Version December 16, 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

Classification: Biological Sciences, Biochemistry/Biophysics and Computational Biology

Abstract

Proteins synthesized in the cell can begin to fold during translation before the entire polypeptide has been produced, which may be particularly relevant to the folding of multidomain proteins. Here, we study the cotranslational folding of adjacent domains from the cytoskeletal protein α-spectrin using Force Profile Analysis (FPA). Specifically, we investigate how the cotranslational folding behavior of the R15 and R16 domains are affected by their neighboring R14 and R16, and R15 and R17 domains, respectively. 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, and may reduce its dependence on chaperones. Furthermore, we directly relate the experimentally observed yield of full-length protein in the FPA assay to the force exerted by the folding protein in pN. By combining pulse-chase experiments to measure the rate at which the arrested protein is converted into full-length protein with a Bell model of force-induced rupture, we estimate that the R16 domain exerts a force on the nascent chain of ~15 pN during cotranslational folding.

Significance

In living cells proteins are produced in a sequential way by ribosomes. This vectoral process allows the growing protein chain to start to fold before translation has been completed. Thereby, cotranslational protein folding can be significantly different than the folding of a full-length protein in isolation. Here we show how structurally similar repeat domains, normally produced as parts of a single long polypeptide, affect the cotranslational folding of their neighbors. This provides insight into how the cell may efficiently produce multidomain proteins, paving the way for future studies *in vivo* or with chaperones. We also provide an estimated magnitude of the mechanical force on the nascent chain incurred by cotranslational folding calculated from biochemical measurements and molecular dynamics simulations.

\body

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 affect folding pathways. 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, establishing 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 were also able to observe that the synthesis of native, likely helical, structures downstream of the folding spectrin domain increases the observed folding force further stabilizing the fold even in the presence of amino acid substitutions that destabilize the final native spectrin structure.

While it has been shown previously that the cotranslational folding of proteins can provide a mechanical force to release SecM-mediated ribosomal stalling (6), the magnitude of the force has never been directly measured by FPA. Using pulse-chase experiments, we have now measured the rate at which arrested protein is converted to full-length protein under standard experimental conditions. These rates were fit to a Bell model (35) for force-induced rupture using parameters elucidated by molecular dynamics simulations (12) to provide an estimate of the magnitude of the pulling force exerted on the nascent chain by cotranslational protein

folding. The determined force is in good agreement we previously published measurements using optical tweezers (6).

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 (36). APs generally have regulatory functions in cells, controlling, *e.g.*, translation initiation on downstream ORFs in polycistronic mRNAs (36). 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, 37, 38). 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, 38, 39).

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 peak amplitude and 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, 40). 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 SecM(*Ec*) 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 PURExfrex *in vitro* translation system (41) 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 S1.

In order to distinguish which spectrin repeat is being assayed in the tandem constructs, we use the following naming convention. Constructs named "nL" (for **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 **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 R16 (9) and newly collected data for R15 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; note that L_{max} values for force profiles that reach saturation ($f_{FL} \approx 1$) cannot be accurately determined. 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, 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 (R15R16R17 nT and R16R17 nT) 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 Table S1). This increase in f_{FL} may thus signal an early interaction between helices A and B in R17.

In order to more precisely define $L_{\max}$ for constructs were f_{FL} approaches 1, we substituted two amino acids in the relatively weak SecM(*Ec*) AP to create the stronger SecM(sup1) variant (42) in R16 nL, R15R16 nL, R16R17 nT, and R15R16R17 nT construtcs (Supplementary Figs. S2 and S3). The more precise L_{max} values confirmed that the presence of R15 shifts the folding of R16 to shorter tether lengths.

As a control, we introduced mutations that are known to inhibit folding of R15 and R16 *in vitro* (43, 44) into the R15 nL, R15R16 nT, R16 nL, R16R17 nT, and R15R16R17 nT constructs, Fig. 4 and Supplementary Fig. S4. As expected, f_{FL} values for the "non-folding" R15(nf) nL and R16(nf) nL constructs were reduced to background levels, but, surprisingly, f_{FL} values remained high for linker lengths between L_{max} and L_{end} for R15(nf)R16 nT, R16(nf)R17 nT, and R15R16(nf)R17 nT constructs. Since the difference between the nL and nT series of constructs is the presence of a C-terminal "linker" derived from the N-terminal part of R16 or R17 in the latter (replacing unstructured segments from LepB), this suggested that the linker itself, perhaps together with C-terminal parts of the upstream spectrin domain, might fold in the nf-mutants.

To investigate this possibility, we started from the R15(nf)R16 nT, L = 40 construct (which has a much higher f_{FL} value than the R15(nf) nL version, Fig. 4a) and successively removed helices A, B, and C from the R15(nf) part, Fig. 4b. We found that removing helix R15-A or both helices R15-A and R15-B, decreased f_{FL}, but only to 0.47 and 0.36, respectively, Fig. 4c. Only with the removal of all of R15, leaving only 19 amino acids of R16 helix A, did f_{FL}

decrease to baseline. Thus, the cotranslational formation of a continuous helical structure encompassing parts of the R15-C and R16-A helices appears to cause the increase in f_{FL} at L = 40; indeed, it is known from previous work that helices can form cotranslationally in the exit tunnel (45). The introduction of helix-breaking residues into the middle of the R16-A segment (see Table S1) in the R16-A/R15-BC and R16-A/R15-ABC constructs decreased the f_{FL} to the same level as R16-A, Fig. 4c (open circles). Similar behavior was also observed for the R16R17 nT constructs, Supplementary Fig. S4, implying that the R17-A helix likewise forms a cotranslational folding intermediate together with the R16-C helix.

We conclude that the cotranslational folding of the R15 domain is affected by its downstream but not by its upstream neighbor domain, and that folding of 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. We attribute the contribution of the downstream neighbor to the formation of helical structure encompassing helix C from the upstream domain and helix A from the downstream domain.

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, 39). 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 (35) 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. S5.

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. S6.

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. 5a, 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 x 10^{-4} s^{-1} (*i.e.*, the same value as in (6) and equal to k_R for R16 nL27, Fig. 5a) and $x^{\ddagger}$ = 0.65 nm (well inside the 95% CI from (6)), Supplementary Fig. S7. With these parameter values and Δt = 550 s, equation [3] nicely captures the relation between the force estimated from the measured k_R and standard f_{FL} values, Fig. 5b, 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 SecM(*Ec*) AP used here, and not for other APs of different stalling strengths (39). 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 (38, 39), 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. S8 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 F estimated from pulse-chase measurements or f_{FL} values represents the constant force which would have the same effect on the escape rate from arrest as the combined effect of the forces from both folded and unfolded states. We can compare this inferred force with 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. S9a, 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: the two forces are not equal 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 and hence to F (a direct comparison between $\langle F_{MD} \rangle$ from the MD simulations and F calculated from [3] is given in Supplementary Fig. S9b).

We note that the forces exerted by a protein folding as it exits the ribosome are qualitatively different from the tensile forces such as those exerted when proteins fold in AFM experiments (as have been performed on spectrin before (46)). Therefore, the force magnitudes probed by the two experiments cannot be directly compared, despite their apparent resemblance. For example, even a small force exerted on the termini of an unfolded protein by an AFM can massively slow the refolding rate because of the large distance the protein must contract against this force in order to fold. The same force magnitude would have a much smaller effect on the refolding rate of a protein attached at one end to the ribosome (47).

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 (48), 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).

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 (43). 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 is stabilized by the presence of the early parts of the R16 A helix, 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 an early folding intermediate in R17. Further work is required to investigate this, but it may be possible that the regular helical structure of the spectrin repeats drives a nearly-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, which could minimize the need for protection of the nascent chain by chaperones.

Finally, a pulse-chase analysis has 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. 5a. 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 protein domains such as spectrin R16 and ADR1a can generate a maximal force of 15-20 pN on the AP, Fig. 5, in line with theoretical estimates based on MD simulations (49, 50).

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 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. All protein sequences used in this publication are presented in Supplementary Table S1. 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. The shortest constructs (nL/nT 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. For some constructs, 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 (51) and yielding only full-length protein. The SecM sup1 variants were created by site-directed mutagenesis of the constructs described above. N-terminal truncations of R15 and R16 were carried out by partially overlapping inverse PCR and insertion of the helix breaking insertions was carried out by site-directed mutagenesis where the inserted residues

were included in the primer sequence. 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. Experiments with the sup1 arrest peptide variant (Supplementary Fig. S2) and the N-terminally truncated and non-folding constructs (Fig. 4 and Supplementary Fig. S4) were carried out using PURExpress following the same protocol above (10 µL reaction mixed according to the manufacturer's instructions).

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 and the N-terminal truncations). 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. S10 for examples of gels. Independent duplicate *in vitro* translation reactions were analyzed for the entire R15 nL and R14R15 nL datasets. As the f_{FL} values were all found to vary by less than 0.07 between duplicate data points (in agreement with our previously published FPA results for spectrin folding) only single reactions were analyzed for R15R16 nT, R16R17nT, R15R16R17 nT, all non-folding variants, the N-terminal truncations, nT insertions. Duplicate data points, which were also measured to vary less than 0.07, were collected for the sup1 SecM variants. This confirms the robustness of the reaction as these profiles indicate the same R16 folding behavior as the WT SecM. The mirrored increase in force observed in all the R15 and R16 non-folding nT constructs further demonstrates this robustness.

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. 5a) 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 Δt = 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. S7).

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 EP, Christodoulou J, Vendruscolo M, Dobson CM New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc* 133(3):513–526. (2011).
2. Nakatogawa H, Ito K Secretion monitor, secM, undergoes self-translation arrest in the cytosol. *Mol Cell* 7(1):185–192. (2001).
3. Narayanan CS, Dubnau D Demonstration of erythromycin-dependent stalling of ribosomes on the ermC leader transcript. *J Biol Chem* 262(4):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(16):10208–10215.

(2015).

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

6. Goldman DH, et al. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science* 348(6233):457–460. (2015).

7. Nilsson OB, et al. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep* 12(10):1533–1540. (2015).

8. Nilsson OB, 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(6):1356–1364. (2016).

9. Nilsson OB, et al. Cotranslational folding of spectrin domains via partially structured states. *Nat Struct Mol Biol* 24(3):221–225. (2017).

10. Farías-Rico JA, Goetz SK, Marino J, von Heijne G Mutational analysis of protein folding inside the ribosome exit tunnel. *FEBS Lett* 591(1):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(48):E11284--E11293. (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 JA, 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(40):E9280–E9287. (2018).

14. Notari L, Martínez-Carranza M, Farías-Rico JA, Stenmark P, von Heijne G Cotranslational Folding of a Pentarepeat β-Helix Protein. *J Mol Biol* 430(24):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).

16. Winkelmann JC, Forget BG Erythroid and nonerythroid spectrins. *Blood* 81(12):3173–85. (1993).

17. Machnicka B, Grochowalska R, Bogusławska DM, Sikorski AF, Lecomte MC Spectrin-based skeleton as an actor in cell signaling. *Cell Mol Life Sci* 69(2):191–201. (2012).

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

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

20. Hill SA, Kwa LG, Shammas SL, Lee JC, Clarke J Mechanism of assembly of the non-covalent spectrin tetramerization domain from intrinsically disordered partners. *J Mol Biol* 426(1):21–35. (2014).

21. Ipsaro JJ, Harper SL, Messick TE, Marmorstein R, Mondragón AE Al Crystal structure and functional interpretation of the erythocyte spectrin tetramerization domain complex. *Blood* 115(23):4843–52. (2010).

22. An X, et al. Thermal Stabilities of Brain Spectrin and the Constituent Repeats of Subunits †. *Biochemistry* 45(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(15):10527–32. (2006).

24. Kusunoki H, Minasov G, MacDonald RI, Mondragón A Independent movement, dimerization and stability of tandem repeats of chicken brain α-spectrin. *J Mol Biol* 344(2):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(1):e0116991. (2015).

26. Scott KA, Batey S, Hooton KA, Clarke J The folding of spectrin domains I: Wild-type domains have the same stability but very different kinetic properties. *J Mol Biol* 344(1):195–205. (2004).

27. MacDonald RI, Cummings JA 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(6):1502–1507. (2004).

28. Kwa LG, 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(7):1600–1610. (2014).

29. Rief M, Pascual J, Saraste M, Gaub HE Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles. *J Mol Biol* 286(2):553–561. (1999).

30. Lenne PF, Raae AJ, Altmann SM, Saraste M, Hörber JKH States and transitions during forced unfolding of a single spectrin repeat. *FEBS Lett* 476(3):124–128. (2000).

31. MacDonald RI, 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(13):3974–3984. (2001).

32. Law R, et al. Cooperativity in forced unfolding of tandem spectrin repeats. *Biophys J* 84(1):533–544. (2003).

33. Wensley BG, et al. Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* 463(7281):685–688. (2010).

34. Wensley BG, 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(44):17795–17799. (2012).

35. Bell G Models for the specific adhesion of cells to cells. *Science* 200(4342):618–627. (1978).

36. Ito K, Chiba S Arrest Peptides: Cis -Acting Modulators of Translation. *Annu Rev Biochem* 82(1):171–202. (2013).

37. Butkus ME, Prundeanu LB, Oliver DB Translocon "Pulling" of Nascent SecM Controls the Duration of Its Translational Pause and Secretion-Responsive secA Regulation. *J Bacteriol* 185(22):6719–6722. (2003).

38. 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(10):1018–1023. (2012).

39. 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(2):145–149. (2015).

40. Batey S, Randles LG, Steward A, Clarke J Cooperative folding in a multi-domain protein. *J Mol Biol* 349(5):1045–1059. (2005).

41. Shimizu Y, et al. Cell-free translation reconstituted with purified components. *Nat Biotechnol* 19(8):751–755. (2001).

42. Yap MN, Bernstein HD The Plasticity of a Translation Arrest Motif Yields Insights into Nascent Polypeptide Recognition inside the Ribosome Tunnel. *Mol Cell* 34(2):201–211. (2009).

43. Scott KA, Randles LG, Clarke J The Folding of Spectrin Domains II: Phi-value Analysis of R16. *J Mol Biol* 344(1):207–221. (2004).

44. Wensley BG, Gärtner M, Choo WX, 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(5):1074–1085. (2009).

45. Tu LW, Deutsch C A Folding Zone in the Ribosomal Exit Tunnel for Kv1.3 Helix Formation. *J Mol Biol* 396(5):1346–1360. (2010).

46. Randles LG, Rounsevell RWS, Clarke J Spectrin domains lose cooperativity in forced unfolding. *Biophys J* 92(2):571–577. (2007).

47. Tian P, et al. Folding pathway of an Ig domain is conserved on and off the ribosome. *Proc Natl Acad Sci U S A* 115(48):E11284--E11293. (2018).

48. 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(2):297–301. (2008).

49. Leininger SE, Trovato F, Nissley DA, O'Brien EP Domain topology, stability, and translation speed determine mechanical force generation on the ribosome. *Proc Natl Acad Sci* 116(12):5523–5532. (2019).

50. Fritch B, et al. Origins of the Mechanochemical Coupling of Peptide Bond Formation to Protein Synthesis. *J Am Chem Soc* 140(15):5077–5087. (2018).

51. Nakatogawa H, Ito K The ribosomal exit tunnel functions as a discriminating gate. *Cell* 108(5):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 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 (red) 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). Arrows indicate the domain boundaries. (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 Supplementary Fig. S1. Construct names are colored the same as the curves in panels a-c.

Figure 4. Linker effects in R15 folding. (a) Force profiles for R15 nL (blue) and R15R16 nT (purple) data from Fig. 3b compared with profiles for a R15 non-folding (nf) mutant (F18D and I55D) in R15(nf) nL constructs (green, dashed line) and R15(nf)R16 nT constructs (black, dashed line). (b) R15R16 nT22 spectrin structure (L = 40) with the A-C helices indicated and the destabilizing mutations F18D and I55D in the R15-A and -C helices shown as yellow spheres. (c) Graph showing f_{FL} values at L = 40 for constructs with increasingly long R15 parts fused to the N terminus of the 19-residue R16 nT linker (black circles). The L = 40 points from panel *a* (same coloring) as well as a structural representation of each construct are shown for reference (spectrin helices are colored as in panel *b* and LepB linker sequences colored green). f_{FL} values for constructs with helix-breaking insertions in the R16-A helix (Supplementary Table S1) are shown as open circles (Gly-Gly), open dotted circles (Pro), and open dashed circles (Gly-Ser-Gly-Ser).

Figure 5. Release rates and estimation of pulling forces. (a) The rate of SecM release (k_R) obtained from pulse-chase experiments (see Supplementary Fig. S5), 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 Figs. S5 and S6. (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.

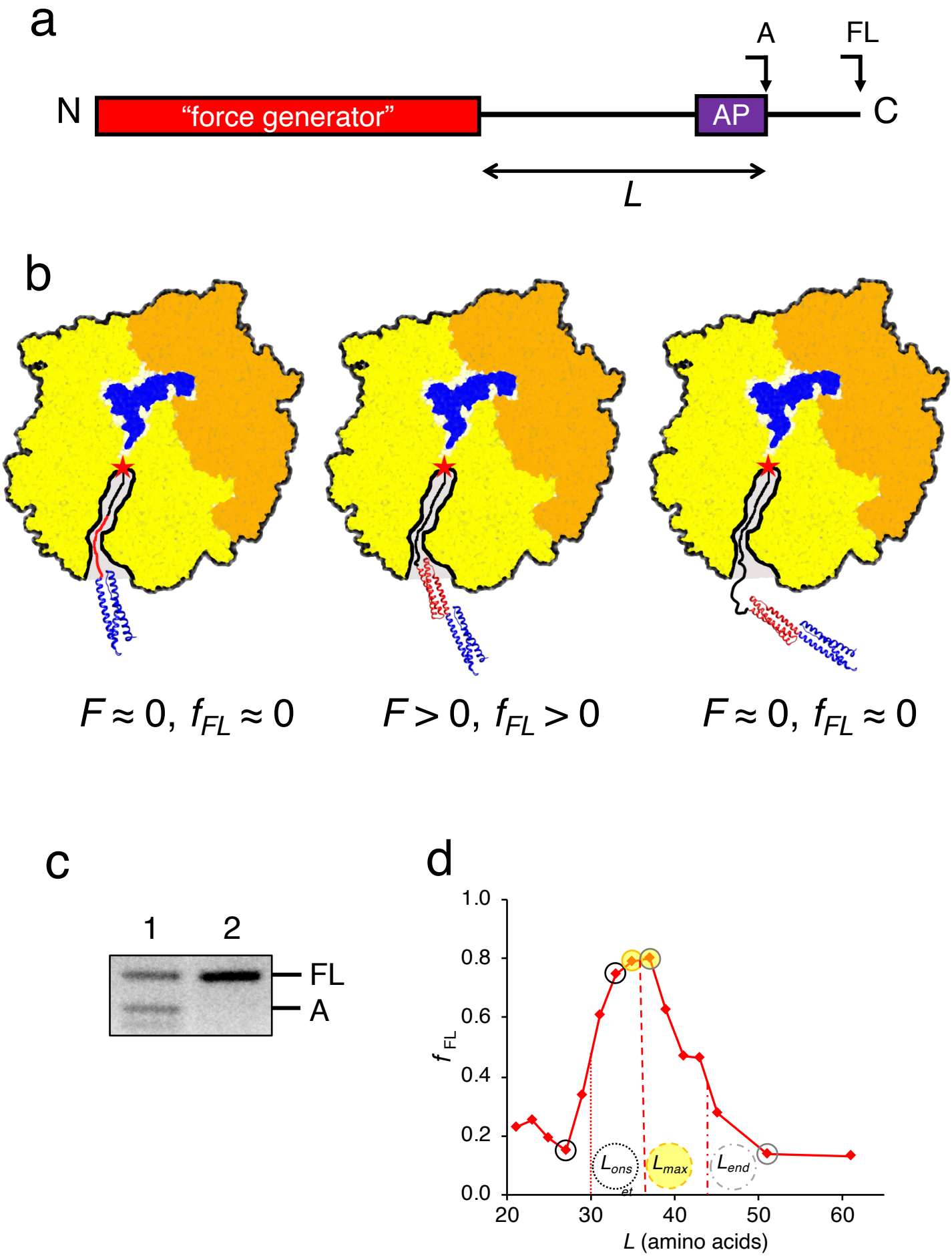

a
A
FL
N
"force generator"
AP
C
L
b
F ≈ 0, f_FL ≈ 0
F > 0, f_FL > 0
F ≈ 0, f_FL ≈ 0
c
1
2
FL
A
d
1.0
0.8
0.6
0.4
0.2
0.0
f_FL
L_ons
L_max
L_end
20
30
40
50
60
L (amino acids)

Figure 1

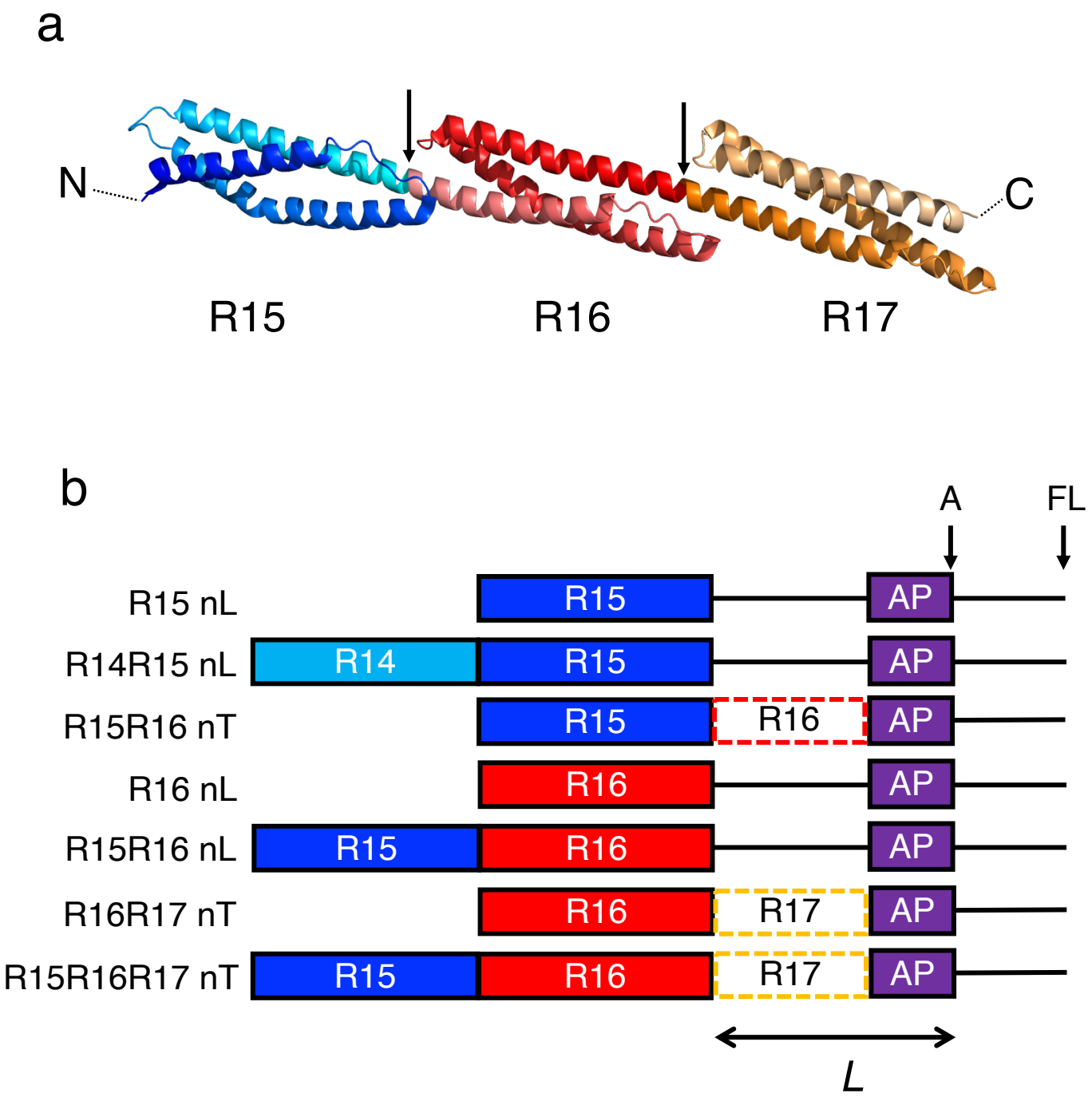
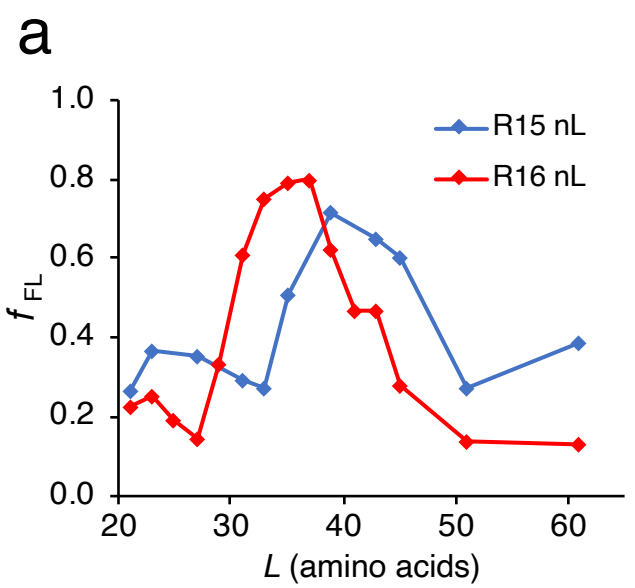
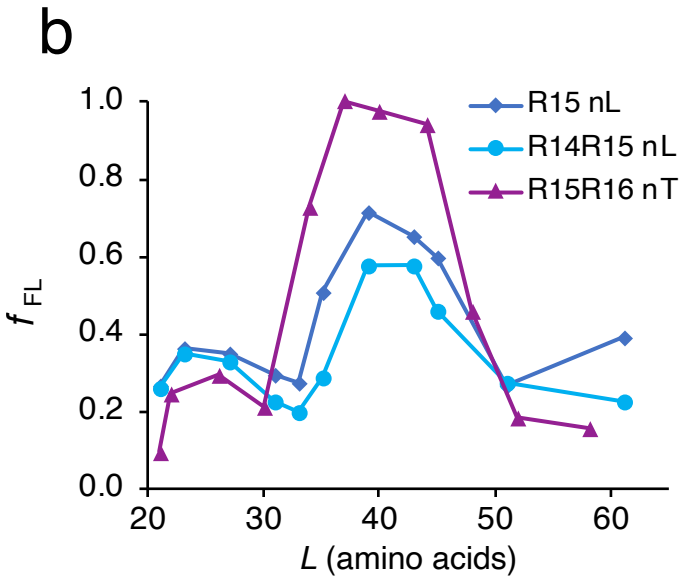
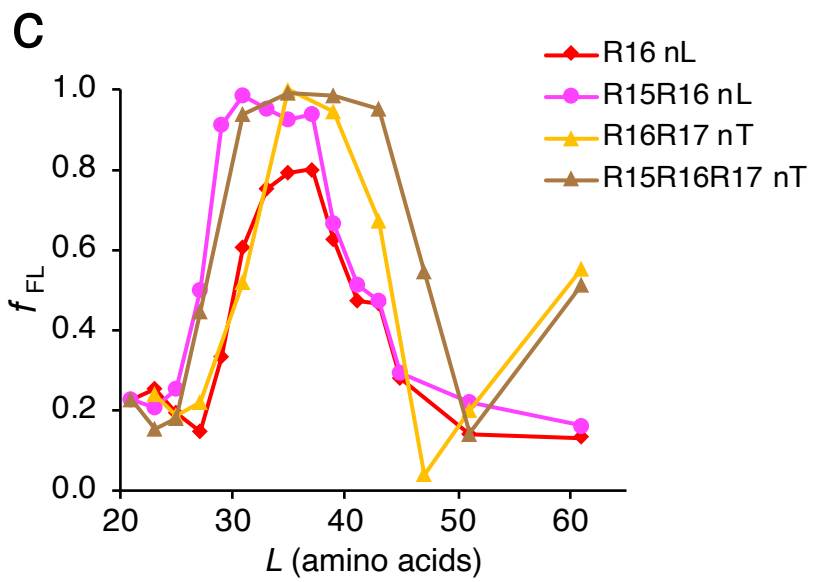

a
N
C
R15
R16
R17
b
A
FL
R15 nL
R14R15 nL
R15R16 nT
R16 nL
R15R16 nL
R16R17 nT
R15R16R17 nT
R14
R15
R16
R17
AP
L

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

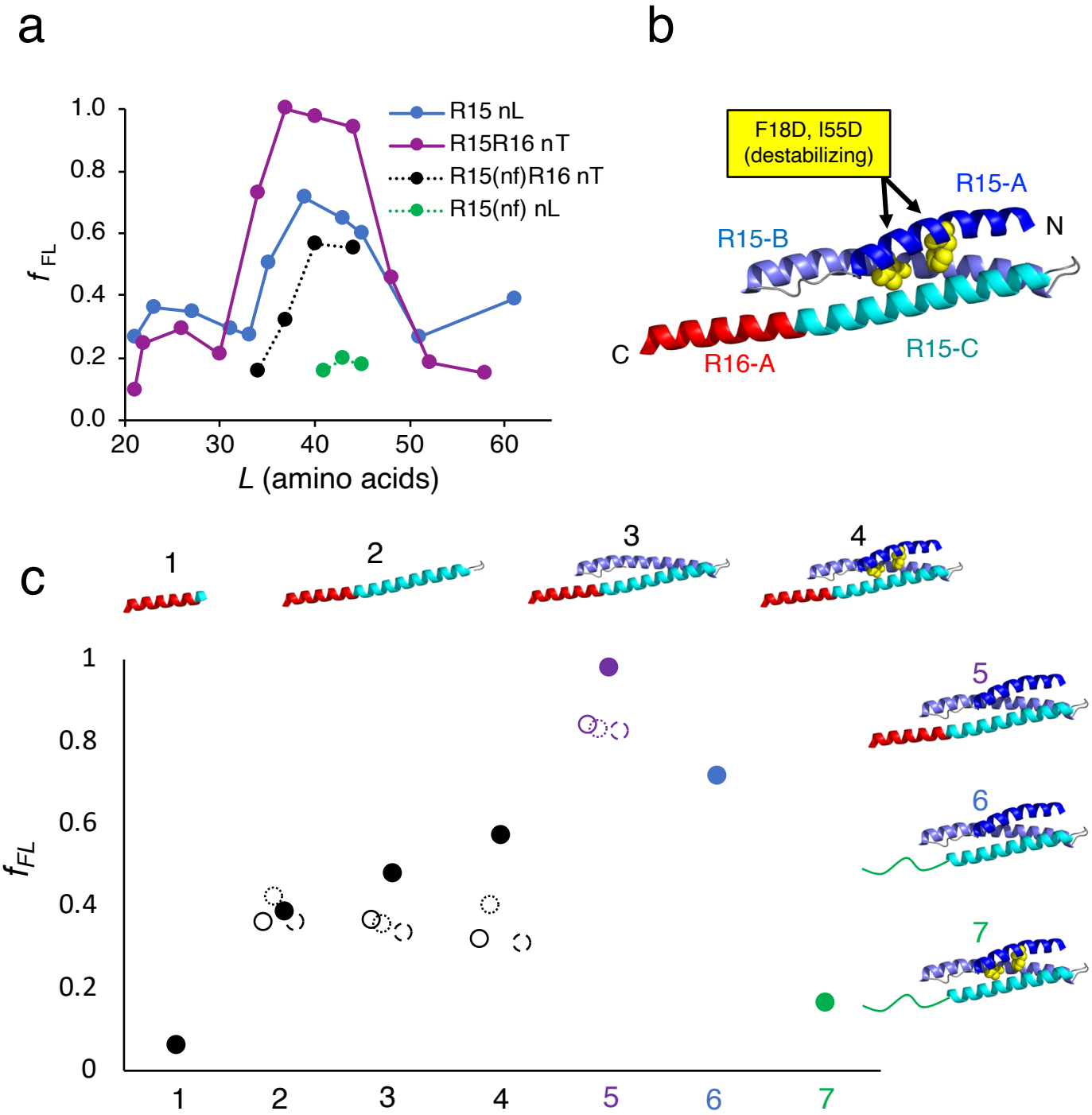

a
1.0
0.8
0.6
0.4
0.2
0.0
f_{FL}
20
30
40
50
60
L (amino acids)
R15 nL
R15R16 nT
R15(nf)R16 nT
R15(nf) nL
b
F18D, I55D
(destabilizing)
R15-A
N
R15-B
C
R16-A
R15-C
c
1
2
3
4
5
6
7
1
0.8
0.6
0.4
0.2
0
f_{FL}
1
2
3
4
5
6
7

Figure 4

a

	k (s^{-1})	f_{FL}	F (pN)
R16 nL27	0.000304	0.19	0.1
R16 nL29	0.000799	0.28	6.5
R16 nL31	0.00141	0.57	10.2
ADR1 L27	0.00168	0.63	11.3
R16 nL33	0.00218	0.70	13.0
R16 nL37	0.00333	0.82	15.9
ADR1 L25	0.00427	0.90	17.5

b

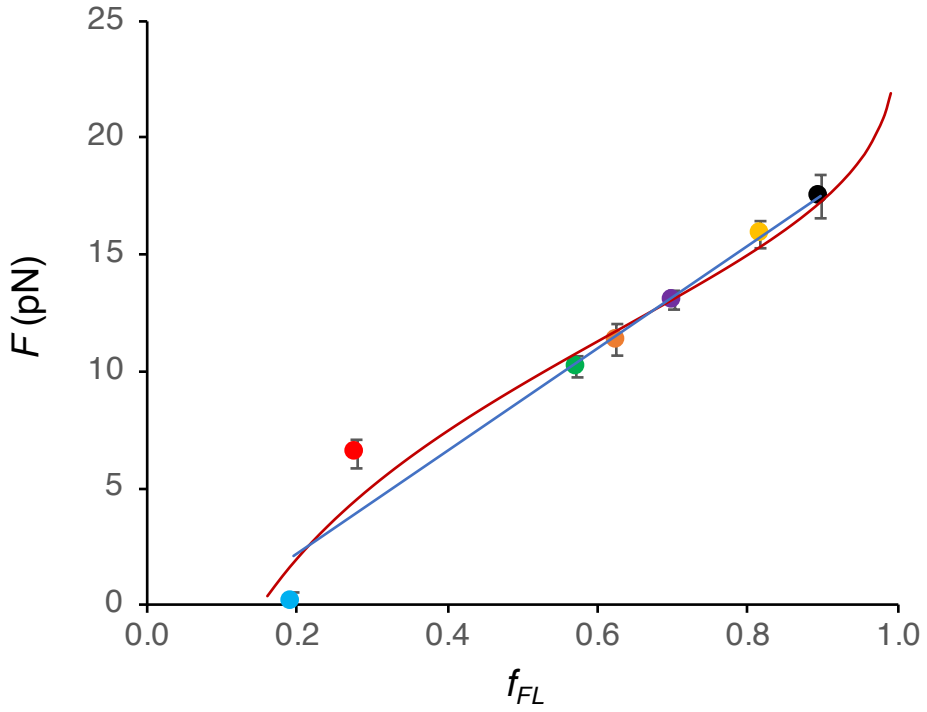
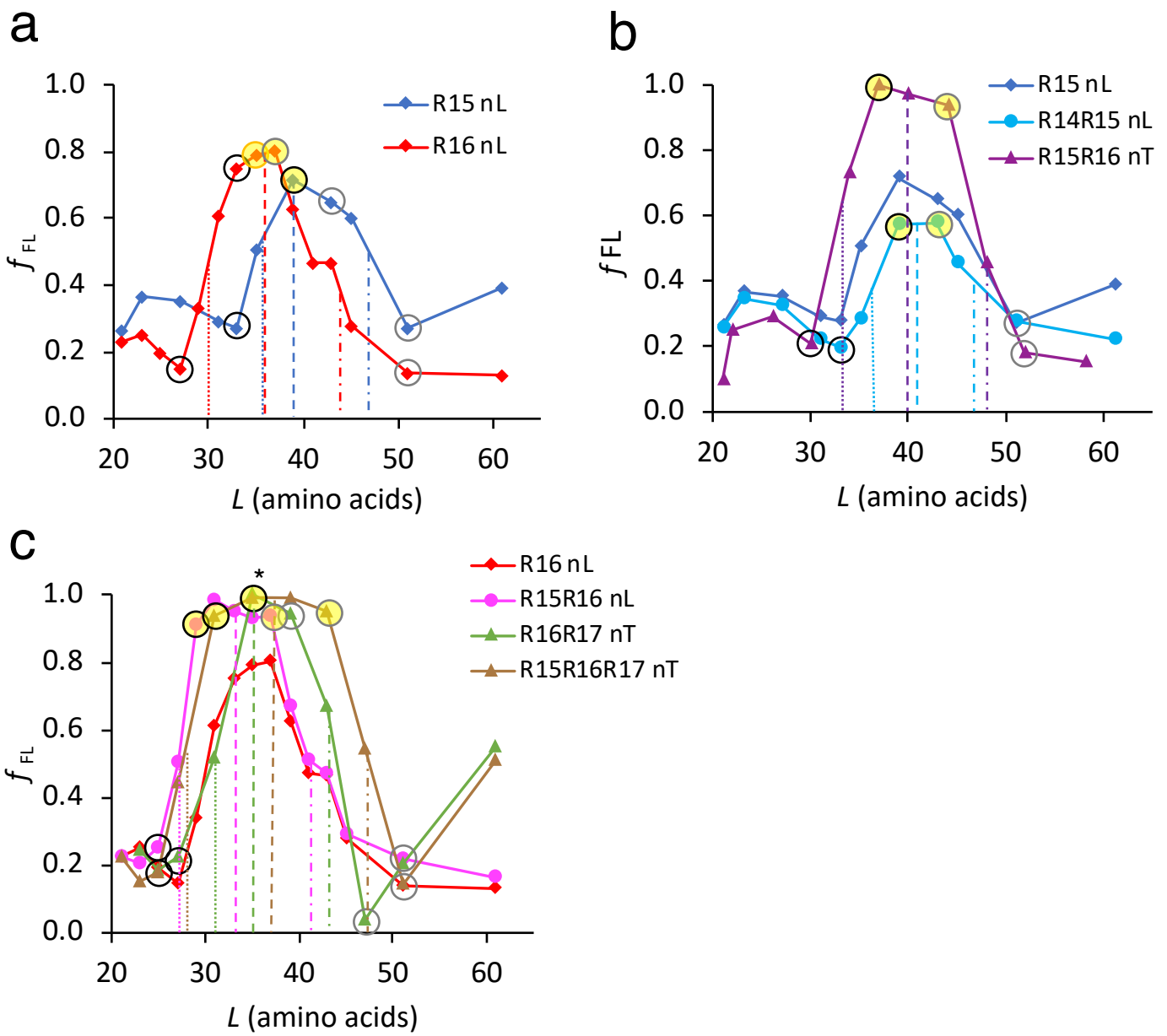
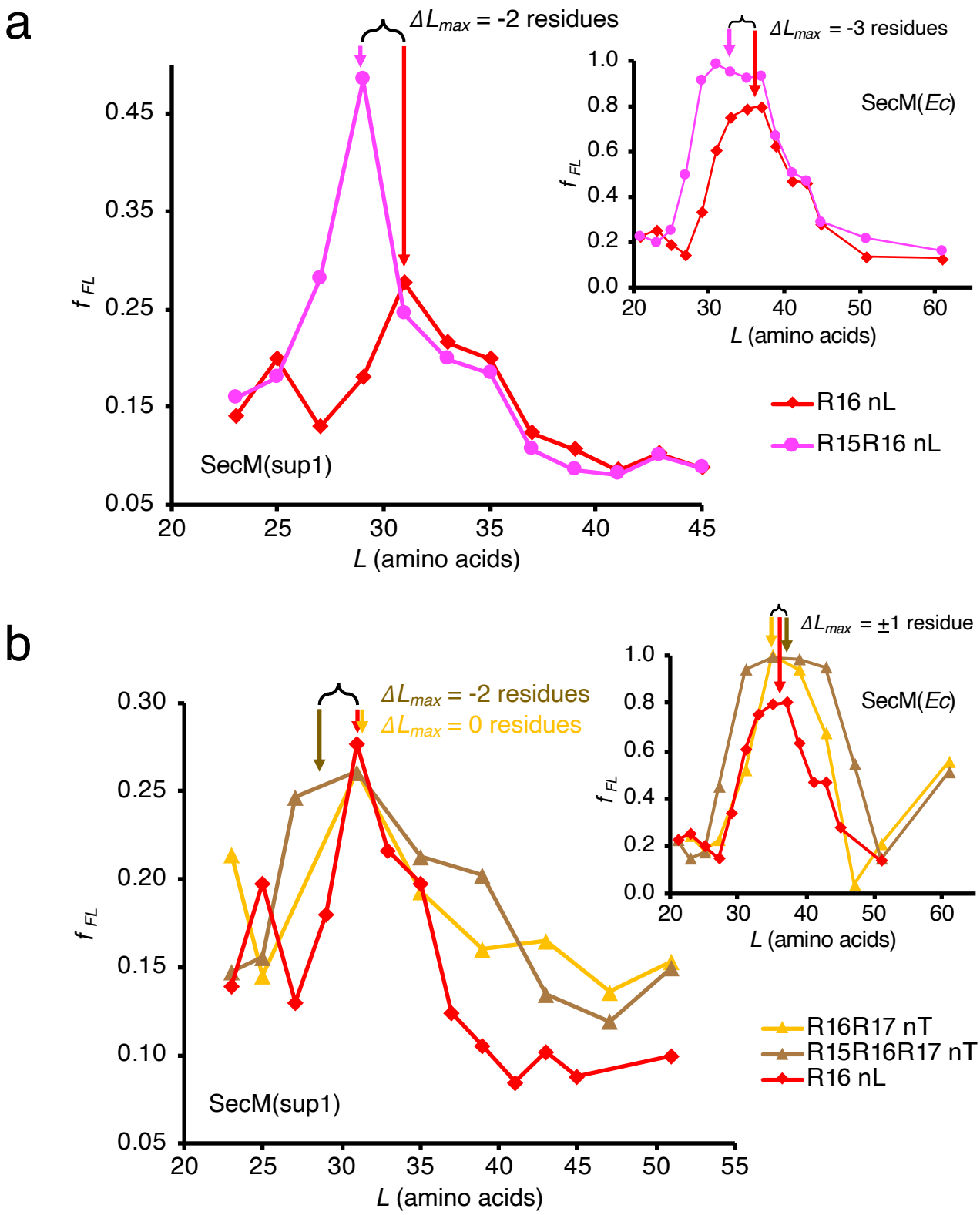


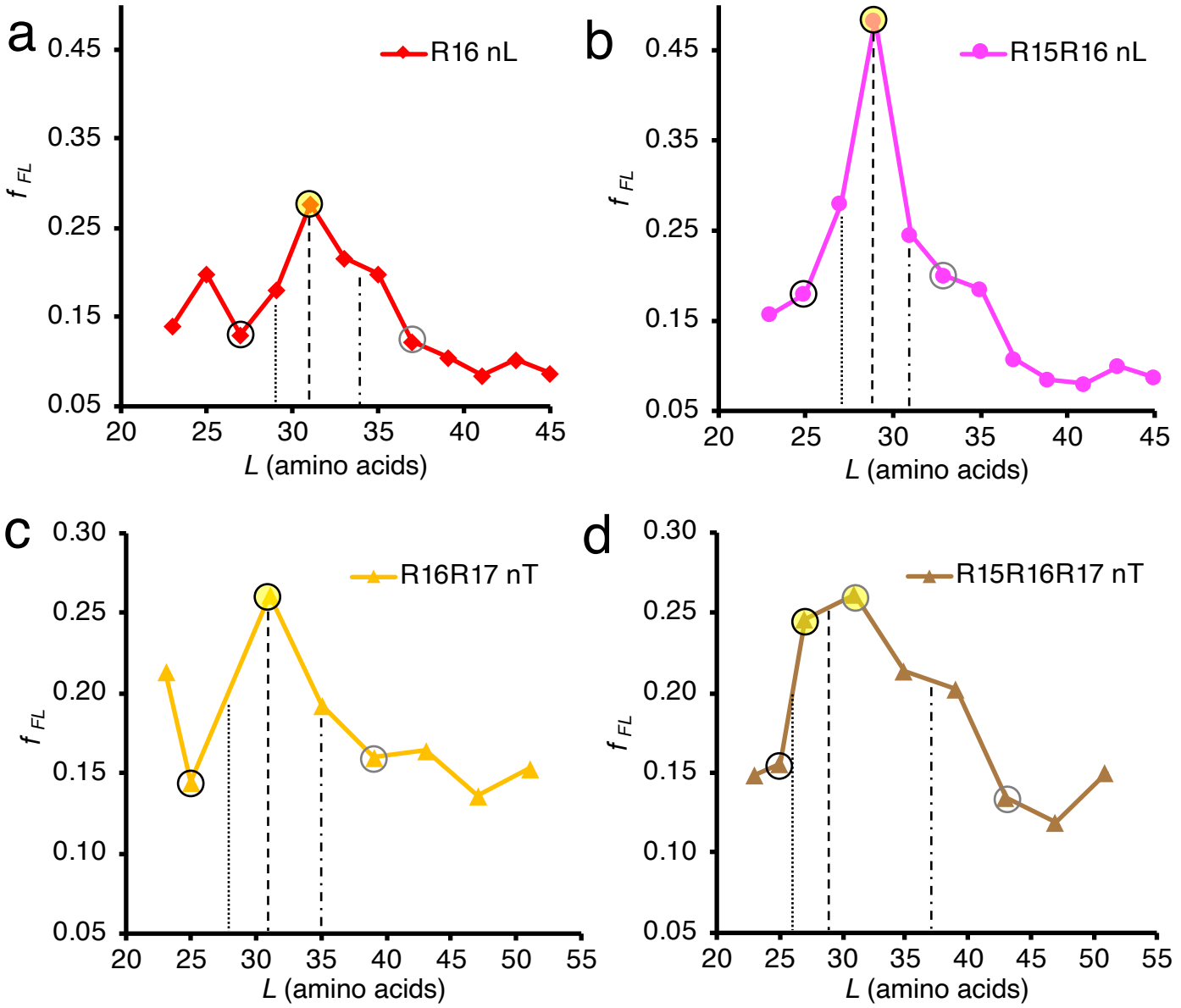
Figure 5



Supplementary Figure S1. Calculation of L_{onset}, L_{max}, and L_{end} for all constructs (annotated version of Fig. 3). In panels *a-c,* the data points used to calculate L_{onset} are indicated by black rings, L_{max} by filled yellow circles, and L_{end} by grey rings for each force profile. The *L* value at the midpoint between the selected data points is indicated by a dotted line (L_{onset}), a dashed line (L_{max}), or a dash-dot line (L_{end}) in the same color as the respective dataset. (a) Force profiles for the R15 nL and R16 nL series of constructs. (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. The (*) indicates L_{max} for R16R17 nT.



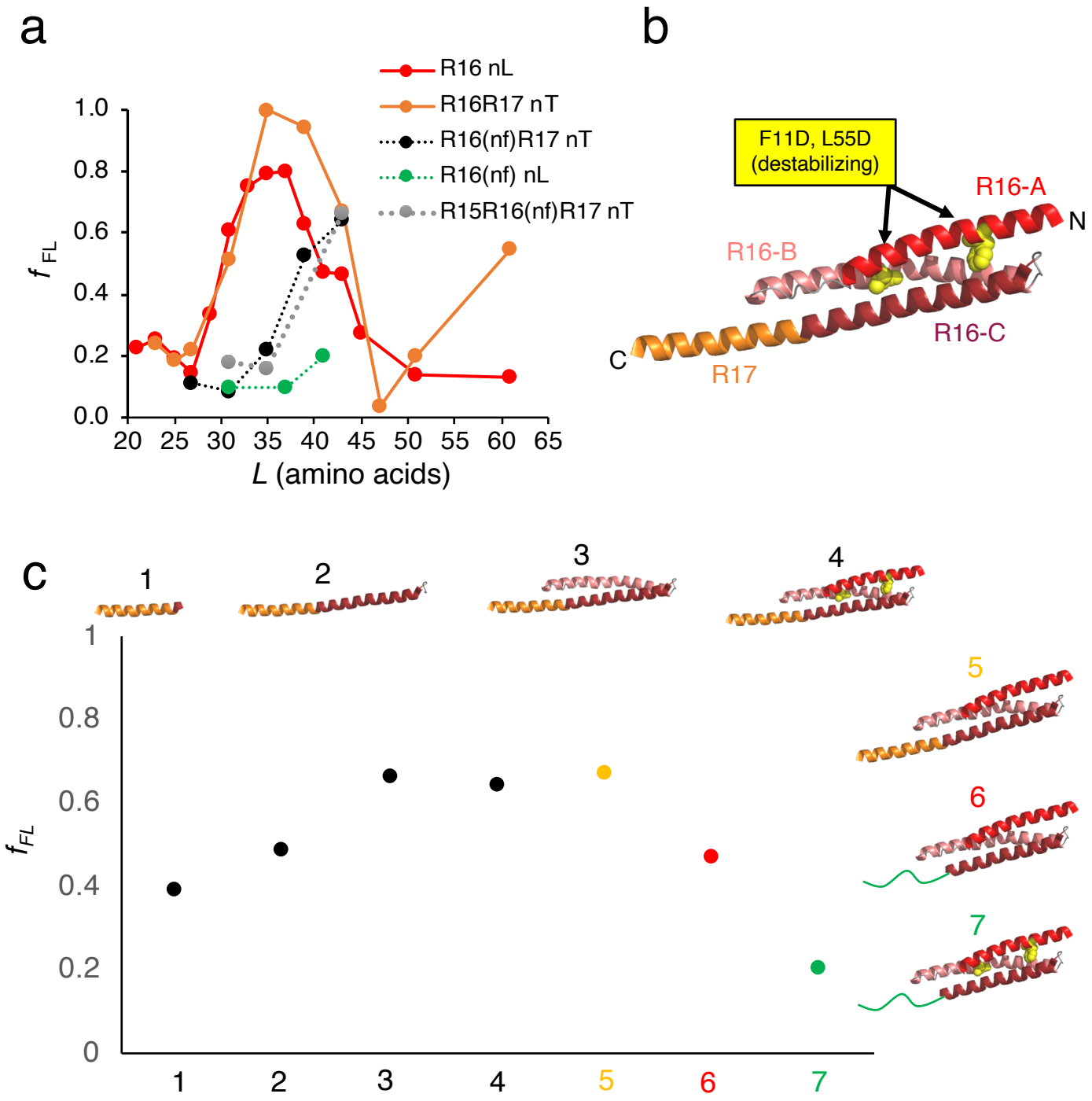
Supplementary Figure S2. FPA analysis using the stronger SecM(sup1) AP (see Supplementary Table S1). f_{FL} is the fraction full-length protein produced and L is linker length (i.e., the number of amino acid residues between the C-terminal end of the spectrin domain and the C-terminal Pro residue in the AP). Data from Fig. 3 (obtained with the weaker SecM(*Ec*) AP) are included as insets for reference. Arrows, colored to match the associated curve, indicate the L_{max} for each variant. The change in the peak location (ΔL_{max}) between R16 nL (red) and the other variants are indicated. See Supplementary Figs. S1 and S10 for more details. (a) Force profiles for the R16 nL and R15R16 nL constructs. (b) Force profiles for the R16 nL (same as in panel *a*; note change in scale), R16R17 nT, and R15R16R17 nT constructs.



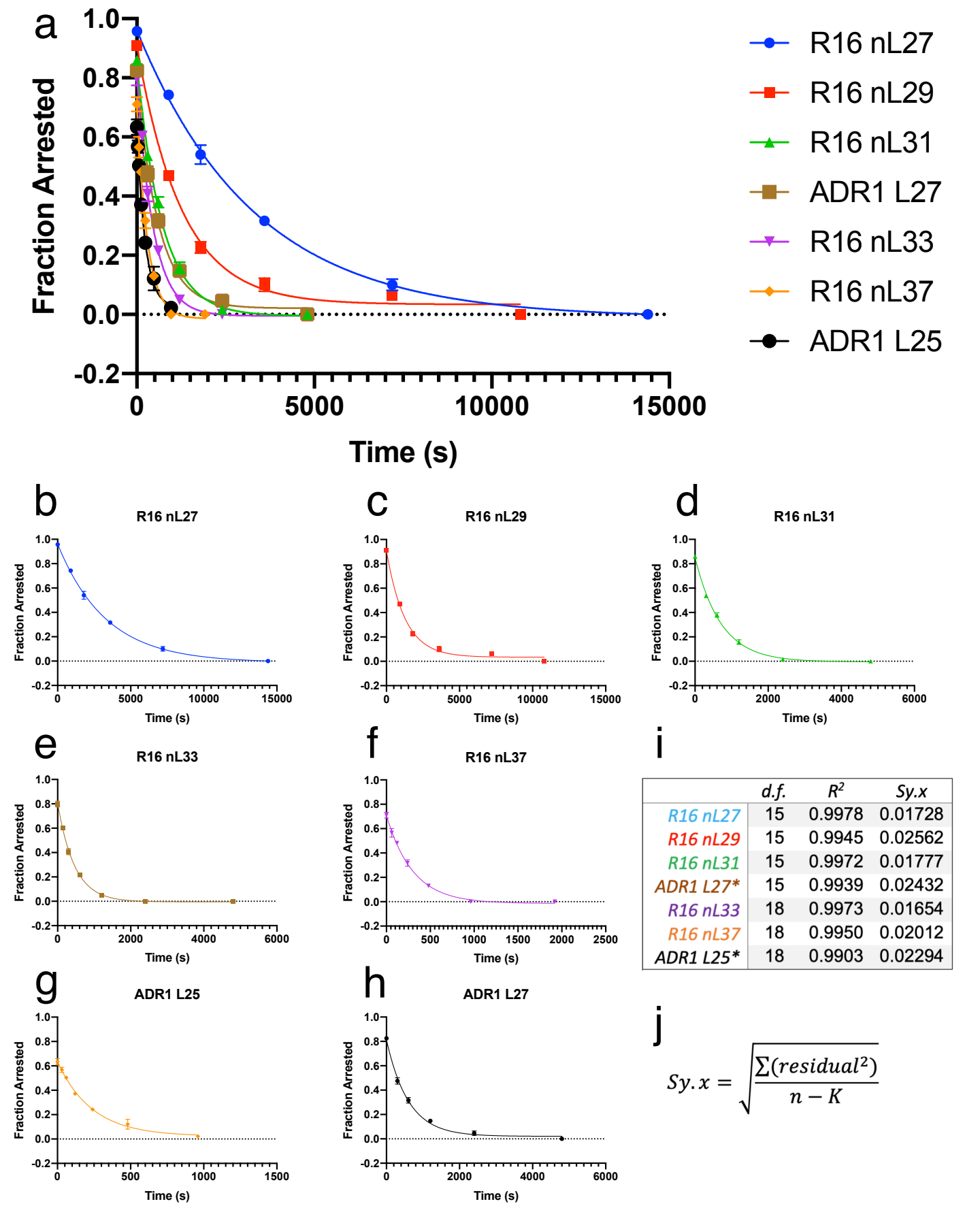
e

SecM(sup1)	L_{onset}	L_{max}	L_{end}
R16 nL	29	31	34
R15R16 nL	27	29	31
R16R17 nT	28	31	35
R15R16R17 nT	26	29	37

Supplementary Figure S3. Calculation of L_{onset}, L_{max}, and L_{end} for all SecM(sup1) AP variants monitoring R16 folding (see Supplementary Fig. S4 and Supplementary Table S1). In panels *a-c* the data points used to calculate L_{onset} are indicated by black rings, L_{max} by filled yellow circles, and L_{end} by grey rings for each force profile. The L value at the midpoint between the selected data points is indicated by a dotted line (L_{onset}), a dashed line (L_{max}), or a dash-dot line (L_{end}). (a) Force profile for R16 nL. (b) Force profile for R15R16 nL. (c) Force profile for R16R17 nT. (d) Force profile for R16R17 nT. (e) Table summarizing the calculated values.



Supplementary Figure S4. Linker effects in R16 folding. (a) Reproduction of the R16 nL (red) and R16R17 nT (orange) data from Fig. 3b compared with profiles for a R16 non-folding (nf) mutatant (F11D and L55D) in R16(nf) nL (green, dashed line), R16(nf)R17 nT (black, dashed line), and R15R16(nf)R17 nT (grey, dashed line). (b) R16R17 nT (L = 43) spectrin structure with the A-C helices indicated and the destabilizing mutations F11D and L55D in the R16-A and -C helices shown as yellow spheres. (c) Graph showing the resultant f_{FL} values at L = 43 (except for R16(nf) nL where L = 41) with increasingly long R16 parts fused to the N terminus of the 22-residue R17 nT linker (black circles). A structural representation of each protein are shown for reference (spectrin helices are colored as in panel *b* and LepB sequences colored green).



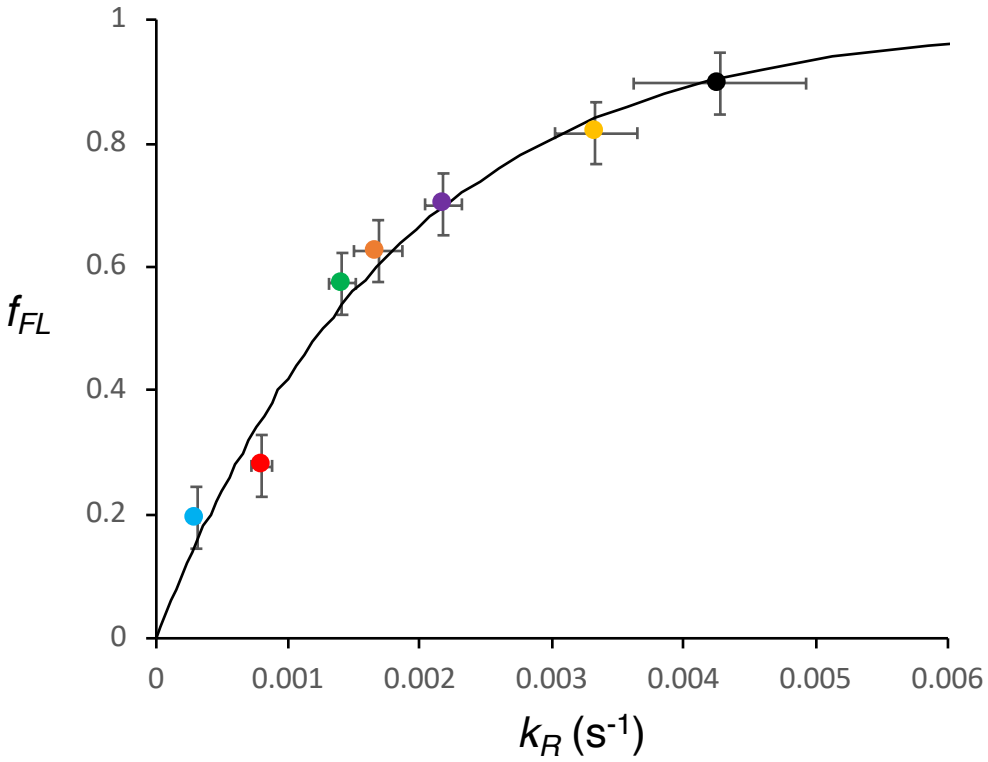
i

	d.f.	R^2	*Sy.x*
R16 nL27	15	0.9978	0.01728
R16 nL29	15	0.9945	0.02562
R16 nL31	15	0.9972	0.01777
*ADR1 L27**	15	0.9939	0.02432
R16 nL33	18	0.9973	0.01654
R16 nL37	18	0.9950	0.02012
*ADR1 L25**	18	0.9903	0.02294

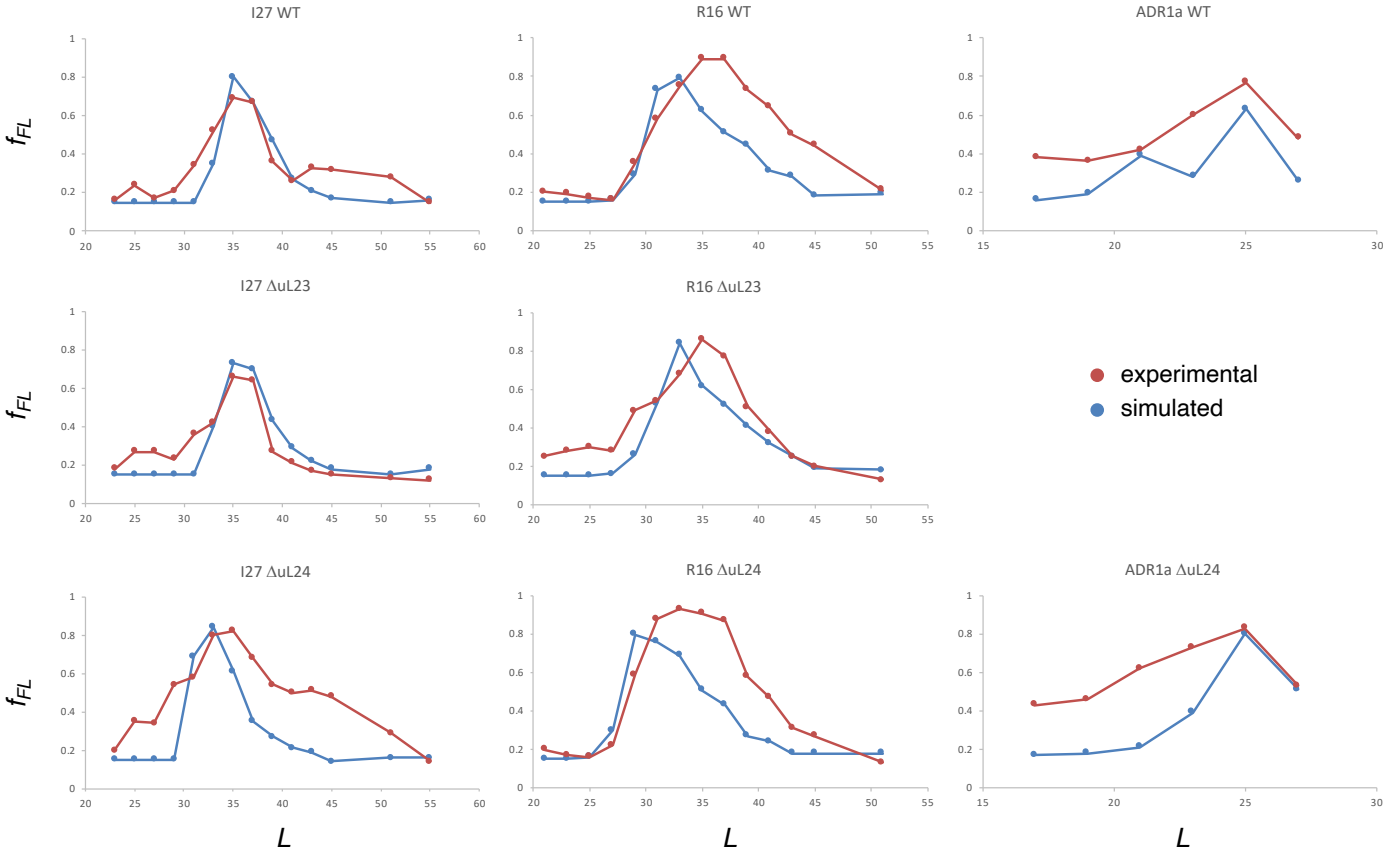
j

$$Sy.x = \sqrt{\frac{\sum(residual^2)}{n-K}}$$

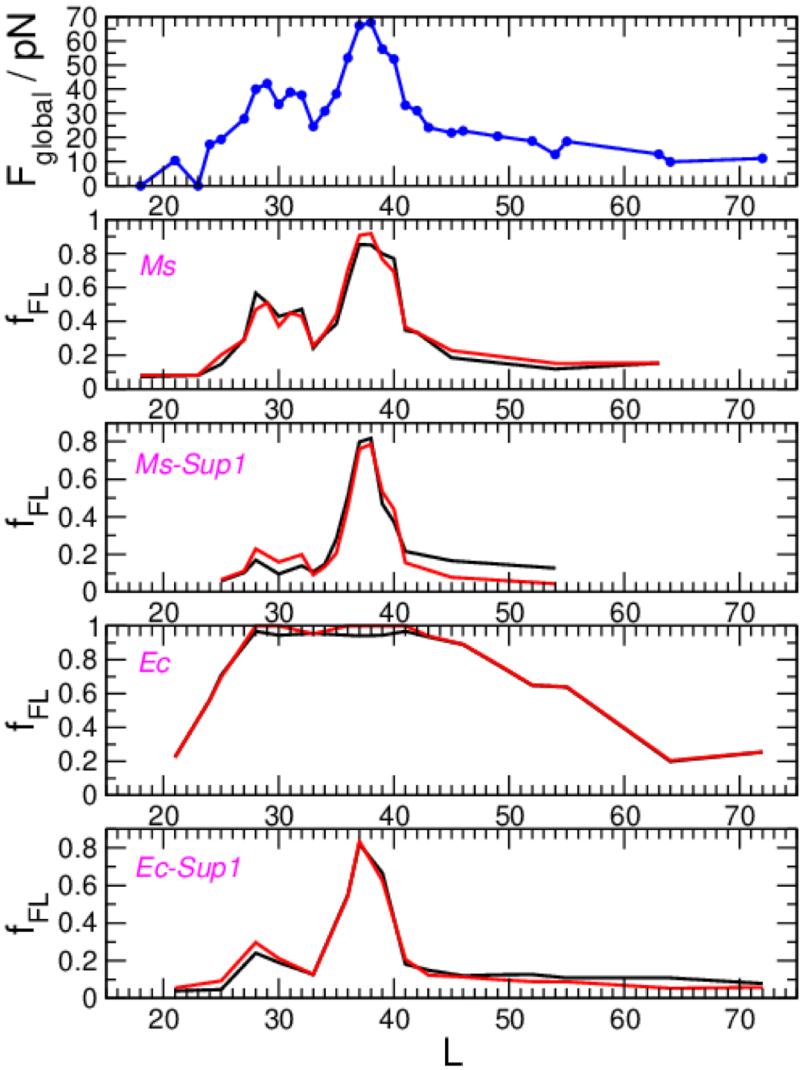
Supplementary Figure S5. Determination of k_R for each construct. (a) Scaled overlay of the individual plots in panels b-h. (b)-(h) Plot of the decrease in fraction arrested protein (f_A) over time after chasing a 5 min PURExpress translation of the indicated construct. The data were fit to the first order equation in the main text to determine the rate of release (k_R) of arrested protein from the ribosome. (i) Summary of the fitness of the equation for each construct: degrees of freedom (*d.f.*), R^2, and goodness-of-fit calculation (*Sy.x*). (j) The goodness-of-fit calculation provided by Prism 8 (Graphpad software) where *n* is the number of data points and *K* is the degrees of freedom.



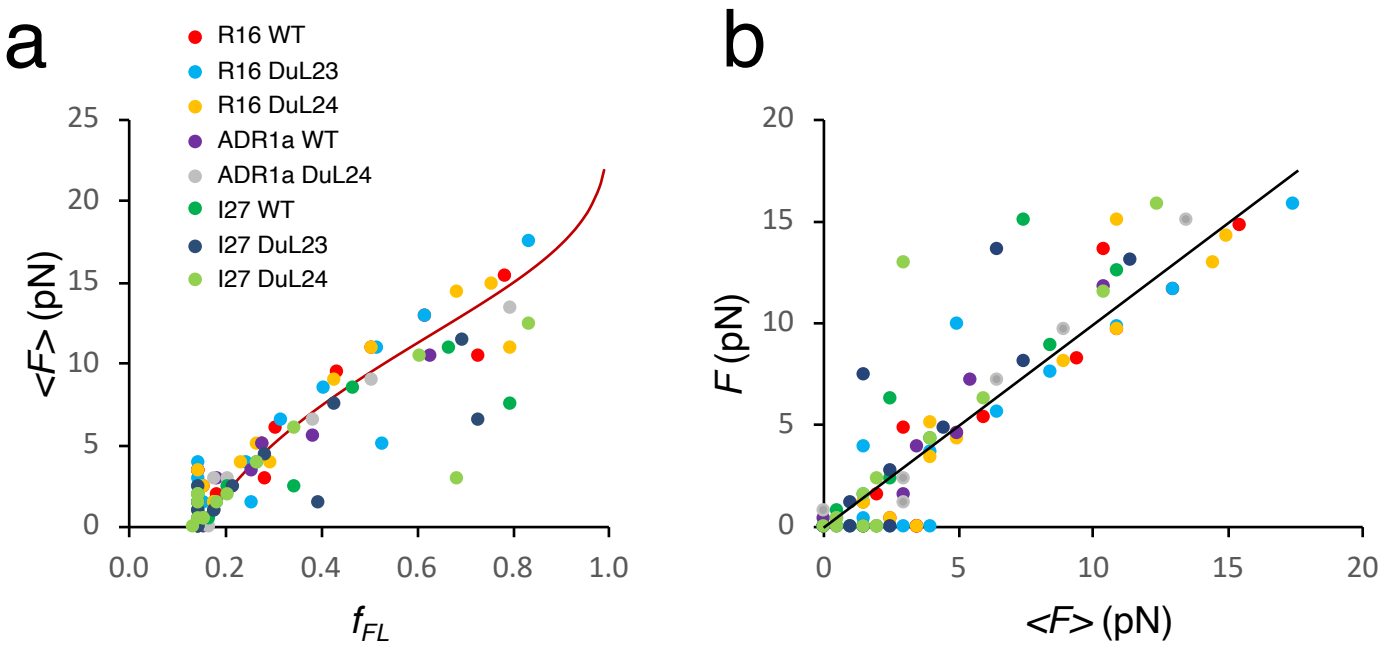
Supplementary Figure S6. Empirical relation between standard f_{FL} values (20 min. continuous [^{35}S]-Met labeling) and the release rate k_R measured in pulse-chase experiments. The data is fitted by the exponential $f_{FL} = 1 - \exp[-k_R \Delta t]$ (black curve) with $\Delta t = 550$ s, i.e., approximately half of the full incubation time. Data points are colored as in Fig. 4. Error bars show 95% CI for k_R and the typical standard error (= 0.05) (Tian et al., *Proc Natl Acad Sci USA* 115, E11284 (2018)) for f_{FL}.



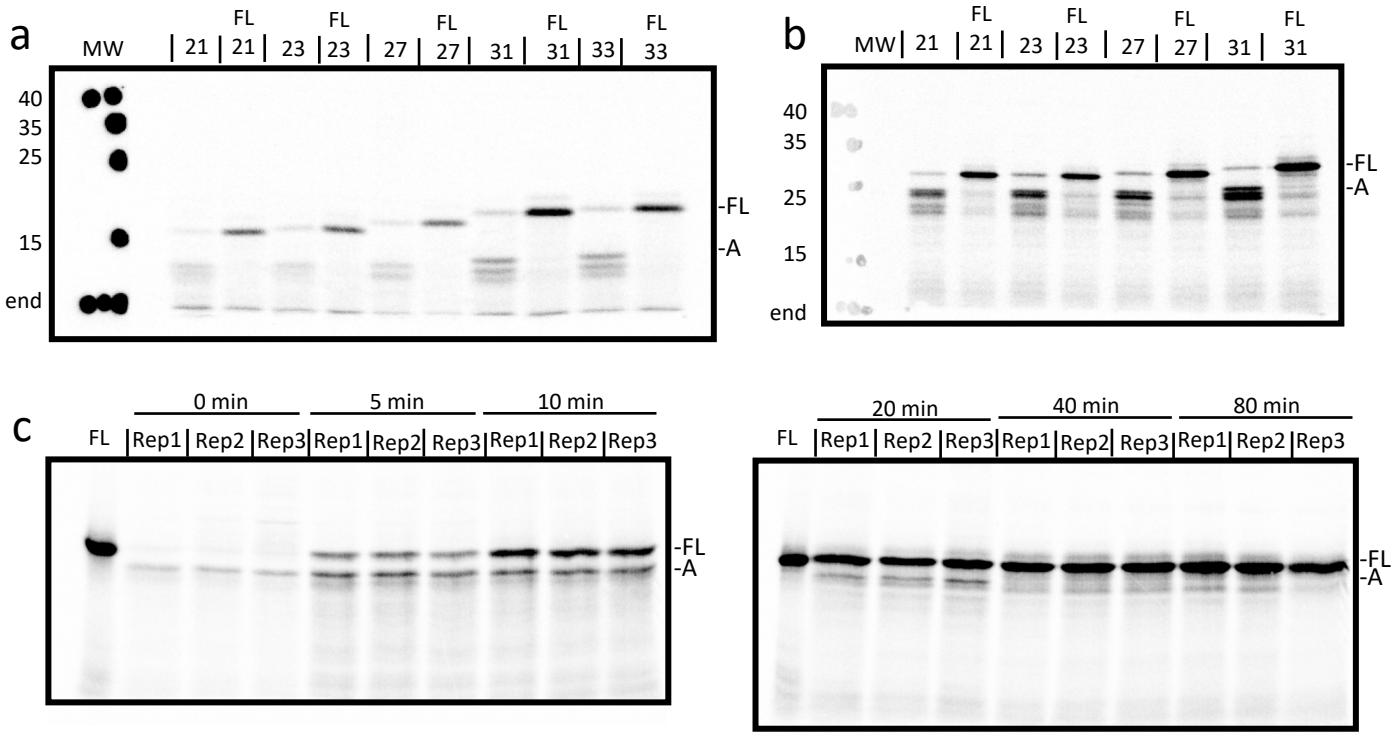
Supplementary Figure S7. Simulated (blue) and experimental (orange) force profiles for I27, R16, and ADR1a measured with wildtype (WT) and mutant (ΔuL23, ΔuL24) ribosomes (Kudva et al., *Elife* 7, 274191 (2018)). The simulated profiles were obtained as described in the *Materials and Methods* section, using the optimized Bell parameters k_0 = 3 x 10^{-4} s^{-1} and $x^{\ddagger}$ = 0.65 nm. As discussed in (Kudva et al., *Elife* 7, 274191 (2018)), the simulations reproduce L_{onset}, L_{max}, and the maximal amplitude well for all the force profiles, but underestimate L_{end} for the I27 ΔuL24, R16 WT, and R16 ΔuL24 ribosomes. Since it is generally difficult to fit the simulated force profile to the experimental ADR1a ΔuL23 force profile described in (Kudva et al., *Elife* 7, 274191 (2018)), this was not included in the optimization of the Bell parameters.



Supplementary Figure S8. Global fit to the experimental data for translocon-mediated insertion of a transmembrane segment of composition [6L, 13 A] into the *E. coli* inner membrane from Ismail et al. (*Nat. Struc. Mol. Biol.* 19, 1018-1022 (2012)) to a common set of pulling forces. For each of the SecM arrest peptide (AP) variants *Ms*, *Ms-Sup1, Ec, Ec-Sup1* used in that work, an AP-specific set of Bell parameters $k_0, \Delta x^{\ddagger}$ is defined. In addition it is assumed that the force exerted on the AP depends only on the linker length *L*, i.e. $F_{global}(L)$. The escape rate for each peptide and linker length can then be determined as $k_R(L) = k_0 \exp[\frac{F_{global}(L)\Delta x^{\ddagger}}{k_B T}]$, and hence the fraction full-length protein f_{FL} is obtained from $f_{FL}(L) = 1 - \exp[-k_R(L)t_i]$, where t_i is the incubation time (120 s in these experiments). The AP-independent forces $F_{global}(L)$ and AP-dependent Bell parameters are globally fitted by minimizing the least-squares difference between experimental and calculated f_{FL} using a basin-hopping algorithm. The Bell parameters for the *Ec* AP were restrained to be similar to those found in Goldman et al (*Science*, 348, 457-460 (2015)). The best-fit global forces are shown in the top panel above, and the experimental (black curves) and calculated (red curves) f_{FL} data are shown in the lower panels for each AP. The optimal Bell parameters were found to be: *Ms:* $k_0 = 7.1\times10^{-4}\text{s}^{-1}$; $x^{\ddagger} = 0.21$ nm; *Ms-Sup1:* $k_0 = 1.64\times10^{-4}\text{s}^{-1}$; $x^{\ddagger} = 0.27$ nm; *Ec:* $k_0 = 3.3\times10^{-4}\text{s}^{-1}$; $x^{\ddagger} = 0.74$ nm; *Ec-Sup1:* $k_0 = 2.4\times10^{-4}\text{s}^{-1}$; $x^{\ddagger} = 0.26$ nm.



Supplementary Figure S9. Force calculations based on coarse-grained molecular dynamics simulations of I27, R16, and ADR1a force profiles measured with wildtype (WT) and mutant (DL23, DL24) ribosomes (Kudva et al., *Elife* 7, 274191 (2018)). (a) Relation between the ensemble average force $\langle F_{MD} \rangle$ and f_{FL} from simulation for the indicated constructs. The red curve is the same as in Fig. 4b, relating the force F from AP measurements to f_{FL}. (b) Comparison of the ensemble average force $\langle F_{MD} \rangle$ with the force F calculated from simulated f_{FL} values using equation [3]. The black line represents $F = \langle F_{MD} \rangle$ The similarity of the forces indicates that they are small enough that $\langle F_{MD} \rangle$ can be well approximated by F. In general F is expected to be slightly larger than $\langle F_{MD} \rangle$.



Supplementary Figure S10. Examples of gels used for quantification of f_{FL}. Above any lane FL indicates a full-length control construct (with an inactivated AP, see *Supplementary Table 1*). Location of full-length (FL) and arrested (A) products from the rightmost construct in each gel are indicated. (a) R15 nL. Construct *L* values are indicated above each lane and MW marker bands (kDa) are denoted as spots and labelled to the left of the gel. (b) R14R15 nL. Construct *L* values are indicated above each lane and MW marker bands (kDa) are denoted as spots and labelled to the left of the gel. (c) ADR1 L27 pulse-chase experiments. Triplicate samples (Rep1, Rep2, Rep3) at various time points (labelled in minutes) are above the lanes.

Supplementary Table 1

Amino acid sequences of the analyzed spectrin constructs Color coding: [R14/R15/R16/R17 sequence]-[LepB derived linker sequence]-[Arrest Peptide]-[LepB derived C-terminal tail]. The critical proline mutated to alanine in the full-length control constructs is in **bold**. The residue *QG* that are mutated to *PP* in the sup1 SecM variant are *italicised*. The destabilizing mutations in R15 (**F**18D and **I**55D) and R16 (**F**11D and **L**55D) that prevent native folding in solution are in **bold**.

Linker length, *L* (aa)	Translated protein sequence
R15 nL 21	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGFSTPVWISQAQG IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 23	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGSSFSTPVWISQA QGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 27	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGDWMSSFSTPVWI SQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 31	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGYFMMGDWMSSFS TPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 33	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGGQYFMMGDWMSS FSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 35	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGPPGQYFMMGDWM SSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 39	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGTWIVPPGQYFMM GDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 41	MKLKEANKQQNFNTGIKD**D**DFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEAD**D**SAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGNATWIVPPGQYF MMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 43	MKLKEANKQQNFNTGIKD**F**DFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEAD**I**SAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGQQNATWIVPPGQ YFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 45	MKLKEANKQQNFNTGIKD**F**DFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEAD**I**SAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGPGQQNATWIVPP GQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 51	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGYQKTLVPGQQNA TWIVPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 nL 61	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGKFAYGIKDPIYQ KTLVPGQQNATWIVPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSR IGGIH*
R14 R15 nL 21	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHAN ADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNT GIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSS AFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGFSTPVWISQAQGIRAG**P**GSSDKQEGE WPTGLRLSRIGGIH*
R14 R15 nL 23	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHAN ADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNT GIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSS AFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGSSFSTPVWISQAQGIRAG**P**GSSDKQE GEWPTGLRLSRIGGIH*
R14 R15 nL 27	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHAN ADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNT GIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSS

	AFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R14 R15 nL 31	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHANADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R14 R15 nL 33	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHANADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R14 R15 nL 35	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHANADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R14 R15 nL 39	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHANADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGTWIVPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R14 R15 nL 43	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHANADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGQQNATWIVPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R14 R15 nL 45	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHANADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGPGQQNATWIVPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R14 R15 nL 51	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHANADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGYQKTLVPGQQNATWIVPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R14 R15 nL 61	MKLGESQTLQQFSRDVDEIEAWISEKLQTASDESYKDPTNIQLSKLLSKHQKHQAFEAELHANADRIRGVIEMGNPLIERGACAGSEDAVKARLAALADQWEFLVQKSSEKSQKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGKFAYGIKDPIYQKTLVPGQQNATWIVPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 21	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNSGSGFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 22	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESGSGFSTPVWISQA*Q*GIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 26	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 30	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 34	MKLKEANKQQNFNTGIKD**F**DFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEAD**I**SAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 37	MKLKEANKQQNFNTGIKD**F**DFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEAD**I**SAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEESGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*

R15 R16 nT 40	MKLKEANKQQNFNTGIKD**F**DFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEAD**I**SAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWISGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 44	MKLKEANKQQNFNTGIKD**F**DFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEAD**I**SAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 48	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 52	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 58	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGG IH*
R16 nL 21	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGFSTPVWISQAQG IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 23	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGSSFSTPVWISQA *QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 25	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGWMSSFSTPVWIS QA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 27	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGGDWMSSFSTPVW ISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 29	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGMMGDWMSSFSTP VWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 31	MKLNESHRLHQ**F**FRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGYFMMGDWMSSFS TPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 33	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGGQYFMMGDWMSS FSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 35	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGPPGQYFMMGDWM SSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 37	MKLNESHRLHQ**F**FRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGIVPPGQYFMMGD WMSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 39	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGTWIVPPGQYFMM GDWMSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 41	MKLNESHRLHQ**F**FRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGNATWIVPPGQYF MMGDWMSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 43	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGQQNATWIVPPGQ YFMMGDWMSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 45	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGVPGQQNATWIVP PGQYFMMGDWMSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 51	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGYQKTLVPGQQNA TWIVPPGQYFMMGDWMSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nL 61	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGKFAYGIKDPIYQ KTLVPGQQNATWIVPPGQYFMMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSR IGGIH*

R15 R16 nL 21	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLR LSRIGGIH*
R15 R16 nL 23	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTG LRLSRIGGIH*
R15 R16 nL 25	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGWMSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWP TGLRLSRIGGIH*
R15 R16 nL 27	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGGDWMSSFSTPVWISQA*QG*IRAG**P**GSSDKQEGE WPTGLRLSRIGGIH*
R15 R16 nL 29	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGMMGDWMSSFSTPVWISQA*QG*IRAG**P**GSSDKQE GEWPTGLRLSRIGGIH*
R15 R16 nL 31	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGYFMMGDWMSSFSTPVWISQA*QG*IRAG**P**GSSDK QEGEWPTGLRLSRIGGIH*
R15 R16 nL 33	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGGQYFMMGDWMSSFSTPVWISQA*QG*IRAG**P**GSS DKQEGEWPTGLRLSRIGGIH*
R15 R16 nL 35	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGPPGQYFMMGDWMSSFSTPVWISQA*QG*IRAG**P**G SSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nL 37	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGIVPPGQYFMMGDWMSSFSTPVWISQA*QG*IRAG **P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nL 39	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGTWIVPPGQYFMMGDWMSSFSTPVWISQA*QG*IR AG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nL 41	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGNATWIVPPGQYFMMGDWMSSFSTPVWISQA*QG* IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nL 43	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGQQNATWIVPPGQYFMMGDWMSSFSTPVWISQA *QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nL 45	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE

	SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGVPGQQNATWIVPPGQYFMMGDWMSSFSTPVWI SQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nL 51	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGYQKTLVPGQQNATWIVPPGQYFMMGDWMSSFS TPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nL 61	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGKFAYGIKDPIYQKTLVPGQQNATWIVPPGQYF MMGDWMSSFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 21	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLESGSGFSTPVWISQAQG IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 23	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESSGSGFSTPVWISQA *QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 25	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLESGSGFSTPVWIS QA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 27	MKLNESHRLHQ**F**FRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQSGSGFSTPVW ISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 31	MKLNESHRLHQ**F**FRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVASGSGFS TPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 35	MKLNESHRLHQ**F**FRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEESG SGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 39	MKLNESHRLHQ**F**FRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEE AWSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 43	MKLNESHRLHQ**F**FRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEE AWINEKSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 47	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEE AWINEKMTLVSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 51	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEE AWINEKMTLVASEDSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 R17 nT 61	MKLNESHRLHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAI QGVLDTGKKLSDDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEE AWINEKMTLVASEDYGDTLAAIQGSGSGFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSR IGGIH*
R15 R16 R17 nT 21	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLESGSGFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLR LSRIGGIH*
R15 R16 R17 nT 23	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTG LRLSRIGGIH*
R15 R16 R17 nT 25	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLESGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWP TGLRLSRIGGIH*

R15 R16 R17 nT 27	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLEYQSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGE WPTGLRLSRIGGIH*
R15 R16 R17 nT 31	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQ**F**FRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVASGSGFSTPVWISQA*QG*IRAG**P**GSSDK QEGEWPTGLRLSRIGGIH*
R15 R16 R17 nT 35	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQ**F**FRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEESGSGFSTPVWISQA*QG*IRAG**P**G SSDKQEGEWPTGLRLSRIGGIH*
R15 R16 R17 nT 39	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEEAWSGSGFSTPVWISQA*QG*IR AG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 R17 nT 43	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQ**F**FRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAE**L**AAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEEAWINEKSGSGFSTPVWISQA *QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 R17 nT 47	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEEAWINEKMTLVSGSGFSTPVW ISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 R17 nT 51	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEEAWINEKMTLVASEDSGSGFS TPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 R17 nT 61	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEE SWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEE IQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEEAWINEKMTLVASEDYGDTLA AIQGSGSGFSTPVWISQAQGIRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R16 nT 40	MESHRLHQFFRDMDDEESWISGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGI H*
R15-B+C R16 nT 40	MASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETIN GRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEESWISGSGFSTPVWISQA*QG*IRAG**P**GSSDK QEGEWPTGLRLSRIGGIH*
R15-C R16 nT 40	MSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDDEESWISGSGFSTP VWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15-B+C R16 nT 40 GG	MASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETIN GRFQRIKSMAAARRAKLNESHRLHQFFRDMDGGDEESWISGSGFSTPVWISQA*QG*IRAG**P**GSS DKQEGEWPTGLRLSRIGGIH*
R15-B+C R16 nT 40 P	MASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETIN GRFQRIKSMAAARRAKLNESHRLHQFFRDMDPDEESWISGSGFSTPVWISQA*QG*IRAG**P**GSSD KQEGEWPTGLRLSRIGGIH*
R15-B+C R16 nT 40 GSGS	MASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRETIN GRFQRIKSMAAARRAKLNESHRLHQFFRDMDGSGSDEESWISGSGFSTPVWISQA*QG*IRAG**P**G SSDKQEGEWPTGLRLSRIGGIH*
R15-C R16 nT 40 GG	MSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDGGDEESWISGSGFS TPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*

R15-C R16 nT 40 P	MSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDPDEESWISGSGFST PVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15-C R16 nT 40 GSGS	MSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDGSGSDEESWISGSG FSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 40 GG	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDGGD EESWISGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 40 P	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDPDE ESWISGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R15 R16 nT 40 GSGS	MKLKEANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRL KDLNSQADSLMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRLHQFFRDMDGSG SDEESWISGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*
R17 nT 43	MESLEYQQFVANVEEEEAWINEKSGSGFSTPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRI GGIH*
R16-B+C R17 nT 43	MSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQGVLDTGKKLSDDNTIGKEEIQQRLAQFV DHWKELKQLAAARGQRLEESLEYQQFVANVEEEEAWINEKSGSGFSTPVWISQA*QG*IRAG**P**GS SDKQEGEWPTGLRLSRIGGIH*
R16-C R17 nT 43	MDNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESLEYQQFVANVEEEEAWINEKSGSGF STPVWISQA*QG*IRAG**P**GSSDKQEGEWPTGLRLSRIGGIH*