Distinctions between Hydrophobic Helices in Globular Proteins and Transmembrane Segments as Factors in Protein Sorting

Fiona Cunningham, Arianna Rath, Rachel M. Johnson, and Charles M. Deber

From the Division of Molecular Structure and Function, Research Institute, Hospital for Sick Children, Toronto, Ontario MSG 1X8, Canada and the Department of Biochemistry, University of Toronto, Toronto, Ontario MSS 1A8, Canada

Received for publication, December 1, 2008, and in revised form, December 18, 2008. Published, JBC Papers in Press, December 18, 2008, DOI 10.1074/jbc.M809017200

Transmembrane (TM) segments in proteins can be distinguished in amino acid sequences as continuous stretches of hydrophobic residues. However, examination of a data base of helical water-soluble (globular) proteins revealed that nearly one-third contained helices of sufficient length to span a bilayer (≥19 residues) that had mean hydrophobicity ≥actual TM segments. We now report that synthetic peptides corresponding to these globular protein sequences, which we termed “α-helices,” behave like native TM sequences and readily insert into membrane mimetic environments in helical conformations. As well, certain α-helix sequences can integrate into the membrane bilayer when placed into a membrane-targeted chimeric protein. We establish that α-helices can be distinguished computationally from bona fide TM segments by the decreased frequency of occurrence of Ile/Val residues and by their relatively decreased solvent accessibilities (versus other globular helices) within tertiary structure. The further observations that (i) α-helices generally contain three or more charged residues and (ii) α-helices display relatively even distribution of these charged residues along their lengths, rather than concentration near their N and C termini as observed for TM segments, may constitute key recognition factors in diverting α-helices from the membrane in vivo. Although a discrete biological role for α-helices remains to be pinpointed, our overall results suggest that such segments may be required for globular protein folding and identify additional factors that may be important in the correct selection of TM segments by the cellular machinery.

Membrane proteins regulate the molecular traffic and information flow across biological membranes, acting as receptors (1, 2), membrane pores (3), ion channels (4, 5), and metabolite transporters (6, 7). Despite their centrality in cellular function, little high resolution structural information is currently available for membrane proteins (currently ~1–2% of the >54,000 structures deposited in the Protein Data Bank). However, from this information two major structural classes of membrane proteins have been elucidated. In one category, proteins embedded in eukaryotic cell membranes and the inner membranes of bacteria and mitochondria generally adopt structures characterized by bundles of α-helical transmembrane (TM) segments. An alternative arrangement where β-strands span the membrane and assemble into barrel-like structures is common to proteins embedded in the bacterial and mitochondrial outer membranes (8).

Because high resolution structural determination of membrane proteins is not yet routine, computer simulation methodologies are often used to evaluate helical protein-protein (9) and protein-lipid interactions (10, 11). As a prerequisite to such studies, protein helical TM4 segments and their boundaries must be defined. Hydrophyt plots are commonly utilized to identify from primary sequence both the location of TM segments and their approximate entry/exit points (12–15). The hydrophyt values assigned to each residue to create such plots are drawn from one or more scales, among them the Liu-Deber values developed in our laboratory (16). TM segments and peptides derived from them that meet or exceed a segmentally averaged “threshold” hydrophobicity on this scale (approximately equivalent to a poly-Ala strand) have been shown to spontaneously insert into micellar media (17). The average hydrophobic levels of ~96% of natural TM segments were also found to exceed this threshold value (18).

From these observations, in conjunction with measurements of residue helical propensity in n-butanol (19), our laboratory developed the TM segment prediction program TM Finder (10) that uses segmental Liu-Deber hydrophobicity and nonpolar phase helicity values to query primary sequences for potential TM segments. TM Finder demonstrated a 98% predictive value in

This work was supported, in part, by grants (to C. M. D.) from the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant A-2807) and Canadian Institutes of Health Research (Grant FRN-5810). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby indicated as paid advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. 

The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1–S3.

1 Holds a post-doctoral research training award from the Hospital for Sick Children.

2 Holds a studentship award from the Canadian Institutes of Health Research Strategic Training Program in the Structural Biology of Membrane Proteins Linked to Disease.

3 To whom correspondence should be addressed: Division of Molecular Structure and Function, Research Institute, Hospital for Sick Children, 555 University Ave., Toronto, Ontario MSG 1X8, Canada. Tel.: 416-813-5924; Fax: 416-813-5005; E-mail: deber@skids.ca.

4 The abbreviations used are: TM, transmembrane; CD, circular dichroism; Pαφ, Chou-Fasman α-helical propensity; GpA, glycoporphin A; CAT, Chloramphenicol acetyltransferase; 1ECA, erythrocruorin; 1MBA, myoglobin; 1HC1, Hemocyanin A chain; 2MNR, mandelate racemase; 5LDH, L-lactate dehydrogenase; Pβφ, Chou-Fasman β-strand structural propensity; ΔGmic, calculated apparent free energy membrane insertion.
pinpointing TM segments in a training set of known membrane proteins (10). TM Finder was additionally specifically trained to limit the occurrence of false positives, i.e. globular (soluble) protein regions mispredicted as membrane-embedded. For this purpose the initial TM Finder code was applied to a sequence data base of globular proteins of known tertiary structure to assess helices that were of sufficient length (estimated at \( \geq 19 \) residues) to span the membrane bilayer. Of a total 174 \( \alpha \)-helical globular protein regions from 134 different proteins in this data base, we observed that \( \sim 30\% \) were identified as potential TM segments from the primary sequence alone but could be separated computationally from \( \delta \)-helices to reflect their intermediary between TM properties and extramembranous localization (20). In the present work we sought to extend this preliminary computational separation of \( \delta \)-helices, globular helices, and TM segments with a view toward understanding the criteria that may serve to distinguish these sequence features of globular proteins from TM segments in vitro or even within the cell.

**EXPERIMENTAL PROCEDURES**

**Data Base Construction**

The databases of globular and \( \delta \)-helices were initially compiled by searching the SwissProt release 34 with the keyword “helix.” The \( \sim 4200 \) sequences returned by this query were subsequently refined to 174 entries (20) as follows: (i) removing all redundant sequences (defined as those \( > 25\% \) identical); (ii) removing sequences with nonstandard amino acids; (iii) removing sequences without high resolution structure coordinates deposited in the Protein Data Bank; (iv) removing segments with \( < 19 \) residues. The 174 remaining sequences were then divided into globular helices (see supplemental Table S1) and \( \delta \)-helices (see supplemental Table S2) via submission to TM Finder, those segments with segmental hydropathy values at or above the Liu-Deber insertion threshold (\( \geq 0.4 \) on the Liu-Deber scale, see Ref. 21) were deemed \( \delta \)-helices. The data base of TM segment sequences was compiled from a non-redundant (defined as those \( > 25\% \) identical) list of TM \( \alpha \)-helices with available high resolution structures. Thirty-seven non-homologous TM proteins were identified containing a total of 212 different TM segments (see supplemental Table S3). Unlike the globular and \( \delta \)-helix databases, segments shorter than 19 amino acids were retained in the TM data base because the membrane localization of each was confirmed in a high resolution structure (16).

**Amino Acid Composition Analysis**

Amino acid composition was determined for each segment in each data base by counting the number of each residue and/or group of residues in each individual helix and dividing by helix length to obtain a normalized residue frequency. Mean residue and/or group residue composition values were then calculated for each of the globular helix, TM helix, and \( \delta \)-helix datasets. For group comparisons, hydrophobic residues were defined as Ala, Cys, Phe, Ile, Leu, Met, Val, Trp, and Tyr, as determined previously (19), polar residues were defined as Asp, Glu, Gly, His, Lys, Asn, Pro, Gln, Arg, Ser, and Thr, and charged residues were defined as Asp, Glu, Lys, and Arg. The overall mean amino acid compositions of globular, TM, and \( \delta \)-helices were compared with an online one-way analysis of variance test. Pair-wise comparisons of amino acid frequencies among the globular, TM, and \( \delta \)-helix datasets were performed using \( t \) tests.

**Solvent Accessibility Analysis**

The solvent accessibility of amino acids in globular helices and \( \delta \)-helices was evaluated by application of the program NACCESS (22) to structure coordinate files obtained from the Protein Data Bank as described previously (16) with the following modifications: (i) a probe size of 1.40 Å was used to approximate the radius of a water molecule, and (ii) the relative solvent accessibility for each amino acid was calculated by comparison of calculated solvent accessible surface areas to the default reference set supplied with NACCESS. Helix solvent accessibility was calculated as the sum of the individual residue analysis of variance values in each helix divided by the helix length. Helix solvent accessibility distributions were determined for the globular and \( \delta \)-helix groups by sorting individual helix solvent accessibility values into bins \((0–5\%), (5–10\%), (10–15\%), (15–20\%), (20–25\%), (25–30\%), (30–35\%), (35–40\%), (40–50\%), (50–55\%), (55–60\%); the mean solvent accessibility values for the globular and \( \delta \)-helix groups were also calculated from the individual solvent accessibility data. Within the globular and \( \delta \)-helix groups, the analysis of variance values of residues in the hydrophobic, polar, and charged residue categories defined above were averaged to obtain mean solvent accessibility values. Comparisons of the mean solvent accessibility of hydrophobic, polar, and charged residues between globular and \( \delta \)-helix segments were performed using \( t \) tests.

**Residue Position Analysis**

Residues in each helix sequence were sequentially numbered from 1 to \( n \), beginning at the N-terminal residue and ending at the C-terminal residue, such that \( n \) represents the total peptide length in residues. The number assigned to each charged residue (Asp, Glu, Lys, and Arg) in each helix sequence was divided by \( n \) to obtain a positional value normalized to helix length. Positional values were sorted into bins corresponding to fractions of helix length (\( \pm 20\% \), \( > 20–40\% \), \( > 40–60\% \), \( > 60–80\% \), \( > 80–100\% \)). Numbers of charged residues in each bin were totaled for the globular helix, \( \delta \)-helix, and TM helix groups. \( \chi^2 \) tests were used to analyze residue distributions.

**Peptide Synthesis and Purification**

Five \( \delta \)-helix segments were selected at random from the data base as candidates for in vitro study based on the presence of an intrinsic Trp residue: 1ECA (residues 114–133 of the full-length protein), 1HC1 (residues 218–236), 1MBA (residues 127–145), 2MNR (residues 96–116), and 5LDH (residues 247–266). The boundaries of the \( \delta \)-helix peptides of 1ECA, 1HC1, 1MBA, 2MNR, and 5LDH were chosen as prescribed by examination of the high resolution structure and the TM Finder output (10). Peptides with sequences corresponding to these \( \delta \)-helix segments were synthesized.
with a PS3 peptide synthesizer (Protein Technologies, Inc.) using standard Fmoc chemistry. Additional lysine residues were incorporated into the peptide sequences to increase aqueous solubility as previously described (23). A 4-fold amino acid excess on a 0.1-mmol scale synthesis was used with the O-(7-azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluroniumhexafluorophosphate; N,N-Diisopropylethylamine activator pair. Synthesis utilized a low load (0.18–0.22 mmol/g) PAL-polyethylene glycol-polystyrene resin that produced an amidated C terminus upon peptide cleavage. Peptide cleavage and deprotection was achieved using a mixture of 88% trifluoroacetic acid, 5% phenol, 5% ultrapure water, 2% trisopropylsilane followed by precipitation with ice-cold diethyl ether, drying, and resuspension in ultrapure water. Crude peptides were purified by reverse phase high performance liquid chromatography on a C4 preparative column (Phenomenex) with a water/acetonitrile gradient in the presence of 0.01% trifluoroacetic acid. Peptide molecular weights were confirmed by mass spectrometry, and the Micro BCA assay (Pierce) was used to determine peptide concentration.

Circular Dichroism (CD) and Fluorescence Spectroscopy

CD spectra were recorded on a Jasco J-720 circular dichroism spectrometer at room temperature. Spectra in aqueous buffer (10 mM Tris, pH 7.2, 10 mM NaCl) and aqueous buffer with 10 mM SDS were taken using a 0.1-cm path length cuvette at peptide concentrations of 25 µM. A 0.01-cm path length cuvette was used for secondary structure determination at peptide concentrations of 100 µM in aqueous buffer with 50 mM sodium perfluorooctanoate. Fluorescence measurements were carried out in the aqueous and SDS buffer conditions described above on a Hitachi F-400 Photon Technology International C-60 fluorescence spectrometer at an excitation wavelength of 295 nm. Emission spectra were recorded at room temperature from 305 to 405 nm using a peptide concentration of 5 µM, and deprotection was achieved using a mixture of 88% trifluoroacetic acid. Peptide molecular weights were confirmed by mass spectrometry, and the Micro BCA assay (Pierce) was used to determine peptide concentration.

TOXCAT Assay

Plasmid Construction—The expression vector pccKAN and the maltose-binding protein-deficient (malE-') Escherichia coli strain NT326 were kindly provided by Dr. Donald M. Engelman, Yale University (24). TOXCAT chimeras fusing the TM sequence of glycophorin A (GpA) and the G83I GpA mutant between the ToxR and maltose-binding protein domains have been previously described (11); chimeras encoding δ-helix sequences instead of the GpA sequence were constructed in an essentially identical manner via restriction digestion of oligonucleotide cassettes encoding each δ-helix segment with NheI and BamHI and subsequent ligation into the NheI and BamHI sites of the pccKAN plasmid. The identity of all constructs was confirmed with DNA sequencing before further characterization.

MalE Complementation Test—NT326 cells expressing TOXCAT chimeras with δ-helix sequences, the wild-type GpA TM sequence, or the G83I GpA mutant were streaked onto M9 minimal plates with 0.4% maltose as the only carbon source. Under these conditions, transformants capable of growth must target a portion of the chimeric TOXCAT protein into the cyto-

### TABLE 1

| Segment       | % Hydrophobic | % Charged | % Polar |
|---------------|---------------|-----------|---------|
| TM helix      | 67.3 ± 6.2    | 59.2 ± 6.2| 47.1 ± 7.8|
| δ-helix       |               |           |         |
| Globular helix| 47.1 ± 7.8    | 59.2 ± 7.7|         |

*Percentage occurrence values may not sum to 100% due to inclusion of residues Asp, Glu, Lys, and Arg in both the polar and charged categories. See “Experimental Procedures” for details.

Statistical significance between categories was determined by analysis of variance testing. Errors represent one S.D. of group compositions from the mean.
Hydrophobic Helices in Globular Proteins

plasmic membrane (24). Transformant growth was evaluated for all constructs after incubation for 2 days at 37 °C.

Chloramphenicol Acetyltransferase (CAT) Enzyme-linked Immunosorbent Assay—NT 326 cells harboring TOXCAT chimeras were grown at 37 °C, harvested into 1-ml fractions at an A400 of 0.6, pelleted, and stored at −80 °C. Cell lysates were prepared from cell pellets as previously described (11) and assayed for CAT concentration using the CAT enzyme-linked immunosorbent assay kit (Roche Applied Science). A standard curve was generated with CAT provided by the manufacturer. Cells expressing the wild-type GpA and G83I GpA sequences were included in each CAT assay as positive and negative controls, respectively. CAT measurements were performed at least triplicate and were normalized for the relative expression of each construct using Western blotting as described (11).

RESULTS

Hydropathy of δ-Helices—The globular protein sequences ≥19 residues in length considered in our initial study were divided into globular helix and δ-helix based on their mean segmental Liu-Deber hydropathy values (see “Experimental Procedures”). A data base of confirmed TM α-helical segments was also assembled from a non-redundant set of protein TM segments with available high resolution structures (supplemental Table S3); this TM helix data base contains 212 TM segments from 37 non-redundant proteins. We note that the length criterion of ≥19 residues was not applied to the TM helix data base because each TM segment was identified as residing in the membrane bilayer via high resolution structure determination. The mean Liu-Deber hydropathy of each sequence in the globular, δ-, and TM helix classes was calculated (see supplemental Tables S1–S3) and averaged for each group. We found that the mean hydropathy values of the δ- and TM-helix classes (0.72 ± 0.26 and 1.31 ± 0.69, respectively) each exceeded the Liu-Deber threshold value for membrane insertion (≥0.4, see Ref. 10), whereas the mean hydropathy value of globular helices (−0.28 ± 0.41) was below this threshold. Therefore, on average, δ-helices have intermediate hydropathy; that is, greater than that observed for globular helices and less than the TM helix value (p ≤ 0.01 in both comparisons).

Hydrophobic and Charged/Polar Residue Content in δ-Helices—The amino acid compositions of δ-, globular, and TM helix groups were determined (see “Experimental Procedures”) to investigate whether the intermediate hydropathy of δ-helices versus other globular helices and TM segments arose from decreased numbers of hydrophobic residues, increased numbers of polar residues, or both. As expected, TM helix segments contained ~1.4- to 2-fold more hydrophobic, ~3-fold fewer charged, and ~2-fold fewer polar residues than globular helices (p < 0.0001, see Table 1). δ-Helix percentage occurrence values in these residue categories, however, presented as different from the TM and globular groups (p < 0.0001) and appeared to be intermediate between them (Table 1). For example, the average percentage occurrence of hydrophobic residues per δ-helix (59.2 ± 6.2%) lies between the globular and TM helix values (47.1 ± 7.8 and 67.3 ± 9.2%, respectively). The distribution of hydrophobic and polar/charged amino acid residue types in δ-helices, thus, appears to be transitional between TM and globular helix segments.

Amino Acid Composition of δ-Helices Versus TM and Other Globular Helices—To further probe the origins of the compositional distinctness of δ-helices versus globular and TM segments, we compared individual amino acid percentage occurrence frequency among the three helix categories (Fig. 1). Consistent with the results of other groups (25), we observed that certain hydrophobic residues (i.e. Phe, Ile, Leu, Val, and Trp; Fig. 1A) were significantly enriched (p ≤ 0.01) in TM helices versus globular helical segments; the situation was reversed for polar and charged residues (i.e. Asp, Asn, Glu, Gln, Gly, His, Lys, and Arg; Fig. 1A). These trends are easily rationalized given the respective intramembranous versus cytoplasmic localization of these sequences in their native proteins. Similarly, we noted that that δ-helices contained significantly more Asp, Glu, Arg, and Lys residues than TM segments (0.05 ≥ p ≤ 0.01, see Fig. 1B), results consistent with our previous observation that the presence of three charged residues could delineate δ-helices as extramembranous (10). Interestingly, the observed decrease in hydropathy of δ-helices versus TM segments could be traced to two individual residues, viz. fewer Ile and Val residues are present in δ-helix sequences than in TM segments (p ≤ 0.01); conversely, the content of all other residues classed as hydrophobic is statistically indistinguishable (p ≥ 0.05, Fig. 1B).

Overall, the individual residue composition of δ-helices appeared to be more similar to globular than TM segments (compare Fig. 1, B and C); however, δ-helices were significantly enriched in the hydrophobic residues Leu and Phe and depleted in Glu, Lys, and Gln compared with their counterparts in globular proteins (Fig. 1C). In terms of hydropathy, δ-helices, thus,
appear to be distinguished from bona fide TM segments based on a decreased content of β-branched hydrophobic residues.

δ-Helices Are More Buried within Their Native Folds Than Other Globular Helices—Because δ-helices contain greater percentages of hydrophobic residues and lower percentages of polar/charged residues than other globular helices, we hypothesized that δ-helices might be more buried within their native protein folds. To determine whether this was the case, the locations of δ- and globular helices within their native structures were mapped using a water-sized probe to calculate residue solvent accessibility (see “Experimental Procedures”). Helices within the TM data base were excluded from this analysis due to their established burial within the membrane bilayer. δ-Helices were found on average to be more buried than other globular helices within their native folds (mean solvent accessibility values of 15–20% solvent-accessible, whereas the majority of globular helices have 25–30% accessibility). Rather than being confined to a single residue category, however, the increased accessibility of globular versus δ-helices is reflected across all residue groupings (Fig. 2B), i.e. hydrophobic, polar, and charged residues are ~1.3–1.4× more exposed in globular versus δ-helices.

Folding of the δ-Helix Peptides in Aqueous and Membrane Mimetic Environments—Because δ-helix segments represent sequences that are mispredicted as transmembranous, we initiated experiments to examine the physical properties of δ-helices versus TM helices in vitro. Accordingly, we synthesized peptides corresponding to five δ-helix segments selected from the database based on the presence of an intrinsic Trp residue for anticipated fluorescence experiments. Segments were selected from 1ECA, 1MBA, 1HC1, 2MNR, and 5LDH; see Table 2 for the corresponding δ-helix sequences. The hydrophobicity of the δ-helix sequences necessitated the introduction of Lys residues at their termini to facilitate synthesis and characterization (23); such “Lys tags” have been shown not to interfere with the core peptide sequence of interest (19, 26). The location of each of these δ-helix segments in their native protein structures is shown in Fig. 3.

CD spectra were obtained for each δ-helix peptide in aqueous buffer and in buffer containing SDS or perfluorooctanoate. Although each sequence demonstrably adopts α-helical structure within its soluble protein tertiary fold, none of the peptides exhibited a large amount of helical structure in aqueous buffer (Fig. 4). The 1MBA and 1HC1 peptides, however, displayed a degree of helical character in aqueous solution (minima at 208 and 222 nm; Fig. 4) that was not observed in the 1ECA, 2MNR, and 5LDH δ-helix peptides. We noted that the helix content of the δ-helix peptides was correlated with their Chou-Fasman secondary structure propensity (Pα) in aqueous solvent (27) (Fig. 5). For example, 1MBA and 1HC1 have the highest calculated α-helical structural propensity (Table 3) and also the greatest amount of helical structure (Fig. 5). The lack of strong aqueous helicity in the 1ECA and 2MNR δ-helix peptides may be similarly rationalized by the relatively high Chou-Fasman β-strand structural propensity (Pβ) predicted for these segments (Table 3). Interestingly, each δ-helix peptide sequence exhibits regions with overlapping α and β secondary structure prediction (Table 3), suggesting that they may represent segments with competing secondary structure preferences.

All δ-helix peptides increased in α-helix content when exposed to the membrane-mimetic environments of SDS or perfluorooctanoate micelles (Fig. 4). Induction of helical structure in apolar media has been observed with TM proteins such as GpA and the epidermal growth factor receptor (26) but can also occur when intact globular proteins are exposed to SDS micelles (for review, see Ref. 28). In both instances, ordered secondary structures such as α-helices are thought to arise in the hydrophobic environment of the detergent micelles to satisfy the hydrogen-bonding requirement of the peptide backbone in the low-dielectric environment of detergent acyl chains. δ-Helix peptide exposure to an environment of reduced polarity in SDS micelles was confirmed by examining the Trp fluorescence emission spectra of the δ-helix peptides. Trp has a characteristic fluorescence emission maximum of ~350 nm in an aqueous environment; blue shifting of this maximum to a lower wavelength accompanies accommodation of the Trp side chain in a more hydrophobic environment (29). Indeed, blue shifts in Trp emission maxima were observed for each δ-helix peptide in the presence of SDS (Table 4), suggesting that each is solvated in the apolar SDS micelle interior.

Competition of δ-Helix Segments for in Vivo Membrane Insertion—The behavior of the δ-helix peptides in SDS micelles implies that these sequences are competent for solva-

**TABLE 2**

| PDB ID | Protein          | Sequence*             | H0b         |
|--------|-----------------|------------------------|-------------|
| 1ECA   | Erythrocrucinin | K-FAGAEAWGATLTDTPFM1P-KK | 1.10        |
| 1HC1   | Hemocyanin A chain | KK-SLFPWVHIQTLKTFDFEKL-K | 0.77        |
| 1MBA   | Myoglobin       | K-ADAQFPGLGTLIDAAEAA-K | 0.96        |
| 2MNR   | Mandelate racemase | K-GLRMAAAGIDMAAWALGK-K | 0.53        |
| 5LDH   | L-Lactate dehydrogenase | KK-GVYAAGLGSVADLIESRKLK | 0.52        |

*a* Lys residues (offset by hyphens) were added to each peptide to enhance aqueous solubility. See “Experimental Procedures” for details.

*b* Mean Liu-Deber segmental hydrophobicity of each peptide, excluding Lys tags.

**FIGURE 3.** Ribbon diagrams of helical globular proteins containing δ-helix regions studied in this work. α-Carbon backbones of each protein are indicated in gray, with δ-helices shaded in green. Proteins and PDB identifiers are 1ECA (erythrocrucinin), 1HC1 (hemocyanin A chain), 1MBA (myoglobin), 2MNR (mandelate racemase), 5LDH (l-lactate dehydrogenase).
tion by micelles in vitro but offers no information as to whether or not they meet the requirements for membrane incorporation. We, therefore, undertook to investigate whether \( \alpha \)-helix segments could mimic TM sequences by assessing their ability to insert and self-associate in the \( E. coli \) inner membrane using the TOXCAT assay (24) (see “Experimental Procedures”). Of the five \( \alpha \)-helix sequences evaluated in the TOXCAT assay, only 1ECA was capable of self-association at levels higher than the monomeric GpA G83I control, with a self-association strength \( \geq 60\% \) of the GpA dimer (Fig. 6). The remaining \( \alpha \)-helix sequences reported lower levels of association strength than the monomeric control, with the possible exception of the 1MBA sequence. Because the correct membrane insertion of each fusion protein was confirmed by growth of NT326 cells expressing each fusion protein construct on M9-maltose plates (not shown), the latter data suggest that whether or not self-association occurs, at least a portion of each

**TABLE 3**

| \( \alpha \)-Helix | Predicted \( \alpha \)-region* | \( P_\alpha \) | Predicted \( \beta \)-region | \( P_\beta \) |
|------------------|-----------------------------|-------------|----------------|-------------|
| 1ECA             | 1–19                        | 1.11        | 5–11           | 1.17        |
| 1HC1             | 1–18                        | 1.17        | 5–10           | 1.17        |
| 1MBA             | 1–18                        | 1.17        | 5–13           | 1.21        |
| 2MNR             | 2–20                        | 1.15        | 2–12           | 1.20        |
| 5LDH             | 5–19                        | 1.13        | 2–18           | 1.09        |

* Residues in each \( \alpha \)-helix peptide are numbered beginning with the first residue after the Lys tag (see Table 2 for sequences). \( P_\alpha \) and \( P_\beta \) were computed using Chou-Fasman structure propensities (27). Lys tags were excluded from propensity calculations.

**TABLE 4**

| \( \alpha \)-Helix | \( \lambda_{\text{max}} \) (nm) | Blueshift\(^a\) |
|------------------|-------------------------------|----------------|
| 1ECA             | 348                           | 336            |
| 1MBA             | 348                           | 330            |
| 1HC1             | 344                           | 335            |
| 2MNR             | 348                           | 337            |
| 5LDH             | 348                           | 335            |

\(^a\) \( \lambda_{\text{max}} \) aqueous – \( \lambda_{\text{max}} \) SDS.

**FIGURE 4.** Circular dichroism spectra of \( \alpha \)-helix peptides in various media. Spectra in aqueous buffer (left), buffer containing SDS (center), and buffer containing sodium perfluorooctanoate (SPFO, right) are shown. Changes in mean residue ellipticity values observed at 208 and 222 nm for each \( \alpha \)-helix peptide observed in detergent solution are consistent with increased \( \alpha \)-helical structure.

**FIGURE 5.** Correlation of \( \alpha \)-helix secondary structure with Chou-Fasman \( \alpha \)-helix propensity. The mean residue ellipticity at 222 nm determined from the CD spectrum of each peptide in aqueous buffer is given as a function of the Chou-Fasman aqueous \( \alpha \)-helix propensity of the 1ECA, 1HC1, 1MBA, 2MNR, and 5LDH peptides. The regression line of best fit, correlation coefficient, and associated \( p \) values is shown; the correlation meets marginal significance levels (0.05 \( \leq p \leq 0.10 \)).

**FIGURE 6.** TOXCAT assay of \( \alpha \)-helix peptides in the \( E. coli \) inner membrane. Mean levels of CAT expression from \( \alpha \)-helices relative to the wild-type GpA dimer are shown \( \pm \) S.D. (top) along with a representative Western blot used to evaluate protein expression levels (bottom). Blot bands excerpted from separate gels are indicated by solid lines between lanes. G83I denotes the GpA G83I mutant, used as a monomeric control (24). The mean CAT expression levels were compared in \( t \) tests; symbols above the bars denote significance level: +, 0.05 \( \leq p \leq 0.10 \); *, \( p \leq 0.05 \); **, \( p \leq 0.01 \). See “Experimental Procedures” for assay details.
Hydrophobic Helices in Globular Proteins

δ-helix TOXCAT constructs is correctly incorporated in the E. coli inner membrane.

Charged Residue Distribution Distinguishes δ-Helix and TM Sequences—Given that δ-helices are comparable in length and in overall hydrophobic/helicity to native TM segments, we inquired whether the TOXCAT results might be explained by other characteristics of δ-helix sequences. For example, the bilayer integration efficiency of model TM segments in vivo has been shown to depend strongly on the position of charged residues, i.e. when they are placed toward the ends of sequences, insertion efficiency is increased (30). As such, it is possible that δ-helix sequences might be fundamentally distinguished from TM sequences based on charged residue positioning. The δ-helix, TM helix, and globular helix databases were queried accordingly for charged residue position (Fig. 7A). We found that charged residues were essentially evenly distributed along the lengths of δ- and globular helix sequences (\( p \approx 0.1 \)) with expected frequencies based on an even distribution. In contrast, charged residues were more abundant within the first 20% or last 20% of segment length than in the middle of TM segments (i.e. near the ends of TM helices). The sequences of δ-helices and TM segments can, thus, be computationally distinguished based on charged residue distribution. The notion that factors other than raw hydrophobicity must influence the membrane insertion propensity of a given δ-helix segment is also supported from the lack of correlation of segmental hydrophobicity versus the calculated apparent free energy membrane insertion (\( \Delta G_{\text{app}} \)) calculated using biological partitioning measurements (30, 31) (Fig. 7B).

DISCUSSION

Role in Globular Proteins—δ-Helices represent ~30% of the globular-based α-helices investigated. This relatively frequent occurrence implies that these segments must be of some utility in the proteins that contain them, and considerable evidence supports a major role for hydrophobic interactions in globular protein folding (32). We nevertheless could discern no trend in terms of localization of δ-helix segments to known surfaces of protein-protein interaction or segregation into a particular globular protein type (data not shown). However, δ-helices are by definition highly apolar and display increased burial within their native protein folds versus other globular helices (Fig. 2, A and B). As such, we suspect that δ-helix sequences may be important to the stability of proteins that contain them, perhaps via sequestration of their hydrophobic residues in the protein interior during folding.

δ-Helix segments are nevertheless unable to adopt their helical native structures without input from the remainder of the protein, perhaps because of their competing secondary structure preferences (Fig. 4, Table 3). In fact, δ-helix sequences generally display strong propensities to exist as β-sheet type segments when considered in the absence of the constraints imposed by globular protein tertiary structure (Table 3); this mixed potential is manifested in CD experiments, where several segments do not develop significant helical structure, even in membrane-mimetic environments. This disparity of structural propensity versus conformation of δ-helix segments mimics that of “discordant helices,” globular protein segments that undergo a α-to-β structure transition and form amyloid-like fibrils (33), and suggests that these sequences may exhibit structural instability in their native folds.

Role of Residue Content—The high intrinsic hydrophobicity of δ-helix segments is imparted by a different set of hydrophobic residues than TM segments. Ile and Val residues are depleted in δ-helices compared with TM segments (Fig. 1), and this may be rationalized in the requirement of such amino acids to be evolutionarily retained in the hydrophobic, restrictive environment of the membrane bilayer versus aqueous solvent. The β-branched residues such as Ile and Val have only one populated rotamer as a result of residing in membrane-induced α-helices, where they are structurally optimized for the folding and helix-helix interaction requirements of membrane proteins (25). There may, thus, be considerable selective pressure to retain Ile and Val in TM segments relative to δ-helices as they may retain a structural role beyond maintaining levels of hydrophobicity. Ile and Val are additionally better β-sheet formers than helix formers in aqueous solvent (27), perhaps necessitating their depletion in natively α-helical δ-helix sequences.

Recognition of Hydrophobic Segments—δ-Helix peptides are sufficiently similar to bona fide TM segments in terms of their segmental hydrophobicity and apolar helicity to be competent for membrane insertion, and our TOXCAT results indicate that certain δ-helix sequences are not only capable of membrane insertion but also of self-association within the bilayer when placed in the correct protein context (Fig. 6). How then are hydrophobic segments destined for the interior of globular

![Graph showing charged residue positioning in δ versus TM helices.](image-url)
Hydrophobic Helices in Globular Proteins

proteins distinguished from those that become incorporated into the interior of the membrane bilayer? Correct δ-helix versus TM segment sorting must rely on factors distinct from bulk biochemical properties. Based on our results, it appears that charged residue distribution and/or protein context may act to exclude δ-helix segments from the bilayer. Thus, authentic TM segments have a skewed distribution of charged residues with such residues appearing at helix termini compared with δ- and soluble α-helical segments (Fig. 7A). As well, proteins containing δ-helix segments appear to lack any additional TM-mimic sequences that could aid their membrane integration in a manner similar, for example, to the bilayer integration of voltage-sensor domains of the voltage-dependent potassium channels via “helper helices” (34); of the 51 globular helical proteins containing δ-helix sequences, only 6 have ≥2 δ-helix regions (Table S2). It is also possible that the potential sequestration of δ-helix sequences in the protein interior at an early stage in folding may prevent recognition of their high intrinsic hydrophobicity.

Examination of the estimated efficiency of δ-helix sequence partitioning into the lipid bilayer in vivo and into membrane-mimetic media in vitro further reinforces the notion that bulk biochemical properties are not sufficient to predict the non-TM versus TM location of a δ- versus TM-helix sequence. The two sets of values do not correlate, although a regression line and correlation coefficient are shown for illustrative purposes (Fig. 7B). Charged residues and their location in the δ-helix segments may, therefore, help to predict these hydrophobic protein regions as destined to be globular.

Conclusions—The present study shows that (i) nearly 30% of globular protein helices of sufficient length to span a membrane bilayer (≥19 amino acids) have mean hydrophathy values equivalent to or greater than known actual TM segments, and (ii) differentiation between these TM-mimicking portions of helical globular proteins and bona fide TM segments could generally be achieved in the first instance by flagging sequences with three or more charged residues as non-TM regions. We further observed that although significant hydrophobicity is absolutely a necessary feature in identifying potential TM segments, additional factors such as the location of the charged residues and an increased occurrence of Ile and Val residues should also be considered. Although δ-helix segments in globular proteins may embody important hydrophobic structural features of the in vivo protein fold, the work described herein provides additional clues as to how proteins may be sorted by the cell. Our current examination of factors that act to divert δ-helices from membrane insertion to an aqueous phase indicate that TM Finder and similar software could exploit these specific features to increase our predictability of transmembrane segments in proteins.

Acknowledgment—We thank Prof. Gunnar von Heijne, Stockholm University, for kindly performing the ∆G_app calculations for the δ-helix database and for helpful discussions.

REFERENCES

1. Okada, T., Sugihara, M., Bondar, A. N., Elstner, M., Entel, P., and Buss, V. (2004) J. Mol. Biol. 342, 571–583

2. Rasmussen, S. G., Choi, H. I., Rosenbaum, D. M., Kobitka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I., and Kobitka, B. K. (2007) Nature 450, 383–387

3. Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A., and Fujiyoshi, Y. (2000) Nature 407, 599–605

4. Eschagi, S., Niegoskwi, D., Kohl, A., Martinez Molina, D., Lesley, S. A., and Nordlund, P. (2006) Science 313, 354–357

5. Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., and MacKinnon, R. (2003) Nature 423, 33–41

6. Khadem, S., O’Connell, J., III, Remis, J., Robles-Colmeneros, Y., Miercke, L. J., and Stroud, R. M. (2004) Science 305, 1587–1594

7. Lupo, D., Li, X. D., Durand, A., Tomizaki, T., Cherif-Zahar, B., Matassi, G., Merrick, M., and Winkler, F. K. (2007) Proc. Natl. Acad. Sci. U. S. A. 104, 19303–19308

8. Galdiero, S., Galdiero, M., and Pedeone, C. (2007) Curr. Protein Pept. Sci. 8, 63–82

9. Cuthbertson, J. M., Bond, P. J., and Sansom, M. S. (2006) Biochemistry 45, 14298–14310

10. Deber, C. M., Wang, C., Liu, L. P., Prior, A. S., Agrawal, S., Muskat, B. L., and Cuticchia, A. J. (2001) Protein Sci. 10, 212–219

11. Johnson, R. M., Rath, A., Melnyk, R. A., and Deber, C. M. (2006) Biochemistry 45, 8507–8515

12. Zhao, G., and London, E. (2006) Protein Sci. 15, 1987–2001

13. Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132

14. Engelman, D. M., Steitz, T. A., and Goldman, A. (1986) Annu. Rev. Biophys. Biophys. Chem. 15, 321–333

15. Cserzo, M., Eisenhaber, F., Eisenhaber, B., and Simon, I. (2002) Protein Eng. 15, 745–752

16. Rath, A., and Deber, C. M. (2008) Proteins 70, 786–793

17. Liu, L. P., Li, S. C., Goto, N. K., and Deber, C. M. (1996) Biopolymers 39, 465–470

18. Liu, L. P., and Deber, C. M. (1999) Bioorg. Med. Chem. 7, 1–7

19. Liu, L. P., and Deber, C. M. (1998) J. Biol. Chem. 273, 23645–23648

20. Wang, C., Liu, L. P., and Deber, C. M. (2000) in Proceeding of the 16th American Peptide Symposium (Fields, G. B., Tam, I. P., and Barany, G., eds) pp. 367–369 Springer-Verlag, Minneapolis, Minnesota, MN

21. Liu, L. P., and Deber, C. M. (1998) Biopolymers 47, 41–62

22. Hubbard, S. L., and Thornton, J. M. (1993) NACCESS, 2.1.1 Ed., London

23. Melnyk, R. A., Partridge, A. W., Yip, J., Wu, Y., Goto, N. K., and Deber, C. M. (2003) Biopolymers 71, 675–685

24. Russ, W. P., and Engelman, D. M. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 863–868

25. Senes, A., Gerstein, M., and Engelman, D. M. (2000) J. Mol. Biol. 296, 921–936

26. Melnyk, R. A., Partridge, A. W., and Deber, C. M. (2001) Biochemistry 40, 11106–11113

27. Chou, P. Y., and Fasman, G. D. (1978) Annu. Rev. Biochem. 47, 251–276

28. Imamura, T. (2006) in Encyclopedia of Surface and Colloid Science (Soma-sundaran, P., ed) pp. 5251–5263, 2nd Ed., Taylor & Francis, New York; Vol. 7

29. Netz, D. J., Bastos Mdo, C., and Sahl, H. G. (2002) Appl. Environ. Microbiol. 68, 5274–5280

30. Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H., and von Heijne, G. (2007) Nature 450, 1026–1030

31. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, I., Andersson, H., Nilsson, I., White, S. H., and von Heijne, G. (2005) Nature 433, 377–381

32. Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008) Annu. Rev. Biophys. 37, 289–316

33. Paivio, A., Nordling, E., Kalberg, Y., Thyberg, J., and Johansson, I. (2004) Protein Sci. 13, 1251–1259

34. Zhang, L., Sato, Y., Hessa, T., von Heijne, G., Lee, J. K., Kodama, I., Sakaguchi, M., and Uozumi, N. (2007) Proc. Natl. Acad. Sci. U. S. A. 104, 8263–8268