# Structures of protein folding intermediates on the ribosome

Sammy H.S. Chan[1*], Julian O. Streit[1*], Tomasz Włodarski[1,2], Alkistis N. Mitropoulou[1], Anaïs M.E. Cassaignau[1], Ivana V. Bukvin[1], Lisa D. Cabrita[1] & John Christodoulou[1,3◊]

[1]Institute of Structural and Molecular Biology, Department of Structural and Molecular Biology, University College London, London, UK

[2]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

[3]School of Natural Sciences, Birkbeck College, London, UK

[*]These authors contributed equally to this work.

[◊]To whom correspondence should be addressed: j.christodoulou@ucl.ac.uk

**Abstract**

The ribosome biases the conformations sampled by nascent polypeptide chains along folding pathways towards biologically active states. A hallmark of the co-translational folding (coTF) of many proteins are highly stable folding intermediates that are absent or only transiently populated off the ribosome, yet persist during translation well-beyond complete emergence of the domain from the ribosome exit tunnel. Intermediates are important for folding fidelity; however, their structures have remained elusive. Here, we have structurally characterised two coTF intermediates of an immunoglobulin-like domain by developing comprehensive $^{19}$F NMR analyses using chemical shifts, paramagnetic relaxation enhancement (PRE), and protein engineering. We integrated these experimental data with extensive molecular dynamics (MD) simulations to obtain atomistic structures of the folding intermediates on the ribosome. The resulting intermediate structures are distinguished by native-like folds initiated from either their N- or C-termini, and reveal parallel folding pathways, which are structurally conserved within the protein domain family, in contrast to their *in vitro* refolding mechanisms. By redirecting proteins to fold along hierarchical, parallel routes, the ribosome may promote efficient folding by avoiding kinetic traps, and regulate nascent chain assembly and targeting by auxiliary factors to maintain cellular proteostasis.

**Main**

During biosynthesis, nascent proteins emerge from the ribosome and can begin to acquire their native biologically active structures[1]. The ribosome itself is thought to regulate coTF by modulating the nascent polypeptide chain's thermodynamic stability[2-8] and (un)folding

rates[2,6,9], through vectorial emergence[10,11], steric restriction of the available conformational space[12], interactions with its highly charged surface[5,13-15] and with specific ribosomal proteins[16-18], as well as the recruitment of molecular chaperones[19-22]. In particular, the physical chemistry of the ribosome results in the structural expansion and entropic destabilisation of unfolded nascent proteins[3], and enthalpic destabilisation of negatively charged folded states by long-range repulsive electrostatic forces[23]. Consequently, for negatively charged proteins (~50% proteome[24]), highly stable folding intermediates are the predominant conformer outside the ribosome exit tunnel[2,3,23], where most tertiary folding occurs. The cellular process of coTF therefore contrasts with studies measuring the refolding of full-length denatured protein, where intermediates are unstable, and is reflected by their folding outcomes, with higher propensities to misfold or aggregate in the absence of the ribosome. Detailed descriptions of coTF pathways are absent, although structures of folding intermediates would provide experimentally accessible snapshots of *de novo* protein folding mechanisms. Such structures could therefore permit insights into how the ribosome biases nascent chain conformations away from hazardous states[25-27] and towards conformations that are biologically active[3] or accessible for co-translational assembly[28,29] or chaperoning processes[19].

Co-translational folding intermediates have generally been reported as compacted[30-32], with some observations of native-like structure formation[2,3] within ensemble conformations[15,33,34]. However, nascent proteins inherently interconvert between multiple conformations[2,35], presenting a technical challenge to detailed, experimental measurements of specific intermediate state conformations. Solution-state NMR spectroscopy of ribosome-nascent chain complexes (RNCs) has enabled residue-specific level measurements of different nascent protein conformations by employing specific isotopic labelling schemes[3,5,13,16,36-38], providing experimental data to guide the determination of accurate structures of unfolded[3,13] and folded[23] RNCs from all-atom MD simulations. The direct observation and resolution of multiple partially folded coTF intermediates has uniquely been enabled by $^{19}$F NMR[2,3,16], which has the potential to extract structural information[39].

Here, we have developed an array of generalisable $^{19}$F NMR analyses for structural biology, and obtain direct structural contacts and intramolecular distances within two distinct coTF

65 intermediate conformations populated by an immunoglobulin-like domain on the ribosome.
66 These $^{19}$F NMR experiments validate atomistic structures of the coTF intermediates, obtained
67 using atomistic folding simulations and extensive unbiased MD simulations, and show their
68 conservation across immunoglobulin folds. By mapping detailed coTF pathways, we thus
69 present the structural basis of how ribosomes guide native structure acquisition of one of the
70 commonest protein folds in the proteome.

71

72 **CoTF folding intermediates identified by $^{19}$F NMR**

73 We previously mapped the folding free energy landscapes of the fifth filamin domain of the
74 *Dictyostelium discoideum* gelation factor, FLN5, both off and on the ribosome, using NMR
75 spectroscopy[2,3,5,11,36]. FLN5 is an immunoglobulin-like domain with two sandwiched anti-
76 parallel β-sheets (**Figure 1a**). In isolation, it is a two-state folder under denaturing conditions[2],
77 and only weakly populates intermediates only when truncated at its C-terminus[2,11]. However,
78 when translation-stalled on the ribosome at specific nascent chain lengths (**Figure 1b**) and
79 fully emerged from the tunnel[36], FLN5 populates two highly stable coTF intermediates (I1 and
80 I2, **Figure 1d-e**) in slow exchange with the unfolded and natively folded states (exchange
81 rate $k_{ex} < 2.4\ s^{-1}$, ref.[2]), and detected as broad NMR resonances by site-specific $^{19}$F-labelling
82 using 4-trifluoromethyl-L-phenylalanine[2] (tfmF, **Figure 1c**), and invisible by other isotopic
83 labelling schemes[11]. Measurement of $^{19}$F NMR signal integrals showed that the coTF
84 intermediates are up to 5 kcal mol$^{-1}$ more stable than those found off the ribosome (relative
85 to their respective unfolded states), and persist at long nascent chain lengths during
86 translation[2] (**Figure 1d**), despite complete availability of the domain for native folding to
87 occur[36]. The coTF intermediates adopt a native-like hydrophobic core, evidenced by their
88 complete destabilisation upon mutation of the natively buried residue Tyr719 (ref.[2]); however,
89 further details of their structures are sparse.

90

91 **Structural contacts revealed by $^{19}$F NMR chemical shifts**

92 To begin to extract structural information by $^{19}$F NMR spectroscopy, we examined $^{19}$F
93 chemical shifts and their relationship to protein structure, exploiting our recent development
94 of solvent-exposed tfmF-labels experiencing ring current effects[39], which are distance- and
95 angle-dependent on nearby aromatic residues. Residue 655 on the A-strand directly contacts

the aromatic ring of Phe675 on the neighbouring B-strand; its tfmF chemical shift is consequently shielded from the random coil value by ~0.8 ppm (**Figure 1f,** ref.[2]) and correlated to the ring current effect as shown by the Phe675Ala mutation[39] (**Figure 1f**). The tfmF655 chemical shift is thus a direct reporter of the inter-residue 655-675 sidechain Van der Waals contact[39] comprising a precise, short-range (4-6 Å) distance and angular preference, as quantifiable by structural modelling[39]. The observed broad I2 resonance of tfmF655-labelled FLN5 RNCs shares its chemical shift with the native state, and thus demonstrates a native inter-residue 655-675 tertiary contact and complete A-B strand pairing within this intermediate structure. Conversely, the random coil tfmF-655 chemical shift of I1 (-61.8 ppm) indicates the absence of the native A-B strand contact (**Figure 1f-g**).

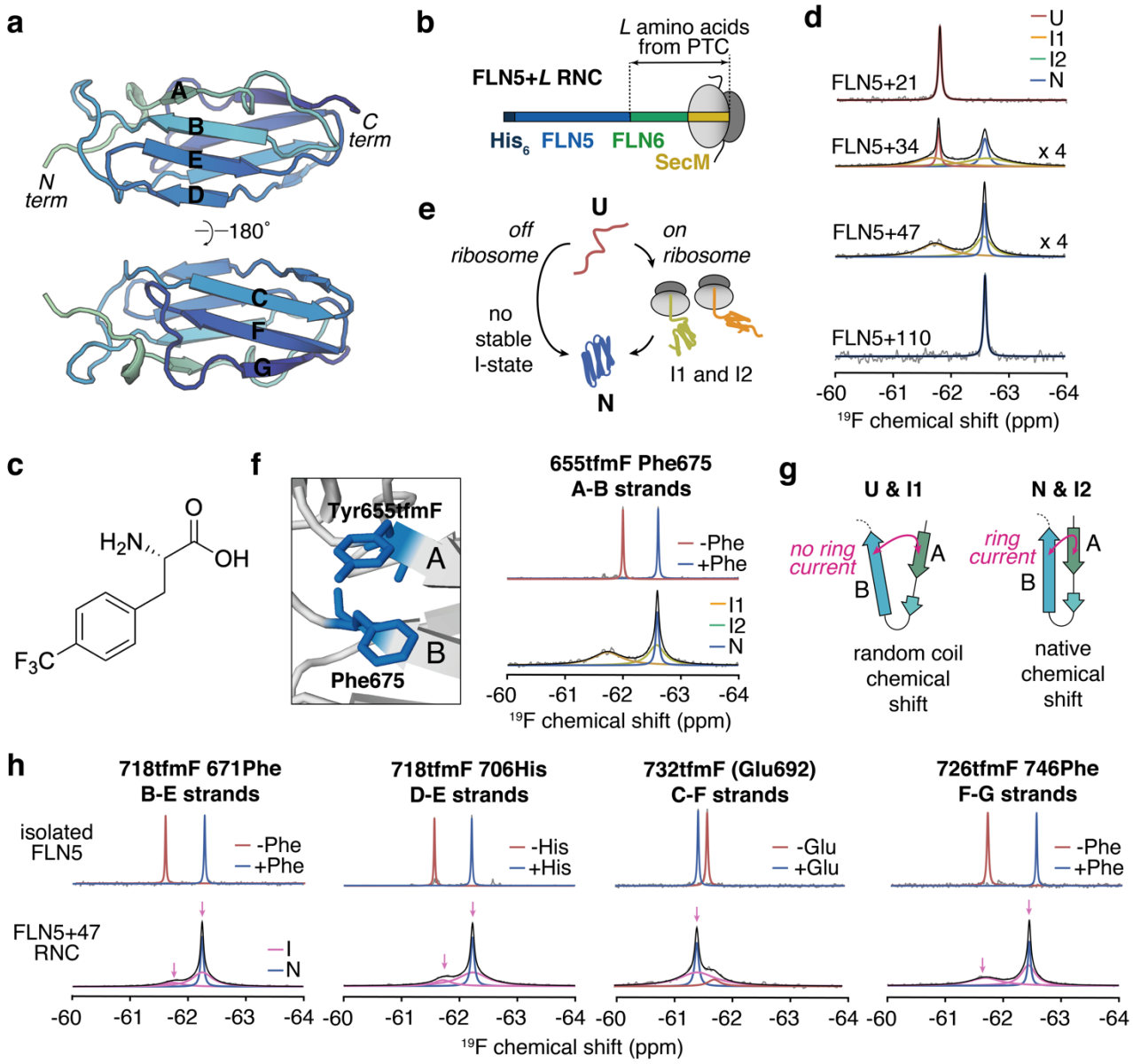


**Figure 1 Structural characterisation of FLN5 coTF intermediates using $^{19}$F NMR chemical shifts. a**, Crystal structure of FLN5 (Protein Data Bank (PDB) code 1QFH) with β-strands labelled. **b**, Schematic of FLN5 RNCs of varying *L* lengths. **c**, Chemical structure of tfmF. **d**, 1D $^{19}$F NMR spectra of different lengths of 655tfmF-labelled FLN5 RNCs. Observed and total fitted spectra are shown in grey and black, respectively. Native (N), unfolded (U) and intermediate state (I1, I2) resonances are indicated. Magnification of spectra by a scaling factor is indicated. **e**, Folding reactions of FLN5 on and off the ribosome. **f**, Left shows magnified view of Tyr655 and Phe675 in the FLN5 crystal structure. Right shows $^{19}$F NMR spectrum of 655tfmF-labelled *(top)* isolated wild-type and Phe675Ala mutant FLN5, and *(bottom)* FLN5+47 RNC. **g**, Schematic depicting ring current between two residues on adjacent β-strands and their corresponding $^{19}$F chemical shifts. **h**, Top shows $^{19}$F NMR spectra of isolated FLN5 with $^{19}$F-label pair, overlaid with that of isolated FLN5 tfmF-labelled without accompanying aromatic/charged residue (left to right, FLN5 718tfmF, FLN5 718tfmF, 732tfmF 692Ala, 726tfmF). Bottom shows $^{19}$F NMR spectra of FLN5+47 RNC with different $^{19}$F-label pairs. Arrows indicate chemical shifts of intermediate states. Observed and total fitted spectra are shown in grey and black, respectively. Assignment of resonances are coloured accordingly. All spectra were recorded at 500 MHz and 298 K.

Additional native ring currents are absent in FLN5, and tfmF-labels at other solvent-exposed sites on FLN5 yield chemical shifts that are poorly resolved between the folded and unfolded (random coil) states[39], and thus do not show clear correlations with (local) protein structure, limiting their use in structural studies. Using ColabFold[40], AlphaFold3 (ref.[41]) and all-atom MD simulations, we therefore rationally designed[39] five additional $^{19}$F-label pairs to engineer ring currents and probe specific sidechain contacts between pairs of β-strands, encompassing the B-E, D-E, and F-G strands (**Figures 1h, S1**), akin to the introduction of (bulkier) Förster resonance energy transfer (FRET) or double electron-electron resonance (DEER) probes. Efforts to generate label pairs for C-F and A'-G using the ring current strategy resulted in more modest chemical shift dispersions insufficient to resolve broad RNC resonances (**Figure S1b**). The successful FLN5 label variants, in isolation of the ribosome, remained two-state folders and showed only mild changes to their folding free energies ($\Delta\Delta G_{N\text{-}U}$ < 0.8 kcal mol$^{-1}$ relative to non-fluorinated, **Figure S1c**), being solvent-exposed; their (de)stabilising effect is expected to be largely suppressed on the RNC due to the ribosome's mutation buffering effect[3,25]. Their $^{19}$F chemical shifts were substantially (de)shielded (> 0.2 ppm) from

140 the random coil and highly sensitive to the presence of the introduced aromatic ring (**Figure**
141 **1h**), as designed. Chemical shifts engineered using ring currents can therefore be interpreted
142 as a residue pair-specific Van der Waals contact[39].

143

144 We introduced each label into FLN5+47 RNC, the length at which the coTF intermediates are
145 maximally populated[2] (44 ± 3% and 38 ± 6% of I1 and I2 respectively for FLN5+47 655tfmF,
146 **Figure 1d**). First, FLN5+47 RNC was labelled to probe the B-E strands (718tfmF 671Phe),
147 yielding $^{19}$F NMR spectra with a sharp (~18 Hz linewidth) resonance at an identical chemical
148 shift to that of the isolated protein (-62.2 ppm), corresponding to the native state as expected
149 (**Figure 1d**), and two additional broad resonances at the native and random coil (-61.8 ppm)
150 chemical shifts attributable to intermediate states (**Supplementary Note 1**). The distinct
151 chemical shifts thus indicate two populations of intermediates, one possessing and one
152 lacking the native B-E strand contact. Similarly, labelling of FLN5+47 RNC at an alternative
153 position across the B-E strands (673tfmF 716His) also produced two broad resonances,
154 having native and random coil chemical shifts (**Figure S1a**), and suggesting that the
155 presence or absence of native contacts persists across the inter-strand network, respectively,
156 and thus complete secondary structure formation.

157

158 When labelled to probe either the D-E (718tfmF 706His) or F-G strands (726tfmF 746Phe),
159 FLN5 RNCs also produced broad resonances attributable to two intermediate states (**Figure**
160 **1h**). In each case, the broad resonances were found at both native and random coil chemical
161 shifts, consistent with alternative label pairs positioned along the same strands (710tfmF
162 716His, and 728tfmF 744His, **Figure S1a**, **Supplementary Note 1**). In addition to the findings
163 for the A-B and B-E strands, native inter-strand contacts across the D-E and F-G strands are
164 therefore only found in one intermediate.

165

166 The 732tfmF (F-strand) chemical shift is highly sensitive to the electrostatic charge of the
167 native neighbouring residue Glu692 (C-strand) in the isolated protein[39] (**Figures 1h, S1a,**
168 **S1e**). In addition to the native state resonance, the $^{19}$F NMR spectrum of FLN5+47 732tfmF
169 showed a single broad resonance (67 ± 11% population) observed at the same chemical shift
170 that was attributable to an intermediate state (**Figure S1a**). Only a lowly populated (11 ± 4%)
171 peak was observed at the random coil chemical shift, having a linewidth (**Table S1**) and small

population (**Table S2**) consistent with an unfolded state, the latter as expected from the more destabilising nature of the 732tfmF-label (relative to other $^{19}$F-label pairs, **Figure S1c**). Similar results were found for the shorter FLN5+34 RNC (**Figure S1a**, **Supplementary Note 1**). These data indicate that both I1 and I2 intermediates possess a native 732-692 contact. Additional structural contacts in the hydrophobic core were measured with natively buried $^{19}$F-labels (**Figure S2**).

Collectively, by using and engineering $^{19}$F chemical shifts that unambiguously probe specific intramolecular contacts, we resolved two distinct intermediate states in slow exchange with the natively folded state (**Supplementary Note 1**), both with native-like structure between the C-F strands, and where additional native contacts across the remaining β-strand pairs and buried core are found in only one intermediate conformation.

**Engineering reversible cross-links stabilises I1**

Equipped with NMR data probing structural features across the intermediate states, we next aimed to attribute the broad resonances to either I1 or I2 to generate a dataset of chemical shifts (i.e. structural contacts) specific to either coTF intermediate. To achieve this, we sought to selectively (de)stabilise one intermediate conformation. Using principles of psi-value analysis[42], we introduced reversible cross-links by engineering bi-histidine metal ion binding sites to stabilise specific β-strands. We experimentally screened for potential inter-strand binding sites by assessing different bi-histidine mutants of isolated FLN5 and their stabilities in increasing concentrations of a range of different metals (**Figure S3a**). Of the seven mutants tested, one variant, 726His 746His (F-G strands), showed a progressive stabilisation of up to -1.45 kcal mol$^{-1}$ in its folding free energy with increasing concentrations of $NiCl_2$ (**Figures 2a-c, S3c**). The metal binding site was confirmed by $^{1}$H,$^{15}$N-NMR, where addition of the paramagnetic Ni(II) metal resulted in increased line broadening localised at the mutation sites (**Figures 2a, S3b-h**), in accord with AlphaFold3 predictions showing coordination between the metal ion and the protein via the engineered histidines and the neighbouring residue Asp744 (**Figure 2a**).

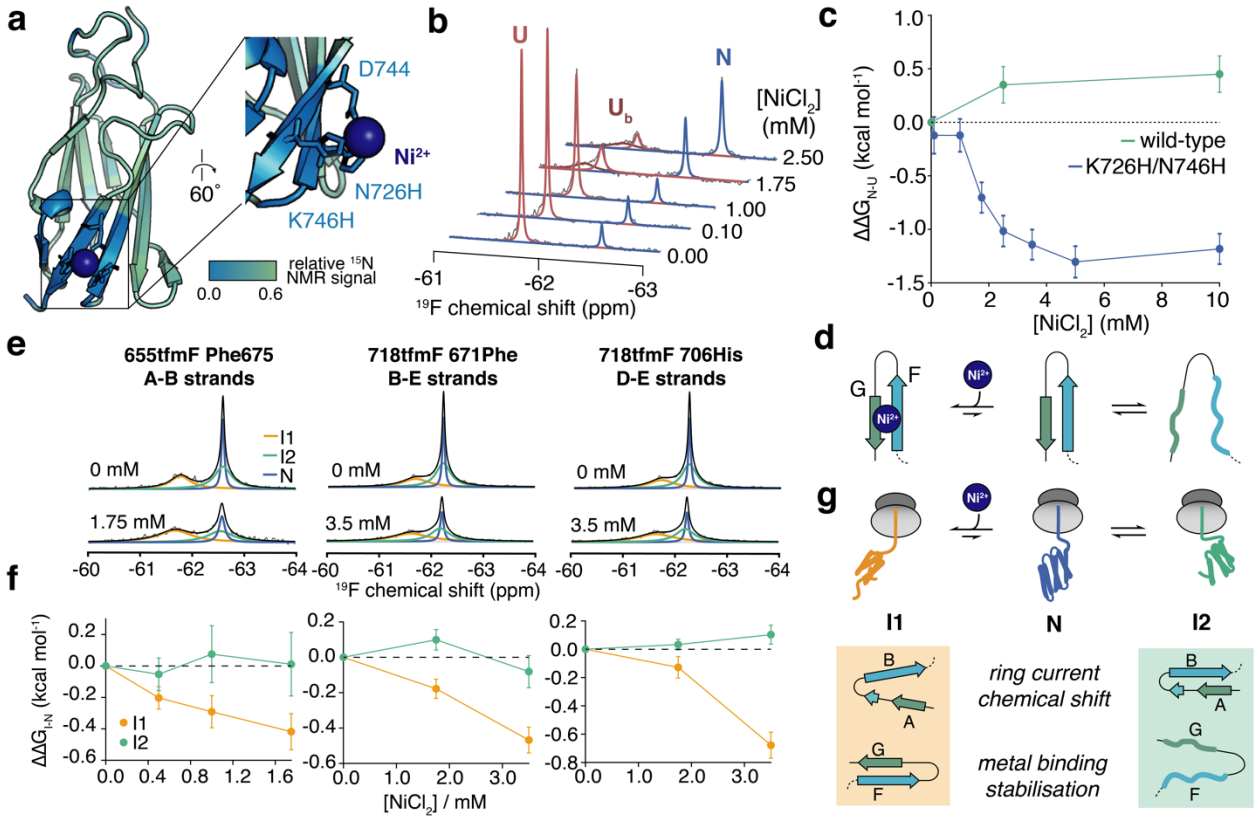



**Figure 2 Stabilising a specific coTF intermediate using an engineered ligand binding site.** **a**, AlphaFold3 predicted structure of FLN5 Asn726His Lys746His in complex with a metal ion, coloured according to the $^{1}$H,$^{15}$N-correlated NMR signal intensity in 2.5 mM $NiCl_2$ relative to without $NiCl_2$ (**Figure S3c** for complete dataset). **b**, $^{19}$F NMR spectra of 655tfmF-labelled FLN5 Asn726His Lys746His incubated in 2 M guanidine hydrochloride and varying $NiCl_2$ concentrations. Observed spectra are shown in grey. Unfolded (U), nickel-bound unfolded ($U_b$, see **Figure S3g**), and native (N) states are in colour and indicated. **c**, Folding free energy changes of isolated wild-type and bi-histidine FLN5 in varying $NiCl_2$ concentrations (relative to without $NiCl_2$), determined by lineshape analysis of spectra shown in **b** and **Figure S3c**. Errors were determined by bootstrapping of residuals from lineshape fits**. d**, Schematic of equilibrium reaction between unfolded and folded β-strand pairs titrated with Ni(II). **e**, $^{19}$F NMR spectra of FLN5+47 Asn726His Lys746His RNCs tfmF-labelled at different positions, with and without 3.5 mM $NiCl_2$ (**Figure S3j** for complete titration data). Observed and total fitted spectra are shown in grey and black, respectively. Assignment of resonances are coloured accordingly. **f**, Changes in Gibbs free energies of intermediate relative to native state, determined by lineshape analysis of spectra shown in e and Figure S3j (**Figure S3j** for complete titration data). Errors were determined by bootstrapping of

221 residuals from lineshape fits**. g**, Schematic of equilibrium reaction between I1, I2 and N
222 titrated with Ni(II). Nickel-binding stabilises I1 more than I2, inferring the presence and
223 absence of the bi-histidine binding site, respectively. I1 and I2 are also distinguished by their
224 ring current-sensitive chemical shifts. All spectra were recorded at 500 MHz and 298 K.

225

226 Having designed a conformation-specific, concentration-dependent ligand binding site
227 (**Figure 2d**), we next introduced the bi-histidine mutations into FLN5+47 RNCs (**Figures 1h,**
228 **S3i-k**). FLN5+47 tfmF655 RNCs (probing A-B strands, **Figure 1f**) yielded resonances
229 corresponding to I1, I2, and N as expected (**Figure 2e**). Addition of increasing amounts of
230 Ni(II) to the RNC sample resulted in increasingly broader resonances, as anticipated from
231 the paramagnetic effect (**Figure S3k**), but critically also a progressive shift in the folding
232 equilibrium towards the I1 state (**Figures 2e-f**, **S3j**), only in the presence of the bi-histidine
233 mutations (**Figure S3j**). An analysis of the signal integrals to determine the populations and
234 Gibb's free energies of each state revealed that I1 was stabilised more than I2 (each relative
235 to N, **Figure 2f**). Given the distinct chemical shifts obtained for the F-G strand $^{19}$F-labels
236 (**Figure 1h**, **S1a**), this indicated the presence of a folded binding site across the F-G strands
237 in I1 that was likely absent in I2 (**Figure 2g**). I1 therefore possesses a native F-G strand
238 conformation, but lacks A-B strand contacts (**Figure 2g**), while I2 exhibits the opposite.

239

240 When the bi-histidine site was introduced in FLN5+47 RNCs possessing the remaining $^{19}$F-
241 label pairs that showed two distinct intermediate resonances (**Figures 1h, S1a**), one
242 intermediate (I1) was consistently stabilised more than the other (I2) when incubated with
243 Ni(II) (**Figures 2e-f, S3j**). These data permitted us to assign the chemical shifts across all the
244 tfmF-labels to either the I1 or I2 intermediate (**Figures 1g, S3j**). Collectively, they show I1
245 likely possesses a natively folded sheet 2 (CFG strands) with native contacts absent across
246 sheet 1 (ABED strands). However, native-like strands are likely to be at least partially formed
247 within sheet 1, since I1 is completely destabilised by the natively buried Tyr719Glu mutation
248 in the core between the two β-sheets[2]. Meanwhile, on the basis of the $^{19}$F NMR contacts, I2
249 appears to closely resemble the intermediate found off in isolated C-terminally truncated
250 variants of FLN5 (ref.[11]), comprising a natively folded A-F strand structure with a disordered
251 C-terminal G-strand; accordingly, native-like chemical shifts for the natively buried sites

252 tfmF715 and tfmF727 also correspond to the I2 intermediate (attributed using the same bi-
253 histidine binding approach, **Figures S2a, S3j-k**).

254

255 **Structures of folding intermediates on the ribosome**

256 Next, we aimed to produce atomistic structural ensembles of the FLN5+47 RNC intermediate
257 states. The structural similarity of I2 to the isolated, truncated intermediate enabled us to
258 produce a starting model of the RNC, by attaching the previously determined structure of the
259 latter[11] to the ribosome (**Supplementary Note 2**). However, in the absence of any analogous
260 structures to I1, we sought to simulate coTF and produce detailed pathways from the
261 unfolded to natively folded state using an all-atom (including ribosome, ions, and solvent)
262 ratchet-and-pawl MD (rMD) approach (**Supplementary Note 2**, **Supplementary Videos 1-**
263 **2**, **Figures S4-6**). The rMD simulations enabled us to successfully capture folding events,
264 which experimentally occur on the timescale of > 0.8 $s^{-1}$ (ref.[2]), beyond the reach of
265 conventional MD simulations. From the obtained coTF trajectories, we identified one on-
266 pathway folding intermediate state (**Figure S5**), which from our experimental dataset
267 appeared to possess structural features common to that of I1. To assess this further, we
268 subsequently subjected the putative I1 and I2 models to extensive atomistic unbiased MD
269 simulations. We found their folded regions to be stable on the microsecond timescale (**Figure**
270 **3**, **Supplementary Videos 3-4**), and therefore compared the atomistic structural ensembles
271 against those previously obtained for the natively folded RNC[23].

272

273 The structures of the putative I1 state (**Figures 3a-b**, **S6a-g**) possess a native-like C-G strand
274 folded core (**Figure 3e-f**); the remaining N-terminal A-, A'- and B-strands are largely detached
275 from the structure, and form either a completely disordered tail, or a β-hairpin (**Figure 3e,**
276 **S6b**). The loss of the A-B strands from the folded structure results in transient non-native
277 contacts, predominantly between the β-sheets at the exposed E-G and A'-F strand interfaces
278 (~20-30%, **Figure S6c-e**). However, we also find a subpopulation (~24%) where the B-strand
279 folds onto the C-G core, with the A-A' strands remaining disordered (**Figure S6b**); this
280 conformational heterogeneity shows multiple accessible pathways from I1 to the native state,
281 as also reflected in our folding simulations (**Figure S5**). Overall, detachment of the N-terminal
282 strands causes an expansion of the structural ensemble, with a larger radius of gyration and
283 increased solvent accessible surface area (SASA) compared to the native state (**Figure 3g**).

Structures of the putative I2 state (**Figures 3c-d, S6a-g**) adopt a more ordered and compact native-like conformation, with a smaller increase in SASA (**Figures 3f-g**). Native contacts are found across the A-F strands (**Figure 3e**), with a non-native trans-Pro742 and disordered G-strand (residues 743-747) detached from the rest of the domain permitting some non-native interactions between A'-F strands (**Figure S6c-f**). We note that energetic calculations show native G-strand folding is dependent on the cis-trans isomerisation state of Pro742 (**Supplementary Note 2, Figure 4**).

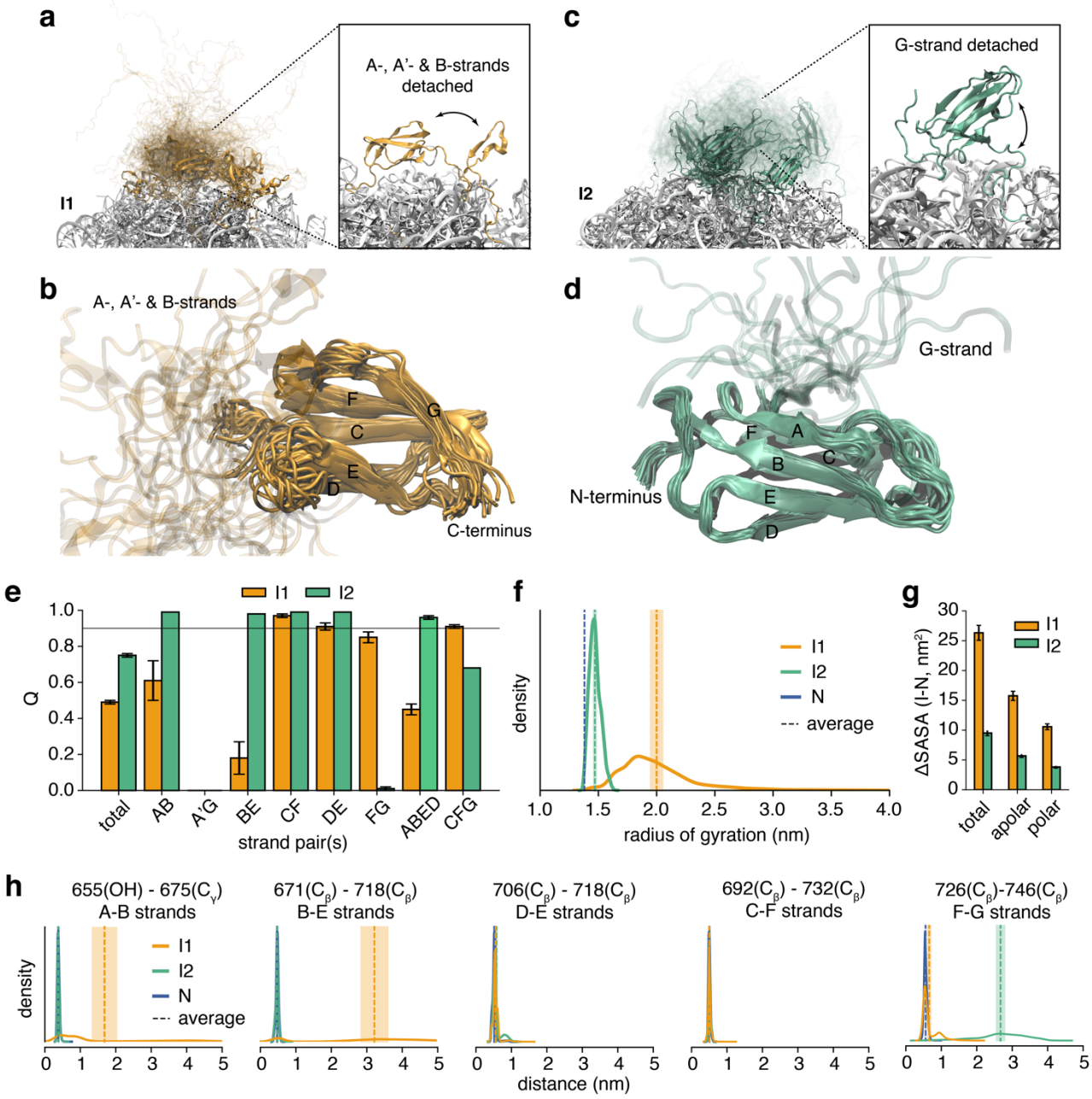


**Figure 3 Atomistic structures of FLN5+47 coTF intermediates on the ribosome.** **a**, Structural ensemble of FLN5+47 I1 on the ribosome, inset shows a representative structure. **b**, Lowest energy structural cluster of I1 (**Table S4**) aligned against the common C-G strand core. **c**, Structural ensemble of FLN5+47 I2 on the ribosome, inset shows a representative structure. **b**, Lowest energy structural cluster of I2 (**Table S4**) aligned against the common A-F strand core. **e**, Average fraction of native contacts (Q) of FLN5 and across different β-strand pairs and β-sheets for I1 and I2 (mean ± s.e.m. from 15 independent simulations of 1 μs). **f**, Radius of gyration probability distribution and mean ± s.e.m. of I1, I2 and native states. **g**, Difference in solvent accessible surface area (SASA) of I1 and I2 relative to the native state. **h**, Distance distributions and mean average ± s.e.m. of residue sidechains corresponding to $^{19}$F-label pairs calculated for I1, I2 and native state structural ensembles.

To assess whether the obtained structures faithfully represent our experimental data (**Figures 1-2**), we examined their intramolecular contacts and compared them against the obtained $^{19}$F chemical shifts, since we engineered the latter to be sensitive reporters of precise, short-range inter-residue distances[39]. We calculated distance distributions between the pairs of residues labelled for NMR analysis (**Methods**, **Figure 3h, S6j**), and found that all could account for the observed chemical shifts of I2: those with a native chemical shift showed short (~0.5 nm), native-like inter-residue distances, consistent with the measured van der Waal's contact, while residues labelling the F-G strands were distanced far (>2 nm) from each other, as reflected by their random coil chemical shifts (**Figures 1f, 1h, S1a**). Analyses for I1 also showed good agreement between inter-residue distances and their ring-current induced chemical shifts (**Figures 3d, S6j**). The exceptions were the two D-E strand labels, but which could instead be resolved and rationalised using more precise, explicit modelling of the $^{19}$F-label pair: the D-E strand pairs form native backbone contacts but lack sidechain interactions in line with the random coil chemical shifts found in this region (**Supplementary Note 3**, **Figure S3l**). Likewise, such simulations entirely validate residue-contacts in the core of the domain (**Supplementary Note 3**, **Figure S6m-n**). In summary, our residue-contact analyses show that both I1 and I2 structures account for all of our $^{19}$F chemical shift data, and thus also that the rMD folding simulations captured an intermediate conformation corresponding to I1.

325 **Interactions and orientations of intermediates with the ribosome**

326 The intermediate state structures show transient interactions with the ribosome surface,
327 predominantly with ribosomal proteins uL24, uL29 and RNA H59 lining the vestibule of the
328 exit tunnel, and more favourably than the native state (**Figure S6h-i**). Their consequently
329 slower calculated rotational correlation times (**Figure S6k**) are reflected experimentally by
330 the increased line broadening of their $^{19}$F NMR resonances[2,3] (relative to the native state,
331 **Table S1**). Locally enhanced binding occurs at their more positively charged N-termini and
332 the loop between the B-C strands, biasing their orientations with the ribosome (**Figure S6h**).
333 This observation is consistent with previous experimental measurements of FLN5 RNCs
334 bearing charge reversal mutations within the two binding sites (Lys646Glu Lys680Glu)
335 showing thermodynamic destabilisation of both I1 and I2 (ref.[2]). The detached C-terminal G-
336 strand in I2 also interacts more strongly with the ribosome surface (**Figure S6h**), a region
337 that has been experimentally shown to bind and stabilise unfolded FLN5 on the ribosome[5].
338 We explored the presence of intermolecular interactions further by directly examining specific
339 contacts between nascent chain residues and ribosomal protein uL24 that lines the exit tunnel
340 using biochemical cross-linking assays. These experiments showed good agreement
341 between (un)observed crosslinks, and intermolecular contacts calculated from the structural
342 ensembles (**Supplementary Note 5**, **Figure S8d-e**).

343

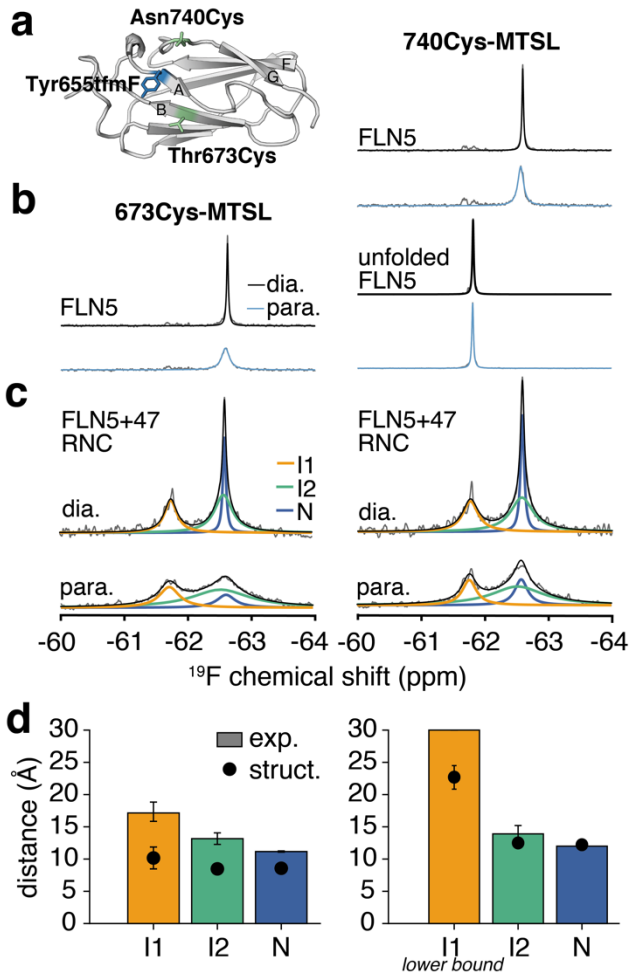
344 Our coTF intermediate structural ensembles were further assessed against our previously
345 obtained cryogenic electron microscopy (cryo-EM) maps of FLN5+47 RNC (Mitropoulou *et*
346 *al.*, in preparation). We performed two orthogonal analyses: firstly by determining cross-
347 correlations of single structures within our intermediate ensembles to the cryo-EM densities,
348 and secondly by producing structural ensembles collectively combining I1, I2 and native state
349 structures (this work, and ref.[23] respectively) reweighted by the cryo-EM densities[43]
350 (**Supplementary Note 5, Figures S8a-c)**. Both approaches showed quantitative
351 improvements to the native state fits when incorporating the intermediate states, and we thus
352 conclude that the cryo-EM maps support the determined intermediate models.

353

354 **$^{19}$F PRE NMR measurements validate intermediate structures**

355 We sought to validate our determined intermediate structures and complement the $^{19}$F
356 chemical shift data by measuring long-range (10-30 Å) distances using $^{19}$F PRE NMR. Having

357 identified detachment of the A-B strands as a key structural feature in I1 (**Figure 3a-b**), we
358 covalently linked paramagnetic spin labels on the adjacent B-strand (Thr673Cys) and F-G
359 loop (Asn740Cys), and measured their effect on the tfmF655 (A-strand) resonance
360 (**Methods**, **Figure 4a**). As expected, the isolated, folded protein showed line broadening in
361 the presence of either paramagnetic label, but only marginal changes when the protein was
362 fully unfolded (**Figures 4b, S1a**).




363

364 **Figure 4 Validation of coTF intermediate structures by $^{19}$F PRE NMR. a**, Location of tfmF-
365 and paramagnetic spin-labels shown on the FLN5 crystal structure (PDB: 1QFH). **b**, $^{19}$F
366 NMR spectra of isolated 655tfmF-labelled FLN5 (pseudo-wild type Cys747Val (**Methods**),
367 and unfolded with destabilising mutations ($A_3A_3$, ref.[36])), spin-labelled at position *(left)* 673Cys
368 and *(right)* 740Cys. Observed spectra are shown in grey. **c**, $^{19}$F NMR spectra of 655tfmF-
369 labelled FLN5+47 RNC (pseudo-wild type Cys747Val), spin-labelled at position *(left)* 673Cys
370 and *(right)* 740Cys. Observed and total fitted spectra are shown in grey and black,
371 respectively. Assignment of resonances are coloured accordingly**. d**, Distances between the
372 $^{19}$F- and spin-labels, bars show experimental values determined by lineshape analysis of

373 spectra shown in **c** (**Methods**), dots show $r^6$-averaged values (mean ± s.e.m.) from the
374 intermediate structures. Errors of experimental values determined by bootstrapping of
375 residuals from lineshape fits. All spectra were recorded at 500 MHz and 298 K.

376

377 The native state resonance was also broadened in paramagnetic spin-labelled FLN5+47
378 RNC. However, lineshape analysis of the resonances (**Methods**) revealed differential effects
379 between the I1 and I2 linewidths and the two spin label sites (**Figures 4c**, **S7b-d**). We used
380 the measured linewidths to calculate distances between the $^{19}$F- and paramagnetic spin
381 labels (**Supplementary Note 4**). Native-like distances between A-B strands and A-loop(FG)
382 were evaluated for I2 (**Figure 4d**), consistent with its determined structural ensemble (**Figure**
383 **3c-d**). In contrast, larger distances were calculated for I1: in particular, a substantial increase
384 (to > 20 Å) for the 740Cys-MTSL spin label (measuring between A-B strands and A-loop(FG))
385 compared to the native state (**Figure 4d**). These data agree with the structures (**Figure 3a-**
386 **b**), where the A-strand remains detached from the C-terminus across all conformations, while
387 the formation of a β-hairpin in subpopulations of I1 rationalises the shorter ensemble-
388 averaged distance between A-B strands.

389

390 Thus collectively, the NMR chemical shifts (**Figures 1-2**), cross-linking (**Figure S8d**), PRE
391 (**Figures 4**, **S7**) and cryo-EM data (**Figure 8a-c**) confirm that the determined structures of I1
392 and I2 are highly representative of the experimental data.

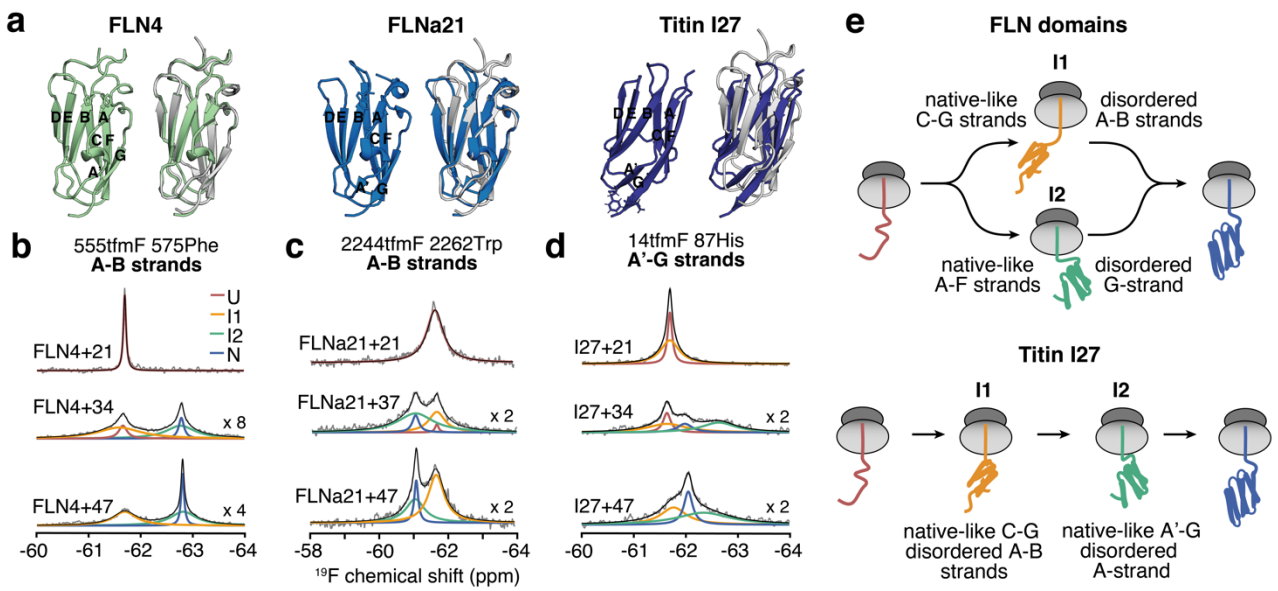
393

394 **Conservation of intermediate structures across filamin domains**

395 Finally, we sought to examine whether the intermediates of FLN5 are conserved across the
396 coTF pathways of domains with a similar fold. We therefore studied the folding of three
397 additional immunoglobulin-like domains (of a multi-domain protein) on the ribosome (**Figure**
398 **5a**): FLN4, the domain preceding FLN5 (47% identity relative to FLN5); FLNa21, the 21$^{st}$
399 domain of human filamin A (the human homologue of gelation factor, 31% identity); and I27,
400 the 27$^{th}$ immunoglobulin-like domain of titin (24% identity), whose smaller size (89 vs 105
401 residues compared to FLN5) is accompanied by a subtly different topology. For each protein,
402 we rationally designed and experimentally validated $^{19}$F-label pairs with ring currents in the

403 natively folded state[39], permitting us to assess inter-strand contacts across each
404 immunoglobulin-like fold.

405




406

407 **Figure 5 $^{19}$F NMR of RNCs of immunoglobulin domains show conservation of filamin**
408 **coTF pathways. a,** Structures of FLN4 (AlphaFold3), FLNa21 (PDB: 2W0P), and titin I27
409 (PDB: 1TIT), and alignment with FLN5 structure (PDB: 1QFH). **b-d**, $^{19}$F NMR spectra of
410 555tfmF 575His-labelled FLN4+*L*, 2244tfmF 2262Trp-labelled FLNa21+*L*, and 14tfmF 87His-
411 labelled I27+*L* RNCs, with linkers of *L* amino acids deriving from the subsequent domain.
412 Observed and total fitted spectra are shown in grey and black, respectively. Assignment of
413 resonances are coloured accordingly. Magnification of spectra by a scaling factor is indicated.
414 **e**, Summary of $^{19}$F NMR data of other labelling sites, shown as schematic of coTF
415 intermediates and pathways of the filamins (FLN4, FLN5, and FLNa21) and I27 (all NMR data
416 shown in **Figure S9a-c**). All spectra were recorded at 500 MHz and 298 K.

417

418 We first aimed to identify nascent chain conformations populated on the ribosome, and
419 therefore produced a library of RNCs varying in length *L* of the subsequent domain (domain+*L*
420 RNC). FLN4 RNCs $^{19}$F-labelled at the A-B strands (555tfmF 575Phe) showed a sharp
421 resonance with a random coil chemical shift at the shortest RNC length (FLN4+21), which
422 reduced in population with increasing RNC length, and thus attributable to the unfolded state;
423 concurrently, a sharp resonance with a native chemical shift progressively increased in signal
424 intensity and was ascribed to the natively folded state (**Figure 5b**, **S9a**). Importantly, two

425 broad resonances, maximally populated at intermediate RNC lengths, were also found at
426 each of the same chemical shifts (**Figure 5b**, **S9a**). Thus, identically to FLN5 (**Figure 1f**),
427 FLN4 populates two coTF intermediates on the ribosome, distinguished by the presence of
428 A-B strand contacts. Indeed, analysis using other $^{19}$F-label pairs (B-E, F-G strands) indicate
429 that the same inter-residue contacts are found in FLN4 as in FLN5 coTF intermediates
430 (**Figure 5e**, **S9a**). We also probed the A'-G strand, using a uniquely successful $^{19}$F-label pair
431 in FLN4 (565tfmF 646His, **Figure S9a**), and found the absence of contacts for either
432 intermediate state; these data are consistent with the determined structures of FLN5 I1 and
433 I2, having detached A'- and G-strands respectively (**Figure 3a-d**). Similar $^{19}$F NMR analyses
434 of FLNa21 RNCs (**Figures 5c, 5e, S9b**) also conclude the population of two coTF
435 intermediates with the same structural features as those of FLN5. We therefore find that the
436 determined structures of FLN5 coTF intermediates are conserved across the folding of filamin
437 domains.

438

439 We have previously identified two coTF intermediates of titin I27 on the ribosome[3], and show
440 here that their populations alter with RNC length, as expected (**Figures 5a, 9c**). However, in
441 contrast to the filamin domains, $^{19}$F NMR spectra of I27 RNCs labelled across the A'-G
442 strands (14tfmF 87His) reveal an intermediate state resonance (I1) at the random coil
443 chemical shift, but interestingly also at a significantly more shielded chemical shift to that of
444 the native state (I2). A weakly populated intermediate having the same chemical shift was
445 also found in a destabilised variant of I27 (Phe73Ala) in isolation[3] (**Figure S9c**), suggesting
446 that its conformation was similar to the ribosome-bound I2 intermediate, and whose structural
447 features have previously been examined[44,45]. Indeed, all-atom MD simulations of the isolated
448 I27 intermediate reconcile the observed chemical shifts and contacts of the I2 intermediate
449 (**Supplementary Note 6**). Overall, the I27 coTF intermediates I1 and I2 are defined by
450 detached A-B and A-strands (respectively) from a common native-like β-sandwich core,
451 comprised of β-strands C-G (**Figure 5e**). Thus, while the filamin I1 state is likely conserved,
452 I2 is subtly distinct to those adopted by filamin domains.

453

454 **Discussion**

455 We have developed a range of $^{19}$F NMR analyses and used these together with MD
456 simulations to determine the first atomistic structural ensembles of coTF intermediates on the

457 ribosome. A particular strength of $^{19}$F NMR, and crucial to obtaining direct, precise structural
458 contacts, is its ability to resolve interconverting conformers[2]. We extracted short- and long-
459 range distances from $^{19}$F NMR chemical shifts (**Figures 1, S1-2**) and PRE measurements
460 (**Figures 4, S7-8**) respectively, and obtained systematic structural analyses specific to each
461 intermediate state using a psi-value-based analysis (**Figures 2, S3**). Our atomistic
462 simulations of FLN5 on the ribosome reveal stable, partially folded intermediate state
463 conformations with common structural features (**Figure 3**), which show striking and entire
464 consistency with the experimental NMR, cryo-EM, and biochemical data (**Figure S8**). Overall,
465 our study advances the utility of $^{19}$F NMR in structural studies, and establishes a
466 generalisable approach to determine and validate atomistic structures of large
467 macromolecular complexes.

468

469 Our structural ensembles of I1 and I2 show a native-like, reduced β-sandwich core, with
470 detached A-B or G strands indicative of folding initiating at either the C- or N-terminus
471 respectively. Thus, the two coTF intermediates are likely populated along parallel, rather than
472 sequential, folding pathways. Indeed, only one intermediate is populated along the same
473 folding trajectory sampled by rMD simulations (**Figure S5**), while NMR data show that both
474 intermediates are simultaneously, rather than successively, populated at varying RNC
475 lengths[2] (**Figure 1d**). We note that energetic calculations (**Figure S4**) and NMR
476 measurements[11] show G-strand folding is dependent on the isomerisation state of the
477 preceding Pro742, which adopts the trans conformation in I2 with a disordered G-strand, but
478 a cis conformation in I1 and N where the G-strands are natively folded (**Figure 3**); thus, the
479 cis-proline likely acts as a molecular switch of the conformational state of the C-terminal
480 strand[11] (**Figure S5**).

481

482 Notably, both the natively cis-Pro742 (ref.[11]) and coTF pathways are conserved across
483 filamin domains (**Figure 5**). In contrast, titin I27 appears to likely fold via two sequential
484 intermediates instead; its structure differs from the filamins (**Figure 5a**), having a smaller
485 size, different topology, and lacking the conserved cis-proline. Thus, the parallel folding
486 pathways identified are (only) conserved across filamins. This observation is reminiscent of
487 the ribosome's ability to thermodynamically[3] and structurally[25] buffer the effects of
488 (de)stabilising point mutations (FLN4 or FLNa21 being substantially mutated variants of

FLN5). The presence of parallel pathways on the ribosome via two possible intermediates has previously been inferred[46], and could provide redundancy of mechanistic routes towards the native state, to improve folding efficiencies and provide resistance against external perturbations. Indeed, the rMD and structural ensemble of I1 reveal two additional pathways to the native fold, either via attachment of an A-B strand β-hairpin, or successive folding of the B- and A-strands to the folded core. Furthermore, the conserved coTF folding pathways are in contrast to un/refolding experiments of the isolated proteins[47], which have shown distinct mechanisms for FLN4 and FLN5, indicating that the ribosome guides folding of similar protein structures via a common, generalised pathway. Interestingly, I27 has been reported to refold via two distinct, parallel transition states and pathways off, in contrast to a sequential pathway on the ribosome (this work), and whose flux is altered with denaturant concentration and mutations[48]; in a similar manner, the ribosome appears to alter the flux towards specific protein folding pathways.

The long-lived coTF intermediates appear to share structural similarities to those found transiently off the ribosome by FLN4 (ref.[47,49]), N-terminal fragments of FLN5 (ref.[2,11]), I27 (ref.[44,45]), and those of refolding transition states of immunoglobulin-like domains[50]. This observation suggests that isolated intermediate conformations are explored on the ribosome, but novel conformational states not observed in isolation also become thermodynamically accessible. The thermodynamic basis for stable coTF intermediates originates from the destabilisation of both the unfolded[3] and folded states[23], and so hierarchical folding via partially folded intermediates is likely one means to minimise repulsive negatively charged globular structures near the ribosome surface (**Table S5**). Our structures also highlight the range of intermediate conformational states that can be stabilised. The high degree of disorder in the I1 coTF intermediate indicates that conformational entropy substantially contributes to its thermodynamic stability. In contrast, the stability of the more native-like I2 structure is likely to benefit from larger enthalpic contributions from intra- and intermolecular interactions. The distinct structures of I1 and I2, yet similar folding free energies[2], indicate that the loss of extensive solvation of the unfolded state to form (partially) folded structures can be compensated through a complex array of different entropic or enthalpic contributions.

520 The thermodynamic stabilisation of such intermediates on the ribosome persists for long
521 durations of translation[2], and for FLN4, whose intermediates are obligate for folding[47], even
522 after translation of the subsequent domain (**Figure S9a**), ensuring step-wise, hierarchical
523 acquisition of native structure under quasi-equilibrium conditions[32]. In the context of multi-
524 domain folding, the population of structurally expanded intermediate states of one domain
525 may be advantageous to the thermodynamic stability of the preceding (fully translated)
526 domain, since this would effectively increase the distance of the preceding domain from the
527 destabilising, repulsive ribosome surface[23], while also limiting exposure of misfolding-prone
528 unfolded states. Such prolonged lifetimes of intermediates, in contrast to highly transient
529 populations off the ribosome, may also provide an opportunity for chaperones[19-21], co-
530 factors[51], enzymes, and quality control machinery[52] to survey, engage, and modify[53] specific
531 nascent chain conformations, or for recruitment of other nascent proteins during co-
532 translational assembly[28,29]. Intriguingly, filamin function is highly regulated by serine
533 phosphorylation at consensus sequences in humans[54], comprising conserved, natively-
534 inaccessible serine-cis-proline motifs preceding the G-strand. Strand-detachment and a
535 trans-proline in the I2 intermediate exposes the phosphorylation site, potentially implicating
536 the formation of stable coTF intermediates to co-translational modifications[55], and could
537 therefore alter the domain's post-translational function by stabilising and trapping the domain
538 in its phosphorylated I2 state.

539

540 Future detailed characterisations of nascent chain structure will permit mechanistic insights
541 into folding pathways of other protein folds, misfolding[25,27,56,57] and other co-translational
542 events, offering prospects for both deeper structural insights and targeted modulation of such
543 processes with therapeutics.

544

545 **Methods**

546 Sample preparation

547 DNA constructs of FLN5, FLN4, FLNa21, and titin I27 were previously described[2,3,36,39].
548 Mutations and amber stop codons were introduced using standard site-directed mutagenesis
549 procedures, and constructs verified by DNA sequencing. For PRE NMR experiments, the
550 additional mutation Cys747Val was introduced to yield a cysteine-free background for site-
551 specific spin-labelling, as previously described[3]. Isolated proteins were expressed as His-

552 tagged proteins (FLNa21 as GST-fused protein[39]), isotopically labelled in *Escherichia coli*
553 BL21 DE3-Gold cells (using amber suppression for incorporation of tfmF), and purified as
554 previously described[2,36,39]. RNC constructs comprised a His-tag for purification and an arrest-
555 enhanced variant of the SecM stalling sequence, FSTPVWIWWWPRIRGPP, as previously
556 described[2]. RNCs were expressed, isotopically labelled, and purified as previously
557 described[2]. For inter-molecular cross-linking experiments, RNCs were expressed in
558 CRISPR-modified *E. coli* BL21 strains with cysteine mutations in uL24, as previously
559 described[3]. MTSL spin-labels for PRE NMR were introduced into purified samples, as
560 previously described[3]. Sample purity and integrity were assessed using SDS-PAGE and
561 western blot analyses with an anti-hexahistidine horseradish peroxide-conjugated antibody
562 (Invitrogen, 1:5000 dilution), as previously described[58].

563 NMR spectroscopy

564 NMR samples were prepared in Tico buffer (10 mM Hepes, 30 mM $NH_4Cl$, 12 mM $MgCl_2$,
565 1 mM EDTA), supplemented with 10% (v/v) $D_2O$ and 0.001% (w/v) DSS as a reference. NMR
566 data were recorded using a 500- ($^{19}F$ and $^{1}H$,$^{15}N$ NMR) or 800-MHz ($^{1}H$,$^{15}N$ NMR) Bruker
567 Avance III spectrometer, both equipped with a TCI cryoprobe, and Topspin 3.5pl2, and at
568 298 K, unless stated otherwise. Data were processed using NMRPipe (v11.7)[59], CCPN
569 (v2.4), MATLAB (R2017a, Mathworks), and Julia 1.5 (ref.[60]).

570

571 1D $^{19}F$ pulse-acquire experiments were recorded using a 350-ms acquisition time and 1.5-s
572 recycle delay. Multiple short experiments were recorded in succession to monitor changes in
573 RNC sample integrity, and data indicating nascent chain release or sample degradation
574 through changes in linewidths, signal intensities or chemical shifts were not used to produce
575 the final summed spectrum, as previously described[2,3]. Processed spectra were baseline-
576 corrected, and peaks fit to Lorentzian lineshapes, with errors of the linewidths and integrals
577 (populations) estimated using bootstrapping (200 iterations, calculating the standard error of
578 the mean), as previously described[2]. The time-domain $^{19}F$ NMR spectra were multiplied with
579 an exponential window function with a line broadening factor of 5 (isolated protein) or 10 Hz
580 (RNC), unless stated otherwise, prior to Fourier transformation. Integrals were used to
581 calculate folding free energies using $\Delta G = -RT\ln(K_{eq})$. 2D $^{1}H$,$^{15}N$ SOFAST-HMQC
582 experiments[61] were recorded using an acquisition time of 50 ms in the direct dimension, and

583 inter-scan delays of 100 ms. Cosine-squared window functions were used to process the
584 spectra.

585

586 $^{19}$F PRE rates were measured using 1D pulse-acquire experiments with samples in the
587 paramagnetic and diamagnetic state, the latter by addition of 2.5 mM (RNC) or 100 x molar
588 excess (isolated protein) sodium ascorbate. For PRE measurements of RNCs, spin-labelled
589 samples were prepared and split into two halves; one half was immediately recorded by NMR
590 at 298 K to produce the paramagnetic spectrum, while the other half was incubated at 277
591 K. Once the paramagnetic spectrum was acquired (typically 16-24 h), the latter half sample
592 was reduced in sodium ascorbate and the diamagnetic spectrum acquired at 298 K (typically
593 also 16-24 h). To improve signal-to-noise, 5-6 separate RNC samples were prepared and the
594 datasets summed together (using only data corresponding to intact samples) to produce the
595 final spectra. The spectra were fitted to Lorentzian lineshapes, as previously described[2],
596 which provide accurate measurements of transverse relaxation rates ($R_2$) and identical to
597 those obtained from direct $^{19}$F $R_2$ measurements[3]. Line broadening resulting from the PRE
598 effect resulted in highly overlapped I2 and native state resonances, which were deconvolved
599 by restraining the lineshape fit of the native state using the expected PRE rate bounds
600 (**Supplementary Note 4**).

601 AlphaFold structure predictions

602 AlphaFold3 (ref.[41], https://alphafoldserver.com) was used to predict structures of FLN5
603 bihistidine mutants in complex with a metal ion, using Co(II) to replace Ni(II) that is otherwise
604 unavailable. AlphaFold3 was also used to predict the structure of flexible regions of FLN4
605 that are absent in its crystal structure (PDB: 1WLH). Default settings were applied.

606 Cross-linking experiments

607 Intermolecular cross-links were introduced into FLN5+47 RNC using a similar approach as
608 described for PRE spin-labelling[3]. Samples were reduced overnight at 277 K in Tico
609 supplemented with 2 mM TCEP. TCEP was then removed by buffer exchange into labelling
610 buffer (50 mM HEPES, 12 mM $MgCl_2$, 20 mM $NH_4Cl$, 1 mM EDTA, pH 8.0). The RNC sample
611 was then incubated with variable concentrations of the cross-linker 1,8-bismaleimido-
612 diethyleneglycol ($BM(PEG)_2$) at 298 K, and the reaction quenched with DTT. Cross-linking
613 was assessed by a mass increase of ~16 kDa (i.e. the mass of uL24) by western blot.

614 Proline free energy calculations

615 We used well-tempered metadynamics[62] (WT-METAD) simulations to calculate the free
616 energy landscapes capturing the isomerisation state of FLN5 proline P742. All-atom
617 simulations were run using the DES-Amber force field and TIP4P-D water model[63,64]. The
618 calculations were prepared and performed with GROMACS (version 2020)[65] and PLUMED
619 (version 2.6)[66,67]. Three different models were used to investigate the effect of local structure
620 on P742 isomerisation energetics: an unfolded peptide with P742 (underlined) in the middle
621 of the sequence corresponding to FLN5 residues N736-I748 (Ac-NPIKNMPIDVKCI-NMe), a
622 C-terminally truncated FLN5 construct, FLN5 Δ6, populating a folded and disordered C-
623 terminus at equilibrium[11] and full-length, folded FLN5 (PDB 1QFH, residues K646-G750)[68].
624 Simulations of the unfolded peptide were initiated from a linear chain built in PyMOL 2.3
625 (Schrödinger, LLC). All systems were initially centred in a dodecahedral simulation box at
626 least 1.2 nm away from the box edge, solvated and neutralised with $MgCl_2$. Energy
627 minimisation of the solvated systems was performed using the steepest descent algorithm
628 using a tolerance of < 1000 kJ $mol^{-1}$ $nm^{-1}$. Equilibration was then performed for 1 ns in the
629 NVT and 1 ns in the NPT ensemble, where the temperature (298 K) was maintained using
630 velocity rescaling[69] with a coupling constant of 0.1 ps and the pressure (1 bar) was regulated
631 with the Berendsen barostat[70] a compressibility of $4.5 \times 10^{-5}$ $bar^{-1}$ and coupling constant of 2
632 ps. During equilibration all protein heavy atoms were position restrained using a force
633 constant of 1000 kJ $mol^{-1}$ $nm^{-2}$. Position restraints were then removed and pressure coupling
634 switched to the Parrinello-Rahman method[71] with the same compressibility and coupling
635 constant. A time step of 2 fs was used together with the leap-frog integrator and the LINCS
636 algorithm[72] to constrain all bonds involving hydrogen atoms. Production simulations with WT-
637 METAD were run in triplicates from different initial velocities for 1 μs each. Van der Waals
638 interactions were computed up to a cut-off distance of 0.9 nm, and short-range electrostatics
639 up to 1 nm. Long-range electrostatics were calculated using the Particle Mesh Ewald (PME)
640 method[73] with cubic interpolation and a grid spacing of 0.125 nm.

641

642 As collective variables for WT-METAD we used two dihedral angles, $\zeta$ ($C\alpha_{i-1} - O_{i-1} - C\delta_I -$
643 $C\alpha_i$) and $\psi$ ($N_i - C\alpha_i - C_i - N_{i+1}$), where *i* corresponds to P742. $\zeta$ describes the isomerisation
644 state (~ 0 for cis and ~± $\pi$ for trans in radians) and $\psi$ describes the the C-terminal amide
645 orientation. Previous work has shown this combination to be efficient for simulating proline

646 isomerisation with WT-METAD[74]. Gaussians were deposited every 1 ps with a width of 0.07
647 rad for each dihedral angle. The Gaussian height and bias factor were set to 0.8 kJ $mol^{-1}$ and
648 10, respectively. We used the method described by Tiwary *et al.* to reweight all snapshots to
649 correct for the WT-METAD bias potential (saved every 5 ps)[75], discarding the first 200 ns
650 (20%) of each trajectory. All ensemble properties including cis/trans populations were then
651 calculated using these weights, with the mean and standard error of the mean calculated
652 from the three independent WT-METAD simulations initiated with different starting velocities.
653 For full-length FLN5 and FLN5 Δ6, we used a collective variable to quantify the fraction of
654 native contacts (Q) formed by the two downstream residues I743 and D744 to correlate C-
655 terminal structure with P742 isomerisation. SMOG2[76] was used to obtain a list of native heavy
656 atom contacts and Q was calculated using the following switching function[77]:

657 $$\boldsymbol{Q(X)} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{ij}} \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{\left(\boldsymbol{\beta}\left(\boldsymbol{r_{i,j}} - \boldsymbol{\lambda r^{0}_{i,j}}\right)\right)}}$$

658

659 The distance between two heavy atoms is defined as $r_{i,j}$, whereas $r^0_{i,j}$ corresponds to the
660 reference distance. $\beta$ and $\lambda$ were set to 50 $nm^{-1}$ and 1.8, respectively. WT-METAD
661 simulations of full-length FLN5 were restrained to have a C-terminal Q (CTERM) of > 0.8
662 using a lower wall restraint in PLUMED (force constant = 10,000 kJ $mol^{-1}$) to maintain the
663 native structure of FLN5 during the simulations.

664

665 All-atom simulations of protein folding

666 We used ratchet-and-pawl MD (rMD) simulations[78-80] to sample folding transition paths
667 starting from unfolded conformations. For isolated FLN5, we used SMOG2[76] to generate a
668 simple structure-based model without any contacts (i.e., no nonbonded attractive potentials)
669 and unfold FLN5. 50 random unfolded conformations were then picked for rMD simulations.
670 Using the DES-Amber force field and TIP4P-D water model[63,64,81], these structures were then
671 energy minimised and equilibrated using an equivalent protocol as detailed above at 350 K.
672 For the FLN5+47 RNC, we performed thermal unfolding simulations starting from 50 different
673 conformations of folded FLN5+47 on the ribosome obtained in previous work[23]. The RNC
674 was simulated with using the DES-Amber force field and TIP4P-D water model[63,64,81],
675 because we have shown that this force field results in good agreement of folded FLN5

simulations with NMR data[23]. For thermal unfolding and subsequent relaxation runs, we used hydrogen mass repartitioning[82] and a 4 fs timestep with LINCS. GROMACS 2020 and PLUMED 2.6 were used for these simulations. Temperature and pressure coupling settings were identical to the production runs used for the proline free energy calculations described above. All ribosome heavy atoms (and C-terminal nascent chain residue) were position restrained at all times using a force constant of 1000 kJ $mol^{-1}$ $nm^{-1}$. Unfolding was performed at 400 K and 1 bar for 25 ns in the presence of lower wall restraints on the radius gyration calculated from residues 646-750 C$\alpha$ atoms at 3 nm with a force constant of 10,000 kJ $mol^{-1}$. The total fraction of native contacts, as defined in the previous section, was restrained to be below 0.02 using a force constant of 10,000 kJ $mol^{-1}$. To ensure all starting structures had a trans peptide bond for P742 (P742 is in the cis conformation in the native structure), we additionally restrained the absolute value of the $\zeta$ dihedral angle to be above 2.62 rad (~150˚). These unfolded structures were then relaxed at 350 K and 1 bar for 1 ns in the absence of wall restraints (but with ribosome position restraints).

Enhanced sampling is achieved with the rMD method (adiabatic-bias MD, ABMD, in PLUMED). We used GROMACS 2020 and PLUMED 2.6 for rMD simulations. In this approach, simulations proceed in the absence of a biasing potential when spontaneous progression towards a pre-defined target state (i.e., the folded state of FLN5) occurs. However, a soft biasing potential is applied when the system moves in the opposite direction[78,79]:

$$\boldsymbol{F_{rMD}} = \begin{cases} \frac{\boldsymbol{\kappa}}{\mathbf{2}}\left(\boldsymbol{\rho(t)} - \boldsymbol{\rho_m(t)}\right)^{\mathbf{2}}, \boldsymbol{\rho(t)} > \boldsymbol{\rho_m(t)} \\ \mathbf{0}, \boldsymbol{\rho(t)} \leq \boldsymbol{\rho_m(t)} \end{cases}$$

$$\boldsymbol{\rho(t)} = \left(\boldsymbol{CV(t)} - \boldsymbol{CV_{target}}\right)^{\mathbf{2}}$$

$$\boldsymbol{\rho_m(t)} = \min_{\mathbf{0}\rightarrow \boldsymbol{t}}(\boldsymbol{CV})$$

The force constant, $\kappa$, controls the magnitude of the bias potential, *CV(t)* is the instantaneous value of a defined collective variable at time, *t*, and $CV_{target}$ is its value in the target (folded) state. rMD therefore effectively promotes transitions from an initial (unfolded) state to the final (native state) using a reaction coordinate. In our rMD simulations, we set $\kappa$ and $CV_{target}$ to

100 kJ mol$^{-1}$ and 0, respectively. As a collective variable, we used the difference contact map between a given conformation and the native state[78,79], defined by heavy atom native contacts, *j,* within a 0.6 nm distance cut-off obtained the shadow algorithm implemented in SMOG2[76]:

$$\boldsymbol{CV} = \sqrt{\sum_{j=1}^{n} \left( \boldsymbol{C_j(X)} - \boldsymbol{C_j(X_{native})} \right)^2}$$

$$\boldsymbol{C_i(X)} = \frac{\boldsymbol{1} - \left(\frac{\boldsymbol{r_j}}{\boldsymbol{r_0}}\right)^{\boldsymbol{6}}}{\boldsymbol{1} - \left(\frac{\boldsymbol{r_j}}{\boldsymbol{r_0}}\right)^{\boldsymbol{10}}}$$

Here, $r_j$ is the native contact distance for atom pair *j* and $r_0$ was set to 0.5 nm. For FLN5, this approach includes 1,009 heavy atom contacts calculated using the energy minimised crystal structure of FLN5. For all 50 starting structures, we ran 36 independent rMD simulations starting from randomly assigned initial velocities for a time of 5 ns, totalling 1800 trajectories and 9 μs of biased sampling. Simulations were run in the NPT ensemble at 350 K, using the same temperature and pressure coupling parameters as described above. The integration time step was 2 fs. For the RNC, position restraints on the heavy atoms of the ribosome and C-terminal NC residue at the peptidyl transferase centre were kept at 1000 kJ mol$^{-1}$ nm$^{-1}$. We then calculated the RMSD over C$\alpha$ atoms of residues 646-750 for all trajectories with respect to the energy minimised crystal structure of FLN5. Trajectories reaching an RMSD of less than 0.2 nm were defined as successful folding transition paths, resulting in 321 and 226 successful folding transitions for isolated FLN5 and FLN5+47 RNC, respectively. These trajectories were analysed further and we calculated the total fraction of native contacts and native contacts between specific structural elements of the protein to project the transition path landscape onto these variables to investigate the folding pathway and intermediate states.

730 Unbiased all-atom simulations of coTF intermediates

731 To select starting structures of the I1 coTF intermediate, we considered all potential
732 structures belonging to the local free energy minimum observed in rMD simulations between
733 a fraction of a global Q of 0.55 < Q < 0.68 (**Figure S5b-f**). Given this intermediate free energy
734 minimum corresponds to structures with a folded G-strand (**Figure S5f**) and our proline free
735 energy calculations show that P742 is in the cis configuration in this case (**Figure S4**), we
736 only selected structures with P742 in cis. To ensure the selection of physically reasonable
737 structures, we sorted the structures by ascending distance RMSD (dRMSD, relative to the
738 energy-minimised native state crystal structure) of all heavy atoms belonging to residues
739 F691-G750. These residues correspond to the natively folded region of this intermediate state
740 predicted by rMD (**Figure S5c**). The top 100 structures were then clustered using the
741 GROMOS algorithm[83] implemented in gmx cluster, using all-atom RMSD of the entire
742 nascent chain without alignment and a cut-off of 0.9 nm to select structures with different
743 orientations on the ribosome. The top 5 cluster centroids (**Figure S5g**) were then used for
744 unbiased MD simulations.

745

746 Since rMD simulations only sampled one intermediate (I1, **Figure S5**) and due to the
747 structural similarity between the I2 coTF intermediate and an intermediate previously
748 characterised for isolated FLN5 $\Delta 6$[3,11,38], we randomly selected five structures from the lowest
749 free energy cluster of the NMR-derived structural ensemble of isolated FLN5 $\Delta 6$[11]. We then
750 built the C-terminal residues of FLN5 in an extended conformation and joined these models
751 with the NC linker for FLN5+47 used in previous work[23]. Unbiased MD simulations were then
752 initiated from these structural models.

753

754 For all input structures of I1 and I2, we placed these complexes in the centre of a
755 dodecahedron box and added solvent atoms and $MgCl_2$ to neutralise the charge of the
756 system. The box size was determined by gmx editconf using a minimum distance of 3 and 2
757 nm from the box edge for I1 and I2, respectively. This resulted in system sizes of 1.2 – 1.7
758 million atoms. After energy minimisation, we first gently heated the simulation boxes from 0
759 to 298 K over a 500 ps using a time step of 0.5 fs and LINCS constraints for all bonds involving
760 hydrogen atoms. The systems were then equilibrated at 298 K in the NVT ensemble (velocity
761 rescaling thermostat with a coupling constant of 0.1 ps) for 500 ps using a 2 fs time step,

762 followed by a 1 ns equilibration step in the NPT ensemble at 1 bar (Berendsen barostat with
763 a time constant of 2 ps). During the equilibration protocol, all RNA and protein heavy atoms
764 were position restrained (1000 kJ $mol^{-1}$ $nm^{-2}$). These restraints were then removed for all NC
765 atoms (except the C-terminal residue at the peptidyl transferase centre) for a further 1 ns
766 equilibration simulation (NPT, Parrinello-Rahman barostat, 2 ps coupling constant) and
767 production simulations. We performed three unbiased simulations starting from different
768 initial velocities for each starting structure lasting 1 μs, resulting in a total of 15 μs for each
769 coTF intermediate state.

770

771 RMSF and SASA analyses were performed with gmx rmsf and sasa[65], respectively.
772 Alignment prior to RMSF calculation was done using Cα atoms of residues in the folded
773 region of the intermediate state, F691-G750 (C-strand to C-terminus) and K646-L733 (N-
774 terminus to F-strand) for I1 and I2, respectively. The fraction of native contacts (Q) was
775 calculated using PLUMED as described above. The radius of gyration was computed using
776 all Cα atoms of the FLN5 domain (residues K646-G750) with PLUMED to compare the
777 compactness of the coTF intermediates with the native state. Contact maps, ribosome
778 interactions, distances, and secondary structure content were calculated using the MDTraj[84]
779 and MDAnalysis[85] Python libraries. The rotational correlation times of the coTF intermediates
780 were calculated as previously described[23]. The rotation of the domain was defined using
781 backbone heavy atoms belonging to the same regions as defined for the RMSF analysis of
782 I1 and I2. Structural analyses were compared with the native state ensemble of FLN5
783 previously obtained[23] with the DES-Amber force field. All ensemble properties were averaged
784 over the entire 15 μs ensemble and errors are reported as the standard error of the mean
785 obtained from the 15 independent 1 μs trajectories.

786

787 The structural ensembles were clustered using for the purpose of structural visualisation
788 using gmx cluster[65] and the GROMOS algorithm[83]. Alignment and RMSD calculations for
789 clustering were performed using the same residue ranges as for the RMSF analysis and all
790 backbone heavy atoms. A similarity cutoff of 0.2 and 0.1 nm were chosen for I1 and I2,
791 respectively, due to these values resulting in similar cluster populations (**Table S4**).
792 Structures from the most populated cluster were then used for visualisation in the main text
793 figure (**Figure 3a-d**).

794

795 Simulations of fluorinated protein variants in isolation

796 We used the CHARMM36m[86] and ff15ipq[87,88] force fields to simulate the following isolated
797 FLN5 variants labelled with 4-trifluoromethyl-L-phenylalanine (tfmF): 655tfmF, 691tfmF,
798 705tfmF, 715tfmF, 719tfmF, 727tfmF, 732tfmF, and 740tfmF. We used an identical
799 simulation protocol as described previously[39]. Production simulations were run for 1 μs,
800 saving coordinates for analysis every 0.1 ns. Titin I27 variants lacking the A-strand (residues
801 1-7 deleted) were simulated with the ff15ipq[87] force field for 10 μs in triplicates to sufficiently
802 sample the folding intermediate. GROMACS tools gmx sasa, rmsd, and rmsf were used for
803 general analyses of the $CF_3$ group (SASA) and protein structure/dynamics (rmsd and rmsf).
804 Hydrogen bonds between fluorine (acceptor) and donor atoms were analysed with gmx
805 hbond. To estimate hydrogen bond lifetimes, we calculated the area under the averaged
806 autocorrelation function from all hydrogen bond existence functions[89,90]. Distances were
807 calculated with in-house Python scripts using MDAnalysis[85], and $CF_3$-aromatic interactions
808 were analysed with an open-source Python script (https://github.com/julian-
809 streit/RingCurrents19F).

810

811 Simulations isolated FLN5 fragments mimicking coTF intermediates I1 and I2 were
812 performed using the ff15ipq force field[87] and minimal constructs comprising the folded core
813 of each intermediate. We used residues Ac-A683-G750-NMe and S645-D744 for I1 and I2,
814 respectively. Three different starting structures were taken from the same rMD-generated
815 models as used for the unbiased RNC simulations (I1) and FLN5 Δ6 structural ensemble[11]
816 (I2). Independent simulations were performed using the same protocol as for full-length
817 fluorinated proteins from these starting structures with random initial velocities to obtain three
818 replicates where the I1/I2 topology was maintained during the 1 μs simulations. This was
819 assessed by having at least one $^{19}F$ NMR residue pair with a native-like distance per β-strand
820 pair (within 0.1 nm of the average distance calculated for the FLN5+47 native state[23]. The
821 native β-strand pairs measured by $^{19}F$ NMR for I1 were CF/FG, and AB/BE/DE/CF for I2.

822 $^{19}$F PRE distance calculations

823 $^{19}$F NMR PRE rates $\Gamma_2$ were converted to distances, *r*, using the Solomon-Bloembergen
824 equation[91], where $\omega_F$ and $\tau_C$ are the Larmor frequency of fluorine-19 and rotational correlation
825 time, respectively:

826 $$r = \left(\frac{K}{\Gamma_2} \times \left(4\tau_c + \frac{3\tau_c}{1 + \omega_F^2 \tau_c{}^2}\right)\right)^{1/6}$$

827 *K* (1.092 x $10^{-44}$ $m^6s^{-2}$) is defined as a product of physical constants including the vacuum
828 permeability, $\mu_0$, gyromagnetic ratio of fluorine-19, $\gamma_F$, electron g-factor, *g*, Bohr magneton,
829 $\mu_B$, and the spin number of fluorine-19, S.

830 $$K = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_F^2\, g^2 \mu_B^2\, S(S+1)$$

831 Values of $\tau_C$ have been determined in previous studies[37], and for RNCs were determined
832 from the linewidths using an empirical relationship recently determined[3]. To extrapolate the
833 expected PRE of native FLN5 on the ribosome, we therefore used the measured PRE
834 distance in isolation and RNC-specific $\tau_C$:

835 $$\Gamma_2 = \frac{K}{r^6} \times \left(4\tau_c + \frac{3\tau_c}{1 + \omega_F^2 \tau_c{}^2}\right)$$

836 **Cryo-EM fitting**

837 We fitted the ribosome model used in the MD simulations to each cryo-EM map representing
838 the four FLN5+47 dataset classes (AP1, AP2, AP5 and P5) (Mitropoulou, et al, in
839 preparation). The obtained rotation-translation matrix was applied in Python using the
840 MDAnalysis library to fit the NC structure from each frame of the MD trajectory (native, I2 and
841 I1 states) into the NC density. Next, we used scripts from cryoENsemble[43] to generate
842 synthetic cryo-EM maps for each MD frame at 8 Å resolution. The correlation coefficient
843 between each synthetic map and the corresponding experimental cryo-EM map was then
844 calculated following the definition of map correlation by ChimeraX[92].

845

846 As an alternative approach, we also used the cryo-EM maps to reweight the combined
847 trajectories corresponding to native, I2 and I1 states using the cryoENsemble iterative
848 strategy[43]. Prior to reweighting we clustered each trajectory using gromos algorithm from the

849 gmx cluster, using a 1-nm cut-off. The obtained clustered trajectory, comprising 2,294
850 conformations (1,091 from I1, 706 from I2 and 497 from native state), was used in four
851 cryoENsemble reweighting runs in each of the cryo-EM maps of FLN5+47, following the
852 previously described protocol[43].

853

854 **Data availability**

855 The structural ensembles of FLN5+47 I1 and I2, and NMR data will be deposited on Zenodo
856 upon publication. This study made use of the following public datasets deposited in the PDB
857 (https://wwww.rcsb.org): 1QFH, 1TIT, 1WLH, and 2W0P.

858

859 **Code availability**

860 Pulse sequences used for the acquisition of NMR data, and Matlab and Julia codes used to
861 fit the $^{19}$F NMR data are available on Github (**https://github.com/shschan/**). Python scripts
862 used to predict aromatic ring interactions for fluorinated protein variants are available on
863 Github (**https://github.com/julian-streit/RingCurrents19F**).

864

865 **Acknowledgements**

866 This work was supported by a Wellcome Trust Investigator Award (to J.C., 206409/Z/17/Z).
867 J.O.S. was supported by a BBSRC London Interdisciplinary Biosciences Doctoral
868 Programme studentship. We acknowledge use of the UCL Biomolecular NMR Centre. This
869 project made use of time on HPC resources on Archer2 (ARCHER2 UK National
870 Supercomputing service, https://www.archer2.ac.uk) granted via the UK High-End
871 Computing Consortium for Biomolecular Simulation, HECBioSim
872 (https://www.hecbiosim.ac.uk), supported by EPSRC (grant no. EP/R029407/1 and
873 EP/X035603/1). We also acknowledge the EuroHPC Joint Undertaking for awarding this
874 project access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland) and the LUMI
875 consortium through a EuroHPC Regular Access call and the Baskerville Tier 2 HPC service
876 (https://www.baskerville.ac.uk/). Baskerville was funded by the EPSRC and UKRI through
877 the World Class Labs scheme (EP/T022221/1) and the Digital Research Infrastructure
878 programme (EP/W032244/1) and is operated by Advanced Research Computing at the
879 University of Birmingham. We additionally acknowledge the use of the UCL Myriad, Kathleen,

and Young High Performance Computing Facility, and associated support services, in the completion of this work.

**Author contributions**

S.H.S.C., J.O.S., and J.C. designed and conceptualised the project. S.H.S.C. and J.O.S. produced the samples and performed the NMR experiments. S.H.S.C. analysed the NMR experiments. J.O.S. performed and analysed the MD simulations. T.W. and A.M. performed the cryo-EM analyses. A.M.E.C. and I.V.B. contributed to the initial C740 PRE experiments. S.H.S.C., J.O.S., L.D.C., and J.C. sourced the computational resources and acquired the funding. S.H.S.C. and J.C. supervised the project. S.H.S.C., J.O.S., and J.C. administered the project. S.H.S.C. wrote the original draft, S.H.S.C., J.O.S., and J.C. edited the manuscript, and all authors reviewed the manuscript.

**Competing interests**

The authors declare no competing interests.

**References**

1 Waudby, C. A., Dobson, C. M. & Christodoulou, J. Nature and Regulation of Protein Folding on the Ribosome. *Trends Biochem Sci* **44**, 914-926, doi:10.1016/j.tibs.2019.06.008 (2019).

2 Chan, S. H. S. *et al.* The ribosome stabilizes partially folded intermediates of a nascent multi-domain protein. *Nature Chemistry* **14**, 1165-1173, doi:10.1038/s41557-022-01004-0 (2022).

3 Streit, J. O. *et al.* The ribosome lowers the entropic penalty of protein folding. *Nature* **633**, 232-239, doi:10.1038/s41586-024-07784-4 (2024).

4 Samelson, A. J., Jensen, M. K., Soto, R. A., Cate, J. H. & Marqusee, S. Quantitative determination of ribosome nascent chain stability. *Proc Natl Acad Sci U S A* **113**, 13402-13407, doi:10.1073/pnas.1610272113 (2016).

5 Cassaignau, A. M. E. *et al.* Interactions between nascent proteins and the ribosome surface inhibit co-translational folding. *Nat Chem* **13**, 1214-1220, doi:10.1038/s41557-021-00796-x (2021).

6 Jensen, M. K., Samelson, A. J., Steward, A., Clarke, J. & Marqusee, S. The folding and unfolding behavior of ribonuclease H on the ribosome. *J Biol Chem* **295**, 11410-11417, doi:10.1074/jbc.RA120.013909 (2020).

7 Liu, K., Rehfus, J. E., Mattson, E. & Kaiser, C. M. The ribosome destabilizes native and non-native structures in a nascent multidomain protein. *Protein Sci* **26**, 1439-1451, doi:10.1002/pro.3189 (2017).

8 Wruck, F. *et al.* The ribosome modulates folding inside the ribosomal exit tunnel. *Commun Biol* **4**, 523, doi:10.1038/s42003-021-02055-8 (2021).

9 Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I., Jr. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723-1727, doi:10.1126/science.1209740 (2011).

10 Duran-Romaña, R. *et al.* Native Fold Delay and its implications for co-translational chaperone binding and protein aggregation. *Nature Communications* **16**, 1673, doi:10.1038/s41467-025-57033-z (2025).

11 Waudby, C. A. *et al.* Systematic mapping of free energy landscapes of a growing filamin domain during biosynthesis. *Proc Natl Acad Sci U S A* **115**, 9744-9749, doi:10.1073/pnas.1716252115 (2018).

12 Bhushan, S. *et al.* alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* **17**, 313-317, doi:10.1038/nsmb.1756 (2010).

13 Deckert, A. *et al.* Common sequence motifs of nascent chains engage the ribosome surface and trigger factor. *Proceedings of the National Academy of Sciences of the United States of America* **118**, doi:ARTN e2103015118

10.1073/pnas.2103015118 (2021).

14 Knight, A. M. *et al.* Electrostatic Effect of the Ribosomal Surface on Nascent Polypeptide Dynamics. *Acs Chemical Biology* **8**, 1195-1204, doi:10.1021/cb400030n (2013).

15 Tan, R. *et al.* Folding stabilities of ribosome-bound nascent polypeptides probed by mass spectrometry. *Proc Natl Acad Sci U S A* **120**, e2303167120, doi:10.1073/pnas.2303167120 (2023).

16 Ahn, M. *et al.* Modulating co-translational protein folding by rational design and ribosome engineering. *Nature Communications* **13**, doi:ARTN 4243

10.1038/s41467-022-31906-z (2022).

17 Kudva, R. *et al.* The shape of the bacterial ribosome exit tunnel affects cotranslational protein folding. *Elife* **7**, doi:10.7554/eLife.36326 (2018).

18 Guzman-Luna, V., Fuchs, A. M., Allen, A. J., Staikos, A. & Cavagnero, S. An intrinsically disordered nascent protein interacts with specific regions of the ribosomal surface near the exit tunnel. *Communications Biology* **4**, doi:ARTN 1236

10.1038/s42003-021-02752-4 (2021).

19 Deuerling, E., Gamerdinger, M. & Kreft, S. G. Chaperone Interactions at the Ribosome. *Cold Spring Harb Perspect Biol* **11**, doi:10.1101/cshperspect.a033977 (2019).

20 Roeselova, A. *et al.* Mechanism of chaperone coordination during cotranslational protein folding in bacteria. *Mol Cell*, doi:10.1016/j.molcel.2024.06.002 (2024).

21 Kramer, G., Shiber, A. & Bukau, B. Mechanisms of Cotranslational Maturation of Newly Synthesized Proteins. *Annu Rev Biochem* **88**, 337-364, doi:10.1146/annurev-biochem-013118-111717 (2019).

22 Oh, E. *et al.* Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147**, 1295-1308, doi:10.1016/j.cell.2011.10.044 (2011).

23 Streit, J. O. *et al.* Long-range electrostatic forces govern how proteins fold on the ribosome. *bioRxiv*, 2025.2002.2010.637539, doi:10.1101/2025.02.10.637539 (2025).

24 Requiao, R. D. *et al.* Protein charge distribution in proteomes and its impact on translation. *PLoS Comput Biol* **13**, e1005549, doi:10.1371/journal.pcbi.1005549 (2017).

25 Plessa, E. *et al.* Nascent chains can form co-translational folding intermediates that promote post-translational folding outcomes in a disease-causing protein. *Nature Communications* **12**, doi:ARTN 6447

962 10.1038/s41467-021-26531-1 (2021).

963 26 Evans, M. S., Sander, I. M. & Clark, P. L. Cotranslational folding promotes beta-helix formation
964 and avoids aggregation in vivo. *J Mol Biol* **383**, 683-692, doi:10.1016/j.jmb.2008.07.035
965 (2008).

966 27 Alexander, L. M., Goldman, D. H., Wee, L. M. & Bustamante, C. Non-equilibrium dynamics of
967 a nascent polypeptide during translation suppress its misfolding. *Nat Commun* **10**, 2709,
968 doi:10.1038/s41467-019-10647-6 (2019).

969 28 Bertolini, M. *et al.* Interactions between nascent proteins translated by adjacent ribosomes
970 drive homomer assembly. *Science* **371**, 57-64, doi:ARTN eabc7151

971 10.1126/science.abc7151 (2021).

972 29 Shiber, A. *et al.* Cotranslational assembly of protein complexes in eukaryotes revealed by
973 ribosome profiling. *Nature* **561**, 268-272, doi:10.1038/s41586-018-0462-y (2018).

974 30 Liutkute, M., Maiti, M., Samatova, E., Enderlein, J. & Rodnina, M. V. Gradual compaction of
975 the nascent peptide during cotranslational folding on the ribosome. *Elife* **9**,
976 doi:10.7554/eLife.60895 (2020).

977 31 Liu, K., Maciuba, K. & Kaiser, C. M. The Ribosome Cooperates with a Chaperone to Guide
978 Multi-domain Protein Folding. *Mol Cell* **74**, 310-319 e317, doi:10.1016/j.molcel.2019.01.043
979 (2019).

980 32 Holtkamp, W. *et al.* Cotranslational protein folding on the ribosome monitored in real time.
981 *Science* **350**, 1104-1107, doi:10.1126/science.aad0344 (2015).

982 33 Wales, T. E. *et al.* Resolving chaperone-assisted protein folding on the ribosome at the peptide
983 level. *Nat Struct Mol Biol*, doi:10.1038/s41594-024-01355-x (2024).

984 34 Pellowe, G. A. *et al.* The human ribosome modulates multidomain protein biogenesis by
985 delaying

986 cotranslational domain docking. *bioRxiv*, doi:10.1101/2024.09.19.613857 (2024).

987 35 Chen, X., Rajasekaran, N., Liu, K. & Kaiser, C. M. Synthesis runs counter to directional folding
988 of a nascent protein domain. *Nat Commun* **11**, 5096, doi:10.1038/s41467-020-18921-8
989 (2020).

990 36 Cabrita, L. D. *et al.* A structural ensemble of a ribosome-nascent chain complex during
991 cotranslational protein folding. *Nat Struct Mol Biol* **23**, 278-285, doi:10.1038/nsmb.3182
992 (2016).

993 37 Burridge, C. *et al.* Nascent chain dynamics and ribosome interactions within folded ribosome-
994 nascent chain complexes observed by NMR spectroscopy. *Chem Sci* **12**, 13120-13126,
995 doi:10.1039/d1sc04313g (2021).

996 38 Chan, S. H. S., Waudby, C. A. & Christodoulou, J. NMR snapshots of nascent chains emerging
997 from the ribosome during biosynthesis. *chemRxiv* (2022).

998 39 Streit, J. O., Chan, S. H. S., Daya, S. & Christodoulou, J. Rational design of $^{19}$F NMR labelling
999 sites to probe protein structure and interactions. *bioRxiv*, 2024.2012.2011.627779,
1000 doi:10.1101/2024.12.11.627779 (2024).

1001 40 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-
1002 682, doi:10.1038/s41592-022-01488-1 (2022).

1003 41 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold
1004 3. *Nature* **630**, 493-500, doi:10.1038/s41586-024-07487-w (2024).

1005 42 Krantz, B. A. & Sosnick, T. R. Engineered metal binding sites map the heterogeneous folding
1006 landscape of a coiled coil. *Nat Struct Biol* **8**, 1042-1047, doi:10.1038/nsb723 (2001).

43 Wlodarski, T. *et al.* Bayesian reweighting of biomolecular structural ensembles using heterogeneous cryo-EM maps with the cryoENsemble method. *Sci Rep* **14**, 18149, doi:10.1038/s41598-024-68468-7 (2024).

44 Fowler, S. B. *et al.* Mechanical Unfolding of a Titin Ig Domain: Structure of Unfolding Intermediate Revealed by Combining AFM, Molecular Dynamics Simulations, NMR and Protein Engineering. *Journal of Molecular Biology* **322**, 841-849, doi:https://doi.org/10.1016/S0022-2836(02)00805-7 (2002).

45 Fowler, S. B. & Clarke, J. Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. *Structure* **9**, 355-366, doi:10.1016/s0969-2126(01)00596-2 (2001).

46 Nilsson, O. B. *et al.* Cotranslational folding of spectrin domains via partially structured states. *Nature Structural & Molecular Biology* **24**, 221-+, doi:10.1038/nsmb.3355 (2017).

47 Schwaiger, I., Schleicher, M., Noegel, A. A. & Rief, M. The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *Embo Reports* **6**, 46-51, doi:10.1038/sj.embor.7400317 (2005).

48 Wright, C. F., Lindorff-Larsen, K., Randles, L. G. & Clarke, J. Parallel protein-unfolding pathways revealed and mapped. *Nat Struct Biol* **10**, 658-662, doi:10.1038/nsb947 (2003).

49 Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A. A. & Rief, M. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat Struct Mol Biol* **11**, 81-85, doi:10.1038/nsmb705 (2004).

50 Clarke, J., Cota, E., Fowler, S. B. & Hamill, S. J. Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure* **7**, 1145-1153, doi:10.1016/s0969-2126(99)80181-6 (1999).

51 Khushoo, A., Yang, Z., Johnson, A. E. & Skach, W. R. Ligand-Driven Vectorial Folding of Ribosome-Bound Human CFTR NBD1. *Molecular Cell* **41**, 682-692, doi:10.1016/j.molcel.2011.02.027 (2011).

52 Duttler, S., Pechmann, S. & Frydman, J. Principles of cotranslational ubiquitination and quality control at the ribosome. *Mol Cell* **50**, 379-393, doi:10.1016/j.molcel.2013.03.010 (2013).

53 Keshwani, M. M. *et al.* Cotranslational cis-phosphorylation of the COOH-terminal tail is a key priming step in the maturation of cAMP-dependent protein kinase. *Proceedings of the National Academy of Sciences* **109**, E1221-E1229, doi:doi:10.1073/pnas.1202741109 (2012).

54 Szeto, S. G. Y., Williams, E. C., Rudner, A. D. & Lee, J. M. Phosphorylation of filamin A by Cdk1 regulates filamin A localization and daughter cell separation. *Experimental Cell Research* **330**, 248-266, doi:https://doi.org/10.1016/j.yexcr.2014.10.024 (2015).

55 Oh, W. J. *et al.* mTORC2 can associate with ribosomes to promote cotranslational phosphorylation and stability of nascent Akt polypeptide. *The EMBO Journal* **29**, 3939-3951-3951, doi:https://doi.org/10.1038/emboj.2010.271 (2010).

56 Urrutia, J. *et al.* An epilepsy-causing mutation leads to co-translational misfolding of the Kv7.2 channel. *BMC Biol* **19**, 109, doi:10.1186/s12915-021-01040-1 (2021).

57 Shishido, H., Yoon, J. S., Yang, Z. & Skach, W. R. CFTR trafficking mutations disrupt cotranslational protein folding by targeting biosynthetic intermediates. *Nat Commun* **11**, 4258, doi:10.1038/s41467-020-18101-8 (2020).

58 Cassaignau, A. M. *et al.* A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nat Protoc* **11**, 1492-1507, doi:10.1038/nprot.2016.101 (2016).

59 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293, doi:10.1007/BF00197809 (1995).

60 Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: A Fresh Approach to Numerical Computing. *Siam Review* **59**, 65-98, doi:10.1137/141000671 (2017).

61 Schanda, P., Kupce, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **33**, 199-211, doi:10.1007/s10858-005-4425-x (2005).

62 Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett* **100**, 020603, doi:10.1103/PhysRevLett.100.020603 (2008).

63 Piana, S., Donchev, A. G., Robustelli, P. & Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B* **119**, 5113-5123, doi:10.1021/jp508971m (2015).

64 Piana, S., Robustelli, P., Tan, D., Chen, S. & Shaw, D. E. Development of a Force Field for the Simulation of Single-Chain Proteins and Protein-Protein Complexes. *J Chem Theory Comput* **16**, 2494-2507, doi:10.1021/acs.jctc.9b00251 (2020).

65 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25, doi:10.1016/j.softx.2015.06.001 (2015).

66 consortium, P. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods* **16**, 670-673, doi:10.1038/s41592-019-0506-8 (2019).

67 Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185**, 604-613, doi:10.1016/j.cpc.2013.09.018 (2014).

68 McCoy, A. J., Fucini, P., Noegel, A. A. & Stewart, M. Structural basis for dimerization of the Dictyostelium gelation factor (ABP120) rod. *Nat Struct Biol* **6**, 836-841, doi:10.1038/12296 (1999).

69 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **126**, doi:Artn 014101,10.1063/1.2408420 (2007).

70 Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **81**, 3684-3690, doi:Doi 10.1063/1.448118 (1984).

71 Parrinello, M. & Rahman, A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *Journal of Applied Physics* **52**, 7182-7190, doi:Doi 10.1063/1.328693 (1981).

72 Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472, doi:Doi 10.1002/(Sici)1096-987x(199709)18:12<1463::Aid-Jcc4>3.0.Co;2-H (1997).

73 Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **98**, 10089-10092, doi:Doi 10.1063/1.464397 (1993).

74 Melis, C., Bussi, G., Lummis, S. C. & Molteni, C. Trans-cis switching mechanisms in proline analogues and their relevance for the gating of the 5-HT3 receptor. *J Phys Chem B* **113**, 12148-12153, doi:10.1021/jp9046962 (2009).

75 Tiwary, P. & Parrinello, M. A time-independent free energy estimator for metadynamics. *J Phys Chem B* **119**, 736-742, doi:10.1021/jp504920s (2015).

76 Noel, J. K. *et al.* SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *Plos Computational Biology* **12**, doi:ARTN e1004794, 10.1371/journal.pcbi.1004794 (2016).

77 Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci U S A* **110**, 17874-17879, doi:10.1073/pnas.1311599110 (2013).

78 Camilloni, C., Broglia, R. A. & Tiana, G. Hierarchy of folding and unfolding events of protein G, CI2, and ACBP from explicit-solvent simulations. *J Chem Phys* **134**, 045105, doi:10.1063/1.3523345 (2011).

79 Tiana, G. & Camilloni, C. Ratcheted molecular-dynamics simulations identify efficiently the transition state of protein folding. *J Chem Phys* **137**, 235101, doi:10.1063/1.4769085 (2012).

80 Wang, F. *et al.* All-Atom Simulations Reveal How Single-Point Mutations Promote Serpin Misfolding. *Biophys J* **114**, 2083-2094, doi:10.1016/j.bpj.2018.03.027 (2018).

81 Tucker, M. R., Piana, S., Tan, D., LeVine, M. V. & Shaw, D. E. Development of Force Field Parameters for the Simulation of Single- and Double-Stranded DNA Molecules and DNA-Protein Complexes. *J Phys Chem B* **126**, 4442-4457, doi:10.1021/acs.jpcb.1c10971 (2022).

82 Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **11**, 1864-1874, doi:10.1021/ct5010406 (2015).

83 Daura, X. *et al.* Peptide folding: When simulation meets experiment. *Angewandte Chemie-International Edition* **38**, 236-240, doi:Doi 10.1002/(Sici)1521-3773(19990115)38:1/2<236::Aid-Anie236>3.0.Co;2-M (1999).

84 McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* **109**, 1528-1532, doi:10.1016/j.bpj.2015.08.015 (2015).

85 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* **32**, 2319-2327, doi:10.1002/jcc.21787 (2011).

86 Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**, 71-73, doi:10.1038/nmeth.4067 (2017).

87 Yang, D. T., Gronenborn, A. M. & Chong, L. T. Development and Validation of Fluorinated, Aromatic Amino Acid Parameters for Use with the AMBER ff15ipq Protein Force Field. *J Phys Chem A* **126**, 2286-2297, doi:10.1021/acs.jpca.2c00255 (2022).

88 Debiec, K. T. *et al.* Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J Chem Theory Comput* **12**, 3926-3947, doi:10.1021/acs.jctc.6b00567 (2016).

89 van der Spoel, D., van Maaren, P. J., Larsson, P. & Timneanu, N. Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media. *J Phys Chem B* **110**, 4393-4398, doi:10.1021/jp0572535 (2006).

90 Luzar, A. & Chandler, D. Hydrogen-bond kinetics in liquid water. *Nature* **379**, 55-57, doi:DOI 10.1038/379055a0 (1996).

91 Solomon, I. & Bloembergen, N. Nuclear Magnetic Interactions in the HF Molecule. *J Chem Phys* **25**, 261-266, doi:https://doi.org/10.1063/1.1742867 (1956).

92 Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **32**, e4792, doi:https://doi.org/10.1002/pro.4792 (2023).