Protein Engineering vol.8 no.11 pp.1081–1092, 1995

Extracting hydrophobicity parameters from solute partition and protein mutation/unfolding experiments

Sandor Vajda, Zhiping Weng and Charles DeLisi

Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215, USA

Hydrophobicity values for amino acids obtained from protein unfolding experiments are about twice as large as those obtained from data on the partitioning of amino acids between water and octanol. Quantitative analyses of several data sets, presented here, indicate that the difference is best explained by the most direct hypothesis, i.e. that the environment of hydrophobic groups in the interior of a protein is poorly modeled by octanol. Instead, we propose—and provide supporting evidence—that hydrocarbons are a more suitable model. First, we reanalyze data from both solute partitioning and protein unfolding experiments, taking account of the effects that were omitted previously, by introducing a volume dependence in the former and a full free energy analysis in the latter. Both changes in evaluation methodology decrease the discrepancy, but the differences remain substantial. The hydrophobicity parameter obtained from side-chain transfers between octanol and water increases from 16.7 to 22 cal/mol/Å^2, while that obtained from protein unfolding decreases from 34.9 to 31.2 cal/mol/Å^2. On the other hand, our analysis of the solubilities of pure hydrocarbons in water provides a hydrophobicity parameter of 30.8 cal/mol/Å^2. This apparent hydrocarbon-like environment of a protein's interior is also suggested more directly by an analysis of the contact environment of hydrophobic side chains in mutation/unfolding experiments, which have polar contact areas that are <2% of the total.

Keywords: hydrophobicity parameters/partition experiments/protein stability/residue mutation/transfer free energy

Introduction

The development of methods to evaluate quantitatively the solvation free energy of a protein and its contribution to protein stability is fundamental to the theory of molecular structure and of some practical importance in computer-assisted drug design. Although molecular dynamics and Monte Carlo simulations with explicit solvent molecules provide powerful methods for estimating solvent–solute free energies, computational costs preclude their use in complex docking and design calculations (Wilson *et al.*, 1991). Consequently, simpler alternatives have been developed, and although they are less rigorous, they have successfully reproduced the bulk properties of solvation (Harvey, 1989) and have been useful in the analysis of a number of experimental systems.

The free energies used to characterize amino acid hydrophobicities are usually obtained from the extent to which they partition between organic liquids and water at a constant temperature and pressure. This requires a theory relating free energy to measurable quantities such as concentration and physical properties of the solutions (see Equations 1–5 in Materials and methods). Under ideal conditions the relationship is especially simple, depending only on the ratio of concentrations, and that is what is typically used. Once the free energies are obtained, a model is required to relate them to properties of the side chains. The simplest model is a linear relationship between free energy and either solvent accessible surface areas (Chothia, 1974; Eisenberg and McLachlan, 1986; Ooi *et al.*, 1987; Cramer and Truhlar, 1992; Wesson and Eisenberg, 1992) or hydration shell volumes (Kang *et al.*, 1987; Colonna-Cesari and Sander, 1990; Stouten *et al.*, 1993). The proportionality constants in units of cal/mol/Å^2 and cal/mol/Å^3, respectively, are usually referred to as the hydrophobicity parameters, or more simply hydrophobicities. Thus the procedure for obtaining hydrophobicity parameters involves two steps, one requiring a theory for estimating the free energy and the other requiring a model relating free energy to side-chain properties.

Apart from questions related to the appropriate choice of theoretical and model-dependent expressions for free energy, the question of which liquid best characterizes a protein core has not been definitively answered (Dill, 1990). Good correlations of partition coefficients with observed amino acid distributions between a protein surface and its interior have been reported for partially polar, hydrogen bonding solvents such as octanol, ethanol and dioxan (Nozaki and Tanford, 1971; Kyte and Doolittle, 1982; Fauchere and Pliska, 1983). The most frequently used hydrophobicity scale (Eisenberg and McLachlan, 1986) is based on the octanol–water partition data of Fauchere and Pliska (1983). We showed previously that these and some 30 other scales all correlate with coefficients >0.95 (Cornette *et al.*, 1987).

Because of uncertainties in using partition experiments to estimate hydrophobicity parameters, several groups introduced an alternative method based on the quantifiable change in protein stability which occurs when a conservative (non-disruptive) point mutation is introduced. Typically, a large hydrophobic side chain is replaced by a smaller one (Matsumura *et al.*, 1988; Kellis *et al.*, 1989; Shortle *et al.*, 1990; Erickson *et al.*, 1992), and the change in stability is related to the difference in the free energy of unfolding. Traditional methods for analyzing these experiments yield hydrophobicity parameters that are about twice as large as those obtained from octanol–water partition experiments (Matsumura *et al.*, 1988; Shortle *et al.*, 1990), and several attempts have been made to rationalize the difference (Kellis *et al.*, 1989; Nicholls *et al.*, 1991; Sharp *et al.*, 1991; Erickson *et al.*, 1992; Lee, 1993; Honig and Yang, 1995) with no firm conclusion.

A number of factors not previously considered in detail can contribute to the free energy change in mutation/unfolding experiments. These include the loss of van der Waals interactions (which can be partially restored by the local readjustment of protein geometry), the change in conformational entropy increase upon unfolding, and the change in the electrostatic interaction energy. As a result, the free energy changes associated with a given mutation (e.g. Ile→Ala) are position dependent, and their evaluation is somewhat more

complicated than past analyses indicate. We show in Materials and methods that when these complications are taken into account, the hydrophobicity value does drop, but only from 34.9 to 31.2 cal/mol/Å^2.

The agreement is much better if the transfer free energy data are reevaluated by including a volume-dependent correction in the thermodynamic expression for the free energy. This takes into account non-ideality in the behavior of the system owing to the differences in size between the solute and solvent (Sharp *et al.*, 1991), which become important for concentrated systems (Flory, 1941, 1953; Huggins, 1941). The effect of this correction is to add a negative volume-dependent term to the free energy, and this forces a doubling of the area-dependent term for non-polar solutes, bringing the hydrophobicity parameter into close agreement with the value obtained from unfolding free energies (Sharp *et al.*, 1991; Pace, 1992). However, the theoretical validity of a Flory–Huggins (FH) 'correction' has been questioned (Holtzer, 1992; Ben-Naim and Mazo, 1993; Ben-Naim, 1994a; Chan and Dill, 1994; Holtzer, 1994; Jackson and Sternberg, 1994; Lee, 1994). Our own analysis presented below is in accord with this conclusion, and indicates a negligible deviation from ideality at the very low concentrations used in hydrocarbon–water and octanol–water partition experiments. If in fact the volume dependence is so small, it leaves us again with a large discrepancy between the hydrophobicities determined from solute transfer and protein unfolding data.

A resolution is suggested when we analyze the hydrocarbon solubilities using the standard transfer free energy expression and a volume-dependent term in the model relating free energy to molecular properties. In this case, we find an area-dependent hydrophobicity parameter of 30.8 cal/mol/Å^2 from transfer free energy data which is in excellent agreement with the 31.2 cal/mol/Å^2 obtained from the corrected analysis of mutation/unfolding experiments. Thus we conclude that the interior of a protein is better modeled by hydrocarbons then by octanol, and support this statement with further observations.

Materials and methods

Transfer free energy

For convenience in the data analysis we define the transfer free energy thermodynamically, and leave for Discussion the connection with its statistical mechanical interpretation (Ben-Naim, 1994a; Chan and Dill, 1994; Holtzer, 1994).

Let x_k and μ_k denote the mole fraction and the chemical potential, respectively, of a solute in solvent k, and let μ_k^0 be the partial molar Gibbs free energy of the solute at $x_k = 0$. Here, the polar solvent (water) will be denoted by $k = 1$ and the non-polar solvent (hydrocarbon or octanol) by $k = 2$. At constant temperature and pressure, the well-known expression (e.g. Hill, 1960) for the chemical potential of the solute is:

$$\mu_k = \mu_k^0 + (RT \ln x_k) + (RT \ln \gamma_k), \quad (1)$$

where γ_k, the activity coefficient, depends on the concentration of solute and approaches 1 as x_k becomes small:

$$\mu_k = \mu_k^0 + (RT \ln x_k). \quad (2)$$

At high concentrations, an equation of form 2 also holds (a form of Raoult's law):

$$\mu = \tilde{\mu}^0 + (RT \ln \bar{x}), \quad (3)$$

where $\tilde{\mu}^0$ is the free energy per mole of the pure solvent. Thus, in the standard state ($\bar{x} = 1$) the interpretation of Equation 3 is that each solvent molecule is surrounded by identical molecules. On the other hand, an evaluation of Equation 2 in the standard state indicates that μ_k^0 represents the free energy per mole of solute when each solute molecule is surrounded by solvent (i.e. the solute is infinitely dilute).

Generally, if two pure and largely immiscible liquids such as water and a hydrocarbon are in contact, after equilibrium is established the hydrocarbon will be in a very dilute solution with water as solvent, and the water will be in a very dilute solution with hydrocarbon as solvent. Thus Henry's law governs the chemical potential of the hydrocarbon as a solute, and Raoult's law governs its chemical potential as a solvent. If x_1 and $\bar{x}$ denote the concentrations of the hydrocarbon as solute and solvent, respectively, then at equilibrium:

$$\mu_1^0 + RT \ln x_1 = \tilde{\mu}^0 + (RT \ln \bar{x}). \quad (4)$$

The quantity of interest is the difference in free energy per mole of hydrocarbon as the result of being in two different environments: one in which it is surrounded entirely by water, and the other in which it is surrounded entirely by other hydrocarbons (Figure 1). Thus, the free energy ΔG^{tr} of transferring one mole of solute (hydrocarbon) from its pure phase into water is just the difference in the standard chemical potentials, i.e. $\Delta G^{tr} = \tilde{\mu}^0 - \mu_1^0$. By Equation 4:

$$\Delta G^{tr} = -RT \ln (\bar{x}/x_1). \quad (5)$$

An alternative situation of interest here is a homogeneous solute that has equilibrated between two (assumed) perfectly immiscible solvents. Then from Equation 1:

$$\Delta\mu^0 = (-RT \ln K) - [RT \ln (\gamma_2/\gamma_1)], \quad (6)$$

where $K = x_2/x_1$ is the partition coefficient and $\Delta\mu^0 = \mu_2^0 - \mu_1^0$. If the solute is sufficiently dilute in both phases, the second term on the right will be negligible.

The rate at which the activity coefficients approach 1 depends to a large extent on the ratio of the sizes of solute and solvent. If the solvent is a high molecular weight polymer, the rate at which its chemical potential approaches Raoult's law behavior (or, equivalently, by the Gibbs–Duhem relation, the rate at which the solute's chemical potential approaches

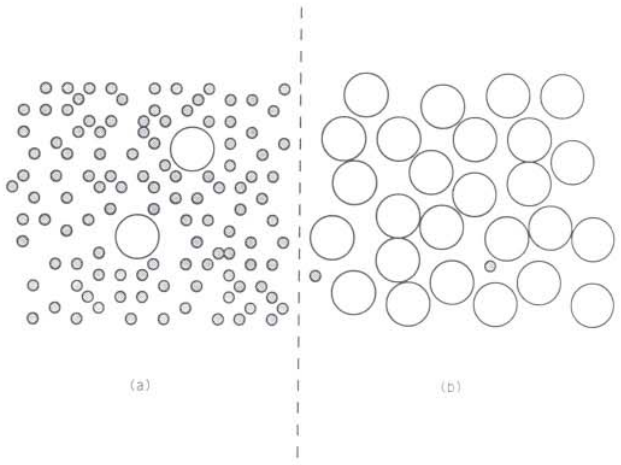

Fig. 1. Illustration of the transfer free energy concept. (**a**) Dilute hydrocarbon solution with water as the solvent. In the standard state, each hydrocarbon molecule is surrounded by water. (**b**) Dilute solution with hydrocarbon as the solvent. In the standard state, each hydrocarbon molecule is surrounded by identical molecules.

Henry's law behavior) as the solute becomes dilute can be surprisingly slow (Hill, 1960). When there is deviation, then a theory for the activity coefficient, such as that due to Flory (1941) and Huggins (1941), is required (see Chan and Dill, 1994). This will be discussed in detail below. Here we note that at sufficiently low concentrations the applicable equation becomes:

$$\Delta\mu^0 = -RT \ln K. \tag{7}$$

The quantity on the left is, by definition, the difference in free energy per mole of solute deriving from the differences in environment, i.e. from being completely surrounded by one or the other solvent. Thus, (Hermann, 1972; Chothia, 1974; Reynolds *et al.*, 1974; Fauchere and Pliska, 1983; Ben-Naim and Marcus, 1984; Ben-Naim and Mazo, 1993; Holtzer, 1994):

$$\Delta G^{tr} = \Delta\mu^0 = -RT \ln K. \tag{8}$$

Volume dependence of hydrophobicity

The transfer free energies of many non-polar solutes, obtained from partition coefficients, have been found empirically to be represented by:

$$\Delta G^{tr} = \sigma_a A + c, \tag{9}$$

where A is the solvent accessible surface area of the molecule, σ_a is a proportionality constant that is frequently used as a convenient measure of hydrophobicity (Hermann, 1972; Chothia, 1974) and c is a constant. The relationship has been extended to the transfer of amino acids by expressing ΔG as:

$$\Delta G^{tr} = \sum_{i=1}^{i=5} \sigma_i A_i, \tag{10}$$

where A_i denotes the solvent accessible surface areas of atom types C, N/O, S, O^- and N^+ (Eisenberg and McLachlan, 1986).

The parameters σ_C, σ_S, $\sigma_{N/O}$, σ_{O-} and σ_{N+} in Equation 10 are frequently called atomic solvation parameters (Smith and Honig, 1994). Because the non-polar area of a protein is essentially determined by the solvent accessible carbon atoms (extended with hydrogens), the parameter σ_C plays the same role as the parameter σ_a in Equation 9. Here we concentrate on those parameters that measure the hydrophobicity of non-polar solutes, and simply use established values for the solvation parameters of polar and charged atoms (Vajda *et al.*, 1994). This simplification is possible because non-polar atoms contribute >90% of the hydrophobic free energy change in protein folding and association.

Although the surface area terms are generally believed to make the largest contribution to the transfer free energy for amino acids under the usual conditions of interest, an additional volume-dependent term may sometimes be important. Thus, Ben-Naim (1994a) has shown, for two exactly solvable models, that the solvation free energy can be written as:

$$\Delta G^{tr} = \sigma_a A + \sigma_v V + c, \tag{11}$$

where V is the molar volume of the solute and c is a constant. For this case, the generalization of Equation 10 is:

$$\Delta G^{tr} = \sum_{i=1}^{i=5} \sigma_i A_i + \sigma_v V. \tag{12}$$

As indicated by Rossky *et al.* (1994), volume dependence appeared in a number of earlier contexts, e.g. in scaled particle theory (Reiss, 1965) with hard sphere pressures (Neff and McQuarrie, 1973), and in solvable approximations of hard sphere liquids (Lebowitz and Rowlinson, 1964). While these theories predict significant volume contributions at high pressures, using a molecular water solvent model and Lennard–Jones spheres with varying sizes as solute, Perkyns and Pettitt (1994) predicted that the volume-dependent terms can be substantial even at 1 atm pressure. However, the question of whether molar volume makes a non-negligible contribution to the transfer free energy under the conditions used in partition experiments has not been answered definitively (Sitkoff *et al.*, 1994).

Evaluation of hydrophobicity from protein unfolding

Here we introduce a detailed and relatively rigorous free energy evaluation model that involves the thermodynamic cycle shown in Figure 2. The major simplifying assumption in our analysis is that mutation affects only electrostatic and hydrophobic interactions, i.e. the side-chain size reduction is accommodated by relaxation of the protein which maintains close packing, and hence van der Waals constancy. This assumption will be examined further in the light of the results.

The quantity of interest is the difference in unfolding free energy, $\Delta\Delta G$, between the mutated and native proteins:

$$\Delta\Delta G = \Delta G^u_w - \Delta G^u_m \equiv (G^u_w - G^f_w) - (G^u_m - G^f_m) = G^f_m - G^f_w + \Delta G^r, \tag{13}$$

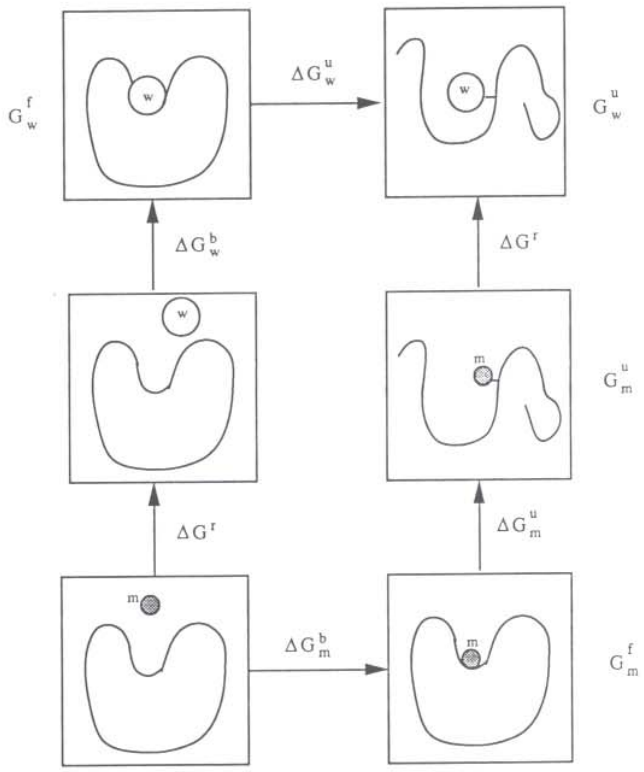

Fig. 2. Thermodynamic cycle used in the calculation of the free energy change $\Delta\Delta G$ due to the mutation of a hydrophobic side chain w to a hydrophobic side chain of smaller size m. The free energy changes are considered in the following processes: ΔG^u_w and ΔG^u_m, respectively, denote the free energy of unfolding of wild-type and mutated proteins. ΔG^b_w and ΔG^b_m are the free energies of binding side-chain analogs in the preformed cavity (see text). G^f_w and G^f_m are the free energies of the folded states, and G^u_w and G^u_m are those of the unfolded states.

where the subscripts w and m denote wild-type and mutant proteins respectively (or wild-type and mutant side chains respectively) and the superscripts f and u denote folded and unfolded states respectively. The 'replacement' free energy ΔG^r is defined as $\Delta G^r = G^u_w - G^u_m$, i.e. it is the free energy difference between the unfolded states. Based on the thermodynamic cycle shown in Figure 2, $G^f_m - G^f_w$ can be obtained as:

$$G^f_m - G^f_w = -\Delta G^b_w - \Delta G^r + \Delta G^b_m, \quad (14)$$

where ΔG^b_w and ΔG^b_m denote the binding free energies of the wild-type and mutant side chains, respectively, to the rest of the protein. Notice that the use of the same 'replacement' free energy ΔG^r in both Equations 13 and 14 assumes that in the unfolded state the only interaction between the side chain and the rest of the protein is due to the covalent bond, and this energy term is the same for wild-type and mutant residues (Figure 1).

The two equations thus yield:

$$\Delta\Delta G = \Delta G^b_m - \Delta G^b_w. \quad (15)$$

The second term on the right can be written as (Vajda *et al.*, 1994):

$$\Delta G^b_w = \Delta E^{el} + \Delta G^h_w - T\Delta S_w, \quad (16)$$

where ΔE^{el} and ΔG^h_w denote the electrostatic and hydrophobic contributions to the binding free energy, and $T\Delta S_w$ represents the free energy change due to the loss of side-chain conformational entropy upon binding. The hydrophobic term ΔG^h_w is obtained by difference:

$$\Delta G^h_w = \Delta G^{tr}_{pw} - \Delta G^{tr}_p - \Delta G^{tr}_w, \quad (17)$$

where ΔG^{tr}_{pw}, ΔG^{tr}_p and ΔG^{tr}_w denote the transfer free energies of the intact protein, the protein with the side chain removed and the separate side chain, respectively. If the protein is rigid (i.e. the mutation does not induce a conformational change), the electrostatic contribution to the binding free energy reduces to the electrostatic interaction energy between the side chain and the rest of the protein. Because the side chain in question is hydrophobic, the latter is close to zero. In the more general flexible case, ΔE^{el} is calculated from the electrostatic energies of the participating molecules as:

$$\Delta E^{el} = E^{el}_{pw} - E^{el}_p - E^{el}_w, \quad (18)$$

where E^{el}_{pw}, E^{el}_p and E^{el}_w denote the electrostatic energies of the intact protein, the protein with the side chain removed and the separate side chain, respectively. Because the side chain is non-polar, $E^{el}_w = 0$. The term E^{el}_p is independent of the mutation and will cancel out when the differential free energy (Equation 15) is computed.

Because equations of the form 16–18 also apply to the mutant protein, Equation 15 becomes:

$$\Delta\Delta G = -(\Delta G^{tr}_m - \Delta G^{tr}_w) + (\Delta G^{tr}_{pm} - \Delta G^{tr}_{pw}) - (T\Delta S_m - T\Delta S_w) + (E^{el}_{pm} - E^{el}_{pw}). \quad (19)$$

By Equation 10, the first and second terms on the right are proportional to the change in surface area of the protein, and to the difference in surface area between the original and substituted side chains, respectively. The third and fourth terms are the change in conformational entropy loss and the change in the electrostatic energy. Although the substitution involves non-polar side chains, the last term may nonetheless be important because the mutation generally affects the structure of the entire protein and thereby its electrostatic energy.

The terms in Equation 19 were evaluated for 55 mutations of aliphatic side chains in staphylococcal nuclease, including the alanine and glycine substitutions of 11 leucine, nine valine and five isoleucine residues, as well as mutations to valine of each isoleucine (Shortle *et al.*, 1990). The structures of mutant proteins were obtained by homologous extension from the X-ray structure of the staphylococcal nuclease (entry 5cyt of the Brookhaven Protein Data Bank) using the side-chain replacement option of the QUANTA molecular modeling program (Molecular Simulations, Inc.). The CHARMm (Brooks *et al.*, 1983) potentials (version 21) of all structures were locally minimized for 40 steps using an adapted basis Newton–Raphson procedure. The solvent accessible surface areas were evaluated by the method of Lee and Richards (1971).

Results

Atomic solvation parameters

Although we have concentrated on the hydrophobicity parameter σ_C, in Table I we list all the atomic solvation parameters obtained from octanol–water data by various methods. The procedure used by Eisenberg and McLachlan (1986) is to fit Equation 10 (transfer free energies assumed to be proportional to the atomic surface areas) to the transfer free energies evaluated from the partition experiments using Equation 7. The resulting parameters, shown in column 1 of Table I, differ slightly from those obtained by Eisenberg and McLachlan (1986) because we: (i) calculated surface areas for the *N*-acetyl amino acid amides that were actually used in the experiment rather than adopting the surface areas calculated

Table I. Estimates of atomic solvation parameters from the octanol–water partition of *N*-acetyl amino acid amides (Fauchere and Pliska, 1983)[a]

Parameter	Units	Surface only ($\sigma_v = 0$)	σ_v as free parameter	$\sigma_v = 29.5$, fixed	FH adjustment (Sharp *et al.*, 1991)
σ_C	cal/mol/Å^2	16.7 ± 1.0	49.2 ± 17.5	31.0 ± 1.1	34.5 ± 1.2
$\sigma_{N/O}$	cal/mol/Å^2	–9.1 ± 2.5	23.1 ± 17.4	4.9 ± 2.2	2.0 ± 3.1
σ_S	cal/mol/Å^2	4.2 ± 9.4	32.0 ± 17.3	16.3 ± 7.4	17.3 ± 11.6
σ_{O^-}	cal/mol/Å^2	–34.5 ± 7.3	–4.1 ± 17.7	–21.2 ± 7.4	–21.4 ± 8.9
σ_{N^+}	cal/mol/Å^2	–38.8 ± 4.5	–14.9 ± 13.5	–28.5 ± 5.8	–27.9 ± 5.6
σ_v	cal/cm^3	–	–69.3 ± 37.1	–29.5[b]	–
s_r	kcal/mol	0.35	0.31	0.32	0.43

[a]All fits are to data of the form $\Delta G^{tr}(X) - \Delta G^{tr}(Gly)$, where $\Delta G^{tr}(X)$ denotes the transfer free energy of the *N*-acetyl amide of amino acid X. The independent variables are differences of the form $A_i(X) - A_i(Gly)$, where $A_i(X)$ denotes the solvent accessible surface area of atom type i (i = C, N/O, S, O^-, N^+) in amino acid X. In the last row, s_r denotes the residual variance of the fit.
[b]Fixed value.

by Shrake and Rupley (1973) for Gly–X–Gly tripeptides; and (ii) disregarded the data for cysteine and histidine. These two residues yield large residual errors probably because no reliable data are available for the –SH group, whereas the ionization state of histidine is not well defined at pH 7.1 used in the experiment. Table I also shows the atomic solvation parameters calculated from the octanol–water partition data adjusted by Sharp *et al.* (1991) and other parameter sets that will be discussed further in the paper.

Volume dependence in octanol–water partition of amino acids

In this section we assume that the transfer free energy depends on both area and volume. A simple physical basis for the effect is that the transfer of a solute from octanol to water requires the creation of a solute-sized cavity. Although a cavity of the same size is lost on the water side, the work involved in the two different solvents will generally differ, resulting in a volume-dependent contribution and a reduction in σ_a. As we show in Discussion, a variety of other phenomena can yield volume-dependent contributions to the transfer free energy. In octanol–water and hydrocarbon–water transfer, this volume-dependent term always turns out to be negative, and hence reduces the positive surface-dependent contribution. Because there is no differential volumetric change in mutation/unfolding experiments, a volume-dependent term in the partition experiments will bring its value closer to that obtained from protein unfolding experiments. As already mentioned, this explains why the FH correction, suggested by Sharp *et al.* (1991), improved the agreement with the latter data.

The volume dependence of hydrophobicity is tested directly by fitting Equation 12 rather than Equation 10 to the transfer free energies. Figure 3 shows the residual variance s_r and the atomic solvation parameters as functions of the volumetric parameter σ_v. At $\sigma_v = 0$, the model reduces to Equation 10 and we obtain the parameter values shown in column 1 of Table I. The residual variance is $s_r = 0.35$. Decreasing σ_v slightly decreases the residual variance s_r. The minimum (s_r = 0.31 kcal/mol) is attained at $\sigma_v = -69.3$ cal/cm^3, resulting in the parameters shown in column 3 of Table I. Thus, assuming that the transfer free energy depends on both surface and volume gives a slightly better fit to the data, but the improvement is not significant. Therefore the data of Fauchere and Pliska (1983) do not confirm the existence of a significant volume dependence, and provide almost no information on the value of σ_v. In fact, as shown in Figure 3, large changes in the value of σ_v leave the residual variance s_r almost invariant, while substantially affecting the atomic solvation parameters.

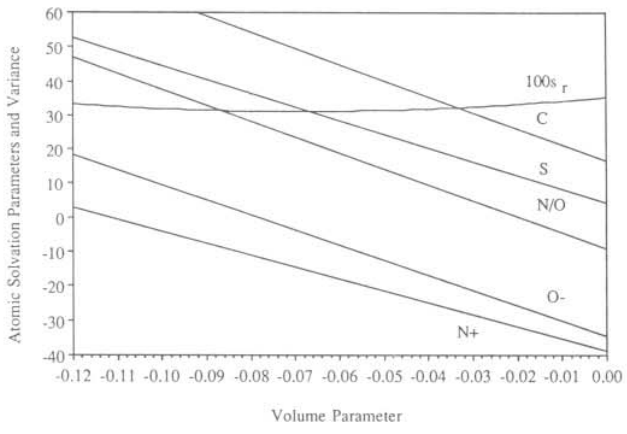

Fig. 3. Residual variance s_r and the atomic solvation parameters σ_C, $\sigma_{N/O}$, σ_S, $\sigma_{\bar{O}}$ and σ_N^+ as functions of the volume dependence parameter σ_v. The variance s_r is multiplied by 100 to show that it slightly varies with σ_v, and attains its minimum at $\sigma_v = -0.069$ kcal/cm^3.

Volume dependence of hydrocarbon transfer to water

For the analysis of volume contributions to the transfer free energies, the hydrocarbon solubilities in water offer more informative data than the Fauchere–Pliska experiments. All atoms are non-polar, and we can restrict consideration to the simple model given by Equation 11. Because the solvent accessible surface area and the volume of small *n*-alkanes are strongly correlated, we include a number of branched, cyclic and aromatic hydrocarbons (Table II) in addition to the 10 normal alkanes used by Sharp *et al.* (1991). All transfer free energies were calculated from the solubility data of McAuliffe (1966) and Ben-Naim and Marcus (1984).

As is well known, for each homologous sequence of hydrocarbons the transfer free energy can be described as a linear function of either molar volume (McAuliffe, 1966) or solvent accessible surface area (Hermann, 1972). The intercept parameters depend on the hydrocarbon type, and the data are on parallel lines. For a given carbon number, branching and ring formation increase the solubility of the hydrocarbon, i.e. decrease its free energy of transfer to water (McAuliffe, 1966).

We assume that the transfer free energy depends on both surface area and volume. Fitting Equation 11 to all the data in Table II simultaneously yields a large residual variance ($s_r = 0.43$) and systematic residual errors. In particular, the free energies are overestimated for all aromatic compounds. Thus we allow the parameters to depend on hydrocarbon type by fitting the following alternative models: (i) a common volume term with coefficient σ_v, and separate surface terms

Table II. Transfer free energies of hydrocarbons to water[a]

Compound	Area (Å^2)[b]	Volume (cm^3/mol)	ΔG^{tr} (kcal/mol)
Methane	130	39	2.59
Ethane	163	65	3.14
Propane	194	81	3.92
n-Butane	224	100	4.79
n-Pentane	252	115	5.74
n-Hexane	282	130	6.60
n-Heptane	311	146	7.32
n-Octane	340	162	8.21
n-Nonane	369	178	9.02
n-Decane	399	195	9.85
2-Methylbutane	240	116	5.61
2,2-Dimethylpropane	237	122	5.51
2-Methylpentane	269	132	6.38
3-Methylpentane	263	130	6.43
2,2-Dimethylbutane	255	133	6.20
2,4-Dimethylpentane	285	149	7.12
2,2,4-Trimethylpentane	297	165	7.43
2,2,5-Trymethylhexane	327	181	7.89
Cyclopentane	224	94	5.02
Cyclohexane	245	108	5.66
Methylcyclopentane	248	112	5.80
Benzene	211	89	3.67
Toluene	240	106	4.40
Ethylbenzene	265	122	5.13
2-Propylbenzene	293	140	5.78

[a]Data are taken from McAuliffe (1966) and Ben-Naim and Marcus (1984).
[b]Surface areas calculated using a carbon radius of 1.9 Å and a hydrogen radius of 0.0 Å. Molar volumes are taken from McAuliffe (1966).

Table III. Estimates of hydrophobicity parameters from solubilities of hydrocarbons[a,b]

Model (i)		Model (ii)			Model (iii)	
σ_v free parameter	$\sigma_v = 0$, fixed	σ_v free parameter	$\sigma_v = 0$, fixed	$\sigma_v = -33.3$, fixed	σ_a free parameter	$\sigma_a = 0$, fixed
$\sigma_{a,n} = 31.2 \pm 4.5$	$\sigma_{a,n} = 27.9 \pm 0.4$	$\sigma_a = 30.6 \pm 3.9$	$\sigma_a = 27.8 \pm 0.4$	$\sigma_a = 47.1 \pm 0.6$	$\sigma_a = 32.5 \pm 3.9$	$\sigma_a = 0$
$\sigma_{a,b} = 32.7 \pm 5.0$	$\sigma_{a,b} = 29.0 \pm 0.5$	$c_n = -1.4 \pm 0.2$	$c_n = -1.3 \pm 0.1$	$c_n = -2.3 \pm 0.2$	$\sigma_{v,n} = -8.1 \pm 9.3$	$\sigma_{v,n} = 48.1 \pm 1.4$
$\sigma_{a,c} = 31.8 \pm 4.5$	$\sigma_{a,c} = 28.5 \pm 0.6$	$c_b = -1.1 \pm 0.2$	$c_b = -1.0 \pm 0.1$	$c_b = -1.5 \pm 0.2$	$\sigma_{v,b} = -4.9 \pm 8.1$	$\sigma_{v,n} = 44.5 \pm 1.4$
$\sigma_{a,a} = 27.4 \pm 7.8$	$\sigma_{a,a} = 24.0 \pm 0.5$	$c_c = -1.3 \pm 0.3$	$c_c = -1.1 \pm 0.1$	$c_c = -2.3 \pm 0.2$	$\sigma_{v,c} = -6.7 \pm 9.3$	$\sigma_{v,n} = 49.8 \pm 2.1$
$c = -1.5 \pm 0.3$	$c = -1.3 \pm 0.2$	$c_a = -2.4 \pm 0.2$	$c_a = -2.2 \pm 0.1$	$c_a = -3.3 \pm 0.2$	$\sigma_{v,a} = -16.3 \pm 9.1$	$\sigma_{v,a} = 39.2 \pm 1.8$
$\sigma_v = -5.7 \pm 7.8$	$\sigma_v = 0$	$\sigma_v = -4.8 \pm 6.7$	$\sigma_v = 0$	$\sigma_v = -33.3$	$c = -1.6 \pm 0.3$	$c = 0.27 \pm 0.1$
$s_r = 0.128$	$s_r = 0.130$	$s_r = 0.125$	$s_r = 0.127$	$s_r = 0.175$	$s_r = 0.137$	$s_r = 0.230$

[a]Data are taken from Table II.
[b]Surface parameters in cal/mol/Å^2, volume parameters in cal/cm^3, additive constants in kcal/mol.

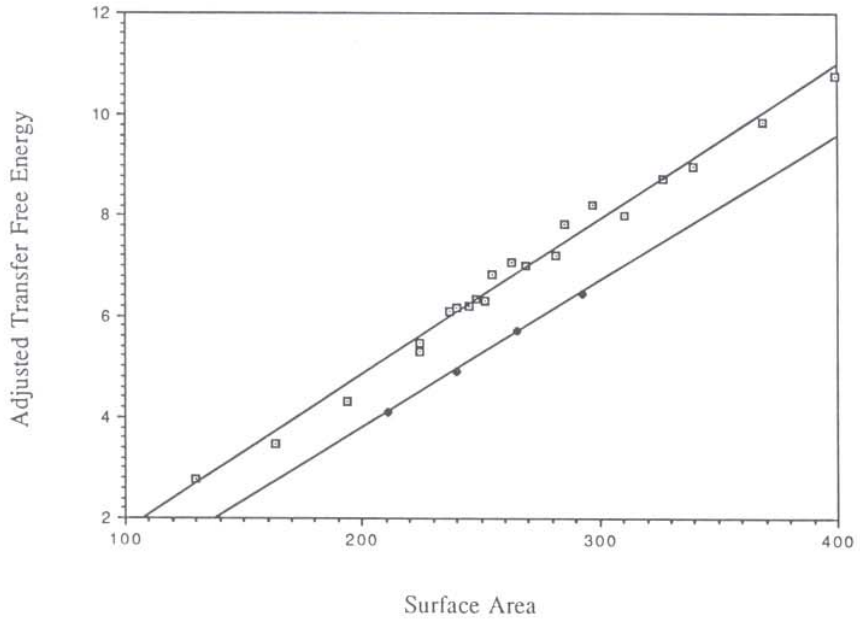

Fig. 4. Surface area dependence of the transfer free energies, calculated from solubility data and adjusted by the small volume-dependent term $-4.8\sigma_v$. Data are pooled for normal, branched and cyclic hydrocarbons. The slope of the line is $\sigma_a = 30.8$ cal/mol/Å^2. The four data points for aromatic compounds are on a separate line.

$\sigma_{a,n}$, $\sigma_{a,b}$, $\sigma_{a,c}$ and $\sigma_{a,a}$ for the normal, branched, cyclic and aromatic hydrocarbons, respectively; (ii) both the surface parameter σ_a and the volume parameter σ_v are common, but Equation 11 is fitted with separate constant terms c_n, c_b, c_c and c_a; and (iii) a common surface term σ_a and separate volume terms $\sigma_{v,n}$, $\sigma_{v,b}$, $\sigma_{v,c}$ and $\sigma_{v,a}$ for the normal, branched, cyclic and aromatic compounds, respectively.

Each of these assumptions can result in a good fit to the data in Table II, with residual variances <0.14 kcal/mol. The residual errors for the three models (i–iii) do not differ significantly from each other. However, the model uncertainty does not lead to any problem because all three models yield very similar values for both volume and surface area parameters (Table III). Apart from the aromatic compounds in model (iii), the σ_v values are between –8.1 and –4.8 cal/cm^3. Because the additional free parameter, σ_v, yields only a small decrease in the residual variance s_r, we are unable to establish the existence of a significant volume dependence. However, the estimates are good enough to show that the volume dependence is substantially smaller than expected on the basis of the FH theory. As will be discussed further, the FH theory yields $\sigma_v = -33.3$ cal/cm^3. Fitting model (ii) with this fixed value of σ_v yields $\sigma_a = 47.1$ cal/mol/Å^2 for the hydrophobicity parameter, exactly the value obtained by Sharp *et al.* (1991) from the analysis of *n*-alkanes, and increases the residual variance to $s_r = 0.175$ from $s_r = 0.125$ (see column 3 and 5 in Table III). By the *F*-test, this increase in the residual variance is significant at the probability level $p = 0.06$ (Rohlf and Sokal, 1969), which implies that the FH model overestimates the volume dependence. This will also affect the hydrophobicity parameter.

Due to the small differences among the constant terms, the data can be pooled for normal, branched and cyclic hydrocarbons. Figure 4 shows the surface area dependence of the free energies already adjusted by the small volume-dependent term $-4.8\sigma_v$. The slope of the line is $\sigma_a = 30.8$ cal/mol/Å^2. The data points for the four aromatic compounds in Table II are on separate lines (Figure 4) and give a slightly different slope of $\sigma_a = 29.0$ cal/mol/Å^2. These values are substantially smaller than the 47.1 cal/mol/Å^2 given by the FH correction (Sharp *et al.*, 1991).

For completeness, we have also estimated the value of σ_a from the cyclohexane–water partition of the analogs of the Ala, Ile, Leu and Val side chains (Figure 5). There are too few points to estimate a volume dependence, and hence we

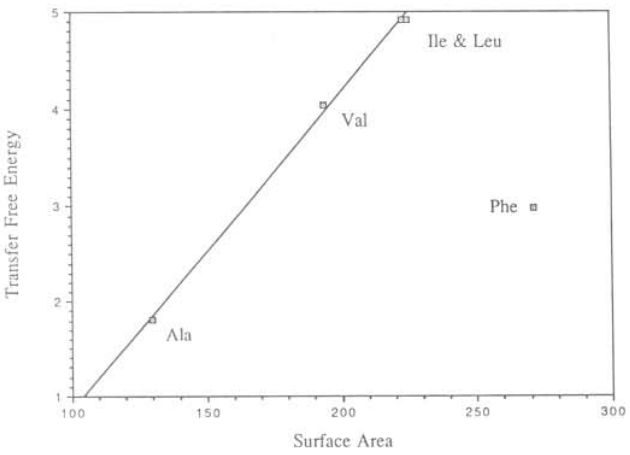

Fig. 5. Surface area dependence of the transfer free energies calculated from the partitioning of the analogs of the Ala, Ile, Leu and Val side chains. The slope of the line is σ_a = 33.1 cal/mol/Å^2. The transfer free energy for phenylalanine differs substantially from the values observed for the other hydrophobic side chains, and has not been taken into account when fitting the line to the data.

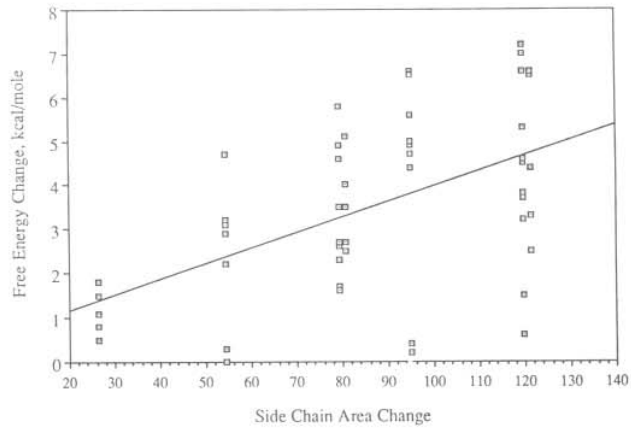

Fig. 6. Observed change in the free energy of unfolding (Shortle *et al.*, 1990) as a function of the change in the side-chain area upon the mutation of the side chain to another of smaller size. The slope of the line is σ_a = 34.9 ± 7.8 cal/mol/Å^2.

set σ_v = 0. The slope of the line is σ_a = 33.1 cal/mol/Å^2. Figure 5 also shows that the transfer free energy for phenylalanine differs substantially from the values observed for the other hydrophobic side chains, and hence has not been taken into account when fitting the line to the data.

Hydrophobicity from mutation data

Figure 6 shows the results of correlating the observed changes in free energy (Shortle *et al.*, 1990) with the differences between the solvent accessible surface areas of mutant and wild-type side chains. The slope of the line is σ_a = 34.9 ± 7.8 cal/mol/Å^2 (Table IV). The large variability of the observations results in a large variance of the hydrophobicity parameter as well. Because the independent variable is the side-chain area change, this traditional method of evaluating the mutation data is equivalent to retaining only the term $-(\Delta G_m^{tr} - \Delta G_w^{tr})$ on the right-hand side of Equation 19 and neglecting all the others. Indeed, the retained term is proportional to the side-chain area change due to the mutation and is always positive because the wild-type side chain is substituted by a smaller one.

Table IV. Hydrophobicity parameters from mutation data[a]

Parameter	Method			
	1[b]	2[c]	3[d]	4[e]
σ_a	34.9 ± 7.8	25.4 ± 5.1	27.7 ± 4.9	31.2 ± 4.7
c	0.45 ± 0.70	0.88 ± 0.55	0.71 ± 0.53	0.45 ± 0.51
s_r	1.69	1.64	1.61	1.49

[a]Data from Shortle *et al.* (1990).
[b]Using only side-chain area.
[c]Using total non-polar area.
[d]Adjustment for entropy change and solvation contribution of polar groups.
[e]Additional correction for electrostatic energy.

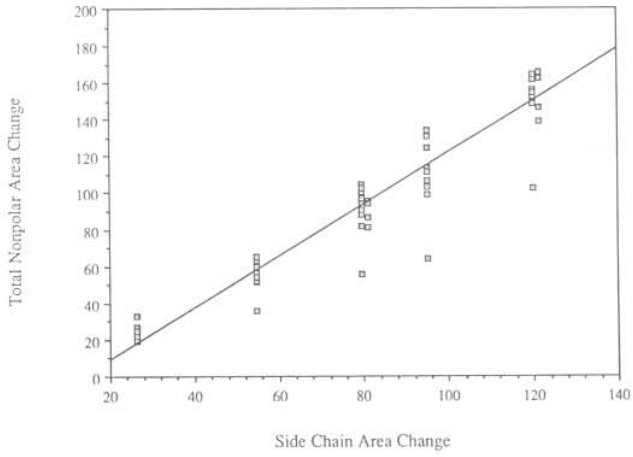

Fig. 7. Change in the total (i.e. side chain plus protein) non-polar surface area as a function of the change in the side-chain area alone. The slope of the line is 1.4.

The change in solvent accessible surface area is not confined only to the side chain, because subsequent relaxation of the molecule can expose previously buried atomic groups and slightly alter the shape of the protein. Indeed, in addition to $-(\Delta G_m^{tr} - \Delta G_w^{tr})$, Equation 19 includes another hydrophobic term $(\Delta G_{pm}^{tr} - \Delta G_{pw}^{tr})$ which is proportional to the change in the accessible area of the protein. Figure 7 shows the change in total (i.e. side chain plus protein) non-polar area upon mutation as a function of the change in the side-chain area alone. The two surface areas are correlated, with a slope of 1.4. Retaining both hydrophobic terms on the right-hand side of Equation 19 reduces σ_a from 34.9 to 25.4 cal/mol/Å^2 because the surface area is increased (Figure 8 and Table IV, column 2). However, this result will change as the further terms of Equation 19 are taken into account. The solvation contribution of the polar groups and the entropic term $(T\Delta S_m - T\Delta S_w)$ have only a small effect (Table IV, column 3), but the inclusion of the electrostatic contributions increases σ_a to 31.2 cal/mol/Å^2 while further improving the fit (Figure 9 and Table IV, column 4).

Discussion

Hydrophobicity from solute partition

We have shown that the effect of including a volume term in the expression for octanol–water transfer free energies can change the surface area hydrophobicity coefficient in a way that reduces its deviation from values obtained from protein unfolding experiments. Similarly, a more complete analysis of

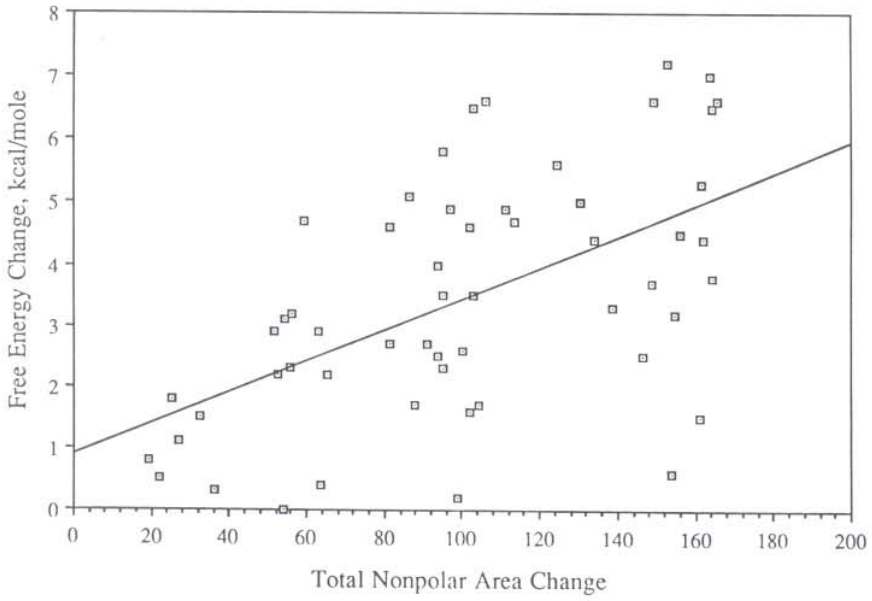

Fig. 8. Free energy change due to mutations as a function of the change in the total non-polar surface area. The slope of the line is 25.4 ± 5.1 cal/mol/Å^2.

unfolding experiments was shown to decrease the discrepancy. However, the deviation remains severe. A direct explanation for the deviation is the possibility that one or both activity coefficients differ from unity and hence the transfer free energy should be determined from Equation 6 rather than Equation 8. This possibility has been suggested by Sharp *et al.* (1991) using the FH theory.

Flory (1941) and Huggins (1941) introduced a widely used theory for taking account of non-ideality in dense polymer solutions. In particular, the generalization of Equation 3 is (Hill, 1960):

$$\mu = \tilde{\mu}^0 + RT\{\ln(1-\phi) + [1-(1/r)]\phi + \chi\phi^2\}. \quad (20)$$

In Equation 20, r is the ratio of solvent to solute molecular volumes, χ is related to the energy of mixing and ϕ is the volume fraction of the solute:

$$\phi = rx/[1 + (r-1)x]. \quad (21)$$

If $\phi << 1$ (or equivalently if $x << 1$), Equation 20 becomes:

$$\mu \approx \tilde{\mu}^0 - RT[(\phi/r) + \phi^2(0.5-\chi) + O(\phi^3)]. \quad (22)$$

For typical solvents $|\chi| \leq 2$. If solute and solvents are approximately the same size, $\phi \approx rx$ and the change in solvent chemical potential is approximately linear in solute concentration even if the solute mole fraction is as high as 0.1. For polymeric solvents, comparable linearity in the chemical potential requires $(r-1)x << 1$, and for high polymers ($r \geq 1000$) this means that attaining the Raoult's law limit requires mole fractions of the solute three orders of magnitude smaller than for monomeric solvents.

In the Fauchere and Pliska (1983) experiments, octanol as the solvent will undoubtedly obey Raoult's law, first because octanol and hydrophobic side-chain volumes are comparable, and more importantly because side-chain concentrations are at 10^{-4} M. Satisfaction of the limiting conditions can be seen by relating the molar concentration, C, to the mole fraction. Because $C = 1000x/[1 + (r-1)x]/V$, where V is the molar volume of the solvent (18 cm^3 for water and 157 cm^3 for octanol), in dilute solutions $x \approx VC/1000$. Evidently, for either octanol or amino acids ($r \leq 10$), the conditions for retaining only the first term in the expansion of Equation 20 are satisfied without an error of any consequence. Thus we expect that on both the water and octanol sides of the partition, solvent will obey Raoult's law and solute will obey Henry's law. These conclusions should hold even if the FH theory is not quantitative because as long as conditions are not close to phase separation, the chemical potential can be written as a well behaved (and therefore expandable) function of concentration.

Transfer free energy and contact energy

Ben-Naim (1980) decomposed the chemical potential of the solute as:

$$\mu_k = -RT \ln q + \mu_k^* + RT \ln(\rho_k \Lambda^3), \quad (23)$$

where q is the internal partition function (electronic, vibrational, rotational, conformational, etc.) of the isolated (gas phase) solute molecule, $\rho_k = N_k/V_k$ is the number density (i.e. N_k solute molecules in volume V_k) and Λ is the solute's DeBroglie wavelength, a constant (Hill, 1960). In Equation 23, μ_k^* represents the contact free energy, i.e. the interaction free energy change when 1 mol of the solute is dispersed in solution k at translationally fixed positions. Since μ_k^* is the binary parameter that describes the solute–solvent interaction, Ben-Naim (1980) defined the solvation free energy as $\Delta\mu^* = \mu_2^* - \mu_1^*$, i.e. the difference in the contact free energies.

Assume first that neither the DeBroglie wavelength Λ nor the internal degrees of freedom of the solute contained in $-RT \ln q$ change upon transfer. The equilibrium condition $\Delta\mu = 0$ then implies $\Delta\mu^* = -kT \ln K_\rho$, where $K_\rho = \rho_2/\rho_1$ is the partition coefficient expressed in number densities. Thus, apart from a constant, this result agrees with the classical expression for the transfer free energy given in Equation 8. To the extent that we are interested in the transfer free energy difference between different amino acids, the constant is unimportant.

However, the change $\Delta\mu^* = \mu_2^* - \mu_1^*$ in the contact energy differs from the transfer free energy if the center of mass degrees of freedom couple with internal degrees of freedom,

rotations or excluded volume, because in this case the term $-RT \ln q$ does not cancel out. While there is no coupling at infinite dilutions in small compact solvents, even where solutes are polymers, it is possible if the solvent consists of chain-like molecules (Chan and Dill, 1994). Such coupling may exist even when the limiting concentrations are achieved (i.e. Raoult's and Henry's law regimes are applicable). It is the presence of the coupling term that may lead to the volume dependence of the contact free energy. While the FH theory is useful to describe the strong coupling of conformational and translational degrees of freedom in concentrated polymer solutions, it is not necessarily the best model for the dilute solutions involved in the partition experiments. As we show here, the volume dependence found is substantially weaker than expected from the FH theory.

Volume dependence of hydrophobicity

As discussed previously, the FH adjustment of transfer free energies would have the same effect on the hydrophobicity parameter as adding (with the opposite sign) a volume-dependent term to Equation 9. In particular, adjusting the free energy of transferring a hydrocarbon from its pure liquid phase into water yields (Sharp *et al.*, 1991):

$$\Delta G^{tr} = -RT \ln K - RT[1 - (V/V_w)], \qquad (24)$$

where V and V_w denote the molar volumes of hydrocarbon and water, respectively. Thus, apart from the constant $-RT$, the adjustment is proportional to the volume of the solute. Because $V_w = 18$ cm^3/mol, the coefficient of proportionality is given by $RT/V_w = 33.3$ cal/cm^3. As we have shown, fitting the transfer free energy obtained from Equation 8 to Equation 11 with fixed $\sigma_v = -33.3$ cal/cm^3 yields the same σ_a (47.1 cal/mol/Å^2; see Sharp *et al.*, 1991) as fitting the transfer free energy obtained from Equation 24 to Equation 9 (which has surface area only). This is an example of the potential complexity that arises in the interpretation of a hydrophobicity parameter that depends on the assumptions of two independent models: one for obtaining free energy and one for fitting the free energy.

However, the direct fit of Equation 11, including a volume-dependent term as a free parameter, to the unadjusted transfer free energies of hydrocarbons revealed that this volume dependence is significantly smaller than the –33.3 cal/cm^3 assumed by Sharp *et al.* (1991). Three slightly different fits to the data shown in Table II yield volume-dependent terms in the range $-8.1V$ to $-4.8V$ (Table III). Even the smallest of these values gives only $\sigma_a = 32.5$ rather than 47.1 cal/mol/Å^2, corresponding to the use of the FH theory. The best agreement with data pooled for normal, branched and cyclic hydrocarbons has been attained at $\sigma_a = 30.8$ cal/mol/Å^2.

Sharp *et al.* (1991) calculated the octanol–water transfer free energies according to the equation:

$$\Delta G^{tr} = -RT \ln K - RTV[(1/V_o) - (1/V_w)], \qquad (25)$$

where V_w and V_o denote the molar volumes for water and octanol, respectively. Because $V_o = 157$ mol cm^{-3} and $V_w = 18$ mol cm^{-3}, the volume-dependent term in Equation 25 is $-29.5V$ cal/mol. As expected, fitting the extended model (Equation 12) with fixed $\sigma_v = -29.5$ cal/cm^3 to unadjusted transfer free energies yields results (Table I, column 3) that are very close to those values obtained by fitting Equation 10 to the free energies adjusted by Sharp *et al.* (1991) (Table I, column 4). The small differences are caused by the use of slightly different volumes for *N*-acetyl amino acid amides.

Table V. Summary of the hydrophobicity values obtained by different methods

System	Method		
	Classical value	FH adjustment	This paper
Octanol–water[a]	16.7 ± 1.0	34.5 ± 1.2	16.7–22.1
Hydrocarbon solubilities[b]	27.9 ± 0.4	47.0 ± 0.4	30.8 ± 1.2
Cyclohexane–water[c]	33.1 ± 1.0	52.0 ± 0.6	–[d]
Mutation and unfolding[e]	34.9 ± 7.8	N/A	31.2 ± 4.6

[a]Unadjusted data are taken from Fauchere and Pliska (1983). Data with FH adjustments are taken from Sharp *et al.* (1991).
[b]Data are taken from Table III.
[c]Unadjusted data for non-polar side-chain analogs are taken from Radzicka and Wolfenden (1988). Data with FH adjustments are taken from Sharp *et al.* (1991).
[d]Too few points, no further analysis.
[e]Data are taken from Shortle *et al.* (1990).

Assuming volume dependence as a free parameter slightly improves the fit to the octanol–water partition data (Table I and Figure 3), but improvement is far from significant. Because the residual variance s_r is nearly invariant to large variations in the value of σ_v (Figure 3), the data provide almost no information about the magnitude of volume dependence.

There are four solutions considered here. Measuring the solubility of a hydrocarbon involves a dilute solution of the hydrocarbon in water and a dilute solution of water in the hydrocarbon. In evaluating the transfer free energies of *N*-acetyl amino acid amides from the experiments by Fauchere and Pliska (1983) one traditionally assumes that water and octanol do not mix, thereby restricting the consideration to solutions of *N*-acetyl amino acid amides in water and octanol. As we have already discussed, the chemical potential is always independent of the solute size in dilute aqueous solutions, but may contain volume dependence if the solvent consists of chain-like molecules (Chan and Dill, 1994; Kumar *et al.*, 1995). Thus, the most likely candidate for a significant volume dependence is the dilute solution of water in hydrocarbons because the ratio of solute:solvent molar volumes is close to unity in solutions of *N*-acetyl amino acid amides in octanol. However, fitting the volume-dependent model to the hydrocarbon–water partition data shows that the dependence on the solute volume is significantly smaller than predicted by the FH theory.

There is no physical reason to assume that the volumetric effects are much larger when transferring amino acids between octanol and water. Assuming a volume dependence parameter σ_v in the same range as for hydrocarbons, i.e. between –10 and 0 cal/cm^3 (see Table III), yields hydrophobicity parameter values between $\sigma_a = 22.1$ and 16.7 cal/mol/Å^2. Even 22.1 cal/mol/Å^2 is much smaller than the value of 34.5 cal/mol/Å^2 calculated from the octanol–water partition data adjusted by Sharp *et al.* (1991). The hydrophobicity parameters derived by the various methods are summarized in Table V.

Hydrophobicity from mutation data

We have reevaluated the mutation data of Shortle *et al.* (1990) using a detailed model that accounts for a variety of contributions to the change in stability, including the change in the surface area of the protein, electrostatic and entropic effects, as well as the small contribution of polar and charged atoms to hydrophobicity. The hydrophobicity parameter obtained when these factors are taken into account is $\sigma_a = 31.2 \pm 4.6$ cal/mol/Å^2. It is interesting that the various

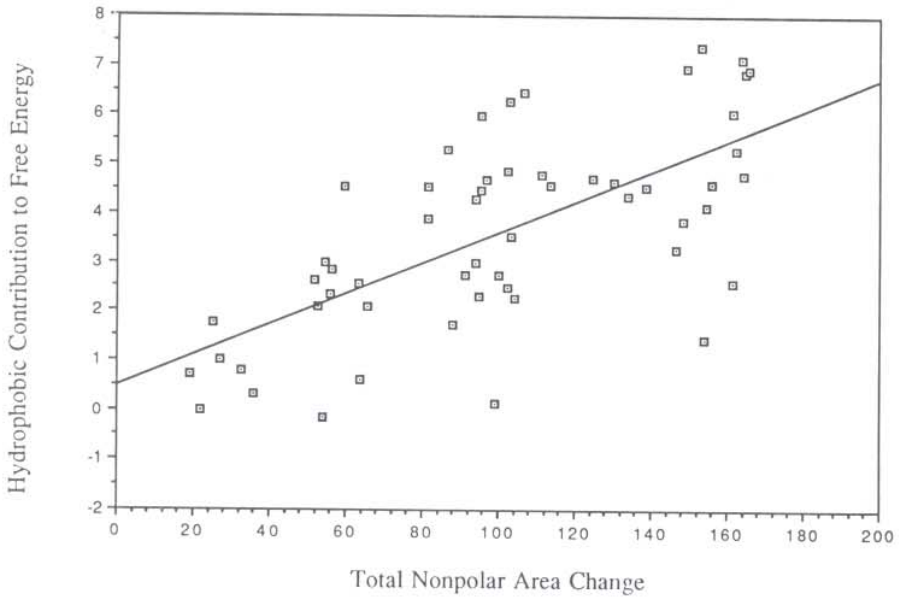

Fig. 9. Hydrophobic contribution to the free energy change resulting from mutations as a function of the change in the total non-polar surface area. The hydrophobic contribution is calculated from the observed free energy change by taking into account the electrostatic and entropic terms in Equation 19, as well as the hydrophobic contribution of polar and charged atoms. The slope of the line is 31.2 ± 4.7 cal/mol/Å^2.

additional contributions largely cancel out, and the hydrophobicity parameter is not very different from the standard value of σ_a = 34.9 cal/mol/Å^2 obtained by dividing the observed $\Delta\Delta G$ value by the difference between the accessible surface areas of wild-type and mutant side chains (Table V). Our final estimate of hydrophobicity has a reasonably small standard error of 4.6 cal/mol/Å^2. However, the large residual errors in Figure 9 reveal that, despite the more detailed model, a number of factors are still unaccounted for, mainly the effect of internal cavities (Nicholls *et al.*, 1991; Erickson *et al.*, 1992; Lee, 1993).

Because a variety of different methods have been presented in the literature for extracting hydrophobicity parameters from mutation data (Matsumura *et al.*, 1988; Kellis *et al.*, 1989; Shortle *et al.*, 1990; Nicholls *et al.*, 1991; Erickson *et al.*, 1992; Pace, 1992; Lee, 1993; Honig and Yang, 1995), and some of the conclusions seem to conflict, effort is required to show that our hydrophobicity parameter value of σ_a = 31.2 cal/mol/Å^2 is actually in good agreement with the previous results. Consider Leu→Ala mutations. The observed $\Delta\Delta G$ values vary between 1.6 and 6.2 kcal/mol, with an average of 3.5 ± 1.1 kcal/mol (Pace, 1992). The octanol–water transfer free energies for Ala and Leu are –2.08 and –0.18 kcal/mol, respectively, resulting in a difference of 1.9 kcal/mol, thus slightly more than half of the average value obtained in the unfolding experiments.

Fersht (1985) suggested that the discrepancy between the transfer free energies and the mutation data can be explained in terms of (i) two surfaces involved in the mutation, and (ii) the loss of van der Waals interactions if a cavity is formed. To check argument (i), we tried to relate the free energy change to the octanol–water transfer free energy. The Leu→Ala mutation reduces the area of the side chain by 79.4 Å^2. The hydrophobicity parameter derived from octanol–water partition varies from 16.7 to 22.1 cal/mol/Å^2 (Table V), resulting in a free energy change of between 1.3 and 1.8 kcal/mol. Doubling the accessible surface area brings the free energy change into the range 2.7–3.6 kcal/mol, which agrees with the observed average value of 3.5 kcal/mol.

Fersht (1985) and later Kellis *et al.* (1989) doubled the change in the side-chain area based on the assumption that, upon burying a hydrophobic side chain in the protein interior, two hydrocarbon–water interfaces are eliminated: that of the side chain and that of the preformed cavity. In contrast, they also assumed that on transferring a side chain into a hydrocarbon liquid, only the surface of the side chain is removed from the water. Thus, the model seems to explain why there should be twice the free energy of transfer to a protein than to a solvent. However, we have shown that the total change in the solvent accessible surface area, on average, is larger than the change in the side-chain area itself only by a factor of 1.4 rather than by a factor of 2.0, and thus the evaluation method proposed by Kellis *et al.* (1989) cannot be justified completely.

Argument (ii) of Fersht (1985), i.e. the contribution of lost van der Waals interactions, has been emphasized by Erickson *et al.* (1992), who observed that the $\Delta\Delta G$ values for a set of mutants of the phage T4 lysozyme correlated with the size of the cavity observed in the crystal structure of these mutants. More specifically, the change in the free energy has been described by equations of the form $\Delta\Delta G = a + b\Delta V_c$ and $\Delta\Delta G = a + c\Delta A$, where a = 1.9 kcal/mol, b = 24 cal/mol/Å^3, c = 20 cal/mol/Å^2, ΔV_c is the increase in cavity volume and ΔA is the increase in the cavity surface area (Erickson *et al.*, 1992). The most likely origin of the cavity-dependent term is that the Leu→Ala replacement destabilizes the protein not only because of the reduction in hydrophobic stabilization, but also because the mutation to a smaller side chain may create a cavity, thereby removing favorable van der Waals interactions. Depending on the flexibility of the local environment, conformational rearrangements reduce the cavity created by the mutation. Because flexibility varies from site to site, the latter effect is primarily responsible for the large variability of the observed free energies (Figure 6).

The equations obtained by Erickson *et al.* (1992) clearly show that, for a given mutation, the smallest change in the free energy of unfolding is attained if the structure is completely relaxed, eliminating any newly created cavity. $\Delta V_c = 0$ implies $\Delta\Delta G = 1.9$ kcal/mol, and values as low as 1.6 kcal/mol have indeed been observed for the Leu→Ala mutation (Shortle *et al.*, 1990; Pace, 1992). Because 1.9 kcal/mol is exactly the difference between the octanol–water transfer free energies of Ala and Leu, accounting for the lost van der Waals effects, the discrepancy between $\Delta\Delta G$ and the octanol–water data seems to be completely reconciled. However, this analysis does not account for the 0.8 kcal/mol change in side-chain conformational entropy associated with the Leu→Ala mutation (Pickett and Sternberg, 1993). Because the entropic term destabilizes the folded conformation of the wild-type protein but does not affect the mutant with Ala, the smallest observed changes in the free energy (1.6–1.9 kcal/mol) indicate hydrophobic contributions in the range 2.4–2.7 kcal/mol. This conclusion follows directly from Equation 19 because the average value of the electrostatic contributions ($E^{el}_{pm} - E^{el}_{pw}$) is very small (0.01 kcal/mol), and we consider here the limiting case of no change in cavity that implies ($\Delta G^{tr}_{pm} - \Delta G^{tr}_{pw}$) = 0. Thus, if the hydrophobic contribution is proportional to the change in the side-chain area (79.4 Å^2), the 2.4–2.7 kcal/mol hydrophobic contributions translate into hydrophobicity values from 30.2 to 33.7 cal/mol/Å^2, in good agreement with the values derived here. We note that Nicholls *et al.* (1991) and Lee (1993) studied the other extreme case, i.e. the effect of mutations in completely rigid proteins. Nicholls *et al.* (1991) concluded that the hydrophobicity parameter derived from octanol–water partitioning is too small to explain the largest possible destabilizing effects of a non-polar mutation. Although this contradicts the results of Lee (1993), the latter study used scaled particle theory, which is of uncertain validity when applied to water or non-spherical molecules such as alkanes, and described energy changes in terms of the work done to create cavities in water rather than in terms of water–solvent transfer free energies.

Consistency of hydrophobicity parameters

As shown in Table V, even with an assumed dependence on the solute size, the hydrophobicity parameter σ_a based on octanol–water transfer data does not exceed 22.1 cal/mol/Å^2. In contrast, the atomic solvation parameter $\sigma_a = 30.8$ cal/mol/Å^2, determined from hydrocarbon solubilities, is in excellent agreement with the hydrophobicity value of 31.2 cal/mol/Å^2, determined from non-disruptive mutations. The value derived from the cyclohexane–water transfer of non-polar side-chain analogs is in the same range (33.1 cal/mol/Å^2). Thus, moving a hydrophobic side chain from water into the interior of a protein is much better described by the transfer between water and hydrocarbon (or water–cyclohexane) than between water and octanol.

It is easy to provide statistics that support the above conclusion. About 11% of the solvent accessible surface area of octanol is polar. Thus, as a first approximation, neglecting the local reorganization of the molecules in the first hydration shell, we can assume that the environment of a solute in octanol includes 11% polar atoms. For comparison, we have calculated the hydrophobicity of the contact environment for the hydrophobic side chains of staphylococcal nuclease, studied by Shortle *et al.* (1990), and found the polar contact area to be <2% of the total (Table VI). In addition, transferring a hydrophobic solute into a hydrogen bonding solvent is likely to reduce the entropy of the solvent. In contrast, while substituting a side chain by a smaller one in an originally well packed protein environment may create cavities that are not completely removed by the structural relaxation, the increase in the entropy of nearby side chains is generally very small. These considerations, and the good agreement of transfer free energies, suggest that the highly structured, hydrophobic environment of a hydrophobic side chain in a protein can be better represented by a hydrophobic liquid than a partially polar, hydrogen bonding solvent like octanol.

Table VI. Contact environment of non-polar side chains in staphylococcal nuclease

Side chain	Contact area[a]		% polar
	Polar	Apolar	
Ile15	3	269	1.1
Ile18	4	277	1.4
Ile72	4	281	1.4
Ile92	0	271	0.0
Ile139	4	289	1.4
Leu7	11	281	3.8
Leu14	7	278	2.5
Leu25	0	269	0.0
Leu36	0	274	0.0
Leu37	0	281	0.0
Leu38	5	271	1.8
Leu89	3	286	1.0
Leu103	3	267	1.1
Leu108	5	276	1.8
Leu125	0	274	0.0
Leu137	31	231	11.8
Val104	0	227	0.0
Val111	0	231	0.0
Val114	6	227	2.6
Val23	0	242	0.0
Val39	0	230	0.0
Val51	14	201	6.5
Val66	0	239	0.0
Val74	0	225	0.0
Val99	0	231	0.0
Average	4	257	1.5

[a]Accessible surface area calculated by the method of Connolly (1981).

The use of $\sigma_a = 31$ cal/mol/Å^2 as the hydrophobicity parameter has allowed us to calculate the binding free energies in a variety of systems (Vajda *et al.*, 1994). Despite reasonable success in the application, the results presented here emphasize that the structured interior of a protein cannot be fully represented by a homogeneous medium. The consequences of this limitation are 2-fold. First, it is unlikely that satisfactory atomic solvation parameters can be extracted from the partitioning of amino acid analogs between water and any single solvent. Here we concentrated on the transfer of an apolar molecule (or the apolar fragment of a molecule) between water and protein versus water and an organic liquid, and established the value of 31 cal/mol/Å^2 for the hydrophobicity parameter $\sigma_a = \sigma_C$. However, because substantially less information is available, no attempt was made to establish similar consensus values for the atomic solvation parameters of polar and charged atoms. Second, from the very outset, the method suggested by Eisenberg and McLachlan (1986) and other techniques based on the use of atomic solvation parameters assume that the protein side chain will be placed in a homogeneous environment. Thus, the failure to identify a suitable liquid that can represent such a homogeneous medium indicates an intrinsic limitation of the approach, and calls for the development of

solvation models that take into account at least the pairwise interactions between polar–polar, polar–apolar and apolar–apolar atoms (Ben-Naim, 1994b; Koehl and Delarue, 1994).

Acknowledgements

This work was supported by grant no. AI30535 from the NIH/NIAID (to C.D., S.V. and Z.W.), and by grant no. DOE-FG02-93ER61656 from the Department of Energy, USA (S.V.).

References

Ben-Naim,A. (1980) *Hydrophobic Interactions*. Plenum Press, New York.
Ben-Naim,A. (1994a) *Biophys. Chem.*, **51**, 203–216.
Ben-Naim,A. (1994b) *Curr. Opin. Struct. Biol.*, **4**, 264–268.
Ben-Naim,A. and Marcus,Y. (1984) *J. Chem. Phys.*, **81**, 2016–2027.
Ben-Naim,A. and Mazo,R.M. (1993) *J. Phys. Chem.*, **97**, 10829–10834.
Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comput. Chem.*, **4**, 187–217.
Chan,H.S. and Dill,K.A. (1994) *J. Chem. Phys.*, **101**, 7007–7026.
Chothia,C. (1974) *Nature*, **248**, 338–339.
Colonna-Cesari,F. and Sander,C. (1990) *Biophys. J.*, **57**, 1103–1107.
Connolly,M.L. (1981) *Molecular Surface Program*. Quantum Chemistry Program Exchange, Program 429.
Cornette,J.L., Cease,K.B., Margalit,H., Spouge,J.L., Berzofsky,J.A. and DeLisi,C. (1987) *J. Mol. Biol.*, **195**, 659–685.
Cramer,C.J. and Truhlar,D.G. (1992) *Science*, **256**, 213–217.
Dill,K.A. (1990) *Biochemistry*, **29**, 7133–7155.
Eisenberg,D. and McLachlan,A.D. (1986) *Nature*, **319**, 199–203.
Erickson,A.E., Baase,W.A., Zhang,X.-J., Heinz,D.W., Blaber,M., Baldwin,E.P. and Matthews,B.W. (1992) *Science*, **255**, 178–183.
Fauchere,J.-L. and Pliska,V. (1983) *Eur. J. Med. Chem.*, **18**, 369–375.
Fersht,A. (1985) *Enzyme Structure and Mechanism*. Freeman, New York.
Flory,P.J. (1941) *J. Chem. Phys.*, **9**, 660–661.
Flory,P.J. (1953) *Principles of Polymer Chemistry*. Cornell University Press, Ithaca, NY.
Harvey,S. (1989) *Proteins*, **5**, 78–92.
Hermann,R.B. (1972) *J. Phys. Chem.*, **76**, 2754–2759.
Hill,T.L. (1960) *An Introduction to Statistical Thermodynamics*. Dover Publications, NY.
Holtzer,A. (1992) *Biopolymers*, **32**, 711–715.
Holtzer,A. (1994) *Biopolymers*, **34**, 315–320.
Honig,B. and Yang,A.-S. (1995) *Adv. Protein Chem.*, **46**, 27–58.
Huggins,M.L. (1941) *J. Chem. Phys.*, **9**, 440–440.
Jackson,R.M. and Sternberg,M.J.E. (1994) *Protein Engng*, **7**, 371–383.
Kang,Y.K., Gibson,G.D., Nemethy,G. and Scheraga,H.A. (1987) *J. Phys. Chem.*, **92**, 4739–4742.
Kellis,J.T.,Jr, Nyberg,K. and Fersht,A.R. (1989) *Biochemistry*, **28**, 4914–4922.
Koehl,P. and Delarue,M. (1994) *Proteins*, **20**, 264–278.
Kumar,S.K., Szleifer,I., Sharp,K., Rossky,P.J., Friedman,R. and Honig,B. (1995) *J. Phys. Chem.*, **99**, 8382–8391.
Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–111.
Lebowitz,J.L. and Rowlinson,J.S. (1964) *J. Chem. Phys.*, **41**, 133–138.
Lee,B. (1993) *Protein Sci.*, **2**, 733–738.
Lee,B. (1994) *Biophys. Chem.*, **51**, 263–269.
Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.
Matsumura,M., Becktel,W.J. and Matthews,B.W. (1988) *Nature*, **334**, 406–410.
McAuliffe,C. (1966) *J. Phys. Chem.*, **70**, 1267–1275.
Neff,R.O. and McQuarrie,D. (1973) *J. Phys. Chem.*, **77**, 413–418.
Nicholls,A., Sharp,K.A. and Honig,B. (1991) *Proteins*, **11**, 281–296.
Nozaki,Y. and Tanford,C. (1971) *J. Biol. Chem.*, **246**, 2211–2217.
Ooi,T., Oobatake,M., Nemethy,G. and Scheraga,H.A. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 3086–3090.
Pace,C.N. (1992) *J. Mol. Biol.*, **226**, 29–35.
Perkyns,J. and Pettitt,B.M. (1994) *Biophys. Chem.*, **51**, 214–214.
Pickett,S.D. and Sternberg,M.J.E. (1993) *J. Mol. Biol.*, **231**, 825–839.
Radzicka,A. and Wolfenden,R. (1988) *Biochemistry*, **27**, 1664–1670.
Reiss,H. (1965) *Adv. Chem. Phys.*, **9**, 1–83.
Reynolds,J.A., Gilbert,D.B. and Tanford,C. (1974) *Proc. Natl Acad. Sci. USA*, **71**, 2925–2927.
Rohlf,F.J. and Sokal,R.R. (1969) *Statistical Tables*. Freeman, San Francisco, CA.
Rossky,P.J., Sharp,K.A., Honig,B. and Kumar,S.K. (1994) *Biophys. Chem.*, **51**, 212–214.
Sharp,K.A., Nicholls,A., Friedman,R. and Honig,B. (1991) *Biochemistry*, **30**, 9686–9697.
Shortle,D., Stites,W.E. and Meeker,A.K. (1990) *Biochemistry*, **29**, 833–841.
Shrake,A. and Rupley,J.A. (1973) *J. Mol. Biol.*, **79**, 351–371.
Sitkoff,D., Sharp,K.A. and Honig,B. (1994) *Biophys. Chem.*, **51**, 397–409.
Smith,K.C. and Honig,B. (1994) *Proteins*, **18**, 119–132.
Stouten,P.F.W., Frommel,C., Nakamura,H. and Sander,C. (1993) *Mol. Simul.*, **10**, 97–120.
Vajda,S., Weng,Z., Rosenfeld,R. and DeLisi,C. (1994) *Biochemistry*, **33**, 13977–13987.
Wesson,L. and Eisenberg,D. (1992) *Protein Sci.*, **1**, 227–235.
Wilson,C., Mace,J.E. and Agard,D.A. (1991) *J. Mol. Biol.*, **220**, 495–506.

Received April 21, 1995; revised August 10, 1995; accepted September 6, 1995