J|A|C|S
COMMUNICATIONS

Published on Web 11/24/2004

Existence of Specific "Folds" in Polyproline II Ensembles of an "Unfolded" Alanine Peptide Detected by Molecular Dynamics

Vibin Ramakrishnan,[§] Ranjit Ranbhor,[§] and Susheel Durani*[,#]

Department of Chemistry and School of Biosciences and Bioengineering, IIT Bombay, Mumbai-400076, India

Received July 14, 2004; E-mail: sdurani@iitb.ac.in

Several recent studies with alanine model peptides have altered dramatically the understanding of the nature of "unordered" peptides and "unfolded" proteins, proving them, in conformity with a long-standing suggestion of Tiffany and Krimm,[1] to be at least partially "ordered" with backbone dihedral angles corresponding predominantly to the left-handed polyproline II (PPII) helix ($\phi \sim -75$ and $\psi \sim 145$). Apparently, the seeds of this PPII-type preference, against the traditional notion of random sampling over α-helical ($\phi \sim -60$ and $\psi \sim -50$) and β-regions ($\phi \sim -135$ and $\psi \sim 145$) of residue level conformational space, are embedded within residue level properties of unfolded peptides in water. This was proven by the observation of strong PPII preference in alanine dipeptides (Ac−Ala−NHMe) based on density functional calculations,[2] NMR,[3] and MD,[4−6] in trialanine model peptides based on vibrational spectroscopy[7] and MD,[8] and in a longer alanine model peptide based on UV−Raman.[9] Although Poon et al.[3] and Shi et al.,[10] on the basis of the results of Han et al.,[2] interpreted PPII preference to be rooted in solvation effect, Pappu and Rose,[11] followed by Drozdov et al.,[12] have demonstrated conclusively that the preference is primarily due to a sterically advantageous "minimization" of interatomic contacts in both alanine dipeptide and heptaalanine models. Longer alanine models with >20 residues have for several years[13] been known to populate in predominantly α-helical states in water. As cooperative "two-state" systems, these peptides populate in folded and unfolded states in equilibrium, with the former "maximized" in intrachain contacts and the latter "maximized" in conformational entropy, as envisaged in so-called "funnel models" of protein folding.[14] As statistical descriptions of folding landscapes, these models have contributed much in highlighting the role of conformational "disorder" as an "expense" in folding free energy and as a "barrier" to folding speed. Since there may well be specific "nonrandom" points from which folding actually initiates, the suggestion of Tiffany and Krimm[1] is with a fundamental bearing on both thermodynamics and kinetics of protein folding. Recent NMR and CD evidence has confirmed that shorter peptides, incapable of "cooperative" ordering, are actually appreciably ordered PPII-like ensembles in water,[9,15] while UV−Raman experiments suggest that a 21-mer α-helical alanine peptide may actually melt into a predominantly PPII-like ensemble.[9] Notably, Shi et al.[10] argue that their model, Ac−$X_2Ala_7O_2$−amide [diaminobutyric acid (X) and ornithine (O) residues are for solubility reasons], is a PPII-like ensemble to at least 90% at 275 K, devoid of even nascent-type α-helices, and in fact displaying further increased β-type content at 325 K. The prevailing understanding of so-called "random coil" ensembles of unfolded peptides and "denatured" proteins is thus brought into serious question with regard to both its statistical and structural descriptions. A few years ago, van Gunsteren and co-workers[16] developed an RMSD-based clustering algorithm for enumeration of equilibrium ensembles prepared by MD and found that the ensemble-averaged properties extracted for a number of small, especially β-peptide, models across several solvents were generally in excellent agreement with NMR data.[17] Here, we use van Gunsteren's clustering algorithm and evaluate equilibrium ensembles of small alanine models, prepared by MD, for their contributing microstates. We find that while the extracted ensemble-averaged properties for an octaalanine model peptide are in excellent agreement with the conclusions from most of above-quoted experiments, the existence of a small number of dynamical but discrete folds are detected and reflect that there could be significant involvement of residues in α-type conformational basins, not easily detectable spectroscopically due to the problems in the accurate assessment of ensemble-averaged ψ torsional angles, highlighting the urgent need for experimental methodologies for their evaluation.

Table 1. Percent Occupancies of Largest Four ϕ, ψ Basins and Ensemble Averaged Torsional Angles ϕ and ψ[a]

peptide	ensemble	% PPII	% β	% α_R	% α_L	$\langle\phi\rangle, \langle\psi\rangle$
Ac−$(Ala)_1$−NHMe	298 K, (25 ns)	40	38	7	3	−87, +109
Ac−$(Ala)_8$−NHMe	298 K, (40 ns)	35	22	13	13	−56, +84
Ac−(Ala)8-NHMe	373 K, (10 ns)	28	24	18	9	−65, +70

[a]Defined in Table S1.

MD trajectories of Ac−Ala_8−NHMe in water, initiated from α-helical and extended β-conformations, generated with GROMOS96 force field[18] in the GROMACS[19] package, were assessed for achievement of equilibrium at both macroscopic levels, in the occupancies of specific Ramachandran basins, and microscopic levels, in the time-dependent accumulation of microstates, as assessed by clustering[16] using a 0.15 nm cutoff. Experiments at 298, 323, and 373 K over 8 ns each established that the relative occupancies of α_R, α_L, β, and PPII basins (Tables S1 and S2 of the Supporting Information) and the progress curves for microstate accumulation during the last 5 ns of each trajectory (Figure S1 of the Supporting Information) were dependent on the choice of the starting structures at 298 and 323 K, but not at 373 K. Prima facie, equilibrium ensembles are approximated in 373 K trajectories but not in the lower-temperature trajectories. Extending the 298 K trajectory initiated in β-conformation and assessing it periodically, we achieved an equilibrium ensemble in this trajectory as well in ∼40 ns (Figure S2 of the Supporting Information). The basin occupancy and microstate accumulation data for 298 and 373 K trajectories are compared in Table 1 and Figures S1 and S2 (Supporting Information). The basin occupancy data were also gathered for Ac−Ala−NHMe in water over 25 ns at 298 K (Figure S3 of the Supporting Information and Table 1). The combined occupancy of β and PPII basins is 78%, while the PPII basin ($\phi = -67$ and $\psi = 130$) is the clear global minimum, in agreement with Han et al.[2] The ensemble-averaged dihedral angles are $\phi = -87$ and $\psi = 109$ (Table 1), as compared with those of the NMR estimates $\phi = -85 \pm 5$ and $\psi = 160 \pm 20$ of Poon et al.[3] The ensemble-averaged ϕ dihedral angles for octaalanine are −56 at 298 K and −65 at 373 K, in excellent agreement with the value

[#] Department of Chemistry.
[§] School of Biosciences and Bioengineering.

10.1021/ja045787y CCC: $27.50

Table 2. Position-Specific Percent Occupancies of Indicated ϕ, ψ Basins in 298 K Equilibrium Ensemble of Octaalanine

	Ala_1	Ala_2	Ala_3	Ala_4	Ala_5	Ala_6	Ala_7	Ala_8
% PPII	44	38	29	28	28	25	35	56
% β	36	18	25	16	7	25	26	21
% α_R	7	22	9	18	23	15	3	5
% α_L	2	9	9	17	19	17	24	9

-70 ± 10 extracted by Shi et al.[10] for their peptide by fitting the $^3J_{NH\alpha}$ to a well-validated Karplus equation. Our ensemble-averaged ψ angles for this system, 84 at 298 K and 70 at 373 K (Table 1), are, on the other hand, significantly lower than the value 145 ± 20, which Shi et al. estimate by indirect reasoning based on NMR and CD data. The large divergence of ψ values may have two origins. First, with its positively charged end residues (X_2 and O_2) at neutral pH, the peptide model of Shi et al. adopted for solubility reasons could be less susceptible to folding and, therefore, less prone toward α_R or α_L basins, which are required for any kind of folding. Second, although there were no helix-specific NOEs observed in the Shi et al. model peptide, a higher involvement of isolated α_R-type residues cannot be entirely ruled out since the ϕ values are identical for α_R and PPII, and any reduction in ψ due to this reason cannot be detected by NMR, unless there are specific folds that populate the ensemble in sufficient concentration for the NOEs to become observable. The reduction of ensemble-averaged ψ from 109 in alanine dipeptide to 84 and 70 in octaalanine at 298 and 373 K, respectively, while the overall ϕ remains PPII-like, is due to an increased occupancy of the α_R basin, from 7% in alanine dipeptide at 298 K to 13 and 18% in octaalanine at 298 and 373 K, respectively (Table 1). Thus, there is an appreciable tendency in octaalanine toward folding. The residue-by-residue basin occupancy data at 298 K (Table 2) reflect that the sampling of the α_R and α_L basins in octaalanine is highly varied. It deviates significantly from the ensemble-averaged values, generally "up" in Ala(4), Ala(5), and Ala(6), but "down" in Ala(1) and Ala(8), suggesting tendencies in these residues for or against localized folding. The clustering analysis at the chosen RMSD cutoff establishes that a rather modest number of similar looking folds of dynamical nature populate the ensemble as a consequence of interactions between the dipeptide units in octaalanine. The largest cluster, identical in 298 and 373 K ensembles (their central members are in <0.05 nm RMSD of each other), comprising ~18% of the 298 K ensemble and ~8% of the 373 K ensemble, is a 2:2 β-hairpin of conventional right-handed twist with a type I′ turn centered on its Ala_4 and Ala_5, accounting for the unusually high occupancy of the α_L basin in these positions. It is noteworthy that with a characteristic ladderlike feature of three hydrogen bonds, as illustrated in Figure S6A (Supporting Information), this β-hairpin is stereochemically equivalent to the most abundant β-hairpin found in protein structures.[20,21] In addition to its canonical interstrand hydrogen bonds, the hairpin features one unusual hydrogen bond between its $-NHMe$ and $-COCH_3$ groups (Figure S6A of the Supporting Information), causing its N-terminal end to mount over the C-terminal end, dragging Ala_2 into the α_R basin and Ala_7 into the α_L basin, accounting for the unusual basin occupancies at these positions, as well (Table 2). Punctuated invariably with one or more α_R- or α_L-type residues, a majority of the next 13 lowest-energy clusters (Figure 1 and Figure S5 and Tables S3 and S4 of the Supporting Information), comprising overall 43% of the 298 K ensemble, are turns, hairpins, or helix nuclei of varied morphologies in overall PPII-like conformation, due to hydrogen-bonding interactions between individual residues in octaalanine. The remarkable position specificity of these folds, even while there is no "sequence context", seems to be promoted by maximization of intramolecular hydrogen bonding in octaalanine. Overall, the 298 and 373 K ensembles populate in 300 and 374 clusters, respectively, which is a modest number against the 1.7×10^4 combinations one might expect over only the four conformational basins that were considered in this study. The existence of specific folds in an unfolded alanine peptide, including inverse turns, β-hairpins, and helix nuclei, suggests that similar sequence context induced folds are likely to occur in "molten globule" or "framework"-like states in unordered peptides and denatured proteins[22] and may be the "seeds" that initiate proteins along their folding pathways.

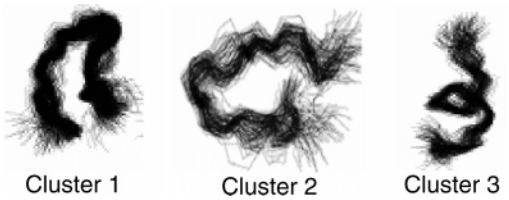

Figure 1. Three most populous clusters in 298 K equilibrium ensemble of octaalanine in water, comprising 18, 6, and 3% of the ensemble, respectively.

Acknowledgment. This work was supported by grants from the Board of Research in Nuclear Sciences (BRNS) and Council of Scientific and Industrial Research (CSIR), Government of India.

Supporting Information Available: Materials and methods, definition and percentage population of ϕ, ψ basins, residue-wise occupancies of ϕ, ψ basins, free energy calculations, stereoviews, and ϕ, ψ values of the 14 most populous clusters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tiffany, M. L.; Krimm, S. *Biopolymers* **1968**, *6*, 1379−1382.
(2) Han, W. G.; Jalkanen, K. J.; Elstner, M.; Suhai, S. *J. Phys. Chem. B* **1998**, *102*, 2587−2602.
(3) Poon, C. D.; Samulski, E. T.; Weise, C. F.; Weishaar, J. C. *J. Am. Chem. Soc.* **2000**, *122*, 5642−5643.
(4) Tobias, D. J.; Brooks, C. L., III. *J. Phys. Chem.* **1992**, *96*, 3864−3870.
(5) Apostolakis, J.; Ferrara, P.; Caflisch, A. *J. Chem. Phys.* **1999**, *110*, 2099−2108.
(6) Smith, P. E. *J. Chem. Phys.* **1999**, *111*, 5568−5578.
(7) Woutersen, S.; Hamm, P. *J. Phys. Chem. B.* **2000**, *104*, 11316−11320.
(8) Mu, Y.; Stock, G. *J. Phys. Chem. B.* **2002**, *106*, 5294−5301.
(9) Asher, S. A.; Mikonin, A. V.; Bykov, S. *J. Am. Chem. Soc.* **2004**, *126*, 8433−8440.
(10) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190−9195.
(11) Pappu, R. V.; Rose, G. D. *Protein Sci.* **2002**, *11*, 2437−2455.
(12) Drozdov, A. N.; Grossfield, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2004**, *126*, 2574−2581.
(13) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286−5290.
(14) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.
(15) Rucker, A. L.; Creamer, T. P. *Protein Sci.* **2002**, *11*, 980−985.
(16) Daura, X.; Gademann, K.; Jann, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(17) Daura, X.; Glattli, A.; Gee, P.; Peter, C.; van Gunsteren, W. F. *Adv. Protein Chem.* **2002**, *62*, 341−360.
(18) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, Switzerland, 1996.
(19) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.
(20) Sibanda, B. L.; Thornton, J. M. *Nature* **1985**, *316*, 170−174.
(21) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759−777.
(22) Baldwin, R. L. *Adv. Protein Chem.* **2002**, *62*, 361−367.

JA045787Y