A Self-Consistent Description of the Conformational Behavior of Chemically Denatured Proteins from NMR and Small Angle Scattering

Pau Bernadó[†*] and Martin Blackledge[‡*]

[†]Institute for Research in Biomedicine, c/ Baldiri Reixac, Barcelona, Spain; and [‡]Protein Dynamics and Flexibility, Institut de Biologie Structurale, UMR 5075, Commissariat à l'Énergie Atomique-Centre National de la Recherche Scientifique-Université Joseph Fourier, Grenoble, France

ABSTRACT Characterization of the conformational properties of unfolded proteins is essential for understanding the mechanisms of protein folding and misfolding. This information is also fundamental to determining the relationship between flexibility and function in the highly diverse families of intrinsically disordered proteins. Here we present a self-consistent model of conformational sampling of chemically denatured proteins in agreement with experimental data reporting on long-range distance distributions in unfolded proteins using small-angle x-ray scattering and nuclear magnetic resonance pulse-field gradient-based measurements. We find that standard statistical coil models, selected from folded protein databases with secondary structural elements removed, need to be refined to correct backbone dihedral angle sampling of denatured proteins, although they appear to be appropriate for intrinsically disordered proteins. For denatured proteins, pervasive increases in the sampling of more-extended regions of Ramachandran space $\{50^\circ < \psi < 180^\circ\}$ throughout the peptide chain are found to be consistent with all experimental data. These observations are in agreement with previous conclusions derived from short-range nuclear magnetic resonance data from residual dipolar couplings, leading the way to a self-consistent description of denatured chains that is in agreement with short- and long-range data measured using both spectroscopic and scattering techniques.

INTRODUCTION

Despite the volume of structural information describing stable protein structures to high atomic resolution, the unfolded state occupies a conformational landscape that remains comparatively poorly understood. Development of an appropriate description of the conformational sampling of denatured proteins is essential for the study of protein folding, whereas the steadily increasing recognition of the biomedical relevance of natively unfolded proteins, for example in cancer-related signaling and neurodegenerative diseases (1–4), reinforce the need for relevant models of disordered molecular ensembles. Many intrinsically disordered proteins (IDPs) only adopt a stable conformation upon binding physiological partners (5–7), underlining the importance of understanding protein folding events occurring in the unfolded state.

Two models have been identified to explain the effects of chemical denaturants on folded proteins. The first invokes the different hydrogen-bonding properties of the denaturant relative to water in the vicinity of hydrophobic groups, such that it acts as a better solvent for hydrophobic groups that are normally buried in water-solvated proteins (8). An alternative model proposes that the denaturant binds directly to the protein backbone, presumably via hydrogen-bonding interaction with the amide groups, thereby destabilizing hydrogen-bonding networks that provide the origin of the stability of the native fold relative to the unfolded state (9). Both models stabilize the unfolded state in the presence of high concentrations of denaturant. The correct description of conformational sampling in the denatured state will surely provide crucial information on the molecular origins of chemical denaturation, thereby contributing to our understanding of protein folding and stability.

It has become clear over the last decade that rather than conforming to a simple random coil description, the conformational behavior of the unfolded protein can be extremely varied. The key question is not whether a protein is unfolded, rather how it is unfolded. To clarify these questions, conformational sampling in denatured proteins has been studied in detail using small-angle scattering (10,11), nuclear magnetic resonance (NMR) spectroscopy (12), and fluorescence correlation spectroscopy (13), as well as ensemble molecular simulation (14–19). Experimental and theoretical studies suggest that apparently diverse descriptions based on diverse experimental techniques probably derive from the sensitivity of the different approaches to local and long-range structure, underlining the strong complementarity of NMR and small-angle scattering techniques for the development of a self-consistent molecular description of the denatured state.

A popular approach used to analyze the conformational properties of disordered proteins is the use of explicit molecular descriptions of unfolded ensembles (17–22). To this end, explicit atomic-resolution ensemble models have been proposed on the basis of statistical coil sampling, extracted from backbone dihedral angles present in coil regions of protein databases. Amino-acid-specific behavior from these conformational distributions is then assumed to represent a Boltzmann distribution of dihedral angles in a statistical description of the unfolded polymer, such that an ensemble

Submitted July 16, 2009, and accepted for publication August 31, 2009.

*Correspondence: pau.bernado@irbbarcelona.org or martin.blackledge@ibs.fr

Editor: Kathleen B. Hall.

0006-3495/09/11/2839/7 $2.00

doi: 10.1016/j.bpj.2009.08.044

of conformers constructed on the basis of random sampling of these distributions can be considered to describe a representative ensemble. For these descriptions to yield useful information, they should be demonstrated to be accurate and relevant representations of the unfolded state. Ab initio and molecular dynamics simulations have indeed suggested that such approximations can be valid (23). In most cases, long-range interactions are reduced to a minimum, steric repulsion or excluded volume simplification, essentially to avoid chain overlap, with electrostatic and amino-acid-specific solvent effects ignored in the interest of efficiency. Nearest-neighbor effects can also be introduced to refine conformational behavior as a function of the nature of the neighboring amino acids.

One such approach, termed Flexible Meccano (FM; (24)), constructs conformational ensembles in this way. The algorithm uses a rudimentary spherical side-chain volume exclusion model, 20-amino-acid-specific conformational wells derived from coil regions of 500 high-resolution crystal structures, and additional conformational basins representing specific dihedral angle distributions such as pre-prolines, or prolines preceding prolines. The model has been validated against numerous types of experimental data: NMR residual dipolar couplings (RDCs) measured between covalently bound spins (24), reporting on orientational properties of internuclear vectors; NMR RDCs measured between protons (25), reporting on both distance and angular internuclear distribution functions; three-bond scalar couplings, reporting directly on backbone dihedral angles distributions (25); and small-angle scattering measurements, reporting on overall pairwise distribution functions and molecular dimensions (24). In addition, state-of-the-art molecular simulation technology, specifically the accelerated molecular dynamics approach (26,27) that allows for statistical sampling of the broad conformational subspace available to unfolded proteins, and critically, the reconstruction of Boltzmann-like sampling on an amino-acid-specific basis, reveals a very close reproduction of the conformational space described by the statistical coil database (28,29). These experimental and theoretical considerations allow us to propose that, although necessarily approximate, the FM approach provides a reasonable tool to allow the refinement of our understanding of the conformational properties of unfolded proteins in general and of denatured proteins in particular.

Partially aligned NMR spectroscopy has been shown to be a particularly powerful tool for studying molecular structure in both folded (30–32) and unfolded states (33–37). In recent years, we have used this parameter as a probe to derive an amino-acid-specific self-excluding model of the unfolded state. Explicit molecular ensemble analysis, using FM, successfully reproduced both spectroscopic and scattering data from protein X from Sendai virus, a protein that is natively partially unfolded (24), and RDCs from both apomyoglobin (34) and staphylococcal nuclease Δ131Δ (12) in high concentration of denaturant. This was subsequently used to identify conformational properties of several important IDPs involved in neurodegenerative diseases, the proteins τ (28) and α-synuclein (38), the tumor-suppressor p53 (29), and the unfolded C-terminal of the nucleoprotein of Sendai virus (39,40). Using the same conformational averaging approach, we have also recently developed further tools for refining the interpretation of small-angle x-ray scattering (SAXS) data from unfolded proteins in terms of average molecular dimensions (41) and the average surface accessibility of amino acids in unfolded proteins (42).

In a more recent study, up to seven RDCs per peptide unit could be determined for urea-unfolded ubiquitin, both between covalently bound and ^{1}H-^{1}H RDCs (25), allowing a detailed description of the angular and distance fluctuations of the unfolded state. The aim in this initial study was the calculation of complete conformational distributions of the unfolded or partially folded polypeptide chain that is in agreement with the available experimental data, resulting in the clear identification of enhanced sampling of two regions of Ramachandran space, in particular β-extended and polyproline II conformations. However, RDCs only report indirectly on the overall dimensions of the protein, thus a realistic description of the unfolded state should also reproduce experimental data reporting more directly on the overall dimensions of the protein. In this study, we demonstrate that this refined coil description provides a consistent description of the conformational behavior of chemically denatured proteins that is compatible with measurement of NMR RDCs, pulsed-field gradient (PFG) hydrodynamic measurements, and SAXS data.

METHODS

The conformations used in this study to calculate the properties of the denatured state were built with Flexible Meccano (FM). To define the amino-acid-specific Ramachandran space, FM uses ϕ/ψ dihedral angles collected from 500 high-resolution (<1.8 Å) x-ray structures (43) where residues belonging to secondary structure elements were removed; this is the so-called standard (coil) library. Side chains for these conformations were added with the program SCCOMP (44).

SAXS properties of the different polypeptide chains were obtained from the averaging the individual scattering curves of 2000 atomic resolution conformations of each of the 23 proteins defined by their amino-acid sequence. The theoretical scattering profiles ($s < 0.1$ Å^{-1}) for each conformation were calculated with the program CRYSOL (45), and averaged to yield a final SAXS curve for each protein. Random Gaussian noise of an amplitude of 2% was added to the curves to produce more experimental-like data. R_g values were derived by analyzing the initial part of the curve ($sR_g < 1.3$) using Guinier's approach with the program PRIMUS (46).

Hydrodynamic properties of the denatured state were computed for eight proteins whose R_h values had been measured by PFG (47,48). For each of the eight proteins, an ensemble with 500 conformations was built as described above, and for each conformation the hydrodynamic radius was computed with a modified version of the program HYDROPRO (49) using an atomic element radius of 3.3 Å to define the volume of the conformation (50). The bead modeling strategy adopted by HYDROPRO has been proven to be efficient to calculate hydrodynamic properties for both rigid (51,52) and flexible (53) molecules. The averaged R_h was obtained as

$$\frac{1}{R_h} = \frac{1}{N}\sum_{i=1,N}\frac{1}{R_{h,i}}, \quad (1)$$

where N is the number of conformations of the ensemble (500) and $R_{h,i}$ is the hydrodynamic radius of the conformation i.

The same calculations of SAXS and hydrodynamic properties were repeated for six additional ensembles with an enhanced population of extended conformations $\{50^\circ < \psi < 180^\circ\}$ in the Ramachandran space. Databases with $X = 1.5, 2.0, 2.5, 3.0, 3.5$, and 4.0 times' higher populations for this range of ψ were created. Explicitly the same conformations with $\{50^\circ < \psi < 180^\circ\}$ were repeated X times in the database to increase their possibility of being selected. Final databases have 68.3%, 74.2%, 78.2%, 81.2%, 83.4%, and 85.2% of the conformations in the extended region, respectively.

RESULTS AND DISCUSSION

The overall dimensions of chemically denatured proteins, expressed in terms of an effective ensemble radius of gyration R_g, has been predicted and then shown to display a simple power-law dependence on chain length N that is in agreement with random-coil polymer behavior (54,55),

$$R_g = R_0 N^{\nu}, \tag{2}$$

where R_0 is a constant that depends on the persistence length and ν is the exponential scaling factor. In a recent article, Kohn et al. measured and compiled the R_g obtained by SAXS for 28 proteins under different denaturing conditions, extracting a value of $\nu = 0.58 \pm 0.11$ (11). We have taken a set of 23 proteins from this list covering a broad range of chain lengths, from eight to 549 amino acids (see Table S1 in the Supporting Material). For each of the 23 amino-acid sequences, an ensemble of 2000 conformations was created using FM, and SAXS profiles were calculated from these ensembles (see Methods). The correlation between the experimental and simulated R_g values is shown in Fig. 1 *a*. A systematic deviation is observed, indicating that simulated ensembles underestimate the overall size of the denatured state.

Hydrodynamic measurements such as analytical ultracentrifugation, sedimentation coefficients, or translational diffusion coefficients have been used to obtain the hydrodynamic radius, R_h, for unstructured proteins (56,57). We have calculated the effective ensemble averaged hydrodynamic characteristics of eight proteins for which experimental data are available. The correlation between experimental and simulated averaged values is shown in Fig. 1 *b*. A systematic deviation on the correlation is observed when increasing the size of the polypeptide chain. This observation is coincident with the SAXS data and indicates an underestimation of the global structural parameters of the disordered chain under denaturing conditions when using the coil library to describe their conformational sampling.

The recent NMR study of urea-unfolded ubiquitin identified the need for an enhanced population of extended conformations, defined as a higher propensity to populate the region of Ramachandran space defined by $\{50^\circ < \psi < 180^\circ\}$ (25). If true, this enhanced extendedness should be reflected in the overall dimensions of denatured proteins and therefore should be detectable by SAXS and hydrodynamic measurements. The overall population of extended conformations was

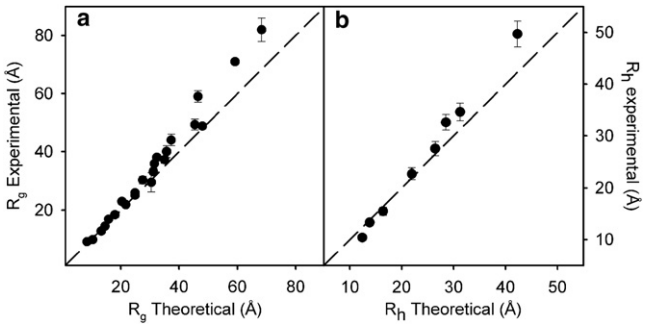

FIGURE 1 Correlation between experimental and theoretical values of (*a*) radius of gyration R_g, and (*b*) hydrodynamic radius R_h for the ensembles corresponding to the standard database. Dashed line represents slope equal to 1.0. Experimental values of R_g were obtained from Kohn et al. (11), and R_h values were obtained from Wilkins et al. (47) and Pan et al. (48).

systematically increased for each amino-acid-specific population in the database, by creating databases with $X = 1.5$, 2.0, 2.5, 3.0, 3.5, and 4.0 times higher populations for this range (see Methods). Taking into account that the coil database has 59.0% of conformations that are in the extended region, the new databases have 68.3%, 74.2%, 78.2%, 81.2%, 83.4%, and 85.2% of the conformations in the extended region, respectively. For each of the six new databases new ensembles of the same size for all proteins were computed and the R_g and R_h values were derived as explained for the standard (coil) database. The capacity of the ensembles to describe experimental R_g and R_h values was monitored via a reduced χ^2 value computed as

$$\chi^2 = \frac{1}{NP} \sum_{i=1,NP} \frac{\left(R_{x,i}^{\exp} - R_{x,i}^{\text{theo}}\right)^2}{\sigma_i^2}, \tag{3}$$

where $R_{x,i}$ is either R_g or R_h for protein i, NP is the total number of proteins for the SAXS (23) and for the PFG data (8). The resulting χ^2 values for each database are shown in Fig. 2. Interestingly, for both experimentally measured parameters, R_g and R_h, the minimum value of the χ^2 function is obtained for the 2.0 database. R_g shows more sensitivity to the degree of extendedness than R_h that has a very flat χ^2 profile in the 68–78% range (1.5–2.5 databases). The presence of larger proteins in the dataset, and the different averaging processes governing the measured parameters, squared distance distribution functions compared to reciprocal, are possible reasons of this higher sensitivity found for R_g. The correlations between the experimental and simulated R_g and R_h values using the optimal database are displayed in Fig. 2, *c* and *d*, where a good agreement is observed. In this optimal database, the extended conformations represent 74.2% of the conformers—a significant enhancement when compared to the coil library, 59%. At this degree of extendedness, no systematic deviations are observed in the correlation plots, and both small peptides and large proteins are well described when using this database. Importantly, this optimal

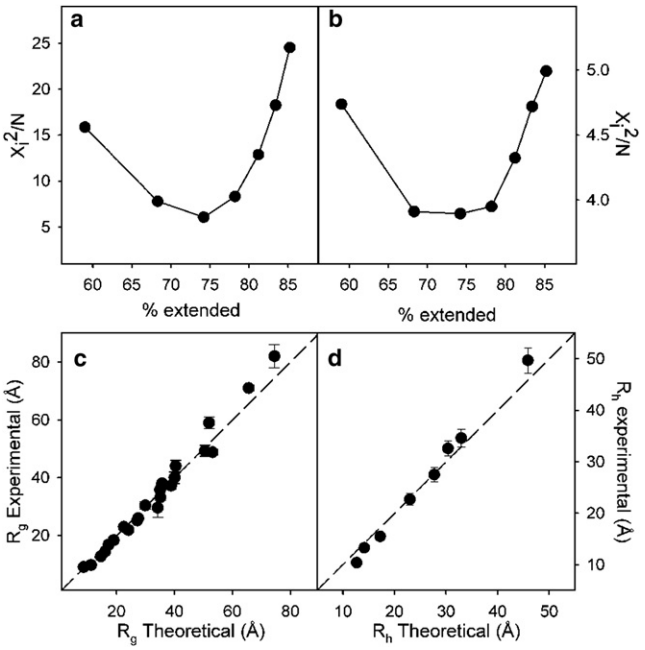

FIGURE 2 Optimal percentage of extended conformations. Agreement of the overall parameters (*a*) R_g and (*b*) R_h upon increasing the percentage of extended conformations in the database. Correlation between the experimental and theoretical (*c*) R_g and (*d*) R_h at the optimal database found, $X = 2.0$, which corresponds to a 74.2% of extended conformations. Dashed line again represents slope equal to 1.0.

value of extendedness found by fitting overall parameters was also found when fitting a large dataset of residue-specific RDCs measured for urea-denatured ubiquitin (Fig. 3). It is also important to note that using this more extended conformational sampling gave very poor agreement for the RDCs ($^1D_{HN}$ and $^2D_{CHN}$) measured in the intrinsically disordered protein X (25). Conversely, the standard coil database provides excellent agreement of both RDCs measured for the same protein. These results strongly suggest the existence of distinct conformational sampling in intrinsically and chemically denatured proteins that probe more extended conformations.

It is interesting to compare the level of extension with similar databases extracted using comparable approaches. Removing sheet and helical regions from a compilation of high-resolution structures resulted in the development of a database with 55% of conformers in the extended region (58), while similar results were found by Griffiths-Jones et al. (54% in this region excluding glycine) (59), and from two similarly constructed earlier, smaller coil databases (60,61). Compared to 59% in our database, these numbers are therefore slightly less extended, indicating that the current findings—that more extended conformers are necessary for chemically denatured proteins—are generally relevant for coil database descriptions.

Our results clearly suggest that parameters describing Flory's relationship (Eq. 2) for chemically denatured proteins are different to these describing IDPs. A parameterization of

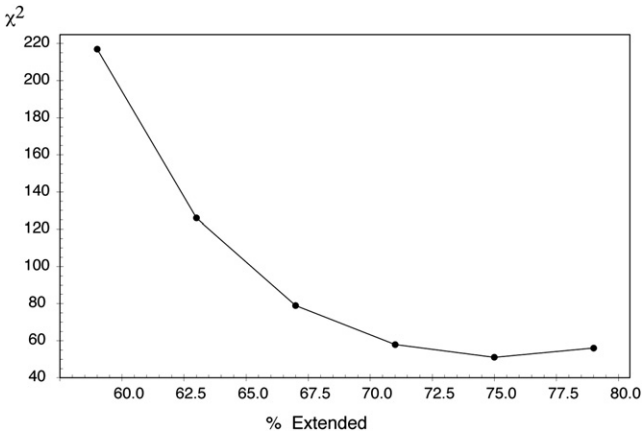

FIGURE 3 Reproduction of experimental residual dipolar couplings measured from urea unfolded ubiquitin at pH2 aligned in polyacrylamide gel using databases sampling different levels of extension. The χ^2 is measured over seven types of coupling: $^1D_{NH}$, $^1D_{C\alpha H\alpha}$, $^1D_{C\alpha C'}$, $D_{HNH\alpha\ (i,i-1)}$, $D_{HNHN\ (i,i+1)}$, $D_{HNHN(i,i+2)}$, and $D_{HNH\alpha}$. Data taken from Meier et al. (25).

Flory's equation for IDPs was obtained from the simulated R_g and R_h values using the coil database with the length of the biopolymer chains (see Fig. 4). From the fitting of the data to the logarithmic form of Eq. 2, ν-values of 0.522 ± 0.01 and 0.449 ± 0.01 for the R_g and R_h parameters were obtained, respectively, and the R_0 values derived were 2.54 ± 0.01 and 3.53 ± 0.01 for the R_g and R_h, respectively. The ν-values obtained from the R_g and R_h data are notably smaller than the 0.6 and 0.588 values predicted for a random coil polymer with excluded volume (62), and experimentally calibrated values that are also close to these numbers (10,11,51).

The R_g values measured for different constructs of protein τ (63) and other IDPs have been collected from literature (see Table S2), and plotted as a function of the protein length in Fig. 4 *c*. The points are scattered around the line representing the previously described parameterization for IDPs indicating a broad agreement with the general law. Departures from the parameterized line are expected for IDPs, as they can often observe structuring phenomena that will modify the R_g values expected for a nonstructured chain of N residues. R_g values measured for different τ-protein constructs (63), solid circles in Fig. 4 *c*, appear above the line indicating that the conformational bias increases the overall stiffness of the protein. This general signature is more evident for K18 and K19 τ constructs that encompass the so-called repeat domain in which extended and turn regions have been identified in different NMR studies (28,64). Conversely, N-tail of measles virus nucleoprotein, a 139-residue-long IDP, appears as a more compact ensemble than the calibrated threshold (65). The SAXS data for this protein presents a Kratky plot with a clear bump at $s = 0.08$ Å^{-1} that indicates a certain degree of compactness in a concrete region of the chain. A bioinformatics analysis predicts a 16-residue-long fragment to have a strong tendency to form an α-helix that is probably the origin of this partial compactation (66).

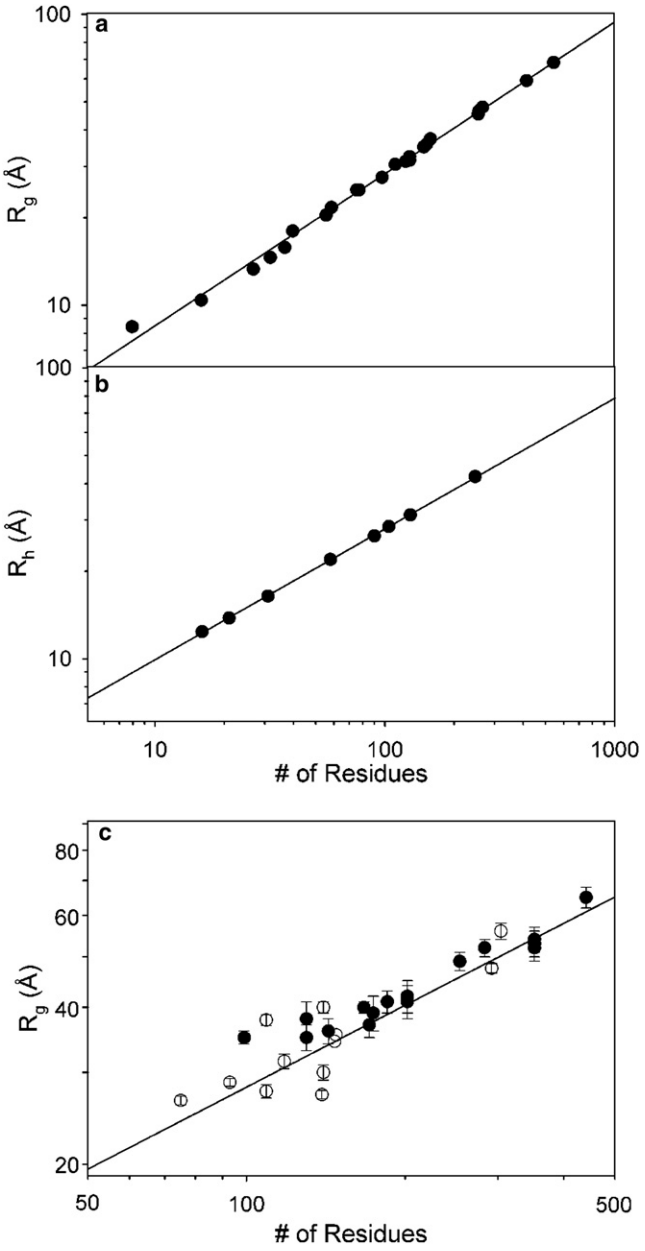

FIGURE 4 Relationship of the R_g (*a*) and R_h (*b*) values calculated using the standard database with the length of the protein chain *N*. From this correlation, the parameters of Flory's relationship (Eq. 1) were derived for both observables. From R_g, $\nu = 0.522 \pm 0.01$ and $R_0 = 2.54 \pm 0.01$; and from R_h, $\nu = 0.449 \pm 0.01$ and $R_0 = 3.53 \pm 0.01$. (*c*) R_g values compiled for protein τ (*solid circles*) and other IDPs (*open circles*), compared with the threshold values according to the calibration found in panel *a* (*solid line*). See the Supporting Material for the list of IDPs used in this plot.

CONCLUSIONS

In conclusion, by comparing explicit molecular simulation with extensive experimental data reporting on the overall dimensions of highly disordered chains, we are able to refine the coil database to a level that agrees simultaneously with different and highly complementary biophysical measurements of local and long-range structure. We find that Ramachandran sampling, described using popular definitions of the statistical coil model that simply remove secondary structural elements from high-resolution crystallographic structures, is insufficient to describe protein conformational sampling under strongly denaturing conditions. On the basis of small-angle scattering and NMR data, we define a self-consistent model that enhances access to β-extended and polyproline II regions of Ramachandran space within native amino-acid-specific conformational wells. It is interesting to speculate that this model agrees qualitatively with a mechanism of denaturation that implicates urea binding to the polypeptide backbone (9), causing steric hindrance that would favor more extended local conformations, although it certainly does not disprove the importance of the side-chain solvation model (8).

SUPPORTING MATERIAL

One table is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01436-2.

Authors acknowledge Prof. José García de la Torre (University of Murcia) for providing a modified version of HYDROPRO.

Financial support from the Spanish Ministry of Education-FEDER (grant No. BIO2007-63458) is gratefully acknowledged. P.B. holds a Ramón y Cajal contract that is partially financed by the Spanish Ministry of Education and the by funds provided to the Institute for Research in Biomedicine by the Generalitat de Catalunya. M.B. acknowledges support from grant No. ANR PCV 2008 PCV07-194985.

REFERENCES

1. Tompa, P. 2002. Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27:527–533.
2. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.
3. Fink, A. L. 2005. Natively unfolded proteins. *Curr. Opin. Struct. Biol.* 15:35–41.
4. Dyson, H. J., and P. E. Wright. 2004. Intrinsically unstructured proteins and their functions. *Chem. Rev.* 104:3607–3622.
5. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* 338:1015–1026.
6. Vacic, V., C. J. Oldfield, A. Mohan, P. Radivojac, M. S. Cortese, et al. 2007. Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Protein Res.* 6:2351–2366.
7. Wright, P. E., and H. J. Dyson. 2009. Linking folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 19:1–8.
8. Tanford, C. 1964. Isothermal unfolding of globular proteins in aqueous urea solutions. *J. Am. Chem. Soc.* 86:2050–2059.
9. Schellman, J. A. 1958. Factors affecting the stability of hydrogen-bonded polypeptide structures in solution. *J. Phys. Chem.* 62:1485–1494.
10. Millett, I. S., S. Doniach, and K. W. Plaxco. 2002. Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv. Protein Chem.* 62:241–262.
11. Kohn, J. E., I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, et al. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 101:12491–12496.

12. Shortle, D., and M. S. Ackerman. 2001. Persistence of native-like topology in a denatured protein in 8 M urea. *Science*. 293:487–489.

13. Lipman, E. A., B. Schuler, O. Bakajin, and W. A. Eaton. 2003. Single-molecule measurement of protein folding kinetics. *Science*. 301:1233–1235.

14. Smith, L. J., K. M. Fiebig, H. Schwalbe, and C. M. Dobson. 1996. The concept of the random coil. Residual structure in peptides and denatured proteins. *Fold. Des.* 1:R95–R106.

15. Pappu, R. V., R. Srinivasan, and G. D. Rose. 2000. The isolated-pair hypothesis is not valid for poly-peptide chains: implications for protein folding. *Proc. Natl. Acad. Sci. USA*. 97:12565–12570.

16. Goldenberg, D. P. 2003. Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* 326:1615–1633.

17. Fitzkee, N. C., and G. D. Rose. 2004. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. USA*. 101:12497–12502.

18. Ding, F., R. K. Jha, and N. V. Dokholyan. 2005. Scaling behavior and structure of denatured proteins. *Structure*. 13:1047–1054.

19. Tran, H. T., X. Wang, and R. V. Pappu. 2005. Reconciling sequence-specific conformational preferences with generic behavior of denatured proteins. *Biochemistry*. 44:11369–11380.

20. Zaman, M. H., M.-Y. Shen, R. S. Berry, K. F. Freed, and T. R. Sosnick. 2003. Investigations into sequence and conformational dependence of backbone entropy, inter-basin dynamics and the Flory isolated-pair hypothesis for peptides. *J. Mol. Biol.* 331:693–771.

21. Avbelj, F., and R. L. Baldwin. 2003. Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of ϕ. *Proc. Natl. Acad. Sci. USA*. 100:5742–5747.

22. Fitzkee, N. C., and G. D. Rose. 2005. Sterics and solvation winnow accessible conformational space for unfolded proteins. *J. Mol. Biol.* 353:873–887.

23. Mackerell, Jnr., A. D., D. Feig, and C. L. Brooks. 2003. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* 126:698–699.

24. Bernadó, P., L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok, et al. 2005. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. USA*. 102:17002–17007.

25. Meier, S., S. Grzesiek, and M. Blackledge. 2007. Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. *J. Am. Chem. Soc.* 129:9799–9807.

26. Hamelberg, D., J. Mongan, and J. A. McCammon. 2004. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys*. 120:11919–11929.

27. Markwick, P. R. L., G. Bouvignies, and M. Blackledge. 2007. Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. *J. Am. Chem. Soc.* 129:4724–4730.

28. Mukrasch, M. D., P. R. L. Markwick, J. Biernat, M. von Bergen, P. Bernadó, et al. 2007. Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc.* 129:5235–5243.

29. Wells, M., H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, et al. 2008. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. USA*. 105:5762–5767.

30. Tjandra, N., and A. Bax. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*. 278:1111–1114.

31. Prestegard, J. H., C. M. Bougault, and A. I. Kishore. 2004. Residual dipolar couplings in structure determination of biomolecules. *Chem. Rev.* 104:3519–3540.

32. Bouvignies, G., P. R. Markwick, R. Brüschweiler, and M. Blackledge. 2006. Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings. *J. Am. Chem. Soc.* 128: 15100–15101.

33. Louhivouri, M., K. Pääkkönen, K. Fredriksson, P. Permi, J. Lounila, et al. 2003. On the origin of residual dipolar couplings from denatured proteins. *J. Am. Chem. Soc.* 125:15647–15650.

34. Mohana-Borges, R., N. K. Goto, G. J. Kroon, H. J. Dyson, and P. E. Wright. 2004. Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.* 340:1131–1142.

35. Fieber, W., S. Kristjansdottir, and F. M. Poulsen. 2004. Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. *J. Mol. Biol.* 339:1191–1199.

36. Dames, S. A., R. Aregger, N. Vajpai, P. Bernadó, M. Blackledge, et al. 2006. Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids. *J. Am. Chem. Soc.* 128:13508–13514.

37. Meier, S., M. Blackledge, and S. Grzesiek. 2008. Conformational distributions of unfolded polypeptides from novel NMR techniques. *J. Chem. Phys.* 128:052204.

38. Bernadó, P., C. W. Bertoncini, C. Griesinger, M. Zweckstetter, and M. Blackledge. 2005. Defining long-range order and short-range disorder in native α-synuclein using residual dipolar couplings. *J. Am. Chem. Soc.* 127:17968–17969.

39. Jensen, M. R., K. Houben, E. Lescop, L. Blanchard, R. W. H. Ruigrok, et al. 2008. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.* 130:8055–8061.

40. Jensen, M. R., and M. Blackledge. 2008. On the origin of NMR dipolar waves in transient helical elements of partially folded proteins. *J. Am. Chem. Soc.* 130:11266–11267.

41. Bernadó, P., E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle x-ray scattering. *J. Am. Chem. Soc.* 129:5656–5664.

42. Bernadó, P., M. Blackledge, and J. Sancho. 2006. Sequence-specific solvent accessibilities of protein residues in unfolded protein ensembles. *Biophys. J.* 91:4536–4543.

43. Lovell, S. C., I. W. Davis, W. B. Arendall, III, P. I. W. de Bakker, J. M. Word, et al. 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins*. 50:437–450.

44. Eyal, E., R. Najmanovich, B. J. McConkey, M. Edelman, and V. Sobolev. 2004. Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J. Comput. Chem.* 25:712–724.

45. Svergun, D. I., C. Barberato, and M. H. J. Koch. 1995. CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst*. 28:768–773.

46. Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun. 2003. PRIMUS—a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.* 36:1277–1278.

47. Wilkins, D. K., S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, et al. 1999. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*. 38:16424–16431.

48. Pan, H., G. Barany, and C. Woodward. 1997. Reduced BPTI is collapsed. A pulsed field gradient NMR study of unfolded and partially folded bovine pancreatic trypsin inhibitor. *Protein Sci.* 6:1985–1992.

49. García de la Torre, J., M. L. Huertas, and B. Carrasco. 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78:719–730.

50. Bernadó, P., J. García de la Torre, and M. Pons. 2002. Interpretation of ^{15}N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J. Biomol. NMR*. 23:139–150.

51. García de la Torre, J., and V. A. Bloomfield. 1981. Hydrodynamic properties of complex, rigid, biological macromolecules. Theory and applications. *Q. Rev. Biophys.* 14:81–139.

52. Carrasco, B., and J. García de la Torre. 1999. Hydrodynamic properties of rigid particles. Comparison of different modeling and computational strategies. *Biophys. J.* 76:3044–3057.

53. García de la Torre, J., H. E. Pérez-Sánchez, A. Ortega, J. G. Hernández, M. X. Fernández, et al. 2003. Calculation of the solution properties of

flexible macromolecules: methods and applications. *Eur. Biophys. J.* 32:477–486.

54. Flory, P. J. 1953. Principles of Polymer Chemistry. Cornell University Press, Ithaca, NY.
55. Tanford, C., K. Kawahara, S. Lapanje, T. M. Hooker, Jr., M. H. Zarlengo, et al. 1967. Proteins as random coils. 3. Optical rotatory dispersion in 6 M guanidine hydrochloride. *J. Am. Chem. Soc.* 89:5023–5029.
56. Jones, J. A., D. K. Wilkins, L. J. Smith, and C. M. Dobson. 1997. Characterization of protein unfolding by NMR diffusion measurements. *J. Biomol. NMR.* 10:199–203.
57. Choy, W.-Y., F. A. A. Mulder, K. A. Crowhurst, D. R. Muhandiram, I. S. Millett, et al. 2002. Distribution of molecular size within an unfolded state ensemble using small-angle x-ray scattering and pulse field gradient NMR techniques. *J. Mol. Biol.* 316:101–112.
58. Jha, A. K., A. Colubri, M. H. Zaman, S. Koide, T. R. Sosnick, et al. 2005. Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry.* 44:9691–9702.
59. Griffiths-Jones, S. R., G. J. Sharman, A. J. Maynard, and M. S. Searle. 1998. Modulation of intrinsic φ,ψ propensities of amino acids by neighboring residues in the coil regions of protein structures: NMR analysis and dissection of a β-hairpin peptide. *J. Mol. Biol.* 284:1597–1609.
60. Swindells, M. B., M. W. MacArthur, and J. M. Thornton. 1995. Intrinsic ϕ,ψ propensities of amino-acids, derived from the coil regions of known structures. *Nat. Struct. Biol.* 2:596–603.
61. Smith, L. J., K. A. Bolin, H. Schwalbe, M. W. MacArthur, J. M. Thornton, et al. 1996. Analysis of main-chain torsion angles in proteins-prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* 255:494–506.
62. LeGuillou, J. C., and J. Zinn-Justin. 1977. Critical exponents for the *n*-vector model in three dimensions from field theory. *Phys. Rev. Lett.* 39:95–98.
63. Mylonas, E., A. Hascher, P. Bernadó, M. Blackledge, E. Mandelkow, et al. 2008. Domain conformation of τ-protein studied by solution small-angle x-ray scattering. *Biochemistry.* 47:10345–10353.
64. Mukrasch, M. D., J. Biernat, M. von Bergen, C. Griesinger, E. Mandelkow, et al. 2006. Sites of τ important for aggregation populate β-structure and bind to microtubules and polyanions. *J. Biol. Chem.* 280:24978–24986.
65. Longhi, S., V. Receveur-Bréchot, D. Karlin, K. Johansson, H. Darbon, et al. 2003. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J. Biol. Chem.* 278:18638–18648.
66. Bourhis, J.-M., V. Receveur-Bréchot, M. Oglesbee, X. Zhang, M. Buccellato, et al. 2005. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci.* 14:1975–1992.