Krimm, S., & Bandekar, J. (1986) *Adv. Protein Chem. 38*, 181.
Kumar, V. D., Lee, L., & Edwards, B. F. P. (1990) *Biochemistry 29*, 1404.
Lee, D. C., Hayward, J. A., Restall, C. J., & Chapman, D. (1985) *Biochemistry 24*, 4364.
Lee, D. C., Haris, P. I., Chapman, D., & Mitchell, R. C. (1990) *Biochemistry 29*, 9185.
Levine, B. A., Delgarno, D. C., Esnouf, M. P., Klevit, R. E., Scott, G. M. M., & Williams, R. J. P. (1983) *Ciba Found. Symp. 93*, 72-97.
Martin, S. M., & Bayley, P. (1986) *Biochem. J. 238*, 485.
Moews, P. C., & Kretsinger, R. H. (1975) *J. Mol. Biol. 91*, 201.
Satyshur, K. A., Rao, S. T., Pyzalska, D., Drendel, W., Greaser, M., & Sundaralingam, M. (1988) *J. Biol. Chem. 263*, 1628.
Surewicz, W. K., & Mantsch, H. H. (1988) *Biochim. Biophys. Acta 952*, 115.
Trewhella, J., Liddle, W. K., Heidorn, D. B., & Strynadka, N. (1989) *Biochemistry 28*, 1294.
Yasui, S. C., Keiderling, T. A., Bonora, G. M., & Toniolo, C. (1986) *Biopolymers 25*, 79.

Extracting Hydrophobic Free Energies from Experimental Data: Relationship to Protein Folding and Theoretical Models†

Kim A. Sharp, Anthony Nicholls, Richard Friedman, and Barry Honig*

Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168 Street, New York, New York 10032

Received March 27, 1991; Revised Manuscript Received June 26, 1991

ABSTRACT: Solubility and vapor pressure measurements of hydrocarbons in water are generally thought to provide estimates of the strength of the hydrophobic effect in the range 20–30 cal/(mol·Å^2). Our reassessment of the solubility data on the basis of new developments in solution thermodynamics suggests that the hydrophobic surface free energy for hydrocarbon solutes is 46–47 cal/(mol·Å^2), although the actual value depends strongly on curvature effects [Nicholls et al. (1991) *Proteins* (in press); Sharp et al. (1991) *Science 252*, 106–109]. The arguments to support such a significant increase in the estimate of the hydrophobic effect stem partly from theoretical considerations and partly from the experimental results of De Young and Dill [(1990) *J. Phys. Chem. 94*, 801–809] on benzene partition between water and alkane solvents. Previous estimates of the hydrophobic effect derive from an analysis of solute partition data, which does not fully account for changes in volume entropy. We show here how the ideal gas equations, combined with experimental molar volumes, can account for such changes. Revised solubility scales for the 20 amino acids, based on cyclohexane to water and octanol to water transfer energies, are derived. The agreement between these scales, particularly the octanol scale, and mutant protein stability measurements from Kellis et al. [(1989) *Biochemistry 28*, 4914–4922] and Shortle et al. [(1990) *Biochemistry 29*, 8033–8041] is good. The increased strength of the hydrophobic interaction has implications for the energetics of protein folding, substrate binding, and nucleic acid base stacking and the interpretation of computer simulations.

The hydrophobic effect has provided a major unifying concept in understanding the structure and function of biological systems. Hydrophobicity can be defined phenomenologically in terms of the low solubility of nonpolar molecules in water. The underlying physical interactions that are responsible for this effect are also thought to yield the attractive forces that cause nonpolar molecules to aggregate in an aqueous medium, producing, for example, what is certainly a major driving force in protein folding (Dill, 1990). A distinction is sometimes made between hydrophobicity as measured from solubilities and from intermolecular association (Wood & Thompson, 1990), but it is likely that they represent different manifestations of the same physical phenomenon: the energetically costly disruption of the hydrogen-bonding network of water by solute molecules that cannot themselves form hydrogen bonds.

A quantitative measure of the hydrophobic effect is usually obtained from the solubilities of nonpolar molecules, particularly hydrocarbons, in water. In this paper it is argued that solvation free energies that have been extracted from such experimental data have been evaluated incorrectly, which has led to a serious underestimate of the magnitude of the effect. The standard form of the expression used to relate hydrophobic free energies to observed concentrations is $\Delta G = -RT \ln K$, where K is a partition coefficient. The problem is that this expression is correct only under special circumstances. As we show in this paper, when the solute and solvent molecules are of different sizes, an additional term [that is widely used in polymer theory (Flory, 1941; Huggins, 1941)] enters into the analysis. This leads to much larger transfer free energies. This finding has a number of important implications, including a requirement for a new hydrophobicity scale, the need to reevaluate the extent of agreement between experiment and computer simulations, and the need to revise the parameters used in protein folding algorithms and in the evaluation of binding energies.

†Support from the NSF (PCM88-05434), ONR (N00014-90-J-1713), and NIH (GM-30518 and GM-41371) is gratefully acknowledged.
* To whom correspondence should be addressed.

0006-2960/91/0430-9686$02.50/0

Solute transfer experiments show that hydrophobic free energy is, to a good approximation, proportional to the surface area of the solute (Hermann, 1977; Tanford, 1980). Aliphatic hydrocarbons are almost totally nonpolar, and the transfer energy per unit area will be referred to here as the hydrophobic strength or hydrophobic surface free energy. A consensus derived from many measurements is that the transfer of an aliphatic hydrocarbon from a nonpolar solvent to water costs about 20–30 cal/(mol·Å^2), where the area is defined as the solvent-accessible area (Lee & Richards, 1971). An alternative, "macroscopic" measure can be obtained from the interfacial free energy (surface tension) between liquid hydrocarbons and water, which ranges from 72 cal/(mol·Å^2) for pentane to 75 cal/(mol·Å^2) for decane (Aveyard & Haydon, 1965). An explanation for these differing values, based on size and curvature effects on surface free energies, has recently been proposed (Nicholls et al., 1991; Sharp et al., 1991). Recent measurements by Fersht, Shortle, and co-workers (Kellis et al., 1989; Shortle et al., 1990) appear to show that hydrophobic amino acids contribute 45–60 cal/(mol·Å^2) to protein stability, i.e., an intermediate value between the microscopic and macroscopic measurements. In order to understand why these types of measurements apparently give very different results, it is necessary to consider in detail what interactions are actually probed by each. In this paper, we reexamine the thermodynamics of solubility measurements and, on this basis, obtain a revised amino acid solubility scale. This scale is consistent with recent protein stability measurements.

Theory

General Considerations. When a solute partitions between two phases, the condition of equilibrium is that its chemical potential is the same in both phases. Since the chemical potential is the partial derivative of the free energy with respect to the number of moles of solute, at equilibrium the partial molar free energy of transferring the solute between the phases is zero. Since the composition of the two phases is different, the intermolecular interactions the solute makes with its surroundings in each phase are different. For the net transfer free energy to be zero, this difference must be counterbalanced by a free energy difference arising just from the different concentration of components. More formally, this concentration-dependent free energy difference arises from the translational partition functions of the components, which in turn depend on the volume in which they are contained (McQuarrie, 1976). This concentration dependence is sometimes ascribed to "entropy of mixing", but Ben-Naim has argued that the use of the term "mixing" is misleading, if not erroneous (Ben-Naim, 1987). For want of a better word, we use the term "volume entropy" (denoted by the superscript "ve") to refer to the volume or concentration-dependent translational term. Since the volume dependence of the translational free energy is also the origin of the ideal gas behavior, one can simply account for the volume entropy changes in any transfer process by calculating the equivalent ideal gas work, as we show below.

In a dilute, or ideal, gas, the solute makes no intermolecular interactions. If we transfer the solute into a liquid or solution phase where it makes interactions with its surroundings, a certain amount of chemical work is done by these interactions. This may be termed the solvation free energy [or the coupling work of introducing the solute into a fixed position in the solution (Ben-Naim, 1978)]. Since this free energy contribution is over and above that of an ideal gas (i.e., in addition to any volume entropy contribution), it is referred to as the excess chemical potential. Borrowing from Gurney (1953), we shall use the term "unitary" (denoted by the superscript "°") to refer to the intermolecular interaction dependent, or excess, free energy contribution—the word unitary being taken to mean that it should depend only on the strength of interaction of a single solute with the solvent, independent of the concentrations of the component in solution. With these definitions, the condition of equilibrium for a solute partitioning between two phases may now be written $\Delta\bar{G}_s^\circ = T\Delta\bar{S}_s^{ve}$, where the unitary term involves intermolecular interactions and is independent of concentration, and the volume entropy term depends on concentration and is independent of intermolecular interactions.

For a nonpolar solute in water, the solute–solvent interaction is the origin of the hydrophobic effect. Thus ability to extract hydrophobic free energies from experimental partition or solubility data is contingent upon being able to correctly calculate the volume entropy contribution from the solution composition to obtain the unitary free energy. In the following sections, we derive expressions that relate the unitary free energies of transfer of a solute from one phase to another to the volume entropy changes and thus to experimentally measurable quantities. Expressions will be derived for (a) condensation of a solute molecule from the gas phase to its pure liquid, (b) mixing pure solute and solvent to form a solution, (c) partitioning of a solute between two solvents, and (d) transfer of a solute from the gas phase to a solution. We assume throughout that solution and gas-phase concentrations are dilute enough that solute–solute interactions can be neglected and that ideal gas and ideal *dilute* solution behavior hold, respectively.

Condensation of a Pure Solute. Consider the partition of a solute (s) between its gaseous (g) and pure liquid (l) phases. In the gas phase, the chemical potential may be written (McQuarrie, 1976)

$$\mu_s^g = RT \ln N\Lambda^3/Vq^{int} = \mu_s^{\phi g} + RT \ln \rho^g \qquad (1)$$

where N is the number of molecules in volume V, $\rho^g = N/V$ is the density, q^{int} is the internal partition function, and Λ is the deBroglie wavelength of the solute. The volume is expressed in the same units as Λ^3. The term $RT \ln (\Lambda^3 N/V)$ arises from the translational partition function and is sometimes referred to as the liberation free energy of the solute (Ben-Naim, 1978). The term $\mu_s^{\phi g}$ contains those contributions that remain constant in any transfer process and may be omitted from the discussion hereon. For the liquid phase, we write

$$\mu_s^l = \mu_s^{\phi g} + \Delta\bar{G}^\circ_s(g \rightarrow l) + RT \ln \rho^l \qquad (2)$$

where $\Delta\bar{G}^\circ_s(g \rightarrow l)$ is partial molar free energy of "coupling" the solute to the liquid through the intermolecular potentials, i.e., the unitary free energy. At equilibrium, the solute chemical potential is the same in both phases: $\mu_s^g = \mu_s^l$. From eqs 1 and 2 we have

$$\Delta\bar{G}^\circ_s(g \rightarrow l) = -RT \ln \rho^l/\rho^g \qquad (3)$$

We can also describe eq 3 in terms of transfer processes as follows (referring to Figure 1): The condensation, or transfer, of a mole of solute (step 1) can be divided into two separate steps. First, the solute is compressed to its liquid density with no intermolecular interactions, i.e., as an ideal gas (step 2). This step involves the isothermal reversible ideal gas work (representing volume entropy changes)

$$-T\Delta\bar{S}^{ve} = RT \ln V_s^g/V_s = RT \ln \rho^l/\rho^g \qquad (4)$$

where V_s^g and V are the initial (gas phase) and final (liquid

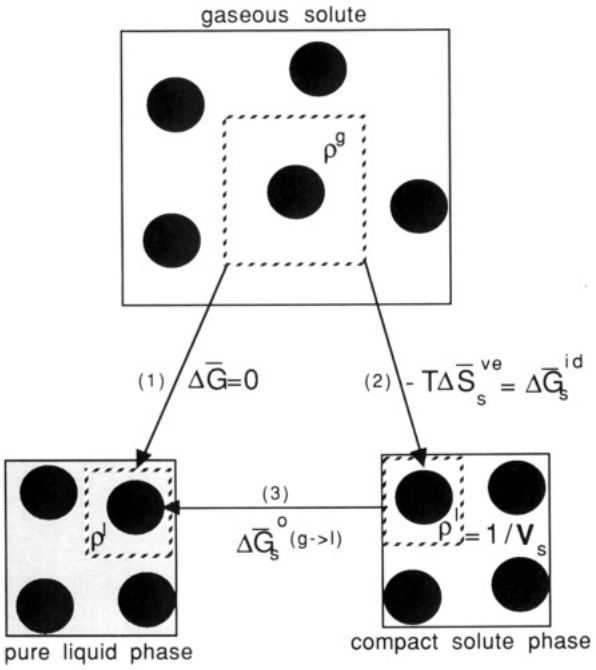

FIGURE 1: Gas to liquid condensation. Solute molecules are shown as black circles, an absence of shading indicates the gas phase (no intermolecular interactions), and shading indicates the liquid phase. Other symbols are defined in the text. (Step 1) Transfer of a molecule of solute from gas to liquid phases at equilibrium. (Step 2) An isothermal ideal gas compression that accounts for the volume entropy change. Dashed boxes illustrate the change in volume available per solute molecule upon transfer. (Step 3) Switching on the intermolecular interactions.

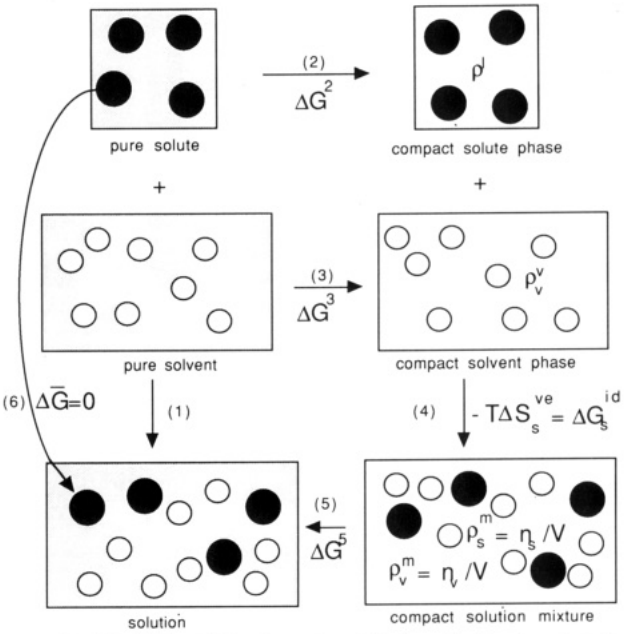

FIGURE 2: Liquid to solution transfer. Solute and solvent molecules are shown as black and white circles, respectively. An absence of shading indicates the gas phase (no intermolecular interactions); shading indicates the liquid phase. Other symbols are defined in the text. (Step 1) Mixing of pure solute and solvent. (Step 2) Switching off the solute–solute interaction. (Step 3) Switching off solvent–solvent interactions. (Step 4) An isothermal ideal gas expansion that accounts for the volume entropy change upon mixing. (Step 5) Switching on solvent–solvent and solvent–solute interactions in solution. (Step 6) Transfer of a solute molecule between phases at equilibrium.

phase) molar volumes occupied by the solute, respectively. This results in a "compact ideal gas state". Note that the ideal gas work is just the change in translational free energy (volume entropy) of the solute that accompanies the actual change in its molar volume upon condensation.

In the next step, intermolecular interactions between the solute molecules are turned on (step 3), with free energy $\Delta\bar{G}^\circ_s(g \rightarrow l)$. Since at equilibrium $\Delta\bar{G}^\circ_s(g \rightarrow l) - T\Delta\bar{S}^{ve} = 0$, we again have eq 3.

Mixing Solute and Solvent To Form a Solution. The treatment for a pure solute may be generalized to any transfer process. Consider the mixing of n_s moles of a pure liquid solute (s) with n_v moles of a pure solvent (v) to form a dilute solution or mixture (m) (Figure 2, step 1). We first switch off the intermolecular interactions in both solute (step 2) and solvent (step 3) to form the compact ideal gas states with molar volumes V_s and V_v for solute and solvent, respectively. These are then mixed to form a compact ideal gas mixture of volume $V = n_sV_s + n_vV_v$ (step 4), where we assume volume additivity. The contribution from volume entropy changes is given by the reversible isothermal ideal gas expansion work done on both solute and solvent upon mixing:

$$-\Delta S^{ve}/R = n_s \ln (n_sV_s/V) + n_v \ln (n_vV_v/V) \quad (5)$$

Finally, the intermolecular interactions are switched on in the mixture (step 5).

The partial molar volume entropy change upon transferring a molecule of solute from the liquid to solution phase, $\Delta\bar{S}^{ve}$, is obtained from the derivative of eq 5 with respect to n_s at constant n_v, P, and T:

$$-T\Delta\bar{S}^{ve} = RT \ln \phi_s + RT(1 - r) = RT \ln (\rho^m/\rho^l) + RT(1 - r) \quad (6)$$

where $\phi_s = n_sV_s/V$ is the volume fraction concentration of solute, $r = V_s/V_v$ is the solute to solvent molar volume ratio, and we have taken the dilute solution limit of $\phi_s \ll 1$. Step 4 is the only process that contributes to the volume entropy since it is the only process involving a volume change. The sum of steps 2, 3, and 5 describe all changes in interaction free energy. The unitary free energy change for the transfer process, $\Delta\bar{G}^\circ_s(l \rightarrow m)$, is thus given by the partial derivative of $\Delta G^2 + \Delta G^3 + \Delta G^5$ with respect to n_s. At equilibrium, $\Delta\bar{G}^\circ_s(l \rightarrow m) - T\Delta\bar{S}^{ve} = 0$ (step 6), giving

$$\Delta\bar{G}^\circ_s(l \rightarrow m) = T\Delta\bar{S}^{ve} = -RT \ln (\rho^m/\rho^l) - RT(1 - r) \quad (7)$$

Note that if the molar volumes of solute and solvent are the same then eq 7 becomes

$$\Delta\bar{G}^\circ_s(l \rightarrow m) = -RT \ln (\rho^m/\rho^l) = -RT \ln K_M \quad (8)$$

where K_M is the molarity partition coefficient. Equation 8 is commonly used for obtaining transfer free energies [e.g., Ben-Naim and Marcus (1984), Radzicka and Wolfenden (1988), and Wolfenden et al. (1981)]. With the assumption that $V_s = V_v$, eq 5 reduces to

$$-\Delta S^{ve}/R = n_s \ln X_s + n_v \ln X_v \quad (9)$$

where X_s and X_v are the mole fractions of solute and solvent in the mixture, respectively. Equation 9 is the standard expression given in physical chemistry texts for "entropy of mixing" in solution theory (Atkins, 1986). From it one derives the familiar equation (Tanford, 1980)

$$\Delta\bar{G}^\circ_s(l \rightarrow m) = \mu^{om} - \mu^{ol} = -RT \ln X_s \quad (10)$$

where μ^{ol} and μ^{om} are identified by Tanford (1980) as the standard-state chemical potentials of the solute in the liquid and solution phases, respectively. Equation 10 can also be obtained by putting $V_s = V_v$ into eq 8. Since the mole fraction of solute in the pure liquid is one, X_s is also the mole fraction partition coefficient, K_X. Equations 8–10, however, are only valid for special cases.[1] Since water is a particularly small

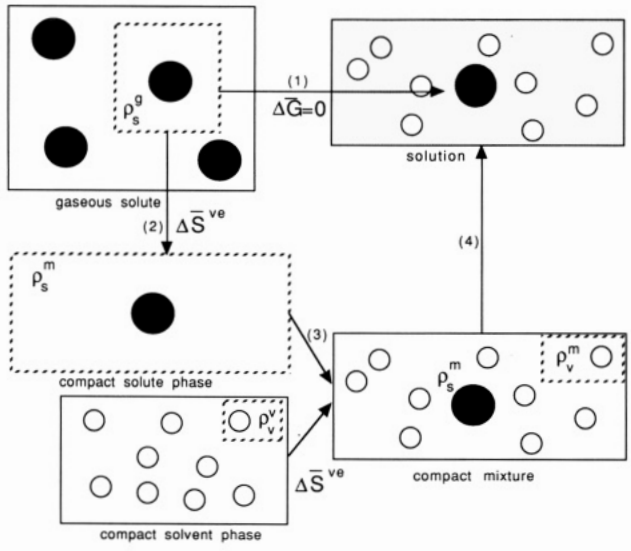

FIGURE 3: Gas to solution transfer. Solute and solvent molecules are shown as black and white circles, respectively. An absence of shading indicates the gas phase (no intermolecular interactions); shading indicates the liquid phase. Other symbols are defined in the text. (Step 1) Transfer of a solute molecule at equilibrium. (Step 2) An isothermal ideal gas expansion of the solute to its solution density. (Step 3) An isothermal ideal gas expansion of the solvent into the final solution volume. (Step 4) Switching on the solvent–solvent and solute–solvent intermolecular interactions (A prior step in which pure solvent interactions are switched off is omitted from the diagram for clarity).

solvent (Lee, 1985), it should be noted that different size solute/solvent solutions are the rule rather than the exception.

Solute Partitioning between Two Solvents. For a solute partitioning between two solvents a and b, with molar volumes V_a and V_b, respectively, we use eq 7 for the solute transfer to each solvent and by subtraction obtain

$$\Delta\bar{G}^\circ_s(a \rightarrow b) = -RT \ln (\rho^b/\rho^a) + RTV_s(1/V_b - 1/V_a) \quad (11)$$

Transfer of Solute from Gas Phase to Solution. The transfer of a solute from gas phase to solution (step 1, Figure 3) is just the sum of gas to liquid and liquid to solution transfers. Adding eqs 3 and 7 gives

$$\Delta\bar{G}^\circ_s(g \rightarrow m) = -RT \ln (\rho^m/\rho^g) - RT(1- r) \quad (12)$$

where $\Delta\bar{G}^\circ_s(g \rightarrow m) = \Delta\bar{G}^\circ_s(g \rightarrow l) + \Delta\bar{G}^\circ_s(l \rightarrow m)$. It might seem strange that the molar volume of the solute as a pure liquid, V_s, enters into the gas to solution transfer expression. This is because under the assumption of volume additivity used in deriving equation 5 (thus neglecting any partial molar volume change upon mixing), the partial molar volume of the solute in solution is V_s. If experimental partial molar volume data are available, they should be used directly in eqs 7, 11, and 12.

Comparing the two processes of condensing a vapor into a liquid and the dissolution of a gaseous solute in a solvent (eqs 3 and 12, referring to step 1 of Figures 1 and 3, respectively), we find that the volume entropy term on the right-hand side in both cases involves the ideal gas work of compressing the solute from its gas phase density to its actual density at equilibrium in the compact phase. The dissolution in solvent contains an extra ideal gas term, however, arising from the volume entropy contribution from the solvent. The significance of this term is illustrated by the following, more heuristic, argument. The solute gas is compressed to its solution density (step 2 of Figure 3), and it is introduced into the solvent (step 3). Assume that the solute is larger than the solvent. The volume available per mole of *solvent* is increased, compared to that of adding an equivalent number of solvent molecules. The physical origin of this term may be associated with the volume entropy change of the solvent expanding into the extra volume. The molar volume difference is $\delta V = V_s - V_v$. Since the solution is dilute, the ideal gas pressure of the solvent before and after the addition is that for the pure solvent density, $P_v^{id} = RT/V_v$, and the volume entropy contribution is

$$-P^{id}\delta V = -RT(V_s - V_v)/V_v = RT(1 - r) \quad (13)$$

We may say that the solvent effectively "expands" into a larger volume, on a molar basis, leading to a more favorable free energy than if the solute and solvent were of the same size. Note that this term favors the dissolution of a large solute in a small solvent.

From the definition of volume entropy and unitary free energies given above, it is clear that a size ratio term $RT(1 - r)$ must generally be included when analyzing any data obtained from experimental measurements on solutions before solute–solvent interaction energies can be extracted. This includes vapor pressure, solubility, and partition coefficient measurements. If the original data were analyzed with the mole fraction concentration scale, concentrations must first be converted into molarities. Although we are concerned here primarily with hydrophobic solutes, these remarks also apply to polar and ionic solutes as well.

To summarize the results of this section, at equilibrium the partial molar free energy of solute transfer between two phases is zero. To extract a meaningful solvation free energy from experimental data, i.e., one that provides information about intermolecular interactions, one must separate out the all contributions to the transfer process that involve volume entropy (i.e., translational free energy) changes. We argue here that one should do this using the ideal gas work corresponding to the actual experimental molar volumes and concentrations.

Experimental Procedures

Calculation of Accessible Surface Areas, Molar Volumes, and Transfer Energies. Accessible areas, as defined by Lee and Richards (1971) of alkanes, *N*-acetyl amino acid amides, and amino acid side-chain analogues (RH) were calculated with the program DELPHI (Gilson et al., 1988; Sharp & Nicholls, 1990) using a radius of 1.4 Å for water and the radii of Bondi (1968) as applied by Richards (1974).

Molar volumes of water, cyclohexane, octanol, liquid alkanes, and amino acids analogues were calculated by dividing the molecular weight by the density obtained from the CRC Handbook of Chemistry and Physics (Lide, 1990). Molar volumes of gaseous alkanes were taken from McAuliffe (1966). If there is an appreciable partial molar volume change upon mixing, then the volume occupied in solution will be different from the molar volume. Data on partial molar volumes of hydrophobic solutes in water is sparse due to the difficulty of such measurements on insoluble substances. The data of Masterton (1954) indicate essentially no partial molar volume change between liquid methane and methane in water and about a 15 cm^3/mol decrease for both ethane and propane. The data of Friedman and Scheraga (1965) show a much smaller decrease of about −2 cm^3/mol per CH_2 group. In either case, the changes are much smaller than the actual

[1] Namely, ideal solutions, i.e., ones that obeys Raoult's law at all concentrations. The chemical potential in an ideal solution is $\mu = \mu^\circ + kT \ln X$, where μ° refers to a standard state of pure solute. Since ideal solution conditions are only approached for molecules of very similar size, shape, and intermolecular interactions, i.e., isotopic mixtures, use of this expression for the chemical potential does not allow a clean separation of translational free energy and intermolecular interactions as does the present treatment. Equations 8–10 therefore do not apply to most real solutions.

Table I: Transfer Energies of *n*-Alkanes to Water[a]

n	ρ (g/cm³)	V (cm³/mol)[b]	area (Å²)	liquid to water transfer energy			vapor to water transfer energy	
				$-RT \ln X$	$-RT \ln K_M$	$-RT [\ln K_M + (1 - r)]$[c]	$-RT \ln K_M$	$-RT [\ln K_M + (1 - r)]$[c]
1	0.41	39 (38)	130	3.06	2.59	3.29 (3.26)	1.93	2.63 (2.60)
2	0.46	65 (52)	163	3.92	3.14	4.72 (4.28)	1.77	3.34 (2.90)
3	0.544	81 (68)	194	4.82	3.92	6.01 (5.58)	1.98	4.07 (3.64)
4	0.579	100	224	5.81	4.79	7.53	2.15	4.89
5	0.626	115	252	6.85	5.74	8.97	2.34	5.57
6	0.660	130	282	7.79	6.60	10.35	2.55	6.29
7	0.684	146	311	8.58	7.32	11.59	2.63	6.90
8	0.703	162	340	9.53	8.21	13.02	2.89	7.70
9	0.718	178	369	10.39	9.02	14.36	3.04	8.38
10	0.730	195	399	11.28	9.85	15.73	3.22	9.10
	slope [cal/(mol·Å²)]			31 ± 0.1	28 ± 0.2	47 ± 0.4	5 ± 0.2	24 ± 0.1
	intercept (kcal/mol)			−1 ± 0.2	−1 ± 0.3	−3 ± 1	1 ± 0.4	−0.6 ± 0.3
	R^2			0.998	0.997	0.996	0.944	0.999

[a] Energies are in kcal/mol. Uncorrected data are taken from McAuliffe (1966) and Ben Naim and Marcus (1984). Symbols: *n*, number of carbon atoms; ρ, liquid density at 20 °C, except for gaseous alkanes; V, molar volume; X, mole fraction of alkane in water at saturation; K_M, molarity partition coefficient; r, molar volume ratio of solute to water. The regression data are for a linear fit to the energy as a function of accessible area. Uncertainties in fitted slopes and intercepts represent one standard deviation. [b] The figures in parentheses are experimental partial molar volumes in water, taken from Masterton (1954). [c] The figures in parentheses were calculated by using the experimental partial molar volume.

molar volumes. We have therefore taken the "size" of a molecule to be its molar volume as a pure component. The advantage is that experimental molar volumes are widely available.

Size corrections were also made by using partial molar volumes of methane, ethane, and propane in water taken from Masterton (1954). Molar volumes of *N*-acetyl amino acid amides whose densities were unavailable were calculated by adding the molar volume of *N*-acetyl-*N*-methylglycine $V \approx 100$ cm³/mol to the molar volume of the side chain analogue. This assumes that molar volumes are approximately additive. Partial molar volume measurements for alcohols in water (Friedman & Scheraga, 1965) indicate that molar volumes are in fact fairly additive. In any case, the resulting size corrections to the transfer energies are insensitive to the exact values ±20 cm³/mol taken for these compounds.

For the liquid alkanes, liquid to vapor transfer energies were taken from Ben-Naim and Marcus (1984). For the gaseous alkanes, liquid to vapor transfer energies were calculated from vapor pressures taken from the CRC Handbook of Chemistry and Physics by using the molar volumes and eq 3. The exception is methane, for which the vapor pressure at 25 °C is unavailable. A linear fit to the data for octane through ethane gives $\Delta G(l \rightarrow g) = 0.68$ kcal/mol·carbon ($r = 0.95$), and extrapolation to methane gives 0.66 kcal/mol. The vapor to water solvation data for alkanes were taken from Ben-Naim and Marcus (1984), based on the data of McAuliffe (1966). Vapor to water and cyclohexane to water data for the amino acid side-chain analogues were taken from Radzicka and Wolfenden (1988). Octanol to water data for the *N*-acetyl amino acid amides were taken from Fauchere and Pliska (1983). All data derived from vapor pressure measurements were corrected by using the molar volumes and eq 12, solubility data were corrected by using eq 7, and the cyclohexane–water and octanol–water partition data were corrected by using eq 11. Solvent molar volumes (V_v) of water, cyclohexane, and octanol are 18, 108, and 157 cm³/mol, respectively (Lide, 1990).

Protein folding can be considered as a conformational rearrangement of a solute; the protein is simply changing shape. As such, there is no transfer between phases, and, to a first approximation, the volume of the solution remains constant. For this reason the unfolding free energies are obtained from denaturation experiments directly with no volume entropy correction.

Results

Alkane Solubility and Partition Coefficients. We first consider the process of alkane dissolution in water. Two independent measures are available, the first using solubilities, which measure the partition of the solute between itself as a pure liquid and water, giving $\Delta\bar{G}°(l \rightarrow w)$. The second are vapor pressure measurements, giving the vapor to water transfer energy, $\Delta\bar{G}°(g \rightarrow w)$. These quantities are related by $\Delta\bar{G}°(l \rightarrow w) - \Delta\bar{G}°(g \rightarrow w) = \Delta\bar{G}°(l \rightarrow g)$, the liquid to vapor transfer, or vaporization, energy. Table I contains the liquid to water transfer energies for various *n*-alkanes derived by using the mole fraction partition coefficient (eq 10), the molarity partition coefficient (eq 8), and the molarity partition coefficient plus the $RT(1 - r)$ correction for the difference in alkane and water size (eq 7). The transfer energies were fit to a linear function of area, the regression data also being tabulated in Table I. In each case the plot is highly linear, but the slopes, which provide a measure of the hydrophobic surface free energy, vary significantly. The molarity partition coefficient gives a slightly lower value than that obtained from the analysis of Tanford (1980), which used the mole fraction partition coefficient, decreasing the slope from 31 to 28.9 cal/(mol·Å²). The volume correction term has the largest effect on transfer energies, and one that increases with area since the correction increases as the molar volume increases. In this case the slope is significantly increased to 46.7 cal/(mol·Å²).

Table I also gives the vapor to water transfer energies for the alkanes. Again the volume correction makes the transfer energies significantly more positive. The major qualitative difference is for the small alkanes. With the uncorrected data (using the molarity partition coefficient) the solvation energies of methane through butane appear first to decrease, then increase to essentially the same as for methane. In contrast, for the corrected data the cost of solvation scales with area.

In Figure 4 the liquid to water and vapor to water transfer energy data from Table I are plotted as a function of accessible surface area. For the liquid to water transfer data, the mole fraction partition coefficient, molarity partition coefficient, and the size corrected lines converge as the size of the solute approaches that of water, intersecting around the accessible area of water (98 Å²). The intercepts at zero area have negative energies in the range −1.2 to −3.5 kcal/mol.

Cyclohexane/water molarity partition coefficients are available from the studies of Wolfenden and co-workers

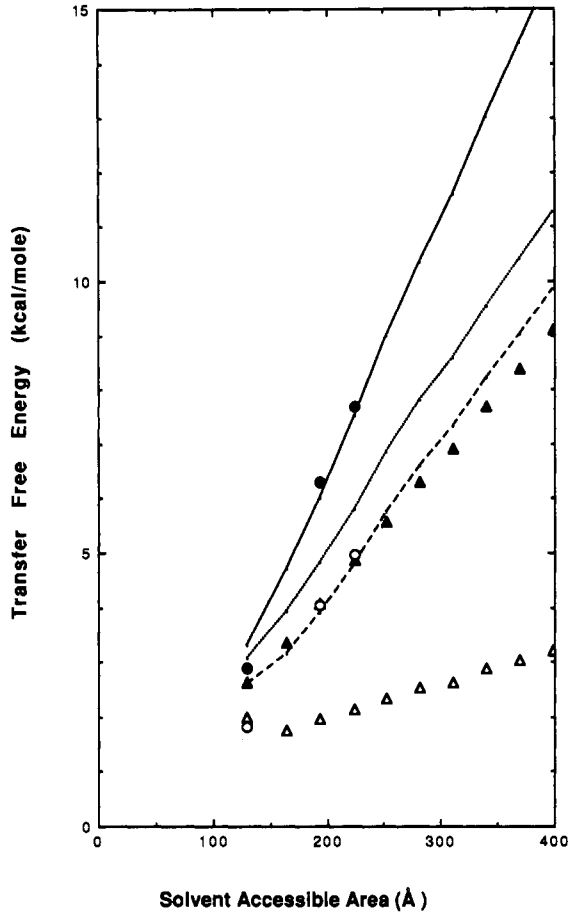

FIGURE 4: Energetics of *n*-alkane hydrocarbon transfer into water. Liquid to water transfer, taken from McAuliffe (1966), with mole fraction partition coefficients (···), molarity partition coefficients (---), and molarity partition coefficients plus size correction (—). Vapor to water transfer energies were taken from Ben-Naim and Marcus (1984) uncorrected (molarity partition coefficients) (△) and corrected for size (▲). Cyclohexane to water transfer energies were taken from Radzicka and Wolfenden (1988) uncorrected (molarity partition coefficients) (○) and corrected for size (●).

(Radzicka & Wolfenden, 1988; Wolfenden et al., 1981) for the amino acid side-chain analogues, methane, butane, and propane. These data, also plotted in Figure 4, fall very close to the alkane molarity partition coefficient line, except for methane, which lies below. When cyclohexane/water partition data are corrected, they fall very close to the corresponding line for pure alkanes, the deviation for methane decreasing. The size correction for the cyclohexane → water data is somewhat larger than that for the pure alkane data. This is because there is a correction for the cyclohexane as well for the water phase. Since the small alkanes are smaller than cyclohexane and larger than water, both corrections increase the transfer energy.

Partition Coefficients of Amino Acids. While a variety of amino acid derivatives and solvents have been examined in the literature, we focus on just three sets of data: (i) octanol → water partition of the derivatized amino acids, *N*-acetyl amino acid amides (Fauchere & Pliska, 1983), (ii) cyclohexane → water partition of amino acid side-chain analogues, i.e., the amino acids truncated at the β-carbon position (Radzicka & Wolfenden, 1988), and (iii) vacuum → water partition of amino acid side-chain analogues (Wolfenden et al., 1981). These data were chosen since they comprise complete sets for all the amino acids, with the same solvent and the same experimental method. Vacuum forms the simplest nonpolar "solvent" from a theoretical point of view since the solute can make no interactions with it. Next in order of simplicity is a fairly rigid hydrocabon solvent such as cyclohexane. Since this can make only van der Waals interactions, it is a good choice of model solvent. It also dissolves very little water. From a theoretical point of view, cyclohexane, which is completely nonpolar, is preferable to alcohols, dioxane, or carbon tetrachloride for studying the purely hydrophobic properties of the amino acids. The other solvents have been used at various times, in part because they are believed to simulate the interior of a protein, which has polar as well as hydrophobic characteristics. However, if one wishes to gain insight by separating variables, these solvents are complicated by hydrogen bonding and different electrostatic moments. Octanol is probably the best choice for a partially polar solvent since it is considerably more hydrophobic than other commonly used alcohols and has the advantage (relative to cyclohexane) of solubilizing blocked amino acids. A disadvantage of octanol is that presence of the polar hydroxyl plus the relatively high solubility of water in octanol complicates the interpretation of polar and charged solute partition data.

Table II: Partition of Amino Acid Analogues between Cyclohexane or Vapor and Water[a]

amino acid	A (Å²)	V (cm³/mol)	transfer free energy (kcal/mol) uncorrected c → w	uncorrected g → w	corrected for size c → w	corrected for size g → w
Ala	130	39	1.81	1.94	2.89	2.63
Arg	286	92	−14.92	−19.92	−12.37	−17.46
Asn	199	59	−6.64	−9.68	−5.00	−8.31
Asp	194	57	−8.72	−10.95	−7.13	−9.64
Cys	176	55	1.28	−1.24	2.82	0.01
Gln	224	79	−5.54	−9.38	−3.35	−7.35
Glu	222	75	−6.81	−10.24	−4.74	−8.36
Gly	58	19	0.94	2.39	1.47	2.42
His	247	79	−4.66	−10.27	−2.47	−8.25
Ile	223	100	4.92	2.15	7.70	4.89
Leu	225	106	4.92	2.28	7.85	5.20
Lys	261	98	−5.55	−9.52	−2.82	−6.84
Met	243	90	2.35	−1.48	4.86	0.93
Phe	271	106	2.98	−0.76	5.93	2.18
Ser	161	40	−3.4	−5.06	−2.28	−4.31
Thr	190	58	−2.57	−4.88	−0.95	−3.54
Trp	321	122	2.33	−5.88	5.73	−2.40
Tyr	285	106	−0.14	−6.11	2.81	−3.17
Val	194	81	4.04	1.99	6.29	4.07

[a] Uncorrected data are taken from Radzicka and Wolfenden (1988) and Wolfenden et al. (1981), with molarity partition coefficients (eq 9). Symbols: A, accessible area; V, molar volume; c → w, cyclohexane to water; g → w, vapor to water. Transfer free energies corrected to account for solute–solvent size differences were obtained by using eqs 11 and 12 for cyclohexane to water and gas phase to water transfers, respectively.

Table II contains the data for the amino acid analogues for transfer from vacuum and cyclohexane, and Table III contains the octanol data. The uncorrected data are taken directly from the published values, which are based on molarity partition coefficients (i.e., eq 8). The transfer energies are corrected for the solute–solvent size difference by reanalyzing the data with eq 11 and 12. This has the most dramatic overall effects for the octanol–water transfer free energies, which all become positive rather than negative. The molar volume correction is sizable for each solvent, especially for the larger amino acids such as tryptophan and tyrosine. Neutral arginine is clearly the least hydrophobic amino acid. Tryptophan is slightly polar in the vapor data, mildly hydrophobic in the cyclohexane data, and the most hydrophobic in octanol. The ambiguous nature of tryptophan has been discussed previously (Kyte & Doolittle, 1982; Radzicka & Wolfenden, 1988). Isoleucine and leucine appear the most hydrophobic in the vapor and cyclohexane

Table III: Partition of *N*-Acetyl Amino Acid Amides between Octanol and Water[a]

amino acid	*V* (cm³/mol)	*A* (Å²)	transfer free energy (kcal/mol) uncorrected	corrected for size
Ala	231	347	−2.08	2.03
Arg	318	468	−3.88	1.78
Asn	264	379	−3.29	1.40
Asp	261	378	−3.55	1.09
Cys	258	360	−0.40	4.19
Gln	297	408	−2.80	2.48
Glu	290	405	−3.37	1.78
Gly	198	323	−2.50	1.01
His	297	432	−2.32	2.95
Ile	332	402	−0.04	5.87
Leu	341	399	−0.18	5.89
Lys	329	445	−3.85	2.01
Met	316	425	−0.82	4.80
Phe	342	429	−0.05	6.03
Ser	233	352	−2.55	1.59
Thr	263	370	−2.14	2.53
Trp	369	470	0.57	7.14
Tyr	342	447	−1.19	4.90
Val	300	375	−0.83	4.51
Pro	300	364	−1.83	3.51

[a] Uncorrected data are taken from Fauchere and Pliska (1983) with molarity partition coefficients (eq 9). Symbols: *A*, accessible area; *V*, molar volume. Equation 11 was used to correct for solute–solvent size differences.

data. The results for the different solvents correlate reasonably well although the octanol data span a much smaller range of free energies. The corrected transfer energies cover a range of 24 kcal/mol for vacuum, 21 kcal/mol for cyclohexane, and only 6 kcal/mol for octanol.

Figure 5 shows a comparison of the different nonpolar solvents for the aliphatic amino acids using corrected transfer energies from Tables I–III. Transfer free energies from octanol (for Gly, Ala, Val, and Ile), from cyclohexane (the analogues H, CH_4, C_3H_8, and C_4H_{10}), and from pure alkane (the alkanes CH_4, C_2H_6, C_3H_8, and C_4H_{10}, there being no equivalent for H/Gly) to water are plotted against solvent-accessible surface area. Since the octanol data are obtained with blocked peptides, the polar portion of the molecule contributes to both the area and transfer energy, so only differences in quantities can be compared. Energies and areas for the blocked peptides in Figure 5 are given relative to the Ile. For comparison with the solubility and cyclohexane data, the origin for the blocked peptides is shifted on the plot so that the Ile point coincides with the butane point. In terms of the hydrophobicity/area relationships, these three hydrophobic solvents look very similar.

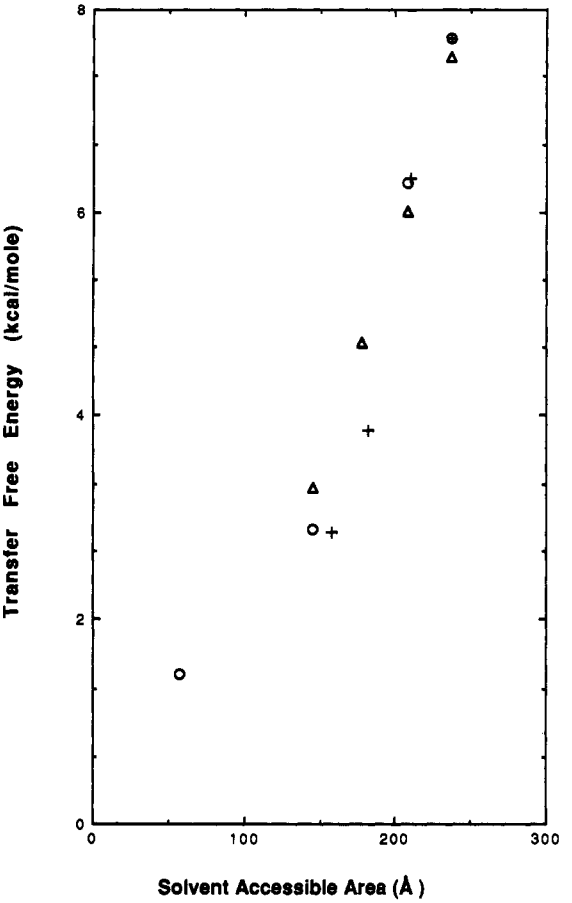

FIGURE 5: Comparison of corrected aliphatic amino acid transfer energies from octanol (+), cyclohexane (O), and alkane (Δ) to water.

Protein Stability Measurements. Table IV compares the solvent partition data with that obtained from three recent series of measurements of the contribution of hydrophobic residues to protein stability (Kellis et al., 1989; Matsumura et al., 1988; Shortle et al., 1990). In these experiments, defined single-residue mutations were made in a protein, and the change in protein unfolding free energy was measured. All the mutations have two features: First the size of the size chain is reduced. Second both wild-type and mutant are hydrophobic residues (the exception is one set of tyrosine mutations). The table gives differences in hydrophobic solvent to water transfer free energies (obtained from Tables II and III) corresponding to these mutations. Each datum from Shortle et al. (1990) represents the averages of several equivalent mutations made

Table IV: Comparison of Mutant Protein Unfolding Energies and Solvent Partition Energies[a]

mutation	change in protein stability (kcal/mol) MBM lysozyme	KNF barnase	SSM staphylococcal nuclease	solvent transfer free energy (kcal/mol) uncorrected o → w	uncorrected c → w	corrected by size o → w	corrected by size c → w
Ile → Val	0.4	0.9–1.4		0.79	0.88	1.36	1.41
Ile → Ala	0.7	3.3–4.2	3.6	2.04	3.11	3.84	4.81
Ile → Gly			4.7	2.46	3.98	4.85	6.23
Leu → Ala			3.1	1.90	3.11	3.87	4.96
Leu → Gly			4.4	2.32	3.98	4.88	6.39
Val → Ala	0.3	2.4–2.8	2.4	1.24	2.23	2.48	3.40
Val → Gly			4.3	1.67	3.10	3.49	4.82
Met → Ala			2.5	1.26	0.54	2.77	1.96
Met → Gly			3.4	1.68	1.41	3.78	3.39
Phe → Ala			3.3	2.02	1.17	4.00	3.03
Phe → Gly			5.2	2.45	2.04	5.02	4.46
Tyr → Ala			2.7	0.89	−1.95	2.87	−0.09
Tyr → Gly			3.4	1.31	−1.08	3.88	1.34

[a] Abbreviations: MBM, data taken from Matsumura et al. (1988); KNF, data taken from Kellis et al. (1989); SSM, data taken from Shortle et al. (1990); o → w, octanol to water partition data taken from Table III; c → w, cyclohexane to water partition data taken from Table II.

Table V: Linear Regression Parameters for Protein Stability Changes vs Solvent Transfer Energies[a]

mutation data compared with	slope	intercept (kcal/mol)	correlation coeff
o → w			
uncorrected	0.46 ± 0.02	0.15 ± 0.1	0.76
corrected by size	0.91 ± 0.04	0.5 ± 0.2	0.86
c → w			
uncorrected	0.64 ± 0.07	0.18 ± 0.3	0.39
corrected by size	1.09 ± 0.1	0.36 ± 0.4	0.65
c → w (no Gly)			
uncorrected	0.82 ± 0.2	−0.25 ± 0.5	0.41
corrected by size	1.32 ± 0.2	−0.22 ± 0.4	0.68

[a] Data taken from Tables II–IV. Symbols: o → w, octanol to water partition data; c → w, cyclohexane to water partition data; no Gly, mutations of the form X → Gly were omitted from the analysis. Errors in the slope and intercept are expressed as one standard deviation from best fit values.

at different sites of the proteins. There is good agreement between these and the data of Kellis et al. (1989), which represent the average of two equivalent sets of mutations. The data of Matsumura et al. (1988), which are taken from a single mutation site, Ile3 in T4 lysozyme, shows significantly smaller effects. Equivalent mutations can vary widely depending on location, as the individual mutation data of Shortle et al. (1990) demonstrate. However, by combining the results of various mutations, local effects tend to be averaged out.

As is evident from Table IV the corrected partition free energies agree quite well with the changes in unfolding free energy obtained from the thermal stability measurements of Shortle et al. (1990) and Kellis et al. (1989). The correlation between the protein stability and octanol solvent transfer data is illustrated in Figure 6 and a linear regression analysis is given in Table V. The data for tyrosine mutations were omitted because this residue is partly polar. For the uncorrected data, both the octanol and cyclohexane transfer energies significantly underestimate the effects of the mutations. This point was noted by both Kellis et al. (1989) and Shortle et al. (1990) in their original analyses. However, for the corrected data the slope in Figure 6 is close to 1. Also, the fit is greatly improved, as indicated by the increase in correlation coefficient (R^2) from 0.76 to 0.86 for octanol and from 0.39 to 0.65 for cyclohexane. The octanol data in particular show a good correspondence with the mutation data, including the tyrosine mutants. The cyclohexane data show more scatter and have the wrong sign for the tyrosine mutants. The cyclohexane partition data are based on amino acid side chain analogues, i.e., they do not include the backbone (Radzicka & Wolfenden, 1988). The glycine analogue is therefore hydrogen, which is a somewhat dubious reference compound, so the data was reanalyzed for just the X → Ala and X → Val mutations (Table V, no Gly). This improves the fit somewhat, but the uncertainty in slope increases, due to the reduced amount of data.

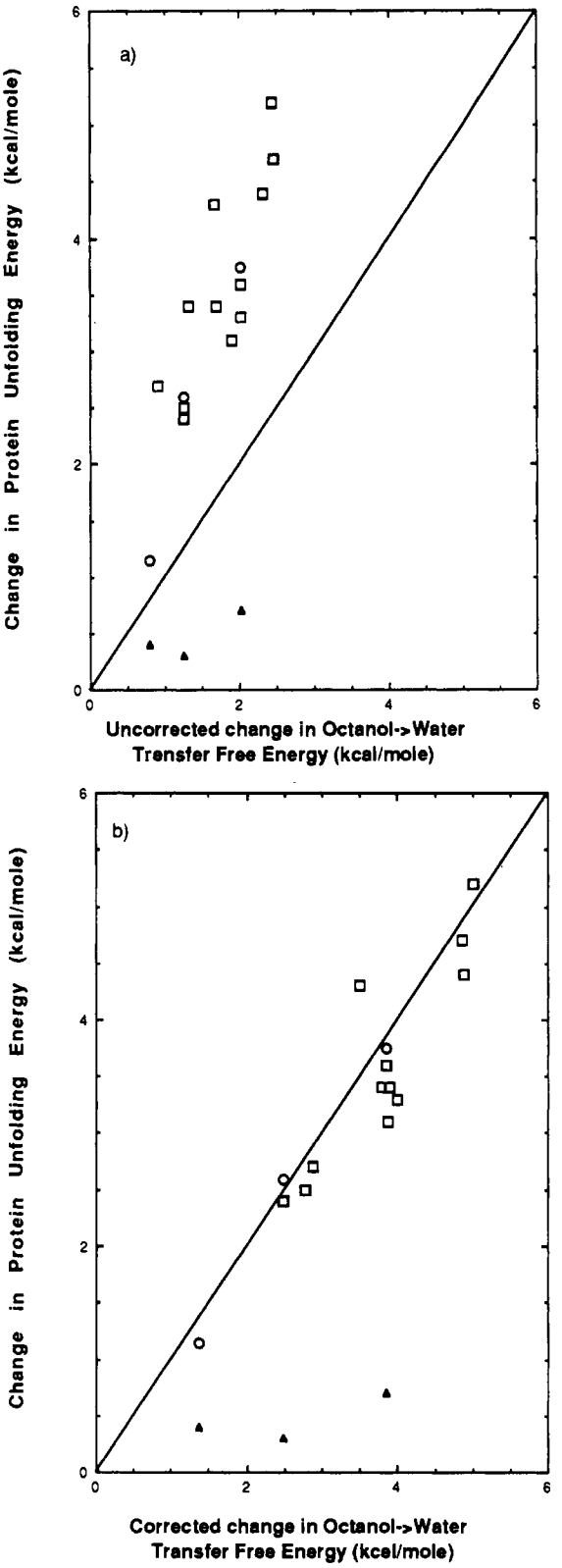

FIGURE 6: Comparison of protein stability ($\Delta\Delta G$ mutation) and nonpolar solvent to water transfer free energy measurements. (▲) Data from Matsumura et al. (1988). (○) Data from Kellis et al. (1989). (□) Data from Shortle et al. (1990). The solvent transfer data are for octanol, with uncorrected values (molarity partition coefficients) (a) and corrected values (b). The line of unit slope is indicated on the figures.

Discussion

Solute Transfer Thermodynamics. In this section we compare our treatment of solute transfer thermodyamics with previous work. We have shown how the transfer process can be split up into contributions due to intermolecular (unitary) interactions and volume entropy terms. The volume entropy term arises from the change in translational free energy upon transfer and so is simply accounted for by using the ideal gas work associated with actual (i.e., experimental) molar volume changes. The volume entropy term thus contains all the concentration-dependent effects (in the limit of dilute gases and solutions). The unitary free energy changes as defined here are changes in the partial molar excess free energies with respect to the ideal gas state. These excess terms arise from intermolecular interactions and include all the solute–solvent

or solvation interactions, such as the hydrophobic interaction. This separation allows one to extract unitary free energies from solubility and partition data and obtain just those quantities of interest that depend on intermolecular interactions.

Our derivation of the volume entropy contribution to solute–solvent mixing parallels that of Hildebrand (1947) but with significant differences. Hildebrand's derivation was aimed at obtaining the entropies of solution and made use of accessible volumes (defined as molar volumes minus the hard-sphere volumes of solute and solvent). In effect it is based on a hard solute model of solutions. In the present work, in contrast, we are not trying to calculate the entropy of mixing per se but to calculate the ideal gas or translational contribution to the transfer process, using the *actual experimental* volumes and concentrations. When subtracted off, this leaves just excess, or intermolecular interaction dependent, terms. In this sense the expression for the volume entropy term used here is exact.

Disagreement as to proper units and standard states has not been uncommon in the area of hydrophobic solute partitioning. As this and previous studies (Ben-Naim, 1978) make clear, the correct concentration scale to use when extracting unitary free energies from partition or solubility data is the molarity scale, not the mole fraction scale. In spite of this, use of mole fraction concentrations is still widespread. More importantly there is an additional volume entropy term, $RT(1 - r)$, that does not arise in Ben-Naim's analysis, which accounts for the solute–solvent size difference. The changes resulting from differences in molar volumes are found to be much larger than those arising from different concentration scales (see Figure 4).

The first direct demonstration that the solute/solvent size difference effect applied to solubility and partition data would lead to a significant increase in the hydrophobic strength was reported by DeYoung and Dill (1990). They studied the partition of benzene between water and a series of *n*-alkane solvents of different chain length. The partition coefficient expressed as either the mole fraction ratio or volume fraction ratio of benzene between the two phases was found to vary as a function of the alkane chain length. Only when the partition coefficient was corrected for the molar volume difference (by using eq 11) was it constant with chain length. Since intuitively one expects the intrinsic hydrophobic interaction to be independent of chain length, this support the idea of a stronger hydrophobic effect. The value they derived of 43 cal/(mol·Å^2) for alkanes is close to that reported in this work. DeYoung and Dill used the Flory–Huggins theory of polymer mixing (Flory, 1941; Huggins, 1941) to obtain unitary transfer free energies from their experimental data. In this theory, the partial molar entropy of mixing of a polymer solute in solution is given by

$$-T\Delta\bar{S}^{mix} = RT \ln \phi_s + RT(1 - r) \qquad (14)$$

which is identical with eq 6.

It is striking that the same expression results from a quite different physical model. This is due to a common assumption regarding random distribution in both the ideal gas and Flory–Huggins theories. In a mixture of ideal gases, the probability of finding a molecule in a small region of space is proportional to its partial pressure, or molar concentration. Similarly in Flory–Huggins theory, the probability of a lattice site being occupied by either a polymer segment or by a solvent molecule is dependent only on the mean fraction of segments or solvent in the mixture, respectively (Flory, 1941). The polymeric nature of the solute per se does not appear in the final expression of Flory–Huggins theory, since only the molar volume of the polymer and the solution volume available to it are important. More significantly, the treatment presented in this work removes the uncertainty involved in applying an expression derived from polymer theory to a general mixing process and, moreover, demonstrates that, at the limit of infinite dilution, the volume correction term is exact.

In the derivation of eqs 3, 7, 11, and 12 for condensation, solubility, partition, and gas to solution transfer, respectively, we have avoided the use of standard states as unnecessary and confusing. Instead, our derivation emphasizes transfer processes; and since we are always dealing with a transfer between two phases, it is quite sufficient to use the same concentration scale, in moles/unit volume, for each phase. Terms such as "infinitely dilute solutions at 1 M concentration", which commonly occur in physical chemistry textbooks, are confusing and have led to errors. For example, the widely used hydrophobicity scale of Kyte and Doolittle (1982) introduced a "standard state" correction to the amino acid side-chain partition data of Wolfenden et al. (1981) and Hine and Mukerjee (1975). This was apparently introduced to account for differences in reference concentrations. However, since the original data are already expressed in ratios of molarity in the vapor and water phases, it is wrong to apply this correction.

From an experimental standpoint, the work of DeYoung and Dill demonstrates that, at least in one important case, ideal gas/Flory–Huggins theories of the volume entropy entropy, as opposed to mole fraction and molarity partition coefficients, provide a physically meaningful prescription for extracting hydrophobic free energies from experimental data. Thus, in the absence of new evidence to the contrary, it appears that when analyzing experimental data for mixtures of small molecules volume entropy changes should be obtained by using molarity partition coefficient plus a term accounting for molar volume differences. We have done this for alkanes and amino acid side-chain partition data and obtained new values for the transfer free energies that differ dramatically from those commonly in use. It is important to emphasize that the raw experimental data are always expressed in terms of concentrations or partial pressures. What are usually termed "experimental values" for free energies are not measured directly but are rather derived from the raw data on the basis of a theoretical model for volume entropy changes. In this work we have used a model that, instead of assuming that the solute and solvent have the same molar volumes, uses the experimental values. This leads to new "experimental free energies".

Hydrophobic Surface Free Energies. Since alkanes are almost completely nonpolar, they provide the most appropriate series of compounds to determine the magnitude of the hydrophobic effect. The generally accepted value, derived from both solubility and partition experiments, is about 25 cal/(mol·Å^2) of solvent accessible area for a purely hydrophobic (i.e., alkane like) surface, although values as low as 16 cal/(mol·Å^2) are reported for carbon atoms in the atomic solubility scale of Eisenberg and McLachlan (1986) and as high as 33 by Hermann (1977). In part, this range of values represents differences in areas, resulting from different atomic radii used in the calculations. More significantly, in some cases the data used to derive transfer free energies are not self-consistent and are presented in different units and with respect to different reference states. Moreover, some of the data are reported for solvents such as methanol and ethanol, which are quite polar and which, even in the case of octanol, may contain a considerable amount of water. Finally, in a number of cases (Guy,

1985; Nozaki & Tanford, 1971), averages of very different partition coefficients for different solvents were used, leading to free energies of uncertain meaning. For example the well-known Nozaki–Tanford hydrophobicity scale contains numbers that are in one case (leucine) the average of quite different partition coefficients in ethanol and dioxane. It can be easily verified that use of ethanol data alone would tend to remove any regularity in the Nozaki–Tanford scale and would not have produced the straight line for free energy vs surface area obtained by Chothia (1976).

Our reanalysis of the original experimental data for the solubilities of *n*-alkanes using the thermodynamically correct units of volume concentrations gives a value for the hydrophobic surface free energy of 28 cal/(mol·Å^2) (Table I). While the data fit a linear relationship, we do not assume, as does Tanford, that the line extrapolates back through the origin. We agree with Hermann (1977) that one should not assume that the energy is proportional to area for very small solutes.

As shown in Table I by using eq 7 to analyze the alkane solubility data yields a value of about 47 cal/(mol·Å^2) for the hydrophobic surface free energy (the limiting value, applicable to an interior methylene group). The data from the partition experiments involving amino acid side chains fall approximately on the same line. It is important to stress that the alkane data, while linear over the size range from 5 to 10 carbons, do not rule out a nonlinear dependence of transfer energy on area for smaller and larger solutes, i.e., a variation in the hydrophobic strength. While the values for propane and butane are similar for transfer from pure solute to water, and from cyclohexane to water, and scale with area, methane is relatively less hydrophobic. In fact, the hydrophobic surface free energy is expected to vary with solute size, increasing with larger solutes, and ultimately reaching a value of $\approx$72 cal/(mol·Å^2) for macroscopic scales (Nicholls et al., 1991; Sharp et al., 1991).

The corrected vapor to water transfer energies for n-alkanes in Table I also are greater, increasing from 5 to 24 cal/(mol·Å^2). The corrected data also make more sense for the series methane through butane. The uncorrected data makes it appear that butane, while much larger than methane, is as easy to transfer to water, while ethane is easier to transfer than methane. This puzzling behavior has previously been attributed to attractive van der Waals interactions, which for these small solutes are postulated to increase more rapidly with size than the unfavorable cavity term. When the size correction is applied, the transfer energies for methane through butane now scale uniformly with size.

Solubility Scales. Partition measurements have been made for different solutes (particularly amino acids and amino acid side chains) in a wide variety of solvents in order to provide insight as to the free energy changes associated with the protein folding process. A review of hydrophobic scales by Nakai et al. (1988) alone mentions several dozen different scales. Actually, the term "hydrophobicity scale" is a misnomer when applied to amino acids since it includes the effects of varying polarity of the side chains. "Solubility scale" is a more appropriate term.

Equations 3 and 11 have been applied to selected amino acid data, with similar conclusions as for alkanes: the energy of transfer from nonpolar media to water is much larger (values are more positive) than previously estimated. Thus nonpolar amino acids are actually more nonpolar, or hydrophobic, than was previously thought, and polar amino acids are less polar. Tables II and III thus provide a new set of data that can be used directly in the analysis of folding experiments, in correlating the burial of amino acids with solubility, and in the derivation of revised solubility scales.

The corrected octanol and cyclohexane solubility scales of amino acids are very similar for the aliphatic amino acids (Figure 5). Overall though, the octanol scale has uniformly positive transfer energies into water, even for polar amino acids, and a smaller difference between polar and nonpolar amino acids than the cyclohexane scale. Both these features can be attributed to the relatively high dielectric constant of octanol (10.3) compared to cyclohexane (2.0) and perhaps to dissolved water in the octanol. Because of this, electrostatic contributions, which would favor the transfer of polar and charged amino acids, will be much smaller for octanol (by close to a factor of 5), leaving only the unfavorable cavity/hydrophobic components. The polar/charged character of amino acids is thus suppressed in the octanol scale.

Simulations. Calculation of solvation or transfer energies by computer simulations of liquids using free energy perturbation or integration methods basically involve calculating the thermodynamic work done on coupling the solute to the surrounding liquid phase (Beveridge & DiCapua, 1989; McQuarrie, 1976; Mezei & Beveridge, 1986). As defined here, the unitary free energies of transfer involve only intermolecular interactions. These simulation methods therefore provide the unitary free energy, or excess chemical potential terms, directly. Since these are defined relative to the ideal gas state, no part of the ideal gas or volume entropy term is included in such simulations. When such simulations are compared with experimental solubilities or partition data, the volume entropy term is standardly obtained by using the mole fraction concentration of solute in solution. This amounts to the assumption that the solute and solvent have the same molar volume.

Since in the free energy perturbation method the solute is "grown" into the solvent during the course of the simulation, it may be argued that the effect of solute–solvent size differences is included in the simulation. It is true that contributions due to the solute–solvent intermolecular *interaction* are included. However, the volume entropy term is not included since no ideal gas work terms enter the simulation directly. This can be seen by the following two "thought" simulations. (a) Two simulations are performed upon the same solute, with N and $2N$ solvent molecules (where N is large enough that cutoff effects are vanishingly small). The difference in volume entropy is $\approx R \ln 2$. If the volume entropy term were included in simulations, there should be a concentration dependence of the result, for any number of solvent molecules. (b) A "simulation" is done on a solute with no interaction potential between it and the solvent (i.e., an "ideal gas solute"), where the solute is introduced into a solution occupying volume V. The simulation is repeated with volume $2V$. Since there is no interaction, or perturbation potential, the result in each case is zero, although the translational or ideal gas work of introducing the solute into the volume is different by $RT \ln 2$. Following this argument, free energy simulations should be compared to solvation energies calculated with eqs 8 and 11, i.e., those that separate out the contribution of the solute–solvent size differences to the volume entropy free energy. Since the size-corrected values in Table II are quite different from the standard values (which appear in the mole fraction column), the extent of agreement between theory and experiment that has been reported in the literature may have to be reevaluated.

Protein Stability Measurements. Consider a protein that upon folding buries an area A of hydrophobic surface, and

assume that the free energy contribution from this process is $\Delta G = -A\gamma_{ow}$, where γ_{ow} is the free energy per unit of surface area of forming a nonpolar solvent or solute–water interface, i.e., this corresponds to what we understand as the "hydrophobic surface free energy" (Nicholls et al., 1991; Sharp et al., 1991). A mutation is made to reduce the surface area of one residue by ΔA. If upon folding the protein interior repacks as well as the wild-type, then the change in folding energy is $\approx\Delta A\gamma_{ow}$. With this assumption about repacking, the measured change in free energy upon folding should be close to the difference in hydrophobic solvation free energies of the wild-type and modified amino acids. In contrast, the original interpretation by Fersht and co-workers of the large effect of mutations on protein stability was based on a model where the mutated protein did not repack, leaving a cavity. In this case, the mutant is destabilized with respect to the wild-type due not only to loss of hydrophobic area but also to less than optimal packing (i.e., the cavity), which would give a larger effect than that predicted on the basis of organic solvent water transfer energies.

On the basis of our reanalysis of solvent transfer thermodynamics, we obtain a substantially larger hydrophobic surface free energy. Consequently, it is not necessary to assume that the mutant proteins contain cavities to explain the large effects of mutations. Instead, we assume that mutant proteins generally repack with comparable efficiency to the wild-type [a more detailed examination of this assumption will appear elsewhere (Nicholls et al., 1991)]. Recent work demonstrates that protein hydrophobic cores show considerable plasticity in accommodating mutations (Lim & Sauer, 1989; Sandberg & Terwilliger, 1991; Wesley Stites, personal communication). We reanalyzed three sets of site-directed mutagenesis data in T4 lysozyme (Matsumura et al., 1988), barnase (Kellis et al., 1989), and staphylococcus nuclease (Shortle et al., 1990). The data include nearly 50 mutations. In each, a selected interior hydrophobic residue was successively mutated to a smaller residue. The data from barnase and staphylococcal nuclease correlate very well with the corrected solubility scales in Tables II and III, with slopes close to one. The fit to the octanol data is particularly good, showing less scatter than with the cyclohexane data. The octanol data are perhaps more appropriate since they are obtained from charge-blocked amino acids, whereas the cyclohexane data are obtained from side-chain analogues, i.e., the C_α and peptide backbone are absent. The octanol transfer energy even agrees well with the tyrosine mutations, the only set that is not purely hydrophobic. This is probably due to the ability of octanol, with its partially polar nature, to mimic the polar interactions that tyrosine must make in the protein.

The question of whether the protein core repacks, and what the energetic cost is, can be settled definitively only by detailed structural and energetic analysis of each mutant. Of the three proteins, structural data were provided only for the T4 lysozyme mutants, which for the I3V and I3A structures were essentially unchanged, i.e., no repacking occurred (Matsumura et al., 1988). In this case one would expect a *larger* decrease in stability. In fact, the data set, while small, shows a substantially smaller effect than expected. This suggests the possibility that the relevant residues are not completely buried in the folded state. Indeed, in the crystal structure Ile3, although apparently mostly buried (≈80%), is close to the surface (Matsumura et al., 1988).

On the basis of the data analyzed to date, the molar volume corrected octanol solubility scale provides the best agreement with changes in protein stability upon mutation. Thus, to a first approximation, the effect of mutating one nonpolar residue to a smaller nonpolar residue on protein stabilization is due to the change in hydrophobicity, as measured by organic solvent to water transfer. Superimposed on this, variations are expected to arise from variations in packing and from local interactions. This is evident in the data from Shortle et al. (1990) on equivalent mutations in different sites, which show standard deviations in the change in stabilization energy of 1–2 kcal/mol. In this case a refinement based on their analysis of local packing density or α-carbon density may improve the accuracy of the fit. An alternative separation of the energetics of these mutations, which we did not pursue here, is that upon folding, first the interactions of the side chain with water are removed, then the interactions with the interior of the protein are made. The first step is independent of the environment of the amino acid side chain and can be obtained from differences in our corrected vacuum to water transfer energies for the corresponding mutations (Table II). The second step then contains all nonpolar interactions plus variations in packing and local environment effects.

The correspondence between solvent transfer energies and protein packing energies appears puzzling from one aspect. The side chains are expected to be more constrained in the protein interior than a solute in a nonpolar solvent. Thus there should be an unfavorable conformational entropy term in the free energy of folding that is not present in solubility experiments. That its effect is not seen may be due to compensating effects from greater packing density in the protein interior than in a nonpolar solvent (Nicholls et al., 1991). It is also possible that the conformational entropy difference between mutated residues is small and lies within the uncertainty of the measurements and the variations in packing interactions. These issues are currently being examined.

ACKNOWLEDGMENTS

We thank Ken Dill for bringing his work to our attention and Mihaly Mezei and Rick Fine for stimulating discussions.

REFERENCES

Atkins, P. W. (1986) in *Physical Chemistry*, W. H. Freeman, New York.
Aveyard, R., & Haydon, D. A. (1965) *J. Colloid Interface Sci. 20*, 2255–2261.
Ben-Naim, A. (1978) *J. Phys. Chem. 82*, 792–803.
Ben-Naim, A. (1987) *Am. J. Phys. 55*, 725–733.
Ben-Naim, A., & Marcus, Y. (1984) *J. Chem. Phys. 81*, 2016–2027.
Beveridge, D., & DiCapua, F. M. (1989) *Annu. Rev. Biophys. Biophys. Chem. 18*, 431.
Bondi, A. (1968) *Molecular Crystals, Liquids and Glasses*, John Wiley and Sons, New York.
Chothia, C. (1976) *J. Mol. Biol. 105*, 1.
DeYoung, L. R., & Dill, K. A. (1990) *J. Phys. Chem. 94*, 801–809.
Dill, K. A. (1990) *Biochemistry 29*, 7133.
Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199.
Fauchere, J. L., & Pliska, V. (1983) *Eur. J. Med. Chem. 18*, 369–375.
Flory, P. J. (1941) *J. Chem. Phys. 9*, 660–671.
Friedman, M. E., & Scheraga, H. A. (1965) *J. Phys. Chem. 69*, 3795–3800.
Gilson, M., Sharp, K. A., & Honig, B. (1988) *J. Comput. Chem. 9*, 327–335.
Gurney, R. W. (1953) in *Ionic Processes in Solution*, McGraw-Hill, New York.
Guy, H. R. (1985) *Biophys. J. 47*, 61.

Hermann, R. B. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4144.
Hildebrand, J. H. (1947) *J. Chem. Phys. 15*, 225-228.
Hine, J., & Mookerjee, P. K. (1975) *J. Org. Chem. 40*, 292.
Huggins, M. L. (1941) *J. Chem. Phys. 9*, 440-449.
Kellis, J. T., Nyberg, K., & Fersht, A. R. (1989) *Biochemistry 28*, 4914-4922.
Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379-400.
Lee, B. K. (1985) *Biopolymers 24*, 813-823.
Lide, D. R., Ed. (1990) *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL.
Lim, W. A., & Sauer, R. T. (1989) *Nature 339*, 31-36.
Masterton, W. L. (1954) *J. Phys. Chem. 22*, 1830-1833.
Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406.
McAuliffe, C. (1966) *J. Phys. Chem. 70*, 1267.
McQuarrie, D. (1976) *Statistical Mechanics*, Harper & Row, New York.
Mezei, M., & Beveridge, D. L. (1986) *Computer Simulations and Biomolecular Systems, Ann. N.Y. Acad. Sci. 494*, 1-23.
Nakai, K., Kidera, A., & Kanehisa, M. (1988) *Protein Eng. 2*, 93.
Nicholls, A., Sharp, K. A., & Honig, H. (1991) *Proteins* (in press).
Nozaki, Y., & Tanford, C. H. (1971) *J. Biol. Chem. 246*, 2211.
Radzicka, A., & Wolfenden, R. (1988) *Biochemistry 27*, 1664-1670.
Richards, F. M. (1974) *J. Mol. Biol. 82*, 1-14.
Sandberg, W. S., & Terwilliger, T. C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1706-1710.
Sharp, K. A., & Nicholls, A. (1990) DELPHI 3.0, available from the Department of Biochemistry and Molecular Biophysics, Columbia University, New York, or from Biosym Corp., San Diego, CA.
Sharp, K. A., Nicholls, A., Fine, R. M., & Honig, B. (1991) *Science 252*, 106-109.
Shortle, D., Stites, W. E., & Meeker, A. K. (1990) *Biochemistry 29*, 8033-8041.
Tanford, C. H. (1980) *The Hydrophobic Effect*, John Wiley and Sons, New York.
Wolfenden, R., Andersson, L., Cullis, P. M., & Southgate, C. C. (1981) *Biochemistry 20*, 849-855.
Wood, R. H., & Thompson, P. T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 946-949.

Primary Structure of Hydrogenase I from *Clostridium pasteurianum*†

Jacques Meyer*,‡ and Jean Gagnon§
DBMS-Métalloprotéines and DBMS-Biologie Structurale, CNRS URA 1333, CENG 85X, 38041 Grenoble, France
Received May 2, 1991; Revised Manuscript Received July 1, 1991

ABSTRACT: Peptides obtained by cleavage of *Clostridium pasteurianum* hydrogenase I have been sequenced. The data allowed design of oligonucleotide probes which were used to clone a 2310-bp *Sau*3A fragment containing the hydrogenase encoding gene. The latter has been sequenced and was found to translate into a protein composed of 574 amino acids (M_r = 63 836), including 22 cysteines. *C. pasteurianum* hydrogenase is homologous to, but longer than, the large subunit of *Desulfovibrio vulgaris* (Hildenborough) [Fe] hydrogenase. It includes an additional N-terminal domain of ca. 110 amino acids which contains eight cysteine residues and which therefore could accommodate two of its postulated four [4Fe-4S] clusters. *C. pasteurianum* hydrogenase is most similar in length, cysteine positions, and sequence altogether to the translation product of a putative hydrogenase encoding gene from *D. vulgaris* (Hildenborough). Comparisons of the available [Fe] hydrogenase sequences show that these enzymes constitute a structurally rather homogeneous family. While they differ in the length of their N-termini and in the number of their [4Fe-4S] clusters, they are highly similar in their C-terminal halves, which are postulated to harbor the hydrogen-activating H cluster. Five conserved cysteine residues occurring in this domain are likely ligands of the H cluster. Possible ligation by other residues, and in particular by methionine, is discussed. The comparisons carried out here show that the H clusters most probably possess a common structural framework in all [Fe] hydrogenases. On the basis of the available data on these proteins and on the current developments in iron–sulfur chemistry, the H clusters possibly contain six to eight iron atoms. The large differences between the DNA compositions of *C. pasteurianum* (30% G + C) and *D. vulgaris* (65% G + C) result in significant differences, not only in codon usage for a given amino acid but also in the amino acid compositions of their respective hydrogenases.

Hydrogenases are iron–sulfur enzymes that catalyze the reaction $H_2 \leftrightarrow 2H^+ + 2e^-$. They are divided into two main groups, on the basis of metal content and sequence homology: those containing only iron ([Fe] hydrogenases) (Adams, 1990) and those containing iron and nickel ([NiFe] hydrogenases) (Fauque et al., 1988). A few members of the second group also contain selenium (Fauque et al., 1988). The widespread distribution of [NiFe] hydrogenases in the microbial world is reflected in the significant number of species from which hydrogenase genes have been cloned and sequenced (Voordouw et al., 1989a; Leclerc et al., 1988; Sayavedra-Soto et al., 1988; Reeve et al., 1989; Menon et al., 1990a,b; Uffen et al., 1990; Tran-Betcke et al., 1990; Alex et al., 1990; Rousset et al., 1990;

† The nucleotide sequence reported in this paper has been submitted to GenBank under Accession Number M62754.
‡ DBMS-Métalloprotéines.
§ DBMS-Biologie Structurale.

0006-2960/91/0430-9697$02.50/0