

Systematic mapping of free energy landscapes of a growing filamin domain during biosynthesis

Christopher A. Waudby^{a,b}, Tomasz Włodarski^{a,b,c}, Maria-Evangelia Karyadi^{a,b}, Anaïs M. E. Cassaignau^{a,b}, Sammy H. S. Chan^{a,b}, Anne S. Wentink^{a,b}, Julian M. Schmidt-Engler^{a,b}, Carlo Camilloni^d, Michele Vendruscolo^c, Lisa D. Cabrita^{a,b}, and John Christodoulou^{a,b,1}

^aInstitute of Structural and Molecular Biology, University College London, London WC1E 6BT, United Kingdom; ^bInstitute of Structural and Molecular Biology, Birkbeck College, London WC1E 7HX, United Kingdom; ^cDepartment of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom; and ^dDipartimento di Bioscienze, Università degli studi di Milano, 20133 Milan, Italy

Edited by Joseph D. Puglisi, Stanford University School of Medicine, Stanford, CA, and approved July 24, 2018 (received for review September 15, 2017)

Cotranslational folding (CTF) is a fundamental molecular process that ensures efficient protein biosynthesis and minimizes the formation of misfolded states. However, the complexity of this process makes it extremely challenging to obtain structural characterizations of CTF pathways. Here, we correlate observations of translationally arrested nascent chains with those of a systematic C-terminal truncation strategy. We create a detailed description of chain length-dependent free energy landscapes associated with folding of the FLN5 filamin domain, in isolation and on the ribosome, and thus, quantify a substantial destabilization of the native structure on the ribosome. We identify and characterize two folding intermediates formed in isolation, including a partially folded intermediate associated with the isomerization of a conserved cis proline residue. The slow folding associated with this process raises the prospect that neighboring unfolded domains might accumulate and misfold during biosynthesis. We develop a simple model to quantify the risk of misfolding in this situation and show that catalysis of folding by peptidyl-prolyl isomerases is sufficient to eliminate this hazard.

cotranslational folding | protein misfolding | nuclear magnetic resonance | molecular dynamics simulations | tandem repeat protein

The misfolding and aggregation of proteins are implicated in a wide range of human disorders (1). It is, therefore, essential for cells to have mechanisms to ensure quality control in protein synthesis and maintain protein homeostasis (2). Increasing evidence indicates that cotranslational folding (CTF) can increase the efficiency of folding compared with the folding of full-length (FL) proteins *in vitro* (3), in some cases modulating the folding process itself (4, 5). In this work, we focus on the folding of multidomain proteins, which account for 70% of eukaryotic proteins, many of which are tandem repeat proteins (TRPs; *i.e.*, are composed of multiple copies of similar domains) (6). Such sequences are at high risk of forming nonnative interdomain contacts during the folding process, potentially leading to misfolded states (7, 8), and therefore, a strong selective pressure exists to minimize the homology between adjacent domains (9).

A major obstacle to the folding of multidomain proteins is the exponential growth in the volume of conformational space that must be explored as the polypeptide chain length increases. CTF is widely believed to play a major role in the solution of this problem, as the sequential folding of emerging domains restricts the volume of conformational space, thus promoting rapid folding of the entire protein to its native state (while also minimizing the risk of interdomain misfolding). Therefore, in the context of the free energy landscape theory of protein folding, to study CTF, one should consider not a single free energy landscape but a series of nested landscapes of increasing dimensionality and polypeptide chain lengths (10). It is our goal here to map out these “cotranslational landscapes” to improve our understanding of the mechanisms by which the ribosome may contribute to efficient folding.

We have developed NMR spectroscopy as a tool for the observation of CTF in translationally arrested ribosome–nascent chain complexes (RNCs) (11, 12), in particular to investigate the emergence and folding of a pair of IgG-like domains, FLN5–6, from the TRP filamin (FLN) (Fig. 1 *A* and *B*) (5). Folding of FLN5 to the native state was observed, and a chemical shift-restrained ensemble structure was determined at a linker length (comprising the subsequent FLN6 domain) well beyond the point at which the entire domain was solvent exposed as probed by PEG mass-tagging assays.

In this work, using a C-terminal truncation system (13–15), we have conducted a systematic exploration of the thermodynamics and kinetics of folding in isolation to create a detailed description of the free energy landscape for CTF, from which the remarkably strong destabilizing effect of the ribosome on the native structure may be quantified. We also identify and determine the structure of an isolated intermediate associated with the isomerization of a conserved native-state cis proline, and we use this model to explore and eliminate the possible role of steric occlusion in perturbing the folding process on the ribosome. Lastly, we apply these data to a simple model to analyze the effect of folding kinetics and proline isomerization on the risk of misfolding in this multidomain protein.

Significance

Efficient synthesis and folding of proteins, avoiding misfolded states, are central to cell function. As folding may be initiated in parallel with translation, key experimental challenges are to map changes that occur in folding free energy landscapes as translation proceeds and to understand how these landscapes might be modulated by the ribosome and auxiliary factors. Here, we study the length-dependent folding of a domain from a tandem repeat protein and solve the structure of a stable folding intermediate. Although destabilized by the ribosome at equilibrium, modeling of the nonequilibrium folding pathway nevertheless indicates a significant role for proline isomerization during translation. We develop a simple model to explore the impact of cotranslational folding kinetics on misfolding hazards.

Author contributions: C.A.W., T.W., M.-E.K., L.D.C., and J.C. designed research; C.A.W., T.W., M.-E.K., A.M.E.C., S.H.S.C., A.S.W., J.M.S.-E., and C.C. performed research; C.A.W., T.W., M.-E.K., A.M.E.C., S.H.S.C., A.S.W., J.M.S.-E., C.C., M.V., L.D.C., and J.C. analyzed data; and C.A.W., T.W., M.-E.K., M.V., L.D.C., and J.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence should be addressed. Email: j.christodoulou@ucl.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716252115/-DCSupplemental.

Published online September 10, 2018.

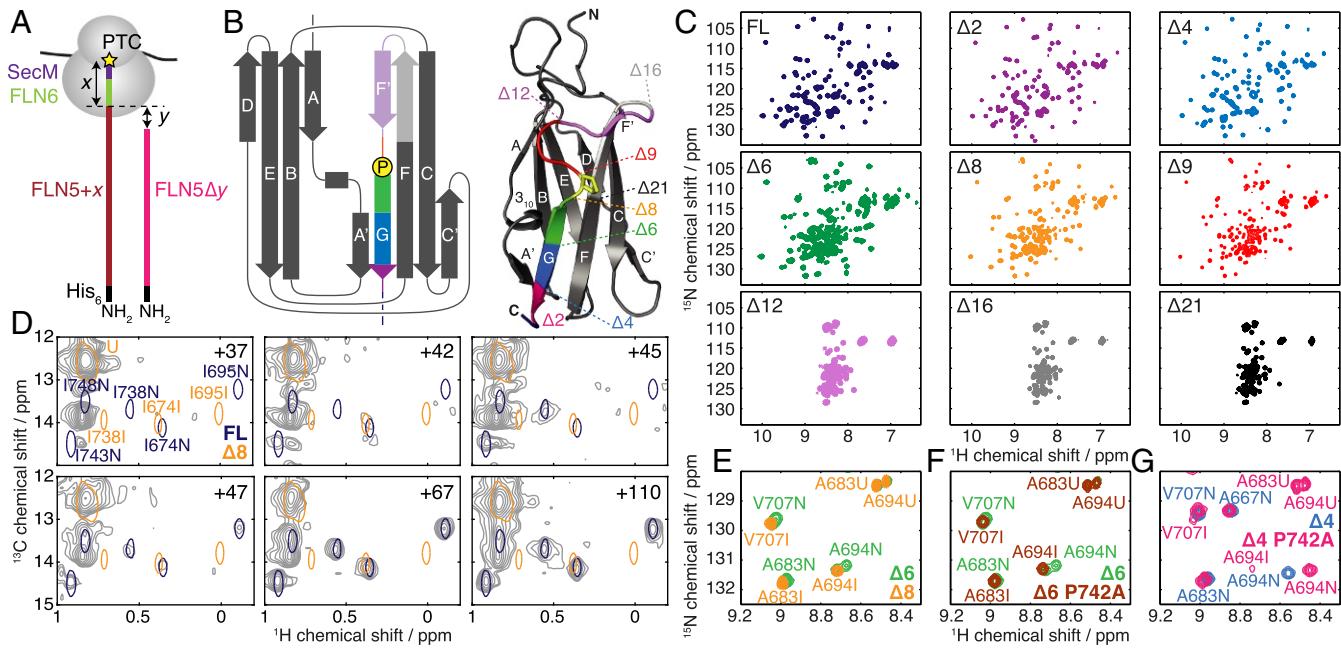


Fig. 1. (A) Schematic and nomenclature of FLN5 RNC (FLN5 + x) and truncation (FLN5Δy) constructs; x and y represent the FLN5 + secM linker length (measured from the PTC), and the extent of truncation, respectively. (B) Topology and crystal structure of FLN5 (1qfh) colored and labeled to indicate the truncated constructs used in this study. The cis proline P742 is highlighted in yellow (stick representation). (C) ^1H , ^{15}N -selective optimized flip-angle short transient heteronuclear multiple quantum coherence (SOFAST-HMQC) spectra of FLN5 truncation variants (298 K, 600 MHz). (D) Comparison of ^1H , ^{13}C -HMBC spectra of FL and $\Delta 8$ with spectra of [$^{13}\text{CH}_3$ -Ile, ^2H]-labeled FLN5 + L RNCs (5) (gray; 298 K, 700 MHz). (E) Comparison of ^1H , ^{15}N -HSQC spectra of FLN5 $\Delta 6$ and FLN5 $\Delta 8$, showing a magnified view of representative residues (283 K, 700 MHz). FLN5 $\Delta 6$ resonance assignments are not shown where overlapping with FLN5 $\Delta 8$ assignments. (F) Comparison of ^1H , ^{15}N -HSQC spectra of FLN5 $\Delta 6$ and FLN5 $\Delta 6$ (P742A) (283 K, 700 MHz). (G) Comparison of ^1H , ^{15}N -HSQC spectra of FLN5 $\Delta 4$ and FLN5 $\Delta 4$ (P742A) (283 K, 700 MHz).

Results

Variants of the FL FLN5 domain were designed to mimic CTF by truncating between 2 and 21 residues from the C terminus, denoted $\Delta 2$ to $\Delta 21$ (Fig. 1 A and B). Native PAGE showed $\Delta 12$ and $\Delta 16$ to have expanded conformations relative to FL, while the migration of $\Delta 6$ to $\Delta 9$ was approximately midway between these states (SI Appendix, Fig. S14). A transition from β -sheet to random coil structure was also observed at these lengths by far-UV CD spectroscopy (SI Appendix, Fig. S1 B and C).

To investigate the length dependence of folding further, 2D NMR correlation spectra of truncation variants were acquired (Fig. 1C and SI Appendix, Fig. S1D). The large ^1H chemical shift dispersion observed in FL $\Delta 2$ and $\Delta 4$ spectra indicates the presence of stable tertiary structure, while the narrow dispersion observed for $\Delta 12$, $\Delta 16$, and $\Delta 21$ is characteristic of a disordered state. At intermediate lengths ($\Delta 6$, $\Delta 8$, and $\Delta 9$), multiple sets of cross-peaks were observed, indicating that both structured and unstructured states are populated in equilibrium. Two states, designated “I” (intermediate) and “U” (unfolded) were populated by $\Delta 8$ and $\Delta 9$, while an additional set of folded resonances [“N” (native like)] was observed in the $\Delta 6$ spectrum (Fig. 1E and SI Appendix, Fig. S2).

We hypothesized that the two sets of folded $\Delta 6$ resonances arise from cis/trans isomerization of the native-state cis proline P742 (Fig. 1B). To test this conjecture, we examined the $\Delta 6$ P742A variant, in which P742 was mutated to favor the nonnative trans conformation. The additional resonances observed in the $\Delta 6$ spectrum were absent in spectra of $\Delta 6$ P742A (Fig. 1F), and the remaining resonances corresponded closely with those observed for $\Delta 9$, in which P742 is absent altogether. We, therefore, conclude that there is a length-dependent transition from an intermediate I, in which P742, if present, is in a trans conformation, to a (near)-native state N, in which P742 is in a cis

conformation. Consistent with this conclusion, in the $\Delta 4$ P742 A variant (Fig. 1G), the N state was destabilized, and both I and U resonances could additionally be observed with the same chemical shifts as observed in WT $\Delta 6$.

Structural Characterization of Folding Pathway. Having identified a series of states populated within isolated FLN5 fragments, we sought to analyze their structures to understand and connect their folding behavior in isolation to that occurring cotranslationally within RNCs. Progressive changes in amide chemical shifts were observed from the FL native state to the $\Delta 6$ I state clustered around the A and F strands adjacent to the truncated G strand (SI Appendix, Fig. S3). The magnitude of chemical shift perturbations increased with the extent of truncation up to the $\Delta 6$ I state, but no large chemical shift differences were observed in shorter lengths, indicating that the intermediate structure does not change significantly beyond this point.

We then used chemical shift-restrained replica-averaged metadynamics simulations (16) to determine an ensemble structure of the $\Delta 6$ intermediate (together, as a control, with the FL native state) (Fig. 2 A and B and SI Appendix, Fig. S4). The quality of the ensembles was comparable with those previously determined by chemical shift methods (17) assessed by back-calculated chemical shift deviations, residual dipolar coupling Q factors (SI Appendix, Table S1), and a backbone rmsd of 1.8 Å between the FL ensemble and a crystal structure (18) (Fig. 2A). We found that the FL ensemble is generally well ordered, with $\text{C}\alpha$ root-mean-square fluctuation (RMSF) values below 0.5 Å, except in some loop regions (Fig. 2A). In contrast, in the $\Delta 6$ intermediate, a native-like core is retained, comprising the A–F strands, but the truncated G strand is unfolded, with increased disorder also found in the adjacent A' and F strands (Fig. 2B). No structure could be determined for the $\Delta 6$ N state due to broadening of resonances around the site of the truncation.

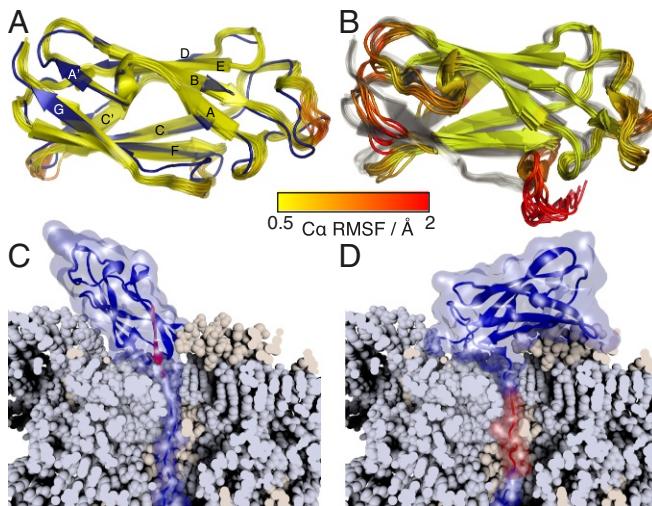


Fig. 2. Structural analysis of FLN5 truncation variants. (A) Ensemble structure of FL FLN5 (yellow to red) aligned against the previously determined crystal structure (blue) (18). The FL ensemble is colored according to the C_α RMSF as indicated by the key. (B) Ensemble structure of the $\Delta 6$ trans intermediate (colored by the C_α RMSF) aligned against the FL ensemble (gray). (C and D) Modeling of the closest possible approach of (C) native and (D) intermediate FLN5 structures tethered to the ribosome (shown in a cutaway view to highlight the NC path through the exit tunnel). The G strand of FLN5 (disordered in the intermediate) is highlighted in red.

relaxation dispersion measurements (19) of adjacent residues indicate that this is due to conformational exchange on a millisecond timescale (SI Appendix, Fig. S5), which we suggest may reflect transient association of the truncated G strand with the F strand that becomes stabilized as the chain length increases.

In these relaxation measurements, small dispersions ($R_{ex} < 4 \text{ s}^{-1}$) were also observed in the unfolded state of $\Delta 6$ (SI Appendix, Fig. S6A). These data could be fitted to two-state exchange with a high-energy intermediate, populated to $0.95 \pm 0.05\%$, and designated “I*” (SI Appendix, Fig. S6B). Chemical shift differences were uniformly small ($< 0.5 \text{ ppm}$) for residues 646–670, but from residue 671 onward, they were correlated with those between $\Delta 6$ U and I states, indicating the formation of intermediate-like structure in this region. Thus, the I* intermediate comprises a native-like core formed from the B to F strands, while the A strand and the truncated G strand remain disordered (SI Appendix, Fig. S6C). The intermediate was also identified in $\Delta 9$, $\Delta 12$, and $\Delta 16$ variants, with a decreased population, $0.4 \pm 0.1\%$, in the shortest length, $\Delta 6$ (SI Appendix, Fig. S6D–J).

In summary, therefore, the length-dependent folding pathway proceeds from the unfolded state via a high-energy intermediate with disordered A and G strands to a stable intermediate with a disordered G strand and P742, if present, in a trans conformation. Isomerization of P742 to cis, in either the intermediate or via the unfolded state, then allows association of the G strand to form the final native state.

If the structure of the intermediate is now considered in the context of CTF, the disordered nature of the C-terminal G strand suggests that this state should be accessible early in the translation process. To explore this, we ran simulations using structure-based models (20) with various linker lengths to study folding of the N and I states on the ribosome. Using this approach, we found that the native state requires a 20-aa linker (in an extended conformation) to fold on the ribosome, whereas only a 14-aa linker was needed for the intermediate (Fig. 2C and D). By comparison with the linker conformation in an all-atom molecular dynamics (MD) simulation of the FLN5 +

110 RNC (5), we further estimated that these native and intermediate structures should be accessible with linker lengths (in a relaxed conformation) of ca. 26 and 18 aa, respectively (SI Appendix, Fig. S7).

The structured states of FLN5 RNCs can be observed with the greatest sensitivity using a specific isoleucine C δ methyl labeling strategy (Fig. 1D) (5). However, no resonances corresponding to the intermediate state were detected in RNC spectra of any linker length (although as it exists in isolation as a minor state in equilibrium with the unfolded state, it is not possible to exclude a small undetected population). Similarly, native resonances were not observed in these data (5) until a linker length of 42 aa (Fig. 1D), well beyond the 26 aa expected from our structural modeling. This indicates that there must be a specific inhibition of folding that arises from ribosomal tethering. To understand this further, we, therefore, sought to complement our structural description of the folding pathway during elongation with a detailed energetic description of the CTF landscape.

Thermodynamic Analysis. A range of experimental techniques have been applied to create a detailed thermodynamic map of the length-dependent folding process. The stabilities of FL, $\Delta 2$, and $\Delta 4$ variants were measured by chemical denaturation (SI Appendix, Fig. S8), while in $\Delta 6$, $\Delta 8$ and $\Delta 9$, where multiple states were observed to be populated at equilibrium, NMR measurements of peak volumes in $^1\text{H}, ^{15}\text{N}$ -heteronuclear single quantum correlation (HSQC) spectra were used to determine the relative populations of folded and unfolded states and hence, calculate free energies of folding directly (Fig. 3 and SI Appendix, Fig. S9). Where only the unfolded state could be observed ($\Delta 12$ and $\Delta 16$), limits on the stability of the intermediate have been calculated based on a population below the ca. 1% limit of detection derived from signal-to-noise levels (dashed lines in Fig. 3). From these measurements, we found that the six C-terminal residues contribute almost the entire stability of the domain, $\Delta G_{N-D} = -7.0 \pm 0.2 \text{ kcal mol}^{-1}$. The stability of the I state did not vary significantly between $\Delta 6$ and $\Delta 9$ ($\Delta G_{I-U} = 0.8 \pm 0.2 \text{ kcal mol}^{-1}$), but in shorter lengths, the intermediate was further destabilized, with a population below the limit of detection.

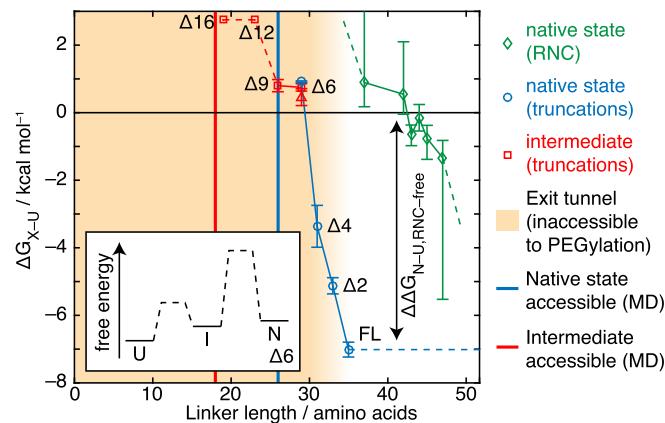


Fig. 3. Thermodynamic characterization of the FLN5 CTF pathway. Free energies of folding to native (cis-P742) and intermediate (trans-P742) states at 298 K are shown in truncated variants and RNCs as indicated. As a specific example, the $\Delta 6$ folding pathway is illustrated in *Inset*. A red triangle indicates $\Delta 6$ intermediate stability as determined by magnetization transfer measurements (Fig. 4 A and B). Yellow shading indicates the extent of the ribosome exit tunnel determined by PEG accessibility (5) and used as the basis for alignment of the isolated and ribosome-associated folding pathways, while solid lines indicate the points at which native and intermediate states were found to be accessible in MD simulations (Fig. 2 C and D).

In a similar manner to the truncated variants, we then used decreases in unfolded-state peak volumes (5) to determine folding free energies in RNCs (Fig. 3), observing a folding midpoint at a 42- to 44-aa linker length, beyond which nascent chain (NC) stability rapidly increases. The alignment of chain lengths shown here is based on the accessibility to PEG observed for RNC residues 36 aa beyond the peptidyl-transferase center (PTC) (5), and therefore, we assigned the FL isolated state to this length. This is a conservative alignment compared with our structure-based modeling (Fig. 2 C and D), which predicts that linker lengths of ca. 18 and 26 aa are required for folding to the intermediate and native states, respectively (Fig. 3, solid lines). In either case, it is clear that the “delay” in NC folding after emergence from the exit tunnel must at the midpoint be associated with a substantial destabilization of the NC, with a $\Delta\Delta G$ comparable with the 7-kcal mol⁻¹ folding free energy of the isolated domain. In contrast, while no intermediate NC resonances have been detected (Fig. 1D), because the intermediate is only marginally stable in isolation, the extent to which it may be destabilized in RNCs is less clear. Nevertheless, its identification in isolation prompted us to consider more carefully the potential consequences of slow proline isomerization, whether occurring via the intermediate or directly from the unfolded state, on the CTF process.

Measurement of Folding Kinetics. CTF is a nonequilibrium process, with folding occurring in parallel with elongation of the NC, and therefore, it is important to understand the rate of NC folding in addition to the thermodynamic characterization above. As direct measurements of NC folding kinetics are experimentally inaccessible, we have used the isolated $\Delta 6$ truncation as a proxy. We first used 2D N_z -exchange measurements (21) to observe reversible folding between $\Delta 6$ U and I states (Fig. 4A), which was determined to occur on a timescale of 0.4 s at 298 K (Fig. 4B). We then carried out real-time NMR temperature jump measurements to measure the proline isomerization limited rate of refolding from a thermally unfolded state at 310 K (SI Appendix, Fig. S10A). At 298 K, trans-to-cis isomerization of P742 was observed with a time constant of 1.5 min (Fig. 4C), while at lower temperatures, an additional faster phase was resolved that we attribute to off-pathway cis-to-trans isomerization of seven native-state trans prolines in the unfolded state (SI Appendix, Fig. S10B). Consistent with this interpretation, trans-to-cis isomerization was not observed in measurements of $\Delta 6$ P742A refolding, while the refolding of $\Delta 4$ P742A at 283 K fitted to a single slow phase with a time constant of 17.3 min, indicating formation of a cis alanine (SI Appendix, Fig. S10 C and D). Activation energies of 21.7 ± 1.0 and 20.0 ± 0.7 kcal mol⁻¹ were determined for the fast and slow phases of $\Delta 6$ refolding (Fig. 4D), respectively, comparable with typical activation energies for proline isomerization (20–22 kcal mol⁻¹) (22).

Lastly, we investigated the effect of the peptidyl-prolyl isomerase (PPIase) cyclophilin A. On addition of this enzyme, folding of $\Delta 6$ was no longer observed in real-time NMR experiments, but instead, new exchange cross-peaks were observed between *N* and *U* states in N_z -exchange measurements (SI Appendix, Fig. S11). We, therefore, conclude that CypA can catalyze proline isomerization and folding to a timescale comparable with the nitrogen T_1 (ca. 1 s) and within the dead time of the real-time experiment. Exchange was not observed between *I* and *N* states, indicating that unfolding may be necessary for P742 to be sterically accessible for catalysis.

Reconstructing the Nonequilibrium CTF Pathway. In this work, we have analyzed the thermodynamics and kinetics of FLN5 folding as a function of polypeptide chain length and identified states that are structurally compatible with tethering to the

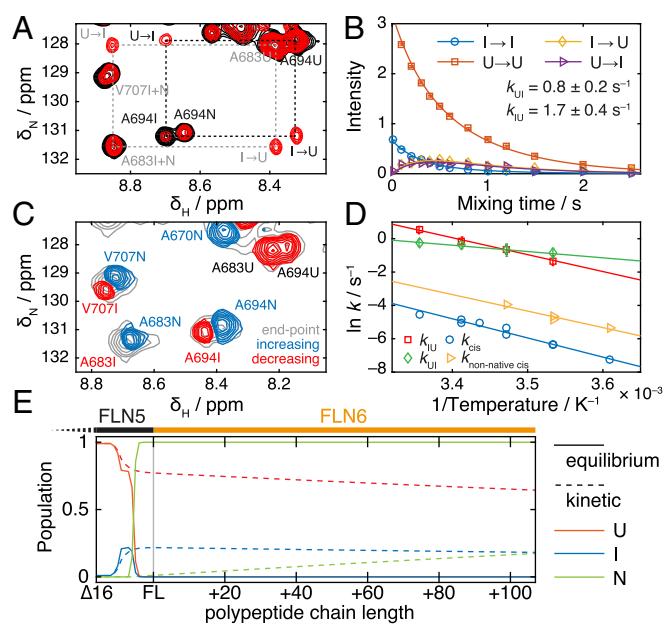


Fig. 4. Analysis of folding kinetics. (A) 2D slices from an N_z magnetization transfer experiment (21) at 298 K, with exchange times of 10 ms (black) and 500 ms (red). Dashed rectangles highlight exchange cross-peaks between unfolded and intermediate states. Gray coloring indicates peaks excluded from additional analysis due to overlap of native and intermediate resonances. (B) Analysis and fitting of A694 cross-peak intensities from the magnetization transfer experiment shown in A. Data from multiple resonances were fitted globally to determine the indicated exchange rates. (C) Component spectrum determined from the analysis of real-time NMR data for the refolding of $\Delta 6$ after a temperature jump from 310 to 298 K, with a fitted time constant of 1.5 min. Peaks that increase or decrease over the reaction time course are shown in blue and red, respectively. The spectrum at equilibrium is shown in gray. (D) Arrhenius plot showing the temperature dependence of $\Delta 6$ folding kinetics. $k_{U \rightarrow I}$ and $k_{I \rightarrow I}$ are exchange rates determined by magnetization transfer measurements; $k_{\text{non-native-cis}}$ and k_{cis} are proline isomerization rates determined by real-time NMR methods. (E) Populations of native, unfolded, and intermediate states based on characterization of isolated truncations plotted against polypeptide chain length at equilibrium (i.e., infinitely slow/stalled translation; solid lines) and under nonequilibrium conditions with a translation rate of 5 amino acids s^{-1} (dashed lines).

ribosome during translation (Figs. 2–4). This description can be combined with a known translation rate (ca. 5 amino acids s^{-1} in eukaryotes) to construct a Markov model of the CTF pathway (SI Appendix, Fig. S12), which we have integrated numerically to obtain quantitative predictions about the fundamentally nonequilibrium CTF process (23). From this model, we observe significant differences between populations calculated at thermodynamic equilibrium and those predicted under nonequilibrium conditions (Fig. 4E). In particular, as translation is more rapid than *trans*cis isomerization (i.e., folding to the native state), only 17% of FLN5 NCs are predicted (based on our analysis of isolated truncations) to be fully folded by the point at which the subsequent FLN6 domain has been translated. Such an accumulation of adjacent unfolded domains seems likely to present a misfolding hazard (7, 8, 24), and this is explored further in the next section.

Assessing the Risk of Cotranslational Misfolding in TRPs. Filamins are TRPs comprising up to 24 IgG domains, with a mean sequence identity between adjacent domains of $31 \pm 7\%$ (SD) and 9% of pairs above the 40% threshold associated with misfolding in other IgG domains (SI Appendix, Table S2) (9). Given that P742 is conserved in a cis conformation in every known

structure (*SI Appendix*, Table S2), the cotranslational accumulation of unfolded domains predicted above may indicate a more widespread risk of interdomain misfolding in filamins. More generally, a systematic analysis of Trp– structures [RepeatsDB (25)] shows that the abundance of cis prolines is comparable with that within the wider proteome [CATH S35 sets (26)] (*SI Appendix*, Fig. S14). Thus, proline isomerization limited CTF, and the potential accumulation of unfolded domains is likely to be an issue of relevance beyond filamins alone.

To explore the misfolding of filamin domains in more detail, we performed metadynamics simulations using structure-based models (7) to determine free energy landscapes for the folding and misfolding of pairs of filamin and titin IgG domains (Fig. 5A and *SI Appendix*, Fig. S13). From these calculations, we found that the population of domain-swapped misfolded states formed between filamin domains was 2.4% (Fig. 5A), comparable with that of the titin I27 dimer (0.7%) (*SI Appendix*, Fig. S13) for which misfolded states have been observed experimentally (7, 8).

To investigate this further, we developed a simple model to quantify the risk of misfolding during biosynthesis (Fig. 5B). By analyzing the timescale of CTF, $\tau_{\text{fold}}^{(i)}$, relative to translation of the following domain, $\tau_{\text{trans}}^{(i+1)}$, we found (*SI Appendix*, *SI Materials and Methods*) that, for a protein with N tandem repeat domains, the misfolding risk, R , given by the probability of populating adjacent unfolded domains during biosynthesis is $R = 1 - \prod_{i=1}^{N-1} \left[1 - \exp \left(-\tau_{\text{trans}}^{(i+1)} / \tau_{\text{fold}}^{(i)} \right) \right]$. This risk is plotted in Fig. 5C for increasing lengths of the filamin tandem repeat region (given a translation rate of 5 amino acids s^{-1}). In isolation, when folding is limited by proline isomerization (Fig. 5C, i), the misfolding risk is high, particularly

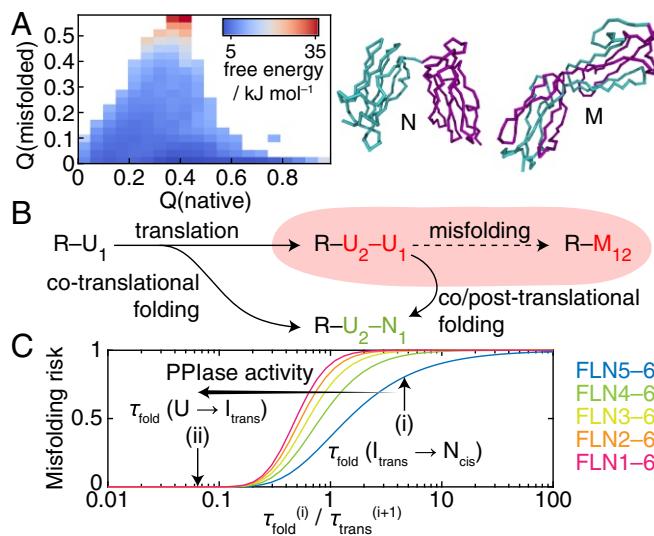


Fig. 5. Analysis of filamin misfolding during protein synthesis. (A) Free energy landscape calculated for an FLN4-5 dimer (cyan to purple) by coarse-grained MD simulations plotted against the fraction of native and misfolded contacts. Representative structures of native and misfolded states are shown as labeled. (B) A model for assessing the risk of cotranslational misfolding in which the rate of folding of a ribosome-bound unfolded NC ($R-U_1$) to its native state (N_1) relative to the rate at which the following domain (U_2) is translated determines the likelihood of populating adjacent unfolded domains ($R-U_1-U_2$), which may be at risk for forming a misfolded state (M_{12}). (C) Estimated risk of misfolding in filamin as a function of the number of tandem repeat domains and the relative rates of folding and translation. Arrows indicate predicted risks where folding is (i) limited by proline isomerization to the native state or (ii) through the action of PPIases, limited by folding to the intermediate.

when the full multidomain protein is considered. However, if as shown this step is accelerated by PPIases (*SI Appendix*, Fig. S11), folding times will instead become limited by the rate of folding from the unfolded state (Fig. 5C, ii). Under this condition, we find that the risk of misfolding can be essentially eliminated.

Modulation of CTF by a Molecular Chaperone. The interaction of molecular chaperones with NCs has previously been identified to be a key mechanism for suppressing long-range misfolding interactions (27) and may provide an additional defense against the misfolding hazard discussed above. Here, we have, therefore, investigated interactions of both $\Delta 6$ and a disordered FLN5 + 21 RNC with the ribosome-associated holdase trigger factor (TF). Due to its marginal stability (Fig. 3), the intensities of N , I , and U resonances in $\Delta 6$ spectra are highly sensitive to intermolecular interactions. However, no changes in these intensities were observed in the presence of TF (*SI Appendix*, Fig. S15A). Likewise, only a small and uniform reduction in ^1H , ^{15}N resonance intensities was observed in an FLN5 + 21 RNC on binding of TF (*SI Appendix*, Fig. S15 B–E), comparable with that previously observed for an intrinsically disordered and weakly interacting α -synuclein RNC (12). The observed absence of interactions for this folding-competent domain—in contrast to those between FLN5 and the ribosome surface (5)—indicates that only in some cases can individual chaperones provide protection against misfolding hazards.

Discussion

During protein synthesis on the ribosome, free energy landscapes for NC folding evolve as the chain length is increased by translation (10). Understanding CTF, therefore, requires characterizing a series of free energy landscapes rather than the single surface associated with traditional folding studies. Here, we have begun this process using observations of translationally arrested RNCs under equilibrium conditions complemented by a systematic C-terminal truncation strategy (13–15).

The truncation model developed in this work has two important roles. First, it provides an essential baseline against which the effects of ribosome interactions can be discerned. It is increasingly evident that the ribosome surface can be an active participant in CTF processes (4, 5), and this strategy has highlighted the large free energy changes ($\Delta\Delta G \sim 7$ kcal mol^{-1}) associated with the delay of FLN5 folding (Fig. 3). This is a substantially greater perturbation than observed using pulse proteolysis measurements in other RNC systems (28), and a key challenge for future work will be to develop a quantitative and structural understanding of these effects.

Second, our truncation model enabled us to apply high-resolution methods not currently feasible directly on the ribosome to analyze structure and folding kinetics. In doing so, we have identified a marginally stable intermediate associated with the isomerization of a native-state cis proline, and we have shown that this isolated structure can be used as a starting configuration for structural modeling of nascent chains (Fig. 2). From these results, we concluded that there are no steric barriers to formation of the intermediate early in the translation process, although its final stability on tethering to the ribosome remains an open question. Nevertheless, the slow trans-to-cis isomerization identified in isolation must still occur at some point during filamin biosynthesis, and a consideration of this led to the development of a simple model to assess the cotranslational misfolding hazard (Fig. 5).

It is still unclear whether ribosome–NC interactions, which necessarily act only over a short range after emergence from the exit tunnel, in anyway assist the folding of filamin. However, other situations can be envisaged where such interactions are more beneficial (e.g., sequestering aggregation or

misfolding-prone sequences before recognition and engagement by chaperones or entry into secretion pathways). Likewise, while we have shown here that the ribosome surface and TF have at least partly different recognition motifs and therefore, client pools, it is possible that other molecular chaperones might effectively suppress long-range interactions and contribute to alleviating the misfolding hazard, even in the absence of PPIase activity.

It is interesting to compare the intermediate identified here with the aggregation-prone intermediate formed during β_2 -microglobulin folding (29). Both domains have IgG folds, and both intermediates are associated with isomerization of native cis proline residues; however, a key difference is that FLN5 is part of a multidomain protein, and therefore, the primary hazard is the intramolecular misfolding of adjacent unfolded states during translation rather than posttranslational intermolecular aggregation. We have found that cis prolines are as abundant within TRPs as within the broader proteome (SI Appendix, Fig. S14), and therefore, this hazard is clearly relevant to more than filamin proteins alone. A simple model (Fig. 5 B and C) has indicated that PPIase activity is necessary but also sufficient to eliminate this risk in filamins (as indeed could be other forms of chaperone intervention), and we expect that this model could be applicable to other cases and forms of cotranslational misfolding. Overall, our findings highlight that thermodynamically favorable CTF is not by itself sufficient to avoid misfolding hazards: kinetic requirements must also be satisfied.

- Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
- Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* 82:323–355.
- Cabrita LD, Dobson CM, Christodoulou J (2010) Protein folding on the ribosome. *Curr Opin Struct Biol* 20:33–45.
- Kaiser CM, Goldman DH, Chodera JD, Tinoco I, Bustamante C (2011) The ribosome modulates nascent protein folding. *Science* 334:1723–1727.
- Cabrita LD, et al. (2016) A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol* 23:278–285.
- Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J (2007) The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* 8:319–330.
- Borgia MB, et al. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474:662–665.
- Borgia A, et al. (2015) Transient misfolding dominates multidomain protein folding. *Nat Commun* 6:8861.
- Wright CF, Teichmann SA, Clarke J, Dobson CM (2005) The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* 438:878–881.
- Clark PL (2004) Protein folding in the cell: Reshaping the folding funnel. *Trends Biochem Sci* 29:527–534.
- Cassaignau AME, et al. (2016) A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nat Protoc* 11:1492–1507.
- Deckert A, et al. (2016) Structural characterization of the interaction of α -synuclein nascent chains with the ribosomal surface and trigger factor. *Proc Natl Acad Sci USA* 113:5012–5017.
- Shortle D, Meeker AK (1989) Residual structure in large fragments of staphylococcal nucleic: Effects of amino acid substitutions. *Biochemistry* 28:936–944.
- De Prat Gay G, et al. (1995) Conformational pathway of the polypeptide chain of chymotrypsin inhibitor-2 growing from its N terminus in vitro. Parallels with the protein folding pathway. *J Mol Biol* 254:968–979.
- Neira JL, Fersht AR (1999) Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments. *J Mol Biol* 285:1309–1333.
- Camilloni C, Vendruscolo M (2014) Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J Am Chem Soc* 136:8982–8991.
- Rosato A, et al. (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. *Structure* 20:227–236.
- McCoy AJ, Fucini P, Noegel AA, Stewart M (1999) Structural basis for dimerization of the Dictyostelium gelation factor (ABP120) rod. *Nat Struct Biol* 6:836–841.
- Hansen DF, Vallurupalli P, Kay LE (2008) An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J Phys Chem B* 112:5898–5904.
- Noel JK, et al. (2016) SMOG 2: A Versatile software package for generating structure-based models. *PLoS Comput Biol* 12:e1004794.
- Farrow NA, Zhang O, Forman-Kay JD, Kay LE (1994) A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J Biomol NMR* 4:727–734.
- Balbach J, Schmid FX (2000) Proline isomerization and its catalysis in protein folding. *Mechanisms of Protein Folding*, ed Pain RH (Oxford Univ Press, Oxford), pp 212–249.
- O'Brien EP, Vendruscolo M, Dobson CM (2012) Prediction of variable translation rate effects on cotranslational protein folding. *Nat Commun* 3:868.
- Zheng W, Schafer NP, Wolynes PG (2013) Frustration in the energy landscapes of multidomain protein misfolding. *Proc Natl Acad Sci USA* 110:1680–1685.
- Paladin L, et al. (2017) RepeatsDB 2.0: Improved annotation, classification, search and visualization of repeat protein structures. *Nucleic Acids Res* 45:D308–D312.
- Dawson NL, et al. (2017) Cath: An expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res* 45:D289–D295.
- Mashaghi A, et al. (2013) Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* 500:98–101.
- Samelson AJ, Jensen MK, Soto RA, Cate JHD, Marqusee S (2016) Quantitative determination of ribosome nascent chain stability. *Proc Natl Acad Sci USA* 113:13402–13407.
- Rennella E, et al. (2012) Real-time NMR characterization of structure and dynamics in a transiently populated protein folding intermediate. *J Am Chem Soc* 134:8066–8069.

Materials and Methods

FLN5 truncations, mutational variants, and TF were cloned, expressed, and purified in Tico buffer as described previously (5, 12). RNCs were prepared, and their stability was monitored during acquisition as described previously (5, 11). CD spectra were acquired at protein concentrations of ca. 0.5 mg mL⁻¹ using a Jasco J-810 spectropolarimeter. To determine protein stability, CD signals were measured at 234 nm for urea concentrations between 0 and 8.2 M and fitted to a two-state unfolding process. NMR data were acquired at 283 K on a 700-MHz Bruker Avance III spectrometer with TXI cryoprobe unless otherwise indicated. Spectra were assigned with standard triple-resonance methods, and magnetization transfer, real-time kinetics, and Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion data were acquired as described in SI Appendix, SI Materials and Methods. FL and Δ 6 FLN5 structures were determined by chemical shift-restrained replica-averaged metadynamics (16) using Gromacs as described in SI Appendix, SI Materials and Methods. Simulations of folding on the ribosome and inter-domain misfolding were conducted in Gromacs with structure-based models generated in SMOG 2 (20) and misfolding potentials derived following the approach of refs. 7 and 8 as described in SI Appendix, SI Materials and Methods.

ACKNOWLEDGMENTS. We thank Dr. Flemming Hansen for helpful discussions, Dr. John Kirkpatrick and Dr. Tom Frenkel for assistance with NMR experiments, and Dr. Leo James for a gift of CypA. We acknowledge the use of Legion@UCL and the University College London NMR Center, the Medical Research Council for access to the Biomedical NMR Center at the Francis Crick Institute, and the staff for their support. This work was supported by a Wellcome Trust Investigator Award (to J.C.). T.W. was supported by an European Molecular Biology Organization fellowship, and M.-E.K. and S.H.S.C. were supported by Biotechnology and Biological Sciences Research Council studentships.