

**Supporting Information:**

**Structural adaptability facilitates histidine heme ligation in a  
cytochrome P450**

John A. McIntosh<sup>‡</sup>, Thomas Heel<sup>‡</sup>, Andrew R. Buller, Linda Chio, Frances H. Arnold\*

Division of Chemistry and Chemical Engineering, California Institute of Technology,  
1200 E. California Blvd, Pasadena, CA, 91125

\*Correspondence addressed to [frances@cheme.caltech.edu](mailto:frances@cheme.caltech.edu)

<sup>‡</sup>These authors contributed equally

## Table of content

<b>General</b>	<b>S2</b>
<b>Protein expression</b>	<b>S2</b>
<b>Protein purification</b>	<b>S3</b>
<b>Protein crystallography</b>	<b>S4</b>
<b>X-ray data collection and protein structure determination</b>	<b>S4</b>
<b>Thermostability measurements</b>	<b>S5</b>
<b>Spectral characterization</b>	<b>S6</b>
<b>Heme quantitation</b>	<b>S6</b>
<b>Supporting References</b>	<b>S45</b>

### General

All gene constructs were obtained from codon-optimized gBlocks<sup>®</sup> (IDT, San Diego, CA). Site-directed mutagenesis was accomplished by Quick-change-style mutagenesis using primers bearing desired mutations (IDT, San Diego, CA). Linear DNA was then cloned into pET22b cut with NdeI/XhoI restriction enzymes using Gibson cloning using the previously reported mixture of Taq ligase, T7 exonuclease, and Phusion polymerase.<sup>1</sup> All enzymes used in cloning procedures were purchased from New England Biolabs. Primer and plasmid sequences available upon request. Hyper Broth<sup>™</sup> was obtained from AthenaES<sup>™</sup> and used according to manufacturer's recommendations.

### Protein expression

CYP119 variants was expressed in BL21(DE3). Seed cultures of 2XYT-amp (50 mL, 100  $\mu$ g/mL ampicillin) were inoculated from glycerol stocks and grown overnight (200 RPM, 30 °C) in Erlenmeyer flasks (125 mL capacity). The resulting cultures were used to inoculate 1 L of Hyper Broth<sup>™</sup> supplemented with ampicillin (100  $\mu$ g/mL) in Fernbach flasks (2.8 L capacity). After inoculation, cultures were grown at 37 °C and 180 RPM for 3.5 hours, then cooled on ice for 10-15 minutes and then induced via addition of IPTG (0.25 mM final concentration) and  $\delta$ -

aminolevulinic acid (0.5 mM final concentration). Induced cultures were then grown overnight at reduced temperature and agitation rate (140 RPM, 25 °C). Following expression, cells were pelleted and frozen at -20 °C until purification.

### **Protein purification**

For purification, frozen cell pellets were resuspended (4 mL/g wet cell weight) in lysis buffer (25 mM tris, 100 mM NaCl, 30 mM imidazole, lysozyme (0.5 mg/mL), DNaseI (0.02 mg/mL), hemin (1 mg/g wet cell weight), pH 7.5). Cells were disrupted by sonication (3 min, output control 1.5, duty cycle: 1 sec on / 1 sec off; Sonicator 3000, Misonix, Inc.). The cell suspension was subsequently incubated for 30 min at 65 °C to precipitate *E. coli* proteins. To pellet insoluble cell debris, lysates were centrifuged (20,000 x *g* for 30 min at 4 °C). Cleared lysates were then loaded onto Ni-NTA columns (5 mL size, HP resin, GE Healthcare) using an AKTExpress purifier FPLC system (GE healthcare). Target proteins were eluted using a linear gradient from 100% buffer A (25 mM tris-HCl, 100 mM NaCl, 10 mM, pH 7.5), 0% buffer B (25 mM tris, 100 mM NaCl, 300 mM imidazole pH 7.5) to 100% buffer B over 10 column volumes.

Proteins used for crystallography were subjected to an additional ion-exchange purification step. For these proteins, fractions were pooled, concentrated and exchanged into anion exchange buffer (25 mM tris-HCl buffer pH 7.5) and loaded onto an anion exchange Q Sepharose column (HiTrap™ Q HP, GE Healthcare) using an AKTExpress purifier FPLC system (GE healthcare). The enzyme was eluted from the Q-column by running a gradient from 0 to 0.5 M NaCl over 10 column volumes. Fractions containing the respective enzyme were pooled, concentrated and exchanged into storage buffer (25 mM tris, pH 7.5). Subsequently, the concentrated protein was aliquoted, flash-frozen on powdered dry ice, and stored at -80 °C. Enzyme concentrations were determined via molecular extinction coefficient from frozen aliquots.

Proteins used for other purposes were eluted from the Ni-NTA column as described above, and then pooled, concentrated to 1 mL, and subjected to three 10-fold dilution and concentration steps using centrifugal spin filters (Vivaspin 20, 30 kDA molecular weight cut-off, GE healthcare), each time diluting into fresh buffer (0.1 M KPi pH 8.0).

## Protein crystallography

CYP119 T213A/C317H was crystalized by vapor diffusion. A 1:1 mixture of protein stock (25 mg/ml enzyme in 50 mM KPi pH 8.0 I + 1 mM 4-phenylimidazole) and mother liquor was combined in 24-well sitting-drop plates (Hampton Research). The crystals were grown at room temperature over a span of 4–10 days. CYP119 T213A/C317H crystals formed in 0.1 M tris-HCl pH 7.25, 5 % PEG 4000.

## X-ray data collection and protein structure determination

Crystals of T213A/C317H grew in 7 days and were cryoprotected with an additional 20 % glycerol before diffraction experiments at the Stanford Synchrotron Light Source. Despite extensive effort, only a single crystal was found to diffract to  $< 3.0 \text{ \AA}$ . Data indexed as spacegroup  $P2_1$  and were integrated using XDS.<sup>2</sup> Images were then scaled using AIMLESS to  $2.7 \text{ \AA}$  resolution using a cutoff of  $CC1/2 > 0.3$ .<sup>3,4</sup>

The structure was determined by molecular replacement using PHASER with the 4-phenylimidazole-bound structure of CYP119 (PDB ID: 1F4T) with all ligands removed as the search model.<sup>5</sup> From the output of this run, we identified and removed the ~110 residues of the initial model that had no density in the T213A/C317H structure. Molecular replacement was repeated with this trimmed model, which resulted in substantially improved initial maps. Model building was performed in Coot with iterative rounds of refinement in Refmac5.<sup>6,7</sup> Each of the four molecules in the asymmetric unit were built independently due to significant structural differences. The register of the G and I helices was set using the unambiguous side chain density of the Gly-Tyr residues at 173-174 and 201-202, respectively. The H-helix showed density extending towards two potential I-helices, one roughly matching the connectivity of the wild-type structure and the other suggesting domain-swapped dimerization had occurred. Using the side chain density at residues Gly201 and Tyr202 to accurately thread the I-helix, we found that the domain swapped dimer would be off register by two residues. We cannot assign the origin of the continuous electron density extending between these regions of the protein. Generally, chain C

had the best density and was used for subsequent analysis and to prepare the figures.

### **Thermostability measurements**

Purified CYP119 variants (10  $\mu$ M, 50  $\mu$ L) were heated in a thermocycler (Eppendorf) over a range of temperatures (20  $^{\circ}$ C range, between 60 - 95  $^{\circ}$ C, e.g. 65 – 85  $^{\circ}$ C) for 10 min followed by rapid cooling to 4  $^{\circ}$ C for 1 min. The precipitate was removed by centrifugation. The concentration of folded CYP119 variant remaining in the supernatant was measured by CO-difference spectroscopy as described elsewhere.<sup>8</sup> Briefly, 40  $\mu$ L of the heat-treated CYP119 preparation was transferred to new wells in a microtiter plate. To this sample, dithionite solution (80 mM Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub> in 1 M KPi pH 8.0, 160  $\mu$ L) was added. The absorbance from 390 to 490 nm was recorded using a Tecan M1000 UV/Vis plate reader in 2 nm steps, and the microtiter plates were placed in a vacuum chamber. The chamber was sealed, evacuated to approximately -30 in Hg, purged with CO gas, and incubated for 30 min. The plates were then removed and the absorbance from 390 to 490 nm was again recorded using the plate reader. The difference spectra were taken by subtracting the absorbance at each wavelength before exposure from the absorbance of the ferrous CO-treated samples (i.e. ferrous CO bound absorbance minus ferrous absorbance). Baselines were normalized by subtracting the absorbance at 490 nm from the Soret maximum in both starting spectra and ferrous CO-bound spectra. Variants C317S, C317T, C317G, and C317P underwent shifts in their ferrous-CO difference spectra upon heating, and thus for these variants the lowest heated condition (which is lower than the 65  $^{\circ}$ C heat-treatment used during lysis) was used to normalize the other Soret signals to determine the fraction of folded enzymes remaining. The observed Soret shifts may reflect the higher protein concentration used for T50 assays (as opposed to lysate), and thus may reflect an aggregation behavior. The temperature at which half of the protein was denatured ( $T_{50}$ ) was determined by fitting the data to the equation:  $f(T) = 1 / (1 + \exp ((T_{50}-T)/C) + C_2$ .

### **Spectral characterization**

Proteins used in spectral characterization were thawed and then subjected to three additional concentration and dilution steps. Concentrations were then determined by the

pyridine-hemochrome assay described above, and then proteins were diluted to 10  $\mu\text{M}$  final concentration and analyzed by UV-visible spectroscopy from 350-600 nm in a plastic cuvette. To reduce enzymes, solid dithionite was added, and the cuvette was covered with parafilm and then read. After a satisfactory reduced spectrum was obtained, samples were bubbled with carbon monoxide for  $\sim 30\text{s}$ , or until a clear spectral shift was observed. Spectra were then exported to excel and analyzed.

### **Heme quantitation**

Holoprotein concentrations were determined on the basis of their heme content, which was determined using the pyridine-hemochrome assay. The assay was performed according to established methods.<sup>9</sup> Briefly, a protein sample (5-20  $\mu\text{L}$ ) was added to a cuvette, along with 2  $\mu\text{L}$  of 0.1 M  $\text{K}_3\text{Fe}(\text{CN})_6$  and pyridine-hemochrome reagent solution (80% water, 20 % pyridine (v/v), 50 mM NaOH). After mixing, 3-5 mg of solid dithionite was added, the cuvette was covered with parafilm and gently flicked to dissolve dithionite. Samples were read immediately and the absorbance spectrum scanned between 600 and 400 nm. The peak absorbance at 418 nm was used to determine concentration using the published extinction coefficient of  $191,500 \text{ M}^{-1} \text{ cm}^{-1}$ .<sup>9</sup>

Relative heme loadings were determined by the following procedure. Enzymes were diluted to 10  $\mu\text{M}$ . A portion of the enzyme solution (100  $\mu\text{L}$  in 0.1 M KPi pH 8.0) was mixed with formic acid solution (5 % v/v, 400  $\mu\text{L}$ ) to denature the protein and release the heme. To correct for pipetting and dilution errors, some samples were diluted more at this stage via addition of buffer (4:1 5% v/v formic acid: 0.1 M KPi pH 8.0) such that the heme absorbance ( $A_{395}$ ) was between 0.13 and 0.15. At this point the protein absorbance peak at 276 nm and the 395 nm absorbance were measured. The ratio of the heme content to the protein content was then used to determine a rank order of heme loadings among different axial variants.

**Table S1. Crystallographic data collection and refinement statistics**

<b>Variants</b>	<b>T213A/C317H</b>
Space group	P2 <sub>1</sub>
Cell dimensions (Å)	a,b,c = 64.5, 137.9, 91.8
Cell angles	$\alpha = \gamma = 90^\circ$ , $\beta =$ 101.18°
<b>Data Collection</b>	
Wavelength (Å)	0.9795
Beamline	SSRL 12.2
Resolution (Å)	50 - 2.7
Last Bin (Å)	2.7 - 2.8
No. Observations	138,213
Completeness (%)	98.5 (97.6)
$R_{pim}$ (%)	0.051 (0.499)
CC(1/2)	0.995 (0.565)
I/ $\sigma$ I	10.7 (1.6)
Redundancy	3.3 (3.1)
<b>Refinement</b>	
Total no. of reflections	40139
Total no. of atoms	9827
No. of waters	
Final bin (Å)	2.7 - 2.77
$R_{work}$ (%)	23.8 (35.0)
$R_{free}$ (%)	27.7 (38.1)
Average B factor (Å <sup>2</sup> )	64.5
Ramachandran plot allowed, %	97.7
Favored, %	99.9
Outliers, %	0.1

Values in parenthesis are for the highest resolution shell

$R_{merge}$  is  $\sum |I_o - \bar{I}| / \sum I_o$ , where  $I_o$  is the intensity of an individual reflection, and  $\bar{I}$  is the mean intensity for multiply recorded reflections

$R_{work}$  is  $\sum |F_o - F_c| / F_o$ , where  $F_o$  is an observed amplitude and  $F_c$  a calculated amplitude;  
 $R_{free}$  is the same statistic calculated over a 5% subset of the data that has not been included during refinement.

**Table S2. Summary of UV-visible spectroscopy of C317X mutants.** Broad peaks marked 'br'.

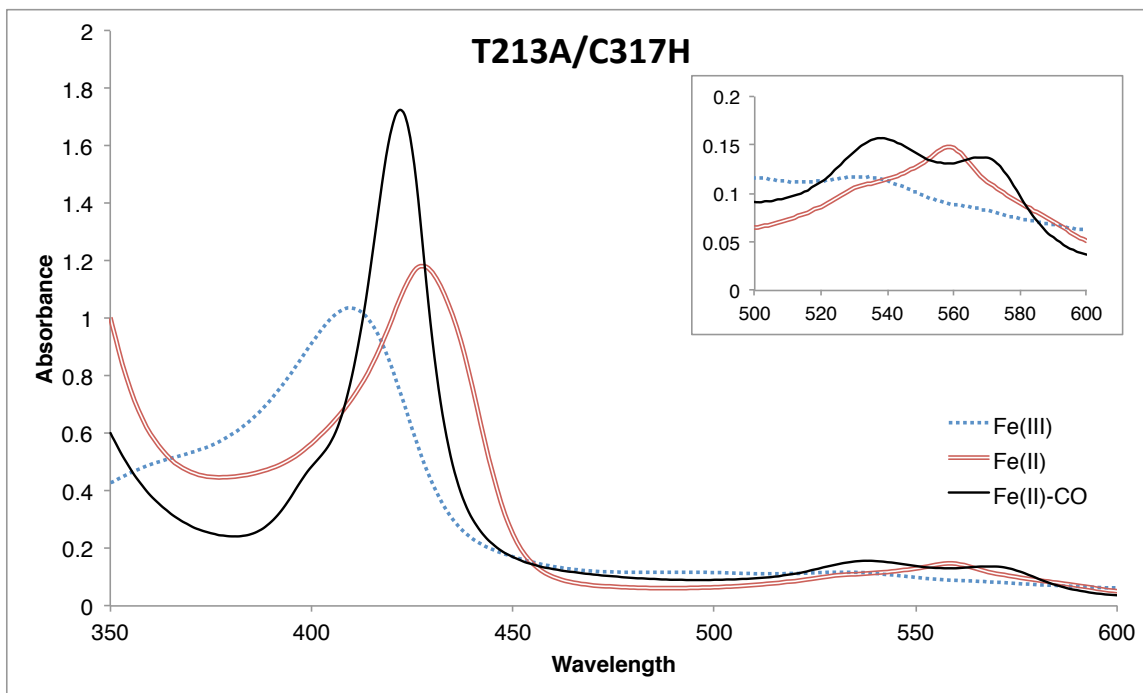
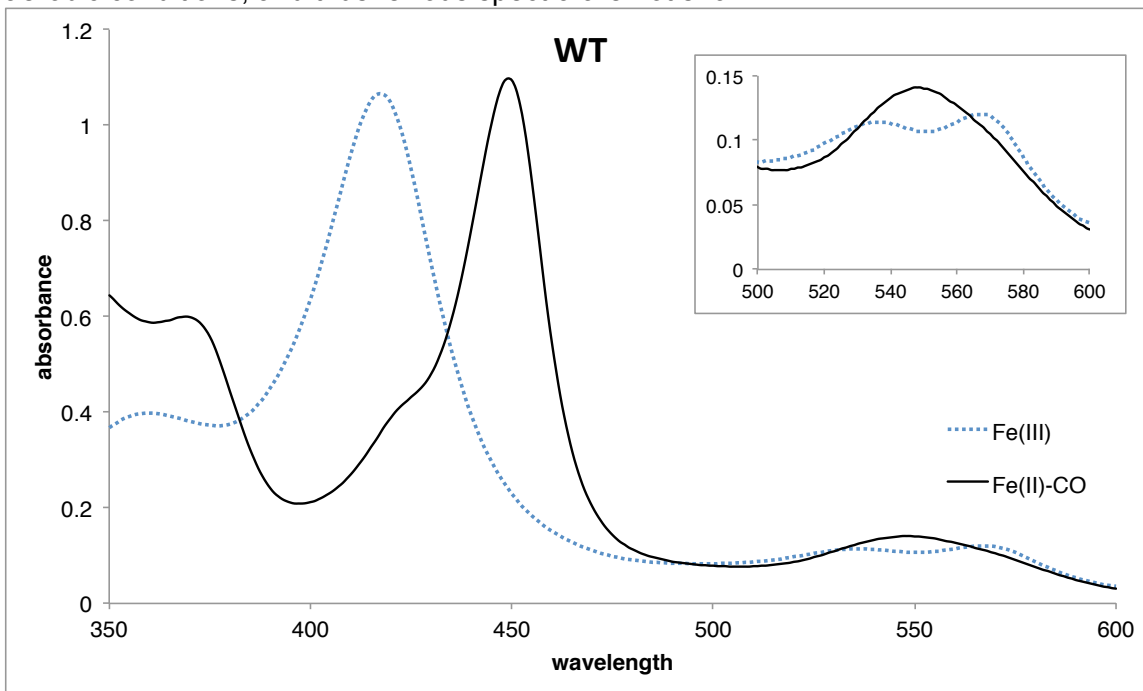
	$\lambda$ (nm)		
	$\alpha$	$\beta$	Soret
WT Fe(III)	568	536	417
WT Fe(II)CO	548 (br)	548 (br)	449
T213A/C317H Fe(III)	-	532 (br)	409
T213A/C317H Fe(II)	559	-	428
T213A/C317H Fe(II)CO	568	538	422
C317H Fe(III)	-	528 (br)	409
C317H Fe(II)	559 (br)	-	428
C317H Fe(II)CO	569	540	422
C317G Fe(III)	578	544	413
C317G Fe(II)	551	-	426
C317G Fe(II)CO	568	540	414
C317A Fe(III)	570	-	405
C317A Fe(II)	551	-	424
C317A Fe(II)CO	553	-	400
C317S Fe(III)	-	-	404
C317S Fe(II)	550 (br)	550 (br)	422
C317S Fe(II)CO	564	528	408
C317T Fe(III)	575 (br)	524 (br)	408
C317T Fe(II)CO	574	536	413
C317P Fe(III)	-	529	411
C317P Fe(II)	560	531	427
C317P Fe(II)CO	567	540	423
C317V Fe(III)	-	532	408
C317V Fe(II)	558 (br)	558 (br)	429
C317V Fe(II)CO	567	539	422
C317L Fe(III)	-	527	409
C317L Fe(II)	559	535	427
C317L Fe(II)CO	567	539	422
C317I Fe(III)	-	525 (br)	409
C317I Fe(II)	559	559 (br)	427
C317I Fe(II)CO	568	539	422

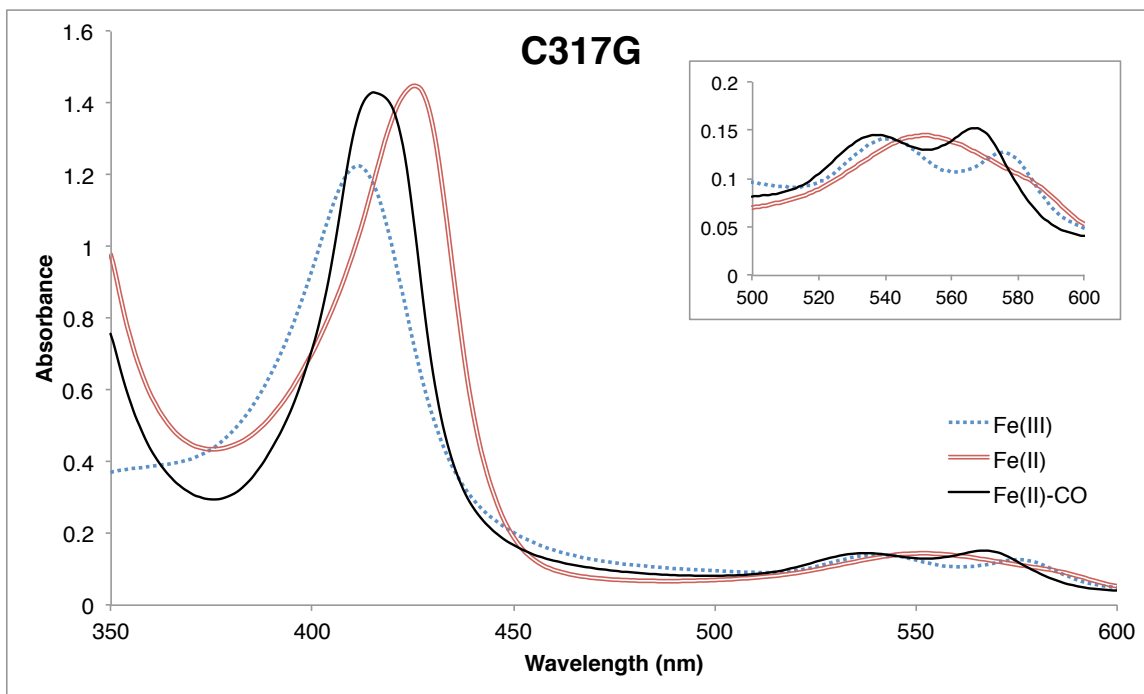
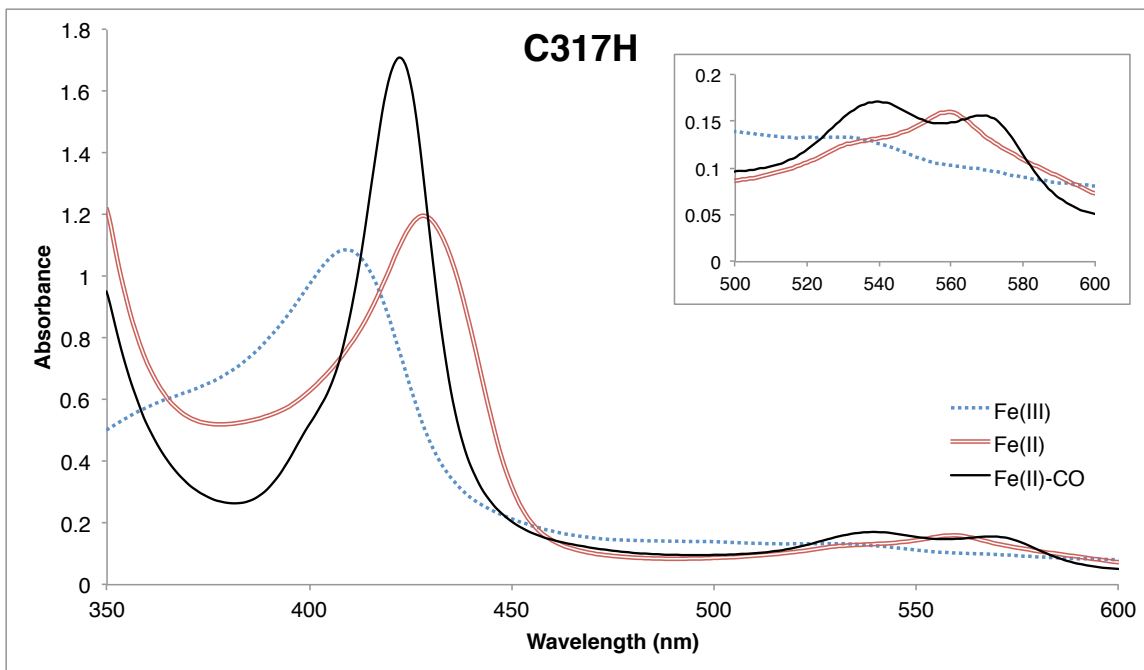


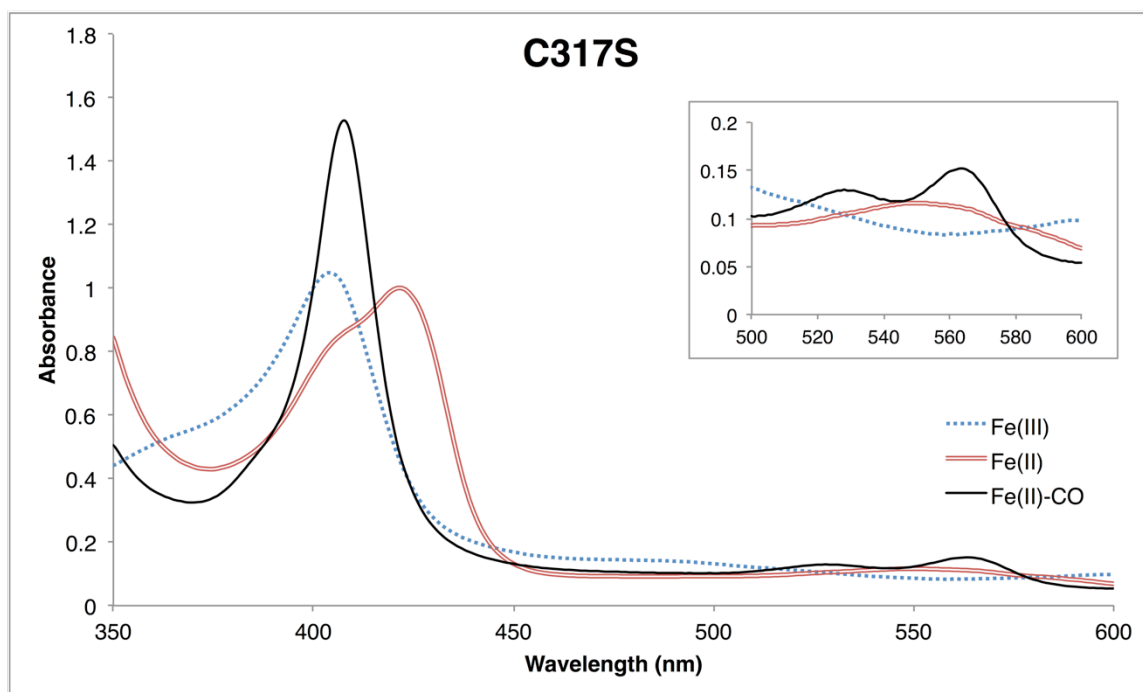
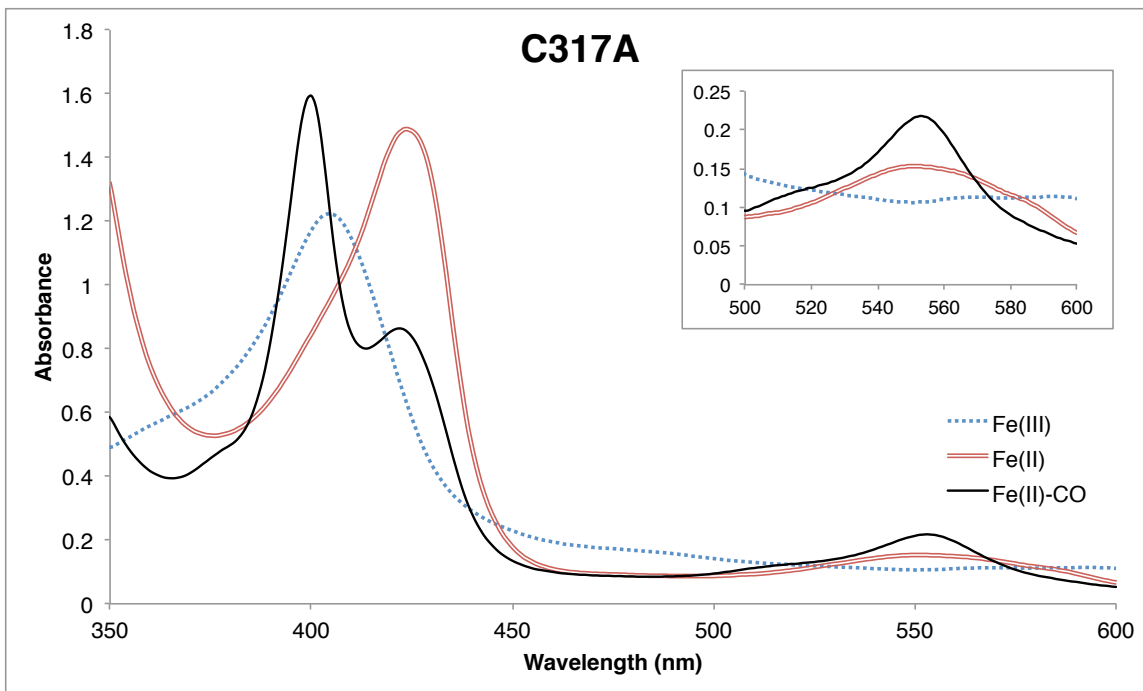
	$\lambda$ (nm)		
	$\alpha$	$\beta$	Soret
C317M Fe(III)	-	528	408
C317M Fe(II)	559	530	427
C317M Fe(II)CO	569	540	422
C317N Fe(III)	-	530	411
C317N Fe(II)	561 (br)	561 (br)	428
C317N Fe(II)CO	568	539	422
C317D Fe(III)	-	531	408
C317D Fe(II)	560 (br)	560 (br)	427
C317D Fe(II)CO	570	539	422
C317Q Fe(III)	-	526 (br)	408
C317Q Fe(II)	558 (br)	558 (br)	428
C317Q Fe(II)CO	567	540	422
C317E Fe(III)	-	530	411
C317E Fe(II)	560 (br)	560 (br)	427
C317E Fe(II)CO	569	540	423
C317K Fe(III)	-	530	410
C317K Fe(II)	559.5 (br)	559.5 (br)	427
C317K Fe(II)CO	569	540	422
C317R Fe(III)	-	532	411
C317R Fe(II)	559 (br)	559 (br)	427
C317R Fe(II)CO	569	540	422
C317F Fe(III)	-	528	409
C317F Fe(II)	558 (br)	558 (br)	428
C317F Fe(II)CO	567	539	422
C317Y Fe(III)	-	526	408
C317Y Fe(II)	558 (br)	558 (br)	428
C317Y Fe(II)CO	567	540	422
C317W Fe(III)	-	529	408
C317W Fe(II)	558 (br)	558 (br)	428
C317W Fe(II)CO	568	540	422

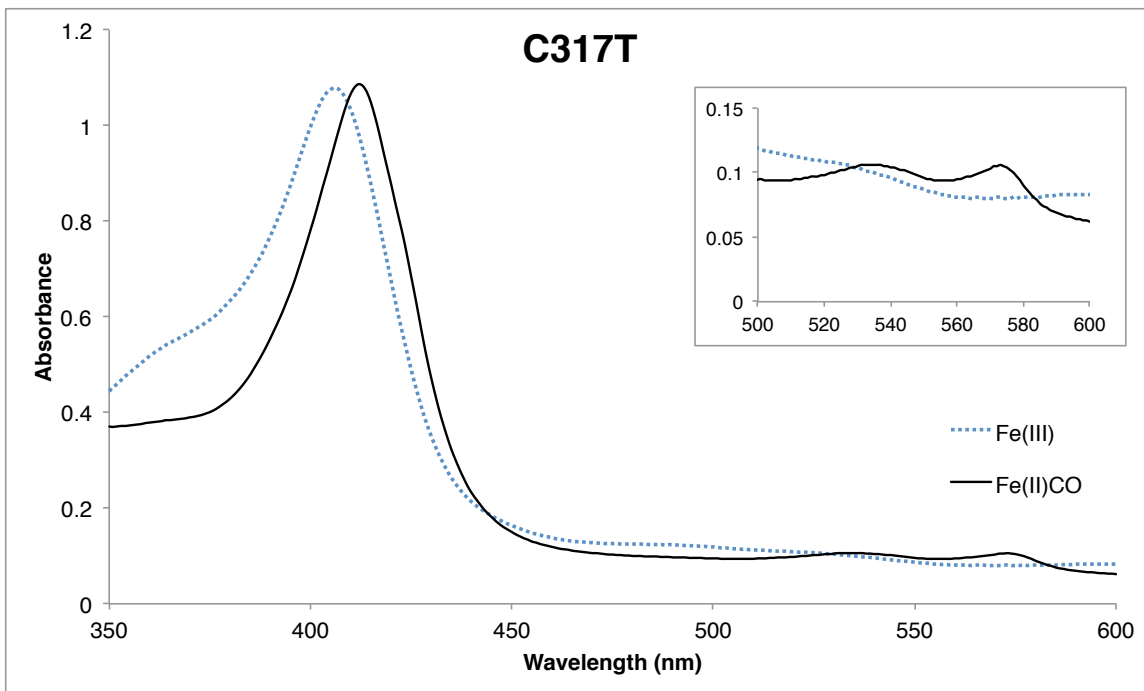
**Figure S1. UV-visible spectra for C317X variants**

Ferric, ferrous, and ferrous CO spectra are shown as dotted, double-line, and solid line traces. Wild-type and C317T variants could not be completely reduced under semi-aerobic conditions, and thus ferrous spectra are not shown.

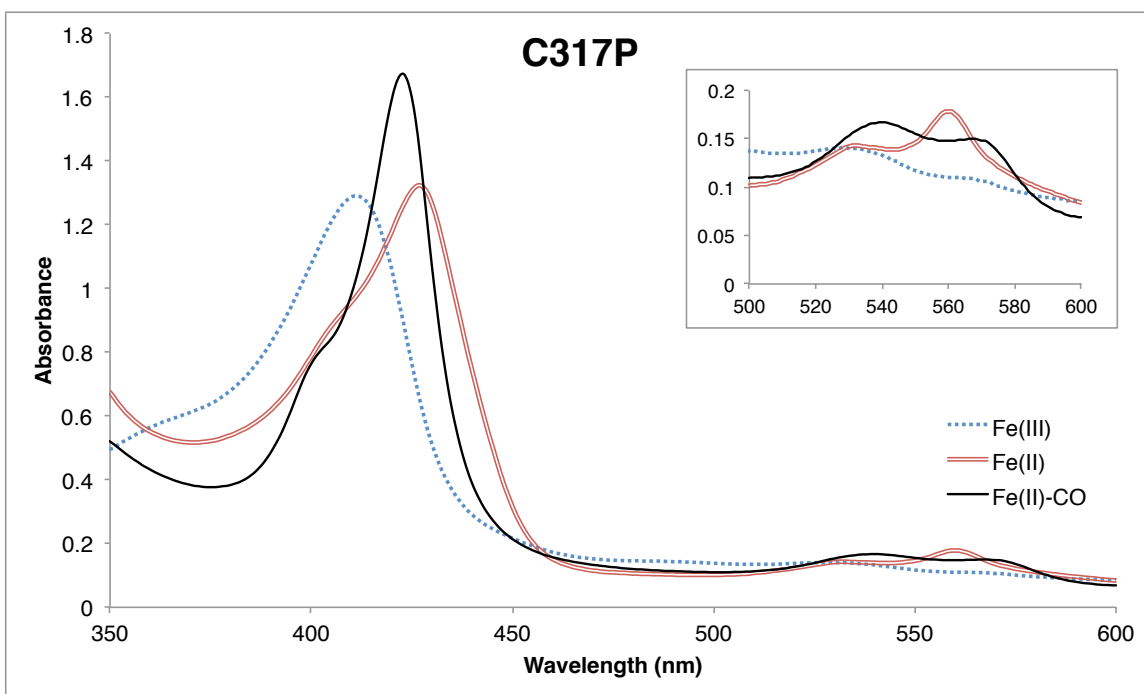


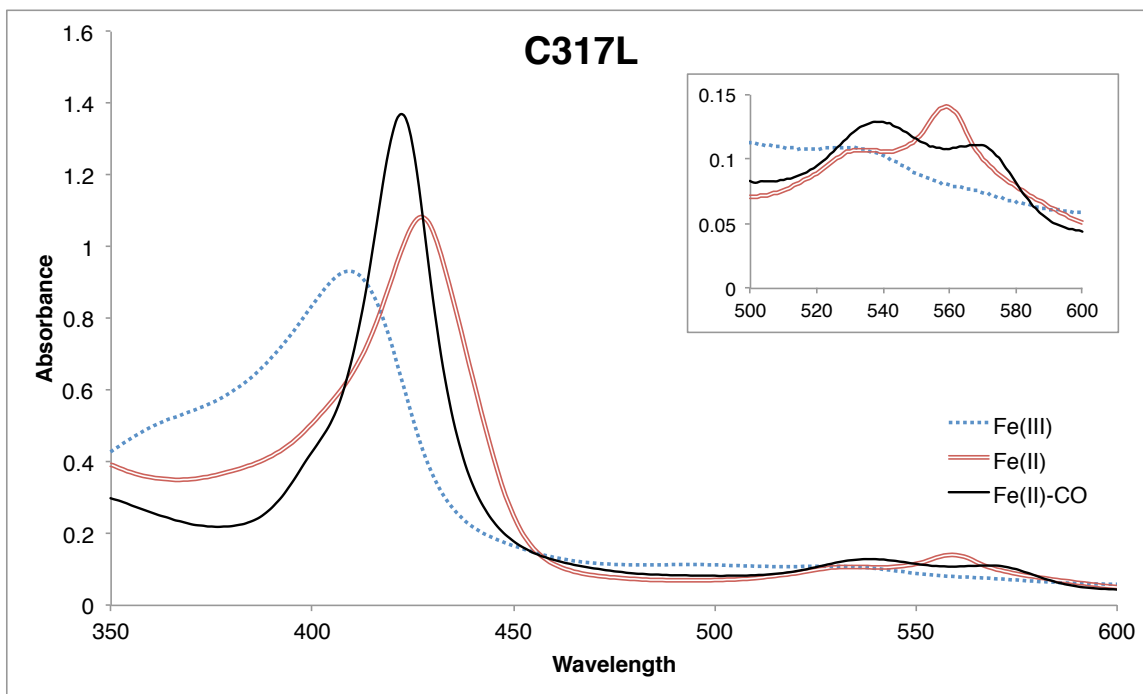
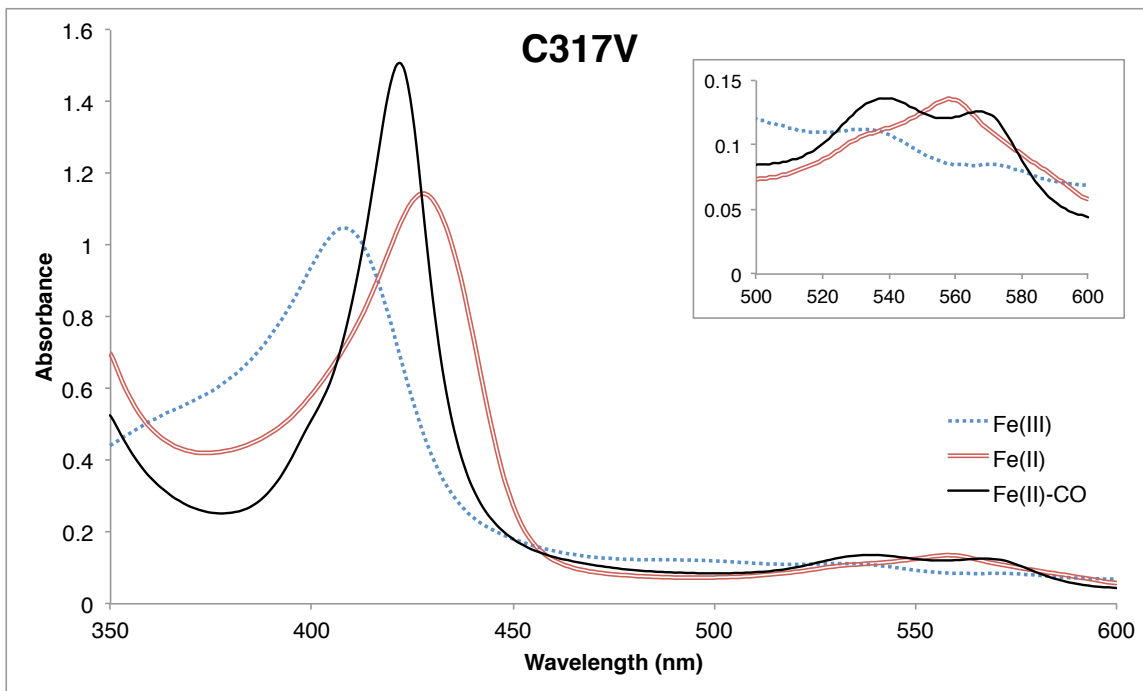


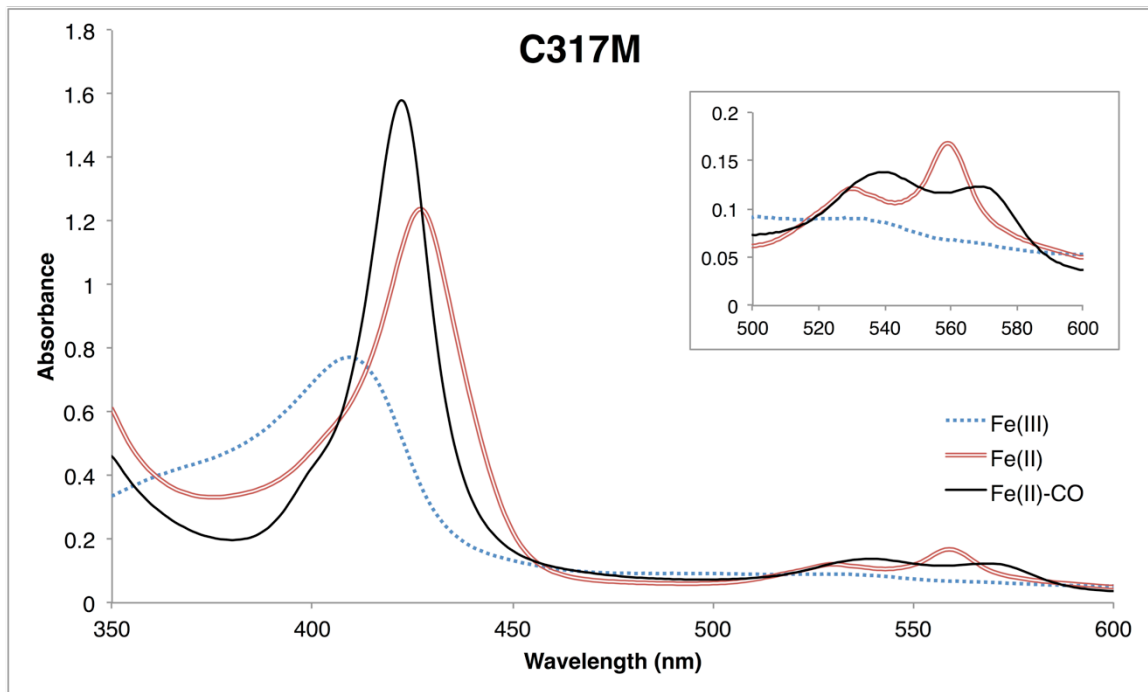
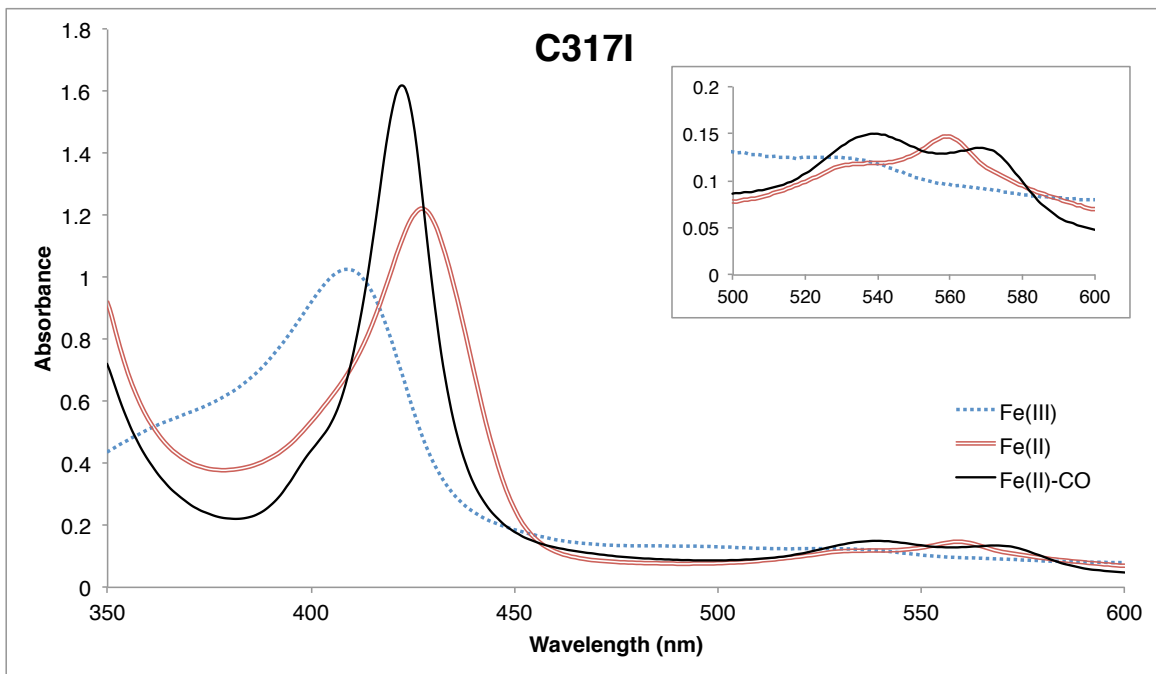


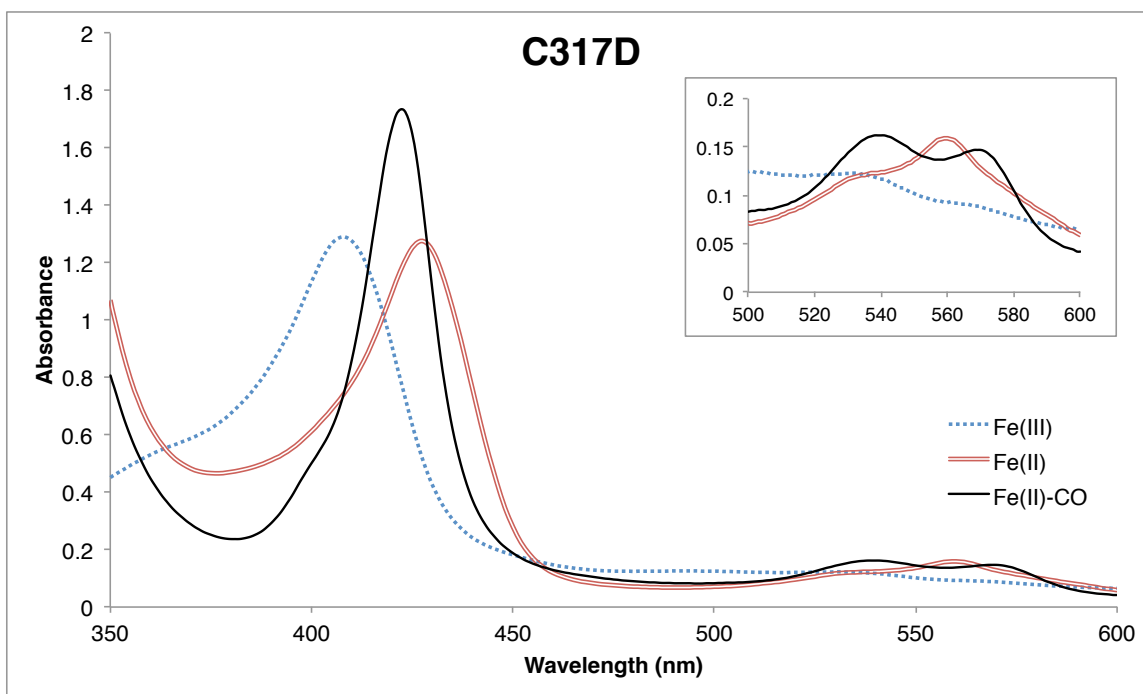
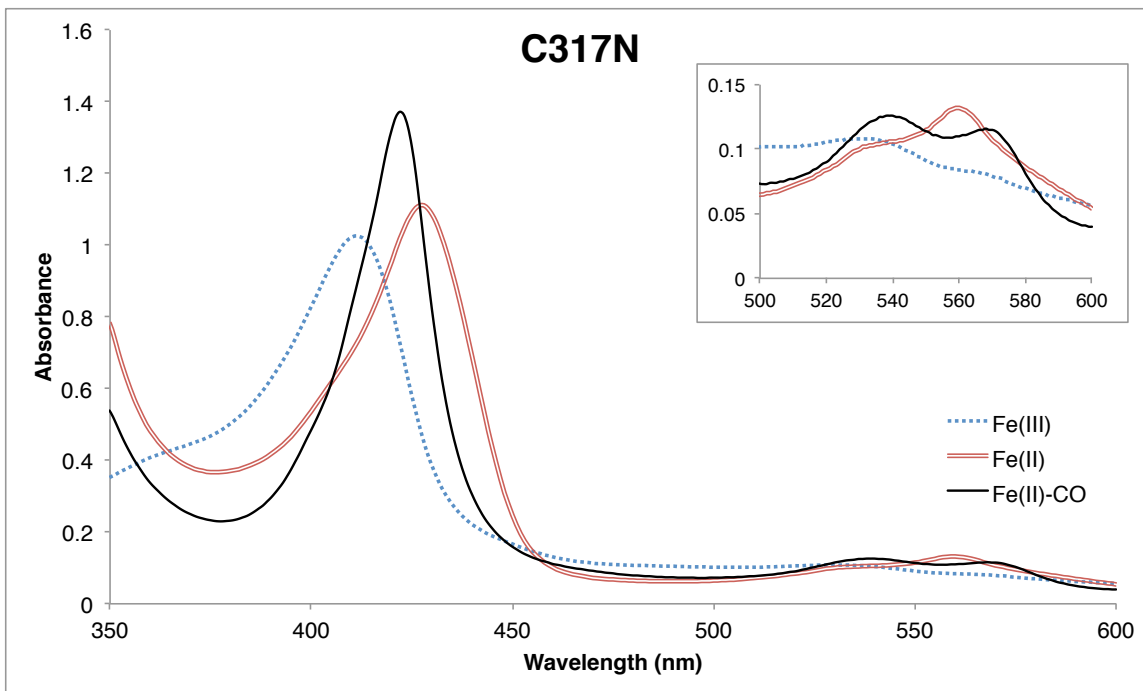


fref

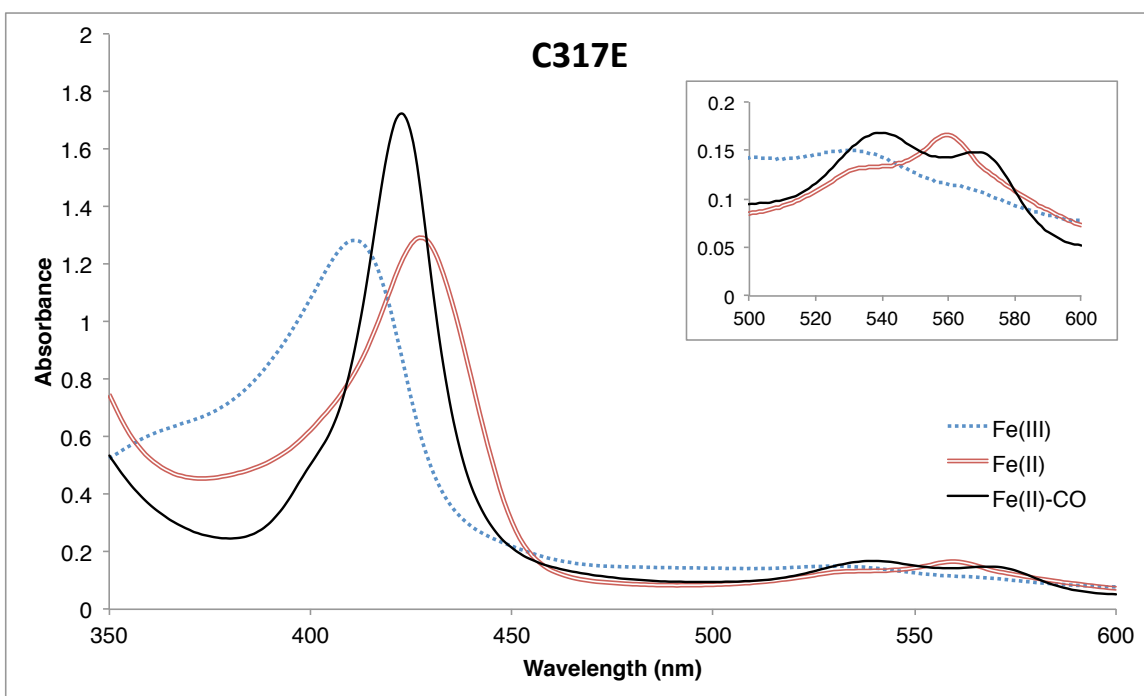
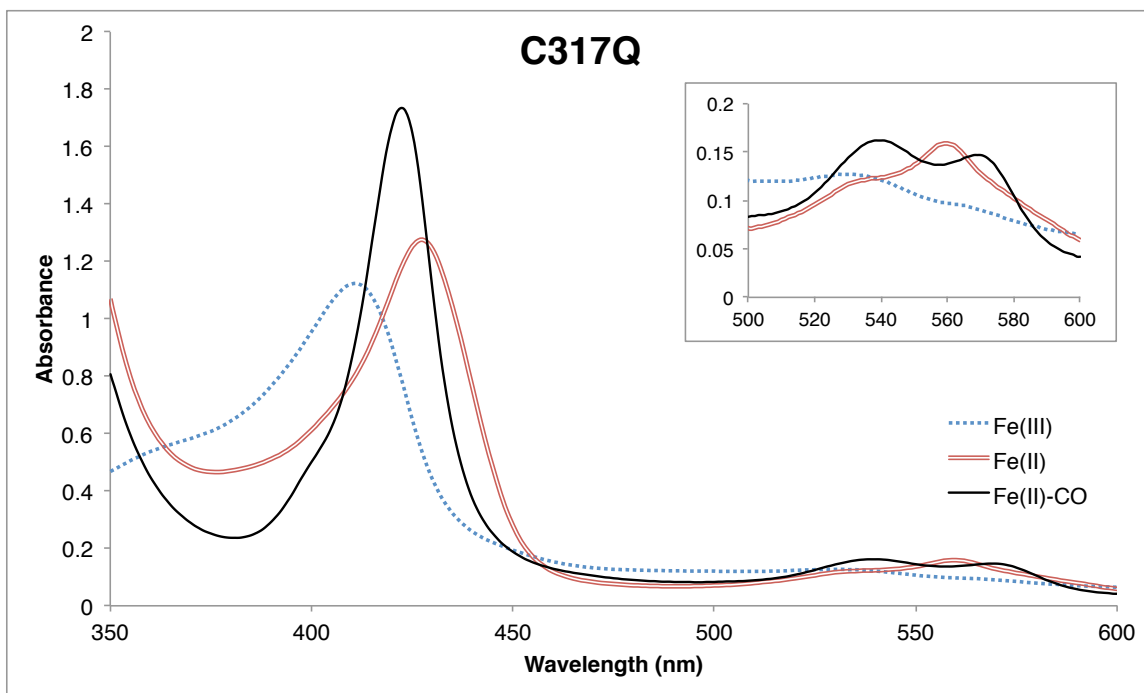


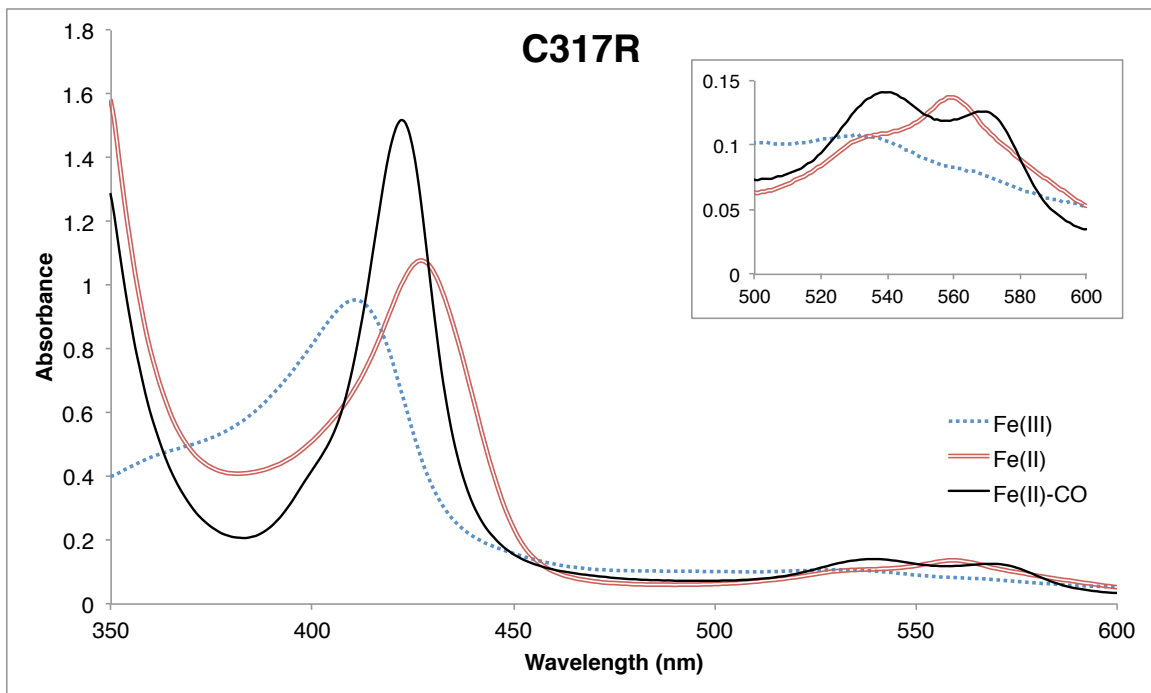
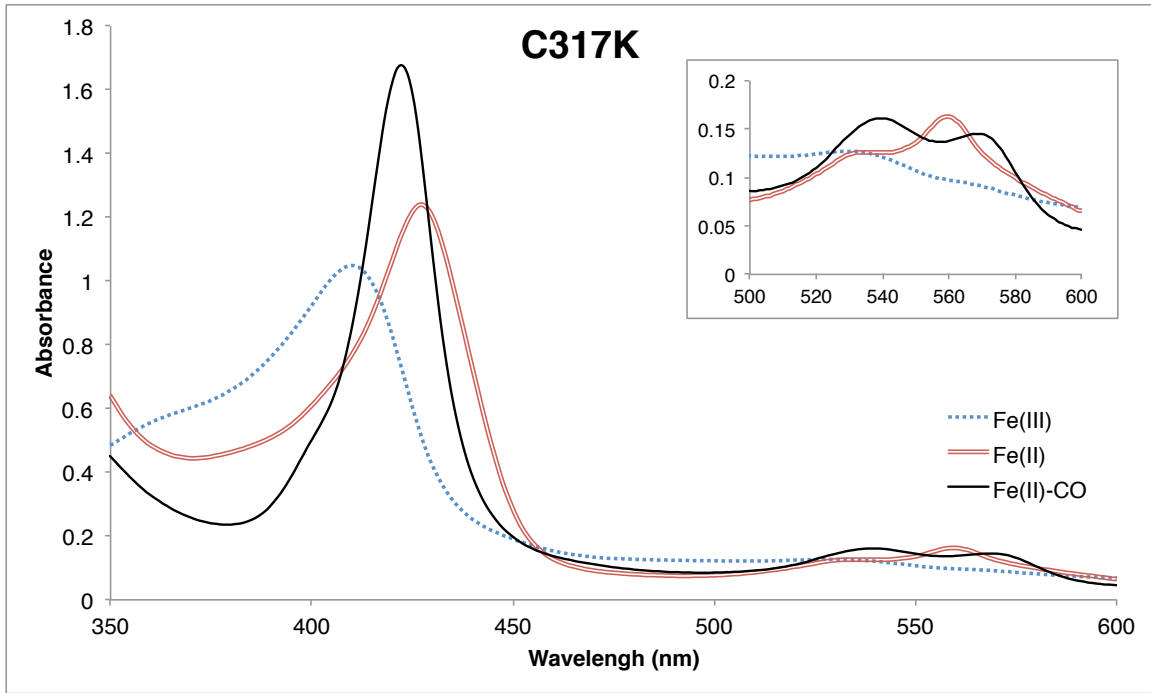


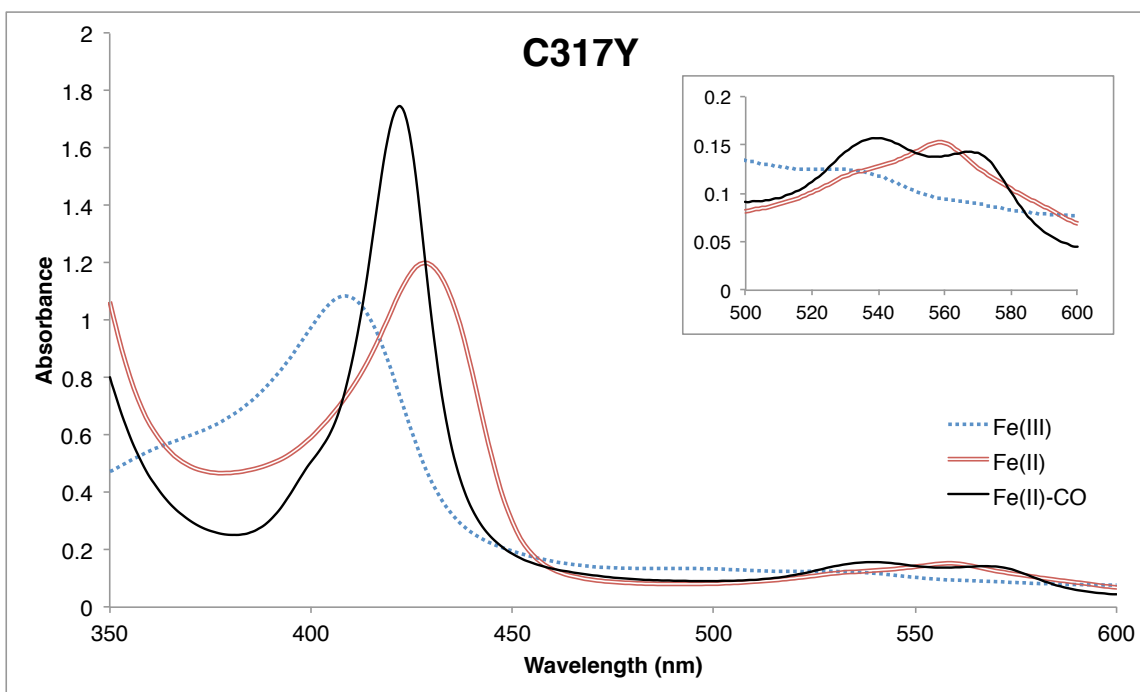
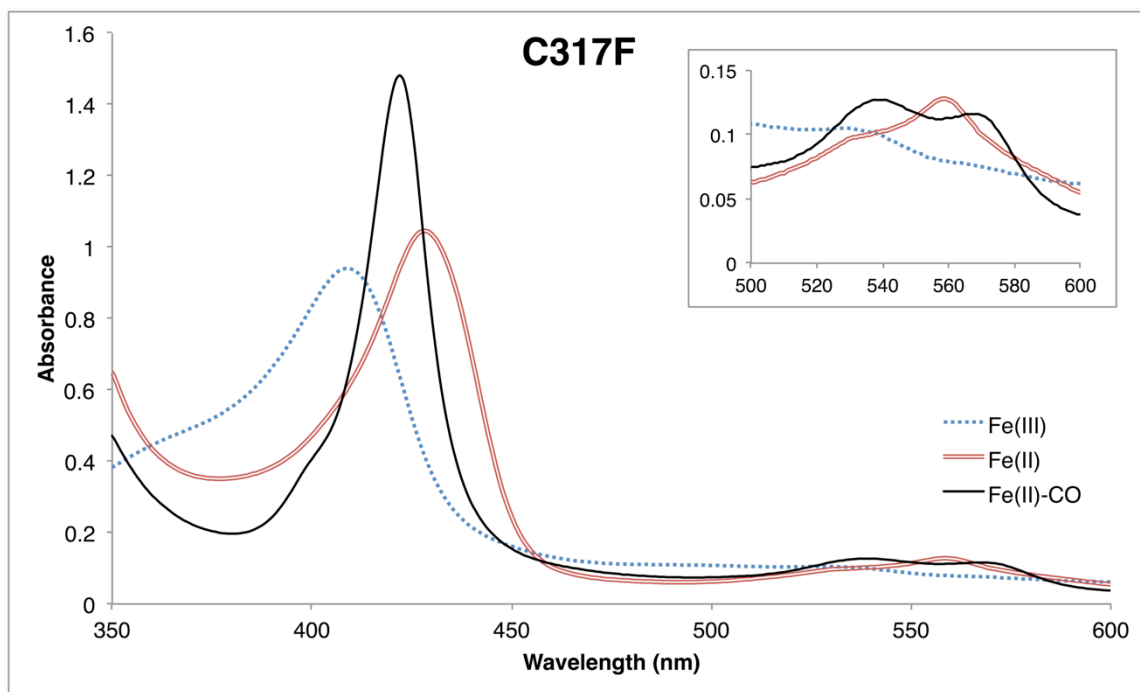


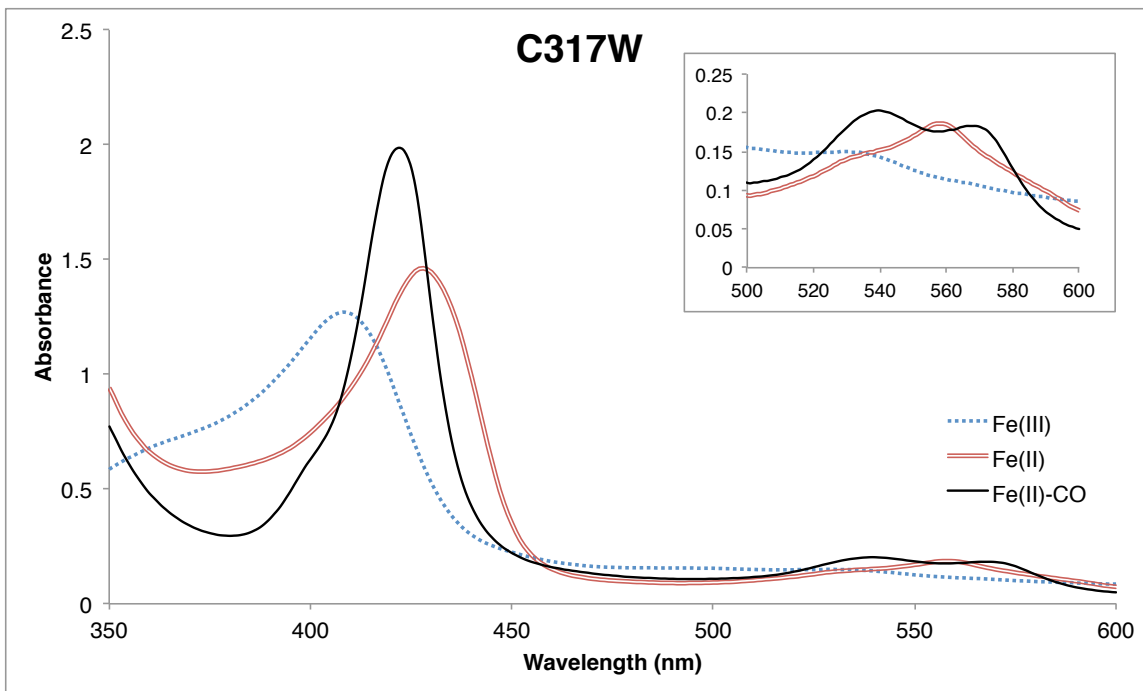












**Table S3. Relative heme incorporation levels for C317X mutants**

Higher heme loading is reflected in increased ratios of 395 nm to 276 nm absorbances for the acid-denatured proteins.

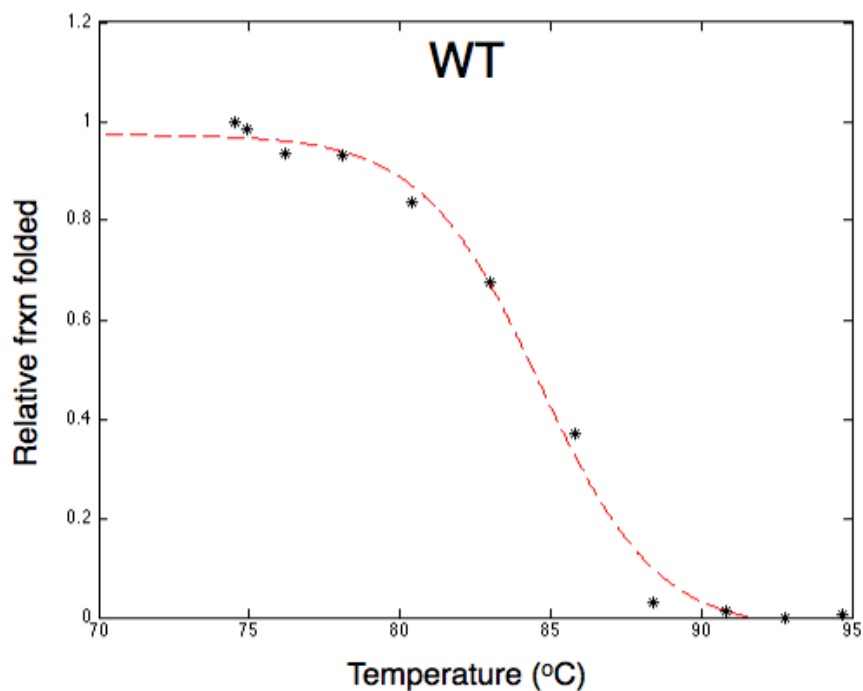
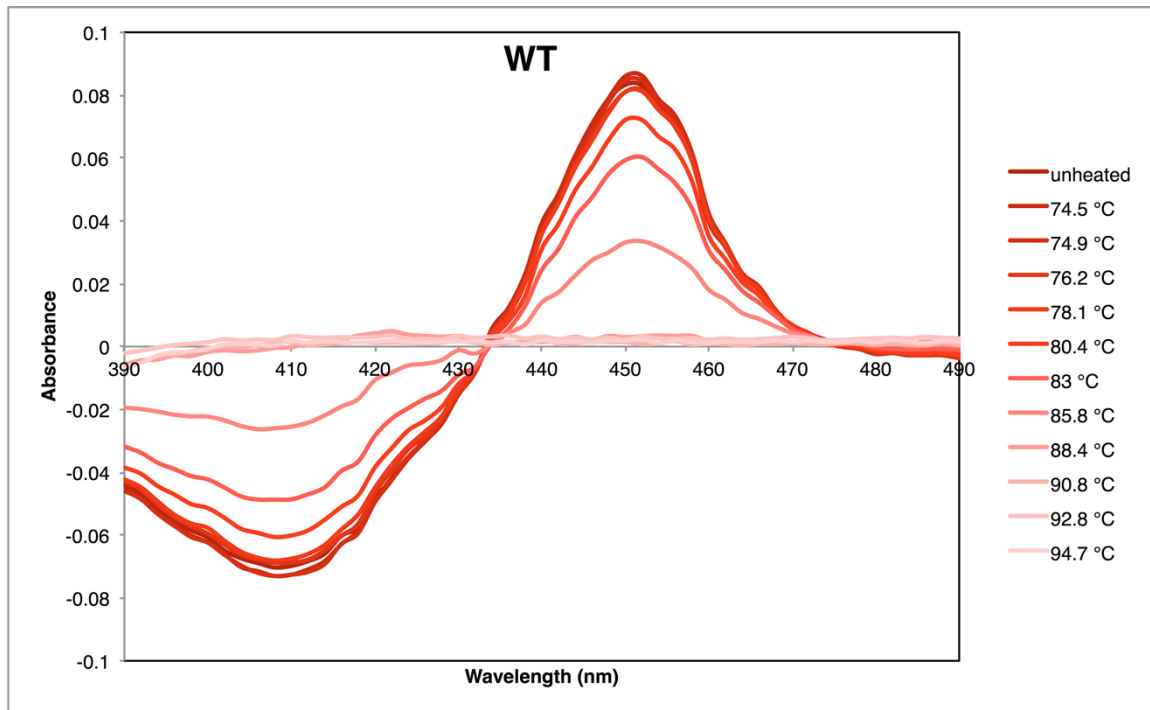
<b>C317X</b>	<b>A395/A276 ratio</b>
Arg	1.5
Lys	1.5
His	1.4
Glu	1.4
Leu	1.4
Trp	1.3
Met	1.3
Ile	1.3
Tyr	1.3
Cys (T213A)	1.2
Gln	1.2
Pro	1.2
His (T213A)	1.2
Asp	1.2
Phe	1.1
Cys (WT)	1.1
Val	1.1
Gly	1.0
Ala	1.0
Asn	1.0
Thr	0.9
Ser	0.8

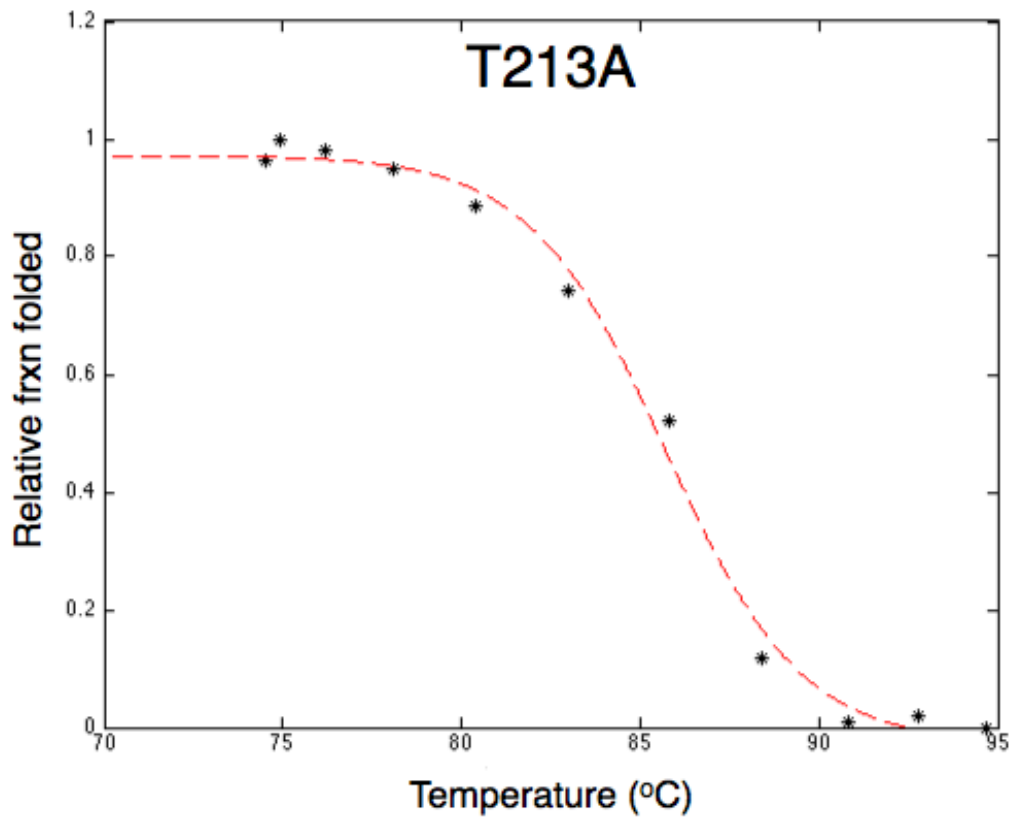
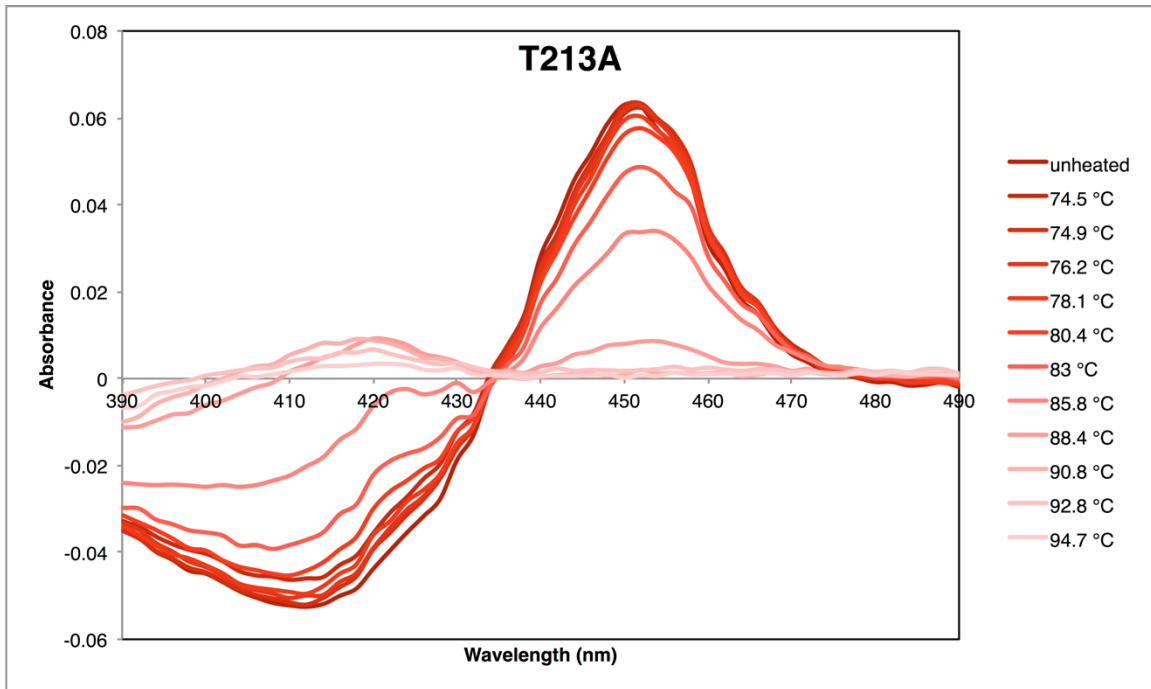
**Table S4. T<sub>50</sub> data for axial mutants described in this study.** Axial mutants are ordered from highest to lowest T<sub>50</sub> (the temperature at which half of the protein is unfolded after a ten minute incubation – see Methods for complete assay description). Additional mutations shown in parentheses.

<b>C317X</b>	<b>T<sub>50</sub> (°C)</b>
Cys (T213A)	85.7
Cys (WT)	<b>84.6</b>
His (T213A)	83.4
Lys	79.6
Gly*	79.3
Glu	79.2
Gln	79.2
Arg	79.1
Tyr	79.1
Thr*	79.1
Leu	78.9
Trp	78.8
Val	78.8
Asp	78.6
Asn	78.5
His	78.5
Ile	78.3
Pro*	78.2
Met	78.1
Phe	78.1
Ser*	77.9
Ala	77.1

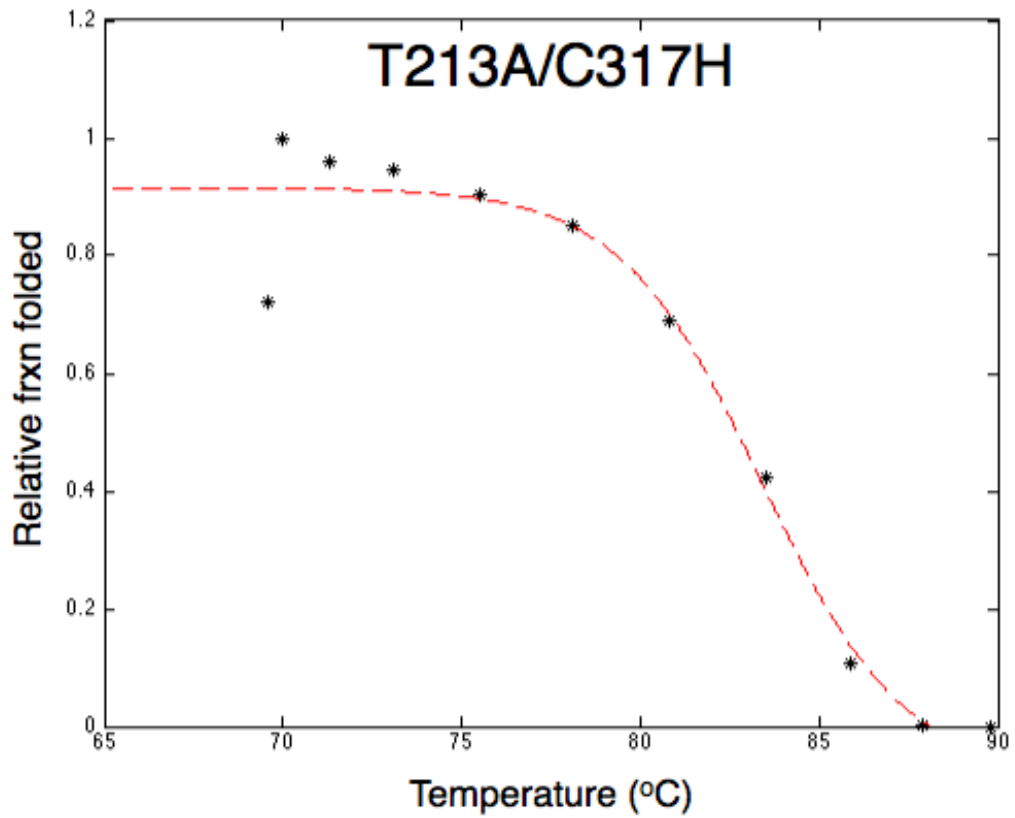
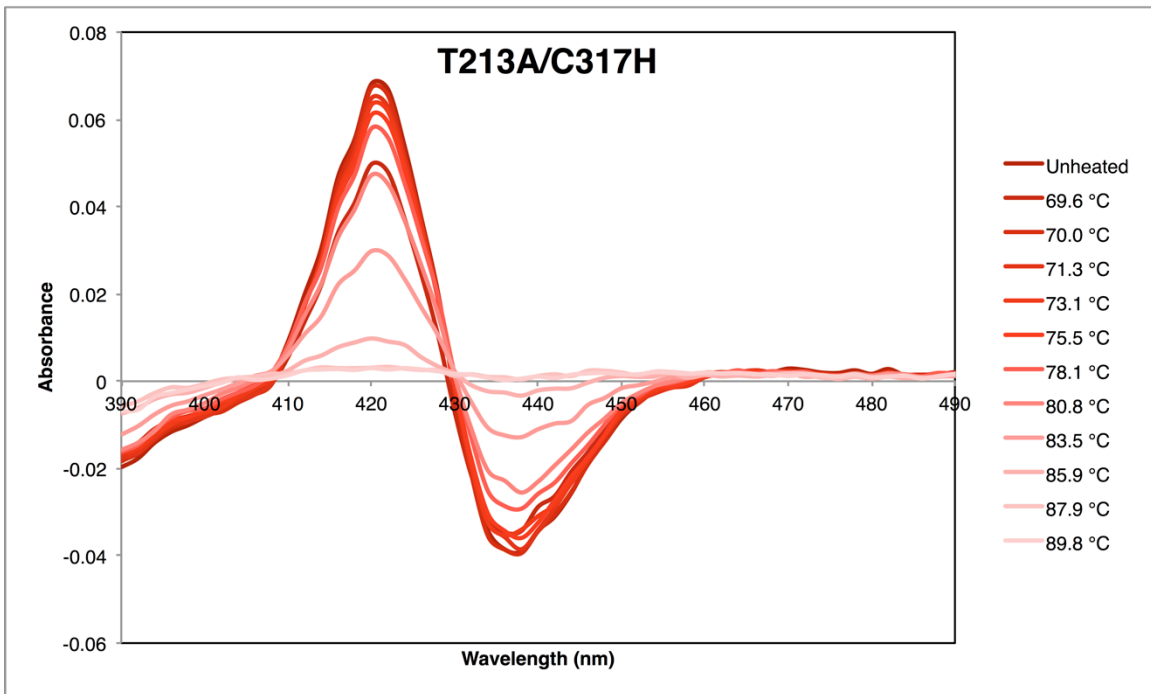
\*Underwent Soret shift upon heating

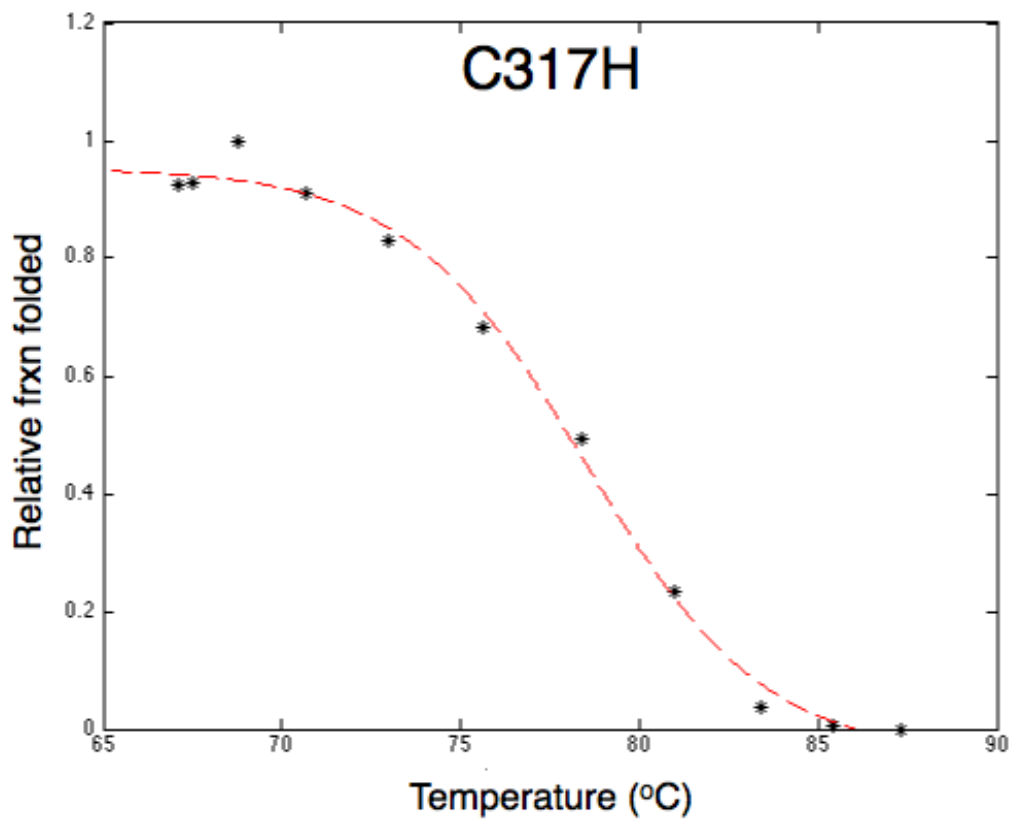
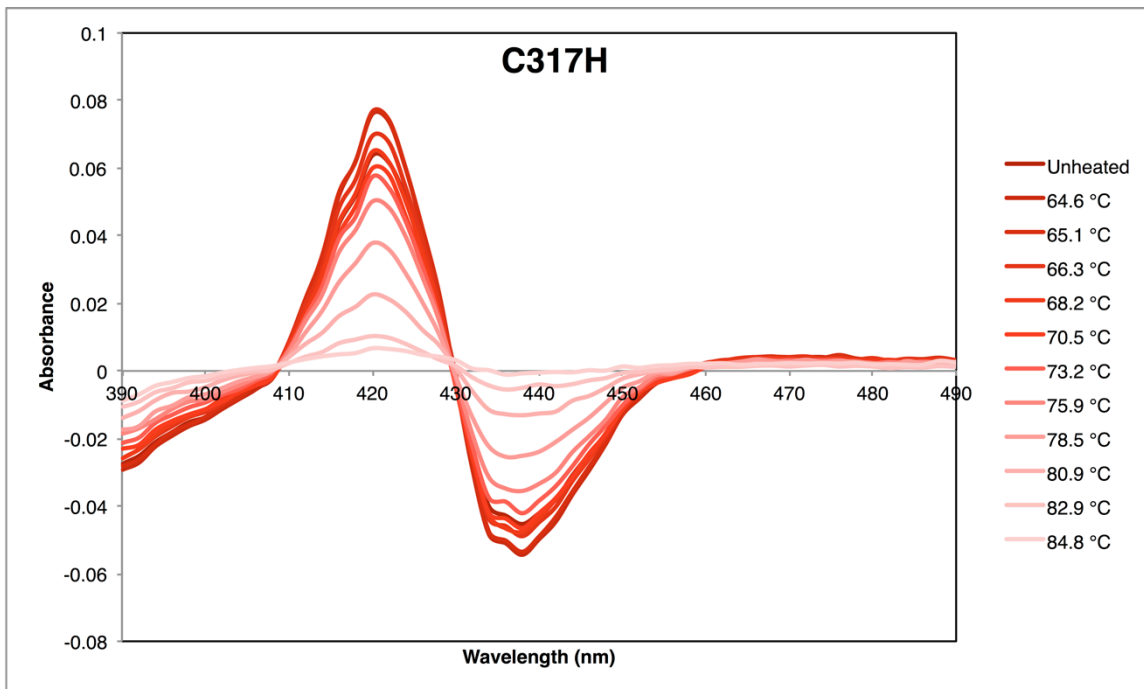
**Figure S2.** C317X  $T_{50}$  curves and raw spectral data (shown in difference mode) used to generate  $T_{50}$  curves. Variants C317G, C317S, C317T, C317P gave shifts in their CO-bound Soret difference maxima upon heating, and thus for these the lowest-heated condition was used to normalize the other ferrous CO-bound Soret signals. The differences in these variants could be traced to overt shifts in either the ferric or ferrous-CO Soret peaks.

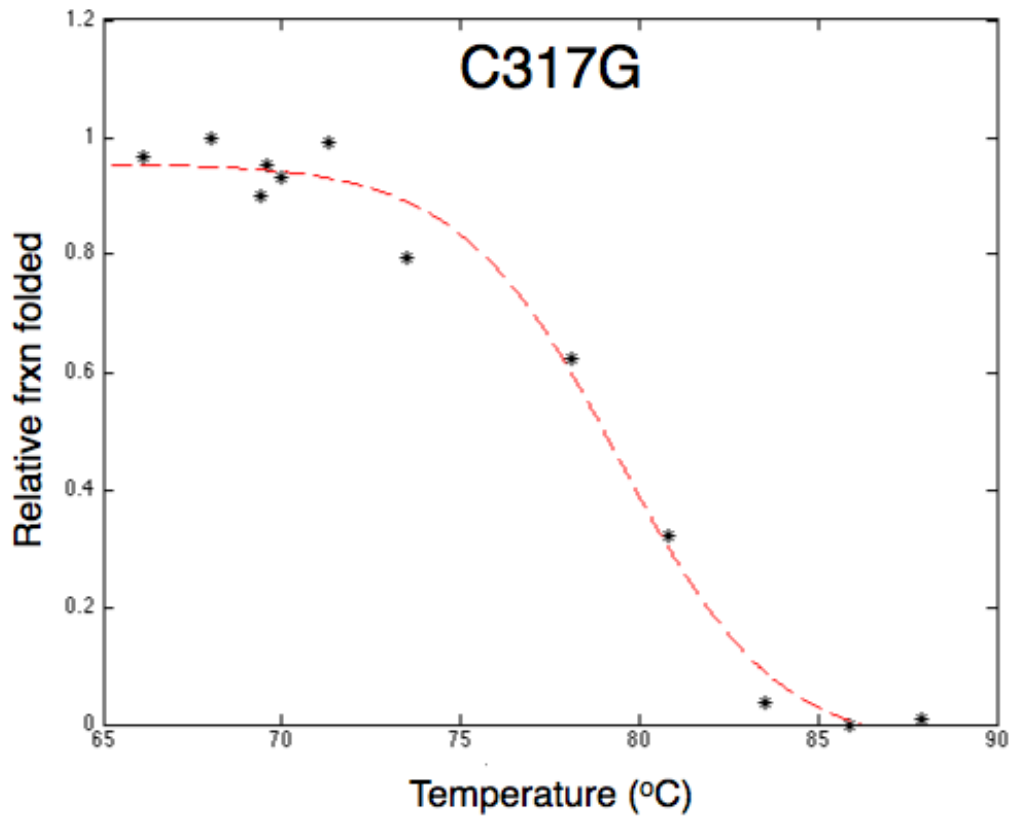
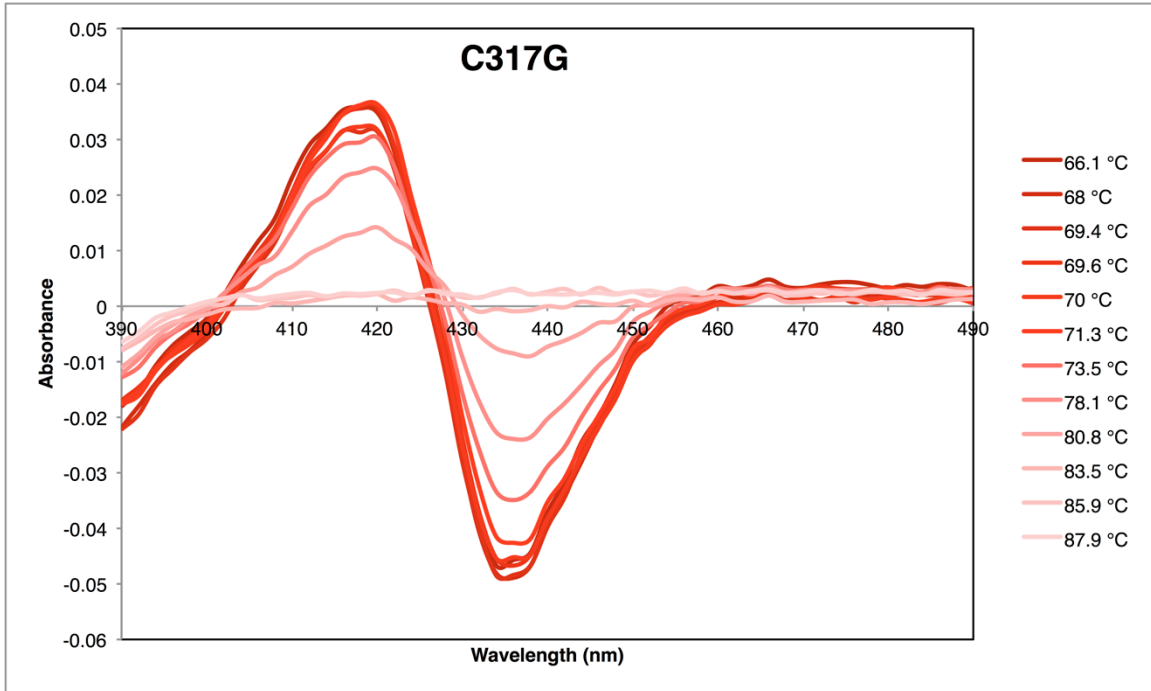


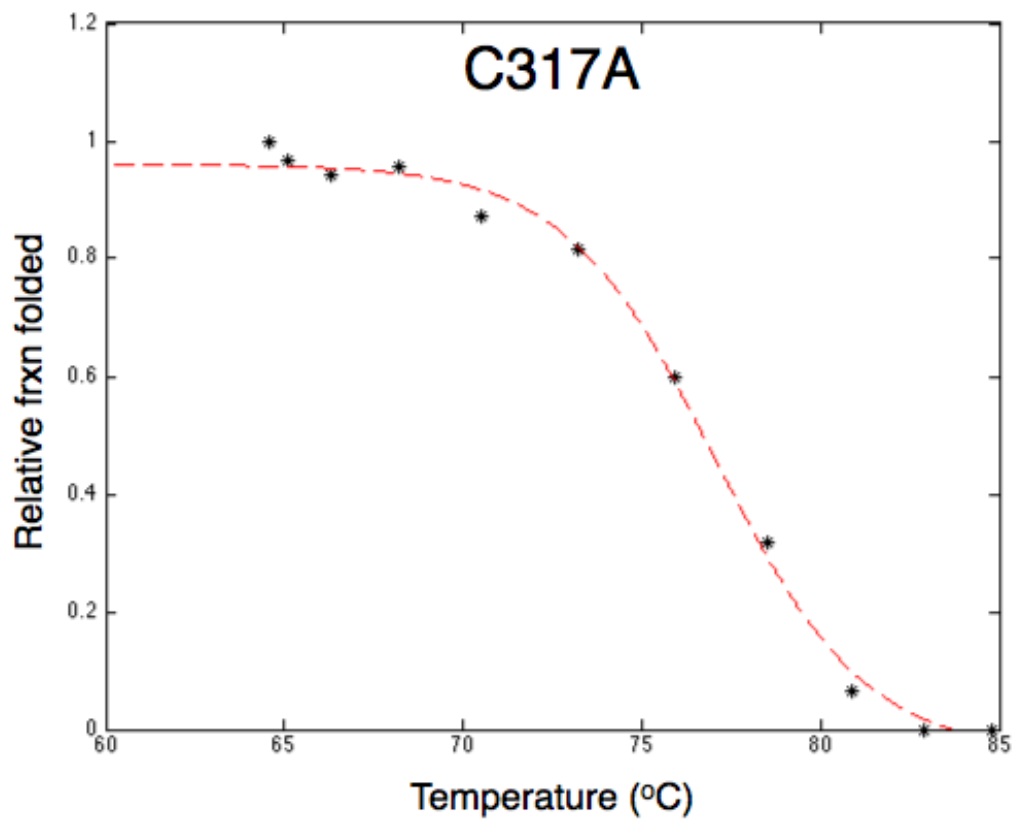
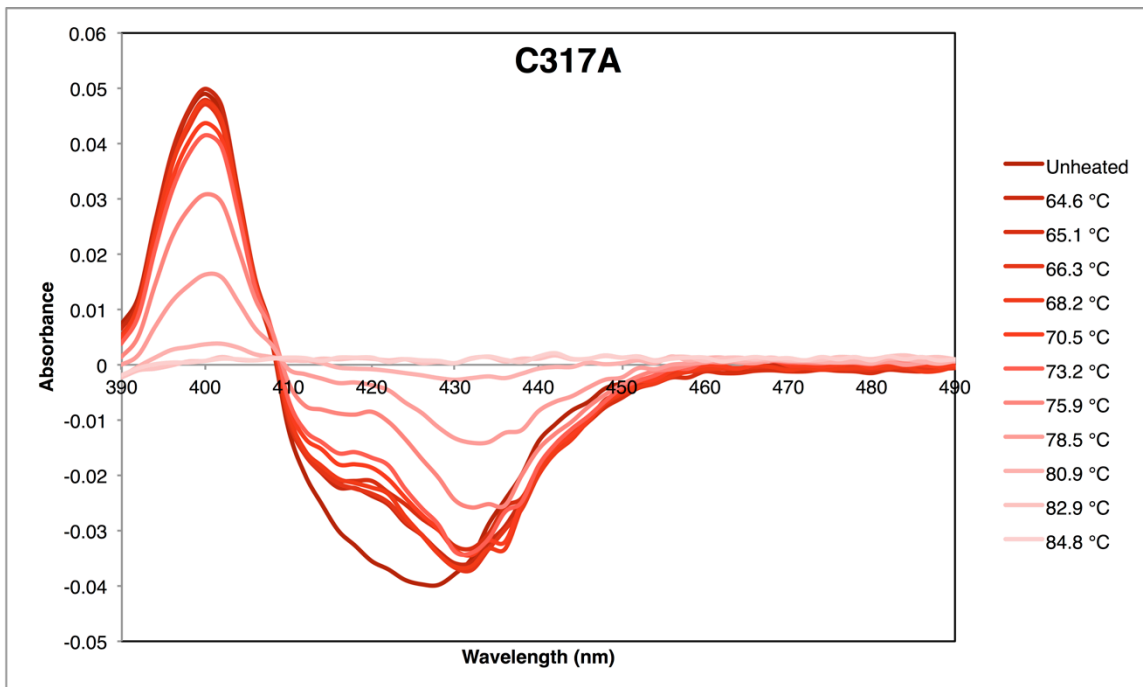


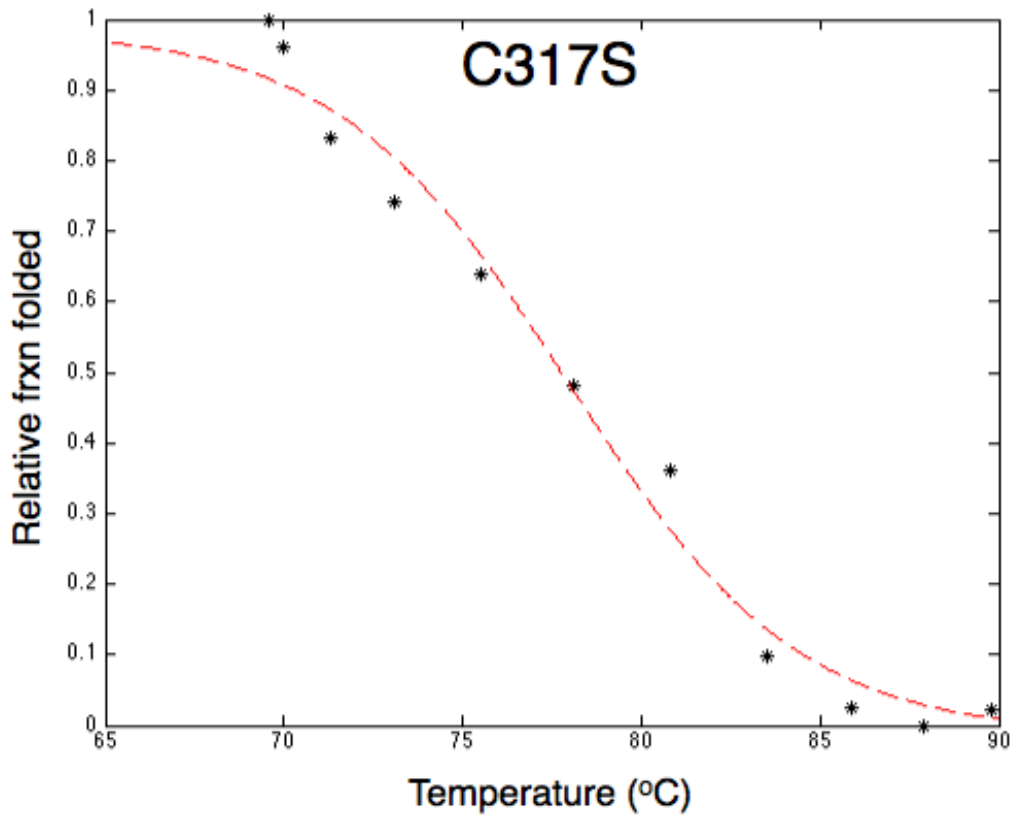
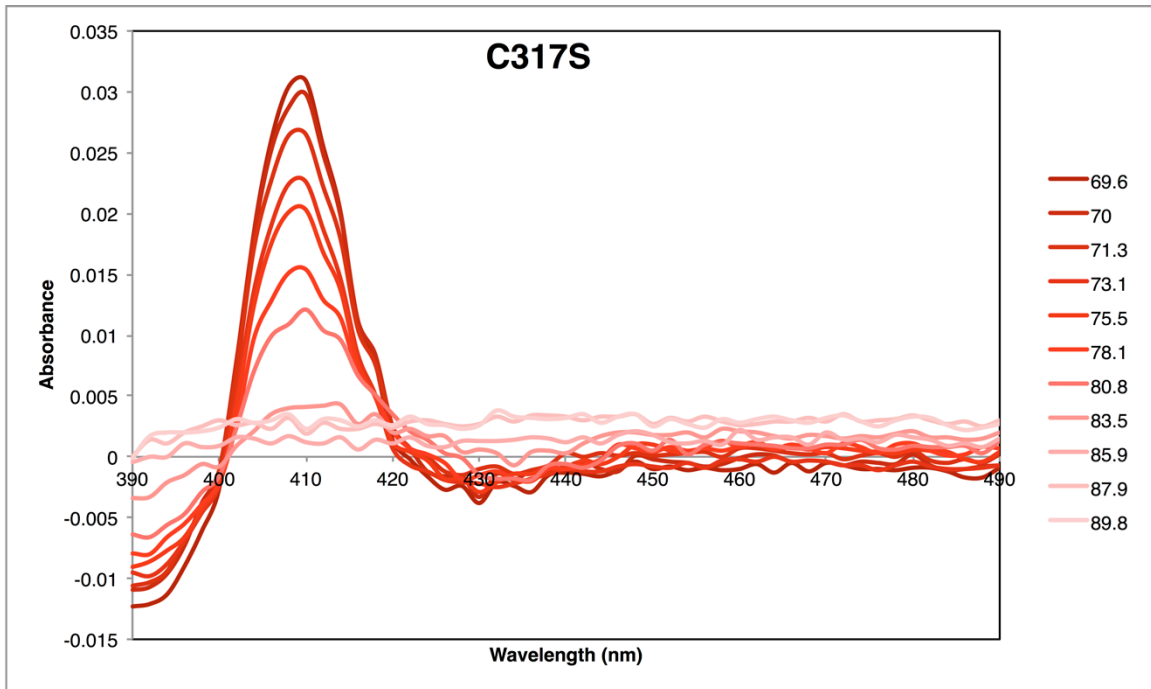


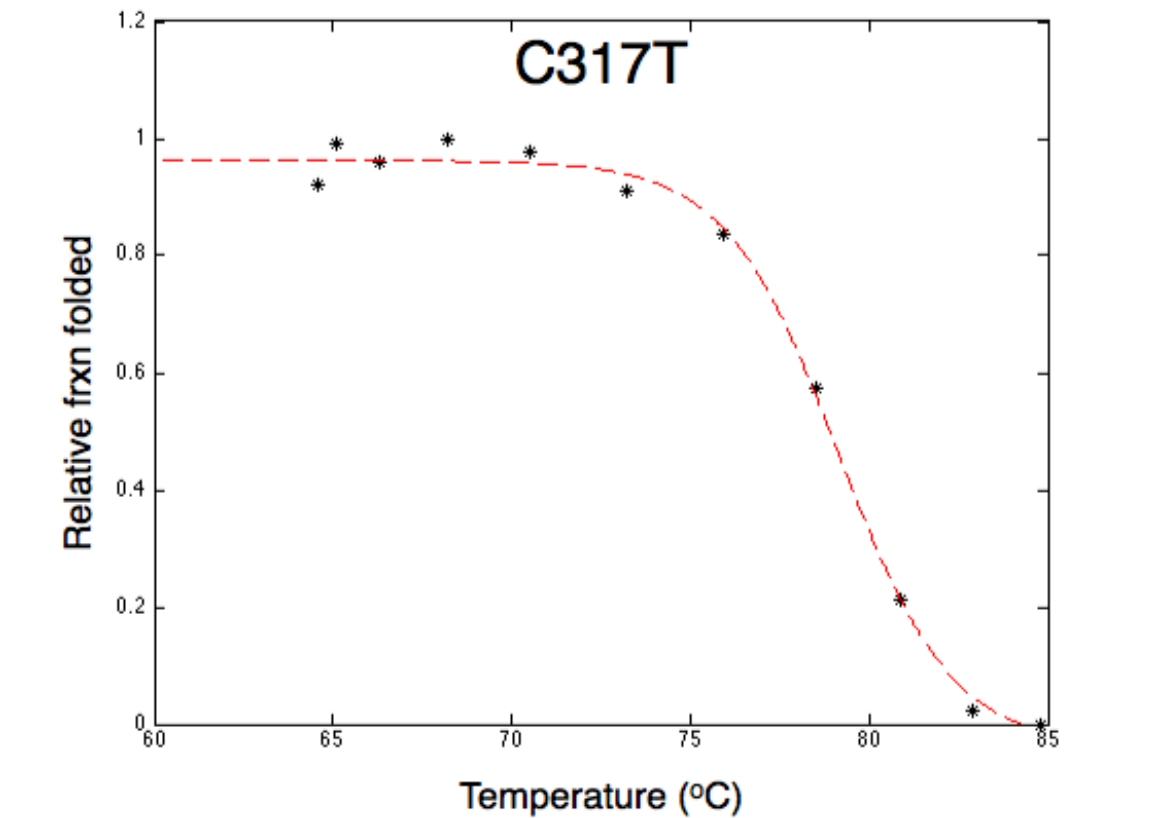
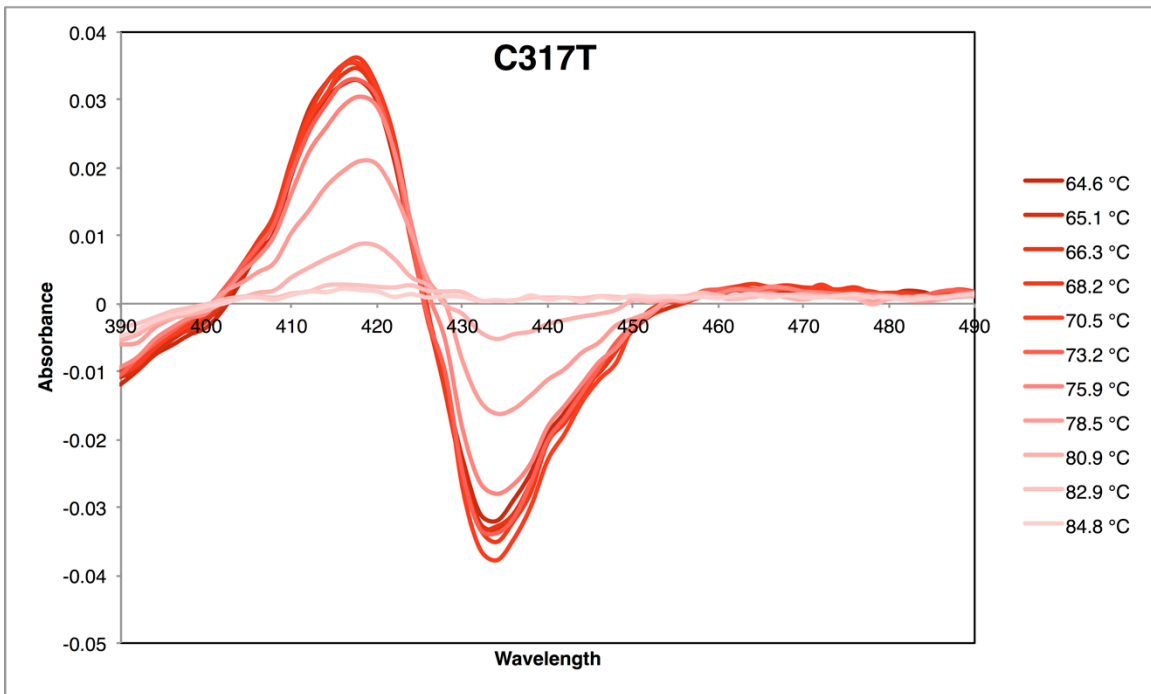


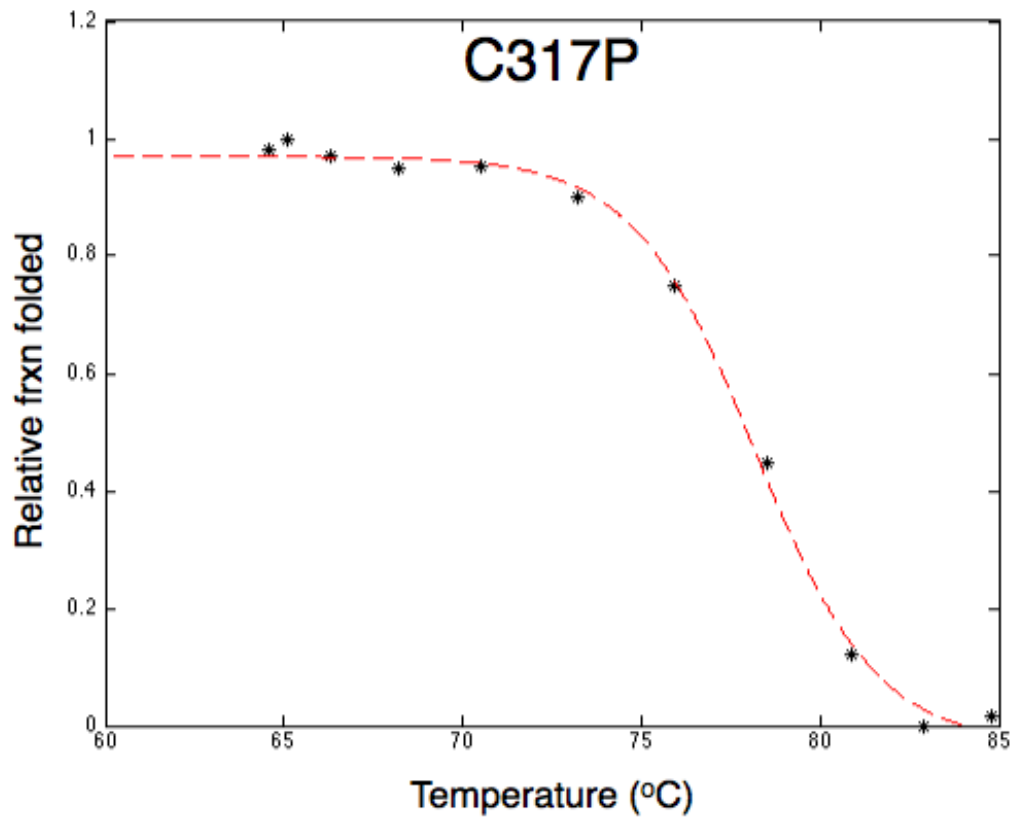
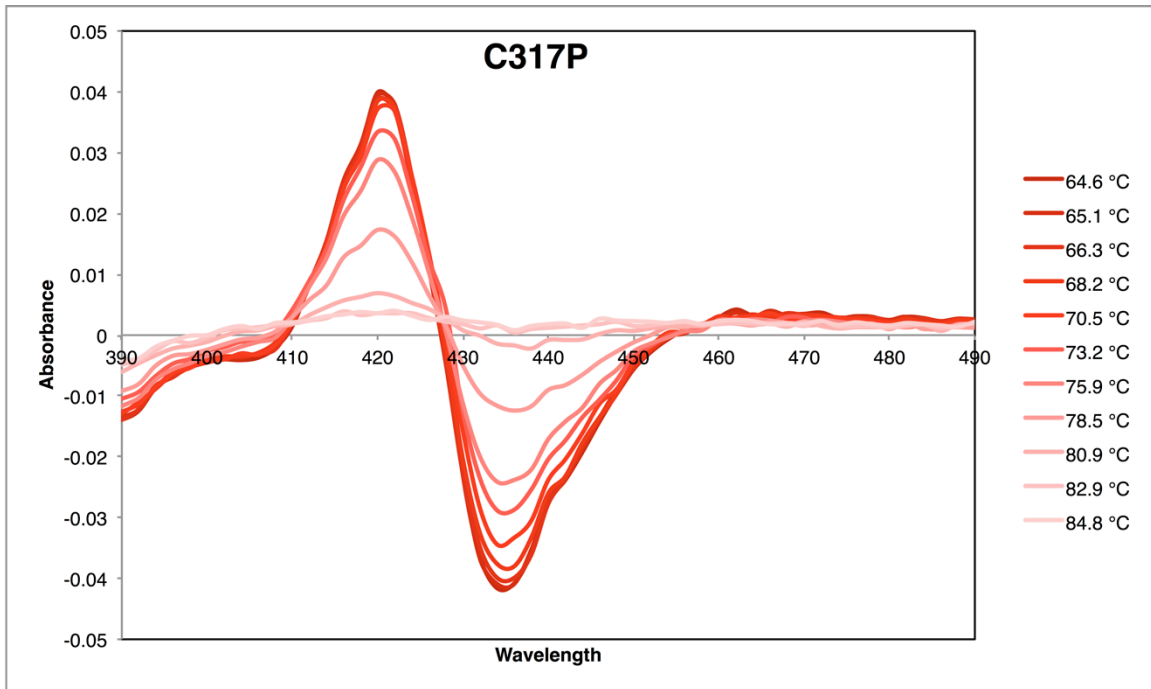


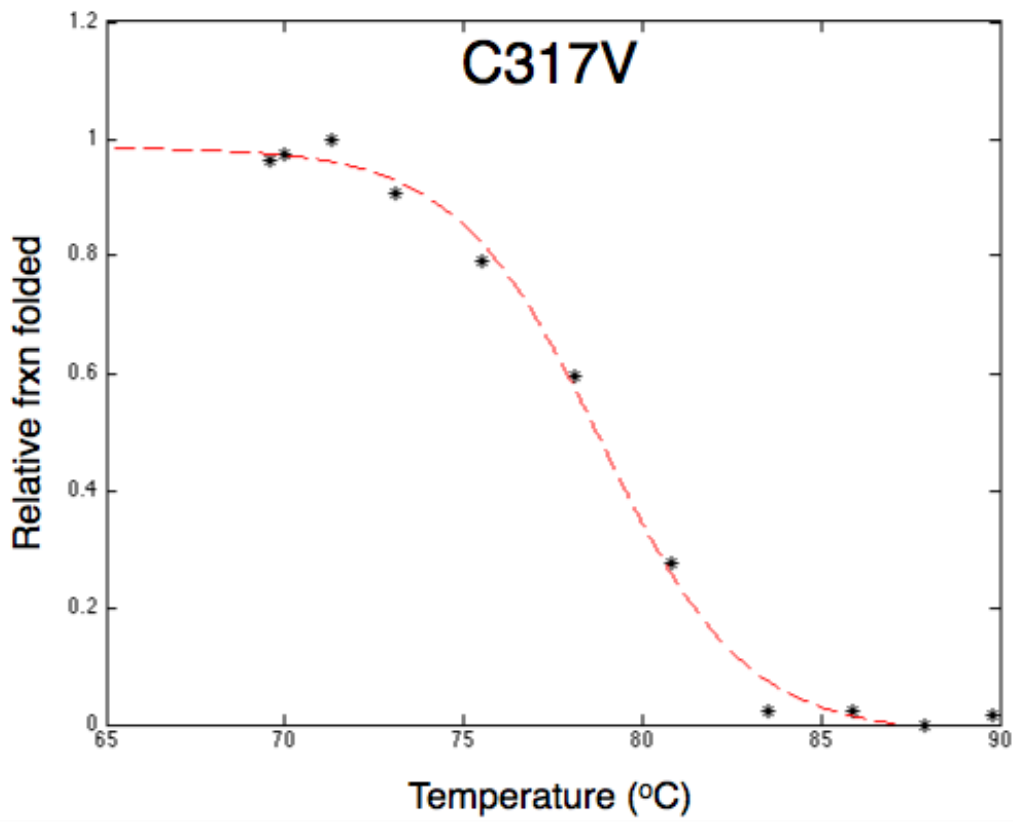
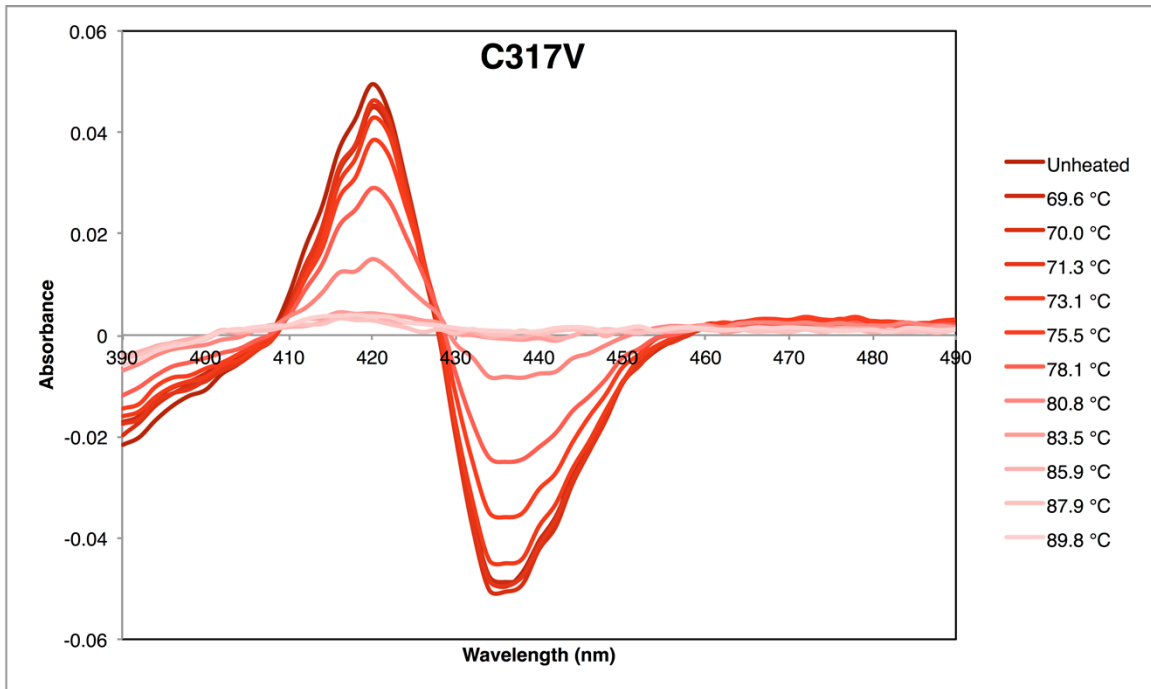




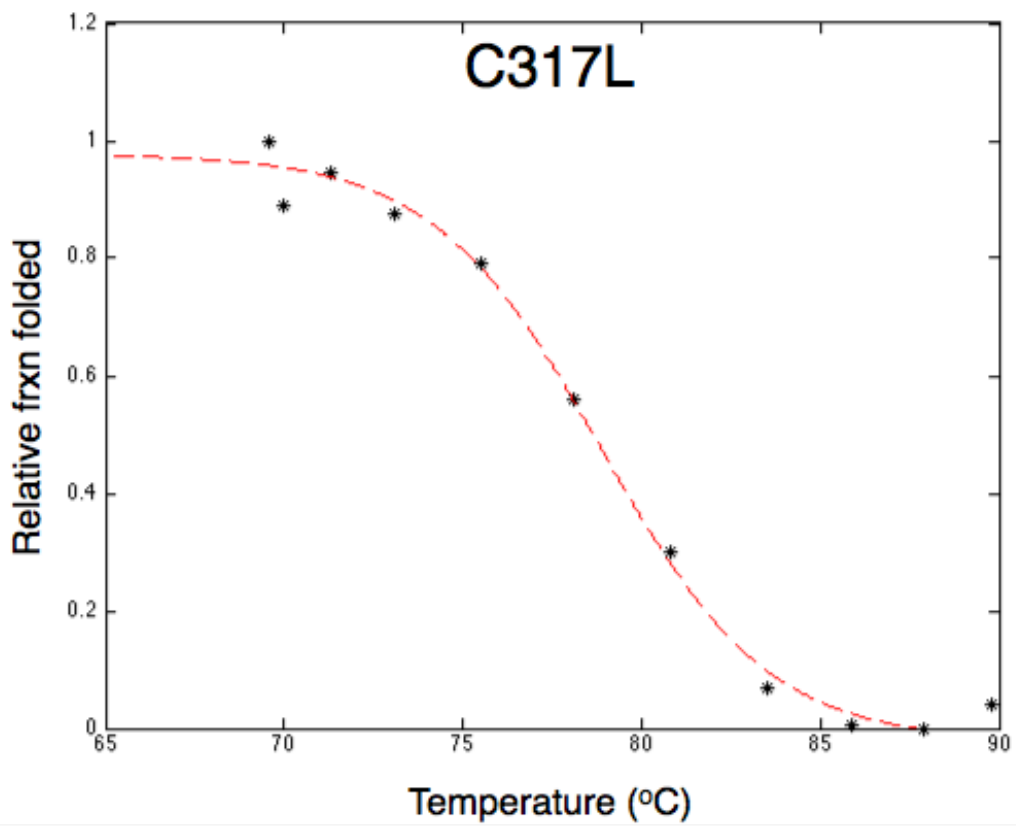
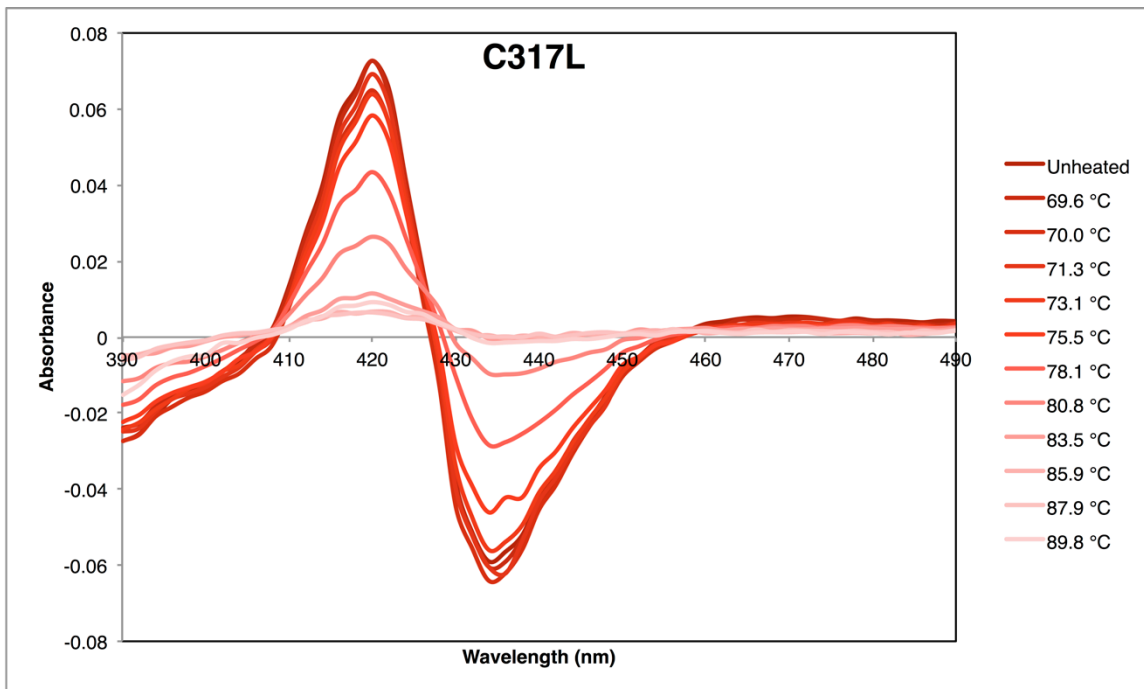


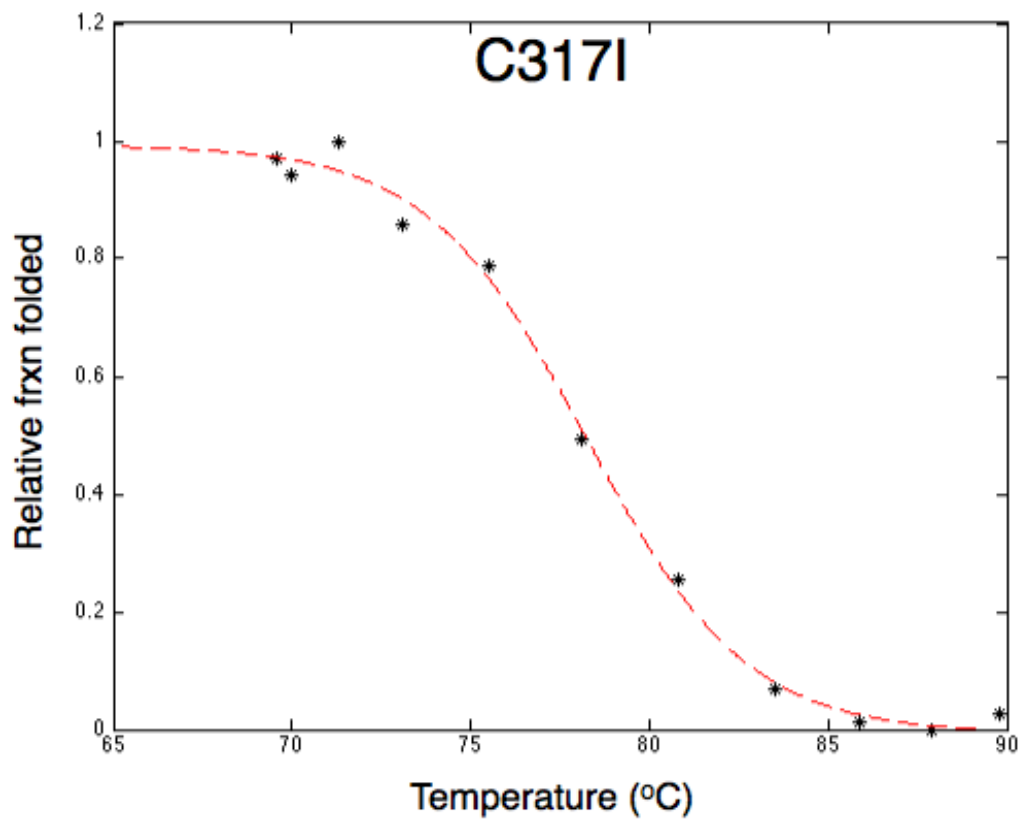
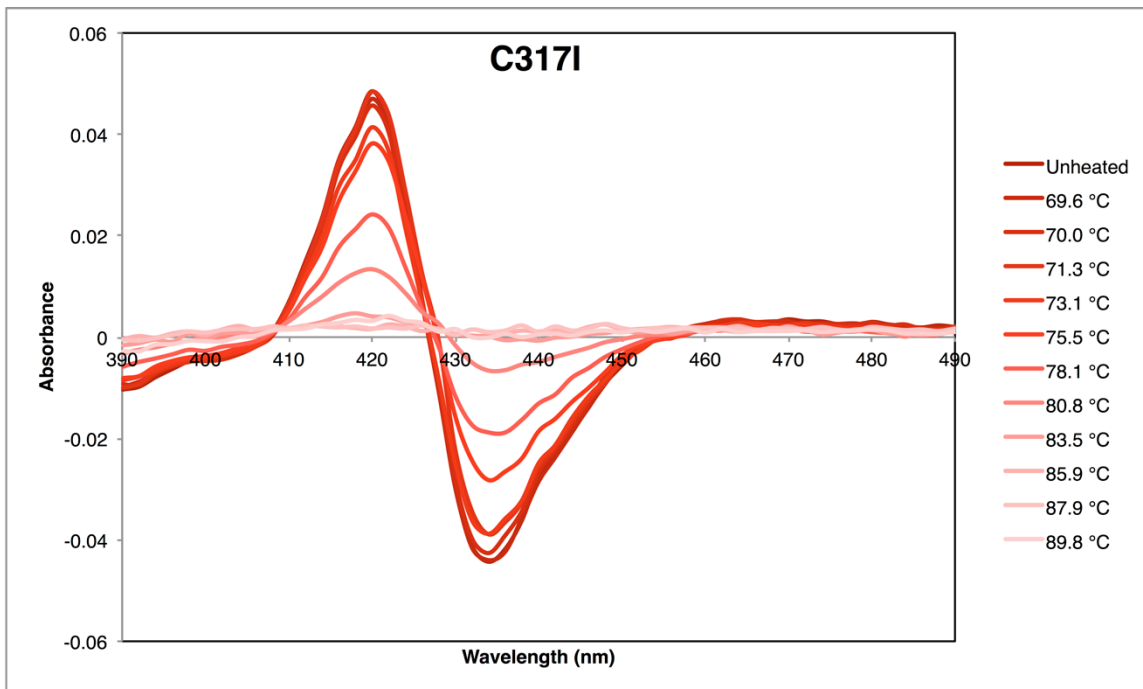


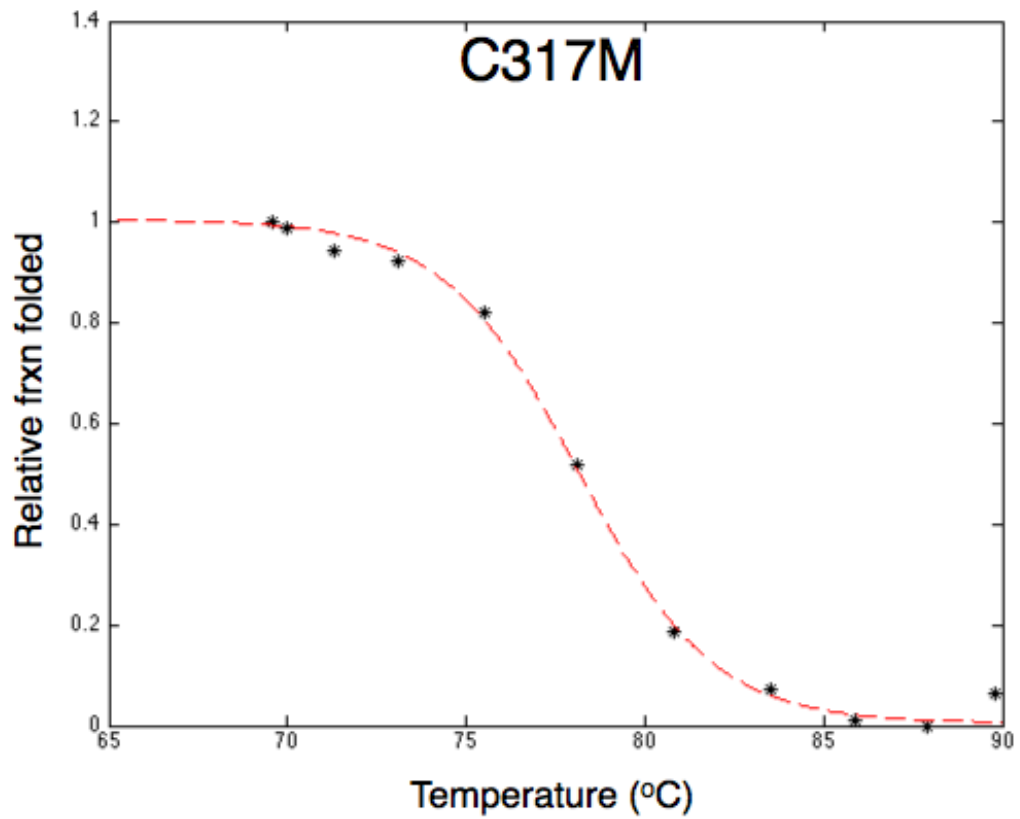
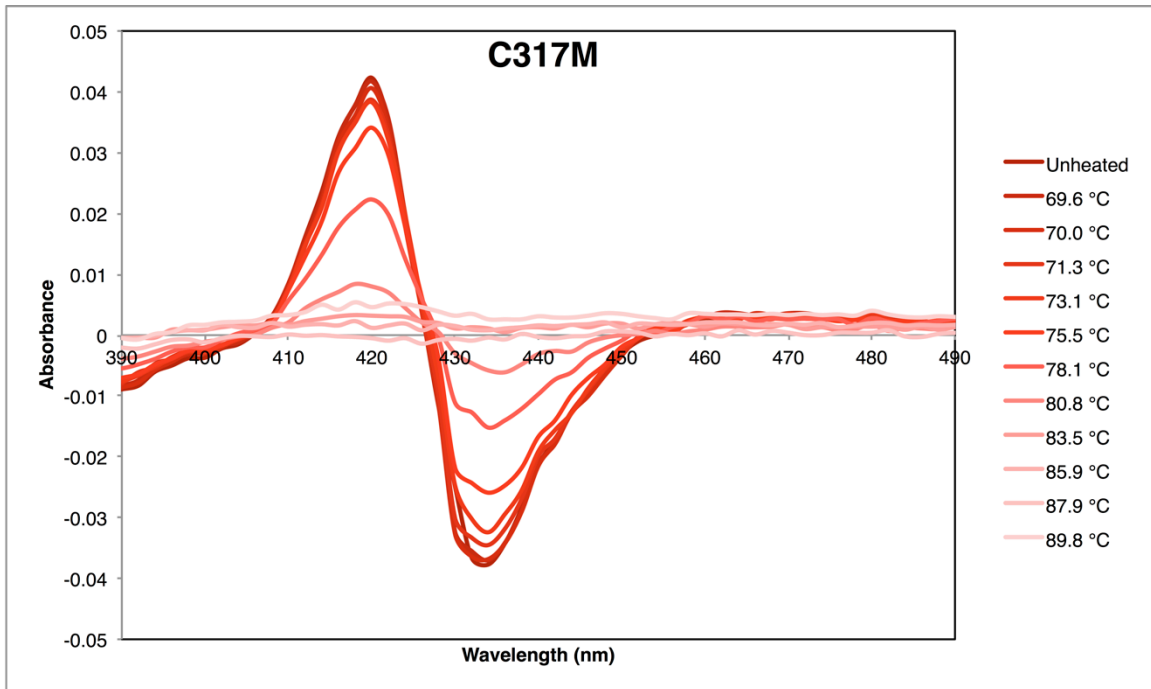


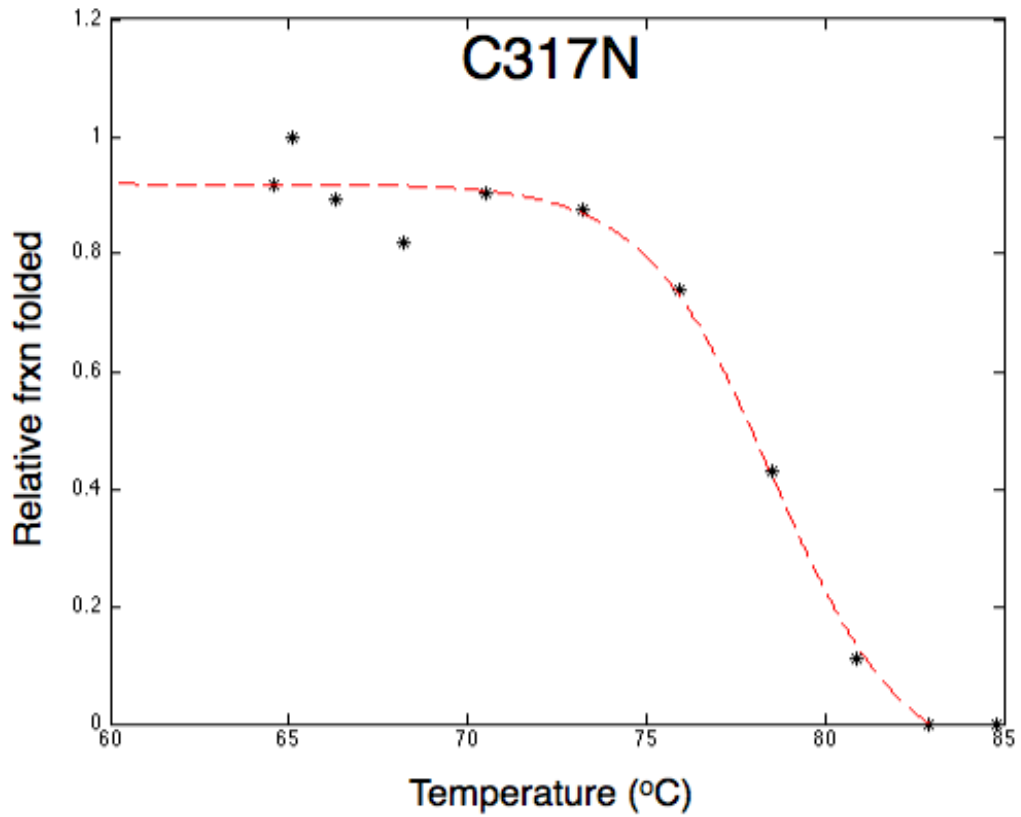
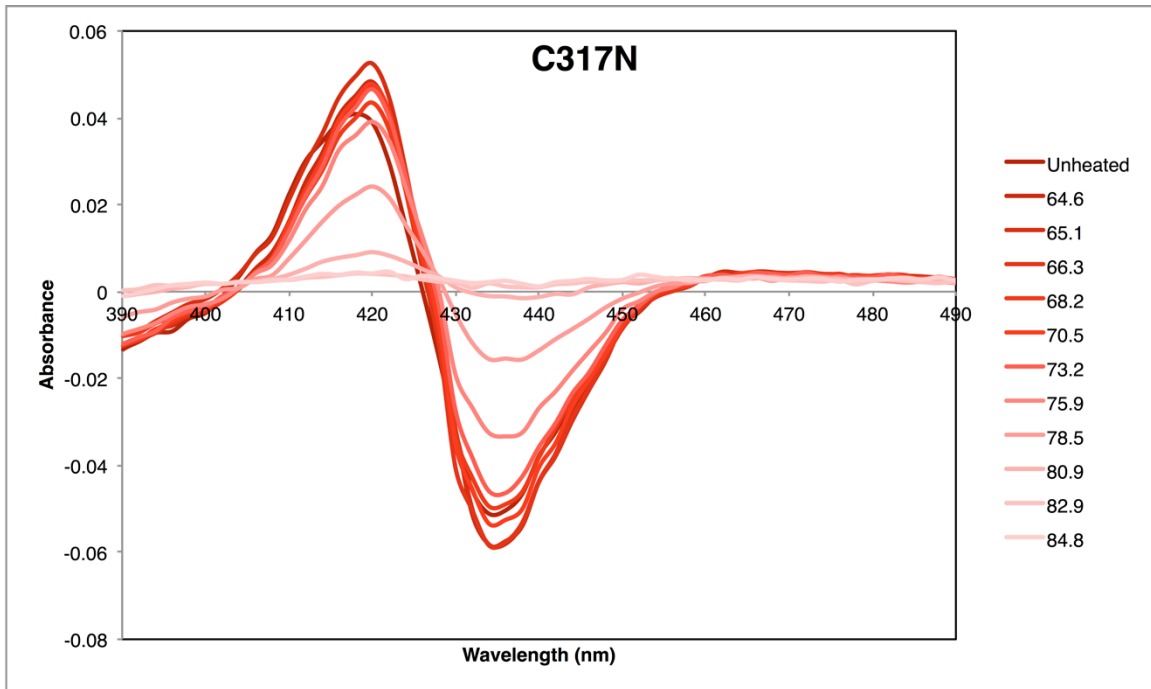


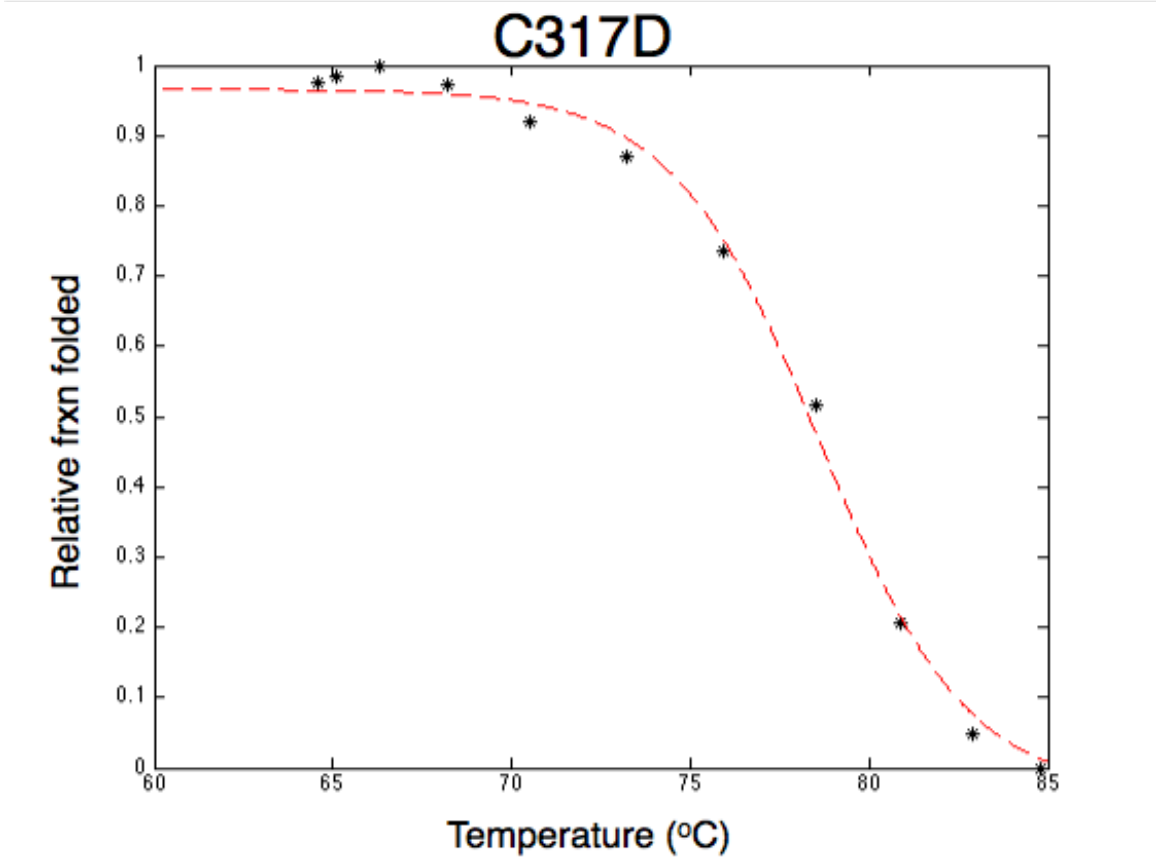
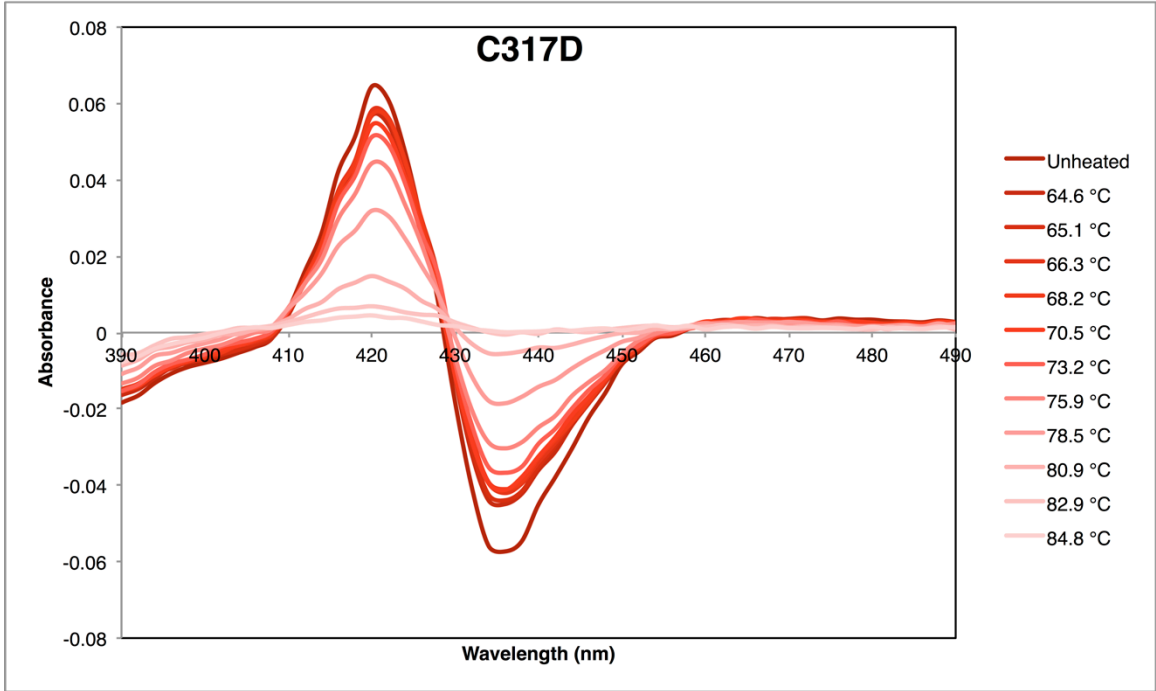


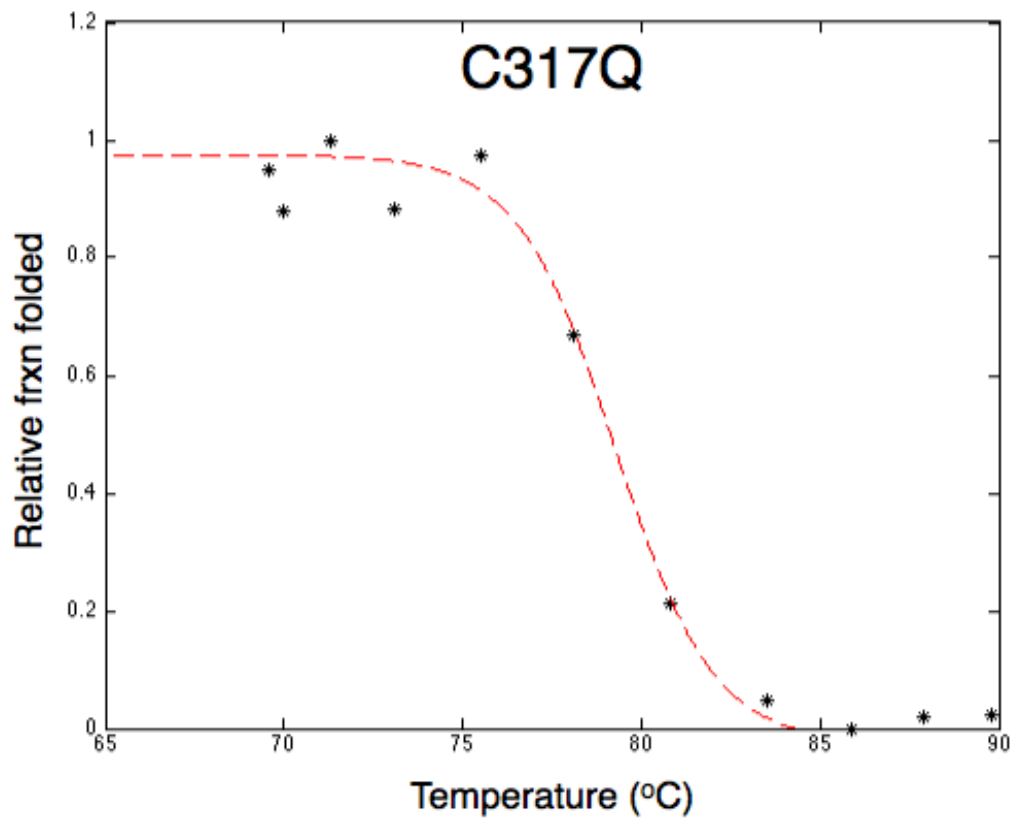
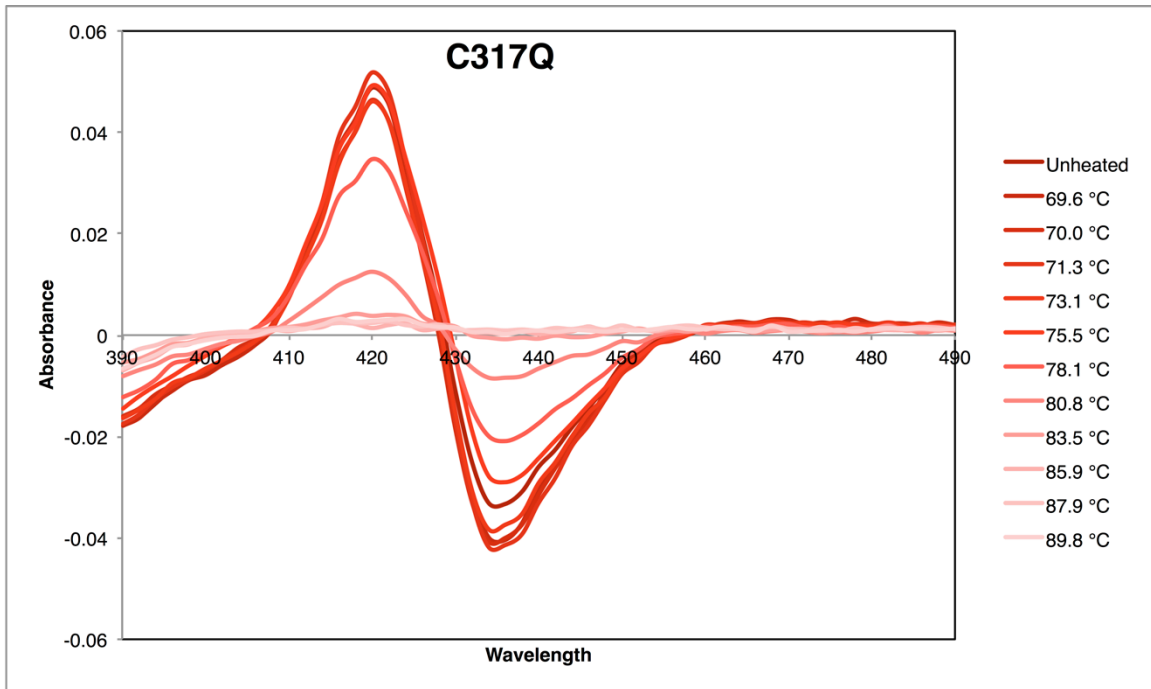


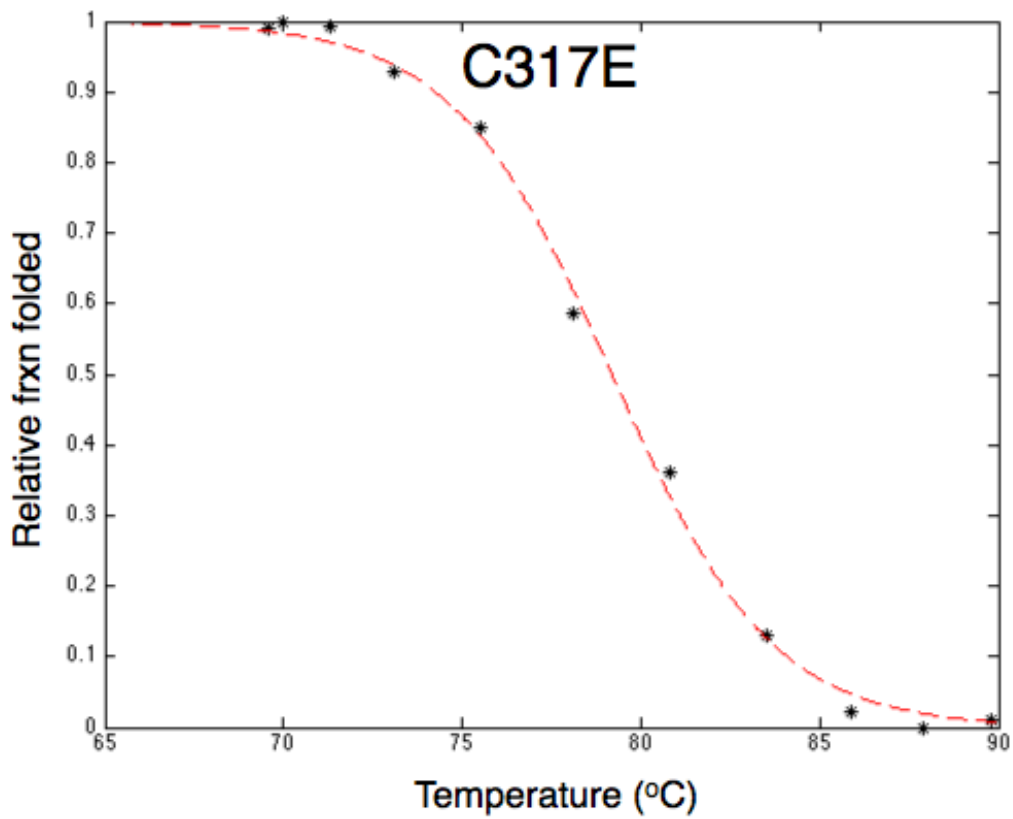
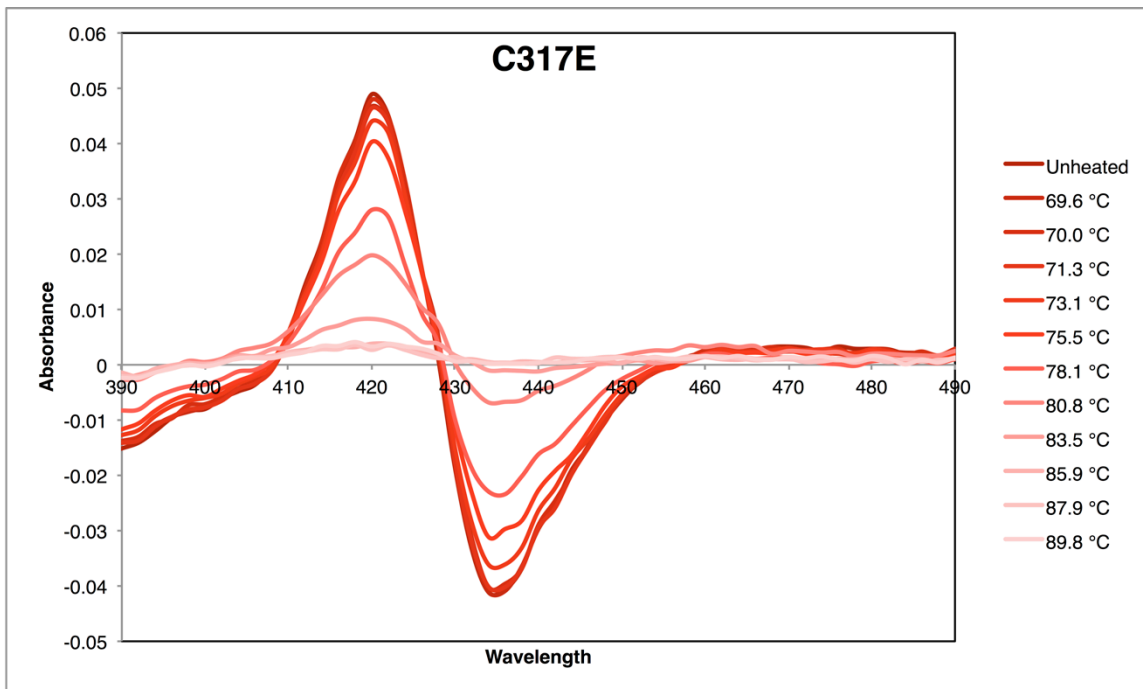


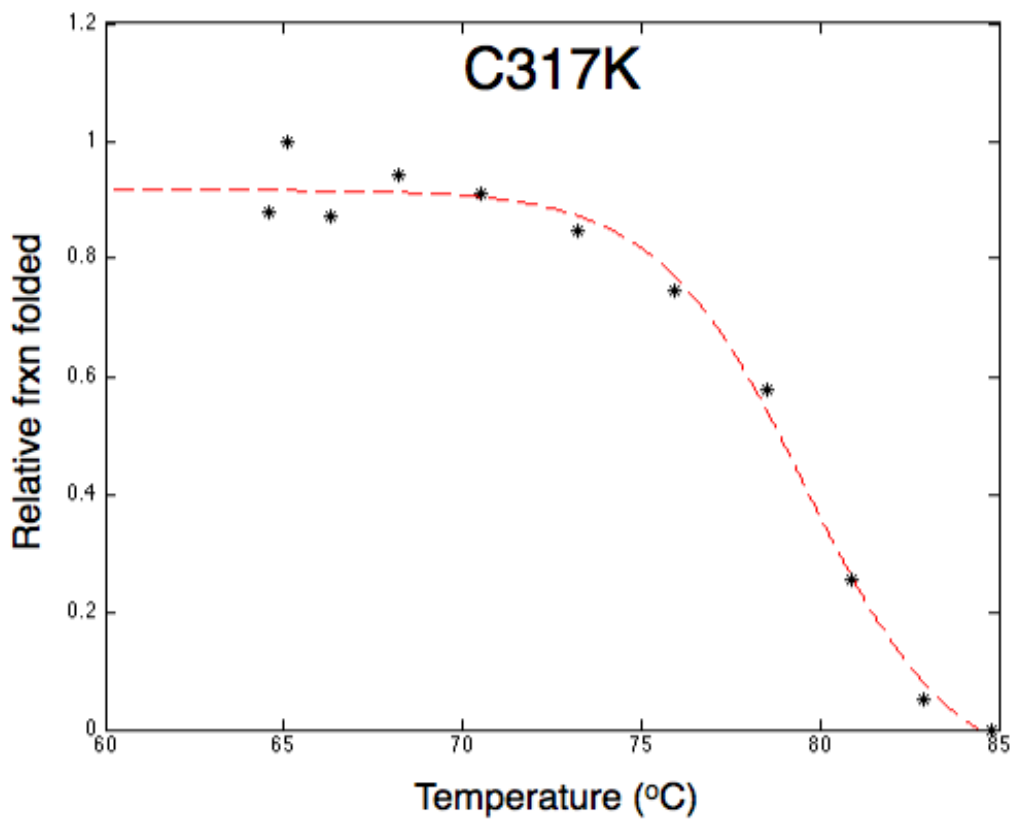
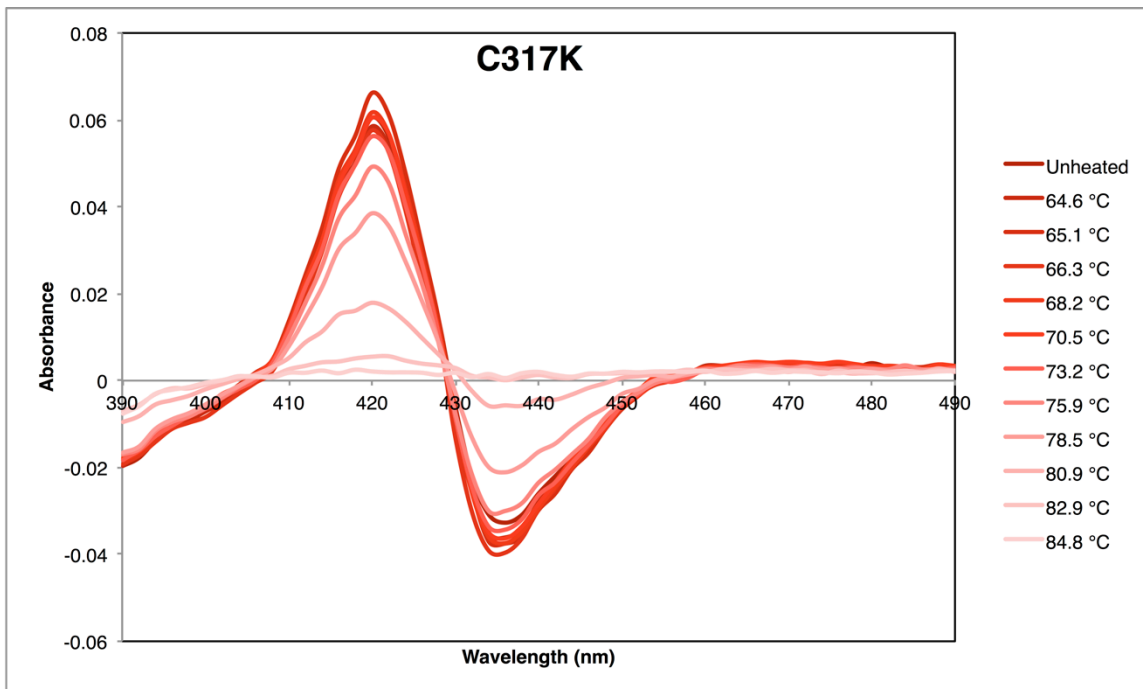




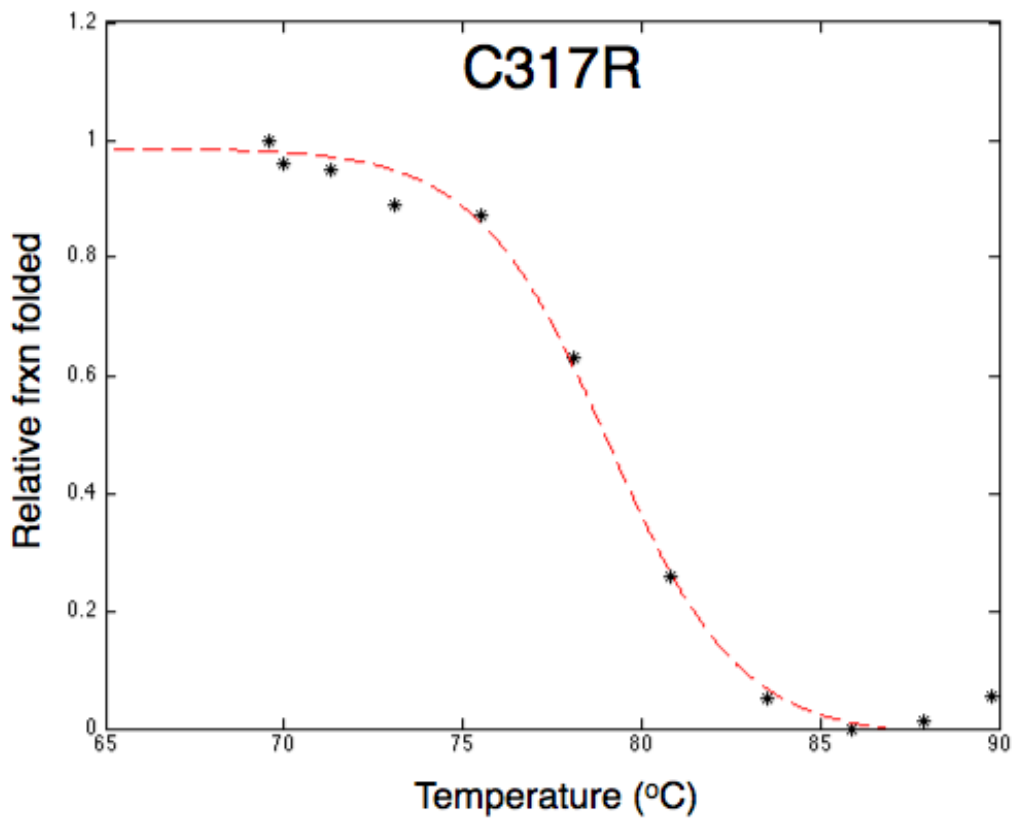
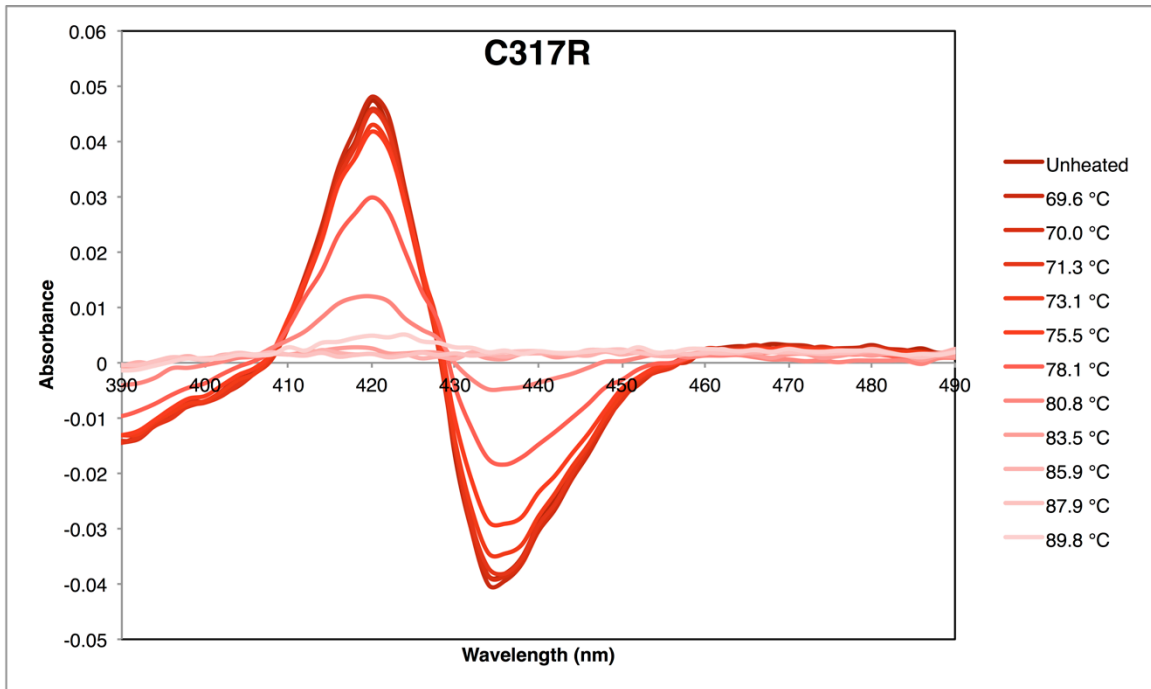


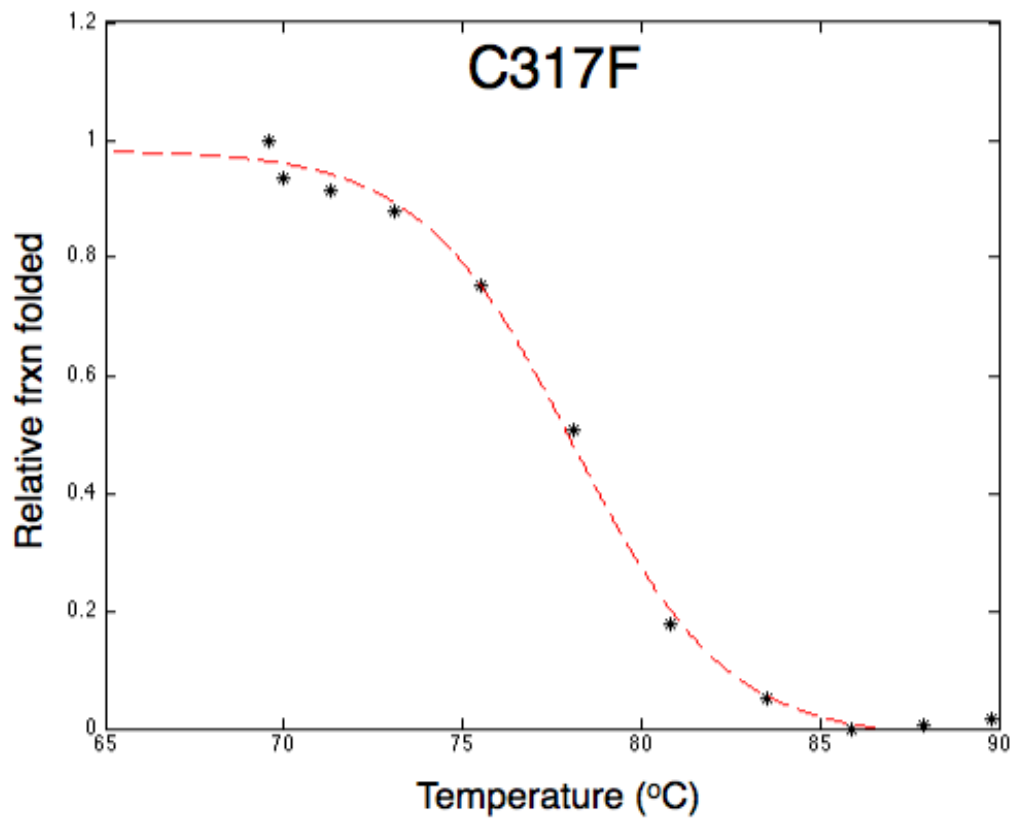
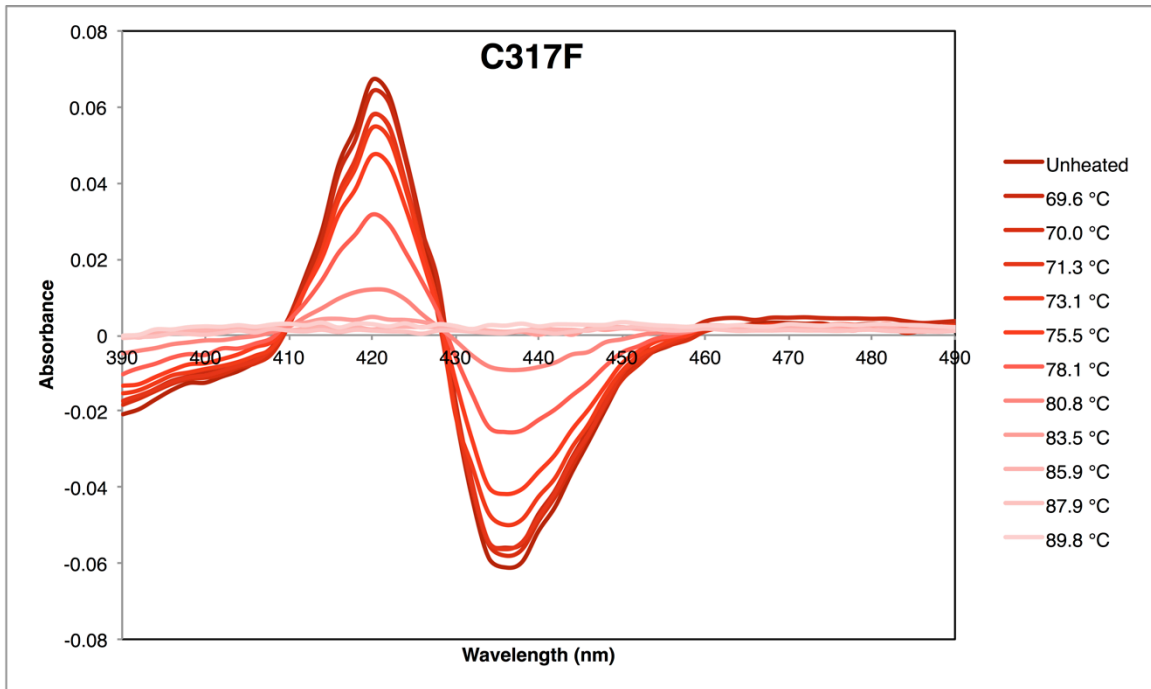


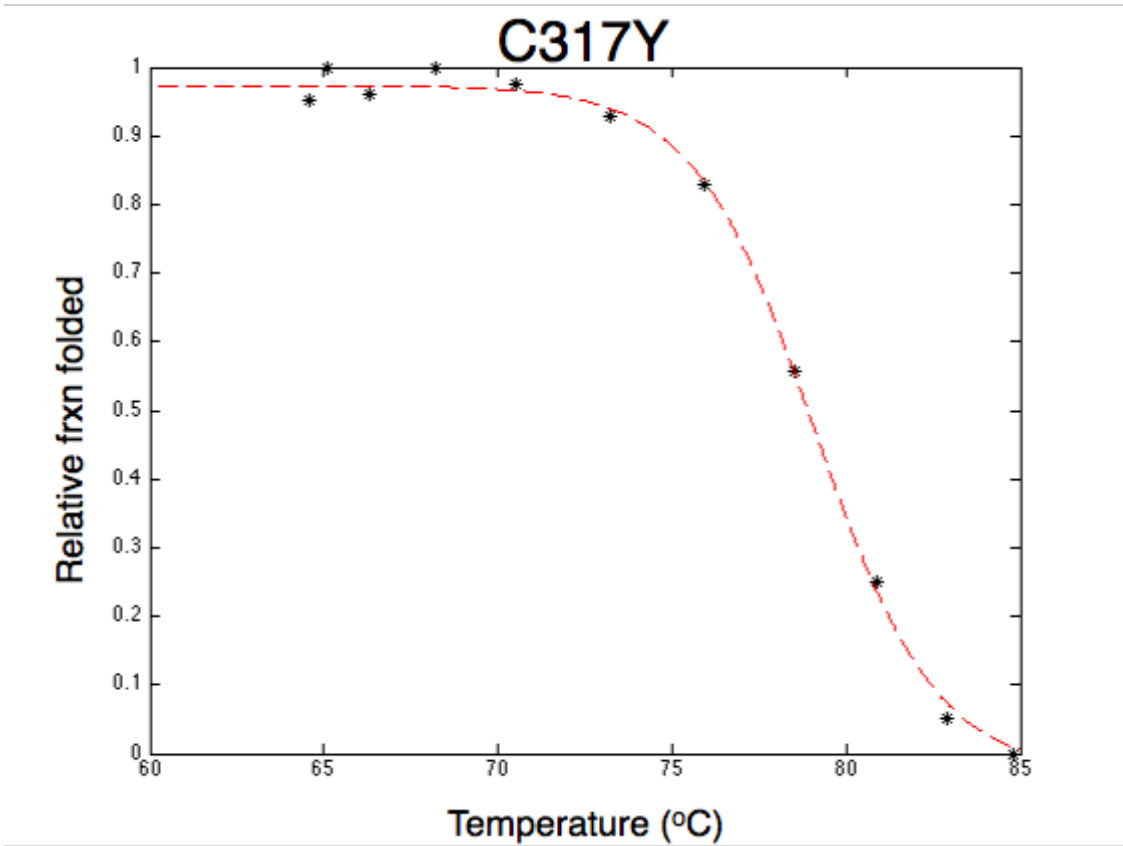
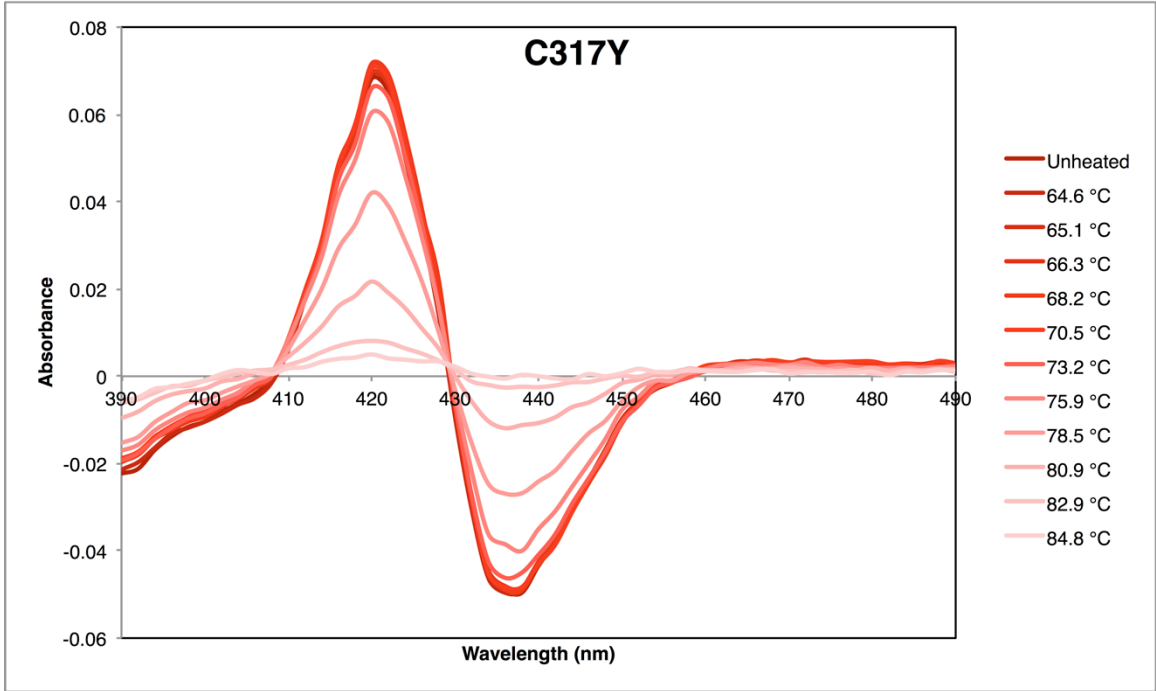


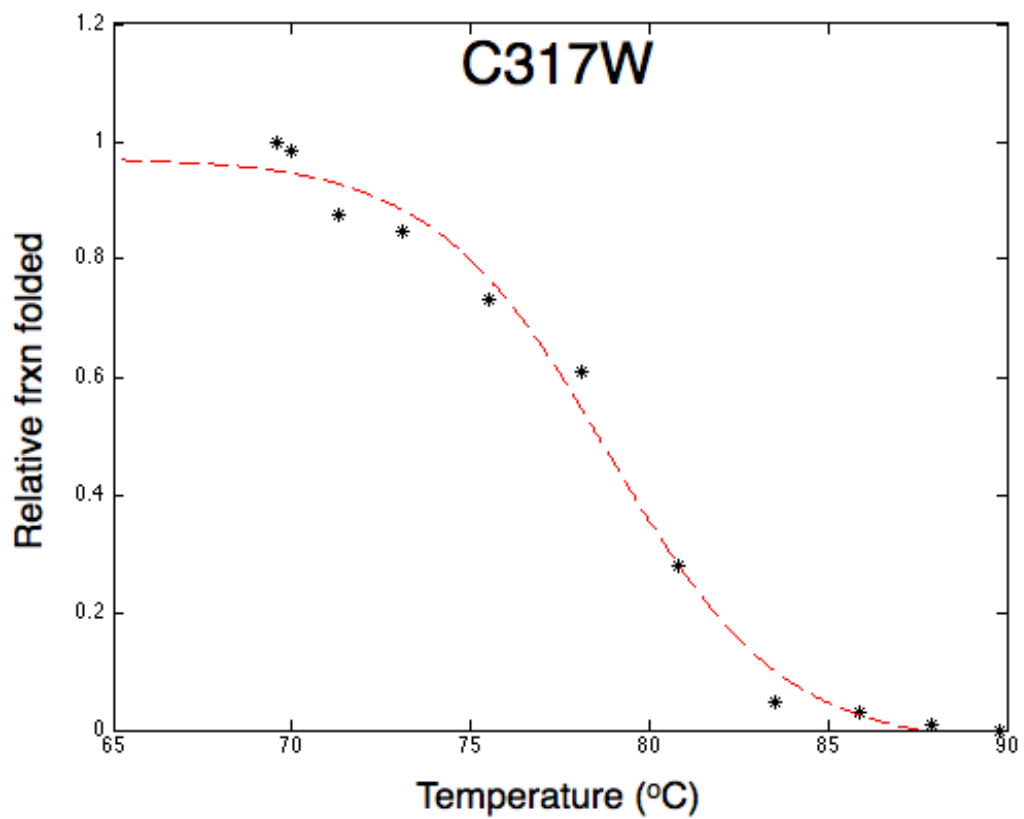
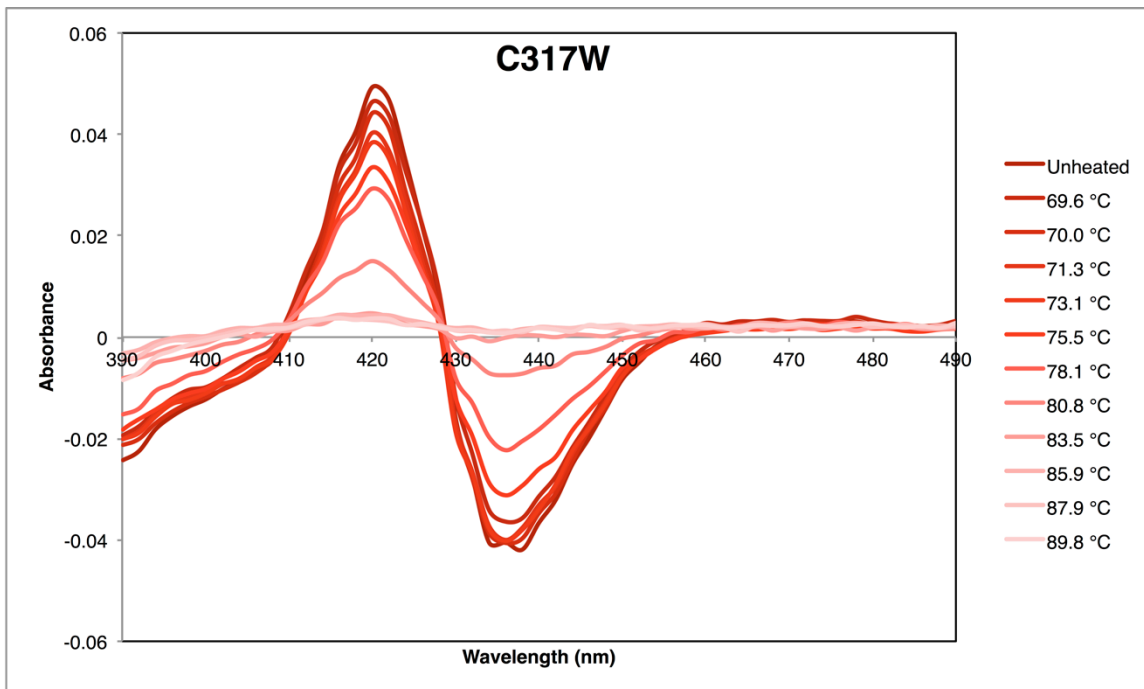












**Supporting references:**

- (1) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchinson, C. A.; Smith, H. O. *Nat. Methods* **2009**, *6*, 343.
- (2) Kabsch, W. *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66*, 125.
- (3) Karplus, P. A.; Diederichs, K. *Science* **2012**, *336*, 1030.
- (4) Evans, P. R.; Murshudov, G. N. *Acta Crystallographica Section D-Biological Crystallography* **2013**, *69*, 1204.
- (5) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Journal of Applied Crystallography* **2007**, *40*, 658.
- (6) Emsley, P.; Cowtan, K. *Acta Crystallographica Section D-Biological Crystallography* **2004**, *60*, 2126.
- (7) Winn, M. D.; Murshudov, G. N.; Papiz, M. Z. *Macromolecular Crystallography, Pt D* **2003**, *374*, 300.
- (8) Otey, C. *Methods Mol. Biol.* **2003**, *230*, 137.
- (9) Fehrhop, J. H.; Smith, K. M. *Porphyrins and metalloporphyrins*; Elsevier: New York, **1975**.