effective FEPs $V(\gamma_n)$ (thick black lines) and $V(\delta_n)$ (thin black lines) along the aminoacid sequence of the Trp-cage (γ_n , $n = 2$ to 18 and δ_n , $n = 1$ to 19) computed as in.¹⁸ FEPs computed from a 400 ns MD trajectory performed at 300 K (A) and 380 K (B) and 450 K (C). All the FEPs are compared to the NMR-derived structural data computed from the 1L2Y PDB files (blue diamonds γ_n and red diamonds δ_n). Residues 2 to 8 form an α -helix and residues 11 to 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).

FIG. 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.