Registry No. ATP synthase, 37205-63-3.


REFERENCES

Bazzo, R., Tappin, M. J., Pastore, A., Harvey, T. S., Carver, J. A., & Campbell, I. D. (1988) *Eur. J. Biochem. 173*, 139-146.

Breeze, A. L., Harvey, T. S., Bazzo, R., & Campbell, I. D. (1991) *Biochemistry 30*, 575-582.

Clore, G. M., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 7387-7401.

Cox, G. B., Fimmel, A. L., Gibson, F., & Hatch, L. (1986) *Biochim. Biophys. Acta 849*, 62-69.

Dempsey, C. E., Bazzo, R., Harvey, T. S., Syperek, I., Boheim, G., & Campbell, I.D. (1991) *FEBS Lett. 281*, 240-244.

Fillingame, R. H. (1976) *J. Biol. Chem. 251*, 6630-6637.

Foster, D. L., & Fillingame, R. H. (1982) *J. Biol. Chem. 257*, 2009-2015.

Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., & Fenney, J. (1990) *J. Magn. Reson. 90*, 420-425.

Gronenborn, A. M., Bax, A., Wingfield, P. T., & Clore, G. M. (1989) *FEBS Lett. 243*, 93-98.

Hertzeberg, E. L., & Hinkle, P. C. (1974) *Biochem. Biophys. Res. Commun. 58*, 178-184.

Hoppe, J., & Sebald, W. (1984) *Biochim. Biophys. Acta 768*, 1-27.

Hoppe, J., Schairer, H. U., & Sebald, W. (1980) *Eur. J. Biochem. 112*, 17-24.

Kagawa, Y., & Racker, E. (1966) *J. Biol. Chem. 241*, 2475-2482.

Kay, L. E., Brooks, B., Sparks, S. W., Torchia, D. A., & Bax, A. (1989a) *J. Am. Chem. Soc. 111*, 5488-5490.

Kay, L. E., Torchia, D. A., Bax, A., & Lewis, E. K. (1989b) *Biochemistry 28*, 8972-8979.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mao, D., Wachter, E., & Wallace, B. A. (1982) *Biochemistry 21*, 4960-4968.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989) *Biochemistry 28*, 6150-6156.

Moody, M. F., Jones, P. T., Carver, J. A., Boyd, J., & Campbell, I. D. (1987) *J. Mol. Biol. 193*, 759-774.

Norwood, T. J., Boyd, J., & Campbell, I. D. (1989) *FEBS Lett. 255*, 369-371.

Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N., & Campbell, I. D. (1990a) *J. Magn. Reson. 87*, 488-501.

Norwood, T. J., Boyd, J., Soffe, N., & Campbell, I. D. (1990b) *J. Am. Chem. Soc. 112*, 9638-9640.

Pastore, A., Harvey, T. S., Dempsey, C. E., & Campbell, I. D. (1989) *Eur. Biophys. J. 16*, 363-367.

Schneider, E., & Altendorf, K. (1985) *EMBO J. 5*, 515-518.

Sebald, W., & Hoppe, J. (1981) *Curr. Top. Bioenerg. 12*, 1-64.

Senior, A. E. (1983) *Biochem. Biophys. Acta 726*, 81-95.

Senior, A. E. (19885) *Physiol. Rev. 68*, 177-231.

Sigrist, H., Sigrist-Nelson, K., & Gitler, C. (1977) *Biochem. Biophys. Res. Commun. 74*, 178-184.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286-298.

Von Meyenberg, K., Jorgenson, B. B., & van Deurs, B. (1984) *EMBO J. 3*, 1791-1797.

Wüthrich, K. (1986) in *NMR of Proteins and Nucleic acids*, Wiley-Interscience, New York.

Yoshida, M., Okamoto, H., Sone, N., Hirata, H., & Kagawa, Y. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 936-940.


# Use of a Potential of Mean Force To Analyze Free Energy Contributions in Protein Folding†


F. Avbelj‡

*Institute of Chemistry, Hajdrihova 19, 61001 Ljubljana, Slovenia*



*Received October 21, 1991; Revised Manuscript Received March 19, 1992*



ABSTRACT: A method for calculation of the free energy of residues as a function of residue burial is proposed. The method is based on the potential of mean force, with a reaction coordinate expressed by residue burial. Residue burials are calculated from high-resolution protein structures. The largest individual contributions to the free energy of a residue are found to be due to the hydrophobic interactions of the nonpolar atoms, interactions of the main chain polar atoms, and interactions of the charged groups of residues Arg and Lys. The contribution to the free energy of folding due to the uncharged side chain polar atoms is small. The contribution to the free energy of folding due to the main chain polar atoms is favorable for partially buried residues and less favorable or unfavorable for fully buried residues. Comparison of the accessible surface areas of proteins and model spheres shows that proteins deviate considerably from a spherical shape and that the deviations increase with the size of a protein. The implications of these results for protein folding are also discussed.


The evaluation of the free energies of a protein in the denatured and folded conformations is crucial for the understanding the protein folding process. The free energy of a protein can be approximated by the contributions arising from hydrophobic interactions, configurational entropy, hydrogen bonding, electrostatic interactions of charged residues, and van der Waals' interactions. Unfortunately, the quantitative contributions of the configurational entropy and of the individual interactions to the free energy of a protein are not well understood.


† Supported by Science Foundation of Slovenia.



‡ Present address: Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, MD 20850.



0006-2960/92/0431-6290$03.00/0

Although the exact origin of the hydrophobic interactions and also its contribution to the free energy of a protein is controversial (Kauzmann, 1954; Dill, 1990; Ben-Naim, 1990), it has been considered as the main driving force in the protein folding process. From analysis of experimental data using transfer experiments, the free energy change due to the hydrophobic interactions has been found to be proportional to the change of accessible surface area ($A$):

$$\Delta G = K\Delta A \quad (1)$$

The constant $K$ has been estimated to be between 0.014 and 0.024 kcal/(mol·Å$^2$) (Chothia, 1974; Eisenberg & McLachlan, 1986; Murphy & Gill, 1991, 1990; Eriksson et al., 1992). A much larger value of the constant $K$ of 0.046–0.047 kcal/(mol·Å$^2$) was also proposed (Sharp et al., 1991).

The configurational entropy is the main interaction that opposes the folding of a protein. With the exception of some free rotating side chain atoms, the configurational entropy of the folded state is much smaller than of the unfolded state. The contribution of the configurational entropy to the free energy change in protein folding has been estimated to be 1.2 kcal/mol of residue (Karplus et al., 1987; Privalov, 1979).

The contribution of hydrogen bonding to the free energy change in protein folding is also controversial (Fersht, 1987; Privalov & Khechinashvili, 1974; Murphy et al., 1990; Privalov & Gill, 1988; Dill, 1990; Ben-Naim, 1991; Murphy & Gill, 1991; Shirley et al., 1992; Scholtz et al., 1991). It has been suggested that the hydrogen bonds do not contribute to the stability of proteins and that they play only a selective role preventing incorrect associations (Chothia & Janin, 1975; Fersht, 1984). The experimental values for the free energy contribution due to the formation of the hydrogen bond obtained from the free energies of the dimerization of small model molecules varies from negative to positive values (−2.8 to +1.9 kcal/mol) depending on the molecules studied (Schellman, 1955; Klotz & Franzen, 1962; Kresheck & Klotz, 1969; Gill & Noll, 1972; Susi, 1969; Doig & Williams, 1992). The stability of the hydrogen bonds also strongly depends on the solvent. For example, the stability of the $N$-methylacetamide dimers is small in aqueous solution (+3.1 kcal/mol), slight in dioxane (+0.39 kcal/mol), and large in carbon tetrachloride (−0.92 kcal/mol) (Klotz & Franzen, 1962).

Salt bridges and other electrostatic interactions of the charged residues are important for the stability of a protein. The contribution of these interactions depends on whether the charged residues are located on the surface or in the interior of a protein. Mutations that affect ion pairing have shown that the salt bridge formation on the surface of a protein stabilize proteins by 1–3 kcal/mol (Fersht, 1972; Perutz & Raidt, 1975), and in one case by as much as 3–5 kcal/mol (Anderson et al., 1990). Burying a salt bridge or a charged residue in the protein interior is very unfavorable. Using theoretical continuum models, the free energy change for burying a charge has been estimated to be ≈+19 kcal/mol and for burying an ion pair ≈+7 kcal/mol (Warshel et al., 1984; Honig & Hubbell, 1984; Honig, et al., 1986).

The main difficulty in obtaining the quantitative values for the free energy contributions of individual interactions in proteins is that small model molecules studied in polar and nonpolar solvents do not faithfully represent the interactions in the environment of a protein. The first reason for this is that the less polar solvents such as ethanol and octanol are a poor approximation for the environment in proteins (Murphy & Gill, 1990). The second and more important reason is that in proteins most of the residues (≈93%) are in contact with solvent to some degree and only ≈7% of the residues are fully buried in the protein interior (on the basis of a calculation of the accessible surface areas of residues in the high-resolution X-ray structures of 83 proteins). This surprisingly large percentage of the residues which are accessible to the solvent and the fact that the free energy contributions of individual interactions, particularly of the electrostatic interactions and hydrogen bonding (Warshel & Russell, 1984), strongly depend on the amount of solvent in close proximity indicate that the free energy contributions of the individual interactions must be described as a function of the amount of solvent in the vicinity.

In this work, a method is introduced which can be used to get the contributions of individual interactions to the free energy of a protein as a function of the amount of solvent in the vicinity. The free energies of residues as a function of the reaction coordinate were calculated using the potential of mean force $\Gamma(R)^r$ (Brooks et al., 1988). Transfer of a residue from the surface to the protein interior was used as a reaction coordinate (Rose et al., 1985; Prabhakaran & Ponnuswamy, 1979). The reaction coordinate was expressed by the residue burial (Rose et al., 1985). For sampling the configurational space along the reaction coordinate, experimental high-resolution X-ray structures were used. X-ray structures of 83 refined proteins with resolution better than 2 Å were taken from the Brookhaven Protein Data Bank. The potential of mean force can be used for calculation of the free energies of residues when two requirements are satisfied: the particles, i.e., residues, are independent of each other and in equilibrium with each other. Using the autocorrelation function of the residue burial along the sequence and the hydrophobic contribution of the large residues, we found out that both requirements are satisfied.

The contributions of the individual interactions to the free energy of residues were assumed to be group additive. The hydrophobic contribution of nonpolar atoms to the free energy of residues was calculated from the accessible surface areas of nonpolar atoms. The contributions of the polar atoms to the free energy of residues were then calculated from the functions $\Gamma(R)^r$ and the hydrophobic contribution of nonpolar atoms.

## Materials and Methods

*Accessible Surface Area, Residue Burial, and Atom Burial.* The accessible surface area $A$ of an atom is defined as the area over which the center of a water molecule (water radius 1.4 Å) can be placed while retaining van der Waals' contact with that atom and not penetrating any other atom. The accessible surface areas were calculated according to Lee and Richards (1971). A spacing between section planes of 0.1 Å was used. The hydrogen atoms were not considered individually but were included in the van der Waals' radii used for non-hydrogen atoms. The van der Waals' radii used were the following: oxygen, 1.40 Å; trigonal nitrogen, 1.65 Å; tetrahedral nitrogen, 1.50 Å; tetrahedral carbon, 1.87 Å; trigonal carbon, 1.76 Å; sulfur, 1.85 Å; water oxygen, 1.40 Å (Chothia, 1975).

The residue burial $R^i$ of the residue $i$ is defined as (Rose et al., 1985)

$$R^i = (1 - (A^i_{\text{protein}}/\bar{A}^r_{\text{standard}})) \quad (2)$$

where $A^i_{\text{protein}}$ is the accessible surface area of residue $i$ in the protein molecule and $\bar{A}^r_{\text{standard}}$ is the mean accessible surface area of a corresponding ensemble of residues of type $r$ in the stochastic standard state calculated by the method of Shrake and Rupley (1973), where it is assumed that the frequency with which a particular conformation is observed in a set of

structures reflects the population of that conformation in the denatured state. The accessible surface area of a residue in the stochastic standard state therefore approximates the average accessible area of a residue in the denatured state of a protein (Shrake & Rupley, 1973). Note that the superscript $i$ in eq 2 represents one particular residue in the protein molecule, while the superscript $r$ represents an ensemble of residues of the same type, for example all alanine residues.

The accessible surface area of the residue in the stochastic standard state (in eq 2) was calculated as a mean accessible surface area of an ensemble of actual tripeptides (U-X-Z) taken from the observed distribution in 83 proteins, rather than the tripeptides Gly-X-Gly used by Shrake and Rupley (1973). The accessible surface area of a residue calculated using tripeptides U-X-Z is ≈8.3% smaller than that using tripeptides Gly-X-Gly.

The atom burial $B^{i,j}$ is defined analogously to the residue burial:

$$B^{i,j} = (1 - (A^{i,j}_{\text{protein}}/\bar{A}^{r,s}_{\text{standard}})) \quad (3)$$

where $A^{i,j}_{\text{protein}}$ is the accessible surface area of atom $j$ of residue $i$ in the protein molecule and $\bar{A}^{r,s}_{\text{standard}}$ is the mean accessible surface area of corresponding atoms $s$ of residues of type $r$ in the stochastic standard state calculated by the method of Shrake and Rupley (1973). The superscript $i$ represents one particular residue in the protein molecule, while the superscript $r$ represents a corresponding ensemble of residues of the same type.

*Free Energy as a Function of Residue Burial.* The free energy $\Gamma(R)^r$ was calculated for residues of type $r$ as a function of the residue burial using the potential of mean force (Brooks et al., 1988):

$$\Gamma(R)^r = -kT \ln (g(R)^r) + C \quad (4)$$

where the superscript $r$ represents an ensemble of residues of the same type. $R$ is the reaction coordinate for burying a residue (eq 2), $g(R)^r$ is the probability distribution function for the occurrence of residues of type $r$ with the value $R$, $T$ is the temperature, $k$ is the Boltzmann constant, and $C$ is an undefined constant. Residue burial $R^i$ was calculated for 15 490 residues in 83 protein structures. The probabilities $g(R)^r$ for all 20 residue types were calculated by dividing the residue burial into 10 equal intervals. To avoid insignificant noise, only those $g(R)^r$ data were used that contained more than 10 residues in each of two consecutive intervals.

The error in $\Gamma(R)^r$ was estimated using two different methods. In the first method, the error was calculated using counting statistics (Wilson, 1952; Mandel, 1964):

$$\sigma_{\Gamma(R)^r} = \frac{kT}{(n(R))^{1/2}} \quad (5)$$

where $T$ is the temperature, $k$ is the Boltzmann constant, and $n(R)$ is the number of residues in an interval. For the least populated interval ($n = 11$) the error calculated by eq 5 is 0.18 kcal/mol. In the second method $\Gamma(R)^r$ was recalculated, with the binning intervals shifted by half of an interval (0.05) along the reaction coordinate. From the maximum differences between these two sets of $\Gamma(R)^r$, the error was estimated to be ≈0.2 kcal/mol.

To test the convergence of the method, we compared the functions $\Gamma(R)^r$ calculated from the set of 83 proteins and a smaller set of 42 proteins. The standard deviations between two corresponding functions were less than 0.1 kcal/mol for the more abundant residues such as Gly, Ala, and Ser and ≈0.3 kcal/mol for the less abundant residues such as His.

The behavior of the functions $\Gamma(R)^r$ was tested on the compact random coil structures (Gregoret & Cohen, 1991). For the compact random coil structures, we would expect that the functions $\Gamma(R)^r$ are independent of $R$ because there are no interactions between residues (except steric exclusions). Indeed, for a wide range of densities of the compact random coil structures which are close to the density of proteins, these functions were found to be independent of $R$. The analysis of these functions will be presented separately.

*Autocorrelation Function.* In order to test the condition of independence of residue burial the autocorrelation function of the residue burial along the sequence acf($u$) was calculated using eqs 6–8 (Jenkins & Watts, 1968):

$$\text{ac}f(u) = \frac{c(u)}{c(0)} \quad (6)$$

$$c(\text{u}) = \frac{1}{N}\sum_{i=1}^{N-u}(R^i - \bar{R})(R^{i+u} - \bar{R}) \quad (7)$$

$$\bar{R} = \frac{1}{N}\sum_{i=1}^{N}R^i \quad (8)$$

where $R^i$ is the residue burial of residue $i$ in the sequence, $\bar{R}$ is the mean residue burial, $u$ is the sequence interval, and $N$ is the number of residues. The autocorrelation function of the residue burial along the sequence was calculated from 79 proteins with more than 52 residues. Averaging over 11 530 origins was performed. The first and last residues were excluded from the averaging.

*Hydrophobic Contribution of Nonpolar Atoms.* For calculation of the free energy contribution of the hydrophobic interactions of nonpolar atoms to the free energy of residues, the constant $K$ in eq 1 was taken as 0.024 kcal/(mol·Å$^2$) (Chothia, 1974). Carbon was considered as a nonpolar atom, and oxygen, nitrogen, and sulfur were considered as polar atoms.

*Hydrogen Bonds.* Because protein structures in the Brookhaven Protein Data Bank usually do not contain hydrogens, the hydrogen atoms were added on the optimal positions according to the known stereochemistry around donor atoms. A contact between donor X and acceptor Y was treated as the hydrogen bond when three geometry requirements for X–H···Y–Z were satisfied (Chothia, 1975):

1. The distance between proton donor X and acceptor Y was less than 3.4 Å.
2. The angle X–H···Y was more than 120°.
3. The angle X···Y–Z was more than 90°.

*Volume of Proteins and Model Spherical Molecules.* The volume of a protein was calculated from the average residue volumes and the composition of the protein. Volumes were those of Chothia (1975), except for arginine which was taken to be 200 Å$^3$ (Teller, 1976).

To get the deviation of the shape of a protein from the ideal spherical shape, we introduced model spherical molecules. Model spherical molecules were cut out from a closed packed face centered cubic lattice. The volume was calculated from the volume occupied per atom in the face centered cubic lattice and the number of atoms making up the model spherical molecule.

## Results and Discussion

*Significance of the Free Energy Functions $\Gamma(R)^r$.* Consider a hypothetical protein folding process in which residues are independent of each other:

$$\text{denatured state} \rightleftharpoons \text{restricted state} \rightleftharpoons \text{folded state} \quad (9)$$

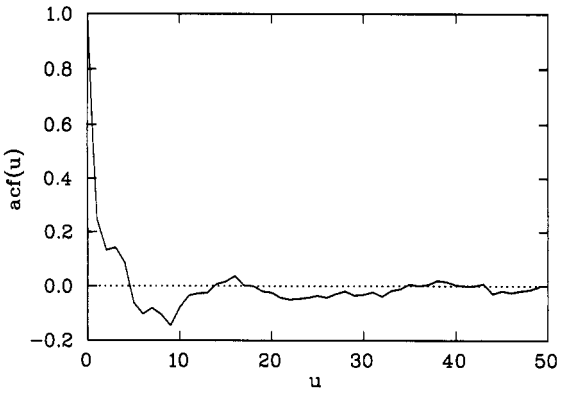


FIGURE 1: Autocorrelation function acf($u$) of the residue burial $R$ along a sequence.

In the first step of this process, all movement of the protein molecule is frozen, making the configurational entropy of the restricted state zero. (We neglect flexibility of some long side chains and exposed loops in folded proteins.) In the second step of this hypothetical process, the restricted state of each residue is transformed to the folded state by allowing interresidue interactions and equilibration along the reaction coordinate, i.e., residue burial. Some residues stay completely exposed, others become completely buried, but most of them adopt an intermediate position on the reaction coordinate. The distribution function along the reaction coordinate thus reflects the free energy of the residues in the folded state.

We express the mean free energies of an ensemble of residues of the type $r$, for example, all alanine residues, in the denatured $\bar{G}^r_{\text{denatured}}$, restricted $\bar{G}^r_{\text{restricted}}$, and folded $\bar{G}(R)^r_{\text{fold}}$ states relative to the restricted state:

$$\Delta\bar{G}^r_{\text{denatured}} = \bar{G}^r_{\text{restricted}} - \bar{G}^r_{\text{denatured}} = \Delta\bar{G}^r_{\text{entropy}} \quad (10)$$

$$\Delta\bar{G}(R)^r_{\text{fold}} = \bar{G}(R)^r_{\text{fold}} - \bar{G}^r_{\text{restricted}} \quad (11)$$

The difference between the mean free energies of the restricted and denatured states is the contribution of the configurational entropy of residues in the first step of the hypothetical protein folding process $\Delta\bar{G}^r_{\text{entropy}}$. We assume that the configurational entropy of a protein can be separated into the contributions from residues of different types $r$. The difference between the free energies of the folded and restricted states, $\Delta\bar{G}(R)^r_{\text{fold}}$, is the mean free energy change of residues of type $r$ in the second step of the hypothetical protein folding process.

The function $\Delta\bar{G}(R)^r_{\text{fold}}$ can in principle be obtained using the potential of mean force. Equation 4, defining the potential of mean force, is valid when the particles, i.e., residues, are independent of each other and in equilibrium with each other. The fulfillment of the first requirement would seem to be questionable because residues in proteins constitute a chain which implies a strong correlation of residue burial along a sequence. The autocorrelation function of the residue burial along a sequence (Figure 1) shows that the correlation coefficient for the first neighbors ($u = 1$) is only 0.24, indicating that the requirement for the independence of residue burial is satisfied. Proteins with known X-ray structures are stable molecules; therefore, we assume that the requirement for the equilibrium between residues is satisfied as well. We will see later that the validity of these requirements can be tested for the large non-polar residues.

For some residues, the requirements of independence and equilibrium may not apply. For example, to maintain specific structural features of a protein, other residues may contribute some free energy to keep the residues involved exposed or buried. Thus, proline is a hydrophobic residue and would therefore be expected to be buried in protein core. However, some proline residues form crucial parts of reverse turns in proteins and are therefore exposed to the solvent.

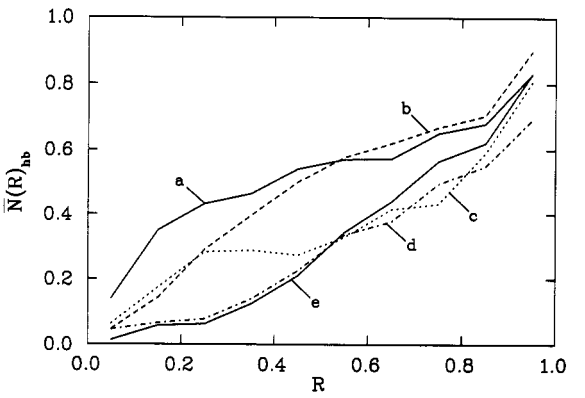


FIGURE 2: Mean number of hydrogen bonds per donor or acceptor $\bar{N}(R)_{hb}$ as a function of residue burial $R$ for (a) main chain N, (b) main chain O, (c) side chain O as donor, (d) side chain O as acceptor, and (e) side chain N.

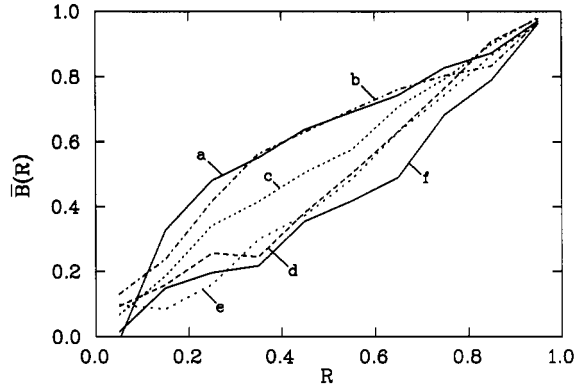


FIGURE 3: Mean atom burial $\bar{B}(R)$ for some types of atoms as a function of residue burial $R$: (a) main chain N, (b) main chain O, (c) side chain C, (d) neutral side chain N and O, (e) charged side chain N and O, and (f) lysine side chain N.

To obtain $\Delta\bar{G}(R)^r_{\text{fold}}$ from the potential of mean force $\Gamma(R)^r$, the constant $C$ in eq 4 has to be determined. If the values of the $\bar{G}^r_{\text{restricted}}$ and $\bar{G}(R)^r_{\text{fold}}$ at $R = 0$ are equal, the function $\Delta\bar{G}(R)^r_{\text{fold}}$ can be obtained from $\Gamma(R)^r$ by setting the constant $C$ in eq 4 to get zero free energy at $R = 0$. To find out whether the folded state at $R = 0$ in real protein structures corresponds to the restricted state, we analyzed the interactions of the most exposed residues ($R = 0$). Hydrogen bonds were assessed by the average number of hydrogen bonds per donor or acceptor $\bar{N}(R)_{hb}$ as a function of residue burial for all residues in 83 proteins studied (Figure 2). The mean number of hydrogen bonds per donor or acceptor for the most exposed residues ($R = 0$) is close to zero (smaller than 0.14); therefore, the contribution of internal hydrogen bonding to the free energy is also close to zero. (Note that zero burial ($R = 0$) is only an approximation because of the finite size of the intervals on the reaction coordinate.) Contributions from hydrophobic interactions and solvation of polar atoms were assessed by the average atom burial $\bar{B}(R)$ (see eq 3) for the different types of atoms as a function of residue burial for all residues in 83 proteins studied (Figure 3). The mean atom burial for the most exposed residues $\bar{B}(0)$ shows that for most atoms the deviations from zero are small (less than 0.1). From these data, it follows that the free energies of the restricted state and the folded state at $R = 0$ of neutral residues are equal, because in both states the contributions of the hydrogen bonding, hydrophobic interactions, and solvation of polar atoms

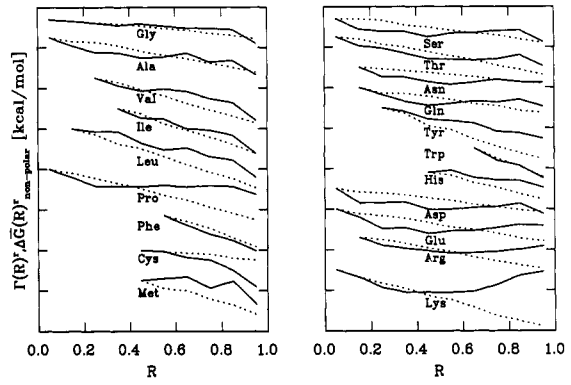


FIGURE 4: Functions $\Gamma(R)^r$ (solid lines) and $\Delta\bar{G}(R)^r_{nonpolar}$ (dashed lines) (in kilocalories per mole) for all residues $r$. We can arbitrarily translate the functions $\Gamma(R)^r$ up or down along the ordinate without changing the meaning of these functions (note constant $C$ in eq 4). The functions $\Gamma(R)^r$ are not defined for the complete reaction coordinate for some residues. The values of $\Gamma(R)^r$ and $\Delta\bar{G}(R)^r_{nonpolar}$ were set to be equal at the smallest value of $R$ for which the function $\Gamma(R)^r$ is defined. The distance between tick marks on the ordinate represents 2 kcal/mol of free energy.

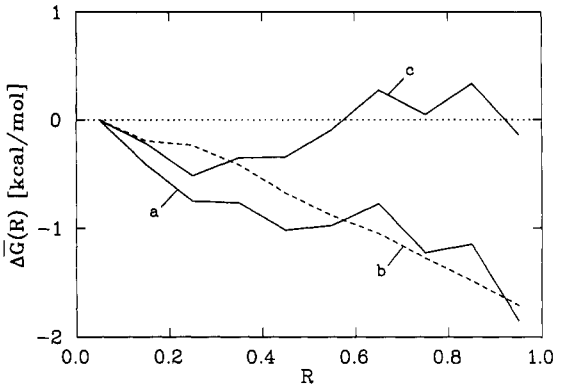


FIGURE 5: Free energy functions for alanine: (a) $\Delta\bar{G}(R)^{Ala}_{fold}$ (kilocalories per mole), (b) $\Delta\bar{G}(R)^{Ala}_{nonpolar}$ (kilocalories per mole), (c) $\Delta\bar{G}(R)^{Ala}_{polar\text{-}main}$ (kilocalories per mole).

are equal. $\Delta\bar{G}(R)^r_{fold}$ can therefore be obtained from $\Gamma(R)^r$.

The free energies of the charged residues for the most exposed residues ($R = 0$) are not necessarily equal to the free energies in the restricted state because the electrostatic interactions of the charged groups are long-range interactions which are strongly dependent on the amount of water in the vicinity (Warshel & Russell, 1984). The amount of water in the vicinity of the most exposed residues in the folded state is smaller than the amount of water in the vicinity of the residue in the denatured or restricted state.

If we assume that the free energy contributions to the total free energy of a residue from individual types of atoms are additive, the function $\Delta\bar{G}(R)^r_{fold}$ can be approximated as a sum:

$$\Delta\bar{G}(R)^r_{fold} = \Delta\bar{G}(R)^r_{nonpolar} + \Delta\bar{G}(R)^r_{polar\text{-}main} + \Delta\bar{G}^r_{polar\text{-}side} \quad (12)$$

where $\Delta\bar{G}(R)^r_{nonpolar}$ is the contribution from nonpolar atoms due to the hydrophobic interactions, $\Delta\bar{G}(R)^r_{polar\text{-}main}$ is the contribution from main chain polar atoms, and $\Delta\bar{G}(R)^r_{polar\text{-}side}$ is the contribution from side chain polar atoms. For charged residues, the contribution from side chain polar atoms $\Delta\bar{G}(R)^r_{polar\text{-}side}$ is replaced by the electrostatic contribution from the charged atoms $\Delta\bar{G}(R)^r_{charge}$.

*The Potential of Mean Force and the Hydrophobic Contribution to the Free Energy Agree for Large Residues.* The functions $\Gamma(R)^r$ and $\Delta\bar{G}(R)^r_{nonpolar}$ are shown in Figure 4. The hydrophobic contributions of the nonpolar atoms $\Delta\bar{G}(R)^r_{nonpolar}$ to $\Delta\bar{G}(R)^r_{fold}$ were calculated using differences between the accessible surface areas of the nonpolar atoms in proteins and the corresponding areas in the stochastic standard state. The constant $K$ in eq 1 was taken as 0.024 kcal/(mol·Å$^2$). For residues for which $\Gamma(R)^r$ at $R = 0$ is not defined, the values of functions $\Gamma(R)^r$ and $\Delta\bar{G}(R)^r_{nonpolar}$ were arbitrarily set to be equal at the smallest value of $R$ for which the function $\Gamma(R)^r$ is defined (Figure 4).

The population of large nonpolar residues on the most exposed part of the reaction coordinate is small, therefore $\Gamma(R)^r$ is not defined in that region of $R$ (Figure 4). This is a result of a large negative contribution of the hydrophobic interactions to the free energies of buried nonpolar residues. In the protein folding, most of the large nonpolar residues are removed from the water and assembled in the protein interior (Kauzmann, 1959).

We assume that for the large nonpolar residues, the free energy contribution of the polar atoms is small compared to the contribution of the nonpolar atoms. The free energy change along the reaction coordinate for these residues may therefore be considered to be dominated by the hydrophobic interactions. Indeed, the slope of the function $\Gamma(R)^r$ is almost equal to the slope of the function $\Delta\bar{G}(R)^r_{nonpolar}$ for the large residues Phe and Trp (Figure 4). The functions $\Gamma(R)^r$ and $\Delta\bar{G}(R)^r_{nonpolar}$ for the large nonpolar residues are not defined along the complete reaction coordinate; therefore, only the slopes of $\Gamma(R)^r$ and $\Delta\bar{G}(R)^r_{nonpolar}$ can be compared. The average slope of $\Gamma(R)^r$ for the large nonpolar residues Leu and Ile is smaller than the average slope of $\Delta\bar{G}(R)^r_{nonpolar}$, indicating that for the nonpolar nonaromatic residues the value of 0.017 kcal/(mol·Å$^2$) for the constant $K$ may be more accurate. This value of $K$ is very close to those obtained from the thermodynamic data on dissolution of the solid cyclic peptides (0.014–0.018 kcal/(mol·Å$^2$)) (Murphy & Gill, 1990, 1991). The agreement between the values of the constant $K$, one based on the thermodynamics of dissolution of the solid cyclic peptides and the other on the potential of mean force $\Gamma(R)^r$, is very important and indicates that the method for the calculation of the free energy functions $\Gamma(R)^r$ introduced in this work is appropriate.

The values of the constant $K$ obtained from the solid cyclic peptides and from the potentials of mean force in proteins do not agree with the recently corrected value of the constant $K$ of 0.046–0.047 kcal/(mol·Å$^2$) on the basis of liquid/liquid and liquid/vacuum transfer experiments (Sharp et al., 1991). This disagreement supports the conclusion that a vacuum or less polar solvents such as ethanol and octanol are a poor approximation for the environment in proteins.

*The Free Energy Contribution of Main Chain Polar Atoms Is Favorable for Partially Buried Residues and Less Favorable or Unfavorable for Fully Buried Residues.* For the nonpolar residues, the free energy contributions of the main chain polar atoms $\Delta\bar{G}(R)^r_{polar\text{-}main}$, which includes the contribution from hydrogen bonding and the free energy of solvation, can be calculated by subtracting $\Delta\bar{G}(R)^r_{nonpolar}$ from $\Delta\bar{G}(R)^r_{fold}$ (see eq 12). This is only possible for those nonpoalr residues for which the function $\Delta\bar{G}(R)^r_{fold}$ is defined along the whole reaction coordinate: Gly, Ala, and Pro (see Figure 4). $\Delta\bar{G}(R)^r_{polar\text{-}main}$ cannot be obtained for Gly and Pro because they are often an integral part of the reverse turns which are usually located on the protein surface. For those residues, the requirements of independence and equilibrium may not apply.

Figure 5 shows the functions $\Delta\bar{G}(R)^r_{fold}$, $\Delta\bar{G}(R)^r_{nonpolar}$, and $\Delta\bar{G}(R)^r_{polar\text{-}main}$ for alanine. The constant $K$ in eq 1 was taken as 0.024 kcal/(mol·Å$^2$). The function $\Delta\bar{G}(R)^r_{polar\text{-}main}$ is neg-

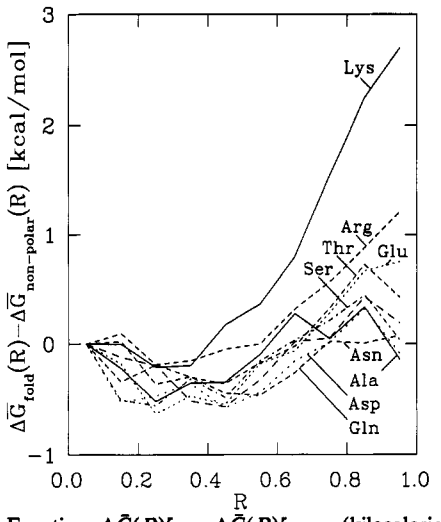


FIGURE 6: Functions $\Delta\bar{G}(R)^r_{fold} - \Delta\bar{G}(R)^r_{nonpolar}$ (kilocalories per mole) for Ala, Arg, Lys, Glu, Ser, Thr, Asn, Gln, and Asp.

ative for partially buried alanine residues ($0 < R < 0.55$) with the minimum free energy of ≈−0.5 kcal/mol. For fully buried residues ($0.55 < R < 0.9$), the function $\Delta\bar{G}(R)^r_{polar\text{-}main}$ is positive with the maximum free energy of ≈+0.3 kcal/mol. The function $\Delta\bar{G}(R)^r_{polar\text{-}main}$ shows that at $K = 0.024$ kcal/(mol·Å$^2$) burying main chain polar atoms is favorable for partially buried alanine residues and unfavorable for fully buried residues. If we use a smaller value of the constant $K$ to calculate $\Delta\bar{G}(R)^r_{nonpolar}$, the function $\Delta\bar{G}(R)^r_{polar\text{-}main}$ would become favorable along the entire reaction coordinate.

A similar functional dependence of the free energy contribution of the polar groups of atoms that are able to form hydrogen bonds was postulated by Ben-Naim (Ben-Naim, 1991), who found out that for the binding in which the two functional groups are only partly removed from the aqueous environment, their contribution to the free energy can be negative and of the order of ≈−1 kcal/mol. In contrast, when the two functional groups are completely removed from the aqueous environment, the contribution of the pair of polar functional groups to the free energy can be as high as +6.4 kcal/mol.

Results of the theoretical study by Sneddon et al. (1989) also show that the free energy change for the hydrogen bond formation between two formamide molecules in water is negative (−0.3 kcal/mol). The amide hydrogen-bonded dimer is more stable by 2.2 kcal/mol in water than in $CCl_4$.

*The Free Energy Contributions of Side Chain Polar Atoms Is Small for Uncharged but Large for Positively Charged Residues.* For the polar and charged residues, the function $\Delta\bar{G}(R)^r_{fold}$ can be approximated as a sum of the contributions from nonpolar atoms $\Delta\bar{G}(R)^r_{nonpolar}$, main chain polar atoms $\Delta\bar{G}(R)^r_{polar\text{-}main}$, and uncharged side chain polar atoms $\Delta\bar{G}(R)^r_{polar\text{-}side}$ or charged side chain polar atoms $\Delta\bar{G}(R)^r_{charge}$ (see eq 12).

In the hypothetical protein folding process (eq 9), we assume that all configurational entropy of the residues is lost in the first step of the process. The contribution of the configurational entropy to the functions $\Gamma(R)^r$ should therefore be zero. In the native structure of proteins, some long side chains and exposed loops are quite flexible, adding a new term to eq 12. To assess the role of the configurational entropy, we compared the functions $\Delta\bar{G}(R)^r_{fold}$ for the residues Asn and Gln, which differ by one methylene group and therefore one internal rotational degree of freedom. The largest difference between their functions $\Delta\bar{G}(R)^r_{fold}$ is only 0.4 kcal/mol (Figure 6). Similar behavior is observed for the charged residues Asp and Glu (Figure 6). We concluded that the contribution of the configurational entropy to the free energy along the reaction coordinate is probably small and that eq 12 without an additional term can be used.

The contribution of the uncharged side chain polar atoms may now be assessed by comparing the contributions of all polar atoms to the total free energy function of a residue $(\Delta\bar{G}(R)^r_{fold} - \Delta\bar{G}(R)^r_{nonpolar})$ for the residues: Ser, Thr, Asn, and nonpolar residue Ala (Figure 6). We found that these functions are very similar. The largest difference between them is smaller than 0.5 kcal/mol. Assuming that the contribution of the main chain polar atoms $\Delta\bar{G}(R)^r_{polar\text{-}main}$ to $\Delta\bar{G}(R)^r_{fold}$ is transferable from alanine to other similar residues, the contribution from the polar side chain atoms $\Delta\bar{G}(R)^r_{polar\text{-}side}$ for these residues must be small.

Comparing the contributions of the polar atoms to the free energy function $\Delta\bar{G}(R)^r_{fold} - \Delta\bar{G}(R)^r_{nonpolar}$ for the charged residues, we found a difference between the behavior of the positively and negatively charged residues. The functions $\Delta\bar{G}(R)^r_{fold} - \Delta\bar{G}(R)^r_{nonpolar}$ for the negatively charged residues are more similar to the free energy functions of the polar residues and alanine (Figure 6) than to the free energy functions of the positively charged residues, indicating that the free energy of burying a negatively charged group $\Delta\bar{G}(R)^r_{charge}$ is much smaller than the free energy of burying a positively charged group.

The functions $\Delta\bar{G}(R)^r_{fold} - \Delta\bar{G}(R)^r_{nonpolar}$ for the totally buried lysine and arginine are positive and large (Figure 6). We estimated the free energy of burying a positively charged group of lysine and arginine to be ≈+3.2 and ≈+2.1 kcal/mol, respectively, assuming that the free energies of a residue in the denatured state and the folded state at $R = 0$ are equal, that the free energy function of the main chain polar atoms of alanine is also valid for lysine and arginine, and that the energies of all hydrogen bonds are equal (including hydrogen bonds involving the charged side chain atoms).

The estimated values for the free energy of burying a charged group must be treated with caution, because the free energy of a residue in the restricted state is not necessarily equal to the free energy of folded state at $R = 0$ for these residues (see above). In addition, using the residue burial as a reaction coordinate for charged residues is an oversimplification. These electrostatic interactions are long-range interactions; therefore, the charged residues with zero accessible surface area can still interact substantially with the water molecules in the vicinity.

*Protein Structures Are Far from Being Globular.* Hydrophobic interactions of the nonpolar atoms tend to make proteins globular. To find out how much the shape of proteins deviates from the spherical shape, we compared the dependence of the accessible surface areas with the volume for monomeric proteins and for model spherical molecules. For a set of solid bodies having similar shape, the relation between the surface area and volume $V$ follows the law

$$\text{surface area} = aV^k \qquad (13)$$

with the coefficient $k$ equal to $^2/_3$. A similar relation between the accessible surface area and molecular weight has been found for proteins. For monomeric proteins the coefficient $k$ was found to be $^2/_3$ (Janin, 1976; Teller, 1976) and 0.73 (Miller et al., 1987).

For 55 monomeric proteins used in this work (Figure 7), the relation between the total accessible surface area $A_{protein}$ and volume $V$ can be expressed by eq 14 and linear eq 15, with residual standard deviations of 732.7 Å$^2$ and 680.7 Å$^2$, respectively. The volume of a protein was calculated from the

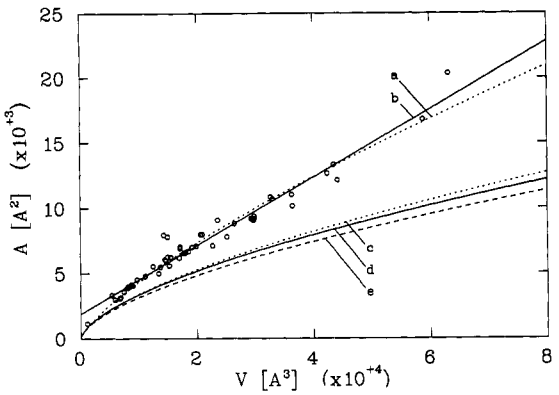


FIGURE 7: Accessible surface area $A$ [Å$^2$] as a function of volume $V$ [Å$^3$]: for 55 monomeric proteins (O), (a) fit for proteins with eq 14, (b) fit for proteins with eq 15, (c) fit for model spheres (atom radius 2.0 Å), (d) fit with eq 16 for model spheres (atoms radius 1.5 Å), and (e) fit for model spheres (atoms radius 1.0 Å).

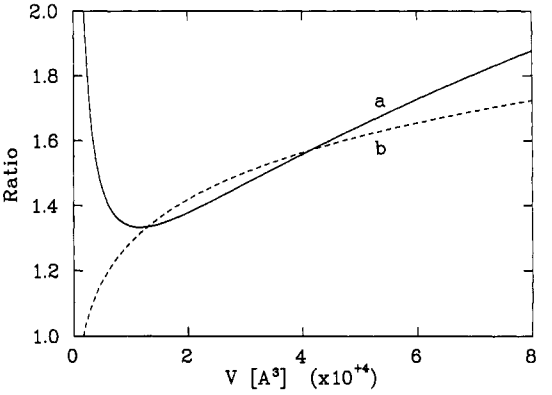


FIGURE 8: Ratio between the accessible surface areas [Å$^3$] of proteins and of model spherical molecules built from atoms with radius 1.5 Å as a function of volume $V$: (a) eq 14 divided by eq 16 and (b) eq 15 divided by eq 16.

average residue volumes and the composition of the protein (Chothia, 1975; Teller, 1976). Equivalent relations also exist between $A_{protein}$ and molecular weight:

$$A_{protein} = 4.03(\pm 0.92)V^{0.76(\pm 0.02)} \qquad \sigma = 732.7 \quad (14)$$

$$A_{protein} = 1885.8(\pm 172.4) + 0.261(\pm 0.007)V \qquad \sigma = 680.7 \quad (15)$$

The linear and power functions fit almost equally well the dependence of the accessible surface area of a protein with volume or molecular weight, although it seems that linear function is more appropriate for large proteins and power function more appropriate for small proteins (Figure 7).

Figure 7 also shows the accessible surface area as a function of volume for the model spherical molecules cut out from the closed packed face centered cubic lattice. The face centered cubic lattice was constructed from atoms with radii 1.0 Å, 1.5 Å, and 2.0 Å. The calculated accessible surface areas as a function of volume were fitted with the power functions (Figure 7) with the residual standard deviations 49.0 Å$^2$, 87.0 Å$^2$ (eq 16), and 144.4 Å$^2$, respectively.

$$A_{model\ sphere} = 11.447(\pm 0.550)V^{0.617(\pm 0.004)} \qquad \sigma = 87.0 \quad (16)$$

The accessible surface area of the model spherical molecules is much smaller than the accessible surface area of the corresponding proteins with the same volume (Figure 7). Excluding small polypeptides the ratio between the accessible surface areas of proteins and of model spherical molecules increases with the volume (Figure 8).

The increasing deviations from the spherical shape for large proteins show that in proteins strong interactions acting against the hydrophobic interactions must exist. The potential of mean force analysis shows that the largest opposing contributions arise from the free energy cost of burying the main chain polar and charged side chain atoms.

*Implications for Protein Folding.* We have found that the largest individual contributions to the free energy change of residues in folding $\Delta\bar{G}(R)'_{fold}$ are due to hydrophobic interactions $\Delta\bar{G}(R)'_{nonpolar}$, main chain polar atoms $\Delta\bar{G}(R)'_{polar\text{-}main}$, and electrostatic interactions $\Delta\bar{G}(R)'_{charge}$ of the residues Arg and Lys.

The contribution of the hydrophobic interactions $\Delta\bar{G}(R)'_{nonpolar}$ to the overall stability of a protein is favorable and large depending on the accessible surface area lost in folding. The contribution of the main chain polar atoms $\Delta\bar{G}(R)'_{polar\text{-}main}$ stabilizes the hydrogen bonded complex for partially buried residues, which may be important in the early stages of the protein folding process (Ben-Naim, 1991). Indeed, experimental data show that secondary structures are formed more rapidly than is the overall folding (Roder et al., 1988; Udgaonkar & Baldwin, 1988; Baum et al., 1989; Bycroft et al., 1990) and that formation of the $\alpha$-helix in synthetic polypeptides occurs on the time scale of $10^{-6}$–$10^{-8}$ s (Hammes & Roberts, 1969; Gruenewald et al., 1979). Because the residues arranged in the secondary structures contain more hydrogen bonds per donor or acceptor than the rest of the residues (helix 0.90 > sheet 0.75 > other 0.48; on the basis of the calculation of the number of hydrogen bonds in the high-resolution X-ray structures of 83 proteins), an early formation of the secondary structures can be explained by the negative value of $\Delta\bar{G}(R)'_{polar\text{-}main}$ for partially buried residues. For fully buried residues in later stages of the protein folding, the interactions of main chain polar and charged side chain atoms prevent formation of structures which are spherical.

## ACKNOWLEDGMENTS

I am very grateful to Prof. Dušan Hadži for his support and encouragement. I thank Prof. John Moult for critical reading of the manuscript. I also thank Osnat Herzberg, Darko Kocjan, and Janez Mavri for helpful discussions. This research was supported by the Science Foundation of Slovenia.

**Registry No.** Gly, 56-40-6; Ala, 56-41-7; Val, 72-18-4; Ile, 73-32-5; Leu, 61-90-5; Pro, 147-85-3; Phe, 63-91-2; Cys, 52-90-4; Met, 63-68-3; Ser, 56-45-1; Thr, 72-19-5; Asn, 70-47-3; Gln, 56-85-9; Tyr, 60-18-4; Trp, 73-22-3; His, 71-00-1; Asp, 56-84-8; Glu, 56-86-0; Arg, 74-79-3; Lys, 56-87-1.

## REFERENCES

Anderson, D. E., Becktel, W. J., & Dalquist, F. W. (1990) *Biochemistry 29*, 2403–2408.

Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7–13.

Ben-Naim, A. (1990) *Biopolymers 29*, 567–596.

Ben-Naim, A. (1991) *J. Phys. Chem. 95*, 1437–1444.

Brooks, C. L., Karplus, M., & Pettitt, B. M. (1988) *Proteins. A theoretical Perspective of Dynamics, Structure, and Thermodynamics*, John Wiley and Sons, Inc., New York.

Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488–490.

Chothia, C. (1974) *Nature 248*, 338–339.

Chothia, C. (1975) *Nature 254*, 304–308.

Chothia, C. H., & Janin, J. (1975) *Nature 256*, 705–708.

Dill, K. A. (1990) *Biochemistry 29*, 7133–7155.

Doig, A. J., & Williams, D. H. (1992) *J. Am. Chem. Soc. 114*, 338–343.
Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199–203.
Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., & Matthews, E. P. B. B. W. (1992) *Science 255*, 178–183.
Fersht, A. R. (1972) *J. Mol. Biol. 64*, 497.
Fersht, A. R. (1984) *Trends Biochem. Sci. 9*, 145.
Fersht, A. R. (1987) *Trends Biochem. Sci. 12*, 301.
Gill, S. J., & Noll, L. (1972) *J. Phys. Chem. 76*, 3065–3068.
Gregoret, L. M., & Cohen, F. E. (1991) *J. Mol. Biol. 219*, 109–122.
Gruenewald, B., Nicola, C. U., Lustig, A., Schwarz, G., & Klump, H. (1979) *Biophys. Chem. 9*, 137.
Hammes, G. G., & Roberts, P. B. (1969) *J. Am. Chem. Soc. 91*, 1812–1816.
Honig, B., & Hubbell, W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5412.
Honig, B., Hubbell, W., & Flewelling, R. F. (1986) *Annu. Rev. Biophys. Biophys. Chem. 15*, 163.
Janin, J. (1976) *J. Mol. Biol. 105*, 13–14.
Jenkins, G. M., & Watts, D. G. (1968) *Spectral Analysis and its Application*, Holden-Day, San Francisco.
Karplus, M., Ichiye, T., & Pettitt, B. M. (1987) *Biophys. J. 52*, 1083–1085.
Kauzmann, W. (1954) in *The Mechanism of Enzyme Action* (McEllroy, W. D., & Glass, B., Eds.) Johns Hopkins Press, Baltimore.
Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1–63.
Klotz, I. M., & Franzen, J. S. (1962) *J. Am. Chem. Soc. 84*, 3461–3466.
Kresheck, G. C., & Klotz, I. M. (1969) *Biochemistry 8*, 8–12.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.
Mandel, J. (1964) *The Statistical Analysis of Experimental Data*, Dover Publications, Inc., New York.
Miller, S., Janin, J., Lesk, A. M., & Chothia, C. (1987) *J. Mol. Biol. 196*, 641–656.
Murphy, K. P., & Gill, S. J. (1990) *Thermochim. Acta 172*, 11–20.
Murphy, K. P., & Gill, S. J. (1991) *J. Mol. Biol. 222*, 699–709.
Murphy, K. P., Privalov, P. L., & Gill, S. J. (1990) *Science 247*, 559–561.
Perutz, M. F., & Raidt, H. (1975) *Nature 255*, 256.
Prabhakaran, M., & Ponnuswamy, P. K. (1979) *J. Theor. Biol. 80*, 485–504.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167.
Privalov, P. L., & Khechinashvili, N. N. (1974) *J. Mol. Biol. 86*, 665–684.
Privalov, P. L., & Gill, S. J. (1988) *Adv. Protein Chem. 39*, 191–234.
Roder, H., Elove, G. A., & Englander, S. W. (1988) *Nature 335*, 700–704.
Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985) *Science 229*, 834–838.
Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg, Ser. Chim. 29*, 230–259.
Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M., & Bolen, D. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2854–2858.
Sharp, K. A., Nicholls, A., Friedman, R., & Honig, B. (1991) *Biochemistry 30*, 9686–9697.
Shirley, B. A., Stanssens, P., Hahn, U., & Pace, C. N. (1992) *Biochemistry 31*, 725–732.
Shrake, A., & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351–371.
Sneddon, S. F., Tobias, D. J., & Brooks, C. L., III (1989) *J. Mol. Biol. 209*, 817–820.
Susi, H. (1969) in *Structure and Stability of Biological Macromolecules* (Timasheff, S. N., & Fassman, G. D., Eds.) pp 575–663, Marcel Dekker, New York.
Teller, D. C. (1976) *Nature 260*, 729–731.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694–699.
Warshel, A., & Russell, S. T. (1984) *Q. Rev. Biophys. 17*, 283–422.
Warshel, A., Russell, S. T., & Churg, A. K. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4785–4789.
Wilson, E. B. (1952) *An Introduction to Scientific Research*, McGraw-Hill, New York.