Title

Water as a Good Solvent for Unfolded Proteins: Folding and Collapse are Fundamentally Different

Journal

Journal of Molecular Biology, 432(9)

ISSN

0022-2836

Authors

Clark, Patricia L
Plaxco, Kevin W
Sosnick, Tobin R

Publication Date

2020-04-01

DOI

10.1016/j.jmb.2020.01.031

Peer reviewed

HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2021 July 05.

Published in final edited form as:

J Mol Biol. 2020 April 17; 432(9): 2882–2889. doi:10.1016/j.jmb.2020.01.031.

Water as a good solvent for unfolded proteins: Folding and collapse are fundamentally different

Patricia L. Clark[a,*], **Kevin W. Plaxco**[b,*], **Tobin R. Sosnick**[c,*]

[a]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556.

[b]Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106

[c]Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637.

Abstract

The argument that the hydrophobic effect is the primary effect driving the folding of globular proteins is near universally accepted (including by the authors). But does this view also imply that water is a "poor" solvent for the unfolded states of these same proteins? Here we argue that the answer is "no." That is, folding to a well-packed, extensively hydrogen bonded native structure differs fundamentally from the non-specific chain collapse that defines a poor solvent. Thus, the observation that a protein folds in water does not necessitate that water is a poor solvent for its unfolded state. Indeed, chain-solvent interactions that are marginally more favorable than non-specific intra-chain interactions are beneficial to protein function because they destabilize deleterious misfolded conformations and inter-chain interactions.

Keywords

protein folding; denatured state; small-angle scattering; collapse; Flory exponent

The major driving force underlying protein folding is the hydrophobic effect. As articulated by Kauzmann, burial of apolar side chains in the interior of a soluble protein reduces the amount of hydrophobic surface area in contact with water [1]. Such burial increases solvent entropy, as burial reduces the number of water molecules that must adopt partially constrained conformations in order to solvate apolar groups. Against the background of this general driving force, conformational specificity is dictated by the precise arrangement of hydrogen bonds, electrostatic interactions and side chain packing. Associated with this paradigm is an oft-held but rarely stated assumption: that non-specific hydrophobic interactions also cause compaction of unfolded proteins. That is, upon transfer of a protein from a chemical denaturant to more physiologically relevant conditions (water, essentially), the energetics of hydrophobic burial are sufficient to drive collapse of the unfolded chain,

*Correspondence should sent to pclark1@nd.edu, kwp@ucsb.edu and trsosnic@uchicago.edu.

shrinking the expanded, self-avoiding random walks (SARWs) populated at high denaturant to something much more compact. Such collapse would significantly impact both folding thermodynamics (by stabilizing the denatured state ensemble, DSE) and folding kinetics (by constraining the DSE).

In the parlance of polymer physics, collapse of the DSE equates to water being a "poor solvent" for the unfolded polypeptide chain [2]. A growing collection of experimental results, however, have demonstrated that the unfolded states of many globular proteins remain highly expanded upon transfer to water (or at least low denaturant concentrations), calling into question the seemingly reasonable paradigm that water is a poor solvent [3–12]. In this light, we discuss here the extent to which the principles of polymer physics can be applied to describe the DSE, examine the experimental evidence for the view that water is a good DSE solvent, and explore the implications of this paradigm shift for the folding, interactions, and evolution of proteins. We restrict our discussion here to the DSEs of soluble, globular proteins. While the properties of intrinsically disordered proteins (IDPs) is a related and fascinating topic, their unusual and diverse sequence compositions can lead to vastly differing physical properties, which may preclude a single physical description.

Adapting polymer physics formalisms to define solvent quality for proteins

While proteins are polymers, they differ in potentially significant ways from the simple homopolymers from which the formalisms of polymer physics were historically derived [13]. Specifically, proteins are relatively short, chiral, sequence-specific heteropolymers, distinctions that raise questions regarding the extent to which the principles of classical polymer physics apply to them. Given this, we begin our discussion by addressing the extent to which proteins obey the same physics as long homopolymers regarding the concept of solvent quality. To do so we consider three types of (qualities of) solvent. The first are solvents that interact more strongly with the monomers in the polymer than one monomer does with another, leading the polymer to adopt expanded, self-avoiding random walk (SARW). In such a solvent, the dimensions of the chain exhibit a power-law dependence on polymer length, with a Flory scaling exponent, ν of 3/5 in the relationship $R_g \propto N^{\nu}$ ($R_{SARW} \propto N^{3/5}$ Fig. 1). A poor solvent, in contrast, interacts more weakly with the polymer than the polymer does with itself, causing the chain to collapse and form a compact globule with a 1/3-power dependence of dimensions on length ($R_{compact} \propto N^{1/3}$). Between these extremes there must necessarily be a point of perfect balance. That is, if one could tune the quality of a solvent from good to poor with arbitrary fineness, there will be a point along that arc at which monomer-monomer attraction perfectly counter-balances excluded volume effects. Under this solvent condition, termed a "θ solvent," polymer dimensions scale with the same square-root length dependence ($R_\theta \propto N^{1/2}$) expected for a non-self-avoiding random coil.

What values does the scaling exponent adopt when proteins are unfolded under various solvent conditions? There is broad consensus that $\nu = 3/5$ for the DSE populated in aqueous solutions containing high concentrations of a chemical denaturant, such as urea or guanidinium hydrochloride (Fig. 1), indicating that such solutions are good solvents for polypeptides [14]. Some controversy remains, however, regarding solvent quality under physiological conditions (i.e., when the concentration of denaturant is reduced to zero).

Below we briefly describe the origins and current status of this controversy and then explore the implications for water being a good DSE solvent on our understanding of folding kinetics and thermodynamics, as well as the function and evolution of proteins.

Experimental evidence indicates water is a good solvent for the unfolded states of many proteins

Two primary arguments motivated the historical model that water is a poor solvent for the DSE. First, as noted above, the fact that proteins *fold* in the absence of denaturant could be taken to imply that water must be a poor solvent for the polypeptide chain. In this context, however, one must distinguish between folding to a native conformation and the non-specific compaction associated with the classical, homopolymer-derived definition of solvent quality. Specifically, the packing within a native protein differs dramatically from that seen during the non-specific collapse of a homopolymer; the interiors of native proteins achieve densities similar to crystalline organic solids [15] and are characterized by near-complete hydrogen bonding of the chain either with itself or with the solvent [16]. These observations suggest that the phenomenon of folding could differ significantly from non-specific collapse. A direct prediction of folding and collapse representing distinct phenomena is that a folded state can be stable under conditions where a non-specifically collapsed state of the same sequence is not. Consistent with this prediction, many proteins remain stably folded in high concentrations of guanidinium hydrochloride [8, 17, 18], conditions that are generally accepted to be a good solvent (e.g., the dimensions of the DSEs of proteins that unfold under these conditions invariably fall on the $\nu = 3/5$ power law line expected for a SARW [14]). Simply put, while water must be a poor solvent in order for the DSE to collapse, it need not be a poor solvent in order for the native states of proteins to be stable.

The second major historical argument in favor of water being a poor solvent is drawn from single-molecule Förster resonance energy transfer (FRET) studies – and associated simulations – of the DSE, which were interpreted to imply that DSEs undergo significant contraction as the concentration of denaturant is reduced from initially high values, leading to ν falling below 0.5 [19–40]. In contrast, small-angle x-ray scattering (SAXS) studies have consistently failed to identify such a large contraction (Fig. 2); this includes both studies of the dimensions of single-domain proteins immediately upon dilution from denaturant and equilibrium studies in the absence of denaturant of the dimensions of IDPs with amino acid compositions akin to stably-folded proteins [3–8, 10–12, 41, 42]. Recent studies have reconciled this crucial discrepancy by developing, for example, more realistic simulations (the force fields employed in prior studies are now known to over-estimate collapse [8, 43]) and new analytical models for analyzing FRET data, which produce closer agreement with SAXS-derived models [9, 41, 42, 44]. In parallel, we recently demonstrated that, in the absence of denaturant, interactions between commonly used FRET fluorophores – either with each other or with the chain – can introduce artifacts, causing the fluorophore-modified DSE to contract relative to the unmodified DSE [10–12, 45], a finding supported by other recent studies [46–48]. Taken together, these results provide further support for the argument that water is a good solvent for the unfolded states of many proteins.

We want to emphasize that we are not proposing that water is a good solvent for all unfolded proteins. It is clear, for example, that low-diversity or poorly mixed sequences sometimes characterized by high local or total hydrophobic content, unusual charge patterns, or specific, stabilizing contacts (such as cation-π interactions) do not always expand to SARW dimensions in water [49]. More generally, a palette of 20 diverse amino acids permits protein sequences to encode a wide range of conformational behaviors. Our perspective, however, is that highly expanded, well-solvated DSEs are common amongst foldable proteins. Consistent with this, protein sequences are typically well-mixed patterns of hydrophobic and polar residues, lacking long stretches of hydrophobic residues [50]. Indeed, the ubiquity of well-mixed sequences suggests that this is a broadly beneficial property.

Implications of water as a good solvent for the DSE

We are arguing that non-specific compaction of the DSE in water is often unfavorable (uphill in free energy). We appreciate that this claim runs counter to common perception, including those inferred (correctly or incorrectly) from the “funnel shaped” energy landscape diagrams commonly used to depict folding, in which initial events including the “beginning of helix formation and collapse” are depicted as “downhill” by virtue of the loss of chain entropy being excluded from the vertical energy axis [51]. We note, however, that when this (sizable) entropic component is included, the ensuing free energy surface can easily contain a sizable barrier to folding and, presumably, non-specific collapse.

In addition to the direct experimental evidence described above, there are numerous favorable physiochemical and biological implications associated with water acting as a good solvent for unfolded proteins. One of the reasons that DSE collapse in water has proven such an appealing idea is that it would help to explain how proteins avoid the Levinthal paradox [52] to fold rapidly. Specifically, if the DSE contracts significantly in water, this would convert folding into a two-step “collapse-then-fold” process that, in theory, would accelerate folding by constricting the search to only relatively compact states. As has been noted previously, however, such collapse could lead to the formation of stable, non-native compact states [53–56] that might serve as kinetic traps, slowing folding rather than accelerating it. From this viewpoint, the fact that water is a good solvent and thus destabilizes compact, non-native conformations helps ensure that folding is rapid. Consistent with this view, simulations of realistically complex folding models indicate that, even under conditions where the folding energy landscape is otherwise optimal (designed sequences folding at their temperature of maximum folding rate), the folding of protein-like heteropolymers is accelerated when thermodynamic cooperativity is enhanced [57]. In parallel, plausible mechanistic descriptions of folding have been put forward that do not invoke collapse as an essential first step in the folding of simple, single-domain proteins (Box 1).

Although the polymer physics concept of solvent quality formally holds only for intra-chain effects, the fact that water is a good solvent for unfolded proteins implies that inter-protein contacts are likewise often weaker than protein-solvent interactions, which in turn has implications regarding the specificity of protein-protein interactions. For example, although the existence of chaperone networks indicates that cells have evolved mechanisms to deal with undesirable interactions that occur within and between proteins at high cellular

concentrations [58], that water serves as a good solvent should reduce the population of misfolded and aggregation-prone states [59, 60]. Furthermore, just as weak intra-molecular interactions require a heavily optimized native state for stability, water-induced weakening of inter-protein interactions will promote binding cooperativity and, with that, specificity. Water acting as a good solvent thus increases the specificity of protein-protein interactions (e.g., intracellular signaling and other desirable binding events) while reducing competing interactions.

Evolution had no say in the selection of water as a solvent for biomolecules, but it did, presumably, have a say in the choice of biopolymer. Given the beneficial effects of water acting as good solvent for unfolded proteins, this property of poly-α-amino acids may have been one of the selective pressures that ensured its adaptation. Specifically, proteins composed of the twenty "proteogenic" α-amino acids strike a delicate balance between having sufficiently strong interactions and large conformational entropy in order to fold and bind specifically and cooperatively while avoiding too many unwanted interactions: they are "just sticky enough" to get the job done, without suffering from undue non-specific interactions.

Acknowledgements

We thank members of our groups and colleagues for stimulating discussions over the past couple of decades. This work was supported by NIH Grants GM055694 (TRS) and GM130122 (TRS, PLC), the W. M. Keck Foundation (PLC) and National Science Foundation MCB 1516959 (CR Matthews).

References

[1]. Kauzmann W Some factors in interpretation of protein denaturation. Adv Prot Chem. 1959;14:1–63.

[2]. Thirumalai D, Samanta HS, Maity H, Reddy G. Universal Nature of Collapsibility in the Context of Protein Folding and Evolution. Trends Biochem Sci. 2019;44:675–87. [PubMed: 31153683]

[3]. Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. Nature Struct Biol. 1999;6:554–6. [PubMed: 10360359]

[4]. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR. Early Collapse is not an Obligate Step in Protein Folding. J Mol Biol. 2004;338:369–82. [PubMed: 15066438]

[5]. Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR. Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. J Mol Biol. 2007;367:609–15. [PubMed: 17292402]

[6]. Yoo TY, Meisburger SP, Hinshaw J, Pollack L, Haran G, Sosnick TR, et al. Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state. J Mol Biol. 2012;418:226–36. [PubMed: 22306460]

[7]. Kathuria SV, Kayatekin C, Barrea R, Kondrashkina E, Graceffa R, Guo L, et al. Microsecond barrier-limited chain collapse observed by time-resolved FRET and SAXS. J Mol Biol. 2014;426:1980–94. [PubMed: 24607691]

[8]. Skinner JJ, Yu W, Gichana EK, Baxa MC, Hinshaw JR, Freed KF, et al. Benchmarking all-atom simulations using hydrogen exchange. Proc Natl Acad Sci U S A. 2014;111:15975–80. [PubMed: 25349413]

[9]. Song JH, Gomes GN, Gradinaru CC, Chan HS. An Adequate Account of Excluded Volume Is Necessary To Infer Compactness and Asphericity of Disordered Proteins by Forster Resonance Energy Transfer. Journal of Physical Chemistry B. 2015;119:15191–202.

[10]. Riback JA, Bowman MA, Zmyslowski AM, Knoverek CR, Jumper JM, Hinshaw JR, et al. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. Science. 2017;358:238–41. [PubMed: 29026044]

[11]. Riback JA, Bowman MA, Zmyslowski A, Knoverek CR, Jumper J, Kaye EB, et al. Response to Comment on “Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water”. Science. 2018;361.

[12]. Riback JA, Bowman MA, Zmyslowski AM, Plaxco KW, Clark PL, Sosnick TR. Commonly used FRET fluorophores promote collapse of an otherwise disordered protein. Proc Natl Acad Sci U S A. 2019;116:8889–94. [PubMed: 30992378]

[13]. Flory PJ. Mean-Square Moments of Chain Molecules. Proc Natl Acad Sci U S A. 1964;51:1060–7. [PubMed: 16578566]

[14]. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, et al. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci U S A. 2004;101:12491–6. [PubMed: 15314214]

[15]. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. J Mol Biol. 1999;290:253–66. [PubMed: 10388571]

[16]. Fleming PJ, Rose GD. Do all backbone polar groups in proteins form hydrogen bonds? Protein Sci. 2005;14:1911–7. [PubMed: 15937286]

[17]. Watters AL, Deka P, Corrent C, Callender D, Varani G, Sosnick T, et al. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. Cell. 2007;128:613–24. [PubMed: 17289578]

[18]. Lai Z, McCulloch J, Lashuel HA, Kelly JW. Guanidine hydrochloride-induced denaturation and refolding of transthyretin exhibits a marked hysteresis: equilibria with high kinetic barriers. Biochemistry. 1997;36:10230–9. [PubMed: 9254621]

[19]. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. Proc Natl Acad Sci U S A. 2007;104:1528–33. [PubMed: 17251351]

[20]. Nettels D, Gopich IV, Hoffmann A, Schuler B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. Proc Natl Acad Sci U S A. 2007;104:2655–60. [PubMed: 17301233]

[21]. Nettels D, Muller-Spath S, Kuster F, Hofmann H, Haenni D, Ruegger S, et al. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. Proc Natl Acad Sci U S A. 2009;106:20740–5. [PubMed: 19933333]

[22]. Sherman E, Haran G. Coil-globule transition in the denatured state of a small protein. Proc Natl Acad Sci U S A. 2006;103:11539–43. [PubMed: 16857738]

[23]. Ziv G, Thirumalai D, Haran G. Collapse transition in proteins. Phys Chem Chem Phys. 2009;11:83–93. [PubMed: 19081910]

[24]. Ziv G, Haran G. Protein folding, protein collapse, and tanford’s transfer model: lessons from single-molecule FRET. J Am Chem Soc. 2009;131:2942–7. [PubMed: 19239269]

[25]. O’Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. Proc Natl Acad Sci U S A. 2008;105:13403–8. [PubMed: 18757747]

[26]. Pugh SD, Gell C, Smith DA, Radford SE, Brockwell DJ. Single-molecule studies of the Im7 folding landscape. J Mol Biol. 2010;398:132–45. [PubMed: 20211187]

[27]. Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. Proc Natl Acad Sci U S A. 2012;109:16155–60. [PubMed: 22984159]

[28]. Magg C, Kubelka J, Holtermann G, Haas E, Schmid FX. Specificity of the initial collapse in the folding of the cold shock protein. J Mol Biol. 2006;360:1067–80. [PubMed: 16815441]

[29]. Waldauer SA, Bakajin O, Lapidus LJ. Extremely slow intramolecular diffusion in unfolded protein L. Proc Natl Acad Sci U S A. 2010;107:13713–7. [PubMed: 20643973]

[30]. Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, Pande VS. Unfolded-state dynamics and structure of protein L characterized by simulation and experiment. J Am Chem Soc. 2010;132:4702–9. [PubMed: 20218718]

[31]. Dasgupta A, Udgaonkar JB. Evidence for Initial Non-specific Polypeptide Chain Collapse During the Refolding of the SH3 Domain of PI3 Kinase. J Mol Biol. 2010;403:430–45. [PubMed: 20837026]

[32]. Huang F, Lerner E, Sato S, Amir D, Haas E, Fersht AR. Time-resolved fluorescence resonance energy transfer study shows a compact denatured state of the B domain of protein A. Biochemistry. 2009;48:3468–76. [PubMed: 19222162]

[33]. Sadqi M, Lapidus LJ, Munoz V. How fast is protein hydrophobic collapse? Proc Natl Acad Sci U S A. 2003;100:12117–22. [PubMed: 14530404]

[34]. Haran G How, when and why proteins collapse: the relation to folding. Curr Opin Struct Biol. 2012;22:14–20. [PubMed: 22104965]

[35]. Voelz VA, Jager M, Yao S, Chen Y, Zhu L, Waldauer SA, et al. Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment. J Am Chem Soc. 2012;134:12565–77. [PubMed: 22747188]

[36]. Teufel DP, Johnson CM, Lum JK, Neuweiler H. Backbone-driven collapse in unfolded protein chains. J Mol Biol. 2011;409:250–62. [PubMed: 21497607]

[37]. Goluguri RR, Udgaonkar JB. Microsecond rearrangements of hydrophobic clusters in an initially collapsed globule prime structure formation during the folding of a small protein. J Mol Biol. 2016;428:3102–17. [PubMed: 27370109]

[38]. Lapidus LJ, Acharya S, Schwantes CR, Wu L, Shukla D, King M, et al. Complex pathways in folding of protein G explored by simulation and experiment. Biophys J. 2014;107:947–55. [PubMed: 25140430]

[39]. Bhatia S, Krishnamoorthy G, Dhar D, Udgaonkar JB. Observation of Continuous Contraction and a Metastable Misfolded State during the Collapse and Folding of a Small Protein. J Mol Biol. 2019;431:3814–26. [PubMed: 31330152]

[40]. Reddy G, Thirumalai D. Collapse Precedes Folding in Denaturant-Dependent Assembly of Ubiquitin. J Phys Chem B. 2017;121:995–1009. [PubMed: 28076957]

[41]. Fuertes G, Banterle N, Ruff KM, Chowdhury A, Mercadante D, Koehler C, et al. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. Proc Natl Acad Sci U S A. 2017;114:242–50.

[42]. Borgia A, Zheng W, Buholzer K, Borgia MB, Schuler A, Hofmann H, et al. Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. J Am Chem Soc. 2016;138:11714–26. [PubMed: 27583570]

[43]. Piana S, Klepeis JL, Shaw DE. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr Op Structural Biology. 2014;24C:98–105.

[44]. Zheng W, Zerze GH, Borgia A, Mittal J, Schuler B, Best RB. Inferring properties of disordered chains from FRET transfer efficiencies. J Chem Phys. 2018;148:123329. [PubMed: 29604882]

[45]. Watkins HM, Simon AJ, Sosnick TR, Lipman EA, Hjelm RP, Plaxco KW. Random coil negative control reproduces the discrepancy between scattering and FRET measurements of denatured protein dimensions. Proc Natl Acad Sci U S A. 2015;112:6631–6. [PubMed: 25964362]

[46]. Takahashi S, Yoshida A, Oikawa H. Hypothesis: structural heterogeneity of the unfolded proteins originating from the coupling of the local clusters and the long-range distance distribution. Biophys Rev. 2018;10:363–73. [PubMed: 29446056]

[47]. Feng H, Zhou B-R, Bai Y. Binding affinity and function of the extremely disordered protein complex containing human linker histone H1.0 and Its chaperone ProTα. Biochemistry. 2018;57:6645–8. [PubMed: 30430826]

[48]. Qu S, Liu C, Liu Q, Wu W, Du B, Wang J. Solvent effect on FRET spectroscopic ruler. J Chem Phys. 2018;148:123331. [PubMed: 29604875]

[49]. Riback JA, Katanski CD, Kear-Scott JL, Pilipenko EV, Rojek AE, Sosnick TR, et al. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. Cell. 2017;168:1028–40 e19. [PubMed: 28283059]

[50]. Schwartz R, King J. Frequencies of hydrophobic and hydrophilic runs and alternations in proteins of known structure. Protein Sci. 2006;15:102–12. [PubMed: 16373477]

[51]. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. Science. 1995;267:1619–20. [PubMed: 7886447]

[52]. Levinthal C Mossbauer Spectroscopy in Biological Systems (proceedings of a meeting held at Allerton House, Monticello, Ill)1969. p. 22–4.

[53]. Sosnick TR, Mayne L, Hiller R, Englander SW. The barriers in protein folding. Nature Struct Biol. 1994;1:149–56. [PubMed: 7656032]

[54]. Sosnick TR, Mayne L, Englander SW. Molecular collapse: the rate-limiting step in two-state cytochrome c folding. Proteins. 1996;24:413–26. [PubMed: 9162942]

[55]. Otzen DE, Oliveberg M. Salt-induced detour through compact regions of the protein folding landscape. Proc Natl Acad Sci U S A. 1999;96:11746–51. [PubMed: 10518521]

[56]. Chen T, Song J, Chan HS. Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. Curr Opin Struct Biol. 2015;30:32–42. [PubMed: 25544254]

[57]. Faisca PF, Plaxco KW. Cooperativity and the origins of rapid, single-exponential kinetics in protein folding. Protein Sci. 2006;15:1608–18. [PubMed: 16815915]

[58]. Braselmann E, Chaney JL, Clark PL. Folding the proteome. Trends Biochem Sci. 2013;38:337–44. [PubMed: 23764454]

[59]. Malhotra P, Udgaonkar JB. How cooperative are protein folding and unfolding transitions? Protein Sci. 2016;25:1924–41. [PubMed: 27522064]

[60]. Clark PL. Protein folding in the cell: reshaping the folding funnel. Trends Biochem Sci. 2004;29:527–34. [PubMed: 15450607]

[61]. Chan HS, Zhang Z, Wallin S, Liu Z. Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. Annual review of physical chemistry. 2011;62:301–26.

[62]. Krantz BA, Srivastava AK, Nauli S, Baker D, Sauer RT, Sosnick TR. Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. Nature Struct Biol. 2002;9:458–63. [PubMed: 11979278]

[63]. Krantz BA, Moran LB, Kentsis A, Sosnick TR. D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. Nature Struct Biol. 2000;7:62–71. [PubMed: 10625430]

[64]. Krantz BA, Dothager RS, Sosnick TR. Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. J Mol Biol. 2004;337:463–75. [PubMed: 15003460]

[65]. Pandit AD, Jha A, Freed KF, Sosnick TR. Small proteins fold through transition states with native-like topologies. J Mol Biol. 2006;361:755–70. [PubMed: 16876194]

[66]. Baxa MC, Freed KF, Sosnick TR. Psi-constrained simulations of protein folding transition states: implications for calculating Phi values. J Mol Biol. 2009;386:920–8. [PubMed: 19244613]

[67]. Baxa M, Freed KF, Sosnick TR. Quantifying the Structural Requirements of the Folding Transition State of Protein A and Other Systems. J Mol Biol. 2008;381:1362–81. [PubMed: 18625237]

[68]. Baxa MC, Yu W, Adhikari AN, Ge L, Xia Z, Zhou R, et al. Even with nonnative interactions, the updated folding transition states of the homologs Proteins G & L are extensive and similar. Proc Natl Acad Sci U S A. 2015;112:8302–7. [PubMed: 26100906]

[69]. Yu W, Baxa MC, Gagnon I, Freed KF, Sosnick TR. Cooperative folding near the downhill limit determined with amino acid resolution by hydrogen exchange. Proc Natl Acad Sci U S A. 2016.

[70]. Makarov DE, Plaxco KW. The topomer search model: A simple, quantitative theory of two-state protein folding kinetics. Protein Sci. 2003;12:17–26. [PubMed: 12493824]

[71]. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol. 1998;277:985–94. [PubMed: 9545386]

[72]. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. Biochemistry. 2000;39:11177–83. [PubMed: 10985762]

[73]. Englander SW, Mayne L. The case for defined protein folding pathways. Proc Natl Acad Sci U S A. 2017;114:8253–8. [PubMed: 28630329]

[74]. Englander SW, Mayne L. The nature of protein folding pathways. Proc Natl Acad Sci U S A. 2014;111:15873–80. [PubMed: 25326421]

[75]. Walters BT, Mayne L, Hinshaw JR, Sosnick TR, Englander SW. Folding of a large protein at high structural resolution. Proc Natl Acad Sci U S A. 2013;110:18898–903. [PubMed: 24191053]

[76]. Hu W, Walters BT, Kan ZY, Mayne L, Rosen LE, Marqusee S, et al. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. Proc Natl Acad Sci U S A. 2013;110:7684–9. [PubMed: 23603271]

[77]. Adhikari AN, Freed KF, Sosnick TR. Simplified protein models: predicting folding pathways and structure using amino acid sequences. Phys Rev Lett. 2013;111:028103. [PubMed: 23889448]

[78]. Adhikari A, Freed KF, Sosnick TR. De novo prediction of protein folding pathways and structure using the principle of sequential stabilization. Proc Natl Acad Sci U S A. 2012;109:17442–7. [PubMed: 23045636]

[79]. Gates ZP, Baxa MC, Yu W, Riback JA, Li H, Roux B, et al. Perplexing cooperative folding and stability of a low-sequence complexity, polyproline 2 protein lacking a hydrophobic core. Proc Natl Acad Sci U S A. 2017;114:2241–6. [PubMed: 28193869]

[80]. Contreras-Martos S, Piai A, Kosol S, Varadi M, Bekesi A, Lebrun P, et al. Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. Sci Rep. 2017;7:4676. [PubMed: 28680062]

[81]. Martin EW, Holehouse AS, Grace CR, Hughes A, Pappu RV, Mittag T. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. J Am Chem Soc. 2016;138:15323–35. [PubMed: 27807972]

[82]. Kolonko M, Ozga K, Holubowicz R, Taube M, Kozak M, Ozyhar A, et al. Intrinsic Disorder of the C-Terminal Domain of Drosophila Methoprene-Tolerant Protein. PLoS One. 2016;11:e0162950. [PubMed: 27657508]

[83]. Shell SS, Putnam CD, Kolodner RD. The N terminus of Saccharomyces cerevisiae Msh6 is an unstructured tether to PCNA. Mol Cell. 2007;26:565–78. [PubMed: 17531814]

[84]. Renner M, Paesen GC, Grison CM, Granier S, Grimes JM, Leyrat C. Structural dissection of human metapneumovirus phosphoprotein using small angle x-ray scattering. Sci Rep. 2017;7:14865. [PubMed: 29093501]

[85]. Watson MC, Curtis JE. Probing the average local structure of biomolecules using small-angle scattering and scaling laws. Biophys J. 2014;106:2474–82. [PubMed: 24896127]

[86]. Yang C, van der Woerd MJ, Muthurajan UM, Hansen JC, Luger K. Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. Nucleic Acids Res. 2011;39:4122–35. [PubMed: 21278419]

[87]. Konno T, Tanaka N, Kataoka M, Takano E, Maki M. A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. Biochim Biophys Acta. 1997;1342:73–82. [PubMed: 9366272]

[88]. Bressan GC, Silva JC, Borges JC, Dos Passos DO, Ramos CH, Torriani IL, et al. Human regulatory protein Ki-1/57 has characteristics of an intrinsically unstructured protein. J Proteome Res. 2008;7:4465–74. [PubMed: 18788774]

[89]. Gazi AD, Bastaki M, Charova SN, Gkougkoulia EA, Kapellios EA, Panopoulos NJ, et al. Evidence for a coiled-coil interaction mode of disordered proteins from bacterial type III secretion systems. J Biol Chem. 2008;283:34062–8. [PubMed: 18836182]

[90]. Lens Z, Dewitte F, Monte D, Baert JL, Bompard C, Senechal M, et al. Solution structure of the N-terminal transactivation domain of ERM modified by SUMO-1. Biochemical and biophysical research communications. 2010;399:104–10. [PubMed: 20647002]

[91]. Boze H, Marlin T, Durand D, Perez J, Vernhet A, Canon F, et al. Proline-rich salivary proteins have extended conformations. Biophys J. 2010;99:656–65. [PubMed: 20643086]

[92]. Alborghetti MR, Furlan AS, Silva JC, Paes Leme AF, Torriani IC, Kobarg J. Human FEZ1 protein forms a disulfide bond mediated dimer: implications for cargo transport. J Proteome Res. 2010;9:4595–603. [PubMed: 20812761]

[93]. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, et al. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A. 2008;105:5762–7. [PubMed: 18391200]

[94]. Moncoq K, Broutin I, Craescu CT, Vachette P, Ducruix A, Durand D. SAXS study of the PIR domain from the Grb14 molecular adaptor: a natively unfolded protein with a transient structure primer? Biophys J. 2004;87:4056–64. [PubMed: 15465854]

Box 1.

Possible origins for rapid folding in a good solvent

When water is a good solvent for the DSE, non-specific chain collapse is thermodynamically unfavorable and cannot serve to prune the complexity of the conformational search for the native structure. Under such circumstances, how can rapid protein folding occur? Thermodynamically, of course, folding occurs because other, favorable energy terms, including the precise arrangement of side chains, hydrogen bonds and electrostatic interactions, offset the unfavorable energetics of chain collapse, including the loss conformational entropy and backbone desolvation. This said, of course, thermodynamic stability alone does not ensure that a chain can find its native conformation on a biologically relevant time scale.

What factors, then, might accelerate folding relative to an exhaustive conformation search (i.e., on a relatively flat energy landscape), enabling the folding of many single domain proteins to occur on the microseconds to seconds timescale? At one level, intrinsic, sequence-dependent conformational biases (for example, dihedral angle biases) likely enhance sampling of local segments of native-like structure, such as nascent helices and hairpins. Although too weak to significantly contract the conformational ensemble on their own, these local biases could help overcome the Levinthal search problem by lowering the energy of some conformations relative to others, confining the search to relatively lower energy (albeit still thermodynamically “uphill”) routes while still avoiding the formation of kinetic traps [61]. The height of the energy barrier is undoubtedly reduced by cooperative effects. That is, subsequent events “take advantage” of either prior contacts (e.g., closing an interior loop within a larger closed loop) or previously folded elements (e.g., docking a nascent helix onto an existing hairpin). Supporting this view, studies of kinetic amide isotope effects [62, 63] and ψ-value analyses [64–69] have found that the transition state ensemble (TSE) often contains native-like hydrogen bonded structure.

The above view of structure formation in folding notwithstanding, a critical step for overcoming the kinetic barrier for folding of single-domain proteins appears to be adoption of a native *topomer*, one of the subset of conformations whose overall topology approximates that of the native structure. These conformations must resemble the native structure sufficiently such that the chain need not cross itself nor break stable interactions (which would be slow) in order to rapidly “zipper up” the remaining structure all the way to the native state [70]. Conversely, the ensemble of conformations in the native topomer must also be sufficiently large, such that access to a native topomer is kinetically accessible. Evidence supporting this model comes from the surprisingly strong correlation between folding rates and various measure of the topological complexity of the native state [70–72] and y-value studies demonstrating the TSE of single-domain proteins achieve a native-like topology [62–64, 69]. As folding progresses further, beyond the TSE, additional elements of structure will build onto folded regions in a process of sequential stabilization, a model supported by considerable hydrogen exchange data [73–76] and simulations [77, 78].

Highlights

- Unfolded states of globular proteins often are expanded in water, suggesting that water is xa “good solvent.”
- This occurs, despite the fact that the hydrophobic effect drives folding, implying that folding differs fundamentally from non-specific collapse.
- Water being a good solvent for unfolded states may be beneficial, as this suppresses misfolding and non-specific interactions.

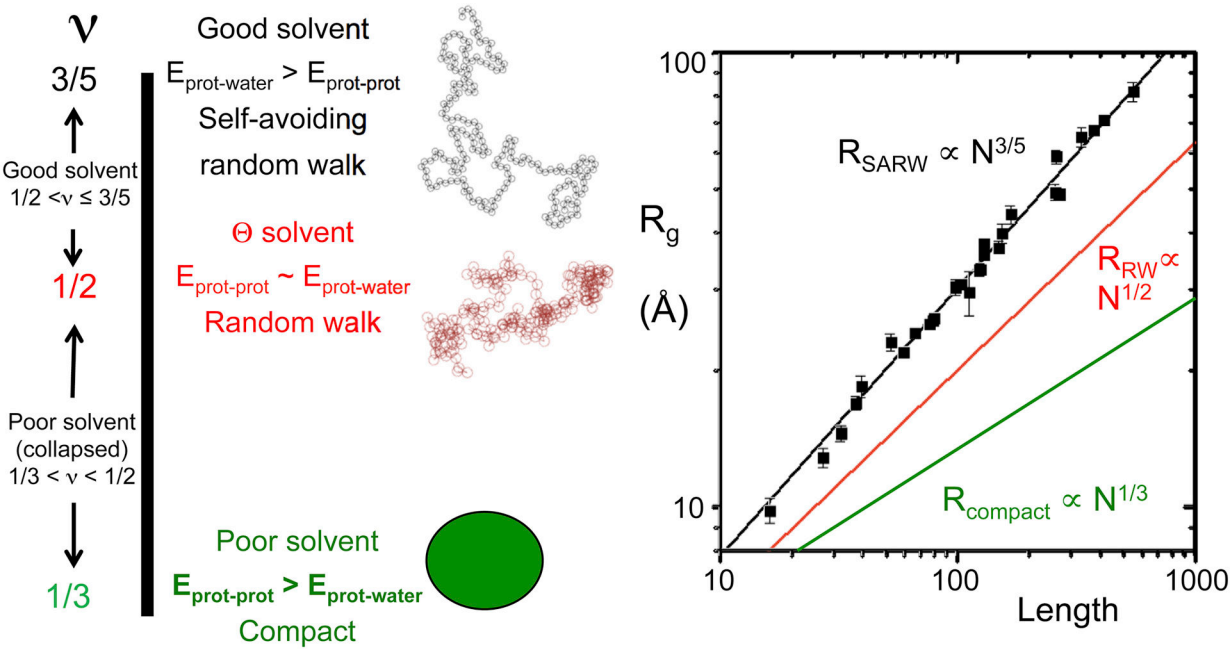

Figure 1. The scaling of chain dimensions with chain length provides a measure of solvent quality.

The Flory exponent, ν, relating chain length to a physical dimension (e.g., R_g or $R_{end\text{-}to\text{-}end}$) canonically adopts values of 1/3, 1/2, and 3/5, corresponding to compact states, random walks or chains at the "θ condition", and self-avoiding random walks (SARW), respectively. As shown, proteins unfolded at high levels of chemical denaturant produce the 3/5 exponent expected for a SARW [14]. Formally, only the three canonical ν values will be observed for chains of infinite length. For chains of finite length, however, ν can assume intermediate values. Here we define a poor solvent as one that produces a ν of less than 1/2 (i.e., below the θ point) and a good solvent as one that produces a ν above this cutoff.

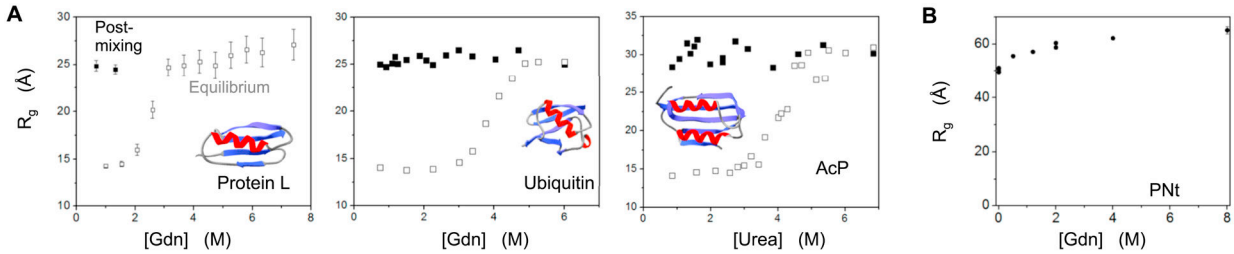

Figure 2. The DSE of many proteins do not contract significantly upon a shift from high to low denaturant.

(**A**) Shown are the SAXS-derived radii of gyration (R_g) of three globular, single domain proteins, collected either at equilibrium at various denaturant concentrations (gray) or immediately after transfer from high levels of denaturant to low (■). Specifically, the dimensions observed immediately after denaturant dilution (i.e., after the 2 to 4 ms dead-time) for Protein L [3, 6], ubiquitin and acylphosphotase (AcP) [4] are within error of those observed at equilibrium at much higher denaturant concentrations. (**B**) Equilibrium SAXS measurements of PNt, a 334 amino acid protein that is intrinsically disordered, likewise indicate no contraction of the chain above 0.5 M Gdn. Mild contraction is seen at still lower denaturant concentrations, but this corresponds to a mild decrease in ν (from 0.57 to 0.54; ref [10], indicating that these conditions remain well above the θ point that defines the cutoff for poor solvent quality.

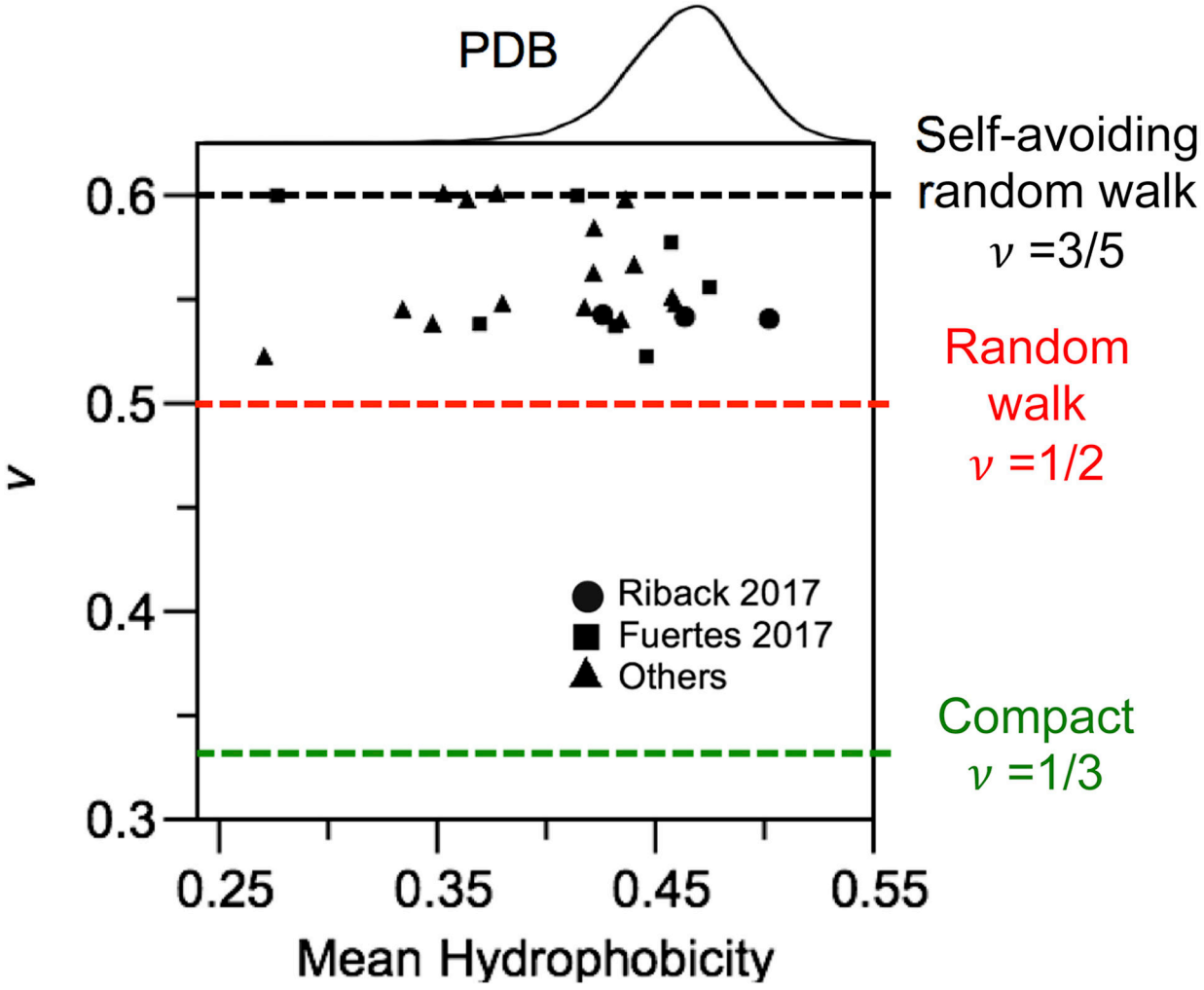

Figure 3. SAXS indicates that water is a good solvent for many DSEs.

Shown are scaling exponents, ν, versus hydrophobicity (Kyte-Doolittle) for foldable protein sequences in the absence of denaturant. These were determined from published SAXS data analyzed using our molecular form factor [10, 12, 41, 79–94]. A histogram of the hydrophobicity of representative proteins from the PDB is presented at the top of the figure (dataset from [10]).