Supporting Information

Fast Protein Footprinting by X-ray Mediated Radical Trifluoromethylation

Ming Cheng,[a] Awuri Asuru,[b] Janna Kiselar,[b] George Mathai,[c] Mark R. Chance,*[,b] and Michael L. Gross*[,a]
[a] Department of Chemistry, Washington University, St. Louis, Missouri 63130, United States
[b] Center for Proteomics & Bioinformatics, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106, United States
[c] Department of Chemistry, Sacred Heart College, Thevara, Kochi, Kerala 682013, India

Corresponding Authors: * MLG: Tel: (314)935-4814. E-mail: mgross@wustl.edu;
*MRC: Phone: (216) 368-4406. E-mail: mark.chance@case.edu

Table of Contents

1. Additional Experiment Procedures

1.1. Materials.

Equine skeletal myoglobin, apo-myoglobin, sodium triflinate, phosphate buffered saline tablets (PBS, pH = 7.4), urea, formic acids (FA), and trifluoracetic acid (TFA), acetonitrile and water were obtained from Sigma-Aldrich Chemical Company (St. Louis, MO, USA). Except for sodium triflinate (≥ 95% purity), all other chemicals were LC-MS grade. Trypsin was purchased from Thermo Fisher Scientific. The Zorbax Eclipse XDB-C8 trap column (2.1 x 20 mm) was obtained from Agilent. The silica capillary C18 column was custom-built with C18 reversed-phase material from Waters Symmetry, 5 μm, 100 Å, 75 μm ×30 cm).

1.2. Intact Protein Analysis

For each run, ~40 pmol of protein sample was passed over a custom-built platform where the protein was captured by a C8 trap column and desalted at 200 μL/min of H_2O containing 0.1% trifluoracetic acid for 3 min. After on-line desalting, the protein was eluted with a 7 min gradient of 5% to 80% acetonitrile in 0.1% formic acid at a flow rate of 200 μL/min.

1.3. LC/MS/MS Analysis of Proteolytic Peptides

After the digestion, an aliquot of 5 μL was loaded onto a custom-built silica capillary column packed with C18 reversed-phase material. The HPLC gradient was: from 2.5% solvent B (80% acetonitrile, 0.1% formic acid) to 18.5% solvent B over 30 min, then increased to 50% solvent B over 30 min and subsequently jumped to 98% solvent B over 2 min and wash the column for 6 min, followed by a 12 min re-equilibration step. A Q Exactive Plus hybrid quadrupole orbitrap mass spectrometer coupled with a Nanospray Flex ion source (Thermo Fisher, Santa Clara, CA) was utilized for MS analyses. The instrumental parameters were spray voltage of 2.5 kV; capillary T = 250 °C. The Q Exactive Plus was operated in the data-dependent acquisition mode. Mass spectra were acquired in the orbitrap (*m/z* 400−1600) with a mass resolving power of 70 000 at *m/z* 400 for MS1 and 17 500 for MS/MS. The twenty most abundant ions were selected for higher energy collisional dissociation at an automatic gain control (AGC) target of 500,000. Ions previously selected for MS/MS were dynamically excluded for 3 s.

1.4. Data Analysis

LC-MS/MS raw files were imported into the Byonic™ Software (Protein Metrics, San Carlos, CA, USA) coupled with an in-house database. Tolerances were chosen of 20 ppm for precursor mass, 60 ppm fragment mass tolerance and for CID/HCD fragmentation. In the modification table of Byonic, oxidation were specified (+15.9949 Da) @ all amino acids, di-oxidations (+31.9898) @ residues C,F,K,M,P,R,W,Y, and carbamyl (+43.0058) @ residues C,K,M,R,S,T,Y as rare modifications. Trifluoromethylation (+67.9874) was specified @ all amino acids. A typical search allowed a total of at most three common modifications and a total of at most two rare modification per peptide. Modification fractions for certain peptides were interrogated with Byologic™ software (Protein Metrics, San Carlos, CA, USA) and double-checked with a custom program Thermo Xcalibur. Modification sites on the peptide were assigned based on product-ion spectra (MS/MS data). Modification ratios were calculated using the following equation:

$$\%\ \text{modified} = \frac{\sum I_{\text{CF3}}}{\sum I_{\text{CF3}} + \sum I}$$

Modified residues were identified via manual validation of product-ion (MS/MS) spectra, and the abundance of unmodified and modified species was determined by manual peak integration of extracted ion chromatograms (EIC) from Xcalibur software. The fraction of remaining unmodified was calculated at each exposure time. Values were plotted against exposure time to produce

dose-response curves that were fit to a single-exponential function to determine reactivity rates for each modified AA. Protection ratios were calculated by dividing the modification rates of holo-myoglobin by the modification rates for apo-myoglobin and used to identify regions with potentially decreased or increased solvent accessibility.

1.5. CD Analysis

Myoglobin samples (Apo and Holo states, 5 uM) were incubated with or without 20 mM $NaSO_2CF_3$ in 1x PBS buffer. Circular dichroism (CD) spectra were measured at room temperature over the wavelength range of 195-340 at 0.5 nm intervals by using a JASCOJ815CD spectrometer (JASCO Analytical Instruments, Tokyo, Japan).

2. Calculation of Trifluoromethylation Modification Ratios for Intact Protein

Radical trifluoromethylation on protein and peptides were exploited recently by several groups.[1-2] In those studies, pathways for modifying various amino acids were proposed, and timescales of labeling were mins to hours. Our results show that extensive $\bullet CF_3$ labeling of proteins can be realized on millisecond timescales on the synchrotron platform (Fig.S1 and Fig.S2). In addition, we see a higher level of trifluoromethylation including mono, di, tri, tetra- CF_3 modification at apo-myoglobin whereas most abundant modified holo-myoglobin contains one CF_3 modification.

Modification ratios were calculated from Extracted-ion chromatograms (EIC) by using the integrated areas of peaks representing CF_3-modification and summed over zero, mono, di, tri, tertra-CF_3 species. The CF_3-modified protein is the dominate modified species according to our analysis. $\bullet CF_3$ modified apo-myoglobin and holo-myoglobin are 71% and 67%, respectively.

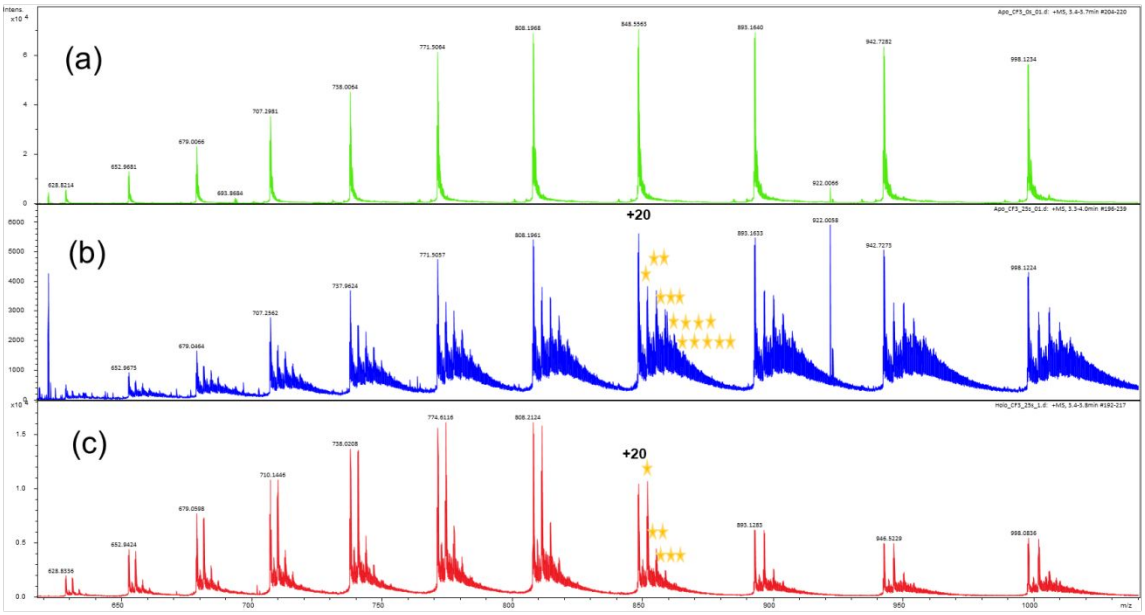

Figure S1. ESI mass spectra of intact protein. (a) ESI mass spectra of intact (not digested) apo myoglobin (aMb) with 0 ms X-ray exposure. (b) Full ESI mass spectrum of CF_3-modified aMb with 25 ms X-ray exposure. (c) CF_3-modified holo-myolgobin (hMb) with 25 ms X-ray exposure. The number of CF-modifications are indicated by stars at +20 charge state (red and blue).

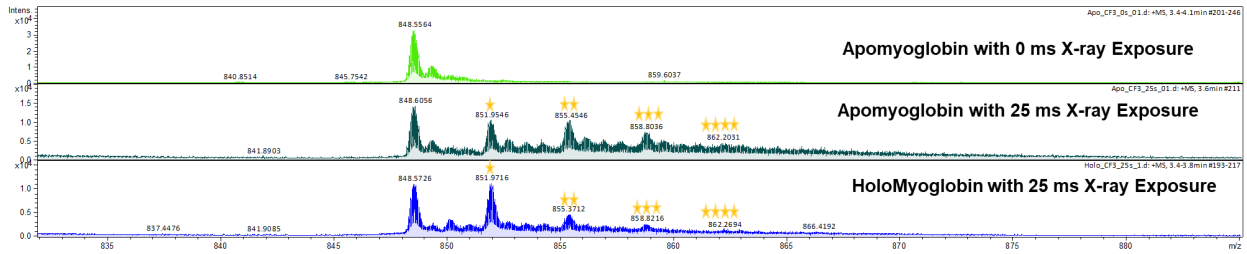

Figure S2. Zoom-in mass spectra of aMb/hMb of +20 charge state. Times of X-ray exposure given in Figure. The number of CF_3-modifications are indicated by stars. Holo state is less susceptible to labeling based on the extent of modification.

3. Measurement of Trifluoromethylation Ratio at Residue Level

The $\bullet CF_3$ is known to react with aromatic and heteroaromatic functionalities with limited equivalents of reagents, as shown in the synthetic literatures.[3-4] 98% sequence coverage of myoglobin was obtained with trypsin digestion (Fig.S3). We calculated the percentage for modified residues in apo myoglobin with 25 ms X-ray irradiation (Table S1). From our LC-MS/MS analysis, we observe that $\bullet CF_3$ not only modifies aromatic residues but also aliphatic residues.

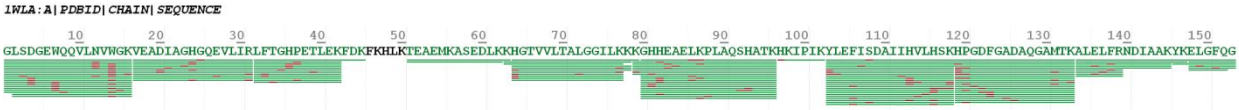

Figure S3. Sequence coverage of myoglobin (hMb): 25 ms x-ray exposure. Search utilizes (PDB 1WLA). The sequence was calculated by Byonic™ Software to afford 98% coverage. The small red sections on the sequence are where modifications (both rare and common) were detected.

Table S1. Modification extent for residues upon 25 ms X-ray irradiation

Residues	Aver % CF_3 Addition	Aver % OH addition	CF_3+ Oxidation (%)
W14	28	8	0.15
A19	0.23	0	0
H24	0.35	0.23	0
L32	0.31	0	0
H36	0.21	0.2	0
H64	0.8	0.05	0
L72, L76, V68	0.3	0.75	0
K77/K78	0.44	0.1	0
H81, H82 and H93	20	0	0
Y103	4	0	0
H119	4.3	0	0
F123	2.5	0	0
M131	0	1.5	0
H119+F123	0	0	0.2
M131+F123	0	0	0.1
F138	2.6	0.1	0
L149	0.2	0	0

3.1. Calculation of Standard Enthalpy of Radical Trifluoromethylation

•CF_3 is known to react with aromatic and heteroaromatic functionalities with limited equivalents of reagents, as shown in the synthetic literature.[3-4] From our LC-MS/MS analysis, however, we observe that •CF_3 not only modifies aromatic residues but also aliphatic residues. This special reactivity may be because:

(1) The energetics, •CF_3 can abstract aliphatic an H-atom from methane that contains the most inert aliphatic C-H bonds (Fig. S4). C-H abstraction reactions by perfluoroalkyl radicals (PFs, CF_3 radical analogies) were previously observed[5-6], and their rate constants are larger in aqueous media than in nonpolar organic solvents.[7] We performed footprinting in aqueous solution. It is likely that CF_3 radicals abstract H from aliphatic residues to produce a protein radical, followed by recombination with another CF_3 radical to generate the CF_3-modified product.

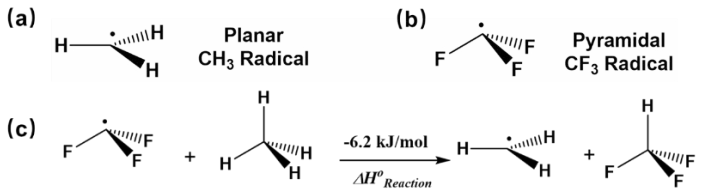

Figure S4. Standard enthalpy of reaction (ΔH^o Reaction) for methyl radical formation by abstraction by CF_3 radicals.[8]

(2) We used excess reagent (usually more than 10^3 equivalents compared to protein) to increase the reaction probability with different residues on the protein surface. Unlike in synthetic chemistry where limited equivalents of reagents are used to improve reaction selectivity, protein footprinting uses a high-dose reagent to ensure a broad-based and fast footprinting.

Third, mass spectrometry has a high dynamic range and can detect low-abundance products. The low-yield product usually cannot be determined by NMR in other studies because the dynamic range is small.

4. Test of Biocompatibility of Langlois' Reagent in Protein Solution by CD

Maintaining a protein native structure is a major concern in protein footprinting. To ensure fast labeling, footprinting usually makes use of excess reagent that serves as the radical precursor. The precursor may perturb the protein high order structure and yield misleading conclusions about protein structure. Various approaches can be employed to check the protein high-order structural integrity: examples are CD,[9-10] reaction kinetics for individual modification sites,[11] modification patterns at the protein level,[12] and activity assays.[13] CD spectroscopy is the most common check because it is sensitive to variations in protein secondary structure, with some sensitivity to tertiary structure. In the development of •CF_3 labeling, we analyzed both holo and apo myoglobin incubated with Langlois' reagent by using CD (Fig. S5 and Fig. S7). No change occurred in the CD spectrum of myoglobin with and without the Langlois reagent, and CD curves showing helical structure were observed for both samples. The result shows that no significant secondary structural changes occur when myoglobin is incubated with 20 mM Langlois' reagent.

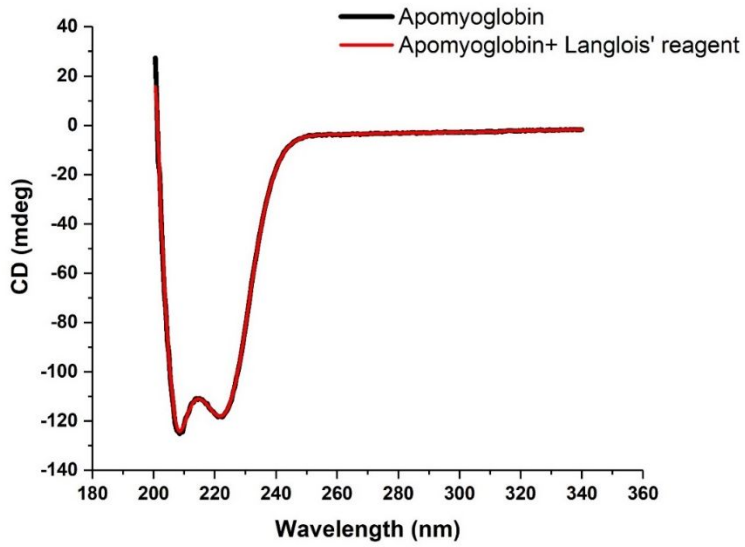

Figure S5. CD spectrum of apo-myolgobin samples. Native apo-Mb is incubated with 20 mM Langlois' reagent (red) and without reagent (black).

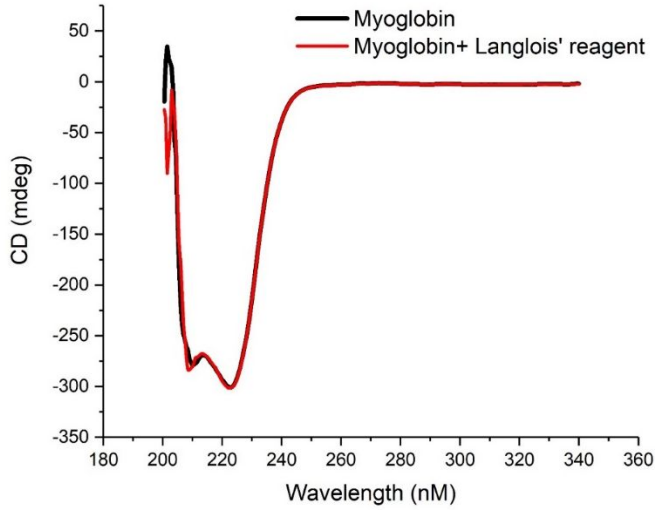

Figure S6. CD spectra of holo-myoglobin samples. Native holo-Mb is incubated with 20 mM Langlois' reagent (red) and without reagent (black).

4. Examples of LC-MS Chromatograms, Mass Spectra, and SASA Calculation

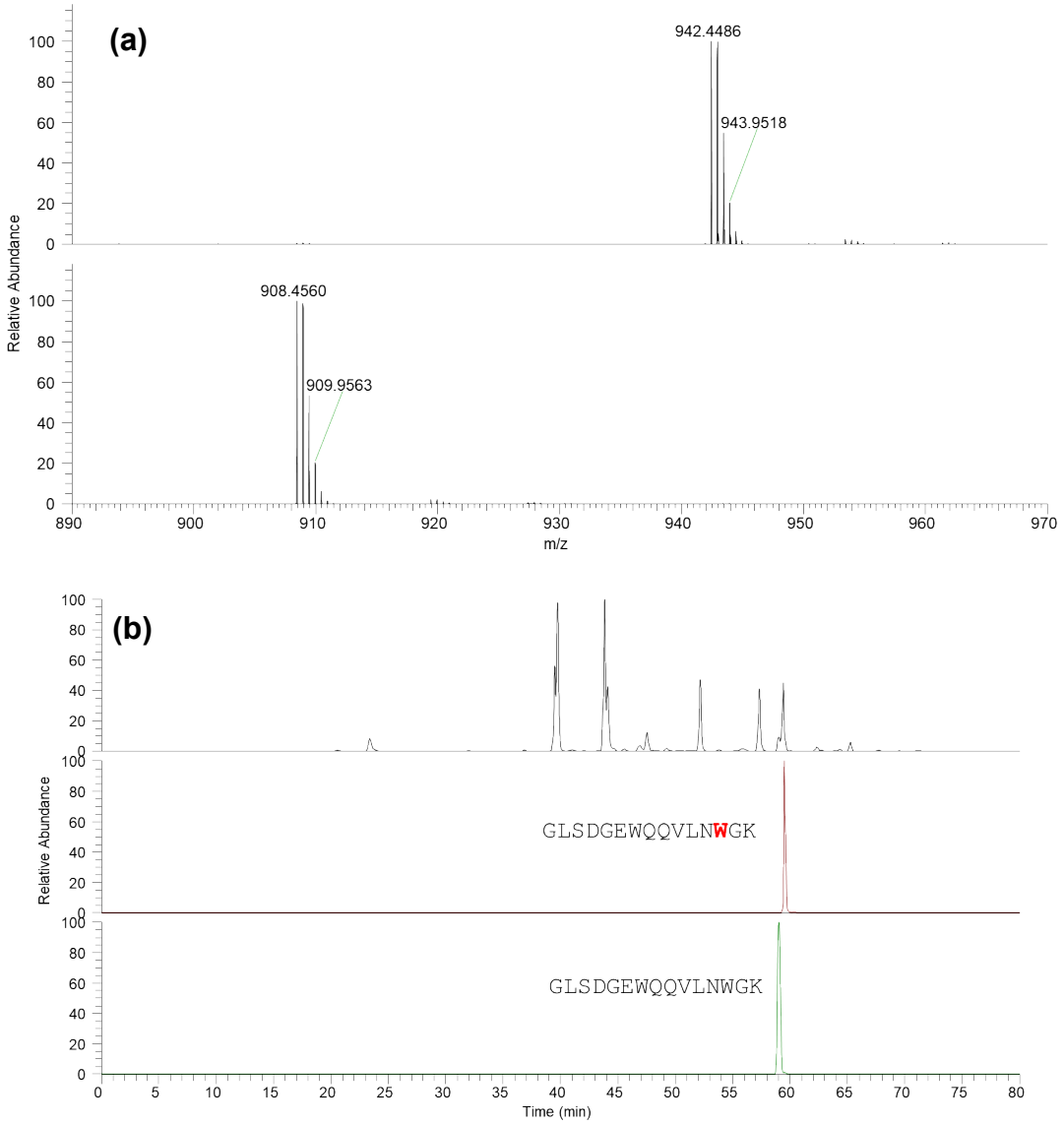

Figure S7. Representative LC-MS results for CF_3 modification at Trp. (a) Mass spectra of peptide 1-16 (+2 charge), unmodified peptide (top), and modified peptide (bottom). (b) Base peak chromatogram from LC-MS experiments with all peptide peaks labeled with retention time (top). EIC for the CF_3-modified peptide 119-133 as doubly charged precursor (middle). EIC for the unmodified peptide 80-96 as a doubly charged precursor (bottom). From product-ion spectra (MS/MS), we assigned modification to be on Trp-14 (see later).

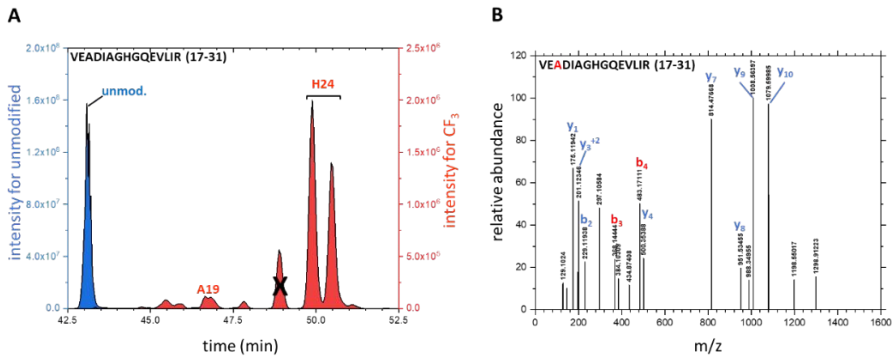

Figure S8. Representative LC-MS results of CF_3 modification at Ala. (A) EIC for peptide 17-31 showing the EIC peaks for unmodified and modified peptides with CF_3 modifications on A19 and H24. EIC peaks marked with an X do not contain the peptide of interest. (B) MS/MS showing the CF_3-modified A19 residue. Product ions b3 and b2 assign CF_3 modification on A19.

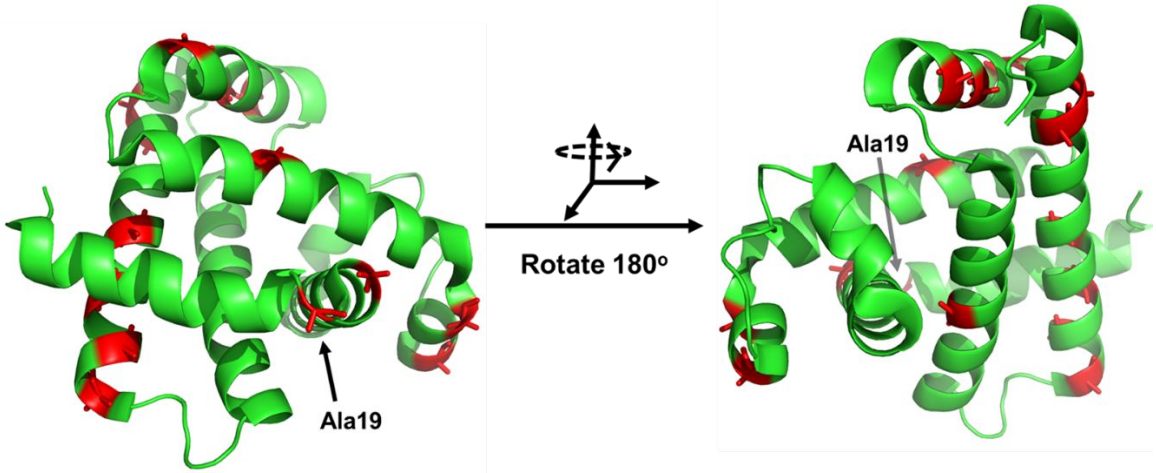

Figure S9. Mapping Ala residues on myoglobin crystal structure (PDB 1WLA). Ala19, the residue being labeled, is the only Ala located in a loop region, whereas the other Ala's are part of a structured helix.

Table S2. The calculation of SASA for Ala made by Getarea_Remote.pl program.

Residues	SASA
Ala19	55.48
Ala22	31.39
Ala53	67.48
Ala57	55.2
Ala71	31.84
Ala84	61.4
Ala90	0
Ala94	0.09
Ala110	2.23
Ala125	77.32
Ala127	7.96
Ala130	1.17
Ala134	0
Ala143	31.26
Ala144	39.72

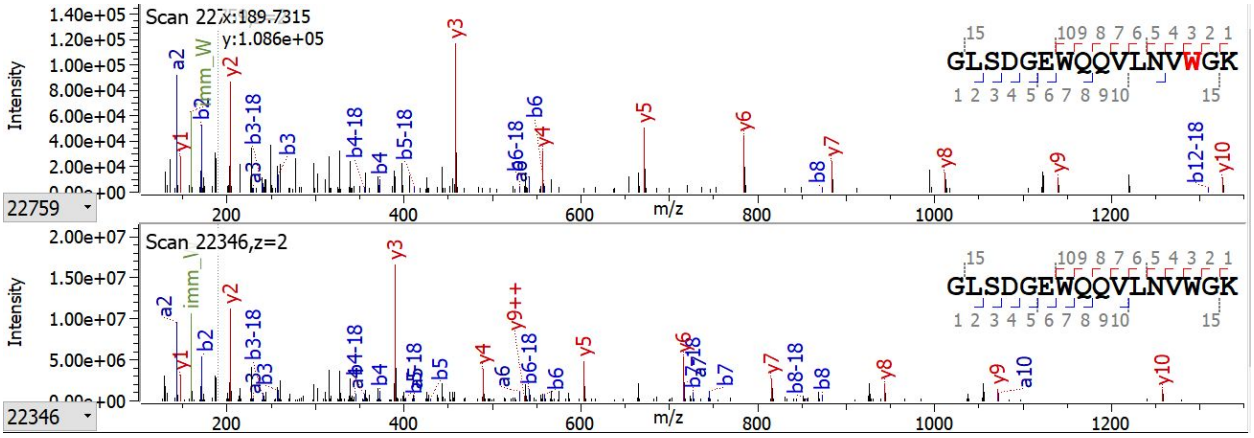

Figure S10. Product-ion spectrum of CF_3-modification on Trp-14. MS/MS spectrum of doubly charged unmodified peptide 1-16 (bottom) and of CF_3-modified peptide 1-16 (top) from apo-myoglobin (PDB3KP9). The y2 and y3 assign the modification on Trp-14.

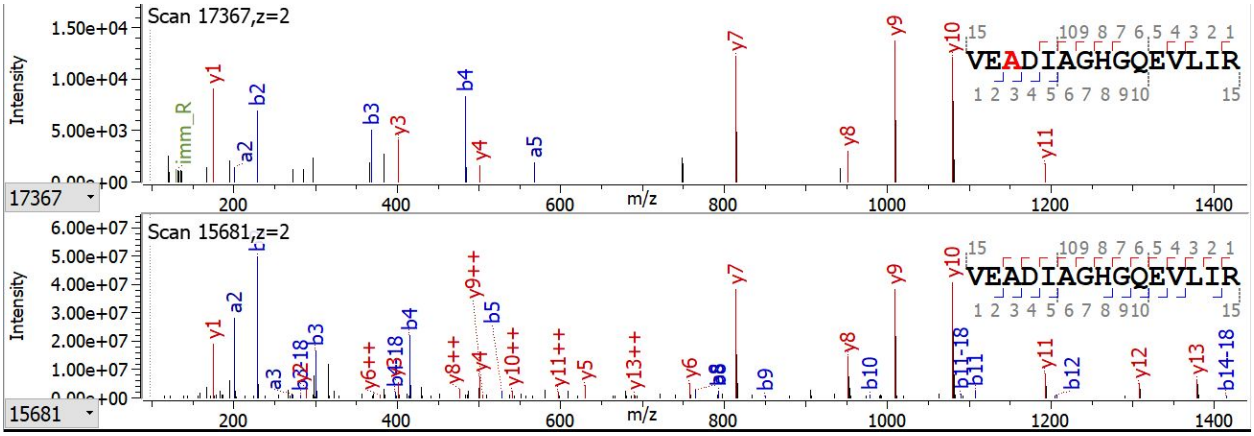

Figure S11. Product-ion spectrum of CF_3-modification on Ala-19 from MS/MS of doubly charged unmodified peptide 17-31 (bottom) and CF_3-modified peptide (top) from apo-myoglobin. The b2 and b3 assign modification as Ala-19.

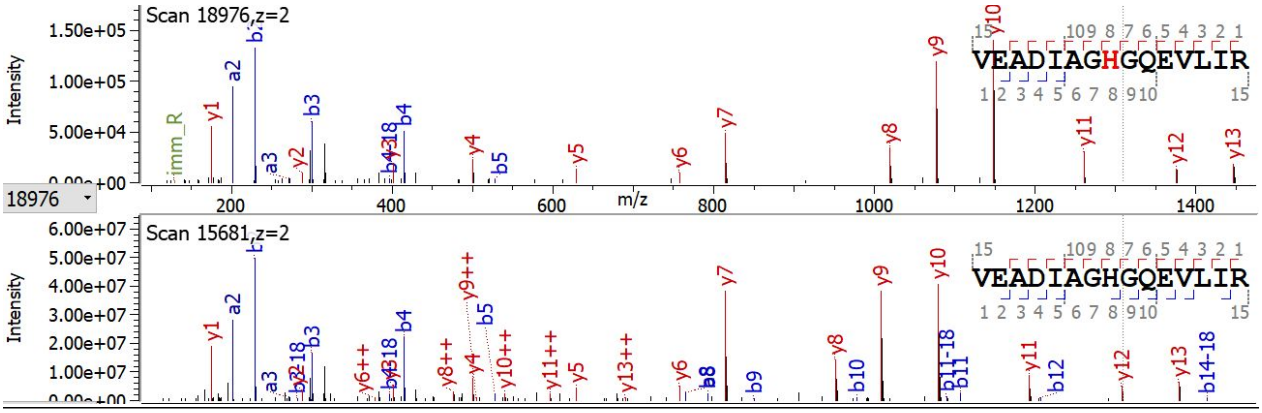

Figure S12. Product-ion spectrum of CF_3-modification on His-24 from MS/MS of doubly charged unmodified peptide 17-31 (bottom) and CF_3-modified peptide (top) from apo-myoglobin. The y8 and y7 assign the modification as His-24.

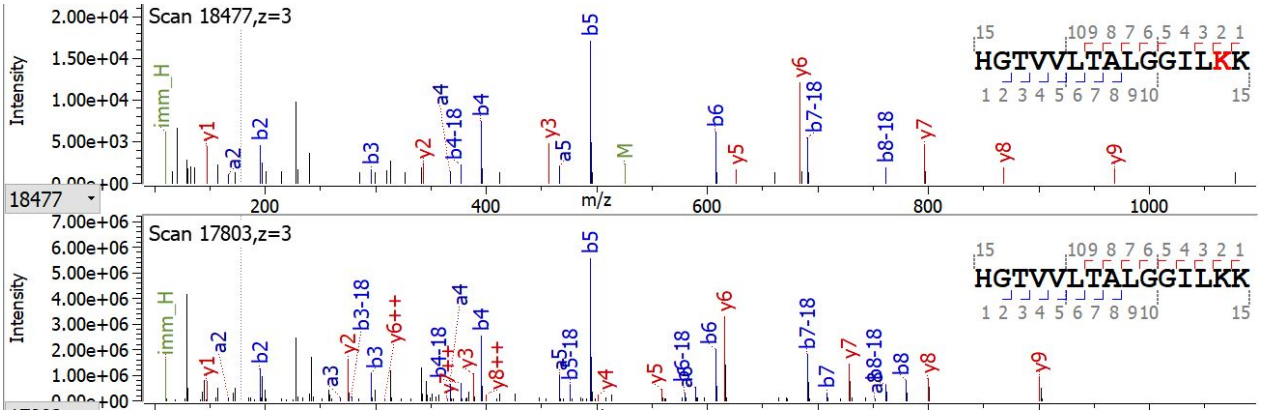

Figure S13. Product-ion spectrum of CF_3-modification on Lys-77 from MS/MS of doubly charged unmodified 64-78 (bottom) and CF_3-modified peptide (top) from apo-myoglobin. The y1 and y2 ions assigned the modification on Lys-77.

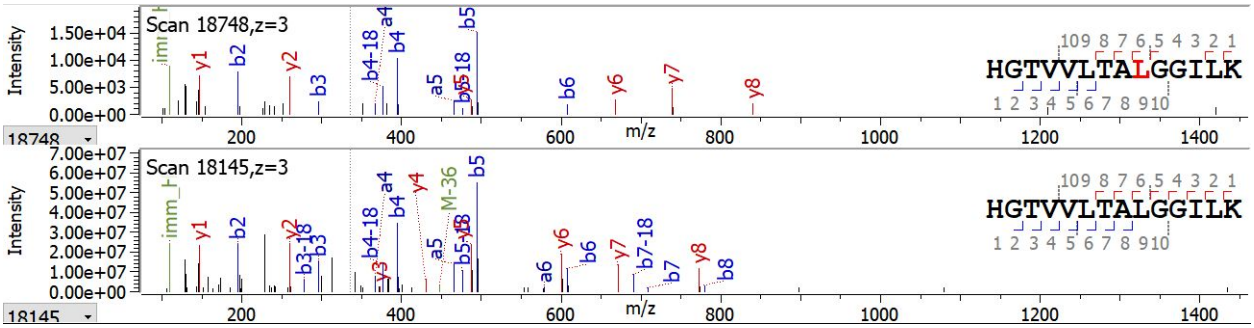

Figure S14. Product-ion spectrum of CF_3-modification on Leu-72 from MS/MS of doubly charged unmodified peptide 64-77 (bottom) and CF_3-modified peptide (top) from apo-myoglobin. Fragments y5 and y6 assigned the modification on Leu-72.

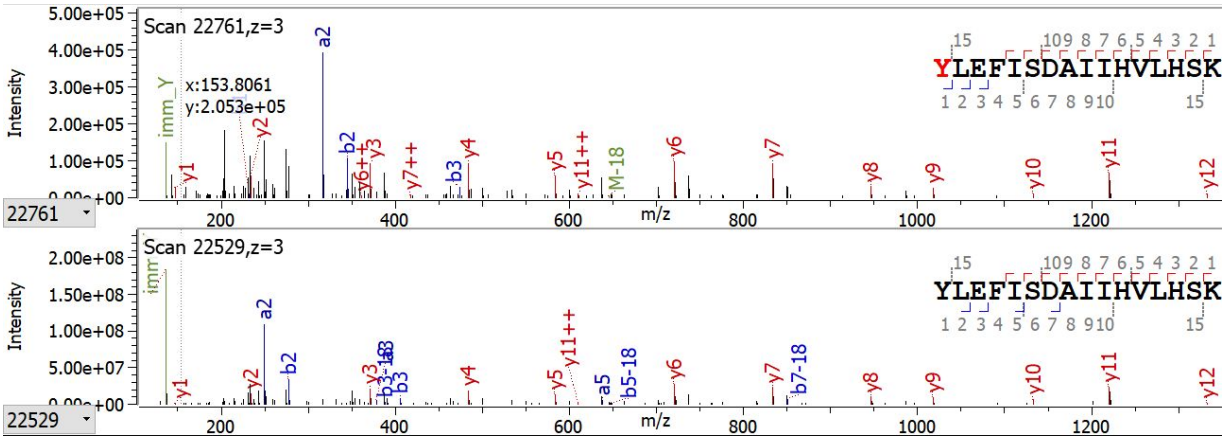

Figure S15. Product-ion spectrum of CF_3-modification on Tyr-103 from MS/MS of doubly charged unmodified peptide 103-118 (bottom) and CF_3-modified peptide (top) from apo-Myb. Fragments b1 and b2 assign the modification on Tyr-103.

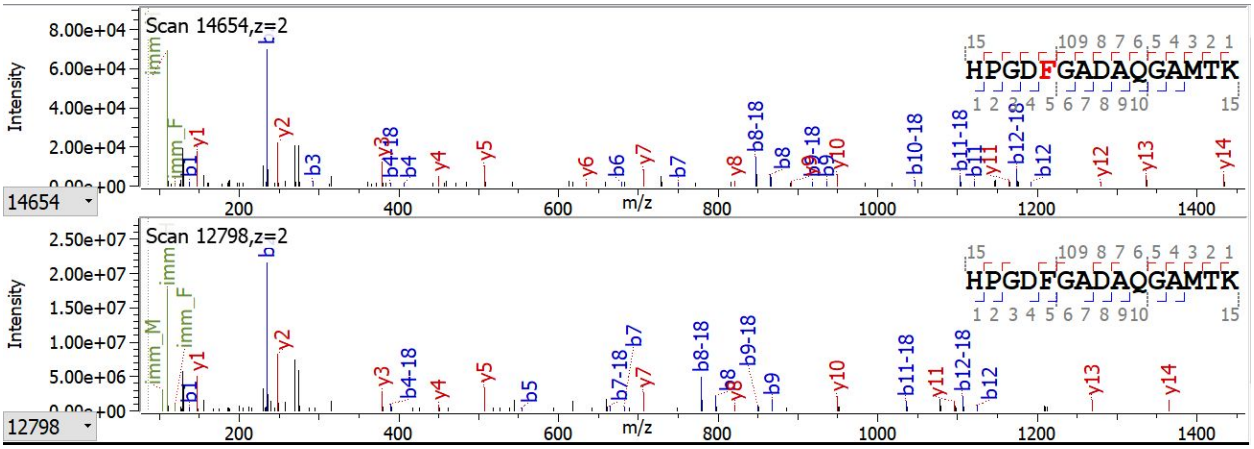

Figure S16. Product-ion spectrum of CF_3-modification on Phe-123 from MS/MS of doubly charged unmodified peptide 119-133 (bottom) and CF_3-modified peptide (top) from apo-myoglobin. The y10 and y11 assign the modification on Phe-123.

5. References

1. Imiołek, M.; Karunanithy, G.; Ng, W.-L.; Baldwin, A. J.; Gouverneur, V.; Davis, B. G., Selective Radical Trifluoromethylation of Native Residues in Proteins. *J. Am. Chem. Soc.* **2018,** *140*, 1568-1571.
2. Ichiishi, N.; Caldwell, J. P.; Lin, M.; Zhong, W.; Zhu, X.; Streckfuss, E.; Kim, H.-Y.; Parish, C. A.; Krska, S. W., Protecting group free radical C–H trifluoromethylation of peptides. *Chem. Sci.* **2018,** *9*, 4168-4175.
3. Li, L.; Mu, X.; Liu, W.; Wang, Y.; Mi, Z.; Li, C.-J., Simple and Clean Photoinduced Aromatic Trifluoromethylation Reaction. *J. Am. Chem. Soc.* **2016,** *138*, 5809.
4. Ji, Y.; Brueckl, T.; Baxter, R. D.; Fujiwara, Y.; Seiple, I. B.; Su, S.; Blackmond, D. G.; Baran, P. S., Innate C-H trifluoromethylation of heterocycles. *Proc. Natl. Acad. Sci. U S A* **2011,** *108*, 14411.
5. Shtarev, A. B.; Tian, F.; Dolbier, W. R.; Smart, B. E., Absolute Rates of Intermolecular Carbon−Hydrogen Abstraction Reactions by Fluorinated Radicals. *J. Am. Chem. Soc.* **1999,** *121*, 7335.
6. Liguori, L.; Bjørsvik, H.-R.; Bravo, A.; Fontana, F.; Minisci, F., A new direct homolytic iodination reaction of alkanes by perfluoroalkyl iodides. *Chem. Commun.* **1997,** (16), 1501.
7. Zhang, L.; Cradlebaugh, J.; Litwinienko, G.; Smart, B. E.; Ingold, K. U.; Dolbier, J. W. R., Absolute rate constants for some hydrogen atom abstraction reactions by a primary fluoroalkyl radical in water. *Org. Biomol. Chem.* **2004,** *2*, 689.
8. Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L., Laser-Initiated Radical Trifluoromethylation of Peptides and Proteins: Application to Mass-Spectrometry-Based Protein Footprinting. *Angew. Chem. Int. Ed. Engl.* **2017,** *56*, 14007.
9. Tong, X.; Wren, J. C.; Konermann, L., Effects of Protein Concentration on the Extent of γ-Ray-Mediated Oxidative Labeling Studied by Electrospray Mass Spectrometry. *Anal. Chem.* **2007,** *79*, 6376-6382.
10. Zhang, H.; Wen, J.; Huang, R. Y. C.; Blankenship, R. E.; Gross, M. L., Mass spectrometry-based carboxyl footprinting of proteins: method evaluation. *Int. J. Mass. Spectrom.* **2012,** *312*, 78-86.
11. Mendoza, V. L.; Vachet, R. W., Protein Surface Mapping Using Diethylpyrocarbonate with Mass Spectrometric Detection. *Anal. Chem.* **2008,** *80*, 2895-2904.
12. Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L., Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* **2009,** *81*, 6563-6571.
13. Willwacher, J.; Raj, R.; Mohammed, S.; Davis, B. G., Selective Metal-Site-Guided Arylation of Proteins. *J. Am. Chem. Soc.* **2016,** *138*, 8678-8681.