Article No. jmbi.1999.3183 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 293, 807–814

JMB

Turns in Transmembrane Helices: Determination of the Minimal Length of a ''Helical Hairpin'' and Derivation of a Fine-grained Turn Propensity Scale

Magnus Monné, IngMarie Nilsson, Arne Elofsson and Gunnar von Heijne*

Department of Biochemistry Stockholm University, S-106 91 Stockholm, Sweden

We have recently reported a first experimental turn propensity scale for transmembrane helices. This scale was derived from measurements of how efficiently a given residue placed in the middle of a 40 residue poly(Leu) stretch induces the formation of a ''helical hairpin'' with two rather than one transmembrane segment. We have now extended these studies, and have determined the minimum length of a poly(Leu) stretch compatible with the formation of a helical hairpin. We have also derived a more fine-grained turn propensity scale by (i) introducing each of the 20 amino acid residues into the middle of the shortest poly(Leu) stretch compatible with helical hairpin formation, and (ii) introducing pairs of residues in the middle of the 40 residue poly(Leu) stretch. The new turn propensities are consistent with the amino acid frequencies found in short hairpin loops in membrane proteins of known 3D structure.

Keywords: membrane protein; protein structure; glycosylation; transmembrane helix; turn propensities

*Corresponding author

Introduction

The relationship between amino acid sequence and 3D protein structure, i.e. the ''second part of the genetic code'', remains to be fully elucidated. It is generally thought that the protein folding problem may be easier for integral membrane proteins than for globular proteins thanks to the limited number of architectures (helix bundles and β-barrels) compatible with a lipid environment (Cowan & Rosenbusch, 1994; von Heijne, 1996). Nevertheless, many rather fundamental questions concerning even such basic structural elements as the transmembrane α-helix itself remain to be addressed.

Transmembrane α-helices vary considerably in length, from less than 15 to over 40 residues (Bowie, 1997). The longest single transmembrane helices are thus longer than the shortest possible transmembrane helix-turn-helix structure (the ''helical hairpin''). As a first step towards understanding the factors responsible for helical hairpin formation, a single proline residue was shown to induce the formation of a helical hairpin when placed in the middle 12 positions in a 40 residue poly(Leu) transmembrane segment (Nilsson & von Heijne, 1998). We have now determined the minimum length of a poly(Leu) stretch compatible with the formation of a proline-induced helical hairpin. The minimum length is 31 residues, suggesting that the minimal helical hairpin is composed of two 14 residue transmembrane helices separated by a three residue turn.

Previously, we also examined the turn propensities of the 20 naturally occurring amino acids when placed in the middle of a 40-residue poly(Leu) segment (Monné *et al*., 1999). These data were consistent with a two-tier system, where residues have either a high or a low turn propensity, but it was not possible to further discriminate between the turn-inducing properties of the different residues. Here, we provide a more fine-grained scale, based on (i) measurements of the turn-inducing potential of single residues placed in the middle of a 31 residue poly(Leu) segment; and (ii) measurements of the turn-inducing potential of pairs of residues placed in the middle of a 40 residue poly(Leu) segment. We have also collected examples of helical hairpins with short loops from the available high-resolution membrane protein structures. Together, these results lead to a considerably sharper picture of the relative turn preferences of the different

E-mail address of the corresponding author: gunnar@biokemi.su.se

0022-2836/99/440807–8 $30.00/0

amino acids and the sequence characteristics responsible for the formation of transmembrane helical hairpins.

Results

Model protein and topology assay

As in our previous study of turn formation in transmembrane helices, we used the well-characterized *Escherichia coli* inner membrane protein leader peptidase (Lep) as a model protein. Lep consists of two transmembrane segments (H1 and H2) connected by a short cytoplasmic loop (P1) and followed by a large C-terminal periplasmic domain (P2). When expressed *in vitro* in the presence of dog pancreas microsomes, Lep adopts the same membrane topology as in its natural environment in the inner membrane of *E. coli* (Wolfe *et al.*, 1983), i.e. with the N and C termini on the lumenal side (Nilsson & von Heijne, 1993) (Figure 1(a), left).

To provide a simplified sequence context for our studies of turn formation, H2 was substituted by poly(Leu) stretches of the general composition $K_4L_nVL_{10}Q_3P$ (subscripts indicate the number of consecutive residues of a given kind). As an easily scored marker for the lumenal or cytoplasmic localization of the P2 domain, an N-glycosylation site (Asn-Ser-Thr) was introduced 20 amino acid residues downstream of H2; if the poly(Leu) stretch spans the membrane only once this site will be glycosylated by the lumenally disposed oligosaccharyl transferase enzyme, while it will not be modified if the poly(Leu) stretch forms a helical hairpin (Figure 1(a), right). We have previously shown that poly(Leu) helical hairpins are assembled into the ER membrane, ruling out the possibility that low levels of glycosylation are caused by a failure in membrane integration rather than helical hairpin formation (Nilsson & von Heijne, 1998). Turn formation can thus be assessed by *in vitro* transcription/translation of the relevant constructs in the presence of microsomes followed by quantification of the efficiency of glycosylation of the engineered N-glycosylation site.

Minimum length of a transmembrane helical hairpin

To determine the minimum length of the poly(Leu) stretch compatible with the formation of a transmembrane helical hairpin, a single proline residue was scanned across the middle positions of poly(Leu) stretches of different lengths. Proline was chosen for this study, since it was previously found to be one of the most effective inducers of helical hairpin formation in a 40 residues poly(Leu) segment (Monné *et al.*, 1999). As exemplified in Figure 1(b), each construct was expressed *in vitro* in the absence and presence of dog pancreas microsomes, analyzed by SDS-PAGE, and the degree of glycosylation was quantified.

As is clear from the quantifications shown in Figure 1(c), the size of the region in the center of the hydrophobic segment where proline induces formation of a helical hairpin decreases in proportion to the overall length of the poly(Leu) stretch. Remarkably, while efficient helical hairpin formation is still seen for the $L_{20}VL_{10}(P^{+2})$ construct (superscripts indicate positions relative to the center of the poly(Leu) stretch, with positive values towards the C-terminal end), there is no indication of helical hairpin formation for any of the $L_{29}V$-based constructs, although they are only one residue shorter. We also note that for all lengths of the poly(Leu) segment, there is a slight asymmetry in that the most efficient helical hairpin formation is seen when the proline residue is placed one or two residues C-terminal of the midpoint. Thus, the shortest possible helical hairpin has the sequence $K_4L_{17}PL_2VL_{10}Q_3P$, with 17 hydrophobic residues N-terminal and 13 C-terminal to the proline residue.

Turn-promoting effects of the 20 naturally occurring amino acids in the context of a 31 residue hydrophobic stretch

Our previous turn propensity scale was based on an analysis of constructs where all the 20 naturally occurring amino acid residues were placed in a central position in the $L_{29}VL_{10}$ poly(Leu) construct (Monné *et al.*, 1999). This scale indicated two groups of residues with, respectively, high and low turn propensities, but could not clearly differentiate between residues within the two groups. We reasoned that differences in turn propensities between residues at the high end of the scale might be more readily observed in a poly(Leu) segment close to the threshold for helical hairpin formation, i.e. in a 31 residues poly(Leu) segment. The 20 different residues (X) were thus introduced in construct $L_{17}XL_7VL_5$, i.e. in position +2 where proline has the strongest effect (c.f. Figure 1(c); note that we chose the $L_{25}VL_5$ context rather than $L_{20}VL_{10}$ to avoid possible complications stemming from having a valine residue rather than leucine residue next to the residue being tested). The glycosylation efficiencies for these substitutions were indeed higher compared to the 40 residues poly(Leu) stretch analyzed previously, and gave a more fine-grained quantitative discrimination between the residues in the group with high propensity for turn-induction (Figure 2(a), black bars).

Turn-promoting effects of residue pairs in the context of a 40 residue hydrophobic stretch

To better discriminate between the turn propensities of residues close to the border between the two groups with high and low turn propensities identified in our previous study (Monné *et al.*, 1999), we introduced a pair of identical amino acid residues in the middle of the $L_{29}VL_{10}$ poly(Leu) stretch to produce a set of constructs of the general

sequence $L_{20}X_2L_7VL_{10}$ where X is Gln, Glu, Gly, Trp, Ser, Tyr, Thr, Cys, Met, or Ile. This allowed a more clear-cut distinction to be made between residues essentially devoid of turn-promoting ability (Met, Ile) and those that are efficient turn-inducers when present as pairs (Glu, Gln, Gly, Trp, Ser, Tyr, Thr) (Figure 2(a), gray bars). The Cys-Cys pair had an intermediate effect.

To confirm that the introduction of turn-inducing residues in the poly(Leu) stretch did in fact result in the topology shown in Figure 1(a) (right) and not in the complete inversion of the topology of the whole molecule (i.e. in a N_{cyt}-C_{cyt} orientation with only two transmembrane segments), we engineered a second Asn-Ser-Thr N-glycosylation site in the N-terminal tail of the $L_{29}VL_{10}$ and $L_{29}VL_{10}(P^{+2})$ constructs (Figure 2(b)). As expected if the turn-inducing residue only affects the topology of the H2 segment, the former construct (which lacks a turn in H2) was glycosylated both on the N-tail and in the P2 domain, while the latter was glycosylated on the N-tail but not in the P2 domain (the $L_{29}VL_{10}(P^{+2})$ construct lacking an N-terminal glycosylation site is very inefficiently glycosylated; Monné *et al.*, 1999).

A final ranking of all the 20 different amino acids was constructed by averaging the results from the three sets of constructs (the $L_{21}XL_7VL_{10}$-series, the $L_{17}XL_7VL_5$-series, and the $L_{20}X_2L_7VL_{10}$-series) (Table 1).

Short loops in membrane proteins of known structure

To be able to compare our results with transmembrane helical hairpins in natural proteins, we collected all loops ⩽seven residues in helical hairpins from eight integral membrane proteins for which high-resolution structures have been reported (see Materials and Methods). A total of 21 such loops were found, and amino acid frequencies were calculated for the loops as well as for the five helical residues on either side of the loops (Figure 3). In Table 2, we show the sequences of all loops ⩽four residues in our collection.

The most striking feature in Figure 3 is the high incidence of Gly (18%) in the short loops. This is due to the very high incidence of Gly in the most N-terminal loop position (48%); in the other positions, Gly has a frequency of 8% (data not shown). This is similar to globular proteins, where Gly often serves as a C-terminal helix breaker (Aurora & Rose, 1998). Pro is found in the loops and among the first few residues in the C-terminal helix, but is absent from the C-terminal five residues in the N-terminal helix of the helical hairpin. Again, this is consistent with what is found for helices in globular proteins (Aurora & Rose, 1998). Leu and Val have high frequencies in the helical segments, but not in the loops. It is also interesting to note that all loops but one (the Gly-Ala loop from mitochondrial cytochrome *c* oxidase, chain A) contain at least two residues with high turn propensities (Table 2). The turn propensities in Table 1 are thus largely consistent with the composition of short loops seen in natural integral membrane proteins.

Table 1. Averaged turn propensities for residues in a transmembrane helix

Residue	Turn propensity	Normalized turn potential
Pro	1.6	2.7
Asn	1.6	2.6
Arg	1.5	2.6
Asp	1.5	2.5
His	1.4	2.3
Gln	1.4	2.3
Lys	1.4	2.3
Glu	1.3	2.1
Gly	1.1	1.9
Trp	0.9	1.6
Ser	0.9	1.6
Tyr	0.9	1.5
Thr	0.7	1.2
Cys	0.7	1.1
Ile	0.5	0.9
Met	0.5	0.8
Val	0.4	0.7
Ala	0.4	0.7
Phe	0.3	0.6
Leu	0.3	0.5

Turn propensities (column 2) were calculated as $[(1-f_{40}^{X})+(1-f_{31}^{X})+(1-f_{40}^{XX})]/[\mu(1-f_{40}^{X})+\mu(1-f_{31}^{X})+\mu(1-f_{40}^{XX})]$, where f_{40}^{X}, f_{31}^{X}, and f_{40}^{XX} are the fraction of glycosylated molecules with amino acid X (or a pair of X residues) placed in the middle of the 40 and 31 residue poly(Leu) stretches. $\mu(1-f_{40}^{X})$ is the mean value of $(1-f_{40}^{X})$ over all 20 residues. When no f_{40}^{XX} value was available, f_{40}^{XX} was put equal to 0.8 (this is the maximum glycosylation efficiency observed for the Lep P2 domain in our *in vitro* system, c.f. Figure 2(a)) for residues with high f_{40}^{X} values (Val, Ala, Phe), and to 0 for those with low f_{40}^{X} values (Pro, Asn, Arg, Asp, His, Lys). Normalized turn potentials (column 3) were calculated by dividing the values in column 2 by 0.6 (a value between that of Cys and Ile) to arrive at a scale where values >1 indicate residues that tend to induce a turn, while values <1 indicate residues that do not induce a turn even as pairs (c.f. Figure 2(a)).

Discussion

Many basic structural characteristics of membrane proteins have not yet been studied experimentally. Here, we have addressed two related questions: What is the shortest length of a hydrophobic region required for the formation of a transmembrane helical hairpin, i.e. two closely spaced transmembrane helices separated by a tight turn? and how efficient are the different amino acids in inducing the formation of the turn in the helical hairpin? In both cases, we have based the analysis on the insertion of model poly(Leu) segments into ER-derived microsomes, i.e. under conditions that closely mimic membrane protein assembly *in vivo*.

The shortest possible poly(Leu)-based helical hairpin was found to be 31 residues, including the central turn-inducing proline residue (Figure 1(c)). Two points are worth noting. First, the transition between a poly(Leu) segment that can and cannot

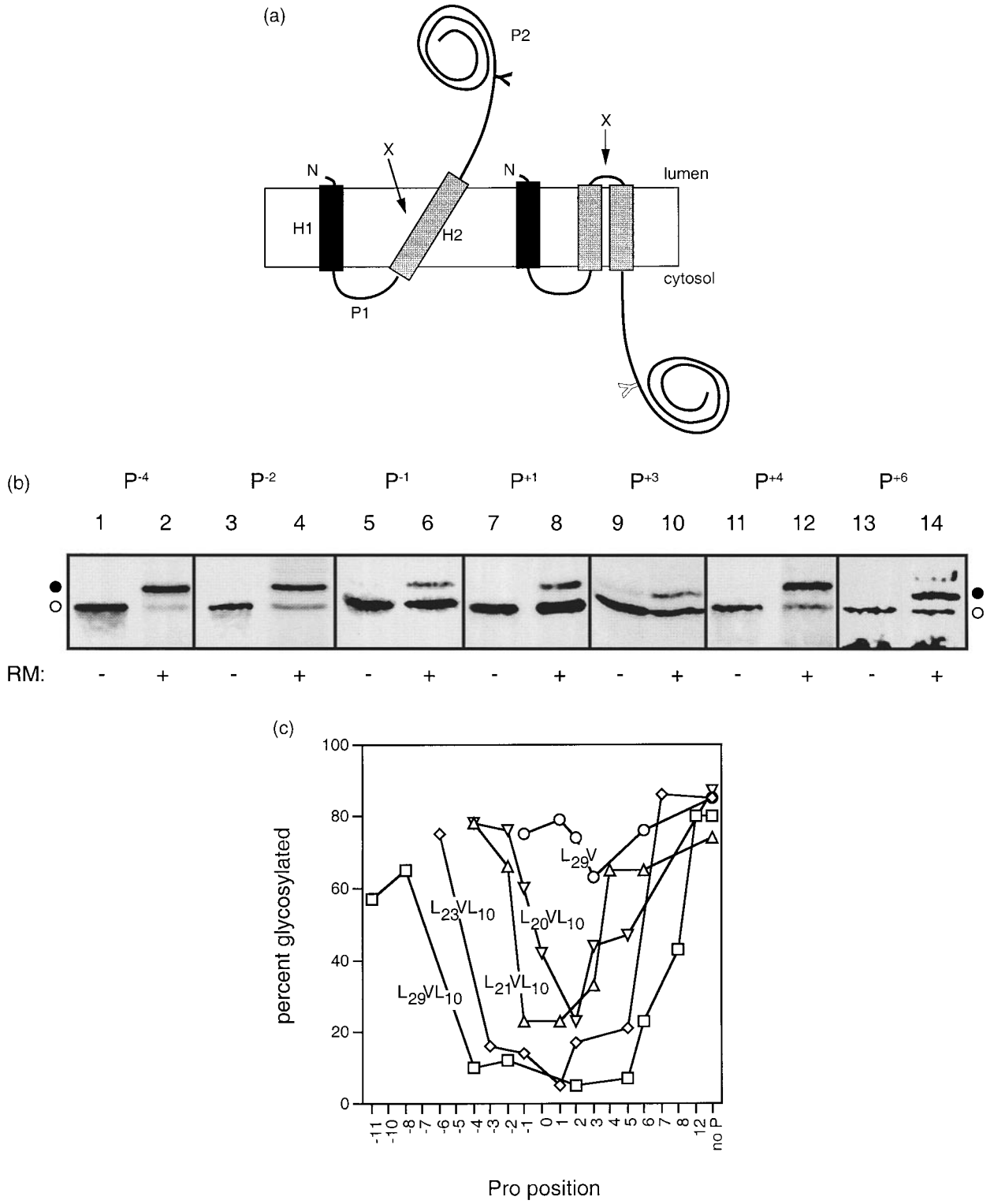

Figure 1 (*legend shown opposite*)

form a helical hairpin is very sharp and essentially corresponds to a length difference of only one residue. Second, in the shortest poly(Leu) stretches, turn-formation is maximally efficient when the proline residue is placed two residues into the C-terminal half of the hydrophobic stretch. Since proline can be accommodated at the N terminus but not at the C terminus of an α-helix (Aurora & Rose, 1998), and since the shortest turn between two α-helices is two to three residues, this suggests that the shortest possible helical hairpin is formed from two 14 residue helices (the C-terminal one

Table 2. Short loops in membrane proteins of known 3D structure

Protein	N-helix	Loop	C-helix
1AIJ (chain L) 107-120	ICRKL	GIGY	HIPFA
1AIJ (chain M) 136-149	RAQAL	GMGK	HTAWA
1OCC (chain C) 33-45	MWFHF	NSM	TLLMI
1OCC (chain C) 103-114	HSLME	GD	RKHML
bc_1 complex (chain C) 336-349	TWIGG	QPVE	HPYIT
mit. cyt *c* oxidase (chain A) 422-433	YVMSL	GA	VFGIF
mit. cyt *c* oxidase (chain C) 153-165	HALVH	ENN	RRDVA

Loops with a length $\leqslant$four residues found in the known structures (see Materials and Methods). The amino acid sequence of the loop as well as the five flanking N and C-terminal helical residues is shown.

with an N-terminal proline residue) and an intervening three-residue turn. This fits well with the 20 Å thickness of the hydrophobic core of the membrane estimated from an analysis of membrane proteins of known 3D structure (Wallin *et al.*, 1997), which would correspond precisely to a 14 residues helix. Shortening the poly(Leu) stretch by just one residue apparently makes it impossible for a sufficient number of the exposed hydrogen bond donor and acceptor groups in the turn and at the ends of the helices to make contact with water molecules or other polar groups in the lipid headgroup region; instead, a single transmembrane helix with a buried proline residue (two unsatisfied hydrogen bonds) becomes the energetically most favorable structure.

In an earlier study of turn formation in a 40 residue poly(Leu) transmembrane segment, we found that Pro was amongst the most efficient turn-inducers, together with Asn, Gln, and the charged residues (Monné *et al.*, 1999). By measuring the efficiency of turn induction in the 31 residue poly(Leu) segment, i.e. in the context of the shortest possible helical hairpin, we now find a clear ranking among these residues: Pro > Asn > Arg > Asp > Gln > His > Lys > Glu (Figure 2(a)). To get a clearer picture of the turn propensities near the middle of our previously derived scale, we also measured the efficiency of turn induction by pairs of residues inserted into the 40 residue poly(Leu) stretch (Figure 2(a)), which allowed us to arrive at a final ranking of all 20 residues (Table 1). Compared to our earlier scale, Trp, Tyr, Ser, and Thr now display clear turn-inducing properties, and even Cys has a slight effect. The only residues that appear totally devoid of turn-inducing potential are Phe and the aliphatic residues, fully consistent with their high abundance in transmembrane helices (Landolt-Marticorena *et al.*, 1993; Wallin *et al.*, 1997).

As a comparison, we have identified a collection of short loops in eight membrane proteins of known 3D structure (Table 1 and Figure 3). In almost every case, these short loops contain at least two turn-promoting residues, and they have a particularly high content of Gly in their most N-terminal position. Just as in our 31 residue helical hairpin, Pro is found quite frequently among the first few residues in the C-terminal helix but is never seen near the end of the N-terminal helix. The high incidence of Gly in the first loop position cannot be fully explained on the basis of the turn propensities measured above, but is consistent with the role of Gly in globular proteins where it often acts as a helix-breaking residue that, thanks to its lack of a side-chain, facilitates a reversal in the chain direction. In the context of a transmembrane helical hairpin, this would presumably facilitate the tight packing of the two transmembrane helices, a feature that is not scored in our poly(Leu)-based experimental system.

In conclusion, we have been able to define the minimum length required for the formation of a poly(Leu)-based helical hairpin, and have defined

Figure 1. (a) Model protein used in this study. The H2 transmembrane segment in Lep was replaced with a stretch of residues of the general design $LIK_4L_iPL_jVL_kQ_3P$, where *i*, *j*, and *k* (indicating the number of consecutive residues of a given kind) were varied to get the desired lengths. A glycosylation acceptor site was placed 20 residues downstream of H2 (counting from the first Gln residue after the hydrophobic stretch). Depending on the lumenal or cytoplasmic localization of the P2 domain, the glycosylation acceptor site will either be modified (**Y**) or not (𝕐). (b) $L_{21}VL_{10}(P^X)$ mutants with a proline residue in position X counting from the central residue of the poly(Leu) stretch (e.g. the P^{-1} mutation has the sequence $L_{15}PL_5VL_{10}$) were translated *in vitro* in the absence (−) and presence (+) of rough microsomes (RM) and analyzed by SDS-PAGE. Filled and open dots indicate the glycosylated and non-glycosylated forms of the proteins, respectively. (c) Quantification of the full set of data. The percentage glycosylation was calculated as $100I_+/(I_+ + I_-)$, where I_+ (I_-) is the intensity of the glycosylated (non-glycosylated) band. From duplicate experiments the typical error in the determination of glycosylation efficiency was $\leqslant$5%. Position zero corresponds to the middle residue of the hydrophobic segment, negative positions to the N-terminal side of position zero, and positive positions to the C-terminal side of position zero.

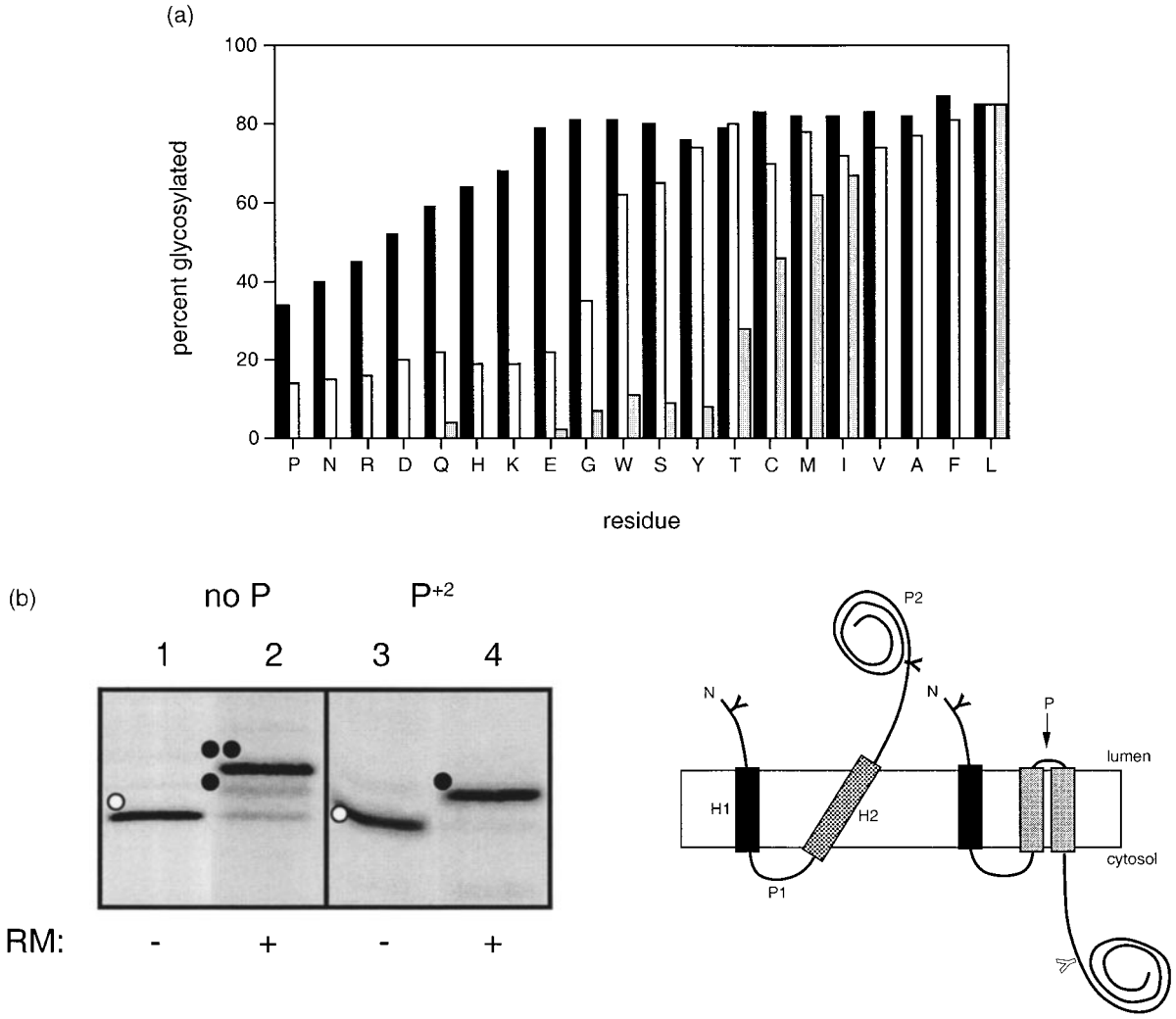

Figure 2. Turn-induction in a poly(Leu) transmembrane segment. (a) Glycosylation efficiencies measured for constructs with the H2 transmembrane segment in Lep replaced by a stretch of residues of the general design $LIK_4L_{17}XL_7VL_5Q_3P$ (black bars) or $LIK_4L_{20}X_2L_7VL_{10}Q_3P$ (gray bars); X is one of the 20 naturally occurring amino acid residues (c.f. Figure 1(a)). Results from our earlier turn propensity study (Monné *et al.*, 1999) where H2 was substituted by a stretch of the design $LIK_4L_{21}XL_7VL_{10}Q_3P$ are included for comparison (white bars). (b) Constructs $L_{29}VL_{10}$ (lanes 1 and 2) and $L_{29}VL_{10}(P^{+2})$ (lanes 3 and 4) with an extra glycosylation site in the N-terminal tail were translated *in vitro* in the absence (−) and presence (+) of rough microsomes (RM) and analyzed by SDS-PAGE. Two filled dots, one filled dot, and one open dot indicate, respectively, doubly glycosylated, singly glycosylated, and non-glycosylated forms of the proteins. The $L_{29}VL_{10}$ construct is efficiently glycosylated on both the N and C-terminal glycosylation sites, whereas the $L_{29}VL_{10}(P^{+2})$ construct is only glycosylated on the N-terminal site.

a clear ranking of the naturally occurring amino acid residues in terms of their ability to induce helical hairpin formation. In both cases, the results were obtained under *in vivo*-like conditions. Beyond their importance for the basic understanding of membrane protein structure, these observations should also help to improve current methods for membrane protein structure prediction.

Materials and Methods

Enzymes and chemicals

Unless otherwise stated, all enzymes were from Promega (Madison, WI, USA). Phage T7 DNA polymerase, *Bcl*I, [^{35}S]Met, ribonucleotides, deoxyribonucleotides, dideoxyribonucleotides, and the cap analog m7G(5′)ppp(5′)G were from Amersham-Pharmacia (Uppsala, Sweden). Plasmid pGEM1, transcription buffer and rabbit reticulocyte lysate were from Promega. Oligonucleotides were from Cybergene (Stockholm, Sweden).

DNA manipulations

For cloning into and expression from the pGEM1 plasmid, the 5′ end of the *lep* gene was modified first by the introduction of an *Xba*I site and, second, by changing the context 5′ to the initiator ATG codon to a ''Kozak consensus'' sequence (Kozak, 1992). Thus, the 5′ region of the gene was modified to: ...ATAACCCTCAGACCACCATGGCGAAT... (the *Xba*I site and initiator codon are underlined). Replacement of the H2 region in

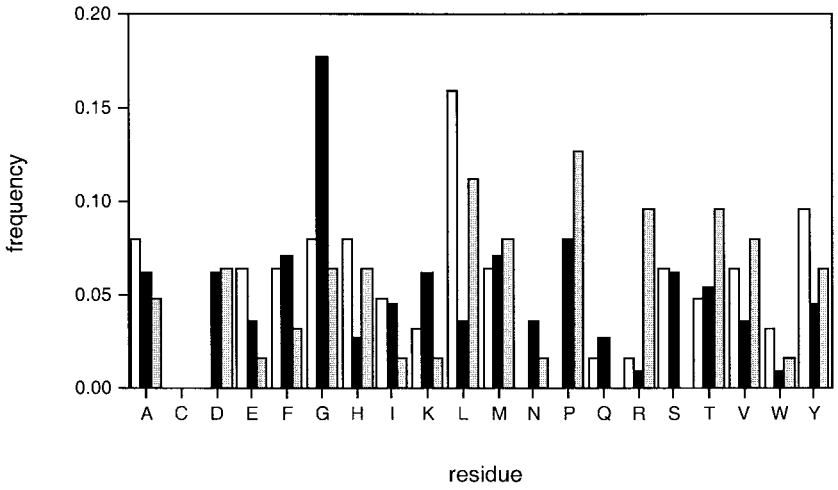

Figure 3. Amino acid frequencies in all loops ⩽seven residues located in transmembrane helical hairpins in eight membrane proteins of known structure (see Materials and Methods for a list). Data for the five N-terminal flanking helical residues are shown as white bars, in the loops as black bars, and in the five C-terminal flanking helical residues as gray bars.

Lep was performed by first introducing *Bcl*I and *Nde*I restriction sites in codons 59 and 80 flanking the H2 region and then replacing the *Bcl*I-*Nde*I fragment by the appropriate double-stranded oligonucleotides. Site-specific mutagenesis used to add *Bcl*I and *Nde*I restriction sites at the 3′ and 5′ ends of H2 in Lep and to introduce an Asn-Thr-Ser acceptor site for N-linked glycosylation was performed according to the method described by Kunkel (Geisselsoder *et al.*, 1987; Kunkel, 1987). The glycosylation acceptor site was designed as described (Nilsson *et al.*, 1994), i.e. by replacing three codons positioned 20 codons downstream of H2 with codons for the acceptor tri-peptide Asn-Ser-Thr. In all constructs, the naturally occurring glycosylation site at Asn214 in Lep was removed by an Asn214 → Gln mutation. For constructs with a glycosylation acceptor site in the N-tail, a previously designed Lep mutant (Nilsson & von Heijne, 1993) with a lengthened N-tail of the sequence MANSTSQGSQPINAQAAP-VAQGGSQGEF[4]... was used (F[4] is the fourth residue in wild-type Lep; the glycosylation site is underlined). Residues 59-81 in H2 were replaced by a poly(Leu) sequence of the design $LIK_4L_{29}VQ_3P$ (subscripts indicate the number of consecutive residues) for the 30 residue poly(Leu) construct and $LIK_4L_{29}VL_{10}Q_3P$ for the 40 residue poly(Leu) construct. In the 34, 32 and 31 residue poly(Leu) constructs, six, eight and nine leucine residues were deleted from the N terminus of the 40 residue poly(Leu) stretch by PCR mutagenesis. The poly(Leu) constructs used for introducing the 20 naturally occurring amino acid residues and the double substitutions had the general design $LIK_4L_{17}XL_7VL_5Q_3P$ and $LIK_4L_{20}X_2L_7VL_{10}Q_3P$. For the L → X and LL → XX substitutions, the QuickChange site-directed mutagenesis kit from Stratagene was used. Some of the primers were designed with a degenerate base in the second position in order to generate multiple amino acid substitutions. All mutants were confirmed by DNA sequencing of plasmid or single-stranded M13 DNA using phage T7 DNA polymerase.

Expression *in vitro*

The constructs in pGEM1 were transcribed by SP6 RNA polymerase for one hour at 37 °C. The transcription mixture was as follows: 1-5 μg of DNA template, 5 μl of 10 × SP6 H-buffer (400 mM Hepes-KOH (pH 7.4), 60 mM magnesium acetate, 20 mM spermidine-HCl), 5 μl of bovine serum albumin (1 mg/ml), 5 μl of m7G(5′)ppp(5′)G (10 mM), 5 μl of DTT (50 mM), 5 μl of rNTP mix (10 mM ATP, 10 mM CTP, 10 mM UTP, 5 mM GTP), 18.5 μl of H_2O, 1.5 μl of RNase inhibitor (50 units), 0.5 μl of SP6 RNA polymerase (20 units). Translation was performed in reticulocyte lysate in the presence and absence of dog pancreas microsomes (Liljeström & Garoff, 1991). Translation products were analyzed by SDS-PAGE and gels were quantified on a Fuji BAS1000 phosphoimager using the MacBAS 2.31 software. The glycosylation efficiency of a given mutant was calculated as the quotient between the intensity of the glycosylated band divided by the summed intensities of the glycosylated and non-glycosylated bands. In general, the glycosylation efficiency varied by no more than ±5 % between different experiments.

Analysis of known membrane protein structures

The following membrane protein structures were used to extract loops between transmembrane helices of length ⩽seven residues (PDB codes in parenthesis): bacterial photosynthetic reaction center (1AIJ) (Allen *et al.*, 1987), bacteriorhodopsin (1AP9) (Pebay-Peyroula *et al.*, 1997), potassium channel protein (1BL8) (Doyle *et al.*, 1998), mitochondrial cytochrome *c* oxidase (1OCC) (Tsukihara *et al.*, 1996), F_1F_o ATPase subunit c (1A91) (Girvin & Fillingame, 1995), mechanosensitive ion channel (1MSL) (Chang *et al.*, 1998), bacterial cytochrome *c* oxidase (Iwata *et al.*, 1995), and the bacterial cytochrome bc_1 complex (Iwata *et al.*, 1998). Secondary structure assignments were taken from the PDB files, or were assigned by STRIDE (Frishman & Argos, 1995) for the two last proteins. A total of 21 loops of length ⩽seven residues were found, and amino acid frequencies in these loops as well as in the five flanking helical residues on either side of the loops were calculated.

Acknowledgments

Drs S. Iwata, M. Iwata, and M. Ek (Uppsala University) kindly made co-ordinates for bacterial cytochrome *c* oxidase and the bacterial bc_1 complex available. Dog pancreas microsomes were a gift from Dr M. Sakaguchi, Fukuoka. This work was supported by grants from the Swedish Cancer Foundation, the Swedish Natural and Technical Sciences Research Councils, and the Göran Gustafsson Foundation to G.v.H.

References

Allen, J., Feher, G., Yeates, T., Komiya, H. & Rees, D. (1987). Structure of the reaction center from *Rhodobacter sphaeroides* R-26: the protein subunits. *Proc. Natl Acad. Sci. USA,* **84**, 6162-6166.

Aurora, R. & Rose, G. (1998). Helix capping. *Protein Sci.* **7**, 21-28.

Bowie, J. U. (1997). Helix packing in membrane proteins. *J. Mol. Biol.* **272**, 780-789.

Chang, G., Spencer, R., Lee, A., Barclay, M. & Rees, D. (1998). Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science,* **282**, 2220-2226.

Cowan, S. W. & Rosenbusch, J. P. (1994). Folding pattern diversity of integral membrane proteins. *Science,* **264**, 914-916.

Doyle, D., Cabral, J., Pfuetzner, R., Kuo, A., Gulbis, J., Cohen, S., Chait, B. & MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of K^+ conduction and sensitivity. *Science,* **280**, 69-77.

Frishman, D. & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins: Struct. Funct. Genet.* **23**, 566-579.

Geisselsoder, J., Witney, F. & Yuckenberg, P. (1987). Efficient site-directed *in vitro* mutagenesis. *BioTechniques,* **5**, 786-791.

Girvin, M. E. & Fillingame, R. H. (1995). Determination of local protein structure by spin label difference 2D NMR: the region neighboring Asp61 of subunit c of the F_1F_0 ATP synthase. *Biochemistry,* **34**, 1635-1645.

Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995). Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature,* **376**, 660-669.

Iwata, S., Lee, J. W., Okada, K., Lee, J. K., Iwata, M., Rasmussen, B., Link, T. A., Ramaswamy, S. & Jap, B. K. (1998). Complete structure of the 11-subunit bovine mitochondrial cytochrome bc_1 complex. *Science,* **281**, 64-71.

Kozak, M. (1992). Regulation of translation in eukaryotic systems. *Annu. Rev. Cell Biol.* **8**, 197-225.

Kunkel, T. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367-382.

Landolt-Marticorena, C., Williams, K. A., Deber, C. M. & Reithmeier, R. A. F. (1993). Non-random distribution of amino acids in the transmembrane segments of human type-I single span membrane proteins. *J. Mol. Biol.* **229**, 602-608.

Liljeström, P. & Garoff, H. (1991). Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. *J. Virol.* **65**, 147-154.

Monné, M., Hermansson, M. & von Heijne, G. (1999). A turn propensity scale for transmembrane helices. *J. Mol. Biol.* **288**, 141-145.

Nilsson, I. & von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* **268**, 5798-5801.

Nilsson, I. & von Heijne, G. (1998). Breaking the camel's back: proline-induced turns in a model transmembrane helix. *J. Mol. Biol.* **284**, 1185-1189.

Nilsson, I., Whitley, P. & von Heijne, G. (1994). The C-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase. *J. Cell Biol.* **126**, 1127-1132.

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. & Landau, E. (1997). X-ray structure of bacteriorhodopsin at 2.5 Ångstrom from microcrystals grown in lipidic cubic phases. *Science,* **277**, 1676-1681.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1996). The whole stucture of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science,* **272**, 1136-1144.

von Heijne, G. (1996). Principles of membrane protein assembly and structure. *Prog. Biophys. Mol. Biol.* **66**, 113-139.

Wallin, E., Tsukihara, T., Yoshikawa, S., von Heijne, G. & Elofsson, A. (1997). Architecture of helix bundle membrane proteins: an analysis of cytochrome *c* oxidase from bovine mitochondria. *Protein Sci.* **6**, 808-815.

Wolfe, P. B., Wickner, W. & Goodman, J. M. (1983). Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. *J. Biol. Chem.* **258**, 12073-12080.

Edited by F. E. Cohen

(Received 4 June 1999; received in revised form 7 September 1999; accepted 7 September 1999)