Structural Origins of FRET-Observed Nascent Chain Compaction on the Ribosome

Daniel A. Nissley¹ and Edward P. O'Brien^{1,2,*}

¹Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA

²Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA

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

Abstract

A fluorescence signal arising from a FRET process was used to monitor conformational changes of a domain within the *E. coli* protein HemK during its synthesis by the ribosome. A sharp increase in fluorescence was observed to begin ~10 s after translation was initiated, indicating the domain became more compact in size. Since fluorescence only reports a single value at each time point it contains very little information about the structural ensemble that gives rise to it. Here, we supplement this experimental information with coarse-grained simulations that describe protein conformations and transitions at a spatial resolution of 3.8 Å. We use these simulations to test three hypotheses for the cause of domain compaction: (1) that poor solvent quality conditions drive the unfolded state to compact, (2) that a change in the dimension of the space the domain occupies upon moving outside the exit tunnel causes compaction, or (3) that domain folding causes compaction. We find that partial folding and dimensional collapse are both consistent with the experimental data while poor-solvent collapse is inconsistent. We identify alternative dye labeling positions on HemK that upon fluorescence can differentiate between the domain folding and dimensional collapse mechanisms. Partial folding of domains has been observed in C-terminally truncated forms of proteins. Therefore, it is likely that the experimentally observed compact state is, according to our simulations, a partially folded intermediate consisting of the first three helices of the HemK N-terminal domain bundled together. With these simulations we identify the possible co-translational folding pathways of HemK.

Introduction

Förster resonance energy transfer (FRET) has recently been used to probe the conformations of nascent proteins on both translationally arrested¹ and continuously translating² ribosomes. The efficiency of energy transfer from the FRET donor dye to the FRET acceptor dye is related to the distance between them, providing a means of monitoring protein conformational states³. In one such study Rodnina and co-workers² used a FRET-based assay, which measures acceptor dye fluorescence but not FRET efficiency, to monitor the synthesis of the N-terminal domain (NTD) of the *E. coli* N5-glutamine methyltransferase protein HemK. The HemK NTD is 73 residues in length and contains five helices that form a bundle in the native state (Fig. 1A). In these experiments the HemK NTD was fused to a 39-residue C-terminal segment composed of the inter-domain linker and a small portion of the HemK C-terminal domain (CTD). A FRET donor/acceptor pair was incorporated into the HemK NTD *via* chemically modified amino acids at positions 1 and 34 (Figs. 1B and C). Time series of acceptor fluorescence due to FRET during continuous synthesis were obtained for six HemK constructs by monitoring the fluorescence in the FRET acceptor channel in a stopped-flow apparatus. The acceptor fluorescence began to increase after 10 s of synthesis, indicating that the FRET dyes moved, on average, closer together due to some type of co-translational HemK NTD compaction. These data, in combination with limited proteolysis assays and photoinduced electron transfer experiments, were used to support the hypothesis that HemK NTD co-translationally folds through a compact state. Subsequent fitting to a chemical kinetic model suggested that partially folded structures can transiently form during synthesis, but the formation of a molten globule or a non-specific collapsed state was not ruled out⁴.

The observed compact state could arise due to partial domain folding⁵ or collapse of the unfolded state in either a good or poor solvent. The ribosome exit tunnel is about 100 Å in length and has an average diameter of 15 Å, which is similar to the 10 Å persistence length of unstructured proteins⁶⁻⁸. For these reasons, nascent protein segments are confined in an effectively 1-dimensional tube inside the tunnel and translocate to a 3-dimensional space outside of it. Polymers tend to collapse when there is an increase in the space dimension in which they exist⁹⁻¹¹. The radius of gyration, R_g , of the unfolded state of a N -monomer chain in a good solvent scales with the space dimension, d , as $R_g \propto N^{\frac{3}{d+2}}$. Thus, as a domain moves from inside to outside the exit tunnel R_g 's scaling switches from N ($d = 1$) to $N^{\frac{3}{5}}$ ($d = 3$), such that the unfolded state of the 73-residue HemK NTD would compact in size by $N^{-\frac{2}{5}}$ or 82%. Indeed, such unstructured collapse has been seen in simulations of other ribosome-nascent chain complexes¹². For a domain in a poor solvent, $R_g \propto N^{\frac{1}{d}}$, and so R_g will go from N ($d = 1$) to $N^{\frac{1}{3}}$ ($d = 3$) when the domain enters the cytosol, such that HemK NTD will collapse by $N^{-\frac{2}{3}}$ or 94%. Such collapse would bring the fluorescent dyes in HemK NTD closer together and therefore has the potential to explain the increased FRET fluorescence.

Here, we present results from coarse-grained, low-friction Langevin dynamics simulations of HemK synthesis that were performed with three different Hamiltonians to assess which driving force yields a compaction process that is most consistent with the experimental data. The first Hamiltonian, which models domain folding, is a Gō-based model previously used to model co-translational folding *in silico*¹³⁻¹⁵. The second Hamiltonian approximates the influence of good solvent on the nascent protein to model dimensional collapse. The third Hamiltonian models the

influence of a poor solvent to permit non-specific collapse. With this approach we are able to provide a molecular interpretation of FRET measurements on ribosome-nascent chain complexes, suggest optimal dye positions for future experiments, and determine the co-translational folding pathway of HemK NTD.

Results

E and $\frac{E}{E_{\text{end}}}$ are equally valid for comparison to experimental fluorescence. The experimentally observed time-dependent fluorescence of the acceptor in the presence of the donor ($F_{\text{AD(A)}}$) divided by the final fluorescence of the acceptor measured for HemK112 ($\frac{F_{\text{AD(A)}}}{F_{\text{AD(A)}}^{\text{end}}}$) cannot be calculated from our simulations. However, we can compare our results to experiment through the FRET efficiency, E , which is the probability of energy transfer from the donor to the acceptor dye, can be calculated from our simulations using Förster's equation, $E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}$, where r is the distance between the dyes and R_0 is the Förster Radius. We can also calculate the related quantity $\frac{E}{E_{\text{end}}}$, which is the time-dependent value of E divided by the final value of E for HemK112 (E^{end}). Both E and E^{end} are directly proportional to $\frac{F_{\text{AD(A)}}}{F_{\text{AD(A)}}^{\text{end}}}$ (see Eq. S2 and S4 and Supplementary Discussion). Under the experimental conditions of constant illumination at the FRET donor excitation wavelength, as $\frac{F_{\text{AD(A)}}}{F_{\text{AD(A)}}^{\text{end}}}$ increases E and $\frac{E}{E_{\text{end}}}$ will also increase, and as $\frac{F_{\text{AD(A)}}}{F_{\text{AD(A)}}^{\text{end}}}$ decreases E and $\frac{E}{E_{\text{end}}}$ will decrease. We emphasize that both E and $\frac{E}{E_{\text{end}}}$ are equally valid to compare to $\frac{F_{\text{AD(A)}}}{F_{\text{AD(A)}}^{\text{end}}}$, and it is not possible to argue that one is better to use than the other. Therefore, in the rest of this study we compare both our simulated E and $\frac{E}{E_{\text{end}}}$ time series to the experimental $\frac{F_{\text{AD(A)}}}{F_{\text{AD(A)}}^{\text{end}}}$ time series and discuss the conclusions from each set of results.

Both partial folding and dimensional collapse explain the experimental data. To test whether partial folding, dimensional collapse, or poor-solvent collapse is most consistent with experiment we simulated the synthesis of the six HemK experimental constructs, which differ in their final length, with three different Hamiltonians (see Methods). A coarse-grain representation of the *E. coli* 50S ribosomal subunit¹⁶ and a dye-modified HemK coarse-grain structure² (Fig. 1D) were used in all synthesis simulations. To model domain folding we use a Gō-based Hamiltonian utilized in previous studies of co-translational folding^{13–15}. To model dimensional collapse we use another Hamiltonian that approximates the influence of good-solvent conditions on the HemK NTD. Finally, we use a third Hamiltonian that approximates the influence of poor-solvent conditions with the introduction of a non-specific attractive interaction determined from simulations in bulk to cause collapse. Note that nascent chains modeled using any of these Hamiltonians will undergo dimensional collapse to some extent; the good-solvent model, however, will *only* collapse due to a change in dimension, while the other two models have additional driving forces for compaction.

The poor-solvent collapse E and $\frac{E}{E_{\text{end}}}$ results are inconsistent with the experimental fluorescence data for each of the six HemK constructs (Table 1 and Figs. 2A and B orange time series, Pearson R^2 range: 0.51 – 0.82). Strong correlations, however, are found between the dimensional collapse E and $\frac{E}{E_{\text{end}}}$ time series (Table 1 and Figs. 2A and B, magenta time series, Pearson R^2 range: 0.76 – 0.99) and the experimental data, as they follow similar trends. The E and $\frac{E}{E_{\text{end}}}$ time series resulting from continuous synthesis with the folding Hamiltonian (Table 1 and Figs. 2A and B, blue time series, Pearson R^2 range: 0.76 – 0.98) are also strongly correlated with the experimental $\frac{F_{\text{AD(A)}}}{F_{\text{AD(A)}}^{\text{end}}}$ time series. On the basis of these comparisons, we conclude that co-translational folding or good-solvent collapse hypotheses are in roughly equivalent agreement with experiment, while poor-solvent collapse is in poor agreement.

To quantify the similarity of the E and $\frac{E}{E_{\text{end}}}$ time series arising from different Hamiltonians we calculated the root-mean-square deviation between pairs of simulation time series, which we denote λ (Table 2 and Methods). Comparison of λ to the time series in Fig. 2 suggests that when λ is ≥ 0.1 the folding and good-solvent time series can be differentiated from one another. We find that for the majority of HemK constructs λ is < 0.1 between the folding and good-solvent Hamiltonians, indicating that in many cases they are highly similar. However, λ is > 0.1 when the poor-solvent Hamiltonian is compared to either the folding or good-solvent Hamiltonian results. These results quantitatively demonstrate that in general the folding and good-solvent Hamiltonians give the same results, meaning that both mechanisms of compaction can provide a reasonable explanation for the experimental data.

Alternative dye positions allow the partial folding and dimensional collapse mechanisms to be tested. With the experimental dye labeling positions of residues 1 and 34 the folding and dimensional collapse mechanisms give similar agreement with experiment. We hypothesized that there exist alternative dye positions within the HemK NTD that would permit these two mechanisms to be tested. We therefore reanalyzed the folding and good-solvent Hamiltonian simulations to determine dye positions that maximize the difference in E and $\frac{E}{E_{\text{end}}}$ between these two mechanisms. We find that alternative dye positions can improve the separation between the folding and good-solvent E time series by 80 to 196%, with all $\lambda > 0.1$, such that folding and dimensional collapse can be differentiated from one another (Fig. 3 and Table 3). The separation between $\frac{E}{E_{\text{end}}}$ time series also improved in five of the six HemK constructs, and λ became > 0.1 for three of the six constructs. We therefore recommend that, in future experiments, to maximize the observable difference between folding and dimensional collapse for HemK112 the donor and acceptor be placed at residues 3 and 66, respectively.

Co-translational folding of HemK NTD proceeds through a partially folded intermediate. Multiple proteins have been observed to fold co-translationally through partially folded intermediates¹⁷, and we found on the basis of our simulations that partial folding is a likely mechanism for HemK NTD compaction. We therefore sought insight into the possible co-translational folding pathways of HemK NTD by analyzing how the fraction of native contacts (Q) changed with time in our folding-Hamiltonian simulations. For HemK112 we observe that helices

1, 2, and 3 form their intra-helix contacts while still buried in the exit tunnel after 7, 12, and 15 s of synthesis, respectively (Fig. 4A; h1, h2, and h3 time series. Fig. 4D; L = 42 aa and L = 56 aa structures). Tertiary contacts begin to form at 15 s and, after 23 s of synthesis, helices 1, 2, and 3 form a stable intermediate at a nascent chain length of 84 aa (Figs. 4B and C; h1h2, h1h3, h2h3, and h123 time series. Fig. 4D; L = 70 aa and L = 84 aa structures). Helix 4 forms secondary structure after 24 s of synthesis and joins to the partially folded intermediate after 27 s of synthesis (Figs. 4A, B, and C; h4, h1h4, h2h4, h1234 time series. Fig. 4D; L = 98 aa structure). Helix 5 forms intra-helix contacts after 31 s, when the full-length protein is complete, and begins to participate in tertiary structure at roughly the same time, forming the native state (Figs. 4A, B, and C; h5, h2h5, h4h5 time series. Fig. 4D; L = 112 aa structure).

The *Q* time series for the shorter HemK constructs are very similar to those for HemK112, but with some exceptions. HemK98 is not long enough for helix 5 to dock with helices 1, 2, 3, and 4, and helix 5 therefore forms secondary but not tertiary contacts (Fig. S1). HemK84 forms intra-helix contacts in helix 4 transiently from 15-23 s, but as tertiary contacts form between helices 1, 2, and 3, it unwinds (Fig. S2). HemK70 displays unstable tertiary contact formation between helices 1 and 2, while HemK56 and HemK42 display no tertiary structure formation (Figs. S3, S4, and S5). In summary, we find that the co-translational folding of HemK NTD proceeds *via* the formation of a three-helix bundle consisting of helices 1, 2, and 3 after 23 s of synthesis followed by a four-helix bundle at 27 s and the native-state five-helix bundle at 31 s.

Discussion

The output of experimental fluorescence assays is a single value per time point, and converting these single values into a meaningful interpretation of the nature of a protein's underlying conformational ensemble is a difficult inverse problem. This issue becomes acute when considering ensembles of asynchronously elongating ribosome-nascent chain complexes containing a range of nascent chain lengths. Our coarse-grain molecular dynamics study of the HemK NTD reveals the molecular details of its co-translational folding at Angstrom spatial resolution to help alleviate this fundamental issue of interpreting fluorescence data. Co-translational folding is an important process by which nascent proteins obtain structure at the earliest possible stage in their biosynthesis. Estimates suggest that at least one third of the *E. coli* proteome folds co-translationally¹⁸, and perturbations to co-translational folding are associated with protein misfolding¹⁹ and disease²⁰. Understanding the molecular details of co-translational folding is therefore important to understanding protein biogenesis at large.

We tested three hypotheses concerning the nature of the compact state observed experimentally during HemK NTD synthesis: (1) that it collapses due to the change in dimension it experiences upon leaving the exit tunnel as it would in a good solvent, (2) that it collapses to an unstructured globule upon leaving the exit tunnel as in a poor solvent, and (3) that it attains a partially folded structure. We simulated HemK NTD synthesis with three different Hamiltonians to test these hypotheses, and found that while non-specific collapse is inconsistent with the experimental data both partial domain folding and dimensional collapse are consistent. Partial folding of HemK NTD by sequential addition of helices to a growing globular core has previously been suggested⁴, and our results are consistent with this mechanism but cannot rule out dimensional collapse. Reanalysis of our good-solvent and folding-Hamiltonian results allowed us to select optimal dye positions that maximize the observable difference between the folding and

dimensional collapse mechanisms. We were able to improve the root-mean-square separation between our folding and good-solvent model results by up to 245% by permuting the dye labeling positions in this way. Use of these dye positions will allow for the good-solvent and folding hypotheses to be tested in future experiments.

The extensive molecular information contained in our simulations of HemK NTD with the folding Hamiltonian also allowed us to investigate its folding pathways by analyzing the time series of native contacts. The native state, observed only for HemK112, is preceded at shorter nascent chain lengths by structures composed of subsets of helices. HemK42 and HemK56 display no tertiary contact formation between helices, while HemK70 begins to form contacts between helices 1 and 2. This is in contrast with a previous study⁴ which suggested tertiary contact formation between helices 1 and 2 within HemK56. HemK84 displays contact formation between helices 1, 2, and 3, and coarse-grain structures indicate the formation of a three-helix bundle intermediate, but helices 4 and 5 remain unable to participate due to confinement in the ribosome exit tunnel. HemK98 is almost fully native, but helix 5 is unable to participate in stable tertiary structure with the other helices. Our results are thus largely consistent with the sequence of partially folded intermediates suggested by chemical kinetic modeling of experimental fluorescence data⁴, with the exception that we see no contact formation between helices 1 and 2 for HemK56.

Calculating fluorescence intensities from molecular dynamics simulation data is problematic. While Förster's Equation provides an easy route from atomic distances to FRET efficiencies, no such equation exists to convert atomic coordinates to a fluorescence intensity. However, E is intrinsically related to the fluorescence of the acceptor molecule during FRET, and comparisons can therefore still be made (see SI discussion). The quantity $\frac{E}{E^{\text{end}}}$, the calculation of which mimics the data processing step of division by the final HemK112 time point value as was done for the experimental fluorescence data, is also intrinsically related to the experimental fluorescence intensity. Both E and $\frac{E}{E^{\text{end}}}$ lead to the same general conclusion: a poor-solvent model is inconsistent with experiment while folding and good-solvent models are in roughly equivalent agreement. The robustness of this result to normalization by E^{end} supports our assertion that both values are equally valid. It is also important to note that the type of coarse-grain model represented by our folding Hamiltonian is only capable of forming native contacts, and thus cannot form alternative tertiary structure. We have therefore not ruled out the possibility that compact structure(s) detected during HemK NTD synthesis *in vitro* are the result of non-native tertiary structure.

Coarse-grain molecular dynamics is a powerful means of investigating molecular mechanisms of co-translational folding. Previous molecular dynamics simulations at both atomic and coarse-grained resolution have been used to gain insight into FRET results with biopolymers²¹⁻²³. We have found that coarse-grain Langevin Dynamics simulations can be used to optimize dye position choices to test specific hypotheses concerning the nature of protein conformations. Without the acceleration of protein dynamics inherent to coarse-grain models and low-friction Langevin dynamics these simulations would be exceedingly time consuming, but we were able to achieve the equivalent of ~35 s of experimental time for 6 HemK constructs in roughly 30 days of computer processor time. Acceleration of these simulations through the use of a further reduced ribosome structure will allow the co-translational folding behavior of even very large

proteins to be studied. This type of continuous synthesis protocol could then be used to investigate co-translational folding of entire proteomes given sufficient computational resources.

Molecular dynamics simulations can help decrease the ambiguity of single-valued measurements such as fluorescence and reveal the molecular details of co-translational folding, and simulations with different Hamiltonians can be used to test different molecular hypotheses for the origin of an experimental signal. If used in a preliminary scan of potential FRET dye labeling positions molecular dynamics can also reveal ideal dye positions that will provide the best signal for testing a particular molecular hypothesis. These results highlight molecular dynamics as a powerful tool for investigating co-translational folding in molecular detail.

Methods

Construction of HemK N-terminal domain folding model. A structure-based Gō model of the HemK N-terminal domain (residues 2-73) was constructed from PDB ID: 1T43 in a manner previously described²⁴. Each amino acid was represented as a single spherical interaction site with its center of mass at the location of the C_α atom in the crystal structure. Transferable bond, angle, and dihedral terms were used. Native contacts were defined based on the PDB ID: 1T43 crystal structure. The ETEN potential was used for all simulations. The stability of this HemK NTD Gō model was tuned with replica exchange simulations in the molecular dynamics package CHARMM²⁵ to reproduce the experimental value of the folded state stability at 298 K within error (Experiment: -4.66 ± 0.10 kcal/mol, Simulation: -4.58 ± 0.11 kcal/mol) when Lennard-Jones well depths were globally increased by a factor of 2.2. A total of 50,000 exchanges were attempted between replicates, the first 5,000 discarded, and the weighted histogram analysis method²⁶ used for analysis. The root-mean-square deviation (RMSD) from the native state was calculated for the five helices of 1T43 as identified by STRIDE. The HemK NTD Gō model was defined as folded when the RMSD was <7 Å in comparison to the native state.

Construction of dye-modified HemK model. Explicit coarse-grained representations of the FRET dyes were constructed². Structures of BodipyFL (BOF, Life Technologies D6140) and Bodipy 576/589 (BOP, Life Technologies D2225) and their linkers were built in Gaussian/g09d01 and their geometry optimized with B3LYP/6-31G^[27]. These structures include all atoms of the dyes and linkers with the C_α position represented by a $-CH_3$ group. The coordinates of the optimized structures were used to generate coarse-grained representations at a similar level as has been done previously for FRET dyes²¹. The approximate locations of the coarse-grain centers are displayed in Fig. 1C. Each of the rings was reduced to a single interaction site at the centroid of the ring's heavy atoms. Heavy atoms in two rings were included in the calculation of both ring's centroids. Linker interaction sites were placed at bond midpoints. The total mass of the dye, linker, and C_α beads was redistributed evenly between all of the beads constituting each dye-modified amino acid representation. BOF-Met and BOP-Lys beads were thus assigned masses of 67.666667 AMU and 48.777778 AMU, respectively, such that the total mass of each dye-modified amino acid was conserved.

A full accounting of the parameters chosen for the BOF-Met and BOP-Lys is provided in Table S1. Bond lengths were extracted by measuring the distances between the coarse-grain beads in the geometry-optimized structures. Force constants of 50.0 kcal/mol*Å² were used for

all dye and linker bonds. Bond angles were extracted from the B3LYP/6-31G-optimized structures, and all force constants were taken as 30 kcal/mol*degree². A double-well angle potential²⁸ was used for all angles in the linker. Bond angles formed by adjacent nascent chain beads, the dye-modified C_α beads, and the first bead of the linker (e.g., the L1-A1-A2 bond angle for BOF-Met) were assigned the average C_α-C_α-C_β bond angle from an ALA-ALA-ALA tripeptide constructed in CHARMM in a right-handed helical conformation. Gly-Gly dihedral terms were used for the linkers. Single-well dihedral potentials were used for the dye dihedrals, with the minimum energy position located at the value of the dihedral calculated from the B3LYP/6-31G-optimized structures. The $r_{\min}/2$ values for the linker beads were taken from Merchant et al. 2007^[21]. The $r_{\min}/2$ of the ribose interaction site was used for each of the five-membered rings, while the $r_{\min}/2$ used for six-membered rings by Merchant et al. 2007 was used for the six-membered rings here. The dye and linker representations interact in a purely repulsive manner with one another, the rest of the protein representation, and the ribosome representation. As the PDB ID: 1T43 structure does not contain an N-terminal Met residue the BOF-Met coarse-grain representation was appended to the N-terminus of the HemK NTD model. A structured linker, consisting of the next 39 residues in HemK's sequence, was appended to the C-terminus of the N-terminal domain model as was done in the original experiments. This coarse-grain structure was used for all continuous-synthesis trajectories and is referred to in following Methods sections as the HemK NTD Gō-dye model.

Construction of good- and poor-solvent collapse Hamiltonian. The folding Hamiltonian was modified to approximate the influence of good solvent by removing all intra-NTD contacts while preserving inter-NTD/linker and inter-NTD/CTD contacts. Similarly, the poor-solvent Hamiltonian was constructed by first removing all intra-NTD contacts while preserving inter-NTD/linker and inter-NTD/CTD contacts and then adding a non-specific pairwise attractive term. The functional form of the potential for these attractive interactions is taken to be the same as for contacts within the folding Hamiltonian. The well-depth for these added interactions was tuned with simulations of the HemK Gō-dye model NTD in bulk solution with well-depths of $\epsilon = \{0.0, -0.2, -0.4, \dots, -19.6, -19.8, -20.0\}$, in units of kcal/mol. Ten trajectories were run for 750 ns at each well-depth and system coordinates recorded every 1.5 ns. The first 150 ns were discarded and the mean radius of gyration calculated over the remaining 600 ns of each simulation. The resulting plot of mean radius of gyration as a function of the well-depth is shown in Fig. S6. The well-depth corresponding to the midpoint of the collapse transition, -3.4 kcal/mol, was used in our continuous-synthesis simulations. An $r_{\min}$ of 6.17 Å, which is the median $r_{\min}$ of all the intra-NTD contacts for the folding Hamiltonian, was used for these non-specific interactions.

Selection of mean *in silico* translation elongation time. The use of low-friction Langevin Dynamics with a coarse-grain model greatly accelerates dynamics. Because the relative timescales of folding and amino acid addition are critical to accurately capturing co-translational folding behavior, we calculated the *in silico* timescale of amino acid addition, $\langle \tau_A^{\text{sim}} \rangle$, as

$$\langle \tau_A^{\text{sim}} \rangle = \langle \tau_F^{\text{sim}} \rangle * \left(\frac{\langle \tau_A^{\text{exp}} \rangle}{\langle \tau_F^{\text{exp}} \rangle} \right), \quad [1]$$

where $\langle \tau_A^{\text{exp}} \rangle$ and $\langle \tau_F^{\text{exp}} \rangle$ are the experimental amino acid addition and folding times, respectively, and $\langle \tau_F^{\text{sim}} \rangle$ is the simulation folding time. This equation ensures that the experimental ratio of

timescales is maintained. The value of $\langle \tau_F^{\text{sim}} \rangle$ was determined with temperature quenching simulations. The HemK NTD Gō-dye model was first equilibrated at 800 K for 40 ns to ensure complete unfolding and then instantaneously cooled to 298 K for 200 ns. Three hundred and fifty trajectories were run with this protocol and the time-dependent survival probability of the unfolded state calculated. The HemK NTD Gō-dye model was considered to have folded during the quench at 298 K if the RMSD of the residues within its five helical elements was ≤ 5 Å. The survival probability curve was fit with a single-exponential function of the form $S_U(t) = \exp(-k_F t)$, where $S_U(t)$ is the survival probability of the unfolded state, t is the time since the start of the quenching period, and k_F is a fit parameter representing the rate of folding. The value of $\langle k_F \rangle$ was determined with curve-fitting in python to be 0.281 ns^{-1} (Pearson R^2 of fit: 0.987), corresponding to a mean folding time of $\langle \tau_F^{\text{sim}} \rangle = 3.56 \text{ ns}$. Holtkamp et al. 2015 determined that $\langle \tau_A^{\text{exp}} \rangle = 0.278 \text{ s}$ and $\langle \tau_F^{\text{exp}} \rangle = 1.95 \times 10^{-4} \text{ s}$; with $\langle \tau_F^{\text{sim}} \rangle = 3.56 \text{ ns}$, Eq. [1] gives $\langle \tau_A^{\text{sim}} \rangle = 5080 \text{ ns}$. Simulating the synthesis of the HemK constructs with an explicit representation of the *E. coli* 50S ribosome is computationally intractable for a dwell time of this duration. To reduce the required computational time, we applied an acceleration factor, α , such that

$$\langle \tau_A^{\text{sim}} \rangle' = \frac{\langle \tau_F^{\text{sim}} \rangle}{\alpha} * \left(\frac{\langle \tau_A^{\text{exp}} \rangle}{\langle \tau_F^{\text{exp}} \rangle} \right) = \frac{1}{\alpha} * \langle \tau_A^{\text{sim}} \rangle. \quad [2]$$

As α increases the simulation mean dwell time decreases linearly. A factor of α speed up is reasonable if $\langle \tau_F^{\text{sim}} \rangle \ll \langle \tau_A^{\text{sim}} \rangle'$, such that the simulation is in a quasi-equilibrium regime. We used Gillespie Algorithm simulations of two-state co-translational folding to determine the highest value of α that does not significantly alter the behavior of the system. In these simulations, the HemK NTD was only permitted to fold once the entire domain and an additional 30 residues, to simulate the effect of the exit tunnel, were synthesized. An initial set of 10,000 trajectories with the experimental rates of folding, unfolding, and amino acid addition were run for comparison. Additional sets of simulations were then run with the *in silico* rates calculated from temperature quenching with $\alpha = \{1, 10, 100, 200, 500, 1000, 10000\}$ (see Fig. S7). These simulations suggest that an α of up to $\sim 1,000$ is reasonable, consistent with the notion that significant departure from quasi-equilibrium behavior is only expected when $\langle \tau_A^{\text{exp}} \rangle \cong \langle \tau_F^{\text{exp}} \rangle$, and there is ~ 1400 -fold difference between $\langle \tau_A^{\text{exp}} \rangle$ and $\langle \tau_F^{\text{exp}} \rangle$. We chose $\alpha = 500$ as a compromise between computational expense and preservation of the experimental ratio of timescales as closely as possible. Thus, $\langle \tau_A^{\text{sim}} \rangle' = \frac{5080 \text{ ns}}{500} = 10.2 \text{ ns}$ was used for all continuous synthesis simulations.

Continuous synthesis simulations. A coarse-grained representation of the *E. coli* ribosome's large subunit was constructed from PDB ID: 3UOS as described previously²⁹. A partial sphere of 26 ribosomal interaction sites within 12 Å of the point (6, 0, 0) and with x-coordinate greater than 3 Å were harmonically restrained using CONS HARM with a force constant of $0.5 \text{ kcal/mol} \cdot \text{Å}^2$. A single ribosome interaction site corresponding to U2585's uracil ring (by PDB ID: 3UOS numbering) was deleted from the PDB structure in order to avoid steric clashes with the nascent chain. All interactions between ribosome sites were deactivated with the BLOCK module in CHARMM. The C-terminal bead of the nascent chain was held at the point (6, 0, 0) by a spherical harmonic restraint with a force constant of $50 \text{ kcal/mol} \cdot \text{Å}^2$. Five planar restraints were positioned on five sides around the nascent chain using the GEO PLANE functionality in the MMFP module of CHARMM for nascent chain lengths up to and including 15 residues to direct the nascent chain

into the exit tunnel. These planar restraints use the potential form EXPONENTIAL and parameters FORCE 50, DROFF 2.0, and P1 0.050. One yz-plane, passing through the point (1, 0, 0), was used to approximate the steric bulk of the P-site tRNA. Four additional planes were used to define a box around the nascent chain, open only on the face pointing out of the exit tunnel, to help guide it into the exit tunnel correctly. Two xz-planes, with one passing through the point (0, 20, 0) and the other through the point (0, -10, 0), were used. Two xy-planes, one passing through the point (0, 0, 20) and the other through the point (0, 0, -20), were also used. For nascent chain lengths greater than 15 residues only the yz-plane through (1, 0, 0) was retained to approximate the steric bulk of the tRNA. This combination of planar and harmonic restraints prevents early termination due to steric clashes that lead to SHAKE algorithm errors.

A modified version of a previously published continuous synthesis protocol for CHARMM was used³⁰. Simulations were begun with the C α bead of the N-terminal residue (BOF-Met) restrained at the point (6, 0, 0). The dwell time at each nascent chain length was randomly sampled from a single-exponential distribution with mean $\langle \tau_A^{\text{sim}} \rangle' = 10.2$ ns. A temperature of 310 K, frictional coefficient of FBETA = 0.050 ps⁻¹, an integration time step of 0.015 ps, and the SHAKE algorithm were used for all continuous synthesis simulations. System coordinates were saved every 5000 integration time steps (75 ps). Two-hundred, one-hundred, and eighty synthesis trajectories were completed using the folding, good-solvent, and poor-solvent Hamiltonians, respectively.

Each poor-solvent Hamiltonian trajectory was initiated from a different configuration from the folding Hamiltonian simulations at nascent chain length 33. Starting at a nascent chain length of 33 ensures that the nascent chain will be equilibrated by the time the BOP-Lys residue is added at position 34 and FRET becomes possible. The time series for continuous synthesis trajectories synthesized with the poor-solvent Hamiltonian were adjusted by appending thirty-two dwell times selected randomly from an exponential distribution with mean of 10.2 ns. This adjustment accounts for the time that would be required to synthesize the first 32 residues of the nascent chain that were not explicitly simulated with Langevin dynamics. The size of the harmonically restrained region of ribosomal interaction sites was also increased to a total of 95 ribosome interaction sites within the box bounded by $x = [3, 30]$, $y = [-15, 10]$, and $z = [-15, 10]$. The poor-solvent simulations were otherwise identical to the folding and good-solvent simulations. Trajectories for HemK98, HemK84, HemK70, HemK56, and HemK42 with each of the three Hamiltonians were extended by an appropriate amount to bring their total duration to ~35 s when mapped to experimental time.

Mapping of simulation timescales. In order to compare our simulation results to experimental time series we mapped our simulation time course to experimental time. We assume that there exists a uniform scaling factor, c , that maps the timescale of our simulations to experimental time:

$$t_{\text{exp}} = c * t_{\text{sim}}. \quad [3]$$

If we assume that $t_{\text{exp}} = \langle \tau_A^{\text{exp}} \rangle$ and $t_{\text{sim}} = \langle \tau_A^{\text{sim}} \rangle$, we have

$$c = \frac{\langle \tau_A^{\text{exp}} \rangle}{\langle \tau_A^{\text{sim}} \rangle}. \quad [4]$$

As we employed an acceleration factor, α , this equation must be adjusted to the form

$$c = \frac{\langle \tau_A^{\text{exp}} \rangle}{\frac{\langle \tau_A^{\text{sim}} \rangle}{\alpha}} = \frac{\langle \tau_A^{\text{exp}} \rangle}{\langle \tau_A^{\text{sim}} \rangle'}. \quad [5]$$

Thus, we mapped our simulated time series in units of nanoseconds onto the experimental time regime in units of seconds by the procedure

$$t_{\text{exp}} = t_{\text{sim}} * \frac{1 \text{ s}}{10^9 \text{ ns}} * \frac{\langle \tau_A^{\text{exp}} \rangle}{\langle \tau_A^{\text{sim}} \rangle}, \quad [6]$$

in which t_{sim} has units of ns and both $\langle \tau_A^{\text{exp}} \rangle$ and $\langle \tau_A^{\text{sim}} \rangle$ have units of s.

Calculation of E and $\frac{E}{E_{\text{end}}}$ ensemble average time series. The Förster equation was used to calculate the FRET efficiency as a function of the inter-dye distance, r , according to Equation 1. A Förster radius, R_0 , of 54 Å was used²⁷. The distance between the donor and acceptor was calculated as the distance between the interaction sites representing the six-membered rings in BOF and BOP at each frame of a trajectory. The FRET efficiency time series for each trajectory was then time averaged into 15-ns bins. Finally, the binned trajectories were averaged together to produce the ensemble average FRET time series. Statistics were generated by determining the standard error of the average for each bin across all trajectories and then calculating the corresponding 95% confidence interval. Time series of $\frac{E}{E_{\text{end}}}$ were produced by dividing the ensemble average E time series for each construct by the value in the final bin for the HemK112 ensemble average time series for the corresponding Hamiltonian. These ensemble average trajectories were then projected onto experimental time by multiplication by the time adjustment factor, c . Calculation of E and $\frac{E}{E_{\text{end}}}$ time series for alternative dye positions was carried out in the same fashion except the inter-dye distance was taken to be the distance between the C_α interaction sites for the dye locations under consideration.

Fraction of native contacts analysis. The fraction of native contacts (Q) formed by the five helices within the HemK Gō-dye model was calculated from the folding Hamiltonian continuous synthesis simulations. First, the set of contacts within the coarse-grain, native-state reference structure was determined. Two interaction sites are considered to share a contact if two criteria are satisfied: (1) both residues are within structured regions and separated by at least three residues in the primary sequence (*i.e.*, $i \rightarrow i + 4$ or greater separation) and (2) the coarse-grain residues in the reference structure are no more than 8 Å apart. A contact was considered to be formed during a continuous synthesis trajectory if it satisfied criteria (1) above and the distance between the residues was less than or equal to $1.2 * d_{\text{ref}}$, where d_{ref} is the distance between the residues in the reference structure and the factor of 1.2 is to adjust for thermal fluctuations. The total number of contacts formed at frame j of the simulation, $N(j)$, was divided by the total number of contacts formed in the native state, N_{ref} , to produce the value of $Q(j)$. Time series of $Q(j)$ were then binned, averaged, and the corresponding statistics determined in the same manner as described for the E and $\frac{E}{E_{\text{end}}}$ time series.

Comparing simulated and experimental time series. Pearson R^2 values were calculated between simulated E or $\frac{E}{E_{\text{end}}}$ time series and experimental $\frac{F_{\text{AD(A)}}}{F_{\text{AD(A)}}^{\text{end}}}$ time series. The experimental time point temporally closest to each of the simulation time point (after appropriate mapping of the simulated time series to the experimental time frame of reference) was determined and the

Pearson R^2 calculated between this set of experimental values and the simulation time series. All p -values are $<10^{-10}$. We also introduce the parameters λ defined as

$$\lambda = \sqrt{\frac{\sum (E_{H1}(k) - E_{H2}(k))^2}{N}}, \quad [7]$$

where the summation is evaluated over all bins, k , in the ensemble-average time series for a particular HemK construct and $E_{H1}(k)$ and $E_{H2}(k)$ are the ensemble average E values in bin k for the Hamiltonian 1 and Hamiltonian 2 time series, respectively, which are selected from the folding, good-solvent, and poor-solvent time series. The term N is equal to the total number of bins across which λ is calculated. When $\frac{E}{E^{\text{end}}}$ time series rather than E time series are compared, the terms $E_{H1}(k)$ and $E_{H2}(k)$ are replaced with $\frac{E_{H1}(k)}{E_{H1}^{\text{end}}}$ and $\frac{E_{H2}(k)}{E_{H2}^{\text{end}}}$, respectively, in which E_{H1}^{end} and E_{H2}^{end} are the final values for the HemK112 time series calculated with the corresponding Hamiltonian's simulation results.

References

1. Kim, S. J. *et al.* Translational tuning optimizes nascent protein folding in cells. *Science*. **348**, 444–448 (2015).
2. Holtkamp, W. *et al.* Cotranslational protein folding on the ribosome monitored in real time. *Science*. **350**, 1104–1107 (2015).
3. Kajihara, D. *et al.* FRET analysis of protein conformational change through position-specific incorporation of fluorescent amino acids. *Nat. Methods* **3**, 923–929 (2006).
4. Mercier, E. & Rodnina, M. V. Co-Translational Folding Trajectory of the HemK Helical Domain. *Biochemistry* **57**, 3460–3464 (2018).
5. Nilsson, O. B. *et al.* Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. Mol. Biol.* **24**, 221–225 (2017).
6. Stirnemann, G., Giganti, D., Fernandez, J. M. & Berne, B. J. Elasticity, structure, and relaxation of extended proteins under force. *Proc. Natl. Acad. Sci.* **110**, 3847–3852 (2013).
7. Schlierf, M., Li, H. & Fernandez, J. M. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl. Acad. Sci.* **101**, 7299–7304 (2004).
8. Walther, K. A. *et al.* Signatures of hydrophobic collapse in extended proteins captured with force spectroscopy. *Proc. Natl. Acad. Sci.* **104**, 7916–7921 (2007).
9. Flory, P. J. Principles of Polymer Chemistry. (1953).
10. Flory, P. J. & Volkenstein, M. Statistical mechanics of chain molecules. *Biopolymers* **8**, 699–700 (1969).
11. Dill, K. A. & Shortle, D. Denatured States of Proteins. *Annu. Rev. Biochem.* **60**, 795–825 (1991).
12. O'Brien, E. P., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. New scenarios of protein folding can occur on the ribosome. *J. Am. Chem. Soc.* **133**, 513–526 (2011).
13. Caniparoli, L. & O'Brien, E. P. Modeling the effect of codon translation rates on co-translational protein folding mechanisms of arbitrary complexity. *J. Chem. Phys.* **142**, (2015).
14. O'Brien, E. P., Vendruscolo, M. & Dobson, C. M. Prediction of variable translation rate effects on cotranslational protein folding. *Nat. Commun.* **3**, 868 (2012).
15. 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.* **134**, 10920–10932 (2012).
16. Fritch, B. *et al.* Origins of the Mechanochemical Coupling of Peptide Bond Formation to Protein Synthesis. *J. Am. Chem. Soc.* **140**, (2018).
17. Thommen, M., Holtkamp, W. & Rodnina, M. V. Co-translational protein folding: progress and methods. *Curr. Opin. Struct. Biol.* **42**, 83–89 (2017).

18. 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.* **110**, E132-40 (2013).
19. Siller, E., DeZwaan, D. C., Anderson, J. F., Freeman, B. C. & Barral, J. M. Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J. Mol. Biol.* **396**, 1310–18 (2010).
20. Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12**, 683–691 (2011).
21. Merchant, K. A., Best, R., Louis, J., Gopich, I. & Eaton, W. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations - SI. *Proc. Natl. Acad. Sci.* **104**, (2007).
22. Best, R. B., Hofmann, H., Nettels, D. & Schuler, B. Quantitative Interpretation of FRET Experiments via Molecular Simulation: Force Field and Validation. *Biophys. J.* **108**, 2721–2731 (2015).
23. Henry, E. R. & Hochstrasser, R. M. Molecular dynamics simulations of fluorescence polarization of tryptophans in myoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6142–6146 (1987).
24. 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.* **105**, 13403–13408 (2008).
25. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
26. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
27. Mittelstaet, J., Konevega, A. L. & Rodnina, M. V. A Kinetic Safety Gate Controlling the Delivery of Unnatural Amino Acids to the Ribosome. *J. Am. Chem. Soc.* **135**, 17031–17038 (2013).
28. Best, R. B., Chen, Y. G. & Hummer, G. Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of Arc repressor. *Structure* **13**, 1755–1763 (2005).
29. 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.* **134**, 10920–32 (2012).
30. Trovato, F. & O'Brien, E. P. Fast Protein Translation Can Promote Co- and Posttranslational Folding of Misfolding-Prone Proteins. *Biophys. J.* **112**, 1807–1819 (2017).

Table 1 | Pearson R^2 values between simulated time series and experimental time series

Hamiltonian	HemK112	HemK98	HemK84	HemK70	HemK56	HemK42
Folding model	0.98	0.97	0.97	0.95	0.78	0.76
Good solvent	0.99	0.99	0.96	0.98	0.88	0.76
Poor solvent	0.65	0.66	0.51	0.58	0.55	0.82

Table 2 | λ values for comparisons between E and E/E^{end} time series

λ values for E time series						
Hamiltonians compared	HemK112	HemK98	HemK84	HemK70	HemK56	HemK42
Folding vs. good solvent	0.162	0.160	0.148	0.054	0.064	0.086
Folding vs. poor solvent	0.228	0.227	0.229	0.164	0.243	0.427
Good solvent vs. poor solvent	0.203	0.215	0.210	0.215	0.305	0.513

λ values for E/E^{end} time series						
Hamiltonians compared	HemK112	HemK98	HemK84	HemK70	HemK56	HemK42
Folding vs. good solvent	0.073	0.075	0.083	0.152	0.076	0.049
Folding vs. poor solvent	0.298	0.316	0.308	0.360	0.482	0.688
Good solvent vs. poor solvent	0.292	0.311	0.302	0.312	0.444	0.736

Table 3 | λ values for comparisons between alternative dye position E and E/E^{end} time series

λ values for alternative dye position analysis of E time series						
	HemK112	HemK98	HemK84	HemK70	HemK56	HemK42
Best location	d3, a66	d3, a64	d3, a56	d42, a64	d4, a31	d3, a29
λ, (% change)	0.365, (+125%)	0.325, (+103%)	0.267, (+80%)	0.159, (+194%)	0.136, (+113%)	0.216, (+151%)

λ values for alternative dye position analysis of E/E^{end} time series						
	HemK112	HemK98	HemK84	HemK70	HemK56	HemK42
Best location	d38, a64	d35, a73	d36, a73	d1, a38	d19, a43	d2, a29
λ, (% change)	0.095, (+30%)	0.159, (+112%)	0.250, (+201%)	0.146, (-4%)	0.087, (+14%)	0.169, (+245%)

Figure 1. Coarse-grain representation of HemK NTD with dye-modified residues. (A) Left: cartoon model superimposed over an all-atom representation of HemK NTD (residues 2-73, PDB ID: 1T43). Right: C_{α} coarse-grain representation of HemK NTD. Helices 1, 2, 3, 4, and 5 are shown in red, blue, yellow, cyan, and silver, respectively. Unstructured regions are shown in ochre. (B) Schematic of the 112-residue HemK construct with dye positions indicated. (C) All-atom and coarse-grained representations of the FRET Donor/Acceptor pair BOF/BOP. Interaction sites are shown in magenta while virtual bonds are shown in green. The leftmost bead in each image corresponds to the residue's C_{α} interaction site. (D) Cross-section of a coarse-grained ribosome-nascent chain complex harboring a 112-residue HemK nascent chain. Ribosomal RNA and ribosomal protein interaction sites are shown in magenta and light purple, respectively. The coloring scheme for the N-terminal domain of HemK is the same as in (A). The unstructured and structured regions of the C-terminal linker are displayed in ochre and green, respectively. The FRET dyes are shown in pink.

Figure 2. Simulation time series compared to experimental results. (A) Simulated E time series are plotted alongside experimental $\frac{F_{AD(A)}}{F_{AD(A)}^{end}}$ time series with experimental fluorescence data plotted as green circles. Ensemble average simulation time series for the folding, good-solvent, and poor-solvent Hamiltonians are displayed in blue, magenta, and orange, respectively. Simulation error bars are 95% confidence intervals that are, for some data points, smaller than the plotted line. (B) Same as (A), but simulation time series are $\frac{E}{E^{end}}$ rather than E . Note that the left set of y axes applies only to the experimental time series.

Figure 3. Alternative dye positions that maximize separation between good-solvent and folding results for each HemK construct. (A) Simulated E time series calculated with the Förster Equation using explicit dye representation coordinates (folding Hamiltonian: blue; good-solvent Hamiltonian: magenta) or calculated using alternative dye position C_α coordinates (folding Hamiltonian: green; good-solvent Hamiltonian: orange). Simulation error bars are 95% confidence intervals that are, for some data points, smaller than the plotted line. (B) Same as (A) but time series are $\frac{E}{E_{\text{end}}}$ rather than E .

Figure 4. Three sets of fraction of native contact, Q , time series are displayed for HemK112 alongside representative coarse-grain nascent-chain structures. (A) Time series of Q representing intra-helix contacts for helices 1, 2, 3, 4, and 5 within HemK NTD. (B) Time series of Q representing inter-helix contacts between all pairs of helices between which contacts form in HemK NTD native state. The plot legend indicates the pair of helices considered for each time series. (C) Time series of Q representing overall folding (i.e., intra- and inter-helix contacts) calculated over either all five helices (h12345), the N-terminal four helices (h1234), or the N-terminal three helices (h123). Error bars are 95% confidence intervals. (D) Structures corresponding to key folding intermediates labeled with the nascent chain length at which they occur. Coloring is the same as in Fig. 1. The orange sphere indicates the position at which the C-terminal residue is restrained. The 50S ribosome representation is removed for visual clarity but was present during simulations.