Towards resolution of ambiguity for the unfolded state of proteins.

Gregory Beaucage

Department of Chemical and Materials Engineering, University of Cincinnati, Cincinnati OH 45221-0012.

The unfolded states in proteins remain weakly understood despite their importance to protein folding; misfolding diseases (Parkinson's & Alzheimer's); and natively unfolded proteins (as many as 30% of eukaryotic proteins, Fink, A. L. *Curr. Op. in Struc. Biol.* 15, 35-41 (2005)). Research has been hindered by the inability to quantify the residual (native) structure present in an unfolded protein. Here, a scaling model is proposed for unfolded proteins that can be used to quantify the *degree of folding*. The model takes a global view of protein structure and can be applied to a number of analytic methods and to simulations. Two examples are given of application to small angle scattering from pressure induced unfolding of SNase and from acid unfolded Cyt c. These examples quantitatively show 3 characteristic unfolded states & the statistical nature of a folding pathway.

Proteins generally display a globular conformation characterized by the presence of a surface with a distinguishable core often maintained by hydrophobic interactions (residue/residue interactions). The globular state can be of fixed structure associated with biological activity, native state (N), or can be somewhat flexible in structure and display only limited functionality, molten globule state (MG)[1,2]. Proteins can also display categories of unfolded states such as the pre-molten globule (PMG), the unfolded state (U) and misfolded states[3] that display characteristics of synthetic polymers and mass fractal aggregates in that a discrete surface and distinguishable core are not observed. The nature of interactions in unfolded states & PMG are not well

defined but may be partly associated with solvation effects (residue/solvent interactions).

The unfolded state is the nascent state[4] but it is usually achieved in vitro through chemical, physical or thermal denaturing[5,6]. In some cases unfolded proteins can display biological activity[2,7-9]; *natively* unfolded proteins. There is dispute concerning the extent of native secondary structure in the unfolded state and especially how the spectroscopic observation of secondary structure in the unfolded state is associated with folding[10,11,12]. For instance, the premolten globule state shows secondary structure associated with the native state but displays variable density with molar mass for different PMG proteins[2,13] similar to the unfolded state.

The unfolded state has been described as a Gaussian structure[14,15], however, it has been pointed out that unfolded proteins are not expected to follow Gaussian scaling since they exist in good solvents where a self-avoiding structure is expected[2,14]. Further, several different unfolded states are known for many proteins[2].

It would be desirable to quantify the *extent of folding* for various denatured states[16,17]. Here a quantitative, statistical description of folding is presented based on a mass-fractal model[18,19]. The description associates branching/bridging in polymers and inorganic aggregates with folding in proteins allowing a quantification of protein folding using a simple analytic method. The result of this approach is a quantitative definition of the unfolded state and a statistical description of the extent of folding.

A Scaling Model for Protein Folding.

There are subtle similarities between our understanding of random synthetic polymers, fractal aggregates of inorganic particles and our understanding of the unfolded states of proteins. Recently, advances have been made in our analytic understanding of these

ramified fractal structures[18,19]. This report seeks to apply these advances in order to quantify topological features of unfolded, misfolded and premolten globule states in proteins[9]. In the fractal model a protein residue is represented by a particle of an aggregate (circles in Fig. 1) allowing a direct analogy between a protein chain and a branched mass-fractal aggregate. In this analogy, bridge sites reflect a variety of possible long-range enthalpic interactions such as hydrophobic interactions, hydrogen bonds, or disulfide bonds between residues at large sequence separations, Fig. 1[2,15]. Such bridge/fold sites will lead to densification of the unfolded structure and might be seen as a quantifiable stage in the process of folding.

From a fractal view the partially folded structure in Fig. 1 displays two 'paths' or sequences of residues, the chain path, z, and the minimum path, p, that short-cuts the main path through long-range interactions, *bridges*, shown as bold lines. The minimum path, p, grey circles in Fig. 1, is important to the static & dynamic mechanical response of the protein since the stress bearing path occurs along p[20] and because motion of the chain and the available conformations are dramatically restricted by shorter minimum paths. The chain path, z, is reflected by the total number of residues in a protein, the sum of the grey and white circles in Fig. 1. For an unfolded chain with no bridges $p = z$. When folding bridges[18] are present, $p < z$. *An unfolded protein is unambiguously defined by, $p = z$ in this model.*

We can relate the C-terminal to N-terminal separation distance of the partially folded protein, R, to p and the residue statistical step length, b, through,

$$R \sim p^{1/d_{\min}} b \tag{1}$$

where d_{min} is the minimum dimension and represents the mass fractal dimension of the minimum path (grey circles in Fig. 1). Similarly, the whole protein can be described in terms of the number of residues, z, and the mass fractal dimension, d_f,

$$R \sim z^{1/d_f} b \tag{2}$$

where d_f is the mass-fractal dimension for the coil and is reported to be generally about 5/3 in the denatured state by Kohn et al.[14], distinctly different from a Gaussian value of 2. Uversky[2] has given values of about $d_f \sim 2.5$ for PMG and values between about 2 and 5/3 for U. For MG and N, $d_f = 3$ and $d_{min} \sim 1$, Table 1.

p can be directly related to z through the connectivity dimension c,

$$z = p^c \qquad \text{where} \qquad c = \frac{d_f}{d_{\min}} \tag{3}$$

For a linear (unfolded) chain the connectivity $c = 1$ and $d_{min} = d_f$. *This serves as an analytic definition for an unfolded protein.* When $d_{min} = 1$ the path p is a straight line in at least one direction (for example along the length of a rod/helix or the diameter of a disk or along the pleat of a β-sheet) and the structure is termed a *regular object*. We can consider a measure of the *degree of folding*, ϕ_{Br}, as the ratio of the residues in loops ($z - p$) to the total residues in the protein, z,

$$\phi_{Br} = \frac{z - p}{z} = 1 - z^{1/c - 1} \tag{4}$$

Table 1 shows the values for c, d_{min}, d_f and ϕ_{Br} for various structures. Although there are a number of methods, including simulation, by which these statistical parameters might be determined for an unfolded or partially folded protein, the most convenient method involves small-angle scattering of x-rays or neutrons from dilute solutions. All of the parameters in Table 1 can be determined from a single static scattering pattern (given sufficient angular range) just as they have been determined in studies of branched ceramic nano-aggregates and for branched synthetic polymers[18,20].

Quantification of Folding by Small-Angle Scattering.

Scattering under the Rayleigh-Gans approximation (x-ray and neutron scattering) is often reduced to generic scattering laws like Guinier's and power laws. Guinier's law[21,22] is given by,

$$I(q) = G\exp\left(\frac{-q^2 R_g^2}{3}\right) \tag{5}$$

where *I(q)* is the scattered intensity, $q = 4\pi\sin(\theta/2)/\lambda$, θ is the scattering angle and λ is the wavelength of radiation, R_g^2 is the mean square radius of gyration and G is a contrast factor. At higher-q the fractal power-law equation[22,23] is appropriate for U and PMG,

$$I(q) = B_f q^{-d_f} \tag{6}$$

where B_f is the power law prefactor. d_{min} can be calculated from scattering using features from (5) and (6) to account for bridged structures[18],

$$d_{\min} = \frac{B_f R_g^{d_f}}{G\tilde{A}\left(\frac{d_f}{2}\right)} \tag{7}$$

where *Γ()* is the gamma function. (3) can then be used to determine *c*. *z* can be considered the number of residues in the protein or *z* can be experimentally measured from a high-q observation of the protein persistence unit[18,24,25].

Using this approach we can quantify the extent of folding in terms of the dimensions *c* and d_{min}. For example, for a fully unfolded structure, that folds to the native state, we should observe $c = 1$, $d_{min} = d_f$ initially, with d_f close to 5/3 for a self-avoiding walk[14]. The folding process should be characterized by an increase in *c* from 1

to d_f as the structure approaches a regular object. The folding process presumably involves densification of the structure so an increase in d_f is also expected. A calculation of the extent of folding, ϕ_{Br}, will show an increase from 0 to 1.

Example Using Staphylococcal Nuclease.

Staphlococcal Nuclease (SNase) is a protein excreted by Staphylococcus Aureus bacterium that cleaves nucleic acids in the extracellular medium by hydrolyzing P-O bonds[26]. SNase is a small monomeric protein, 17.5 kDa, z = 149 residues and a single tryptophan residue with no disulfide bonds that has served as a model for protein folding studies[27,28]. The native state structure (grown at pH 8.15 in low salt buffer[28]) is composed of a 5 member beta barrel and three alpha helices, Fig. 2d. The tertiary structure displays a nucleotide binding pocket with two flexible regions associated with biological function[29]. Panick et al.[29] have studied the pressure dependence of small-angle x-ray scattering, SAXS, from staphylococcal nuclease in 1% solutions at a pH of 5.5 (using a recombinant nuclease produced from Escherichia coli). A denatured state and a folded state were measured as well as defolding kinetics. Pressure induced changes in the protein conformation are of particular interest due to the weak perturbation. Here, SAXS data of Panick et al.[29] that was digitized is reanalyzed using the procedure described above.

Fig. 2a shows a low pressure dilute solution scattering curve at pH 5.5 (100 bar) and a pressure denatured curve (2800 bar) for SNase. SNase displays a native state in SAXS at pH 7.5, as reported by Uversky et al[30] and unfolds in the presence of acid below pH of 4[5,6,30]. For this reason, Panick used a pH of 5.5 for these studies, just above the pH for acid unfolding. At 100 bar the sample displays the characteristics of a *regular object*, that is a low-dimensional solid object similar to a sheet. The low dimensional *regular* structure at 100 Bar pH 5.5 has not been previously reported since the radius of gyration at pH 5.5 is indistinguishable from that at pH 7, though the

structure is significantly different by this dimensional analysis. Such regular structures, displaying 2-dimensional sheet structures, have been associated with misfolded proteins (for example 3). The rounded rectangle in Fig. 4b highlights a convergence of c and d_f indicating this *regular structure* with $d_f \sim 2.2$. This initial pH 5.5 structure, indicated in Fig. 4b, may be related to the pH denatured states shown by Whitten et al. for SNase[31], associated with a framework model for folding where elements of secondary structure fluctuate around native tertiary locations[30]. This is also reminiscent of the PMG described by Uversky[2]. Similar molecular mobility was previously reported by Wall[26] form XRD measurements on SNase.

At higher pressure c drops towards 1 for an unfolded, linear chain as might be expected for the pressure unfolding process. Simultaneously, d_{min} increases towards about 1.8 and d_f drops towards the same value with a projected unfolded structure just below 4000 bar (circle in Fig. 4b). The increase in d_{min} indicates that the minimum path becomes more convoluted and longer as the folds are released.

While the dimensional analysis displays a gradual unfolding with pressure between an asymmetric *regular-object*, possibly misfolded[3] state ($d_{min} = 1$) and a totally unfolded state at high pressures ($c = 1$), R_g (Fig. 2b) as well as p and ϕ_{Br} (Fig. 2c) display an exponential dependence on pressure that might be interpreted as a rapid transition such as in a two state model. This discrepancy is due to the exponential relationships between the dimensions of Fig. 2b and R_g, p and z.

Example Using Cytochrome c.

SAXS studies of Cyt c were first performed by Kataoka et al.[32]. Kataoka studied acid unfolded horse ferricytochrome c in an investigation of the unfolded-molten globule-native transition as previously studied by CD and NMR[5,33]. Cyt c has been used as a model protein for folding studies[34-37]. Kataoka denatured Cyt c by reduction of pH to

near 2 (acid unfolded). He then induced folding from this acid-unfolded state by two methods 1) addition of salt (NaCl) and 2) acetylation. Kataoka's SAXS data was digitized & fit using the unified function[18,22,25]. Considering the acid unfolded Cyt c, SAXS shows a mass-fractal dimension of 2, a radius of gyration of 20.9 Å and a connectivity dimension, *c*, of 1.1, Table 2, indicating a close to linear chain with a sparse amount of folding (*c* would be 1 for unfolded) that leads to a fractal dimension slightly higher than the minimum path dimension of 1.9. The coil R_g is the largest measured by Kataoka supporting an almost unfolded structure. Cyt c has z = 105 residues so the degree of folding for the acid unfolded state is 0.34 from (4).

Two quite different mechanisms for collapse of the unfolded state are indicated by the fractal approach, Table 2:

1) The addition of salt (0.02 M NaCl) induces collapse by making the chain path more convoluted (d_{min} increases to ~ 2) making d_f higher while the chain folding is identical to the acid unfolded state (c, ϕ_{Br}). This is essentially coil collapse driven by reduction in solvent quality (residue-solvent interactions) since d_{min} approaches 2 for θ-coils in a poor solvent similar to solution phase separation in synthetic polymers.

2) For acetylation (2), the protein is misfolded in a sheet-like structure similar to pH 5.5 SNase above, with d_{min} close to 1 (a close to linear path exists through the structure). So, despite a mass fractal dimension, d_f, close to the acid unfolded state, the structure is quite different and more compact reflecting an asymmetric, crumpled sheet-like, misfolded structure. This is essentially coil collapse by formation of secondary structure (residue-residue interaction) similar to SNase.

All three results in Table 2 display what would appear to be Gaussian scaling in a Kratky plot of Iq^2 versus q, since d_f is close to 2, but *none of these conditions produce truly random walks*!

The Cyt c results show 3 possible unfolded states: good-solvent unfolded expanded coil; poor-solvent unfolded coil; and a low-dimensional (sheet-structure), solvated misfolded chain. The SNase results indicate how the fractal model can view structural changes along a folding, or in this case, an unfolding pathway from a misfolded state. It is demonstrated that some structural measures can display a two-state model, while scaling parameters show a continuous transition. A statistical interpretation of proteins can yield unique information concerning the extent of folding as well as defining the unfolded state in terms of quantitative, statistical features that can be determined by a number of techniques including small-angle scattering. This information will be of use in the cure of misfolding diseases such as Parkinson's and Alzheimer's Diseases & Type II Diabetes and may prove helpful in understanding "natively" unfolded proteins as well as folding pathways, generally.

1. Pande, V.S. & Rokhsar, D. S. Is the molten globule a third phase of proteins? *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1490-1494 (1998).

2. Uversky, V. N. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? *Cell. Mol. Life Sci.* **60**, 1852-1871 (2003).

3. Chaudhuri, T. K., Paul, S. Protein-misfolding diseases and chaperone-based therapeutic approaches *Febs J.* **273**, 1331-1349 (2006).

4. Nissen, P., Hansen, J., Ban, N., Moore, P. B., Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis *Science* **289** 920-930 (2000).

5. Goto, Y., Takahashi, N., Fink, A. L. Mechanism of acid-induced folding of proteins *Biochemistry* **29,** 3480-3488 (1990).

6. Tanford, C. Protein denatureation. *Advan. Protein Chem.* **23,** 121-282 (1968).

7. Uversky, V. N. Natively unfolded proteins: A point where biology waits for physics. *Pro. Sci.* **11**, 739-756 (2002).

8. Fink, A. L. Natively unfolded proteins. *Curr. Op. in Struc. Biol.* **15**, 35-41 (2005).

9. Receveur-Brechet, V., Bourhis, J-M., Uversky, V. N., Canard, B., Longhi, S. Assessing protein disorder and induced folding. *Pro.: Str. Func. Bioinfo.* **62**, 24-45 (2006).

10. McCarney, E. R., Werner, J. H., Bernstein, S. L., Ruczinski, I., Makarov, D. E., Goodwin, P. M., Plaxco, K. W. Site-specific dimensions across a highly denatured protein; a single molecule study *J. Mol. Biol.* **52,** 672-682 (2005).

11. McCarney, E. R., Kohn, J. E., Plaxco, K. W. Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins *Crit. Rev. in Biochem. and Mol. Biol.* **40,** 181-189 (2005).

12. Jha, A. K., Colubri, A., Freed, K. F., Sosnick, T. R. Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13099-13104 (2005).

13. Uversky, V. N., Ptitsyn, O. B. Further evidence on the equilibrium "pre-molten globule state": Four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.,* **255** 215-228 (1996).

14. Kohn, J. E., Millet, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., et al. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12491-12496 (2004).

15. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., Schwalbe, H. Long-Range Interactions within a nonnative protein. *Science* **295**, 1719-1722 (2002).

16. Shortle, D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27-34 (1996).

17. Ramos, C. H. I., Ferreira, S. T. Protein folding, misfolding and aggregation: Evolving concepts and conformational diseases. *Pro. Pept. Lett.* **12**, 213-222 (2005).

18. Beaucage, G. Determination of branch fraction and minimum dimension of mass-fractal aggregates *Phys. Rev. E.* **70**, 031401 (2004).

19. Beaucage, G., Kammler, H. K., Mueller, R., Strobel, R., Agashe, N., Pratsinis, S. E., Narayanan, T. Probing the dynamics of nanoparticle growth in a flame using synchrotron radiation. *Nature Mater.* **3**, 370-374 (2004).

20. Witten, T. A., Rubinstein, M., Colby, R. H. Reinforcement of rubber by fractal aggregates *J. Phys. II* **3**, 367-383 (1993).

21. Guinier, A., Fournet, G. (1955) *Small Angle Scattering of X-rays.* (John Wiley, New York).

22. Beaucage, G. Approximations leading to a unified exponential/power law approach to small angle scattering. *J. Appl. Crystallogr.* **28**, 717-728 (1995).

23. Meakin, P. Fractal structures. *Prog. Solid State Chem.* **20**, 135-233 (1990).

24. Beaucage, G., Rane, S., Sukumaran, S., Satkowski, M. M., Schechtman, L. A., Doi, Y. Persistence length of isotactic poly(hydroxy butyrate) *Macromolecules* **30**, 4158-4162 (1997).

25. Beaucage, G. Small-Angle Scattering from Polymeric Mass Fractals of Arbitrary Mass-Fractal Dimension. *J. Appl. Crystallogr.* **29**, 134-146 (1996).

26. Wall, M. E., Ealick, S. E., Gruner, S. M. Three-dimensional diffuse x-ray scattering from crystals of Staphylococcal nuclease. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6180-6184 (1997).

27. Chatfield, D. C., Szabo, A., Brooks, B. R. Molecular Dynamics of Staphylococcal Nuclease: Comparison of Simulation with 15N and 13C NMR Relaxation Data *J. Am. Chem. Soc.* **120**, 5301-5311 (1998).

28. Hynes, T. R., Fox, R. O. The crystal structure of staphylococcal nuclease refined at 17 Ang Resolution Hynes TR Fox RO Proteins: structure, function and genetics *Pro.: Str. Func. Gen.* **10**, 92-105 (1991).

29. Panick, G., Malessa, R., Winter, R., Rapp, G., Frye, K. J. & Royer, C. A. Structural characterization of the pressure-denatured state and unfolding/refolding kinetics of staphylococcal nuclease by synchrotron small-angle X-ray scattering and Fourier-transform infrared spectroscopy. *J. Mol. Biol.* **275**, 389-402 (1998).

30. Uversky, V. N., Karnoup, A. S., Segel, D. J., Seshadri, S., Doniach, S., Fink, A. L. Anion-induced folding of staphylococcal nuclease: Characterization of multiple equilibrium partially folded intermediates. *J. Mol. Biol.* **278**, 879-894 (1998).

31. Whitten, S. T., Garcia-Moreno, E. B., Hilser, V. J. Local conformational fluctuations can modulate the coupling between proton binding iand global structural transitions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4282-4287 (2005).

32. Kataoka, M., Hagihara, Y., Mihara, K., Goto, Y. Molten globule of cytochrome-c studied by small angle X-ray scattering. *J. Mol. Biol.* **229**, 591-596 (1993).

33. Ohgushi, M., Wada, A. 'Molten-globule state': a compact form of globular proteins with mobile side-chains *FEBS Letters* **164**, 21-24 (1983).

34. Pletneva, E. V., Gray, H. B., Winkler, J. R. Nature of the cytochrome c molten globule *J. Am. Chem. Soc.* **127**, 15370-15371 (2005).

35. Tezcan, F. A. A., Findley, W. M., Crane, B. R., Ross, S. A., Lyubovitsky, J. G., Gray, H. B. Using deeply trapped intermediates to map the cytochrome c folding landscape *PNAS* **99**, 8626-8630 (2002).

36. Lyubovitsky, J. G., Gray, H. B., Winkler, J. R. Mapping the cytochrome c folding landscape *J. Am Chem Soc.* **124**, 5481-5485 (2002).

37 Lyubovitsky, J. G., Gray, H. B., Winkler, J. R. Structural features of the cytochrome c molten globule revealed by fluorescence energy transfer kinetics *J. Am. Chem. Soc.* **124**, 14840-14841 (2002).

Correspondence should be addressed to G.B. (beaucag@uc.edu).

Tables

Table 1. Dimensional Description of Objects.

Object	c ($\leq d_f$)	d_{min} ($\leq d_f$)	d_f	ϕ_{Br}
Sphere/MG/N	**3**	**1**	**3**	**1**
Disk	**2**	**1**	**2**	**$(1 - 1/\alpha)$**
Rod	**1**	**1**	**1**	**0**
Gaussian Chain	**1**	**2**	**2**	**0**
Self-Avoiding Walk	**1**	**5/3**	**5/3**	**0**
Randomly Branched Chain	**5/4**	**2**	**5/2**	**$(1 - z^{-1/5})$**
PMG	**≥3/2**	**≤5/3**	**5/2**	**(4)**
U	**≥3/5 d_f**	**≤5/3**	**1-2**	**(4)**

c, d_{min}, d_f and the *extent of folding*, ϕ_{Br}, are defined in the text. (d_f for PMG and U from [2], others from [18].) *c* & d_{min} are ≤ d_f following (4) as noted in the column title. α is the aspect ratio for the disk.

Table 2. Unfolded and PMG States for Cytochrome c form Kataoka's data [32].

Sample	R_g, Å	d_f	d_{min}	c	ϕ_{Br}	Description	Thermodynamic State
Acid Unfolded	**20.9**	**2**	**1.9**	**1.1**	**0.34**	**Weakly Folded Good Solvent**	**Expanded Coil** d_{min} **=> 5/3** [18]
0.02 M NaCl	**17.9**	**2.3**	**2**	**1.1**	**0.34**	**Weakly Folded Nearly Poor Solvent**	**θ-state** d_{min} **=> 2** [18]
2 Acetylation	**17.4**	**1.9**	**1.2**	**1.6**	**0.83**	**Misfolded**	**Sheet-like State** d_{min} **=> 1** [18]