Characterization of Alkaline Transitions in

Ferricytochrome *c* Using Carbon-Deuterium IR Probes

Patrick Weinkam[‡§], Jörg Zimmermann[‡], Laura B. Sagle[‡], Shigeo Matsuda[‡],

Philip E. Dawson[‡], Peter G. Wolynes[§‖], Floyd E. Romesberg[‡*]

Contribution from Department of Chemistry[‡], The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA, 92037, and the Center for Theoretical Biological Physics, Department of Chemistry and Biochemistry[§], Department of Physics[‖], University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093.

AUTHOR EMAIL ADDRESS: floyd@scripps.edu

Deconvolution of the UV/vis Spectra

We deconvoluted the pH-dependant UV/vis spectra in order to determine the intrinsic UV/vis spectra for each state that is observed at alkaline pH: $\mathbf{III} \rightleftharpoons \mathbf{3.5} \rightleftharpoons \mathbf{IV_a} \rightleftharpoons \mathbf{IV_b} \rightleftharpoons \mathbf{V} \rightleftharpoons \mathbf{U}$. Using the equilibrium parameters listed in Table S1 and Eqn S1, we obtained fractional concentrations for each state (Figure S8). The calculated fractional amplitudes using Table S1 closely reproduce the fractional concentrations in Figure 8 in the main text. Therefore it is appropriate to use the parameters determined from fits of the observed transitions to determine fractional concentrations. The UV/vis spectra as a function of pH are calculated by: $\epsilon_{tot}(\lambda, pH) = \sum_i \epsilon_i(\lambda) c_i(pH)$ where $\epsilon_i(\lambda)$ is the intrinsic extinction coefficient of state i and $c_i(pH)$ is the calculated pH dependent fractional concentration of state i. The function $\epsilon_{tot}(\lambda, pH)$ was fit to many UV/vis spectra spanning pH 6.0 to 13.5 in which $\epsilon_{III}(\lambda)$ and $\epsilon_V(\lambda)$ were set equal to the spectra at pH 6.0 and 13.5 respectively and the values of $\epsilon_i(\lambda)$ for the four intermediates were freely varied. As discussed in the main text, the fractional amplitudes are calculated using the Boltzmann distribution with the highly alkaline state used as the reference state:

$$c_i(pH) = \frac{exp\left[\Delta G^o_{S_i \rightarrow U}/kT\right]}{\sum_j^{N_j} exp\left[\Delta G^o_{S_j \rightarrow U}/kT\right]} \qquad \text{(S1)}$$

Assignment of residue signals to global intermediates

Table S2 utilizes both experimental data and simulation results to assign the signals observed at specific residues to the conformational states of cytochrome *c*. Due to the structural localization of the conformational changes, no single residue shows unique signals corresponding to all six global states. However, using chemical reasoning and structures obtained from simulation we can determine whether a residue is solvent-exposed or buried in a given state. Most of the site-specific signals are sensitive to conformational transitions involving either deprotonation, a change in solvent exposure, or a change in ligation and are assigned to the involved states as discussed in the main text. In the case

of Lys72, the observed signals suggest that the residue is always solvent-exposed and that the observed deprotonation therefore has no consequence in the observed global conformational transitions. In the case of Met80, the I_2 signal was assigned to $\mathbf{IV_b}$ by comparing the solvent exposure of that residue between different structural ensembles calculated by simulation.

Alkaline transition of the protein with trimethylated lysines

We synthesized cyt *c* with trimethylated (N_ϵ-$(CH_3)_3$) lysines (Me_3-Lys) incorporated at positions 72, 73, and 79 and (d_3)methionine incorporated at Met80. The Me_3-Lys72/73/79-(d_3)Met80 protein should prevent heme misligation and allow us to monitor the alkaline transition. The signals observed at (d_3)Met80 were well fit by a sum of the folded and solvent-exposed signals for all pH values observed (Figure S9). The lack of an intermediate validates the assignment of the higher pH intermediate signal observed at Met80 in the wild-type protein, to a lysine-misligated intermediate. The FT IR signals and the 695 nm both band show transitions with a midpoint of pH 10.1 (Figure S9) as opposed to 8.8 in the wild-type protein. The shift of midpoints suggests that trimethylation perturbs the alkaline transition, but since the transition is well below the expected pK_a for a free lysine, another chemical entity may be playing a role in the trimethylated protein. The seemingly minor perturbation of trimethylation causes a significantly different heme ligation state as evident by the absorbance band at 600 nm which appears coincidental with the shift of the IR signal and the loss of the 695 nm signal (Figure S9). The absorbance band at 600 nm, which does not appear in the wild-type transition, is characteristic of a high spin heme and rises to a maximum around pH 11 and disappears by pH 13 where the UV/vis spectra is identical to the wild-type pH 13 spectra. The data suggests that an intermediate, unique from the wild-type intermediates, is populated between pH 10 and pH 12 where global unfolding occurs. More experiments are needed to fully understand the alkaline transitions of the Me_3-Lys72/73/79-(d_3)Met80 protein, but the results support the assignments of the Met80 signals in the wild-type transition.

Preparation of Nα-tert-butyloxycarbonyl-Nϵ-(2-chlorobenzyloxycarbonyl)-d_8-L-lysine

Basic cupric carbonate powder (629 mg, 2.85 mmol) was added to a solution of (d_8)lysine hydrochloride (Cambridge Isotope Laboratories, Inc.; 1.02 g, 5.37 mmol) in water (11 mL). The mixture was heated at reflux for 30 min, then filtered and washed with hot water (13 mL). To the blue filtrate was added sodium bicarbonate (480 mg, 5.71 mmol), 1,4-dioxane (10 mL), 2-chlorobenzyl chloroformate (0.986 mL, 6.44 mmol), and 4 M KOH (1.6 mL). The mixture was stirred overnight at 4^oC. The blue precipitate was collected, washed with cold water, ethanol and ether (each 10 mL), and vacuum dried to afford the copper salt of Nϵ-(2-chlorobenzyloxycarbonyl)-(d_8)L-lysine. This powder was added to a rapidly stirring suspension of ethylenediamine tetraacetic acid disodium salt (1.20 g, 3.22 mmol) in 0.33 M aqueous hydrochloric acid (40 mL). The mixture was stirred for 2 h and filtered. The filter cake was suspended in 9:1 (v/v) ethanol-water (30 mL), heated to reflux for 30 min, and filtered. The filtrate was stored overnight at 4^oC to give white solids. The mixture was filtered and washed with 90% ethanol to afford Nϵ-(2-chlorobenzyloxycarbonyl)-(d_8)L-lysine (781 mg, 2.42 mmol, 45% yield). ^1H NMR (400 MHz, MeOH-d_4) δ 7.33-7.36 (m, 1 H), 7.28-7.30 (m, 1 H), 7.19-7.22 (m, 2 H), 5.06 (s, 2 H), 3.40 (s, 1 H). HRMS (ESI-TOF) calculated for $C_{14}H_{11}N_2O_4Cl_1D_8$ (MH^+) 323.160, found

323.1601. Without further characterization, to a solution of N-(2-chlorobenzyloxycarbonyl)-(d_8)L-lysine (763 mg, 2.37 mmol) in 1,4- dioxane (5 mL) and water (2.5 mL) was added 1M NaOH aq. (2.37 mL) and di-tert-butyl dicarbonate (570 mg, 2.61 mmol). The mixture was stirred for 3 h at room temperature and evaporated to remove dioxane. The remained aqueous solution was adjusted to pH 2.3 with 10% citric acid. The solution was extracted with EtOAc (30 mL x 3). The collected EtOAc was washed with H_2O (50 mL), dried over anhydrous Na_2SO_4, filtered and evaporated to give Nα-tert-butyloxycarbonyl-Nϵ-(2-chlorobenzyloxycarbonyl)-(d_8)lysine (900 mg, 2.13 mmol, 90% yield) as a white foam. ^1H NMR (500 MHz, MeOH-d_4) δ 7.33-7.35 (m, 1 H), 7.27-7.28 (m, 1 H), 7.18-7.21 (m, 2 H), 5.06 (s, 2 H), 3.96 (s, 1 H), 1.33 (s, 9 H). ^{13}C NMR (125 MHz, MeOH-d_4) δ 176.2, 158.5, 158.0, 135.8, 134.0, 130.4, 130.3, 128.1, 80.4, 64.5, 54.6, 28.7. HRMS (ESI-TOF) calculated for $C_{19}H_{19}N_2O_6Cl_1D_8$ (MNa^+) 445.1944, found 445.1951. This material was used for peptide synthesis without further purification.

pH Correction

A transition curve for the 695 nm absorption was obtained using higher volume samples of cyt c under the same solvent conditions as the FT IR experiment such that more precise pH measurements could be taken. Using a sigmoidal function fitted to this data (with parameters n(695) and $\mathsf{pH_m}$(695)), the 695 nm absorption can be used to determine the pH of the FT IR samples. However, the pH correction can only be reliably performed close to the transition midpoint around which the absorption is most sensitive to changes in pH. Thus the pH for the FT IR samples was determined using the following equation employing a weighted average of pH as measured using the 695 nm absorption ($pH(695)$) and pH measured using the pH probe ($pH(probe)$).

$$pH = exp\left[\frac{(-2.3n_{695}kT(pH_{probe}-pH_{m,695}))^6}{20kT}\right]pH(695)+$$
$$(1-exp\left[\frac{(-2.3n_{695}kT(pH_{probe}-pH_{m,695}))^6}{20kT}\right])pH(probe) \qquad \text{(S2)}$$

Simulations of the Misligated States

Our simulations are based on the associative memory Hamiltonian (AMH) [1, 2]. By using only a single "memory" protein conforming to the native X-ray structure, we ensure a perfect funnel. Details of the simulations were carried out almost identically to previous work [3]. Contact potentials are represented by Gaussian well potentials similar to the protein-protein contacts described in earlier studies [4]: $\epsilon_{ij}(r_{ij}) = -\left|\frac{\epsilon}{a}\right|^{1/p}\boldsymbol{\gamma}_{ij}exp\left[\frac{-(r_{ij}-r_{ij}^{nat})^2}{2\sigma_{ij}^2}\right]$. The indices i and j run over all of the C$^\alpha$ and C$^\beta$ atoms for protein-protein interactions and all heme and C$^\alpha$ atoms for the protein-heme interactions. We define interactions based on distances in the crystal structure, r_{ij}^{nat}. These occur between atoms with $r_{ij}^{nat} < 8\mathring{A}$ for protein-protein interactions and $r_{ij}^{nat} < 12\mathring{A}$ for protein-heme interactions. The value of $\boldsymbol{\gamma}_{ij}$ is uniformly set to 0.290 which results in a homogeneous contact potential and r_{ij} is the distance between atoms i and j. The contact variance σ_{ij} is equal to $|i-j|^{0.3}$ for protein-protein interactions and $|12|^{0.3}$ for protein-heme interactions. The interaction energies are scaled to account for non-additivity as in the study of Eastwood and Wolynes [4]. The scaling factor a is used to maintain that the unit of energy is $1kT$: $\frac{1}{8N}\sum_i\left|\sum_j\boldsymbol{\gamma}_{ij}\right|^p$. The energy per residue is given by: $E_i = \sum_j\epsilon_{ij}$. Non-additivity comes about because of the cross terms in the expression $E^{na} = -\frac{1}{2}\sum_i|E_i|^p$. There are

no cross terms for the $p = 1$ case which corresponds to a purely additive potential. The cross terms for the $p = 2$ case allows for three body interactions and in general there are $(p + 1)$ body interactions accounted for in the energy function. Non-additivity was added by setting $p = 1.3$ so that about 40 percent of the total energy is due to three body interactions.

References

[1] Friedrichs, M. S.; Wolynes, P. G. *Science*, **1989**, *246*, 371–373.

[2] Sasai, M.; Wolynes, P. G. *Phys. Rev. A*, **1992**, *46*, 7979–7997.

[3] Weinkam, P.; Zong, C.; Wolynes, P. G. *Proc. Natl. Acad. Sci. USA*, **2005**, *102, 35*, 12401–12406.

[4] Eastwood, M. P.; Wolynes, P. G. *J. Chem. Phys.*, **2001**, *114*, 4702–4716.

Table S1: Fit Parameters for IR Absorptions

	First Gaussian			Second Gaussian		
	Relative Amplitude	Frequency (cm^{-1})	FWHM (cm^{-1})	Relative Amplitude	Frequency (cm^{-1})	FWHM (cm^{-1})
Leu68						
Neutral/buried (N)	1.00	2215.5	13.9	0.85	2204.0	13.9
Alkaline/solvent exposed(A)	1.00	2213.0	12.9	0.45	2227.0	13.9
Lys72						
Alkaline/deprotonated/ solvent exposed (A)	1.00	2108.0	17.5	0.24	2120.0	13.4
Alkaline/deprotonated/ solvent exposed (A)	1.00	2104.0	20.7	0.52	2215.0	19.0
Lys73						
Neutral/protonated/ solvent exposed (N)	1.00	2107.8	18.2	0.21	2120.0	10.5
Alkaline/deprotonated/ solvent exposed (A)	1.00	2103.0	18.2	0.60	2115.5	18.6
Intermediary/ heme misligated (I)	1.00	2100.0	16.2	NA	NA	NA
Lys79						
Neutral/protonated/ solvent exposed (N)	1.00	2110.7	20.1	0.26	2122.0	9.9
Alkaline/deprotonated/ solvent exposed (A)	1.00	2104.7	21.4	0.37	2115.4	19.7
Intermediary/ heme misligated (I)	1.00	2100.0	16.2	NA	NA	NA
Met80						
Neutral/buried/heme ligated (N)	1.00	2127.8	7.8	NA	NA	NA
Intermediary+Alkaline/ solvent exposed (I_1, A)	1.00	2134.5	12.3	NA	NA	NA
Intermediary/buried (I_2)	1.00	2127.8	9.9	NA	NA	NA

[a]Gaussians have the form $A*\exp[-(x-F)^2/2(W/2.35482)^2]$ where F is the frequency (cm^{-1}), W is the FWHM (cm^{-1}), and A is the relative amplitude

Table S2: Assignment of Site-Specific Signals to Observed Transitions

Site Specific Signals[a]	Observed States					
	III	3.5	IV_a	IV_b	V	U
Leu68	N	N	N	N	N	A*
Met80	N	I_1*	I_1*	I_2	A*	A*
Lys72	N*	N*	N*⇌A*	N*⇌A*	A*	A*
Lys73	N*	N*	I	N*⇌A*	A*	A*
Lys79	N*	N*	N*⇌A*	I	A*	A*
695 nm Abs	N	A	A	A	A	A

[a]Signals correspond to the folded (N), high pH, solvent-exposed (A), and any intermediate states (I). The equilibrium between N and A implies that there is a balance of protonation states in a given intermediate but it is unclear which protonation state dominates in that conformation. *indicates solvent exposed signals

Table S3: Parameters used for the decomposition of the UV/vis spectra.

Observed Transitions	Residue Signals[a]	pH_m	n
III ⇌ 3.5	$(N \rightleftharpoons I_1)^{695}$	8.8	1.2
3.5 ⇌ IV_a	$(N \rightleftharpoons I)^{Lys73}$	10.5	0.5
IV_a ⇌ IV_b	$(N \rightleftharpoons I)^{Lys79}$ – $(N \rightleftharpoons I)^{Lys73}$	NA[b]	0.1
IV_b ⇌ V	$(I \rightleftharpoons A)^{Lys79}$	9.0	0.4
V ⇌ U	$(N \rightleftharpoons A)^{Leu68}$	12.7	1.3

[a]Certain residue-specific transitions were selected to approximate the transitions in the protein. [b]Due to the similar pH dependence of IV_a/IV_b, the midpoint value is calculated using the equation $\Delta G/2.3nkT$, where ΔG is determined using the Boltzmann equation and the relative concentrations of IV_a/IV_b at pH 10.5, at which the intermediates are at maximal concentration

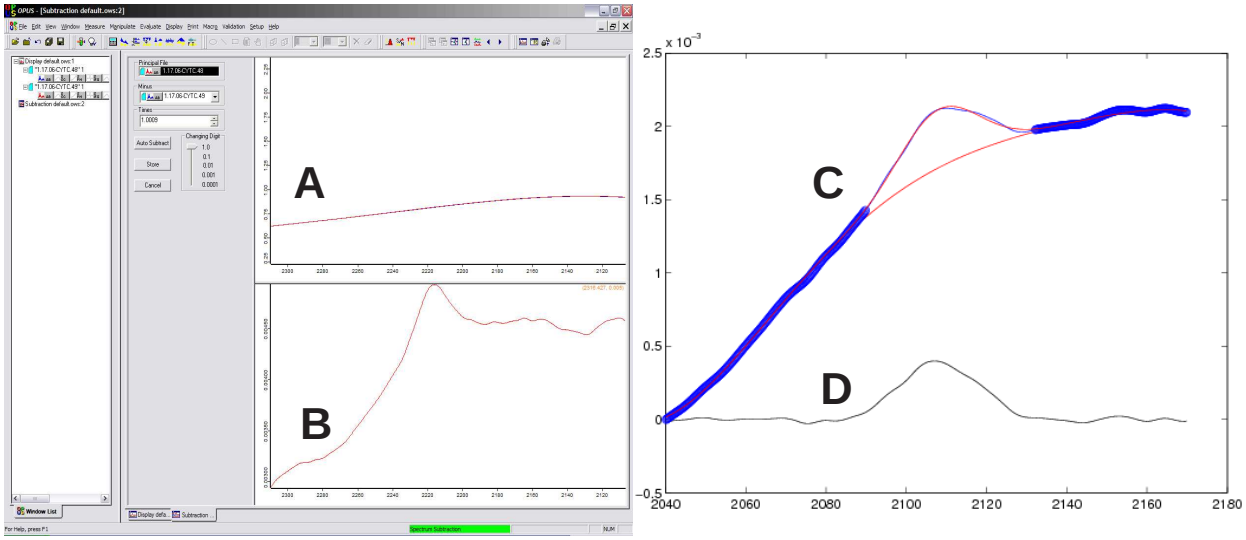

Figure S 1: For any given residue, the FT IR absorption spectra of non-deuterated samples were subtracted from those of the deuterated samples to ensure flatness in non-saturated regions of the spectra using the OPUS program from Bruker. **A** Screenshot of the OPUS program from Bruker used for subtraction of the proteo spectrum from the deuterated (Lys79) spectrum at pH = 7. **B** Difference spectrum resulting from using a subtraction factor of 1.0009. **C** Matlab was used to fit the sum of a 5th order polynomial plus a Gaussian to the difference spectrum. The blue lines correspond to the difference spectrum with dark, thick blue lines used as a visual reference. The red line below the Gaussian is the 5th order polynomial used for subtraction while the red line fit to the Gaussian represents the sum of the 5th order polynomial and the the Gaussian. **D** After verification that the 5th order polynomial is smooth under the Gaussian, subtraction results in the final spectrum.

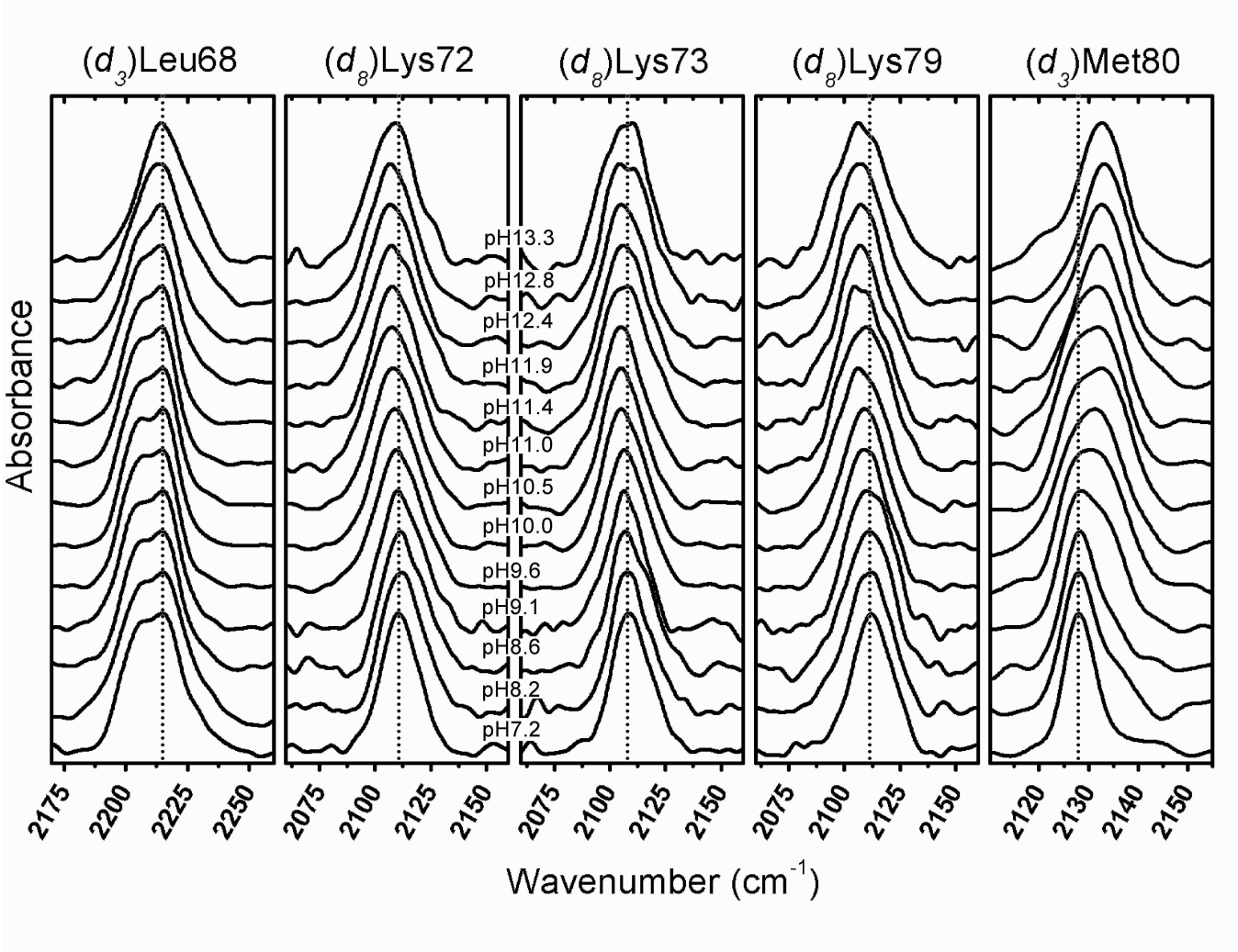

Figure S 2: Spectra of the observed carbon-deuterium stretching absorptions for d_3-Leu68, d_8-Lys72, d_8-Lys73, d_8-Lys79, and d_3-Met80 at different pH values.

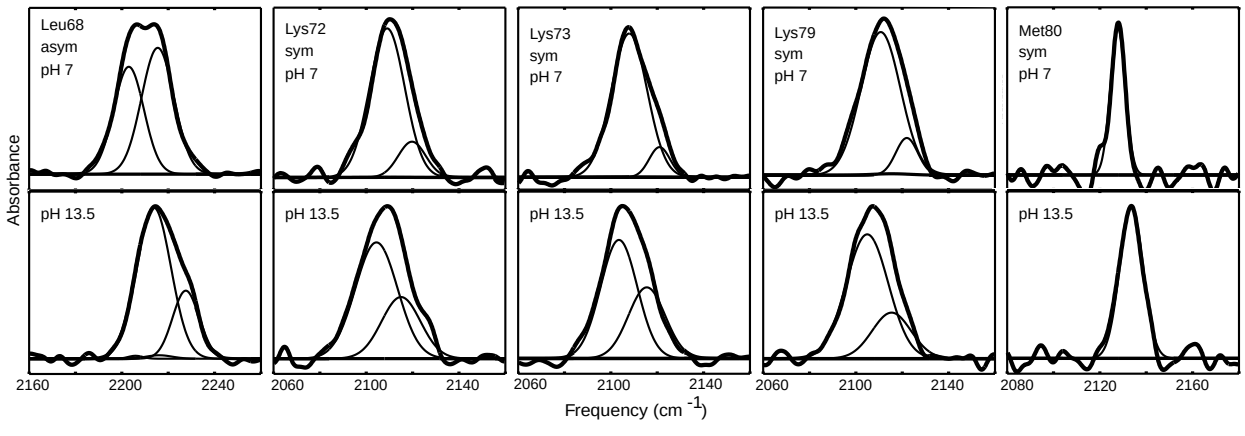

Figure S 3: Gaussian deconvolution of the IR absorption spectra obtained from the indicated deuterium-labeled residues in pH 7 (upper panels) and highly alkaline (lower panels) conditions (sym/asym refers to the symmetric/asymmetric stretch).

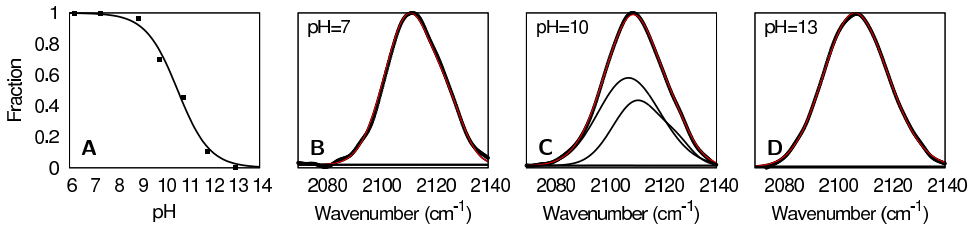

Figure S 4: The assignment of the transition observed at (d_8)Lys72 to deprotonation is supported by its overall similarity to the transition observed for a boc-protected, deuterated free lysine in solution. **A** Fractional concentrations of the protonated and deprotonated lysine species with a midpoint of 10.7 and n value of 0.7. **B-D** The spectra at pH 7, 10, and 13 fit to the sum of the protonated and deprotonated signals.

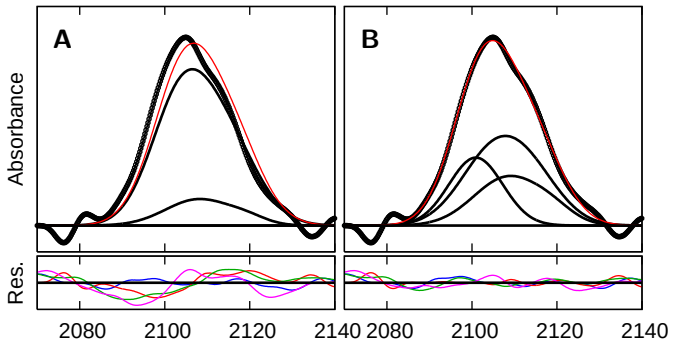

Figure S 5: Residuals at several pH values and a representative spectrum of (d_8)Lys79 cytochrome c (pH 10.4) resulting from fits using the superposition of (A) the folded and high pH/solvent-exposed signals and (B) the folded, high pH/solvent-exposed, and intermediate signals. Residuals and spectra are shown for pH 7.0 (blue), 10.4 (red), 10.9 (green), and 11.5 (magenta). F-tests yielded an F value of 8.2 with a critical F value of 1.2 at the 95% confidence level for the (d_8)Lys79 fit at pH 10.4.

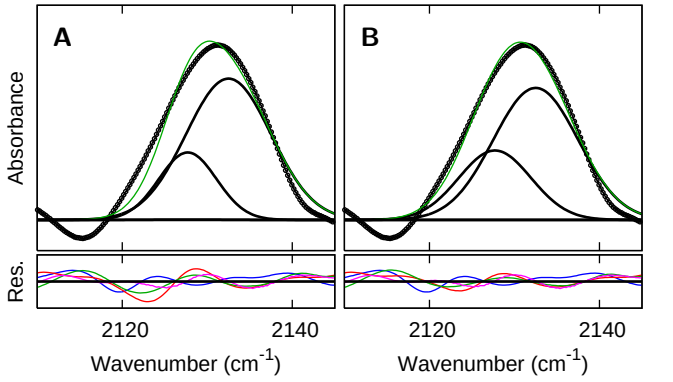

Figure S 6: Residuals at several pH values and a representative spectrum of (d_3)Met80 cytochrome c (pH 10.9) resulting from fits using the superposition of (A) the folded and high pH/solvent-exposed signals and (B) the folded, high pH/solvent-exposed, and intermediate signal Residuals and spectra are shown for pH 7.0 (blue), 10.4 (red), 10.9 (green), and 11.5 (magenta). F-tests yielded an F value of 1.7 with a critical F value of 1.4 at the 95% confidence level for the (d_3)Met80 fit at pH 10.9.

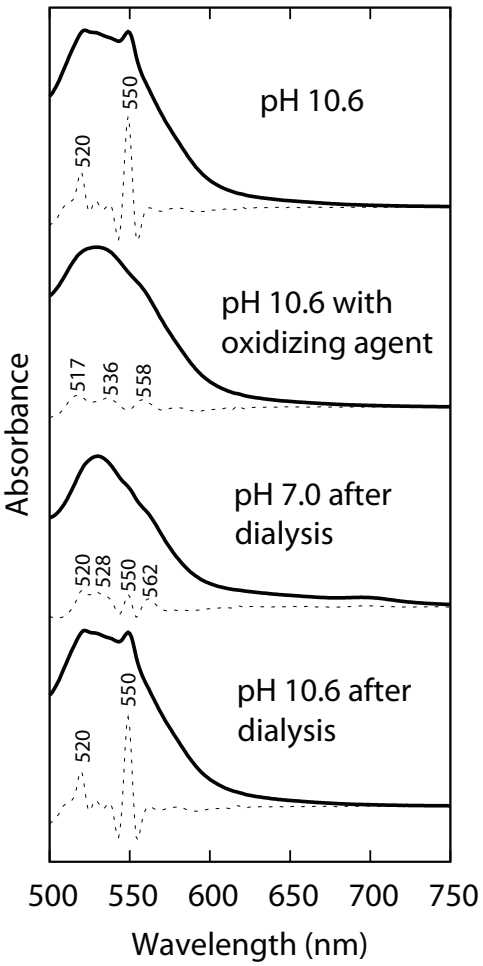

Figure S 7: We examined the UV/vis spectra prior to and after the addition of oxidizing agent (bis(dipicolinoato) Co(III)) for evidence of the signals associated with the reduced state (peaks occuring with a maxima of 520 and 550 nm). Plotted are the UV/vis spectra (solid lines) and the negative second derivatives (dotted lines) of cytochrome *c* equilibrated at pH 10.6, equilibrated at pH 10.6 with 2 fold oxidizing agent, equilibrated and refolded at pH 7 after dialysis, and equilibrated at pH 10.6 after dialysis. The plots clearly show that oxidizing the alkaline state eliminates the 520 and 550 nm peaks. These peaks return in full at pH 10.6 but do not return at pH 7.

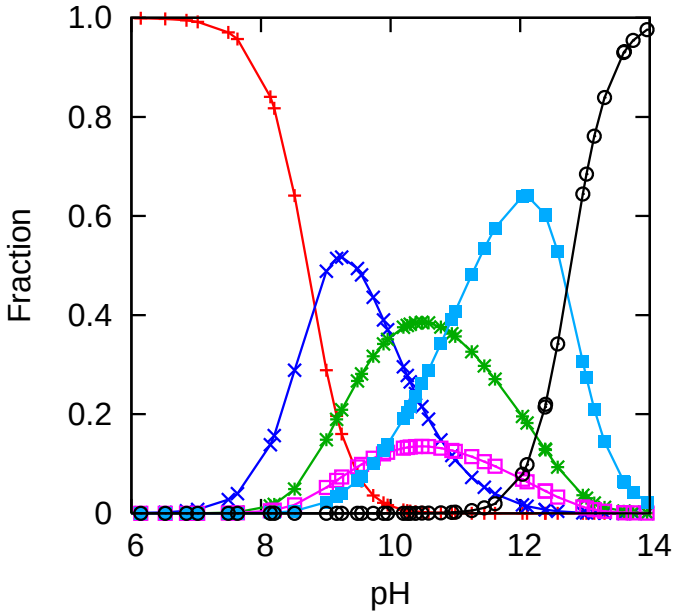

Figure S 8: Fractional concentrations of the intermediates calculated using a sequential model and parameters in Table S1. Compared to Figure 8 in the main text, **III** (+ native), **3.5**, (× neither methionine nor lysine ligated), $\mathbf{IV_a}$ (∗ Lys73 ligated), $\mathbf{IV_b}$ (□ Lys79 ligated), **V** (■ hydroxide ligated), **U** (○ unfolded).

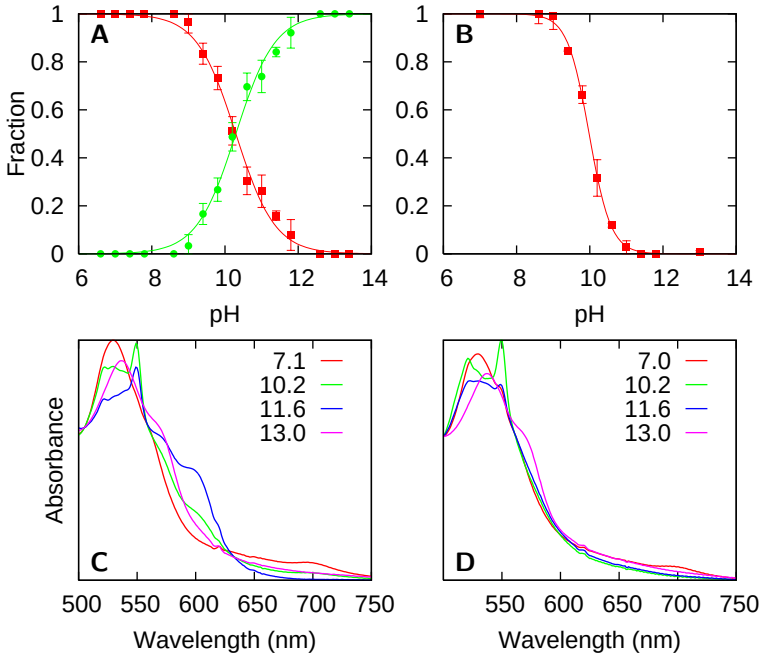

Figure S 9: Fractional concentrations of (A) the folded and unfolded signals observed by FT IR and (B) the 695 nm absorbance obtained from experiments with the Me_3-Lys72/73/79-(d_3)Met80 protein. UV/vis absorption spectra of the Q-band region at several pH values for (C) the Me_3-Lys72/73/79-(d_3)Met80 protein and (D) the wild-type protein.

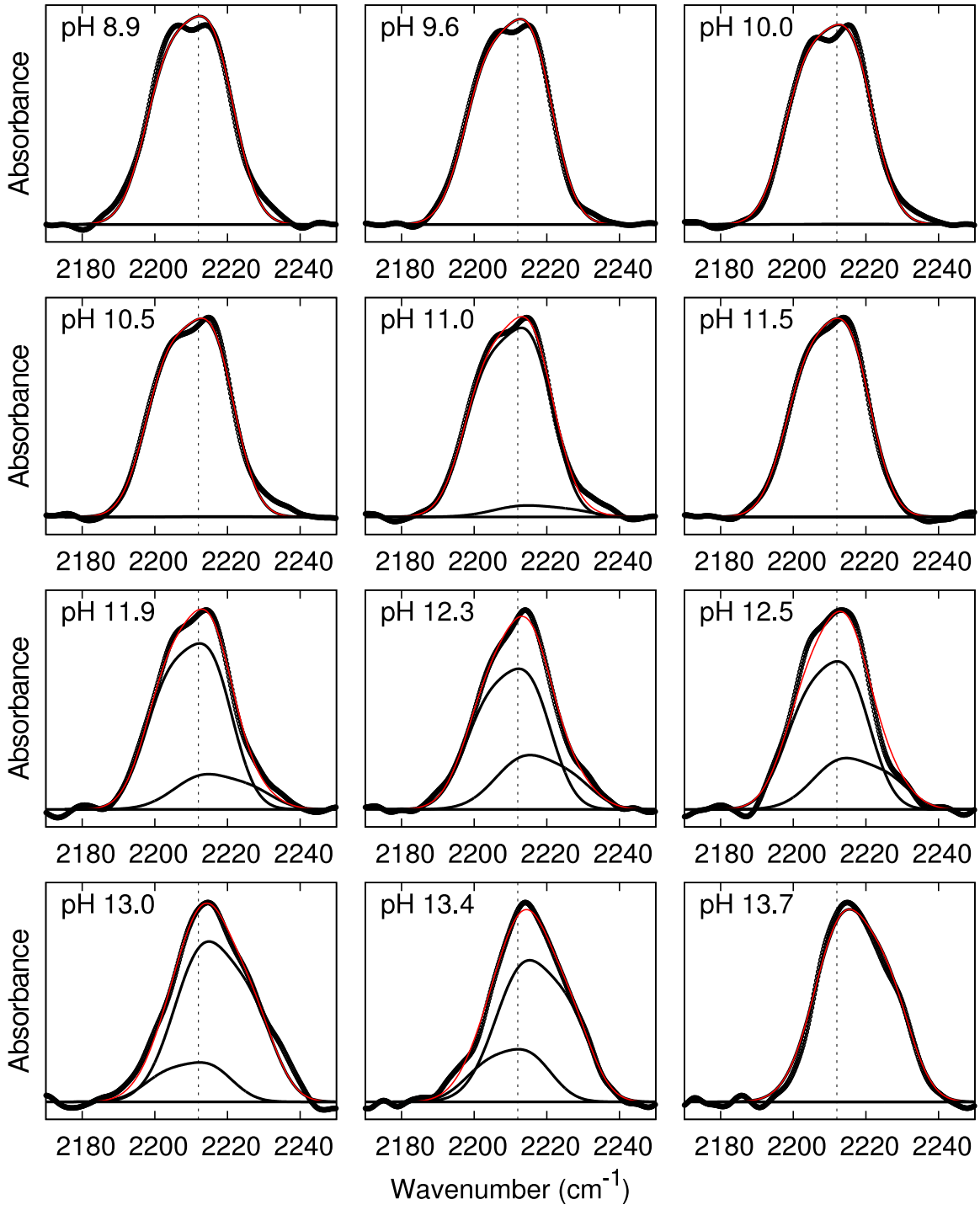

Figure S 10: Representative spectra showing the observed transition of Leu68 showing raw data (thick black lines), folded, intermediary, and highly alkaline signals (thin black lines), and fits (red lines).

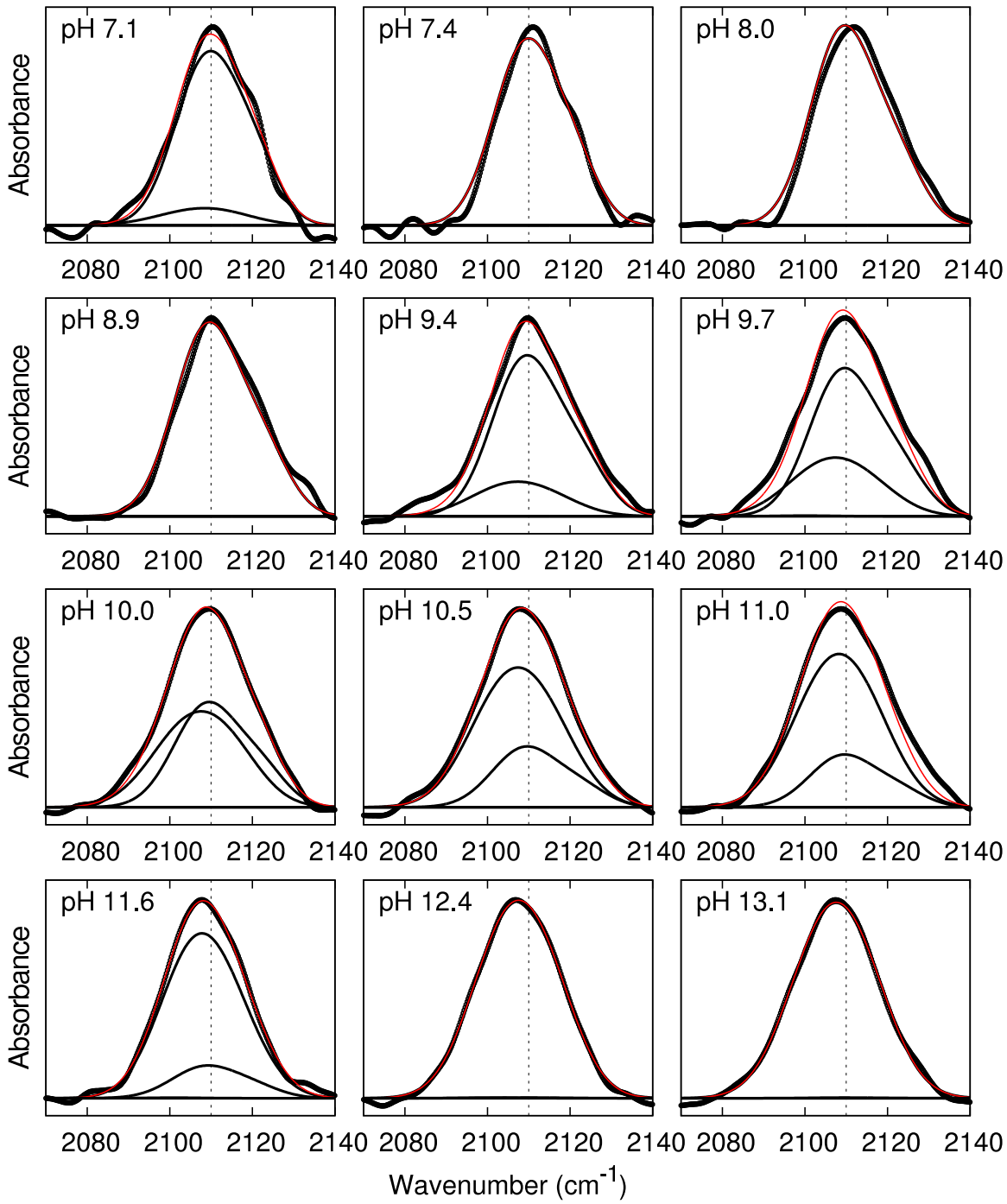

Figure S 11: Representative spectra showing the observed transition of Lys72 showing raw data (thick black lines), folded, intermediary, and highly alkaline signals (thin black lines), and fits (red lines).

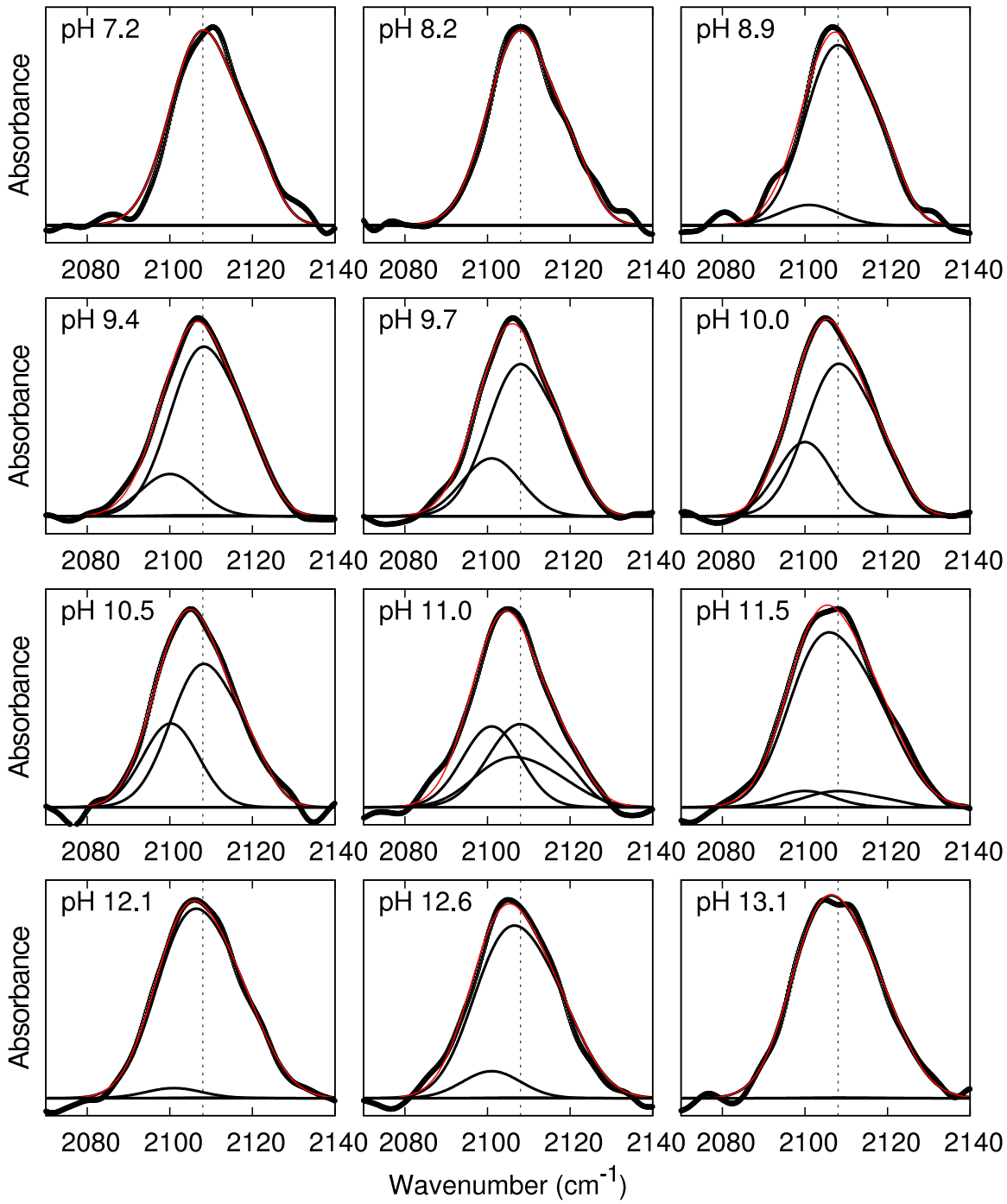

Figure S 12: Representative spectra showing the observed transition of Lys73 showing raw data (thick black lines), folded, intermediary, and highly alkaline signals (thin black lines), and fits (red lines).

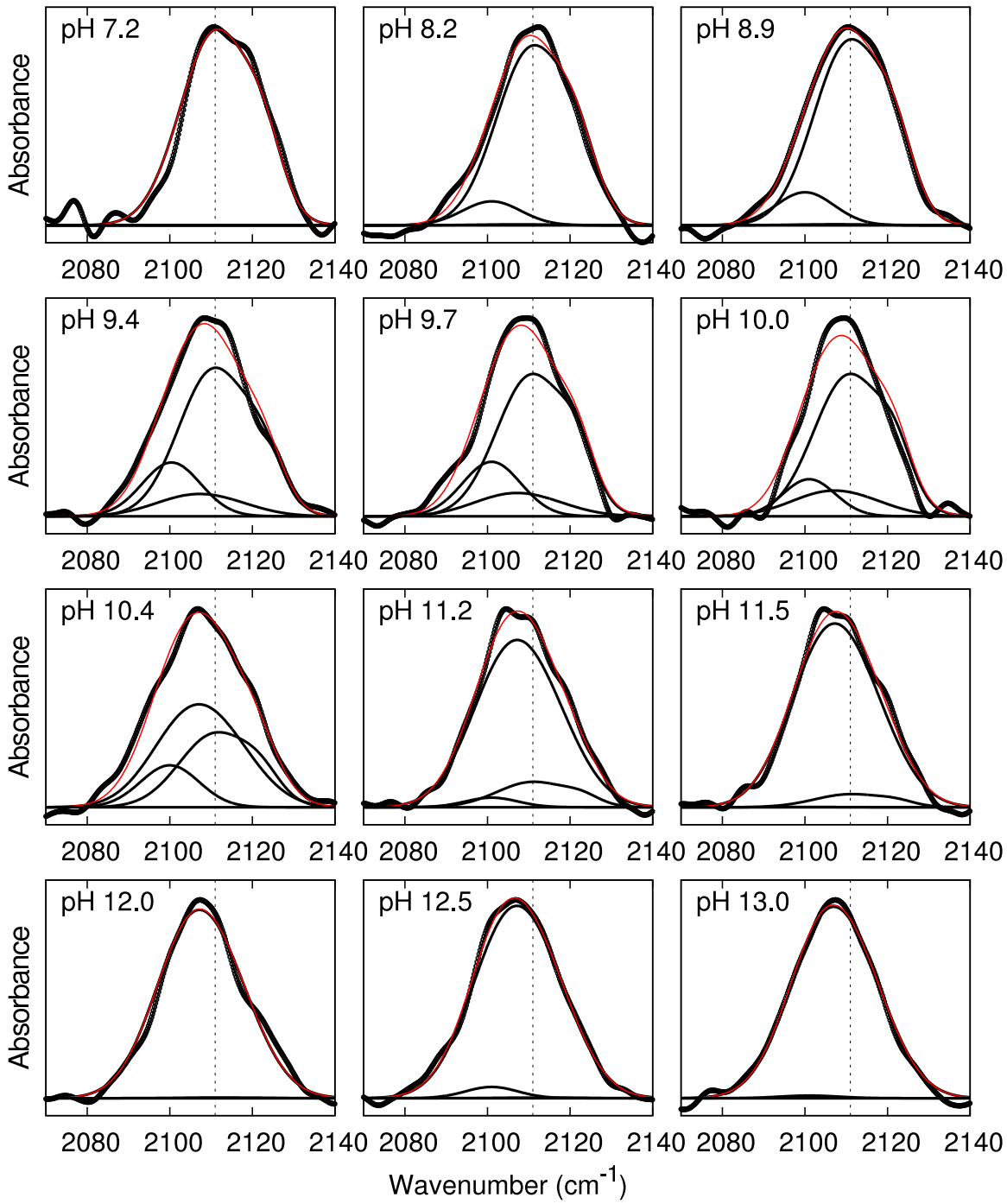

Figure S 13: Representative spectra showing the observed transition of Lys79 showing raw data (thick black lines), folded, intermediary, and highly alkaline signals (thin black lines), and fits (red lines).

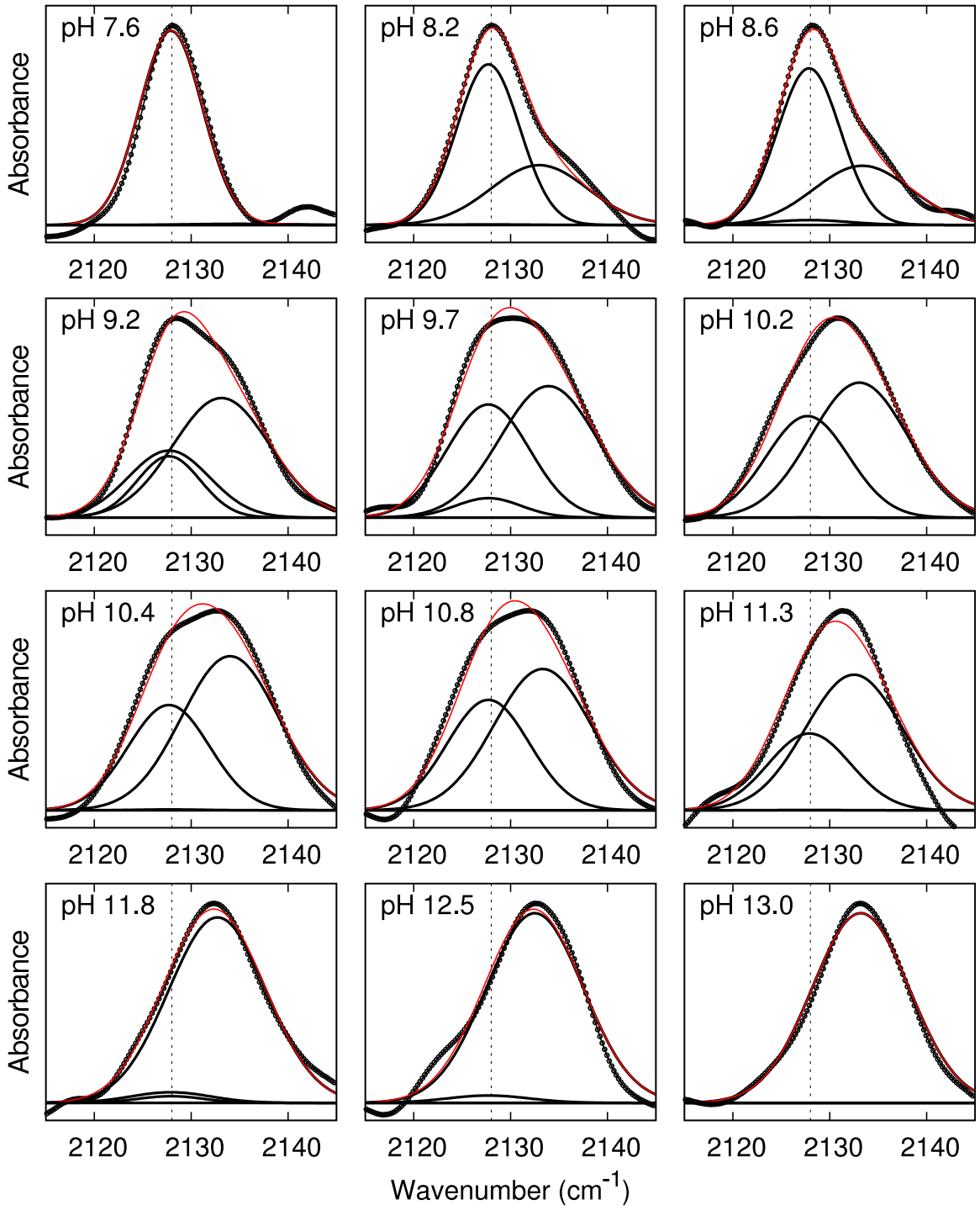

Figure S 14: Representative spectra showing the observed transition of Met80 showing raw data (thick black lines), folded, intermediary, and highly alkaline signals (thin black lines), and fits (red lines).