Published in final edited form as:

Curr Opin Struct Biol. 2012 February ; 22(1): 4–13. doi:10.1016/j.sbi.2011.09.002.

Residual structure in unfolded proteins

Bruce E. Bowler

Department of Chemistry and Biochemistry and Center for Biomolecular Structure and Dynamics,
The University of Montana, Missoula, MT 59812 USA

Abstract

The denatured state ensemble (DSE) of unfolded proteins, once considered to be well-modeled by an energetically featureless random coil, is now well-known to contain flickering elements of residual structure. The position and nature of DSE residual structure may provide clues toward deciphering the protein folding code. This review focuses on recent advances in our understanding of the nature of DSE collapse under folding conditions, the quantification of the stability of residual structure in the DSE, the determination of the location and types of residues involved in thermodynamically significant residual structure and advances in detection of long-range interactions in the DSE.

Introduction

For many years, the unfolded states of proteins were considered to behave as unstructured polymers with no persistent nonrandom interactions along the length of the polypeptide chain [1]. This random coil model of the unfolded state held sway until around 1990 when thermodynamic studies of site-directed variants of staphylococcal nuclease (SNase) [2] and NMR studies of proteins under strongly denaturing conditions [3] began to show that the unfolded state was considerably more complex. Unlike the native state of a protein, which has a unique structure, the unfolded state is comprised of a broad structural ensemble and thus is considerably more difficult to characterize. To reflect this diversity of structure, the unfolded state will be referred to as the denatured state ensemble (DSE) in this review. The complexity of this ensemble has required characterization by a breadth of methods including ensemble and single molecule fluorescence and fluorescence resonance energy transfer (FRET) methods and small angle X-ray scattering (SAXS) to define the dimension of and nature of long-range interactions in the DSE, NMR methods to define both local and long-range structural interactions, thermodynamic methods to define the strength and nature of interactions in the DSE and molecular dynamics (MD) and Monte Carlo simulations to provide detailed structural insight into this complex state.

As the starting point for the folding of a protein, the structural and thermodynamic biases of the DSE may hold important clues into the “folding code”, which unlike the genetic code has proven difficult to decipher because of its redundancy. With this in mind, this review focuses on advances in our understanding of the nature and specificity of collapse in the DSE when it is switched from denaturing to folding conditions, advances in our ability to quantify the strength of and to identify the location of thermodynamically significant

© 2011 Elsevier Ltd. All rights reserved.

Corresponding author: Bowler, Bruce E. (bruce.bowler@umontana.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

interactions along a polypeptide chain, and advances in our ability to identify long-range residual structure in the DSE which may be important in setting up the topology of a fold. Literature from the last two to three years will be emphasized with reference to earlier literature when important for context.

Effects of solvent quality on polypeptides

The effect of solvent quality on the DSE of proteins has been an area of considerable interest. Collapse of the unfolded state under folding conditions reduces conformational space and could induce formation of ordered structure [4]. Both these factors are widely believed to be important for efficient folding. Small-angle X-ray scattering (SAXS) and various fluorescence methods have been the primary techniques used to assess the compactness of unfolded or disordered proteins. While fluorescence methods generally show compaction of the DSE as solvent quality becomes poorer [5], several proteins studied by SAXS do not appear to collapse immediately upon transfer to folding conditions [6]. In the case of protein L, the two methods disagree [6]. A recent SAXS study on the folding of barnase (110 amino acids) addressed the effect of polypeptide length on degree of collapse [7], given that all proteins >100 amino acids in length appear to collapse, as assessed by SAXS, upon transfer to folding conditions, whereas smaller proteins often do not [8]. For barnase, dilution to folding conditions produced only a modest decrease in the radius of gyration, R_g , ($26.9 \pm 0.7 \text{ \AA}$ to $23.9 \pm 0.2 \text{ \AA}$), much less than the decrease to $R_g \approx 19 \text{ \AA}$ predicted based on the behavior of proteins >100 amino acids in length [7,8]. These data suggest that the collapse behavior of proteins near 100 amino acids may be more dependent on folding mechanism than that of either smaller or larger proteins. In the case of barnase, the folding nucleus primarily involves the N-terminal region of the protein and thus R_g might be expected to be less prone to decrease early in folding.

Another issue of keen interest is whether collapse early in folding is mediated by backbone hydrogen bonding or interactions between hydrophobic side chains. A comprehensive study using fluorescence correlation spectroscopy (FCS) has provided important new insights into this question [9]. In aqueous solution, the hydrodynamic radius, R_h , of a G_{20} polypeptide with all amide NH groups methylated (NMe- G_{20}) was found to be significantly larger than that of G_{20} . In 8 M guanidine hydrochloride (GdnHCl), both polypeptides were found to have similar R_h with the expansion of the N-methylated peptide being modest (Figure 1A). These results provide direct experimental evidence of the importance of backbone hydrogen bonds in collapse and are consistent with the primary interaction of denaturants being with the backbone [10,11]. Simulations of G_{15} also demonstrate that non-specific backbone hydrogen bonds lead to a highly collapsed structure in water, whereas in 8 M urea intramolecular hydrogen bonds are replaced with hydrogen bonds to urea leading to an extended structure [12]. FCS studies on the 28-residue intrinsically-disordered protein (IDP), kinase-inducible activation domain (KID), show that it also expands in 8 M GdnHCl [9]. Interestingly, conversion of its 7 hydrophobic residues to serine (KID-noHP) had little effect on R_h in water indicating non-polar residues make a relatively small contribution to collapse of the DSE under folding conditions (Figure 1B). By contrast, conversion of its 11 charged residues to serine led to substantial compaction, consistent with the strong influence of electrostatics in the DSE [13]. Surprisingly, loop formation kinetics measured by photo-induced electron transfer-FCS (PET-FCS) in water do not correlate well with R_h [9]. Side chains appear to slow loop formation through intrachain interactions [9,14]. By contrast, the faster loop formation kinetics of (GS) $_{10}$ relative to NMe- G_{20} suggest that backbone hydrogen bonding enhances the rate constant for first contact, k_c , a conclusion that is supported by molecular dynamics simulations [15]. However, the smallest k_c was observed with G_{20} , which is the most compact of this set of 20-residue polypeptides (Figure 1A), can form backbone hydrogen bonds, yet has no side chains. Our understanding of diffusion in

compact polymers remains incomplete. Recent work indicates that diffusion in a compact DSE may be considerably slower than in an expanded DSE [16]. Since loop formation is critical early in folding, a better understanding of how solvent quality affects loop formation is needed, particularly for a compact DSE under folding conditions.

If collapse of the DSE early in folding is dominated by the backbone, how the side chains and their order along a polypeptide mediate the specificity needed to achieve a unique topology remains in question. A variant of the fyn SH3 domain lacking the 4 C-terminal residues, which is unfolded in water, and a sequence randomized variant of the fyn SH3 domain both have similar natively compactness [17]. Thus, compactness alone in the absence of the sequence-specific ordering of the side chains is insufficient to specify a unique fold. As we will see below, for foldable sequences, the DSE is often biased toward the topology of the native state.

Recent simulations, however, suggest that backbone hydrogen bonding could mediate conformational specificity during the collapse of the DSE early in folding [18,19]. Poorer solvent conditions cause a redistribution in ϕ, ψ space that could be important for establishing the gross features of fold topology during collapse. In particular, in a good solvent the “bridge region” of the Ramachandran plot between the β -basin and the α -region, which includes ϕ, ψ combinations needed for the type I β -turn, is disfavored because the amide NH at the $i+1$ position of the turn cannot be solvated [19]. Under poor solvent conditions, the “bridge region” becomes favorable. Thus, amino sequences that favor type I β -turns could be important in establishing the gross features of fold topology during compaction of the DSE [18], suggesting the possibility of a backbone-mediated component to the “folding code”.

Thermodynamic characterization of residual structure

Thermodynamic methods have long been important in evaluating residual structure in the DSE. Shortle’s observation of large changes in denaturant m -values (slope of a plot of free energy of unfolding versus denaturant concentration, $\delta\Delta G_u/\delta C$, which is proportional to the change in solvent accessible surface area, ΔSASA , associated with unfolding) resulting from single amino acid mutations to SNase provided the early impetus for this approach [2]. Studies on the pH dependence of protein stability have shown that electrostatic interactions modulate the free energy of the DSE by up to 4 kcal/mol [13,20,21]. Earlier work on the thermodynamics of residual structure in the DSE has been summarized in detail [20].

More recently, Tanford’s transfer model (TM) [1] has been an important focus of work on the DSE [10]. In this model, the free energy of unfolding, ΔG_u , depends on the favorable transfer free energy, ΔG_{tr} , of the polypeptide chain from water to a denaturing solvent. ΔG_{tr} can be broken down into components for the individual side chains and the backbone, $\Delta g_{tr,i}$, with the contribution of each component depending on its average fractional ΔSASA , α_i , when the protein unfolds (Eq 1, where $\Delta G_u^\circ(\text{H}_2\text{O})$ is the free energy of unfolding in the absence of

$$\Delta G_u = \Delta G_u^\circ(\text{H}_2\text{O}) + \sum_i \alpha_i n_i \Delta g_{tr,i} \quad (1)$$

denaturant and n_i is the number of groups of type i). Simulations using the TM are able to replicate FRET data for the R_g of the DSE, show that collapse of the DSE under poor solvent conditions leads to secondary structure, that α -helical structure in particular persists at high denaturant concentrations and that the response of a given type of side chain to solvent conditions is context-dependent [22–24]. It is now possible to use the TM to

quantitatively predict m -values and to dissect the contributions of individual side chains and the backbone to the stability of the DSE using $\Delta g_{tr,i}$ corrected for the activity coefficients of glycine in water versus 1 M urea [10,25,26]. Using a truncated version of the *Drosophila* notch receptor ankyrin repeat protein, Nank4-7*, Bolen and coworkers showed that the DSE is stabilized by 13.1 kcal/mol upon transfer from water to 6 M urea. The majority of this stabilization is due to the backbone, consistent with Molecular Dynamics (MD) simulations which quantitatively reproduce this strong stabilization of the backbone by urea [27]. Although the simulations show that urea hydrogen bonds well to the backbone (see also ref. [11]), the stabilization mainly results from better van der Waals interactions of urea with the backbone relative to water. By contrast, the interaction of the side chains of Nank4-7* with 6 M urea is small and unfavorable with nonpolar side chains only modest contributors. Thus, urea does not strongly perturb hydrophobic interactions. Structural studies on proteins in high denaturant concentrations clearly show that nonpolar interactions persist [3,20]. Thus, the contribution of nonpolar residues to the specificity of the “folding code” is not abrogated at high denaturant concentration supporting the notion that denaturants may simply scale protein stability through interactions with the main chain without modifying the determinants of the “folding code” [28]. Although transiently populated, these specificities as we show below can be detected. Given the denaturant dependence of the Ramachandran plot discussed above [18,19], it will be important to better discern how the “folding code”, as expressed in DSE biases, partitions between sequence-dependent backbone bias and the biases of side chain interactions [29,30].

A recent survey of m -values from urea unfolding compared experimental m -values to m -values calculated with the TM using compact versus extended models to calculate the solvent accessible surface area of the denatured state [31]. The analysis indicates that residual structure in the DSE varies widely for urea-unfolded proteins with proteins such as barstar and CheY having a maximally compact DSE and even proteins with the most unfolded DSE at pH 7, SNase and barnase, retaining considerable residual structure. Urea unfolding at low pH produced m -values most consistent with the extended model for the DSE, consistent with NMR studies, which show that low pH and high urea concentration produce a DSE with the least evidence for residual structure [32]. Urea unfolding of variants of ribonuclease Sa (RNase Sa) with high positive charge had m -values at pH 3 consistent with the highest degree of solvent exposure based on the TM [31]. CD experiments showed a correlation between high m -values and high polyproline II (PPII) structure. For proteins with high m -values, the presence of low energy pathways through the “bridge region” of the Ramachandran plot from the PPII to turn and α -helix regions will be important for efficient folding [19].

The heat capacity increment, ΔC_p , associated with the temperature-dependent unfolding of a protein, like the denaturant m -value, is proportional to the Δ SASA associated with protein unfolding. Thus, ΔC_p is also a sensitive monitor of the compactness of the denatured state [20]. Recently, measurements of ΔC_p coupled to site-directed mutagenesis have been used to define loci of residual structure in the DSE. For RNase Sa, a D79F mutation decreases ΔC_p from 1.68 to 1.07 kcal mol⁻¹K⁻¹ [33]. Second site variants indicate that F79, I92 and Y80 stabilize a natively-like hydrophobic cluster in the DSE. For ribonuclease H1 (RNase H1), the lower ΔC_p in RNase H1 from the moderate thermophile, *Chlorobium tepedium*, and the thermophile, *Thermus thermophilus*, compared to RNase H1 from the mesophile, *E. coli*, localizes to the folding core versus the periphery of the protein [34]. Mutagenesis work showed that natively-like isoleucine, leucine, valine (ILV) clusters stabilize residual structure and decrease ΔC_p for RNase H1 from the thermophiles. Thus, thermodynamically significant residual structure can be induced by clusters of predominately aromatic or large aliphatic side chains.

Equilibrium His-heme loop formation in denaturing concentrations of GdnHCl has been used to probe deviations from random coil behavior along the sequence of the four-helix bundle protein, cytochrome *c'* (Cyt*c'*). Substantial scatter about a log-log plot of loop stability versus loop size (Figure 2) demonstrates that there is a high degree of sequence-dependent variability in loop stability – up to 10-fold for adjacent portions of the sequence [35,36]. Interestingly, the pattern of deviations from random coil behavior along the sequence is identical at 3 M and 6 M GdnHCl (Figure 2A), consistent with results from the TM [10,25,26] show that nonpolar interactions persist at high denaturant concentration. By contrast, polyalanine sequences engineered into iso-1-cytochrome *c* (iso-1-Cyt*c*) adhere very closely to a the linear log-log dependence of loop stability on loop size expected for a random coil [37] (Figure 2A). Thus, relative to a homopolymer of alanine, the side chains of a foldable heteropolymer like Cyt*c'* lead to sequence-dependent biases that could seed fold topology. Given the recent results on solvent quality effects on backbone conformational biases [18,19], it will be important to understand how the sequence-dependent biases of the DSE of Cyt*c'* partition between the backbone and the side chains. In this regard, comparison of His-heme loop formation with iso-1-Cyt*c* containing polyglycine sequences with the data for polyalanine sequences shows that the scaling exponent for loop formation, ν_3 , decreases for polyglycine as GdnHCl concentration decreases whereas ν_3 for polyalanine increases as GdnHCl concentration decreases [38]. Thus, the nature of the backbone driven collapse is different for polymers of these two amino acids. Simulations indicate that collapse of polyglycine [12] and polyglutamine [39] results from non-specific backbone hydrogen bonding, whereas that for polyalanine favors γ -turns [18,40]. Identifying which amino acids tend to cause ordered versus non-specific collapse also could be important in understanding how sequence specifies structure.

Similarly, determining the relative ability of different amino acids to stabilize residual structure in the DSE is essential for understanding the role of each amino acid in the “folding code”. Using iso-1-Cyt*c* with an (AAAXAK) insert, the effect of changing X from A to Y, W, F and L on the stability of a 22-residue His-heme loop was measured in 3 M GdnHCl [41]. All 3 aromatic residues stabilized the loop by 0.4 to 0.5 kcal/mol in 3 M GdnHCl, whereas leucine had a negligible effect on loop stability. While the changes in loop stability are small in magnitude, the fact that single aromatic or aliphatic to alanine mutations are adequate to break up hydrophobic clusters in the DSE of RNase Sa [33] and RNase H1 [34] suggests that interactions of modest magnitude are sufficient to change the structural bias of the DSE. For all three alanine to aromatic substitutions, the stabilization of the His-heme loop was due primarily to a decrease in the rate constant for His-heme loop breakage. Thus, it is possible that biases of backbone collapse may initiate the search to find the correct fold topology and that hydrophobic interactions are used as a second filter to select the correct topology, as has been suggested recently [9].

Recently, an innovative mutant cycle method has been developed to probe the importance of different amino acids in stabilizing a collapsed DSE (Figure 3A). The method, which combines mutagenesis and addition of NaCl to perturb the stability of the compact DSE, has been applied to nucleophosmin C-terminal domain (Cter-NPM1), a small 3-helix bundle protein [42,43]. The results show that individual amino acids in helices 2 and 3 contribute 0.5 to 1.6 kcal/mol to the stability of the collapsed DSE [43]. Aromatic and large aliphatic residues are the largest contributors to the stability of the collapsed state (Figure 3B, C), consistent with the types of residues that lead to hydrophobic clusters that lower ΔC_p in proteins from thermophiles [33,34] and stabilize His-heme loop formation in the DSE [41]. An Ala to Gly mutation near the N-terminus of helix 3 produced one of the larger effects on the stability of the compact DSE. Given the different ways in which Gly and Ala appear to mediate backbone collapse [12,18,38] this alanine may be particularly important for backbone mediated control of the topology of DSE collapse for Cter-NPM1. Many of the

residues important in stabilizing the compact DSE of Cter-NPM1 also yield negative kinetic ϕ values, consistent with the effect of these residues on folding kinetics originating from the DSE [44].

Structural characterization of the denatured state

NMR methods have been particularly useful in characterizing residual structure in the DSE [45,46]. NMR data in combination with R_h or R_g data from pulsed-field gradient NMR and SAXS, respectively, have provided constraints for developing reasonable structural models of the DSE [47,48] for the drkN SH3 domain [49] and α -synuclein [50]. In both cases, the ensembles generated contained both compact and extended structures. For the drkN SH3 domain, natively like secondary structure is evident, as well as a non-native hydrophobic cluster centered around Trp 36 and a small segment of non-native helix. The observation of both native and non-native structure in the DSE is typical [6]. Recent NMR studies show that both the cold denatured state [51] and the denatured state at pH 3.8 [52] of the C-terminal domain of protein L9 contain both native and non-native structure. Studies on the c-src SH3 domain, an all- β fold, show that residual structure in the DSE is primarily α -helical, is located in loop regions and is not conserved relative to residual structure in other SH3 domains despite the conserved nature of the transition state ensemble of this fold [53]. Consistent with this observation, theoretical modeling of the thermodynamics of the DSE indicates that relative to regions of the native state that are helical, regions of the native state that form β -sheet tend have a much lower structure forming propensity [54]. It will be interesting to see if greater diversity in DSE structure emerges as a general observation for all- β versus all- α folds.

Thermodynamic and kinetic studies coupled to MD simulations have also proven useful in understanding the relationship between structure and stability in the DSE. MD simulations have shown that the DSE of Cyt c' is compact [35] (Figure 2D). In particular, dynamic hydrophobic clusters in the DSE, with both native and non-native interactions, maintain the general topological features of a short loop that connects helices 1 and 2 in the native state, as well as a long 20-residue Ω -loop at the base of helices 2 and 3 (Figure 2D). These persistent chain reversals in the MD simulations of the DSE correspond to portions of the Cyt c' sequence which form His-heme loops with slow breakage rates (small k_b – see Figure 2B, C). Thus, local sequence appears to be selected such that nonpolar interactions bias the DSE of Cyt c' toward its native topology. The persistence of a 20-residue loop in the DSE given the entropic cost relative to smaller loops [55] is somewhat surprising, however, long-range natively like interactions have been detected very early in the folding of adenylate kinase using FRET methods [56]. A combination of thermodynamic and kinetic methods and MD simulations also has shown that the DSE plays a key role in defining the topologies of the designed proteins, G $_A$ 88 (all- α , three helix bundle) and G $_B$ 88 (protein G α + β fold), which differ by only 5 residues [57]. As with Cyt c' [35], essential elements of the native topology are evident in the DSE of G $_A$ 88 and G $_B$ 88. In both cases, long-range side chain mediated hydrogen bonds bias the DSE toward natively like β -hairpins (G $_B$ 88) or natively like helical structure (G $_A$ 88). Recent work on the folding of proteins with deeply knotted native state topologies has shown that the knotted topology is maintained in the DSE even in 6 M GdnHCl [58]. Thus, an increasing number of examples of proteins that maintain the gross features of their native topology in the DSE have emerged in the last couple of years.

Measurement of ^{15}N NMR transverse relaxation rates coupled to mutational analysis [59] and measurement of paramagnetic relaxation enhancement (PRE) of ^1H NMR nuclei [45] have been particularly effective qualitative tools for measuring long-range interactions in the DSE. Strategies that allow better definition and quantification of long-range interactions detected by NMR have emerged recently. The precision and accuracy with which NMR

chemical shifts can be measured has been used to detect and quantify tertiary interactions in the DSE by coupling mutational analysis to measurements of changes in secondary shifts [60,61]. Use of truncated forms of apoMyoglobin (apoMb) show that the presence of helix H, but not helix G, enhances secondary structure in the helices A, B and C of the pH 2.4 acid denatured state [61]. Long-range interactions in the pH 2.4 acid denatured state of acyl coenzyme A binding protein (ACBP) were assessed by measuring the effects of single-site mutations dispersed throughout the sequence on secondary shifts [60]. Both native and non-native interactions between helices 2, 3 and 4 were observed (Figure 4). Small decreases in the secondary shift of a helix produced similar effects on secondary shifts in distant helices, consistent with transient formation of cooperative tertiary interactions in the DSE. The changes in secondary shifts were consistent with up to a 7% change in the helicity of distant helices. Detailed modeling of an extensive set of PRE data also has provided estimates of the population of species with long-range contacts in the DSE of apoMb. The analysis is consistent with <5% of the DSE forming collapsed structures with long-range contacts in acid denatured apoMb [62]. Thus, both the apoMb and ACBP results indicate that long-range contacts with natively-like topology are flickering components of the DSE.

Role of the DSE in folding kinetics

Does DSE bias make protein folding more efficient? In the few cases, where the thermodynamics of residual structure in the DSE has been perturbed in a known direction, the folding rate constant moves in the expected direction [20]. However, in other cases disruption of natively-like residual structure has no apparent effect on folding [20]. The recent studies on Cter-NPM1 indicate that stabilization of the compact DSE speeds folding [43]. Non-random natively-like structure in the DSE of fast-folding proteins is also suggestive of an important role for the DSE in efficient folding [14,63]. In the case of ACBP, kinetic ϕ value analysis indicates that the residual structure in the DSE and the residues which participate in the transition state (TS) are from different parts of the protein. Flickering structure in the DSE could directly provide the elements of the TS, equally well residual structure in the DSE could provide a template for assembling more disordered parts of the DSE in the TS. The uncertainties surrounding the role of residual structure in the DSE in promoting efficient folding indicate that this is an area in need of further investigation.

Conclusion

Significant advances in our understanding of the nature and specificity of DSE collapse in poor solvents have been made in the last several years. These data suggest that backbone hydrogen bonding mediates this process in a manner that allows collapse to a globular structure with some side chains favoring ordered hydrogen-bonded structure and others non-specific hydrogen bonding. Thorough studies are only available for Ala, Gly and Gln. Studies on other homopolymers both experimentally and by simulation could provide important insights into the specificity of backbone mediated collapse and its role in establishing fold topology.

Hydrogen-bond mediated collapse to a globular structure, however, is insufficient to establish a unique fold. Studies on the thermodynamics of the DSE point to aromatics and large aliphatic residues as important in stabilizing hydrophobic residual structure. Earlier work indicates that electrostatic interactions are also important in nonrandom behavior in the DSE. New NMR methods are beginning to allow detection of long-range interactions in the DSE, which combined with our growing knowledge of which amino acids stabilize residual structure, may yield insights into how the “folding code” provides smooth landscapes that lead to unique structures. The database of proteins with well-characterized DSE's is small and the available data suggest that the DSE's of all- α and all- β proteins may behave

differently. The recent observation that the DSE's of a number of proteins are biased – if only transiently – toward native topology suggests that detailed thermodynamic and structural characterization of the DSE could provide important insights into the “folding code”. Thus, development of a larger database of well-characterized DSEs from a broader variety of fold topologies could provide important advances in our understanding of how a specific amino acid sequence folds to a unique structure.

Acknowledgments

The author acknowledges the support of the National Institutes of Health for the on-going support of his work on protein denatured states most recently through R01GM074750 and the efforts of the many excellent students who have made the work possible over the years.

References and recommended reading

Papers of particular interest published within the review period have been highlighted as:

- of special interest
- of outstanding interest

1. Tanford C. Protein denaturation. *Adv Protein Chem.* 1968; 23:121–282. [PubMed: 4882248]
2. Shortle D. Staphylococcal nuclease: a showcase of *m*-value effects. *Adv Protein Chem.* 1995; 46:217–245. [PubMed: 7771319]
3. Neri D, Billeter M, Wider G, Wüthrich K. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science.* 1992; 257:1559–1563. [PubMed: 1523410]
4. Dill KA. Dominant forces in protein folding. *Biochemistry.* 1990; 29:7133–7155. [PubMed: 2207096]
5. Schuler B, Eaton WA. Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol.* 2008; 18:16–26. [PubMed: 18221865]
6. Sosnick TR, Barrick D. The folding of single domain proteins - have we reached a consensus? *Curr Opin Struct Biol.* 2011; 21:12–24. [PubMed: 21144739]
7. Konuma T, Kimura T, Matsumoto S, Goto Y, Fujisawa T, Fersht AR, Takahashi S. Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase. *J Mol Biol.* 2011; 405:1284–1294. [PubMed: 21146541]
8. Uzawa T, Kimura T, Ishimori K, Morishima I, Matsui T, Ikeda-Saito M, Takahashi S, Akiyama S, Fujisawa T. Time-resolved small-angle X-ray scattering investigation of the folding dynamics of heme oxygenase: implication of the scaling relationship for the submillisecond intermediates of protein folding. *J Mol Biol.* 2006; 357:997–1008. [PubMed: 16460755]
- 9•• Teufel DP, Johnson CM, Lum JK, Neuweiler H. Backbone driven collapse in unfolded protein chains. *J Mol Biol.* 2011; 409:250–262. An important contribution providing firm experimental evidence for the conclusion that backbone hydrogen bonding mediates collapse of the DSE. [PubMed: 21497607]
10. Bolen DW, Rose GD. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem.* 2008; 77:339–362. [PubMed: 18518824]
11. Lim WK, Rösgen J, Englander SW. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc Natl Acad Sci USA.* 2009; 106:2595–2600. [PubMed: 19196963]
12. Tran HT, Mao A, Pappu RV. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc.* 2008; 130:7380–7392. [PubMed: 18481860]
13. Cho J-H, Sato S, Horng J-C, Anil B, Raleigh DP. Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability. *Arch Biochem Biophys.* 2008; 469:20–28. [PubMed: 17900519]

14. Neuweiler H, Johnson CM, Fersht AR. Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. *Proc Natl Acad Sci USA*. 2009; 106:18569–18574. [PubMed: 19841261]
15. Daidone I, Neuweiler H, Doose S, Sauer M, Smith JC. Hydrogen-bond driven loop-closure in unfolded polypeptide chains. *PLoS Comp Biol*. 2010; 6:1–9.
16. Waldauer SA, Bakajin O, Lapidus LJ. Extremely slow intramolecular diffusion in unfolded protein L. *Proc Natl Acad Sci USA*. 2010; 107:13713–13717. [PubMed: 20643973]
17. Kohn JE, Gillespie B, Plaxco KW. Non-sequence-specific interactions can account for the compaction of proteins unfolded under native conditions. *J Mol Biol*. 2009; 394:343–350. [PubMed: 19751743]
18. Gong H, Porter LL, Rose GD. Counting peptide hydrogen bonds in unfolded proteins. *Protein Sci*. 2010; 20:417–427. [PubMed: 21280132]
- 19•. Porter LL, Rose GD. Redrawing the Ramachandran plot after inclusion of hydrogen-bonding interactions. *Proc Natl Acad Sci USA*. 2011; 108:109–113. An intriguing simulation study showing that ϕ, ψ space depends on solvent conditions and that low energy routes from extended to globular conformations are possible upon switching to folding conditions. [PubMed: 21148101]
20. Bowler BE. Thermodynamics of protein denatured states. *Mol BioSyst*. 2007; 3:88–99. [PubMed: 17245488]
21. Arbely E, Rutherford TJ, Neuweiler H, Sharpe TD, Ferguson N, Fersht AR. Carboxyl pK_a values and acid denaturation of BBL. *J Mol Biol*. 2011; 403:313–327. [PubMed: 20816989]
- 22•. O'Brien EP, Brooks BR, Thirumalai D. Molecular origin of constant m -values, denatured state collapse, and residue-dependent transition midpoints in globular proteins. *Biochemistry*. 2009; 48:3743–3754. Detailed presentation of a simulation method built on the TM model. Important new insights into DSE properties include demonstration that the response of an amino acid to denaturant concentration in the DSE depends on sequence context. [PubMed: 19278261]
23. 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 USA*. 2008; 105:13403–13408. [PubMed: 18757747]
24. Liu Z, Reddy G, O'Brien EP, Thirumalai D. Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc Natl Acad Sci USA*. 2011; 108:7787–7792. [PubMed: 21512127]
- 25•. Holthauzen LMF, Roesgen J, Bolen DW. Hydrogen bonding progressively strengthens upon transfer of the protein urea-denatured state to water and protecting osmolytes. *Biochemistry*. 2010; 49:1310–1318. A beautifully-designed experimental study that directly shows the response of the DSE to solvent conditions and the importance of the TM model in analysis of the DSE. [PubMed: 20073511]
26. Auton M, Holthauzen LMF, Bolen DW. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc Natl Acad Sci USA*. 2007; 104:15317–15322. [PubMed: 17878304]
27. Hu CY, Kokubo H, Lynch GC, Bolen DW, Pettitt BM. Backbone additivity in the transfer model of protein solvation. *Protein Sci*. 2010; 19:1011–1022. [PubMed: 20306490]
- 28•. Lattman EE, Rose GD. Protein folding--what's the question? *Proc Natl Acad Sci USA*. 1993; 90:439–441. It's old, but classic. Read it if you haven't, it's prescient. [PubMed: 8421673]
29. Chakrabarti P, Bhattacharyya R. Geometry of nonbonded interactions involving planar groups in proteins. *Prog Biophys Mol Biol*. 2007; 95:83–137. [PubMed: 17629549]
30. Saha RP, Bahadur RP, Chakrabarti P. Interresidue contacts in proteins and protein-protein interfaces and their use in characterizing the homodimeric interface. *J Proteome Res*. 2005; 4:1600–1609. [PubMed: 16212412]
- 31•. Pace CN, Huyghues-Despointes BMP, Fu H, Takano K, Scholtz JM, Grimsley GR. Urea denatured state ensembles contain extensive secondary structure that is increased in hydrophobic proteins. *Protein Sci*. 2010; 19:929–943. An analysis of urea m -values in terms of the TM model for a set of 39 proteins with important implications for interpretation of m -values in terms of DSE structure. [PubMed: 20198681]

32. Shan B, Bhattacharya S, Eliezer D, Raleigh DP. The low-pH unfolded state of the C-terminal domain of the ribosomal protein L9 contains significant secondary structure in the absence of denaturant but is no more compact than the low-pH urea unfolded state. *Biochemistry*. 2008; 47:9565–9573. [PubMed: 18707127]
33. Fu H, Grimsley G, Scholtz JM, Pace CN. Increasing protein stability: importance of ΔC_p and the denatured state. *Protein Sci*. 2010; 19:1044–1052. [PubMed: 20340133]
34. Ratcliff K, Marqusee S. Identification of residual structure in the unfolded state of ribonuclease H1 from the moderately thermophilic *Chlorobium tepidum*: comparison with thermophilic and mesophilic homologs. *Biochemistry*. 2010; 49:5167–5175. [PubMed: 20491485]
35. Dar TA, Schaeffer RD, Daggett V, Bowler BE. Manifestations of native topology in the denatured state ensemble of *Rhodopseudomonas palustris* cytochrome *c'*. *Biochemistry*. 2011; 50:1029–1041. A combined experimental and simulation study showing that cytochrome *c'* is predisposed toward its native topology by residual structure in the DSE. [PubMed: 21190388]
36. Rao KS, Tzul FO, Christian AK, Gordon TN, Bowler BE. Thermodynamics of loop formation in the denatured state of *Rhodopseudomonas palustris* cytochrome *c'*: scaling exponents and the reconciliation problem. *J Mol Biol*. 2009; 392:1315–1325. [PubMed: 19647747]
37. Tzul FO, Bowler BE. Denatured states of low complexity polypeptide sequences differ dramatically from those of foldable sequences. *Proc Natl Acad Sci USA*. 2010; 107:11364–11369. [PubMed: 20534569]
38. Finnegan, ML.; Bowler, BE. Experimental measurement of the effect of glycine on main chain flexibility and scaling properties. 2011. Submitted
39. Vitalis A, Wang X, Pappu RV. Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *J Mol Biol*. 2008; 384:279–297. [PubMed: 18824003]
40. Gong H, Rose GD. Assessing the solvent-dependent surface area of unfolded proteins using an ensemble model. *Proc Natl Acad Sci USA*. 2008; 105:3321–3326. [PubMed: 18305164]
41. Finnegan ML, Bowler BE. Propensities of aromatic amino acids versus leucine and proline to induce residual structure in the denatured-state ensemble of iso-1-cytochrome *c*. *J Mol Biol*. 2010; 403:495–504. A host-guest application of the His-heme loop formation method that allows quantitative measurement of the ability of different amino acids to stabilize residual structure in the DSE. [PubMed: 20850458]
42. Scaloni F, Gianni S, Federici L, Brunangelo F, Brunori M. Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. *FASEB J*. 2009; 23:2360–2365. [PubMed: 19289604]
43. Scaloni F, Federici L, Brunori M, Gianna S. Deciphering the folding transition state structure and denatured state properties of nucleophosmin C-terminal domain. *Proc Natl Acad Sci USA*. 2011; 107:5447–5452. A novel mutant cycle approach that permits quantification of the magnitude of stabilizing interactions in a compact DSE. [PubMed: 20212148]
44. Cho J-H, Raleigh DP. Denatured state effects and the origin of nonclassical ϕ values in protein folding. *J Am Chem Soc*. 2006; 128:16492–16493. [PubMed: 17177385]
45. Eliezer D. Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol*. 2009; 19:23–30. [PubMed: 19162471]
46. Bowler, BE. Globular proteins: characterization of the denatured state. In: Egelman, E., editor. *Comprehensive Biophysics*. Elsevier; 2012. in press
47. Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol*. 2007; 17:3–14. [PubMed: 17250999]
48. Vendruscolo M. Determination of conformationally heterogeneous states of proteins. *Curr Opin Struct Biol*. 2007; 17:15–20. [PubMed: 17239581]
49. Marsh JA, Forman-Kay JD. Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J Mol Biol*. 2009; 391:359–374. [PubMed: 19501099]
50. Allison JR, Varnai P, Dobson CM, Vendruscolo M. Determination of the free energy landscape of alpha-synuclein using spin label Nuclear Magnetic Resonance measurements. *J Am Chem Soc*. 2009; 131:18314–18326. [PubMed: 20028147]

51. Shan B, McClendon S, Rospigliosi C, Eliezer D, Raleigh DP. The cold denatured state of the C-terminal domain of protein L9 is compact and contains both native and non-native structure. *J Am Chem Soc.* 2010; 132:4669–4677. [PubMed: 20225821]
52. Shan B, Eliezer D, Raleigh DP. The unfolded state of the C-terminal domain of the ribosomal protein L9 contains both native and non-native structure. *Biochemistry.* 2009; 48:4707–4719. [PubMed: 19301913]
53. Rösner HI, Poulsen FM. Residue-specific description of non-native transient structures in the ensemble of acid-denatured structures of the all- β protein c-src SH3. *Biochemistry.* 2010; 49:3246–3253. [PubMed: 20218679]
54. Wang S, Gu J, Larson SA, Whitten ST, Hilser VJ. Denatured-state energy landscapes of a protein structural database reveal the energetic determinants of a framework model for folding. *J Mol Biol.* 2008; 381:1184–1201. An innovative application of the Corex algorithm that shows that DSE thermodynamic attributes are better at fold prediction than native state thermodynamic attributes. [PubMed: 18616947]
55. Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. *Annu Rev Biophys.* 2008; 37:289–316. [PubMed: 18573083]
56. Orevi T, Ishay EB, Pirchi M, Jacob MH, Amir D, Haas E. Early closure of a long loop in the refolding of adenylate kinase: a possible key role of non-local interactions in the initial folding steps. *J Mol Biol.* 2009; 385:1230–1242. [PubMed: 19013178]
57. Morrone A, McCully ME, Bryan PN, Brunori M, Daggett V, Gianni S, Travaglini-Allocatelli C. The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. *J Biol Chem.* 2011; 286:3863–3872. [PubMed: 21118804]
58. Mallam AL, Rogers JM, Jackson SE. Experimental detection of knotted conformations in denatured proteins. *Proc Natl Acad Sci USA.* 2010; 107:8189–8194. [PubMed: 20393125]
59. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H. Long-range interactions within a nonnative protein. *Science.* 2002; 295:1719–1722. [PubMed: 11872841]
60. Bruun SW, Ieřmantavius V, Danielsson J, Poulsen FM. Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proc Natl Acad Sci USA.* 2010; 107:13306–13311. A sensitive new method combining mutagenesis with measurement of NMR secondary shifts that provides site-specific insight into long-range interactions in the DSE. [PubMed: 20624986]
61. Fedyukina DV, Rajagopalan S, Sekhar A, Fulmer EC, Eun Y-J, Cavagnero S. Contributions of long-range interactions to the secondary structure of an unfolded globin. *Biophys J.* 2010; 99:L37–L39. [PubMed: 20816043]
62. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE. Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci USA.* 2008; 105:6278–6283. [PubMed: 18434548]
63. Meng W, Shan B, Tang Y, Raleigh DP. Native like structure in the unfolded state of the villin headpiece helical subdomain, an ultrafast folding protein. *Protein Sci.* 2009; 18:1692–1701. [PubMed: 19598233]

Highlights

- Denatured state ensemble collapse is mediated by backbone hydrogen bonds
- Denaturant concentration affects the Ramachandran plot
- Site-specific determination of residual structure stability is now possible
- Long-range nativelylike topology exists in the denatured state ensemble

Figure 1.

Effect of solvent quality on R_h of polypeptides. (A) Comparison of R_h in water (solid bar) and 8 M GdnHCl (open bar) for the 20 residue polypeptides G_{20} , NMe- G_{20} and $(GS)_{10}$. R_h in water (solid bar) and 8 M GdnHCl (open bar) for 20 residue Trp-cage miniprotein is shown on the right. (B) Comparison of R_h in water (solid bars) and 8 M GdnHCl (open bars) for the IDP, KID, and the variants with all nonpolar (KID-noHP) or all charged (KID-noCH) residues mutated to serine. Adapted from ref. [9] with permission.

Figure 2. Characterization of the DSE of Cyt c' by thermodynamic and kinetic methods and MD simulations. (A) Plot of loop stability, $pK_{\text{loop}}(\text{His})$ versus loop size, n , in 3 M (●) and 6 M (Δ) GdnHCl for His-heme loop formation in the DSE of Cyt c'. Each data point is labeled with the site of the histidine mutation used to form the loop. Red circles are data for polyalanine sequences. The solid and dashed lines are fits to the dependence of loop stability on loop size expected for a random coil: $pK_{\text{loop}}(\text{His}) = pK_{\text{loop}}(\text{His})_{\text{ref}} + \nu_3 \log(n)$, where ν_3 is the scaling exponent, n is the number of monomers in the loop and $pK_{\text{loop}}(\text{His})_{\text{ref}}$ is $pK_{\text{loop}}(\text{His})$ for $n = 1$. (B) Plot of k_b versus loop size, n , for breakage of His-heme loops in 3 M GdnHCl. Loops with the smallest k_b are circled in cyan. (C) Structure of Cyt c' showing the sites of single histidine variants used for His-heme loop formation in the DSE of Cyt c'. The mutation sites with the smallest values of k_b are shown in cyan. (D) MD simulation of the DSE of Cyt c' showing residual structure in the region including the Ω loop between helices 2 and 3. Residues 50–78 are colored red and shown for the final 60 ns structure of an MD unfolding simulation at 498 K (left). Mutation sites are shown in cyan. Side chain

positions of hydrophobic cluster participants are shown in detail for the final structure (right). Adapted from ref. [35] with permission.

Figure 3.

Denatured state structure of Cter-NPM1 as obtained from protein engineering. (A) Double perturbation cycle. The cube depicts the different states of the native (N) and denatured (D) conformations populated as a result of a double perturbation (mutagenesis, $\Delta\Delta G_u^{\circ\text{mut}} = \Delta G_u^{\circ\text{mut}} - \Delta G_u^{\circ\text{WT}}$ and addition of stabilizing salt, $\Delta\Delta G_u^{\circ\text{salt}} = \Delta G_u^{\circ\text{mut,salt}} - \Delta G_u^{\circ\text{WT,salt}}$). The stabilization of the DSE (B and C) is given by the coupling free, $\Delta\Delta\Delta G_u^{\circ\text{mut,salt}} = \Delta\Delta G_u^{\circ\text{salt}} - \Delta\Delta G_u^{\circ\text{mut}}$. The method assumes that the salt perturbation acts entirely on the DSE. $\Delta\Delta\Delta G_u^{\circ\text{mut,salt}}$ is mapped onto the structure in two orientations. Color-coding is Black, $\Delta\Delta\Delta G_u^{\circ\text{mut,salt}} < 0.5 \text{ kcal mol}^{-1}$; Green, $0.5 \text{ kcal mol}^{-1} < \Delta\Delta\Delta G_u^{\circ\text{mut,salt}} < 1 \text{ kcal mol}^{-1}$; Blue, $\Delta\Delta\Delta G_u^{\circ\text{mut,salt}} > 1 \text{ kcal mol}^{-1}$. Used with permission from ref. [43].

Figure 4.

Mapping of the mutation-induced chemical shift changes observed in the pH 2.4 DSE of ACBP onto the peptide backbone of the native structure of ACBP. The color code for the magnitude of the mutation induced changes in the secondary shifts is shown in the color code bar on the right. Local and long-range stabilizing (red) and destabilizing (blue) effects on secondary structure in the DSE are shown for 4 variants of ACBP. The structures have been labeled with the sites of the mutation V77A, I27A, I39A, and P44A, and their positions in the structures are each shown in green as a stick model of the native residue. The positions of the four helices A1–A4 in the structure are marked in each structure. Used with permission from ref. [60].