The role of domain interactions in the aggregation of full-length immunoglobulin light chains

Enrico Rennella[1,*], Gareth J Morgan[2,*], Jeffery W Kelly[2,*], and Lewis E Kay[1,3,*]

Supporting Information

Production of variable and constant domains. BL21(DE3) *E. Coli* cells were transformed with a pET-22 plasmid containing the DNA sequence for the variable or constant domain with a *pelB* leader sequence for periplasmic expression. Cells were grown at 37 °C (in all cases expect WIL-V_L) or 33 °C (WIL-V_L) in M9 medium containing $^{15}NH_4Cl$ (1 g/L) and ^{12}C-glucose (4 g/L) as the sole nitrogen and carbon sources until OD ~ 0.6, then induced overnight with 0.2 mM IPTG at 20 °C. Cells were harvested, resuspended in 50 mL of 50 mM TrisHCl pH 8, 0.5 M sucrose, 5 mM EDTA in the presence of SigmaFAST protease inhibitors and incubated for 30 minutes at 4 °C to release the periplasmic content by osmotic shock. 100 mL of Milli-Q water was added followed by a further incubation period of 30 minutes. The soluble material containing the protein was separated from insoluble cellular debris by centrifugation at 10,000 g for 20 minutes. Subsequently, a 2-step ammonium sulfate precipitation procedure was performed at 4 °C. The first step was performed under 25% saturation conditions, with the light-chain domain residing in the soluble fraction, which was then isolated by centrifugation at 10,000 g for 20 minutes. The second step involved 80% saturation with the light-chain domain residing in the insoluble fraction that was obtained by centrifugation (10,000 g) for 60 minutes. This fraction was resuspended in 25 mM

TrisHCl pH 8 and dialyzed overnight against the same buffer. The final purification procedure involved anion exchange and size-exclusion chromatographic steps. Fractions containing the protein were identified via SDS-PAGE and the identity of each protein confirmed by mass spectrometry.

Production of full length light chains. BL21(DE3) *E. Coli* cells were transformed with a modified pDEST-14 plasmid containing the DNA sequence for the full-length protein. Cells were grown at 37 °C in M9 medium containing $^{15}NH_4Cl$ (1 g/L) and ^{12}C-glucose (4 g/L) as the sole nitrogen and carbon sources until OD ~ 0.6, followed by overnight induction with 0.5 mM IPTG at 20-25 °C. Cells were harvested, resuspended in 30 mL PBS (phosphate buffered saline, pH 7.4) in the presence of SigmaFAST protease inhibitors and trace DNase I and sonicated at 4 °C. The insoluble material containing the protein was collected by centrifugation at 35,000 g for 20 minutes. The pellet was washed with PBS containing 1% v/v Triton X-100 to solubilize and remove inner membrane proteins, followed by a wash with PBS without the detergent. The pellet was then resuspended in 4 M GdnHCl, 25 mM TrisHCl pH 8, 5 mM DTT for several hours. Refolding was achieved by a slow 20X dropwise dilution into 25 mM TrisHCl pH 8, 5 mM reduced glutathione and 0.5 mM oxidized glutathione at 4 °C. The final purification steps, involving ammonium sulfate precipitation and chromatography are the same as for the individual domains, discussed above.

Sample conditions for NMR. In all NMR experiments, except hydrogen/deuterium exchange, the protein was dissolved in a 50 mM Bis-Tris buffer, pH 6.4, 1 mM EDTA, 10% D_2O. The buffer for hydrogen/deuterium exchange experiments was 20 mM sodium citrate, pD 5, in D_2O. Protein concentrations varied between 25 μM and 2 mM,

depending on the application. To avoid precipitation during sample preparation, FL light-chain samples were dialyzed against Milli-Q water, lyophilized and dissolved into the NMR buffer. NMR samples were stable over the time-course required to record datasets, typically at least for 24 hours.

CPMG experiments. ^{15}N CPMG experiments were acquired on a 14.1 T Bruker AVANCE III HD spectrometer equipped with a cryogenically cooled x,y,z gradient probe. Dispersion profiles for V_L and C_L domains were recorded using a pulse scheme that measures the decay of transverse in-phase ^{15}N magnetization as a function of the number of refocusing pulses (1). A constant-time relaxation interval (2), T = 35 ms, was employed during which ^{1}H continuous-wave decoupling was applied with a 17 kHz field (3). For the analysis of FL light chains ^{15}N TROSY CPMG experiments (4) were acquired with a constant-time relaxation interval of 25 ms. Average ^{15}N R_2 values (Fig. 5e) were also obtained from these experiments, considering only residues in β-strand regions. Intensities of cross-peaks in dispersion data sets were converted into $R_{2,eff}$ rates according to the relation $R_{2,eff} = \frac{-1}{T}\ln(\frac{I(\upsilon_{CPMG})}{I_o})$ where $I(\nu_{CPMG})$ and I_o are peak intensities from spectra recorded in the presence and absence of the T delay, respectively, and ν_{CPMG} = 1/(2δ), where δ is the delay between successive refocusing pulses (2, 5). Approximately 20 ν_{CPMG} values were recorded, between 0 and 1000 Hz, with a pair of repeats for error analysis (6).

Global fits of CPMG dispersion profiles and chemical shifts for JTO-V_L. Concentration dependent ^{15}N CPMG relaxation dispersion data and ^{15}N chemical shift changes were globally fit to a two-site monomer-dimer model of chemical exchange using the *chemex* software package (http://www.github.com/gbouvignies/chemex (7))

that numerically integrates the Bloch-McConnell equations (8), as described in detail previously (6). Only nuclei with dispersion profile sizes, $\Delta R_{2,eff} = R_{2,eff}(33\ \text{Hz}) - R_{2,eff}(1000\ \text{Hz}) > 10\ \text{s}^{-1}$, were included in the analysis, corresponding to 12 residues. For these residues concentration dependent chemical shifts were also included as restraints. These were simulated by first calculating free induction decays (FIDs) using simplified Bloch-McConnell equations that consider only transverse x and y components of in-phase magnetization:

$$\frac{d}{dt}\begin{bmatrix} M_x \\ M_y \\ D_x \\ D_y \end{bmatrix} = \begin{bmatrix} -k_1' & -\varpi_M & k_{-1} & 0 \\ \varpi_M & -k_1' & 0 & k_{-1} \\ k_1' & 0 & -k_{-1} & -\varpi_D \\ 0 & k_1' & \varpi_D & -k_{-1} \end{bmatrix} \cdot \begin{bmatrix} M_x \\ M_y \\ D_x \\ D_y \end{bmatrix}, \qquad \text{[S1]}$$

In Eq. [S1] M_j and D_j are the j-magnetization components for monomer and dimer (j={x,y}), respectively, $k_1' = 2k_1[M]$ is the pseudo-first order rate constant where [M] is the concentration of free monomer (9), k_{-1} is the dissociation rate constant (s^{-1}), and $\varpi_{M,D}$ is the chemical shift (rad/s) of the exchanging spin in the monomer (ϖ_M) or dimer (ϖ_D) state. Initial values of M_x and D_x were set to the equilibrium fractional populations of monomer (p_M) and dimer (p_D), respectively, that were calculated during the fitting process from the total protein concentration and optimized dissociation constant, $K_D = k_{-1}/k_1$; p_M = [M]/([M]+2[D]), p_D=1-p_M. Simulated FIDs were obtained by integrating Eq. [S1] to t = 90 ms and were then multiplied by a Gaussian function prior to Fourier transformation and fitting of peak positions. Finally, a target function $\chi^2_{red} = \chi^2_{red,CPMG} + \chi^2_{red,chemical\_shift}$ was minimized to obtain exchange rates and chemical shift differences between nuclei in the exchanging states.

Correction of relaxation rates for concentration differences. In general, ^{15}N R_2 values (as shown in Fig. 5e) were measured on samples with protein concentrations of 0.2 mM,

except for WT JTO-FL (0.4 mM), F101D JTO-FL (1 mM), WT WIL-FL (0.4 mM), and P145A+C215S WIL-FL (0.4 mM). R_2 values were corrected for the concentration dependent viscosity effect according to the relation $R_2^{\mathrm{C}} = R_2^0 \cdot (1 + \varepsilon \cdot \mathrm{C})$, where C is the protein concentration, R_2^0 is the extrapolated R_2 rate to C=0, and ε is the fractional increase in the effective viscosity upon increase in C. The value of ε was calculated to be 900 M^{-1} from R_2 values of the monomeric F101D+C215S JTO-FL construct measured at two different protein concentrations. A similar value of ε (800 M^{-1}) was obtained using R_2 values of dimeric C215S JTO-FL measured at two different protein concentrations.

Global fits of chemical shifts for WIL-V_L. Concentration dependent chemical shift changes were fit to a monomer-dimer model of chemical exchange, assuming fast exchange on the NMR chemical shift timescale. Only the 13 nuclei with well-resolved chemical shifts at the lowest total protein concentration used (0.025mM) were included in the fits. It is worth noting that accurate K_D values can be obtained most simply from the position of peaks (δ) relative to those for the fully monomer (δ_M) or dimeric (δ_D) species, via the relations (9):

$$\delta = p_M \cdot \delta_M + p_D \cdot \delta_D, \qquad \text{[S2]}$$

where $p_M = \frac{\sqrt{K_D^2 + 8 \cdot K_D \cdot P_T} - K_D}{4 \cdot P_T}$,

and P_T is the total protein concentration.

Reference values, δ_M and δ_D, were taken from spectra of 0.025 mM WIL-V_L (δ_M) and WT WIL-FL (δ_D).

Global fits of chemical shifts for V_L domains of JTO- and WIL-FL proteins. Chemical shifts were analyzed using a two-state model (Eq. [S2]) for exchange assuming fast interconversion between open and closed V_L-V_L interfaces (see Fig. 4). For JTO-FL the same 12 residues as for the isolated JTO-V_L domain were considered. Reference chemical shifts were set to the values measured for WT JTO-FL for the dimeric state, while monomer shifts were set either to the values extrapolated to zero protein concentration from the concentration dependent shifts for the isolated JTO-V_L domain or to the measured values for F101D JTO-FL. The two different choices for monomer shifts led to % closed values that were within error; the average value, rounded to the nearest 5% is reported in Figure 4f. For the WIL-FL protein residues were selected if the ^{15}N chemical shift differences between WT WIL-FL and 0.025 mM WIL-V_L were larger than 0.4 ppm. Reference chemical shifts for dimeric and monomeric states were set to measured shift values for WT WIL-FL and 0.025 mM WIL-V_L, respectively.

Global fits of chemical shifts for constant domains, either isolated C_L or within FL proteins. Only those nuclei were included in the analysis if their chemical shift differences were bigger than 0.2 ppm in the ^{15}N dimension or 0.03 ppm in the ^{1}H dimension in a comparison of spectra recorded on C215S C_L at concentrations of 1.76 and 0.025 mM. In this manner 15 ^{15}N nuclei and 11 ^{1}H nuclei were selected for global fits of all C_L and FL proteins, expect for P145A+C215S C_L and P145A+C215S FL proteins where only 4 ^{15}N spins and 4 ^{1}H spins were used. An additional constraint was also added for the P145A proteins that shift changes between P145A+C215S C_L and C215S C_L must be smaller than 5% of the difference between shifts of nuclei in C215S C_L and WT C_L constructs. This ensures that chemical shift changes reflect only the

monomer-dimer equilibrium and not the additional P145A mutation. Chemical shifts were analyzed using a two-state model (Eq. [S2]) for exchange assuming fast interconversion between open and closed C_L-C_L interfaces. Reference chemical shifts for the dimeric species were taken from a spectrum of WT JTO-FL in all analyses (C_L and FL proteins), while reference values for the monomeric species were from a spectrum of 0.025 mM C215S C_L when analyses did not involve P145A proteins and from a 0.025 mM data set of P145A+C215S C_L when the concentration dependencies of chemical shifts in P145A+C215S C_L and P145A+C215S FL proteins were quantified. Five concentrations were analyzed for C215S C_L (0.025 mM, 0.2 mM, 0.5 mM, 1.0 mM, 1.8 mM), 4 for P145A+C215S C_L (0.025 mM, 0.4 mM, 0.75 mM, 1.25 mM) and 1 (0.2 mM) for all FL proteins. In Figure 3f the values of the combined $\Delta\varpi_{HN}$ are calculated as:

$$\Delta\varpi_{HN} = \sqrt{\Delta\varpi_{1H}^2 + \left(\frac{\Delta\varpi_{15N}}{6}\right)^2} \quad [S3]$$

Hydrogen/deuterium exchange. Prior to performing experiments the protein was desalted by dialysis in Milli-Q water and lyophilized. Lock, shims and probe tuning were optimized on an NMR sample containing only the buffer. After dissolution of the protein, the hydrogen exchange reaction was followed by a series of 2D ^{1}H,^{15}N HSQC experiments, each lasting 5 minutes. Enhanced-sensitivity (10) and TROSY-based (11) HSQC experiments were recorded for variable domains and full-length proteins, respectively. Free energy values were obtained via (12):

$$\Delta G = -R \cdot T \cdot ln\frac{k_{hx}}{k_{rc}}, \quad [S4]$$

where R is the gas constant, T the temperature in Kelvin (310.15 K), k_{hx} is the decay rate of each amide, obtained from a mono-exponential fit of the exchange data, and k_{rc} is the

intrinsic rate of hydrogen exchange in D_2O, calculated using software available on the web, http://hx2.med.upenn.edu/download.html.

Aggregation assays. LCs (10 μM monomer equivalent) were incubated in black, clear-bottomed microwell plates (Corning #3631) in PBS, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, and 2.7 mM KCl, pH 7.4 containing 1 μM ThT. Plates were sealed with clear film and covered with a lid held in place by tape to minimize evaporation and shaken at 1000 rpm at 37 °C. ThT fluorescence was measured at intervals in a Molecular Devices Spectramax Gemini EM platereader, λ_{ex} = 440 nm, λ_{em} = 480 nm. Typically, 4 independent plates of 3 replicates each were used and in all cases, expect for data shown in Figure 2g, replicates derive from at least two plates. Only the replicates shown in Figure 2g are from a single plate. At the end of each time course, samples were taken for electron microscopy or filter trap assay. Insoluble material was quantified using a differential filtration ("filter trap") assay (13). LC solutions were drawn through a cellulose acetate membrane, which traps large aggregates, and a cellulose nitrate membrane, to which soluble protein is adsorbed, using a vacuum filter. After washing with PBS, the filters were stained in 0.1% (w/v) amido black in 10% acetic acid, then destained in 5% acetic acid. Membranes were scanned and protein quantified by densitometry using ImageJ (14). Aggregation results from the filter trap approach are in good agreement with ThT analyses.

Electron microscopy. Electron micrographs were recorded by the microscopy core facility at The Scripps Research Institute. Samples for electron microscopy were deposited onto glow-discharged copper grids (carbon-coated 400 mesh; Electron

Microscopy Sciences, Hatfield, PA) and stained with 2% phosphotungstic acid for 2 min. Grids were examined on a Philips CM100 electron microscope (FEI, Hillsbrough OR).

Analytical ultracentrifugation. 20 μM LCs in PBS were analyzed by sedimentation velocity analytical ultracentrifugation at 20 °C in a Beckmann XL-1 ultracentrifuge. PBS was used as a buffer blank. Samples were centrifuged at 50,000 rpm in an AN-60 Ti rotor and sedimentation was measured by absorbance at 280 nm. Data were fit to a c(S) model using Sedfit (15).

Measurement of free energy differences for the unfolding of V_L (ΔG_V) and C_L (ΔG_C) domains of FL proteins (Fig. 6). ΔG values were generally calculated from hydrogen exchange data (using Eq. [S4]) as the average over ΔGs from amides exchanging with the largest protection factors (Fig. S11). In the three cases, where domains are severely destabilized, extraction of ΔG from HX data was not possible because amides exchanged faster than the dead time of the NMR experiments (*i.e.*, time before acquisition of signal). These constructs include: (1) the V_L domain of L76G+C215S JTO-FL, where $\Delta G_V < -2$ kcal/mol, since native resonances are not visible in HSQC spectra acquired at equilibrium (Fig. S6); (2) the C_L domain of L121G+C215S JTO-FL, for which ΔG_C is close to zero since two peaks of approximately equal intensity are observed for S13 at the V_L-C_L interface, one derived from the folded portion of the C_L domain that engages in the V_L-C_L interface, the other from the unfolded fraction of the C_L domain (Fig. S9); (3) the C_L domain of P145A+C215S JTO-FL. An upper estimate for ΔG_C of +1.5 kcal/mol is calculated, since amides from the C_L domain exchanged faster than the dead time of the NMR experiments (300 s) and the intrinsic hydrogen exchange values are on the order of 0.1 s^{-1}.

Contribution of the monomer-dimer equilibrium to measured hydrogen exchange rates in the V_L domain. Consider first a hydrogen exchange reaction at a buried site in a protein (*N*), with exchange only occurring when the site becomes accessible to solvent (*U*). The reaction of interest is given by

$$N^H \underset{k_{UN}}{\overset{k_{NU}}{\rightleftarrows}} U^H \xrightarrow{k_{rc}} U^D \qquad \text{[S5]}$$

In the EX2 limit ($k_{UN} >> k_{rc}$) the exchange rate is given by (16)

$$k_{ex} = \frac{k_{NU}}{k_{NU}+k_{UN}} k_{rc} \ . \qquad \text{[S6]}$$

In the case of dimeric JTO-V_L we assume that the amide site in question is both protected by dimerization and by the native structure so that the dimer first dissociates to monomers that then exchange according to the scheme in Eq. [S5]. Assuming further that the monomer-dimer equilibrium is fast compared to either the subsequent folding/unfolding or hydrogen exchange steps, the exchange rate becomes

$$k_{ex}^D = \alpha \frac{k_{NU}}{k_{NU}+k_{UN}} k_{rc} \ ,$$

$$\text{where } \alpha = \frac{[N^H]}{[N^H]+2[N_2^{HH}]+[N_2^{HD}]+[N_2^{DH}]} \ . \qquad \text{[S7]}$$

In Eq. [S7] α is the fraction of protons from the exchanging site in question in the monomeric native state, and N_2^{HH}, N_2^{HD}, N_2^{DH} denote dimeric states where both monomers of the dimer are protonated (N_2^{HH}) or only one of the protomers in the dimer is protonated at the position of interest (N_2^{HD}, N_2^{DH}). The factor α can in turn be written as

$$\alpha = \frac{K_D}{K_D+2[N_T]} \qquad \text{[S8]}$$

where K_D is the dimer dissociation constant and N_T is the total concentration of monomeric protein, $[N_T]=[N^H]+[N^D]$. Note that for the JTO-V_L domain a dimer-monomer

exchange rate of 1200 s^{-1} is measured from CPMG relaxation dispersion experiments under our experimental conditions (total protein concentration of 0.5 mM) while folding rates are on the order of 3 s^{-1} that are an order of magnitude higher than k_{rc} values at pH 5, 37 ^{o}C. In a similar manner the monomer-dimer exchange rate for WIL-V_L is greater than 1000 s^{-1}, with a similar unfolding rate as for JTO-V_L (the folding rate could not be accurately measured by our chevron-based approach as the protein is not sufficiently stable upon addition of denaturants). Thus, the assumptions made in the derivation of the above equations (EX2 regime) are valid. Recasting k_{ex}^{D} in terms of a free energy change, ΔG_{HX},

$$\Delta G_{HX} = -RTln\left(\frac{k_{ex}^{D}}{k_{rc}}\right) = -RTln(\alpha) + \Delta G_{NU} \qquad [S9]$$

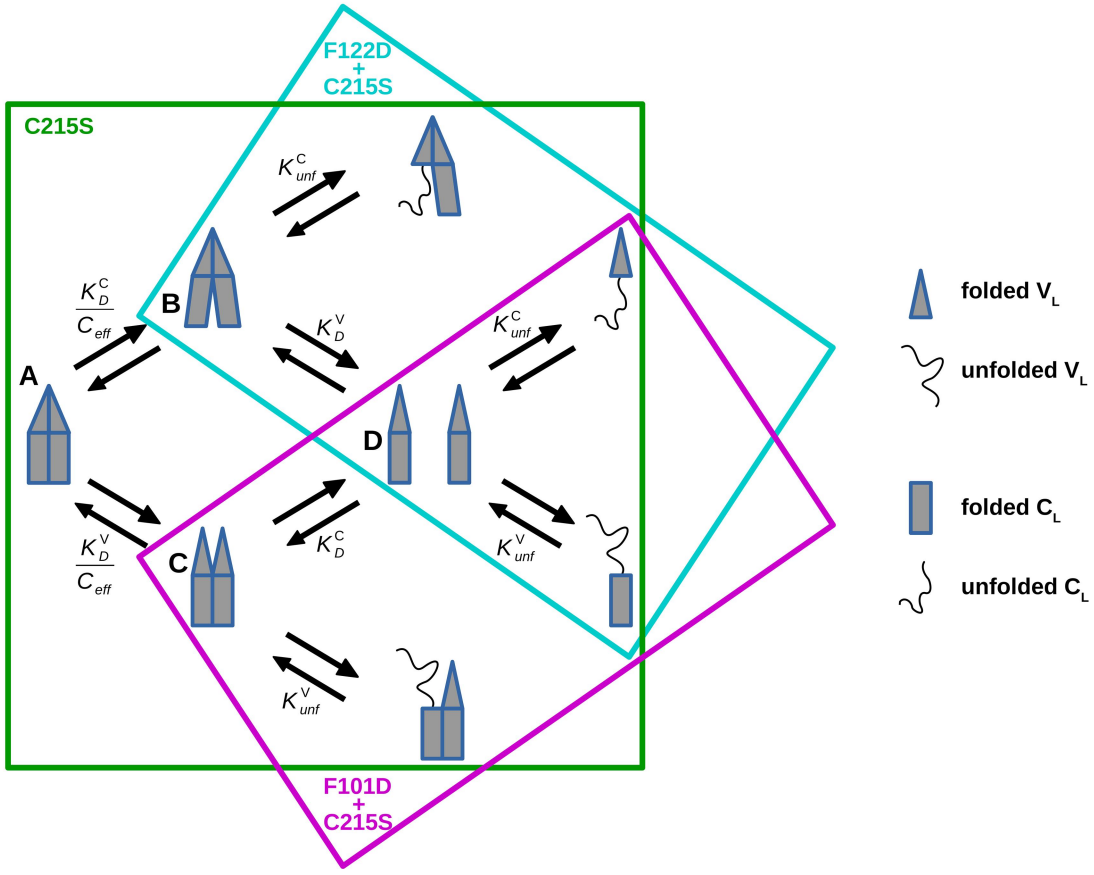
where $\Delta G_{NU} = -RTln(\frac{k_{NU}}{k_{UN}})$

and we have used the relation $k_{UN}>>k_{NU}$ that is appropriate for a stably folded domain. Thus, dimerization introduces an additional term ($-RTln(\alpha)$) to the free energy equation for hydrogen exchange, corresponding to 1.4 and 0.14 kcal/mol of stabilization for JTO-V_L and WIL-V_L, respectively, at protein concentrations of 0.5 and 0.8 mM. It is worth emphasizing that the $-RTln(\alpha)$ term only applies for amide protons where dissociation precedes exchange. For amides that are far outside of the dimer interface $\alpha = 1$.

The above discussion has focused on hydrogen exchange in V_L or C_L domains using a simple model that takes into account the monomer-dimer equilibrium. The situation can be more complex for FL light chains, described below, because monomerization of one interface can occur while the second interface remains intact.

A multi-step model for the unfolding of FL light chains with a C215S mutation. Schematic 1, below, shows an unfolding model that is applicable for C215S FL light chain mutants (*i.e.*, where the dimers are not covalently attached). In the scheme A, B, C and D denote molecules where both V_L and C_L domains are folded, as indicated, but where V_L-V_L or C_L-C_L interfaces are either formed or not present. Here we consider the following FL C215S variants: C215S JTO-FL (green), where the dominant conformation is one where all interfaces are present; F101D+C215S JTO-FL (dark magenta) where all species with a closed V_L-V_L interface are absent (Figure 4); F122D+C215S JTO-FL (cyan) where the closed C_L-C_L interface is effectively absent (Figure 5). NMR spectra for these variants establish that V_L and C_L domains are well folded (Fig. S3) so that the total protein concentration (monomer) in each case is well approximated by 2[A]+2[B]+2[C]+[D]. In Scheme 1 the bimolecular dissociation constants and unimolecular unfolding equilibrium constants are denoted by K_D^i and K_{unf}^i, respectively, where the superscript i distinguishes between the V_L and C_L domains ($i \in [V, C]$). Unimolecular dissociation constants, for example A to C, are related to the bimolecular constants, for example B to D, by an effective molarity, C_{eff} (17, 18). Here we have assumed that dissociation off-rates are independent of whether the process is uni- or bimolecular, but that the on-rates vary according to the relation $k^{bi} C_{eff} = k^{uni}$. Under these conditions it is straightforward to show that $K_D^{bi} / C_{eff} = K^{uni}$ (Scheme 1) and that the C_{eff} values as related to the A,C and A,B equilibria are identical. The latter results from the fact that ΔG=0 for the $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$ pathway. In what follows we will determine C_{eff} and subsequently show that the simple thermodynamic model presented

explains the measured hydrogen exchange data for these mutants, as a means of validating the proposed unfolding scheme.



Scheme 1. Schematic of the relevant unfolding equilibria for C215S JTO-FL (green), F122D+C215S JTO-FL (cyan) and F101D+C215S JTO-FL (magenta) LCs. Details are as described in the SI text.

As described in the main text, we have measured $K_D^V \sim 10$ μM from an analysis of CPMG relaxation dispersion data recorded on JTO-V_L (Fig. 2e) and a value of $K_D^C \sim 10$ mM for C215S C_L (Fig. 3e) from fits of concentration dependent chemical shift data. Using $K_D^C = 10$ mM it is straightforward to calculate the fraction of protein monomer in species C for F101D+C215S JTO-FL (corresponding to the C_L closed state) at a total

protein concentration, P_T, of 0.23 mM, as $f_C^{F101D+C215S}$= 2[C]/([D] + 2[C]) = 4% in good agreement with $f_C^{F101D+C215S}$ = 5% that is measured from the position of chemical shifts of amide probes that are sensitive to the C,D equilibrium (Fig. 5). In a similar manner, using K_D^V ~ 10 μM, a value of $f_B^{F122D+C215S}$ = 83% (corresponding to the fraction of protein monomer in state B for the F122D+C215S double mutant that is the closed V_L state) is calculated for P_T = 0.15 mM. This value is consistent with $f_B^{F122D+C215S}$ ~ 80% that is measured from relative peak positions in spectra recorded of F122D+C215S JTO-FL (see main text and Fig. 4).

Having established that the measured equilibrium constants for the isolated JTO-V_L and C215S JTO-C_L domains predict relative concentrations of species for both F101D+C215S and F122D+C215S JTO-FL LCs that are in agreement with experimental observations we now wish to calculate C_{eff}. This will then allow the evaluation of [A]-[D] for C215S JTO-FL and, as we show below, relative hydrogen exchange rates that can be compared with experiment. Our starting point is the observation that at a C215S JTO-FL concentration of 0.2 mM the C_L-C_L interface is 93% closed, f_{CC} = 0.93 (Fig. 5), where f_{CC} is the fraction of molecules with a closed C_L-C_L interface. Thus, with reference to Scheme 1 we can write

$$f_{CC} = \frac{2[\mathrm{A}] + 2[\mathrm{C}]}{P_T},\ P_T = 2[\mathrm{A}] + 2[\mathrm{B}] + 2[\mathrm{C}] + [\mathrm{D}]$$
$$\frac{f_{CC} P_T}{2} = [\mathrm{A}] + [\mathrm{C}] \qquad \text{[S10]}$$

where the factors of 2 are included because states A, B and C are dimeric. From the equilibria in Scheme 1,

$$\frac{K_D^V}{C_{eff}} = K_D'^V = \frac{[C]}{[A]},\ \frac{K_D^C}{C_{eff}} = K_D'^C = \frac{[B]}{[A]},\ K_D^C = \frac{[D]^2}{[C]},\ K_D^V = \frac{[D]^2}{[B]}, \qquad [S11]$$

from which

$$\frac{[B]}{[C]} = \frac{K_D^C}{K_D^V} = 1000\,. \qquad [S12]$$

From Eqs. [S10] and [S11] it follows that

$$[B] = \frac{4\beta + K_D^V - \sqrt{(4\beta + K_D^V)^2 - 16\beta^2}}{8}, \beta = P_T(1 - f_{CC}) \qquad [S13]$$

and [A], [C] and [D] are easily obtained from the value of [B] and Eqs. [S10] – [S12]. Finally, C_{eff} = 240 mM is calculated directly from either of the first two expressions in Eq. [S11] using P_T=0.2 mM, f_{CC} = 0.93 (see above). Thus, the equilibrium constants $\{K_D\}$ for the equilibria connecting states A, B, C and D are known.

It is possible to validate the model of Scheme 1 and the obtained $\{K_D\}$ values for the FL LC by comparing relative unfolding free energies for the F101D+C215S JTO-FL and F122D+C215S JTO-FL mutants, as measured from the slowest hydrogen exchange rates, with predicted values calculated from the K_Ds. The most slowly exchanging hydrogens report on unfolding and we assume in what follows that this class of hydrogen only exchanges with solvent once the dimeric interface connecting either C_L-C_L domains (for exchange involving protons from the C_L domain) or V_L-V_L domains (exchange from protons resident in V_L domains) is broken. The open interface must then unfold prior to the exchange event. Following from Eq. [S7] above (EX2 limit) we can write

$$k_{ex} = f_o^{i,j} \frac{K_{unf}^j}{K_{unf}^j + 1} k_{rc} \qquad [S14]$$

where $f_o^{i,j}$ is the fraction of molecules with open interface *j*, the superscript $i \in$ (C215S JTO-FL, F101D+C215S JTO-TL, F122D+C215S JTO-FL) denotes the mutant considered and $j \in \{V,C\}$ refers to exchange from either the V_L or C_L domain. Note that $f_o^{i,j}$ is analogous to α in Eq. [S7]. For the case of stably folded domains (*i.e.*, K_{unf}^{j}<<1) unfolding free energy differences can be expressed as

$$\Delta G_{HX}^{i,j} = \Delta G_{unf}^{i,j} - RT \ln f_o^{i,j} \qquad \text{[S15]}$$

where $\Delta G_{unf}^{i,j}$ is the unfolding free energy for domain *j* when the interface between *j* domains is absent (*i.e.*, unfolding from the monomer). The difference in $\Delta G_{HX}^{i,j}$ values calculated for different mutants *i* = *x* and *y* is thus given by

$$\Delta G_{HX}^{i=x,j} - \Delta G_{HX}^{i=y,j} = -RT \ln \frac{f_o^{i=x,j}}{f_o^{i=y,j}} \qquad \text{[S16]}$$

assuming that $\Delta G_{unf}^{i=x,j} = \Delta G_{unf}^{i=y,j}$. This latter assumption is reasonable in our analyses because we compare $\Delta G_{HX}^{i,j}$ values for domain *j* in pairs of constructs where substitutions differ only in domain *k*, where *j*≠*k* (for example, $\Delta G_{HX}^{F101D+C215S\ JTO-FL,C} - \Delta G_{HX}^{C215S\ JTO-FL,C}$, see below). Finally, values of $f_o^{i,j}$ can be readily calculated by recognizing that

$$\begin{aligned}
f_o^{C215S\ JTO-FL,C} &= \frac{2[\mathrm{B}]+[\mathrm{D}]}{P_T = (2[\mathrm{A}]+2[\mathrm{B}]+2[\mathrm{C}]+[\mathrm{D}])} \\
f_o^{C215S\ JTO-FL,V} &= \frac{2[\mathrm{C}]+[\mathrm{D}]}{P_T} \\
f_o^{F101D+C215S\ JTO-FL,C} &= \frac{[\mathrm{D}]}{2[\mathrm{C}]+[\mathrm{D}]} \\
f_o^{F122D+C215S\ JTO-FL,V} &= \frac{[\mathrm{D}]}{2[\mathrm{B}]+[\mathrm{D}]}
\end{aligned} \qquad \text{[S17]}$$

The concentrations of [A]-[D] for *i*=C215S JTO-FL are obtained from

$$[\mathrm{C}] = \frac{4\gamma P_T + K_D^C - \sqrt{(4\gamma P_T + K_D^C)^2 - 16\gamma^2 P_T^2}}{8\gamma^2}, \gamma = \frac{1}{K_D'^V} + \frac{K_D'^C}{K_D'^V} + 1. \quad [S18]$$

and relations in Eq. [S11], while concentrations for *i*=F101D+C215S FL-JTO and *i*=F122D+C215S FL-JTO are calculated from the total protein concentrations used and the known values of K_D^C and K_D^V, respectively. Below calculated and experimental $\Delta\Delta G_{HX}$ values are compared, focusing on the C_L domain using data from C215S JTO-FL and F101D+C215S JTO-FL (top, at a pair of concentrations) and the V_L domain based on measurements using C215S JTO-FL and F122D+C215S JTO-FL (bottom, at a pair of concentrations), with good agreement obtained in all cases.

Data for the C_L domain:

Protein (*i*)	$f_o^{i,C}$	$-RT\ln(f_o^{i,C})$	Calculated $\Delta\Delta G_{HX}^C$	Experimental $\Delta\Delta G_{HX}^C$
C215S JTO-FL, 0.67 mM	5.6%	1.7		
F101D+C215S JTO-FL, 0.23 mM	95.8%	0.0	-1.7	-1.8 ± 0.2

Protein (*i*)	$f_o^{i,C}$	$-RT\ln(f_o^{i,C})$	Calculated $\Delta\Delta G_{HX}^C$	Experimental $\Delta\Delta G_{HX}^C$
C215S JTO-FL, 0.17 mM	7.2%	1.6		
F101D+C215S JTO-FL, 0.23 mM	95.8%	0.0	-1.6	-1.4 ± 0.2

Data for the V_L domain:

Protein (*i*)	$f_o^{i,V}$	$-RT\ln(f_o^{i,V})$	Calculated $\Delta\Delta G_{HX}^V$	Experimental $\Delta\Delta G_{HX}^V$
C215S JTO-FL, 0.67 mM	1.7%	2.5		
F122D+C215S JTO-FL, 0.15 mM	16.7%	1.1	-1.4	-1.8 ± 0.2

Protein (*i*)	$f_o^{i,V}$	$-RT\ln(f_o^{i,V})$	Calculated $\Delta\Delta G_{HX}^V$	Experimental $\Delta\Delta G_{HX}^V$
C215S JTO-FL, 0.17 mM	3.4%	2.1		
F122D+C215S JTO-FL, 0.15 mM	16.7%	1.1	-1.0	-1.4 ± 0.2

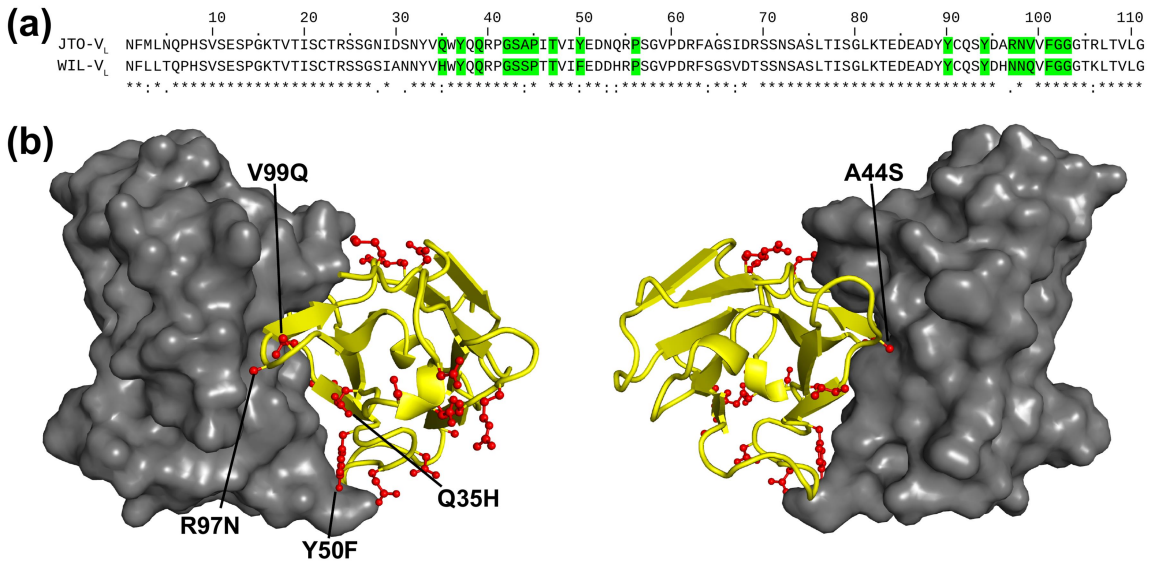


Figure S1. (a) Clustal Omega (19) sequence alignment of WIL-V_L and JTO-V_L; the two sequences are 84% identical and 95% similar. Residues at the V_L-V_L dimer interface are indicated in green. (b) Location of the mutated residues (red), with those in the V_L-V_L interface highlighted (for example V99Q), shown on the crystal structure of JTO-V_L (PDB 1CD0 (20)).

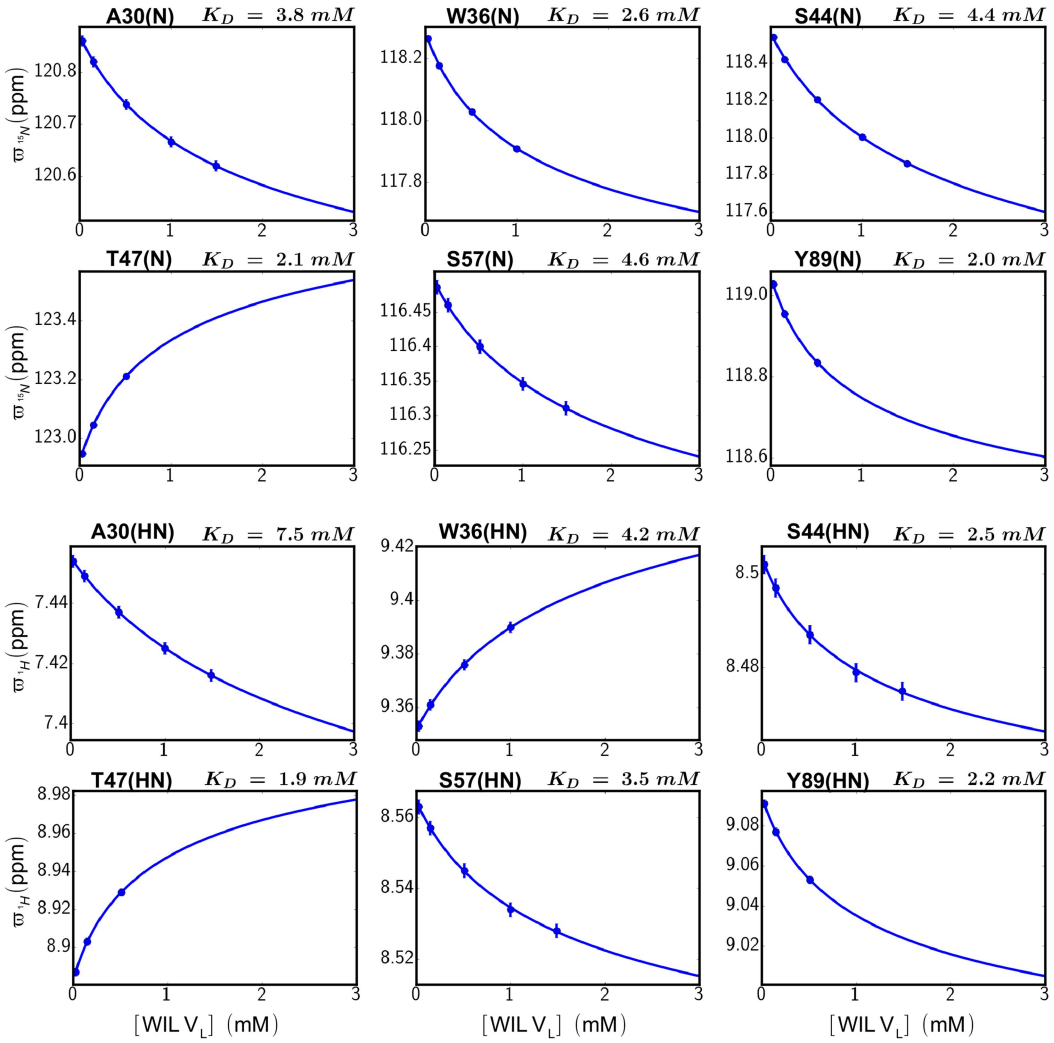


Figure S2. Fits of concentration dependent chemical shift changes for selected resonances of WIL-V_L. The fitted K_D value for each residue, along with the identity of each residue, is shown at the top of each plot. K_D values were estimated via global fits of chemical shifts at different concentrations (Eq. [S2]).

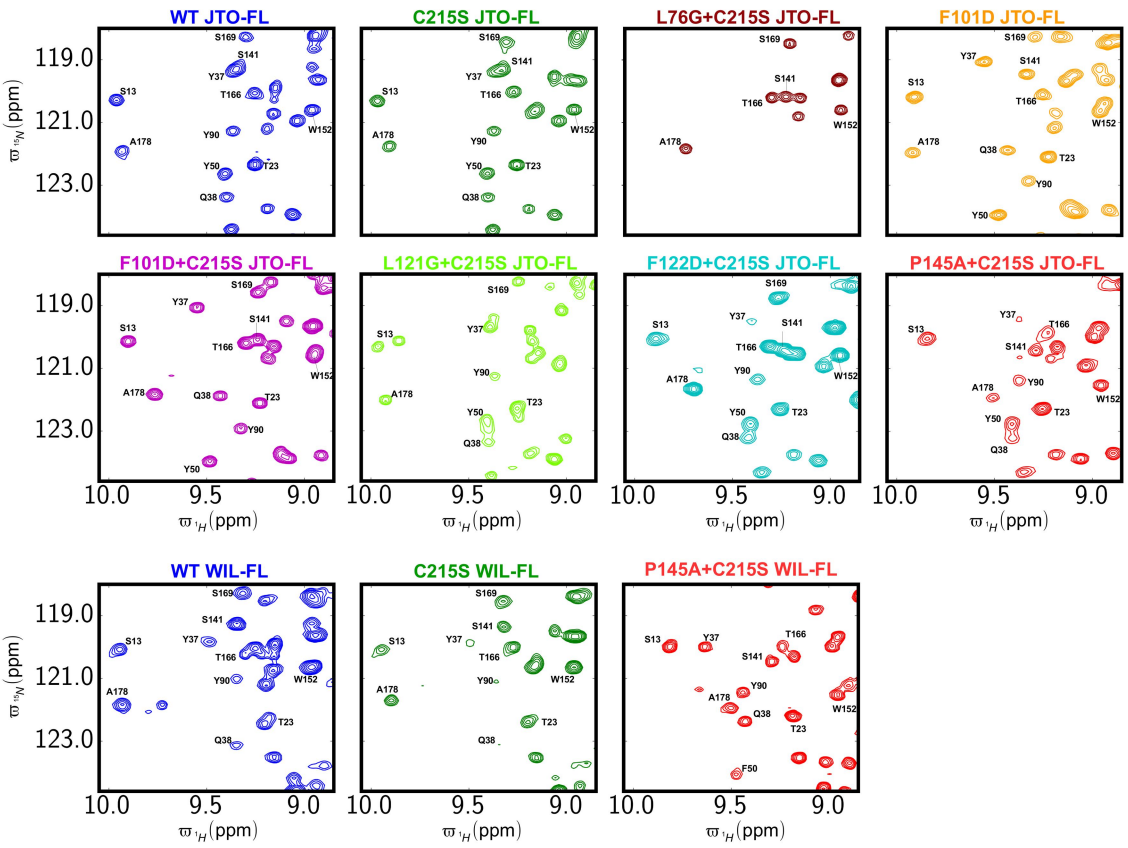


Figure S3. Selected regions from 2D $^{1}H,^{15}N$ correlation maps (37°C and 14.1 T) of JTO-FL and WIL-FL constructs used in the present work. The well-dispersed spectra indicate that the domains are folded.

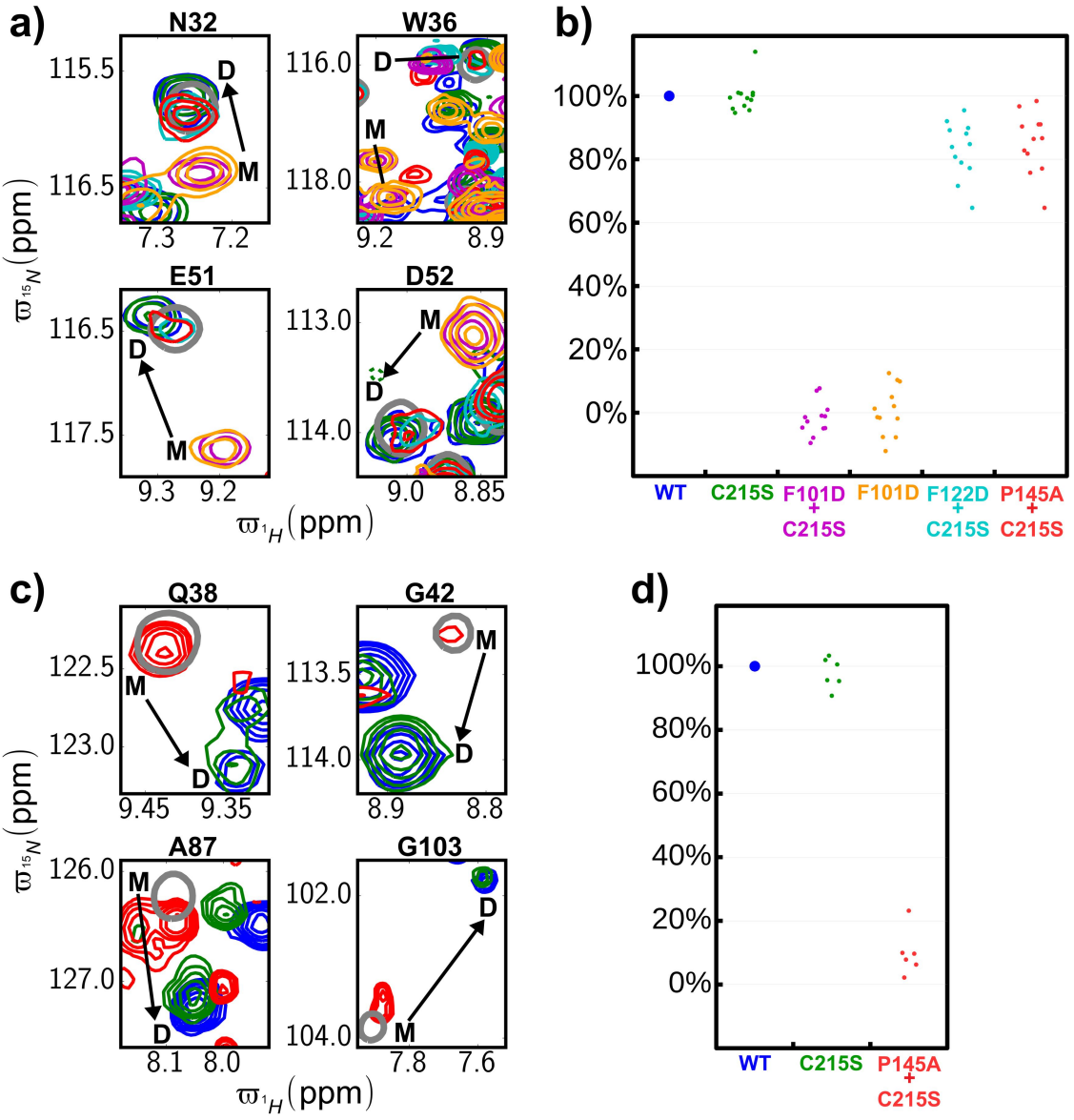


Figure S4. Selected regions of $^{1}H,^{15}N$ spectra of JTO-FL (**a**) and WIL-FL (**c**) constructs. Data were recorded at 37 ^{o}C and 14.1 T. Color-coding is as in Figure 4 and indicated in (**b**) and (**d**); grey contours derive from V_L domains. Percentage of closed (WT-like) V_L-V_L interfaces for variants of JTO-FL and WIL-FL are highlighted in panels (**b**) and (**d**), respectively. Each point corresponds to one residue from which the % closed value was estimated (see above). M and D denote the positions of cross-peaks without (M) and with (D) a WT V_L-V_L interface.

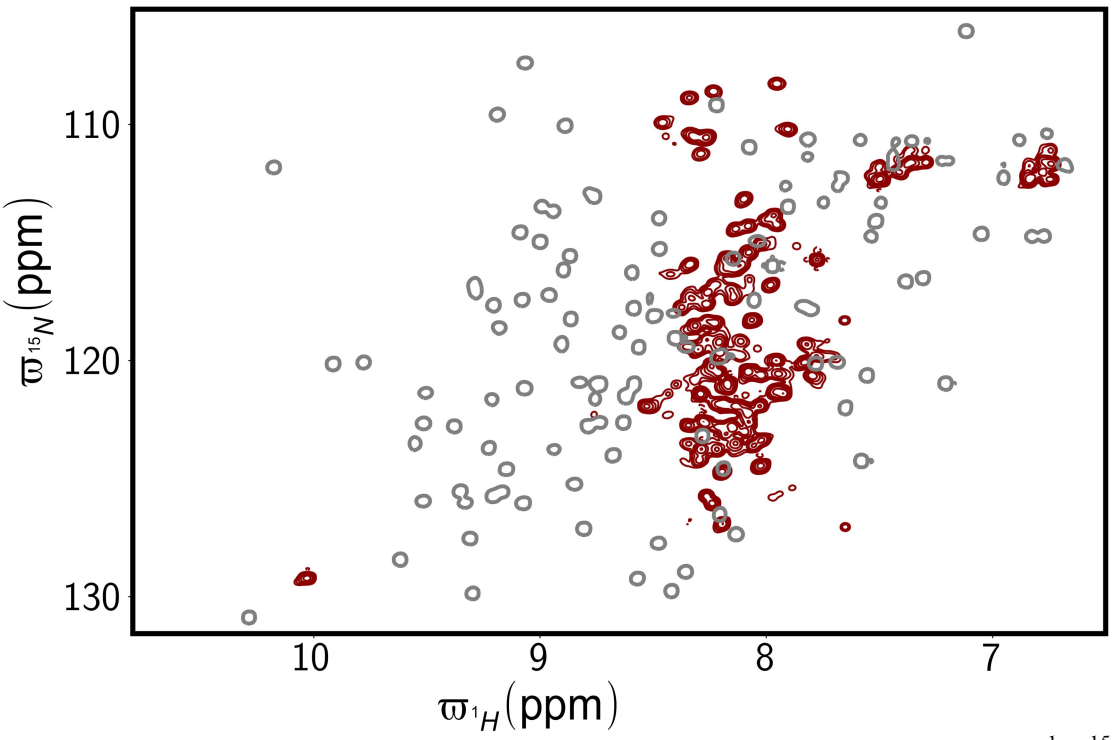


Figure S5. The L76G mutation severely destabilizes the V_L domain. (**a**) The $^{1}H,^{15}N$ HSQC spectrum of the L76G 6aJL2-V_L domain (dark red contours) at 37 °C is characteristic for an unfolded domain, with the cross-peaks for backbone amides in the 7.7-8.6 ppm range in the ^{1}H dimension. This is in contrast with the well-dispersed resonances for the WT 6aJL2-V_L domain (single grey contour). Both HSQC spectra were recorded at 0.1 mM protein concentration.

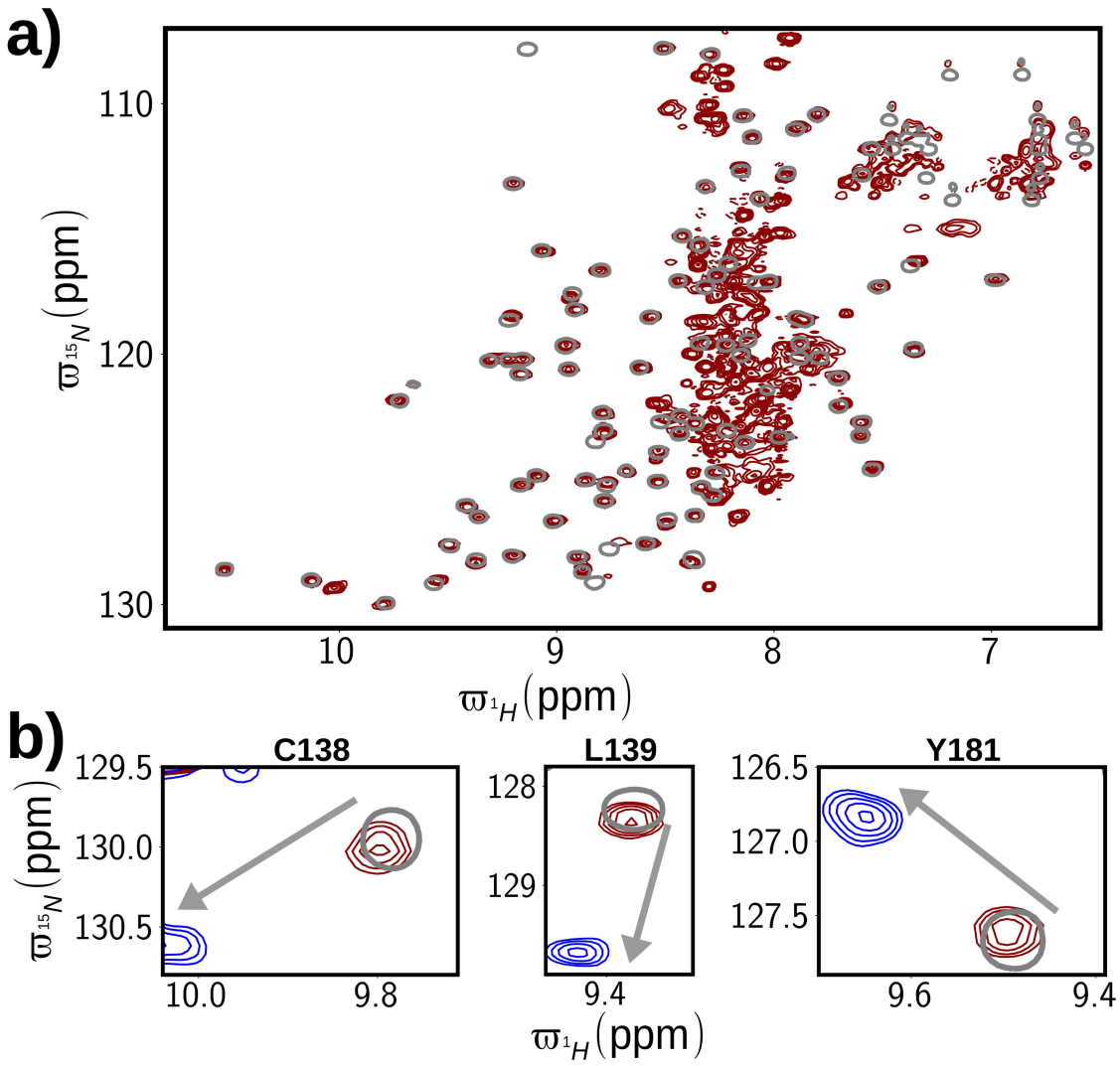


Figure S6. The L76G mutation severely destabilizes the V_L domain. (**a**) The $^{1}H,^{15}N$ HSQC spectrum of L76G+C215S JTO-FL (dark red contours) shows a combination of cross-peaks with large chemical shift dispersions, typical of a folded domain, and cross-peaks centered around 7.7-8.6 ppm in the ^{1}H dimension, typical of disordered regions. All peaks outside the 7.7-8.6 ppm region correspond to the C_L domain, as shown by the almost perfect overlap with the HSQC spectrum of the isolated C215S C_L domain (single grey contour). Therefore L76G+C215S JTO-FL has an unfolded V_L domain and a folded C_L domain. (**b**) Selected regions of $^{1}H,^{15}N$ spectra of L76G+C215S JTO-FL (dark red

contours), wild-type JTO-FL (blue contours) and C215S C_L (thick gray contour), showing that the C_L-C_L interface is absent for L76G+C215S JTO-FL.

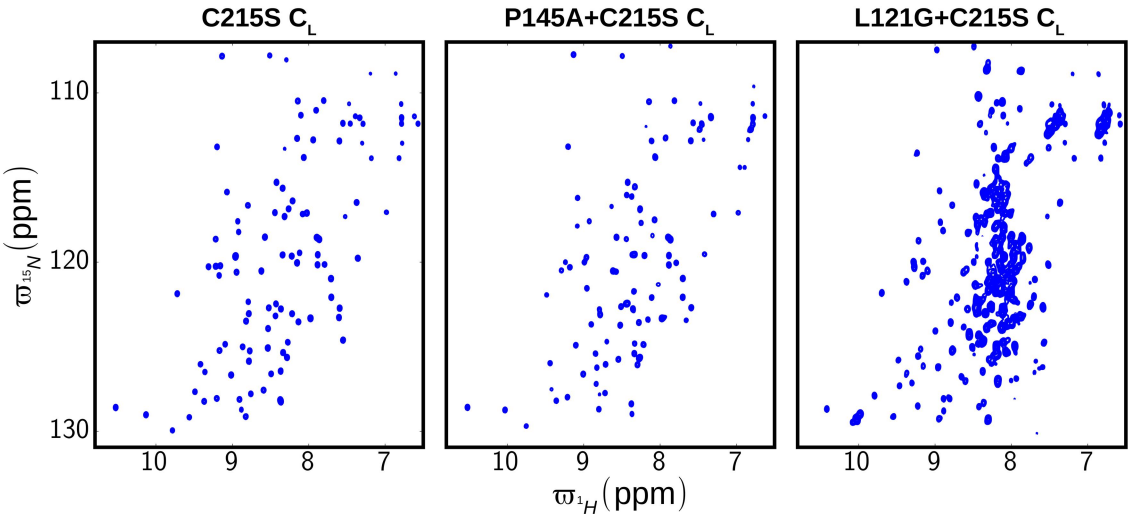


Figure S7. The L121G+C215S C_L domain is in equilibrium between folded and unfolded conformations. $^{1}H,^{15}N$ HSQC spectra of isolated C_L domains at 0.025 mM protein concentration (37 ^{o}C and 14.1 T). In addition to resonances from the folded state, intense correlations are also observed over a narrow region in the ^{1}H dimension for the L121G+C215S C_L double mutant that derive from unfolded protein.

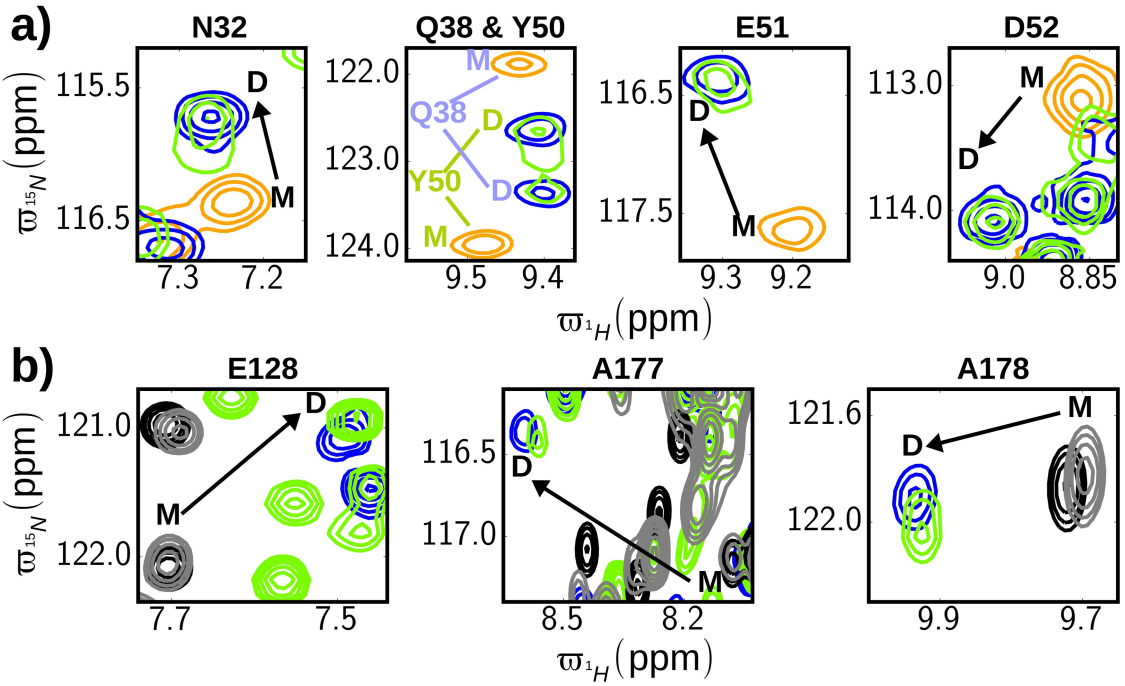


Figure S8. Selected regions of 2D $^{1}H,^{15}N$ spectra of L121G+C215S JTO-FL showing intact V_L-V_L and C_L-C_L interfaces. (**a**) Cross-peaks reporting on the V_L-V_L interface from L121G+C215S JTO-FL (lawn green) are compared with the corresponding peaks from WT JTO-FL (blue; closed V_L-V_L interface) and F101D JTO-FL (orange; open interface). (**b**) Comparison of peaks reporting on the C_L-C_L interface, with color-coding as above, including correlations from C215S C_L (black; open interface) and L121G+C215S C_L (grey; open interface) domains. M and D denote the positions of cross-peaks without (M) and with (D) a WT C_L-C_L interface. Note, ~ 50% of the L121G+C215S FL molecules do not have a folded C_L domain; these also show no C_L-C_L interface, although WT-like V_L-V_L interfaces are present.

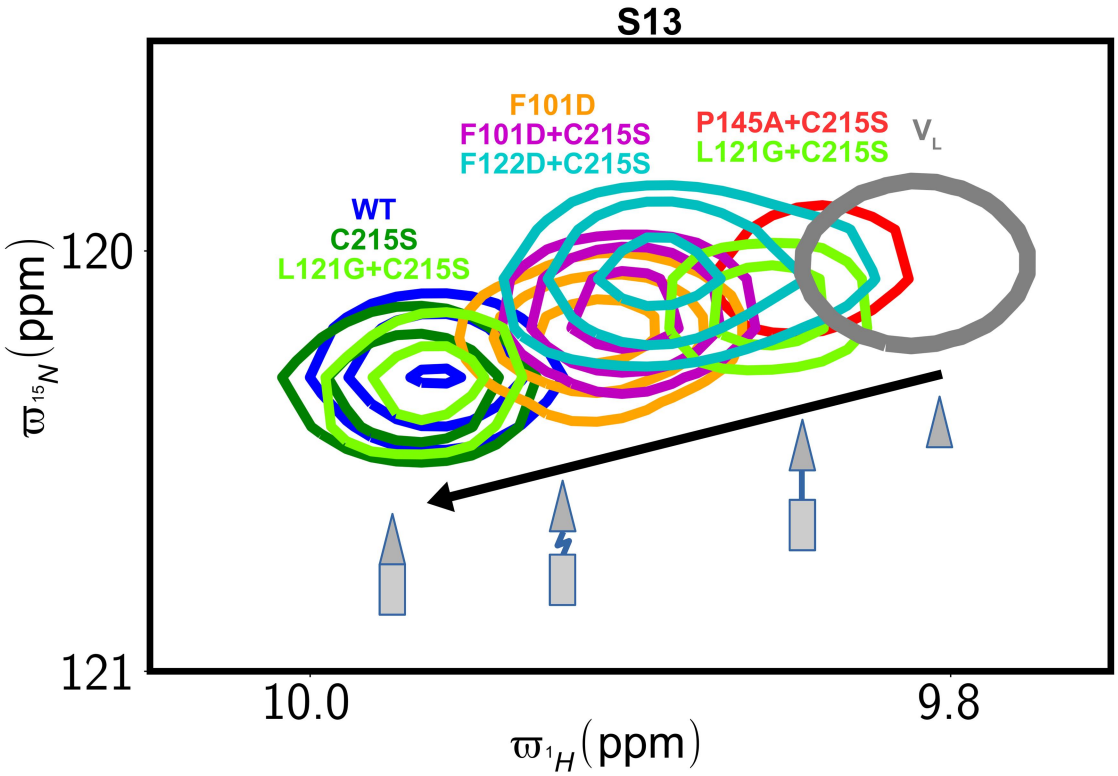


Figure S9. The V_L-C_L interface is only partially formed in a number of JTO-FL mutants. Regions of ^{1}H,^{15}N spectra focusing on S13 that is a reporter of the V_L-C_L interface. Note that there are two populations of L121G+C215S molecules, corresponding to those with and without folded C_L domains; cross-peaks from both class of molecule are shown with about the same intensity. Also indicated are schematics of a single monomer ranging from an intact interface (WT, C215S, L121G+C215S with folded C_L domains) to no interface (V_L). The most stable V_L-C_L interface is observed in cases where both V_L-V_L and C_L-C_L interfaces are closed.

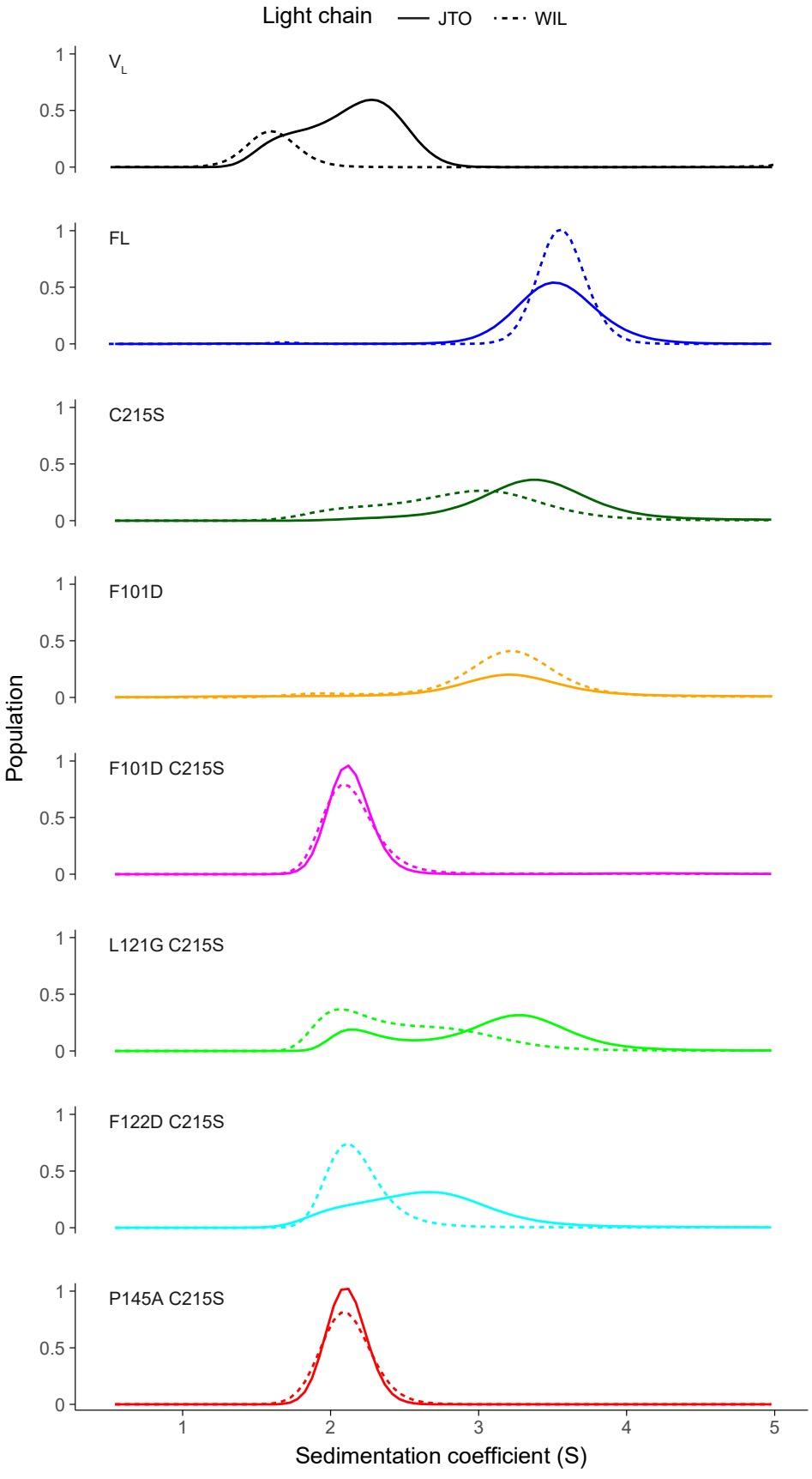

Light chain
JTO
WIL
V_L
FL
C215S
F101D
F101D C215S
L121G C215S
F122D C215S
P145A C215S
Population
0
0.5
1
1
2
3
4
5
Sedimentation coefficient (S)

Figure S10. Analytical ultracentrifugation profiles for JTO- and WIL-FL mutants at 20 μM protein concentration in PBS, 20 °C.

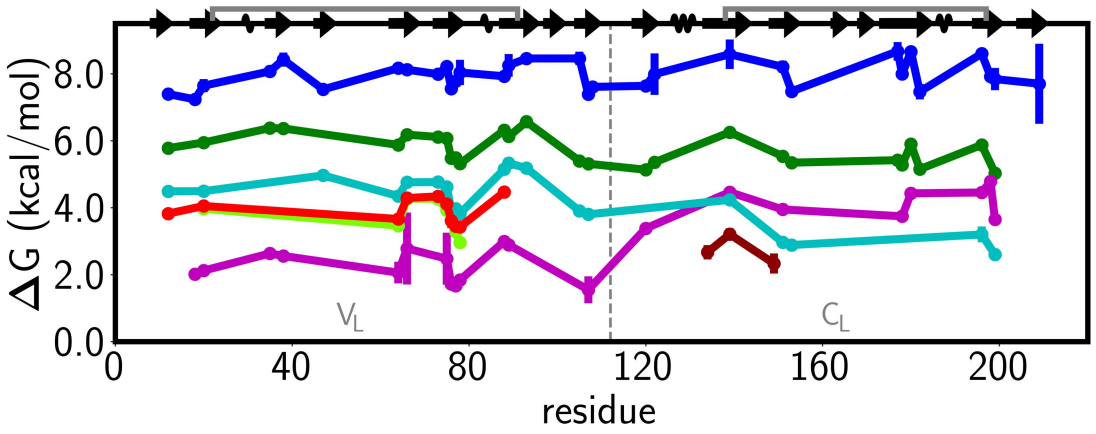


Figure S11. Free energy profiles as a function of residue in JTO-FL variants, as measured by solvent hydrogen exchange. Shown are ΔG values for the amides exchanging with the largest protection factors in each of the V_L and C_L domains of the FL constructs, so as to provide an estimate of unfolding free energies (21). From these data average ΔG values for V_L and C_L domains for each JTO-FL variant were calculated (data reported in Fig. 6a). Protein concentrations were 0.2 mM (monomer), with the exception of WT (where values are concentration independent as the LC dimers are covalently linked via the C215 disulfide) and L76G+C215S where concentrations of 1.1 mM and 0.4 mM were used, respectively. Color code is: WT, blue; C215S, green; L76G+C215S, brown; F101D+C215S, magenta; L121G+C215S, lawn green; F122D+C215S, light blue; P145A+C215S, red.

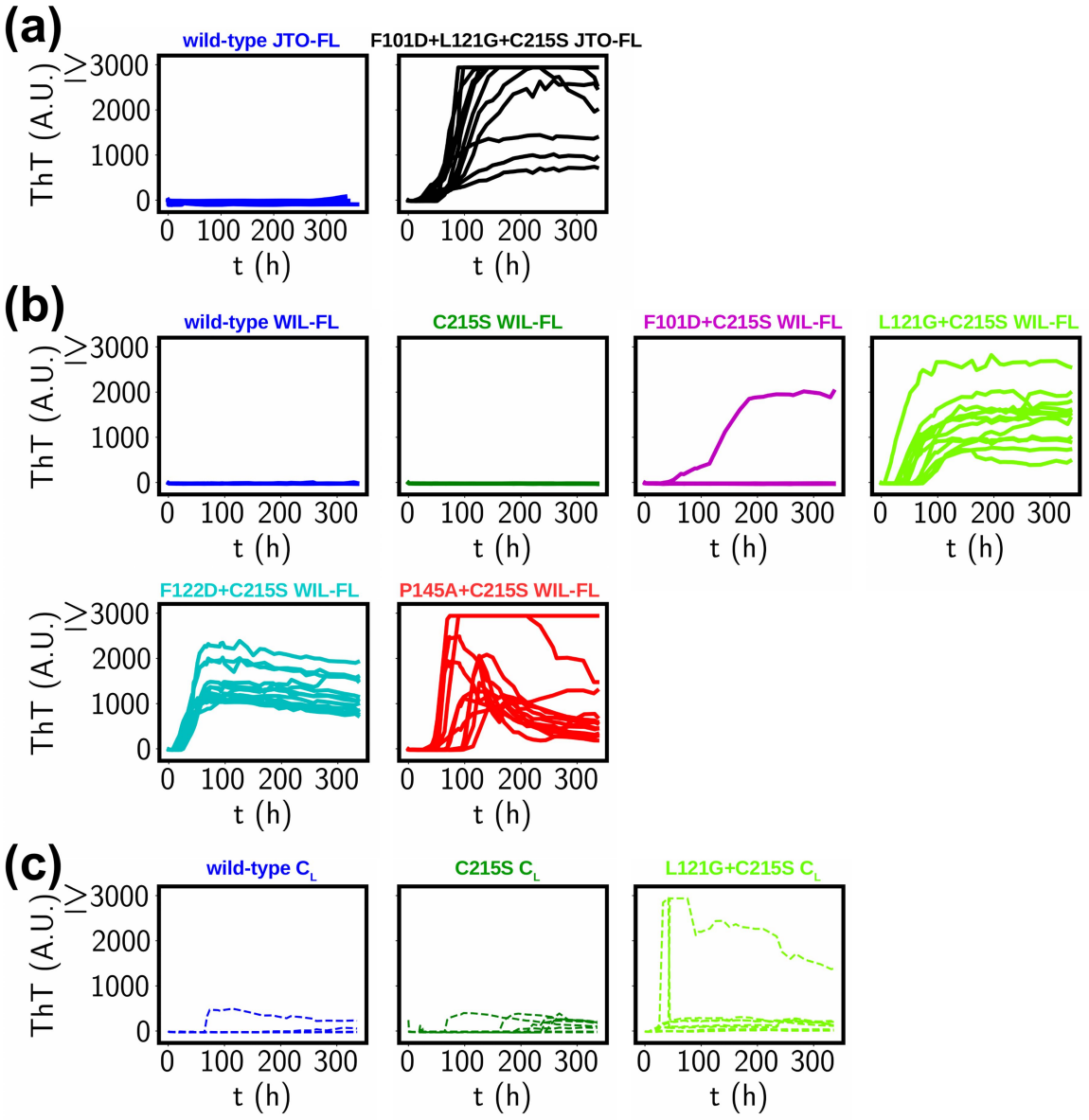


Figure S12. (a) Thioflavin T fluorescence kinetics of the aggregation of WT (10 profiles) and F101D+L121G+C215S (12 profiles) JTO-FL. (b) Thioflavin T fluorescence kinetics of the aggregation of the following WIL-FL constructs: WT (4 profiles), C215S (12 profiles), F101D+C215S (12 profiles; only 1 of the 12 shows aggregation), L121G+C215S (12 profiles), F122D+C215S (12 profiles) and P145A+C215S (12 profiles) WIL-FL. (c) Thioflavin T fluorescence kinetics of the aggregation of WT (12 profiles), C215S (12 profiles) and L121G+C215S (12 profiles) C_L domains.

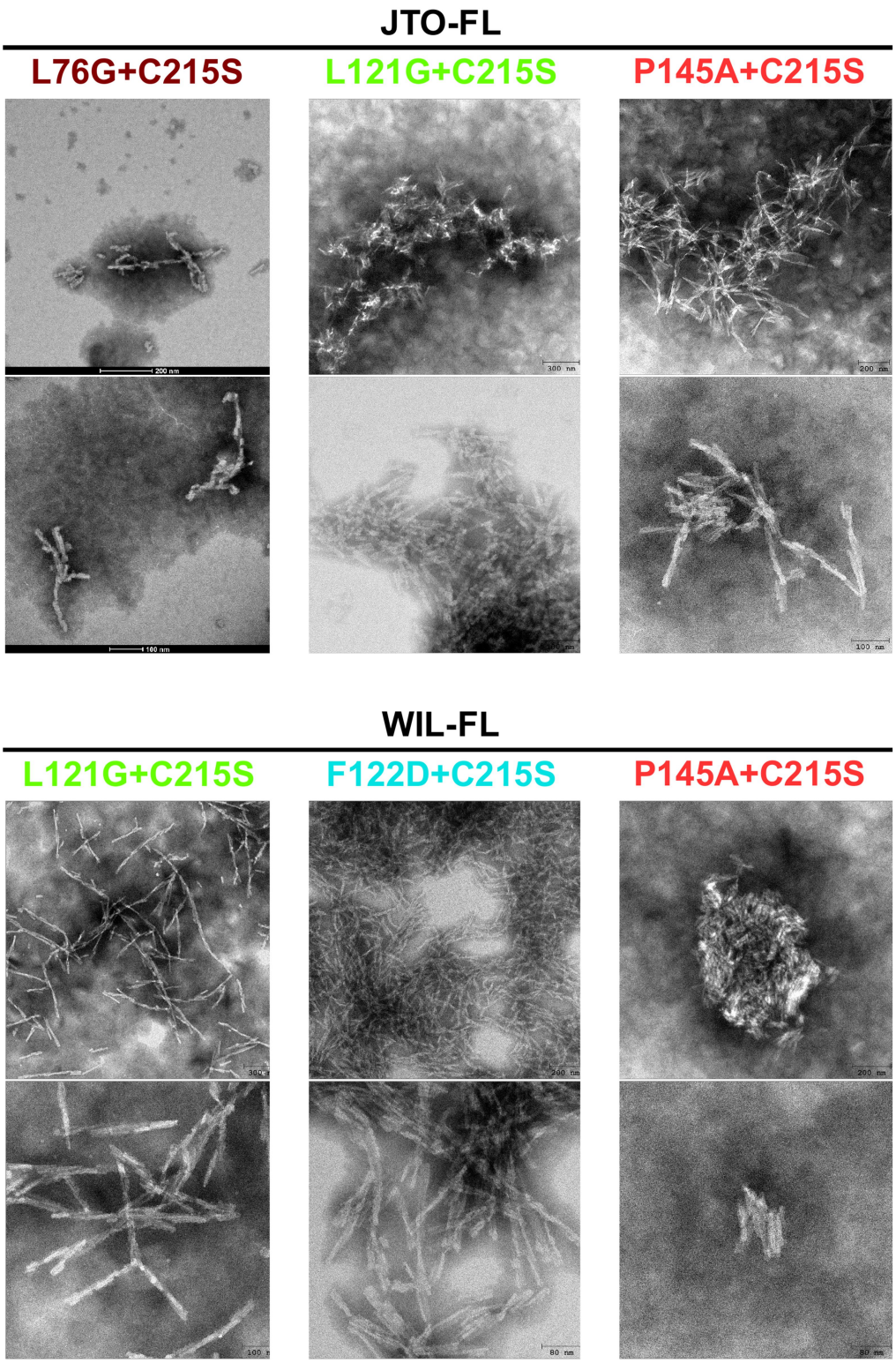


Figure S13. Electron micrographs of aggregates from JTO-FL and WIL-FL variants showing the formation of fibrillar structures.

Table S1. Proteins and techniques used to acquire the data shown in this paper

construct		HSQC	CPMG	HX[a]	AUC[b]	Aggregation assay
JTO-V_L	wild-type	✓	✓	✓	✓	✓
	F101D	✗	✗	✗	✗	✓
WIL-V_L	wild-type	✓	✓	✓	✓	✓
6aJL2-V_L	wild-type	✓	✗	✗	✗	✗
	L76G	✓	✗	✗	✗	✗
C_L	wild-type	✓	✗	✗	✗	✓
	C215S	✓	✓	✓	✗	✓
	L121G+C215S	✓	✗	✗	✗	✓
	P145A+C215S	✓	✗	✗	✗	✗
JTO-FL	wild-type	✓	✓	✓	✓	✓
	F101D	✓	✓	✗	✓	✗
	C215S	✓	✓	✓	✓	✓
	L76G+C215S	✓	✗	✓	✗	✓
	F101D+C215S	✓	✓	✓	✓	✓
	L121G+C215S	✓	✗	✓	✓	✓
	F122D+C215S	✓	✓	✓	✓	✓
	P145A+C215S	✓	✓	✓	✓	✓
	F101D+L121G+C215S	✗	✗	✗	✗	✓
WIL-FL	wild-type	✓	✓	✓	✓	✓
	F101D	✗	✗	✗	✓	✗
	C215S	✓	✓	✓	✓	✓
	F101D+C215S	✗	✗	✗	✓	✓
	L121G+C215S	✗	✗	✗	✓	✓
	F122D+C215S	✗	✗	✗	✓	✓
	P145A+C215S	✓	✓	✗	✓	✓

[a]Hydrogen exchange

[b]Analytical ultracentrifugation

References

1. 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.
2. Mulder FA, Skrynnikov NR, Hon B, Dahlquist FW, & Kay LE (2001) Measurement of slow (micros-ms) time scale dynamics in protein side chains by (15)N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J Am Chem Soc* 123(5):967-975.
3. Jiang B, Yu B, Zhang X, Liu M, & Yang D (2015) A (15)N CPMG relaxation dispersion experiment more resistant to resonance offset and pulse imperfection. *J Magn Reson* 257:1-7.
4. Vallurupalli P, Hansen DF, Stollar E, Meirovitch E, & Kay LE (2007) Measurement of bond vector orientations in invisible excited states of proteins. *Proc Natl Acad Sci U S A* 104(47):18473-18477.
5. Tollinger M, Skrynnikov NR, Mulder FA, Forman-Kay JD, & Kay LE (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *J Am Chem Soc* 123(46):11341-11352.
6. Korzhnev DM*, et al.* (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430(6999):586-590.
7. Bouvignies G & Kay LE (2012) Measurement of proton chemical shifts in invisible states of slowly exchanging protein systems by chemical exchange saturation transfer. *J Phys Chem B* 116(49):14311-14317.
8. McConnell HM (1958) Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.* 28(3):430-431.
9. Palmer AG, 3rd, Kroenke CD, & Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* 339:204-238.
10. Kay LE, Keifer, P., Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114(26):10663-10665.
11. Pervushin K, Riek R, Wider G, & Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* 94(23):12366-12371.
12. Hvidt A & Nielsen SO (1966) Hydrogen exchange in proteins. *Adv Protein Chem* 21:287-386.
13. Wanker EE*, et al.* (1999) Membrane filter assay for detection of amyloid-like polyglutamine-containing protein aggregates. *Methods Enzymol* 309:375-386.
14. Abramoff MDM, Paulo J.; Ram, Sunanda J. (2004) Image processing with ImageJ. *Biophotonics International* 11(7):36-42.
15. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78(3):1606-1619.

16. Bai Y, Sosnick TR, Mayne L, & Englander SW (1995) Protein folding intermediates: native-state hydrogen exchange. *Science* 269(5221):192-197.
17. Hunter CA & Anderson HL (2009) What is cooperativity? *Angew Chem Int Ed Engl* 48(41):7488-7499.
18. Mazor MH, Wong, C. F., McCammon, A., Deutch, J. M., Whitesidesd, G. (1990) Effective molarity in diffusion-controlled reactions. *J Phys Chem* 94(9):3807-3812.
19. Sievers F & Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 1079:105-116.
20. Pokkuluri PR, Solomon A, Weiss DT, Stevens FJ, & Schiffer M (1999) Tertiary structure of human lambda 6 light chains. *Amyloid* 6(3):165-171.
21. Bai Y, Milne JS, Mayne L, & Englander SW (1994) Protein stability parameters measured by hydrogen exchange. *Proteins* 20(1):4-14.