Version May 28, 2019

Cotranslational folding cooperativity of contiguous domains of α-spectrin

Grant Kemp[1], Ola B. Nilsson[1,2], Pengfei Tian[4], Robert B. Best[5], Gunnar von Heijne[1,3*]

[1]Department of Biochemistry and Biophysics,
Stockholm University, SE-106 91 Stockholm, Sweden
[2]Department of Clinical Neuroscience, Therapeutic Immune Design Unit,
Karolinska Institutet, SE-171 77 Solna, Sweden
[3]Science for Life Laboratory Stockholm University, Box 1031, SE-171 21 Solna, Sweden
[4]Protein Engineering, Novozymes A/S, Krogshøjvej 36, DK-2880 Bagsvaerd, Denmark
[5]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[*]Corresponding author: GvH. Phone: Int+46-8-16 25 90, E-mail: gunnar@dbb.su.se

Abbreviations: PCR, polymerase chain reaction; TCA, trichloroacetic acid; SDS-PAGE, sodium dodecylsulphate polyacrylamide gel electrophoresis; PTC, peptidyl transferase center

Keywords: protein folding, spectrin, repeat protein, translational arrest peptide, SecM, molecular dynamics

Abstract

Proteins synthesized in the cell can begin to fold during translation, before the entire polypeptide has been produced. Previously, we investigated the cotranslational folding of three different repeats of α-spectrin by Force Profile Analysis (FPA), and compared the results to *in vitro* folding data obtained with purified proteins. We observed that cotranslational folding of the 16th α-spectrin repeat (the R16 domain) commenced at a shorter nascent chain length when it was preceded by the R15 domain compared to when being produced alone. Here, we extend this result to study how the cotranslational folding behavior of the R15 and R16 domains are affected by their neighboring R14 and R17 domains. Our results show that the domains impact each other's folding in distinct ways, that may be important for the efficient assembly of α-spectrin. We further estimate that the R16 domain can exert a force on the nascent chain of ~15 pN during cotranslational folding.

Introduction

Classically, protein folding has been studied *in vitro* using purified proteins, providing a detailed picture of how the biophysical properties of a given polypeptide chain contribute to the final structure. However, studies of purified proteins do not reflect the vectorial nature of biological protein synthesis, where each amino acid is added sequentially on a time scale that allows the growing protein to explore its continually expanding energy landscape [1].

The discovery [2,3] and engineering [4,5] of translational arrest peptides (APs) has provided us with a new tool to examine how proteins fold cotranslationally. To date, the folding of several small proteins and protein domains has been studied using the force-sensitive AP from the *E. coli* SecM protein with a method we have called Force-Profile Analysis (FPA) [6–15]. Recently, we demonstrated that FPA faithfully picks up cotranslational protein folding events observed by direct biophysical measurements [15].

Spectrin, the primary structural component of the cellular cytoskeleton [16–19], is composed of two heterodimers of α- and β-spectrin that come together in a coiled tetramer [20]. The 285 kDa non-erythroid αII isoform of spectrin contains 20 three-helix bundle domains, which provide interesting model systems for protein folding. This is because, while they are structurally nearly identical [21], they vary in stability [22–25], folding rate [26,27], and folding mechanism [28]. In particular, folding of the 15th, 16th, and 17th α-spectrin domains (herein called R15, R16, and R17, respectively) has been extensively studied *in vitro* [24,27,29–32].

Previously, using FPA, we demonstrated that R15 and R16 are able to fold cotranslationally and that the point at which this folding begins does not seem to be influenced by the *in vitro* folding rate [9]. Specifically, while R15 folds very quickly *in vitro*, cotranslational folding of R16 starts at shorter chain lengths than does folding of R15. Indeed, even when the folding core of R15 was substituted into R16, which gives R16 the folding properties of R15 *in vitro*

[33], the protein was still able to fold earlier than R15 during translation. When the minimal R15 folding nucleus (five residues) [34] was substituted into R16, this led to late cotranslational folding, but early folding was restored when this mutant was expressed in tandem with R15, as it would be *in vivo*. This demonstrated that the mutant R16 variants fold cotranslationally via early intrachain contacts, at chain lengths where the R15 folding nucleus cannot yet form. This was not observed *in vitro* because the purified mutant R16 protein would preferentially fold using the "fast" R15 nucleus rather than the slower R16 contacts. Cotranslationally, R16 folds using the first available pathway, i.e., the slower R16 folding contacts.

Here, we have extended our studies on spectrin to determine how the presence of neighboring domains affect the folding of R15 and R16. We find that the cotranslational folding of these two domains is differently affected by their respective up- and downstream neighbor domains, a previously unappreciated feature of spectrin folding. We further provide an estimate for the magnitude of the pulling force exerted on the nascent chain by cotranslational protein folding.

Results

The force profile analysis (FPA) assay

APs are short stretches of polypeptide that interact with the ribosome exit tunnel in such a way that translation is stalled when the last codon in the portion of the mRNA that codes for the AP is located in the ribosomal A-site [35]. APs generally have regulatory functions in cells, controlling, *e.g.*, translation initiation on downstream ORFs in polycistronic mRNAs [35]. Interestingly, the stalling efficiency of many APs has been shown to be sensitive to pulling forces exerted on the nascent polypeptide chain, with high pulling force leading to reduced stalling [6,36,37]. APs can therefore be used as force sensors, reporting on cotranslational events that in one way or another generate pulling forces on the nascent chain [7,37,38].

In FPA, a force-generating domain in a polypeptide is placed at increasing distances upstream of an AP, Fig. 1a, and the degree of translational stalling is measured for the corresponding series of protein constructs. Figure 1b shows how FPA can be applied to study the cotranslational folding of soluble protein domains, *e.g.*, tandem spectrin domains. In the construct shown on the left, the chain is long enough for one of the two spectrin domains to have already folded, while the second spectrin domain is largely buried in the exit tunnel and thus cannot fold when the ribosome reaches the C-terminal end of the AP. Therefore, little force is generated on the AP, and stalling is efficient. In the right-hand construct, the chain is so long that both spectrin domains have folded prior to the ribosome reaching the C-terminal end of the AP, and again little force is generated on the AP. In the middle construct, however, the chain is just long enough that the second spectrin domain can begin to fold if the linker that connects it to the AP is stretched beyond its equilibrium length. Under this condition, some of the free energy gained upon folding will be converted to tension in the linker and generate a pulling force on the AP, resulting in reduced stalling and increased production of full-length protein.

The cotranslational folding process can hence be followed by measuring the amount of full-length and arrested protein product at each linker length, Fig. 1c. Folding transitions will appear as peaks in a "force profile" plot of the fraction full-length protein (f_{FL}) vs. linker length (L), Fig. 1d, and can be described by the linker lengths that define the onset (L_{onset}), maximum (L_{max}), and end (L_{end}) of the peak, as shown.

Cotranslational folding of tandem spectrin domains

In the intact spectrin protein, the C-terminal α-helix in one domain is continuous with the N-terminal α-helix in the next domain, Fig. 2a. Consecutive domains are thus intimately connected to each other, and can influence each other's thermodynamic stability [24,27,31,39]. In order to better understand the cotranslational folding of spectrin, we decided to systematically

evaluate the effects of up- and downstream domains on the force profiles of the R15 and R16 spectrin domains by analyzing the two- and three-domain combinations shown in Fig. 2b. For each domain combination, a force profile was recorded by generating a series of constructs where the 17-residue AP from the *E. coli* SecM protein was placed at different linker lengths *L* from the C-terminal end of the R15 or R16 domain, translating each construct for 20 minutes in the PURExpress *in vitro* translation system [40] in the continuous presence of [^{35}S]-methionine, separating the translation products by SDS-PAGE, and calculating f_{FL} from the intensities of the bands corresponding to the full-length and arrested forms of the protein. The full amino acid sequences of all constructs are given in Supplementary Table 1.

In order to distinguish which spectrin repeat is being assayed in the tandem constructs, we use the following naming convention. Constructs named "nL" (for ***n*** residues of **L**epB linker) are used to follow the folding of most C-terminal spectrin repeat (i.e, R15R16 nL follows the folding of R16) and constructs named "nT" (for ***n*** residues of **T**runcated spectrin linker) follow the folding of the penultimate spectrin repeat (i.e. R15R16 nT follows the folding of R15).

The results are shown in Fig. 3, with the previously reported results for the individual R15 and R16 domains [9] included for reference in Fig. 3a. To facilitate comparison of the results, the L_{onset}, L_{max}, and L_{end} values extracted from each force profile (see Supplementary Fig. S1) are summarized in Fig. 3d. As seen in Fig. 3b, the cotranslational folding of the R15 domain is unaffected by the presence of the upstream R14 domain (dark blue and light blue curves), while the presence of the N-terminal part of the downstream R16 domain induces a reduction in L_{onset} from 36 to 33 residues and a marked increase in the amplitude of the peak (compare the dark blue and purple curves). Thus, somewhat counterintuitively, the onset of folding of R15 is sensitive to the presence of the early parts of the N-terminal α-helix of the downstream R16 domain, but is not affected by the presence of the upstream R14 domain.

For the R16 domain, the effects of the neighboring R15 and R17 domains are more dramatic, Fig. 3c. The presence of the upstream R15 domain leads to a reduction in L_{onset} from 30 to 27 residues and an increase in the amplitude of the peak (compare the red and purple curves), as shown before [9]. The presence of the downstream R17 domain does not appreciably affect L_{onset} but leads to an increase in amplitude and a shift to a higher L_{end} value (compare the red and orange curves). Finally, when flanked by both the R15 and R17 domains, the R16 folding transition starts at a lower L_{onset}, ends at a higher L_{end}, and has a higher amplitude than for the isolated R16 domain (compare the red and brown curves).

Interestingly, both R16 nT constructs that include parts of R17 show an increase in f_{FL} at the longest linker length (L = 61 residues). At this point the entire R17 helix A, the loop, and the very beginning of R17 helix B have been translated (see Supplementary Table 1). This increase in f_{FL} may thus signal an early interaction between helices A and B in R17.

We conclude that the cotranslational folding of the R15 domain is affected by its downstream but not by its upstream neighbor domain, and that R16 is affected both by its upstream and downstream neighbors. Different spectrin domains thus not only fold via different folding mechanisms, but their cotranslational folding is also differently impacted by their up- and downstream neighbor domains, a previously unappreciated feature of spectrin folding.

A quantitative relation between f_{FL} and pulling force

In all FPA studies published to date, the quantitative relation between the calculated f_{FL} values and the underlying pulling force acting on the nascent chain has remained undefined (although attempts have been made to derive it from simulations or other kinds of theoretical modeling [11,38]. Using the PURE translation system, Goldman *et al.* [6] showed that the interactions between the SecM AP and the ribosome exit tunnel can be disrupted by a mechanical force applied through optical tweezers, and that the rate by which the

translational stall induced by the SecM AP is released, k_R, increases in step with the amount of force, F, applied to the nascent chain, in a way that can be approximated by the Bell model [41] for force-induced rupture:

$$k_R \approx k_0 \exp[\frac{F\Delta x^{\ddagger}}{k_B T}] \qquad [1]$$

where k_0 = 3 x 10^{-4} s^{-1} (95% confidence interval (CI): 0.5 x 10^{-4} s^{-1}, 20 x 10^{-4} s^{-1}) is the release rate at zero pulling force, and $\Delta x^{\ddagger}$ = 0.4 nm (95% CI: 0.1 nm, 0.8 nm) is the distance to the transition state.

The work by Goldman *et al.* [6] suggested to us that approximating k_R with the rate of conversion of the arrested form of a given construct to the full-length form as measured in a pulse-chase experiment would allow us to estimate the corresponding pulling force using relation [1] between F and k_R. To this end, we did pulse-chase experiments on a range of spectrin R16 and ADR1a [7] constructs that have f_{FL} values between 0.2 and 0.9, and fit the pulse-chase data to a first-order kinetic equation, Supplementary Fig. S2.

To derive an analytical relation between f_{FL} and F, we reasoned that if f_{FL} were measured at a single delay time (Δt) in a pulse-chase experiment, rather than under the standard continuous-labeling experimental conditions, the release rate (and hence F, according to Eqn. [1]) could be estimated from:

$$f_{FL} = 1 - e^{-k_R \Delta t} \qquad [2]$$

Since in the standard experiment translation can be initiated at any time during the 20 min. labelling period, the delay time Δt varies from one ribosome-nascent chain complex to another. Although many factors could in principle contribute to the distribution of delay times, it turns out empirically that using an average delay time Δt = 550 s, approximately equal to half of the total incubation time, describes remarkably well the relation between the

standard f_{FL} values (20 min. continuous [^{35}S]-Met labeling) and the release rates measured by pulse-chase experiments, Supplementary Fig. S3.

Combining equations [1] and [2], one can solve for the pulling force F:

$$F \approx \frac{k_B T}{\Delta x^{\ddagger}} \ln\left[\frac{\ln(1-f_{FL})}{-k_0 \Delta t}\right] \qquad [3]$$

To obtain good estimates of the Bell parameters k_0, $\Delta x^{\ddagger}$ we started from the values obtained in [6] and performed a local optimization by requiring that k_0 (the release rate at zero force) is close to k_R measured for the "zero-force" construct R16 nL27, Fig. 4a, and that the titin I27, spectrin R16, and ADR1a force profiles in [12] are well reproduced by the molecular dynamics simulation also described in [12]. These conditions are both fulfilled by setting $k_0 = 3 \times 10^{-4}$ s^{-1} (*i.e.*, the same value as in [6] and equal to k_R for R16 nL27, Fig. 4a) and $x^{\ddagger} = 0.65$ nm (well inside the 95% CI from [6]), Supplementary Fig. S4. With these parameter values and $\Delta t = 550$ s, equation [3] nicely captures the relation between the force estimated from the measured k_R and standard f_{FL} values, Fig. 4b, and can hence be used to predict F from f_{FL}. Over the interval $0.2 < f_{FL} < 0.9$, the experimental data is also well approximated by the simple linear fit $F = 22 f_{FL} - 2.2$ (blue line). Obviously, [3] holds only for the 17-residue *E. coli* SecM AP used here, and not for other APs of different stalling strengths [38]. Presumably, [3] may be applied also in other contexts where the SecM AP is used to measure pulling forces, such as during membrane protein synthesis or translocation of charged residues across energized membranes [37,38], although the parameters would first need to be verified to hold for *in vivo* experiments. We note that another approach to determining Bell parameters under a given set of conditions would be to repeat the FPA experiments using multiple APs with different resistance to force. A global fit of f_{FL} obtained for each AP as a function of linker length L would allow simultaneous determination of the forces exerted by the protein (independent of AP) as well as AP-dependent Bell parameters. In Supplementary Fig. S5 we illustrate this

approach using data for translocon-mediated transmembrane helix insertion into the inner membrane of *E. coli* from *in vivo* AP measurements.

The force estimated from pulse-chase measurements or f_{FL} values is an effective force which contains contributions from both folded and unfolded states. It is not exactly the same thing as the ensemble average force calculated from MD simulations, $\langle F_{MD} \rangle = p_U F_U + p_F F_F$, where $p_{U(F)}$ and $F_{U(F)}$ are, respectively, the population of, and force exerted by, the unfolded (folded) state in the simulation. In previous work [12], we calculated both f_{FL} as well as $\langle F_{MD} \rangle$ directly from MD simulations. In Supplementary Fig. S6a, we show the relation between $\langle F_{MD} \rangle$ and f_{FL} from our previous simulations [12] , which is in good agreement with the relation between F and f_{FL} calculated from [3]. This agreement also confirms that the pulling forces are small enough that the force F calculated from simulated f_{FL} values using [3] is close to the ensemble-average force $\langle F_{MD} \rangle$ determined by MD simulations. This would not be true in general because of the non-linear relation between release rate and force, such that states exerting a larger force (e.g. the folded state) contribute disproportionately to the average release rate (a direct comparison between $\langle F_{MD} \rangle$ from the MD simulations and F calculated from [3] is given in Supplementary Fig. S6b).

Discussion

α-spectrin contains over 20 repeat domains and is produced as a single, long polypeptide. Previously, we observed that the R16 domain starts to fold while a part of its C-terminal α-helix is still in the ribosome exit tunnel [9]. We now find that the cotranslational folding of R16 is affected both by the presence of its upstream R15 domain, and when the N-terminal part of the following R17 domain is present in the exit tunnel. In contrast, folding of the R15 domain is not affected by the upstream R14 domain, but starts at shorter linker lengths when the N-terminal end of the following R16 domain is present.

In vitro, R16 is stabilized by ~1.7 kcal/mol by the presence of R15 [42], meaning that, during cotranslational folding, R16 can "sacrifice" a few interactions in the folding nucleus and start to fold at a shorter linker length in the R15R16 nL constructs [9]. Likewise, the early part of the N-terminal α-helix (helix A) in R17 makes a few interactions with the tip of the R16 domain [43,44], Fig. 2a, presumably explaining why L_{end} is increased in the R16R17 nT constructs.

For R15, the situation is a bit different because, in contrast to R16, its folding nucleus involves residues located at the C-terminal end of helix C [45]. The presence of the upstream R14 domain thus may have little impact on L_{onset}, because the whole R15 domain anyway must have emerged from the exit tunnel before folding can start. On the other hand, the R15 C helix may be stabilized by the presence of the early parts of the R16 N-terminal α-helix (helix A), possibly allowing the folding nucleus to form at a shorter linker length and reducing L_{onset} in the R15R16 nT constructs.

The picture that emerges is that the spectrin domains fold one after the other as they emerge from the exit tunnel, but not completely independently of one another. As seen for the R16 domain, the presence of an already folded N-terminal upstream neighbor domain not only can increase the thermodynamic stability of the folded state of an emerging domain, but can also allow the emerging domain to start folding while still partly buried in the exit tunnel. Likewise, when the C-terminal α-helix in the emerging domain is extended by a part of the N-terminal α-helix in the following domain, the folding transition can persist to longer linker lengths (as seen for R16) or start at shorter linker lengths (as seen for R15). Finally, the increase in f_{FL} seen for the R16R17 nT and R15R16R17 nT constructs at L = 61 residues (Fig. 3c) is suggestive of the formation of an early folding intermediate in R17. The cotranslational folding of spectrin may thus be seen as an almost continuous folding reaction, punctuated by periods when the nascent chain is lengthened but no residues are added to the folded part emerging from the ribosome. In this scenario, at most a short stretch of unfolded

polypeptide would be exposed outside the ribosome at any one time, minimizing the need for protection of the nascent chain by chaperones.

Finally, a pulse-chase analysis allowed us to measure the release rate k_R from the stalled state for different R16 and ADR1a constructs, making it possible to estimate the magnitude of the pulling forces exerted on the AP for different linker lengths, Fig. 4a. This in turn makes it possible to derive expressions for how the magnitude pulling force F depends on f_{FL} (as obtained from our standard continuous-labeling experimental protocol), Eqns. [2] and [3]. In general, we find that the folding of typical protein domains such as spectrin R16 and ADR1a can generate a maximal force of 15-20 pN on the AP, Fig. 4, in line with theoretical estimates based on MD simulations[46,47].

Materials and Methods

All the enzymes used for molecular biology were obtained from New England Biolabs. The PURExpress *in vitro* transcription-translation system was produced by Eurogentec and purchased via BioNordika. The PURExpress *in vitro* transcription-translation system used for translation rate determinations was purchased from NEB. GeneJet PCR clean-up kit was purchased from Thermo-Fisher Scientific. Precast NuPAGE gels and running buffers were purchased from Invitrogen. Homemade gels were cast using acrylamide-bisacrylamide mix, Tris, and glycine from VWR. Filter paper for gel drying was from Whatman. All other chemicals were purchased from Sigma.

Cloning

The cDNAs for spectrin R15 and R16 domains were kindly provided by Dr. Jane Clarke and the spectrin R17 and R14 domains were ordered as DNA fragments from Eurofins Genomics GmbH. We used two different construct types: those with an unrelated linker sequence,

labelled nL, and those with a truncated spectrin linker sequence, labelled nT (Fig. 2b). For nL constructs, the respective spectrin domain or domains were cloned into the pET19b plasmid upstream from a Ser-Gly-Ser-Gly sequence attached to a linker sequence derived from the P2 domain of leader peptidase (LepB) followed by the 17-residue AP from the *E. coli* SecM protein (FSTPVWISQAQGIRAGP), and a further 23 amino acids derived from LepB (protein sequence in Supplementary Table 1). The number following "nL" of each construct denotes the number of amino acids from the start of the linker sequence to the C-terminal Pro residue of the SecM AP. The shortest constructs (nL 21) include only the SGSG sequence fused to the AP. For nT constructs, the respective spectrin domains were cloned into the same pET19b plasmid without a LepB linker sequence, but including the SGSG sequence immediately before the SecM AP. The most C-terminal spectrin domain was then sequentially truncated via partially overlapping inverse PCR. The number following "nT" denotes the number of residues remaining in the C-terminal spectrin domain. For each construct, a full-length (FL) control was created by changing the critical Pro at the C-terminal end of the SecM AP to Ala, thereby abolishing arrest [48] and yielding only full-length protein. All sequences were confirmed by DNA sequencing (Eurofins Genomics GmbH).

In vitro expression

Expression and analysis was carried out as described previously [9]. Briefly, a linear DNA product is created from each construct plasmid by PCR using Q5 polymerase with forward and reverse primers that anneal to the T7 promoter and terminator regions, respectively. Following PCR clean-up (using the manufacturer's instructions) and confirmation of product by agarose gel electrophoresis, 1 µL of the PCR product is added to a 10 µL PURExpress reaction (mixed according to the manufacturer's recommendations) containing ~10 µCi [^{35}S]-methionine, followed by incubation at 37°C for 20 min at 600 rpm shaking. Translation is halted by the addition of 15 µL of ice-cold 10% TCA, followed by incubation on ice for at

least 30 min. Total protein is sedimented by centrifugation at 4°C for 5 min at 20,000x*g*. The supernatant is carefully removed and the pellet is resuspended in a suitable volume of 1X SDS-PAGE sample buffer (62.5 mM Tris-HCl pH 6.8, 10% glycerol, 2.5% SDS, 5% β-ME, 0.02% bromophenol blue, and 25 mM NaOH (NaOH is added to neutralize any remaining TCA) by shaking at 37°C and 1,000 rpm for at least 5 min. The prolyl-tRNA that remains attached due to SecM arrest is digested by the addition of 4 µg of RNAse I (2 µL of a 4 µg/µL solution), followed by incubation at 37°C and 600 rpm for 30 min.

Quantifications

Following a brief centrifugation to remove any remaining insoluble material, the sample is loaded onto an appropriate SDS-PAGE gel (12% tris-glycine gels were used for two- and three-spectrin domain constructs and 16 or 18% tris-tricine gels were used for single-spectrin domains). Following electrophoresis, the gels are dried onto thick filter paper by heating under vacuum (BioRad Model 583 or Hoefer GD 2000), a radioactive molecular weight ladder included in the gel is visualized by spotting the filter paper with a ~1:1,000 solution of [^{35}S]-methionine in 1X SDS-PAGE sample buffer, and the gel is exposed to a phosphorimager screen for 12 to 72 hours depending on the strength of the signal. The screen is imaged using a Fujifilm FLA9000 (50 µm pixels), and densitometry analysis on the resultant raw image (TIFF format) file is carried out using FIJI (ImageJ) software. The densitometry values are quantified using our in-house EasyQuant software and the fraction full-length protein, $f_{FL} = \frac{I_{FL}}{I_{FL}+I_A}$, is calculated from the intensities of the full-length (I_{FL}) and arrested (I_A) bands. See Supplementary Fig. S7 for examples of gels. Independent duplicate PURE reactions were analyzed for R15 nL and R14R15 nL datasets, and the f_{FL} values were found to vary by less than 0.07 between duplicate data points. For R15R16 nT and R15R16R17 nT, only single reactions were analyzed.

Pulse-chase experiments

The rate at which translation recommences following the arrest of various constructs was measured for five R16 constructs and two ADR1 constructs [7] chosen to represent the range of f_{FL} values measured during typical experimental conditions. Pulse-chase experiments were carried out using the PURExpress system. 75 µL reactions were mixed according to the manufacturer's instructions, with the addition of [^{35}S]-methionine. Following a 5 min incubation at 37°C (pulse) an excess of unlabeled methionine was added and incubation at 37°C was continued. At discrete time points, a 10 µL aliquot of the reaction was removed and mixed with 15 µL of ice-cold 10% TCA and further processed as above. Three independent replicates were collected for each construct. The rate calculated is for the conversion of arrested protein to full-length protein:

$$A \xrightarrow{k_R} FL$$

The rate of this irreversible conversion can be calculated using a 1st order equation:

$$f_A(t) = f_A(0)\, e^{-k_R t}$$

Where $f_A(\mathrm{t})$ is the fraction arrested protein at a given time, t, and $f_A(0)$ is the fraction arrested protein at $t = 0$. Prism 8 (GraphPad Software) was used to calculate the non-linear regression fit of the kinetic equation to each data set (R16 nL-27, -29, -31, -33 and -37, and ADR1 L25 and 27[7]). The rates were used to calculate the force exerted by the folding domain on the arrested nascent chain as detailed in the main text.

Optimization of the Bell parameters

The coarse-grained molecular dynamics (MD) simulations of cotranslational folding of R16 and ADR1a constructs in WT, ΔuL23, and ΔuL24 ribosomes used here were originally reported in [12]. For each construct and each linker length, these simulations were used to

calculate the ensemble average force $\langle F_{MD} \rangle = p_U F_U + p_F F_F$, where $p_{U(F)}$ and $F_{U(F)}$ are, respectively, the population of, and force exerted by, the unfolded (folded) state in the simulation. In order to determine the Bell parameters k_0 and $\Delta x^{\ddagger}$ in equation [1] in the main text, we set $k_0 = 3 \times 10^{-4}$ s^{-1} (*i.e.*, the same value as estimated in [6], and equal to k_R for the "zero force" construct R16 nL27, Fig. 4a) and explored a range of $\Delta x^{\ddagger}$ values around the approximate value $\Delta x^{\ddagger} = 0.4$ nm given in [6]. For each $\Delta x^{\ddagger}$ value, we calculated an average k_R over unfolded and folded states as $k_R = p_U k_0 \exp[F_U \Delta x^{\ddagger}/k_B T] + p_F k_0 \exp[F_F \Delta x^{\ddagger}/k_B T]$ and then obtained f_{FL} (setting $\Delta t = 550$ s) from equation [2] in the main text. The simulated f_{FL} values (*i.e*, the simulated force profiles) obtained in this way were then compared with the experimental force profiles reported in [12] for the I27 and R16 domains expressed with WT, ΔuL23, and ΔuL24 ribosomes, and for the ADR1a domain expressed with WT and ΔuL24 ribosomes. The parameter values $k_0 = 3 \times 10^{-4}$ s^{-1}, $\Delta x^{\ddagger} = 0.65$ nm were found to give the best fit between the simulated and experimental force profiles (see Supplementary Fig. S4).

Author contributions

GK, OBN, and GvH conceived the project. GK and OBN designed and performed the experiments. PT and RB performed the theoretical calculations. All authors contributed to writing the manuscript.

Competing Interests

The authors declare no competing interests.

Acknowledgements

This work was supported by grants from the Knut and Alice Wallenberg Foundation (2012.0282), the Swedish Cancer Society (15 0888), and the Swedish Research Council (621-2014-3713) to GvH, and from the Intramural Research program of the National Institute of Diabetes and Digestive and Kidney Diseases of the NIH to RBB, and from NIH (grant R35GM122543) to FPA. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov). We thank Dr. Jane Clarke and Dr. Adrian Nickson for providing plasmids with spectrin DNA and Dr. Rickard Hedman for programming and maintenance of the EasyQuant software.

Data Availability

The data that support the findings of the study are available from the corresponding author upon reasonable request.

References

1. 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).
2. Nakatogawa, H. & Ito, K. Secretion Monitor, SecM, Undergoes Self-Translation Arrest in the Cytosol. *Mol. Cell* **7**, 185–192 (2001).
3. Narayanan, C. S. & Dubnau, D. Demonstration of erythromycin-dependent stalling of ribosomes on the ermC leader transcript. *J. Biol. Chem.* **262**, 1766–1771 (1987).
4. Cymer, F., Hedman, R., Ismail, N. & Von Heijne, G. Exploration of the arrest peptide sequence space reveals arrest-enhanced variants. *J. Biol. Chem.* **290**, 10208–10215

(2015).

5. Schiller, N., Magoulopoulou, A., Cymer, F. & Heijne, G. Von. Mutational analysis of the human Xbp1 translational arrest peptide and construction of arrest-enhanced variants. *bioRxiv* 201913 (2017). doi:10.1101/201913

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

7. Nilsson, O. B. *et al.* Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep.* **12**, 1533–1540 (2015).

8. 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).

9. Nilsson, O. B. *et al.* Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. Mol. Biol.* **24**, 221–225 (2017).

10. Farías-Rico, J. A., Goetz, S. K., Marino, J. & von Heijne, G. Mutational analysis of protein folding inside the ribosome exit tunnel. *FEBS Letters* **591**, 155–163 (2017).

11. Tian, P. *et al.* Folding pathway of an Ig domain is conserved on and off the ribosome. *Proc. Natl. Acad. Sci.* **115**, 201810523 (2018).

12. Kudva, R. *et al.* The shape of the bacterial ribosome exit tunnel affects cotranslational protein folding. *Elife* **7**, 274191 (2018).

13. Farías-Rico, J. A., Ruud Selin, F., Myronidi, I., Frühauf, M. & von Heijne, G. Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proc. Natl. Acad. Sci.* **115**, E9280–E9287 (2018).

14. Notari, L., Martínez-Carranza, M., Farías-Rico, J. A., Stenmark, P. & von Heijne, G. Cotranslational Folding of a Pentarepeat β-Helix Protein. *J. Mol. Biol.* **430**, 5196–5206 (2018).

15. Kemp, G., Kudva, R., de la Rosa, A. & von Heijne, G. Force-Profile Analysis of the Cotranslational Folding of HemK and Filamin Domains: Comparison of Biochemical and Biophysical Folding Assays. *J. Mol. Biol.* 1308–1314 (2019). doi:10.1016/j.jmb.2019.01.043

16. Winkelmann, J. C. & Forget, B. G. Erythroid and nonerythroid spectrins. *Blood* **81**, 3173–85 (1993).

17. Machnicka, B., Grochowalska, R., Bogusławska, D. M., Sikorski, A. F. & Lecomte, M. C. Spectrin-based skeleton as an actor in cell signaling. *Cellular and Molecular Life Sciences* **69**, 191–201 (2012).

18. Xu, K., Zhong, G. & Zhuang, X. Actin, Spectrin, and Associated Proteins Form a Periodic Cytoskeletal.pdf. *Science* **339**, 452–6 (2013).

19. Liem, R. K. H. Cytoskeletal Integrators: The Spectrin Superfamily. *Cold Spring Harb. Perspect. Biol.* **8**, a018259 (2016).

20. 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).

21. Ipsaro, J. J., Harper, S. L., Messick, T. E., Marmorstein, R. & Mondragón, A. E. Al. Crystal structure and functional interpretation of the erythocyte spectrin tetramerization domain complex. *Blood* **115**, 4843–52 (2010).

22. An, X. *et al.* Thermal Stabilities of Brain Spectrin and the Constituent Repeats of Subunits †. *Biochemistry* **45**, 13670–13676 (2006).

23. An, X. *et al.* Conformational stabilities of the structural repeats of erythroid spectrin and their functional implications. *J. Biol. Chem.* **281**, 10527–32 (2006).

24. Kusunoki, H., Minasov, G., MacDonald, R. I. & Mondragón, A. Independent movement, dimerization and stability of tandem repeats of chicken brain α-spectrin. *J.*

Mol. Biol. **344**, 495–511 (2004).

25. Patra, M., Mukhopadhyay, C. & Chakrabarti, A. Probing Conformational Stability and Dynamics of Erythroid and Nonerythroid Spectrin: Effects of Urea and Guanidine Hydrochloride. *PLoS One* **10**, e0116991 (2015).

26. Scott, K. A., Batey, S., Hooton, K. A. & Clarke, J. The folding of spectrin domains I: Wild-type domains have the same stability but very different kinetic properties. *J. Mol. Biol.* **344**, 195–205 (2004).

27. MacDonald, R. I. & Cummings, J. A. Stabilities of folding of clustered, two-repeat fragments of spectrin reveal a potential hinge in the human erythroid spectrin tetramer. *Proc. Natl. Acad. Sci.* **101**, 1502–1507 (2004).

28. Kwa, L. G. *et al.* The folding of a family of three-helix bundle proteins: Spectrin R15 has a robust folding nucleus, unlike its homologous neighbours. *J. Mol. Biol.* **426**, 1600–1610 (2014).

29. Rief, M., Pascual, J., Saraste, M. & Gaub, H. E. Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles. *J. Mol. Biol.* **286**, 553–561 (1999).

30. Lenne, P. F., Raae, A. J., Altmann, S. M., Saraste, M. & Hörber, J. K. H. States and transitions during forced unfolding of a single spectrin repeat. *FEBS Lett.* **476**, 124–128 (2000).

31. MacDonald, R. I. & Pozharski, E. V. Free energies of urea and of thermal unfolding show that two tandem repeats of spectrin are thermodynamically more stable than a single repeat. *Biochemistry* **40**, 3974–3984 (2001).

32. Law, R. *et al.* Cooperativity in forced unfolding of tandem spectrin repeats. *Biophys. J.* **84**, 533–544 (2003).

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. Wensley, B. G. *et al.* Separating the effects of internal friction and transition state energy to explain the slow, frustrated folding of spectrin domains. *Proc. Natl. Acad. Sci.* **109**, 17795–17799 (2012).

35. Ito, K. & Chiba, S. Arrest Peptides: Cis -Acting Modulators of Translation. *Annu. Rev. Biochem.* **82**, 171–202 (2013).

36. 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.* (2003). doi:10.1128/JB.185.22.6719-6722.2003

37. Ismail, N., Hedman, R., Schiller, N. & Von Heijne, G. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat. Struct. Mol. Biol.* **19**, 1018–1023 (2012).

38. 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**, 145–9 (2015).

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

40. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).

41. Bell, G. Models for the specific adhesion of cells to cells. *Science* **200**, 618–627 (1978).

42. 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).

43. Grum, V. L., Li, D., MacDonald, R. I. & Mondragón, A. Structures of two repeats of

spectrin suggest models of flexibility. *Cell* **98**, 523–535 (1999).

44. Kusunoki, H., MacDonald, R. I. & Mondragón, A. Structural insights into the stability and flexibility of unusual erythroid spectrin repeats. *Structure* **12**, 645–656 (2004).

45. 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).

46. Leininger, S. E., Trovato, F., Nissley, D. A. & O'Brien, E. P. Domain topology, stability, and translation speed determine mechanical force generation on the ribosome. *Proc. Natl. Acad. Sci.* **116**, 5523–5532 (2019).

47. Fritch, B. *et al.* Origins of the Mechanochemical Coupling of Peptide Bond Formation to Protein Synthesis. *J. Am. Chem. Soc.* **140**, 5077–5087 (2018).

48. Nakatogawa, H. & Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629–636 (2002).

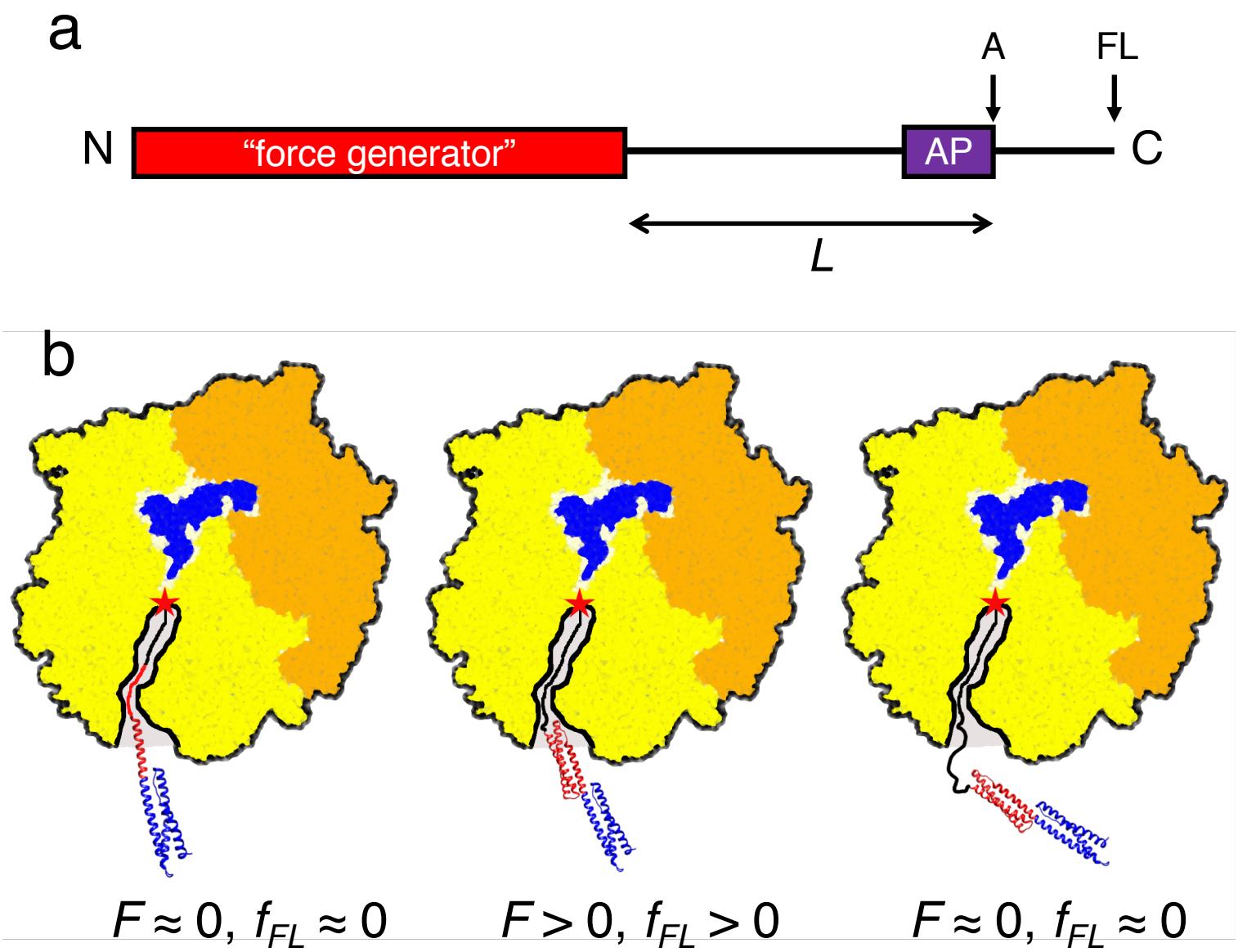
Figure legends

Figure 1. Force profile analysis. (a) Schematic representation of a typical construct. A "force-generator," which here is one or more spectrin domains, is connected by a linker sequence of length *L* to the *E. coli* SecM arrest peptide (AP). Following the AP is an unrelated protein sequence from the *E. coli* LepB protein that is added to allow the arrested (A) and full-length (FL) forms to be separable by SDS-PAGE. (b) Schematic representation of cotranslationally arrested ribosome-nascent chain complexes with increasing linker lengths (left to right). Represented here is the R15R16 nL construct (colored as in Fig. 2a). Left: at short linker lengths, when the folding of the R16 domain (blue) has not yet begun, the force acting on the nascent chain is minimal, and the arrest is efficient ($F \approx 0, f_{FL} \approx 0$). Middle: at linker lengths that allow the folding of the R16 domain, the force acting on the nascent chain is increased and the arrest is less efficient ($F > 0, f_{FL} > 0$). Right: at linker lengths that arrest the ribosome after the folding of the R16 domain has already occurred, the force acting on the nascent chain is again minimal ($F \approx 0, f_{FL} \approx 0$). (c) Determination of f_{FL} by SDS-PAGE. Lane 1 is the R15 nL39 construct and lane 2 is the R15 nL39 full-length control construct. The control construct has the final (critical) proline residue of the AP mutated to alanine, preventing arrest and producing only full-length protein. (d) A plot showing the calculation of L_{onset}, L_{max}, and L_{end}. The data points used for the calculation are denoted by black, yellow, and grey circles, respectively. The interpolated midpoint between each pair of selected point gives the L_{onset} (vertical dotted line), L_{max} (dashed line), and L_{end} (dash-dot line) values.

Figure 2. Construct design. (a) Structure of the R15 (blue), R16 (red), and R17 (orange) domains of chicken brain α-spectrin (PDB:1U4Q)[24]. (b) Constructs analyzed in this study. A solid black line represents sequence derived from the LepB protein and dashed boxes represent truncated spectrin sequences. The ends of the arrested (A) and full-length (FL) forms are indicated.

Figure 3. Cotranslational folding of spectrin domains. (a) Force profiles for the R15 nL and R16 nL constructs. f_{FL} is the fraction full-length protein produced and L is linker length (*i.e.*, the number of amino acid residues between the C-terminal end of the spectrin domain and the C-terminal Pro residue in the AP). Data for R16 was reproduced from [9]. (b) Force profiles showing the influence of R14 and R16 on the folding of R15. (c) Force profiles showing the influence of R15 and R17 on the folding of R16. Data for the R15R16 nL constructs was reproduced from [9]. (d) Summary of the force profile data. L_{onset}, L_{max}, and L_{end} values were calculated as shown in see Supplementary Fig. S1. Construct names are colored the same as the curves in panels a-c.

Figure 4. Release rates and estimation of pulling forces. (a) The rate of SecM release (k_R) obtained from pulse-chase experiments (see Supplementary Fig. S2), the fraction full-length protein (f_{FL}) measured under standard experimental conditions (20 min. incubation in PURExpress in the continuous presence of [^{35}S] Met), and the pulling force F calculated using equation [1] ($k_0 = 3.0 \times 10^{-4}$ s^{-1}, $\Delta x^{\ddagger} = 0.65$ nm). The constructs are from Kudva *et al.* [12], and are colored to match those in panel *b* and Supplementary Fig. S2 and S3. (b) F values calculated from equation [1] plotted against the standard f_{FL} values, with constructs colored as in panel *a*. The least-squares fit line is indicated by the blue line, and the analytic relation [3] between F and f_{FL}, assuming an average delay time Δt = 550 s (approximately equal to half the standard incubation time), is shown as a red curve.


a
A
FL
N
"force generator"
AP
C
L
b
F ≈ 0, f_FL ≈ 0
F > 0, f_FL > 0
F ≈ 0, f_FL ≈ 0

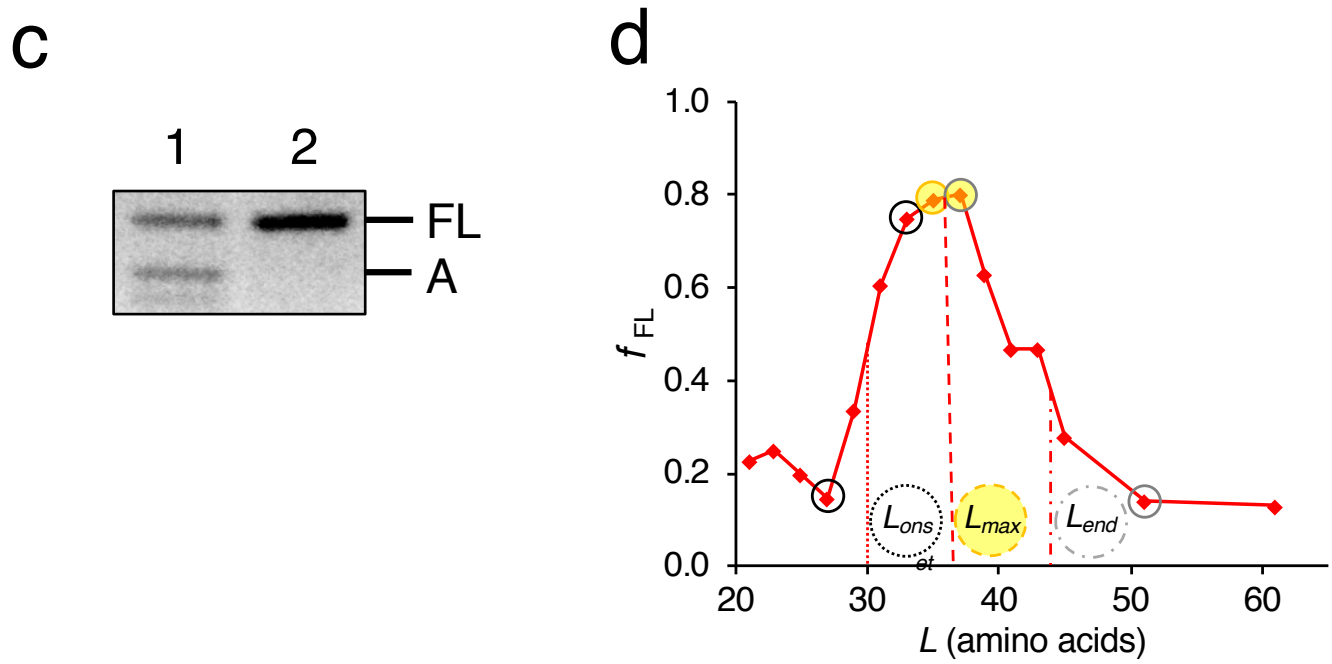

c
1
2
FL
A
d
1.0
0.8
0.6
0.4
0.2
0.0
f_FL
L_ons
L_max
L_end
20
30
40
50
60
L (amino acids)

Figure 1

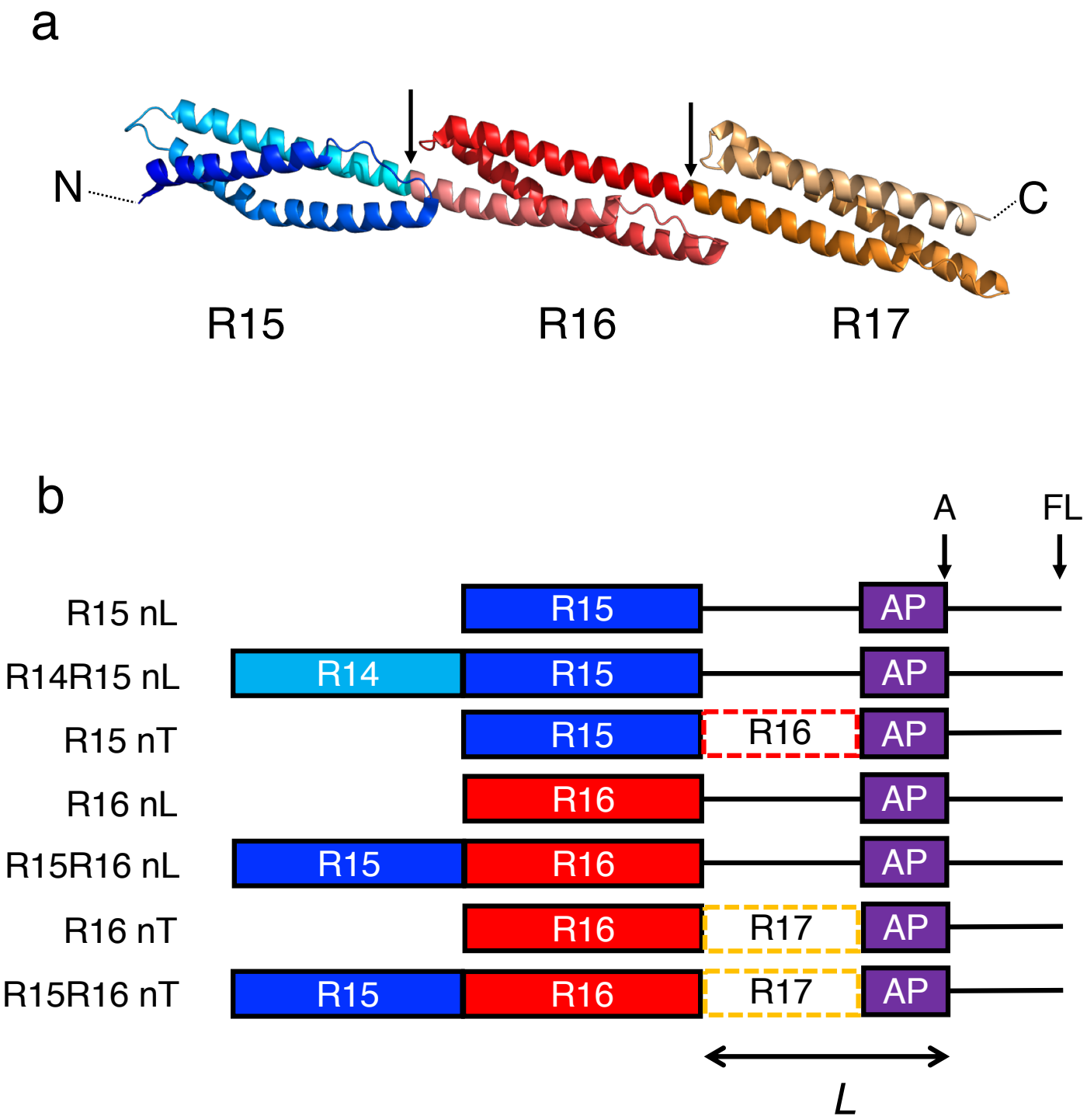
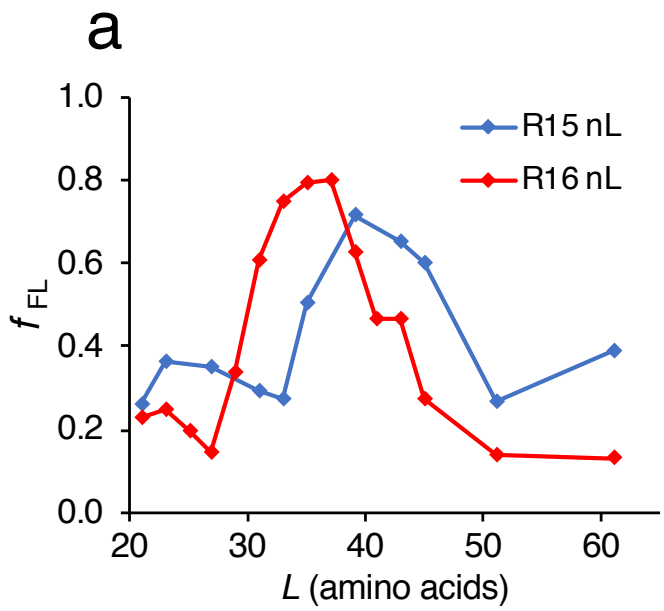
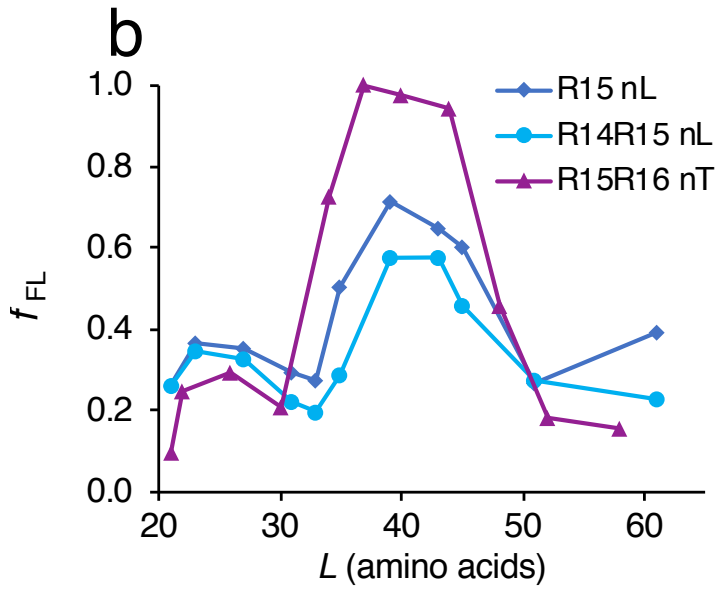
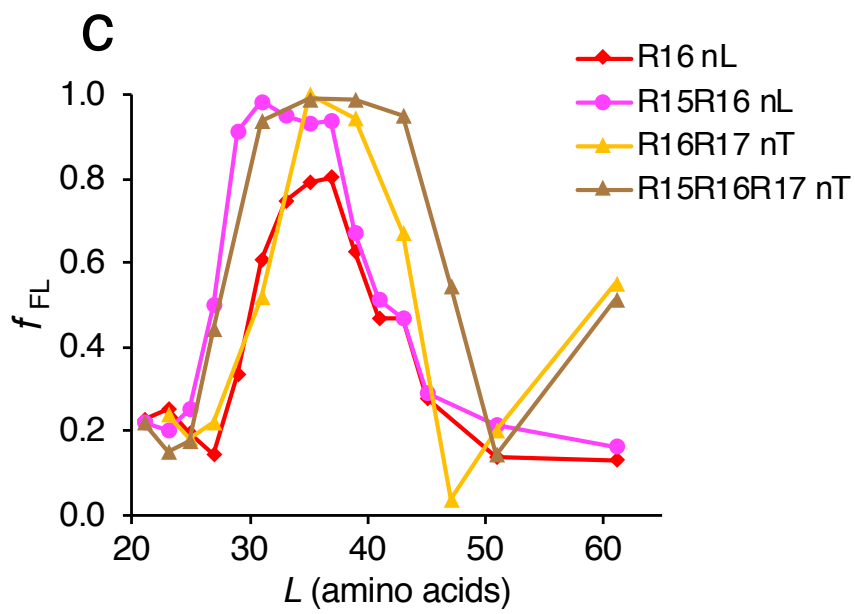

a
N
C
R15
R16
R17
b
A
FL
R15 nL
R14R15 nL
R15 nT
R16 nL
R15R16 nL
R16 nT
R15R16 nT
R14
R15
R16
R17
AP
L

Figure 2







d

	L_{onset}	L_{max}	L_{end}	
R15 nL	36	39	47	R15 folding
R14R15 nL	36	41	47	
R15R16 nT	33	40	48	
R16 nL	30	36	44	R16 folding
R15R16 nL	27	33	41	
R16R17 nT	31	35	43	
R15R16R17 nT	28	37	47	

Figure 3

a

	k (s^{-1})	f_{FL}	F (pN)
R16 nL27	0.00030355	0.19	0.1
R16 nL29	0.00079915	0.28	6.5
R16 nL31	0.0014105	0.57	10.2
ADR1 L27	0.0016815	0.63	11.3
R16 nL33	0.0021765	0.70	13.0
R16 nL37	0.003334	0.82	15.9
ADR1 L25	0.0042715	0.90	17.5

b

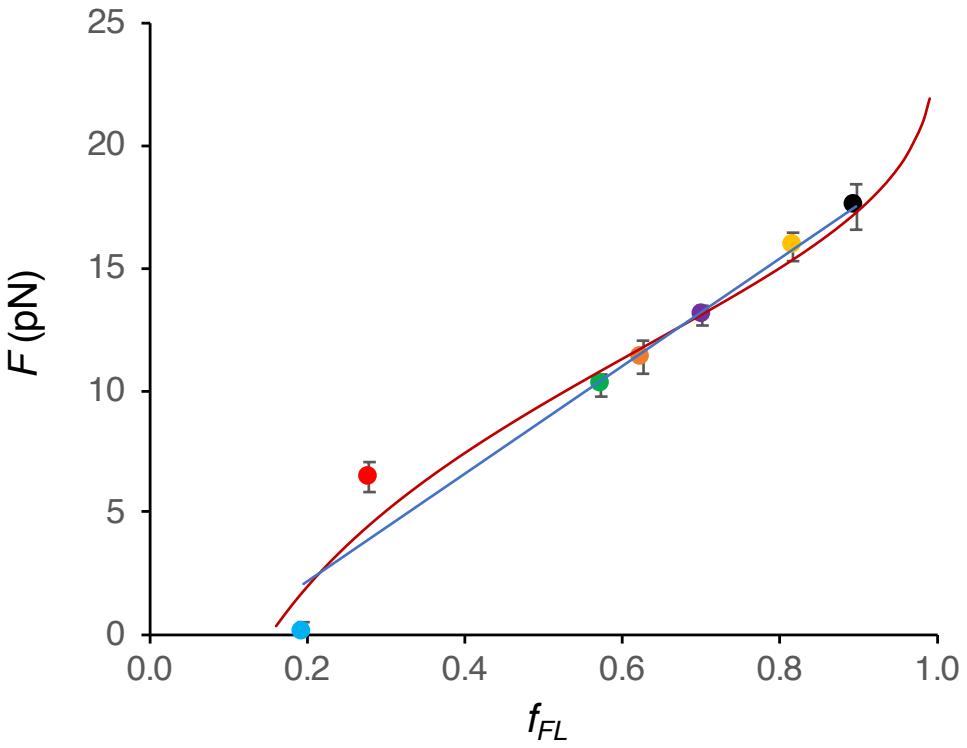


Figure 4