Annu. Rev. Biophys. Biomol. Struct. 1993. 22:381–415

HYDROGEN BONDING, HYDROPHOBICITY, PACKING, AND PROTEIN FOLDING

George D. Rose

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, Box 8231, 660 South Euclid Avenue, St. Louis, Missouri 63110

Richard Wolfenden

Department of Biochemistry and Biophysics, CB # 7260 Faculty Laboratory Office Building, University of North Carolina, Chapel Hill, North Carolina 27599–7260

KEY WORDS: secondary structure, protein engineering, solvent entropy, solute entropy, stereochemical code

---

CONTENTS

1056–8700/93/0610–0381$02.00

PERSPECTIVES AND OVERVIEW

The literature abounds with arguments about whether a protein hydrogen bond is stabilizing or destabilizing. Controversy also surrounds the hydrophobic effect. Long thought to be the principal stabilizing force in folded proteins, it recently was reexamined and found to oppose folding; meanwhile other studies reached an opposite conclusion, finding that hydrophobicity is even more stabilizing than previously believed. In the case of packing, the ubiquitous jigsaw puzzle–like fit of side chains in the protein interior suggests structural restriction and close tolerances, whereas recent mutational studies foster a contrasting view of structural malleability and forgiving tolerances.

The dialectic is often heated, as familiar premises are questioned and alternative ones proposed. The purpose of this review is to delineate these topical issues and perhaps contribute to their resolution. First, the protein-folding problem is defined and placed in historical perspective.

The Protein-Folding Problem

A protein molecule adopts its native three-dimensional structure spontaneously under normal physiological conditions (2). Despite intense research, a general mechanistic understanding of the folding transition remains obscure; this important problem is called the protein-folding problem.

The protein-folding problem was recognized clearly at least half a century ago in work of Anson & Mirsky, who observed that denaturation is a reversible process (4), and Mirsky & Pauling (108), who hypothesized that native proteins have a characteristic structure that is abolished upon denaturation. Such observations culminated ultimately in the experiments of Anfinsen and coworkers (3), who showed that reduced, denatured ribonuclease will renature spontaneously in vitro, with full restoration of enzymatic activity and return of its four native disulfide bridges. Such work has led to the contemporary view that protein tertiary structure is dictated by the amino acid sequence, although molecular chaperones may influence the folding kinetics in some cases (58).

A solution to the folding problem remains one of the principal challenges of Twentieth Century chemistry and biology. Many recent (1, 27, 34, 36, 49, 60, 77, 78, 111, 138, 152) and earlier (74, 76, 137) reviews have been written on these topics.

Stability of the Native Structure—the Evolving View

Current perspectives in protein folding have been conditioned by ideas developed during the preceding several decades of research. The driving

force for folding was initially thought to be intramolecular hydrogen bonding. This was Pauling's view, and it led him to the crucial role ascribed to hydrogen bonding in model structures for α-helix and β-sheet (121, 122). Such models, which were constrained by the known geometry of the peptide unit (101), sought to optimize both the number and the geometry of $\rangle$N-H$\cdots$O=C$\langle$ hydrogen bonds. The existence of these predicted structures was soon confirmed in ongoing X-ray studies of proteins (124), bolstering the assumption that hydrogen bonds play the formative role in folding and stability.

Less than a decade later, this point of view was to change dramatically. In a seminal review, Kauzmann, reasoning from model compounds, showed that burial of apolar groups must be a significant source of the stabilization energy in proteins (74). It had long been observed that liquid water dissolves polar substances readily but apolar substances only sparingly (168). Upon mixing, water squeezes out hydrophobic molecules, resulting in a segregation into polar and nonpolar phases. The spontaneous separation of oil and water after mixing is a familiar example of this phenomenon. Thus, that hydrophobicity plays a key role in organizing the self-assembly of protein molecules is entirely plausible because some amino acid residues are abundantly water soluble while others are only sparingly so (167, 180).

Kauzmann's observations have evolved into the contemporary textbook view that the hydrophobic effect serves as the driving force for protein folding. According to this popular idea (referred to as the oil-drop model), the protein interior is enriched in apolar (oily) residues that are expelled from water and engender, in effect, a separate organic phase (101a, 175). However, protein folding cannot be simply a matter of burying apolar residues while exposing polar ones. Even residues with hydrophobic side chains have pronounced hydrogen-bonding capacity because of the presence of backbone $\rangle$N-H and $\rangle$C=O groups. Were such residues unable to realize hydrogen bonds within the molecular interior, then hydrogen bonding would favor denaturation, since presumably these same groups could hydrogen bond readily to water in the unfolded state. For this reason, hydrogen bonding came to be regarded as energetically neutral, or even unfavorable (79, 166), with respect to folding, with intramolecular hydrogen bonds in the native structure supplanted by intermolecular hydrogen bonds in the unfolded state. A neutral or unfavorable hydrogen bonding balance sheet leaves the hydrophobic effect as the presumed driving force toward the folded state.

The hydrophobic effect alone seems insufficient to account for the existence of a unique equilibrium structure in proteins because aggregates of oil, which form spontaneously in water, lack specific internal architecture.

Specific packing interactions within the molecular interior are believed to be a major source of structural specificity in proteins (136). Richards (161) showed that globular proteins are packed, on average, as well as crystals of small organic molecules, with packing densities that are more reminiscent of solids than of oil. The inside of a typical protein contains side chains that fit together with striking complementarity, like pieces of a three-dimensional jigsaw puzzle.

The high packing densities seen in globular proteins have been interpreted to mean that protein conformation is linked tightly to internal packing. Thus, for example, lysozyme does not have the same folded conformation as ribonuclease, although both proteins have approximately the same size and composition, because the lysozyme sequence cannot achieve efficient internal packing when organized into a ribonuclease fold. This interpretation of packing is consistent with classical studies of protein evolution in which the most conserved residues are found in the buried interior (153).

In summary, the hydrophobic effect came to be viewed as the principal force that drives the protein toward globular collapse and engenders a solvent-shielded molecular interior. Within this interior, specific packing interactions are thought to be the primary determinant of structural specificity.

Scope and Purpose of this Review

Many recent developments have appeared in each of these areas: hydrogen bonding, the hydrophobic effect, and packing. One purpose of this review is to assess whether and how our understanding of these issues has changed during the 15 years since Richards' influential review in this series (137). Also, the final section introduces a novel model for protein folding that is consistent with data from earlier sections.

H-BONDS

The importance of the hydrogen bond was made apparent in Pauling's text (120) and in the early treatise of Pimentel & McClellan (125). Of more recent interest is the monograph by Jeffrey & Saenger (71), which emphasizes the role of hydrogen bonding in biological systems.

In proteins, the patterns and principles of hydrogen bonding have been analyzed by surveying X-ray-elucidated molecules to identify recurrent themes. Baker & Hubbard used such an approach in an influential review (8) that laid the groundwork for many later studies. Hydrogen-bond geometry in proteins has been analyzed in both individual molecules (7) and systematic surveys (8, 10, 66, 71, 160, 163). Small-molecule crystal

structures have also provided a wealth of information about H-bond geometry (23, 70, 110, 169, 170, 174).

How Strong Are H-Bonds Between Polar Groups in Proteins?

The thermodynamics of hydrogen bonding in proteins have been controversial and remain so. Since early work of Schellman (151), Susi et al (166), and Klotz & Franzen (79), studies have found that the stability of hydrogen bonds between polar groups in water is marginal, at best. Thus, it came as a surprise when Privalov & Gill, in a recent review (130), concluded that hydrogen bonding is the principal source of stabilization energy in folded proteins. Earlier studies by Fersht and coworkers had reported favorable H-bond energies in protein engineering experiments, with values ranging from -0.5 to -1.8 kcal/mol for polar partners and -3.5 to -4.5 kcal/mol when one of the partners bears a charge (52). More recently, Williams reckoned the amide-amide H-bond to be favorable by -1 to -4 kcal/mol (38) by analyzing the binding between antibiotics (e.g. vancomycin and ristocetin A) and suitable peptides. Also, Pace and coworkers (158), using protein engineering, found that, on average, an H-bond contributes -1.3 kcal/mol to the stability of the protein ribonuclease T1.

The lack of suitable model compounds can often confound thermodynamic analysis. In particular, the dimerization of two independent, freely diffusing molecules A and B to AB is not entropically analogous to formation of an intramolecular H-bond within a protein. According to Stahl & Jencks (162, p. 4201):

> This does not mean that hydrogen bonding is not an important driving force for the maintenance of the native structure of proteins and nucleic acids. Most of the hydrogen bonds in proteins and other macromolecules are formed in intramolecular reactions, so that the equilibrium constants for hydrogen-bond formation can be much more favorable than for bimolecular reactions because of the smaller loss of translational and rotational entropy in intramolecular reactions.

How strong are H-bonds between polar groups? In proteins, the inherent strengths of individual intramolecular H-bonds are modulated by direct competition with water molecules that supply alternative binding partners and by other polar groups in the surroundings. For estimates of the intrinsic strengths of individual H-bonds that are unperturbed by these additional effects, the only uncomplicated reference state is the dilute vapor phase, in which intermolecular interactions are absent. In the vapor phase, mass spectroscopic methods show that groups carrying a single positive or negative charge form single H-bonds to water molecules with the release of 16–23 kcal/mol of enthalpy (105). Second virial coefficients indicate that between molecules in which no net charge is present, formation of single

H-bonds is accompanied by the release of 3–6 kcal/mol of enthalpy (84). Figure 1 presents a scale of vapor phase values, showing second virial coefficients from cluster ion experiments.

In principle, such high free energies of formation of H-bonds can be realized not only in the vapor phase, but also in condensed phases if

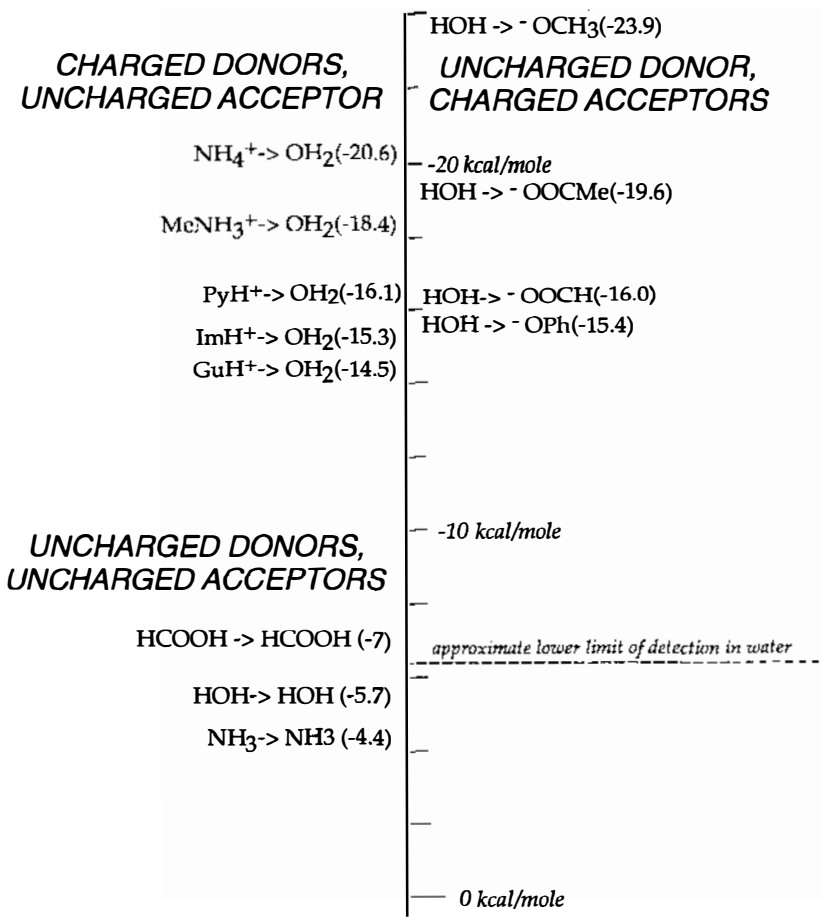

Figure 1 Scale of vapor phase H-bond values. Groups carrying a single positive or negative charge form single H-bonds to water molecules with the release of 16–23 kcal/mol of enthalpy (105). When no net charge is present, formation of single H-bonds is accompanied by the release of 3–6 kcal/mol of enthalpy (84).

elimination of a group from one of the H-bonding partners leaves a gap (in their complex) that cannot be filled by solvent water. Thus, high levels of equilibrium-binding discrimination, favoring OH-substituted compounds over their H-containing counterparts by factors as large as 10^7, have been observed in inhibitory complexes formed by adenosine and cytidine deaminases (183). After correction for differences in free energies of solvation between H- and OH-containing ligands, the desolvated OH-containing ligand is bound roughly 10^{12} times more favorably than the desolvated H-containing ligand. If the site itself must lose solvent to bind either ligand, then the difference in binding affinities could in principle approach the value that would be observed for H-bonding in the vapor phase. In fact, the observed difference in enthalpies is approximately -17 kcal/mol, approaching the value observed in the vapor phase for H-bonding between a water molecule and a partner bearing a single positive or negative charge (Figure 1). In the actual structures, a charged group is present but additional bonds are also present.

In practice, these extremes are moderated by multiple factors, including competition from solvent, molecular flexibility, cooperativity, and any constraints imposed by excluded volume. When solvent water has free access to both binding partners, as presumably tends to be the case in a fully unfolded protein, the effective strengths of their H-bonding interactions are expected to be greatly attenuated. Nevertheless, substantial free energies of interaction have been observed in enzyme-ligand interactions, even in the presence of water. For example, in Fersht's experiments involving mutations in tyrosyl-tRNA synthetase, which compare ligand affinities as reflected by k_{cat}/K_m, H-bonds between neutral partners typically seem to have free energies of formation of -0.5 to -1.8 kcal, and H-bonds involving one charged partner attain effective values of -3.5 to -4.5 kcal/mol (51). These values may not correspond to true equilibrium-binding constants, because they are based on kinetic constants that may describe reactions having transition states at different stages of advancement along the reaction coordinate.

Recent efforts to analyze the strengths of H-bonds in water compared true binding affinities of small model peptides, using the antibiotics ristocetin and vancomycin (38). Reassuringly, after correction for losses in rotational and translational entropy, the resulting negative free energies of H-bond formation appear to be comparable to those inferred from earlier mutagenesis experiments (179). However, Williams' analysis underscores the degree to which the apparent intrinsic enthalpy of H-bond formation depends upon underlying assumptions about differences in configurational entropy between free monomers and the H-bonded dimer. Related experimental observations concerning effects of mutations on

protein stabilities (see above/below) suggest that individual H-bonds are not usually very directional in their preferences, so that bond rotors may not in fact be fully frozen in these complexes. Figure 2, a "blurogram" that depicts the ensemble of conformations over which a glutamate side chain can maintain an H-bond with a given backbone >N-H group, illustrates this point pictorially.

Another interesting question raised by Williams' observations is the degree to which variations in the binding contributions of individual H-bonds are entropic, rather than enthalpic, in origin. The suggestion has been made that H-bond formation between amides is accompanied by a major gain in entropy resulting from the release of bound water molecules, which had been immobilized when they were H-bonded to the amides (38). Vapor-to-water transfer equilibria of simple solutes indicate, however, that for molecules of similar size, differences in polarity appear to be reflected almost entirely in enthalpies of solvation (Table 1). Thus, any entropy that is gained probably requires some other explanation.

Several investigators have used protein engineering to toggle selected hydrogen bonds off and on by replacing a native hydrogen-bonding side chain with a steric homologue that lacks a polar moiety (e.g. Asn → Val), followed by determination of the $\Delta\Delta G$ between native and mutant structures (1, 13, 17, 22, 155, 158, 159). Such experiments, while facile, cannot distinguish cleanly between energy differences resulting from H-bonds and from complicating side effects of comparable magnitude due to attendant changes in configurational entropy, hydrophobic burying, or van der Waals interactions.

In molecules capable of forming several H-bonds, one must also consider the possibility of competition from other groups that may furnish alternative binding partners. This effect is clearly illustrated by the solvation of ethylene glycol, which is much weaker than expected (16). Microwave spectroscopy shows that this diminution results from intramolecular H-bonding in the vapor phase. Alternatively, formation of one H-bond can affect the strength of another by electronic effects transmitted through the compound in question, and such effects may work in either direction. The existence of cooperativity has been demonstrated in the solvation of molecules such as p-nitrophenol by water, and anticooperativity is evident in the solvation of the p-nitrophenolate ion (182a). The structures of protein complexes with carbohydrates suggest that secondary -OH groups of sugars are typically involved in cooperative H-bonds with the protein (134). Finally, volumes of ionization of carboxylic acids (75), and of covalent hydration of aliphatic aldehydes (93), show that the solvation requirements of different parts of the same solute can be in conflict, leading to marked departures from additivity. The systematic approach of

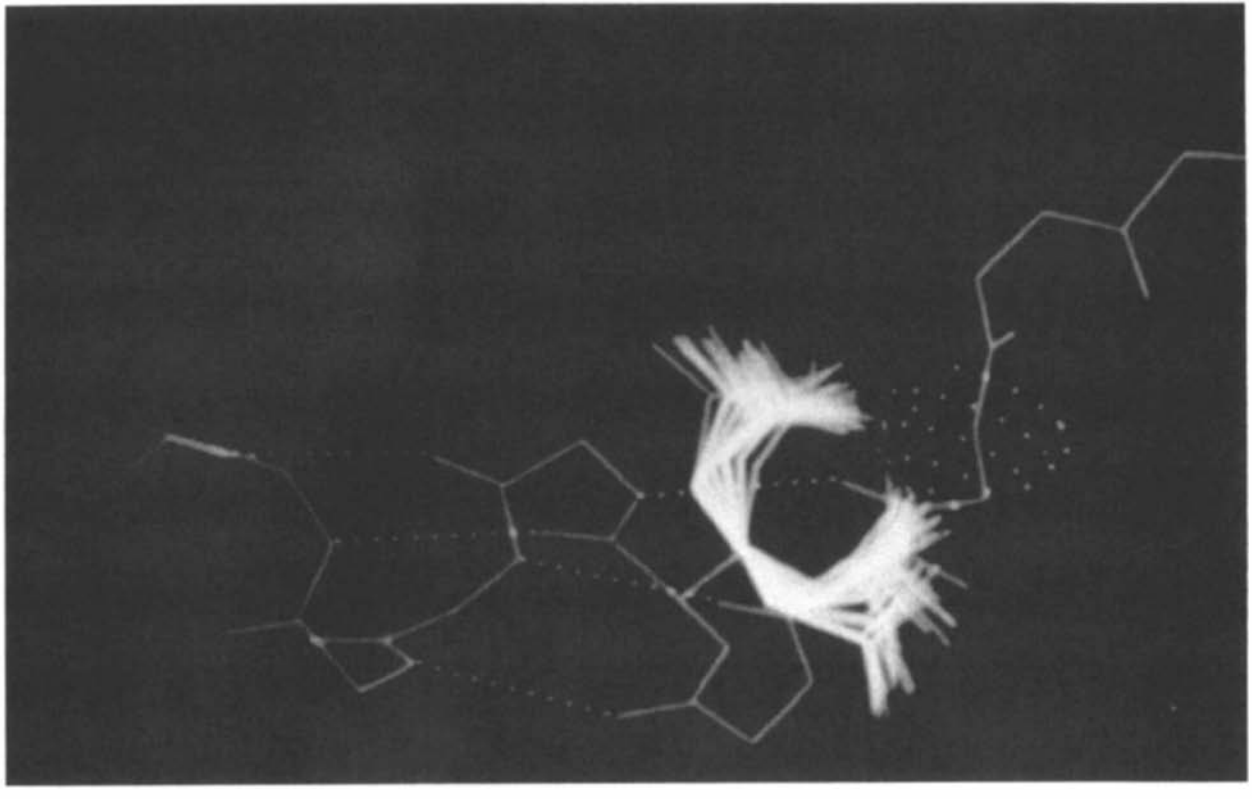

Figure 2 Hydrogen-bond rotors need not be fully frozen. The blurogram shows a glutamate side chain at the N3 position of a helix hydrogen bonded to the backbone >N-H at the N-cap position (128). Stippled surface represents the van der Waals radius around the >N-H donor; solid lines mark acceptable conformations (163) for an oxygen acceptor.

Table 1 Equilibria of transfer from the vapor phase to water (25°C)[a]

	K_{hyd} vapor → water	ΔG kcal/mol	ΔH kcal/mol	$T\Delta S$ kcal/mol
H_2	1.7×10^{-2}	+2.4	−1.3	−3.7
He	8.5×10^{-3}	+2.7	−0.8	−3.5
Ne	1.0×10^{-2}	+2.7	−1.9	−4.6
Ar	3.1×10^{-2}	+2.1	−2.7	−4.8
N_2	1.4×10^{-2}	+2.5	−2.1	−4.6
O_2	2.9×10^{-2}	+2.1	−3.0	−5.1
Cl_2	2.0	−0.4	−5.0	−4.6
CO_2	0.8	+0.2	−4.7	−4.9
H_2O	5.1×10^{4}	−6.4	−11.9	−5.4
CH_4	3.8×10^{-2}	+1.9	−3.2	−5.1
CH_3-CH_3	4.6×10^{-2}	+1.8	−4.4	−6.2
CH_3-OH	5.3×10^{3}	−5.1	−11.2	−6.1
CH_3-CH_2-OH	4.7×10^{3}	−5.0	−12.9	−7.9
CH_3-CH_2-NH_2	2.4×10^{3}	−4.6	−12.9	−8.3
CH_3-CO-CH_3	7.7×10^{3}	−3.8	−10.1	−6.3
CH_3-CO-OH	1.0×10^{5}	−6.8	−12.6	−5.8
CH_3-CO-OCH_3	7.7×10^{2}	−3.9	−10.1	−6.2
CH_3-CO-OC_2H_5	1.8×10^{2}	−3.1	−11.7	−8.6

[a] Data from Butler & Ramchandani (20a) and Frank & Evans (55). For an extensive collection of data for organic compounds, see Cabani et al (21a).

Gellman and coworkers (57), using NMR to study solution equilibria of model amides, has been designed to recognize and correct for these complicating factors.

Does the Hydrogen Bond Play a Directing Role in Protein Folding?

Complementary base-pairing in DNA represents a discriminatory mechanism based upon hydrogen bonding (176), tantamount to a stereochemical code. Does a corresponding code exist in proteins, and, if so, where is it found?

Regular secondary structure—α-helix (122) and β-sheet (121)—does not represent such a code because discrimination is not at issue. Indeed, one reason why these structural motifs can recur with such high frequency in proteins is that the hydrogen bonding in question is between backbone amides (>N-H···O=C<), and all residues (except proline) are indiscriminate in this respect.

A candidate stereochemical code with hydrogen bonding as a discriminatory mechanism is found in helix caps (128, 141). In the α-helix, the initial four N-H groups and final four >C=O groups necessarily lack

intrahelical hydrogen-bonding partners. These eight polar groups account for half of the total backbone hydrogen bonds in a protein helix of average length (~12 residues), with intrahelical Pauling-Corey-Branson H-bonds (122) accounting for the remaining half. It has been hypothesized (128) that hydrogen bond partners (i.e. caps) for these initial four amide hydrogens and final four carbonyl oxygens are provided characteristically by side chains of polar residues that flank the helix termini—Asp, Glu, Ser, Thr, Asn, Gln, and neutral His at the N terminus and Lys, Arg, Ser, Thr, Asn, Gln and His(+) at the C terminus.

To what extent do the data support this hypothesis? By exhaustive stereochemical modeling, Presta & Rose (128) showed that the helices in 13 proteins of known structure are flanked either by residues that could provide complete capping (i.e. all four backbone groups at either end of the helix) or by combinations that include Gly and/or Pro. Do these potential capping interactions actually occur? In X-ray-elucidated proteins, approximately one-half of the initial four amide hydrogens and one-third of the final four carbonyl oxygens are indeed capped by side-chain partners of residues nearby in sequence. An additional 5% of the amide hydrogens and 9% of the carbonyl oxygens are satisfied by partners that are distant in sequence. In some of the remaining instances, capping is provided by lattice neighbors, and in these cases, the X-ray structure does not model an isolated molecule in solution. Any ostensibly uncapped groups, which are presumed to be solvated, may be capped during the folding process and subsequently liberated, once the nascent helix is fixed within the tertiary fold, although an alternative explanation is raised by the discussion that follows.

In their analysis of X-ray-elucidated proteins, Richardson & Richardson (141) found sharply differentiated residue preferences at positions that flank helix termini. For example, Asn in the N-cap position occurs 3.5 times more frequently than expected by chance. Their statistics as well as the stereochemical analysis of Presta & Rose (128) were truncated at two residues beyond the helix proper, within the adjacent peptide-chain turn. Figure 3 extends these statistical data in a series of histograms derived from helices in proteins of known structure. Each histogram represents the normalized frequency of a single residue type at the six positions on either side of both termini. The position-dependent frequencies in Figure 3 are similar, though not identical, to those of Richardson & Richardson (141). Differences probably resulted because Richardson & Richardson and Presta & Rose used different criteria to identify helix termini from X-ray coordinates.

It is often argued that intramolecular hydrogen bonding of this type is unlikely because solvent water is so concentrated, and so efficient in form-

ing H-bonds to solutes, that solute-solute interactions cannot compete. That is, for a donor (D) and an acceptor (A) at the surface of the protein, the solvation reaction

$$D{\cdot}A + \text{solvent} \rightarrow D{\cdot}\text{solvent} + A{\cdot}\text{solvent}$$

is always favored. Whereas the preceding section cites examples of favorable intramolecular H-bonds, it might still be argued that these are typically buried within the protein where they are effectively shielded from solvent water.

Extending the argument, the attraction between any H-bonding donor or acceptor and solvent water is substantial. The vapor-to-water distribution coefficient of methanol, for example, is about 10^5-fold more favorable than that of methane, indicating the existence of an attraction between the hydroxyl group of methanol and solvent water, with a free energy of -7 kcal/mol.

However, solvent water does not appear to be extremely efficient at forming hydrogen bonds with solutes. Experimentally determined enthalpies of transfer, from vapor to water, are for methanol -11.24, ethanol -12.88, propanol -14.42, butanol -15.94, amyl alcohol -17.50, or $-CH_2$ -1.56. Extrapolation gives an enthalpy of -9.7 kcal for transfer of a hydroxyl group from vapor to water. The hydroxyl group of an alcohol should in principle be able to form three hydrogen bonds to solvent water, and the enthalpies of formation of those bonds, based on calculations by Kollman et al (177), should be worth ~ -12.3 kcal. In the actual case, the shortfall is about 20%, suggesting that the average hydroxyl group makes roughly 2.4 H-bonds to solvent water. The preexisting structure of bulk water is expected to be incompatible, at least to some extent, with the

Figure 3 Histograms derived from helices in Presta & Rose (128). Each histogram represents the normalized frequency of a single residue type at the six positions on either side of both termini. Nomenclature is as follows: N^{-6}-N^{-5}-N^{-4}-N^{-3}-N^{-2}-N^{-1}-N-cap-N^1-N^2-N^3-N^4-N^5-C^{-5}-C^{-4}-C^{-3}-C^{-2}-C^{-1}-C-cap-C^1-C^2-C^3-C^4-C^5-C^6, with the helix proper extending from N-cap to C-cap. To normalize a frequency, the number of instances of residue X at position *i* was divided by the total number of residues at position *i*, yielding the fraction of times that residue X was found at position *i*. This fraction, converted to a percentage, was then divided by the percentage of residue X in the total data base. For example, at the N^{-6} position: among the 73 residues *in toto*, 6 are Ala, representing 8.22%. There are 294 Ala in the data base of 3423 residues, so in general, the percent composition of Ala is 8.59%. Thus, the normalized frequency of occurrence for Ala at N^{-6} is 8.11/8.59 = 0.94. A normalized value equal to 1 means that the frequency of occurrence of residue X at position *i* is the same as the frequency of occurrence of residue X at large. A value of 2, for example, would then mean that the residue occurs at this position twice as often as would be expected from the empirically determined percentage composition in the data base.

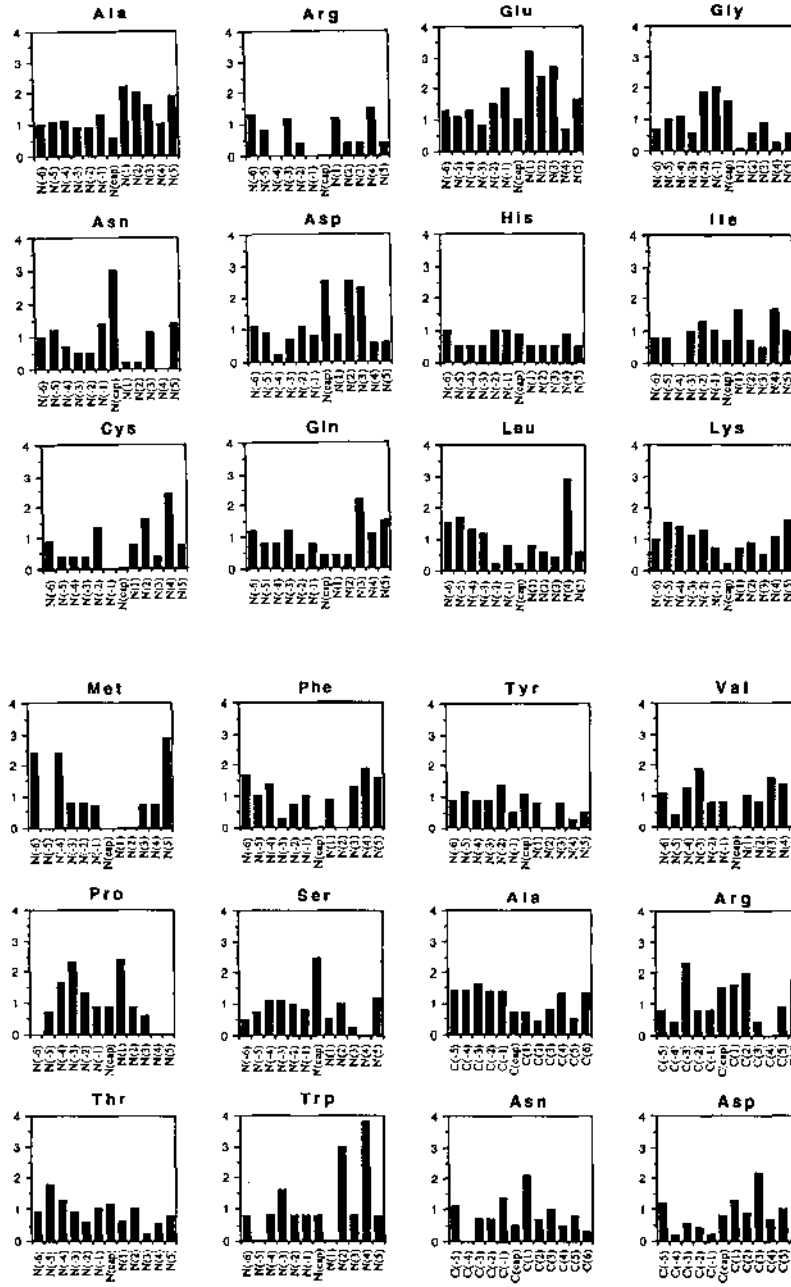

Figure 3

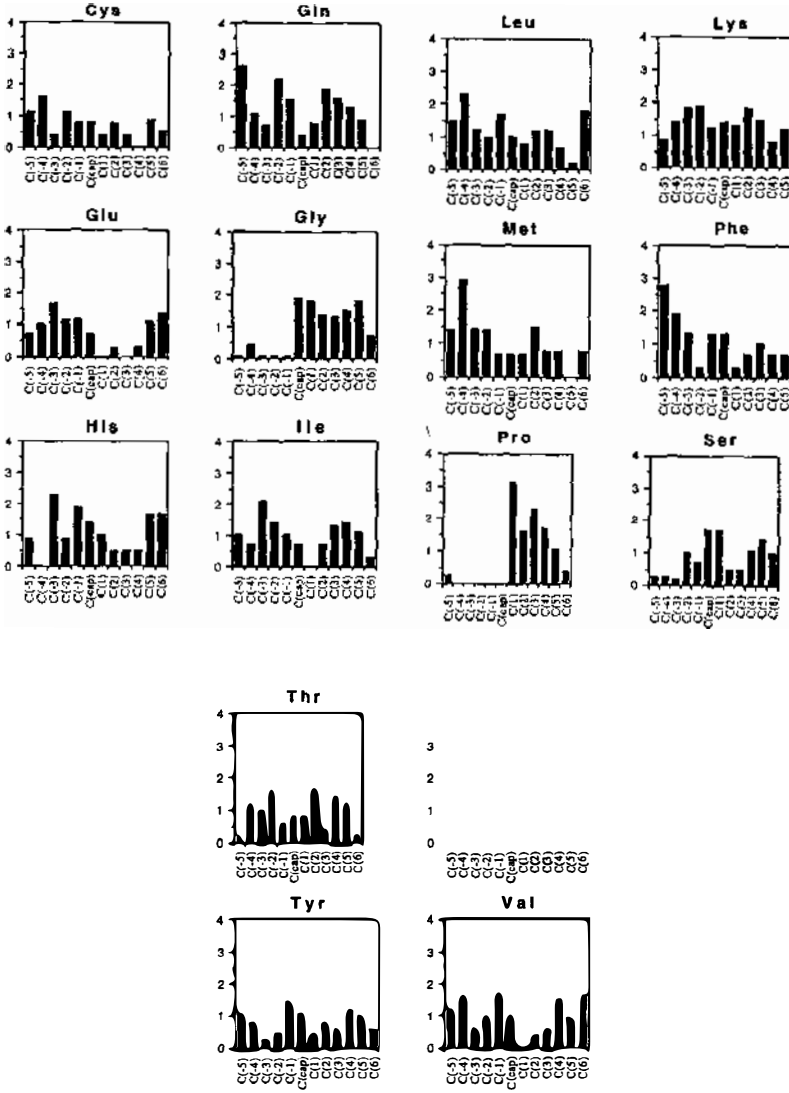

Figure 3 (*continued*)

introduction into that bulk water of almost any solute, and in simple organic compounds, a $-CH_2-$ increment is typically observed to make an unfavorable entropic contribution of ~ 1.6 kcal to the free energy of transfer from vapor to water. As in other groups of similar size, a foreign hydroxyl group contributes $+1.0$ to $+1.6$ kcal to the observed free energy of transfer (21).

This rationale for favorable intramolecular H-bonding, even at the solvent-accessible molecular surface, is fortified by experimental results from several groups. In a mutational analysis of the protein barnase, Serrano & Fersht (154) found that a side chain to backbone H-bond at the N-terminus of either helix stabilizes the protein by up to ~ 2.5 kcal/mol relative to a non-hydrogen-bonding residue in the corresponding position. (Of course, the conclusion is subject to caveats already mentioned.) Bruch et al (20) synthesized a peptide with the sequence of a carboxypeptidase helix and its flanking polar groups, and compared the helicity of the native-like sequence with variants in which alanine substituted for the flanking polar groups. Using circular dichroism to assess helicity, Bruch et al found that substitution of Ala for the three residues at either end reduced θ_{222} to two-thirds the value measured for the native peptide. A telling example is provided by the finding that a probable pH-dependent switch in which protonation of an H-bonded histidine situated at the N-cap position of a helix in apo-cytochrome b_5 lowers the helix$\rightleftharpoons$coil equilibrium constant by more than an order of magnitude (86a). On the other hand, Fairman et al (48) replaced the amide-blocking group of a model peptide (which can potentially H-bond to one of the four terminal main-chain $>C=O$ groups) with its methyl-ester (in which no H-bond is possible), with no observed effect on stability. In this case, either the intramolecular H-bond is unfavorable or, more likely, the C terminus of the peptide helix is frayed.

Kallenbach and coworkers conducted a more extensive test of helix capping using a reference peptide together with a series of derivatives (99). In the reference peptide, each position bears the residue found to occur with highest frequency at the corresponding position among protein helices (141). Derivatives are devised in matched pairs; each pair quantifies the effect of a single substitution, made either at the N-cap or at a central position. For example, Ala, the most helix-stabilizing residue in a central position (98, 106, 115, 119), diminishes helicity by a factor of two when at the N-cap position. Conversely, Ser, a statistically preferred residue at the N-cap position, diminishes helicity by a factor of two when in a central position. The study shows that statistical preferences seen in X-ray studies (5, 141) correspond to structural differences in a related series of peptides.

Taken in sum, these experimental findings suggest widespread helix capping in natural proteins. This conclusion is supported by a recent study

of Stickle et al (163), who conducted a global census of hydrogen bonding in high resolution, X-ray-elucidated proteins. In their study, the set of all side chain–to-backbone hydrogen bonds is subdivided into (*a*) backbone donors with side-chain acceptors and (*b*) side-chain donors with backbone acceptors. Histograms of these two groups show that pronounced peaks are situated at helix-capping loci, against an otherwise undifferentiated, low-level background. In view of these histograms, a directed analysis of existing X-ray structures may disclose the existence of further, previously overlooked, interactions of this type.

HYDROPHOBIC EFFECT

The hydrophobic effect has been a topic of chemical interest for more than a century (172). Early on, Edsall noted that the transfer of an apolar compound from an organic medium to water is accompanied by a large positive change in heat capacity (41). Proteins exhibit similar, striking differences in heat capacity between their folded and unfolded states (129). Prompted by this similarity, Kauzmann proposed that the transfer of small apolar solutes from water to liquid hydrocarbon could model the process in which apolar side chains in a protein are sequestered from solvent upon folding (74). However, as X-ray-elucidated structures became available, Richards surveyed protein interiors and found them to be remarkably close-packed in a manner more reminiscent of a solid than of liquid hydrocarbon, and he questioned the suitability of the liquid hydrocarbon model (137). Since that time, the field has been mired in controversy, with disagreement over the meaning (63), sign (37, 109, 131), and magnitude (156) of the hydrophobic contribution to folding.

Despite ongoing controversy, there can be no disagreement about the tendency of saturated hydrocarbons to leave water and enter other solvents. Dispute arises over whether solute molecules tend to leave water and enter less polar solvents primarily because they are repelled by water, or because they are attracted to the less-polar solvent. This question can be analyzed by using some absolute standard of reference such as the vapor phase that neither attracts nor repels solutes. Framing the question in these terms, one finds that methane exhibits an appreciable water-leaving tendency, with an equilibrium distribution between water and the vapor phase of 27 in favor of the vapor phase. This tendency, expressed as an equilibrium constant for transfer from dilute aqueous solution to the dilute vapor phase, increases in the normal alkanes as follows: methane 27, ethane 20, propane 29, butane 38, pentane 51, hexane 74, heptane 83. Thus, for each methylene increment ($-CH_2-$), transfer to the vapor phase is enhanced very gradually by an average factor of 1.3, equivalent to 0.14

kcal in free energy. This increment is about the same in the normal series of hydrocarbons, acetic acid alkyl esters, primary amines, and primary alcohols. In contrast, each methylene increment increases the distribution coefficient for transfer from water to a nonpolar solvent by a factor of 4.0, equivalent to 0.8 kcal in free energy (32, 72, 182).

Summarizing, saturated hydrocarbon molecules have an appreciable tendency to leave water, and are thus hydrophobic in any sense of the word. As the size of the hydrocarbon increases (addition of a methylene increment), this tendency is enhanced only gradually, while the corresponding tendency to enter a nonpolar solvent, such as a hydrocarbon, is considerably stronger.

With this background in mind, the definition of hydrophobicity used here is the operational process in which an apolar group is transferred from a polar or neutral phase to an apolar phase. The effect is pertinent to protein folding to the degree that residues with apolar side chains are expelled from water and engender a solvent-shielded molecular interior.

Many investigators have studied the transfer process (36, 74, 167, 180). The effect of a methylene increment on the free energy associated with simple removal of a solute from water is small at room temperature, but major compensating changes occur in other thermodynamic parameters. Removal of nonpolar molecules or groups from water is accompanied by increases in entropy (and volume) and by a compensating uptake of heat from the surroundings. Thus, hydrophobic bonds are distinguished by a tendency to become stronger with increasing temperature (74, 112). Because the properties of liquid water, and of water of solvation in particular, are still not fully understood, an ongoing discussion has examined the likely origins of the entropy increases that accompany the formation of hydrophobic bonds, on the removal of hydrocarbons from water to the vapor phase. Conditions have been found (9, 132, 184) under which the major changes in entropy and heat capacity disappear.

The self-cohesive properties of water, although not unique (47), are unusual (42). A reasonable assumption is that changes in the properties of water in the immediate neighborhood of the solute could account for the loss of entropy that accompanies introduction of a nonpolar molecule or methylene increment into water from the vapor phase or from a nonpolar solvent. Frank & Evans (55) suggested that the observed changes in entropy and heat capacity might imply the formation around solutes of a kind of clathrate or iceberg structure, in which water molecules were more ordered than in the bulk solute, but did not resemble Ice I in any literal sense.

More recent evidence suggests that entropic effects associated with introducing nonpolar solutes into water may arise, not only from restrictions

on the mobility of water molecules, but also from restrictions on the mobility of solutes when they are introduced into the aqueous environment (6). Solutes experience a three- to fivefold enhancement in ^{13}C spin-lattice relaxation times when they are transferred to water from nonhydroxylic solvents of similar viscosity, so that water appears to be unusual in the restrictions that it imposes on the motion of dissolved solutes (65). In normal aliphatic compounds of increasing size, solubility might (according to this view) be reduced by progressive restrictions on internal rotation. Compounds with internal rotations already restricted would not be affected to the same extent. It is therefore of interest that steroids and cycloalkanes display considerably lower activity coefficients in water than nonrigid compounds, with reference both to nonpolar solvents and the vapor phase (116). For similar reasons, the hydrophilic character of proline is greater than that of acyclic amino acids of similar size; this effect has been shown to be entirely entropic in origin (59). Such disparate solutes as argon, CO_2, and water itself show almost the same entropy of solution, despite major differences in their affinities for solvent water and the restrictions that their polar interactions might have been expected to impose on the solvent.

Scales of Hydrophobicity

The free energy required to transfer amino acid solutes from an aqueous to an organic phase, $\Delta G^0_{\text{water}\rightarrow\text{organic}}$, is related to their relative solubility in either phase. As shown by Cohn & Edsall (30), the free energy of transfer, $\Delta G^0_{p\rightarrow p'}$, between phases p and p′ is related to the partition coefficient between the phases, $K_{p\rightarrow p'}$, by the equation $\Delta G^0_{p\rightarrow p'} = -RT\ln K_{p\rightarrow p'}$. Cohn & Edsall provided the solubilities for many amino acids in both water and ethanol.

Beginning with an early compendium based on the relative solubilities of some amino acids in water compared with dioxan or ethanol (113), several experimental scales have been developed for comparing the affinities of amino acid side chains for solvent water (31, 147). The simplest of these uses the gas phase, devoid of attractive or repulsive forces, as a reference phase in water-to-vapor distribution measurements at infinite dilution (181). More complex scales use a second solvent, whose solvation properties must be considered. Of these, one of the least complicated is cyclohexane, with a dielectric constant of 2 (135). Water-to-vapor and water-to-cyclohexane distributions (Figure 4) are related to each other by a third scale (not shown) of cyclohexane-to-vapor distribution coefficients. Positions of amino acids on the latter scale are closely related to surface area, consistent with the view that cyclohexane can participate in van der Waals' interactions with solutes and not much else. Other solvents, such as 1-octanol (50), dissolve substantial quantities of water and are capable

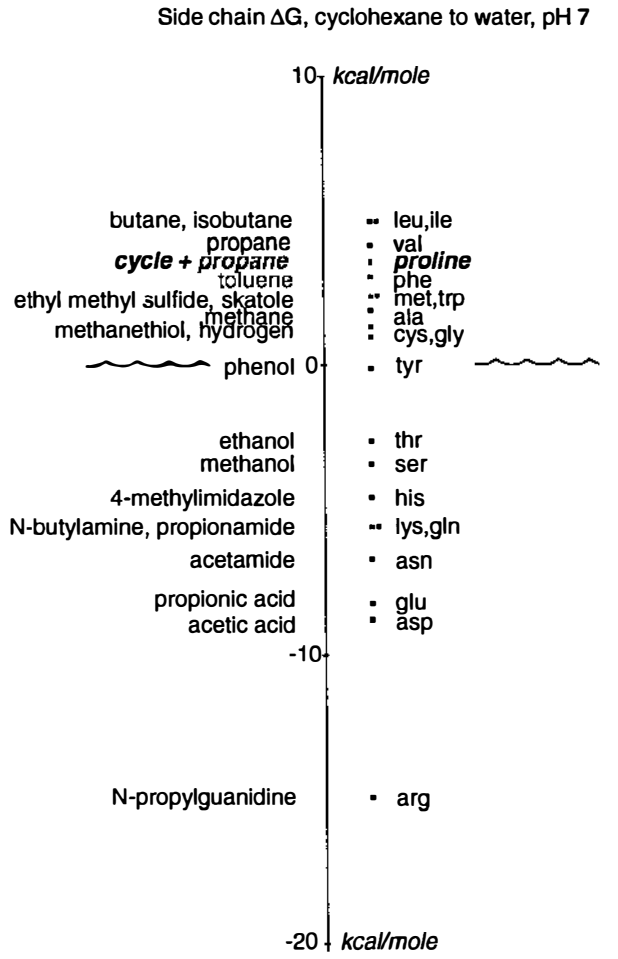

Figure 4 Cyclohexane-to-water distribution coefficients. The scale includes a recent value for proline (*bold*) that was generated by a cycle of transformations of the side chain of norvaline, as discussed in the text.

of specific polar interactions with solutes such as indole, resulting in the misleading impression that tryptophan, for example, is extremely non-polar.

The scale shown in Figure 4 includes a recent value for proline, which does not have a side chain in the ordinary sense. The value for proline was generated by a cycle of transformations of the side chain of norvaline,

for which exact experimental information could be obtained. The results indicate that proline is more hydrophilic than would have been expected from simple considerations of surface area. Proline exemplifies a general tendency of ring compounds, mentioned previously, to enter water more readily than their acyclic counterparts. The exceptional hydrophilic character of proline is entropic rather than enthalpic in origin. Ring compounds have less freedom of rotation than open-chain compounds, and may experience fewer constraints upon entering the relatively structured environment of solvent water. However, ring compounds have slightly less surface area than their acyclic homologues; e.g. the standard-state area of the prolyl side chain is 111.0 $Å^2$ while that of the valyl side chain is 128.4 $Å^2$ (90).

Generalizations for Protein Folding

How are these scales related to protein structure? The above distribution coefficients of amino acid side chains can be related to various measures of the tendencies of these side chains to appear on the surfaces or in the interiors of proteins (107, 146). However, a note of caution remains in order. Despite the strengths of these correlations, the tendencies of residues of each kind to be exposed to the solvent in real proteins vary as an extremely shallow function of their water-to-vapor or water-to-cyclohexane distribution coefficients, i.e. the former values, expressed as equilibrium constants, vary over a numerical range of 15 orders of magnitude, whereas the latter values vary over a range of only 2 orders of magnitude. This insensitivity is hardly surprising: the interiors of proteins tend to be more polar than cyclohexane, with evident hydrogen-bond partners for virtually every buried polar group (8, 163), and most surface locations may be less polar than bulk water.

Nevertheless, it has been possible to arrive at generalizations that are valuable for understanding the effects of mutations and for predicting protein structure. After correction for accessibility, the mean change in stability (i.e. $\Delta\Delta G$) for a large collection of hydrophobic mutants was found to scale linearly with the free energy of transfer from water to *n*-octanol (118). In a mutational study of myoglobin, where Ala was substituted at 16 selected sites, approximately 80% of the measured change in stability ($\Delta\Delta G$) can be accounted for by the difference in surface area between the native residue and the alanine substitution (126). Table 2 shows energy differences, expressed as $\Delta\Delta G$ values, between a set of mutations of large nonpolar residues and those of smaller residues upon denaturation of bacterial nucleases by guanidine. These are seen to be numerically equivalent, within experimental error, to values based on the cyclohexane-to-water distribution coefficients. This near-identity suggests

Table 2 Effects of mutations on the stability of staphylococcal nuclease[a]

Mutation	Denaturation equilibrium, GU$^+$ ΔΔG (kcal)	Cyclohexane → water ΔΔG (kcal)	Octanol → water ΔΔG (kcal)
Leu → Ala (11 mutants)	-3.10 ± 1.5	-3.11	-1.90
Val → Ala (9 mutants)	-2.40 ± 1.5	-2.23	-1.24
Ile → Ala (5 mutants)	-3.60 ± 1.0	-3.11	-2.04
Ala → Gly (5 mutants)	-0.95 ± 0.2	-0.87	-0.42

[a] Data from Shortle et al (159), Radzicka & Wolfenden (135), Eisenberg & McLachlan (43), and Fauchère & Pliska (50).

that when staphylococcal nuclease is denatured, these nonpolar residues pass from a cyclohexane-like to a water-like environment. Evidently, these nonpolar residues can find environments in the native protein with dielectric constants of approximately 2, and denaturation involves complete exposure of their side chains to solvent water. This relationship fails for the more-polar amino acids, in such a way as to suggest that they find more-polar environments in the native protein, as might be expected from the observation that buried polar groups are usually hydrogen bonded.

The distribution of hydrophobic groups along the linear sequence of a polypeptide chain is often used in a predictive manner. A plot of the average hydrophobicity per residue against the sequence number reveals loci of minima and maxima in hydrophobicity. Such a plot, termed a hydrophobicity profile, has been applied to proteins of known sequence but unknown structure to predict (*a*) the probable location of peptide-chain turns (144), (*b*) segmentation of the molecule into interior/exterior regions (83, 148), (*c*) likely antigenic sites (64), and (*d*) the location of membrane-spanning segments (44, 83). Any profile of this type is based upon some scale of hydrophobicity for the naturally occurring amino acid residues, of which there are many (31, 43, 147).

Perhaps the most widely applied generalization is that made by Chothia, who noted that free energies of transfer of side chains of nonpolar amino acids to water, from a diverse collection of organic solvents, show a slope of 25 cal of free energy per squared Ångstrom of buried surface area (24). However, this value was called into question recently. From consideration of the effects of the relative molecular volumes of the solute and water, Honig and coworkers (157) concluded that concave regions such as protein-binding sites (or sites in a protein from which a residue is removed by

mutation) can be more hydrophobic than had been realized, resulting in a slope of 46 cal of free energy per squared Ångstrom of buried surface area. Debate about this topic is ongoing (165).

On consideration, one would not expect that any single solvent or physical environment could, except by an occasional coincidence, represent the variety of environments experienced by residues within a protein. Thus, correlations between various thermodynamic variables, e.g. heat capacity or free energy of transfer from water to nonaqueous solvent, may seem surprising and, by inference, informative. Again, a note of caution is in order. Some time ago, Bigelow showed that the ratio of hydrophilic to hydrophobic groups in globular proteins is confined within a narrow range of values (15). Later, Janin (67) and Teller (171) demonstrated that the surface area of a protein monomer is a simple function of its molecular weight, i.e. upon folding, proteins bury a constant fraction of their available surface. Chothia (25) found that this fraction can be subdivided into characteristic contributions from polar and apolar residues. Extending these observations to an atomic level, Lesser & Rose (90), using the algorithm of Lee & Richards (87), calculated the mean area buried upon folding for every atom in a data base of X-ray-elucidated proteins and found that, on average, each atom type buries a constant fraction of its standard-state area (Figure 5). Surprisingly, the mean area buried by most (though not all) residues can be closely approximated by summing contributions from three characteristic parameters corresponding to three generic atom types: (*a*) carbon or sulfur, which are 86% buried, on average; (*b*) neutral oxygen or nitrogen, which are 40% buried, on average; and (*c*) charged oxygen or nitrogen, which are 32% buried, on average.

Although such characteristic behavior seems to confer welcome predictability on the folding process, its existence leads inescapably to a complex of confounding functional interrelationships. For example, if proteins bury a well-behaved and constant fraction of their apolar surface and also have a constant specific heat capacity, then any strong correlation between these properties need not be causal, as noted by Record and coworkers (96).

Indeed, to the extent that the mean behavior of individual hydrophobic groups can be predicted by a single line, they are indistinguishable from aliphatic compounds, which exhibit analogous linear behavior in both liquid and condensed phases (161). Although aliphatic compounds do have characteristic thermodynamic and geometric properties, they lack the unique molecular architecture that is the very hallmark of globular proteins. Thus, the regularities described in the preceding paragraphs suggest that the hydrophobic effect may push the chain to collapse to a compact, water-excluding state as an effective means to sequester apolar

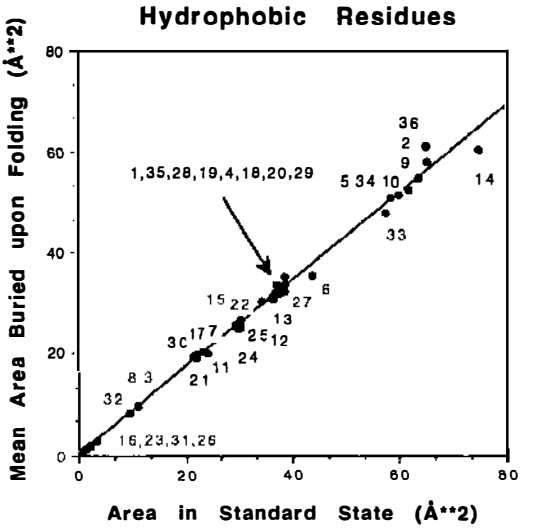

Figure 5 Plot of the mean area buried upon folding vs the standard-state area for 36 atom types from 8 apolar residues in a data base of 61 proteins of known structure (90). Residue types included are Val, Ile, Leu, Met, Phe, Trp, Cys, and 1/2-Cys. The identity of individual groups (which can be obtained from the original paper) is of less significance than the fact that all are well-described by a single straight line, as shown. Groups 2 and 36, the worst outliers, are -SH and -S- atoms, respectively.

groups from solvent access. However, the reason why the chain adopts a specific, unique, equilibrium fold instead of an ensemble of micelle-like structures is unlikely to be the hydrophobic effect per se, although Lau & Dill (86) have argued to the contrary.

PACKING

In the 1970s, after the first several protein structures had been elucidated by X-ray crystallography, Richards made a fundamental observation (137). Upon calculating the packing density—the ratio of the summed atom volumes to the molecular volume—he determined that the interiors of proteins are packed as well, on average, as crystals of small organic molecules with few cavities of atomic dimension, more reminiscent of a solid than of liquid hydrocarbon. A typical cross-section through a protein molecule reveals lumpy, irregular-residue side chains that fit together with exquisite complementarity, almost like pieces of a three-dimensional jigsaw puzzle. This surprising observation prompted a general reassessment and

altered prevailing notions about protein structure. Richards' seminal review appeared earlier in this series (137).

Richards' work led to the appealing idea that interior packing is the property that distinguishes a protein from an oil drop. According to this view, lysozyme, for example, adopts the lysozyme fold and not the ribonuclease fold because efficient internal packing of the lysozyme sequence can be achieved only in the former case. Like a jigsaw puzzle, the side chains fail to fit together unless the conformation is "right." This interpretation motivated many subsequent attempts to predict and/or characterize conformation based on internal packing constraints (88, 127, 188, 189; J. S. Fetrow & S. Bryant, submitted).

Internal Packing and Protein Folding

Richards' observations are consistent with a restricted architecture in which constituent parts fit together with snug tolerances. However, that picture must be incomplete because further evidence supports a somewhat different view. Both NMR (185) measurements and molecular-dynamics calculations (104) indicate that buried aromatic rings experience a highly mobile microenvironment, tight packing notwithstanding. Furthermore, the ensemble of methyl rotors, which is coextensive with the molecular interior, is mobile and preferentially populates staggered (i.e. relaxed) configurations (80), suggestive of more fluid-like surroundings. Indeed, side chains in X-ray-elucidated protein structures populate staggered conformations (69, 100, 127), which is again consistent with internal relaxation.

Compounding the paradox, recent mutational studies of proteins have demonstrated that the buried interior can tolerate a broad diversity of residue substitutions, usually with only minor effects on structure, stability, or function (1, 46, 95, 102, 103). Thus, although static measurements of packing suggest a tight particle with constrained tolerances, both dynamics and mutagenesis studies portray the molecule as malleable with forgiving tolerances.

This conundrum prompted Behe et al (12) to conduct a simple test of whether protein conformation is determined primarily by packing interactions. In this test, they analyzed proteins of known structure for the presence of preferred interactions either between residue pairs or among larger composites, reasoning that if side-chain complementarity is an important source of conformational discrimination, then sets of residues that interact favorably should be readily apparent. Also, if no interactions are especially favorable, then efficient packing—an undeniable experimental fact—is achieved without severe limitation of the individually allowed side-chain orientations. The analysis of packing was conducted in

two parts. First, all residue pairs, X-Y, were assayed for the existence of particularly favorable interactions between X and Y (i.e. binary interactions). None were found. However, preferred higher-level packing arrangements (i.e. tertiary, quaternary, etc) may exist despite an absence of preferred binary interactions. To address this remaining possibility, the interaction of all X-Y pairs with the rest of the protein was assessed. Again, no preferred interactions were found. This analysis leads to the unexpected conclusion that high packing densities so characteristic of globular proteins are readily attainable among clusters of the naturally occurring residues.

These two ostensibly conflicting views of protein architecture—tight versus malleable—would be reconciled if tight packing could be achieved primarily by local adjustment of the structure, without global rearrangement. In this case, a perturbation event—either kinetic (e.g. a ring flip) or structural (e.g. a mutation)—would be followed by local relaxation of the structure, with minimal impact beyond the immediate microenvironment (56). Mounting evidence supports the view that the ordering imposed by packing requirements does not propagate beyond a highly local neighborhood. For example, a single mutation in a protein typically results in many small structural changes, but all are confined within a several-Ångstrom sphere centered about the site of mutation (1). Such mutations do measurably affect packing (173), usually reducing stability (150), but the changes are generally quite small. However, even in extreme cases where a mutation results in a substantial packing fault (45), the overall conformation of the molecule does not undergo global rearrangement to eliminate the fault.

Because a single mutation is most often destabilizing, one might think the combined effect of several such mutations, each disruptive of packing, could cause gross conformational rearrangement. However, the cumulative effect on stability is unlikely to be realized in this fashion for reasons described below:

According to Lattman & Rose (85, p. 439):

> The folding reactions of small, globular proteins typically exhibit two-state kinetics, in which the folded and unfolded states interconvert readily without observable intermediates (76).[1] The free energy difference, ΔG, between the native and denatured states of such proteins is quite small, lying in the range of ~ -5 to -15 kcal/mol (117). In the usual thermodynamic interpretation, if a single mutation destabilizes a protein by ~ 1–2 kcal/mol (a typical change), and if the difference in free energy between native and denatured states is ~ 8 kcal/mol (a typical ΔG), then a few such mutations might be expected to destabilize the molecule altogether (i.e. $\Delta G > 0$).
>
> While mutations may well result in structural changes, they are not expected to affect two-state behavior materially. Logically, the persistence of two-state behavior implies that a population of native-like molecules will exist, even in the presence of the most

[1] Citation numbers correspond to those used in this review.

destabilizing mutations. To emphasize this point, suppose that several destabilizing mutations were introduced, shifting the equilibrium constant such that $\Delta G = +5.5$ kcal/mol. Assuming two-state behavior at physiological RT, in a population of 10,000 molecules, ~9999 would be unfolded, but the remaining molecule would still adopt the native-like conformation.

Thus, jigsaw-puzzle packing is unlikely to be the main factor that governs conformational specificity, i.e. the stereochemical code that discriminates the native conformation from other conceivable chain folds. The two arguments in summary are:

1. X-ray studies indicate that mutations do alter packing slightly, but the changes are accommodated locally, without gross rearrangement of the conformation.
2. From folding studies, one can infer that packing faults push the folding equilibrium toward the unfolded state, not toward a novel conformation.

The conclusion that packing is not a principal source of conformational specificity prompts two obvious questions: given the close-packed nature of the protein core, how can perturbations that affect packing be accommodated without global disruption, and if packing is not the source of specificity, what is? We turn now to the former question; the next section examines the latter one.

Generalizing from current studies, a structural perturbation, such as a mutation, can be accommodated without global rearrangement. This finding appears enigmatic, in view of the close-packed nature of the protein core; were a tennis ball added to the middle of a stack of close-packed tennis balls, the perturbation would be propagated throughout the stack. How can the perturbing effect of adding a carbon atom to the close-packed protein core be absorbed with only local rearrangement?

We propose an explanation. The interior of a globular protein is comprised largely of assemblages of repetitive secondary structure (92)—helix, sheet, and their superstructures—with turns and loops relegated to the exterior (81, 91, 144). At least in part, the high packing densities observed in proteins are a consequence of the fact that segments of repetitive secondary structure pack together tightly: helix against helix (28), strand-of-sheet against strand-of-sheet (189), and helix against strands-of-sheet (68). Even turns and loops, though typically situated on the molecular exterior, are known to be intrinsically well packed (91).

In general, perturbation of a residue (e.g. by mutation) will involve some segment of secondary structure that includes the affected residue, together with other impinging segments. In the case of interacting helices, the side chain–into-groove packing should be somewhat insensitive to the precise

details of the perturbation. The globins, being predominantly helical, are a useful model for accommodation of this kind. As shown by Lesk & Chothia (89), natural mutations do not promote global change. Instead, the axis of the involved helix realigns slightly, with adjacent turns acting as flexible joints. Sheet is also highly flexible within its allowed modes of deformation, as demonstrated by Salemme (149) and Chothia (26).

In summary, the high packing densities of globular proteins are an undeniable experimental fact (137). Given the lumpy, idiosyncratic shapes of residue side chains, this fact begs rationalization, and an explanation has been proposed. However, regardless of the path by which high packing densities are realized, the molecular interior of a folded protein should be well packed. Otherwise, dispersion forces would favor denaturation, because presumably efficient packing can be achieved between protein and solvent in the denatured state. High packing densities have been interpreted to mean that internal packing is a principal source of conformational specificity, but recent structural and folding studies imply that such a conclusion may be unwarranted.

A MODEL FOR PROTEIN FOLDING

We propose a hierarchic framework model of protein folding, an expanded hybrid of two earlier models (76, 145), in which nascent elements of secondary structure associate in step-wise fashion, leading to larger modules and further association. Proteins are known to exhibit hierarchic architecture (35, 145), and numerous examples of independently folding protein domains and subdomains have been documented, both in natural (73, 178) and synthetic (114) systems. Hierarchic organization is a familiar and simplifying pattern in the macroscopic world, so the existence of hierarchic architecture and autonomous domains in proteins has been a suggestive finding. The following paragraphs describe major features of the model.

HYDROPHOBIC COLLAPSE Akin to a critical micelle concentration (168), under physiological conditions, a polypeptide with both a sufficient fraction of hydrophobic residues and sufficient chain length to enclose a suitable volume will undergo spontaneous collapse, because constituent hydrophobic groups are a better self-solvent than bulk water.

MOLTEN GLOBULE Under conditions that promote the molten-globule state (39, 62, 82), which includes some extraneous water and precedes fixed tertiary interactions, the usual cohesive properties of water would be reduced or eliminated because of the high local concentration of protein. For this reason, the hydrophobic effect is expected to be markedly di-

minished once the molten globule has formed, and further folding will arise from the drive to return interior waters to the bulk phase.

PACKING Because the molecule traverses paths leading to the folded state, from either an unfolded state or the molten globule, the search for close-packed interactions would not be rate-limiting. This critical feature arises from the fact that protein structure is reducible, almost entirely, to just four secondary-structure formats: α-helix, β-sheet, turns, and loops (54, 139). The core is comprised primarily of the two repetitive structures—α and β—with connecting elements of nonrepetitive structure turns and loops situated typically on the outside. Both α and β have extensible, open-ended structural formats, like tile patterns (123). These elements can pack together tightly in any pairwise combination (92)—α with α, β with β, and α with β—resulting in the high packing densities that are characteristic of globular proteins.

HYDROGEN BONDS Because the core is comprised largely of α-helix and β-sheet, many polar groups will necessarily be buried. Hydrogen-bonding requirements for these polar groups must be satisfied within the molecular interior, or hydrogen bonding would favor denaturation. An effective way to satisfy such requirements is to establish H-bonds prior to or concomitant with formation of the core. The finding that H-bonds are predominantly local (8, 163) is consistent with this order of events.

AUTONOMY OF SECONDARY STRUCTURE Several years ago, Baldwin and coworkers (14), expanding upon work of Brown & Klee (19), demonstrated the existence of stable, short, isolated helices in water, under near-physiological conditions. Though not as well studied, stable, isolated turns (40) and loops (61) have also been reported. Studies of β-sheet in isolation have lagged behind other categories, because sheet, by its very nature, tends to aggregate in isolation. Unlike earlier studies of homopolymers (164), contemporary studies utilize sequences drawn from natural proteins (98, 106, 115, 119); it is plausible that factors that promote secondary structure in isolation will also do so in proteins.

HIERARCHY Once secondary structure is fixed, only a few assembly patterns will both satisfy steric constraints and bury exposed apolar surfaces effectively (29, 133, 142). The additional imposition of hierarchic constraints (35, 145) would further winnow this set, possibly to uniqueness. In such a folding process, segments of secondary structure associate, spawning supersecondary structure. In turn, these small hierarchic units coalesce into larger modules, in stepwise progression, leading to the hierarchic architecture that is observed in globular proteins.

COOPERATIVITY In globular proteins, folding is accompanied by a hierarchy of energetically favored and mutually reinforcing conformational restrictions. Dipeptides preferentially populate discrete regions of a Ramachandran (ϕ,ψ) plot (18). Larger structures, e.g. helix and sheet, are found within these preferred regions (140), although their characteristic hydrogen bonding is not a factor in determining the preferred conformation of a dipeptide. Supersecondary structure involves favorable packing between segments of helix and sheet (28); again at this level, the burial of hydrophobic surface promotes segment association but is not a factor in segment formation. Thus, at each successive level in the accretion of structure—dipeptide, secondary structure, and supersecondary structure—new quasi-independent, stabilizing factors come into play that serve to enhance those of earlier steps, with no energetic compromises required at any level.

These levels of structure emerge regardless of their order of formation. Early steps in the process, with individual stabilities approximating physiological *RT* (18), should be readily reversible. However, as secondary structure is nucleated and structures of persisting stability form, the folding reaction proceeds spontaneously, leading to the observed cooperativity of protein folding under physiological conditions.

Central to any model is the question of why proteins have a unique equilibrium structure. In the features described above, the set of allowed conformations would be restricted to a small and possibly unique number, once segmentation of the polypeptide chain into elements of secondary structure has been fixed. As such, the formation of secondary structure plays a pivotal role in our model. In light of recent work, the secondary-structure format may be governed by an underlying stereochemical code.

A Stereochemical Code for Protein Folding

To a close approximation, protein structure can be dissected into just four categories of secondary structure: helix, sheet, turns, and loops (54, 138). If so, then factors governing secondary structure formation need only be sufficient to discriminate among these four formats. What factors determine which of the four available structural formats will be adopted by an arbitrary chain segment?

Selection of a secondary structure format is likely mediated by a redundant stereochemical code that arises from the interplay between the shape and polarity of residue side chains and secondary-structure conformation. Initial evidence for such a code is now emerging in the case of helices. Capping of helix termini involves residue polarity (128, 141), while residue preferences at central positions of the helix are governed largely by side-chain shape (33, 186, 187), as described next.

The key feature of the Pauling-Corey-Branson α-helix is a repeating

pattern of main-chain hydrogen bonds between each amide hydrogen and the carbonyl oxygen located four residues upstream (122). However, this feature fails to discriminate between helix and other categories of secondary structure because all residues (except proline) have identical backbones.

A likely source of structural discrimination is found in helix capping (128, 141). In the helix backbone, the first four amide hydrogens and last four carbonyl oxygens necessarily lack intrahelical hydrogen-bond partners. Since the mean length of a helix in a globular protein is 12 residues, these 8 unsatisfied groups represent 50% of main-chain hydrogen bonding, on average. It has been hypothesized that hydrogen bonds are essential for these otherwise unsatisfied backbone amides and, typically, are provided by side chains of polar residues that flank the helix termini. Experimental evidence in favor of this hypothesis has been accumulating, as described in the section on hydrogen bonding. Recent NMR studies of peptides (97) have been particularly convincing in this regard.

In addition to capping at helix termini, side-chain shape appears to play a discriminatory role in central residues of the helix (33, 186, 187). For example, valine can populate all three side-chain conformers (*gauche*+, *gauche*−, *trans*) in the unfolded state, but this β-branched side chain is restricted to essentially one conformer (*trans*) in a helix, as shown in Figure 6. The corresponding energy loss ($T\Delta S$), due solely to the reduction in

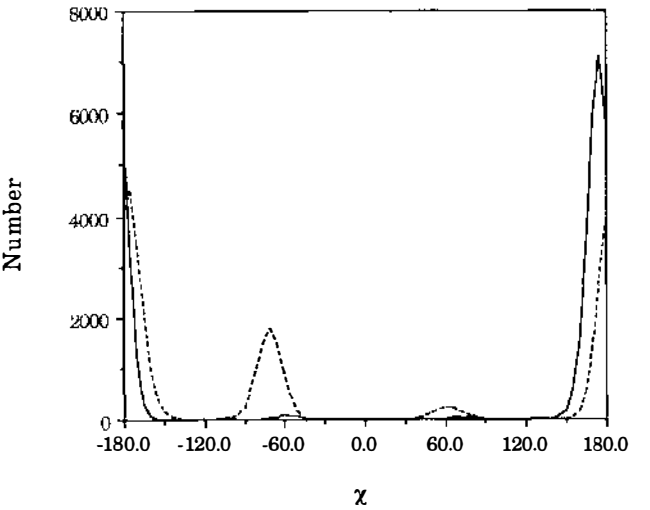

Figure 6 Rotamer distributions for Val in both the helical state (*solid line*) and the standard state (*dashed line*) (33). See text for further details.

side-chain configurational entropy, is $\sim RT$ ln 3, on the order of physiological RT. Such energy terms can cause residues to favor one type of secondary structure over others, and, although modest individually, they can be substantial in the aggregate.

In a stereochemical code, such as the one described here for helices, each residue makes an energetically small contribution to conformation: e.g. a single H-bond for a polar residue at a helix cap or an (unfavorable) entropic contribution of $\sim RT$ ln 3 for a Val in the center of the helix. It is the aggregate effect of these individual terms, summed over the entire molecule, that determines (and probably overdetermines) conformation. One prediction from this type of distributed code is a scarcity of conformational tender spots wherein single mutations result in major disruption of the overall fold. Currently, the prediction is in good agreement with experiment (1, 46, 95, 102, 103). Even in extreme cases, the protein apparently can tolerate a major packing fault without precipitating global rearrangement (45).

This postulated stereochemical code, with the protein fold determined by many individually small terms, represents a robust folding mechanism. One might think that the existence of conserved residues in proteins argues against these ideas, but on further consideration, the opposite appears to be true. Of approximately 450 globins of known sequence, only one residue (i.e. the proximal histidine) is absolutely conserved (11), ostensibly for functional, not conformational, reasons. Typically, the mutational experiments of nature, as well as those in the laboratory, do not dead-end in conformational catastrophe, thereby allowing evolution to proceed apace.

ACKNOWLEDGMENTS

We thank Eaton Lattman for helpful discussion, Don Engelman for forbearance, and Robert Baldwin and Neville Kallenbach for preprints of papers. This review is dedicated to Kensal E. Van Holde on the occasion of his sixty-fifth birthday. Support from grants NIH (GM 29459 to G. R.) and NSF (DMB 8705085 to R. W.) is acknowledged.

Literature Cited

1. Alber, T. 1989. *Annu. Rev. Biochem.* 58: 765–98
2. Anfinsen, C. B. 1973. *Science* 181: 223–30
3. Anfinsen, C. B., Haber, M., Sela, M., White, F. H. Jr. 1961. *Proc. Natl. Acad. Sci. USA* 47: 1309–14
4. Anson, M. L., Mirsky, A. E. 1925. *J. Gen. Physiol.* 9: 169–79
5. Argos, P., Palau, J. 1982. *Int. J. Pept. Protein Res.* 19: 380–93
6. Aronow, R. H., Witten, L. 1960. *J. Phys. Chem.* 64: 1643–48
7. Artymiuk, P. J., Blake, C. C. F. 1981. *J. Mol. Biol.* 152: 737–62
8. Baker, E. N., Hubbard, R. E. 1984. *Prog. Biophys. Mol. Biol.* 44: 97–179
9. Baldwin, R. L. 1986. *Proc. Natl. Acad. Sci. USA* 83: 8069–72
10. Barlow, D. J., Thornton, J. M. 1988. *J. Mol. Biol.* 201: 601–19
11. Bashford, D., Chothia, C., Lesk, A. M.

1987. *J. Mol. Biol.* 196: 199–216
12. Behe, M. J., Lattman, E. E., Rose, G. D. 1991. *Proc. Natl. Acad. Sci. USA* 88: 4195–99
13. Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A., Matthews, B. W. 1992. *Biochemistry* 31: 3590–96
14. Bierzynski, A., Kim, P. S., Baldwin, R. L. 1982. *Proc. Natl. Acad. Sci. USA* 79: 2470–74
15. Bigelow, C. C. 1967. *J. Theor. Biol.* 16: 187–211
16. Bone, R., Cullis, P., Wolfenden, R. 1983. *J. Am. Chem. Soc.* 105: 1339–43
17. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., Sauer, R. T. 1990. *Science* 247: 1306–10
18. Brant, D. A., Miller, W. G., Flory, P. J. 1967. *J. Mol. Biol.* 23: 47–65
19. Brown, J. E., Klee, W. A. 1971. *Biochemistry* 10: 470–76
20. Bruch, M. D., Dhingra, M. M., Gierasch, L. M. 1991. *Proteins* 10: 130–39
20a. Butler, J. A. V., Ramchandani, C. N. 1935. *J. Chem. Soc.* pp. 952–55
21. Cabani, S., Gianni, P. 1979. *J. Chem. Soc. Faraday Trans.* 1 75: 1184–95
21a. Cabani, S., Gianni, P., Mollica, L. 1981. *J. Sol. Chem.* 10: 563
22. Carter, P., Wells, J. A. 1987. *Science* 237: 394–99
23. Ceccarelli, C., Jeffrey, G. A., Taylor, R. 1981. *J. Mol. Struct.* 70: 255–71
24. Chothia, C. 1974. *Nature* 248: 338–39
25. Chothia, C. 1976. *J. Mol. Biol.* 105: 1–14
26. Chothia, C. 1983. *J. Mol. Biol.* 163: 107–17
27. Chothia, C., Finkelstein, A. V. 1990. *Annu. Rev. Biochem.* 59: 1007–39
28. Chothia, C., Levitt, M., Richardson, D. 1977. *Proc. Natl. Acad. Sci. USA* 74: 4130–34
29. Cohen, F. E., Kuntz, I. D. 1989. See Ref. 49, pp. 647–705
30. Cohn, E. J., Edsall, J. T. 1943. *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions.* Princeton, NJ: Van Nostrand–Reinhold
31. Cornette, J. L., Cease, K. B., Margalit, H., Spouge, J. L., Berzofsky, J. A., DeLisi, C. 1987. *J. Mol. Biol.* 195: 659–85
32. Cramer, R. D. 1977. *J. Am. Chem. Soc.* 99: 5408–12
33. Creamer, T. P., Rose, G. D. 1992. *Proc. Natl. Acad. Sci. USA* 89: 5937–41
34. Creighton, T. E. 1990. *Biochem J.* 270: 1–16
35. Crippen, G. M. 1978. *J. Mol. Biol.* 126: 315–32
36. Dill, K. A. 1990. *Biochemistry* 29: 7133–55
37. Dill, K. A. 1990. *Science* 250: 297
38. Doig, A. J., Williams, D. H. 1992. *J. Am. Chem. Soc.* 114: 338–43
39. Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., et al. 1981. *FEBS Lett.* 136: 311–15
40. Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E., Lerner, R. A. 1985. *Nature* 318: 480–83
41. Edsall, J. T. 1935. *J. Am. Chem. Soc.* 57: 1506–7
42. Edsall, J. T., McKenzie, H. A. 1978. *Adv. Biophys.* 10: 137–207
43. Eisenberg, D., McLachlan, A. D. 1986. *Nature* 319: 199–203
44. Engelman, D. M., Steitz, T. A., Goldman, A. 1986. *Annu. Rev. Biophys. Biophys. Chem.* 15: 321–53
45. Eriksson, A. E., Baase, W. A., Wozniak, J. A., Matthews, B. W. 1992. *Nature* 355: 371–73
46. Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, D. B., et al. 1986. *Science* 233: 659–63
47. Evans, D. F., Chen, S., Schriver, G. W., Arnett, E. M. 1981. *J. Am. Chem. Soc.* 103: 481–82
48. Fairman, R., Shoemaker, K. R., York, E. J., Steward, J. M., Baldwin, R. L. 1989. *Proteins* 5: 1–7
49. Fasman, G. D., ed. 1989. *Prediction of Protein Structure and the Principles of Protein Conformation.* New York: Plenum
50. Fauchère, J.-L., Pliska, V. E. 1983. *Eur. J. Med. Chem. Chim. Ther.* 18: 369–75
51. Fersht, A. R. 1987. *Trends Biochem. Sci.* 12: 301–4
52. Fersht, A. R., Shi, J.-P., Jack, K.-J., Lowe, D. M., Wilkinson, A. J., et al. 1985. *Nature* 314: 235–38
53. Deleted in proof
54. Fetrow, J. S., Zehfus, M. H., Rose, G. D. 1988. *Bio/Technology* 6: 167–71
55. Frank, H. S., Evans, M. W. 1945. *J. Chem. Phys.* 13: 507–32
56. Frauenfelder, H., Sligar, S. G., Wolynes, P. G. 1991. *Science* 254: 1598–1603
57. Gellman, S. H., Adams, B. R., Dado, G. P. 1990. *J. Am. Chem. Soc.* 112: 460–61
58. Gething, M.-J., Sambrook, J. 1992. *Nature* 355: 33–45
59. Gibbs, P. R., Radizick, A., Wolfenden, R. 1991. *J. Am. Chem. Soc.* 113: 4714–15
60. Gierasch, L. M., King, J., eds. 1990. *Protein Folding.* Washington, DC: Am. Assoc. Adv. Sci.

61. Gooley, P. R., Carter, S. A., Fagerness, P. E., MacKenzie, N. E. 1988. *Proteins* 4: 48–55
62. Goto, Y., Fink, A. L. 1989. *Biochemistry* 28: 945–52
63. Hildebrand, J. H. 1978. *Proc. Natl. Acad. Sci. USA* 76: 1–194
64. Hopp, T. P., Woods, K. R. 1981. *Proc. Natl. Acad. Sci. USA* 78: 3824–28
65. Howarth, O. W. 1975. *J. Chem. Soc. Faraday Trans.* 1 71: 2303–9
66. Ippolito, J. A., Alexander, R. S., Christianson, D. W. 1990. *J. Mol. Biol.* 215: 457–71
67. Janin, J. 1976. *J. Mol. Biol.* 105: 13–14
68. Janin, J., Chothia, C. 1980. *J. Mol. Biol.* 143: 95–128
69. Janin, S., Wodak, S., Levitt, M., Maigret, B. 1978. *J. Mol. Biol.* 125: 357–86
70. Jeffrey, G. A., Maluszynska, H. 1982. *Int. J. Biol. Macromol.* 4: 173–85
71. Jeffrey, G. A., Saenger, W. 1991. *Hydrogen Bonding in Biological Structures.* New York: Springer-Verlag
72. Jencks, W. P. 1966. In *Current Aspects of Biochemical Energetics*, ed. N. O. Kaplan, E. P. Kennedy, p. 273. New York: Academic
73. Jeng, M.-F., Englander, S. W. 1991. *J. Mol. Biol.* 221: 1045–61
74. Kauzmann, W. 1959. *Adv. Protein Chem.* 16: 1–64
75. Kauzmann, W., Bodanszky, A., Rasper, J. 1962. *J. Am. Chem. Soc.* 84: 1777–88
76. Kim, P. S., Baldwin, R. L. 1982. *Annu. Rev. Biochem.* 51: 459–89
77. Kim, P. S., Baldwin, R. L. 1990. *Annu. Rev. Biochem.* 59: 631–60
78. King, J. 1989. *Chem. Eng. News* 67: 32–54
79. Klotz, I. M., Franzen, J. S. 1962. *J. Am. Chem. Soc.* 84: 3461–66
80. Kossiakoff, A. A., Shteyn, S. 1984. *Nature* 311: 582–83
81. Kuntz, I. D. 1972. *J. Am. Chem. Soc.* 94: 8568–72
82. Kuwajima, J. 1989. *Proteins* 6: 87–103
83. Kyte, J., Doolittle, R. F. 1983. *J. Mol. Biol.* 157: 105–32
84. Lambert, J. D. 1953. *Discuss. Faraday Soc.* 15: 226–36
85. Lattman, E. E., Rose, G. D. 1993. *Proc. Natl. Acad. Sci. USA* 90: 439–41
86. Lau, K. F., Dill, K. A. 1989. *Macromolecules* 22: 3986–97
86a. Lecomte, J. T. J., Moore, C. D. 1991. *J. Am. Chem. Soc.* 113: 9663–65
87. Lee, B. K., Richards, F. M. 1971. *J. Mol. Biol.* 55: 379–400
88. Lee, C., Subbiah, S. 1991. *J. Mol. Biol.* 217: 373–88
89. Lesk, A. M., Chothia, C. 1980. *J. Mol. Biol.* 136: 225–70
90. Lesser, G. J., Rose, G. D. 1990. *Proteins* 8: 6–13
91. Leszczynski, J. F., Rose, G. D. 1986. *Science* 234: 849–55
92. Levitt, M., Chothia, C. 1976. *Nature* 261: 552–58
93. Lewis, C. A. Jr., Wolfenden, R. 1973. *J. Am. Chem. Soc.* 95: 6685–88
94. Deleted in proof
95. Lim, W. A., Sauer, R. T. 1991. *J. Mol. Biol.* 219: 359–76
96. Livingstone, J. R., Spolar, R. S., Record, M. T. Jr. 1991. *Biochemistry* 30: 4237–44
97. Lyu, P., Wemmer, D. E., Zhou, H. X., Pinker, R. J., Kallenbach, N. R. 1992. *Biochemistry.* In press
98. Lyu, P. C., Liff, M. I., Marky, L. A., Kallenbach, N. R. 1990. *Science* 250: 669–73
99. Lyu, P. C., Zhou, H. X., Jelveh, N., Wemmer, D. E., Kallenbach, N. R. 1992. *J. Am. Chem. Soc.* 114: 6560–62
100. MacGregor, M. J., Islam, S. A., Sternberg, M. J. E. 1987. *J. Mol. Biol.* 198: 295–310
101. Marsh, R. E., Donohue, J. 1967. *Adv. Protein Chem.* 22: 235–56
101a. Mathews, C. K., van Holde, K. E., 1990. *Biochemistry.* Redwood City, CA: Benjamin/Cummings
102. Matouschek, A., Kellis, J. T., Serrano, L., Fersht, A. R. 1989. *Nature* 340: 122–26
103. Matthews, B. W. 1987. *Biochemistry* 26: 6885–88
104. McCammon, J. A., Harvey, S. C. 1987. *Dynamics of Proteins and Nucleic Acids.* Cambridge: Cambridge Univ. Press
105. Meot-ner, M., Sieck, L. W. 1986. *J. Am. Chem. Soc.* 108: 7525–29
106. Merutka, G., Lipton, W., Shalongo, W., Park, S.-H., Stellwagen, E. 1990. *Biochemistry* 29: 7511–15
107. Miller, S., Janin, J., Lesk, A. M., Chothia, C. 1987. *J. Mol. Biol.* 196: 641–56
108. Mirsky, A. E., Pauling, L. 1936. *Proc. Natl. Acad. Sci. USA* 22: 439–47
109. Murphy, K. P., Privalov, P. L., Gill, S. J. 1990. *Science* 247: 559–61
110. Murray-Rust, P., Glusker, J. P. 1984. *J. Am. Chem. Soc.* 106: 1018–25
111. Nall, B. T., Dill, K. A., eds. 1991. *Conformations and Forces in Protein Folding.* Washington, DC: Am. Assoc. Adv. Sci.
112. Nemethy, G., Scherega, H. A. 1962. *J. Chem. Phys.* 36: 3382–3400
113. Nozaki, Y., Tanford, C. 1971. *J. Biol. Chem.* 246: 2211–17

114. Oas, T. G., Kim, P. S. 1988. *Nature* 336: 42–48
115. O'Neil, K. T., DeGrado, W. F. 1990. *Science* 250: 646–51
116. Osinga, M. 1979. *J. Am. Chem. Soc.* 101: 1621–22
117. Pace, C. N. 1990. *TIBS* 15: 14–17
118. Pace, C. N. 1992. *J. Mol. Biol.* 226: 29–35
119. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., Baldwin, R. L. 1990. *Nature (London)* 344: 268–70
120. Pauling, L. 1939. *The Nature of the Chemical Bond.* Ithaca, NY: Cornell Univ. Press
121. Pauling, L., Corey, R. B. 1951. *Proc. Natl. Acad. Sci. USA* 37: 251–56
122. Pauling, L., Corey, R. B., Branson, H. R. 1951. *Proc. Natl. Acad. Sci. USA* 37: 205–11
123. Penrose, R. 1989. *Aperiodicity and Order* 2, ed. M. Jaric. New York: Academic
124. Perutz, M. F. 1951. *Nature* 167: 1053–54
125. Pimentel, G. C., McClellan, A. L. 1960. *The Hydrogen Bond.* San Francisco: Freeman
126. Pinker, R. J., Rose, G. D., Kallenbach, N. R. 1993. *J. Mol. Biol.* Submitted
127. Ponder, J. W., Richards, F. M. 1987. *J. Mol. Biol.* 193: 775–91
128. Presta, L. G., Rose, G. D. 1988. *Science* 240: 1632–41
129. Privalov, P. L. 1979. *Adv. Protein Chem.* 33: 167–241
130. Privalov, P. L., Gill, S. J. 1988. *Adv. Protein Chem.* 39: 191–243
131. Privalov, P. L., Gill, S. J., Murphy, K. P. 1990. *Science* 250: 297–98
132. Privalov, P. L., Khechinashvili, N. N. 1974. *J. Mol. Biol.* 86: 665–84
133. Ptitsyn, O. B., Rashin, A. A. 1975. *Biophys. Chem.* 3: 1–20
134. Quiocho, F. A. 1986. *Annu. Rev. Biochem.* 55: 287–315
135. Radzicka, A., Wolfenden, R. 1988. *Biochemistry* 27: 1664–70
136. Richards, F. M. 1974. *J. Mol. Biol.* 82: 1–14
137. Richards, F. M. 1977. *Annu. Rev. Biophys. Bioeng.* 6: 151–76
138. Richards, F. M. 1991. *Sci. Am.* 264: 54–63
139. Richards, F. M., Kundrot, C. E. 1988. *Proteins Struct. Funct. Genet.* 3: 71–84
140. Richardson, J. S. 1981. *Adv. Protein Chem.* 34: 167–339
141. Richardson, J. S., Richardson, D. C. 1988. *Science* 240: 1648–52
142. Richmond, T. J., Richards, F. M. 1978. *J. Mol. Biol.* 119: 537–55
143. Deleted in proof
144. Rose, G. D. 1978. *Nature* 272: 586–90
145. Rose, G. D. 1979. *J. Mol. Biol.* 134: 447–70
146. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., Zehfus, M. H. 1985. *Science* 229: 834–38
147. Rose, G. D., Gierasch, L. M., Smith, J. A. 1985. *Adv. Protein Chem.* 37: 1–109
148. Rose, G. D., Roy, S. 1980. *Proc. Natl. Acad. Sci. USA* 77: 4643–47
149. Salemme, F. R. 1983. *Prog. Biophys. Mol. Biol.* 42: 95–133
150. Sandberg, W. S., Terwilliger, T. C. 1990. *Science* 245: 54–57
151. Schellman, J. A. 1955. *C. R. Trav. Lab. Carlsburg Ser. Chim.* 29: 223–29
152. Scholtz, J. M., Baldwin, R. L. 1992. *Annu. Rev. Biophys. Biomol. Struct.* 21: 95–118
153. Schulz, G. E., Schirmer, R. H. 1979. In *Principles of Protein Structure*, pp. 166–205. New York: Springer-Verlag
154. Serrano, L., Fersht, A. R. 1989. *Nature* 342: 296–99
155. Serrano, L., Kellis, J. T. Jr., Cann, P., Matouschek, A., Fersht, A. R. 1992. *J. Mol. Biol.* 224: 783–804
156. Sharp, K. A., Nicholls, A., Fine, R. F., Honig, B. 1991. *Science* 252: 106–9
157. Sharp, K. A., Nicholls, A., Friedman, R., Honig, B. 1991. *Biochemistry* 30: 9686–97
158. Shirley, B. A., Stanssens, P., Hahn, U., Pace, C. N. 1992. *Biochemistry* 31: 725–32
159. Shortle, D., Stiles, W. E., Meeker, A. K. 1990. *Biochemistry* 29: 8033–41
160. Singh, J., Thornton, J. M., Snarey, M., Campbell, S. F. 1987. *FEBS Lett.* 224: 161–71
161. Small, D. M. 1986. *The Physical Chemistry of Lipids.* New York: Plenum
162. Stahl, N., Jencks, W. P. 1986. *J. Am. Chem. Soc.* 108: 4196–4205
163. Stickle, D. F., Presta, L. G., Dill, K. A., Rose, G. D. 1992. *J. Mol. Biol.* 226: 1143–59
164. Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Konishi, Y., Scheraga, H. A. 1984. *Macromolecules* 17: 148–55
165. Sun, Y., Spellmeyer, D., Pearlman, D. A., Kollman, P. 1992. *J. Am. Chem. Soc.* 114: 6798–6801
166. Susi, H., Timasheff, S. N., Ard, J. S. 1964. *J. Biol. Chem.* 239: 3051–54
167. Tanford, C. 1978. *Science* 200: 1012–18
168. Tanford, C. 1980. *The Hydrophobic Effect.* New York: Wiley
169. Taylor, R., Kennard, O. 1984. *Acc. Chem. Res.* 17: 320–26

170. Taylor, R., Kennard, O., Versichel, W. 1983. *J. Am. Chem. Soc.* 105: 5761–66
171. Teller, D. C. 1976. *Nature* 260: 729–31
172. Traube, J. 1891. *Liebigs Ann. Chem.* 265: 27–55
173. Varadarajan, R., Connelly, P. R., Sturtevant, J. M., Richards, F. M. 1992. *Biochemistry* 31: 1421–26
174. Vedani, A., Dunitz, J. D. 1985. *J. Am. Chem. Soc.* 107: 7653–58
175. Voet, D., Voet, J. G. 1990. *Biochemistry*. New York: Wiley & Sons
176. Watson, J. D., Crick, F. H. C. 1953. *Nature* 171: 737–38
177. Weiner, S. J., Kollman, P. A., Case, D. A., Chandra Singh, U., Ghio, C., et al. 1984. *J. Am. Chem. Soc.* 106: 765–84
178. Wetlaufer, D. B. 1981. *Adv. Protein Chem.* 34: 61–92
179. Williams, D. H. 1992. *Aldrichimica Acta* 25: 9
180. Wolfenden, R. 1983. *Science* 222: 1087–93
181. Wolfenden, R., Andersson, L., Cullis, P. M., Southgate, C. B. 1981. *Biochemistry* 20: 849–55
182. Wolfenden, R., Lewis, C. A. Jr. 1976. *J. Theor. Biol.* 59: 231–35
182a. Wolfenden, R., Liang, Y. L., Matthews, M., Williams, R. 1987. *J. Am. Chem. Soc.* 109: 463–66
183. Wolfenden, R. V., Kati, W. M. 1991. *Acc. Chem. Res.* 24: 209–15
184. Woolfson, D. N., Cooper, A., Harding, M. M., Williams, D. H., Evans, P. A. 1993. *Science*. In press
185. Wüthrich, K., Wagner, G. 1978. *TIBS* 3: 227–30
186. Yun, R. H., Anderson, A., Hermans, J. 1991. *Proteins* 10: 219–28
187. Yun, R. H., Hermans, J. 1991. *Protein Eng.* 4: 761–66
188. Zehfus, M. H. 1987. *Proteins* 2: 90–110
189. Zehfus, M. H., Rose, G. D. 1986. *Biochemistry* 25: 5759–65

Annual Review of Biophysics and Biomolecular Structure
Volume 22, 1993

CONTENTS

(*continued*) v