Published in final edited form as:

Arch Biochem Biophys. 2008 January 1; 469(1): 118–131. doi:10.1016/j.abb.2007.07.007.

Minimizing Frustration by Folding in an Aqueous Environment

Carla Mattos* and A. Clay Clark*

Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695

Abstract

Although life as we know it evolved in an aqueous medium, the properties of water are not completely understood. In this review, we focus on the role of water in guiding protein folding and stability. Specifically, we discuss the mechanisms of protein folding in an aqueous environment, the effects of water on the folding energy landscape as well as the transition state ensemble, and interactions of water with the folded state. We show that water cannot be viewed as a passive solvent, but rather, plays a very active role in the life of a protein.

Keywords

protein folding; funnel; energy landscape; transition state ensemble; minimal frustration; water; hydrogen bonding

1. Introduction

“If there is magic on this planet, it is contained in water.”

- Loren Eiseley [1]

It seems unlikely that life as we know it today could have evolved in a solvent other than water. As described by Pace [2], the solvent of life must meet two essential criteria in regard to protein structure and function. First, the native ensemble of conformational states must be favored thermodynamically over non-native or unfolded states. Second, the native state, and perhaps the unfolded state as well, must be soluble in that solvent. Water meets both criteria in this regard, where the native conformation is favored thermodynamically and is soluble. Based on measurements of the transfer free energies (ΔG_{tr}) of proteins from water to ethanol [3], cyclohexane [4], or a vacuum [5], proteins are predicted to be most stable in a vacuum, very stable in non-polar solvents such as cyclohexane, and unstable in polar solvents such as ethanol. This prediction is in line with the crystal structures of proteins solved in organic solvents, where regions that are disordered in aqueous solution become highly ordered in the presence of hexanediol or trifluoroethanol, but not in the presence of small alcohols [6]. The stability of proteins in different solvents or environments is due to the interplay between the contributions of the burial of non-polar side-chains *versus* the burial of peptide groups in the protein interior. These two features are significant because, on average, 83% of most non-polar side chains are

*Address correspondence to A. Clay Clark (clay_clark@ncsu.edu) or Carla Mattos (carla_mattos@ncsu.edu) at the Department of Molecular and Structural Biochemistry, 128 Polk Hall, North Carolina State University, Raleigh, NC 27695–7622. Phone: (919) 515-5805 (ACC), 919-513-2556 (CM). Fax: (919) 515-2047.

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.

buried, or inaccessible to the solvent, in the native conformation, and 82% of the peptide backbone groups are buried as well [7]. In a vacuum, for example, no intermolecular hydrogen bonds or van der Waals interactions are possible, so the peptide groups prefer to be in water (ΔG_{tr} is positive). On the other hand, the unfavorable hydrophobic effect does not occur for non-polar groups in a vacuum (ΔG_{tr} is negative), so non-polar groups prefer to be in a vacuum rather than in water. Indeed, a protein will remain folded in a vacuum unless the net charge is large enough that unfavorable charge-charge interactions overcome the stabilizing interactions and cause the protein to unfold [2]. In contrast, proteins have very low solubility in non-polar solvents [2,8], so although they are more stable in nonpolar solvents *versus* water, the lower solubility severely limits their function.

This review discusses the protein water interactions that emerge from the very delicate balance between minimizing the exposure of hydrophobic groups to water and maximizing the pairing of protein polar groups in hydrogen bonding interactions. This balance has been optimized throughout evolution in such a way that water drives and at the same time facilitates the transition from the unfolded polypeptide chain to the marginally more stable folded native state. This is a very complex process that passes through a transition state, the nature of which is currently under intense investigation. An overview of this process will begin with a discussion of the properties of water and its relation to proteins (section 2) followed by a general characterization of protein folding (section 3) and a description of how water affects the energy landscape sampled by a protein in aqueous solution (section 4). The current understanding of the nature of the transition state is presented with focus on the role of water in determining the properties of the transition state energy barrier (section 5). The review continues with a characterization of protein-water interactions in the folded state that are responsible for stability, dynamics and functional properties (section 6) with particular emphasis on the role of buried water molecules in cavities in the protein core (section 7). The conclusion (section 8) summarizes the most important aspects of the review.

2. General Properties of Water and Its Interaction with Proteins

As a general solvent, water has a number of unusual properties. It has a large heat capacity, high melting and boiling points compared to other common liquids, high thermal conductivity and surface tension, and it shrinks on melting [9,10]. In addition, water has a small size (11.5 Å³ *versus* 89 Å³ for alanine), contains a dipole moment, is polarizable and is capable of forming up to four hydrogen bonds simultaneously. In practice, bulk water molecules form 3.5 hydrogen bonds on average. Importantly, water can serve as donor and acceptor of hydrogen bonds. As a result of its hydrogen bonding ability, liquid water has large internal cohesion due to the strong attractive forces with neighboring water molecules. Although all of these features have important biological consequences, we still have much to learn about this solvent. For example, there is not a satisfactory description of how a proton moves in water, it currently is not possible to calculate with a satisfactory degree of accuracy many of the properties of water listed above over large ranges of conditions, and the nature of the surfaces of ice and liquid water are not fully understood [11].

Water participates in many hydrogen bonding networks and in screening electrostatic interactions. In protein structure, for example, some hydrogen bonds can be considered as containing three centers focused on the protein polar groups, including the amide nitrogen, the carbonyl oxygen and/or the various side chain donor and acceptor groups, and the water oxygen or protons (Figure 1). These bonds lower the potential energy barrier in general dynamic processes, including protein folding, unfolding, ligand binding, and protein-protein interactions [12]. In lowering the energy barrier between two dynamic states, water acts as a lubricant in these reactions. In terms of protein folding, water catalyzes the rapid interconversion between conformational states of a given amino acid residue. In addition, the

expulsion of water from hydrophobic regions, described below, may reduce rapid fluctuations, thereby stabilizing transient secondary structure that forms during the initial collapse of the protein. Evidence for this stabilization is shown by the fact that protein structures are very rigid in the absence of water because the lower dielectric environment of the nonpolar solvent results in strong intra-protein electrostatic interactions [13]. As a result, the drive to unfold is large in hydrophobic organic solvents, but the protein lacks the conformational flexibility necessary to unfold. On the other hand, in more polar organic solvents such as dimethylformamide or dimethylsulfoxide, the protein tends to unfold due to the combined effects of somewhat increased protein flexibility and favorable interactions between solvent and hydrophobic amino acid side chains in the absence of the hydrophobic effect induced by water [14,15].

Water plays a critical role not only in protein folding, but also in maintaining the structural integrity of the native state. It defines the three dimensional structure of a protein primarily due to two factors. First, water plays a fundamental role in favoring the weak van der Waals attractions between hydrophobic residues at the core through the hydrophobic effect [15–17]. This is a result of the nonpolar groups on the protein surface promoting the interactions among the water molecules themselves [18]. Second, direct interactions between water molecules and polar groups on the protein result in a hydration shell such that hydrogen bond donors or acceptors are paired in areas where the protein packing is not optimal. In general, the water molecules adjacent to the protein surface bridge between the exposed surface residues and the bulk solvent.

High-resolution X-ray crystal structures provide a partial view of the so-called first hydration shell in proteins, which shows preferential interaction with charged and polar groups. This conventional hydrogen bond is the type thought to contribute most prominently in the transient protein folding process, both in terms of direct intra-molecular hydrogen bonding and those mediated by water. However, the conventional hydrogen bond is by no means the only type found in proteins. The folded state is stabilized by a variety of hydrogen bonding types involving water, including those that are weaker in energy but common enough to make a significant contribution. These include the oxygen-aromatic interactions, where side chain aromatic rings serve as hydrogen bond acceptors or the ring hydrogen atoms serve as donors [19,20] and the more recently observed interactions between water and the heterocyclic nitrogen atoms of histidine and tryptophan side chains [21].

The physical principles that apply to protein folding and stability of the native state also apply to protein-protein and protein-ligand interactions. Thus, issues of solvation, hydration/dehydration and specific roles played by water in mediating structural features and dynamics, so important in protein folding, also are critical for binding. The two processes are distinct, however, in that folding of globular proteins typically involves a highly cooperative process greatly enhanced by proximity due to covalent bonding, whereas binding often relies on other mechanisms, such as electrostatic steering [22]. Since the main focus of the present review is the role of water in protein folding, the reader is referred to reviews regarding the role of water in molecular assemblies involving proteins [23–25], as well as to recent database studies that reveal the very prominent role that water plays in protein-protein [26,27] and protein-ligand interactions [28].

3. Protein Folding in an Aqueous Environment

Protein folding can be characterized in general terms as the initial collapse of the extended and hydrated polypeptide chain to an overall native topology followed by structural rearrangements that form the native conformation with a densely packed protein interior [29]. Thus, globular proteins fold by minimizing the nonpolar surface area exposed to water while at the same time providing hydrogen bonding interactions for buried backbone polar groups [30]. Optimization

of these two factors leads to the high cooperativity and specificity observed in the protein folding process. Both the increase in hydrophobic exposed surface area and the breaking of intramolecular hydrogen bonds are responsible for the increased heat capacity typically observed for proteins in the unfolded relative to the folded state and contribute to the large increase in heat capacity that is the hallmark of the unfolding/folding transition and of other highly cooperative processes [31].

The role of water as the major determinant of the hydrophobic effect is vital. The unfavorable entropy decrease due to formation of pentagonal water rings or clathrates around the nonpolar groups [32,33] favor the van der Waals interactions between hydrophobic amino acid side chains such that the solvent exposed hydrophobic surface is minimized [24,34]. Thus, the discord between hydrophobic side chains and polar water is a central driving force for protein folding. Water also drives folding by the gain of translational entropy upon its release from the unfolded polypeptide chain [24].

As a result of the initial collapse driven by the hydrophobic effect, the polar backbone is sequestered from bulk solvent in the nonpolar interior of the protein, a situation that is sustainable only if the buried polar groups can participate in hydrogen bonding interactions. In fact, surveys of the structural database demonstrate that approximately 90% of buried polar groups are hydrogen bonded (see Fleming and Rose [35] and references therein). Taken together, the structural and thermodynamic databases have led to the hydrogen bonding hypothesis, which states that “all potential hydrogen bond donors and acceptors in proteins are satisfied a significant fraction of the time, either by intermolecular hydrogen bonds or by hydrogen bonds to solvent water” [35]. The formation of secondary structure usually is the means for optimizing hydrogen bonding within the protein core. Pace estimates that each hydrogen bond contributes ~1–2 kcal/mol net stabilization [36], and the enthalpic cost of breaking a hydrogen bond in the protein interior is even higher (~4–6 kcal/mol) [35]. Therefore, while the hydrophobic effect is a major driving force for protein folding [34], the additive contributions of hydrogen bonds cannot be ignored.

Packing of the protein interior is more efficient than packing in organic crystals [37]. Because proteins are heterogeneous, the different sizes and shapes of the amino acid residues allow them to pack more tightly than homogeneous objects of the same size and shape. In addition, proteins are not limited by the necessity to form a crystal lattice, so they can rearrange their component parts relative to each other in order to maximize the packing density. But, packing still is imperfect, leading to unpaired polar groups in cavities in the protein interior. The presence of water molecules fulfills the available hydrogen bonds at those sites. Therefore, as a general solvent, water not only drives the interactions of hydrophobic amino acids (the hydrophobic effect), but also promotes favorable enthalpic interactions between polar or charged residues.

Exactly when the majority of the water molecules are removed during folding is a topic of debate that has led to the proposal of two distinct models: the drying mechanism and the expulsion mechanism (Figure 2). According to the “dewetting” or drying mechanism [24,38, 39], the initiation of folding is accompanied by a decrease in water density. The subsequent collapse of the polypeptide chain stabilizes the protein by reducing the solvent-accessible area of the residues found in the core of the protein. The drying effect results from the movement of water away from sites that nucleate folding, resulting in the formation of a vapor bubble, or vacuum, around those sites. In essence, the removal of water from key nucleation sites is followed by the collapse of the extended polypeptide chain. Alternatively, in the expulsion mechanism, the initial collapse of the extended, yet hydrated, polypeptide chain is followed by desolvation of the core hydrophobic amino acids [24,40]. The initial hydrophobic collapse is accompanied by substantial dehydration where most of the water molecules are excluded

from the polar backbone upon its burial. Regardless of the mechanism of the initial collapse, there is direct experimental evidence that the hydration pattern of the molten globule state of proteins very much resembles that of the native state [41].

During the folding of a protein, the water hydrating the peptide backbone assists in assembly of secondary and supersecondary structure as the protein progresses toward the native state [42]. In addition, the cooperative formation of hydrogen bonds by the buried polar groups in the backbone results in a cooperative removal of water from the protein core [43], which is necessary for the native contacts to form. In general, water can guide folding and packing of the nascent structural elements by mediating long-range electrostatic interactions between polar groups [17]. Because of their ability to form bridging hydrogen bonds with polar and charged groups, water molecules act as a lubricant to confer on proteins the flexibility they need for rearrangements of peptide amide-carbonyl hydrogen bonds during conformational changes. Importantly, the lubrication conferred by water enables the hydrophobic core to find its optimally packed state. In this regard, water serves as a bridge between backbone hydrogen bonded residues in the core, thus playing a role in the search for the native conformation. Water also can play a role in preventing nonnative contacts from forming, consequently minimizing frustration, as described below [24].

Each step in the folding process can occur only if the solvent molecules move. Therefore, the rate of the protein movement is proportional to the Debye or dielectric relaxation rate coefficient, k_{α} , also known as the α -relaxation [44–46]. In this way, protein folding is slaved to the movement of water molecules. The so-called slaving model of protein folding is based on three concepts [47]. First, proteins assume a large number of different conformational substates. Second, proteins are organized by a hierarchical energy landscape, as described below. Third, large-scale protein conformational changes follow the α -relaxation of the bulk solvent. Regarding the last point, one observes that there is a correlation between protein motions and water dynamics [48,49]. So, the fluctuations in the hydration water can slave the protein dynamics and thus affect protein function. Interestingly, while the overall protein dynamics are correlated with those of the bulk water, the residence times of individual water molecules on protein surfaces have been determined by NMR to be in the picosecond to nanosecond time scale and even the most buried water molecules exchange in the microsecond time scale [50,51]. The exchange with bulk solvent is much faster than protein folding, which typically occurs in the millisecond to second time scale. This means that while the overall folding dynamics are slaved to the dynamic fluctuations of the solvent, no individual water molecule is trapped long enough to be rate limiting in the folding process [16,52]. The critical importance of the overall bulk properties of water in facilitating rapid exchange of protein-bound water molecules, and thus its properties as a lubricant in dynamical processes, is made clear in an NMR study of subtilisin Carlsberg in the presence of minute quantities of water in tetrahydrofuran (THF) [53]. Under the bulk properties of THF, the water molecules are observed to remain tightly bound to the protein, with no detectable exchange into solvent, which is consistent with a rigid and immobile hydration shell, at least on the NMR time scale.

Overall, the role of hydration in influencing the folding mechanism for a specific protein is complicated by the protein sequence in two ways. First, the mechanisms of hydration/dehydration that govern the pathways of initial chain collapse are sequence dependent since nucleation sites are not conserved in all proteins. Second, the mechanisms of hydration/dehydration that govern water expulsion and packing of the protein interior following the initial chain collapse depend on unresolved polar groups in the interior of the protein, as described above, which depend on the specific protein sequence.

4. Water Affects the Energy Landscape

Protein folding is not viewed as a process in which all possible conformations available to a particular amino acid sequence are sampled until the thermodynamically most stable conformation of the protein is found (shown schematically in Figure 3A) [54]. Rather, the energy landscape for protein folding is described as a multidimensional funnel in which the unfolded ensemble, representing the highest energy states that have large conformational freedom, resides at the top of the funnel (Figure 3B). The native ensemble, representing the lowest energy states with less conformational freedom, is found at the bottom of the funnel [55–58]. “Ensembles” of conformations represent closely related sets of structures that fluctuate around a local, or in the case of the native protein, global, energy minimum. The energy landscape consists of small energy barriers that are overcome by the available thermal energy [42]. An unfolded protein begins folding by making a random (Brownian) walk in the unfolded ensemble until it reaches the transition state ensemble (TSE), characterized as an ensemble of structures that have the same probability of folding or unfolding.

In this context, protein folding can be described as the progressive organization of an ensemble of partially folded structures that minimize frustration [24]. The protein folds guided by the principle of “minimal frustration” in which natural protein sequences are selected by evolution to minimize interactions that are in conflict [59]. Here, “frustration” refers to conflicting interactions in the protein that result in a competition between two or more states that minimize a local part of the free energy [24]. One likely goal in the evolutionary utilization of protein sequence space is to limit conflicting interactions that lead to alternate conformations, thus limiting kinetic traps that may occur during folding. In other words, alternate conformations may exist, but their stability is low enough to allow escape in a reasonable time period. The primary effect of the high energy states is to slow conformational diffusion, which effectively decreases the overall rate of folding [59].

Inherent in this model for protein folding are flexible, transient, interactions that may be mediated by water. Water is ideally suited to facilitate these reactions for several reasons [59,60]. First, water decreases the activation energy between local energy minima, which effectively smoothes the energy landscape. Second, water helps rescue misfolded protein from conformational traps by providing conformational flexibility to the protein. Third, as described below, water stabilizes the native conformation. These features depict water as intimately involved in the folding process.

A subtle balance occurs during protein folding regarding the removal of water. If too much water is removed during folding, that is the interior is left dry, then the protein would be frozen from making further conformational transitions. On the other hand, removal of too little water during folding would result in structural ambiguity, thereby increasing the search of conformational space for the native ensemble. In either case, folding to the native ensemble would be impaired. A series of molecular dynamics simulations where two monomers of BphC come together to form a dimer with a largely hydrophobic interface illustrates the complexity of this balance in hydrophobic collapse involving proteins, compared to the idealized paraffin-like plates typically used as a model to study desolvation [61]. The results show that electrostatic interactions between water molecules at the interface and the polar groups of the protein play a crucial role in the kinetics of the hydrophobic collapse in a way that cannot be captured by the paraffin model. Turning off the electrostatics results in much faster dewetting at the interface, which during protein folding would disturb the water balance that must exist in order for the protein to fold properly.

A full molecular understanding of the folding landscape would require complete thermodynamic and kinetic descriptions of the ensembles of structures available to a protein

sequence as well as the mechanisms of their interconversion. Such a description also should provide a detailed mechanism for the role of water molecules in facilitating transitions between ensembles, thereby guiding the folding pathways. However, it is difficult to display possible transitions between free energy minima on simplified representations of the energy landscape as well as to obtain information on the TSE. As a result, Csermely [60] and others [62,63] describe the energy landscape as a network of conformational space, where various nodes represent local energy minima (conformations), and links between the nodes correspond to transitions between the conformations. Various initial conditions may lead to folding by different trajectories, or heterogeneous routes, leading to structures that coalesce into the TSE [63]. Although these are challenging methods, potentially one can examine the free energy minima and the connectivity between them. In these models, the conformational space of a two-dimensional lattice representation of the polypeptide chain is mapped onto a network where the native ensemble is only a few conformational transitions, or nodes, apart from any other energy minimum of the landscape. Sudden jumps over an activation energy barrier lead to the population of one of the neighboring conformational states. The role of water is fundamental in these transitions. Water is thought to lower the activation energy between two local minima, resulting in a smoother energy landscape [60]. This occurs because the hydrogen bonds between water molecules and either main chain or side chain atoms fluctuate. As a result of the fluctuations, changes occur in the energy level of the protein conformation, thereby allowing for a transient decrease in the activation energy between two conformations. Recently, it was shown that the network representation also is effective in obtaining a quantitative description of the energy basins and barriers of the landscape [62].

This and other work based on simplified models are beginning to elucidate specific ways in which water is involved in guiding protein folding by minimizing frustration, that is, by linking folding pathways that are unproductive with high energy transition state barriers. As two nonpolar groups approach each other in aqueous solution, the free energy of interaction is characterized by two energy minima. One is at the optimal van der Waals contact distance and the other is at the distance where the two groups are separated by a single water molecule (Figure 4). The solvation/desolvation barrier lies between the two minima, and represents the unfavorable situation where the nonpolar groups have lost (or not yet gained) interaction with each other and with water, creating an energetically unfavorable cavity. Within the context of a protein, the nature of these local transition barriers can vary depending on the groups near the interacting atoms in such a way to favor some conformations over others (see Figure 4 for a comparison of the solvation/desolvation barriers for two approaching methane molecules *versus* alanine molecules). In this manner, solvation/desolvation barriers collectively can have a profound effect on the rate-limiting step in protein folding [64], affecting the energy landscape as to minimize frustration [65]. The effect of the solvation/desolvation barriers in protein folding is most obvious when comparing simulations with and without the explicit inclusion of the barriers in the potential energy function. Incorporation of solvation/desolvation barriers in folding simulations based on a semi-realistic off-lattice protein model for several peptides with well defined native states, results in an increase of the average energy for the non-native states, increased stability of the native state and an increase in folding cooperativity, relative to the same simulations in the absence of the desolvation barriers [65]. In all cases, the folding rate is increased when the desolvation barriers are included. It is clear that water has an explicit effect on both the thermodynamics and the kinetics of protein folding, and a mechanistic picture is emerging of how water molecules participate in decreasing energetic frustration during the protein folding process.

5. Wet or Dry Transition State Ensemble?

The initial collapse of the unfolded ensemble of high energy states is accompanied by a reduction in conformational entropy to a more compact ensemble of states. The restriction of

conformational freedom at the rate-limiting step gives rise to the empirical free energy barrier to folding [66]. Likewise, a solvation/desolvation barrier arises during folding due to the asynchrony between the escape of water and the formation of internal contacts in the protein [67,68]. This solvation barrier is thought to contribute to the kinetic stability of proteins [68] because the breaking of internal contacts and the penetration of water during unfolding are not coincident. The simultaneous removal of a large number of water molecules would be required if folding involves the formation of interactions among many residues in a specific manner. In such a highly cooperative folding reaction, a large activation barrier would be expected due to a significant desolvation barrier [69]. On the other hand, the desolvation barrier is expected to disappear in so-called downhill folding, where the funnel landscape is smooth and the search for the TSE is not uphill energetically relative to the unfolded ensemble [67,70]. As a result of the solvation/desolvation barrier, the TSE is characterized by networks of partially formed, and consequently partially broken, internal contacts [68]. The unsatisfied polar interactions likely contribute, at least partially, to the higher energetic state of the TSE, indicating that the solvation barrier contributes to the activation enthalpy of the reaction. Indeed, solvation and desolvation effects may be the origin of the intrinsic enthalpy barriers observed by Arrhenius analysis of folding rate [67]. Liu and Chan have shown that small pairwise desolvation barriers, which are correlated to the nature of water, act cooperatively to give rise to an overall enthalpic barrier to folding [66]. Based on a small data set of proteins, Sanchez-Ruiz and coworkers estimate that approximately 80% of the activation energy is attributed to the solvation/desolvation barrier for folding [67].

Considerable differences exist in the specific contacts that occur in the TSE of various proteins. However, TSEs are thought, in general, to be dominated by native-like structures with many side chains fully dehydrated but with few of them forming their native contacts [71]. In a process referred to as ϕ -value analysis, the TSE can be probed indirectly by measuring the kinetic and thermodynamic effects of mutations, hydrophobic-to-alanine substitutions for example, in different regions of the protein [72–74]. The ϕ -value represents the change in stability of the TSE relative to that of the native ensemble as a result of the mutation of a residue, using the following relationship.

$$\phi(x) = \Delta\Delta G^\ddagger(x) / \Delta\Delta G_{eq}(x)$$

Where, $\Delta\Delta G_{eq}(x) = \Delta G_{mutant} - \Delta G_{wild-type}$ and $\Delta\Delta G^\ddagger(x) = RT \ln[k_f(mutant)/k_f(wild-type)]$

In this case, ΔG_{mutant} and $\Delta G_{wild-type}$ refer to the conformational free energies of the mutant and wild-type proteins, respectively, and k_f refers to the rate of folding for the mutant or wild-type protein extrapolated to zero denaturant. Importantly, there is considerable controversy regarding this analysis because it reports on side chain-side chain interactions, or in rare cases side chain-backbone interactions, rather than backbone-backbone interactions, and it does not take into account potential changes in the unfolded ensemble [75,76]. The analysis may lead to an ambiguous interpretation of the results because fundamentally it uses changes in energetic properties to infer structural properties of the TSE. Nevertheless, ϕ -value analysis has been used to examine the TSE of several small proteins (see reviews by Zarrine-Afsar & Davidson [77] and Daggett & Fersht [78]), and the results indicate that there are two general classes of TSEs [79,80]. In a diffuse TSE, almost all side chains have similarly low ϕ values, indicating that the native contacts are, at best, only partially formed in the TSE. A diffuse TSE is consistent with a nucleation-condensation mechanism for folding in which the folding nucleus is located diffusely throughout the protein [78,80]. In a polarized TSE, distinct substructures are present with high ϕ values, indicating that native contacts are formed in the substructures, whereas

other regions of the protein have low ϕ values, indicating that native contacts have not formed in those regions [81].

Illustrating the two classes of TSEs, Wittung-Stafshede and coworkers used a modified ϕ -value analysis, hydrophobic-to-polar substitutions, to suggest that the polarized TSE of apoazurin is completely dehydrated, while the diffuse TSE of zinc-substituted azurin contains water molecules involved in partially formed interactions [82–85]. Likewise, SH3 protein has been shown to contain a highly polarized TSE in which water molecules are removed after the rate-limiting step for folding [86], whereas pressure jump relaxation studies of Staphylococcal Nuclease (SNase) show that most of the water molecules are expelled from the hydrophobic core before the TSE is reached [71]. Cold shock protein B (CspB) also appears to have a dehydrated TSE [87], although many other proteins, such as tryptophan repressor [88] and CI2 for example [89], exhibit a hydrated TSE. In short, the role of water molecules in guiding folding through the TSE, particularly how hydration may affect folding rates and mechanisms, is a subject of intense study [71,90,91], but no general principles have been elucidated at present. The major conclusion from these studies is that the role of water in the folding reaction differs from one protein to the next.

6. Water in the Protein Folded State

Protein-water interactions in the folded state can be described from a variety of perspectives that together provide a picture of the importance of water in maintaining protein structure, function, dynamics and other features that contribute to the integrity of the native state. It is clear from a variety of methods that the water structure in the immediate vicinity of a protein has properties different from those of bulk water. To date, several reviews of protein hydration have been published summarizing information from NMR [50], X-ray crystallography [92], and neutron scattering [93], and progress continues to be made in other experimental areas such as Raman spectroscopy [94]. In addition, it has been well established by molecular dynamics simulations that the local solvent properties, such as density, compressibility and electrostatic screening are strongly altered by features of the protein surface [18,22,95,96]. Together the various approaches to the study of protein-water interactions using different techniques sample a range of time scales and provide an increasingly comprehensive view of the protein-water interface.

Although water binding sites appear in a wide variety of local environments on the protein, the estimated binding energies for crystallographically visible protein-bound water molecules in the first hydration shell is estimated to be within a narrow range between -0.38 and -0.56 kcal/mol, while those in the second hydration shell interact with an average energy of -0.04 kcal/mol [97]. These weak binding energies are consistent with the observation made by NMR spectroscopy that water molecules in the first hydration shell have low residence times and are in constant exchange with the bulk solvent (see discussion above). It is customary to refer to the crystallographically visible water molecules as the first hydration shell as if it were a well determined entity. However, a theoretical thermodynamic treatment of protein hydration shows how the presence of co-solvents and solutes can lead to preferential hydration, altering the number of protein bound water molecules according to the nature of the co-solvent, thus changing the composition of the first hydration shell [98]. This effect has been clearly documented experimentally for protein structures solved in the presence of high concentrations of organic solvents [16]. Elastase, for example, was solved in 11 different organic solvent conditions [99]. Superposition of these structures reveals over 400 unique water binding sites, only about 10% of which are found in all 11 structures, with almost half of the binding sites occupied in only one of the structures [92]. Another example of preferential hydration is found in the structures of extreme halophiles, where multiple layers of well-ordered water molecules can be observed in a network that forms a protective layer around the protein in an otherwise

dehydrating environment [100]. Molecular dynamics (MD) simulations also can demonstrate the effect of preferential hydration in proteins. MD studies of the 23-residue antimicrobial peptide magainin show that in trifluoroethanol(TFE)/water mixtures the TFE molecules accumulate close to the protein, excluding water and enhancing the intra-protein hydrogen bonds present in secondary structure, particularly α -helices [101]. In contrast, the preferential hydration in the presence of urea leads to direct hydrogen bonding interactions between the co-solvent and the protein, disruption of intra-protein interactions and swelling of the peptide, facilitating the penetration of water into the core and unfolding of the protein. An even more extensive molecular dynamics study looked at hydration of a serine protease in five organic solvents with different polarities and showed that the nature of the co-solvent determines the structure and dynamics of water at the protein surface [102]. Thus, in thinking about the first hydration shell, it is important to keep in mind that it is a very malleable entity that is highly influenced by the various components of the environment in which the protein is found.

In general, water molecules at the protein interface are seen to interact directly with the protein as expected from the stereochemistry requirements of the donor or acceptor atoms that are unpaired on the protein surface [103]. Many of these interactions are with protein side chain atoms, but water also plays an important role in fulfilling hydrogen bonding interactions in secondary structure. These occur primarily at the edges or ends of β -sheets, as well as at the N- and C-termini of α -helices, although water also is observed to bridge between carbonyl groups that already are involved in secondary structure [104]. Turns are most often found at the protein surface rather than at the core and therefore tend to be very well hydrated. The pattern of hydration varies with the type of turn, and hydrogen bonding occurs mainly to exposed carbonyl or amide groups, but bridging between two main chain atoms or a side chain to main chain often is observed, particularly in the more open turns. From the perspective of protein tertiary structure, water is most frequently found to be well-ordered, and therefore observed crystallographically, in grooves, which account for one quarter of the protein surface and bind half of all observed water molecules [105]. Of course, the first hydration shell is in contact with all areas of the protein surface [106], and its inhomogeneous distribution gives rise to an optimal balance of interactions that favors the net stabilization of the native state.

From the perspective of interactions made by water molecules with the protein, they have been classified into four categories based on information from crystallographic data: surface, crystal contacts, channel, and buried [92]. Surface water molecules typically make one or sometimes two hydrogen bonds with protein atoms, and their position with respect to the protein is highly sensitive to changes in the environment [16]. These water molecules are in constant and very rapid exchange with bulk water and are at the front line in a shell 6–8 Å of water with highly perturbed properties in the immediate vicinity of the protein [107]. Surface water molecules interact with polar atoms or charged side chains on the protein surface and often accompany conformational changes, most likely facilitating these motions [108]. In addition, nonpolar atoms on the protein surface contribute to the ordering of water molecules around the hydrophobic group in so-called hydrophobic interactions. These interactions, though energetically less favorable than those with polar groups, are critical in maintaining functionally important surfaces essential for macromolecular interactions [109,110]. A subgroup of the surface water molecules are in crystal contacts and are typically found within 4 Å of a symmetry-related protein molecule in the crystal. These water molecules have essentially the same properties as those of surface water molecules except that some of the hydrogen bonds occur with the symmetry related molecule across the crystal contact.

Ultra-high resolution crystal structures show a significant amount of connectivity between water molecules on the protein surface, forming continuous networks that can be linked with functional roles. The 1.1 Å resolution structure of the SRP GTPase Ffh, for example, shows three distinct water networks (Figure 5) [108]. The first occupies the nucleotide-binding site

and mediates interactions between fifty-four polar residues in that region, sixteen of which are strictly conserved in SRP GTPases (Figure 5A). The second network covers a hydrophobic pocket known to be involved in protein-protein interactions, forming the typical water rings previously discussed (Figure 5B). This pocket contains seven polar residues that contribute to the network, four of which are conserved in SRP GTPases. The third water network is found at the interface between the α -helical N-terminal domain and the G-domain (Figure 5C) and is thought to facilitate mobility at the interface during heterodimer formation [111]. There are eight evolutionarily conserved residues that are part of this network. This study provides a visual picture of how water molecules on the protein surface can mediate communication between functionally important protein residues and of how evolution has conserved residues that are critical in maintaining water binding sites involved in this process.

A smaller number of water molecules can be considered to be internal, in that they are found either in very deep crevices, that is channels, or buried in protein cavities. Water molecules found in channels make hydrogen bonds with at least two other water molecules in a groove on the protein and may facilitate large-scale, low frequency protein motions [16]. The channel water binding sites can be distinguished most easily from the surface type when multiple structures of a protein of interest are superimposed and analyzed together. Collectively, the water molecules found in the elastase channels fulfill all of the available hydrogen bonds within these sites, although in a single structure only a fraction of those are present [92]. Buried water molecules generally make three or four hydrogen bonds with main chain carbonyl or amide groups and typically are found either at the base of a water channel or bridging elements of secondary or tertiary structure, especially in more irregular structure such as turns [92]. Many buried water molecules are found to be conserved and are considered fundamental structural components of proteins [112]. Some buried water molecules are so important for either structural integrity or function that they are conserved across entire families of related proteins [113–115]. Not all conserved water molecules are buried, although many thought to stabilize significantly the native state are found in the interior of the protein to bridge polar groups that would otherwise be too far apart to interact. The 1.35 Å crystal structure of the Kelch domain of Keap1 provides a prime example of how conserved water molecules participate in maintaining the structural integrity of the protein [116]. The Kelch domain is composed of a six-bladed β -propeller where each blade consists of a four-stranded antiparallel twisted β -sheet containing ten strictly conserved water molecules (Figure 6). When the six blades are superimposed, three of the conserved waters can be seen in the central cavity of the propeller, interacting with backbone amide or carbonyl groups in the innermost β -strand. These water molecules play an important role in bridging the inner blades, which would otherwise be too far apart to interact. Two other conserved water molecules bridge consecutive blades of the propeller at the outermost edge. The remaining six waters bridge between various highly conserved residues in the loops or strands at the internal parts of the propeller. The strict conservation of the water binding sites reflected in the residues with which they interact, suggest that these water molecules play a prominent role in stabilizing the folded state and are likely to be important in the protein folding process [116].

7. Water at the Protein Core

Buried water molecules normally observed in crystal structures very often interact with main chain atoms to fulfill unsatisfied hydrogen bonds in the protein interior [117]. This is because the nature of the side chains can be optimized during evolution for packing in the hydrophobic core, and in the process the polar amide and carbonyl groups of the backbone are sometimes left unpaired. Another feature of buried water molecules is that the B-factor, a crystallographic descriptor of motion in an atom, decreases with an increase in the number of hydrogen bonds, up to three [112]. In fact, the B-factors of crystallographically visible buried water molecules typically are similar to those of the protein atoms with which they interact, demonstrating that

buried water molecules observed in crystal structures are generally well ordered. RNase T1, for example, contains a buried and evolutionarily conserved water molecule (Wat1) that forms four hydrogen bonds to several side chains in the protein [118]. These interactions may affect the dynamics of the active site, so evolution has chosen to preserve the water molecule rather than use an additional amino acid side-chain to contribute to the hydrogen bonding network [118,119].

Not all cavities in the protein core, however, contain crystallographically visible water molecules. A database study of cavities in 121 proteins found 265 cavities with crystallographically visible water, while many more, 383, contained no visible solvent [120]. These latter cavities have a high content of hydrophobic surface and tend to be on average smaller than the polar cavities. Several molecular dynamics studies of internal protein hydration have been performed in order to obtain a more thorough picture of the role of water in protein structure, dynamics and stability [121–123]. An excellent example of how MD simulations can be used to complement experiments is a study of a relatively large hydrophobic cavity in human interleukin-1 β , where the presence of water can be detected by NMR spectroscopy, but not observed in the crystal structure [124]. Up to four water molecules are found in the cavity at the core of the protein (Figure 7). They are dynamically hydrogen bonded to each other in various configurations and make oxygen-aromatic interactions with several phenylalanine residues lining the cavity. The simulations reveal correlated motions between distant but functionally important regions of the protein only when the cavity is hydrated, strongly suggesting that the buried water molecules are critical for information transfer during signal transduction.

As described by Janin, the number of buried water molecules can be predicted empirically by the relationship that large proteins contain quantifiably more buried water molecules than do small proteins [112,125]. However, the position of buried water molecules cannot be predicted based on protein size because buried water molecules are not distributed randomly in the protein interior. In fact, the *a priori* prediction of water in protein cavities and the way they interact within those sites has been somewhat problematic, primarily because the time scale for entrance of water into the core is longer than that sampled in molecular dynamics simulations. Thus, computational approaches other than molecular dynamics simulations are being developed and tested for *a priori* prediction of the presence of buried water molecules. For example, a recent adaptation of the reference interaction site model (RISM), based on integral equation theories of molecular solution, has been used successfully to predict all known internal water binding sites in hen egg white lysozyme [126].

Overall, the available structural data suggest that a water molecule that forms multiple hydrogen bonds to protein atoms is important in maintaining the protein structure. The implication from this conclusion is that the loss of such a water molecule should affect the structure, stability, and/or function of the protein. However, the experimental database shows that the effects of introducing or removing water binding sites are complex. Mutations that affect the positions of water molecules can be neutral energetically or can result either in a decrease or an increase in protein stability, depending on the location of the water and the properties, particularly the hydrophobicity, of the surrounding environment [127–134]. For example, in the energetically neutral mutations, a hydrogen bond in the native structure is replaced in the unfolded structure by an energetically equivalent hydrogen bond to water. On the other hand, new solvent binding sites were created in T4 and human lysozymes, for example, by site-directed point mutations in which water-bound cavities were introduced into the protein [135,136]. There were no mutations found in those proteins, however, that removed an internal water molecule, a fact that may demonstrate the importance of maintaining the hydrogen bonds contributed by water. Overall, the results with lysozyme showed that removal of a water molecule, resulting in a dry cavity in the protein, is energetically unfavorable because

the hydrogen bonding partners in the protein are left unsatisfied. Thus, the binding of water mitigates the destabilizing effect of introducing a dry cavity in the protein. Although the structural database suggests that buried water molecules should help stabilize protein structure, the limited experimental database suggests that stabilization or destabilization is context or position dependent, so no general principles have been established from those studies. For RNase T1, as well as other enzymes, the contributions of buried water molecules on preserving enzyme activity through affecting the protein dynamics may be a larger consideration than protein stability [118,119].

8. Conclusion

The role of water in guiding protein folding and in conferring stability as well as functional meaning to the native state is of such monumental importance that the current understanding of proteins cannot be imagined in its absence. Water is the medium in which the protein evolved, and water provides the context in which the protein's properties have been optimized. Yet water is so ubiquitous that for years it was "transparent" in the study of protein folding as well as in the analysis of protein structure and function, simply being taken for granted as a passive container of life.

The current literature has accumulated a large amount of evidence that water indeed plays a very active role in all aspects of the existence of a protein. It solubilizes the nascent polypeptide chain and as the protein folds it provides the major driving force for achieving the native state. During this process, water minimizes frustration, constraining the conformational freedom of the polypeptide chain, thus smoothing the folding funnel. As the nature of the local desolvation/solvation barriers varies with the surrounding environment, water effectively serves as an editor of the conformations sampled during protein folding to guide the ensemble of states toward the transition state, leading to the native state ensemble. In the process, water facilitates the organization and packing of structural elements in the protein core by solvating polar groups as they find their native contacts. The constant and rapid exchange with the bulk water mediates the expulsion of bound water as those native contacts are formed. This highly dynamic process of rapid exchange of protein-bound water with the bulk solvent underlies the ease with which water fulfills hydrogen bonding where needed and is primarily responsible for the flexibility of proteins both during folding and in the native state. In the presence of hydrophobic solvents, water is not able to undergo this exchange, and the protein is essentially frozen.

In the folded state, water continues to play its role in fulfilling hydrogen bonds where needed, both at the surface and at the core of the protein. Some buried water molecules are found in cavities that are formed due to the imperfections in the packing of the core and thus play an essential role in protein stability. This is exemplified in Figure 6, where conserved water molecules, many at the protein core, bridge polar groups in the β -propeller structure of the Kelch domain. Other water molecules are in cavities in which they are highly disordered due to the hydrophobic nature of the lining residues. This is exemplified in Figure 7 by a cavity in the core of interleukin-1 β thought to have evolved to facilitate communication between distant parts of the protein. Yet another function of buried water molecules may be in facilitating dynamics in the active site, as mentioned above for the case of Wat1 in RNase T1. The majority of water molecules, however, are at the surface where they mediate interactions of the protein with its solvent environment, facilitating protein motions, participating in enzyme reactions, mediating interaction with other molecules and fulfilling hydrogen bonds in a way that contributes to the stability and function of the native state.

Acknowledgments

This work was supported by a grant from the National Institutes of Health (GM065970 to ACC) and by a grant from the National Science Foundation (0237297 to CM).

References

1. Eiseley, L. The Immense Journey. Random House; New York: 1957.
2. Pace CN, Trevino S, Prabhakaran E, Scholtz JM. Philosophical Transactions of the Royal Society of London B: Biological Sciences 2004;359:1225–1235.
3. Nozaki Y, Tanford C. Journal of Biological Chemistry 1971;246:2211–2217. [PubMed: 5555568]
4. Radzicka A, Wolfenden R. Biochemistry 1988;27:1664–1670.
5. Privalov PL, Makhatadze GI. Journal of molecular biology 1993;232:660–679. [PubMed: 8393941]
6. Buhrman G, de Serrano V, Mattos C. Structure 2003;11:747–751. [PubMed: 12842038]
7. Lesser GJ, Rose GD. Proteins Structure Function and Genetics 1990;8:6–13.
8. Ruckenstein E, Shulgin IL. Advances in Colloid and Interface Science 2006;123–126:97–103.
9. Despa F, Fernandez A, Berry RS. Physical Review Letters 2004;93:228104–228104–228104–228104. [PubMed: 15601122]
10. Dill KA, Truskett TM, Vlachy V, Hribar-Lee B. Annual Review of Biophysics and Biomolecular Structure 2005;34:173–199.
11. Keutsch FN, Saykally RJ. Proceedings of the National Academy of Sciences USA 2001;98:10533–10540.
12. Barron LD, Hecht L, Wilson G. Biochemistry 1997;36:13143–13147. [PubMed: 9376374]
13. Klibanov AM. Nature 2001;409:241–246. [PubMed: 11196652]
14. Knubovets T, Osterhout JJ, Klibanov AM. Biotechnol Bioeng 1999;63:242–248. [PubMed: 10099601]
15. Mattos C, Ringe D. Current opinion in structural biology 2001;11:761–764. [PubMed: 11751059]
16. Mattos C. Trends in biochemical sciences 2002;27:203–208. [PubMed: 11943548]
17. Papoian GA, Ulander J, Eastwood MP, Luthey-Schulten Z, Wolynes PG. Proceedings of the National Academy of Sciences USA 2004;101:3352–3357.
18. Raschke TM. Current opinion in structural biology 2006;16:152–159. [PubMed: 16546375]
19. Burley SK, Petsko GA. Advances in protein chemistry 1988;39:125–189. [PubMed: 3072867]
20. Levitt M, Perutz MF. Journal of molecular biology 1988;201:751–754. [PubMed: 3172202]
21. Stollar EJ, Gelpi JL, Velankar S, Golovin A, Orozco M, Luisi BF. Proteins 2004;57:1–8. [PubMed: 15326588]
22. Hildebrandt A, Blossey R, Rjasanow S, Kohlbacher O, Lenhof HP. Bioinformatics 2006;23:e99–e103. [PubMed: 17237112]
23. Cozzini P, fornabai M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A. Current Medicinal Chemistry 2004;11:3093–3118. [PubMed: 15579003]
24. Levy Y, Onuchic JN. Annual Review of Biophysics and Biomolecular Structure 2006;35:389–415.
25. Whitesides GM, Krishnamurthy VM. Quarterly Reviews of Biophysics 2005;38:385–395. [PubMed: 16817982]
26. Rodier F, Bahadur RP, Chakrabarti P, Janin J. Proteins 2005;60:36–45. [PubMed: 15856483]
27. Teyra J, Pisabarro MT. Proteins 2007;67:1087–1095. [PubMed: 17397062]
28. Lu Y, Wang R, Yang CY, Wang S. J Chem Inf Model. 2007
29. Banerjee R, Sen M, Bhattacharya D, Saha P. Journal of molecular biology 2003;333:211–226. [PubMed: 14516754]
30. Dyson JH, Wright PE, Scheraga HA. Proceedings of the National Academy of Sciences USA 2006;103:13057–13061.
31. Cooper A. Biophysical chemistry 2005;115:89–97. [PubMed: 15752588]
32. Nakasako M. Journal of molecular biology 1999;289:547–564. [PubMed: 10356328]
33. Teeter MM. Proc Natl Acad Sci U S A 1984;81:6014–6018. [PubMed: 16593516]

34. Kauzmann W. Advances in protein chemistry 1959;14:1–62. [PubMed: 14404936]
35. Fleming PJ, Rose GD. Protein Science 2005;14:1911–1917. [PubMed: 15937286]
36. Myers JK, Pace CN. Biophysical journal 1996;71:2033–2039. [PubMed: 8889177]
37. Harpaz Y, Gerstein M, Chothia C. Structure 1994;2:641–649. [PubMed: 7922041]
38. Chandler D. Nature 2005;437:640–647. [PubMed: 16193038]
39. Huang X, Margulis CJ, Berne BJ. Proceedings of the National Academy of Sciences USA 2003;100:11953–11958.
40. Cheung MS, Garcia AE, Onuchic JN. Proceedings of the National Academy of Sciences USA 2002;99:685–690.
41. Denisov VP, Jonsson BH, Halle B. Nature structural biology 1999;6:253–260.
42. Chaplin M. Nature Reviews Molecular Cell Biology 2006;7:861–866.
43. Levy Y, Onuchic JN. Proceedings of the National Academy of Sciences USA 2004;101:3325–3326.
44. Fenimore PW, Frauenfelder H, McMahon BH, Parak FG. Proceedings of the National Academy of Sciences USA 2002;99:16047–16051.
45. Fenimore PW, Frauenfelder H, McMahon BH, Young RD. Proceedings of the National Academy of Sciences USA 2004;101:14408–14413.
46. Lubchenko V, Wolynes PG, Frauenfelder H. Journal of Physical Chemistry 2005;109:7488–7499. [PubMed: 16851860]
47. Frauenfelder H, Fenimore PW, Chen G, McMahon BH. Proceedings of the National Academy of Sciences USA 2006;103:15469–15472.
48. Teeter MM, Yamano A, Stec B, Mohanty U. Proc Natl Acad Sci U S A 2001;98:11242–11247. [PubMed: 11572978]
49. Vitkup D, Ringe D, Petsko GA, Karplus M. Nature structural biology 2000;7:34–38.
50. Halle, B. Hydration Processes in Biology: Theoretical and Experimental Approaches Bellissent-Funel MC, editor. 1999. p. 143-155.
51. Otting G, Liepinsh E, Wuthrich K. Science 1991;254:974–980. [PubMed: 1948083]
52. Otting G. Prog Nuclear Magn Reson Spectrosc 1997;31:259–285.
53. Lee CS, Ru MT, Haake M, Dordick JS, Reimer JA, Clark DS. Biotechnology and bioengineering 1998;57:686–693. [PubMed: 10099248]
54. Levinthal, C. Proceedings of a Meeting held at Allerton House. DeBrunner, JTP.; Munck, E., editors. University of Illinois Press; Monticello: 1969. p. 22-24.
55. Dill KA, Chan HS. Nature structural biology 1997;4:10.
56. Gruebele M. Current opinion in structural biology 2002;12:161–168. [PubMed: 11959492]
57. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Annual Review of Physical Chemistry 1997;48:545–600.
58. Wolynes P, Lathey-Schulten Z, Onuchic J. Chemical Biology. 1996
59. Wolynes PG. Quarterly Reviews of Biophysics 2005;38:405–410. [PubMed: 16934172]
60. Kovacs IA, Szalay MS, Csermely P. FEBS Letters 2005;579:2254–2260. [PubMed: 15848154]
61. Zhou R, Huang X, Margulis CJ, Berne BJ. Science (New York, NY) 2004;305:1605–1609.
62. Gfeller D, de los Rios P, Caflisch A, Rao F. Proceedings of the National Academy of Sciences USA 2007;104:1817–1822.
63. Rao F, Caflisch A. Journal of molecular biology 2004;342:299–306. [PubMed: 15313625]
64. Rank JA, Baker D. Protein Sci 1997;6:347–354. [PubMed: 9041636]
65. Sessions RB, Thomas GL, Parker MJ. Journal of molecular biology 2004;343:1125–1133. [PubMed: 15476826]
66. Liu Z, Chan HS. Journal of molecular biology 2005;349:872–889. [PubMed: 15893325]
67. Rodriguez-Larrea D, Ibarra-Molero B, Sanchez-Ruiz JM. Biophysical journal 2006;91:L48–L50. [PubMed: 16798814]
68. Rodriguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM. Journal of molecular biology 2006;360:715–724. [PubMed: 16784752]

69. Bai Y. Biochemical and Biophysical Research Communications 2006;340:976–983. [PubMed: 16405866]
70. Gruebele M. CR Biologies 2005;328:701–712.
71. Brun L, Isom DG, Velu P, Garcia-Moreno B, Royer CA. Biochemistry 2006;45:3473–3480. [PubMed: 16533028]
72. Goldenberg DP, Frieden RW, Haack JA, Morrison TB. Nature 1989;338:127–132. [PubMed: 2465497]
73. Matouschek A, Kellis JT Jr, Serrano L, Bycroft M, Fersht AR. Nature 1990;346:440–445. [PubMed: 2377205]
74. Matouschek A, Kellis JT Jr, Serrano L, Fersht AR. Nature 1989;340:122–126. [PubMed: 2739734]
75. Raleigh DP, Plaxco KW. Protein and Peptide Letters 2005;12:117–122. [PubMed: 15723637]
76. Sanchez IE, Kiefhaber T. Journal of molecular biology 2003;334:1077–1085. [PubMed: 14643667]
77. Zarrine-Afsar A, Davidson AR. Methods 2004;34:41–50. [PubMed: 15283914]
78. Daggett V, Fersht AR. Trends in biochemical sciences 2003;28:18–25. [PubMed: 12517448]
79. Goldenberg DP. Nature structural biology 1999;6:987–990.
80. Itzhaki LS, Otzen DE, Fersht AR. Journal of molecular biology 1995;254:260–288. [PubMed: 7490748]
81. Petrovich M, Jonsson AL, Ferguson N, Daggett V, Fersht AR. Journal of molecular biology 2006;360:865–881. [PubMed: 16784750]
82. Wilson CJ, Apiyo D, Wittung-Stafshede P. Protein Science 2006;15:843–852. [PubMed: 16522792]
83. Wilson CJ, Wittung-Stafshede P. Biochemistry 2005;44:10054–10062. [PubMed: 16042382]
84. Wilson CJ, Wittung-Stafshede P. Proceedings of the National Academy of Sciences USA 2005;102:3984–3987.
85. Zong C, Wilson CJ, Shen T, Wolynes PG, Wittung-Stafshede P. Biochemistry 2006;45:6458–6466. [PubMed: 16700556]
86. Fernandez-Escamilla AM, Cheung MS, Vega MC, Wilmanns M, Onuchic JN, Serrano L. Proceedings of the National Academy of Sciences USA 2004;101:2834–2839.
87. Jacob MH, Saudan C, Holtermann G, Martin A, Perl D, Merbach AE, Schmid FX. Journal of molecular biology 2002;318:837–845. [PubMed: 12054827]
88. Desai G, Panick G, Zein M, Winter R, Royer CA. Journal of molecular biology 1999;288:461–475. [PubMed: 10329154]
89. Mohana-Borges R, Silva JL, Ruiz-Sanz J, de Prat-Gay G. Proceedings of the National Academy of Sciences USA 1999;96:7888–7893.
90. Bezsonova I, Korzhnev DM, Prosser RS, Forman-Kay JD, Kay LE. Biochemistry 2006;45:4711–4719. [PubMed: 16605239]
91. Pradeep L, Udgaonkar JB. Journal of molecular biology 2007;366:1016–1028. [PubMed: 17188296]
92. Mattos C, Ringe D. International Tables for Crystallography F 2001:623–640.
93. Niimura, N. Neutron Scattering in Biology, Techniques and Applications. Fitter, J.; Gutberlet, T.; Katsaras, J., editors. 2006. p. 43-62.
94. Hunt NT, Kattner L, Shanks RP, Wynne K. Journal of the American Chemical Society 2007;129:3168–3172. [PubMed: 17315992]
95. Dadarlat VM, Post CB. Biophysical journal 2006;91:4544–4554. [PubMed: 16997864]
96. Kim B, Young T, Harder E, Friesner RA, Berne BJ. The journal of physical chemistry 2005;109:16529–16538. [PubMed: 16853101]
97. Amadasi A, Spyarakis F, Cozzini P, Abraham DJ, Kellogg GE, Mozzarelli A. Journal of molecular biology 2006;358:289–309. [PubMed: 16497327]
98. Timasheff SN. Biochemistry 2002;41:13473–13482. [PubMed: 12427007]
99. Mattos C, Bellamacina CR, Peisach E, Pereira A, Vitkup D, Petsko GA, Ringe D. Journal of molecular biology 2006;357:1471–1482. [PubMed: 16488429]
100. Britton KL, Baker PJ, Fisher M, Ruzhenikov S, Gilmour DJ, Bonete MJ, Ferrer J, Pire C, Esclapez J, Rice DW. Proc Natl Acad Sci U S A 2006;103:4846–4851. [PubMed: 16551747]
101. Mehrnejad F, Naderi-Manesh H, Ranjbar B. Proteins 2007;67:931–940. [PubMed: 17357162]

102. Micaelo NM, Soares CM. The FEBS journal 2007;274:2424–2436. [PubMed: 17419728]
103. Thanki N, Thornton JM, Goodfellow JM. Journal of molecular biology 1988;202:637–657. [PubMed: 3172231]
104. Thanki N, Umrana Y, Thornton JM, Goodfellow JM. Journal of molecular biology 1991;221:669–691. [PubMed: 1920440]
105. Kuhn LA, Siani MA, Pique ME, Fisher CL, Getzoff ED, Tainer JA. Journal of molecular biology 1992;228:13–22. [PubMed: 1447777]
106. Levitt M, Park BH. Structure 1993;1:223–226. [PubMed: 8081736]
107. Lounnas V, Pettitt BM, Phillips GN Jr. Biophysical journal 1994;66:601–614. [PubMed: 8011893]
108. Ramirez UD, Freymann DM. Acta crystallographica 2006;62:1520–1534.
109. Clackson T, Wells JA. Science (New York, NY 1995;267:383–386.
110. Thorn KS, Bogan AA. Bioinformatics 2001;17:284–285. [PubMed: 11294795]
111. Gawronski-Salerno J, Freymann DM. Journal of structural biology 2007;158:122–128. [PubMed: 17184999]
112. Park S, Saven JG. PROTEINS: Structure, Function, and Genetics 2005:450–463.
113. Bottoms CA, White TA, Tanner JJ. Proteins 2006;64:404–421. [PubMed: 16700049]
114. Loris R, Stas PP, Wyns L. The Journal of biological chemistry 1994;269:26722–26733. [PubMed: 7929406]
115. Sreenivasan U, Axelsen PH. Biochemistry 1992;31:12785–12791. [PubMed: 1334431]
116. Beamer LJ, Xuchu L, Bottoms CA, Hannink M. Acta crystallographica 2005;D61:1335–1342.
117. Williams MA, Goodfellow JM, Thornton JM. Protein Sci 1994;3:1224–1235. [PubMed: 7987217]
118. Langhorst U, Loris R, Denisov VP, Doumen J, Roose P, Maes D, Halle B, Steyaert J. Protein Science 1999;8:722–730. [PubMed: 10211818]
119. Langhorst U, Backmann J, Loris R, Steyaert J. Biochemistry 2000;39:6586–6593. [PubMed: 10828976]
120. Hubbard SJ, Gross KH, Argos P. Protein engineering 1994;7:613–626. [PubMed: 8073031]
121. Damjanovic A, Garcia-Moreno B, Lattman EE, Garcia AE. Proteins 2005;60:433–449. [PubMed: 15971206]
122. Garcia AE, Hummer G. Proteins 2000;38:261–272. [PubMed: 10713987]
123. Sterpone F, Ceccarelli M, Marchi M. Journal of molecular biology 2001;311:409–419. [PubMed: 11478869]
124. Somani S, Chng CP, Verma CS. Proteins 2007;67:868–885. [PubMed: 17380484]
125. Janin J. Nature 1979;277:491–492. [PubMed: 763335]
126. Imai T, Hiraoka R, Kovalenko A, Hirata F. Proteins 2007;66:804–813. [PubMed: 17186526]
127. Berndt KD, Beunink J, Schroder W, Wuthrich K. Biochemistry 1993;32:4564–4570. [PubMed: 7683491]
128. Buckle A, Cramer P, Fersht AR. Biochemistry 1996;35:4298–4305. [PubMed: 8605178]
129. Das G, Hickey DR, McLendon D, McLendon G, Sherman F. Proceedings of the National Academy of Sciences USA 1989;86:496–499.
130. Hickey DR, Berghuis AM, Lafond G, Jaeger JA, Cardillo TS, McLendon D, Das G, Sherman F, Brayer GD, McLendon G. Journal of Biological Chemistry 1991;266:11686–11694. [PubMed: 1646814]
131. Lett CM, Berghuis AM, Frey HE, Lepock JR, Guillemette JG. Journal of Biological Chemistry 1996;271:29088–29093. [PubMed: 8910563]
132. Pedersen JT, Olsen OH, Betzel C, Eschenburg S, Branner S, Hastrup S. Journal of molecular biology 1994;242:193–202. [PubMed: 8089841]
133. Takano K, Funahashi J, Yamagata Y, Fujii S, Yutani K. Journal of molecular biology 1997;274:132–142. [PubMed: 9398521]
134. Xu J, Baase WA, Baldwin E, Matthews BW. Protein Science 1998;7:158–177. [PubMed: 9514271]
135. Takano K, Yamagata Y, Yutani K. Protein engineering 2003;16:5–9. [PubMed: 12646687]

136. Xu J, Baase WA, Quillin ML, Baldwin EP, Matthews BW. Protein Science 2001;10:1067–1078. [PubMed: 11316887]
137. Sarkhel S, Desiraju GR. PROTEINS: Structure, Function, and Bioinformatics 2004;54:247–259.

Figure 1.

Examples of three-centered hydrogen bonds. (A) Backbone carbonyl, amide proton, and water proton hydrogen bonds. (B) Hydrogen bonds between amino acid side chains tyrosine and glutamate with water. For A and B, the yellow sphere represents a water molecule (PDB code 2J30). Additional examples of three-centered hydrogen bonds can be found in Sarkhel and Desiraju [137].

Figure 2.
Comparison of water removal during folding via the drying (“dewetting”) or expulsion mechanisms.

Figure 3.

Funnel energy landscapes. (A) The “golf-course” energy landscape representing a random search of conformational space for the thermodynamically stable native conformation. (B) Smooth energy landscape funnel representing folding to the native conformation. Images were kindly provided by Dr. Ken Dill (www.dillgroup.ucsf.edu/).

Figure 4.

Representative desolvation barriers. (A) Schematic representation of desolvation to form native contacts. (B) Potential of mean force (PMF) energy (E) *versus* sphere center-to-center distance (D) for the association of two methane molecules in water. (C) PMF for the association of two alanine methyl groups modeled in RAFT (reduced Ramachandran angle and folding force field for tertiary structure prediction) with (continuous line) and without (broken line) a desolvation barrier. For B and C, green spheres represent methane molecules or methyl groups, blue spheres represent water molecules, and blue lines represent a stick representation of the peptide backbone. Panels B and C were reprinted with permission from Sessions *et al.* [65].

Figure 5.

Examples of functionally important hydrogen bonding networks facilitated by water molecules in SRP Ffh. The central ribbon diagram shows the orientation of the three panels relative to each other. (A) In the GTPase active site, water molecules hydrogen bond to residues in the GTPase motifs I–IV. Water molecules are shown in blue. Conserved waters are shown with a larger diameter. (B) Water molecules in the conformationally sensitive pocket central to the “latch” interaction of the targeting heterodimer. Waters that hydrogen bond to residues in motifs I and III are in yellow. (C) Waters (red) that contribute to the conformational flexibility between the N and G subdomains. Reprinted with permission from Ramirez and Freymann [108].

Figure 6.

Kelch domain of human Keap1. (A) Ribbon diagram showing the six-bladed β -propeller structure. Consecutive blades are numbered 1–6, and the strands of blade 1 are labeled A–D. (B) Sixty conserved water molecules (ten per blade) are shown as spheres for the Kelch domain. Yellow waters occupy the central channel. Magenta waters interact with the loop between strands A and B, orange waters interact with the loop between strands C and D, and green and white waters line the solvent-exposed edge of strand D on the outside of the β -propeller. Reprinted with permission from Beamer *et al.* [116].

Figure 7.

Four water molecules in the central cavity of human interleukin-1 β . (A) Ribbon diagram of human interleukin-1 β . Water molecules are shown as red spheres. (B) Hydrophobic residues lining the cavity (sticks) are shown packing against the water molecules (spheres). Reprinted with permission from Somani *et al.* [124].