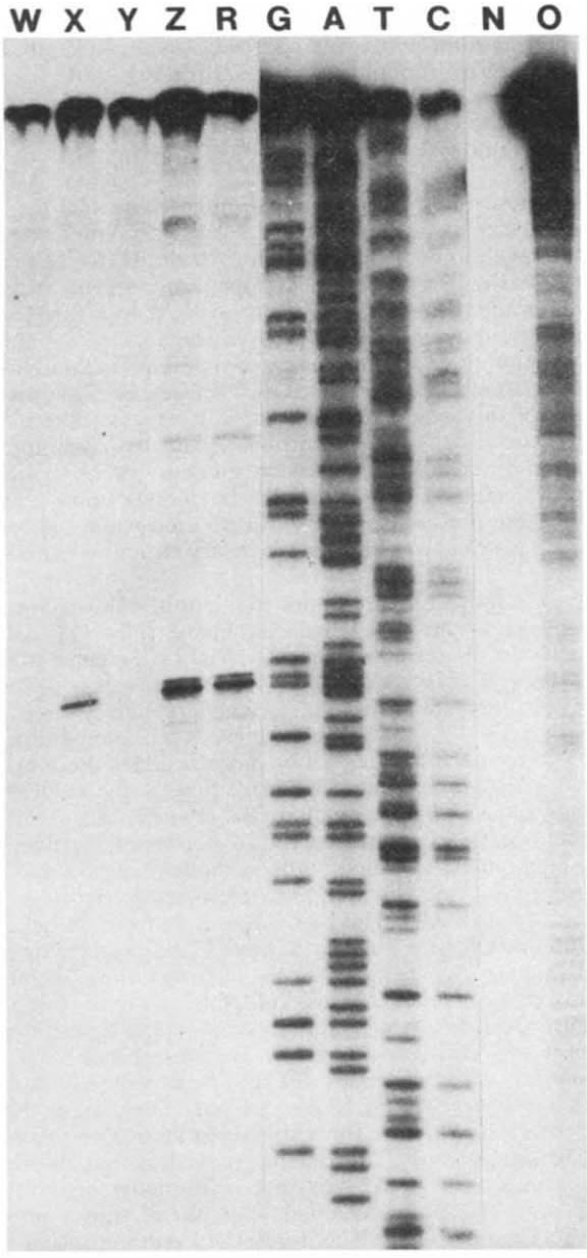

Fig. 4 Replication-uncoupled transcription. Probe 3 was used (gene 23). Lanes G, A, T, C, N, O and R are as described in Fig. 2. Lanes W, X, Y and Z are S_1 mapping experiments performed as described in Fig. 2 using RNA extracted from *E. coli* infected with the T4 mutant *tsL56 amH39X amN130*. Lanes W and X are from an infection of *E. coli* B^F at 30 °C with RNA extracted at 5 and 20 min after infection, respectively. Lanes Y and Z are from an infection of *E. coli* B^E at 43 °C with RNA extracted at 3 and 18 min after infection, respectively.

Transcriptional specificity in prokaryotes is conferred in large part by DNA sequences at −10 (Pribnow box) and −35 relative to the initiation site[15–17]. Temporal control of gene expression in the bacteriophage life cycle can be achieved by synthesis of a new RNA polymerase molecule with different specificity. For example, T7 directs the synthesis of an RNA polymerase which does not recognize the −35 and −10 regions of *E. coli* or T7 class I promoters, but instead recognizes a specific 23-base pair sequence[18]. Alternatively, phage-specified proteins can modify the existing RNA polymerase, altering its DNA sequence specificity. This phenomenon is well documented in *Bacillus subtilis*. For example, phage SP01 synthesizes gp 28 which displaces the σ subunit of the host RNA polymerase. This activates middle promoters which have −35 and −10 sequences different from those of *B. subtilis* promoters or SP01 early promoters[19]. T4 is similar to SP01 in that it modifies the host RNA polymerase rather than synthesizing a new enzyme. Like SP01 middle promoters, the DNA sequence which is well conserved among the T4 late transcripts examined here is different from the sequences recognized by the unmodified RNA polymerase. Although the first five bases of the conserved sequence are identical to the first five bases of the Pribnow box (TATAA), the sixth base of the Pribnow box is a 100 per cent conserved T (refs 15–17) while the 100 per cent conserved A found here is actually more similar to eukaryotic promoter sequences[20]. The consensus sequence is also longer and closer to the 5′ end of the mRNA than the Pribnow box recognized by the host enzyme and there is apparently no specific sequence recognized at −35. The change in transcriptional specificity that occurs at late times is thus not only a substitution of one set of specific DNA–protein interactions with another, but also a fundamental change in the geometry of the interactions.

We thank M. Parker, G. Kassavetis, T. Elliott, E. P. Geiduschek, M. Showe, G. Smith and H. Krisch for communicating results before publication, G. Stormo and L. Gold for access to T4 DNA sequences, M. Parker, A. H. Doermann and C. Furlong for help and technical advice, and the NIH for support (grant AI-09456).

Received 28 April; accepted 13 July 1982.

1. Rabussay, D. & Geiduschek, E. P. in *Comprehensive Virology* Vol. 8 (eds Fraenkel-Conrat, H. & Wagner, R. R.) 1–196 (Plenum, New York, 1977).
2. Snyder, L., Gold, L. & Kutter, E. M. *Proc. natn. Acad. Sci. USA* **69**, 603–607 (1972).
3. Wu, R., Geiduschek, E. P. & Cascino, A. *J. molec. Biol.* **96**, 539–562 (1975).
4. Young, E. T., Menard, R. C. & Harada, J. *J. Virol.* **40**, 790–799 (1981).
5. Kassavetis, G. A. & Geiduschek, E. P. *EMBO J.* **1**, 107–114 (1982).
6. Parker, M., Christensen, A. C., Young, E. T. & Doermann, A. H. (in preparation).
7. King, J. & Laemmli, U.K. *J. molec. Biol.* **62**, 465–477 (1971).
8. Oliver, D. B. & Crowther, R. A. *J. molec. Biol.* **153**, 545–568 (1981).
9. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
10. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C. & Smith, M. *Cold Spring Harb. Symp. quant. Biol.* **45**, 961–981 (1980).
11. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
12. Sollner-Webb, B. & Reeder, R. H. *Cell* **18**, 485–499 (1979).
13. Emrich-Owen, J., Schultz, D. W., Taylor, A. & Smith, G. R. *J. molec. Biol.* (submitted).
14. Lewis, M. K. & Burgess, R. R. *J. biol. Chem.* **255**, 4928–4936 (1980).
15. Pribnow, D. in *Biological Regulation and Development* Vol. 1 (ed. Goldberger, R. F.) 219–278 (Plenum, New York, 1979).
16. Rosenburg, M. & Court, D. *A. Rev. Genet.* **13**, 319–353 (1979).
17. Siebenlist, U., Simpson, R. B. & Gilbert, W. *Cell* **20**, 269–281 (1981).
18. Rosa, M. D. *Cell* **16**, 815–825 (1979).
19. Lee, G. & Pero, J. *J. molec. Biol.* **152**, 247–265 (1981).
20. Grosveld, G. C., deBoer, E., Shewmaker, C. K. & Flavell, R. A. *Nature* **295**, 120–126 (1982).
21. Mattson, T., Van Houwe, G., Bolle, A., Selzer, G. & Epstein, R. *Molec. gen. Genet.* **154**, 319–326 (1977).
22. Young, E. T. *et al. J. molec. Biol.* **183**, 423–445 (1980).
23. Völcker, T. A. & Showe, M. K. *Molec. gen. Genet.* **177**, 447–452 (1980).
24. Showe, M. K., Isobe, E. & Onorato, L. *J. molec. Biol.* **107**, 35–54 (1976).
25. Hagen, F. & Young, E. T. *J. Virol.* **26**, 793–804 (1978).
26. Krisch, H. *et al. Proc. natn. Acad. Sci. USA* (in the press).

The helical hydrophobic moment: a measure of the amphiphilicity of a helix

David Eisenberg, Robert M. Weiss & Thomas C. Terwilliger

Molecular Biology Institute and Department of Chemistry, University of California at Los Angeles, Los Angeles, California 90024, USA

The spatial distribution of the hydrophobic side chains in globular proteins is of considerable interest. It was recognized previously[1] that most of the α-helices of myoglobin and haemoglobin are amphiphilic; that is, one surface of each helix projects mainly hydrophilic side chains, while the opposite surface projects mainly hydrophobic side chains. To quantify the amphiphilicity of a helix, here we define the mean helical hydrophobic moment, $\langle\mu_H\rangle = |\sum_{i=1}^{N} \vec{H}_i|/N$, to be the mean vector sum of the hydrophobicities $\vec{H}_i$ of the side chains of a helix of N residues. The length of a vector $\vec{H}_i$ is the signed numerical hydrophobicity associated with the type of side chain, and its direction is determined by the orientation of the side chain about the helix axis. A large value of $\langle\mu_H\rangle$ means that the

0028-0836/82/380371—04$01.00

helix is amphiphilic perpendicular to its axis. We have classified α-helices by plotting their mean helical moment versus the mean hydrophobicity of their residues, and report that transmembrane helices, helices from globular proteins and helices which are believed to seek surfaces between aqueous and non-polar phases, cluster in different regions of such a plot. We suggest that this classification may be useful in identifying helical regions of proteins which bind to the surface of biological membranes. The concept of the hydrophobic moment can be generalized also to non-helical protein structures.

Schiffer and Edmundson[2] represented helix amphiphilicity by a two-dimensional 'helical wheel' diagram: a projection down the idealized helix axis shows side chains protruding from a circle every 100°. Non-polar residues were mainly on one side of the circle, polar and charged residues on the other. Helical wheels have since been used to represent other amphiphilic helices: for example, it has been noted that portions of apolipoprotein chains[3–5] and a synthetic melittin-like peptide[6] can be built into helices having one polar face and one non-polar face. The notion of amphiphilic helices has also been helpful in studying the folding of proteins (for example, see ref. 7).

The concept of amphiphilic helices can be quantified by combining a hydrophobicity scale with the helical wheel. Figure 1 shows portions of two helices viewed in projection down their axes. The hydrophobicity of each amino acid residue is represented by a vector directed radially from the projected helix centre to the idealized projected α-carbon position. The length of each vector is H_i, the hydrophobicity of the residue; as this is a signed quantity, hydrophobic residues on one face of the helix reinforce contributions of hydrophilic residues on the other face. For the calculations described in Figs 1 and 2, the values of H_i are those proposed by Janin[8] on the basis of observed distributions of each amino acid between the surface and interior of proteins. We also performed calculations with other sets of hydrophobicity values, including those proposed by Wolfenden *et al.*[9], von Heijne and Blomberg[10] and Chothia[11], as well as with our own set. The general features of the results described below do not depend on the hydrophobicity scale used.

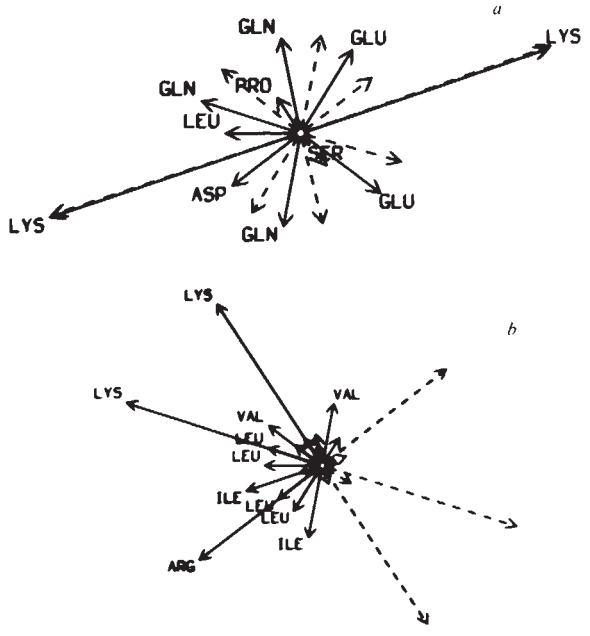

Fig. 1 Graphical representation of the contributions, by residue, to the helical hydrophobic moment, $\vec{\mu}_H$. *a*, The 11-residue stretch of a helix in lactate dehydrogenase starting at residue 308. Hydrophobic residues have positive values for the hydrophobicity, H_i, and are represented as vectors extending out from the centre of the axial projection of the α-helix. Hydrophilic residues have negative values of H_i and their positions about the helix axis are represented by dashed vectors extending from the centre. Their vector contributions are represented by solid lines 180° away. The vector sum of the $\vec{H}_i$ is $\vec{\mu}_H$. As the distribution of solid lines is relatively symmetric, this segment has a small hydrophobic moment, $\langle\mu_H\rangle = 0.10$. *b*, Residues 5–22 of melittin, the most amphiphilic 18-residue segment ($\langle\mu_H\rangle = 0.40$). Residues of small contribution are represented by vectors but are not labelled. The resulting hydrophobic moment would be off the page to the upper left.

We define the degree of amphiphilic character of a helix perpendicular to its axis by the helical hydrophobic moment, μ_H, the magnitude of the vector sum of the $\vec{H}_i$ for N residues of an α-helix: $|\vec{\mu}_H| = \sum_{i=1}^{N} \vec{H}_i$. To compare moments of helices of different lengths, it is convenient to work with the intensive form of the moment, $\langle\mu_H\rangle = |\vec{\mu}_H|/N$, which we will sometimes refer to simply as the hydrophobic moment. In much of our work, a stretch of 18 residues was chosen because an ideal α-helix of this length makes exactly five turns; thus the side chains are distributed uniformly about its circumference. However other lengths were also studied. Although the real vectors calculated from atomic coordinates (see below) can also be used, the present approach can be used when the tertiary structure is unknown, but the secondary structure can be predicted.

Figure 2 shows a hydrophobic moment plot in which $\langle\mu_H\rangle$ is plotted against the mean hydrophobicity, $\langle H\rangle = (\sum_{i=1}^{N} H_i)/N$, for helices of N residues, where N is always greater than 10. For helices of $\geqslant 19$ residues, the largest value of $\langle\mu_H\rangle$ for any 18-residue segment is also plotted with a related symbol. Thus, for each helix of 11–18 residues, there is a single point on the plot, and for each helix of 19 or more residues there are two points. The abscissa value of Fig. 2 reflects the solubility of each helix in a non-polar medium, the points falling to the right representing helices which prefer a non-polar medium to a polar medium. The ordinate reflects the tendency of a helix to assume a preferred orientation at an interface between polar and non-polar media. The plot shows data from 64 helices in 26 proteins. Of these helices, 28 have 19 or more residues (up to a maximum of 53 residues for the haemagglutinin membrane glycoprotein of influenza virus) and 36 have 11–18 residues.

Figure 2 shows that different types of helices cluster in different regions of the diagram. Transmembrane sequences are found in the lower right; this is to be expected because, as noted by others[12], such sequences have high mean hydrophobicities which render them soluble in lipid. They also have small helical hydrophobic moments, indicating that all orientations about the helix axis are essentially energetically equivalent. We have assumed that these transmembrane sequences are in α-helical conformations on the basis of Henderson's argument[13] that this conformation is strongly favoured in a membrane. The common length of 18–24 hydrophobic residues also suggests that these sequences bridge the apolar portion of the bilayer ($\sim$30 Å) in α-helical conformations (1.5 Å per residue along the helix axis).

In the left and lower central regions of Fig. 2 are clustered α-helical segments from globular proteins. These segments (Table 1) include every representative α-helix of known conformation having 18 or more residues, and every helix of 11–17 residues in the same proteins, after Feldman[14]. Of course, other regions of these polypeptide chains are generally not α-helical (and therefore are not plotted). Because these helices are from soluble proteins, it is not surprising that their mean hydrophobicities are smaller than those of membrane-penetrating sequences (that is, they are more hydrophilic). The helical hydrophobic moments for these helices vary substantially, but are often small.

In the upper right of Fig. 2 is a group of helices (represented by triangles and arrowheads) which have both high helical moments and large mean hydrophobicities. The main members of this group are 26-residue peptides known to be both lytic and surface-active. Two are melittins, the main protein components of venom from bees[15,16], another is a synthetic melittin-like peptide[6] and two others are δ-haemolysins, lytic peptides secreted by *Staphylococcus aureus* which are similar to melittin

Table 1 Helices included in the hydrophobic moment plot of Fig. 2

Protein	Organism/organ	1st residue	No. of residues	Evidence for helix*	Protein type	Ref. for sequence or structure
Adenylate kinase	Pig	69(Leu)	15	X ray	Globular	14
		123(Glu)	11			
		144(Glu)	21			
		179(Val)	16			
Alcohol dehydrogenase	Horse	170(Cys)	18	X ray	Globular	14
		202(Gly)	11			
		324(Ser)	13			
		353(Glu)	13			
Carboxy-peptidase A	Cow	14(Thr)	15	X ray	Globular	14
		112(Asn)	11			
		173(Glu)	15			
		215(Asp)	17			
		285(Gln)	22			
Chymotrypsin	Cow	230(Arg)	16	X ray	Globular	14
HA2 haemagglutinin	Influenza virus	77(Glu)	53	X ray	Globular	22, 23
Lactate dehydrogenase	Dogfish	33(Val)	12	X ray	Globular	14
		55(Met)	16			
		107(Glu)	13			
		120(Phe)	11			
		165(Cys)	17			
		249(Trp)	15			
		308(Lys)	22			
Myoglobin	Sperm whale	1(Val)	19	X ray	Globular	14
		20(Asp)	16			
		58(Ser)	20			
		100(Pro)	19			
		125(Ala)	24			
Myohaemerythrin	Marine worm	18(Tyr)	21	X ray	Globular	14
		40(Ser)	23			
		69(Glu)	19			
		93(Ala)	18			
Ribonuclease S	Cow	3(Thr)	11	X ray	Globular	14
		24(Asn)	11			
Thermolysin	*Bacillus thermo-proteolyticus*	67(Asp)	21	X ray	Globular	14
		137(Ile)	14			
		160(Glu)	20			
		235(Gly)	12			
		260(Arg)	15			
		281(Phe)	16			
		301(Gln)	12			
TMV coat protein	Tobacco mosaic virus	20(Pro)	13	X ray	Globular	24, 25
		38(Gln)	11			
		74(Ala)	15			
		114(Val)	21			
Triose phosphate isomerase	Chicken	17(Lys)	15	X ray	Globular	14
		44(Pro)	12			
		105(Ser)	16			
		138(Ile)	17			
		177(Thr)	20			
		213(Thr)	11			
Glycophorin	Human erythrocyte	12(Ile)	23	Length	Membrane	26
Glycoprotein	Vesicular stomatitis virus	51(Ser)	24	Length	Membrane	27
HA2 haemagglutinin	Influenza A/Victoria or A/Aichi	185(Trp)	24	Length	Membrane	28
HA2 haemagglutinin	Influenza A/Japan	527(Val)†	24	Length	Membrane	23
IgM	B lymphocyte	569(Asn)	26	Length	Membrane	29
Isomaltase	Small intestine	10(Ile)	22	Length	Membrane	30
M13 coat	M13 virus	20(Tyr)	18	Length	Membrane	31
M13 procoat	M13 virus	−20(Ser)	21	Length	Membrane	31
δ-Haemolysin	*S. aureus*	1(Met)	26	Analogy to melittin	Surface	17
	S. aureus (canine strain)	1(Met)‡	26		Surface	17
Melittin	*Apis mellifera*	1(Gly)	26	X ray	Surface	15, 19
	Apis florea	1(Gly)§	26	Analogy to above	Surface	16
Cytotoxic peptide with melittin-like activity	Synthetic	1(Leu)	26	CD	Surface	6
Diphtheria toxin peptide	Diphtheria toxin	7(Leu)	26	Prediction from sequence	Surface	20

* X ray means that an α-helical conformation has been established by X-ray crystallographic studies; CD, evidence for helicity comes from examination of circular dichroism; length, for a transmembrane protein, the length of the hydrophobic sequence is suitable for spanning the hydrophobic portion of a lipid bilayer as an α-helix.
† Corresponds to residue 185 in the enumeration of ref. 28.
‡ The sequence differs from that of the above strain of *S. aureus* in four residues.
§ The sequence differs from that of *A. mellifera* in five residues.

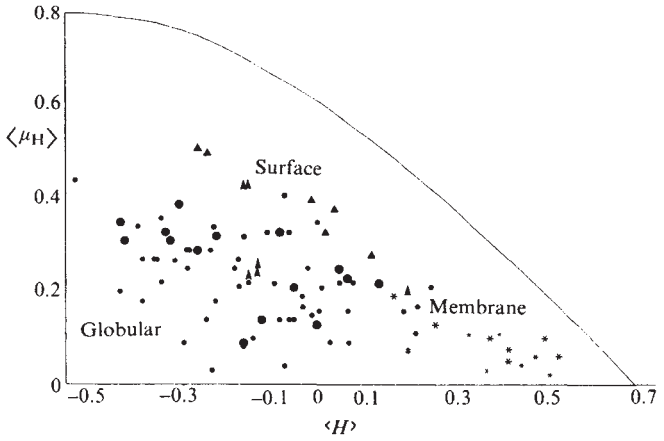

Fig. 2 A hydrophobic moment plot for 64 α-helical segments from the 26 proteins listed in Table 1. The abscissa gives the mean hydrophobicity of each segment and the ordinate gives the corresponding value of $\langle\mu_H\rangle$, as defined in the text. Helices from proteins having different functions plot in various regions of the diagram, as explained in the text. Membrane, membrane-penetrating helices: * represents the average of the entire helix (present for all helices) and ⁎ the 18-residue segment having the largest hydrophobic moment (present only for helices of ≥ 19 residues). Globular, helical segments from globular proteins: ●, average and ●, largest moment, Surface, surface-seeking helices: ▲ and ▲. The curve shows the largest possible value of $\langle\mu_H\rangle$ for each value of $\langle H\rangle$, as described in the text.

in both properties and sequence[17]. One of the melittins is known from X-ray diffraction studies to have an α-helical conformation[18,19]; the synthetic melittin is α-helical as judged by circular dichroism[6]. By analogy to these two peptides, we assume here that the δ-haemolysins are also α-helical. The final member of this group is a 26-residue peptide, fragment B from diphtheria toxin. This peptide has been predicted to have an α-helical conformation and is part of a larger CNBr peptide that induces conductance changes in lipid bilayers[20]. We include it in this group only to suggest that some peptides from larger proteins may have relatively large hydrophobic moments related to special functions.

Also shown in Fig. 2 is the maximum possible helical hydrophobic moment for each value of the hydrophobicity, represented by a curve which meets the abscissa at 0.70, the value for a hypothetical polyisoleucine α-helix. (Cysteine was excluded from the calculations of the maximum hydrophobic moment because the procedure used[8] to determine hydrophobicities may overestimate that of cysteine.) Note that even the highly amphiphilic δ-haemolysins are far less amphiphilic than some hypothetical helices.

The amphiphilic helical portions of melittin, as well as similar helical regions in other proteins, might be termed 'surface-seeking helices' because their large hydrophobic moments tend to align them at the surface between polar and non-polar phases[19]. Another class of proteins believed to function as amphiphilic helices is the apolipoproteins[3,4]; these have not been included here because even less direct information on their three-dimensional structures is available than for the melittins.

Although proteins of different functional types cluster in different regions of Fig. 2, these regions do not have absolute boundaries. This is expected as the various types of helices do not have entirely distinct functions. For example, the solubility of melittin in aqueous solution is believed[18] to be achieved in part by the highly charged C-terminal segment of the helix. This segment projects polar groups uniformly to all sides[19] and hence lowers the value of $\langle\mu_H\rangle$ for the entire melittin chain below that of the most amphiphilic 18-residue segment. For the two types of melittin and the synthetic melittin-like peptide, this results in arrowheads near the coordinates $\langle H\rangle = -0.15$, $\langle\mu_H\rangle = 0.25$. Other helical segments that plot in intermediate regions include some of those from the purple membrane protein model described by Engelman *et al.*[21]: these plot between the 'membrane' and 'surface' regions because of the charged and polar residues in the sequence (they are not included in Fig. 2). Such membrane helices with large hydrophobic moments may pair in the membrane, thereby reducing the net hydrophobic moment of the pair. More generally we suspect that segments from a helix that plot in a region of Fig. 2 occupied mainly by another type of helix may have a specialized function.

A potential limitation in the calculation of the helical hydrophobic moment $\langle\mu_H\rangle$ is that amino acid side chains deviate significantly from the idealized positions we have used (spaced 100° apart perpendicular to the helix axis). To clarify this point, we have investigated the properties of the structural hydrophobic moment, $\vec{\mu}_s$; the most convenient definition for this comparison is $\vec{\mu}_{s1} = \sum_{i=1}^{N} H_i\vec{s}_i$, where $\vec{s}_i$ is a unit vector pointing from the α-carbon atom of the ith residue to the centre of the residue side chain. This quantity can be calculated for any protein segment whose structure is known; when calculated for the idealized helices used in obtaining the mean hydrophobic moment, the angular deviation between the idealized $\vec{s}_i$ and the corresponding vectors perpendicular to the helix axis for a real helix averaged 30°. The difference in magnitude between the mean and structural moments, after scaling the latter by the length of the helix, averaged 25%. When computed for a helix, the structural hydrophobic moment $\vec{\mu}_{s1}$, like $\vec{\mu}_H$, emphasizes the amphiphilicity perpendicular to the helix axis. An alternative definition is $\vec{\mu}_{s2} = \sum_{i=1}^{N}(H_i\vec{R}_i - \langle H\rangle\vec{R}_i)$, in which $\vec{R}_i$ is a vector from any origin to the centre of the side chain of the ith residue. This definition of the hydrophobic moment represents amphiphilicity in directions both parallel and perpendicular to a helix axis and may have applications in energetics calculations.

We conclude that hydrophobic moment plots can reveal correlations between amino acid sequence and protein function; in particular, they can identify sequences particularly well suited for forming surface-seeking helices.

We thank L. Aha for discussions, and acknowledge support from the NSF (PCM 80-03725) and the NIH (GM 16925). T.C.T. was the recipient of a NSF graduate fellowship.

Received 25 February; accepted 24 June 1982.

1. Perutz, M. F., Kendrew, J. C. & Watson, H. C. *J. molec. Biol.* **13,** 669–678 (1965).
2. Schiffer, M. & Edmundson, A. B. *Biophys. J.* **7,** 121–135 (1967).
3. Morrisett, J. D., Jackson, R. L. & Gotto, A. M. Jr *Biochim. biophys. Acta* **472,** 93–133 (1977).
4. Segrest, J. P., Jackson, R. L., Morrisett, J. D. & Gotto, A. M. Jr *FEBS Lett.* **38,** 247–253 (1974).
5. Segrest, J. P. & Feldman, R. J. *Biopolymers* **16,** 2053–2065 (1977).
6. DeGrado, W. F., Kezdy, F. J. & Kaiser, E. T. *J. Am. chem. Soc.* **103,** 679–681 (1981).
7. Ptitsyn, O. B. & Rashin, A. A. *Biophys. Chem.* **3,** 1–20 (1975).
8. Janin, J. *Nature* **277,** 491–492 (1979).
9. Wolfenden, R., Andersson, L., Cullis, P. M. & Southgate, C. C. B. *Biochemistry* **20,** 849–855 (1981).
10. von Heijne, G. & Blomberg, C. *Eur. J. Biochem.* **97,** 175–181 (1979).
11. Chothia, C. *J. molec. Biol.* **105,** 1–14 (1976).
12. Segrest, J. P. & Feldman, R. J. *J. molec. Biol.* **87,** 853–858 (1974).
13. Henderson, R. *Soc. gen. Physiol.* **33,** 3–15 (1979).
14. Feldman, R. J. *Atlas of Macromolecular Structure on Microfiche* (Tracor Jitco, Rockville, Maryland, 1976).
15. Habermann, E. *Science* **177,** 314–322 (1972).
16. Kriel, G. *FEBS Lett.* **33,** 241–244 (1973).
17. Fitton, J. E., Dell, A. & Shaw, W. V. *FEBS Lett.* **115,** 209–212 (1980).
18. Terwilliger, T. C. & Eisenberg, D. *J. biol. Chem.* **257,** 6016–6022 (1982).
19. Terwilliger, T. C., Weissman, L. & Eisenberg, D. *Biophys. J.* **37,** 353–361 (1982).
20. Kayser, G. *et al. Biochem. biophys. Res. Commun.* **99,** 358–363 (1981).
21. Engelman, D. M., Henderson, R., McLachlan, A. D. & Wallace, B. A. *Proc. natn. Acad. Sci. U.S.A.* **77,** 2023–2027 (1980).
22. Wilson, I. A., Wiley, D. C. & Skehel, J. J. *Nature* **289,** 366–373 (1981).
23. Porter, A. G. *et al. Nature* **282,** 471–477 (1979).
24. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. *Nature* **276,** 362–368 (1978).
25. Dayhoff, M. O. *Atlas of Protein Sequence and Structure* Vol. 5, D283 (National Biomedical Research Foundation, Washington DC, 1972).
26. Furthmayr, H., Galardy, R. E., Tomita, M. & Marchesi, V. T. *Archs Biochem. Biophys.* **185,** 21–29 (1978).
27. Rose, J. K., Welch, W. J., Sefton, B. M., Esch, F. S. & Ling, N. C. *Proc. natn. Acad. Sci. U.S.A.* **77,** 3884–3888 (1980).
28. Verhoeyen, M. *et al. Nature* **286,** 771–776 (1980).
29. Rogers, J. *et al. Cell* **20,** 303–312 (1980).
30. Frank, G. *et al. FEBS Lett.* **96,** 183–188 (1978).
31. Wickner, W. *Science* **210,** 861–868 (1980).