Supporting Information

Characterising side chain conformations in proteins using ^{19}F-Gd^{3+} ENDOR and new fluorinated amino acid spin labels

Martyna Judd,[1,2*] Elwy H. Abdelkader,[3] Haocheng Qianzhu,[1] Edan Habel,[1] Caitlin Lindsay,[1] Anthony Brancewicz[1], Thomas Huber,[1] Gottfried Otting,[3] and Nicholas Cox*[3]

[1]Research School of Chemistry, The Australian National University, Canberra, ACT, 2601, Australia
[2]Department of Chemistry, Weinberg College of Arts and Sciences, Northwestern University, Evanston, IL 60208, USA
[3]ARC Centre of Excellence for Innovation in Peptide and Protein Science, The Australian National University, Canberra, ACT, 2601, Australia

Table of Contents

SI 1. Experimental – Genetic encoding and protein expression

4-F-3,5-Cl_2Phe instead of 4-F-Phe was targeted for genetic encoding to avoid misincorporation of 4-F-Phe in response to the common phenylalanine codons. For the site-specific incorporation of 4-F-3,5-Cl_2Phe, we screened a library of variants of the pyrrolysyl-tRNA synthetase derived from the methanogenic archaeon ISO4-G1 (G1PylRS) for mutants capable of recognizing 4-F-3,5-Cl_2Phe as a specific substrate.[1] A previously established library of G1PylRS mutants, encoded on the pBK-G1RS plasmid, was transformed into *E. coli* DH10B cells harbouring the pBAD-H6RFP reporter plasmid. After transformation, the culture was directly inoculated into 25 mL LB medium supplemented with 100 mg/L carbenicillin, 50 mg/L kanamycin, 0.2% L-arabinose and 1 mM 4-F-3,5-Cl_2Phe. This culture served as the sample for the first round of positive selection (**1P+**). Overnight expression at 37 °C led to readily detectable level of RFP expression.

Following the overnight incubation, the cells were harvested, resuspended in 5 mL of PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) and diluted 100-fold to a concentration suitable for fluorescence-activated cell sorting (FACS). FACS was performed using a FACSAria Fusion cell sorter (BD Biosciences, USA; Figure S1). Cells with high RFP levels were collected from the **1P+** sample, representing 0.8% of the total population (indicated by violet shades in Figure S1), and subjected to a subsequent round of negative selection (**2N–**) in the absence of 4-F-3,5-Cl_2Phe. Cells exhibiting low RFP expression (67.2%) were collected from the **2N–** sample and aliquoted to inoculate media under positive (**3P+**, with 4-F-3,5-Cl_2Phe) and negative (**3P–**, without 4-F-3,5-Cl_2Phe) conditions. The top RFP-expressing cells from the **3P+** sample (8.4%) were collected and underwent another negative selection round (**4N–**, without 4-F-3,5-Cl_2Phe). From the **4N–** sample, cells showing low RFP expression (33.1%) were collected and aliquoted again to inoculate media under positive (**5P+**, with 4-F-3,5-Cl_2Phe) and negative (**5P–**, without 4-F-3,5-Cl_2Phe) conditions. The cultures demonstrated a clear response to the presence of 4-F-3,5-Cl_2Phe and both the top positive (5.1%) and total positive (65.7%) of RFP-fluorescent cells from the **5P+** sample were collected and recovered by plating on LB agar plates containing 100 mg/L carbenicillin and 50 mg/L kanamycin. Isolated colonies were analyzed in 96-well plates. 120 enzyme candidates from both the top and total positive populations were inoculated into media under both positive and negative conditions. The red fluorescence intensity was measured as an indicator of RFP expression and normalized to the OD_{600} of the cell culture using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland). DNA sequence analysis identified the best candidate, G1-dClFFRS, as a novel G1PylRS mutant incorporating 4-F-3,5-Cl_2Phe, carrying mutations N165S, A221R, and W237S. A previously published G1PylRS mutant, G1-F5FRS (G1-F5F29), was used to recognize both 3,5-F_2-Phe and 2,6-F_2Phe as substrate.[1]

Protein Expression. The G1-dClFFRS and G1-F5FRS enzymes were individually cloned into the high-copy number pRSF plasmid for use in our previously reported dual-plasmid system for efficient amber stop codon suppression.[1] C_{D9k} samples with pcAAs were produced by co-transforming *E. coli* B-95DAD*fabR* cells with pCDF-CD9k-F50TAG or pCDF-CD9k-F66TAG and the corresponding pRSF plasmids. The recovered cells were grown at 37 °C in LB medium containing 50 mg/L kanamycin and 50 mg/L spectinomycin. A 2 mL overnight culture was used to inoculate 100 mL LB medium supplemented with the antibiotics and 1 mM pcAA. The cells were grown at 37 °C to an OD_{600} of 0.6–1, where the temperature was reduced to 25 °C and protein expression induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

Following protein expression, cells were harvested by centrifugation at 4,000 *g* for 15 minutes at 4 °C. The cell pellet was resuspended in buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM imidazole) and lysed on ice by sonication using 50% power and 50% pulse length for 10 min (ultrasonic homogenizer Omni-Ruptor 4000, Omni International, USA). The cell lysate was centrifuged at 30,000 *g* for 1 h at 4 °C. The clarified supernatant was then loaded onto a 1 mL His GraviTrap column (Cytiva, USA). After washing with 20 column volumes of buffer B (same as buffer A but with 20 mM imidazole), the target protein was eluted using 5 column volumes of buffer C (same as buffer A but with 500 mM imidazole). Finally, buffer exchange to buffer D (50 mM Tris-HCl pH 7.5, 300 mM sodium chloride, 1 mM dithiothreitol (DTT)) was performed using an Amicon ultrafiltration centrifugal filter with a 10 kDa molecular weight cut-off, and the TEV proteolytic cleavage was conducted by incubation with TEV protease at 4 °C for 16 h. The cleaved C_{D9k} protein was recovered by a reverse IMAC step.

Intact protein analysis was performed on an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific, USA) connected to a Thermo Fisher Scientific UltiMate 3000 HPLC system equipped with a ZORBAX 300SB-C3, 3.5 µm, 4.6 x 50 mm HPLC column (Agilent Technologies, USA). Approximately 30 pmol of sample was injected using a 500 µL/min linear gradient of solvent A (0.1% (v/v) formic acid in water) and solvent B (0.1% (v/v) formic acid in acetonitrile), ramping solvent B from 5% at 2 min to 80% after 7 min. Data were collected using an electrospray ionization (ESI) source in positive ion mode. Protein intact mass was determined by deconvolution using the program Xcalibur 3.0.63 (Thermo Fisher Scientific, USA). The results of intact protein mass spectrometry are shown in Figure S1

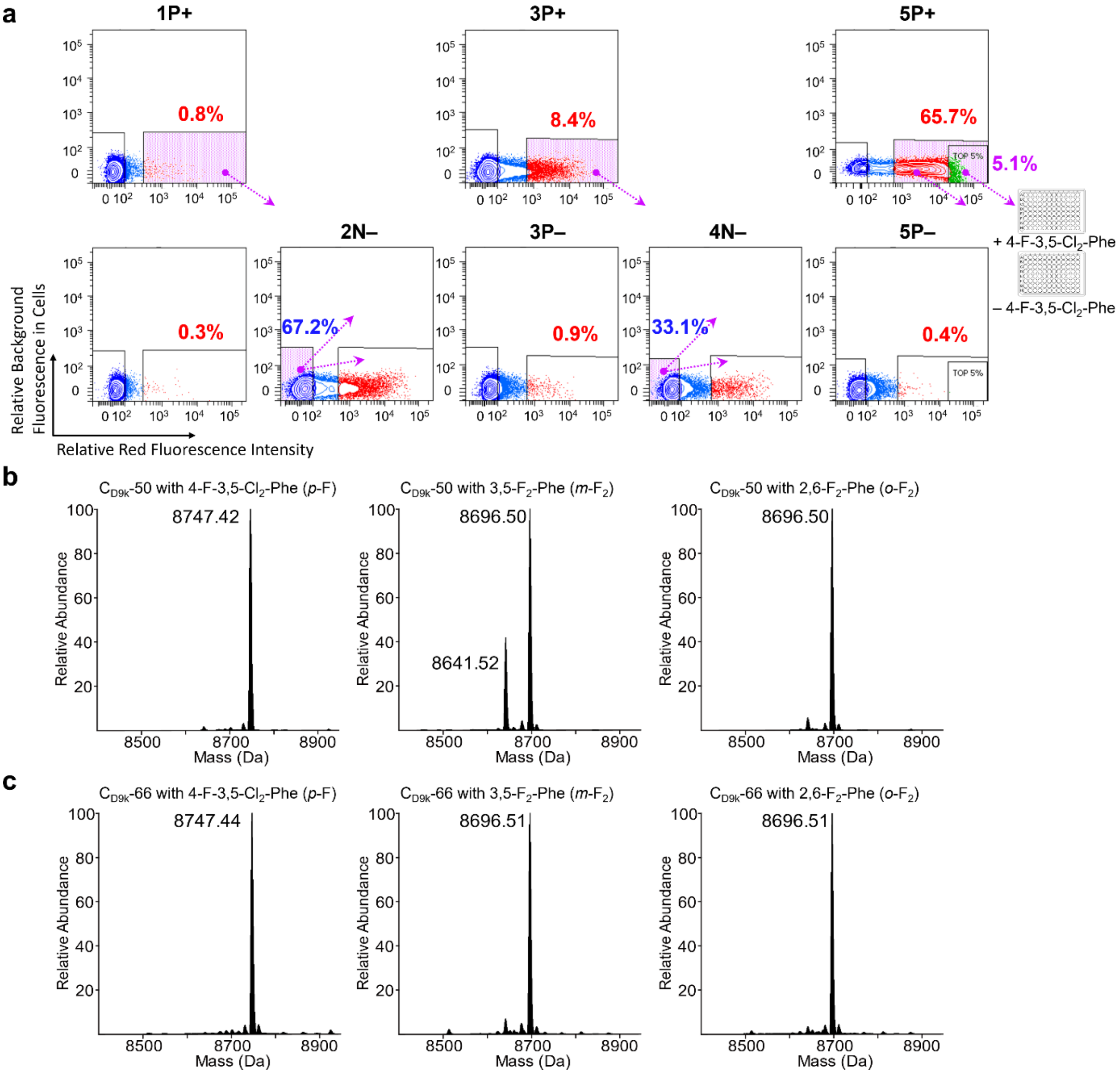

Figure S1. (a) FACS screening of G1PylRS variants for activity and specificity in recognizing 4-F-3,5-Cl$_2$Phe. The horizontal axis of the scatter plots represents red fluorescence intensity (excitation at 560 nm), while the vertical axis indicates background fluorescence in cells excited at 488 nm. Positive (**P**) and negative (**N**) selection rounds are indicated with "+", denoting growth conditions with 1 mM 4-F-3,5-Cl$_2$Phe, and "-" for conditions without 4-F-3,5-Cl$_2$Phe. Violet-shaded regions highlight the collected cell populations. Arrows point to the subsequent selection strategy applied after amplification by culturing. (b) Intact protein mass spectrometric analysis results of C$_{D9k}$-50. The expected mass is 8747.66 Da with 4-F-3,5-Cl$_2$Phe incorporated or 8696.76 Da with either 3,5-F$_2$-Phe or 2,6-F$_2$Phe incorporated. The peak of 8641.52 Da corresponds to the product of amber codon suppression by glutamine (expected mass: 8641.73 Da). (c) Intact protein mass spectrometric analysis results of C$_{D9k}$-66.

SI 2. EPR T_1 and T_M relaxation measurements of C_{D9k} proteins

All EPR and ^{19}F ENDOR measurements were performed at 10 K on a modified W-band 94 GHz E680 Bruker spectrometer[2] equipped with a TE_{011}-mode EPR/ENDOR resonator in a helium gas flow cryostat (Oxford Instruments).

All C_{D9k} proteins were measured at 100 µM concentration in a buffer of 50 mM MES in D_2O, pH 6.5, 50 mM NaCl containing 10% glycerol-d_8. The GB1 proteins were ere measured at 50 µM concentration in a buffer of 50 mM MES in D_2O, pH 6.5, 50 mM NaCl containing 10% glycerol-d_8. In addition to ^{19}F ENDOR measurements, we performed EPR field-sweep lineshape measurements to determine the magnetic field for the maximum Gd^{3+} echo intensity, as well as inversion recovery and electron spin echo decay measurements to characterise the relaxation properties of each sample. The relaxation data for the GB1 **Gd.C1**-labelled proteins have been reported previously.[3,4] The relaxation data for the C_{D9k} proteins is discussed below.

Gd^{3+} field-sweep spectra were recorded with the two-pulse Hahn-echo sequence $\pi/2 - t - \pi - t -$ echo, with $\pi/2 = 20$ ns, $\pi = 40$ ns and $\tau = 300$ ns. The echo-detected EPR lineshapes for each protein are shown in Figure S2.1, annotated with the full width at half maximum of the central Gd^{3+} $\left|-\frac{1}{2}\right\rangle \leftrightarrow \left|+\frac{1}{2}\right\rangle$ transition. In general, the line shapes look as expected: the linewidths of the **Gd.C1**-labelled proteins are significantly narrower than those of the protein-bound Gd^{3+} centre in C_{D9k}, reflecting a smaller zero field splitting (ZFS) in the symmetric, homogeneous environment of the tag compared to the protein metallocentre. The exception is the spin-labelled protein **2b,** where we observe a much broader linewidth compared to the other GB1 constructs. We suspect that this is due to part of the protein being partially misfolded or degraded. In Figure S2.1 we decompose the echo-detected spectrum of **2b** to show that it can be modelled as a sum of a broad population and a narrow component (in about 40% population), the latter being the same as that seen in other GB1 proteins. Note that the ^{19}F-ENDOR data collected on **2b** dominantly samples the narrow component, as ENDOR spectra are collected on the field maximum. Thus, the effect of the second broader population is to reduce the S/N of collected ENDOR data. It also makes analysis of the ^{19}F-ENDOR intensity for **2b** impossible.

T_1 relaxation for all samples was measured by the saturation–recovery sequence $\pi - (t + dt) - \pi/2 - \tau - \pi -$ echo, with $\pi = 40$ ns, $\tau = 300$ ns, and a delay step $dt = 1$ µs. The integrated echo intensity was recorded as a function of the delay time t (incremented by the time step dt). T_M relaxation was measured by recording the decay of the integrated echo intensity with time, using the pulse sequence $\pi/2 - (\tau + \mathrm{d}\tau) - \pi - (\tau + \mathrm{d}\tau) -$ echo, using the same pulse settings as above but with $\mathrm{d}\tau = 24$ ns. The relaxation data were fit using a MATLAB least squares fitting algorithm to the stretched exponential functions as detailed in Table S2.1. The plotted curves and fits are shown in Figure S2.2 for the C_{D9k} proteins. The relaxation data for the GB1 proteins has been reported in ref.[4] with the exception of **2b**. We reproduce the fitting parameters in Table S2.1 for the convenience of the reader, and add the fitting parameters for **2b** as well.

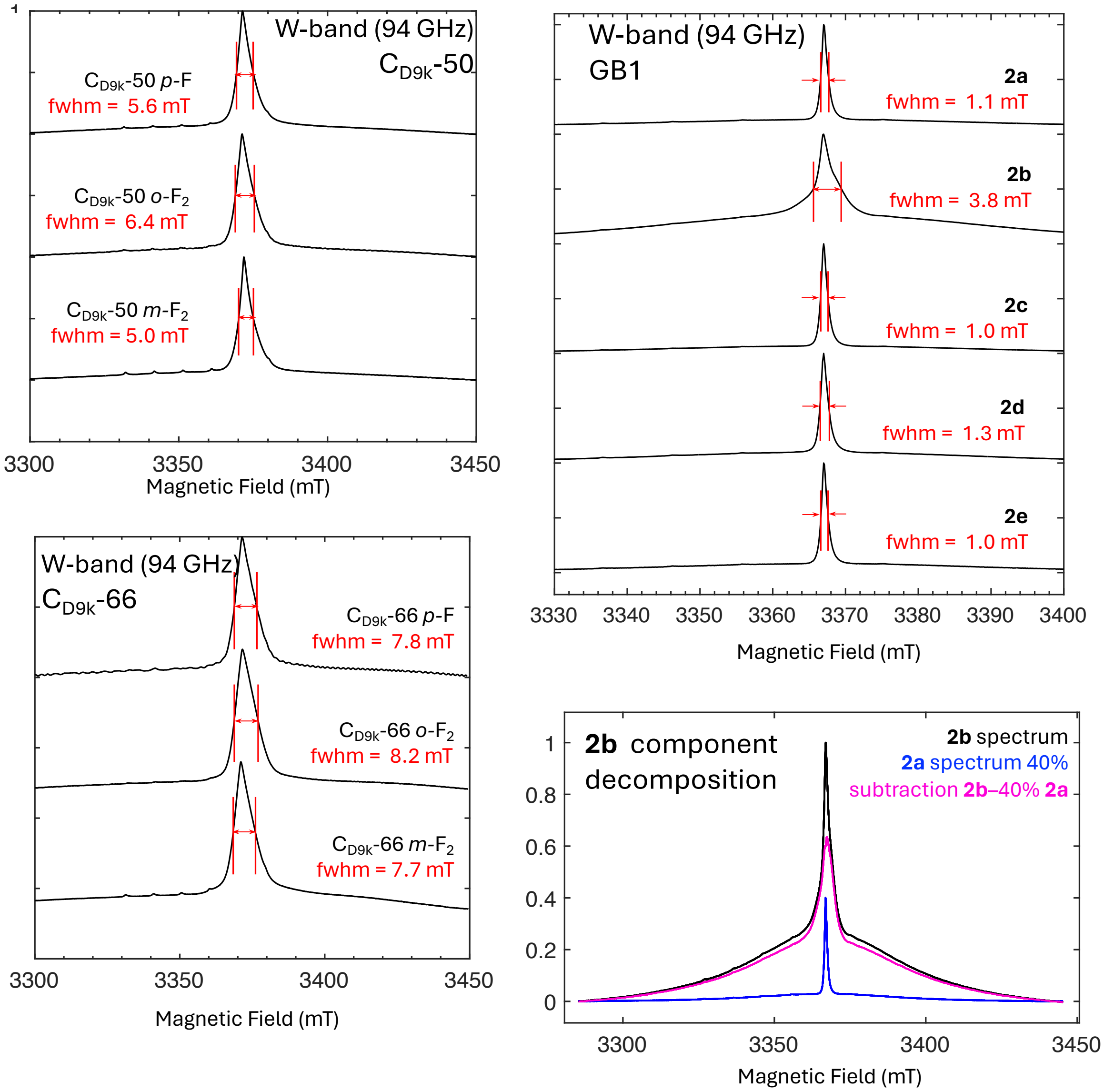

Figure S2.1 W-band (94 GHz) echo-detected field sweep of the C_{D9k}-50, C_{D9k}-66 and GB1 proteins investigated in this work. The full widths at half maximum (fwhm) of the Gd^{3+} central transitions are annotated next to each trace. In the bottom ride hand side we show the decomposition of the echo-detected spectrum of **2b**, showing it can be decomposed into a broad component (pink) and the same narrow component displayed by the other proteins **2**.

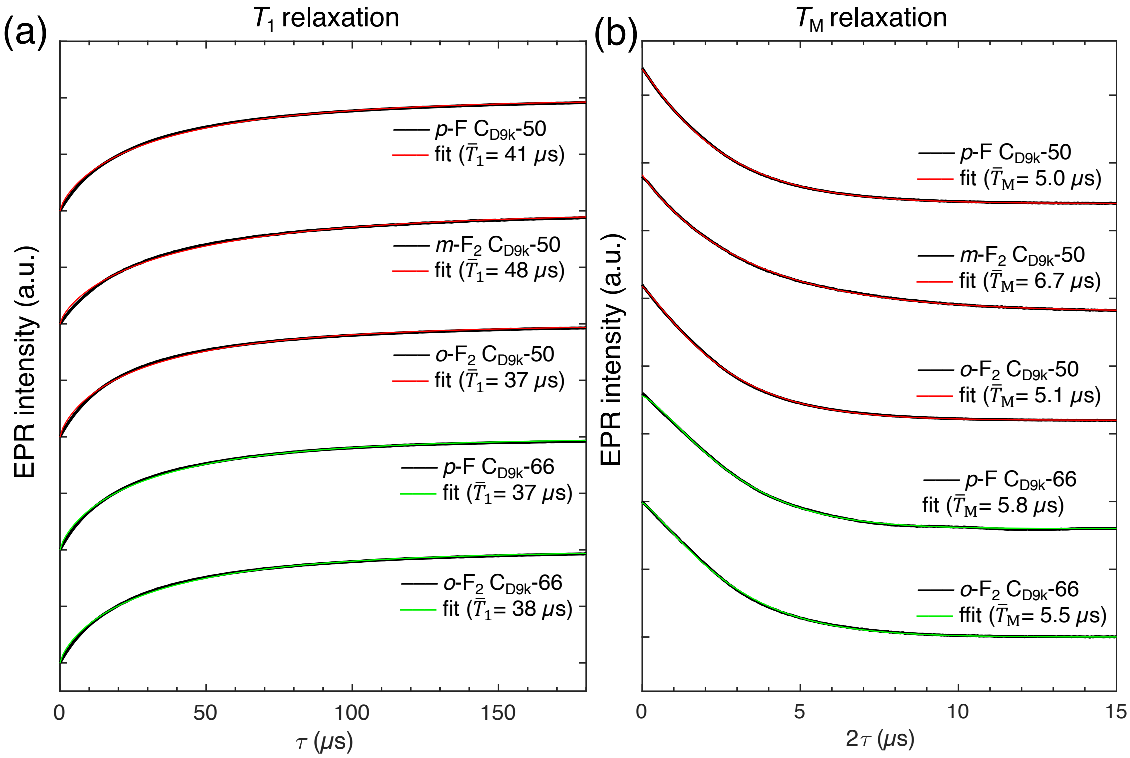

Figure S2.2 Relaxation curves (black traces) and fits (colored traces) of the C_{D9k} proteins labelled with *p*-F, *m*-F_2 and *o*-F_2. (a) T_1 relaxation data and fits using the stretched exponential fitting function $f_{T1} = A\left(1 - \exp\left[-\left(\frac{x}{\tau_1}\right)^{\alpha}\right]\right)$ as described in Table S2.1. The mean best-fit T_1 value is annotated beside each curve. (b) T_M relaxation data and fits using the stretched exponential fitting function $f_{TM} = \mathrm{A}\exp\left[-\left(\frac{2x}{\tau_2}\right)^{\alpha}\right]$ as described in Table S2.1. The mean best-fit T_M value is annotated beside each curve.

Table S2.1. Summary of the fitting parameters for the T_1 and T_M relaxation time data.[a]

Protein	$\overline{T}_1(\tau_1, \alpha)$ / μs	$\overline{T}_M(\tau_2, \alpha)$ / μs
p-F C_{D9k}-50	41 (34, 0.73)	5.0 (5.1, 1.1)
m-F_2 C_{D9k}-50	48 (40, 0.73)	6.7 (6.6, 1.0)
o-F_2 C_{D9k}-50	37 (30, 0.73)	5.1 (5.4, 1.2)
p-F C_{D9k}-66	37 (31, 0.74)	5.8 (6.2, 1.3)
o-F_2 C_{D9k}-66	38 (32, 0.74)	5.5 (5.8, 1.2)
2a (GB1 Q32CF_3Phe/K28**Gd.C1**)	95 (76, 0.71)	10.3 (11.0, 1.2)
2b (GB1 F30CF_3Phe/K28**Gd.C1**)	44 (37, 0.75)	7.6 (8.1, 1.3)
2c (GB1 V21CF_3Phe/K28**Gd.C1**)	82 (67, 0.73)	11.6 (12.7, 1.3)
2d (GB1 F52CF_3Phe/K28**Gd.C1**)	82 (73, 0.80)	10.2 (10.8, 1.2)
2e (GB1 G14CF_3Phe/K28**Gd.C1**)	99 (81, 0.73)	9.8 (10.3, 1.2)

[a] T_1 data were fitted to the stretched exponential function $f_{T1} = A\left(1 - \exp\left[-\left(\frac{x}{\tau_1}\right)^{\alpha}\right]\right)$ and the average T_1 time was calculated using the gamma function distribution $\overline{T}_1 = \frac{\tau_1}{\alpha} \cdot \boldsymbol{\Gamma}\left(\frac{1}{\alpha}\right)$. T_M data were fit to the stretched decay exponential function $f_{TM} = A\exp\left[-\left(\frac{2x}{\tau_2}\right)^{\alpha}\right]$ and the average T_M time for each stretch exponential fit was also calculated assuming a gamma function distribution, $\overline{T}_2 = \frac{\tau_2}{\alpha} \cdot \boldsymbol{\Gamma}\left(\frac{1}{\alpha}\right)$.

SI 3. ENDOR measurements and data analysis

^{19}F Mims ENDOR experiments

^{19}F Mims ENDOR spectra were acquired at the field maximum of the Gd^{3+} EPR-spectrum, using the Mims ENDOR pulse sequence[5] $\pi/2 - \tau - \pi/2 - t_{\pi,\mathrm{RF}} - \pi/2 - t -$ echo, with a microwave pulse length of $\pi/2 = 12$ ns, and an RF pulse length of $t_{\pi,\mathrm{RF}} = 60$ µs for GB1 protein measurements and $t_{\pi,\mathrm{RF}} = 48$ µs for the $\mathrm{C_{D9k}}$ proteins with the RF pulse power adjusted to achieve the maximum ENDOR response (calibrated by RF pulse length nutation measurements). Because of the stimulated echo detection sequence (i.e. three π/2 pulses), the Mims ENDOR spectrum is subject to "blindspotting" artefacts, which manifest in missing intensity in the ENDOR lineshape, which depends sinusoidally on the interpulse delay time τ:

$$F_{\mathrm{ENDOR}} = \tfrac{1}{2} \sin^2(2\pi \frac{A\tau}{2}) \tag{1}$$

where F_{ENDOR} is the ENDOR signal response, and A is the electron–nuclear hyperfine coupling, which, in the dipolar coupling limit, is given by:

$$A(\theta) = \frac{\mu_0}{4\pi h} \frac{g_e \mu_B g_n \mu_n}{r^3} (3 \cos^2 \theta - 1) = (3 \cos^2 \theta - 1)\, A_\perp \tag{2}$$

where μ_0 is the permeability of free space, g_e and g_n are the electron and nuclear g-values respectively, μ_B and μ_n are the Bohr and nuclear magnetons respectively, r is the electron-nuclear distance, θ is the angle of the vector r with respect to the orientation of the external magnetic field B_0, and $A_\perp$ is the hyperfine interaction at the perpendicular orientation at $\theta =$ 90°, which corresponds to the value T_{read} (i.e. the ENDOR line splitting we refer to in the main text).

The experimental ^{19}F ENDOR spectra were measured for different τ values (τ = 0.5, 1.0, 1.5 and 2.0 µs) and summed to properly reproduce the ENDOR lineshape. All the experimental spectra presented in the main text show the sum of the individual τ-varied measurements. For long distances with couplings below about 50 kHz, the splitting falls outside the blindspotting range according to Eq. 1 (as long as $A \ll \frac{\pi}{\tau}$), and the ENDOR spectrum can be reliably acquired using a single long τ value. Additionally, the choice of τ needs to be balanced against the electron-spin phase memory time T_{M}, which accounts for signal decay during the delay interval. For $\mathrm{C_{D9k}}$ produced with p-F in position 66 ($T_{\mathrm{M}} \approx 6$ µs) we used τ = 1.5 µs. For the GB1 constructs **2b**–**2e** ($T_{\mathrm{M}} \approx 10$ µs) we used τ = 2.0 µs.

^{19}F-ENDOR analysis and simulations

The resolved splitting (T_{read}) was determined visually from each spectrum. The uncertainty in T_{read} was estimated as three times the resolution of the ENDOR frequency-axis;[6] e.g. for an RF pulse swept in 5 kHz increments, the uncertainty would be $\Delta T_{\text{read}} = \pm 15$ kHz. The uncertainty in the frequency was then converted to an uncertainty in r_{read} using the dipolar coupling equation above (Eq. 2).[3,6]

^{19}F-ENDOR simulations were performed using the MATLAB program *Easyspin*, employing the *saffron* script for Mims ENDOR simulations.[7] The simulations were parameterised for the geometries of the C_{D9k} and GB1 protein systems as described in the main text (more detail is provided below). Best-fit parameters were determined by the nonlinear optimisation function *fminsearch* in MATLAB. Where Mims ENDOR data were collected for different τ values, the same parameter set was used to fit all spectra simultaneously. All fitting was performed on symmetrised, unsmoothed experimental data.

Rotamer library-based ^{19}F-ENDOR simulations were performed by first loading the coordinates generated by the PyParaTools rotamer modelling Python module[8,9] into MATLAB. The ^{19}F–Gd^{3+} distances were calculated from the coordinates of each rotamer and entered into a *saffron* Mims ENDOR simulation. The ENDOR spectra calculated for all the rotamers were then summed and overlayed with the experimental data for comparison.

3.1 Ring orientation determined from T_{read} values

First, the resolved ENDOR splittings (T_{read}) were visually determined from the spectrum to establish the relative ordering of the ^{19}F–Gd^{3+} distances for each of the p-F, m-F$_2$ and o-F$_2$ pcAAs. The relative ordering gives an indication of the likely ϕ and θ ring orientation, as outlined in Figure 2 of the main text and in Figure S3.1a below. In the case of the C_{D9k}-50 protein labelled with p-F, m-F$_2$ and o-F$_2$, the distances determined from the T_{read} splittings were $r_p = 7.1 \pm 0.3$ Å, $r_m = 7.4 \pm 0.3$ Å, and $r_o = 8.9 \pm 0.6$ Å, respectively. The relationship $r_p < r_m < r_o$ corresponds to the (ϕ, θ) combination in the red region of Figure S3.1a. At the extremes of the uncertainty range ($r_p = 7.4$ Å $> r_m = 7.1$ Å), however, the relative ordering of the distances could become $r_m < r_p < r_o$, corresponding to the green region of Figure S3.1a.

For the C_{D9k}-66 sample, only the o-F$_2$ splitting was resolved which indicates that the r_m and r_p distances are definitively longer than r_o. According to the calculation in Figure S3.1a, this means the phenyl ring must be tilted away from the Gd^{3+} ion (blue region), which corresponds to ring orientation parameters $90^\circ \leq \theta \leq 180^\circ$, and $0^\circ \leq \phi \leq 30^\circ$.

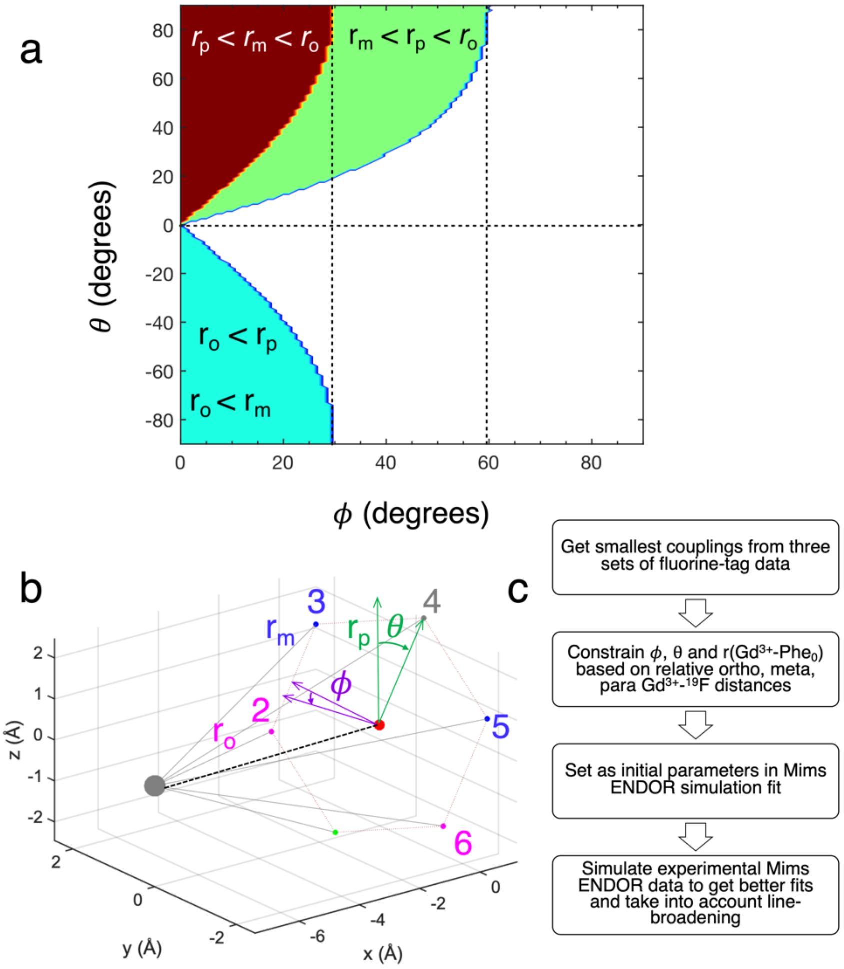

Figure S3.1 Outline of the simulation procedure used to determine the Phe ring orientation. (a) The relative ordering of the T_{read} and corresponding r_{read} values of the p-F, m-F$_2$ and o-F$_2$-labelled protein data can be used to determine the possible range of θ and ϕ angles of the phenyl ring. (b) The specific θ and ϕ angle values as well as the distance from Gd^{3+} to the ring center r_c can be further narrowed down by fitting the r_{read} values in MATLAB using nonlinear optimisation (*fminsearch*). The ^{19}F positions around the phenyl ring are simulated as shown in the plot, where the ring geometry is fixed with its centre at (0,0,0), while the Gd^{3+} position is varied along the x axis to define the distance vector r_c. The Euler rotation matrix R = erot[ϕ , θ, 0] is then applied to the ^{19}F coordinates around the ring to vary the individual ^{19}F-Gd^{3+} distances r_F until they fit the r_{read} values. (c) Flowchart outlining the overall analysis protocol to perform the simultaneous ENDOR lineshape fits with informed initial guess values for the fitted r_c, θ and ϕ parameters.

3.2 Fitting Mims ENDOR data in Easyspin – additional remarks

As described above, *EasySpin* simulations using the *saffron* function were performed on the data sets of all three p-F, m-F$_2$ and o-F$_2$-labelled variants of one protein simultaneously to determine a common r_c, θ and ϕ parameter set. Initial guess values for these parameters are used as starting points for the *fminsearch* MATLAB function. *fminsearch* then varies these parameters until the objective error function is minimised. A fit was accepted as best fit within a tolerance of 10^{-6}. Without an informed initial guess, however, we found that the optimisation algorithm easily got stuck in a local minimum that was often, visually, a poor fit to the experimental data. This was particularly the case for variations of the angles θ and ϕ, where changes in the first digit after the decimal point could have a large effect on the simulated lineshape.

Consequently, we used an informed approach to the simultaneous parameter fit, where the initial guess values for r_c, θ and ϕ were restricted by simulating the phenyl ring in MATLAB based on the r_{read} values. First, a phenyl ring centred at [0,0,0] was constructed, which determined the five ^{19}F positions. Next, a Gd^{3+} atom was positioned at a distance [-r_c, 0, 0] from the ring. Applying an Euler angle rotation (*erot* in EasySpin) to the phenyl ring coordinates then defines the [ϕ, $\theta, 0$] rotation of the ring relative to the Gd^{3+} ion. The three parameters r_c, ϕ and θ were varied in a nonlinear optimisation algorithm (*fminsearch*) until the differences between the corresponding simulated r_F values and the r_{read} values are minimised. Figure S3.1b illustrates this process when plotted in MATLAB. As outlined in the flowchart in Figure S3.1c, the resulting best-fit r_c, ϕ and θ parameters where then used as starting points for the ENDOR spectral simulations.

To estimate the uncertainty in the final best-fit parameter set (i.e. r_c = 8.0 ± 0.4 Å, θ = 23° ± 10°, and ϕ = 17° ± 17° for C_{D9k}-50), we tested different initial guess values for the ENDOR spectral simulations. The range of guess values was sampled by taking all combinations of the T_{read} values r_p =7.1 ± 0.3 Å, r_m= 7.4 ± 0.3 Å, and r_o = 8.9 ± 0.6 Å i.e. at either +, – or at the centre of their respective uncertainties ranges (giving 27 different possible starting points). Each starting point resulted in a final simulation that converged to a slightly different {r_c, θ, ϕ} best fit. The final uncertainty ranges represent the total span of each parameter in the collection of these converged results.

3.3 τ-varied Mims ENDOR experimental data and simulations

The Mims ENDOR pulse sequence is subject to blindspotting artefacts and therefore must be acquired with different values of the interpulse delay τ and summed in order to accurately reflect the real ENDOR lineshape. The main text (Figures 4 and 5) shows the summed spectra. The individual τ-dependent ^{19}F ENDOR lineshapes and the best-fit simulations performed in *EasySpin* for each are shown in Figure S3.2

Table S3.1 Summary of the distances and phenyl ring parameters determined from ^{19}F ENDOR lineshape analysis.

Protein	r_{read} **(Å)**	r_c **(Å)**	r_F **(Å)**	θ, ϕ	lw **(kHz)**
p-F C_{D9k}-50	7.1 ± 0.3	8.0	7.3	23°, 12°	41
m-F_2 C_{D9k}-50	7.4 ± 0.3		7.3, 8.4		
o-F_2 C_{D9k}-50	8.9 ± 0.6		8.5, 9.4		
p-F C_{D9k}-66	min.13.6	11.7	10.1, 11.2	-97°, 12°	40
o-F_2 C_{D9k}-66	10.2 +0.9/-0.6		14.5		

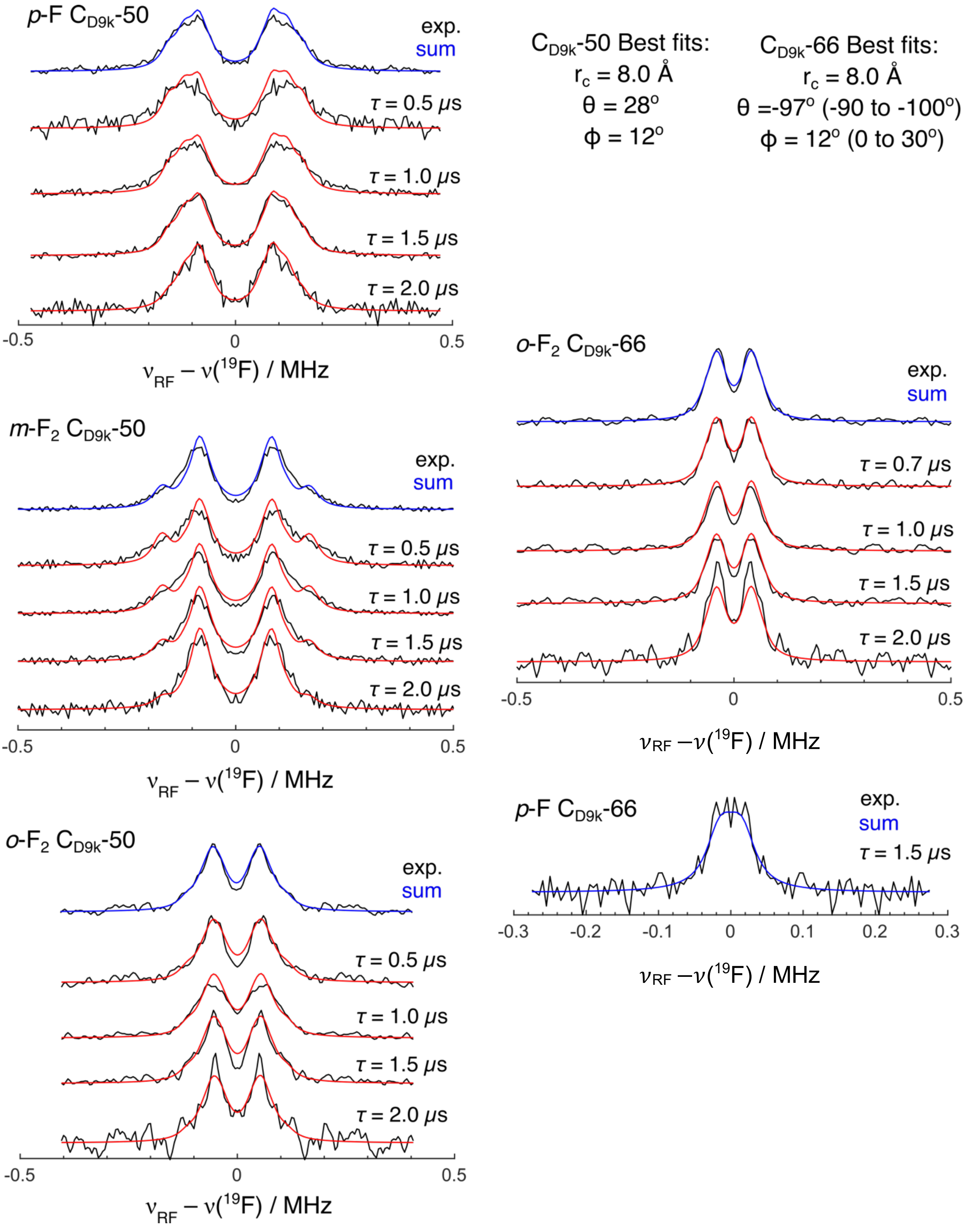

Figure S3.2. Processed and experimental ^{19}F Mims ENDOR data for all five C$_{D9k}$ proteins measured at different τ values. The experimental data (black traces) are shown symmetrised. The *Easyspin* saffron calculations are shown in red. The summed spectra presented in Figures 4 and 5 of the main text are shown in the top of each panel, along with the blue summed simulation trace. These fits are for the best-fit parameters of the protocol outlined in Figure S3.1.

SI 4. ENDOR fits of the C_{D9k} proteins using rotamer simulations

4.1 Rotamer definitions *t, g-* and *g+*

Throughout the main text we use the definitions *t* (trans), *g-* (gauche minus) and *g+* (gauche plus) to refer to the rotamer conformations in terms of their χ_1 angles. This terminology is adapted from ref.[10] In brief, the χ_1 angle is assigned to the dihedral angle defined by the atoms N-C^{α}-C^{β}-C^{γ}. The plots in Figure S4.1 provide Newman projections looking from the C^{α} to the C^{β} atom. χ_1 changes the position of the phenyl ring relative to the backbone atoms. In the *t* rotamers ($\chi_1 = 180^o$) the Phe ring is opposite the H atom of the C^{α}. In the *g-* rotamers ($\chi_1 = -60^o$) the phenyl ring is opposite the carbonyl carbon, in the *g+* rotamers ($\chi_1 = +60^o$) the phenyl ring is opposite the N atom.

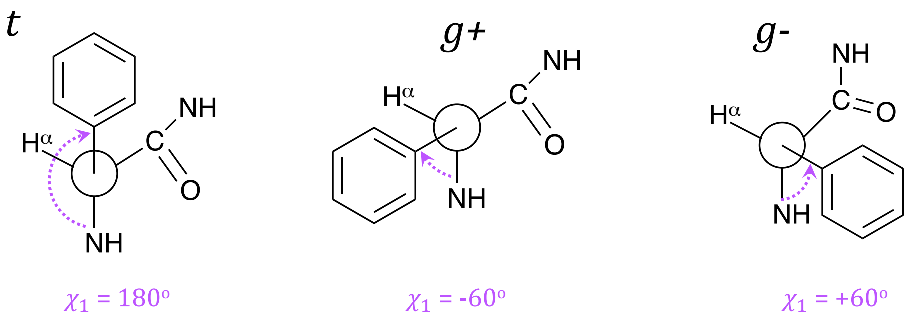

Figure S4.1. Definitions of the *t*, *g-* and *g+* rotamers. Figure adapted from ref.[11]

4.2 Complete unconstrained rotamer simulations of Phe in $\mathrm{C_{D9k}}$-50 and $\mathrm{C_{D9k}}$-66

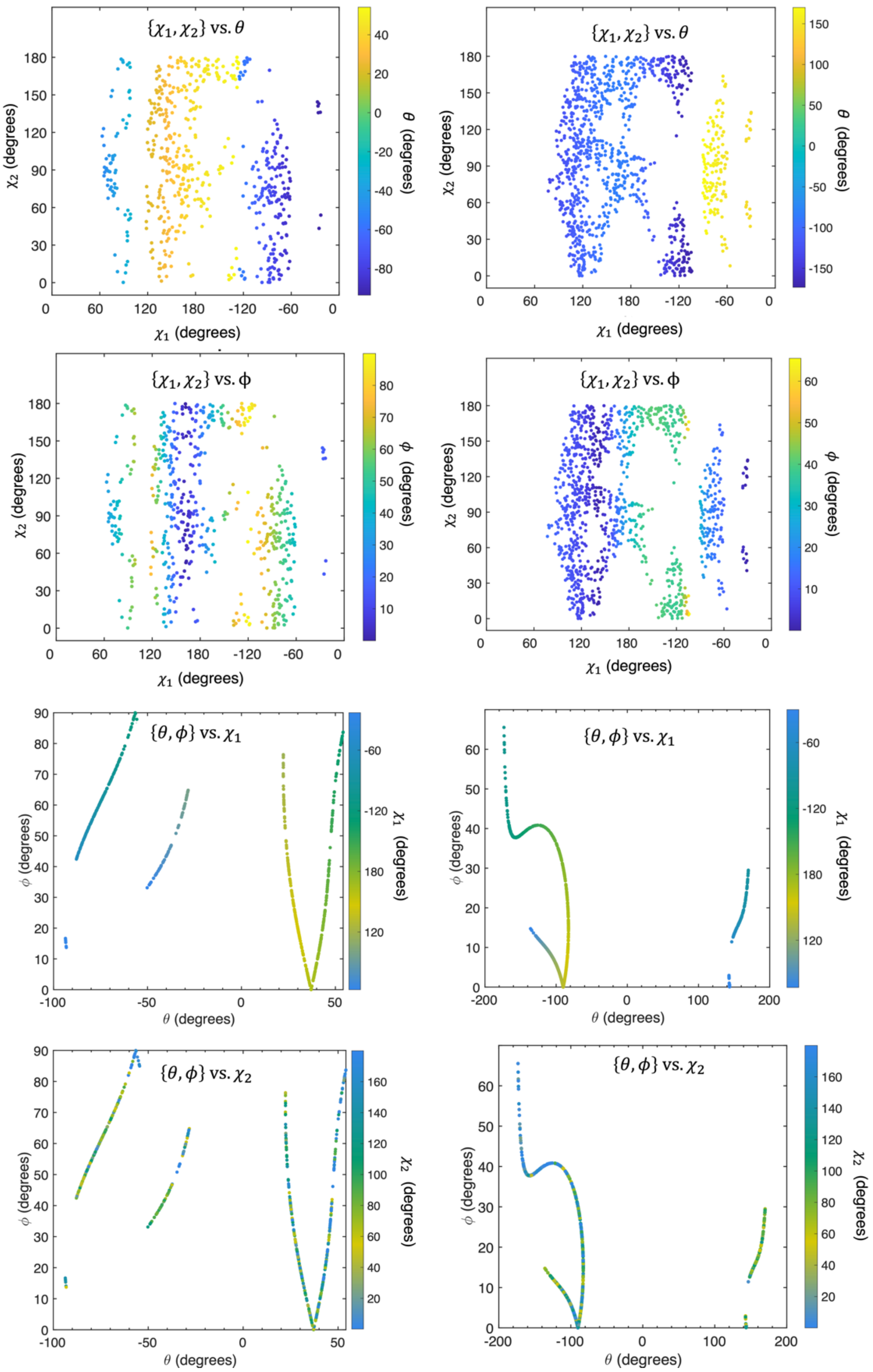

Figure S4.2 Complete overview of the phenyl ring parameters and phenylalanine χ angles resulting from unconstrained rotamer simulations. The rotamer simulations above were performed for the unfluorinated Phe residue at either position 50 in $\mathrm{C_{D9k}}$ (left hand side panel) or position 66 (right hand side panel). In the main text we plot the different ring and side chain angles against the $\mathrm{Gd^{3+}}$–ring center distance r_c. In this figure plotting the rotamers in terms of different combinations of χ_1, χ_2 and θ and ϕ further helps in visualising the relationship between the ENDOR ring parametrisation and conventional $\{\chi_1, \chi_2\}$ parameterisation.

4.3 ENDOR simulations using unfiltered vs. filtered rotamer libraries.

The PDB file generated in PyParaTools[8] containing the rotamers simulated for each C_{D9k} mutant was entered into MATLAB to generate a list of the ^{19}F–Gd^{3+} distances. These distance lists were then used to simulate the corresponding ENDOR spectra in *EasySpin* using the *saffron* function for Mims ENDOR. We explored three staged approaches to fitting the ENDOR data from the rotamers: (i) the whole, unconstrained rotamer library (all χ_1 and χ_2 angles were allowed except for eclipsed orientations of χ_1 and angle combinations generating steric clashes of the residue with the protein) shown in red in Figure S4.3, (ii) only the rotamers in a χ_1 subpopulation with similar ring-orientation parameters to the ones identified from ENDOR modelling, shown in blue in Figure S4.3. In both C_{D9k}-50 and C_{D9k}-66, this happened to be the *t* subpopulation, and (iii) using the best-fit ENDOR parameters explicitly to further filtering the dominant conformer subset.

Neither of the approaches (i) or (ii) delivered a particularly good fit to the experimental data, especially for C_{D9k}-50. As shown in the main text, filtering of the rotamers as per approach (iii) i.e. based on the ^{19}F ENDOR r_F distance cut-offs, was needed to better approximate the measured lineshape. In the case C_{D9k}-66, the fits using only $\chi_1 = t$ more closely reproduced the experimental spectra. However, this is in part due to the reduced ENDOR sensitivity and resolution of the longer ^{19}F–Gd^{3+} distances in the C_{D9k}-66 proteins.

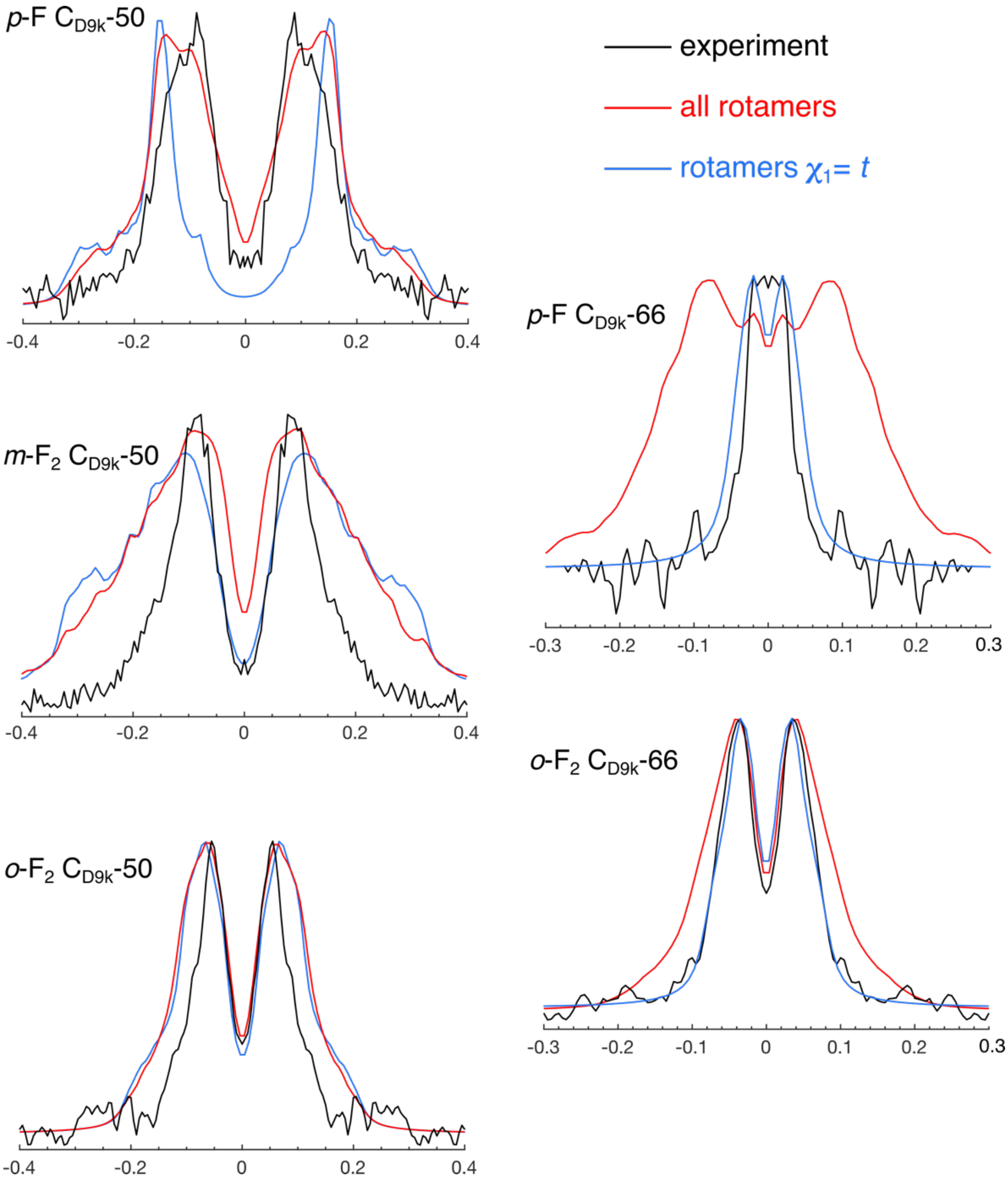

Figure S4.3. ENDOR simulations of the $\mathrm{C_{D9k}}$ mutants with *p*-F, *m*-$\mathrm{F_2}$ or *o*-$\mathrm{F_2}$ using the rotamer libraries generated in PyParaTools as input into *EasySpin* Mims ENDOR calculations. The symmetrised, τ-summed experimental data are shown in black. The red traces correspond to simulations using the whole unconstrained rotamer library, while the blue traces depict the ENDOR simulations using only the *t* rotamer subpopulations.

4.4 ENDOR simulations of $\mathrm{C_{D9k}}$-50 proteins using unfiltered rotamer libraries

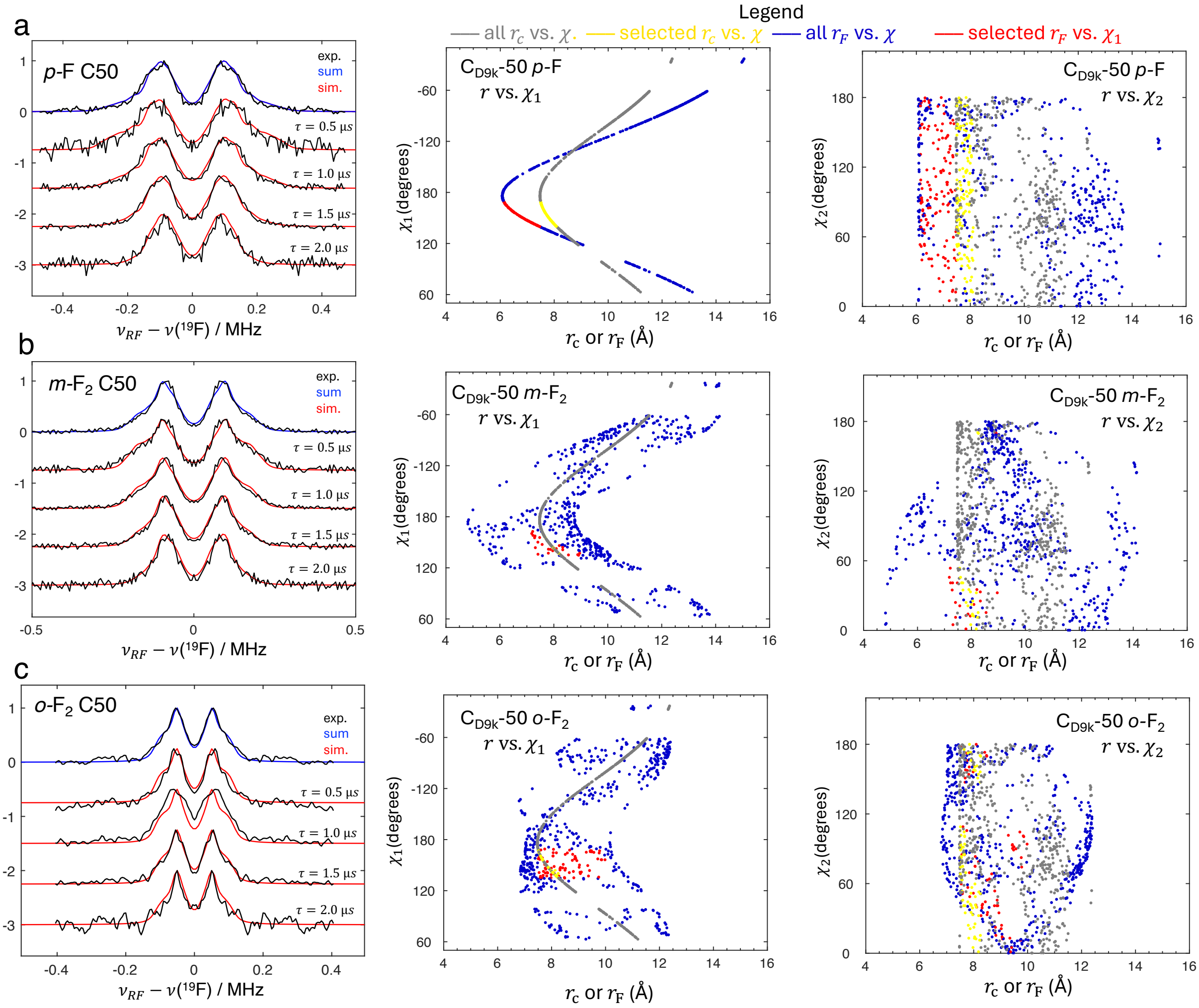

Figure S4.4. Rotamer selection for the $\mathrm{C_{D9k}}$-50 data and simulations using the restraints determined from ^{19}F ENDOR modelling ($r_c = 8.0$ Å, $\theta > 0^o$) with the minimum r_F distance cutoffs of 6.3 Å for *p*–F, 7.2 Å for *m*–F$_2$ and 8.4 Å for *o*–F$_2$. Including distances below these cutoffs in the simulation would result in significant deviations of the simulated lineshape (red) from the experimental spectrum (black), shown in the left-hand side panels. The rotamer calculations are shown in the right hand side panel. The full rotamer simulations are indicated in grey (r_c distances) and blue (r_F distances). The selected simulations are indicated in yellow (r_c distances) and red (r_F distances).

4.5 ENDOR simulations of C_{D9k}-66 proteins using unfiltered rotamer libraries

We applied the same filtering approach to the C_{D9k}-66 mutant made with o-F_2. The minimum r_c distance was set to the r_{read} distance of 10.2 Å, and the θ angle was constrained to <0°. The best approximation of the experimental lineshape was obtained when the minimal limit of the r_F distances was set to 10 Å. The conformations selected in this way correspond to the subset of *t* conformations, with a χ_1 angle of 150-180°, highlighted in red/yellow in Figure S4.5.

Additionally, in the main text, for simplicity, we show constrained rotamer libraries where the χ_1 angle is varied $\pm 10^{\circ}$ about the staggered rotamer conformations (see Figure S4.1). i.e. $-70^{o} < \chi_1 < -50^{o}, 170 < \chi_1 < 180^{o}$ and $50 < \chi_1 < -70^{o}$. In figure S4.5 we show the r_c/r_F vs.χ_1 and χ_2 plots for both the constrained χ_1 rotamer libraries and the unconstrained rotamer libraires $(-180^{o} < \chi_1 < 180^{o})$. We see that the unconstrained libraries produce a lot of (unrealistic) conformations close to $\chi_1 = 120^{o}$, which is why we opted to show the simpler, constrained library in the main text. Regardless of the $\chi_1$constraints though, the *t* conformers with χ_1 around 180° is observed.

For the C_{D9k}-66 mutant made with p-F, the lack of a resolved splitting meant that further filtering of the rotamer subpopulations within the *t* conformations added no significant restraints and therefore was not pursued.

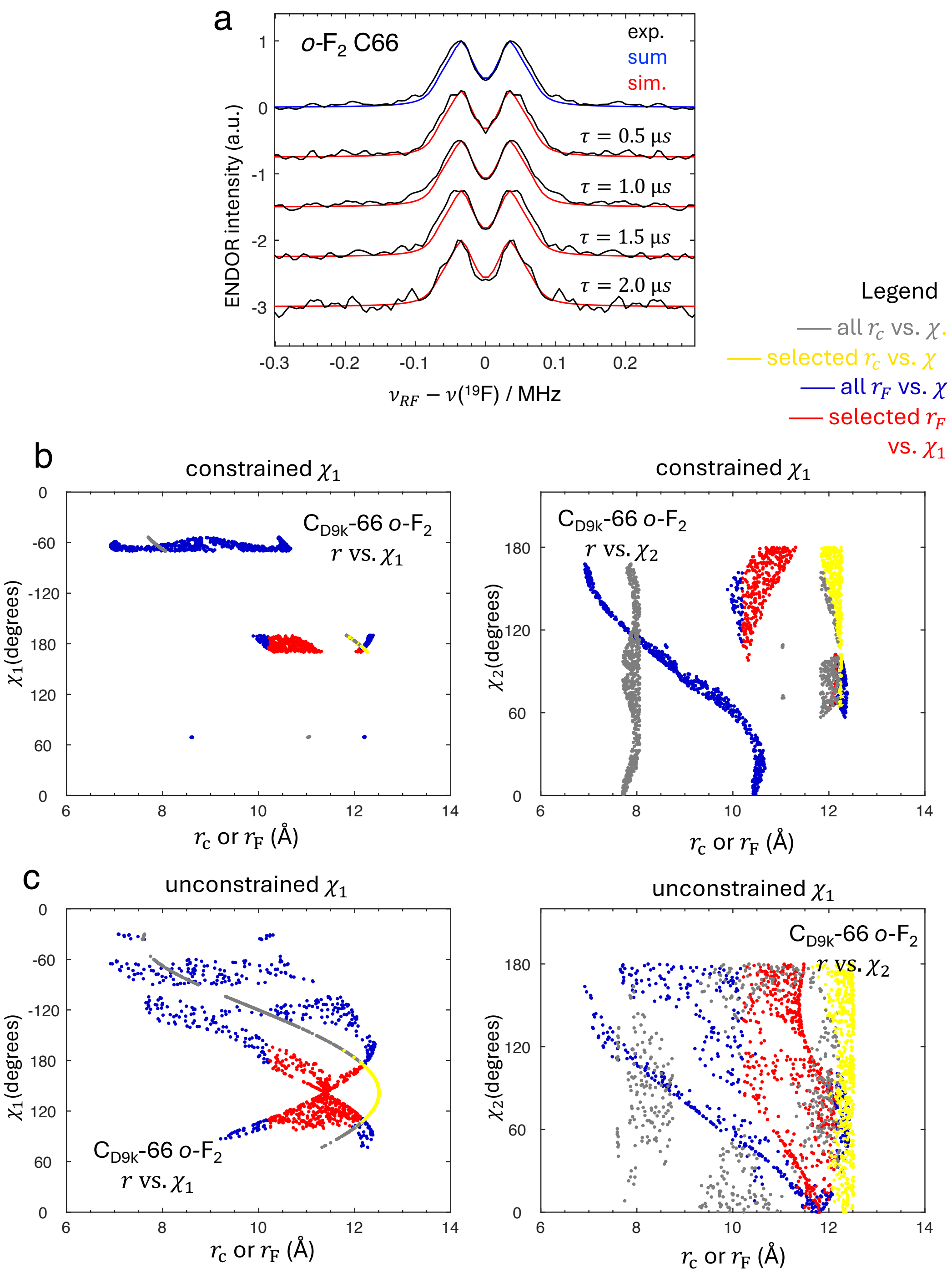

Figure S4.5. Rotamer selection for the C$_{D9k}$-66 mutant made with *o*-F$_2$, using the restraints determined from ^{19}F ENDOR modelling ($r_c \approx 10.2$ Å, $\theta < 0^{o}$) with a minimum r_F distance cutoff of 10.0 Å. Including distances below these cutoffs in the simulation would result significant deviation of the simulated lineshape (red) from the experimental spectrum (black), shown in the top panel a. The middle panel b indicates the constrained rotamer library with the χ_1 angle variation restricted to $-70^{o} < \chi_1 < -50^{o}, 170 < \chi_1 < 180^{o}$ and $50 < \chi_1 < -70^{o}$. The bottom panel c shows the unconstrained rotamer simulations with $-180^{o} < \chi_1 < 180^{o}$ The full rotamer simulations are indicated in grey (r_c distances) and blue (r_F distances). The filtered simulations are indicated in yellow (r_c distances) and red (r_F distances).

SI 5. Comparing C_{D9k}-50 ENDOR restraints to the X-ray crystal structure

The main text compares the phenyl ring conformations in C_{D9k}-50 and C_{D9k}-66 obtained with the ENDOR restraints, with the high-resolution (1.6 Å) crystal structure 4ICB.[12] For C_{D9k}-50 the crystal structure parameters for the phenyl ring are $r_c = 7.6$ Å, $\theta = 47^{\circ}$, $\phi = 28^{\circ}$. These values fall approximately within the uncertainty range of the single parameters determined from ENDOR simulations. However, when we enter just the crystal-structure values (i.e. single parameter set), into a Mims ENDOR simulation, the calculated spectra poorly fitted the experimental data particularly for the protein made with m-F_2 (Figure S5.1), presumably pointing to a degree of flexibility in the real system that is not reflected in the crystal structure.

In the crystal structure, the average B-factor of the phenyl ring is about 15–16 $Å^2$, which corresponds to a mean displacement of each atom by about 0.4–0.5 Å. This structure also predicts an r_c distance between the Gd^{3+} and the center of the Phe50 ring of approximately 7.6 Å, which is in close agreement to the distance we obtained previously in the CF_3-Phe labelled C_{D9k}-50 protein in ref.[3] There, we fitted a value of 6.9 Å as the distance between the Gd^{3+} ion and the CF_3 group, and the tilt angle of the CF_3 group relative to the main ^{19}F–Gd^{3+} vector was approximately 90°.

Although the crystal structure does not fit the ENDOR data very well, the variation in the fitted parameters between the crystallographic and ENDOR data indicates acceptable agreement within the experimental resolution limits. Additionally, our ENDOR modelling assumes that the Gd^{3+} ion position is rigid and only the phenyl ring orientation varies. However, the metal-binding loop may also vary in conformation to some degree. The distribution of r_c distances contributing to the ^{19}F ENDOR data therefore contains a “background” from the global flexibility of the protein structure.

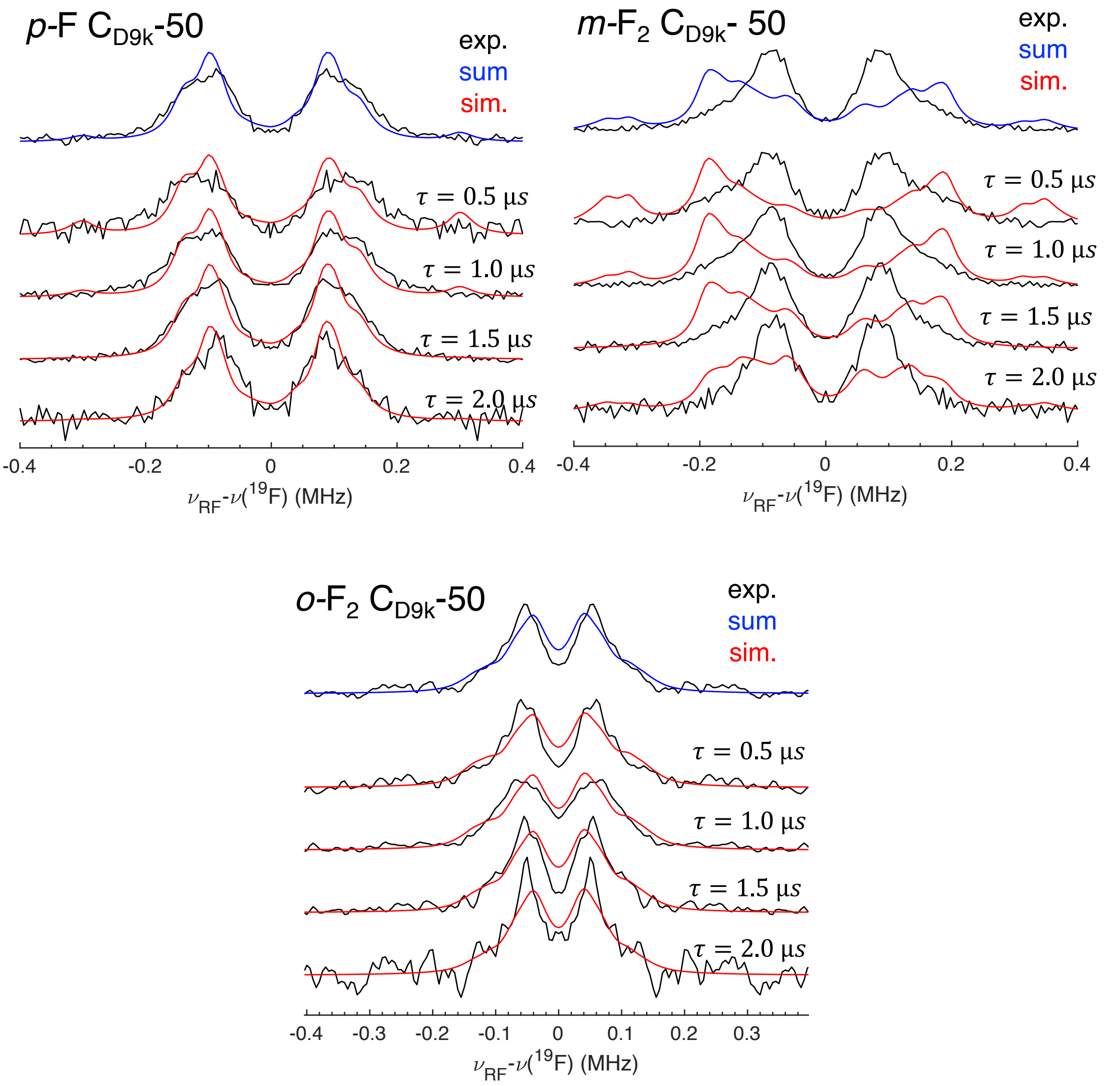

Figure S5.1 ^{19}F ENDOR fits of the C_{D9k}-50 mutants made with *p*-F, *m*-F_2 and *o*-F_2, using the phenyl ring conformation from the crystal structure 4ICB ($r_c = 7.6$ Å, $\theta = 47^o$, $\phi = 28^o$).

SI 6. Rotamer simulations and ENDOR fits of the GB1 proteins

6.1 Rotamer simulations of the GB1 proteins

The GB1 proteins **2a-e** labelled with the **Gd.C1** tag at position 28 and varied CF_3Phe positions were also analysed via rotamer simulations. These were performed using the crystal structure (PDB: 1PGB)[13] and the program PyParaTools[8] as described previously.[3] To recap, in brief, the dihedral angles annotated in Fig. S6.1 were varied in random combinations. Conformations that generated steric clashes, i.e., distances <0.9 times the sum of the van der Waals radii of a pair of two atoms, were discarded.

Heavy atom representations of the **Gd.C1** tag and CF_3Phe residue are shown in Fig. S6.1. For the **Gd.C1** tag, the dihedral angle χ_1 was varied by ±30° around the staggered rotamers, χ_2 was varied by ±30° around the 90° and -90° values, while the angles χ_3– χ_6 were allowed to vary freely. For the CF_3 group, the dihedral angle was varied by ±10° about the staggered rotamers, while the angle χ_2 was allowed to vary freely.

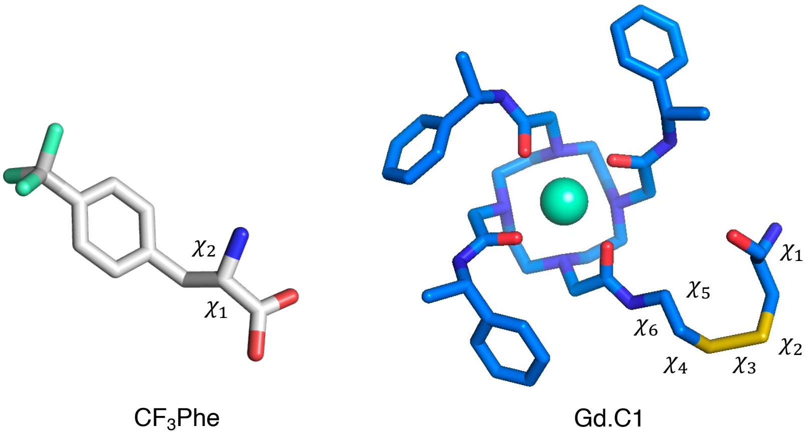

Figure S6.1 Heavy atom representations of the CF_3Phe amino acid and the **Gd.C1** tag annotated with the dihedral angles varied to generate rotamer libraries using the program PyParaTools. The rotamer libraries were established by modeling the **Gd.C1** tag and CF_3Phe side chain onto the crystal structure of GB1 (1PGB). The Gd^{3+} ion is indicated by a teal sphere, the ^{19}F atoms are indicated in cyan and the varied dihedral angles are annotated on the structures.

6.2 Lineshape Model Parameters of the GB1 proteins

The lineshape model of the **Gd.C1** spin-labelled GB1 proteins was carried out as described previously in ref.[3,4] In brief, we assume a Gaussian ^{19}F-Gd^{3+} distance distribution centred about a mean distance $\bar{r}_{\mathrm{ls}}$ with a Gaussian width δ_{ls}. Distance distributions with guess $\bar{r}_{\mathrm{ls}}$ and δ_{ls} values as well as a guess ENDOR spectral linewidth *lw* (Lorenztian) are used as inputs into *EasySpin saffron* Mims ENDOR calculations to produce fits to the experimental ^{19}F ENDOR spectra. The best-fit parameter sets that minimise the residuals between simulation and experiment are found using the MATLAB least-squares minimisation algorithm *fminsearch.* In Figure S6.2 we reproduce the simulated fits to the experimental data from the main text figure 5 but show (i) the unsmoothed, unsymmetrised data overlayed with the smoothed, symmetrised spectra for **2b-e**, (ii) the full τ-dependent set of measurements collected for **2a** (where because of the larger dipolar coupling blindspotting of the ENDOR lineshape has to be taken into account) and the individual fits for each different τ spectrum.

Table S6.1 is partially reproduced from ref.[4] and details the best-fit lineshape model parameters found for each protein **2a-e**. The error bounds were determined from the 95% confidence intervals using covariance analysis as described previously in the Supporting Information of ref.[4] We also include the distances determined from the relative ^{19}F ENDOR intensities scaled to the **2d** protein as the reference intensity as described previously.[1] Note that a distance cannot be reliably predicted for **2b** from the intensity reliably since, as outlined in SI 2, we predict two distinct populations – intact and misfolded protein, are present in this sample.

Table S4.1 ^{19}F–Gd^{3+} distances in proteins **2** predicted by from (i) the lineshape model, and (ii) $1/r^6$ dependence of the relative integrated signal intensities referenced to the intensity of **1d** (16 Å).

Protein	Lineshape Model			Integrated Intensity $\bar{r}_{\mathrm{int}}$ (Å)
	$\bar{r}_{\mathrm{ls}}$ (Å)	δ_{ls} (Å)	lw (kHz)	
2a	8.2 ± 0.1	1.2 ± 0.1	24 ± 3	8.4 +0.4/-0.4
2b	10.9 ± 0.5	2.1 ± 0.3	35 ± 5	12.8 +0.7/-0.6*
2c	15.5 ± 0.1	1.8 ± 0.7	38 ± 6	15.3 +0.4/-0.3
2d	16.0 ± 0.5	2.2 ± 0.3	32 ± 6	**ref**^{b}. +0.7/-0.6
2e	18.5 ± 0.9	2.2 ± 0.8	25 ± 3	18.5 +1.0/-1.1

*The **2b** distance predicted from intensity is an estimate - as outlined in SI 2, **2b** the EPR lineshape of this sample suggests that there are two distinct populations present in the sample – the target, intact protein, and an aggregated population: if not all of the Gd^{3+} tag is seeing the "real" CF$_3$Phe conformation due to disorder, then the integrated ^{19}F ENDOR will be lower than the maximum achievable response. The fact that the distances estimated from the lineshape model and from the intensity differ so vastly, does suggest that only a fraction of the sample is the target **2b** conformation.

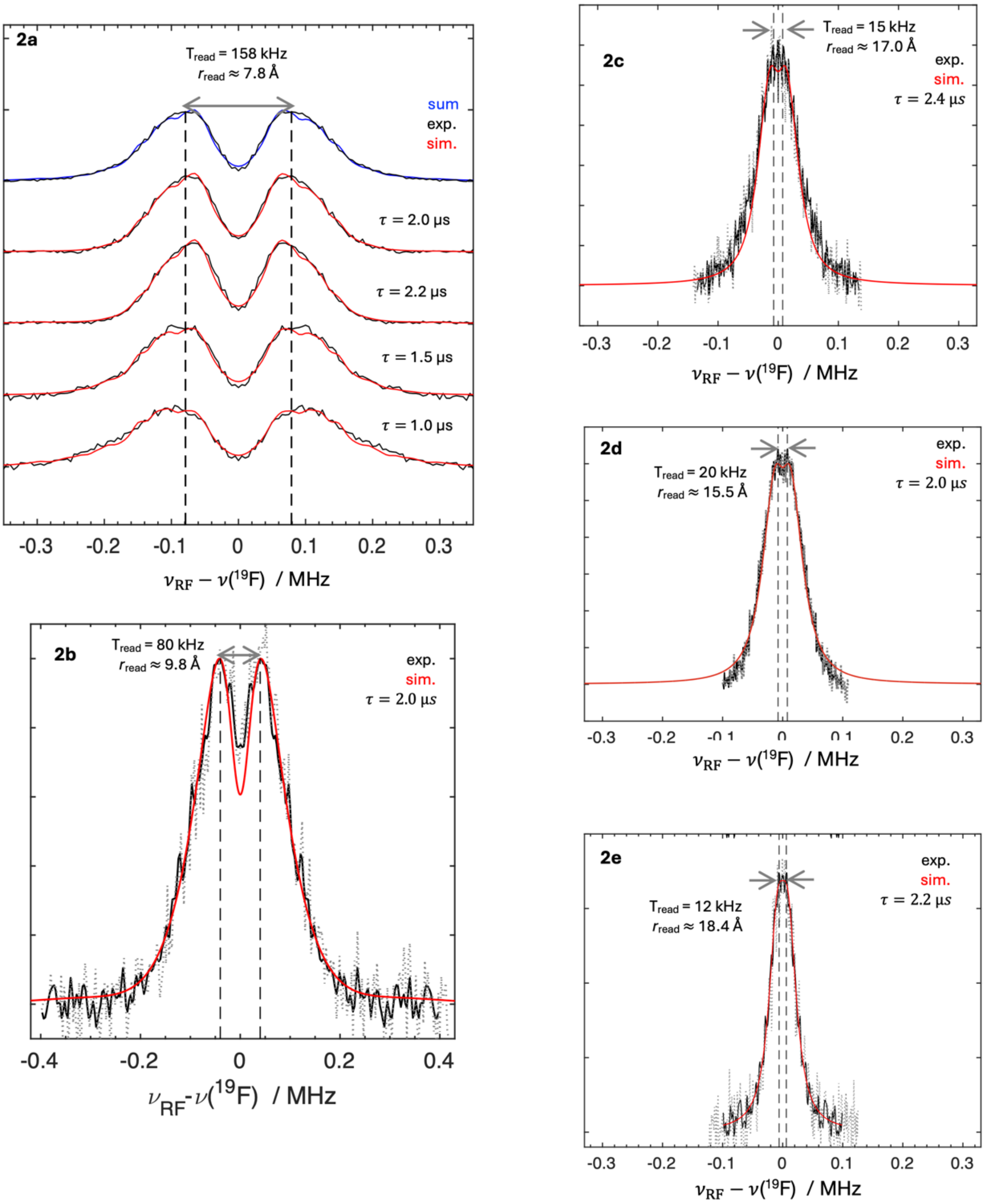

Figure S6.2 ^{19}F ENDOR experimental data for GB1 protein samples **2a-e**, showing the raw, unsymemetrised, unsmoothed data (grey dotted traces) overlayed with the smoothed, symmetrised data (black traces). The simulated spectra are shown in red. For **2a** we show the individual spectra measured at different interpulse delay values τ. The τ value used to measure each spectrum **2b-e** is annotated in each figure.

6.3 A closer look at the ENDOR-filtered rotamers vs. all rotamers

In figures S6.3-6.7 we take a closer look at the rotamers identified by filtering the full library according to the ENDOR fits (see Figure 5 of main text). In these figures the **Gd.C1** tag at position 28 is shown in purple, the preferred CF$_3$Phe conformation identified from ENDOR-based filtering is shown in pink and the other rotamers shown in grey – the χ_1 conformations *g*-, *g*+ or *t* are annotated beside each rotamer. The native conformation of the substituted residue is shown in green, the hydrophobic residues (Ala, Gly, Cal, Ile, Leu, Phe or Me) are shown in orange.

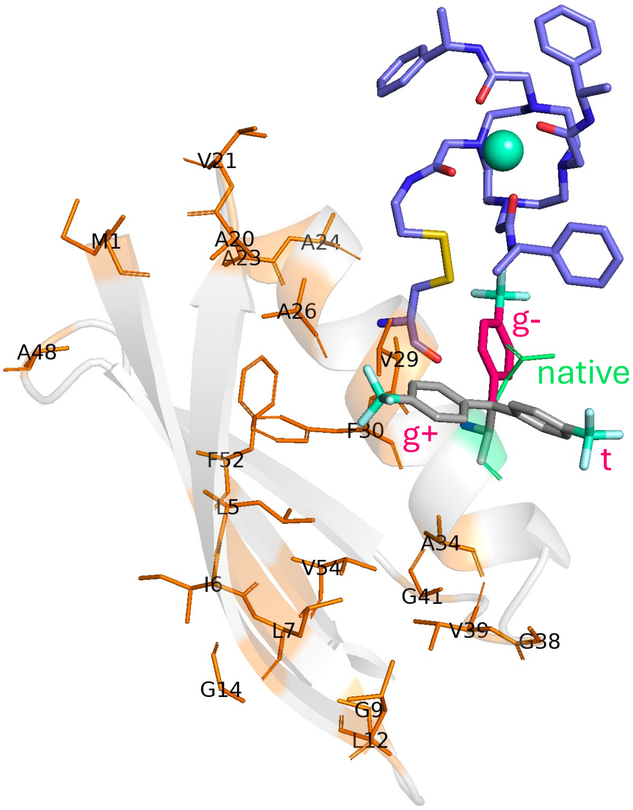

Fig. S6.3 Rotamers of GB1 construct **2a.** *g*+ would minimises hydrophobic Phe ring exposure to the solvent but does not preserve native χ_1 angle and produces strong clash with neighbouring V29. *g*-: Phe ring is somewhat exposed to solvent, but no clashes with protein residues and native χ_1 angle is preserved. However, clashed can occur between **Gd.C1** and the *g*- CF$_3$Phe which presumably modulates the conformational space sampled by both. *t*: Phe ring is fully solvent exposed, native χ_1 angle not preserved, but no clear proneness to clashes with other protein residues.

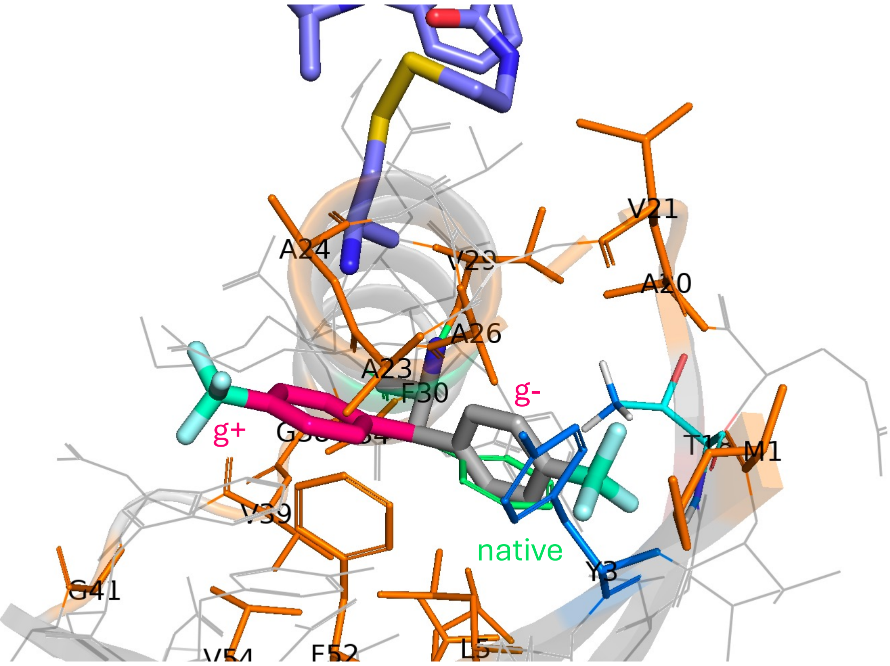

Fig S6.4 Rotamers of GB1 construct **2b.** Interestingly the *g*+ conformation identified as being preferrable from ENDOR-filtering of the rotamer library does not preserve the *g*- χ_1 of the native F30 residue. We think this is because the *g*- CF_3Phe results in very close proximity of the CF_3 group to the T18 methyl group or that there is not enough space for the additional bulk of the CF_3 group in this part of the hydrophobic core. This may result not only in preference for the *g*+ conformer but also in overall destabilisation of the protein structure, which could explain the much broader EPR lineshape of the Gd^{3+} tag for **2b** (Figure S2.1). The full rotamer simulations already eliminate the possibility of a *t* conformer due to strong clashes with the rest of the residues in the core (e.g. F52).

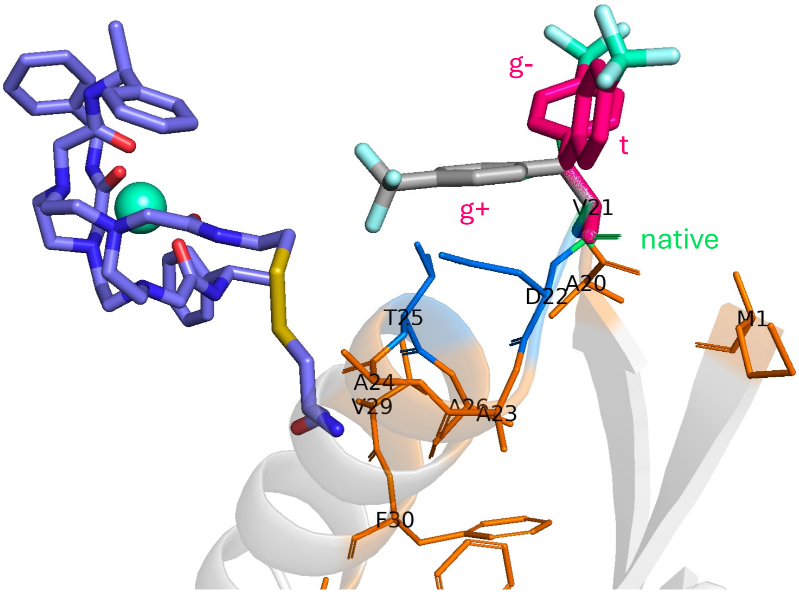

Fig S6.5 Rotamers of GB1 construct **2c.** The native residue V21 is on a flexible loop and completely solvent exposed. Therefore, there are no original hydrophobic interactions to preserve. The native crystal structure V21 rotamer is *g*-. Based on the ENDOR-filtered rotamer conformations it the *g*- and *t* rotamers of CF_3Phe appear to be equally likely to be sampled. The *g*+ rotamer is most likely not sampled: as shown in Figure 5 of the main text, *g*+ would produce short ^{19}F–Gd^{3+} distances which are clearly not sampled in the experimental ^{19}F ENDOR spectra. By examining the nearby GB1 residues we can also infer that the *g*+ CF_3Phe most likely also produces clashes with the T25 and D22 residues (highlighted in blue). Indeed, even in the full rotamer library *g*+ is significantly more sparsely sampled than the *g*- and *t* rotamers indicating that clash interactions are at play.

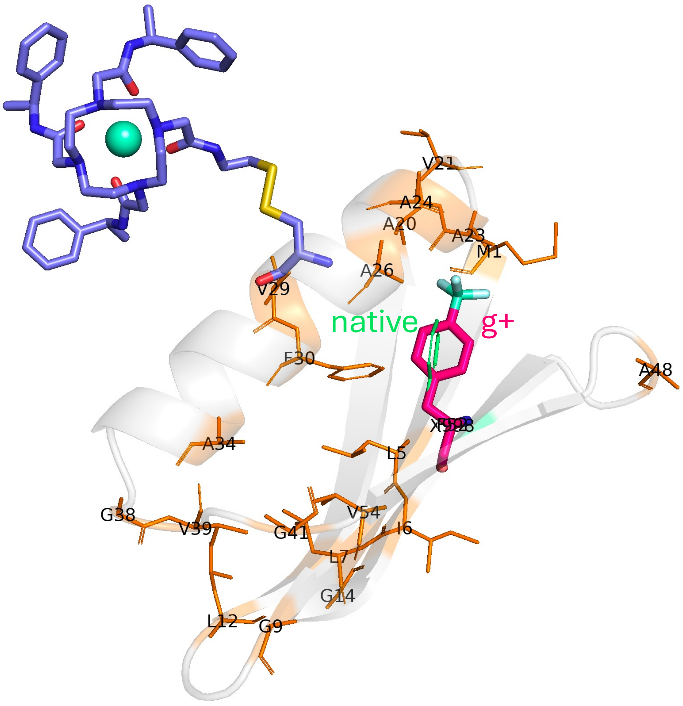

Fig S6.6 Rotamers of GB1 construct **2d.** In agreement with the native crystal structure, only the native *g*$^+$ conformation is sampled by the CF$_3$Phe due to steric clashes. Unlike **2b** which was the only other construct where the unnatural amino acid CF$_3$Phe substituted its native analogue Phe, in **2d** the native χ_1 angle is preserved, and the addition of the CF$_3$ group to the phenyl ring does not appear to perturb the structure of the protein core.

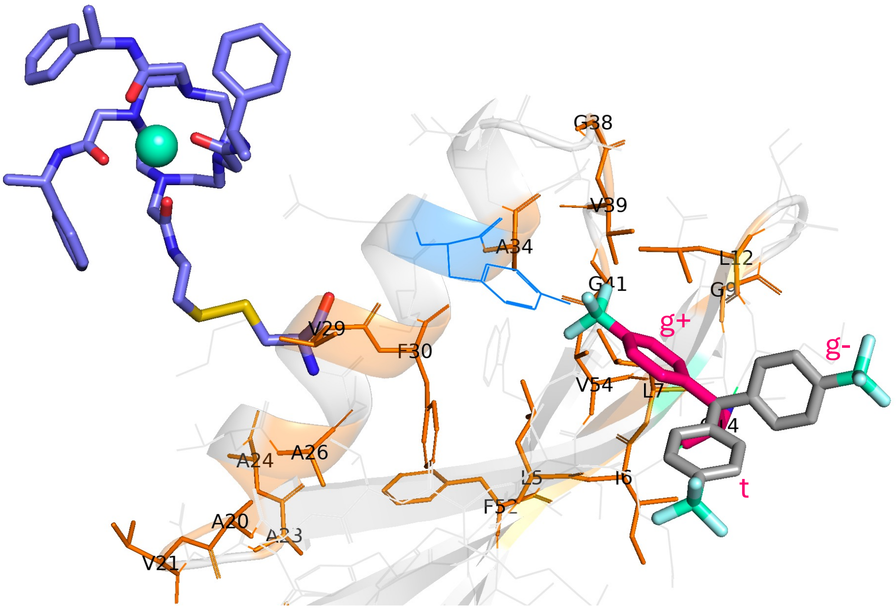

Fig S6.7 Rotamers of GB1 construct **2e.** The native glycine (G14) residue does not have a side chain and therefore there is no native χ_1 angle to preserve. In the ENDOR-filtered rotamer the *g*+ rotamers population must be sampled in order to reproduce the linewidth of the experimental ^{19}F ENDOR spectrum. The *g*+ rotamer positions the phenyl ring close to other hydrophobic residues in the protein core (highlighted orange), however it does result in the CF_3 group being in close proximity to the tyrosine Y33 residue (highlighted blue) which could lead to rearrangement of these residues due to clashes or stabilising interactions between the polar CF_3 and -OH groups of the two residues. It is unclear if *g*- or *t* are additionally sampled – as shown in Figure 5 of the main text their retention or exclusion from the rotamer library fit to the ENDOR data makes no difference to the fit's quality since they sample longer ^{19}F–Gd^{3+} distances than *g*+. As the ENDOR intensity drops off as $1/r^3$, these longer distances contributions are obscured by the larger signal intensity produced by the shorter ^{19}F–Gd^{3+} distances from *g*+.

6.4 Confirming that Gd.C1 χ_1 angles cannot be discerned from ENDOR fits.

As stated in the main text, we know from the fit of **2d** (where the CF_3Phe residue is mostly rigid in space and therefore represents fixed ^{19}F coordinates) that the **Gd.C1** conformation distribution is Gaussian, as the lineshape model and rotamer-predicted distributions agree closely with each other. Since the **Gd.C1** tagging position is identical across all five protein constructs, we do not expect conformation space of the Gd^{3+} tag to change from sample to sample. The exception is protein **2a** where our analysis (Figure 5 main text) indicates that the *g*+ rotamer is most likely dominant for the CF_3Phe residue. As these fluorine atoms also are much closer to the **Gd.C1** tag, the conformation space of the paramagnetic tag in this sample may be slightly more constrained than in the others to avoid clashes with the CF_3Phe residue. When we calculated a rotamer library that kept the *g*- rotamer fixed for the CF_3Phe residue while allowing full conformational freedom for the **Gd.C1** tag, the χ_1 angle of the **Gd.C1** tag (which should produce the largest change in the Gd^{3+} ion position) could assume only the *g*+ conformation, while the other tether conformations produced clashes with the CF_3Phe residue. Nevertheless, the ^{19}F-Gd^{3+} distance distribution for the CF_3Phe(*g*-)/**Gd.C1**(*g*-) combination is only slightly narrower than for the full **Gd.C1** conformational space (Figure 6.8).

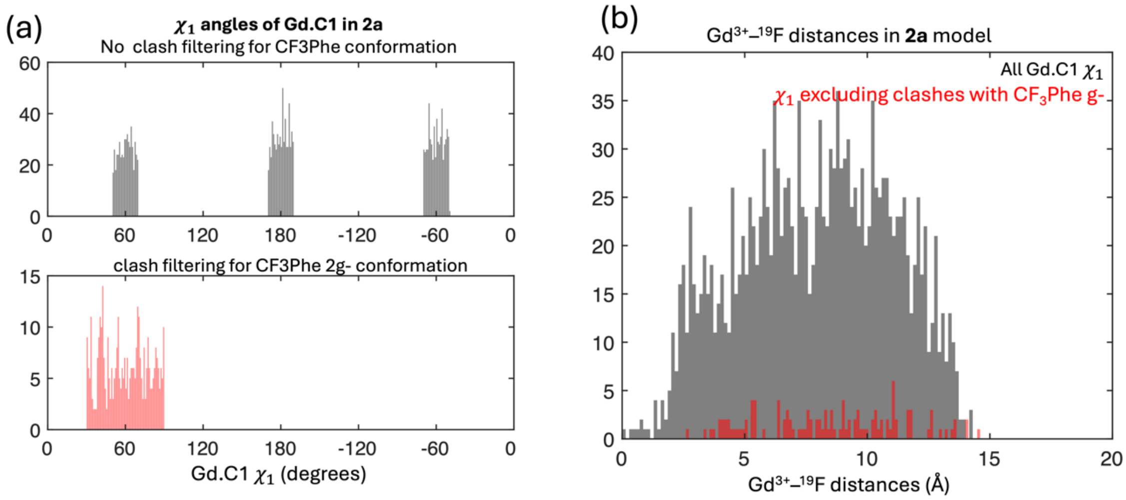

Figure S6.8 (a) Comparison of the χ_1 angles of the **Gd.C1** tag sampled for **2a** when clashes between the tag and the *g*- rotamer of the CF_3Phe residue are taken into account (red histogram in the lower panel) or not considered (grey histogram in the upper panel). (b) Corresponding ^{19}F-Gd^{3+} distance distributions when the *g*- rotamer is set for the CF_3Phe residue and either all rotamers (grey histogram) or only the *g*+ rotamers (red histogram) are allowed for the c_1 angle of the **Gd.C1** tag.

As a further check we confirmed that the conformations of the **Gd.C1** tag cannot be resolved in the same way as the conformations of the CF_3Phe residue by analysing the ^{19}F–Gd^{3+} distance distributions produced between all non-clashing rotamers of CF_3Phe and either the *g*+, *g*- or *t* rotamers of the **Gd.C1** tag. As seen in Figure S6.9 below, regardless of the χ_1 angle of the **Gd.C1** tag, the distance distributions remain broad and do not fit the lineshape models predicted for the GB1 constructs.

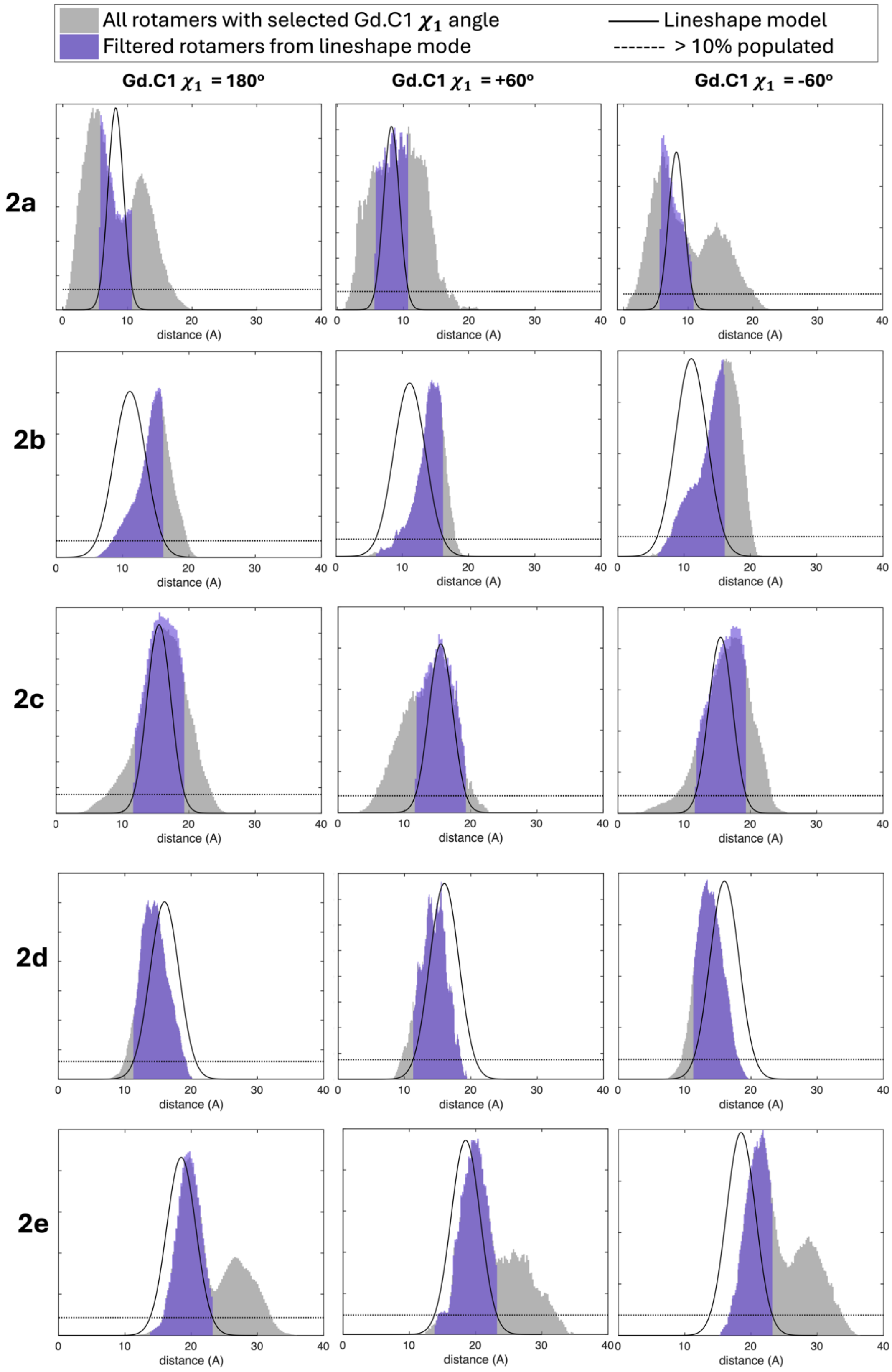

Figure S6.9. Conformational analysis of the CF3Phe residue in the GB1 constructs **2a** – **2e** in terms of the χ_1 angles of the **Gd.C1** tag calculated from rotamer libraries using constrained tag conformations. The χ_1 angle of the tag was set either to *g*+ ($30^o < \chi_1 < 90^o$), *g*- ($-90^o < \chi_1 < -30^o$) or *t* ($150^o < \chi_1 < 210^o$). The results are shown as grey histograms. The black traces indicate the ^{19}F–Gd^{3+} distances predicted by the lineshape model. The portions of the rotamer library that match the lineshape model distribution (cut-off set to 10% of the population maximum, indicated by the dashed line) are highlighted in purple.

References

(1) Qianzhu, H.; Tan, Y. J.; Abdelkader, E. H.; Huber, T.; Otting, G. Genetic encoding of fluorinated analogues of phenylalanine for ^{19}F NMR spectroscopy: detection of conformational heterogeneity in flaviviral NS2B-NS3 proteases. *ACS. Sens.* **2025**, *10*, 3152–3161. https://doi.org/10.1021/acssensors.5c00432.

(2) Nalepa, A.; Möbius, K.; Lubitz, W.; Savitsky, A. High-field ELDOR-detected NMR study of a nitroxide radical in disordered solids: towards characterization of heterogeneity of microenvironments in spin-labeled systems. *J. Magn. Reson.* **2014**, *242*, 203–213. https://doi.org/10.1016/j.jmr.2014.02.026.

(3) Judd, M. M.; Abdelkader, E. H.; Qi, M.; Harmer, J.; Huber, T.; Godt, A.; Savitsky, A.; Otting, G.; Cox, N. Short-range ENDOR distance measurements between Gd(III) and trifluoromethyl labels in proteins. *Phys. Chem. Chem. Phys.* **2022**, *24*, 25214–25226. https://doi.org/10.1039/D2CP02889A.

(4) Judd, M.; Qi, M.; Abdelkader, E. H.; Qianzhu, H.; Savitsky, A.; Huber, T.; Harmer, J. R.; Godt, A.; Otting, G.; Cox, N. Measuring nanometer distances in proteins and rigid rulers between ^{19}F and Gd^{3+} by integration of ^{19}F-ENDOR signal intensities. *J. Am. Chem. Soc.* **2025**, *147*, 16826–16835. https://doi.org/10.1021/jacs.4c13530.

(5) Mims, W. B. pulsed ENDOR experiments. *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.* **1965**, *283*, 452–457. https://doi.org/10.1098/rspa.1965.0034.

(6) Meyer, A.; Dechert, S.; Dey, S.; Höbartner, C.; Bennati, M. Measurement of Ångstrom to nanometer molecular distances with ^{19}F nuclear spins by EPR/ENDOR spectroscopy. *Angew. Chem. Int. Ed.* **2020**, *59*, 373–379. https://doi.org/10.1002/anie.201908584.

(7) Stoll, S.; Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **2006**, *178*, 42–55. https://doi.org/10.1016/j.jmr.2005.08.013.

(8) Stanton-Cook, M; Su, X.-C.; Otting, G.; Huber, T. PyParaTools. https://compbio.anu.edu.au/mscook/PPT.

(9) Yagi, H.; Banerjee, D.; Graham, B.; Huber, T.; Goldfarb, D.; Otting, G. Gadolinium tagging for high-precision measurements of 6 nm distances in protein assemblies by EPR. *J. Am. Chem. Soc.* **2011**, *133*, 10418–10421. https://doi.org/10.1021/ja204415w.

(10) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins Struct. Funct. Bioinforma.* **2000**, *40*, 389–408. https://doi.org/10.1002/1097-0134(20000815)40:3<389::AID-PROT50>3.0.CO;2-2.

(11) Arseniev, A.; Schultze, P.; Wörgötter, E.; Braun, W.; Wagner, G.; Vašák, M.; Kägi, J. H. R.; Wüthrich, K. Three-dimensional structure of rabbit liver [Cd7]metallothionein-2a in aqueous solution determined by nuclear magnetic resonance. *J. Mol. Biol.* **1988**, *201*, 637–657. https://doi.org/10.1016/0022-2836(88)90644-4.

(12) Svensson, L. A.; Thulin, E.; Forsén, S. Proline cis-trans isomers in Calbindin D$_{9k}$ observed by X-Ray crystallography. *J. Mol. Biol.* **1992**, *223*, 601–606. https://doi.org/ 10.1016/0022-2836(92)90976-Q.

(13) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **1994**, *33*, 4721–4729. https://doi.org/10.1021/bi00181a032.