1

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

3

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#], Jane Clarke[2#] and Robert B. Best[1#]

7

8

[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.

19

20

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

[#]Corresponding authors

23

Classification: Biological Sciences, Biophysics and Computational Biology

25

26

27

28

29 **Abstract**

30

31 Proteins that fold cotranslationally may do so in a restricted configurational space, due to the

32 volume occupied by the ribosome. How does this environment, coupled with the close

33 proximity of the ribosome, affect the folding pathway of a protein? Previous studies have

34 shown that the cotranslational folding process for many proteins, including small, single

35 domains, is directly affected by the ribosome. Here, we investigate the cotranslational folding

36 of an all-β immunoglobulin domain, titin I27. Using an arrest peptide-based assay and

37 structural studies by cryo-EM, we show that I27 folds in the mouth of the ribosome exit

38 tunnel. Simulations that use a kinetic model for the force-dependence of escape from arrest,

39 accurately predict the fraction of folded protein as a function of length. We used these

40 simulations to probe the folding pathway on and off the ribosome. Our simulations - which

41 also reproduce experiments on mutant forms of I27 - show that I27 folds, while still

42 sequestered in the mouth of the ribosome exit tunnel, by essentially the same pathway as free

43 I27, with only subtle shifts of critical contacts from the C to the N terminus.

44

45

46 **Significance Statement**

47 Most proteins need to fold into a specific three-dimensional structure in order to function.

48 The mechanism by which isolated proteins fold has been thoroughly studied by experiment

49 and theory. However, in the cell proteins do not fold in isolation, but are synthesized as linear

50 chains by the ribosome during translation. It is therefore natural to ask at which point during

51 synthesis proteins fold, and whether this differs from the folding of isolated protein

52 molecules. By studying folding of a well characterized protein domain, titin I27, stalled at

53 different points during translation, we show that it already folds in the mouth of the ribosome

54 exit tunnel, and that the mechanism is almost identical to that of the isolated protein.

55

Introduction

57   To what extent is the cotranslational folding pathway of a protein influenced by the presence

58   of the ribosome and by the vectorial emergence of the polypeptide chain during translation?

59   Recent studies have shown that small proteins can fold inside the ribosome exit tunnel (e.g.,

60   the small zinc finger domain ADR1a) (1), while other proteins can fold at the mouth of the

61   tunnel (e.g., the three-helix bundle spectrin domains) (2); however some proteins may be

62   simply too large to fold within the confines of the ribosome (e.g., DHFR) (3). The nature of

63   cotranslational protein folding is determined by a number of biophysical factors, including

64   the folding properties of the isolated protein (4-9), together with the effects the ribosome

65   itself may have on the folding process (10-16). Due to the spatial constraints imposed upon

66   the nascent chain by the confines of the tunnel, and effects due to the close proximity of the

67   ribosome itself, the ribosome has been shown to influence directly the cotranslational folding

68   of small proteins and single domains: The stability of folded or partly folded states may be

69   reduced when folding occurs close to, or within the confines of, the ribosome (17, 18); the

70   folding kinetics are expected to be correspondingly altered, with the rate of folding likely to

71   be decreased and the unfolding rate increased, in close proximity to the ribosome(18).

72   Interactions of the folded state or nascent polypeptide with the ribosome may also be either

73   stabilising or destabilising (19, 20). Since translation is vectorial in nature, it is possible that

74   when proteins fold cotranslationally they fold via different pathways than those used when

75   proteins fold outside the ribosome, or when isolated proteins fold *in vitro* (2, 11, 21-24).

76   However, addressing these issues is challenging, because standard protein folding methods

77   are not directly applicable to cotranslational folding.

78   The folding of the protein close to the ribosome generates a pulling force on the

79   nascent chain. This force has been probed by single molecule (25) as well as arrest peptide

80   (AP) experiments (1-3). In this work, we use such arrest peptide-based cotranslational force-

81   measurement experiments, simulations, and structural studies to investigate how the

82   ribosome affects the folding of titin I27, a small all-β immunoglobulin domain with a

83   complex greek-key fold; the stability, kinetics and folding pathway of I27 has been

84   extensively characterized in previous studies of the isolated domain (26, 27). In this study we

85   investigate whether I27 can begin to fold in the confines of the ribosome, and if the folding

86   pathway observed in the isolated domain is conserved during cotranslational folding. Results

87   from all three techniques show that I27 folds in the mouth of the ribosome exit tunnel; our

88   simulations correctly capture the onset of folding in I27 and three mutant variants, allowing

89   us to predict how destabilisation of regions that fold early and late in the isolated domain

90   affect folding on the ribosome. Our simulations further show that the folding pathway of I27

91   is largely unaffected by the presence of the ribosome, except for small but significant changes

92   observed for contacts near the N and C termini.

93   **Results**

94   **I27 folds close to the ribosome**

95   In order to gain insight into when I27 can commence folding on the ribosome, we employed

96   an arrest peptide force-measurement assay (28) carried out using the PURE *in vitro*

97   translation system, as described in (1-3). In these experiments, the *E. coli* SecM arrest peptide

98   (AP) is used to stall the nascent protein chain temporarily during translation. The yield of

99   full-length protein which escapes stalling in a defined time interval (f_{FL}), determined from

100   SDS-PAGE gels, provides a proxy for the pulling force exerted on the nascent chain by the

101   protein as it folds (1-3) (Figure 1A). By measuring f_{FL} for a set of constructs where the length

102   L of the linker between the target protein and the SecM AP is systematically varied, a force

103   profile can be recorded that reflects the points during translation where the folding process

104   starts and ends. Previous work has shown that the location of the main peak in a force profile

105   correlates with the acquisition of protease resistance in an on-ribosome pulse-proteolysis

106   assay (17, 29) and that the amplitude of the main force peak correlates with the

107   thermodynamic stability of the protein (29, 30), indicating that the main peak represents a

108   *bona fide* folding event rather than, *e.g.*, the formation of a molten globule. The sharp onset

109   of the main force peak observed for most proteins analysed thus far (29) is also as expected

110   for a cooperative folding event.

111   The force profile for wild-type I27 (Figure 1B) has a distinct peak at L = 35-38

112   residues (see Methods for sequences of the constructs). This peak is absent from the force

113   profile for the mutant I27[W34E], a non-folding variant of I27, demonstrating that the peak is

114   due to a folding event and not, for example, to non-specific interactions of the unfolded

115   nascent chain with the ribosome. The non-zero f_{FL} for the non-folding mutant is attributed to

116   the spontaneous rate of escape from arrest in the absence of acceleration by forces associated

117   with folding. Since it takes ~35 residues in an extended conformation to span the ~100 Å

118   long exit tunnel (31), the critical length $L \approx 35$ residues suggests that I27 starts folding while

119   in mouth of the exit tunnel.

4

120
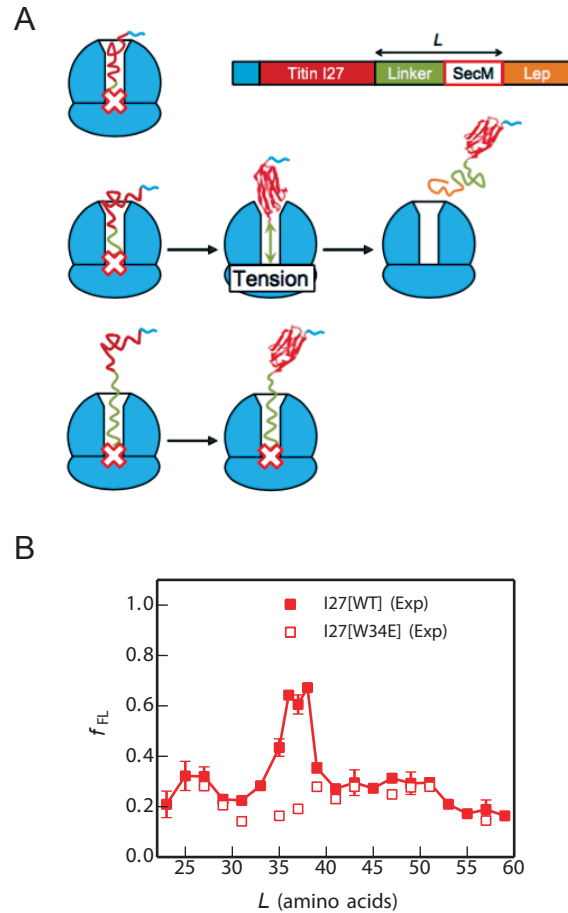
Figure 1. Cotranslational folding of the titin I27 domain by force-profile analysis. (A) The force-measurement assay (modified from (2)). I27, preceded by a His-tag, is placed L residues away from the last amino acid of the SecM AP, which in turn is followed by a 23-residue C-terminal tail derived from *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.

136

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 (32-34) and purified stalled ribosome-nascent chain complexes (RNCs) carrying an N-terminally His-tagged I27[L=35]

141   construct (see Methods). The construct was expressed in *E. coli*, RNCs were purified using

142   the N-terminal His-tag, and an RNC structure with an average resolution of 3.2 Å (SI

143   Appendix, Fig. S1) was obtained by cryo-EM. In addition to the density corresponding to the

144   TnaC AP, a well-defined globular density (~4.5-9 Å resolution) was visible protruding from

145   the exit tunnel (Figure 2A). Given the flat ellipsoidal shapes of the protruding density and of

146   the I27 structure, there is only one way to fit the NMR structure of I27 (PDB 1TIT (35)) that

147   gives a good Fourier-shell correlation between the isolated I27 density and the map generated

148   from the I27 PDB model (SI Appendix, Fig. S2). In the fitted model, the C-terminal end of

149   I27 extends into the exit tunnel and a β-hairpin loop on ribosomal protein uL24 is lodged in a

150   cavity in I27 (Figure 2B and Supporting Video S1). The I27 domain further packs against

151   ribosomal protein uL29 and ribosomal 23S RNA (Figure 2C), as if it is being pulled tight

152   against the ribosome by the nascent chain. We conclude that the peak at $L = 35-38$ residues in

153   the force profile indeed represents the cotranslational folding of the I27 domain at the tunnel

154   exit.

155
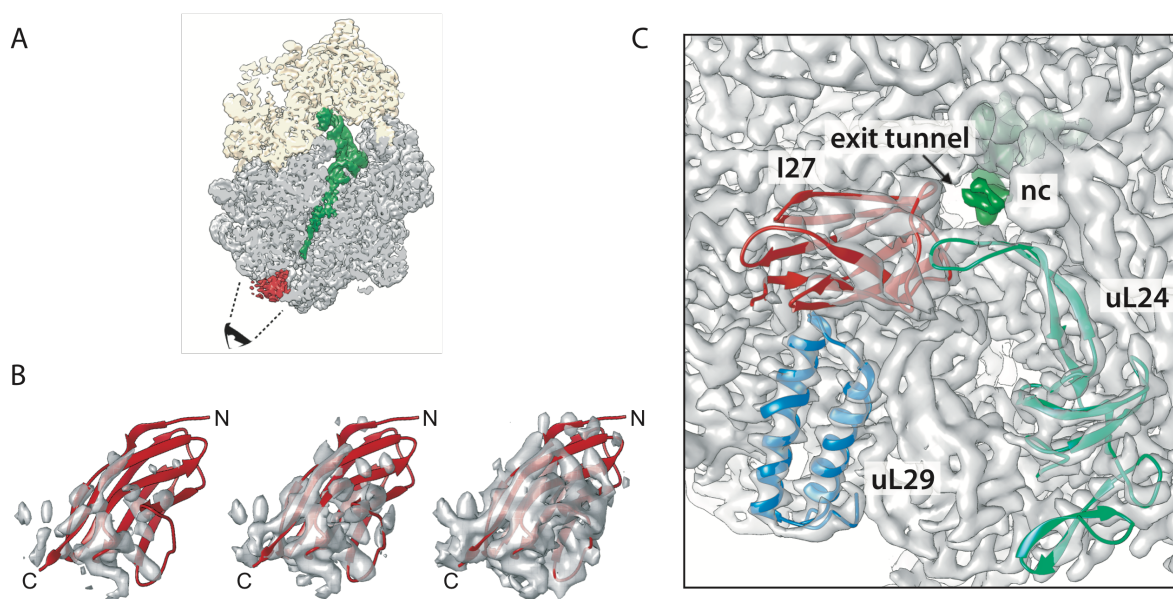
Figure 2. Cryo-EM structure of I27[L=35] RNCs. (A) Cryo-EM reconstruction of the I27–
TnaC[L = 35] RNC. The ribosomal small subunit is shown in yellow, the large subunit in
grey, the peptidyl-tRNA with the nascent chain in green, and an additional density
corresponding to I27 at the ribosome tunnel exit in red. The black cartoon eye and dash lines
indicate the angle of view in 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

6

167   fitted I27 domain (red) are shown in cartoon mode; 23S RNA and proteins not contacting I27
168   are shown as density only. The density contour level is at 5 RMSD excluding tRNA, nascent
169   chain and I27 domain, which are displayed at 1.7 RMSD.

170

171   **Coarse-grained molecular dynamics simulations recapitulate I27 folding on the**

172   **ribosome**

173   The yield of folded protein in arrest peptide experiments has been used as a proxy for the

174   pulling forces that are exerted on the nascent chain at different points during translation in all

175   studies to date (1, 2, 29). Here, to further elucidate the molecular origins of these forces and

176   provide a quantitative interpretation of the observed folding yield of I27, we have calculated

177   force profiles based on coarse-grained MD simulations (see Methods). Briefly, in the MD

178   model, the 50S subunit of the *E. coli* ribosome (36) (PDB 3OFR) and the nascent chain are

179   explicitly represented using one bead at the position of the Cα atom per amino acid, and three

180   beads (for P, C4', N3) per RNA base (Figure 3A). The interactions within the protein were

181   given by a standard structure-based model (37-39), which allowed it to fold and unfold.

182   Interactions between the protein and ribosome beads were purely repulsive (40) and the

183   ribosome beads were fixed in space, as in previous simulation studies (18). I27 was

184   covalently attached to unstructured linkers having the same sequences as those used in the

185   force-profile experiments (Figure 3B) and the C terminus of the linker was tethered to the last

186   P atom in the A-site tRNA (41) with a harmonic potential, allowing the force exerted by the

187   folding protein to be directly measured. The potential chosen was stiff enough that

188   displacements caused by typical pulling forces were smaller than 1 Å. For each linker length

189   *L*, we used umbrella sampling to determine the average force exerted on the AP by the

190   protein in the folded and unfolded states while arrested, as well as the populations of those

191   two states (Figure 3C). We also estimated the folding and unfolding rates directly from

192   folding/unfolding simulations.

193
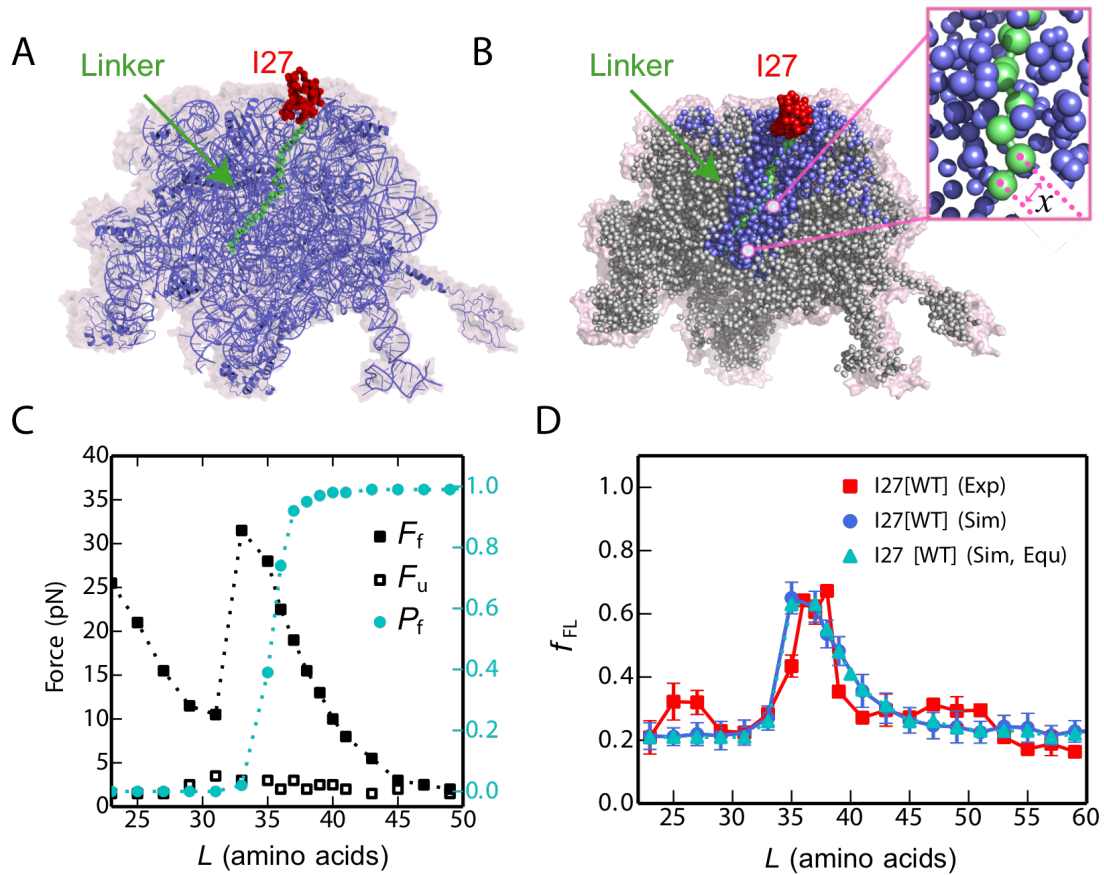
Figure 3. MD simulations of cotranslational folding of I27. (A) 50S subunit of the *E. coli* ribosome (PDB 3OFR) with I27[L=35] attached via an unstructured linker. (B) Coarse-grained model for I27 (red) and linker (green), with surrounding ribosomal pseudo-atoms in blue. Pseudo-atoms with grey colour 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, empty symbols) and folded state (F_f, filled symbols) of I27 at different linker lengths L. The average fraction folded I27 for different L, P_f, is shown in cyan on the right axis. Free energy profiles at each linker length are shown in SI Appendix, Fig. S4. (D) Experimental (red square) force profiles for cotranslational folding of I27. Force profiles calculated from simulations using full kinetic scheme or pre-equilibrium model are shown in blue circle and cyan triangle respectively. The RMSD of the f_{FL} between experiment and simulation is 0.08.

Given the experimentally-determined force-dependence of the escape rate $k(F)$ (25), here approximated by a Bell-like model (42), we can calculate 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 from a kinetic model, as described in Methods. The calculated f_{FL} profile for I27 is shown in Figure 3D (see also SI

215 Appendix, Fig. S5) for the full solution of the kinetic model, as well as for an approximation

216 in which the folding and unfolding rates are assumed to be faster than the escape rate ("pre-

217 equilibrium"). Both results are very consistent with each other, as well as with the

218 experimental profile. The peak in the folding yield arises as consequence of two opposing

219 effects, the force exerted by the folding protein and population of the folded state, which

220 respectively decrease and increase as the linker length increases. In the simulations with the

221 I27[L=35] construct, the folded I27 domain is seen to occupy positions that largely overlap

222 with the cryo-EM structure (Supporting Video S2). Overall, these results suggest that the MD

223 model provides a good representation of the folding behaviour of the I27 domain in the

224 ribosome exit tunnel. To show that the simulation model is not specific to I27, we have also

225 applied it to another two proteins with different topologies for which experimental force

226 profiles have been recorded, Spectrin R16 (all-α fold) and S6 (α/β fold) (2, 29). In these

227 cases, we also obtain force profiles similar to experiment (SI Appendix, Fig. S6 and S7).

228

Force profiles of I27 variants probe the folding pathway

230 To test whether the cotranslational folding pathway is the same as that observed for the

231 isolated I27 domain *in vitro*, we investigated three destabilised variants of I27, both by

232 simulation and experiment. One mutation in the core, Leu 58 to Ala (L58A), located in β-

233 strand E (Figure 4A) destabilizes the protein by 3.2 kcal mol^{-1}, and removes interactions that

234 form early during folding of the isolated domain, playing a key role in formation of the

235 folding nucleus (ϕ-value = 0.8) (26). Two further mutations, M67A and deletion of the N-

236 terminal A-strand, remove interactions that form late in the folding of I27 (*i.e.*, both mutants

237 have low ϕ-values (26, 27)). The A-strand is the first part of I27 to emerge from the

238 ribosome, while M67 is located in a part of I27 that is shown by cryo-EM to be located in

239 very close proximity to a β hairpin loop of ribosomal protein uL24 in I27-TnaC[L=35] RNCs

240 (SI Appendix Fig. S3A). The interaction with the I27 domain shifts the tip of this uL24

241 hairpin by about 6 Å compared to its location in other RNC structures (SI Appendix Fig.

242 S3B).
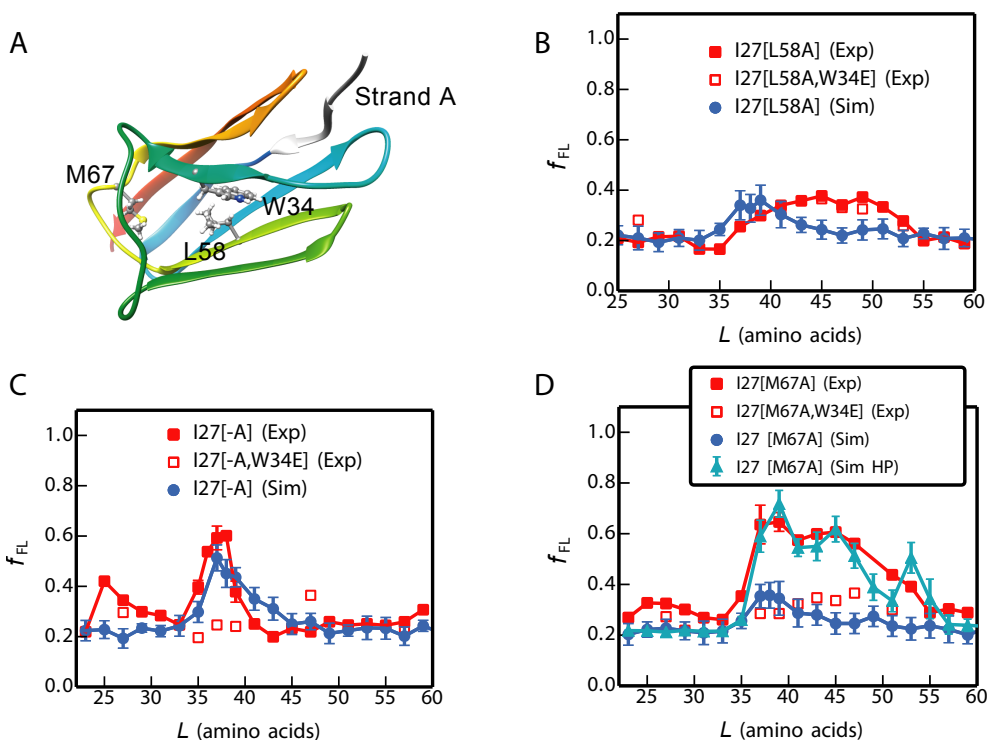
Figure 4. Simulations capture the experimental force profiles for mutant I27 domains. (A) Mutated residues in I27 (sticks). (B-D) Experimental (red) force profiles and calculated ones from full kinetic scheme (blue) for (B) I27[L58A], (C) A-strand deletion mutant I27[-A], (D) I27[M67A]. I27[M67A] (Sim HP) represents a simulation in which hydrophobic interactions between I27[M67A] and ribosome proteins uL23/uL29 are included. Experimental force profiles for non-folding mutants that contain an additional W34E mutation are shown as red open squares. The RMSD of the f_{FL} between experiment and simulation for I27[L58A], I27[-A] are 0.07 and 0.08 respectively. For I27[M67A], the f_{FL} RMSD is 0.07 between experiment and simulation (Sim HP).

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). Non-folding control experiments

10

266   for variants I27[-A,W34E], and I27[M67A,W34E] (Figure 4C and D) show that the peaks in

267   the force profiles for these variants are due to a folding event. These results show that

268   deletion of the A-strand and destabilisation of the E-F loop do not affect the onset of

269   cotranslational folding of I27, but that the M67A mutation increases the width of the folding

270   transition. The simulation model used for the other mutants does not predict such a broad

271   peak, suggesting that it may be necessary to include additional factors to reproduce the data

272   for M67A. One possibility which may explain the result would be favourable interactions of

273   the folded M67A mutant with the ribosome surface. The ribosomal surface proteins uL23 and

274   uL29 have been suggested to form a potential interaction site for nascent proteins such as

275   trigger factor (43), signal recognition particle (44) and SecYE (45). Here we have explored

276   the hypothesis that the broad force peak of mutant M67A might due to interactions between

277   an exposed hydrophobic cavity on I27[M67A] resulting from the mutation, and hydrophobic

278   surface residues of ribosomal proteins uL23 and uL29. By introducing such interactions into

279   the model, we are able to obtain a broad peak in the force profile very similar to that seen in

280   experiment (Figure 4D).

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

282   To compare the folding pathways when the protein is folding near the tunnel exit or outside

283   the ribosome, we estimated ϕ-values based on the transition paths of I27 folding on the

284   ribosome from our coarse-grained simulations, using a method introduced previously(46).

285   The transition paths are those regions of the trajectory where the protein crosses the folding

286   barrier, here defined as crossing between $Q = 0.3$ and $Q = 0.7$. For each linker length, 30

287   transition paths were collected from MD simulations. To reduce the uncertainty in the

288   experimental reference data, we only compared with experimental ϕ-values if the change in

289   folding stability between the mutant and the wild type is sufficiently large ($|\Delta\Delta G| > 7$ kJ/mol)

290   (47). As seen in Figure 5A, when the linker is long ($L = 51$ residues) and I27 is allowed to

291   fold outside the ribosome, the calculated ϕ-values are consistent (Spearman correlation

292   r=0.80) with the experimental values obtained for the folding of isolated I27 *in vitro* (26). For

293   shorter linker lengths ($L = 31$ and 35 residues), calculated ϕ-values remain largely unchanged

294   except for a slight increase near the N terminus (around residues 3-6) and a slight decrease

295   near the C terminus (around residues 72-74) (Figure 5B and C).

11

296
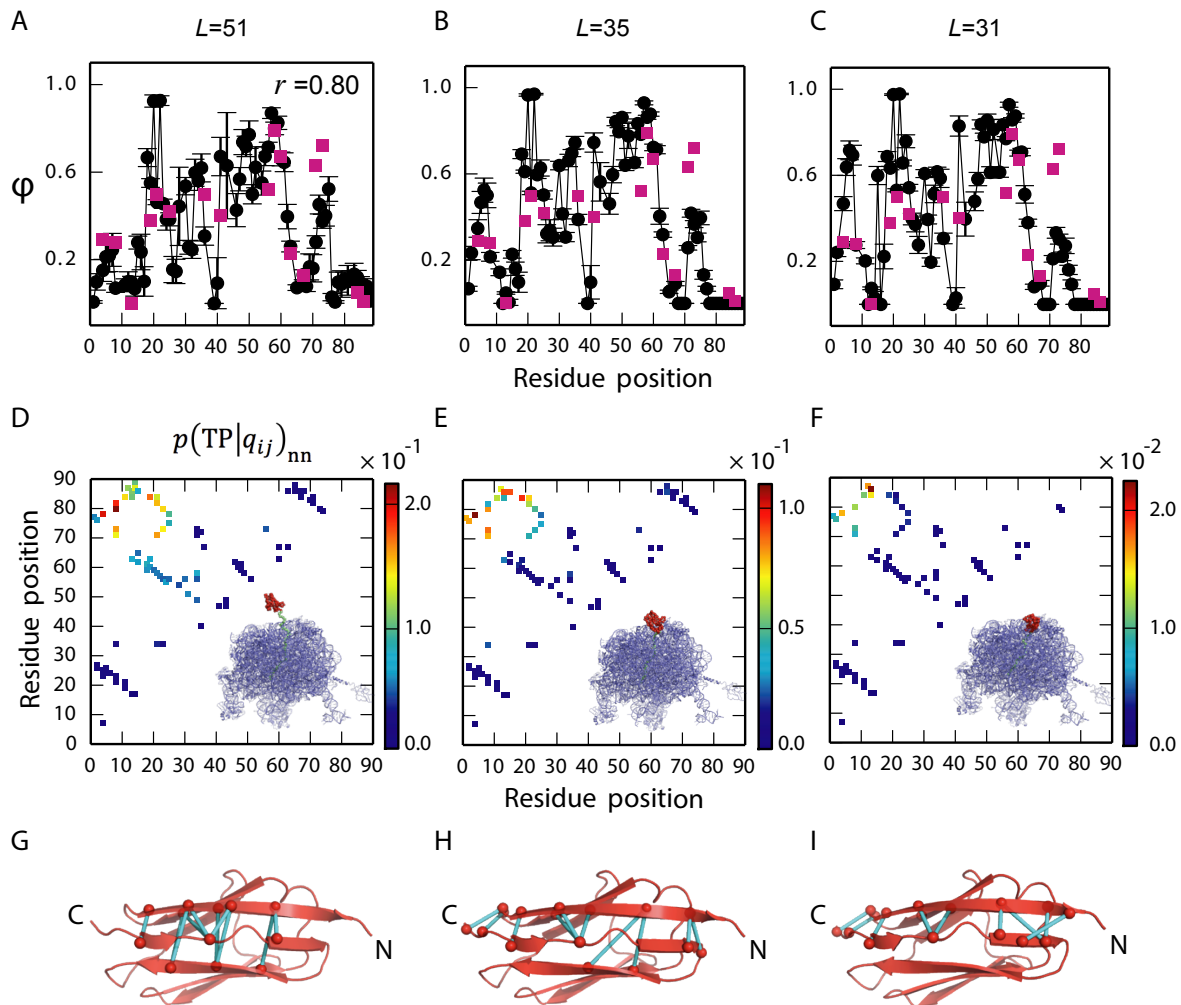
Figure 5. Simulated folding pathways for ribosome-tethered I27. LH column, $L=51$; middle column, $L=35$; RH column, $L=31$. Top panels: Simulated ϕ-values for I27 (black). ϕ-values determined by *in vitro* folding of purified I27 are shown as red squares. At $L=51$ the simulated ϕ-values match well with experiment (spearman correlation $r=0.80$). 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, $p(\text{TP}|q_{ij})_{\text{nn}}$. 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.

309

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,

$p(\text{TP}|q_{ij})_{\text{nn}}$ (48). This quantity indicates which native contacts are most important for

314 determining a successful folding event. $p(\text{TP}|q_{ij})_{\text{nn}}$ is closely related to the frequency of the

315 contact q_{ij} on transition paths $p(q_{ij}|\text{TP})$, but is effectively normalized by the probability that

316 the contact is formed in non-native states $p(q_{ij})_{\text{nn}}$, and can be expressed as:

$$317 \qquad p(\text{TP}|q_{ij})_{\text{nn}} = \frac{p(q_{ij}|\text{TP})p(\text{TP})_{\text{nn}}}{p(q_{ij})_{\text{nn}}} \qquad [1]$$

318 where $p(\text{TP})_{\text{nn}}$ is the fraction of non-native states which are on transition paths at

319 equilibrium. The subscript nn means that only the non-native segments of a trajectory are

320 included, *i.e.*, unfolded states and transition paths; the native, folded state is not included in

321 the calculation since native contacts are always formed in this state. The simulations suggest

322 that formation of native contacts between the N and C termini is somewhat more important

323 when folding takes place in the mouth of the exit tunnel ($L = 31$ residues) than far outside the

324 ribosome ($L = 51$ residues) (Figure 5D-F, upper left-hand corner in the panels). This is likely

325 due to the greater difficulty of forming these contacts (examples are shown in Figure 5G-I)

326 under ribosomal confinement; therefore, forming them becomes more critical in enabling the

327 protein to fold.

328

Discussion

330 Using a combination of MD simulation, force-profile measurements and cryo-EM, we have

331 investigated the cotranslational folding pathway of the 89-residue titin I27 domain. I27 has

332 been extensively characterised in previous *in vitro* folding studies (26, 27, 49-60). Results

333 from all three techniques show that wild-type I27 folds in the mouth of the ribosome exit

334 tunnel; in the cryo-EM structure of I27-TnaC[L=35] RNCs, I27 packs against ribosomal

335 proteins uL24, uL29, and ribosomal 23S RNA. This is in apparent contrast to a previous

336 NMR study on another Ig-like protein, in which the domain was shown to acquire its native

337 fold (as reflected in the NMR spectrum) only when fully outside the ribosome tunnel, at L =

338 42-47 residues linker length (20).

339   In order to determine the molecular origin of the measured force profile, we

340 performed molecular dynamics simulations of I27 folding on the ribosome, varying the

341 length of the linker sequence between the arrest peptide and the I27 domain. We calculated

342 the pulling force directly from the simulations and translated this into yield of folded protein

343 using a kinetic model parameterized based on known release kinetics of the SecM AP. This

344 enabled us to recapitulate the experimental arrest peptide force measurement profile, and

345   therefore relate f_{FL} directly to the force exerted on the arrest peptide. Our simulations

346   demonstrate the direct effect that the restoring force of the nascent chain can have on

347   determining when the protein folds on the ribosome. We show that f_{FL} depends upon a

348   combination of the force exerted by the folded protein and the fraction of folded protein at

349   the given linker length L.

350   In order to relate how destabilization of regions that fold early and late in the isolated

351   domain affects folding on the ribosome, we used simulations to predict the onset of folding in

352   three mutant variants of I27. A previous ϕ-value analysis of I27 (26) showed that early

353   packing of the structurally central β-strands drives the folding of this domain, while

354   peripheral strands and loop regions pack later in the folding process. Mutations in the folding

355   core (such as L58A) slow folding, whereas mutations in the periphery have no effect on

356   folding rates (26). L58 is a key residue in the critical folding nucleus and almost fully packed

357   in the transition state, in isolated domain studies. The simulated and experimental force

358   profiles of I27 [L58A] show that this variant does not fold in or near the exit tunnel; hence,

359   destabilisation of the central folding core prevents folding close to the ribosome. Since

360   isolated I27[L58A] is fully folded, it is likely that this variant can only fold cotranslationally

361   at longer linker lengths, when it is no longer in close proximity to the ribosome and exerts

362   little force on the nascent chain.

363   Our experiments show that I27 variants destabilized in regions of the protein that are

364   unstructured, or only partially structured, in the transition state, are still able to commence

365   folding close to the ribosome. The force profiles reveal that the onset of folding of mutants

366   with the A-strand deleted, or with the Met 67 to Ala mutation in the E-F loop, is the same as

367   for wild-type although these have a similar destabilisation as L58A (Figure 4). The broader

368   peak observed experimentally for M67A is harder to interpret. A plausible explanation is that

369   the mutation introduces non-specific interactions of the folded domain with the ribosome

370   surface, and we have shown that incorporating such interactions into the simulations could

371   reproduce the results. An additional factor may be that that the mutation is in a region that

372   interacts closely with ribosomal protein uL24 in the wild-type cryo-EM structure (SI

373   Appendix Fig. S3).

374   Our simulations reproduce the onset of folding in the three mutant variants of I27 (Figure 4),

375   and so give us the confidence to investigate how confinement within the ribosome affects the

376   folding pathway of I27. We used simulations to investigate the folding of I27 arrested on the

377   ribosome at various linker lengths, using a Bayesian method for testing the importance of

378   specific contacts on the folding pathway, as well as by computing ϕ-values (Figure 5).

379   Overall, we find that the mechanism and pathway of folding are robust towards variation in

380   linker length and relatively insensitive to the presence of the ribosome; small but significant

381   changes are observed only for contacts near the N and C termini. These shifts are consistent

382   with the greater importance of forming N-terminal contacts when the C terminus is

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

384   terminus.

385   In our kinetic modelling, we found that we obtained similar results with or without the

386   assumption that folding and unfolding are fast relative to the escape rate, suggesting that this

387   "pre-equilibrium" assumption is justified, at least for this protein. The reason for its validity

388   in the case of I27 can be seen by comparing the folding and unfolding rates with the force-

389   dependent escape rate of $\sim 2.4 \times 10^{-3}$ s^{-1} obtained at the highest forces of ~20 pN (c.f., Fig.

390   3C). Folding and unfolding rates at different linker lengths can be obtained by combining the

391   linker-length dependence of the rates from simulation with the known folding/unfolding rates

392   for isolated I27 from experiment (SI Appendix, Fig. S9). The presence of the ribosome

393   increases the unfolding rate at shorter linker lengths so that it is faster than the maximum

394   escape rate, while not slowing the folding rate sufficiently for it to drop below the escape

395   rate. Note that the unfolding rate does drop below the maximum escape rate at larger linker

396   lengths, but by that point the folded population is already almost 100%, so the pre-

397   equilibrium assumption still gives accurate results. Although this assumption appears to be

398   justified in the case of I27, it is probably not true in general, and it will be interesting to

399   investigate for slower-folding proteins in future.

400   The arrest peptide experiments, in which a protein exerts a force due to folding in

401   some ways resemble atomic force microscopy or optical tweezer experiments in which an

402   external force is applied to the protein termini. It is important to note, however, that the

403   nature and effect of the forces exerted on the folding protein by tethering to the ribosome are

404   very different than is the case for pulling on both termini by an external force. For example,

405   forces of the magnitude seen in this work (up to ~20 pN) tend to have very little effect on the

406   unfolding rate when applied to the termini of I27, due to the similarity in extension of the

407   folded and transition states (61); by contrast folding rates are dramatically slowed, even by

408   very small forces, due to the large difference in extension of between unfolded and transition

409   states (56). The forces arising from tethering to the ribosome are due to the folding of the

410   protein itself rather than an external device. They arise from the constriction of available

411   configuration space, particularly for folded and partially folded states, as well as from any

412   attractive interactions between the protein and the ribosome. Our simulations suggest that for

413   I27, reducing the linker length speeds up unfolding and slows folding rates by similar factors.

414   Thus, it seems that comparisons to the effects of forces exerted by AFM and optical tweezer

415   experiments need to be performed with care.

416   We have previously shown that α-helical proteins can fold co-translationally (2),

417   perhaps unsurprising since helical structures are dominated by short-range interactions and

418   helices can form within the ribosome tunnel itself (62, 63). Here, our equilibrium arrest-

419   peptide assay and structural studies reveal that an all-β protein, titin I27, is able to fold within

420   the mouth of the ribosome exit tunnel, despite its folding being dominated by long-range

421   interactions. Molecular simulations, accounting for the effect of the entropic restoring force

422   on protein stability, reproduce the yield of protein from experiments remarkably well. These

423   simulations reveal that I27 folds on the ribosome by the same pathway as when the protein

424   folds away from the confines of the ribosome. We note that a similar conclusion has been

425   reached by Guinn et al. (64) for another small protein, src SH3, using a completely different

426   experimental approach which combines optical tweezer experiments and chemical denaturant

427   to characterize the folding pathway of src SH3. Thus, the evidence so far suggests that single-

428   domain proteins, both α-helical and β-sheet, can fold close to the ribosome. On the other

429   hand, while all-β proteins appear to fold by a similar pathway with or without the ribosome

430   present, there is evidence for α-helical proteins forming partially structured cotranslational

431   intermediates (11, 65) or folding by different pathways on the ribosome (2). This mechanistic

432   difference may relate partly to the small contact order of helical proteins, allowing partially

433   folded states to be more stable than for all-β proteins. The situation for multidomain proteins

434   is likely to be still more complicated, as some studies have already indicated (11, 23, 66, 67).

435

16

443   Diseases of the National Institutes of Health; JC is a Wellcome Trust Senior Research

444   Fellow. The cryo-EM data were collected at the Swedish National Cryo-EM Facility funded

445   by the Knut and Alice Wallenberg Foundation, the Family Erling Persson Foundation and the

446   Science for Life Laboratory. This work utilized the computational resources of the NIH HPC

447   Biowulf cluster. (http://hpc.nih.gov)

448

Materials and Methods

Enzymes and chemicals

451   All enzymes were obtained from Thermo Scientific. Oligonucleotides were purchased from

452   Life Technologies. In-Fusion Cloning kits were obtained from Clontech and DNA

453   purification kits were purchased from Qiagen. PUREfrex cell-free translation system was

454   obtained from Eurogentec. [35S]-methionine was purchased from Perkin Elmer. Instant Blue

455   protein stain was purchased from Expedeon.

456

DNA manipulation

458   Titin I27 constructs for *in vitro* translation were generated in pRSET A plasmid (Invitrogen)

459   (previously modified to remove the sequence including the entire T7 gene 10 leader and EK

460   recognition site up to, but not including, the *Bam*H I site and replaced with a sequence

461   encoding residues L, V, P, R, G, S) carrying the *E. coli* SecM arrest peptide

462   (FSTPVWISQAQGIRAGP) and a truncated *E. coli lepB* gene, under the control of a T7

463   promoter. Increasing linker lengths were generated in pRSET A by PCR; linear pRSET A

464   constructs (containing the SecM AP and truncated lepB, but lacking I27) were generated by

465   PCR using primers which extended the linker from 23 aa to 63 aa (in steps of 2 aa) from the

466   direction of the C to the N terminus. I27 flanked by GSGS linkers was amplified by PCR

467   with overhanging homology to the plasmid containing the desired linker length. Cloning was

468   performed using the In-Fusion system (Takara Bio USA, Inc.), according to the

469   manufacturer's instructions. The final two C-terminal residues (EL) of the 89 aa Titin I27

470   construct are not structured in the PDB file 1TIT, and are therefore included in the linker

471   region. The amino acid sequence of the construct I27[*L*=63] is as follows (I27 in bold and

472   SecM AP underlined):

473   MRGSHHHHHHGLVPRGSGS**LIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWK**

474   **LKGQPLAASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANTKSAANLKVK**EL

475   SGSGKFAYGIKDPIYQKTLVPGQQNATWIVPPGQYFMMGDWMSS<u>FSTPVWISQAQG</u>

476   <u>IRAGP</u>GSSDKQEGEWPTGLRLSRIGGIH**

477   The mutants I27[–A] (lacking β-strand A), I27[L58A] and I27[M67A] were generated for

478   each linker length by site-directed mutagenesis. For the wild-type I27 and I27[–A] constructs

479   with $L = 27, 35, 37, 39, 47$ and 57 residues, site-directed mutagenesis was performed to

480   generate constructs with the non-functional FSTPVWISQAQGIRAG<u>A</u> arrest peptide

481   (mutated residue underlined) as full-length controls, and constructs with the crucial Pro, at

482   the end of the AP, substituted with a stop codon as arrest controls. Site-directed mutagenesis

483   was performed to generate W34E variants as non-folding (nf) controls at $L = 27, 29, 31, 35,$

484   37, 39, 41, 43, 47, 49, 51 and 57 for wild-type I27; $L = 27, 35, 37, 39, 47$ and 57 residues for

485   I27[–A]; $L = 27, 41, 45, 47, 49$ and 53 residues for I27[L58A]; $L = 27, 29, 37, 39, 41, 43, 45,$

486   47 and 51 residues for I27[M67A]. All constructs were verified by DNA sequencing.

487   **In vitro transcription and translation**

488   Transcription and translation were performed using the commercially available PUREfrex *in*

489   *vitro* system (GeneFrontier Corporation), according to the manufacturer's protocol, using 250

490   μg plasmid DNA as template. Synthesis of [^{35}S]-Met-labeled polypeptides was performed at

491   37 °C, 500 r.p.m. for exactly 15 min. The reaction was quenched by the addition of an equal

492   volume of 10% ice-cold trichloroacetic acid (TCA). The samples were incubated on ice for

493   30 min and centrifuged for 5 min at $20,800 \times g$ and 4 °C. Pellets were dissolved in sample

494   buffer and treated with RNase A (400 μg ml^{-1}) for 15 min at 37 °C before the samples were

495   resolved by SDS-PAGE and imaged on a Typhoon Trio or Typhoon 9000 phosphorimager

496   (GE Healthcare). Bands were quantified using ImageJ to obtain an intensity cross section,

497   (http://rsb.info.nih.gov/ij/), which was subsequently fit to a Gaussian distribution using in-

498   house software (Kaleidagraph, Synergy Software). The fraction full-length protein, f_{FL}, was

499   calculated as $f_{FL} = I_{FL}/(I_{FL}+I_A)$, where I_{FL} and I_A are the intensities of the bands representing

500   the full-length and arrested forms of the protein. For wild-type I27 and six nf control samples

501   ($L = 27, 35, 37, 39, 47$ and 57 residues), *in vitro* transcription and translation were also

502   performed at 37 °C, 500 r.p.m. for exactly 30 min. The resultant force profile was slightly

503   higher than that obtained at 15 min but has essentially the same shape (SI Fig. S5).

504   The reproducibility of force profile data has been discussed previously (2). For wild-type I27,

505   data points $L = 61$ and 63 residues are a single experiment; $L = 33, 36, 38, 45, 53, 55$ and 59

506   residues are an average of 2 experiments; all other values of L are an average of at least 3

507   experiments. For I27[–A] strand, $L = 23, 25, 33, 41, 43, 51, 53, 55$ residues are a single

508   experiment; all other values of L are an average of 2 experiments, except L = 35, 37 and 39

509   residues which are an average of at least 3 experiments. For I27[L58A], all data points are a

510   single experiment except L = 27, 37, 41, 45, 47, 49 and 53 residues, which are an average of

511   2 experiments. For I27[M67A], L = 23, 25 and 51 – 63 residues are a single experiment; L =

512   29 – 35 residues are an average of 2 experiments; L = 27, 37 – 47 and 51 residues are an

513   average of at least 3 experiments. For wild-type I27 samples incubated for 30 min, all data

514   points are a single experiment except L = 27, 35, 37, 39, 47 and 57 residues, which are an

515   average of 2 experiments. For non-folding controls, all data points are a single experiment

516   except for wild-type I27 L = 29, 31, 39, 43 and 47 residues which are an average of 2

517   experiments.

518   **Cloning and purification of ribosome-nascent chain complexes**

519   The I27 construct at L = 35, which is at the peak of f_{FL} (Figure 1B), was studied by cryo-EM.

520   The SecM AP in these constructs was substituted with the TnaC AP (34) for more stable

521   arrest, and the constructs were engineered to maintain a linker length of 35 amino acid

522   residues. An N-terminal 8X His tag was introduced to enable purification. The amino acid

523   sequence of the construct used was (I27 in bold and TnaC AP underlined):

524   MDMGHHHHHHHHDYDIPTTLEVLFQGPGT**LIEVEKPLYGVEVFVGETAHFEIELS**

525   **EPDVHGQWKLKGQPLAASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANTKS**

526   **AANLKVK**ELSGSGSGSGGPNILHISVTSKWFNIDNKIVDHRP**

527   The construct was engineered into a pBAD expression vector, under the control of an

528   arabinose-inducible promoter. The translation-initiation region was optimized as described in

529   (68). The plasmid was transformed into the *E. coli* KC6 *ΔsmpB ΔssrA* strain. 4 colonies were

530   picked and tested for expression of the RNCs at 37°C in Lysogeny broth (LB).

531   Large-scale purification of RNCs was carried out based on a protocol described in (34).

532   Briefly, a single colony of the KC6 cells found to express the RNCs was picked and cultured

533   in LB at 37°C to an A_{600} of 0.5. Expression was induced with 0.3% arabinose and was carried

534   out for 1 hour. Thereafter, the cells were chilled on ice, harvested by centrifugation, and

535   resuspended in Buffer A at pH 7.5 (50 mM HEPES-KOH, 250 mM KOAc, 2 mM

536   Tryptophan, 0.1% DDM, 0.1% Complete protease inhibitor). Cell lysis was carried out by

537   passing the cell suspension thrice through the Emulsifex (Avestin) at 8000 psi at 4°C. The

538   lysate was cleared of cell debris by centrifugation at 30,000xg for 30 min in the JA25-50

539   rotor (Beckman Coulter). The supernatant obtained was loaded on a 750 mM sucrose cushion

540 (in Buffer A) and centrifuged at 45, 000 x g for 24 hours in a Ti70 rotor (Beckman Coulter)

541 to obtain a crude ribosomal pellet, which was resuspended in 200 μl Buffer A by shaking

542 gently on ice.

543 RNCs from the crude suspension were purified via their His tags by affinity purification

544 using Talon (Clontech) beads, which was pre-incubated with 10 μg/ml tRNA to reduce

545 unspecific binding of ribosomes. The suspension was incubated with the beads for 1 hour at

546 4°C and subsequently washed with 20 column volumes of Buffer B at pH 7.5 (50 mM

547 HEPES-KOH, 10 mM $Mg(OAc)_2$, 0.1% Complete Protease Inhibitor, 250 mM sucrose, 2

548 mM Tryptophan). RNCs were eluted by incubating the Talon beads with Buffer C at pH 7.5

549 (50 mM HEPES, 150 mM KOAc, 10 mM $Mg(OAc)_2$, 0.1% Complete protease inhibitor, 150

550 mM imidazole, 250 mM sucrose) for 15 minutes and subsequently collecting the flow-

551 through. Elution was carried out thrice and the eluents were concentrated by centrifugation at

552 40,000 rpm for 2.5 hours in a TLA 100.3 rotor (Beckman Coulter). The pellet obtained at the

553 end of this step was gently suspended in a minimal volume of Buffer D at pH 7 (20 mM

554 HEPES-KOH, 50 mM KOAc, 5 mM $Mg(OAc)_2$, 125 mM sucrose, 2 mM Trp, 0.03% DDM).

555 **Cryo-EM sample preparation, data collection, processing and accession codes**

556 Approximately 4 A_{260}/ml units of RNCs were loaded on Quantifoil R2/2 grids coated with

557 carbon (3 nm thick) and vitrified using the Vitrobot Mark IV (FEI-Thermo) following the

558 manufacturer's instructions. Cryo-EM data was collected at the Cryo-EM National Facility at

559 the Science for Life Laboratory in Stockholm, Sweden.

560 Data was acquired on a 300 keV Titan Krios microscope (FEI) equipped with a K2 camera

561 and a direct electron detector (both from Gatan). The camera was calibrated to achieve a

562 pixel size of 1.06 Å at the specimen level. 30 frames were acquired with an electron dose

563 0.926 e^-/Å^2/frame and a total dose of 27.767 e^-/Å^2 and defocus values between -1 to -3 μm.

564 The first two frames were discarded and the rest were aligned using MotionCor2 (69). Raw

565 images were cropped into squares by RELION 2.1 beta 1 (70). Power-spectra, defocus values

566 and estimation of resolution were determined using the Gctf software (71) and all 2,613

567 micrographs were manually inspected in real space, in which 2,613 were retained. 468,015

568 particles were automatically picked by Gautomatch (http://www.mrc-lmb.cam.ac.uk/kzhang/)

569 using the *E. coli* 70S ribosome as a template. Single particles were processed by RELION 2.1

570 beta 1 (70). After 80 rounds of 2D classification, 384,039 particles were subjected to 3D

571 refinement using the *E. coli* 70S ribosome as reference structure, followed by 160 rounds of

20

572   3D classification without masking and 25 rounds of tRNA-focused sorting. One major class

573   containing 301,510 particles (64% of the total) was further refined including using a 50S

574   mask, resulting in a final reconstruction with an average resolution of 3.2 Å (0.143 FSC). The

575   local resolution was calculated by ResMap (72). Finally, the final map was obtained by local

576   B-factoring followed by low-pass filtering to 4.5 Å by RELION 2.1 beta 1 (70) in order to

577   best demonstrate the I27 domain.

578   For interpretation of the cryo-EM density, the cryo-EM structure model (PDB 4YU8) of *E.*

579   *coli* TnaC-stalled ribosome was fitted into corresponding density using UCSF Chimera (73).

580   The NMR model (PDB 1TIT) of I27 domain was fitted into the extra density of TnaC-stalled

581   ribosome using UCSF Chimera (73). Since the I27 domain represents a flat ellipsoid, we used

582   all four major and minor axes covering all possible orientations of the model fitting within

583   the density to validate the orientation of the fitted I27 model. Briefly, the model with four

584   different orientations were converted into densities (8 Å) by UCSF Chimera, and the cross-

585   correlation coefficients of each model map and the isolated I27 density were calculated by

586   RELION 2.1 beta 1 (70). Finally, uL24 β hairpin was remodeled as the tip of the hairpin is

587   shifted due to the existence of I27 domain.

588   Figures showing electron densities and atomic models were generated using UCSF Chimera

589   (73). Electron densities are shown at multiple contour levels in Figure 2 and SI Appendix, Fig.

590   S1. The contour levels relative to the root-mean-square deviation (RMSD) were calculated

591   from the final map values. Final map contains the volume for the entire RNC including the

592   I27 domain.

593   Coordinates for the cryo-EM map of the ribosome with the I27 domain density have been

594   deposited at the EMDataBank under accession code EMD-*xxxx*. Coordinates of fitted *E.coli*

595   TnaC-stalled ribosome (PDB 4UY8; uL24 remodeled) and I27 domain (PDB 1TIT) models

596   for interpreting the cryo-EM map have been deposited at the ProteinDataBank under

597   accession code *xxxx*.

598   **Coarse-grained molecular simulations**

599   The 50S subunit of the *E. coli* ribosome (PDB 3OFR (36)) and the nascent chain are

600   explicitly represented using one bead at the position of the α-carbon atom of each amino

601   acid, and three beads (for P, C4', N3) per RNA residue (Figure 2A). The interactions within

602   the protein were given by a standard structure-based model (37-39), which allowed it to fold

21

603 and unfold. Interactions between the protein and ribosome beads were purely repulsive (40)

604 and given by the same form of potential as for the structure-based model(37-39),

605
$$V_{ij} = \varepsilon_{ij} \left[\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{10}} + \frac{C}{r_{ij}^{6}} \right] \quad [2]$$

606 where r_{ij} is the distance between two beads i and j, ε_{ij} (=0.001 kJ/mol) sets the strength of

607 the repulsive interactions. The amino acid, phosphate, sugar and base are assigned radii $\sigma_i =$

608 4.5, 3.2, 5.1 and 4.5 Å respectively, and coefficients in Eq. 2 for interactions between protein

609 and ribosome beads i,j are obtained from the mixing rules $A = \sqrt{\sigma_i^{12} \sigma_j^{12}}$, $B = \frac{2}{\sigma_i^{-10} + \sigma_j^{-10}}$

610 and $C = \sqrt{\sigma_i^{6} \sigma_j^{6}}$.

611 During the simulations, the positions of the ribosome atoms were fixed in space, as in

612 previous studies (18). The linker between the AP and I27 was tethered by its C terminus to

613 the last P atom of the A-site tRNA, but was otherwise free to fluctuate. The trajectory was

614 propagated via Langevin dynamics, with a friction coefficient of 0.1 ps^{-1} and a time step of

615 10 fs, at 291 K in a version of the Gromacs 4.0.5 simulation code, modified to implement the

616 potential given by Eq. 2 (74). All bonds (except the one used to measure force, below) were

617 constrained to their equilibrium length using the LINCS algorithm (75). The attractive

618 interactions between I27[M67A] and the hydrophobic residues (A, V, L, I, F, M, Y, W) on

619 the surface of uL23 and uL29 are modelled as (76):

620
$$V_{ij} = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right], \quad [3]$$

621 where r_{ij} is the distance between residues i and j, σ is the range of the interaction and ε

622 represent the strength of the interaction. σ and ε are fixed at 6 Å and 5 kJ/mol respectively.

623 Residues of I27[M67A] which are involved in the attractive interactions are defined as the

624 ones whose heavy atoms are within 4.5 Å of any heavy atoms from residue 67 in the native

625 state.

626 To calculate the pulling force exerted on the nascent chain by the folding of I27, the bond

627 between the last and the second last amino acid of the SecM AP was modelled by a harmonic

628 potential as a function the distance between these two atoms, x (Figure 3B):

629
$$E = \frac{1}{2} k_s (x - x_0)^2 \quad [4]$$

22

630   where x_0 is a reference distance. Here x_0 is set to 3.8 Å, which is the approximate distance

631   between adjacent Cα atoms in protein structures and k_s is a spring constant, set to 3000

632   kJ.mol^{-1}.nm^{-2}. The value of k_s was chosen so that the average displacement $x - x_0$ remains

633   below 1 Å for forces up to ~500 pN, which is much larger than the forces actually exerted by

634   the folding protein. The pulling force on the nascent chain was measured by the extension of

635   this bond as $F = -k_s(x - x_0)$.

636   I27 was covalently attached to unstructured linkers having the same sequences as used in the

637   force-profile experiments (see Figure 2B). Linker amino acids are repulsive to both the

638   ribosome and I27 beads, with interaction energy as described in Eq. 2.

639   The protein in its arrested state is subject to force $F(t)$, which will fluctuate, for example

640   when the protein folds or unfolds. The rate of escape from arrest has been shown to be force-

641   dependent (25); here we approximate the sensitivity to force using the phenomenological

642   expression originally proposed by Bell (42)

$$k(F) = k_0 e^{\beta F \Delta x^{\ddagger}}, \quad [5]$$

644   where k_0 is a zero-force rupture rate, $\Delta x^{\ddagger}$ is the distance from the free energy minimum to

645   the transition state, $\beta = 1/k_B T$ where k_B is Boltzmann's constant and T the absolute

646   temperature. While there are functions to describe force-dependent rates with stronger

647   theoretical basis, we use the Bell equation due to its simplicity and because its parameters

648   have previously been estimated from optical tweezer experiments for the SecM AP. In all

649   cases, we set k_0 (Eq. 5) to 3.4 ×10^{-4} s^{-1} and $\Delta x^{\ddagger}$ to 3.2 Å, based on the values determined by

650   Goldman *et al.* (they estimated k_0 and $\Delta x^{\ddagger}$ to be in the range of 0.5 ×10^{-4} to 20 ×10^{-4} s^{-1} and

651   1-8 Å, respectively) (25).

652   We assume the probability of remaining on the ribosome $S(t) = 1 - f_{FL}(t)$ assuming that

653    $\dot{S} = -k(F(t))$, hence

$$S(t) = \exp[-\int_0^t k(F(t))dt]. \quad [6]$$

655   The escape of I27 from the ribosome can be described using kinetic model shown in SI

656   Appendix, Fig. S8 which explicitly takes into account the linker length-dependent

657   folding/unfolding rates of the I27 nascent chain, $k_u(L)$ and $k_f(L)$, on the ribosome, and the

658   force-dependent rate of escape from ribosome: $k(F_u(L))$ and $k(F_f(L))$. To estimate k_f at

659   different linker lengths, we first carried out unbiased MD simulations to estimate the mean

23

660   first passage time for folding t_F^{mfpt}, from which the folding rate can be calculated as

661      $k_f^{wt} = 1/t_F^{mfpt}$. Similarly, the unfolding rate can be calculated from unfolding simulations as

662      $k_u^{wt} = 1/t_u^{mfpt}$. Since the rates in coarse-grained simulations are naturally much faster than

663   in experiment, we globally scale the unfolding rates k_u^{wt}(L=21, 23 …61) at different linker

664   lengths so that k_u^{wt} at very long linker lengths (L=61) is equal to the unfolding rate of isolated

665   I27 (4.9×10⁻⁴ s⁻¹). Similarly, k_f^{wt}(L=21, 23 …61) is scaled so that the k_f^{wt} of I27 RNC[L=61]

666   is equal to the unfolding rate of isolated I27 (SI Appendix, Fig. S9). For consistency with our

667   pre-equilibrium solution, we further scale k_f^{wt} to match the stability of I27 RNC[L=61] in

668   our simulation model, yielding k_u^{wt} and k_f^{wt} at L=61 of 4.9×10⁻⁴ s⁻¹ and 0.14 s⁻¹ (SI Appendix,

669   Fig. S9) respectively. The same scaling method has been applied to the folding and

670   unfolding rates of all mutants (I27[L58A], I27[M67A] and I27 [-A]) so that the

671   folding/unfolding rates of the mutant RNC are consistent with the relative experimental

672   values measured for the isolated mutants(26).

673

674   With the rates obtained from above, the time dependent survival probability $S(t)$ is estimated

675   by the kinetic Monte Carlo method (the Bortz-Kalos-Lebowitz algorithm (77)). The system is

676   initialized at the state when the unfolded nascent chain just emerges from the ribosome tunnel

677   (UA, SI Appendix, Fig. S8) at time t=0. At each Monte Carlo step, a uniform random number

678   δ between 0 and 1 is chosen, and a transition from the current state state s to state j will occur

679   for the state j which satisfies $\sum_{i=1}^{j-1} k_{si} < \delta \sum_{i=1}^{N} k_{si} < \sum_{i=1}^{j} k_{si}$, where k_{si} represents the

680   transition rate from state s to state i. The time is updated by $t = t + \Delta t$, where $\Delta t =$

681      $-(\delta')/\sum_{i=1}^{N} k_{si}$. δ' is a new number randomly chosen between 0 to 1.

682

683   The solution to the kinetic model can be simplified if we further assume that the escape from

684   the ribosome is slow relative to the folding and unfolding of the protein. In this situation, we

685   can approximate $S(t)$ in terms of the mean forces experienced when the protein is unfolded,

686   F_u, or folded, F_f, and the unfolded and folded populations of P_u and P_f respectively,

687      $$S(t) \approx \exp\left[-t[P_u k(F_u) + P_f k(F_f)]\right]. \quad [7]$$

688   The equilibrium properties of the system for each linker length were obtained from umbrella

689   sampling using the fraction of native contacts Q as the reaction coordinate, allowing P_u, P_f

690   and F_u, F_f to be determined (Figure 3C). The details of the definition of Q have been

691   previously described (48); in short, Q is defined as

$$Q = \frac{1}{N} \sum_{(i,j)} \frac{1}{1 + e^{(r_{ij} - \lambda r_{ij}^0)}}, \quad [8]$$

693   where the sum runs over the N pairs of native contacts (i,j), r_{ij} is the distance between i and

694   j in configuration, r_{ij}^0 is the distance between i and j in the native state, λ =1.2 which

695   accounts for fluctuations when the contact is formed. A boundary of Q = 0.5 is used to

696   separate folded from unfolded states.

697   In order to characterize folding mechanism, we used transition paths from folding

698   simulations for the L = 51 case at 291 K. 50 independent simulations, each started from fully

699   extended configurations, were carried out for 4 microseconds. The folding barriers for the L =

700   31 and 35 cases are very high at the same temperature, therefore the transition paths are

701   obtained from unfolding simulations instead. Starting from native-like folded

702   configurations, 50 unfolding simulations were carried out, with each trajectory being 4

703   microseconds long. Transition paths were defined as those portions of the simulation

704   trajectory from the last time I27 samples the configuration with Q < 0.3 till the first time it

705   samples a configuration with Q > 0.7 (in the folding direction; opposite for unfolding). ϕ-

706   Values were computed from the transition paths using the approximation:

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

708   In which $p(q_{ij}|TP)$ is the probability that the native contact q_{ij} between residues i and j is

709   formed on transition paths as defined above. We also characterized the importance of

710   individual contacts in determining the folding mechanism using $p(\text{TP}|q_{ij})_{nn}$, defined in Eq.

711   1 of the main text, i.e. the probability of being on a transition path given that contact q_{ij} is

712   formed and the protein is not yet folded. Having already calculated $p(q_{ij}|TP)$ above,

713   evaluating $p(TP|q_{ij})_{nn}$ required $p(q_{ij})_{nn}$, the probability of a contact being formed in all

714   non-native fragments of the trajectory, and $p(\text{TP})$, the fraction of time spent on transition

715   paths. For L=51, we obtained $p(q_{ij})_{nn}$ directly from unbiased folding simulations, using the

716   portion of the trajectory up to the first folding event (i.e. the first time Q > 0.7). For L=31 or

25

717   35, where the protein is still relatively unstable, we determined it from unfolding simulations

718   by computing $p(q_{ij})$ separately for the unfolded and transition-path portions of the trajectory

719   and combining them weighted by $p(\text{TP})_{nn}$. We determined $p(\text{TP})_{nn}$ via folding ($L = 51$

720   case) and unfolding ($L = 31$ and $L = 35$ cases) simulations (described above). For the $L = 51$

721   case, $p(\text{TP})_{nn} = \frac{2t_{\text{TP}}}{2t_{\text{TP}} + t_{\text{F}}^{\text{mfpt}}}$, where t_{TP} is the mean transition path time and $t_{\text{F}}^{\text{mfpt}}$ is the mean

722   first passage time for folding obtained from the maximum likelihood estimator $t_{\text{F}}^{\text{mfpt}} =$

723    $[N_{\text{fold}} t_{\text{fold}} + (N - N_{\text{fold}}) t_{\text{sim}}]/N_{\text{fold}}$, where N is the total number of trajectories ($N = 50$), N_{fold}

724   is the number of trajectories folding within 4 μs, t_{fold} is the average folding time (of the

725   trajectories which fold), and t_{sim} is the length of the simulations (4 μs). For the $L = 31$ and L

726    $= 35$ cases, it is less efficient to obtain the folding time $t_{\text{F}}^{\text{mfpt}}$ directly, therefore we estimate it

727   based on the mean first passage time for unfolding, $t_{\text{U}}^{\text{mfpt}}$, from unfolding simulations.

728    $p(\text{TP})_{nn} = \frac{2t_{\text{TP}}}{2t_{\text{TP}} + \frac{p_U}{p_F} t_{\text{U}}^{\text{mfpt}}}$, where p_U and p_F are the equilibrium populations of the unfolded

729   and folded respectively determined from umbrella sampling.

730

731   **References**

732   1.   Nilsson OB, Hedman R, Marino J, Wickles S, Bischoff L, Johansson M, Muller-
733   Lucks A, Trovato F, Puglisi JD, O'Brien EP, Beckmann R, & von Heijne G (2015)
734   Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep*
735   12(10):1533-1540.
736   2.   Nilsson OB, Nickson AA, Hollins JJ, Wickles S, Steward A, Beckmann R, von
737   Heijne G, & Clarke J (2017) Cotranslational folding of spectrin domains via partially
738   structured states. *Nat Struct Mol Biol* 24(3):221-225.
739   3.   Nilsson OB, Muller-Lucks A, Kramer G, Bukau B, & von Heijne G (2016) Trigger
740   Factor Reduces the Force Exerted on the Nascent Chain by a Cotranslationally
741   Folding Protein. *J Mol Biol* 428(6):1356-1364.
742   4.   Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science*
743   181(4096):223-230.
744   5.   Onuchic JN, Luthey-Schulten Z, & Wolynes PG (1997) Theory of protein folding: the
745   energy landscape perspective. *Annu Rev Phys Chem* 48(1):545-600.
746   6.   Dill KA & Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol*
747   4(1):10-19.
748   7.   Tycko R (2004) Progress towards a molecular-level structural understanding of
749   amyloid fibrils. *Curr Opin Struct Biol* 14(1):96-103.
750   8.   Clark PL (2004) Protein folding in the cell: reshaping the folding funnel. *Trends*
751   *Biochem Sci* 29(10):527-534.
752   9.   Dobson CM (2001) Protein folding and its links with human disease. *Biochem Soc*
753   *Symp* (68):1-26.

754   10.   Gloge F, Becker AH, Kramer G, & Bukau B (2014) Co-translational mechanisms of
755   protein maturation. *Curr Opin Struct Biol* 24:24-33.
756   11.   Holtkamp W, Kokic G, Jager M, Mittelstaet J, Komar AA, & Rodnina MV (2015)
757   Cotranslational protein folding on the ribosome monitored in real time. *Science*
758   350(6264):1104-1107.
759   12.   Thommen M, Holtkamp W, & Rodnina MV (2017) Co-translational protein folding:
760   progress and methods. *Curr Opin Struct Biol* 42:83-89.
761   13.   Krobath H, Shakhnovich EI, & Faísca PF (2013) Structural and energetic
762   determinants of co-translational folding. *J Chem Phys* 138(21):215101.
763   14.   Tanaka T, Hori N, & Takada S (2015) How co-translational folding of multi-domain
764   protein is affected by elongation schedule: molecular simulations. *PLOS Comput Biol*
765   11(7):e1004356.
766   15.   Wruck F, Katranidis A, Nierhaus KH, Buldt G, & Hegner M (2017) Translation and
767   folding of single proteins in real time. *Proc Natl Acad Sci U S A* 114(22):E4399-
768   E4407.
769   16.   Jacobs WM & Shakhnovich EI (2017) Evidence of evolutionary selection for
770   cotranslational folding. *Proc Natl Acad Sci U S A* 114(43):11434-11439.
771   17.   Samelson AJ, Jensen MK, Soto RA, Cate JH, & Marqusee S (2016) Quantitative
772   determination of ribosome nascent chain stability. *Proc Natl Acad Sci U S A*
773   113(47):13402-13407.
774   18.   O'Brien EP, Christodoulou J, Vendruscolo M, & Dobson CM (2011) New scenarios
775   of protein folding can occur on the ribosome. *J Am Chem Soc* 133(3):513-526.
776   19.   Knight AM, Culviner PH, Kurt-Yilmaz N, Zou T, Ozkan SB, & Cavagnero S (2013)
777   Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS*
778   *Chem Biol* 8(6):1195-1204.
779   20.   Cabrita LD, Cassaignau AME, Launay HMM, Waudby CA, Wlodarski T, Camilloni
780   C, Karyadi ME, Robertson AL, Wang X, Wentink AS, Goodsell L, Woolhead CA,
781   Vendruscolo M, Dobson CM, & Christodoulou J (2016) A structural ensemble of a
782   ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct*
783   *Mol Biol* 23(4):278-285.
784   21.   Fedorov AN & Baldwin TO (1995) Contribution of cotranslational folding to the rate
785   of formation of native protein structure. *Proc Natl Acad Sci U S A* 92(4):1227-1231.
786   22.   Nicola AV, Chen W, & Helenius A (1999) Co-translational folding of an alphavirus
787   capsid protein in the cytosol of living cells. *Nat Cell Biol* 1(6):341-345.
788   23.   Ugrinov KG & Clark PL (2010) Cotranslational folding increases GFP folding yield.
789   *Biophys J* 98(7):1312-1320.
790   24.   Evans MS, Clarke TF, & Clark PL (2005) Conformations of co-translational folding
791   intermediates. *Protein Pept Lett* 12(2):189-195.
792   25.   Goldman DH, Kaiser CM, Milin A, Righini M, Tinoco I, Jr., & Bustamante C (2015)
793   Mechanical force releases nascent chain-mediated ribosome arrest *in vitro* and *in vivo*.
794   *Science* 348(6233):457-460.
795   26.   Fowler SB & Clarke J (2001) Mapping the folding pathway of an immunoglobulin
796   domain: structural detail from Phi value analysis and movement of the transition state.
797   *Structure* 9(5):355-366.
798   27.   Fowler SB, Best RB, Toca Herrera JL, Rutherford TJ, Steward A, Paci E, Karplus M,
799   & Clarke J (2002) Mechanical unfolding of a titin Ig domain: structure of unfolding
800   intermediate revealed by combining AFM, molecular dynamics simulations, NMR
801   and protein engineering. *J Mol Biol* 322(4):841-849.

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. Farias-Rico JA, Selin FR, Myronidi I, & Von Heijne G (2018) Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *bioRxiv*:303784.

30. Marino J, Heijne G, & Beckmann R (2016) Small protein domains fold inside the ribosome exit tunnel. *FEBS letters* 590(5):655-660.

31. Kramer G, Rauch T, Rist W, Vorderwulbecke S, Patzelt H, Schulze-Specking A, Ban N, Deuerling E, & Bukau B (2002) L23 protein functions as a chaperone docking site on the ribosome. *Nature* 419(6903):171-174.

32. Gong F & Yanofsky C (2003) A transcriptional pause synchronizes translation with transcription in the tryptophanase operon leader region. *J Bacteriol* 185(21):6472-6476.

33. Seidelt B, Innis CA, Wilson DN, Gartmann M, Armache JP, Villa E, Trabuco LG, Becker T, Mielke T, Schulten K, Steitz TA, & Beckmann R (2009) Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* 326(5958):1412-1415.

34. Bischoff L, Berninghausen O, & Beckmann R (2014) Molecular basis for the ribosome functioning as an L-tryptophan sensor. *Cell Rep* 9(2):469-475.

35. Improta S, Politou AS, & Pastore A (1996) Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* 4(3):323-337.

36. 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.

37. 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.

38. 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.

39. 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.

40. Elcock AH (2006) Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLOS Comput Biol* 2(7):e98.

41. Muto H, Nakatogawa H, & Ito K (2006) Genetically encoded but nonpolypeptide prolyl-tRNA functions in the A site for SecM-mediated ribosomal stall. *Mol Cell* 22(4):545-552.

42. Bell GI (1978) Models for the specific adhesion of cells to cells. *Science* 200(4342):618-627.

43. Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E, & Ban N (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* 431(7008):590.

44. Wild K, Halic M, Sinning I, & Beckmann R (2004) SRP meets the ribosome. *Nat Struct Mol Biol* 11(11):1049.

45. Frauenfeld J, Gumbart J, Van Der Sluis EO, Funes S, Gartmann M, Beatrix B, Mielke T, Berninghausen O, Becker T, & Schulten K (2011) Cryo-EM structure of the

ribosome–SecYE complex in the membrane environment. *Nat Struct Mol Biol* 18(5):614.

46. Best RB & Hummer G (2016) Microscopic interpretation of folding φ-values using the transition path ensemble. *Proc Natl Acad Sci U S A* 113(12):3263-3268.

47. Sánchez IE & Kiefhaber T (2003) Origin of unusual φ-values in protein folding: evidence against specific nucleation sites. *J Mol Biol* 334(5):1077-1085.

48. 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.

49. Best RB, Fowler SB, Herrera JL, Steward A, Paci E, & Clarke J (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.

50. 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.

51. Williams PM, Fowler SB, Best RB, Toca-Herrera JL, Scott KA, Steward A, & Clarke J (2003) Hidden complexity in the mechanical properties of titin. *Nature* 422(6930):446-449.

52. Wright CF, Lindorff-Larsen K, Randles LG, & Clarke J (2003) Parallel protein-unfolding pathways revealed and mapped. *Nat Struct Biol* 10(8):658-662.

53. 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.

54. 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.

55. Botello E, Harris NC, Sargent J, Chen WH, Lin KJ, & Kiang CH (2009) Temperature and chemical denaturant dependence of forced unfolding of titin I27. *J Phys Chem B* 113(31):10845-10848.

56. Chen H, Yuan G, Winardhi RS, Yao M, Popa I, Fernandez JM, & Yan J (2015) Dynamics of equilibrium folding and unfolding transitions of titin immunoglobulin domain under constant forces. *J Am Chem Soc* 137(10):3540-3546.

57. 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.

58. Nunes JM, Mayer-Hartl M, Hartl FU, & Muller DJ (2015) Action of the Hsp70 chaperone system observed with single proteins. *Nat Commun* 6:6307.

59. Yagawa K, Yamano K, Oguro T, Maeda M, Sato T, Momose T, Kawano S, & Endo T (2010) Structural basis for unfolding pathway-dependent stability of proteins: vectorial unfolding versus global unfolding. *Protein Sci* 19(4):693-702.

60. Zheng W, Schafer NP, & Wolynes PG (2013) Frustration in the energy landscapes of multidomain protein misfolding. *Proceedings of the National Academy of Sciences* 110(5):1680-1685.

61. Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, & Fernandez JM (1999) Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci U S A* 96(7):3694-3699.

62. Su T, Cheng J, Sohmen D, Hedman R, Berninghausen O, von Heijne G, Wilson DN, & Beckmann R (2017) The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. *eLife* 6:e25642.

63. Ziv G, Haran G, & Thirumalai D (2005) Ribosome exit tunnel can entropically stabilize α-helices. *Proc Natl Acad Sci U S A* 102(52):18956-18961.

901 64. Guinn EJ, Tian P, Shin M, Best RB, & Marqusee S (2018) A small single-domain
902   protein folds through the same pathway on- and off- the ribosome. *bioRxiv*:347864.
903 65. Mercier E & Rodnina MV (2018) Co-translational Folding Trajectory of the HemK
904   Helical Domain. *Biochemistry*.
905 66. Samelson AJ, Bolin E, Costello SM, Sharma AK, O'Brien EP, & Marqusee S (2018)
906   Kinetic and structural comparison of a protein's cotranslational folding and refolding
907   pathways. *Sci Adv* 4(5):eaas9098.
908 67. Evans MS, Sander IM, & Clark PL (2008) Cotranslational folding promotes β-helix
909   formation and avoids aggregation in vivo. *J Mol Biol* 383(3):683-692.
910 68. Mirzadeh K, Martinez V, Toddo S, Guntur S, Herrgard MJ, Elofsson A, Norholm
911   MH, & Daley DO (2015) Enhanced Protein Production in Escherichia coli by
912   Optimization of Cloning Scars at the Vector-Coding Sequence Junction. *ACS Synth
913   Biol* 4(9):959-965.
914 69. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, &
915   Cheng Y (2013) Electron counting and beam-induced motion correction enable near-
916   atomic-resolution single-particle cryo-EM. *Nat Methods* 10(6):584-590.
917 70. Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM
918   structure determination. *J Struct Biol* 180(3):519-530.
919 71. Zhang K (2016) Gctf: Real-time CTF determination and correction. *J Struct Biol*
920   193(1):1-12.
921 72. Kucukelbir A, Sigworth FJ, & Tagare HD (2014) Quantifying the local resolution of
922   cryo-EM density maps. *Nat Methods* 11(1):63-65.
923 73. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, &
924   Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and
925   analysis. *J Comput Chem* 25(13):1605-1612.
926 74. Hess B, Kutzner C, Van Der Spoel D, & Lindahl E (2008) GROMACS 4: algorithms
927   for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory
928   and Comput* 4(3):435-447.
929 75. Hess B, Bekker H, Berendsen HJ, & Fraaije JG (1997) LINCS: a linear constraint
930   solver for molecular simulations. *J Comput Chem* 18(12):1463-1472.
931 76. Sirur A, Knott M, & Best RB (2014) Effect of interactions with the chaperonin cavity
932   on protein folding and misfolding. *Phys Chem Chem Phys* 16(14):6358-6366.
933 77. Bortz AB, Kalos MH, & Lebowitz JL (1975) A new algorithm for Monte Carlo
934   simulation of Ising spin systems. *J Comput Phys* 17(1):10-18.
935 78. Öhman A, Öman T, & Oliveberg M (2010) Solution structures and backbone
936   dynamics of the ribosomal protein S6 and its permutant P54-55. *Protein Sci*
937   19(1):183-189.
938 79. Pascual J, Pfuhl M, Walther D, Saraste M, & Nilges M (1997) Solution structure of
939   the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil1. *J Mol Biol*
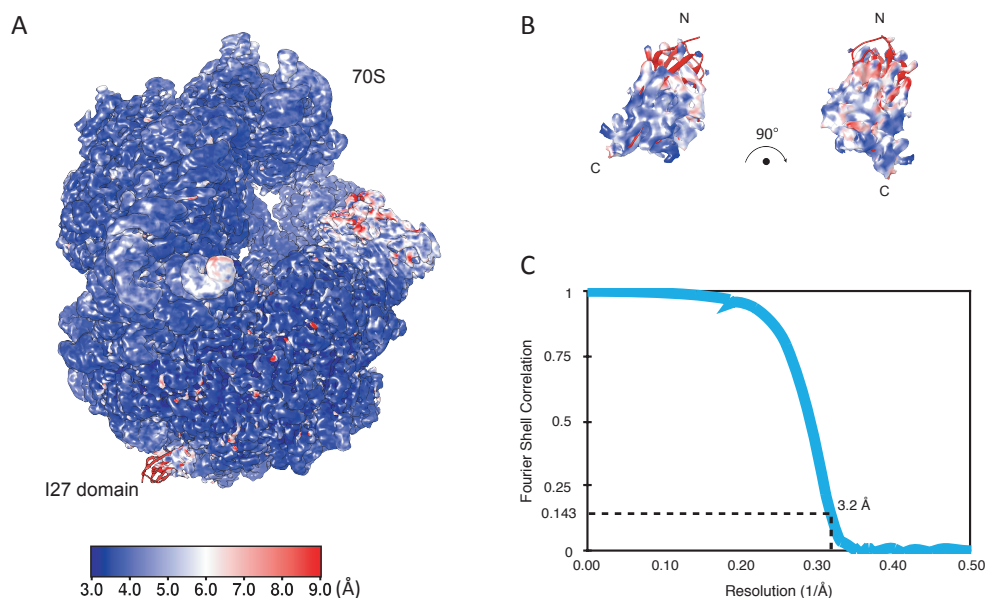940   273(3):740-751.
941
942
943

944 **Supplementary Figures**

945

946
947

948

949 **Figure S1.** Resolution of the ribosome-nascent chain complex (RNC). (A) Calculation of the

950 local resolution using Resmap (Kucukelbir, A. et al. Nat Methods 11, 63-65, 2014). The RNC

951 density is displayed at 1.7 RMSD. (B) local resolution of the I27 domain. The I27 domain

952 density is displayed at 2 RMSD. N and C termini are indicated. (C) Fourier-shell correlation

953 (FSC) curve of the refined final map of the RNC, indicating the average resolution of 3.2 Å

954 (at 0.143).

955
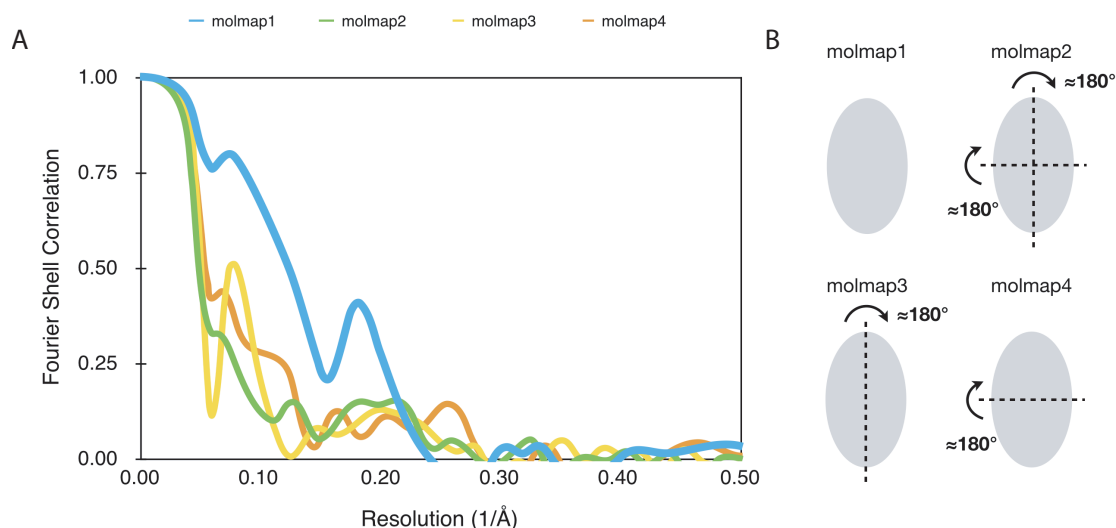
Figure S2. Validation of model orientation for I27 domain. To validate the orientation of the I27 domain model (PDB 1TIT) to its corresponding density, four possible orientations were tested. (A) The Fourier-shell correlations between the isolated I27 density and the map generated from the model of the final orientation (molmap1, blue) and the models fitted with the other three possibilities (molmap2, green; molmap3, yellow; molmap4, orange) were plotted. In the frequency range 0 to 0.2 (1000 to 5 Å) the correlation of molmap1 is significantly higher compared to all other orientation molmaps. (B) The illustration showing the relationship among the four model orientations. Since the density represents a flat ellipsoid, we used all four major and minor axes covering all possible orientations of the model fitting within the density.

967

968
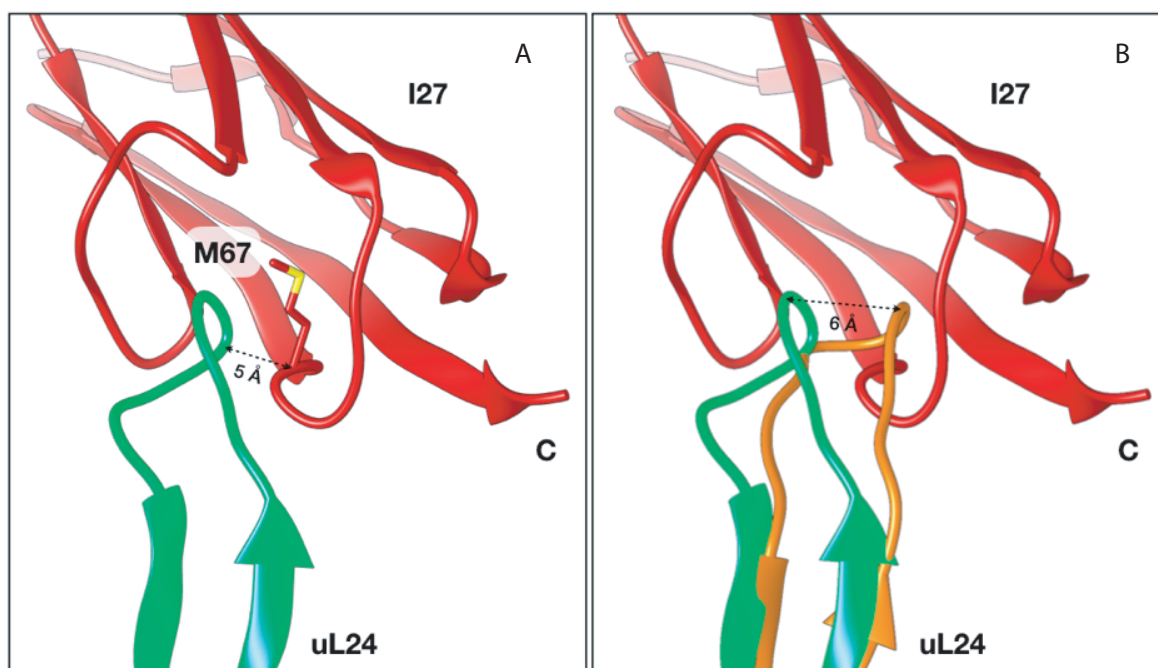
969

970   **Figure S3.** The I27 domain and a β hairpin in ribosomal protein uL24 close to the ribosomal

971   exit tunnel. (A) Residue M67 in the I27 domain is located in close proximity to a β hairpin

972   loop in uL24 in the cryo-EM structure of I27-TnaC[L=35] RNCs. (B) The uL24 β hairpin in

973   the I27-RNC (light green; re-modeled based on PDB 5NWY) is ~ 6 Å shifted (distance

974   measured via the backbone of Pro50) compared to its location in the VemP-RNC (orange;

975   PDB 5NWY) and the TnaC-RNC (PDB 4UY8, not shown). C represents the C terminus of

976   the I27 domain.

977

978
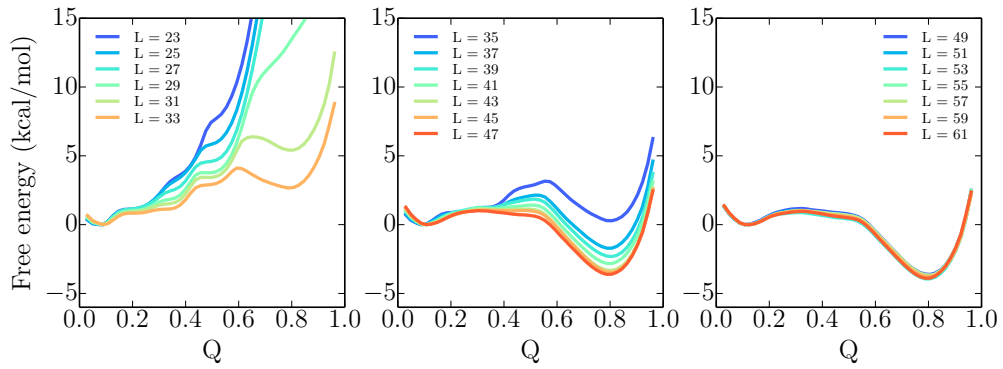
979

980   **Figure S4**. Simulation free energy $F(Q)$ projected on the fraction of native contacts, Q, for

981   I27 folding with different linker lengths (as indicated in legend) at 291K.

982
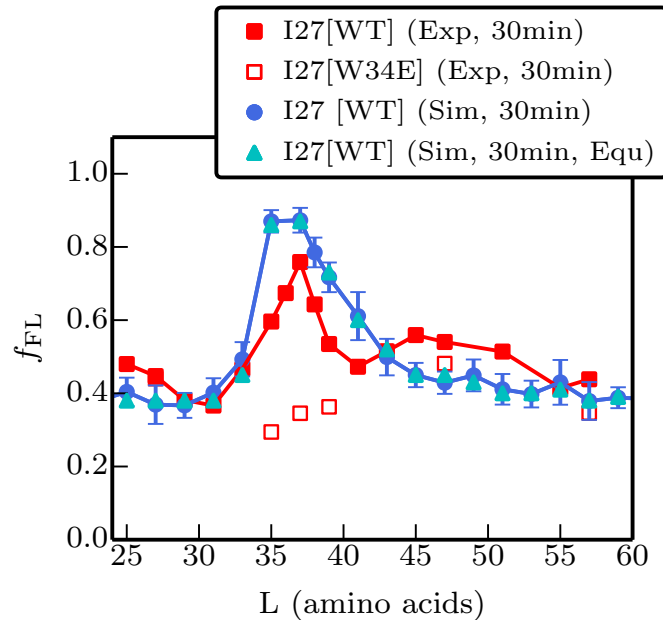
983

984   **Figure S5.** Experimental (red) and simulated profiles of fraction full length protein, f_{FL},

985   obtained with a 30 min incubation. Note the higher background values compared to main text

986   Figures 1 and 3D. Force profiles calculated from simulations using full kinetic scheme and

987   pre-equilibrium model are shown in blue circles and cyan triangles respectively. The RMSD

988   of the f_{FL} between experiment and simulation is 0.12.

989

990
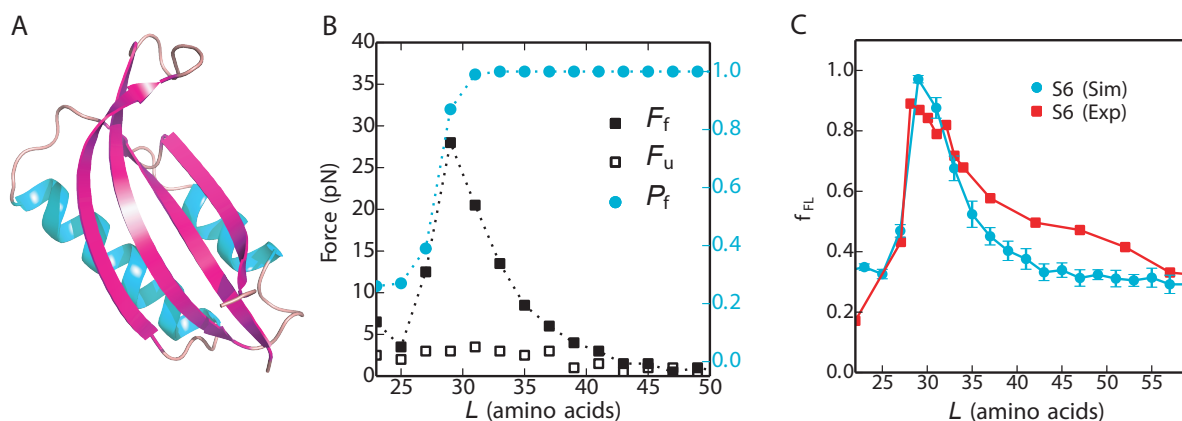
Figure S6. (A) Native structure of protein S6 (pdb code: 2KJV (78)) B) Average forces exerted on the AP by the unfolded state (F_f, filled black symbols) and folded state (F_u, empty black symbols) of S6 at different linker lengths L. The average fraction folded S6 for different L, P_f, is shown in cyan on the right axis. (C) Experimental (red) and simulated (cyan) force profiles for cotranslational folding of S6 based on pre-equilibrium kinetic solution.
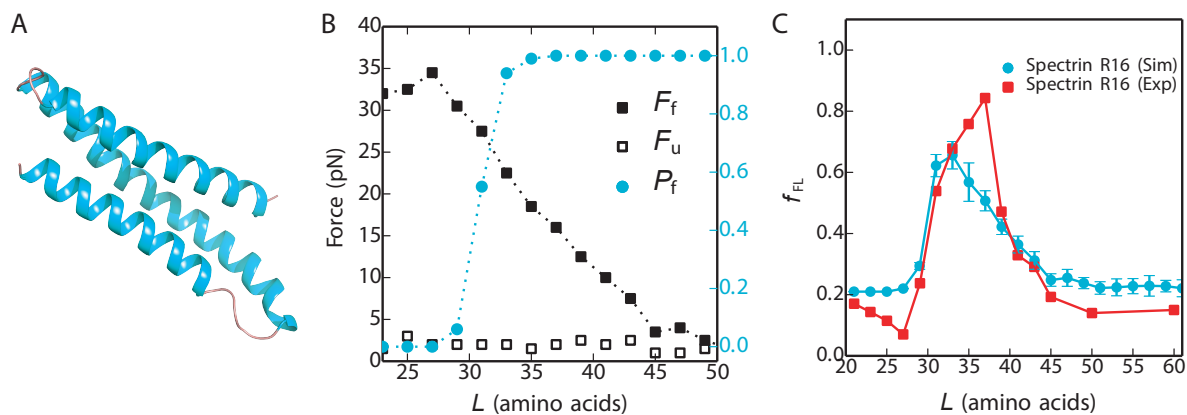

Figure S7. (A) Native structure of Spectrin R16 domain (PDB 1AJ3 (79)). (B) Average forces exerted on the AP by the unfolded state (F_f, empty black symbols) and folded state (F_u, filled black symbols) of R16 at different linker lengths L. The average fraction folded R16 for different L, P_f, is shown in cyan on the right axis. (C) Experimental (red) and simulated (cyan) force profiles for cotranslational folding of R16 based on the pre-equilibrium kinetic solution.

35
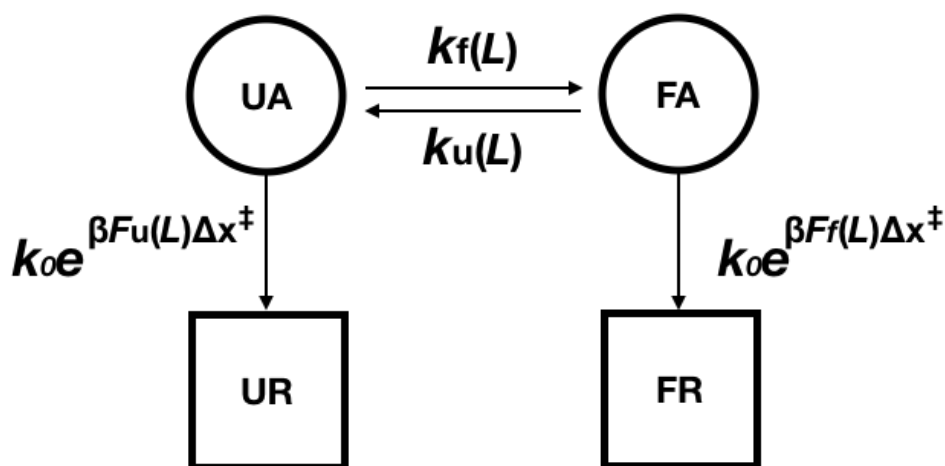
1008

Figure S8. Schematic for the full kinetic model which describes the escape pathway of I27 from the ribosome. k_f and k_u are the linker length-dependent folding and unfolding rates respectively. UA: I27 is unfolded and arrested by ribosome. FA: I27 is folded and arrested by ribosome. UR: I27 is unfolded and released from ribosome. FR: I27 is folded and released from ribosome.

1014

1015
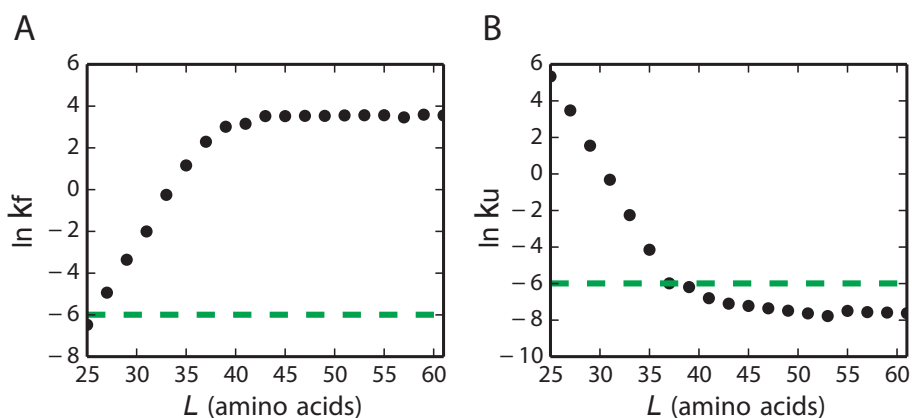
1016

Figure S9. Dependence of folding rate k_f (left) and unfolding rate k_u (right) on the length of the linker between the AP and I27. Rates determined directly from simulations have been scaled so that k_f and k_u at large linker lengths are equal to the experimental values determined for the isolated protein. The green dashed line indicate the force-dependent escape rate of ~ 2.4 x 10^{-3} s^{-1} obtained at the force of 20 pN.
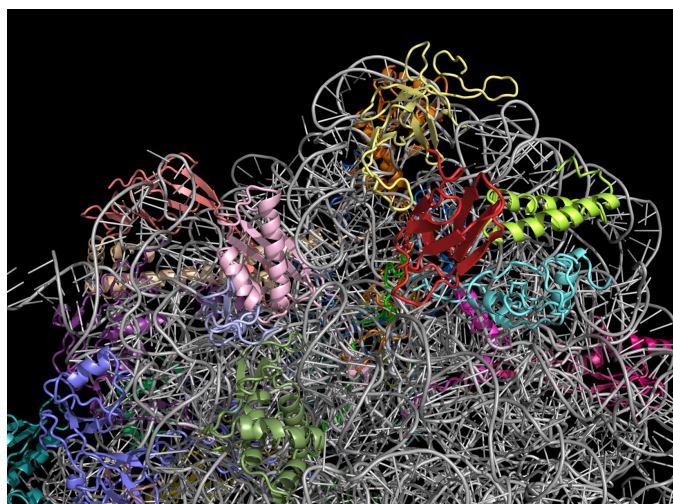
1022

1023
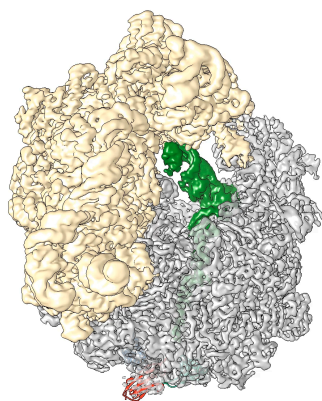
1024

1025 **Supporting Video S1**. Cryo-EM density of ribosome and I27 (one static frame of the video).
1026 Video showing cryo-EM map for I27[L=35] RNCs. 30S in yellow, 50S and I27 domain in
1027 grey, tRNA and nascent chain in green, the model (PDB 1TIT) of I27 domain in red.

1028

1029

1030

1031

1032 **Supporting Video S2.** MD folding simulation (one static frame of the video). Video showing
1033 an unbiased 1.8 μsec fragment of an MD trajectory of I27 folding and unfolding at linker
1034 length L=35. Ribosomal 23s rRNA is shown in white cartoon mode, ribosomal proteins
1035 uL24, uL29, uL23 are shown in yellow, lime and cyan respectively. I27 and linker are in red
1036 and green cartoon mode respectively.

1037