Supporting Information to

Transient Protein Encounters Characterized by Paramagnetic NMR

Karen Van de Water, Nico A. J. van Nuland, and Alexander N. Volkov

Supplementary Figures

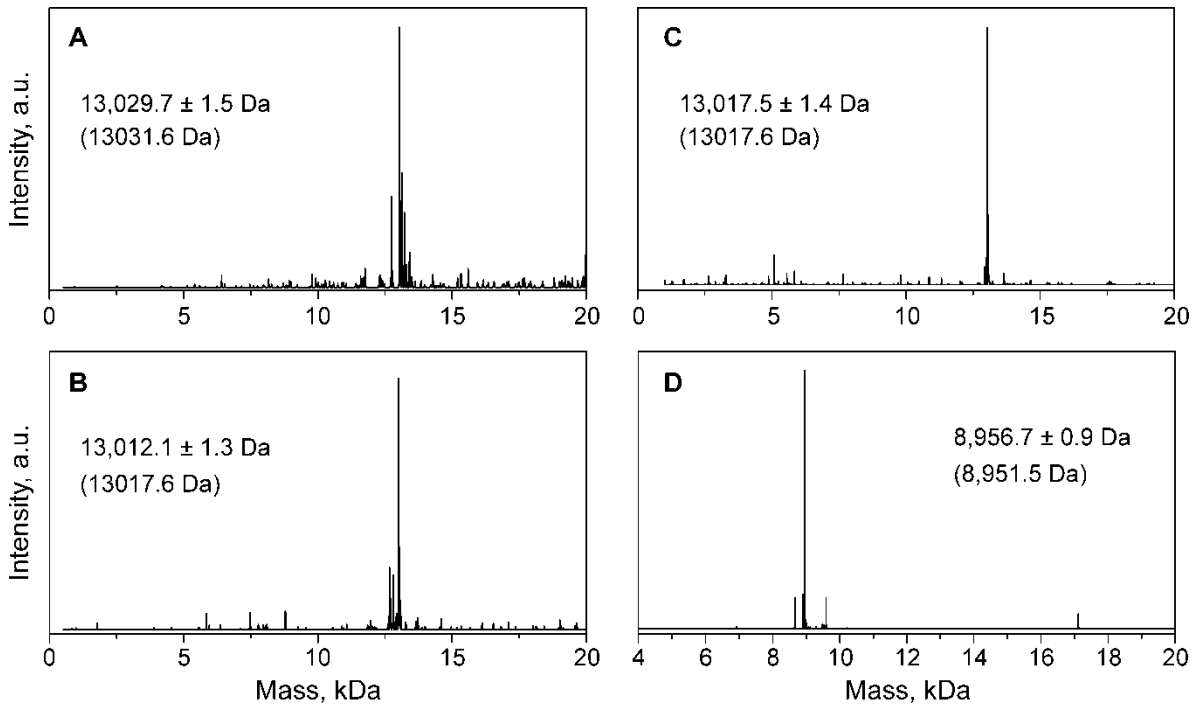

Fig. S1. Deconvoluted electrospray ionization mass spectra of (A) D50C Cc, (B) E66C Cc, (C) E88C Cc, and (D) D32C ubiquitin conjugated to EDTA(Mn^{2+}). The mass of the major peak is indicated in the graphs; the expected, theoretical mass is given in parentheses.

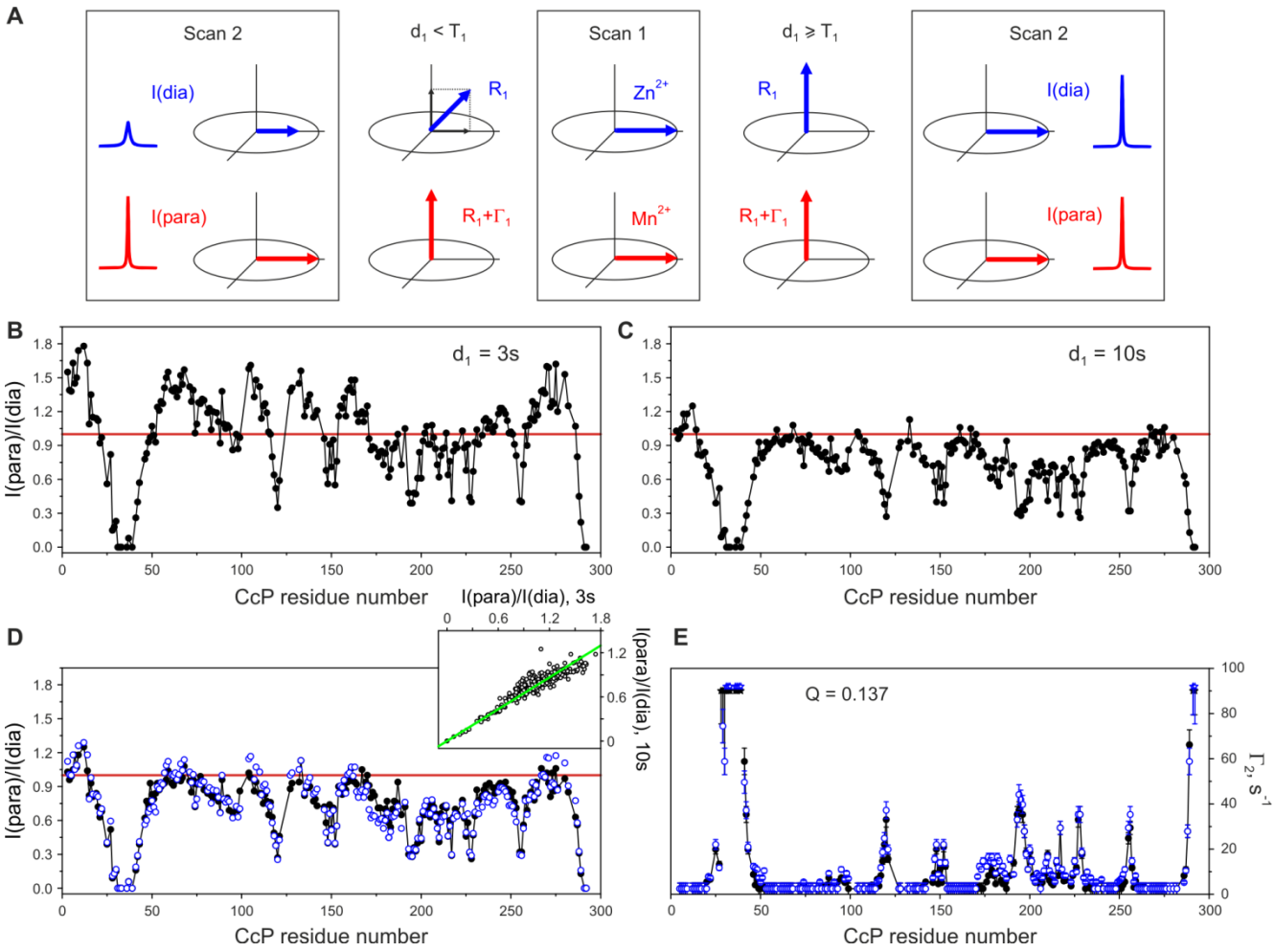

Fig. S2. Intensity-based PRE measurements. (A) Schematic description of the nuclear magnetization during the HSQC experiment. Found in the transverse plane at the end of the first scan (middle, boxed), the magnetization vector returns to its equilibrium, longitudinal position during the repetition delay, d_1. When d_1 is shorter than the amide relaxation time ($d_1 < T_1$, left hand side), the magnetization recovery levels in the paramagnetic (red) and diamagnetic (blue) samples differ, leading to a decreased signal intensity of the latter in the next scan (boxed; also see text). Increasing the repetition delay to $d_1 \geq T_1$ (right hand side) equalizes the recovery levels and yields the full signal for both samples in the subsequent scans. (B)-(C) Ratio of the signal intensities in the HSQC spectra of [D, ^{15}N] CcP(CN) bound to E88C-EDTA Cc^{red} harboring paramagnetic (Mn^{2+}) or diamagnetic (Zn^{2+}) metal ion, recorded with (B) d_1 = 3 s and (C) d_1 = 10 s. The panels (B) and (C) correspond to the left and the right parts of the scheme (A), respectively. (D) Comparison of the signal intensity ratios for the experiments with d_1 = 10 s [black circles, the same as in (C)] and the rescaled d_1 = 3 s dataset (blue circles). The scaling factor is derived from the slope of the linear fit in the corresponding correlation plot (inset; green line with the slope of 1.38 ± 0.14, Pearson r^2 = 0.9). The red lines in (B)-(D) indicate I(para)/I(dia) = 1.0. (E) The Γ_2 values calculated from the intensity ratios in (D) for the d_1 = 10 s (black symbols) and the rescaled d_1 = 3 s datasets (blue symbols). Stars indicate the residues whose resonances disappear in the paramagnetic spectrum. The errors are standard deviations. The comparative Q factor for the two PRE sets is indicated in the plot.

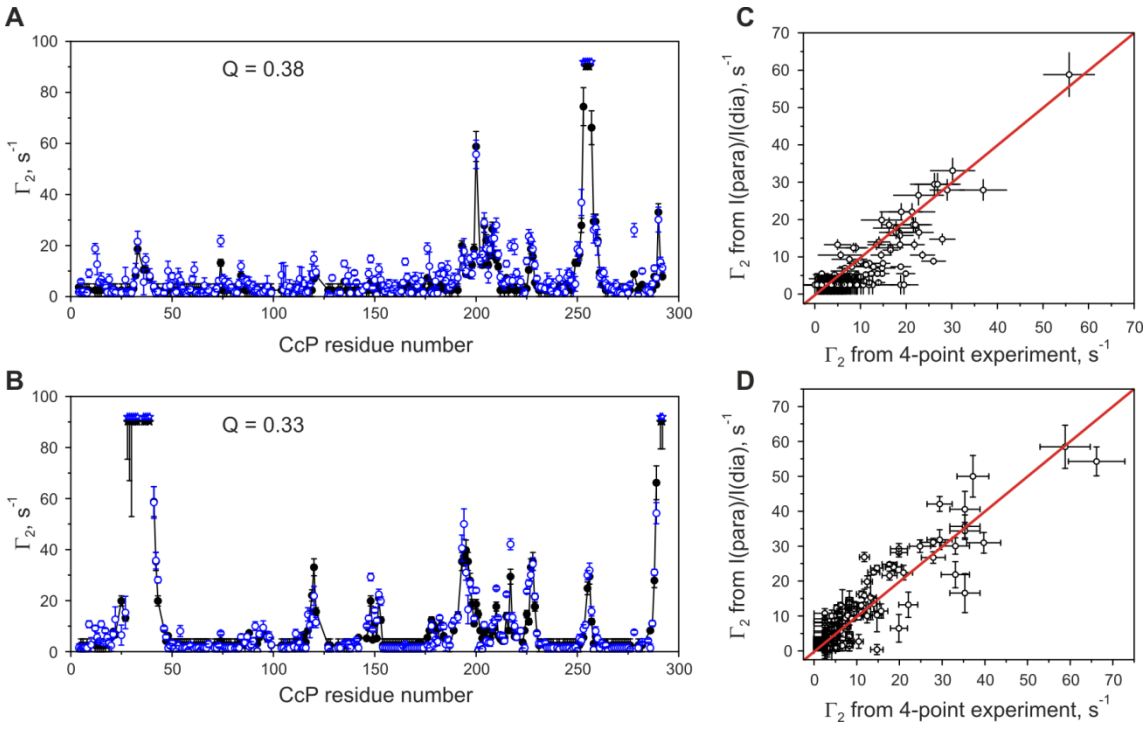

Fig. S3. Comparison of the Γ_2 values for [D, ^{15}N] CcP(CN) in complex with (A) D50C-EDTA(Mn^{2+}) or (B) E88C-EDTA(Mn^{2+}) Cc^{red} calculated from the HSQC intensity ratios (d_1 = 10 s, black symbols) or obtained from the 4-point Γ_2 measurement experiments (blue symbols). Stars indicate the residues whose resonances disappear in the paramagnetic spectrum. The errors are standard deviations. The comparative Q factors are indicated in the plots. (C)-(D) Corresponding correlation plots of the Γ_2 values calculated from the HSQC intensity ratios with d_1 = 10 s (ordinate) and the four-point Γ_2 measurement experiments (abscissa). The red line indicates the perfect match. The samples contained 0.4 mM CcP and 1 molar equivalent of the labeled D50C or E88C Cc in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0 at 303 K.

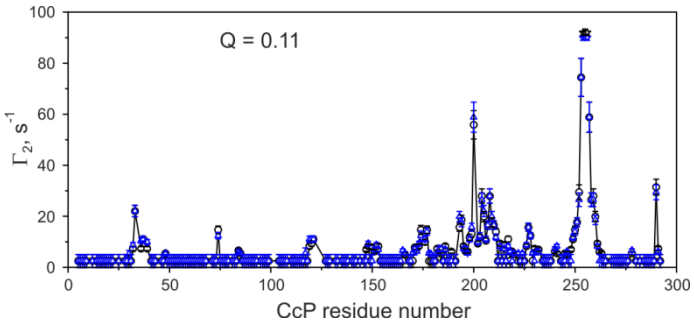

Fig. S4. Reproducibility of the PRE measurements. Comparison of the Γ_2 values for [D, ^{15}N] CcP(CN) in complex with D50C-EDTA(Mn^{2+}) Cc^{red} obtained from the HSQC spectra of two independently labeled and separately prepared sets of samples (black and blue symbols). Stars indicate the residues whose resonances disappear in the paramagnetic spectrum. The errors are standard deviations. The comparative Q factor for the two PRE sets is indicated in the plot. The samples contained 0.4 mM CcP and 1 molar equivalent of the labeled D50C Cc in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0. The spectra were recorded at 303 K with d_1 = 10 s.

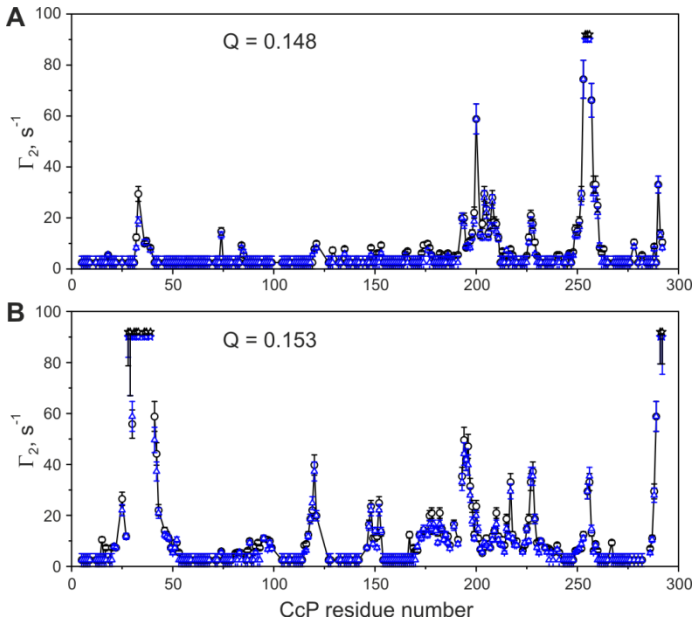

Fig. S5. Validation of the diamagnetic reference. Comparison of the Γ_2 values for [D, ^{15}N] CcP(CN) in complex with (A) D50C-EDTA(Mn^{2+}) or (B) E88C-EDTA(Mn^{2+}) Cc^{red} obtained from the HSQC spectra using the wt Cc (black) or the corresponding EDTA(Zn^{2+}) derivative (blue) as the diamagnetic reference. Stars indicate the residues whose resonances disappear in the paramagnetic spectrum. The errors are standard deviations. The comparative Q factors are indicated in the plots. The samples contained 0.4 mM CcP and 1 molar equivalent of the Cc in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0. The spectra were recorded at 303 K with d_1 = 10 s.

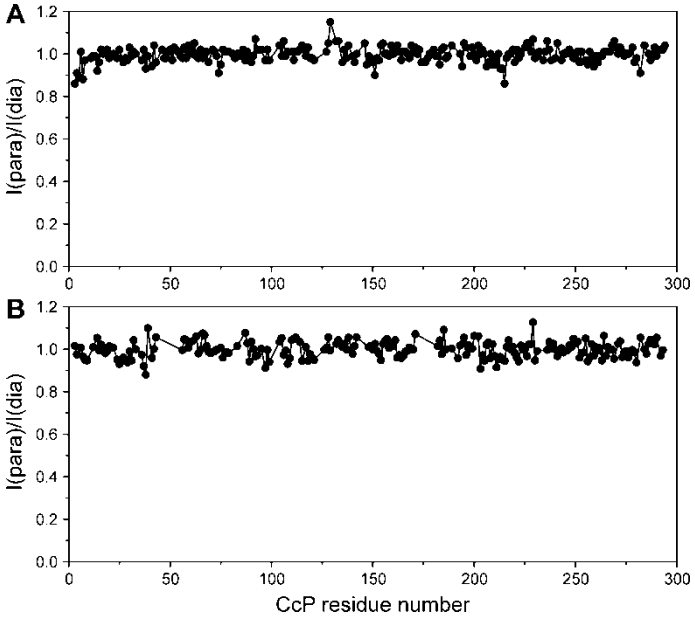

Fig. S6. Ubiquitin-EDTA(Mn^{2+}) control PRE experiments. (A)-(B) Ratio of the signal intensities in the HSQC spectra of (A) [D, ^{15}N] CcP(CN) or (B) [D, ^{15}N] CcP(RS) bound to the wt and D32C-EDTA(Mn^{2+}) ubiqutin. The samples contained 0.4 mM CcP and 1 molar equivalent of the ubiquitin in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0. The spectra were recorded at 303 K with (A) d_1 = 10 s and (B) d_1 = 3 s. The flat profiles with I_{para}/I_{dia} > 0.85 values indicate the absence of any intermolecular PREs.

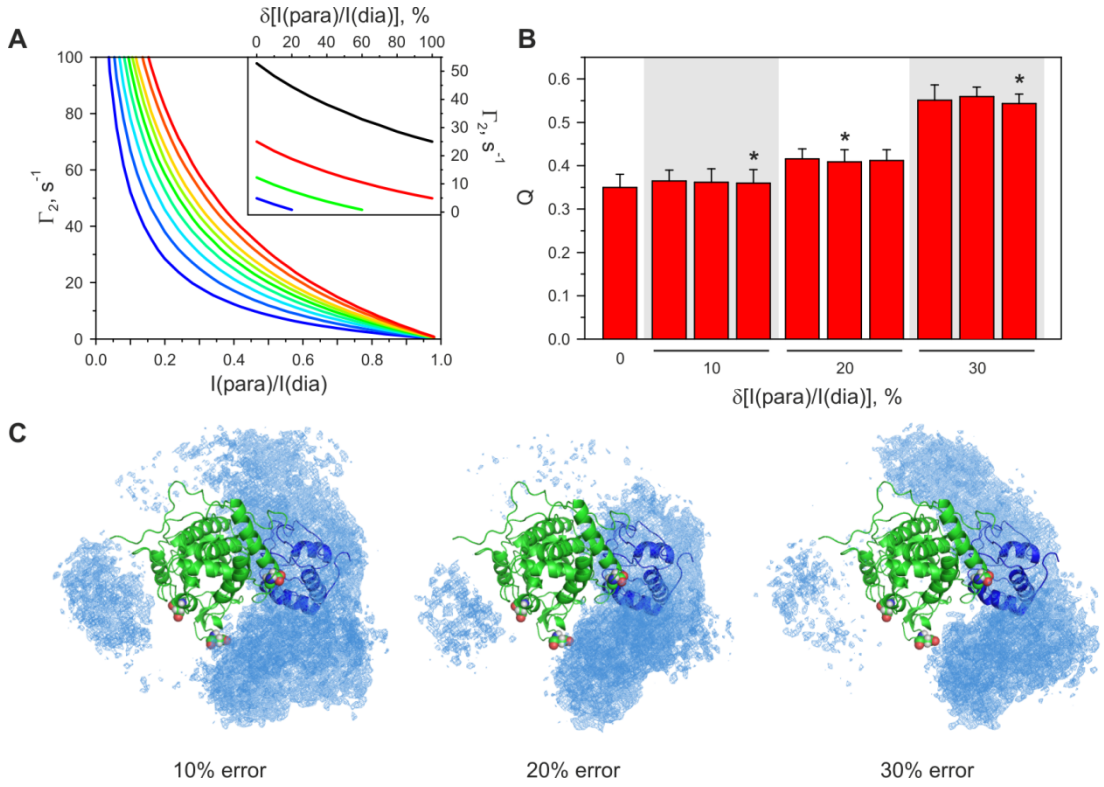

Fig. S7. Error analysis of the ensemble refinement against intermolecular PREs obtained from the single-time-point measurements. (A) Simulated curves for the Γ_2 as a function of I(para)/I(dia) for the HSQC peaks with varying $R_{2,dia}$ (from blue to red, $R_{2,dia}$ = 10, 15, 20, 25, 30, 35, 40, 50, and 60 s^{-1}) calculated from Eq. 2 in the main text. The inset shows changes in the Γ_2 values upon increase in the intensity ratios, δ[I(para)/I(dia)], for the peaks with $R_{2,dia}$ = 25 s^{-1} and the initial values of I(para)/I(dia) = 0.2 (black), 0.4 (red), 0.6 (green), and 0.8 (blue). (B)-(C) Results of the CcP(CN)-Cc^{red} ensemble refinement (p = 0.4, N = 5) against the Γ_2^{obs} obtained from the experimental I(para)/I(dia) dataset (d_1 = 10 s), whose values were randomly increased or decreased by 10, 20, and 30 %. (B) Comparison of the Q factors for the ensembles refined against the original data (δ[I(para)/I(dia)] = 0) and those including 10, 20, and 30 % of uncertainty. In the case of δ[I(para)/I(dia)] > 0, three bars correspond to separate calculations from three independent datasets with randomly introduced errors. The asterisks indicate the calculation runs whose solutions are visualized in the panel (C). (C) Reweighted atomic probability density map for the overall distribution of the Cc molecules obtained from the ensemble refinement against the PRE data with the introduced uncertainties (indicated in the plots). The CcP and Cc in the crystallographic orientation are shown in green and blue cartoons, respectively. The CcP residues D34, D148, and D217 are spacefilled. The proteins are in the same orientation as in Fig. 6A.

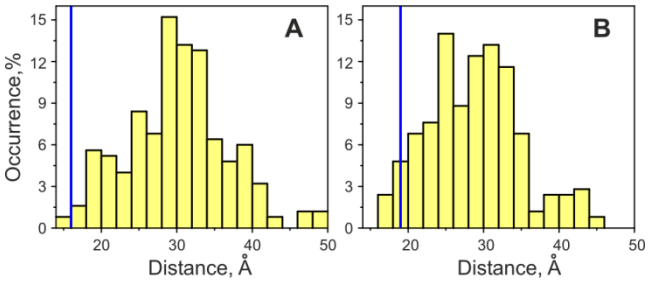

Fig. S8. Intermolecular ET donor-acceptor distances in the CcP(CN)-Cc^{red} encounter state produced in a control refinement run with the van der Waals potential set to zero for all side-chain atoms extending beyond Cβ. The histograms show distributions of the edge-to-edge (A) W191 (CcP) – heme (Cc) and (B) heme-heme distances in the encounter ensemble of the control run (Q = 0.36 ± 0.03). Solid vertical lines mark the corresponding ET distances in the crystallographic orientation.

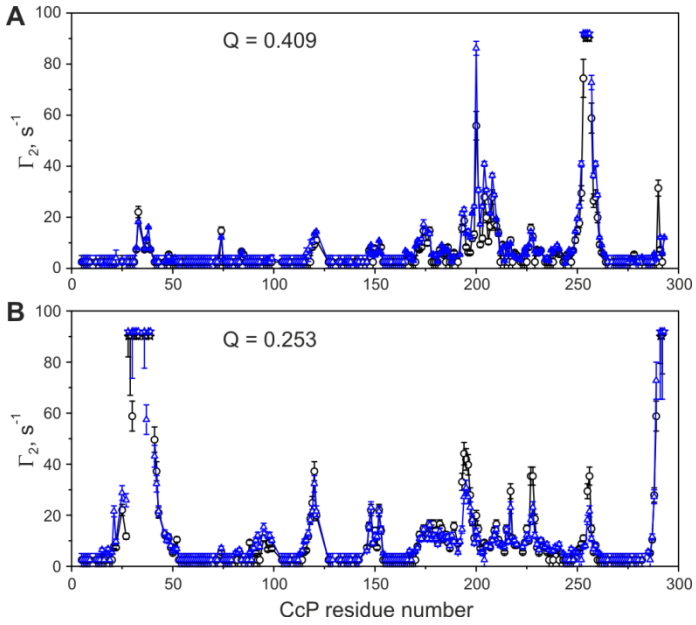

Fig. S9. Effect of the Cc redox state on the intermolecular PREs. Comparison of the Γ_2 values for [D, ^{15}N] CcP(CN) in complex with (A) D50C-EDTA(Mn^{2+}) or (B) E88C-EDTA(Mn^{2+}) Cc^{red} (black) or Cc^{ox} (blue). Stars indicate the residues whose resonances disappear in the paramagnetic spectrum. The errors are standard deviations. The comparative Q factors are indicated in the plots. The samples contained 0.4 mM CcP and 1 molar equivalent of the Cc in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0. The spectra were recorded at 303 K with d_1 = 3 s. Prior to the Γ_2 conversion, the I_{para}/I_{dia} peak intensity ratios were rescaled by a factor derived from the comparison of the CcP(CN)-Cc^{red} data acquired with d_1 = 3 s and 10 s. Note that, while both PRE profiles in (A) show the same pattern, they appear to be uniformly shifted relative to each other along the vertical axis. Likely caused by an imperfect scaling of the two datasets, such systematic shift accounts for an elevated comparative Q factor.

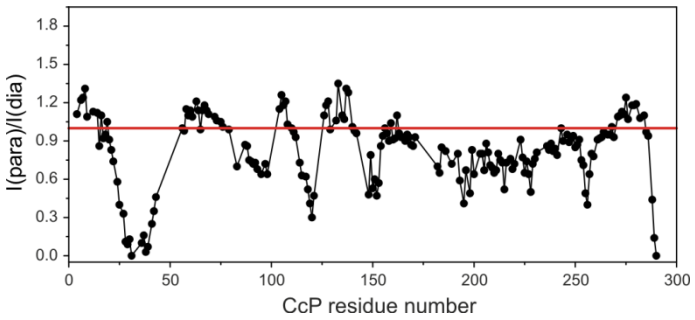

Fig. S10. Ratio of the signal intensities in the HSQC spectra of [D, ^{15}N] CcP(RS) bound to E88C-EDTA Cc^{ox} harboring paramagnetic (Mn^{2+}) or diamagnetic (Zn^{2+}) metal ion. The red line indicates I(para)/I(dia) = 1.0. The samples contained 0.4 mM CcP and 1 molar equivalent of the labeled E88C Cc in 20 mM sodium phosphate, 100 mM NaCl, pH 6.0. The spectra were recorded at 303 K with d_1 = 3 s.

Supplementary Text

Practical aspects of the single time-point HSQC-based PRE measurements

In a typical HSQC experiment, the ^{15}N chemical shift evolution time is progressively incremented, and multiple scans at each increment are acquired to increase the signal-to-noise ratio. These single-scan experiments are separated by a repetition delay (d_1), during which the transverse magnetization – prepared and detected in each scan – relaxes back to its equilibrium, longitudinal position with a rate $R_1 = 1/T_1$, where T_1 is the longitudinal relaxation time. For the 38 kDa perdeuterated protein studied here, the amide ^1H T_1 relaxation times are expected to be very long (> 3 s), which is larger than d_1 = 1-2 s generally used in biomolecular HSQC experiments.[31] As a consequence, at $d_1 < T_1$ the magnetization recovery is incomplete, and the d_1 delay dictates the recovery level, which in turn determines the size of the signal generated in the next scan (Fig. S2A).

For an accurate estimate of the Γ_2 PREs using the HSQC-based, single-time-point approach, the magnetization recovery levels in the paramagnetic and diamagnetic samples should be the same. However, due to the PRE contribution to the R_1 relaxation rate, the signal recovery will be higher in the paramagnetic sample, resulting in increased apparent I_{para}/I_{dia} values (manifested by the profiles shifted upwards of the I_{para}/I_{dia} = 1.0 baseline, Fig. S2A-B), which yield underestimated Γ_2 PREs.[31] This problem can be avoided by substantially increasing the d_1 repetition delay (allowing for a fuller magnetization recovery in the diamagnetic sample) or employing the multiple-time-point Γ_2 measurement scheme pioneered by Kay and co-workers[30] and further developed by Clore's group.[31] Indeed, acquiring HSQC spectra with d_1 = 10 s yields I_{para}/I_{dia} profiles with the ≈ 1.0 baseline (Fig. S2C).

As discussed by Clore and co-workers,[31] the multiple time-point Γ_2 measurement is the method of choice for the quantitative PRE analysis. Unfortunately, because of a lower sensitivity of the multiple-time-point experiment compared to the conventional HSQC (employed here to derive the I_{para}/I_{dia} values) and practical considerations (such as limited stability of the samples encountered before),[12] which prevented acquisition for longer periods of time, we could not obtain high-quality data from the four-point Γ_2 measurements. As can be seen in Fig. S3, the four-point Γ_2 profiles feature a number of values with large errors, miss data for several resonances that were too weak for reliable analysis, and are generally noisy. Still, despite these shortcomings, the PREs obtained in the four-time-point Γ_2 measurements tend to reproduce the Γ_2 values calculated from the intensity ratios in the HSQC spectra with d_1 = 10 s (Fig. S3). Albeit imperfect, the match of the Γ_2 data obtained with the two different schemes reinforces our conclusion that the d_1 repetition delay of 10 s is long enough for the equal magnetization recovery in the paramagnetic and diamagnetic samples (also seen from the I_{para}/I_{dia} profiles with the ≈ 1.0 baseline, e.g. Fig. S2C), which enables the use of the I_{para}/I_{dia} intensity ratios for the quantitative Γ_2 analysis carried out in this work.

To enable a qualitative analysis of the PREs obtained from the HSQC spectra acquired with d_1 = 3 s, the $(I_{para}/I_{dia})_{3s}$ data were divided by the scaling factor estimated from the linear slope of the $(I_{para}/I_{dia})_{10s}$ vs $(I_{para}/I_{dia})_{3s}$ correlation plots (the subscript after the parentheses indicates the d_1 delay) (Fig. S2D). The Γ_2 values obtained from the rescaled $(I_{para}/I_{dia})^*_{3s}$ profiles are in excellent agreement with those calculated from the $(I_{para}/I_{dia})_{10s}$ data (Fig. S2E) and, thus, can be used for the comparison among different systems studied (e.g. CcP-Cc^{ox} vs CcP-Cc^{red} complexes, see below).

Due to the slow electron relaxation time of the high-spin (S = 5/2) heme iron ($\tau_e = 10^{-9} – 10^{-11}$ s^{-1}), the resting-state CcP, CcP(RS), experiences a strong intramolecular PRE.[14] The resulting fast T_1

relaxation leads to the nearly-complete magnetization recovery in both the CcP(RS)-Cc(Mn^{2+}) and CcP(RS)-Cc(Zn^{2+}) samples even at short d_1 delays. Thus, contrary to the situation described above for the weakly paramagnetic, low-spin (S = 1/2) CcP(CN), the HSQC spectra of the CcP(RS) acquired with d_1 = 3 s yield the profiles with little deviation from the I_{para}/I_{dia} = 1.0 baseline (Fig. S10) and can be used for the quantitative Γ_2 analysis.

Error analysis of the ensemble refinement against intermolecular PREs obtained from the single-time-point measurements

In order to assess the performance of the HSQC-based single-time-point PRE measurement scheme employed here for the quantitative structural analysis of the protein encounters, we investigated the impact of the errors in the measured peak intensity ratios on the derived intermolecular PREs and, subsequently, the results of the Cc-CcP ensemble refinement. As can be seen in Fig. S2D, the largest discrepancies between the data acquired with different relaxation delays (d_1 = 3 and 10 s) are exhibited by the residues with large intensity ratios, I(para)/I(dia). Inspection of the curves generated from Eq. 2 shows that the peaks with large I(para)/I(dia) are less sensitive to absolute errors than those with the smaller ratios (Fig. S7A). For example, for a peak with $R_{2,dia}$ = 25 s^{-1}, increasing the intensity ratio from 0.7 to 0.8 translates to decrease in Γ_2 of only 3.3 s^{-1}, while changing it from 0.2 to 0.3 leads to the substantial difference of Γ_2 = 17.2 s^{-1}. Similarly, the signals with larger intensity ratios are also less sensitive to the relative errors (inset in Fig. S7A). For instance, a 20 % increase in I(para)/I(dia) = 0.8 results in Γ_2 decrease of 4.1 s^{-1}, while the same increase in I(para)/I(dia) = 0.2 would diminish the Γ_2 by 8.3 s^{-1}.

To investigate the effect of the errors in the measured I(para)/I(dia) on the outcome of the ensemble refinement, we have 1) generated the datasets with the original, experimental peak intensity ratios randomly increased or decreased by 10, 20, and 30 %; 2) converted those into the Γ_2 values; and 3) performed ensemble refinement with (p, N) = (0.4, 5) in the same way as was done before with the original data. To assess the reproducibility, the entire procedure (i.e. steps 1-3) was repeated three times. To quantify the agreement with the experimental data, the back-calculated PREs for each generated ensemble were compared to the original Γ_2^{obs} values by calculating the Q factor. As can be seen in Fig. S7B, there is a gradual increase in the Q factors with the increasing error. Compared to the Q = 0.35 ± 0.03 for the ensembles refined with the original Γ_2^{obs} data, the Q factors for the runs with 10 % error increase only slightly (average Q = 0.36 ± 0.03), followed by the larger Q = 0.41 ± 0.03 for the data with 20 % errors, finally reaching the value of Q = 0.55 ± 0.03 for the 30 % error dataset. Taken at the face value, these results indicate that the errors of 10-20 % in the HSQC peak intensities are well tolerated and have little effect on the outcome of the ensemble refinement.

This conclusion is further confirmed by inspection of the corresponding probability density maps (Fig. S7C), showing that the spatial distributions of the generated ensembles closely resemble that of the solutions obtained with the original PRE data (Fig. 6A). Although the exact extent of the conformational space occupied by the interacting molecules in the original and control simulations is not the same, the location and the overall distribution of protein encounters remain similar even for the data with the 30 % error. Taken together, these findings suggest that the ensemble refinement procedure used in this work is relatively insensitive to the specific Γ_2 values used in the calculations and, instead, is driven by the set of intermolecular PREs as a whole.