Energetics of side-chain snorkeling in transmembrane helices probed by nonproteinogenic amino acids

Karín Öjemalm*a,1, Takashi Higuchib,1, Patricia Laraca, Erik Lindhalla, Hiroaki Sugab, and Gunnar von Heijne*a,c,2

*Department of Biochemistry and Biophysics, Center for Biomembrane Research, Stockholm University, SE-106 91 Stockholm, Sweden; †Department of Chemistry, Graduate School of Science, University of Tokyo, 113-0033 Tokyo, Japan; and ‡Science for Life Laboratory, Stockholm University, SE-171 21 Solna, Sweden

Edited by Barry Honig, Howard Hughes Medical Institute, Columbia University, New York, NY, and approved August 9, 2016 (received for review April 28, 2016)

Cotranslational translocon-mediated insertion of membrane proteins into the endoplasmic reticulum is a key process in membrane protein biogenesis. Although the mechanism is understood in outline, quantitative data on the energetics of the process is scarce. Here, we have measured the effect on membrane integration efficiency of nonproteinogenic analogs of the positively charged amino acids arginine and lysine incorporated into model transmembrane segments. We provide estimates of the influence on the apparent free energy of membrane integration (∆Gapp) of “snorkeling” of charged amino acids toward the lipid–water interface, and of charge neutralization. We further determine the effect of fluoroine atoms and backbone hydrogen bonds (H-bonds) on ∆Gapp. These results help establish a quantitative basis for our understanding of membrane protein assembly in eukaryotic cells.

Almost all helical integral membrane proteins are inserted cotranslationally into the endoplasmic reticulum (ER) membrane in eukaryotic cells or the cytoplasmic membrane in prokaryotic cells (1). Insertion is mediated by Sec-type translocons: the Sec61 complex in eukaryotes and the homologous SecYEG complex in prokaryotes. The Sec translocon serves as a protein-conducting channel through which polypeptide can be translocated across the membrane, but it can also open laterally toward the surrounding lipid bilayer, thereby allowing membrane integration of hydrophobic transmembrane helices (1).

Using a cotranslational insertion assay, we previously measured the contribution of each of the 20 natural amino acids to the membrane integration efficiency of model transmembrane segments, and derived a “biological” hydrophobicity scale that assigns an apparent free-energy of membrane integration (∆Gapp) of “snorkeling” of charged amino acids toward the lipid–water interface, and of charge neutralization (2). To further probe the physicochemical basis for membrane insertion, we also analyzed a series of aliphatic and aromatic nonproteinogenic amino acids, using the same insertion assay (3). This approach made it possible to quantitate the “hydrophobic effect” during membrane protein insertion, giving nonpolar solvation energy parameters values of −10 cal/(mol·Å2) and −7 cal/(mol·Å2) for aliphatic and aromatic surface area, respectively.

Here, we extend our analysis of nonproteinogenic amino acids to analogs of the positively charged amino acids Arg and Lys. We provide estimates of the influence on the apparent free energy of membrane integration (∆Gapp) of “snorkeling” of charged amino acids toward the lipid–water interface, and of charge neutralization. We further determine the effect of fluoroine atoms and backbone H-bonds on ∆Gapp.

Results

Membrane-Insertion Assay. To examine the effects of nonproteinogenic amino acid analogs on Sec61-mediated insertion of model hydrophobic segments (H segments) into the ER membrane, we used a previously described Lep construct (3) translated in vitro in the presence of ER-derived dog pancreas rough microsomes (Fig. L4). In this assay, the relative extents of mono- and di-glycosylation provide a direct measure of the efficiency of translocon-mediated insertion of a given hydrophobic segment (H segment). A typical experiment is shown in Fig. 1B; note that the suppression efficiency is ~50% (Fig. 1B, lane 3). An apparent equilibrium constant for membrane insertion of a given H segment, Kapp, is defined as Kapp = Igg/Igg, where Igg and Igg are the respective intensities of the bands representing the singly and doubly glycosylated forms of the protein (the two uppermost bands in Fig. 1B, lane 4). Kapp is then converted to an apparent free energy difference ∆Gapp = −RT ln (Kapp), where R is the gas constant and T the absolute temperature (T = 293 K).

A single nonproteinogenic amino acid “X” was inserted in defined positions in H segments of composition [6L,12A,1X] or [4L,14A,1X] by introduction of an amber stop codon (TAG) in the corresponding position in the gene, and carrying out in vitro translation in the presence of a suppressor tRNA_Sup charged with the nonproteinogenic amino acid in question. Charging of the tRNA_Sup was done using Flexizyme (4) (Materials and Methods).

Snorkeling in Arginine Analogs. The guanidinium group on the Arg side chain has such a high pKₐ that it remains protonated even when buried in a hydrophobic environment (5). To explore the combined effects of charge and side-chain length [i.e., snorkeling (6, 7)] on membrane insertion, we analyzed a series of Arg analogs with increasingly long side chains of three to eight methyl groups, as well as an analog (l- canavanine), where the δ carbon is replaced by an oxygen atom, thereby lowering the pKₐ of the guanidinium group on the Arg side chain has such a high pKₐ that it remains protonated even when buried in a hydrophobic environment (5).

Significance

Membrane proteins are central players in all cells, and their structure and function are under intense study. However, we still lack a detailed understanding of the process whereby they are integrated into biological membranes. Most membrane proteins are integrated cotranslationally into the membrane bilayer. Although the energetics that drive membrane protein integration are known in outline, detailed studies are difficult because the naturally occurring amino acids represent only a limited set of side-chain chemistries. Here we use synthetic, nonproteinogenic amino acids engineered into a transmembrane segment to systematically probe the energetics of membrane insertion in a way not possible with the set of natural amino acids.


This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. Email: gunnar@dbb.su.se.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1606776113/-/DCSupplemental.
methyl groups in the side chain increases. In particular, addition of one methyl group to an Arg side chain placed in position 10 in an H segment of composition [6L,12A,1R] reduces \( \Delta G_{app} \) by \(-0.30\) kcal/mol. Going from four to six methyl groups leads to a further reduction by \(-0.75\) kcal/mol (\(-0.38\) kcal/mol per carbon), and from six to seven methyl groups by \(-0.63\) kcal/mol. Longer side chains with seven and eight methyl groups promote >90% insertion of the H segment, precluding the calculation of accurate \( \Delta G_{app} \) values, and were therefore reanalyzed in H segments of composition [4L,14A,1R]; going from seven to eight methyl groups reduces \( \Delta G_{app} \) by \(-0.46\) kcal/mol when the Arg analog is in position 10 (Fig. 2B). These stepwise reductions in \( \Delta G_{app} \) are larger than the value obtained previously for linear aliphatic side chains (\(-0.26\) kcal/mol per carbon) (3). Subtracting the aliphatic contribution, we find that the \( \Delta \Delta G_{app} \) for moving the charged guanidinium group one methyl group (\(-1.2\) Å) closer to the membrane interface varies, as shown in Fig. 2C, i.e., the polarity gradient in the membrane is maximal in a region located approximately half-way between the center of the membrane and the membrane–water interface. This finding is consistent with neutron and X-ray diffraction measurements of the water distribution across lipid bilayers (9), indicating that the ER membrane has a similar polarity gradient as a model lipid bilayer.

The uncharged carbamide analog of Arg has a \( \Delta G_{app} \) that is \(-0.30\) kcal/mol lower than that of Arg when placed in position 10 in the H segment (Fig. 2D). Lengthening of the side chain of one carbamide analog by one methyl group leads to a further reduction of \( \Delta G_{app} \) by \(-0.24\) kcal/mol, very similar to the value found for purely aliphatic side chains (3). The most notable difference between the Arg \( \Delta G_{app} \) profile and the profiles for the two carbamide analogs, however, is that the latter are almost perfectly symmetric around position 10, whereas the C-terminal half of the Arg profile has lower \( \Delta G_{app} \) values than the N-terminal half. This finding is consistent with the previously demonstrated ability of cytoplasmically located positively charged residues to lower \( \Delta G_{app} \) (10), a reflection of the so-called “positive inside rule” (11).

The Arg analog l-canavanine, finally, has a substantially lower \( \Delta G_{app} \) than Arg (\( \Delta \Delta G_{app} = -0.64 \) kcal/mol) when placed in position 10 (Fig. 2D), but retains the asymmetric shape of the \( \Delta G_{app} \) profile. With a \( pK_a \) of 7.0, the formal charge of l-canavanine at pH 7 is +0.5; apparently, this is enough to make it behave as a positively charged residue, but with a significantly lower cost for membrane insertion than has Arg.

**Snorkeling in Lysine Analogs.** Because of the difficulty in shortening the side-chain length of Lys without inducing the formation of lactones, we were unable to investigate such lysine analogs containing different numbers of methyl groups. Instead, we modified the terminal amino group by acetylation and methylation. Acetylation removes the positive charge, and reduces \( \Delta G_{app} \) (\( \Delta \Delta G_{app} = -0.67 \) kcal/mol in position 10 in the H segment) (Fig. 3A). Remarkably, shortening of the acetylated Lys side-chain by one to three methyl groups only marginally increases \( \Delta G_{app} \); in position 10, \( \Delta G_{app} \) changes by +0.23 kcal/mol when one methyl group is removed (as expected from the reduced aliphatic surface area), but then remains constant upon removal of the second methyl group and even decreases slightly when the third methyl group is removed (Fig. 3A). Thus, for AcLys(2) and AcLys(3), the increase in \( \Delta G_{app} \) expected from the removal of aliphatic surface area and deeper burial of the polar moiety is compensated for by some other interaction, possibly H-bonding of the side chain to the backbone as often seen for Ser, Thr, and Cys residues (12). To test whether this is a physically plausible explanation, we performed molecular dynamics (MD) simulations of AcLys and AcLys(C2–C4) introduced in position 10 in a [6L,12A,1X] H segment embedded in a palmitoyloleoylphosphatidylcholine (POPC) bilayer. In support of our hypothesis, and as illustrated by the representative snapshots in Fig. 3C, the side chain of AcLys makes H-bonds to interfacial water with
AcLys(C3) is alternately H-bonded to water and to the backbone carbonyl of the Ala residue in position 6, both with time-averaged occupancies of 0.3. For AcLys(C2), the H-bonds to water and backbone have time-averaged occupancies of 0.6 and 0.4, respectively, and AcLys(C1) forms a H-bond only with the backbone with a time-averaged occupancy of 0.9. This result is also in agreement with earlier measurements on Gln and Asn, which differ by one methyl group but have nearly overlapping ΔG_{app} profiles (2). In fact, the profiles for AcLys(C2), AcLys(C1), Gln, and Asn in H segments of composition [4L,14A,1X] (2) are all very similar (the ΔG_{app} values for position 10 are within 0.17 kcal/mol), reflecting the similarities in side-chain composition (one oxygen, one nitrogen, two to four methyl groups).

As seen in Fig. 3B, methylation of the Lys side chain reduces ΔG_{app} by only −0.13 kcal/mol in position 10. MD simulations of H segments with Lys and CH$_3$-Lys in position 10 embedded in a POPC bilayer (Fig. 3C) show that the charged amino group snorkels toward the lipid–water interface in both cases, and hence is located in a region where the introduction of additional nonpolar surface area in the form of an extra methyl group contributes little to the overall ΔG_{app}.

Fluorination Increases the Hydrophobicity of a Transmembrane Segment. Fluorine is often used in medicinal chemistry, one reason being that its lipophilic character can make drugs better soluble in lipid membranes and thereby improve their ability to penetrate into cells (13). With this motivation in mind, we measured the effect of fluorine on ΔG_{app} by changing the terminal methyl group on AcLys and AcLys(C2) to −CF$_3$. This modification reduced ΔG_{app} by −0.47 kcal/mol for AcLys and by −1.25 kcal/mol for AcLys(C2) when the fluorinated residue was placed in position 10 in the H segment (Fig. 4). These values compare reasonably well to the literature value of −1.0 kcal/mol for the relative hydration-free energies of CH$_4$ vs. CHF$_3$ (14, 15). The difference between CF$_3$-AcLys and CF$_3$-AcLys(C2) presumably reflects the deeper burial in the membrane of the −CF$_3$ group in CF$_3$-AcLys(C2). The fact that corresponding AcLys and AcLys(C2) H-segments have almost the same ΔG_{app} values
Fig. 3. $\Delta G_{\text{app}}$ profiles for Lys analogs. (A) Acetylated Lys analogs were introduced into H segments of composition [4L, 14A, 1X], where X is the Lys analog. (B) Methylated Lys was introduced into H segments of composition [6L, 12A, 1X]. SDs obtained from three independent experiments are indicated. (C) Representative snapshots of Lys analogs introduced in position 10 in a [6L, 12A, 1X] H segment embedded in a POPC bilayer (headgroups cyan, tails gray). Hydrogen bonds are indicated by yellow dashed lines.

(Fig. 3A) has the interesting consequence that $\Delta G_{\text{app}}$ for CF$_3$-AcLys (C2) H-segments is lower than for the corresponding CF$_3$-AcLys H-segments, despite the longer side chain of the latter.

A Backbone H-Bond Contributes $\sim 0.7$ kcal/mol to $\Delta G_{\text{app}}$. Although the charged and polar side chains studied above do not have much influence on the helical conformation of the peptide backbone of membrane-integrated H segments, proline specifically disrupts helical structure. Proline prevents the formation of backbone H-bonds and kinks the H segment transmembrane helix, and therefore behaves as a polar residue when inserted into an H segment (2). To better understand the energetics of backbone H-bond formation, we tested the Pro analog carboxypiperidine [Pro(C5)], which has an additional $-$CH$_2$ group in the side chain. Unexpectedly, the Pro(C5) H segments had higher $\Delta G_{\text{app}}$ values than the corresponding Pro H segments for all tested positions except the most C-terminal one (Fig. 5A), despite the larger aliphatic surface area of Pro(C5). MD simulations of H segments dissolved in cyclohexane (to mimic a membrane environment) containing either a Pro or a Pro(C5) residue in position 10 show that Pro(C5) H segments had higher $\Delta G_{\text{app}}$ values than the corresponding Pro H segments for all tested positions except the most C-terminal one (Fig. 5A), despite the larger aliphatic surface area of Pro(C5). The average difference in exposed aliphatic surface area between the simulated helices with Pro or Pro(C5) is 33 Å$^2$, corresponding to a predicted difference in $\Delta G_{\text{app}}$ of $\sim 0.33$ kcal/mol. Experimentally, the difference in $\Delta G_{\text{app}}$ between the two H segments is $+0.26$ kcal/mol, hence the loss of 0.8 main-chain H-bonds (mainly between residues 4 and 8) (Fig. 5B) contributes $+0.6$ kcal/mol to $\Delta G_{\text{app}}$, corresponding to $\sim 0.7$ kcal/mol per H-bond. From previous measurements on natural amino acids in position 10, we know that Val and Thr contribute, respectively, $-0.31$ and $+0.52$ kcal/mol to $\Delta G_{\text{app}}$ (16); that is, because the surface areas of a $-\text{CH}_3$ and an $-\text{OH}$ group are similar, the H-bond on Thr contributes $\sim 0.8$ kcal/mol to $\Delta G_{\text{app}}$. The two measures agree that the loss of an H-bond in the interior of the ER membrane is thus associated with an increase in $\Delta G_{\text{app}}$ of 0.7–0.8 kcal/mol, in the same range as observed for the contribution of a membrane-buried H-bond to the stability of membrane proteins in detergent solution (0.5–1.0 kcal/mol) (17).
Uncharged carbamide analogs of Arg nicely illustrate both the effect of aliphatic surface area and the strong effect of positively charged residues located near the cytoplasmic end of a transmembrane helix on $\Delta G_{\text{app}}$. When placed in the center of the H segment (position 10), the uncharged carbamide analog reduces $\Delta G_{\text{app}}$ by $\sim0.30$ kcal/mol compared with a charged Arg residue. In contrast to the $\Delta G_{\text{app}}$ profile for Arg, the profile for the carbamide analog is perfectly symmetric, which has the interesting consequence that Arg has a higher cost of membrane insertion than the uncharged carbamide analog in the luminal two-thirds of the H segment, but a lower cost of insertion in the cytoplasmic third. Adding an extra methyl group to the side chain of the carbamide analog reduces $\Delta G_{\text{app}}$ by the amount expected from the added aliphatic surface area ($\sim0.24$ kcal/mol), while maintaining the symmetric shape of the $\Delta G_{\text{app}}$ profile. Reducing the $pK_a$ of the Arg side chain by replacing the C6 by O markedly reduces $\Delta G_{\text{app}}$, especially in the cytosolic half of the H segment where the side chain is likely snorkeling toward the cytosol.

The Lys analogs that we have analyzed behave as expected: acetylation of the side chain (rendering it uncharged) reduces $\Delta G_{\text{app}}$ by $\sim0.67$ kcal/mol in position 10 of the H segment, twice the amount seen for carbamidation of Arg in the same position. The delocalized charge on the Arg guanidinium group is thus less costly to bury in the membrane than is the charge on the $\sim\text{NH}_2$ group on Lys. Methylation of the amino group on the Lys side chain (that retains its charge) has only a marginal effect on $\Delta G_{\text{app}}$.

Analogs of acetylated lysine (AcLys) with shortened side chains show an interesting behavior: the removal of one methyl group from AcLys leads to the expected increase of $0.23$ kcal/mol in $\Delta G_{\text{app}}$ caused by the reduction in aliphatic surface area, but further shortening of the side chain does not cause any further increase in $\Delta G_{\text{app}}$. This finding is reminiscent of our earlier observation that the $\Delta G_{\text{app}}$ profiles of Asn and Gln are nearly identical (2), despite the difference in charge. An additional methyl group in the Gln side chain, H-bond formation between the shorter side chains and the helix backbone seems to offset the loss of aliphatic surface area in these cases.

Finally, replacing the terminal $\sim\text{CH}_3$ group in the AcLys and AcLys(C2) side chains by $\sim\text{CF}_3$ reduces $\Delta G_{\text{app}}$ by $\sim0.5$ and $\sim1.2$ kcal/mol, respectively, when the residue is in position 10 in the H segment, clearly attesting to the lipophilicity of fluorine in the context of the ER membrane. In addition, a comparison of the energetic effects of introducing either Pro or the Pro(C5) analog allows us to derive a value of $0.7\sim0.8$ kcal/mol for the energy of a backbone H-bond in the membrane environment.

Discussion

Nonproteinogenic amino acids are powerful biochemical tools, in part because they make it possible to vary systematically the physico-chemical characteristics of amino acid side chains in ways that the restricted set of natural amino acids does not. In an earlier study, we used nonproteinogenic amino acids to measure nonpolar solvation energy parameters for aliphatic and aromatic surface areas that apply during cotranslational insertion of transmembrane $\alpha$-helices into the ER membrane (3). We have now used the same approach to study the energetics of membrane insertion of positively charged amino acids, fluorine atoms, and backbone H-bonds.

Our analysis of how charged Arg analogs of increasing side-chain length affects $\Delta G_{\text{app}}$ shows that snorkeling toward the membrane–water interface can significantly reduce the cost of burying a charged guanidinium group in the membrane; we estimate that moving the guanidinium group $\sim1.2$ Å closer to the interface (i.e., a distance corresponding to the addition of one methyl group to the side chain) can lower $\Delta G_{\text{app}}$ by up to $\sim0.4$ kcal/mol. The maximal effect is seen when the guanidinium group is located approximately half-way between the center of the membrane and the membrane–water interface.

Materials and Methods

Enzymes and Chemicals. All enzymes were purchased from Fermentas, except Phusion DNA polymerase from Finnzyme and SP6 RNA polymerase from Promega. The QuikChange Site-directed Mutagenesis kit and deoxyribonucleotides were from Stratagene, and the Megascript T7 kit was from Ambion. The plasmid pGEM1, the rabbit reticulocyte lysate system and the RNasin were from Promega. Oligonucleotides were from MWG Operon. All chemicals were from Sigma-Aldrich, except DMSO from J. T. Baker Chemicals, ethanol from Kemetyl, and [35S]-methionine from PerkinElmer.

DNA Manipulations. For cloning, a modified version of the Escherichia coli lepB gene (signal peptide) in a pGEM1 vector was used (16). This version harbors an engineered test segment (H segment) in the coding region of the P2 domain between a SpeI cleavage site in codons 226–227 and a KpnI cleavage site in codon 253 (WT lepB codon positions) as well as two glycosylation acceptor sites for N-linked glycosylation at codons 96–98 (G1: Asn–Ser–Thr) and codons 258–260 (G2: Asn–Ala–Thr). The sequences of the H-segment’s flanking regions are: QETKENGIRLSFORESGPGP(G–H segment)—GPGVPPQQAQTVIPPP(SpeI and KpnI cleavage sites underlined). The introduction of amber stop codons (TAG) into the termini of the H-segment-encoding sequence was done by site-directed mutagenesis using Pfu Turbo polymerase. To introduce amber stop codons in or near the middle of the H segment, double-stranded oligonucleotides encoding the H segment with the amber stop (including also GPG/GPGG) and flanked by N-terminal SpeI and C-terminal KpnI sticky ends were first generated by annealing of two pairs of complementary oligonucleotides.
oligonucleotides with overlapping overhangs (each 18–45 nucleotides long), followed by annealing of the pairs via the complementary overhangs and cloning into the LEPI gene between the Spe1 and Kpn1 cleavage sites (16).

**Flexizyme dFx and eFx and ML-tRNA<sup>Sup</sup>** Preparation. Preparation of Flexizyme (dFx, eFx) and mycobacteriophage LS (ML)-derived ML-tRNA<sup>Sup</sup> (tRNA<sup>Sup</sup>) were done using the same protocol. First, double-stranded DNA templates encoding the RNA species and a N-terminal T7 promoter sequence (18, 19) were generated by PCR extension of annealed overlapping oligonucleotides. DNA templates were then amplified by PCR using primers complementary to both ends of the templates, followed by phenol/chloroform extraction and ethanol precipitation. The DNA was used in a second step for transcription by T7 polymerase using the Ambion Megashortscript T7 kit, and the RNA product was isopropanol precipitated and purified over 12% denaturing PAGE. After cutting out the RNA band, RNA was eluted for 2 h in 0.3 M NaCl, ethanol-precipitated, and dissolved in dH<sub>2</sub>O (70–250 μM final concentration).

**Synthesis and Characterization of Amino Acid Derivatives.** Amino acid derivatives were synthesized as described previously (18, 20). For NMR characterization of the amino acid derivatives, see the Supporting Information.

**Acylation of ML-tRNA<sup>Sup</sup>** Flexizyme (dFx in all cases, except for aromatic amino acids where eFx was used) and ML-tRNA<sup>Sup</sup> (tRNA<sup>Sup</sup>) each 250 μM in 71 mM Hepes-K buffer (pH 7.5), with volumes 7 μL, were heated at 95 °C for 2 min and cooled to room temperature over 5 min. One micromolar MgCl<sub>2</sub> (200 mM for dFx and 3 M for eFx reactions) and 2 μL of amino acid-DBE (25 mM, nonaromatic amino acids) or amino acid-CME (25 mM, aromatic amino acids) were added. The reaction was carried out for 2 h on ice and then stopped with 40 μM AcONa (0.3M, pH 5.2). RNA was ethanol-precipitated twice, once with a 0.1 M ethanol AcONa solution (pH 5.2) and once with 95% (v/vol) ethanol. More detailed information can be found elsewhere (18).

**Expression in Vitro and Quantification of Membrane Insertion Efficiency.** All constructs were transcribed for 60 min at 37 °C using a standard SP6 polymerase transcription protocol (21). Resulting mRNA was translated for 80 min at 30 °C in rabbit reticulocyte lysate (40 μL of amino acid-DBE (25 mM, nonaromatic amino acids) or amino acid-CME (25 mM, aromatic amino acids) were added). The reaction was carried out for 2 h on ice and then stopped with 40 μM AcONa (0.3M, pH 5.2). RNA was ethanol-precipitated twice, once with a 0.1 M ethanol AcONa solution (pH 5.2) and once with 95% (v/vol) ethanol. More detailed information can be found elsewhere (18).