Pergamon

Bioorganic & Medicinal Chemistry 7 (1999) 1–7

BIOORGANIC & MEDICINAL CHEMISTRY

Combining Hydrophobicity and Helicity: A Novel Approach to Membrane Protein Structure Prediction

Li-Ping Liu and Charles M. Deber*

Division of Structural Biology and Biochemistry, Research Institute, Hospital for Sick Children, Toronto, Ontario, M5G 1X8 Canada and Department of Biochemistry, University of Toronto, Ontario, M5G 1X8 Canada

Received 19 May 1998; accepted 27 August 1998

Abstract—In spite of the overwhelming numbers and critical biological functions of membrane proteins, only a few have been characterized by high-resolution structural techniques. From the structures that are known, it is seen that their transmembrane (TM) segments tend to fold most often into α-helices. To evaluate systematically the features of these TM segments, we have taken two approaches: (1) using the experimentally-measured residence behavior of specifically designed hydrophobic peptides in RP-HPLC, a scale was derived based directly on the properties of individual amino acids incorporated into membrane-interactive helices; and (2) the relative α-helical propensity of each of the 20 amino acids was measured in the organic non-polar environment of *n*-butanol. By combining the resulting hydrophobicity and helical propensity data, in conjunction with consideration of the 'threshold hydrophobicity' required for spontaneous membrane integration of protein segments, an approach was developed for prediction of TM segments wherein each must fulfill the dual requirements of hydrophobicity and helicity. Evaluated against the available high-resolution structural data on membrane proteins, the present combining method is shown to provide accurate predictions for the locations of TM helices. In contrast, no segment in soluble proteins was predicted as a 'TM helix'.

Introduction

Genome sequencing projects are generating thousands of sequences of new proteins from diverse organisms, of which a surprisingly high number contain membrane-embedded domains. For example, bacterium *Haemophilus influenzae*, the first complete genome sequence, contains only one type of membrane, but has devoted 140 (30%) of its 482 genes to encoding membrane-inserted proteins.[1] Membrane proteins such as receptors,[2–4] pore-forming proteins,[5–7] ion pumps and metabolite transporters,[8–11] and photosynthetic proteins,[12–14] are absolutely essential for the cell communicating with the outside world. As genome projects progress, one can expect that the identification of new membrane proteins and their corresponding primary sequences will increase exponentially.

In spite of the overwhelming numbers and critical biological functions of membrane proteins, little structural information is available for this key group of proteins. Compared to the >2000 non-membrane protein structures that have been solved, only a handful of membrane proteins (ten or so) have been characterized by high-resolution structural techniques (see ref 15 for a list). From the structures that are known, it is found that the transmembrane (TM) segments of these proteins tend to fold in relatively simple ways. There are two major categories through which a protein can adapt its structure to the requirements of a lipid environment: by forming apolar α-helices or by forming β-sheet 'barrels' with an apolar external surface. In both cases, all backbone hydrogen bonds are internally satisfied and polar surfaces are shielded from contact with lipids. Proteins with TM α-helices are found in the plasma membrane of both prokaryotes and eukaryotes,[14,16,17] whereas the β-type structures appear to be associated exclusively with the outer membranes of bacteria and mitochondria.[10,18] The toxin α-hemolysin—produced by a pathogenic bacterium—was found to penetrate into the lipid bilayer of its host via the formation of β-barrels.[7]

Many algorithms toward prediction of membrane protein structure and topology have been developed over the past few years. In assessing the origins of these approaches, presently available algorithms can be roughly classified into the following two groups:

(1) *Hydropathy analysis*, which is based on the recognition of the existence of long hydrophobic segments in membrane proteins. Typically, such an analysis will

Key words: membrane proteins; transmembrane segments; predictive analysis; hydrophobicity; helix propensity.
*Corresponding author. Tel.: +1 (416) 813-5924; fax: +1 (416) 813-5005; E-mail: deber@sickkids.on.ca

0968-0896/99/$ - see front matter
PII: S0968-0896(98)00233-8

scan the protein sequence and locate segments of suitable length with demonstrably high average hydrophobicity that can be considered as 'TM' segments. One of the most widely used is that of Kyte and Doolittle,[19] where mean residue hydrophobicity values are calculated for consecutive 19-residue sequence spans; segments with hydrophobicity above an average value are predicted to be membrane-spanning. A similar approach is adopted by Engelman et al., but using a different hydropathy scale.[20] To refine these predictive approaches, many subsequent hydrophobic scales for amino acid residues have been proposed [for a review see ref 21], in which correlations amongst the various amino acid hydrophobicity scales were examined, and the accuracies of various predictions compared. A trapezoidal sliding window was used by von Heijne in his hydrophobic analysis of TM sequences in conjunction with a consideration of positively-charged resides interior to the membrane (the 'positive-inside rule'),[22,23] which correctly predicted the topology for 23 out of 24 bacterial inner membrane proteins. These rules were also applied to a number of eukaryotic membrane proteins. Clearly, the ultimate hydrophobicity analysis method would be able to discriminate 'perfectly' between TM and non-TM segments in any protein chain.

(2) *Conformational preference analysis*, which is based mainly on the conformational preference(s) of individual amino acids in a membrane protein sequence. Assuming that soluble protein structures contain statistical information about folding of hydrophobic domains, Juretic et al. extracted a set of simple preference functions from the database of such proteins and applied this information to the prediction of folding motifs in the membrane environment.[24] Persson et al. calculated the residue distribution in a TM base[25] extracted from Swissprot database. Two sets of propensity values were established: one for the 21-residue hydrophobic region of TM segments, and one for the flanking residues at both termini corresponding to those amino acids likely to interact with the phosphate groups of the membrane lipids. An algorithm was then developed to predict TM helices based upon multiple sequence alignments (see refs 26 and 27 for details). Jones et al. used a dynamic programming algorithm to recognize the secondary structure and topology of integral membrane proteins based on the statistical data derived from well-characterized membrane proteins.[28] Compared to hydropathy analysis, these latter prediction approaches tend to provide a better understanding of membrane protein structure. However, the complexity of these algorithms has restricted their widespread use.

To aid in elucidation of membrane protein topology, particularly for proteins where structural information is not yet available, hydrophobicity scales are of great value. Indeed, such scales might be numerically among the most abundant ever developed in protein science, with 82 hydropathy scales collected in a review by Nakai et al. in 1988.[29] The proliferation of hydrophobicity scales appears to derive principally from the diverse definitions of hydrophobicity, as well as the various determination methods. The very fact that so many reports exist attests to the challenge of specifying protein transmembrane segments in the absence of hard experimental data.

Previous studies have suggested that the hydrophobic interaction between the peptide hydrophobic face with the hydrophobic stationary phase of reversed-phase high performance liquid chromatography (RP-HPLC) is a reasonable mimic for the interaction of peptides with membrane environments (Fig. 1).[30] In the present work, through evaluating the residence behavior of hydrophobic peptides in RP-HPLC, we have derived a scale that appears to be the first experimental scale based directly on the properties of individual amino acids incorporated into membrane-interactive helices. At the same time, we have determined experimentally the relative α-helical propensity of each of the 20 amino acids as measured in a non-polar environment. Through combining the resulting hydrophobicity and helical propensity data, we have developed a novel approach for prediction of transmembrane segments, viz. each TM segment must fulfill the dual requirements of hydrophobicity and helicity. Evaluated against the available high-resolution structural data on membrane proteins, the present combining method is shown to provide accurate predictions for the location of TM helices.

Results and Discussion

Peptide design and synthesis

De novo designed peptides with prototypical sequence Lys-Lys-Ala-Ala-Ala-**X**-Ala-Ala-Ala-Ala-Ala-**X**-Ala-Ala-Trp-Ala-Ala-**X**-Ala-Ala-Ala-Lys-Lys-Lys-Lys-amide were synthesized by continuous-flow Fmoc solid-phase chemistry, where X residues were substituted with each of the 20 commonly-occurring amino acids.[31] The rationale for the peptide design is given as follows: (i) the hydrophobic segment of peptide is comprised of 19 amino acids, which when folded into an α-helical conformation provides a hydrophobic core of sufficient length to span a phospholipid bilayer;[32] (ii) distributions of the three guest residues 'X' have been designed to preserve both angular and longitudinal symmetry around the helix, thereby minimizing any bias from

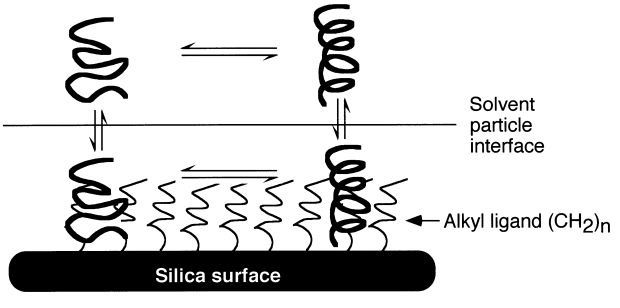

Figure 1. Schematic representation of a peptide in equilibrium between coil and helical conformations, upon interaction with the hydrophobic surface of a stationary nonpolar phase in HPLC experiments.

amphipathic character that may arise when X is substituted by polar and charged residues. This will guarantee that all peptides approach the membrane environment in an equivalent manner. In addition, triple substitutions of guest residue 'X' in the hydrophobic core amplify the effect of varying replacements, which enhances the reliability of the spectroscopic measurements; (iii) Ala, the most appropriate background residue as demonstrated by previous studies,[33] was chosen as the 'host' residue; (iv) a Trp residue was incorporated into the hydrophobic segment as a fluorescence probe to monitor characteristics of the local microenvironment; and (v) Lys residues were added at N- and C-termini to enhance the water solubility of peptides. Further, a cluster of Lys residues should be efficient in preventing peptide–peptide aggregation through the resulting high local positive charges. For the same purpose, the C-terminus was amidated to eliminate potential electrostatic attractions that might occur inter- or intra-molecularly. Amino acid analysis and mass spectrometry were used to characterize synthetic peptides, which verified the success of synthesis and purification.

HPLC-Derived hydrophobicity scale for amino acid residues

The retention time of a peptide interacting with an HPLC hydrocarbon phase is based on the competing attraction of the peptide molecule and the solvent for the nonpolar stationary phase. As peptides pass through the column, the hydrophobic examples interact more strongly with the hydrophobic stationary phase, to produce a longer retention time; oppositely, the hydrophilic peptides' retention times are relatively shorter. If we assume that the overall hydrophobicity of a given peptide is directly related to the summed hydrophobicity/hydrophilicity of the amino acid residues in that peptide, the measured retention times of each peptide can be converted to a relative hydropathy scale corresponding to the substituted 'X' residue. The following equation was used in the data conversion:

$$H = 10 \times (\Delta tR_{X-Lys}/\Delta tR_{Phe-Lys}) - 5.00$$

where ΔtR_{X-Lys} is the retention time difference between the 'X' peptide and the most hydrophilic 'Lys' peptide, and $\Delta tR_{Phe-Lys}$ is the retention time difference between the most hydrophobic 'Phe' peptide and the least hydrophobic 'Lys' peptide. Through this calculation, Phe was assigned a value of +5, and Lys a value of −5. The remaining residues were scaled proportionally (the resulting values are listed in Table 7 of ref 38).

Intrinsic α-helical propensity of individual amino acids in a nonpolar environment

Although all amino acid residues in TM segments take the helical conformation, intuitively it appeared to us that they were unlikely to contribute equally to helix formation/stability in the membrane environment. Organic solvents with dielectric constants between pure water and the hydrocarbon interior of biological membranes have been widely used to mimic the nonpolar environments of membranes.[34–37] In addition, studies in organic phases are free of complications from water/membrane transfer effects and lipid head group interactions. Accordingly, the intrinsic conformational preference (α-helical) for individual amino acids in membrane environments can be approached by dissolving them in *n*-butanol, a moderate nonpolar solvent with dielectric constant 17.8 at 25 °C. All 20 peptides formed predominantly α-helical conformations in *n*-butanol (Table 1). CD spectra were found to be independent of peptide concentration over the range 7.5 μM to 120 μM,[38] and thus the helicity differences observed among the peptides are not caused by their differential solubility, but by their intrinsic propensities to form the α-helical conformation in a nonpolar environment. The range of helicity likely has its origin in the intrinsic ability of the nonpolar solvent medium to solvate a given side chain; more polar residues may engage in side chain–backbone interactions to varying extents, which in turn may perturb the α-helical structure. Since the only difference among these peptides is the substitution of 'X' residues, it is reasonable to assume that variations in peptide helicity is attributed to the 'X' amino acid residue.

Previous CD analyses with membrane proteins have shown that for a full transmembrane α-helix, the intensities of negative bands at 208 nm and 222 nm are as high as −50,000 and −60,000 deg cm^2/dmol, respectively.[39] Accordingly, for a protein segment to adopt 50% α-helical conformation in a nonpolar (np) environment, the magnitude at 222 nm can be approximated by −30,000 deg cm^2/dmol. If we assume that 50% α-helix corresponds to a helical preference defined in *n*-butanol as $P_\alpha(np) = 1$, then the experimentally-determined helicity of each peptide can be converted to

Table 1 Helicity of model peptides in *n*-butanol and residue helical preference values.

Peptide	$\theta_{222\,nm} \times 10^{-4}$ (deg cm^2/dmol)	Pα(np)[a]
I	3.88	1.29
L	3.84	1.28
V	3.82	1.27
F	3.78	1.26
A	3.72	1.24
M	3.67	1.22
G	3.44	1.15
Y	3.33	1.11
T	3.27	1.09
W	3.20	1.07
S	2.99	1.00
H	2.90	0.97
Q	2.87	0.96
R	2.84	0.95
N	2.82	0.94
D	2.66	0.89
K	2.65	0.88
E	2.54	0.85
C[b]	3.28	0.79
P	1.70	0.57

[a]Helical preference parameter in a non-polar (np) environment.
[b]For the Cys peptide, only the middle 'X' residue was substituted by Cys, while the other two 'X' were replaced by Leu, and the Pα(np) of Cys was calibrated proportionately.

the conformational preference of each amino acid in membrane-mimetic environments. The resulting P_α(np) values for each of the 20 commonly-occurring amino acids are listed in Table 1. Values >1 indicate a preference for helicity in a membrane environment, while for values <1, an avoidance is implied.

Identification of TM segments by combining hydropathy and intrinsic α-helical preference

Hydropathy analysis has been used extensively as a standard approach to predict TM segments from primary sequence in the absence of experimental data. However, there is a recurrent phenomenon in previous prediction methods, that is, Kyte and Doolittle found their most hydrophobic 19-residue segment in the water-soluble lactate dehydrogenase.[19] To effectively discriminate between membrane-spanning and non-membrane-spanning segments, in present approach we introduce an experimentally-determined 'threshold hydrophobicity' as the cutoff value. The definition of 'threshold hydrophobicity' is based on the observation that there is a minimum hydrophobicity requirement for peptide incorporation into the membrane-mimetic environment;[40] above this threshold, peptide segments spontaneously integrate into membrane environment. The numerical value of the threshold is determined, in turn, from the HPLC-derived hydropathy parameters. The general applicability of this threshold hypothesis has been assessed by statistical analyses with databases derived from TM segments and non-TM helices[38,41] where it is found that parameters for membrane insertion established by model peptides can be applied with 96–97% success to natural TM segments contained in both single- and multi-spanning membrane proteins.

While one is able to identify highly hydrophobic segments with the potential to insert into membranes from threshold hydrophobicity data alone, essential structural information as to how the sequence becomes folded in membranes is not yet considered. Since the TM helices are basic building blocks for the majority of known membrane proteins, we can use the helical propensity scale developed above for each of 20 amino acids to perform a 'helical preference analysis' as a complementary approach to the hydropathy analysis (i.e. those segments of ca. 20 residues above the arbitrarily-chosen P_α(np) value of 1.10) (see Materials and Methods for details). Ideally, then, a TM α-helix should jointly satisfy the two criteria: hydrophobicity and helicity. The advent of several high-resolution structures of membrane proteins provides the opportunity to evaluate the accuracy of this combined predictive approach. For this purpose, we undertook 'predictions' of the actual TM helices (total = 78) in those membrane proteins for which high-resolution structural information is available, viz. bacteriorhodopsin,[16,17] the photosynthetic reaction center,[14] the light harvesting complex,[42,43] photosystem I,[44] cytochrome C oxidase[45–47] and the potassium ion channel[48] (Table 2). The combining predictive analysis is illustrated in Figure 2 for cytochrome C oxidase.[46] When Figure 2(a) (HPLC scale, using a 'threshold' of 0.4) and 2(b) [P_α(np) helix propensity values] are superimposed (using a window of 19 residues), the overlapping regions shown in Figure 2(c) are predicted as TM helices. The segments identified in cytochrome C oxidase as TM helices by X-ray crystallography versus those predicted by the present combining approach are delineated in Table 3. Similar analyses performed on the potassium ion channel from *Streptomyces lividans*[48] are shown in Figure 3. Using the method proposed by Ponnuswamy & Gromiha,[49] the accuracy of a theoretical prediction can be quantified by the equation:

$$\text{Accuracy (\%)} = [N_t - (N_o + N_u)]/N_t \times 100$$

where N_t, N_o, and N_u are, respectively, the total number of residues experimentally observed, the number of residues over-predicted, and the number of residues underpredicted or missed in the TM domain upon comparison to the experimentally-obtained structural data. Using this equation to analyze the data in Table 3, the present combined approach correctly predicts ca. 80% of overall cytochrome C oxidase TM residues. For all 78 TM segments contained in the proteins listed in

Table 2 Membrane proteins with known high-resolution structures

No.	PDB code	Protein name	Resolution (Å)	Refs.
1	2BRD	Bacteriorhodopsin	3.5	16, 17
2	1AIG	Photosynthetic reaction center	2.6	14
3	1LGH	Light harvesting complex	2.4	42, 43
4	2PSS	Photosystem I	4.0	44
5	1OCC	Cytochrome C oxidase	2.8	45–47
6	n/a	Potassium ion channel	3.4	48

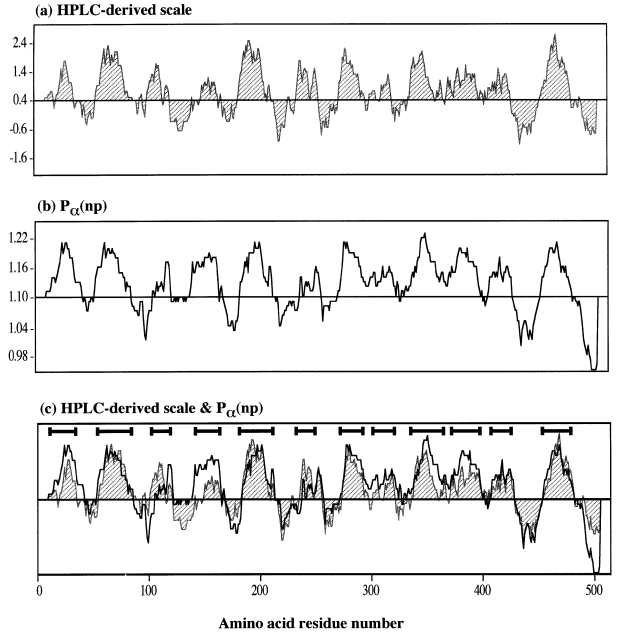

Figure 2. The sequence of cytochrome C oxidase[46] examined using (a) the HPLC-derived hydropathy scale;[38] and (b) helical propensity data derived from Pα(np) values (Table 1). In panel (c), the HPLC data from panel (a) are superimposed with the helical propensity data Pα(np) from panel (b). The solid lines indicate regions that are predicted as TM helices.

Table 3 Comparison of experimentally-determined TM helices with predicted TM helices in cytochrome C oxidase

Protein	TM segments (experimentally-determined)	TM segments (predicted, hydropathy & helicity)
Cytochrome C Oxidase	12–41	11–43
(A chain, I)	51–87	54–86
	95–116	95–121
	141–170	142–168
	183–212	183–213
	228–262	228–257
	270–285	272–297
	299–327	299–325
	336–359	335–364
	371–401	374–400
	407–433	407–430
	445–478	453–482
(B chain, II)	15–45	24–50
	60–87	62–83
(C chain, III)	16–37	22–40
	41–66	41–62
	73–106	76–105
	129–153	129–148
	156–183	157–182
	191–223	191–223
	233–255	236–252
(D chain, IV)	77–102	79–100
(G chain, VI-A)	13–37	11–36
(I chain, VI-C)	12–53	15–42
(J chain, VII-A)	26–54	27–50
(K chain, VII-B)	9–35	17–34
(L chain, VII-C)	18–44	19–40
(M chain, VIII)	12–35	18–35

Table 2, the combining approach predicts ca. 77% of overall TM residues (not shown); no actual TM segments are missed, but the fact that entry/exit points of TM segments did not always coincide lowered the final % accuracy level. Prediction based solely on hydropathy or solely on helical preference gave accuracies of 67% and 66%, respectively (not shown).

As controls, we randomly chose the following soluble proteins from the protein data bank: carboxypeptidase A,[50] thioredoxin[51] *staphylococcal nuclease*,[52] hydrolase,[53] cytochrome C peroxidase,[54] lysozyme,[55] and hemoglobin[56] (Table 4), and performed the combined analysis as described above. No segment in any of the above soluble proteins fulfilled the dual requirements of hydrophobicity and helicity. Results for carboxypeptidase A are graphed in Figure 4, where it is seen that no overlap of hydrophobicity data Fig. 4(b) and helicity data Fig. 4(c) occurs above the threshold levels. The Kyte–Doolittle hydropathy plot is shown for comparison Fig. 4(a).

Conclusion

Accurate prediction of membrane protein folding is a challenging task for structural biologists. For soluble (globular) proteins, secondary structure predictions are widely applied whenever a new sequence is determined, and many predictive algorithms have been developed on the basis of structural preferences (α-helix, β-sheet, β-turn, random) of the 20 commonly-occurring amino

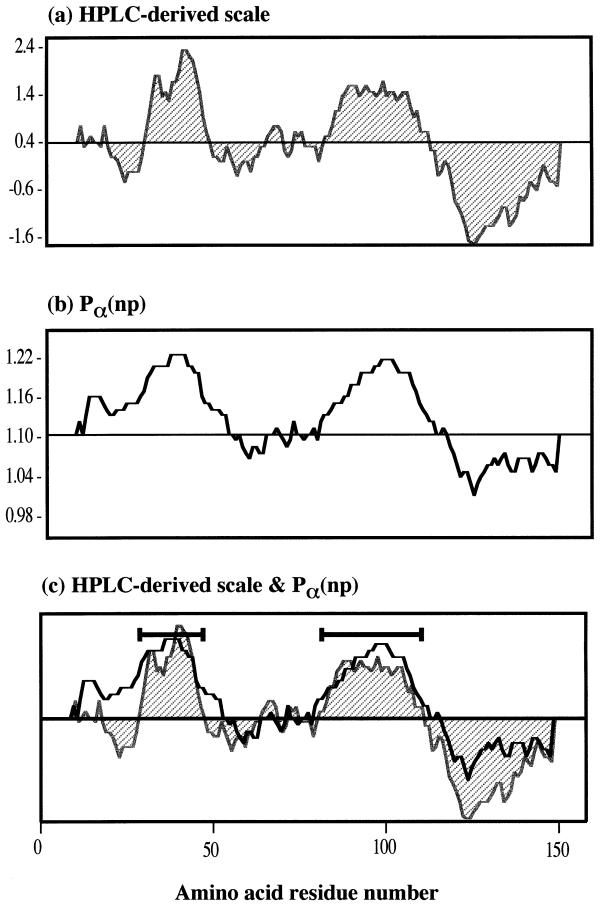

Figure 3. The sequence of the potassium ion channel[48] examined using the HPLC-derived hydropathy scale and the helical propensity values of Pα(np). In panel (c), data from the HPLC scale (panel (a)) are superimposed with Pα(np) data (panel (b)). The solid lines indicate regions that are predicted as TM helices.

acids. However, for membrane proteins, use of prediction strategies derived from globular proteins eventually becomes inappropriate,[57] as the low dielectric environment of the membrane profoundly affects protein folding preferences.

To distinguish reliably between TM and non-TM segments in any given protein, we have introduced two additional factors into previous hydropathy analysis: (1) a quantitative 'threshold hydrophobicity' for the spontaneous insertion of polypeptide segments into membranes used as cut-off value in hydrophobicity analysis; and (2) a requirement for helical preference in the non-polar environment. Predictive analyses presented

Table 4 Selected soluble proteins

No.	PDB code	Protein name	Resolution (Å)	Refs.
1	5CPA	Carboxypeptidase A	1.54	50
2	2TRX	Thioredoxin	1.68	51
3	1STN	*Staphylococcal* nuclease	1.70	52
4	1FSU	Hydrolase	2.5	53
5	2CYP	Cytochrome C peroxidase	1.7	54
6	2LZM	Lysozyme	1.70	55
7	4HBB	Hemoglobin	1.74	56

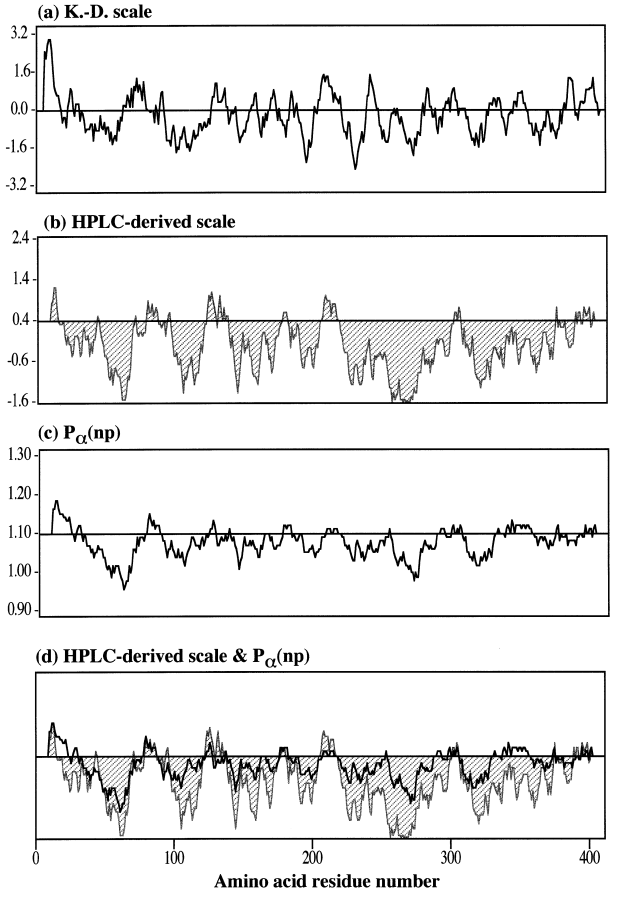

Figure 4. The sequence of carboxypeptidase A[50] examined through (a) the Kyte–Doolittle scale (see text for a discussion); (b) the HPLC-derived hydropathy scale;[38] and (c) the helical propensity values of Pα(np) (Table 1). In panel (d), data from the HPLC scale (panel (b)) are superimposed with Pα(np) data (panel (c)). No region in the carboxypeptidase A sequence fulfills the dual requirements of hydrophobicity and helicity.

herein demonstrate that TM helices and the great majority of their residues can be identified accurately. Despite the success of the present approach in TM helix prediction, the approach has yet to be employed to identify the other important stucture of membrane proteins: TM β-barrels. Future work focusing on the design of model TM β-barrel peptides should allow us to gain a more complete understanding as to how protein segments fold into membranes.

Materials and Methods

Peptide synthesis

Peptides were synthesized by the continuous-flow Fmoc solid-phase method.[31] C-Termini of peptides were aminated after cleavage from NovaSyn KR 125 resin. Purification of peptides was carried out on a reverse-phase Vydac-C4 semi-preparative HPLC (10×250 mm, 300 Å), using a linear gradient of acetonitrile in 0.1% trifluoroacetic acid (TFA). Purified peptides were characterized by analytical HPLC, amino acid analysis and mass spectrometry. Concentrations of peptides were determined in triplicate through quantitative amino acid analysis using Ala recovery as the standard. Peptides were stored as solid powders at −20 °C. To avoid the complexity of synthesizing a multiple Cys-containing peptide, for cysteine peptide, only the middle 'X' residue was substituted by Cys, and the other two 'X' residues were replaced by Leu. Amino acids are abbreviated by one-letter symbols: A-Ala, C-Cys, D-Asp, E-Glu, F-Phe, G-Gly, H-His, I-Ile, K-Lys, L-Leu, M-Met, N-Asn, P-Pro, Q-Gln, R-Arg, S-Ser, T-Thr, V-Val, W-Trp, Y-Tyr.

Spectroscopic measurements

Circular dichroism (CD) measurements were performed on a Jasco-720 spectropolarimeter using 1-mm path length quartz cell at 25 °C. Each spectrum was the average of four scans with buffer background subtracted. Peptide concentration was typically 30 μM in aqueous buffer, in *n*-butanol. The aqueous buffer was prepared from 10 mM Tris-HCl, 10 mM NaCl, pH 7.0. Peptides remained soluble and monomeric in both aqueous and membrane-mimetic media over a wide concentration range (5–250 μM), based on circular dichroism measurements and size exclusion HPLC chromatography.

Reverse-phase HPLC

The retention time of each 'X' peptide was determined on a C4 reverse-phase column (4.6×250 mm, 300 Å pore size, 10 μ particle size). Equal amounts of each peptide were injected into the column and eluted at a flow rate of 1 ml/min, utilizing a linear AB gradient (2% B/min), where buffer A was 0.1% TFA/ddH_2O, and buffer B was 0.1% TFA/acetonitrile. The retention time of each peptide is reported as the average of triplicate measurements.

Structural data base

The membrane protein data sample consists of 78 TM helices derived from six different membrane proteins (a total of 4816 residues) known with resolution equal to, or greater than, 4 Å (Table 2) selected from the Brookhaven Protein Data Bank (PDB). Soluble proteins (Table 4) were randomly chosen from the PDB as controls.

Prediction of membrane spanning segments

Membrane spanning segments were predicted by scanning the protein sequence within a 19 amino acid 'window'. The mean residue hydrophobicity and helicity were calculated by the algorithm available in GeneWork (version 2.2.1, IntelliGenetics, Inc., 1992) using the present scales. When the average hydrophobicity or helicity is plotted against residue number, the resulting curves identify potential membrane spanning segments as broad peaks with high positive values. For hydropathy analysis, previous studies with this series of peptides has identified the existence of a 'threshold hydrophobicity' (ca. 0.4, calculated as the mean residue hydrophobicity

of the X = Ala peptide, which contains 18 Ala's and one Trp in its hydrophobic core),[38,40] which controls the spontaneous incorporation of protein segments into membrane;[41] this experimentally-determined 'threshold hydrophobicity' was used as the cut-off point. For helicity analysis, peak values above 1.10 were used to determine the entry and exit points of the transmembrane region.

Acknowledgements

This work was supported, in part, by grants to C.M.D. from the Natural Sciences and Engineering Research Council of Canada, and the Medical Research Council of Canada. L.-P. L. thanks the Hospital for Sick Children Research Training Committee for a post-doctoral training award.

References

1. Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; Kerlavage, A. R.; Bult, C. J.; Tomb, J. F.; Dougherty, B. A.; Merrick, J. M.; et al. *Science* **1995**, *269*, 496.
2. Berlose, J. P.; Convert, O.; Brunissen, A.; Chassaing, G.; Lavielle, S. *Eur. J. Biochem.* **1994**, *225*, 827.
3. Schubert, U.; Ferrer-Montiel, A. V.; Oblatt-Montal, M.; Henklein, P.; Strebel, K.; Montal, M. *FEBS Lett.* **1996**, *398*, 12.
4. Schwartz, T. W. *Curr. Opin. Biotech.* **1994**, *5*, 434.
5. Bhakdi, S.; Bayley, H.; Valeva, A.; Walev, I.; Walker, B.; Kehoe, M.; Palmer, M. *Archives of Microbiology* **1996**, *165*, 73.
6. Lee, S.; Kiyota, T.; Kunitake, T.; Matsumoto, E.; Yamashita, S.; Anzai, K.; Sugihara, G. *Biochemistry* **1997**, *36*, 3782.
7. Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. *Science* **1996**, *274*, 1859.
8. Arkin, I. T.; Adams, P. D.; Brunger, A. T.; Smith, S. O.; Engelman, D. M. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 157.
9. Arteca, G. A.; Van Allen, D. R. *J. of Mol. Graphics* **1996**, *14*, 235.
10. Cowan, S. W.; Schirmer, T.; Rummel, G.; Steiert, M.; Ghosh, R.; Pauptit, R. A.; Jansonius, J. N.; Rosenbusch, J. P. *Nature (London)* **1992**, *358*, 727.
11. Martoglio, B.; Hofmann, M. W.; Brunner, J.; Dobberstein, B. *Cell* **1995**, *81*, 207.
12. Boekema, E. J.; Hankamer, B.; Bald, D.; Kruip, J.; Nield, J.; Boonstra, A. F.; Barber, J.; Rogner, M. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 175.
13. Davis, C. M.; Bustamante, P. L.; Todd, J. B.; Parkes-Loach, P. S.; McGlynn, P.; Olsen, J. D.; McMaster, L.; Hunter, C. N.; Loach, P. A. *Biochemistry* **1997**, *36*, 3671.
14. Deisenhofer, J.; Michel, H. *Science* **1989**, *245*, 1463.
15. Preusch, P. C.; Norvell, J. C.; Cassatt, J. C.; Cassman, M. *Nature Struct. Biol.* **1998**, *5*, 12.
16. Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckman, E.; Downing, K. H. *J. Mol. Biol.* **1990**, *213*, 899.
17. Pebay-Peyroula, E.; Rummel, G.; Rosenbusch, J. P.; Landau, E. M. *Science* **1997**, *277*, 1676.
18. Cowan, S. W.; Garavito, R. M.; Jansonius, J. N.; Jenkins, J. A.; Karlsson, R.; Konig, N.; Pai, E. F.; Pauptit, R. A.; Rizkallah, P. J.; Rosenbusch, J. P.; Rummel, G.; Schirmer, T. *Structure* **1995**, *3*, 1041.
19. Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105.
20. Engelman, D. M.; Steitz, T. A.; Goldman, A. *Annu. Rev. Biophys. Biophys. Chem.* **1986**, *15*, 321.
21. Degli Esposti, M.; Crimi, M.; Venturoli, G. *Eur. J. Biochem.* **1990**, *190*, 207.
22. Claros, M. G.; von Heijne, G. *Comput. App. Biol. Sci.* **1994**, *10*, 685.
23. von Heijne, G. *J. Mol. Biol.* **1992**, *225*, 487.
24. Juretic, D.; Lee, B.; Trinajstic, N.; Williams, R. W. *Biopolymers* **1993**, *33*, 255.
25. Hofmann, K.; Stoffel, W. *Bio. Chem. Hoppe-Seyler* **1993**, *374*, 166.
26. Persson, B.; Argos, P. *J. Mole. Biol.* **1994**, *237*, 182.
27. Persson, B.; Argos, P. *Prot. Sci.* **1996**, *5*, 363.
28. Jones, D. T.; Taylor, W. R.; Thornton, J. M. *Biochem.* **1994**, *33*, 3038.
29. Nakai, K.; Kidera, A.; Kanehisa, M. *Protein Eng.* **1988**, *2*, 93.
30. Horath, C. S.; Melander, W.; Molnar, I. *J. Chromatogr.* **1976**, *125*, 129.
31. Liu, L.-P.; Deber, C. M. *Biochem.* **1997**, *36*, 5476.
32. Reithmeier, R. A. F. *Curr. Opin. in Struct. Biol.* **1995**, *5*, 491.
33. Li, S. C.; Deber, C. M. *Nature Struct. Biol.* **1994**, *1*, 368.
34. Chung, L. A. *Anal. Biochem.* **1997**, *248*, 195.
35. Li, S.-C.; Goto, N. K.; Williams, K. A.; Deber, C. M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6676.
36. Ben-Efraim, I.; Bach, D.; Shai, Y. *Biochem.* **1993**, *32*, 2371.
37. Vinogradova, O.; Badola, P.; Czerski, L.; Sonnichsen, F. D.; Sanders, C. R. *Biophys. J.* **1997**, *72*, 2688.
38. Liu, L.-P.; Deber, C. M. *Biopolymers* **1998**, *47*, 41.
39. Park, K.; Perczel, A.; Fasman, G. D. *Protein Sci.* **1992**, *1*, 1032.
40. Liu, L.-P.; Li, S.-C.; Goto, N. K.; Deber, C. M. *Biopolymers* **1996**, *39*, 465.
41. Liu, L.-P.; Wang, C.; Goto, N. K.; Reithmeier, R. A. F.; Deber, C. M. (submitted).
42. Koepke, J.; Hu, X.; Muenke, C.; Schulten, K.; Michel, H. *Structure* **1996**, *15*, 581.
43. Kuhlbrandt, W.; Wang, D. N.; Fujiyoshi, Y. *Nature (London)* **1994**, *367*, 614.
44. Krauss, N.; Schubert, W. D.; Klukas, O.; Fromme, P.; Witt, H. T.; Saenger, W. *Nature Struct. Biol.* **1996**, *3*, 965.
45. Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yoshikawa, S. *Science* **1995**, *269*, 1069.
46. Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yoshikawa, S. *Science* **1996**, *272*, 1136.
47. Iwata, S.; Ostermeier, C.; Ludwig, B.; Michel, H. *Nature (London)* **1995**, *376*, 660.
48. Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science* **1998**, *280*, 69.
49. Ponnuswamy, P. K.; Gromiha, M. M. *Int. J. Pept. Protein Res.* **1993**, *42*, 326.
50. Rees, D. C.; Lewis, M.; Lipscomb, W. N. *J. Mol. Biol.* **1983**, *168*, 367.
51. Katti, S. K.; LeMaster, D. M.; Eklund, H. *J. Mol. Biol.* **1990**, *212*, 167.
52. Hynes, T. R.; Fox, R. O. *Proteins* **1991**, *10*, 92.
53. Bond, C. S.; Clements, P. R.; Ashby, S. J.; Collyer, C. A.; Harrop, S. J.; Hopwood, J. J.; Guss, J. M. *Structure* **1997**, *5*, 277.
54. Finzel, B. C.; Poulos, T. L.; Kraut, J. *J. Biol. Chem.* **1984**, *259*, 13027.
55. Weaver, L. H.; Matthews, B. W. *J. Mol. Biol.* **1987**, *193*, 189.
56. Fermi, G.; Perutz, M. F.; Shaanan, B.; Fourme, R. *J. Mol. Biol.* **1984**, *175*, 159.
57. Wallace, B. A.; Cascio, M.; Mielke, D. L. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 9423.