NIH Public Access
Author Manuscript
Proteins. Author manuscript; available in PMC 2011 October 1.

Published in final edited form as:
Proteins. 2010 October ; 78(13): 2725–2737. doi:10.1002/prot.22803.

Dry Molten Globule Intermediates and the Mechanism of Protein Unfolding

Robert L. Baldwin[1], **Carl Frieden**[2], and **George D. Rose**[3]

[1]Department of Biochemistry Stanford University Medical Center Beckman Center; School of Medicine 300 Pasteur Drive Stanford, CA 94305-5307

[2]Department of Biochemistry and Molecular Biophysics Washington University School of Medicine 660 South Euclid Avenue St. Louis, MO 63110

[3]Jenkins Department of Biophysics Johns Hopkins University Jenkins Hall 3400 N. Charles Street Baltimore, MD 21218

Abstract

New experimental results show that either gain or loss of close packing can be observed as a discrete step in protein folding or unfolding reactions. This finding poses a significant challenge to the conventional two-state model of protein folding. Results of interest involve dry molten globule intermediates, an expanded form of the protein that lacks appreciable solvent. When an unfolding protein expands to the dry molten globule state, side chains unlock and gain conformational entropy, while liquid-like van der Waals interactions persist. Four unrelated proteins are now known to form dry molten globules as the first step of unfolding, suggesting that such an intermediate may well be commonplace in both folding and unfolding. Data from the literature show that peptide amide protons are protected in the dry molten globule, indicating that backbone structure is intact despite loss of side chain close packing. Other complementary evidence shows that secondary structure formation provides a major source of compaction during folding. In our model, the major free-energy barrier separating unfolded from native states usually occurs during the transition between the unfolded state and the dry molten globule. The absence of close packing at this barrier provides an explanation for why ϕ-values, derived from a Brønsted-Leffler plot, depend primarily on structure at the mutational site and not on specific side chain interactions. The conventional two-state folding model breaks down when there are dry molten globule intermediates, a realization that has major implications for future experimental work on the mechanism of protein folding.

Protein folding: the historical mindset

Experimental evidence that small proteins fold via a 2-state reaction, U(nfolded) *f* N(ative), has become an anchoring idea in protein folding. Early work with RNase A showed that unrelated probes of the native structure (e.g., optical rotation, ultraviolet absorption, viscosity, etc.) describe the same equilibrium folding curve after suitable normalization. This finding could be rationalized if folding is regarded as a cooperative, all-or-none process. Specifically, en route from U to N, the population would consist of two subpopulations, one predominantly unfolded, the other predominantly folded, with a negligible population of transitional intermediate forms. Many later studies reinforced this early finding. To physicists this type of process suggested a first order phase transition, like the freezing of a liquid. The thermodynamically simplifying realization that protein folding

Corresponding author: Robert L. Baldwin, BaldwinB@Stanford.edu, 650-723-6168 (voice), 650-723-6783 (fax).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

could be accurately modeled as a 2-state reaction set the stage for many subsequent contributions, such as m-values[1], transition states and ϕ-values [2], and chevron plots [3].

Building on the 2-state model, work of Tanford [4] and others led to the conclusion that the U state is structurally featureless because its energy landscape is consistent with a vast number of accessible conformers separated by small energy barriers, of order k_BT. Under these conditions any given molecule in the population can visit accessible conformers readily. Consequently, persisting structural features would have to emerge in the N state, and therefore it seemed likely that those interactions which stabilize the native fold over other alternatives remain apparent in the X-ray structure, as in a Go model [5] (a popular energy function in folding simulations that rewards native contacts but does not penalize non-native contacts).

For authentic 2-state folders the free energy difference between the two states can be determined from the U *f* N equilibrium constant, K = [N]/[U], with $\Delta G° = -RT \ln K$. The folding process then involves selecting the correct constellation of favorable interactions from the vast number of possibilities present in the U state. How this happens has been a longstanding puzzle [6] that has given rise to current landscape views of protein folding [7,8].

The interpretation that the folding process lacks measureable structural intermediates – at least in small proteins – has become an entrenched conviction. We question that conviction in this perspective.

It is our conjecture that dry molten globules are commonplace structural intermediates in protein folding and unfolding. A dry molten globule (DMG) intermediate is an expanded form of the native fold in which solvent water has been expelled from the protein core but buried sidechains lack the close-packed character that is a familiar hallmark of the native state [9]. When unfolding, the side chains unlock, the protein gains conformational entropy, and the strength of the van der Waals interactions is reduced about 2-fold, as discussed below. We review evidence for the existence of DMG intermediates in four different proteins, where instances of this previously overlooked feature are detected during both unfolding and refolding.

The existence of on-pathway DMG intermediates implies that the folding (or unfolding) process can be separated into discrete steps. In particular, steps resulting in super-secondary structure formation and solvent squeezing need not require concomitant sidechain close packing. Consequently, factors that promote formation of these structures may not be apparent in the X-ray structure, contrary to earlier expectations. The search for additional DMG intermediates in other proteins is likely to guide future experimental directions in protein folding and, if productive, to result in a radical reinterpretation of the protein folding problem.

Existence of Dry Molten Globule Intermediates at the Start of Protein Unfolding

The first reports of DMG intermediates in 1995 [10,11] were received with general skepticism, and the protein folding literature was silent about them until 2009 -10, when two new reports were published [12,13]. Two main reasons explain the long period of skepticism: first, during this 14-year interval, experimentalists focused on the folding kinetics of small, fast-folding proteins that were consistent with the 2-state model, and they failed to find evidence for intermediates of any kind. Second, molecular dynamics (MD) simulations of unfolding generally failed to detect DMG-initiated unfolding. However, in 2008 an extended (1-μsec) simulation of urea-induced unfolding of hen lysozyme at 37° C reported 2-stage unfolding in which the fast initial stage has properties expected for DMG-initiation of unfolding [14]. Earlier MD simulations of unfolding were often performed at high temperature (above 100°

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

C), a protocol that speeds unfolding but with the unsuspected consequence of destabilizing possible DMG intermediates. Important features of the DMG reported [14] from simulations of the first stage of urea unfolding of hen lysozyme are: (1) urea molecules invade the protein and interact with the peptide backbone but (2) water does not invade the protein in a comparable manner.

When unfolding is monitored by standard optical probes such as Trp fluorescence or far-UV CD, the failure to detect DMG intermediates is readily explained. These probes detect water access to previously buried groups. When a buried Trp residue is exposed to water, the position of its fluorescence spectrum shifts, and the maximum fluorescence intensity decreases markedly. Correspondingly, when a solvent-shielded helix is exposed to water, its stabilizing side chain interactions with the protein interior are broken, and the far-UV CD spectrum diminishes or disappears entirely. Such changes in the solvent environment do not accompany formation of a DMG intermediate. Instead, the protein merely expands somewhat from a close-packed (locked) to a loose-packed (unlocked) state, but liquid-like van der Waals (vdW) interactions persist and water does not yet enter the core.

A comparable picture of the DMG was proposed in 1989 [15] on the basis of classical physics arguments, not simulations. However, there is a critical difference between this earlier proposal and later experimental work. As first proposed [15], the DMG was hypothesized to be the highly unstable transition state (denoted here as $I^{\ddagger}$) for a 2-state unfolding reaction, not an observable intermediate. Finding instead that DMGs are populated intermediates in unfolding[10-13] is a welcome game-changer because it will be possible to learn their properties through experimentation. It is anticipated that properties ascribed earlier to $I^{\ddagger}$ are likely now to be measurable properties of the DMG (see below).

Two classes of probes have been used to detect DMGs: (i) 1D-NMR spectra, either ^{1}H [10] or ^{19}F [11], that monitor loss of close-packed (sharp) NMR resonance lines, and (ii) energy transfer probes that either measure distance between donor and acceptor groups using fluorescence (FRET) [12] energy transfer or that detect contact between these groups using triplet-triplet energy transfer [13]. These energy transfer probes have been used to detect the structural rearrangement that occurs as a DMG is formed [12] and to detect a reversible reaction between locked and unlocked states [13], corresponding to the loss and regain of close-packing.

Now it is important to test the proposed existence of DMGs by using new probes and other approaches. Our view is that the evidence for DMGs and DMG-initiated unfolding is persuasive. It seems certain that new work will be undertaken to test this proposal. Although only four examples of DMGs have been reported, we suspect these intermediates have evaded detection because they are unseen by the standard probes of unfolding, as explained above. In fact, DMGs were found when suitable probes were used. This perspective assumes that DMG-initiated unfolding is a general property of at least one major class of proteins, and we discuss the consequences for the folding mechanism of this class.

Evidence for DMG Intermediates

The four reports of DMG unfolding intermediates are summarized next.

First report—The first study was a two-part search in 1995 for intermediates in the unfolding of ribonuclease A (RNase A) by using novel probes [10,16]. In the first paper [16], hydrogen exchange (HX) data were taken during unfolding in EX1 exchange conditions (fig. 1). Under these conditions the HX rate constant measures the unfolding rate constant. The HX measurements of unfolding kinetics were then compared with corresponding measurements using near-UV and far-UV CD. In the second paper [10], real-time 1D ^{1}H-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NMR spectra were taken during unfolding in the same conditions as the previous HX study. No unfolding intermediates were found in the HX study, but in fact, a rapidly formed intermediate was found in the NMR study. The HX study is of interest in itself because it provides basic information about water access to the H-bonded peptide backbone in a DMG (fig. 1). In these two studies the RNase A unfolding conditions were pH 8.0, 10° C, 4.5 M guanidinium chloride (GdmCl).

This study was repeated in 2002 under almost the same conditions [17]. Their HX experiments were made by pulse labeling, which allows more direct comparison between the unfolding rate constants measured by HX and by optical probes. In agreement with the earlier studies [10,16], the pulse labeling HX rate constant agreed with the near-UV CD rate constant, and no difference was found between the unfolding rate constants measured by near-UV and far-UV CD at 4.4M, although a slight difference was seen at 5.2 M.

The 1995 search for unfolding intermediates in RNase A [10,16] was motivated by an intriguing difference between unfolding and refolding kinetics. Specifically, medium-size proteins like RNase A (124 residues) were widely believed not to show intermediates in unfolding, although intermediates in refolding could be observed readily, for example by using the ratio test. In this test, reaction kinetics – either folding or unfolding – are compared by monitoring two very different probes, one probing secondary structure (far-UV CD) and the other probing tertiary structure (near-UV CD). In the absence of observable intermediates, the kinetics measured by the two probes are superimposable. The failure of the ratio test to detect intermediates is illustrated in the unfolding of RNase A [16], where probes of both secondary and tertiary structure follow the same single-exponential kinetics of unfolding. Yet when medium-size proteins refold, intermediates are routinely observed by the ratio test – as in the refolding kinetics of RNase [10] – because secondary structure forms before tertiary structure. This order of events gave rise in 1982 to the framework model of folding [18].

The intermediate (I) found in the NMR study of RNase A unfolding [10] was classified as a dry molten globule based on the proposed properties of DMGs that were outlined in the 1989 hypothesis for the mechanism of heat-induced protein unfolding [15]. In detail, the experimentally-determined RNase A unfolding intermediate failed to give the resolved methyl resonance line of Val 63 shown by N, so that the amount of N remaining after I was formed could be determined by measuring the intensity of this resonance line, and, in turn, the ratio [I]/[N]. The protein interior of I was clearly dry because the H-bonded NH protons of the peptide backbone were highly protected against exchange [16], and the sharp, well-resolved methylene resonance lines of N were no longer observed in I [10]. When [GdmCl] was varied, I was not detected until the GdmCl concentration reached the transition zone for equilibrium unfolding.

Second report—The second report of an intermediate in the unfolding of a medium-size protein, dihydrofolate reductase (DHFR) with 159 residues, was based on two important technical advances, stopped-flow mixing and ^{19}F labeling of the 5 Trp residues. The stopped-flow mixing time was 1.5 sec, and the 5 ^{19}F-Trp resonance lines, which are sharp and well-resolved in both N and U, are not observable in the intermediate. Urea-induced unfolding was monitored at pH 7.2, 22° C. The resonance lines of N disappeared shortly after mixing, but the lines of U appeared only slowly; consequently, an intermediate that was not detected by NMR must have been formed at the start of the slow phase. The fast phase of unfolding has a half-time ~1 sec while the major slow phase has a half-time ~50 sec according to the unfolding kinetics monitored by optical probes. The slow phase monitored by ^{19}F-NMR has kinetics similar to those monitored by either far-UV CD or Trp fluorescence. The intermediate formed in the slow phase was suggested to be a molten

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

globule because its 5 resonance lines were too broad to be observable. Further, the intermediate was dry as judged both by Trp fluorescence and far-UV CD intensity. As discussed later, side chain stabilization during refolding was highly cooperative, compatible with the formation of the native structure from a DMG. Unlike the unfolding behavior of RNase A, N was completely converted to the intermediate at the start of the slow phase of unfolding.

Third report—Following the two 1995 studies there was a lapse of 14 years before publication of the next paper reporting a DMG intermediate at the start of unfolding. This third report [12] analyzed unfolding kinetics in single-chain monellin (MNEI, 97 residues), measured at pH 8, 25° C, 3-6 M GdmCl. Unfolding kinetics were monitored by FRET (dead-time 6 msec), with the sole Trp residue (W4) serving as donor and with an acceptor dye placed at either of two locations. Kinetic measurements of near-UV CD demonstrated the rapid formation of an intermediate with an altered CD spectrum at the start of unfolding, and Trp fluorescence data indicated that the Trp chromophore was not exposed to water. The FRET measurements revealed changes in the distance between the donor probe (W4) and the acceptor dye, attached either at the C-terminus of the protein (Cys 97) or at the C-terminus of the single long helix (Cys 29). From these distances it is clear that the structure of the intermediate differs from N. The distances indicate displacement of the single helix and an overall expansion of the protein, suggesting a molten globule intermediate. That the molten globule is, in fact, dry was confirmed by its failure to bind ANS (8-anilino-1-naphthalenesulfonic acid), which does bind to equilibrium (wet) molten globules. The ratio [DMG]/[N] at the start of unfolding was simulated but not measured directly.

Fourth report—The fourth example of a DMG intermediate [13] was found by monitoring conformational fluctuations from equilibrium in the villin headpiece (HP35), a small (35-residue), fast-folding protein whose folding and unfolding kinetics have been studied intensively by other approaches [19,20]. The conformational fluctuations were detected by triplet-triplet energy transfer (TTET) that occurs only when donor (D) and acceptor (A) groups are in van der Waals (vdW) contact. A reversible equilibrium was found between locked (close packed) and unlocked (loosely packed) forms of the native protein. The N- and C-termini are close together in HP35, so when D and A are attached to the N- and C-terminal residues, fast triplet-triplet transfer occurs in the locked form of HP35 with slower triplet-triplet transfer (30 nsec) in the unlocked form. The unlocking mechanism is detected by triplet-triplet transfer between D and A attached to either end of helix 3 (fig. 2). No triplet-triplet transfer occurs in the locked form, while in the unlocked form triplet-triplet transfer occurs by partial unfolding of helix 3 (time constant 170 nsec). The time range of TTET is μsec - nsec, and conformational fluctuations could be detected only in this fast time range. The rate constants and the kinetic amplitudes of the steps involved in the conformational fluctuations were measured as functions of temperature (5° - 30° C) and [GdmCl], and the results were used to deduce the nature of the conformational reactions. A notable result is that the time constant of the close packing reaction is extremely fast, ~1 μsec at 5° C.

Another notable result is that the equilibrium constant K for the unlocking reaction could be measured. The changes in enthalpy and entropy for unlocking were measured from the dependence of ln K on (1/T): $\Delta H° = 8.37$ kcal mol^{-1}, $\Delta S° = 26.8$ cal mol^{-1} K^{-1}. Thus, the unlocking reaction is entropy-driven and is opposed by a substantial enthalpy barrier, as expected from the crystal-melting model below. The enthalpy change for unlocking is temperature-independent (i.e., $\Delta Cp°$ is 0 within error), as expected for breaking vdW interactions. If nonpolar side chains become solvated significantly as unlocking occurs, $\Delta Cp°$ would be measurable (see below). The stability of the DMG relative to the native protein depends on $\Delta G° = \Delta H° - T\Delta S°$, and $\Delta G°$ is small compared to $T\Delta S°$ (entropy-enthalpy

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

compensation) so that the equilibrium between N and DMG is measurable within the range studied, 5° - 30° C. Because ΔH° is large and unfavorable, increasing temperature causes unlocking to occur in a fairly narrow transition zone. The large values of ΔH° and ΔS° suggest that unlocking is fully cooperative for HP35, which is a very small protein; later work will show whether larger proteins share this property.

A Model System for the Energetics of Close Packing

The observation that proteins are close-packed and that close packing must an important factor in the energetics of folding has been known since 1971 [21]. In 1988, the proposal [22] was made that Kauzmann's model [23] for estimating the magnitude of the hydrophobic interaction – a major factor in the energetics of protein folding – should be modified to include close packing as a second major factor. Kauzmann's model is based on the unfavorable energetics of dissolving a nonpolar molecule in water. He proposed that the free energy change for burying solvent-exposed nonpolar side chains inside a protein could be estimated from the energetics of transferring hydrocarbon solutes between water and a suitable nonaqueous solvent. The transfer thermodynamics are simplest when the nonaqueous solvent is the hydrocarbon solute itself, in liquid form (i.e., the liquid hydrocarbon model, see [24]). For example, using toluene to model the Phe side chain, the free energy change for transferring this side chain from water to the interior of a protein can be estimated from the solubility of toluene in water. At 25° C this free energy change is −5.2 kcal/mol, using solubility on the mole fraction scale. Once Kauzmann pointed out the huge free energy changes produced by removing nonpolar side chains from water [23], protein chemists became convinced that the hydrophobic interaction is the dominant factor in the energetics of protein folding [25-27]. When linear alkanes of varying chain lengths are used in the liquid hydrocarbon model, ΔG° is −0.9 kcal/mol per buried carbon atom; this number can be calculated from published data in reference [28] by using solubility on the mole fraction scale.

The vdW interactions made by an alkane with either liquid alkane or liquid water are nearly the same [27,29], so that the vdW interactions essentially cancel out in this liquid-liquid transfer process. What matters energetically is the unfavorable free energy of making a water cavity large enough to accommodate the hydrocarbon solute [27,29]. Consequently, the basic physics of the hydrophobic interaction applies equally to other nonpolar molecules such as the noble gases. When two hydrocarbon side chains become juxtaposed in a refolding protein, their contacts with each other replace unfavorable contacts with water, and in this sense the term *hydrophobic interaction* has a straightforward meaning as applied to protein folding.

To include close packing in the energetics of folding, a desolvation-packing model [22] was proposed: the energetics of the desolvation step, in which nonpolar groups are removed from water, can be modeled from data for the transfer of hydrocarbons between the gas phase and water, with solubility calculated on the molarity scale for gas-liquid transfer [29,30]. In principle the energetics of the packing step can be assessed from the structural coordinates of a protein by computing the vdW interaction energies between neighboring nonbonded atoms, but in 1988 a suitable algorithm had yet to be developed.

In 1991 it was pointed out that thermal melting of alkane crystals provides a model system for estimating the energetics of close packing [31]; ample data are available for alkanes of varying chain lengths. This crystal-melting model gives ΔH = −0.6 kcal/mol per buried carbon atom for the enthalpy change when a liquid alkane crystallizes [31]. Comparing this value for ΔH with ΔG = −0.9 kcal/mol per buried carbon for the hydrophobic interaction (above) suggests that the energetics of close packing are indeed comparable to the energetics of the hydrophobic interaction in protein folding. However, there is substantial entropy-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

enthalpy compensation in the energetics of melting hydrocarbon crystals [31], and ΔG in this process is always much smaller than ΔH.

We note that close packing is an important factor in analyzing experiments with native proteins but not in analyzing folding intermediates that are not close-packed. According to the crystal-melting model, formation of a DMG from N would be opposed by a large enthalpy barrier, as observed, but nevertheless the DMG is stabilized by a large compensating gain in conformational entropy.

In 1995 an algorithm was developed for computing the vdW interaction energies within native proteins [32] based on the Lennard-Jones formula for the vdW interaction energy. The algorithm was tested by calculating the vdW interaction energies in the liquid and crystalline forms of linear alkanes. The attractive interaction between a pair of atoms varies as the inverse sixth power of the distance between them, and although the average change in distance between nonbonded atoms that occurs during close packing is small, this contribution approximately doubles the interaction energy. When the vdW interactions between heavy atoms (C, N, O, S) inside proteins were tabulated [32], polar-nonpolar contacts were found to be the major class; thus, close packing involves not only side chains but also side chain-backbone interactions.

Significance of DMG-Initiated Unfolding for Understanding the Rate-Limiting Step of Unfolding

In 1989 the dry molten globule was proposed to be the transition state species, $I^{\ddagger}$, a highly unstable species in 2-state unfolding [15]. Instead, the DMG turns out to be a moderately stable species that exists in measurable equilibrium with N [10,12,13], although it remains true today that DMG formation initiates the unfolding process. A main reason for expecting the DMG to be unstable upon unfolding is the large unfavorable enthalpy change, which increases with the number of buried residues. However, there is also a large, concomitant gain in conformational entropy, so the DMG can attain a stable form, like the liquid hydrocarbon produced upon melting a hydrocarbon crystal.

If the DMG proves to be a commonplace structural intermediate in protein unfolding and refolding, its existence will prompt a major reevaluation of the unfolding/refolding reaction. Evidence from the four proteins reviewed here suggests the generality of a fast initial equilibrium between N and the DMG. Unfolding then proceeds from the DMG; conversely, the last step in refolding is from the DMG to N. Written as an equation:

$$\mathrm{N}f \quad \mathrm{DMG}f \quad \mathrm{I}^{\ddagger} \rightarrow \ldots \mathrm{U} \qquad [1]$$

There are several basic implications of this equation for understanding the nature of the rate-limiting step in unfolding. One important implication is that the measured rate constant for unfolding refers to the unfolding process that starts from a dynamic mixture of N and DMG, not simply from N, and this must be taken into account when using unfolding rate constants to determine activation volumes, enthalpies, etc. Another basic implication is that side chain interactions in $I^{\ddagger}$ are not close packed, and consequently they are weaker and more labile than in close-packed native proteins. This realization challenges conclusions about side chain interactions drawn from mutant studies using native proteins. It also calls into question the many speculations about the nature of $I^{\ddagger}$ that are based on the erroneous assumption of a close-packed structure.

There are two basic reasons for the weaker and more labile side chain interactions found in the absence of close packing. First, the vdW interactions between neighboring nonbonded

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

atoms are weaker because the distances separating atoms are larger than in close-packed proteins. Second, the rigidity of close packing eliminates many conformations that are otherwise possible, and therefore less conformational entropy is lost on forming the interaction in the absence of close packing. This type of behavior is well illustrated by a study of overpacking mutations in apomyoglobin (apoMb) [33] in which each mutant has a larger nonpolar side chain than its wild-type counterpart. Overpacking mutations made in apoMb do not destabilize the molten globule that forms at pH 4, but they do destabilize native apoMb, which forms at pH 7 [33].

Several types of pairwise side chain interactions have been quantified in both experiments [34] and simulations [35] using an alanine peptide helix as the host. Like the DMG, these peptide helices are not close-packed [35]. The $-\Delta G°$ values for these side-chain interactions are in the range ≤ 1 kcal/mol. Various types of side chain-side chain interactions have been measured, including salt bridges (H-bonded ion pairs), H-bonds, hydrophobic interactions, and cation-pi interactions (cation = Arg^+, His^+, Lys^+; pi = Trp, Tyr, Phe). A statistical study of how side chain interactions increase the stability of peptide helices, based on literature data[36], also gives $-\Delta G°$ values in the same size range as the interactions measured directly with alanine peptide helices [34].

Close packing is present only at the beginning of unfolding or the end of refolding. These strong vdW close packing interactions are lost at the first step in unfolding and incorporated at the last step in refolding, but weaker, liquid-like vdW interactions persist during intermediate steps. Understanding this basic framework for the folding process helps to clarify several puzzles. The assumption is widely made in the literature that packing interactions are synonymous with close packing, and consequently that wet molten globules lack specific tertiary structures because they are not yet close-packed. From this point of view, it is difficult to explain why the pH 4 molten globule of apomyoglobin (apoMb) includes the A and GH helices of Mb [37], situated on opposite ends of the polypeptide chain, if there are no specific packing interactions that organize the intervening regions. Recognizing that, in fact, there are specific vdW interactions, but they are liquid-like, provides the likely answer.

The idiosyncratically shaped side protein chains pack together in the native protein with few voids the size of a carbon atom, rather like a three-dimensional jigsaw puzzle [9]. In 1987 a widely discussed model considered whether the complementary shapes and close tolerances required for such packing could determine the backbone fold [38]. In this model the backbone was held rigid, and close-packed arrangements of the side chains were tested systematically by simulation and rejected if they showed steric clash. However, when acceptable packing arrangements were determined either experimentally in a mutational study[39] or by analyzing proteins of known structure in a model study [40,41], side chain packing was found to be more flexible than expected.

The side chains are not close-packed during the folding process, and this extra flexibility may be an important component of protein “foldability” by reducing entropy loss en route to the native fold. Folding flexibility and fold stability both may be factors that exert selection pressure during the course of protein evolution.

Complementary information confirming the important role of vdW interactions in the energetics of folding has also been obtained from mutational studies on proteins. Several early studies of “large-to-small” mutations, e.g., Val to Ala, were made with the hope that the mutation would introduce a small cavity but leave the protein structure unaffected otherwise. This experimental design was intended to allow measurement of the $\Delta\Delta G$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

resulting from deletion of an individual buried hydrophobic group, like -CH_3, providing a way to tie data on protein stability together with model compound studies.

Deletion of a buried -CH_3 was expected to produce an unfavorable entropy change because the hydrophobic interaction is entropy-driven at 25° C. However, when the experiments were analyzed by calorimetry, the entropy change for deleting a buried -CH_3 group turned out to be favorable for folding; evidently cavity creation resulted in increased side chain conformational entropy. As anticipated, an overall decrease in protein stability was observed, but it was caused instead by an unfavorable enthalpy change, evidently caused by a loss of vdW interactions. This same conclusion was reached in two separate studies, one on RNase S mutants using titration calorimetry [42] and another on ubiquitin mutants using differential scanning calorimetry [43]. A different approach to understanding how vdW interactions influence protein stability made use of a protocol that involved surface mutations in staphylococcal nuclease, selection of thermostable mutants, and analysis of their X-ray structures [44]. Using the Voronoi volume of the major hydrophobic core as an index of packing, a clear correlation was found between thermostability and good packing, this despite the fact that no mutations were made in the core.

Solvent Access to the Peptide Backbone in a DMG

In the early model [15], it was plausible that water would enter a DMG readily, destabilizing the structure and leading directly to the unfolding transition state, $I^{\ddagger}$. A common early proposal about the nature of $I^{\ddagger}$ [16,17] was that in the transition state for unfolding, water reaches the peptide backbone and destabilizes the H-bonded secondary structure. Even much earlier this proposal seemed plausible to pioneers like Kauzmann and Tanford. This model is still current, although today it has been largely superseded by an energy landscape model, in which the rate-limiting step in unfolding involves finding a low-energy gap among the hills of the energy landscape as the polypeptide chain undergoes major conformational change.

The HX study of RNase A [16] shows, however, that the DMG is stable to exchange, which occurs only after the rate-limiting step during unfolding. Whether or not exchange occurs in $I^{\ddagger}$ unfortunately is not known because, for HX to be a measurable reaction, it must occur within a populated species such as U. However, the detailed findings of the initial HX study [16] did provide a definitive answer to the question of whether exchange occurs in the DMG. The findings also showed that exchange occurs in an all-or-none fashion: EX1 exchange rates (which measure the unfolding rate constant) were measured at pH 8.0 for 42 of the 49 slow-exchanging backbone NH protons, and all 42 protons gave the same exchange rate constant, within error (see fig. 1). Recently, a similar result was found for the small SH3 domain of P13 kinase where 14 of the 19 slow-exchanging backbone protons undergo exchange only after the rate-limiting step in unfolding [45].

In seeking to understand the mechanism of protein unfolding it is important to understand why in these two examples – RNase A and the SH3 domain of P13 – the entire peptide backbone is highly protected until the rate-limiting step in unfolding occurs. A corollary question is why the DMG of RNase A has a highly protected peptide backbone.

DMG Intermediates and the Interpretation of Phi-Values

Close packing is absent in $I^{\ddagger}$, and specifically, strong, close-packed vdW interactions between side chains are absent. This statement pertains to both unfolding and refolding because $I^{\ddagger}$ is the same in both directions under the same conditions. The absence of close packing in $I^{\ddagger}$ provides a missing link in the interpretation of ϕ-values.

The ϕ-value is defined as $[\Delta\Delta G(I^{\ddagger}\text{-}U)]/[\Delta\Delta G(N\text{-}U)]$, where $\Delta\Delta G$ is the change in folding free energy produced by a mutation in either the equilibrium folding reaction (N-U) or the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

kinetic refolding reaction ($I^{‡}$-U) [2,46]. The refolding rate constant is used to find the apparent ΔG for the kinetic refolding reaction.

The ϕ-value method of characterizing $I^{‡}$ is based on the proposal [2,46] that strong interactions between specific side chains are likely to be critical in stabilizing $I^{‡}$, and those key side chains can then be identified by their high ϕ-values (~1.0). A ϕ-value of 1.0 is interpreted to mean that the wild-type side chain stabilizes $I^{‡}$ to the same extent that it stabilizes N. Conversely, residues that play only minor roles in the stability of $I^{‡}$ during refolding would have low (~0) ϕ-values. In practice, the determination of ϕ-values involves measuring changes in protein stability, and an important caveat is that such changes must be large enough for reliable analysis [47].

Two classes of $I^{‡}$, diffuse and polarized, have been found from ϕ-value studies [47,48]. A recent survey of the literature emphasizes the common occurrence of diffuse $I^{‡}$ [49], in which the ϕ-values of residues throughout the protein are fairly uniform and have intermediate ϕ-values (ϕ ~ 0.3). A plausible explanation for intermediate ϕ-values in diffuse transition states is that almost all of the polypeptide chain participates in forming $I^{‡}$, resulting in widespread albeit modest stabilization. The fact that hydrophobic interactions in $I^{‡}$ are nearly as strong as those in N provides a reasonable explanation for such widespread stabilization. Measurement of $\Delta Cp^{‡}$, taken to be diagnostic of the hydrophobic interaction, is discussed in the following section.

An important insight into the nature of ϕ-values was found by making an entire set of mutations at a single site [50-52] and then displaying the data as a Brønsted-Leffler plot [53,54] with ΔΔG($I^{‡}$-U) plotted against ΔΔG(N-U). These plots are straight lines for which the slope, ϕ(β), is a property of the structural site and not of the wild-type residue, i.e., any of the mutants could be used as the reference without changing the value of ϕ(β). For example, when ϕ(β) = 0.4, any mutant that fits the line is stabilized 40% as much in $I^{‡}$ as it is in N.

In polarized $I^{‡}$ the ϕ-values are high (often ϕ ~ 0.8) in residues within one region of the native protein but low (ϕ ~ 0.1) elsewhere. The Fyn SH3 domain has a polarized $I^{‡}$, and the 4 ϕ(β) values that have been determined vary strikingly among the 4 sites, from ϕ(β) = 0.08 at site 41 to 0.86 at site 40 [50,51]. This surprising variation suggests that ϕ(β) values can provide a powerful tool for dissecting the structure of $I^{‡}$. The fact that ϕ-values determined in this way are a general property of the structural site, not of any specific side chain interactions made by the wild type residue, is consistent with the argument presented here that strong side chain interactions in $I^{‡}$ are unlikely in the absence of close packing.

Characterization of $I^{‡}$ from $\Delta Cp^{‡}$ and $\Delta H^{‡}$

Both $\Delta H^{‡}$ for unfolding (N-$I^{‡}$) and $\Delta Cp^{‡}$ for refolding ($I^{‡}$-U) have surprisingly large values which provide important information about the nature of $I^{‡}$. $\Delta H^{‡}$ and $\Delta Cp^{‡}$ values are determined by measuring the 2-state unfolding and refolding rate constants as a function of temperature, and plotting them by the kinetic equivalent of a van't Hoff plot, ln K versus 1/T, where K is the equilibrium constant and T is absolute temperature. In typical chemical reactions at equilibrium, ΔH is nearly independent of temperature, and the plot is a straight line with slope $-1/R\Delta H$ (where R is the gas constant). However, for protein folding reactions the plot is significantly curved [26,27] because for hydrophobic interactions ΔH depends strongly on temperature, while ΔCp, which determines the curvature, is nearly independent of temperature.

Two remarkable results were found in 1984 in a pioneering study of the temperature dependence of ln k versus 1/T for the unfolding and refolding kinetics of hen lysozyme [55]. First, the curvature in this plot appears entirely in refolding, and the plot could be fitted with

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$\Delta Cp^{\ddagger}$ equal to the equilibrium value of ΔCp. Second, the unfolding plot is a straight line with a huge value for $\Delta H^{\ddagger}$, 50 kcal/mol. Protein chemists were mystified by these results. What kind of interaction could give such a large and temperature-independent enthalpy change in the unfolding kinetics? The interactions could not be hydrophobic interactions because of the very small value of $\Delta Cp^{\ddagger}$. Moreover, the interactions are broken early in unfolding when N is only slightly exposed to water according to the small m-value that is found from the slope of ln k versus [GdmCl]. How could nearly all of the native hydrophobic interactions be formed in $I^{\ddagger}$, as implied by fitting the refolding plot with the equilibrium value for ΔCp? These 1984 results gave the first clear indication that forming a dry molten globule initiates unfolding, although this interpretation came later.

Further work was needed on two problems to make these results convincing today. First, the initial results were obtained with hen lysozyme, a medium-size protein (129 residues) that shows refolding intermediates in most conditions, but not in the pH 2.6 conditions of the 1984 study [55], which have not been studied further. Perhaps the unexpected features of the unfolding and refolding plots are caused by undetected intermediates. Second, unfolding and refolding rate constants depend on solvent viscosity [56]. Because viscosity is not a thermodynamic property, the rate constants must be corrected to obtain valid values of $\Delta Cp^{\ddagger}$ and $\Delta H^{\ddagger}$. Viscosity dependence of reaction rates is found also in some ordinary chemical reactions such as proton transfer reactions in aqueous solution. Typically a viscosity-dependent reaction rate means that a diffusional process is rate-limiting. In 2002 [56] a study of the unfolding and refolding kinetics of cold shock protein B (Csp B), from both mesophilic (*B. subtilis*) and thermophilic organisms, solved the problem of viscosity-dependent reaction rates and gave results for a protein with 2-state folding kinetics over a wide range of conditions [57]. Plots of ln k versus 1/T have the same two basic features found in the earlier hen lysozyme study [55]. Tests for a DMG intermediate have not yet been made in the Csp B system. Correcting the kinetic results for the viscosity effect does not change the value of $\Delta Cp^{\ddagger}$ and has only a small effect on $\Delta H^{\ddagger}$ [56]. The ratio $\Delta Cp^{\ddagger}/\Delta Cp$ (refolding/equilibrium) is 0.64 for Csp B from the thermophile and 0.84 from the mesophile [56].

The proposal that the ratio $\Delta Cp^{\ddagger}/\Delta Cp$ reliably measures the hydrophobic interactions in $I^{\ddagger}$ relative to N rests on two points. First, when the hydrophobic interaction is measured in the liquid hydrocarbon system, ΔG is proportional to ΔCp [24,58], and therefore ΔCp may be used to quantify ΔG. Second, the hydrophobic interaction is the major contributor to ΔCp in protein folding reactions, and it may well be the only major factor. A second factor, the formation of peptide H-bonds, has been suggested to be important [59], but this proposal is based on model compound data, and today group additivity is known to be invalid when applied to the interaction between water and the peptide group [60]. Moreover, direct measurements of ΔCp for the formation of peptide H-bonds in peptide helices have found that it is too small to measure by titration calorimetry [61,62].

DMG Intermediates and Activation Volumes for Unfolding and Refolding

Clear evidence for the expansion expected when a DMG is formed from N in unfolding comes from the study of activation volumes for unfolding and refolding. These volumes are found by measuring the unfolding and refolding rate constants in pressure-induced unfolding and plotting ln k versus pressure. The results give the volume difference between N and $I^{\ddagger}$ in unfolding experiments and between $I^{\ddagger}$ and U in refolding experiments. The measured volumes of N, U and $I^{\ddagger}$ are thermodynamic (partial molar) volumes, and they are strongly influenced by the packing of water molecules around the protein. N has a substantially larger volume than U chiefly because of the void volume (small, sub-atomic size cavities) in the close-packed structure of N, and also because water molecules pack well around the unfolded protein, much as they pack well around hydrocarbon molecules dissolved in water. Initially the volume of $I^{\ddagger}$ was anticipated to be substantially smaller than N because $I^{\ddagger}$ was

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

expected to resemble a wet molten globule, with dissipation of the native state voids and efficient water packing around newly exposed hydrophobic groups.

The surprising result found in 1995 for staphylococcal nuclease (SNase) [63] was that the volume of $I^{\ddagger}$ is much closer to N than to U. The authors interpreted this result to mean that $I^{\ddagger}$ resembles a dry molten globule. More recently, the activation volumes of two proteins were measured for which $I^{\ddagger}$ has an even larger volume than N, tendamistat [64] and a variant of notch ankyrin [65]. These results lend strong support to the proposals that $I^{\ddagger}$ has a DMG-like structure in typical proteins[63] and that an expansion occurs as the DMG is formed from N [15].

Mutants of SNase are known in which either basic or acidic residues (Lys, Asp, Glu) are buried inside the protein [66], and their activation volumes have been measured [67,68]. For these mutants the volume of $I^{\ddagger}$ is closer to U than N, while the volumes of U and N are practically unaffected by the mutation. This result suggests there is some water inside $I^{\ddagger}$ for these mutants [67,68]. Assuming, as proposed here, that the DMG and $I^{\ddagger}$ are formed in separate steps of unfolding, it will be important to study water access to the peptide backbone of the DMG in such intermediates, for example with the HX methods used to study RNase A [16,17] and the SH3 domain of P13 kinase [45].

Role of Secondary Structure in Causing Compaction during Refolding

We conclude above that $I^{\ddagger}$ is well structured in refolding and that much of the structure must be H-bonded secondary structure. In this context, it is important to emphasize recent work showing that the formation of secondary structure can account for a large part of the compaction that occurs during folding. Numerous examples of partly folded protein structures (PS = partial structure) are available in the literature. Often they are found in conditions, such as acidic pH, where the native tertiary structure is no longer stable. A survey of 41 PSs in 2002 [69] compared their inverse hydrodynamic volumes (measured by size-exclusion chromatography (SEC)) with the amounts of secondary structure measured by CD at 222 nm, the optimal wavelength for α-helices. A surprisingly good correlation was found between $[CD_{PS}/CD_{U}]$ and $[V_{U}/V_{PS}]$; the best-fit line is close to a direct proportionality with slope 1.00 and intercept 0.0. The authors conclude that formation of secondary structure has a much larger role than commonly realized in causing compaction during folding.

Recently, experiments with a single partially folded protein have shown directly that forming secondary structure can cause compaction during folding [70]. The protein is a deletion variant of notch ankyrin, a repeat protein in which each repeat unit consists chiefly of an α-helix, a β-hairpin and a β-turn. These are locally formed, hydrogen-bonded, backbone structures. When thermally unfolded at 55° C, the protein has substantial secondary structure, and the amount can be varied by adding either trimethylamine-*N*-oxide (TMAO) or sarcosine (stabilizing osmolytes) or urea (a denaturing osmolyte). Compaction was measured by the hydrodynamic volume, determined by SEC, and secondary structure was measured by CD at 228 nm. When the data are plotted in the same manner as in [69] (i.e., inverse relative hydrodynamic volume versus CD measured at 228 nm), the plot shows a direct proportionality between the change in hydrodynamic volume and the change in secondary structure. When the results are compared with a line drawn between values for U (the thermally unfolded protein at 55° C) and N (native ankyrin), the results fall on this line within error. This finding indicates that the compaction of ankyrin during folding is driven by secondary structure formation. Note that addition of urea, TMAO, or sarcosine does not affect the strength of the hydrophobic interactions appreciably, either in these experiments [70] or in general [71,72]. Instead, these osmolytes act by either strengthening (TMAO and sarcosine) or weakening (urea) the H-bonded backbone structure [71,72].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Similar observations of compaction accompanying the formation of secondary structure have been made in a study of the pressure dependence of folding for this same notch ankyrin variant [65]. In this case, compaction was measured by small-angle x-ray scattering and secondary structure was measured by FTIR spectroscopy. Both pressure and urea were used to promote partial unfolding. These further observations illustrate the generality of the relation between compaction and secondary structure.

Tanford's classic early work [73] showed that residual secondary structure is likely to remain after thermal unfolding, and he emphasized that 6 M GdmCl is needed to ensure complete unfolding. The assumption has often been made that urea, like GdmCl, also produces complete unfolding. However, it is now known that urea unfolds proteins by a different mechanism [74], and a recent paper emphasizes that extensive secondary structure may persist [75] following urea-induced unfolding. Provocatively, the extent of the residual secondary structure is correlated with both hydrophobicity and net charge per residue, the same variables that were used initially to identify intrinsically disordered proteins [76].

Taken together, the current papers show that the presence of residual secondary structure is by no means limited to thermal unfolding. Rather, it is a general phenomenon, observed as well when unfolding is induced by pressure or urea.

Secondary Structure as the Platform for Hydrophobic Interactions

During refolding, the presence of pervasive hydrophobic interactions at $I^{\ddagger}$, before close packing occurs, strongly suggests that extensive secondary structure provides the supporting platform [77]. Hydrogen-bonded backbone structures – α-helices, β-hairpins and β-turns – can form fast and in isolation. In a helix, the side chains are forced to protrude from the backbone cylinder, such that they are hyper-exposed in comparison to the coil state [78,79]. Yet despite destabilizing hyper-exposure of apolar groups, peptide helices can persist in isolation [34], indicating that backbone interactions must be sufficient to maintain the structure [80]. This same reasoning holds for β-hairpins and hydrogen-bonded reverse turns. In a protein, these individual scaffold elements can engage in mutually stabilizing interactions by burying hydrophobic groups between two or more hydrogen-bonded backbone segments, thereby converting unfavorable hyper-exposure to favorable displacement of solvent from their respective hydrophobic surfaces. In a survey of 112 non-redundant, ultra-high resolution (≤ 1Å) proteins, approximately two-thirds of the α-helices or β-strands bury at least 5% of their available sidechain surface area in a pairwise interaction with the next consecutive α-helix or β-strand in the sequence (fig. 3). Such interactions are not arrested at the pairwise level but continue iteratively, resulting in supersecondary structure and, in stepwise fashion, to ever larger hierarchically organized aggregates [81] with ever- increasing hydrophobic burial.

^{19}F-NMR Studies of Refolding

There is a basic problem in reconciling the villin headpiece results [13] with earlier ^{19}F-NMR studies of refolding. The ^{19}F-NMR studies began in 1996 [82] and were reviewed several years later [83]. They suggest that close packing may be observed by ^{19}F-NMR as a final slow step in the refolding reactions of proteins such as DHFR [82] (159 residues) and other proteins[83]. Native-like resonance lines, as judged by chemical shift, appear early in refolding and then gradually become as narrow as those of the native protein. The slow line narrowing is fairly uniform throughout the protein, suggesting cooperative close packing. Importantly, the rate of side chain stabilization is independent of side chain location in the structure, providing a strong argument for the collapse of the DMG to the native structure during refolding. Similar results have also been obtained for other proteins studied by ^{19}F-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NMR [83,84], and in some cases, slow line narrowing is also detected by conventional folding probes such as Trp fluorescence.

The basic problem in connecting these ^{19}F-NMR studies with close packing in the villin headpiece [13] is that close packing in HP35 is an extremely fast reaction (~ 1 μsec). The other DMG intermediates formed at the start of unfolding [10-12] are also formed rapidly, within the mixing dead time. One possible explanation for the slow line-narrowing observed by ^{19}F-NMR [82,83] involves proline isomerization. Examples are known in which a wrong proline isomer within a partly folded protein is slowly converted to the native isomer after folding is nearly complete [85]. It is plausible that slow proline isomerization would be reflected in narrowing of the ^{19}F resonance lines.

However, a different, more interesting explanation [82,83] is that regain of close packing can become a slow final step in the folding of larger proteins. Currently, villin headpiece is the most informative system for understanding the kinetics of close packing, but HP35 may be anomalous. The protein has only 35 residues and its three helices resemble an isolated segment of super-secondary structure. Furthermore, helix three, which is highly unstable [86], is anchored to the remaining two helices by a single hydrophobic residue and can detach readily to form a two helix + coil intermediate [86] (fig. 2). HP35 refolds in microseconds [13], rivaling its close-packing rate, and the authors suggest that the rate of close packing may even become rate-limiting in such a small system. In contrast, close packing in larger systems may involve extensive intra-molecular interfaces between different regions, causing close packing to occur cooperatively throughout the protein. If so, coordinating close packing among different segments may slow the overall process in a major way.

Concluding Comments and Questions

It seems likely that formation of a dry molten globule at the start of unfolding will come to be realized as the standard mechanism of protein unfolding, at least in one class of proteins. Four protein examples are known now, ranging in size from villin headpiece (35 residues) to dihydrofolate reductase (159 residues), and no case has yet been reported of testing for this unfolding mechanism and with a negative result.

One might think that DMG intermediates would be conspicuous because their formation involves a major structural change. Yet these structures have gone largely unnoticed, probably because they are not detected by standard probes such as Trp fluorescence, which responds to hydration of the protein interior. Both DMG intermediates and the native state have dry interiors, which are indistinguishable using Trp fluorescence.

There are major consequences for analyzing the mechanism of protein folding if the dry molten globule turns out to be the canonical mechanism of unfolding. Properties that have been attributed previously to the folding transition state are likely to be strongly influenced by a rapid equilibrium between the native protein and a dry molten globule formed at the start of unfolding. Examples discussed here include ϕ-values derived from Brønsted-Leffler plots and activation volumes derived from pressure studies.

It is critical now to find more examples of proteins that form DMG intermediates in unfolding and to investigate their properties. Two larger proteins that are being studied are: a heat-shock protein, Hsp 15, 133 residues (T. Kiefhaber, personal communication) and murine adenosine deaminase, 351 residues, (C. Frieden, personal communication). Both proteins develop unfolding intermediates before the major unfolding barrier, and detailed analysis is currently underway to learn if the intermediates have other characteristics expected of DMGs.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Many further questions invite study. What are the necessary and sufficient conditions for close packing to take place? How dry is the interior of a dry molten globule? Are there mutations – like the introduction of Glu or Lys into the protein core – that can render the dry molten globule water-accessible? Is an intact, or near-intact, hydrogen-bonded peptide backbone a pre-condition for forming a dry molten globule? Can conditions be found that stabilize the dry molten globule for analysis? Such questions are now experimentally accessible, and we suspect that the answers will be transformative.

Acknowledgments

We gratefully acknowledge the discussion of Wayne Bolen, Thomas Kiefhaber, Cathy Royer and Jayant Udgaonkar. Supported in part by NIH grant R01-DK13332 (to CF) and the Mathers Foundation (to GDR).

References

1. Greene RF Jr. Pace CN. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. J Biol Chem. 1974; 249:5388–5393. [PubMed: 4416801]
2. Matouschek A, Kellis J Jr, Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. Nature. 1989; 340:122–126. [PubMed: 2739734]
3. Matthews CR. Effect of point mutations on the folding of globular proteins. Methods Enzymol. 1987; 154:498–511. [PubMed: 3431461]
4. Tanford C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. Adv Protein Chem. 1970; 24:1–95. [PubMed: 4912353]
5. Go N. Theoretical studies of protein folding. Annu Rev Biophys Bioeng. 1983; 12:183–210. [PubMed: 6347038]
6. Levinthal, C. How to fold graciously. In: Debrunner, P.; Tsibris, JCM.; Münck, E., editors. Mössbauer Spectroscopy in Biological Sytems. Univ. of Illinois Press; Urbana: 1969. p. 22-24.
7. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins. 1995; 21:167–195. [PubMed: 7784423]
8. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol. 1997; 4:10–19. [PubMed: 8989315]
9. Richards FM. Areas, volumes, packing, and protein structure. Ann Rev Biophys Bioeng. 1977; 6:151–176. [PubMed: 326146]
10. Kiefhaber T, Labhardt AM, Baldwin RL. Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. Nature. 1995; 375:513–515. [PubMed: 7777063]
11. Hoeltzli SD, Frieden C. Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-19F-tryptophan-labeled Escherichia coli dihydrofolate reductase. Proc Natl Acad Sci U S A. 1995; 92:9318–9322. [PubMed: 7568125]
12. Jha SK, Udgaonkar JB. Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. Proc Natl Acad Sci U S A. 2009; 106:12289–12294. [PubMed: 19617531]
13. Reiner A, Henklein P, Kiefhaber T. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. Proc Natl Acad Sci U S A. 2010; 107:4955–4960. [PubMed: 20194774]
14. Hua L, Zhou R, Thirumalai D, Berne BJ. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. Proc Natl Acad Sci U S A. 2008; 105:16928–16933. [PubMed: 18957546]
15. Shakhnovich EI, Finkelstein AV. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. Biopolymers. 1989; 28:1667–1680. [PubMed: 2597723]
16. Kiefhaber T, Baldwin RL. Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. Proc Natl Acad Sci U S A. 1995; 92:2657–2661. [PubMed: 7708700]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

17. Juneja J, Udgaonkar JB. Characterization of the unfolding of ribonuclease a by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding. Biochemistry. 2002; 41:2641–2654. [PubMed: 11851411]
18. Kim PS, Baldwin RL. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem. 1982; 51:459–489. [PubMed: 6287919]
19. Brewer SH, Vu DM, Tang Y, Li Y, Franzen S, Raleigh DP, Dyer RB. Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. Proc Natl Acad Sci U S A. 2005; 102:16662–16667. [PubMed: 16269546]
20. Kubelka J, Henry ER, Cellmer T, Hofrichter J, Eaton WA. Chemical, physical, and theoretical kinetics of an ultrafast folding protein. Proc Natl Acad Sci U S A. 2008; 105:18655–18662. [PubMed: 19033473]
21. Klapper MH. On the nature of the protein interior. Biochim Biophys Acta. 1971; 229:557–566. [PubMed: 5555208]
22. Privalov PL, Gill SJ. Stability of protein structure and hydrophobic interaction. Adv Protein Chem. 1988; 39:191–234. [PubMed: 3072868]
23. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem. 1959; 14:1–63. [PubMed: 14404936]
24. Baldwin RL. Temperature dependence of the hydrophobic interaction in protein folding. Proc Natl Acad Sci USA. 1986; 83:8069–8072. [PubMed: 3464944]
25. Tanford, C. The Hydrophobic Effect. 2nd edition. John Wiley & Sons; New York: 1980.
26. Dill KA. Dominant forces in protein folding. Biochemistry. 1990; 29:7133–7155. [PubMed: 2207096]
27. Baldwin RL. Energetics of protein folding. J Mol Biol. 2007; 371:283–301. [PubMed: 17582437]
28. Sharp KA, Nicholls A, Friedman R, Honig B. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theorectical models. Biochemistry. 1991; 30:9686–9697. [PubMed: 1911756]
29. Lee B. Solvent reorganization contribution to the transfer thermodynamics of small nonpolar molecules. Biopolymers. 1991; 31:993–1008. [PubMed: 1782360]
30. Ben-Naim A, Marcus Y. Solvation thermodynamics of nonionic solutes. J Chem Phys. 1984; 81:2016–2027.
31. Nicholls A, Sharp KA, Honig B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Structure, Function, and Genetics. 1991; 11:281–296.
32. Lazaridis T, Archontis G, Karplus M. The enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. Advances in Protein Chemistry. 1995; 47:231–306. [PubMed: 8561050]
33. Bertagna AM, Barrick D. Nonspecific hydrophobic interactions stabilize an equilibrium intermediate of apomyoglobin at a key position within the AGH region. Proc Natl Acad Sci U S A. 2004; 101:12514–12519. [PubMed: 15314218]
34. Chakrabartty A, Baldwin RL. Stability of α-helices. Adv Protein Chem. 1995; 46:141–176. [PubMed: 7771317]
35. Creamer TP, Rose GD. Interactions between hydrophobic side chains within alpha-helices. Protein Sci. 1995; 4:1305–1314. [PubMed: 7670373]
36. Lacroix E, Viguera AR, Serrano L. Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. J Mol Biol. 1998; 284:173–191. [PubMed: 9811549]
37. Hughson FM, Wright PE, Baldwin RL. Structural characterization of a partly folded apomyoglobin intermediate. Science. 1990; 249:1544–1548. [PubMed: 2218495]
38. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol. 1987; 193:775–791. [PubMed: 2441069]
39. Lim WA, Sauer RT. Alternative packing arrangements in the hydrophobic core of lambda repressor. Nature. 1989; 339:31–36. [PubMed: 2524006]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

40. Behe MJ, Lattman EE, Rose GD. The protein-folding problem: the native fold determines packing, but does packing determine the native fold? Proc Natl Acad Sci U S A. 1991; 88:4195–4199. [PubMed: 2034665]
41. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A. 2000; 97:10383–10388. [PubMed: 10984534]
42. Ratnaparkhi GS, Varadarajan R. Thermodynamic and structural studies of cavity formation in proteins suggest that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics. Biochemistry. 2000; 39:12365–12374. [PubMed: 11015216]
43. Loladze VV, Ermolenko DN, Makhatadze GI. Thermodynamic consequences of burial of polar and non-polar amino acid residues in the protein interior. J Mol Biol. 2002; 320:343–357. [PubMed: 12079391]
44. Chen J, Lu Z, Sakon J, Stites WE. Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. J Mol Biol. 2000; 303:125–130. [PubMed: 11023780]
45. Wani AH, Udgaonkar JB. Native state dynamics drive the unfolding of the SH3 domain of PI3 kinase at high denaturant concentration. Proc Natl Acad Sci U S A. 2009
46. Serrano L, Matouschek A, Fersht AR. The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. J Mol Biol. 1992; 224:805–818. [PubMed: 1569558]
47. Sanchez IE, Kiefhaber T. Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. J Mol Biol. 2003; 334:1077–1085. [PubMed: 14643667]
48. Raleigh DP, Plaxco KW. The protein folding transition state: what are Phi-values really telling us? Protein Pept Lett. 2005; 12:117–122. [PubMed: 15723637]
49. Naganathan AN, Munoz V. Insights into protein folding mechanisms from large scale analysis of mutational effects. Proc Natl Acad Sci U S A. 2010
50. Northey JG, Maxwell KL, Davidson AR. Protein folding kinetics beyond the phi value: using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. J Mol Biol. 2002; 320:389–402. [PubMed: 12079394]
51. Zarrine-Afsar A, Dahesh S, Davidson AR. Protein folding kinetics provides a context-independent assessment of beta-strand propensity in the Fyn SH3 domain. J Mol Biol. 2007; 373:764–774. [PubMed: 17850820]
52. Mok YK, Elisseeva EL, Davidson AR, Forman-Kay JD. Dramatic stabilization of an SH3 domain by a single substitution: roles of the folded and unfolded states. J Mol Biol. 2001; 307:913–928. [PubMed: 11273710]
53. Anslyn, EV.; Dougherty, DA. Modern Physical Organic Chemistry. University Science Books; Sausalito, CA: 2006.
54. Leffler JE. Parameters for the Description of Transition States. Science. 1953; 117:340–341. [PubMed: 17741025]
55. Segawa S, Sugihara M. Characterization of the transition state of lysozyme unfolding. I. Effect of protein-solvent interactions on the transition state. Biopolymers. 1984; 23:2473–2488. [PubMed: 6518262]
56. Perl D, Jacob M, Bano M, Stupak M, Antalik M, Schmid FX. Thermodynamics of a diffusional protein folding reaction. Biophys Chem. 2002; 96:173–190. [PubMed: 12034439]
57. Schindler T, Schmid FX. Thermodynamic properties of an extremely rapid protein folding reaction. Biochemistry. 1996; 35:16833–16842. [PubMed: 8988022]
58. Spolar RS, Ha JH, Record MT Jr. Hydrophobic effect in protein folding and other noncovalent processes involving proteins. Proc Natl Acad Sci U S A. 1989; 86:8382–8385. [PubMed: 2813394]
59. Privalov PL, Makhatadze GI. Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding. J Mol Biol. 1992; 224:715–723. [PubMed: 1314903]
60. Avbelj F, Baldwin RL. Origin of the change in solvation enthalpy of the peptide group when neighboring peptide groups are added. Proc Natl Acad Sci U S A. 2009; 106:3137–3141. [PubMed: 19202077]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

61. Lopez MM, Chin DH, Baldwin RL, Makhatadze GI. The enthalpy of the alanine peptide helix measured by isothermal titration calorimetry using metal-binding to induce helix formation. Proc Natl Acad Sci U S A. 2002; 99:1298–1302. [PubMed: 11818561]
62. Goch G, Maciejczyk M, Oleszczuk M, Stachowiak D, Malicka J, Bierzynski A. Experimental investigation of initial steps of helix propagation in model peptides. Biochemistry. 2003; 42:6840–6847. [PubMed: 12779338]
63. Vidugiris GJ, Markley JL, Royer CA. Evidence for a molten globule-like transition state in protein folding from determination of activation volumes. Biochemistry. 1995; 34:4909–4912. [PubMed: 7711012]
64. Pappenberger G, Saudan C, Becker M, Merbach AE, Kiefhaber T. Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements. Proc Natl Acad Sci U S A. 2000; 97:17–22. [PubMed: 10618363]
65. Rouget J-B, Aksel T, Barrick D, Royer CA. Unique features of the folding landscape of a repeat protein revealed by pressure perturbation. Biophys J. 2010
66. Isom DG, Cannon BR, Castaneda CA, Robinson A, Garcia-Moreno B. High tolerance for ionizable residues in the hydrophobic interior of proteins. Proc Natl Acad Sci U S A. 2008; 105:17784–17788. [PubMed: 19004768]
67. Brun L, Isom DG, Velu P, Garcia-Moreno B, Royer CA. Hydration of the folding transition state ensemble of a protein. Biochemistry. 2006; 45:3473–3480. [PubMed: 16533028]
68. Mitra L, Hata K, Kono R, Maeno A, Isom D, Rouget J-B, Winter R, Akasaka K, Royer CA. V(i)-value analysis: a pressure-based method for mapping the folding transition state ensemble of proteins. J Am Chem Soc. 2007; 129:14108–14109. [PubMed: 17960902]
69. Uversky VN, Fink AL. The chicken-egg scenario of protein folding revisited. FEBS Lett. 2002; 515:79–83. [PubMed: 11943199]
70. Holthauzen LM, Rosgen J, Bolen DW. Hydrogen bonding drives contraction of the protein urea-denatured state upon transfer to water and poorer solvents. Biochemistry. 2010; 49:1310–1318. [PubMed: 20073511]
71. Auton M, Bolen DW. Predicting the energetics of osmolyte-induced protein folding/unfolding. Proc Natl Acad Sci U S A. 2005; 102:15065–15068. [PubMed: 16214887]
72. Auton M, Holthauzen LM, Bolen DW. Anatomy of energetic changes accompanying urea-induced protein denaturation. Proc Natl Acad Sci U S A. 2007; 104:15317–15322. [PubMed: 17878304]
73. Tanford C. Protein denaturation. Adv Protein Chem. 1968; 23:121–282. [PubMed: 4882248]
74. Lim WK, Rosgen J, Englander SW. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. Proc Natl Acad Sci U S A. 2009; 106:2595–2600. [PubMed: 19196963]
75. Pace CN, Fu H, Takano K, Scholtz JM, Grimsley GR. Urea denatured state ensembles contain extensive secondary structure that is increased in hydrophobic proteins. Prot Science. 2010; 19:929–943.
76. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? Proteins. 2000; 41:415–427. [PubMed: 11025552]
77. Rose GD, Fleming PJ, Banavar JR, Maritan A. A backbone-based theory of protein folding. Proc Natl Acad Sci U S A. 2006; 103:16623–16633. [PubMed: 17075053]
78. Chothia C. The nature of the accessible and buried surfaces in proteins. J Mol Biol. 1976; 105:1–14. [PubMed: 994183]
79. Creamer TP, Srinivasan R, Rose GD. Modeling unfolded states of peptides and proteins. Biochemistry. 1995; 34:16245–16250. [PubMed: 8845348]
80. Bolen DW, Rose GD. Structure and energetics of the hydrogen-bonded backbone in protein folding. Annu Rev Biochem. 2008; 77:339–362. [PubMed: 18518824]
81. Baldwin RL, Rose GD. Is protein folding hierarchic? I. Local structure and peptide folding. Trends Biochem Sci. 1999; 24:26–33. [PubMed: 10087919]
82. Hoeltzli SD, Frieden C. Real-time refolding studies of 6-19F-tryptophan labeled Escherichia coli dihydrofolate reductase using stopped-flow NMR spectroscopy. Biochemistry. 1996; 35:16843–16851. [PubMed: 8988023]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

83. Frieden C. The kinetics of side chain stabilization during protein folding. Biochemistry. 2003; 42:12439–12446. [PubMed: 14580188]

84. Li H, Frieden C. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and 19F NMR. Proc Natl Acad Sci U S A. 2007; 104:11993–11998. [PubMed: 17615232]

85. Cook KH, Schmid FX, Baldwin RL. Role of proline isomerization in folding of ribonuclease A at low temperatures. Proc Natl Acad Sci U S A. 1979; 76:6157–6161. [PubMed: 293712]

86. 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]

87. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]

88. Aurora R, Srinivasan R, Rose GD. Rules for alpha-helix termination by glycine. Science. 1994; 264:1126–1130. [PubMed: 8178170]

89. Baldwin RL, Rose GD. Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci. 1999; 24:77–83. [PubMed: 10098403]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

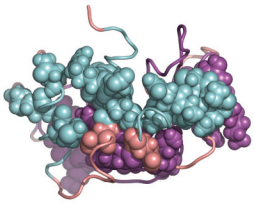

Fig. 1.

Unfolding of RNase A monitored by hydrogen exchange. Data were taken during unfolding in EX1 exchange conditions, where the HX rate constant measures the unfolding rate constant [16]. Exchange rates were determined for 42 slow-exchanging backbone NH proteins. Sidechain atoms of these 42 residues are shown as CPK spheres on the backbone ribbon of 5rsa.pdb [87]. The molecule is colored by secondary structure type (cyan = α-helix; purple = β-sheet; pink = turn/coil). It is apparent that these 42 residues span almost the entire molecule. All 42 protons were found to exchange with the same rate constant, within error, indicating that RNase A unfolds in a discrete, all-or-none step.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

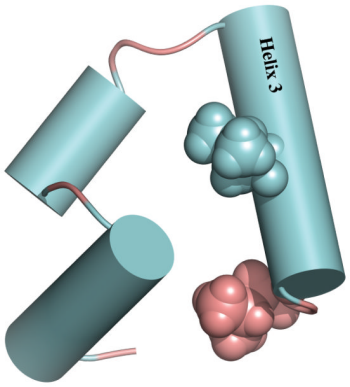

Fig. 2.

Native state ⟨ DMG equilibrium in villin headpiece. The protein, which consists of three helices, undergoes an extremely fast unlocking/relocking equilibrium (~1 μsec at 5° C) as monitored by triplet-triplet transfer [13]. The three helices of 1yrf.pdb [87] are displayed as cylinders. Helix three has only one hydrophobic residue, Leu 28 (atoms shown as CPK spheres), that interacts with the other two helices. This helix is highly unstable, and the molecule can unfold to form a helix 1 + helix 2 + coil intermediate [86], most likely via the unlocked (i.e., DMG) form. Helix three terminates in a Schellman helix-capping motif [88], where Leu 34 (CPK spheres) at the C″ position both stabilizes the helix and provides an additional hydrophobic anchor to the helix 1 + helix 2 complex.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

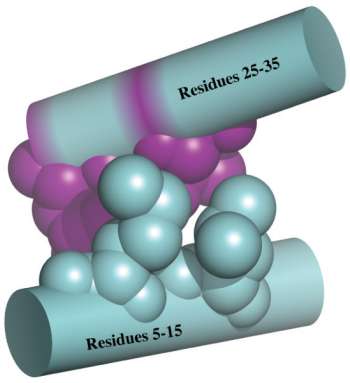

Fig. 3.

Supersecondary structure in hen lysozyme. Sidechains in an isolated helix are hyper-exposed in comparison to the coil state [35,78], but supersecondary structure formation is accompanied by substantial hydrophobic burial, as illustrated here for 8lyz.pdb [87]. In this case, two apolar residues in helix 1 (Leu 8 and Met 12, in cyan) and two apolar residues in helix 2 (Leu 25 and Val 29, in purple), engender a solvent-shielded interface. Such interactions accumulate iteratively, from supersecondary structure to ever larger hierarchically organized aggregates [81,89] with ever-increasing hydrophobic burial.