Supplementary information

Supplementary Tables 1-5

www.pnas.org/cgi/doi/10.1073/pnas.1812756115

Supplementary Table 1. Proteins tested in the study. (#) EHEE is an artificial protein for which only a theoretical 3D model is available. ($) Highlighted in green are the two ILBP proteins used in this study, for which we made homology models based on PDB 2LBA. Highlighted in gray are S6 and four of its variants; there are no structures available for the charge mutants.

Protein	**Reference**	**PDB ID**	**Length (residues)**	**Molecular weight (kDa)**	**Net charge**
FSD1	(1)	1fsd	28	3.48	+5
WW domain	(2)	2m8i	35	4.08	+4
EHEE	(3)	model[#]	40	4.94	+1
Protein G	(4)	3gb1	56	6.1	-4
Cam Frag	(5)	2hf5	69	7.83	-9
SOD1	(6)	4bcz	110	11.10	-3
ILBP	(7)	model[$]	128	14.40	+1
ILBP-tm	(7)	model[$]	105	11.51	+2
S6-wt-(+16/-16)	(8, 9)	2kjv	101	11.97	0
S6-(+16/-9)	(8, 9)	----	101	11.96	+7
S6-(0)	(8, 9)	----	101	10.96	0
S6-(+9/-16)	(8, 9)	----	101	11.54	-7
S6-(-16)	(8, 9)	----	101	10.97	-16

Supplementary Table 2. Sequences of the longest construct tested for each protein domain (orange). Shorter constructs were obtained by shortening of the $(GS)_n$ linker (purple) For small domains, an unstructured segment from the *E. coli* LepB protein (light blue) was appended to the N terminus in order to make the constructs easy to identify by SDS-PAGE. The arrest peptide (green) is followed by a 23-residue C-terminal tail (light blue), also derived from the LepB protein.

>Lep-154-1fsd-SecM-L60 (Not listed in Uniprot, artificial protein, PDB;1fsd)
MANRSFIYEPFQIPSGSMMPTLNSTDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDP
KLDYIKRAVGLPGDKVTYDPVSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQTFSRRNGGEAT
SGFFEVPKQETKENGIRLSETSGSGSQQYTAKIKGRTFRNEKELRDFIEKFKGRGSGSGSGSGSGS
GSGGSGSGSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSR
IGGIH*

Number of amino acids: 28
Molecular weight: 3.48 kDa
Arg+Lys=4+5=9
Glu+Asp=3+1=4
Net charge at pH 7:+5

>WW domain UniProtKB - Q13526 (PIN1_HUMAN)
MANRSFIYEPFQIPSGSMMPTLNSTDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDP
KLDYIKRAVGLPGDKVTYDPVSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQTFSRRNGGEAT
SGFFEVPKQETKENGIRLSETSGSGSKLPPGWEKRMSRSSGRVYYFNHITNASQWERPSGGSGSGS
GSGSGSGSGGSGSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGEWPT
GLRLSRIGGIH*

Number of amino acids: 35
Molecular weight: 4.08 kDa
Arg+Lys=4+2=6
Glu+Asp=2+0=2
Net charge at pH 7:+4

>EHEE (Not listed in Uniprot, artificial protein,(3))
MANRSFIYEPFQIPSGSMMPTLNSTDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDP
KLDYIKRAVGLPGDKVTYDPVSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQTFSRRNGGEAT
SGFFEVPKQETKENGIRLSETSGSGSTTRYRFTDEEEARRAAKEWARRGYQVHVTQNGTYWEVEVR
GSGSGSGSGSGSGSGGSGSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQ
EGEWPTGLRLSRIGGIH*

Number of amino acids: 40
Molecular weight: 4.94 kDa
Arg+Lys=7+1=8
Glu+Asp=6+1=7
Net charge at pH 7:+1

>Protein G UniProtKB - P06654 (SPG1_STRSG)
MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGSGS
GSGSGSGSGSGGSGSGSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSD
KQEGEWPTGLRLSRIGGIH*

Number of amino acids: 56
Molecular weight: 6.1 kDa
Arg+Lys=0+6
Glu+Asp=5+5
Net charge at pH 7:-4

>CamFrag UniProtKB - G3V361 (G3V361_HUMAN)
MAELQDMINEVDADGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAE
LRHVMTNLGSGSGSGSGSGSGSGSGSGSGSGGSGSGSGSGSGSGSGSGSGFSTPVWIS
QAQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH*

Number of amino acids: 69
Molecular weight: 7.83 kDa
Arg+Lys=4+3=7
Glu+Asp=8+8=16
Net charge at pH 7:-9

>SOD1 UniProtKB - P00441 (SODC_HUMAN)
MATKAVAVLKGDGPVQGIINFEQKESNGPVKVWGSIKGLTEGLHGFHVHGAGGDLGNVTA
DKDGVADVSIEDSVISLSGDHSIIGRTLVVHEKAGAGAGSRLASGVIGIAQMMGSGSGSG
SGSGSGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGE
WPTGLRLSRIGGIH*

Number of amino acids: 110
Molecular weight: 11.10 kDa
Arg+Lys=2+7=9
Glu+Asp=5+7=12
Net charge at pH 7:-3

>ILBP UniProtKB - P50119 (FABP6_RABIT)
MAFTGKFEMESEKNYDEFMKLLGLPSDVVEKSRNIKIVTEIKQDGQDFTWSHHYSGGQIM
TNKFTIGKESEIQTFGGKKFKAVVNMEGGKVVANFPNYQHTSEIKGDKLVEVSSIGGVTY
ERVSKRLAGSGSGSGSGSGSGSGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQ
AQGIRAGPGSSDKQEGEWPTGLRLSRIGGIH*

Number of amino acids: 128
Molecular weight: 14.40 kDa
Arg+Lys=3+15=18
Glu+Asp=12+5=17
Net charge at pH 7:+1

>ILBP-tm UniProtKB - P50119 (FABP6_RABIT) mutant
MAFTGKFEMGGSGIVTEIKQDGQDFTWSHHYSGGQIMTNKFTIGKESEIQTFGGKKFKAV
VNMEGGKVVANFPNYQHTSEIKGDKLVEVSSIGGVTYERVSKRLAGSGSGSGSGSGSGSG
GSGSGSGSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLR
LSRIGGIH*

Number of amino acids: 105
Molecular weight: 11.51 kDa
Arg+Lys=2+11=13
Glu+Asp=8+3=11
Net charge at pH 7:+2

>S6-SecM(*Ec*) AP UniProtKB - P23370 (RS6_THETH)
MRRYEVNIVLNPNLDQSQLALEKEIIQRALENYGARVEKVEELGLRRLAYPIAKDPQGYF
LWYQVEMPEDRVNDLARELRIRDNVRRVMVVKSQEPFLANAGSGSGSGSGSGSGSGGSGS
GSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRI
GGIH*

Number of amino acids: 101
Molecular weight: 11972.77
Arg+Lys=12+4=16
Glu+Asp=11+5=16
Net charge at pH 7:0

>S6^{+16-9}-SecM(*Ec*) AP UniProtKB - P23370 (RS6_THETH) mutant
MRRYEVNIVLNPNLDQSQLALEKQIIQRALENYGARVQKVQELGLRRLAYPIAKDPQGYF
LWYQVEMPQNRVNDLARQLRIRNNVRRVMVVKSQEPFLANAGSGSGSGSGSGSGSGGSGS
GSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRI
GGIH*

Number of amino acids: 101
Molecular weight: 11965.88
Arg+Lys=12+4=16
Glu+Asp=6+3=9
Net charge at pH 7:+7

>S6^{0}-SecM(*Ec*) AP UniProtKB - P23370 (RS6_THETH) mutant
MSSYQVNIVLNPNLNQSQLALQSQIIQSALQNYGASVQSVQQLGLSSLAYPIASNPQGYF
LWYQVQMPQNSVNNLASQLSISNNVSSVMVVSSQQPFLANAGSGSGSGSGSGSGSGGSGS
GSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRI
GGIH*

Number of amino acids: 101
Molecular weight: 10963.32
Arg+Lys=0+0=0
Glu+Asp=0+0=0
Net charge at pH 7:0

> S6^{+9-16}-SecM(*Ec*) AP UniProtKB - P23370 (RS6_THETH) mutant

MSRYEVNIVLNPNLDQSQLALESEIIQRALENYGARVESVEELGLSRLAYPIAKDPQGYF
LWYQVEMPEDRVNDLASELRISDNVRSVMVVKSQEPFLANAGSGSGSGSGSGSGSGGSGS
GSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRI
GGIH*

Number of amino acids: 101
Molecular weight: 11545.03
Arg+Lys=7+2=0
Glu+Asp=11+5=16
Net charge at pH 7: -7

>S6^{-16}-SecM(*Ec*) AP UniProtKB - P23370 (RS6_THETH) mutant
MSSYEVNIVLNPNLDQSQLALESEIIQSALENYGASVESVEELGLSSLAYPIASDPQGYF
LWYQVEMPEDSVNDLASELSISDNVSSVMVVSSQEPFLANAGSGSGSGSGSGSGSGGSGS
GSGSGSGSGSGSGSGSGSGSGSGSFSTPVWISQAQGIRAGPGSSDKQEGEWPTGLRLSRI
GGIH*

Number of amino acids: 101
Molecular weight: 10979.08
Arg+Lys=0+0=0
Glu+Asp=11+5=0
Net charge at pH 7: -16

>S6-SecM(*Ms*) AP UniProtKB - P23370 (RS6_THETH) mutant
MRRYEVNIVLNPNLDQSQLALEKEIIQRALENYGARVEKVEELGLRRLAYPIAKDPQGYF
LWYQVEMPEDRVNDLARELRIRDNVRRVMVVKSQESGSGSGSGSGSGSGSGSGSGSGG
SGSGSGSGSGSGSGSGSGSGSGSGSGSHAPIRGSPGSSDKQEGEWPTGLRLSRIGGIH*

Number of amino acids: 95
Molecular weight: 11359.06
Arg+Lys=12+4=16
Glu+Asp=11+5=16
Net charge at pH 7:0

>S6^{-16}-SecM(*Ms*) AP UniProtKB - P23370 (RS6_THETH)
MSSYEVNIVLNPNLDQSQLALESEIIQSALENYGASVESVEELGLSSLAYPIASDPQGYF
LWYQVEMPEDSVNDLASELSISDNVSSVMVVSSQESGSGSGSGSGSGSGSGSGSGSGG
SGSGSGSGSGSGSGSGSGSGSGSGSGSHAPIRGSPGSSDKQEGEWPTGLRLSRIGGIH*

Number of amino acids: 95
Molecular weight: 10365.36
Arg+Lys=0+0=0
Glu+Asp=11+5=16
Net charge at pH 7:-16

Supplementary Table 3. Thermodynamic stabilities *(*$\Delta G_{D\text{-}N}$) in kcal/mol and *log* k_f values (measured at 25 ºC in 50 mM MES buffer, pH 6.2) for 16 single-point mutations in protein S6 (10), and fraction full length (f_{FL}) for the corresponding nascent chain mutants obtained with the SecM(*Ms*) AP at at L = 30 residues. The f_{FL} measurements are averages of 3 independent replicates and their corresponding standard deviations are shown.

	$\Delta G_{D\text{-}N}$	*log* k_f	f_{FL}	ST dev.
WT	8.97	2.64	0.84	0.005
V6A	5.34	1.86	0.28	0.012
I8A	4.82	1.61	0.24	0.011
L10A	4.37	1.83	0.16	0.011
L19A	6.36	2.42	0.53	0.031
I26A	5.94	1.86	0.33	0.038
L30A	5.31	2.03	0.41	0.026
V37A	6.31	2.39	0.42	0.034
L61A	5.55	2.4	0.20	0.012
Y63A	5.08	2.3	0.55	0.024
V65A	5.64	2.16	0.39	0.051
V72A	7.56	2.59	0.80	0.008
L75A	6.89	2.3	0.60	0.011
L79A	4.21	2.27	0.23	0.011
V85A	5.51	2.51	0.54	0.007
V88A	6.88	2.39	0.69	0.009
V90A	6.12	2.42	0.67	0.013

Supplementary Table 4 Thermodynamic stabilities *(*$\Delta G_{D\text{-}N}$) in kcal/mol, *log* k_f , and *log* k_u values for S6 variants of different net charge (9). In the *in vitro* studies, the $S6^0$ mutant (called $S6^{+1}$ in Ref. 9) was obtained by analyzing the $S6^{-16}$ mutant (called $S6^{+1\text{-}17}$ in Ref. 9) at pH 2.3.

S6 version	*log* k_f	*log* k_u	$\Delta G_{D\text{-}N}$ kcal/mol
WT	2.53	-3.51	8.24
+17	3.34	-0.86	5.72
0	3.28	0.35	3.99
-16	1.48	-0.64	2.89

Supplementary Table 5. Proteins tested in previous studies

Protein	Reference	PDB ID	Length (residues)	Molecular weight (kDa)	Net charge
DHFR	(11)	8dfr	189	21.64	+1
PENT	(12)	6fls	105	12.03	-10
Titin I27	(13)	1tit	89	9.78	-6
R15	(14)	1u5p	109	12.99	+1
R16		1aj3	109	12.73	+1
β16		3lbx	109	13.43	0
ADR1a	(15)	2adr	29	3.47	+5
FLN5	(16)	1qfh	106	11.22	-9
TOP7	(17)	1qys	95	10.75	-4

Supplementary References

1. Dahiyat B, Sarisky C, & Mayo S (1997) De novo protein design: towards fully automated sequence selection. *J Mol Biol* 273:789-796.
2. Luh LM, *et al*. (2013) Molecular crowding drives active Pin1 into nonspecific complexes with endogenous proteins prior to substrate recognition. *J Am Chem Soc* 135:13796-13803.
3. Rocklin GJ, *et al*. (2017) Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* 357:168-175.
4. Kuszewski J, Gronenborn AM, & Clore GM (1996) Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci* 5:1067-1080.

5. Lakowski TM, Lee GM, Okon M, Reid RE, & McIntosh LP (2007) Calcium-induced folding of a fragment of calmodulin composed of EF-hands 2 and 3. *Protein Sci* 16:1119-1132.

6. Danielsson J, *et al*. (2013) Global structural motions from the strain of a single hydrogen bond. *Proc Natl Acad Sci U S A* 110:3829-3834.

7. Rea AM, Thurston V, & Searle MS (2009) Mechanism of ligand-induced folding of a natively unfolded helixless variant of rabbit I-BABP. *Biochemistry* 48:7556-7564.

8. Ohman A, Oman T, & Oliveberg M (2010) Solution structures and backbone dynamics of the ribosomal protein S6 and its permutant P(54-55). *Protein Sci* 19:183-189.

9. Kurnik M, Hedberg L, Danielsson J, & Oliveberg M (2012) Folding without charges. *Proc Natl Acad Sci U S A* 109:5705-5710.

10. Haglund E, Lindberg MO, & Oliveberg M (2008) Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. *J Biol Chem* 283:27904-27915.

11. 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.

12. Notari L, Martinez-Carranza M, Farias-Rico JA, Stenmark P, & von Heijne G (2018) Cotranslational folding of a pentarepeat β-helix protein. *bioRxiv* https://doi.org/10.1101/255810.

13. 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.

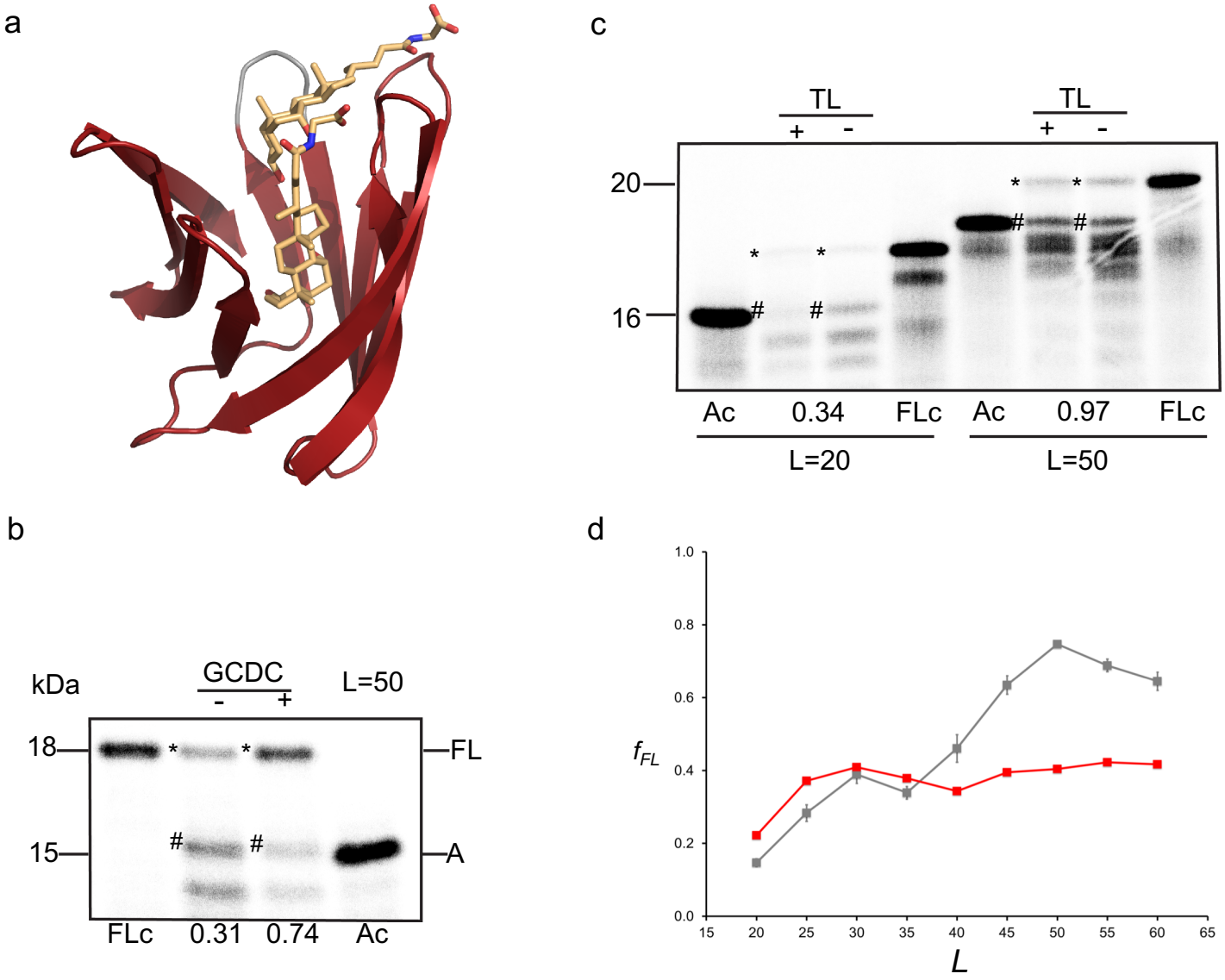
14. Nilsson OB, *et al*. (2017) Cotranslational folding of spectrin domains via partially structured states. *Nat Struct Mol Biol* 24:221-225.

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

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

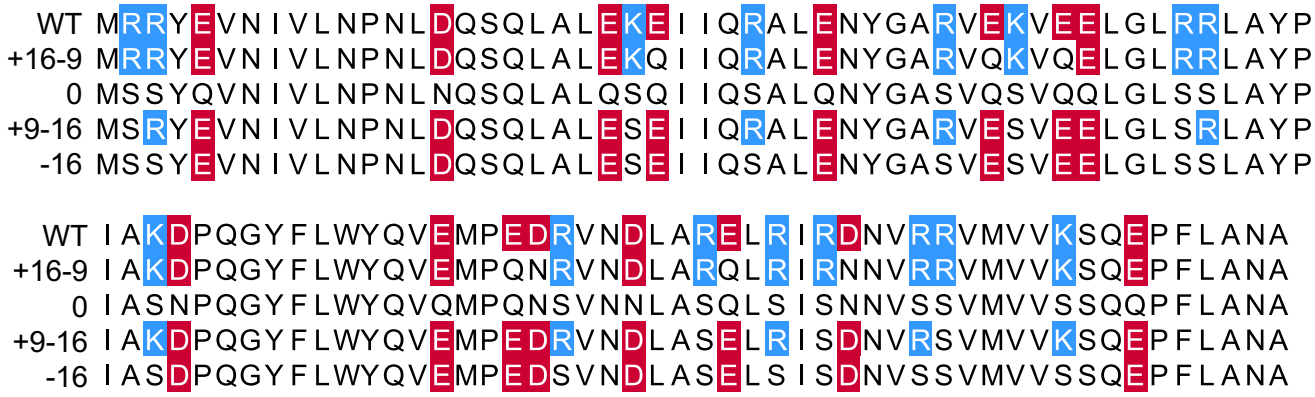
17. Goldman DH, *et al*. (2015) Mechanical force releases nascent chain-mediated ribosome arrest *in vitro* and *in vivo*. *Science* 348:457-460.

Supplementary Figures S1-S3

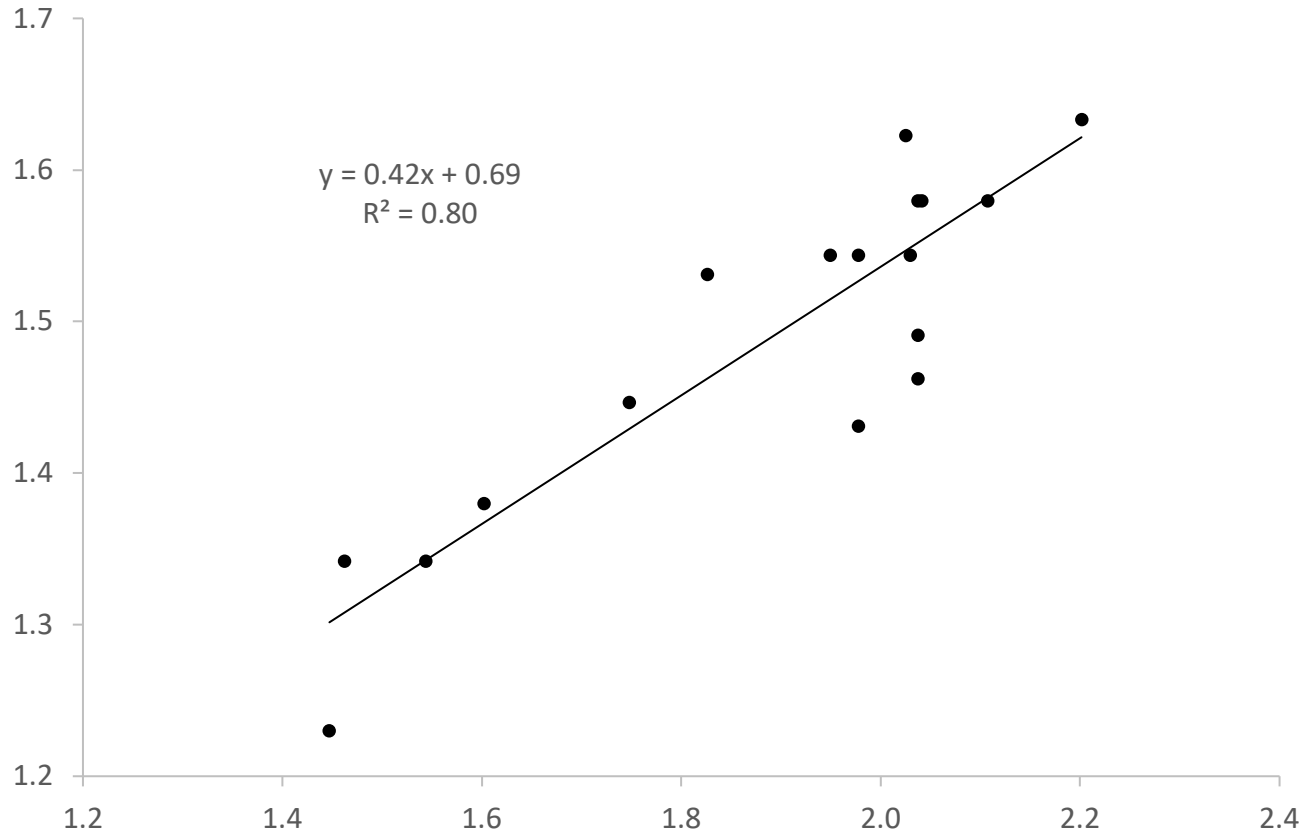


Supplementary Figure S1

Supplementary Figure S1, continued. **(**a) Model for a loop-truncated ILBP mutant used to test the effects of ligand binding on the f_{FL} profile. The model is based on PDB 2lba, the modeling template showed 60 % sequence identity with the modeled targets and the position of the ligand reflects an estimation of the position in the modeled using as reference the position in the NMR structure. Biochemical data supports the binding of this ligand in all versions of the protein used for this study [17]. (b) *In vitro* translation of the loop-truncated ILBP[L=50] construct with the SecM(*Ms*) AP, showing that the protein exerts a stronger pulling force on the nascent chain in the presence (+) than in the absence (-) of the ligand GCDC (f_{FL} = 0.74 *vs*. 0.31; $f_{FL} = I_{FL}/(I_{FL}+I_A)$ where I_{FL} (or I_A) is the intensity of the band marked * (or #). Full-length (FLc) and arrested (Ac) controls are indicated; the FLc construct has a P-to-A mutation the critical Pro at the end of the AP and does not give any arrested product, the Ac construct has a stop codon inserted directly after the AP. (c) *In vitro* translation and pulse proteolysis of wildtype ILBP at L = 20 and L = 50 residues. The arrested ILBP nascent chain is susceptible to thermolysin at L = 20 residues (f_{TR} = 0.34) but not at L = 50 residues (f_{TR} = 0.97), compare bands marked # in the **±**TL lanes. Full length (FLc) and arrested (Ac) controls are indicated. (d) f_{FL} profiles for loop-truncated ILBP (red curve) and loop-truncated ILBP translated in the presence of 400 μM of the ligand glycochenodeoxycholic acid (GCDC; grey curve).



Supplementary Figure S2. Sequences of the protein S6 charge variants analyzed in Figure 4.



Supplementary Figure S3. Scatter plot and linear fit of $log_{10}(L_{onset})$ vs $log_{10}(size)$