Amino Acid Side-chain Hydrophobicity

Hue Sun Chan, *University of Toronto, Toronto, Canada*

Secondary article

Hydrophobicity is the unfavourable energetics of dissolving nonpolar compounds in water. The hydrophobicities of the 20 amino acid side-chains are currently described by hydrophobicity scales derived primarily from solubility studies; these scales have provided semiquantitative rationalizations of some properties of native (folded) proteins.

Introduction

Hydrophobicity is a term used to describe the unfavourable energetics of dissolving nonpolar (e.g. carbon) compounds in water: while nonpolar compounds readily dissolve in many nonpolar solvents, they have very low solubilities in water. Another defining characteristic of hydrophobic compounds is their tendency to self-associate in aqueous solutions. A generally accepted explanation of this phenomenon is that it results largely from the more favourable interactions among the water molecules achieved when the nonpolar groups cluster together, rather than from a direct van der Waals attraction between nonpolar groups.

What is the physical origin of hydrophobicity? There are extensive hydrogen-bonding interactions among water molecules (H_2O) in liquid water. Hydrophobic groups, being nonpolar, cannot form hydrogen bonds with the water molecules in their vicinity. However, it would be energetically costly if these water molecules are not hydrogen bonded; therefore, in the presence of a nonpolar solute, water molecules are driven to either adopt more specific orientations to avoid losing hydrogen bonds, or to minimize the loss, depending on which alternative is allowed by the geometry of the hydrophobic surface (**Figure 1**). The presence of special water configurations near a hydrophobic surface, with average properties distinguishable from that of bulk water, is often referred to as 'hydrophobic hydration'. The statistical mechanical consequences are either a restriction of water configurational freedom (entropy) or a loss of favourable (enthalpic) hydrogen bonds, or both. These effects translate into an unfavourable free energy change associated with exposing hydrophobic groups to water (**Figure 1**).

Liquid state configurations are constantly fluctuating, not static. **Figure 1** illustrates only the dominant configurations; it is possible to have other water configurations near hydrophobic surfaces. However, the dominant configurations in **Figure 1** show how unfavourable free energy changes are caused by hydrophobic surfaces' disruption of water structure relative to that in the bulk. Consistent with this molecular picture, it has been empirically observed that the magnitude of hydrophobic effect, as measured by transfer free energies (see below), is approximately proportional to the water-accessible surface area (Lee and Richards, 1971) of the hydrophobic groups (Tanford, 1980). In this perspective, the association of hydrophobic groups in water is viewed as driven by the need to reduce energetically unfavourable exposure of nonpolar surface to water. Typically, transferring hydrophobic groups from water to a nonpolar environment is believed to be favoured by approximately 105–142 J mol^{-1} (25–34 cal mol^{-1}) per square angstrom of water-accessible area of the hydrophobic group (reviewed by Chan and Dill, 1997). (In the biophysics literature, water-accessible surface area is often called 'solvent-accessible surface area' (SASA) without specifying explicitly that water is the solvent in question.)

Experimentally Derived Scales from Free Energies of Transfer

It is a common practice to use experimentally determined free energies of transfer of various solutes, from nonpolar solvents to water, to quantify the extent of hydrophobicity, i.e. the degree to which each solute 'dislikes' water, relative to the nonpolar phase (**Figure 2**). In these measurements, a solute is partitioned between two different solvents according to its solubility in each. The experimental observable is the equilibrium concentration of the solute in each phase. Transfer free energies are determined from the measured concentrations, and contact energies are deduced using thermodynamic arguments and statistical mechanical models (reviewed by Chan and Dill, 1997).

Because the process of folding a protein involves removing a large fraction of amino acid surface from water to the protein interior, it is on some level analogous to transferring a model compound from water to a nonpolar phase. Consequently, many researchers have obtained experimental transfer data for amino acids, their derivatives and other model compounds with the goal of using such data to understand the origins of protein stability. While the analogy to water phase is direct,

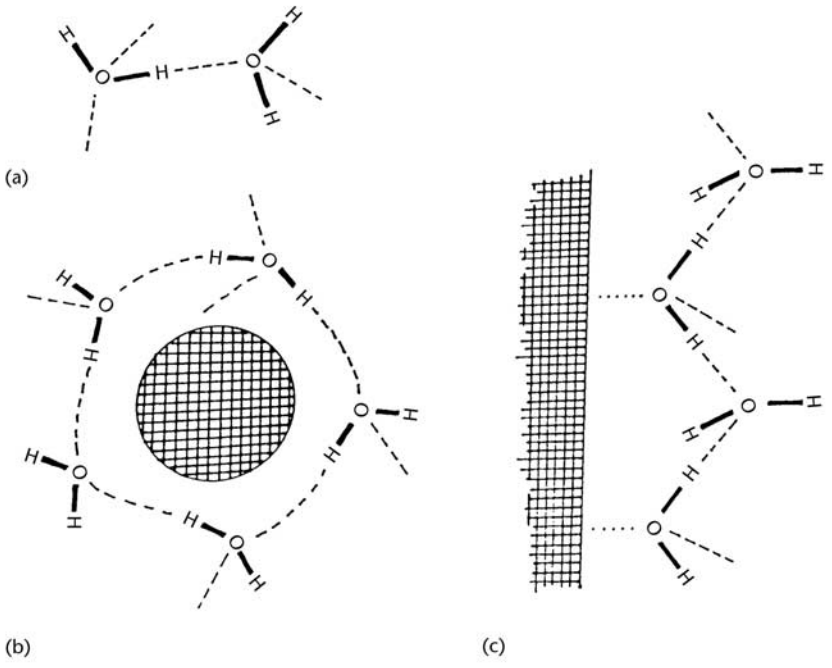

Figure 1 Molecular origins of hydrophobicity. Typical hydrogen bonding (dashed lines) pattern among water molecules H_2O: (a) in the bulk; (b) around a small nonpolar solute (shaded circle); and (c) near an extended nonpolar surface. The hydrogen bonding geometry in (b) is distorted relative to that in (a) to maintain an interaction strength among water molecules comparable to that in the bulk. Water molecules around a small solute with a convex nonpolar surface are oriented to avoid directing their hydrogen-bonding groups (donor or acceptor) towards the solute. This arrangement is not possible near a flat extended nonpolar surface. In this case there are 'dangling' hydrogen bonds, i.e. potentially hydrogen-bonding groups (dotted lines) oriented towards the nonpolar surface (Lee *et al.*, 1984).

discerning which nonpolar phase best models a protein interior has been less obvious. As a result, a variety of different nonpolar phases have been employed. These include octanol, linear alkanes, cyclohexane, bilayer, grafted alkyl phase on chromatographic columns, among others.

Hydrophobicity scales have been determined from free energies of transfer of amino acids or their chemical derivatives from water to various nonpolar phases (for example, an organic solvent, see **Figure 2**). In analysing transfer data, it is often assumed that contributions to the transfer free energy from different parts (e.g. a methyl or amide group) of a molecule are additive. When group additivity is assumed, contributions to the free energy of transfer of individual amino acids can be calculated from differences in transfer free energies of various amino acid derivatives. By the same token, some scales give hydrophobicities of individual amino acids in terms of whether they are more hydrophobic or less hydrophobic than glycine, the smallest amino acid that does not have a side-chain (**Table 1**). The assumption of group additivity is quite reasonable, and is often useful. None the less, it should be noted that, ultimately, group additivity and treatments based solely on water-accessible surface area are only approximations. Calculations at the atomic level (Lee *et al.*, 1984) show that the underlying molecular basis of hydrophobicity is more intricate than these simplified descriptions.

Dependence of Experimental Scale on Phase, Solvent

Given the central importance of understanding what determines protein stability and structure, numerous amino acid hydrophobicity scales have been obtained by many research groups during the last few decades. Representative scales are given in **Table 1**. Information regarding other scales and recent reviews is provided in **Table 2**. In general, there has been a lack of quantitative agreement between hydrophobicity scales. **Table 1** demonstrates that different nonpolar phases and different techniques give rise to different amino acid transfer free energies. **Figure 3a** provides a visual comparison of the level of quantitative agreement between scales.

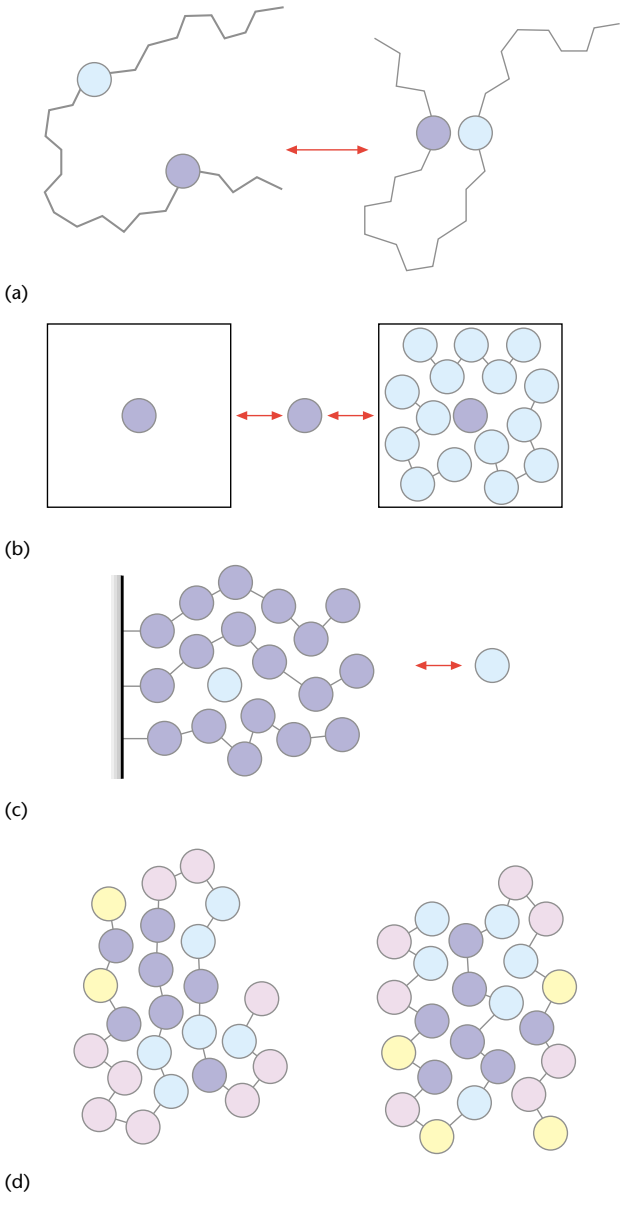

Figure 2 Experimental and statistical procedures for estimating amino acid interaction parameters. (a) Formation of a contact when a biomolecule undergoes conformational changes, as in protein folding. The aqueous solvent is not depicted explicitly. (b) Modelling contact formation by transferring a model compound (small solute) from an aqueous phase (left) to a nonpolar phase (right). Hydrophobic hydration is also studied by transferring small solutes from a gaseous phase (middle) to water. (c) Modelling contact formation by studying the partitioning of solutes into aligned nonpolar phases such as bilayers and in reversed-phase liquid chromatography experiments. (d) Some interaction parameters are deduced from the statistics of contacts among different amino acid types in the database of protein native structures.

The charged and polar (hydrogen-bonding) side-chains are particularly sensitive to the nonpolar phase, and their transfer free energies show large deviation from the expected proportionality to water-accessible surface area. This may be caused by a number of factors. For instance, octanol, one of the commonly used nonpolar phase solvents, is somewhat polar and contains a significant amount of dissolved water, whereas other nonpolar phase solvents, such as liquid alkanes, have very little dissolved water. On the other hand, the hydrophobicities of the hydrophobic/nonpolar amino acid side-chains (aliphatic nonhydrogen bonding, aromatic nonhydrogen bonding and sulfur-containing) show less variation among different scales (Karplus, 1997). The temperature dependence of transfer free energies can also vary significantly. In some experiments involving partially aligned alkyl chains, such as reverse-phase liquid chromatography stationary phases (**Figure 2c**), nonpolar solutes are driven mainly by enthalpy into the partially aligned alkyl phase, instead of by entropy as in transfer between water and bulk nonpolar phases. This observation is sometimes called 'nonclassical' hydrophobic effect or 'bilayer effect' (discussed in DeVido *et al.* (1998) and White and Wimley (1999); **Table 2**).

Given the variability between different hydrophobicity scales, it is perhaps not surprising that the rank ordering of amino acid hydrophobicities shows marked variation from scale to scale. For example, in the 16 scales tabulated by Wilce *et al.* (1995), phenylalanine is ranked first (most hydrophobic) by four of the scales, yet one scale ranks it as 16th, i.e. close to being the least hydrophobic (20th). In **Table 1**, tryptophan is ranked first by two scales (b and c), whereas it is the third most hydrophobic according to scale (a). These observations highlight the limitations of amino acid hydrophobicity scales in quantitative applications. Unfortunately, no one set of 20 numbers (i.e. a single generic hydrophobicity scale) exists that is capable of accurately predicting protein stability and/or ligand-binding energetics.

Theoretical Scales Derived from Protein Structures

As discussed above, most hydrophobicity scales use a nonpolar phase to model a generic protein core. As such, they cannot account for the specific interactions between amino acid side-chains in protein interiors.

In principle, these interactions could be studied at a more fundamental level by using potentials for each atom. However, for many applications, an amino acid-based approach is still preferred, because a calculation involving the pairwise interactions between the thousands of atoms in a given protein is often not tractable with currently available computational power. A logical next step, therefore, is to determine a set of 210 pairwise amino acid contact energies.

To date, no systematic, direct experimental measurements of all pairwise amino acid contact energies exists.

Table 1 Free energies of transfer of the amino acids from water to nonpolar environments, relative to that of glycine

		Free energy transfer[a] (kJ mol^{-1})		
Amino acid	Code	(a)	(b)	(c)
Phenylalanine	F	–12.3	–10.1	–4.79
Isoleucine	I	–11.4	–10.2	–1.34
Tryptophan	W	–11.3	–12.8	–7.81
Leucine	L	–8.95	–9.64	–2.39
Tyrosine	Y	–8.70	–5.44	–3.99
Methionine	M	–8.23	–6.97	–1.01
Valine	V	–7.80	–6.91	+0.25
Proline	P	–7.15	–4.08	+1.85
Cysteine	C	–4.05	–8.73	–1.05
Glutamate	E	–3.02	+3.63	+8.44
Alanine	A	–2.67	–1.76	+0.67
Threonine	T	–2.50	–1.47	+0.55
Glutamine	Q	–1.95	+1.25	+2.39
Aspartate	D	–1.03	+4.36	+5.12
Glycine	G	0.00	0.00	0.00
Serine	S	+0.42	+0.23	+0.50
Asparagine	N	+1.25	+3.40	+1.72
Arginine	R	+1.75	+5.72	+3.36
Lysine	K	+3.00	+5.61	+4.12
Histidine	H	+4.22	–0.74	+3.99

[a]These hydrophobicity scales were obtained by using different chemical derivatives of the amino acids and the following different nonpolar phases: (a) alkyl chains in a reversed-phase liquid chromatography stationary phase (DeVido *et al*., 1998), (b) bulk octanol (Fauchère and Pliška, 1983), and (c) large unilamellar vesicle membranes (Wimley and White, 1996; with ionized side-chains E, D, R, K and H).

Instead, these energies have been obtained by knowledge-based or statistical methods from databases of protein native structures. (Hence they are called statistical potentials.) There are a number of slightly different approaches, but they all share the same basic assumption that the statistical distribution of contacts among native structures of proteins can be related to the underlying contact energies by some very simple mathematical relations (**Figure 2d**). Tanaka and Scheraga first advanced this idea in 1976. A comprehensive analysis was provided by Miyazawa and Jernigan in 1985 (Miyazawa and Jernigan, 1996), and subsequently developed by many researchers.

The formulation of Miyazawa and Jernigan (1996) is a good illustration of the statistical potential method. They define the energy for a contact between amino acid residue types i and j by $e_{ij} = -RT \ln [(\bar{n}_{ij}\bar{n}_{00})/(\bar{n}_{i0}\bar{n}_{j0})]$, where $\bar{n}$s are average numbers of contacts observed in a given database of native structures. The subscript 0 denotes solvents, 00 and $i0$ represent a solvent–solvent contact and a contact between residue type i and solvent (when i is exposed to solvent), respectively. By this simple formula, more favourable (lower) contact energies e_{ij} are assigned to pairs of amino acid side-chains that occur more frequently together in spatial contact (i.e. larger $\bar{n}_{ij}$) in the given protein native structure database.

This and similar approaches assume that the contacts in the protein native structure database follow a Boltzmann distribution at temperature T. In reality, however, the set of native structures is not a Boltzmann ensemble. Native structures of different proteins are not in thermal equilibrium because they cannot interconvert into one another. Nevertheless, in test studies using simple model systems, these simple statistical procedures have been shown to be reasonably accurate in extracting contact energies, provided that certain special conditions are satisfied. It should also be pointed out, however, that these procedures have been shown to be not generally valid. A rigorous connection has yet to be established between statistical potentials and the true physical interactions.

Comparison Between Prediction and Experiment

What can be learned about protein structure and stability from hydrophobicity scales and pairwise amino acid statistical potentials? At a qualitative or semiquantitative level, predictions by some scales are consistent with protein experiments. For example, amino acid residues that are deemed more hydrophobic in transfer experiments are more likely to be buried in native structures of globular proteins. In membrane proteins, contiguous stretches of hydrophobic residues are useful for identifying membrane-spanning segments. More quantitatively, **Figure 3b** shows that knowledge-based statistical potentials correlate well with hydrophobicities of individual amino acids determined by nonpolar/water transfers. The statistical potentials are derived from experimentally determined protein native structures. Hence, the correlation in **Figure 3b** indicates that some scales derived from model compound transfer experiments are predictive, in the sense that they are able to provide a reasonable account of the structural organization of native proteins (see also Rose *et al.* (1985) and Eisenberg and McLachlan (1986); **Table 2**).

Amino acid hydrophobicity scales have been used to rationalize changes in protein stability caused by changes in amino acid sequence (mutations); however, the results are not straightforward to interpret. **Figure 3c** shows that there are two sources of uncertainties in relating experimental mutagenesis data to hydrophobicity scales: (1) variation between different hydrophobicity scales, as noted above; and (2) variation in folding free energy changes caused by mutations between the same two amino acid types, which turn out to depend on the protein and the particular location of the mutation site. This observation implies that the protein environment of a given amino acid residue can strongly affect its interactions. Twenty-parameter hydrophobicity scales cannot accurately predict mutational effects on native stability because these scales effectively assume a single uniform generic protein core environment.

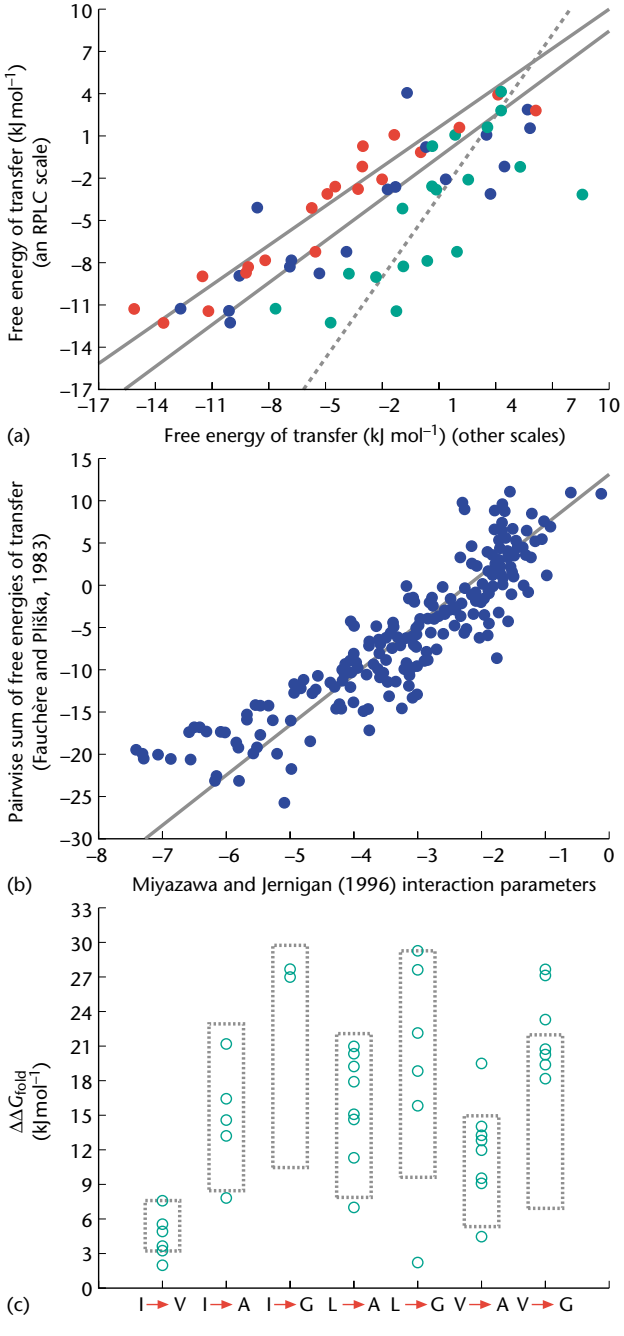

Figure 3 Hydrophobicity scales obtained from different techniques and their applications to protein folding. (a) Correlation between the reversed-phase liquid chromatography (RPLC) scale in Table 1(a) and one of the Wimley, Creamer and White (1996) scales (red dots), the scale of Fauchère and Pliška (1983) (blue dots, Table 1 (b)), and that of Wimley and White (1996) (green dots, Table 1 (c)). Least square fits are given by the upper and lower solid lines and the dashed line, with correlation coefficient $r = 0.96$, 0.87 and 0.72, respectively; see DeVido *et al.* (1998) in Table 2 for details. (b) A set of 210 interaction parameters between pairs of amino acids determined statistically from a database of protein native structures (Miyazawa and Jernigan, 1996) are plotted against pairwise sums of hydrophobicities (Fauchère and Pliška, 1983) of the corresponding amino acids; solid line is the least square fit, $r = 0.90$. (c) $\Delta\Delta G_{fold}$ is the folding free-energy change caused by the type of single-site mutation given below the horizontal axis. A larger $\Delta\Delta G_{fold}$ means that the mutation results in a less stable native structure. A total of 48 different experimental values of $\Delta\Delta G_{fold}$ are plotted (open circles). The same mutation can produce very different changes in folding free energy in different proteins or at different sites of the same protein. The ranges of corresponding free-energy changes predicted by small model compound results and the analysis of Lee are indicated by the dashed boxes. Part (c) of this figure is adapted from Lee (1993); more details are given in this reference.

Table 2 References for representative hydrophobicity scales obtained by different experimental methods

Experimental method	Reference
Transfer experiments	Nozaki and Tanford (1971) reviewed in Tanford (1980) (Table 13-1, p. 140)
	Fauchère and Pliška (1983) (see **Table 1**)
	Kim and Szoka (1992)
	Wimley, Creamer and White (1996)
Comparing protein structure data with transfer experiments	Rose *et al.* (1985)
	Eisenberg and McLachlan (1986)
Mutation studies	Yutani *et al.* (1987)
	Eriksson *et al.* (1992)
Bilayer studies	Wimley and White (1996) (see **Table 1**)
	Thorgeirsson *et al.* (1996)
	White and Wimley (1999)
Chromatography	Meek and Rosetti (1981)
	Guo *et al.* (1986)
	Wilce *et al.* (1995)
	DeVido *et al.* (1998) (see **Table 1**)

More comprehensive lists of hydrophobicity scales can be found in Wilce *et al.*, 1995 (16 scales are compared) and DeVido *et al.*, 1998 (more than 40 scales are considered and cited in Tables 3 and 4 of this reference).

Statistical potentials have been used in fold identification for predicting protein structure from sequence (Sippl, 1995). Given an amino acid sequence, a fold identification technique attempts to recognize the native structure among alternate folds (decoys). Typically, a total energy or score is computed as a sum of pairwise contact energies for each of the conformations to be considered. The technique is successful if the native conformation has an energy lower (more favourable) than all the decoys. When these procedures are tested using protein sequences with known native structures, they often fail when the decoy set used to challenge the identification technique contains a wide range of compact conformations. These failures suggest that some essential aspects of protein interactions are missing in current pairwise contact energies.

In summary, while low-resolution amino acid-based contact energies and hydrophobicity scales have provided insight into protein energetics, their utility in quantitative predictions has proved limited. Advances in both experimental and theoretical treatments beyond these empirical approaches are needed to provide more accurate accounts of the interactions among amino acid side-chains, and hence a more detailed understanding of protein stability and structure.

References

Chan HS and Dill KA (1997) Solvation: how to obtain microscopic energies from partitioning and solvation experiments. *Annual Review of Biophysics and Biomolecular Structure* **26**: 425–459.

DeVido DR, Dorsey JG, Chan HS and Dill KA (1998) Oil/water partitioning has a different thermodynamic signature when the oil solvent chains are aligned than when they are amorphous. *Journal of Physical Chemistry B* **102**: 7272–7279.

Eisenberg D and McLachlan AD (1986) Solvation energy in protein folding and binding. *Nature* **319**: 199–203.

Eriksson AE, Baase WA, Zhang X-J *et al.* (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**: 178–183.

Fauchère J-L and Pliška V (1983) Hydrophobic parameters Π of amino-acid side chains from the partitioning of *N*-acetyl-amino-acid amides. *European Journal of Medicinal Chemistry – Chimie Thérapeutique* **18**: 369–375.

Guo D, Mant CT, Taneja AK *et al.* (1986) Prediction of peptide retention times in reversed-phase high-performance liquid chromatography. I. Determination of retention coefficients of amino acid residues of model synthetic peptides. *Journal of Chromatography* **359**: 499–517.

Karplus PA (1997) Hydrophobicity regained. *Protein Science* **6**: 1302–1307.

Kim A and Szoka FC Jr (1992) Amino acid side-chain contributions to free energy of transfer of tripeptides from water to octanol. *Pharmaceutical Research* **9**: 504–514.

Lee B (1993) Estimation of the maximum change in stability of globular proteins upon mutation of a hydrophobic residue to another of smaller size. *Protein Science* **2**: 733–738.

Lee B and Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology* **55**: 379–400.

Lee CY, McCammon JA and Rossky PJ (1984) The structure of liquid water at an extended hydrophobic surface. *Journal of Chemical Physics* **80**: 4448–4455.

Meek JL and Rosetti ZL (1981) Factors affecting retention and resolution of peptides in high-performance liquid chromatography. *Journal of Chromatography* **211**: 15–28.

Miyazawa S and Jernigan RL (1996) Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *Journal of Molecular Biology* **256**: 623–644.

Rose GD, Geselowitz AR, Lesser GJ *et al*. (1985) Hydrophobicity of amino acid residues in globular proteins. *Science* **229**: 834–838.

Sippl MJ (1995) Knowledge-based potentials for proteins. *Current Opinion in Structural Biology* **5**: 229–235.

Tanford C (1980) *The Hydrophobic Effect: Formation of Micelles and Biological Membrane*s, 2nd edn. New York: Wiley.

Thorgeirsson TE, Russell CJ, King DS and Shin YK (1996) Direct determination of the membrane affinities of individual amino acids. *Biochemistry* **35**: 1803–1809.

White SH and Wimley WC (1999) Membrane protein folding and stability: physical principles. *Annual Review of Biophysics and Biomolecular Structure* **28**: 319–365.

Wilce MCJ, Aguilar M-I and Hearn MTW (1995) Physiochemical basis of amino acid hydrophobicity scales: evaluation of four new scales of amino acid hydrophobicity coefficients derived from RP-HPLC of peptides. *Analytical Chemistry* **67**: 1210–1219.

Wimley WC, Creamer TP and White SH (1996) Solvation energies of amino acid side chains and backbone in a family of host–guest pentapeptides. *Biochemistry* **35**: 5109–5124.

Wimley WC and White SH (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology* **3**: 842–848.

Yutani K, Ogasahara K, Tsujita T and Sugino Y (1987) Dependence of conformational stability on hydrophobicity of the amino acid residue in a series of variant proteins substituted at a unique position of tryptophan synthase α subunit. *Proceedings of the National Academy of Sciences of the USA* **84**: 4441–4444.

Further Reading

Cheng Y-K and Rossky PJ (1998) Surface topography dependence of biomolecular hydrophobic hydration. *Nature* **392**: 696–699.

Dill KA (1990) Dominant forces in protein folding. *Biochemistry* **29**: 7133–7155.

Hummer G, Garde S, Garcia AE, Paulaitis ME and Pratt LR (1998) Hydrophobic effects on a molecular scale. *Journal of Physical Chemistry B* **102**: 10469–10482.

Kyte J and Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**: 105–132.

Liu L-P and Deber CM (1998) Uncoupling hydrophobicity and helicity in transmembrane segments: α-helical propensities of the amino acids in non-polar environments. *Journal of Biological Chemistry* **273**: 23645–23648.

Park BH, Huang ES and Levitt M (1997) Factors affecting the ability of energy functions to discriminate correct from incorrect folds. *Journal of Molecular Biology* **266**: 831–846.

Privalov PL and Gill SJ (1988) Stability of protein structure and hydrophobic interaction. *Advances in Protein Chemistry* **39**: 191–234.

Robertson AD and Murphy KP (1997) Protein structure and the energetics of protein stability. *Chemical Reviews* **97**: 1251–1267.

Scheraga HA (1998) Theory of hydrophobic interactions. *Journal of Biomolecular Structure and Dynamics* **16**: 447–460.

Skolnick J, Jaroszewski L, Kolinski A and Godzik A (1997) Derivation and testing of pair potentials for protein folding. When is the quasichemical approximation correct? *Protein Science* **6**: 676–688.