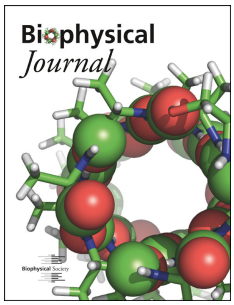
Accepted Manuscript



Forces on nascent polypeptides during membrane insertion and translocation via the Sec translocon

M.J.M. Niesen, A. Müller-Lucks, R. Hedman, G. von Heijne, T.F. Miller, III

PII: S0006-3495(18)31115-9

DOI: 10.1016/j.bpj.2018.10.002

Reference: BPJ 9207

To appear in: *Biophysical Journal*

Received Date: 16 August 2018

Accepted Date: 2 October 2018

Please cite this article as: Niesen MJM, Müller-Lucks A, Hedman R, von Heijne G, Miller III TF, Forces on nascent polypeptides during membrane insertion and translocation via the Sec translocon, *Biophysical Journal* (2018), doi: https://doi.org/10.1016/j.bpj.2018.10.002.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Forces on nascent polypeptides during membrane insertion and translocation via the Sec translocon

M.J.M. Niesen[1], A. Müller-Lucks[2], R. Hedman[2], G. von Heijne[2], and T.F. Miller III[1]

[1]California Institute of Technology, Division of Chemistry and Chemical Engineering, Pasadena, CA 91125, USA
[2]Stockholm University, Department of Biochemistry and Biophysics, SE-106 91 Stockholm, Sweden

[*]tfm@caltech.edu

Abstract

During ribosomal translation, nascent polypeptide chains (NCs) undergo a variety of physical processes that determine their fate in the cell. This study utilizes a combination of arrest peptide (AP) experiments and coarse-grained molecular dynamics (CGMD) to measure and elucidate the molecular origins of forces that are exerted on NCs during co-translational membrane insertion and translocation via the Sec translocon. The approach enables deconvolution of force contributions from NC-translocon and NC-ribosome interactions, membrane partitioning, and electrostatic coupling to the membrane potential. In particular, we show that forces due to NC-lipid interactions provide a read-out of conformational changes in the Sec translocon, demonstrating that lateral gate opening only occurs when a sufficiently hydrophobic segment of NC residues reaches the translocon. The combination of experiment and theory introduced here provides a detailed picture of the molecular interactions and conformational changes during ribosomal translation that govern protein biogenesis.

Introduction

Co-translational protein biogenesis is tightly regulated to ensure that newly synthesized proteins are correctly targeted and folded within the cellular environment. Throughout this process, a nascent polypeptide chain (NC) is exposed to a complex range of forces and interactions, the study of which is complicated by the crowded, stochastic nature of the cell. The current work combines arrest peptide (AP) experiments and simulation to connect the pulling forces experienced by a NC to the underlying molecular processes associated with membrane integration and translocation via the Sec translocon.

Most membrane proteins and many secretory proteins are targeted to the Sec translocon during ribosomal translation (reviewed in Refs. (1–6)). The translocon is a protein-conducting transmembrane channel that is ubiquitous across all kingdoms of life. The ribosome docks onto the cytosolic opening of the translocon, cotranslationally inserting the NC into the translocon channel. The central pore of the translocon facilitates the translocation of hydrophilic loops across the cell membrane, and a lateral gate enables passage of transmembrane domains into the cell membrane (7). The components of the translocon have been characterized structurally (7–11) and biochemically (12–16), and extensive work has focused on the role of the translocon on regulating NC translocation versus membrane integration (17–23). Nonetheless, open questions remain about the nature of the transient interactions between the NC and the translocon channel interior and membrane environment.

AP experiments probe the co-translational forces that act on the NC, providing a signature of the underlying interactions between the NC and the translocon during cotranslational membrane integration. Once an AP is synthesized by the ribosome, it stalls further NC translation (24); the stall is released with a rate that is dependent on the pulling forces that are experienced by the NC (25). APs are used in nature to control NC translation (24) and have recently been applied to gain insight into physical processes such as integration into the cell membrane (26), co-translational folding (27, 28), and electrostatic interactions (29). In this study, we use AP experiments with engineered NCs to measure the forces exerted during membrane integration and translocation. To complement the AP experiments, simulations are performed using a recently developed structurally detailed coarse-grained molecular dynamics (CGMD) model that provides nm-lengthscale resolution (30), allowing for the direct computation of the NC dynamics, interactions, and resulting pulling forces. The

combination of simulation and experiment elucidates the diverse interactions and forces acting on the NC at specific lengths during translation.

Materials and Methods

Experimental methods

All plasmids used in this work were designed as in (26) i.e., "H segment" of different amino acid composition and flanked by GPGG….GGPG segments were inserted into the periplasmic P2 domain of the E. coli inner membrane protein LepB. The 8-residue "arrest peptide" HAPIRGSP from the *Mannheimia succiniciproducens* SecM protein (26) was inserted at varying distances downstream of the C-terminal end of the H segment, leaving a 23-reside C-terminal tail after the AP to ensure that arrested and full-length protein products were of sufficiently different molecular weight to allow separation by SDS-PAGE. Constructs with poly-leucine H segment of composition 5L, 8L and 10L were expressed in E. coli, and analyzed by pulse-labeling (2 min pulse, no chase), immunoprecipitation, and SDS-PAGE as described in ref (26). See Figure S1 for sequences and example SDS-PAGE gels. The fraction full-length protein, f_{FL}, was calculated as $f_{\mathrm{FL}} = I_{\mathrm{FL}}/(I_{\mathrm{FL}} + I_{\mathrm{A}})$, where I_{FL} and I_{A} are the intensities of the bands corresponding to, respectively, the full-length and arrested forms of the protein on the SDS-PAGE gel. Experiments were performed in at least three independent trials, reported results are averages with the error bars indicating the standard error of the mean.

Computational methods

We employ a previously developed CGMD approach (21, 22, 30, 31) to measure co-translational forces acting on a NC during Sec-facilitated integration into the lipid membrane or translocation across the lipid membrane. CGMD calculates the dynamics of the NC at a μs time-resolution and a nm length-resolution, with explicit ribosomal translation and lateral gating of the Sec translocon (30). Residue-specific interactions between the translocon and NC have been parameterized using over 200 μs of residue-based coarse grain simulations (MARTINI FF). The lipid membrane and water are included implicitly using a position dependent energy function based on the Wimley-White octanol/water scale (32), and the effect of counter-ions is accounted for using Debye-Hückel electrostatic interactions with a screening length corresponding to an ionic strength of 150 mM. Despite the high level of coarse-graining employed, the CGMD method has been thoroughly tested and demonstrated to capture experimental integration probabilities and topology distributions of single transmembrane domains (30), such as the ones investigated in the current study. The only modification of the CGMD method from previous work (30) is the inclusion of a membrane electrostatic potential, as described below. All CGMD simulations are performed using in-house software, as previously detailed (30).

CGMD Method

For a given NC, the protein sequence is mapped into the coarse-grained (CG) representation (Figure 1a, top), with one CG bead representing three amino-acid residues (30). NC beads interact with CG beads representing the translocon and ribosome via pair-wise interaction that depend on the charge and hydrophobicity of the NC bead, and the charge and location of the translocon bead. The NC is treated as an ideal polymer chain with excluded volume interaction and is fully dynamic during the CGMD simulations. The full geometric coordinates for the CG representation of the ribosome, translocon, and membrane environment are provided in Ref. (30). The interaction between the CG beads and the lipid membrane is accounted for by a water-lipid transfer free energy assigned to each CG bead, derived from the Wimley-White water-octanol transfer free energy (32) of the underlying amino-acid residues. Overdamped Langevin dynamics for the CG beads is simulated with an isotropic diffusion coefficient of $253.0\ \mathrm{nm}^2/\mathrm{s}$ and a timestep of 300 ns. The translocon occupies two discrete conformations (closed and open), with stochastic transitions between the conformations governed by the free-energy difference between the two states as a function of the NC configuration (30). The ribosome beads are in a single discrete conformation (coordinates in Ref. (30)).

To simulate AP stalling, translation is halted after a given number of CG beads have been translated. The length *L* of the NC upon stalling corresponds to the number of amino-acid residues counted from the C-terminus of the H segment to the C-terminus of the AP (Figure 1a). The instantaneous pulling force on the NC, F_z, is calculated as the component of the force along the translocon channel axis that acts on the most C-terminal bead in the CGMD; this C-terminal bead is held fixed to mimic translation arrest (Figure 1a, top). The value of *L* in the CGMD is then defined as the number of residues that are explicitly represented as CG beads plus a constant correction of 27 residues, accounting for the amino-acid residues between the most C-terminal bead in the CGMD and the end of the AP.

Simulations without residue-specific interactions (Figures 1g and 2d) are performed in exactly the same manner, except with modified interactions for the CG beads. Instead of having interaction parameters as based on the underlying amino-acid sequence (30), parameters are set to a constant value irrespective of the amino-acid sequence. Specifically, for simulations without specific NC-translocon interactions, all NC beads interact with the translocon channel using the parameters for a QQQ tri-peptide ($\lambda_c = 0.75$ and $\lambda_o = 0.78$); and for simulations without residue-specific lipid interactions, all NC beads employ a water-lipid transfer free energy of 5ϵ.

Inclusion of the membrane potential

To investigate coupling of the charged residues in the NC to the membrane electrostatic potential in *E. coli* (Figure 3), a membrane potential is included in the CG model. Following previous work (29), the potential energy function, U_{mp}, associated with the additive interaction of the membrane potential with each charged CG bead, *i*, is described using

$$U_{mp}(z_i; q_i, \Delta\Psi) = q_i \frac{\Delta\Psi}{1+e^{\kappa z_i}}, \quad (1)$$

where z_i is the position of the bead along the channel axis, q_i is the charge of the bead, $\kappa = 1.6\sigma^{-1}$ (0.2Å^{-1}) is the reciprocal lengthscale of the membrane potential drop, and $\Delta\Psi$ is the value of the maximum potential drop of -3.74ϵ (-100 mV) used for the results in Figure 3b, g, and h or 0ϵ (0 mV) used elsewhere.

Calculation of fraction full-length protein

To compare with the fraction full length, f_{FL}, determined via AP experiments, the pulling forces calculated from the CGMD must be converted to a prediction of f_{FL}. Following previous work (25, 29), the AP-stalled ribosome is assumed to restart translation with a force-dependent rate, k_{FL}, which is calculated assuming Bell's model,

$$k_{FL} = k_0 \langle e^{\beta \Delta x^{\ddagger} F_z} \rangle \quad (2)$$

where k_0 is the rate without an applied force, $\beta = 1/k_B T$, $\Delta x^{\ddagger}$ is an AP-dependent characteristic distance, F_z is the previously defined instantaneous pulling force on the NC obtained directly from CGMD, and $\langle .. \rangle$ indicates ensemble averaging over the CGMD trajectory data. The employed value of $\Delta x^{\ddagger}$=0.5 nm for all sequences is based on previous work (25, 29); the value for k_0 is described below.

The ensemble average in Eq. 2 is obtained from CGMD sampling trajectories, with the NC stalled at a given length *L*. For each NC sequence and each value of *L*, the ensemble average is obtained by averaging 100 independent sampling trajectories of length 15 s in time. Note that the time required for equilibration of the ensemble average is much shorter than the total time required for AP stall breaking and full-length protein translation.

The force dependent rate for breaking the translation arrest, k_{FL}, is then used to calculate the experimentally observable fraction of full-length protein, f_{FL}, using

$$f_{FL} = 1 - \exp[-k_{FL}t] = 1 - \exp[-k_0 t \langle e^{\beta \Delta x^{\ddagger} F_z} \rangle]. \quad (3)$$

The only undetermined parameter in this equation is $(k_0 t)$, which depends on the details of the AP, the background pulling-force in the experimental system, and the observation time. We determine a value for $(k_0 t)$ in this work by fitting the calculated baseline of f_{FL} to that observed in experiment (using the data in Figure 1c for $L >= 51$). We emphasize that this fit is done once, yielding a value of $(k_0 t) = 3.7*10^{-12}$ that is held fixed for all other reported results.

The CG mapping of three amino-acid residues to a single CG bead allows for three possible frame-shifts between the amino-acid and CG sequences. At each reported value of *L*, f_{FL} is separately calculated for all three possible frame-shifts (with associated lengths $L-1$, *L*, and $L+1$) and the result is averaged. When comparing to experimental data at a given *L*, we refer to the CGMD results for which *L* equals the nearest multiple of three (i.e., experiments at $L = 28$ are compared to CGMD simulations for which the middle frame is $L = 27$). Altogether, 4500 s of CGMD simulation time is performed for each reported value of f_{FL}.

The main assumption employed in Eq. 3 is a first-order kinetic scheme in which AP stalling is overcome with a force-dependent rate k_{FL}, with no off-target pathways. We also considered a slightly more complex kinetic scheme in which the stalled ribosomes experience an additional degradation pathway with a fixed rate; this more complex scheme led to no substantial changes in the results.

Results

We consider a series of NC substrates to validate the combined simulation and experimental approach and to investigate the molecular interactions that govern co-translational NC integration and translocation. All NC substrates described in this work utilize a well-established model system for which CGMD has been previously validated to correctly capture experimental integration probabilities (30), with an engineered domain (H segment) incorporated into the leader peptidase (LepB) protein (Figure 1a, bottom) (17, 26, 29). We study the forces exerted on the NC during *(i)* the integration of a model transmembrane domain, *(ii)* translocation and integration of non-spanning hydrophobic segments, and *(iii)* the translocation of model hydrophilic and charged domains. CGMD simulations are compared with both previously published (26, 29) and new AP experimental data, providing validation for the computational method and yielding insight into the interactions that govern co-translational NC integration and translocation via the Sec translocon.

Forces on integrating hydrophobic segments, and the mechanism of the biphasic pulling force

We begin by investigating the forces of co-translational integration, with the H segment comprised of a model transmembrane domain (Figure 1a, bottom). Previously published AP experiments (26) reveal the points during translation at which increased pulling forces are exerted on the NC (Figure 1b). In these experiments, an AP is inserted downstream of the H segment, and the number of residues between the C-terminal end of the AP and the C-terminal end of the H segment, *L*, is varied (Figure 1a, bottom). The H segment has a fixed length of 19 residues that are either leucine or alanine, and various H segment compositions are tested. The fraction of full-length protein, f_{FL}, is experimentally quantified as a proxy for the pulling force acting on the AP, with greater forces leading to increased f_{FL} (see Materials and Methods) (25, 26). Two peaks in the f_{FL} profile (henceforth called pulling-force profile) are observed at $L = 28$ and $L = 39$ (Figure 1b). For comparison with the experiment, CGMD simulations are used to calculate the co-translational forces

exerted on the NC for the same sequences as those experimentally tested. The protein sequences are mapped into a coarse-grained representation (Figure 1a, top) and forces acting on the end of the NC that is tethered to the ribosome are directly calculated (Figure 1a, white arrow). Calculated forces are converted to f_{FL} assuming Bell's model to relate force to the force-dependent rate of stall release (see Materials and Methods) (25). The CGMD successfully captures peaks in f_{FL} at the same values of *L* (Figure 1c, dashed vertical lines) as previously observed experimentally. Consistent with the experiment, the peaks in f_{FL} are dependent on the number of leucine-residues, $nLeu$, in the H segment (Figure 1c).

To identify the physical processes that underlie the observed peaks in both the experimental and simulated pulling-force profiles, we analyze the CGMD trajectories. A characteristic MD configuration for $L = 28$ is shown in Figure 1d. At this NC length, the N-terminus of the H segment first reaches the interior of the translocon, allowing for attractive, residue-specific interactions; this interpretation of the first peak in the pulling-force profile is consistent with experimental data on the effects of point mutations in the H segment (26). In Figure 1e, a characteristic MD configuration for $L = 39$ is shown. At this point, the N-terminus of the H segment is first able to partition from the interior of the translocon channel into the interior of the lipid membrane via the open lateral gate; this interpretation of the second peak is again consistent with available experimental mutagenesis data (26). Finally, in Figure 1f, a characteristic configuration associated with larger values of *L* is presented; at these NC lengths, the H segment has completed integration into the lipid membrane and the NC is no longer under tension.

The molecular origin of the observed peaks is further confirmed by additional CGMD simulations with modified interactions. Considering first the H segment with nine leucine residues, Figure 1g shows the pulling-force profile calculated from simulations for which (purple) the residue-specificity of the interactions between the NC and the translocon are eliminated or (teal) the residue-specificity of the water-lipid transfer free energies are eliminated (see Materials and Methods). The simulations without residue-specific interactions between the NC and the translocon channel do not display the pulling-force peak at $L = 28$, confirming that the first peak reports on the specific interactions between the H segment and the residues of the translocon interior. Similarly, the simulations without residue-specific water-lipid transfer free energies do not display the pulling-force peak at $L = 39$, confirming that this second peak arises from the partitioning of the H segment from the translocon interior into the membrane interior. Similar results are obtained for all tested H segments, with the pulling-force profiles consistently comprised of two underlying peaks (Figure S3). These results provide clear validation of the CGMD simulations in comparison to experiment, as well as direct evidence of the physical origins of the observed features in the pulling-force profiles.

Finally, to examine the dependence of the pulling forces on the sequence of the H segment, we examine the height of the peaks in the pulling-force profiles as a function of the number of leucine residues in the H segments, $nLeu$. For each value of $nLeu$, we calculate the pulling-force profile either using simulations with the non-specific lipid interactions or using simulations with the non-specific channel interactions(Figure 1g). Given that these modified interactions lead to pulling-force profiles with only a single peak, we can unambiguously determine the peak height, max(f_{FL}), for the peak near $L = 28$ (Figure 1h, teal) and the peak near $L = 39$ (Figure 1h, purple). The figure shows that with increasing $nLeu$, the peak associated with the NC-translocon interactions (teal) remains relatively unchanged, while the peak associated with NC-lipid interactions increases. Experimentally determined peak heights are shown as dashed lines for comparison ($L = 28$ in teal and $L = 39$ in purple). Qualitative agreement between simulation and experiment is obtained in these results. In particular, both simulation and experiment predict that the peak at $L = 39$ rises sharply with respect to $nLeu$, whereas the peak at $L = 28$ does not. The rightward shift in the simulation curve for the peak at $L = 39$ arises from an underestimation of the affinity of the hydrophobic TMD with the membrane interior, which is a source of error that has previously been identified and noted in the context of stop-transfer experiment (30). As shown in Figure S2, if the results in Figure 1h are scaled

by a constant factor to account for this source of error in the CG model, then the agreement between theory and experiment becomes quantitative.

Forces on hydrophobic segments of variable length

To examine the relation between the size of the hydrophobic segment and the forces exerted on NC, we investigate hydrophobic poly-leucine H segments of varying length using both CGMD simulations and new AP experiments. This "variable-length assay" allows for comparison of short hydrophobic segments (illustrated in Figure 2a) that primarily undergo membrane translocation versus longer hydrophobic segments that primarily undergo membrane integration; it contrasts with the "fixed-length assay" from the previous section in which all H segments were the same length and sufficiently long to span the membrane.

Figures 2b and c present pulling-force profiles for the poly-leucine H segments of various lengths, *n*, obtained using CGMD and experiments, respectively (see Figure S1 for examples of SDS-PAGE gels). Results are plotted as a function of the length of the NC chain from the C-terminus of the AP to the N-terminus of the variable-length H segment, $L + n$, where *L* is defined as before and *n* is the fixed length of the H segment associated with each curve (Figure 2a, bottom). With this choice for the x-axis, the expected position for the peaks associated with the NC-translocon interactions and the NC-lipid interactions from the fixed-length assay in the previous section (teal and purple vertical lines, respectively) are independent of the variable length of the H segment. Both simulation and experiment show a single broad peak in the pulling-force profile (Figures 2b and c), compared to the two distinct peaks observed for model transmembrane domains in Figure 1. With increasing length of the hydrophobic segment, the observed single peak broadens and increases in height.

To deconvolute the role of NC-translocon versus NC-lipid interactions in Figures 2b and c, CGMD simulations with modified interactions are performed, as before. Considering first the H segment with eight leucine residues, Figure 2d contrasts the results obtained using non-specific lipid interactions (teal) versus non-specific translocon interactions (purple). Consistent with the fixed-length assay (Figure 1g), the simulations with non-specific lipid interactions (teal) yield a peak at the expected NC length due to residue-specific interactions between the NC and the translocon. However, the simulations in Figure 2d with non-specific translocon interactions (purple) yield a peak at shorter NC lengths than expected from the fixed-length assay. Figure S4 presents the analog of Figure 2d for the NC sequences with different H segment length and demonstrates consistent results. To further validate the deconvolution of the pulling-force profile in Figure 2d into two distinct peaks with different physical origins, Figure 2e presents the calculated peak heights across the various H segment lengths as a function of the number of leucine residues, $nLeu$, revealing a trend that is consistent with the fixed-length assay (Figure 1h); specifically, the peak height associated with the NC-translocon interactions (teal) remains unchanged, while the peak height associated with NC-lipid interactions increases. Finally, Figure 2f contrasts the position of the peak associated with NC-lipid interactions from the variable-length assay (purple) versus the corresponding results from the fixed-length assay from the previous section (black). In contrast to the fixed-length assay, for which the NC-lipid peak position is relatively invariant with respect to the increasing number of leucine residues, the results from the variable-length assay find that the peak position steadily increases with the number of leucine residues.

The contrasting behavior of the fixed- versus variable-length assays in Figure 2f provides insight into the mechanism by which hydrophobic portions of the NC sample the lipid membrane, an issue that has been the focus of considerable discussion (4, 33–37). Observation of the peak in the pulling-force profile associated with the NC-lipid interaction requires that the lateral gate of the translocon be in the open conformation, to allow for contact of the NC with the lipid environment. Previous work has suggested that opening of the translocon lateral gate is stabilized when hydrophobic NC residues reside in the translocon channel interior (36, 37). Note that for a given number of leucine residues, $nLeu$, the variable-length assay prescribes that those hydrophobic residues appear consecutively in the H segment sequence, whereas the fixed-length assay dilutes the hydrophobic leucine residues over a total of 19 residues. If a threshold number of hydrophobic residues is needed to stabilize the free-energy of opening of the translocon lateral gate (37), then the variable-length assay will reach that

threshold at shorter lengths of the NC than the fixed-length assay. Consistent with this mechanism for lateral gating, the lateral gate in the CGMD is found to open at shorter NC lengths for the H segments used in the variable-length assay (Figure S5). This explains why the onset of the lipid-interaction peak for small values of $nLeu$ appears at shorter NC lengths in the variable-length assay than in the fixed-length assay. The experimental and simulation results presented in Figure 2 thus provide evidence in support of the hydrophobic stabilization of the open-state state for the lateral gate of the translocon (37), as well as the prediction that a sufficiently hydrophobic H segment samples the membrane environment across the lateral gate as it passes down the axis of the translocon channel (31, 38); it is likewise consistent with the "sliding" model for transmembrane helix integration, which posits that hydrophobic segments in the NC slide along the lateral gate of the translocon, with one side exposed to lipid (4).

Pulling forces on hydrophilic segments

Previously published AP experiments indicate that significant pulling forces act on hydrophilic segments of the NC during translocation in *E. coli* (29). These forces were attributed to the coupling of negatively charged residues on the NC with the membrane electrostatic potential. Here, we investigate this underlying mechanism using CGMD, finding broad agreement with the previously proposed mechanism, as well as identifying additional features in the pulling-force profiles that are attributed to interactions of charged residues in the NC with the charges on the ribosome and to changes in the NC solvation environment.

Figure 3a presents pulling-force profiles for three distinct hydrophilic H segments (D_5, Q_5, and K_5) obtained from previous AP experiments (29). Results are plotted as a function of $L+n$, with $n=5$ for all considered cases. A dominant peak at $L+n\approx 50$ in the pulling-force profile is observed for the negatively charged D_5 H segment (orange); the peak was found to reduce in magnitude, in a concentration dependent manner, when indole was added to the growth medium, which suggests that the peak is due to the membrane electrostatic potential (29). The corresponding peak in the pulling-force profile is not found for the charge-neutral H segment (Q_5, teal) nor the positively charged H segment (K_5, purple). Interestingly, the negatively charged H segment also exhibits a somewhat larger value for the pulling force at shorter NC lengths ($L+n<45$) in comparison to the other sequences; this feature was found to recur in a variety of negatively charged NC sequences (Figure S4a of Ref. (29)) although a mechanistic explanation was not provided.

To explore the mechanistic origin of these pulling-force features, CGMD pulling-force profiles are obtained using the same protein sequences, calculated with (Figure 3b) and without (Figure 3c) the approximate *E. coli* membrane potential ($\Delta\Psi=-100$ mV). For the CGMD results obtained in the presence of the membrane potential, the calculated pulling-force profiles are in good agreement with experiment, showing a dominant peak at $L+n\approx 50$ for the negatively charged H segment and no such features for the other H segments; additionally, it is seen that the pulling-force profile for the negatively charged H segment at short NC lengths ($L+n<45$) is increased in comparison to the other sequences, albeit to a greater degree than is observed experimentally. Figure 3c shows that removing the membrane potential in the CGMD simulations leaves all features of the pulling force unchanged, except for the dominant peak in the D_5 profile (orange) at $L+n\approx 50$. The membrane-potential sensitivity of the D_5 peak at $L+n\approx 50$ is in good agreement with the previous experimental studies of indole concentration dependence (29). From both CGMD and experiment, these results suggest that the dominant D_5 peak at $L+n\approx 50$ arises from coupling of the negatively charged residues to the membrane potential, whereas a different mechanism leads to the greater pulling forces on D_5 for shorter NC lengths in comparison to the other H segments.

To illustrate the interactions of the H segment at various NC lengths, Figures 3d-f present snapshots of the CGMD simulations for the sequence with the D_5 H segment (indicated in red beads). At short

NC lengths (part d), the H segment remains in close proximity to the ribosome, and it does not extend to the membrane interior regions where the membrane potential significantly varies; this is consistent with the finding that the membrane potential exhibits minimal forces on the H segment at these NC lengths. For NC lengths associated with the dominant peak in the D_5 pulling force profile at $L+n \approx 50$ (part e), the H segment extends to the membrane interior, where the membrane potential is most rapidly varying and will exert the largest pulling forces on the negatively charged residues. Finally, for even larger NC lengths (part f), the hydrophilic H segment is fully translocated across channel and is favorably solvated in the hydrophilic environment of the periplasm.

While the CGMD pulling-force profiles (Figures 3b and c) and the simulation snapshots at $L+n \approx 50$ (Figure 3e) are completely consistent with the interpretation that the dominant peak in the D_5 pulling force profile is due to the membrane potential, the CGMD simulations provide additional insight into the mechanistic features of the pulling-force profile at both shorter and longer NC lengths. In particular, at shorter NC lengths ($L+n<45$), it is observed in both CGMD and experiment that D_5 exhibits larger pulling forces than the other two H segments. The simulation snapshot at Figure 3d suggests that this feature in the D_5 profile may arise from repulsive electrostatic interactions between the negatively charged NC and the negatively charged ribosomal RNA. Figure 3g tests this hypothesis using the CGMD model, comparing the original D_5 pulling-force profile (orange) with that obtained in the absence of charges on the ribosome (black); clearly, the pulling force profile at short NC lengths in the absence of ribosomal charges returns to the baseline of the other H segments, supporting the hypothesis. This feature at short NC lengths is more accentuated in the CGMD than in the experiment, possibility due to the mean-field description in the CGMD of Mg^{2+} counter-ions near the ribosomal RNA that could lead to locally increased electrostatic screening.

Finally, we investigate the rise in the pulling force profile that is observed for all three hydrophilic H segments at long NC lengths ($L+n>60$), in both experiment (Figure 3a) and in the CGMD simulations (Figure 3b and even more clearly in Figure 3c). This feature is relatively independent of the charge of the hydrophilic segment; it was therefore not explained by previous work that focused on the role of charged residues (29). From the CGMD snapshot in Figure 3f, it is clear that the H segment has extended beyond the translocon interior at these NC lengths, such that the H segment has been replaced by C-terminal residues of the NC in the channel interior. This suggests a hypothesis in which the observed pulling forces at long NC lengths is due to the favorable free energy associated with transferring the hydrophilic H segment from the amphiphilic (or weakly hydrophilic) interior of the channel to the strongly hydrophilic environment of the periplasm. To test this hypothesis, Figure 3h compares the original D_5 pulling-force profile (orange) with that obtained by increasing the hydrophilicity of the residues in the C-terminal tail (C-tail) of the NC, thus counterbalancing the favorable free energy of transferring the D_5 from the channel interior to the periplasm with an unfavorable free energy of transferring the C-tail from the hydrophilic cytosol the channel interior. This alteration of the C-tail sequence leads to a reduction of the pulling force profile at long NC lengths, confirming that the the increased pulling forces at long NC lengths arise from the favorable free energy of transferring the hydrophilic H segment from the channel interior to the periplasm. Further experimental work on mutated sequences, such as the one employed here, would enable confirmation of this proposed driving force for hydrophilic domain translation.

Discussion

Membrane integration and protein translocation via the Sec translocon are critical steps in the biosynthesis and targeting of proteins in cells. The current work probes the fundamental interactions and conformational changes associated with these processes, using AP experiments to measure the pulling forces associated with interactions of the NC with its environment (25, 39). We present new AP experiments to obtain the pulling-force profiles as a function of NC length during translation, as

well as detailed analysis of these and previously reported AP experiments (26, 29) using long-timescale CGMD simulations (30) that allow for the direct computation of the pulling-force profiles.

Engineered NC sequences allow for the investigation of co-translational forces that act on transmembrane hydrophobic segments (Figure 1), non-membrane-spanning hydrophobic segments (Figure 2), and translocating hydrophilic segments (Figure 3). Despite their simplicity, these engineered sequences exhibit characteristic features of naturally occurring protein sequences. For each NC sequence, experimental pulling-force profiles are directly compared with those obtained using CGMD, validating the simulation method. Analysis of the microscopically detailed CGMD simulations provides insight into the mechanistic origins for the experimentally observed features pulling-force profiles. It is striking that the simple description of the co-translational integration machinery employed in CGMD, without system-specific modifications, can accurately capture the diverse experiments investigated here and in previous work (30). This strongly suggests that the key interactions and processes that underly the experimental observations are reliably captured by the employed CGMD approach.

Several conclusions emerge from this work. Firstly, a detailed analysis of the mechanistic origin of biphasic pulling-force profiles for transmembrane hydrophobic segments is provided (26); deconvolution of the pulling-force profiles using the CGMD (Figure 1) confirm that the peak at shorter NC lengths arises from translocon-NC interactions, while the peak at longer lengths is associated with NC-lipid interactions during membrane integration. Secondly, consideration of hydrophobic segments of variable length (Figure 2) elucidates the effect of hydrophobic segments on the conformational state of the translocon. The combined experimental and theoretical analysis confirms predictions that hydrophobic segments of the NC stabilize the open state of the tranlsocon lateral gate (37), as well as that even during translocation, sufficiently hydrophobic segments sample the membrane interior as they pass down the axis of translocon channel (4, 38). Finally, investigation of translocating hydrophilic segments (Figure 3) using CGMD confirms that the dominant peak in the pulling-force profile arises from the coupling of charged residues to the membrane electrostatic potential (29); however, the CGMD additionally suggest that previously unexplained features of the pulling-force profiles arise from electrostatic repulsion between negatively charged residues in the NC and the ribosomal RNA (at short NC lengths) and from forces associated with partitioning of hydrophilic segments of the NC from the translocon channel interior to the more hydrophilic environment of the periplasm (at long NC lengths).

Conclusion

Taken together, the results presented here demonstrate that AP experiments – combined with long-timescale CGMD simulations that enable the interpretation and deconvolution of the experimentally observed pulling-force profiles – provide rich detail on the interactions and conformational changes associated with Sec-facilitated membrane integration and protein translocation.

Author Contributions

M.J.M.N, T.F.M., and G.v.H. conceived the experiments, M.J.M.N, A.M-L, and R.H. conducted the experiments, All authors analyzed the results and reviewed the manuscript.

Acknowledgments

GvH acknowledges support from the Knut and Alice Wallenberg Foundation, the Swedish Cancer Foundation, and the Swedish Research Council. TFM acknowledges support from the National Institutes of Health (R01GM125063) and the Office of Naval Research (N00014-16-1-2761). Computational resources were provided by the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and the Extreme Science and Engineering Discovery Environment (XSEDE) (40), which is supported by National Science Foundation grant number ACI-1548562.

References

ACCEPTED MANUSCRIPT

1. Rapoport, T. A., 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450:663–669.

2. Driessen, A. J. M., and N. Nouwen, 2008. Protein translocation across the bacterial cytoplasmic membrane. *Annu. Rev. Biochem.* 77:643–667.

3. Shao, S., and R. S. Hegde, 2011. Membrane protein insertion at the endoplasmic reticulum. *Annu. Rev. Cell Dev. Biol.* 27:25–56.

4. Cymer, F., G. Von Heijne, and S. H. White, 2015. Mechanisms of integral membrane protein insertion and folding. *J. Mol. Biol.* 427:999–1022.

5. Heinrich, S. U., W. Mothes, J. Brunner, and T. A. Rapoport, 2000. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* 102:233–44.

6. White, S. H., and G. von Heijne, 2005. Transmembrane helices before, during, and after insertion. *Curr. Opin. Struct. Biol.* 15:378 – 386. `http://www.sciencedirect.com/science/article/pii/S0959440X05001260`, membranes/Engineering and design.

7. Berg, B. V. D., W. M. Clemons Jr, I. Collinson, Y. Modis, E. Hartmann, S. C. Harrison, T. A. Rapoport, B. Van den Berg, and W. M. Clemons, 2004. X-ray structure of a protein-conducting channel. *Nature* 427:36–44.

8. Gogala, M., T. Becker, B. Beatrix, J.-P. Armache, C. Barrio-Garcia, O. Berninghausen, and R. Beckmann, 2014. Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* 506:107–10.

9. Egea, P. F., and R. M. Stroud, 2010. Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc. Natl. Acad. Sci.* 107:17182–17187.

10. Frauenfeld, J., J. Gumbart, E. O. van der Sluis, S. Funes, M. Gartmann, B. Beatrix, T. Mielke, O. Berninghausen, T. Becker, K. Schulten, and R. Beckmann, 2011. Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. *Nat. Struct. Mol. Biol.* 18:614–621.

11. Voorhees, R. M., I. S. Fernández, S. H. W. Scheres, and R. S. Hegde, 2014. Structure of the mammalian ribosome-Sec61 complex to 3.4 A resolution. *Cell* 157:1632–1643.

12. Junne, T., T. Schwede, V. Goder, and M. Spiess, 2007. Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. *J. Biol Chem* 282:33201–33209.

13. Junne, T., L. Kocik, and M. Spiess, 2010. The Hydrophobic Core of the Sec61 Translocon Defines the Hydrophobicity Threshhold for Membrane Integration. *Mol. Biol. Cell* 21:1662–1670.

ACCEPTED MANUSCRIPT

14. Demirci, E., T. Junne, S. Baday, S. Bernèche, and M. Spiess, 2013. Functional asymmetry within the Sec61p translocon. *Proc. Natl. Acad. Sci. U. S. A.* 110:18856–61.

15. Stirling, C. J., J. Rothblatt, M. Hosobuchi, R. Deshaies, and R. Schekman, 1992. Protein translocation mutants defective in the insertion of integral membrane proteins into the endoplasmic reticulum. *Mol. Biol. Cell* 3:129–142.

16. Trueman, S. F., E. C. Mandon, and R. Gilmore, 2012. A gating motif in the translocation channel sets the hydrophobicity threshold for signal sequence function. *J. Cell Biol* 199:907–918.

17. Hessa, T., H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S. H. White, and G. von Heijne, 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433:377–381.

18. Goder, V., and M. Spiess, 2003. Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J.* 22:3645–3653.

19. Woodall, N. B., Y. Yin, and J. U. Bowie, 2015. Dual-topology insertion of a dual-topology membrane protein. *Nat. Commun.* 6:8099.

20. Elazar, A., J. Weinstein, I. Biran, Y. Fridman, E. Bibi, and S. J. Fleishman, 2016. Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *eLife* 5:e12125.

21. Marshall, S. S., M. J. M. Niesen, A. Müller, K. Tiemann, S. M. Saladi, R. P. Galimidi, B. Zhang, W. Clemons, and M. T. F., 2016. A Link Between Integral Membrane Protein Expression and Simulated Integration Efficiency. *Cell Rep.* 16:2169–2177.

22. Niesen, M. J. M., S. S. Marshall, T. F. Miller, and W. M. Clemons, 2017. Improving membrane protein expression by optimizing integration efficiency. *J. Biol. Chem.* 292:19537–19545. `http://www.jbc.org/content/292/47/19537.abstract`.

23. Van Lehn, R. C., B. Zhang, and T. F. Miller, 2015. Regulation of multispanning membrane protein topology via post-translational annealing. *eLife* 4:1–23.

24. Ito, K., S. Chiba, and K. Pogliano, 2010. Divergent stalling sequences sense and control cellular physiology. *Biochem. Biophys. Res. Commun.* 393:1 – 5. `http://www.sciencedirect.com/science/article/pii/S0006291X10001221`.

25. Goldman, D. H., C. M. Kaiser, A. Milin, M. Righini, I. Tinoco, and C. Bustamante, 2015. Mechanical force releases nascent chain–mediated ribosome arrest in vitro and in vivo. *Science* 348:457–460. `http://science.sciencemag.org/content/348/6233/457`.

26. Ismail, N., R. Hedman, N. Schiller, and G. von Heijne, 2012. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat. Struct. Mol. Biol.* 19:1018–1022.

27. Cymer, F., and G. von Heijne, 2013. Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proc. Natl. Acad. Sci.* 110:14640–5.

28. Nilsson, O. B., A. A. Nickson, J. J. Hollins, S. Wickles, A. Steward, R. Beckmann, G. von Heijne, and J. Clarke, 2017. Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. Mol. Biol.* 24:221–225. http://dx.doi.org/10.1038/nsmb.3355.

29. Ismail, N., R. Hedman, M. Lindén, and G. von Heijne, 2015. Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat. Struct. Mol. Biol.* 22:145–149.

30. Niesen, M. J. M., C. Y. Wang, R. C. Van Lehn, and T. F. Miller, III, 2017. Structurally detailed coarse-grained model for Sec-facilitated co-translational protein translocation and membrane integration. *PLOS Comp. Biol.* 13:1–26. https://doi.org/10.1371/journal.pcbi.1005427.

31. Zhang, B., and T. F. Miller, 2012. Long-Timescale Dynamics and Regulation of Sec-Facilitated Protein Translocation. *Cell Rep.* 2:927–937.

32. Wimley, W. C., and S. H. White, 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature* 3:842–848.

33. Gumbart, J. C., and C. Chipot, 2016. Decrypting protein insertion through the translocon with free-energy calculations. *BBA -Biomembranes* 1858:1663 – 1671. New approaches for bridging computation and experiment on membrane proteins.

34. Fleming, K. G., 2014. Energetics of Membrane Protein Folding. *Annu. Rev. Biophys.* 43:233–255. https://doi.org/10.1146/annurev-biophys-051013-022926, pMID: 24895854.

35. Rychkova, A., and A. Warshel, 2013. On the nature of the apparent free energy of inserting amino acids into membrane through the translocon. *J. Phys. Chem. B* 117:13748–13754.

36. Ge, Y., A. Draycheva, T. Bornemann, M. V. Rodnina, and W. Wintermeyer, 2014. Lateral opening of the bacterial translocon on ribosome binding and signal peptide insertion. *Nat. Commun.* 5:5263.

37. Zhang, B., and T. F. Miller, 2010. Hydrophobically stabilized open state for the lateral gate of the Sec translocon. *Proc. Natl. Acad. Sci.* 107:5399–5404.

38. Zhang, B., and T. F. Miller, 2012. Direct simulation of early-stage sec-facilitated protein translocation. *J. Am. Chem. Soc.* 134:13700–13707.

ACCEPTED MANUSCRIPT

39. Dudko, O. K., G. Hummer, and A. Szabo, 2008. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci.* 105:15755–15760. http://www.pnas.org/content/105/41/15755.abstract.

40. Towns, J., T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr, 2014. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* 16:62–74.

Supplementary Material

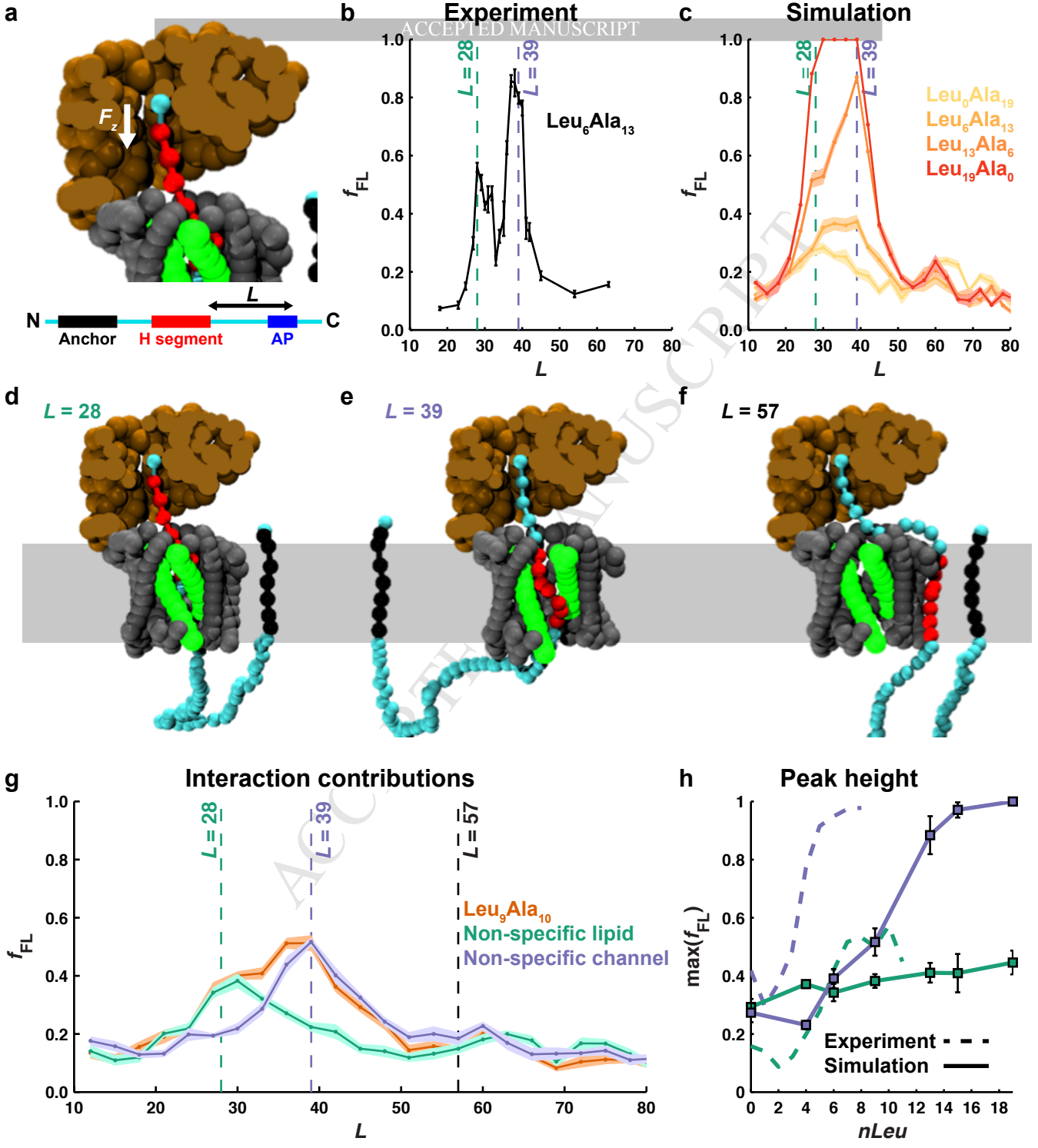
An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

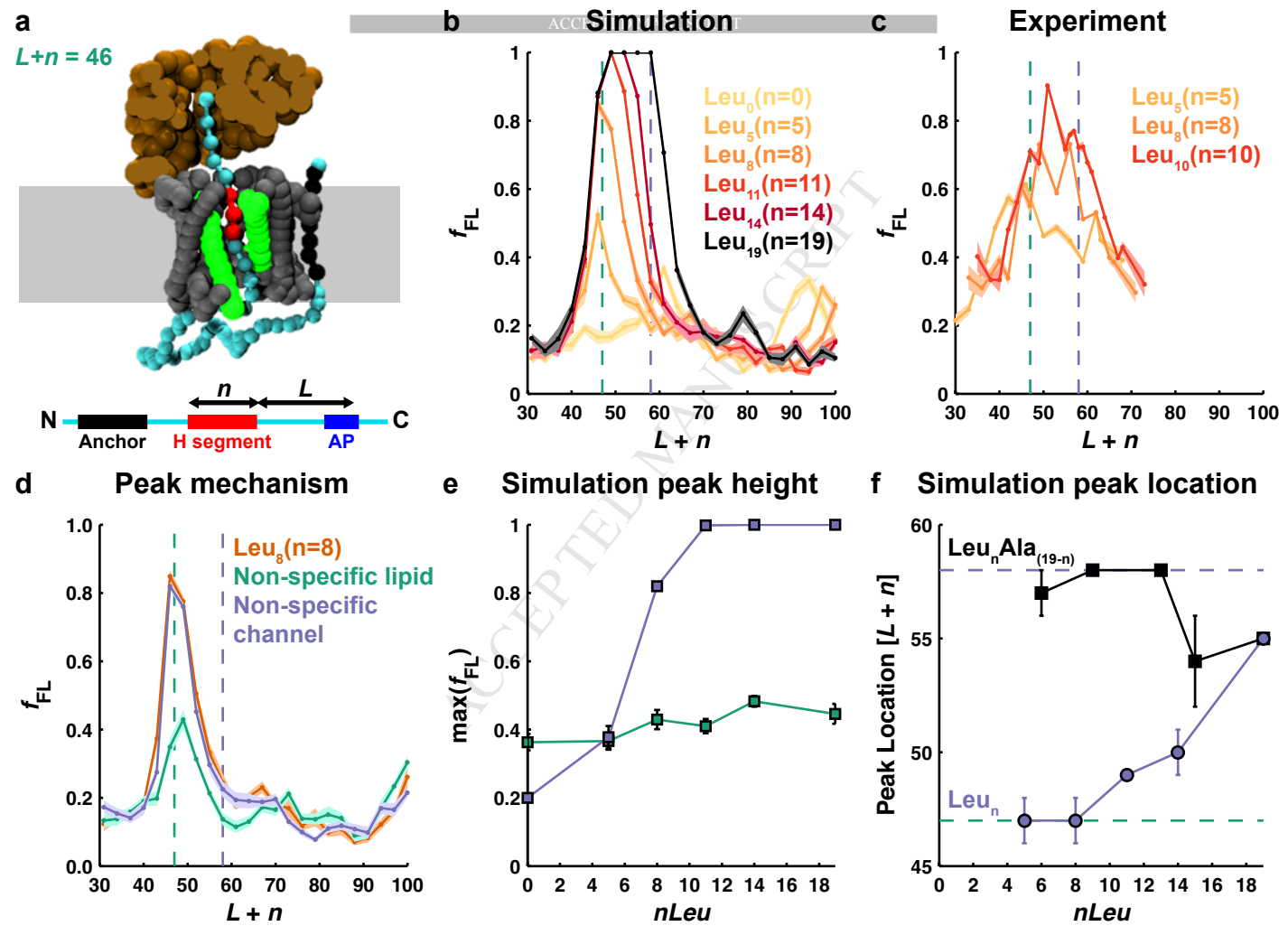
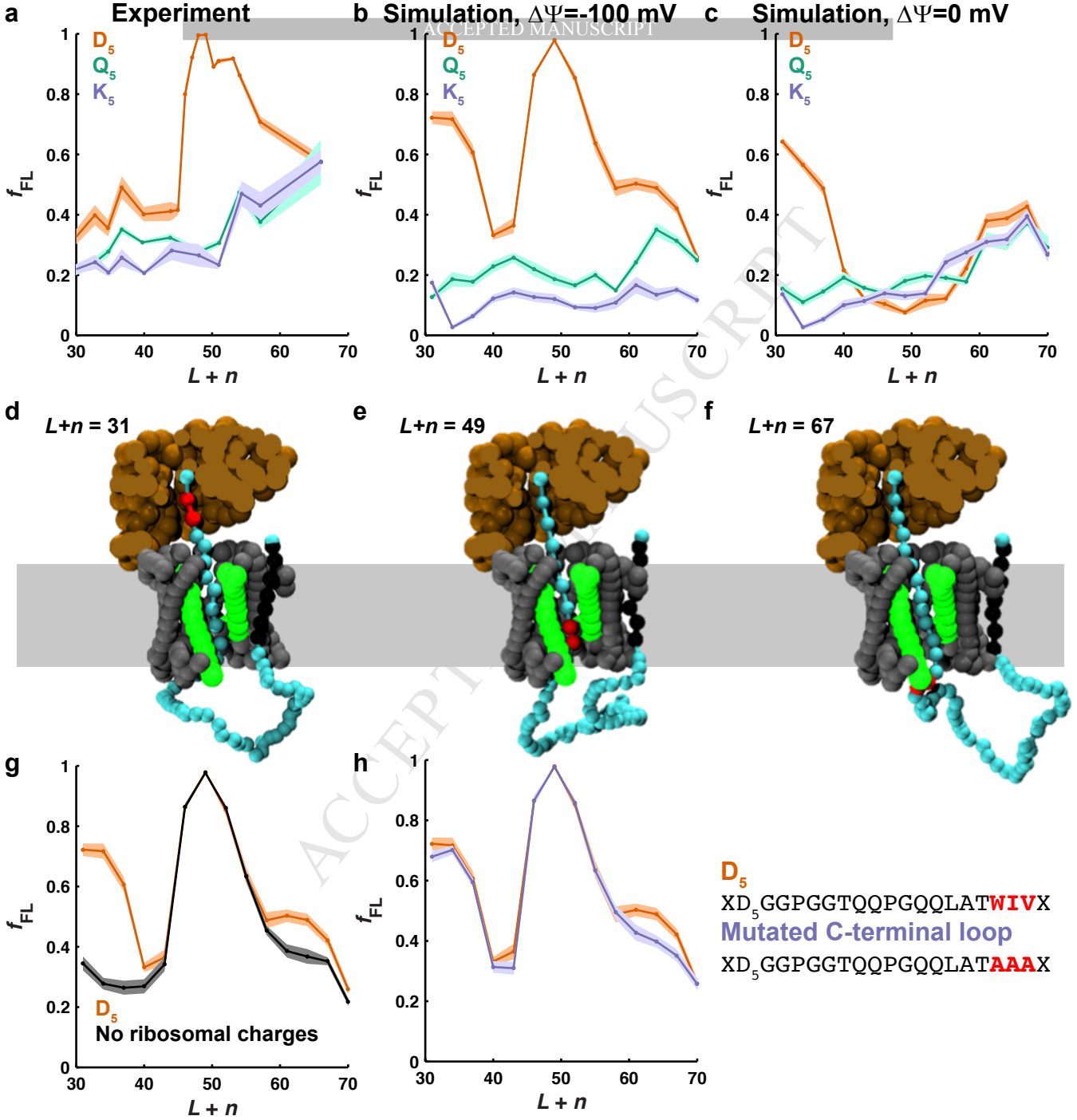
Figure 1: Characterization of the physical processes that drive integration of a hydrophobic transmembrane domain. **(a)** CGMD simulation setup used to calculate pulling forces acting on an engineered hydrophobic H segment (orange) during co-translational integration. Shown is a CGMD snapshot at $L = 28$; the C-terminal bead is held fixed and forces exerted by the nascent protein on that bead are calculated. **(b)** Experimental data reproduced from ref (26). Two peaks in the pulling-force profile are observed during the co-translational integration of the hydrophobic H segment. **(c)** CGMD data for H segments of varying Leucine content. Vertical dashed lines indicate the position of the corresponding peaks in the experimental results. **(d-f)** Representative CGMD configurations at $L = 28$ (d), $L = 39$ (e), and $L = 57$ (f). **(g)** CGMD pulling-force profiles for an H segment with nine leucine residues with default interactions (orange), non-specific lipid interactions (teal), and non-specific channel interactions (purple). **(h)** The maximum value of f_{FL} for the peak near $L = 28$ (teal) and the peak near $L = 39$ (purple), obtained from CGMD (solid lines) and experiment (26) (dashed lines). Error bars indicate the standard error of the mean. Figure S2 provides an alternate version of this figure for which the simulated curves are scaled to enable easier comparison.

Figure 2: Forces exerted on hydrophobic segments of variable length. **(a)** CGMD snapshot for an H segment with eight leucine residues (orange), stalled at $L + n = 46$. The pulling-force profile determined from CGMD **(b)** and from experiment **(c)** for poly-leucine H segments with increasing numbers of leucine residues. **(d)** CGMD pulling-force profile for an H segment with eight leucine residues with default interactions (orange), non-specific lipid interactions (teal), and non-specific channel interactions (purple). **(e)** The maximum value of f_{FL} from CGMD in which the peaks were isolated as shown (d). **(f)** Location of the lipid-interaction peak in the CGMD pulling-force profile as a function of $nLeu$. For poly-leucine H segments (purple) and for 19-residue H segments consisting of alanine and leucine (black). The dashed lines correspond to the $L + n$ value at which the channel interaction peak (teal) and the lipid interaction peak (purple) are observed for the fully spanning transmembrane domains in the fixed-length assay. Error bars indicate the standard error of the mean.

Figure 3: Forces exerted on hydrophilic H segments. The pulling-force profile determined from experiment (29)**(a)** and from CGMD **(b)** for negatively charged (D_5, orange), positively charged (K_5, purple), and neutral (Q_5, teal) 5-residue H segments. **(c)** As in **(b)**, but for CGMD simulations without a membrane potential. **(d-f)** CGMD snapshot for a D_5 H segment (orange), stalled at; $L + n = 31$ (d), $L + n = 49$ (e), and $L + n = 67$ (f). **(g)** The pulling-force profile for a D_5 H segment with (orange) and without (black) ribosomal charges. **(h)** The pulling-force profile for a D_5 H segment with the original C-terminal loop (orange) and with a mutated C-terminal loop that is more hydrophilic (purple). Error bars indicate the standard error of the mean.

ACCEPTED MANUSCRIPT




a
L+n = 46
n
L
N
C
Anchor
H segment
AP
b Simulation
Leu0(n=0)
Leu5(n=5)
Leu8(n=8)
Leu11(n=11)
Leu14(n=14)
Leu19(n=19)
fFL
L + n
c Experiment
Leu5(n=5)
Leu8(n=8)
Leu10(n=10)
fFL
L + n
d Peak mechanism
Leu8(n=8)
Non-specific lipid
Non-specific channel
fFL
L + n
e Simulation peak height
max(fFL)
nLeu
f Simulation peak location
LeunAla(19-n)
Leun
Peak Location [L + n]
nLeu

ACCEPTED MANUSCRIPT