Hydrophobicity and Amphiphilicity in Protein Structure

David Eisenberg, William Wilcox, and Andrew D. McLachlan

Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024 (D.E., W.W.) and MRC Laboratory of Molecular Biology, MRC, Cambridge CB2 2QH, United Kingdom (A.D.M.)

The importance of the hydrophobic interaction in stabilizing native protein structure has long been appreciated. However, more than other component forces, this one has resisted quantitative description. We present two approximate methods of assessing the hydrophobic component to the free energy of protein folding. Both are expressed in terms of what can be called *hydrophobic moments* of the protein. The first method is intended to yield an approximate value for the hydrophobic energy. This energy is calculated from a set of atomic coordinates in terms of the hydrophobicity (or 0th hydrophobic moment) of each amino acid residue and its accessibility or lack of it to aqueous solvent. The second method considers the first moment of the hydrophobicity of a group of residues, the hydrophobic moment. Segments of secondary structure in folded proteins tend to have hydrophobic moments that oppose each other. For example, α-helices on the protein surface tend to have one hydrophobic face and one hydrophilic face, with the hydrophilic face out towards the solvent. This pattern of organization is often apparent from a computer model of the protein that shows the magnitude and direction of the hydrophobic moment of each segment of secondary structure. Examples are given for the incorrectly folded structures of Novotný et al [J Mol Biol 177:787, 1984] and for the correct structures to which they correspond.

The energy of folding a protein from a disorganized coil to a native structure is often considered to be a sum of terms involving various covalent and noncovalent forces [1–4]. Of these terms, the hydrophobic interaction has been the most elusive for treatment with standard potential energy functions. The reason is that this contribution depends in a complex way on the arrangements of water about both the native and the unfolded structures [5–7]. Because of this, the hydrophobic contribution to the free energy cannot be expressed simply as an interaction between pairs of protein atoms.

The importance of the hydrophobic interaction in maintaining the stabilities of proteins was first clearly stated by Kauzmann [5]. Since then, there have been many

contributions to understanding this force [eg, 6–11]. An important step in providing a quantitative link between protein structure and energy was made by Richards and others [12–15], who emphasized that the accessibility to aqueous solvent of protein atoms can be expressed in terms of the areas of atoms adjacent to the protein surface. As a protein folds, carbon and sulfur atoms lose a greater fraction of their area than do oxygen and nitrogen atoms. However, the problem remains of establishing a relationship between accessible area and energy.

This link can be provided by considering the hydrophobicity of exposed and buried amino acid residues. The hydrophobicity, H, of a residue is a measure of its free energy of transfer from a relatively apolar phase, such as the protein interior, to aqueous solution. In a pioneering paper, Nozaki and Tanford [16] measured hydrophobicities for several amino residues, and their work has been followed by many useful studies, which have gradually provided values of hydrophobicities for all the 20 coded amino acid residues of proteins [eg, 17–20].

By combining measurements of accessible surface areas of proteins with numerical hydrophobicities, we present here a method for estimating the hydrophobic energy of protein folding. A second method for assessing the hydrophobic interaction in protein folding combines residue hydrophobicities with known features of protein structure. In this method, the directions and magnitudes of the hydrophobic moments of the various segments of secondary structure of a protein are considered. It can be illustrated with the incorrect protein folds devised by Novotný et al [1] and their correctly folded analogs.

METHODS
The Hydrophobic Residue Method

It has been suggested that a rough estimate for the free energy of a protein is given by [20]:

$$G_H = \sum_{\text{residues, } i} H_i M(x_i),$$ (1)

in which H_i is the hydrophobicity of the ith residue and $M(x_i)$ represents the hydrophobicity of the environment. Suppose that some fraction A/A^0 of the surface of a residue is available to solvent and that the rest of the residue $[1 - (A/A^0)]$ is buried within the protein, hence shielded from solvent. Further suppose that we can represent the hydrophobicity of the environment of the aqueous solvent by $+1/2$ and the hydrophobicity of the environment of the interior of the protein by $-1/2$. The justification for the values of $\pm 1/2$ is implicit in Equation 1 if G_H is regarded as the free energy of transfer of a single residue of hydrophobicity H_i. Then we can write:

$$G_H = \sum_{\text{residues}} H_i \left[\frac{A_i}{A_i^0} - 1/2 \right].$$ (2)

For the calculation of G_H by Equation 2, we have used values of A_i/A_i^0 tabulated by Novotný et al [1]. For values of H, we have adopted the consensus scale of Eisenberg et al [20]. It is also possible to extend Equation 2 to use atomic accessibilities rather than residue accessibilities [21].

Hydrophobic Moments

Structural hydrophobic moments were calculated for segments of secondary structure in the example proteins using formula 2 from Eisenberg et al [20]:

$$\underset{\sim}{\mu}_H = \sum_{\text{residues, } i} H_i \, \underline{s}_i, \qquad (3)$$

in which the sum is over all residues in a given segment of secondary structure, H_i is the hydrophobicity of the ith residue, and $\underline{s}_i$ is a unit vector pointing from the alpha carbon atom of the ith residue to the center of the residue's side chain. The $\underline{\mu}_H$ vector for each segment is displayed as pointing from the center of the segment toward the direction of greater hydrophobicity.

RESULTS

Hydrophobic Moments of Segments of Protein Secondary Structure

In a previous study, we inspected the relative directions and magnitudes of hydrophobic moments of the segments of secondary structure in folded globular proteins [20]. We found that they tend to cancel. The reason for this is that hydrophobic side chains of different segments tend to face each other. For example (see Fig. 1 in Eisenberg et al [20]), the hydrophobic moments of the four α-helices of the melittin molecule point inward towards each other and effectively cancel.

A similar pattern is found in the four α-helical structure of hemerytherin from T dyseritum [22] as is shown in Figure 1A. This is one of the two 113-residue proteins whose energetics of folding were compared by Novotný et al [1]. As in the case of melittin, the hydrophobic moments of the four helices tend to oppose those of neighboring helices. This pattern of opposing hydrophobic moments of helices is generally observed in antiparallel α-helical bundles of the type discussed by Weber and Salemme [23].

In two-layer β-sheet structures, as in the α-helical structures, the hydrophobic moments of individual strands tend to point inwards. This is so in the variable domain of the mouse immunoglobulin κ-chain, here termed VL, whose structure was determined by Segal et al [24] and whose energetics folding were analyzed by Novotný et al [1]. The nine strands and their associated moments are shown in Figure 2A.

These coherent patterns of residue organization that give rise to the opposing hydrophobic moments of these actual protein structures are not present in the incorrectly folded protein molecules devised by Novotný et al [1]. These authors provided useful but improperly folded protein models by replacing the side chains of hemerythrin by those of VL and vice versa. Then, using the program CHARMM, they made small adjustments in coordinates to reduce the potential energies of the two structures to a local minimum. In the resulting incorrect fold of VL arranged as hemerythrin (Fig. 2B), all the moments point out towards aqueous solution. The pattern is unlike the one observed in real helical proteins.

A similar lack of coherence is seen in their incorrectly folded hemerythrin molecule arranged in the pattern of the VL domain (Fig. 2B). The hydrophobic moments of the segments of secondary structure do not point generally inwards, as they do in the VL structure; all are small or point outwards.
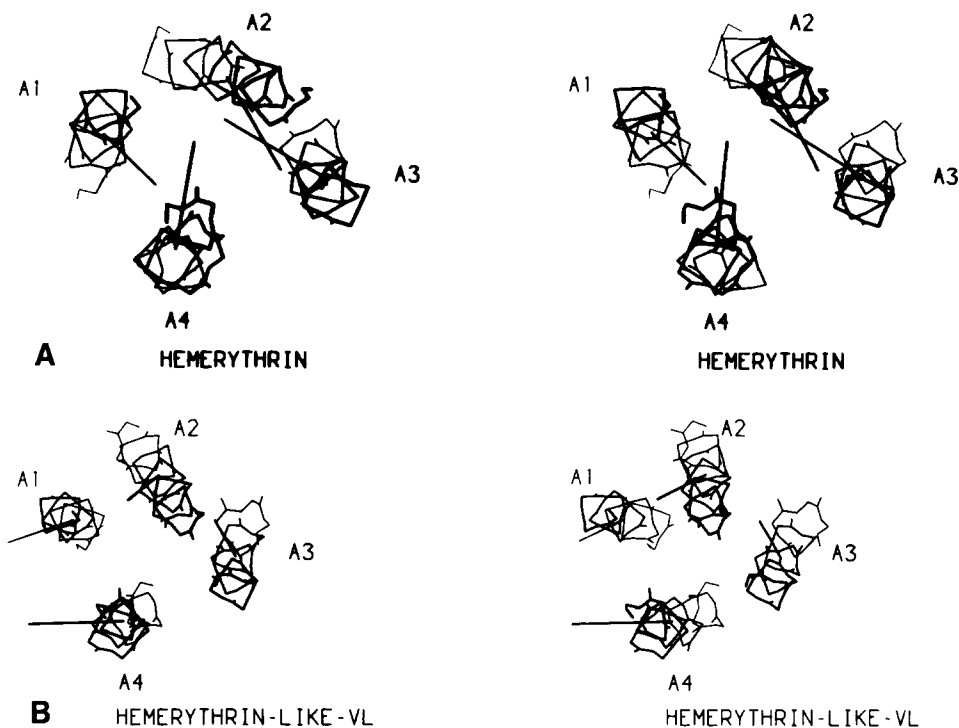
Fig. 1. Hydrophobic moments of helices in a real and in a misfolded protein. A1 is the first helix in the chain, A2 is the second, and so forth. A) The four α-helical segments from the structure of hemerythrin from T dyscritum [22]. Note that the hydrophobic moments from the four helices point inwards and oppose each other. Each moment is represented as a line drawn from the center of its segment toward the direction of greater hydrophobicity. B) The incorrectly folded structure of a mouse κ-chain VL domain [24] arranged by Novotný et al [1] into the structure of hemerythrin. The hydrophobic moments of the four helices are shown by lines emerging from the centers of the helices. These are scaled by a factor of 2 over those in A for better visibility.

Calculation of Hydrophobic Energies

A highly approximate estimate of the hydrophobic component of the free energy of folding can be made from Equation 2 and from a knowledge of the accessible surface areas of residues in a protein. For the two incorrectly folded models, and their correct analogues, Novotný et al [1] have tabulated the fractional accessibility of each type of residue in the structure. These accessibilities can be combined with the number and hydrophobicity of each type of residue in Equation 2 to calculate a crude hydrophobic free energy. For these four structures, we have done this, and the energies are given in Table I. Both actual structures are more stable than their incorrectly folded analogs by about 9 kcal mol^{-1}.

DISCUSSION
Energetics of Correctly and Incorrectly Folded Structures

The two incorrectly folded structures, hemerythrin and the immunoglobulin variable domain VL, devised by Novotný et al [1] offer a convenient testing ground
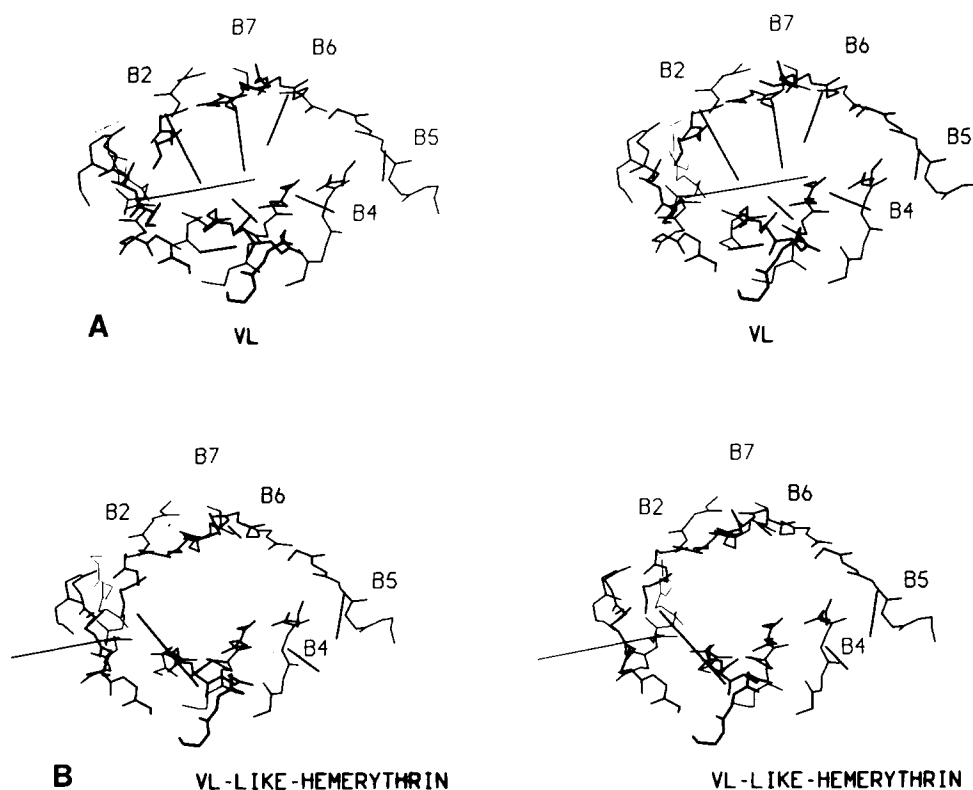
Fig. 2.   Hydrophobic moments of β-sheets in a real and in a misfolded protein. B1 is the first strand of sheet in the polypeptide chain, B2 is the second, and so forth. A) the nine-stranded β-sheet of a mouse κ-chain VL domain [24]. The hydrophobic moment of each strand of the sheet is shown as a line emerging from the center of the strand. The lengths are scaled by a factor of 1.5 over those in Figure 1A. B) the incorrectly folded structure of hemerythrin from T dyscritum [22] as arranged by Novotný et al [1]. The moments are on a scale two times those in Figure 1A.

TABLE I. Hydrophobic Contribution to the Free Energy of Protein Folding, G_H, Computed From Equation 2 as Described in the Text

Protein	Type of structure	G_H (kcal mol^{-1})
Hemerythrin	Actual	−6.4
Hemerythrin-like VL domain	Misfolded	+2.8
Net stabilization of actual structure		−9.2
VL domain	Actual	−7.6
VL-like hemerythrin	Misfolded	+1.4
Net stabilization of actual structure		−9.0

for quantitative theories of protein folding. Novotný et al note in their analysis of these structures that side-chain nonpolar surface area of the incorrectly folded structures is greater than for the correctly folded structures. This finding is reflected in Table I in terms of hydrophobic free energies computed from Equation 2. Notice that both actual structures are stabilized by their hydrophobic free energies and that both hypothetical, misfolded structures are destabilized.

The origin of the higher (destabilizing) energy in the misfolded structures is a combination of greater solvent accessibility of hydrophobic residues and diminished solvent accessibility of hydrophilic residues. In the VL structure, all residue types other than Gly, His, and Arg lower the hydrophobic free energy. However, in hemerythrin-like VL, a positive contribution to the hydrophobic free energy is contributed by the following residue types: Ala, Cys, Gln, Gly, His Lys, Phe, Tyr, and Val.

It should be emphasized that the hydrophobic free energies in Table I rest on many assumptions and are rough estimates at best. Among other factors, they neglect the amphiphilic character of amino acids such as Lys, Arg, Glu, Tyr, and Trp in which there are both hydrophilic and hydrophobic portions.

Assessing and Designing Structures With the Aid of the Hydrophobic Moment

The examples in Figures 1 and 2 as well as those of others [20] suggest that actual protein structures are characterized by coherent patterns of opposed hydrophobic moments in their segments of secondary structure. In contrast, the two hypothetical, misfolded structures of Figures 1B and 2B lack these coherent patterns. Thus a visual inspection of a protein in terms of the directions of the hydrophobic moments of its segments may be useful for assessment and design of proposed structures.

ACKNOWLEDGMENTS

REFERENCES

1. Novotný J, Bruccoleri R, Karplus M: J Mol Biol 177:787, 1984.
2. Karplus M, McCammon A: Annu Rev Biochem 52:263, 1983.
3. Némethy G, Scheraga HA: Quart Rev Biophys 10:239, 1977.
4. Levitt M, Warshel A: Nature 253:692, 1975.
5. Kauzmann W: Adv Protein Chem 14:1, 1959.
6. Némethy G, Scheraga HA: J Phys Chem 66:1773, 1962.
7. Klapper MH: Prog Bioorganic Chem 2:55, 1973.
8. Pratt L, Chandler D: J Chem Phys 67:3683, 1977.
9. Pratt L, Chandler D: J Chem Phys 73:3434, 1980.
10. Tanford C: Proc Natl Sci USA 76:4175, 1979.
11. Tanford C: "The Hydrophobic Effect, 2nd ed." New York: Wiley, 1980.
12. Lee B, Richards RM: J Mol Biol 55:379, 1971.
13. Shrake A, Rupley JA: J Mol Biol 79:351, 1973.
14. Richmond TJ: J Mol Biol 178:63, 1984.
15. Richards FM: Annu Rev Biophys Bioeng 6:151, 1977.
16. Nozaki Y, Tanford C: J Biol Chem 246:2211, 1971.

17. Wolfenden R, Anderson L, Cullis PM, Southgate CCB: Biochemistry 20:849, 1981.
18. Flauchere J-L, Pliska V: Eur J Med Chem-Chem Ther 18:369, 1983.
19. Janin J: Nature 277:491, 1979.
20. Eisenberg D, Weiss RM, Terwilliger TC, Wilcox W: Faraday Symp Chem Soc 17:109, 1982.
21. Eisenberg D, McLachlin AD: Nature 319:199, 1986.
22. Stenkamp RE, Sieker LC, Jensen LH: Acta Crystallogr Sect B 38:784, 1978.
23. Weber P, Salemme FR: Nature 287:82–84, 1980.
24. Segal DM, Padlan EA, Cohen GH, Rudikoff S, Potter M, Davies DR: Proc Natl Acad Sci USA 71:4298, 1974.