Non-equilibrium dynamics of a nascent polypeptide during translation suppress its misfolding

Lisa M. Alexander1, Daniel H. Goldman1,7, Liang M. Wee2,3,4 & Carlos Bustamante1,2,3,4,5,6

Protein folding can begin co-translationally. Due to the difference in timescale between folding and synthesis, co-translational folding is thought to occur at equilibrium for fast-folding domains. In this scenario, the folding kinetics of stalled ribosome-bound nascent chains should match the folding of nascent chains in real-time. To test if this assumption is true, we compare the folding of a ribosome-bound, multi-domain calcium-binding protein stalled at different points in translation with the nascent chain as it being synthesized in real-time, via optical tweezers. On stalled ribosomes, a misfolded state forms rapidly (1.5 s). However, during translation, this state is only attained after a long delay (63 s), indicating that, unexpectedly, the growing polypeptide is not equilibrated with its ensemble of accessible conformations. Slow equilibration on the ribosome can delay premature folding until adequate sequence is available and/or allow time for chaperone binding, thus promoting productive folding.
T

ough folding, polypeptides adopt the specific three-dimensional structures required for their function. In the cell, some of these structures begin to form cotranslationally1,2. However, the overall stability of the folded protein is reduced with respect to the folded state off the ribosome3–5, and the rate of folding can be slower6. Such suppression of folding rates during vectorial synthesis by the ribosome is thought to increase the amount of active protein by promoting folding in a domain-wise manner and by preventing off-pathway misfolded states7,10.

Studies have shown that smaller domains can fold inside the ribosome exit tunnel starting when the ribosome has decoded 25 amino acids downstream of the domain boundary11,12, whereas larger nascent domains require synthesis of ~40 amino acids past the domain boundary to fold13,14. Overall, these results suggest that the nascent chain can fold as soon as there is enough space in the tunnel or at the ribosome surface such that the folded state is not sterically precluded15. Accordingly, most co-translational folding studies employ stalled ribosome-bound nascent chain complexes (RNCs), which also ensure a uniform population during analysis6,8,12. However, whether the stalling is permanent (e.g., generated by mRNA truncation) or temporary (e.g., induced by ribosome stalling sequences such as SecM), the implicit assumption has been that the observed folding dynamics faithfully reflect the dynamics that occur during synthesis, as the transitions are thought to be fast relative to the rate of translation and the folding process is presumed to occur at equilibrium16,17. Although this assumption may be justified in some cases18 it presupposes that the rates of elongation and folding are the only relevant ones controlling the overall behavior of the nascent chain. In fact, other processes may occur contemporaneously and the chain may not begin to fold immediately upon synthesis.

To test this hypothesis, we investigate the folding of a multidomain protein, calerythrin, on an actively translating ribosome and compare it to the protein’s folding on the surface of stalled RNCs and its folding off the ribosome. We show on stalled complexes that the ribosome can decrease the kinetics of folding and increase the unfolding rate of an off-pathway misfolded state. We also demonstrate that the ribosome reduces the stability of an N-terminal intermediate, precluding its formation during active translation, despite the fact that the N terminus is the earliest independently folding unit to emerge from the ribosome. Strikingly, we find that the folding pathway during active elongation is distinct from that derived from our stalled RNC data. Specifically, calerythrin exhibits non-equilibrium folding on actively translating ribosomes, whereby the onset of co-translational folding is delayed, suppressing the misfolded state. Together, our results indicate that a co-translational folding pathway is not necessarily determined by the rates of elongation and equilibrium folding alone, and that non-equilibrium contributions can enhance the probability of proper native folding.

Results

Full-length calerythrin folds through the C domain. To investigate the folding of a multi-domain protein in solution, on stalled RNCs, and during synthesis, we used an optical tweezers (OT) assay that enables us to selectively denature the polypeptide (using force) while leaving the ribosome intact. With this assay, we can directly detect folding intermediates and measure the rates of folding and unfolding. We selected as a model protein calerythrin, a two-domain calcium-binding protein from the bacterium *Saccharopolyspora erythraea*19,20. Calerythrin is a member of the EF-hand family of proteins. It contains four EF hands, each of which consists of a helix-loop-helix motif with a calcium-binding site in the loop. One pair of hands (EF1 and EF2) forms the N domain and another pair (EF3 and EF4) the C domain (Fig. 1a).

Nuclear magnetic resonance and circular dichroism studies suggest that the C domain folds more quickly than the N domain upon addition of calcium21. EF2 is noncanonical and does not bind calcium22. Other members of the EF-hand family have been previously studied using OT and these studies have yielded detailed measurements of folding networks off the ribosome23–25. In order to provide a benchmark to interpret our real-time elongation results, we first characterized the folding of calerythrin off the ribosome and the equilibrium folding kinetics of stalled RNCs.

We tethered the isolated full-length (FL) protein by attaching the N and C termini to DNA handles (Fig. 1a, b) and collected data via force–extension measurements, where the traps move apart at a fixed rate. In this measurement, applying load (increasing force) leads to unfolding of calerythrin (detected as increasing extension). In a typical force–extension curve, at 13 ± 1.3 pN (s.d.), FL calerythrin unfolds in a single rip, indicating cooperative unfolding (Fig. 1c). The unfolding transition is 65 ± 2 nm, or 177 ± 5 amino acids, as expected for the full unfolding of calerythrin, which is 177 amino acids in length. Although both domains unfold in one transition, the domains refold sequentially in two transitions (Fig. 1c). To investigate the refolding in more detail, we collected data in passive mode (where the traps remain stationary, as contrasted with force–extension measurements where the traps move apart at a fixed rate). In this mode, the force suddenly increases if folding occurs. The passive data show hoppings between unfolding and an intermediate state before FL folding (Fig. 1d). FL folding is long-lived at this force and must be unfolded at high force to observe further transitions. We used a Hidden Markov Model (HMM)24,26 to extract the lifetimes of the unfolded and intermediate states as a function of force (Fig. 1e). Because the HMM with only one intermediate converges to single exponential distributions (Supplementary Fig. 1a) and this state’s kinetics follow Bell’s model as expected, we conclude that there is only one intermediate (FLint), and the protein refolds every time through this same state before reaching the fully folded state (Fig. 1d, e).

Next, we measured the folding of each domain individually by preparing truncated proteins of either the N domain (amino acids 1–90) or the C domain (91–177) with the same tags on the termini. Folding of the isolated N domain involves more amino acids, is slower, and occurs at lower forces than FLint (two-tailed Welch’s t test, *p* < 0.05) (Fig. 1c–e, Supplementary Table 1). In contrast, folding of the isolated C domain is indistinguishable from FLint in number of amino acids and the folding and unfolding kinetics (Fig. 1c–e, Supplementary Table 1), suggesting that FLint corresponds to the folding of the C domain.

To confirm this assignment, we individually mutated the conserved 12th position glutamate/aspartate to glutamine in the EF-hand loops of hands 1, 3, or 4, thus impairing calcium binding and destabilizing one EF-hand fold at a time27,28 (EF hand 2 does not bind calcium in the natural sequence29). The glutamine mutation in EF1 (belonging to the N terminus domain) led to a significant change in the unfolding transition of the FL protein, which is no longer cooperative and the domains unfold in two sequential steps. However, this mutation does not affect the folding or unfolding kinetics of the first intermediate state, which is indistinguishable from the wild-type FLint (Supplementary Fig. 2a). In contrast, mutation of this residue in EF3 or EF4 destabilizes the intermediate so that it only folds at lower forces and unfolds more rapidly (Supplementary Fig. 2b, c). These mutant studies confirm that the FLint intermediate corresponds to the folded C domain. Accordingly, the folding pathway of the isolated FL molecule proceeds from the unfolded state through the C domain to the native structure. Folding through the C
domain exclusively, rather than through either domain as in calmodulin, has been reported for another EF-hand protein NSC125, which like calerythrin also has a noncanonical EF hand in the N domain.

N domain folding is suppressed on the ribosome. Having established the off-ribosome folding pathway of calerythrin, we next sought to characterize how the surface of the ribosome modifies the folding of the protein in stalled complexes, wherein the process occurs at equilibrium. We prepared RNCs stalled at codons 135–177 in 10–11 codon increments (Fig. 2a) and tethered them via the N terminus of the nascent chain to one DNA handle and a ybbR tag on protein L17 of the ribosome to the other DNA handle (Fig. 2b). At codon 135, the N domain is completely outside of the exit tunnel. However, we do not observe folding for this complex (RNC135), nor for RNC146 or RNC157 (Fig. 2c, Supplementary Fig. 3). The isolated N domain off the ribosome folds at a rate of $k_0 = 660 \pm 370 \text{ s}^{-1}$, so we infer that the ribosome destabilizes the N domain by reducing its folding rate or increasing its unfolding rate relative to the isolated N domain such that it is not sampled. A decrease in folding rate of up to two orders of magnitude has been reported for other proteins such as T4 lysozyme on the ribosome, although in that case folding was reduced but not eliminated. The N domain intermediate observed here has lower stability than many full proteins or domains, and the destabilization may be sufficient to prevent folding entirely. Thus, even though the N domain is readily sampled for the isolated protein off the ribosome, and the vectorial N-to-C nature of protein synthesis precludes early formation of the C domain, folding on the ribosome does not proceed through an N-terminal intermediate.

**Stalled RNCs misfold.** We next extended the polypeptide further to generate RNC167 and RNC177 (RNC177 is FL but with part of the C domain in the tunnel). Surprisingly, although we still failed...
to observe folding of the N domain, we observed a misfolded state, M, of 130 ± 15 amino acids, significantly larger than either the N or C domains alone, for both RNC167 and RNC177 (Fig. 2c–e, also compare to Fig. 1d and Supplementary Table 1). Only by extending the entire protein sequence outside of the exit tunnel with a 45 amino acid artificial linker (RNC177 +45) is native-like folding restored (Fig. 2c, d). At this stage, the misfolded state is not observed because native folding through the C domain occurs at higher forces and is more stable, outcompeting the misfolded state. These observations explain why the misfolded state is not seen in isolated FL protein experiments.

The misfolded state has altered kinetics on the ribosome. Although the misfolded state observed with RNC167 and RNC177 is not observed for the fully synthesized and released protein, it is accessible to a truncated version of the protein lacking the 4th EF hand (EF123, amino acids 1–134), owing to the lack of sequence required for C domain folding. The isolated EF123 can fold to either state M or the N domain in solution (Fig. 3a, Supplementary Fig. 4). Moreover, the misfolded and folded N domain states of EF123 do not interchange without first reverting to the fully unfolded state, as shown from our HMM analysis and analysis of the point-spread function31 (Supplementary Fig. 4). This observation suggests that the misfolded state is off-pathway for the nascent chain, involving non-native interdomain contacts that preclude the correct folding of the individual domains. Misfolding of EF-hand proteins has been reported previously, either through mispairing EF hands or via

Fig. 2 Stalled RNCs misfold. a The construct design for stalling. Using truncated mRNAs that lack a stop codon, we are able to generate homogeneously stalled ribosomes. The last codon added (in the P site) corresponds to the name of the RNC; less sequence than this will be outside the tunnel. b The OT geometry uses an N-terminal Avi tag and a tag on the ribosome itself. c Refolding force–extension curves of our RNCs do not show any folding until codon 167 or 177. WLC models are shown as dotted black or red lines. The WLC model assumed 40 amino acids in the ribosomal exit tunnel. For RNC167 and RNC177, folded WLCs show a transition of 135 amino acids. For RNC177 +45, folded WLCs correspond to a transition of 75 and 177 amino acids. See Supplementary Fig. 3 for further testing. Representative force–extension curves from n = 3, 10, 5, 10, 5 molecules, respectively. d An example of the folding of RNC177 and RNC177 +45. RNC177 has a misfolded state larger than the C domain (75 amino acids) but smaller than the FL protein (177 amino acids). RNC177 +45 folds to the FL state. The dotted lines are WLC models to 177, 135, 75, and 0 amino acids folded (black, magenta, black, black, respectively). e Passive mode data of RNC167 and RNC177 shows folding transitions to state M. HMM fit shown as dashed lines (red for RNC167 and black for RNC177). Source data are provided as a Source Data file.
off-pathway compaction of partially structured intermediates. As for calerythrin, since the misfolded state forms directly from the unfolded state and involves the entire available sequence of the truncated protein, state M could involve mispaired EF hands, similar to a previously reported misfolded state of calmodulin. EF2 is also a noncanonical EF hand that does not bind calcium, potentially allowing EF3 folding to preempt EF2 folding. Thus, although the FL protein off the ribosome usually avoids the misfolded state owing to the faster folding of the C domain, M is accessible on the surface of the ribosome when the sequence needed to form the C domain is still partly sequestered in the ribosomal exit tunnel and the ribosome has destabilized the N domain (as shown with RNC135).

The misfolding occurs at much lower forces on the ribosome than on the isolated polypeptide, as observed in the overlaid force–extension relaxation curves of RNC177 and EF123 (Fig. 3b). Significantly, the ribosome decelerates the folding kinetics of the misfolded state at codons 167 and 177 by 104-fold and accelerates the unfolding rate by 90-fold at zero force compared with the free truncation EF123 (Fig. 3b, c, Supplementary Table 2). These trends reduce the probability that the RNC attains the misfolded state and increase the probability that the molecule escapes it to recover productive folding if it does form. Notice, however, that the time until misfolding on the ribosome at zero force, whereas longer than for the free EF123 construct, is still quite short (6 ± 3 × 10^−4 s) and the time it remains trapped in this state (1.9 ± 0.5 s) is still relatively long. Assuming that our equilibrium rates on room temperature are the only determinants of co-translational folding, the misfolded state would be expected to form readily during translation in spite of the relative destabilization by the ribosome. In real-time elongation conditions, assuming the rates we have measured do not change substantially with chain length (as is the case between codons 167 and 177), we would expect on average one misfolding event per 1.5 sec at 4.0 pN.

Translating RNCs adopt non-equilibrium trapped states. Next, we asked: are the kinetics to-and-from the misfolded state—as determined at equilibrium on the stalled RNCs—predictive of the order of folding events during active translation? To answer this question, we needed to follow the growth of calerythrin in real-time while simultaneously monitoring its co-translational folding. Accordingly, we generated stalled RNCs by amino-acid omission (Supplementary Fig. 5), and restarted peptide elongation by introducing an in vitro translation mix (Fig. 4a, Supplementary Table 3). The mRNA was engineered to contain 225 codons to translate in the tweezers: a three-valine stall site, all 177 amino acids of native calerythrin, and the same C-terminal linker used for generating the stalled RNC177+45 (see Fig. 2a). As these experiments are conducted in passive mode, active elongation is observed as a drop in force concomitant with the polypeptide length increase (Fig. 4b). Folding transitions appear as a sudden increase of the tension across the tether. We conducted translation at 4.5–3.5 pN, so that it is in the similar range of forces as probed using stalled RNCs.

We converted the change in force to amino-acid position using a worm-like chain (WLC) model for the polypeptide and an extensible WLC for the DNA handles, and reported cellular rates of elongation (4–6 s⁻¹ at room temperature) are the only determinants of cotranslational folding, the misfolded state would be expected to form readily during translation in spite of the relative destabilization by the ribosome. In real-time elongation conditions, assuming the rates we have measured do not change substantially with chain length (as is the case between codons 167 and 177), we would expect on average one misfolding event per 1.5 sec at 4.0 pN.
point out that our in vitro system does not perfectly recreate the cellular environment— for example, it does not include chaperones and does not mimic the extent of molecular crowding in the cell.

At forces below 5 pN, we observe transitions to the folded C domain or to the misfolded state (Fig. 4d, red arrows mark the M state or C domain). Every molecule exhibits misfolding or C domain folding, but N domain folding is not observed, as expected from the stalled RNC data (Fig. 2c). Most surprisingly, however, although the misfolded state is detected during real-time translation, it does not form readily when the ribosome has reached codon 167, as would be expected from our experiments with equilibrated stalled complexes. Instead, we observe a delay until the onset of folding, after which the chain (mis)folds and unfolds multiple times (hopping) as previously observed with the stalled RNCs (Fig. 5a–c; the delay time \( \tau \) is indicated, at the end of which the hopping is observed).

We measured the initial delay as the time elapsed between the moment when the ribosome reaches codon 167 to the time when we first observed a folding transition (Supplementary Fig. 6), either to the misfolded state (75% of the molecules) or to the folded C domain (25%) (Fig. 5b, c). In 8% of molecules, the delay is essentially indistinguishable from the lifetime of the unfolded state in stalled complexes \( \tau_{\text{unfold}} \), demonstrating that this outcome is observable in our assay, though not very common. The distribution of measured \( \tau \) values is exponential (Supplementary Fig. 7), with the characteristic delay until the onset of folding \( \tau_{\text{delay}} = 63 \pm 12 \) s from the fit to the cumulative distribution (Fig. 5d). This exponential distribution supports the idea that, as it is being synthesized, the polypeptide is kept out of equilibrium in the unfolded state until it undergoes a stochastic equilibration step. The characteristic delay is 40 times longer than the measured lifetime of the unfolded state at equilibrium in RNC167 and RNC177 \( \tau_{\text{unfold}} = 1.5 \pm 0.3 \) s at 4.0 pN, the average force at which the elongating ribosomes reach codon 167). This delay is not owing to interactions between components of the translation mix and the nascent chain, nor to linker effects, as experiments with stalled RNCs in translation mix or added linkers did not display a delay (Supplementary Fig. 8).

This behavior indicates that prior to the onset of folding, the newly synthesized polypeptide resides in a trapped unfolded state that is not folding-competent. Although there is a proline in the sequence of the C domain, this trapped state is unlikely to involve a set of local interactions with the ribosome surface or exit tunnel that were established during its growth in real time or, alternatively may form only in the context of actively elongating ribosome surface dynamics. For example, in real time, there will be local dynamics of ribosomal components as factors bind/unbind and ribosomal elements alter conformation. It is also

**Fig. 4** Set-up for real-time elongation OT experiments. a The sequence for stalling and restart. RNCs were stalled at a three-valine repeat (see Supplementary Fig. 5) and restarted with the addition of translation mix in the OT (Supplementary Table 3). The force–extension curves show the lengthening of the tether after translating (red line denotes difference between blue and black force–extension curves) as well as folding of the full-length protein, which confirms the ribosome reached the end of the mRNA. b An example of ribosome translating. As each amino acid is added, the force decreases as the tether gets longer. The fit is shown as a red dotted line. c The trajectory is converted to amino acids using a worm-like-chain model. d A zoom of the panel in b that shows the folding region. The red arrows show the misfolded state and the dotted line the expected size for state M. e A zoom of the panel in b that shows the folding region later in time. At this point, both C and M are accessible (dotted lines show expected sizes, arrows mark some the transitions). Representative molecule from \( n = 12 \). Source data are provided as a Source Data file.
possible that the chain may form local structures when it emerges from the tunnel that are different from a random coil unfolded state, that may be folding incompetent. For example, small helical sections are known to fold in the tunnel that may prevent the folding process. Once the chain manages to escape from this trapped state, however, it is seen to (mis)fold and unfold without visiting again the trapped state, becoming equilibrated within its energy landscape. In support of this interpretation, once the misfolded state forms, ramping the force to unfold it and returning it to forces as low as 2 pN, we observe no subsequent delay; instead, hopping between unfolded and (mis)folded states resumes, as in stalled complexes (Supplementary Fig. 9).
assays where ribosomes are not synchronized\textsuperscript{13}, which could make deviations from equilibrium behavior difficult to detect, depending on the magnitude of the delay and the folding rates in question. Using a real-time single-molecule assay we detect a clear signature of non-equilibrium dynamics that may play an important role in the fate of the newly synthesized polypeptide. The polypeptide is not equilibrated with the ensemble of accessible conformations measured in stalled complexes; rather, it persists in the unfolded state until some equilibration step occurs. Our results show that comparing equilibrium-derived folding rates from stalled RNCs with translation rates is not always sufficient to predict co-translational folding of a polypeptide. Indeed, in this case we find that the peptide product must be observed in real-time as it is being generated, as other processes may also be at play that are obscured in stalled complex measurements.

Stalled complexes require, for most experimental techniques, at minimum tens of minutes to prepare, which allows time for equilibration processes to occur. If these processes could be induced in stalled complexes, we would expect to be able to recreate the co-translational behavior during discontinuous synthesis; however, as the exact nature of this process is unknown and was not previously predicted, the measurement must be conducted in real-time. The real-time folding assay developed in this study will also be useful in other situations where measurements of stalled RNCs may not capture the full dynamics of folding, such as length-dependent protein conformational changes, dynamic interactions with chaperones, or folding near domain boundaries.

Discussion

Since translation in our system is on a similar timescale (1.0 amino acids/s) as folding for calerythrin measured in the stalled RNCs under force (misfolds in 1.5 s at 4.0 pN, although this would be faster at zero force), we expected the kinetics of folding during real-time elongation experiments to display a quasi-equilibrium behavior that reflects that of stalled, equilibrated RNCs. Using our measured parameters, we performed simulations of translation and folding behavior under equilibrium and non-equilibrium assumptions. We would expect the first misfolding transition to occur in 3.5 sec or less for 90% of ribosomes in equilibrium (Fig. 5d). In terms of codon position, we would expect >99% of RNCs to fold before codon 177 in real-time conditions (Fig. 5e). In this scenario, the RNC would not be able to avoid the misfolded state and the misfolded product would be released in solution (Fig. 6a). Yet, despite the relatively slow rate of translation in our experiments—rates that should even more strongly favor a quasi-equilibrium behavior because there is more time for folding at each subsequent amino acid added—we see that the nascent polypeptide folds only after a significant time delay. Given the known elongation rate of 12–18 amino acids/s in vivo at 37 degrees\textsuperscript{33} and a measured \( k_{cat} \) of 1.5 s\textsuperscript{-1} for peptide release\textsuperscript{43}, the measured delay of 63 s is long enough to avert misfolding for the remaining duration of elongation until the release of the nascent polypeptide. After release, the nascent chain would fold through the previously described off-the-ribosome pathway (Fig. 6b). Such a long delay could also allow time for chaperones to bind to the unfolded polypeptide and prevent the formation of the misfolded state on the ribosome.

Kinetic modeling suggests that not all proteins are equilibrated on a translational timescale, and the percent of proteins that are predicted to fold co-translationally can vary from 37% to 59% of the proteome when altering translation speed\textsuperscript{16}. However, experimental evidence that folding is not necessarily limited by the translation rate and folding takes place out of equilibrium has been indirect, for example, through decreases in protein activity after increasing the translation rate via synonymous codon mutations or changing tRNA levels\textsuperscript{3,4,44}. Other in vitro experiments that incorporate active translation employ ensemble methods.

Methods

Constructs and purification of proteins. The calerythrin gene was purchased (UniProtKB—P06495, IDT gene block), amplified, digested (Ndel, Nhel from NEB), and inserted into a plasmid with flanking Avi and ybbR tags, with a serine–glycine linker after the AviTag (TSGGGGSGGGGSAGGGGSGGGG).

Truncated constructs for released proteins were prepared via PCR and digestion with the same set of enzymes followed by cloning into the same backbone vector. Site-directed mutagenesis for calcium mutants and insertion of the valine stall site was conducted via whole-plasmid PCR amplification with mismatches in the primers, followed by DpnI and PPKK treatment and blunt-end ligation with T4 Ligase. Constructs were transformed into DH5\textalpha{} cells (Fisher Scientific, 18265017) and plasmid DNA was harvested via mini-prep. All constructs were verified by plasmid sequencing. The primers for all cloning steps are listed in Supplementary Table 4.

Wild-type calerythrin, calerythrin truncations, and calerythrin mutants were expressed in BL21 DE3 Escherichia coli cells (NEB, catalog number C2527H) with isopropyl β-D-1-thiogalactopyranoside induction (1 mM for 2 hrs at 37 °C), lysed, and biotinylated. Biotinylation of the lysate was conducted with 1 μM recombinant birA, 25 μM biotin, 5 μM ATP, and 5 mM Mg(ACaCl\textsubscript{2}) at room temperature for 40’. Excess biotin was removed by overnight dialysis in 1× PBS with 0.5 mM phenylmethylsulfonyl fluoride (PMSF). Biotinylated protein was purified on monomeric avidin resin using standard elution procedures (Pierce). Excess biotin was removed by overnight dialysis in 20 mM HEPES, 100 mM KCl, and 0.5 mM CaCl\textsubscript{2}. The protein sample was concentrated via Amicon centrifugal filters (Millipore, MWCO 10 kD or 3.5 kD) and stored at 80.

Purification of the tRNA synthetases was based on reported procedures\textsuperscript{45,46}. In brief, each tRNA synthetase was auto-induced in and extracted from BL21 strains by sonication. The enzymes were then captured on nickel sepharose columns (HisTrap\textsuperscript{®} HP, GE Healthcare) in binding buffer (30 mM HEPES-KOH, pH 7.5, 150 mM Na\textsubscript{2}SO\textsubscript{4}, 5 mM imidazole, and 100 mM KCl) and isolated from the column using a linear gradient (0-100%) of elution buffer (50 mM HEPES-KOH, pH 7.5, 100 mM KCl, 10 mM MgCl\textsubscript{2}, 1x imidazole and 7 mM β-mercaptoethanol). Fractions that contained the enzymes were pooled, dialyzed into storage buffer (30 mM HEPES, pH 7.5, 100 mM KCl, 10 mM MgCl\textsubscript{2}, 1x imidazole, and 7 mM β-mercaptoethanol), concentrated and supplemented with glycerol to 30% (v/v) before storage at –80 °C. Aspartyl-tRNA synthetase and glutamyl-tRNA synthetase were further purified by anion-exchange (HiTrap Q HP, GE Healthcare) prior to dialysis in storage buffer. The buffers for ion-exchange chromatography consisted of 50 mM HEPES-KOH, pH 7.5, 10 mM MgCl\textsubscript{2}, and 7 mM β-mercaptoethanol with either 50 mM (binding buffer) or 1x buffer (KCl). The gene for nucleotides diphosphate kinase (NDK) was amplified from bacteria strain MG1655 (The Coli Genetic Stock Center #6300, Yale) and cloned into pET His6 TEV LIC cloning vector (Plasmid #29653, Addgene) to introduce an N-terminal His-tag for affinity purification. Purification used the same binding and elution buffer for tRNA synthetases except that the starting final imidazole concentration in the binding buffer was increased to 40 mM. The purification of NDK ended with size exclusion chromatography (HiPrep\textsuperscript{TM} 16/60 Sephacryl\textsuperscript{®} S-100 HR, GE Healthcare) using
Real-time elongation OT samples. Samples for real-time elongation were stalled via omission of the amino-acid valine. Other amino acids that were not coded included cysteine, proline, tyrosine, and valine). 0.6 µM mRNA and 0.2 µM-labeled ribosomes. Our experiments made via a 524-Met incorporation timecourse (same conditions as above except 25 µM 3-mercaptopropionyl glycine instead of 300 µM and 2.5 µM 335-Met) indicate the stalled product is only stable for ~10 min before read-through if stalling at a single codon (Supplementary Fig. 5a). Presumably, after this time, the ribosome misincorporates a competing amino acid and continues translating until the end of the sequence. Thus, we used a triple codon repeat which prolongs this stalled state to >30 min (Supplementary Fig. 5b). We also extended the serine–glycine linker prior to the valine site to ensure that the AValTag was completely outside of the ribosomal exit tunnel and thus accessible for biotinylation and tethering. The inserted sequence was SGSGSGSGSGSGSGSGSVV.

To avoid prolonged translation reactions and sucrose cushioning, the reactions were incubated 20 min at 25 degrees. During this time, 2 µl of His-tag Dynabeads were prepared by washing 2× with 1× polymix (20 mM HEPES, 95 mM KCl, 5 mM MgCl2, 5 mM NH4Cl, 0.5 mM MgCl2, 50 mM GDP, and 2 mM 2-mercaptoethanol over 12 CV. Fracton were pooled and dialyzed overnight into 50 mM HEPES, 100 mM KCl, and 10 mM 2-mercaptoethanol. The purified protein was concentrated and supplemented with buffered glycerol to 30% final concentration (buffer composition unchanged), then stored at ~80. All proteins were checked for purity via SDS-PAGE.

Free protein and stalled RNC OT samples. The optical tweezzer sample consists of a protein or RNC tethered between two DNA handles. One handle is biotinylated, and through a streptavidin bridge it connects to the biotinylated substrate. The other handle is restriction digested to provide an overhang which is complementary to an oligo handle. Both handles have internal modifications to attach to antioxidantdigoxigenin-covalently attached polystyrene beads. The DNA handle ligated directly to the sample was generated via PCR of a 2 kb region of the lambda phage genome with a digoxigenin-modified reverse primer and an inserted Bsal site in the forward primer. The primers for the handle generation are listed in Supplementary Table 4. The PCR product was bsaI digested and BsaI generated via a run-off in vitro transcription (MEGAscript kit). The transcription template was generated via PCR and gel purifications then gel purified on a Qfi DNA-Sepharose 4 Fast flow. These were then added to in vitro translation reactions with purified Rs to maintain the equilibrium population of charged tRNAs. Although this step should not be necessary, it removes any lag time while waiting for initial charging to occur and places less stress on the regenerating system since synthetases are ATP dependent. Thus, we then act to recharging any deacylated tRNA, either due to activity of the ribosome or spontaneous deacylation. The required concentration of each synthetase was determined by measuring charging activity with [32P]-ATP47.

Data collection and processing. Data were collected on either of two optical tweezzer set-ups, both previously described48,49 and set to the same trap stiffness of 0.25 pN/nm. The traps are generated from 1064 nm lasers focused through high-numerical aperture objectives, allowing us to stably catch particles and measure their position and the force applied. Tethers were formed by catching a bead with sample deposited in one trap and a bead with the biotinylated handle in the other trap. By monitoring the bend angle of pulling apart the bead, the tether is formed. We used both contour length and curvature of the resulting tethers to ensure single molecules were analyzed.

Data were collected at 2500 Hz for equilibrium measurements and 1333 Hz for elongation measurements, which occur on a longer timescale. Data were used at concentrations of 0.001–0.008 M for all fitting experiments. This optimization and ligated to the DNA handles. The ligation to DNA handles was conducted in 1× T4 ligase buffer (NEB) with 6 U/µL T4 ligase at 16 degrees for 1 h. The ligated sample was stored at 80 for future experiments or further diluted two to tenfold.

All data were collected in passive mode to avoid artifacts of missed folding transitions and depressed apparent folding rates53. In real-time elongation
experiments, one individual amino acid is below our resolution (calculated step size for 1 amino acid is ~0.008 pN at an initial force 4.5 pN and no added amino acid) which is less than the measured step size for 224 amino acids at 3.5 pN (~0.003 pN). Empirically, our noise at 4 pN is 0.25 pN at 1333 Hz. Assuming a desired signal-to-noise of 2–2.5, this would be equivalent to 10 steps at 13.33 Hz. Thus, although we cannot say where steps occurred, we can conclude that the chain has elongated by 10 amino acids within a certain window.

When one amino acid is below our resolution, the exponential distribution as the null hypothesis as it is the simplest model developed for identifying transitions in force is assumed to be normal54. This method scores short events, as long-lived or frequent events are not significant. The reported translation rate is from the median value of the time to cross a 10-amino acid window, while the error is the standard deviation. For determining \( r \), the first point of the fit after the specified codon position (codon 167) is used as the time of arrival. The first folding transition is identified by one of two methods. First, the data are divided into 1 sec windows (where the baseline is approximately flat and Gaussian distributed) and downsampled by a factor of 10 in order to better differentiate folding from baseline. Owing to this downsampling, our limit of detection will be ~8 ms. Each point in the window is compared with a Gaussian CDF with the mean and standard deviation of the window to compute a probability of observing an event of this magnitude or larger. Points which cross a threshold of 1 × 10^{-5} were scored as folding events (corresponding to 99.99% CI) (Supplementary Fig. 6). This scoring method is similar to scoring that has been developed for identifying transitions in force–extension curves, which used a cutoff based on the standard deviation of point-wise difference, where the difference was assumed to be normal23. This method scores short events, as long-lived or frequent events are not significant. These events can be identified by eye in the window.

Then, the translation rate is related to the cumulative distribution function (CDF), the integral of the probability distribution function (PDF). The simplest model would correspond to an exponential fit to \( e^{-t/\tau} \). This is a single step stochastic process. We also tested a variety of other potential fits for added complexity: a double exponential fit where the chain has elongated by 10 amino acids, and an Erlang distribution where the chain has elongated by 2 of each codon or 10 amino acids. These fits were compared via an F test with the exponential distribution as the null hypothesis as it is the simplest model (Supplementary Fig. 7). The exponential fit has an \( R^2 = 0.91 \).

Using our fit value to \( \tau_{\text{day}} \), our translation rate, and our stalled folding rates, we performed a simulation of ribosome translation. We simulated trajectories of 10,000 ribosomes with and without a delay step prior to folding. Translation times at each codon were drawn from an exponential distribution of 1.0 aa/second. Folding times were drawn from an exponential distribution of \( k_{\text{fold}} = \tau_{\text{day}} \) (from our folding rate at 4.0 pN). As the folding rate is measured as constant between 0.1 and 1.0 sec, we maintained a constant folding rate, which was independent of translation. This assumption will likely not hold for the full linker region. The delay step was drawn from an exponential distribution for \( \tau_{\text{day}} = 63 \) sec. For a simulated CDF, we used the fraction of ribosomes, which had folded at or prior to a given time (or codon position). Translation was stopped at 222 codons, at which time all of the RNAs will fold eventually, leading to a sharp increase in the CDF after codon 221.

Statistical analysis. The number of molecules is indicated in the figure legends as \( n = m \) molecules, \( y \) rate measurements. Each rate measurement represents an independent fitting of many folding or unfolding events to an exponential cumulative distribution. Error on Bell plots is SE; 95% CIs for the fit parameters are shown in the supplementary tables. Force–extension or passive displays are representative examples of each construct.

Once the model was excluded from folding kinetics fitting of EF123 owing to increased baseline noise that skewed the unfolded lifetime owing to missing folded events to the N domain. The resulting HMM was non-exponential, although EF123 unfolding detection was unaffected and was included. For all protein and stalled simulation trajectories of ribosome translation, we performed a simulation of ribosome translation. We simulated trajectories of 10,000 ribosomes. The trap stiffness and bead offset were required in our analysis for calculations. No C domain folding can be owing to bead size variations. The trap stiffness and bead offset files owing to trap cross-talk were calibrated for each bead pair after each experiment.

For composite binned data as shown in the Bell plots, the rate data for all molecules for each condition was binned into equally spaced bins along the force axis (x-axis). The number of bins was determined by the square root of the number of rate measurements. Then, the linear fit to \( \ln(k_F) = \ln(k_0) - \Delta E/k_BT \) was performed with weighting on the binned data that was \( 1/n^2 \) of each bin or \( 1/n \) from the lifetime fits within the bin, whichever was a greater source of variation (usually the spread of the bin is greater, except in bins at very high or low force, which have more error in the fit and are also less populated). Although more exact models for fitting \( \ln(k) \) versus force exist that are non-linear24, they do not converge well to our data set likely because it only covers a small range of forces (−2 pN) such that we do not sufficiently sample the non-linear regions. To test if values from the linear fits (\( \ln(k_0) \) or \( \Delta E \)) are equivalent for different experimental conditions, we used a two-tailed Welch’s t test for two values with unequal variances. The variances were derived from the fit parameters in Matlab. The number of binned data points in the fit and the variances determined the degrees of freedom according to the Welch-Satterthwaite equation. The null hypothesis is that they are from the same distribution. A p value of <0.05 indicates they are significantly different values.

Data availability

Data supporting the findings of this manuscript are available from the corresponding author upon reasonable request. A reporting summary for this article is available as a Supplementary Information file. The source data underlying Figs 1c–e, 2c and e, 3a–c, 4a and b and 5a, b, d and e, Supplementary Figs 1, 2, 3, 5, 8 and 9 and Supplementary Tables 1 and 2 are provided as a Source Data file.

Received: 29 January 2019 Accepted: 7 May 2019
Published online: 20 June 2019

References

1. Crombie, T., Swoffield, J. C. & Brown, J. A. P. Protein folding within the cell is influenced by controlled rates of polypeptide elongation. J. Mol. Biol. 228, 7–12 (1992).
2. Komar, A. A., Lesnik, T. & Reiss, C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vivo translation. FEBS Lett. 462, 387–391 (1999).
3. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol. 16, 274–280 (2009).
4. Kim, S. J. et al. Translational tuning optimizes nascent protein folding in cells. Science 348, 444–448 (2015).
5. Cabrita, L. D. et al. A structural ensemble of a ribosome-nascent chain complex during in vivo co-translational protein folding. Nat. Struct. Mol. Biol. 23, 278–285 (2016).
6. Samelson, A. J., Jensen, M. K., Soto, R. A., Cate, J. H. D. & Marques, S. Quantitative determination of ribosome nascent chain stability. Proc. Natl. Acad. Sci. 113, 201610272 (2016).
7. Liu, K., Rehfus, J. E., Mattson, E. & Kaiser, C. M. The ribosome destabilizes native and non-native structures in a nascent multidomain protein. Protein Sci. 26, 1439–1451 (2017).
8. Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I. & Bustamante, C. The ribosome modules nascent protein folding. Science 334, 1723–1727 (2011).
9. Frydman, J., Erdjument-Bromage, H., Tempst, P. & Harrl, F. U. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. Nat. Struct. Biol. 6, 697–705 (1999).
10. Jacobson, G. N. & Clark, P. L. Quality over quantity: optimizing co-translational protein folding in the Escherichia coli cytosolic proteome. Proc. Natl. Acad. Sci. 105, 22810–22815 (2008).
11. Nilsson, O. B. et al. Cotranslational protein folding inside the ribosome exit tunnel. Cell Rep. 12, 1533–1540 (2015).
12. Marino, J., Heijne, Gvón & Beckmann, R. Small domains fold inside ribosome exit tunnel. FEBS Lett. 590, 655–660 (2016).
13. Mercier, E. & Rodnina, M. V. Co-translational folding trajectory of the HemK helical domain. Biochemistry 57, 3460–3464 (2018).
14. Goldman, D. H. et al. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. Science 348, 457–460 (2015).
15. Farias-Rico, J. A., Ruad Selim, F., Myronidi, I., Fruhau, M. & von Heijne, G. Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. Proc. Natl. Acad. Sci. 115, E9280–E9287 (2018).
16. Chyram, 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 cytotoxic proteome. PNAS 110, E132–E140 (2012).
