SUPPORTING INFORMATION

Confinement and stabilization of a Fyn SH3 folding intermediate mimetics within the cavity of the chaperonin GroEL demonstrated by relaxation-based NMR

David S. Libich,[1] Vitali Tugarinov,[1] Rodolfo Ghirlando,[2] and G. Marius Clore[1,*]

Laboratories of Chemical Physics[1] and Molecular Biology,[2] National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

Cloning, expression and purification of truncated Fyn SH3 Domains

Plasmids encoding the wild type (SH3[WT]) and triple A39V/N53P/V55L mutant (SH3[Mut]) SH3 domains (residues 85-142, numbered 2-59 in the current work) of the *Gallus gallus* tyrosine kinase Fyn[1] were synthesized by Genscript. The gene sequences were codon optimized for expression in *E. coli* and included an N-terminal His$_6$-tag and a tobacco etch virus (TEV) protease cleavage site[2] as previously described.[3] The codon encoding Ala55 or Pro56 was replaced by a stop codon using the QuickChange mutagenesis kit to generate the Δ56 and Δ57 variants respectively, for both SH3[WT] and SH3[Mut]. The resulting plasmids were transformed into *E. coli* BL21 Star (DE3) competent cells (Life Technologies). Expression and purification of the truncation mutants followed the procedure described previously for full length SH3[Mut].[3] Briefly, *E. coli* containing the appropriate plasmid was grown at 37°C in M9 minimal medium, supplemented with either $^{15}NH_4Cl$ and/or $^{13}C_6$-glucose as the sole nitrogen and carbon sources, respectively, to an OD$_{600}$ ~ 0.6, induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and grown for a further 4 h. Cells were harvested via centrifugation, resuspended in a denaturing buffer, lysed by two passes through a high-pressure homogenizer (Avestin), and initially fractionated by passing the cleared lysate over a 5 mL HisTrap column (GE Healthcare). Fractions containing the SH3 protein were pooled, refolded in a step dialysis, and had the His$_6$-tag removed by TEV protease. A second pass over a HisTrap column removed the His$_6$-tag remnant, and the flow-through, containing the SH3 domain, was concentrated and applied to a Superdex 75 HiLoad size exclusion column as a final purification step. All purified SH3 domains were subjected to liquid chromatography/mass spectrometry to confirm correct molecular weight, mutations and isotopic incorporation.

Expression and purification of GroEL and ribulose-1,5-bisphosphate carboxylase

GroEL and ribulose-1,5-bisphosphate carboxylase (Rubisco) were expressed in an *E. coli* recombinant system and purified as previously described.[3-4] For the quantitative binding studies described in this work, endogenous *E. coli* proteins remaining in the cavity of GroEL after initial purification were removed by subjecting purified GroEL to two rounds of acetone (45% v/v) precipitation.[3,4b] The removal of contaminating proteins was monitored by tryptophan fluorescence, a sensitive diagnostic of extraneous protein contamination since GroEL does not contain any native tryptophans. Purification of Rubisco followed a previously published protocol.[4a] Purified Rubisco was buffer exchanged into 2 mM Tris pH

7.4 (using a G25 HiPrep Desalting column, GE Healthcare), aliquoted, and flash frozen in liquid nitrogen. Prior to titration into GroEL, Rubisco was acid denatured by adding equal volumes of 20 mM HCl.[3-4]

NMR sample preparation

The four SH3 truncation mutants (SH3^{WT}Δ56, SH3^{WT}Δ57, SH3^{Mut}Δ56 and SH3^{Mut}Δ57) have the potential to form protein fibrils.[1b] Preliminary NMR experiments were therefore conducted on samples of these four constructs at concentrations ranging from 100 to 500 µM in NMR buffer comprising 50 mM sodium phosphate, pH 7.0, 0.2 mM EDTA, 0.5% (w/v) NaN$_3$ and 90%(v/v) H$_2$O/10% (v/v) D$_2$O. Temperatures were varied between 283 K and 303 K. Fibril formation was monitored by measuring the reduction in ^{1}H-^{15}N cross-peak intensity in ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) correlation spectra and confirmed by both analytical ultracentrifugation and negative stain electron microscopy. These initial experiments showed that the SH3^{WT}Δ56 and SH3^{Mut}Δ57 constructs exhibit characteristics of slow to intermediate chemical exchange resulting in significant cross-peak line broadening, missing cross-peaks and overall poor spectral quality. For the SH3^{WT}Δ57 and SH3^{Mut}Δ56 constructs, however, no detectable fibril formation or aggregation occurred at 283 K over the course of the NMR experiments. Consequently all experiments with GroEL were conducted exclusively with the SH3^{WT}Δ57 and SH3^{Mut}Δ56 constructs. As an additional preventative measure against fibril contamination, NMR samples were only used for approximately two weeks and continuously monitored during that time from ^{1}H-^{15}N cross-peak intensities. NMR samples of either SH3^{WT}Δ57 or SH3^{Mut}Δ56 were carefully prepared in matched pairs consisting of a reference and GroEL-containing sample as described previously.[3] An aliquot of the appropriate SH3 domain stock solution in NMR buffer was diluted with the same buffer to 900 µL. This aliquot was divided into two equal fractions and added to either 100 µL NMR buffer or 100 µL of 0.66 mM GroEL (subunit concentration) in NMR buffer resulting in 550 µL SH3 concentration-matched reference and GroEL samples. In experiments designed to specifically block substrate access to the GroEL cavity, a stock solution of GroEL was saturated with stoichiometric amounts of acid-denatured Rubisco subsequent to the addition of the SH3 domains, as described previously.[3]

NMR spectroscopy

All NMR experiments were recorded on Bruker Avance III spectrometers operating at ^{1}H frequencies of 900.27, 800.1 or 600.13 MHz, each equipped with Bruker TCI z-axis gradient cryogenic probes. All experiments were recorded on samples dissolved in NMR buffer at 283 K unless otherwise noted. Temperature differences were corrected by matching the chemical shift difference between residual water and the methyl resonance of 4,4-dimethyl-4-silapentance-1-sulfonic acid (DSS) set at 0 ppm in a sample containing 50 mM sodium phosphate, pH 7.0, 0.2 mM EDTA, 0.5% (w/v) NaN$_3$, 99% (v/v) D$_2$O. All spectra were processed using the NMRPipe/NMRDraw software package.[5]

Backbone assignment and structure determination of SH3^{WT}Δ57 and SH3^{Mut}Δ56

Backbone (^{13}Cα, ^{13}Cβ, ^{13}C', N and H$_N$) sequential resonance assignments were derived from analysis of a set of standard three-dimensional triple resonance experiments (HNCO, CBCACONH and HNCACB).[6] Backbone amide ($^1D_{NH}$) residual dipolar couplings (RDCs) were obtained from the difference in $^1J_{NH}$ couplings between aligned (~11 mg/ml phage pf1, ASLA Biotech)[7] and isotropic media. $^1J_{NH}$ couplings were determined from the doublet splitting observed in ^{1}H-^{15}N HSQC in-phase/anti-phase (IPAP) spectra measured at 600 MHz.[8] Three-dimensional structures were calculated from the backbone chemical shifts and $^1D_{NH}$ RDCs using CS-Rosetta.[9] To reduce any possible bias from known SH3 domain structures, all structures containing an SH3 domain in the PDB were excluded from the fragment generation step in the

CS-Rosetta protocol. The backbone chemical shifts for SH3^{Mut}Δ56 and SH3^{WT}Δ57 have been deposited in the BMRB (accession numbers 26954 and 26955, respectively).

^{15}N Relaxation Measurements

The ^{15}N-$R_{1\rho}$ and R_1 relaxation rates were measured on 100 µM SH3^{WT}Δ57 and SH3^{Mut}Δ56 ^{15}N-labeled samples free and in the presence of either 120 µM (in subunits) GroEL or 120 µM (in subunits) GroEL and 19 µM acid-denatured Rubisco. The pulse schemes used for measuring ^{15}N $R_{1\rho}$ and ^{15}N R_1 rates[10] were HSQC-based, heat compensated and recorded in an interleaved manner as described previously.[3] An effective spin-lock field of 1.8 kHz was used in the ^{15}N-$R_{1\rho}$ experiments to completely suppress contributions from chemical exchange-induced line broadening. The relaxation decays were fit to a single exponential function $A\mathrm{e}^{-RT}$, where R is a relaxation rate and T the relaxation delay, and errors were estimated by Monte Carlo simulations. ^{15}N-R_2 (in-phase) values were calculated using the relationship $R_2 = (R_{1\rho} - R_1\cos^2\theta)/\sin^2\theta$, where θ is the angle between the effective spin-lock field and the external magnetic (B$_0$) field. ^{15}N-ΔR_2 values were calculated as the differences in ^{15}N-R_2 values measured in the presence and absence of GroEL (or GroEL + Rubisco) with appropriate propagation of errors.

^{15}N-DEST measurements

^{15}N-DEST experiments were recorded at 900 MHz on samples containing 100 µM ^{15}N-labeled SH3^{WT}Δ57 and SH3^{Mut}Δ56 samples in the presence of 120 µM (in subunits) GroEL using our previously described pulse sequence.[11] DEST profiles were acquired in an interleaved manner using ^{15}N saturation field strengths of 500 or 750 Hz applied for 0.7 s at ±20, ±15, ±10, ±8, ±6, ±5, ±4, ±3, ±2, ±1 and 0 kHz from the ^{15}N carrier frequency. An addition two reference experiments were recorded at ±20 kHz with the ^{15}N saturation field strength set to 0 Hz.

^{15}N Relaxation Dispersion Measurements

^{15}N Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersions were acquired at 600 and 900 MHz on samples of 100 µM ^{15}N-labeled SH3^{WT}Δ57 and SH3^{Mut}Δ56 free and in the presence of either 120 µM (in subunits) GroEL or 120 µM (in subunits) GroEL + 19 µM Rubisco. To quantify the extent of off-pathway dimerization of the SH3^{Mut}Δ56 construct, ^{15}N relaxation dispersion data were also recorded on 350 and 600 µM samples of SH3^{Mut}Δ56 (in the absence GroEL). The ^{15}N relaxation dispersion data were obtained using a pulse scheme with amide proton decoupling to measure the rates of in-phase ^{15}N coherences.[12] CPMG field strengths ($\nu_{CPMG} = 1/2\tau_{CP}$ where τ_{CP} is the time between 180° pulses) of 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000 Hz were applied for a constant relaxation period of 40 ms. ^{1}H$_N$-CW decoupling was applied at a radiofrequency strength of 11 kHz during the constant relaxation period. One experiment with the relaxation period omitted was also recorded and served as a reference for the calculation of effective R_2 as a function of applied CPMG field strength.[12]

Sedimentation velocity

Sedimentation velocity experiments were conducted at 50,000 rpm and 10°C on a Beckman Coulter ProteomeLab XL-I analytical ultracentrifuge following standard protocols.[13] A stock solution of ^{15}N-labeled SH3^{Mut}Δ56 in 50 mM sodium phosphate, pH 7.0, 0.2 mM EDTA, 0.05% (w/v) sodium azide and 90% (v/v) H$_2$O/10% (v/v) D$_2$O along with matching solvent buffer was used to prepare a series of SH3^{Mut}Δ56 samples at different concentrations. Samples at concentrations of 40, 80, 165 and 470 µM were loaded into 2-channel, 3 mm path-length sector shaped cells and thermally equilibrated at zero speed. Absorbance and interference velocity scans were subsequently acquired at approximately 3.5 min.

intervals – absorbance data were collected in a continuous mode as single measurements at 280 nm using a radial spacing of 0.003 cm. Absorbance data were not collected for the sample at 470 μM. Time corrected[14] data were analyzed in SEDFIT 15.01c[15] in terms of a continuous $c(s)$ distribution of sedimenting species using an s range of 0 to 4 with a linear resolution of 100 and a maximum entropy regularization confidence interval of 0.68. In all cases, excellent fits were observed with root mean square deviations ranging from 0.0039 – 0.0072 absorbance units and 0.0015 – 0.0076 fringes. The partial specific volume of SH3^{Mut}Δ56 was calculated based on its amino acid composition using SEDNTERP[16] (http://sednterp.unh.edu/) and corrected to account for the ^{15}N labeling and partial deuteration of exchangeable protons. The solvent density ρ and viscosity η were determined experimentally at 20°C on an Anton Paar DMA 5000 density meter and Anton Paar AMVn rolling ball viscometer, respectively, and corrected to values at 10°C using standard tables. Sedimentation coefficients were corrected to standard conditions in water at 20°C, $s_{20,w}$.

Simultaneous analysis of ^{15}N-CPMG relaxation dispersion, DEST and ΔR_2 data

The general kinetic scheme used to analyze the SH3^{Mut}Δ56 experimental data is a 5-state scheme represented by (see also Fig. 3 of main text):
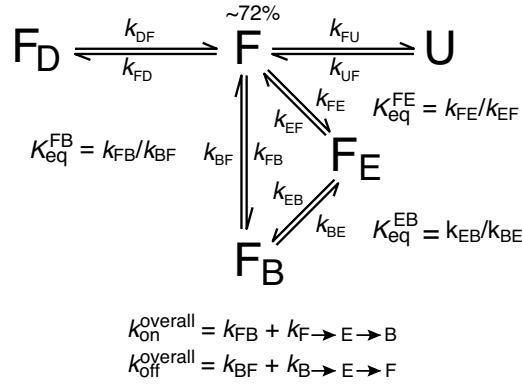
Scheme S1

All the experimental data for SH3^{Mut}Δ56 in the presence and the absence of GroEL were fit simultaneously by minimizing the following sum of squared differences between the observed and calculated values of the experimental observables, using an in-house MatLab program:

$$\chi^2 = \alpha_1 \sum_i \sum_k \sum_{j=1}^2 \left(\frac{R_{2,\text{Mut}\Delta56+\text{GroEL}}^{\text{obs},i,k,j} - R_{2,\text{Mut}\Delta56+\text{GroEL}}^{\text{calc},i,k,j}}{\sigma_{R_{2,\text{Mut}\Delta56+\text{GroEL}}^{\text{obs},i,k,j}}} \right)^2 + \alpha_2 \sum_i \sum_k \sum_{j=1}^2 \left(\frac{R_{2,\text{Mut}\Delta56\ \text{free}}^{\text{obs},i,k,j} - R_{2,\text{Mut}\Delta56\ \text{free}}^{\text{calc,3state},i,k,j}}{\sigma_{R_{2,\text{Mut}\Delta56\ \text{free}}^{\text{obs},i,k,j}}} \right)^2$$

$$+\alpha_3 \sum_i \sum_{j=1}^2 \left(\frac{\Delta R_{2,\text{Mut}\Delta56+\text{GroEL}}^{\text{obs},i,j} - \Delta R_{2,\text{Mut}\Delta56+\text{GroEL}}^{\text{calc},i,j}}{\sigma_{\Delta R_{2,\text{Mut}\Delta56+\text{GroEL}}^{\text{obs},i,j}}} \right)^2 + \alpha_4 \sum_i \sum_l \sum_{m=1}^2 \left(\kappa_{\text{Mut}\Delta56+\text{GroEL}}^{\text{obs},i,l,m} - \kappa_{\text{Mut}\Delta56+\text{GroEL}}^{\text{calc},i,l,m} \right)^2$$

$$(S1)$$

where $R_{2,n}(\Delta R_{2,n})$ denotes the $R_2(\Delta R_2)$ of state (sample) n; κ denotes normalized experimental intensities of DEST profiles at each offset frequency; the subscripts i, j, k, l, and m refer to residue number, ^{1}H spectrometer frequency (600 and 900 Hz), CPMG RF field strength, DEST offset, and DEST saturation RF field strength (500 and 750 Hz), respectively; and α_1, α_2, α_3, and α_4 represent empirically determined factors used to appropriately weight the different data types and have numerical values of 1, 3, 0.3, and 1, respectively. Thus, the CPMG data obtained on SH3$^{Mut}\Delta$56 in the absence of GroEL ($R_{2,\ \Delta56\ free}$; the second term in Eq. S1) are used to characterize the transition $\mathbf{F_D} \leftrightarrow \mathbf{F} \leftrightarrow \mathbf{U}$ via a 3-state fit providing the rate constants {k_{FU}, k_{UF}, k_{FD}, k_{DF}} as well as the (absolute values of) chemical shift differences between the folded and unfolded states ($\Delta\varpi^{F\text{-}U}$) and between the folded monomer and the dimer ($\Delta\varpi^{F\text{-}F_D}$). These parameters are dynamically 'shared' within the set of optimized parameters in the combined minimization of the target function for all the data for the SH3$^{Mut}\Delta$56+GroEL sample (Eq. S1). The space of residue-specific fitted parameters included: {R_{2,F_B}^{900}; R_{2,F_B}^{600}; $R_{2,F;CPMG}^{900}$; $R_{2,F;CPMG}^{600}$; R_{2,F_E}^{900}; R_{2,F_E}^{600}; $\Delta\varpi^{F\text{-}U}$; $\Delta\varpi^{F\text{-}F_D}$; $\Delta\varpi^{F\text{-}F_B}$}, where $R_{2,n}$ denotes the R_2 of state n and the superscripts denote spectrometer fields; $\Delta\varpi^{mn}$ are the differences in chemical shifts between states n and m. The residue-specific parameters $R_{2,F;CPMG}^{900}$ and $R_{2,F;CPMG}^{600}$ were allowed to vary because the CPMG-derived relaxation rates can differ from those measured by $R_{1\rho}$ techniques as supplied for analysis of DEST and ΔR_2 measurements. Importantly, the confined state $\mathbf{F_E}$ was treated as an *observable* state in analyses of all types of data included in Eq. S1, *i.e.* the intensities of the states $\mathbf{F}$ and $\mathbf{F_E}$ were summed up in the calculations of effective R_2 rates (CPMG), and the values of ΔR_2 and κ (DEST) as explained in more detail below.

The maximal set of global variable parameters thus includes: {k_{FU}, k_{UF}, k_{FD}, k_{DF}, k_{FE}, k_{EF}, k_{EB}, k_{BE}, k_{FB}}, where all the rate constants are defined in Scheme S1 (as well as in Fig. 3 of the main text) and the remaining rate constant k_{BF} is calculated from the material balance relationship, $k_{BF} = (k_{FB}k_{BE}k_{EF})/(k_{FE}k_{EB})$. (Note that k_{FD}, k_{FE} and k_{FB} are pseudo-first order rate constants; k_{FD} for the formation of the dimer species is given by $2k_{FD}^*[F]$ where k_{FD}^* is the second-order association rate constant for dimer formation, and [F] is the concentration of free F.). It is clear that the complexity of this binding scheme does not allow us to determine all the involved rate constants from the available data. Nevertheless, the equilibrium constants for the $\mathbf{F} \leftrightarrow \mathbf{F_E} \leftrightarrow \mathbf{F_B}$ and $\mathbf{F} \leftrightarrow \mathbf{F_B}$ transitions can be defined with reasonable accuracy with the proviso that value of $<^{15}\text{N-}R_{2,F_B}>$ is restrained within a range consistent with the molecular weight (~800 kDa) of GroEL (i.e. ~950 s^{-1} at 900 MHz and 10 °C). The overall forward ($k_{on}^{overall}$) and backward ($k_{off}^{overall}$) rate constants for the formation of the GroEL-bound state $\mathbf{F_B}$ is given by the sum of the rate constants for the direct ($\mathbf{F} \leftrightarrow \mathbf{F_B}$) and indirect ($\mathbf{F} \leftrightarrow \mathbf{F_E} \leftrightarrow \mathbf{F_B}$) pathways:

$$k_{on}^{overall} = k_{FB}^{direct} + k_{F\to E\to B}^{indirect} = k_{FB} + k_{FE}k_{EB}/(k_{EF} + k_{EB}) \tag{S2.1}$$

$$k_{off}^{overall} = k_{BF}^{direct} + k_{B\to E\to F}^{indirect} = k_{BF} + k_{EF}k_{BE}/(k_{EF} + k_{EB}) \tag{S2.2}$$

While the relative contributions of the direct and indirect two pathways cannot be determined from the data, multiple minimizations of the target function in Eq. S1 with two of the rate constants fixed, shows that ~ 4 s^{-1} $< k_{on}^{overall} < \sim 6$ s^{-1}, and ~ 100 s^{-1} $< k_{off}^{overall} < \sim 160$ s^{-1}. Typically, minimizations were conducted fixing the rate constant k_{EF} at a number of values ranging from very high (3000 s^{-1}) to very low (15 s^{-1}) values, while k_{FB} was fixed at values ranging from 1 to 5 s^{-1}, effectively ensuring that the results with very different relative contributions of the two pathways are obtained (the upper limit of ~ 5 s^{-1} for $k_{on}^{overall}$ derives from the $\Delta R_{2,max}$ in the exchange regime where ΔR_2 values are nearly field-independent; cf. Fig. 1B in main text). It is interesting to note that when k_{FE} is set equal to k_{FB} during minimization, the flux through the direct binding pathway ($\mathbf{F} \leftrightarrow \mathbf{F_B}$) is dominant ($\sim 90\%$). The equilibrium constants and populations for confinement by and binding to GroEL reported in Fig. 3 and the main text ($p_E \sim 21 \pm 3$ % and $p_B \sim 2.6 \pm 0.3$ %) are maintained to within errors in all these calculations.

The set of populations of all states in Scheme S1 (Fig. 3 main text) were expressed through the rate constants as follows. First, the population of state $\mathbf{F_E}$, p_E, is given by the relationship:

$$p_E = \left[1 + \frac{k_{EB}}{k_{BE}} + \frac{1}{K_{eq}^{FE}} \left(1 + \frac{k_{FU}}{k_{UF}} + \frac{k_{FD}}{k_{DF}} \right) \right]^{-1} \tag{S3}$$

Then, the populations of the remaining states $\mathbf{F}$, $\mathbf{U}$, $\mathbf{F_D}$, and $\mathbf{F_B}$ (denoted by p_F, p_U, p_D and p_B) were calculated from the relationships: $p_F = p_E / K_{eq}^{FE}$, $p_U = p_F k_{FU} / k_{UF}$, $p_D = p_F k_{FD} / k_{DF}$, and $p_B = p_E k_{EB} / k_{BE}$.

The expression for χ^2 for the fitting of the SH3$^{WT}\Delta57$ data to a 2-state model ($\mathbf{F} \leftrightarrow \mathbf{F_B}$) is given by,

$$\chi^2 = \alpha_1 \sum_i \sum_{j=1}^2 \left(\frac{\Delta R_{2,WT\Delta57+GroEL}^{obs,i,j} - \Delta R_{2,WT\Delta57+GroEL}^{calc,i,j}}{\sigma_{\Delta R_{2,WT\Delta57+GroEL}^{obs,i,j}}} \right)^2 + $$
$$\alpha_2 \sum_i \sum_l \sum_{m=1}^2 \left(\kappa_{WT\Delta57+GroEL}^{obs,i,l,m} - \kappa_{WT\Delta57+GroEL}^{calc,i,l,m} \right)^2 \tag{S4}$$

as neither dimerization nor unfolding were detectable in the CPMG relaxation dispersion data recorded on the SH3$^{WT}\Delta57$ sample. The DEST and ΔR_2 data for SH3$^{WT}\Delta57$ were acquired at 600 and 800 MHz ^{1}H spectrometer frequencies ($j = \{600; 800\}$ MHz) and the same saturation RF field strengths ($l = \{500; 750\}$ Hz), while α_1 and α_2 had empirically determined numerical values of 0.1 and 1, respectively. The space of residue-specific fitted parameters included: { R_{2,F_B}^{800}; R_{2,F_B}^{600} }, while the set of global fitted parameters was comprised of the rate constants describing the apparent association (forward) and dissociation (backward) processes, $\{k_{FB}; k_{BF}\}$. The minimization of the target function in Eq. S4 yielded $k_{FB} \sim 7.3 \pm 1$ s^{-1} and $k_{BF} \sim 500 \pm 100$ s^{-1} with $< R_{2,F_B}^{800} > \sim 870$ s^{-1} corresponding to a population of the bound state $p_B \sim 1.4$ % (see below and Fig. S9).

The uncertainties in the values of the optimized parameters, corresponding to confidence intervals of ±1 S.D. were determined from the Jacobian matrix of the non-linear fit. Convergence of the solution was confirmed by varying initial values for all parameters and obtaining the same solution within reported uncertainties. Below, we describe the details of how each of the data types (CPMG, DEST and ΔR_2) used in the calculation of the error functions in Eqs. S1 and S4, was modeled for comparison with the observed experimental values.

The evolution of magnetization during the CPMG constant time period for Scheme S1 can be represented by:

$$M(t) = (AA*A*A)^n M(0) \qquad (S5)$$

where $M = [M^{\mathrm{F}};M^{\mathrm{U}};M^{\mathrm{F_E}};M^{\mathrm{F_B}};M^{\mathrm{F_D}}]^T$; M^N denotes the transverse magnetization of state N; T denotes transposition; $A = \exp(-R\tau_{CP}/2)$; $A*$ is the complex conjugate of A, n is the number of CPMG cycles employed, τ_{CP} is the distance between two consecutive 180° pulses in the CPMG pulse train, and $R = R^{cs} + R^{rel} + R^{ex}$, where

$$R^{cs} = i\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \Delta\varpi^{\mathrm{F\text{-}U}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Delta\varpi^{\mathrm{F\text{-}F_B}} & 0 \\ 0 & 0 & 0 & 0 & \Delta\varpi^{\mathrm{F\text{-}F_D}} \end{bmatrix} \qquad (S6.1)$$

$$R^{rel} = \begin{bmatrix} R_2^{\mathrm{F}} & 0 & 0 & 0 & 0 \\ 0 & R_2^{\mathrm{U}} & 0 & 0 & 0 \\ 0 & 0 & R_2^{\mathrm{F_E}} & 0 & 0 \\ 0 & 0 & 0 & R_2^{\mathrm{F_B}} & 0 \\ 0 & 0 & 0 & 0 & R_2^{\mathrm{F_D}} \end{bmatrix} \qquad (S6.2)$$

$$R^{ex} = \begin{bmatrix} k_{\mathrm{FU}}+k_{\mathrm{FE}}+k_{\mathrm{FB}}+k_{\mathrm{FD}} & -k_{\mathrm{UF}} & -k_{\mathrm{EF}} & -k_{\mathrm{BF}} & -k_{\mathrm{DF}} \\ -k_{\mathrm{FU}} & k_{\mathrm{UF}} & 0 & 0 & 0 \\ -k_{\mathrm{FE}} & 0 & k_{\mathrm{EF}}+k_{\mathrm{EB}} & -k_{\mathrm{BE}} & 0 \\ -k_{\mathrm{FB}} & 0 & -k_{\mathrm{EB}} & k_{\mathrm{BE}}+k_{\mathrm{BF}} & 0 \\ -k_{\mathrm{FD}} & 0 & 0 & 0 & k_{\mathrm{DF}} \end{bmatrix} \qquad (S6.3)$$

where the chemical shifts of states $\mathbf{F}$ and $\mathbf{F_E}$ are assumed to be identical, R_2^n is the transverse relaxation rate of state n in the absence of exchange (the R_2 values for the dimeric state $\mathbf{F_D}$ are assumed to be twice

those of state **F**), and all the rate constants in the exchange matrix (Eq. S6.3) are defined in Scheme 1. Note that the R_2 values for the confined state $\mathbf{F_E}$ are variable parameters in the fit and found to be close to those for state **F**. Interestingly, Horwich and co-workers, using fluorescence anisotropy measurements, observed that hindered rotational tumbling in a putative GroEL-confined state leads to ~4-fold higher rotational correlation times relative to the free state.[17] Only states **F**, **U**, $\mathbf{F_E}$ and $\mathbf{F_D}$ were assumed to be present at the start of the CPMG train, $M(0) = (p_\mathrm{F}, p_\mathrm{U}, p_\mathrm{E}, 0, p_\mathrm{D})$. The R_2 values for state **F** and the differences in chemical shifts between states **F** and **U**/$\mathbf{F_D}$, as well as the rate constants k_FU, k_UF, k_FD, and k_DF, were shared in the minimization of the target function with the 3-state fits of the CPMG profiles obtained on the sample of free SH3$^\mathrm{Mut}$Δ56 at the same protein concentration (see Eq. S1). Of note, the confined state $\mathbf{F_E}$ was considered *observable* for the calculation of effective R_2 values, namely the *sum* of signal intensities of states **F** and $\mathbf{F_E}$ was used in the calculations of effective relaxation rates after the evolution of M calculated at each applied CPMG field strength (number of CPMG cycles n).

The time dependent magnetization in DEST and ΔR_2 experiments performed on the SH3$^\mathrm{Mut}$Δ56+GroEL and SH3$^\mathrm{WT}$Δ57+GroEL samples for a single isolated spin in exchange between 5 (Scheme S1) and 2 states, respectively, was represented by the appropriate set of homogeneous McConnell equations.[18] The DEST experimental observables, κ, namely the ratio of the signal intensity of the resonance as function of saturation offset and saturation field to that without saturation, were calculated as described previously,[3,11a] except that *both* states **F** and $\mathbf{F_E}$ were treated as observable states, *i.e.* the sum of the signal intensities of the states **F** and $\mathbf{F_E}$ was used in the calculation of κ with and without saturation. In the same manner, ΔR_2 values were calculated as described previously[11] using a simple two-time point single-exponential decay:

$$\Delta R_2 = \frac{\ln\{[I_x^\mathrm{F}(\tau_1) + I_x^{\mathrm{F_E}}(\tau_1)]/[I_x^\mathrm{F}(\tau_2) + I_x^{\mathrm{F_E}}(\tau_2)]\}}{(\tau_2 - \tau_1)} - R_2^\mathrm{F} \tag{S7}$$

where $I_x^\mathrm{N}(\tau)$ is the intensity of transverse magnetization of state N at time τ, the delays τ_1 and τ_2 were set to 10 and 100 ms, respectively (τ_1 was chosen to remove any small deviations from exponential behavior at very short delays, and τ_2 to match the order of magnitude of the experimental delays used to measure R_2). For the fitting of ΔR_2, all the states were assumed to have the same chemical shifts as contributions to R_2 arising from chemical exchange are to a good approximation completely suppressed in ^{15}N-$R_{1\rho}$ measurements.

Analysis of ^{15}N-DEST profiles and ΔR_2 values for both SH3$^\mathrm{Mut}$Δ56+GroEL and SH3$^\mathrm{WT}$Δ57+GroEL samples indicates that exchange between NMR visible (**F** and $\mathbf{F_E}$ with small R_2) and NMR "dark" ($\mathbf{F_B}$ with large R_2) states is in the slow-to-intermediate exchange regime ($k_\mathrm{ex} = k_\mathrm{on}^\mathrm{overall} + k_\mathrm{off}^\mathrm{overall} < R_{2,\mathrm{F_B}}$), as manifested by a weak dependence of ΔR_2 values in both samples on the spectrometer magnetic field strength (see Figs. 1B and S7A). In this regime, the DEST and ΔR_2 data are relatively insensitive to the

value of $k_{off}^{overall}$ (and, by extension, the population of the bound state $\mathbf{F_B}$), *i.e.* the values of $k_{off}^{overall}$ and R_{2,F_B} cannot be simultaneously defined without the use of additional assumptions about R_{2,F_B}. As the molecular size of the GroEL is known, we can estimate that the transverse relaxation rates (R_{2,F_B}) in the bound F_B state have values of ~950 s^{-1} at 10 °C and 900 MHz.[3] Fig. S9 shows the results of minimizations of the target functions defined in Eqs. S1 (red) and S4 (black) for the SH3$^{Mut}\Delta56$+GroEL and SH3$^{WT}\Delta57$+GroEL data, respectively, with the average $<R_{2,F_B}>$ values plotted as a function of the fractions of the bound state, p_B, that were fixed in the fitting. The regions highlighted in Fig. S9 show the ranges of p_B that correspond to realistic $<R_{2,F_B}>$ values for the complexes at 800/900 MHz and 10 °C. Note that the binding of SH3$^{WT}\Delta57$ to GroEL is about 2-fold weaker than that of SH3$^{Mut}\Delta56$.

**Stabilization of the folded state within the GroEL cavity.**

The thermodynamics of stabilization of protein folded states upon confinement in the GroEL cavity, in the absence of any compensatory attractive interactions between the unfolded state and the walls of the cavity, can be estimated using the random-flight Gaussian chain model proposed by Zhou and Dill.[19] For small proteins (less than 350 residues), the energy of stabilization upon confinement ($\Delta\Delta G$) is essentially independent of cavity shape. The value of $\Delta\Delta G$ for SH3$^{Mut}\Delta56$ can therefore be estimated to be about -2.2 kcal·mol^{-1} using the relationship $\Delta\Delta G^{sphere}/RT = -2p^2Nb^2/3d^2 - 3\ln(1 - 2R_{gyr}/d) + \ln(6/\pi^2)$ for a spherical cavity, where d is the diameter of the sphere (69.4 Å), b the effective virtual bond length for the unfolded state (8 Å), R_{gyr} the radius of gyration of the folded state (=3.73$N^{1/3}$ = 14.3 Å, where the number of residues N = 56), R the gas constant, and T the temperature in Kelvin (283 K).[19] Using the relationship $K_{eq}^{F-U} / K_{eq}^{F_E-U_E} = \exp(-\Delta\Delta G / RT)$, where K_{eq}^{F-U} and $K_{eq}^{F_E-U_E}$ are the equilibrium constants for the folding-unfolding transition in the free and confined states, respectively, one can estimate the ratio of the equilibrium constants to be ~50, and, hence, the population of the confined unfolded state U_E (Fig. 3, main text) is predicted to be only ~0.017%.

Alternative kinetic scheme

An alternative scheme in which the confined states F_E and U_E are omitted, can partially account for the observed decrease in R_{ex} in the presence of apo GroEL (see Fig. 2B of main text) if the folded state is dramatically destabilized upon binding GroEL and the equilibrium between folded and unfolded GroEL-bound states concomitantly shifted by a factor of >100 towards the unfolded state. This alternative scheme has a higher value for the penalty function than that shown in Scheme S1 (Fig. 3 of main text) and fails to account simultaneously for the observed ΔR_2 and DEST data; in addition, such a scheme would also be inconsistent with the ^{15}N-ΔR_2 and DEST data for SH3$^{WT}\Delta57$ that shows binding of the folded state to GroEL since there is no evidence for the existence of an unfolded state from relaxation dispersion.
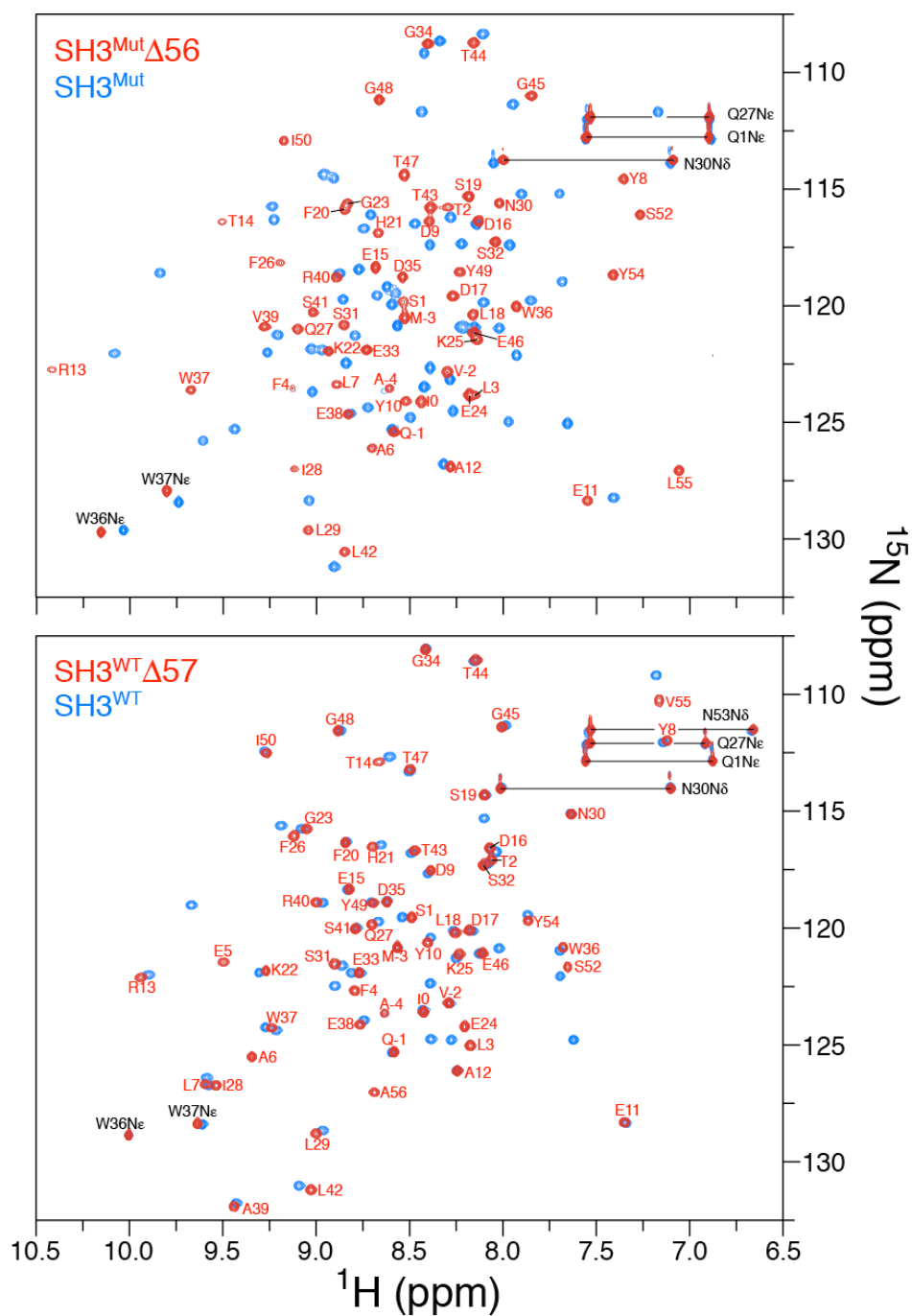
Figure S1. ¹H-¹⁵N HSQC spectra of ¹⁵N-labeled SH3^Mut^Δ56 (top panel, red contours) and ¹⁵N SH3^WT^Δ57 (bottom panel, red contours) overlaid on the spectra of the corresponding full-length homologues for each respective protein (blue contours). Labels in red are backbone assignments for the truncated constructs; labels in black relate to side chains. Sample concentrations were 100 µM and the temperature was 10°C.
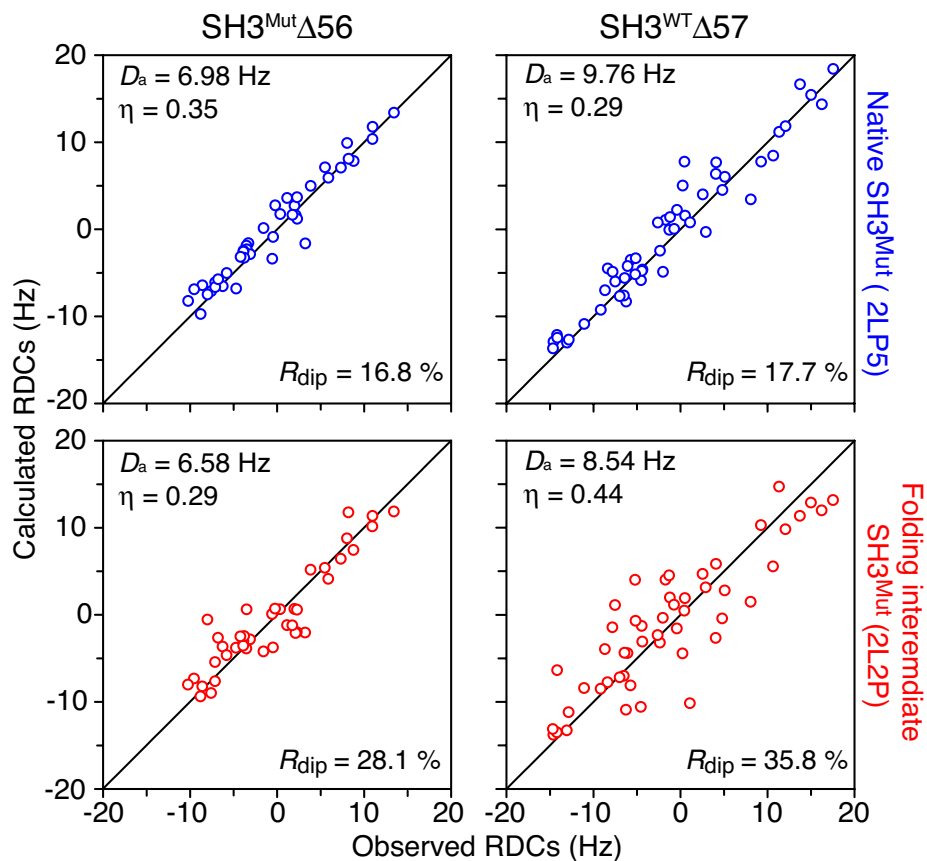
Figure S2. Backbone RDC analysis of SH3^{Mut}Δ56 and SH3^{WT}Δ57 constructs. Best-fit agreements obtained by singular value decomposition (SVD) between observed backbone amide RDCs measured in ~11mg/ml phage pf1 (10°C) and calculated RDCs derived from the coordinates of the folded (PDB 2LP5, top panels) and folding intermediate (PDB 2LP2, bottom panels) states of SH3^{Mut}.[1b] The RDC R factor (R_{dip}) is given by $\{5<(D_{obs}-D_{calc})^2>/[2D_a^2(4+3\eta^2)]\}^{1/2}$, where D_{obs} and D_{calc} are the measured and calculated RDC values respectively, and D_a and η are the magnitude and rhombicity of the alignment tensor obtained by SVD.[20]
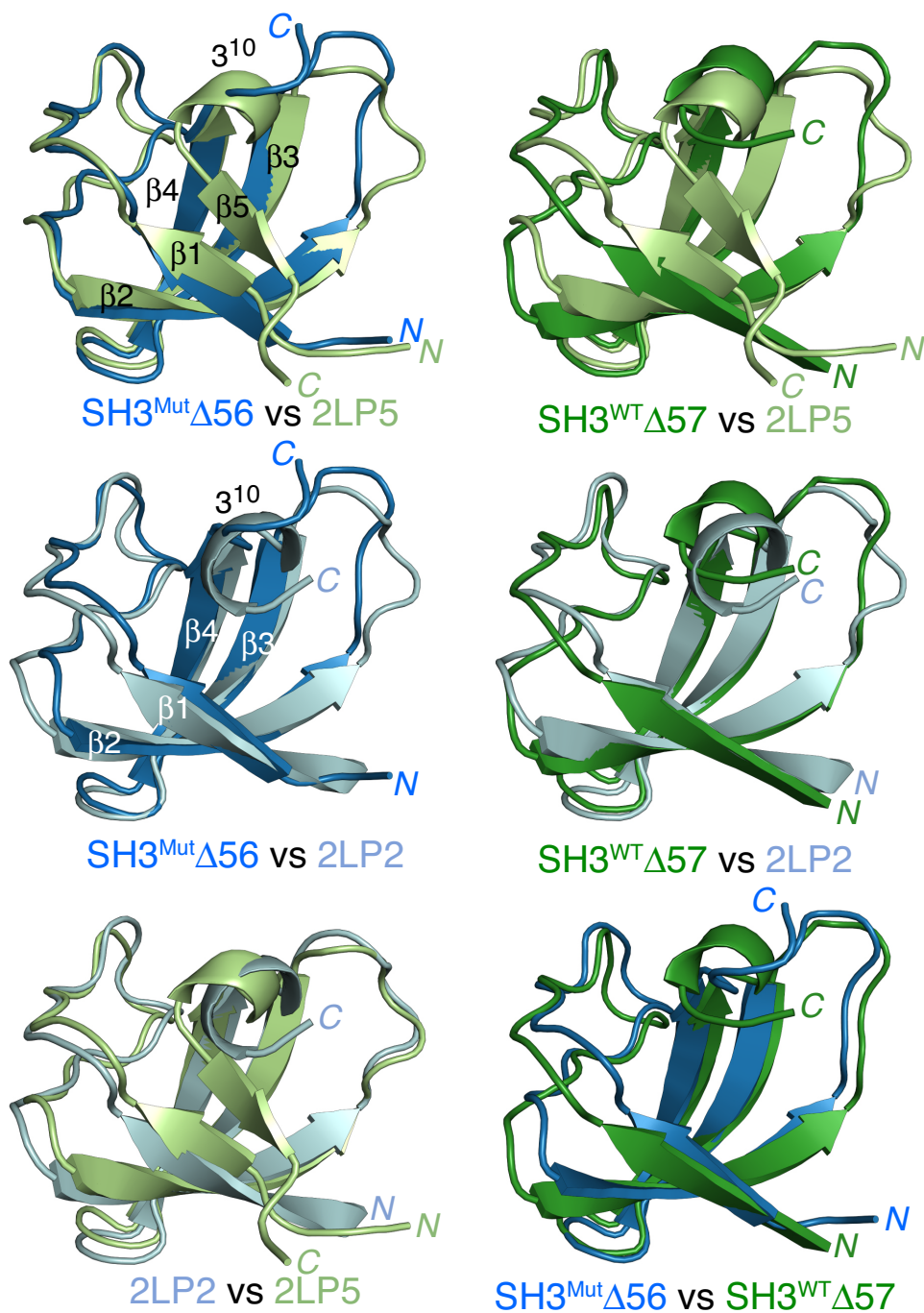
Figure S3. Comparison of SH3$^{\text{Mut}}\Delta$56 and SH3$^{\text{WT}}\Delta$57 structures calculated using CS-Rosetta on the basis of backbone chemical shifts (^{15}N, ^{13}Cα, ^{13}Cβ, ^{13}C', ^{1}H$_{N}$) and backbone amide residual dipolar couplings[9] with those of the native (PDB ID 2LP5) and folding intermediate (PDB 2LP2) states of SH3$^{\text{Mut}}$.[1b]
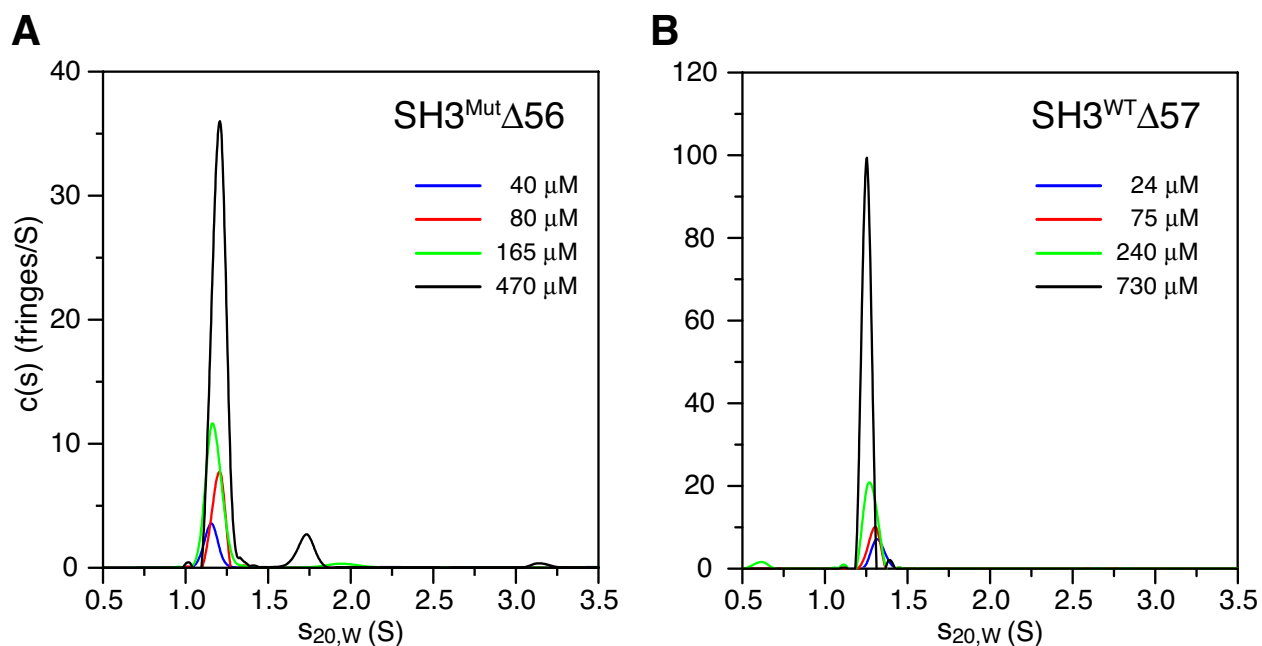
Figure S4. Sedimentation velocity measurements of SH3^{Mut}Δ56 and SH3^{WT}Δ57 at various loading concentrations. The data confirm the presence and absence of a populated folded dimeric state for SH3^{Mut}Δ56 and SH3^{WT}Δ57, respectively. (A) Interference c(s) profiles for samples of ^{15}N SH3^{Mut}Δ56 at 40 (blue), 80 (red), 165 (green) and 470 (black) μM. Similar absorbance profiles are noted for samples at 40, 80 and 165 μM. Absorbance data were not collected at the highest loading concentration. Both the absorbance and interference c(s) profiles at 40 (blue) and 80 (red) μM show the presence of single species at 1.18 S with an estimated molar mass of 7.6±0.3 kDa indicative of a monomer. At these concentrations, there does not appear to be any evidence for faster sedimenting material. At 165 μM (green), sedimentation data indicate the presence of trace amounts (~5% of the total interference signal) of faster sedimenting material, whereas at 470 μM (black) another species is observed at 1.72 S with an estimated mass of 13.7 kDa indicative of a dimer. Trace amounts (~1%) of faster sedimenting material are also observed at the highest concentration. (B) Interference c(s) profiles for samples of ^{15}N SH3^{WT}Δ57, on the other hand, shows the presence of only a single species at 1.13S with an estimated molecular mass of ~7.2 kDa, indicative of a stable monomer, over a concentration range of 24 to 730 μM. All experiments were conducted at 10°C.
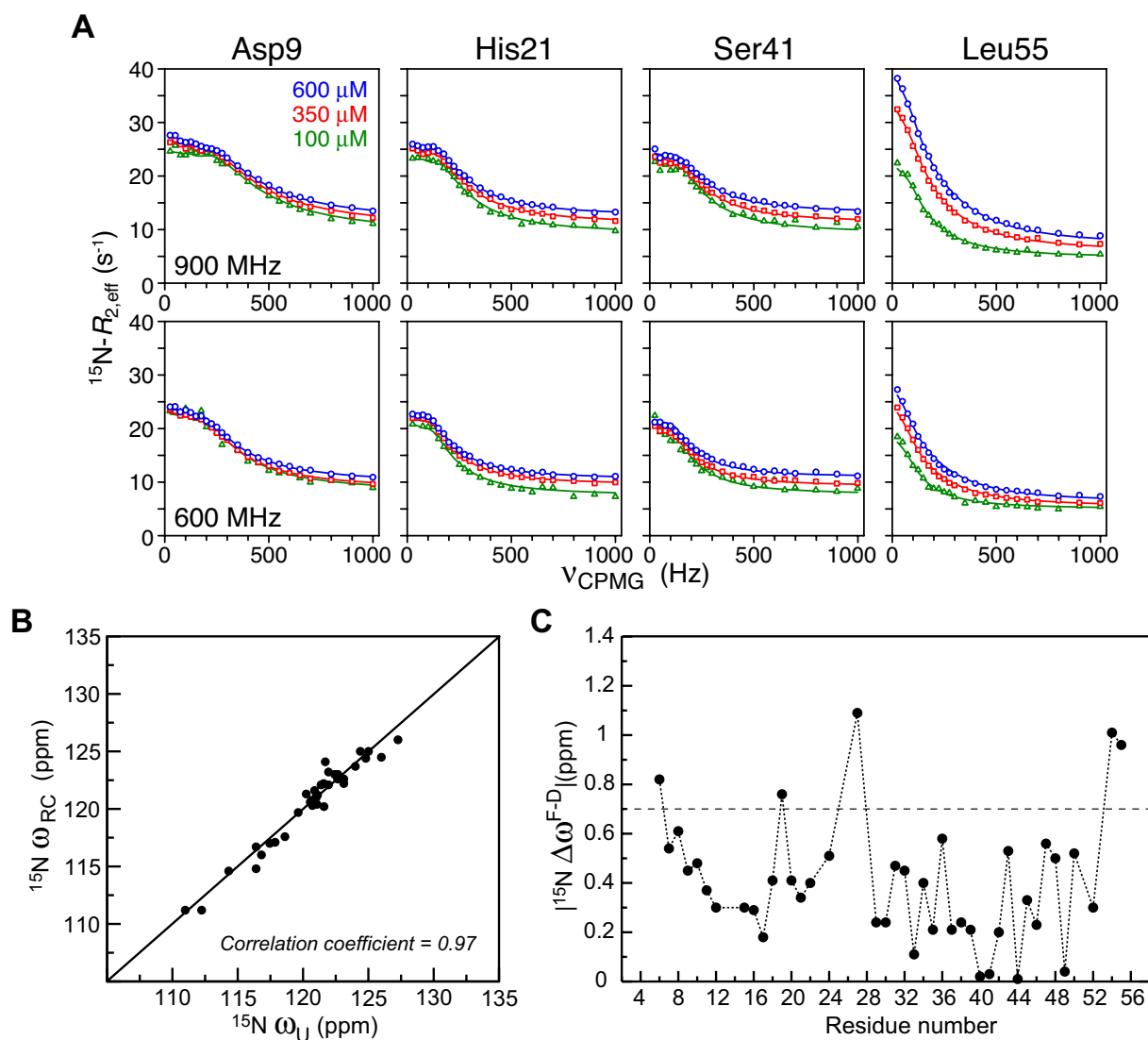
Figure S5. ^{15}N CPMG relaxation dispersion for free SH3^{Mut}Δ56. (A) Examples of ^{15}N in-phase CPMG relaxation dispersion profiles measured for free SH3^{Mut}Δ56 at three concentrations (blue, 0.6 mM; red, 0.35 mM; and green, 0.1 mM). The best-fit curves to a three state model comprising folded dimer, folded monomer and unfolded monomer are shown as solid lines. The CPMG curves for 42 residues were fit simultaneously. Excluded residues comprised those with either no significant dispersion, poorly resolved cross-peaks or poor signal-to-noise. Experiments were conducted at 10°C. The values of the rate constants (cf. scheme in Fig. 3 and Scheme S1) are as follows: k_{DF} ~1400 s^{-1}; k_{FU} ~15 s^{-1}; k_{UF} ~ 370 s^{-1}; k_{FD} ~ 280, 150 and 55 s^{-1} at SH3^{Mut}Δ56 concentrations of 600, 350 and 100 μM, respectively. No dispersions of any significance ($R_{ex} > 1$ s^{-1}) are observed for SH3^{WT}Δ57 indicative of a stable monomer and of the absence of any detectable dimer or unfolded species. (B) Correlation between the calculated values of the ^{15}N chemical shifts for the unfolded state **U** versus the corresponding predicted random coil (RC) shifts. The former are directly obtained from the optimized absolute values of $|\Delta\varpi^{F-U}|$ with the signs taken from those reported previously for the chemical shift differences between folded and unfolded states of the full length SH3^{Mut}.[1a] The predicted random coil shifts are taken from ref. 1a based on the work of Wishart et al.[21] (C) Optimized values of the absolute chemical shift differences, $|\Delta\varpi^{F-D}|$, between folded monomeric (F) and dimeric (F$_D$) states as a function of sequence. Five residues (A6, S19, Q27, Y54 and Y55) have optimized values of $|\Delta\varpi^{F-D}| > 0.7$ ppm.
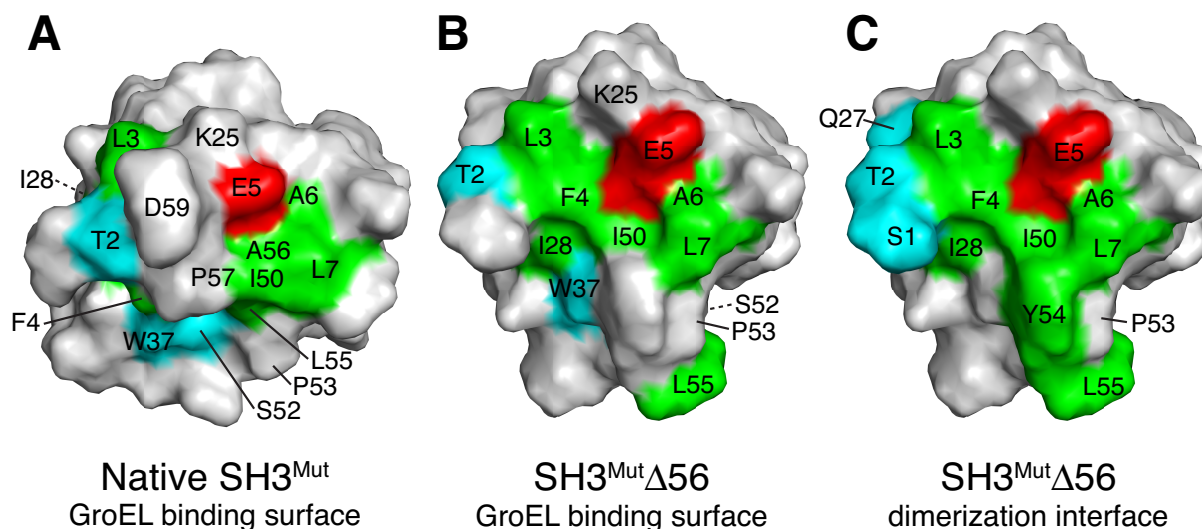
Figure S6. Molecular surface representations showing a comparison of the GroEL-binding surface and dimerization interface for SH3^{Mut}Δ56. The GroEL-binding interface for the folding intermediate of SH3^{Mut}, determined previously from combined analysis of relaxation dispersion, ΔR_2 and DEST data on the SH3^{Mut}/GroEL system,[3] mapped onto the molecular surface of the structures of (A) the folded (native) state of SH3^{Mut} and (B) SH3^{Mut}Δ56 (a molecular mimetic of the folding intermediate of SH3^{Mut}). (C) The dimerization interface for SH3^{Mut}Δ56 determined from relaxation dispersion (see Fig. S5C, residues A6, S19, Q27, Y54 and L55) and the observation of severe cross-peak line broadening in the ^{1}H-^{15}N HSQC spectrum of a concentrated (0.6 mM) sample of ^{15}N-labeled SH3^{Mut}Δ56 (residues S1, T2, L3, F4, E5, A6, L7, I28; note that apart from A6 and L7, the cross-peaks for the other residues were too broad in the concentrated sample to permit analysis of their relaxation dispersion profiles owing to low signal-to-noise). The color-coding used for the residues of the GroEL-binding surface and the dimerization interface is as follows: green, hydrophobic; cyan, polar; and red, negatively charged. Note that residues V54 and L55 are not surface exposed in SH3^{WT}Δ57 providing a rationalization as to why no dimeric form is observed for SH3^{WT}Δ57; even at high concentration (0.6 mM) no broadening of ^{1}H/^{15}N cross-peaks is detected.
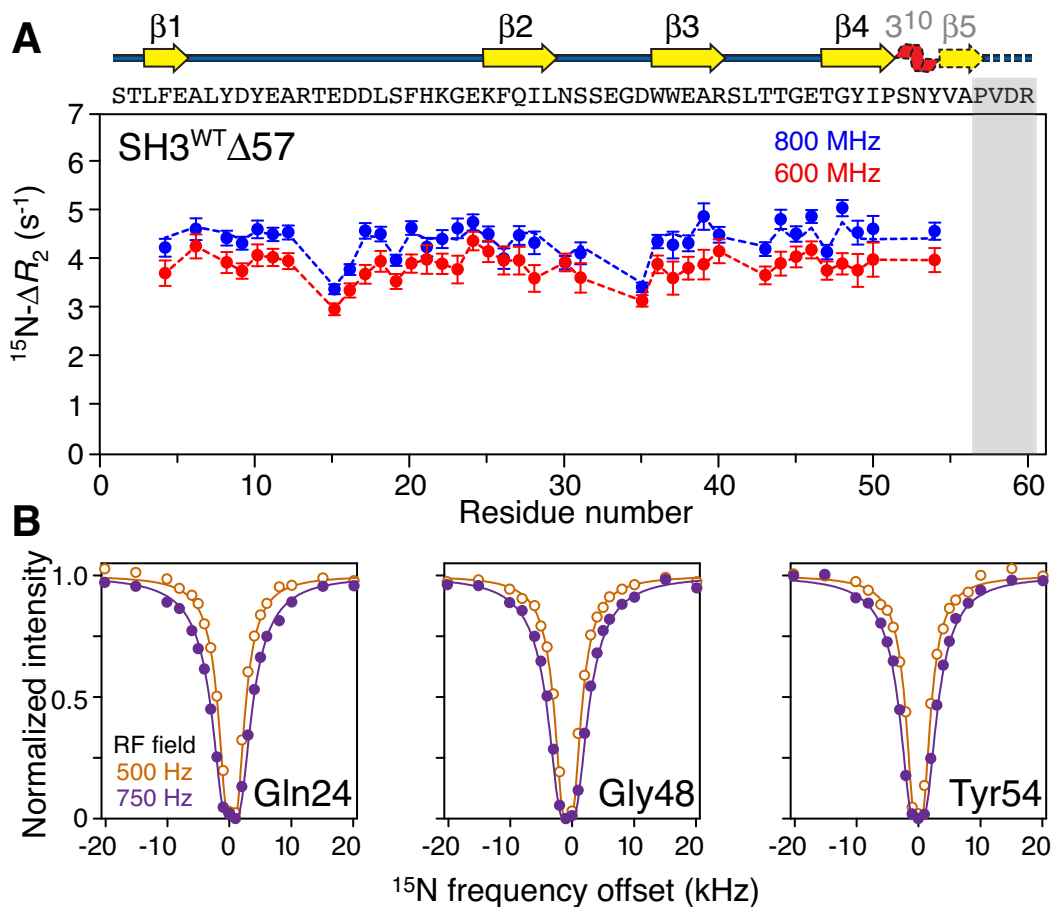
Figure S7. (A) ^{15}N-ΔR_2 lifetime line broadening and (B) ^{15}N-DEST profiles measured for SH3^{WT}Δ57 (100 μM) in the presence of GroEL (120 μM in subunits). The dashed and solid lines in panels A and B, respectively, are the best-fits obtained using a two-state model involving exchange between NMR-visible SH3^{WT}Δ57 in bulk solution and NMR-invisible ("dark") SH3^{WT}Δ57 bound to GroEL. The grey bar in panel A indicates the residues of the full length SH3 domain that are deleted in the SH3^{WT}Δ57 construct. Experiments were conducted at 10°C. The ^{15}N-ΔR_2 and ^{15}N-DEST data for 48 residues were fit simultaneously (excluding residues with either cross-peak overlap or very poor signal-to-noise). The amino acid sequence and secondary structure are displayed above panel A. Error bars = 1 S.D. The data are described by two-site exchange between F and F$_B$ states with k_{FB} and k_{BF} values of ~7 and 500 s^{-1} (and $< R_{2,F_B}^{800} > \sim 870$ s^{-1}). The values of the rate constants are slightly larger than the corresponding values of $k_{on}^{overall}$ (4-6 s^{-1}) and $k_{off}^{overall}$ (100-160 s^{-1}) for SH3^{Mut}Δ56 (see Fig. 3 main text), accounting for the larger magnetic field dependence of ΔR_2 observed for SH3^{WT}Δ57 (panel A) compared to SH3^{Mut}Δ56 (Fig. 1B, main text).
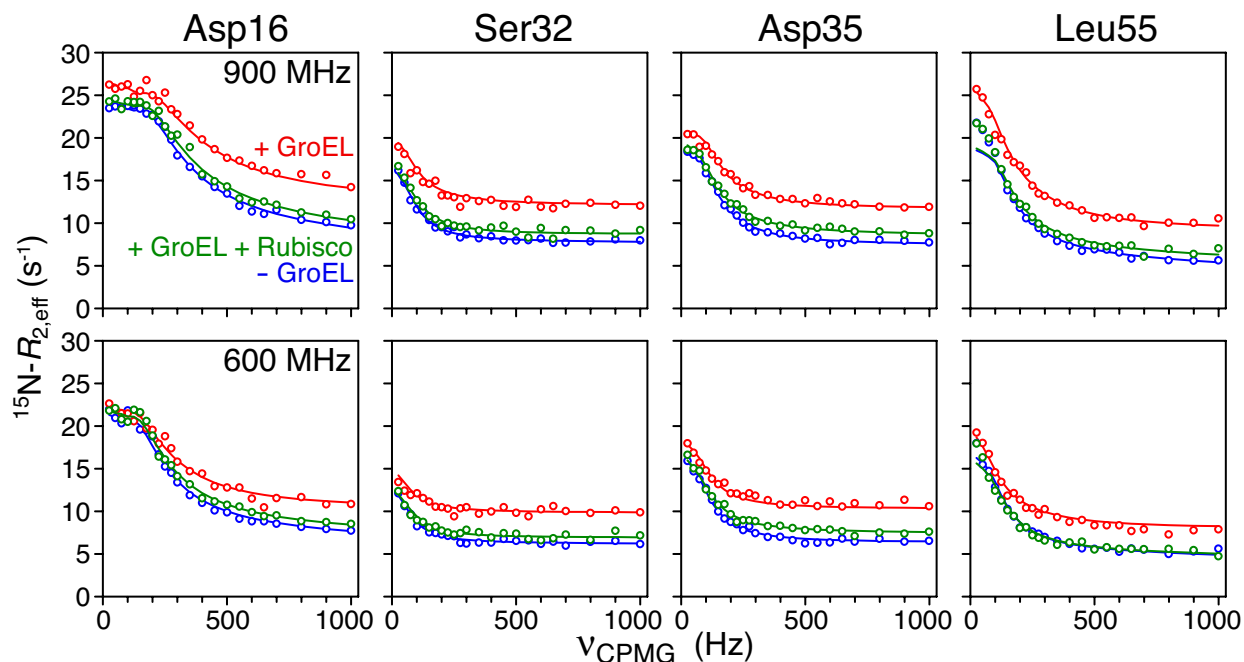
Figure S8. Examples of ^{15}N relaxation dispersion profiles for $SH3^{Mut}\Delta56$ free (blue), in the presence of acid denatured Rubisco-blocked GroEL (green), and in the presence of GroEL (red). The best-fits to five-state ($SH3^{Mut}\Delta56$ + GroEL, red) and three-state kinetic models (free $SH3^{Mut}\Delta56$, and $SH3^{Mut}\Delta56$ + GroEL + acid denatured-Rubisco) are shown as solid lines. The kinetic models with rate constants and populations are shown in Scheme S1 as well as in Fig. 3 of the main text. Experiments were conducted at 10°C. The concentrations are as follows: 100 µM $SH3^{Mut}\Delta56$ free and in the presence of either 120 µM (in subunits) GroEL or 120 µM (in subunits) GroEL + 19 µM acid-denatured Rubisco. There are no significant differences between the rate constants obtained for free $SH3^{Mut}\Delta56$ versus those in the presence of GroEL blocked by acid-denatured Rubisco.
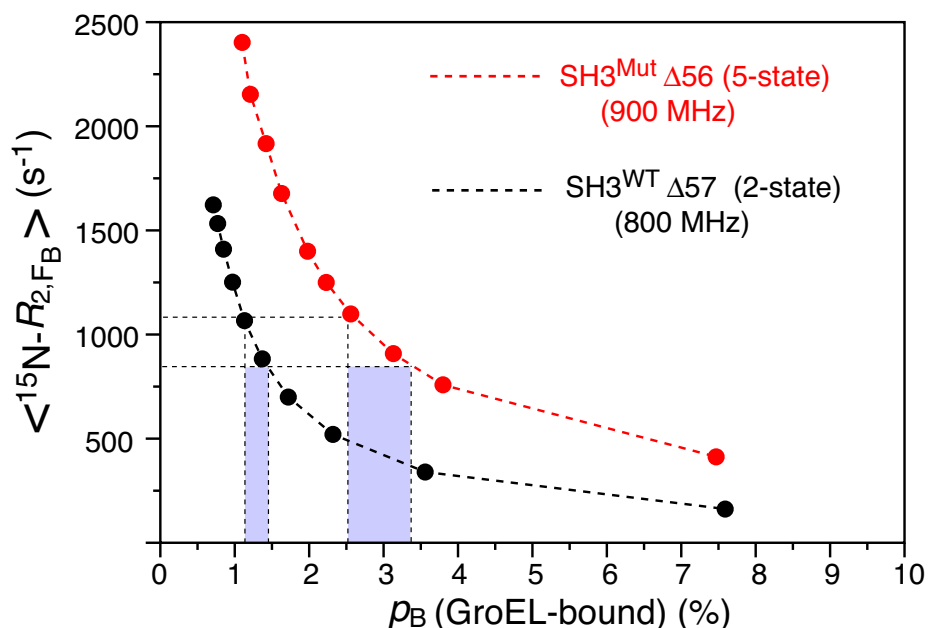
Figure S9. Dependence of the average $<^{15}N\text{-}R_{2,F_B}>$ values for SH3$^{Mut}\Delta56$ (red) and SH3$^{WT}\Delta57$ (black) bound to GroEL as a function of the population p_B of the GroEL-bound state, **F$_B$**. The $<^{15}N\text{-}R_{2,F_B}>$ values for the **F$_B$** state were determined from minimization of the target functions for either SH3$^{Mut}\Delta56$ (red, 900 MHz) or SH3$^{WT}\Delta57$ (black, 800 MHz) at fixed values of the population of SH3 bound to GroEL (p_B). For simplicity, the minimizations were performed with either $\{k_{FB}; k_{BF}\}$ or $\{k_{EB}; k_{BE}\} = 0$, effectively eliminating either the direct or indirect pathways for the formation of state F$_B$, respectively, leading to virtually identical results. Based on molecular weight of GroEL (~800 kDa), $<^{15}N\text{-}R_{2,F_B}>$ is calculated to be ~950 s^{-1} at 900 MHz and 10°C. The regions highlighted in grey denote the ranges of p_B that correspond to realistic $<^{15}N\text{-}R_{2,F_B}>$ values for the SH3-GroEL complexes in the F$_B$ state at 10°C.

Supplementary references

(1)   (a) Neudecker, P.; Zarrine-Afsar, A.; Choy, W. Y.; Muhandiram, D. R.; Davidson, A. R.; Kay, L. E. *J. Mol. Biol.* **2006**, *363*, 958-976. (b) Neudecker, P.; Robustelli, P.; Cavalli, A.; Walsh, P.; Lundstrom, P.; Zarrine-Afsar, A.; Sharpe, S.; Vendruscolo, M.; Kay, L. E. *Science* **2012**, *336*, 362-366.

(2)   Kapust, R. B.; Tozser, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S. *Prot. Eng.* **2001**, *14*, 993-1000.

(3)   Libich, D. S.; Tugarinov, V.; Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A* **2015**, *112*, 8817-8823.

(4)   (a) Van der Vies, S. M.; Viitanen, P. V.; Gatenby, A. A.; Lorimer, G. H.; Jaenicke, R. *Biochemistry* **1992**, *31*, 3635-3644. (b) Grason, J. P.; Gresham, J. S.; Lorimer, G. H. *Proc. Nat. Acad. Sci. U. S. A.* **2008**, *105*, 17339-17344.

(5)   Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.

(6)   (a) Bax, A.; Grzesiek, S. *Acc. Chem. Res.* **1993**, *26*, 131-138. (b) Clore, G. M.; Gronenborn, A. M. *Trends Biotech.* **1998**, *16*, 22-34.

(7)   Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571-10572.

(8)   Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373-378.

(9)   Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685-4690.

(10)   Lakomek, N. A.; Ying, J.; Bax, A. *J. Biomol. NMR* **2012**, *53*, 209-221.

(11)   (a) Fawzi, N. L.; Ying, J.; Ghirlando, R.; Torchia, D. A.; Clore, G. M. *Nature* **2011**, *480*, 268-272. (b) Fawzi, N. L.; Ying, J.; Torchia, D. A.; Clore, G. M. *Nature Protocols* **2012**, *7*, 1523-1533.

(12)   Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Phys. Chem. B* **2008**, *112*, 5898-5904.

(13)   Zhao, H.; Brautigam, C. A.; Ghirlando, R.; Schuck, P. *Curr. Protoc. Protein Sci.* **2013**, *20.12*.

(14)   Zhao, H. Y.; Ghirlando, R.; Piszczek, G.; Curth, U.; Brautigam, C. A.; Schuck, P. *Anal. Biochem.* **2013**, *437*, 104-108.

(15)   Schuck, P. *Biophys. J.* **2000**, *78*, 1606-1619.

(16)   Cole, J. L.; Lary, J. W.; Moody, T. P.; Laue, T. M. *Methods Cell. Biol.* **2008**, *84*, 143-179.

(17)   Weissman, J. S.; Rye, H. S.; Fenton, W. A.; Beechem, J. M.; Horwich, A. L. *Cell* **1996**, *84*, 481-490.

(18)   Helgstrand, M.; Härd, T.; Allard, P. *J. Biomol. NMR* **2000**, *18*, 49-63.

(19)   Zhou, H. X.; Dill, K. A. *Biochemistry* **2001**, *40*, 11289-11293.

(20)   Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 9008-9012.

(21)   Wishart, D. S.; Bigam, C. G., Holm, A., Hodges, R. S., Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67-81.