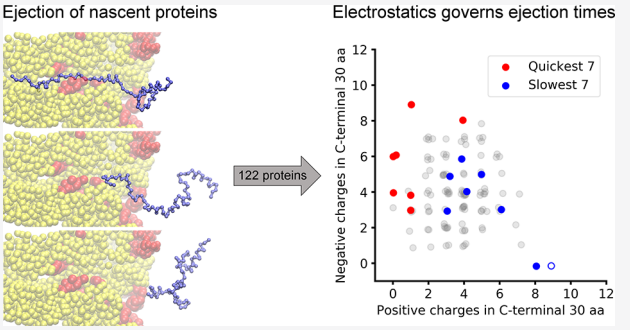
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

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

ABSTRACT: The ejection of nascent proteins out of the ribosome exit tunnel, after their covalent bond to transfer-RNA has been broken, 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.



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

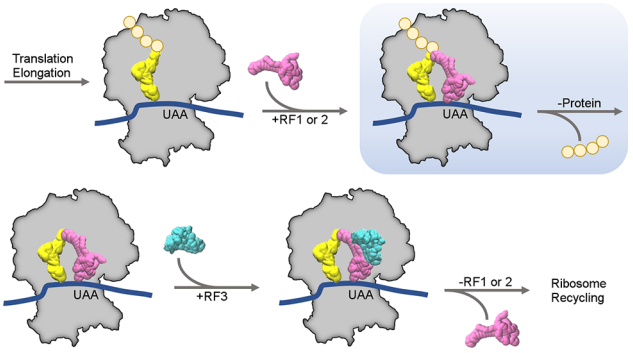


Figure 1. Translation termination in *E. coli*. Translation termination begins after translation elongation ends when a stop codon (e.g., UAA) enters the ribosome's A-site (ribosome shown as a gray outline). Release factor (RF) 1 or 2 (magenta) binds the stop codon in the A-site and catalyzes the hydrolysis 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 2H5E.

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 vitro[10] 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.

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 STRIDE[12] to be in secondary structural elements were in α-helixes 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] PREDATOR[16] and IUPRED[17] 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 $\leq$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·Å^2). All reconstructions, minimizations, and molecular dynamics simulations were performed using CHARMM with the par27 force field.[14] The minimized structures were solvated in TIP3P[18] 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

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

$$E = \sum_i k_b(r_i - r_0)^2 + \sum_i \sum_{j=1}^{4} k_{\varphi,ij}(1 + \cos[j\varphi_i - \delta_{ij}])$$
$$+ \sum_i -\frac{1}{\gamma} \ln\{\exp[-\gamma(k_\alpha(\theta_i - \theta_\alpha)^2 + \varepsilon_\alpha)] + \exp[-\gamma k_\beta(\theta_i - \theta_\beta)^2]\}$$
$$+ \sum_{ij} \frac{q_i q_j e^2}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \exp\left[-\frac{r_{ij}}{l_D}\right] + \sum_{ij\in\{NC\}} \epsilon_{ij}^{NC}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$
$$+ \sum_{ij\notin\{NC\}} \epsilon_{ij}^{NN}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

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 ϵ_{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 set[24] 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 model[26] 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 × 8.44869 × 8.18680 nm^3. 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 field[27] with all C_α positions restrained. Backbone and side-chain all-atom structures were then rebuilt using Prodart2[28] 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 $MgCl_2$ 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·nm^2) 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·nm^2). All simulations were carried out with GROMACS 2018[31] using the AMBER99SB[32] force field and the TIP3P[18] water model. The particle mesh Ewald method[33] 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 barostat[36] were employed to maintain the temperature and pressure at 310 K and 1 atm, respectively. The LINCS algorithm[37] 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 3IV5) 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

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. XXXX, XXX, XXX–XXX

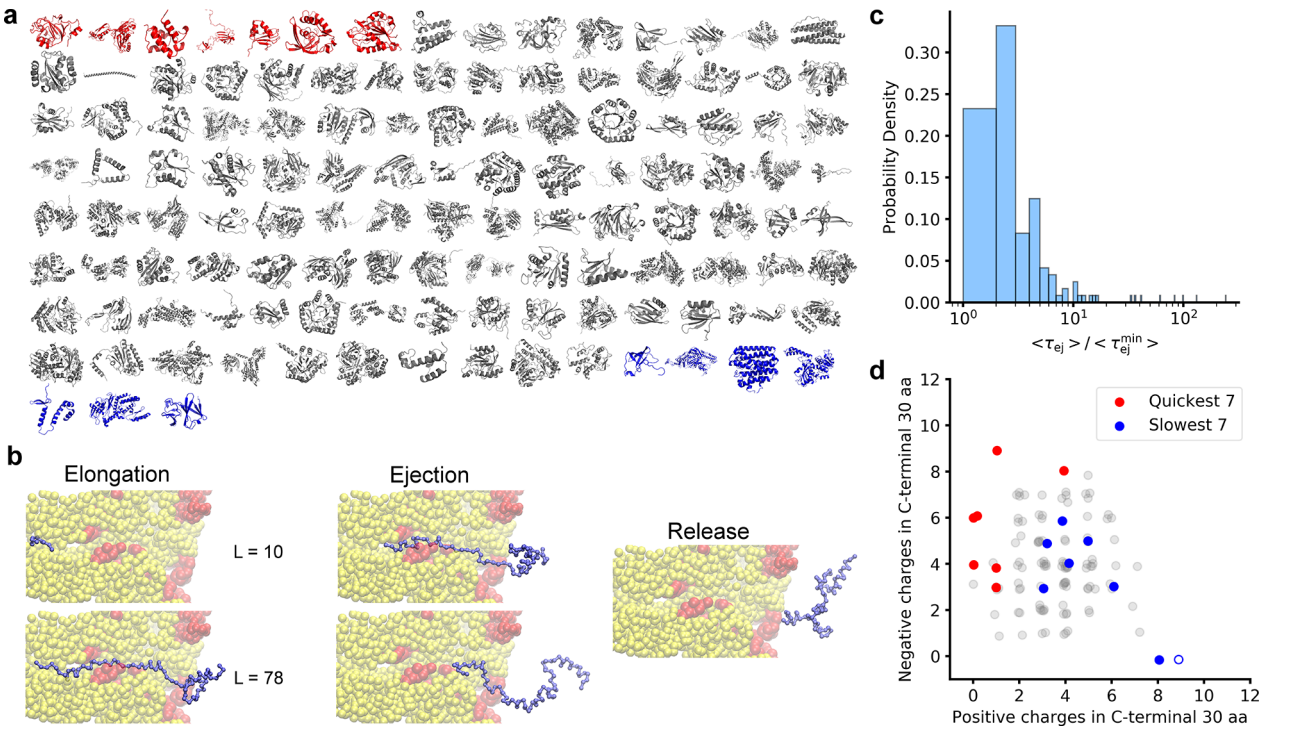


Figure 2. The 242-fold variation in ejection times is related to the presence of charged residues in the C-termini of proteins. (a) The set of 122 *E. coli* proteins that were simulated is shown from top left to bottom right by increasing ejection time. The top 5% fastest and slowest ejecting proteins 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.

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

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
3IV5	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

[a]Columns 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.

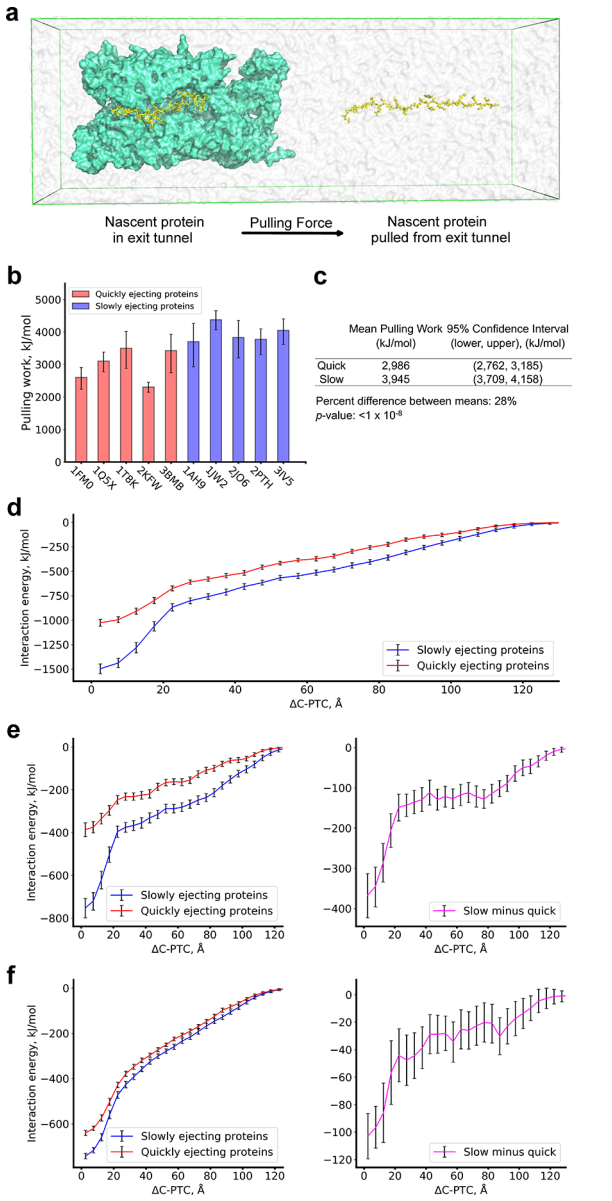


Figure 3. Slowly ejecting proteins are more electrostatically attracted to the ribosome exit tunnel. (a) Initial (left) and final (right) conformations from all-atom, steered molecular dynamics simulations of the extraction of a nascent protein (yellow) from the ribosome (cyan). (b) Mean pulling work 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 ΔC − 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.

While coarse-grained models can simulate larger systems for longer times in comparison to all-atom simulations, they leave out atomic details that have the potential to influence these results. It is currently not possible to simulate the complete ejection process of nascent chains from ribosomes using unrestrained all-atom molecular dynamics simulations. Therefore, to qualitatively test the robustness of the conclusions from our coarse-grained model, we carried out nonequilibrium all-atom steered molecular dynamics simulations in which the nascent protein is pulled from the ribosome exit tunnel using an external pulling force applied to the N-terminus of the protein (Figure 3a). If the coarse-grained model results are correct, then we predict that it will be harder to pull (as measured by the nonequilibrium work) the slowly ejecting chains out of the exit tunnel as compared to the quickly ejecting chains due to differential electrostatic interactions with the ribosome exit tunnel. Twenty-one independent trajectories were run for each of five quickly ejecting and five slowly ejecting proteins drawn from the bottom and top 10% of the distribution of ejection times, respectively. A cantilever speed of 0.25 nm/ns was used for all simulations. In these all-atom simulations, we find that the slowly ejecting nascent proteins require 28% (95% CI [19%, 36%] computed from bootstrapping; $p < 1 \times 10^{-8}$ computed from the permutation test) more work on average to be extracted from the exit tunnel than quickly ejecting nascent proteins (Figure 3b,c). Decomposing the intermolecular interactions in these simulations, we find that slowly ejecting nascent proteins have stronger interactions

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

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

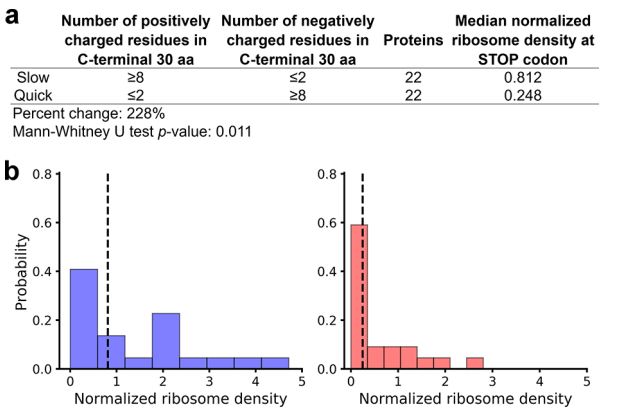


Figure 4. Presence of slowly ejecting proteins in ribosome-nascent chain complexes correlated with longer ribosome dwell times at the stop codon. (a) Proteins with ≥8 positive and ≤2 negative residues in their C-terminus as well as proteins with ≥8 negative and ≤2 positive residues in their C-terminus 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

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

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

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/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) Kuhlenkoetter, 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) Fluitt, 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.

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Supplementary Information

Electrostatic interactions govern extreme nascent protein ejection times from ribosomes and can delay ribosome recycling

Daniel A. Nissley[1, †], Quyen V. Vu[2], Fabio Trovato[1, ‡], Nabeel Ahmed[1,4], Yang Jiang[1], Mai Suan Li[2, 3], Edward P. O'Brien[1, 4, 5, *]

[1]Department of Chemistry, Penn State University, University Park, Pennsylvania, United States

[2]Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland

[3]Institute for Computational Sciences and Technology, Ho Chi Minh City, Vietnam

[4]Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, Penn State University, University Park, Pennsylvania, United States

[5]Institute for Computational and Data Sciences, Penn State University, University Park, Pennsylvania, United States

[†]**Present Address:** Department of Statistics, University of Oxford, Oxford, United Kingdom

[‡]**Present Address**: Department for Mathematics and Computer Science, Freie Universität, Berlin, Germany

*__To whom correspondence should be addressed:__ epo2@psu.edu

Supplementary Methods

Setting realistic stabilities for the single- and multi-domain proteins. Choosing η values that result in realistic stabilities for coarse-grain protein models is critical to ensuring they behave reasonably in simulations. The most rigorous method of determining η for a given protein model is to run multiple replica exchange simulations to find the value of η that produces a model for which the ΔG between the folded and unfolded states matches an experimentally determined value. However, we could not feasible perform this computationally expensive procedure for each of our 122 proteins; further, most proteins do not have an experimentally determined ΔG. We therefore selected η for our single- and multi-domain proteins based on a previously published training set of 6 α, 6 β, and 7 α/β proteins for which this rigorous procedure was performed[1]. The average values of η from this training set for each of the three structural classes are reported in Table S5 and additional related information is reported in Table S6 (these tables are based on data in Tables S3, S4, and S5 of Leininger *et al.* 2019[1]). In Tables S5 and S6, $\langle\eta\rangle_{\text{class}}$ is the mean η found for training set proteins of a particular structural class. These values of $\langle\eta\rangle_{\text{class}}$ were used as the basis for setting the intra- and inter-domain η values for our data set.

Values of η for single-domain proteins were first set to the value of $\langle\eta\rangle_{\text{class}}$ corresponding to each domain's structural class. A domain is considered to be stable if its fraction of native contacts, Q, is greater than the average Q_{kin} from the training set of $\langle Q_{\text{kin}}\rangle = 0.691$ for at least 98% of the simulation frames in each of ten 1-μs Langevin Dynamics trajectories (see original paper for a description of $\langle Q_{\text{kin}}\rangle$)[1]. All trajectories were run at 310 K with a 0.015 ps integration time step and the SHAKE algorithm[2] applied to all bonds. Coordinate information was saved every 5000 integration time steps (75 ps). Inspection of the $\Delta G_{\text{class}} = m * \langle\eta\rangle_{\text{class}} + b$ values in Table S6 indicates that $\langle\eta\rangle_{\text{class}}$ results in under-stabilized protein models in some cases. To account for this variation, simulations were also run with η increased by $\langle\Delta\Delta G/\Delta G_{\text{exp}}\rangle_{-} * 100\% = 22\%$, where $\Delta\Delta G = \Delta G_{\text{exp}} - \Delta G_{\text{class}}$ for a given protein and $\langle\ \rangle_{-}$ indicates that the average is calculated only for negative values of $\Delta\Delta G$ (*i.e.*, over values of $\Delta\Delta G$ corresponding to proteins under-stabilized by $\langle\eta\rangle_{\text{class}}$). This increase of 22% is thus the average relative increase in $\langle\eta\rangle_{\text{class}}$ necessary to stabilize a protein not stable at $\langle\eta\rangle_{\text{class}}$. Finally, a third set of simulations were run with $\langle\eta\rangle_{\text{class}}$ increased by the largest destabilization relative to the average ΔG_{exp}, calculated as $\Delta\Delta G_{\text{min}}/\langle\Delta G_{\text{exp}}\rangle$*100% = 72% where $\Delta\Delta G_{\text{min}}$ is the most-negative value of $\Delta\Delta G$ within the training set. These values are summarized in Table S5. The lowest value of η at which a particular domain was stable was selected in all cases to prevent over-stabilization, which tends to increase topological frustration and give unrealistic results. Seven of the 72 single-domain proteins were found to be unstable at each of the three η values tested. For these proteins a fourth round of simulations were run with $\eta = 2.480$, the overall largest value of η suggested by the training set (corresponding to the +72% value for β domains). None of the seven proteins were found to be stable at $\eta = 2.480$, suggesting that they are inherently structurally unstable. Therefore, to avoid over-stabilization, these domains were assigned the median η from the set of stable domains of 1.170. The final values of η selected for each single-domain protein are listed in Table S2.

Values of η for multi-domain proteins were determined by an analogous procedure. Three initial sets of ten 1-μs simulations were run with $\eta = \langle\eta\rangle_{\text{class}}$, $\langle\eta\rangle_{\text{class}} + 22\%$, or $\langle\eta\rangle_{\text{class}} + 72\%$ applied to all domains within a protein. Interface contacts were assigned $\langle\eta\rangle_{\text{overall}}$, $\langle\eta\rangle_{\text{overall}} + 22\%$, or $\langle\eta\rangle_{\text{overall}} + 72\%$ (see Table S5), where $\langle\eta\rangle_{\text{overall}}$ indicates an average over all training set η values. So, for example, the +72% simulation for PDB ID: 1CLI was run with $\eta = 1.916, 1.916,$ and 2.124 for domain 1, domain 2, and the domain 1/domain 2 interface, respectively.

A second round of ten 1-μs simulations were then run with each domain and interface assigned the minimum η at which it was found to be stable in the initial round of simulations. In many cases domains and interfaces previously stable at a particular η value became unstable in these "mixed-η" simulations. For example, PDB ID: 1D2F domain 1, domain 2, and the domain 1/domain 2 interface were initially found to be stable at $\eta = 1.427$, 1.359, and 2.124, respectively. However, when these η values were used in the mixed-η simulations the domain 1/domain 2 interface became unstable. In these situations, η values were increased by one level (*e.g.*, from $\langle\eta\rangle_{\text{class}}$ to $\langle\eta\rangle_{\text{class}} + 22\%$) and an additional set of ten 1-μs simulations run with the updated contact energies. As was done for the single-domain proteins, all domains and interfaces that were found to be unstable at each of the three η values for their respective structural class were also run with $\eta = 2.480$. If a domain or interface was unstable even at $\eta = 2.480$ it was assigned the median η from the subsets of domains or interfaces found to be stable of 1.170 and 1.507, respectively. The only exception to these selection rules is PDB ID: 2KX9 domain 2; this domain was found to be unstable at all tested values of η, but when it was assigned the median $\eta = 1.170$ value domains 1 and 3 also became unstable. Due to this strong correlation between the stability of domain 2 and the stabilities of domains 1 and 3, domain 2 was assigned $\eta = 2.480$ to ensure domains 1 and 3 remain well folded despite the fact that domain 2 is unstable at this value and the selection rules dictate it be assigned $\eta = 1.170$.

A total of 59/110 interfaces and 14/213 domains (not including domain 1 of PDB ID: 1FTS, which appears to be intrinsically disordered and contains no native contacts) were found to be unstable at any of the η values calculated from the training set. Unstable interfaces tend to be smaller, consist of fewer inter-domain contacts, and to be less hydrophobic than stable interfaces (Fig. S4). These unstable interfaces most likely represent crystal packing interfaces that are not present in the soluble native state of the protein. Similarly, unstable domains tend to be small and to contain a smaller number of intra-domain contacts than stable domains (Fig. S5). As previously noted, increasing η beyond the values suggested by the training set to stabilize these interfaces and domains would likely result in an exaggerated incidence of topological frustration as more and more energy is required to break a given contact, making contacts longer lived on average. We therefore choose to use the median values for interfaces and domains we identified to be inherently unstable in order to preserve a realistic energy scale.

Construction of 50S *E. coli* truncated ribosome. Simulating the synthesis of each of the 122 proteins in the multi- and single-domain data sets would be prohibitively time consuming if the entire 50S ribosome were to be explicitly represented. To increase computational efficiency the structure of the 50S ribosome contained in PDB ID: 3R8T was reduced to a cutout of the exit tunnel and surface near the exit tunnel opening. The full 50S structure was first coarse-grained and oriented as previously described[3] with the origin of the simulation coordinate system placed at the position of the N6 atom of A2602 and the positive x-axis pointing from this origin towards the exit tunnel opening. The positive x-axis therefore lies along the long axis of the ribosome exit tunnel. A single ribosomal interaction site corresponding to the uracil ring of U2585 was removed to prevent steric clashes with the nascent chain. Initial continuous synthesis simulations with PDB ID: 2QVR were run with the resulting CG structure, which contained the entire 50S subunit minus U2585's uracil ring, and a trajectory in which the nascent chain was inserted in an extended conformation through the exit tunnel selected. All ribosome interaction sites within 30 Å of the nascent chain or with an *x*-coordinate greater than 60 Å were kept and all other interaction sites were deleted. Residues with an *x*-coordinate greater than 60 Å but with zero solvent accessible surface area, as calculated with the COOR SURF functionality of CHARMM with RPROBE = 1.8

Å, were also removed. This probe size is significantly smaller than the smallest nascent chain interaction site, meaning that only ribosome sites with which the nascent chain cannot interact were removed. An 18-residue loop of ribosomal protein L24, which extends out over the exit tunnel opening, was allowed to fluctuate. The resulting ribosome cutout, shown in Fig. S6, consists of 3,800 interaction sites.

Calculation of interaction energy as a function of distance between the nascent protein and peptidyl transferase center. The interaction energy of each peptide with the ribosome was calculated as a function of the distance between the N6 atom of nucleotide A2602 in the ribosome peptidyl-transferase center and the C-terminal C_α atom of the 30-aa protein segment. Trajectory frames from quickly or slowly ejecting protein simulations were then placed into 5-Å bins based on this distance. Only those bins containing data for at least 10 of the 105 trajectories in a set (= 21 trajectories per protein x 5 proteins per set, quick or slow) were included in the analysis. The interaction energies between the ribosome and nascent protein in the simulation frames in a given distance bin were then averaged to produce the plots shown in Figures 3, S2, and S3 for simulations run with cantilever speeds of 0.25, 1, and 5 nm/ns, respectively. Confidence intervals were calculated using the bootstrap sampling method.

Ribosome profiling analysis. *E. coli* ribosome profiling data from Mohammed and co-workers[4] were processed as described previously[5]. This data set was chosen as it minimizes the technical biases associated with the ribosome profiling protocol in *E. coli*. Transcripts encoding proteins with ≥8 positively and ≤2 negatively charged residues in their C-terminal 30 aa were classified as putative slowly ejecting proteins. Transcripts encoding proteins with ≤2 positively and ≥8 negatively charged residues in their C-terminal 30 aa were classified as putative quickly ejecting proteins. 22 proteins occur in each of the two categories. The statistical significance of the difference between the median normalized ribosome density at the stop codon between these two populations was tested using the Mann-Whitney U test.

Gene ontology analysis. Gene ontology term enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) web server[6,7] (https://david.ncifcrf.gov/). A Benjamini-corrected p-value of 0.05 was used to determine enriched GO terms. Results are summarized in Table S9 for putative slowly ejecting proteins. No significant enrichment in any biological process was found for quickly ejecting proteins.

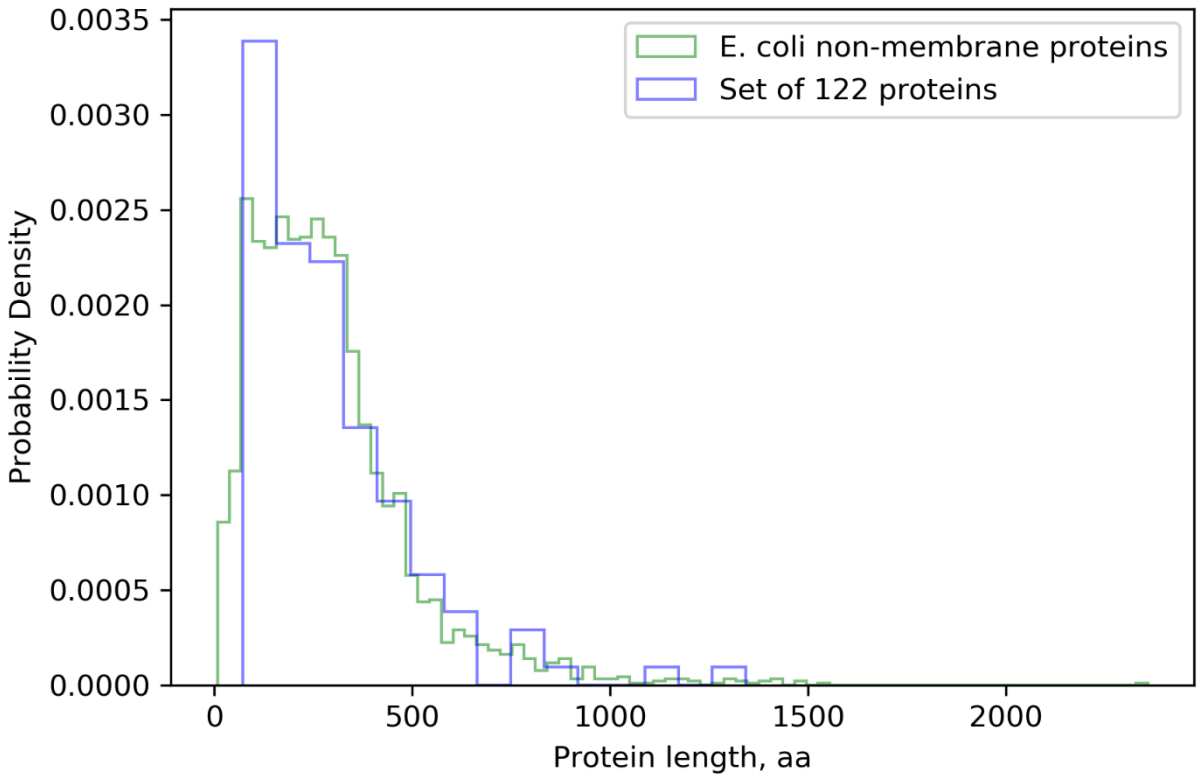
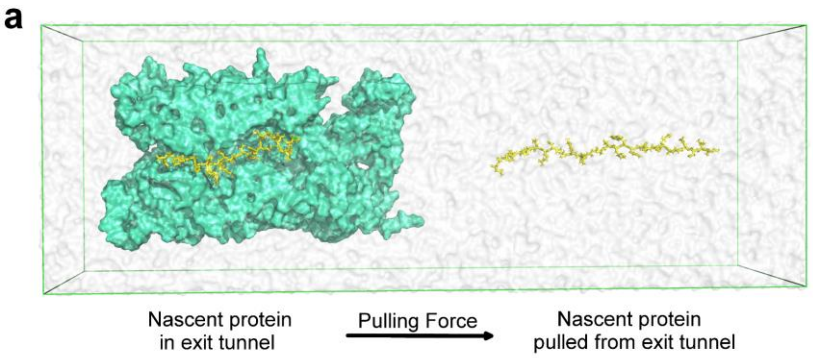
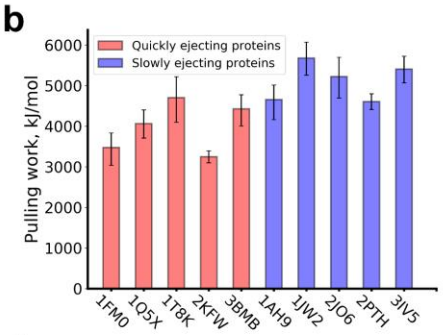


Figure S1. Distributions of protein lengths from our data set of 122 proteins and from all non-membrane proteins on Uniprot[8] for *E. coli* strain K-12 ($n = 3{,}139$).

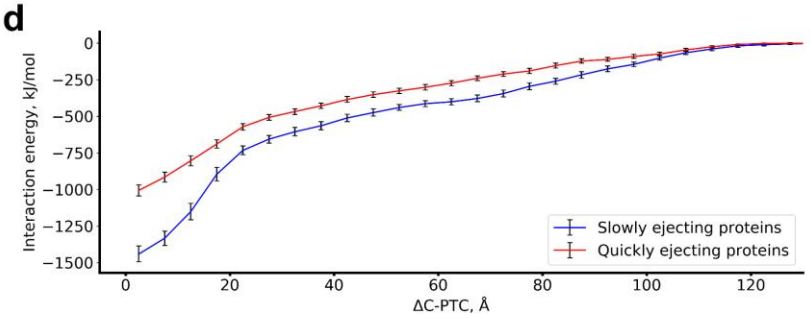


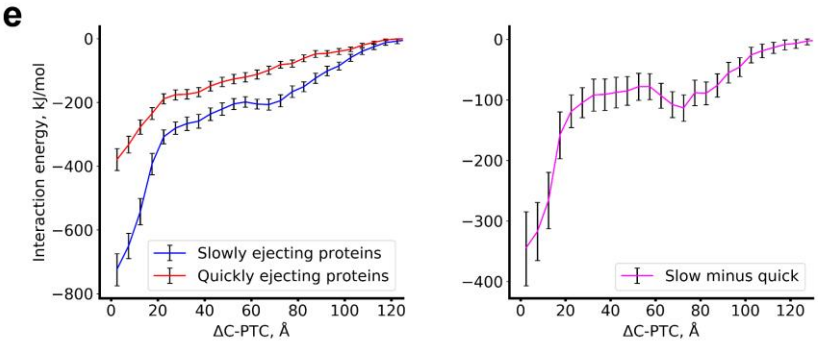


c

	Mean Pulling Work (kJ/mol)	95% Confidence Interval (lower, upper), (kJ/mol)
Quick	3,983	(3,771, 4,180)
Slow	5,114	(4,917, 5,300)

Percent difference between means: 25%
p-value: <1 x 10^{-8}





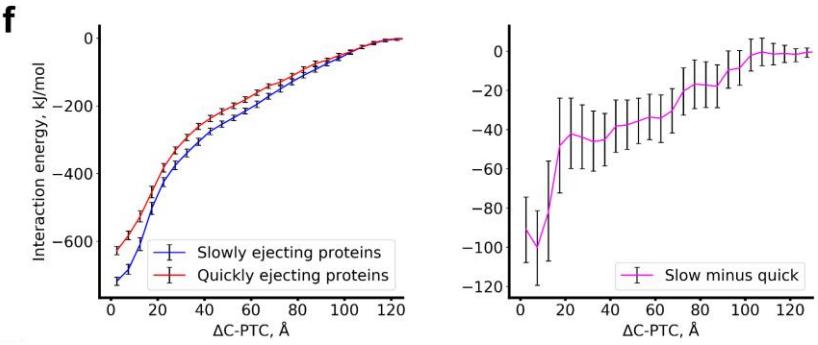
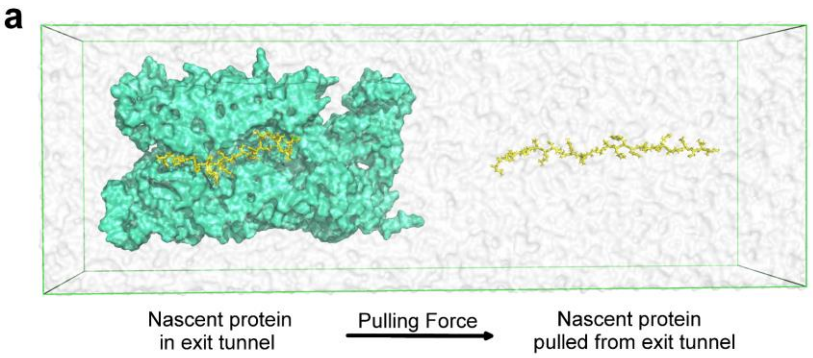
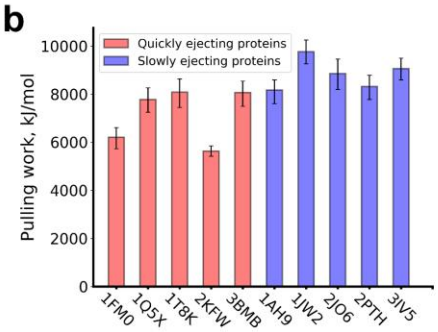


Figure S2. **Slowly ejecting proteins are more electrostatically attracted to the ribosome exit tunnel – 1 nm/ns**. (A) Initial (left) and final (right) conformations from all-atom, steered molecular dynamics simulations of the extraction of a nascent protein (yellow) from the ribosome (cyan). (B) Mean pulling work required to extract ten 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 ΔC-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) Same as (D) but for van der Waals interaction energy. These results were obtained using a cantilever speed of 1 nm/ns.

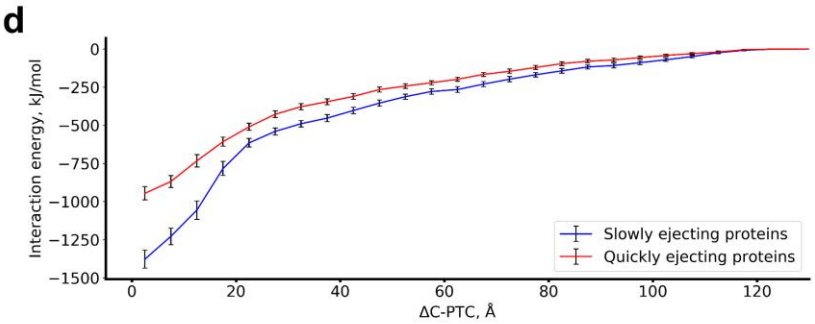


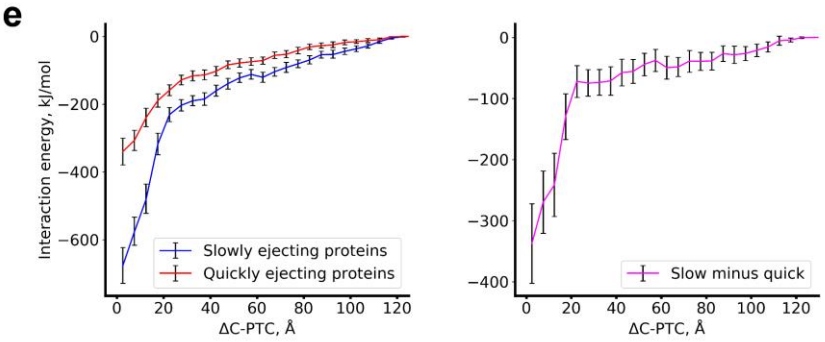


c

	Mean Pulling Work (kJ/mol)	95% Confidence Interval (lower, upper), (kJ/mol)
Quick	7,149	(6,853, 7,432)
Slow	8,828	(8,566, 9,080)

Percent difference between means: 21%
p-value: $<1 \times 10^{-8}$





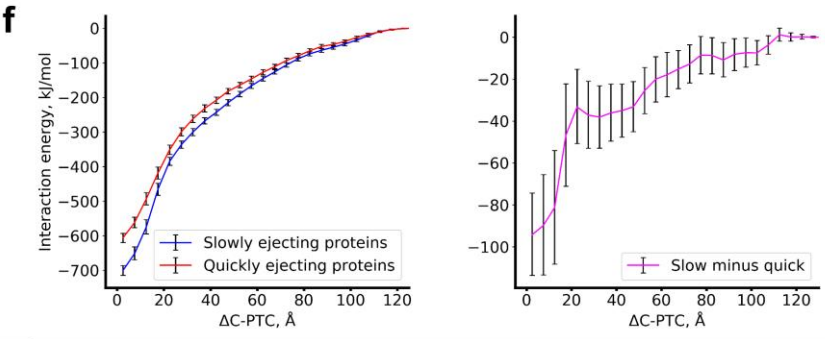


Figure S3. **Slowly ejecting proteins are more electrostatically attracted to the ribosome exit tunnel – 5 nm/ns**. (A) Initial (left) and final (right) conformations from all-atom, steered molecular dynamics simulations of the extraction of a nascent protein (yellow) from the ribosome (cyan). (B) Mean pulling work required to extract ten 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 ΔC-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) Same as (D) but for van der Waals interaction energy. These results were obtained using a cantilever speed of 5 nm/ns.

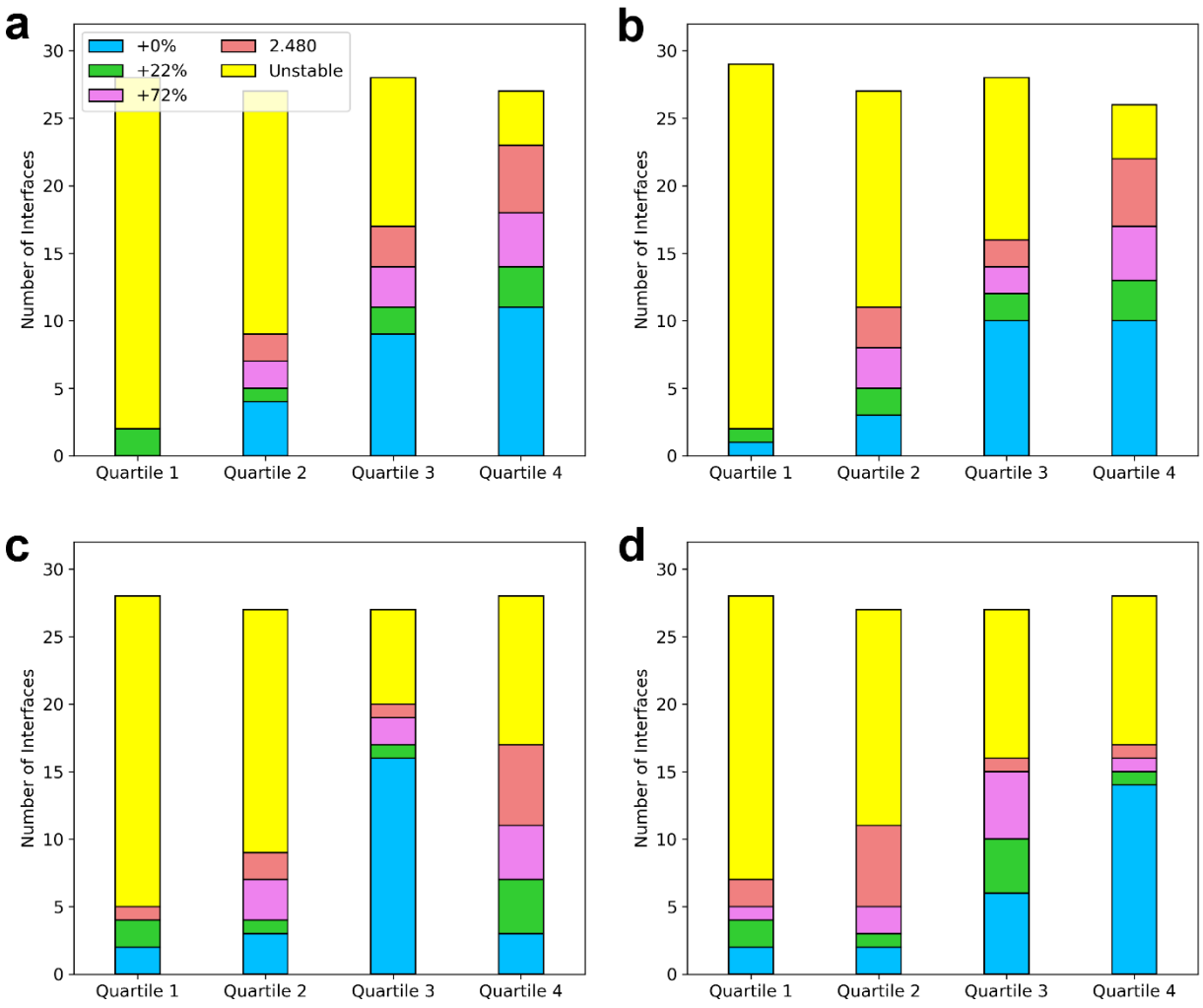


Figure S4. Relationship between interface characteristics and η used in coarse-grained models. Stacked boxplots for (a) number of residues involved in interface contacts, (b) number of inter-domain contacts per interface, (c) number of contacts per interface residue, and (d) mean hydropathic score of residues involved in interface contacts based on Kyte-Doolittle[9] scale. Boxplots were produced by rank ordering interfaces based on a given metric (*e.g.*, number of contacts at interface), splitting the ordered list into quartiles, and then color-coded based on the η values used to scale each interface's inter-domain contact energies in the quartile.

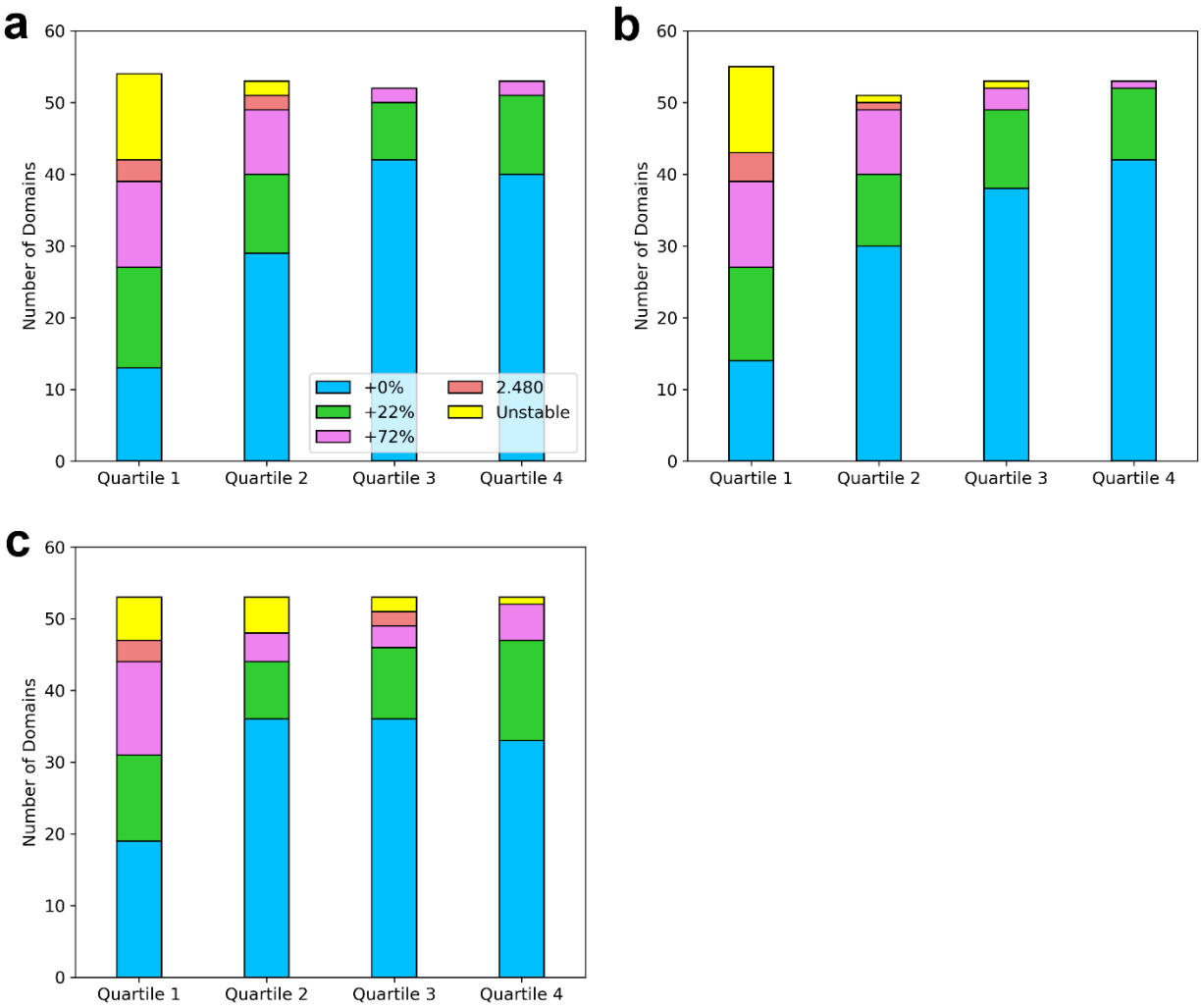


Figure S5. Relationship between domain characteristics and η used in coarse-grained models. Stacked boxplots for (a) number of residues per domain, (b) number of intra-domain contacts, and (c) number of intra-domain contacts per residue in the domain. Boxplots were produced by rank ordering domains based on a given metric (*e.g.*, number of contacts in domain), splitting the ordered list into quartiles, and then color-coded based on the η values used to scale each domain's inter-domain contact energies in the quartile. Plots include domains from both single- and multi-domain proteins except for domain 1 of PDB ID: 1FTS which is predicted to be intrinsically disordered.

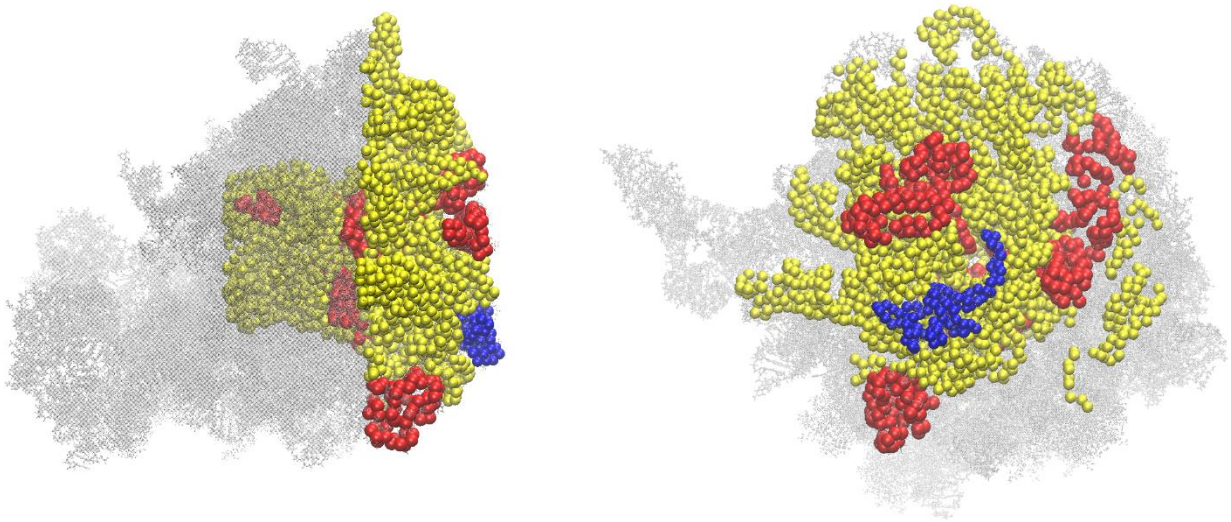


Figure S6. Side and top-down views of the ribosome exit tunnel and surface coarse-grain model used in all synthesis and ejection simulations. The full 50S ribosome subunit (PDB ID: 3R8T) is shown in gray with the coarse-grain representation superimposed. Ribosomal proteins (except L24), ribosomal RNA, and the L24 protein are shown in red, yellow, and blue, respectively.

Table S1. Structural class proportions over subset of 122 proteins and full database

	Full database		Subset of 122 proteins	
Structural Class	**Number of domains**	**Proportion of domains**	**Number of domains**	**Proportion of domains**
α	761	0.45	74	0.35
β	152	0.09	16	0.08
α/β	783	0.46	122	0.58
Total	1,696		212	

Table S2. Single-domain protein data set information

PDB ID	Chain in PDB	Number of residues	Name	Structural Class	η used in final model†
1A69	A	239	Purine nucleoside phosphorylase DeoD-type	α/β	1.359
1A6J	B	163	Nitrogen regulatory IIA protein	α/β	1.114
1A82	A	225	ATP-dependent dethiobiotin synthetase BioD	α/β	1.114
1AG9	A	176	Flavodoxin 1	α/β	1.114
1AH9	Model 1	72	Translation initiation factor 1	β	2.480
1AKE	A	214	Adenylate kinase	α	1.170
1B9L	A	120	Dihydroneopterin triphosphate 2'-epimerase	α/β	1.359
1DCJ	Model 1	81	Sulfur carrier protein TusA	α/β	1.359
1DFU	P	94	50S ribosomal protein L25	α/β	1.916
1DXE	A	256	5-keto-4-deoxy-D-glucarate aldolase	α	1.170
1EIX	C	245	Orotidine 5'-phosphate decarboxylase	α	1.170
1EM8	A	147	DNA polymerase III subunit chi	α/β	1.114
1EUM	A	165	Bacterial non-heme ferritin	α	1.170
1FJJ	A	158	UPF0098 protein YbhB	β	1.442
1FM0	D	81	Molybdopterin synthase sulfur carrier subunit	α/β	1.359
1GQT	B	309	Ribokinase	α/β	1.114
1GT7	A	274	Rhamnulose-1-phosphate aldolase	α	1.170
1H16	A	760	Formate acetyltransferase 1	α	1.170
1H75	A	81	Glutaredoxin-like protein NrdH	α/β	1.359
1I6O	B	220	Carbonic anhydrase 2	α	1.170
1JNS	Model 1	93	Peptidyl-prolyl cis-trans isomerase C	α/β	1.170*
1JW2	A	72	Hemolysin expression-modulating protein Hha	α	2.012
1JX7	A	117	Protein YchN	α/β	1.114
1K7J	A	206	Uncharacterized protein YciO	α/β	1.114
1KO5	A	175	Thermoresistant gluconokinase	α	1.170
1L6W	A	220	Fructose-6-phosphate aldolase 1	α	1.170
1M3U	A	264	3-methyl-2-oxobutanoate hydroxymethyltransferase	α	1.170
1MZG	B	138	Cysteine desulfuration protein SufE	α	1.170
1NAQ	A	112	Divalent-cation tolerance protein CutA	α/β	1.916
1ORO	A	213	Orotate phosphoribosyltransferase	α/β	1.114
1P91	A	269	23S rRNA (guanine(745)-N(1))-methyltransferase	α/β	1.359
1PF5	A	131	RutC family protein YjgH	α/β	1.916
1PMO	B	466	Glutamate decarboxylase beta	α	1.170
1PSU	B	140	Acyl-coenzyme A thioesterase PaaI	α/β	1.916
1Q5X	A	161	Regulator of RNase E activity A	α/β	1.359
1QTW	A	285	Endonuclease 4	α	1.170
1RQJ	A	299	Farnesyl diphosphate synthase	α	1.170
1SG5	Model 1	84	Protein rof	β	1.170*
1SV6	A	269	2-keto-4-pentenoate hydratase	α/β	1.359
1T8K	A	78	Acyl carrier protein	α	1.427
1U60	A	310	Glutaminase 1	α	1.170
1W8G	A	234	Pyridoxal phosphate homeostasis protein	α	1.170
1WOC	C	104	Primosomal replication protein N	β	2.480
1XN7	Model 1	78	Probable [Fe-S]-dependent transcriptional repressor FeoC	α	1.170*
1YQQ	A	277	Purine nucleoside phosphorylase 2	α/β	1.114
1ZYL	A	328	Stress response kinase A	α	1.170
1ZZM	A	259	Uncharacterized metal-dependent hydrolase YjjV	α	1.170
2A6Q	E	84	Toxin YoeB	α/β	1.359
2AXD	Model 1	76	DNA polymerase III subunit theta	α	1.170*
2D1P	B	119	Protein TusC	α	1.170
2FEK	Model 1	147	Low molecular weight protein-tyrosine-phosphatase Wzb	α	1.170
2GQR	A	237	Phosphoribosylaminoimidazole-succinocarboxamide synthase	α/β	1.359
2HD3	K	95	Ethanolamine utilization protein eutN	β	1.759
2HGK	Model 1	109	Uncharacterized protein YqcC	α	1.427
2HNA	A	147	Protein MioC	α/β	1.916
2HO9	Model 1	167	Chemotaxis protein CheW	β	1.442
2JEE	C	81	Cell division protein ZapB	α	1.170*
2JO6	Model 1	108	Nitrite reductase (NADH) small subunit	β	2.480
2JRX	Model 1	75	UPF0352 protein YejL	α	1.170*
2KC5	Model 1	162	Hydrogenase-2 operon protein HybE	α/β	1.359
2O1C	A	150	Dihydroneopterin triphosphate diphosphatase	α/β	1.114
2OQ3	Model 1	147	Mannitol-specific cryptic phosphotransferase enzyme IIA component	α/β	1.359

Table S2 continued

PDB ID	Chain in PDB	Number of residues	Name	Structural Class	η used in final model†
2PTH	A	194	Peptidyl-tRNA hydrolase	α/β	1.114
2UYJ	A	129	Putative reactive intermediate deaminase TdcF	α/β	1.359
2V81	A	205	2-dehydro-3-deoxy-6-phosphogalactonate aldolase	α	1.170
2YVA	A	196	DnaA initiator-associating protein DiaA	α	1.170
3ASV	B	248	Short-chain dehydrogenase/reductase SDR	α	1.170
3BMB	B	136	Regulator of nucleoside diphosphate kinase	α/β	1.114
3HWO	A	391	Isochorismate synthase EntC	α/β	1.359
3IV5	B	98	DNA-binding protein fis	α	1.170*
3N1S	J	119	Purine nucleoside phosphoramidase	α/β	1.114
4A2C	A	346	Galactitol 1-phosphate 5-dehydrogenase	α/β	1.359

*indicates an η value that does not result in a stable domain or interface based on the criteria described in Supplementary Methods

†values chosen based on 1-μs simulations as described in Supplementary Methods

Table S3. Information on multi-domain proteins used in this study

PDB ID	Chain in PDB	Number of residues	Name	Domains and Interfaces	Structural Class	η used in model building†
1CLI	A	345	Phosphoribosylformylglycinamidine cyclo-ligase	Domain 1: 1-170	α/β	1.359
				Domain 2: 171-345	α/β	1.114
				1|2 Interface	–	1.235
1D2F	A	390	Protein MalY	Domain 1: 1-41; 285-390	α	1.427
				Domain 2: 42-284	α/β	1.359
				1|2 Interface	–	2.480
1DUV	G	334	Ornithine carbamoyltransferase subunit I	Domain 1:1-136; 320-334	α	1.170
				Domain 2: 137-319	α	1.170
				1|2 Interface	–	1.235
1EF9	A	261	Methylmalonyl-CoA decarboxylase	Domain 1: 1-202	α/β	1.114
				Domain 2: 203-261	α	1.170*
				1|2 Interface	–	1.507*
1FTS	A	497	Signal recognition particle receptor FtsY	Domain 1: 1-200	ID	1.170
				Domain 2: 201-284	α	1.170
				Domain 3: 285-497	α/β	1.114
				1|2 Interface	–	1.507*
				1|3 Interface	–	1.507*
				2|3 Interface	–	1.507*
1FUI	A	591	L-fucose isomerase	Domain 1: 1-175	α/β	1.114
				Domain 2: 176-340	α	1.170
				Domain 3: 341-591	α/β	1.114
				1|2 Interface	–	1.235
				1|3 Interface	–	1.235
				2|3 Interface	–	2.124
1GER	B	450	Glutathione reductase	Domain 1: 1-143; 262-337	α/β	1.114
				Domain 2: 144-261	α/β	1.114
				Domain 3: 338-450	α/β	1.114
				1|2 Interface	–	1.507*
				1|3 Interface	–	2.124
				2|3 Interface	–	1.235
1GLF	O	502	Glycerol kinase	Domain 1: 1-254	α/β	1.114
				Domain 2: 255-502	α/β	1.114
				1|2 Interface	–	1.235
1GQE	A	365	Peptide chain release factor RF2	Domain 1: 1-120	α	1.170
				Domain 2: 121-225; 317-365	α/β	1.114
				Domain 3: 226-316	α/β	1.114
				1|2 Interface	–	1.507*
				1|3 Interface	–	1.507*
				2|3 Interface	–	1.507*
1GYT	L	503	Cytosol aminopeptidase	Domain 1: 1-180	α/β	1.114
				Domain 2: 181-503	α/β	1.114
				1|2 Interface	–	1.235
1GZ0	C	243	23S rRNA (guanosine-2'-O-)-methyltransferase RlmB	Domain 1: 1-78	α/β	1.114
				Domain 2: 79-243	α/β	1.114
				1|2 Interface	–	1.507*
1KSF	X	758	ATP-dependent Clp protease ATP-binding subunit ClpA	Domain 1: 1-155	α	1.427
				Domain 2: 156-350	α/β	1.916
				Domain 3: 351-438	α	1.427
				Domain 4: 439-652	α/β	1.359
				Domain 5: 653-758	α/β	1.916
				1|3 Interface	–	2.480
				2|3 Interface	–	1.507*
				3|4 Interface	–	1.507*
				4|5 Interface	–	1.507*

Table S3 continued

PDB ID	Chain in PDB	Number of residues	Name	Domains and Interfaces	Structural Class	η used in final model†
1NG9	A	853	DNA mismatch repair protein MutS	Domain 1: 1-127	α/β	1.114
				Domain 2: 128-265	α/β	1.114
				Domain 3: 266-384; 538-566	α	1.170
				Domain 4: 385-537	α	1.427
				Domain 5: 567-811	α/β	1.359
				Domain 6: 812-853	α	2.012
				1|2 Interface	–	1.507
				1|3 Interface	–	1.507*
				2|3 Interface	–	1.235
				2|5 Interface	–	1.507*
				3|4 Interface	–	1.235
				3|5 Interface	–	1.235
				5|6 Interface	–	1.507*
1P7L	A	384	S-adenosylmethionine synthase	Domain 1: 1-10; 137-233	α/β	1.114
				Domain 2: 11-105; 234-270	α/β	1.359
				Domain 3: 106-136; 271-384	α/β	1.114
				1|2 Interface	–	1.507*
				1|3 Interface	–	1.235
				2|3 Interface	–	2.480
1QF6	A	642	Threonine-tRNA ligase	Domain 1: 1-64	α/β	2.480
				Domain 2: 130-188	α/β	1.114
				Domain 3: 65-129; 189-234	α/β	1.114
				Domain 4: 235-534	α/β	1.114
				Domain 5: 535-642	α/β	1.114
				1|2 Interface	–	1.507*
				1|3 Interface	–	1.507*
				2|3 Interface	–	1.235
				3|4 Interface	–	1.507
				4|5 Interface	–	1.235
1SVT	J	548	60 kDa chaperonin (GroEL)	Domain 1: 1-135; 411-548	α	1.427
				Domain 2: 136-191; 373-410	α	1.427
				Domain 3: 192-372	α/β	1.114
				1|2 Interface	–	2.124
				2|3 Interface	–	1.507*
1T4B	A	367	Aspartate-semialdehyde dehydrogenase	Domain 1: 1-134; 352-367	α/β	1.114
				Domain 2: 135-351	α/β	1.114
				1|2 Interface	–	1.235
1U0B	B	461	Cysteine-tRNA ligase	Domain 1: 1-305	α/β	1.114
				Domain 2: 306-392	α	1.170
				Domain 3: 393-461	α	2.480
				1|2 Interface	–	1.235
				1|3 Interface	–	1.507*
				2|3 Interface	–	1.235
1UUF	A	349	Aldehyde reductase YahK	Domain 1: 1-158; 315-349	α/β	1.114
				Domain 2: 159-314	α/β	1.114
				1|2 Interface	–	1.507*
1W78	A	422	Dihydrofolate synthase/folylpolyglutamate synthase	Domain 1: 1-287	α/β	1.114
				Domain 2: 288-422	α/β	1.114
				1|2 Interface	–	2.480
1XRU	A	278	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase	Domain 1: 1-7; 30-143	β	1.442
				Domain 2: 8-29; 144-278	β	1.442
				1|2 Interface	–	1.507
1XVI	A	271	Mannosyl-3-phosphoglycerate phosphatase	Domain 1: 1-92; 187-271	α/β	1.916
				Domain 2: 93-186	α/β	1.916
				1|2 Interface	–	2.480
2FYM	A	432	Enolase	Domain 1: 1-127	α/β	1.359
				Domain 2: 128-432	α/β	1.114
				1|2 Interface	–	2.480
2H1F	A	326	Lipopolysaccharide heptosyltransferase-1	Domain 1: 1-163	α/β	1.114
				Domain 2: 164-326	α/β	1.114
				1|2 Interface	–	1.235

Table S3 continued

PDB ID	Chain in PDB	Number of residues	Name	Domains and Interfaces	Structural Class	η used in final model†
2HG2	A	479	Lactaldehyde dehydrogenase	Domain 1: 1-253; 450-479	α/β	1.114
				Domain 2: 254-449	α/β	1.114
				1|2 Interface	–	1.507*
2HNH	A	1160	DNA polymerase III subunit alpha	Domain 1: 1-280	α	1.170
				Domain 2: 281-404	α	1.170
				Domain 3: 1079-1160	α/β	1.916
				Domain 4: 558-931	α	1.170
				Domain 5: 932-1078	α/β	1.916
				Domain 6: 405-557	α/β	1.114
				1|2 Interface	–	1.507*
				1|6 Interface	–	2.480
				2|4 Interface	–	1.235
				2|6 Interface	–	1.235
				3|5 Interface	–	1.507*
				4|5 Interface	–	1.507*
				4|6 Interface	–	1.507*
2ID0	A	644	Exoribonuclease 2	Domain 1: 1-82	α/β	1.916
				Domain 2: 83-172	α/β	1.359
				Domain 3: 173-557	α/β	1.114
				Domain 4: 558-644	β	1.442
				1|2 Interface	–	2.480
				2|3 Interface	–	2.124
				2|4 Interface	–	1.507*
				3|4 Interface	–	1.507*
2KFW	Model 1	196	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD	Domain 1: 1-71; 126-196	α/β	2.480
				Domain 2: 72-125	β	1.170*
				1|2 Interface	–	1.507*
2KX9	A	575	Phosphoenolpyruvate-protein phosphotransferase	Domain 1: 1-20; 147-229	α/β	2.480
				Domain 2: 21-146	α	2.480*
				Domain 3: 230-575	α	1.170
				1|2 Interface	–	1.507*
				1|3 Interface	–	1.507*
2PTQ	A	456	Adenylosuccinate lyase	Domain 1: 1-117	α	1.170
				Domain 2: 118-381; 446-456	α	1.170
				Domain 3: 382-445	α	1.170
				1|2 Interface	–	1.235
				2|3 Interface	–	2.124
2QCU	B	501	Aerobic glycerol-3-phosphate dehydrogenase	Domain 1: 1-387	α/β	1.114
				Domain 2: 388-501	α	1.170
				1|2 Interface	–	1.507*
2QVR	A	332	Fructose-1,6-bisphosphatase	Domain 1: 1-194	α/β	1.114
				Domain 2: 195-332	α	1.170
				1|2 Interface	–	1.507
2R5N	A	663	Transketolase 1	Domain 1: 1-326	α	1.170
				Domain 2: 327-537	α	1.170
				Domain 3: 538-663	α/β	1.114
				1|2 Interface	–	1.235
				2|3 Interface	–	1.507*
2WIU	A	440	Serine/threonine-protein kinase toxin HipA	Domain 1: 1-183; 217-250	α/β	1.114
				Domain 2: 184-216; 251-440	α	1.170
				1|2 Interface	–	1.507
2WW4	A	283	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	Domain 1: 1-164	α/β	1.114
				Domain 2: 165-283	α/β	1.114
				1|2 Interface	–	2.124
3BRQ	B	336	HTH-type transcriptional regulator AscG	Domain 1: 1-59	α	1.170*
				Domain 2: 60-161	α/β	1.114
				Domain 3: 162-336	α	1.170
				1|2 Interface	–	1.507*
				2|3 Interface	–	1.235

Table S3 continued

PDB ID	Chain in PDB	Number of residues	Name	Domains and Interfaces	Structural Class	η used in final model†
3GN5	B	131	Antitoxin MqsA	Domain 1: 1-68	α/β	2.480
				Domain 2: 69-131	α	1.170
				1|2 Interface	–	1.507*
3M7M	X	292	33 kDa chaperonin	Domain 1: 1-178	α/β	1.359
				Domain 2: 179-229	α	2.012
				Domain 3: 230-292	α	1.170*
				1|2 Interface	–	2.480
				1|3 Interface	–	1.507*
3NXC	A	198	Nucleoid occlusion factor SlmA	Domain 1: 1-62	α	1.427
				Domain 2: 63-198	α	1.170
				1|2 Interface	–	2.124
3OFO	D	206	30S ribosomal protein S4	Domain 1: 1-96; 191-206	α	2.012
				Domain 2: 97-190	α/β	1.916
				1|2 Interface	–	1.507*
3PCO	D	795	Phenylalanine--tRNA ligase beta subunit	Domain 1: 1-38; 155-190	α/β	1.359
				Domain 2: 39-154	β	1.759
				Domain 3: 191-398	α/β	1.359
				Domain 4: 399-480	α/β	1.359
				Domain 5: 481-697	α/β	1.916
				Domain 6: 698-795	α/β	1.916
				1|2 Interface	–	1.507*
				1|3 Interface	–	1.507*
				1|4 Interface	–	1.507*
				2|3 Interface	–	1.507*
				3|4 Interface	–	1.507
				5|6 Interface	–	1.507*
3QOU	A	284	Chaperedoxin	Domain 1: 1-110	α/β	1.114
				Domain 2: 111-198	α	1.170
				Domain 3: 199-284	α	1.170
				2|3 Interface	–	1.507*
4DCM	A	378	Ribosomal RNA large subunit methyltransferase G	Domain 1: 1-185	α/β	1.114
				Domain 2: 186-378	α/β	1.114
				1|2 Interface	–	1.507*
4DZD	A	199	CRISPR system Cascade subunit CasE	Domain 1: 1-74	α/β	1.359
				Domain 2: 75-199	α/β	1.359
				1|2 Interface	–	1.507*
4E8B	A	243	Ribosomal RNA small subunit methyltransferase E	Domain 1: 1-71	β	1.442
				Domain 2: 72-243	α/β	1.114
				1|2 Interface	–	1.507*
4FZW	A	255	2,3-dehydroadipyl-CoA hydratase	Domain 1: 1-197	α/β	1.114
				Domain 2: 198-255	α	1.170*
				1|2 Interface	–	1.507*
4HR7	A	449	Biotin carboxylase	Domain 1: 1-85	α/β	1.114
				Domain 2: 131-203	α/β	1.114
				Domain 3: 86-130; 204-449	α/β	1.114
				1|3 Interface	–	1.235
				2|3 Interface	–	1.507*
4IM7	A	486	Hypothetical oxidoreductase ydfI	Domain 1: 1-279	α/β	1.114
				Domain 2: 280-486	α	1.170
				1|2 Interface	–	2.124
4IWX	A	300	Ribosomal protein S6-L-glutamate ligase	Domain 1: 1-114	α/β	1.114
				Domain 2: 115-181	α/β	1.916
				Domain 3: 182-300	α/β	1.114
				1|2 Interface	–	1.507*
				1|3 Interface	–	1.235
				2|3 Interface	–	1.507*

Table S3 continued

PDB ID	Chain in PDB	Number of residues	Name	Domains and Interfaces	Structural Class	η used in final model†
4KN7	C	1342	DNA-directed RNA polymerase subunit beta	Domain 1: 1-17; 796-825; 1060-1242	α/β	1.359
				Domain 2: 18-152; 445-585; 657-713	α/β	1.916
				Domain 3: 826-935; 1041-1059	β	1.759
				Domain 4: 586-656	β	1.759
				Domain 5: 714-795	β	2.480
				Domain 6: 153-444	α/β	1.359
				Domain 7: 936-1040	α	1.427
				Domain 8: 1243-1342	α	1.170*
				1|2 Interface	–	2.124
				1|3 Interface	–	1.507
				1|5 Interface	–	1.507*
				1|8 Interface	–	1.507*
				2|3 Interface	–	1.507*
				2|4 Interface	–	2.480
				2|5 Interface	–	1.507
				2|6 Interface	–	1.507*
				2|7 Interface	–	1.507*
				3|5 Interface	–	1.507*
				3|7 Interface	–	1.507*
				3|8 Interface	–	1.507*
				5|7 Interface	–	1.507*

*indicates an η value that does not result in a stable domain or interface based on the criteria described in Supplementary Methods

†values chosen based on 1-μs simulations as described in Supplementary Methods

Table S4. Multi-domain protein model building information

pdb:chain	missing res	other pdb	Modeling strategy	Notes
1svt:J	526-548		Rebuild and minimize. MD to relax the missing residues.	Missing residues are predicted to be disordered.
1ksf:X	143-167 611-623		Rebuild and minimize. MD to relax the missing residues.	Missing residues are predicted to be disordered.
1brm:A	232-242	1t4b:B	Substitution with alternative structure. Rebuild and minimize.	1t4b represents the same gene.
3brq:B	1-58	3oqo:A	Modeling of N-ter domain of 3brq. Rebuild and minimize. MD to relax the relative orientation of the N-terminal domain relative to the rest	3oqo (different gene) is used as a template to model the missing N-terminal domain of 3brq. 3oqo and 3brq are structurally aligned and 3oqo coordinates of residues 1 to 58 merged with 3brq.
1fts:A	1-200		Rebuild and minimize.	Missing residues are predicted to be disordered. Implicit solvent MD is performed due to the fragment length.
3dnt:A	<10	2wiu:A	Substitution with alternative structure. Rebuild and minimize.	2wiu represents the same gene.
1hw7:A	233-292	3m7m:X 1xjh:A	Substitution with alternative structures 3m7m. Modeling of C-terminal domain of 3m7m. Rebuild and minimize. MD to relax the relative orientation of the N and C-ter domains.	3m7m (same gene) is used because it is the closed state of hsp33, which occurs under no-stress conditions. 1xjh is used to model the missing C-ter domain. 3m7m and 1xjh are structurally aligned to 1vzy (other organism) and then merged together.
1mxa:A	<10	1p7l:A	Substitution with alternative structure. Rebuild and minimize.	1p7l represents the same gene.
1d2f:A	1-29	4dgt:A	Substitution with alternative structure. Rebuild and minimize. MD to relax the missing residues.	Coordinates missing from 1d2f are taken from 4dgt (different organism), after structural alignment of 1d2f and 4dgt.
1xvi:A	236-271		Predict model using I-TASSER[10]. MD to relax the missing residues.	Missing residues are partly structured; I-TASSER predicted model is used as a template.
1ng9:A	801-853	3zlj:C	Modeling of C-ter domain of 1ng9. Rebuild and minimize. MD simulation to find the relative orientation of the N- and C-terminal domains.	3zlj (same gene) corresponds to residues 823 to 853. Linker residues 801 to 822 are missing in 1ng9 and 3zlj, and are predicted to be unstructured. The linker is added to 1ng9; 3zlj is re-oriented in vmd to be attached to linker.
1zym:A	250-575	2kx9:A	Substitution with alternative structure. Rebuild and minimize	2kx9 represents the same gene.
2hnh:A	911-1160	5fku:A	Modeling of missing residues. Rebuild and minimize. MD to relax the missing residues.	Coordinates missing from 2hnh are taken from 5fku (same gene) after structural alignment of 5fku and 2hnh. Residues 927-937 are missing in both 2hnh and 5fku. They are predicted as unstructured.
Multi-domain proteins subject to rebuilding and minimization only (pdb:chain)				
2kfw:A 3ofo:D 4e8b:A 4kn7:C 3nxc:A 3gn5:B 1u0b:B 3pco:D 2fym:A 1qf6:A 2r5n:A 1uuf:A 3qou:A 4im7:A 4hr7:A 2hg2:A 1duv:G 1gyt:L 4dzd:A 1fui:A 2qvr:A 1w78:A 2ww4:A 1xru:A 4fzw:A 1cli:A 2ptq:A 1gqe:A 2h1f:A 4iwx:A 1gz0:C 4dcm:A 2id0:A 1ef9:A 2qcu:B 1glf:O 1ger:B				

Table S5. η values used for each structural class and the training set overall

Structural Class	$\langle\eta\rangle_{\text{class}}$	$\langle\eta\rangle_{\text{class}}$+22%	$\langle\eta\rangle_{\text{class}}$+72%
α	1.170	1.427	2.012
β	1.442	1.759	2.480
α/β	1.114	1.359	1.916
Overall	1.235	1.507	2.124

Table S6. Additional training set parameters

α	η^*	m	b	ΔG_{exp}	$\Delta G_{\mathrm{class}} = m * \langle \eta \rangle_{\mathrm{class}} + b$	$\Delta\Delta G = \Delta G_{\mathrm{exp}} - \Delta G_{\mathrm{class}}$
EC298	1.080	-23.456	22.609	-2.72	-4.82	2.10
λ-repressor	1.163	-20.230	19.278	-4.25	-4.38	0.13
bACBP	1.304	-21.941	22.216	-6.40	-3.44	-2.96
IM7	0.998	-16.745	13.832	-2.88	-5.75	2.87
IM9	1.298	-18.480	17.167	-6.81	-4.45	-2.36
cytochrome-256b	1.174	-28.079	26.378	-6.60	-6.46	-0.14

$\langle \eta \rangle_{\mathrm{class}}$, α	1.170

β	η^*	m	b	ΔG_{exp}	$\Delta G_{\mathrm{class}} = m * \langle \eta \rangle_{\mathrm{class}} + b$	$\Delta\Delta G = \Delta G_{\mathrm{exp}} - \Delta G_{\mathrm{class}}$
ABP1 SH3	1.628	-4.590	4.403	-3.07	-2.22	-0.85
Fyn SH3	1.467	-11.149	9.680	-6.68	-6.40	-0.28
CspB Bc	1.298	-13.749	16.035	-1.81	-3.79	1.98
CspA	1.445	-15.226	18.998	-3.00	-2.96	-0.04
Tenascin	1.457	-28.220	34.408	-6.71	-6.29	-0.42
Twitchin	1.357	-20.195	22.402	-5.00	-6.72	1.72

$\langle \eta \rangle_{\mathrm{class}}$, β	1.442

α/β	η^*	m	b	ΔG_{exp}	$\Delta G_{\mathrm{class}} = m * \langle \eta \rangle_{\mathrm{class}} + b$	$\Delta\Delta G = \Delta G_{\mathrm{exp}} - \Delta G_{\mathrm{class}}$
Hpr	1.164	-21.377	19.419	-5.46	-4.40	-1.06
Urm1	1.075	-22.836	21.078	-3.48	-4.36	0.88
Src SH2	1.298	-20.051	18.778	-7.24	-3.56	-3.68
Azurin (apo)	1.163	-25.190	23.997	-5.29	-4.07	-1.22
CheY	1.001	-58.105	52.554	-5.60	-12.18	6.58
Ribonuclease H	1.047	-38.735	33.369	-7.20	-9.79	2.59
Dihydrofolate reductase	1.051	-29.666	24.797	-6.37	-8.26	1.89

$\langle \eta \rangle_{\mathrm{class}}$, α/β	1.114

Note: All free energies, *m*, and *b* have units of kcal/mol. See Leininger *et al*. 2019 for training set analysis details.

Table S7. Mean *in silico* dwell times calculated from Fluitt-Viljoen model

Index	Codon	Mean decoding time from Fluitt and Viljoen, ms	Mean *in silico* decoding time, 0.015 ps time steps
1	UUU	136	846908
2	UUC	195	1214317
3	UUG	50	311363
4	UUA	157	977681
5	UCU	55	342500
6	UCC	246	1531908
7	UCG	96	597818
8	UCA	106	660090
9	UGU	75	467045
10	UGC	109	678772
11	UGG	168	1046181
12	UGA	12	74727
13	UAU	53	330045
14	UAC	77	479500
15	UAG	19	118318
16	UAA	11	68500
17	CUU	260	1619090
18	CUC	204	1270363
19	CUG	35	217954
20	CUA	286	1780998
21	CCU	143	890499
22	CCC	197	1226772
23	CCG	134	834454
24	CCA	237	1475862
25	CGU	28	174363
26	CGC	35	217954
27	CGG	397	2472225
28	CGA	34	211727
29	CAU	296	1843271
30	CAC	222	1382453
31	CAG	231	1438499
32	CAA	179	1114681
33	GUU	26	161909
34	GUC	208	1295272
35	GUG	42	261545
36	GUA	73	454591
37	GCU	39	242863
38	GCC	415	2584316
39	GCG	44	274000
40	GCA	83	516863
41	GGU	35	217954
42	GGC	49	305136
43	GGG	81	504409
44	GGA	324	2017635
45	GAU	77	479500
46	GAC	116	722363
47	GAG	36	224182
48	GAA	57	354954
49	AUU	97	604045
50	AUC	128	797090
51	AUG	266	1656453
52	AUA	128	797090
53	ACU	55	342500
54	ACC	153	952772
55	ACG	129	803318
56	ACA	178	1108454
57	AGU	85	529318
58	AGC	127	790863
59	AGG	461	2870770
60	AGA	190	1183181
61	AAU	109	678772
62	AAC	161	1002590
63	AAG	102	635181
64	AAA	76	473272

Table S8. Mean ejection times for 122 *E. coli* proteins from a coarse-grain model

PDB ID	Mean simulated ejection time (ns)
1Q5X	0.3090
3M7M	0.3120
1T8K	0.3150
2KFW	0.3240
1FM0	0.3630
3BMB	0.3735
1AG9	0.3885
2HGK	0.4020
1FJJ	0.4095
2HO9	0.4140
1SVT	0.4170
1SG5	0.4455
1PF5	0.4485
1CLI	0.4575
1EUM	0.4815
2OQ3	0.4815
2JEE	0.4875
1A69	0.4920
1QTW	0.4920
1SV6	0.5265
1UUF	0.5265
2YVA	0.5385
3ASV	0.5505
2QVR	0.5565
1FTS	0.5595
2HG2	0.5715
1ORO	0.6150
1L6W	0.6165
1K7J	0.6345
2O1C	0.6420
1ZYL	0.6450
2D1P	0.6600
4KN7	0.6600
2ID0	0.6615
4E8B	0.6615
4IWX	0.6735
1M3U	0.6825
3NXC	0.6900
1H16	0.6915
2V81	0.6915
1NAQ	0.6930
2UYJ	0.6930
3N1S	0.7035
1XVI	0.7080
1KSF	0.7095
2JRX	0.7185
1JX7	0.7275
2GQR	0.7275
1P7L	0.7320
1GQE	0.7395
1P91	0.7575
1H75	0.7590
1ZZM	0.7695
1AKE	0.7740
2KC5	0.7755

Table S8 continued

PDB ID	Mean simulated ejection time (ns)
4A2C	0.7785
2FEK	0.7875
2HNH	0.7965
1XN7	0.8280
2H1F	0.8370
4HR7	0.8385
1GER	0.8715
2HD3	0.8715
1EF9	0.8805
1PMO	0.8805
2KX9	0.9000
2QCU	0.9000
1W78	0.9060
1B9L	0.9840
1XRU	1.0305
1EIX	1.0350
1DUV	1.0710
1DFU	1.1085
1WOC	1.1145
1A82	1.1220
3BRQ	1.1280
1A6J	1.1550
1GT7	1.1700
1MZG	1.2405
2WW4	1.2600
1GQT	1.2750
3HWO	1.2840
3PCO	1.2885
1EM8	1.3245
2A6Q	1.3425
2R5N	1.3485
1GYT	1.3500
1GZ0	1.3755
1U60	1.3875
2WIU	1.3995
4DZD	1.4085
2PTQ	1.4460
2AXD	1.4490
1JNS	1.5480
1W8G	1.6050
3QOU	1.6290
1KO5	1.6305
1YQQ	1.6770
2HNA	1.8720
1GLF	2.0235
1PSU	2.0835
1DCJ	2.1420
3GN5	2.2935
4FZW	2.4750
2FYM	2.5815
1I6O	3.2775
1FUI	3.2820
1QF6	3.3300
1DXE	3.5025
4IM7	3.9180
1JW2	4.3905

Table S8 continued

PDB ID	Mean simulated ejection time (ns)
2PTH	4.7535
1D2F	5.0160
3OFO	10.2225
1AH9	11.2770
1NG9	12.6645
1RQJ	18.7995
1T4B	25.6635
3IV5	30.4695
1U0B	40.7490
2JO6	74.7270
4DCM	>2172.7170

Table S9. GO pathway analysis from DAVID webserver for proteins predicted to be slowly ejecting in *E. coli*.

Annotation Cluster 1, Enrichment score: 15.94						
Category	**Term**	**Count**	**Fold Enrichment**	**p-value**	**Benjamini corrected p-value**	**Proteins**
UP_KEYWORDS	Ribonucleoprotein	14	46.21	2.77E-20	1.30E-18	P68679, P0A7K6, P0A7P5, P0A7Q6, P0A7S3, P0A7Q1,P0AGD7,P0A7S9, P0A7T7, P61175, P60422, P0AG59, P0A7X3, P0AG48
UP_KEYWORDS	Ribosomal protein	13	43.67	2.87E-18	6.74E-17	P68679, P0A7K6, P0A7P5, P0A7Q6, P0A7S3, P0A7Q1,P0A7S9,P0A7T7, P61175, P60422, P0A7X3, P0AG59, P0AG48
GOTERM_MF_DIRECT	GO:0003735~ structural constituent of ribosome	13	33.03	8.17E-17	4.00E-15	P68679,P0A7K6,P0A7P5, P0A7Q6, P0A7S3, P0A7Q1, P0A7S9, P0A7T7, P61175, P60422, P0A7X3, P0AG59, P0AG48
GOTERM_BP_DIRECT	GO:0006412~ translation	12	27.93	1.95E-14	5.86E-13	P0A7K6, P68679, P61175,P0A7P5,P0A7Q6,P0A7S3, P0A7Q1, P0A7S9, P0AG59, P0A7X3, P0A7T7, P0AG48
KEGG_PATHWAY	eco03010: Ribosome	13	15.33	1.57E-13	1.57E-12	P68679, P0A7K6, P0A7P5, P0A7Q6,P0A7S3,P0A7Q1, P0A7S9, P0A7T7, P61175, P60422, P0A7X3, P0AG59, P0AG48
Annotation Cluster 2, Enrichment score: 4.98						
Category	**Term**	**Count**	**Fold Enrichment**	**p-value**	**Benjamini corrected p-value**	**Proteins**
UP_KEYWORDS	rRNA-binding	7	29.27	5.62E-08	6.61E-07	P0A7K6, P61175, P0A7S3, P60422, P0A7S9, P0A7T7, P0AG48
GOTERM_MF_DIRECT	GO:0019843~ rRNA binding	7	25.34	1.25E-07	2.26E-06	P0A7K6, P61175, P0A7S3, P60422, P0A7S9, P0AG59, P0AG48
UP_SEQ_FEATURE	sequence variant	3	3.95	0.162193	0.994096	P61175, P0A7S3, P0A7S9
Annotation Cluster 3, Enrichment score: 3.38						
Category	**Term**	**Count**	**Fold Enrichment**	**p-value**	**Benjamini corrected p-value**	**Proteins**
GOTERM_CC_DIRECT	GO:0022627~ cytosolic small ribosomal subunit	6	32.08	5.40E-07	7.56E-06	P68679, P0A7S3, P0A7S9, P0AG59, P0A7X3, P0A7T7
UP_KEYWORDS	tRNA-binding	3	24.54	0.005818	0.053372	P0A7S3, P0A7S9, P0A7X3
GOTERM_MF_DIRECT	GO:0000049~tRNA binding	3	12.07	0.022785	0.241625	P0A7S3, P0A7S9, P0A7X3

References

1. 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.* **116**, 5523–5532 (2019).

2. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

3. Nissley, D. A. & O'Brien, E. P. Structural Origins of FRET-Observed Nascent Chain Compaction on the Ribosome. *J. Phys. Chem. B* **122**, 9927–9937 (2018).

4. Mohammad, F., Green, R. & Buskirk, A. R. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *Elife* 1–25 (2019).

5. Ahmed, N. *et al.* Identifying A- and P-site locations on ribosome-protected mRNA fragments using Integer Programming. *Sci. Rep.* **9:6256**, 1–14 (2019).

6. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2008).

7. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).

8. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, 158–169 (2017).

9. Kyte, J. & Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **157**, 105–132 (1982).

10. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).