Dimensional Scaling of the Large Majority of Chemically Unfolded Proteins Is Indistinguishable from Random Coil Behavior

J.E. Kohn,[1] I.S. Millett,[2] J. Jacob,[3,4] T.M. Dillon,[5]
N. Cingel,[5] S. Seifert,[6] P. Thiyagarajan,[4] T.R. Sosnick,[3,7]
M.Z. Hasan,[8] I. Ruczinski,[9] S. Doniach,[2] K.W. Plaxco[1,5]

[1]Interdepartmental Program in Biomolecular Science and Engineering, University of California, Santa Barbara, CA, U.S.A.
[2]Department of Applied Physics, Stanford University, Stanford, CA, U.S.A.
[3]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, U.S.A.
[4]Intense Pulsed Neutron Source (IPNS), Argonne National Laboratory, Argonne, IL, U.S.A.
[5]Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA, U.S.A.
[6]Advanced Photon Source (APS), Argonne National Laboratory, Argonne, IL, U.S.A.
[7]Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, U.S.A.
[8]Department of Physics, Princeton University, Princeton, NJ, U.S.A.
[9]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, U.S.A.

Introduction

In recent years, there have been numerous and compelling reports of residual denatured state structure in even the most highly denatured protein. For example, the significant formation of local secondary structure and long-range hydrophobic clustering has been observed in nonnuclear magnetic resonance (NMR)-based studies of proteins denatured in high concentrations of urea or guanidine hydrochloride (GuHCl) [1-5]. Similarly, recent reports have suggested that even the most highly denatured proteins exhibit residual long-range order similar to the native topology [6, 7]. This residual denatured-state structure is widely thought to play a significant role in the thermodynamics and kinetics of protein folding [8]. Thus a better understanding of its magnitude may be a key to understanding the folding process.

Because it represents a deviation from purely random-coil behavior, the significant residual structure should affect the average dimensions of the unfolded ensemble. A hallmark of random-coil behavior is a simple power-law relationship between a polymer's length and its ensemble average radius of gyration (R_g):

$$R_g = R_0 N^{\nu} \quad (1)$$

where N is the number of monomers in the polymer chain; R_0 is a "constant" that is a function of, among other things, the persistence length of the polymer; and ν is an exponential scaling factor that varies with solvent quality. For the compact, globular state, ν = 1/3. For an ideal (infinitely thin) random-coil chain in a "good" solvent, ν = 1/2. For an excluded volume polymer (i.e., a real polymer with nonzero thickness and nontrivial interactions between monomers) in a good solvent, Flory has estimated that ν expands to ~3/5, and more precise follow-on estimates stemming from renormalization group models indicate that ν = 0.588. The formation of persistent denatured-state structure, however, should lead to perturbations from these ideal behaviors. For example, given these arguments, the formation of hydrophobically stabilized clusters should lead to net compaction and a smaller exponent than the ν = 0.588 proposed for an inert, purely random-coil polymer. Similarly, the presence of persistent local structure could increase (or decrease) the polymer's mean persistence length, leading to expansion (or contraction) of the denatured state relative to unfolded states lacking such structure. Differences in the magnitude of the residual denatured-state structure from one unfolded protein to another would thus be expected to obscure the otherwise systematic power-law relationship between the dimensions and lengths across a diverse set of unfolded proteins.

Methods and Materials

We have determined the R_g's of nine new proteins and peptides and remeasured the R_g's of five previously characterized proteins under strongly denaturing conditions. We have also collected the R_g's of 10 additional cross-link-free and prosthetic group-free, chemically denatured proteins from the literature.

The Proteins

We have determined or redetermined the R_g's of 14 proteins and peptides under strongly denaturing conditions (Table 1). Expression and purification of these proteins are discussed elsewhere [9-12].

Excluding RNase A and the GCN4-p2′ terminal dimer, all cysteine-containing proteins were treated with 20 mM iodoacetamide and 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) for 2 hours at room temperature

TABLE 1. Experimental conditions employed here.

Protein/Peptide	Flow Cell	Beamline
Reduced RNase	1.5-mm	APS/Bio-CAT
Snase	1-mm	APS/BESSRC
α-TS	1-mm	APS/BESSRC
Creatine kinase	1-mm	APS/BESSRC
GCN4-p2'	1.5-mm	APS/Bio-CAT
OspA	1-mm	APS/BESSRC
ctAcP	1-mm	APS/Bio-CAT
Ubiquitin	1-mm	APS/Bio-CAT
Fyn SH3	1-mm	APS/BESSRC
pI3K SH3	Static	SSRL/4-2
pI3K SH2	Static	SSRL/4-2
Cyt C peptide	1-mm	APS/BESSRC
AK-16 peptide	1-mm	APS/BESSRC
Angiotensin	1-mm	APS/BESSRC

before being dialyzed exhaustively against 10 mM ammonium bicarbonate and lyophilized. RNase A was dissolved in 6 M GuHCl and 50 mM TRIS (pH 8) at a concentration of ~40 mg/mL. After ~30 minutes, a 30-fold excess of TCEP was added as a reductant. After an additional 30 minutes, the pH was reduced to 2.5 to inhibit reoxidation. This sample was diluted with GuHCl at various concentrations to a final protein concentration of 2 mg/mL and 10 concentrations of GuHCl ranging from 3.25 to 6 M in the presence of approximately 30-fold excess TCEP, 50 mM tris (pH 2.5).

Fyn SH3, α-TS, creatine kinase, Snase, and OspA were denatured by dissolving the lyophilized powder to a concentration of 10 mg/mL in 6 M GuHCl, 10 mM EDTA, and 20 mM tris (pH 7). The radical scavenger N-tert-butyl-α-(4-pyridyl)nitrone N′-oxide (Fluka, Inc.), which reduces radiation damage, thus allowing for increased x-ray flux, was also included at a concentration of 5 mM. The proteins were incubated in this solution for more than 24 hours at room temperature to ensure complete equilibration.

Denaturation Conditions

Heteropolymer theory suggests that the dimensions of an excluded volume random-coil polymer will change as solvent quality improves [14-16]. How, then, can we compare R_g*'s* collected under a wide range of differing solvent conditions? Empirical observations suggest that once a protein is unfolded (i.e., at denaturant concentrations well past the denaturation transition midpoint), any increase in R_g with increasing solvent quality is unmeasurable [16]. We have confirmed this observation by monitoring the R_g*'s* of ubiquitin, RNase A, ctAcP, and GCN4-p2′ over wide ranges of denaturant concentration. We observed no statistically significant variation in R_g for any of these proteins as the denaturant concentration was increased beyond the end of the unfolding transition (Table 2). Similarly, previous reports indicate that the R_g*'s* of denatured states produced by high levels of urea are indistinguishable from those produced by high levels of GuHCl [16]. Thus it appears that we can directly compare R_g*'s* collected with either denaturant and at any denaturant concentration that is well beyond the end of the folding transition.

Scattering Experiments

Small-angle x-ray scattering (SAXS) experiments were conducted on the BESSRC-CAT beamlines at the APS and on beamline 4-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) (Table 1). All SAXS experiments were conducted with the sample cuvette held at 25 ±1°C. R_g*'s* were determined by using Guinier analysis [13].

Results and Discussion

We have constructed a comprehensive and consistent data set of the R_g*'s* of 24 chemically denatured, cross-link-free and prosthetic-group-free proteins and polypeptides (Table 2).

We performed a weighted least squares regression by using the inverse of the reported standard errors as case weights. The $\log_{10}$ of R_g was employed as the dependent variable, and the $\log_{10}$ of the peptide or protein length (in residues) was used as the independent variable. Of particular interest is the estimate for the slope, since this parameter is predicted by theory to be 0.588. The results of this weighted fitting are effectively identical to the ones obtained from ordinary least squares regression ($\nu = 0.598 \pm 0.027$) (Fig. 1).

Only the R_g of creatine kinase and angiotensin II fall outside the 95% confidence intervals calculated for the observed dimensional scaling. To determine whether these measurements reflect statistically significant deviations from the theoretically predicted scaling relationship, we calculated 95% predictive intervals for future observations (i.e., prediction intervals for R_g given any arbitrary N; see Fig. 1, gray). The R_g of both proteins fall well outside this region, strongly suggesting that they do not obey the expected power-law relationship. Formal hypothesis testing, based on the assumption that data points were observations stemming from the model, yields p-values of 10^{-6} and 2×10^{-4}, respectively. Thus the R_g*'s* of only 2 of the 24 chemically denatured, cross-link-free and prosthetic-group-free peptides and proteins deviate significantly from the predicted random-coil behavior.

While the observation of random-coil dimensional scaling across the large majority of chemically denatured proteins is consistent with the results of previous scattering studies that indicate the chemically denatured

TABLE 2. Chemically denatured, cross-link-free and prosthetic-group-free proteins.

Protein	Length	R_g (Å)[a]	Conditions	Reference
GroEL	549	82 ±4	4 M urea	18
yPGK	416	71 ±1	2 M GuHCl	19
Creatine kinase	380	46.0 ±1.5	6 M GuHCl	This work
α-TS	268	47.7 ±2.3	6 M GuHCl	This work
Carbonic anhydrase	260	59 ±(~2)	6 M GuHCl	20
OspA	257	50.4 ±2.5	6 M GuHCl	This work
Apomyoglobin	154	40 ±(~2)	6 M GuHCl	20
Snase	149	37.2 ±1.2	6 M GuHCl	This work
Lysozyme, reduced	129	35.8 ±0.5	4 M GuHCl	21
CheY	129	38.0 ±1.0	5-7 M urea	2
RNase A, reduced	124	33.2 ±1.0	3.25-6 M GuHCl	This work
pI3K SH2	112	29.6 ±3.3	3 M GuHCl	This work
pI3K SH3	103	30.9 ±0.3	2.67 M GuHCl	This work
mAcP	98	30.4 ±1.3	6.5-8 M urea	22
ctACP	98	30.5 ±0.4	5.5-6.8 M urea	This work
Protein L	79	26.0 ±0.6	4-5 M GuHCl	23
Fyn SH3	78	25.7 ±0.5	6 M GuHCl	This work
Ubiquitin	76	25.2 ±0.2	4.9-6 M GuHCl	This work
GCN4-p2′	66	24.1 ±0.9	4.2-6 M GuHCl	This work
drK SH3	59	21.9 ±0.5	2 M GuHCl	24
Protein G	52	23 ±1	2.3 M GuHCl	25
N-term cyt C	39	18.4 ±1.0	4 M urea	This work
AK-16 peptide	16	9.8 ±0.6	4 M urea	This work
Angiotensin	8	9.1 ±0.3	4 M urea	This work

[a] Standard errors are indicated. These were derived by using a variety of approaches and widely varying numbers of observations; therefore, they provide, at best, only a qualitative indicator of experimental precision.

[b] Source for solvent conditions and approximate standard errors: G. Semisotnov, personal communication.

ensemble is entirely unstructured, it is seemingly inconsistent with compelling spectroscopic-based and simulations-based studies that suggest that many, if not all, chemically denatured proteins populate significant residual structure.

Reconciliation of short-range (i.e., sequence-local) order with near random-coil behavior may lay in the observation that the R_g of a denatured ensemble can be surprisingly independent of the detailed structure of the unfolded chain at the residue-to-residue level [16, 17]. For example, while sequence-local structure would be expected to alter a polypeptide's persistence length, altering R_0 and thus R_g, modest changes in such structure might not be apparent if the dependence of R_0 on persistence length is weak. We note, however, that classic polymer theory predicts a 2/5 power dependence, and that even this relatively weak dependence places seemingly strong constraints on the population of sequence-local residual structure. The observation that the majority of proteins fall within 2.7% of the best-fit power law suggests that if the 2/5 power dependence is correct, their mean persistence lengths are identical to within just 7% (i.e., $1.07^{2/5} = 1.027$).

Acknowledgments

O. Bilsel, P. Dawson, S. Koide, D. Shortle, and B. Matthews generously and without hesitation provided multimilligram samples of several of the peptides and proteins characterized here. The authors also wish to thank B. Zagrovic, K. Dill, C. Dobson, C. Freiden, and V. Pande for critical commentary and/or helpful clarifying discussions. This work was supported by National Institutes of Health (NIH) Grant Nos. R01GM62868-01A2 to K.W. Plaxco and R01GM55694 to T.R. Sosnick, funds from the Packard Foundation Interdisciplinary

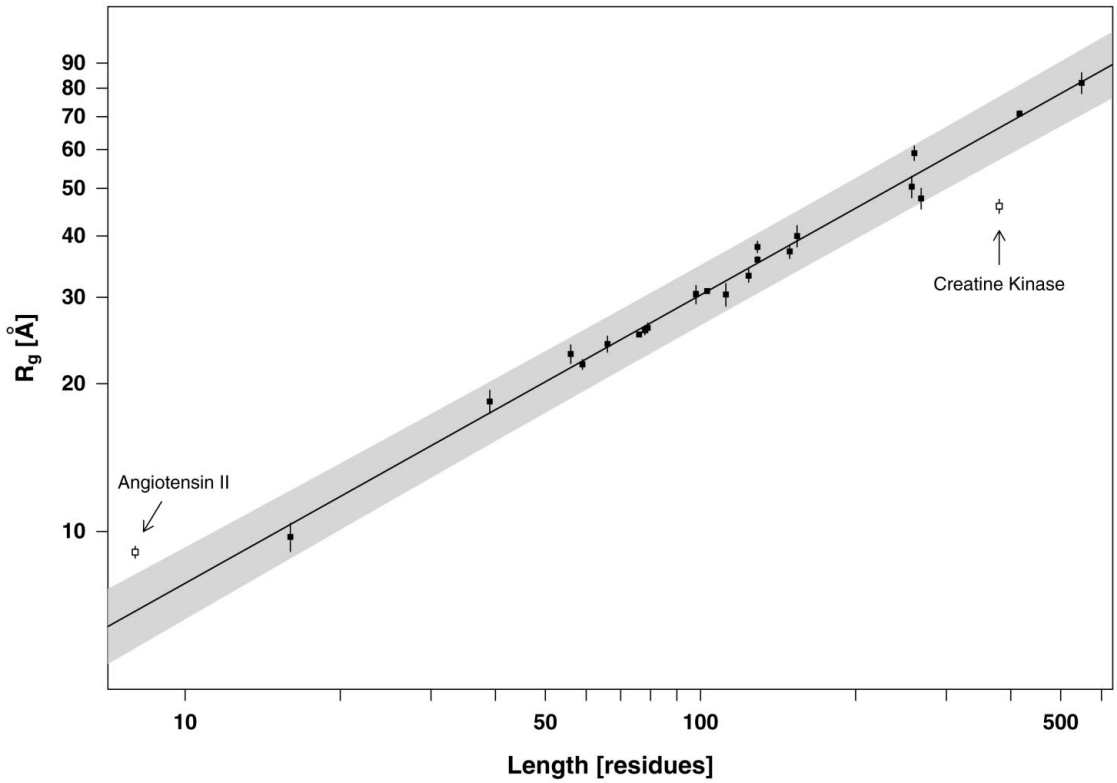

FIG. 1. The R_g's of the large majority of chemically denatured unfolded proteins scale with polymer length N via the power-law relationship $R_g = R_0N^{\nu}$. Two potentially significant outliers, creatine kinase and angiotensin II, are plotted as open squares. The solid line, which is the least squares fit ignoring the two potential outliers, produces an exponent $\nu = 0.587 \pm 0.030$ (95% confidence interval) that is indistinguishable from the 0.588 predicted for an excluded volume random coil. The shaded region represents the 95% confidence intervals for future measurements when it is assumed that the (log) R_g's are normally distributed around the fitted relationship. Only the measurements for creatine kinase and angiotensin II fall outside this predictive interval; thus, only these measurements can be said to represent unambiguously significant deviations. The vertical bars indicate the reported experimental (i.e., standard) errors. Since these were derived by using a variety of approaches and widely varying numbers of observations, they provide only a qualitative indicator of experimental precision.

Science Program to T.R. Sosnick and P. Thiyagarajan, and a faculty innovation award from Johns Hopkins University to I. Ruczinski. The SSRL is supported by the U.S. Department of Energy (DOE) and the NIH. Use of the IPNS and APS are supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-ENG-38.

References

[1] S.L. Kazmirski, K.B. Wong, S.M.V. Freund, Y.J. Tan, A.R. Fersht, and V. Daggett, "Protein folding from a highly disordered denatured state: The folding pathway of chymotrypsin 2 at atomic resolution," Proc. Natl. Acad. Sci. U.S.A. **98**, 4349-4354 (2001).

[2] P. Garcia, L. Serrano, D. Durand, M. Rico, and M. Bruix, "NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: Implications for protein folding initiation," Protein Sci. **10**, 1100-1112 (2001).

[3] M.E. Hodsdon and C. Frieden, "Intestinal fatty acid binding protein: The folding mechanism as determined by NMR studies," Biochemistry **40**, 732-742 (2001).

[4] J. Klein-Seetharaman, M. Oikawa, S.B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L.J. Smith, C.M. Dobson, and H. Schwalbe, "Long-range interactions within a nonnative protein," Science **295,** 1719-1722 (2002).

[5] R.E. Baldwin, "Protein folding — Making a network of hydrophobic clusters," Science **295**, 1657-1658 (2002).

[6] D. Shortle, "Composites of local structure propensities: Evidence for local encoding of long-range structure," Protein Sci. **11**, 18-26 (2002).

[7] D. Shortle and M.S. Ackerman, "Persistence of native-like topology in a denatured protein in 8 M urea," Science **293**, 487-489 (2001).

[8] D. Shortle, "The denatured state (the other half of the folding equation) and its role in protein stability," FASEB J. **10**, 27-34 (1996).
[9] G.W. Booker, A.L. Breeze, A.K. Downing, G. Panayotou, I. Gout, M.D. Waterfield, and I.D. Campbell, "Structure of an SH2 domain of the P85-alpha subunit of phosphatidylinositol-3-OH kinase," Nature **358**, 684-687 (1992).
[10] J.I. Guijarro, C.J. Morton, K.W. Plaxco, I.D. Campbell, and C.M. Dobson, "Folding kinetics of the SH3 domain of PI3 Kinase by real-time NMR combined with optical spectroscopy," J. Mol. Biol. **276**, 657-667 (1998).
[11] N. Taddei, F. Chiti, T. Fiaschi, M. Bucciantini, C. Capanni, M. Stefani, L. Serrano, C.M. Dobson, and G. Ramponi, "Stabilisation of alpha-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding," J. Mol. Biol. **300**, 633-647 (2000).
[12] C.N. Larsen, B.A. Krantz, and K.D. Wilkinson, "Substrate specificity of deubiquitinating enzymes: Ubiquitin C-terminal hydrolases," Biochemistry **37**, 3358-3368 (1998).
[13] O. Glatter and O. Kratky, *Small Angle X-ray Scattering* (Academic Press, London, England, 1982).
[14] K.A. Dill and D. Shortle, "Denatured States of Proteins," Ann. Rev. Biochem. **60**, 795-825 (1991).
[15] D.O. Alonso and K.A. Dill, "Solvent Denaturation and Stabilization of Globular-Proteins," Biochemistry **30**, 5974-5985 (1991).
[16] I. Millett, S. Doniach, and K.W. Plaxco, Adv. Protein Chem. **62**, 241-262 (2002).
[17] W.G. Miller and C.V. Goebel, "Dimensions of protein random coils," Biochemistry **7**, 3925-3935 (1968).
[18] M. Arai et al., "Denaturation and reassembly of chaperonin GroEL studied by solution x-ray scattering," Protein Sci. **12**, 672-680 (2003).
[19] V. Receveur, P. Garcia, D. Durand, P. Vachette, and M. Desmadril, "Role of hydrophobic interactions in yeast phosphoglycerate kinase stability," Proteins **38**, 226-238 (2000).
[20] E.V. Semisotnov et al., "Protein globularization during folding. A study by synchrotron small-angle x-ray scattering," J. Mol Biol. **262**, 559-574 (1996).
[21] M. Hoshino, Y. Hagihara, D. Hamada, M. Kataoka, and Y. Goto, "Trifluoroethanol-induced conformational transition of hen egg-white lysozyme studied by small-angle x-ray scattering," FEBS Lett. **416**, 72-76 (1997).
[22] I.S. Millet, L. Townsley, F. Chiti, S. Doniach, K.W. Plaxo, "Equilibrium collapse and the kinetic 'foldability' of proteins," Biochemistry **41**, 321-325 (2002).
[23] K.W. Plaxco, I.S. Millett, D.J. Segel, S. Doniach, and D. Baker, "Chain collapse can occur concomitantly with the rate-limiting step in protein folding," Nat. Struct. Biol. **6**, 554-556 (1999).
[24] W.Y. Choy et al., "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 (2002).
[25] C.K. Smith, Z.M. Bu, K.S. Anderson, J.M. Sturtevant, D.M. Engelman, and L. Regan, "Surface point mutations that significantly alter the structure and stability of a protein's denatured state," Protein Sci. **5**, 2009-2019 (1996).