J|A|C|S
COMMUNICATIONS

Published on Web 08/17/2006

Simulated Unfolded-State Ensemble and the Experimental NMR Structures of Villin Headpiece Yield Similar Wide-Angle Solution X-ray Scattering Profiles

Bojan Zagrovic*,[†] and Vijay S. Pande*,[‡]

*Physical Chemistry Institute, ETH Zurich, 8093 Zurich, Switzerland, and Department of Chemistry,
Stanford University, Stanford, California 94305*

Received June 9, 2006; E-mail: zagrovic@igc.phys.chem.ethz.ch; pande@stanford.edu

With the advent of powerful synchrotron sources, solution X-ray scattering is being increasingly used to get basic information about the structure of proteins in the native aqueous milieu.[1] In addition to analysis of the long-range structure of polypeptides (such as their radius of gyration), solution scattering has also been used to study interdomain correlation and intradomain geometry, including the secondary structural content of some proteins.[2] However, it should be emphasized that the technique provides essentially one-dimensional data which can be interpreted in terms of a three-dimensional structure only through model building and comparison.[1] It is possible that several structural models agree equally well with a given solution scattering pattern. Furthermore, particular care needs to be taken in the process of data interpretation in order to avoid problems related to the dynamical nature of proteins and issues of conformational averaging.

To analyze how diverse two ensembles can be and still yield similar solution wide-angle scattering patterns, we calculate here scattering profiles for an ensemble of simulated unfolded structures and the native experimental NMR structural ensemble of villin headpiece.[3] Using worldwide distributed computation techniques, we generated thousands of long (tens of nanoseconds) trajectories of villin headpiece. All simulations were initiated from the extended conformation ($\phi = -135°$, $\psi = 135°$) with *N*-acetyl and C-amino caps, each started with a different random number seed. The simulations, carried out using the Tinker simulation package, were implemented using Langevin dynamics in implicit GB/SA solvent[4a] (friction coefficient $\gamma = 91$ ps^{-1}, to match that of water) with a 2 fs integration step, at 300 K, using the OPLS-UA force field.[4b] The analysis given here is carried out on an ensemble consisting of 5200 structures at the 27 ns time point. As this time is 2 orders of magnitude shorter than the folding time of villin headpiece,[3b] our ensemble can be thought of as a model for the unfolded state of the molecule under folding conditions, that is, in the absence of a denaturant. The simulated ensemble collapses to locally converged radius of gyration, solvent accessible surface area, secondary structure content, and rmsd from the native structure in about 20 ns.[4d,e] After this, the exact time point or the size of the analyzed ensemble has no significant effect on our conclusions. Further details about the simulation setup are given elsewhere.[4d,e]

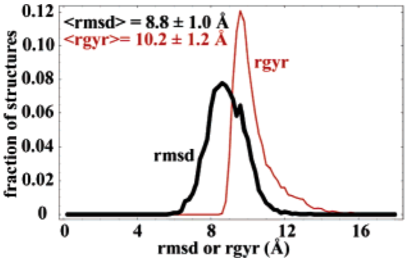

Figure 1. Distributions of all-atom root mean square deviation (rmsd) from the average experimental NMR structure (in black) and of the radius of gyration, rgyr (in red), for the simulated unfolded ensemble of villin headpiece.

How diverse and non-native is the simulated unfolded state ensemble? Figure 1 shows the distribution of the all-atom root mean square deviation (rmsd) from the average experimental NMR structure of villin headpiece[3a] (PDB code 1VII) for all the members of the unfolded state ensemble. The unfolded state ensemble shares very little similarity with the native structure on the level of tertiary structure, as demonstrated by the large value of $\langle\text{rmsd}\rangle = 8.8 \pm 1.0$ Å. In addition, the unfolded state ensemble has negligible levels of native secondary structure as well; while there are 19 α-helical residues in the native villin structure, the average number of α-helical residues per member of the unfolded ensemble is only 3.4 ± 3.6, according to DSSP.[4d,5] If, as a definition of α-helicity, we take a range of backbone dihedral angles surrounding the average α-helical values,[6] $\phi = -62 \pm 20°$ and $\psi = -41 \pm 20°$, this number rises to 4.8 ± 2.2, again fairly low. Overall, less than 1% of residues in the simulated unfolded ensemble on average can be classified as belonging to the same DSSP secondary structure category as the equivalent residue in the experimental native structure.[4d] However, it is important to emphasize that the unfolded ensemble is as compact as the native structure in terms of its average radius of gyration ($\langle\text{rgyr}\rangle = 10.2 \pm 1.2$ Å vs 9.5 Å for the native structure) and solvent-accessible surface area ($\langle\text{SASA}\rangle = 3123 \pm 193$ Å^2 vs 3221 Å^2 for the native structure). More details about the characterization of the simulated unfolded state ensemble are given elsewhere.[4d,e]

Solution X-ray scattering has extensively been used to study the chemically denatured state of proteins.[1] Due to its fleeting nature, much less is known about the solution scattering properties of the unfolded state of proteins *under folding conditions*. This is particularly true for the ultrafast folding proteins, such as villin headpiece. In Figure 2, we show the solution wide-angle scattering profiles calculated for the experimental NMR structural ensemble of villin headpiece (Figure 2a) and the simulated unfolded state ensemble (Figure 2b). Scattering profiles were calculated using the Debye formula:[7a]

$$I(q) = \sum_i \sum_j f_i f_j \frac{\sin(qr_{ij})}{qr_{ij}}$$

where f_i and f_j are the atomic scattering factors of the *i*th and the *j*th atoms, respectively, calculated for each atom type using Cromer−Mann coefficients;[7b] r_{ij} is the distance between atoms *i*

[†] ETH Zurich.
[‡] Stanford University.

10.1021/ja0640694 CCC: $33.50 © 2006 American Chemical Society

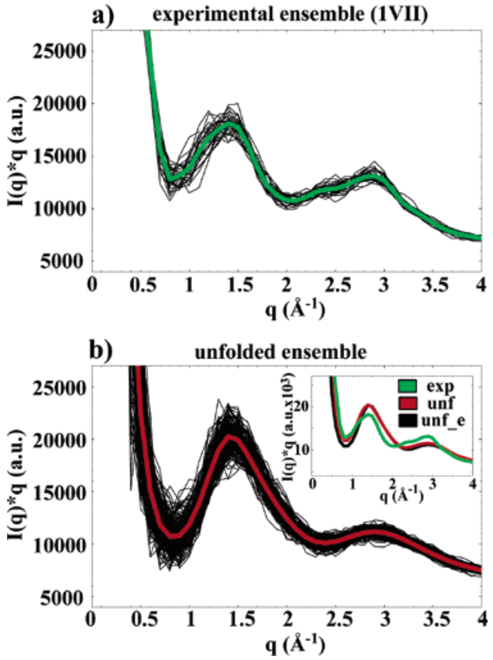

Figure 2. Calculated solution scattering patterns for the experimental NMR ensemble (a) and the simulated unfolded ensemble (b). In (b), scattering curves for only 200 randomly chosen members of the ensemble are shown for clarity. Average curves for complete ensembles are shown in green (exp) and red (unf), respectively, and are for clarity reproduced in the inset in (b). The average scattering curve for the extended, noncompact members of the simulated unfolded state ensemble (rgyr > 15 Å) is shown in the inset in black (unf_e): it largely overlaps with the unfolded ensemble curve (red).

and j, and q is the scattering vector defined as

$$q = \frac{4\pi \sin(2\theta)}{\lambda}$$

where 2θ is the scattering angle and λ is the radiation wavelength. The scattering profiles were calculated for $q_{\max} = 4$ Å^{-1} using the step size of 0.1 Å^{-1}. The scattering profiles were calculated for each member of the ensemble separately and then averaged.

How do the two scattering profiles compare? The principal features in the two profiles are similar across a range of scattering vector values. In particular, the dominant peaks around 1.4 and 2.9 Å^{-1}, as well as the valleys around 0.8 and 2.0 Å^{-1}, are equally prominent in both profiles, with minor offsets from one another. The relative intensities of the peaks are comparable in the two ensembles, with the main difference being that the 2.9 Å^{-1} peak is less prominent in the case of the unfolded state ensemble.

Muroga[2b,c] has developed an analytical theory for solution scattering of α-helices and showed how a 1.4 Å^{-1} peak is a signature motif of such profiles. This peak has also been observed in scattering profiles of several exclusively α-proteins.[8] Our result suggests that it is possible that the 1.4 Å^{-1} peak (corresponding to distances of ~4.5 Å) arises from other repetitive arrangements in the peptide besides α-helices. More importantly, our results suggest that it is not at all necessary that a given protein has a defined three-dimensional structure with stable side-chain contacts or even stable secondary structure for this peak, or the valley around 0.8 Å^{-1}, to arise. The agreement of the two curves for higher values of the scattering vector is easier to understand, as this region is dominated by structural features on the length scale of a single residue or less.

To test the effects of compaction induced by hydrophobic collapse on our conclusions, we have calculated an average scattering curve for a small subset of the simulated unfolded ensemble consisting of all the structures whose radius of gyration is greater than 15 Å (a total of 26 structures). Interestingly, this curve (Figure 2b inset, black curve) superimposes well with the average scattering curve calculated for the entire unfolded state ensemble. The reason for this is that, for the q values analyzed here, scattering is not sensitive to the long-range structure of molecules, while structural features on the length scale of 5 Å or less are similar in both compact and noncompact members of the ensemble. Regarding the utility of using chemically denatured proteins to model unfolded proteins under folding conditions, one of the problems is that denatured proteins may not be as compact as the unfolded proteins under folding conditions.[9] We should note that our results, as evidenced in Figure 2b (inset), do not provide significant clues either way because of the short length scale that we focus on.

Recently, we have shown that the average Cα−Cα distance matrix based on the simulated unfolded ensemble of villin is native-like (“the mean-structure hypothesis”).[4d] It is possible that the native-like scattering curves calculated here are a consequence of the same property. However, native-like scattering curves are also obtained for the relatively noncompact members of the ensemble (Figure 2b, inset), for which the average distance matrix is not native-like. This suggests that there need not be direct connection between the two findings. However, the fact remains that the scattering profile of a highly non-native ensemble of structures is similar to the one based on the experimental NMR structure of the molecule. This result should serve as a caveat demonstrating that solution scattering in the wide-angle limit, by itself, provides very little information about the secondary structure content of a polypeptide or its side-chain packing.[2a]

Acknowledgment. B.Z. acknowledges support of an EMBO postdoctoral fellowship. This work was supported by grants from the NSF (MCB-0317072), NIH (GM062868), and NSF MRSEC CPIMA (DMR-9808677).

References

(1) Doniach, S.*Chem. Rev.* **2001**, *101*, 1763−1778.

(2) (a) Hirai, M.; Koizumi, M.; Hayakawa, T.; Takahashi, H.; Abe, S.; Hirai, H.; Miura, K.; Inoue, K. *Biochemistry* **2004**, *43*, 9036−9049. (b) Muroga, Y. *Biopolymers* **2000**, *54*, 58−63. (c) Muroga, Y. *Biopolymers* **2001**, *59*, 320−329. (d) Fischetti, R. F.; Rodi, D. J.; Mirza, A.; Irving, T. C.; Kondrashkina, E.; Makowski, L. J. *J. Synchron. Radiat.* **2003**, *10*, 398−404. (e) Svergun, D. I.; Petoukhov, M. V.; Koch, M. H. J. *Biophys. J.* **2001**, *80*, 2946−2953.

(3) (a) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. *Nat. Struct. Biol.* **1997**, *4*, 180−184. (b) Kubelka, J.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2003**, *329*, 625−630.

(4) (a) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014. (b) Jorgensen, W. L.; Tirado-Rives J. *J. Am. Chem. Soc.* **1988**, *110*, 1666−1671. (c) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24−34. (d) Zagrovic, B.; Snow, C. D.; Khaliq, S.; Shirts, M. R.; Pande, V. S. *J. Mol. Biol.* **2002**, *323*, 153−164. (e) Zagrovic, B.; Snow, C. D.; Shirts, M. R.; Pande, V. S. *J. Mol. Biol.* **2002**, *323*, 927−937.

(5) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

(6) Creighton, T. E. *Proteins*, 2nd ed.; W. H. Freeman: New York, 1992.

(7) (a) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; W. H. Freeman: New York, 1980; Part II. (b) International Tables for Crystallography, Vol. C, Tables 6.1.1.4 and 6.1.1.5; 2004.

(8) Hirai, M.; Iwase, H.; Hayakawa, T.; Miura, K.; Inoue, K *J. Synchron. Radiat.* **2002**, *9*, 202−205.

(9) (a) Richards, F. M., Eisenberg, D. S., Rose, G. D., Kuriyan, J., Eds.; Unfolded Proteins. In *Advances in Protein Chemistry*; Academic Press: San Diego, CA, 2002. (b) Li, Y.; Picart, F.; Raleigh, D. P. *J. Mol. Biol.* **2005**, *349*, 839−46. (c) Tran, H. T.; Pappu, R. V. *Biophys. J.* **2006**, 16766618.

JA0640694