Version August 19, 2024

Single-residue effects on the behavior of a nascent polypeptide chain inside the ribosome exit tunnel

Fátima Pardo-Avila[1†], Renuka Kudva[2,3†], Michael Levitt[1], Gunnar von Heijne[*2,3]

[1]Department of Structural Biology, Stanford University, Palo Alto, CA, USA

[2]Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden

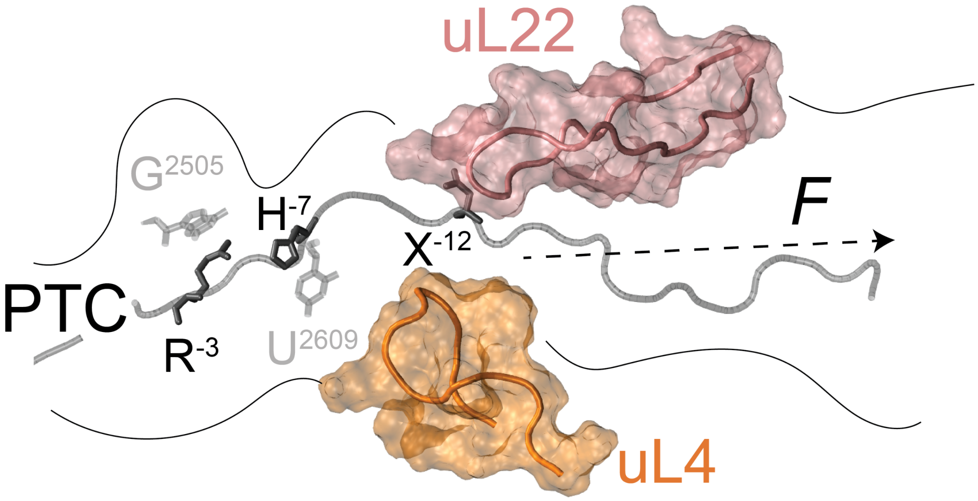
[3]Science for Life Laboratory Stockholm University, Box 1031, SE-171 21 Solna, Sweden

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

[†]These two authors contributed equally to this work.

Keywords: Translational arrest; ribosome; SecM; Force Profile Analysis; molecular dynamics

Graphical Abstract



Abstract

Nascent polypeptide chains (NCs) are extruded from the ribosome through an exit tunnel (ET) traversing the large ribosomal subunit. The ET's irregular and chemically complex wall allows for various NC-ET interactions. Translational arrest peptides (APs) bind in the ET to induce translational arrest, a property that can be exploited to study NC-ET interactions by Force Profile Analysis (FPA). We employed FPA and molecular dynamics (MD) simulations to investigate how individual residues placed in a glycine-serine repeat segment within an AP-stalled NC interact with the ET to exert a pulling force on the AP and release stalling. Our results indicate that large and hydrophobic residues generate a pulling force on the NC when placed ≳10 residues away from the peptidyl transfer center (PTC). Moreover, an asparagine placed 12 residues from the PTC makes a specific stabilizing interaction with the tip of ribosomal protein uL22 that reduces the pulling force on the NC, while a lysine or leucine residue in the same position increases the pulling force. Finally, the MD simulations suggest how the *Mannheimia succiniproducens* SecM AP interacts with the ET to promote translational stalling.

Introduction

During translation, the elongating polypeptide chain traverses the ribosome through a ~100 Å long exit tunnel (ET) in the large ribosomal subunit (1-4). While the ET was initially thought to have a "Teflon-like" surface that would have minimal interactions with the nascent polypeptide chain (NC) (3), it is by now well-established that there are ample possibilities for such interactions in specific regions along its length (5-14). The best-studied examples where interactions between the ET and the NC modulate translation elongation are the so-called translational arrest peptides (APs), relatively short stretches of sequences in NCs that interact with the proximal parts of the ET to induce translational arrest (15,16). Detailed information on AP-ET interactions has been obtained by cryo-EM and molecular dynamics (MD) simulations of ribosome-nascent chain complexes (RNCs) and by extensive mutagenesis studies (12-14,17-24), but studies of how individual residues may engage in NC-ET interactions in more distal parts of the exit tunnel are less common (25-29).

Here, we use Force Profile Analysis (FPA), a method based on the observation that the translational arrest efficiency of an AP is sensitive to external pulling forces acting on the NC, to map residue-specific interactions between the NC and the ET. Along with all-atom MD simulations, we examine how individual amino acids in a 19-residue segment comprising alternating glycine (G) and serine (S) residues, engineered next to the eight-residue long SecM AP from *Mannheimia succiniproducens* (SecM(*Ms*))(8) interact with the ET. We examine the interactions of single charged (K, D), polar (N), and hydrophobic (W, P, L) residues at different locations in the GS-segment, corresponding to a part of the ET that is ~20 Å to ~70 Å distant from the peptidyl transferase center (PTC) of the ribosome. Overall, our results indicate that the introduction of larger residues in this area of the NC generates a pulling force on the AP in a direction away from the PTC. Moreover, placing an L or D residue in the nascent chain at position -18 (relative to the PTC), near the constriction site formed by the loops of ribosomal proteins uL4 and uL22 protruding into the ET (30), leads to a specific increase in the pulling force. In contrast, an N residue placed at position -12 relative to the PTC (i.e., at the constriction site) makes a specific stabilizing interaction with the tip of protein uL22 that reduces the pulling force on the AP, while a K or L residue in the same position increases the pulling force. In addition, data from the MD simulations suggest how residues in the SecM(*Ms*) AP can interact with rRNA within the ET to potentially affect translational stalling.

Materials and methods

Enzymes and chemicals

All enzymes used in this study were purchased from Thermo Fisher Scientific and New England Biolabs. T7-RNA polymerase was purchased from Promega. All reagents and chemicals except for Bacto-yeast extract and Bacto-peptone were purchased from Sigma Aldrich (now Merck); Bacto-yeast extract and Bacto-peptone to culture *E.coli* cells for the S30 extract, were purchased from BD-Biosciences. Oligonucleotides were from Eurofins Genomics and gene fragments from Thermo Fisher Scientific (GeneArt). DNA isolation/purification kits and precast polyacrylamide gels were from Thermo Fisher Scientific. L-[^{35}S]-methionine was obtained from PerkinElmer.

Cloning and Mutagenesis

The ADR1a constructs that were sub-cloned into the pET19b expression plasmid in a previous study (31) were used as a parental construct for this project. Gene fragments with DNA regions encoding for a total of forty GS repeats and the SecM(*Ms*) AP were designed and ordered from GeneArt (ThermoFisher Scientific) – the codons encoding G and S were varied to avoid repeats.

This fragment was introduced downstream of the gene encoding for ADR1a by_Gibson assembly®. The shorter constructs for the initial screen (ADR1 linker length scan available in Supplementary Excel file on Zenodo: 10.5281/zenodo.13244596) were generated by truncating the GS linker to the desired length by inverse PCR using phosphorylated primers, ligation with T4 DNA ligase and transformed into DH5α chemical competent cells. The different single mutants (single amino acid substitutions) were generated from the construct with 19 amino acid GS linker (as described in the results) by site directed mutagenesis using partially overlapping primers. All cloning and mutagenesis products were confirmed by DNA sequencing. The amino acid sequences for all the constructs can be found in the Supplementary files on Zenodo (10.5281/zenodo.13244596).

Coupled in vitro transcription and translation and quantitation

The constructs used in this study were translated in an *E.coli* Zn-free S30 cell extract (31). The S30 extract was prepared from *E.coli* MRE 600 cells following the protocol detailed in (32), with some modifications to rid the lysate of endogenous $Zn^{2+.}$ Specifically, cells were cultured to an A600 of 1.2, following which they were treated with 100 uM TPEN for one hour, and harvested by centrifugation and prepared as described (32).

300 ng of plasmid DNA were used as templates for polypeptide synthesis, and translation was carried out in the presence of [^{35}S] Methionine at 37°C for 15 min and shaking at 300 r.p.m. (33). Since the S30 extract used was depleted of endogenous Zn^{2+}, for the reactions + Zn^{2+}, the reactions included 50 µM zinc acetate. The translation reaction was terminated after 15 min by treating the samples with a final concentration of 5% trichloroacetic acid (TCA), followed by a 30 min incubation on ice. They were subsequently centrifuged at 20,000 g for 10 min in a

tabletop centrifuge (Eppendorf) and the pellet obtained was solubilized in Laemmli buffer, supplemented with RNaseA (400 µg/ml), and denatured at 37°C for 15 min.

The samples were resolved on 12% Bis-Tris gels (Thermo Scientific) in MOPS buffer. Gels were dried and subjected to autoradiography and scanned using the Fujifilm FLA-9000 phosphorimager for visualization of radioactively labeled translated proteins. The one-dimensional intensity profiles corresponding to the protein bands on each individual lane in the gel images were extracted using ImageGauge (Fujifilm). The output .txt files (deposited on Zenodo: 10.5281/zenodo.13244596) were used as inputs to visualize and fit to a Gaussian distribution using EasyQuant (Rickard Hedman, Stockholm University). The sum of the arrested and full-length bands was calculated, and this was used to estimate the fraction full-length protein for each construct as exemplified in Figure 1.

Model building

Our ribosome-AP model was based on the PDB structure *3jbu* (19). This structure is based on a cryo-electron density map of the *Escherichia coli* ribosome stalled during translation by the 17 amino acid long arrest peptide SecM. We used PyMOL (34) to mutate the residues in the nascent chain to match the sequence used in the experimental setup (Fig. 2). The nascent peptide chain modelled in *3jbu* contains 4 peptide bonds with the omega dihedral in the *cis* configuration (the cis peptide bonds in the nascent peptide are shown in bold: KLISE**ED**LFSTP**VW**ISQAQ**GI-RA**G) (SI Fig. SI2). It is highly unlikely to have omega dihedrals in the *cis* configuration, so we changed all these bonds to the trans configuration. For a couple of residues, it was possible to fix the peptide bonds by “flipping” the oxygen of the peptide bond (manually change the coordinates of the atom), followed by energy minimization to correct the geometry. To fix the remaining bonds, we used the GeneralizedKIC mover (35) implemented in Rosetta (36,37). We defined the loops as 4 residues centred at the *cis* bond. We fixed one bond at a time in a sequential manner. We generated multiple models for each position and selected the ones with the best score.

Building nascent peptide bound to tRNA

The nascent chain with the fixed peptide bonds was concatenated with the tRNA chain present in *3jbu*, and the residues were renumbered. A new bond had to be created between the Ser $^{-1}$ of the nascent peptide and RA104 of the tRNA. We obtained the missing parameters for the bond between the nucleic acid and the amino acid (Ser^{-1}) using ACPYPE (38). The parameters necessary to specify the bond were added to the final topology files.

Molecular Dynamics Simulations

Molecular Dynamics simulations were performed using the Amber99SB-ILDN forcefield (39) with GROMACS 2018.4 (40). Each protein was solvated in a dodecahedron box of explicit TIP3P (41) water, and neutralized with 4115 K, 258 Cl, and 129 Mg ions. The ions chosen were included to try to mimic the experimental conditions.

We followed an equilibration protocol based on a previously described protocol for ribosome simulations (42):

0–5 ns: NVT and position restraints on all ribosomal heavy atoms (force constant of 1000 kJ $\mathrm{mol}^{-1}\mathrm{nm}^{-2}$).

5–10 ns: NVT, the position restraints force constant was linearly decreased to zero.

10–20 ns: NPT with a Berendsen barostat (43) with a coupling constant $\tau p = 1$ps and an isotropic compressibility of $4.5 \cdot 10^{-5} \mathrm{bar}^{-1}$

After equilibration, for each system (Control, G^{-12}, K^{-12}, N^{-12}), 5 simulations of 50 ns were performed in the NVT ensemble, with periodic boundary conditions. A 10 Å cut-off was used for van der Waals and short-range electrostatic interactions. The Particle-Mesh Ewald (PME) summation method was used for long-range electrostatic interactions (44). Verlet cut-off scheme was used (45). Covalent bonds were constrained using the LINCS algorithm (46). The integration time-step was 2 fs for all steps.

Hydrogen bonds calculation

We calculated the prevalence of hydrogen bonds present in the 334-residue trajectories and in a subsystem of 78 residues near the PTC (for all systems). We used the baker_hubbard function (47) implemented in the MDTraj Python library (48). The function identifies hydrogen bonds based on cutoffs for the Donor – H…Acceptor distance and angle:

$$\theta > 120 \text{ and } r_{H...Acceptor} < 2.5\text{Å}$$

We used three frequency thresholds: 25%, 50%, 75%, and 90% of the simulation time.

Cross-correlation analysis

The Pearson correlation of the covariance matrix (cross-correlation) allows us to quantify the similarity between two time-series datasets, by measuring the extent to which changes in one dataset corresponds to changes in another dataset, over a range of time. We can use cross-correlation to identify the coupling of the motion of the atoms in molecular dynamics simulations (49). Thus, being able to uncover hidden patterns or relationships. The correlation coefficient (C_{ij}) between atom i and atom j is defined as:

$$C_{ij} = \frac{< \Delta X_i \cdot \Delta X_j >}{\left[< \Delta X_i^2 >< \Delta X_j^2 >\right]^{1/2}}$$

Where ΔX_i is the fluctuation of the position of atom i with respect to its mean position. $C_{ij} = 1$ denotes positively correlated motion, while $C_{ij} = -1$ denotes negatively correlated motion. We calculated the cross-correlation between the center-of-mass of all the residues in the 334-residue filtered trajectories. To calculate the cross-correlation, we first calculated the covariance matrix using the “gmx covar” function implemented in GROMACS (40), and then using the “cov2cor” function implemented in R 3.5.0 (50), to transform the covariance matrix to the cross-correlation matrix.

Root Mean Square Deviation (RMSD) measurements (in layers)

The RMSD is the measure of the average distance between atoms of superimposed structures. We used the pre-fitted c.o.m. trajectories (created as described in a previous section). The initial model after energy minimization is the reference structure. To calculate the RMSD by layers, we divided the ribosome in 10 Å layers, using the nascent peptide as the centre. All the layers were 10 Å, except for the last one, which included all atoms from 80 to 200 Å. We used the RMSD Trajectory Tool (RMSDTT v3.0 plugin) implemented in VMD (51).

Results

Force Profile Analysis

FPA is based on the observation that the translational arrest efficiency of certain APs, including the SecM AP, is sensitive to external pulling forces acting on the NC (15,52,53). In general, stronger pulling forces result in low levels of translational arrest, and *vice versa*. Thus, by engineering an AP into a protein and measuring the arrest efficiency, one can obtain a proxy for the pulling force acting on the protein NC at the moment when the ribosome reaches the last codon in the AP (52). In FPA, a series of protein constructs are made where an AP followed by a short C-terminal tail is fused to a protein of interest, Fig. 1A. Each construct is translated either *in vivo* or *in vitro* (here, we use *in vitro* translation in a modified *E. coli* S30 cytosolic extract that is depleted of endogenous Zn^{2+} and devoid of membranes (32)), and subjected to a short pulse of [^{35}S]-Met, followed by analysis by SDS-PAGE and quantitation of the amount of product representing NCs arrested at the AP (I_A) and the amount representing non-arrested, full-length chains (I_{FL}), Fig. 1C. Finally, the fraction full-length product, $f_{FL} = I_{FL}/(I_A+I_{FL})$, is calculated as a measure of the pulling force exerted on the AP in that particular construct (52). FPA has been used to probe cotranslational processes such as protein folding (31,54-62), membrane protein biogenesis (52,63-65), and protein translocation across membranes (58,59,66). Here, we use FPA to probe NC-ET interactions by measuring how the pulling force exerted on the NC is modified by single point mutations introduced in different locations along a model, ribosome-embedded NC composed of glycine-serine (GS) repeats (referred to as the GS-linker).

In order to probe the effects on f_{FL} of single charged (K, D), polar (N), and hydrophobic (W, L, P) residues in different positions along the NC, we analyzed six series of constructs, Fig. 1A. The constructs have a common design that is based on previous studies on the cotranslational folding of the small Zn-finger domain ADR1a (31,62,67,68): a 150 residue long N-terminal unstructured segment derived from the periplasmic domain of the LepB protein followed by a 6-residue linker, the 29-residue ADR1a Zn-finger domain, a 19-residue GS-linker composed of 10 Gly and 9 Ser residues, the 8-residue *Mannheimia succiniciproducens* SecM (SecM(*Ms*)) AP (of sequence HAPIRGSP), and a 23-residue C-terminal tail. Zn^{2+}-induced co-translational folding of ADR1a in the ET has previously been shown to generate a strong pulling force on the NC and to reduce the degree of translational stalling on the related *Escherichia coli* SecM AP (SecM(*Ec*)), while in the absence of Zn^{2+}, ribosome stalling on the SecM(*Ec*) AP is efficient (31,62). The SecM(*Ms*) AP is the AP of choice in the present study since it is considerably shorter and more resilient to pulling forces than the SecM(*Ec*) AP (52), making it possible to probe locations in the NC closer to the PTC.

In each of the six scan-series, the G residues in the 19-residue GS linker were individually replaced by the specific tested residue type, and f_{FL} was determined for each construct. For each of the six scan-series, we could thus determine the position-specific effect on f_{FL} of the residue in question in locations 8 to 25 residues (~20 Å to ~70 Å) away from the PTC, both in the presence (+Zn^{2+}) and in the absence (-Zn^{2+}) of a strong external pulling force on the NC.

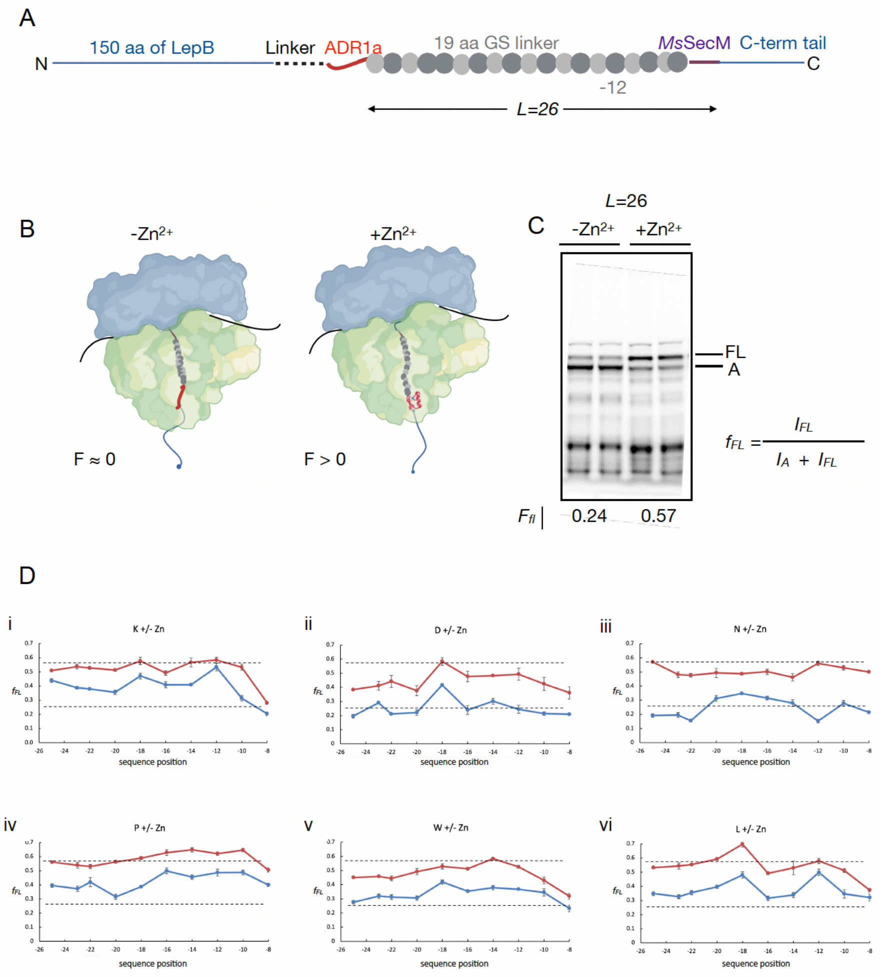


Figure 1. Diagrammatic representation of constructs used. The Zinc-finger domain of ADR1a was engineered to the 8 amino acid arrest peptide of SecM from *Mannheimia succiniciproducens* via a 19 amino acid linker consisting of GS-repeats (GS-linker). 150 amino acids of the periplasmic domain of LepB were introduced at the N-terminus of ADR1 and 23 amino acids at the C-terminus of SecM to be able to resolve the arrested and full-length protein products by SDS-PAGE and autoradiography (B) A schematic of ribosomes stalled by the SecM AP and how an N-terminal pulling force generated by the folding of ADR1a can result in a resumption of translation. In the panel on the left, ADR1a (red) does not fold due to the absence of Zn^{2+} whereas ADR1a folds inside the ET in the presence of Zn^{2+} , generating a pulling force on the NC, in the panel on the right. (C) Autoradiographs of unfolded (-Zn^{2+}) and folded

(+Zn^{2+}) ADR1a constructs stalled by the SecM(*Ms*) AP after radioactive pulse-labelling *in vitro* and SDS-PAGE. The linker length (*L*) corresponds to the 19 amino acid GS-repeats and the 7 amino acids of the SecM(*Ms*) AP (not counting the C-terminal Pro residue which is bound to the A-site tRNA during arrest). The relative amounts of arrested (*A*) and full-length (*FL*) product were estimated by quantification of the protein bands in the autoradiographs, and the fraction full-length was calculated as $f_{FL}=I_{FL}/(I_A+I_{FL})$. Two repeat experiments are shown for each construct. (D) Force profile analysis of constructs with (i) K; (ii) D; (iii) N; (iv) P; (v) W, and (vi) L at different positions within the Gly-Ser linker.

The results for the six scan-series are shown in Fig. 1D, where f_{FL} is plotted against the position of the mutated residue relative to the C-terminal Ser residue in the SecM(*Ms*) AP (which is attached to the P-site tRNA in the stalled ribosome-NC complex, counted as position -1). As expected, f_{FL} values are always higher in the presence of Zn^{2+} (red curves) than in its absence (blue curves). Interestingly, the shapes of the f_{FL} curves are similar ±Zn^{2+}. However, in the presence of Zn^{2+}, all f_{FL} values are lower or equal to the f_{FL} value for the unperturbed GS-linker (dotted line at f_{FL} = 0.58), except for P in positions -10 to -16 and L in position -18. In the absence of Zn^{2+}, f_{FL} values are generally increased compared to the unperturbed GS-linker (dotted line at f_{FL} = 0.26) for the K, W, L, and P series, but not for constructs in the D and N series. Finally, for all scans across the different series, f_{FL} values drop in position -8, with the strongest drop seen in the scans recorded for the larger residues K, W, and L.

Comparing the different residue types, it is notable that the f_{FL} values for the K series are higher than those for the D series, both in the presence and absence of Zn^{2+} (most easily seen in SI Fig. S1a). Although, at first sight, this might suggest that a positively charged residue exerts a stronger pulling force on the NC than a negatively charged one, this is not consistent with the observation that the f_{FL} values for the D and N series are similar throughout the range of positions tested (Supplementary Fig. S1b). Instead, the best correlation is with the size of the residue: the series for larger and hydrophobic residues (K, W, L, P) generally have higher f_{FL}

values than those for the smaller ones (D, N). Overall, the P series has the highest f_{FL} values of all the series in the region -8 to -16, i.e., in the proximal part of the ET relatively close to the peptidyl transferase center (PTC).

Beyond these general trends, a few constructs have notably high or low f_{FL} values in certain positions: D and L (and to some degree K and W) in position -18 (high f_{FL}), L and K in position -12 (high f_{FL}), and N in position -12 (low f_{FL}). The latter result is particularly interesting, as it suggests that an N residue at position -12 forms a specific interaction with the exit tunnel that increases the stalling strength of the AP, thereby reducing f_{FL}. We explored this observation using the all-atom MD simulations described next.

All-atom MD simulations

To help identify and visualize the molecular interactions that result in the observed differences in f_{FL} values obtained in the absence of Zn^{2+} (*i.e.*, under low pulling-force load) in the FPA experiments, we used all-atom MD simulations (Fig. 2A). We focused on the G→N and G→K mutations in position -12 (referred to as N^{-12} and K^{-12}, respectively) as these two constructs represent the biggest observed difference in f_{FL} values in any of the analyzed positions, Fig. 1D. The N^{-12} and K^{-12} mutants were compared to the unperturbed G^{-12} sequence, and to an additional control construct where the SecM(*Ms*) AP was replaced by an equally long stretch of GS-repeats Fig. 2B.

Since there are currently no structures available for the *E. coli* ribosome with a stalled SecM(*Ms*) AP, we generated a molecular model based on the PDB structure of an *E. coli* ribosome stalled by the closely related SecM(*Ec*) AP (PDB ID *3jbu*; EMDB ID EMD-6483) (19). When setting up our system, *3jbu* was the highest resolution structure available for the SecM(*Ec*) AP. We mutated the NC residues in the PDB structure to match the sequence used in the experimental set-up described above: 26 GS residues in the Control model, the SecM(*Ms*) AP (with the penultimate S residue in the P-site), plus a 19-residue GS linker in the G^{-12} model, and the N and K mutations at position -12 for the mutated models (Fig. 2B, C). Since the MD simulations were employed to probe the interactions of the individual residues within the GS-linker with the ET (independent of additional Zn^{2+} in the translation extract), and therefore of any potential pulling force generated due to the folding of ADR1a, we excluded ADR1a from the model.

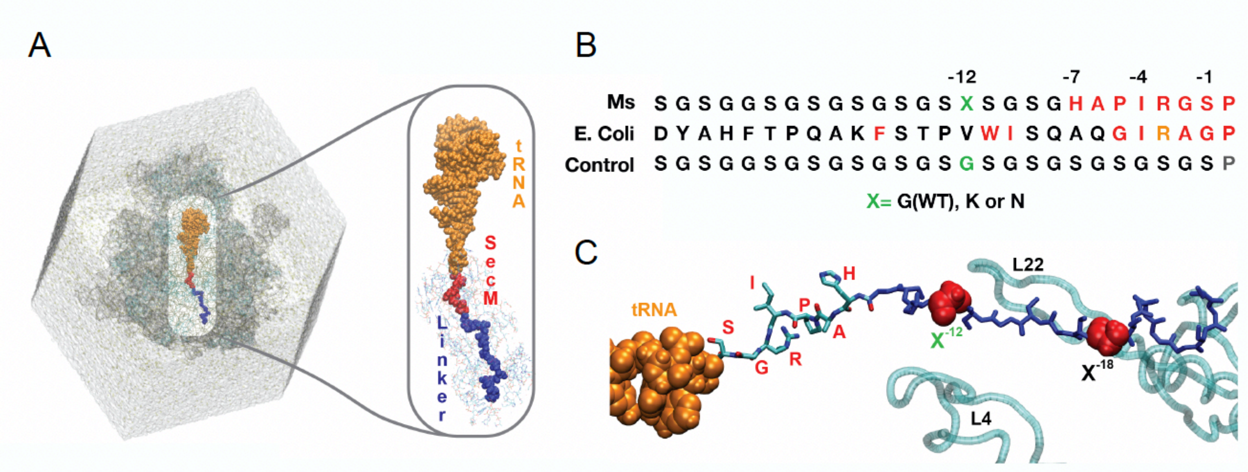

A
tRNA
SecM
Linker
B
-12
-7
-4
-1
Ms S G S G G S G S G S G S G S X S G S G H A P I R G S P
E. Coli D Y A H F T P Q A K F S T P V W I S Q A Q G I R A G P
Control S G S G G S G S G S G S G S G S G S G S G S G S G S P
X= G(WT), K or N
C
tRNA
S
G
R
I
P
A
H
L22
X-12
X-18
L4

Figure 2. Molecular Dynamics setup. (A) Ribosome in a solvated dodecahedron box. The ribosome is shown as a cartoon, and the solvent as a grey surface. The exit tunnel is highlighted. The tRNA is shown as orange spheres, SecM(*Ms*) AP as red spheres, and GS-repeat linker as blue spheres. In the close-up, the nucleic acids within 15 Å of the nascent peptide are shown as lines. (B) SecM AP and GS-linker sequences. Residues critical for stalling in the SecM(*Ms*) and SecM(*Ec*) APs are highlighted in red. Position -12 is highlighted in green; X is either Gly (in WT), Lys, or Asn. The *E. coli* SecM(*Ec*) AP sequence is shown for reference. In the Control sequence, the SecM(*Ms*) AP is replaced by an equally long GS-repeat sequence. (C) Close-up of the nascent peptide (WT sequence) inside the exit tunnel. The loops of the uL4 and uL22 proteins are shown in cyan as references. Positions -12 and -18 are shown as red spheres.

The *3jbu* model that we used as a basis for our SecM(*Ms*) AP model contained 4 peptide bonds with the omega dihedral in the *cis* configuration in the NC (peptide bonds shown in bold: KLISE**ED**LFSTP**VW**ISQAQ**GI**-**RA**G). We converted all the *cis* peptide bonds to the *trans* configuration, as it is unlikely that the NC would contain *cis* peptide bonds (69,70). Remodeling the backbone resulted in changes in sidechain orientation. In particular, R^{-3} no longer pointed towards the A-site (the A-site tRNA was not included in our model, SI Fig. S2).

Of note, in a new, recently published high-resolution ribosome structure with a stalled SecM(*Ec*) AP (12) (PDB ID *8qoa* ; EMDB ID EMD-18534), the proximal SecM(*Ec*) region between F^{150}-G^{165} forms a compact α-helix, not seen in the *3jbu* model. This α-helix was also predicted by AlphaFold2 (71) and the secondary structure prediction program PSIPRED (72). However, neither AlphaFold2 nor PSIPRED predict a helical structure for the SecM(*Ms*) AP (SI Fig. S3). Therefore, the more extended conformation of the AP in *3jbu* (with the *cis-trans* corrections noted above) seems to represent a better starting point for modelling the SecM(*Ms*) AP and the GS-repeat segment than the compact helical conformation seen in *8qoa* .

For each of the four different systems (G^{-12}, K^{-12}, N^{-12}, Control), we simulated five replicates of 50 ns production runs each. We focused the analysis on the residues within a 15 Å radius of the NC (Fig. 2A).

We did not observe major differences in the ribosome structure between the replicas and between the four simulated systems . From visual inspection of the simulations, we noticed that the L1 stalk of the large ribosomal subunit was very flexible. Thus, we calculated the root mean squared distance (RMSD) by dividing the ribosome into concentric atom layers, starting from the nascent peptide (SI Fig. SI4). The RMSD plots show the similarity to a reference structure; in this case, the reference is the original ribosome model (used to start all the simulations) after energy minimization. As seen in the SI Fig. SI4A, all the simulations were stable and behaved similarly.

We further investigated the stability of the residues of the ET by calculating the root mean squared fluctuation (RMSF), which measures how much a residue moves during the simulation (fluctuation around the average position) for the components of the exit tunnel (nucleic acids of the tunnel, P-site tRNA, the loops of the proteins uL22, uL4, and uL23 that protrude into the ET), as well as for the NC. The residues in the protein loops have low RMSF values, and the values are similar for all simulations (SI Fig. SI5). We observed higher RMSF values for the nucleic acids in the ET than for the amino acids in the uL22 and uL24 loops; the values were again similar among all four systems. Notably, the RMSF values for the K^{-12} and N^{-12} systems are very similar along the NC, except in the region between the SecM(*Ms*) AP and the mutation site (residues -12 to -7), where the K^{-12} NC appears more mobile than the N^{-12} NC (SI Fig. SI5A).

In short, the ET components behave similarly in all the simulations and are stable regardless of the specific sequence of the nascent peptide.

NC-ET interactions

In order to better understand how the K^{-12} and N^{-12} mutations affect the NC-ET interactions, we calculated the fraction of the total simulation time that specific nucleotides or amino acid residues in the ET spend within a distance of 4 Å of residues S^{-11}, X^{-12}, and S^{-13} in the NC. The observed interactions involve a few distinct regions in the ET (Fig. 3). For position -12 in the NC, the only residue that does not interact at all with uL4 is N^{-12}. Both the K^{-12} and N^{-12} residues interact with uL22, A751, and A752 for more than 50% of the simulation time. However, the interaction between N^{-12} and uL22 is due to sidechain-sidechain interactions, while the interaction between K^{-12} and uL22 is through the K backbone atoms, while its sidechain points towards A751 and A752 (SI Fig. SI6). The K^{-12} sidechain can also reach a pocket formed by A789/A790 and U1781.

The interaction between S^{-13} and the exit tunnel follows a similar pattern, with the N^{-12} mutation strongly promoting interactions between S^{-13} and uL22/A751. In contrast, for S^{-11}, we observed that consistent high-probability interactions are lacking, as most interactions are observed only in one single simulation (SI Fig. SI7). The N^{-12} mutation promotes a medium-probability interaction between S^{-11} and A751/A752, which is not seen in the other systems.

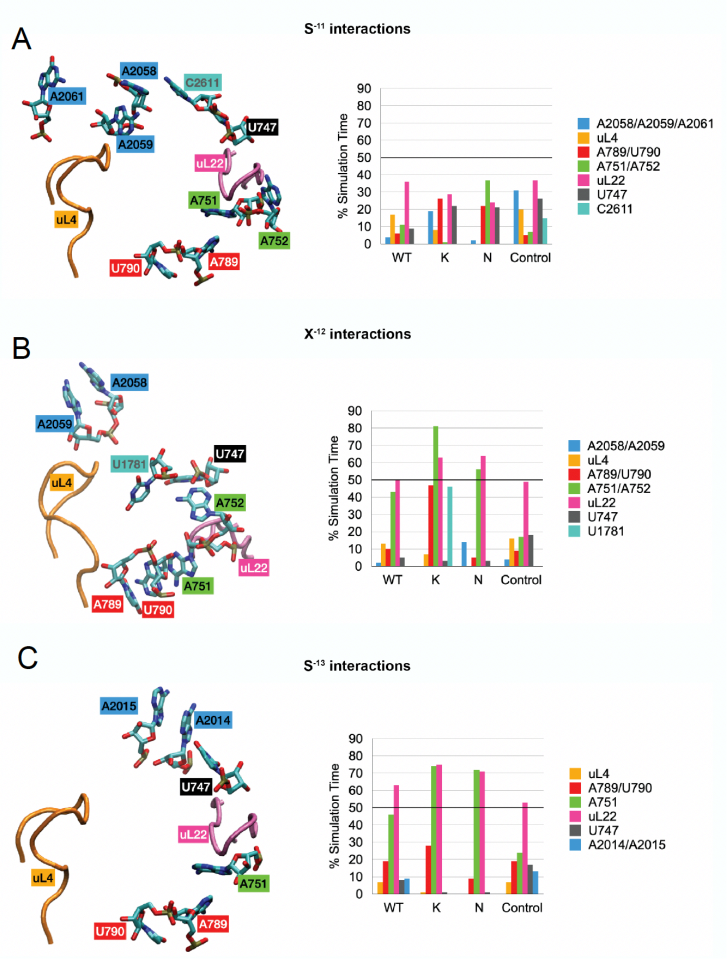


Figure 3. NC-ET interactions for residues S^{-11} (A), X^{-12} (B), and S^{-13} (C). Residues are defined as interacting when the minimum distance between them is less than 4 Å. The bar graphs show the percentage of the simulation time that the indicated residues interact. The interacting residues are shown in stick representations (left panels).

The 23S A751/A752 loop has been previously implicated in the ribosomal response to the SecM(*Ec*) and TnaC APs (73). Another study suggested the importance of stable aromatic interactions between W^{155} in the SecM(*Ec*) AP and A^{751} for ribosomal stalling (18).

Overall, we observe that N^{-12} and S^{-13} in the N^{-12} NC interact almost exclusively with uL22 and A751/A752. The N^{-12}-uL22 interactions are mediated through the N^{-12} sidechain. In contrast, we observe that K^{-12} explores more extensive regions within the ET and interacts with uL22 via its backbone. These results are consistent with the higher RMSF values observed for the K^{-12} NC in the region around the mutation site (SI Fig. S5).

A persistent interaction between N^{-12} *and uL22.*

To further characterize the interactions between the NC and the ET, we calculated the Pearson correlation of the covariance matrix of the center-of-mass of all residues (nucleotides and amino acids) within a 15 Å radius from the NC. We used the correlation to assess if the motion of a pair of residues is coordinated and to identify clusters of residues that could form a network of interactions that may affect translational arrest.

For the N^{-12} NC simulations, we found a strong correlation between N^{-12} and residues in the uL22 loop (Fig. 4A, SI Fig. SI10 and SI11). G^{-12} in the unperturbed NC variant also had a fairly strong correlation with the uL22 loop, whereas K^{-12} showed no such correlation. In contrast, none of the NC variants showed a strong correlation with the loop of uL4, which also forms part of the constriction site. Notably, mutations in Gly^{91} and Ala^{93} at the tip of the uL22 loop suppress the translational arrest induced by the SecM(*Ec*) AP (5). At the same time, uL4 has been suggested to be less involved in interactions that promote translational arrest by *E. coli* SecM (74), and mutations in uL4 had little or no effect on the SecM AP response (75).

Thus, we propose that the low f_{FL} value recorded for the N^{-12} mutation (Fig. 1D) results from a strong interaction between N^{-12} and uL22 that makes the NC bind more tightly to the ET and hence increases the arrest potency of the AP. While G^{-12} in the unperturbed system also has a fairly strong cross-correlation with the uL22 loop, K^{-12} has a much weaker cross-correlation with the uL22 loop, in agreement with the higher f_{FL} value of the latter.

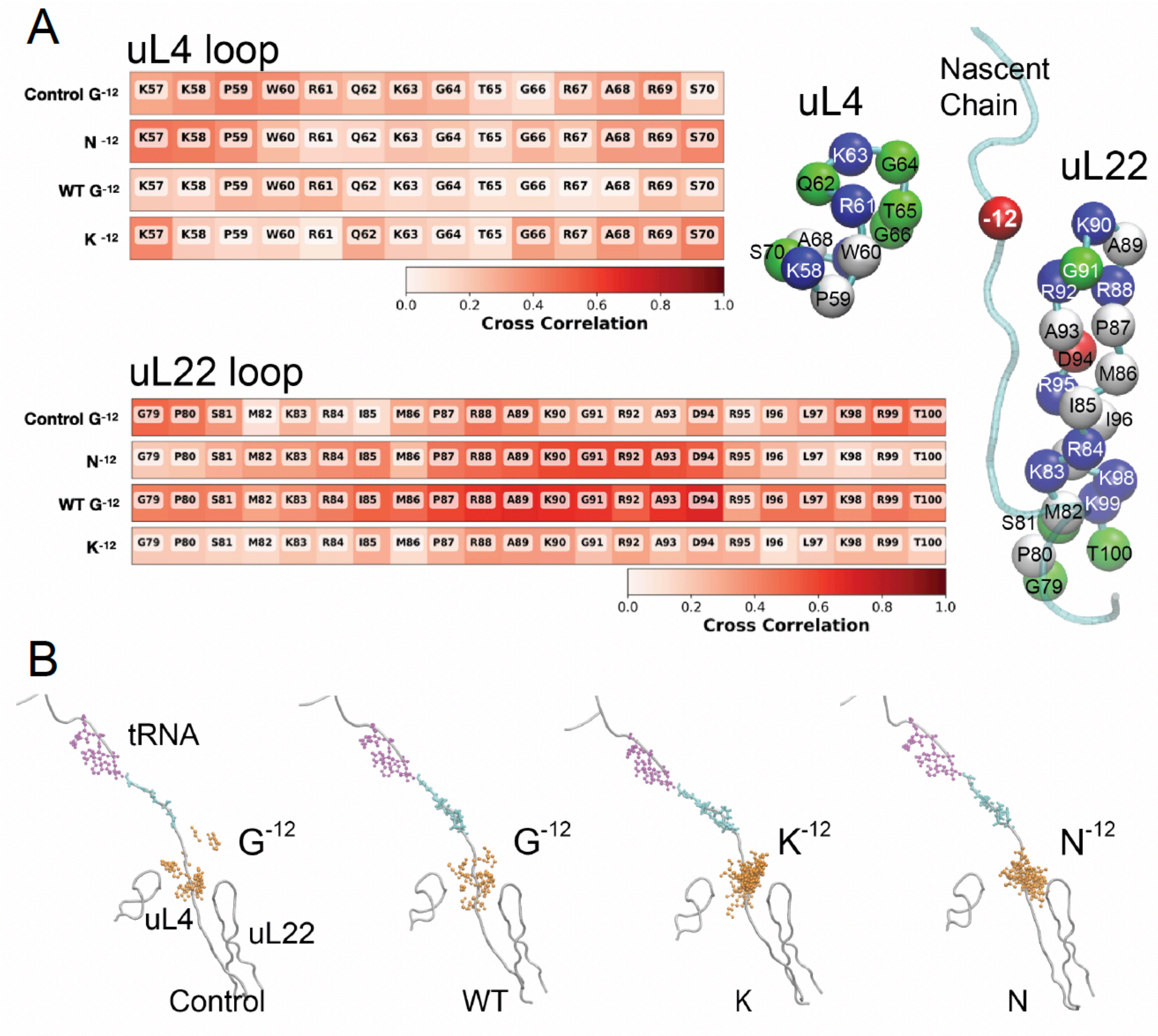


Figure 4. NC-ET interactions. (A) Cross-correlation between the residue of the nascent peptide in position -12 (for four systems) and the loops of proteins uL22 (residues G^{79} to T^{100}) and uL4 (residues K^{57} to S^{70}). The image on the right shows the loops of proteins uL22 (residues G^{79} to T^{100}) and uL4 (residues K^{57} to S^{70}). Each residue is represented as a sphere located at the center of mass of each residue. (B) The location within the ET of the nascent peptide residue at position -12. For each system, the average position of the loops of proteins uL4 and uL22 and the nascent peptide are shown in grey as cartoon representations. The SecM(*Ms*) AP residues are shown as cyan sticks, and the CCA terminus of the tRNA

is shown as magenta sticks. To visualize the regions that the residue at position -12 visits during the simulation, this residue is shown as a CPK model (in orange) from overlapping frames of the sub-sampled trajectories.

Hydrogen bonds and stacking interactions in the ET.

In a final analysis (Fig. 5), we identified all the hydrogen bonds present within a 15 Å radius of the NC. The hydrogen bonds that are present >50% of the simulation time are mostly between nucleotides. By lowering the threshold to 25% of the simulation time, we could identify a few hydrogen bond interactions between the NC and the ET in the region near the mutation site, including K^{-12} and N^{-12}, that can both form a hydrogen bond with the backbone of K^{90} of uL22. As noted above, these interactions differ because the bond formed with K^{-12} is through its backbone, while N^{-12} interacts with uL22 via its sidechain.

We also identified hydrogen bonds near the PTC (Supplementary Tables 1 to 3). Three hydrogen bond clusters were found in this region, Fig. 5F. Cluster 2 is more extensive in the N^{-12} NC variant, including hydrogen bonds between U2506 and G2583, nucleotides that have been previously associated with stalling (76). Major differences can be observed between the SecM(*Ms*) AP simulations and the Control simulations (Fig. 5A-D). While in the SecM(*Ms*) AP simulations, R^{-3} in the SecM(*Ms*) AP and G2505 interact, either through hydrogen bonds or by stacking (SI Fig. S12), for the Control system hydrogen bonds are also formed between G2505 and G2581. Notably, in the SecM(*Ms*) AP simulations, rRNA bases U1782-U2586 and U1779-A1784 near the PTC interact, while these interactions are not seen in the Control system. U2586 has been previously proposed as relevant for SecM-mediated stalling (19). We also identified two regions that are stable for all four systems (low RMSF values and multiple hydrogen bonds):

residues C2507-G2582 (near the A-site), and C2064-G2446, C2065-G2445, and C2066-G2444 (near the uL4 loop).

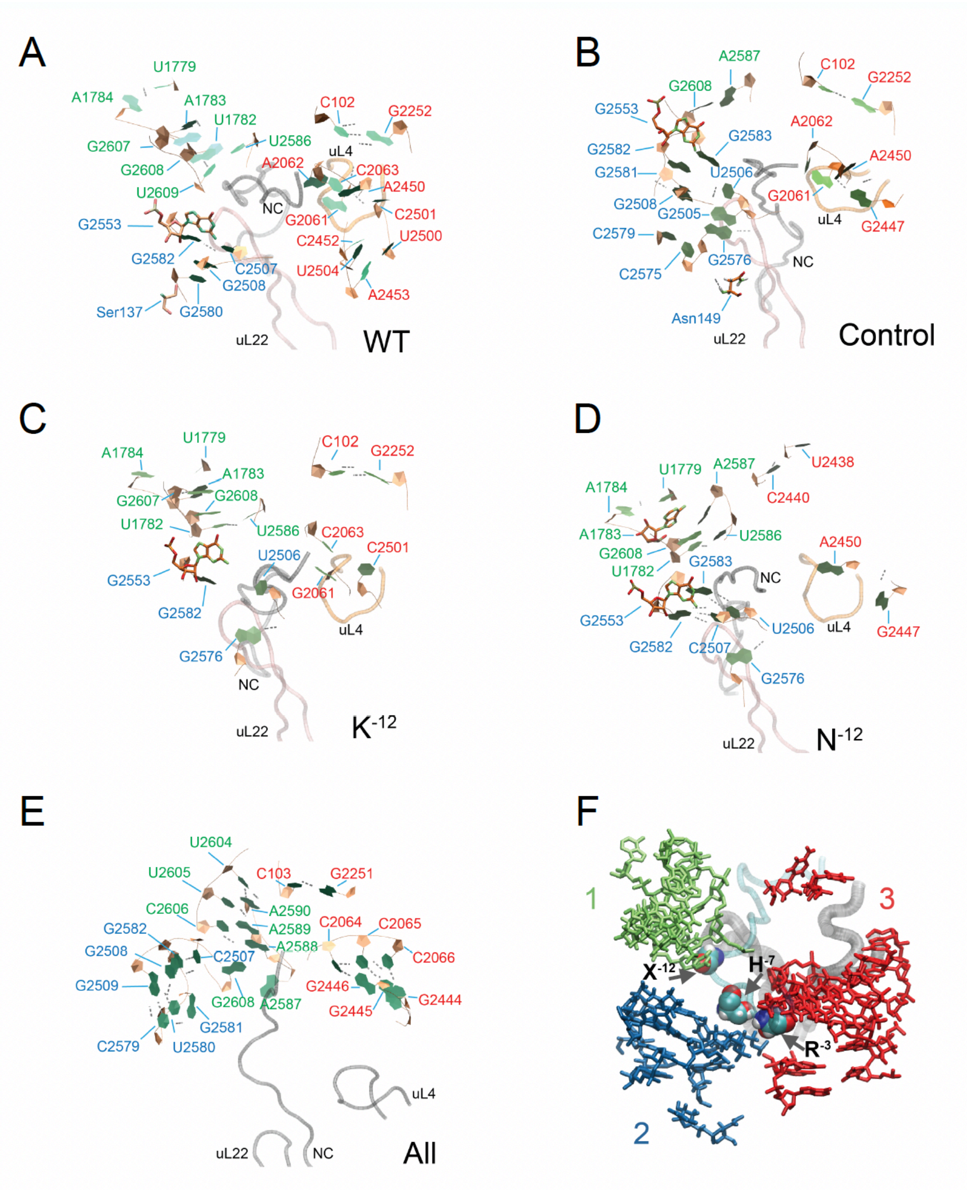

A
U1779
A1784
A1783
U1782
G2607
U2586
G2608
U2609
G2553
G2582
C2507
G2508
Ser137
G2580
C102
G2252
uL4
A2062
C2063
A2450
NC
G2061
C2501
C2452
U2500
U2504
A2453
uL22
WT
B
A2587
G2608
G2553
G2582
G2581
G2583
U2506
G2508
G2505
C2579
G2576
C2575
Asn149
C102
G2252
A2062
A2450
G2061
uL4
G2447
NC
uL22
Control
C
U1779
A1784
A1783
G2607
G2608
U1782
G2553
G2582
G2576
U2586
U2506
C102
G2252
C2063
C2501
G2061
uL4
NC
uL22
K^{-12}
D
U1779
A2587
U2438
A1784
C2440
A1783
U2586
G2608
G2583
U1782
A2450
NC
G2553
G2582
C2507
U2506
uL4
G2447
G2576
uL22
N^{-12}
E
U2604
U2605
C103
G2251
C2606
A2590
G2582
A2589
C2064
C2065
G2508
C2507
A2588
C2066
G2509
G2446
G2608
A2587
G2444
G2445
G2581
C2579
U2580
uL4
uL22
NC
All
F
1
3
X^{-12}
H^{-7}
R^{-3}
2

Figure 5. Hydrogen bond networks near the PTC. (A-E) The structures show the nucleic acids (orange and green, NewRibbons VMD representation) that form hydrogen bonds present at least 50% of the simulation time. Grey dashed lines between residues represent hydrogen bonds between residues. The loops of proteins uL4 (orange) and uL22 (pink), as well as the nascent chain (grey) are shown as tubes. Note that the frame selected for each system may show only some of all the possible hydrogen bonds. For a list of the hydrogen bonds present 50%, 75%, and 90% of the simulation time, refer to SI tables 1-3. To facilitate visualization, the nucleic acids that were involved in hydrogen bonds more than 50% of the time and were present in all the systems are shown on panel E and are not shown in panel A to D (A: WT; B:

Control; C: K^{-12}; D: N^{-12}). (F) The h-bonds are localized in three major regions, as shown in the blue, green, and red sticks. The loops of proteins uL22 and uL4, tRNA CCA fragment, and nascent peptide are shown as ribbons. SecM residues R^{-3}, H^{-7}, and X^{-12} are shown as spheres. The viewpoint is different from the one shown for panels A to E. The structure was rotated ~90 degrees so the tRNA would come off the page towards the reader. The colors of the residue labels in panels A to E correspond to the three major regions highlighted in this panel.

We further observed a stacking interaction between H^{-7} in the SecM(*Ms*) AP and U2609, present mainly in the N^{-12} NC variant, Fig. 6. Even though it only accounts for 20% of the frames, it represents another possible stabilizing interaction between the N^{-12} NC and the ET. U2609 has been previously implicated in the ribosomal response to the SecM(*Ec*) AP (73), and the recent SecM(*Ec*) AP model *8qoa* shows a stacking interaction between the similarly placed F^{150} and U2609. In the TnaC structure (14), it was observed that U2609 is part of the binding pocket where the L-Trp molecule binds and has been found essential for stalling.

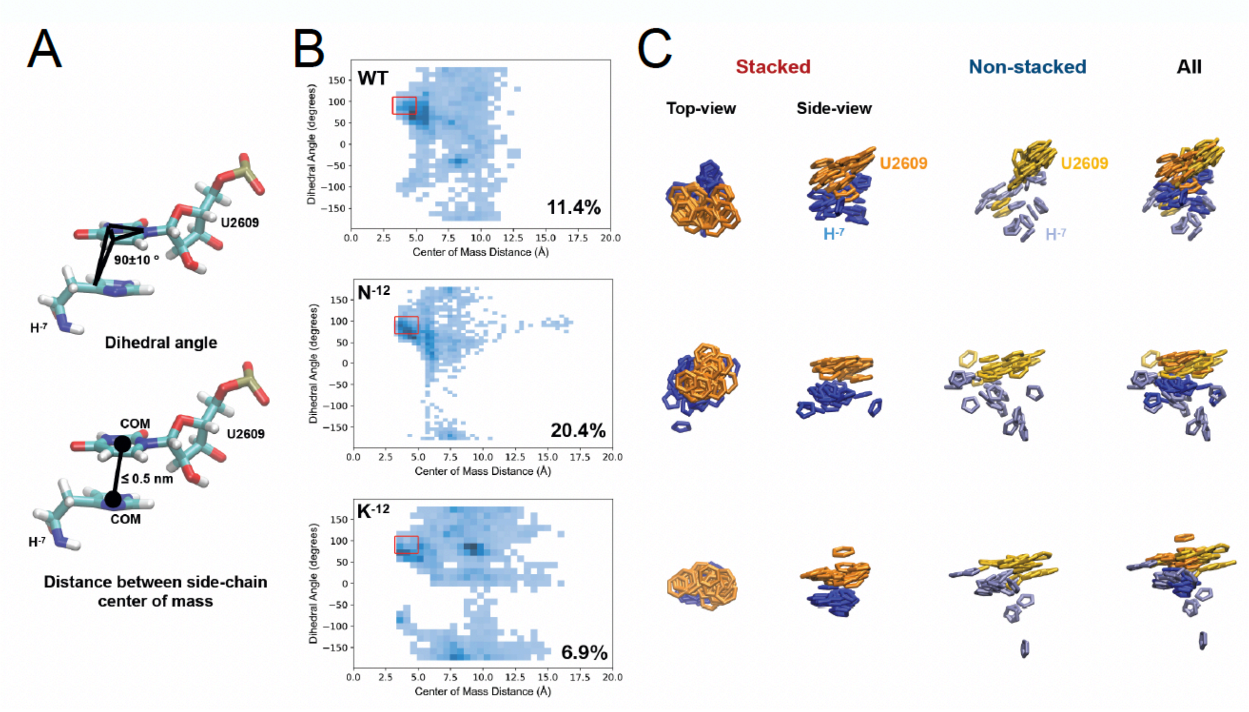


Figure 6. Stacking between H^{-7} and U2609. (A) Two reaction coordinates are used to define stacking between H^{-7} and U2609 (left panel). The distance between the centre of mass (c.o.m.) of the H^{-7} sidechain and the base ring of U2609 is less than 0.5 nm, and the dihedral angle formed between the base ring of

U2609 and the C2 atom of H^{-7} has a value between 70 and 110 degrees. All the frames were projected onto these two reaction coordinates (dihedral angle vs c.o.m. distance). (B) 2D histograms of the projections onto the reaction coordinates described in A. The red rectangle shows the region where the stacking conditions are met, and the percentage of frames that fall within this region is shown in the bottom right of each plot. (C) Extracted structures from the stacked and non-stacked regions highlighted in B. The aromatic rings of H^{-7} and U2609 are shown as blue and orange sticks, respectively.

Discussion

In summary, under conditions of high external pulling force on the NC (in the presence of Zn^{2+}), mutation of the Gly residues in the "skinny" GS-repeat sequence in positions -8 to -25 have no or a slightly reducing effect on f_{FL}. Thus, not unexpectedly, the force generated by the cotranslational folding of ADR1a largely swamps out the effects of single point mutations in the GS-repeat sequence. In the absence of a strong external pulling force (in the absence of Zn^{2+}), however, differences between the different types of residues start to appear. Mutation of Gly residues in the GS-repeat segment to any of the larger residues (K, W, L) leads to increases in f_{FL}; *i.e.*, large residues tend to pull the NC towards the tunnel exit, away from the PTC. As the ET gets progressively wider as one approaches the tunnel exit from the constriction site, sidechain entropy would favor the movement of large residues in this direction. Mutation of Gly residues to medium-size residues (D, N) does not, in general, affect f_{FL} except in certain positions: -18 for D and N (high f_{FL}) and -12 for N (low f_{FL}). Previous experimental work (25,28,57) and theoretical calculations (77-81) have suggested that the average electrostatic potential varies along the ET; we do not see much evidence for an electrostatic effect (compare the f_{FL} plots for the negatively charged D and its neutral analog N, Supplementary Fig. SI1b), implying that, on the single-residue scale, local residue-residue interactions (hydrogen bonds, salt bridges, stacking) dominate over electrostatic gradients.

Gersteuer et al. have proposed that SecM(*Ec*)-mediated arrest comprises two modules: an arrest module and a regulator module (12). The arrest module is formed by residues RAG/P (with the P residue attached to the A-site tRNA) and is directly involved in preventing peptide bond formation. Meanwhile, the more N-terminal regulator module can modulate the strength of stalling. In the SecM(*Ms*) AP, the arrest module would correspond to the RGS/P sequence. A mutagenesis scan showed that the SecM(*Ms*-Sup1) AP sequence is a stronger staller than SecM(*Ec*), as mutations that match SecM(*Ec*) ($G^{-2} \rightarrow$ A and $S^{-1} \rightarrow$ G) are significantly weaker stallers than SecM(*Ms*-Sup1). The mutagenesis scan also showed that R^{-3} is crucial for stalling (21).

U2504 has been proposed to control the access to the cavity where the incoming A-site amino acid would bind (82). In the old SecM(*Ec*) AP structure *3jbu* (19), the sidechain of R^{-3} (R^{165}) also points towards this pocket, potentially blocking the A-site. However, in the recent, high-resolution structure of the SecM(*Ec*) AP *8qoa* (12), R^{165} forms a stacking interaction with U2504. Interestingly, even though we base our model of the SecM(*Ms*) AP on *3jbu*, our correction of a couple of presumably incorrectly modeled *cis* peptide bonds led to a starting structure for the MD simulations where the sidechain of R^{-3} no longer points to the same location as in *3jbu*. Instead, it points towards the pocket formed by the 23S rRNA residues A2503-U2506, similar to its location in the recent cryo-EM structure. R^{-3} in the SecM(*Ms*) AP remains within this pocket in the MD simulations and forms a stable stacking interaction with U2505 (>50% of simulation time in the unperturbed G^{-12} system; SI Fig. SI12).

The mutations in position -12 fall in the proposed regulator module. Our results show two different and opposite behaviors between the N^{-12} and K^{-12} systems. On the one hand, we observe stabilizing interactions and a high correlation between N^{-12} and the uL22 loop. Furthermore, a stacking interaction observed between H^{-7} and U2609 (Fig. 6) could also contribute to the increase in stalling efficiency of the N^{-12} system. On the other hand, we observe increased flexibility for the K^{-12} system with the lysine side-chain moving between different interaction sites, none of which is particularly stable, in line with the low stalling efficiency seen for this system.

Our results demonstrate that FPA is sufficiently sensitive to detect position- and residue-specific differences in how individual amino acids in a nascent chain interact with the ribosomal exit tunnel, and that all-atom MD simulations can be used to pin-point the relevant interactions.

Data availability

The data underlying this article are available and deposited in Zenodo: 1) Molecular dynamics simulations, filtered to include all the residues within 1.5 nm of the nascent chain, including the nascent chain and the complete tRNA. The initial models with mutated nascent chains and fixed *cis* peptide bonds are also included (doi.org/10.5281/zenodo.13248385); 2) Setup of Molecular dynamic simulation and model (doi.org/10.5281/zenodo.13340587); 3) Molecular dynamics analysis scripts (doi.org/10.5281/zenodo.13259735); 4) EasyQuant files, Excel sheet with the analysis of the .txt files from EasyQuant, and amino acid sequences of all constructs (doi.org/10.5281/zenodo.13244596).

Supplementary Data

Supplementary Figures and Tables are available in the SupplementaryMaterial.pdf

Author contributions

RK performed the force profile analysis experiments. FPA performed Molecular Dynamics simulations. FPA, RK, ML, GvH discussed the results and commented on the manuscript.

Acknowledgments

We thank Dr. Rickard Hedman (Stockholm University) for programming and maintenance of the EasyQuant software. We thank Dr. Hanlun Jiang (University of California, Berkeley) for his assistance in remodeling the SecM(*Ms*) nascent chain. Fig 1A and B were made using BioRender.

Funding

This work was supported by grants from the Knut and Alice Wallenberg Foundation (2017.0323), the Novo Nordisk Fund (NNF18OC0032828), and the Swedish Research Council (621-2014-3713) to GvH, by the Stanford Data Science Initiative to FPA and by the National Institute of Health (R35 GM122543) to ML.

Conflict of interest

The authors declare no competing interests.

References

1. Bernabeu, C. and Lake, J.A. (1982) Nascent polypeptide chains emerge from the exit domain of the large ribosomal subunit: immune mapping of the nascent chain. *Proc Natl Acad Sci U S A*, **79**, 3111-3115. 10.1073/pnas.79.10.3111.
2. Milligan, R.A. and Unwin, P.N. (1986) Location of exit channel for nascent protein in 80S ribosome. *Nature*, **319**, 693-695. 10.1038/319693a0.
3. Nissen, P., Hansen, J., Ban, N., Moore, P.B. and Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920-930.
4. Wilson, D.N. and Beckmann, R. (2011) The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Curr Opin Struct Biol*, **21**, 274-282. 10.1016/j.sbi.2011.01.007.
5. Nakatogawa, H. and Ito, K. (2002) The ribosomal exit tunnel functions as a discriminating gate. *Cell*, **108**, 629-636.
6. Woolhead, C.A., McCormick, P.J. and Johnson, A.E. (2004) Nascent membrane and secretory proteins differ in FRET-detected folding. *Cell*, **116**, 725-736.
7. Lu, J. and Deutsch, C. (2005) Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol*, **12**, 1123-1129. 10.1038/nsmb1021.
8. Yap, M.N. and Bernstein, H.D. (2009) The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel. *Mol Cell*, **34**, 201-211. 10.1016/j.molcel.2009.04.002.
9. Baram, D. and Yonath, A. (2005) From peptide-bond formation to cotranslational folding: dynamic, regulatory and evolutionary aspects. *FEBS Lett*, **579**, 948-954. 10.1016/j.febslet.2004.11.063.

10. Muta, M., Iizuka, R., Niwa, T., Guo, Y., Taguchi, H. and Funatsu, T. (2020) Nascent SecM chain interacts with outer ribosomal surface to stabilize translation arrest. *Biochem J*, **477**, 557-566. 10.1042/BCJ20190723.
11. Tsai, A., Kornberg, G., Johansson, M., Chen, J. and Puglisi, J.D. (2014) The dynamics of SecM-induced translational stalling. *Cell reports*, **7**, 1521-1533. 10.1016/j.celrep.2014.04.033.
12. Gersteuer, F., Morici, M., Gabrielli, S., Fujiwara, K., Safdari, H.A., Paternoga, H., Bock, L.V., Chiba, S. and Wilson, D.N. (2024) The SecM arrest peptide traps a pre-peptide bond formation state of the ribosome. *Nat Commun*, **15**, 2431. 10.1038/s41467-024-46762-2.
13. Su, T., Cheng, J., Sohmen, D., Hedman, R., Berninghausen, O., von Heijne, G., Wilson, D.N. and Beckmann, R. (2017) The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. *eLife*, **6**, e25642.
14. Su, T., Kudva, R., Becker, T., Buschauer, R., Komar, T., Berninghausen, O., von Heijne, G., Cheng, J. and Beckmann, R. (2021) Structural basis of l-tryptophan-dependent inhibition of release factor 2 by the TnaC arrest peptide. *Nucleic Acids Res*, **49**, 9539-9547. 10.1093/nar/gkab665.
15. Ito, K. and Chiba, S. (2013) Arrest peptides: *cis*-acting modulators of translation. *Annu Rev Biochem*, **82**, 171-202. 10.1146/annurev-biochem-080211-105026.
16. Chiba, S., Fujiwara, K., Chadani, Y. and Taguchi, H. (2023) Nascent chain-mediated translation regulation in bacteria: translation arrest and intrinsic ribosome destabilization. *J Biochem*, **173**, 227-236. 10.1093/jb/mvad007.

17. Bhushan, S., Meyer, H., Starosta, A.L., Becker, T., Mielke, T., Berninghausen, O., Sattler, M., Wilson, D.N. and Beckmann, R. (2010) Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell*, **40**, 138-146.

18. Gumbart, J., Schreiner, E., Wilson, D.N., Beckmann, R. and Schulten, K. (2012) Mechanisms of SecM-mediated stalling in the ribosome. *Biophys J*, **103**, 331-341.

19. Zhang, J., Pan, X., Yan, K., Sun, S., Gao, N. and Sui, S.F. (2015) Mechanisms of ribosome stalling by SecM at multiple elongation steps. *eLife*. 10.7554/eLife.09684.

20. Sohmen, D., Chiba, S., Shimokawa-Chiba, N., Innis, C.A., Berninghausen, O., Beckmann, R., Ito, K. and Wilson, D.N. (2015) Structure of the *Bacillus subtilis* 70S ribosome reveals the basis for species-specific stalling. *Nat Commun*, **6**, 6941. 10.1038/ncomms7941.

21. Cymer, F., Hedman, R., Ismail, N. and von Heijne, G. (2015) Exploration of the arrest peptide sequence space reveals arrest-enhanced variants. *J Biol Chem*, **290**, 10208-10215. 10.1074/jbc.M115.641555.

22. Shanmuganathan, V., Schiller, N., Magoulopoulou, A., Cheng, J., Braunger, K., Cymer, F., Berninghausen, O., Beatrix, B., Kohno, K., von Heijne, G. *et al.* (2019) Structural and mutational analysis of the ribosome-arresting human XBP1u. *eLife*, **8**. 10.7554/eLife.46267.

23. Di Palma, F., Decherchi, S., Pardo-Avila, F., Succi, S., Levitt, M., von Heijne, G. and Cavalli, A. (2022) Probing Interplays between Human XBP1u Translational Arrest Peptide and 80S Ribosome. *J Chem Theory Comput*, **18**, 1905-1914. 10.1021/acs.jctc.1c00796.

24. Kolář, M.H., Nagy, G., Kunkel, J., Vaiana, S.M., Bock, L.V. and Grubmüller, H. (2022) Folding of VemP into translation-arresting secondary structure is driven by the ribosome exit tunnel. *Nucleic Acids Res*, **50**, 2258-2269. 10.1093/nar/gkac038.

25. Lu, J., Kobertz, W.R. and Deutsch, C. (2007) Mapping the electrostatic potential within the ribosomal exit tunnel. *J Mol Biol*, **371**, 1378-1391. 10.1016/j.jmb.2007.06.038.

26. Leininger, S.E., Rodriguez, J., Vu, Q.V., Jiang, Y., Li, M.S., Deutsch, C. and O'Brien, E.P. (2021) Ribosome Elongation Kinetics of Consecutively Charged Residues Are Coupled to Electrostatic Force. *Biochemistry*, **60**, 3223-3235. 10.1021/acs.biochem.1c00507.

27. Zimmer, M.H., Niesen, M.J.M. and Miller, T.F., 3rd. (2021) Force transduction creates long-ranged coupling in ribosomes stalled by arrest peptides. *Biophys J*, **120**, 2425-2435. 10.1016/j.bpj.2021.03.041.

28. Lu, J. and Deutsch, C. (2008) Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol*, **384**, 73-86. 10.1016/j.jmb.2008.08.089.

29. Po, P., Delaney, E., Gamper, H., Szantai-Kis, D.M., Speight, L., Tu, L., Kosolapov, A., Petersson, E.J., Hou, Y.M. and Deutsch, C. (2017) Effect of Nascent Peptide Steric Bulk on Elongation Kinetics in the Ribosome Exit Tunnel. *J Mol Biol*, **429**, 1873-1888. 10.1016/j.jmb.2017.04.019.

30. Dao Duc, K., Batra, S.S., Bhattacharya, N., Cate, J.H.D. and Song, Y.S. (2019) Differences in the path to exit the ribosome across the three domains of life. *Nucleic Acids Res*, **47**, 4198-4210. 10.1093/nar/gkz106.

31. Nilsson, O.B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., Muller-Lucks, A., Trovato, F., Puglisi, J.D., O'Brien, E.P. *et al.* (2015) Cotranslational protein

folding inside the ribosome exit tunnel. *Cell reports*, **12**, 1533-1540. 10.1016/j.celrep.2015.07.065.

32. Kudva, R. (2019) Preparation of a cell-free expression system from *Escherichia coli*. *Protocols IO*. dx.doi.org/10.17504/protocols.io.2xkgfkw.
33. Moser, M., Panahandeh, S., Holzapfel, E. and Muller, M. (2007) In vitro analysis of the bacterial twin-arginine-dependent protein export. *Methods Mol Biol*, **390**, 63-79. 10.1007/978-1-59745-466-7_5.
34. Schrödinger, LCC. The PyMOL Molecular Graphics System v. 2.0.
35. Bhardwaj, G., Mulligan, V.K., Bahl, C.D., Gilmore, J.M., Harvey, P.J., Cheneval, O., Buchko, G.W., Pulavarti, S.V., Kaas, Q., Eletsky, A. *et al.* (2016) Accurate de novo design of hyperstable constrained peptides. *Nature*, **538**, 329-335. 10.1038/nature19791.
36. Fleishman, S.J., Leaver-Fay, A., Corn, J.E., Strauch, E.M., Khare, S.D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G. *et al.* (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One*, **6**, e20161. 10.1371/journal.pone.0020161.
37. Leman, J.K., Weitzner, B.D., Lewis, S.M., Adolf-Bryfogle, J., Alam, N., Alford, R.F., Aprahamian, M., Baker, D., Barlow, K.A., Barth, P. *et al.* (2020) Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods*, **17**, 665-680. 10.1038/s41592-020-0848-2.
38. Sousa da Silva, A.W. and Vranken, W.F. (2012) ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, **5**, 367. 10.1186/1756-0500-5-367.

39. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. and Shaw, D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950-1958. 10.1002/prot.22711.

40. Abraham, M., Murtola, T., Schulz, R., Pall, S., Smith, J., Hess, B. and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1**, 19-25.

41. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, **79**, 926-935. 10.1063/1.445869.

42. Bock, L.V., Blau, C., Schroder, G.F., Davydov, II, Fischer, N., Stark, H., Rodnina, M.V., Vaiana, A.C. and Grubmuller, H. (2013) Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat Struct Mol Biol*, **20**, 1390-1396. 10.1038/nsmb.2690.

43. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, **81**, 3684-3690. 10.1063/1.448118.

44. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A Smooth Particle Mesh Ewald Method. *J Chem Phys*, **103**, 8577-8593. Doi 10.1063/1.470117.

45. Pall, S., Zhmurov, A., Bauer, P., Abraham, M., Lundborg, M., Gray, A., Hess, B. and Lindahl, E. (2020) Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J Chem Phys*, **153**, 134110. 10.1063/5.0018516.

46. Hess, B. (2008) P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput*, **4**, 116-122. 10.1021/ct700200b.

47. Baker, E.N. and Hubbard, R.E. (1984) Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol*, **44**, 97-179. 10.1016/0079-6107(84)90007-5.

48. McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernandez, C.X., Schwantes, C.R., Wang, L.P., Lane, T.J. and Pande, V.S. (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J*, **109**, 1528-1532. 10.1016/j.bpj.2015.08.015.

49. Da, L.T., Pardo-Avila, F., Xu, L., Silva, D.A., Zhang, L., Gao, X., Wang, D. and Huang, X. (2016) Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nat Commun*, **7**, 11244. 10.1038/ncomms11244.

50. R Development Core Team. (2005) R: A Language and Environment for Statistical Computing. R foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

51. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, **14**, 33-38.

52. Ismail, N., Hedman, R., Schiller, N. and von Heijne, G. (2012) A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat Struct Mol Biol*, **19**, 1018-1022. 10.1038/nsmb.2376.

53. Goldman, D.H., Kaiser, C.M., Milin, A., Righini, M., Tinoco, I. and Bustamante, C. (2015) Mechanical force releases nascent chain-mediated ribosome arrest *in vitro* and *in vivo*. *Science*, **348**, 457-460. 10.1126/science.1261909.

54. Kemp, G., Nilsson, O.B., Tian, P., Best, R.B. and von Heijne, G. (2020) Cotranslational folding cooperativity of contiguous domains of alpha-spectrin. *Proc Natl Acad Sci USA*, **117**, 14119-14126. 10.1073/pnas.1909683117.

55. Nilsson, O.B., Nickson, A.A., Hollins, J.J., Wickles, S., Steward, A., Beckmann, R., von Heijne, G. and Clarke, J. (2017) Cotranslational folding of spectrin domains via partially structured states. *Nat Struct Mol Biol*, **24**, 221-225. 10.1038/nsmb.3355.

56. Notari, L., Martinez-Carranza, M., Farias-Rico, J.A., Stenmark, P. and von Heijne, G. (2018) Cotranslational Folding of a Pentarepeat β-Helix Protein. *J Mol Biol*, **430**, 5196-5206. 10.1016/j.jmb.2018.10.016.

57. Farias-Rico, J.A., Ruud Selin, F., Myronidi, I., Frühauf, M. and von Heijne, G. (2018) Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proc Natl Acad Sci U S A*, **115**, E9280-E9287. 10.1073/pnas.1812756115.

58. Sandhu, H., Hedman, R., Cymer, F., Kudva, R., Ismail, N. and von Heijne, G. (2021) Cotranslational translocation and folding of a periplasmic protein domain in *Escherichia coli*. *J Mol Biol*. 10.1016/j.jmb.2021.167047.

59. Elfageih, R., Karyolaimos, A., Kemp, G., de Gier, J.W., von Heijne, G. and Kudva, R. (2020) Cotranslational folding of alkaline phosphatase in the periplasm of Escherichia coli. *Protein Sci*, **29**, 2028-2037. 10.1002/pro.3927.

60. Kemp, G., Kudva, R., de la Rosa, A. and von Heijne, G. (2019) Force-Profile Analysis of the Cotranslational Folding of HemK and Filamin Domains: Comparison of Biochemical and Biophysical Folding Assays. *J Mol Biol*, **431**, 1308-1314. 10.1016/j.jmb.2019.01.043.

61. Tian, P., Steward, A., Kudva, R., Su, T., Shilling, P.J., Nickson, A.A., Hollins, J.J., Beckmann, R., von Heijne, G., Best, R.B. *et al.* (2018) The folding pathway of an Ig domain is conserved on and off the ribosome. *Proc Natl Acad Sci USA*, , . doi.org/10.1073/pnas.1810523115.

62. Kudva, R., Tian, P., Pardo-Avila, F., Carroni, M., Best, R.B., Bernstein, H.D. and von Heijne, G. (2018) The shape of the bacterial ribosome exit tunnel affects cotranslational protein folding. *eLife*, **7**, e36326. 10.7554/eLife.36326.

63. Nicolaus, F., Metola, A., Mermans, D., Liljenström, A., Krc, A., Abdullahi, S.M., Zimmer, M., Miller Iii, T.F. and von Heijne, G. (2021) Residue-by-residue analysis of cotranslational membrane protein integration in vivo. *eLife*, **10**. 10.7554/eLife.64302.

64. Nicolaus, F., Ibrahimi, F., den Besten, A. and von Heijne, G. (2022) Upstream charged and hydrophobic residues impact the timing of membrane insertion of transmembrane helices. *FEBS Lett*, **596**, 1004-1012. 10.1002/1873-3468.14286.

65. Mermans, D., Nicolaus, F., Fleisch, K. and von Heijne, G. (2022) Cotranslational folding and assembly of the dimeric *E. coli* inner membrane protein EmrE. *Proc Natl Acad Sci USA*, **119**, e2205810119. 10.1073/pnas.2205810119.

66. Ismail, N., Hedman, R., Lindén, M. and von Heijne, G. (2015) Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat Struct Mol Biol*, **22**, 145-149. 10.1038/nsmb.2940.

67. Farias-Rico, J.A., Goetz, S.K., Marino, J. and von Heijne, G. (2017) Mutational analysis of protein folding inside the ribosome exit tunnel. *FEBS Lett*, **591**, 155-163. 10.1002/1873-3468.12504.

68. Wruck, F., Tian, P., Kudva, R., Best, R.B., von Heijne, G., Tans, S.J. and Katranidis, A. (2021) The ribosome modulates folding inside the ribosomal exit tunnel. *Commun Biol*, **4**, 523. 10.1038/s42003-021-02055-8.

69. Joseph, A.P., Srinivasan, N. and de Brevern, A.G. (2012) Cis–trans peptide variations in structurally similar proteins. *Amino Acids*, **43**, 1369-1381. 10.1007/s00726-011-1211-9.

70. Stewart, D.E., Sarkar, A. and Wampler, J.E. (1990) Occurrence and role of *cis* peptide bonds in protein structures. *J Mol Biol*, **214**, 253-260. https://doi.org/10.1016/0022-2836(90)90159-J.

71. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589. 10.1038/s41586-021-03819-2.

72. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, **292**, 195-202. 10.1006/jmbi.1999.3091.

73. Vazquez-Laslop, N., Ramu, H., Klepacki, D., Kannan, K. and Mankin, A.S. (2010) The key function of a conserved and modified rRNA residue in the ribosomal response to the nascent peptide. *EMBO J*, **29**, 3108-3117. 10.1038/emboj.2010.180.

74. Woolhead, C.A., Johnson, A.E. and Bernstein, H.D. (2006) Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol Cell*, **22**, 587-598. 10.1016/j.molcel.2006.05.021.

75. Lawrence, M.G., Lindahl, L. and Zengel, J.M. (2008) Effects on translation pausing of alterations in protein and RNA components of the ribosome exit tunnel. *J Bacteriol*, **190**, 5862-5869. 10.1128/jb.00632-08.

76. Guzel, P., Yildirim, H.Z., Yuce, M. and Kurkcuoglu, O. (2020) Exploring Allosteric Signaling in the Exit Tunnel of the Bacterial Ribosome by Molecular Dynamics Simulations and Residue Network Model. *Front Mol Biosci*, **7**, 586075. 10.3389/fmolb.2020.586075.

77. Nissley, D.A., Vu, Q.V., Trovato, F., Ahmed, N., Jiang, Y., Li, M.S. and O'Brien, E.P. (2020) Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling. *J Am Chem Soc*, **142**, 6103-6110. 10.1021/jacs.9b12264.

78. Petrone, P.M., Snow, C.D., Lucent, D. and Pande, V.S. (2008) Side-chain recognition and gating in the ribosome exit tunnel. *Proc Natl Acad Sci U S A*, **105**, 16549-16554. 10.1073/pnas.0801795105.

79. Bui, P.T. and Hoang, T.X. (2021) Hydrophobic and electrostatic interactions modulate protein escape at the ribosomal exit tunnel. *Biophys J*, **120**, 4798-4808. 10.1016/j.bpj.2021.09.027.

80. Wang, J., Karki, C., Xiao, Y. and Li, L. (2020) Electrostatics of Prokaryotic Ribosome and Its Biological Implication. *Biophys J*, **118**, 1205-1212. 10.1016/j.bpj.2020.01.014.

81. Joiret, M., Kerff, F., Rapino, F., Close, P. and Geris, L. (2022) Ribosome exit tunnel electrostatics. *Phys Rev E*, **105**, 014409. 10.1103/PhysRevE.105.014409.

82. Gurel, G., Blaha, G., Moore, P.B. and Steitz, T.A. (2009) U2504 determines the species specificity of the A-site cleft antibiotics: the structures of tiamulin, homoharringtonine, and bruceantin bound to the ribosome. *J Mol Biol*, **389**, 146-156. 10.1016/j.jmb.2009.04.005.

Figure legends

Figure 1. (A) Diagrammatic representation of constructs used. The Zinc-finger domain of ADR1a was engineered to the 8 amino acid arrest peptide of SecM from *Mannheimia succiniciproducens* via a 19 amino acid linker consisting of GS-repeats (GS-linker). 150 amino acids of the periplasmic domain of LepB were introduced at the N-terminus of ADR1 and 23 amino acids at the C-terminus of SecM to be able to resolve the arrested and full-length protein products by SDS-PAGE and autoradiography (B) A schematic of ribosomes stalled by the SecM AP and how an N-terminal pulling force generated by the folding of ADR1a can result in a resumption of translation. In the panel on the left, ADR1a (red) does not fold due to the absence of Zn^{2+} whereas ADR1a folds inside the ET in the presence of Zn^{2+} , generating a pulling force on the NC, in the panel on the right. (C) Autoradiographs of unfolded ($-Zn^{2+}$) and folded ($+Zn^{2+}$) ADR1a constructs stalled by the SecM(*Ms*) AP after radioactive pulse-labelling *in vitro* and SDS-PAGE. The linker length (*L*) corresponds to the 19 amino acid GS-repeats and the 7 amino acids of the SecM(*Ms*) AP (not counting the C-terminal Pro residue which is bound to the A-site tRNA during arrest). The relative amounts of arrested (*A*) and full-length (*FL*) product were estimated by quantification of the protein bands in the autoradiographs, and the fraction full-length was calculated as $f_{FL}=I_{FL}/(I_A+I_{FL})$. Two repeat experiments are shown for each construct. (D) Force profile analysis of constructs with (i) K; (ii) D; (iii) N; (iv) P; (v) W, and (vi) L at different positions within the Gly-Ser linker.

Figure 2. Molecular Dynamics setup. (A) Ribosome in a solvated dodecahedron box. The ribosome is shown as a cartoon, and the solvent as a grey surface. The exit tunnel is highlighted. The tRNA is shown as orange spheres, SecM(*Ms*) AP as red spheres, and GS-repeat linker as blue spheres. In the close-up, the nucleic acids within 15 Å of the nascent peptide are shown as

lines. (B) SecM AP and GS-linker sequences. Residues critical for stalling in the SecM(*Ms*) and SecM(*Ec*) APs are highlighted in red. Position -12 is highlighted in green; X is either Gly (in WT), Lys, or Asn. The *E. coli* SecM(*Ec*) AP sequence is shown for reference. In the Control sequence, the SecM(*Ms*) AP is replaced by an equally long GS-repeat sequence. (C) Close-up of the nascent peptide (WT sequence) inside the exit tunnel. The loops of the uL4 and uL22 proteins are shown in cyan as references. Positions -12 and -18 are shown as red spheres.

Figure 3. NC-ET interactions for residues S^{-11} (A), X^{-12} (B), and S^{-13} (C). Residues are defined as interacting when the minimum distance between them is less than 4 Å. The bar graphs show the percentage of the simulation time that the indicated residues interact. The interacting residues are shown in stick representations (left panels).

Figure 4. NC-ET interactions. (A) Cross-correlation between the residue of the nascent peptide in position -12 (for four systems) and the loops of proteins uL22 (residues G^{79} to T^{100}) and uL4 (residues K^{57} to S^{70}). The image on the right shows the loops of proteins uL22 (residues G^{79} to T^{100}) and uL4 (residues K^{57} to S^{70}). Each residue is represented as a sphere located at the center of mass of each residue. (B) The location within the ET of the nascent peptide residue at position -12. For each system, the average position of the loops of proteins uL4 and uL22 and the nascent peptide are shown in grey as cartoon representations. The SecM(*Ms*) AP residues are shown as cyan sticks, and the CCA terminus of the tRNA is shown as magenta sticks. To visualize the regions that the residue at position -12 visits during the simulation, this residue is shown as a CPK model (in orange) from overlapping frames of the sub-sampled trajectories.

Figure 5. Hydrogen bond networks near the PTC. (A-E) The structures show the nucleic acids (orange and green, NewRibbons VMD representation) that form hydrogen bonds present at least 50% of the simulation time. Grey dashed lines between residues represent hydrogen bonds

between residues. The loops of proteins uL4 (orange) and uL22 (pink), as well as the nascent chain (grey) are shown as tubes. Note that the frame selected for each system may show only some of all the possible hydrogen bonds. For a list of the hydrogen bonds present 50%, 75%, and 90% of the simulation time, refer to SI tables 1-3. To facilitate visualization, the nucleic acids that were involved in hydrogen bonds more than 50% of the time and were present in all the systems are shown on panel E and are not shown in panel A to D (A: WT; B: Control; C: K^{-12}; D: N^{-12}). (F) The h-bonds are localized in three major regions, as shown in the blue, green, and red sticks. The loops of proteins uL22 and uL4, tRNA CCA fragment, and nascent peptide are shown as ribbons. SecM residues R^{-3}, H^{-7}, and X^{-12} are shown as spheres. The viewpoint is different from the one shown for panels A to E. The structure was rotated ~90 degrees so the tRNA would come off the page towards the reader. The colors of the residue labels in panels A to E correspond to the three major regions highlighted in this panel.

Figure 6. Stacking between H^{-7} and U2609. (A) Two reaction coordinates are used to define stacking between H^{-7} and U2609 (left panel). The distance between the centre of mass (c.o.m.) of the H^{-7} sidechain and the base ring of U2609 is less than 0.5 nm, and the dihedral angle formed between the base ring of U2609 and the C2 atom of H^{-7} has a value between 70 and 110 degrees. All the frames were projected onto these two reaction coordinates (dihedral angle vs c.o.m. distance). (B) 2D histograms of the projections onto the reaction coordinates described in A. The red rectangle shows the region where the stacking conditions are met, and the percentage of frames that fall within this region is shown in the bottom right of each plot. (C) Extracted structures from the stacked and non-stacked regions highlighted in B. The aromatic rings of H^{-7} and U2609 are shown as blue and orange sticks, respectively.