

## Supporting Information

### Heterogeneous and Highly Dynamic Interface in Plastocyanin-Cytochrome *f* Complex Revealed by Site-Specific 2D IR Spectroscopy

Sashary Ramos, Amanda L. Le Sueur, Rachel E. Horness, Jonathan T. Specker, Jessica A. Collins, Katherine E. Thibodeau, Megan C. Thielges\*

*Department of Chemistry, Indiana University, 800 E. Kirkwood, Bloomington, Indiana, 47405, United States*

\*Email: thielges@indiana.edu

#### Table of Contents

<b>I. Experimental Methods</b>	<b>S2</b>
Expression and purification of plastocyanin (Pc) variants and cytochrome <i>f</i> (cyt <i>f</i> )	S2
Characterization of Pc variants	S3
Sample preparation for IR spectroscopy	S4
FT IR Spectroscopy and analysis	S5
2D IR spectroscopy and data analysis	S5
<b>II. Supplemental Figures</b>	<b>S8</b>
Figure S1. Mass spectra of Pc variants	S8
Figure S2. Circular dichroism spectra of Pc variants	S8
Figure S3. UV/Vis spectra of Pc variants	S9
Figure S4. UV/visible spectrum of cyt <i>f</i>	S9
Figure S5. Fluorescence binding curves of Pc variants to cyt <i>f</i>	S10
Figure S6. Representative 2D spectra	S10
<b>III. References</b>	<b>S11</b>

## I. Experimental Methods

### *Expression and purification of plastocyanin (Pc) variants and cytochrome f (cyt f)*

Plasmid construction, expression, and purification of plastocyanin (Pc) was performed as previously reported with modifications made for expression of variants with selectively introduced *p*-cyanophenylalanine (CNF).<sup>1</sup> The plasmid pEAP containing the *petE* gene encoding for Pc from *Nostoc* PCC7119 constructed by Dr. Miguel A. De la Rosa (University of Seville)<sup>2</sup> was provided by Dr. Marcellus Ubbink (Leiden University). The amino acid numbering for the modified Pc follows that of PDB entry 1TU2 and excludes the initial methionine. Standard site-directed mutagenesis (QuikChange II Site-Directed Mutagenesis Kit, Agilent Technologies) was used to introduce an amber (stop) codon (TAG) at positions 36 (CNF36), 88 (CNF88) and 90 (CNF90) in the wild type (WT) sequence to direct incorporation of CNF via amber suppression. Mutagenesis was successful for 10 other Pc positions (3, 16, 31, 54, 60, 63, 64, 67, 70, and 85), however initial expression tests resulted in either no or misfolded protein.

The WT and variant Pc were expressed in BL21(DE3) in LB at 37 °C with 250 rpm shaking for 20 hours. All plasmids were co-expressed with a plasmid containing the gene for methionine aminopeptidase, which we empirically found improved expression yields. To introduce CNF, the plasmids were co-expressed with pULTRA-CNF, provided by Dr. Peter Schultz (The Scripps Research Institute), a plasmid encoding for an orthogonal tRNA/aminoacyl-tRNA synthetase pair evolved for *in vivo* CNF incorporation in response to an amber (stop) codon.<sup>3</sup> CNF was added to the culture media at 1 mM prior to induction. All cultures were grown until OD<sub>600</sub> reached 0.6, then protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation at 6400 x g for 20 min. The pellet was resuspended in 1 mM sodium phosphate (NaP), pH 6.8, 1 mM PMSF, 250 μM CuSO<sub>4</sub>, and supplemented with 0.2 mg/mL DNase I. The cells were lysed by sonication and centrifuged at 12,000 x g for 50 min. The supernatant was dialyzed in 1 mM NaP, pH 6.8, 250 μM CuSO<sub>4</sub>, clarified by centrifugation at 12,000 x g for 40 minutes, then applied to a S cellulose cation exchange column (Bio-Rad), equilibrated with 1 mM NaP, pH 6.8, and eluted using a linear gradient of 1–200 mM NaP, pH 6.8. Fractions containing Pc were collected, concentrated, and further purified by size exclusion chromatography (Sephacryl S-100 HR, GE Healthcare Life Sciences, 125mL). The concentration of the purified Pc ( $A_{280/597} < 1.5$ ) was determined by visible spectroscopy ( $\epsilon_{597} = 4.5 \text{ mM}^{-1} \text{ cm}^{-1}$ ).<sup>4</sup> Protein yields for CNF36, CNF88, and CNF90 Pc were 3.2 mg/L, 2.4 mg/L, and 3.1 mg/L of

culture, respectively. Purified Pc was used immediately or brought to 25% glycerol, flash frozen in liquid N<sub>2</sub>, and stored at -80°C. UV/visible spectroscopy was used to determine protein concentration ( $\epsilon_{597} = 4.5 \text{ mM}^{-1} \text{ cm}^{-1}$ )<sup>4</sup> and monitor protein integrity before and after 2D IR measurements.

The plasmid (pEAF-WT) containing the gene coding for the soluble domain of cytochrome *f* (*cyt f*) *Nostoc* sp. PCC7119 constructed by Dr. Miguel A. De la Rosa (University of Seville) was generously provided by Dr. Marcellus Ubbink (Leiden University). *Cyt f* was expressed and purified as previously reported with minor modifications.<sup>5</sup> Briefly, *cyt f* was expressed in DH5- $\alpha$  in LB supplemented with ampicillin (100  $\mu\text{g/mL}$ ), chloramphenicol (20  $\mu\text{g/mL}$ ), and ammonium iron (III) citrate (6  $\mu\text{g/L}$ ) at 25°C with 150 rpm shaking for 40 hours. Expression was induced by addition of IPTG when OD<sub>600</sub> was equal to 0.6, at which point the cultures were brought to 0.5 mM  $\delta$ -aminolevulinic acid. The cells were harvested by centrifugation at 6400 x g for 20 minutes and resuspended in 10 mM Tris-Cl, pH 8.2, at which point PMSF was added to a final concentration of 1 mM. Cells were lysed using three freeze-thaw cycles and the lysate was centrifuged at 12,000 x g for 50 minutes. The supernatant was treated with benzonase (2 U/mL) and then dialyzed into 10 mM Tris-Cl, pH 8.2. *Cyt f* was purified using a 0 – 500 M NaCl gradient elution from a DEAE column (Bio-Rad) in 10 mM Tris-Cl, pH 8.2 followed by a Q column (Bio-Rad) in 5 mM MES buffer, pH 6. Additional purification was performed by passage over a Sephacryl S-100 HR gel filtration column (GE Life Sciences, 125 mL). Purified *cyt f* ( $A_{280/556} < 1.5$ ) was concentrated and used immediately or brought to 25% glycerol, flash-frozen in liquid N<sub>2</sub> and stored at -80°C. The protein yield was approximately 0.4 mg/L of culture. Protein concentration was determined by UV/Visible spectroscopy ( $\epsilon_{556} = 31.5 \text{ mM}^{-1} \text{ cm}^{-1}$ ).<sup>6</sup>

### ***Characterization of Pc variants***

Pc variants were characterized by mass spectrometry with a Waters LCT Classic electrospray/time-of-flight LC-MS system. All variants exhibited the mass/charge as predicted. For CNF36Pc, CNF88Pc, and CNF90Pc, respectively: 11318 *m/z* (calculated: 11450.0 *m/z* – Met1 = 11318.8 *m/z*), 11254 *m/z* (calculated: 11385.2 *m/z* – Met1 = 11254 *m/z*), 11289 *m/z* (calculated: 11420 *m/z* – Met1 = 11288.8 *m/z*) (Figure S1). The total percentage of CNF incorporation was ~70%, ~65%, and ~75% for CNF36Pc, CNF88Pc, and CNF90Pc respectively (Figure S1), with the native residue found otherwise. The native residue incorporation does not interfere with

spectroscopic measurements but does lead to overall reduced concentration of the *CNF* probe compared to total protein concentration.

All proteins were characterized by circular dichroism spectroscopy on a Jasco J-715 spectrometer. The spectra (Figure S2) indicate no perturbation to the secondary structure by *CNF* incorporation. In addition, each protein was characterized by UV/visible spectroscopy on an Agilent Cary 300 UV/visible spectrometer (Figure S3). All spectra show a band at 600 nm associated with the charge transfer band of the Cu site, indicating no perturbation to the Cu coordination from *CNF* incorporation.

The binding constants for each of the Pc variants to *cyt f* were determined via fluorescence spectroscopy of the sole intrinsic tryptophan residue (W4) of *cyt f* located on the binding ridge and heme face of the protein. The experiment was performed with an ISS PC1 photon counting spectrophotometer with excitation and emission slits set to 0.5 mm, an excitation wavelength of 295 nm and emission wavelength monitored at 307-370 nm. Aliquots of Pc were titrated into 400 nM of *cyt f* and the decrease in fluorescence was monitored as a function of the Pc concentration. Sodium ascorbate was used to reduce the Pc copper center to eliminate effects from ET between redox centers. Data was fit to a standard two-state binding model to determine the  $K_D$  (Figure S5). Values were calculated as  $72.9 \pm 8.4 \mu\text{M}$ ,  $48.9 \pm 9.6 \mu\text{M}$ ,  $9.6 \pm 5.5 \mu\text{M}$  and  $1.2 \pm 1.1 \mu\text{M}$  for WT, *CNF36*, *CNF88*, and *CNF90* respectively.

#### ***Sample preparation for IR spectroscopy***

For FT IR spectroscopy, the unligated Pc variants were dialyzed into 1 mM NaP, pH 6.8 and concentrated to 2 mM, samples of the Pc-*cyt f* complex were prepared at 1 mM Pc and 1.5 mM *cyt f* for *CNF36* and *CNF88* or 0.5 mM Pc and 0.5 mM *cyt f* for *CNF90*. A volume of 5  $\mu\text{L}$  of unligated Pc or the Pc-*cyt f* complex was then sealed between two  $\text{CaF}_2$  windows separated by a 38.1  $\mu\text{m}$  Teflon spacer.

For 2D experiments, the Pc variants were dialyzed into 1 mM NaP, pH 6.8 and concentrated by filtration to 3-5 mM. Based on mass spectrometry, the samples were calculated to be 2.8 mM, 3.1 mM, and 2.4 mM with respect to the *CNF* probe for the *CNF36*, *CNF88*, and *CNF90* variants, respectively. Samples of the Pc-*cyt f* complex were prepared at 1.5 mM Pc and 2.2 mM *cyt f* for *CNF36*, 1.5 mM Pc and 1.6 mM *cyt f* for *CNF88*, and 1.7 mM Pc and 1.6 mM *cyt f* for *CNF90*. Given the measured  $K_D$  values, all samples were calculated to be at least 94% bound. The samples

were then sealed between two CaF<sub>2</sub> windows separated by a 76 (CNF36, CNF88) or 127 (CNF90)  $\mu\text{m}$  Teflon spacer.

### ***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 detector (MCT) at 4  $\text{cm}^{-1}$  resolution. WT Pc was used to generate a reference spectrum for the unligated samples, while the WT Pc-cyt *f* complex was used to generate a reference spectrum for the complex. For both reference and sample, 10,000 scans were averaged after purging the chamber with dry nitrogen for 30 min. Experiments were performed using a bandpass filter (Thorlabs FB4500-500) with a center frequency of 4500 nm and full width half max of 500 nm. All spectra were processed using a 4-term Blackman-Harris apodization function, zero-filling factor of 8, and Mertz phase correction algorithm. A residual slowly varying baseline in the absorption spectra was removed by fitting a polynomial to a spectral region of  $\sim 200 \text{ cm}^{-1}$  excluding the CN absorption bands (Matlab 7.8.0). The absorption bands of the baseline-corrected spectra were then fit to Gaussian functions to determine the center frequencies and line widths. All experiments were performed in at least triplicate with independently prepared protein samples; averages and standard deviations from the fits are given in Table 1 of the main text.

### ***2D IR spectroscopy and data analysis.***

2D IR spectra were obtained as previously described in the literature.<sup>7,8</sup> A Ti:Sapphire oscillator/regenerative amplifier (Spectra Physics) producing  $\sim 75$  fs pulses centered at 800 nm with 1 kHz repetition rate was used to pump a home-built optical parametric amplifier to generate  $\sim 170$  fs pulses centered at  $2225 \text{ cm}^{-1}$  ( $80 \text{ cm}^{-1}$  FWHM), which encompassed both the 0-1 and 1-2 vibrational transitions of the CNF probe. The beam was split into three beams of approximately equal energy, with total energy of 18  $\mu\text{J}$  for all three beams. The three beams were focused into the sample in a 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-correlation using a non-resonant signal from CCl<sub>4</sub>. Chirp was determined by frequency-resolved optical grating experiments and corrected to less than  $\pm 0.02 \text{ fs/cm}^{-1}$  by placement of germanium and CaF<sub>2</sub> 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 is

heterodyne-detected by overlapping with a fourth beam, the local oscillator, to provide amplification and phase information. The combined beam is dispersed by a spectrograph onto a 32-element MCT detector. The generation of a single 2D spectrum involves scanning the time between the first two pulses ( $\tau$ ) with the time period between the second and third pulses ( $T_w$ ) held constant. The heterodyned third-order signal generated at a time  $\leq \tau$  after the application of the third pulse produces an interferogram along  $\tau$ , which is Fourier transformed to produce the  $\omega_1$  axis. Frequency-resolved detection with the MCT array generates the  $\omega_m$  axis of the 2D spectrum. A reference beam, split from the local oscillator, is 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  $T_w$ . Data was smoothed along the  $\omega_m$  axis using a cubic smooth spline function and a smoothing factor of 0.01. The 2D IR spectra were corrected for phase errors and inner filter effects as described previously.<sup>7,9</sup> Phase correction was performed on the basis of the projection slice theorem using the pump probe spectra of the amino acid, CNF, to model those for the protein samples. Briefly, small adjustments to pulse timing and chirp correction were introduced to match the projection of the 2D spectrum onto the  $\omega_m$  axis with the modeled pump probe spectrum. A representative 2D spectrum collected at  $T_w$ s of 0.25 ps and 3 ps with an expanded frequency range to show relative noise compared to the actual peak is shown in Figure S6.

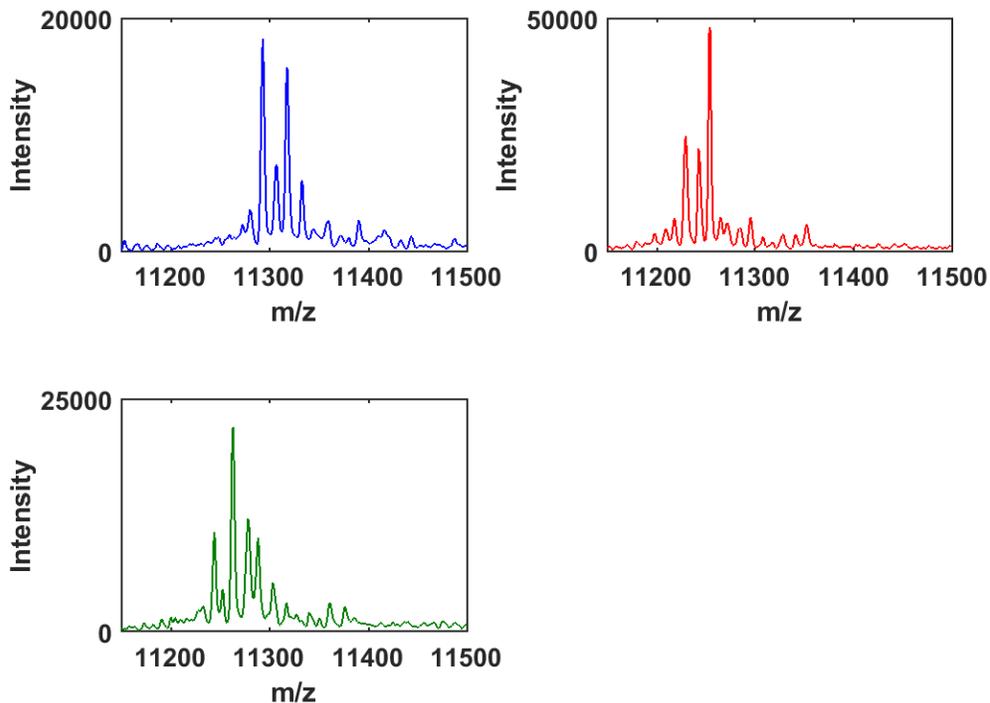
The normalized inhomogeneous part of the frequency-frequency correlation function (FFCF) was extracted from the  $T_w$ -dependent changes in 2D spectral line shapes via the center line slope (CLS) method.<sup>10</sup> Briefly, 1D slices along the  $\omega_\tau$  axis at each  $\omega_m$  frequency were fit to Gaussian functions to determine the  $\omega_\tau$  of maximum absorbance. These  $\omega_\tau$  were plotted as a function of  $\omega_m$  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 is equal to the normalized FFCF in the absence of homogenous broadening.<sup>10</sup> The complete FFCF including frequency fluctuation amplitudes is obtained by simultaneous fitting of the CLS decay and linear IR spectrum. As described in the main text, the FFCFs were analyzed according to the Kubo model<sup>11</sup> using the equation

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

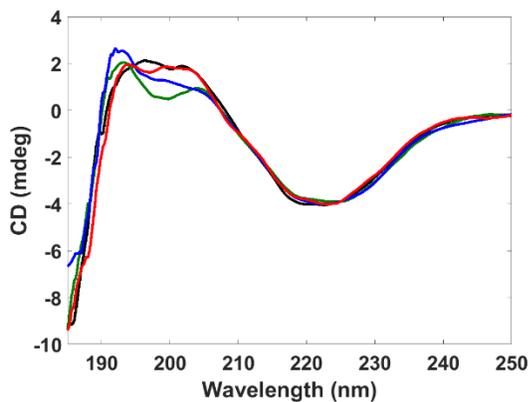
The latter two terms describe the dynamics among the inhomogeneous distribution of frequencies underlying the absorption bands. The inhomogeneous dynamics are separated into two timescales, where  $\Delta_1$  is the frequency fluctuation amplitude sampled on the faster timescale  $\tau_1$ , and the static

term  $\Delta_s$ , 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 previously measured to be 4.2 ps.<sup>1</sup> 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,  $\Gamma^* = 1/\pi T_2^*$ .

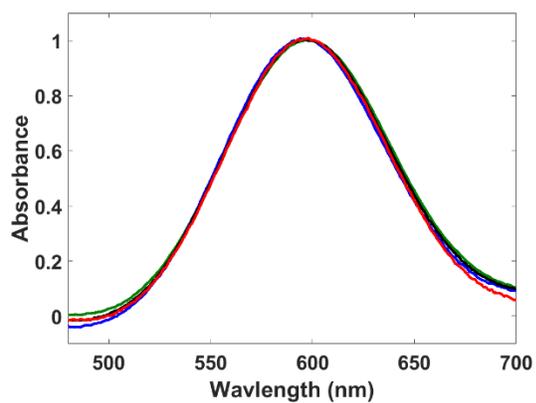
## II. Supplemental Figures



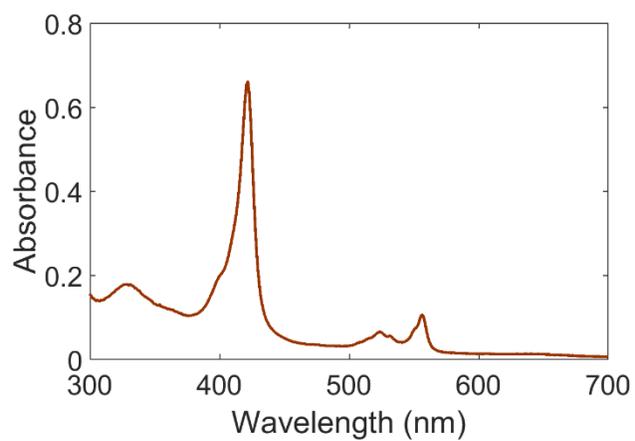
**Figure S1.** Mass spectra of Pc variants: *CNF36* (blue), experimental: 11318  $m/z$ , calculated: 11318.8  $m/z$ ; *CNF88* (red), experimental: 11254  $m/z$ , calculated: 11254  $m/z$ ; *CNF90* (green), experimental: 11289  $m/z$ , calculated: 11288.8  $m/z$ .



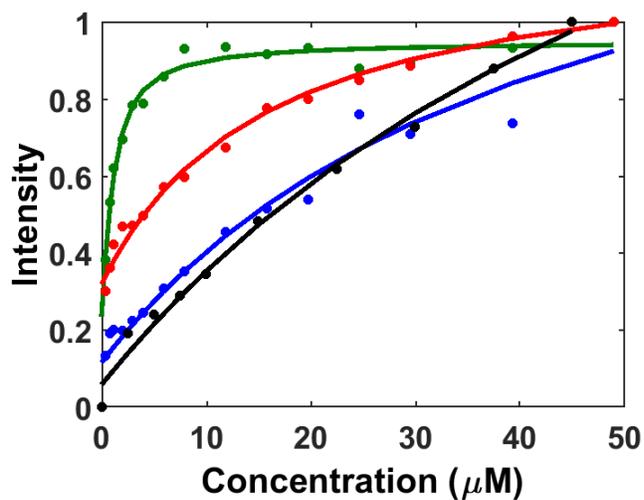
**Figure S2.** Circular dichroism spectra of *CNF36* (blue), *CNF88* (red), *CNF90* (green), and WT Pc (black). The spectra have been normalized to the ellipticity at 225 nm.



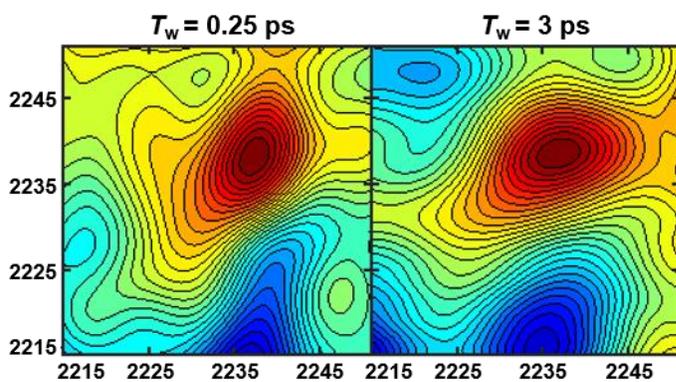
**Figure S3.** Normalized UV/visible absorbance spectra showing charge-transfer transition of the Cu site for *CNF36* (blue), *CNF88* (red), *CNF90* (green), and WT Pc (black).



**Figure S4.** UV/visible absorbance spectrum of *cyt f*.



**Figure S5.** Fluorescence-based assay of binding cyt *f* for WT (black), CNF36 (blue), CNF88 (red), and CNF90 (green) *Pc*.



**Figure S6.** Representative 2D spectra at 0.25 ps and 3 ps with an expanded frequency range to show the relative noise to peak ratio. The data shown is of CNF88 when bound to cyt *f*.

### III. References

1. Le Sueur, A. L.; Ramos, S.; Ellefsen, J. D.; Cook, S. P.; Thielges, M. C., Evaluation of *p*-(<sup>13</sup>C<sup>15</sup>N-cyano)Phenylalanine as an Extended Timescale 2D IR Probe of Proteins. *Anal. Chem.* **2017**, *89*, 5254-5260.
2. Molina-Heredia, F. P.; Hervas, M.; Navarro, J. A.; De la Rosa, M. A., Cloning and Correct Expression in Escherichia Coli of the *petE* and *petJ* Genes Respectively Encoding Plastocyanin and Cytochrome *c*<sub>6</sub> from the Cyanobacterium *Anabaena* sp. PCC 7119. *Biochem. Biophys. Res. Commun.* **1998**, *243*, 302-306.
3. Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G.; A Genetically Encoded Infrared Probe. *J. Am. Chem. Soc.* **2006**, *128*, 13984-13985.
4. Solomon, E. I.; Hare, J. W.; Gray, H. B., Spectroscopic Studies and a Structural Model for Blue Copper Centers in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 1389-1393.
5. Sueur, A. L. L.; Schaugaard, R. N.; Baik, M.-H.; Thielges, M. C., Methionine Ligand Interaction in a Blue Copper Protein Characterized by Site-Selective Infrared Spectroscopy. *J. Am. Chem. Soc.* **2016**, *138*, 7187-7193.
6. Albarran, C.; Navarro, J. A.; Molina-Heredia, F. P.; Murdoch, P. S.; De la Rosa, M. A.; Hervas, M., Laser Flash-Induced Kinetic Analysis of Cytochrome *f* Oxidation by Wild-Type and Mutant Plastocyanin from the Cyanobacterium *Nostoc* sp. PCC 7119. *Biochemistry* **2005**, *44*, 11601 - 11607.
7. Park, S.; Kwak, K.; Fayer, M. D., Ultrafast 2D-IR Vibrational Echo Spectroscopy: A Probe of Molecular Dynamics. *Laser Phys. Lett.* **2007**, *4*, 704-718.
8. Basom, E. J.; Spearman, J. W.; Thielges, M. C., Conformational Landscape and the Selectivity of Cytochrome P450cam. *J. Phys. Chem. B* **2015**, *119*, 6620-6627.
9. Asbury, J. B.; Steinel, T.; Kwak, K.; Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L.; Fayer, M. D., Dynamics of Water Probed with Vibrational Echo Correlation Spectroscopy. *J. Chem. Phys.* **2004**, *121*, 12431-12446.
10. Kwak, K.; Park, S.; Finkelstein, I. J.; Fayer, M. D., Frequency-Frequency Correlation Functions and Apodization in Two-Dimensional Infrared Vibrational Echo Spectroscopy: A New Approach. *J. Chem. Phys.* **2007**, *127*, 124503.
11. Kubo, R. In *Advances in Chemical Physics: Stochastic Processes in Chemical Physics*; Schuler, K. E.; John Wiley & Sons: New York, 1969; Vol. 15, pp 101-127.