Supplemental Information for

Increased Protein Structural Resolution from Diethylpyrocarbonate-based Covalent Labeling and Mass Spectrometric Detection

Yuping Zhou and Richard W. Vachet*

Department of Chemistry, University of Massachusetts, Amherst

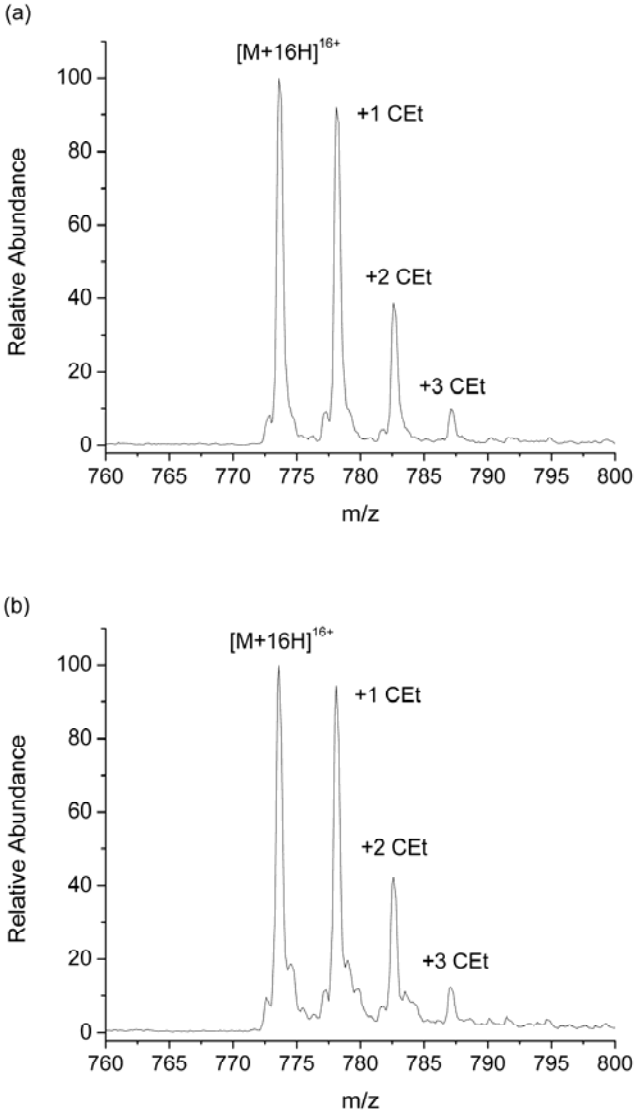

Figure S1: Expanded view of mass spectra showing the extent of DEPC modification for the +16 charge state of cytochrome c (a) 10 min after quenching the reaction and (b) 24 hrs after quenching the reaction. CEt refers to the carbethoxy group that is added to the protein upon reaction with DEPC. The modification percentage of the intact protein is determined from the sum of the ion abundances for the modified protein divided by the total ion abundances for all the peaks (modified and

unmodified) for all the charge states. In (a) the modification percentage is 56 ± 3% based on three separate experiments. In (b) the modification percentage is 57 ± 2% based on three separate experiments.

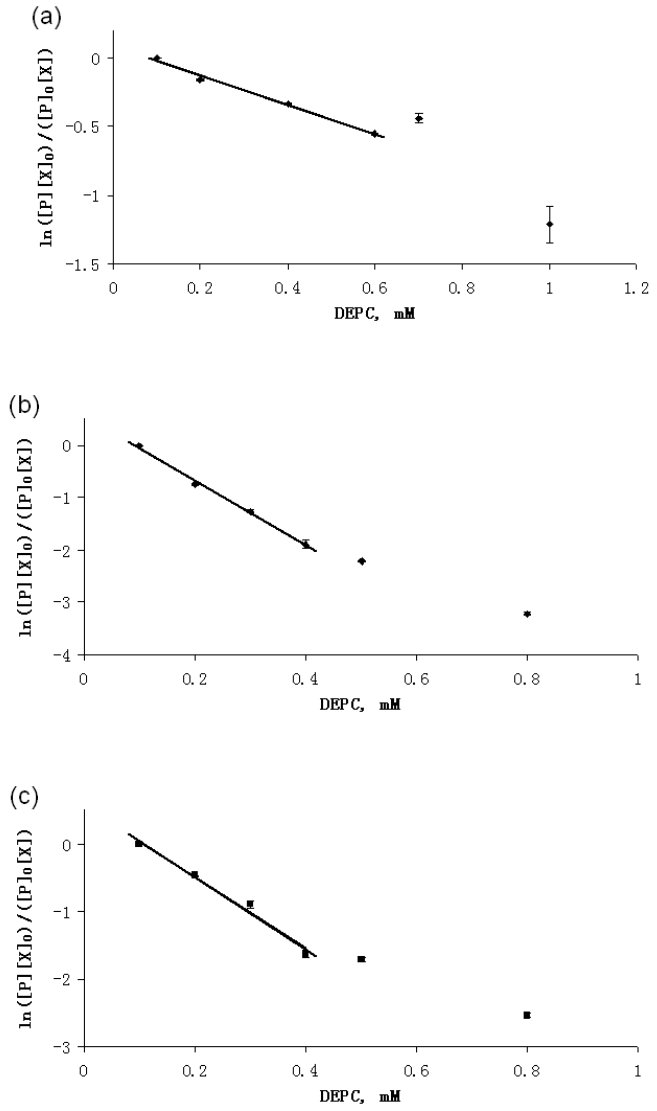

Figure S2: Kinetics plots for the reaction of DEPC with (a) cytochrome c, (b) myoglobin, and (c) β-2-microglobulin. The plot for each protein was generated from ESI-MS data of the DEPC-reacted proteins. In previous work [1], we demonstrated that deviations from linearity indicate reagent concentrations at which protein structural changes begin to occur. Based on these plots, proteins are reacted at DEPC concentrations below the linearity break point to ensure the structural integrity of the protein. On these second-order kinetics plots, the x axis is DEPC concentration, and the y axis corresponds to $\ln([P][X]_0)/([P]_0[X])$. The $[P]/[P]_0$ ratio is calculated by dividing the peak area of the unmodified protein by the sum of the peak areas of the

modified and unmodified protein. The concentration of DEPC, [X] is obtained by the difference between the [P] and $[P]_0$ values.

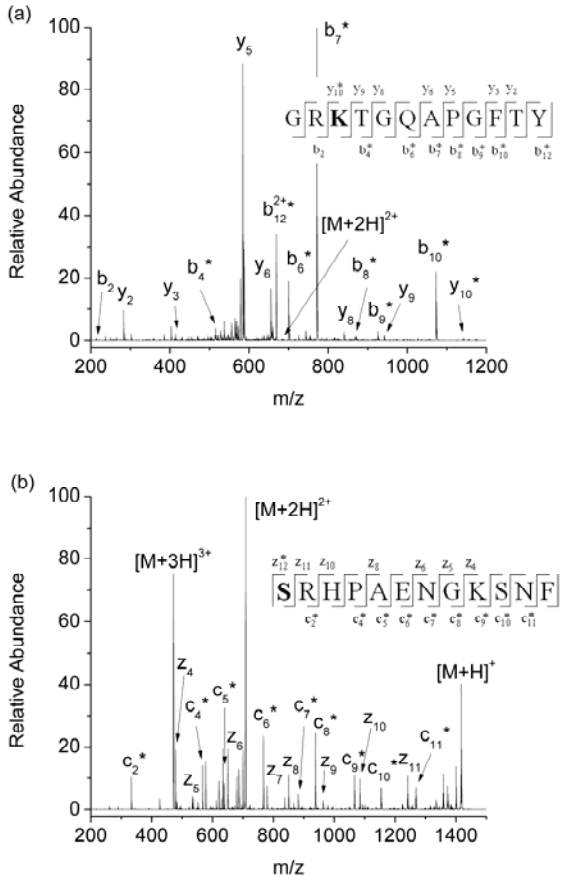

Figure S3: (a) Tandem mass spectrum acquired after CID of the $[M+2H]^{2+}$ ion of the Gly37-Tyr48 proteolytic fragment from cytochrome c. A series of unmodified y ions from y_2 to y_9, a series of modified b ions from b_4 to b_{12}, a modified y_{10} ion, and an unmodified b_2 ion confirm that Lys39 is the site of modification. (b) Tandem mass spectrum acquired after ETD of the $[M+3H]^{3+}$ ion of the Ser11-Phe22 fragment of β2m. A series of modified c ions and a series of unmodified z ions confirm that Ser11 is the site of modification. The product ions with an asterisk are the product ions that contain the DEPC modification.

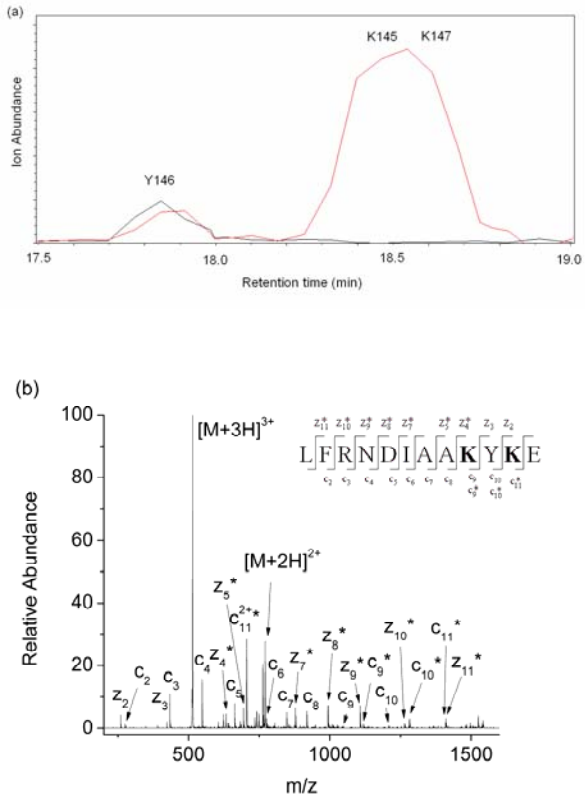

Figure S4**:** (a) Extracted ion chromatograms (EIC) of m/z 514 (red) and 385.7 (black), which are the +3 and +4 charge state of the DEPC-modified peptide ^{137}LFRNDIAAKYKE^{148} from myoglobin. (b) An example ETD tandem mass spectrum of the second chromatographic peak of the modified Leu137-Glu148 fragment, illustrating how ETD can identify two modified sites. A series of unmodified c_2 to c_8 ions, a mix of modified and unmodified c_9 and c_{10} ions, and a modified c_{11} ion indicate that both K145 and K147 are modified. Note: modified z_2 and z_3 ions are not observed because when K147 is modified these products are no longer charged.

Table S1. Solvent accessible surface area (SASA) ratios for DEPC-modifiable amino acids in cytochrome c, myoglobin and β-2-microglobulin. The modification status under two different digestion conditions is indicated for cytochrome c and myoglobin.

Cytochrome c[1]			
residue	% SASA ratio[2]	2 h	Glu-C
His18	15.2	--[3]	--
His26	34.2	--	--
His33	48.7	--	--
Tyr48	25.9	--	--
Tyr76	10.3	--	--
Tyr74	27.3	--	--
Tyr97	8.0	--	--
Thr19	3.0	--	--

Thr28	41.5	--	--
Thr40	0.0	--	--
Thr47	100.0	--	--
Thr49	40.9	--	--
Thr58	48.1	--	--
Thr63	12.0	--	--
Thr78	8.9	--	--
Thr89	64.9	--	--
Thr102	1.1	--	--
N-terminus	53.9	--	--
Lys5	78.8	--	Labeled
Lys7	69.8	--	Labeled
Lys8	75.4	--	Labeled
Lys13	40.8	--	--
Lys22	84.0	--	Labeled
Lys25	90.7	Labeled	Labeled
Lys27	63.4	--	Labeled
Lys39	72.9	Labeled	Labeled
Lys53	63.0	Labeled	Labeled
Lys55	36.5	Labeled	Labeled
Lys60	50.0	Labeled	Labeled
Lys72	75.3	Labeled	Labeled
Lys73	62.6	Labeled	Labeled
Lys79	68.3	--	Labeled
Lys86	41.6	Labeled	Labeled
Lys87	86.8	Labeled	Labeled
Lys88	94.0	Labeled	Labeled
Lys99	44.2	Labeled	Labeled
Lys100	50.4	Labeled	Labeled
Cys17[4]	44.6	--	--
% Labeled[5]		45%	67%
Myoglobin			
residue	% SASA ratio	2 h	Glu-C
His24	2.7	Labeled	--
His36	32.6	Labeled	Labeled
His48	64.7	--	Labeled
His64	26.6	--	Labeled
His81	89.1	Labeled	Labeled
His82	4.9	--	--
His93	35.3	--	--
His97	43.3	--	Labeled
His113	43.2	--	Labeled
His116	43.9	Labeled	Labeled

His119	20.0	--	--
Tyr103	15.2	--	Labeled
Tyr146	4.2	--	Labeled
Ser3	63.8	--	--
Ser58	9.1	--	--
Ser92	0.0	--	Labeled
Ser108	5.4	--	--
Ser117	51.3	Labeled	Labeled
Thr34	33.9	--	--
Thr39	1.9	--	--
Thr51	61.8	--	--
Thr66	40.2	--	--
Thr70	71.8	--	--
Thr95	76.2	--	Labeled
Thr132	45.8	--	--
N-terminus	97.8	Labeled	Labeled
Lys16	22.1	Labeled	--
Lys42	33.0	Labeled	Labeled
Lys45	54.9	Labeled	Labeled
Lys47	0.0	--	--
Lys50	83.4	--	Labeled
Lys56	58.3	--	--
Lys62	48.0	--	Labeled
Lys63	75.3	--	Labeled
Lys77	50.5	Labeled	Labeled
Lys78	46.2	Labeled	Labeled
Lys79	53.6	Labeled	Labeled
Lys87	52.0	--	Labeled
Lys96	84.2	--	Labeled
Lys98	65.0	--	Labeled
Lys102	63.6	--	--
Lys118	23.5	--	Labeled
Lys133	44.8	--	--
Lys145	16.4	--	Labeled
Lys147	97.0	--	Labeled
% Labeled[5]		39%	81%
β-2-microglobulin			
residue	% SASA ratio	2 h	
His13	50.8	Labeled	
His31	32.7	Labeled	
His51	54.7	Labeled	
His84	0.0	--	
Tyr10	41.9	Labeled	

Tyr26	25.8	--	
Tyr63	27.1	--	
Tyr66	7.0	--	
Tyr67/Thr68	22.5/2.5	Labeled	
Tyr78	6.0	Labeled	
Ser11	11.4	Labeled	
Ser20	65.0	Labeled	
Ser28	19.7	Labeled	
Ser33	66.4	Labeled	
Ser52	43.8	--	
Ser55	55.9	Labeled	
Ser57/Lys58	35.2/96.3	Labeled	
Ser61	27.1	--	
Ser88	90.1	--	
Thr4	83.8	--	
Thr71	36.0	--	
Thr73	61.4	--	
Thr86	17.8	--	
N-terminus	85.0	Labeled	
Lys6	80.1	Labeled	
Lys19	69.0	Labeled	
Lys41	14.1	Labeled	
Lys48	81.8	Labeled	
Lys75	92.3	Labeled	
Lys91	61.5	Labeled	
Lys94	67.8	Labeled	
Cys25	0.0	--	
Cys80	0.0	--	
% Labeled[5]		95%	

1. The PDB IDs for cytochrome c, myoglobin and β-2-microglobulin that were used to determine SASA were 1AKK, 1DWR and 1JNJ, respectively. For β-2-microglobulin, 1JNJ consists of 20 NMR structures, and so the reported SASA values are the average from these 20 structures.
2. SASA was calculated using GETAREA 1.1 [2]. 1.4 Å was used as the probe radius, and the calculated SASA percentage is the ratio of the SASA of the side chain in the protein to the SASA of the side chain (X) in the unstructured Gly-X-Gly tripeptide. Residues with %SASA values that exceed 50% are typically considered to be solvent exposed, and residues with ratios less than 20% are typically considered to be buried. We chose 30% as the cutoff for determining if a residue is solvent exposed.
3. -- indicates that modified peptide is not detected.
4. GETAREA ignores the heme in cytochrome c, so it calculates Cys17 to be solvent exposed when it is really not because it is covalently bound to the heme.

5. % labeled corresponds to the percentage of the surface exposed modifiable (i.e. His, Lys, Cys, Ser, Thr, and Tyr) residues that are found to be labeled.

Reference

1. Mendoza, V. L., Vachet, R. W.: Protein Surface Mapping Using Diethylpyrocarbonate with Mass Spectrometric Detection. *Anal. Chem.* **80,** 2895-2904 (2008)

2. Fraczkiewicz, R., Braun, W.: Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19,** 319-333 (1998)