Exploring the nature of the translocon-assisted protein insertion

Anna Rychkova and Arieh Warshel1

Department of Chemistry, University of Southern California, Los Angeles, CA 90089

Contributed by Arieh Warshel, November 23, 2012 (sent for review October 16, 2012)

The elucidation of the molecular nature of the translocon-assisted protein insertion is a challenging problem due to the complexity of this process. Furthermore, the limited availability of crucial structural information makes it hard to interpret the hints about the insertion mechanism provided by biochemical studies. At present, it is not practical to explore the insertion process by brute force simulation approaches due to the extremely lengthy process and very complex landscape. Thus, this work uses our previously developed coarse-grained model and explores the energetics of the membrane insertion and translocation paths. The trend in the calculated free-energy profiles is verified by evaluating the correlation between the calculated and observed effect of mutations as well as the effect of inverting the signal peptide that reflects the “positive-inside” rule. Furthermore, the effect of the tentative opening induced by the ribosome is found to reduce the kinetic barrier. Significantly, the trend of the forward and backward energy barriers provides a powerful way to analyze key energetics information. Thus, it is concluded that the insertion process is most likely a nonequilibrium process. Moreover, we provided a general formulation for the analysis of the elusive apparent membrane insertion energy, \( \Delta G_{\text{app}} \), and conclude that this important parameter is unlikely to correspond to the free-energy difference between the translocon and membrane. Our formulation seems to resolve the controversy about \( \Delta G_{\text{app}} \) for Arg.

coarse-grain modeling | hydrophobicity scale | topology

1To whom correspondence should be addressed. E-mail: warshel@usc.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1220361110

PNAS | January 8, 2013 | vol. 110 | no. 2 | 495–500

Author contributions: A.W. designed research; A.R. and A.W. performed research; A.R. and A.W. analyzed data; and A.R. and A.W. wrote the paper.

The authors declare no conflict of interest.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1220361110/-/DCSupplemental.
biological conditions. These workers generated a construct with two TM helices of this protein (TM1 and TM2) and an additional helix (the H helix), flanked by two acceptor sites for N-linked glycosylation. The degree of membrane integration of the H helix was then determined by the number of glycosylated sites and the apparent equilibrium constant, $K_{app} = f_1/f_2$ (where $f_1$ and $f_2$ are the fractions of singly and doubly glycosylated proteins, respectively). This value was then converted to the relevant apparent free energy, $\Delta G_{app} = -RT \ln K_{app}$ which was then decomposed to the contribution from each of the 20 naturally occurring amino acids placed in the middle of the H helix.

The basic question in exploring the molecular meaning of $\Delta G_{app}$ and the TR-mediated insertion process is the nature of the relevant energetics and insertion paths. The original implication has been that the translocation effect is relatively small and that the experimental findings are related to the energetics of the specific amino acid in the center of the membrane. This issue is important in view of attempts to point out the possibility that the experiments have not established that the controversial low $\Delta G_{app}$ for a positively charged Arg corresponds to Arg in the center of the membrane, since we have an obvious possibility that the Arg side chain is tilted toward the membrane surface and that this explains the low $\Delta G_{app}$ (9, 10, 17). We point out in this respect that one can use a more trivial suggestion, just saying that the center of the helix can move and place the charge closer to the membrane surface. However, it seems to be clear from the heated arguments in the field that the insertion is related to the Arg energy when it is in the center of the membrane. In fact, once it is asserted that $\Delta G_{app}$ corresponds to the water membrane partition, then this must be related to the energy of being at the center of the membrane. This issue is further addressed in the SI Text.

We would also like to clarify that our CG finding about the importance of helix–helix interaction in reducing the $\Delta G_{app}$ for Arg insertion should not be overlooked by arguing that the experiment by Meindl-Beinker et al. (18) shows very little effect of mutating the Arg side chain to another residue. Unfortunately, as stated above, we are not aware of a clear justiﬁcation for this assumption. Thus, the relatively clearly deﬁned issue of the energetic of charges in the center of the membrane forces one to move to the far more challenging issue of the energetics of the whole insertion process.

The pioneering studies of von Heijne (e.g., ref. 19) have established the positive-inside rule, which identiﬁed the retention of the positively charged N terminus on the intracellular part of the membrane. However, the exact structure-based origin of this effect remains an open question.

In both prokaryotic and eukaryotic cells, proteins are allowed entry into the secretory pathway only if they are endowed with a specific targeting signal—a SP (20), Goder and Spiess (1) performed systematic and elegant studies of the in/out ratio in the insertion of SPs with a helical segment and a tail of different length and provided intriguing information about the insertion process. In particular (Fig. S1), it has been found that increasing the positive charge in the N-terminal increases the percentage of the C-translocated peptides ($N_{trans}$), whereas increasing the length of the helix decreases this percentage. Furthermore, the fraction of C-translocated peptides increases upon moving from short tails to longer tails, until reaching a fixed fraction (SI Text). Reproducing and rationalizing these experimental trends is clearly a worthwhile challenge.

Results

Our analysis starts with a focus on the SP sequences studies by Goder and Spiess (1). These SPs generated mixed topologies in experiment where the partitioning depended on the flanking charges and the signal hydrophobicity. In our modeling, we truncated the C-terminal to four residues and used the sequences shown in Fig. S2 (we also performed calculations that considered the tail).

The TR was modeled by using as a starting point the SecY structure of the SecA-bound form (5) (Protein Data Bank ID code 3DIN). The SP structure was created in PyMOL (21) from an arbitrary helix of 30 residues by mutating the amino acids to the required sequence. The generated SP was inserted into the TR by first placing it in the lateral gate [according to the recent electron cryomicroscopy data (22, 23)] and then relaxing the system, applying torsion and hydrogen bond constraints on the helical part of the SP (22 oligo-leucine) to prevent it from unfolding. The resulting model is depicted in Fig. 1.

To explore the origin of the observed (1) dependence of the in/out ratio on the flanking charges of the SP, we calculated the barriers for the SP insertion into the TR for three SPs, which are named here RR, PR, and PH (Fig. S2). We started with a targeted molecular dynamics (TMD) treatment in which we pulled the SP from inside the TR to the cytoplasm (to reflect the reversed TR insertion process), to the membrane and to the exoplasm (to reflect the translocation process). All of the TMD runs constrained the helical region of the SP to prevent it from unwinding. At any rate, after obtaining the SP insertion path we evaluated the CG energies along this path (see justification in SI Text) and also used a specialized treatment to obtain stable results for mutational studies (see SI Text). In imposing the helical constraint, we considered the fact that exploring the insertion of a nonhelical system will drastically reduce the reliability of the calculations. Thus, we focused on the helical system whose energetics provides a qualitative limit to the energetics of other insertion processes (Fig. S3). We also note that a recent work (24) concluded, by length estimates, that the inserted system in related cases is at least partially helical. It was also found that the energetics of a helical construct and a construct in which the helix is perturbed by a central proline residue is quite similar, indicating that nonhelical and helical insertion should have similar barriers.

The reaction coordinate (X) for the TR insertion and translocation was taken as the rms of the first and last amino acid of the SP, respectively, and as the rms of the first and last residues of the helical part of the SP for the membrane insertion. Zero reaction coordinate corresponds to the SP positioned inside the lateral gate of the TR, $X_{memtrans} < 0$ — SP inside the cytoplasm,

![Fig. 1. The CG model of the RR SP inside the TR (at X = 0). The SP is shown in purple, the TR in orange with the TM 2b and 7 forming lateral gate in green, and the membrane in gray. Part of the membrane is removed for better protein visibility.](image-url)
Although the insertion process is complicated and is driven by a vectorial process (e.g., the ribosome), we started examining the energetics of the insertion without the ribosome and then explored some aspect of the ribosome-assisted insertion and the effect of the tail. Furthermore, we focused on the difference between the effects of mutations on each of SP helical configurations (rather than on the difference between the energies of the relaxed configuration of each mutant).

**Energetics of the Translocon-Assisted Membrane Insertion and Translocation.** The calculated profile for translocon-assisted membrane insertion and translocation of RR SP is shown in Fig. 2, where the data points are plotted relative to the \( \Delta G_{\text{in}}(A_{\text{in}}) \) of the corresponding SP in the \( N_{\text{in}} \) orientation. It should be noted that \( \Delta G_{\text{in}}(A_{\text{in}}) \) appears to be shifted up for RR SP, although the free energy of the system where the SP is fully positioned in water should be the same independently of the SP orientation. The shift is due to the electrostatic interaction of the SP with the overall field from the TR. This effect decreases when the charge of the SP is reduced. Overall, we find that two positive charges on the N terminus side of the RR SP contribute to the highest barrier for the insertion into the TR. We can also see the reduction in the free energy of the SP positioned inside the TR (state B) when the charges are eliminated (Supporting Information).

To explore the consistency of the different models, we compared the difference in the barrier heights for the TR insertion [\( \Delta \Delta G^I(A_{\text{out}} \rightarrow A_{\text{in}}) \)] and the difference in the free energies inside the TR [\( \Delta \Delta G(B_{\text{out}} \rightarrow B_{\text{in}}) \)] of the SPs in the \( N_{\text{in}} \) and \( N_{\text{out}} \) orientations with the corresponding experimental values [\( \Delta \Delta G_{\text{exp}}(N_{\text{in}} \rightarrow N_{\text{out}}) \)] (Fig. S4 and Supporting Information). Both sets have a reasonable correlation with the experiment. However, the calculated effects are significantly larger than the corresponding observed effect. This trend is due in part to the missing compensating contributions associated with the use of a fixed helix and an identical path for \( N_{\text{in}} \) and \( N_{\text{out}} \) [note that the apparent dielectric effect reflects factors that are not explicitly included in the simulations (25)], as well as to the missing effect of the tail, which will be considered below.

**Effect of Mutations on the Energetics of the TR-Assisted SP Insertion.** Next, we calculated the effect of TR mutations on the energetics of the SP insertion, considering the experimental data from refs. 6 and 26. The SP models were derived from 60[H1]+1) and 40 [Leu16(+5) peptides, while truncating the N- and C-terminals of the SPs (Fig. S2). The peptide models were built in a way similar to that described above. The experimental results for the TR mutations were taken from studies of the yeast *Saccharomyces cerevisiae* Sec61 TR (6, 26). To check the effect of similar mutations in the Thermotoga maritima SecY system, we also used the sequence alignment data from ref. 27, considering data from conserved residues in both organisms. The calculated

\[ X_{\text{in}} > 0 - \text{SP inside membrane, and } X_{\text{transl}} > 0 - \text{SP in the extracellular part.} \]
results are summarized in the Supporting Information and the correlations of $\Delta \Delta g(A_{\text{out}} \rightarrow A_{\text{in}})$ and $\Delta \Delta g(B_{\text{out}} \rightarrow B_{\text{in}})$ with $\Delta \Delta G_{\exp}(N_{\text{out}} \rightarrow N_{\text{in}})$ depicted in Fig. 3 B and C. As can be seen from Fig. 3B, $\Delta \Delta g(A_{\text{out}} \rightarrow A_{\text{in}})$ of the 60[H1](+1) SP has better correlation with the experiments than $\Delta \Delta G(B_{\text{out}} \rightarrow B_{\text{in}})$. Moreover, for the 40[Leu16](+5)CPY, $\Delta \Delta G(B_{\text{out}} \rightarrow B_{\text{in}})$ is uncorrelated with $\Delta \Delta G_{\exp}(N_{\text{out}} \rightarrow N_{\text{in}})$ (see discussion below). Finally, we found that the effect of the TR mutations that change the in/out ratio is independent from the tail effect.

**Effect of the Ribosome on the Energetics of the TR-Assisted SP Insertion.** In the absence of direct estimate of the time for the insertion process, we took the estimate (2) of a translation rate of $\sim 5$ aa per second. We note, however, that this estimate does not tell us what exactly the actual barrier in the TR is and that it is taken in the absence of alternative information. At any rate, this time constant can be converted to an activation barrier of 20 kcal/mol based on transition state theory (see SI Text for clarifications). This barrier is most probably much lower than the barrier without the external help of activating systems such as the ribosome. The ribosome can act just by its direct electrostatic effect or by changing the structure of the TR and thus reducing the barrier. To explore these options, we examined first how the ribosome charges may affect the barrier for the SP insertion. This was done by adding 16 negative charges 5 Å away from the membrane surface in a position similar to rRNA H59 helix of the ribosome using the data from the cryo-EM structure of the ribosome–SecY complex (22). The addition of the negative charges reduces the barriers for RR TR insertion in $N_{\text{in}}$ and $N_{\text{out}}$ orientations by $\sim 2$ kcal/mol. Next, we explored the possible indirect effect of the ribosome-induced TR structural changes on the SP insertion. For this purpose, we used the cryo-EM structure of the SecY. Now (Fig. 4) we obtained very significant reduction in the barrier heights for the RR SP insertion into the ribosome-bound TR: $\Delta \Delta \Delta g(A_{\text{in}}) = 6.3$ kcal/mol and $\Delta \Delta \Delta g(A_{\text{out}}) = 9.3$ kcal/mol. Although the high-resolution structure of the ribosome–TR complex is not available and our modeling does not involve the simulation of a protein chain growing inside the ribosome, we believe that we captured the trend in the real effect. We also note that despite the fact that the X-ray structure with SecA does not show the large structural change used here it may also have an activated open structure (5). However, it is very likely that the movement from the SecA-bound to the ribosome-bound TR structure presents a significant overestimate of the actual opening. Thus, we considered in Fig. 4 (and in the subsequent discussion) a profile that takes 55% of the SecA-bound and 45% of the ribosome-bound TR insertion profiles to reproduce the rough estimate of $\Delta \Delta \Delta g(A_{\text{in}})$ (20 kcal/mol) of the barrier of the insertion process. The idea is that in this way we capture the some of the tentative effect of the ribosome while still having a reasonable barrier. We are also well aware that the role of SecA includes an active ATP-driven process (see figure S2 of ref. 11). However, in the present case, we only consider this effect implicitly.

**Conceptual Analysis**

Although the present CG profile does not provide a quantitative tool of estimating the activation barriers for the insertion process, there are elements of the calculations that are quite reliable and robust, at least in establishing the relative trends. The current work allows us to use an energy-based formulation in relating the insertion problem to the available experiments. This can be done by considering the energy diagrams of Fig. 5 and the more complete description in Figs. S4 and S5, in which we provide estimates of the barriers for the different feasible paths. Here, we used the scaled results mentioned above with $\Delta \Delta \Delta g(A_{\text{out}})$ set at 20 kcal/mol. Furthermore, the energy values in the membrane were treated as discussed in SI Text, leading to the second set of energies used in Fig. S5.

In considering Figs. S4 and S5, we can reach several tentative conclusions. The first is that the insertion is most probably irreversible because the barrier for going back from the membrane to the initial state is too high. That is, the forward barrier in the ribosome-assisted process can be estimated to be $\sim 20$ kcal/mol, whereas the stabilization by the membrane is most probably $>10$ kcal/mol. Thus, the back reaction is at the range of 30 kcal/mol, which is not accessible at biological times. Note that we define here a limiting barrier $\Delta \Delta \Delta g$, which correspond to the time of the experiment or the time when competing processes stop the insertion. Our conclusion about the irreversibility is supported by the analysis of the mutation experiments where we found much better correlation between the calculated and observed results for the forward activation barrier that for the equilibrium free energy (e.g., Fig. 3C).

Now, despite the above conclusion, we still have to ask whether the equilibration between the in and out configurations may involve state B as an intermediate. To explore this issue, we note that the barrier for moving to the membrane $B_{\text{in}} \rightarrow C_{\text{in}}$ or $B_{\text{out}} \rightarrow C_{\text{out}}$ is relatively small (in agreement with other calculations (8)). Thus, the retention model (Fig. 5) leads to the conclusion that state $C_{\text{out}}$ can move to state D. In fact, the model also allows for movement from D to $C_{\text{in}}$ through $B_{\text{in}}$. However, this is inconsistent with the fact that the in/out equilibrium is not determined by the final states $C_{\text{out}}$ and $C_{\text{in}}$ because these states are likely to have very similar energies (unless somehow the SP inserted to the membrane stays near the
TR or the ribosome). It is also likely that we underestimate the D to B barrier (this possibility is discussed in SI Text and also indicated in the figure), but it would not change our conclusions. Considering the alternative inversion model (Fig. S4B), we can also see the same problem (equilibrium between the two inserted configurations). Interestingly, the second inversion model (Fig. S4C) seems to be inconsistent with the observed in/out partition, because the barrier for A to B (or A') is identical for the in and out pathways and this should lead to 50% ratio, which is not observed experimentally in most cases. However, it may be possible that the second barrier becomes higher than the first barrier in the case of a short tail. In such a case, we will have a very small in/out ratio for a short tail and a ratio of 50% (which is the result of having the same rate limiting barrier) only for a longer tail. However, because we did not observe exactly 50%, this would mean that such a scenario is only possible if the path of Fig. S4C is not the only path. At any rate, the crucial point is that all models are consistent with a kinetic control by $\Delta g_j$ (A in/out), and none seems to reflect the equilibrium between the TR and membrane (see also concluding discussion).

The analysis of Fig. S4 is not directly related to the experiment that determined $\Delta G_{app}$ (which used a different system). Thus, we consider in Figs. 6 and 7 the process that corresponds to the measurements of $\Delta G_{app}$. In generating this figure, we took into account the fact that both state H and G are likely to have similar values to within a few kilocalories per mole, because otherwise (with the exception discussed below) the population in water would not be observed experimentally. Thus, we considered the limit of the second set of the membrane water energies (see above and SI Text) but also kept in mind the actual CG results. Using these two limits, we can reach the following conclusions (Fig. 7): (i) If $\Delta G_{H\rightarrow G}$ is small and if the back barriers from H to F ($\Delta g_{H\rightarrow F}$) or G to F ($\Delta g_{G\rightarrow F}$) are lower than the limiting barrier, $\Delta g_l$, we come back to the original idea that $\Delta G_{app}$ reflects equilibrium between the membrane and the water systems (Fig. 7A). In this case, the equilibrium is independent of the energy of state F so it is not an equilibrium between the TR and the membrane. (ii) If $\Delta G_{H\rightarrow G}$ is large and $\Delta g_{H\rightarrow F}$ and $\Delta g_{G\rightarrow F}$ are higher than $\Delta g_l$, then we have a kinetic control, where we might have a linear free-energy relationship (LFER), where the product distribution is correlated with the activation barriers (SI Text and Fig. 6b), so that the forward rates are correlated with $\Delta G_{app}$ (Fig. 7B). (iii) In the less likely case that the $\Delta G$'s of H and G are very different [$\Delta G_{H\rightarrow G}$] > 4 kcal/mol], the forward barriers from F must determine the populations of G and F (otherwise one of them will not be observed) and in this case $\Delta G_{app}$ must be determined by LFER. The seemingly alternative option that $\Delta G_{H\rightarrow G}$ > 4 kcal/mol and that the barriers $\Delta g_{H\rightarrow F}$ and $\Delta g_{G\rightarrow F}$ are lower than $\Delta g_l$ would lead to the finding of all of the population in G (which is inconsistent with the experimental finding) (Fig. 7C). (iv) Finally, in the case when the barriers are sufficiently small, we can just focus on the equilibrium problem with $k_1 = W/TR$ and $k_2 = M/TR$ (where M, W, and TR correspond to G, H, and F, respectively), assuming that the experiment cannot determine whether the H helix is in the membrane or in the TR region, so that $K_{app} = (TR + M)/W$. Now we can show (SI Text) that $K_{app} = (1 + k_1)/k_2$, and thus when $k_1$ and $k_2$ are much larger than 1, we have $K_{app} = k_2/k_1 = M/W$, and when $k_2 << 1$, we have $K_{app} = 1/k_1 = TR/W$. Overall, none of the considered options is consistent with the currently popular assumption that $\Delta G_{app}$ is determined by the equilibrium between the TR and membrane.

**Concluding Remarks**

The present study focused on the qualitative CG exploration of the insertion free-energy landscape using the hints provided by biochemical studies. The relative heights and positions of the calculated CG barriers were found to be consistent with key mutational information and with the positive-inside rule. Furthermore, the tentative effect of the TR opening induced by the ribosome is found to reduce the kinetic barrier. Equally important is the fact that our systematic analysis indicated that the mutation studies of the insertion process are much better correlated with $\Delta g_F$ than $\Delta g_H$, indicating that we have a kinetic control.

Our finding can be explained in rather clear qualitative terms, starting from our view that the knowledge of the energetics of $\Delta g_F$ (water/membrane) can represent water/membrane equilibrium problem with $K_1 = C/V$ and $K_2 = C/W$. Overall, none of the considered options is consistent with the currently popular assumption that $\Delta G_{app}$ is determined by the equilibrium between the TR and membrane.
the system should provide the clearest way of describing the kinetics and the partition results. For example, it must be obvious that the positive-inside rule is related to the interaction of the charge of the SP with some regions of the overall system, but elucidating the relevant energy contributions is crucial for a concrete understanding. Here, the seemingly obvious suggestion would be the interaction of positive charges with the negative ribosome charges that stabilizes the barrier for the $N_{in}$ path. However, this cannot explain why mutations of the TR change the in/out ratio. In this case, our calculations established that the TR electrostatic potential stabilizes a positive SP charge near the top of the free-energy profile [the TS at $A_{(out)}$ of Fig. 5]. Thus, the most likely possibility is that the electric potential of the TR, at $X$ of approximately $-40$ Å, is responsible to the positive-inside rule. Now, if this is true, then the insertion is controlled by the height of the barrier, which is a nonequilibrium kinetic control (SI Text). Furthermore, our study seems to indicate that the effect of the mutations that change the in/out distribution is independent on the ribosome effect.

This work reproduced the opposing trends in the effect of the hydrophobicity and polarity on the in/out ratio. That is, the increase in positive charge increases the barrier for insertion of the positive head and thus reduces the $N_{out}$ fraction. However, increasing the hydrophobicity of the SP helix reduces the barriers for both the $N_{out}$ and $N_{in}$, but does it to a lesser extent in the case of $N_{in}$, where the tail must also pass near the inserted helix. This opposing trend is indicated in Fig. S1 and also discussed in SI Text.

Finally, we may also speculate on the possible reason for the increase of the in/out ratio for long tails (2) (without performing actual simulations). That is, with a long tail we probably have an increase in hydrophobicity in the $N_{in}$ case, and this is likely to lead to a decrease of the $N_{out}$ barrier and an increased in/out ratio.

The most important advance in the present work is not so much in providing qualitative free-energy profiles but in forcing us to look at the alternative kinetic options in a well-defined energy-based logical way and to be able to incorporate experimental and conceptual constraints in the overall analysis. In particular, the trend of the forward and backward energy barriers provide a powerful way of analyzing key energetics information such as the apparent membrane insertion energy $\Delta G_{app}$. It is concluded that $\Delta G_{app}$ is unlikely to correspond to the difference between the free energies of the protein inside the translocon and the membrane, but in most limiting cases to the equilibrium between the membrane and water or the equilibrium between the TR and water. The use of our formulation seems to resolve the controversy about $\Delta G_{app}$ of Arg (SI Text).

Interestingly, our calculated profile seems to provide a rationale to the results found in the recent exciting experiment of ref. 24. Overall, we view the present study as a demonstration of the need of a clear mechanistic formulation in the study of the translocon-mediated insertion and of the ability of CG modeling to augment available experimental information and to provide further constraints on the kinetic analysis.

Methods

The present work uses a CG model that describes the main chains by an explicit model and represents the side chains by a simplified united atom model. The model has a unique treatment of the electrostatic energy that increases its reliability. The details of the model are given elsewhere (28) and in SI Text.

ACKNOWLEDGMENTS. We thank Martin Spiess and Toby Allen for stimulating discussions and the University of Southern California’s High Performance Computing and Communication Center for computer time. This work was supported by National Institutes of Health Grant GM40283.

Supporting Information

Rychkova and Warshel 10.1073/pnas.1220361110

SI Text

SI Additional Background

On ΔG_{app}. This paper put significant emphasis on the origin of the ΔG_{app}. Here, it is important to avoid some assertions that were discussed in the main text, and because of space limitation we add here some additional discussion. For example, we would like to point out that our coarse-grained (CG) finding about the importance of helix–helix interaction in reducing the ΔG_{app} for Arg cannot be overlooked by arguing that the experiments by Meindl-Beinker et al. (1) show very little effect of mutating the helices. In fact, the experiments were done with Asp and Asn but never with Arg, which would have led to a more conclusive finding. That is, in the case of an Asp residue, the desolvation penalty could easily lead to the protonation of an Asp, because (pK_w – pH) is relatively small. In fact, in our previous study (2), we found Glu residue being protonated inside the membrane at pH 7. Thus, it seems to us that the experiments of Meindl-Beinker et al. (1) are not conclusive enough in this respect and cannot be used to rule out the importance of helix–helix interaction for insertion of charged residues into the membrane.

It is also important to further clarify the controversy about the ΔG_{app} proposal as this adds important background for the motivation of this work. The original implications that ΔG_{app} for Arg corresponds to the equilibrium between a charge in the center of the membrane and in water has been rather obvious, because otherwise it does not make sense to talk on the thermodynamic scale of insertion of amino acids from water to membrane. Of course, one can offer many suggestions of how the Arg in the actual experiment cannot be at the center of the membrane (e.g., refs. 3 and 4) or to argue that the system cannot hold the charges in the center of the membrane. However, as pointed out in the main text, the question has been logically related to the rationalization of very small apparent energy of a charged Arg at the center of the membrane.

As to the more recent proposal in which it is implied that we have equilibrium between the membrane and the translocon (TR), we are not aware of any logical kinetic scheme that can justify this assertion (see main text).

On the "Positive-Inside" Rule and on Topology Constraints. In considering the positive-inside rule, we also note that, for the endoplasmic reticulum (ER) signals, the charge difference between the two flanking regions of the signal core, rather than the positive charge per se, correlates with the signal sequence orientation (5). Because there is no general electrical potential across the ER membrane, the positive inside rule is likely to be due to the interactions of charges in the signal with those at or near the translocon (6).

It is interesting to mention that the positive-inside rule applies even to organisms with reversed membrane potential (i.e., inside-positive instead of negative) (7). In these cases, the retention of positive charges in the cytoplasm cannot be determined by membrane potential. Cases like these suggest the importance of the insertion process by itself in determining the final membrane protein topology.

Goder et al. (6) argued that the TR charges are not solely responsible for the positive-inside rule, and that other subunits of the translocon complex may as well contribute to the charge rule. Due to the lack of the electrical potential across the ER membrane in eukaryotes, lipids may still influence the positive-inside rule by asymmetrically accessing the lateral gate of the TR, thus creating the charge imbalance between the two sides of the membrane. Some experiments suggest even more tricky tools, like changing lipid composition and Ca^{2+} concentration, that the cell might use in establishing the charge rule (8).

The proteins synthesized by the ribosome are allowed to enter into the secretory pathway only if they are endowed with a specific targeting signal—a signal peptide (SP) (9). The SP is in most cases a transient extension to the amino terminus of the protein and is removed by one of a small class of enzymes known as signal peptidases once its targeting function has been carried out. The studies of the SP insertion performed by Goder and Spiess (10) led to the interesting suggestion of the retention and inversion models. In the first model, the positive side of the SP is retained in the cytoplasm by interaction with the negative charges at or near the TR; thus, the SP inserts into the TR with predefined orientation and does not change it during the protein synthesis. In the second model, the SP enters the TR with either N_{in} or N_{out} orientation and then inverts its orientation to position positive part of the SP facing the cytoplasm. There are several experimental studies that support both retention and inversion models (10–13). The recent multiple CG trajectories analysis of Zhang and Miller (14) suggested that the SP may use both types of mechanism.

More evidence appears in the literature suggesting that the first signal peptide is not the sole determinant of the whole polypeptide membrane protein topology. As was shown in ref. 15, the topology of the protein containing five transmembrane helices could be controlled by a single positively charged residue placed at the very end of the protein. This result suggested that the protein remains “topologically uncommitted” until the last residue is synthesized. Because the size of the translocon pore seems to only allow the accommodation of one helix, membrane proteins should use other mechanisms to flip their orientation in the lipid bilayer. Analysis of membrane protein structure database also suggested that transmembrane helices must reposition during folding and oligomerization (16).

Multiple studies conducted by Dowhan’s group (e.g., refs. 17 and 18) indicated that lipid composition is important for the membrane protein orientation. It seems that the interaction of the nascent polypeptide chain with the translocon only determines the initial orientation of the TM domain. The final protein orientation depends on short-range interactions between the protein and the lipid environment and long-range interactions within the protein during final folding events. This means that the membrane protein organization can be changed post-assembly depending on changes in the lipid environment. This leads to an interesting possibility of lipid environment, which changes along the protein secretory pathway, to activate initially latent membrane proteins.

Some TR mutations are known to affect the topology of membrane proteins (19). For example, Spiess and coworkers identified a class of such mutants and found that all of them have a pfr phenotype. This type of mutants allows proteins with defective or absent signal sequence to be translocated. Based on the translocon structure information, Smith et al. (20) proposed that pfr mutations either destabilize the closed state of the channel or stabilize the open form, thus allowing the channel to open without triggering events of the signal sequence binding.

On the Effective Barrier. In this work, we assess the barrier for the TR insertion starting from the estimate of ref. 11 that the SP reaches its final orientation in ~50 s during the ribosome-driven insertion process. Unfortunately, this estimate is based on the ribosome
translation rate of ~5 aa/s, thus implying that the TR barrier is designed so that the rate of transfer through the TR will be similar to the rate of exit from the ribosome. Although this is clearly a very tentative assumption, it represents all of the currently available information, and thus we consider here an estimate of a transfer of ~3 Å in (1/5) s. Now assuming that the TR barrier has an effective width of around 15 Å, we can obtain a rough estimate of the effective activation barrier by using transition state theory (TST) with a transmission factor correction (e.g., ref. 21):

\[ k \approx \kappa \cdot 6 \cdot 10^{-12} \exp(-\Delta g^f / RT), \]  

[S1]

where \( \Delta g^f \) is the activation barrier in kilocalories per mole and RT is 0.6 kcal/mol at room temperature. The transmission factor, \( \kappa \), is given by the number of times the target ecty, which arrives to the transition state (TS) and ends at the product state, moves back and forth on the TS. The treatment of Eq. SI can become less valid when we have a low activation barrier and when we have a diffusive process. In the case of the insertion process, we believe that the activation barrier is sufficiently high to be treated by TST with some transmission factor coefficient. We also note that the diffusion constant used by ref. 14 (which is related to the transmission factor) presents a very significant overestimate. At any rate, with a transmission factor of 0.5, we estimate the activation barrier to be in the range of 20 kcal/mol. If the transmission factor is around 0.01, we have to reduce the activation barrier by ~2.5 kcal/mol, but it is still in the range used in the present work. Obviously, unless we have a realistic structural model with the ribosome, we cannot get a sufficiently accurate potential of mean force (PMF) to push for more well-defined activation barrier. However, at some stage it will be worthwhile to use our renormalization approach (22) and to obtain a realistic effective friction and effective activation barrier.

SI CG Method, Energy Contributions, and Calibration

CG models appeared to be very effective in elucidating protein functions (e.g., ref. 22) and the CG model has already been used by us in a study (2) of the assumption that \( \Delta G_{app} \) reflects an equilibrium process. A recent work (14) used a drastically simpler CG model (turning each three residues into one effective residue in a 2-D model). This CG study captured effects that are probably determined by very coarse features. However, this model has not explored the relevant energetics, and we believe that understanding the nature of the free-energy landscape rather than dynamical features is crucial for a better description of the control of the insertion process and for the elucidation of the nature of \( \Delta G_{app} \). In fact, except in the very unlikely case that the insertion is a completely diffusive event, the corresponding kinetics is controlled by the free-energy barriers and minima whose knowledge is a key to a clear description. Thus, it seems to us that our CG that is focused on a realistic description of the electrostatic energetics provides the optimal current strategy and such a model was used here.

Our CG model, whose details were given elsewhere (23), has been continuously refined (24, 25), considering a benchmark of absolute protein stability (an extension of the set considered in ref. 16 where we used the PDL/S model) and also other relevant features such as pK_a values and the energies of inserting charges into membrane. Recent examination of the effect of the main chain and hydrogen bond terms on the agreement between calculated and observed absolute stability of proteins led to a modification of the terms given in ref. 23. The current terms and parameters are given in the MOLARIS manual (26).

Because the energy of insertion to the membrane plays an important role in our analysis, we explored the performance of the CG model in evaluating the water to membrane insertion energy and report the corresponding results in Table S1. Unfortunately, there is only one direct experimental observation and a couple of direct simulation results (Table S1), and the calculated results significantly overestimate the values obtained from \( \Delta G_{app} \). The same well-known problem occurs with other CG or physically based models (27, 28). At present, it is unclear what the origin of this problem is and it might reflect the following three factors. (i) The value of \( \Delta G_{app} \) does not correspond to water–membrane equilibrium (this issue is analyzed in great length in the main text and in the Supporting Information). (ii) The overestimate of water–membrane free-energy difference can be in part due to the missing entropic effects (28) and possibly due to having the helix in the water phase stabilized by the membrane surface (27). (iii) In the case of the TR-mediated experiments, it is possible that the bound glycan in the glycosylation site stabilizes the helix in the water phase by around 5 kcal/mol. The corresponding effects were taken into account in part as a feasible limit in our second estimate in Fig. S5. However, we must mention that our main conclusions do not change by this estimate.

SI Effect of the Tail

Moving from a short helix to a long one without the tail gives an equal decrease in the calculated hydrophobicity (about ~4 kcal/mol) to \( N_{in} \) and \( N_{out} \). Furthermore, the addition of the tail with its charges increases the energy of the TS for \( N_{in} \) (where the tail passes near the helix) by ~5 kcal/mol for both the short and long helix (note that the electrostatic effect of the tail is an overestimate due to the incomplete stabilization of the charges of the tail where a major sampling including exploring ion pair formation between the tail charges would be required for improved absolute values). However, the addition of the tail reduces the hydrophobicity effect in moving from a sort to a long helix by ~1 kcal/mol in the \( N_{in} \) case. Thus, the calculated \( \Delta \Delta g^f \) for \( N_{in} \) going to \( N_{out} \) and moving from a short to a long helix decreases by 1 kcal/mol, which is consistent with the experimental trend of the reduction in the in/out ratio.

We have not performed simulation studies in the long tail limit. However, it is reasonable to assume that a longer tail would increase the hydrophobicity effect on the \( N_{in} \) path and stabilize the top of the barrier more than with a short tail, for a fixed helix length.

Another interesting issue is the nature of the barrier for moving the tail to the endoplasm in the \( N_{in} \) case. The corresponding barrier is unlikely to be rate limiting, as the effect of the helix charges changes the in/out ratio and this cannot happened at the stage when the tail goes to the outside, unless the helix is still in the TR (or between the TR and the membrane) in the stage of the tail exit. However, this would reflect the interaction between the helix charges and the tail rather than between the helix charge and the TR, which is the effect found here. In particular, the fact that the in/out ratio does not change when the tail becomes long enough, indicates that the helix insertion is the rate limiting in these cases.

SI Rationales for the Simplifications and the Robustness of the Calculations and the Corresponding Conclusions

Justification of the Model. The present study may seem to some readers as an extremely oversimplified collection of qualitative assumptions. In fact, some readers may assume that all-atom targeted molecular dynamics (TMD) studies must be much more reliable than the current calculations. This, however, may reflect the familiar confusion of the rigor of a model with its ability to converge and to give meaningful results. In fact, CG models are at present significantly more reliable that full atomistic models in capturing the action of, for example, molecular motors (29, 30).

In fact, when we started the present study, we explored the performance of the PMF of the CG model and found them (not surprisingly) to involve major instabilities. The problem is that we are looking on almost a folding problem and that we are interested in small energy differences such as the effects of
mutations (see also below). Attempts to take the points generated in the preparing the PMF calculations and to average the corresponding CG energies provided better but still unstable results for our purpose of estimating mutational effect. Here, we note that the main strength of the CG is in obtaining the electrostatic energy of the simplified side chains because the main chain is described explicitly with the regular instability of all-atom models. In fact, the main power of the CG model is in looking on stabilities of proteins in folded structures or in exploring the energetic in well-defined paths without significant fluctuations of the main chain [e.g., the remarkable successes in modeling the landscape of F1-ATPase (30)]. It also appeared that the interaction with the L6/7 and L8/9 loops of the TR added to the instability, because in the SecA-bound TR structure these loops locate on top of the TR channel (in contrast to ribosome-bound TR structure, where these loops are tilted away from the channel axis by the interaction with the ribosome). Eventually, we found out that the instability can be reduced drastically by performing a short TMD without significant relaxation and then evaluating the CG effective energy (which actually corresponds to the free energy) at each point along the TMD path. The next crucial step in obtaining stable results was made by fixing all of the TR residues, because otherwise the calculated results become rather unstable and less informative. Similarly in exploring the in and out energies we found it useful to use the same geometries of the SP main chain and only mutated the side chains to reverse their order.

Of course, we are very well aware of the need for sampling; we also introduced many key studies of sampling with explicit (e.g., ref. 31) and implicit (e.g., ref. 32) protein model, but the whole point of the CG model is that it already captures a large part of the averaging process (the same is true with the use of a proper effective dielectric). At any rate, the present study did focus on the most effective current ways of obtaining stable results and reproducing mutational trend rather than on the most rigorous treatment (which at present cannot provide what is needed here).

Of course, part of the difficulties in moving to more explicit treatment reflected the fact that the ribosome was not included explicitly, the fact that the tail was sometimes truncated, and the fact that we considered just the insertion of a preformed helix. The effect of the ribosome was estimated by considering the effect of its charges, as well as its effect on the TR structure. It was found that the effect of the ribosome on the in/out distribution is not likely to account for the mutational effects considered here. Similarly, we have considered the effect of the tail and showed that it also does not account for the mutational effects either. The justification of using a preformed helix model was given in Results in the main text.

Overall, we believe that calculations are sufficiently reliable to (i) establish qualitatively the general shape of the insertion profile and (ii) show that the calculated electrostatic effect of some key mutations occurs at or near the barrier of the insertion profile.

Analyzing the Origin of ΔG_app. In the main text, we explored some of the main options for the meaning of ΔG_app. Here, we expand this issue, starting by examining the effect of replacing central Leu residue on the H helix designed in ref. 33 by Arg and Glu residues and calculating the TR insertion, translocation, and membrane insertion profiles for each protein using our CG approach (Fig. S6A). Using the data for the relative free energies of states G and H [ΔG(G→F−G) and ΔG(F−H)] and the barrier heights for F−G and F−H parts of the profile [ΔG^eff(F−G) and ΔG^eff(F−H)], we compare the changes in these values when central Leu is mutated to Arg or Glu. The correlation plot for ΔG^eff versus ΔΔG (Fig. S6B) showed a linear dependence, which is an indication of LFER. The results are summarized in Fig. S6 and Table S3. As can be seen from Fig. S6, we have a reasonable LFER, which may well be the origin of ΔG_app. Note that LFER is a well-accepted concept in studies of the related problem of protein folding (34) and in studies of chemical and enzymatic reactions (21, 35). Thus, one possibility is that ΔG_app reflects LFER (if the back barriers from H and G are very high).

Of course, we would also like to further clarify and extend the options considered in the main text. To simplify the analysis, we can just concentrate on the part of moving from the TR to the membrane and from the TR to water. In this case (if the activation barriers of going forward and backward are not too large), we will obtain a partition between the H helix in water and the H helix in both the TR and the membrane, which will be better described as equilibrium between the TR and water and that between the TR and the membrane. We can, in fact, obtain a general solution once we assume that the forward and backward rate constants from the TR to both the membrane (M) and to water (W) are sufficiently large so that we have equilibration for each process. In this case, we can use TR, M, and W for F, G, and H, respectively, and write the following:

\[
K_1 = W/TR \\
K_2 = M/TR \\
TR + M + W = 1.
\]

We can also write the following:

\[
K_{app} = (M + TR)/W. \tag{S3}
\]

Note that here we added the fraction that is still in the TR as part of the membrane because we cannot separate them experimentally. Solving Eqs. S2 and S3, we obtain:

\[
K_{app} = (1 + K_2)/K_1. \tag{S4}
\]

This important result (which has not been pointed out before) corresponds to the water/membrane equilibrium and not to the membrane/TR equilibrium. In the limit where K_2 and K_1 are significantly larger than 1, we find that K_{app} can be approximated by K_2/K_1. In the other limit of K_2 << 1, we find that K_{app} = 1/K_1. Interestingly, when K_{app} is ~1, which is the range chosen in designing the H helix of ΔG_app, we can have either K_2 ~ K_1 >> 1, or K_2 << 1 and K_1 ~ 1. Thus, it seems to us that the attempts to invoke the membrane/TR equilibrium as the factor that determines ΔG_app (K_{app} = K_2) is not justified.

As much as the elusive question about ΔG_app of Arg is concerned, we can make the following analysis. (i) In the range of K_2 ~ K_1 >> 1, we have ΔG_app ~ RT ln(K_2/K_1). For the apparent free energy of Arg relative to Leu, we can get the following:

\[
\Delta G_{app} (\text{Leu} \rightarrow \text{Arg}) = \Delta G_{app} (\text{Arg}) - \Delta G_{app}^0 (\text{Leu}) \approx RT [\ln(K_2/K_1) - \ln(K_0^0/K_0^1)], \tag{S5}
\]

where K_1, K_2, and K_0^0, K_0^1 are the equilibrium constants for the H helix with central Arg and Leu, respectively. Using data from Table S3, we obtain the following:

\[
\Delta G_{app} (\text{Leu} \rightarrow \text{Arg}) = \Delta G(W \rightarrow M)_{\text{Arg}} - \Delta G(W \rightarrow M)^0 = 3.21 \text{ kcal/mol.} \tag{S6}
\]

This situation corresponds to the equilibrium between water and membrane. This might be the case when Arg in the center of the membrane is stabilized by other helices in the way considered in our previous work (2), where we estimated the energy of moving Arg from water to the membrane to be around 3–5 kcal/mol. A similar value will be obtained if we have the combination of snorkeling, helix tilting, and helix sliding (3, 4). (ii) In the range of K_2 << 1, we have K_{app} = 1/K_1 and ΔG_{app} ∼ RT ln(K_1). Now the apparent free energy of Arg relative to Leu will be given by:

\[
\Delta G_{app} (\text{Leu} \rightarrow \text{Arg}) = \Delta G(W \rightarrow M)_{\text{Arg}} - \Delta G(W \rightarrow M)^0 = 3.21 \text{ kcal/mol.} \tag{S6}
\]
\[ \Delta \Delta G_{\text{app}} (\text{Leu} \rightarrow \text{Arg}) = \Delta G_{\text{app}} (\text{Arg}) - \Delta G_{\text{app}}^0 \approx RT \ln \left[ \frac{[\text{K}_1]}{[\text{K}^0_1]} \right], \]  

and using data from Table S3, we will get the following:

\[ \Delta \Delta G_{\text{app}} (\text{Leu} \rightarrow \text{Arg}) = \Delta \Delta G (\text{W} \rightarrow \text{TR})_{\text{Arg}} - \Delta \Delta G (\text{W} \rightarrow \text{TR})^0 = 6.56 \, \text{kcal/mol}. \]  

In this case, we have an equilibrium between water and the TR. This corresponds to the situation when Arg is in the center of the membrane and is not stabilized by another helix. In this case, the energy of moving from water to membrane is \(\sim 20 \, \text{kcal/mol}\), which has been the origin of the controversy in the field. However, as we show here, the measurements cannot determine cases of extreme instability in the membrane because the H helix will be in the TR. In both cases, our estimates are at the observed range.

At any rate, our main point is that any actual proposal that the equilibrium between the TR and the membrane determines \(\Delta \Delta G_{\text{app}}\) should be formulated in a similar way to the present analysis, because otherwise such a proposal cannot be verified. Here, obtaining the free-energy profile even in a very qualitative way forces one to formulate the problem in a clear way.
Fig. S1. The main trend obtained in the experiment of ref. 1. The figure describes the observed dependence of the percentage of the translocated C terminus, which is proportional to the in/out ratio, and also presents the corresponding calculated trend, which is rationalized in SI Text. The calculated effect of moving down vertically (from a short helix to a long one) is consistent with the reduction of the in/out ratio, whereas moving horizontally is consistent with the observed increase in the in/out ratio (see main text).


**H1ΔLeu22 mutants:**

RR: +MGRR[LLLLLLLLLLLLLLLLLL]GSQN
PR: +MGPR[LLLLLLLLLLLLLLLLLL]GSQN
PH: +MGPH[LLLLLLLLLLLLLLLLLL]GSQN
21-residue tail: SQLQEELRGLETFSNSTAST

**60[H1](+1):**

PPPPQPLQLRLCSGPR[LLLLLGLSLLLVVVCVI]GSQNSQKLRL

**40[Leu16](+5):**

QPQLQDLCSPD[LLLLLLLLLLLLLL]GSQNSQKLRL

Fig. S2. The sequences of the model proteins used to study the membrane protein topogenesis. RR, PR, and PH are the mutants of H1ΔLeu22 SP truncated to 30 residues. The extra 21 residues were added to the C terminus of H1ΔLeu22 mutants to estimate the effect of the tail on the CG profile. The names of the experimental constructs 60[H1](+1) and 40[Leu16](+5) SPs were kept, although they were truncated to 40 and 51 residues, respectively. Charged residues are shown in bold. The N-terminal charge of the H1ΔLeu22 mutants was added to the first amino acid.
Fig. S3. Justifying the use of the helical model. The picture depicts tentative free-energy profiles for constrained helical peptide (solid line), for unconstrained peptide (dashed and dot-dashed lines), and for TR insertion in the presence of the ribosome (dashed line). In the case of the unconstrained peptide, we depicted two possible trends. Overall, the figure reflects the assumption that the helical model provides a qualitative way of estimating the trend in the profile for the real process that involves relaxation from the helical structure.
Fig. S4. Qualitative free-energy profiles for different insertion pathways. (A) The profiles for the retention model. The SP insertion in \(N_{in}\) (solid line) and \(N_{out}\) (dashed line) orientations consist of TR insertion (in blue), membrane insertion (in red), and translocation (in green) parts. State A corresponds to the SP in the cytoplasm, state B to the SP inside the TR, state C to the SP inserted into the membrane and state D to the SP in the exoplasm. (B) SP insertion model in which inversion to \(N_{in}\) orientation occurs simultaneously with polypeptide synthesis [step \(A\rightarrow B\)]. (C) A model where the head-on insertion [through \(A\)] is followed by the inversion to \(N_{in}\) orientation [through \(A\)].
Fig. S5. Kinetic schemes for different insertion models. (A) A scheme for the retention model. The numbers are taken from the calculation for RR SP after scaling them by 0.68 to fit $\Delta g^\circ(\text{A}_{\text{out}}\rightarrow\text{B}_{\text{out}})$ to 20 kcal/mol. The numbers in parentheses represent the estimated free energy after taking into account the missing effects discussed in SI Text. (B and C) The kinetic schemes for the two inversion models depicted in Fig. S4.
**Fig. S6.** (A) The CG free-energy profiles for H helices with a central Arg, Glu, and Leu (denoted as cR, cE, and cL, respectively). The simulated H helix has the following sequence: AAAALALALXLALALAAAA, where X is R, E or L. The energy in the membrane is most probably too negative as indicated by the tentative dashed lines. (B) Using the correlation between $\Delta G^\ddagger$ and $\Delta G$ to examine the LFER idea. $\Delta G^\ddagger$ is the change in the barrier height of the F $\rightarrow$ G and F $\rightarrow$ H parts of the profile when cL is mutated to cR or cE peptide. $\Delta G$ is the change in the relative free energies of states G and H when cL is mutated to cR or cE.

### Table S1. Examination and calibration of the helix insertion energy

<table>
<thead>
<tr>
<th>Model</th>
<th>$\Delta G_{\text{elec}}$</th>
<th>$\Delta G_{\text{hyd}}$</th>
<th>$\Delta G_{\text{solv}}$</th>
<th>$\Delta G_{\text{HB}}$</th>
<th>$\Delta G_{\text{tot}}$</th>
<th>$\Delta G_{\text{tot}}$ (wat $\rightarrow$ mem)</th>
<th>$\Delta G_{\text{est}}$ (wat $\rightarrow$ mem)</th>
</tr>
</thead>
<tbody>
<tr>
<td>polyA_wat</td>
<td>0.00</td>
<td>-4.44</td>
<td>-12.65</td>
<td>-5.12</td>
<td>-22.21</td>
<td>-</td>
<td>-4</td>
</tr>
<tr>
<td>polyA_mem</td>
<td>0.00</td>
<td>-16.36</td>
<td>-0.49</td>
<td>-9.90</td>
<td>-26.75</td>
<td>-4.55</td>
<td>-</td>
</tr>
<tr>
<td>M2δ_wat</td>
<td>-0.74</td>
<td>-4.96</td>
<td>-19.24</td>
<td>-0.23</td>
<td>-25.17</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M2δ_mem</td>
<td>0.16</td>
<td>-26.13</td>
<td>-6.25</td>
<td>-2.17</td>
<td>-34.40</td>
<td>-9.22</td>
<td>-8.7</td>
</tr>
<tr>
<td>polyL_wat</td>
<td>0.00</td>
<td>-5.24</td>
<td>-8.52</td>
<td>-2.15</td>
<td>-15.91</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>polyL_mem</td>
<td>0.00</td>
<td>-34.72</td>
<td>-0.39</td>
<td>-5.11</td>
<td>-40.23</td>
<td>-24.32</td>
<td>-3.0</td>
</tr>
<tr>
<td>RR_wat</td>
<td>0.45</td>
<td>-11.99</td>
<td>-19.82</td>
<td>-3.91</td>
<td>-35.27</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RR_mem</td>
<td>1.18</td>
<td>-54.82</td>
<td>-7.47</td>
<td>-7.88</td>
<td>-68.99</td>
<td>-33.72</td>
<td>-</td>
</tr>
</tbody>
</table>

Energies are in kilocalories per mole; $\Delta G_{\text{tot}}$, $\Delta G_{\text{elec}}$, $\Delta G_{\text{hyd}}$, $\Delta G_{\text{solv}}$, and $\Delta G_{\text{HB}}$ are the total free energy (relative to state A), the electrostatic energy, the hydrophobic contribution, the main chain solvation, and the hydrogen bond contribution, respectively. $\Delta G_{\text{tot}}$ (wat $\rightarrow$ mem) is the free-energy change for moving the peptide from water to membrane. $\Delta G_{\text{est}}$ (wat $\rightarrow$ mem) designates estimated values of the free-energy change, which includes the following: the observed value for the 20-mer poly(A) (1, 2), the value for the 23-aa M2δ segment of the nicotinic acetylcholine receptor, estimated as the average of two CG calculations (3, 4), and the value for the 12-mer poly(L) estimated from microscopic simulations (5).

Table S2. Energetics of the insertion process for the different systems

<table>
<thead>
<tr>
<th>SP</th>
<th>System</th>
<th>ΔG‡(A→N)</th>
<th>ΔG‡(A→out)</th>
<th>Δ∠G‡(A→out-A→N)</th>
<th>ΔG(B→in)</th>
<th>ΔG(B→out)</th>
<th>Δ∆G(B→out-B→in)</th>
<th>Δ∆Gexp (Nout→Nin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1ΔLeu22</td>
<td>RR</td>
<td>25.11</td>
<td>29.49</td>
<td>-4.38</td>
<td>-7.03</td>
<td>-6.56</td>
<td>-1.38</td>
<td>-0.82</td>
</tr>
<tr>
<td></td>
<td>PR</td>
<td>23.72</td>
<td>26.36</td>
<td>-2.64</td>
<td>-9.17</td>
<td>-9.11</td>
<td>-0.66</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>21.42</td>
<td>22.25</td>
<td>-0.83</td>
<td>-11.01</td>
<td>-12.06</td>
<td>1.05</td>
<td>1.30</td>
</tr>
<tr>
<td>60[H1]+1</td>
<td>R74(74)E</td>
<td>15.91</td>
<td>23.71</td>
<td>-7.80</td>
<td>-7.94</td>
<td>-2.08</td>
<td>-5.86</td>
<td>-0.62</td>
</tr>
<tr>
<td></td>
<td>K264(284)E</td>
<td>17.76</td>
<td>25.12</td>
<td>-7.36</td>
<td>-6.06</td>
<td>-0.67</td>
<td>-5.39</td>
<td>-0.50</td>
</tr>
<tr>
<td></td>
<td>Wt</td>
<td>17.73</td>
<td>24.57</td>
<td>-6.84</td>
<td>-6.64</td>
<td>-1.10</td>
<td>-5.54</td>
<td>-0.26</td>
</tr>
<tr>
<td></td>
<td>Q234(261)R</td>
<td>17.51</td>
<td>24.07</td>
<td>-6.56</td>
<td>-7.10</td>
<td>-1.59</td>
<td>-5.51</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>Q93(93)R</td>
<td>19.37</td>
<td>24.95</td>
<td>-5.58</td>
<td>-5.98</td>
<td>-0.59</td>
<td>-5.39</td>
<td>0.17</td>
</tr>
<tr>
<td>40[Leu16]+5</td>
<td>K37(65)E</td>
<td>31.11</td>
<td>29.04</td>
<td>2.07</td>
<td>-3.50</td>
<td>5.37</td>
<td>-8.87</td>
<td>-1.02</td>
</tr>
<tr>
<td></td>
<td>D6(34)R</td>
<td>31.33</td>
<td>26.17</td>
<td>5.16</td>
<td>-2.90</td>
<td>3.44</td>
<td>-6.33</td>
<td>-0.98</td>
</tr>
<tr>
<td></td>
<td>K264(284)E</td>
<td>30.54</td>
<td>23.46</td>
<td>7.08</td>
<td>-6.86</td>
<td>0.35</td>
<td>-7.20</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>R74(74)E</td>
<td>27.11</td>
<td>22.95</td>
<td>4.16</td>
<td>-9.07</td>
<td>1.90</td>
<td>-10.96</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>Wt</td>
<td>29.28</td>
<td>22.26</td>
<td>7.02</td>
<td>-7.89</td>
<td>0.80</td>
<td>-8.68</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Q234(261)R</td>
<td>28.87</td>
<td>21.65</td>
<td>7.22</td>
<td>-8.66</td>
<td>1.13</td>
<td>-9.78</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Q93(93)R</td>
<td>29.58</td>
<td>21.85</td>
<td>7.73</td>
<td>-7.23</td>
<td>1.75</td>
<td>-8.97</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Evaluating the energetics of mutations in different systems. The three mutants of the H1ΔLeu22 peptide studied were: RR, PR, and PH (see sequence in Fig. S2). The system name indicates the relevant TR mutation. The residue number represents its position in SecY, where the corresponding position in Sec61 is given in parentheses. For the 40[Leu16]+5 peptide, the effect of the SP mutations was studied in addition to the TR mutations. The residue number for the SP mutation is for the simulated SP (see sequence in Fig. S2) where the corresponding position in the original peptide is given in parentheses. ΔG‡(A→out) and ΔG‡(A→out) are the barrier heights for the SP insertion into the TR in Nout and Nin orientations, respectively. ΔG(B→out) and ΔG(B→out) are the free energies of the SP inside the TR in Nin and Nout orientations, respectively. ΔGexp(Nout→Nin) is the preference of the Nin orientation versus the Nout orientation obtained from the experimental results of Goder and Spiess (1), who reported the percentage data for C terminus translocated SPs. To get the free energy, we first converted the insertion probability \(P = f(N_{in})/f(N_{in}) + f(N_{out})\), where \(f(N_{in})\) and \(f(N_{out})\) are the fractions of SPs with Nin and Nout orientations, respectively, to the equilibrium constant, \(K = f(N_{in})/f(N_{out})\). The result can be converted to the free energy using the relationship \(\Delta G = -RT \ln K\). The resulting free energy would represent the free energy of inserting SP in Nin orientation with respect to the one in Nout orientation, which we denote as \(\Delta G_{exp}(N_{out}→N_{in})\). It should be noted that experiment was done with yeast Sec61p translocon, whereas in calculations we used bacterial SecY structure from ref. 2 (Protein Data Bank ID code 3DIN). Energies are in kilocalories per mole.


Table S3. Examining possible LFER in the insertion process

<table>
<thead>
<tr>
<th>System</th>
<th>ΔΔG(F→G)</th>
<th>ΔΔG_{C→I}(F→G)</th>
<th>ΔΔG(F→H)</th>
<th>ΔΔG_{C→I}(F→H)</th>
<th>Δg exploited in F→G</th>
<th>Δg exploited in F→H</th>
</tr>
</thead>
<tbody>
<tr>
<td>cR</td>
<td>-19.76</td>
<td>-3.35</td>
<td>-6.56</td>
<td>-6.56</td>
<td>1.20</td>
<td>1.20</td>
</tr>
<tr>
<td>cE</td>
<td>-18.04</td>
<td>-1.63</td>
<td>-4.24</td>
<td>-4.24</td>
<td>1.31</td>
<td>1.31</td>
</tr>
<tr>
<td>cL</td>
<td>-16.41</td>
<td>2.18</td>
<td>1.45</td>
<td>1.45</td>
<td>8.97</td>
<td>8.97</td>
</tr>
</tbody>
</table>

ΔΔG(F→G) and ΔΔG(F→H) are the free energies of states G and H relative to state F, respectively. ΔΔG_{C→I}(F→G) and ΔΔG_{C→I}(F→H) are, respectively, the differences in the relative free energies of states G and H when cL is mutated to either cR or cE (indicated as i). Δg exploited in F→G and Δg exploited in F→H are the barriers for moving H helix from the TR into the membrane and to the exoplasm. Δg exploited in F→G and Δg exploited in F→H are the changes in barriers when cL is mutated to either cR or cE. Energies are in kilocalories per mole.