Classification: Biological Sciences - Biochemistry

Cotranslational folding of a pentarepeat β-helix protein

Luigi Notari[1#], Markel Martínez-Carranza[1#], Pål Stenmark[1*], Gunnar von Heijne[,1,2*]

[1]Department of Biochemistry and Biophysics
Stockholm University, SE-106 91 Stockholm, Sweden
[3]Science for Life Laboratory Stockholm University, Box 1031, SE-171 21 Solna, Sweden

[*]Corresponding authors. GvH Phone: Int+46-8-16 25 90, E-mail: gunnar@dbb.su.se.
PS Phone: Int+46-8-16 37 29, E-mail: stenmark@dbb.su.se.

[#] These authors contributed equally to this work.

Keywords: pentapeptide-repeat protein, beta-helix, cotranslational folding, X-ray structure, *Clostridium botulinum*

Abstract

It is becoming increasingly clear that many proteins start to fold cotranslationally, even before the entire polypeptide chain has been synthesized on the ribosome. One class of proteins that *a priori* would seem particularly prone to cotranslational folding is repeat proteins, *i.e*., proteins that are built from a linear array of nearly identical folding units. However, while the folding of repeat proteins has been studied extensively *in vitro* with purified proteins, only a handful of studies have addressed the issue of cotranslational folding of repeat proteins. Here, we have determined the structure and studied the cotranslational folding a β-helix pentarepeat protein from *Clostridium botulinum*, using an assay in which the SecM translational arrest peptide serves as a force sensor to detect folding events. We find that the folding nucleus involves the first four of the eight β-helix coils in the protein, and that folding starts when this folding nucleus is ~35 residues away from the P-site, near the mouth of the exit tunnel. The early cotranslational formation of a folded nucleus from which the β-helix can grow may be important to avoid misfolding events *in vivo*.

Significance statement

It has long been thought that many proteins start to fold cotranslationally as they come off the ribosome, before the entire chain has been synthesized. However, the overwhelming majority of experimental studies of protein folding have been carried out *in vitro*, using purified proteins. Here, we use an assay that allows us to detect the force that a cotranslationally folding protein exerts on the nascent polypeptide chain to follow the folding of a β-helix repeat protein as it emerges from the ribosome exit tunnel. We find that only four of the eight β-coils of the protein need to be present for the protein to start folding, and that this folding intermediate is formed near the mouth of the exit tunnel.

\body

Introduction

With their simple, repetitive architecture composed of a linear array of nearly identical folding units, repeat proteins represent an important paradigm for protein folding studies (1). Typical repeat proteins are ankyrin repeat proteins, tetratricorepeat proteins, heat repeat proteins, leucine-rich repeat proteins, and various kinds of β-helix proteins (2). In general, the folding of an individual repeat is thermodynamically unfavorable while the interaction between successive repeats is favorable, meaning that a critical number of folded neighboring repeats need to interact in order for the folded state to become more stable than the unfolded state (1).

As for protein folding studies in general, folding of repeat proteins has mainly been analyzed *in vitro*, using purified proteins. *In vivo*, however, proteins can start to fold cotranslationally (3). For repeat proteins in particular, cotranslational folding, where repeats are added to a growing folding nucleus as they emerge from the ribosome exit tunnel, seems a likely scenario. Still, only a handful of studies have addressed the issue of cotranslational folding of repeat proteins (4, 5).

The pentapeptide repeat protein (PRP) family belongs to the class of β-helix proteins. In PRPs, four pentapeptide repeats form an approximately square repeating unit (a "coil"), and a string of coils form the β-helix (6). The hydrophobic core inside the β-helix contains conserved Phe and Leu residues while polar and charged residues decorate its surface, in many cases mimicking a DNA double helix (7). Here, we analyze the cotranslational folding of a PRP from *Clostridium botulinum* (PENT), a 217-residue polypeptide that forms a highly regular eight-coil β-helix, similar to that seen in *M. tuberculosis* MfpA (7, 8). Using a force-sensing assay based on the SecM

translational arrest peptide (AP), we find that the main folding transition takes place when approximately four N-terminal coils have emerged from the ribosome exit tunnel, i.e., when only about half the protein has been synthesized. We see no evidence of a second folding transition, suggesting that the remaining coils are added continuously to the growing β-helix, as they emerge from the exit tunnel. Since incompletely folded β-helix proteins are prone to aggregation (9), the early cotranslational formation of a folded nucleus from which the β-helix can grow may be important to avoid misfolding events *in vivo*.

Results

PENT forms a highly regular β-helix

To provide a structural context for our cotranslational folding studies, we first determined the crystal structure of the *Clostridium botulinum* pentapeptide repeat-containing protein PENT (UniProtKB A0A0M0A2X5; see Supplementary Table 1 for refinement statistics and Supplementary Fig. S1 for B-values). The protein crystallizes as a dimer and adopts a right-handed quadrilateral β-helix fold, Fig. 1a, first observed in the *M. tuberculosis* MfpA PRP (8). Four parallel β-sheets constitute the sides of the square-shaped helix, and its diameter varies between 20 to 30 Å through the squared shape of every coil. The parallel nature of the β-sheets gives a slight left-handed helicity to the β-helix, each coil being slightly offset from the previous one. The N- and C-terminal caps adopt an α-helical fold that is perpendicular to the axis of the β-helix, and seal off its hydrophobic core. A dimer interface is formed between the C-terminal caps of two PENT monomers, with the α-helices oriented perpendicularly to each other. The two monomers are coaxial, forming a 100 Å long dimer.

Each β-helix coil is formed by four pentapeptide repeat units. The central residue of each unit is most often a phenylalanine, and occasionally a methionine, Fig. 1b. These residues are designated with the *i* position in the repeat, and their side chains point inwards in the β-helix. Residues in the i^{-2} position also point inwards, and constitute the corners of the quadrilateral β-helix. Cysteine and serine residues populate this position most often. It is worth noting that no disulfide bonds are formed between any cysteine residues, which might otherwise slow down the folding process. No proline residues are found in the coils, which would also impair folding. The side chains of residues in position i^{-1}, i^{+1} and i^{+2} point outwards, and are often populated by amino acids with charged side chains (6, 7).

Force-profile analysis of cotranslational folding

Translational arrest peptides (APs) are short stretches of polypeptide that interact with the ribosome exit tunnel in such a way that translation is stalled when the ribosome reaches the last codon in the AP (10). The stall can be overcome by external forces pulling on the nascent chain (11), and the stalling efficiency of a given AP is reduced in proportion to the magnitude of the external pulling force (12, 13). APs can therefore be used as force sensors to follow a range of cotranslational processes such as membrane protein biogenesis (12, 14), protein translocation (15), and protein folding (16-18).

A schematic representation of how cotranslational protein folding generates force on the nascent chain is shown in Fig. 2a. For short constructs for which there is not enough room in the ribosome exit tunnel for the protein to fold at the point when the ribosome reaches the end of the AP, or for long constructs where the protein has already folded when the ribosome reaches the end of the AP, little force is generated.

However, for constructs of intermediate length where there is just enough space in the tunnel for the protein to fold if the tether is stretched out from its equilibrium length, some of the free energy gained upon folding will be stored as elastic energy (increased tension) in the nascent chain, reducing stalling. By measuring the stalling efficiency for a series of constructs of increasing length, a force profile can be generated that shows how the folding force varies with the location of the protein in the ribosome tunnel (18), and hence when during translation the protein starts to fold.

The basic construct used in the force-measurement experiment is shown in Fig. 2b. The C-terminal end of the force-generating moiety (PENT in the present case) is placed 21 residues upstream of the C-terminal end of the SecM AP (19), which in turn is followed by a 23-residue C-terminal tail. Constructs are translated for 20 min. in the PURExpress™ coupled *in vitro* transcription-translation system (20) in the presence of [^{35}S]-Met, the radiolabeled protein products are analyzed by SDS-PAGE, and bands are quantitated on a phosphoimager, Fig. 2c. For constructs where little pulling force is exerted on the nascent chain, stalling is efficient and the arrested (A) form of the protein dominates. In contrast, for constructs experiencing a high pulling force there is little stalling and the full-length (FL) form dominates. We use the fraction full-length protein, $f_{FL} = I_{FL}/(I_{FL}+I_A)$ (where I_i is the intensity of band i = A, FL), as a measure of the force exerted on the nascent chain (12).

PENT force profile

To generate the full force profile for PENT, a series of constructs where the PENT-moiety was progressively truncated from its C-terminal end were translated in the PURE system, and f_{FL} was plotted as a function of the length of the truncated PENT part, Fig. 3. The force profile has two minor peaks corresponding to PENT truncations at residues ~50 (constructs T47, T51; T*n* denotes a construct that includes PENT

residues 1 to *n*) and ~95 (constructs T93, T96), and a major, 20-residue wide peak corresponding to PENT truncations at residues ~120-140 (constructs T122-T140).

Effects of mutations in the hydrophobic core of PENT

To ascertain whether the peaks in the force profile reflect partial folding of the PENT β-helix or may be caused by, *e.g.*, minor alterations in the way the nascent chain interacts with the ribosome exit tunnel, we mutated hydrophobic core residues in relevant β-helix coils in constructs T93 and T129, Fig. 4a. Simultaneous mutation of four core residues in the N-terminal cap and coils 2-3 to Ala (V15A, F18A, F57A, F67A) in construct T93 had no significant effect on f_{FL} (the f_{FL} value changed from 0.38 to 0.44, data not shown), meaning that this peak probably does not represent a partially folded intermediate.

In contrast, mutations in core residues of the T129 construct had strong, cumulative effects on f_{FL}. The double mutation V15A+F18A in the N-terminal cap domain led to a significant decrease in f_{FL} when compared to the wildtype sequence, Fig. b, suggesting that the cap plays an important role in the folding process and stabilizes the folding intermediate (21). The cumulative introduction of F→A mutations within the β-helix coils (F47A, F57A, F67A, F97A) also led to strong reductions in f_{FL}., Fig. 4c.

To determine the role of each individual β-helix coil, we further introduced paired F→A mutations in the hydrophobic core of the T129 construct, starting in the N-terminal coil and moving progressively towards the C-terminal coil. Paired mutations closer to the C-terminal end of T129 had stronger effects on f_{FL}, Fig. 4d, except when placed in the fifth coil (F122A+F127A), where they had no effect. Thus, the fifth coil

is not part of the β-helix formed in the T129 construct, presumably because it remains buried within the exit tunnel.

Finally, to better determine the end of the folded β-helix in the T129 construct, starting from its C-terminal end we replaced five amino acids at a time with alternating Gly and Ser residues. Replacement of up to 20 residues had little effect on f_{FL}, but when the C-terminal 25 residues were replaced there was a significant drop in f_{FL}, similar to the one seen for the F87A+F97A double mutant, Fig. 5 and Supplementary Fig. S2.

We conclude that the T129 construct contains a folded β-helix domain that extends from the N-terminal cap to the fourth coil, ending around residue 107±3. The onset of the folding transition is between constructs T118 and T122, Fig. 3, meaning that folding commences when residue 107 is approximately (120-107) + 21 = 34 residues away from the peptidyl-tRNA site (P-site) in the ribosome. This places the C-terminus of the folded β-helix domain in a similar location as we previously determined by cryo-EM for spectrin and titin domains tethered 33-35 residues away from the P-site (16, 22), *i.e.*, at the distal end of the exit tunnel. A model of the 107-residue PENT folding intermediate based on the cryo-EM structure of a ribosome-nascent chain complex with a folded titin domain (22) is shown in Fig. 6.

Discussion

Despite being obvious candidates to be cotranslationally folding proteins, the only β-helix protein for which there is experimental cotranslational folding data available is the phage P22 tailspike protein, a complex, trimeric protein with a central, 13-rung β-helix domain in the monomer. Using a panel of conformation-specific monoclonal antibodies to analyze the cotranslational folding of P22 tailspike, it was concluded

that more than ~3 but less than ~7 rungs of the β-helix must be exposed outside the ribosome for the protein to start to fold into a native-like structure (4).

Here, we have determined the structure and analyzed the cotranslational folding of the much simpler and more regular pentapeptide repeat β-helix protein PENT from *Clostridium botulinum.* The structure of PENT is broadly similar to that of MfpA (8), the founding member of the PRP family, except that MfpA lacks the N-terminal α-helix present in PENT. PENT is also structurally similar to the fourth luminal domain of the human synaptic vesicle protein 2C (SV2C), which is a receptor for botulinum neurotoxin serotypes A, D and F from *Clostridium botulinum* (23). The fact that PENT is present in this bacterium while being similar to the receptor for the neurotoxin it produces is interesting. The fold might hint at the function of the extracellular domain of SV2; it is possible that this domain "mimics" polynucleotides and interacts with DNA- and RNA-binding proteins.

Using an arrest-peptide assay that allows us to measure the force generated on the nascent chain by cotranslational folding of PENT, we detect a clear folding transition that involves PENT residues 1 to ~107, *i.e.*, approximately the first four coils of the eight-coil β-helix. The force on the nascent chain persists as the length of the nascent chain is increased by another ~20 residues, *i.e.*, the fifth β-helix coil also appears to contribute to the initial cooperative folding event. The subsequent appearance of the final three coils outside the exit tunnel does not generate appreciable force on the nascent chain, suggesting that they do not contribute to the cooperative folding transition but are only added residue-by-residue to the already nucleated β-helix. It is also clear from the force-profile analysis that the β-helix starts to fold when the C-

terminal end of the fourth coil is ~34 residues away from the P-site, right in the mouth of the exit tunnel.

Pentapeptide repeat proteins are prone to aggregation when folded *in vitro* (9), and cotranslational folding of the β-helix as it emerges from the exit tunnel would minimize the risk of aggregation *in vivo*. By sequestering the C-terminal end of the growing β-helix in the exit tunnel, the ribosome may further protect the exposed hydrophobic surface at this end of the β-helix until the C-terminal α-helix finally caps the monomer (21).

Given the difficulties of mimicking the cotranslational folding of repeat proteins *in vitro*, we believe that arrest-peptide based force-profile analysis will provide a good alternative to the further study of this important class of proteins.

Materials and Methods

Enzymes and chemicals. All restriction and DNA modifying enzymes were purchased from Thermo Scientific (Waltham, MA, USA) and New England Biolabs (Ipswich, MA, USA). Oligonucleotides for cloning and mutagenesis were obtained from MWG Eurofins (Ebersberg, Germany). DNA/RNA purification kits were from Thermo Scientific (Waltham, MA, USA) and the PUREexpress™ cell-free coupled transcription-translation system was from New England Biolabs (Ipswich, MA, USA). [^{35}S]-Methionine was purchased from PerkinElmer (Waltham, MA, USA). All other reagents were from Sigma-Aldrich (St. Louis, MO, USA).

DNA manipulations. All PENT constructs were obtained from a previously described pET19b-derived plasmid (Novagen, Madison, WI, USA) containing a soluble, non-membrane targeted LepB derivative of the *lepB* gene, the small zinc finger domain ADR1a, a 21 residues long linker, the *E. coli* SecM arrest peptide, FSTPVWISQAQGIRAGP, and a 23-residue C-terminal tail under the control of a T7 promoter (18). The GSGS-flanked ADR1a gene was replaced with the 217-residue-long *M. tuberculosis* PENT gene, using the megaprimer approach. C-terminal truncations of PENT (Supplementary Table S1) were generated by PCR using partially overlapping primers, as previously described (12).Site-directed mutagenesis was performed to generate constructs with the non-functional FSTPVWISQAQGIRAGA SecM AP as full-length controls for all constructs, to replace the Pro at the end of the AP with a stop codon as arrest controls for all constructs, and on the truncated PENT constructs T129 and T140 to introduce Phe→Ala mutations in the PENT hydrophobic core. Cumulative mutations were introduced by site-directed mutagenesis using partially overlapping primers; mutations of paired Phe residues in the hydrophobic core of PENT were introduced

by Gibson assembly (24) using synthetic gene fragments encompassing the mutations (GeneArt Strings, Thermo Scientific, Waltham, MA, USA). All constructs were verified by DNA sequencing.

Expression in vitro. In vitro coupled transcription-translation was performed in PURExpress™ (New England Biolabs, Ipswich, MA, USA), according to the manufacturer's instructions, using PCR products as templates for the generation of truncated nascent chains. Briefly, 1 μl (50 ng) of PCR template and 10 μCi (1 μl) [^{35}S]-Methionine were added to a final volume of 10 μl reaction components. Transcription-translation was carried out for 20 min. at 750 rpm in a bench-top tube shaker at 37°C. Reactions were stopped by adding equal volumes of ice-cold 30% trichloacetic acid, incubated on ice for 30 min. and then centrifuged at 14,000 rpm (20,800 x g*)* at 4°C for 10 min. in an Eppendorf centrifuge. Pellets were resuspended in 20 μl of SDS Sample Buffer and treated with RNase A (400 μg ml^{-1}) for 15 min. at 37 °C. Proteins were separated by SDS-PAGE, visualized on a Fuji FLA-9000 phosphoimager, and quantified using ImageGauge software V4.23 (FujiFilm Corporation). Analysis of quantified bands was performed using EasyQuant (in-house developed quantification software). Values of f_{FL} were calculated as $f_{FL} = I_{FL}/(I_{FL} + I_A)$, where I_{FL} is the intensity of the band corresponding to the full-length protein, and I_A is the intensity of the band corresponding to the arrested form of the protein. Experiments were repeated at least three times, and SEMs were calculated.

Protein expression and purification. PENT was expressed in *E. coli* BL21(DE3) T1R pRARE2, grown in TB supplemented with 8 g/l glycerol and 0.4% glucose, with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol at 37°C. After OD_{600} of 2 was reached the temperature was lowered to 18°C and expression was induced with 0.5 mM of IPTG. Protein expression continued overnight, and cells were harvested by

centrifugation (10 min. at 4,500 g) the next morning. Cell pellets were resuspended 100 mM HEPES pH 8.0, 500 mM NaCl, 10% glycerol, 10 mM imidazole, 0.5 mM TCEP. A tablet of Complete EDTA-free protease inhibitor cocktail (Roche) and 5 μl/ml benzonase nuclease (250 U, Sigma) were added to the resuspension, and cells were lysed by pulsed sonication (4 s on/4 s off for 3 min., 80% amplitude in a Vibra-Cell Sonics sonicator). The lysate was centrifuged for 20 min. at 49,000 g and the supernatant was collected and filtered through a 0.45 μm pore size filter.

A 2 ml HisTrap HP column (GE Healthcare) was used for the first purification step. After running the lysate through, the column was washed with IMAC buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.5 mM TCEP) containing 10 mM imidazole. A second wash was performed with IMAC buffer containing 50 mM imidazole, and the protein was eluted with IMAC buffer containing 500 mM imidazole. The elution fraction was loaded onto a HiLoad 16/60 Superdex 75 column (GE Healthcare) as a second purification step. Fractions were examined on an SDS-PAGE gel, and TCEP was added up to 2 mM concentration before pooling and concentrating with a centrifugal concentrator.

Crystallization and structure determination. PENT crystals were found in the A6 condition of the PACT premier screen (Molecular Dimensions). In order to optimize these, 150 nl of protein solution was mixed with 50 nl of reservoir solution (0.1 M SPG buffer pH 9.5, 25% PEG 1500) in a sitting-drop crystallization plate. Long thin rod-shaped crystals grew overnight, and were flash frozen in liquid nitrogen after transfer to a drop of cryo-protecting solution (0.1 M SPG buffer pH 9.5, 6.6 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 25% w/v PEG 1500). X-ray diffraction data collection was performed at beamline i03, Diamond Light Source (Oxford, UK). The data were processed using DIALS (25), molecular replacement was performed

using Phaser (26) with a model provided by the Phyre^2 server (27). The model was built using Phenix Autobuild (28) and Coot (29), and the structure was refined using Refmac5 (30). Data processing and refinement statistics are presented in Supplementary Table 1. Coordinates for the PENT structure have been deposited in the PDB with the accession code 6FLS.

Acknowledgements

We thank the beamline scientists at BESSY, Berlin, ESRF, Grenoble, Max-Lab, Lund and SLS, Villigen for their support in data collection and PSF for protein production.

This work was supported by grants from the Knut and Alice Wallenberg Foundation, the Swedish Cancer Foundation, and the Swedish Research Council to GvH, and by the Swedish Cancer Foundation and the Swedish Research Council (2014-5667) to PS.

Author contributions

GvH and PS conceived the project. LN and MMC designed and performed the experiments. LN, MMC, PS, and GvH wrote the manuscript.

References

1. Kloss E, Courtemanche N, & Barrick D (2008) Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch Biochem Biophys* 469:83-99.
2. Main ER, Jackson SE, & Regan L (2003) The folding and design of repeat proteins: reaching a consensus. *Curr Opin Struct Biol* 13:482-489.
3. Nissley DA*, et al.* (2016) Accurate prediction of cellular co-translational folding indicates proteins can switch from post- to co-translational folding. *Nat Commun* 7:10341.

4. Evans MS, Sander IM, & Clark PL (2008) Cotranslational folding promotes β-helix formation and avoids aggregation *in vivo*. *J Mol Biol* 383:683-692.

5. Lee W*, et al.* (2010) Full reconstruction of a vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations. *J Biol Chem* 285:38167-38172.

6. Vetting MW*, et al.* (2006) Pentapeptide repeat proteins. *Biochemistry* 45:1-10.

7. Shah S & Heddle JG (2014) Squaring up to DNA: pentapeptide repeat proteins and DNA mimicry. *Appl Microbiol Biotechnol* 98:9545-9560.

8. Hegde SS*, et al.* (2005) A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA. *Science* 308:1480-1483.

9. Khrapunov S, Cheng H, Hegde S, Blanchard J, & Brenowitz M (2008) Solution structure and refolding of the *Mycobacterium tuberculosis* pentapeptide repeat protein MfpA. *J Biol Chem* 283:36290-36299.

10. Ito K, Chiba S, & Pogliano K (2010) Divergent stalling sequences sense and control cellular physiology. *Biochem Biophys Res Comm* 393:1-5.

11. Butkus ME, Prundeanu LB, & Oliver DB (2003) Translocon "pulling" of nascent SecM controls the duration of its translational pause and secretion-responsive secA regulation. *J Bacteriol* 185:6719-6722.

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

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

14. Cymer F & von Heijne G (2013) Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proc Natl Acad Sci U S A* 110:14640-14645.

15. Ismail N, Hedman R, Lindén M, & von Heijne G (2015) Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat Struct Mol Biol* 22:145-149.

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

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

18. Nilsson OB*, et al.* (2015) Cotranslational protein folding inside the ribosome exit tunnel. *Cell reports* 12:1533-1540.

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

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

21. Bryan AW, Jr., Starner-Kreinbrink JL, Hosur R, Clark PL, & Berger B (2011) Structure-based prediction reveals capping motifs that inhibit beta-helix aggregation. *Proc Natl Acad Sci U S A* 108:11099-11104.

22. Tian P*, et al.* (2018) The folding pathway of an Ig domain is conserved on and off the ribosome. *BioRxiv* doi: https://doi.org/10.1101/253013.

23. Benoit RM*, et al.* (2014) Structural basis for recognition of synaptic vesicle protein 2C by botulinum neurotoxin A. *Nature* 505:108-111.

24. Gibson DG, *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343-345.

25. Waterman DG, *et al.* (2016) Diffraction-geometry refinement in the DIALS framework. *Acta Crystallogr D Struct Biol* 72:558-575.

26. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.

27. Kelley LA, Mezulis S, Yates CM, Wass MN, & Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845-858.

28. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213-221.

29. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126-2132.

30. Murshudov GN, *et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355-367.

31. Goddard TD, *et al.* (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 27:14-25.

Figure legends

Figure 1. Structural representation of the *C. botulinum* PENT protein. (a) Crystal structure of the PENT dimer. Chains are colored in rainbow from N terminus (blue) to C terminus (red). (b) View along the central axis of the PENT monomer, highlighting the Phe residues in the hydrophobic core. N-terminal residues 1-30 have been removed for clarity.

Figure 2. Arrest-peptide based force-measurement assay. (a) Schematic scenario for constructs generating ($F > 0$) or not generating ($F \approx 0$) pulling force depending on the location of the PENT domain relative to the arrest peptide. (b) The force generating PENT domain is placed 21 residues upstream of the critical C-terminal proline of the SecM arrest peptide. A 153-residue long domain from the *E.coli* LepB protein is included in all constructs (except for the full-length PENT construct), and a 23-residue C-terminal tail (also from LepB) is appended at the C terminus. The hatched area highlights the β-helix fold within the full-length construct. The lengths of the different parts of the construct are indicated. (c) SDS-PAGE gels showing full-length (FL, black arrowhead) and arrested (A, white arrowhead) species for the PENT T30 and T129 constructs. Lanes 2 and 4 show control constructs with a Pro→Ala mutation at the C-terminal end of the AP that prevents translational arrest and serve as markers for the full-length forms of the protein. Calculated f_{FL} values are shown below the gels.

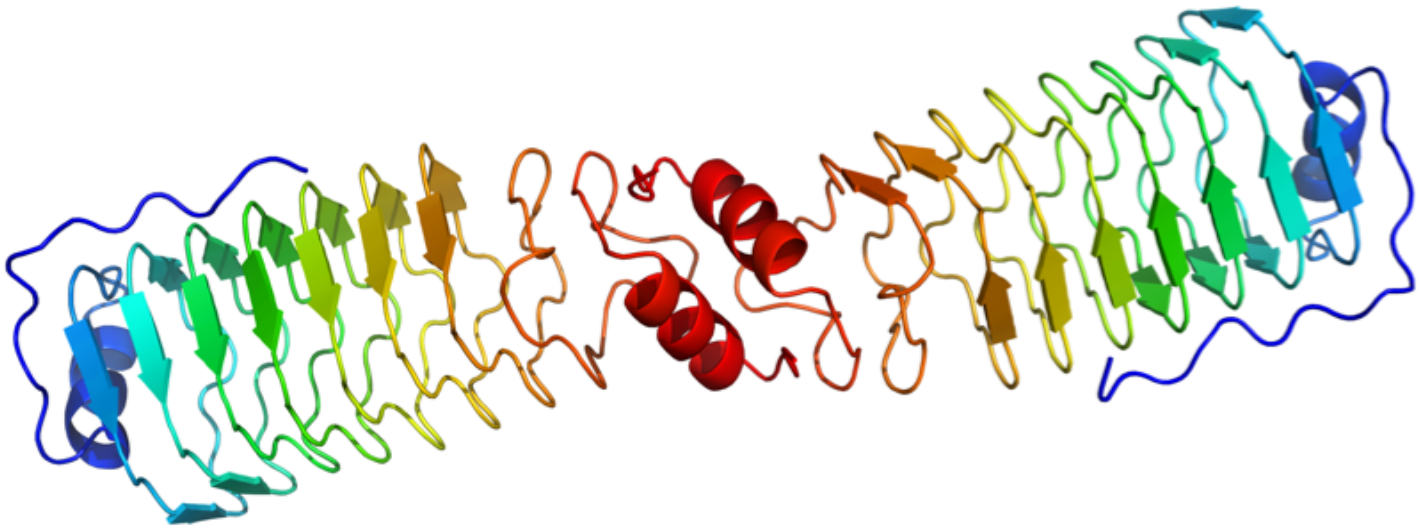
Figure 3. Force profile for the full set of C-terminal truncations of PENT. Three peaks discussed in the text are indicated. Truncations are made from the C-terminal end of PENT, as indicated by the arrow in the cartoon. The sequence of the SecM AP is shown in red.

Figure 4. Effects of mutations in the hydrophobic core of PENT T129. (a) The N-terminal cap and 5 N-terminal β-helix coils of PENT (residues 1-129; residues 108-129 are in grey), highlighting the hydrophobic core mutations analyzed: Val15, Phe18, 47, 57, 67, 77, 87, 97, 122, and 127. All amino acids were mutated to Ala in a cumulative or paired fashion. (b) Effect of the double mutant V15A+F18A in the N-terminal cap. (c) Effects of cumulative Phe→Ala mutations (as indicated) in the hydrophobic core of PENT. (d) Effects of paired Phe→Ala mutations in the hydrophobic core of PENT. All experiments were repeated at least 3 times; averages ± SE are shown. N.S. not significant, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. Note that the average f_{FL} value for the T129 construct (included as a control in all experiments) is somewhat variable, due to different batches of PURE being used for panels b-d.

Figure 5. Effect of cumulative replacement of amino acids at the C-terminal end of PENT T129. (a) Schematic representation of the replacement strategy with alternating Gly and Ser residues, five at a time. (b) Effects of the cumulative replacements. All experiments were repeated at least 3 times; averages ± SE are shown. ** $p \leq 0.01$.

Figure 6. Model for the PENT 1-107 folding intermediate at the lateral end of the exit tunnel based on the cryo-EM structure of a ribosome-nascent chain complex with a folded titin I27 domain and a 35-residue tether (22). The PENT domain (orange) was overlapped with the I27 domain (which is of very similar size and shape, not shown), with its C-terminal end pointing into the exit tunnel. An *E. coli* 70S ribosome structure (PDB 3JBU) was fit into the cryo-EM density and rendered in surface mode (grey), using ChimeraX (31).

a



b

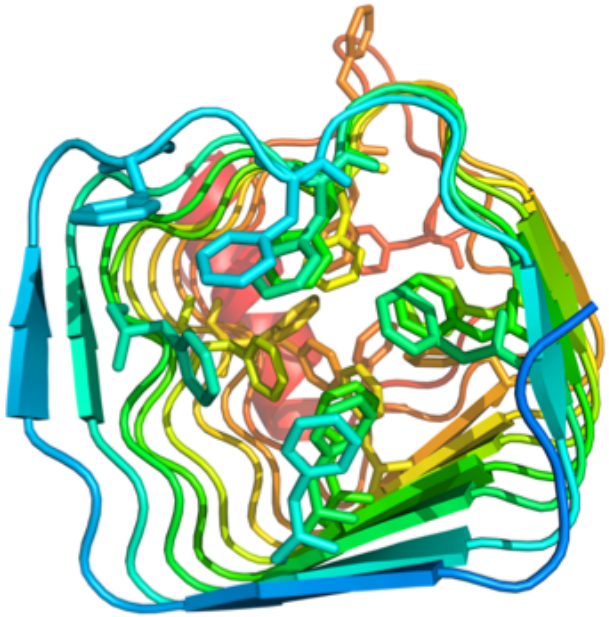


Figure 1

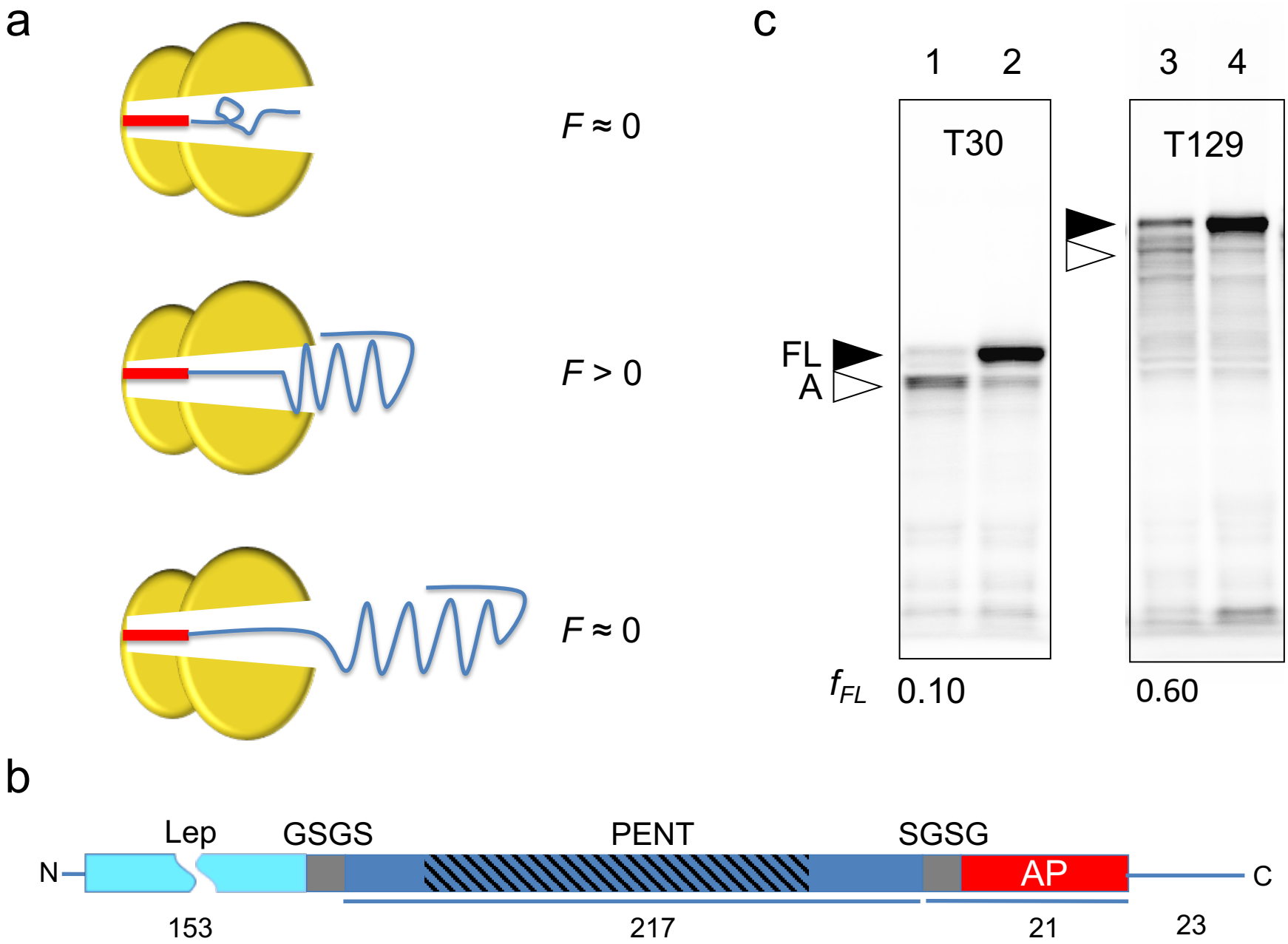

a
F ≈ 0
F > 0
F ≈ 0
c
1
2
3
4
T30
T129
FL
A
f_{FL}
0.10
0.60
b
Lep
GSGS
PENT
SGSG
N
AP
C
153
217
21
23

Figure 2

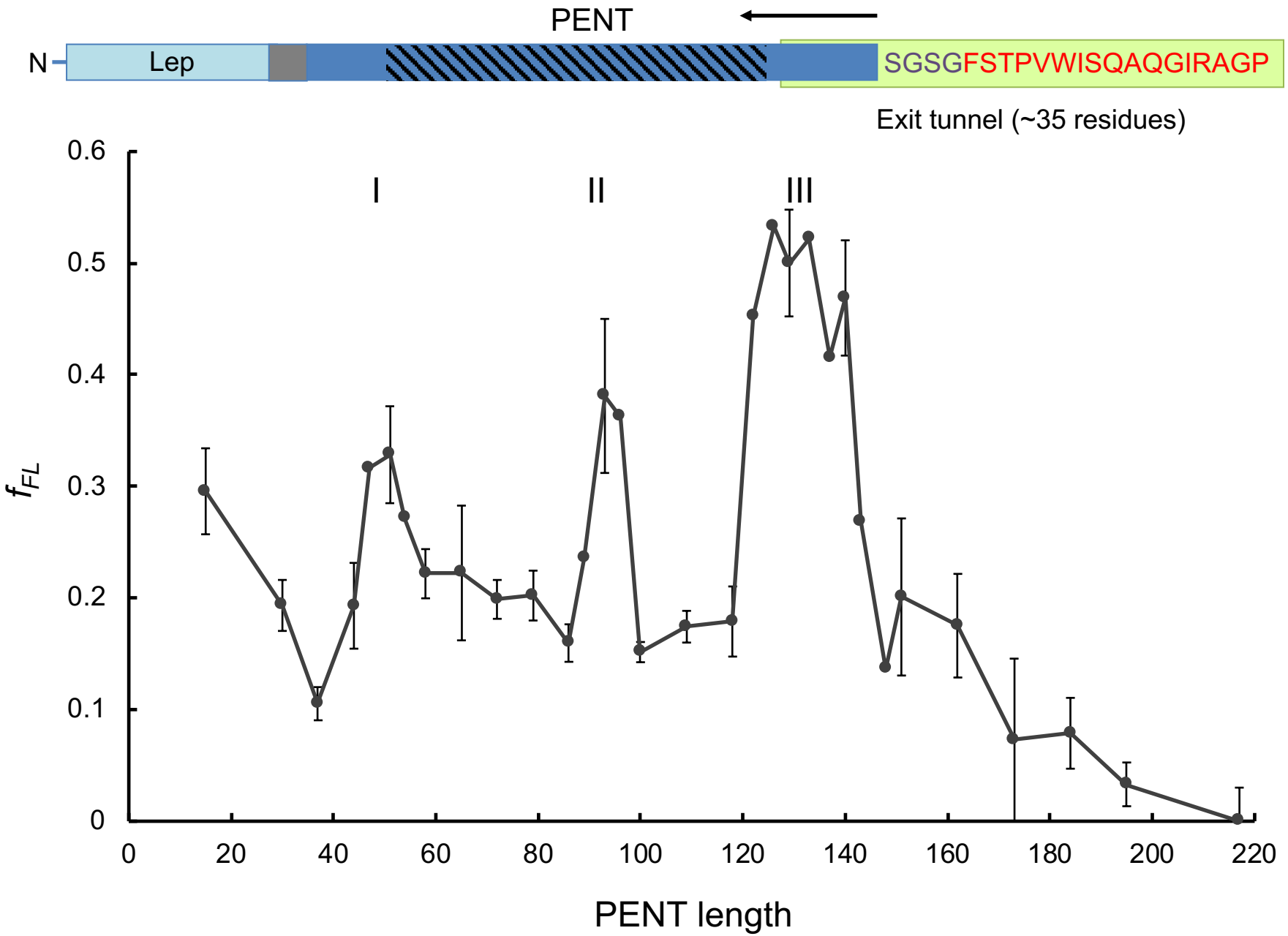

PENT
N
Lep
SGSGFSTPVWISQAQGIRAGP
Exit tunnel (~35 residues)
I
II
III
f_{FL}
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120
140
160
180
200
220
PENT length

Figure 3

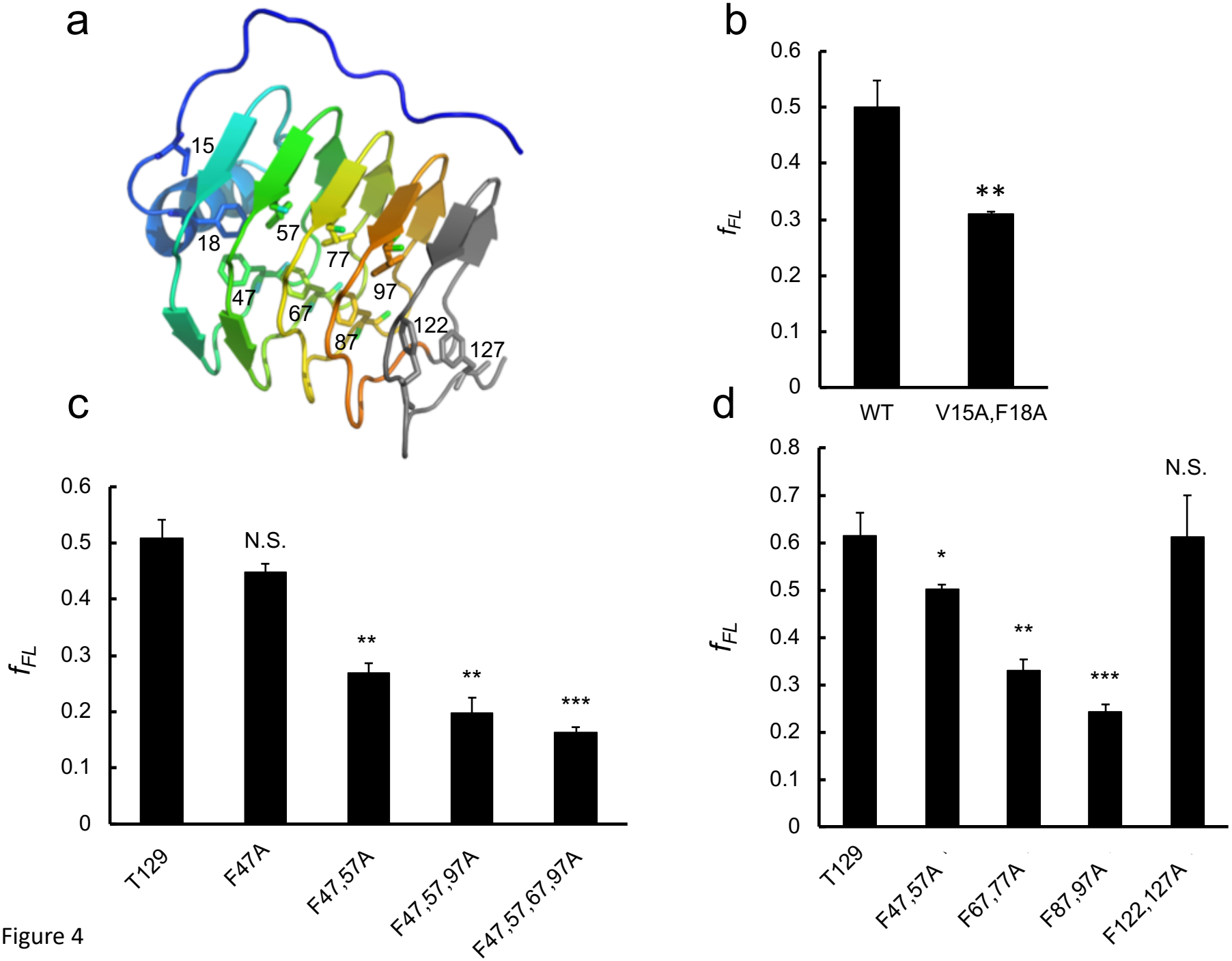

a
15
18
57
77
47
97
67
122
87
127
b
0.6
0.5
0.4
0.3
0.2
0.1
0
f_{FL}
**
WT
V15A,F18A
c
0.6
0.5
0.4
0.3
0.2
0.1
0
f_{FL}
N.S.
**
**
***
T129
F47A
F47,57A
F47,57,97A
F47,57,67,97A
d
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
f_{FL}
N.S.
*
**
***
T129
F47,57A
F67,77A
F87,97A
F122,127A

Figure 4

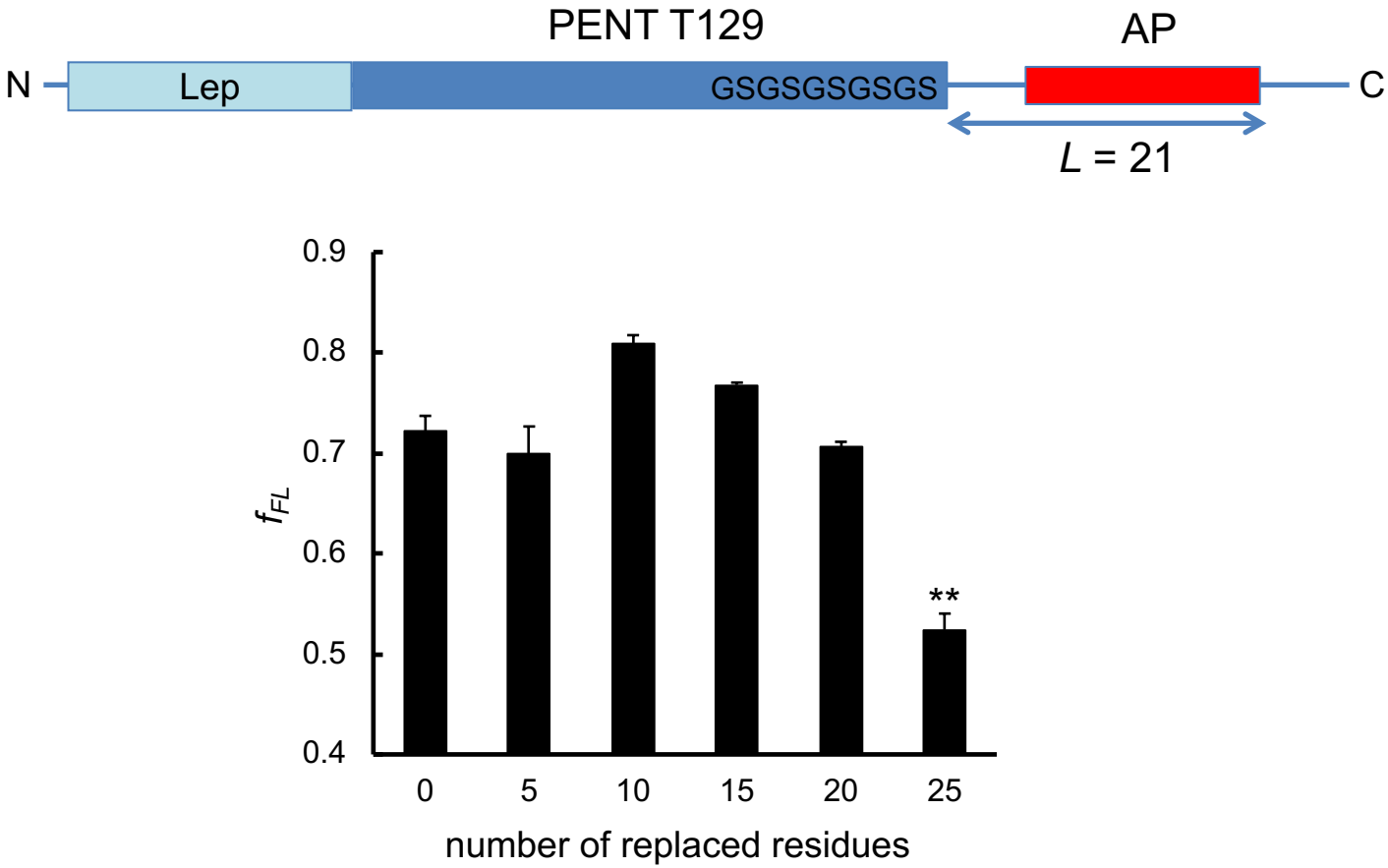

PENT T129
AP
N
Lep
GSGSGSGSGS
C
L = 21
0.9
0.8
0.7
0.6
0.5
0.4
fFL
**
0
5
10
15
20
25
number of replaced residues

Figure 5

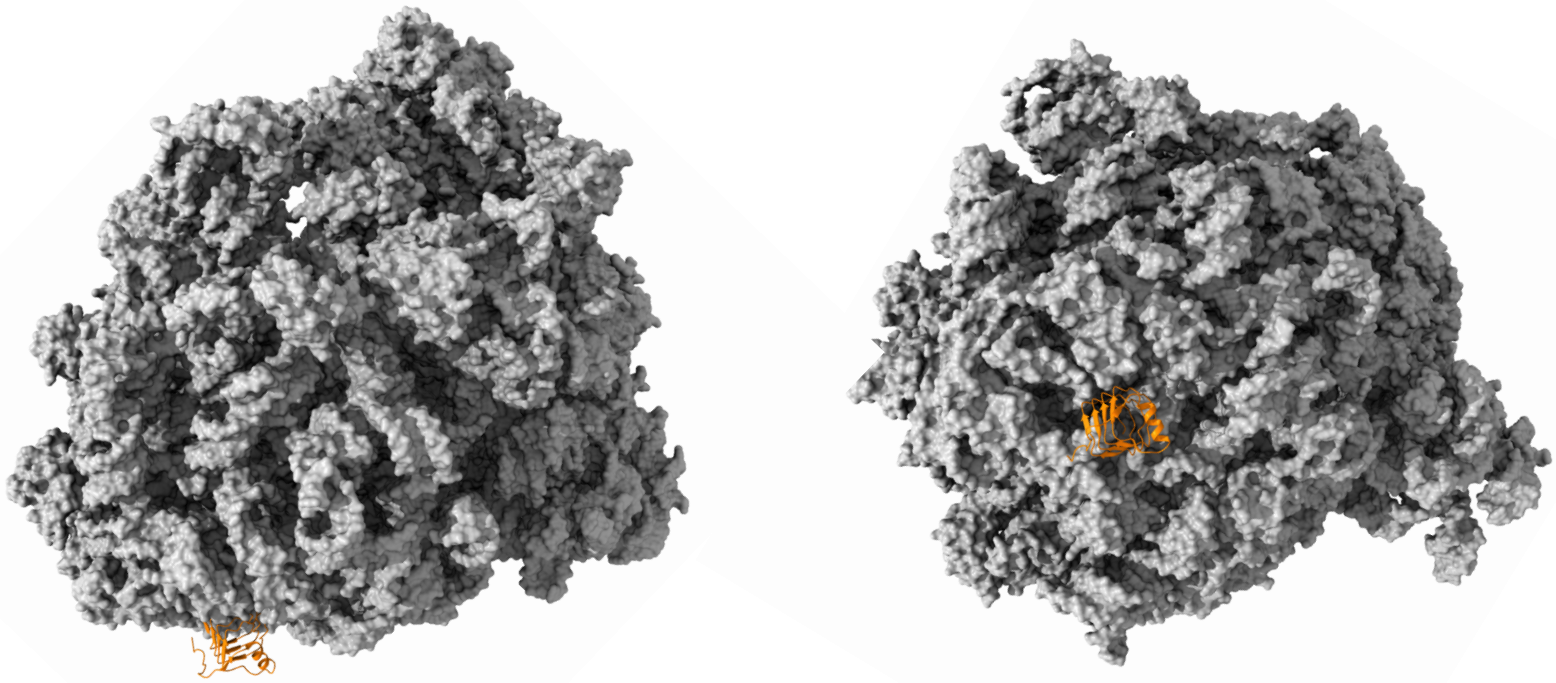


Figure 6

Supplementary Table 1

Data processing and refinement statistics for PENT (PDB access code 6FLS).

Space group	P 4_1
Unit cell parameters *a, b, c* (Å)	107.3, 107.3, 110.3
Resolution (Å)	3
$I/\sigma(I)$	6.42 (0.26)
Completeness (%)	100 (99.4)
Multiplicity	7.4 (7.5)
R_{merge}	0.148 (4.678)
R_{cryst} (%)	21.0
R_{free} (%)	24.0
RMSD in bond length (Å)	0.024
RMSD in bond angle (Å)	2.67
Ramachandran plot (%)	
In preferred regions	89.93
In allowed regions	7.96
Outliers	2.11

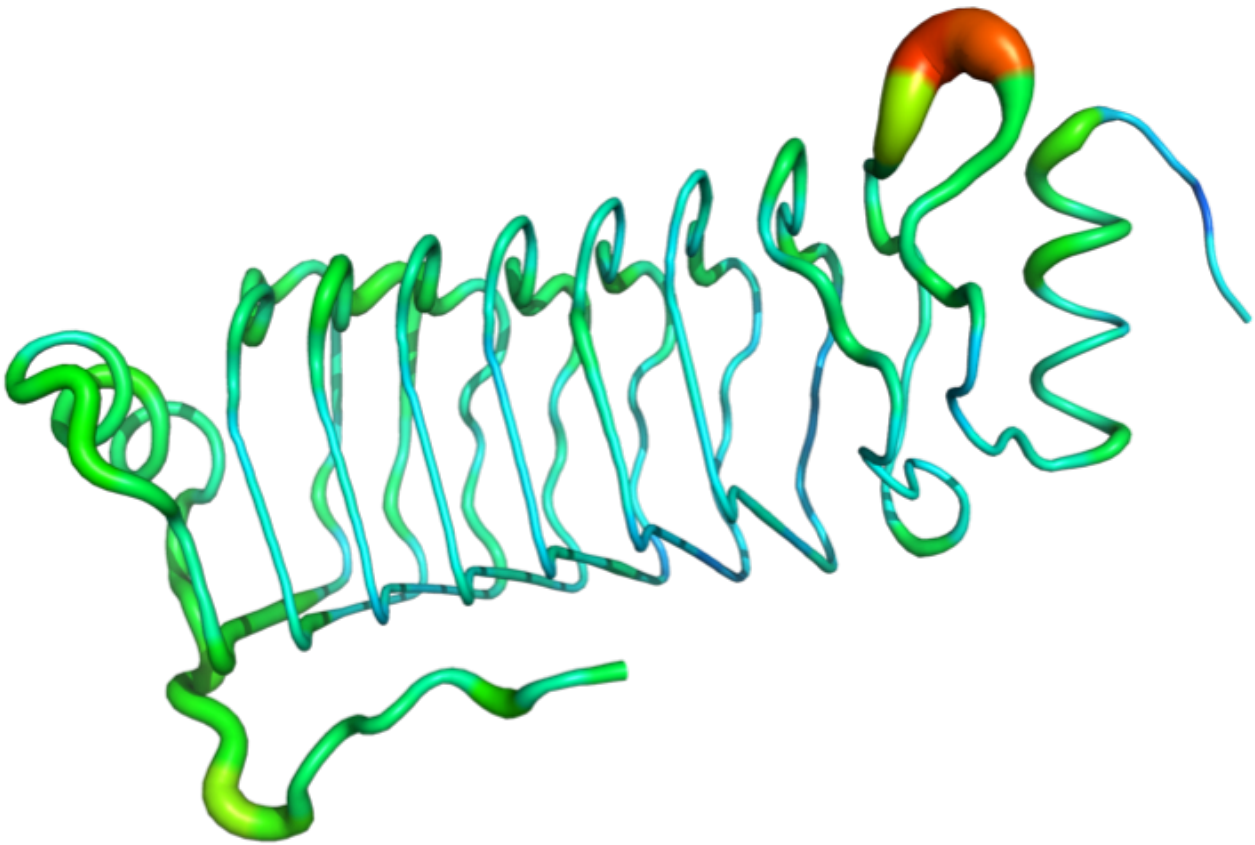
Supplementary Table 2

Amino acid sequences of full-length [Lep154] – PENT – SecM – C-terminal tail and of the C-terminal PENT truncations analysed in the paper. Note that the N-terminal Lep-part is not present in the full-length PENT 217 construct.

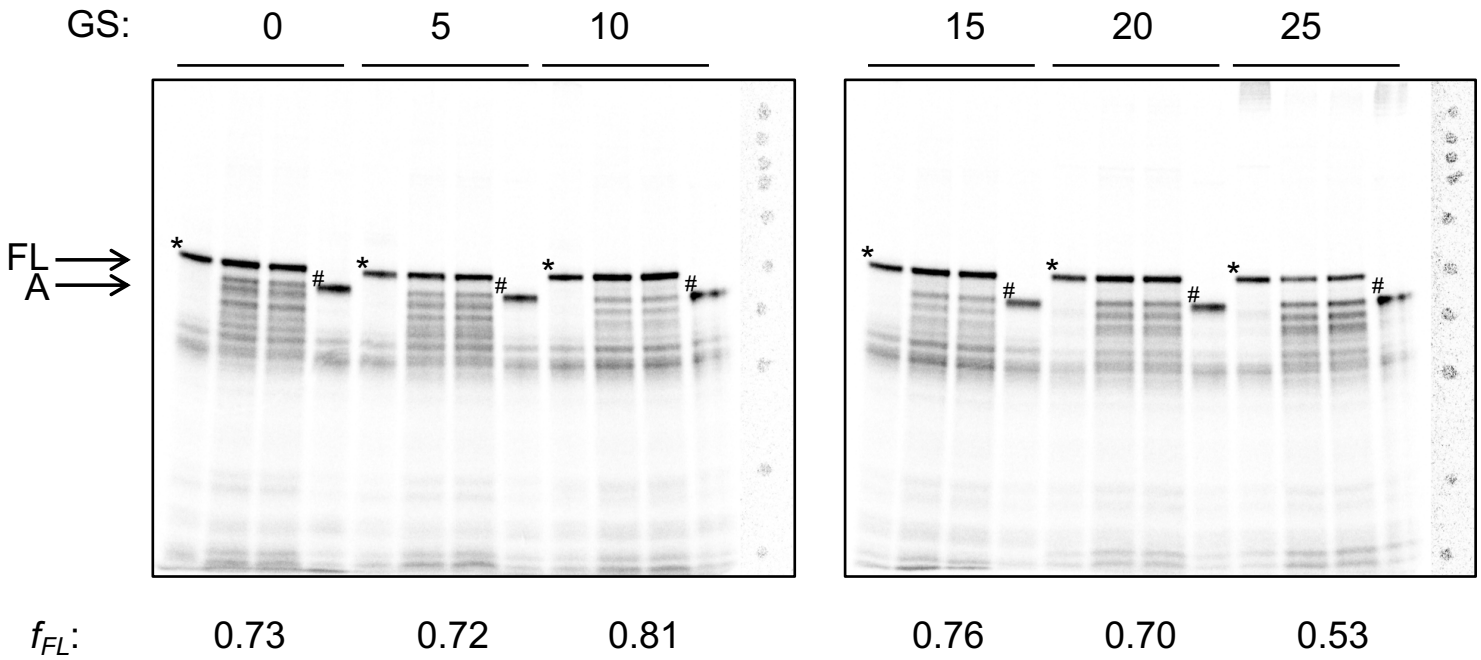
PENT 217 (full-length)	MSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDIDFTEVSITEAKLKRFKAKNSHFIKNNFFKTMLTGVDFTKNELVAPTVSSPPIEFQGAKISMVQAADLIGLWGIIVEQSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T195	MANRSFIYEPFQIPSGSMMPTLNSTDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLPGDKVTYDPVSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQTFSRRNGGEATSGFFEVPKQETKENGIRLSETSGSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDIDFTEVSITEAKLKRFKAKNSHFIKNNFFKTMLTGVDFTKNELVAPTVSSPPIEFQSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T184	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDIDFTEVSITEAKLKRFKAKNSHFIKNNFFKTMLTGVDFTKNELVASGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T173	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDIDFTEVSITEAKLKRFKAKNSHFIKNNFFKTMLTSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T162	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDIDFTEVSITEAKLKRFKAKNSHFSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T151	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDIDFTEVSITEAKSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T148	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDIDFTEVSITSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T143	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDIDFTSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T140	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFEDISGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T137	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTDIAFSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T133	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKAKMTSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH

T129	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYFDKSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T126	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFQYSYSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T122	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQRSNFSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T118	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDTIFKQTSMQSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T109	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACKCVGVNMIDSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T100	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQFIACSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T96	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFERCQSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T93	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDAYFESGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T89	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNFTDSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T86	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNCNLSNSNSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T79	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFVDVVFQNSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T72	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSFEGASFSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T65	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFENCTFLDCSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T58	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSIFESGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T54	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSEINTSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T51	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSKSESGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T47	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYFDFSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH

T44	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKIYGEHIEYSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T37	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEIPITQIKISGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T30	[Lep154]GSGSMSISNPRIPADLIMVDDFSSYAQGYLYEEISGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH
T15	[Lep154]GSGSMSISNPRIPADLIMVSGSGFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH



Supplementary Figure S1. Cartoon putty representation of chain A in the crystal structure of PENT, coloured by B-factors (low B-factors represented in blue and narrow radius, high B-factors in red and wide radius).



Supplementary Figure S2. SDS-PAGE analysis of PENT T129 constructs with 5-25 C-terminal residues of the PENT part replaced with alternating Gly-Ser residues. Two independent translation reactions are shown for each construct (middle lanes), and the reported f_{FL} values are averages for the two lanes. A full-length control (*) where the AP has been inactivated by a Pro-to-Ala mutation and an arrest control (#) with a stop codon replacing the Pro residue at the end of the AP are included for each construct. The ladder of bands running immediately below the arrested (A) form of the protein are most likely caused by ribosome stacking.