Hydrophobic Interactions in Proteins

Brian W Matthews, *University of Oregon, Eugene, Oregon, USA*

Proteins fold spontaneously into complicated three-dimensional structures that are essential for biological activity. Much of the driving energy for this folding process comes from the hydrophobic effect, i.e. the removal of nonpolar amino acids from solvent and their burial in the core of the protein.

Introduction

The hydrophobic (literally, 'water-hating') effect is named for the tendency of certain oil-like substances to avoid contact with water (proverbially, 'oil and water don't mix'). It is generally understood to be the driving force responsible for the folding of proteins.

Proteins are synthesized in the cell as polymers made up of linked units – amino acids. In nature there are 20 possible amino acids that are used for this purpose (**Table 1**). Any given protein is characterized by the number of amino acids that it contains, together with the sequence in which the amino acids are arranged. It is the sequence of amino acids that determines the active three-dimensional shape of the protein. How the sequence information is used to define the structure is not understood in detail. What is clear, however, is that substantial free energy is required to drive the polymer chain into a well-defined structure, and to prevent it from unravelling. It is also generally accepted that the energy for this process comes primarily from the hydrophobic effect.

Polar and Nonpolar Amino Acids

About a quarter of the amino acids have side-chains that are normally charged. These are called 'hydrophilic' (water loving) and prefer to be in an aqueous environment. In contrast, about another third have side-chains that are made up of hydrocarbon atoms. Three examples are alanine, leucine and phenylalanine (**Table 1**). These side-chains are 'oily' in character and prefer to be segregated in contact with each other, and out of contact with water. The amino acids that contain such nonpolar side-chains are described as 'hydrophobic'. The side-chains of the remaining amino acids typically contain both polar and nonpolar atoms and have properties that reflect the characteristics of both.

When a protein folds into a well-defined three-dimensional structure, the majority of the hydrophobic side-chains cluster together within the core of the protein (something like a tiny oil-drop). This removal of the hydrophobic side-chains from contact with solvent is highly favourable and generates sufficient free energy to maintain the folded structure of the protein.

Forces that Stabilize Protein Structures

Even simple proteins have complicated three-dimensional shapes that can include α helices, β strands, turns and irregular segments (**Figure 1**). A number of different types of interaction help define the structure. These include hydrogen bonds, electrostatic interactions, van der Waals interactions and hydrophobic interactions.

Because proteins fold in an aqueous environment, the contribution of a given interaction to the folding of the protein depends not so much on the strength of interaction within the protein but on the difference between the strength of the interaction within the protein and the strength of interaction of the same groups with water. A hydrogen bond, for example, may occur in the folded protein between a hydrogen bond donor and a hydrogen bond acceptor. In the unfolded protein, however, both the donor and acceptor will make more-or-less equivalent hydrogen bonds to water. Thus the net energetic contribution to protein folding from hydrogen bonding tends to be rather weak. (Hydrogen bonds are, however, thought to be especially important in discriminating between the correct folded structure and incorrect ones.) Similarly, van der Waals interactions occur throughout folded proteins, but equivalent interactions with solvent can occur in the unfolded form.
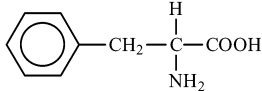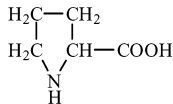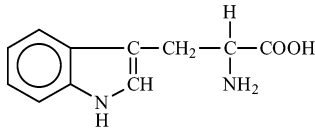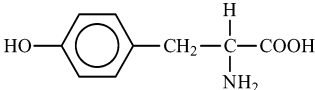
Hydrophobic interactions, however, are different. As noted above, protein folding occurs in the presence of water and the properties of water are dominated by its propensity to form hydrogen bonds. Polar compounds such as sugars can share hydrogen bonds with water and, for this reason, are readily soluble. In contrast, when a hydrophobic (nonpolar) surface is introduced into an aqueous environment it precludes hydrogen bonding. This preclusion of hydrogen bonding to the hydrophobic surface forces the water molecules to adopt alternative

Table 1 List of 20 fundamental amino acids and their abbreviations

Amino acid	Three-letter abbreviation	One-letter symbol	Formula
Alanine	Ala	A	$CH_3-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Arginine	Arg	R	$H_2N-\overset{\overset{}{}}{\underset{\underset{\displaystyle NH_2}{\parallel}}{C}}-NH-CH_2-CH_2-CH_2-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Asparagine	Asn	N	$H_2N-\overset{\overset{}{}}{\underset{\underset{\displaystyle O}{\parallel}}{C}}-CH_2-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Aspartic acid	Asp	D	$HOOC-CH_2-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Cysteine	Cys	C	$HS-CH_2-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Glutamic acid	Glu	E	$HOOC-CH_2-CH_2-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Glutamine	Gln	Q	$H_2N-\overset{\overset{}{}}{\underset{\underset{\displaystyle O}{\parallel}}{C}}-CH_2-CH_2-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Glycine	Gly	G	$H-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Histidine	His	H	$HC\!=\!\overset{}{\underset{\underset{\displaystyle N}{\parallel}}{C}}-\overset{}{\underset{\underset{\displaystyle NH}{\mid}}{C}}-CH_2-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$ (ring closed by $\overset{\displaystyle C}{\underset{\displaystyle H}{}}$)
Isoleucine	Ile	I	$CH_3-CH_2-\overset{}{\underset{\underset{\displaystyle CH_3}{\mid}}{CH}}-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$
Leucine	Leu	L	$\overset{\displaystyle H_3C}{\underset{\displaystyle H_3C}{>}}CH-CH_2-\overset{\overset{\displaystyle H}{\mid}}{\underset{\underset{\displaystyle NH_2}{\mid}}{C}}-COOH$

continued

Table 1 – *continued*

Amino acid	Three-letter abbreviation	One-letter symbol	Formula
Lysine	Lys	K	$H_2N{-}CH_2{-}CH_2{-}CH_2{-}CH_2{-}\underset{NH_2}{\overset{H}{C}}{-}COOH$
Methionine	Met	M	$CH_3{-}S{-}CH_2{-}CH_2{-}\underset{NH_2}{\overset{H}{C}}{-}COOH$
Phenylalanine	Phe	F	$C_6H_5{-}CH_2{-}\underset{NH_2}{\overset{H}{C}}{-}COOH$
Proline	Pro	P	$\begin{matrix}H_2C{-}CH_2\\ H_2C\quad CH{-}COOH\\ \underset{H}{N}\end{matrix}$
Serine	Ser	S	$HO{-}CH_2{-}\underset{NH_2}{\overset{H}{C}}{-}COOH$
Threonine	Thr	T	$CH_3{-}\underset{OH}{CH}{-}\underset{NH_2}{\overset{H}{C}}{-}COOH$
Tryptophan	Trp	W	$\begin{matrix}C{-}CH_2{-}\underset{NH_2}{\overset{H}{C}}{-}COOH\\ CH\\ \underset{H}{N}\end{matrix}$
Tyrosine	Tyr	Y	$HO{-}C_6H_4{-}CH_2{-}\underset{NH_2}{\overset{H}{C}}{-}COOH$
Valine	Val	V	$\begin{matrix}H_3C\\ \quad CH{-}\underset{NH_2}{\overset{H}{C}}{-}COOH\\ H_3C\end{matrix}$

arrangements that permit hydrogen bonding to other water molecules. This imposed restriction on the alignment of the water molecules (strictly speaking, a reduction in their entropy) has an energetic cost and is the physical basis of the hydrophobic effect. Because the folding of a protein includes the removal of many nonpolar side-chains from an aqueous environment and their sequestration from solvent, the energy benefit can be very substantial.

Estimation of the Strength of the Hydrophobic Effect

The classical way to estimate the magnitude of the hydrophobic effect for a given compound is to measure the free energy of transfer, ΔG_{tr}, of the compound from the gas, liquid or solid phase to water. A positive value for ΔG_{tr} means that the molecule prefers a nonaqueous environ-
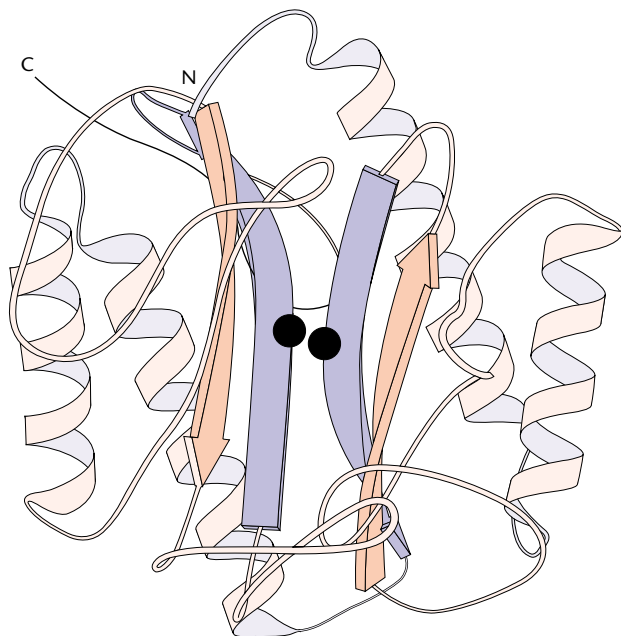
Figure 1 Sketch of the backbone structure of the protein methionine aminopeptidase from *Escherichia coli*. The chain begins at the amino-terminus (N), includes 264 amino acids, and ends at the carboxy-terminus (C). The protein includes α helices and β sheet strands as well as two metal ions (spheres) at the active site. Reprinted from Bazan JF *et al*. (1994) *Proceedings of the National Academy of Sciences of the USA* **91**: 2473–2477, figure 2.

ment. In the case of the amino acids, measurements can be made with the free amino acid or with variants modified to better represent the amino acids incorporated within the protein chain.

As well as choosing the particular compounds to be investigated, one also has to decide which transfer medium best mimics the interior of a protein. For this reason many different hydrophobicity scales are available. **Table 2** shows hydrophobicities of the 20 amino acids based on transfer between water and octanol. It is the relative value of the hydrophobicity that is relevant. For this reason it is often convenient to define the hydrophobicity of glycine to be zero and to quote values for the other amino acids relative to this reference.

As can be seen in **Table 2**, amino acids with large, nonpolar or largely nonpolar side-chains such as leucine and tryptophan (**Table 1**) are most hydrophobic. The least hydrophobic amino acids are those that are charged and those such as asparagine that are largely polar.

Relationship between the Hydrophobic Effect and Surface Area

It was noted above that the hydrophobic effect is due to the effect of nonpolar atoms on surrounding water molecules. As such, one would expect the magnitude of the hydrophobic effect for a given amino acid to be proportional to the surface area of the nonpolar atoms that it contains.

That this is the case is illustrated in **Figure 2**. The amino acids alanine, valine, leucine and phenylalanine have side-chains that are made up of non-polar hydrocarbon atoms (**Table 1**). These lie on one straight line. The other amino acids shown in **Figure 2** have side-chains that are partly made up of hydrocarbon atoms but also include some polar atoms as well. These lie on a second line.

The Hydrophobic Moment

As shown in **Table 2**, the hydrophobicities of the given amino acids vary substantially. Some are strongly hydrophobic while others are strongly hydrophilic. A molecule

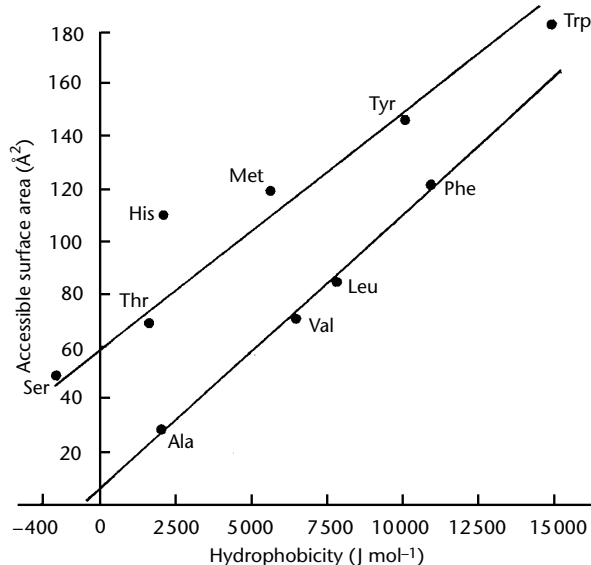**Table 2** Hydrophobicities[a] of the 20 naturally occurring amino acids

Amino acid	Hydrophobicity	
	(kJ mol^{-1})	(kcal mol^{-1})
Tryptophan	9.41	2.25
Phenylalanine	7.49	1.79
Isoleucine	7.53	1.80
Leucine	7.11	1.70
Cysteine	6.44	1.54
Methionine	5.14	1.23
Valine	5.10	1.22
Tyrosine	4.02	0.96
Proline	3.01	0.72
Alanine	1.30	0.31
Threonine	1.09	0.26
Glycine	0.00	0.00
Serine	–0.17	–0.04
Histidine	0.54	0.13
Glutamine	–0.92	–0.22
Asparagine	–2.51	–0.60
Glutamic acid	–2.68	–0.64
Aspartic acid	–3.22	–0.77
Lysine	–4.14	–0.99
Arginine	–4.23	–1.01

[a]The hydrophobicities are based on the solvent transfer free energies from octanol to water.

Table 3 Changes in protein stability resulting from the replacement of larger hydrophobic amino acids with smaller ones

Substitution	Number of examples	$\Delta\Delta G$ (kJ mol^{-1})		
		Low	High	Average
Ile → Val	9	2.1	7.6	5.4±1.7
Ile → Ala	9	4.6	21.3	15.9±2.9
Leu → Ala	17	7.1	25.9	14.6±4.6
Val → Ala	11	0.0	19.7	10.5±3.8
Met → Ala	4	8.8	19.2	12.5±3.8
Phe → Ala	4	14.6	18.4	15.9±1.2

that includes both hydrophobic and hydrophilic parts is called amphiphilic. For such amphiphilic molecules it is sometimes useful to define a hydrophobic moment, which is analogous to a dipole moment. For a single amino acid, the hydrophobic moment can be defined as a line that points from the Cα atom to the middle of the side-chain, and whose length is proportional to the hydrophobicity of the side-chain. For a protein or part of a protein, the dipole moment is obtained by summing the individual vectors (in magnitude and direction) corresponding to the constituent amino acids. As an example, an α helix located on the surface of a protein will have one side of the helix exposed to solvent and the other side facing the interior of the protein. The amino acids that comprise the buried side of the α helix will, in general, be much more hydrophobic than those on the solvent-exposed side of the helix. Because of this asymmetry the α helix will have a large hydrophobic moment directed towards the centre of the protein.

Core Packing and the Effects of Mutations

Attempts have been made to estimate the magnitude of the hydrophobic effect by substituting one nonpolar amino acid for another within the core of a protein and measuring the resultant change in stability of the protein. One difficulty, however, is that the cores of proteins are tightly packed and the substitution of amino acids of different shapes and sizes tends to introduce steric clashes that complicate the interpretation of the experiment. A possible way to avoid such steric interference is to only use replacements in which a larger nonpolar residue is replaced
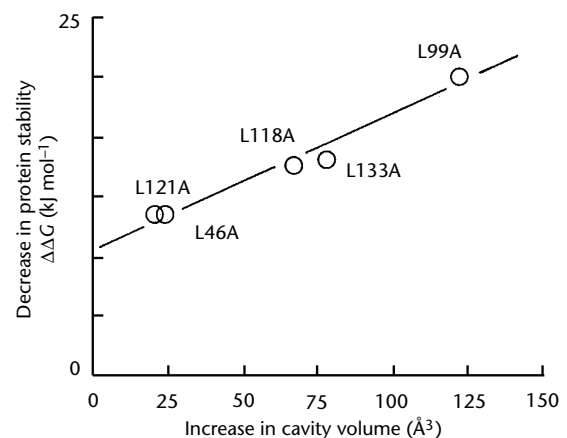
Figure 2 Relationship between the hydrophobicities of the amino acids and the solvent-accessible areas of their side-chains. Reprinted and adopted from Chothia C (1974) *Nature* **248**: 338–339, figure 1.

Figure 3 Loss of protein stability, $\Delta\Delta G$, for a series of leucine-to-alanine substitutions within T4 lysozyme. 'L99A', for example, denotes the mutant in which leucine 99 is replaced by alanine. Mutations that result in cavities of the largest volume cause the greatest loss of stability. At a cavity volume of zero, the loss of stability can be attributed to the difference between the hydrophobicity of leucine and alanine. Reprinted and adapted from Xu J *et al.* (1998) *Protein Science* **7**: 158–177, figure 17A.

with a smaller one. **Table 3** summarizes the results of a series of experiments of this type. If we consider, for example, the leucine-to-alanine substitutions included in the table, the average loss in stability is $14.6 \pm 4.6\,\text{kJ mol}^{-1}$. As can be seen, however, there is a very large spread in the individual measurements, ranging from 7.1 to $25.9\,\text{kJ mol}^{-1}$. This variation occurs because a given protein may respond in different ways to leucine-to-alanine substitutions at different sites. Sometimes the protein structure surrounding the replacement will hardly change at all, with the result that a cavity will be formed. In other cases the atoms surrounding the smaller replacement will collapse or partly collapse to occupy the space vacated by the large side-chain. Mutations that lead to the creation of larger cavities tend to be more destabilizing than those that lead to small cavities (**Figure 3**). This in turn suggests that the loss in stability resulting from large-to-small substitutions such as leucine-to-alanine is due to two different factors: (1) the hydrophobic effect and (2) van der Waals interactions between the large side-chain and the atoms it contacts. If the protein structure remains the same, the replacement of the large side-chain with a smaller one will result in the loss of some of these favourable van der Waals contacts. However, if the structure relaxes in response to the mutation, it will regenerate new van der Waals contacts that will tend to offset those present in the parent structure. This van der Waals term, which varies from one mutation site to another, explains why the loss of stability depends on the size of the cavity that is created (**Figure 3**).

If the straight line shown in **Figure 3** is extrapolated to zero cavity volume, one can define this intercept as ΔG_0. Here the protein structure relaxes to completely fill any space created by the mutation. The intercept value is the difference between the energy of burying an alanine rather than a leucine in the core of the protein. In other words, the intercept measures the difference between the hydrophobic stabilization of leucine and alanine. The value of $8.8\,\text{kJ mol}^{-1}$ obtained from **Figure 3** corresponds only moderately well to the difference between the octanol-to-water transfer free energies of the same two amino acids ($5.81\,\text{kJ mol}^{-1}$; **Table 2**). While the approach illustrated in **Figure 3** illustrates, in principle, how amino acid hydrophobicities can be determined from mutant proteins, its use is limited in practice by a number of technical factors. For this reason the amino acid hydrophobicities obtained from solvent transfer experiments (such as in **Table 2**) are recommended for everyday use.

Further Reading

Creighton TE (1992) *Protein Folding*. New York: WH Freeman.

Richards FM (1991) The protein folding problem. *Scientific American* (January 1991), pp. 54–63.

Matthews BW (1996) Structural and genetic analysis of the folding and function of T4 lysozyme. *FASEB Journal* **10**: 35–41.

Pace CN, Shirley BA, McNutt M and Gajiwala K (1996) Forces contributing to the conformational stability of proteins. *FASEB Journal* **10**: 75–83.

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.

Xu J, Baase WA, Baldwin E and Matthews BW (1998) The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Science* **7**: 158–177.