# Supporting Information

**Waudby et al.**

## SI Materials and Methods

**Protein expression and purification.** FLN5 truncations were prepared by site-directed mutagenesis (Stratagene) by replacing the residue of choice with a stop codon. Isotopically unlabeled and uniformly $^{15}$N, $^{13}$C/$^{15}$N and $^{2}$H/$^{13}$C/$^{15}$N labeled proteins were overexpressed in *E. coli* BL21-GOLD cells (Stratagene) and purified as previously described(1). Trigger factor was expressed and purified as previously described (2). Purified proteins were exchanged into Tico buffer (10 mM HEPES, 30 mM $NH_4Cl$, 6 mM $MgCl_2$, 0.02% sodium azide, 1 mM EDTA, pH 7.5). 1 mM BME was also included in samples of full-length FLN5 and the $\Delta$2 truncation to inhibit formation of disulphide-bonded dimers at Cys747.

**Sequence analysis.** Multiple sequence alignments were generated using the MATLAB Bioinformatics Toolbox (2016b, The MathWorks Inc., MA) using the BLOSUM80 scoring matrix. Percent identities are reported relative to the number of non-gap positions.

**Analysis of *cis*-proline abundance.** PDB entries from the RepeatsDB (v2.0, (3)) and CATH database (S35 sets, (4)) were systematically surveyed for the presence of *cis*-proline residues. All repeating units were counted for a given chain within RepeatsDB, but where multiple chains were present (for entries in both RepeatsDB and CATH databases) only a single chain was analysed and counted. *Cis* conformations were defined as having $\omega$ dihedral angles between $-30$ and $+30$ degrees.

**CD spectroscopy.** CD spectra of truncation variants were acquired at 283 and 298 K using a Jasco J-810 spectropolarimeter (Jasco Corporation, Tokyo), with protein concentrations of ca. 0.5 mg mL$^{-1}$ and urea concentrations varied between 0 and 8.2 M. Samples were pre-equilibrated for 1–4 hours in a water bath and CD signals at 234 nm were then measured as a function of the denaturant concentration and globally fitted to a two-state unfolding model using a shared $m$-value, in order to determine the free energy of folding (5).

**NMR spectroscopy.** Unless otherwise noted, NMR data were acquired at 283 K on a 700 MHz Bruker Avance III spectrometer equipped with a TXI cryoprobe, using uniformly $^{15}$N or $^{15}$N/$^{13}$C labeled protein at concentrations of 0.2–1 mM in Tico buffer supplemented with 10% $D_2O$ (v/v) and 0.001% DSS (w/v) as an internal chemical shift reference. Data were processed with nmrPipe and cross-peak intensities were measured using FuDA (http://www.biochem.ucl.ac.uk/hansen/fuda/).

***Resonance assignment.*** FL, $\Delta$4, $\Delta$6 and $\Delta$9 FLN5 spectra were assigned by standard triple-resonance methods. $^{1}$H,$^{15}$N difference spectra obtained from real-time NMR measurements of $\Delta$6 refolding were used to assist the assignment of closely overlapping resonances arising from *cis* and *trans* intermediates. Additionally, a 0.2 mM sample of uniformly $^{2}$H/$^{15}$N-labeled $\Delta$6 was used to record a 2D $N_z$ magnetization exchange spectrum (6) (mixing time 1 s) to cross-validate the assignment of unfolded and intermediate resonances. Only small chemical shift changes were observed between the unfolded and intermediate states of $\Delta$6 and $\Delta$9, and cross-peak assignments of $\Delta$9 were therefore transferred directly from those of $\Delta$6 and cross-validated using a 2D $N_z$ magnetization transfer measurement. The assignments of $^{1}$H,$^{15}$N-HSQC resonances of unfolded $\Delta$12 and $\Delta$16 variants were similarly transferred from the complete $\Delta$6 and $\Delta$9 assignments.

***Intensity analysis.*** HSQC spectra were acquired of uniformly $^{15}$N-labeled proteins at concentrations of 300–350 $\mu$M and processed with exponential apodization in both dimensions. Resonances were fitted with Lorentzian lineshapes and observed volumes were corrected for the effects of transverse relaxation during INEPT transfer periods(7). For each spectrum, $\Delta G_{\mathrm{a-b}} = -RT\ln(V_{\mathrm{a}}/V_{\mathrm{b}})$ was calculated independently for several residues using cross-peaks that were well resolved in all states, and stabilities have been reported as the mean ± s.e.m. of these measurements.

In $\Delta$12 and $\Delta$16 $^{1}$H,$^{15}$N-HSQC spectra, the only resonances that could be detected were those arising from the U state, and therefore it is only possible to place a limit on the maximum stability using the signal-to-noise ratio in the spectrum to estimate the maximum possible population of unobserved structured state. The observed signal-to-noise ratios in $\Delta$12 and $\Delta$16 spectra were between 250 and 300, but to reflect the additional uncertainty arising from possible changes in the position and line-widths of folded cross-peaks, the limit of detection was capped at 1%. From this, a minimum $\Delta G_{I-U}$ of 2.7 kcal mol$^{-1}$ was calculated.

RNC stabilities were calculated as a function of linker length using the intensity of selected $^{1}$H,$^{15}$N-SOFAST-HMQC resonances (8), measured relative to the partially emerged and fully unfolded +21 RNC (1). As the selected resonances were previously shown not to exhibit length-dependent changes in linewidth (1), their relative intensity was used directly as a measure of the unfolded state population and used to determine free energies of folding. Uncertainties were propagated using a Monte Carlo approach.

***Magnetization transfer experiments.*** 2D $N_z$ longitudinal magnetization transfer experiments(6) were acquired for $\Delta$6, $\Delta$6 P742A and $\Delta$9 with magnetization transfer delays between 0.01 and 2.5 s. The intensities of residues for which complete sets of exchange cross-peaks could be observed between folded (F) and unfolded (U) states were fitted to a numerical solution of the exchange process:

$$\begin{pmatrix} I_{FF}(t) & I_{FU}(t) \\ I_{UF}(t) & I_{UU}(t) \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \cdot \exp\left[ - \begin{pmatrix} R_{1,F} + k_{FU} & -k_{UF} \\ -k_{FU} & R_{1,U} + k_{UF} \end{pmatrix} t \right] \cdot \begin{pmatrix} I_{F,0} & 0 \\ 0 & I_{U,0} \end{pmatrix} \qquad [1]$$

www.pnas.org/cgi/doi/10.1073/pnas.1716252115

where $I_{FU}$ is the amplitude of the F-to-U cross-peak etc., $I_{F,0}$ and $I_{U,0}$ are the amplitudes of the folded and unfolded states at the start of the magnetization transfer period, and $\alpha$ and $\beta$ are factors accounting for relaxation following this period. Data for multiple residuals were fitted globally with shared exchange rates $k_{FU}$ and $k_{UF}$. Uncertainties in the fitted exchange parameters were determined by bootstrap resampling of residuals.

***Real-time temperature-jump kinetics.*** Real-time NMR temperature-jump measurements were acquired for samples of Δ6, Δ6 P742A and Δ4 P742A. Cooling gas was provided using a BCU-05 unit with a flow rate of 935 L hr$^{-1}$. The time required for the temperature within NMR samples to re-equilibrate after lowering the temperature from 310 K was measured by repeated acquisition of 1D $^1$H NMR spectra of a sample of d$_4$-methanol during a programmed temperature change (9), and we found that the internal temperature stabilized to within 1 K within 80–120 s depending on the final temperature (277–298 K). We observed little difference in the rate of temperature change in 3 mm vs 5 mm diameter NMR tubes, and therefore 5 mm shigemi tubes were used to maximize sensitivity while minimizing formation of temperature gradients within the sample.

$^1$H,$^{15}$N-SOFAST-HMQC experiments (2 scans, 1024 × 32 complex points, 15.7 s per spectrum) were prepared and calibrated with a sample incubated at the target temperature (8). The sample temperature was then raised to 310 K and equilibrated for at least 15 min to induce unfolding. We note that the extrapolated timescale of cis/trans isomerization at this temperature is 20 s (Figure 4D), so the system is expected to be fully equilibrated under these conditions. The repeated acquisition of 2D experiments was then begun and a temperature change was initiated. Acquisition was continued for 1–2 hr until the re-equilibration process was complete. This process was repeated a number of times to increase the signal to noise ratio of the final data. Sample integrity during these temperature cycles was monitored using 1D $^1$H spectra and the first increments of the 2D spectra.

Experimental time-courses were fitted globally to a constant plus one or two exponential phases using a variable projection algorithm(10). After Fourier transformation in the direct dimension and extraction of the amide region, the pseudo-3D data $Y_{mij}(\tau, t_1, \omega_2)$ were reshaped into an $M \times N$ matrix $y_{mn}$, where $M$ is the number of time points and $N$ the total number of points in each 2D spectrum. We then attempted to find solutions of the form $y = CA + \epsilon$, where $C$ is an $M \times L$ matrix describing the changing concentration of spectral components over time, $A$ is a $L \times N$ matrix describing the spectral amplitudes of these components, and $\epsilon$ is the error term. $C$ is parameterized by the rate constants $k_l$ for each kinetic process, $C_{ml}(k) = \exp(-k_l \tau_m)$, and we fix $k_1 = 0$ to represent the equilibrium end-point. Thus, we attempt to find $\min_{A,k} \|C(k)A - y\|_F$. This is a difficult non-linear optimization problem in a high-dimensional space, but if $k$ is fixed then the solution is simply $A = C^+(k)y$, where $C^+$ is the Moore-Penrose pseudoinverse. By exploiting this conditionally linear structure, the initial problem can be reduced to $\min_k \|C(k)C^+(k)y - y\|_F$. This is the far simpler problem of non-linear minimization in the rates alone, which we solve using a Trust-Region-Reflective algorithm. Reported uncertainties in the rate constants are calculated from the Jacobian approximation to the covariance matrix. The amplitudes $A$ of the component spectra can then be found by projection, and these are reshaped back into 2D spectra and Fourier transformed in the indirect dimension for further visualization and analysis.

***CPMG relaxation dispersion.*** $^{15}$N CW-CPMG relaxation dispersion experiments(11) were acquired at 500 and 700 MHz. Uniformly $^{15}$N-labeled samples of Δ9, Δ12 and Δ16 were prepared, while measurements of Δ6 were conducted using uniform $^2$H/$^{15}$N-labeling. Experiments were performed at 283 K, using a ca. 10 kHz $^1$H spin-lock, 5.56 kHz $^{15}$N CPMG refocusing pulses, and a 40 ms relaxation delay with $^1$H and $^{15}$N temperature compensation elements. Changes in the sample temperature during acquisition were measured using the changes in the $H_2O$ chemical shift and the external temperature was adjusted to maintain the desired internal temperature. Dispersion profiles were fitted to two-state exchange models using CATIA (http://www.biochem.ucl.ac.uk/hansen/catia/) in a two step process: firstly, residues with large exchange contributions were fitted globally to determine the global exchange parameters ($k_{\text{ex}}$ and $p_B$); these parameters were then held constant and used to fit chemical shift differences for residues showing weaker dispersions.

***Trigger factor.*** Changes in $^1$H,$^{15}$N HSQC peak intensities were measured following addition of an equimolar amount of TF to a sample of 50 $\mu$M $^{15}$N-labelled Δ6 (283 K). Similarly, a 10 $\mu$M $^{15}$N-labelled FLN5+21 RNC sample was prepared (as described previously, (1)) and changes in peak intensities in $^1$H,$^{15}N$ SOFAST-HMQC spectra were measured following the addition of 3 equivalents of TF. The stability of the RNC was monitored by a dual strategy of western blot analysis using an anti-His$_6$ antibody and monitoring of NMR translational diffusion measurements (Fig. S15D,E) as previously described (12), and all measurements were completed within the first nine hours of acquisition.

**Structure determination.** Structures of the ground states were obtained from chemical shift-restrained replica-averaged metadynamics (13) performed using the Amber99SB*-ILDN force field (14, 15). All the simulations were run in GROMACS 4.6.5 (16) using PLUMED 1.3 (17) and Almost (18) to introduce metadynamics and chemical shift restraints. The particle-mesh Ewald method was used for long-range electrostatic interactions, with a short range cut-off of 0.9 nm, and for Lennard-Jones interactions a 1.2 nm cut-off was used. Simulations were run in the canonical ensemble using velocity rescaling with a stochastic term for temperature coupling (19). Simulations were carried out for FL and Δ6 I (with P742 in a *trans* conformation). Initial structures for each of the systems were prepared from the crystal structure, 1qfh (20), with addition of N-terminal hexahistidine tags, truncation of the C-terminus as required, and changing the isomeric state of P742 to *trans* in the Δ6 model. After initial equilibration runs, each system was simulated for ca. 450 ns in four replicas in parallel at 283 K, with a harmonic restraint applied on the average value of CamShift back-calculated chemical shifts (21). In addition, each replica was sampling one of the four collective variables (CVs) applied to: function of the $\phi$ and $\psi$ dihedral angles (ALPHABETA1), the $\chi_1$ and $\chi_2$ dihedral angles for hydrophobic and polar amino acids (ALPHABETA2), the fraction of the antiparallel $\beta$-sheet conformations (ANTIBETARMSD), and the radius of gyration (RGYR). Exchanges between replicas were attempted periodically every 50 ps according to a replica-exchange scheme and well-tempered scheme(22), with a bias-factor of 8.0 used to rescale the added bias potential. Convergence of the simulations was reached after 350 ns which resulted in free energy landscapes with statistical uncertainties of $\leq$2.0 kJ mol$^{-1}$ for free energies up to 10 kJ mol$^{-1}$ (Figure S3g-h). Analysis of the metadynamics trajectory and the assignment of microstates was carried out in VMD using the METAGUI plugin (23, 24). Structural representatives of each ground state were chosen from the lowest free energy microstate using criteria of the lowest MolProbity score. Residual dipolar couplings were back-calculated from structural ensembles using PALES and ensemble averaged values were used to calculate quality factors, $Q = \sqrt{\sum (D_\text{obs} - D_\text{calc})^2 / \sum D_\text{obs}^2}$ (25). Hexahistidine purification tags were excluded from analyses of RMSD/RMSF and are not shown in figures.

Folding simulations on the ribosome for I and N states were conducted in Gromacs 4.6.5 (16) with all-atom structure-based models generated in SMOG 2 (26). Simulated systems combined a fragment of the ribosome around the exit tunnel and the FLN5 domain in an extended conformation tethered to the PTC via linker with variable length from 15 aa to 31 aa. The linker sequence was the same as used in FLN5 RNCs studies(1). Ribosome atoms were frozen during the simulations, whereas for FLN5 contacts from N or I states were applied to simulate folding; linker residues were not restrained.

**Coarse-grained simulations of interdomain misfolding.** Simulations of interdomain misfolding in FLN4-5 and titin I27 dimers (Fig. 5A and S13A) were performed using a variant of the approach described in (27). Structure-based models (26) were generated using one bead centered at the C$\alpha$ position of each residue. Lennard-Jones 10–12 potentials, with relative strengths given by the Miyazawa-Jernigan matrix (28), were used to model native contacts and homologous contacts between domains, whereas all other non-bonded interactions were modelled by repulsive terms ($r^{-12}$). The temperature of each simulation was chosen to be close to the $T_m$ of the domains. 500 ns trajectories were calculated for FLN4-5 and I27 dimers using parallel bias metadynamics, with two collective variables: $Q_\text{native}$, the fraction of native (intradomain) contacts, and $Q_\text{misf}$, the fraction of misfolded (interdomain) contacts. All simulations were carried out in Gromacs 4.5.7. with PLUMED 2.1 (29). In order to asses convergence of these simulations we plotted changes in one-dimensional free energy landscapes (Fig. S13B-E) and the free energy difference between unfolded and native states (Fig. S13F) as a function of simulation time, finding in both cases a high level of convergence. We then calculated two-dimensional free energy landscapes as a function of the collective variables for FLN4-5 (Fig. 5A) and I27 dimers (Fig. S13A). Finally, to calculate the populations of misfolded states structures obtained from simulations were clustered with a density peak clustering algorithm (30), and misfolded conformations were defined as members of the unfolded cluster with $Q_\text{misf} > 0.16$ (as this threshold value was not present in clusters in which one or both domains were folded).

**Markov chain modeling.** Transition matrices were calculated by linear interpolation of measured stabilities as a function of chain length. Rate constants were interpolated assuming that folding rates are approximately independent of chain length, with the unfolding rate being modulated according to the measured stability, $k_u = k_f \exp(\Delta G/RT)$. We also assumed that folding to the native state occurs directly from the intermediate, although given the separation in the timescales of folding and proline isomerization, $I$ and $U$ are effectively in pre-equilibrium and the impact of this choice is minimal. Populations (folding probabilities) were then calculated as a function of time using the matrix exponential of the transition matrix, given an initial population of Δ16 in the unfolded state. Populations were also calculated as a function of polypeptide chain length by insertion of absorbing states after each translation step to quench further reactions, an approach equivalent to that of O'Brien et al. (31).

**Calculation of misfolding risk.** Following synthesis of domain $i$, the expected translation time of the following domain is $\tau_\text{trans}^{(i+1)} = k_\text{trans} L^{(i+1)}$, where $k_\text{trans}$ is the translation rate and $L^{(i+1)}$ is the number of residues in domain $i+1$. As the sum of a series of identical exponential processes (i.e. averaging out sequence-specific variations in $k_\text{trans}$), $\tau_\text{trans}$ has an Erlang distribution with rate $k_\text{trans}$ and shape parameter $L$. $\tau_\text{trans}$ therefore has a narrow distribution, with coefficient of variation $L^{-1/2}$, and so will be approximated as a constant hereonin. The probability of domain $i$ folding during translation of domain $i+1$ is then $p_F^{(i)} = 1 - \exp\left(-\tau_\text{trans}^{(i+1)}/\tau_\text{fold}^{(i)}\right)$, where $\tau_\text{fold}^{(i)}$ is the time constant for folding, in the present case given by $k_\text{UI}^{-1} + k_\text{IN}^{-1}$ (approximating folding with single exponential kinetics). Therefore, for a protein with $N$ tandem repeat domains, the misfolding risk, $R$, given by the probability of populating adjacent unfolding domains during biosynthesis, is $R = 1 - \prod_{i=1}^{N-1} p_F^{(i-1)}$.

## References

1. Cabrita LD, et al. (2016) A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat. Struct. Mol. Biol.* 23(4):278–285.
2. Deckert A, et al. (2016) Structural characterization of the interaction of $\alpha$-synuclein nascent chains with the ribosomal surface and trigger factor. *Proc. Natl. Acad. Sci. U.S.A.* 113(18):5012–5017.
3. Paladin L, et al. (2017) RepeatsDB 2.0: improved annotation, classification, search and visualization of repeat protein structures. *Nucleic Acids Res.* 45(D1):D308–D312.
4. Dawson NL, et al. (2017) CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* 45(D1):D289–D295.
5. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Meth. Enzymol.* 131:266–280.
6. Farrow NA, Zhang O, Forman-Kay JD, Kay LE (1995) Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry* 34(3):868–878.
7. Christodoulou J, et al. (2004) Heteronuclear NMR investigations of dynamic regions of intact Escherichia coli ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* 101(30):10949–10954.
8. Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR* 33(4):199–211.
9. Findeisen M, Brand T, Berger S (2007) A 1H-NMR thermometer suitable for cryoprobes. *Magn Reson Chem* 45(2):175–178.
10. Golub GH, Pereyra V (1973) The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate. *SIAM J. Numer. Anal.* 10(2):413–432.
11. Hansen DF, Vallurupalli P, Kay LE (2008) An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J Phys Chem B* 112(19):5898–5904.

12. Cassaignau AME, et al. (2016) A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nat Protoc* 11(8):1492–1507.
13. Camilloni C, Vendruscolo M (2014) Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J. Am. Chem. Soc.* 136(25):8982–8991.
14. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113(26):9004–9015.
15. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78(8):1950–1958.
16. Pronk S, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854.
17. Bonomi M, et al. (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* 180(10):1961–1972.
18. Fu B, et al. (2014) ALMOST: an all atom molecular simulation toolkit for protein structure determination. *J Comput Chem* 35(14):1101–1105.
19. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1):014101.
20. McCoy AJ, Fucini P, Noegel AA, Stewart M (1999) Structural basis for dimerization of the Dictyostelium gelation factor (ABP120) rod. *Nat. Struct. Biol.* 6(9):836–841.
21. Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M (2009) Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.* 131(39):13894–13895.
22. Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100(2):020603.
23. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33–8– 27–8.
24. Biarnés X, Pietrucci F, Marinelli F, Laio A (2012) METAGUI. A VMD interface for analyzing metadynamics and molecular dynamics simulations. *Computer Physics Communications* 183(1):203–211.
25. Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc* 3(4):679–690.
26. Noel JK, et al. (2016) SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput. Biol.* 12(3):e1004794.
27. Borgia MB, et al. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474(7353):662–665.
    Borgia A, et al. (2015) Transient misfolding dominates multidomain protein folding. *Nat Commun* 6:8861.
28. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* 256(3):623–644.
29. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) PLUMED 2: New feathers for an old bird. *Computer Physics Communications* 185(2):604–613.
30. Rodriguez A, Laio A (2014) Machine learning. Clustering by fast search and find of density peaks. *Science* 344(6191):1492–1496.
31. O'Brien EP, Vendruscolo M, Dobson CM (2012) Prediction of variable translation rate effects on cotranslational protein folding. *Nat Commun* 3:868.
32. Neira JL, Fersht AR (1999) Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments. *J. Mol. Biol.* 285(3):1309–1333.

**Table S1. NMR and refinement statistics for protein structures.**

| | FL | $\Delta$6 |
|---|---|---|
| Total NMR chemical shift constraints | 521 | 467 |
| $\delta(^{1}$HN) r.m.s. deviation[a] (ppm) | 0.43 | 0.36 |
| $\delta(^{1}$H$\alpha$) r.m.s. deviation[a] (ppm) | 0.26 | 0.22 |
| $\delta(^{15}$N) r.m.s. deviation[a] (ppm) | 2.97 | 2.26 |
| $\delta(^{13}$C') r.m.s. deviation[a] (ppm) | 1.11 | 0.97 |
| $\delta(^{13}$C$\alpha$) r.m.s. deviation[a] (ppm) | 1.15 | 0.76 |
| $\delta(^{13}$C$\beta$) r.m.s. deviation[a] (ppm) | 1.18 | 1.04 |
| Deviations from idealized geometry | | |
| Bond lengths (Å) | 0.011 | 0.011 |
| Bond angles (°) | 3.5 | 3.5 |
| Total RDCs (HN–N) | 84 | 56 |
| $Q_{\mathrm{free}}$ | 0.64 | 0.52 |

[a] Deviations from experimental chemical shifts were back-calculated from 10 refined structures using SPARTA+ (reported accuracy: 0.49, 0.25, 2.45, 1.09, 0.94, 1.14 ppm for $^{1}$HN, $^{1}$H$\alpha$, $^{15}$N, $^{13}$C', $^{13}$C$\alpha$ and $^{13}$C$\beta$ respectively).

**Table S2. Sequence identity (with preceding domain) and structural analysis of conserved proline conformation across filamin domains.**

| | FLN (*D. discoideum*) | | FLNA (human) | | FLNB (human) | | FLNC (human) | |
|---|---|---|---|---|---|---|---|---|
| domain | identity (%) | Pro conformation | identity (%) | Pro conformation | identity (%) | Pro conformation | identity (%) | Pro conformation |
| 1 | | | | | | cis (4b7l) | | |
| 2 | 41 | | 36 | | 40 | | 40 | |
| 3 | 37 | | 34 | cis (4m9p) | 34 | | 35 | |
| 4 | 41 | cis (1wlh) | 34 | cis (4m9p) | 29 | | 32 | cis (3v8o) |
| 5 | 44 | cis (1wlh) | 27 | cis (4m9p) | 26 | | 30 | cis (3v8o) |
| 6 | 31 | cis (1wlh) | 32 | | 30 | | 31 | |
| 7 | | | 35 | | 35 | | 30 | |
| 8 | | | 31 | | 36 | | 31 | |
| 9 | | | 22 | | 32 | cis (2di9) | 34 | |
| 10 | | | 36 | cis (3rgh) | 36 | cis (2dia) | 33 | |
| 11 | | | 33 | | 29 | cis (2dib) | 35 | |
| 12 | | | 37 | | 34 | cis (2dic) | 33 | |
| 13 | | | 41 | | 37 | cis (2dj4) | 37 | |
| 14 | | | 45 | | 38 | cis (2e9j) | 47 | cis (2d7m) |
| 15 | | | 31 | | 42 | cis (2dmb) | 35 | |
| 16 | | | 22 | cis (2k7p) | 23 | cis (5dcp) | 22 | cis (2d7n) |
| 17 | | | 26 | cis (2k7p) | 32 | cis (5dcp) | 31 | cis (2d7o) |
| 18 | | | 20 | cis (2k7q) | 18 | cis (2dmc) | 21 | |
| 19 | | | 20 | cis (2j3s) | 23 | cis (2di8) | 21 | |
| 20 | | | 20 | cis (2j3s) | 20 | cis (2dlg) | 31 | |
| 21 | | | 25 | cis (2j3s) | 23 | cis (2ee6) | 34 | |
| 22 | | | 24 | | 26 | cis (2eeb) | 24 | cis (2d7p) |
| 23 | | | 28 | cis (2k3t) | 29 | cis (2eec) | 24 | cis (2nqc) |
| 24 | | | 27 | cis (3cnk) | 29 | cis (2eed) | 33 | cis (1v05) |

**Table S3. Analysis of proline and *cis*-proline occurrence within repeat proteins (RepeatsDB v2.0) and the CATH database of protein domains (CATH v4.2 S35 set).**

| | class | description | entries | repeating units | residues | prolines | *cis*-prolines | %pro | %cis |
|---|---|---|---|---|---|---|---|---|---|
| RepeatsDB | III.1 | $\beta$-solenoid | 293 | 2473 | 58357 | 1871 | 179 | 3.2 | 9.6 |
| | III.2 | $\alpha/\beta$-solenoid | 246 | 3610 | 89629 | 3728 | 72 | 4.2 | 1.9 |
| | III.3 | $\alpha$-solenoid | 815 | 7285 | 277481 | 8940 | 205 | 3.2 | 2.3 |
| | III.4 | $\beta$-trefoil/hairpins | 67 | 448 | 9363 | 267 | 21 | 2.9 | 7.9 |
| | III.5 | anti-parallel $\beta$-layer/hairpins | 46 | 486 | 11181 | 239 | 6 | 2.1 | 2.5 |
| | III.6 | box | 31 | 144 | 8699 | 288 | 2 | 3.3 | 0.7 |
| | IV.1 | TIM barrel | 557 | 4452 | 123180 | 5001 | 161 | 4.1 | 3.2 |
| | IV.2 | $\beta$-barrel | 442 | 1881 | 63814 | 1459 | 34 | 2.3 | 2.3 |
| | IV.3 | $\beta$-trefoil | 13 | 39 | 1902 | 134 | 5 | 7.0 | 3.7 |
| | IV.4 | $\beta$-propeller | 698 | 4695 | 222786 | 9163 | 461 | 4.1 | 5.0 |
| | IV.5 | $\alpha/\beta$-prism | 82 | 477 | 31838 | 1287 | 88 | 4.0 | 6.8 |
| | IV.6 | $\alpha$-barrel | 19 | 114 | 5210 | 150 | 2 | 2.9 | 1.3 |
| | IV.7 | $\alpha/\beta$-barrel | 2 | 10 | 437 | 12 | 0 | 2.7 | 0.0 |
| | IV.8 | $\alpha/\beta$-propeller | 119 | 754 | 39850 | 1891 | 113 | 4.7 | 6.0 |
| | IV.9 | $\alpha/\beta$-trefoil | 34 | 102 | 4497 | 156 | 7 | 3.5 | 4.5 |
| | IV.10 | aligned prism | 24 | 72 | 3129 | 148 | 24 | 4.7 | 16.2 |
| | V.1 | $\alpha$ beads | 11 | 48 | 3154 | 62 | 1 | 2.0 | 1.6 |
| | V.2 | $\beta$ beads | 37 | 156 | 8954 | 678 | 41 | 7.6 | 6.0 |
| | V.3 | $\alpha/\beta$ beads | 20 | 76 | 3874 | 143 | 1 | 3.7 | 0.7 |
| | V.4 | $\beta$-sandwich beads | 43 | 160 | 15937 | 1015 | 83 | 6.4 | 8.2 |
| | V.5 | $\alpha/\beta$-sandwich beads | 48 | 220 | 22450 | 1340 | 90 | 6.0 | 6.7 |
| | III | elongated structures | 1498 | 14446 | 454710 | 15333 | 485 | 3.4 | 3.2 |
| | IV | closed structures | 1990 | 12596 | 496643 | 19401 | 895 | 3.9 | 4.6 |
| | V | beads on a string | 159 | 660 | 54369 | 3238 | 216 | 6.0 | 6.7 |
| CATH | 1 | mainly $\alpha$ | 7903 | N/A | 1032279 | 34581 | 983 | 3.3 | 2.8 |
| | 2 | mainly $\beta$ | 6879 | N/A | 915980 | 40778 | 2374 | 4.5 | 5.8 |
| | 3 | $\alpha/\beta$ | 16194 | N/A | 2745914 | 118861 | 5565 | 4.3 | 4.7 |

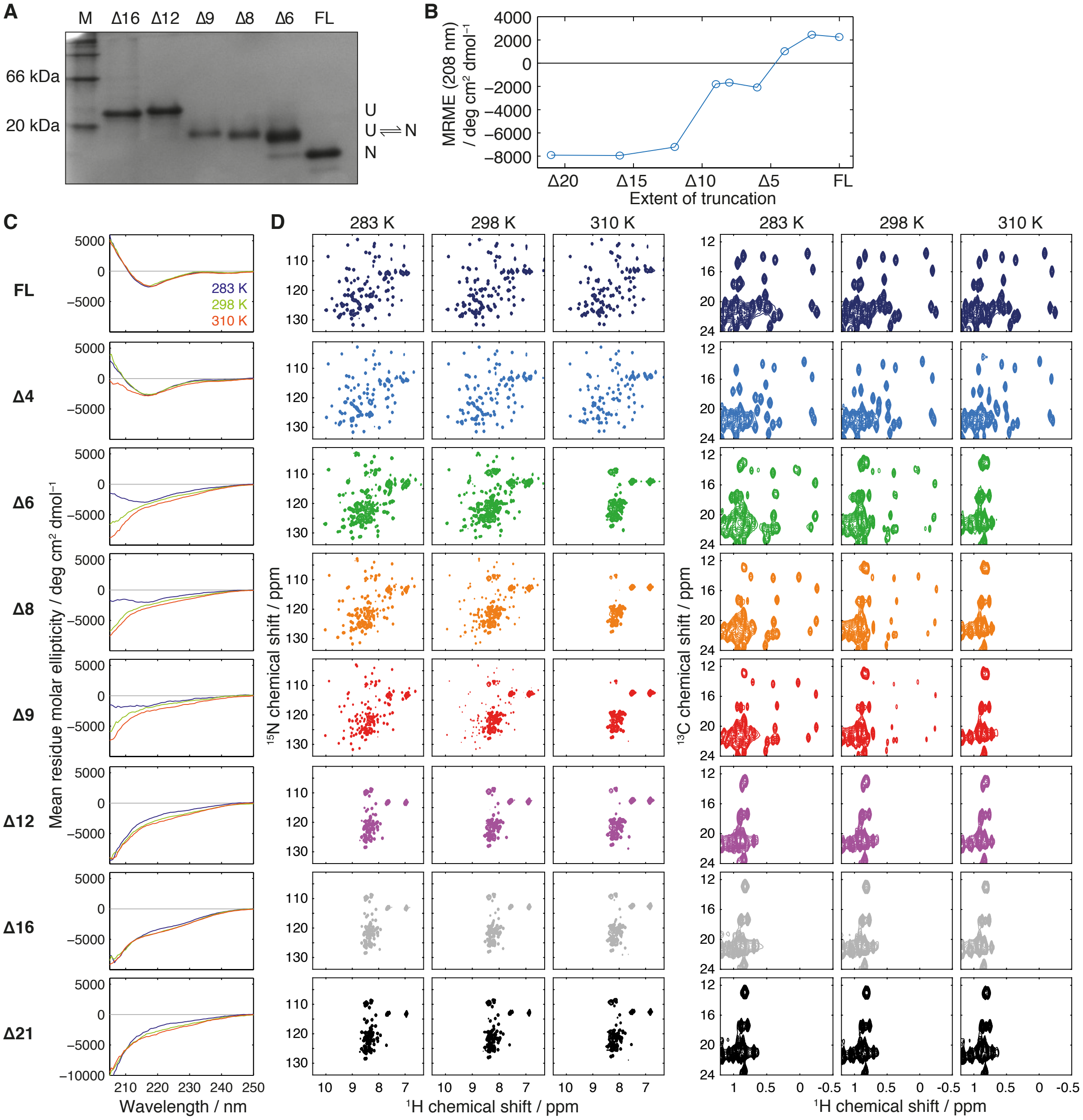

**Fig. S1.** Biochemical and spectroscopic characterization of truncation variants. (A) Native PAGE analysis of truncation variants at 283 K. (B) CD measurements of the folding of FLN5 truncation variants at 283 K (MRME, mean residue molar ellipticity). (C) CD spectra of truncation variants at 283 K (blue), 298 K (green) and 310 K (red). (D) $^{1}$H,$^{15}$N and $^{1}$H,$^{13}$C-SOFAST-HMQC spectra (600 MHz) of truncation variants, at multiple temperatures as indicated.

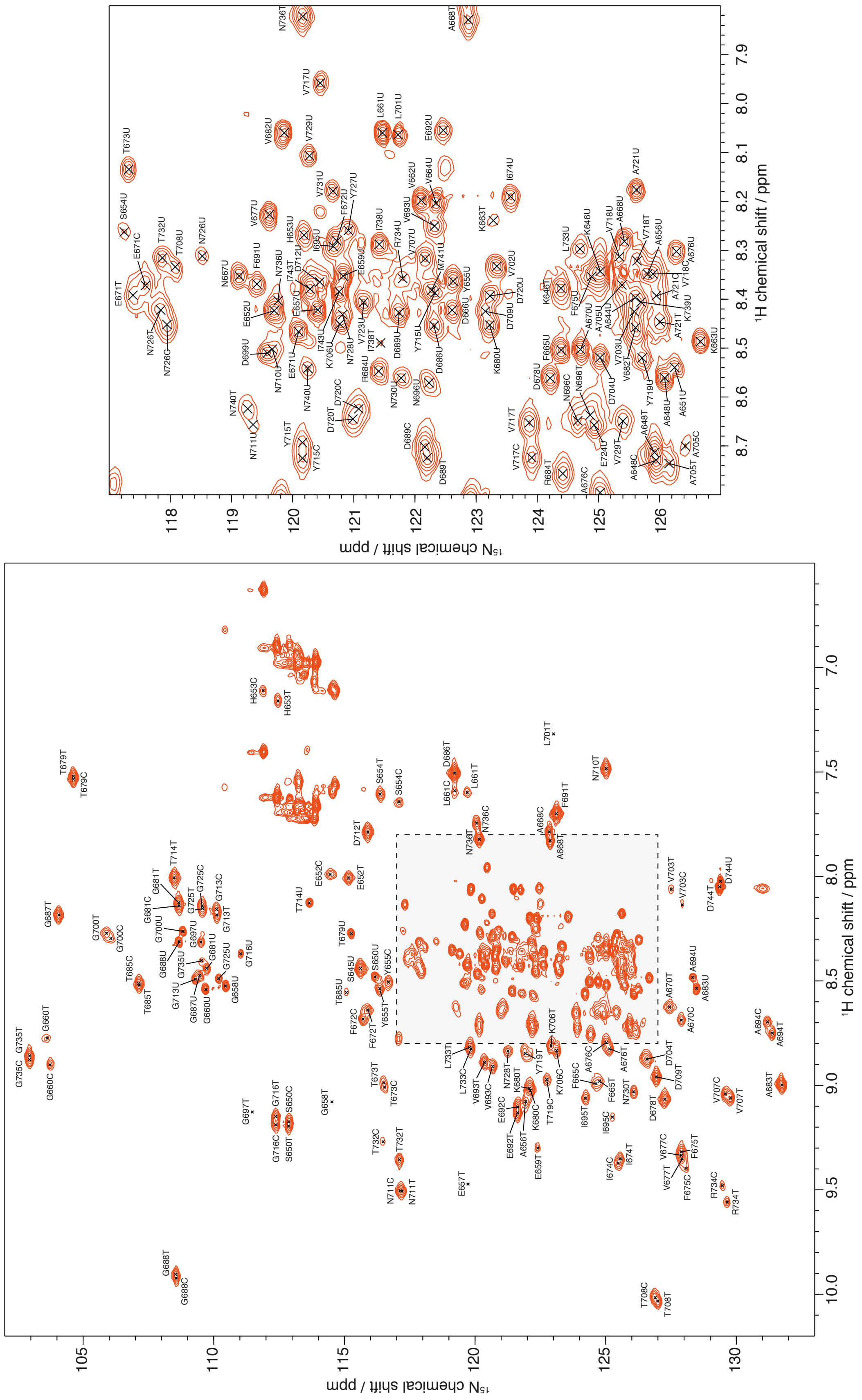

**Fig. S2.** $^{1}$H,$^{15}$N HSQC spectrum (283 K, 700 MHz) of FLN5△6 showing resonance assignments.

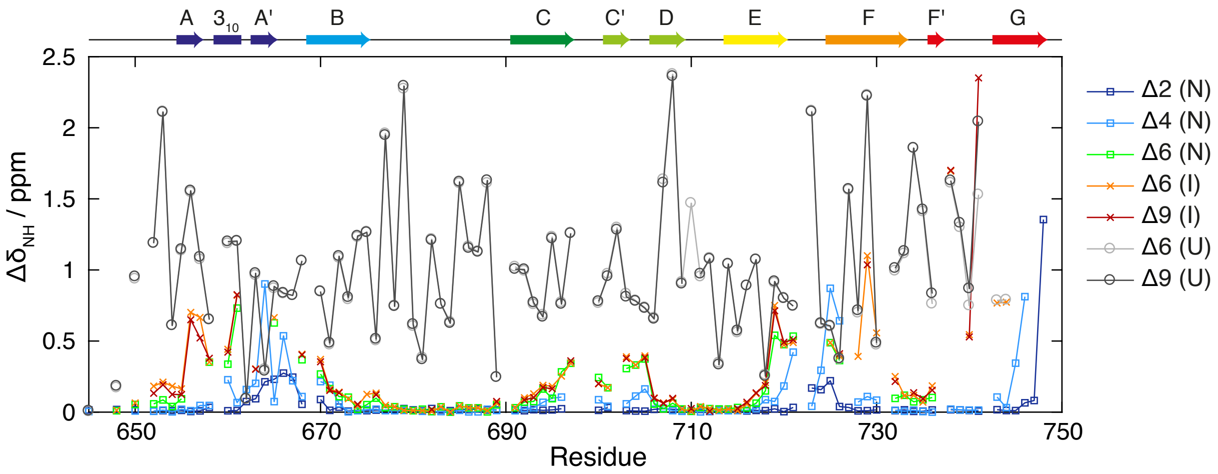

**Fig. S3.** Amide chemical shift changes, relative to FL FLN5, calculated as the weighted average $\Delta\delta_{\mathrm{NH}} = \sqrt{\Delta\delta_{\mathrm{H}}^2 + (\Delta\delta_{\mathrm{N}}/5)^2}$. The structural schematic above shows secondary structure elements of FL FLN5 (Figure 1B).

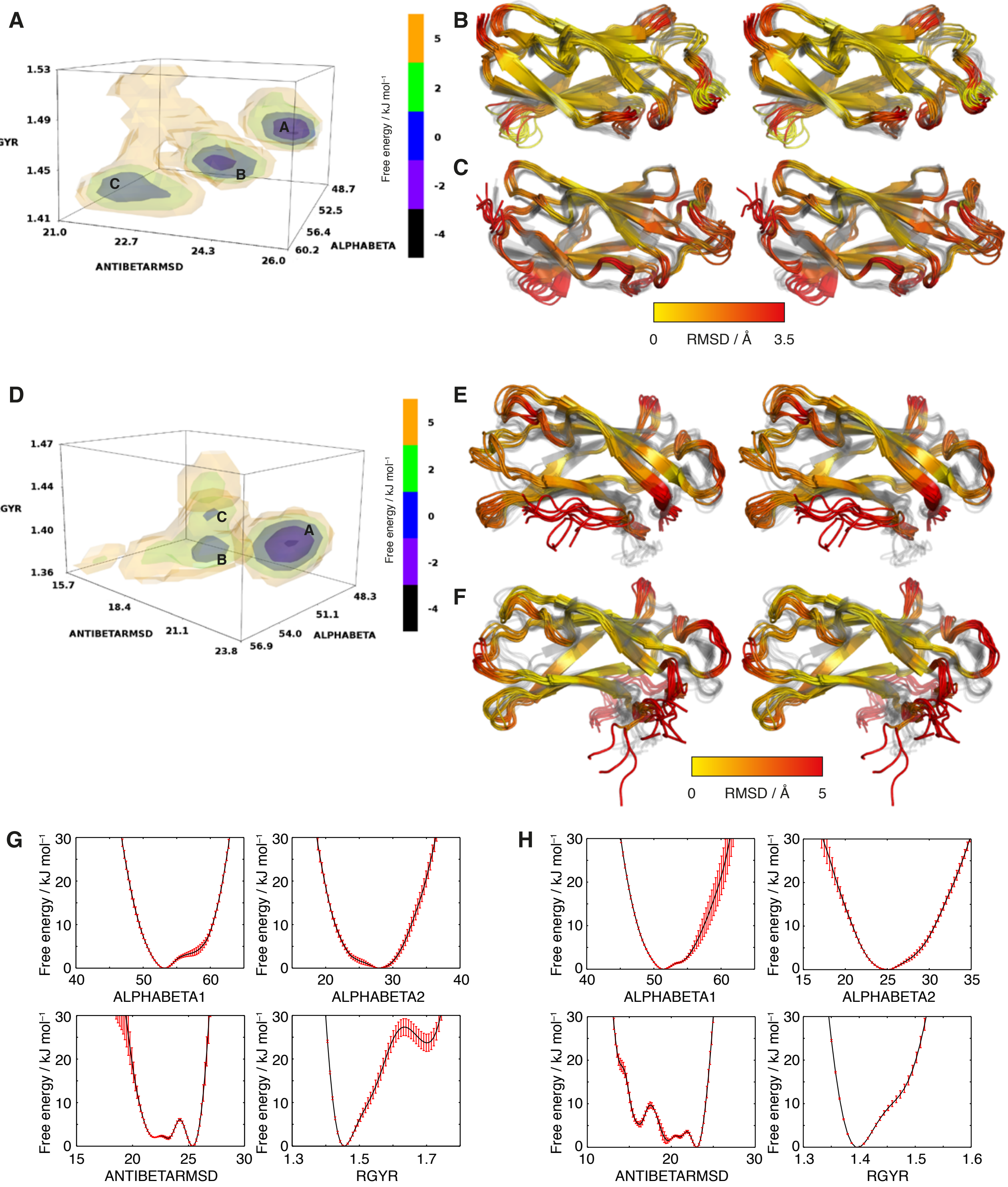

**Fig. S4.** Free energy landscapes determined by chemical shift-restrained replica-averaged metadynamics. (A) Free energy landscape of the FL metadynamics ensemble, as a function of the collective variables ANTIBETARMSD, ALPHABETA1 and RGYR (see Methods). The ensemble structure of the ground state 'A' (–1.0 kcal mol$^{-1}$) is shown in Figure 2A. (B) Ensemble structure of the first excited state 'B' (–0.74 kcal mol$^{-1}$), colored yellow-red by the RMSD from the ground state 'A', shown in grey. (c) Ensemble structure of the second excited state 'C' (–0.45 kcal mol$^{-1}$), colored yellow-red by the RMSD from the ground state 'A', shown in gray. (D) Free energy landscape of the $\triangle 6$ metadynamics ensemble, as a function of the collective variables ANTIBETARMSD, ALPHABETA1 and RGYR (see Methods). The ensemble structure of the ground state 'A' (–1.0 kcal mol$^{-1}$) is shown in Figure 2B. (E) Ensemble structure of the first excited state 'B' (–0.36 kcal mol$^{-1}$), colored yellow-red by the RMSD from the ground state 'A', shown in gray. (F) Ensemble structure of the second excited state 'C' (–0.07 kcal mol$^{-1}$), colored yellow-red by the RMSD from the ground state 'A', shown in grey. (G,H) Convergence of the free energy calculations. One-dimensional free energy landscapes of the four collective variables (see Methods), for (G) FL and (H) $\triangle 6$, averaged over the final segment (the last 80 ns) of the metadynamics simulations, with standard deviations reported as error bars.

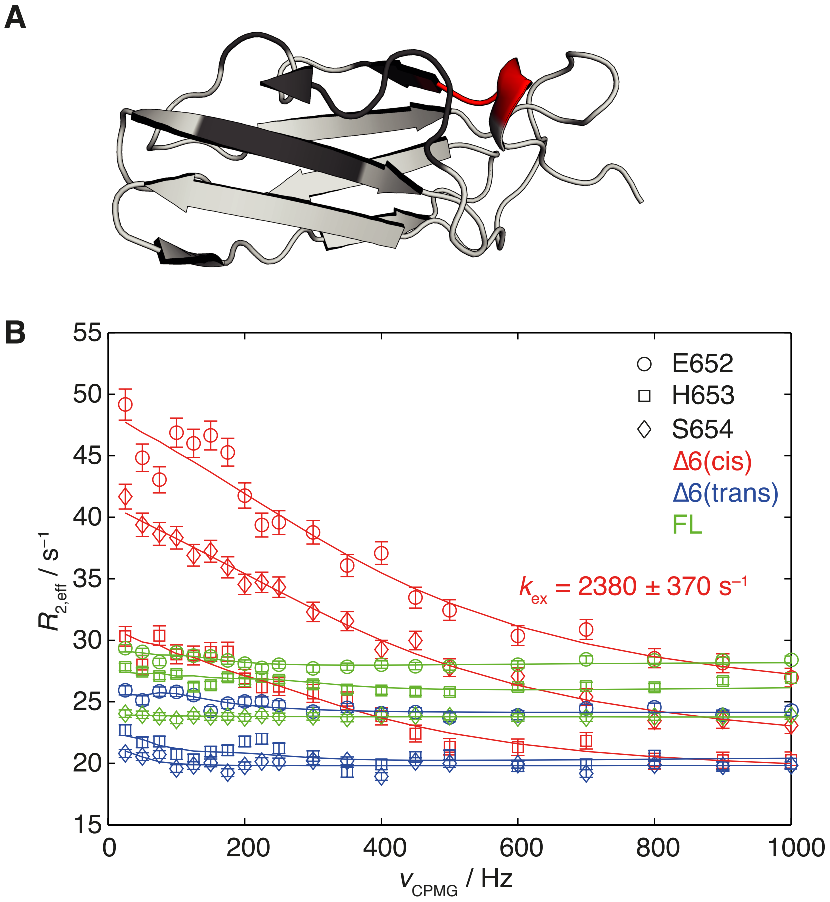

**Fig. S5.** Millisecond timescale dynamics in the FLN5 Δ6 *cis* state. (A) FL crystal structure[20] showing residues 652–654 (red), and residues broadened beyond detection in the Δ6 N state (black). (B) $^{15}$N CPMG relaxation dispersion profiles (700 MHz, 283 K) for residues E652, H653 and S654 in FL FLN5 (green) and Δ6 I and N states (blue and red respectively). Solid lines show global fits of each state to a two-state exchange process.

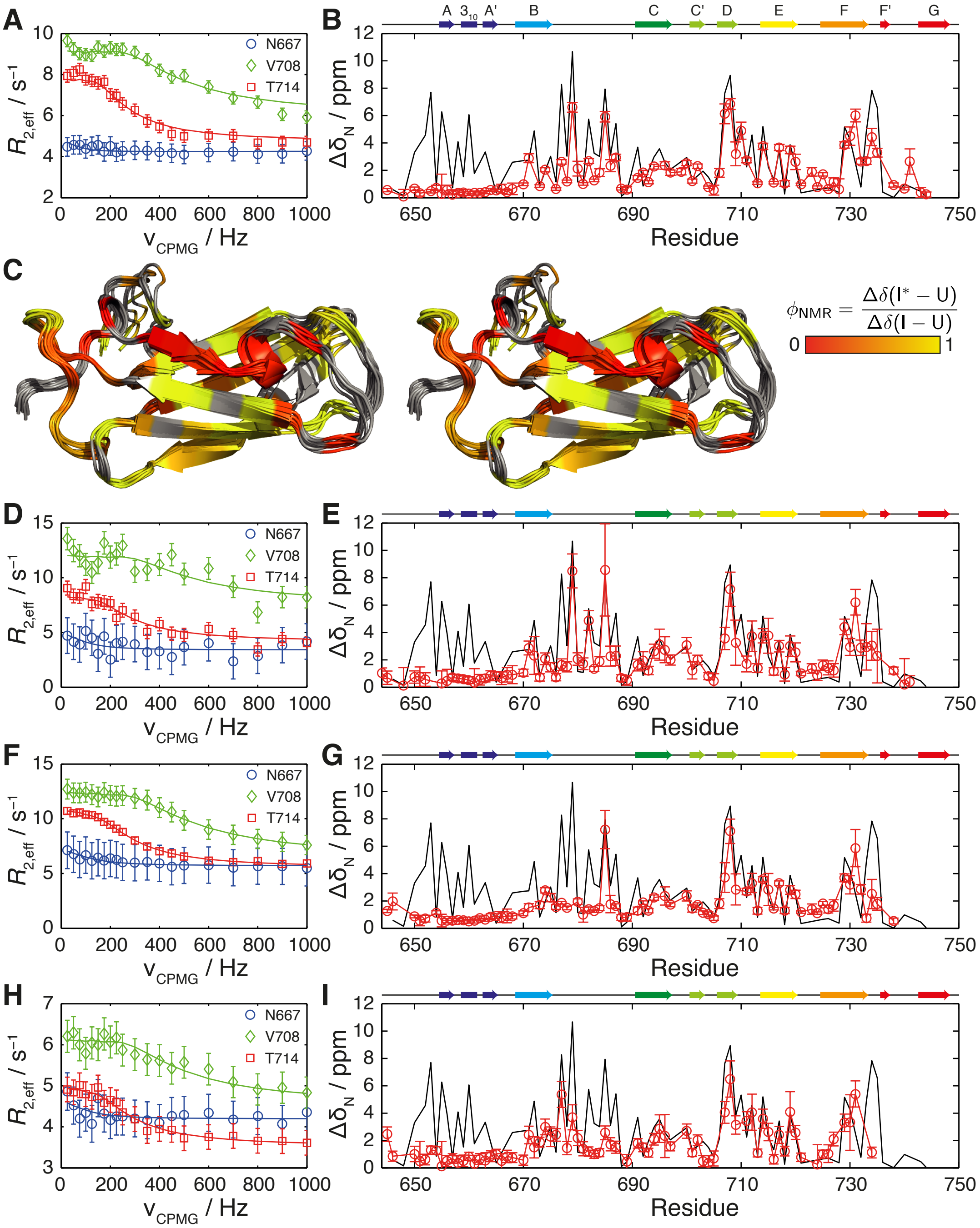

**Fig. S6.** CPMG relaxation dispersion measurements of unfolded states. (A,B) Δ6, (D,E) Δ9, (F,G) Δ12, (H,I) Δ16. (A,D,F,H) $^{15}$N CPMG relaxation dispersion profiles (700 MHz, 283 K) of representative residues. (B,E,G,I) Fitted $^{15}$N chemical shift differences between U and I* states. Chemical shift differences between Δ6 I and U states are plotted in black for reference. (C) Stereoview of the Δ6 intermediate ensemble, colored as indicated according to $\phi_{\mathrm{NMR}}$, the relative chemical shift difference between I* and I states and the unfolded U state(32).

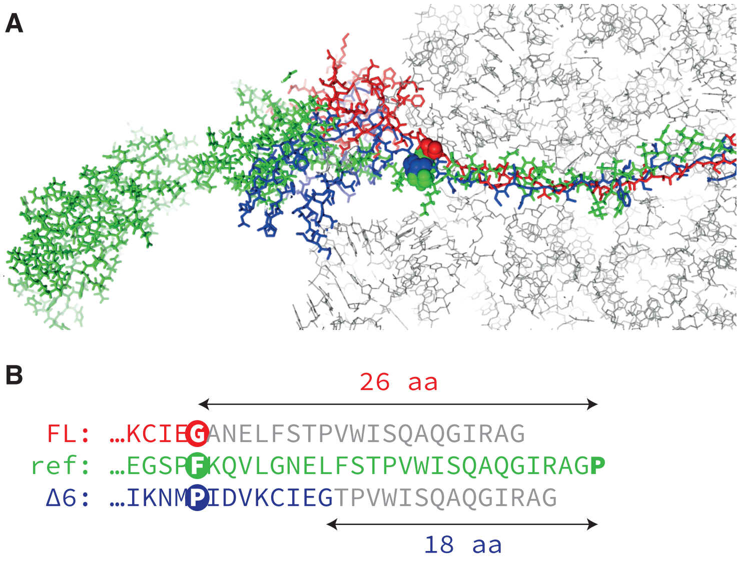

**Fig. S7.** (A) Comparison of NC linker conformations in structure-based models of the native (FL, red, 20 aa linker) and intermediate (Δ6, blue, 14 aa linker) states, with an all-atom, chemical shift-restrained molecular dynamics simulation of the FLN5+110 RNC (NC pdb 2n62, ribosome pdb 4ybb) (1). The locations of the last structured residues in the native and intermediate states are marked with spheres, as is the closest corresponding residue from the MD simulation. (B) C-terminal sequences of the NCs, aligned to the location of the last structured residues highlighted in (A). On the basis of this, the linker lengths that would be required for formation of the native and intermediate states, assuming a relaxed linker conformation comparable to that the of reference MD simulation, can be determined as indicated.

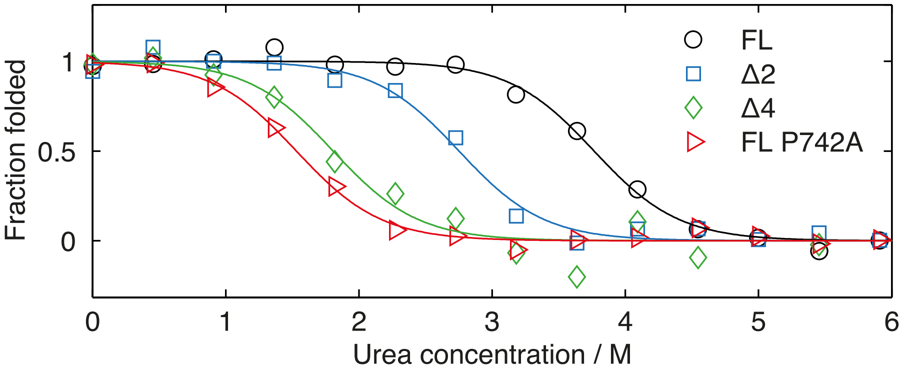

**Fig. S8.** Measurement of FLN5 FL and truncation stability by chemical denaturation in urea at 298 K. Measured CD signals at 234 nm were fitted globally to a two-state folding transition (5), with a shared $m$-value, $\langle m_{D-N} \rangle = 1.9 \pm 0.2$ kcal mol$^{-1}$.

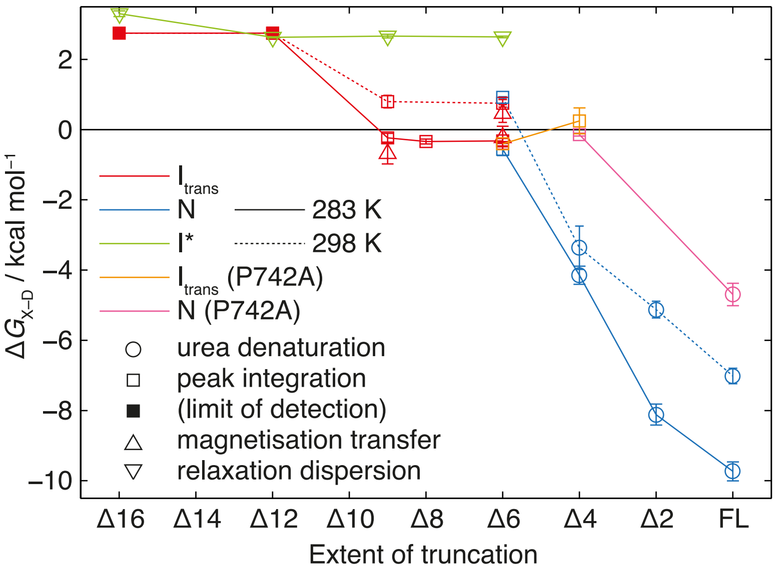

**Fig. S9.** Thermodynamic characterization of folding in FLN5 truncation variants at 283 K. Free energies of folding to N, I and I* states, in WT and P742A FLN5, were determined by a range of experimental methods, shown by symbols as indicated in the legend.

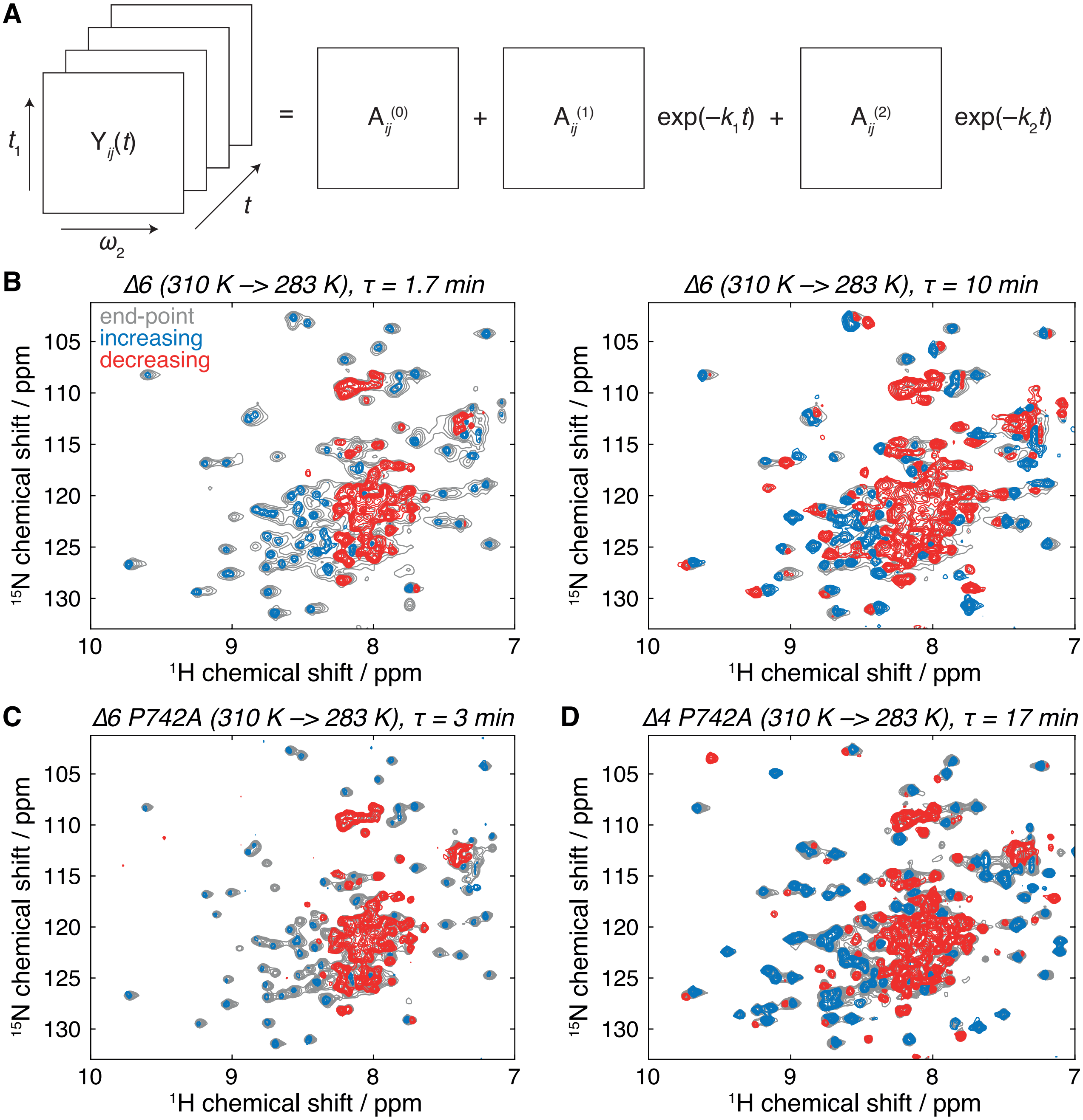

**Fig. S10.** Real-time NMR measurement of proline isomerization kinetics following temperature jumps. (A) Schematic illustration of the analysis method, in which spectra over time are decomposed into component spectra and associated rate constants. (B) Component amplitude spectra determined from the analysis of $\Delta$6 refolding following a temperature jump from 310 K to 283 K. Two phases were identified, with time constants of $1.8 \pm 0.1$ min and $9.8 \pm 0.1$ min. Peaks that increase or decrease over the reaction time-course are shown in blue or red respectively. The spectrum at equilibrium is shown in gray. (C) Component amplitude spectrum determined from the analysis of $\Delta$6 P742A refolding following a temperature jump from 310 K to 283 K. A single phase was identified with a time constant of $2.7 \pm 0.3$ min. (D) Component amplitude spectrum determined from the analysis of $\Delta$4 P742A refolding following a temperature jump from 310 K to 283 K. A single phase was identified with a time constant of $17.3 \pm 0.1$ min.

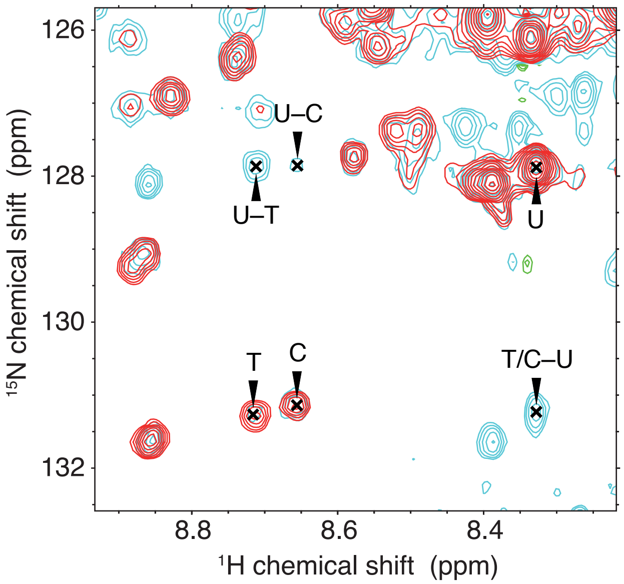

**Fig. S11.** $N_z$ magnetisation transfer experiment ([6]) acquired at 298 K for $^{15}$N-labelled $\Delta 6$ (150 $\mu$M) and 0.13 equivalents unlabelled cyclophilin A (a.k.a. peptidyl-prolyl isomerase A, 20 $\mu$M) with exchange times of 10 ms (cyan) and 400 ms (red). Arrows indicate cross-peaks and exchange between unfolded (U), *trans* intermediate (I) and native *cis* (N) state resonances of A694.

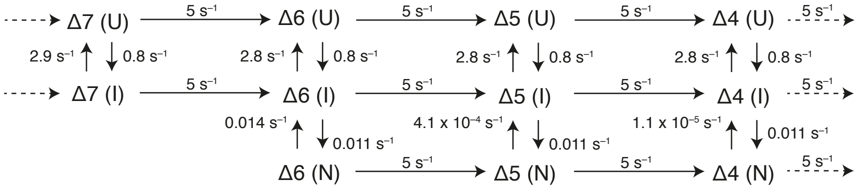

**Fig. S12.** Markov chain modelling of the non-equilibrium co-translational folding pathway. Fragment of kinetic pathway based upon measured folding rates, with linear interpolation between measured stabilities, and incorporating an irreversible translation process ($k_{\mathrm{trans}} = 5$ amino acids s$^{-1}$).

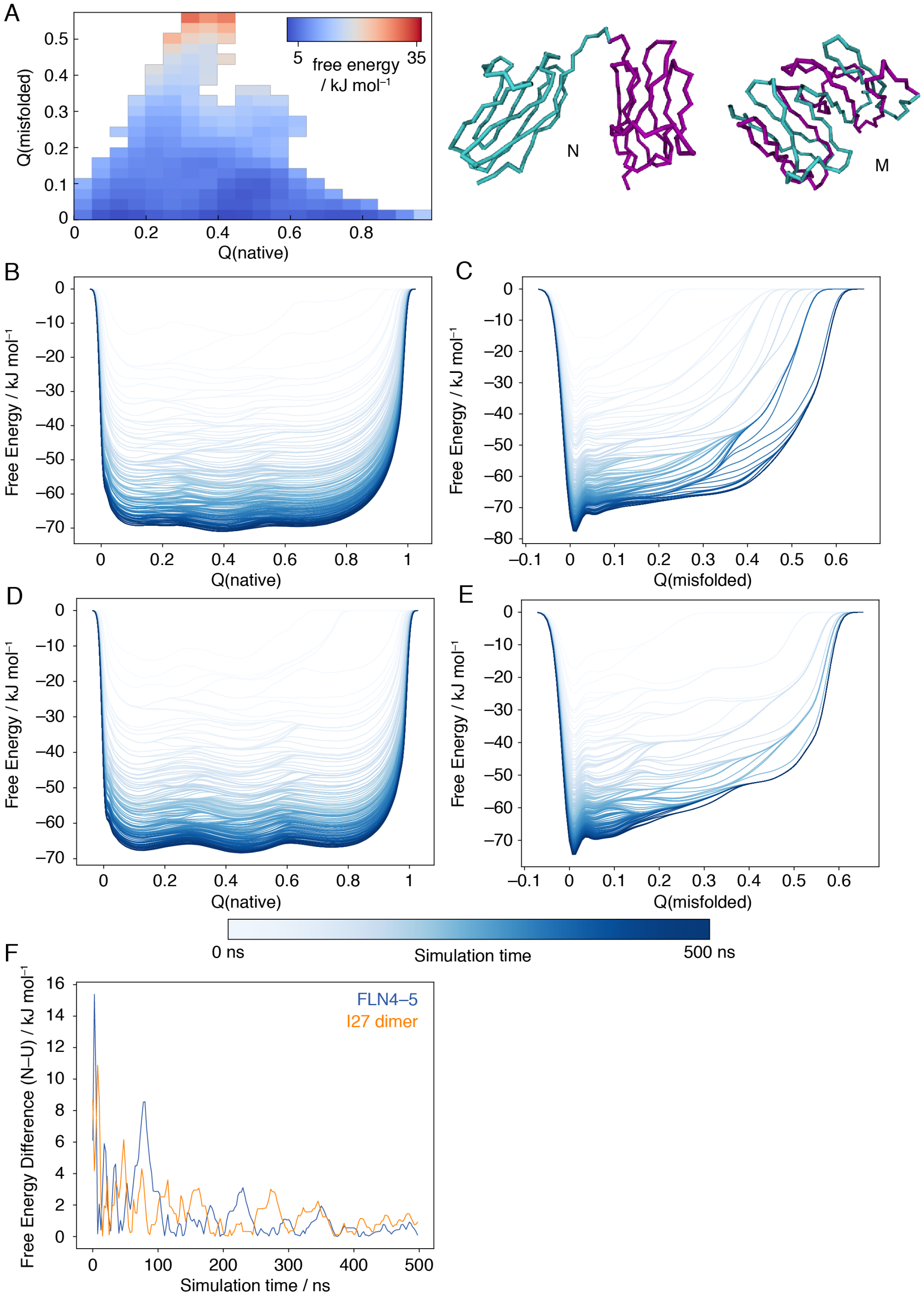

**Fig. S13.** MD simulations of interdomain misfolding. (A) Free energy landscape calculated for a titin I27 dimer by coarse-grained MD simulations, plotted against the fraction of native and misfolded contacts. Representative structures of native and misfolded states are shown as labelled (NTD, cyan; CTD, purple). (B-E) Changes in projected one-dimensional free energy landscapes as a function of simulation time for FLN4-5 (B,C) and I27 dimers (D,E). (F) Free energy differences between the native basin $(Q_{\mathrm{native}} > 0.6)$ and unfolded basin $(Q_{\mathrm{native}} < 0.3)$ calculated for FLN4-5 and I27 dimers as a function of simulation time.

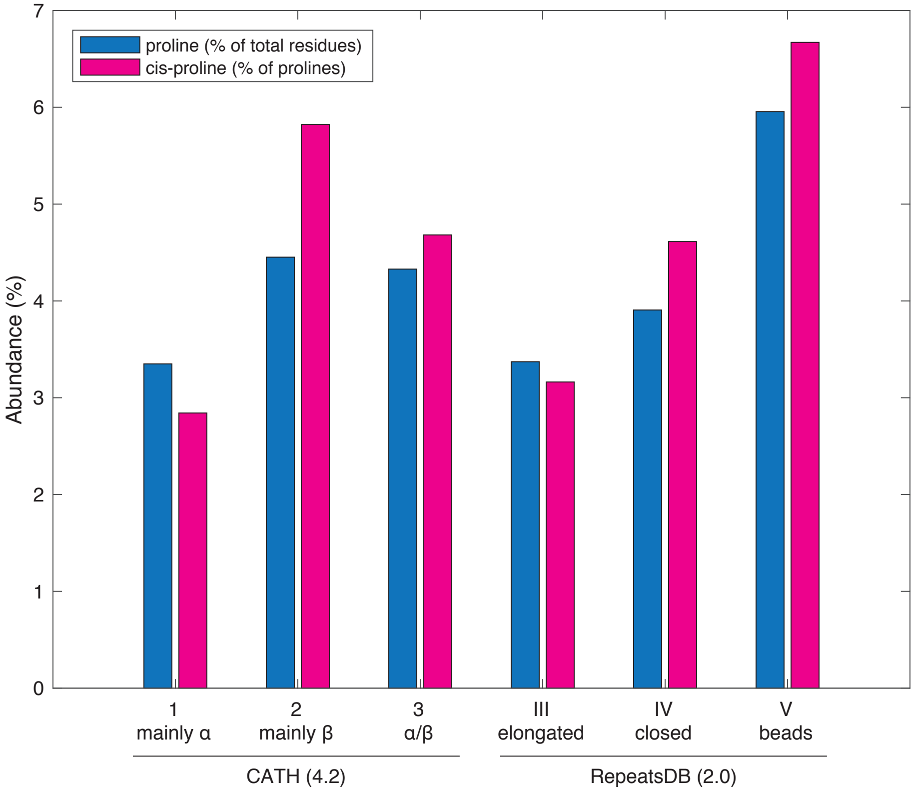

**Fig. S14.** Frequency of proline and *cis*-proline occurrence within repeat proteins (RepeatsDB v2.0) and the CATH database of protein domains (CATH v4.2 S35 set) (Table S3).

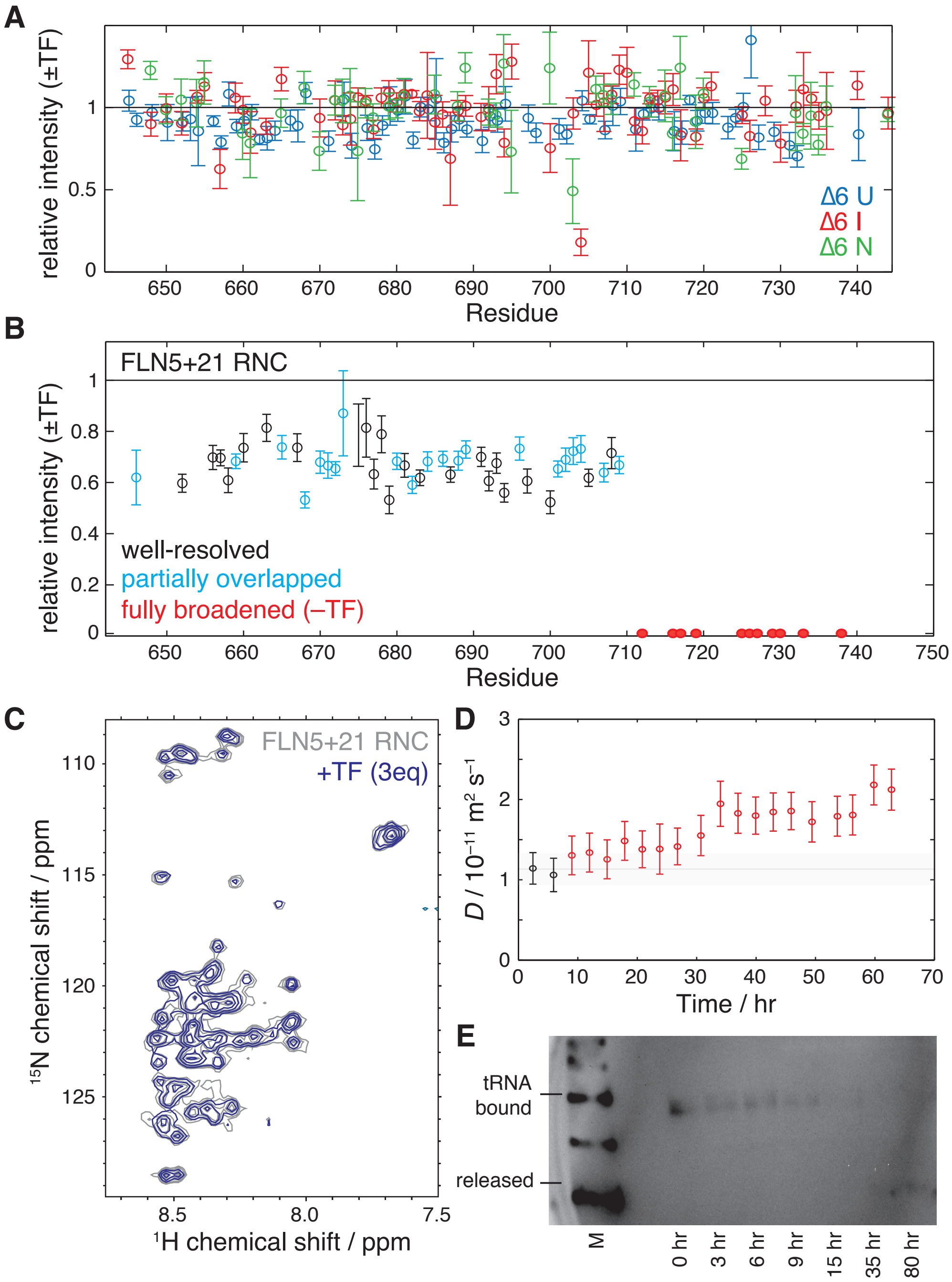

**Fig. S15.** Interaction of TF with FLN5 Δ6 and +21 RNC. (A) Relative changes in $^1$H,$^{15}$N HSQC resonance intensities of unfolded, intermediate and native states of Δ6 (50 $\mu$M) upon addition of 1 equivalent TF (283 K). (B) Relative changes in $^1$H,$^{15}$N SOFAST-HMQC resonance intensities of a disordered FLN5+21 RNC (10 $\mu$M) upon addition of 3 equivalents TF (283 K). Uncertainties are calculated from spectrum noise levels, but resonances partially overlapped with other NC or 70S background resonances are indicated and may therefore have larger uncertainties. The location of well-resolved NC resonances that were broadened and undetectable in the NC prior to addition of TF are marked in red, and should not be interpreted as having zero relative intensity following addition of TF. (C) $^1$H,$^{15}$N SOFAST-HMQC spectrum of the +21 RNC before (grey) and following (blue) addition of 3 equivalents TF. (D) RNC diffusion coefficients (measured with interleaved $^{15}$N SORDID experiments) observed over the acquisition timecourse. Black points indicate the data used in calculating relative intensities in panel B (12). (E) Anti-His$_6$ detected western blot of the +21 RNC stability over time (12).