© 2017. Published by The Company of Biologists Ltd | Journal of Cell Science (2017) 130, 372-381 doi:10.1242/jcs.194472

RESEARCH ARTICLE

Integration of transmembrane domains is regulated by their downstream sequences

Tina Junne and Martin Spiess*

ABSTRACT

The Sec61 translocon catalyzes translocation of proteins into the endoplasmic reticulum and the lateral integration of transmembrane segments into the lipid bilayer. Integration is mediated by the hydrophobicity of a polypeptide segment consistent with thermodynamic equilibration between the translocon and the lipid membrane. Integration efficiency of a generic series of increasingly hydrophobic sequences (H-segments) was found to diverge significantly in different reporter constructs as a function of the ∼100 residues that are C-terminal to the H-segments. The hydrophobicity threshold of integration was considerably lowered through insertion of generic ∼20-residue peptides either made of flexible glycine–serine repeats, containing multiple negative charges, or consisting of an oligoproline stretch. A highly flexible, 100-residue glycine–serine stretch maximally enhanced this effect. The apparent free energy of integration was found to be changed by more than 3 kcal/mol with the downstream sequences tested. The C-terminal sequences could also be shown to affect integration of natural mildly hydrophobic sequences. The results suggest that the conformation of the nascent polypeptide in the protected cavity between the ribosome and translocon considerably influences the release of the H-segment into the bilayer.

KEY WORDS: Endoplasmic reticulum, Membrane insertion, Protein translocation, Sec61, Translocon

INTRODUCTION

The Sec61/SecY translocon in the endoplasmic reticulum (ER) membrane of eukaryotes and in the plasma membrane of bacteria performs the dual function of translocating hydrophilic sequences of secretory and membrane proteins across the membrane, and of mediating the integration of transmembrane segments into the lipid bilayer (Park and Rapoport, 2012; Shao and Hegde, 2011). Structure determination (e.g. van den Berg et al., 2004) has revealed the protein-conducting pore to be generated from a single copy of the 10-transmembrane-domain helix-bundle protein SecY/Sec61α (encoded by *SEC61A1* in mammals), which associates with a β and a γ subunit to form a heterotrimeric complex SecYEG in bacteria and Sec61αβγ in eukaryotes (Sec61p, Sbh1p and Sss1p in yeast). Polypeptides inserted into the channel can exit laterally into the lipid membrane through a gate formed by transmembrane domains 2 and 3, and 7 and 8. Partial opening of the lateral gate has been displayed in fortuitous crystal structures (Egea and Stroud, 2010; Tsukazaki et al., 2008). Recent cryo-electron microscopy structures with stalled or arrested nascent chains show a signal sequence or a transmembrane segment intercalated between the gate helices (Gogala et al., 2014; Park et al., 2014; Voorhees and Hegde, 2016), representing intermediate states of membrane insertion.

The main requirement for membrane integration and its energetic driving force is the hydrophobicity of a potential transmembrane sequence. Systematic quantitative analyses of the insertion of mildly hydrophobic sequences (H-segments) into the membrane in a reporter protein yielded a 'biological hydrophobicity scale', similar to scales determined by physical partitioning (Hessa et al., 2005, 2007, 2009; Öjemalm et al., 2013, 2011). The analyses suggested that membrane insertion is a purely thermodynamic equilibration process for each polypeptide segment entering the translocon between the lipid phase and the more hydrophilic environment of the pore. These studies allowed quantification of the apparent energy contributions (ΔG_{app}) of each amino acid at any position in an H-segment to membrane insertion and thus prediction of transmembrane segments, at least for single-spanning membrane proteins, where the process is not complicated by specific transmembrane– transmembrane interactions (Hedin et al., 2010).

In comparison with the biophysical scales, the apparent insertion-free-energy (ΔG_{app}) scale was found to be compressed and shifted (Gumbart et al., 2013; MacCallum and Tieleman, 2011; Schow et al., 2011), which is likely to reflect the fact that the transmembrane sequence does not equilibrate between lipid and free solution, but instead between lipid and the interior of the translocation pore, which forms a constricted space with limited water content. Most prominently, the six mostly hydrophobic residues that form the central constriction point act as a gasket around the polypeptide substrate and exclude water (Park and Rapoport, 2011). Mutation of these constriction residues to more hydrophilic amino acids, thus increasing polarity and hydration inside the pore, reduce the hydrophobicity threshold for membrane integration considerably (Demirci et al., 2013; Junne et al., 2010), perfectly in line with the equilibration model. In addition, water molecules inside the translocon show anomalous behavior departing from that in bulk water in molecular dynamics simulations (Capponi et al., 2015).

The biological hydrophobicity scale, including position dependence (based on Hessa et al., 2007), is a useful tool for transmembrane domain prediction of single-spanning membrane proteins. As illustrated in Schow et al. (2011), it improves the prediction of transmembrane domains in comparison to a biophysical scale, but at the expense of mis-predicting relatively hydrophobic translocated sequences. The simple equilibration model of membrane integration expects the process to be largely independent of all but the immediate flanking sequences of an H-segment, a prerequisite for general prediction. In this study, we tested in yeast (*Saccharomyces cerevisiae*) the hydrophobicity threshold of identical H-segments in the context of two different reporter proteins derived from dipeptidyl aminopeptidase B (DPAPB; also known as DAP2) or

Biozentrum, University of Basel, Klingelbergstrasse 70, Basel CH-4056, Switzerland.

*Author for correspondence (martin.spiess@unibas.ch)

M.S., 0000-0001-7139-0550

Received 27 June 2016; Accepted 10 November 2016

Journal of Cell Science

Journal of Cell Science (2017) 130, 372-381 doi:10.1242/jcs.194472

carboxypeptidase Y (CPY; also known as PRC1). DPAPB is a signal-recognition particle (SRP)-dependent type II membrane protein, whereas CPY is an SRP-independent secretory and vacuolar protein. A striking difference was observed for these two reporter proteins. The sequence responsible for this difference could be localized to approximately 100 residues C-terminal to the H-segment. Our findings suggest that the hydrophobicity-based membrane integration mechanism is superimposed by conformational effects of the downstream sequence modulating the release of the H-segment into the lipid bilayer.

RESULTS

The integration threshold of identical H-segments differs in distinct reporter proteins

We have previously determined the hydrophobicity threshold for 50% membrane integration in *Saccharomyces cerevisiae* by inserting a series of H-segments into diaminopeptidylaminopeptidase B (DPAPB; schematically shown in Fig. 1, top). The inserted H-segments comprise a 19-residue oligoalanine host sequence with 0–8 alanine residues replaced by leucine residues and flanked by GGPG and GPGG insulators (Fig. 1, bottom), following the example of Hessa et al. (2005). Upon expression in yeast, labeling with [^{35}S] methionine for 5 min, immunoprecipitation, separation by SDS-gel electrophoresis and autoradiography, the topology of the products can be derived from the glycosylation pattern (Fig. 1, box). The signal-anchor sequence targets the protein to the ER and initiates translocation of the downstream sequence across the membrane. The H-segments then either stop further translocation through membrane integration, resulting in a product glycosylated only in the lumenal sequence that connects the signal with the H-segment, or is transferred across the membrane followed by the rest of the protein, causing full glycosylation. As expected, integration efficiency increased with increasing hydrophobicity (Fig. 2A, construct DHDD). Considering the results to be the outcome of an equilibration process with an apparent equilibrium constant $K_{app}=f_I/f_T$ (the ratio of the integrated to translocated fraction) as proposed by Hessa et al. (2005, 2007), the apparent free energies of integration, $\Delta G_{app}=-RT\ln K_{app}$, can be calculated. There was an overall linear relationship between the number of leucine residues in the H-segment and ΔG_{app} (Fig. 2), based on which the probability distribution for integration was plotted. As shown previously (Demirci et al., 2013; Junne et al., 2010), the threshold for 50% integration (i.e. $\Delta G_{app}=0$) for this DPAPB-derived model protein was approximately four leucine residues, similar to that also observed for almost identical H-segments inserted into *Escherichia coli* leader peptidase constructs and analyzed by *in vitro* translation and translocation into dog pancreas microsomes (Hessa et al., 2005, 2007).

DPAPB is an established SRP-dependent cotranslationally translocated substrate. In order to test the integration behavior of the same H-segments in an entirely different reporter sequence, they were inserted into the corresponding position of carboxypeptidase Y (CPY), an established SRP-independent protein capable of post-translational (i.e. ribosome-independent) translocation (Ng et al., 1996). To avoid possible effects of sequences immediately flanking the potential transmembrane sequences, the H-segments were transplanted together with the 12 N- and 16 C-terminally adjacent

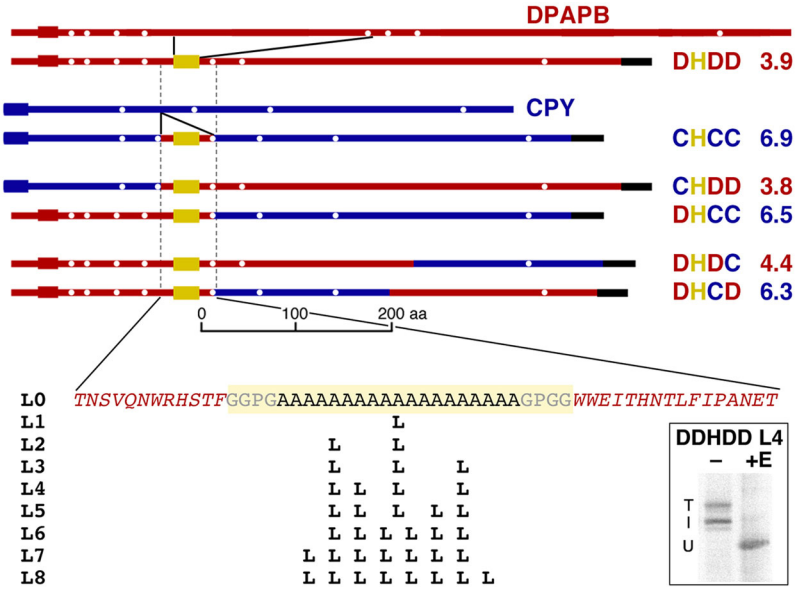

Fig. 1. Model proteins to determine the hydrophobicity threshold for membrane integration. The wild-type sequences of DPAPB and CPY are schematically shown as red and blue lines, respectively, with white dots indicating potential glycosylation sites. The signal-anchor of DPAPB and the cleaved signal of CPY are indicated as red and blue boxes, respectively. H-segments, comprising a 19-residue oligoalanine sequence, flanked by glycine–proline insulators, with 0–8 alanine residues replaced by leucine residues to increase the hydrophobicity as listed below, are shown as yellow boxes. The black bar indicates C-terminal triple-HA epitope tags. The H-segments were initially inserted into the DPAPB sequence replacing residues 170–378 (Junne et al., 2010). These H-segments, including the immediate DPAPB flanking sequences (12 and 16 residues at the N- and C-termini, respectively) were transplanted as a cassette into the coding sequence of CPY after codon 160. Portions of the two H-segment-containing model proteins corresponding to the sequence N-terminal to the H-segment, including the signal or signal-anchor sequence, the proximal ~200 residues downstream of the H-segment and the distal C-terminal sequence, were swapped as indicated. The constructs were named according to the origins of these three portions before and after the H-segment. The numbers next to the constructs indicate the number of leucine residues in the H-segment required for 50% membrane integration as determined in Fig. 2. The box inset below shows an example of DHDD with an L4 H-segment expressed in yeast cells, labeled with [^{35}S]methionine for 5 min, immunoprecipitated, and incubated with and without endoglycosidase H (+ or –E) before SDS-gel electrophoresis and autoradiography. It allows distinction of translocation (full glycosylation; T) and integration (partial glycosylation; I) of the H-segment, as well as identification of uninserted unglycosylated products (U). All protein sequences are listed in Table S1.

Journal of Cell Science (2017) 130, 372-381 doi:10.1242/jcs.194472

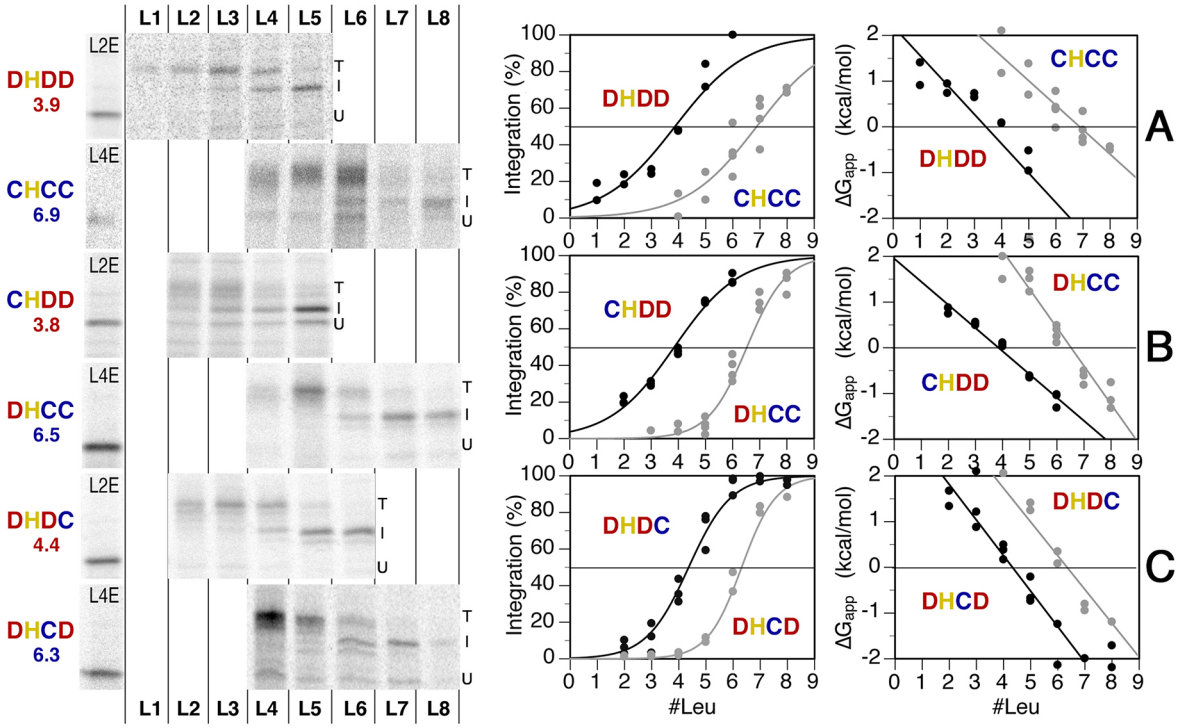

Fig. 2. Integration of H-segments in the context of DPAPB and CPY sequences differ dependent on the proximal downstream sequence. (A–C) The various reporter constructs with H-segments containing 1–8 leucine residues (L1–L8) were expressed in yeast cells, labeled with [^{35}S]methionine for 5 min, immunoprecipitated and analyzed by SDS-gel electrophoresis and autoradiography (left). Based on the glycosylation pattern, translocation ('T') or integration ('I') of the H-segment, as well as uninserted unglycosylated products (U) were distinguished. To identify the position of the unglycosylated polypeptide, a matching sample after deglycosylation by endoglycosidase H is shown on the left (L#E, with # indicating the number of leucine residues in the protein analyzed). The fraction of products with integrated H-segments as a percentage of the total membrane-targeted proteins was plotted versus the number of leucine residues in the H-segment (middle column). Curve fitting was performed by plotting the data as apparent free energies ΔG_{app} of membrane insertion and analyzing the linear regression (right). The number of leucine residues required for 50% integration (ΔG_{app}=0) determined in this way is indicated below the construct names on the left. All constructs were analyzed in 2–4 independent experiments corresponding to the individual values shown. Data from panel B are also shown in Fig. 3 for comparison.

residues of DPAPB (Fig. 1). The resulting CPY-derived model construct was named CHCC as opposed to DHDD for the original DPAPB construct, where 'C' and 'D' indicate the CPY or DPAPB origin of the sequence N-terminal to the H-segment, and of the proximal and distal sequence C-terminal to the H-segment cassette, respectively. Upon expression in yeast, the hydrophobicity threshold for H-segment integration in CHCC was increased significantly to almost seven leucine residues, corresponding to an apparent free-energy shift of almost 2 kcal/mol. This result indicates that membrane integration does not depend on the H-segment alone.

The H-segment integration threshold is influenced by the proximal downstream sequence

To test whether the threshold difference is due to the sequence N-terminal to the H-segment – i.e. the signal and the translocated sequence – or that C-terminal to the potential transmembrane segment, we analyzed constructs CHDD and DHCC in which these portions had been exchanged. Surprisingly, the integration threshold did not correlate with the signal sequence (Jan et al., 2014), but did with the identity of the C-terminal sequence (Fig. 2B). Within the C-terminal sequence, the proximal half determined the integration behavior – the threshold of the DPAPB-based model protein was increased to more than six leucine residues when only the proximal 200 residues were replaced by the CPY sequence in DHCD, but not when this was done for the distal ∼200 residues in DHDC (Fig. 2C).

Replacing all the C-terminal sequences but the ∼100 residues adjacent to the H-segment for CPY sequences in construct $DHD_{116}C$ retained the lower threshold of four leucine residues (Fig. 3). Switching from the DPAPB to the CPY sequence at various positions in this 100-amino-acid region downstream of the H-segment cassette produced intermediate integration behaviors with thresholds that gradually increased with proximity of the switch point to the H-segment (Fig. 3). These results demonstrate that the requirements for transmembrane integration are influenced by downstream sequences that in part are not yet synthesized when the H-segment first enters the translocon. They suggest that the insertion process extends at least to the time when the proximal 100 residues are translated or have even emerged from the ribosome exit tunnel.

Membrane integration is not affected by codon usage

The two original proteins differ significantly in their natural expression levels in yeast. DPAPB and CPY are produced naturally at ∼600 and ∼12,000 molecules/cell, respectively (Kulak et al., 2014). Accordingly, they also differ in their codon usage with CPY predominantly using the most frequently used codons (Sharp et al., 1988) for which there are typically higher concentrations of matching tRNAs in the cell for more rapid synthesis. No such bias is found for the codons encoding DPAPB. To test the possibility that the rate of translation of the nascent polypeptide during H-segment integration and translocation affects this process, the codons 49–70 positions C-terminal to the H-segment (which are

Journal of Cell Science (2017) 130, 372-381 doi:10.1242/jcs.194472

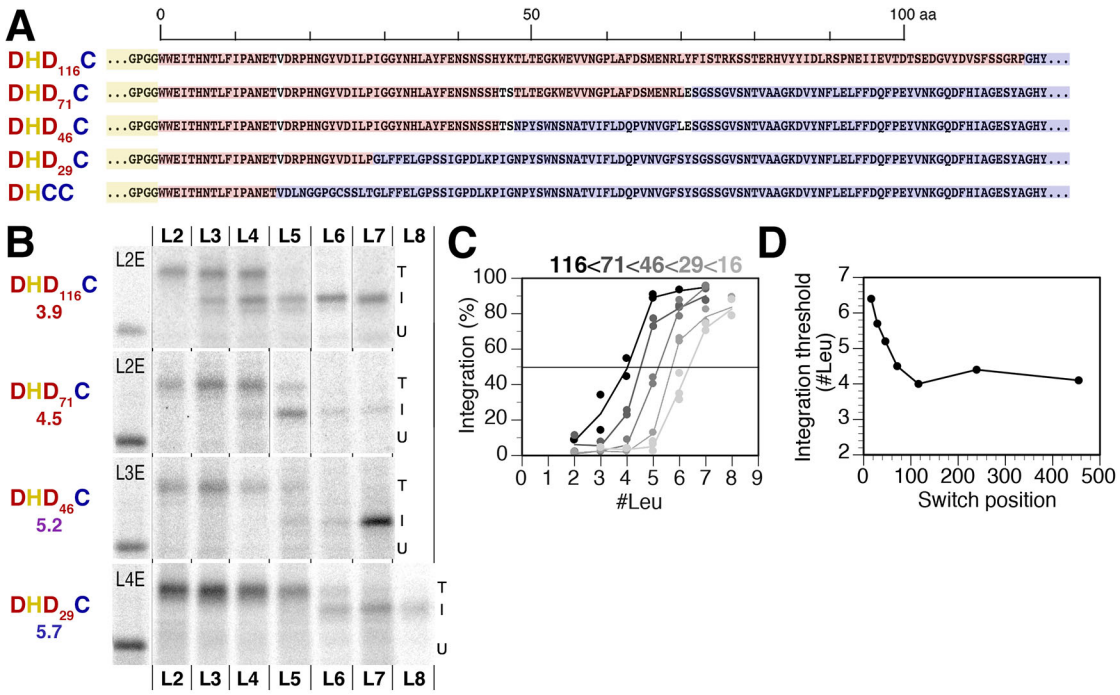

Fig. 3. The hydrophobicity threshold of membrane integration is affected by sequences within a range of 100 residues C-terminal to the H-segment. (A) Reporter proteins were constructed that switched from the DPAPB sequence (red) to the CPY sequence (blue) at positions 116, 71, 46 and 29 C-terminal to the H-segment. (B) The proteins were labeled and analyzed as described in Fig. 2. (C) H-segment integration was plotted together with that of DHCC (switching at position 16) from Fig. 2B. Individual measurements are plotted with a line connecting the means. Data for $DHD_{29}C$ are also re-shown in Fig. 5B. Some data from B and C is also shown in Fig. 7 for comparison. (D) The integration thresholds expressed as the number of leucine residues in the H-segment required for 50% integration of these constructs as well as of DHDC and DHDD are plotted against the position of switching from the DPAPB to the CPY sequence following the H-segment.

translated during and shortly after the H-segment enters the translocon) were modified in construct $DHD_{71}C$ by using synonymous substitutions from the original relatively 'slow' composition of the DPAPB sequence to the most frequent ('fast') composition, and in construct $DHD_{46}C$ from the original composition of frequently used codons of CPY to the rarest ('slow') codons (Fig. 4A). H-segment integration with these constructs was not significantly affected (Fig. 4B), indicating that the observed threshold difference is not due to codon usage and translation rate.

Negatively charged and oligoproline sequences reduce the hydrophobicity threshold for membrane integration

How do protein sequences C-terminal to the H-segment affect the insertion process? Comparison of the DPAPB and CPY sequences in this region of our constructs did not reveal conspicuous features.

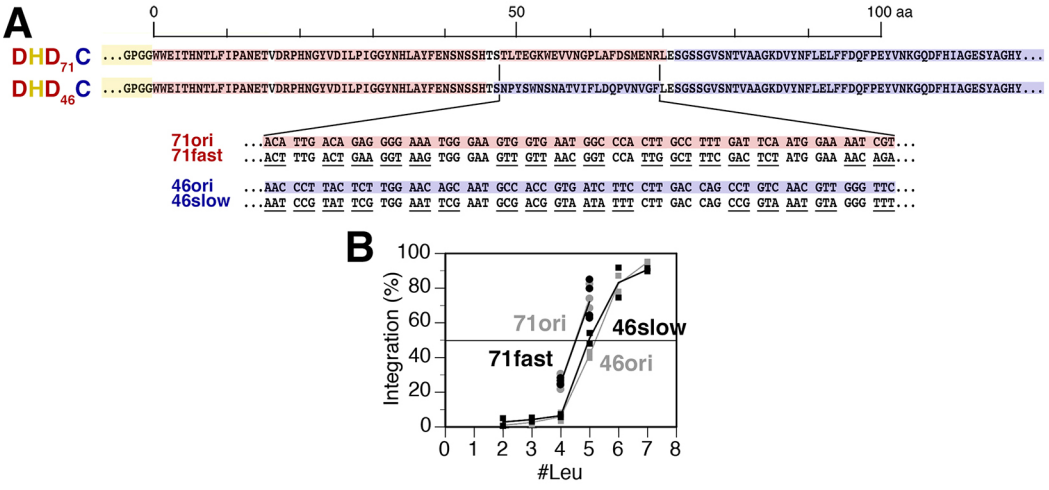

Fig. 4. Membrane integration of the model constructs depends on protein sequence, not codon usage. (A) In the illustrated segments of $DHD_{71}C$ and $DHD_{46}C$, the original codons were exchanged wherever possible (underlined) to the most or least frequently used ones, respectively (Sharp et al., 1988), thus allowing for faster and slower translation, respectively. The end of the H-segment is highlighted in yellow, and sequences derived from DPAPB and CPY are shown in red and blue, respectively. (B) H-Segment integration was determined as described in Fig. 2. Individual measurements are plotted with a line connecting the means ($n \geq 3$).

Journal of Cell Science (2017) 130, 372-381 doi:10.1242/jcs.194472

We therefore decided to test the effect of generic sequences with predictable properties. Because exchanging segments of ∼20 residues already produced reproducible differences, we exchanged segments of this size, taking advantage of existing constructs in which sequences had been swapped that also contained convenient restriction sites. As illustrated in Fig. 5A, sequences were inserted in positions 31–49 C-terminal to the H-segment. When the H-segment enters the translocon, these residues might still reside entirely within the ribosomal exit tunnel. The sequences tested include a very hydrophilic and flexible segment consisting entirely of alternating glycine and serine residues (abbreviated GS; Fig. 5B). This sequence was used as a background to analyze the effect of charged residues – i.e. six lysine residues (6K), six or 12 arginine residues (6R, 12R), six aspartate residues (6D), six glutamate residues (6E), 12 or even 18 negative charges (both aspartate and glutamate residues; 12– and 18–), as well as sequences containing both negative and positive amino acids (4+/4− and 8+/8−). In addition, six proline residues were inserted into the glycine–serine background and a 19-residue oligoproline stretch (19P) was tested.

$DHD_{29}C$ can be considered the parental construct of this series with the CPY-derived sequence LFFELGPSSIGPDLKPIGNPY as residues 31–51 and produced an integration threshold of 5.7 leucine residues (Fig. 3). With the GS sequence in this position, the threshold was reduced to 4.1 leucine residues (Fig. 5B). A minor threshold reduction was observed for the 6K construct, but no significant effect for 6R and 12R constructs. In contrast, negative charges as in the 6D and 6E mutants further facilitated integration with thresholds of 2.4 and 2.1 leucine residues, respectively. Additional negative amino acids in 12– and 18– proteins further enhanced this effect and produced an increasingly steep transition for H-segments with one and two leucine residues from no to almost complete integration, respectively. Neutral sequences containing four or eight negative, and as many positive amino acids, also showed clear threshold reduction, indicating that negative charges are largely dominant over positive ones. Negative charges more than 30 residues downstream of the H-segment thus promoted membrane integration in a dose-dependent manner, as is summarized in Fig. 5C (middle panel). Surprisingly, positive charges did not have an opposite or neutralizing effect (Fig. 5C, left panel).

Six proline residues in the context of a flexible sequence of glycine and serine residues (6P) showed no difference to glycine and serine sequences alone (GS). However, a 19-proline stretch (19P) stimulated integration, reducing the threshold to fewer than two leucine residues (Fig. 5B,C).

To test whether the observed effects of these relatively short sequences are not specific for the particular location in the polypeptide downstream of the H-segment, we placed the 18– and 19P sequences even further away from the H-segment at positions 52–60 (Fig. 6). In both cases, the integration threshold was again strongly reduced, although somewhat less than in the previous position at residues 31–49. The effect therefore might get weaker with the distance from the H-segment.

A flexible and hydrophilic downstream sequence strongly stimulates membrane integration

The effect of insertion of a short generic sequence of 19 glycine and serine residues facilitated H-segment integration, possibly owing to its flexibility and lack of potentially clustering hydrophobic surfaces. With construct GS100 (Fig. 7A), we tested the effect of a much larger segment of ∼100 glycine and serine residues after position 30 from the H-segment, replacing the same number of CPY-derived residues. This further enhanced H-segment integration with a reduction of the 50% threshold to ∼1.2 leucine residues, supporting the notion that conformational flexibility in the polypeptide C-terminal to the H-segment promotes membrane integration of mildly hydrophobic sequences.

Integration or translocation of natural sequences depends on their context

Our finding that the sequences downstream of a potential transmembrane domain influence integration efficiency might explain a fraction of stop-transfer sequences that are incorrectly predicted to be translocated based on the biological hydrophobicity scale that has been determined with a single reporter construct (http://dgpred.cbr.su.se; Hessa et al., 2007) as illustrated by Schow et al. (2011, figure 8 in that article). For example, human CD99 is an established type I single-spanning membrane protein; its transmembrane domain, however, is predicted to have a positive ΔG_{app} of +1.0 kcal/mol corresponding to an integration probability of less than 20% (Table 1). Similarly, the transmembrane domain of human STIM2 is predicted to have a ΔG_{app} close to 0 kcal/mol, yet the natural protein is known to span the membrane. To test the dependence on the downstream sequence for insertion of natural sequences, we inserted the transmembrane segments of CD99 and STIM2 (as listed in Table 1 as sequence X) into constructs DHCC (integration threshold of 6.7 leucine residues) and DHD_{29}[GS100]C (threshold of 1.2 leucine residues), replacing the H-segment between the flanking GGPG and GPGG insulators. With the C-terminal GS100 sequence, both transmembrane sequences integrated efficiently into the membrane (Fig. 7B). With the terminal CPY sequence, however, the CD99 transmembrane sequence was completely translocated and the STIM2 sequence integrated only to ∼50%. This result confirms the effect of the downstream polypeptide on integration of natural hydrophobic sequences.

We further expanded our analysis to relatively hydrophobic portions within natural secretory proteins, as listed in Table 1. Their predicted ΔG_{app} values ranged 0.9–2.2 kcal/mol, corresponding to membrane integrations of 18–3%. No integration was detected in the DXCC context (where X indicates replacement of the H-segment with the potential transmembrane sequences as listed in Table 1) for any one of them (Fig. 7B). However, the GS stretch in the DXD_{29}[GS100]C construct induced almost complete integration of the most hydrophobic segment of rabbit α_1-antiproteinase F with ΔG_{pred}=0.9 kcal/mol and ∼50% integration of a segment of angiotensinogen with ΔG_{pred}=1.5 kcal/mol. No integration was detected for a second angiotensinogen sequence with ΔG_{pred}=2.2 kcal/mol. These results support the conclusion that natural sequences of intermediate hydrophobicity could be induced to be efficiently integrated or translocated in appropriate C-terminal sequence contexts.

DISCUSSION

It has been previously shown for multispanning membrane proteins that integration of transmembrane domains depends on their sequence context (Hedin et al., 2010). At least in part, this is due to interaction between helices that evolved to pack together in the membrane, thus shifting the equilibrium towards integration. Transmembrane segments in multispanning proteins are also frequently very closely spaced so that their integration is not independent, but must occur jointly as hairpins. In our study, however, we investigated the integration of individual generic H-segments separated by more than 120 residues from the preceding signal or signal-anchor sequence of the reporter protein. To find an effect of sequences of the reporter protein on H-segment integration was unexpected, given that the prevailing model proposed that each

Journal of Cell Science (2017) 130, 372-381 doi:10.1242/jcs.194472

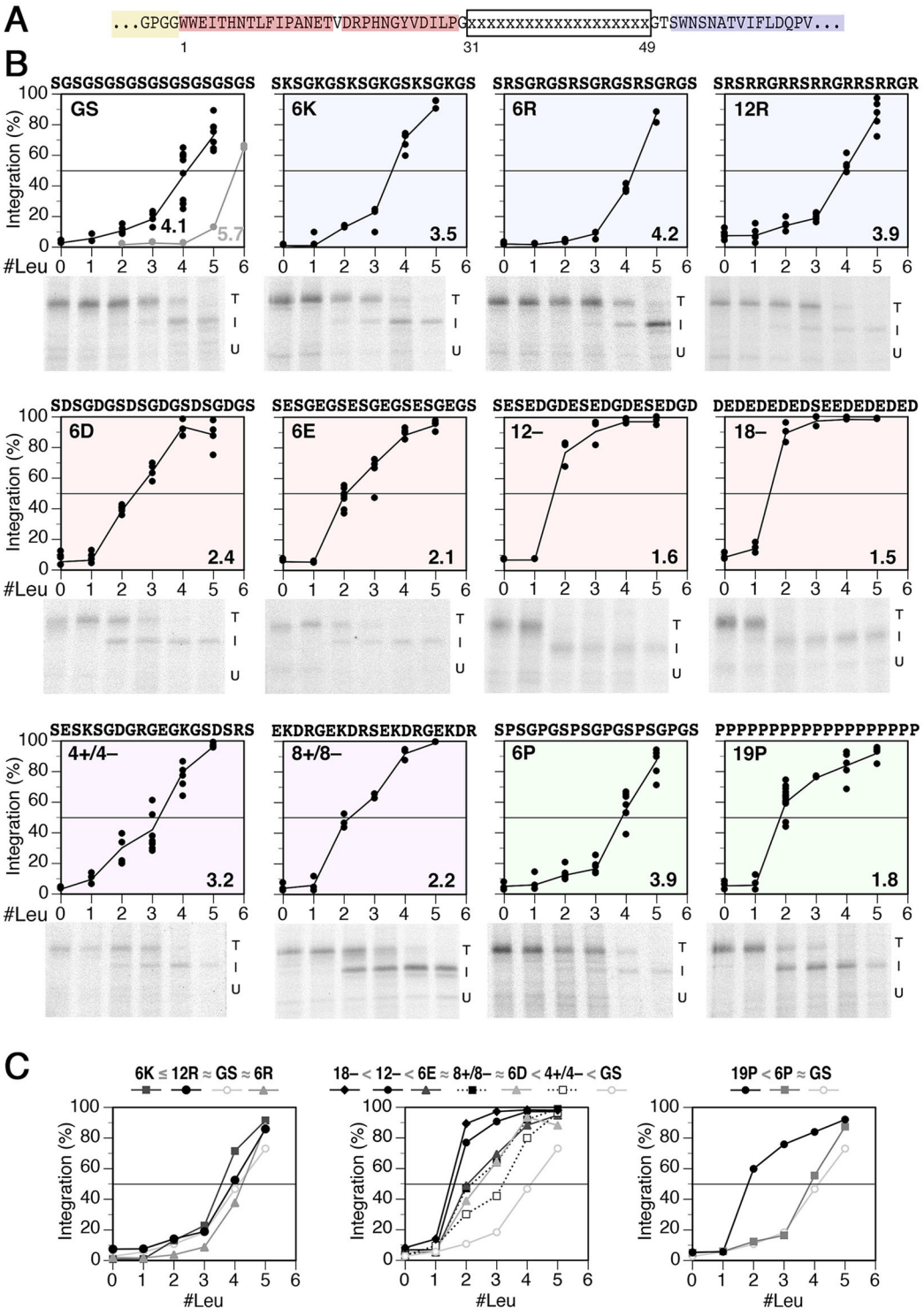

Fig. 5. H-segment integration is strongly affected by generic short sequences inserted 30 residues downstream. (A) Generic sequences of 19 amino acids (as shown in panel B) were inserted at positions 31–49 C-terminal to the H-segment into the reporter construct $DHD_{29}C$ (Fig. 3). The end of the H-segment is highlighted in yellow, and sequences derived from DPAPB and CPY are shown in red and blue, respectively. (B) Constructs containing the indicated sequences and H-segments with 0–5 leucine residues were expressed in yeast. The proteins were labeled and analyzed as described in Fig. 2. Individual determinations of H-segment integration are plotted with a line connecting the means ($n \geq 3$). In the first panel (GS), the results obtained in Fig. 3C for the parental construct $DHD_{29}C$ are shown again in gray for comparison. The interpolated hydrophobicity threshold is indicated as the number of leucine residues required for 50% integration. The graphs are color-coded for positive (blue), negative (red) and mixed charges (purple), as well as for proline residues (green). The data for constructs with the 18– and P19 sequences at position 31–49 are also shown in Fig. 6, and the data for the construct with the GS sequence is shown in Fig. 7 for comparison. (C) To directly compare the effects of positive or negative charges, as well as of prolines in the test sequence, compilations of the respective integration curves are plotted. 6K, six lysine residues; 6R and 12R, six and 12 arginine residues; 6D, six aspartate residues; 6E, six glutamate residues; 12– and 18–, 12 and 18 negative charges (both aspartate and glutamate residues); 4+/4– and 8+/8–, sequences containing both negative and positive amino acids; 6P, insertion of six proline residues into the glycine–serine background; 19P, a 19-residue oligoproline stretch. 'I', insertion; 'T', translocation; 'U', uninserted.

Journal of Cell Science (2017) 130, 372-381 doi:10.1242/jcs.194472

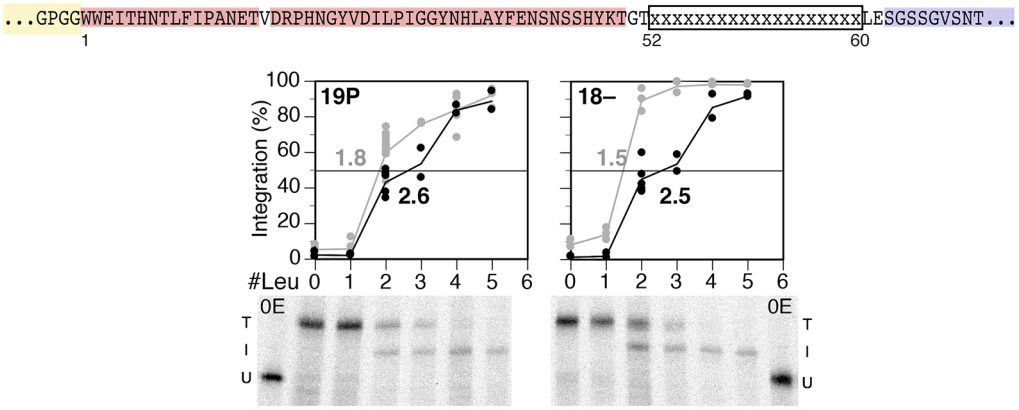

Fig. 6. Position dependence of 19-residue sequences affecting the integration threshold of H-segments. The 18– and 19P sequences that had been inserted 30 amino acids C-terminal to the H-segment and analyzed in Fig. 5 were also inserted after residue 51 as shown. These proteins with H-segments of increasing hydrophobicity were expressed and H-segment integration was analyzed as before. An autoradiograph of 19P and 18– sequences inserted at position 52–60 is shown, together with an L0 sample that had been deglycosylated by endoglycosidase H (L0E). Percentage of integration was quantified and plotted in the graphs above (in black). The values for the constructs with 18– and P19 sequences at position 31–49 (data from Fig. 5B) are re-shown in gray. Individual measurements are plotted with a line connecting the means. The interpolated hydrophobicity threshold is indicated as the number of leucine residues required for 50% integration.

segment of a translocating polypeptide has the opportunity to independently sample the conditions in and out of the translocon through the lateral gate. This notion seemed to be confirmed by the relative success of transmembrane prediction based on the apparent free energy of integration of isolated H-segments determined in a constant context. It was also supported by the fact that the biological integration of H-segments extensively paralleled the spontaneous insertion of corresponding peptides into liposomal membranes (Ulmschneider et al., 2014).

Our main finding is that sequences of the reporter protein downstream of the H-segment can substantially affect membrane integration. The observed thresholds for 50% integration of the standard oligoalanine and leucine H-segment series varied from fewer than two to around seven leucine residues, thus covering an apparent free-energy range of 3–4 kcal/mol. The effect is not limited to sequences in the immediate vicinity of the H-segment, where interaction with lipid head groups might modulate integration. Instead, sequences extending up to 100 residues C-terminal to the H-segment were effective. Most of them have not yet exited the peptide tunnel of the ribosome or even been synthesized when the H-segment first enters the translocon. This argues against a mechanism in which the H-segment in the nascent chain moves through the translocon uniformly at the speed of translation. Rather, the H-segment remains at the translocon–membrane interface until these downstream sequences emerge and can exert their effect. The H-segment, typically flanked by much more hydrophilic sequences, is at an energetic minimum at the translocon and does not spontaneously move forward into the ER lumen. It thus requires energy to pull the H-segment out and the following more hydrophilic sequence into the translocon, most likely through Brownian motion of the already translocated polypeptide and associated chaperones. Indeed, it has been shown experimentally that marginally hydrophobic segments stall at the translocon and cause the subsequent chain to accumulate on the cytosolic side, and to eventually loop out into the cytosol (Kida et al., 2016; Onishi et al., 2013). Accordingly, the sequence C-terminal to the H-segment has time and opportunity to influence H-segment integration.

As the H-segment stalls in the translocon exploring the lipid environment, the C-terminal sequence first gathers in the space formed by the vestibule of the ribosome exit tunnel and the cytosolic cavity of the translocon. This area is open toward the cytosol by a gap of ∼20 Å, and is clearly visible in structures of the ribosome–translocon complex (see Beckmann et al., 1997; Ménétret et al., 2000). This space, and even part of the tunnel itself, have been shown to be large enough to allow folding of a 29-residue zinc finger domain (Conti et al., 2014; Nilsson et al., 2015). The cavity between the ribosome and translocon accommodates a volume of roughly 20 nm^3. Nascent polypeptides have been shown in biophysical experiments to assume compact non-native conformations with partial secondary structures already in the ribosomal tunnel and the exit vestibule (Rodnina and Wintermeyer, 2016). Assuming a protein density of 50% [as has been shown for the 'fluid globule' of OmpX inside the periplasmic chaperone Skp (Burmann et al., 2013)], the cavity could accommodate up to ∼75 residues.

If the peptide connecting the H-segment to the ribosome assumes a compact conformation in the space created between the ribosome and translocon (illustrated in Fig. 7C, left), the diffusion of the H-segment into the lipid bilayer is limited by the short tether. Because the sequences of DPAPB and CPY downstream of the H-segments in our model constructs display no dramatic differences in amino acid or charge composition, conformational dissimilarities that differentially influence translocation versus integration of the H-segment offer themselves as an explanation of our observations. Consistent with this notion, replacement of even a short 20-residue segment 30 residues C-terminal to the H-segment by a sequence comprising only glycine and serine residues produced a clear reduction of the integration threshold from 5.7 to 4.1 leucine residues (Fig. 5). This effect was maximized by insertion of a 100-residue glycine–serine sequence (Fig. 7A). Glycine and serine residues make for a hydrophilic and highly flexible chain that, rather than to compact on itself, might easily diffuse through the gap into the cytosolic space (Fig. 7C, right). By extending the effective length of the leash on which the H-segment is tethered to the ribosome–translocon complex, this flexible sequence increases the probability of the H-segment to dissociate into the bilayer and to be lost from the equilibrium at the translocon. The longer the tether to the ribosome–translocon complex, the higher the entropy is of the H-segment when in the lipid bilayer and the lower the probability of its return.

The effect of charged residues in our model constructs is not entirely obvious. Negative charges within the glycine–serine sequence enhanced integration even further, possibly by extending the peptide owing to charge repulsion. Alternatively, repulsion of

Journal of Cell Science (2017) 130, 372-381 doi:10.1242/jcs.194472

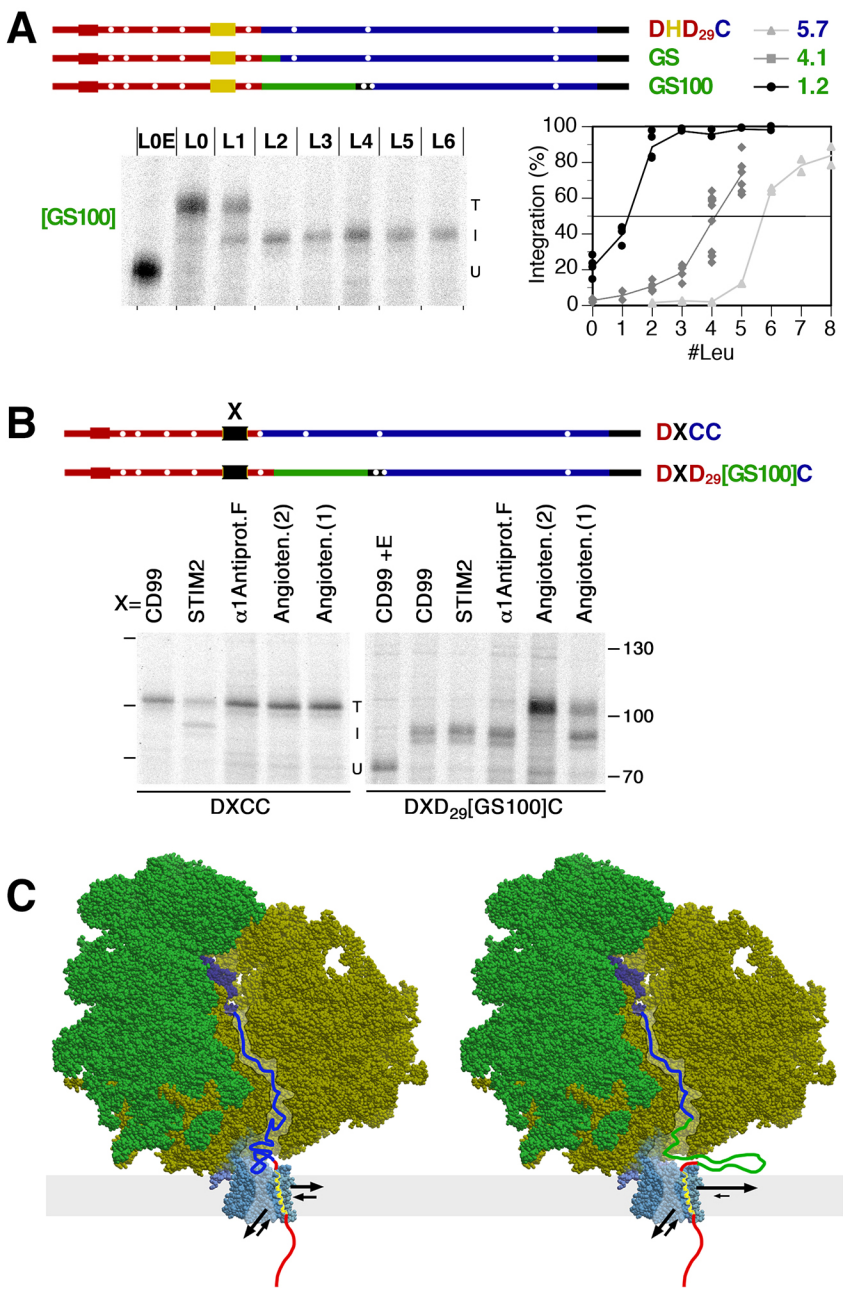

Fig. 7. A long flexible glycine–serine sequence strongly enhances integration of H-segments and natural marginally hydrophobic sequences. (A) A segment of 100 glycine and serine residues was inserted after position 30 following the H-segment in construct GS100 as illustrated in comparison to the GS construct analyzed in Fig. 5B and $DHD_{29}C$ of Fig. 3B and C. GS100 variants with H-segments of increasing hydrophobicity were expressed, and H-segment integration was analyzed as described in Fig. 2. An L0 sample was deglycosylated by endoglycosidase H (L0E). Integration was quantified (in black) and is shown in comparison to that for GS and $DHD_{29}C$ constructs (in gray, from Figs 5B and 3B,C, respectively). Individual measurements are plotted with a line connecting the means ($n \geq 3$). The interpolated hydrophobicity threshold is indicated in panel A as the number of leucine residues required for 50% integration. (B) Marginally hydrophobic natural sequences known to be either integrated into or translocated across the ER membrane as listed in Table 1 were inserted in place of the H-segment in constructs DxCC and DxD_{29}[GS100]C to be analyzed for membrane integration. Constructs were expressed and analyzed as above. The position of molecular mass markers are indicated with their mass in kDa. (C) A model of how conformational properties of the C-terminal sequence affect membrane integration. Ribosome–nascent chain–translocon complexes are shown with the nascent chain extending from the peptidyl t-RNA (purple) through the exit tunnel and the translocon. The polypeptide sequences are colored as in panel B. The H-segment is shown as a yellow helix in the lateral gate of the translocon. Sequences from globular proteins – shown on the left in blue as for construct $DHD_{29}C$ – tend to compact on themselves to various degrees within the protected cavity between ribosome and translocon, shortening the tether connecting the H-segment to the ribosome and limiting its dispersion into the membrane. Highly flexible and hydrophilic sequences like glycine–serine-rich polypeptides – shown on the right in green as for construct DxD_{29}[GS100]C – easily diffuse into the cytosol and thus facilitate H-segment movement into the bilayer and reduce the probability of return to the pore as indicated by the arrows.

these segments from the negatively charged ribosomal RNA away from the ribosome might facilitate membrane insertion of the H-segment. However, positively charged residues do not show a significant effect on their own and do not clearly neutralize negative charges in a mixed sequence, as one would have expected. Positive charges could have additional effects, preferentially hindering translocation and thereby concealing an effect on integration. It has been shown that clusters of positive charges inhibit polypeptide

Table 1. Context-dependent integration of natural hydrophobic sequences

Protein	Sequence X	Wild-type behavior	Predicted ΔG_{app} (kcal/mol) [predicted integration]	Integration with DXCC	Integration with DXD_{29}[GS100]C
CD99	VIPGIVGAVVVAVAGAISSFIAY	Integration	1.0 [16%]	3%±1	92%±7
STIM2	WMKDFILTVSIVIGVGGCWFAYT	Integration	−0.2 [58%]	44%±3	91%±3
α1-Antiproteinase F	IFFSPVSIALAFAMLSLGA	Translocation	0.9 [18%]	3%±1	84%±6
Angiotensinogen (1)	ATVLSPTAVFGTLASLYLGAL	Translocation	1.5 [8%]	0%±2	47%±6
Angiotensinogen (2)	AQLLLSTVVGVFTAPGLHL	Translocation	2.2 [3%]	0%±2	4%±2

The transmembrane sequences of human CD99 and STIM2, as well as relatively hydrophobic segments from the secretory proteins α1-antiproteinase F and angiotensinogen (two sequences labeled 1 and 2) are listed. ΔG_{app} was predicted using the ΔG prediction server http://dgpred.cbr.su.se based on Hessa et al. (2005, 2007). The extent of integration as a percentage of the total product was determined as shown in Fig. 7B (mean±s.d.; n=3).

translocation at least *in vitro* (Fujita et al., 2012) and thus might neutralize an additional inhibitory effect on integration.

Finally, a longer oligoproline peptide is likely to assume its sterically preferred conformation as a polyproline type II helix, which is very extended with 3.1-Å spacing per residue. Similarly, its effect could counteract compaction of the peptide or extend tether length, and thus favor membrane integration versus translocation as is experimentally observed.

Molecular dynamics simulations, although limited to rather short time intervals, suggest that H-segments are rather stably positioned at the interface between the lipid and pore (Demirci et al., 2013). The hydrophobicity of the H-segment might determine how easily it can be pulled through Brownian motion of the nascent chain from its interface position either into the ER lumen or into the lipid bilayer. We propose that the conformational state of the C-terminal sequence (exemplified by compact collapsed states versus freely diffusing flexible chain) differentially affects the probabilities of H-segment transport in these two directions.

Even though the exact mechanism of how different downstream sequences affect the integration process remains to be established, the fact that they do has important consequences. It explains the limitations of transmembrane domain prediction based on the biological scale (or any other hydrophobicity scale). Marginally hydrophobic sequences, as they evolve with their native contexts to be efficiently integrated or translocated, end up having different hydrophobicities to accomplish the same purpose. We confirmed this by placing natural sequences into two reporters that are either restrictive or permissive of membrane integration (DxCC or DxD_{29}[GS100]C, respectively). Depending on the type of downstream polypeptides, normally translocated segments could be induced to integrate to a significant extent and, conversely, normally integrated transmembrane sequences could be mostly translocated.

Our results highlight that the decision between translocation and integration is more complex than simple partitioning between two phases. H-segments are depleted from the thermodynamic equilibrium in either direction over time. Features that affect removal of the H-segment away from the translocon into either the ER lumen or the lipid bilayer deplete the equilibrium differentially and thus distort the original distribution.

MATERIALS AND METHODS

Yeast strain and model proteins

Yeast strain RSY1293 (*mata*, *ura3-1*, *leu2-3,-112*, *his3-11,15*, *trp1-1*, *ade2-1*, *can1-100*, *sec61::HIS3*, [pDQ1]) was used (Pilon et al., 1997). The coding sequences of dipeptidyl aminopeptidase B (DPAPB; *DAP2*) and carboxypeptidase Y (CPY; *PRC1*) were modified by adding a C-terminal triple-HA tag, inserting different H-segments (as introduced by Hessa et al., 2005), and exchanging various parts between them as shown in Fig. 1, and expressed using the plasmid pRS426 (*URA3 2 μ*) with a glyceraldehyde-3-phosphate dehydrogenase promoter. DNA constructs were generated using PCR to introduce restriction sites for fragment exchange. Short segments encompassing synonymous codons or encoding generic peptide sequences were inserted as annealed complementary oligonucleotides with appropriate protruding ends. The glycine–serine repeat sequence GS100 was produced by amplification of 85 glycine–serine codons of 87XH1-4-L13[460] (Kocik et al., 2012) and fusion with the glycine–serine insert of construct GS (Fig. 5). Complete protein sequences of the model proteins analyzed in this study are provided in Table S1.

From the protein sequences analyzed previously by Schow et al. (2011, figure 8 in that article), two natural stop-transfer sequences with high predicted ΔG_{app} values for membrane insertion (human CD99 and STIM2) were selected. Similarly, three sequences from within translocated domains of secretory proteins (rabbit α1-antiproteinase F and human angiotensinogen) with relatively low predicted ΔG_{app} values were chosen. Pairs of oligonucleotide primers were designed – a sense primer and an antisense primer comprising ∼40 nucleotides corresponding to the C-terminal and N-terminal sequences, respectively, of these peptides of interest (denoted by x) followed by sequences matching GPGGWWE and STFGGPG (the C- and N-terminal flanking sequence of the H-segment in the template construct DHCC), respectively. The primer pairs contain a complementary overlap of ∼20 nucleotides that was used to fuse the PCR products obtained by amplification of the 5′ portion of DHCC using the antisense primer and of the 3′ portion using the sense primer, each in combination with an appropriate vector primer, by overlap extension using the two vector primers. In the resulting construct encoding DXCC, the segment downstream of X was exchanged using restriction sites to generate DXD_{29}[GS100]C.

Labeling and immunoprecipitation

Overnight yeast cultures were diluted and grown to OD_{600} ∼1. Equivalents of 1.5 OD were resuspended in 200 µl of synthetic defined (SD) medium lacking uracil, incubated for 15 min at 30°C and labeled for 5 min (which is too short to allow CPY constructs to be affected by potential modifications at the Golgi) with 150 µCi/ml [^{35}S]methionine–cysteine (Hartmann Analytics). At 4°C, cells were supplemented with 10 mM sodium azide, pelleted, resuspended in 50 mM Tris-HCl, pH 7.5, 5 mM EDTA, 1 mM phenylmethylsulfonyl fluoride and lysed with glass beads for 7 min in a bead-beater, supplemented with 1% SDS and heated at 95°C for 5 min. After removal of cell remnants by centrifuging for 10 min in a microfuge, the supernatant was used for immunoprecipitation with mouse monoclonal anti-HA antibodies (HA.11, clone 16B12; 1:1000) and protein-A–Sepharose and analysis by SDS-gel electrophoresis and autoradiography. Signals were quantified using a phosphorimager. Each replicate was performed using cells from separate transformation events. Each result is displayed individually in the figures, unless stated otherwise. For deglycosylation, immune complexes were boiled in 50 mM sodium citrate, pH 6 with 1% SDS, and incubated with 1 mU endo-β-D-*N*-acetyl glucosaminidase H for 1 h at 37°C before gel electrophoresis.

Acknowledgements

We thank Salvatore Assenza, Simon Bernèche, Flavienne Bieri, Marco Janoschke, Gabriel Studer and Mirjam Zimmermann for valuable discussions.

Competing interests

The authors declare no competing or financial interests.

Author contributions

T.J. and M.S. conceived and designed the experiments. T.J. performed the experiments. T.J. and M.S. analyzed the data and prepared the manuscript.

Funding

This work was supported by Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (the Swiss National Science Foundation) (31003A-162643).

Supplementary information

Supplementary information available online at
http://jcs.biologists.org/lookup/doi/10.1242/jcs.194472.supplemental

References

Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G. and Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science* **278**, 2123-2126.

Burmann, B. M., Wang, C. and Hiller, S. (2013). Conformation and dynamics of the periplasmic membrane-protein–chaperone complexes OmpX–Skp and tOmpA–Skp. *Nat. Struct. Mol. Biol.* **20**, 1265-1272.

Capponi, S., Heyden, M., Bondar, A.-N., Tobias, D. J. and White, S. H. (2015). Anomalous behavior of water inside the SecY translocon. *Proc. Natl. Acad. Sci. USA* **112**, 9016-9021.

Conti, B. J., Elferich, J., Yang, Z., Shinde, U. and Skach, W. R. (2014). Cotranslational folding inhibits translocation from within the ribosome–Sec61 translocon complex. *Nat. Struct. Mol. Biol.* **21**, 228-235.

Demirci, E., Junne, T., Baday, S., Bernèche, S. and Spiess, M. (2013). Functional asymmetry within the Sec61p translocon. *Proc. Natl. Acad. Sci. USA* **110**, 18856-18861.

Egea, P. F. and Stroud, R. M. (2010). Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc. Natl. Acad. Sci. USA* **107**, 17182-17187.
Fujita, H., Yamagishi, M., Kida, Y. and Sakaguchi, M. (2012). Positive charges on the translocating polypeptide chain arrest movement through the translocon. *J. Cell Sci.* **124**, 4184-4193.
Gogala, M., Becker, T., Beatrix, B., Armache, J.-P., Barrio-Garcia, C., Berninghausen, O. and Beckmann, R. (2014). Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* **506**, 107-110.
Gumbart, J. C., Teo, I., Roux, B. and Schulten, K. (2013). Reconciling the roles of kinetic and thermodynamic factors in membrane–protein insertion. *J. Am. Chem. Soc.* **135**, 2291-2297.
Hedin, L. E., Öjemalm, K., Bernsel, A., Hennerdal, A., Illergård, K., Enquist, K., Kauko, A., Cristobal, S., von Heijne, G., Lerch-Bader, M. et al. (2010). Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J. Mol. Biol.* **396**, 221-229.
Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I. M., White, S. H. and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-381.
Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H. and von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-1030.
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.
Jan, C. H., Williams, C. C. and Weissman, J. S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521-1257521.
Junne, T., Kocik, L. and Spiess, M. (2010). The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Mol. Biol. Cell* **21**, 1662-1670.
Kida, Y., Ishihara, Y., Fujita, H., Onishi, Y. and Sakaguchi, M. (2016). Stability and flexibility of marginally hydrophobic-segment stalling at the endoplasmic reticulum translocon. *Mol. Biol. Cell* **27**, 930-940.
Kocik, L., Junne, T. and Spiess, M. (2012). Orientation of internal signal-anchor sequences at the Sec61 translocon. *J. Mol. Biol.* **424**, 368-378.
Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319-324.
MacCallum, J. L. and Tieleman, D. P. (2011). Hydrophobicity scales: a thermodynamic looking glass into lipid–protein interactions. *Trends Biochem. Sci.* **36**, 653-662.
Ménétret, J.-F., Neuhof, A., Morgan, D. G., Plath, K., Radermacher, M., Rapoport, T. A. and Akey, C. W. (2000). The structure of ribosome-channel complexes engaged in protein translocation. *Mol. Cell* **6**, 1219-1232.
Ng, D. T., Brown, J. D. and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* **134**, 269-278.
Nilsson, O. B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., Müller-Lucks, A., Trovato, F., Puglisi, J. D., O'Brien, E. P. et al. (2015). Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep.* **12**, 1533-1540.
Öjemalm, K., Higuchi, T., Jiang, Y., Langel, U., Nilsson, I., White, S. H., Suga, H. and von Heijne, G. (2011). Apolar surface area determines the efficiency of translocon-mediated membrane-protein integration into the endoplasmic reticulum. *Proc. Natl. Acad. Sci. USA* **108**, E359-E364.
Öjemalm, K., Botelho, S. C., Stüdle, C. and von Heijne, G. (2013). Quantitative analysis of SecYEG-mediated insertion of transmembrane α-helices into the bacterial inner membrane. *J. Mol. Biol.* **425**, 2813-2822.
Onishi, Y., Yamagishi, M., Imai, K., Fujita, H., Kida, Y. and Sakaguchi, M. (2013). Stop-and-move of a marginally hydrophobic segment translocating across the endoplasmic reticulum membrane. *J. Mol. Biol.* **425**, 3205-3216.
Park, E. and Rapoport, T. A. (2011). Preserving the membrane barrier for small molecules during bacterial protein translocation. *Nature* **473**, 239-242.
Park, E. and Rapoport, T. A. (2012). Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu. Rev. Biophys.* **41**, 21-40.
Park, E., Ménétret, J.-F., Gumbart, J. C., Ludtke, S. J., Li, W., Whynot, A., Rapoport, T. A. and Akey, C. W. (2014). Structure of the SecY channel during initiation of protein translocation. *Nature* **506**, 102-106.
Pilon, M., Schekman, R. and Römisch, K. (1997). Sec61p mediates export of a misfolded secretory protein from the endoplasmic reticulum to the cytosol for degradation. *EMBO J.* **16**, 4540-4548.
Rodnina, M. V. and Wintermeyer, W. (2016). Protein elongation, co-translational folding and targeting. *J. Mol. Biol.* **428**, 2165-2185.
Schow, E. V., Freites, J. A., Cheng, P., Bernsel, A., von Heijne, G., White, S. H. and Tobias, D. J. (2011). Arginine in membranes: the connection between molecular dynamics simulations and translocon-mediated insertion experiments. *J. Membr. Biol.* **239**, 35-48.
Shao, S. and Hegde, R. S. (2011). Membrane protein insertion at the endoplasmic reticulum. *Annu. Rev. Cell Dev. Biol.* **27**, 25-56.
Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H. and Wright, F. (1988). Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. *Nucleic Acids Res.* **16**, 8207-8211.
Tsukazaki, T., Mori, H., Fukai, S., Ishitani, R., Mori, T., Dohmae, N., Perederina, A., Sugita, Y., Vassylyev, D. G., Ito, K. et al. (2008). Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* **455**, 988-991.
Ulmschneider, M. B., Ulmschneider, J. P., Schiller, N., Wallace, B. A., von Heijne, G. and White, S. H. (2014). Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion. *Nat. Commun.* **5**, 4863.
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.
Voorhees, R. M. and Hegde, R. S. (2016). Structure of the Sec61 channel opened by a signal sequence. *Science* **351**, 88-91.