Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information

Site Specific 2D IR Spectroscopy: A General Approach for the Characterization of Protein Dynamics with High Spatial and Temporal Resolution

Sashary Ramos, Rachel E. Horness, Jessica A. Collins, David Haak, and Megan C. Thielges

Corresponding Author: Megan C. Thielges
Email: thielges@indiana.edu

Experimental Methods

Expression and purification of *CN*F-labeled SH3^{Sho1}.

Expression and purification of SH3^{Sho1} proceeded as previously described.[1] The plasmid for expression of SH3^{Sho1} was kindly provided by the laboratory of Alan Davidson (University of Toronto).[2] The gene of interest is ligated into a pet21d+ vector (Novagen) between the Nco1 and Xho1 restriction sites such that protein is expressed with a C-terminal hexahistidine (His6) tag. Phusion Site-Directed Mutagenesis (Thermo Scientific) was used to incorporate a thrombin cleavage site immediately N-terminal to His6 such that the tag could be removed during purification. A TAG codon was introduced at the codons for each of the residues 2, 8, 10, 16, 20, or 54 via standard site-directed mutagenesis using a Stratagene Site-Directed Mutagenesis kit (Agilent). The pUltraCNF plasmid that encodes the orthogonal tRNA synthetase and tRNA for incorporation of *CN*F was generously provided by Peter Schultz (The Scripps Research Institute).[3]

The expression of each SH3^{Sho1} variant proceeded as previously described for the wild-type protein domain with minor modifications.[1] Briefly, plasmids containing the SH3^{Sho1} gene were co-transformed with pUltraCNF into BL21 (DE3) *E. coli*. A single colony was cultured in 4 mL of Luria-Bertani (LB) medium (100 μg/mL ampicillin, 60 μg/mL streptomycin) at 37 °C for 12 hours. Starter culture was used to inoculate 50 mL of LB supplemented with antibiotics at a 1:1000 ratio and allowed to grow overnight at 37 °C. Five mL of overnight culture were added to 1 L of Terrific Broth supplemented with antibiotics. Cells were grown to OD_{600} 0.1-0.2 at which time cultures were supplemented with 1 mM *CN*F (ChemPep Inc.). Expression was induced at OD_{600} 0.6-0.8 via addition of IPTG to a final concentration of 0.2 mM for *CN*F8, *CN*F10, *CN*F20, and *CN*F54 or 1 mM for *CN*F2 and *CN*F16 SH3^{Sho1}. Expression was allowed to continue at 37 °C for 13 hours (*CN*F20), 20 hours (*CN*F2), or 24 hours (*CN*F8, *CN*F10, *CN*F54). For *CN*F16 SH3^{Sho1}, the temperature during expression was reduced to 30 °C for 18-20 hours.

Cells were isolated by centrifugation, resuspended in 50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8.0, and lysed via lysozyme treatment and sonication. The lysate was treated with DNaseI (New England Biolabs) or benzonase (Santa Cruz Biotech), clarified by centrifugation, combined with NiNTA resin (GoldBio), and rocked gently on ice for 45-60 minutes. The media was subsequently washed with three volumes of 50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0, followed by an equal volume of the same buffer containing 250 mM imidazole, to elute bound SH3. Eluent was dialyzed into phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) before 16-hour cleavage at room temperature by thrombin (Novagen) at a concentration of 1 U enzyme per mg of protein followed by inhibition with 1 mM phenylmethylsulfonyl fluoride (PMSF). Any uncleaved protein was removed by passage over NiNTA media, as described above. The protein was further purified via size exclusion chromatography (S100HR Sephacryl, GE Life Sciences). Protein concentrations were determined using calculated extinction coefficients of ε_{280} = 17,810 $M^{-1}cm^{-1}$ for all variants except *CN*F20, for which ε_{280} = 19,300 $M^{-1}cm^{-1}$ was used.

Synthesis and characterization of pPbs2.

Proline-rich recognition sequence pPbs2 (Ac-VNKPLPPLPVA-NH2) was synthesized via standard (fluorenylmethoxy)carbonyl (Fmoc) solid-phase peptide synthesis (Applied Biosystems 433A peptide synthesizer) using H-Rink Amide-ChemMatrix resin and 10-20 molar equivalents of amino acids. Cleavage from the solid support and simultaneous side chain deprotection was accomplished via treatment with trifluoroacetic acid containing 2% (v/v) triisopropylsilane and 2% water for 2 hours. Crude peptide was recovered by diethyl ether precipitation, dissolved in water, and subsequently lyophilized. The identity of the crude peptide was confirmed by mass spectrometry before purification on a Luna C18 reversed-phase HPLC column (Phenomenex). Purified peptide was lyophilized and stored at -20 °C. A sample of purified peptide was analyzed by both absorbance spectroscopy and amino acid analysis (AstraOmics, UC Davis Genome Center) to determine the extinction coefficient, ε_{205} = 43,319 $M^{-1}cm^{-1}$, which was used for all calculations of peptide concentration.

Characterization of *CN*F-labeled SH3^{Sho1}

Mass Spectrometry.

To verify *CN*F incorporation, each variant of SH3^{Sho1} was digested by trypsin (Thermo Scientific) with solutions of 0.1 mg/mL protein with 30:1 mass ratio of trypsin:SH3 and 25% acetonitrile. After incubation overnight at 37 °C, samples were desalted with C4 or C18 Zip Tips (EMD Millipore) before analysis by MALDI-TOF on a Bruker Autoflex III (Billerica, MA). Representative mass spectra are shown in Fig. S1 and summarized in Table S1. For only *CN*F20 SH3^{Sho1} do the mass spectra show the presence of a significant amount of wild-type protein along with the labeled domain. Samples of *CN*F20 SH3^{Sho1} showed 34-50% labeling efficiency, depending on the preparation.

Circular dichroism spectroscopy.

Circular dichroism spectra were acquired on a Jasco J-715 spectrometer from 270 to 190 nm for 0.01 mM protein solutions in 10 mM sodium phosphate, pH 7.0 (Fig. S2).

Fluorescence-based binding assays.

The dissociation constants for binding of each of the SH3^{Sho1} variants to pPbs2 were determined via intrinsic tryptophan fluorescence spectroscopy as described previously.[1] Briefly, a set of solutions were prepared containing a fixed SH3^{Sho1} concentration (3 or 0.5 μM) and variable pPbs2 concentration (0-50 μM). All solutions equilibrated for a minimum of 12 hours at 4 °C before analysis. Emission was recorded from 290 to 470 nm (5 nm slit width) with an excitation wavelength of 280 nm (5 nm slit width). The barycentric mean of each spectrum was calculated between 300 and 400 nm and the change in this value with pPbs2 concentration (Fig. S3) was fit to a standard two-state binding model to determine the K_d (Table S2).

Sample preparation for IR spectroscopy.

For FT IR spectroscopy, the SH3^{Sho1} variants were exchanged and concentrated by filtration into 50 mM sodium phosphate, pH 7.0, 100 mM NaCl. A volume of 5 μL was

loaded between two, 2 mm thick, CaF_2 windows separated by a 38.1 μm or 76 μm Teflon spacer for 1D or 2D experiments, respectively. The 1D spectra of the unligated protein were collected with samples of 1.5 mM $SH3^{Sho1}$, while spectra of the complex with pPbs2 were collected with 1.5 mM $SH3^{Sho1}$ and 2 mM pPbs2. For 2D IR spectroscopy, all samples, with the exception of *CN*F20 $SH3^{Sho1}$, contained 4 mM $SH3^{Sho1}$ and for the complex with pPbs2 contained 4.8 mM pPbs2. For *CN*F20 $SH3^{Sho1}$, samples contained 5 mM protein, while 6 mM pPbs2 was added to prepare the complex.

FT IR spectroscopy and analysis.

FT IR spectra were collected on an Agilent Cary 670 FT IR spectrometer using a liquid nitrogen cooled mercury-cadmium-telluride (MCT) detector at 2 cm^{-1} resolution. Wild-type $SH3^{Sho1}$ was used to generate reference spectra. For characterization of both reference and sample, 5,000 scans were averaged after purging the chamber with dry nitrogen for 10 min. All spectra were processed using a four-term Blackman-Harris apodization function, zero-filling factor of eight and a Mertz phase correction algorithm. A residual, slowly varying, baseline in the absorption spectra was removed by fitting a polynomial to a spectral region of ~ 200 cm^{-1}, excluding the CN absorption band (Matlab R2016b). The first moment of each spectrum was calculated and all were found to be similar, if not identical, to the frequencies of maximum absorbance (Table S3), suggesting a single underlying state. The absorption bands were then fit to single Gaussian functions to determine the center frequencies and line widths. All experiments were performed in triplicate with independently prepared protein samples; averages and standard deviations from the fits are given in Table 1 of the main text. To further visualize the linear data, the area of each spectrum was normalized to unity and difference spectra were generated of the unligated proteins and pPbs2 complexes (Figure S4).

2D IR spectroscopy and data analysis.

2D IR spectra were obtained as previously described.[4, 5] A Ti:Sapphire oscillator/regenerative amplifier (Spectra Physics) producing ~75 fs pulses centered at 800 nm with 1 kHz repetition rate was used to pump a home-built optical parametric amplifier to generate ~170 fs pulses centered at 2225 cm^{-1} (FWHM = 80 cm^{-1}) to encompass both the 0-1 and 1-2 vibrational states for the *CN*F probe in the protein/aqueous solution. The beam was split into three beams of approximately equal energy, with total excitation energy of 25 μJ applied for the 2D IR experiments of the proteins. The three, all parallel-polarized, beams were focused into the sample in a conventional BOXCARS geometry. Beam three, which is fixed in time, was chopped at 500 Hz to remove scattered light from the other excitation beams. Pulse temporal overlap was set by cross-correlations using a non-resonant signal from CCl_4. Chirp was determined by frequency-resolved optical grating experiments and corrected to less than ±0.02 fs/cm^{-1} by placement of Ge and CaF_2 substrates in the beam paths.

In a 2D IR vibrational echo experiment, application of the three pulses leads to the generation of a third order signal in the phase-matched ($-k_1 + k_2 + k_3$) direction. This signal was heterodyne-detected by overlap with a fourth beam, the local oscillator, to provide amplification and phase information. The combined beam was dispersed by a spectrograph onto a 32-element MCT detector (Infrared Systems Development). A single 2D spectrum was generated by scanning the time between the first two pulses (τ) while the time period

between the second and third pulses (T_w) was held constant. The heterodyned third-order signal generated at a time $\leq \tau$ after the application of the third pulse produced an interferogram along τ, which was Fourier transformed to produce the ω_τ axis. Frequency-resolved detection with the MCT array generated the ω_m axis of the 2D spectra. A reference beam, split from the LO, was detected on a single element MCT detector and used to correct for shot-to-shot laser fluctuations. 2D IR spectra were acquired as a function of T_w (Fig. S5) The 2D IR spectra were corrected for phase errors and inner filter effects as described previously.[4, 6]

The normalized FFCFs were extracted from the T_w-dependent changes in 2D spectral line shapes via the center line slope (CLS) method.[7] Briefly, 1D slices along the ω_τ axis at each ω_m were fit to Gaussian functions to determine the ω_τ of maximum absorbance. These ω_τ were plotted as a function of ω_m to reveal the slope of the center line. These slopes were then plotted as a function of T_w to obtain the CLS decay, which approximates the normalized, inhomogeneous part of the FFCF.[7] The complete FFCF including frequency fluctuation amplitudes was obtained by simultaneous fitting of the CLS decay and linear IR spectrum.[4] The FFCFs were analyzed as described in the main text according to the Kubo model[8] with the equation

$$FFCF = \frac{\delta(t)}{T_2} + \Delta_1^2 e^{-t/\tau_1} + \Delta_s^2$$

in which the latter two terms describe the dynamics among the inhomogeneous distribution of frequencies underlying the absorption bands. The inhomogeneous dynamics are separated into two timescales, where Δ_1 is the frequency fluctuation amplitude sampled on the faster timescale τ_1, and the static term Δ_s, is the frequency fluctuation amplitude sampled more slowly than the experimental time window. The first term, $\delta(t)/T_2$, where $1/T_2 = (1/T_2^*) + (1/2T_1)$, accounts for the homogeneous contribution to the FFCF. T_1 is the vibrational lifetime, which was set to a value of 4.2 ps previously measured for the amino acid.[9] The pure dephasing time, $T_2^* = (\Delta^2\tau)^{-1}$, describes very fast fluctuations that are in the motionally narrowed limit on the IR timescale, where the frequency amplitude and timescale cannot be separated ($\Delta\tau \ll 1$). The homogeneous dynamics lead to a Lorentzian contribution to the line shape, $\Gamma^* = 1/\pi T_2^*$.

The T_w-dependent change in the 2D band amplitudes were analyzed to determine the lifetime (T_1) of the CN probe for each sample. Slices along the diagonal of the 2D spectra were extracted and the change in the maximum absorbance was fit to an exponential function (Table S4). The error in the calculation precludes any significant conclusions to be drawn, but all lifetimes are similar to that of free amino acid in aqueous solution.

Molecular dynamics (MD) simulations and analysis.

Charges for *CN*F were derived for the amino acid capped with acetyl and N-methylamide groups via the R.E.D. Server.[10-13] The Gaussian geometry optimization and charge fitting were performed at the B3LYP/6-31G(d) level with 10 layers and 17 point density. The total charge for each cap was restrained to zero and the charges of the backbone and beta carbon atoms of *CN*F were fixed to those found in the Amber ff14SB force field for Tyr and Phe. In the absence of these restraints, the calculations yielded charges of relatively large magnitude (~0.2-0.4) for the backbone and beta carbon atoms of *CN*F. Calculations performed with higher levels of theory generally increased the

magnitude of these charges. The partial charges used for the cyano carbon and nitrogen were 0.32 and -0.43, respectively.

*CN*F was introduced into the crystal structural model of SH3^{Sho1} bound to pPbs2 (PDB ID 2VKN) at residues 10, 16, 20, 54 or 2 and 8 using Chimera (UCSF).[14] *CN*F was placed at both residues 2 and 8 in one structure to minimize computational effort, as residue 2 is considered a control residue on the opposite side of the protein as the peptide binding site and is not expected to perturb the dynamics at residue 8. Force field library files for the *CN*F and Amber topology and coordinate files for *CN*F-labeled SH3^{Sho1} with and without pPbs2 were generated via the LEaP program. The protein was solvated by a periodic 12 Å octahedron of TIP3 water and Na^+ counter ions were added to neutralize the charges of the system. Additional Na^+ and Cl^- ions were then added to make the system 150 mM in Na^+ concentration according to the total number of water molecules. For the SH3^{Sho1}-pPbs2 complex, ~6000 water molecules and eight Na^+ ions were added to solvate and neutralize the charge, then an additional eight Na^+ and Cl^- ions were added to make the ionic strength 150 mM. For the free SH3^{Sho1}, ~5000 waters and a total of 13 Na^+ and four Cl^- ions were added.

Energy minimization and MD simulations were performed using the Amber16 package on BigRed2 at Indiana University. The Particle Mesh Ewald summation method with a non-bonded cut-off of 10 Å was employed for long-range interactions and the SHAKE procedure was used to constrain all bonds involving hydrogen atoms. A Langevin thermostat was used with a collision frequency of 2 ps^{-1}. Each system was first energy-minimized by 2000 steps with the protein/peptide atoms highly restrained (force constant 500 kcal/mol-$Å^2$), then energy-minimized unrestrained by 2000 steps. The temperature of the system was increased to 300 K over 50 ps at constant volume with moderate restraints on the protein/peptide (force constant 10 kcal/mol-$Å^2$). The unrestrained system was then equilibrated at 300 K for 50 ps. The system then was equilibrated at constant pressure (1 bar, isotopic position scaling, Berendsen barostat) at 300 K for 500 ps. Self-guided Langevin dynamics were then run for 1 ns with a 2 ps local averaging time and target guiding temperature of 450 K. Ten frames from this trajectory (separated by 100 ps) were extracted and used to start two sets of 5 ns production simulations with 1 fs time steps, saving the coordinates every 100 fs.

MD trajectories were analysed using the CPPTRAJ program of Amber16.[15] To investigate the parts of the protein influencing the environment of each *CN*F residue, the MD trajectories were analyzed to find the relative frequency that atoms approach the cyano nitrogen within 4 Å (Table S6), and radial distribution functions (RDFs) were determined for the distance of the cyano nitrogen to select sets of atoms (*e.g.* all atoms, all heavy atoms, all protein heteroatoms, etc.; Fig. S9, Fig. S10). Analysis of the hydrogen bonding of the CN to solvent or protein moieties was conducted for each residue using a distance cutoff of 3 Å and no angle restriction, hydrogen bonds with angles less than 90° were manually excluded. The frequency that each CN engaged in a hydrogen bonding interaction during the trajectories and the hydrogen bond time correlation function were determined for each variant in the unligated state and complex with pPbs2 (Table S7; Fig. S8). The electric field (EF) along the CN bond vector was calculated for each frame of the trajectories as previously described.[16] Briefly, for each frame of the MD trajectories, the EF at the carbon and nitrogen atoms of the *CN*F were determined from the partial charges of and distances from all other atoms, excluding the *CN*F side chain; the EFs were projected onto the CN

bond vector; and the projected EF at the carbon and nitrogen atoms were averaged. Histograms and the time autocorrelation of the average EF at each CN were determined for all sets of trajectories (Fig. S7 & 6). Both the electric field and hydrogen bonding time correlation functions were fit to biexponential decays with an offset (Table S8).

Supplemental Figures and Tables

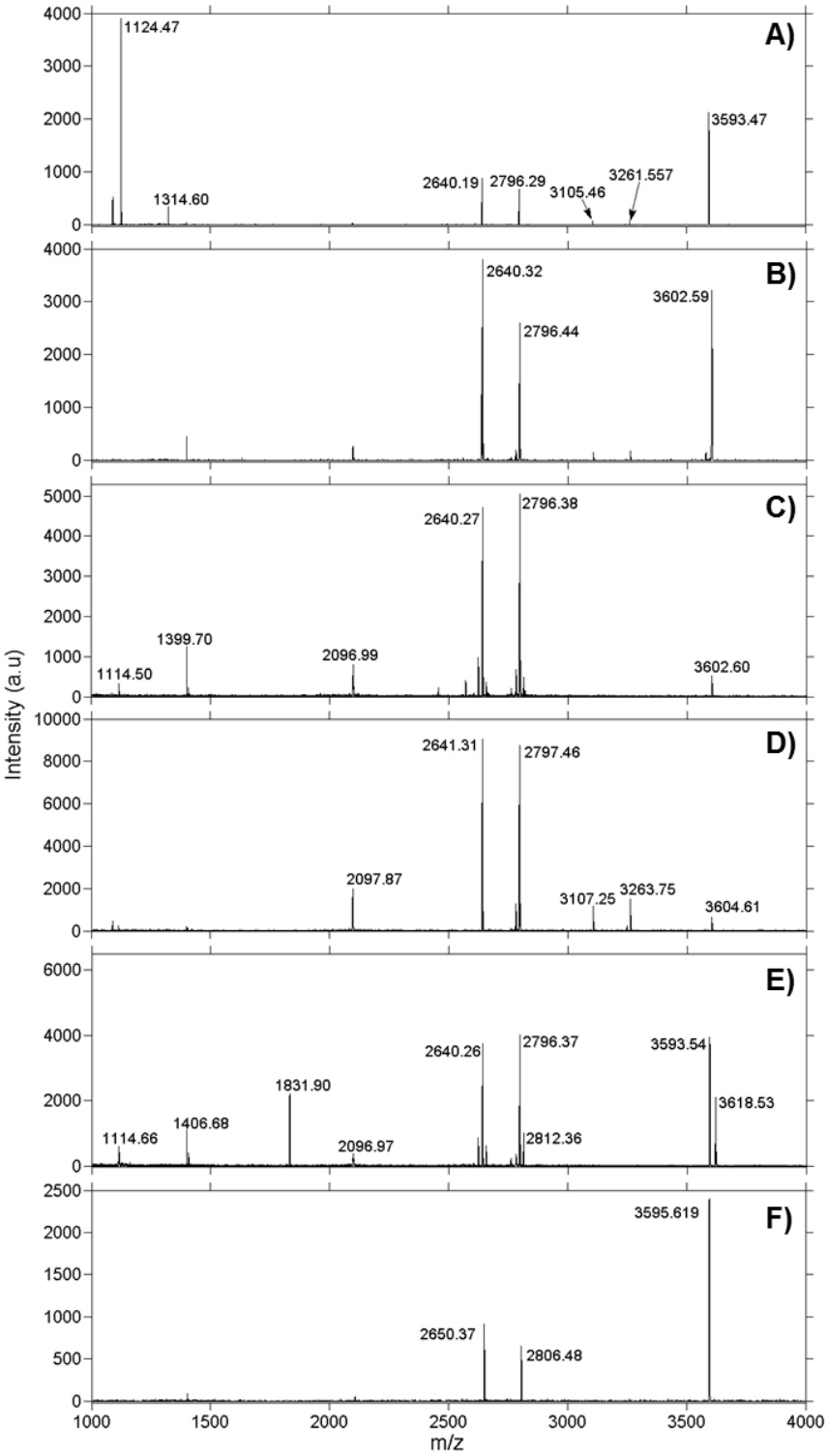

Fig. S1. Representative mass spectra of trypsin digests. (A) *CNF*2, (B) *CNF*8, (C) *CNF*10, (D) *CNF*16, (E) *CNF*20, and (F) *CNF*54 SH3^{Sho1}.

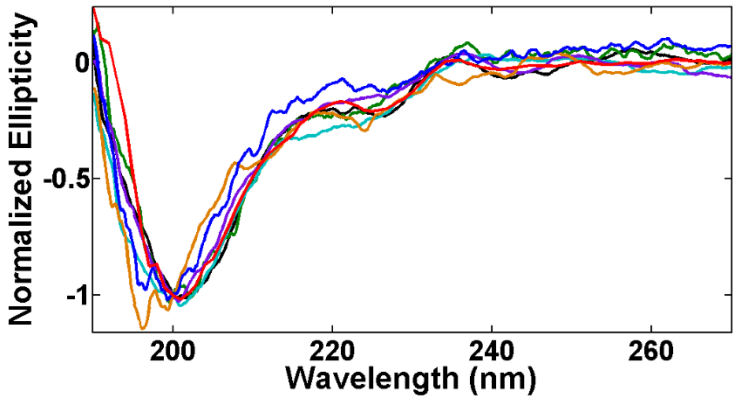

Fig. S2. Circular dichroism spectra normalized to ellipticity at 200 nm for unlabeled (black), *CN*F2 (orange), *CN*F8 (teal), *CN*F10 (blue), *CN*F16 (red), *CN*F20 (green), and *CN*F54 (purple) SH3^{Sho1}.

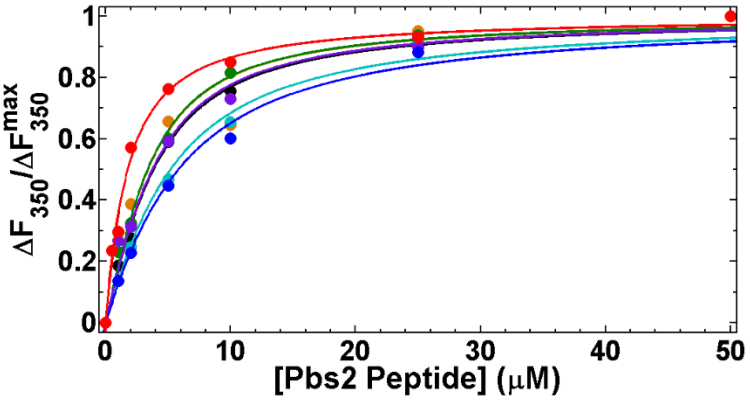

Fig. S3. Fluorescence-based assay (points) and fits to two-state model (lines) for binding of pPbs2 to unlabeled (black), *CN*F2 (orange), *CN*F8 (teal), *CN*F10 (blue), *CN*F16 (red), *CN*F20 (green), and *CN*F54 (purple) SH3^{Sho1}.

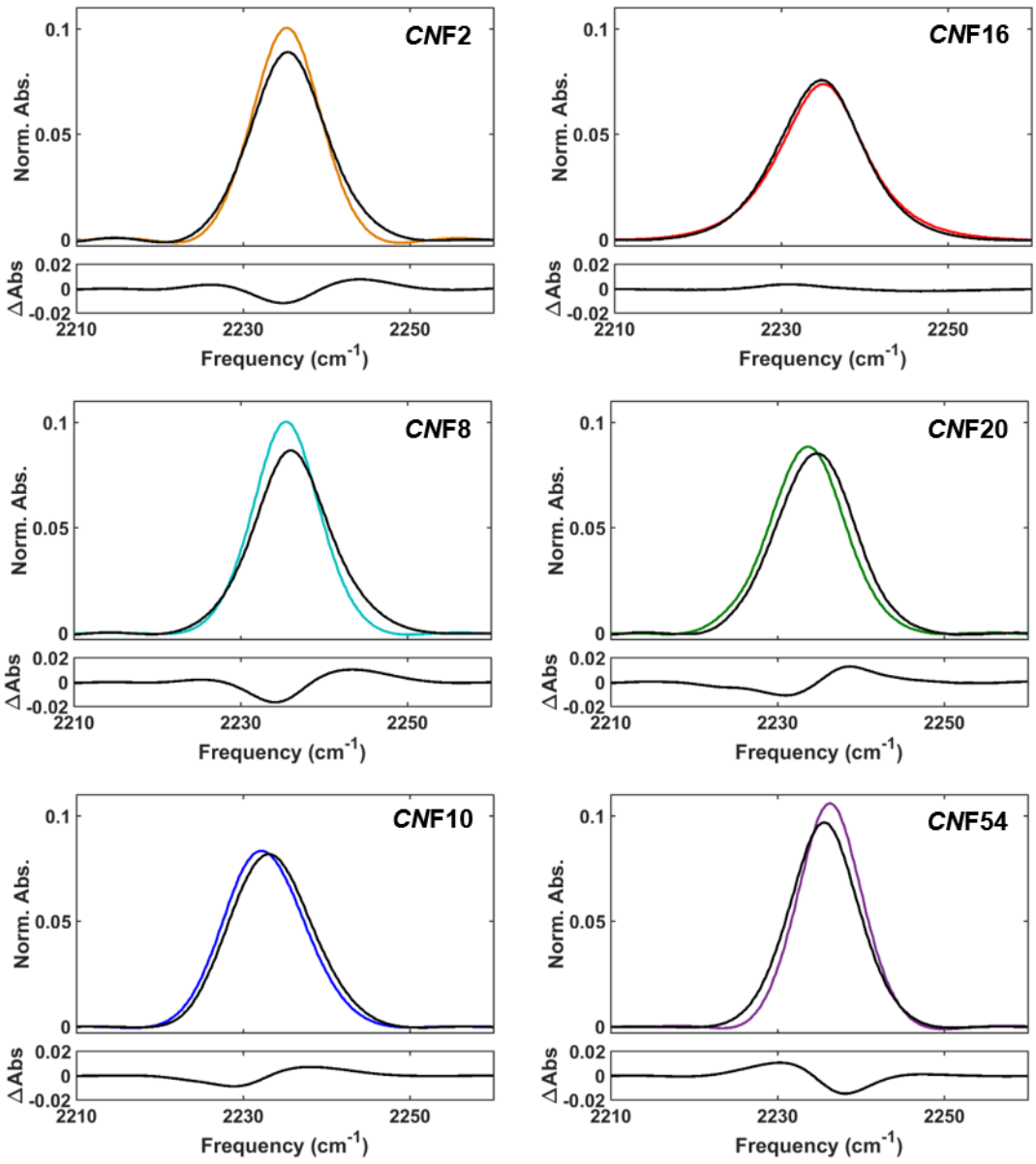

Fig. S4. FT IR spectra of the unligated state (colored lines) and the pPbs2 complex (black lines) with the area normalized to unity; difference spectra in black are displayed below each set of spectra.

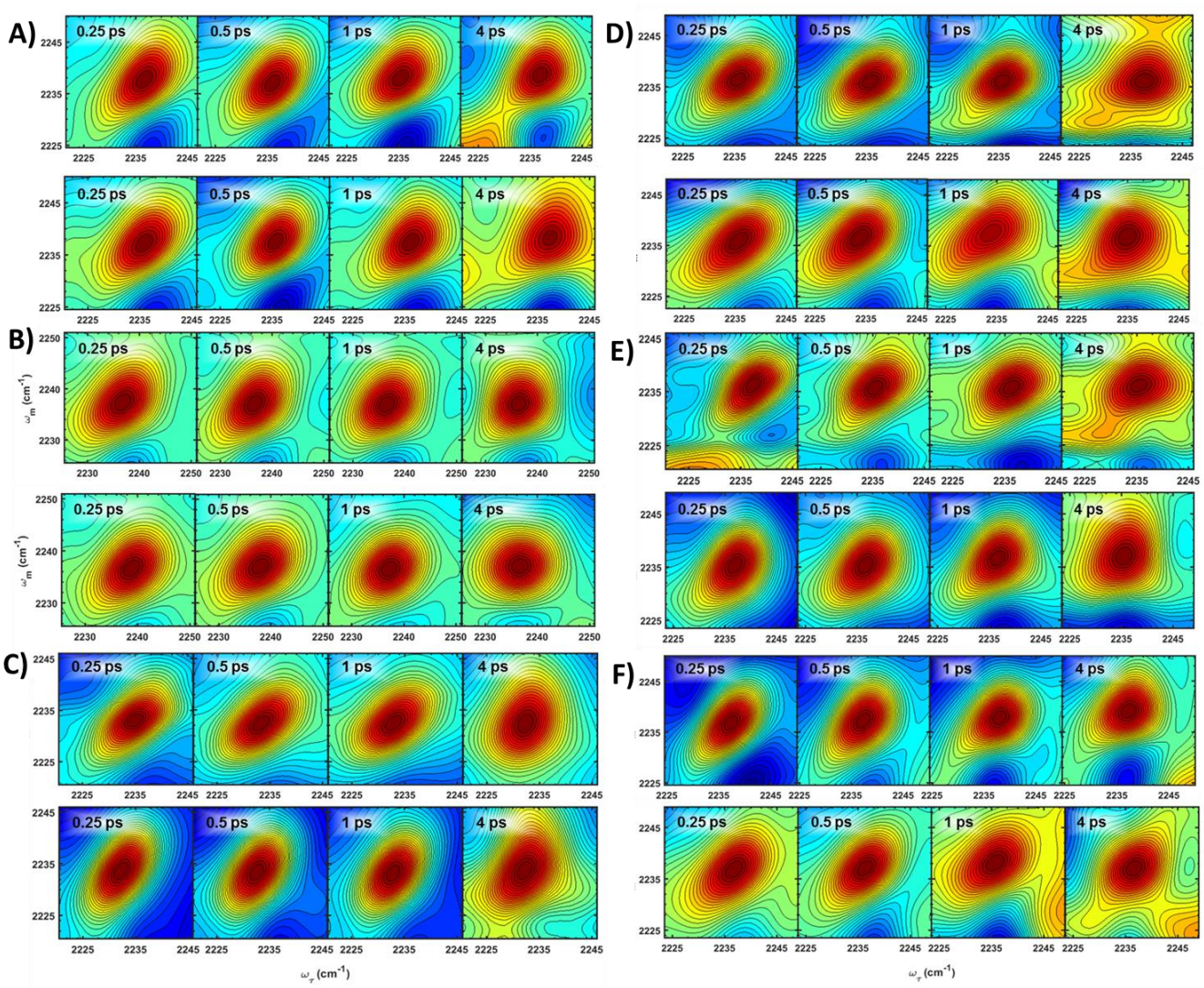

Fig. S5. Representative T_w-dependent 2D IR spectra of the unligated state (top row) and the pPbs2 complex (bottom row). (A) *CN*F2, (B) *CN*F8, (C) *CN*F10, (D) *CN*F16, (E) *CN*F20, and (F) *CN*F54 SH3^{Sho1}.

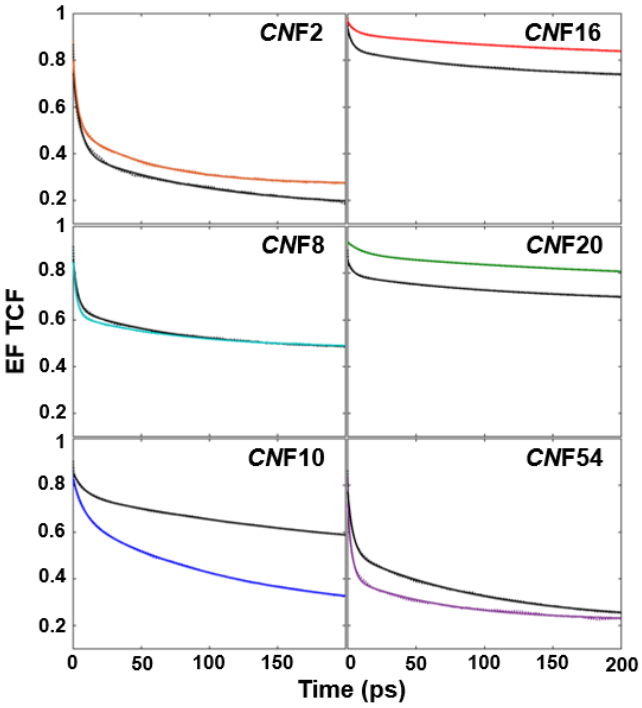

Fig. S6. EF TCFs determined from MD simulations of the unligated state (colored lines) and in complex with pPbs2 (black lines).

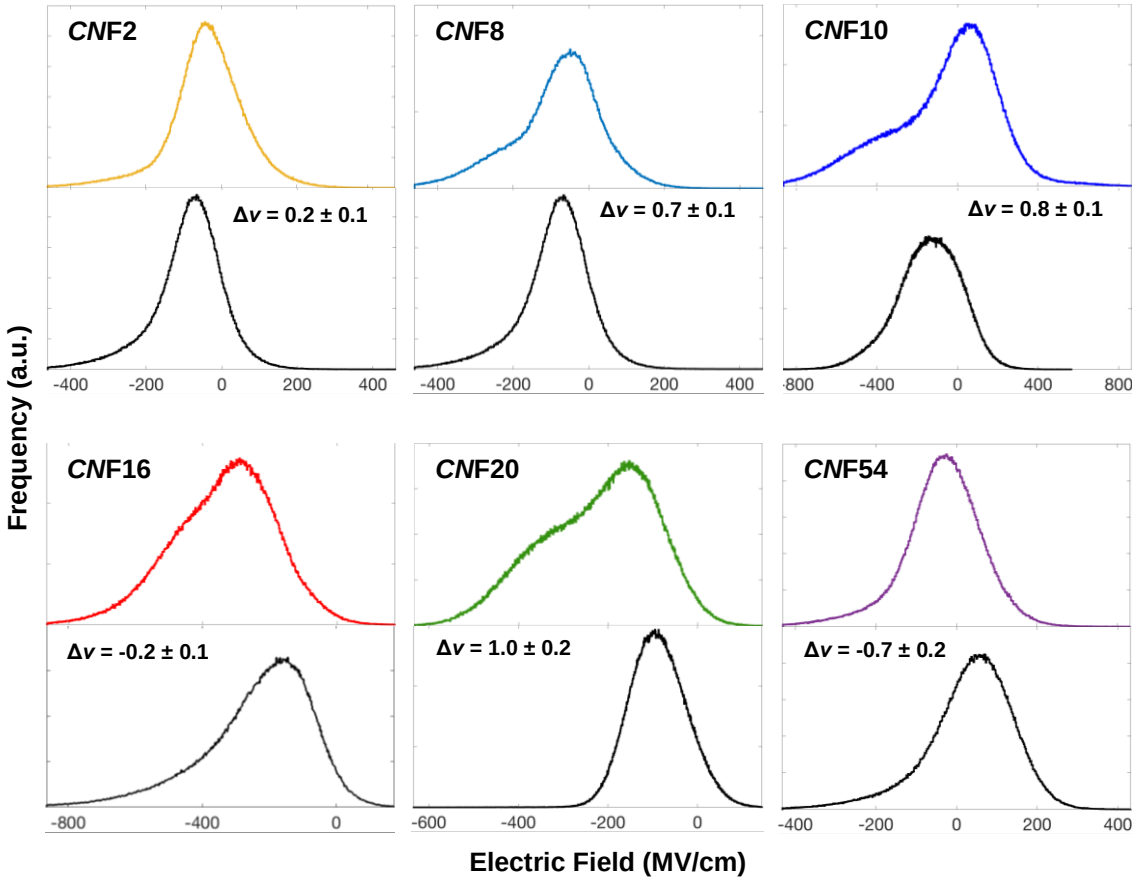

Fig. S7. Histograms of the electric field (EF) along the CN bond from the MD trajectories for the unligated state (upper panel, colored) and the complex with pPbs2 (lower panel, black). The EF strengths are scaled lower by a factor of 2.5 as in reference 17.[16] The changes in the vibrational frequency observed upon ligand binding determined from analysis of the FT IR spectra are shown in the lower panel.

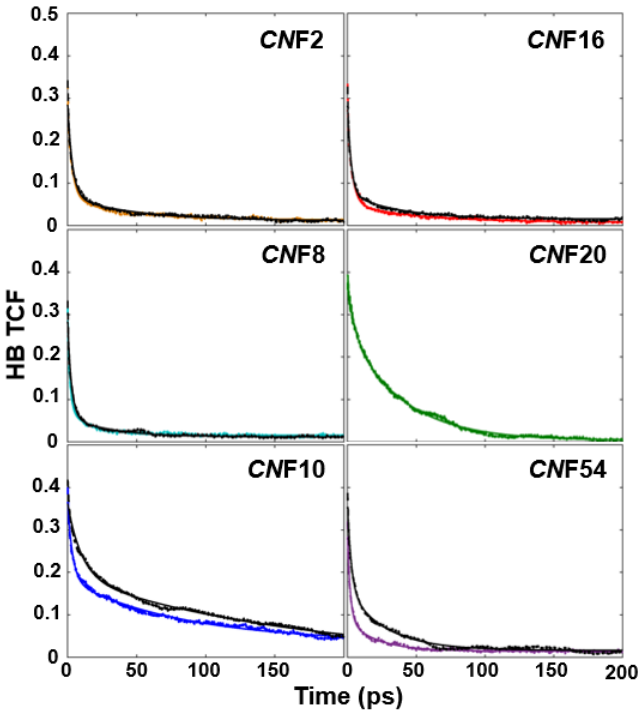

Fig. S8. HB TCFs determined from MD simulations of the unligated state (colored lines) and complex with pPbs2 (black lines).

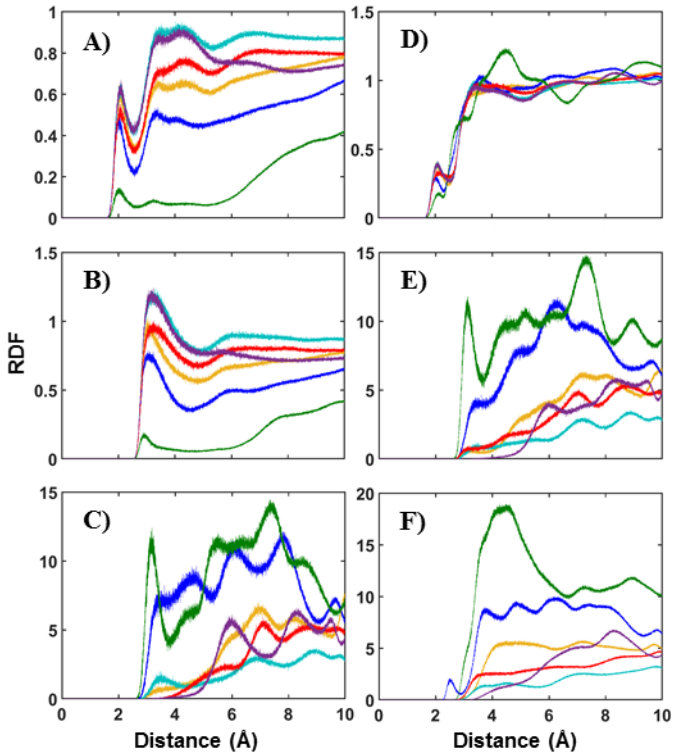

Fig. S9. RDFs of distance of the cyano nitrogen of *CN*F2 (teal), *CN*F8 (blue), *CN*F10 (orange), *CN*F16 (red), *CN*F20 (green), and *CN*F54 (purple) for unligated SH3^{Sho1} to (A) hydrogen atoms of water, (B) water oxygen atoms of water, (C) oxygen atoms of protein, (D) all atoms excluding *CN*F, (E) all heteroatoms of protein, and (F) all ions and heavy atoms of protein excluding *CN*F.

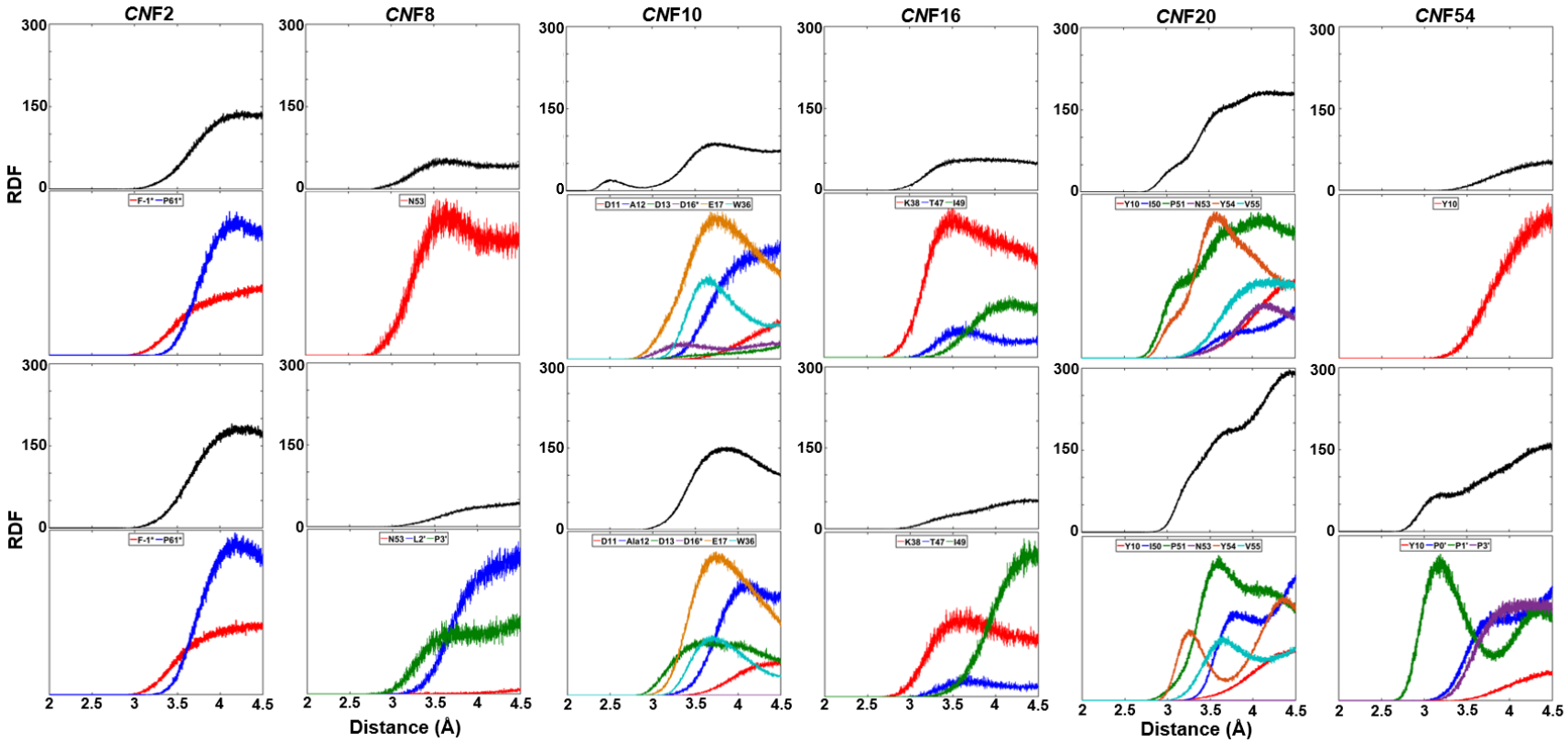

Fig. S10. RDFs for distance of the cyano nitrogen to the heavy atoms of all residues found within 4 Å for >10% of the MD trajectories (black lines) and to the heavy atoms for each of the individual residues (colored lines) for the unligated protein (upper two panels) and the Pbs2 complex (lower two panels). A list of included residues can be found in Table S5.

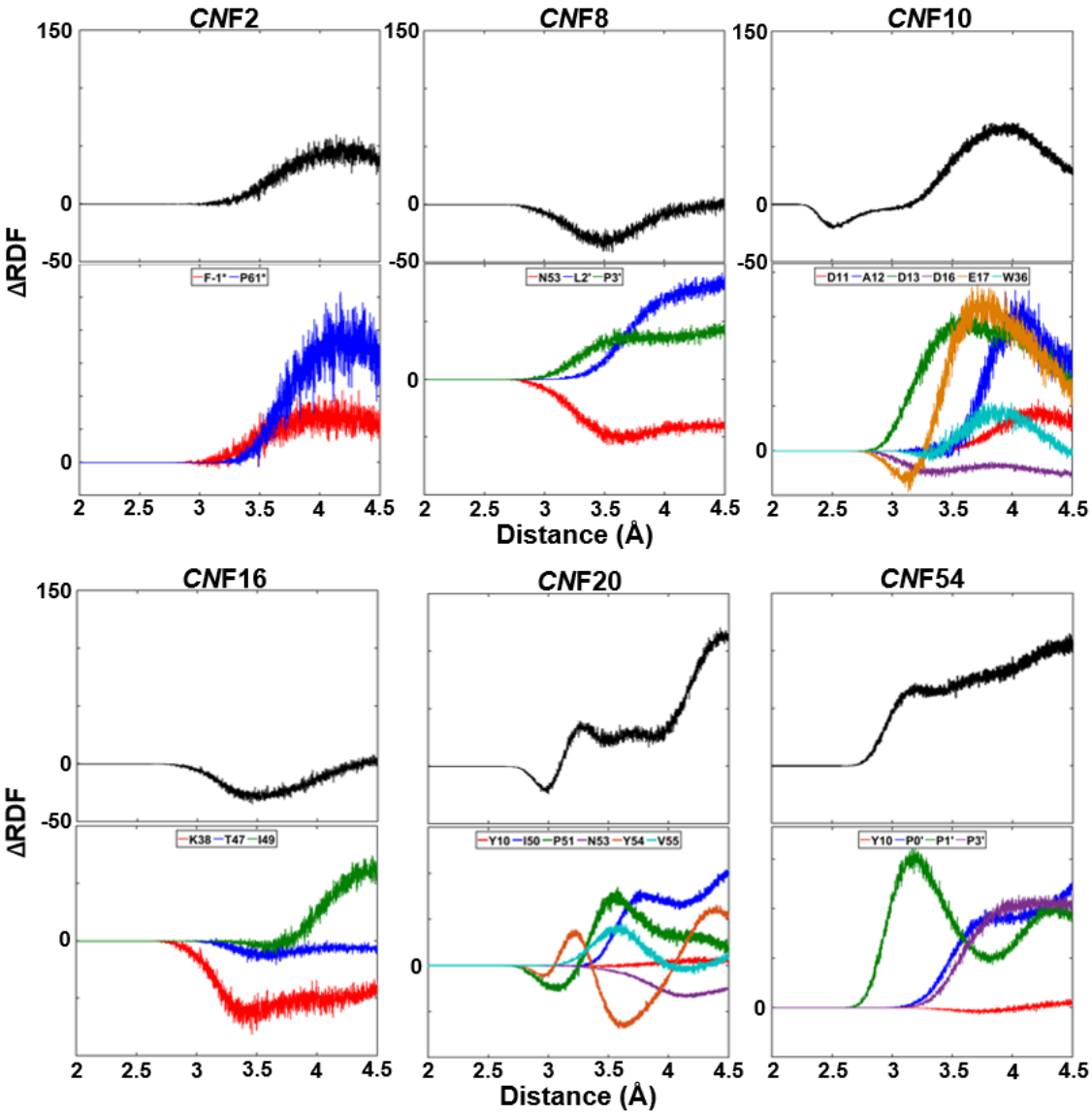

Fig. S11. Change upon pPbs2 binding in the RDFs for distance of the cyano nitrogen to the heavy atoms of all residues found within 4 Å for >10% of the MD trajectories (upper panels, black lines) and to the heavy atoms for each of the individual residues (lower panels, colored lines). A list of included residues can be found in Table S5.

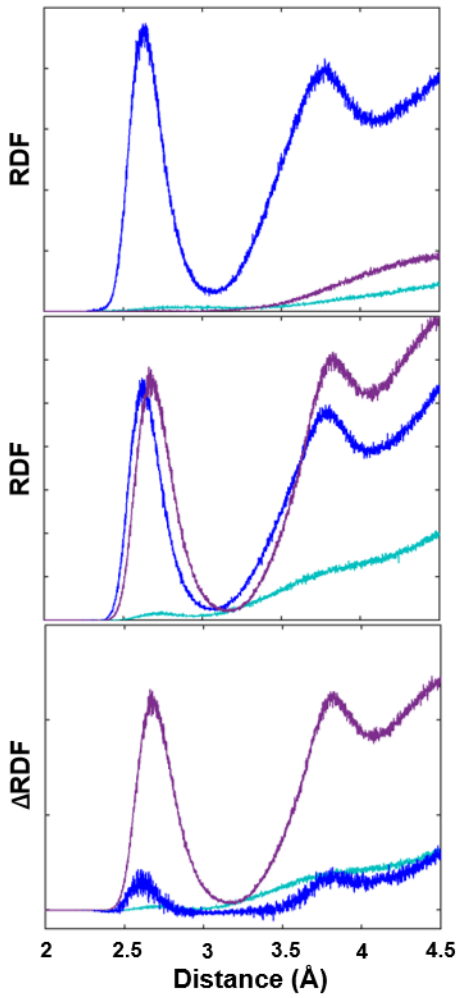

Fig. S12. RDF for distance of the hydroxyl oxygen atom of Tyr8 (teal), Tyr10 (blue) and Tyr54 (purple) to all heavy atoms, excluding water and the Tyr of interest, for the unligated protein (top), the pPbs2 complex (middle), and the difference upon binding (bottom).

Table S1. Summary of mass spectrometry data for tryptic digests of *CN*F-labeled SH3^{Sho1}.

Expected Fragment Mass for Unlabeled Protein	Modifications	Observed *CN*F2 fragments	Observed *CN*F8 fragments	Observed *CN*F10 fragments	Observed *CN*F16 fragments	Observed *CN*F20 fragments	Observed *CN*F54 fragments
1115.50	1 Met-loss	1124.47**		1114.5		1114.66	
1304.55	1 Acetyl 1 Oxidation	1314.60**[a]					
1314.63	1 Met-loss	1314.60[a]					
2640.27		2640.19	2640.32	2640.27	2641.31	2640.26	2650.37**
2796.37		2796.29	2796.44	2796.38	2797.46	2796.37	2806.48**
2812.37	1 Oxidation					2812.36, 1406.68	
3105.58		3105.46			3107.25		
3261.68		3261.56			3263.75		
3593.61		3593.47	3602.59**	3602.60**	3604.61**	3593.54, 3618.53**	3595.62

**Indicates mass shift observed due to incorporation of *CN*F at the desired location.
[a]Mass could be due to the presence of either or both of two fragments, one which exhibits the desired mutation, and one which corresponds to the wild-type mass.

Table S2. Dissociation constants for binding of the *CN*F-labeled SH3^{Sho1} and pPbs2.

Variant	K_d (μM)
Wild-type	2.3 ± 0.7
*CN*F2	1.9 ± 0.3
*CN*F8	3.9 ± 0.6
*CN*F10	4.5 ± 0.5
*CN*F16	1.5 ± 1.3
*CN*F20	2.0 ± 0.7
*CN*F54	3.4 ± 1.6

Table S3. Analysis of 1D IR spectra of SH3^{Sho1} variants.

	1st moment (cm^{-1})	Frequency (cm^{-1})*
***CN*F2**	2235.2	2235.2 ± 0.1
***CN*F2-Pbs2**	2235.6	2235.6 ± 0.1
***CN*F8**	2235.4	2235.6 ± 0.3
***CN*F8-Pbs2**	2236.3	2236.1 ± 0.2
***CN*F10**	2232.7	2232.5 ± 0.2
***CN*F10-Pbs2**	2233.5	2233.3 ± 0.2
***CN*F16**	2234.9	2235.1 ± 0.2
***CN*F16-Pbs2**	2234.6	2234.7 ± 0.1
***CN*F20**	2233.5	2233.6 ± 0.2
***CN*F20-Pbs2**	2234.5	2234.7 ± 0.2
***CN*F54**	2236.3	2236.3 ± 0.2
***CN*F54-Pbs2**	2235.6	2235.6 ± 0.1

*Frequency of maximum absorbance

Table S4. Vibrational lifetimes of CN stretch calculated from 2D IR spectra.

	T_1 (ps)
***CN*F2**	4.0 ± 2.5
***CN*F2-Pbs2**	3.9 ± 1.9
***CN*F8**	4.1 ± 2.0
***CN*F8-Pbs2**	5.0 ± 2.8
***CN*F10**	4.3 ± 1.6
***CN*F10-Pbs2**	4.4 ± 2.6
***CN*F16**	3.9 ± 2.2
***CN*F16-Pbs2**	5.4 ± 2.8
***CN*F20**	5.0 ± 4.3
***CN*F20-Pbs2**	4.2 ± 2.8
***CN*F54**	4.5 ± 4.0
***CN*F54-Pbs2**	4.1 ± 3.2

Table S5. Frequency of hydrogen bonding of *CN*F in MD trajectories.

	Water	Protein	Total
***CN*F2**	7.6 %	2.1 %	9.7 %
***CN*F2-Pbs2**	9.9 %	0.3 %	10.2 %
***CN*F8**	8.4 %	0.2 %	8.6 %
***CN*F8-Pbs2**	9.0 %	0.5 %	9.5 %
***CN*F10**	7.4 %	0.0 %	7.4 %
***CN*F10-Pbs2**	11.6 %	1.8 %	13.4 %
***CN*F16**	7.2 %	3.4 %	10.5 %
***CN*F16-Pbs2**	10.5 %	3.3 %	13.8 %
***CN*F20**	2.2 %	6.4 %	8.6 %
***CN*F20-Pbs2**	0.0 %	0.0 %	0.0 %
***CN*F54**	9.0 %	0.0 %	9.0 %
***CN*F54-Pbs2**	10.2 %	0 %	10.2 %

Table S6. Atoms within 4 Å of cyano nitrogen in >10% of the MD trajectories.

	Unligated			Bound		
	Atom	**Residue**	**Occurrence (%)**	**Atom**	**Residue**	**Occurrence (%)**
***CN*F2**						
	CD2	PHE(-1)†	22.3	CG	PRO61	23.4
	CG	PRO61	21.0	CD2	PHE(-1)†	19.5
	CD	PRO61	16.4	CD	PRO61	19.0
				CE2	PHE(-1)†	18.3
***CN*F8**						
	CB	ASN53	11.5	O	PRO3’	14.4
	O	ASN53	10.9	CD1	LEU2’	12.3
***CN*F10**						
	CG	GLU17	43.9	CG	GLU17	73.8
	CZ3	TRP36	35.3	CZ3	TRP36	44.9
	CB	GLU17	33.8	CB	GLU17	44.4
	OE2	GLU17	26.7	N	ASP13	35.7
	OE1	GLU17	25.3	CD	GLU17	33.3
	CD	GLU17	23.4	CB	ASP13	24.7
	CB	ALA12	17.5	CA	ALA12	23.8
	CA	ALA12	15.4	OE2	GLU17	21.6
	OD2	ASP16†	14.2	OE1	GLU17	13.5
	Na+	Na+68	13.0	CB	ALA12	12.5
	Na+	Na+76	12.1	O	ASP11	10.1
***CN*F16**						
	CE	LYS38	37.0	CE	LYS38	12.5
	CD	LYS38	22.3			
	NZ	LYS38	16.5			
	CG	LYS38	15.8			
	CG2	ILE49	11.0			
	CG2	THR47	10.7			
***CN*F20**						
	CB	TYR54	83.5	CB	TYR54	99.7
	O	PRO51	69.1	CD	PRO51	90.6
	CG	PRO51	51.3	CB	ILE50	76.7
	CD	PRO51	50.0	O	PRO51	74.0
	N	TYR54	46.6	CG1	VAL55	73.4
	CA	TYR54	42.8	CG	PRO51	45.2
	CD2	TYR54	36.5	CD1	ILE50	21.0
	CG	TYR54	33.3	CG2	ILE50	18.8
	C	TYR54	30.1	N	PRO51	17.3
	CG1	VAL55	28.0	C	TYR54	17.3
	CD1	TYR54	27.5	CG	TYR54	12.2
	CD1	TYR10	16.6	CD1	TYR10	11.8
	CB	ILE50	16.4	CB	TYR10	11.4
	N	ASN53	16.1	CA	TYR54	10.6
	N	VAL55	14.6			
	C	PRO51	14.2			
	CG2	VAL55	13.0			
	CG	TYR10	11.0			
	O	TYR54	10.6			
***CN*F54**						
	CE1	TYR10	10.5	O	PRO1’	94.8
				CB	PRO0’	36.5
				CD	PRO3’	28.0
				CG	PRO3’	14.3
				C	PRO1’	11.5
				CG	PRO0’	11.1

†Unique, consecutive numbering according to the consensus sequence of SH3 domains is not possible for Phe(-1) and Asp16 due to a two-residue insertion at this site in the sequence of SH3^{Sho1} and other yeast SH3 domains.[17]

Table S7. Parameters from fit of EF TCFs to biexponential decay with offset.

	a_1	τ_1	a_2	τ_2	a_3
***CN*F2**	0.3	3.6	0.2	57.2	0.3
***CN*F2-Pbs2**	0.3	5.3	0.2	96.8	0.2
***CN*F8**	0.2	2.7	0.1	82.2	0.5
***CN*F8-Pbs2**	0.2	3.3	0.2	77.5	0.5
***CN*F10**	0.2	8.3	0.4	117	0.3
***CN*F10-Pbs2**	0.1	8.2	0.3	220	0.5
***CN*F16**	0.05	5.7	0.1	183	0.8
***CN*F16-Pbs2**	0.1	3.8	0.1	113	0.7
***CN*F20**	0.5	10	0.1	201	0.8
***CN*F20-Pbs2**	0.06	4.0	0.1	126	0.7
***CN*F54**	0.3	3.2	0.2	71.6	0.2
***CN*F54-Pbs2**	0.3	4.0	0.3	103	0.2

Table S8. Parameters from fit of HB TCFs to biexponential decay with an offset.

	a_1	τ_1	a_2	τ_2	a_3
***CN*F2**	0.23	3.1	0.05	38.4	0.02
***CN*F2-Pbs2**	0.23	3.5	0.06	48.6	0.01
***CN*F8**	0.24	1.7	0.06	19.2	0.02
***CN*F8-Pbs2**	0.24	2.7	0.05	32.0	0.01
***CN*F10**	0.20	3.6	0.15	83.2	0.04
***CN*F10-Pbs2**	0.17	10.0	0.19	157.8	0.00
***CN*F16**	0.23	3.3	0.04	47.5	0.01
***CN*F16-Pbs2**	0.22	2.6	0.06	36.6	0.02
***CN*F20**	0.10	5.3	0.27	39.3	0.00
***CN*F54**	0.21	2.1	0.07	21.6	0.02
***CN*F54-Pbs2**	0.21	2.9	0.14	28.4	0.02

Table S9. Frequency of Tyr as a hydrogen bonding acceptor in MD simulations.

	Hydrogen Bond Occurrence Water	**Hydrogen Bond Occurrence Protein**	**Hydrogen Bond Occurrence Total**
Tyr8	18.5 %	0.7 %	19.2 %
Tyr8-Pbs2	23.8 %	0.0 %	23.8 %
Tyr10	10.6 %	0.0 %	10.6 %
Tyr10-Pbs2	19.9 %	1.1 %	21.0 %
Tyr54	18.2 %	0.1 %	18.3 %
Tyr54-Pbs2	10.8 %	0.0 %	10.8 %

References

1. R. E. Horness, E. J. Basom and M. C. Thielges, *Anal. Methods*, 2015, **7**, 7234-7241.
2. J. A. Marles, S. Dahesh, J. Haynes, B. J. Andrews and A. R. Davidson, *Mol. Cell*, 2004, **14**, 813-823.
3. A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao and P. G. Schultz, *Biochemistry*, 2013, **52**, 1828-1837.
4. S. Park, K. Kwak and M. D. Fayer, *Laser Phys. Lett.*, 2007, **4**, 704-718.
5. E. J. Basom, J. W. Spearman and M. C. Thielges, *J. Phys. Chem. B*, 2015, **119**, 6620-6627.
6. J. B. Asbury, T. Steinel, K. Kwak, S. A. Corcelli, C. P. Lawrence, J. L. Skinner and M. D. Fayer, *J. Chem. Phys.*, 2004, **121**, 12431-12446.
7. K. Kwak, S. Park, I. J. Finkelstein and M. D. Fayer, *J. Chem. Phys.*, 2007, **127**, 124503.
8. R. Kubo, *Adv. Chem. Phys.*, 1969, **15**, 101-127.
9. A. L. Le Sueur, S. Ramos, J. D. Ellefsen, S. P. Cook and M. C. Thielges, *Anal. Chem.*, 2017, **89**, 5254-5260.
10. C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269-10280.
11. F. Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7821-7839.
12. E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak and F. Y. Dupradeau, *Nucleic Acids Res.*, 2011, **39**, W511-517.
13. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. B. Scalmani, V.; , G. A. Petersson, H. Nakatsuji, X. C. Li, M.; Marenich, J. A. V.; Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr.; , J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 09.
14. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.
15. D. A. Case, R. M. Betz, D. S. Cerutti, I. T.E. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, AMBER 2016. (University of California, San Francisco).

16. S. D. Fried, L. P. Wang, S. G. Boxer, P. Ren and V. S. Pande, *J. Phys. Chem. B*, 2013, **117**, 16236-16248.
17. T. Brown, N. Brown and E. J. Stollar, *PLoS One*, 2018, **13**, e0193128.