Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for N_{in}–C_{out} and N_{out}–C_{in} transmembrane helices

Carolina Lundin*, Hyun Kim*, IngMarie Nilsson*, Stephen H. White†, and Gunnar von Heijne*‡

*Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden; and †Department of Physiology and Biophysics and the Center for Biomembrane Systems, University of California, Irvine, CA 92697-4560

Edited by Ramanujan S. Hegde, National Institutes of Health, Bethesda, MD, and accepted by the Editorial Board September 3, 2008 (received for review May 19, 2008)

Transmembrane α-helices in integral membrane proteins can have two orientations in the membrane: N_{in}–C_{out} or N_{out}–C_{in}. Previous studies of model N_{out}–C_{in} transmembrane segment have led to a detailed, quantitative picture of the ''molecular code'' that relates amino acid sequence to membrane insertion efficiency *in vivo* [Hessa T, *et al*. (2007) Molecular code for transmembrane helix recognition by the Sec61 translocon. *Nature* 450:1026–1030], but whether the same code applies also to N_{in}–C_{out} transmembrane helices is unknown. Here, we show that the contributions of individual amino acids to the overall efficiency of membrane insertion are similar for the two kinds of helices and that the threshold hydrophobicity for membrane insertion can be up to ≈1 kcal/mol lower for N_{in}–C_{out} compared with N_{out}–C_{in} transmembrane helices, depending on the neighboring helices.

membrane protein | positive-inside rule | *Saccharomyces cerevisiae* | topology | translocon

In mammalian cells, most membrane proteins are cotranslationally inserted into the membrane of the endoplasmic reticulum (ER) by the Sec61 translocon (1). During the insertion process, hydrophobic segments in the nascent polypeptide are dispelled laterally from the translocon to form transmembrane (TM) α-helices with either an N_{in}–C_{out} (i.e., with the N terminus in the cytosol) or N_{out}–C_{in} orientation relative to the membrane (2). From the point of view of the translocon, N_{out}–C_{in} TM helices enter the translocation channel as part of a translocating nascent chain, whereas N_{in}–C_{out} TM helices must gate the channel open and presumably remain in or very near the channel during translocation of the C-terminal parts of the nascent chain (Fig. 1). In principle, this means that the sequence requirements for membrane insertion may be different for N_{in}–C_{out} and N_{out}–C_{in} TM helices.

In previous work (3–5), we have carried out a detailed analysis of how different amino acids contribute to the overall efficiency of membrane insertion of TM helices with an N_{out}–C_{in} orientation (also called stop-transfer sequences). Here, we present a similar analysis, but for TM helices of the opposite orientation, i.e., N_{in}–C_{out}. The analysis is done both by using *in vitro* translation of model constructs in the presence of dog pancreas rough microsomes (RMs) and by expression in the yeast *Saccharomyces cerevisiae*. We find that individual amino acids affect membrane insertion in much the same way irrespective of the orientation of the TM helix. We also show that the hydrophobicity required for 50% membrane insertion for a N_{in}–C_{out} TM helix can be as much as ≈1 kcal/mol less than for a N_{out}–C_{in} TM helix; this difference can be explained in part by the influence from a neighboring N_{out}–C_{in} TM helix on the membrane insertion efficiency of the N_{in}–C_{out} helix.

Results

Model Protein and Membrane Insertion Assay. In our previous studies of N_{out}–C_{in} TM helices, we used a construct derived from the *Escherichia coli* leader peptidase (Lep) protein shown in Fig. 2*A Left*. Lep has two transmembrane helices (TM1 and TM2) and a large C-terminal luminal domain (P2). Test segments (H-segments) are inserted into the P2 domain where they are flanked by two engineered acceptor sites for N-linked glycosylation (G1 and G2). The acceptor sites provide a convenient way to measure insertion efficiency into the ER: Lep constructs with H-segments that insert into the membrane are only glycosylated by the luminally disposed

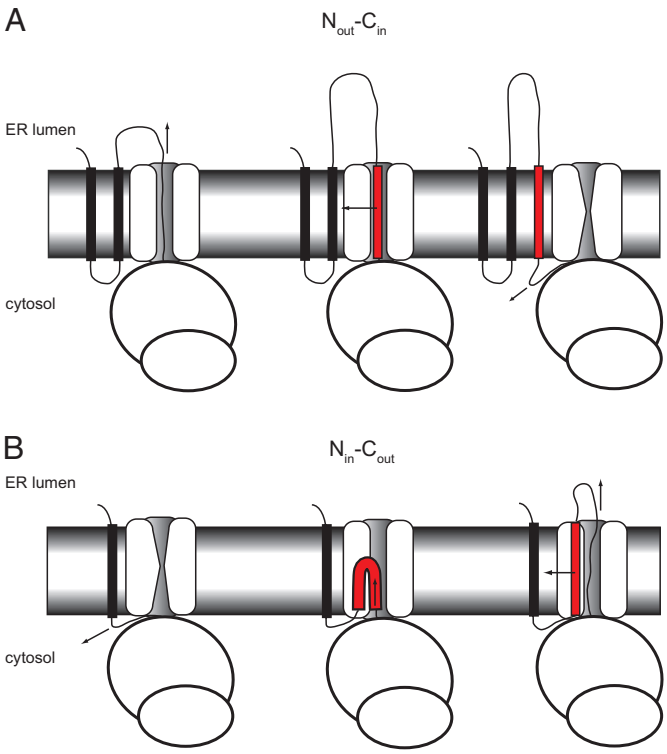

Fig. 1. Membrane insertion of N_{out}–C_{in} and N_{in}–C_{out} transmembrane helices. (*A*) The H-segment (red) is fed into the translocon as part of a translocating nascent chain and integrates into the membrane with an N_{out}–C_{in} orientation. The nascent chain downstream of the H-segment remains in the cytosol. (*B*) The entry of the H-segment into the translocon triggers translocation of the downstream part of the nascent chain. The H-segment integrates into the membrane with an N_{in}–C_{out} orientation.

Author contributions: C.L., H.K., I.N., S.H.W., and G.v.H. designed research; C.L., H.K., and I.N. performed research; C.L., H.K., I.N., S.H.W., and G.v.H. analyzed data; and C.L., H.K., I.N., S.H.W., and G.v.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.S.H. is a guest editor invited by the Editorial Board.

‡To whom correspondence should be addressed. E-mail: gunnar@dbb.su.se.

This article contains supporting information online at www.pnas.org/cgi/content/full/0804842105/DCSupplemental.

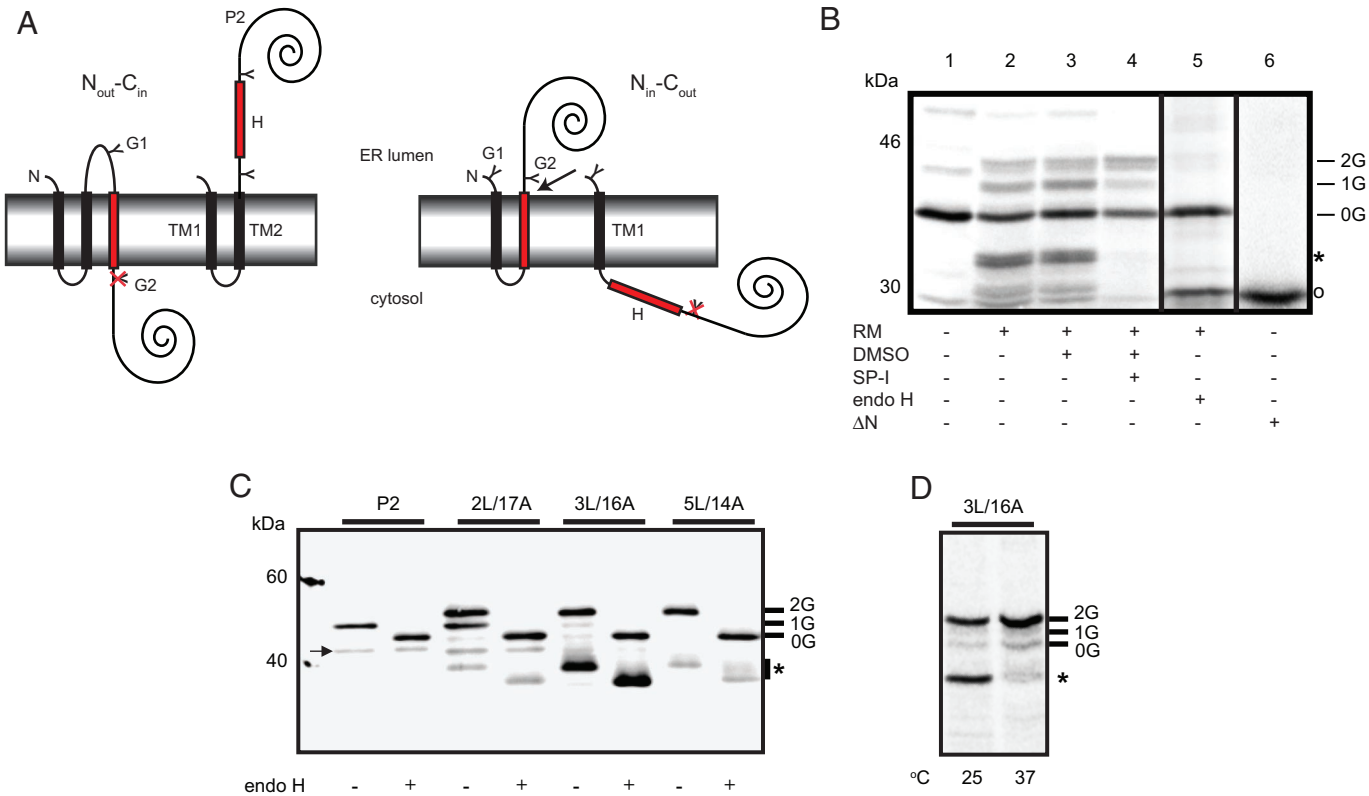

Fig. 2. Membrane integration assay. (*A*) Model proteins. (*Left*) In the construct used to analyze N_{out}–C_{in} H-segments, membrane insertion of the H-segment (red) gives rise to molecules glycosylated only on the G1 site, whereas translocation of the H-segment across the membrane gives rise to molecules glycosylated on both the G1 and G2 sites (5). (*Right*) In the construct used here to analyze N_{in}–C_{out} H-segments, both the G1 and G2 acceptor sites are modified when the H-segments insert into the membrane, whereas only the G1 site is modified when the H-segment remains in the cytosol. The arrow indicates the location of the signal peptidase cleavage site. (*B*) Integration into dog pancreas RMs of an H-segment with the composition 1D/4L/14A (construct 46 in Table S1). A plasmid encoding the Lep construct was transcribed and translated *in vitro* in the absence (−) or presence (+) of RMs, DMSO, the signal peptidase inhibitor *N*-methoxysuccinyl-Ala-Ala-Pro-Val-chloromethyl ketone (SP-I) dissolved in DMSO, or after endo H treatment. An *in vitro* translated version of construct 46 lacking the N-terminal domain up to the GPGG flank at the C-terminal end of the H-segment (construct ΔN) is included in lane 6. (*C*) Integration *in vivo* into the yeast ER of HA-tagged Lep containing H-segments of the indicated compositions (constructs 86, 88–90 in Table S1; the H-segment in the P2 construct is a 19-residue peptide composed of mostly polar amino acids from the Lep P2 domain). One-half of each sample was digested by endo H to remove N-linked glycans before SDS/PAGE and Western blotting with an HA antiserum. (*D*) [^{35}S]Met pulse-labeling of the 3L/16A H-segment construct (89; Table S1) in strain HFY406 that carries a *ts* mutation in the Spc3p subunit of the signal peptidase complex. Cells were pretreated and pulse-labeled at either the permissive (25°C) or nonpermissive (37°C) temperature, and HA-tagged Lep was immunoprecipitated with an HA antiserum. (*B–D*) Unglycosylated (0G), singly glycosylated (1G), and doubly glycosylated (2G) products are indicated. Glycosylated and unglycosylated products resulting from cleavage by signal peptidase are indicated by * and o, respectively, and an unglycosylated cleaved product seen only in yeast is indicated by an arrow in *C*.

oligosaccharide transferase on the G1 acceptor site, whereas constructs in which the H-segment translocates across the membrane are modified on both the G1 and G2 sites.

In the present work, we have used the construct shown in Fig. 2*A Right*. The H-segment now replaces the TM2 helix in Lep, and one acceptor site has been moved from the P2 domain to the N-terminal tail. The membrane-inserted and noninserted forms of the H-segment can again be distinguished by their glycosylation status: when the H-segment forms a TM helix, both the G1 and G2 sites are modified; when it fails to insert into the membrane, only the G1 site is modified. The singly and doubly glycosylated forms of the protein can be readily distinguished and quantified by SDS/PAGE and PhosphorImager analysis of radiolabeled Lep protein translated *in vitro* in the presence of RMs (5) or by Western blotting of HA-tagged protein expressed in yeast. The presence of the TM1 helix ensures that the H-segment is not required for binding to the signal-recognition particle (SRP) and targeting the ribosome–nascent chain complex to the Sec61 translocon (6, 7). The H-segments analyzed here are similar to those used in our previous studies and are composed of a central 19-residue-long stretch of varying hydrophobicity flanked by GGPG…GPGG tetrapeptides intended to "insulate" the central stretch from the surrounding Lep sequence.

Typical gels are shown in Fig. 2 *B* and *C*. When translated in the presence of RMs (Fig. 2*B*), a fraction of the Lep molecules are cleaved to a smaller, glycosylated species (marked *); Treatment with endoglycosidase H (endo H) yields uncleaved and cleaved unglycosylated products (marked 0G and o in lane 5). Cleavage is prevented by inclusion of the signal peptidase inhibitor *N*-methoxysuccinyl-Ala-Ala-Pro-Val-chloromethyl ketone (8) in the translation mix (lane 4), with a concomitant increase in the fraction of doubly glycosylated molecules. The cleaved, unglycosylated product comigrates with a truncated Lep molecule that lacks all N-terminal residues up to the C-terminal end of the H-segment (construct ΔN, lane 6). The signal peptidase active site is located near the luminal surface of the ER membrane (9), and the cleaved form of Lep thus originates from molecules in which the H-segment is inserted into the membrane and the P2 domain is in the lumen. Signal peptidase-mediated cleavage of "internal" signal peptides in multispanning membrane proteins has been reported before in, e.g., viral polyproteins (10) and engineered constructs (11).

Expression in yeast gives similar results (Fig. 2 *C* and *D*). When

the H-segment is not inserted into the ER membrane, only the singly glycosylated form of Lep is seen (construct P2), whereas the doubly glycosylated form predominates for sufficiently hydrophobic H-segments (construct 5L/14A). H-segments of intermediate hydrophobicity give rise to a mixture of singly and doubly glycosylated Lep molecules (construct 2L/17A). A singly glycosylated, truncated form (marked by *) is also seen. This form is absent in a construct expressed at the nonpermissive temperature in a strain carrying a *ts* mutation in the Spc3p component of the signal peptidase complex (12) (Fig. 2*D*). The truncated form thus results from signal peptidase-mediated cleavage of the membrane-inserted H-segment from the luminal side of the membrane. In addition, a weak band corresponding to an unglycosylated cleaved product (marked by an arrow) is visible in constructs with H-segments of low hydrophobicity; we assume that this product results from partial proteolysis of molecules where TM1 but not the H-segment has inserted into the membrane.

To facilitate the comparison between the data reported below and our previous studies of N_{out}–C_{in} TM helices, we express the membrane insertion efficiency of an H-segment as the apparent free energy difference between the inserted and noninserted species:

$$\Delta G_{app} = -RT\ln K_{app} = -RT\ln\left(\frac{f_{2x}+f_c}{f_{1x}}\right),$$

where K_{app} is the apparent equilibrium constant of membrane insertion and f_{1x}, f_{2x}, and f_c denote the fractions of singly glycosylated, doubly glycosylated, and cleaved glycosylated molecules (* in Fig. 2 *B* and *C*), respectively. For constructs expressed in yeast, the fraction of cleaved unglycosylated molecules (arrow in Fig. 2*C*) is included in f_{1x}. Unglycosylated molecules that have not been targeted to the RMs or the yeast ER are ignored, hence $f_{1x} + f_{2x} + f_c = 1$.

Membrane Insertion of Leu/Ala-Based N_{in}–C_{out} H-Segments. For previously analyzed N_{out}–C_{in} H-segments composed of n Leu and $(19 - n)$ Ala residues, ΔG_{app} values measured with RMs depends linearly on n, and 50% insertion (i.e., $\Delta G_{app} = 0$ kcal/mol) is observed for $n \approx 3$ (5). As seen in Fig. 3*A*, the relationship between ΔG_{app} and n is linear also for the N_{in}–C_{out} H-segments, both for RMs (solid line) and yeast (dot-dash line). Interestingly, for the N_{in}–C_{out} H-segments in RMs, $\Delta G_{app} = 0$ kcal/mol for $n \approx 1$–2 rather than 3, i.e., the "threshold hydrophobicity" is somewhat lower in this case. The slope of the curve $\Delta G_{app} = f(n)$ is more negative for yeast than for the mammalian RMs: $\Delta G_{app} = -1.7$ kcal/mol vs. -0.5 kcal/mol per Ala → Leu replacement. The precise yeast value is somewhat uncertain because both the 1L/17A and 3L/16A H-segments have ΔG_{app} values outside the range where our measurements have a high precision ($|\Delta G_{app}| < 1.0$ kcal/mol).

To check whether the identity of TM1 affects these results, we replaced TM1 of Lep in the 1L/18A H-segment construct with more hydrophobic stretches composed of either 18 leucines or 10 leucines and 9 alanines. In the *in vitro* system, the 18L and 10L/9A TM1 segments caused a noticeable reduction (from 60% to 43% or 37%, respectively) in the fraction of molecules with a membrane-inserted H-segment [corresponding to an increase in ΔG_{app} from -0.2 to $+0.2$ or $+0.3$ kcal/mol; see constructs 83–84 in supporting information (SI) Table S1].

We also tested a construct with a 107- instead of a 37-residue loop between TM1 and a 1L/18A H-segment. There was no significant change in ΔG_{app} values between the two constructs (see Table S1 constructs 2 and 85).

Contribution of Different Amino Acids to Membrane Insertion Efficiency. How much do different amino acids contribute to the membrane insertion efficiency of an N_{in}–C_{out} H-segment, and to what extent do these contributions depend on the position of the residue within the H-segment? To address these questions, we

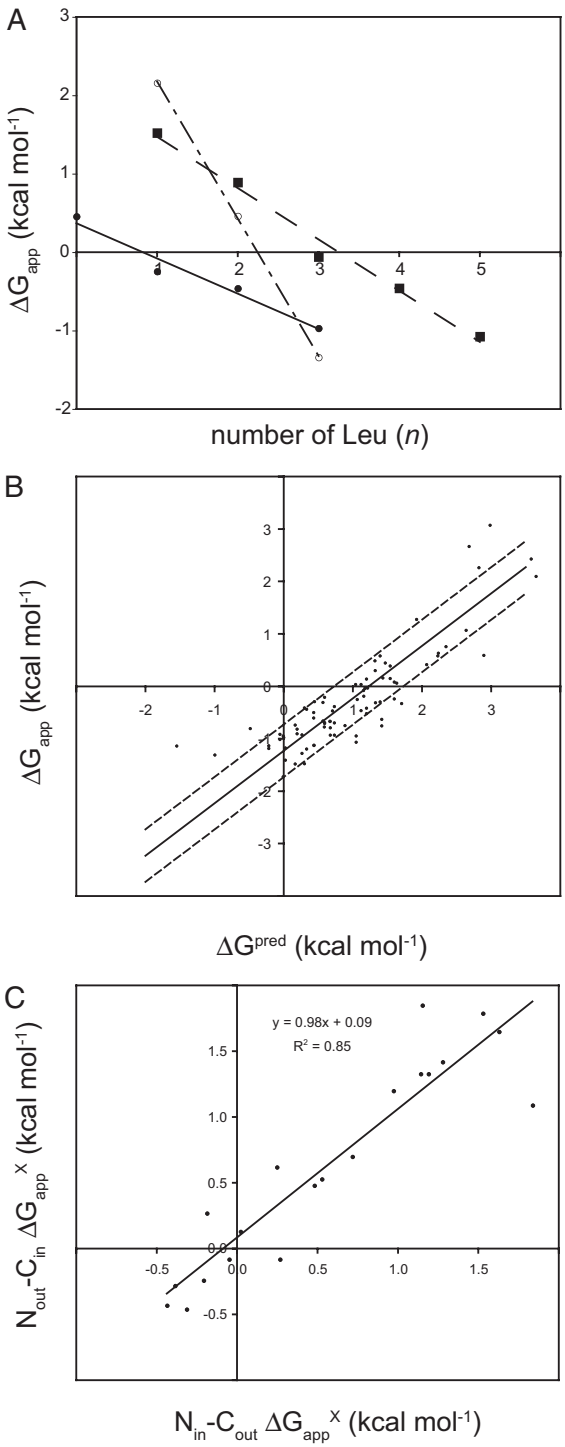

Fig. 3. Residue-specific contributions to ΔG_{app}. (*A*) ΔG_{app} values for N_{in}–C_{out} H-segments of the composition *nL*/(19-*n*)A measured by *in vitro* translation in the presence of RMs (n = 0–3; constructs 1–4, Table S1; solid line) and by expression in yeast (n = 1–3; constructs 87–89, Table S1; dot-dashed line). For comparison, ΔG_{app} values for the same H-segments in the N_{out}–C_{in} orientation (*cf.* Fig. 2*A*) are also shown (dashed line) (5). (*B*) Experimentally measured ΔG_{app} values plotted against predicted ΔG^{pred} values calculated as described in ref. 3 for all N_{in}–C_{out} H-segments analyzed with the RM system listed in Table S1. The solid line has the equation $\Delta G_{app} = \Delta G^{pred} - 1.23$ kcal/mol, where 1.23 kcal/mol is the mean of the differences between the ΔG_{app} and the corresponding ΔG^{pred} values in the plot. Dashed lines indicate an interval ±0.5 kcal/mol from the solid line. (*C*) Correlation between ΔG^{X}_{app} values for the 20 amino acids when placed in the middle of the H-segment determined for N_{in}–C_{out} H-segments (Table S2) and N_{out}–C_{in} H-segments (3).

placed each of the 20 naturally occurring amino acids in the middle of an H-segment of composition 1X/2L/16A or 1X/4L/14A when X was a polar or charged residue, or 1X/18A when X was a nonpolar residue. We further "scanned" selected residues (Asp, Lys, Pro, Trp) across a 1X/4L/14A or 1X/1L/16A (for Trp) H-segment. All constructs were translated *in vitro* in the presence of RMs, and ΔG_{app} values were determined. As seen in Table S1, charged (Asp, Lys, Arg, Glu) and highly polar (His, Asn, Gln) residues together with Pro strongly reduce insertion, weakly polar residues (Tyr, Met, Gly, Thr, Ser) have only a small reducing effect, nonpolar residues (Leu, Phe, Ile, Val, Trp) promote membrane insertion, and Ala and Cys are indifferent. It is noteworthy that Pro is better tolerated near the N terminus of the N_{in}–C_{out} H-segment than near the C terminus (Fig. S1) just as seen for N_{out}–C_{in} H-segments (5), strongly suggests that the adoption of an α-helical structure is important for efficient membrane insertion.

The contributions from the individual amino acids, ΔG^{X}_{app}, to the overall ΔG_{app} were estimated as described in Table S2. There is a good correlation between these values and the "optimized" ΔG^{X}_{app} values determined for N_{out}–C_{in} H-segments (3) (Fig. 3*C*).

Using data from an analysis of an extensive set of N_{out}–C_{in} H-segments, we derived an expression for predicting ΔG_{app} values for N_{out}–C_{in} H-segments from sequence (3). This expression reproduces the current results for N_{in}–C_{out} H-segments to within ±0.5 kcal/mol, i.e., equally well as it reproduces the original N_{out}–C_{in} H-segment data, but with an offset of 1.2 kcal/mol (Fig. 3*B* and Fig. S1). This is the same offset between N_{in}–C_{out} and N_{out}–C_{in} H-segments as found in Fig. 3*A*, i.e., the experimental ΔG_{app} values are 1.2 kcal/mol lower than the predicted ΔG^{pred} values. The different amino acids thus affect relative ΔG_{app} values to the same extent in N_{in}–C_{out} and N_{out}–C_{in} H-segments.

We also tested a few constructs in yeast (constructs 92–95; Table S1). The ΔG_{app} values for constructs with Asp, Lys, Pro, and Trp in or next to the middle position of the H-segment are quite well reproduced by ΔG^{pred}, again with an offset of ≈1 kcal/mol.

Targeting and Membrane Insertion of a Single N-terminal H-Segment. In all constructs discussed above, SRP-dependent targeting to the ER is ensured by the TM1 helix in Lep. Targeting may well have different requirements in terms of overall hydrophobicity than does gating of the Sec61 translocon and membrane insertion. To test this possibility, we designed a construct where TM1 and the loop between TM1 and the H-segment were deleted (Fig. 4*A*). The G1 site was retained in the N-terminal tail, and two sites (G2 and G3) were engineered into the P2 domain. Depending on its orientation in the membrane, the construct will be either singly (N_{out}–C_{in} orientation) or doubly (N_{in}–C_{out} orientation) glycosylated; noninserted molecules will remain unglycosylated. A similar approach has been used by Spiess and co-workers (13, 14) to study the membrane orientation of proteins with a single N-terminal transmembrane segment.

Fig. 4*B* shows that the H-segment inserts with increasing efficiency into RMs in the singly glycosylated N_{out}–C_{in} orientation for the 4L/15A–6L/13A constructs and then starts to switch toward the doubly glycosylated N_{in}–C_{out} orientation. Nearly 50% insertion (corresponding to ΔG_{app} = 0 kcal/mol) is seen for the 7L/12A H-segment (Fig. 4*C*). This threshold is much higher than is required for membrane insertion of N_{out}–C_{in} and N_{in}–C_{out} H-segments in the presence of TM1 (Fig. 3*A*). If targeting is considered as an equilibrium process, the increase in insertion efficiency between the 6L/13A and 7L/12A H-segments corresponds to a $\Delta\Delta G_{app} = -0.5$ kcal/mol for the Ala → Leu replacement.

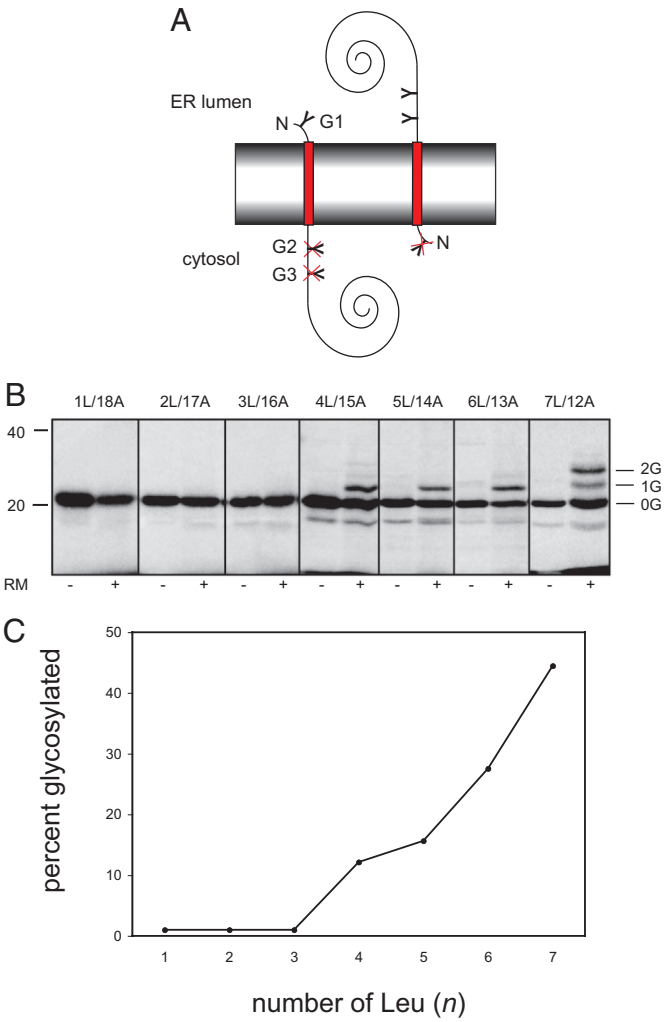

Fig. 4. Targeting and insertion of constructs with an N-terminal H-segment. (*A*) Model protein. If the H-segment forms an N_{out}–C_{in} transmembrane helix, only the G1 acceptor site will modified (*Left*); if it inserts in the opposite N_{in}–C_{out} orientation, the G2 and G3 sites will be modified instead (*Right*). If the protein fails to insert into the membrane, unglycosylated protein will be produced. (*B*) *In vitro* translation of the indicated constructs (96–102, Table S1) in the absence (−) or presence (+) of RMs. Unglycosylated (0G), singly glycosylated (1G), and doubly glycosylated (2G) products are indicated. (*C*) Percentage glycosylated product (1G + 2G) from the gel in *B*.

Discussion

TM helices in integral membrane proteins can be orientated in two ways relative to the membrane: N_{out}–C_{in} or N_{in}–C_{out}. The former kind has been studied extensively, both in the mammalian RM system and in *E. coli* (15, 16), and a detailed quantitative description of how amino acid sequence relates to membrane insertion efficiency of N_{out}–C_{in} TM helices has been achieved (3, 17).

Here, we report data on the sequence requirements for membrane insertion of N_{in}–C_{out} TM helices, by using both *in vitro* translation in the presence of RMs and expression in *S. cerevisiae*. We have mainly studied a model N_{in}–C_{out} TM helix (H-segment) that is placed downstream of an N-terminal N_{out}–C_{in} oriented TM helix (TM1) with an intervening 37-residue loop (Fig. 2). In this construct, SRP-dependent targeting to the Sec61 translocon is mediated by TM1 (7), and hence we are assaying the ability of the H-segment to reinitiate polypeptide translocation through the translocon channel and form a TM helix. Based on data from Kuroiwa *et al.* (6), we expect the reinitiation reaction to be independent of SRP. In contrast, N_{out}–C_{in} H-segments enter the translocon channel as part of a translocating nascent chain and hence trigger channel closure rather than opening (Fig. 1).

BIOCHEMISTRY

Given this mechanistic difference between how the translocon handles N_{out}–C_{in} and N_{in}–C_{out} H-segments, the results for the two kinds of TM helices are surprisingly similar. For our basic GGPG-[*n*L/(19-*n*)A]-GPGG H-segments, 50% membrane insertion (ΔG_{app} = 0 kcal/mol) is obtained for $n \approx 1$–2 for N_{in}–C_{out} H-segments (Fig. 3*A*), and for $n \approx 3$ for N_{out}–C_{in} H-segments (5). The change in insertion free energy ($\Delta\Delta G_{app}$) for an Ala → Leu replacement in an N_{in}–C_{out} H-segment is −0.5 kcal/mol (RM) or approximately −1.7 kcal/mol (yeast), compared with −0.7 kcal/mol (RM) and −1.0 kcal/mol (BHK cells) for N_{out}–C_{in} H-segments (5).

The different amino acids affect ΔG_{app} for N_{in}–C_{out} H-segments in much the same way that they do for N_{out}–C_{in} H-segments (Fig. 3*C*), and a derived expression for predicting ΔG_{app} values from sequence based on data from an extensive set of N_{out}–C_{in} H-segments (3) reproduces the current results for N_{in}–C_{out} H-segments to within ±0.5 kcal/mol but with an offset of 1.2 kcal/mol (Fig. 3*B*). The molecular origin of this offset clearly merits further study. In part, it may be explained by a slight but detectable dependence of ΔG_{app} on the identity of TM1 in the model protein because the replacement of the natural TM1 sequence in Lep by a 18L or 10L/9A stretch increases ΔG_{app} for a 1L/18A H-segment by ≈0.5 kcal/mol. Lengthening the loop between TM1 and the H-segment from 37 to 107 residues does not seem to affect ΔG_{app} to any significant degree.

Finally, we have analyzed constructs with a single, N-terminal GGPG-[*n*L/(19-*n*)A]-GPGG H-segment (Fig. 4). In this case, the threshold hydrophobicity required for targeting and membrane insertion is much higher ($n \approx 7$), presumably reflecting the sequence characteristics that promote efficient binding to SRP rather than membrane insertion *per se*. This compares well with a study of "idealized" Leu/Ala-based signal peptides in *E. coli* where a stretch of 7 leucines and 3 alanines was identified as the minimal functional SRP-binding signal peptide (18–20).

We conclude that the sequence requirements for membrane insertion of N_{out}–C_{in} and N_{in}–C_{out} H-segments are quite similar. In our standard Lep construct, they differ by the equivalent one of 1–2 Leu → Ala replacements in terms of the threshold hydrophobicity required for 50% insertion, and the relative contributions to the overall insertion efficiency provided by the different amino acids are nearly the same for the two orientations. This finding suggests that the mechanism of translocon-mediated recognition of N_{out}–C_{in} and N_{in}–C_{out} TM helices is basically the same, even though N_{in}–C_{out} TM helices must reinitiate translocation of the downstream part of the protein whereas N_{out}–C_{in} TM helices enter the translocon channel as part of a translocating nascent chain.

In contrast, when the TM helix also serves as an ER-targeting signal and must interact productively with SRP before encountering the translocon, the threshold hydrophobicity is markedly higher. This implies on the one hand that many TM helices in multispanning membrane proteins would not by themselves be able to trigger SRP-mediated targeting to the ER translocon, and on the other that segments that are not sufficiently hydrophobic to bind SRP but that could form TM helices if introduced into an ER-targeted protein may exist in soluble cytosolic proteins (3).

Materials and Methods

Enzymes and Chemicals. Unless otherwise stated, all enzymes, plasmid pGEM1, and the TNT Quick-coupled transcription/translation system were from Promega. [^{35}S]Met, ^{14}C-methylated marker proteins, and deoxynucleotides were from GE Healthcare. The BigDye Terminator v1.1 cycle sequencing kit was from Applied Biosystems, and oligonucleotides were from CyberGene AB. The signal peptidase inhibitor *N*-methoxysuccinyl-Ala-Ala-Pro-Val-chloromethyl ketone was from Sigma–Aldrich.

DNA Manipulations. For expression of Lep constructs from the pGEM1 plasmid, the 5′ end of the *lepB* gene from *E. coli* was modified by the introduction of an XbaI site and by changing the context 5′ to the initiator ATG codon to a Kozak consensus sequence (21). Lep constructs used to analyze N_{in}–C_{out} oriented H-segments (Fig. 2*A Right*) carried one acceptor site for N-linked glycosylation in positions 3–5 (Asn-Ser-Thr; G1) included in an extended sequence of 24 residues (Met-Ala-Asn^3-Ser-Thr-Ser-Gln-Gly-Ser-Gln-Pro-Ile-Asn-Ala-Gln-Ala-Ala-Pro-Val-Ala-Gln-Gly-Gly-Ser-Gln-Gly-Glu-Phe^5) inserted between Asn^3 and Phe^5 in the wild-type sequence of Lep, and a second acceptor site in positions 97–99 (Asn-Ser-Thr; G2). Oligonucleotides encoding the different H-segments were introduced between a SpeI site in codons 60–61 and a KpnI site in codon 80 in the *lepB* gene (22). Both the G1 and G2 sites were placed >15 residues away from the nearest transmembrane segment (23, 24). Oligonucleotides encoding the 18L and 10L/9A TM1 segments (constructs 83–84) were introduced between an *MfeI* site in codons 26–27 and an AvrII site in codons 46–47 in the extended N-terminal tail.

Lep constructs used to analyze N-terminal H-segments (Fig. 4*A*) carried one acceptor site for N-linked glycosylation in positions 3–5 (Asn-Ser-Thr; G1) included in an extended sequence of 12 residues (Met-Ala-Asn^3-Ser-Thr-Ser-Gln-Gly-Ser-Gln-Pro-Ile-Asn-Ala-Gln) between Asn^3 and Phe^5 in the wild-type sequence of Lep. Two additional glycosylation acceptor sites were located in positions 110–112 (Asn-Ser-Thr; G2) and 136–138 (Asn-Ser-Thr; G3). Finally, residues Ala^6–Gly^{72} were removed. Oligonucleotides encoding the different H-segments were introduced between an SpeI site in codons 72–73 and a KpnI site in codon 98 in the *lepB* gene. The G1, G2, and G3 sites were all placed >15 residues away from the H-segment.

The construct lacking the N-terminal part of Lep up until the C-terminal end of the H-segment just before the GPGG flank (Fig. 2*B*, lane 6) was made by using a PCR-amplified DNA fragment from the pGEM1 plasmid as transcription template and the T7 polymerase. The 5′ PCR primer had the sequence 5′-GGATCCTAATACGACTCACTATAGGGAGCCACCATGGGACCTGGTGGGGT-ACCG-3′ containing the T7 promoter, a ribosome-binding site, and the initiator codon (underlined). The reverse primer was complementary to the 3′ end of the *lepB* gene and had the sequence 5′-GATGGCTATTAATGGATGCCGC-CAATGCG.

The loop extension in construct 85 was introduced between an engineered AvrII site in codons 46–47 and an SpeI site in codons 60–61. The full loop sequence is V^{22}-PSLLFGQPVGISAVHQPVASAPVPAKGLQCRGCSKLSAPIRFRAVDARHV-HVGAGILFFIMNYLRLGVGRQGPEEGRRVGPVGVQRGGQLHQRLHDALQVLQL-LAPS-W^{60}, where the numbered residues refer to the wild-type Lep sequence.

H-segments and TM1 segments were constructed by using two or three double-stranded oligonucleotides (18–48 nt long) with overlapping overhangs at the ends. Pairs of complementary oligonucleotides were first incubated at 85°C for 10 min followed by slow cooling to 30°C, after which the two or three annealed double-stranded oligonucleotides 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.

Expression *in Vitro* and Quantification of Membrane Insertion Efficiency. Constructs cloned in pGEM1 were transcribed and translated in the TNT Quick-coupled transcription/translation system. One microgram of DNA template, 1 μl of [^{35}S]Met (5 μCi), and 1 μl of dog pancreas RMs were added at the start of the reaction, and samples were incubated for 90 min at 30°C. When relevant, signal peptidase inhibitor dissolved in DMSO was included in the translation mix at a final concentration of 1.4 mM (8). For endo H treatment, 5 μl of TNT lysate mix was resuspended in sample buffer and mixed with 10 μl of dH_2O, 2.4 μl of buffer (800 mM sodium acetate, pH 5.7), and 1.5 μl of endo H (5 units/ml; Roche) or dH_2O (mock sample), and the sample was incubated at 37°C for 2 h. To stop the reaction, the sample was incubated at 90°C for 2 min before loading on a 10% SDS/polyacrylamide gel.

The N-terminally truncated construct in Fig. 2*B*, lane 6, was transcribed by using TNT T7 Quick for PCR DNA and 5 μl of purified PCR product.

Translation products were analyzed by SDS/PAGE, and gels were quantified on a Fuji FLA-3000 PhosphorImager with the Image Reader 1.8J/Image Gauge V 4.22 software. The degree of membrane integration of each H-segment was quantified from SDS/polyacrylamide gels by calculating an apparent equilibrium constant between the membrane-integrated and nonintegrated forms:

$$K_{app} = \frac{f_{2x} + f_c}{f_{1x}},$$

where f_{1x}, f_{2x}, and f_c denote the fractions of singly glycosylated, doubly glycosylated, and cleaved singly glycosylated molecules, respectively (unglycosylated molecules that have not been targeted to the RMs or the yeast ER are ignored, hence $f_{1x} + f_{2x} + f_c = 1$). For constructs expressed in yeast, the fraction of cleaved unglycosylated molecules (arrow in Fig. 2*C*) was included in f_{1x}. The results were then converted to apparent free energies, $\Delta G_{app} = -RT \ln K_{app}$. All ΔG_{app} values were measured as mean values from at least three independent

experiments; for H-segments with $\Delta G_{app} \in [-1, +1]$ kcal/mol, the precision in the ΔG_{app} values is approximately ± 0.2 kcal/mol (5).

ΔG^{X}_{app} values for the individual amino acids were calculated as described in Table S2.

Theoretical ΔG^{pred} values were calculated as described in ref. 3 by using the ΔG prediction server (v1.0) at www.cbr.su.se/DGpred/.

Expression in *S. cerevisiae*. All yeast plasmids were constructed from p424GPD (25) by homologous recombination (26). For subcloning of a C-terminal triple hemagglutinin (HA) tag into p424GPD, the HA sequence was PCR-amplified by using pJK90 (27) as a template and two primers 5′-TCTAGAACTAGTGGATCCCCCGGGCTGCAGCCATCTTACCCATACGATG3-′ and 5′-CTCGAGGTCGACGGTATCGATAAGCTTGATATCCTAATTACATGACTCGAG3-′ (underlined sequences are the sequences complementing upstream and downstream sequences of the EcoRI site in p424GPD for homologous recombination). W303-1a (*MAT a*, *ade2*, *can1*, *his3*, *leu2*, *trp1*, *ura3*) was transformed with the 3×HA PCR fragment and EcoRI-linearized p424GPD. Yeast transformants were selected on −Trp plates. Positive clones were selected by yeast colony PCR, purified, and confirmed by DNA sequencing. The 3×HA-containing plasmid was named p424GPDHA. For subcloning of *E. coli lepB* constructs from the pGEM1 vectors described above into p424GPDHA, the *lepB* gene (including the relevant H-segment) was PCR-amplified from pGEM1 by using two primers, 5′-GTTTCGACGGATTCTAGAACTAGTGGATCCATGGCGAATATGTTTGCC3-′ and 5′-AGGAACATCGTATGGGTAAGATGGCTGCAG-ATGGATGCCGCCAAT 3′ (underlined sequences are the sequences complementing upstream and downstream sequences of the SmaI site in p424GPDHA for homologous recombination). W303-1a (*MAT a*, *ade2*, *can1*, *his3*, *leu2*, *trp1*, *ura3*) was transformed with *lepB* PCR fragment and SmaI-linearized p424GPDHA. Positive clones were selected by either yeast colony PCR or Western blotting with an anti-HA antibody. Plasmids were isolated from yeast transformants, and DNA sequencing confirmed that Lep was fused in-frame to HA. Yeast transformants expressing Lep constructs were grown overnight in 5 ml of −Trp medium at 30°C. Preparation of whole-cell lysates was done as described in ref. 28. Whole-cell lysates were subjected to SDS/PAGE and Western blotting with an HA antiserum. Endo H (Roche) treatment was carried out as described in ref. 29. Western blots were quantified by using the Image Reader 8.1j software.

Expression of a Lep construct in the signal peptidase mutant strain was carried out by [^{35}S]Met radiolabeling. Construct 89 (Table S1) was transformed into strain HFY406 (*MATa*, *spc3-4*, *ura3-52*, *leu2-3*, *112 his3-delta200*, *trp1-delta901*, *suc2-delta9*, *lys2-80*) (12) and selected on a −Trp plate. Transformants were grown to an A_{600} of 0.4, and 1.5 A_{600} units of cells were harvested, washed twice with −Met medium not containing ammonium sulfate, and starved at 25°C for 10 min. Cells were harvested and resuspended in 150 μl of −Met medium not containing ammonium sulfate. One sample was preincubated at the permissive temperature (25°C) and the other at the nonpermissive temperature (37°C) for 15 min before labeling with [^{35}S]Met (50 μCi/A_{600} unit of cells) for 5 min at 25°C and 37°C, respectively. Lep protein was immunoprecipitated with an anti-HA antiserum and subjected to SDS/PAGE.

ACKNOWLEDGMENTS. We thank Prof. Dr. M. Sakaguchi, Hyogo University, for providing rough microsomes. This work was supported by grants from Magnus Bergvalls Stiftelse, Henrik Granholms Stiftelse, and Carl Tryggers Stiftelse (to I.N.), from the Swedish Cancer Foundation, the Swedish Research Council, and the Swedish Foundation for Strategic Research (to G.v.H. and I.N.), and from the U.S. National Institute of General Medical Sciences (to S.H.W.).

1. Rapoport TA (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450:663–669.
2. von Heijne G (2006) Membrane-protein topology. *Nature Rev Mol Cell Biol* 7:909–918.
3. Hessa T, *et al*. (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450:1026–1030.
4. Meindl-Beinker NM, Lundin C, Nilsson I, White SH, von Heijne G (2006) Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *EMBO Rep* 7:1111–1116.
5. Hessa T, *et al*. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433:377–381.
6. Kuroiwa T, Sakaguchi M, Omura T, Mihara K (1996) Reinitiation of protein translocation across the endoplasmic reticulum membrane for the topogenesis of multispanning membrane proteins. *J Biol Chem* 271:6423–6428.
7. Heinrich S, Mothes W, Brunner J, Rapoport T (2000) The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* 102:233–244.
8. Nilsson I, Johnson AE, von Heijne G (2003) How hydrophobic is alanine? *J Biol Chem* 278:29389–29393.
9. Paetzel M, Dalbey RE, Strynadka NCJ (1998) Crystal structure of a bacterial signal peptidase in complex with a β-lactam inhibitor. *Nature* 396:186–190.
10. Liljeström P, Garoff H (1991) Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. *J Virol* 65:147–154.
11. Nilsson IM, von Heijne G (1991) A *de novo* designed signal peptide cleavage cassette functions *in vivo*. *J Biol Chem* 266:3408–3410.
12. Fang H, Mullins C, Green N (1997) In addition to *SEC11*, a newly identified gene, *SPC3*, is essential for signal peptidase activity in the yeast endoplasmic reticulum. *J Biol Chem* 272:13152–13158.
13. Wahlberg JM, Spiess M (1997) Multiple determinants direct the orientation of signal-anchor proteins: The topogenic role of the hydrophobic signal domain. *J Cell Biol* 137:555–562.
14. Higy M, Junne T, Spiess M (2004) Topogenesis of membrane proteins at the endoplasmic reticulum. *Biochemistry* 43:12716–12722.
15. Kuroiwa T, Sakaguchi M, Mihara K, Omura T (1991) Systematic analysis of stop-transfer sequence for microsomal membrane. *J Biol Chem* 266:9251–9255.
16. Chen HF, Kendall DA (1995) Artificial transmembrane segments: Requirements for stop transfer and polypeptide orientation. *J Biol Chem* 270:14115–14122.
17. Bernsel A, *et al*. (2008) Prediction of membrane-protein topology from first principles. *Proc Natl Acad Sci USA* 105:7177–7181.
18. Doud SK, Chou MM, Kendall DA (1993) Titration of protein transport activity by incremental changes in signal peptide hydrophobicity. *Biochemistry* 32:1251–1256.
19. Valent QA, *et al*. (1995) Early events in preprotein recognition in *E. coli*: Interaction of SRP and trigger factor with nascent polypeptides. *EMBO J* 14:5494–5505.
20. Valent QA, *et al*. (1997) Association of nascent membrane and presecretory proteins with signal recognition particle and trigger factor. *Mol Microbiol* 25:53–64.
21. Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene* 234:187–208.
22. Sääf A, Wallin E, von Heijne G (1998) Stop-transfer function of pseudo-random amino acid segments during translocation across prokaryotic and eukaryotic membranes. *Eur J Biochem* 251:821–829.
23. Nilsson I, von Heijne G (1993) Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J Biol Chem* 268:5798–5801.
24. Popov M, Tam LY, Li J, Reithmeier RAF (1997) Mapping the ends of transmembrane segments in a polytopic membrane protein: Scanning N-glycosylation mutagenesis of extracytosolic loops in the anion exchanger, band 3. *J Biol Chem* 272:18325–18332.
25. Mumberg D, Müller R, Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156:119–122.
26. Oldenburg KR, Vo KT, Michaelis S, Paddon C (1997) Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. *Nucleic Acids Res* 25:451–452.
27. Kim H, Park H, Montalvo L, Lennarz WJ (2000) Studies on the role of the hydrophobic domain of Ost4p in interactions with other subunits of yeast oligosaccharyl transferase. *Proc Natl Acad Sci USA* 97:1516–1520.
28. Österberg M, *et al*. (2006) Phenotypic effects of membrane protein overexpression in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 103:11148–11153.
29. Kim H, Österberg M, Melén K, von Heijne G (2006) A global topology map of the *Saccharomyces cerevisiae* membrane proteome. *Proc Natl Acad Sci USA* 103:11142–11147.