

Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling

Daniel A. Nissley, Quyen V. Vu, Fabio Trovato, Nabeel Ahmed, Yang Jiang, Mai Suan Li, and Edward P. O'Brien*



Cite This: <https://dx.doi.org/10.1021/jacs.9b12264>



Read Online

ACCESS |



Metrics & More



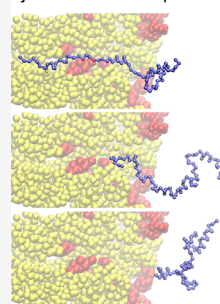
Article Recommendations



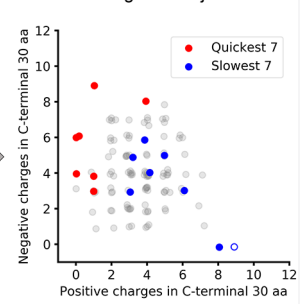
Supporting Information

ABSTRACT: The ejection of nascent proteins out of the ribosome exit tunnel after their covalent bond to transfer-RNA has been broken and has not been experimentally studied due to challenges in sample preparation. Here, we investigate this process using a combination of multiscale modeling, ribosome profiling, and gene ontology analyses. Simulating the ejection of a representative set of 122 *E. coli* proteins we find a greater than 1000-fold variation in ejection times. Nascent proteins enriched in negatively charged residues near their C-terminus eject the fastest, while nascent chains enriched in positively charged residues tend to eject much more slowly. More work is required to pull slowly ejecting proteins out of the exit tunnel than quickly ejecting proteins, according to all-atom simulations. An energetic decomposition reveals, for slowly ejecting proteins, that this is due to the strong attractive electrostatic interactions between the nascent chain and the negatively charged ribosomal-RNA lining the exit tunnel, and for quickly ejecting proteins, it is due to their repulsive electrostatic interactions with the exit tunnel. Ribosome profiling data from *E. coli* reveals that the presence of slowly ejecting sequences correlates with ribosomes spending more time at stop codons, indicating that the ejection process might delay ribosome recycling. Proteins that have the highest positive charge density at their C-terminus are overwhelmingly ribosomal proteins, suggesting the possibility that this sequence feature may aid in the cotranslational assembly of ribosomes by delaying the release of nascent ribosomal proteins into the cytosol. Thus, nascent chain ejection times from the ribosome can vary greatly between proteins due to differential electrostatic interactions, can influence ribosome recycling, and could be particularly relevant to the synthesis and cotranslational behavior of some proteins.

Ejection of nascent proteins



Electrostatics governs ejection times



INTRODUCTION

Translation is the process by which a protein is synthesized from an mRNA and is carried out by the ribosome molecular machine. The four phases of translation (initiation, elongation, termination, and ribosome recycling) are areas of intense research due to the essential role of protein synthesis in life. Each phase is composed of multiple steps, many of which have been characterized in terms of the structures adopted by the molecules involved, the mechanisms of conformational and chemical transitions, and the rates associated with these transitions.¹ Translation termination in *E. coli*, for example, consists of some four steps: the binding of a release factor to a stop codon in the A-site of the ribosome, the hydrolysis of the covalent bond connecting the C-terminus of the nascent chain to the P-site tRNA, the ejection of the nascent protein out of the exit tunnel, and the dissociation of the release factor from the ribosome (Figure 1). While rates for release factor binding and hydrolysis have been measured,^{2–6} the diffusion of the nascent chain out of the exit tunnel has not been experimentally characterized due to challenges with sample preparation. Additionally, the presumably fast time scales of

nascent chain ejection make experimental measurement a challenge.

The ribosome exit tunnel is composed of both rRNA and ribosomal protein and is therefore a chemically heterogeneous environment through which nascent proteins pass into the cytosol. At approximately 10 nm long and roughly 1.5 nm in diameter, the interactions between the exit tunnel and nascent chains can exhibit the full range of intermolecular forces, including charge–charge, hydrogen bonding, and hydrophobic interactions. Highly attractive forces can exist between some regions of the exit tunnel and some nascent peptide sequences.⁷ Indeed, peptide sequences known as stalling sequences have evolved to take advantage of these interactions and bind to the tunnel wall so tightly that they drastically slow

Received: December 2, 2019

Published: March 6, 2020



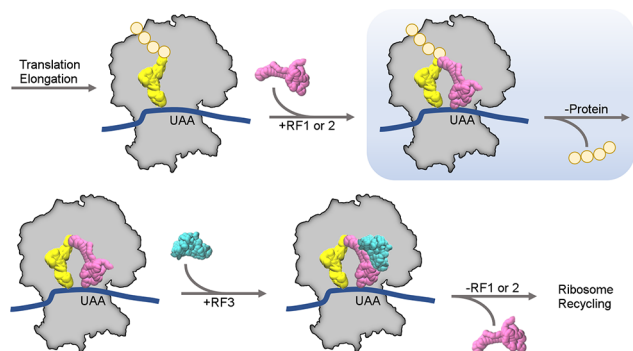


Figure 1. Translation termination in *E. coli*. Translation termination begins after translation elongation ends when a stop codon (e.g., UAA) enters the ribosome's A-site (ribosome shown as a gray outline). Release factor (RF) 1 or 2 (magenta) binds the stop codon in the A-site and catalyzes the hydrolysis of the peptidyl-tRNA bond between the nascent protein (orange spheres) and P-site tRNA (yellow). The nascent protein then diffuses out of the exit tunnel of the ribosome, which is around 10 nm in length. Following ejection, RF3 (cyan) catalyzes the release of RF1 or 2, allowing translation to proceed to the final phase, ribosome recycling. This study focuses on the second panel in this figure, highlighted in light blue. P-site tRNA and RF1 structures are generated from PDB ID 3OSK. The RF3 structure is generated from PDB ID 2HSE.

translation elongation,^{8,9} the biological benefit of which is to regulate downstream protein synthesis. When stretches of positively charged residues are present in the exit tunnel, they slow protein elongation under both in vitro¹⁰ and in vivo conditions. Since nascent protein elongation and nascent protein ejection both involve the passage of nascent protein segments through the exit tunnel, the interactions that can be large and impactful during elongation also have the potential to be important during ejection.

In this study, we use a combination of coarse-grained and all-atom simulations, ribosome profiling data, and gene ontology analysis to estimate the relative range of ejection time scales that can occur across the cytosolic proteome of *E. coli*, determine the intermolecular forces that give rise to the extremes of ejection times, find experimental evidence that slowly ejecting sequences can delay later stages of translation, and identify nascent ribosomal proteins as some of the slowest ejecting proteins from the ribosome.

79 ■ SIMULATION METHODS

Single-Domain Protein Selection and Model Building. The database from which the 122 proteins were selected contains 1014 cytosolic protein structures, 598 of which were single-domain proteins and the rest multidomain proteins.¹¹ A domain in this database is classified as either α or β if more than 70% of its residues identified by STRIDE¹² to be in secondary structural elements were in α -helices or β -strands, respectively. Domains that simultaneously had α -helical and β -strand content greater than 30% were classified as α/β .¹¹ Given the sequence length distribution of these proteins, we determined that we could feasibly simulate the synthesis and ejection of 122 proteins in total. To maintain the ratio of single- to multidomain proteins in the database, we selected 72 single-domain proteins and 50 multidomain proteins. Of the 598 single-domain proteins in the database there are 250 α , 55 β , and 293 α/β domains; this ratio of structural classes was reproduced in the subset of 72 single-domain proteins by randomly selecting 30 α , 7 β , and 35 α/β proteins (Tables S1 and S2). PDB files for each single-domain protein were retrieved from the Protein Data Bank,¹³ and their corresponding mRNA sequences (NCBI assembly eschColi_K12) were retrieved using the University of California Santa

Cruz microbe table browser (<http://microbes.ucsc.edu/>). Randomly selected PDBs from the database were accepted only if the crystallized sequence had no amino acid mutations in comparison to the amino acid sequence which would result from the translation of the eschColi_K12 mRNA. However, small sections of amino acids (12 or less) or small numbers of heavy atoms (less than 10) that were not resolved in the experimental structure were rebuilt on the basis of the reference genome sequence and minimized in CHARMM.¹⁴

Multidomain Protein Selection and Model Building. Fifty multidomain proteins were selected randomly from the same previously published database of *E. coli* globular proteins from which single-domain proteins were selected.¹¹ These multidomain proteins are listed in Table S3. The amino acid sequence of each PDB was aligned to the translated sequence of the corresponding gene in NCBI assembly eschColi_K12, and missing residues and domains were identified. Some PDBs contained large missing sections; to fill in these regions with reasonable structures, PDBs representing the same gene product were used to reconstruct missing sections after structural alignment in VMD.¹⁵ PREDATOR¹⁶ and IUPRED¹⁷ were used to predict whether those residues not resolved in any other PDB were intrinsically unstructured, in which case they were rebuilt and minimized in CHARMM rather than templated using other structures. When homologous structures from *E. coli* were not available, homologous structures from other organisms were used as a template for the protein model, provided the sequence similarity was greater than 30% and the backbone RMSD between the regions common to the two structures was ≤ 2 Å.

Reconstructing missing domains or sections of the multidomain proteins in this way resulted in models that still had, in some cases, sections of missing atoms or mismatched amino acids relative to the consensus mRNA sequence. All proteins were therefore subjected to a rebuilding phase to add missing atoms and correct mutations or sequence mismatches (Table S4). Because the reconstructed segments were generated in an extended conformation, minimization was performed in vacuo for 200 steps. This short minimization was sufficient to resolve steric clashes. For proteins with short stretches of missing residues (less than 10), this minimized configuration was accepted as the final atomistic model. If a protein contained one or more long stretches of missing residues (more than 10) or disconnected domains, then the minimized protein structure was subjected to additional dynamics at 310 K. In this phase, the reconstructed atoms within each templated domain were left free to move, thereby allowing the structure to locally equilibrate. The smallest domain in each protein was also left unrestrained in order to allow it to reorient with respect to all other domains into a favorable conformation. All other atoms were either held fixed or harmonically restrained to the experimentally solved structure with a force constant of 1 kcal/(mol·Å²). All reconstructions, minimizations, and molecular dynamics simulations were performed using CHARMM with the par27 force field.¹⁴ The minimized structures were solvated in TIP3P¹⁸ water and 150 mM NaCl, gradually heated to 310 K for 100 ps, and then equilibrated for 1.5 ns at the same temperature. Production runs had different durations, spanning from 20 to 50 ns. Langevin dynamics with a friction coefficient of 1.0 ps⁻¹ and a time step of 1.5 fs were used. For each protein, the conformation with the lowest potential energy was selected as the final atomistic model. We emphasize that the purpose of the molecular dynamics simulations was not to thoroughly explore the conformational space of the multidomain proteins but rather to provide reasonable atomistic conformations for building coarse-grained models.

Domains in the multidomain proteins were initially defined according to CATH,¹⁹ which is also used in the original database.¹¹ Domain residue numberings were shifted to match the translated sequence and modeling of the missing atoms performed (Table S4). The final domain definitions reported in Table S3 include residues and domains that were modeled as described in Table S4.

Coarse-Grained Force Field and Model Construction. The potential energy for a given configuration of the C_α coarse-grained model is calculated using the equation

$$\begin{aligned}
 E = & \sum_i k_b(r_i - r_0)^2 + \sum_{i,j=1}^4 k_{\varphi,ij}(1 + \cos[j\varphi_i - \delta_{ij}]) \\
 & + \sum_i -\frac{1}{\gamma} \ln\{\exp[-\gamma(k_\alpha(\theta_i - \theta_\alpha)^2 + \epsilon_\alpha)] + \exp[-\gamma k_\beta(\theta_i - \theta_\beta)^2]\} \\
 & + \sum_{ij} \frac{q_i q_j e^2}{4\pi\epsilon_0\epsilon_r r_{ij}} \exp\left[-\frac{r_{ij}}{l_D}\right] + \sum_{ij \in \{NC\}} \epsilon_{ij}^{NC} \left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] \\
 & + \sum_{ij \notin \{NC\}} \epsilon_{ij}^{NN} \left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right]
 \end{aligned}$$

The terms in this equation represent, from left to right, summations over the contributions from C_α – C_α bonds, dihedral angles, bond angles, electrostatic interactions, Lennard-Jones-like native interactions, and repulsive non-native interactions to the total potential energy. The bond, dihedral, and angle terms have been described in detail elsewhere.^{20,21} Electrostatics are treated using Debye–Hückel theory with a Debye length, l_D , of 10 Å and a dielectric of 78.5; lysine and arginine C_α sites are assigned $q = +e$, glutamic acid and aspartic acid are assigned $q = -e$, and all other interaction sites are uncharged.²² The contribution from native interactions is computed using the 12–10–6 potential of Karanicolas and Brooks.²⁰ The value of ϵ_{ij}^{NC} , which sets the depth of the energy minimum for a native contact, is calculated as $\epsilon_{ij}^{NC} = n_{ij}\epsilon_{HB} + \eta\epsilon_{ij}$. Here, ϵ_{HB} , and ϵ_{ij} represent energy contributions arising from hydrogen bonding and van der Waals contacts between residues i and j identified from the all-atom structure of the protein, respectively. n_{ij} is the number of hydrogen bonds formed between residues i and j and $\epsilon_{HB} = 0.75$ kcal/mol. The value of ϵ_{ij} is set on the basis of the Betancourt–Thirumalai pairwise potential.²³ The scaling factor η is determined for each of our 122 proteins (Tables S2 and S3) based on a previously published training set²⁴ to reproduce realistic protein stabilities for different structural classes (Table S5, Table S6, and Supplementary Methods). A single value of η is applied to all native contacts for a given single-domain protein and for each individual domain and interface in multidomain proteins. Collision diameters, σ_{ij} , between C_α interactions sites involved in native contacts are set equal to the distance between the C_α of the corresponding residues in the crystal structure divided by $2^{1/6}$. For non-native interactions, ϵ_{ij}^{NN} is set to 0.000132 kcal/mol and σ_{ij} is computed as previously reported.²⁰

Simulations of Nascent Protein Synthesis and Ejection. The coarse-grained model of each protein in the single- and multidomain protein data sets was synthesized starting from a single residue (see the two exceptions below) using a modified version of a previously published protocol on a coarse-grained representation of the 50S *E. coli* ribosome (details of the ribosome model can be found in Supplementary Methods).²⁵ The dwell time at a particular nascent chain length was randomly selected from an exponential distribution with a mean equal to the average decoding time of the codon in the A-site. Average decoding times are taken from the Fluitt-Viljoen model²⁶ and scaled to reproduce an overall average of 12.6 ns (840 000 integration time steps of 0.015 ps duration; see Table S7) based on a previously published training set.²⁴ A planar restraint in the yz plane through the point (58, 0, 0) Å is used to prevent the nascent chain from contacting the underside of the ribosome cutout. Fifty trajectories were run for each of the 122 proteins in the data set. After synthesis was completed for a given trajectory, the harmonic restraint on the C-terminal bead to model the covalent bond between the nascent protein and the P-site tRNA was removed. Simulations of termination were run until the C-terminal residue of each trajectory reached an x coordinate of 100 Å or greater, indicating that the protein exited the tunnel. Ejection times are calculated as the time between when the C-terminal harmonic restraint is removed and when the C-terminal residue reaches an x coordinate of ≥ 100 Å. Two proteins (PDB IDs 2KFW and 3GN5) became stalled in the exit tunnel when synthesis was begun from a single residue; synthesis for these two proteins was therefore initiated from a nascent chain length of 50 residues. One protein (PDB ID 4DCM) did not eject from the

exit tunnel in 27 of 50 trajectories during 25 days of CPU time when its wild-type C-terminal charges were used; the ejection time for this protein is therefore reported as a lower bound. Mean ejection times for all 122 proteins are listed in Table S8.

All-Atom Steered Molecular Dynamics Simulations. The 50S subunit of the *E. coli* ribosome (PDB ID 3R8T) was aligned with the long axis of the exit tunnel, defined to be between atom N6 of nucleotide A2602 and the C_β atom of Ala50 in ribosomal protein L24, along the x axis of the simulation coordinate system. The ribosome was then cropped to form a rectangular box around the exit tunnel with dimensions of $13.10590 \times 8.44869 \times 8.18680$ nm³. Coarse-grained structures of the C-terminal 30 aa of nascent proteins from the final time step of synthesis simulations were backmapped to atomistic resolution for use as starting structures. The first step in backmapping is the insertion of coarse-grained sites representing amino acid side chains near their corresponding C_α beads followed by energy minimization in the C_α side-chain model force field²⁷ with all C_α positions restrained. Backbone and side-chain all-atom structures were then rebuilt using Prodact2²⁸ and Pulchra,²⁹ respectively, on the minimized C_α side-chain model. The final backmapped structure was obtained after energy minimization within the generalized Born (GB) implicit water environment.³⁰ The N-terminus of the segment was capped by the N-terminal acetyl capping group (ACE) and the atomistic protein structure inserted into the atomistic exit tunnel structure.

A simulation box was constructed with a minimum of 1 nm between the edge of the cropped ribosome and the periodic boundary wall in all dimensions and then extended 15 nm in the positive x dimension to accommodate the nascent protein when fully extracted from the exit tunnel at the end of the steered molecular dynamics simulation. The system was neutralized with Na⁺ before adding 5 mM MgCl₂ and 100 mM NaCl. Next, the system was minimized in the gas phase with the steepest-descent algorithm. Harmonic restraints on all C_α atoms of the nascent peptide and all heavy atoms of the ribosome with a force constant of 1000 kJ/(mol·nm²) were employed to prevent the nascent protein from moving during minimization. The system was then equilibrated in the gas phase for 300 ps to allow ions to rapidly find binding sites on the ribosome, with harmonic restraints again applied to C_α atoms of the nascent chain and all heavy atoms of the ribosome.

The cropped ribosome and nascent protein were then solvated and equilibrated. First, 1 ns of dynamics was carried out in the NVT ensemble, followed by 1 ns of dynamics in the NPT ensemble, with the temperature and pressure held at 310 K and 1 atm, respectively. To allow the nascent protein and the ribosome exit tunnel to reach equilibrium in the all-atom model, we performed a second NPT simulation for 10 ns with harmonic restraints applied to P and C_α atoms of the ribosome that were more than 28 Å from the x axis and all C_α atoms of the nascent protein. The center of mass of the N-terminal ACE residue was then pulled from the exit tunnel with a cantilever speed of 0.25, 1, or 5 nm/ns and a spring constant of 600 kJ/(mol·nm²). All simulations were carried out with GROMACS 2018³¹ using the AMBER99SB³² force field and the TIP3P¹⁸ water model. The particle mesh Ewald method³³ was used to calculate the long-range electrostatic interactions beyond 1.2 nm. Lennard-Jones interactions were calculated within a distance of 1.2 nm. The Nose–Hoover thermostat^{34,35} and Parrinello–Rahman barostat³⁶ were employed to maintain the temperature and pressure at 310 K and 1 atm, respectively. The LINCS algorithm³⁷ was used to constrain all bonds, and the integration time step was set to 2 fs.

Simulations were carried out using this protocol for 5 quickly ejecting proteins (PDB IDs 1FM0, 1Q5X, 1T8K, 2KFW, and 3BMB) and 5 slowly ejecting proteins (PDB IDs 1AH9, 1JW2, 2JO6, 2PTH, and 3IVS) using 21 different initial configurations from the coarse-grained synthesis simulations for each different protein.

RESULTS AND DISCUSSION

To estimate the range of nascent chain ejection times across different proteins, we simulated the synthesis and ejection of

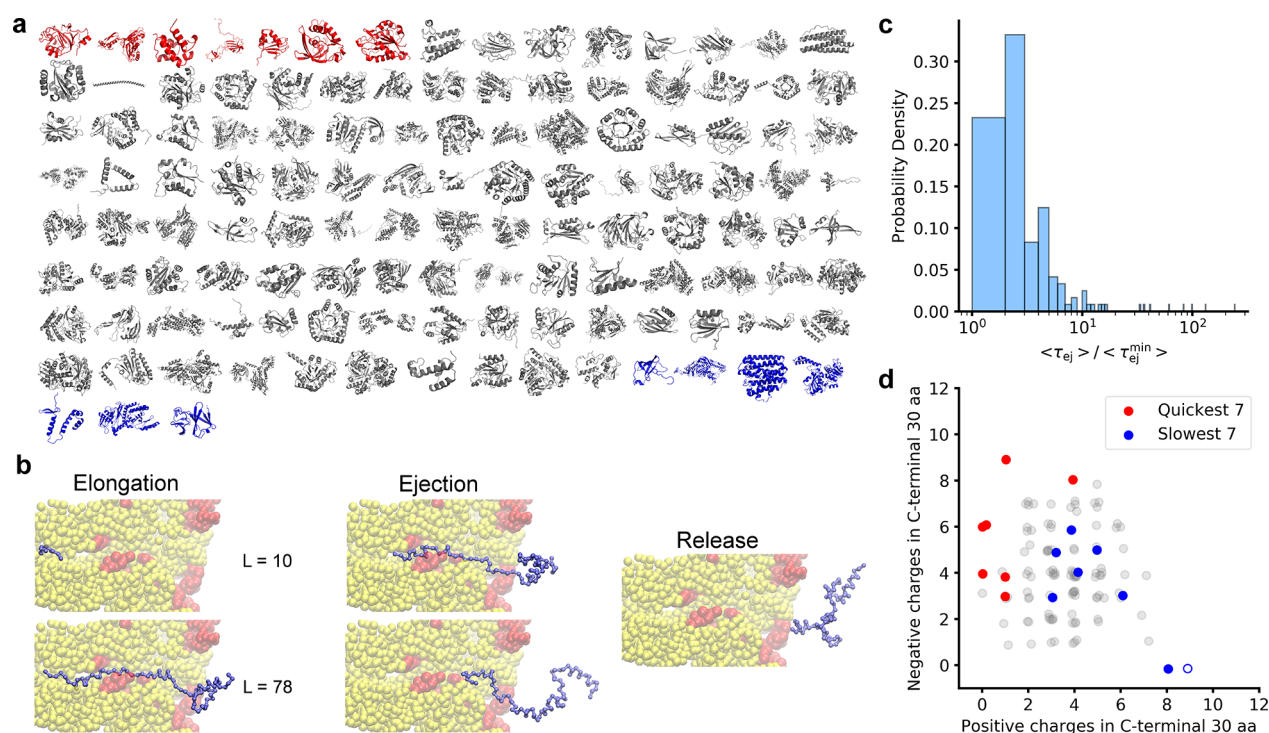


Figure 2. The 242-fold variation in ejection times is related to the presence of charged residues in the C-termini of proteins. (a) The set of 122 *E. coli* proteins that were simulated is shown from top left to bottom right by increasing ejection time. The top 5% fastest and slowest ejecting proteins are colored red and blue, respectively, while the middle 90% are colored gray. (b) Coarse-grained simulations begin with the elongation phase, during which the protein (blue) is synthesized on the ribosome (rRNA and protein colored yellow and red, respectively). Once the full-length protein is synthesized, ejection occurs. Ejection is complete once the C-terminal residue is 100 Å from the peptidyl transferase center of the ribosome. (c) Distribution of mean ejection times, $\langle \tau_{ej} \rangle$, for 121 of the 122 proteins shown in (a) (excluding 4DCM), normalized by the factor $\langle \tau_{ej}^{min} \rangle$ which is the smallest $\langle \tau_{ej} \rangle$ found in the set of proteins. (d) Number of positive and negative charges in each protein's C-terminal 30 residues. The fastest and slowest ejectors (bottom and top 5% of the distribution in panel c) are colored red and blue, respectively, and exist as separate, nonoverlapping populations along these metrics. Values from other proteins are displayed in transparent gray. Random noise (jitter) has been added to minimize overlapping points. As discussed in the main text, the single unfilled blue data point at 9 positive charges and zero negative charges is for PDB ID 4DCM, for which an exact ejection time could not be calculated because not all of its trajectories were released from the ribosome in the simulation. Therefore, this protein was excluded from the distribution in panel c.

122 full-length *E. coli* proteins using a coarse-grained representation of the ribosome nascent chain complex (Figure 2a,b).^{25,27,38,39} This set of 122 proteins is representative of the globular *E. coli* cytosolic proteome as a whole because it reproduces the proteome-wide protein size and structural class distributions (Figure S1, Table S1, and Simulation Methods). Fifty statistically independent synthesis and ejection trajectories were run for each protein. We find that the mean protein ejection time, defined as the average time it takes for the C-terminal nascent chain residue to reach the end of the exit tunnel after the bond between the protein and P-site tRNA is broken, varies 242-fold across 121 of these proteins (Figure 2c). We observed that the proteins at the extremes of this ejection time distribution, that is, those proteins in the top and bottom 5%, have markedly different electrostatic characteristics in their C-termini (Figure 2d), the last 30 residues of which are in the exit tunnel. Quickly ejecting proteins tend to have abundant negatively charged residues and few positively charged residues in their C-terminal 30 residues (red dots in Figure 2d). In contrast, slowly ejecting proteins tend to have fewer negatively charged residues and more positively charged residues (blue dots in Figure 2d). We note the exceptional case of the protein with PDB ID 4DCM (unfilled blue point in Figure 2d), which is the 122nd protein in our set. (Note that complete protein names are provided in Tables S2 and S3.) While complete ejection occurred for all other proteins, only

23 out of the 50 simulation trajectories of 4DCM were fully ejected from the exit tunnel. Under a conservative estimate, this protein's average ejection time is 7031-fold slower than the fastest ejecting protein in our data set. Consistent with electrostatics being important, 4DCM also has the greatest positive charge density in our set of proteins. These results indicate that there is a 3-order-of-magnitude spread in ejection times across *E. coli* cytosolic proteins and suggest that the very fast ejectors are fast because they are electrostatically repelled by the exit tunnel, which is lined with negatively charged rRNA, while the very slow ejectors are slow because they are electrostatically attracted to the exit tunnel wall.

To test this electrostatic hypothesis within our coarse-grained model, we set to zero all negative or positive charges of amino acids in the C-terminal 30 residues of the quickly or slowly ejecting proteins, respectively. All other interactions and charges involving the ribosome and nascent chain remained the same. Rerunning the ejection simulations for these sequences, we find that removing negative charges from the set of quick ejectors slowed the ejection process by 5–98% (average 44%) and removing positive charges from the set of slowly ejecting proteins sped up the ejection process by 48–99% (average 82%, 4DCM results excluded) (Table 1). These results are consistent with the hypothesis that electrostatic interactions are a causal factor in influencing extremely fast or extremely slow ejection times out of the ribosome tunnel.

Table 1. Ejection Times upon Neutralization of C-Terminal Positive or Negative Residues

quickly ejecting peptides			
PDB ID	wild type (ns)	no (–) charges (ns) ^a	% change
1Q5X	0.31	0.47	51.9
3M7M	0.31	0.41	30.8
1T8K	0.32	0.42	31.9
2KFW	0.32	0.41	27.3
1FM0	0.36	0.38	4.55
3BMB	0.37	0.57	51.8
1AG9	0.39	0.76	95.8
2HGK	0.40	0.65	62.3
1FJJ	0.41	0.44	8.06
2HO9	0.41	0.60	43.8
1SVT	0.42	0.50	19.4
1SG5	0.45	0.88	98.0
slowly ejecting peptides			
PDB ID	wild type (ns)	no (+) charges (ns) ^a	% change
4IM7	3.92	0.69	–82.3
1JW2	4.39	1.94	–55.9
2PTH	4.75	0.86	–81.9
1D2F	5.02	0.64	–87.2
3OFO	10.22	5.28	–48.3
1AH9	11.28	0.59	–94.8
1NG9	12.66	2.31	–81.8
1RQJ	18.80	0.84	–95.5
1T4B	25.66	2.58	–90.0
3IVS	30.47	0.45	–98.5
1U0B	40.75	1.04	–97.5
2JO6	74.73	23.31	–68.8
4DCM	>2170	0.54	–100.0

^aColumns labeled “no (–) charges” and “no (+) charges” are ejection times from simulations in which negative or positive charges in C-terminal 30 aa, respectively, are made electrically neutral.

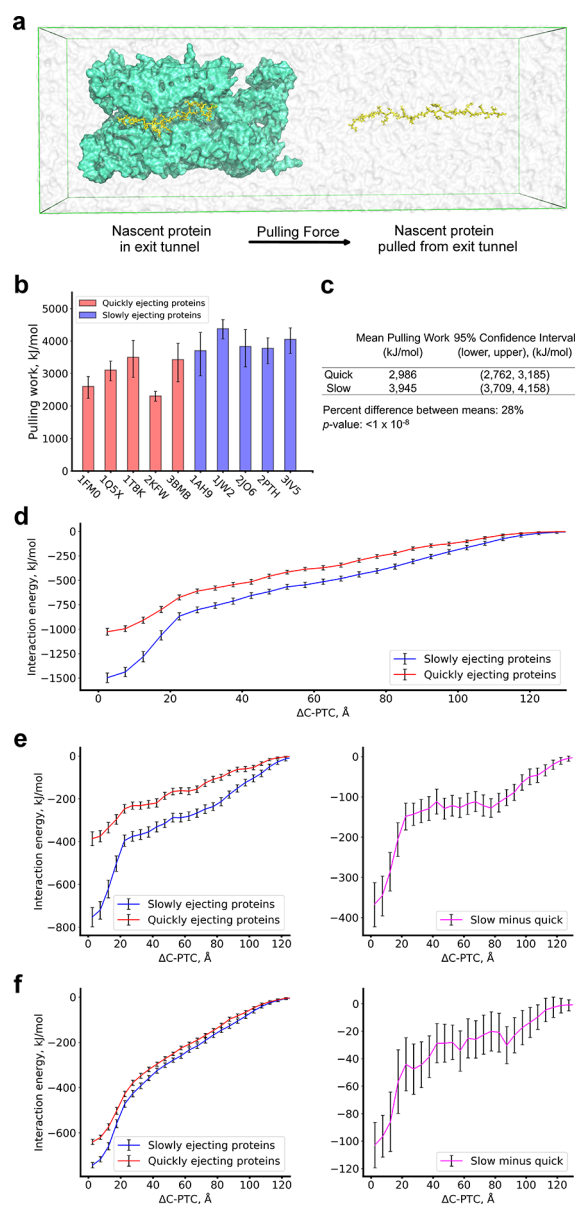


Figure 3. Slowly ejecting proteins are more electrostatically attracted to the ribosome exit tunnel. (a) Initial (left) and final (right) conformations from all-atom, steered molecular dynamics simulations of the extraction of a nascent protein (yellow) from the ribosome (cyan). (b) Mean pulling work required to extract 10 different nascent proteins from the ribosome exit tunnel from 21 statistically independent simulations per protein. Error bars are 95% confidence intervals calculated by bootstrapping. (c) Results from the statistical comparison between the overall means of the slowly and quickly ejecting sets. Confidence intervals are calculated as in (b). The p value is estimated using a permutation test. (d) Total interaction energy between the ribosome and nascent protein as a function $\Delta C - PTC$, the distance between the C_{α} atom of the C-terminal residue of the nascent protein and the N6 atom of nucleotide A2602 in the peptidyl transferase center of the ribosome. (e) Electrostatic contribution to the total interaction energy (left) and the difference between the slowly and quickly ejecting data set mean electrostatic interaction energies (right). (f) The same as in (d) but for the van der Waals interaction energy. These results were obtained using a cantilever speed of 0.25 nm/ns.

the intermolecular interactions in these simulations, we find that slowly ejecting nascent proteins have stronger interactions

with the ribosome tunnel wall than quickly ejecting proteins (Figure 3d–f), with the majority of this energy difference due to electrostatic rather than van der Waals interactions (Figure 3e,f). Qualitatively equivalent results are obtained when the cantilever speed is increased 4- or 20-fold to 1 or 5 nm/ns, respectively (Figures S2 and S3). Thus, the all-atom results and coarse-grained results are consistent, lending further support to the hypothesis that electrostatic interactions between the nascent chain and ribosome govern the extremes of nascent chain ejection times.

An important biological question is whether there are any downstream consequences of this broad range of ejection times. We hypothesized that the slowest ejecting sequences might delay the onset of the next and final step of translation (Figure 1), ribosome recycling, during which molecular factors interact with the ribosome to aid the dissociation of the small and large ribosomal subunits. This hypothesis predicts that ribosomes will dwell for longer at stop codons when a slowly ejecting sequence is present compared to when a quickly ejecting sequence is present. To test this hypothesis, we analyzed ribosome profiling data from *E. coli*⁴⁰ of those cytosolic proteins that have the highest charge density at their C-terminus. Ribosome profiling is an experimental technique that measures a signal, called the “reads”, that is proportional to the number of ribosomes sitting at a particular codon position on the various cellular copies of an mRNA transcript.⁴¹ As such, the greater the normalized ribosome density at a codon, the longer the ribosome spent at that codon position. The normalized ribosome density at a codon position is the number of reads at that codon divided by the average number of reads per codon arising from the coding sequence of the transcript. Therefore, our hypothesis predicts that there will be greater ribosome density at the stop codon for proteins that have the highest number of positive charges in their C-terminus compared to those that have high negative charge density. Therefore, we restricted our analysis to high-coverage transcripts (Supplementary Methods) encoding proteins with either ≥ 8 positive and ≤ 2 negative residues in their 30 C-terminal residues, which we predict to be slowly ejecting proteins ($n = 22$ proteins), or with ≥ 8 negative residues and ≤ 2 positively charged residues ($n = 22$ proteins) in their 30 C-terminal residues, which we predict to be quickly ejecting proteins. We could not include all of the fastest and slowest ejecting proteins from our simulations in this analysis because the read coverage of their transcripts was very sparse in the ribosome profiling data, meaning that their signal-to-noise ratio is too low to be useful. However, two proteins (PDB IDs 1T8K and 3IV5) for which we simulated ejection times did have sufficient read coverage and are included in this analysis. We observe that the putative slow ejectors have on average 3.3-fold higher ribosome density at the stop codon compared to the fast ejectors (Figure 4a,b; median ribosome densities across fast and slow ejector sets are, respectively, 0.248 and 0.812; the difference between medians is significant based on the Mann–Whitney U Test, $p = 0.011$). These results are consistent with the hypothesis that the slowest ejecting nascent chains tend to delay ribosome recycling.

To further explore the potential biological ramifications of very fast or slow ejection times, we carried out a gene ontology analysis to determine whether putative slowly and quickly ejecting proteins are more likely than random chance to be associated with particular cellular or biochemical processes (Supplementary Methods). The putative quickly ejecting

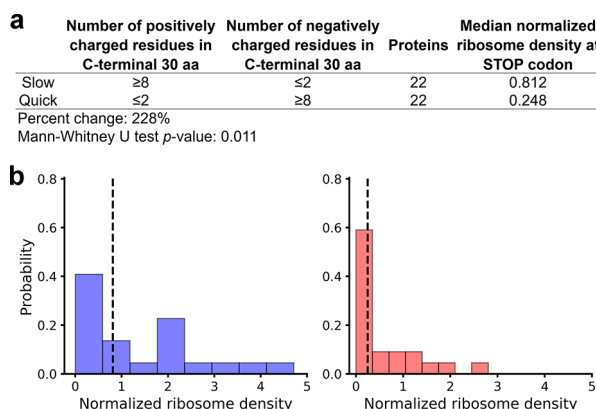


Figure 4. Presence of slowly ejecting proteins in ribosome-nascent chain complexes correlated with longer ribosome dwell times at the stop codon. (a) Proteins with ≥ 8 positive and ≤ 2 negative residues in their C-terminus as well as proteins with ≥ 8 negative and ≤ 2 positive residues in their C-terminus were selected from the *E. coli* ribosome profiling data from ref 40. A total of 22 proteins fit into each category. The median normalized ribosome density is higher for proteins enriched in positive charge at the C-terminus (p value 0.011, Mann–Whitney U test). (b) Histograms of normalized ribosome density at the stop codon for the subsets of proteins enriched in positive (left, blue histogram) or negative (right, red histogram) amino acids.

proteins exhibit no significant relationship to any particular biological processes. However, 14 of the 22 potentially slowly ejecting proteins are associated with translation, and 13 are ribosomal subunit proteins. Two hypotheses can explain this observation. First, slow ejection increases the time a nascent protein is available for cotranslational assembly,^{42–44} suggesting that these highly positively charged C-terminal segments might have evolved to aid in the efficient cotranslational assembly of ribosomes in the *E. coli* cytosol. Second, ribosomal proteins may have evolved positively charged segments solely to aid in their interactions with rRNA in the context of a fully assembled ribosomal subunit, with their slow ejection times being a biologically irrelevant consequence of this fact. Indeed, each of the 13 ribosomal proteins identified by this analysis is in contact with rRNA based on the analysis of a crystal structure of the *E. coli* ribosome in the nonrotated conformation (PDB ID 4V9D). It will be an interesting area of future research to test these distinct hypotheses.

CONCLUSIONS

Our results indicate that nascent protein ejection times are very broad, that the extremes are primarily driven by interactions of the high charge density of either positive or negative residues near the nascent protein’s C-terminus with the negatively charged ribosome exit tunnel, and that very slowly ejecting chains can delay ribosome recycling. The fact that ribosomal proteins have some of the most highly positively charged C-termini across the *E. coli* proteome suggests the intriguing possibility that their charge density did not evolve just to strengthen their binding affinity for rRNA but could also be beneficial by making them slow ejectors, thereby affording more time for potential cotranslational assembly processes to occur. While we have demonstrated that electrostatics are essential for extreme ejection times by running simulations without the charges present in the nascent chain C-termini, other factors must also play a role in determining ejection times. As can be seen in Figure 2d, some

proteins with typical ejection times (gray dots) have a similar number of positive charges in the C-terminus as proteins with very slow ejection times. We speculate that other factors that influence the ejection time could include the backbone structural propensity and the size of the amino acids in the protein sequence. Helical backbone preferences and large amino acids are more likely to sterically clash with the walls of the exit tunnel, while extended strand backbone preferences and small amino acids might make diffusion out of the tunnel sterically easier. This study is the first to our knowledge to provide evidence that the seemingly mundane act of diffusion of nascent proteins out of the exit tunnel can vary greatly between proteins, have downstream cellular consequences, and might be particularly biologically relevant to the synthesis of ribosomal proteins.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.9b12264>.

Full details of coarse-grained model parametrization and simulations, additional all-atom steered molecular dynamics results with different cantilever speeds, and GO pathway analysis details (PDF)

Fast ejection (AVI)

Intermediate ejection (AVI)

Slow ejection (AVI)

■ AUTHOR INFORMATION

Corresponding Author

Edward P. O'Brien – Department of Chemistry, Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, and Institute for Computational and Data Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, United States; orcid.org/0000-0001-9809-3273; Email: epo2@psu.edu

Authors

Daniel A. Nissley – Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States

Quyen V. Vu – Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland; orcid.org/0000-0002-9863-0486

Fabio Trovato – Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States

Nabeel Ahmed – Department of Chemistry and Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, United States

Yang Jiang – Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States; orcid.org/0000-0003-1100-9177

Mai Suan Li – Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland; Institute for Computational Sciences and Technology, Ho Chi Minh City, Vietnam; orcid.org/0000-0001-7021-7916

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/jacs.9b12264>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

E.P.O. acknowledges support from the National Science Foundation (MCB-1553291) for the simulation component and (ABI-1759860) for the bioinformatics component of this study, as well as the National Institutes of Health (R35-GM124818). Portions of numerical computations and data analysis in this work have been carried out on the CyberLAMP cluster, which is supported by NSF-MRI-1626251 and operated by the Institute for CyberScience at The Pennsylvania State University. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. M.S.L. acknowledges that this work was supported by Narodowe Centrum Nauki (grant no. 2015/19/B/ST4/02721), PLGrid Infrastructure in Poland, and the Department of Science and Technology, Ho Chi Minh City, Vietnam.

■ REFERENCES

- (1) Rodnina, M. V. Translation in Prokaryotes. *Cold Spring Harbor Perspect. Biol.* **2018**, *10*, No. a032664.
- (2) Trappl, K.; Mathew, M. A.; Joseph, S. Thermodynamic and Kinetic Insights into Stop Codon Recognition by Release Factor 1. *PLoS One* **2014**, *9*, e94058.
- (3) Pierson, W. E.; et al. Uniformity of peptide release is maintained by methylation of release factors. *Cell Rep.* **2016**, *17*, 11–18.
- (4) Zavialov, A. V.; Mora, L.; Buckingham, R. H.; Ehrenberg, M. Release of Peptide Promoted by the GGQ Motif of Class 1 Release Factors Regulates the GTPase Activity of RF3. *Mol. Cell* **2002**, *10*, 789–798.
- (5) Kuhlkoetter, S.; Wintermeyer, W.; Rodnina, M. V. Different substrate-dependent transition states in the active site of the ribosome. *Nature* **2011**, *476*, 351–355.
- (6) Shoemaker, C. J.; Green, R. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, E1392–E1398.
- (7) Petrone, P. M.; Snow, C. D.; Lucent, D.; Pande, V. S. Side-chain recognition and gating in the ribosome exit tunnel. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16549–16554.
- (8) Murakami, A.; Nakatogawa, H.; Ito, K. Translation arrest of SecM is essential for the basal and regulated expression of SecA. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12330–12335.
- (9) Gumbart, J.; Schreiner, E.; Wilson, D. N.; Beckmann, R.; Schulten, K. Mechanisms of SecM-mediated stalling in the ribosome. *Biophys. J.* **2012**, *103*, 331–341.
- (10) Lu, J.; Deutsch, C. Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. *J. Mol. Biol.* **2008**, *384*, 73–86.
- (11) Ciryam, P.; Morimoto, R. I.; Vendruscolo, M.; Dobson, C. M.; O'Brien, E. P. In vivo translation rates can substantially delay the cotranslational folding of the Escherichia coli cytosolic proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E132–E140.
- (12) Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 566–579.
- (13) Berman, H. M.; et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (14) Brooks, B. R.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (15) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (16) Frishman, D.; Argos, P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng., Des. Sel.* **1996**, *9*, 133–142.
- (17) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433–3434.

- (18) Jorgensen, W. L.; et al. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (19) Sillitoe, I.; et al. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res.* **2013**, *41*, D490–D498.
- (20) Karanicolas, J.; Brooks, C. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **2002**, *11*, 2351–2361.
- (21) Best, R. B.; Chen, Y. G.; Hummer, G. Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of Arc repressor. *Structure* **2005**, *13*, 1755–1763.
- (22) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Trigger factor slows Co-translational folding through kinetic trapping while sterically protecting the nascent chain from aberrant cytosolic interactions. *J. Am. Chem. Soc.* **2012**, *134*, 10920–10932.
- (23) Betancourt, M. R.; Thirumalai, D. Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **1999**, *8*, 361–369.
- (24) Leininger, S. E.; Trovato, F.; Nissley, D. A.; O'Brien, E. P. Domain topology, stability, and translation speed determine mechanical force generation on the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 5523–5532.
- (25) Nissley, D. A.; O'Brien, E. P. Structural Origins of FRET-Observed Nascent Chain Compaction on the Ribosome. *J. Phys. Chem. B* **2018**, *122*, 9927–9937.
- (26) Fluit, A.; Pienaar, E.; Viljoen, H. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.* **2007**, *31*, 335–346.
- (27) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13403–13408.
- (28) Moore, B. L.; Kelley, L. A.; Barber, J.; Murray, J. W.; Macdonald, J. T. High-Quality Protein Backbone Reconstruction from Alpha Carbons Using Gaussian Mixture Models. *J. Comput. Chem.* **2013**, *34*, 1881–1889.
- (29) Rotkiewicz, P.; Skolnick, J. Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* **2008**, *29*, 1460.
- (30) Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2000**, *56*, 275–291.
- (31) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E.; et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.
- (32) Hornak, V.; et al. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712–725.
- (33) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.
- (34) Nosé, S. A unified formulation of the constant temperature molecular dynamics. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (35) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.
- (36) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (37) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (38) Fritch, B.; et al. Origins of the Mechanochemical Coupling of Peptide Bond Formation to Protein Synthesis. *J. Am. Chem. Soc.* **2018**, *140*, 5077–5087.
- (39) O'Brien, E. P.; Vendruscolo, M.; Dobson, C. M. Prediction of variable translation rate effects on cotranslational protein folding. *Nat. Commun.* **2012**, *3*, 868.
- (40) Mohammad, F.; Green, R.; Buskirk, A. R. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *eLife* **2019**, *8*, 1–25.
- (41) Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R. S.; Weissman, J. S. Genome-Wide Analysis of in Vivo Translation with Nucleotide Resolution Using Ribosome Profiling. *Science (Washington, DC, U. S.)* **2009**, *324*, 218–224.
- (42) Kamenova, I.; et al. Co-translational assembly of mammalian nuclear multisubunit complexes. *Nat. Commun.* **2019**, *10*, 1740.
- (43) Shiber, A.; et al. Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature* **2018**, *561*, 268.
- (44) Natan, E.; Wells, J. N.; Teichmann, S. A.; Marsh, J. A. Regulation, evolution and consequences of cotranslational protein complex assembly. *Curr. Opin. Struct. Biol.* **2017**, *42*, 90–97.