

**JMB**Available online at [www.sciencedirect.com](http://www.sciencedirect.com) ScienceDirect

## Repositioning of Transmembrane $\alpha$ -Helices during Membrane Protein Folding

Anni Kauko<sup>†</sup>, Linnea E. Hedin<sup>†</sup>, Estelle Thebaud, Susana Cristobal, Arne Elofsson\* and Gunnar von Heijne\*

Center for Biomembrane Research and Stockholm Bioinformatics Center, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden

Received 23 November 2009; received in revised form 19 January 2010; accepted 19 January 2010  
Available online 25 January 2010

Edited by J. Bowie

We have determined the optimal placement of individual transmembrane helices in the *Pyrococcus horikoshii* Glt<sub>PH</sub> glutamate transporter homolog in the membrane. The results are in close agreement with theoretical predictions based on hydrophobicity, but do not, in general, match the known three-dimensional structure, suggesting that transmembrane helices can be repositioned relative to the membrane during folding and oligomerization. Theoretical analysis of a database of membrane protein structures provides additional support for this idea. These observations raise new challenges for the structure prediction of membrane proteins and suggest that the classical two-stage model often used to describe membrane protein folding needs to be modified.

© 2010 Elsevier Ltd. All rights reserved.

**Keywords:** membrane protein; Glt<sub>PH</sub>; protein folding; membrane insertion

### Introduction

How do integral helix-bundle membrane proteins fold? Classically, the rather simple structural principles gleaned from the first available high-resolution three-dimensional structures—canonical hydrophobic transmembrane (TM)  $\alpha$ -helices packed against each other to shield polar residues—have been interpreted in terms of a two-stage folding model,<sup>1</sup> where the insertion of individual TM helices into the lipid bilayer in their energetically most favored position is followed by a folding process where preformed helices find their optimal packing interactions.

While the two-stage folding model is still a useful first-order approximation to the folding process, both structural and biochemical studies have begun to unravel a more complex reality.<sup>2</sup> TM helices are no longer seen as rigid rods, but often contain kinks and other kinds of nonhelical irregularities.<sup>3</sup> Reentrant loops dip into the plane of the membrane, but do not span its entire width.<sup>4</sup> During cotranslational membrane insertion, TM helices do not necessarily

exit the translocon one by one and can remain close to the translocon channel until the ribosome terminates translation,<sup>5</sup> may change their orientation in the membrane after chain termination,<sup>6,7</sup> or may insert into the membrane only at a late state during folding.<sup>5,8</sup>

Even the basic assumption that TM helices in the folded protein are individually in their equilibrium positions relative to the lipid bilayer—an assumption that often underlies attempts to predict the three-dimensional structure of membrane proteins from their TM topology—may not always hold.<sup>9</sup> We were alerted to this possibility by an analysis of TM helices in the Glt<sub>PH</sub> glutamate transporter homolog from the bacterium *Pyrococcus horikoshii* and now show that the TM2, TM4, and TM7 helices in Glt<sub>PH</sub> in fact become dramatically repositioned relative to the membrane during the folding and oligomerization process. These findings highlight a hitherto neglected aspect of membrane protein structure prediction, namely, that the positions in the membrane of TM helices in the folded structure do not always correspond to the thermodynamically favored positions in the membrane of the isolated helices. Instead, long-range tertiary interactions might make it more energetically favorable for TM helices to alter their position relative to the membrane during folding, thereby providing a way to introduce polar and charged residues into the membrane domain.

\*Corresponding authors. E-mail addresses: [arne@bioinfo.se](mailto:arne@bioinfo.se); [gunnar@dbb.su.se](mailto:gunnar@dbb.su.se).

<sup>†</sup> A.K. and L.E.H. contributed equally to this work.

Abbreviations used: TM, transmembrane; RM, rough microsome; MGD, minimal glycosylation distance; PDB, Protein Data Bank.

## Results

### The hydrophobicity of TM helices does not match membrane location in Glt<sub>Ph</sub>

Glt<sub>Ph</sub> forms a trimer with a large central cavity that extends partway across the membrane<sup>10</sup> (Fig. 1a). Each monomer is composed of eight TM helices and two reentrant loops, and both the N-terminus and the C-terminus face the cytoplasm. Recent structural studies show that TM1, TM2, TM4, and TM5 in each monomer together form a ‘trimerization domain’ that does not move relative to the membrane during the transport cycle,<sup>12</sup> confirming an earlier study on the Glt<sub>Ph</sub> homolog GltT where disulphide cross-linking was used to show that the trimerization domain remains rigid during the transport cycle.<sup>13</sup> TM4 and TM7 both contain short coil segments that break the helical structure near the middle of the membrane. However, as seen by comparing the left and right panels in Fig. 1b, the membrane-buried segments of TM2, TM4, TM7, and TM8 do not coincide with the most hydrophobic segments identified using the experimentally based ‘biological’ hydrophobicity scale, as embodied in the ‘ $\Delta G$  predictor’ program.<sup>14</sup> In addition, for the membrane-buried segments of TM7 and TM8, the predicted apparent free energy of insertion ( $\Delta G_{app}^{pred}$ ) is remarkably high and not typical of a TM helix. As the ‘ $\Delta G$  predictor’ has been shown to quite accurately predict the membrane-insertion efficiency of isolated TM helices,<sup>14–16</sup> this suggests that there may be some important rearrangements in the membrane-embedded part of Glt<sub>Ph</sub> during folding and trimerization.

To better understand the possible rearrangements of the TM segments in Glt<sub>Ph</sub>, we experimentally determined the apparent free energy of insertion into the endoplasmic reticulum membrane of the membrane-embedded part of each TM segment ( $\Delta G_{app}^S$ ), as defined in the OPM membrane protein structure database,<sup>17</sup> and the apparent free energy of insertion of the most hydrophobic segment corresponding to each TM segment ( $\Delta G_{app}^P$ ), as identified by the ‘ $\Delta G$  predictor.’

### Repositioning of TM helices in Glt<sub>Ph</sub>

We used a previously described assay<sup>15,18</sup> for measuring  $\Delta G_{app}^S$  and  $\Delta G_{app}^P$ . Briefly, the relevant Glt<sub>Ph</sub> segment (H-segment) is introduced into the ‘host’ protein Lep, as shown in Fig. 1c; note that two different Lep constructs are used, depending on the orientation of the TM helix in Glt<sub>Ph</sub> (N<sub>in</sub>-C<sub>out</sub> or N<sub>out</sub>-C<sub>in</sub>). Lep has two N-terminal TM helices (TM1 and TM2) and a large C-terminal domain (P2). When expressed in a rabbit reticulocyte *in vitro* transcription/translation system in the presence of dog pancreas rough microsomes (RMs), Lep inserts into the microsomal membrane with both the short N-terminal tail and the large P2 domain located in the lumen of the microsome.<sup>19</sup> In Lep<sup>I</sup> construct

(used for even-numbered Glt<sub>Ph</sub> TM helices), the H-segment is placed near the middle of the P2 domain and is flanked by two engineered Asn-X-Thr acceptor sites for N-linked glycosylation (G1 and G2). If the H-segment inserts efficiently into the membrane, only the G1 site will be modified by the luminal oligosaccharyl transferase; if, on the other hand, the H-segment is translocated across the membrane, both the G1 site and the G2 site will receive a glycan. Quantification of the fractions of singly glycosylated ( $f_{1x}$ ) and doubly glycosylated ( $f_{2x}$ ) molecules makes it possible to calculate an apparent equilibrium constant,  $K_{app}$ , for the membrane insertion of a given H-segment,  $K_{app} = \frac{f_{1x}}{f_{2x}}$ . The  $K_{app}$  value can be converted into an apparent free-energy difference between the noninserted state and the inserted state in the usual way:  $\Delta G_{app} = -RT \ln K_{app}$ , where  $R$  is the gas constant and  $T$  is the absolute temperature ( $T = 303$  K).

In Lep<sup>II</sup> construct (used for odd-numbered TM helices), the G2 site will be modified only if the H-segment inserts across the membrane, while the G1 site is always glycosylated; hence,  $K_{app} = \frac{f_{2x}}{f_{1x}}$ . Since we have found previously that  $\Delta G_{app}$  values determined using the Lep<sup>II</sup> construct are  $\sim 1$  kcal/mol lower than those determined using the Lep<sup>I</sup> construct and depend on the sequence of the H1 TM helix,<sup>15</sup> all Lep<sup>II</sup> values reported below have been increased by 1 kcal/mol to make the data obtained with the two Lep constructs comparable and also to make them comparable with the results from the ‘ $\Delta G$  predictor.’

We used the Lep-based glycosylation assay to measure  $\Delta G_{app}^S$  for all the structurally defined membrane-embedded TM helices and  $\Delta G_{app}^P$  for the most hydrophobic regions overlapping TM2, TM4, TM7, and TM8 (Fig. 1d) (see Table 1 and Fig. S1 for sequences). The experimentally measured  $\Delta G_{app}^S$  and  $\Delta G_{app}^P$  values and the corresponding predicted  $\Delta G_{app}^{pred,S}$  and  $\Delta G_{app}^{pred,P}$  values agree well in most cases. For TM2, TM4, and TM7,  $\Delta G_{app}^P < \Delta G_{app}^S$ , as expected; for TM8,  $\Delta G_{app}^S$  is surprisingly low (given  $\Delta G_{app}^{pred,S}$  for TM8) and even slightly lower than  $\Delta G_{app}^P$ . One possible explanation for the anomalous behavior of TM8 could be charge-pairing between residues D394 and R397, but further studies will be needed to clarify this issue.

As seen in Fig. 1d, the segments with the lowest  $\Delta G_{app}^P$  values are significantly displaced relative to the structurally defined membrane-embedded TM helices for TM2, TM4, and TM7. This suggests that, for these TM helices, the segment that initially inserts into the membrane is different from the membrane-embedded segment that forms the corresponding TM helix in the folded trimer.

To further substantiate this conclusion, we used a ‘glycosylation mapping’ approach to map the ends of the membrane-embedded TM4 segment, both when inserted into the Lep constructs and when present in its normal context within Glt<sub>Ph</sub>. This approach is based on the observation that the oligosaccharyl transferase active site sits at a well-

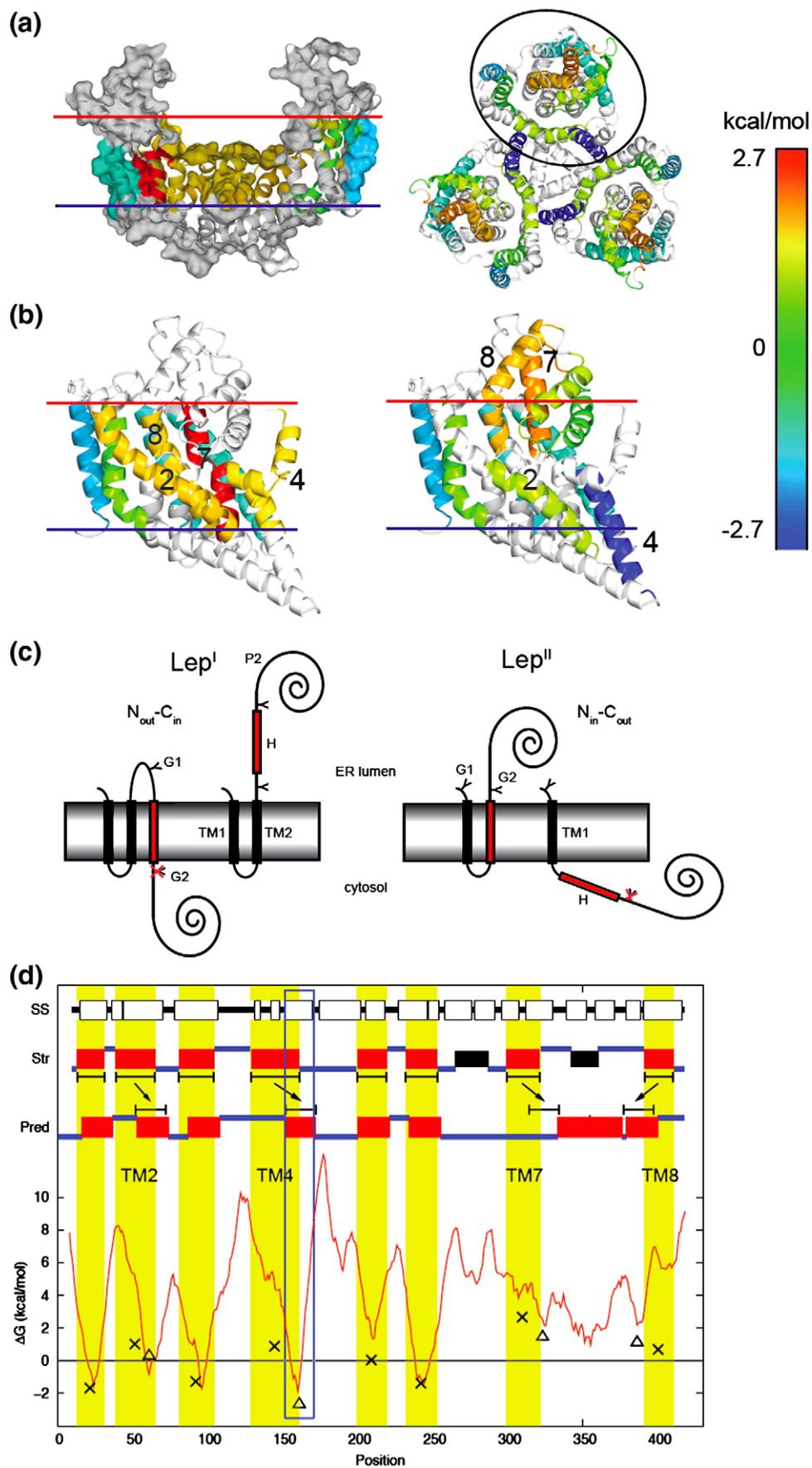


Fig. 1 (legend on next page)

**Table 1.** Apparent free energies of membrane insertion for the investigated segments of Glt<sub>Ph</sub>

TM segment	'Structural' H-segment			'Predicted' H-segment			Sequence
	Residue#	$\Delta G_{app}^s$	$\Delta G_{app}^{pred,s}$	Residue#	$\Delta G_{app}^p$	$\Delta G_{app}^{pred,p}$	
TM1	13-30	-2.7	-0.3	-	-	-	<b>L<sub>13</sub>QKILIGLILGAIVGLIL<sub>30</sub></b>
TM2	39-64	1.0	2.1	51-71	0.3	0.0	<b>V<sub>39</sub>HTYVKPFGDLFVRLKMLVMPIVFASLVVGA<sub>71</sub></b>
TM3	81-103	-2.3	0.5	-	-	-	<b>V<sub>81</sub>GVKIVVYLLTSAFAVTLGIIM<sub>103</sub></b>
TM4	129-160	0.9	3.1	151-171	-2.7	-2.0	<b>P<sub>129</sub>LVHILLDIVPTNPFALANGQVLPITIFFAILGLIAITYLMNS<sub>171</sub></b>
TM5	200-218	-1.0	1.2	-	-	-	<b>G<sub>200</sub>VMQYAPIGVFALIAVMA<sub>218</sub></b>
TM6	232-252	-1.4	-0.6	-	-	-	<b>T<sub>232</sub>AAVYVGLTLQILLVYFVLLK<sub>252</sub></b>
TM7	299-320	2.7	4.1	313-333	0.4	0.4	<b>Y<sub>299</sub>SFTLPLGATINMDGTALYQGVCFFIANALGSHL<sub>333</sub></b>
TM8	391-409	0.7	6.2	376-396	1.1	-0.1	<b>D<sub>391</sub>PNVAAAYAMILGIDAILDMGRITMVNVTGDLTGT<sub>396</sub></b>

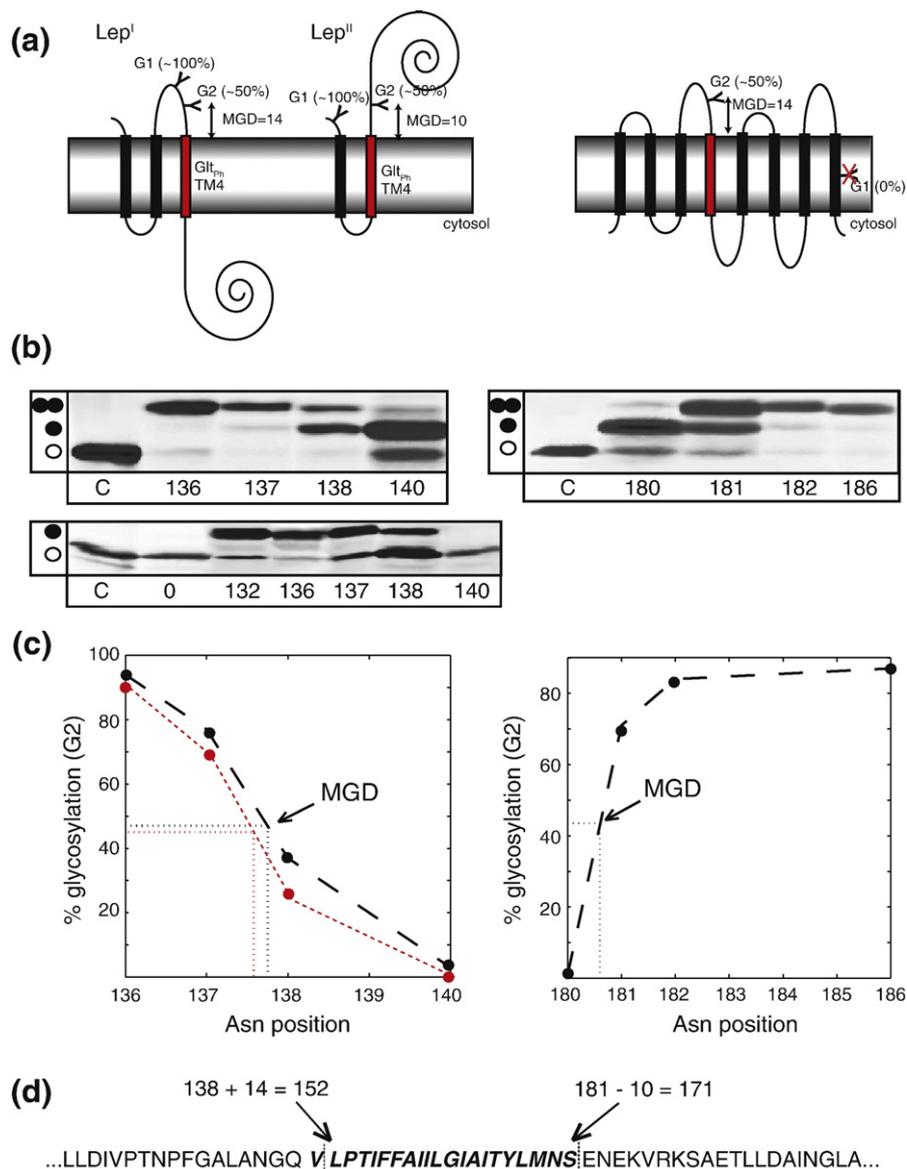
Predicted and experimental values for the insertion efficiency of the investigated TM segments. Each segment was introduced as an H-segment (with GGPG...GPGG flanks) into the Lep<sup>I</sup> or Lep<sup>II</sup> vector, and  $\Delta G_{app}$  values were obtained as described in the main text.  $\Delta G_{app}^s$  and  $\Delta G_{app}^p$  denote experimentally determined insertion free energies.  $\Delta G_{app}^{pred,s}$  and  $\Delta G_{app}^{pred,p}$  are values predicted by the 'ΔG predictor' program. Note that the structural segments (bold underlined sequence) for TM1, TM3, TM5, and TM6 were not predicted to have an overlapping segment (shadowed sequence) with a substantially lower  $\Delta G_{app}$ .

defined distance from the end of a membrane-embedded hydrophobic TM helix<sup>20</sup> (Fig. 2a). Calibration against membrane-embedded TM helices known from X-ray crystallography or NMR has shown that this 'minimal glycosylation distance' (MGD) is ~14 residues at the N-terminal end of the TM helix and ~10 residues at the C-terminal end;<sup>21-25</sup> note that both ends of the TM helix can be mapped by inverting its orientation relative to the membrane in one of the two Lep constructs. As shown in Fig. 2b-d, glycosylation mapping places the N-terminal end of the membrane-embedded TM4 segment around residue 152 (irrespective of whether the measurement is performed in the Lep<sup>I</sup> construct or in Glt<sub>Ph</sub>) and the C-terminal end around residue 171 (blue box in Fig. 1c; the C-terminal measurement cannot be performed for TM4 in the intact Glt<sub>Ph</sub> protein). This is in perfect agreement with the segment predicted by the 'ΔG predictor' (residues 151-171), but not with the membrane-embedded part of TM4 as seen in the folded structure (residues 129-160). As N-linked glycosylation of membrane proteins is generally a cotranslational

event,<sup>26</sup> the result obtained for TM4 in the context of the full Glt<sub>Ph</sub> protein shows directly that the most hydrophobic segment in TM4 is the initial membrane-spanning part, and that TM4 shifts in position by 12-13 residues relative to the membrane during the subsequent folding and oligomerization process. We cannot formally rule out the alternative possibility that repositioning takes place already within the translocon, after the glycosylation event but before TM4 is inserted into the lipid bilayer. We consider this scenario very unlikely, however, given that TM4 is intimately involved in the formation of the trimerization domain, and that highly polar parts of it would be lipid-exposed if the initial membrane-inserted form of the monomer were to have the same structure as it has in the fully assembled trimer.

In summary, the experimental results confirm our initial prediction that three membrane-embedded TM helices (TM2, TM4, and TM7) are inefficiently recognized by the translocon, but that overlapping, more hydrophobic segments have higher insertion efficiencies. Moreover, the borders of the membrane-

**Fig. 1.** The Glt<sub>Ph</sub> glutamate transporter homolog. (a) Left: A 16-Å-thick slice through the middle of the trimeric structure (PDB code 2NWL). TM helices are shown in a color code reflecting their hydrophobicity as predicted by the 'ΔG predictor' (see the scale), and the approximate boundaries of the membrane are shown as red (periplasmic side) and blue (cytoplasmic side) lines. Right: The position of a monomer (encircled) in the trimer as viewed perpendicular to the membrane. Hydrophobic segments are colored as in (b) (right). (b) Structure of a Glt<sub>Ph</sub> monomer. Left: The structurally defined TM helices colored according to hydrophobicity. Right: The most hydrophobic segments identified by the 'ΔG predictor' as overlapping the TM helices colored according to hydrophobicity. (c) The Lep<sup>I</sup> and Lep<sup>II</sup> constructs used to measure membrane-insertion efficiencies for even-numbered N<sub>out</sub>-C<sub>in</sub> and odd-numbered N<sub>in</sub>-C<sub>out</sub> Glt<sub>Ph</sub> TM helices, respectively. The apparent free energy of membrane insertion ( $\Delta G_{app}$ ) of the TM segment in question (H-segment) is quantified by comparing the fractions of singly and doubly glycosylated molecules, as described in the main text. (d) Experimentally measured and predicted insertion efficiencies for TM segments in Glt<sub>Ph</sub>. The top line shows the secondary structure (SS; helices in white) of the protein as annotated in the PDB file; the second line shows the topology based on the X-ray structure (Str; TM helices in red and reentrant loops in black); and the third line shows the topology as predicted by SCAMPI<sup>11</sup> (Pred). The ΔG graph (red line) shows the predicted membrane-insertion efficiency along the sequence. The experimentally measured insertion efficiencies for the membrane-embedded TM helices ( $\Delta G_{app}^s$ ) are shown as crosses, and those measured for the most hydrophobic regions predicted by the 'ΔG predictor' as overlapping TM2, TM4, TM7, and TM8 ( $\Delta G_{app}^p$ ) are shown as triangles (black bars in the Str and Pred plots indicate the segments used in the experiments). The membrane-embedded segment of the isolated TM4 based on MGD measurements is shown as a blue box.



**Fig. 2.** Helix boundaries of the TM4 helix determined by glycosylation mapping. (a) Left: The Lep<sup>I</sup> and Lep<sup>II</sup> constructs used to determine, respectively, the N-terminal and C-terminal ends of the membrane-embedded segment of Glt<sub>Ph</sub> TM4. In both constructs, the G2 glycosylation site is moved stepwise relative to the TM segment to determine the MGD (i.e., the number of residues away from the TM segment where the glycosylation efficiency is one-half of the maximal glycosylation efficiency). For these experiments, Glt<sub>Ph</sub> residues A104-N199 encompassing TM4 and its flanking loops (see Fig. S1) were cloned into Lep<sup>I</sup> and Lep<sup>II</sup>. Right: Full-length Glt<sub>Ph</sub>. The positions of a nonglycosylated NVT sequon (G1) present in TM8 and the engineered G2 glycosylation site used to determine the MGD for the N-terminal end of TM4 are indicated. (b) SDS-PAGE gels of relevant constructs. Top left: Lep<sup>I</sup>; bottom left: full-length Glt<sub>Ph</sub>; right: Lep<sup>II</sup>. Doubly glycosylated molecules are indicated by two black dots; singly glycosylated molecules are indicated by one black dot; and unglycosylated molecules are indicated by a white dot. The sequence position of the engineered glycan acceptor Asn residue is indicated below each lane. Lane C is a control translation in the absence of RMs, and lane 0 in the Glt<sub>Ph</sub> panel is a translation of full-length Glt<sub>Ph</sub> with no added glycosylation site; note that the G1 sequon in Glt<sub>Ph</sub> is not glycosylated (compare lanes C and 0) and therefore does not interfere with the MGD measurement. (c) Quantification of glycosylation efficiencies and determination of MGD values (i.e., the position of the glycan acceptor Asn residue at half-maximal glycosylation). Left: Lep<sup>I</sup> (black) and full-length Glt<sub>Ph</sub> (red). Right: Lep<sup>II</sup>. (d) Partial sequence of the segment A104-N199 used in the Lep<sup>I</sup> and Lep<sup>II</sup> constructs to map the MGD of TM4. The membrane-embedded helix in the X-ray structure is underlined, the membrane-embedded segment predicted by the ‘ $\Delta$ G predictor’ is in bold italics, and the ends of the membrane-embedded segment determined by glycosylation mapping are indicated by arrows. Note that the Asn positions determined in (c) are ~14 residues away from the N-terminal end of the TM segment and ~10 residues away from the C-terminal end, respectively.<sup>21–25</sup>

inserted part of the isolated TM4 helix are essentially identical with the ones predicted by the ‘ $\Delta$ G predictor.’ Thus, it is likely that in all three cases,

the most hydrophobic segment is initially inserted into the membrane. Tertiary and quaternary interactions during folding then induce shifts in the

positions of the helices relative to the membrane. These shifts pull the coil segments, as well as charged and polar residues within TM4 and TM7, into the membrane, presumably in such a way that they do not come into direct contact with the lipid bilayer. While TM7 is part of a domain that moves relative to the membrane during the transport cycle, TM2 and TM4 remain in a fixed position.<sup>12</sup> Thus, for the two latter TM helices, the reorientation relative to the membrane is part of the folding and oligomerization process, while it is related to the functional dynamics of the protein for TM7. Indeed, in a recent structure obtained for an inward-facing conformation of Glt<sub>Ph</sub>,<sup>12</sup> the most hydrophobic parts of TM7 and TM8 (orange in the right-hand panel of Fig. 1b) have become relocated into the plane of the membrane.

Based on these data, we propose a model for the membrane insertion and folding of Glt<sub>Ph</sub>, where the six N-terminal TM helices insert cotranslationally. In the earliest membrane-inserted state, the most hydrophobic segments of TM2 and TM4 span the membrane, and the more polar parts of TM2 and TM4 likely form amphiphilic interfacial helices. Judging from the structure, TM2 and TM4 probably shift to their final positions in the membrane only upon formation of the trimer. The insertion and folding of the C-terminal region, including the formation of the two reentrant loops, are more difficult to envision, but may involve an early—presumably cotranslational—interaction between TM7 and TM8 and rather dramatic structural rearrangements, both within the domain itself and in its positioning relative to the membrane, as the protein folds.

### Repositioning of TM helices in other membrane proteins

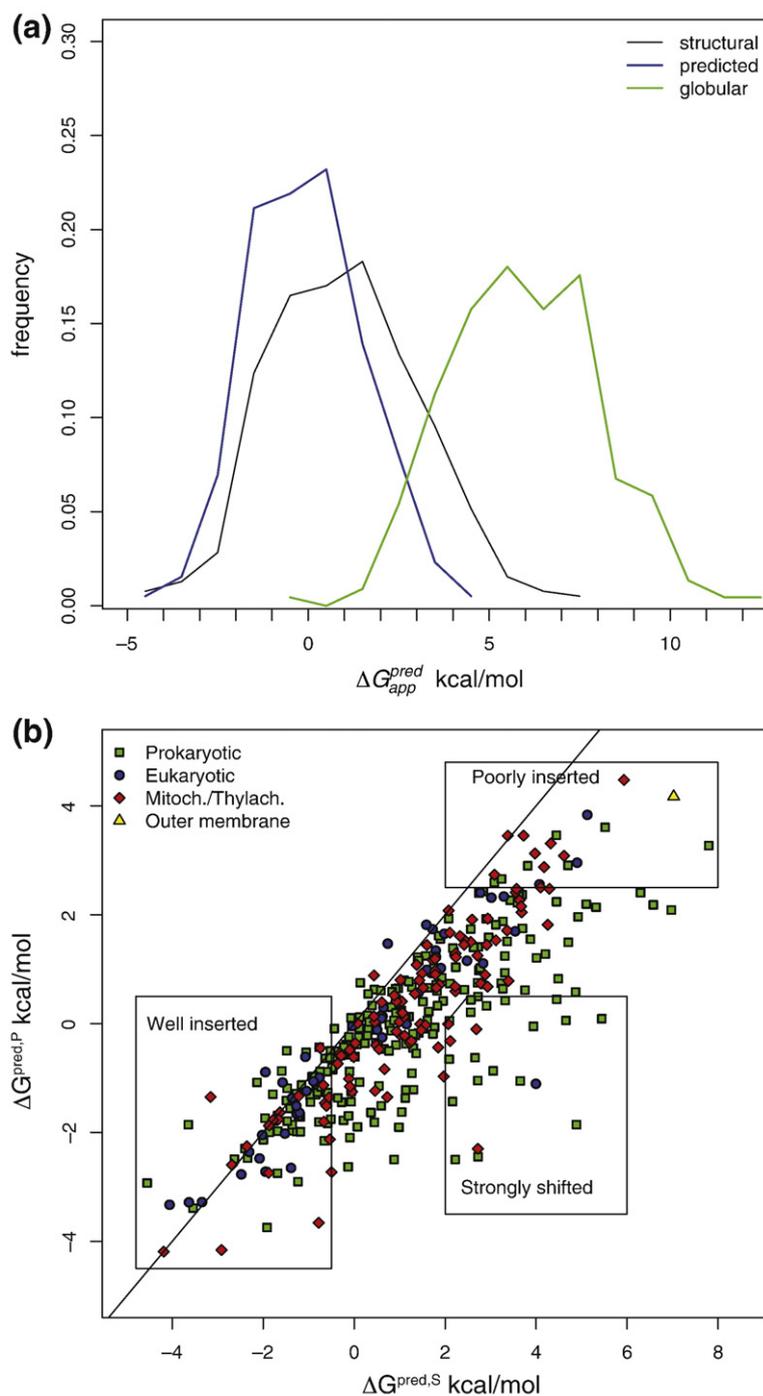
To assess the generality of the observations made on Glt<sub>Ph</sub>, we used the ‘ $\Delta G$  predictor’ to analyze a nonredundant subset of the helix-bundle integral membrane protein structures in the Protein Data Bank<sup>27</sup> (PDB). As observed before,<sup>14</sup> a surprisingly large fraction of the TM helices seem not to be hydrophobic enough to be inserted into the membrane by themselves (Fig. 3a). However, if overlapping, more hydrophobic segments are included, the fraction of helices with high insertion propensity ( $\Delta G_{\text{app}}^{\text{pred}} < 0$  kcal/mol) is increased, suggesting that repositioning of TM helices relative to the membrane occurs rather frequently during folding.

We further searched the membrane protein structures in the PDB for three extreme kinds of TM helices: ‘well-inserted’ helices ( $\Delta G_{\text{app}}^{\text{pred,S}} < -0.5$  kcal/mol), ‘strongly shifted’ helices ( $\Delta G_{\text{app}}^{\text{pred,P}} < 0.5$  kcal/mol;  $\Delta G_{\text{app}}^{\text{pred,S}} > 2$  kcal/mol; and  $\Delta G_{\text{app}}^{\text{pred,S}} - \Delta G_{\text{app}}^{\text{pred,P}} > 2$  kcal/mol), and ‘poorly inserted’ helices ( $\Delta G_{\text{app}}^{\text{pred,P}} > 2.5$  kcal/mol), where S and P denote the TM helix found in the structure and the TM helix as predicted by the ‘ $\Delta G$  predictor,’ respectively (Fig. 3b). As seen in Table 2, well-inserted TM helices are relatively more exposed to the lipid in

the folded structure and have a low coil content. Strongly shifted TM helices, on the other hand, have a high coil content, and many are found in channels and transporters. As the presence of polar residues and short coils often indicates functionally important segments<sup>28</sup> (Illergård *et al.*, submitted), repositioning of TM helices during folding could be a means for bringing functionally important regions of low hydrophobicity into the membrane. Poorly inserted TM helices are mainly found in some nuclearly encoded mitochondrial inner membrane proteins and in the only helix-bundle outer membrane protein known to date (Wza) (see the text below) (Fig. 3b). The relatively low hydrophobicity of many imported mitochondrial inner membrane proteins has been noted before.<sup>29,30</sup> In proteins with only one to two TM helices (either monomers or dimeric one-TM-helix proteins), nearly all of the TM helices belong to the well-inserted group and are not predicted to shift (data not shown).

In Glt<sub>Ph</sub>, the strongly shifted segments form extensions of the TM helices and protrude from the membrane (Figs. 1b and 4a). This is not the case for all strongly shifted segments, however. For example, the potentially shifted, more hydrophobic part of TM3 in the MalG subunit of the maltose transporter<sup>32</sup> forms an interfacial helix (Fig. 4b). In the multidrug transporter EmrD<sup>33</sup> (Fig. 4c), TM9 and TM10 seem to have shifted in such a way that a segment that may initially have formed the connecting loop between TM8 and TM9 (red) has become part of TM9, while a segment that may have been part of the initially inserted segment of TM9 (green) has become the connecting loop between TM9 and TM10 in the folded structure. A fourth case is found in the ClcA H<sup>+</sup>/Cl<sup>-</sup> exchange transporter,<sup>34</sup> where a hydrophobic ‘helical hairpin’ appears to have shifted partly out of the membrane in the folded structure (Fig. 4d). This latter case also serves to illustrate the fine line between a reentrant loop and a pair of closely spaced TM helices; obviously, the distinction is somewhat fuzzy.

Among the poorly inserted TM helices, the case of the homooctameric  $\alpha$ -helical outer membrane protein Wza<sup>35</sup> (Fig. 4e) is particularly instructive, since the C-terminal TM helix in the Wza monomer must not be recognized by the SecYEG translocon as a TM helix during its passage through the inner bacterial membrane. Instead, this highly amphiphilic helix forms a helix bundle with an apolar outer surface only during oligomerization in the outer membrane. Still, roughly half of the poorly inserted segments are from bacterial inner membrane proteins inserted by the SecYEG translocon. Many of these, such as TM11 from EmrD (Fig. 4f), contain several glycines, raising the question of whether the cumulative penalty of multiple Gly residues may be overestimated by the ‘ $\Delta G$  predictor.’ In ClcA, a markedly polar reentrant loop formed by TM3 and TM4 is probably posttranslationally inserted<sup>38</sup> (Fig. 4g). The highly charged S4 voltage sensor helix in the Kv2.1 K<sup>+</sup>-channel<sup>36</sup> (Fig. 4h) may require the presence of neighboring TM helices for efficient assembly.<sup>39,40</sup>



**Fig. 3.** Analysis of TM helices in known structures. (a)  $\Delta G_{app}^{pred}$  distributions for membrane-embedded TM helices as defined in the OPM structural database (black), for the overlapping TM segments as predicted by the ' $\Delta G$  predictor' (red), and for the segment with the lowest  $\Delta G_{app}^{pred}$  value in each of 222 non-membrane segments longer than 15 residues in the OPM structures (green). (b)  $\Delta G_{app}^{pred,P}$  versus  $\Delta G_{app}^{pred,S}$  values for the TM helices in the proteins listed in Table S1. Three classes of TM helices ('well inserted,' 'strongly shifted,' and 'poorly inserted') are indicated on the plot. The provenance of each helix is indicated by the color code.

While there are quite a few examples of strongly shifted and poorly inserting TM helices in known membrane protein structures, we have found only one case where a segment with  $\Delta G_{app}^{pred} < 0$  kcal/mol does not form a TM helix in our data set. This is an interfacial helix in chain D of *Thermosynechococcus elongatus* photosystem II<sup>37</sup> (Fig. 4i).

## Discussion

Our analysis of the *P. horikoshii* Glt<sub>Ph</sub> glutamate transporter homolog strongly suggests that TM

helices may undergo a rather dramatic repositioning in the membrane during the folding and oligomerization process; theoretical analysis suggests that similar phenomena may not be uncommon among membrane proteins in general. Large-scale conformational changes, including repositioning of TM helices during the catalytic cycle of the Ca<sup>2+</sup>-ATPase, have been documented by X-ray crystallography,<sup>41</sup> and our observations generalize this dynamic picture of membrane protein structure to include also the folding stages.

Repositioning of TM helices relative to the lipid bilayer during posttranslational folding and oligo-

**Table 2.** Properties of ‘well inserted,’ ‘strongly shifted,’ and ‘poorly inserted’ TM segments

Type	Number of helices	$\langle \Delta \Delta G \rangle$ (kcal/mol)	$\langle \text{Shift} \rangle$	$\langle \text{RSA} \rangle$ ( $\text{\AA}^2$ )	$\langle \text{Coil} \rangle$ (%)
Well inserted	99	-0.2	1.9	34	6
Strongly shifted	24	-3.7	7.6	27	13
Poorly inserted	20	-1.4	4.6	25	9
Other	245	-1.1	3.6	25	8
Total	388	-1.1	3.4	27	8

The different groups are defined in Fig. 3b.  $\langle \Delta \Delta G \rangle$  is the average difference  $\Delta G_{\text{app}}^{\text{pred,P}} - \Delta G_{\text{app}}^{\text{pred,S}}$  for the TM segments.  $\langle \text{Shift} \rangle$  is the difference in the average number of residues between the center of the membrane-embedded TM segment as annotated in the OPM structure database and the center of the membrane-embedded TM segment as predicted by the ‘ $\Delta G$  predictor.’  $\langle \text{RSA} \rangle$  is the average exposed surface area of the TM helices.  $\langle \text{Coil} \rangle$  is the average percentage of residues in a coil conformation.

merization processes provides a convenient way for nonhydrophobic polypeptide segments to become lodged within the membrane without being exposed to lipid and may in part explain the puzzling observation that many TM helices in the known membrane protein structures are not sufficiently hydrophobic to insert into the membrane by themselves.<sup>14</sup>

From the point of view of structure prediction, repositioning of TM helices may explain why topology predictors are, at best, mediocre at predicting the exact length and location of TM helices as they are found in high-resolution structures,<sup>42</sup> and raises the specter that prediction of membrane protein three-dimensional structure will have to allow for substantial repositioning of the TM helices compared to a starting model where the TM helices are placed in the membrane based on hydrophobicity alone.

## Materials and Methods

### Enzymes and chemicals

Oligonucleotides were obtained from Sigma-Aldrich, Cybergene AB, and MWG Biotech AG. All enzymes were obtained from Fermentas, except for Phusion DNA polymerase (Finnzymes Oy). The plasmid pGEM-1 and the TNT<sup>®</sup> Quick coupled transcription/translation system were obtained from Promega. [<sup>35</sup>S]Met was obtained from Perkin-Elmer.

### DNA manipulation

Double-stranded oligonucleotides encoding the different Glt<sub>ph</sub> H-segments were introduced into the *lepB* gene as

amplified PCR fragments using primers complementary to the 5'-end and the 3'-end of the selected part of the *glt<sub>ph</sub>* gene. Fragments were amplified using Phusion DNA polymerase (Finnzymes Oy) and cloned into pGEM1 containing the *Lep* constructs as a SpeI-KpnI fragment.<sup>15,18</sup>

The full-length *glt<sub>ph</sub>* gene, with the seven-His codon substitutions described<sup>10</sup> and with the last eight codons added, was similarly cloned into a NcoI/SmaI-restricted pGEM1 vector. PDB structure 2NWL<sup>31</sup> has an unreported substitution (Q14I) compared to this sequence. The amplified DNA products were purified using the QIAquick PCR Purification kit from QIAGEN (Hilden, Germany). All constructs were confirmed by sequencing of plasmid DNA at Eurofins MWG Operon (Ebersberg, Germany).

To measure the MGD of Glt<sub>ph</sub> TM4, we introduced the segment A104-N199 from Glt<sub>ph</sub> (including TM4 and its flanking loops) as an H-segment in the *Lep*<sup>I</sup> and *Lep*<sup>II</sup> constructs. The experiment was also performed in the intact Glt<sub>ph</sub> protein. The C-terminal-engineered glycosylation acceptor site in the *Lep* constructs was removed, and new sites were introduced in different positions N-terminal of TM4 in the *Lep*<sup>I</sup> construct and in full-length Glt<sub>ph</sub> and C-terminal of TM4 in the *Lep*<sup>II</sup> construct.<sup>23</sup> Any proline residues that would have been present in engineered YNX(T/S)Y sequons were mutated to S (X position) or A (Y positions).<sup>43</sup>

### Expression *in vitro*

Constructs in pGEM1 were transcribed and translated using the TNT<sup>®</sup> Quick coupled system (Promega). DNA template (200 ng), 1  $\mu$ l of [<sup>35</sup>S]Met (15  $\mu$ Ci), and 1  $\mu$ l of dog pancreas RMs were added to 10  $\mu$ l of lysate at the start of the reaction, and samples were incubated for 90 min at 30 °C.<sup>18</sup>

Translation products were analyzed by SDS-PAGE, and proteins were visualized in a Fuji FLA-3000 Phosphor-Imager using the Image Reader V1.8J/Image Gauge V 3.45 software. Further analysis was carried out using Multi-

**Fig. 4.** Examples of potential rearrangements. Segments predicted by the ‘ $\Delta G$  predictor’ to span the membrane (extended up to the membrane boundary) are shown color-coded according to  $\Delta G_{\text{app}}^{\text{pred}}$  values. The color reflects the  $\Delta G_{\text{app}}^{\text{pred}}$  value of the segment centered on the residue in question. Segments that are closer to the protein’s N-terminus or C-terminus than to the half-width of the sliding window used by the predictor are shown in black. The approximate location of the membrane (according to the OPM database) is indicated by red (‘outside’) and blue (‘inside’) lines; ‘inside’ is taken as the side of the membrane from which the protein becomes inserted. In some cases, parts of the protein have been removed to show the relevant segments more clearly. (a) TM2 and TM4 in Glt<sub>ph</sub><sup>31</sup> (2NWL). (b) TM3 in the maltose transporter chain G<sup>32</sup> (2R6G). (c) TM9, TM10, and TM12 in the EmrD multidrug transporter<sup>33</sup> (2GFP). (d) TM8 and TM9 in the ClcA H<sup>+</sup>/Cl<sup>-</sup> exchange transporter<sup>34</sup> (1OTS). (e) C-terminal helix in the Wza outer membrane protein<sup>35</sup> (2J58). (f) TM11 in the EmrD multidrug transporter<sup>33</sup> (2GFP). (g) TM3 and TM4 in the ClcA H<sup>+</sup>/Cl<sup>-</sup> exchange transporter<sup>34</sup> (1OTS). (h) The Kv2.1 S4 voltage sensor helix in the chimeric Kv2.1-1.2 K<sup>+</sup> channel<sup>36</sup> (2R9R). (i) Interfacial helix in photosystem II, chain D<sup>37</sup> (2AXT). (e–h) The key hydrophilic residues are shown as green sticks.

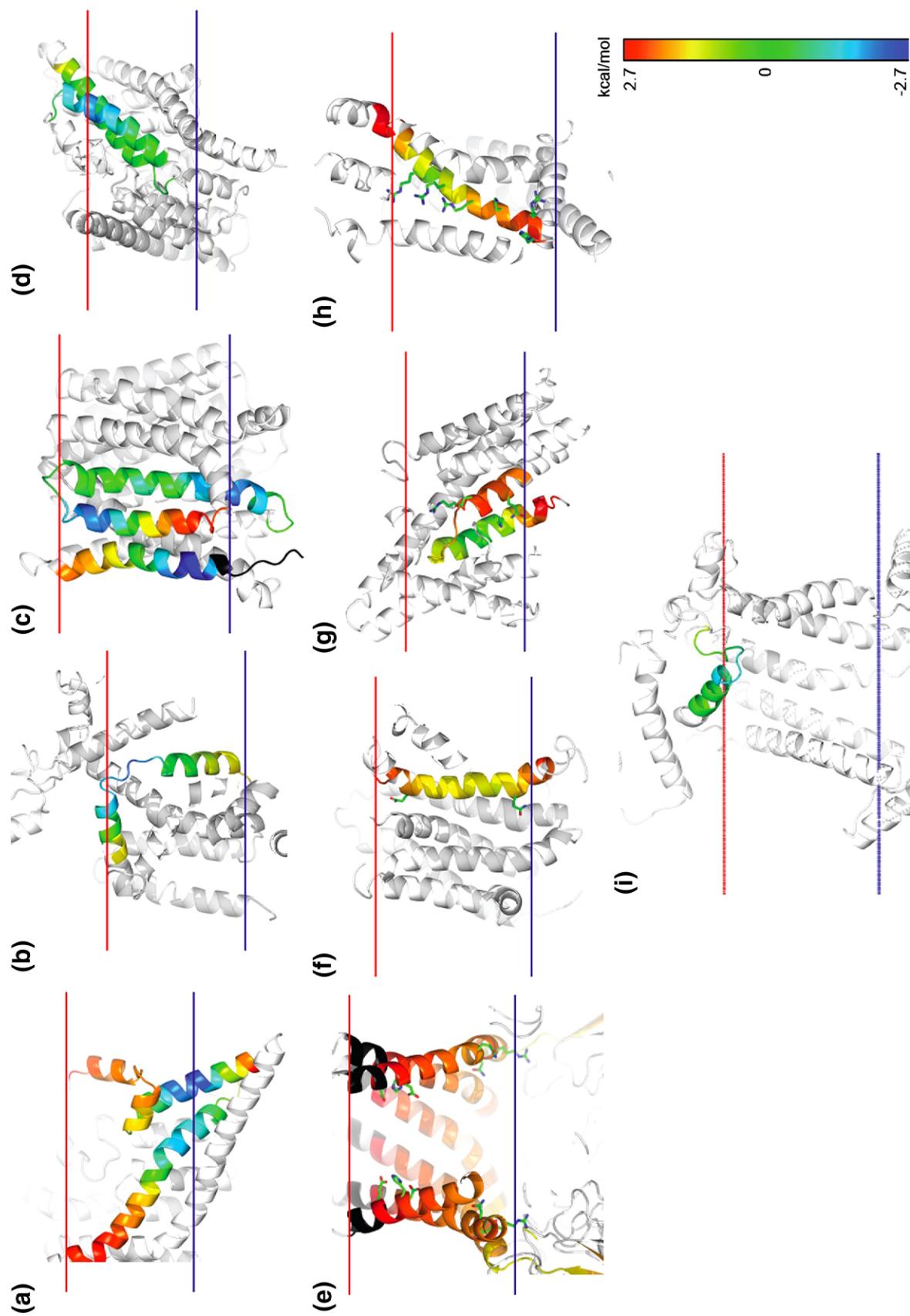


Fig. 4 (legend on previous page)

Gauge (Fujifilm) and Qtiplot<sup>‡</sup>. The apparent membrane-insertion free energy ( $\Delta G_{\text{app}}$ ) of a given H-segment was calculated as described in the main text. On average, the glycosylation levels vary by no more than  $\pm 5\%$  between repeat experiments, corresponding to a standard deviation of  $\pm 0.25$  kcal/mol in the  $\Delta G_{\text{app}}$  values. All values were calculated as the mean values from at least two independent experiments.  $\Delta G_{\text{app}}$  values obtained with the Lep<sup>II</sup> construct are generally  $\sim 1$  kcal/mol lower than those obtained with the Lep<sup>I</sup> construct,<sup>15</sup> and all values obtained with the former were increased by 1 kcal/mol. Table 1 presents both experimentally measured and predicted  $\Delta G_{\text{app}}$  values, with the latter obtained using the ‘ $\Delta G$  predictor<sup>§</sup>.’

## Bioinformatics

Initially, 136  $\alpha$ -helical TM protein structures containing 601 polypeptide chains with TM segments were downloaded from the OPM database<sup>17</sup> (version April 2008). Polyalanine chains, theoretical models, and obsolete entries (as defined in PDB) were excluded. In addition, fragments (1D6G, 1ORS, 2AHY, 1R3J, and 1S5H), very-low-resolution structures ( $>4.2$  Å; 1IFK), and otherwise problematic structures (2QFI, 1ORQ, 2A0, 1YEW, 2OAU) were removed. From this data set, the highest-resolution structure for each OPM family was chosen, resulting in 59 PDB files containing 236 chains. This data set was homology-reduced at the 40% sequence identity level by Cd-hit,<sup>44</sup> leaving a total of 101 chains (Table S1). The corresponding amino acid sequences were downloaded from the PDB.<sup>27</sup> The membrane-embedded part of each TM helix was obtained from the OPM database, and topologies were assigned as described.<sup>45</sup>

$\Delta G_{\text{app}}^{\text{pred}}$  values were calculated using the ‘ $\Delta G$  predictor.’<sup>14</sup> In short, position-dependent contributions to the overall insertion free energy for each amino acid are summed across the sequence of a TM segment, after which corrections for segment length and hydrophobic moment are applied. In the calculation of  $\Delta G_{\text{app}}^{\text{pred}}$  for experimental constructs, the ‘allow subsequence’ option was turned on. For the theoretical analysis of known structures, the ‘allow subsequence’ option was turned on or off, depending on segment type (see the details below).

To generate the  $\Delta G_{\text{app}}^{\text{pred}}$  profile in Fig. 1d, we used the ‘full protein scan’ version of the ‘ $\Delta G$  predictor,’ with the window length varying from 15 to 25 residues.

$\Delta G_{\text{app}}^{\text{pred,S}}$  values were calculated for the membrane-embedded parts of the TM helices as annotated in the OPM database. To find segments that are likely to be initially inserted into the membrane, we added 14-residue-long flanks to each end of membrane-embedded TM segments, and the 15- to 21-residue-long subsequence with the lowest predicted  $\Delta G_{\text{app}}^{\text{pred,P}}$  values was selected. TM segments with an annotated length of  $<10$  residues were excluded from the analysis.  $\Delta G_{\text{app}}^{\text{pred}}$  values were also calculated for reentrant loops and globular regions. For reentrant loops, 14 residue flanks were included at each end; for globular regions, each nonmembrane segment with a length of  $>15$  residues annotated in the OPM database was used as a starting segment. Also here, length limitations of 15–21 residues were used in the identification of the segment with the lowest  $\Delta G_{\text{app}}^{\text{pred}}$  value.

Cellular localization annotation was extracted from the OPM database. Secondary structure was assigned by

DSSP.<sup>46</sup> A residue was defined as a coil if the assignment was C, S, T, or B, and as a helix if the assignment was H, G, or I. Relative surface accessibility was calculated by NACCESS 2.1.1<sup>47</sup> using a probe size of 1.4 Å (corresponding to the size of a water molecule; i.e., the same probe size was used for lipid-accessible and water-accessible regions).

The topology prediction for Glt<sub>Ph</sub> in Fig. 1d was obtained with SCAMPI 1.0<sup>11</sup> run in single-sequence mode with default parameters.

All data plots were generated with R<sup>48</sup> and gnuplot<sup>ll</sup>. Three-dimensional structure illustrations were made using PyMOL.<sup>49</sup>

## Acknowledgements

This work was supported by grants from the European Research Council (G.v.H.; ERC-2008-AdG 232648), the Swedish Cancer Foundation (G.v.H.), the Swedish Research Council (G.v.H., A.E., and S.C.), the Swedish Foundation for Strategic Research (G.v.H., A.E., and S.C.), the Academy of Finland (A.K.), and the European Union ERASMUS program (E.T.). We gratefully thank Prof. Bernhard Dobberstein (University of Heidelberg) for providing RMs, Prof. Eric Gouaux (Oregon Health and Science University) for the *glt<sub>Ph</sub>* gene, and Dr. IngMarie Nilsson, Dr. Mirjam Lerch-Bader, Karin Öjemalm, Dr. Andreas Bernsel, and Kristoffer Illergård for technical advice.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.01.042

## References

1. Popot, J. L. & Engelman, D. M. (1990). Membrane protein folding and oligomerization—the 2-stage model. *Biochemistry*, **29**, 4031–4037.
2. Engelman, D. M., Chen, Y., Chin, C. N., Curran, A. R., Dixon, A. M., Dupuy, A. D. *et al.* (2003). Membrane protein folding: beyond the two stage model. *FEBS Lett.* **555**, 122–125.
3. Elofsson, A. & von Heijne, G. (2007). Membrane protein structure: prediction *versus* reality. *Annu. Rev. Biochem.* **76**, 125–140.
4. Viklund, H., Granseth, E. & Elofsson, A. (2006). Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes. *J. Mol. Biol.* **361**, 591–603.
5. Sadlish, H., Pitonzo, D., Johnson, A. E. & Skach, W. R. (2005). Sequential triage of transmembrane segments by Sec61 $\alpha$  during biogenesis of a native multi-spanning membrane protein. *Nat. Struct. Mol. Biol.* **12**, 870–878.

<sup>‡</sup> www.qtiplot.ro

<sup>§</sup> http://www.cbr.su.se/DGpred/

<sup>ll</sup> http://www.gnuplot.info

6. Lu, Y., Turnbull, I. R., Bragin, A., Carveth, K., Verkman, A. S. & Skach, W. R. (2000). Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum. *Mol. Biol. Cell*, **11**, 2973–2985.
7. Bogdanov, M., Heacock, P. N. & Dowhan, W. (2002). A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. *EMBO J.* **21**, 2107–2116.
8. Young, R. (2000). Holins: the protein clocks of bacteriophage infections. *Annu. Rev. Microbiol.* **54**, 799–825.
9. Carveth, K., Buck, T., Anthony, V. & Skach, W. R. (2002). Cooperativity and flexibility of cystic fibrosis transmembrane conductance regulator transmembrane segments participate in membrane localization of a charged residue. *J. Biol. Chem.* **277**, 39507–39514.
10. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. (2004). Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature*, **431**, 811–818.
11. Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G. & Elofsson, A. (2008). Prediction of membrane-protein topology from first principles. *Proc. Natl Acad. Sci. USA*, **105**, 7177–7181.
12. Reyes, N., Ginter, C. & Boudker, O. (2009). Transport mechanism of a bacterial homologue of glutamate transporters. *Nature*, **462**, 880–885.
13. Groeneveld, M. & Slotboom, D. J. (2007). Rigidity of the subunit interfaces of the trimeric glutamate transporter GltT during translocation. *J. Mol. Biol.* **372**, 565–570.
14. Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M. *et al.* (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, **450**, 1026–1030.
15. Lundin, C., Kim, H., Nilsson, I., White, S. & von Heijne, G. (2008). The molecular code for protein insertion in the ER membrane is similar for  $N_{in}$ - $C_{out}$  and  $N_{out}$ - $C_{in}$  transmembrane helices. *Proc. Natl Acad. Sci. USA*, **105**, 15702–15707.
16. Enquist, K., Fransson, M., Boekel, C., Bengtsson, I., Geiger, K., Lang, L. *et al.* (2009). Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *J. Mol. Biol.* **387**, 1153–1164.
17. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. (2006). OPM: orientations of proteins in membranes database. *Bioinformatics*, **22**, 623–625.
18. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H. *et al.* (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, **433**, 377–381.
19. Johansson, M., Nilsson, I. & von Heijne, G. (1993). Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in *Escherichia coli* than in mammalian microsomes. *Mol. Gen. Genet.* **239**, 251–256.
20. 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.
21. Nilsson, I., Sääf, A., Whitley, P., Gafvelin, G., Waller, C. & von Heijne, G. (1998). Proline-induced disruption of a transmembrane  $\alpha$ -helix in its natural environment. *J. Mol. Biol.* **284**, 1165–1175.
22. Armulik, A., Nilsson, I., von Heijne, G. & Johansson, S. (1999). Determination of the border between transmembrane and cytoplasmic domains of human integrin subunits. *J. Biol. Chem.* **274**, 37030–37034.
23. Stefansson, A., Armulik, A., Nilsson, I., von Heijne, G. & Johansson, S. (2004). Determination of N- and C-terminal borders of the transmembrane domain of integrin subunits. *J. Biol. Chem.* **279**, 21200–21205.
24. Lau, T. L., Dua, V. & Ulmer, T. S. (2008). Structure of the integrin  $\alpha$ IIb transmembrane segment. *J. Biol. Chem.* **283**, 16162–16168.
25. Lau, T. L., Partridge, A. W., Ginsberg, M. H. & Ulmer, T. S. (2008). Structure of the integrin  $\beta$ 3 transmembrane segment in phospholipid bicelles and detergent micelles. *Biochemistry*, **47**, 4008–4016.
26. Chavan, M. & Lennarz, W. (2006). The molecular basis of coupling of translocation and N-glycosylation. *Trends Biochem. Sci.* **31**, 17–20.
27. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids. Res.* **28**, 235–242.
28. Kauko, A., Illergard, K. & Elofsson, A. (2008). Coils in the membrane core are conserved and functionally important. *J. Mol. Biol.* **380**, 170–180.
29. von Heijne, G. (1986). Why mitochondria need a genome. *FEBS Lett.* **198**, 1–4.
30. Claros, M. G., Perea, J., Shu, Y. M., Samatey, F. A., Popot, J. L. & Jacq, C. (1995). Limitations to *in vivo* import of hydrophobic proteins into yeast mitochondria—the case of a cytoplasmically synthesized apocytochrome *b*. *Eur. J. Biochem.* **228**, 762–771.
31. Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K. & Gouaux, E. (2007). Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature*, **445**, 387–393.
32. Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L. & Chen, J. (2007). Crystal structure of a catalytic intermediate of the maltose transporter. *Nature*, **450**, 515–521.
33. Yin, Y., He, X., Szewczyk, P., Nguyen, T. & Chang, G. (2006). Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science*, **312**, 741–744.
34. Dutzler, R., Campbell, E. B. & MacKinnon, R. (2003). Gating the selectivity filter in ClC chloride channels. *Science*, **300**, 108–112.
35. Dong, C., Beis, K., Nesper, J., Brunkan-Lamontagne, A. L., Clarke, B. R., Whitfield, C. & Naismith, J. H. (2006). Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein. *Nature*, **444**, 226–229.
36. Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. (2007). Atomic structure of a voltage-dependent  $K^+$  channel in a lipid membrane-like environment. *Nature*, **450**, 376–382.
37. Loll, B., Kern, J., Saenger, W., Zouni, A. & Biesiadka, J. (2005). Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature*, **438**, 1040–1044.
38. Cassel, M., Seppälä, S. & von Heijne, G. (2008). Confronting fusion protein-based membrane protein topology mapping with reality: the *Escherichia coli* ClcA  $H^+$ /Cl $^-$  exchange transporter. *J. Mol. Biol.* **381**, 860–866.
39. Hessa, T., White, S. H. & von Heijne, G. (2005). Membrane insertion of a potassium channel voltage sensor. *Science*, **307**, 1427.
40. Zhang, L., Sato, Y., Hessa, T., von Heijne, G., Lee, J. K., Kodama, I. *et al.* (2007). Contribution of hydrophobic and electrostatic interactions to the membrane integration of the Shaker  $K^+$  channel voltage sensor domain. *Proc. Natl Acad. Sci. USA*, **104**, 8263–8268.
41. Olesen, C., Picard, M., Winther, A. M., Gyrop, C., Morth, J. P., Oxvig, C. *et al.* (2007). The structural basis of calcium transport by the calcium pump. *Nature*, **450**, 1036–1042.
42. Papatoukas, C., Granseth, E., Viklund, H. & Elofsson,

- A. (2008). Estimating the length of transmembrane helices using Z-coordinate predictions. *Protein Sci.* **17**, 271–278.
43. Gavel, Y. & von Heijne, G. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites—implications for protein engineering. *Protein Eng.* **3**, 433–442.
44. Li, W. & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
45. Viklund, H. & Elofsson, A. (2008). OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics*, **24**, 1662–1668.
46. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
47. Hubbard, S. J. & Thornton, J. M. (1993). NACCESS, Computer Program Department of Biochemistry and Molecular Biology, University College London, London, UK.
48. R Development Core Team. (2005). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria; <http://www.R-project.org>.
49. DeLano, W. L. (2002). The PyMOL Molecular Graphics System DeLano Scientific, Palo Alto, CA, USA.