Upstream charged and hydrophobic residues impact the timing of membrane insertion of transmembrane helices

Felix Nicolaus¹, Fatima Ibrahimi¹, Anne den Besten¹, Gunnar von Heijne^{*1,2}

¹Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden.

²Science for Life Laboratory Stockholm University, Box 1031, SE-171 21 Solna, Sweden

*Correspondence to: Gunnar von Heijne, gunnar@dbb.su.se.

Keywords: membrane protein biogenesis, transmembrane helix, cotranslational, BtuC

Running head: Cotranslational insertion of transmembrane helices

Abstract

During SecYEG-mediated cotranslational insertion of membrane proteins, transmembrane helices (TMHs) first make contact with the membrane when their N-terminal end is ~45 residues away from the peptidyl transferase center. However, we recently uncovered instances where the first contact is delayed by up to ~10 residues. Here, we recapitulate these effects using a model TMH fused to two short segments from the BtuC protein: a positively charged loop and a re-entrant loop. We show that the critical residues are two Arg residues in the positively charged loop and four hydrophobic residues in the re-entrant loop. Thus, both electrostatic and hydrophobic interactions involving sequence elements that are not part of a TMH can impact the way the latter behaves during membrane insertion.

Introduction

Nearly all helix-bundle type membrane proteins in both prokaryotic and eukaryotic cells are inserted cotranslationally into their target membrane [1,2]. While the energetics of membrane insertion of transmembrane helices (TMHs) has been extensively studied both *in vitro* and *in vivo* [3-6], it is only recently that methodologies have been developed that make it possible to follow step-by-step the membrane insertion of individual TMHs as they come out of the ribosome exit tunnel. Two methods in particular can provide this kind of information: *in vitro* real-time FRET analysis of cotranslational membrane insertion [7] and *in vivo* Force Profile Analysis (FPA) where a translational arrest peptide (AP) engineered into the polypeptide chain is used to detect the force generated on the nascent chain during membrane insertion [8,9].

In a recent study, we applied FPA to three *E. coli* inner membrane proteins: EmrE, GlpG, and BtuC [9]. For all three proteins (a total of 20 TMHs), we found that, on average, a TMH starts to generate force on the nascent chain when its N-terminal end is ~45 residues away from the polypeptide transferase center (PTC) in the large ribosome subunit. A real-time FRET study of the N-terminal TMH in another *E. coli* inner membrane protein (LepB) has shown that its N-terminal end reaches the vicinity of the entry to the SecYEG translocon channel when 40-50 residues away from the PTC [7], confirming that the increase in force is caused by the encounter of the TMH with the hydrophobic translocon/lipid environment. However, FPA further revealed that not all TMHs follow the ‘45-residue rule’, and that the fine details of the force profiles (FPs) reflect context-dependent effects that promise to provide further insights into the molecular interactions that drive and control membrane insertion of TMHs [9].

Here, we focus on two such unexpected features in the FP recorded for the 10-TMH BtuC protein [9]: (i) TMH2 starts to generate force only when its N-terminal end is ~55 rather than

the expected ~45 residues away from the PTC, and (ii) two narrow peaks appear in the force profile flanking the location where a single peak generated by TMH6 was expected. In particular, we ask whether these effects are peculiar to BtuC, or whether they are also seen when the segments that flank BtuC TMH2 and TMH6 are fused to a model TMH composed of Leu and Ala residues. Indeed, we find that the effects are seen also with model TMHs, and further pin-point the implicated sequence determinants to two Arg residues located in the periplasmic loop upstream of BtuC TMH2 and a four-residue hydrophobic stretch located in a re-entrant loop upstream of BtuC TMH6. We hypothesize that the positively charged Arg residues are temporarily prevented from translocating through the ‘hydrophobic gasket’ [10] in the middle of the SecYEG channel by the membrane potential and thereby delay membrane insertion of the following TMH, and that the partitioning of the hydrophobic stretch into the periplasmic leaflet of the inner membrane somehow weakens the initial interaction between the following TMH and the lipid bilayer.

Materials and methods

Enzymes and chemicals

Enzymes and other reagents were purchased from Thermo Fisher Scientific, New England Biolabs, and Sigma-Aldrich. Oligonucleotides were ordered from Eurofins Genomics. L-[³⁵S]-methionine was provided by PerkinElmer.

Cloning and Mutagenesis

In order to generate the constructs used in this study the previously described pING1 plasmid harboring a shortened LepB sequence from *E. coli* with an inserted hydrophobic H-segment (4L/15A or 6L/13A) followed by a variable LepB-derived linker (between 4 and 47 residues long), the 17-residue long SecM(*Ec*) AP from *E. coli*, and a C-terminal extension containing

23 residues derived from *E. coli* LepB was employed [8]. All sequences are summarized in Supplementary Fig. S1. The DNA sequences encoding the flanking regions (RGELFVWQIRLPR; LMYWMMGGFGGVDWRQS) derived from *E. coli* BtuC (periplasmic loop connecting TMH1 and TMH2; re-entrant loop connecting TMH5 and TMH6, respectively) were amplified by PCR from a full-length LepB-BtuC-SecM(*Ec*) construct used in a previous study [9] and cloned upstream of 4L/15A or 6L/13A using Gibson assembly® [11]. Site-specific DNA mutagenesis was performed to change Arg residues to Gln residues (mutants QRR, RRQ, QRQ, QQQ) in RGELFVWQIRLPR, and to replace *n* residues in LMYWMMGGFGGVDWRQS with Gly-Ser (GS) repeats (mutants *n* = 5, *n* = 9, *n* = 13, *n* = 17, and LMYW→GSGS). All cloning and mutagenesis products were confirmed by DNA sequencing.

In vivo pulse-labeling analysis

E. coli MC1061 cells [12] transformed with the pING1 vector harboring the respective gene of interest were grown overnight at 37°C in M9 minimal medium supplemented with 19 amino acids (1 µg/ml) lacking Met, 100 µg/ml thiamine, 0.4% (w/v) fructose, 100 mg/ml ampicillin, 2 mM MgSO₄, and 0.1mM CaCl₂. Overnight cultures were diluted into fresh M9 medium to an OD₆₀₀ of 0.1 and cell growth was continued until an OD₆₀₀ of 0.3 was reached. Induction of protein expression was started by adding arabinose to a final concentration of 0.2% (w/v) and continued for 5 min at 37°C. After radiolabeling with [³⁵S]-methionine for 2 min at 37°C the reaction was stopped by adding ice-cold trichloroacetic acid (TCA) to a final concentration of 10%. Proteins were precipitated on ice for 30 min and were spun down for 10 min at 20,000 g at 4°C. Samples were washed with ice-cold acetone, centrifugation was repeated and pellets were solubilized in Tris-SDS buffer (10 mM Tris-Cl pH 7.5, 2% (w/v) SDS) for 5 min at 1,400 rpm at 37°C. After another centrifugation step for 5 min at 20,000 g, the supernatant was incubated in TSET buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1

mM EDTA-KOH, 2% (v/v) triton X-100) supplemented with Pansorbin® for 15 min on ice. Proteins non-specifically bound to Pansorbin® were removed by centrifugation at 20,000 g. The proteins of interest in the supernatant were further used for immunoprecipitation with Pansorbin® and LepB antisera (in-house rabbit polyclonal IgG), which was carried out for at least 1 hour at 4°C whilst rolling. Immunoprecipitates were spun down and washed with 10 mM Tris-Cl pH 7.5, 150 mM NaCl, 2 mM EDTA, and 0.2% (v/v) triton X-100 and subsequently with 10 mM Tris-Cl pH 7.5. Pellets were solubilized in SDS sample buffer (67 mM Tris, 33% (w/v) SDS, 0.012% (w/v) bromophenol blue, 10 mM EDTA-KOH pH 8.0, 6.75% (v/v) glycerol, 100 mM DTT) for 10 min while shaking at 1,400 rpm. Solubilized proteins were incubated with 0.25 mg/ml RNase for 30 min at 37°C to hydrolyze tRNA and subsequently separated by SDS-PAGE. Gel fixation was done in 30% (v/v) methanol and 10% (v/v) acetic acid. After incubation in Gel-Dry™ Drying Solution for 30 min, gels were dried in a Bio-Rad gel dryer model 583. By exposing dried gels to phosphorimaging plates, radiolabeled proteins were detected using a Fuji FLA-3000 scanner. Band intensity profiles were obtained using the Fiji (ImageJ) software and quantified with the in-house software EasyQuant in order to determine the fraction full-length protein $f_{FL} = I_{FL}/(I_{FL} + I_A)$, where I_A , I_{FL} are the intensities of the A and FL bands, respectively. A_c and/or FL_c size controls were included in the SDS-PAGE to confirm the identity of the A and FL bands. Each data point in the force profile represents the average of three independent biological replicates and includes the standard error of the mean (SEM).

Results

Force Profile Analysis

FPA takes advantage of the fact that a translational arrest peptide (AP) can induce a temporary pause in translation by binding in the upper parts of the ribosome exit tunnel when its last codon is positioned in the ribosomal A-site [13]. The AP-induced stalling can be overcome by pulling forces acting on the nascent chain [14,15], with different APs being sensitive to different force levels [16]. Such pulling forces can be generated by, *e.g.*, cotranslational protein folding in the cytoplasm as well as in the periplasm, or cotranslational insertion of transmembrane segments into the membrane [9,17,18]. Therefore, APs can be conveniently used as force sensors to study cotranslational events *in vitro* and *in vivo*.

Here, we have used a series of constructs harboring a force-generating model TMH (H-segment) of composition $nL/(19-n)A$ ($n = 4, 6$) placed L residues away from the C-terminal end of the 17-residue *E. coli* SecM (SecM(*Ec*)) AP followed by a 23-residue C-terminal tail [8], Figure 1a. The backbone of the design is from *E. coli* LepB, an inner membrane protein with two N-terminal TMHs and a periplasmic C-terminal domain [19]. The model 19-residue TMH of composition 4L/15A or 6L/13A (H-segment) followed by the SecM(*Ec*) AP is fused to the C terminus of the LepB part. A flanking region (FR), composed either of the positively charged loop between BtuC TMH1 and TMH2 (BtuC residues 47-59; Fig. 2a) or of the re-entrant loop between BtuC TMH5 and TMH6 (BtuC residues 175-191; Fig 3a), was inserted immediately upstream of the H-segment (the amino acid sequences of all constructs are shown in Supplementary Fig. S1).

By varying L , the model TMH will be positioned in different locations in the ribosome exit tunnel or SecYEG translocon channel at the point when the ribosome reaches the last codon of the AP, and will therefore generate pulling forces of different magnitude. In a construct

where membrane integration of the H-segment happens at the exact same time as the ribosome reaches the last codon of the AP, a strong pulling force F will be exerted on the nascent chain, which in turn prevents the translational pausing at the AP and results in mostly full-length protein (including the C-terminal tail) being produced during a short pulse with [^{35}S]-Met, Figure 1b (middle). In contrast, in constructs where the H-segment has not yet reached the translocon, or has already been inserted into the membrane at the point when the ribosome reaches the end of the AP, little force is exerted on the AP, translational pausing will be efficient, and only the arrested form of the protein will be produced during a short pulse with [^{35}S]-Met, Figure 1b (left and right). After immunoprecipitation with LepB antisera, full-length (FL) and arrested (A) species are separated on an SDS-PAGE gel and the fraction full-length protein is calculated, $f_{FL} = I_{FL}/(I_{FL}+I_A)$, where I_{FL} and I_A are the intensities of the bands representing FL and A species, respectively (Figure 1c). Because it is sensitive to pulling force, f_{FL} can be used as a proxy for F [15,20-22]. A force profile (FP), in which f_{FL} is plotted against L then can be used to follow, step-by-step, the membrane integration of model or natural TMHs; FPs for the 4L/15A and 6L/13A H-segments as well as for BtuC TMH2 and TMH6 are shown in Fig. 1d (see Supplementary Fig. S1 for amino acid sequences of all constructs). For a given peak in the FP, we define L_{start} as the L value for which f_{FL} reaches its half-maximal value ($L_{start} \approx 25$ residues for the 4L/15A and 6L/13A H-segments, $L_{start} \approx 22$ residues for BtuC TMH2 and TMH6, Fig. 1d)

According to the ΔG predictor [4], BtuC TMH2 and TMH6 are predicted to form TMHs with ΔG values of -0.13 and -0.79 kcal/mol, respectively, and both generate strong pulling force from $L=24$ to $L=52$ [9], Fig 1d. The model H-segments 4L/15A and 6L/13A [8] were chosen based on similar ΔG values of -0.49 and -1.54 kcal/mol, respectively. As expected, the model and natural TMHs have very similar FPs, with slightly earlier L_{start} values for the latter. An

FP was also obtained for the 6L/13A H-segment using the stronger AP variant SecM(*Ec*-Sup1) [23], showing that the maximal force is exerted at $L_{max} = 37$ residues, Fig 1d.

Upstream Arg residues delay membrane insertion of a model TMH

In our recent FPA study of the cotranslational membrane insertion of BtuC, we observed that TMH2 initiated membrane insertion when its N-terminal end is ~55 rather than the customary ~45 residues away from the PTC [9], and provisionally attributed this unexpected finding to the presence of the highly positively charged periplasmic loop between TMH1 and TMH2, Fig. 2a. According to the ‘positive-inside rule’ [24], short periplasmic loops between TMHs typically contain only 0-2 positively charged residues, and the TMH1-TMH2 loop that contains 3 Arg residues is thus an unusual feature in bacterial inner membrane proteins.

In order to simplify the experimental system and focus on the possible effects on membrane insertion of the positively charged loop, we used the construct design shown in Fig. 1a. To obtain the FP for the membrane insertion of the H-segment, the length L of the linker between the H-segment and the AP was increased stepwise from $L = 21$ to $L = 64$ residues, and f_{FL} was measured for each construct. As shown in Fig. 2b-c, in the absence of the positively charged loop, the 4L/15A and 6L/13A TMHs exert half-maximal pulling force at $L_{start} \approx 25$ residues, *i.e.*, when their N-terminal ends are $25+19 = 44$ residues from the PTC, as expected. When preceded by the positively charged loop, however, half-maximal pulling force is reached only at $L_{start} \approx 41$ residues for 4L/15A and $L \approx 35$ residues for 6L/13A. Indeed, f_{FL} remains at its background value until $L \approx 30$ for both H-segments in the presence of the positively charged loop. Using the much stronger SecM(*Ec*-Sup1) AP [23], we see that not only is membrane insertion of the 6L/13A H-segment delayed by the positively charged loop, but the maximal pulling force exerted during the insertion process is also reduced, Fig. 2d (orange curve).

To better appreciate the role of the three different Arg residues, we made four R→Q mutants in the 4L/15A[L=30] construct (green dots in Fig.2b). The R⁴⁷→Q mutation (QRR) had almost no effect on f_{FL} , while the R⁵⁹→Q mutation, either alone (RRQ) or together with R⁴⁷→Q (QRQ), gave rise to a substantial increase in f_{FL} . A very strong effect on f_{FL} was seen when all three Arg residues, including R⁵⁶, were mutated (QQQ). Thus, the main culprits in the positively charged loop are R⁵⁶ and R⁵⁹, *i.e.*, the two Arg residue located closest to the H-segment. We conclude that positively charged residues located close to the N terminus of a N_{out}-oriented TMH can substantially delay its insertion into the membrane.

A short upstream hydrophobic segment delays membrane insertion of a model TMH

In our recent study of the membrane insertion of BtuC [9], we also found that the insertion of TMH6 gave rise to two peaks in the FP, one that appeared when the N-terminal end of TMH6 was only 37 residues away from the PTC and one when it was 63 residues away, and we provisionally attributed these unexpected observations to the presence of an upstream re-entrant loop located between TMH5 and TMH6 (Fig. 3a) and an amphipathic surface helix directly downstream of TMH6. As seen in Fig. 1d, when TMH6 (BtuC residues 175-191) is analyzed in the context of the LepB construct (Fig. 1a), half-maximal pulling force is seen at $L_{start} \approx 22$ residues, *i.e.*, when the N-terminal end of TMH6 is 41 residues from the PTC, similar to the results for the 4L15A and 6L/13A H-segments.

To directly probe the role of the re-entrant loop (BtuC residues 175-191) we inserted it upstream of the 4L/15A and 6L/13A H-segments ('FR' in Fig 1a) and recorded the respective FPs, Fig. 3b-c. For the 4L/15A series of constructs, $L_{start} \approx 38$ residues, *i.e.*, the re-entrant loop delays the insertion of the H-segment by ~13 residues. A smaller delay of ~3 residues is seen with the more hydrophobic 6L/13A H-segment, Fig 3c. The effect of the re-entrant loop on the insertion of TMH6 seen in the intact BtuC protein can thus be reproduced with the model H-segments. As for the positively charged loop, an FP obtained with the stronger

SecM(*Ec*-Sup1) AP shows both slightly delayed membrane insertion of the 6L/13A H-segment and a reduction in the maximal pulling force exerted during the insertion process in the presence of the re-entrant loop, Fig. 2d (red curve).

To pin-point the critical residues in the re-entrant loop, we replaced an increasing number of loop residues in the 4L/15A[L=30] construct, first five and then four at a time, by GS-repeats, starting from the C-terminal end of the loop, Fig. 3d. Thirteen loop residues could be replaced by GS repeats with no effect on f_{FL} ; only when the entire re-entrant loop segment was replaced by GS repeats did f_{FL} return to the high value seen for the 4L/15A H-segment itself. The four most N-terminal residues in the re-entrant loop (LMYW) thus seem to be the main determinants of the delay in membrane insertion of the H-segment and, indeed, when these four residues were mutated to GSGS while keeping the other residues in the re-entrant loop there was a clear increase in f_{FL} compared to re-entrant loop construct (+RE loop), albeit not as strong as for the $n = 17$ mutant. The re-entrant loop, and especially the short hydrophobic helix at its N-terminal end, thus causes a marked reduction of the f_{FL} values in the early part of the FP not only for TMH6 in BtuC but also for the model 4L/15A and 6L/13A H-segments in the LepB constructs. We conclude that a short hydrophobic stretch located ~15 residues upstream of the N terminus of a N_{out} -oriented TMH can substantially modify the behavior of the TMH during the membrane insertion process.

Discussion

The 'sliding' models for SecYEG-mediated cotranslational insertion of TMHs into the *E. coli* inner membrane posits that the hydrophobic segment makes contact with membrane lipids through a more or less open lateral gate as soon as it enters the SecYEG channel [25]. The current best estimate is that this happens when the N-terminal end of the hydrophobic

segment is ~45 residues away from the PTC [7,9]. However, our recent study of the cotranslational membrane insertion of three multi-spanning *E. coli* inner membrane proteins [9] uncovered a couple of instances where membrane insertion of a particular TMH appeared to be delayed to markedly longer nascent chain lengths. In order to better understand these exceptional cases, we have tried to reproduce them in a simplified system where short loop segments presumed to be responsible for the delay are inserted upstream of model TMHs (H-segments) composed only of Leu and Ala residues.

Our results show that positively charged Arg residues located close to the N terminus of an N_{out} -oriented H-segment appear to completely prevent the latter from making hydrophobic interactions during an early stage of passage into the SecYEG channel, as seen in the FPs in Fig. 2b-c ($L = 20$ -30 residues). A rough estimate suggests that this early stage corresponds to a situation where the Arg residues are near the hydrophobic gasket [10] that seals off the cytoplasmic from the periplasmic parts of the channel, Fig. 4a. Once a sufficiently long stretch of the H-segment has entered the channel ($L > 30$ residues) there is an increase in f_{FL} , indicating that the H-segment now engages in hydrophobic interactions and starts to insert into the membrane. Even in these later stages, however, the positively charged residues cause a persistent reduction in the f_{FL} values. We propose that the electrical membrane potential pushes back on the positively charged residues and prevents them from translocating across the hydrophobic gasket (in turn preventing the H-segment from entering the SecYEG channel) until a sufficiently long segment of the H-segment has emerged from the ribosome exit tunnel such that the free energy gained upon its membrane insertion compensates for the energetic cost of moving the Arg residues across the gasket. The reduction in pulling force caused by positively charged residues is a mirror image of the increase in pulling force seen when negatively charged residues translocate across the SecYEG channel [21].

In contrast to the effect on insertion caused by the positively charged residues, the re-entrant loop does not reduce f_{FL} to background values in the early part of the FP, but only attenuates the f_{FL} values in this region, Fig. 3b-c. A rough model of the nascent chain at $L = 25$ residues, Fig. 4b, suggests that at this point the critical hydrophobic segment in the re-entrant loop (LMYW) has just reached the periplasmic exit from the SecYEG channel and is able to partition into the periplasmic leaflet of the membrane, while the TMH is just entering the translocon channel. We hypothesize that, when in this location, the re-entrant loop interacts with the lateral gate in such a way that it shifts the equilibrium towards the closed state [26] and thereby reduces the interaction between the TMH and the membrane.

In summary, we find that both charged and hydrophobic upstream segments can impact the timing of cotranslational insertion of model TMHs into the *E. coli* inner membrane, presumably reflecting both the transmembrane electric field and the dynamics of the lateral gate in the SecYEG channel.

Acknowledgments

We thank Dr. Rickard Hedman (Stockholm University) for programming and maintenance of the EasyQuant software, and for providing the 4L/15A SecM(*Ec*), 6L/13A SecM(*Ec*) and 6L/13A SecM(*Ec*-Sup1) constructs used in this study. This work was supported by grants from the Knut and Alice Wallenberg Foundation (2017.0323), the Novo Nordisk Fund (NNF18OC0032828), and the Swedish Research Council (621-2014-3713) to GvH, and by a Marie Curie Initial Training Network Grant (Horizon 2020, ProteinFactory 642863) to FN.

Competing interests

The authors declare no competing interests.

References

- [1] Chitwood, P.J., Juskiewicz, S., Guna, A., Shao, S. and Hegde, R.S. (2018). EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell* **175**, 1507-1519 e16.
- [2] Rapoport, T.A., Li, L. and Park, E. (2017). Structural and Mechanistic Insights into Protein Translocation. *Annu Rev Cell Dev Biol* **33**, 369-390.
- [3] Hessa, T., Reithinger, J.H., von Heijne, G. and Kim, H. (2009). Analysis of transmembrane helix integration in the endoplasmic reticulum in *S. cerevisiae*. *J Mol Biol* **386**, 1222-1228.
- [4] Hessa, T. et al. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-1030.
- [5] Harris, N.J. et al. (2017). Structure formation during translocon-unassisted co-translational membrane protein folding. *Sci Rep* **7**, 8021.
- [6] Janoschke, M., Zimmermann, M., Brunauer, A., Humbel, R., Junne, T. and Spiess, M. (2021). Efficient integration of transmembrane domains depends on the folding properties of the upstream sequences. *Proc Natl Acad Sci U S A* 10.1073/pnas.2102675118
- [7] Mercier, E., Wintermeyer, W. and Rodnina, M.V. (2020). Co-translational insertion and topogenesis of bacterial membrane proteins monitored in real time. *EMBO J* 10.15252/emj.2019104054
- [8] Ismail, N., Hedman, R., Schiller, N. and von Heijne, G. (2012). A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nature Struct Molec Biol* **19**, 1018-1022.

- [9] Nicolaus, F. et al. (2021). Residue-by-residue analysis of cotranslational membrane protein integration *in vivo*. *eLife* 10.7554/eLife.64302
- [10] van den Berg, B., Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C. and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. *Nature* **427**, 36-44.
- [11] Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345.
- [12] Dalbey, R.E. and Wickner, W. (1986). The role of the polar, carboxyl-terminal domain of *Escherichia coli* leader peptidase in its translocation across the plasma membrane. *J Biol Chem* **261**, 13844-13849.
- [13] Ito, K. and Chiba, S. (2013). Arrest peptides: cis-acting modulators of translation. *Annu Rev Biochem* **82**, 171-202.
- [14] Goldman, D.H., Kaiser, C.M., Milin, A., Righini, M., Tinoco, I. and Bustamante, C. (2015). Mechanical force releases nascent chain-mediated ribosome arrest *in vitro* and *in vivo*. *Science* **348**, 457-460.
- [15] Kemp, G., Nilsson, O.B., Tian, P., Best, R.B. and von Heijne, G. (2020). Cotranslational folding cooperativity of contiguous domains of alpha-spectrin. *Proc Natl Acad Sci U S A* **117**, 14119-14126.
- [16] Cymer, F., Hedman, R., Ismail, N. and von Heijne, G. (2015). Exploration of the arrest peptide sequence space reveals arrest-enhanced variants. *J Biol Chem* **290**, 10208-10215.
- [17] Nilsson, O.B. et al. (2015). Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep* **12**, 1533-1540.

- [18] Sandhu, H., Hedman, R., Cymer, F., Kudva, R., Ismail, N. and von Heijne, G. (2021). Cotranslational Translocation and Folding of a Periplasmic Protein Domain in *Escherichia coli*. *J Mol Biol* 10.1016/j.jmb.2021.167047
- [19] Wolfe, P.B., Wickner, W. and Goodman, J.M. (1983). Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. *J Biol Chem* **258**, 12073-12080.
- [20] Niesen, M.J.M., Muller-Lucks, A., Hedman, R., von Heijne, G. and Miller, T.F., 3rd. (2018). Forces on Nascent Polypeptides during Membrane Insertion and Translocation via the Sec Translocon. *Biophys J* **115**, 1885-1894.
- [21] Ismail, N., Hedman, R., Lindén, M. and von Heijne, G. (2015). Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat Struct Mol Biol* **22**, 145-149.
- [22] Leininger, S.E., Trovato, F., Nissley, D.A. and O'Brien, E.P. (2019). Domain topology, stability, and translation speed determine mechanical force generation on the ribosome. *Proc Natl Acad Sci U S A* **116**, 5523-5532.
- [23] Yap, M.N. and Bernstein, H.D. (2009). The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel. *Mol Cell* **34**, 201-211.
- [24] von Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J* **5**, 3021-3027.
- [25] Cymer, F., von Heijne, G. and White, S.H. (2015). Mechanisms of integral membrane protein insertion and folding. *J Mol Biol* **427**, 999-1022.

- [26] Mercier, E., Wang, X., Maiti, M., Wintermeyer, W. and Rodnina, M.V. (2021). Lateral gate dynamics of the bacterial translocon during cotranslational membrane protein insertion. *Proc Natl Acad Sci U S A* 10.1073/pnas.2100474118
- [27] Hvorup, R.N., Goetz, B.A., Niederer, M., Hollenstein, K., Perozo, E. and Locher, K.P. (2007). Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. *Science* **317**, 1387-1390.
- [28] Frauenfeld, J. et al. (2011). Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. *Nat Struct Mol Biol* **18**, 614-621.
- [29] Whitley, P., Nilsson, I.M. and von Heijne, G. (1996). A nascent secretory protein may traverse the ribosome/ER translocase complex as an extended chain. *J Biol Chem* **271**, 6241-6244.
- [30] Bano-Polo, M. et al. (2018). Transmembrane but not soluble helices fold inside the ribosome tunnel. *Nat Commun* **9**, 5246.
- [31] Lu, J. and Deutsch, C. (2005). Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry* **44**, 8230-43.
- [32] Tu, L.W. and Deutsch, C. (2010). A folding zone in the ribosomal exit tunnel for Kv1.3 helix formation. *J Mol Biol* **396**, 1346-60.
- [33] Zhang, J., Pan, X., Yan, K., Sun, S., Gao, N. and Sui, S.F. (2015). Mechanisms of ribosome stalling by SecM at multiple elongation steps. *eLife* 10.7554/eLife.09684

Figure legends

Figure 1. *The force profile assay.* (a) Basic construct. The LepB transmembrane helices TMH1 and TMH2, the flanking region (FR; positively charged loop or re-entrant loop), the H-segment, and the SecM(*Ec*) AP are indicated, together with the arrested (*A*) and full-length (*FL*) products. *L* denotes the number of residues between the C-terminal end of the H-region

and the last residue in the AP. (b) At construct length L_1 , the H-region has not yet entered the SecYEG channel, no pulling force F is generated, and little full-length product is produced ($F \approx 0, f_{FL} \approx 0$). At L_2 , the H-region is integrating into the membrane, generating a strong pulling force ($F \gg 0, f_{FL} \approx 1$). At L_3 , the H-region is already integrated and little pulling force is generated ($F \approx 0, f_{FL} \approx 0$). (c) SDS-PAGE gels showing A and FL products for [^{35}S]-Met labelled and immunoprecipitated 4L/15A constructs with different L -values (c.f., panel *d*). Control constructs A_C and FL_C have, respectively, a stop codon and an inactivating Ala codon replacing the last Pro codon in the AP, leading to the production of A -sized and FL -sized proteins. Mw markers are indicated on the left. (d) FPs obtained with the SecM(*Ec*) AP for constructs lacking the FR segment and with H-segments composed of BtuC TMH2 (green), BtuC TMH6 (magenta), 4L/15A (blue), 6L/13A (dashed blue), and the 6L/13A construct but with the SecM(*Ec*-Sup1) AP (dark grey), c.f. [8,9]. The L_{start} value for 4L/15A is indicated by the arrow. Error bars show SEM values ($n = 3$).

Figure 2. *Analysis of the positively charged BtuC TMH1-TMH2 loop.* (a) Structure of the BtuC TMH1-TMH2 region (PDB 2QI9) [27] with the three Arg residues in spacefill. The sequence of the loop is shown on top. (b) FPs for 4L/15A constructs lacking (blue) and including (orange) the positively charged loop (FR in Fig. 1a). Results for R→Q mutants are shown in green. (c) FPs for 6L/13A constructs lacking (blue) and including (orange) the positively charged loop (FR in Fig. 1a). (d) FPs obtained with the SecM(*Ec*-Sup1) AP for 6L/13A (dark grey), 6L/13A with the positively charged loop (orange), and 6L/13A with the re-entrant loop (red). Error bars show SEM values ($n = 3$).

Figure 3. *Analysis of the TMH5-TMH6 re-entrant loop.* (a) Structure of the BtuC TMH5-TMH6 region (PDB 2QI9) [27] with the LMYW residues in spacefill. The sequence of the re-entrant loop is shown on top. (b) FPs for 4L/15A constructs lacking (blue) and including (red) the re-entrant loop (FR in Fig. 1a). (c) FPs for 6L/13A constructs lacking (blue) and

including (red) the re-entrant loop (FR in Fig. 1a). (d) f_{FL} values obtained with the SecM(*Ec*) AP for mutants of the re-entrant (RE) loop-4L/15A[L=30] construct where $n = 5, 9, 13$, and 17 residues in the re-entrant loop are replaced by GSGS repeats starting from the C-terminal end, or where the LMYW segment in the re-entrant loop is replaced by GSGS (see Supplementary Fig. S1 for sequences). The -RE and +RE values are from panel *b*. Error bars show SEM values ($n = 3$).

Figure 4. *Models of ribosome-nascent chain-SecYEG complexes.* (a) Model for the positively-charged loop $L = 28$ construct with the H-segment in blue and the three Arg residues in the positively charged loop in spacefill. The location of the hydrophobic gasket is indicated in orange. The model is based on PDB 4V6M [28], assuming that the H-segment is helical [29-32] and that the nascent chain model in the ribosome exit tunnel in the PDB file is stretched by ~ 2 residues compared to the SecM AP [33]. (b) Same as in panel *a*, but for the re-entrant loop $L = 25$ construct. The LMYW residues are in spacefill.

Figure 1

Figure 2

Figure 3

Figure 4

Supplementary Figure S1

Design of protein sequences

LepB TMH1 TMH2 - RRR loop - test TMH - linker (4-47aa- SecM(Ec) AP) - *LepB* (23aa)

MANMFALILVIATLVTGILWCVDKFFFPAPKRRERQAAAQAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLI
VRSFIYEPFQIPSGSMPTLNSTDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLP
GDKVTYDVPVSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQTFSSRRNGGEATSGFFEVPKQETKENGIRLSETS
GGPG-**RGELFVWQIRLPR**-**AAAAALALAAAAALALAAAA**-(4-47aa)-**FSTPVWISQAQGIRAGP**-
GSSDKQEGEWPTGLRLSRIGGIH

LepB TMH1 TMH2 - re-entrant loop - test TMH - linker (4-47aa- SecM(Ec) AP) - *LepB* (23aa)

MANMFALILVIATLVTGILWCVDKFFFPAPKRRERQAAAQAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLI
VRSFIYEPFQIPSGSMPTLNSTDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLP
GDKVTYDVPVSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQTFSSRRNGGEATSGFFEVPKQETKENGIRLSETS
GGPG-**LMYWMMGGFGGVDWRQS**-**AAAAALALAAAAALALAAAA**-(4-47aa)-**FSTPVWISQAQGIRAGP**-
GSSDKQEGEWPTGLRLSRIGGIH

Flanking region:

RGELFVWQIRLPR (RRR loop)
QGEFVWQIRLPR (QRR loop)
RGELFVWQIRLPQ (RRQ loop)
QGEFVWQIRLPQ (QRQ loop)
QGEFVWQIQLPQ (QQQ loop)

LMYWMMGGFGGVDWRQS (Re-entrant loop)
LMYWMMGGFGGVGSGSG (Re-entrant loop(Cterm 5GS))
LMYWMMGGGSGSGSGSG (Re-entrant loop(Cterm 9GS))
LMYWGGSGSGSGSGSGSG (Re-entrant loop(Cterm 13GS))
GSGSGSGSGSGSGSGSG (Re-entrant loop(Cterm 17GS))
GSGSMMGGFGGVDWRQS (Re-entrant loop(Nterm 4GS))

Hydrophobic test TMH:

AAAAALALAAAAALALAAAA (4L/15A)

AAAAALALALALALALAAAA (6L/13A)

Linker sequence:

	<i>L</i>
GPGG FSTPVWISQAQGIRAGP	21
GPGGVPG FSTPVWISQAQGIRAGP	24
GPGGWMSS FSTPVWISQAQGIRAGP	25
GPGGVPGWMSS FSTPVWISQAQGIRAGP	28
GPGGVPGQQWMSS FSTPVWISQAQGIRAGP	30
GPGGVPGQQNATWIMSS FSTPVWISQAQGIRAGP	33
GPGGVPGQQNATWIAWMSS FSTPVWISQAQGIRAGP	36
GPGGVPGQQNATWIVPSS FSTPVWISQAQGIRAGP	37
GPGGVPGQQNATWIVPPGWMSS FSTPVWISQAQGIRAGP	39
GPGGVPGQQNATWIVPPGQYWMSS FSTPVWISQAQGIRAGP	41
GPGGVPGQQNATWIVPPGQYFWMSS FSTPVWISQAQGIRAGP	43
GPGGVPGQQNATWIVPPGQYFMMGDWMSS FSTPVWISQAQGIRAGP	46
GPGGVPGQQNATWIVPPGQYFMMGDNRDWMSS FSTPVWISQAQGIRAGP	49
GPGGVPGQQNATWIVPPGQYFMMGDNRDNSAWSS FSTPVWISQAQGIRAGP	52
GPGGVPGQQNATWIVPPGQYFMMGDNRDNSADSRWMSS FSTPVWISQAQGIRAGP	55
GPGGVPGQQNATWIVPPGQYFMMGDNRDNSADSRWGFVPEANWMSS FSTPVWISQAQGIRAGP	64