Cotranslational folding of spectrin domains via partially structured states

Ola B. Nilsson[1‡], Adrian A. Nickson[2‡], Jeffrey J. Hollins[2], Stephan Wickles[3], Annette Steward[2], Roland Beckmann[3], Gunnar von Heijne[1,4]*, Jane Clarke[2]*

[1]Department of Biochemistry and Biophysics Stockholm University, SE-106 91 Stockholm, Sweden

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

[3]Gene Center and Center for Integrated Protein Science Munich, CiPS-M, Feodor-Lynen-Strasse 25, University of Munich, 81377 Munich, Germany

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

[‡]O.B.N. and A.A.N. contributed equally to this work.

* Correspondence and requests for materials should be addressed to J.C. (jc162@cam.ac.uk) and G.vH. (gunnar@dbb.su.se)

Summary

How do the key features of protein folding, elucidated from studies on native, isolated proteins, manifest in cotranslational folding on the ribosome? Using a well-characterized family of homologous α-helical proteins, with a range of biophysical properties, we show that spectrin domains can fold vectorially on the ribosome and may do so via a pathway different from that of the isolated domain. We use cryo-electron microscopy to reveal a folded or partially folded structure, formed in the vestibule of the ribosome. Our results reveal that it is not possible to predict which domains will fold within the ribosome from analysis of the folding behaviour of isolated domains; instead we propose that a complex balance between the rate of folding, the rate of translation and the lifetime of folded or partly folded states will determine whether folding occurs cotranslationally on actively translating ribosomes.

Our knowledge about protein folding, misfolding and aggregation stems mainly from *in vitro* investigations of small, soluble, isolated protein domains that fold reversibly[1]. Combined with simulations and theory, such studies have explained how proteins can fold on a biological timescale and have led to many recent successes in protein engineering and *de novo* protein design[2-4]. Nonetheless, there is some question as to whether the key features of protein folding observed for isolated, native domains are relevant in the context of cotranslational folding on the ribosome and *in vivo*[5].

The vectorial nature of protein synthesis means that N-terminal regions of the protein are potentially able to fold before C-terminal regions have emerged from the ribosome[6-8]. It is estimated that 30 – 40 residues of the nascent chain are sequestered within the ribosome during translation in a "tunnel" that extends some 100 Å from the peptidyl transfer center (PTC)[9,10]; helix formation has been observed in the lower ribosome tunnel[11] (~50-80Å from the PTC) where very small proteins can even achieve their final folded structure[12]. Tertiary interactions have been observed to form within the last 20 Å, in an area of the exit tunnel referred to as the "vestibule"[13-15]. In a recent study a poorly characterized compact state was shown to form transiently in the *E. coli* HemK protein, when parts of the nascent protein were still enclosed in the exit tunnel[16]. Molecular dynamics simulations suggest that sequential folding is promoted by the ribosome[17]. Thus, it is still an open question as to whether cotranslational folding pathways are the same as those observed for isolated, native proteins, where the entire unfolded protein chain is synchronously transferred into folding conditions.

Here we investigate the influence of protein stability, folding kinetics and pathway on cotranslational folding, using an *in vitro* arrest peptide-based force measurement assay on a family of all α-helical spectrin domains. Our results suggest

that for this group of spectrin domains, the ability to fold cotranslationally is not determined directly by the intrinsic stability, folding rate or pathway of the isolated, native protein. Instead we propose that it is the relative stability of folded or partly folded structures, which can form when a proportion of C-terminal residues are still sequestered by the ribosome, that determines when folding can occur cotranslationally. We also observe for these spectrin domains, that the pathway of folding on stalled ribosomes may be different from that observed for the native, isolated protein. Data provided by cryo-EM shows a partially or fully folded structured state, formed in the vestibule of the ribosome and we suggest that the stability of the structured state may be enhanced by interactions with the ribosome.

We discuss these results in the context of cotranslational folding *in vivo*. We note that, although these proteins can start to fold before the domain is fully emerged from the ribosome, it is not possible to predict which domains will fold within the ribosome from analysis of the folding behavior of the isolated, native proteins.

Results

Experimental Set-up. We employed six homologous 3-helix bundle spectrin domains (R15, R16, β16 and three variants of R16) as model proteins (**Fig. 1)** in our *in vitro* arrest peptide-based force measurement assay (AP assay). Such assays have been recently developed to study the forces generated on the nascent polypeptide chain during cotranslational membrane protein insertion and protein folding[15,18-22]. Translational arrest peptides (APs) induce translational stalling when synthesized on a ribosome[23-26]. Mechanical pulling forces acting on the nascent chain can weaken or even abolish stalling[20,27]. When translation is deliberately stalled by a bacterial SecM

AP, the stall can be released by the application of moderate force (around 10 pN) to the nascent chain[21]; the probability of release, and hence the yield of full-length versus stalled protein, is proportional to the force applied. The fraction of released, full-length protein observed, f_{FL}, can therefore serve as a proxy for the force acting on the nascent chain[20]. By measuring f_{FL} as a function of the linker length (L) between the protein and the arrest site, it is possible to generate a profile that represents the force generated by the protein as it folds, as a function of the location of the protein in the ribosome tunnel (for more details see **Fig. 2, Supplementary Fig. S1**). Each data point on the force profiles is a single experiment. We know from previous studies that the SDs for f_{FL} values obtained by translation of arrest-peptide constructs in the PURE system is on the order of 0.03[12]. Here we show that replicate measurements of the same construct produce reproducible data (see **Supplementary Fig. S2**). demonstrating the robustness of the assay. The spectrin domains employed in the AP assay have a range of intrinsic stabilities, rates of folding and unfolding and different folding pathways, as determined from experiments on the isolated, native domains (**Table 1**). (Note that we define intrinsic as relating to the properties of the native domains in isolation). We relate the force profiles obtained by the AP assay to these differences.

The linker length at which folding can occur does not correlate with the stability or folding kinetics of the isolated native protein. To determine whether the linker length at which folding can occur is related to intrinsic thermodynamic stability or folding rate constant of the isolated domains, force profiles were obtained for spectrins R15, R16 and β16 by *in vitro* translation in the PURE system[28] (**Fig. 3a**). Importantly, non-folding variants of R15 and R16 (R15nf, R16nf) did not release translational arrest,

demonstrating that the force generated is due to cotranslational folding and not to interactions of the unfolded nascent chain with the ribosome. Both β16 and R16 start to fold at L = 27-29 residues, *i.e.*, while a part of the nascent chain is still sequestered in the ribosome, despite possessing quite different intrinsic thermodynamic stabilities (**Table 1**). R15 starts to fold only at L = 37 residues, although it has the same intrinsic stability and folds significantly faster than R16 (**Table 1**). Thus there is no apparent correlation between either intrinsic stability of the isolated domains nor rate of folding, and the linker length at which folding occurs on the ribosome.

A stabilizing neighbor results in early folding. Spectrin R16 begins to exert a force at $L \approx 29$ residues (**Fig. 3a**). At this linker length some C-terminal residues will be sequestered in the ribosome and thus unavailable to form structure. Is it possible that R16 can fold into a more stable state than R15, when an equivalent number of residues are in the ribosome exit tunnel, and that this allows R16 to start folding earlier than R15? To gain an insight into the effect upon protein stability of sequestration of C-terminal residues in the ribosome, equilibrium denaturation studies were performed on isolated, native domains R15 and R16 truncated at the C-terminus (**Supplementary Table S1**). Our results show R15 is destabilized earlier and to a greater extent by C-terminal truncation than is R16. Perhaps the stability of structures that can form, in the absence of some C-terminal interactions, is the key to explaining the differences between the spectrin domain force profiles? If this were the case, then stabilization of R16 should cause it to fold at even shorter linker lengths. We have previously shown that R16 is stabilized by ~1.7 kcal mol^{-1} by the presence of an N-terminal, folded R15 domain, in the tandem protein R15R16[29,30]. We therefore recorded a force profile for R15R16. The results show that in the tandem protein, R16

starts to fold at significantly shorter linker lengths than does R16 alone (**Fig. 3b**); we ascribe this to the extra stability provided by the N-terminal R15 domain. When the cooperativity between R15 and R16 in the tandem protein is removed by the introduction of 3 proline residues at the domain boundary[31], the force profile of the resulting R153ProR16 construct reverts to that of R16 alone.

Simple sequence changes can affect when the protein folds. We recorded force profiles for two variants of R16, originally designed to incorporate the fast-folding nucleus of R15 into R16[32]. These both have five residues from the A helix of R15 substituted into R16 to generate the R16m5 variant (**Fig. 1, bottom**); an additional, conservative substitution in the B helix (V65L) confers an extra ~1 kcal mol^{-1} of stability, to generate the protein R16m6 (**Fig. 1, bottom**). Folding studies have previously shown that introduction of these five nucleating residues causes both R16 variants to fold more like R15 in terms of rapid kinetics and pathway (**Table 1**)[32]. We find that the force profile of R16m5 does indeed follow that of R15, but stabilization by just ~1 kcal mol^{-1} switches the force profile so that R16m6 now resembles that of R16 **(Fig. 3c)**. Remarkably, the addition of the stabilizing N-terminal R15 domain to R16m5 also leads to an R15R16-like force profile in the tandem R15R16m5 (**Fig. 3b)**. We also investigated another spectrin variant named R16outside15core (R16o15c), which contains the entire hydrophobic core of R15 and shares only surface residues with R16. As an isolated domain this variant has the same intrinsic rapid folding kinetics and folding pathway as R15[33] (**Table 1**), yet it behaves indistinguishably from R16 in our AP assay **(Fig. 3d)**.

Cryo-EM reveals spectrin R16 is folded in the ribosome vestibule. To ascertain the location of structured states of R16 in the ribosome tunnel, cryo-EM studies were performed on the ribosome-bound R16 [*L*=33] construct, with the SecM AP replaced by the TnaC AP, a stronger stall sequence that retains the nascent chain bound to the ribosome. A ~4 - 8 Å resolution structure was determined from arrested ribosome-nascent chain complexes purified from *E. coli*. In addition to density corresponding to parts of the AP and linker segment of the nascent chain in the ribosome tunnel, a globular density protruding from the tunnel exit was apparent (**Fig. 4, Supplementary Video** and **Supplementary Fig. S3**)**.** The core of the R16 solution NMR structure (PDB: 1AJ3) fits into this density, as does most of helix C, while the regions of helices A and B closest to the ribosomal protein uL24 (C-terminal region of helix A, N-terminal region of helix B and the A-B loop) are not visible.

Discussion

Using a combination of the AP-based force-measurement assay and structural analysis by cryo-EM on a family of all α-helical spectrin domains, we have shown that spectrin domains can fold cotranslationally, while a proportion of C-terminal residues remain sequestered by the ribosome; cryo-EM data on the stalled R16 [*L*=33] shows spectrin at least partly folded in the vestibule of the ribosome tunnel.

The stability of structures that can form before the protein has entirely emerged is key to determining when the protein can fold on the ribosome. We find that the point at which spectrin domains can fold does not correlate with the intrinsic stability, folding rate or pathway of the isolated, native proteins. This is reflected in the early-onset force peaks of β16, R16, R16m6 and R16o15c, all of which can fold (and thus exert a

force to release the stall) before the C-terminal residues are fully emerged from the tunnel and thus available to participate in structure formation (L < 30 residues). Yet these proteins share many of the intrinsic folding properties of R15 which folds only at long linker lengths.

A marginal increase in stability can affect the position of the onset of folding: A simple V to L substitution allows R16m6 to fold significantly earlier than R16m5, and addition of the stabilizing R15 domain to R16 or R16m5 allows the protein to fold even earlier. Thus we propose that the probability that a force sufficient to release the stall is generated correlates with the stability of folded, or potentially partly-folded, states that form cotranslationally and thus probably with the length of time these states can persist (if a folded structure that forms cotranslationally lacks sufficient stability, it will unfold before it has time to release the stall).

Our results also suggest that the ribosome itself may influence the stability and lifetime of structured or partly structured species which may form cotranslationally: R15 can apparently fold only when the protein is almost entirely emerged from the exit tunnel. A simple explanation is that folded or partly folded structures of R15, formed within the confines of the ribosome, are not as stable as similar R16 structures. Studies on the isolated proteins show that R15 is more destabilized by truncation than R16 (**Supplementary Table S1**), but this difference is not sufficient to fully explain the ~8-residue difference in the onset of cotranslational folding (**Fig. 3a**). In the cryo-EM structure, R16 appears to interact with the ribosome (**Fig. 4b**); we infer that stabilizing interactions with the ribosome may play a role in determining the onset of cotranslational folding. Further evidence is provided by R16o15c (R16outside15core) which folds in isolated protein studies like R15, but follows the force profile of R16 on the ribosome. It is possible that stabilizing interactions

between surface residues and the ribosome allow R16o15c to fold cotranslationally like R16. Interactions with the ribosome have previously been shown to influence the formation of stable tertiary interactions in the nascent chain[34-36]. In these cases, however, folding commenced only when the proteins were well outside the ribosome exit tunnel and interactions with the ribosome reduced rather than increased the stability of the folded state.

Relationship between onset of folding and folding pathways. Our previously published results on folding of the isolated domains of R15 and R16[37,38] reveal that the key interactions that form early when R16 folds are between the N-terminal regions of helices A and C (regions of structure that can likely form when a significant part of the C helix is still inside the ribosome tunnel); however in R15 the interactions essential for folding are between the C-terminal regions of helices A and C (**Table 1** and **Fig. 1, top**). From the cryo-EM data it is clear that at length *L*=33, R16 can form a structured state within the confines of the ribosome. There is a significant region of structure in the portion of the molecule distal from the tunnel exit. This density can be fitted using the NMR structure of R16, and is consistent with the packing of the N-terminal portions of all three helices (see **Figs 4, 5** and **Supplementary Video**). However, in the region closest to the tunnel the observed density is only consistent with formation of a single helix (assigned to the terminal C helix as it is contiguous with the nascent chain observed inside the tunnel, colored green in **Fig. 4**). Thus, at this linker length, where the protein is constrained within the vestibule of the ribosome tunnel, stable tertiary structure is not observed in the C-terminal portion of the molecule, exactly where we would expect nucleating packing interactions to be formed in R15 (between the C-terminal regions of helices A and C, shown mapped

onto the model of the partly folded structure in **Fig. 5**). This may give an alternative explanation why R15 can only fold at long linker lengths.

The ribosome can affect the folding pathway. R16m5 can potentially fold via two alternative pathways: via the 'R15 pathway' (by the 5 key residues substituted towards the C-terminal region) or via the 'R16 pathway' (by key residues located towards the N-terminal region) (**Fig. 1, top**). In the isolated domain, R16m5 folds via the C-terminal pathway, which is the lowest energy (faster folding) route. On the ribosome R16m5 still behaves like R15, folding only when almost fully emerged from the ribosome, when the C-terminal folding route becomes available. A single residue substitution (V65L) generates R16m6, stabilized by an extra ~1 kcal mol^{-1}. Now on the ribosome R16m6 is observed to fold like R16, commencing folding at shorter lengths, when the C-terminal pathway is still unavailable. Thus, this extra stabilization allows the protein to switch folding routes and access the 'R16 (N-terminal) pathway' (**Fig. 3c**). Importantly, the mutation that confers the extra stability is in the N-terminal region of the protein that is predicted to be structured in our cryo-EM model (colored green in **Fig. 5**). An even greater stabilizing effect is conferred on R16m5 by the R15 domain, in the tandem protein R15R16m5, which enables folding to begin at even shorter linker lengths (**Fig. 3b**). Thus we infer that if an alternative, sufficiently low-energy pathway is available, a protein may fold cotranslationally by a different pathway than that used by the full-length protein (provided that any folded or partly folded states have sufficient stability).

Implications for cotranslational folding in vivo. Our results provide insights into the factors affecting when a spectrin domain can fold cotranslationally, in a stalled ribosome system. Whether or not folding actually occurs within the confines of the tunnel or vestibule, on actively translating ribosomes, will depend critically upon the

255 balance between the rate of folding compared with the rate of translation and also the
256 lifetime of any structured or partly structured states that form[17]. Presumably, where
257 folding is fast compared to the translation rate (~5 codons s^{-1} in mammalian cells,
258 ~10-20 codons s^{-1} in *E. coli*)[39,40], there is a significant probability that a folding event
259 will occur at each step during translation. Whether folding continues to completion at
260 this chain length will depend on the lifetime of any structured states that are formed.
261 However, although the folding rates for all these spectrin domains, determined for the
262 isolated, native proteins, are significantly faster than the rate of translation, it is not
263 possible to determine the rates of folding (or, indeed unfolding) within the confines of
264 the ribosome. Since on both stalled and freely translating ribosomes the domains are
265 folding against the entropic restoring force of the tethered chain, it seems likely that
266 folding rates are significantly decreased and the probability of unfolding is larger on
267 the ribosome than off[21]. Our data suggest, however, that it is indeed possible that R16
268 can start to fold in the ribosome, and that spectrin domains are apparently able to fold
269 via a different pathway from that observed in the isolated domain, when confined in
270 the ribosome.

271 The relationship between folding of the isolated protein on and off the
272 ribosome is demonstrably complex: Some small domains, such as R15, will fold only
273 outside the ribosome exit tunnel, and it is not possible to predict which domains will
274 fold within the ribosome from analysis of the folding behavior of the isolated, native
275 proteins. Furthermore, the probability that a protein will fold within the ribosome
276 tunnel can be altered by seemingly insignificant conservative sequence changes. Early
277 onset of folding inside the exit tunnel may be particularly important in multidomain
278 proteins, like spectrin, in order to avoid misfolding by domain swapping[41,42]; indeed,

279 we have seen that stabilizing interactions with a neighboring domain can result in

280 early onset of cotranslational folding.

281

Accession codes. Coordinates for the cryo-EM map of the ribosome with the spectrin density have been deposited at the Electron Microscopy Data Bank under accession code EMD 3451. Coordinates of the spectrin model used for fitting and interpretation of the cryo-EM map have been deposited at the Protein Data Bank under accession code PDB 5M6S.

Data Availability: The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Supplementary Information is linked to the online version of the paper at www.nature.com/nsmb.

Acknowledgements

We thank A. Heuer for his help preparing the Figures. Supported by: Grants from the Swedish Cancer Foundation, the Swedish Research Council, and the Knut and Alice Wallenberg Foundation (to GvH); the Wellcome Trust (WT095195, to JC); a grant from the European Research Council (ERC-2008-AdG 232648, to RB). JC is a Wellcome Trust Senior Research Fellow

Author contributions: OBN and AAN designed and carried out the experiments. JJH and AS characterized the purified proteins, SW and RB were responsible for the cryo-EM experiments, AS wrote the manuscript, GvH and JC conceived and planned the investigation and wrote the manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

References

1. Braselmann, E., Chaney, J.L. & Clark, P.L. Folding the proteome. *Trends Biochem Sci* **38**, 337-344 (2013).
2. Joh, N.H. et al. De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle. *Science* **346**, 1520-1524 (2014).
3. Song, W.J. & Tezcan, F.A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **346**, 1525-1528 (2014).
4. Koga, N. et al. Principles for designing ideal protein structures. *Nature* **491**, 222-227 (2012).
5. Hingorani, K.S. & Gierasch, L.M. Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge. *Curr Opin Struct Biol* **24**, 81-90 (2014).
6. Frydman, J., Erdjument-Bromage, H., Tempst, P. & Hartl, F.U. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol* **6**, 697-705 (1999).
7. Kim, S.J. et al. Translational tuning optimizes nascent protein folding in cells. *Science* **348**, 444-448 (2015).
8. Evans, M.S., Sander, I.M. & Clark, P.L. Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J Mol Biol* **383**, 683-692 (2008).
9. Voss, N.R., Gerstein, M., Steitz, T.A. & Moore, P.B. The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol* **360**, 893-906 (2006).
10. Bhushan, S. et al. SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol* **9**, e1000581 (2011).

11. Bhushan, S. et al. α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* **17**, 313-317 (2010).
12. Nilsson, O.B. et al. Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep* **12**, 1533-1540 (2015).
13. O'Brien, E.P., Hsu, S.T., Christodoulou, J., Vendruscolo, M. & Dobson, C.M. Transient tertiary structure formation within the ribosome exit port. *J Am Chem Soc* **132**, 16928-16937 (2010).
14. Tu, L., Khanna, P. & Deutsch, C. Transmembrane segments form tertiary hairpins in the folding vestibule of the ribosome. *J Mol Biol* **426**, 185-198 (2014).
15. Kosolapov, A. & Deutsch, C. Tertiary interactions within the ribosomal exit tunnel. *Nat Struct Mol Biol* **16**, 405-411 (2009).
16. Holtkamp, W. et al. Cotranslational protein folding on the ribosome monitored in real time. *Science* **350**, 1104-1107 (2015).
17. O'Brien, E.P., Christodoulou, J., Vendruscolo, M. & Dobson, C.M. New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc* **133**, 513-526 (2011).
18. Ismail, N., Hedman, R., Lindén, M. & von Heijne, G. Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat Struct Mol Biol* **22**, 145-149 (2015).
19. Cymer, F. & von Heijne, G. Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proc Natl Acad Sci U S A* **110**, 14640-14645 (2013).

20. Ismail, N., Hedman, R., Schiller, N. & von Heijne, G. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nature Struct Molec Biol* **19**, 1018-1022 (2012).

21. Goldman, D.H. et al. Mechanical force releases nascent chain-mediated ribosome arrest *in vitro* and *in vivo*. *Science* **348**, 457-460 (2015).

22. Nilsson, O.B., Müller-Lucks, A., Kramer, G., Bukau, B. & von Heijne, G. Trigger factor reduces the force exerted on the nascent chain by a cotranslationally folding protein. *J Mol Biol* **428**, 1356-1364 (2016).

23. Nakatogawa, H. & Ito, K. Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. *Mol Cell* **7**, 185-192. (2001).

24. Yap, M.N. & Bernstein, H.D. The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel. *Mol Cell* **34**, 201-211 (2009).

25. Tsai, A., Kornberg, G., Johansson, M., Chen, J. & Puglisi, J.D. The dynamics of SecM-induced translational stalling. *Cell Rep* **7**, 1521-1533 (2014).

26. Gumbart, J., Schreiner, E., Wilson, D.N., Beckmann, R. & Schulten, K. Mechanisms of SecM-mediated stalling in the ribosome. *Biophys J* **103**, 331-341 (2012).

27. Butkus, M.E., Prundeanu, L.B. & Oliver, D.B. Translocon "pulling" of nascent SecM controls the duration of its translational pause and secretion-responsive secA regulation. *J Bacteriol* **185**, 6719-6722 (2003).

28. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* **36**, 299-304 (2005).

29. Batey, S. & Clarke, J. The Folding Pathway of a Single Domain in a Multidomain Protein is not Affected by Its Neighbouring Domain. *J Mol Biol*. **378**, 297-301 (2008)

30. Batey, S. & Clarke, J. Apparent cooperativity in the folding of multidomain proteins depends on the relative rates of folding of the constituent domains. *Proc Natl Acad Sci U S A* **103**, 18113-18118 (2006).

31. Batey, S., Randles, L.G., Steward, A. & Clarke, J. Cooperative folding in a multi-domain protein. *J Mol Biol* **349**, 1045-1059 (2005).

32. Wensley, B.G., Kwa, L.G., Shammas, S.L., Rogers, J.M. & Clarke, J. Protein folding: adding a nucleus to guide helix docking reduces landscape roughness. *J Mol Biol* **423**, 273-283 (2012).

33. Wensley, B.G. et al. Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **463**, 685-688 (2010).

34. Cabrita, L.D. et al. A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol* **23**, 278-285 (2016).

35. Cabrita, L.D., Dobson, C.M. & Christodoulou, J. Protein folding on the ribosome. *Curr Opin Struct Biol* **20**, 33-45 (2010).

36. Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco, I., Jr. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723-1727 (2011).

37. Scott, K.A., Randles, L.G. & Clarke, J. The folding of spectrin domains II: phi-value analysis of R16. *J Mol Biol* **344**, 207-221 (2004).

38. Wensley, B.G., Gartner, M., Choo, W.X., Batey, S. & Clarke, J. Different members of a simple three-helix bundle protein family have very different

folding rate constants and fold by different mechanisms. *J Mol Biol* **390**, 1074-1085 (2009).

39. Ingolia, N.T., Lareau, L.F. & Weissman, J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802 (2011).

40. Young, R. & Bremer, H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *Biochem J* **160**, 185-194 (1976).

41. Borgia, M.B. et al. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* **474**, 662-665 (2011).

42. Borgia, A. et al. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat Commun* **3**, 1195 (2012).

43. Hill, S.A., Kwa, L.G., Shammas, S.L., Lee, J.C. & Clarke, J. Mechanism of assembly of the non-covalent spectrin tetramerization domain from intrinsically disordered partners. *J Mol Biol* **426**, 21-35 (2014).

Figure legends

Figure 1. Three-helix bundle structure of spectrin domains. (**Top**) Location of key residues for folding. R16 folds via early formation of stabilizing interactions between the N-terminal regions helices of A and C (red). R15 folds via interactions between the C-terminal regions of helices A and C (blue and magenta). (**Bottom**) Positions of the five residues from R15 substituted into the A-helix of R16, to form R16m5 (green) and position of the V65L substitution in the B-helix of R16m5, to form R16m6 (gold).

Figure 2. Investigating cotranslational folding using an arrest-peptide assay. (**a**) (**Left**) Linker length (*L*) is short: the domain is unable to fold, translation arrest persists and no full-length protein is produced. (**Right**) Design of construct used in the force profile assays. (**b**) Intermediate linker length: the domain folds against the ribosome, translation arrest is released and full-length protein is produced. (**c**) Linker length is long: the domain folds away from the ribosome, the arrest persists and no full-length protein is produced.

Figure 3. Force Profiles. (**a**) β16, R15, R16 with non-folding controls. β16 (light blue), R15 (black), R16 (red). Non-folding (nf) variants of R15 and R16 do not produce a force. (**b**) Comparison of R16 with tandem proteins. R16 (red), R15R16 (blue), R153ProR16 (red with dashed line), R15R16m5 (purple with dashed line). (**c**) Comparison of R15 and R16 with R16m5 and R16m6. R15 (black), R16 (red), R16m5 (purple), R16m6 (green). (**d**) Comparison of R15 and R16 with R16o15c. R15 (black), R16 (red), R16o15c (green). For all plots each point corresponds to a single measurement. We have shown that measurements are reproducible: data collected using *in vitro* translation kits with different lot numbers and from different labs overlay (see **Supplementary Fig. S2).**

Figure 4. Visualization of the R16 spectrin domain at the ribosomal tunnel exit. (**a**) Cryo-EM reconstruction of the R16-TnaC [*L*=33] ribosome-nascent chain complex. The small ribosomal subunit in yellow, large subunit in grey, the peptidyl-tRNA with the nascent chain in green and additional density at the ribosomal tunnel exit in red. A

cross-section through the ribosome shows density for the nascent chain in the ribosomal tunnel and non-ribosomal density corresponding to the R16 domain (PDB: 1AJ3) at the tunnel exit. (**b**) Schematic representation of the construct used for cryo-EM is shown on the top. A close-up of the ribosomal tunnel exit shows density for the R16 domain located between ribosomal helix H59 and ribosomal protein uL24. (**c**) Rigid body fit of the NMR structure of the R16 domain colored according to RMSD **(blue: 0.5 to 1.9; white: 2–3.9; red: ≥ 4.0 Å)** to the cryo-EM density map at different contour levels.

Figure 5. Rigid body fit of the NMR structure of the R16 domain to the cryo-EM density map showing equivalent locations of R15 key folding residues. Regions that appear to be structured (for which there is density in the EM) are colored pink, regions that are apparently unstructured (for which there is no density in the EM) are colored yellow. The residues in the A-helix important for nucleating folding of R15 are shown in light blue; the V65L substitution in the B-helix is shown in green.

Table 1: Stability and folding rate constants of the spectrin domains used in this study.

	R15	**R16**	**β16**	**R16m5** (A version of R16 with 5 residues from R15 substituted in the A helix)	**R16m6** (As R16m5 but with an additional conservative, stabilizing mutation in the B helix)	**R16outside15core**[a] (R16o15c, a protein with all the core residues from R15 substituted into R16)
Stability ΔG (kcal mol^{-1})	Moderate (5.2 ± 0.1)	Moderate (5.0 ± 0.1)	High (9.2 ± 0.2)	Moderate (4.1 ± 0.2)	Moderate (4.8 ± 0.1)	Moderate (5.5 ± 0.2)
Folding rate constant k_f (s^{-1})	Very Fast (60,000 ± 20,000)	Slow (135 ± 27)	Fast (5200 ± 800[b])	Fast (4,300 ± 700)	Fast (13,700 ± 100)	Fast (2200 ± 300)
Region of the protein that folds earliest	C-terminal regions of helices A & C	N-terminal regions of helices A & C	ND	C-terminal regions of helices A & C	C-terminal regions of helices A & C	C-terminal regions of helices A & C

These data are for the 109 amino acid versions of spectrin used in the AP experiments, which are somewhat shorter than those used previously. Note that this truncation has lead to a loss of stability of ~ 0.5 - 1 kcal mol^{-1} but no change in folding rates. Extension of the domains by the linker restores the stability close to that of the original, longer constructs.

[a] Stability data for R16o15c are taken from Wensley *et al*[32] for the original construct which has an extra 7 residues at the C-terminus.

[b] Folding rate constant for β16 is taken from Hill *et al*[43] for the original construct which has an extra 7 residues at the C-terminus.

Online Methods

Enzymes and chemicals

All enzymes were obtained from Thermo Scientific (Waltham, MA, USA) and New England Biolabs (Ipswich, MA, USA). Oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg, Germany). DNA/RNA purification kits were purchased from Qiagen (Hilden, Germany) and the PURExpress™ cell-free translation system was obtained from BioNordika (Stockholm, Sweden). [^{35}S]-methionine was purchased from PerkinElmer (Waltham, MA, USA). All other reagents were from Sigma-Aldrich (St. Louis, MO, USA).

DNA manipulations

All spectrin constructs were synthesized from the previously described pET19b plasmid (Novagen, Madison, WI, USA) carrying a truncated *lepB* gene containing a [6L,13A] H segment insert and the *Escherichia coli* SecM arrest peptide, FSTPVWISQAQGIRAGP, under the control of a T7 promoter[20]. The soluble, non-membrane targeted LepB derivative was generated by a deletion of codons 4-77 using PCR, corresponding to the removal of transmembrane segments 1 and 2. The resulting plasmid was digested with *SpeI* and *KpnI* to release the [6L, 13A] segment and this was substituted by segments encoding the 109 residues of spectrin β16, R15, R16, R16m5, R16m6, R16o15c[32,37,43] or the 215 residues of spectrin R15R16[30], using the megaprimer approach[44]. The spectrin genes were all truncated C-terminally (relative to previous studies) to end one amino acid after the last conserved leucine residue (http://pfam.xfam.org/): 5'-WERLR-3' for β16, 5'-RAKLN-3' for R15 and 5'-GQRLE-3' for R16, R16m5, R16m6, R16o15c and R15R16. Starting with the longest linker, truncations were generated by PCR using partially overlapping oligonucleotides containing the respective sequence alteration[45]. By the same

514 truncation method, the constructs shown in **Fig. 3b** (R16, R15R16, R153ProR16,
515 R15R16m5) were generated with the N-terminal LepB leader segment removed; the
516 R15R16m5 construct was created by ligating R16m5 in place of R16 using the
517 megaprimer method. All constructs were verified by DNA sequencing. For each
518 linker length, *L*, site-directed mutagenesis was performed to generate constructs with
519 the non-functional FSTPVWISQAQGIRAGA arrest peptide as full-length controls.
520 At linker lengths exhibiting substantial force, (*L* = 37 and 41 residues, **Fig. 3a**) site-
521 directed mutagenesis was performed to generate non-folding double mutants of R15
522 (F18D+I55D) and R16 (F11D+L55D)[37,38].
523 *In vitro transcription and translation*
524 For the generation of nascent chains, all genes were first PCR-amplified as DNA
525 constructs containing a 5' T7 promoter. mRNA transcripts were subsequently
526 generated by *in vitro* transcription using T7 RNA polymerase according to the
527 manufacturer's instructions (Promega). RNA was purified using the RNeasy Mini Kit
528 (Qiagen). Translation was performed using the commercially available PURExpress™
529 *in vitro* system according to the manufacturers' protocol (New England Biolabs).
530 Synthesis of [^{35}S]-Met labeled polypeptides was performed at 37 °C, 500 rpm for
531 exactly 15 minutes. The reaction was quenched by the addition of an equal volume of
532 10% ice-cold trichloroacetic acid (TCA). The samples were incubated on ice for 30
533 min and centrifuged for 5 min at 20,800 x *g* and 4°C. Pellets were dissolved in sample
534 buffer and treated with RNase A (400 μg ml^{-1}) for 15 min at 37 °C before the samples
535 were resolved by SDS-PAGE and imaged. Bands were quantified using ImageJ to
536 obtain an intensity cross section, (http://rsb.info.nih.gov/ij/), which was subsequently
537 fit to a Gaussian distribution using in-house software.

Each data point on the force profiles is a single experiment. The data are reproducible; data collected using *in vitro* translation kits with different lot numbers and from different labs overlay (**Supplementary Fig. S2**). We also point to the many different forms of R16 shown in **Fig. 3:** the profiles for R16 (**Fig. 3a**), R16 no Lep leader, R153ProR16 (**Fig. 3b**), and R16m6 all overlay.

Cloning and purification of R16-TnaC [L=33] ribosome-nascent chain complexes

The construct used for cryo-EM was derived from the spectrin R16 construct used for *in vitro* force profile analysis. The SecM stalling sequence was replaced by an equivalent TnaC[46] stalling sequence conserving a linker length of (*L*=33) to the spectrin domain. An N-terminal His-tag was introduced for purification. The construct was subsequently cloned into the pBad vector for *in vivo* expression of R16-TnaC RNCs. *In vivo* expression and purification was carried out as described previously[46]. The final spectrin-TnaC sequence used was HGHHHHHHHHDYDIPTTLEVLFQGPGTKLNESHRLHQFFRDMDDEESWIKEK KLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDN TIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGSGSGSGGPNILHISVTSK WFNIDNKIVDHRP, including an N-terminal His_8-tag for purification and 3C cleavage site (spectrin R16 sequence underlined).

Cryo-EM specimen preparation, data collection and processing

Freshly prepared spectrin-TnaC RNCs (~ 5 OD_{260}) were applied to 2 nm precoated Quantifoil R3/3 holey carbon supported grids and vitrified using a Vitrobot Mark IV (FEI Company). Cryo-EM data was automatically collected on a Titan Krios TEM (FEI, USA) operated at 300 kV and equipped with a back-thinned Falcon II (FEI, USA) direct electron detector. Ribosomal particles on the micrographs were detected using SIGNATURE[47]. All single-particle processing was performed using the

563 SPIDER software package[48]. The data-set contained a total of 177,543 particles.
564 Classes were obtained by competitive projection matching[48,49], and sorted for t-RNA
565 presence as well as extra density near the tunnel exit. The final class contained 46,067
566 particles and the corresponding electron density map showed strong density for t-
567 RNA in the P-Site as well as extra density at the ribosomal tunnel exit. Different
568 folding intermediates could not be enriched despite different classification attempts.
569 The final dataset was refined to a final average resolution of 4.8 Å ($FSC_{0.143}$) (See
570 **Supplementary Fig. S3**). To exclude potential overfitting, the data were processed
571 using a frequency limited refinement protocol by truncating high frequencies (low-
572 pass filter at 8 Å) during the whole refinement process[50]. For interpretation of the
573 cryo-EM density, we fitted the structure of the *E. coli* 70S ribosome (PDB ID: 3OFR),
574 using UCSF Chimera[51]. The fit was assessed by comparing the cross-correlation of
575 the EM map and the density of the model map. Density not corresponding to proteins
576 or RNA of the 50S subunit within the ribosomal tunnel was identified as nascent
577 chain. The non-ribosomal density at the ribosomal tunnel exit was compared with
578 density maps of the spectrin R16 domain (PDB ID: 1AJ3) at different resolutions and
579 contour levels. The parts of the spectrin domain having low RMSDs in the NMR
580 structure show corresponding density in the EM map. Thus, the NMR model of the
581 spectrin domain was rigid-body fitted according to the highest cross-correlation
582 between the density model maps and the non-ribosomal electron density.
583 All figures showing molecular models and electron densities were prepared
584 with the software UCSF Chimera[51]. Segmentation of the maps was performed after
585 docking a ribosomal model into the density and using the command color zone in
586 Chimera. Some parts of the map are shown at different contour levels.

Expression and Purification of Isolated Spectrin Domains

This was performed as described previously[38].

Measurement of Stability and Kinetics of Isolated Spectrin Domains

Stability and kinetics measurements of the isolated 109 amino acid spectrin domains and C-terminally truncated variants were performed in 50 mM phosphate buffer (pH 7.0) at 25±°C as described previously[38].

Accession codes. Coordinates for the cryo-EM map of the ribosome with the spectrin density have been deposited at the Electron Microscopy Data Bank under accession code EMD 3451. Coordinates of the spectrin model used for fitting and interpretation of the cryo-EM map have been deposited at the Protein Data Bank under accession code PDB 5M6S.

Data Availability: The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Methods-only References

44. Ke, S.H. & Madison E.L. Rapid and efficient site-directed mutagenesis by single-tube "megaprimer" PCR method. *Nucleic Acids Res* **25**, 3371-3372 (1997).

45. Zheng, L., Baumann, U. & Reymond, J.L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**, e115 (2004).

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

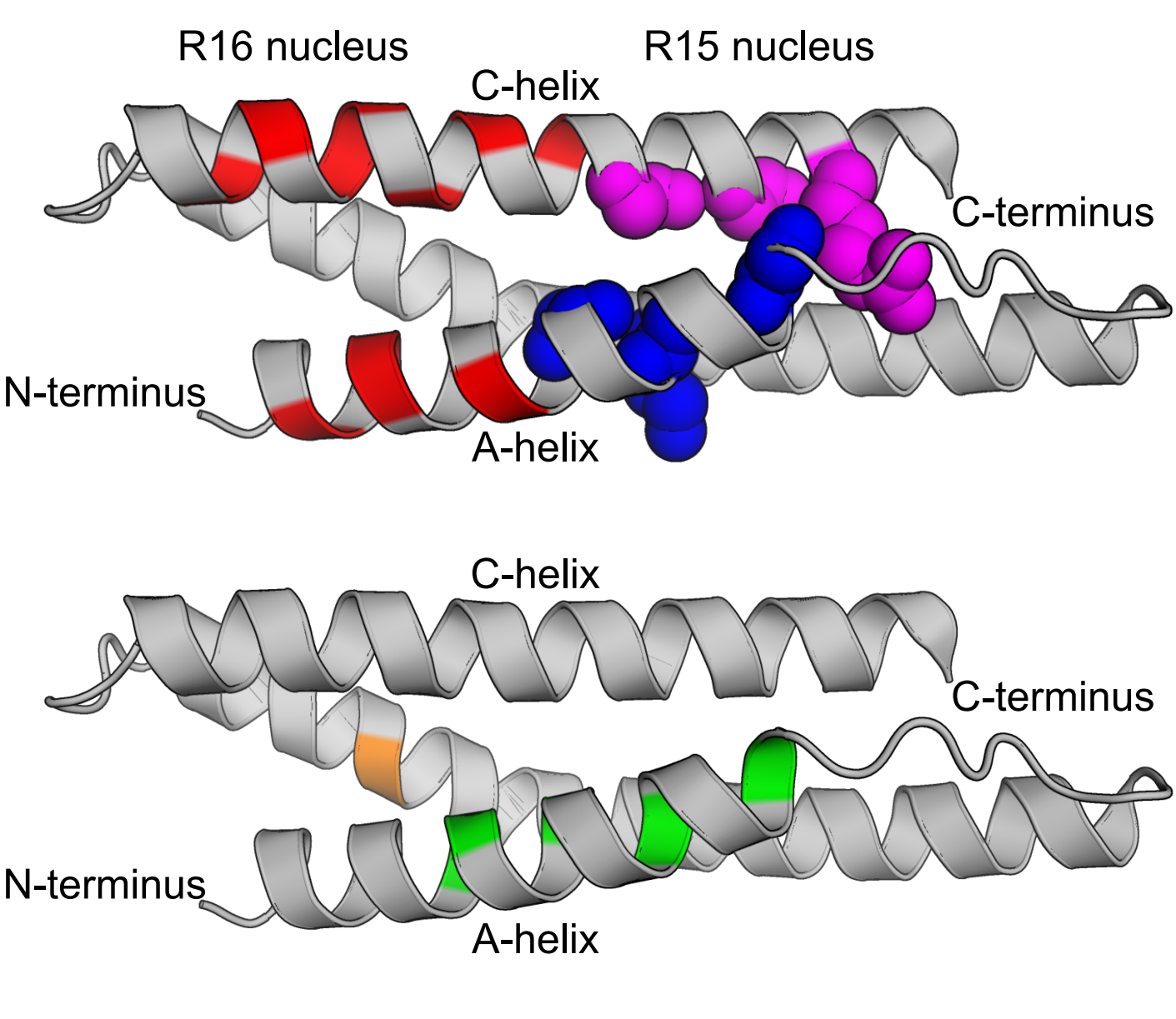
47. Chen, J.Z. & Grigorieff, N. SIGNATURE: a single-particle selection system for molecular electron microscopy. *J Struct Biol* **157**, 168-173 (2007).

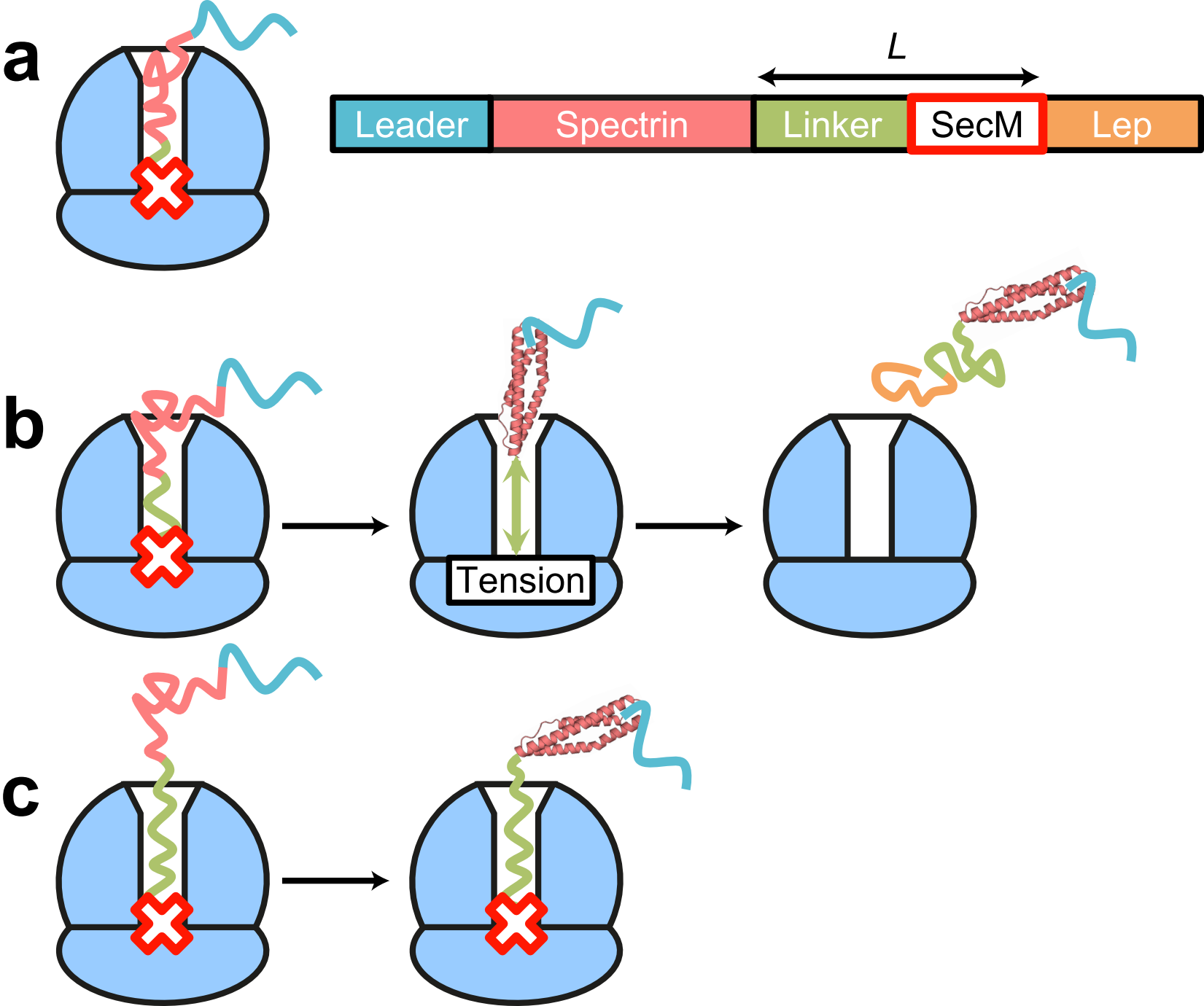
48. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190-199 (1996).

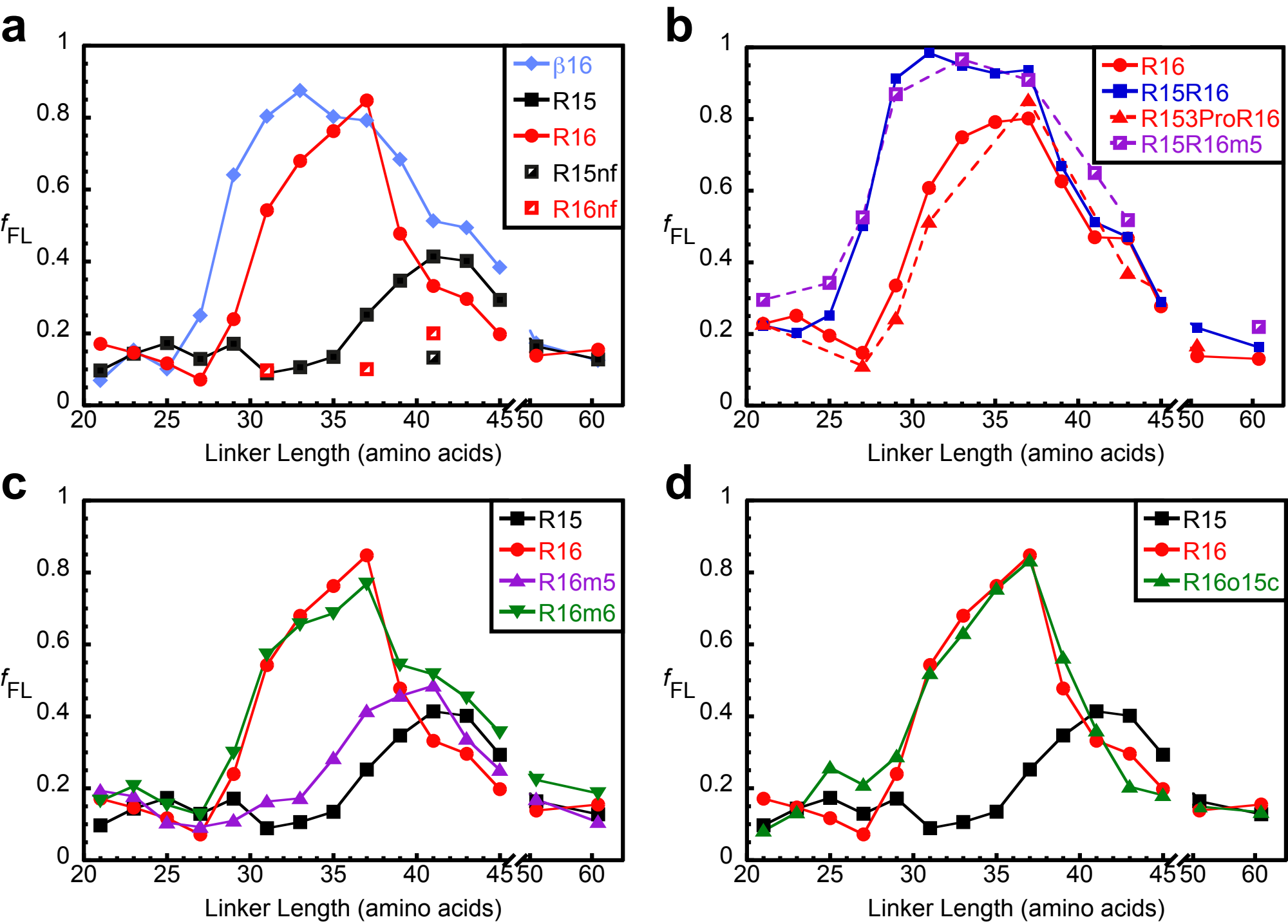
49. Penczek, P.A., Frank, J. & Spahn, C.M. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J Struct Biol* **154**, 184-94 (2006).

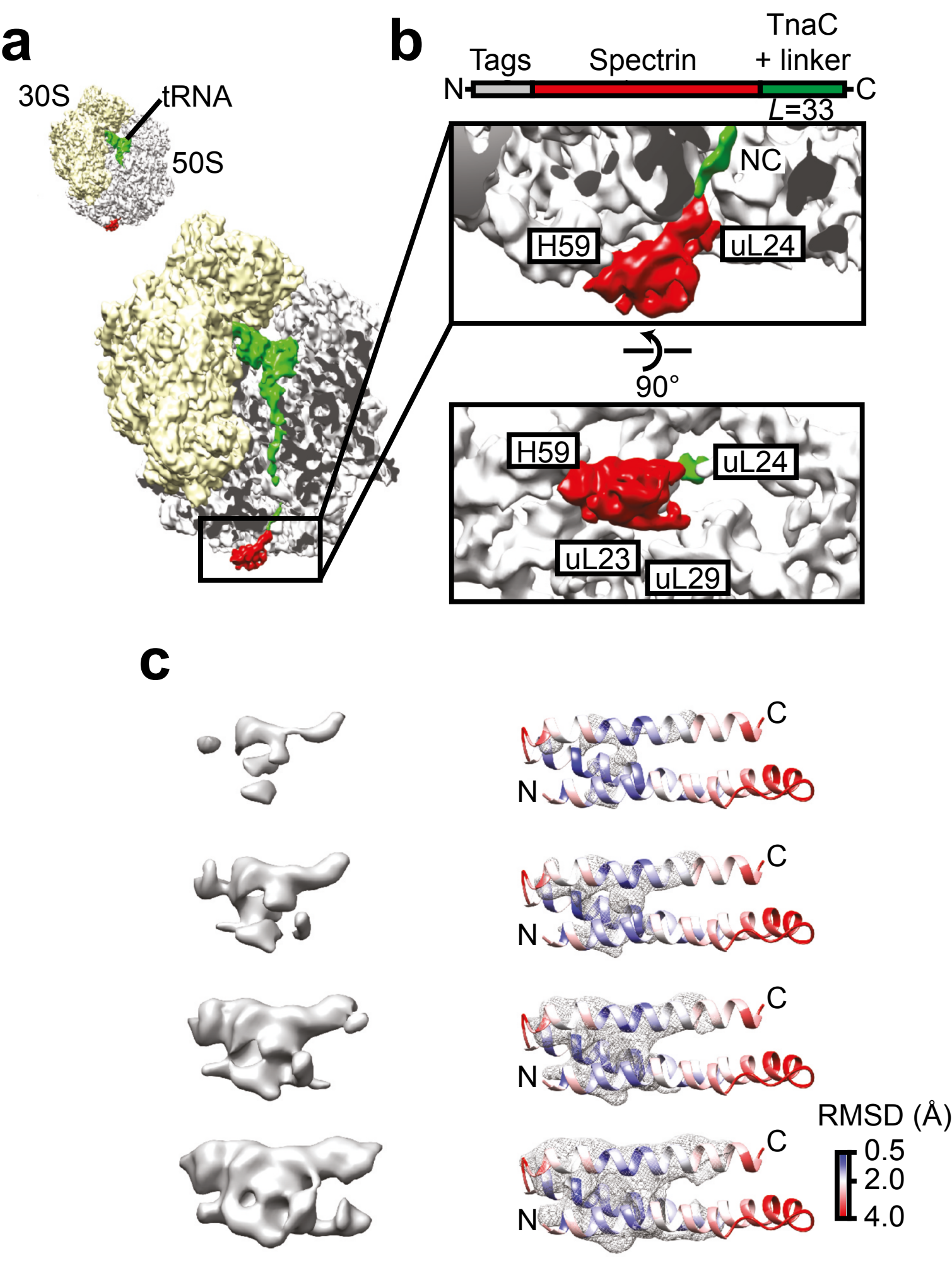
50. Scheres, S.H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat Methods* **9**, 853-854 (2012).

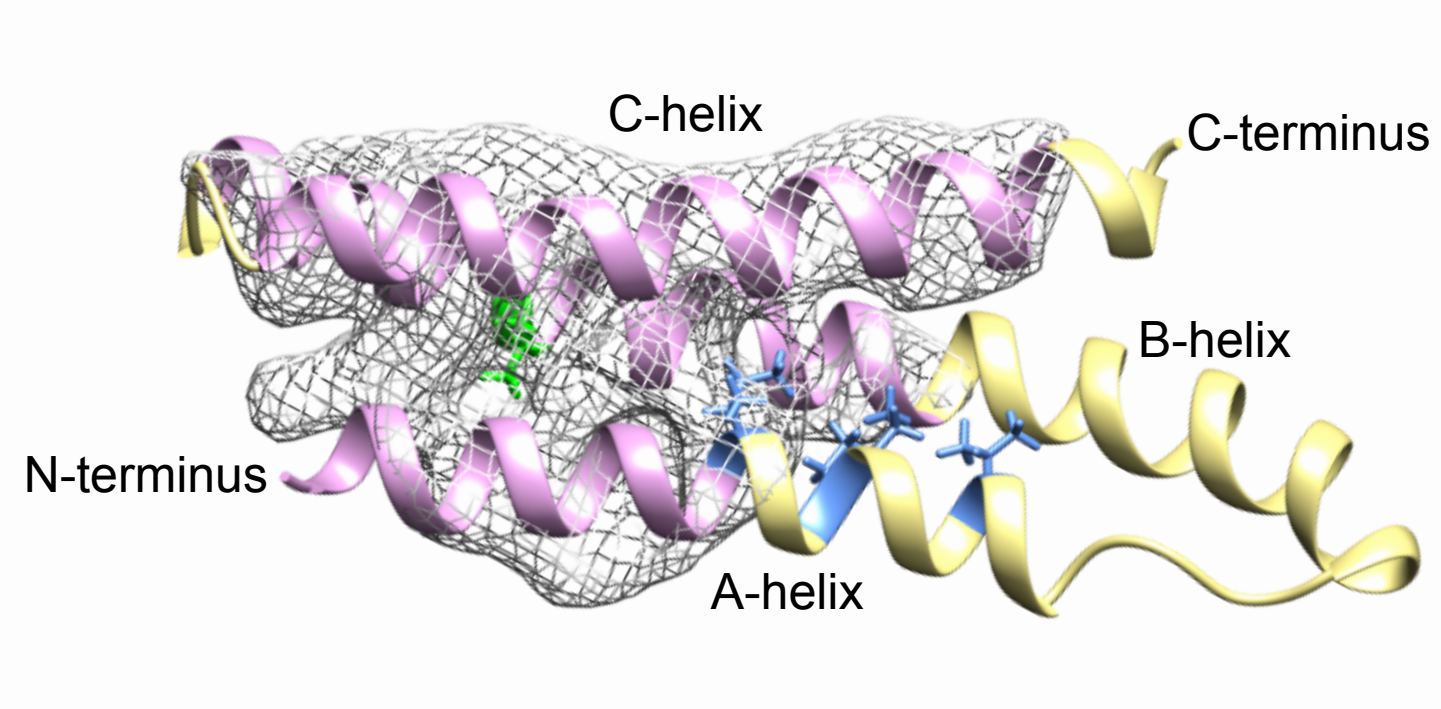
51. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).


R16 nucleus
R15 nucleus
C-helix
C-terminus
N-terminus
A-helix
C-helix
C-terminus
N-terminus
A-helix


a
L
Leader
Spectrin
Linker
SecM
Lep
b
Tension
c


a
b
c
d
f_{FL}
Linker Length (amino acids)
β16
R15
R16
R15nf
R16nf
R15R16
R153ProR16
R15R16m5
R16m5
R16m6
R16o15c


a
30S
tRNA
50S
b
Tags
Spectrin
TnaC
+ linker
N
C
L=33
NC
H59
uL24
90°
H59
uL24
uL23
uL29
c
C
N
C
N
C
N
C
N
RMSD (Å)
0.5
2.0
4.0


C-helix
C-terminus
B-helix
N-terminus
A-helix