The Folding Pathway of an Ig Domain is Conserved On and Off the Ribosome

Pengfei Tian[1‡], Annette Steward[2‡], Renuka Kudva[3‡], Ting Su[5‡], Patrick J. Shilling[3], Adrian A. Nickson[2], Jeffrey J. Hollins[2], Roland Beckmann[5], Gunnar von Heijne[3,4#], Robert B. Best[1#] and Jane Clarke[2#]

[1]Laboratory of Chemical Physics, NIDDK, National Institutes of Health, 5 Memorial Drive, Bethesda, MD 20892-0520, USA.

[2]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.

[3]Department of Biochemistry and Biophysics, Stockholm University, SE-10691 Stockholm, Sweden.

[4]Science for Life Laboratory Stockholm University, Box 1031, SE-171 21 Solna, Sweden.

[5]Gene Center, Department of Biochemistry and Center for Integrated Protein Science Munich, CiPS-M, Feodor-Lynen-Strasse 25, Ludwig Maximilian University of Munich, 81377 Munich, Germany.

[‡]P.T., A.S., R.K. and T.S. contributed equally to this work.

[#]Corresponding authors

Summary

Proteins that fold cotranslationally do so in a restricted configurational space, due to the volume occupied by the ribosome. Here, we investigate the cotranslational folding of an all-β immunoglobulin domain, titin I27, whose intrinsic folding mechanism has been extensively characterized. Using an arrest peptide-based assay and structural studies by cryo-EM, we show that I27 folds in the mouth of the ribosome exit tunnel. Simulations that use a kinetic model for the force-dependence of escape from arrest, accurately predict the fraction of folded protein as a function of length. We used these simulations to probe the folding pathway on and off the ribosome. Our simulations - which also reproduce experiments on mutant forms of I27 - show that I27 folds, while still sequestered in the ribosome, by essentially the same pathway as free I27, with only subtle shifts of critical contacts from the C to the N terminus.

Introduction

The nature of cotranslational protein folding is likely determined by a number of biophysical factors, in particular the intrinsic folding properties of the protein under consideration (1-6), together with the effects the ribosome itself may have on the folding process (7-13). Small proteins can fold inside the ribosome exit tunnel (e.g., the small zinc finger domain ADR1a) (14), other proteins can fold at the mouth of the tunnel (e.g., the three-helix bundle spectrin domains) (15), and yet others may be simply too large to fold within the confines of the ribosome (e.g., DHFR) (16). But, in addition to the spatial constraints imposed upon the nascent chain by the confines of the tunnel, the ribosome may itself affect the folding process (15, 17-20). The stability of folded or partly folded states may be reduced when folding occurs close to, or within the confines of, the ribosome simply due to steric exclusion effects. Interactions of the folded state with the ribosome itself may also be stabilising, or indeed, destabilising (21, 22). Finally, since translation is vectorial in nature, it is possible that when proteins fold cotranslationally they fold via different pathways than those used when proteins fold outside the ribosome or when isolated proteins fold *in vitro* (15, 23-26).

The folding of the protein close to the ribosome generates a pulling force on the nascent chain. This force has been probed by single molecule (27) as well as arrest peptide (AP) experiments (14-16). The folding kinetics are expected to be correspondingly altered, with the rate of folding likely to be decreased and the unfolding rate increased, in close proximity to the ribosome (which again result in destabilization of the protein). In this study, we use such arrest peptide-based cotranslational force-measurement experiments, simulations, and structural studies to investigate how the ribosome affects the folding of the small all-β immunoglobulin domain, titin I27. Results from all three techniques show that I27 folds in the mouth of the ribosome exit tunnel. Our simulations correctly capture the onset of folding in I27 and three mutant variants, allowing us to predict how destabilisation of regions that fold early and late in the isolated domain affect folding on the ribosome. Our simulations further show that the folding pathway of I27 is largely unaffected by the presence of the ribosome, except for small but significant changes observed for contacts near the N and C termini.

Results

I27 folds close to the ribosome

In order to gain insight into when I27 can commence folding on the ribosome, we employed an arrest peptide force-measurement assay (28) carried out using the PURE *in vitro* translation system, as described in (14-16). In these experiments, the *E. coli* SecM arrest peptide (AP) is used to stall the nascent protein chain temporarily during translation. The yield of full-length protein which escapes stalling in a defined time interval (f_{FL}), determined from SDS-PAGE gels, provides a proxy for the pulling force exerted on the nascent chain by the protein as it folds (14-16) (Figure 1A). By measuring f_{FL} for a set of constructs where the length L of the linker between the target protein and the SecM AP is systematically varied, a force profile can be recorded that reflects the points during translation where the folding process starts and ends.

The force profile for wild-type I27 (Figure 1B) has a distinct peak at L = 35-38 residues (see Methods for sequences of the constructs). This peak is absent from the force profile for the mutant I27[W34E], a non-folding variant of I27, demonstrating that the peak is due to a folding event and not, for example, to non-specific interactions of the unfolded nascent chain with the ribosome. The non-zero f_{FL} for the non-folding mutant is attributed to the spontaneous rate of escape from arrest in the absence of acceleration by forces associated with folding. Since it takes ~35 residues in an extended conformation to span the ~100 Å long exit tunnel (29), the critical length $L \approx 35$ residues suggests that I27 starts folding while in mouth of the exit tunnel.

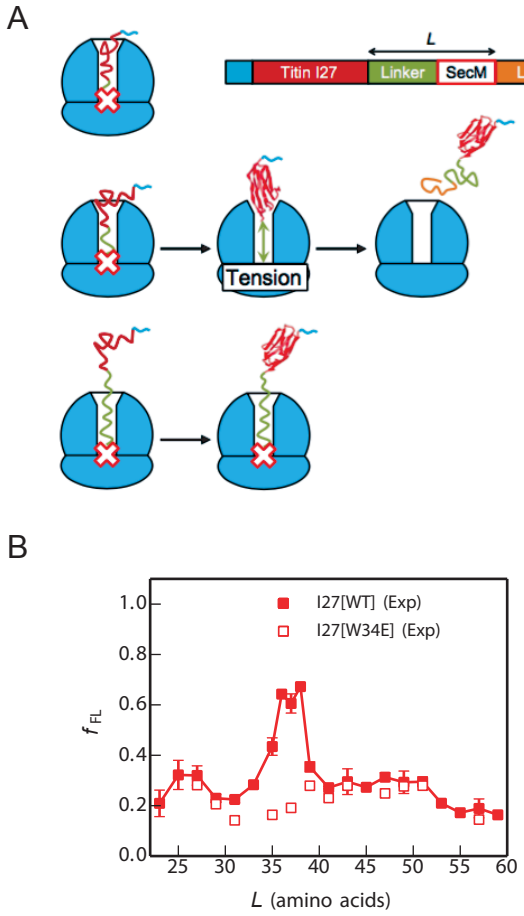


Figure 1. Cotranslational folding of the titin I27 domain by force-profile analysis. (A) The force-measurement assay (modified from (15)). I27, preceded by a His-tag, is placed *L* residues away from the last amino acid of the SecM AP, which in turn is followed by a 23-residue C-terminal tail derived from *E. coli* LepB. Constructs are translated for 15 min. in the PURE *in vitro* translation system, and the relative amounts of arrested and full-length peptide chains produced are determined by SDS-PAGE. The fraction full-length protein, f_{FL}, reflects the force exerted on the AP by the folding of I27 at linker length *L*. At short linker lengths (top), there is not enough room in the exit tunnel for I27 to fold, little force is exerted on the AP, and the ribosome stalls efficiently on the AP ($f_{FL} \approx 0$). At intermediate linker lengths (middle), there is enough room for I27 to fold but only if the linker segment is stretched, force is exerted on the AP, and stalling is reduced ($f_{FL} > 0$). At long linker lengths (bottom), I27 has already folded when the ribosome reaches the last codon in the AP, and again little force is exerted on the AP ($f_{FL} \approx 0$). (B) Force profiles for the I27 domain (solid squares) and the non-folding (nf) mutant I27[W34E] (open squares). The standard error of f_{FL} is calculated for values of *L* where three or more experiments were performed.

Cryo-EM shows that I27 folds in the mouth of the exit tunnel

To confirm that the peak in the force profile corresponds to the formation of a folded I27 domain, we replaced the SecM AP with the stronger TnaC AP (30-32) and purified stalled ribosome-nascent chain complexes (RNCs) carrying an N-terminally His-tagged I27[*L*=35]

construct (see Methods). The construct was expressed in *E. coli*, RNCs were purified using the N-terminal His-tag, and an RNC structure with an average resolution of 3.2 Å (Figure 2-figure supplement 1) was obtained by cryo-EM. In addition to the density corresponding to the TnaC AP, a well-defined globular density (~4.5-9 Å resolution) was visible protruding from the exit tunnel (Figure 2A). The NMR structure of I27 (PDB: 1TIT (33)) was fit into this density (Figure 2B and Video 1) with the C-terminal end extending into the exit tunnel. Orientation validation shows that this is the most reasonable way of fitting, (Figure 2-figure supplement 2). The I27 domain packs against ribosomal proteins uL24, uL29, and ribosomal 23S RNA (Figure 2C) as if it is being pulled tight against the ribosome by the nascent chain. We conclude that the peak at L = 35-38 residues in the force profile indeed represents the cotranslational folding of the I27 domain at the tunnel exit.

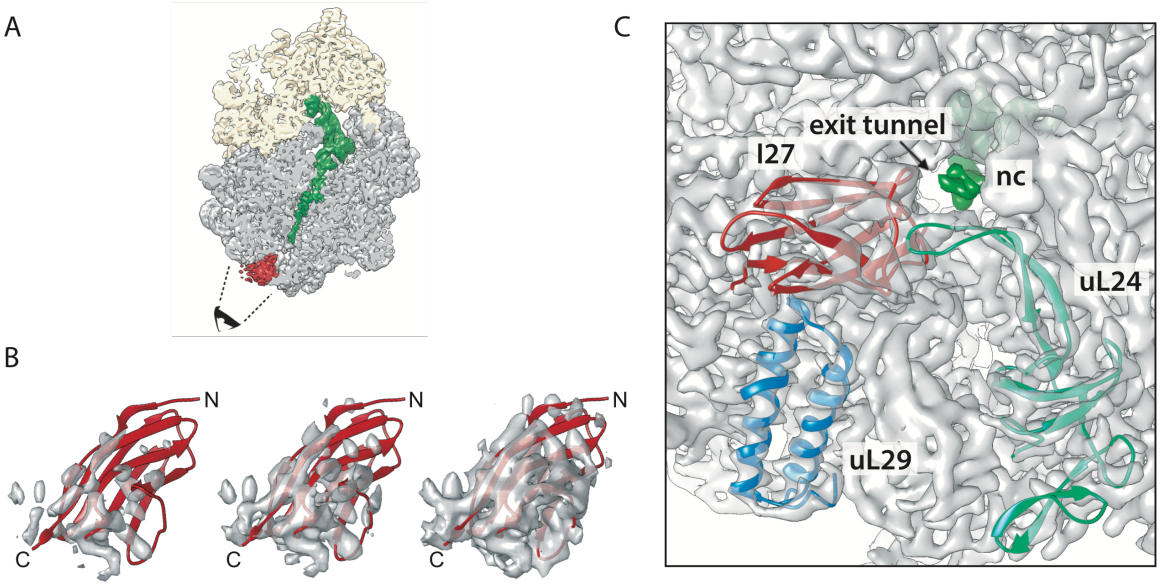
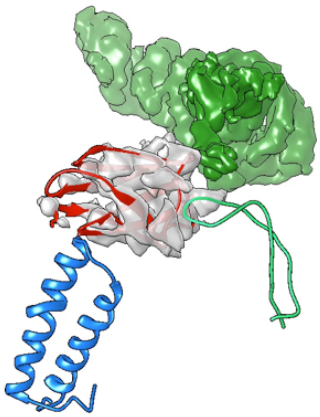


Figure 2. Cryo-EM structure of I27[L=35] RNCs. (A) Cryo-EM reconstruction of the I27–TnaC[L = 35] RNC. The ribosomal small subunit is shown in yellow, the large subunit in grey, the peptidyl-tRNA with the nascent chain in green, and an additional density corresponding to I27 at the ribosome tunnel exit in red. The black cartoon eye and dash lines indicate the view of angle of panel (C). The density contour level for feature visualization is at 1.7 times root-mean-square deviation (1.7 RMSD). (B) Rigid-body fit of the I27 domain (PDB: 1TIT) to the cryo-EM density map displaying from high (left) to low (right) contour levels at 2.6, 2.0 and 1.4 RMSD, respectively. N and C represent the N and C termini of the I27 domain, respectively. (C) View looking into the exit tunnel (arrow) with density for the nascent chain (nc) in dark green. Ribosomal proteins uL29 (blue; PDB: 4UY8), uL24 (light green; the β hairpin close to I27 domain was re-modelled based on PDB: 5NWY) and the fitted I27 domain (red) are shown in cartoon mode; 23S RNA and proteins not contacting I27 are shown as density only. The density contour level is at 5 RMSD excluding tRNA, nascent chain and I27 domain, which are displayed at 1.7 RMSD.



Video 1. Cryo-EM density of ribosome and I27 (one static frame of the video). Video showing cryo-EM map for I27[L=35] RNCs. 30S in yellow, 50S and I27 domain in grey, tRNA and nascent chain in green, the model (PDB: 1TIT) of I27 domain in red.

Coarse-grained molecular dynamics simulations recapitulate I27 folding on the ribosome

In order to provide an independent means of estimating the force exerted on the nascent chain by the cotranslational folding of the I27 domain, we calculated force profiles by coarse-grained MD simulations (see Methods). Briefly, in the MD model, the 50S subunit of the *E. coli* ribosome (34) (PDB: 3OFR) and the nascent chain are explicitly represented using one bead at the position of the Cα atom per amino acid, and three beads (for P, C4, N3) per RNA base (Figure 3A). The interactions within the protein were given by a standard structure-based model (35-37), which allowed it to fold and unfold. Interactions between the protein and ribosome beads were purely repulsive. The ribosome beads were fixed in space, as in previous simulation studies (38). I27 was covalently attached to unstructured linkers having the same sequences as those used in the force-profile experiments (Figure 3B) and the C terminus of the linker was tethered to the last P atom in the A-site tRNA with a harmonic potential, allowing the force exerted by the folding protein to be directly measured. For each linker length L, we used umbrella sampling to determine the average force exerted on the AP by the protein in the folded and unfolded states while arrested, as well as the populations of those two states (Figure 3C).

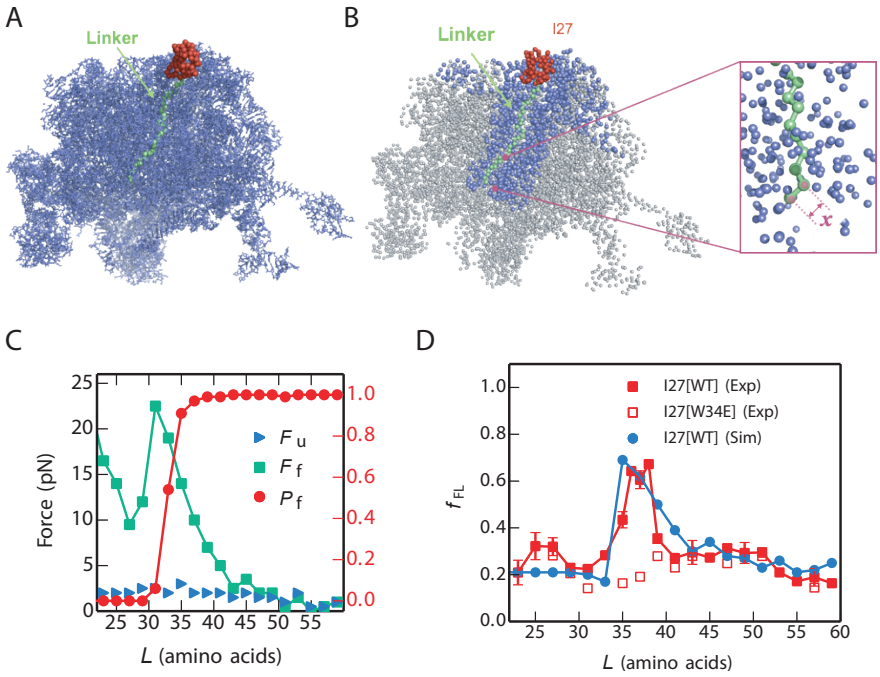
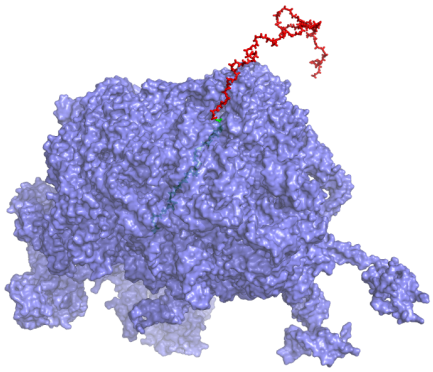


Figure 3. MD simulations of cotranslational folding of I27. (A) 50S subunit of the *E. coli* ribosome (PDB: 3OFR) with I27[L=35] attached via an unstructured linker. (B) Coarse-grained model for I27 (red) and linker (green), with surrounding ribosomal pseudo-atoms in blue. Pseudo-atoms with grey colour are not used in the simulations. The instantaneous force exerted on the AP is calculated from the variation in the distance x between the C-terminal Pro pseudo-atom and the next pseudo-atom in the linker (see inset). (C) Average forces exerted on the AP by the unfolded state (F_u, blue) and folded state (F_f, green) of I27 at different linker lengths L. The average fraction folded I27 for different L, P_f, is shown in red on the right axis. Free energy profiles of each linker length are shown in Figure 3-supplement figure 1. (D) Experimental (red) and simulated (blue) force profiles for cotranslational folding of I27.



Video 2. MD folding simulation (one static frame of the video). Video showing an unbiased 1.8 μsec fragment of an MD trajectory of I27 folding and unfolding at linker length L=35. Ribosome shown in blue surface representation, I27 and linker in red and green wireframe respectively.

Given the empirically-determined force-dependence of the escape rate $k(F)$ (27), here approximated by a Bell-like model (39), we can determine the expected escape rate while the protein is in the unfolded or folded state, from which the fraction full-length protein obtained with a given linker length and incubation time can be determined, as described in Methods. The calculated f_{FL} profile for I27 is shown in Figure 3D (see also Figure 1-figure supplement 1). It matches the experimental profile remarkably well. In the simulations with the I27[L=35] construct, the folded I27 domain is seen to occupy positions that largely overlap with the cryo-EM structure (Video 2). Overall, these results suggest that the MD model provides a good representation of the folding behaviour of the I27 domain in the ribosome exit tunnel.

Force profiles of I27 variants probe the folding pathway

To test whether the cotranslational folding pathway is the same as that observed for the isolated I27 domain *in vitro*, we decided to investigate three destabilised variants of I27, both by simulation and experiment. One mutation in the core, Leu 58 to Ala (L58A), located in β-strand E (Figure 4A) destabilizes the protein by 3.2 kcal mol^{-1}, and removes interactions that form early during folding of the isolated domain, playing a key role in formation of the folding nucleus (ϕ-value = 0.8) (40). Two further mutations, M67A and deletion of the N-terminal A-strand, remove interactions that form late in the folding of I27 (*i.e.*, both mutants have low ϕ-values (40, 41)). The A-strand is the first part of I27 to emerge from the ribosome, while M67 is located in a part of I27 that is shown by cryo-EM to be located in very close proximity to a β hairpin loop of ribosomal protein uL24 in I27-TnaC[L=35] RNCs (Figure 2-figure supplement 3A). The interaction with the I27 domain shifts the tip of this uL24 hairpin by about 6 Å compared to its location in other RNC structures (Figure 2-figure supplement 3B).

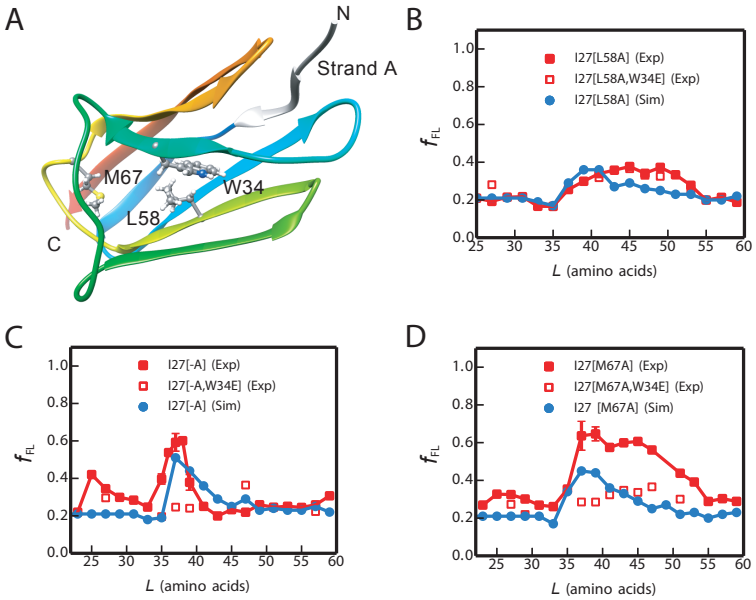


Figure 4. Simulations capture the experimental force profiles for mutant I27 domains. (A) Mutated residues in I27 (sticks). (B-D) Experimental (red) and calculated (blue) force profiles for (B) I27[L58A], (C) A-strand deletion mutant I27[-A], (D) I27[M67A]. Experimental force profiles for non-folding mutants that contain an additional W34E mutation are shown as red open squares.

The simulated force profile for the L58A variant predicts a much lower force peak than for wild-type I27; likewise, the experimental force peak is lower and broader than for wild-type, extending from L = 37-53 residues (Figure 4B). The f_{FL} values are very similar to those obtained for I27[L58A,W34E], a non-folding variant of I27[L58A]. Therefore, the weak forces seen at $L \approx$ 40-50 residues are not due to a folding event, indicating that I27[L58A] does not exert an appreciable force due to folding near the ribosome.

The A-strand comprises the first seven residues of I27 and removal of this strand, I27[-A], results in a destabilisation of 2.78 kcal mol^{-1}; however, both the simulated and experimental force profiles for I27[-A] are very similar to those for wild-type I27 (Figure 4C). Residue M67 is located in the E-F loop, and mutation to alanine results in a destabilisation of 2.75 kcal mol^{-1}; for this variant, folding commences at $L \approx 35$ residues as for wild-type I27, but the peak is much broader (Figure 4D). In this case, simulation does not predict the full height of the peak observed in the experimental f_{FL} profile, but is suggestive of a broader peak. Non-folding control experiments for variants I27[-A,W34E], and I27[M67A,W34E] (Figure 4C and D) show that the peaks in the force profiles for these variants are due to a folding event. These results show that deletion of the A-strand and destabilisation of the E-F loop do not

affect the onset of cotranslational folding of I27, but that the M67A mutation increases the width of the folding transition.

The folding pathway is only subtly affected by the presence of the ribosome.

To compare the folding pathways when the protein is folding near the tunnel exit or outside the ribosome, we estimated ϕ-values based on the transition path of I27 folding on the ribosome from our coarse-grained simulations, using a method introduced previously(42). For each linker length, 30 transition paths were collected from MD simulations. To reduce the uncertainty from experimental reference data, the experimental ϕ-values were chosen if the thermodynamic stability change between the mutant and the wild type is sufficiently large ($|\Delta\Delta G| > 7$ kJ/mol) (43). As seen in Figure 5A, when the linker length is long (L = 51 residues) and I27 is allowed to fold outside the ribosome, the calculated ϕ-values are consistent with the experimental values obtained for the folding of isolated I27 *in vitro* (40). For shorter linker lengths (L = 31 and 35 residues), calculated ϕ-values remain largely unchanged except for a slight increase near the N terminus (around residues 3-6) and a slight decrease near the C terminus (around residues 72-74) (Figure 5B and C).

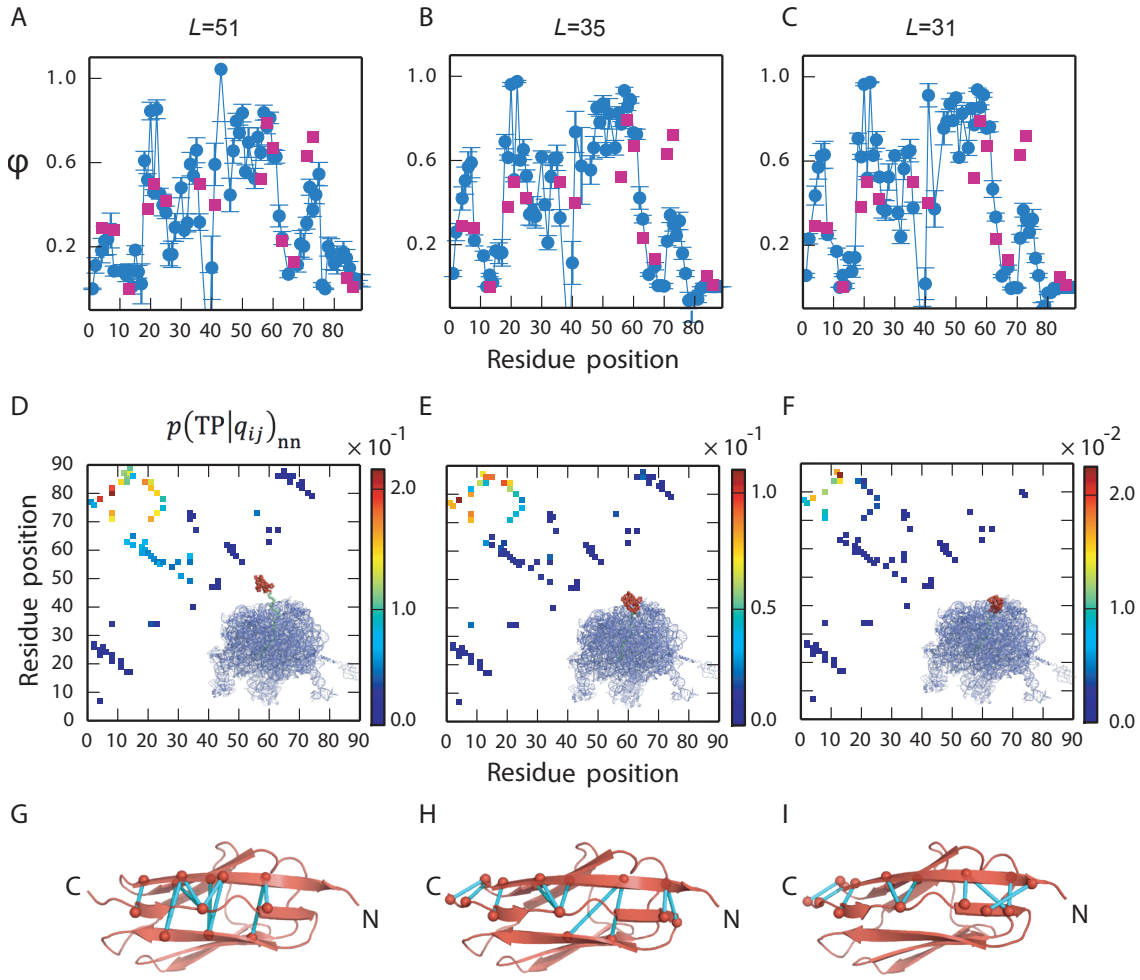


Figure 5. Simulated folding pathways for ribosome-tethered I27. LH column, L=51; middle column, L= 35; RH column, L=31. Top panels: Simulated ϕ-values for I27 (blue). ϕ-values determined by *in vitro* folding of purified I27 are shown as red squares. At L=51 the simulated ϕ-values match well with experiment. At L=35 and L=31 the simulated ϕ-values are higher at the N terminus and lower at the C terminus, than the experimental values, reflecting a change in importance of these regions when I27 folds in the confines of the ribosome. Middle row: Relative probability that if a particular contact is formed then the protein is on a folding trajectory, $p(\mathrm{TP}|q_{ij})_{\mathrm{nn}}$. When the protein is constrained the limiting factor is formation of a few key contacts. A cartoon of the ribosome with I27 in red is shown on each panel. Bottom row: The top ten most important contacts are coloured in cyan on the native structure.

To obtain a more detailed picture regarding the relative importance of different native contacts in the folding mechanism, we computed the conditional probability of being on a transition path (TP), given the formation of a contact q_{ij} between residues i and j, $p(\mathrm{TP}|q_{ij})_{\mathrm{nn}}$ (44). This quantity indicates which native contacts are most important for determining a successful folding event. $p(\mathrm{TP}|q_{ij})_{\mathrm{nn}}$ is closely related to the frequency of the

contact q_{ij} on transition paths $p(q_{ij}|\mathrm{TP})$, but is effectively normalized by the probability that the contact is formed in non-native states $p(q_{ij})_{\mathrm{nn}}$, and can be expressed as:

$$p(\mathrm{TP}|q_{ij})_{\mathrm{nn}} = \frac{p(q_{ij}|\mathrm{TP})p(\mathrm{TP})_{\mathrm{nn}}}{p(q_{ij})_{\mathrm{nn}}} \quad [1]$$

where $p(\mathrm{TP})_{\mathrm{nn}}$ is the fraction of non-native states which are on transition paths at equilibrium. The subscript nn means that only the non-native segments of a trajectory are included, *i.e.*, unfolded states and transition paths; the native, folded state is not included in the calculation since native contacts are always formed in this state. The simulations suggest that formation of native contacts between the N and C termini is somewhat more important when folding takes place in the mouth of the exit tunnel (L = 31 residues) than far outside the ribosome (L = 51 residues) (Figure 5D-F, upper left-hand corner in the panels). This is likely due to the greater difficulty of forming these contacts (examples are shown in Figure 5G-I) under ribosomal confinement; therefore, forming them becomes more critical in enabling the protein to fold.

Discussion

Using a combination of MD simulation, force-profile measurements and cryo-EM, we have investigated the cotranslational folding pathway of the 89-residue titin I27 domain. I27 has been extensively characterised in previous *in vitro* folding studies (40, 41, 45-55). Results from all three techniques show that wild-type I27 folds in the mouth of the ribosome exit tunnel; in the cryo-EM structure of I27-TnaC[L=35] RNCs, I27 packs against ribosomal proteins uL24, uL29, and ribosomal 23S RNA. This is in apparent contrast to a previous NMR study on another Ig-like protein, in which the domain was shown to acquire its native fold (as reflected in the NMR spectrum) only when fully outside the ribosome tunnel, at L = 42-47 residues linker length (22).

In order to determine the molecular origin of the measured force profile, we performed molecular dynamics simulations of I27 folding on the ribosome, varying the length of the linker sequence between the arrest peptide and the I27 domain. We calculated the pulling force directly from the simulations and translated this into yield of folded protein using a kinetic model parameterized based on known release kinetics of the SecM AP. This enabled us to recapitulate the experimental arrest peptide force measurement profile, and therefore relate f_{FL} directly to the force exerted on the arrest peptide. Our simulations demonstrate the

direct effect that the restoring force of the nascent chain can have on determining when the protein folds on the ribosome. We show that f_{FL} depends upon a combination of the force exerted by the folded protein and the fraction of folded protein at the given linker length L.

In order to relate how destabilization of regions that fold early and late in the isolated domain affects folding on the ribosome, we used simulations to predict the onset of folding in three mutant variants of I27. A previous ϕ-value analysis of I27 (40) showed that early packing of the structurally central β-strands drives the folding of this domain, while peripheral strands and loop regions pack later in the folding process. Mutations in the folding core (such as L58A) slow folding, whereas mutations in the periphery have no effect on folding rates (40). L58 is a key residue in the critical folding nucleus and almost fully packed in the transition state, in isolated domain studies. The simulated and experimental force profiles of I27 [L58A] show that this variant does not fold in or near the exit tunnel; hence, destabilisation of the central folding core prevents folding close to the ribosome. Since isolated I27[L58A] is fully folded, it is likely that this variant can only fold cotranslationally at longer linker lengths, when it is no longer in close proximity to the ribosome and exerts little force on the nascent chain.

Our experiments show that I27 variants destabilized in regions of the protein that are unstructured, or only partially structured, in the transition state, are still able to commence folding close to the ribosome. The force profiles reveal that the onset of folding of mutants with the A-strand deleted, or with the Met 67 to Ala mutation in the E-F loop, is the same as for wild-type although these have a similar destabilisation as L58A (Figure 4). We cannot at present explain the broad force peak for the M67A mutant, but note that the mutation is in a region that interacts closely with ribosomal protein uL24 in the wild-type cryo-EM structure.

Our simulations reproduce the onset of folding in the three mutant variants of I27 (Figure 4), and so give us the confidence to investigate how confinement within the ribosome affects the folding pathway of I27. We used simulations to investigate the folding of I27 arrested on the ribosome at various linker lengths, using a Bayesian method for testing the importance of specific contacts on the folding pathway, as well as by computing ϕ-values (Figure 5). Overall, we find that the mechanism and pathway of folding are robust towards variation in linker length and relatively insensitive to the presence of the ribosome; small but significant changes are observed only for contacts near the N and C termini. These shifts are consistent with the greater importance of forming N-terminal contacts when the C terminus is

sequestered within the exit tunnel, possibly to compensate for loss of contacts at the C terminus.

We have previously shown that α-helical proteins can fold co-translationally (15), perhaps unsurprising since helical structures are dominated by short-range interactions and helices can form within the ribosome tunnel itself (56, 57). Here, our equilibrium arrest-peptide assay and structural studies reveal that an all-β protein, titin I27, is able to fold within the mouth of the ribosome exit tunnel, despite its folding being dominated by long-range interactions. Molecular simulations, accounting for the effect of the entropic restoring force on protein stability, reproduce the yield of protein from experiments remarkably well. These simulations reveal that I27 folds on the ribosome by the same pathway as when the protein folds away from the confines of the ribosome.

Acknowledgements

This work was supported by grants from the Knut and Alice Wallenberg Foundation, the Swedish Cancer Foundation, and the Swedish Research Council to GvH, by grants from the Deutsche Forschungsgemeinschaft (DFG) GRK 1721 and FOR1805 to RB, by a DFG fellowship through the Graduate School of Quantitative Biosciences Munich (QBM) to TS, and by the Wellcome Trust (WT095195, to JC); PT and RB were supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health; JC is a Wellcome Trust Senior Research Fellow. The cryo-EM data were collected at the Swedish National Cryo-EM Facility funded by the Knut and Alice Wallenberg Foundation, the Family Erling Persson Foundation and the Science for Life Laboratory. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov)

Methods

Enzymes and chemicals

All enzymes were obtained from Thermo Scientific. Oligonucleotides were purchased from Life Technologies. In-Fusion Cloning kits were obtained from Clontech and DNA purification kits were purchased from Qiagen. PURExpress cell-free translation system was obtained from Eurogentec. [35S]-methionine was purchased from Perkin Elmer. Instant Blue protein stain was purchased from Expedeon.

DNA manipulation

Titin I27 constructs for *in vitro* translation were generated in pRSET A plasmid (Invitrogen) (previously modified to remove the sequence including the entire T7 gene 10 leader and EK recognition site up to, but not including, the *Bam*H I site and replaced with a sequence encoding residues L, V, P, R, G, S) carrying the *E. coli* SecM arrest peptide (FSTPVWISQAQGIRAGP) and a truncated *E. coli lepB* gene, under the control of a T7 promoter. Increasing linker lengths were generated in pRSET A by PCR; linear pRSET A constructs (containing the SecM AP and truncated lepB, but lacking I27) were generated by PCR using primers which extended the linker from 23 aa to 63 aa (in steps of 2 aa) from the direction of the C to the N terminus. I27 flanked by GSGS linkers was amplified by PCR with overhanging homology to the plasmid containing the desired linker length. Cloning was performed using the In-Fusion system (Takara Bio USA, Inc.), according to the manufacturer's instructions. The final two C-terminal residues (EL) of the 89 aa Titin I27 construct are not structured in the PDB file 1TIT, and are therefore included in the linker region. The amino acid sequence of the construct I27[*L*=63] is as follows (I27 in bold and SecM AP underlined):

MRGSHHHHHHGLVPRGSGS**LIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWK LKGQPLAASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANTKSAANLKVK**EL SGSGKFAYGIKDPIYQKTLVPGQQNATWIVPPGQYFMMGDWMSSFSTPVWISQAQG IRAGPGSSDKQEGEWPTGLRLSRIGGIH**

The mutants I27[–A] (lacking β-strand A), I27[L58A] and I27[M67A] were generated for each linker length by site-directed mutagenesis. For the wild-type I27 and I27[–A] constructs with *L* = 27, 35, 37, 39, 47 and 57 residues, site-directed mutagenesis was performed to generate constructs with the non-functional FSTPVWISQAQGIRAGA arrest peptide (mutated residue underlined) as full-length controls, and constructs with the crucial Pro, at the end of the AP, substituted with a stop codon as arrest controls. Site-directed mutagenesis was performed to generate W34E variants as non-folding (nf) controls at *L* = 27, 29, 31, 35, 37, 39, 41, 43, 47, 49, 51 and 57 for wild-type I27; *L* = 27, 35, 37, 39, 47 and 57 residues for I27[–A]; *L* = 27, 41, 45, 47, 49 and 53 residues for I27[L58A]; *L* = 27, 29, 37, 39, 41, 43, 45, 47 and 51 residues for I27[M67A]. All constructs were verified by DNA sequencing.

In vitro transcription and translation

Transcription and translation were performed using the commercially available PURExpress *in vitro* system (GeneFrontier Corporation), according to the manufacturer's protocol, using 250 µg plasmid DNA as template. Synthesis of [^{35}S]-Met-labeled polypeptides was performed at 37 °C, 500 r.p.m. for exactly 15 min. The reaction was quenched by the addition of an equal volume of 10% ice-cold trichloroacetic acid (TCA). The samples were incubated on ice for 30 min and centrifuged for 5 min at 20,800 × *g* and 4 °C. Pellets were dissolved in sample buffer and treated with RNase A (400 µg ml^{-1}) for 15 min at 37 °C before the samples were resolved by SDS-PAGE and imaged on a Typhoon Trio or Typhoon 9000 phosphoimager (GE Healthcare). Bands were quantified using ImageJ to obtain an intensity cross section, (http://rsb.info.nih.gov/ij/), which was subsequently fit to a Gaussian distribution using in-house software (Kaleidagraph, Synergy Software). The fraction full-length protein, f_{FL}, was calculated as $f_{FL} = I_{FL}/(I_{FL}+I_A)$, where I_{FL} and I_A are the intensities of the bands representing the full-length and arrested forms of the protein. For wild-type I27 and six nf control samples (L = 27, 35, 37, 39, 47 and 57 residues), *in vitro* transcription and translation were also performed at 37 °C, 500 r.p.m. for exactly 30 min. The resultant force profile was slightly higher than that obtained at 15 min but has essentially the same shape (Figure 1-figure supplement 1).

The reproducibility of force profile data has been discussed previously (15). For wild-type I27, data points L = 61 and 63 residues are a single experiment; L = 33, 36, 38, 45, 53, 55 and 59 residues are an average of 2 experiments; all other values of L are an average of at least 3 experiments. For I27[–A] strand, L = 23, 25, 33, 41, 43, 51, 53, 55 residues are a single experiment; all other values of L are an average of 2 experiments, except L = 35, 37 and 39 residues which are an average of at least 3 experiments. For I27[L58A], all data points are a single experiment except L = 27, 37, 41, 45, 47, 49 and 53 residues, which are an average of 2 experiments. For I27[M67A], L = 23, 25 and 51 – 63 residues are a single experiment; L = 29 – 35 residues are an average of 2 experiments; L = 27, 37 – 47 and 51 residues are an average of at least 3 experiments. For wild-type I27 samples incubated for 30 min, all data points are a single experiment except L = 27, 35, 37, 39, 47 and 57 residues, which are an average of 2 experiments. For non-folding controls, all data points are a single experiment except for wild-type I27 L = 29, 31, 39, 43 and 47 residues which are an average of 2 experiments.

Cloning and purification of ribosome-nascent chain complexes

The I27 construct at L = 35, which is at the peak of f_{FL} (Figure 1B), was studied by cryoEM. The SecM AP in these constructs was substituted with the TnaC AP (32) for more stable arrest, and the constructs were engineered to maintain a linker length of 35 amino acid residues. An N-terminal 8X His tag was introduced to enable purification. The amino acid sequence of the construct used was (I27 in bold and TnaC AP underlined):

MDMGHHHHHHHHDYDIPTTLEVLFQGPGT**LIEVEKPLYGVEVFVGETAHFEIELS EPDVHGQWKLKGQPLAASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANTKS AANLKVK**ELSGSGSGSGGPNILHISVTSKWFNIDNKIVDHRP**

The construct was engineered into a pBAD expression vector, under the control of an arabinose-inducible promoter. The translation-initiation region was optimized as described in (58). The plasmid was transformed into the *E. coli* KC6 *ΔsmpB ΔssrA* strain. 4 colonies were picked and tested for expression of the RNCs at 37°C in Lysogeny broth (LB).

Large-scale purification of RNCs was carried out based on a protocol described in (32). Briefly, a single colony of the KC6 cells found to express the RNCs was picked and cultured in LB at 37°C to an A_{600} of 0.5. Expression was induced with 0.3% arabinose and was carried out for 1 hour. Thereafter, the cells were chilled on ice, harvested by centrifugation, and resuspended in Buffer A at pH 7.5 (50 mM HEPES-KOH, 250 mM KOAc, 2 mM Tryptophan, 0.1% DDM, 0.1% Complete protease inhibitor). Cell lysis was carried out by passing the cell suspension thrice through the Emulsifex (Avestin) at 8000 psi at 4°C. The lysate was cleared of cell debris by centrifugation at 30,000xg for 30 min in the JA25-50 rotor (Beckman Coulter). The supernatant obtained was loaded on a 750 mM sucrose cushion (in Buffer A) and centrifuged at 45, 000 x g for 24 hours in a Ti70 rotor (Beckman Coulter) to obtain a crude ribosomal pellet, which was resuspended in 200 µl Buffer A by shaking gently on ice.

RNCs from the crude suspension were purified via their His tags by affinity purification using Talon (Clontech) beads, which was pre-incubated with 10 µg/ml tRNA to reduce unspecific binding of ribosomes. The suspension was incubated with the beads for 1 hour at 4°C and subsequently washed with 20 column volumes of Buffer B at pH 7.5 (50 mM HEPES-KOH, 10 mM $Mg(OAc)_2$, 0.1% Complete Protease Inhibitor, 250 mM sucrose, 2 mM Tryptophan). RNCs were eluted by incubating the Talon beads with Buffer C at pH 7.5 (50 mM HEPES, 150 mM KOAc, 10 mM $Mg(OAc)_2$, 0.1% Complete protease inhibitor, 150 mM imidazole, 250 mM sucrose) for 15 minutes and subsequently collecting the flow-

through. Elution was carried out thrice and the eluents were concentrated by centrifugation at 40,000 rpm for 2.5 hours in a TLA 100.3 rotor (Beckman Coulter). The pellet obtained at the end of this step was gently suspended in a minimal volume of Buffer D at pH 7 (20 mM HEPES-KOH, 50 mM KOAc, 5 mM $Mg(OAc)_2$, 125 mM sucrose, 2 mM Trp, 0.03% DDM).

CryoEM sample preparation and data collection

Approximately 4 A_{260}/ml units of RNCs were loaded on Quantifoil R2/2 grids coated with carbon (3 nm thick) and vitrified using the Vitrobot Mark IV (FEI-Thermo) following the manufacturer's instructions. CryoEM data was collected at the CryoEM National Facility at the Science for Life Laboratory in Stockholm, Sweden.

Data was acquired on a 300 keV Titan Krios microscope (FEI) equipped with a K2 camera and a direct electron detector (both from Gatan). The camera was calibrated to achieve a pixel size of 1.06 Å at the specimen level. 30 frames were acquired with an electron dose 0.926 $e^-/Å^2$/frame and a total dose of 27.767 $e^-/Å^2$ and defocus values between -1 to -3 µm. The first two frames were discarded and the rest were aligned using MotionCor2 (59). Raw images were cropped into squares by RELION 2.1 beta 1 (60). Power-spectra, defocus values and estimation of resolution were determined using the Gctf software (61) and all 2,613 micrographs were manually inspected in real space, in which 2,613 were retained. 468,015 particles were automatically picked by Gautomatch (http://www.mrc-lmb.cam.ac.uk/kzhang/) using the *E. coli* 70S ribosome as a template. Single particles were processed by RELION 2.1 beta 1 (60). After 80 rounds of 2D classification, 384,039 particles were subjected to 3D refinement using the *E. coli* 70S ribosome as reference structure, followed by 160 rounds of 3D classification without masking and 25 rounds of tRNA-focused sorting. One major class containing 301,510 particles (64% of the total) was further refined including using a 50S mask, resulting in a final reconstruction with an average resolution of 3.2 Å (0.143 FSC). The local resolution was calculated by ResMap (62). Finally, the final map was obtained by local B-factoring followed by low-pass filtering to 4.5 Å by RELION 2.1 beta 1 (60) in order to best demonstrate the I27 domain.

Model docking and validation

The NMR model (PDB: 1TIT) of I27 domain was fitted into the corresponding density using UCSF Chimera (63). To validate the orientation of the fitted model, all the four possible orientations were compared. Briefly, the model with four different orientations were

converted into densities (8 Å) by UCSF Chimera, and the cross-correlation coefficients of each model map and the isolated I27 density were calculated by RELION 2.1 beta 1 (60).

Coarse-grained molecular simulations

The 50S subunit of the *E. coli* ribosome (PDB: 3OFR (34)) and the nascent chain are explicitly represented using one bead at the position of the C□ atom per amino acid, and three beads (for P, C4, N3) per RNA base (Figure 2A). The interactions within the protein were given by a standard structure-based model (35-37), which allowed it to fold and unfold. Interactions between the protein and ribosome beads were purely repulsive and given by the same form of potential as for the structure-based model(35-37), but with the coefficients determined from a mixing rule,

$$V_{ij} = \varepsilon_{ij}\left[\frac{A}{r_{ij}{}^{12}} - \frac{B}{r_{ij}{}^{10}} + \frac{C}{r_{ij}{}^{6}}\right] \quad [2]$$

where r_{ij} is the distance between two beads i and j, ε_{ij} (=0.001 kJ/mol) sets the strength of the repulsive interactions. The amino acid, phosphate, sugar and base are assigned with collision radii σ_i = 4.5, 3.2, 5.1 and 4.5 Å respectively, and A = $\sqrt{\sigma_i{}^{12}\sigma_j{}^{12}}$, B = $\frac{2}{\sigma_i{}^{-10}+\sigma_j{}^{-10}}$ and C = $\sqrt{\sigma_i{}^{6}\sigma_j{}^{6}}$.

During the simulations, the ribosome atoms were held immobile, as in previous studies (38). The linker between the AP and I27 was tethered by its C terminus to the last P atom of the A-site tRNA, but was otherwise free to fluctuate. The trajectory was propagated via Langevin dynamics, with a friction coefficient of 0.1 ps^{-1} and a time step of 10 fs, at 291 K in a version of the Gromacs 4.0.5 simulation code, modified to implement the potential given by Equation 2 (64). All bonds (except the one used to measure force, below) were constrained to their equilibrium length using the LINCS algorithm (65).

To calculate the pulling force exerted on the nascent chain by the folding of I27, the bond between the last and the second last amino acid of the SecM AP was modelled by a harmonic potential as a function the distance between these two atoms, x (Figure 3B):

$$E = \frac{1}{2}k_s(x - x_0)^2 \quad [3]$$

where x_0 is a reference distance. Here x_0 is set to 3.8 Å, which is the approximate distance between adjacent Cα atoms in protein structures, k_s is a spring constant, set to 3000 kJ.mol.nm^{-2} so that the average displacement $x - x_0$ remains below 1 Å for forces up to ~500

pN, which is much larger than the forces actually exerted by the folding protein. The pulling force on the nascent chain was measured by the extension of this bond as $F = -k_s(x - x_0)$.

I27 was covalently attached to unstructured linkers having the same sequences as used in the force-profile experiments (see Figure 1B). Linker amino acids are repulsive to both the ribosome and I27 beads, with interaction energy as described in Equation 2.

The protein in its arrested state is subject to force $F(t)$, which will fluctuate, for example when the protein folds or unfolds. The rate of escape from arrest has been shown to be force-dependent (27); here we approximate the sensitivity to force using the phenomenological expression originally proposed by Bell (39)

$$k(F) = k_0 e^{\beta F \Delta x^{\ddagger}}, \quad [4]$$

where k_0 is a zero-force rupture rate, $\Delta x^{\ddagger}$ is the distance from the free energy minimum to the transition state, $\beta = 1/k_B T$ where k_B is Boltzmann's constant and T the absolute temperature. While there are functions to describe force-dependent rates with stronger theoretical basis, we use the Bell equation due to its simplicity and because its parameters have previously been estimated from experiment for the SecM AP. In all cases, we set k_0 (Equation 4) to 3 $\times 10^{-4}$ s^{-1} and $\Delta x^{\ddagger}$ to 7 Å, based on the values determined by Goldman *et al.* (they estimated $\Delta x^{\ddagger}$ to be 1-9 Å) (27).

We assume the probability of remaining on the ribosome $S(t) = 1 - f_{FL}(t)$ assuming that $\dot{S} = -k(F(t))$, hence

$$S(t) = \exp[-\int_0^t k(F(t))dt]. \quad [5]$$

If we further assume that folding is two-state, and that the escape from the ribosome is slow relative to the folding and unfolding of the protein, we can approximate $S(t)$ in terms of the mean forces experienced when the protein is unfolded, F_u, or folded, F_f, with unfolded and folded populations of P_u and P_f respectively,

$$S(t) \approx \exp\left[-t[P_u k(F_u) + P_f k(F_f)]\right]. \quad [6]$$

The equilibrium properties of the system for each linker length were obtained from umbrella sampling using the fraction of native contacts Q as the reaction coordinate, allowing P_u, P_f and F_u, F_f to be determined (Figure 3C). The details of the definition of Q have been previously described (44); in short, Q is defined as

$$Q = \frac{1}{N}\sum_{(i,j)} \frac{1}{1 + e^{(r_{ij} - \lambda r_{ij}^0)}}, \quad [7]$$

where the sum runs over the N pairs of native contacts (i, j), r_{ij} is the distance between i and j in configuration, r_{ij}^0 is the distance between i and j in the native state, λ =1.2 which accounts for fluctuations when the contact is formed. A boundary of Q = 0.5 is used to separate folded from unfolded states.

In order to characterize folding mechanism, we used transition paths from folding simulations for the L = 51 case at 291K. 50 independent simulations, each started from fully extended configurations, were carried out for 4 microseconds. The folding barriers for the L = 31 and 35 cases are very high at the same temperature, therefore the transition paths are obtained from unfolding simulations instead. Starting from native-like folded configurations, 50 unfolding simulations were carried out, with each trajectory being 4 microseconds long. Transition paths were defined as those portions of the simulation trajectory from the last time I27 samples the configuration with $Q < 0.3$ till the first time it samples a configuration with Q > 0.7 (in the folding direction; opposite for unfolding). ϕ-Values were computed from the transition paths using the approximation:

$$\phi(i) \approx \sum_{j:(i,j)\in \text{native}} p(q_{ij}|TP)$$

In which $p(q_{ij}|TP)$ is the probability that the native contact q_{ij} between residues i and j is formed on transition paths as defined above. We also characterized the importance of individual contacts in determining the folding mechanism using $p\left(\text{TP}|q_{ij}\right)_{\text{nn}}$, defined in Equation 1 of the main text, i.e. the probability of being on a transition path given that contact q_{ij} is formed and the protein is not yet folded. Having already calculated $p(q_{ij}|TP)$ above, evaluating $p(TP|q_{ij})_{\text{nn}}$ required $p\left(q_{ij}\right)_{\text{nn}}$, the probability of a contact being formed in all non-native fragments of the trajectory, and $p(\text{TP})$, the fraction of time spent on transition paths. For L=51, we obtained $p\left(q_{ij}\right)_{\text{nn}}$ directly from unbiased folding simulations, using the portion of the trajectory up to the first folding event (i.e. the first time $Q > 0.7$). For L=31 or 35, where the protein is still relatively unstable, we determined it from unfolding simulations by computing $p(q_{ij})$ separately for the unfolded and transition-path portions of the trajectory and combining them weighted by $p(\text{TP})_{\text{nn}}$. We determined $p(\text{TP})_{\text{nn}}$ via folding (L = 51

case) and unfolding (L = 31 and L = 35 cases) simulations (described above). For the L = 51 case, $p(\mathrm{TP})_{\mathrm{nn}} = \frac{2t_{\mathrm{TP}}}{2t_{\mathrm{TP}}+t_{\mathrm{F}}^{\mathrm{mfpt}}}$, where t_{TP} is the mean transition path time and $t_{\mathrm{F}}^{\mathrm{mfpt}}$ is the mean first passage time for folding obtained from the maximum likelihood estimator $t_{\mathrm{F}}^{\mathrm{mfpt}} = [N_{\mathrm{fold}}t_{\mathrm{fold}} + (N - N_{\mathrm{fold}})t_{\mathrm{sim}}]/N_{\mathrm{fold}}$, where N is the total number of trajectories ($N = 50$), N_{fold} is the number of trajectories folding within 4 μs, t_{fold} is the average folding time (of the trajectories which fold), and t_{sim} is the length of the simulations (4 μs). For the L = 31 and L = 35 cases, it is less efficient to obtain the folding time $t_{\mathrm{F}}^{\mathrm{mfpt}}$ directly, therefore we estimate it based on the mean first passage time for unfolding, $t_{\mathrm{U}}^{\mathrm{mfpt}}$, from unfolding simulations. $p(\mathrm{TP})_{\mathrm{nn}} = \frac{2t_{\mathrm{TP}}}{2t_{\mathrm{TP}}+\frac{p_U}{p_F}t_{\mathrm{U}}^{\mathrm{mfpt}}}$, where p_U and p_F are the equilibrium populations of the unfolded and folded respectively determined from umbrella sampling.

Figure preparation

Figures showing electron densities and atomic models were generated using UCSF Chimera (63). Electron densities are shown at multiple contour levels in Figure 2 and Figure 2 - figure supplement 1. The contour levels relative to the root-mean-square deviation (RMSD) were calculated from the final map values. Final map contains the volume for the entire RNC including the I27 domain.

References

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181(4096):223-230.
2. Onuchic JN, Luthey-Schulten Z, & Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Annual review of physical chemistry* 48(1):545-600.
3. Dill KA & Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* 4(1):10-19.
4. Tycko R (2004) Progress towards a molecular-level structural understanding of amyloid fibrils. *Curr Opin Struct Biol* 14(1):96-103.
5. Clark PL (2004) Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci* 29(10):527-534.
6. Dobson CM (2001) Protein folding and its links with human disease. *Biochem Soc Symp* (68):1-26.
7. Gloge F, Becker AH, Kramer G, & Bukau B (2014) Co-translational mechanisms of protein maturation. *Curr Opin Struct Biol* 24:24-33.

8. Holtkamp W, *et al.* (2015) Cotranslational protein folding on the ribosome monitored in real time. *Science* 350(6264):1104-1107.
9. Thommen M, Holtkamp W, & Rodnina MV (2017) Co-translational protein folding: progress and methods. *Curr Opin Struct Biol* 42:83-89.
10. Wruck F, Katranidis A, Nierhaus KH, Buldt G, & Hegner M (2017) Translation and folding of single proteins in real time. *Proc Natl Acad Sci U S A* 114(22):E4399-E4407.
11. Jacobs WM & Shakhnovich EI (2017) Evidence of evolutionary selection for co-translational folding. *arXiv preprint arXiv:1703.10948*.
12. Krobath H, Shakhnovich EI, & Faísca PF (2013) Structural and energetic determinants of co-translational folding. *The Journal of chemical physics* 138(21):06B601_601.
13. Tanaka T, Hori N, & Takada S (2015) How co-translational folding of multi-domain protein is affected by elongation schedule: molecular simulations. *PLoS computational biology* 11(7):e1004356.
14. Nilsson OB, *et al.* (2015) Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep* 12(10):1533-1540.
15. Nilsson OB, *et al.* (2017) Cotranslational folding of spectrin domains via partially structured states. *Nat Struct Mol Biol* 24(3):221-225.
16. Nilsson OB, Muller-Lucks A, Kramer G, Bukau B, & von Heijne G (2016) Trigger Factor Reduces the Force Exerted on the Nascent Chain by a Cotranslationally Folding Protein. *J Mol Biol* 428(6):1356-1364.
17. Clark PL & King J (2001) A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early in vitro refolding intermediates. *J Biol Chem* 276(27):25411-25420.
18. Kaiser CM, Goldman DH, Chodera JD, Tinoco I, Jr., & Bustamante C (2011) The ribosome modulates nascent protein folding. *Science* 334(6063):1723-1727.
19. Javed A, Christodoulou J, Cabrita LD, & Orlova EV (2017) The ribosome and its role in protein folding: looking through a magnifying glass. *Acta Crystallogr D Struct Biol* 73(Pt 6):509-521.
20. Rodnina MV (2016) The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci* 25(8):1390-1406.
21. Knight AM, *et al.* (2013) Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem Biol* 8(6):1195-1204.
22. Cabrita LD, *et al.* (2016) A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol* 23(4):278-285.
23. Fedorov AN & Baldwin TO (1995) Contribution of cotranslational folding to the rate of formation of native protein structure. *Proc Natl Acad Sci U S A* 92(4):1227-1231.
24. Nicola AV, Chen W, & Helenius A (1999) Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat Cell Biol* 1(6):341-345.
25. Ugrinov KG & Clark PL (2010) Cotranslational folding increases GFP folding yield. *Biophys J* 98(7):1312-1320.
26. Evans MS, Clarke TF, & Clark PL (2005) Conformations of co-translational folding intermediates. *Protein Pept Lett* 12(2):189-195.
27. Goldman DH, *et al.* (2015) Mechanical force releases nascent chain-mediated ribosome arrest *in vitro* and *in vivo*. *Science* 348(6233):457-460.
28. Ismail N, Hedman R, Schiller N, & von Heijne G (2012) A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat Struct Mol Biol* 19(10):1018-1022.

29. Kramer G, *et al.* (2002) L23 protein functions as a chaperone docking site on the ribosome. *Nature* 419(6903):171-174.
30. Gong F & Yanofsky C (2003) A transcriptional pause synchronizes translation with transcription in the tryptophanase operon leader region. *J Bacteriol* 185(21):6472-6476.
31. Seidelt B, *et al.* (2009) Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* 326(5958):1412-1415.
32. Bischoff L, Berninghausen O, & Beckmann R (2014) Molecular basis for the ribosome functioning as an L-tryptophan sensor. *Cell Rep* 9(2):469-475.
33. Improta S, Politou AS, & Pastore A (1996) Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* 4(3):323-337.
34. Dunkle JA, Xiong L, Mankin AS, & Cate JH (2010) Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc Natl Acad Sci U S A* 107(40):17152-17157.
35. Karanicolas J & Brooks CL, 3rd (2003) Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J Mol Biol* 334(2):309-325.
36. Karanicolas J & Brooks CL, 3rd (2003) The importance of explicit chain representation in protein folding models: an examination of Ising-like models. *Proteins* 53(3):740-747.
37. Karanicolas J & Brooks CL, 3rd (2003) The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design? *Proc Natl Acad Sci U S A* 100(7):3954-3959.
38. O'Brien EP, Christodoulou J, Vendruscolo M, & Dobson CM (2011) New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc* 133(3):513-526.
39. Bell GI (1978) Models for the specific adhesion of cells to cells. *Science* 200(4342):618-627.
40. Fowler SB & Clarke J (2001) Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. *Structure* 9(5):355-366.
41. Fowler SB, *et al.* (2002) Mechanical unfolding of a titin Ig domain: structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J Mol Biol* 322(4):841-849.
42. Best RB & Hummer G (2016) Microscopic interpretation of folding ϕ-values using the transition path ensemble. *Proceedings of the National Academy of Sciences* 113(12):3263-3268.
43. Sánchez IE & Kiefhaber T (2003) Origin of unusual φ-values in protein folding: evidence against specific nucleation sites. *Journal of molecular biology* 334(5):1077-1085.
44. Best RB, Hummer G, & Eaton WA (2013) Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci U S A* 110(44):17874-17879.
45. Best RB, *et al.* (2003) Mechanical unfolding of a titin Ig domain: structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *J Mol Biol* 330(4):867-877.
46. Scott KA, Steward A, Fowler SB, & Clarke J (2002) Titin; a multidomain protein that behaves as the sum of its parts. *J Mol Biol* 315(4):819-829.
47. Williams PM, *et al.* (2003) Hidden complexity in the mechanical properties of titin. *Nature* 422(6930):446-449.

48. Wright CF, Lindorff-Larsen K, Randles LG, & Clarke J (2003) Parallel protein-unfolding pathways revealed and mapped. *Nat Struct Biol* 10(8):658-662.
49. Wright CF, Steward A, & Clarke J (2004) Thermodynamic characterisation of two transition states along parallel protein folding pathways. *J Mol Biol* 338(3):445-451.
50. Borgia MB, Nickson AA, Clarke J, & Hounslow MJ (2013) A mechanistic model for amorphous protein aggregation of immunoglobulin-like domains. *J Am Chem Soc* 135(17):6456-6464.
51. Botello E*, et al.* (2009) Temperature and chemical denaturant dependence of forced unfolding of titin I27. *J Phys Chem B* 113(31):10845-10848.
52. Chen H*, et al.* (2015) Dynamics of equilibrium folding and unfolding transitions of titin immunoglobulin domain under constant forces. *J Am Chem Soc* 137(10):3540-3546.
53. Lu H, Isralewitz B, Krammer A, Vogel V, & Schulten K (1998) Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys J* 75(2):662-671.
54. Nunes JM, Mayer-Hartl M, Hartl FU, & Muller DJ (2015) Action of the Hsp70 chaperone system observed with single proteins. *Nat Commun* 6:6307.
55. Yagawa K*, et al.* (2010) Structural basis for unfolding pathway-dependent stability of proteins: vectorial unfolding versus global unfolding. *Protein Sci* 19(4):693-702.
56. Su T*, et al.* (2017) The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. *eLife* 6:e25642.
57. Ziv G, Haran G, & Thirumalai D (2005) Ribosome exit tunnel can entropically stabilize α-helices. *Proceedings of the National Academy of Sciences of the United States of America* 102(52):18956-18961.
58. Mirzadeh K*, et al.* (2015) Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector-Coding Sequence Junction. *ACS Synth Biol* 4(9):959-965.
59. Li X*, et al.* (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10(6):584-590.
60. Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519-530.
61. Zhang K (2016) Gctf: Real-time CTF determination and correction. *J Struct Biol* 193(1):1-12.
62. Kucukelbir A, Sigworth FJ, & Tagare HD (2014) Quantifying the local resolution of cryo-EM density maps. *Nat Methods* 11(1):63-65.
63. Pettersen EF*, et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.
64. Hess B, Kutzner C, Van Der Spoel D, & Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory and Comput* 4(3):435-447.
65. Hess B, Bekker H, Berendsen HJ, & Fraaije JG (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18(12):1463-1472.

Supplementary Figures

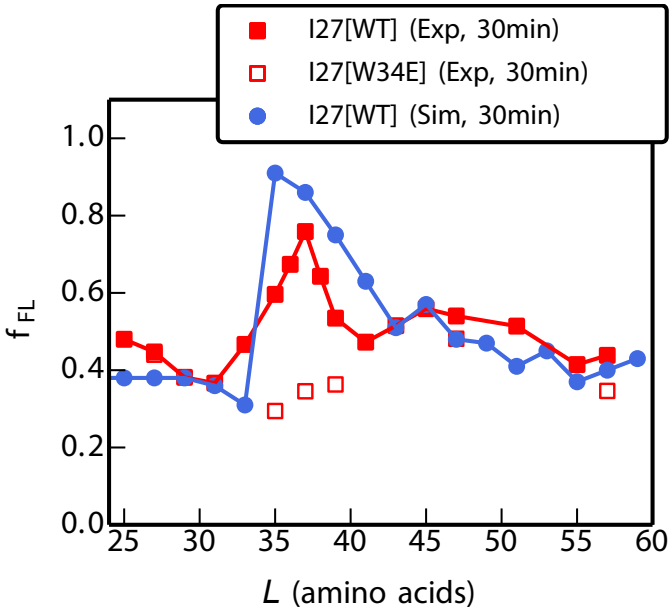


Figure 1 - figure supplement 1. Experimental (red) and simulated (blue) profiles of fraction full length protein, f_{FL}, obtained with a 30 min incubation. Note the higher background values compared to main text Figures 1 and 3D.

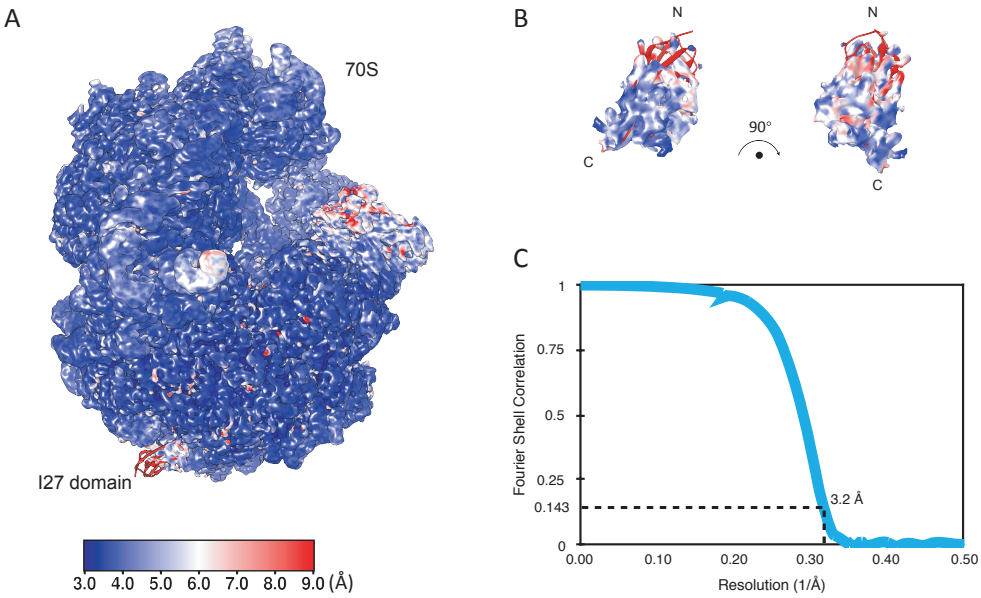
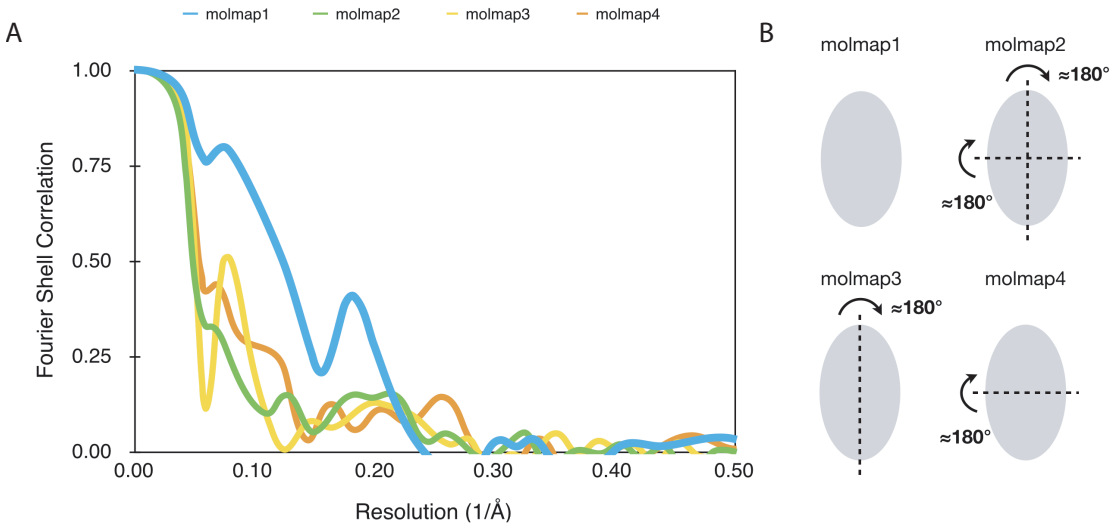


Figure 2 - figure supplement 1. Resolution of the ribosome-nascent chain complex (RNC). (A) Calculation of the local resolution using Resmap (Kucukelbir, A. et al. Nat Methods 11, 63-65, 2014). The RNC density is displayed at 1.7 RMSD. (B) local resolution of the I27 domain. The I27 domain density is displayed at 2 RMSD. N and C termini are indicated. (C) Fourier-shell correlation (FSC) curve of the refined final map of the RNC, indicating the average resolution of 3.2 Å (at 0.143).



764

765 **Figure 2 - figure supplement 2.** Validation of model orientation for I27 domain. To validate
766 the orientation of the I27 domain model (PDB: 1TIT) to its corresponding density, three other
767 possible orientations were tested. (A) The Fourier-shell correlations between the isolated I27
768 density and the map generated from the model of the final orientation (molmap1, blue) and
769 the models fitted with the other three possibilities (molmap2, green; molmap3, yellow;
770 molmap4, orange) were plotted. (B) The illustration showing the relationship among the four
771 model orientations. Major and minor axes were used as the principle of finding possible
772 orientations for the overall ellipsoid shaped map and model.

773

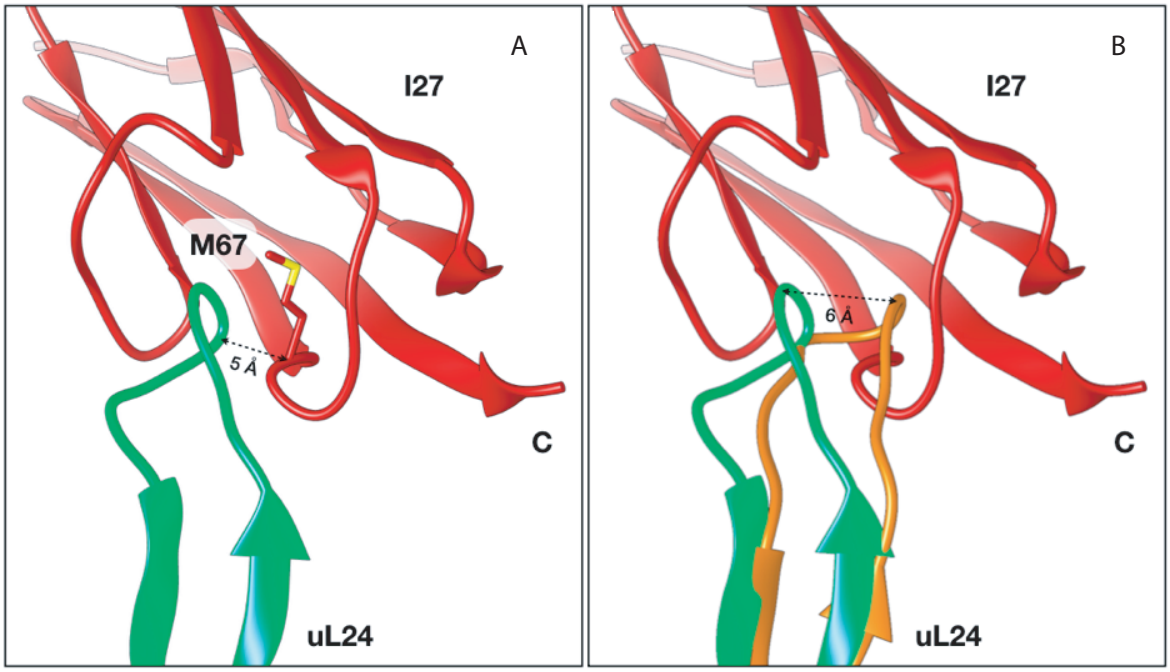
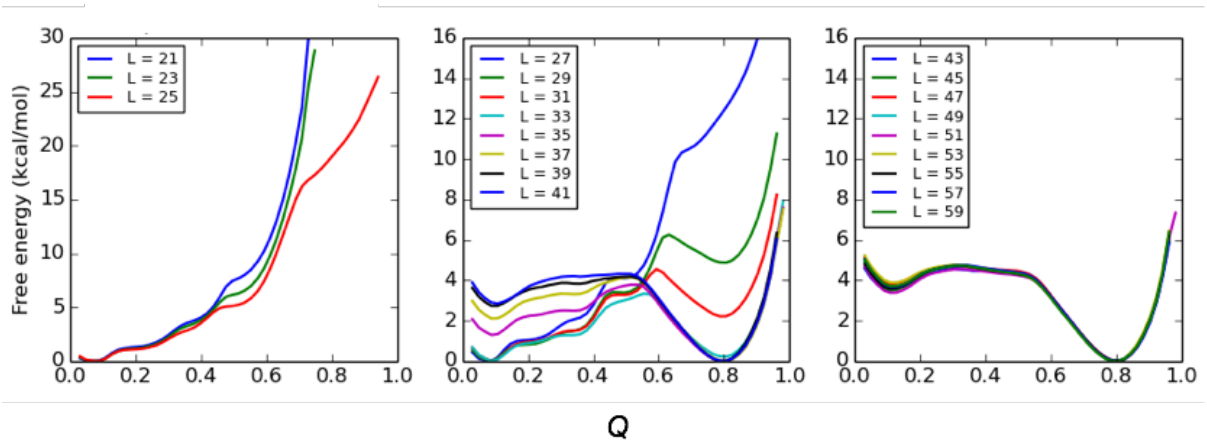


Figure 2 - figure supplement 3. The I27 domain and a β hairpin in ribosomal protein uL24 close to the ribosomal exit tunnel. (A) Residue M67 in the I27 domain is located in close proximity to a β hairpin loop in uL24 in the cryo-EM structure of I27-TnaC[*L*=35] RNCs. (B) The uL24 β hairpin in the I27-RNC (light green; re-modeled based on PDB: 5NWY) is ~ 6 Å shifted (distance measured via the backbone of Pro50) compared to its location in the VemP-RNC (orange; PDB: 5NWY) and the TnaC-RNC (PDB: 4UY8, not shown). C represents the C terminus of the I27 domain.



786

787 **Figure 3 - figure supplement 1**. Simulation free energy *F*(*Q*) projected on the fraction of

788 native contacts, *Q,* for I27 folding with different linker lengths (as indicated in legend) at

789 291K.

790

791