Multiplex chemical labeling of amino acids for protein footprinting structure assessment

Rohit Jain[1,2], Nanak Dhillon[3], Erik R. Farquhar[1], Benlian Wang[2], Xiaolin Li[2], Janna Kiselar[2,3], Mark R. Chance[1,2,3*]

[1]Center for Synchrotron Biosciences, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; [2]Center for Proteomics and Bioinformatics, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; [3]Department of Nutrition, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

ABSTRACT: Protein footprinting with mass spectrometry is an established structural biology technique for mapping solvent accessibility and assessing molecular-level interactions of proteins. In hydroxyl radical protein footprinting (HRPF), hydroxyl (OH) radicals generated by water radiolysis or other methods covalently label protein side chains. Due to the wide dynamic range of OH reactivity, not all side chains are easily detected in a single experiment. Novel reagent development and the use of radical chain reactions for labeling, including trifluoromethyl radicals, is a potential approach to normalize the labeling across a diverse set of residues. HRPF in the presence of a trifluoromethylation reagent under the right conditions could provide a "one-pot" reaction for multiplex labeling of protein side chains. Towards this goal, we have systematically evaluated amino acid labeling with the recently investigated Langlois' reagent activated by X-ray mediated water radiolysis, followed by three different mass spectrometry methods to compare the reactivity of CF_3 and OH radical labeling for all 20 protein side chains. Our investigations provide the evidence and knowledge set to perfect hydroxyl radical activated trifluoromethyl chemistry as "one-pot" reaction for multiplex labeling of protein side chains to achieve higher resolution in HRPF.

INTRODUCTION

Footprinting is an increasingly popular technique employed in biophysics and structural biology studies for understanding macromolecular structure, macromolecular recognition and dynamics.[1-4] One of the most widely used footprinting techniques, Hydroxyl Radical Footprinting (HRF), employs hydroxyl radicals (•OH) to probe solvent accessible surfaces of macromolecules, with successful application to a wide range of protein and nucleic acid systems.[5-9] Some of the most appealing features of HRF are that it penetrates all solvent-accessible areas of a macromolecule and provides irreversible covalent labeling which enhances detection for modified species and allows quantification in a wide range of solution contexts. The HRF workflow for proteins (HRPF) can be divided into several steps, where the first step involves the generation of •OH, the second step comprises a controlled •OH induced covalent modification of protein side chains, with subsequent steps including accurate detection and precise quantification of modifications by mass spectrometry.[10] Many methods of generating hydroxyl radicals have been developed over recent decades and are currently routinely employed, including radiolysis of water by X-rays, gamma rays, electron beams, electric discharge or a plasma source, decomposition of hydrogen peroxide using transition metal based Fenton chemistry and photolysis of hydrogen peroxide using lasers or a high-pressure flash oxidation lamp.[11-15]

Independent of the method of production, •OH are highly reactive species and have van der Waals surface area and solvent properties similar to those of water molecules.[14] •OH are polar moieties, react irreversibly with side chains on the order of microseconds and can routinely modify 12 out of the 20 amino acids in HRPF experiments. These properties make HRPF a highly valuable tool in structural biology. However, challenges to the approach are still significant. For example, the •OH modifications are chemically complex with multiple mass changes, and have at least 1000-fold range of reactivity towards the 20 amino acid side chains.[15] In particular, the side chains of threonine, serine, asparagine, glutamine, alanine, and glycine, as well as aspartate and glutamate amino acids exhibit a lower apparent reactivity to •OH and modifications are typically less frequently observed relative to much more reactive sulfur-containing and aromatic amino acids, while aliphatic amino acids such as valine lie in between these extremes. Additionally, the •OH labeling of Met-containing sequences produces predominantly modification of Met residue across these sequences, effectively "silencing" the ability to detect any modifications of other residues. Although Met is not a common residue in proteins (~2% frequency), its presence on 1 out of 5 typical tryptic peptides makes quantification of other residues on those peptides challenging.[16] Further, the over labeling of Met in HRF experiments due to its high reactivity and high background oxidation makes even the Met data of lower quality, resulting in effectively no labeling coverage for such peptides.[17-18] Overall, low labeling propensity of amino acids with low reactivity to •OH reagent and the high labeling propensity of Met residues both contribute to decreasing the overall impact of HRPF.

Hydroxyl radical labeling is in a class of chemistries that target a wide range of amino acids, and selective amino acid chemistries are available to expand labeling coverage to almost every side chain.[19-23] However, coverage at the peptide and protein level using a single reagent, even hydroxyl radical, is modest with only 5-20% of residues detected as labeled in typical experiments. Chemical reagents and workflows that can produce multiple labels in a "one-pot synthesis" could help further advance HRPF towards being an accessible high-resolution structural biology technique. Towards this end, electrophilic and radical mediated trifluoromethyl (TFM, - CF_3) labeling reagents may provide a simplified, uniform $+CF_3$ labeling (+68 Da) with an overall dynamic range of reactivity with side-chains that permits "low reactivity" side chains to be more easily observed.[24-26] The TFM group is an important synthetic motif in pharmaceuticals, agrochemicals, and specialty materials because CF_3-substituted molecules often possess improved physical and chemical properties to lend stability to molecules in which they are introduced.[25-26] Thus, a wide range of interesting TFM reagents are available for research use.[27]

TFM labeling of large biomolecules for purposes of HRPF has been shown using both radiolytic and photolytic labeling approaches providing motivation for evaluating TFM chemistry as a tool for improving HRPF.[28-34] This potential value of TFM to improving the structural resolution of protein footprinting justifies a rigorous analysis of TFM labeling of free side chains of amino acids. Herein, we demonstrate the •CF_3 and •OH based multiplex labeling of 20 natural amino acids in biocompatible conditions using synchrotron mediated water radiolysis. Our investigations will help drive adoption of TFM chemistry for different HRPF based platforms to ultimately perfect one-pot multiplex labeling reactions for high resolution HRPF.

METHODS

Materials. 17 amino acid amides of form L-XXX-NH₂ (XXX represents all amino acids except Cys, Lys, Pro), L-Cys-OH, L-Lys-OH and L-Pro-OH were purchased from Chem-Impex International Inc. (Wood Dale, IL). All had a purity at least 96% and were used without further purification. Langlois' reagent (Sodium triflinate, CF_3SO_2Na) was purchased from Millipore Sigma (Burlington, MA) and had a purity of at least 95%. B & J Brand HPLC grade water and acetonitrile (ACN) were purchased from Honeywell International Inc. (Charlotte, NC).

Exposure of samples to X-rays. Amino acid samples were exposed to X-rays on the XFP beamline at port 17-BM at the National Synchrotron Light Source II (NSLS-II, Brookhaven National Laboratory, Upton, NY) at 25 µM concentration by dissolving amino acids in 2 mM ammonium acetate solution ($NH_4CH_3CO_2$). Solution samples were exposed in 5 µl sample droplets held by surface tension on the bottom of 200 µl PCR tubes, producing a 2.5 mm diameter droplet well matched to the X-ray beam size. A fixed 76 µm thick aluminum attenuation was selected to provide high photon flux on the sample. X-ray exposures were performed at 25 °C for 30 ms at a constant NSLS-II ring current of 400 mA using a recently reported 96-well HT device.[35-36] In some cases, X-ray exposures were performed at multiple exposure times of 0 ms, 12 ms, 20 ms, and 30 ms to calculate the rate constants in dose-response studies.

After irradiation, all samples were flash-frozen in liquid nitrogen and then stored at -80 °C prior mass spectroscopic analysis.

Mass spectrometry analysis. For direct infusion, the irradiated samples were analyzed by electrospray mass spectrometry (ESI-MS) without chromatographic separation. The diluted samples containing 50% Acetonitrile (ACN) and 0.1 % Formic acid (FA) were infused at a flow rate of 10 µl/min. Full mass spectra were acquired in positive polarity on Thermo Q Exactive™ Plus hybrid quadrupole Orbitrap™ mass spectrometer coupled with a ESI source. UPLC techniques employing Hydrophilic Interaction Liquid Chromatography (HILIC) and nonpolar reverse phase (C_{18}) columns were used to separate the nonvolatile Langlois' reagent post X-ray exposure from amino acids.

For HILIC-MS analysis, a Waters ACQUITY UPLC was coupled online to a Thermo Q Exactive™ Plus hybrid quadrupole Orbitrap™ mass spectrometer. Separation was performed using a Waters ACQUITY BEH Amide column (130 Å, 1.7 µm, 2.1 mm * 150 mm) and two mobile phase systems. Mobile phase A (0.1% FA in water) and mobile phase B (0.1% FA in ACN) were utilized to design a LC-gradient. The 1 µl of sample (~ 0.2 - 0.5 ng) was separated on the column using a flow rate of 100 µl/min. The LC gradient was design as following: gradient started with 98% B at 0.1 min followed by a gradient change to 50% B until 7 min and remained at 50% B from 7 min to 12 min. From 12-13 min the gradient was changed to 98 % B solvent and then the column was equilibrated to 98 % B solvent for 17 min. During the first 6 min of the run, the flow was diverted to waste using a six-port external valve on the instrument and then re-directed to the MS system for the duration of the LC-MS run. The column temperature was maintained at 25 °C. Full MS spectra were acquired in a positive polarity and the resolution was set to 70,000, with an AGC target of $5 * 10^5$ ions and maximum injection time was 100 ms.

For C_{18} chromatographic separation, we utilized Waters ACQUITY Peptide BEH column (300 Å, 1.7 µm, 75 µm * 250 mm) and two mobile phase systems similar to HILIC-MS analysis. Samples were analyzed by coupling Waters nanoACQUITY UPLC to a Thermo Orbitrap™ Eclipse Tribrid mass spectrometer. The 15 µl of sample (~ 13 - 37 ng) containing irradiated amino acid was desalted and concentrated using Waters ACQUITY C_{18} trap column (100 Å, 5 µm, 180 µm * 20 mm) for 4 min at a flow rate of 10 µl/min with 0.1% FA. The amino acids then were eluted from Waters ACQUITY C_{18} analytical column (100 Å, 75 µm, 250 mm) using LC gradient from 0% B to 25 % for 38 min. The gradient was held at 25% B from 38 to 43 min to allow complete elution of the sample from C18 column. The column was then washed with 95% B from 43 to 49 min and re-equilibrated with 1% B solvent for 11 min. The analytical column temperature was maintained at 40 °C. Full MS spectra were acquired in a positive polarity and the resolution was set to 120,000, with an AGC target of $4 * 10^5$ ions and maximum injection time was 50 ms.

Data Processing. The MS data were manually searched for CF_3 and OH modifications for 20 amino acids, which were calculated based on literature parameters.[18, 29-30] The extent of modification for each amino acid was measured using peak intensity of mass spectral signals for unmodified and their modified ion

signal from infusion mass spectrometry analysis. The modification rate constants in dose response study were obtained by exponential fitting of intensity of unmodified amino acids as a function of exposure time to a first-order kinetics equation as previously described.[18] The data from LC-MS analysis were used to measure the extent of modification (modified fraction) of amino acids for CF$_3$ and OH. The modified fraction was calculated as the ratio of LC peak areas of CF$_3$ or OH modifications to the sum of LC peak areas of unmodified and all modified products. The raw MS data files were acquired in Xcalibur and data visualization was carried out using OriginLab software.

RESULTS AND DISCUSSION

Generation of CF$_3$ radicals from sodium triflinate. Synchrotron footprinting typically employs highly focused X-ray beams for water radiolysis without addition of reagents and the generated •OH can activate Langlois' reagent to form •CF$_3$.[29] To monitor •OH flux, we have previously described a dose-response reporter assay, in which the loss of Alexa488 fluorescence due to modification of the fluorophore correlates to total •OH dose as defined by the exposure time and X-ray flux.[37] We used this Alexa488 based fluorescence assay to monitor the •OH scavenging effect of varying levels of Langlois' reagent under different X-ray beam conditions. X-ray beam flux was varied via use of an eight-position attenuator wheel (762 µm – 0 µm) at 17-BM beamline (NSLS II) with beam directed onto samples loaded onto a recently described 96-well high through-put apparatus.[36] Using this apparatus, we can vary incident X-ray flux on the sample nearly 20-fold and with resulting •OH yield in 30 ms exposure varying 40-fold by selecting different aluminum attenuator thicknesses (Table S1).
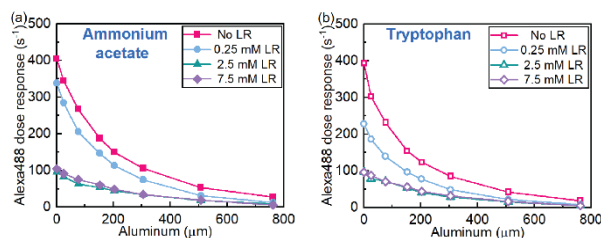

Figure 1. The loss of fluorescence of dosimeter dye (Alexa488) was monitored for (a) 2 mM Ammonium acetate (filled symbols) and (b) 25 µM Tryptophan (empty symbols) in 2 mM ammonium acetate for different Langlois' reagent (LR) concentrations and at varying aluminum attenuator thicknesses (and thus X-ray flux conditions). The decrease in Alexa488 fluorescence against increase in X-ray exposure time was fitted to exponential first order rate equation at an aluminum attenuator thickness and calculated rate constants were plotted or all conditions.

The •OH scavenging effect of the Langlois' reagent revealed by decreases of Alexa488 fluorescence decay constants was examined at various concentrations (0.25 mM – 7.5 mM) and in moderate (203 µm Al) to highest X-ray flux (No Al) conditions (Figure 1a). Minimal changes in the •OH scavenging effect are seen for Langlois' reagent concentrations greater than 2.5 mM. This represent a reasonable limit beyond which additional reagent is considerably less effective. The •OH scavenging effect is slightly higher in the presence of 25 µM Trp, both in absence and presence of Langlois' reagent due to competition for •OH

by Trp (Figure 1b). These observations provide guidance in selecting an optimal •OH dose for exposure of amino acid samples. Overall these data suggest that Alexa488 fluorescence decay as described here can be utilized for optimizing •OH dose to generate •CF$_3$ from Langlois' reagent.

X-rays mediated multiplex labeling on amino acid side chains. Three amino acids (Trp, His and Tyr) were labeled with synchrotron radiolysis in the presence of Langlois' reagent and analyzed by direct infusion ESI-MS. Labeling conditions were optimized by first testing different X-ray flux (203 µm AL and 76 µm AL) and then by using different amino acid concentrations (100 µM and 25 µM).
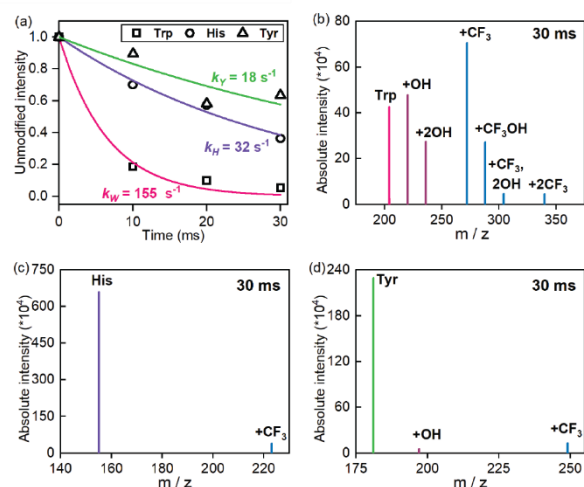

Figure 2. (a) The decrease in unmodified intensity of selected amino acids (Tyr, His and Tyr) was plotted against increase in X-ray exposure time and rate constants were calculated by fitting data to exponential first order rate equation. Infusion MS spectra of CF$_3$ and OH modified (b) Tryptophan, (c) Histidine and (d) Tyrosine amino acids show detected TFM and OH modifications. Amino acids at 25 µM concentrations were exposed with X-ray in presence of 7.5 mM Langlois' reagent.

The decrease of unmodified amino acid fraction for each amino acid was plotted against X-ray exposure time and fitted to a first order rate equation. Trp, His and Tyr amino acids in Langlois' reagent exhibited an increased trifluoromethylation and oxidation as a function of increased X-ray exposure. However, at 25 µM amino acid concentration and 76 µm AL, the modification rate of Trp (155 s^{-1}) was 4.8 times higher than His (32 s^{-1}) and 8.6 times higher than Tyr (18 s^{-1}) (Figure 2a). Figure 2 shows the mass spectra of 25 µM Trp-NH2, His-NH2 and Tyr-NH2 following independent exposure to X-rays for 30 ms in the presence of 7.5 mM Langlois' reagent. Trp shows a complex series of modified products, where both •CF$_3$ and •OH modified products are present in substantial quantities (Figure 2b). His shows •CF$_3$ modification while •OH modified products for His were not detected by direct infusion ESI-MS (Figure 2c). For Tyr, both •CF$_3$ and •OH modified products were detected (Figure 2d).

The modification rate constants for all three amino acids increased at higher X-ray flux conditions (203 µm Al to 76 µm Al) and at lower amino acid concentrations (100 µM to 25 µM)

3

(Figures S1-2). The measured reactivity order of Trp > His > Tyr was observed at all experimental conditions. We noted that both TFM and OH modifications on amino acids occur on the same timescale of X-ray exposures (Table S2). The higher modification rate of amino acids (Trp, His and Tyr) at higher X-ray flux (76 µm Al) in comparison to lower X-ray flux (203 µm Al) was driven mostly by CF₃ modifications and could be due to the increased generation of CF_3 radicals from Langlois' reagent at higher X-ray flux, providing evidence that TFM chemistry can be tuned (Table S2). These observations indicate conditions to base development of a "one-pot" reaction for multiplex labeling of protein side chains to achieve higher resolution in HRPF. Further experiments included exposing all amino acid samples to X-rays on the XFP beamline at port 17-BM at the National Synchrotron Light Source II (NSLS-II, Brookhaven National Laboratory, Upton, NY) at 25 µM concentration by dissolving amino acids in 2 mM ammonium acetate solution ($NH_4CH_3CO_2$). Solution samples were exposed in 5 µl sample droplets held by surface tension on the bottom of 200 µl PCR tubes, producing a 2.5 mm diameter droplet well matched to the X-ray beam size. A fixed 76 µm thick aluminum attenuation was selected to provide high photon flux on the sample. X-ray exposures were performed at 25 °C for 30 ms at a constant NSLS-II ring current of 400 mA using a recently reported 96-well HT device. In some cases, X-ray exposures were performed at multiple exposure times of 0 ms, 12 ms, 20 ms, and 30 ms to calculate the rate constants in dose-response studies. After irradiation, all samples were flash-frozen in liquid nitrogen and then stored at -80 °C prior mass spectroscopic analysis.

Chromatographic analysis of TFM and hydroxyl modifications for 20 amino acids. All 20 amino acids were exposed to x-rays for 30 ms (76 µm Al) where sodium triflinate was added to a final concentration of 7.5 mM based on the prior optimization. Although direct infusion sufficed as a method for examining Trp, Tyr, and His reactivity analysis, we observed that during mass spectrometry measurements, Langlois' reagent was found to significantly reduce the sensitivity of the ion source because of its nonvolatile nature. Consequently, we did not conduct detailed reactivity assessments for the remaining amino acid samples using direct infusion ESI-MS methods. Instead, we used two separate chromatographic techniques in combination with ESI-MS to separate the nonvolatile sodium triflinate salt from X-ray exposed amino acids to increase the detection sensitivity of modified fractions at lower abundance or with reduced ionization efficiency. First, we used Hydrophilic interaction chromatography (HILIC), which has hydrophilic stationary phases where analytes elute in order of increasing polarity or decreasing hydrophobicity.[38] Secondly, we used reverse phase chromatography (C₁₈), which has hydrophobic alkyl chains as stationary phases, in which case analytes elute in order of decreasing polarity or increasing hydrophobicity.[39]

The TFM and OH modified amino acids elute separately on HILIC and C₁₈ chromatographic columns (and separately from unmodified amino acids) due to their opposite polarity (Figure 3, Tables S3-4). As an example, for an experiment with tyrosine using C₁₈ separation, the hydroxyl modified tyrosine (RT = 4.6, 5.6 min) eluted just prior to unmodified tyrosine (RT = 5.5 min) due to higher polarity. In contrast, CF₃ modified tyrosine (RT =

21.5 min) eluted at a significant higher retention time on the C₁₈ column due to its higher hydrophobicity. The order of OH modified (RT = 7.1 min), unmodified (RT = 7.8 min) and CF₃ modified (RT = 8.1 min) peaks of tyrosine reverses on a HILIC column in comparison to C₁₈. This order reversal occurs due to the hydrophilic stationary column of HILIC column. All the CF₃ modified amino acids elute at a higher retention time on a C₁₈
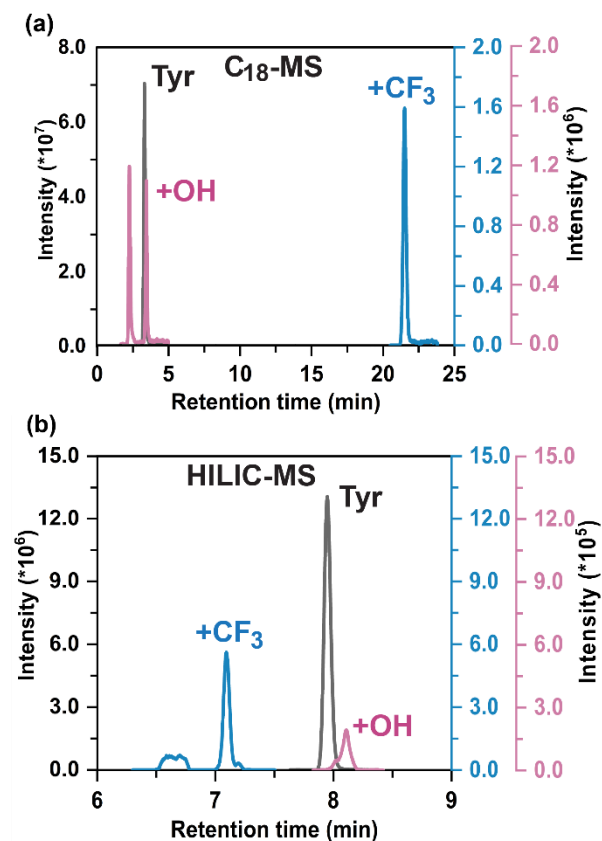

Figure 3. Extracted chromatographic profile of Tyrosine (Tyr) for unmodified (black), CF₃ (blue) and OH (pink) modified peaks on (a) nonpolar reverse phase C₁₈ and (b) polar HILC columns. Left *y-axis* represent unmodified intensity and right *y-axis* represent CF₃ (blue) and OH (pink) MS intensity.

column in comparison to their unmodified and hydroxyl modified counterparts. The retention times for different CF₃ modified amino acids can vary from one another, presumably due to differences in the hydrophobicity of different amino acids. The marked differences in retention times for TFM and hydroxyl modifications provide an additional mechanism for increasing the resolution of HRPF using TFM-based "one-pot" labeling, as •CF₃ modified peaks (in the context of peptide based elution profiles) will likely have elution profiles distinct from the unmodified and the OH modified forms.

Reactivity of 20 protein side chains towards TFM and hydroxyl radicals. We observed complex patterns in labeling and retention behavior of amino acids for TFM and hydroxyl modifications across the C₁₈-MS and HILIC-MS setups (Figures 4, S3, Tables S3-4). Out of 20 amino acids analyzed with C₁₈-MS setup, TFM modifications were detected for 19 amino acids and hydroxyl modifications were detected for 14 amino

4

acids (Figure 4). In comparison, out of 18 amino acids analyzed with HILIC-MS setup, TFM modifications were detected for 7 amino acids and hydroxyl modifications were detected for 14 amino acids (Figures S3a-b). The observed differences for labeling of amino acids between separation approaches could be due to the need for loading low sample amounts on the HILIC column, low retention of hydrophobic TFM modifications on the polar HILIC column compared to C_{18} column and the diversion of LC flow to waste for first 6 min of the run to wash out nonvolatile Langlois' reagent for HILIC-MS (not needed for C_{18}-MS).
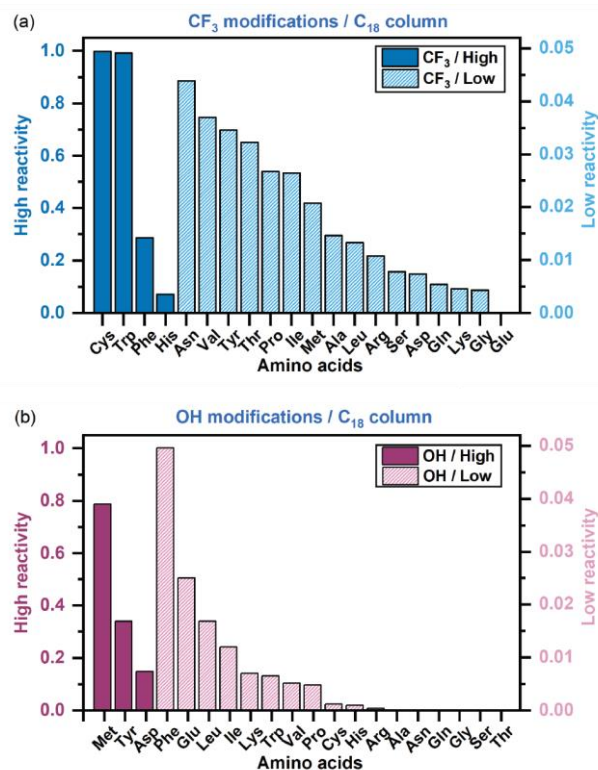

Figure 4. Reactivity of 20 natural amino acids towards (a) CF_3 (blue bars) and (b) OH modifications (pink bars) as analyzed via nonpolar reverse phase (C_{18}) column coupled mass spectrometer. Amino acids at 25 μM concentrations were exposed with X-ray for 30 ms in presence of 7.5 mM Langlois' reagent. The high reactivity (solid bars) on left *y-axis* represent modification fraction between 1.0 – 0.05; the low reactivity (pattern bars) on right *y-axis* represent modification fraction between 0.05 – 0.0 and they were plotted separately for clarity.

Overall, in this example of "one-pot" multiplex chemistry on both C_{18}-MS and HILIC setups, we found that all 20 amino acids exhibit either TFM or OH labeling in the presence of Langlois' reagent (Figure 5). 19 out of 20 amino acids show TFM reactivity, and more impressively 17 exhibit both CF_3 and OH reactivity. Glutamic acid was the only amino acid not found to show CF_3 modifications, but it exhibited OH reactivity in the presence of Langlois' reagent. The tabulated retention times for both HILIC-MS and C_{18}-MS alongside reactivity data for TFM and hydroxyl modifications of 20 protein side chains in the presence of Langlois' reagent will provide important clues for detecting peptide based modifications.

We can compare the intrinsic reactivity of 20 amino acids on the basis of their TFM and hydroxyl modification fraction after C_{18}-MS analysis (Figure 4). In the presence of Langlois' reagent, the amino acids with the highest reactivity towards TFM modifications were Cys > Trp > Phe > His, while Met > Tyr > Asp showed the highest OH modifications. The dynamic range of reactivity for these seven highly reactive amino acids is within 14 for either the •CF_3 or •OH channel. For example, the TFM modification fraction of the most reactive amino acid, Cys (0.99), is 14 times higher than the TFM modification fraction of His (0.07) at same X-ray exposure conditions. Amino acids that normally have the highest reactivity towards •OH in HRPF (Cys, Trp, Phe and His) were effectively directed towards TFM modifications in the presence of Langlois' reagent. Aliphatic amino acids (Val > Ile > Leu) and the basic amino acid, Arg show intermediate reactivity towards •CF_3 and have a dynamic range within 100 for •CF_3 channel by the above logic of analysis. Amino acids with low OH reactivity in HRPF (Ala, Asn, Glu, Thr) show intermediate reactivity in Langlois' reagent and have a dynamic range within 100 for •CF_3 or •OH. In summary, 15 amino acids out of 20 protein side chains on C_{18}-MS setup showed a dynamic range of < 100 for either •CF_3 or •OH in Langlois' reagent. In addition, amino acids with undetected OH reactivity (Ala, Asn, Gln, Gly, Ser and Thr) all show CF_3 reactivity in Langlois' reagent.

We have earlier alluded that high reactivity of Met amino acid towards •OH in HRPF dominates oxidation to the extent of effectively "silencing" the ability to detect any modifications of other residues. We observed that Met in the presence of Langlois' reagent shows very low CF_3 modified fractions (2%) compared to the OH modified fraction (79%) (Figure 4). The low overall Met reactivity will allow both OH and CF_3 modifications on other residues in Met-containing sequences to be efficiently detected.
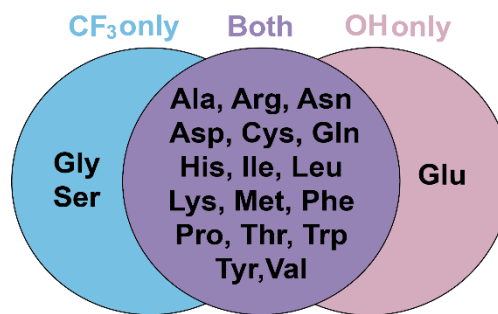

Figure 5. Venn diagram showing the distribution of 20 amino acids for their CF_3 and OH modifications.

In our experiments, we have used a 300-fold excess of Langlois' reagent relative to the amino acid and •OH concentrations to favor •CF_3 vs •OH modifications. Varying the concentration ratio of Langlois' reagent and/or photon flux can tune reactivity between labeling channels, providing an opportunity for further experimentation and improvements. The current results provide a baseline dataset for understanding the intrinsic reactivity of freely accessible side chains of amino acids for •CF_3 and •OH modifications in the presence of Langlois' reagent. This is essential to an understanding of these modifications in the protein

environment, where solvent accessibility is another key criterion for radiolytic modifications.

CONCLUSIONS

A TFM reagent, Langlois' reagent, can be used for efficient multiplex labeling of amino acids within proteins, representing structural probes for protein footprinting experiments. The •CF_3 radical modifications complement •OH radical footprinting and surmount some of the reactivity challenges inherent to using •OH alone. The knowledge of modifications and retention time from our work can be used to locate the modifications for future protein footprinting experiments. In addition to expanding coverage in footprinting experiments, X-ray mediated water radiolysis of TFM reagents could be an approach to readily incorporate the TFM moieties into native peptides and proteins under biocompatible conditions to, for instance, enable structural investigations with ^{19}F NMR spectroscopy or synthesize relevant peptidomimetics with unique bio-medicinal and pharmaceutical properties.[31, 40-41]

ASSOCIATED CONTENT

Supporting Information

The comparison of modification rate between high and low X-ray flux, (Figure S1, S-2) comparison of modification rate between 100 μM and 25 μM amino acids (Figure S2, S-3), reactivity of 20 natural amino acids towards •CF_3 and •OH modifications as analyzed by polar HILIC column (Figure S3, S-4), calculated photon flux, beam power, and •OH yield at the XFP beamline for several attenuator thicknesses (Table S1, S-5), comparison of CF_3 and OH modification rates of 3 amino acids between low and high X-ray flux (Table S2, S-6), retention time along with CF_3 and OH modifications of amino acids as detected via nonpolar reverse phase (C_{18}) coupled mass spectrometry (Table S3, S-7, S-8) and retention time along with CF_3 and OH modifications of amino acids as detected via polar HILIC column coupled mass spectrometry (Table S4, S-9).

The Supporting Information is available free of charge on the ACS Publications website.

SI Multiplex chemical labeling of amino acids (PDF)

AUTHOR INFORMATION

Corresponding Author

*Mark R. Chance - Center for Synchrotron Biosciences, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; Center for Proteomics and Bioinformatics, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; Department of Nutrition, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA;
*Email: mrc16@case.edu

Authors

Rohit Jain - Center for Synchrotron Biosciences, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; Center for Proteomics and Bioinformatics, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; Department of Nutrition, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

Nanak Dhillon – Department of Nutrition, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

Erik R. Farquhar – Center for Synchrotron Biosciences, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; Department of Nutrition, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

Benlian Wang – Center for Proteomics and Bioinformatics, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; Department of Nutrition, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

Xiaolin Li – Center for Proteomics and Bioinformatics, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; Department of Nutrition, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

Janna Kiselar - Center for Proteomics and Bioinformatics, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA; Department of Nutrition, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

Author Contributions

R.J., E.R.F., B.W., X.L. J.K. and M.R.C. designed the methods. R.J. prepared and analyzed samples. R.J. and N.D. performed data analysis. R.J., N.D., J.K. and M.R.C. contributed to writing and editing the manuscript. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

REFERENCES

1. Chance, M. R.; Farquhar, E. R.; Yang, S.; Lodowski, D. T.; Kiselar, J., Protein Footprinting: Auxiliary Engine to Power the Structural Biology Revolution. *Journal of molecular biology* **2020,** *432* (9), 2973-2984.
2. Jain, R.; Muneeruddin, K.; Anderson, J.; Harms, M. J.; Shaffer, S. A.; Matthews, C. R., A conserved folding nucleus sculpts the free energy landscape of bacterial and archaeal orthologs from a divergent TIM barrel family. *Proceedings of the National Academy of Sciences of the United States of America* **2021,** *118* (17).
3. Liu, X. R.; Zhang, M. M.; Gross, M. L., Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chemical reviews* **2020,** *120* (10), 4355-4454.

4.   Vierstra, J.; Stamatoyannopoulos, J. A., Genomic footprinting. *Nature methods* **2016,** *13* (3), 213-21.

5.   Adilakshmi, T.; Soper, S. F.; Woodson, S. A., Structural analysis of RNA in living cells by in vivo synchrotron X-ray footprinting. *Methods in enzymology* **2009,** *468*, 239-58.

6.   Du, Y.; Duc, N. M.; Rasmussen, S. G. F.; Hilger, D.; Kubiak, X.; Wang, L.; Bohon, J.; Kim, H. R.; Wegrecki, M.; Asuru, A.; Jeong, K. M.; Lee, J.; Chance, M. R.; Lodowski, D. T.; Kobilka, B. K.; Chung, K. Y., Assembly of a GPCR-G Protein Complex. *Cell* **2019,** *177* (5), 1232-1242.e11.

7.   Kiselar, J.; Chance, M. R., High-Resolution Hydroxyl Radical Protein Footprinting: Biophysics Tool for Drug Discovery. *Annual review of biophysics* **2018**.

8.   Schoof, M.; Faust, B.; Saunders, R. A.; Sangwan, S.; Rezelj, V.; Hoppe, N.; Boone, M.; Billesbølle, C. B.; Puchades, C.; Azumaya, C. M.; Kratochvil, H. T.; Zimanyi, M.; Deshpande, I.; Liang, J.; Dickinson, S.; Nguyen, H. C.; Chio, C. M.; Merz, G. E.; Thompson, M. C.; Diwanji, D.; Schaefer, K.; Anand, A. A.; Dobzinski, N.; Zha, B. S.; Simoneau, C. R.; Leon, K.; White, K. M.; Chio, U. S.; Gupta, M.; Jin, M.; Li, F.; Liu, Y.; Zhang, K.; Bulkley, D.; Sun, M.; Smith, A. M.; Rizo, A. N.; Moss, F.; Brilot, A. F.; Pourmal, S.; Trenker, R.; Pospiech, T.; Gupta, S.; Barsi-Rhyne, B.; Belyy, V.; Barile-Hill, A. W.; Nock, S.; Liu, Y.; Krogan, N. J.; Ralston, C. Y.; Swaney, D. L.; García-Sastre, A.; Ott, M.; Vignuzzi, M.; Walter, P.; Manglik, A., An ultrapotent synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive Spike. *Science (New York, N.Y.)* **2020,** *370* (6523), 1473-1479.

9.   Tetter, S.; Terasaka, N.; Steinauer, A.; Bingham, R. J.; Clark, S.; Scott, A. J. P.; Patel, N.; Leibundgut, M.; Wroblewski, E.; Ban, N.; Stockley, P. G.; Twarock, R.; Hilvert, D., Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science (New York, N.Y.)* **2021,** *372* (6547), 1220-1224.

10.   Wang, L.; Chance, M. R., Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. *Analytical chemistry* **2011,** *83* (19), 7234-41.

11.   Li, K. S.; Shi, L.; Gross, M. L., Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization. *Accounts of chemical research* **2018,** *51* (3), 736-744.

12.   Minkoff, B. B.; Blatz, J. M.; Choudhury, F. A.; Benjamin, D.; Shohet, J. L.; Sussman, M. R., Plasma-Generated OH Radical Production for Analyzing Three-Dimensional Structure in Protein Therapeutics. *Scientific reports* **2017,** *7* (1), 12946.

13.   Sharp, J. S.; Chea, E. E.; Misra, S. K.; Orlando, R.; Popov, M.; Egan, R. W.; Holman, D.; Weinberger, S. R., Flash Oxidation (FOX) System: A Novel Laser-Free Fast Photochemical Oxidation Protein Footprinting Platform. *Journal of the American Society for Mass Spectrometry* **2021,** *32* (7), 1601-1609.

14.   Takamoto, K.; Chance, M. R., Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annual review of biophysics and biomolecular structure* **2006,** *35*, 251-76.

15.   Xu, G.; Chance, M. R., Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical reviews* **2007,** *107* (8), 3514-43.

16.   Xu, G.; Takamoto, K.; Chance, M. R., Radiolytic modification of basic amino acid residues in peptides: probes for examining protein-protein interactions. *Analytical chemistry* **2003,** *75* (24), 6995-7007.

17.   Xu, G.; Chance, M. R., Radiolytic modification of sulfur-containing amino acid residues in model peptides: fundamental studies for protein footprinting. *Analytical chemistry* **2005,** *77* (8), 2437-49.

18.   Xu, G.; Chance, M. R., Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Analytical chemistry* **2005,** *77* (14), 4549-55.

19.   Chen, J.; Cui, W.; Giblin, D.; Gross, M. L., New protein footprinting: fast photochemical iodination combined with top-down and bottom-up mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2012,** *23* (8), 1306-18.

20.   Gau, B. C.; Chen, H.; Zhang, Y.; Gross, M. L., Sulfate Radical Anion as a New Reagent for Fast Photochemical Oxidation of Proteins. *Analytical chemistry* **2010,** *82* (18), 7821-7827.

21.   Guo, C.; Cheng, M.; Gross, M. L., Protein-Metal-Ion Interactions Studied by Mass Spectrometry-Based Footprinting with Isotope-Encoded Benzhydrazide. *Analytical chemistry* **2019,** *91* (2), 1416-1423.

22.   Jumper, C. C.; Schriemer, D. C., Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Analytical chemistry* **2011,** *83* (8), 2913-20.

23.   Kaur, P.; Kiselar, J.; Yang, S.; Chance, M. R., Quantitative protein topography analysis and high-resolution structure prediction using hydroxyl radical labeling and tandem-ion mass spectrometry (MS). *Molecular & cellular proteomics : MCP* **2015,** *14* (4), 1159-68.

24.   Imiołek, M.; Karunanithy, G.; Ng, W. L.; Baldwin, A. J.; Gouverneur, V.; Davis, B. G., Selective Radical Trifluoromethylation of Native Residues in Proteins. *Journal of the American Chemical Society* **2018,** *140* (5), 1568-1571.

25.   Shibata, N.; Matsnev, A.; Cahard, D., Shelf-stable electrophilic trifluoromethylating reagents: A brief historical perspective. *Beilstein journal of organic chemistry* **2010,** *6*.

26.   Xu, X. H.; Matsuzaki, K.; Shibata, N., Synthetic methods for compounds having CF3-S units on carbon by trifluoromethylation, trifluoromethylthiolation, triflylation, and related reactions. *Chemical reviews* **2015,** *115* (2), 731-64.

27.   Cheng, M.; Guo, C.; Gross, M. L., The Application of Fluorine-Containing Reagents in Structural Proteomics. *Angewandte Chemie (International ed. in English)* **2020,** *59* (15), 5880-5889.

28.   Chachignon, H.; Guyon, H.; Cahard, D., CF(3)SO(2)X (X = Na, Cl) as reagents for trifluoromethylation, trifluoromethylsulfenyl-, -sulfinyl- and -sulfonylation and chlorination. Part 2: Use of CF(3)SO(2)Cl. *Beilstein journal of organic chemistry* **2017,** *13*, 2800-2818.

29.   Cheng, M.; Asuru, A.; Kiselar, J.; Mathai, G.; Chance, M. R.; Gross, M. L., Fast Protein Footprinting by X-ray Mediated Radical Trifluoromethylation. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (5), 1019-1024.

30.   Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L., Laser-Initiated Radical Trifluoromethylation of Peptides and Proteins: Application to Mass-Spectrometry-Based Protein

Footprinting. *Angewandte Chemie (International ed. in English)* **2017,** *56* (45), 14007-14010.

31.   Guerrero, I.; Correa, A., Site-Selective Trifluoromethylation Reactions of Oligopeptides. *Asian Journal of Organic Chemistry* **2020,** *9* (6), 898-909.

32.   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. *Chemical science* **2018,** *9* (17), 4168-4175.

33.   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. *Proceedings of the National Academy of Sciences* **2011,** *108* (35), 14411-14415.

34.   Meng, H.; Kumar, K., Antimicrobial activity and protease stability of peptides containing fluorinated amino acids. *Journal of the American Chemical Society* **2007,** *129* (50), 15615-22.

35.   Asuru, A.; Farquhar, E. R.; Sullivan, M.; Abel, D.; Toomey, J.; Chance, M. R.; Bohon, J., The XFP (17-BM) beamline for X-ray footprinting at NSLS-II. *Journal of synchrotron radiation* **2019,** *26* (Pt 4), 1388-1399.

36.   Jain, R.; Abel, D.; Rakitin, M.; Sullivan, M.; Lodowski, D. T.; Chance, M. R.; Farquhar, E. R., New high-throughput endstation to accelerate the experimental optimization pipeline for synchrotron X-ray footprinting. *Journal of synchrotron radiation* **2021,** *28* (Pt 5), 1321-1332.

37.   Gupta, S.; Sullivan, M.; Toomey, J.; Kiselar, J.; Chance, M. R., The Beamline X28C of the Center for Synchrotron Biosciences: a national resource for biomolecular structure and dynamics experiments using synchrotron footprinting. *Journal of synchrotron radiation* **2007,** *14* (Pt 3), 233-43.

38.   Buszewski, B.; Noga, S., Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Analytical and Bioanalytical Chemistry* **2012,** *402* (1), 231-247.

39.   Bupp, C. R.; Wirth, M. J., Making Sharper Peaks for Reverse-Phase Liquid Chromatography of Proteins. *Annual review of analytical chemistry (Palo Alto, Calif.)* **2020,** *13* (1), 363-380.

40.   Pomerantz, W. C.; Wang, N.; Lipinski, A. K.; Wang, R.; Cierpicki, T.; Mapp, A. K., Profiling the dynamic interfaces of fluorinated transcription complexes for ligand discovery and characterization. *ACS chemical biology* **2012,** *7* (8), 1345-50.

41.   Zhu, W.; Wang, J.; Wang, S.; Gu, Z.; Aceña, J. L.; Izawa, K.; Liu, H.; Soloshonok, V. A., Recent advances in the trifluoromethylation methodology and new CF3-containing drugs. *Journal of Fluorine Chemistry* **2014,** *167*, 37-54.