Faraday Symp. Chem. Soc., 1982, **17**, 109–120

Hydrophobic Moments and Protein Structure

BY DAVID EISENBERG, ROBERT M. WEISS, THOMAS C. TERWILLIGER AND WILLIAM WILCOX

Institute of Molecular Biology and Department of Chemistry,
University of California, Los Angeles,
California 90024, U.S.A.

Received 9th August, 1982

The structure of a protein can be analysed in terms of what may be called the " hydrophobic moments " of (1) the entire molecule and (2) of the segments of secondary structure that make up the polypeptide chain. The zeroth moment is defined as the sum of the hydrophobicities of the amino-acid residues of the structure under consideration; it is the analogue of the net charge of a cluster of point charges. The first moment, or hydrophobic dipole moment, is the analogue of the electric dipole moment of a cluster of charges. Just as the electric dipole moment measures the asymmetry of the charge distribution, the hydrophobic dipole moment measures the amphiphilicity (asymmetry of hydrophobicity) of the structure. A large hydrophobic dipole moment indicates that a structure is predominantly hydrophobic on one side and predominantly hydrophilic on the other. A quadrupole hydrophobic moment may be similarly defined. It indicates whether a protein is more hydrophobic in its interior (as for a globular protein in aqueous solution) or at its surface (as for a membrane protein).

The hydrophobic dipole moment is useful in two separate applications. The first, which we have discussed elsewhere [D. Eisenberg, R. M. Weiss and T. C. Terwilliger, *Nature (London)*, 1982, **299**, 371] relates the function and secondary structure of a region of protein structure to its amino-acid sequence. For example, sequences that form surface-seeking helices have large hydrophobic dipole moments. The second application is in the analysis of interactions of a segment or domain of a protein with neighbouring regions in the protein and with other parts of the environment. In this paper we examine the hydrophobic dipole moments in the known structures of nine globular proteins and find that the moments of neighbouring segments tend to point towards each other. This suggests that the hydrophobic dipole can serve as a simple and pictorial summary of some of the forces at work in the folding of proteins. We have also found that the interactions of some macromolecules with an apolar–polar interface can be conveniently described in terms of the hydrophobic moments of the molecule and a " hydrophobic field " which reflects the hydrophobicity of the environment.

Since Kauzmann's review of 1959 [1] it has been widely appreciated that the hydrophobic interaction is of importance in the folding of proteins. In more recent years there have been several attempts to describe aspects of protein folding quantitatively in terms of the relative hydrophobic or hydrophilic character of the various amino-acid residues. These attempts involve assigning a numerical " hydrophobicity " to each type of amino acid, and then relating these hydrophobicities in a particular protein or fragment with some aspect of structure or function. For example, Capaldi and Vanderkooi [2] noted that strongly bound membrane proteins are especially rich in highly hydrophobic amino-acid residues, and Segrest and Feldman [3] pointed out that trans-membrane segments of proteins have a continuous segment of residues that are particularly hydrophobic. These studies have dealt with what we call the zeroth hydrophobic moment; we show here that the higher moments of the hydrophobicity are also useful in characterizing protein structure and folding.

CONCEPTS

HYDROPHOBICITY SCALES

The hydrophobicity of an amino-acid residue is not a property that can be defined easily or measured simply. Nevertheless, several groups have attempted to derive numerical hydrophobicity scales using a variety of experimental and computational methods. The resulting values generally correspond to the free energy of transfer of the side chain of the amino acid from water to an apolar environment. This free energy thus excludes that due to main-chain atoms in the protein structure. Among recent comparisons of such hydrophobicity-scales are those of Meirovitch *et al.*,[4] Kyte and Doolittle [5] and Edsall and McKenzie.[6] Because each of the methods used to derive such a scale reflects some aspect of the hydrophobicity, and because not one of the methods seems vastly more reliable than the others, our approach has been to take a consensus of the values determined by some of the more reliable methods.

Our " consensus " hydrophobicity scale is given in table 1, together with the five scales from which it has been inferred. One is the well known scale of Nozaki and Tanford [7] as modified by Segrest and Feldman,[3] based on free energies of transfer of amino-acid side chains from water to ethanol. Another is the scale of Wolfenden *et al.*[8] for the transfer from water to vapour. The scales of Chothia [9] and Janin [10] are from calculated free energy changes for the transfer of amino-acid side chains from the surface to the interior of a protein, based on the observed distribution (between surface and interior) of each residue type for globular proteins of known structure. Another set of semi-empirical calculated values was published by von Heijne and Blomberg,[11] who considered the free energy changes associated with the transfer from aqueous to non-aqueous media of the amino-acid side chains. They attempted to include effects of buried non-polar atoms, removal of charges and decrease of hydrogen bonding during burial.

The consensus scale was derived as follows. We noted that for each of the four complete hydrophobicity scales on the left of table 1, the value for serine lies at the mean, or very close to it. For each scale, the standard deviation of the hydrophobicities was determined. The " normalized " hydrophobicity of a residue in each scale was defined as the number of standard deviations that its hydrophobicity lay above or below the mean. The scales were combined by averaging the normalized hydrophobicities for each residue over the five scales. To convert this averaged scale to the units of kcal mol^{-1}, all hydrophobicities were arbitrarily multiplied by the standard deviation of the scale of Janin [10] and added to its mean. The result is the final column in table 1. We used this consensus scale of hydrophobicities for all the calculations reported in this paper.

ESTIMATION OF THE HYDROPHOBIC DIPOLE MOMENT OF AN AMINO-ACID SEQUENCE

The hydrophobic dipole moment can be estimated from a known amino-acid sequence if it is assumed that the polypeptide backbone follows some periodic arrangement such as an alpha helix [11] or a strand from a beta sheet. For this calculation we also assumed that the hydrophobicity of each residue i can be represented by a vector of length H_i, having a direction perpendicular to the axis of the helix or strand of beta structure [see fig. 1 of ref. (1)]. Then the magnitude of the estimated hydrophobic dipole moment is

$$\mu_H = \{[\sum_i H_i \cos(i\delta)]^2 + [\sum H_i \sin(i\delta)]^2\}^{\frac{1}{2}} \quad (1)$$

TABLE 1.—HYDROPHOBICITY SCALES FOR AMINO-ACID RESIDUES

The order of residues is by decreasing hydrophobicity on the consensus scale, and the magnitudes may be considered roughly in kcal mol^{-1} for transfer from a hydrophobic to a hydrophilic phase.

	scale and reference					
residue	Tanford [3,7]	von Heijne–Blomberg [12]	Janin [10]	Chothia [9]	Wolfenden [8]	consensus (this paper)
Ile	5.0	4.4	0.7	0.24	2.15	0.73
Phe	5.0	5.2	0.5	0.0	−0.76	0.61
Val	3.0	3.9	0.6	0.09	1.99	0.54
Leu	3.5	4.2	0.5	−0.12	2.28	0.53
Trp	6.5	3.9	0.3	−0.59	−5.88	0.37
Met	2.5	2.1	0.4	−0.24	−1.48	0.26
Ala	1.0	2.9	0.3	−0.29	1.94	0.25
Gly	0.0	1.9	0.3	−0.34	2.39	0.16
Cys	0.0	−0.08	0.9	0.0	−1.24	0.04
Tyr	4.5	3.6	−0.4	−1.02	−6.11	0.02
Pro	1.5	1.1	−0.3	−0.90	—	−0.07
Thr	0.5	1.2	−0.2	−0.71	−4.88	−0.18
Ser	−0.5	0.36	−0.1	−0.75	−5.06	−0.26
His	1.0	−1.5	−0.1	−0.94	−10.3	−0.40
Glu	—	−4.0	−0.7	−0.90	−10.2	−0.62
Asn	−1.5	−1.0	−0.5	−1.18	−9.68	−0.64
Gln	−1.0	−0.52	−0.7	−1.53	−9.38	−0.69
Asp	—	−5.6	−0.6	−1.02	−10.9	−0.72
Lys	—	−2.3	−1.8	−2.05	−9.52	−1.1
Arg	—	−9.4	−1.4	−2.71	−19.9	−1.8

where δ is the angle separating side chains along the backbone (*e.g.* $\delta = 100°$ for an alpha helix). This is a quantitative expression of the helical wheel proposed by Schiffer and Edmundson.[13]

CALCULATION OF HYDROPHOBIC DIPOLE MOMENTS FROM COORDINATES

When the three-dimensional structure of a protein is known, the hydrophobic dipole moment can be computed from the coordinates (the " structural " moment). We have used two different functions for this purpose.[1] The first is

$$\mu_{s1} = \sum_i H_i s_i \tag{2}$$

in which s_i is a unit vector pointing from the alpha carbon atom of the ith residue to the centre of the residue's side chain. With this definition, the hydrophobic dipole moment can be calculated for any segment of a protein structure, whether periodic or irregular. When computed for a periodic structure, however, like μ_H it emphasizes the amphiphilicity perpendicular to the axis of the helix or strand of beta structure.

A second definition of the structural hydrophobic dipole moment is

$$\mu_{s2} = \sum_i H_i r_i - \langle H_i \rangle \sum r_i \tag{3}$$

in which r_i is a vector from any origin to the centre of the chain of the ith residue,

and $\langle H_i \rangle$ is the average hydrophobicity for the amino-acid side chains in the structure. This moment, like μ_{s1}, is independent of the choice of origin. In contrast to μ_{s1}, it represents amphiphilicity in directions both parallel and perpendicular to the axis of secondary structure.

In our analysis of the structures of folded proteins (below) we have used the moment μ_{s1}. The sum is normally taken over the residues of one segment of secondary structure at a time. We find that the values of μ_{s1} for the corresponding segments of secondary structure in closely related proteins, such as myoglobin and the alpha and beta chains of haemoglobin, are similar in magnitude. Also we find that the moments estimated from the amino-acid sequence (μ_H) are reasonably close to those calculated from the actual structure (μ_{s1}). These points are illustrated for myoglobin and haemoglobin in fig. 1.

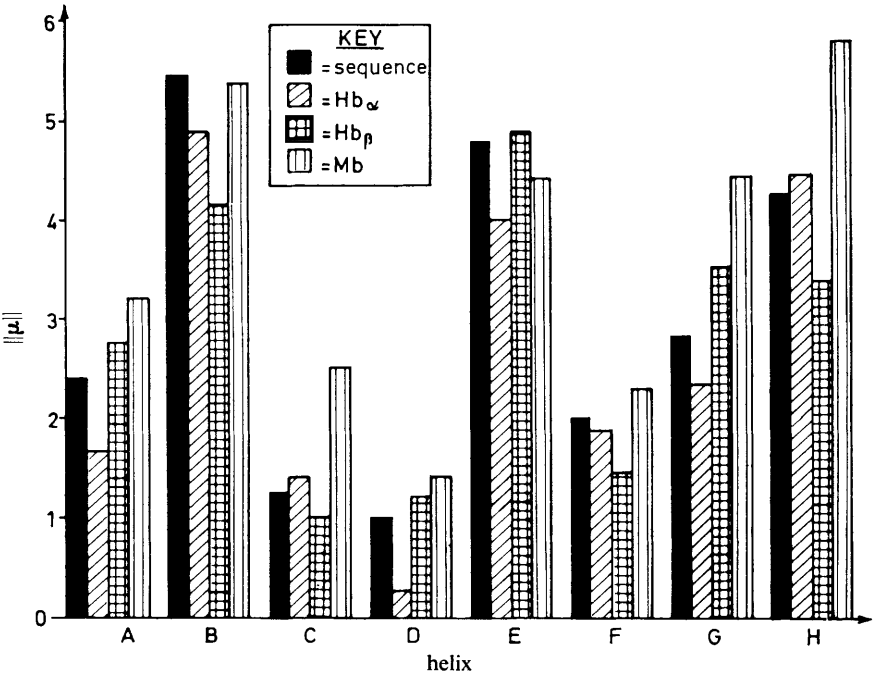

FIG. 1.—Comparison of hydrophobic dipole moments of the alpha helices of the myoglobin and haemoglobin polypeptide chains. The left-hand bar of the histogram gives the magnitude of the moment calculated from the amino-acid sequence by eqn (1). The right-hand three bars give the magnitudes of the structure moments computed by eqn (2) for both the alpha and beta chains of haemoglobin, and for the myoglobin chain.

HYDROPHOBIC INTERACTIONS OF SEGMENTS OF SECONDARY STRUCTURE IN PROTEINS

HYDROPHOBIC DIPOLE MOMENTS IN FOLDED PROTEINS

We have computed the hydrophobic dipole moments for each segment of secondary structure in several proteins, and have examined their magnitudes and directions. For each of the proteins listed in table 2, we have computed the hydrophobic dipole moment for each alpha helix, each strand of beta sheet, and each stretch of irregular backbone that separates any pair of regular structures or any regular segment from a

TABLE 2.—CANCELLATION OF THE HYDROPHOBIC DIPOLE MOMENTS OF SEGMENTS OF SECONDARY STRUCTURE (α HELICES, β STRANDS AND IRREGULAR REGIONS) IN FOLDED PROTEINS

Column 1 shows the root-mean-square sum of the lengths of the moments for each protein. This is the expected length for moments having random directions. Column 2 shows the length of the vector sum, the actual sum of moments.

protein	expected length of random directions	vector sum, the actual length
melittin	11.4	1.9
myoglobin	12.0	8.8
haemoglobin α	9.6	2.0
haemoglobin β	10.0	2.2
haemoglobin dimer (α + β)	13.8	2.6
triosephosphate isomerase	12.1	5.2
carboxypeptidase A	13.2	13.4
flavodoxin	10.1	8.6
lactate dehydrogenase	13.0	3.1
prealbumin monomer 1	7.7	0.1
rhodanese	14.6	3.0
thermolysin	12.7	1.8

terminus. The moments are calculated from the expression above for μ_{s1}. Atomic coordinates and definitions of the segments of secondary structure were taken from the compilation of the Brookhaven Protein Data Bank.[14]

A simple example of such a calculation is that for melittin [17] shown in fig. 2. Melittin consists of four 26 residue peptides, each a bent alpha helix. The hydrophobic dipole moment for each helix is large, and points towards the centre of the tetramer. This illustrates that the centre of the melittin tetramer is far more hydrophobic than the outside, and that the moments of the four polypeptide chains nearly cancel when they join to form a tetramer (see also the first entry in table 2).

A similar tendency for opposition of dipole moments on neighbouring alpha helices is observed in sperm whale myoglobin [16] (fig. 3). In this protein there are 8

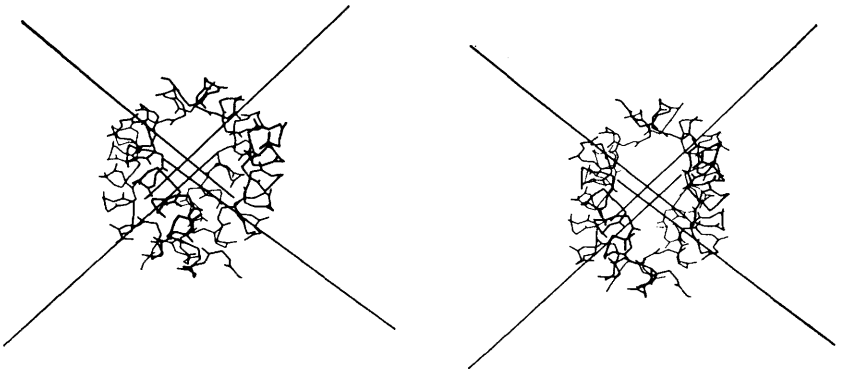

FIG. 2.—The melittin tetramer with the hydrophobic dipole moment (μ_{s1}) of each of the four alpha helical polypeptide chains. As in the following stereo pairs, the origin of each hydrophobic moment vector is placed at the centre of the corresponding segment of secondary structure, and the moment extends in the direction of greater hydrophobicity. Moments are illustrated with lengths (in Å) numerically six times their actual magnitudes (nominally in kcal).

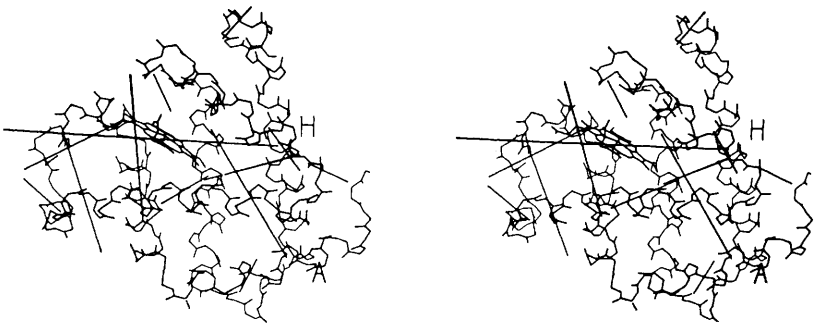

FIG. 3.—Myoglobin, with the hydrophobic dipole moments of each of the 8 helices and 7 irregular segments. The haeme group can be seen edge-on at the top centre of the molecule; helices A and H are labelled.

alpha helices and 7 irregular segments of polypeptide chain that separate them. Although the pattern of the 15 hydrophobic dipole moments is more complicated than for melittin, several points are clear: (1) moments tend to oppose moments from neighbouring segments of secondary structure; (2) moments from turns and other irregular regions are not negligible compared to those from helices, although the largest moments are associated with the long helices B, G and H; (3) the net direction of all 15 moments is toward the haeme pocket, where the hydrophobic edge of the haeme touches the protein.

In proteins of the alpha/beta classification of Richardson,[17] the beta sheet often has

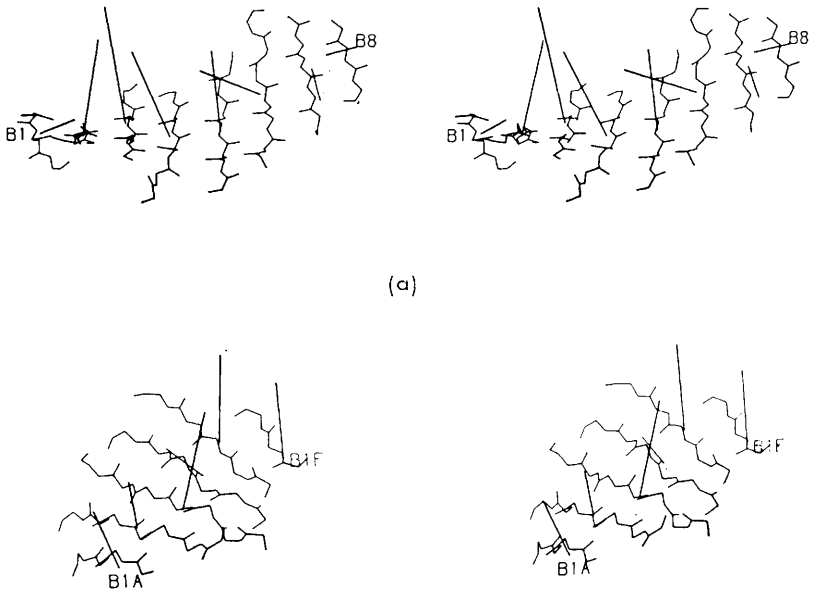

FIG. 4.—Hydrophobic moments of strands in two beta sheets. (*a*) 8 strands in carboxypeptidase A. (*b*) 6 stands in lactate dehydrogenase. In each figure two strands are labelled according to the nomenclature of the Brookhaven Protein Data Bank.

one polar side and one non-polar side, as judged by the directions of the hydrophobic moments of the individual strands. This is illustrated in fig. 4 for the beta sheets in carboxypeptidase A and lactate dehydrogenase. In the former sheet, six of eight strands have their moments pointing toward the same side, and in the latter, five of six strands have moments towards the same side.

This amphiphilicity of beta sheets is not observed in all such proteins, however. In flavodoxin,[16] for example, the beta sheet does not have a uniformly polar side. The hydrophobic dipole moments project from both side of the sheet, and are generally opposed by moments extending from neighbouring alpha helices. Similarly, in rhodanese the directions of the hydrophobic moments associated with the strands of the beta sheets are not regular; moreover the magnitudes of the moments are small.

Some protein structures contain beta sheets twisted into barrels, and we have examined the hydrophobic moments in two of these. One is prealbumin (fig. 5), which Richardson [17] classifies as an " antiparallel beta barrel ". Of the eight

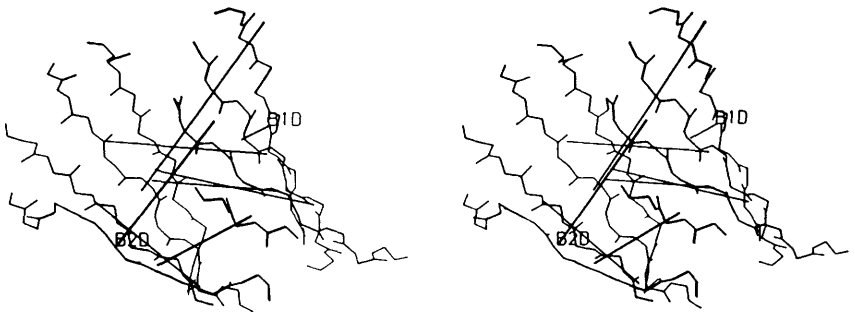

FIG. 5.—Hydrophobic moments of the 8 strands of beta sheet forming the beta barrel of prealbumin. The region between the sheets if hydrophobic; nearly all of the moments point into this region.

strands of the barrel, seven have hydrophobic moments that point toward the interior of the barrel, although one of the moments is small. In triose phosphate isomerase (fig. 6), the moments of the strands forming the barrel tend to point outwards, as though the interior is more polar than the outside. The outside is surrounded by alpha helices whose hydrophobic dipole moments tend to point in.

INTERPRETATION OE THE DIRECTIONS OF THE HYDROPHOBIC MOMENTS IN FOLDED GLOBULAR PROTEINS

The hydrophobic moments of segments of secondary structure at the surface of globular proteins tend to point inwards. This is illustrated in fig. 7, which shows for two proteins the angle between the hydrophobic moments of segments of secondary structure and the vector from the centre of the molecule to the segment. These angles are represented by their cosines, weighted by the magnitude of the moments, and plotted against the distance from the centre of the segment of secondary structure to the centre of the molecule. A negative value means that the hydrophobic moment vector points inwards, toward the centre. In myoglobin, the hydrophobic moments of 6 out of 8 alpha helices point inwards. Also 5 of 7 of the hydrophobic moments of the irregular segments between helices point inwards. In triose phosphate isomerase, the moments of most of the alpha helices point inwards. In contrast, the moments of the beta strands (which are more nearly toward the centre of the molecule) point outwards.

The subjective impression of directions of the hydrophobic moments in folded proteins is that they tend to cancel. Statistics presented in table 2 confirm this impression. If the directions and magnitudes of the hydrophobic moments of segments of secondary structure were truly random, we would expect that the length of the vector sum of the moments could be represented as the result of a random walk in three dimensions. The total expected length of the walk can be estimated as the root-mean-square sum of the lengths of the steps, in the case that the distribution of step

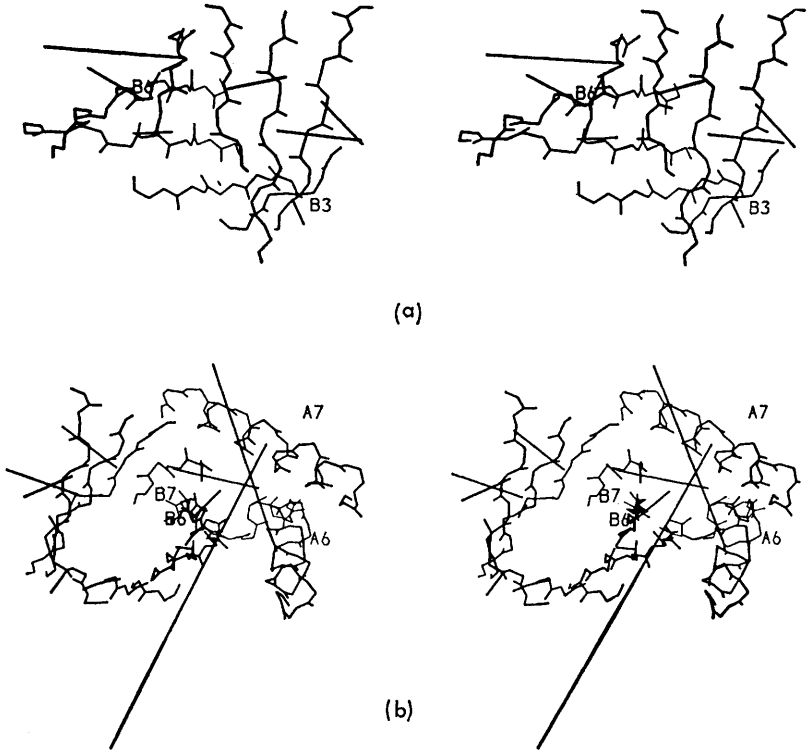

FIG. 6.—Hydrophobic moments in triose phosphate isomerase. (*a*) The beta barrel, showing that the moments of 6 of the 8 strands point outwards; (*b*) moments of both the beta and some of the alpha segments.

sizes is Gaussian, or the number of steps is very large.[15] We assume that one of these conditions holds approximately. The expected lengths are given in the first column of table 2, and the magnitude of the vector sum of the moments is given in the second column. For all proteins other than carboxypeptidase A the actual length is smaller than the expected length, and for all but three it is smaller than half the expected length. We conclude that the hydrophobic moments of elements of secondary structure tend to oppose each other in folded globular proteins.

While we have examined hydrophobic moments in only a few proteins, several patterns seem to be common. In clusters of alpha helices, moments of neighbouring helices tend to oppose each other. For beta sheets sandwiched between alpha helices, the moments of the helices tend to oppose the moments of strands in the sheets. In some cases, but not all, the strands within the sheet have moments that project mainly to one side of the sheet.

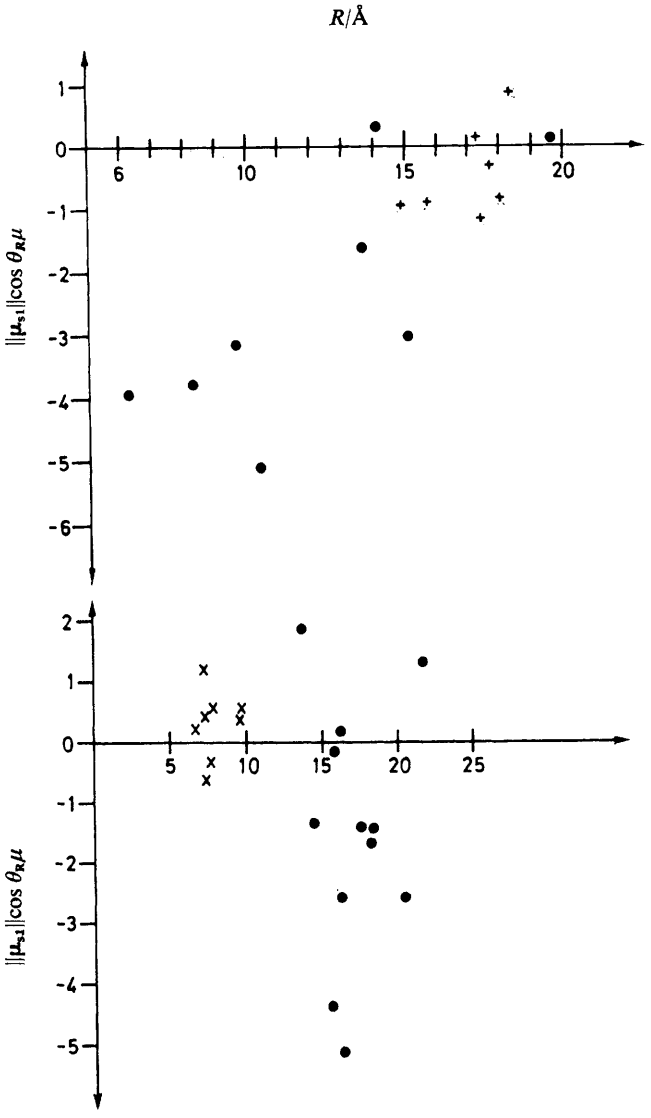

FIG. 7.—Orientation of hydrophobic dipole moments in myoglobin (top) and triose phosphate isomerase (bottom), plotted as a function of the distance R of the corresponding segment of secondary structure from the centre of the protein. Negative values mean that the dipole points towards the centre of the protein. ●, Helix; ×, β-strand; +, irregular.

INTERACTIONS OF HYDROPHOBIC MOMENTS WITH THE ENVIRONMENT

ORIENTATION OF A HYDROPHOBIC DIPOLE AT A SURFACE

It is possible to develop a very crude quantitative description for the energy of orientation of a hydrophobic dipole. We can think of a surface between a hydrophobic phase and a hydrophilic phase, such as the surface of a lipid membrane, as being described by a hydrophobicity function, $M(x)$, as shown in fig. 8. Because of

thermal motion of lipid and water molecules at the surface, we expect that, averaged over the time of diffusion, this surface is not sharp, but may more accurately be thought of as a smoothly varying gradient in the hydrophobicity. This diffusionally averaged structure is similar in conception to the D-averaged structure for a liquid.[18] The value of the hydrophobicity function $M(x)$ is defined as $+1$ for an aqueous environment and -1 for an entirely hydrophobic one. As discussed above, the

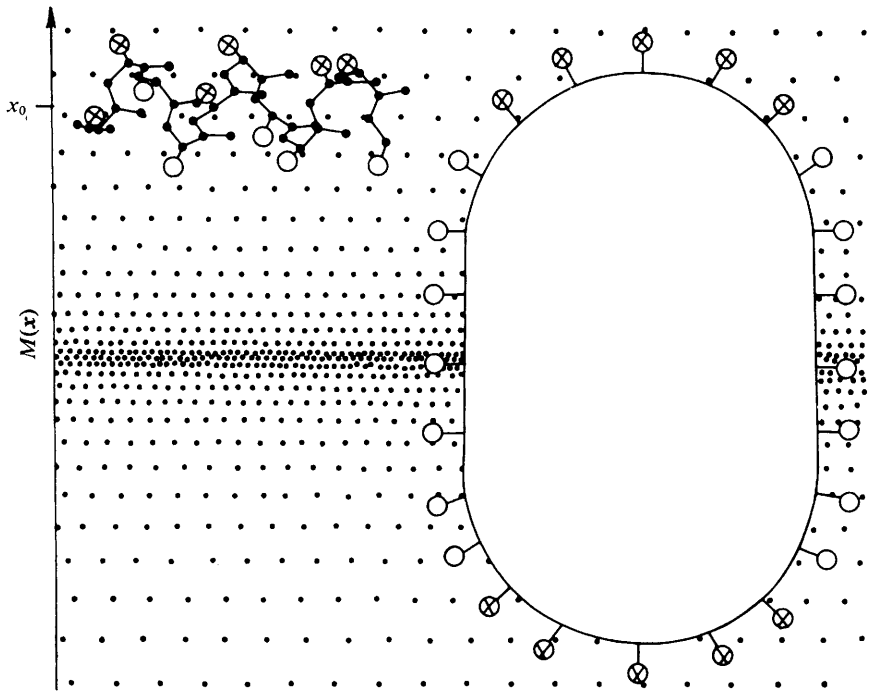

FIG. 8.—A highly schematic representation of the hydrophobicity function, $M(x)$, plotted against position perpendicular to a membrane. $M(x)$ is represented by the density of points: it is highest at the centre of the membrane, and falls off smoothly towards the aqueous interface at the top and bottom of the figure. At the top left is a surface-seeking helix, positioned so that its hydrophilic side chains (×) are mainly in contact with the aqueous solvent and its hydrophobic side chains (○) are mainly in contact with the more hydrophobic region. At the right is a schematic representation of a membrane protein with hydrophobic side chains (○) in contact with the more hydrophobic region.

" hydrophobicities " of the various amino acids are related to the free energies of transfer of their side chains from non-polar to polar environments. Therefore the product of the hydrophobicity of the environment $M(x)$ and the hydrophobicity of an amino-acid yields, on a relative scale, a very rough estimate of the free energy of the amino-acid side chain in this environment. Low free energies (negative on this scale) result from hydrophobic side chains in apolar environments, and from hydrophilic side chains in polar environments. Thus a rough estimate for the free energy of a fixed protein structure in an environment with hydrophobicity described by the function $M(x)$ is

$$G \approx \sum_i H_i M(x_i) \tag{4}$$

in which the summation is over all amino-acid side chains in the structure. This expression neglects contributions from main-chain atoms, and assumes that all side-

chain atoms are exposed to solvent. This assumption is not a good one in general, and it is clear that a more accurate estimate of the free energy could be obtained by summing only over side chains at least partially exposed to solvent. Alternatively, one could weight each amino-acid side chain by the fraction of its surface area exposed to solvent.

If the environment of a protein structure varies slowly and smoothly from apolar to polar, then we can rewrite eqn (4) in a simpler form. Let us denote the gradient in the hydrophobicity function as $\boldsymbol{F}(\boldsymbol{x})$, so that

$$\boldsymbol{F}(\boldsymbol{x}) = -\nabla M(\boldsymbol{x}). \tag{5}$$

If the hydrophobicity varies smoothly we can write that, near a point $\boldsymbol{x}_0$, the hydrophobicity function is given roughly by

$$M(\boldsymbol{x}) \approx M(\boldsymbol{x}_0) - (\boldsymbol{x} - \boldsymbol{x}_0) \cdot \boldsymbol{F}(\boldsymbol{x}_0). \tag{6}$$

Using eqn (6) and the definitions of the hydrophobic moment μ_{s2}, eqn (3) and the hydrophobicity $M(\boldsymbol{x})$, eqn (4) can then be rewritten as

$$G \approx \sum_i H_i M(\boldsymbol{x}_0) - \mu_{s2} \cdot \boldsymbol{F}(\boldsymbol{x}_0) \tag{7}$$

where $\boldsymbol{x}_0$ is the " centre " of the structure, defined as the mean position of the centres of the amino-acid side chains. This expression is analogous to that for the potential energy of a charged object in an electric field. In a slowly varying " hydrophobic field," $M(\boldsymbol{x})$, the free energy of a structure depends on the location of the " centre " of the structure ($\boldsymbol{x}_0$) as well as on its orientation (the direction of μ_{s2}). The free-energy difference between the state in which the hydrophobic dipole moment is oriented parallel to the hydrophobic field $\boldsymbol{F}(\boldsymbol{x})$ and the state in which they are anti-parallel is then

$$\Delta G \approx -2\|\mu_{s2}\| \cdot \|\boldsymbol{F}(\boldsymbol{x}_0)\|. \tag{8}$$

Two important assumptions used in this analysis are that (1) the solvent (water and lipid molecules) diffuses rapidly relative to the protein molecule so that the diffusionally averaged solvent structure is applicable and (2) the hydrophobicity function $M(\boldsymbol{x})$ varies in roughly linear fashion over the region in which the protein structure is located. Because the first assumption requires that the protein structure be not too small and the second requires that it be not too large, this analysis will not apply for all protein structures at a polar–apolar interface. We suspect that it is most useful in describing the interaction of a single alpha helical protein structure (such as melittin) with a lipid–water or air–water interface.

HYDROPHOBIC QUADRUPOLE MOMENTS

The hydrophobic quadrupole moments are tensor quantities, analogues of the electric quadrupole moments of a charge distribution.[19] Their simplest definition is

$$Q_{\alpha\beta} = \sum_i H_i r_{i\alpha} r_{i\beta}. \tag{9}$$

Thus $Q_{xx} = \sum_i H_i x_i^2$. A related definition which leads to values independent of the choice of origin is

$$Q_{\alpha\beta} = \sum H_i r_{i\alpha} r_{i\beta} - \langle H_i r_{i\alpha} \rangle \sum_i r_{i\beta} - \langle H_i r_{i\beta} \rangle \sum_i r_{i\alpha} + \langle H_i \rangle \sum_i r_{i\alpha} r_{i\beta}. \tag{10}$$

After diagonalization, it is clear from eqn (10) that a globular protein should have three negative quadrupole components (since it has predominantly hydrophilic side-

chains, with negative hydrophobicities, on its surface at the greatest distances from the origin). In contrast, membrane proteins, which are expected [20] to have hydrophobic residues inside and hydrophobic out, are likely to have at least some positive components of the hydrophobic quadrupole. A protein which spans a membrane would be expected to have one negative and two positive diagonal components.

We thank Walter Kauzmann for the suggestion to consider quadrupole moments, and we thank the N.S.F. and N.I.H. for support.

[1] W. Kauzmann, *Adv. Protein Chem.*, 1959, **14**, 1.
[2] R. A. Capaldi and G. Vanderkooi, *Proc. Natl Acad. Sci. USA*, 1972, **69**, 930.
[3] J. P. Segrest and R. J. Feldman, *J. Mol. Biol.*, 1974, **87**, 853.
[4] H. Meirovitch, S. Rackovsky and H. A. Scheraga, *Macromolecules*, 1980, **13**, 1398.
[5] J. Kyte and R. F. Doolittle, *J. Mol. Biol.*, 1982, **157**, 105.
[6] J. T. Edsall and H. A. McKenzie, *Adv. Biophys.*, 1982, in press.
[7] Y. Nozaki and C. Tanford, *J. Biol. Chem.*, 1971, **246**, 2211.
[8] R. Wolfenden, L. Anderson, P. M. Cullis and C. C. B. Southgate, *Biochemistry*, 1981, **20**, 849.
[9] C. Chothia, *J. Mol. Biol.*, 1976, **105**, 1.
[10] J. Janin, *Nature (London)*, 1979, **277**, 491.
[11] D. Eisenberg, R. M. Weiss and T. C. Terwilliger, *Nature (London)*, 1982, in press.
[12] G. von Heijne and C. Blomberg, *Eur. J. Biochem.*, 1979, **97**, 175.
[13] M. Schiffer and A. B. Edmundson, *Biophys. J.*, 1967, **7**, 121.
[14] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.*, 1977, **112**, 535.
[15] J. S. Richardson, *Adv. Protein Chem.*, 1981, **34**, 167.
[16] J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips and V. C. Shore, *Nature (London)*, 1960, **185**, 422.
[17] T. C. Terwilliger and D. Eisenberg, *J. Biol. Chem.*, 1982, **257**, 6010; 6016.
[18] R. M. Burnett, G. D. Darling, D. S. Kendall, M. E. LeQuesne, S. G. Mayhew, W. W. Smith and M. L. Ludwig, *J. Biol. Chem.*, 1974, **249**, 4383.
[19] M. N. Barber and B. W. Ninham, *Random and Restricted Walks* (Gordon and Breach, New York, 1970), pp. 5-15.
[20] D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (The Clarendon Press, Oxford, 1969).
[21] D. M. Engelman, R. Henderson, A. D. McLachlan and B. A. Wallace, *Proc. Natl Acad. Sci. USA*, 1980, **77**, 2023.