Insertion of short transmembrane helices by the Sec61 transloco

Simon Jauda, Mónica Fernández-Vidalb,1, IngMarie Nilssonb, Nadja M. Meindl-Beinkehr, Nadja C. Hübnerb, Douglas J. Tobiasa, Gunnar von Heijnea, and Stephen H. Whiteb,2

aDepartment of Chemistry and the Center for Biomembrane Systems, University of California, Irvine, CA 92697-2025; bDepartment of Physiology and Biophysics and the Center for Biomembrane Systems, University of California, Irvine, CA 92697-4560; and cDepartment of Biochemistry and Biophysics and the Center for Biomembrane Research, Stockholm University, SE-106 91 Stockholm, Sweden

Edited by Donald M. Engelman, Yale University, New Haven, CT, and approved May 28, 2009 (received for review January 19, 2009)

The insertion efficiency of transmembrane (TM) helices by the Sec61 translocon depends on helix amino acid composition, the positions of the amino acids within the helix, and helix length. We have used an in vitro expression system to examine systematically the insertion efficiency of short polyoleucine segments (Lept, n = 4 ... 12) flanked at either end by 4-residue sequences of the form XXPX-LXpXPXX with X = G, N, D, or K. Except for X = K, insertion efficiency (p) is <10% for n < 8, but rises steeply to 100% for n = 12. For X = K, p is already close to 100% for n = 10. A similar pattern is observed for synthetic peptides incorporated into oriented phospholipid bilayer arrays, consistent with the idea that recognition of TM segments by the translocon critically involves physical partitioning of nascent peptide chains into the lipid bilayer. Molecular dynamics simulations suggest that insertion efficiency is determined primarily by the energetic cost of distorting the bilayer in the vicinity of the TM helix. Very short lysine-flanked leucine segments can reduce the energetic cost by extensive hydrogen bonding with water and lipid phosphate groups (snorkeling) and by partial unfolding.

hydrophobic mismatch | membrane protein synthesis | membrane proteins | molecular dynamics simulation | lipid bilayer

The thickness of the hydrocarbon core of bilayer membranes is ~30 Å, which leads to the expectation that transmembrane (TM) helices of helix-bundle membrane proteins (MPs) should be ~20-aa long to span the hydrocarbon core. Consistent with this expectation, the average helix length of α-helical MPs of known 3D structure is 26 ± 5 residues (http://blanco.biomol.ucr.edu/mptopo). A few membrane proteins, most notably the aquaporins (1) and H+]-Cl- exchangers (2) have membrane-buried helices as short as n = 10 residues, corresponding to a helix length of (n - 1) × 1,5 = 13.5 Å. Although there is ample evidence that Sec61/SecY translocons can insert short helices, the biophysical basis of short-helix insertion has not been explored systematically.

Kuroiwa et al. (3) found that the Sec61 translocon could efficiently insert artificial stop-transfer polyoleucine sequences as short as 9 residues across rough microsomal membranes. Chen and Kendall (4) subsequently found that the Escherichia coli SecY translocon is also able to insert 10-residue polyoleucine segments across the inner membrane. In a detailed examination of the molecular code that the Sec61 translocon uses for identifying TM segments, Hessa et al. (5) found that efficient stop-transfer segments comprised of short polyoleucine helices could also be engineered into a multistaping membrane protein, specifically the E. coli leader peptidase (Lep). In a recent model-peptide study, Krishnakumar and London (6) determined the minimum hydrophobic length required for the formation of poly(Leu) or poly(LeuAla) TM helices in small unilamellar vesicles formed from diacyl phosphatidylcholines with chain acyl lengths of 14 to 24 carbons. Their results were consistent with the TM insertion of hydrophobic sequences as short as 11–12 leucine residues in vesicle membranes formed from dioleoylphosphatidylcholine (DOPC).

A limitation of the biochemical and model-peptide studies to date is that they can only be inferentially compared with each other. To overcome this limitation, we have used the same set of TM helices (H-segments) for biochemical, biophysical, and computational studies. By using a microsomal, in vitro expression system to examine systematically the Sec61 insertion efficiency of short polyoleucine segments flanked at either end by 4-residue sequences of the form XXPX-LXpXPXX (n = 4–12; X = G, N, D, or K), we found that helices with n = 10–12 were readily inserted into the membrane. We also synthesized selected sequences from the XXPX-LXpXPXX family and examined by oriented circular dichroism (OCD) the incorporation of the peptides into oriented palmitoyloleoyl phosphopholine (POPC) and POPC:POPG (palmitoyloleoyl phosphatidylglycerol) multibilayers at various concentrations and hydrations. We found that short helices of the XXPX family were indeed stable in a transmembrane configuration in oriented multibilayers. Finally, to explore the physical principles underlying the stability of short helices under conditions of severe hydrophobic mismatch, we carried out molecular dynamics (MD) simulations of some of the peptides in transmembrane orientations. Taken together, the results from this multifaceted approach suggest that the biophysics of hydrophobic mismatch controls the insertion of short helices across the endoplasmic reticulum (ER) membrane.

Results

Integration of Polyoleucine H-Segments Across ER Membranes by the Sec61 Transloco. We first determined insertion efficiencies of systematically designed H-segments of the form XXPX-LXpXPXX (X = G, N, D, or K) across the endoplasmic membrane by using the approach of Hessa et al. (5, 7), summarized in Fig. 1A. In brief, the H-segments were introduced near the middle of the large luminal P2 domain of the model protein, Lep. The protein was then expressed in vitro in the presence of ER-derived dog pancreas rough microsomes (RMs). The insertion probability (efficiency) p was calculated on the basis of the fractions of singly (f1) and doubly (f2) glycosylated proteins by using p = f1/[f1 + f2] determined from quantitative phosphomager scans of SDS/PAGE gels. Hessa et al. (5, 7) showed that p is proportional to the Boltzmann function [1 + exp(ΔGmismatch/RT)]-1, which we consequently used to characterize the probability-of-insertion data.

The central Lept stretch was flanked by the tetrapeptides XXPX and –XPXX to “insulate” it from the surrounding sequence in the Lep model protein. For the present work, we took H-segments with...
X = G flanks as the reference segments. The immediate sequence environment outside the H-segment is quite polar (ENGRILSETSV/H-segment/VPGQONATWI), hence there is no additional hydrophobic flanking segment that could be ‘recruited’ into the membrane-spanning H-segment. Examples of SDS/PAGE gels showing the translation products of GGPG-L-x-GPGG constructs with n = 8, 10, and 12. Plasmids encoding the Lep/H-segment constructs were transcribed and translated in vitro in the presence (−) and absence (+) of dog pancreas rough microsomes (DRM). Translation products were analyzed by SDS/PAGE. Bands of unglycosylated protein are indicated by white dots; singly (1g) and doubly (2g) glycosylated proteins are indicated by 1 and 2 red dots, respectively. The data were quantitated by scanning the gels in a phosphorimager. The probability (efficiency) of membrane insertion is given by P = 1g/(1g + 2g). Mean values from 2 independent experiments were used for computing P values. On average, glycosylation levels vary by about ±2% between repeat experiments.

Fig. 2A shows plots of the probability of insertion as a function of n for X = G, N, D, and K. The probabilities of insertion for the reference set of H-segments with X = G are plotted as solid black curves. This distribution means that the insertion process has the appearance of equilibrium between the inserted and translocated states measured by the relative amounts of singly and doubly glycosylated translation products. The data show that the probability of insertion is ~0 for n = 8 or fewer leucines, but rises steeply to ~90% for n = 12 or greater. These data confirm the earlier observations of Kuroiwa et al. (3) and Chen and Kendall (4). Assuming an α-helix, 12 leucines corresponds to a length of 16.5 Å, which cannot span the typical 30-Å-thick hydrocarbon core of lipid bilayers without considerable negative hydrophobic mismatch.

Fig. 2B shows wavelength minima of the OCD spectra as a function of the number of leucines for X = G and X = K. Approximate representative positions of the minima are shown in Fig. 3 A and B. Although the interpretation of the OCD spectra for n = 8 (Fig. 3) is uncertain, this plot reveals a structural change that approximates the shape of the probability-of-insertion curves shown here. All OCD spectra for n = 10 are consistent with TM α-helices, as shown by the examples in Fig. 3.

Fig. 2C shows relative free-energy costs of deforming a POPC bilayer to accommodate XXPX-L-x-XPXX peptides of different lengths and compositions. Relative free energy is the computed free energy for a given n divided by the free energy computed for GL20. By using the fits of Eq. S1 to the δf(t) curves of Fig. 5, we calculated the differences in deformation free energies of the lipid-protein system by using Eq. S2. We calculated the difference in free energy of all systems relative to data from a simulation of a GL20 peptide in a bilayer, ΔGprotein/ΔGGL20 = GL20 was chosen as a reference because the hydrophobic mismatch is small. For this reason, the relative free energies for n = 20 are zero in this figure.

Replacement of the flanking Gly residues with Asn (Fig. 2, green triangles) had little effect on the probability of insertion, whereas
replacement with Asp (Fig. 2, red circles) caused a small shift of the insertion probabilities toward longer runs of leucines. In contrast, replacement of the flanking glycines with Lys residues (Fig. 2, inverted blue triangles) caused a shift of the probability curve to shorter lengths so that 90% efficiency was achieved for \( n = 10 \) leucines. If these results are due to physical interactions of the peptides with the membrane, then it should be possible to construct model membrane systems in which short synthetic peptides are stable across lipid bilayer membranes.

**Integration of Synthetic Polyleucine H-Segments Across Oriented Phospholipid Membranes.** We synthesized 2 families of H-segments, each containing \( n = 6, 8, 10, \) or 12 leucines. The glycine family (designated GL\( n \)) had X = Gly flanks; the lysine family (designated KL\( n \)) had X = Lys flanks. In methanol and trifluoroethanol solutions, the GL\( n \) and KL\( n \) peptide families were \( \alpha \)-helical for all values of \( n \) as determined by solution CD spectra (Fig. S1). The conformations of the GL\( n \) family in phosphate buffer depended on \( n \) (Fig. S2A). For \( n = 6 \) or 8, the CD spectra were consistent with \( \beta \)-sheet structure, probably due to aggregation, whereas for \( n = 10 \) or 12, the conformations were predominantly \( \alpha \)-helical. In contrast, the CD spectra of the KL\( n \) family in phosphate buffer were consistent with random coil for \( n = 6 \) or 8 and predominately \( \alpha \)-helical for \( n = 10 \) or 12 (Fig. S2B). We conclude from these measurements that a conformational transition to \( \alpha \)-helical occurs for both peptide families between \( n = 8 \) and \( n = 10 \).

The ability of the peptides to insert across a lipid bilayer was examined by means of OCD spectroscopy of peptides in oriented multibilayer arrays that were formed by depositing methanol solutions of lipid and peptide on quartz substrates. Highly oriented arrays form spontaneously after slow evaporation of the methanol followed by hydration via vapor-phase equilibration with saturated salt solutions at 66% or 93% relative humidity (RH) (9, 10). Although the bilayers are not fully hydrated at these RHs, the disposition of incorporated peptides is generally the same as in excess water (9, 11). Bilayers were composed of either POPC or 1:1 POPC:POPG; peptide:lipid concentrations ranged from 1:12 to 1:100. Bilayer/peptide multilayers (consisting of stacks of \( \approx 1,000 \) bilayers) formed in this way are highly aligned parallel to the quartz substrate as determined by X-ray diffraction.

OCD spectra were obtained for all peptides in both POPC and POPC:POPG at all peptide:lipid ratios (Fig. S3). Representative OCD data for the GL\( n \) and KL\( n \) peptide families (\( n = 6 \) or 12) are shown in Fig. 3. Shown as solid curves in each panel are the theoretically expected spectra for TM helices aligned perpendicular and parallel to the bilayer. The nearly identical spectra for GL12 and KL12 (Fig. 3A and B, respectively) show that the H-segments are predominately helical and aligned normal to the bilayer plane. The GL12 spectra adhere slightly less accurately to the ideal helix-normal spectra than the KL12 peptides, possibly indicating improved helicity due to the lysine residues. The spectra for GL6 and KL6, on the other hand, are not readily interpretable; OCD spectra such as these have not been previously reported as far as we know.

We characterized the changes in the OCD spectra empirically by plotting the positions of wavelength minima as a function of \( n \) (Fig. 2B). The curves are remarkably similar to the probability curves of

---

**Fig. 3.** Oriented circular dichroism (OCD) spectra of synthetic XXPX-L\( X \)-XPX peptides in oriented multimembranellar POPC bilayer arrays for \( X = G \) and K. Oriented multimembranellar bilayer arrays were formed on 2.5-cm-diameter quartz plates, as described in detail elsewhere (9, 10, 27). Solid curves (labeled in A) represent theoretical spectra for transmembrane helices oriented normal and parallel to the membrane plane, as indicated. Red arrows in A and D indicate the positions of the minima in OCD curves. (A) Spectra for GL12 at various peptide/lipid mole ratios. The spectra indicate that the peptides are predominantly \( \alpha \)-helical and normal to the bilayer plane. (B) Spectra for KL12 at various peptide/lipid mole ratios. As for GL12, the spectra are consistent with TM \( \alpha \)-helices. (C and D) Spectra for GL6 and KL6, respectively. GL12: X = G, \( n = 12 \); GL6: X = G, \( n = 6 \); KL12: X = K, \( n = 12 \); KL6: X = K, \( n = 6 \).

---

**Molecular Dynamics Simulations of Transmembrane Polyethylene H-Segments.** What is the conformation of the synthetic peptides in lipid bilayers, and how does the lipid bilayer adapt to them? To examine these questions, we carried out all-atom molecular dynamics simulations of the GL\( n \) and KL\( n \) peptides in POPC lipid bilayers for \( n = 6, 8, 10, 12, \) and 20 (Figs. S4 and S5). The transbilayer phosphate-to-phosphate thickness of ER lipid membranes (12) is \( \approx 38 \) Å, similar to the thicknesses of DOPC and POPC at moderate hydinations (13). The number of atoms, the simulation box size (\( a \times b \times c \)) at equilibrium, the simulation time \( t_e \) (ns) required to reach equilibrium, and the lateral area per lipid (\( A_{\text{lipid}} \)) for each system are reported in Table 1. If it were technically feasible to run all-atom simulations for milliseconds or seconds, one might observe transitions between membrane-parallel interfacial helices and transmembrane helices. However, all-atom simulations on millisecond or longer time scales are not yet feasible. Nonetheless, our simulations provide critical insights into the behavior of bilayers in the vicinity of our H-segment transmembrane helices.
The importance of water is revealed by snapshots of the 4 systems at the end of the 10-ns production run (Fig. 4 Left). Water (orange) penetrates deeply into the bilayer for GL6 and even through the bilayer for KL6, because of hydration of the peptide backbone. Water penetration is much less for the longer GL12 and KL12 peptides. This water penetration, shown quantitatively in Fig. 4 Right, results from the formation of cavities around each peptide due to the bilayer collapsing around the peptides (see below) through hydrophobic mismatch. Lysine groups are much longer and more polar than glycine, and these properties must necessarily have an important effect on a peptide’s ability to stabilize itself through hydrogen-bonding interactions, as suggested by Fig. 4 and quantified by the number of hydrogen bonds \( n_{\text{HB}} \) in Table 1. There are nearly twice as many hydrogen-bonding interactions for the KL12 peptides, which means that the lysine residues add additional stability that backbone interactions alone cannot provide in the case of the GL12 peptides. These lysine interactions are a complex form of snorkeling (14, 15) between lysine sidechains and lipid phosphate groups. A complete analysis of the hydrogen bonding of the peptides is included in Fig. S7.

The shape of the bilayer surrounding the GL12 and KL12 peptides reveals the adaptation of bilayer and peptide-to-hydrophobic mismatch (Fig. 5). The mismatch is made possible by changes in the lipid order parameters in the peptide neighborhood (Fig. S6). The shape of the bilayer around each peptide was analyzed by using a model based on the Landau-de Gennes theory, in which the free energy of lipid–protein interactions is determined by minimizing the functional integral of the local free energy density (16) (see SI Text). The variation of membrane thickness \( \zeta \) as a function of the radial distance \( r \) from the peptide obtained from the shape analysis (Fig. 5) provides the entrée for parameterizing and computing the free energy change \( \Delta G \) associated with mismatch deformation relative to GL20 and KL20 (see SI Text). In simple terms, the energetic cost of deformation increases as the square of the difference between the hydrophobic mismatch between the peptide and the unperturbed lipid bilayer. Qualitatively, Fig. S7 shows that the GL12 and GL6 mismatch is higher than the mismatch between KL12 and KL6, meaning that it is energetically more costly to deform the bilayer around the GLn peptides than that around the KLn peptides. The computed relative free energies of deformation for GLn and KLn are compared in Fig. 2C. For values of \( n \) between 20 and 12, the mismatch energetic cost is about the same for the 2 peptide families. But for \( n < 12 \), the free energy cost for GLn is persistently greater than that for KLn. This result implies that it should be energetically easier for the translocon to insert short KLn peptides than to insert GLn peptides, which is in fact the case (Fig. 2A).

### Discussion

We have used a consistent set of model TM segments (H-segments) for molecular biological, biophysical, and computational studies to gain insights into the mechanism of insertion of short polypeptide segments across the ER membrane by the Sec61 translocon. We used a microsomal in vitro expression system to examine systematically the Sec61 insertion efficiency of short polypeptide segments \( L_n, n = 4 \ldots 12 \) flanked at either end by 4-residue sequences of the form XXPX-Lx-PXXX with \( X = G, N, D, \) or \( K \) (Fig. 1). With the exception of \( X = K \), insertion efficiency rises steeply between \( n = 8 \) to 12. For \( X = K \), the range of steep rise shifts to shorter lengths \( (n = 8 \to 10) \).

To explore the question of whether or not simple physical principles can explain the in vitro Sec61 results, we synthesized members of the GLn and KLn families of sequences, and examined by OCD their incorporation into POPC- and POPC:POPG-oriented multibilayers at various protein/lipid ratios and hydrations. The results were largely independent of lipid and hydration (Fig. S3). A comparison of the OCD data for GL6 and GL12 with KL6 and KL12 was revealing (Figs. 3 and 4). For \( n \approx 10 \), peptides incorporated into oriented lipid bilayers had OCD spectra unambiguously characteristic of transmembrane helices. This result shows that insertion of TM helices as short as 10 leucines by the Sec61 translocon is consistent with a physical partitioning of TM segments between translocon and lipid bilayer (5, 7).

OCD spectra for \( n = 8 \) were ambiguous with respect to orientation and secondary structure, but solution CD spectra suggested that that the peptides are likely rich in \( \beta \) or extended conformations. A plot of the OCD wavelength minima against \( n \) (Fig. 4) revealed a sharp break between \( n = 8 \) and \( n = 10 \), mirroring the length-dependent Sec61 probability of insertion (Fig. 2). The solution spectra (Fig. S2) and the MD simulations (see below) raise the possibility that the shorter sequences cross the membrane with imperfect \( \alpha \)-helical or extended conformations. But whatever their

### Table 1. Summary of the molecular dynamics simulations for various properties of GLn and KLn simulations in POPC bilayers

<table>
<thead>
<tr>
<th>Peptide</th>
<th>( n_{\text{atoms}} )</th>
<th>( a \times b \times c, \text{Å}^3 )</th>
<th>( t_e, \text{ns} )</th>
<th>( A_{\text{lipid}}, \text{Å}^2 )</th>
<th>( n_{\text{HB}} )</th>
<th>( n_{\text{HB}}^\text{a} )</th>
<th>( n_{\text{hbN}} )</th>
<th>( n_{\text{HB}}^\text{b} )</th>
<th>( n_{\text{HB}}^\text{c} )</th>
<th>Overlap</th>
<th>( n_{\text{HB}}^\text{d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GL6</td>
<td>62,915</td>
<td>82.4 × 83.8 × 84.6</td>
<td>27</td>
<td>49.3 ± 0.2</td>
<td>13</td>
<td>4.2</td>
<td>6.7</td>
<td>—</td>
<td>1.4</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td>GL8</td>
<td>62,951</td>
<td>84.0 × 83.4 × 83.4</td>
<td>25</td>
<td>50.1 ± 0.2</td>
<td>8</td>
<td>4.5</td>
<td>7.9</td>
<td>2.0</td>
<td>10.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL10</td>
<td>62,989</td>
<td>83.4 × 85.9 × 82.7</td>
<td>23</td>
<td>50.6 ± 0.2</td>
<td>3</td>
<td>5.5</td>
<td>8.5</td>
<td>3.3</td>
<td>10.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL12</td>
<td>63,027</td>
<td>86.6 × 81.1 × 81.5</td>
<td>17</td>
<td>51.4 ± 0.2</td>
<td>3</td>
<td>7.3</td>
<td>10.9</td>
<td>3.8</td>
<td>14.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL20</td>
<td>63,179</td>
<td>88.4 × 87.5 × 77.0</td>
<td>15</td>
<td>55.2 ± 0.4</td>
<td>0</td>
<td>10.2</td>
<td>13.8</td>
<td>—</td>
<td>7.6</td>
<td>16.3</td>
<td></td>
</tr>
<tr>
<td>KL6</td>
<td>63,009</td>
<td>83.9 × 83.2 × 83.9</td>
<td>18</td>
<td>49.9 ± 0.2</td>
<td>6</td>
<td>4.1</td>
<td>7.9</td>
<td>11.9</td>
<td>0.9</td>
<td>23.0</td>
<td></td>
</tr>
<tr>
<td>KL8</td>
<td>63,047</td>
<td>86.1 × 83.8 × 83.1</td>
<td>7</td>
<td>51.5 ± 0.2</td>
<td>13</td>
<td>3.9</td>
<td>7.1</td>
<td>12.0</td>
<td>1.6</td>
<td>21.3</td>
<td></td>
</tr>
<tr>
<td>KL10</td>
<td>63,085</td>
<td>86.7 × 82.9 × 81.5</td>
<td>23</td>
<td>51.4 ± 0.2</td>
<td>11</td>
<td>4.9</td>
<td>10.3</td>
<td>11.8</td>
<td>2.3</td>
<td>24.8</td>
<td></td>
</tr>
<tr>
<td>KL12</td>
<td>63,123</td>
<td>84.0 × 85.5 × 81.8</td>
<td>11</td>
<td>51.3 ± 0.2</td>
<td>3</td>
<td>6.0</td>
<td>8.6</td>
<td>11.7</td>
<td>3.4</td>
<td>22.8</td>
<td></td>
</tr>
<tr>
<td>KL20</td>
<td>63,275</td>
<td>88.4 × 89.0 × 79.0</td>
<td>24</td>
<td>53.8 ± 0.3</td>
<td>1</td>
<td>11.1</td>
<td>14.5</td>
<td>12.7</td>
<td>6.6</td>
<td>31.7</td>
<td></td>
</tr>
</tbody>
</table>

The parameter \( n_{\text{atoms}} \) is the total number of atoms; \( a, b \) and \( c \), the size along the \( x, y \)- and \( z \)-axes, respectively, averaged over the 10-ns equilibrated run (the \( z \)-axis is the transmembrane axis); \( t_e \), the time to reach equilibrium, i.e. the time required before the system reached a stable area (\( a \times b \)) for 10 consecutive ns; \( A_{\text{lipid}} \), the mean area per lipid averaged over the 10-ns equilibrated run; \( n_{\text{HB}} \sim 280 \). Various properties of \( Gn \) and \( Kn \) systems related to hydrogen-bonding are given by following tabulated values: \( n_{\text{HB}} \), the average number of water molecules within 10 Å of the center of the bilayer; \( n_{\text{hbN}} \), and \( n_{\text{HB}}^\text{a} \) are the total number of hydrogen-bonds involving backbone nitrogens as donors, backbone oxygens as acceptors, and side chain nitrogens as donors, respectively; overlap, the number of hydrogen-bonding interactions that are counted twice, because \( n_{\text{HB}}^\text{a} \) and \( n_{\text{HB}}^\text{b} \) sometimes include the same combination of donor and acceptor; \( n_{\text{HB}}^\text{a} \), the total number of hydrogen-bonds, given by \( n_{\text{HB}}^\text{a} + n_{\text{HB}}^\text{b} + n_{\text{HB}} \)-overlap; —, no side-chain nitrogens.
conformations and orientations may be, one can conclude that the shorter sequences interact differently with bilayers than do the longer ones.

Molecular dynamics simulations (Fig. 5) show that the $\alpha$-helical conformity is stable for all of the peptides examined, except for the KL6 peptide, which partially unfolds. The unfolding is stabilized because the cost of deforming the bilayer was smaller than the cost of exposing the peptide bonds to the bilayer interior.

An analysis of the shape of the bilayer in the vicinity of the peptide revealed consistent hydrophobic mismatch that, as expected, increased in proportion to the difference between the length of the leucine helices and the unperturbed bilayer thickness (Fig. 5). The shape analysis allowed the positive free energy cost of the hydrophobic mismatch for each peptide to be computed relative to the GL20 and KL20 peptides, which had essentially no mismatch (Fig. 2C). The energetic cost arises from the energy required to distort the bilayer to minimize the exposure of the nonpolar leucines to the aqueous phase. The most significant observation may be that the energetic costs for the GL peptides rise roughly in parallel as $n$ decreases to 10 or 12; but for $n < 10$, the energetic cost of mismatch becomes significantly smaller for the KL peptides.

because the cost of deforming the bilayer was smaller than the cost of exposing the peptide bonds to the bilayer interior.
The consistent picture that emerges from these studies is that lipid bilayers with acyl chains of 16–18 carbons easily adapt to TM helices as short as 10–12 leucines without too severe an energy penalty. The energetic cost of inserting helices shorter than ~10 residues rises dramatically and prohibitively unless the leucines are flanked by a sufficient number of lysine residues, whose additional hydrogen bonding interactions with the phospholipid headgroups (snorkeling) reduce hydrophobic mismatch and consequently the energetic cost of short-helix insertion. What do these findings say about the mechanism of insertion of TM segments by the Sec61 translocon? First, no new insertion mechanism need be invoked to explain the insertion of short TM segments. The ability of the translocon to guide the insertion of short helices is completely consistent with the physical principles elucidated by our synthetic peptide studies and our molecular dynamics simulations. Second, despite the complicated inner workings of the ribosome–translocon complex (19, 20), the final “selection rules” for transmembrane insertion of nascent chains by the translocon are broadly consistent with physical partitioning of single or possibly multimers of TM helices between the translocon and membrane bilayer (21–23).

Methods

Expression in Vitro. Constructs in pGEM1 were transcribed and translated in the TNT SP6 Quick Coupled Transcription/Translation System. One microgram DNA template, 1 μL 15S-Met (5 μCi), and 1 μL microsomes (a gift from Dr. M. Sakaguchi, Hyogo University, Hyogo, Japan) were added at the start of the reaction, and samples were incubated for 90 min at 30 °C. Translation products were analyzed by SDS/PAGE and gels were quantitated on a Fuji FLA-3000 digital laser scanner. Translation products of the wild-type and the Sec61Δ mutants were incubated for 90 min at 30 °C. Translation products from the wild-type and Sec61Δ mutants were incubated for 90 min at 30 °C. Translation products were analyzed by SDS/PAGE and gels were quantitated on a Fuji FLA-3000 digital laser scanner.

Correction for the amount of doubly glycosylated protein was used as described by Hessa et al. (7). On average, the glycosylation levels vary by no more than ±2% between repeat experiments.

Pepptide Synthesis. The sequences of the peptides were 6, 8, 10, and 12 Leu flanked by the groups GGGP or KKKP. In addition, each peptide carried a Trp residue on the C terminal to facilitate measurements of concentration. The peptides were synthesized on a 433A Applied Biosystems automatic synthesizer with stepwise solid-phase procedures (24) by using fluorenylmethoxycarbonyl chemistry and trifluoroacetic cleavage.

Solution Circular Dichroism. CD measurements were performed by using an upgraded Jasco-720 spectropolarimeter (Japan Spectroscopic). Normally, 10 to 30 scans were recorded between 190 and 260 nm at ambient temperature (~25 °C), using a 1 mm optical path. All spectra were corrected for background scattering by subtracting a vesicle-only spectrum measured with an appropriate concentration of vesicles in buffer, without the peptide. See SI Text for more details.

Oriented Circular Dichroism. Oriented CD measurements were performed by using the Jasco-720 spectropolarimeter. Oriented CD spectra were measured by using oriented multibilayers deposited on a quartz slide according to the procedures of Huang and colleagues (25–27).

Molecular Dynamics Simulations. All-atom MD simulations of the GLn and KLn peptides with n = 6, 8, 10, 12, and 20 were carried out in POPC bilayers hydrated with 30 water molecules per lipid. Each system consisted of a single peptide in a TM configuration, 280 POPC lipids, and 8,400 water molecules. The simulations were run at a constant temperature of 300 K and a constant pressure of 1 bar until the dimensions of the simulation cell remained stationary (fluctuated around a stable average) for 10 ns. The details of the system setup, equilibration, and simulation protocols are provided in the SI Text.

ACKNOWLEDGMENTS. This work was supported by the National Institute of General Medical Sciences Grants GM46823, GM68002, GM74637 to S.J.T. and the National Science Foundation Grants CHE-0417158 and CHE-0750175 (to D.J.T.); as well as grants from the Swedish Cancer Foundation, the Swedish Foundation for Strategic Research, the Marianne and Marcus Wallenberg Foundation, and the Swedish Research Council (to G.v.H.), and grants from the Magnus Bergvall Foundation (to L.N.). We thank Le Fond Québécois de la Recherche sur la Nature et les Technologies for a graduate fellowship to S.J.

Supporting Information

Jaud et al. 10.1073/pnas.0900638106

SI Text

Molecular Dynamics (MD): Analysis of HydrophobicMismatch Energetics. To quantify the shape of the bilayer around every peptide, we have selected a model of lipid–protein interaction based on the Landau-de Gennes theory, in which the free energy of lipid–protein interactions is determined by minimizing the functional integral of the local free energy density (1). Doing so requires solving the lipid thickness as a function of the radial distance from the protein, \( r \). We must first assume that the peptide is fixed in the membrane and is only acting to provide a boundary condition for the lipid. Under relaxed boundary conditions \( d_l(0) = d_p \), and by using a free energy density model with terms for elastic deformation and local thickness fluctuation, one finds (2):

\[
 d_l(r) = d_l^0 - (d_l^0 - d_p) \exp(-r/\xi_p) \tag{S1}
\]

where \( d_p \) is the hydrophobic length of the peptide, \( d_l^0 \) is the unperturbed lipid thickness (far from the peptide), and \( \xi_p \) is the coherence length of the peptide perturbation.

We used this model to parameterize membrane perturbation around each simulated system, as follows. We first quantified the shape of the bilayer. Taking periodic boundary conditions into account, we selected choline groups within 50 Å of the center-of-mass (c-o-m) of the helix (orange dots, Fig. 5 Left). We then calculated the average positions of 2-Å-thick rings of cholines centered on the helix c-o-m (black dots, Fig. 5 Left). Finally, we calculated the average position of Leu sidechains every 1 Å along the z-axis (red circles, Fig. 5). Using these data, we subtracted the average lower bilayer surface from average upper surface to obtain the hydrophobic thickness as a function of the radial distance \( r \) from the peptide c-o-m. We ignored data for \( r < 5 \) Å, because the average radial position of the Leu sidechains is \( \sim 5 \) Å (Fig. 5 Left). Our analysis illustrates the extensive bilayer rearrangement occurring around the peptides to maximize favorable hydrophobic interactions. The bilayer collapses around peptides to form cavities that are typically deepest for the shortest peptides.

Information gained by using Eq. S1 can then be used to calculate the difference in free energy of deformation of the lipid-protein system:

\[
 \Delta G = k \left( \frac{2r_p}{\xi_l} + 1 \right) (d_l^0 - d_p)^2 \tag{S2}
\]

where \( k \) is a phenomenological constant related to the bilayer area compressibility modulus, \( r_p \) is the radius of the peptide, which is assumed to have a cylindrical shape, and \( \xi_l \) is the persistence length of the lipid-bilayer fluctuations. Hence, \( \Delta G \) is proportional to the square of the hydrophobic mismatch between the peptide and the unperturbed lipid, and calculating it requires the knowledge of \( k \) and \( \xi_l \), which are not readily available. After assuming that \( r_p = 5 \) Å, which is the average radial distance of helical side chains from the c-o-m of the helix, that \( k \) is constant in all systems, which are made of the same type of lipid, and that \( \xi_l \) can be used as a fair approximation for \( \xi_l \), we calculated the difference in free energy of all systems relative to G20, \( \Delta G_{\text{peptide}}/\Delta G_{\text{GL20}} \).

In the case of GL6, this ratio is remarkably linear, with a slope of \(-1.8 \) units per residue, an intercept of 35 units (\( R = 0.966 \)). After ignoring KL6, it is also linear for KL1, with a slope of \(-1.3 \) units per residue and an intercept of 27 units (\( R = 0.976 \)). Unsurprisingly, the longest peptides have the lowest free-energy of bending because they do not require their bilayers to bend as much as those around the shortest peptides.

Methods: In Vitro Expression. Enzymes and chemicals. All enzymes, plasmid pGEM1, DTT (DTT), and the TNT SP6 Quick Coupled Transcription/Translation System Kit were from Promega. [35S]-Met, ribonucleotides, and deoxyribonucleotides were from Amersham Pharmacia. BigDye Terminator v1.1 Cycle Sequencing Kit was from Applied Biosystems. Oligonucleotides were from Cybergene AB and MWG Biotech.

DNA manipulations. For expression of H-segment-containing Lep constructs from the pGEM1 plasmid, the 5’ end of the lep gene from Escherichia coli was modified by the introduction of an XbaI site and by changing the context of 5’ to the initiator ATG codon to a Kozak consensus sequence (3). Site-directed mutagenesis was performed by using the QuikChange Site-Directed Mutagenesis Kit from Stratagene to introduce acceptor sites for N-linked glycosylation in positions 96–98 (Asn-Ser-Thr) and 258–260 (Asn-Ala-Thr).

Oligonucleotides encoding the different H-segments were introduced between a SpeI-site in codons 226–227 and the KpnI site in codon 253 in the Lep gene (4). H-segments were constructed by using 1 or 2 double-stranded oligonucleotides with overlapping overlaps at the ends. Pairs of complementary oligonucleotides were first annealed at 85 °C for 10 min followed by slow cooling to 30 °C, after which the 2 annealed double-stranded oligos were mixed, incubated at 65 °C for 5 min, cooled slowly to room temperature, and ligated into the vector. All H-segment inserts were confirmed by sequencing of plasmid DNA at BM Labnet AB.

Peptide Synthesis. The sequences of the peptides were 6, 8, 10, and 12 Leu flanked by the groups GGPG or KKPK. In addition, each peptide carried a Trp residue on the C terminal to facilitate measurements of concentration. The peptides were synthesized on a 433A Applied Biosystems automatic synthesizer with stepwise solid-phase procedures (5) by using fluorenlymethoxycarbonyl (Fmoc) chemistry and trifluoroacetic cleavage. Syntheses were performed on a 0.1-mmol scale starting with Fmoc-Rink amide 4-methylbenzhydrylamine hydrochloride resin. Fmoc-protected amino acids were used for all coupling reactions. The crude peptides were purified by reverse-phase HPLC on a preparative C18 reverse-phase column with gradients of acetonitrile in 0.1% TFA. Peptide identities were confirmed by MALDI mass spectrometry (Chemistry Department, University of California, Irvine). Fmoc amino acids and resins for peptide synthesis were obtained from NovaBiochem (EMD Biosciences). All chemicals were of analytical reagent grade.

Circular Dichroism Spectroscopy. Materials. Palmitoyloleoyl phosphocholine (POPC) and palmitoyloleoyl phosphatidylglycerol (POPG) were purchased from Avanti Polar Lipids.

Analysis of solution CD data. Measured values of ellipticity (\( \theta \)) were converted into the ellipticity per amino acid residue [\( \theta ] \) by using

\[
 [\theta] = \frac{\Theta M W_R}{l c} \tag{S3}
\]

where \( l \) is the optical path-length of the cell, \( c \) the peptide concentration, and \( MW_R \) the average mass per amino acid residue of the peptide used. UV absorbance was measured with a Cary 3E spectrophotometer (Varian Analytical Instruments).
Molar concentrations were determined by using a molar extinction coefficient of ε_{280 nm} \approx 5600 M^{-1} cm^{-1} for tryptophan (a Trp was added to the C terminal of all peptides). Peptide concentrations were typically between 10 and 30 µM in phosphate buffer.

**Oriented CD measurements.** The peptide and the lipid were codissolved in methanol at different molar ratios (R = 1:12, 1:25, 1:50, and 1:100) and carefully added drop-wise onto a 1-cm-diameter circular area in the center of a 2.5-cm-diameter quartz plate. After solvent removal under a stream of nitrogen followed by final removal under vacuum, samples were hydrated with warm air at 100% relative humidity. The quartz plate was mounted (sample side inward) on the end of a tube sealed at the opposite end with a second quartz plate. Samples were hydrated to their final hydration through the vapor phase by placing a drop of saturated salt solution in the sample holder before sealing. The sample was equilibrated overnight. The sample was then placed in a Jasco J720 CD spectrometer such that the optical axis was normal to the 2 parallel quartz plates. Spectra were recorded at each of 8 rotations of 45° around the optical axis, coinciding with the beam and normal to the lipid multilayers, and then averaged. The background signal was determined with the same amount of lipid, without the peptide.

**All-Atom Molecular Dynamics Simulations.** Using the Maestro molecular modeling environment (Schrodinger LLC), we built 2 sets of leucine α-helices (L_{n}) of varying length n, capped on each end with either weakly polar (GGPG) or positively charged (KPKP) groups in extended conformations. We first constructed 8 GGPG-L_{6}-GGPG peptides, the so-called GLn set, with n = 6, 8, 10, 12, 14, 16, 18, and 20. We then mutated each glycine into lysine residues to generate 8 KPKP-L_{6}-KPKP peptides, the KLn set. Finally, we used psfgen (6) to acetylate the N termini and amidate the C termini of these 16 different systems. After applying strong harmonic constraints (100 kcal mol^{-1}) on the position of all heavy atoms, we used Nanoscale Molecular Dynamics (NAMD) software (6) to relax each peptide through application of all-atom forces. Moreover, because of sheer computing cost and time restrictions, we only retained the backbone from 100 to 20 to 5 to 0 kcal mol^{-1} over three 1-ns NPT (constant number of particles, pressure, and temperature) runs (with electrostatic interactions) at 300 K and 1 bar. We then replicated the G20 system and proceeded to build the G18 system by replacing G20 with G18. We progressively removed constraints on the backbone from 100 to 20 to 5 to 0 kcal mol^{-1} during 10,000 steps at 0 K, in absence of electrostatic interactions. We followed this with a 10-ps NVT run at 310 K, and then unfroze the lipids and the water molecules. We progressively removed constraints on the backbone from 100 to 20 to 5 to 0 kcal mol^{-1} during 3 one-ns NPT (constant number of particles, pressure, and temperature) runs (with electrostatic interactions) at 300 K and 1 bar. We then replicated the G20 system and proceeded to build the G18 system by replacing G20 with G18. We treated this new system in exactly the same way we relaxed the G20 system. After the peptide was unconstrained, we duplicated the G18 system to generate a starting point in which to insert G16, etc. Hence, we recursively built all systems down to G6. The KL20 system was built from the G20 system, to which we added 6 Cl^- counterions. Other KLn systems were generated recursively from K20, K18, etc.

Following peptide insertion into the bilayer, all MD simulations were performed with periodic boundary conditions by using a multiple time step integrator (7, 8) with an elementary time-step of 1 fs. Nonbonded and electrostatic interactions were calculated every 2 and 4 time-steps, respectively, with a cutoff of 11 Å. The SHAKE algorithm (9) was used to constrain the length of the bonds involving hydrogen atoms. The particle mesh Ewald summation (10) was used in the calculation of Coulomb interactions. Water was modeled by using the TIP3P representation (11). The temperature was kept constant by using Langevin dynamics, and a Nosé-Hoover Langevin piston (12, 13) was used for pressure control (NPT, 1 bar and 300K). The MD simulations were performed with the NAMD program (6) using the all-atom CHARMM22 (14, 15) potential and CHARMM27 (16) lipid force fields. Molecular graphics images were prepared by using the Visual Molecular Dynamics (VMD) program (17).

We removed excess water from our systems to speed up simulations by reducing the aqueous content to 30 per lipid, which corresponds to the minimum water content of fully hydrated phosphatidycholine bilayers (18). Hence, every system contained n_{lip} = 280 POPC molecules and 8,400 water molecules. Moreover, because of shear computing cost and time constraints, we only retained the n = 6, 8, 10, and 12 systems, to study the effect of short helices on interactions, and the n = 20 systems, to use as controls. We ran MD simulations on unconstrained systems until they were equilibrated, i.e., until their area per lipid was stable for 10 consecutive nanoseconds. The number of atoms, the size, the length of the equilibration run, and the area per lipid for each system are reported in Table 1.
Fig. S1. CD spectra of synthetic peptides in methanol. Peptide concentrations are generally 40 μM.
Fig. S2. CD spectra of synthetic peptides in 0.1 M phosphate buffer. Peptide concentrations were typically 25 μM. (A) Spectra of GGPG-(Leu),-GPGG peptides. (B) Spectra of KKPK-(Leu),-KKPK peptides.
Fig. S3. Oriented CD data for the GPG-Leu<sub>n</sub>-GPG (GL) and KKPK-Leu<sub>n</sub>-KPK (KL) peptides under various experimental conditions. The expected spectra for helices oriented normal and parallel to bilayers oriented normal to the CD spectrometer optical axis are shown in the Upper Right (GL6 in POPC).
Fig. S3. Continued.
Fig. S3. Continued.
Fig. S3. Continued.
Fig. S4. Overview of molecular dynamics simulations. GGPG-Ln-GPGG (GLn) and KKPK-Ln-KPKK (KLn) systems for \( n = 6, 8, 10, 12, \) and 20. Every system contained 280 POPC molecules and 30 water molecules per lipid. Systems were run at 300 K and 1 bar until their area per lipid was found to be stable for 10 consecutive nanoseconds. Information about each system is reported in Table 1. Color code: Green, glycine; blue, lysine; red, proline; and white, leucine. (Upper) G20 and K20 peptides in their respective POPC membrane before equilibration. Systems with shorter peptides were constructed from these bilayers. (Lower) All Gn and Kn peptides, averaged over the equilibrated runs. There are no major differences between the structure of the backbone of Gn and Kn systems. They all retained their helicity, except for K6.
Fig. S5. Some features of the GL6, GL20, and KL6 systems observed from a slice through the mean systems. (Upper and Left) A cut through the average GL20, GL6, and KL6 systems in van der Waals representation. The orange and light-blue lipid sections represent the polar headgroup and the apolar core, respectively. The acyl chains of GL6 and KL6 are strongly curled in an effort to accommodate the apolar helices. This is not the case of GL20 and KL20 (not shown here), around which the acyl chains are significantly straighter and more stretched. (Lower and Right) A cut through the KL6 backbone, surrounding water molecules, and the lipid in licorice representation, during the simulation. The KL6 peptide killed its helicity to maximize interactions between its capping groups and the bilayer's headgroup. Hence, intramolecular hydrogen bonds were replaced for interactions with water molecules, resulting in a trail of water going through the bilayer.
Fig. S6. Orientational order parameters for the GL\(n\) (a and b) and KL\(n\) (c and d) systems. Lipids are separated in 2 regions: Within 15 Å of the center of mass of the peptide's helix (region 1), lipids are less ordered than farther than 15 Å (region 2).
Fig. S7.  Time-averaged number of hydrogen bonds in GLn and KLn systems as a function of the residue. Hydrogen-bonding occurs when 2 electro-negative atoms, usually nitrogen or oxygen, compete for the same hydrogen atom. The hydrogen atom is bonded covalently to the donor, D, and is being shared with the acceptor, A. In ideal cases, D-H-A form a 180° angle, but values can deviate by as much as 30° depending on the systems. Moreover, D and A are typically within 3 Å of each other. We used a purely geometrical procedure that is blind to the identity of the atoms interacting with each other, as long as D and A are within 3 Å, and as long as the D-H-A angle is within 150–210°. Nitrogen and oxygen are the only atoms that can participate in hydrogen-bonding in our systems. Our analysis has revealed that there was no occurrence of interactions in which nitrogen was an acceptor. Therefore, the only possibilities involve the backbone nitrogen as a donor and, as an acceptor: (i) backbone oxygen, (ii) phosphate oxygen, (iii) carbonyl oxygen, and (iv) water oxygen. In this figure, we labeled those combinations “backbone N as a donor, acceptor is oxygen, phosphate, carbonyl, and water”. There are 4 more, nearly identical combinations in which the donor is the nitrogen from a lysine side chain instead. We labeled these “side chain N as a donor, acceptor is oxygen, phosphate, carbonyl, and water”. Moreover, we wanted to consider backbone oxygens as acceptors, and their interactions with peptide nitrogens and water acceptors. We labeled those combinations “side chain N as a donor, acceptor is nitrogen and water”. Finally, the amidated and acylated termini can also interact with other species, so we added “T” at the end of the residue names when these groups were involved, on the y axis. The analysis is shown for (a) GL6, (b) GL20, (c) KL6, and (d) KL20. Information regarding the average number of hydrogen-bonding for all systems is presented in Table 1.