Structure and Energetics of the Hydrogen-Bonded Backbone in Protein Folding

D. Wayne Bolen¹ and George D. Rose²

¹Department of Biochemistry and Molecular Biology and The Sealy Center for Structural Biology, The University of Texas Medical Branch, Galveston, Texas 77555-1052; email: dwbolen@utmb.edu
²T.C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218; email: grose@jhu.edu

Key Words
m value, protein denaturation, organic osmolyte, solvent quality, Tanford Transfer Model

Abstract
We seek to understand the link between protein thermodynamics and protein structure in molecular detail. A classical approach to this problem involves assessing changes in protein stability resulting from added cosolvents. Under any given conditions, protein molecules in aqueous buffer are in equilibrium between unfolded and folded states, U(nfolded) ⇄ N(ative). Addition of organic osmolytes, small uncharged compounds found throughout nature, shift this equilibrium. Urea, a denaturing osmolyte, shifts the equilibrium toward U; trimethylamine N-oxide (TMAO), a protecting osmolyte, shifts the equilibrium toward N. Using the Tanford Transfer Model, the thermodynamic response to many such osmolytes has been dissected into groupwise free energy contributions. It is found that the energetics involving backbone hydrogen bonding controls these shifts in protein stability almost entirely, with osmolyte cosolvents simply dialing between solvent-backbone versus backbone-backbone hydrogen bonds, as a function of solvent quality. This reciprocal relationship establishes the essential link between protein thermodynamics and the protein's hydrogen-bonded backbone structure.
**INTRODUCTION**

“Function follows structure” is an oft-repeated truism in biology. A vast number of biological components attain their structure via spontaneous, hierarchic self-assembly, a modular, bottom-up process rootest in the spontaneous folding of individual proteins (1). Under suitable conditions, a globular protein will experience a spontaneous drive toward its folded native conformation from any accessible initial conformation (2). Although not all proteins work in exactly this way (3, 4), most do, or almost so. Without the spontaneous disorder ⇀↽ order folding reaction, life as we know it could not exist.

The driving force for this protein folding reaction, $U(\text{folded}) \rightleftharpoons U(\text{native})$, is exerted by a gradient in Gibbs free energy (2), like a ball rolling downhill under the influence of a gravitational potential. This thermodynamic description has characterized our thinking about protein folding in buffered solution for decades. Arguably, thermodynamics is our most powerful tool for understanding biological processes, and it has been applied to protein folding for more than 75 years (5, 6). Yet, the field still lacks physical-chemical understanding of protein folding at a molecular level. Are we missing something essential?

Over the past half century, views about the important forces in the thermodynamics of protein folding have been continuously changing targets. The driving force in folding was initially thought to be intramolecular hydrogen bonding (7), then the hydrophobic effect (8). In recent times, it has been argued that intramolecular hydrogen bonding is destabilizing (9, 10), partially stabilizing and destabilizing (11) and, once again, an important driving force (12, 13). During the past two decades, the advent of protein engineering brought a hope that the newfound ability to introduce site-directed mutants at will would provide ready answers to such unresolved questions (14). However, despite thousands of mutational experiments, disagreement about the energetic role of hydrogen bonding remains, and the likelihood of success in resolving these questions using previous approaches is fading. This inability to resolve the energetic role of hydrogen bonding in protein folding has been a lingering conceptual obstacle that

---

**Contents**

INTRODUCTION 340
Two Paradigms for Analyzing Protein Folding 342
TERMS AND CONCEPTS 343
Solvent Quality and Protein Conformation 343
Organic Osmolytes and Solvent Quality 344
Transfer Free Energies: Quantifying Solvent Quality Effects on Proteins 345
The Transfer Model: Identifying Driving Forces and Predicting Protein Behavior 345
WHICH CONTRIBUTES MORE TO THE $m$ VALUE, BACKBONE OR SIDE CHAINS? 346
Dissection of Backbone and Side Chain $m$-Value Contributions: A Case Study 346
DOES THE PEPTIDE HYDROGEN BOND STABILIZE PROTEINS? 349
Selected Highlights from Seven Decades of Hydrogen-Bonding History 349
Lingering Doubts and Alternative Explanations 351
THE SOLVENT QUALITY OF WATER 353
How Solvated Is the Peptide Unit in Water? 354
A SIMPLE STRUCTURAL ORIGIN FOR OBSERVED OSMOLYTE THERMODYNAMICS 355
continues to impede our understanding of protein stability.

The strategy that nature uses to regulate protein stabilization/destabilization introduces a viewpoint that differs radically from familiar ideas about protein folding in buffer solution, so well explored during the last half century. Nature’s strategy operates predominantly on the backbone in the unfolded state, with much less involvement of the native state. When analyzed from this perspective, the pivotal role of intramolecular hydrogen bonding in both stability and structure formation is revealed with clarity.

Throughout the course of evolution, nature has successfully modulated protein folding using organic osmolytes, small organic molecules whose intracellular presence protects cells from environmental stress conditions that would otherwise threaten survival (15). Temperature, pressure, and denaturing cosolvents can unfold proteins in a natural setting, just as they do in the laboratory. For example, thermal stresses subject intracellular proteins to destabilization from hot or cold denaturation, or desiccation. In such cases, a repertoire of polyols with the ability to counteract denaturation under such conditions has been selected for intracellular accumulation. Similarly, chemical stress is commonplace in the urea-rich cells of sharks and rays, and even more so in the kidney cells of mammals. Urea is a denaturing osmolyte, and in these cases, methylamine osmolytes with the ability to counteract solvent denaturation have been selected for intracellular accumulation (16, 17). In general, naturally occurring osmolytes are used extensively throughout all three kingdoms of life, and understanding their mode of action can provide insight into the way nature regulates folding and stability so successfully.

Accordingly, this review departs from the typical retrospective of protein folding energetics in buffer solution. Taking our cue instead from the time-tested rigors of natural selection, we review protein folding primarily in relation to the way that organic osmolytes modulate the $U \leftrightarrow N$ equilibrium.

The overall goal of this review is to relate the thermodynamics of folding to the organization of molecular structure in mechanistic detail. Our inquiry into the molecular origin of protein structure is guided by a large body of experimental evidence, much of it involving the relationship between solvent quality and protein stability. Ultimately, we suggest that seemingly disparate data in the literature can be rationalized by a single mechanism that originates primarily in the hydrogen-bonded backbone.

Putting the “bottom line” at the top of the review, we argue that the half century quest to solve the protein folding problem has been impeded primarily by failure to correctly assess the energy of the peptide hydrogen bond in water. Water is a poor solvent for unfolded proteins, i.e., intramolecular protein interactions are favored over protein-water interactions. Were this not so, proteins in water would not fold spontaneously. Recent experimental evidence adds a specific and surprising new twist to this self-evident inference: Predominantly, water is a poor solvent for the peptide backbone, as discussed at length in this review. Of course, water is also a poor solvent for apolar side chain groups. Indeed, the ongoing focus on the hydrophobic effect as the driving force in protein folding has tended to obscure the crucial role of the protein backbone, an oversight we seek to remedy.

In particular, evidence from solution thermodynamics suggests that intramolecular hydrogen bonds are marginally favored over water-backbone hydrogen bonds (18). In folded proteins, approximately two-thirds of the intramolecular hydrogen bonds are within repetitive elements of secondary structure (19). From these observations, it is plausible that desolvation of the peptide unit in water is more than compensated by formation of intramolecular peptide hydrogen bonds in the $\alpha$-helix and $\beta$-sheet, the hydrogen-bonded scaffold elements upon which proteins are built. If true, this plausible mechanism would constitute a fundamental link between

---

**Protein folding:** the spontaneous acquisition of native structure under solution conditions that favor the native state

**Organic osmolyte:** small uncharged molecules, found throughout nature, which are involved in cell volume regulation and modulate solvent quality

**Urea:** a denaturing osmolyte, which is highly effective in inducing protein unfolding

**Solvent quality:** the character of any solvent relative to a reference solvent, often described on a scale ranging from good to poor
solvation thermodynamics and the protein’s hydrogen-bonded backbone structure (20).

Contrary to this hypothesis, many papers in the literature conclude that, in fact, a large desolvation penalty overwhelms the transfer of a peptide unit from water into a folded protein, and therefore without additional, compensating energetic factors, the protein would not fold. This conclusion has prompted many complex ideas about folding; for insightful discussions, see several recent reviews by Kallenbach and coworkers (21), Baldwin (22) and Bowler (23). Here, we review literature pertaining to a simple hydrogen bond-based mechanism for protein folding.

Two Paradigms for Analyzing Protein Folding

As defined by Anfinsen (2) and Mirsky & Pauling before him (5), protein folding is inherently a thermodynamic problem. The core of this problem is to enumerate and quantify the factors that link protein energetics to protein structure and stability.

Since Kauzmann’s seminal review (8), the “energy-ledger” paradigm has conditioned our thinking about the relationship between protein energetics and protein structure. In the energy-ledger approach, the folding process is partitioned into additive components that can be quantified or estimated independently. In a two-state folding reaction, $U \rightleftharpoons N$, the folding equilibrium constant, $K = [N]/[U]$, can be measured under some set of physical-chemical conditions and used to derive a free energy difference between the unfolded and folded populations, $\Delta G_{U\rightarrow N} = -RT \ln K$ (24). Many contributions to $\Delta G_{U\rightarrow N}$ can then be estimated using analogous model compounds, as Kauzmann showed.

However, there is a practical stumbling block in the energy-ledger approach, as revealed in Brandts’ early papers on the energetics of chymotrypsinogen stability (25, 26). $\Delta G_{U\rightarrow N}$ is the small difference between large, opposing terms. Under folding conditions, entropy ($T \Delta S_{U\rightarrow N}$) favors the unfolded state, enthalpy ($\Delta H_{U\rightarrow N}$) favors the native state, and each contributes $\sim 100$–200 kcal/mol (see figure 4 in Reference 26), resulting in $\Delta G_{U\rightarrow N} = \Delta H_{U\rightarrow N} - T \Delta S_{U\rightarrow N}$ of order $-10$ kcal/mol. Upon evaluating these quantities, even small errors could easily result in the wrong sign at the bottom of the energy ledger, as Brandts pointed out.

The energy ledger is a tried-and-true strategy to assess a state function by partitioning it into separable components, measuring them individually, and then summing these values to obtain net system behavior. Hess’s law for the heat of combustion is a familiar example. However, unlike Hess’s law, where the data can be measured at high accuracy, the experimental quantities in protein folding are estimates and averages that depend on the microenvironment, and significant inaccuracies cannot be eliminated. Beset by such practicalities, a pure energy-ledger approach is ill matched to the protein folding problem.

Of even greater concern though, the energy-ledger approach may also be inherently ill matched to the question of interest. In this review, that question is whether the transition between the unfolded and folded populations can be ascribed to a predominant molecular origin. In a two-state folding reaction, $U \rightleftharpoons N$, a population of folded molecules still persists under destabilizing conditions ($\Delta G > 0$), and therefore the information needed to encode the fold must survive energetic variation within the range set by $\Delta G_{U\rightarrow N}$ (27). This analysis suggests that some energy-ledger quantities are decisive and that others are only secondary.

This abstract conclusion is nicely illustrated by a concrete example: The Protein Data Bank (28) holds $\sim 400$ lysozyme mutants spanning a range of energies, but all have essentially identical overall conformation (29). A straightforward interpretation of this phenomenon is that the fold is encoded robustly, i.e., small changes in free energy have little impact on the gross native structure. In general,
the fidelity of the fold is maintained by a robust hydrogen-bonded network that typically survives conformational changes arising from small, localized perturbations, as is shown below.

Indeed, nature can compensate for a broad diversity of destabilizing conditions simply by shifting the folding equilibrium to favor this hydrogen-bonded network. Nature achieves this result by use of organic osmolytes, a natural strategy that has now been deciphered, as discussed next.

In addition to the energy ledger, another time-tested way to assess a system is the stimulus-response approach. Here, one introduces a perturbation and then tries to intuit whether one or a few principal underlying factors are responsible for the system’s response to that perturbation. When most successful, this approach facilitates construction of a predictive model of system behavior.

The two approaches are of opposite type: The energy ledger is from the bottom up and reductive, whereas stimulus-response is from the top down and inductive. Many—perhaps most—folding studies are of the latter type, involving perturbations of either physical (temperature, volume, pressure) or chemical (pH, mutations, solvent quality) factors, with results often quantified as $\Delta G$

In particular, much of the thermodynamic data covered in this review comes from experiments involving the measurement and interpretation of osmolyte-induced changes in solvent quality, using the Tanford Transfer Model (or, simply, the Transfer Model) (30, 31). These are stimulus-response-type experiments in which cosolvent additions affect protein stability in an informative way. Taken together, data from these experiments prompt an obvious folding mechanism. Before proceeding, it is necessary to define a few terms and concepts.

**TERMS AND CONCEPTS**

The “protein folding problem,” an ongoing challenge for protein folders, is not a problem for nature. Natural conditions subject cells to a wide range of physical and chemical extremes, stress conditions that can denature most proteins. To counteract such effects, nature utilizes a diverse assortment of organic osmolytes, as mentioned previously. A central purpose of this review is to distill insights from osmolyte-induced protein folding and use them to cast light on the energetics of classical protein folding studies. To this end, we introduce concepts from adaptive biology and the theory of polymer solutions that are needed for this inquiry.

**Tanford Transfer Model:** a thermodynamic cycle comparing the difference between the native ⇔ unfolded state equilibrium in osmolyte and in buffer.

**Solvent Quality and Protein Conformation**

In this review, water or dilute buffer at pH 7 is defined as a reference solvent for the protein. The term *solvent quality*, from polymer science, describes the character of some other solvent with respect to this reference condition, and the terms “poor” and “good” are used to designate any change in solvent character (32). In a *poor solvent*, protein solubility is reduced, and the mean radius of gyration, $\langle R_G \rangle$, of the unfolded ensemble contracts. Conversely, in a *good solvent*, protein solubility is increased, and the $\langle R_G \rangle$ of the unfolded ensemble expands. Protein in poor solvent is said to be *solvophobic*, and in a good solvent, *solvophilic*. The protein’s solvent response indicates that in a good solvent, protein-solvent interactions are enhanced at the expense of protein intramolecular interactions and/or solvent-solvent interactions, whereas in a poor solvent, protein intramolecular interactions predominate.

The protein’s solvent response can be measured without regard to the molecular origin of the response, and we employ the solvent quality paradigm for just this reason. However, ultimately we seek to interpret such measurements at the molecular level by decomposing the overall solvent response into the respective contributions made by individual chemical groups.
Organic Osmolytes and Solvent Quality

A basic stress response in living systems is to adjust protein stability by regulating solvent quality. Understanding this natural process can provide deep insights into the mechanism of protein folding.

In greater detail, essentially all organisms can experience various types of water stress, i.e., stresses such as high or low temperature, desiccation, and external osmotic pressure (33). To avoid osmotic catastrophe under such conditions, cell volume is maintained osmotically via carefully controlled changes in the intracellular concentrations of organic osmolytes, small uncharged molecules that can also modulate solvent quality (15, 34–36). Osmolytes may be classified as either denaturing or protecting. In the equilibrium folding reaction, $U \rightleftharpoons N$, the denaturing osmolyte urea stabilizes $U$ relative to $N$; protecting osmolytes do the opposite, stabilizing $N$ relative to $U$ (38, 39). Often, water stress is coupled with protein-denaturing stress, e.g., from either elevated temperature or urea. In such cases, natural selection has provided a repertoire of protecting osmolytes, which do double duty—both regulating cell volume and counteracting denaturation (15). A clear example of the latter occurs in the urea-rich cells of sharks and rays, where the protecting osmolyte trimethylamine-N-oxide (TMAO) is used to counteract protein denaturation (16). In the papilla of mammalian kidney, urea and salt concentrations can reach several molar in some species (38), and correspondingly high intracellular concentrations of protecting osmolytes are deployed in such cases (39, 40). These examples underscore the vital role that solvent plays in protein stability and folding.

How does solvent quality modulate protein stability? It is important to realize that the folding equation, $U \rightleftharpoons N$, is not an ordinary chemical reaction; no covalent bonds are made or broken. Instead, the folded fraction is simply dialed up or down in response to physicochemical conditions such as solvent quality. In thermodynamic terms, protecting osmolytes like TMAO are solvophobic; they effect a substantial increase in the free energy of the $U$ state above that of the $N$ state relative to the protein’s free energy in buffer (41, 42). This increase shifts the $U \rightleftharpoons N$ folding equilibrium toward $N$ by changing the ground states of $N$ and $U$ on transfer from water to osmolyte solution, as illustrated in Figure 1 (37, 43). Conversely, denaturing osmolytes like urea are solvophilic; they effect a substantial decrease in the free energy of the $U$ state below that of the $N$ state relative to the free energy in buffer, shifting the folding equilibrium toward $U$. The thermodynamic consequence of this natural mechanism is to repopulate denatured or intrinsically disordered states relative to the native state and, importantly, without altering any information encoded in the native fold (34).

Figure 1

Solvent quality, free energy and protein folding. Predominantly, organic osmolytes affect the unfolded state ($U$), much more than the folded state ($N$). (left) Relative to the protein in buffer, addition of a protecting osmolyte, e.g., trimethylamine-N-oxide (TMAO), raises the free energy of $U$ above that of $N$, shifting the $U \rightleftharpoons N$ equilibrium toward $N$. (right) Conversely, addition of the denaturing osmolyte, urea, lowers the free energy of $U$ below that of $N$, shifting the $U \rightleftharpoons N$ equilibrium toward $U$. Abbreviations: $N_{aq}$ represents the native state in aqueous buffer; $U_{aq}$ represents the unfolded state in aqueous buffer.
Transfer Free Energies: Quantifying Solvent Quality Effects on Proteins

A basic level of quantifying protein-solvent interactions involves the use of transfer free energies. The transfer free energy, $\Delta G_t$, of any solute (protein in this case) from water to a second solvent system will be either favorable or unfavorable. By definition, an unfavorable transfer free energy, $\Delta G_t > 0$, means the protein becomes solvophobic on transfer to a poor solvent, whereas a favorable transfer free energy, $\Delta G_t < 0$, means the protein becomes solvophilic on transfer to a good solvent. The sign and magnitude of the measured $\Delta G_t$ quantifies the protein’s response to changes in solvent quality.

However, as in all thermodynamic descriptions, $\Delta G_t$ values do not address the molecular details that account for these solvent quality changes. Clues to the underlying molecular origins can be discovered by dissecting the overall $\Delta G_t$ values into component contributions corresponding to individual protein groups. Accordingly, apparent transfer free energies from water to 1 M osmolyte solutions have been determined for the amino acids, their side chains, and the peptide backbone unit (44–48). Such data can then be applied to molecular hypotheses using energy-ledger approaches.

The Transfer Model: Identifying Driving Forces and Predicting Protein Behavior

The Transfer Model, shown in Scheme 1, compares the extent to which the native (N) and denatured (U) state equilibrium in 1 M osmolyte (given by $\Delta G_{t,N}$) differs from native protein from buffer solution to 1 M osmolyte. From the fact that the Transfer Model is a thermodynamic cycle, it follows that $\Delta G_{t,N} = \Delta G_{t,U}$ and $\Delta G_{t,U} = \Delta G_{t,N}$, and therefore, the free energy difference of the perpendicular reactions, $\Delta G_{t,U} - \Delta G_{t,N}$, is also equal to the $m$ value.

Assuming additivity, $\Delta G_{t,U}$ and $\Delta G_{t,N}$ can be evaluated by summing their components, the experimentally obtained individual apparent group transfer free energies (GTFEs) of side chains and the peptide transfer free energies of unfolded and native protein from buffer solution to 1 M osmolyte. Scheme 1 is a thermodynamic cycle, it follows that $\Delta G_{t,N} = \Delta G_{t,U}$ and $\Delta G_{t,U} = \Delta G_{t,N}$, and therefore, the free energy difference of the perpendicular reactions, $\Delta G_{t,U} - \Delta G_{t,N}$, is also equal to the $m$ value.

$m$ value: an experimentally obtained measure of osmolyte efficacy in folding/unfolding a protein; also a measure of the folding cooperativity

GTFE: group transfer free energy
backbone. For the N state, the accessible surface area (ASA) of the native structure is computed (49) from atomic coordinates (28), using a solvent probe of radius 1.4 Å. The calculated ASA of each constituent backbone and side chain group is normalized by its standard state ASA, defined as the area of that group in a Gly-X-Gly tripeptide (50). These normalized values are summed by group, and the groupwise sums are then multiplied by their corresponding GTFEs, providing groupwise free energy subtotals that total $\Delta G_{N,U}$. Further details of the procedure are described in Auton & Bolen (45).

Similarly for the U state, evaluation of $\Delta G_{U,N}$ requires additivity of component groups and, in this case, a denatured state model from which accessible surface areas can be obtained. Creamer et al. (51) proposed two extreme models that are intended to bracket denatured state solvent accessibilities between credible boundaries: (a) a random coil in a good solvent, representing an expanded, solvent-exposed limit, and (b) a compact coil in a poor solvent, representing a collapsed, solvophobic limit. Following Schellman (52), Creamer’s groupwise limits were averaged, multiplied by their corresponding GTFEs, and summed to obtain $\Delta G_{U,N}$. Reassuringly, the numerical average of Creamer’s two extremes closely resembles the ASA of Goldenberg’s self-avoiding random coil model (53).

These m-value calculations constitute a rigorous test of the Transfer Model. GTFEs are measurements—not fitted parameters, and the model requires that they be strictly additive. These are stringent constraints, especially so considering that GTFE values differ from one osmolyte to another and that both osmolyte-induced folding and urea-induced unfolding are encompassed within a single, invariant framework. Accurate m-value prediction would be impossible if either of the assumptions, additivity or Schellman’s model, is invalid or if the measurements are in error. At this writing, predicted m values, covering a range of nine kcal/mol M$^{-1}$, have been obtained for osmolyte-induced folding and urea-induced unfolding of over 40 proteins using seven different osmolytes, with correlation coefficient of 0.97 between the predicted and experimental m values. Accurate prediction over this large range provides persuasive evidence for the validity and utility of the Transfer Model.

**WHICH CONTRIBUTES MORE TO THE m VALUE, BACKBONE OR SIDE CHAINS?**

As described above, the Transfer Model provides a practical procedure to predict the osmolyte-induced m value of a protein, and in doing so, it partitions the overall free energy of transfer, $\Delta G_{N,U}$, into groupwise components. When $\Delta G_{N,U}$ is partitioned in this way, it becomes apparent that the backbone alone controls the $U \rightleftharpoons N$ folding transition, a conclusion that has far-reaching implications for the mechanism of protein folding, as is shown below. We begin by dissecting the m value for the Notch ankyrin domain, a representative protein (54).

**Dissection of Backbone and Side Chain m-Value Contributions: A Case Study**

Notch ankyrin domain m values in urea and TMAO have been measured by Mello & Barrick (55) and predicted using the Transfer Model. Experimental and predicted m values are shown in Table 1.

Applying the unfolded state model (described above in The Transfer Model: Identifying Driving Forces and Predicting $\Delta G_{U,N}$) to the Notch ankyrin domain, a representative protein (54), we can calculate the m value for this protein in urea and TMAO. The predicted m values, shown in Table 1, agree well with the experimental m values.

**Table 1** m values for ankyrin in kcal/mol M$^{-1}$

<table>
<thead>
<tr>
<th></th>
<th>Urea</th>
<th>Trimethylamine N-oxide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>−2.76</td>
<td>6.52</td>
</tr>
<tr>
<td>Tanford Transfer Model</td>
<td>−2.98</td>
<td>6.11</td>
</tr>
</tbody>
</table>
The difference in surface area between native and unfolded states, \( \Delta \text{ASA} \), is 15,084 \( \text{Å}^2 \), summed over both backbone and side chains. Groupwise changes in area and the resultant group transfer free energies are depicted graphically in Figure 2 for urea (panel a) and TMAO (panel b). In particular, the abscissa denotes \( \Delta \text{ASA} \), and the ordinate measures the component m value (\( \Delta \Delta g \) transfer) per unit surface area. Filled rectangles, colored by class, show the total free energy contribution of each respective component. By way of mental calibration, the free energy of transfer of a backbone unit from water to 1 M urea is \(-40\) cal/mol from Auton & Bolen (44), and the standard state area of the backbone, averaged over the 20 residue types, is \(-40\) \( \text{Å}^2 \) from Lesser & Rose (50). The resulting \( \Delta \Delta g \) transfer is \(-1\) cal/mol M\(^{-1}\)Å\(^2\), corresponding to the value of the backbone component indicated on the ordinate in panel a. By far, the largest free energy contributions to the m value are from the backbone, as described in the text.
for the backbone and each of the 18 types of side chains (Notch ankyrin domain lacks Cys residues and Gly lacks a side chain); the ordinate denotes the m value per unit surface area (kcal/mol M\(^{-1/2}/\text{Å}^2\)) contributed by these components upon transfer from water to osmolyte; and therefore, the filled rectangles (colored by polarity class) denote the total free energy contribution of each respective component.

Groupwise dissection reveals a remarkable fact: The backbone free energy contribution alone controls the direction of the U ⇔ N folding transition in both urea-induced unfolding and TMAO-induced folding. It is a striking realization that net side chain contributions oppose the overall direction of the reaction, disfavoring unfolding in urea and disfavoring folding in TMAO. Furthermore, hydrophobic groups (shown in Figure 2) contribute negligibly to the m value in either osmolyte.

Elaborating, the overall urea m value is the sum of both side chain and backbone contributions. The net free energy contributed by all side chains is small and positive; collectively, Notch ankyrin domain side chains oppose denaturation, though only slightly. In opposition, the free energy contributed by the newly exposed backbone is large and negative, overcoming side chain contributions and driving urea-induced unfolding (48, 56–60). Correspondingly, in the overall TMAO m value, the net free energy contributed by side chains is small and negative; in this case, side chains oppose folding but, again, only slightly. Here too, it is the free energy of the backbone, now large and positive, that overcomes side chain contributions and drives the equilibrium toward the native state.

In the general case, the backbone/side chain proportions of energies and areas in globular proteins resemble those in the Notch ankyrin domain (45, 46). As in the Notch ankyrin domain, the backbone contributes most of the free energy, but the side chains contribute most of the area. Specifically, side chains contribute about 75% of the newly exposed surface area, with the backbone contributing the remaining 25% (45, 46).

Proportionately, free energy contributions are the other way around. In the Notch ankyrin domain, the peptide backbone contributes −3.49 kcal/mol M\(^{-1}\) to the predicted urea m value, amounting to 117% of the net total, whereas side chains contribute +0.51 kcal/mol M\(^{-1}\), a −17% offset. In the corresponding case of TMAO, the peptide backbone contributes 8.06 kcal/mol M\(^{-1}\) to the m value, 132% of the total, and side chains contribute −1.95 kcal/mol M\(^{-1}\), a −32% offset. In either case, backbone contributions alone control the direction of the U ⇔ N transition.

We note in passing that data depicted in Figure 2 challenges the widely held proposition that the m value is highly correlated with ΔASA, the surface area newly exposed on urea-induced protein unfolding. Myers et al. (61) calculated that this relationship has a correlation coefficient of 0.84. However, the ostensible correlation is deceptive. In fact, the largest contributor to ΔASA on unfolding (e.g., side chains) correlates negatively with the m value; it is only the smaller ΔASA—but larger transfer free energy—of the peptide backbone that contributes positively to the urea m value and accounts for urea’s ability to unfold proteins. Correspondingly, side chain burial correlates negatively with the m value on TMAO-induced folding of the Notch ankyrin domain.

With regard to the mechanism of protein folding, the opposing contributions of backbone and side chains serve to resolve any ambiguity about which chemical groups control the folding reaction. In both urea-induced unfolding and TMAO-forced folding, the non-covalent interactions that control the U ⇔ N equilibrium are localized to the polar backbone, where hydrogen bonds are the most likely candidate involved in determining the relative populations of [N] and [U]. Accordingly, we turn now to the peptide hydrogen bond.
DOES THE PEPTIDE HYDROGEN BOND STABILIZE PROTEINS?

Few topics in biochemistry are as confused, as confusing, or as important as the contribution that hydrogen bonds make to protein stability. In his introduction to a recent monograph on this topic (62), Baldwin asserted, “…the drive for continued rapid progress in protein structure prediction, which requires a fuller understanding of protein-folding energetics, brings peptide H-bonds and peptide solvation into central focus.”

Undoubtedly, there is a favorable distance- and orientation-dependent interaction between an electro-negative atom and a hydrogen covalently bonded to another electro-negative atom (e.g., C=O···H–N). However, the comparison between a hydrogen bond to water in the unfolded polypeptide chain and a corresponding intramolecular hydrogen bond in the folded protein (i.e., –O–H···O = C + H–O···H–N ⇔ C = O···H–N + 2H₂O) has been highly controversial: Is it stronger, weaker, or equivalent? An energy-ledger approach to answering this question requires (a) quantitative consideration of the enthalpy of hydrogen bond formation in water (a high-dielectric medium) and in the protein (a lower-dielectric medium), and accurate assessment of (b) the desolvation penalty paid by a peptide unit upon transfer from the solvent-accessible aqueous environment to the solvent-shielded protein environment, (c) the changes in configurational entropy in both the polypeptide chain and in water molecules, (d) the conformational dependence of intramolecular hydrogen bond strength, and (e) hydrogen bond cooperativity. All are complex and elusive issues.

In this as in all venues, our mindset conditions our thinking. Seeking perspective, we first sketch a history of the hydrogen bonding in protein chemistry. Then, we attempt to identify—and possibly to resolve—some sources of confusion. Often, a zeal for energy-ledger clarity has led the field to disregard inferential evidence from stimulus-response experiments, as discussed above.

Selected Highlights from Seven Decades of Hydrogen-Bonding History

Contemporary ideas about hydrogen bonding date back to Bernal & Megaw (63) and Huggins (64) in the 1930s, but more than others, it was Pauling (65) who brought these concepts to the forefront. For a scholarly account, see Jeffrey & Saenger (66).

Guided by his ideas about the formative role of hydrogen bonding, Pauling & Corey proposed models for protein secondary structures, the α-helix (7) and β-sheet (67), arguing, “The energy of an N–H···O = C hydrogen bond is of the order of 8 kcal mol⁻¹, and such great instability would result from the failure to form these bonds that we may be confident of their presence” (7). See Eisenberg (68) for an insightful account of Pauling’s methods and discoveries. Soon after, Schellman (69) estimated the peptide hydrogen bond enthalpy in water to be approximately −1.5 kcal/mol, a value obtained by analyzing the thermodynamics of dilute urea solutions under the plausible assumption that deviations from ideality are caused by hydrogen-bonded dimerization. This early estimate is remarkably close to contemporary values (see below).

Kauzmann’s influential review (8) in 1959 overturned the existing mindset. Part III of his classic energy-ledger-motivated analysis of intramolecular forces argued persuasively that the hydrophobic effect drives protein folding. Also, see Dill’s later review (9). Kauzmann’s conclusions were soon corroborated by experimental results of Klotz & Franzen (70) and Susi et al. (71). Using the dimerization of N-methyl acetamide (NMA) monomers as a model for hydrogen bonding between peptide units, Klotz & Farnam (72) found that the hydrogen-bonding reaction in water is,
in fact, disfavored ($\Delta G^\circ = +3.1$ kcal/mol). Kauzmann’s analysis and Klotz’s measurements anchored the plausible idea that folding is driven largely by the burial of hydrophobic groups in the protein interior. These ideas were widely accepted, and they conditioned thinking in the field for decades to come.

If the hydrogen bond in water is disfavored, then one would expect that short, protein-length helices lack autonomous stability. Consistent with this expectation, helix propagation parameters (73) near unity (i.e., neither helix favoring nor helix disfavoring) were found in an experimental host-guest system (see Reference 74 and the references therein). A clear inference from these measurements was that 100 or more residues would be needed to stabilize a helix in water, well beyond the 12-residue average length of a protein helix. Yet, contrary to this expectation, Brown & Klee (75), using circular dichroism, demonstrated a measurable helix content in C-peptide, the cyanogen-bromide cleavage product of residues 1–13 of ribonuclease A. Importantly, residues 2–12 of C-peptide are helical in the intact protein (76), implying that forces that stabilize the helix in the protein resemble those at work in the isolated peptide. Regrettably, the impact of this paper was slow in coming: Of 254 total references to date, only 7 appeared in the first two years, and only 21 within the first five years.

If the peptide hydrogen bond in water is sufficiently favorable, then quasi-stable helices of protein-length should be detectable. Over much of the 1980s, Baldwin’s lab analyzed the stability and determinants of short helices in water (see Reference 77), starting with C-peptide (78) and culminating in alanine-based peptides (79). Among the naturally occurring amino acids, alanine has both the highest helix propensity (80–83) and the least-complicated chiral side chain, making polyalanine an attractive candidate for experimentation. However, pure polyalanine is insoluble in water, requiring alanine-based peptides to be punctuated by solubilizing polar residues and thereby complicating the analysis of helix-stabilizing factors (84).

Eventually, Kallenbach and coworkers (85) eliminated most of these complications by designing an uninterrupted 13-residue polyalanine sequence with charged residues relegated to flanking positions; this peptide is soluble in water, and it is helical. Other evidence, from binding studies involving tRNA synthetase, also demonstrated that hydrogen bonding is favorable by approximately $-1$ kcal/mol (86). Additionally, $\alpha$-helices often terminate in stabilizing capping motifs (87, 88), and these too are hydrogen-bonded structures (89).

Summarizing, it is now known that short polyalanyl peptides can populate helical conformations in water. In such a simple homopolymer, what energetic factors are available to buttress helical structure against the disorder engendered by conformational entropy (90)? The most likely explanation is an energetically favorable peptide hydrogen bond, although other explanations cannot be ruled out, including van der Waals and/or side chain interactions (91, 92). These issues are revisited in the following section.

In the 1990s, Scholtz et al. (93) and later Lopez et al. (94) determined the enthalpy change for the helix to coil transition of a 50-residue alanine-based peptide in water. If this measured enthalpy change is attributed to hydrogen bonding, then the peptide hydrogen bond in water is favorable, with an enthalpy approximating $-1$ kcal/mol, in good agreement with values obtained by Fersht (86) and, much earlier, by Schellman (69). Meanwhile, in an extensive series of mutations, Pace and coworkers (13, 95) replaced dozens of polar residues with apolar residues of similar size and shape (e.g., Tyr $\rightarrow$ Phe, Thr $\rightarrow$ Val) and measured the conformational free energy differences, $\Delta \Delta G_{U \rightarrow N}$, between the protein with the polar residue and its apolar counterpart after suitable correction for conformational entropy, volume changes, polar groups stranded without partners, and the like. From this series, it can be concluded that buried hydrogen bonds stabilize proteins substantially, by as much as $-3$ kcal/mol in some cases.
Measurements of hydrogen bond strength over the decades prior to 2000 were extracted from systems in which the possibility of cryptic variables could not be easily eliminated. Ideally, one would like the means to selectively toggle a hydrogen bond in an experimental system without perturbing any other aspect of that system. That ideal could only be realized in computational systems, which come with their own set of inherent uncertainties. However, experiments using time-resolved fluorescence resonance energy transfer (FRET) are now providing a less ambiguous way to quantify intramolecular interactions, e.g., Reference 96. Recently, Kiefhaber and coworkers (97) were able to detect nonspecific, intramolecular hydrogen bonds in an unfolded polypeptide chain using FRET. In their study, the unfolded state was modeled by a long Gly-Ser polymer, too flexible to adopt a folded conformation. In this model system, intramolecular hydrogen bonds form in water and break upon addition of GdmCl, a good solvent. This work provides solid evidence that intramolecular hydrogen bond formation is governed by solvent quality, and that, in comparison, water is a poor solvent. The same conclusion was recently drawn for the collapsed conformation of monomeric polyglutamine in water (98).

Lingering Doubts and Alternative Explanations

Against this historical backdrop, it now seems likely that, in fact, the peptide hydrogen bond is somewhat stabilizing in water. Yet, arguments to the contrary persist, many of them tracking back to Klotz and coworkers (70, 72, 99), who used the dimerization of NMA monomers as a model system for hydrogen bonding. Among these arguments, a common denominator is the use of Klotz’s thermodynamic cycle (Scheme 2), which compares the solvation of a donor (D) and acceptor (A) in water (Dw, Aw) and in a nonpolar solvent (Dn, An) with the hydrogen-bonded donor-acceptor pair in the same medium (Dw⋯Aw and Dn⋯An, respectively).

In an influential paper, Roseman (100) proposed a correction to Klotz & Farnham (72), used it to reanalyze their thermodynamic cycle, and set the stage for later conclusions by others (e.g., 9–11). Taking CCl4 as the nonpolar solvent and values from Klotz & Farnham (72), the measured free energies in Scheme 2 are as follows: \[ \Delta G_1 = +2.4 \text{ kcal/mol}, \]
\[ \Delta G_2 = -6.12 \text{ kcal/mol (Roseman’s corrected value)}, \]
\[ \Delta G_3 = +3.1 \text{ kcal/mol}, \]
\[ \Delta G_4 = +0.62 \text{ kcal/mol}. \]

Consequently, transfer of the hydrogen-bonded species between water and nonpolar solvent—or by inference, the interior of a protein—is essentially neutral, or even slightly disfavored, as Roseman points out.

The largest free energies in this cycle involve desolvation penalties: \(-\Delta G_2\) measures the cost of drying up two polar groups upon transfer to nonpolar solvent, and its magnitude rivals the overall free energy, \(\Delta G_{\text{U→N}}\), of a typical protein. Although smaller in magnitude, \(\Delta G_1\) measures a desolvation penalty that is nevertheless sufficient to strongly disfavor the hydrogen-bonded species in water. Taken at face value, these substantial desolvation penalties...
penalties indicate that peptide hydrogen bond formation in water is substantially disfavored (9–11).

How can the observed stability of isolated polyalanyl helices in water be reconciled with a desolvation penalty that exceeds the hydrogen bond energy? Explanations invoke stabilizing interactions from hydrophobic side chains and/or stabilizing dispersion forces within the close-packed, hydrogen-bonded helical core. The former explanation seems unlikely, especially for a polyalanyl helix. Relative to the coil state, a Cβ atom buries little, if any, surface upon helix formation, and longer side chains, forced to protrude from the bulky backbone cylinder, can actually gain surface (101). Furthermore, the distance between any two Cβ atoms exceeds 5.4 Å at the distance of closest approach, a separation of almost two water diameters, precluding the possibility of hydrophobic burial between or among them.

An N–H···O = C hydrogen bond favors an N to O distance that is closer than the sum of their van der Waals radii. Consequently, elements of hydrogen-bonded secondary structure are expected to be tightly packed, and detailed analysis of protein structure confirms this expectation (102). However, tight packing would serve to stabilize the folded state only if it exceeds corresponding protein:water packing interactions in the unfolded state. Whether or not this is the case remains an open question. With the benefit of hindsight, it is clear that conclusions about hydrogen bonding modeled on NMA are misleading for many reasons, some of which are now discussed.

1. Owing to well-known end-group effects in polymers, small molecules, such as NMA, are unsuitable models for the solvation energy of an internal peptide backbone unit in a longer polypeptide. Avbelj et al. (103) drive this point home in figures 1 and 2 of their paper, and they state, “...amides are not close models for the interaction of the peptide group with water. The ESF (electrostatic solvation free energy) value for N-methylacetamide differs by $-0.4$ kcal/mol from the value for an interior alanine peptide group.” Avbelj et al.’s conclusions are based on calculations. An equivalent point, on the basis of measurements, is made for short peptides by Auton & Bolen in figure 3 of their paper (44), and they state “For peptides with small chain lengths, acetamide and N-acetylglycinamide, the solubility is high, resulting in large differences in $\Delta G_{\text{int}}$ between concentration scales, while at longer chain lengths, the lower solubility of the peptide causes $\Delta G_{\text{int}}$ values to converge.”

2. Protein–water interaction energy is conformation dependent in polypeptide chains (104–111) but not in rigid molecules like NMA. In grand canonical ensemble Monte Carlo simulations using explicit solvent, Mezei et al. (105) computed the per residue energy difference between left-handed polyproline II (PII) and a β-strand to be 0.7 kcal/mol, approximating physiological RT (where R is the gas constant, and T is the temperature in degrees Kelvin). Both PII and the β-strand lack intramolecular hydrogen bonds. Yet, changes in backbone dihedral angles, from $\phi, \psi = (-139^\circ, 135^\circ)$ to $(-78^\circ, 149^\circ)$, were found to make a 7 kcal/mol difference in solvation free energy for the central 10 residues in a longer polyalanyl peptide (105).

3. Unlike NMA dimers, hydrogen bond energies in longer polymers can be highly cooperative. In his review of cooperative interactions (112), Dannenberg points out that the length and strength of hydrogen bonds in an α-helix are a function of residue position: As chain length increases, bond length decreases, and bond strength increases. Dannenberg reports that the enthalpy of adding an alanine to the α-helix,
ΔH, can be as much as 4.2 kcal/mol. His findings are from calculations using density functional theory, with the helix treated as a solid and the coil as a liquid in the helix ⇀↽ coil equilibrium. In yet another recent density functional theory calculation, Baker and colleagues (113) report a high degree of hydrogen bond cooperativity in amyloid fiber formation, with an astonishing hydrogen bond energy of −9.1 kcal/mol.

4. NMA molecules in solution are freely diffusible, and their dimerization would come at a higher entropy penalty than an intrachain N–H⋯O = C hydrogen bond between partners for which the loop-closing entropy is facilitated by covalent constraints (114).

5. The positive sign of ΔG in Scheme 2 is derived from the other legs in the thermodynamic cycle and involves the small difference between substantially larger measured quantities. Unavoidable experimental error could easily result in a change of sign from positive (i.e., unfavorable) to negative (i.e., favorable) (25, 26).

6. The measured free energies for NMA are based on concentration differences, uncorrected for activity coefficients, a nontrivial issue in many cases, e.g., Reference 46.

It often happens that newly introduced models are carefully qualified by their originators; then, with the passage of time, the models persist but the cautions fade. The use of NMA dimerization as a model for peptide hydrogen bond formation in water is subject to this caveat. Essentially, the issues and arguments raised in this section refocus the discussion on one central question: Is water a good solvent or a poor solvent for backbone polar groups?

THE SOLVENT QUALITY OF WATER

For polypeptide chains, water is neither as good a solvent as urea nor as poor a solvent as TMAO. This comparison serves to bracket water’s relative position on a solvent quality scale (Figure 3) but fails to resolve its absolute position.

Indeed, most thermodynamic quantities in experimental systems are inaccessible as absolute energy-ledger values. Instead, these quantities are typically calibrated as differences in a stimulus-response-type paradigm. Often, one obtains a ΔG_{N→U}^f for the system of interest and a corresponding ΔG_{N→U}^f for a perturbed version of the system (where the prime designates the perturbed system). Then ΔΔG = ΔG_{N→U}^f − ΔG_{N→U}^f measures the degree to which the perturbation either stabilizes or destabilizes the original system.

The m value, as determined by the free energy difference ΔG_{N→U}^f − ΔG_{N→U}^f, is such a ΔΔG; it measures the degree to which added cosolvents dial solvent quality either up or down. Using Scheme 1, m values obtained in this way have been dissected, showing that osmolytes operate predominantly on the
A broad panel of osmolyte cosolvents, ranging in solvent quality from TMAO to urea, can be described by a single backbone-dependent, hydrogen-bonding mechanism. Water falls within this solvent-quality continuum, and it too should operate according to this same mechanism.

A helix of average length in proteins is approximately 12 residues. At this length, short polyalanyl-based peptides populate helical conformations in water. This observation indicates that, by definition, water is a poor solvent relative to a neutral solvent in which protein:solvent interactions are neither favored nor disfavored. An implicit assumption undergirds this conclusion: Essentially, all backbone polar groups form hydrogen bonds, either to solvent or to protein partners, just as Pauling et al. anticipated. A completely unsatisfied backbone polar group would come at a free energy cost of a magnitude that rivals $\Delta G_{U \rightarrow N}$ for a typical protein.

Even low-complexity polypeptide chains, too soluble and too flexible to promote helix formation, nevertheless form nonspecific intramolecular hydrogen bonds in water, as described above. Such intrachain hydrogen bonding is unencumbered by any of the energetic complexity that might be thought to accompany helix formation. Using fluorescence correlation spectroscopy, Pappu and colleagues determined that polyglutamine, another low-complexity polymer, forms collapsed structures in water, despite the absence of any hydrophobic groups.

These results, and many others, provide clear evidence that water is a comparatively poor solvent for polypeptide chains. Why is this so? In $\beta$-strand or $\alpha$H conformation, the chain has a large negative enthalpy of solvation, an ostensible indication that water is, in fact, a good solvent. Yet, under folding conditions in water, the chain also visits other conformations for which the solvation energy is much less favorable, as described above. Osmolytes operate on this equilibrium distribution between conformers that favor self-interaction and those that favor chain:solvent interaction, dialing it either up or down. To understand these thermodynamic issues at a molecular level, the interactions among water, osmolyte, and the polypeptide chain must be examined in detail.

We begin by reassessing the number of water molecules that are stripped away upon chain desolvation.

### How Solvated Is the Peptide Unit in Water?

The thermodynamics of protein solvation has long been a topic of keen interest. In numerous studies of peptide and protein folding, it is often assumed, usually implicitly, that each peptide unit in the unfolded state is hydrogen-bonded to three water molecules, one to the amide hydrogen and one to each carbonyl oxygen lone pair. This assumption has persisted over many decades and has anchored numerous assessments of the desolvation energy. Is it valid?

Conclusions about hydration of the peptide unit are often based on simple models, such as NMA (see above) or small-molecule crystal structures, which can be dominated by nonadditive end effects. Water molecules at polypeptide termini have a larger effective interaction volume than they do at interior peptide groups, where covalent chain connectivity imposes additional excluded volume restrictions. Enhanced solvation at the chain termini is apparent in both measurements and calculations, suggesting that a polypeptide chain will experience partial dewetting of its interior peptide groups. If so, an assumption of three hydrogen-bonded water molecules per peptide unit would lead to an overestimate of the desolvation penalty.

Furthermore, given that water is a somewhat poor solvent for polypeptide chains, intramolecular hydrogen bonds are expected to compete with solvent hydrogen bonds, stripping some water from the backbone. The calculated difference in solvation free energy...
energy between a residue in an isolated β-strand versus an isolated polyproline helix is approximately 0.7 kcal/mol (103, 105) (as described above in Lingering Doubts and Alternative Explanations). Clearly, this energy difference is not large enough to trap a conformer in the PII basin exclusively (125), and with ambient temperature fluctuations, the chain would visit other alternatives readily (106, 126, 127), adjusting its hydration state accordingly.

A SIMPLE STRUCTURAL ORIGIN FOR OBSERVED OSMOLYTE THERMODYNAMICS

At long last, Tanford’s Transfer Model (30, 31) has been validated by successful prediction of m values for both protecting osmolytes (45) and urea (46)—a stringent test. Using the Transfer Model, the $\Delta G_v$ of protein solutes between buffer and osmolyte-containing solvents can be reliably dissected into group-wise components, $\Delta g_v$ (45). Equally, the same, self-consistent thermodynamic framework holds for urea-induced denaturation, as shown recently (46). Thus, the Transfer Model accounts for protecting and denaturing effects on protein stability, both lying on the same solvent-quality continuum, with peptide hydrogen bonding as the dominant variable.

This thermodynamic framework lends itself to a straightforward structural interpretation: Protecting osmolytes, such as TMAO, promote intramolecular hydrogen bonding in the unfolded protein. Several early folding models are based on explicit intramolecular hydrogen bonding in unfolded states, e.g., References 128–130. In common among these models, flickering elements of secondary structure initiate a native-like scaffold within the denatured species (see References 131–133). In accord with these ideas, incipient hydrogen bonding can be observed experimentally (97), and it is known that TMAO promotes helix formation in alanine-based peptides (134). In contrast, denaturing osmolytes, like urea, work in the opposite way, disrupting incipient intramolecular hydrogen bonding and suppressing secondary structure formation; although even in this case, evidence of incipient secondary structure in the presence of high denaturant concentrations has been reported (135).

Extending previous models, we hypothesize that the molecular origin of the osmolyte effect can be ascribed to backbone hydrogen bonding (20, 116). In brief, the addition of a protecting osmolyte reduces solvent quality for the backbone, decreasing the equilibrium population of solvent:backbone hydrogen bonds in U, but not in N, and raising the free energy of U relative to N accordingly. Given the high energetic cost of even one completely unsatisfied hydrogen bond (119), loss of hydrogen-bonding capacity in U would shift the $U \rightleftharpoons N$ equilibrium markedly toward N, where almost all backbone polar groups are satisfied within units of local, hydrogen-bonded secondary structure: α-helix, β-sheet, $\beta_{10}$-helix, and β-turns (136). To a first approximation, TMAO simply dial down the number of solvent:backbone hydrogen bonds, relative to buffer (see Figure 1). Reciprocally, urea dials up the number of solvent:backbone hydrogen bonds, again relative to buffer. This reciprocal relationship—the osmolyte hypothesis for protein folding—establishes the fundamental link between solvent thermodynamics and the protein’s hydrogen-bonded backbone structure.

Anfinsen’s thermodynamic hypothesis states that “the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, and other) is the one in which the Gibbs free energy of the whole system is the lowest” (2). The thermodynamic hypothesis posits a link between free energy and native structure, and using the Transfer Model, the osmolyte hypothesis describes that link in quantitative detail.

It is important to emphasize that shifting hydrogen bonding between U and N is an
exchange reaction, i.e., solvent hydrogen bonds in U are exchanged for peptide hydrogen bonds in N. No covalent bonds are made or broken in this reaction, individual hydrogen bonds need not be intrinsically strengthened or weakened, and the structure of water is not altered in any significant way (59, 116). Although cooperative strengthening of hydrogen bonds may further stabilize secondary structures (112), the basic exchange reaction does not depend upon this. If it did, one would not expect the percentage of residues involved in hydrogen-bonded \( \beta \)-turns to exceed that of residues in \( \alpha \)-helices, as it does in globular proteins (137).

The structure-promoting efficacy of backbone hydrogen bonding is especially conspicuous in the forced folding of intrinsically disordered proteins. To date, at least six proteins have been studied (Table 2), some of particular physiological relevance (e.g., \( \alpha \)-synuclein, AF1, and RNase P protein), in which the addition of a protecting osmolyte can shift the population from a disordered ensemble to the native one. Clearly with such proteins, the hydrophobic interaction, although in water and under usual folding conditions, is nevertheless insufficient to stabilize the folded form (138). But, upon addition of TMAO, which promotes folding by increasing polar backbone interactions, the folded population emerges spontaneously despite modest energetic opposition from side chain burial.

Proteins are poised between order and disorder. Under folding conditions, the stabilization free energy, \( \Delta G_{U \rightarrow N} \), of a typical globular protein ranges between \(-5 \) to \(-15 \) kcal/mol (139), the energetic equivalent of one or two water:water hydrogen bonds. Consequently, small changes in energy can leverage substantial changes in chain organization. Osmolytes work in exactly this way. According to table 2 in Reference 44, the free energy of transferring a backbone unit from water to 1 M urea is favored by 40 cal/mol. This seemingly modest contribution amounts to \(-4 \) kcal/mol when summed over 100 residues in 1 M urea, approaching the total \( \Delta G_{U \rightarrow N} \) of a characteristic protein. With the urea concentration further increased to typical denaturing conditions (e.g., 6 M), destabilization becomes overwhelming.

On consideration, it is apparent that hydrogen bonding acts as the conformational pivot in the \( U \rightleftharpoons N \) equilibrium. On one side of the balance, the decrease in conformational entropy that accompanies the folded conformation favors the U state under all conditions. On the opposing side of the balance, the sequestering of apolar residues from water access favors the N state under all conditions. It is only hydrogen bonding that pivots between unfolding and folding conditions with changes in solvent quality from osmolyte, and this shifting balance can be monitored experimentally using hydrogen exchange and NMR (140–142).

We conclude this review with the recognition that the osmolyte effect is universal throughout all three kingdoms of life (15). A broad repertoire of biologically active osmolytes has been assembled via natural selection, enabling each organism to select for variants that are best suited to its cellular microenvironment and external conditions (143). Since Darwin, we have come to regard macroscopic characteristics, such as organelles or opposable thumbs, as the province of evolutionary biology. But osmolyte adaptation shows that natural selection is also at work on a strictly physicochemical level as well (15). Nature has been practicing successful bioengineering since the beginning of life on Earth.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )-Synuclein</td>
<td>144</td>
</tr>
<tr>
<td>T62P staphylococcal nuclease</td>
<td>37</td>
</tr>
<tr>
<td>Reduced carboxyamidated RNase T1</td>
<td>37</td>
</tr>
<tr>
<td>Truncated notch ankyrin protein</td>
<td>145</td>
</tr>
<tr>
<td>Glucocorticoid receptor AF1 domain</td>
<td>146</td>
</tr>
<tr>
<td>RNase P protein</td>
<td>147</td>
</tr>
</tbody>
</table>
SUMMARY POINTS

1. The relationship between the thermodynamic forces responsible for protein folding and the spontaneous emergence of protein structure is still not well understood at the molecular level. In this review, we focus on the peptide hydrogen bond and its role in protein folding. Our analysis reveals that the transition between the unfolded and folded populations, $U(\text{unfolded}) \rightleftharpoons N(\text{native})$, is mediated predominantly via peptide hydrogen bonding.

2. Organic osmolytes, found throughout nature, modulate protein folding. The denaturing osmolyte urea shifts the $U \rightleftharpoons N$ equilibrium toward $U$, and protecting osmolytes shift the equilibrium toward $N$. This universal osmolyte effect provides essential clues to the nature of the folding reaction.

3. The Tanford Transfer Model is used to dissect stability differences between a protein in an osmolyte solution and in buffer alone. Specifically, the Transfer Model quantifies the degree to which an osmolyte either stabilizes or destabilizes the protein relative to buffer, and it partitions the overall stability difference into energetic contributions from different side chains and the backbone.

4. From the Transfer Model, the free energy differences that control the folding reaction are contributed overwhelmingly by the backbone, with only small contributions from side chains and negligible contributions from hydrophobic groups. The exclusive role of the peptide backbone in the energetics of osmolyte-induced folding and unfolding identifies peptide hydrogen bonding as the control point in these processes.

5. Conclusions in the literature about the energetics of peptide hydrogen bonding have often been based on simple monomers, which turn out to be misleading models.

6. Aqueous buffer alone is a somewhat poor solvent for the protein backbone. As a result, intramolecular peptide hydrogen bonds are marginally favored over backbone:solvent hydrogen bonds in water. Addition of protecting osmolytes, such as TMAO, further diminish water solvent quality, increasing the population of peptide backbone hydrogen-bonded species. Conversely, addition of urea increases solvent quality, favoring protein:solvent hydrogen bonds at the expense of intramolecular hydrogen bonds. In either case, hydrogen-bonding is pivotal in shifting the $N \rightleftharpoons U$ equilibrium.

7. The repertoire of sterically accessible hydrogen-bonded scaffold elements in proteins is limited largely to $\alpha$-helices, $\beta$-sheets, and $\beta$-turns, and all globular proteins are built upon such scaffolds. Consequently, solvent quality that disfavors backbone:water hydrogen bonds necessarily favors the hydrogen-bonded scaffold, thereby promoting protein structure. Shifting hydrogen bonding between $U$ and $N$ is an exchange reaction in which solvent:protein hydrogen bonds in $U$ are exchanged for peptide hydrogen bonds in $N$. This reciprocal relationship establishes the fundamental link between protein thermodynamics and the protein’s hydrogen-bonded backbone structure.

FUTURE ISSUES

1. Where does aqueous buffer fall on the solvent quality scale? Is water a good solvent or a poor solvent for the protein backbone?
2. How solvated is the peptide unit under unfolding conditions? Many studies make an implicit assumption that each peptide unit in the unfolded state is hydrogen bonded to three water molecules, one to the amide hydrogen and one to each carbonyl oxygen lone pair. Is this assumption correct? If not, how does solvation vary with peptide chain length?

3. How organized is the polypeptide backbone under unfolding conditions of interest? What fraction of the backbone participates in intramolecular hydrogen bonds under unfolding conditions, and how does this fraction vary with solvent quality?

4. To what extent is hydrogen bonding a paramount driving force in protein folding?

5. What is the relationship between osmolyte-induced protein stabilization/destabilization and the past 70 years of classical protein folding studies?

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Buzz Baldwin, Neville Kallenbach, and Jörg Rösgen for many formative discussions; Neville Kallenbach and Jörg Rösgen for reading the manuscript, and Buzz Baldwin for sending his review prior to publication. Support from the NIH GM49760 (D.W.B.) and the Mathers Foundation (G.D.R.) is gratefully acknowledged.

LITERATURE CITED

## Contents

### Prefatory Chapters

Discovery of G Protein Signaling  
*Zvi Selinger* ................................................................. 1

Moments of Discovery  
*Paul Berg* ................................................................. 14

### Single-Molecule Theme

*In singulo* Biochemistry: When Less Is More  
*Carlos Bustamante* .......................................................... 45

Advances in Single-Molecule Fluorescence Methods  
for Molecular Biology  
*Chirlmin Joo, Hamza Balci, Yuji Ishitsuka, Chittanon Buranachai, and Taeckjip Ha* .......................................................... 51

How RNA Unfolds and Refolds  
*Pan T.X. Li, Jeffrey Vieregg, and Ignacio Tinoco, Jr.* .......................................................... 77

Single-Molecule Studies of Protein Folding  
*Alessandro Borgia, Philip M. Williams, and Jane Clarke* .......................................................... 101

Structure and Mechanics of Membrane Proteins  
*Andreas Engel and Hermann E. Gaub* .......................................................... 127

Single-Molecule Studies of RNA Polymerase: Motoring Along  
*Kristina M. Herbert, William J. Greenleaf, and Steven M. Block* .......................................................... 149

Translation at the Single-Molecule Level  
*R. Andrew Marshall, Colin Echeverría Aitken, Magdalena Dorywalska, and Joseph D. Puglisi* .......................................................... 177

Recent Advances in Optical Tweezers  
*Jeffrey R. Moffitt, Yann R. Chemla, Steven B. Smith, and Carlos Bustamante* .......................................................... 205

### Recent Advances in Biochemistry

Mechanism of Eukaryotic Homologous Recombination  
*Joseph San Filippo, Patrick Sung, and Hannah Klein* .......................................................... 229
Structural and Functional Relationships of the XPF/MUS81 Family of Proteins
Alberto Ciccia, Neil McDonald, and Stephen C. West ........................................ 259

Fat and Beyond: The Diverse Biology of PPARγ
Peter Tontonoz and Bruce M. Spiegelman .............................................................. 289

Eukaryotic DNA Ligases: Structural and Functional Insights
Tom Ellenberger and Alan E. Tomkinson ............................................................... 313

Structure and Energetics of the Hydrogen-Bonded Backbone in Protein Folding
D. Wayne Bolen and George D. Rose ........................................................................ 339

Macromolecular Modeling with Rosetta
Rhiju Das and David Baker ..................................................................................... 363

Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry
Benjamin F. Cravatt, Aaron T. Wright, and John W. Kozarich ............................... 383

Analyzing Protein Interaction Networks Using Structural Information
Christina Kiel, Pedro Beltrao, and Luis Serrano ....................................................... 415

Integrating Diverse Data for Structure Determination of Macromolecular Assemblies
Frank Alber, Friedrich Förster, Dmitry Korkin, Maya Topf, and Andrej Sali .......... 443

From the Determination of Complex Reaction Mechanisms to Systems Biology
John Ross .................................................................................................................. 479

Biochemistry and Physiology of Mammalian Secreted Phospholipases A2
Gérard Lambeau and Michael H. Gelb .................................................................. 495

Glycosyltransferases: Structures, Functions, and Mechanisms
L.L. Lairson, B. Henrissat, G.J. Davies, and S.G. Withers ................................. 521

Structural Biology of the Tumor Suppressor p53
Andreas C. Joerger and Alan R. Fersht .................................................................. 557

Toward a Biomechanical Understanding of Whole Bacterial Cells
Dylan M. Morris and Grant J. Jensen ..................................................................... 583

How Does Synaptotagmin Trigger Neurotransmitter Release?
Edwin R. Chapman ................................................................................................. 615

Protein Translocation Across the Bacterial Cytoplasmic Membrane
Arnold J.M. Driessen and Nico Nouwen ............................................................... 643
Maturation of Iron-Sulfur Proteins in Eukaryotes: Mechanisms, Connected Processes, and Diseases
Roland Lill and Ulrich Mühlenhoff ................................................................. 669

CFTR Function and Prospects for Therapy
John R. Riordan ..................................................................................................... 701

Aging and Survival: The Genetics of Life Span Extension by Dietary Restriction
William Mair and Andrew Dillin ........................................................................... 727

Cellular Defenses against Superoxide and Hydrogen Peroxide
James A. Imlay ...................................................................................................... 755

Toward a Control Theory Analysis of Aging
Michael P. Murphy and Linda Partridge .............................................................. 777

Indexes
Cumulative Index of Contributing Authors, Volumes 73–77 .............................. 799
Cumulative Index of Chapter Titles, Volumes 73–77 ........................................... 803

Errata
An online log of corrections to *Annual Review of Biochemistry* articles may be found at http://biochem.annualreviews.org/errata.shtml