Mass Spectrometry-based Structural Proteomics: Methodology and Application of Fast Photochemical Oxidation of Proteins (FPOP)

Ben Niu

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Biochemistry Commons, Bioinformatics Commons, and the Chemistry Commons

Recommended Citation


https://openscholarship.wustl.edu/art_sci_etds/1238

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Mass Spectrometry-based Structural Proteomics: Methodology and Application of Fast Photochemical Oxidation of Proteins (FPOP)

By

Ben Niu

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2017
St. Louis, Missouri
Table of Contents

List of Figures .......................................................................................................................... vi

List of Tables ........................................................................................................................... xii

Acknowledgements ................................................................................................................. xiii

Chapter 1 A Review of Mass Spectrometry-based Hydroxyl Radical Footprinting: Methodology and Application of Fast Photochemical Oxidation of Proteins (FPOP)........................................................................................................... 1

Introduction ............................................................................................................................... 2

Generation of Hydroxyl Radicals ............................................................................................. 6

Fast Photochemical Oxidation of Proteins (FPOP) ................................................................. 18

Applications of FPOP ............................................................................................................. 31

Conclusion ............................................................................................................................... 49

References ............................................................................................................................... 50

Chapter 2 Dosimetry Determines the Initial Hydroxyl Radical Concentration in Fast Photochemical Oxidation of Proteins (FPOP) ......................................................................................................................... 72

Abstract ................................................................................................................................. 73

Introduction ............................................................................................................................. 73

Experimental .......................................................................................................................... 75
Chapter 3 Incorporation of a Reporter Peptide in FPOP Compensates for Adventitious Scavengers and Permits Time-dependent Measurements

Results and Discussion

Conclusions

References

Chapter 4 Using R for Mass Spectrometry-based Proteomics Data Visualization and Analysis

Abstract

Introduction

Unified API (Application Programming Interface) by mzR

Custom-built Web-based Applications

Conclusion and Future Work

References
Chapter 5 Fast Photochemical Oxidation of Proteins (FPOP) Locates the Unfolding of ALS-inducing Mutant Variant G93A SOD1 Caused by Destabilization with Trifluoroethanol………………………………………………………………………………..131

Abstract………………………………………………………………………………………132

Introduction……………………………………………………………………………………132

Experimental Section………………………………………………………………………..138

Results and Discussion……………………………………………………………………...143

Conclusions…………………………………………………………………………………...162

References……………………………………………………………………………………164

Chapter 6 Mass Spectrometry and Protein Footprinting in Characterizing EGF Receptor Intracellular Domain (EGFR-ICD).…………………………………………………………..170

Abstract………………………………………………………………………………………171

Introduction…………………………………………………………………………………..171

Experimental…………………………………………………………………………………..175

Results and Discussion……………………………………………………………………...178

Conclusions…………………………………………………………………………………...200

Future Work…………………………………………………………………………………..201

References……………………………………………………………………………………203
<table>
<thead>
<tr>
<th>Chapter 7 Protein Footprinting Elucidates the Protein-Membrane Association of Nanodiscs-Incorporated KRAS4b Proteins</th>
<th>208</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>209</td>
</tr>
<tr>
<td>Introduction</td>
<td>209</td>
</tr>
<tr>
<td>Experimental and Methods</td>
<td>215</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>223</td>
</tr>
<tr>
<td>Conclusion</td>
<td>244</td>
</tr>
<tr>
<td>References</td>
<td>246</td>
</tr>
<tr>
<td>Chapter 8 Conclusions and Perspectives</td>
<td>252</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1-1: (A) Photo of National Synchrotron Light Source II (NSLS-II) in Brookhaven National Laboratory; (B) Schematic showing the NSLS-II beamline hutches .................................................................................................................................15

Figure 1-2: Coupling of synchrotron hydroxyl radical labeling and mass spectrometry for protein footprinting......................................................................................................................................................16

Figure 1-3: Schematic showing the apparatus for conducting an FPOP experiment ........................................................................................................................................................................20

Figure 1-4: Numerical simulations of HO• concentration versus time........................................26

Figure 1-5: Schematic of the custom-built flow system for carbene footprinting on an FPOP platform ......................................................................................................................................................................31

Figure 1-6: Crystal structure of the exEGFR (colored in grey) complexed with Adnectin (colored in red), PDB code: 3QWQ..............................................................................................................................................36

Figure 1-7: Mass spectra of intact, FPOP-labeled Aβ1-42 (+5 charge) as a function of incubation time.........................................................................................................................................................40

Figure 1-8: Schematic of a custom-build two laser pump/probe system on an FPOP platform ......................................................................................................................................................................46

Figure 1-9: (A) Schematic illustration of a Nanodisc with a transmembrane protein embedded; (B) Representation of Nanodiscs-incorporated, FPOP-labeled LH2 transmembrane protein (PDB 1NKZ)........................................................................................................48

Figure 2-1: Workflow of FPOP dosimetry experiment......................................................................76

Figure 2-2: Mass spectra of trimethylsilyl derivatives for dσ-Phe and dδ-Phe, molecular ions (M+) were not discernable, the peaks representing [M+-CH3] were found at m/z 294.1 and 299.2, respectively...............................................................................................................79

Figure 2-3: GC/MS detection of BSTFA-derivatized dσ-phenylalanine (m/z 294.1340) and dδ-phenylalanine (m/z 299.1654)...........................................................................................................................................80

Figure 2-4: Standard calibration curves of dσ-phenylalanine based on fragment ions of: (A) m/z 294.1340 and 299.1654; (B) m/z 91.0542 and 96.0856; (C) m/z 266.1391 and 271.1705......................................................................................................................81
Figure 2-5: Dose-response curves determined from fragment ions of $m/z$ 294.1340 and 299.1654 for Met, His and Gln..........................................................83

Figure 2-6: Dose response curves determined by using fragment ions of (A) $m/z$ 266.1391 and 271.1705, and (B) $m/z$ 91.0542 and 96.0856.................................84

Figure 2-7: Numerical simulation of HO• concentration vs. time. The simulation includes the HO• reaction with scavengers, the recombination of HO•, and the Haber-Weiss chain reaction..........................................................84

Figure 3-1: Numerical simulations of HO• concentration versus time. Lifetime of HO• (spanning from ~60 ns to more than 10 µs) is dependent on the concentration of histidine (from 0 to 40 mM).........................................................95

Figure 3-2: (A) Mass spectra from intact measurements of FPOP-labeled WT SOD1 (+16 charge) obtained by adjusting [His]; (B) Intact protein fraction modified vs. reporter fraction modified under various scavenging conditions.........................................................95

Figure 3-3: Extracted ion chromatograms (EICs) of leu-enkephalin, both unmodified and modified (+16 Da) peptides were extracted with MS1 tolerance of 15 ppm ...............95

Figure 3-4: Response curves of each peptide from WT and A4V SOD1 in an FPOP kinetics experiment performed by adjusting different histidine concentrations..........99

Figure 3-5: Residue level kinetic curves of WT and A4V SOD1..........................100

Figure 3-6: Crystal structure of thermostable pseudo WT SOD1, containing the C6A/C111S mutations..........................................................100

Figure 4-1: The structure of an R Shiny web-based application.........................111

Figure 4-2: Proteomics Toolkit application. Shown is the display of computed results, under the sub-tab of Computations, based on the amino acid sequence input of YGGFLHPERTYNNAASDKL..........................................................113

Figure 4-3: Proteomics Toolkit application under the sub-tab of MS Spectral Simulation. Shown is the interactive display of simulated isotopic pattern of amino acid sequence YGGFLHPERTYNNAASDKLM with a charge state of +3...............................114

Figure 4-4: Proteomics Toolkit application under the sub-tab of in silico Digestion.....114

Figure 4-5: Proteomics Toolkit application under the sub-tab of Targeted m/z......115

Figure 4-6: The screenshot of HDX Easy Plotting application...........................123
Figure 4-7: Typical workflow of data processing of HDX-MS experiments (yellow arrows) and automation facilitated by HDX Easy Plotting application..................................................123

Figure 4-8: An example of the HDX kinetic curves generated by using HDX Easy Plotting application..........................................................................................................................124

Figure 4-9: A screenshot of MS/MS Spectrum Analysis & Assignment application.....126

Figure 4-10: By providing a spectrum scan number, the module of Peaks Matching can display the corresponding spectral information including retention time, m/z value and charge state of precursor ion..........................................................................................................................126

Figure 4-11: An example of annotated MS/MS spectrum using MS/MS Spectrum Analysis & Assignment application.................................................................126

Figure 4-12: Workflow of prediction of important binding interfaces from a set of HDX kinetic curves..........................................................................................................................128

Figure 5-1: Crystal structure of human SOD1 (PDB code 1N18), a native dimer......137

Figure 5-2: Electron microscopy of in vivo SOD1 aggregates induced by TFE (A-C) and heat (D-F)..........................................................................................................................138

Figure 5-3: Native ESI spectra of G93A SOD1 with and without adding of TFE........145

Figure 5-4: Mass spectra of intact, FPOP-labeled G93A SOD1 (+15 charge state) with 0.2 mM histidine as scavenger.................................................................................................147

Figure 5-5: (A-E) Mass spectra of intact, TFE-untreated G93A SOD1 footprinted by FPOP at different labeling times as adjusted by varied histidine concentrations; (F) Time-dependent curve obtained by plotting the fraction unmodified of TFE-untreated G93A vs. reporter peptide fraction modified (black circles); the point representing TFE-treated G93A (blue triangle) is not on the curve.................................................................150

Figure 5-6: Kinetic curves of 10 peptides from TFE-untreated G93A SOD1 in a time-dependent FPOP measured in triplicates, performed by adjusting different histidine concentrations..........................................................................................................................151

Figure 5-7: (A) Averaged residue-specific b-factor values plotted versus the residue index. Two regions showing high b-factor values are the zinc binding loop (residues 50-84) and the electrostatic loop (residues 122-143); (B, C) Zoom-in views of the SOD1 crystal structure (3GZP) show that the side chains of Trp32, Lys122, Lys128, and Lys 136 point to solution..........................................................................................................................153
Figure 5-8: (A) Comparison of FPOP yields of TFE-treated and untreated G93A SOD1 at peptide level under the same labeling time, as guaranteed by the same reporter fraction modified; (B) FPOP results mapped onto the crystal structure of dimeric G93A SOD1 (PDB code 3GZP).…………………………………………………………………………………………………………………...154

Figure 5-9: Structure, extracted ion chromatograms, and FPOP quantitation of peptide 80-91..........................................................................................................................................................158

Figure 5-10: Product-ion spectra corresponding to +16 modified peptide 80-91 locating the FPOP modification on His80 (A) and Lys91 (B).………………………………………………………………………………………………………………………….159

Figure 5-11: Structure, extracted ion chromatograms, and FPOP quantitation of peptide 92-115..........................................................................................................................160

Figure 5-12: Secondary structure representation of SOD1, the eight anti-parallel β strands are shown as arrows with numbers representing the starting and ending residue index of each strand...........................................................................................................................................162

Figure 6-1: (A) Total ion chromatogram (TIC) of EGF receptor trypsin digests without sample cleanup, loaded amount 2.5 pmol; (B) Mass spectrum containing all scans from 25 to 40 min elution window in the TIC of EGF receptor without sample cleanup; (C) TIC of EGF receptor trypsin digests after acetone precipitation, loaded amount 2.5 pmol...........................................................................................................................................................................................................180

Figure 6-2: Representation of activated and dimerized EGF receptor.......................183

Figure 6-3: Product-ion (MS/MS) spectra of EGFR-ICD tryptic peptide GSTAENAEYLR, amino acid 1165-1175..........................................................................................................................184

Figure 6-4: Comparison of extracted ion chromatograms (XICs) of peptide 1165-1175 between unphosphorylated sample and 10-min phosphorylated sample.................184

Figure 6-5: Bar-graph showing the differences in phosphorylation extents for major identified residues for 0-min and 10-min reaction.........................................................185

Figure 6-6: (A-C) HDX result of EGFR-ICD binding to Cbl under three different incubation times (5s, 10s, and 30s)........................................................................................................189

Figure 6-7: (A) XICs of unmodified and +16 modified Leu-enkephalin when there is no buffer exchange, the +16 species is minor compared to the unmodified species; (B) Same as (A), but after buffer exchange prior to FPOP labeling; (C) Quantification of FPOP yields on reporter peptide shows dramatic increase from 3% to 28% with buffer exchange...........................................................................................................................................................191
Figure 6-8: Quantification of FPOP yields for EGFR-ICD tryptic peptides with and without buffer exchange..................................................................................................................191

Figure 6-9: Solvent accessibility output of EGFR-ICD measured by FPOP with the reactivity normalization with buffer exchange (A) and without buffer exchange (B) prior to labeling.................................................................................................................................194

Figure 6-10: Extracted ion chromatograms (XICs) of unmodified, +16, and +32 modified peptide 976-1007.................................................................................................................................198

Figure 6-11: Solvent accessibility output of EGFR-ICD measured by FPOP with normalization of reactivity of peptides containing no methionine residue.........................200

Figure 7-1: (A) Crystal structure of KRAS4b with HVR truncated (amino acid 1-166), PDB 4l8g; (B) Sequence alignment of amino acids in NRAS, HRAS, KRAS4a, and KRAS4b.........................................................................................................................214

Figure 7-2: Post-translationally lipidated cysteine residues in HVR of NRAS, HRAS, KRAS4a, and KRAS4b..........................................................................................................................215

Figure 7-3: (A) A schematic representation of a Nanodisc; (B) A cartoon representation of the process of nanodisc assembly.................................................................215

Figure 7-4: Fraction modified of Leu-enkephalin (reporter peptide) under different FPOP labeling times as adjusted by varying scavenger concentrations (histidine concentrations: 0.1, 1.0, 5.0, 20 mM)........................................................................................................225

Figure 7-5: FPOP labeling extent quantitation of Leu-enkephalin..............................................227

Figure 7-6: Mass spectra from measurements of intact, FPOP-labeled KRAS4b-FMe (+27 charge) obtained by adjusting histidine concentrations......................................................228

Figure 7-7: Response curves of each peptide from Nanodiscs-bound and unbound KRAS4b-FMe in FPOP time-dependent measurements performed by varying the histidine concentrations (0.1, 1.0, 5.0, and 20 mM)................................................................................................................232

Figure 7-8: MS/MS of peptide 20-40 illustrates the MS detection of unmodified and modified peptide species, the chromatographic separation of various modified species, and localization of the modified residue........................................................................................................233

Figure 7-9: Residue level kinetic curves of Nanodiscs-bound and unbound KRAS4b-FMe.................................................................................................................................234
Figure 7-10: Product-ion (MS/MS) spectra of various peptides used to locate the sites of FPOP modification............................................................................................................................................................................235

Figure 7-11: Annotated product-ion (MS/MS) spectra corresponding to the FPOP-modified peptide 171-185 for which the Cys185 is both farnesylated and methylated.............................................................................................................................................................................................................................................................................236

Figure 7-12: Time-dependent GEE labeling results of Nanodiscs-bound (dark grey) and unbound (light grey) KRAS4b-FMe presented as bar-graph..........................................................239

Figure 7-13: Response curves of each peptide from Nanodiscs-bound and unbound KRAS4b-FMe in a GEE kinetics measurements performed by controlling different labeling times (2, 6, and 10 min) ..........................................................................................................................................................................................................................................................................................................................240

Figure 7-14: Extracted ion chromatograms (XICs) of unmodified and GEE-labeled peptide 54-64..........................................................................................................................................................................................................................................................................................................................................................................................240

Figure 7-15: (A) Representation of Nanodiscs-bound KRAS4b-FMe; (B) Zoom-in view of HVR interacting with the Nanodiscs determined by FPOP and GEE labeling........241

Figure 7-16: (A) Schematic showing the tethered lipid bilayers prepared for neutron reflectivity experiments; (B) The geometry of a typical reflectivity experiment showing the incident and reflected angles, and the incident and reflected wave factors..........................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................243

Figure 7-17: (A) Electron structure of the Gd atom; (B) Schematic of Nanodiscs containing Gd-lipid chelate and $^{15}$N labeled KRAS4b-FMe..........................243
List of Tables

Table 1-1: Rate constants of amino acids towards hydroxyl radical and mass shifts for primary oxidative products..................................................................................................................6

Table 1-2: Rate constants of amino acids towards hydrated electrons..............................11

Table 4-1: The different file extensions of major vendors of LC-MS/MS systems........109

Table 6-1: The numbers of hydroxyl radical reactive residues in kinase domain and C-terminal tail domain of EGFR-ICD. The kinase domain contains a large number of FPOP-reactive residues................................................................................................................195
ACKNOWLEDGEMENTS

Thanks to my dissertation committee members, including Prof. Michael Gross, Prof. Jay Ponder, Prof. Dewey Holten, Prof. Robert Blankenship, and Prof. Gaya Amarasinghe, for their advice and expertise in guiding my research. I thank Mike particularly, who’s my advisor, but more like a friend, for his tremendous effort in maintaining a world-class, multi-disciplinary research facility, and for his instruction and support in broadening my horizons.

I would thank all my current and former colleagues, with whom I've had the good fortune to cross path, Henry Rohrs, Don Rempel, Weidong Cui, Daryl Giblin, Joyce Neff, Jun Zhang, Lisa Jones, Manolo Plasencia, Brian Gau, Hao Zhang, Ying Zhang, Yuetian Yan, Jing Li, Yining Huang, Elizabeth Bergman, Eva Illes-Toth, Ken Chanthamontri, Jagat Adhikari, Saketh Chemuru, Rachel Shi, Hanliu Wang, Yue Lu, Ke Li, Bojie Zhang, Ming Cheng, and those more junior students, for being part of this lovely group, and part of so many memorable moments that I have cherished. Special thanks to Don, for his endless bags of chocolate, jars of peanuts shared with me, and equally endless conversations he had with me, that included, but not limited to, scientific research. Thanks to Henry, for sharing his knowledge in mass spectrometry and chromatography, turning me successfully from a total novice to an apprentice, and for providing me the opportunities to accrue my required daily steps (and pocket money) while bringing his needed trenta iced green tea or whipped hot chocolate. Thanks to Ying, Yuetian, and Hao, for their generous support and encouragement when I needed cheering up, for
their tremendous help in my job-hunting, and for their long-term “collaboration” with me, to play Puzzle & Dragons.

I want to acknowledge my collaborators as well, including Dr. Andy Stephen and Xiaoying Ye from NCI, Dr. Kathy Brendza from Gilead, Dr. Jill Zitzewitz et al. from University of Massachusetts medical school, and Dr. Linda Pike and Kamaldeep Gill from Washington University medical school, for their advice and expertise, and fruitful, pleasant collaborations.

I thank all my friends, for sharing countless laughter and tears, teaching me life’s lessons, guiding me through, and backing me up. Those memories are amazing and will be with me forever. I thank pastor Bob Lange, for organizing the Saturday afternoon bible discussion session, five years in a row. It has taught me something bright and essential, including being myself, treating others, and loving people. I thank Sammi, my beloved kitten, who’s right now crouching onto my knees while I’m writing this. To me, every one of you is a lifelong treasure with whom God has generously endowed me.

A special shout-out to my fiancée, Susan Ma. I cannot imagine how anybody else could possibly put up with me like you did, nor can I imagine my life without you. Thank you for entering my life. Sometimes words pale and are insufficient to express feelings. Heartfelt thanks also to my mom and dad, for their unconditional support, caring, and love for me.

I also acknowledge the funding of NIH NIGMS 5P41GM103422, and NIH NIGMS 1S10OD016298 that supported the research associated with this thesis.
Chapter 1

A Review of Mass Spectrometry-based Hydroxyl Radical Footprinting: Methodology and Application of Fast Photochemical Oxidation of Proteins (FPOP)
1.1 Introduction

The conformations of proteins when alone or complexed with other molecules (e.g., metal ion, small peptide, protein) are an essential basis for understanding many biological processes and developing protein therapeutics. Traditional approaches for assessing protein conformations include X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, both of which provide high resolution structural information. Developments in cryo-electron microscopy (cryo-EM), more recently, provides unprecedented opportunities for high-resolution characterization of protein conformations. Some “global” information can also be obtained by low resolution methods (e.g., fluorescence, infrared, UV-Vis, circular dichroism, and differential scanning calorimetry). Recent advances in mass spectrometry (MS)-based approaches have paved a new avenue to elucidate protein structures and interactions, providing coarse-grained protein conformations by fast-turnaround, high sensitivity, reproducible methods that require relatively small sample sizes. Protein footprinting coupled to mass spectrometry is one of the most developed of these MS-based strategies.

1.1.1 General Approaches for Mapping Protein Conformations

Although there are several approaches affording low resolution protein structural information [1-5], high order structures (HOS) of proteins and their interactions require more detailed characterizations with higher resolution. Techniques including X-ray crystallography [6], NMR [7], and cryo-electron microscopy (cryo-EM) [8] provide structural information of macromolecules at atomic or near-atomic resolution. Mass spectrometry-based techniques, in contrast, offer intermediate-level resolution [9].
Although they are more recently developed, ultimately they may provide constraints for high resolution structure.

X-ray crystallography generates a static view of three-dimensional electron density maps of protein structures [10]. It is currently the favored techniques extensively used in determining protein structures and protein-protein interactions, since its debut in late 1950s when it was first used to reveal the 3-D structure of myoglobin [11]. The use of X-ray crystallography is, nevertheless, impeded by a large amount of highly homogeneous protein sample [12, 13], and slow turnaround [14]. Further, proteins with high flexibility, low stability, or are intrinsic membrane proteins are difficult to crystalize [15].

NMR spectroscopy complements X-ray crystallography by allowing measurements of three-dimensional protein structures in solution [16]. Because NMR is often applied to proteins in solution, it is a robust tool for investigating protein folding/unfolding [17], fast dynamic motions [18], and conformational equilibria [19]. Similar to X-ray crystallography, NMR demands protein samples at high amounts [20, 21]. There are problems, however, in resolving overlapping peaks of large proteins, restricting applications of NMR to relatively small proteins [22, 23].

1.1.2 Mass Spectrometry-based Approach

Mass spectrometry was restricted only to small and thermostable molecules, owing to a lack of gentle ionization and ion transportation [9]. The marriage of electrospray ionization (ESI) and mass spectrometry [24, 25], however, has greatly advanced mass spectrometry-based approaches as powerful tools in assessing larger molecules including peptides and proteins. This new capability was recognized by the 2002 Nobel
prize in chemistry. Today, mass spectrometry has increasingly become the method of choice for complicated proteomic analyses, owing to the multi-disciplinary improvements including instrumentation [26-28], sample preparation [29-32], and informatics [33-36]. Not only does mass spectrometry access protein structural information from a global level (whole protein), but it can screen peptides, upon proteolysis, with high sensitivity (fmol level) and throughput when coupled to liquid chromatography [37]. With fragmentation methods including collision-induced dissociation (CID) and electron-capture/transfer dissociation (ECD and ETD), sequence information at residue level from the product-ion spectra can be obtained [38-40]. This sequence information is useful for establishing peptide identity and locating post-translational modifications (PTMs) [41].

Mass spectrometry-based protein footprinting has evolved as an effective analytical technique for characterizing protein HOS, albeit with lower resolution than X-ray crystallography and NMR spectroscopy [42, 43]. MS typically examines conformational changes of proteins, by monitoring solvent accessibility of either backbones or sidechains, based on their sensitivity to chemical labeling [44]. Such experiments are best utilized in a comparative analysis where the extent of labeling is indicative of the changes of solvent accessibility of two or more protein states (e.g., bound and unbound).

Protein footprinting techniques can be classified based on the nature of modifications (e.g., reversible labeling and irreversible labeling). Numerous footprinting techniques have emerged over the past few decades [45-51], among which hydrogen deuterium exchange (HDX) is probably the most widely used and extensively developed [46, 52-56]. HDX probes the solvent accessibility and hydrogen bonding environments of amide
backbone hydrogens [57, 58]. Differential deuterium uptake of a protein, upon folding, ligand-binding, or oligomerization, can be monitored by MS measurements at both the global and peptide levels. It is noteworthy, however, that by taking advantage of the sequentially overlapping peptide fragments from pepsin digestion, resolution level of individual amino acid can also be achieved in HDX [59, 60].

Hydroxyl radical (HO•) footprinting, which covalently labels protein sidechains, represents the second class of protein footprinting techniques. It is emerging for studying protein structure and dynamics in part because hydroxyl radical is readily formed and highly reactive [61]. The size of a HO• closely approximates that of a water molecule, enabling its reactivity to be a faithful map of protein solvent accessibility. Advantageously, the irreversible nature of HO• modifications permits more rigorous sample handling procedures downstream of the labeling process and prior to MS detection, including sample cleanups, proteolysis, and chromatography. Hydroxyl radicals label protein sidechains with a residue-specific mechanism, which leads to various products, and rate constants [62]. Nevertheless, the major modification products are the +16, +32, and +48 species, etc., corresponding to the labeled protein with one or more substitutions of H for OH [63]. Shown in Table 1-1 are the rate constants of various amino-acid residues towards HO• and the mass shifts associated with primary oxidative modifications observed in HO• footprinting.
<table>
<thead>
<tr>
<th>Residue</th>
<th>Rate Constant (M⁻¹s⁻¹)</th>
<th>Mass shifts (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys</td>
<td>3.5 × 10⁻⁶</td>
<td>+31.9898</td>
</tr>
<tr>
<td></td>
<td>-15.9772</td>
<td>+47.9847</td>
</tr>
<tr>
<td>Trp</td>
<td>1.3 × 10⁻⁶</td>
<td>+15.9949</td>
</tr>
<tr>
<td></td>
<td>+3.9949</td>
<td>+31.9898</td>
</tr>
<tr>
<td>Tyr</td>
<td>1.3 × 10⁻⁶</td>
<td>+31.9898</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td>+47.9847</td>
</tr>
<tr>
<td>Met</td>
<td>8.5 × 10⁻⁹</td>
<td>+15.9949</td>
</tr>
<tr>
<td></td>
<td>-32.0085</td>
<td>+31.9898</td>
</tr>
<tr>
<td>Phe</td>
<td>6.9 × 10⁻⁹</td>
<td>+15.9949</td>
</tr>
<tr>
<td></td>
<td>+31.9898</td>
<td>+47.9847</td>
</tr>
<tr>
<td>His</td>
<td>4.8 × 10⁻⁹</td>
<td>-23.0160</td>
</tr>
<tr>
<td></td>
<td>-22.0320</td>
<td>-10.0320</td>
</tr>
<tr>
<td></td>
<td>+4.9789</td>
<td>+15.9949</td>
</tr>
<tr>
<td>Arg</td>
<td>3.5 × 10⁻⁹</td>
<td>-43.0534</td>
</tr>
<tr>
<td></td>
<td>+13.9793</td>
<td>+15.9949</td>
</tr>
<tr>
<td>Cys</td>
<td>2.1 × 10⁻⁹</td>
<td>+31.9898</td>
</tr>
<tr>
<td></td>
<td>+47.9847</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>1.8 × 10⁻⁹</td>
<td>+13.9793</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>1.7 × 10⁻⁹</td>
<td>+13.9793</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>8.5 × 10⁻⁸</td>
<td>+13.9793</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>6.5 × 10⁻⁸</td>
<td>+13.9793</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>5.4 × 10⁻⁸</td>
<td>+13.9793</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>5.1 × 10⁻⁸</td>
<td>-2.0157</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>3.5 × 10⁻⁸</td>
<td>+13.9793</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>3.2 × 10⁻⁸</td>
<td>-2.0157</td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>2.3 × 10⁻⁸</td>
<td>-30.0106</td>
</tr>
<tr>
<td></td>
<td>+13.9793</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+15.9949</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>7.7 × 10⁻⁷</td>
<td>+15.9949</td>
</tr>
<tr>
<td>Asp</td>
<td>7.5 × 10⁻⁷</td>
<td>-30.0106</td>
</tr>
<tr>
<td>Asn</td>
<td>4.9 × 10⁻⁷</td>
<td>+15.9949</td>
</tr>
<tr>
<td>Gly</td>
<td>1.7 × 10⁻⁷</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1-1. Rate constants of amino acids towards hydroxyl radical and mass shifts for primary oxidative products [44, 64].

1.2 Generation of Hydroxyl Radicals

1.2.1 Fenton and Fenton-like Chemistry

In 1876, H. J. H. Fenton observed a violet-colored product obtained by mixing tartaric acid with hydrogen peroxide and a low concentration of ferrous salt [65]. This experiment is known as “Fenton chemistry” today and is one of the earliest approaches to generate hydroxyl radicals. In Fenton chemistry, tartaric acid is oxidized to dihydroxymaleic acid by hydrogen peroxide under the catalysis of ferrous ion(II), and it is the complex formed between iron and dihydroxymaleic acid that’s responsible for the “violet color” in Fenton’s
observation. Fenton established the molecular formula of this complex [65, 66], but did not propose any reaction schemes involving hydroxyl radicals. The notion of free radicals in solution was not generally accepted until 1930s, when Michaelis et al. [67-69] characterized radicals from one-electron reduction of quinones, flavins and bipyridinium compounds. The reaction described in Eq. 1 is typically referred to as the “Fenton reaction”, which shows the oxidation (abstraction of H from tartaric acid) initiated by HO• generated in the reaction between Fe(II) and H₂O₂. The Fenton reaction has been mechanistically studied later by Haber, Weiss, and Willstätter [70, 71].

\[
\text{Fe(II) + H}_2\text{O}_2 \rightarrow \text{Fe(III)} + \text{HO}^\cdot + \text{OH}^- \quad (1)
\]

Besides Fe(II), many other transition metal ions in lower oxidation states (e.g., Ti(III), Cr(II), Co(II), Cu(I)) are capable of “Fenton-like” chemistry, as generalized in Eq. 2. This oxidation can occur naturally in the microenvironment of cells and is associated with aging and various diseases [72-74].

\[
\text{M}^{(n)} + \text{H}_2\text{O}_2 \rightarrow \text{M}^{(n+1)} + \text{HO}^\cdot + \text{OH}^- \quad (2)
\]

Based on Eq. 2, one might speculate that HO• can be generated for oxidative footprinting by simply mixing hydrogen peroxide with the salt of transition metals in their lower oxidation states, preferably under acid conditions [75]. The secondary order rate of these reactions, however, is approximately 60 M⁻¹s⁻¹[70], which is practically too slow for any oxidative footprinting experiments. To address this issue, Tullius and Dombroski [76, 77] developed a Fenton system that incorporates Fe(II) chelation, by EDTA for example, that exhibits a rate increases of approximately two orders of magnitude [78]. Under
appropriate conditions, other transition metals with proper chelates could also undergo this Fenton-like chemistry at increased rates [79-81].

Various Fenton approaches were adopted to produce HO• for biological footprinting studies in nucleic acid/ligand interactions [82-84]. In 2003, Sharp et al. [85] introduced protein footprinting with a Fenton reagent. A 5-min-oxidation was required to obtain sufficient oxidative labeling for mass spectrometry analysis. Fenton chemistry can also probe protein-metal interactions whereby the protein-bound metal can reduce H₂O₂ to HO• for footprinting nearby [86-89]. These approaches, however, suffer from slow rate in generating radicals and lack of control in the reaction time. Brenowitz et al. [90] developed a faster Fenton footprinting method that incorporates a stop-flow apparatus to initiate and quench the Fenton reaction with 44 mM H₂O₂ and 5 mM Fe²⁺-EDTA, on the ms timescale. The high concentrations of metal salts and EDTA required for such reactions, however, raise concerns about maintaining protein structural integrity during the reaction [91, 92]. Moreover, the interactions between metal ions and biological samples, or between chelates and biological samples, can bias the reactivity [93, 94].

1.2.2 Electron Pulse Radiolysis

Electron pulse radiolysis of water is an ionizing process that generates hydroxyl radicals (and many other radicals) by accelerating electrons to the MeV range [95-98]. The first step of this process is the ionization of water, producing a water radical cation and an electron (Eq 3).

$$\text{H}_2\text{O} \rightarrow \text{H}_2\text{O}^* + \text{e}^- \quad (3)$$
The products in Eq 3 will then quickly undergo the following reactions (on the picosecond time scale) and turn to primary radicals that are highly reactive. The electron may recombine with a water molecule and form an excited water molecule.

\[ \text{H}_2\text{O}^{++} + \text{H}_2\text{O} \rightarrow \text{HO}^+ + \text{H}_3\text{O}^+ \] (4)

\[ \text{e}^- + \text{H}_2\text{O} \rightarrow \text{e}_{\text{aq}}^- \] (5)

Given the short time scale \((10^{-12} \text{ s})\) and the typical width of each pulse (ranging from 1 to 100 ns), the radiation window should be mostly filled with hydroxyl radicals \((\text{HO}^+)\) and hydrated electrons \((\text{e}_{\text{aq}}^-)\) after each pulse. By definition, \(\text{e}_{\text{aq}}^-\) is free radical because it is an unpaired electron. These two radicals recombine at diffusion-controlled rate under the condition of pure water [64, 95]:

\[ \text{HO}^+ + \text{e}_{\text{aq}}^- \rightarrow \text{OH}^- \] (6)

\[ \text{HO}^+ + \text{HO}^+ \rightarrow \text{H}_2\text{O}_2 \] (7)

Besides hydroxyl radicals and hydrated electrons, a small amount of hydrogen radical may also be formed during radiolysis, owing to deactivation of an excited water molecule [95]. In practice, aqueous solutions are saturated with dissolved oxygen \((\sim 0.3 \text{ mM})\), which itself is a reactive di-radical. The hydrogen radicals and hydrated electrons can react with dissolved oxygen, generating hydroperoxyl radicals and superoxide radicals, respectively, with rate constants close to the diffusion-rate limit [97].

\[ \text{H}^+ + \text{O}_2 \rightarrow \text{HO}_2^+ \] (8)

\[ \text{e}_{\text{aq}}^- + \text{O}_2 \rightarrow \text{O}_2^{2-} \] (9)
Therefore, in most oxygenated biological systems, the primary free radical is superoxide radical, which is in equilibrium with its protonated form—hydroperoxyl radical [99]. Both radicals are not as reactive towards amino acids, as is hydroxyl radical [44, 100, 101]. The relative amount of these two is determined by its acid-base equilibrium equation with a pK\textsubscript{a} of 4.8 [102].

\[
\text{HO}_2^- \leftrightarrow H^+ + O_2^* \tag{10}
\]

As shown in Table 1-2, except for cysteine and cystine, hydrated electrons (\(e_{aq}^-\)) react with most amino acids far slower than reacting with oxygen (Eq 9). Alanine is the major product of hydrated electrons reacting with cysteine; however, in the presence of oxygen, alanine hydroperoxide also forms [103, 104].

The co-presence of hydrated electrons and hydroxyl radicals in pulse radiolysis of water can be problematic for protein footprinting because each radical undergoes multiple reactions, which could either potentially complicate the reaction pathways, or interfere with the measurements of free-radical products. To address this issue, hydrated electrons are converted to hydroxyl radicals in the presence of gaseous N\textsubscript{2}O per Eq 11:

\[
e_{aq}^- + N_2O \rightarrow HO^* + OH^- + N_2 \tag{11}
\]

The rate constant of this conversion is \(9 \times 10^9 \text{ M}^{-1}\text{s}^{-1}\) [95], rendering a solution of hydroxyl radicals on the ns timescale. The rapid generation of hydroxyl radical is overwhelmingly advantageous for protein footprinting to insure there is minimal perturbation of the protein structure.
<table>
<thead>
<tr>
<th>Residue</th>
<th>Rate Constant $(M^{-1}s^{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cystine</td>
<td>$1.5 \times 10^{16}$</td>
</tr>
<tr>
<td>Cys</td>
<td>$1.0 \times 10^{16}$</td>
</tr>
<tr>
<td>Gly</td>
<td>$8.0 \times 10^{8}$</td>
</tr>
<tr>
<td>Trp</td>
<td>$3.0 \times 10^{8}$</td>
</tr>
<tr>
<td>Tyr</td>
<td>$2.8 \times 10^{8}$</td>
</tr>
<tr>
<td>Arg</td>
<td>$1.5 \times 10^{8}$</td>
</tr>
<tr>
<td>Asn</td>
<td>$1.5 \times 10^{8}$</td>
</tr>
<tr>
<td>His</td>
<td>$6.0 \times 10^{7}$</td>
</tr>
<tr>
<td>Met</td>
<td>$4.5 \times 10^{7}$</td>
</tr>
<tr>
<td>Ser</td>
<td>$&lt; 3.0 \times 10^{7}$</td>
</tr>
<tr>
<td>Thr</td>
<td>$2.0 \times 10^{7}$</td>
</tr>
<tr>
<td>Lys</td>
<td>$2.0 \times 10^{7}$</td>
</tr>
<tr>
<td>Pro</td>
<td>$2.0 \times 10^{7}$</td>
</tr>
<tr>
<td>Asp</td>
<td>$1.8 \times 10^{7}$</td>
</tr>
<tr>
<td>Phe</td>
<td>$1.6 \times 10^{7}$</td>
</tr>
<tr>
<td>Glu</td>
<td>$1.2 \times 10^{7}$</td>
</tr>
<tr>
<td>Ala</td>
<td>$1.2 \times 10^{7}$</td>
</tr>
<tr>
<td>Leu</td>
<td>$&lt; 1.0 \times 10^{7}$</td>
</tr>
<tr>
<td>Val</td>
<td>$&lt; 5.0 \times 10^{6}$</td>
</tr>
<tr>
<td>Ile</td>
<td>–</td>
</tr>
<tr>
<td>Gln</td>
<td>–</td>
</tr>
</tbody>
</table>

**Table 1-2.** Rate constants of amino acids towards hydrated electrons [44].

### 1.2.3 High Voltage Electrical Discharge

Downard and coworkers [105, 106] developed a source of hydroxyl radicals by using high voltage electrical discharge in a mass spectrometer ion source and used it to footprint proteins. In the discharge, the protein solution (µM concentration) is introduced to conventional atmospheric pressure electrospray ionization (ESI) source of with typical flow rates between 1-5 µL/min. High electrical voltage ranging from 4 to 8 kV is applied to electrospray emitter so the discharge occurs, generating HO• in solution [107]. Oxygen is typically used as the nebulizer gas at a regular pressure of 250-350 kPa, corresponding to an approximate flow rate of 10 L/min [108]. After oxidation, the droplets are introduced
into the mass spectrometer or are collected by condensation for subsequent enzymatic digestion and mass spectrometry analysis.

Using oxygen as nebulizer gas greatly enhances the yield of oxygen-containing reactive species, including the hydroxyl radical and the hydroperoxyl radical [106]. The production rates of these radicals under various conditions are known [109, 110]. By adjusting the discharge conditions (e.g., voltage of emitter, flow rate of protein solution, nature of nebulizer gas), one can control the production of radicals, although a systematic dose-response study has yet been done.

Although the high-voltage electrical discharge method of protein footprinting can probe protein structures and protein-protein interactions [108, 111, 112], there are caveats. Protein structure is determined by not only its primary sequence but also by other factors including the hydrogen-bonding network [113, 114], disulfide-bonding pattern [115-117], external ligand binding condition [118-120], and the solution environment the protein stays [119, 121-125]. Our view is that protein footprinting should be performed in a near-native environment where the labeling can faithfully report the protein solvent accessibility. The effect of high voltage on proteins, however, raise concerns regarding the protein structural integrity and conformational equilibrium. For example, numerous voltage-sensitive proteins can undergo conformational shifts within a voltage increase of 1 kV [126-128]. It’s also clear that, the electrospray process unfolds some proteins while others stay properly folded [129-131], indicating that unfolding depends on the protein and on the experimental conditions as many proteins analyzed by electrospray appear to maintain their structural integrity under carefully controlled conditions [132]. The fact that
large protein complexes can be transferred intact into the gas phase indicates some level of preservation of solution conformation [133, 134]. Moreover, ion mobility MS has demonstrated the conformational integrity of proteins footprinted by high voltage electrical discharge [112]. It’s likely, however, that some proteins do not maintain their structures during electrospray ionization, and one cannot yet predict the effects of ESI on protein in advance. We prefer a method that’s free from these uncertainties.

**1.2.4 Synchrotron X-ray Radiolysis of Water**

One of the most effective ways of generating hydroxyl radicals, pioneered by Chance and coworkers [44, 62, 135, 136], is the radiolysis of water by high-energy X-ray synchrotron beamline that produces approximately $10^{14}$ to $10^{15}$ photons per second with energies ranging from 3 to 30 keV. During this process, the energy of an incoming photon is transferred to ionize water molecules or other targets. The electrons are thermalized and deposit their energy to ionize other water molecules. As shown in 2.2 Eq 4, an ionized water molecule reacting with another water molecule generates the hydroxyl radical for the subsequent oxidative reactions in footprinting to occur. Similar to electron-pulse radiolysis, the oxidation reactions by synchrotron radiolysis of water involves multiple radical species, including hydroxyl radical, hydroperoxyl radical, and superoxide radical; among which hydroxyl radical is the most reactive and predominant species. Although the photon flux from the synchrotron is continuous, the exposure time and dose is tunable via an X-ray electronic shutter that controls the exposure time yielding a steady-state, micromolar dose of HO• within ms [137] [138].
Synchrotron X-ray radiolysis of water has been extensively used for footprinting, owing to its high flux density, short labeling timescale, high reproducibility and ease of varying the dose [139-142]. Chance et al. [143, 144] first used synchrotron HO• footprinting to probe the solvent exposure of oligonucleotides and their protein interactions (e.g., with RNA polymerase) in real time. The radical cleaves the phosphodiester backbone of nucleotides by abstracting hydrogen from the C-4 position of ribose, giving a free tertiary hydroxyl that leads to backbone cleavage [145]. Gel electrophoresis can capture the radical-cleaved nucleotides, the gel strands of which can be used to distinguish the protected and unprotected regions on nucleic acids, allowing the mapping of interaction interfaces.

The coupling of synchrotron radiolysis and mass spectrometry has expanded the synchrotron footprinting to more structural proteomics studies with improved robustness and sensitivity (Figure 1-2). This method has been extensively used to monitor protein folding and unfolding [143, 146, 147], to measure the protein-protein interaction interfaces and to determine the solvent-accessible regions and side chains [148, 149]. Regions involved in interactions typically show decreased reactivity. Remarkably, this approach can locate water molecules within a transmembrane GPCR protein, rhodopsin [150], providing opportunities to both in vivo and in vitro studies of membrane proteins.

Nevertheless, the synchrotron footprinting method is limited by its lack of accessibility to the ever-increasing number of users and the high cost of operation [90].
Figure 1-1. (A) Photo of National Synchrotron Light Source II (NSLS-II) in Brookhaven National Laboratory; (B) Schematic showing the NSLS-II beamline hutch. Retrieved from https://www.bnl.gov/ps/.
**Figure 1-2.** Coupling of synchrotron hydroxyl radical labeling and mass spectrometry for protein footprinting. (A) Hydroxyl radical generation by X-ray radiolysis and labeling on proteins; (B) Subsequent proteolysis and using of liquid chromatography for peptides separation; (C) Mass spectrometry detection and identification of HO• modified and unmodified peptide; (D) Quantitation of modification extent using extracted ion chromatograms (XICs); (E) Kinetic curve plotted as the fraction unmodified of each peptide versus the exposure time. Reprinted with permission from Ref. [150]

### 1.2.5 Photolysis of Hydrogen Peroxide

A more convenient way of generating hydroxyl radicals is through the photolysis of hydrogen peroxide. The light absorption by hydrogen peroxide (H$_2$O$_2$) reaches maximum at wavelength of 253.7 nm, leading to homolytic dissociation of hydrogen peroxide that generates two hydroxyl radicals per molecule with primary quantum yield of 0.4-0.5 [151, 152]. Different from radiolysis approaches where energy is deposited in the solvent water, the photolysis relies on absorption of photons by hydrogen peroxide per Eq 12:

$$H_2O_2 \xrightarrow{hv} 2HO^*$$  \hspace{1cm} (12)

The generated hydroxyl radicals undergo Haber-Weiss reactions as shown in Eq. 2 and 3, at rate constants of $2.7 \times 10^7$ and $7 \times 10^9$ M$^{-1}$s$^{-1}$, respectively [153, 154] or undergo the diffusion-controlled self-quenching reaction (Eq 13) with a rate constant of $4.7 \times 10^9$ M$^{-1}$s$^{-1}$ at room temperature [155, 156].

$$2HO^* \rightarrow H_2O_2$$  \hspace{1cm} (13)

The UV light-induced photolysis of hydrogen peroxide was first used for nucleic acids footprinting studies [157, 158]. Sharp and Hettich [159] demonstrated that this method can be extended to protein footprinting studies. In their method, apomyoglobin with 5 M
H$_2$O$_2$ in an Eppendorf tube was subjected to a 254-nm UV-light irradiation for 5 min. The high concentration of H$_2$O$_2$, however, can trigger direct protein oxidation [160]; and the long exposure time may compromise the integrity of protein structures, owing to radical-induced protein degradation or conformational changes [161].

To overcome this, two pulsed-laser approaches were developed in 2005 independently. Aye et al. [162] developed an UV laser-induced photodissociation method for protein footprinting, where a pulsed Nd:YAG laser (operating at 266 nm and 30 Hz) with an output of 2 mJ/pulse was used to photolize H$_2$O$_2$ at final concentration of 0.3-1.0% to generate hydroxyl radicals. The use of the UV laser enhances the yield of photolysis, allowing a significant decrease of H$_2$O$_2$ concentration (from 5 M to 100 mM). The protein solution, however, was still held in an Eppendorf tube and subjected to 1-100 laser shots, resembling Sharp’s method [159]. The oxidation of protein was quenched by freezing in liquid nitrogen and lyophilizing in a vacuum chamber at 10$^{-3}$ Torr for 30 min. The results show that, compared to non-oxidized ubiquitin, the singly oxidized ubiquitin was more susceptible to further oxidations by subsequent laser shots, owing to a conformational change of ubiquitin induced by oxidation. This experiment underscores the importance of using single laser shot and rapid generation of hydroxyl radicals to avoid secondary protein oxidations that can give unreliable footprinting [160, 161, 163].

Hambly and Gross [164], in the same year, developed a similar methodology that was later elaborated as “fast photochemical oxidation of proteins (FPOP)”. FPOP highlights the generation of hydroxyl radicals through homolytic cleavage of H$_2$O$_2$ in a flow cell, permitting footprinting of protein to take place in a “microenvironment” or a “plug”, and
allows multiple laser shots for sequential plugs of protein solution. Such design effectively reduces the hydrogen peroxide concentration in solution even further (15-20 mM H₂O₂). More importantly, at any given laser frequency, nearly single irradiation of fresh protein solution can be achieved by adjusting flow rates.

Both the latter two methods involve pulsed, nanosecond fluxes of UV laser to generate rapidly HO•; there are some experimental differences, however, between them. Aye’s method requires 100 mM of H₂O₂ to guarantee sufficient yield of HO• with single laser pulse; those extra HO• are solely quenched via self-recombination [155, 156, 165]. The method by Hambly and Gross uses lower concentrations of H₂O₂ and its design permits even further minimizing of the exposure by mixing the H₂O₂ just prior to irradiation region [166]. Importantly, it also introduces the idea of a radical scavenger to limit the lifetime of primary radical to microsec [164]. Further, the adjustment of scavenger identity and/or concentration makes possible the tuning of reaction times [165, 167].

### 1.3 Fast Photochemical Oxidation of Proteins (FPOP)

FPOP utilizes a 248 nm KrF excimer laser (GAM Laser Inc., Orlando, FL, USA) to photolyze H₂O₂ (15 mM) with an energy of approximately 45 mJ per pulse. The UV spectrum of H₂O₂ shows that its UV absorption wavelength spans from 300 nm to the beginning of the Rydberg transition region at 124 nm, with the absorption maximum around 250 nm [151, 152]. Thus, the 248 nm KrF excimer laser is an ideal light source for FPOP. We recently acquired an optical parametric oscillation (OPO) Nd:YAG laser, the fourth harmonic of which can yield UV light with wavelength of 266 nm. To conduct FPOP experiment, typically 50 µL of solution with protein concentrations ranging from 2
to 10 μM in PBS buffer (10 mM, pH 7.0) is submitted to approximately 1000 laser shots (Figure 1-3 A). The laser beam is focused with a convex lens (Edmunds Optics, Barrington, NJ, USA) onto a 150 μm i.d. fused silica tubing (Polymicro Technologies, Phoenix, AZ, USA), giving a 2.5-3.0 mm irradiation window (Figure 1-3 B). The pulse frequency was externally controlled by a pulse generator (B&K Precision, Yorbal Linda, CA, USA). The irradiation window width, pulse frequency, and sample flow rate are adjusted prior to irradiation to insure a fraction of 20-25% of protein solution unexposed to light, avoiding repeated exposure of radicals (Figure 1-3 B) [164, 168]. To limit the labeling time (i.e., lifetime of primary radicals), a radical scavenger is added to the protein solution. Typically, we chose a simple amino acid (e.g., histidine, glutamine, phenylalanine) that have a variety of reactivities towards HO• as radical scavenger; different identity and concentration of scavengers can be adjusted to control the FPOP labeling time, ranging from hundreds of ns to a few μs [165, 167, 169]. The labeled samples are collected in an Eppendorf containing 50 nM catalase and 70 mM methionine, preventing further oxidation by residual H₂O₂ and other remaining reactive oxygen species (ROS) (Figure 1-3 A). We always incorporate a positive control that is handled in the same manner without laser irradiation to keep track of background oxidations. Post footprinting, the sample is usually subjected to sample cleanup, proteolysis, chromatography, and mass spectrometry analysis.
Figure 1-3. Schematic showing the apparatus for conducting an FPOP experiment. (A) The protein solution with H$_2$O$_2$ and radical scavenger in syringe was advanced by a syringe pump, irradiated by pulsed KrF laser, and collected in Eppendorf containing methionine and catalase; (B) Zoom-in view of the irradiation window on silica tubing where the laser passes through. Each plug of irradiated solution is ideally sandwiched by non-irradiated plugs to guarantee an exclusion volume fraction of 20-25%.

1.3.1 FPOP Footprints Faster than Protein Folding/Unfolding
The rates of protein folding/unfolding are protein-specific and depend on protein size and structure. Protein folding into its native state is a complicated process that takes multiple intermediate steps; each of these steps occurs on a different timescale [170-172]. For instance, some proteins form secondary structures in an early step, which is followed by the formation and consolidation of tertiary structures as the late step. The formation of the secondary structures takes ns to μs; that of the tertiary structures, however, spans from tens of μs and even longer [173, 174]. In a fast relaxation kinetics measurement, Vu et al. [175] observed a two-step folding process for a protein, B-domain of staphylococcal protein A (BdpA), in which the first step is dominated by the fast formation of individual helices (90 ns), followed by the second step representing the packing of the helices together with squeezing out of water molecules to form the three-helix bundle (9 μs). For some proteins, the folding steps include hydrophobic collapse (μs) and tertiary structure formation (tens of μs and longer) [176]. Protein unfolding is likely the reverse of protein folding, including the breakdown of the tertiary structures and the admission of water molecules [177, 178]. The timescale of this process can be tens of μs or even longer.

To corroborate experimentally the claim that FPOP labels a protein faster than its unfolding, Gau et al. [179] applied FPOP to three oxidation-sensitive proteins (i.e., β-lactoglobulin, apo-calmodulin, and lysozyme), to test whether the population distribution of 0, +16, +32…products constitutes a Poisson distribution. The hypothesis is that if a single conformation of the protein exists during the labeling, the distribution of products should constitute a Poisson distribution. The fit to a Poisson distribution was poorer in the absence of the radical scavenger or if residual H₂O₂ was not removed after laser
irradiation; pointing to sampling of multiple protein conformations induced by repeated oxidation. The results strongly suggest that the Poisson model is applicable to protein footprinting when the protein population has an invariant conformation and undergoes non-cooperative modification, verifying that footprinting is faster than protein conformational change.

1.3.2 FPOP Dosimetry

Although the timescale of FPOP is sufficiently short to sample a single conformational state of the protein, as shown by Gau et al. [179], this approach doesn’t provide a quantitative description of the radical concentration or lifetime. To assess the radical lifetime, knowledge of the initial concentration of hydroxyl radical produced by laser photolysis is a prerequisite to adjust the yield of HO•, to control the extent of modifications, and to provide a basis for simulations and kinetic measurements [165].

A dosimetry experiment that measures the dose delivered to samples can be exploited for hydroxyl radical quantitation. Chance et al. [44] performed a systematic dosimetry study using a fluorescent dye, Alexa Fluor 488, as a chemical dosimeter to investigate the effects of experimental and sample conditions on the generation of hydroxyl radicals by synchrotron radiolysis. The assay depends on the dosimeter’s fluorescence intensity of, which decreases linearly with exposure time. This approach allows the measurement of the rate at which the dosimeter is oxidized (the rate at which HO• is generated); this rate is a reference rate that can be compared to those measured under other conditions. Sharp et al. [180] used a similar approach, which incorporates an adenine-fluorescence-based dosimeter to measure quantitatively the effective hydroxyl radical dose.
These methods afford convenient and reliable means to compare the HO• dose from sample to sample, enabling reproducible experimental results across a variety of platforms and applications. A defined initial concentration of hydroxyl radicals, however, is not obtainable by them. Quantitation of HO• generated in FPOP is difficult because the radicals are short-lived. Hambly and Gross [164] estimated the initial concentration of hydroxyl radicals to be 1 mM, allowing for the molar absorptivity of H₂O₂ and the quantum yield of HO•. Chen et al. [181] reported an experimental measurement of the initial concentration of HO• by LC/MS using phenylalanine as a dosimeter, and measured in a preliminary way a concentration of 0.42 mM. Recently, Niu et al. [165] described a determination of the initial hydroxyl radical concentration upon laser photolysis by using an isotope dilution GC/MS method, which provides good accuracy and sensitivity [182], and corrects for analyte loss during handling [183] (see details in Chapter 2). The method utilizes undeuterated phenylalanine (d₀-Phe) and deuterated phenylalanine (d₅-Phe) as dosimeter molecule and internal standard, respectively. Upon laser photolysis, we monitored and quantified the decrease of d₀-Phe signals, which correspond to the amount of HO• generated, as no radical scavenger was present. We determined quantitatively that the initial HO• concentration (from 15 mM H₂O₂ for a laser pulse of 40 mJ) in FPOP was 0.95 mM, a dose comparable to our previous estimate [164].

Further with this isotope dilution GC/MS approach, we were able to measure the effective hydroxyl radical concentrations in the presence of different radical scavengers (i.e., histidine, glutamine, methionine), and probed the changes of [HO•] as a function of scavenger concentration and identity. The outcome provides direct evidence that the
radical dose, and therefore the primary radical lifetime, is adjustable by using radical scavengers.

### 1.3.3 Primary Radical Lifetime and Adjustment of Radical Scavengers

In the absence of radical scavenger, most hydroxyl radicals generated after each laser pulse undergo self-recombination, with a rate constant of $4.7 \times 10^9$ M$^{-1}$s$^{-1}$ [155, 156]. Hambly and Gross [164] demonstrated that the protein was heavily modified if no scavenger was present, indicating a prolonged lifetime of primary radicals. In the presence of 20 mM glutamine as scavenger, however, the extent of protein modification became less, consistent with a reduced radical lifetime. The attenuation of protein modification extent continued when 20 mM phenylalanine, instead of glutamine, was used as radical scavenger. The reaction time under this condition must be extremely short, because the rate constant of phenylalanine reacting with HO• is approximately 10 times greater than that of glutamine [44]. The time courses of HO• concentration can be obtained by numerical simulations, which are based on a series of non-linear differential equations solved with an adaptive step-size Runge-Kutta method implemented in MathCAD software [184]. The simulations model three reactions with differential equations: the first reaction is that of the hydroxyl radical with scavenger (e.g., glutamine, histidine, phenylalanine), taking advantage of the known scavenger rate constant and initial concentration; the second reaction is the self-recombination of hydroxyl radical with a rate constant of $4.7 \times 10^9$ M$^{-1}$s$^{-1}$ and an initial concentration that can be determined from the dosimetry experiment (details see Chapter 2) [165]; the third reaction is the Haber-Weiss chain reaction that generates HO• from H$_2$O$_2$ [71, 153, 154].
Figure 1-4 shows the plot of simulated HO• concentration versus time based on an initial HO• concentration of 0.95 mM in FPOP. From the plot, we can interrogate the lifetime of primary radical (we assign the lifetime to be that at which the concentrations of •OH are 100 times less than that of the protein). For example, we show that with 20 mM glutamine as radical scavenger, the concentration of hydroxyl radical is less than $10^{-8}$ M in about 1 µs, similar to the earlier estimation of Hambly and Gross [164]. Glutamine at a concentration of 20 mM was originally used as radical scavenger in FPOP [116, 185, 186]. The scavenger, however, is not limited to glutamine. We show in Figure 1-4 that, with varied identity/concentration of radical scavengers, the time range available to primary radicals in FPOP can be shifted from either ~1 µs to ~100 ns (using scavengers of higher reactivity or concentration), or to high µs (using scavengers of lower reactivity or concentration). For example, with 20 mM of histidine or methionine as scavenger, the lifetime of primary radical can be as low as 0.1 µs.

Using amino acid as radical scavenger adds great flexibility in FPOP, as the rate constants of amino acids towards HO• can vary by up to three orders of magnitude [44]. We took advantage of this flexibility and tuned the FPOP labeling times by adjusting the identity and/or the concentration of scavengers [165]. As shown in Figure 1-4, with 2 mM phenylalanine as radical scavenger, we can achieve nearly the same labeling time as that from 20 mM glutamine. This observation can be significant, because in some scenarios the identity of scavenger must be different while the labeling time needs to be maintained constant. For example, histidine, often used in protein formulation buffers in the biopharmaceutical industry, owing to the “histidine stabilizing effect” [187, 188], is now
commonly used as radical scavenger as a surrogate for glutamine [169, 185, 189]. Histidine is a more efficient radical scavenger compared to glutamine because it is intrinsically more reactive with HO• [44]. Yan et al. [169] in a series of FPOP comparisons of WNV E DIII, found that, the FPOP labeling time with 350 µM histidine as scavenger is very similar to that with 20 mM glutamine.

Adding radical scavengers in FPOP to limit the reaction time minimizes protein conformational isomerism while being oxidized [190]. The adjustment of radical scavengers (i.e., concentration, identity) enables footprinting over a wide timescale, providing a flexible approach suitable for various protein conformational characterization, and a foundation for compensating for differential scavenging.

Figure 1-4. Numerical simulations of HO• concentration versus time. FPOP labeling time is adjustable depending on the identity and concentration of radical scavenger. Reprinted with permission from Ref. [165]
1.3.4 Differential Scavenging and Reporter Peptide in FPOP

Hydroxyl radicals are highly reactive with many species [191, 192]. In FPOP, our preference is they react mainly with proteins under study in competition with the radical scavengers (e.g., histidine, glutamine, phenylalanine). Given their reactive nature, they also modify other species in solution including formulation buffer components, protein ligands, and sample detergents [167]. The presence of these species, sometimes adventitious, affects the lifetime of hydroxyl radicals, giving rise to unpredictable and misleading extents of modifications. In Chapter 5, for example, we report that trifluoroethanol (TFE) in solution is an adventitious radical scavenger that causes biased FPOP comparisons between TFE-treated and untreated samples [193]; in Chapter 7, we describe the footprinting of proteins associated with a Nanodisc, which contains abundant phospholipids and membrane scaffold proteins [194, 195]. The presence of these materials also results in differential scavenging between Nanodiscs-bound and unbound protein.

The consequence of differential scavenging is an unpredictable number of hydroxyl radicals delivered to the protein samples under comparison [44, 167, 180, 196], affording a misleading FPOP fraction modified. The extent of protein modified is not only determined by its solvent accessibility and residue-specific reactivity but also by the effective dose of hydroxyl radicals, as shown in Eq 14.

\[
\text{Fraction modified} \propto (\text{Solvent accessibility}) \ast (\text{Sequence}) \ast (\text{Effective dose}) \quad \text{Eq 14}
\]
We recently incorporated Leu-enkephalin as a reporter peptide to disentangle the complication of differential scavenging, because the extent of modification of the reporter is a measure of the radical dose for the protein under study. Furthermore, its modification extent at various labeling times, achieved by varying concentrations of histidine as scavenger [167] (details discussed in Chapter 3) is an approximate measure of the time of reaction. That is, the fraction modified of Leu-enkephalin serves as a “timer” of the primary radical lifetime. Longer reaction times, as controlled by the concentration of the scavenger, afford more FPOP modifications on the reporter peptide, and on the protein under study as well. By plotting the fraction modified of Leu-enkephalin at each histidine concentration versus the corresponding fraction modified of peptide from protein digest on the $y$-axis, we obtain a time-dependent output similar to that of HDX kinetic curves. The dependence between the time and yield of reporter peptide, however, does not follow a simple, linear relationship [167].

A time-dependent measurement of FPOP provides assurance that the modification reactions are occurring normally, and adds statistical weight to the data and confidence to the measurements [167, 197]. It is especially useful when adventitious radical scavengers are present. We aim to enlarge the library of reporter peptides, based on the following criteria: the reporter should be (1) soluble, with some hydrophobicity, to avoid eluting with solvent front; (2) unstructured, so that its fraction modified is not confounded by structural changes; (3) moderately reactive towards HO$, to monitor its lifetime; (4) non-interacting with the proteins under study; and (5) readily available. Ideally, a short peptide with one FPOP-reactive site is preferred, because this can simplify the
quantitation of reporter modification and the normalization. Xie and Sharp [180] recently reported a fluorescence-based reporter, adenine, which also allows monitoring of radical dose and correcting the differential scavenging in FPOP experiments.

The incorporation of a reporter peptide in FPOP paves the way to quantitative measures of residue and peptide reactivity, analogous to the “protection factors” in HDX [198]. This will require calibration of a reporter peptide’s rate constant against a standard reference reaction. Chance and co-workers have identified a similar goal for the synchrotron footprinting [199]. This would facilitate cross-laboratory comparisons of FPOP results and enable its use to determine coarse-grained, high-order structures of proteins.

1.3.5 New Reagents for FPOP Platform

To footprint a protein with chemical reactions, it is critically important to initiate and quench the reaction rapidly to ensure that the footprinted protein is not additionally modified [200]. The FPOP platform largely meets this requirement by using a flow system and a pulsed laser to footprint; when the laser irradiation is off, no additional radical is formed and the solution moves out of the laser beam. The reaction is quenched chemically by scavengers in solution. To establish a footprinting platform applicable to multiple types of protein problems (e.g., transmembrane protein, globular protein), and to afford a toolbox with various radicals targeting different amino acids with different selectivity and specificity, we seek to expand the current FPOP platform and increase its versatility by exploring reagents other than hydroxyl radicals.

Gau et al. [49] described a new footprinting design with the FPOP platform using the sulfate radical anion SO$_4^{2-}$, generated photochemically from a low amount of sodium
persulfate (NaO₃SO-OSO₃Na). This persulfate system is more reactive in footprinting, requiring 3-5 times less persulfate to give similar extent of protein modifications as seen with the original HO• design, probably owing to the high reduction potential of SO₄⁻ compared to that of HO• (2430 mV vs. 1900 mV) [201]. Gau et al. showed that, with apomyoglobin and calmodulin as model proteins, the modification products and residue selectivity of this approach highly resemble that of hydroxyl radical footprinting, including the substitution by OH, not SO₄, with minor exception.

Chen et al. [202] developed, contemporaneously, an FPOP-based footprinting method by photochemically dissociating iodobenzoic acid at 248 nm to give •I and •C₆H₄COOH. The iodine radical species generated in this method is relatively unreactive, long-lived, and targets only histidine and tyrosine residues of protein, as opposed to •OH that modifies three quarters of the 20 amino acids [44, 164]. The high specificity greatly simplifies the post-footprinting processes, including LC/MS feature detection, data processing, and labeling-site identification.

Carbene chemistry has recently been used as a labeling method for structural proteomics [203, 204]. Zhang et al. [205] adapted the current FPOP platform to utilize carbenes as the footprinting reagent in flow cell, where a Nd-YAG laser, at its third harmonic (wavelength of 355 nm) photolyzes photoleucine as the carbene precursor (Figure 1-5). This carbene FPOP platform highlights the complementarity of those residues that are relatively “FPOP-silent”, including Ser, Thr, Glu, and Asp. Further, owing to the rapid quenching with water in protein solution, the lifetime of carbene diradicals during
footprinting is estimated to be on ns timescale, obviating the need to add any chemical scavenger [206].

**Figure 1-5.** Schematic of the custom-built flow system for carbene footprinting on an FPOP platform where a pulsed Nd:YAG laser is used to photolyze carbene precursor. Reprinted with permission from Ref. [205]

### 1.4 Applications of FPOP

#### 1.4.1 Protein-protein Interactions and Protein Therapeutics: Epitope Mapping

Now that we’ve covered described various methods of footprinting and their advantages and methodology, we turn to various applications to illustrate problem solving in
biochemistry and biomedicine. We will restrict the application to those with FPOP, and applications with other approaches were reviewed elsewhere.

Protein-protein interactions (PPIs) play a central role in regulating many biological processes [207]. Malfunction of PPIs leads to multiple diseases [208, 209]. Characterization of PPIs provides pivotal insights about the protein-protein interface, which is essential in guiding drug discovery and development [210]. Determining protein-protein interfaces can be challenging, because the interfaces typically incorporate large amorphous surface area [211, 212], comprise a linear (continuous sequence) or conformational (discontinuous residues proximal in tertiary structure) epitope [213, 214], and are generally less well-defined. There are increasing interests in designing protein therapeutics (e.g., monoclonal antibodies, a.k.a., mAbs) to modulate PPIs [215-218], as the interacting region of an antigen with mAb (i.e., epitope) is the key to successfully disrupt and inhibit PPIs [213, 219, 220].

Owing to the high sensitivity, fast turnaround, and low sample consumption [9, 37, 221], mass spectrometry-based methods have been recently used for characterizing protein therapeutics, by monitoring the changes of high order structures (HOS) and the effects of post-translational modifications (PTMs). Although top-down approaches for protein therapeutics are emerging [222], the more accepted and well-developed strategy is protein footprinting coupled to bottom-up approaches (e.g., HDX, FPOP), which allow interrogation of solution-phase protein conformations. HDX has been widely used in protein therapeutics to, for example, provide HOS information of mAbs [223-226]. FPOP footprinting, which is a more recent development compared to HDX, has emerged as an
alternative to not only complement the HDX results, but also to afford potentially more reproducible and reliable outcome, owing to the irreversible nature of modifications on protein. The short labeling time ($\mu$s timescale) of FPOP makes it advantageous to probe dynamic conformational changes, which might not be resolved using HDX.

To demonstrate that FPOP is sensitive to conformational changes in protein therapeutics, we chose disulfide isomeric IgG2 antibodies, and used FPOP to assess the different conformations between the wild-type IgG2 and disulfide mutants [186]. The footprinting results were further corroborated with top-down experiments including ion-mobility (IM)-MS and native ESI with ECD fragmentation [222]. FPOP shows that the light-chain region of IgG2 antibodies undergoes more modifications, suggesting a higher solvent accessibility. This finding concurs well with the top-down ECD results, where we observed extensive fragmentations of the light-chain region. Moreover, FPOP reveals the conformational differences of the hinge region between the WT and mutant IgG2 antibodies. This hinge region, interestingly, is exactly where the major differences of disulfide bonding pattern occur. The decreased FPOP yield in hinge region of disulfide mutants indicates more compact structures for mutants as opposed to WT. Such compact structures of mutants were further confirmed using IM-MS that showed apparently longer drift times for mutants compared to WT IgG2.

Epitope mapping, which is the process of experimentally determining the binding sites of antibodies on their target antigens, aids in the discovery and development of new therapeutics and diagnostics [227]. The epitope of serine protease thrombin, when complexed with antibody, was footprinted using FPOP and analyzed by mass
spectrometry-based bottom-up approach, marking the first time hydroxyl radical labeling is applied to epitope mapping [116]. Both antibody-bound and unbound protein samples were subjected to FPOP footprinting and proteolysis, resulting in 34 tryptic peptides with a sequence coverage of 86%. The FPOP results show, at residue-level, two regions protected upon binding to antibody, consistent with the HDX results by Komives and co-workers [228]. FPOP further identified, however, two loop regions (i.e., 99-loop and 148-loop) with increased solvent accessibility, as represented by increased FPOP modifications upon binding. This observation was not previously seen with HDX, there are several aspects of HDX and FPOP footprinting, however, pertaining to ascribing differences in their outcomes. First, the timescale in between HDX and FPOP is significantly different: FPOP snapshots the protein conformation on µs timescale, whereas HDX takes an averaged ensemble of protein conformations over seconds. Second, HDX monitors the backbone of amides, whereas FPOP focuses on protein sidechains. We attribute this increased solvent accessibility, exclusively observed by FPOP, to the ligand-induced allosteric changes of thrombin loop regions, which likely switch a number of different conformations in a short time and therefore, difficult to be captured by HDX [229, 230]. An alternative possibility is the changes of sidechain orientations of the loop regions where sidechain rotations are generally facile and can only be reported by FPOP. This unique observation by FPOP underscores the level of sensitivity FPOP can reach when applied to epitope mapping, making FPOP a promising tool in identifying those “hidden” epitopes.
For example, in the scenario where we employed FPOP to characterize the epitope of human exEGFR-Adnectin 1 complex [169], we compared FPOP yield between the two states of exEGFR (Adnectin 1-bound and unbound). The differential FPOP yield should reflect the changes of solvent accessibility of human exEGFR, corresponding to either intermolecular interactions or allosteric changes. We identify, at residue level, 5 amino acids with decreased FPOP modification upon binding (Figure 1-6), 4 of them (Leu14, Leu17, Phe20, and Leu69) fall into the putative EGFR-Adnectin 1 interface according to the previously determined X-ray structure [231]. The additionally identified residue Phe24, although locates proximally to Leu17 and Phe20, was not revealed as epitope site in X-ray structure. The structures by X-ray crystallography are often deemed as the “reference” structures, the static contact interfaces determined as such, however, might not necessarily be identical to those identified with solution-phase methods (e.g., FPOP) [232, 233].

We recently presented the mapping of interface between vascular endothelial growth factor (VEGF) and a fragment antigen binding region of an antibody (Fab-1) using FPOP [185]. The results of FPOP, complemented by previous carboxyl-group footprinting and native top-down MS analysis [234, 235], are consistent with the predicted binding interface as determined by crystallography and alanine-scanning [236]. We also observed, however, residues with reduced modification that are distal to the binding region, suggesting occurrence of remote conformational changes. That both FPOP and carboxyl-group footprinting showing epitopes and remote conformational changes reinforces the
capability of protein footprinting techniques of revealing structural mechanisms that may not be resolved by static techniques such as X-ray crystallography.

**Figure 1-6.** Crystal structure of the exEGFR (colored in grey) complexed with Adnectin (colored in red), PDB code: 3QWQ. The five residues with reduced FPOP modifications upon binding are shown and colored in yellow. Reprinted with permission from Ref. [169]

### 1.4.2 Protein Aggregation / Oligomerization

The phenomenon of protein aggregation is a special case of protein-protein interactions where, instead of interacting with other proteins or ligands, the protein self-associates and forms higher molecular-weight molecules. The accumulation of protein aggregates is often related with diseases [237-241]. In fact, protein aggregates have been implicated in a wide variety of diseases known as amyloidosis, including Alzheimer’s disease (AD), Parkinson’s disease (PD), Huntington’s disease (HD), amyotrophic lateral sclerosis (ALS), and prion diseases [242, 243]. Currently, there are no effective cures for amyloidosis,
partially owing to a lack of means to characterize the protein aggregates at different stages, and a poor understanding of the aggregation mechanisms [244]. Therefore, a platform that enables fast and accurate characterization of protein aggregates, an in-depth knowledge of protein-specific aggregation pathway, and a more detailed understanding of the factors that trigger and stabilize aggregate formation are required for future development of disease-curing therapies.

Owing to the vast heterogeneity and propensity to oligomerize, aggregation-prone proteins are challenging amenable to traditional high-res approaches including X-ray crystallography and NMR. Fluorescence-based methods, and atomic force microscopy (AFM) have been recently utilized [245-247], but only low structural resolutions are attainable. Mass spectrometry-based techniques have emerged as alternative approaches to characterize protein aggregations. Top-down analysis coupled to native ESI, which preserves the solution-conformation and non-covalent interactions of protein aggregates by spraying the sample at near-native environment, has enabled detection of protein aggregates as large as hundreds of kDa [248-250]. Protein footprinting approaches interfaced with bottom-up analysis, on the other hand, provide opportunities to obtain regional information on aggregates. Two exemplary approaches are HDX [251, 252] and oxidative footprinting. For example, we have implemented the pulsed HDX technique to probe the conformational changes of amyloid beta (Aβ) peptides aggregation at various stages [52]. The successful use of HDX requires constrained downstream processes including sample cleanup, enzymatic proteolysis, and peptides separation, in order to prevent back-exchange issues [253]. Hydroxyl radical footprinting,
advantageously, is free of such concerns, owing to the nature of irreversible modifications. Chance and co-workers have used synchrotron radiolysis-based hydroxyl radical footprinting method for structural analysis of multiple protein aggregates, including proinsulin hexamer assembly [254], Aβ1-40 fibril [255].

We recently described a platform with fast photochemical oxidation of proteins (FPOP) and mass spectrometry to follow a time-dependent aggregation of Aβ1-42, putatively the most pathologically relevant species in Alzheimer’s disease [256]. Essentially, pretreated samples, at various incubation time points with different extents of oligomerization, were submitted to FPOP for fast, irreversible labeling. We hypothesized that the FPOP modifications respond sensitively to the solvent accessibility changes of Aβ protein as it aggregates progressively. Mass spectrometry analysis of intact, HO• footprinted Aβ samples show that the fraction modified decreases as incubation time increases (Figure 1-7). Peptide and residue-level information were achieved by implementation of Lys-N rapid digestion and LC-MS/MS. The FPOP results suggest that the middle domain and C-terminal domain of Aβ1-42, as represented by peptides 16-27 and 28-42 that show significant decrease of FPOP yield over time, are involved in the self-association. Specifically, we observed a 6-fold decrease in rate from Aβ1-42 monomers to fibrils for peptide 16-27, which convincingly suggests the pivotal role of this region in forming the aggregates. A recent MD simulation study also suggests that the middle region of Aβ1-42 is involved in the formation of nucleation interface [257]. In contrast, the FPOP yield of the N-terminal domain (represented by peptide 1-15) stays high (~85%) throughout the aggregation, indicating this region is structurally flexible with little self-association and
little loss of solvent accessibility as Aβ1-42 self-associates. This finding agrees well with
the previous solid-state NMR results that show a disordered N-terminal region in multiple
Aβ1-42 oligomers and fibrils [258].

Most aggregation-prone proteins follow a nucleation-dependent pathway to aggregate
[259-261]. The protein-protein interface that induces the aggregation functions like a
“scaffold core”, excludes solvent, and promotes the formation of larger molecular weight
molecules. The aggregation interface should exhibit decreased FPOP modifications as it
is less solvent accessible. For instance, we reported the successful FPOP
characterization of structural differences between tetrameric apolipoprotein of WT ApoE3
and its monomeric mutant (ApoE3MM), aiming to identify the aggregation interface of
ApoE3 protein [197]. FPOP localizes, at residue level, that the C-terminal tail domain is
involved in the ApoE3 oligomeric interaction, because the FPOP yield of monomer is
significantly higher than that of tetramer in C-terminal tail. We also employed, parallelly,
the carboxyl-group footprinting (i.e., GEE), which specifically modifies solvent-accessible
carboxyl sidechain of aspartate and glutamate [262], and reached a similar conclusion as
FPOP, although the two footprinting approaches are different regarding the reaction
timescale and residue specificity.

Reinforced by numerous biophysical techniques [263-265], the conclusion that the C-
terminal domain is the aggregation interface of ApoE3 is demonstrated, at amino acid
resolution, with two protein footprinting techniques (i.e., FPOP and GEE). FPOP further
identifies, however, a dynamic hinge region (amino acids 183-205) that is possibly also
involved in the oligomeric interaction, evidenced by several residues in this region with
increased FPOP modifications for monomer. Some previous studies indicated that the hinge region is required for the inter-domain interactions [266, 267]. The carboxyl-group footprinting, however, is not sensitive enough to resolve changes in dynamic region, likely owing to the dramatically longer labeling timescale (a few minutes) compared to FPOP (µs). As discussed in previous section (4.1), a longer labeling time renders sampling of an ensemble of protein conformations over time, and compromises the capability of probing protein dynamic motions.

![Mass spectra of intact, FPOP-labeled Aβ1-42 (+5 charge) as a function of incubation time: (A) Control: Aβ1-42 monomer with all reagents, including H2O2, flowed through the FPOP tubing but without laser irradiation; (B) Extensively hydroxyl radical-modified, unstructured Aβ1-42 monomers; (C–E) decreasing FPOP modification extents of Aβ1-42 aggregates, reflecting increasing structural protection to FPOP modification; (F) minimal Aβ1-42 modification, reflecting solvent-inaccessible, highly ordered core structure of the fibrillar aggregates that resist FPOP modification. Reprinted with permission from Ref. [256]]
1.4.3 FPOP Probing Dynamic Motions of Proteins

As discussed earlier, FPOP footprinting platform harbors the advantages of ultrafast labeling, high sensitivity, and low selectivity, which have enabled the finding in ApoE3MM that a dynamic hinge region (amino acids 183-205) is also involved in oligomerization, besides the C-terminal region (amino acids 232-251). It has been well acknowledged that protein functions are not only determined by its structure, but also by the dynamic motions it undergoes [268-270]. The dynamic properties intrinsic to a protein structure may provide information on the location and the energetics of the conformation change process, and are thus the focus of many biophysical studies [16]. These dynamics span from secondary to higher order structures, and range from picoseconds to hours in timescale [271, 272].

Despite the obvious importance of dynamic motions to protein function, determining the dynamics of protein motion has long been an area largely unstudied, owing to lack of suitable experimental and theoretical probes. Relaxation experiments, such as the use of temperature-jump relaxation spectroscopy, can provide a means for protein dynamic studies with low structural resolution [273, 274]. X-ray crystallography, on the other hand, presents atomic-level structural information; it is not amenable, however, to protein dynamics because it typically represents an averaged protein conformations over seconds to hours [10, 14]. NMR spectroscopy techniques, in contrast, can be used to monitor the dynamic behaviors of proteins with high resolution on a broad range of timescales [16]. Paralleling the NMR developments, multiple mass spectrometry-based methods have also been developed to monitor protein dynamics [275]. Contemporaneous
with these experimental methodologies, recent advances in computations have led to the development of increasingly successful molecular simulations of protein structural dynamics that are intrinsic to biological process [276, 277].

TEM β-lactamase (TEM) endows bacteria with resistance to numerous antibiotics, by rapidly evolving to mutant TEM variants that are capable of degrading new drugs [278]. TEM-52, for example, which carries mutations of E104K, G238S, and M182T, can hydrolyze cefotaxime 2300-fold more efficiently than the wild-type TEM-1 [279]. Despite mutational changes in amino acid sequence, the crystal structures of TEM-52 and TEM-1, however, are surprisingly similar [280]. We hypothesized, based on the prediction of extensive MD simulations, that there are hidden dynamic conformations of TEM, eluded by X-ray crystallography, accounting for the functions of TEM with these effects of mutations [281]. To experimentally confirm the presence of hidden conformational states of TEM, we resorted to rapid mass spectrometric footprinting (FPOP), because the μs labeling timescale of FPOP is much faster than TEM unfolding [282], and similar to the timescale of some secondary structural motions of WT and mutant TEM [281]. Specifically, FPOP identified restricted motions of Ω-loop in TEM-52, as represented by reduced modification extents in regions proceeding the Ω-loop and the Ω-loop itself. This finding supports the modelling prediction that the increased cefotaxime hydrolysis activity of TEM-52 is associated with the reduced Ω-loop mobility.

1.4.4 FPOP Accessing Protein Folding

The fast labeling of FPOP not only allows probing of protein dynamic motions, but also paves the avenue to elucidate protein folding mechanisms, which remain challenging in
structural biology, owing partially to the short-lived folding intermediates formed along the pathway [283]. Conventional means such as rapid mixing have been used to study the folding of many two-state or multi-state systems [284, 285]. The mixing dead time, however, presents a major limitation for many fast-folding events. Recent advances in mixing devices lower the mixing dead times to low μs level [286-288]. An alternative to assess fast folding is the temperature-jump relaxation [273], as discussed in the previous section 4.3.

The folding kinetics can be tracked by spectral probes that detect various features of folding intermediates. For example, one could use NMR to probe the hydrogen/deuterium exchange at amino-acid level after rapid mixing [289]. Recent advances in mass spectrometry-based protein footprinting have enabled protein folding studies by examining solvent accessibility as an indicator of protein conformational changes, as monitored by changes in labeling extents measured by LC/MS [290, 291].

We recently described a pump/probe platform that combines T-jump relaxation and FPOP footprinting, allowing tracking of protein fast folding on submillisecond timescale [292, 293]. The platform incorporates two lasers, with the first laser supplying a fast temperature-jump (ns) and the second laser generating HO• to footprint the protein irreversibly within 1 μs (Figure 1-8). The delay time between the two lasers, however, is tunable and reflects the protein folding time. The platform, interfaced with proteolysis and LC-MS/MS analysis, reveals protein folding kinetics at peptide/residue level. To demonstrate, a model protein, barstar, was used because it denatures at 0 °C and folds with T-jump. Eight different time delays between the two lasers were selected to follow
the folding kinetics (i.e., 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, and 2.0 ms). We observed that the helix1 of barstar, as represented by three residues H17, L20, and L24, exhibits decreased HO• modifications during the first 0.1-1 ms of folding; while the modification extent of other regions remains unchanged. This observation indicates the involvement of helix1 in barstar early folding, as reinforced by previous study that barstar fast folding follows a nucleation-condensation mechanism with the nucleus centered on helix1 formed in a folding intermediate [294]. The NMR structure (PDB 1BTA) also shows that the three FPOP-identified residues are tightly buried in a hydrophobic core in the native state of barstar [295]. Further, an intermediate folding state formed 2 ms after initiation of folding was also identified with this approach.

The combination of pump/probe platform and mass spectrometry provides insights into the protein folding/unfolding conformational changes occurring in submillisecond timescale at residue-specific resolution. This platform can either probe even faster biological processes, by narrowing the time delay between the two lasers to even closer, or probe slower processes (e.g., late stages of protein folding) by displacing the probe laser from the pump. The flexibility of adjustment in time frame allows drawing of a more complete picture of protein folding event.

FPOP has also demonstrated its capability to track folding conformational intermediates of large proteins. For instance, in the event of virus infecting a cell, a receptor-binding protein and a metastable fusion protein (pre-fusion state) of virus would coordinate effectively to facilitate membrane fusion and genomic transfer [296-299]. The receptor-binding protein, upon activated by receptor, often induces extensive refolding event of
fusion protein to form a stable post-fusion state [298, 300]. Understanding of the refolding process can be critically important in inhibiting virus infections. Apart from some structures available for the pre-fusion and post-fusion states, however, little is known about the structural details for the folding intermediates. Recently, a soluble form of parainfluenza virus 5 (PIV5) fusion protein was triggered to refold with heating and footprinted along the refolding pathway using FPOP [301]. To characterize the conformations of refolding intermediates, high-resolution tandem mass spectrometry was utilized. The use of heating as the surrogate to induce refolding, and FPOP to footprint, make it possible to take “snapshots” of intermediate folding structures. Intact mass spectrometry analysis suggests that the metastable pre-fusion state (before heating) is overall more solvent exposed compared to post-fusion state, as evidenced by the higher FPOP yield in pre-fusion state. Bottom-up analysis monitors the refolding at peptide and residue levels, permitting the creation of a detailed model of the refolding event. FPOP reveals that, interestingly, changes in solvent accessibility of fusion protein refolding are region-dependent. For example, FPOP identifies two regions (represented by peptides 79-91 and 130-141) with significantly increased modification extent, upon heating from 21 °C to 45 °C. The product-ion spectra showed, with high confidence that, Leu87, Thr89, Ile90, Ile137, and Leu138 are accountable for the differential modifications, indicative of the increased solvent accessibility of those residues during refolding, consistent with their exposure upon release of fusion protein and the extension towards the target membrane [302-304].
To summarize, FPOP footprinting provides structural characterizations with medium-resolution suitable for both fast protein folding kinetics and complicated transient protein refolding events. The marriage of folding triggers (e.g., T-jump) and FPOP allows probing of the dynamic folding movements under physiologic conditions.

Figure 1-8. Schematic of a custom-build two laser pump/probe system on an FPOP platform. The time between the two lasers is adjustable. Reprinted with permission from Ref. [292]

1.4.5 Membrane protein

Membrane proteins interact with, or are part of, biological membranes. They are targets of more than half of all modern medicinal drugs [305]. Approximately 20-30% of all genes in most genomes encode membrane proteins [306, 307]. Most membrane proteins function by interacting with binding partners (other proteins, peptides, or themselves) for signaling and intracellular communication, vesicle trafficking, ion transport, and protein translocation, the conformations of membrane proteins change during these processes.
The determination of membrane protein conformations, however, has remained challenging largely due to their hydrophobicity, which makes it difficult to establish experimental conditions where the native protein structures are preserved.

The use of Nanodiscs as a carrier to incorporate membrane proteins into lipid bilayers provides near-native membrane environment and controllable stoichiometry of target membrane proteins (Figure 1-9 A) [194, 309-311]. This has enabled various biophysical analysis including mass spectrometry-based approaches to study membrane proteins [312-314]. HDX has been used to characterize membrane proteins in the presence of detergent micelles provided the protocols encompass fast cleanup, effective digestion, and good solubility of proteins [315-317]. Hydroxyl radical footprinting, on the other hand, is not so much constrained because the irreversible labeling provides flexibility in proteolysis and chromatography as there is no concern for back-exchange, and more importantly, allows efficient lipid removal prior to MS analysis.

Bearing these merits, hydroxyl radical labeling is emerging as an alternative footprinting approach for membrane protein studies. Konermann et al. [318] first demonstrated the feasibility of FPOP on a membrane protein, bacteriorhodopsin, in a native lipid bilayer environment. Chance and co-workers [150], in the meantime, footprinted membrane proteins using synchrotron radiolysis with HO• to locate structural water and characterize conformational changes. We recently described an FPOP footprinting of a Nanodiscs-incorporated membrane protein, light harvesting complex 2 (LH2), to probe its structure and topology [196]. Most of LH2 structures are transmembrane helices, the outer-membrane regions, however, are short and have little higher order structures. FPOP
results show that the outer-membrane regions are more heavily labeled by HO• compared to the regions embedded in the lipid bilayers (Figure 1-9 B).

More recently, we reported the use of FPOP to footprint the full-length, post-translationally modified oncogenic protein, KRAS4b, in a near-native environment facilitated by incorporation of lipid Nanodiscs (see details in Chapter 7). We employed Leu-enkephalin to normalize the differential radical dose induced by Nanodiscs, and conducted a time-dependent FPOP experiment. FPOP suggests the protection of C-terminal tail region of Nanodisc-bound KRAS4b, owing to the reduced extent of modifications observed for several peptides on the tail upon Nanodiscs binding. Other regions of KRAS4b, however, do not show changes of solvent accessibility. This convincingly indicates the proximity of C-terminal region to the membrane whereas the majority of the KRAS4b is outside. Such protein-membrane association is corroborated by complementary biophysical means including neutron reflectivity, NMR, SAX, and SANS.
Figure 1-9. (A) Schematic illustration of a Nanodisc with a transmembrane protein embedded. Retrieved from https://duonglabdrop.wordpress.com; (B) Representation of Nanodiscs-incorporated, FPOP-labeled LH2 transmembrane protein (PDB 1NKZ). Regions outside the membrane are heavily modified, whereas regions inside the membrane are almost inert to the HO• modifications.

1.5 Conclusion

FPOP footprinting allows fast, irreversible, and flexible mapping of protein surface solvent accessibility. The fast-labeling permits interrogation of protein conformations even if the protein unfolds after the reaction because of further oxidation; the irreversible nature of hydroxyl radical labeling enables more rigorous sample handling processes downstream the footprinting; and the flexibility endowed by incorporation of radical scavengers and reporter peptide makes FPOP more quantitative and tunable. Additionally, the current FPOP platform can be adapted for protein footprinting with new reagents, affording a footprinting toolbox with different labeling reagents targeting different amino acids with different selectivity and specificity. FPOP could be broadly applied to a variety of biochemical and biomedical areas, including problems associated with protein therapeutics, protein aggregation, protein fast folding/unfolding and dynamic motions, and membrane proteins.
References


88. Bridgewater, J.D., Vachet, R.W.: Using microwave-assisted metal-catalyzed oxidation reactions and mass spectrometry to increase the rate at which the copper-binding sites of a protein are determined. Analytical chemistry. 77, 4649-4653 (2005)
115. Lightle, S., Aykent, S., Lacher, N., Mitaksov, V., Wells, K., Zobel, J., Oliphant, T.: Mutations within a human IgG2 antibody form distinct and homogeneous disulfide isomers but do not affect Fc gamma receptor or C1q binding. Protein science: a publication of the Protein Society. 19, 753-762 (2010)


187. Falconer, R.J., Chan, C., Hughes, K., Munro, T.P.: Stabilization of a monoclonal antibody during purification and formulation by addition of basic amino acid excipients. Journal of Chemical Technology and Biotechnology. 86, 942-948 (2011)


Chapter 2

Dosimetry Determines the Initial Hydroxyl Radical Concentration in Fast Photochemical Oxidation of Proteins (FPOP)
2.1 Abstract

Fast photochemical oxidation of proteins (FPOP) employs laser photolysis of hydrogen peroxide to give hydroxyl radicals that label amino acid side-chains of proteins on the microsecond time scale. A method for quantitation of hydroxyl radicals after laser photolysis is of importance to FPOP because it establishes a means to adjust the yield of HO•, offers the opportunity of tunable modifications, and provides a basis for kinetic measurements. The initial concentration of hydroxyl radicals has yet to be measured experimentally. We report here an approach using isotope dilution gas chromatography/mass spectrometry (GC/MS) to determine quantitatively the initial HO• concentration (we found ~ 0.95 mM from 15 mM H₂O₂) from laser photolysis and to investigate the quenching efficiencies for various HO• scavengers.

2.2 Introduction

Protein footprinting examines protein structure and conformational changes by monitoring solvent accessibility and/or H-bonding by using either modification or cleavage reactions [1]. The marriage of mass spectrometry (MS) proteomics methods and protein footprinting allows protein structure, function, dynamics, interactions with ligands and stoichiometry to be examined [2]. Hydroxyl-radical oxidation is one method of “protein footprinting”, a term coined by Chance [3] for synchrotron-induced labeling. Hettich and Sharp [4] had earlier explored slow labeling via Fenton chemistry. Although there are other ways to produce HO• [5-7], the “fast photochemical oxidation of proteins” (FPOP) method shares the advantages that a stable, irreversible covalent modification is installed in nearly 3/4th of the amino-acid residues, providing higher coverage than most labeling methods. Given
that HO• is comparable in size to a water molecule, FPOP is a probe of solvent accessibility. Moreover, FPOP provides a fast “snapshot” of protein structure (~ 1 µs) without concern for rearrangements or loss of label during subsequent sample handling and proteolysis, and it can accommodate other radical reagents [8, 9].

FPOP utilizes a pulsed laser to photolyze hydrogen peroxide to generate two HO• that rapidly modify proteins in a flow system [10]. The laser provides a spatially small, high flux of light, maximizing the exposure of small plugs of protein solution to radicals and ensuring all but a small fraction of the protein in the flow is irradiated only once [11]. FPOP is a practical approach, requiring only a modest laser and syringe pump [12-14].

Missing thus far in the development of FPOP is a method that gives the initial concentration of HO•, allowing rational means for tuning. The concentration of HO• is also an input to any kinetic analysis to control the approach. Quantitation of HO• generated in FPOP is difficult because the radicals are short-lived. Hambly et al. [10] estimated the initial concentration of HO• to be 1 mM, consistent with the molar absorptivity of H₂O₂ and the quantum yield of HO•. Chen, using LC/MS, measured preliminarily the [HO•] to be 0.42 mM [15]. The high speed of oxidative modifications in FPOP, however, can be placed on a firmer ground by measuring the initial [HO•] with a routine method.

Here we report a determination of the initial concentration of HO• upon laser photolysis by using isotope dilution GC/MS, which is sensitive, specific [16, 17], and gives good accuracy, precision, and correction for analyte loss during handling [18]. We selected unlabeled phenylalanine as a “dosimeter molecule” to quantify the HO• and d₅-phenylalanine as the internal standard. Phenylalanine is suitable because it is reactive
towards HO• (rate constant of $6.9 \times 10^9 \text{M}^{-1}\text{s}^{-1}$ [19]), and it yields simple oxidation products by a mechanism discussed elsewhere [1].

We measured the initial HO• concentration in FPOP and found that it is comparable to our previous estimate [10]. Further, we measured [HO•] available to label proteins as a function of both the nature and concentration of the scavenger (i.e., methionine, histidine and glutamine).

### 2.3 Experimental

#### 2.3.1 Materials and Reagents

Unlabeled phenylalanine and $d_5$-phenylalanine were provided by Jan Crowley. $L$-Glutamine, $L$-methionine, $L$-histidine, catalase, 30% hydrogen peroxide, $N,O$-bis(trimethylsilyl)fluoroacetamide (BSTFA), phosphate-buffer saline (PBS) and HPLC-grade solvents were from Sigma Aldrich (St. Louis, MO).

#### 2.3.2 Calibration

A series of solutions containing increasing amounts of $d_0$-phenylalanine (0, 10, 25, 50, 100 nmol) were dissolved in PBS buffer (10 mM phosphate buffer, 138 mM NaCl, and 2.7 mM KCl, pH 7.4), mixed with $\text{H}_2\text{O}_2$, and submitted to FPOP without pulsing the laser to correct for any analyte losses. Methionine and catalase were added to each vial immediately following to simulate FPOP conditions. A constant amount of $d_5$-phenylalanine internal standard (100 nmol) was then added to each vial.

After removing the solvent by speed-vac, the samples were treated with BSTFA to form trimethylsilyl derivatives that were analyzed by GC/high resolution MS with extracted ion
chromatography and integration of ion peaks for the standard and unknown to give an intensity ratio of the analyte and the internal standard. The ratio was plotted as a function of the amount of \(d_0\)-phenylalanine initially added.

2.3.3 FPOP Dosimetry Experiment

Our goal was to determine the yield of HO• when there was no scavenger in the system (Figure 2-1) by incorporating the laser pulse for FPOP, as described previously [10]. \(d_0\)-Phenylalanine (100 nmol) as dosimeter was dissolved in PBS buffer, and \(H_2O_2\) peroxide was added just prior to syringe infusion; their concentrations were 2 and 15 mM, respectively. The sample solution was advanced at a rate of 23 µL/min by a syringe pump (Harvard Apparatus, Holliston, MA). The excimer laser power (GAM Laser Inc, Orlando, FL) at 248 nm was adjusted to \(~40\) mJ/pulse at a frequency of 7.0 Hz, affording an exclusion volume fraction of 20% (the volume not irradiated by laser, sandwiched between plugs of irradiated solution). All collections were in vials containing 500 nM catalase and 70 mM methionine, preventing further oxidation by \(H_2O_2\). A solution containing 100 nmol of \(d_5\)-phenylalanine was added to each vial, the solution dried on a speed vac, derivatized with BSTFA, and analyzed by GC/MS to give the amount of \(d_0\)-phenylalanine reacted which equals the amount of HO• generated.
Figure 2-1. Workflow of FPOP dosimetry experiment. (1) Without scavenger, 100 nmol of dosimeter (d₀-phenylalanine) was subjected to FPOP, generating tyrosine-like products from HO• attack on the aromatic ring; (2) Internal standard (d₅-phenylalanine) added to vial; (3) BSTFA added for derivatization of both dosimeter and internal standard; (4) Products analyzed by GC/MS.

2.3.4 Gas Chromatography/Mass Spectrometry

Samples were diluted 10:1 before injecting into an Agilent 7200 Q-TOF GC/MS (Santa Clara, CA) with a split of 10:1, to avoid column overload and detector saturation. The inlet temperature was 280 °C. The gas chromatograph was an Agilent 7890 equipped with an Agilent 19091S HP-5ms column (30 m, 0.25 mm id., 0.25 µm 5% diphenyl 95% dimethylpolysiloxane film coating). The GC oven temperature was at 80 °C for 2 min after sample auto-injection and then ramped at a rate of 10 °C/min to a final temperature of 300 °C, which was held for 6 min. All spectra were acquired in the positive-ion EI mode.

2.4 Results and Discussion

2.4.1 Dosimetry

To measure the [HO•], we derivatized the carboxyl and amine groups of phenylalanine with trimethylsilyl groups, taking advantage of this well-known protocol [20]. We chose fragment ions of m/z 294.1340 and 299.1654 (derivatized d₀-phenylalanine and d₅-phenylalanine) because the molecular ions (M⁺) of m/z 309.1575 and 314.1889 were not detectable [21] (structures and EI mass spectra are in Figure 2-2). Other appropriate fragment ions are of m/z 266.1391 and 271.1705; and of m/z 91.0542 and 96.0856.

To conduct the analysis, we held constant the amount of d₅-phenylalanine derivative in each sample (100 nmol), so there was little change in its ion-current peak integral (see
Figure 2-3), and increased $d_0$-phenylalanine from 0 to 100 nmol. The ratio of the ion-current peak integrals ($d_0/d_5$) was obtained from extracted ion chromatograms and plotted as a function of amount of added $d_0$-phenylalanine. The ions of m/z 294.1340 and 299.1654 gave a ratio of peak integrals ($d_0/d_5$) as a linear function of the amount of $d_0$-phenylalanine over 0-100 nmol (Figure 2-4). The curve (slope of 0.0061, y-axis intercept of 0.015, and an $R^2$ of 0.9970) provided the interpolated concentration of remaining $d_0$-phenylalanine. By subtracting this amount of $d_0$-phenylalanine from the initial amount, we obtained the amount of consumed $d_0$-phenylalanine, which is the amount of HO• from each laser pulse that reacted with the dosimeter. From these experiments, the HO• concentration that reacted with the dosimeter from each laser pulse was 0.67 ± 0.08 mM from a starting solution of 15 mM H$_2$O$_2$, an amount close to our earlier estimate [10].

We selected two other fragment-ion pairs (m/z 266.1391 and 271.1705; and m/z 91.0542 and 96.0856) to build two other calibrations (see Figure 2-4) that afforded measured [HO•] of 0.69 ± 0.12 and 0.68 ± 0.12 mM, respectively. Two control experiments were included, one without laser and the other without H$_2$O$_2$.

The initial [HO•] should be slightly higher because the measured value is reduced by HO• self-recombination, which competes with the reaction of HO• and dosimeter.

We simulated by kinetic modeling the amount of phenylalanine product to predict an initial [HO•]. A search varied the postulated initial [HO•] until the calculated HO• concentration that reacted with the dosimeter matched the experimentally determined value (0.67 mM). The best fit showed the initial [HO•] to be 0.95 mM. The simulations also included the
Haber-Weiss chain reaction, which has little effect. We did not model, however, any reaction of •OOH produced in the final step of phenylalanine modification [1].

**Figure 2-2.** Mass spectra of trimethylsilyl derivatives for $d_0$-Phe and $d_5$-Phe, molecular ions (M$^+$) were not discernable, the peaks representing [M$^+$--CH$_3$] were found at m/z 294.1 and 299.2, respectively.
Figure 2-3. GC/MS detection of BSTFA-derivatized $d_0$-phenylalanine ($m/z$ 294.1340) and $d_5$-phenylalanine ($m/z$ 299.1654). Each panel represented the EIC of trimethylsilyl...
From A to E, amount of $d_0$-phenylalanine was increasing whereas $d_5$-phenylalanine remained constant.

**Figure 2-4.** Standard calibration curves of $d_0$-phenylalanine based on fragment ions of: (A) $m/z$ 294.1340 and 299.1654; (B) $m/z$ 91.0542 and 96.0856; (C) $m/z$ 266.1391 and 271.1705. The error bars correspond to the ± SE of triplicate measurements.

### 2.4.2 Effect of Scavenger

In a typical FPOP experiment, scavengers limit the available [HO•] for footprinting. Thus, it is of interest to measure the [HO•] when using scavengers of different nature or concentrations, facilitating creation of an improved FPOP platform with scavenger-tunable [HO•]. Tuning the [HO•] is important because different proteins usually require different [HO•] to achieve modifications that are sufficient to reveal structural information, yet do not give overoxidation. Moreover, varying the scavenger provides the means for
measuring the kinetics of radical modification reactions. To test this, we used different concentrations of glutamine (10, 20, 40 mM) as scavenger and measured [HO•] to establish a relationship between measured [HO•] and scavenger concentration. Furthermore, histidine and methionine were also tested as a function of concentration (0.2, 0.8, 2.0 mM) (Figure 2-5). For each scavenger (His, Met, Gln), the measured [HO•] decreases as scavenger concentration increases. Moreover, extrapolation of the curves to the y-axis, gives approximately the same intercept, [HO•] with no scavenger (i.e., 0.67 mM) (see Figure 2-5 and Figure 2-6). The equation for each scavenger response curve is also given in those figures. The ratios of the rate constants of glutamine, histidine and methionine towards HO• (5.4×10^8 M⁻¹s⁻¹, 4.8×10^9 M⁻¹s⁻¹, and 8.5×10^9 M⁻¹s⁻¹ respectively [1]) should predict the ratios of the line slopes, and this is approximately correct. Discrepancies may be due to differences in our methods and conditions compared to the original kinetic studies.

For different amino-acid scavengers, the rate constants differ by three orders of magnitude [1]. Taking advantage of this, we can adjust the [HO•] available for footprinting by varying either the nature or concentration of the scavenger. The consequence of this adjustment may be significant because the time range available to FPOP can be shifted from either ~1 µs to ~100 ns (scavengers of higher reactivity or concentration), or to near ms (scavengers of lower reactivity or concentration). A numerical simulation of [HO•] concentrations by using different scavengers (Figure 2-6) shows that the lifetime is approximately 1 µs when using 20 mM glutamine as scavenger but adjustable to ~0.1 µs by using methionine or histidine as scavenger or to 100 µs with alanine as scavenger (we
assign the lifetime of the radicals to the time at which their concentration is 100 times less than that of the protein. Adjusting the lifetime of HO• enables FPOP labeling over a wider time scale, making FPOP a flexible tool to investigate various folding/unfolding events. For example, some proteins fold to a native state by first forming secondary and then tertiary structure [22]; whereas other proteins go through hydrophobic collapse first and then form tertiary structure [23]. The time-course changes for FPOP is in the range of the time scales for the major steps of protein folding, allowing future studies on the details of protein folding.

Figure 2-5. Dose-response curves determined from fragment ions of \( m/z \) 294.1340 and 299.1654 for Met, His and Gln. See SI for dose-response curves determined from other fragment ions (i.e., \( m/z \) 266.1391 and 271.1705; and \( m/z \) 91.0542 and 96.0856). The error bars correspond to the ± SE of triplicate measurements. Note that the lifetime of the radicals is \( \sim 1 \) µs when glutamine is 20 mM—see Figure 2-7.
Figure 2-6. Dose response curves determined by using fragment ions of (A) m/z 266.1391 and 271.1705, and (B) m/z 91.0542 and 96.0856. The error bars correspond to the ± SE of triplicate measurements.

Figure 2-7. Numerical simulation of HO• concentration vs. time. The simulation includes the HO• reaction with scavengers, the recombination of HO•, and the Haber-Weiss chain reaction. Lifetime of HO• (~100 ns to more than 10 µs) depends on the nature and concentration of the scavenger.

2.5 Conclusion
The specificity, sensitivity, and precision of isotope dilution GC/MS to identify and quantify the dosimeter recommend it for quantifying the initial HO• concentration in FPOP. The approach permits optimization of FPOP conditions and adjustment of experimental parameters to realize tunable extents of modifications. Our next step is to study the effect of tuning, aiming for secure “hard” kinetic data that will allow comparison of results day-to-day and lab-to-lab.
References


Chapter 3

Incorporation of a Reporter Peptide in FPOP Compensates for Adventitious Scavengers and Permits Time-dependent Measurements
3.1 Abstract

Incorporation of a reporter peptide in solutions submitted to fast photochemical oxidation of proteins (FPOP) allows for the correction of adventitious scavengers and enables the normalization and comparison of time-dependent results. Reporters will also be useful in differential experiments to control for the inclusion of a radical-reactive species. This incorporation provides a simple and quick check of radical dosage and allows comparison of FPOP results from day-to-day and lab-to-lab. Use of a reporter peptide in the FPOP workflow requires no additional measurements or spectrometers while building a more quantitative FPOP platform. It requires only measurement of the extent of reporter-peptide modification in a LC-MS/MS run, which is performed by using either data-dependent scanning or an inclusion list.

3.2 Introduction

Interest in elucidating protein higher order structures and the conformational changes connecting them continues to grow for both understanding fundamental biophysics and the development of therapeutic proteins. Protein conformational changes can be induced, for example, by posttranslational modifications (PTMs) [1], mutagenesis [2], oligomerization [3], and binding to metal ions, drugs, or to other proteins [4]. Multiple biophysical approaches, including circular dichroism (CD), fluorescence spectroscopy, and isothermal titration calorimetry, can report that a conformational change has occurred [5], but these approaches afford limited structural resolution. X-ray crystallography and NMR can provide site-specific information, but the turnaround is slow and the required sample quantity is large. The mass spectrometry (MS)-based method of hydroxyl-radical

FPOP utilizes a pulsed laser to photolyze hydrogen peroxide to generate two hydroxy radicals (HO•) that rapidly modify protein side chains in a flow system. The laser provides a spatially small spot but high light flux, maximizing the exposure of small plugs of protein solution to radicals and ensuring that most of the protein in the flow is irradiated only once [8]. FPOP is easily implemented, requiring only a laser, optics, syringe pump, and flow system [9, 10]. By comparing the extent of modification for two samples (e.g., wild-type vs. mutant or apo vs. holo), conformational differences can be established at the peptide and even amino-acid levels [11, 12].

There is always a need to normalize the yields of oxidative modification. For example, the presence of adventitious HO• scavengers (e.g., DTT, ATP, TFE, lipids, DMSO) in a protein solution may significantly reduce the FPOP-induced modification extent. If the scavenger amounts differ between a sample and control, the outcomes may be difficult to compare. In antigen-epitope mapping where one compares modification extent of an antigen with and without an antibody, one can use a non-interacting (“dummy”) antibody to compensate for the many reactions of HO• with the antibody, but reliable comparisons are needed to compensate for changes in the test vs. control solutions or when a non-interacting protein is not available.

Here we report that the use of a reporter peptide to resolve discrepancies in HO• dosage and “normalize” the results. We used leu-enkephalin as the reporter, where the fraction
modified is a “ruler” of radical dosage to the protein samples. This enables an important extension of FPOP, namely the acquisition of FPOP time-dependent data acquired by varying the scavenger concentration to modify the primary HO• lifetime. The extent of modification of the sample vs. that of the reporter provides time-dependent data and helps resolve conformational differences from readouts of peptide/residue responses at each time point.

A kinetics-type approach should provide assurance that the modification reactions are occurring normally and also add statistical weight to the data and confidence to comparisons of higher order protein structure, affording an output that is similar to that of HDX. Another approach that accommodates adventitious scavengers is by Xie and Sharp [13], who proposed an adenine-based dosimeter to monitor and correct the radical dose from sample-to-sample in FPOP experiments.

The reporter-peptide approach should be useful for protein therapeutics where higher order structure must be verified [14]. The ALS-linked dimeric protein Cu,Zn superoxide dismutase 1 (SOD1) is used here as test case for developing a robust method for comparing conformational changes between WT and mutant.

3.3 Experimental

3.3.1 Materials

All experiments were performed using recombinant SOD1 protein expressed and purified from BL21-Gold(DE3) PLysS E. coli cells (Stratagene, Inc., Cedar Creek, TX) as previously described [15]. The free cysteines in WT SOD1 were replaced with the
C6A/C111S mutations to avoid inter-molecular disulfide-scrambling; the alanine to valine mutation (A4V), the most common ALS variant in North America, was introduced into this pseudoWT background [16]. All proteins in this study are in apo-form. Leucine enkephalin, catalase, L-histidine, L-methionine, 30% hydrogen peroxide, phosphate-buffer saline (PBS), triethylammonium bicarbonate buffer (TEAB), HPLC-grade solvents, dithiothreitol (DTT), dimethyl sulfoxide (DMSO), trifluoroacetic acid (TFE), iodoacetamide (IAM), and sequencing-grade modified trypsin were purchased from Sigma Aldrich (St. Louis, MO, USA).

3.3.2 FPOP Experiment

A total volume of 100 µL of sample at a final concentration of 8 µM protein, 5 µM leu-enkephalin, 22.5 mM freshly prepared hydrogen peroxide and at various concentrations (0, 0.2, 2.0, 40 mM) of histidine in PBS (10 mM phosphate buffer, 138 mM NaCl, and 2.7 mM KCl, pH 7.4), was loaded in a 100 µL syringe (Hamilton, Reno, NV, USA) and advanced through 150 µm i.d. fused silica tubing (Polymicro Technologies, Pheonix, AZ, USA) at a flowrate of 27 µL/min by a syringe pump (Harvard Apparatus, Holliston, MA, USA). H$_2$O$_2$ was added just prior to syringe infusion. The excimer laser power (GAM Laser Inc., Orlando, FL, USA) at 248 nm was adjusted to ~35 mJ/pulse at a frequency of 7.4 Hz controlled by an external pulse generator (B&K Precision, Yorbal Linda, CA, USA). The laser beam was focused with a convex lens (Edmunds Optics, Barrington, NJ, USA) to an irradiation window of ~2.6 mm, affording an exclusion volume fraction of 22.5% (the volume not irradiated, sandwiched between plugs of irradiated solution). Samples were
collected in low protein-binding Eppendorfs containing 50 nM catalase and 70 mM methionine, preventing further oxidation by residual H$_2$O$_2$ and other oxidative species.

### 3.3.3 LC-MS/MS Analysis

Aliquot of 5 µL of sample was loaded onto a C18 reversed-phase desalting column (nanoViper, 100 µm x 2 cm, 5 µm, 100 Å; Thermo Fisher Scientific, Waltham, MA, USA), desalted with water in 0.1% formic acid at 4.5 µL/min for 10 min, and eluted to a silica capillary analytical column that was custom-packed with C18 reverse-phase material (Magic, 100 µm x 180 mm, 5 µm, 120 Å; Michrom Bioresources, Inc., Auburn, CA), which was connected with a metal emitter and mounted on a Nanospray Flex ion source. The gradient was delivered at a flow rate of 0.4 µL/min from 2% to 45% solvent B (80% of acetonitrile with 0.1% formic acid) with an increment of 0.67%/min, then ramped quickly from 45% to 90% at 3%/min, and slowly from 90% to 95% solvent B in 5 min and re-equilibrated to solvent A (water with 0.1% formic acid) for 6 min. The top 18 abundant charged ions were subjected to the high-energy C-trap dissociation (HCD) in the HCD cell with data-dependent scanning.

### 3.3.4 Data Analysis

The *.mgf files, converted from the *.raw MS files using ProteoWizzard 3.0, were searched using MASCOT (Matrix Science, London, UK) against a custom-built database containing the sequence of WT and A4V SOD1. All known hydroxyl radical side-chain reaction products [17] with a carbamidomethyl modification (57.0214 Da) to Cys-containing peptides were added as variable modifications. Quantitation of the extent of modification at the peptide or residue level was calculated. Signals from LC-MS for each
unmodified peptide (I_{un}) and its modified species (I_{ox}) were extracted from raw data files using Thermo Xcalibur 2.2 software (Thermo Fisher Scientific, Waltham, MA, USA) with a mass tolerance of 15 ppm. Calculation of the extent of modification was performed as previously reported [10].

3.4 Results and Discussion

FPOP labels rapidly owing to the high reactivity of HO• [7, 8]. Because the rate constants for amino acids reacting with HO• are known [17], the lifetime of HO• is adjustable by either scavenging with different amino acids or with a single amino acid at different concentrations (see Figure 3-1 for an illustrative numerical simulation of HO• lifetime as a function of time). The FPOP reaction time was varied by using histidine as radical scavenger for the OH primary radical at 0, 0.2, 2.0, 40 mM. We chose histidine because it is a simple amino acid, bio-compatible, soluble, and reactive with radicals. To normalize the results, we introduced leu-enkephalin as a reporter peptide prior to irradiation.

Following FPOP, we analyzed the undigested protein by LC/MS, integrating TI signals over an elution time of 0.5 min. The unmodified protein is the most abundant followed by a singly oxidized species (+15.9949 Da), a doubly oxidized species (+31.9898 Da), etc. (Figure 3-2 A) according to the fraction modified [18]. The kinetics of oxidative modification at the protein level are relative to the oxidative modification of reporter leu-enkephalin (M + 15.9949), determined from the extracted ion chromatograms (EICs) of both unmodified and modified leu-enkephalin (mass tolerance of 15 ppm; see Figure 3-3). As the HO• lifetime becomes longer, the extent of modification of both the reporter leu-enkephalin and the SOD1 protein increase. Plotting the protein-fraction modified versus
the reporter-fraction modified generates a curve whose x-axis reflects the varying HO• lifetime and y-axis tracks the protein response as labeling time changes (Figure 3-2 B).

**Figure 3-1.** Numerical simulations of HO• concentration versus time. Lifetime of HO• (spanning from ~60 ns to more than 10 µs) is dependent on the concentration of histidine (from 0 to 40 mM).

**Figure 3-2.** (A) Mass spectra from intact measurements of FPOP-labeled WT SOD1 (+16 charge) obtained by adjusting [His]; (B) Intact protein fraction modified vs. reporter fraction modified under various scavenging conditions. The resulting curve is independent of the presence of FPOP adventitious scavengers, it demonstrates that incorporation of reporter peptide could normalize (or correct for) the FPOP outcome, making possible unbiased comparison between one state of protein and another (e.g., apo and holo). The error bars correspond to the ± SE of triplicate measurements.
Figure 3-3. Extracted ion chromatograms (EICs) of leu-enkephalin, both unmodified and modified (+16 Da) peptides were extracted with MS1 tolerance of 15 ppm. Different isomers of +16 modified species eluted at different retention times, MS/MS spectra can provide the sequence information, which can help to identify the residue that was modified. Here we observed that Tyr1, Phe4 and Leu5 of leu-enkephalin were modified.

3.4.1 Correction for Adventitious Scavengers

Protein samples in FPOP are studied in buffer solutions with near-physiological pH, permitting FPOP to be used under reasonably biologically relevant conditions. The lifetime of HO• is tunable and should predominantly be controlled by radical scavenger concentrations [19]. Sometimes, one is unaware of the presence of adventitious HO• scavengers (e.g., DTT, ATP, TFE, lipids, DMSO) that unpredictably scavenge the radical dosage and bias comparisons of the modification extents of the two protein states. Incorporating the reporter peptide into FPOP workflow allows for correction. Although FPOP experiments in the presence of adventitious radical scavengers would show reduced protein modification, the corresponding decrease of reporter-fraction modified maintains the protein-reporter outcome on the curve, correcting for differential scavenging (Figure 3-2 B).

3.4.2 FPOP Kinetics at the Peptide/Residue Levels
To demonstrate time-dependent measurements, we plotted the fraction of leu-enkephalin modified at each histidine concentration vs. the corresponding fraction modified for each peptide of the protein (here SOD1) on the y-axis. The reporter-fraction modified reports \( \text{HO} \cdot \) lifetime (i.e., longer times give more modification), and the plot shows the time-dependent SOD1 modification, affording a similar output to that of HDX kinetics data for peptides. The response of each peptide of SOD1 as a function of reporter-fraction modified allow us to distinguish those reactive or solvent-accessible regions from the more protected regions (Figure 3-4). For example, peptides 24-36, 37-69, and 116-128 exhibit greater fractions modified for both WT and A4V compared to most other peptides, indicating that those regions are more solvent-exposed as expected for these loop regions of SOD1. The curves for peptide regions of WT and A4V SOD1 should overlap if the solvent accessibilities of these peptides are the same for WT and mutant. The nonoverlapping curves for peptides 1-9 and 123-136 (Figure 3-4) indicate a conformational change in these regions upon mutation, likely due to destabilization of the monomer-dimer equilibrium for the A4V variant [16].

Although a time dependence is achieved by this simple approach, the increasing yield for the reporter does not have a simple relationship to time. Nevertheless, the approach is simple and effective for identifying protein regions that change upon perturbation.

The functional dependence of sample-fraction modified v.s. reporter-fraction modified was described using the following Eq 3-1:

\[
y = p(1 - (1 - x)^\frac{k_p}{k_{pr}})
\]  

Eq 3-1
where \( y \) is the sample peptide-fraction modified, \( x \) is the reporter-fraction modified, \( p \) is the asymptotic sample peptide response (assume to be 1), \( k_p/k_{rpt} \) is the ratio between rate constants of sample peptide and reporter towards \( \text{HO}^\cdot \). As described in Eq 3-1, when \( k_p \) is close to \( k_{rpt} \), the relationship between \( y \) and \( x \) is almost linear.

The reporter response \( x \), is related to the lifetime of \( \cdot \text{OH} \) following Eq 3-2:

\[
x = r(1 - e^{-k_{rpt}OHT})
\]

Eq 3-2

where \( r \) is the asymptotic reporter response, \( OHT \) is the \( \text{HO}^\cdot \)-time area (see Figure 3-1), which equals to the product of initial \( \text{HO}^\cdot \) concentration and \( \text{HO}^\cdot \) effective lifetime:

\[
OHT = [\text{HO}^\cdot]_{initial} \times t_{eff}
\]

Eq 3-3

Residue-level analysis can also be normalized by using the reporter peptide. We identified and quantified with high confidence eleven residues that undergo oxidative modification (see Figure 3-5). Referring to the X-ray crystallography structure (1N18), we see that most peptides/residues that exhibit higher modification extents (F50 in peptide 37-69, K122 in peptide 116-128 and K136 in peptide 123-136) are located on loop IV (Zn-binding loop—residue 49-83) and loop VII (electrostatic loop—residues 121-142) (see Figure 3-6) [20]. A detailed interpretation of the results will be part of a more extensive biochemical study to be published in the future.
Figure 3-4. Response curves of each peptide from WT and A4V SOD1 in an FPOP kinetics experiment performed by adjusting different histidine concentrations. The fraction modified for peptides from WT and A4V SOD1 were normalized by the reporter fraction modified under each scavenging condition. The error bars correspond to the ± SE of triplicate measurements.
Figure 3-5. Residue level kinetic curves of WT and A4V SOD1. Eleven residues were identified with high confidence and plotted as the fraction modified of each residue vs. the fraction modified of reporter peptide under each histidine concentration. The error bars correspond to the ± SE of triplicate measurements.

Figure 3-6. Crystal structure of thermostable pseudo WT SOD1, containing the C6A/C111S mutations. Loop IV, i.e., the Zinc binding loop (residue 49-83, yellow) and loop VII, i.e., the electrostatic loop (residue 121-142, red) are shown. These two loops exhibited a higher extent of modification by FPOP. Residues that are significantly modified by FROP are labeled (cyan).
3.5 Conclusions

In conclusion, an advantage of this simple approach is it allows quick location of conformational changes when comparing two samples, especially when adventitious HO• scavengers are present. Although we used leu-enkephalin, many peptides are candidate reporters. Ideally, they should be (1) mildly reactive towards HO•, so as not to affect the protein radical dosage while remaining sensitive to the HO• lifetime, (2) minimally structured so that its fraction modified won’t be confounded by structural changes; (3) soluble but not too hydrophilic to avoid eluting with the solvent front; (4) non-interacting with the proteins under study, and (5) readily available. A short peptide with one FPOP-reactive residue may be preferred to simplify the quantitation of reporter modifications and the normalization of the extent of modification.

The reporter-peptide approach is the first step to realize our vision to obtain quantitative measures of residue and peptide reactivity, analogous to “protection factors” in HDX. This will require calibrating a reporter peptide’s rate constant against a standard reaction. Chance has identified a similar goal for synchrotron footprinting [21]. This would facilitate cross-laboratory comparisons of FPOP results and enable its use to determine coarse-grained HOS of proteins.
References

2. Furukawa, Y., O'Halloran, T.V.: Amyotrophic lateral sclerosis mutations have the greatest destabilizing effect on the apo- and reduced form of SOD1, leading to unfolding and oxidative aggregation. The Journal of biological chemistry. 280, 17266-17274 (2005)


Chapter 4

Using R for Mass Spectrometry-based Proteomics Data

Visualization and Analysis
4.1 Abstract

Mass spectrometry-based proteomics has developed to the point where time-efficient, accurate data processing appears to be one of the bottlenecks. The R programming language and its accompanying packages are promising for providing a solution, by creating free analytical tools that are dynamic, interactive, and adaptive to the downstream analytical processes. The ease of coding, sharing, and communication of R Shiny web-based applications have greatly expanded the applicability, allowing people specialized in areas other than computer science to participate in creating useful software toolkits.

4.2 Introduction

Proteomics is the large-scale study of proteins [1, 2]. It has yielded many biological insights into diseases, as most biological functions are transmitted through proteins [3]. Mass spectrometry-based proteomics is evolving rapidly owing to the development of new technologies [4-6]. The data output, however, is becoming increasingly complicated [7-9]. The analysis of the large amount, complicated data generated in mass spectrometry-based proteomic experiments highlight one of the most significant challenges, and is a bottleneck in many current proteomic researches [8].

HPLC coupled to tandem mass spectrometry is quickly emerging as a method of choice for large-scale proteomic analysis [10, 11]. These instruments and the accompanying workflows can be used to identify and track the relative abundance of molecules, with up to tens of thousands of fragment ion spectra per hour of data acquisition [6, 12]. Typically,
in a standard bottom-up experiment, peptide mixtures derived from proteolysis of the proteins present in the sample are spatially separated by chromatography prior to injection into the mass spectrometer. This generates information-rich, time-dependent data, consisting of both the unique elution profile (column retention times) as well as mass-to-charge ratios (m/z) of individual peptide. Those peptides of interest are subjected to fragmentation (e.g., collision induced dissociation (CID), electron transfer dissociation (ETD)) followed by database searching for sequence identification; whereas the recorded elution profiles are used to determine relative abundances. The fast assignment of the fragmentation information, correct inference of the protein species, and faithful determination of the abundances in the analyzed samples, however, are subject to a number of challenging analytical issues in structural proteomics [4, 13, 14], including identification of less-abundant sequence variants, alignment of different experimental runs with systematic variations, and development of statistical tools for large-scale data interpretation. The work presented in this chapter focuses on a particular type of software, namely R, with its add-on packages that extend its functionalities, to show the usefulness of R for mass spectrometry-based proteomic data analysis.

R is a free, open-sourced, comprehensive statistical environment and programming language for professional data analysis and graphic display [15]. It was originally created by Ihaka and Gentleman at the University of Auckland [16], and since then, been developed and maintained by the R-core group. R is currently widely used in public and private in a broad range of fields. Unlike other closed-source software, the validity of R has been reviewed by many internationally renowned statisticians and computational
scientists. The popularity of R can be ascribed to its numerous features including the huge collection of statistical algorithms, the capability to easily model and handle data with various formats, the cross-platform compatibility, and well-designed documentation. The biggest advantages of using R in bioinformatics is its high-quality, dynamic extension packages available from multiple repositories specializing in various topics including data subsetting, data mining, and graphic analysis [17]. In addition to those packages, there are numerous R-based projects concentrating on different life science areas including genomics and proteomics [18]. The Bioconductor project, for example, initiated by Gentleman [19], is specially focused on computational biology and bioinformatics and represents a central repository for more than 100 software, data processing and annotating packages, dedicated to the thorough analysis and comprehension of high-throughput biological data. Additionally, recent developments of R Shiny package (https://cran.r-project.org/web/packages/shiny/index.html) have enabled creation of powerful interactive web applications without the need to delve into other programming languages such as HTML, CSS, and JavaScript. R Shiny allows the development and redistribution of software that other developers and users can easily access, making it possible to create a user-friendly, interactive, and easy-to-maintain data-processing platform for mass spectrometry-based proteomic analysis.

In this chapter, we discuss several custom-built computational and analytical tools that can be considered when performing mass spectrometry-based proteomics analyses, for the purpose of achieving more convenient computations, more reliable quantitative and
qualitative measurements, and more adaptive data input and output that can facilitate the downstream data analyses.

4.3 Unified API (Application Programming Interface) by mzR

An LC-MS system incorporates: (1) a chromatography column that separates peptide mixtures based on physicochemical properties (e.g., hydrophobicity, hydrophilicity, total charge, size) prior to mass spectrometer detection; (2) an ionization source that generates gas-phase ions directly from solution; (3) one or more mass analyzers that record ions based on \( m/z \); and (4) a detector that registers the relative abundance of ions at discrete \( m/z \). For tandem MS, precursor ions are recorded in a full-scan mode, followed by conditional ion isolation and fragmentation for product-ion information. The typical workflow includes protein digestion by enzymes, peptide fractionation by chromatography, and mass spectrometer analysis. The raw data from LC-MS experiment generally consist of a list of variables including the scan number (spectra number), the LC retention time, the precursor-ion and product-ion \( m/z \) values, and the corresponding signal intensities. The format of these data, however, is vendor-dependent and not interchangeable (Table 4-1). Recently, several research groups and HUPO Proteomics Standards Initiative (PSI) have developed open XML-based standards, formats and libraries to advocate vendor-agnostic tools and analytical pipelines [20]. R unifies the data through a package, namely mzR [21], by providing an unified programming interface for common file formats including mzXML [22], mzML [23], mzData [24], and netCDF formats. A connection to any of these file types can be established by using the openMSfile function, which enables queries of instrument information and raw data in a consistent way.
Table 4-1. Different file extensions of major vendors of LC-MS/MS systems.

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Extension</th>
<th>File type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermo Scientific</td>
<td>.RAW*</td>
<td>Thermo Xcalibur</td>
</tr>
<tr>
<td>Micromass**</td>
<td></td>
<td>Micromass (Waters) MassLynx</td>
</tr>
<tr>
<td>Waters</td>
<td></td>
<td>PerkinElmer TurboMass</td>
</tr>
<tr>
<td>PerkinElmer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agilent</td>
<td>.D</td>
<td>Instrument data format</td>
</tr>
<tr>
<td>Bruker</td>
<td>.BAF, .FID, .YEP</td>
<td>Instrument data format</td>
</tr>
<tr>
<td>ABI/Sciex</td>
<td>.WIFF</td>
<td>QSTAR and QTRAP file format</td>
</tr>
<tr>
<td>ABI/Sciex</td>
<td>.t2d</td>
<td>4700 and 4800 file format</td>
</tr>
<tr>
<td>Waters</td>
<td>.PKL</td>
<td>MassLyns associated format</td>
</tr>
<tr>
<td>Shimadzu</td>
<td>.qgd</td>
<td>Instrument data format</td>
</tr>
<tr>
<td>Shimadzu</td>
<td>.spc</td>
<td>Library data format</td>
</tr>
</tbody>
</table>

(*) RAW formats of each vendor are not interchangeable
(**) Micromass was acquired by Waters in 1997

4.4 Custom-built Web-based Applications

A web-based application is any application that uses a website as the interface or front-end. In contrast to traditional desktop applications, which are installed on a local computer, users can access the web application easily from any device connected to the internet by using a standard web browser [25]. Web-based applications have become increasingly popular, with the rise of widely available internet connection. To create a web-based application is, however, not simple, as lots of programming experience in HTML and JavaScript is needed. This presents a hurdle in creating web applications, especially for people who are not professional programmers.

To overcome, RStudio released Shiny, which is a means of creating web-based applications entirely in R. The client-server communication, HTML, layout and JavaScript programming is completely handled by Shiny. Given the nature of the R language, it is
possible to program interactive applications with a Shiny Web-browser interface that can be accessed either on the localhost (users’ own device) or on another computer accessed by means of the internet. The R Shiny applications have the potential to provide great convenience and feasibility of data processing to a broader audience who are not necessarily experienced web-developers [26].

4.4.1 The Structure of R Shiny Applications

Each web-based application comes with an URL. By inputting the specific URL of the web application, a reaction from the server is triggered, sending the HTML, CSS, and JavaScript code needed to run the application back to the client. The web application then executes by itself inside the web browser of the client. A web-based application typically consists of a variety of user-interface (UI) elements (e.g., a button, a checkbox, and a drop-down list). Each of these elements is interactive. For example, pushing a button can trigger an action, allowing the execution of a piece of code, which in turn, changes the status of the web application, directing it to retrieve a piece of information from the server or draw a graph within the application. This process is commonly known as event-driven programming; that piece of code executed based on an event is known as event handler.

R Shiny web applications follow this typical structure of web-based applications (Figure 4-1). It is, however, much simpler. Shiny applications are built using two R scripts that communicate with each other (i.e., a user-interface script and a server script). The user-interface script (ui.R) controls layout and appearance; the server script (server.R) contains content codes for user input and output by R language, and functions by using
user-installed packages. RStudio has provided detailed tutorials and training exercises for building Shiny applications [27].

The construction of HTML, CSS, and JavaScript codes is, conveniently, done by using these two R scripts, as is the generation of event handlers. In addition, upon an update of an UI element, all the relevant content code is updated automatically. Therefore, a fully functional web-based application can be completely written in R Shiny.

![Figure 4-1. The structure of an R Shiny web-based application.](image)

4.4.2 Example of an R Shiny Application: Proteomics Toolkit

4.4.2.1 Overview of Proteomics Toolkit

Herein, we discuss the core elements of the ui.R and server.R scripts necessarily needed to create a web-based application, by using a custom-built application, namely, Proteomics Toolkit (https://benniu-wustl.shinyapps.io/proteomics-toolkit/), as an example. This multi-functional application aims at facilitating the mass spectrometry-based proteomics and peptidomics analysis and computation. The functionalities include
monoisotopic m/z calculation with site-specific PTMs, *in silico* mass spectrum isotopic pattern simulation, *in silico* protein digestion, and generation of an inclusion list for mass spectrometry data-dependent-acquisition (targeted MS analysis, see Chapter 7 where we used this kind of MS acquisition mode) [28]. The complete code of Proteomics Toolkit is available online through GitHub (https://github.com/benniu720/Proteomics-Toolkit).

Shown in Figure 4-2 is a screenshot of the user interface panel of Proteomics Toolkit. The two tabs on the top (i.e., Peptide Analysis and Protein Analysis, feature the two categories of utilities of this web application). Within each tab, there are several sub-tabs (e.g., Computations, MS Spectral Simulation, *in silico* Digestion, Targeted m/z), representing the different functionalities this software can achieve. For example, we selected the Computations sub-tab from the tab of Protein Analysis in Figure 4-2. By inputting the protein amino acid sequence on the top-left (in this case, YGGFLHPERTYNNASDKLM), detailed information including the elemental composition (C101 H150 N28 O31 S1), total residue number (20 residues), monoisotopic m/z values at various charge states (from +2 to +6), and the amino acid composition are computed and output automatically. With this amino acid sequence, we can go to the sub-tab of MS Spectral Simulation and calculate the isotopic pattern by clicking the “Simulate” button, specifying a user-defined charge state. Figure 4-3 shows its normalized isotopic pattern in the +3 charge state. The simulation pattern can change accordingly if we choose a different charge state or add a PTM to the molecule (from the panel on the left). Additionally, this application allows generation of *in silico* peptides (Figure 4-4), and calculations of the monoisotopic m/z values of these peptides with and without user-defined PTMs at various charge states.
(Figure 4-5). An output of this peptide list, which can be useful for generating targeted inclusion lists [28], is obtainable by clicking the “Download Targeted List!” button. Instead of going through a series of repetitive calculations and simulations, users are allowed to control and adjust all parameters in Shiny, and, interactively, access the results simultaneously.

**Figure 4-2.** Proteomics Toolkit application. Shown is the display of computed results, under the sub-tab of Computations, based on the amino acid sequence input of the peptide sequence YGGFLHPERTYNNAAASDKLM. The computation includes elemental composition, total residue number, monoisotopic m/z from +2 to +6 charge state, and the amino acid composition.
Figure 4-3. Proteomics Toolkit application under the sub-tab of MS Spectral Simulation. Shown is the interactive display of simulated isotopic pattern of the amino acid sequence YGGFLHPERTYNNAAASDKLM with a charge state of +3. Placing the cursor onto each isotopic peak (in red) displays its accurate $m/z$ value and the normalized intensity.

Figure 4-4. Proteomics Toolkit application under the sub-tab of *in silico* Digestion. Shown is the peptide list from an *in silico* proteolysis, where the users can specify parameters including cleavage site(s), cleaving terminal (C-/N- terminal), number of missed
cleavages, and the minimally allowed peptide length (defaulted as 4). The output is a list with peptide sequence, starting and ending residue number, missed cleavage number, and the length of peptide.

**Figure 4-5.** Proteomics Toolkit application under the sub-tab of Targeted m/z. Shown is a table with the monoisotopic m/z values of the *in silico* generated peptides with and without user-defined PTMs at various charge states (+1 to +5 in this case). The output of this table facilitates the generation of an inclusion list in a targeted MS acquisition mode.

### 4.4.2.2 Structure of ui.R Script and server.R Script of Proteomics Toolkit
The ui.R script of Proteomics Toolkit encodes instructions for the application’s layout, appearance, and interactive widgets (e.g., sliders, selection boxes, buttons), as shown below:

```r
library(shiny)
library(DT)
library(plotly)
shinyUI(fluidPage(
    navbarPage(title=span(strong("Proteomics Toolkit"), style="font-family:'rockwell extra bold';color:slateblue; font-size:20pt"),
        ## The 1st tabPanel is for peptide analysis, see below.
        tabPanel(title=span(strong("Peptide Analysis"),style="font-family:'luminari';color:darkgreen;font-size:13pt"),icon=icon("leaf",lib="font-awesome"),
            sidebarLayout(
                sidebarPanel(
                    textInput("seq.pp", label=div(em("Input peptide sequence"),style="font-family:'marker felt';color:darkgreen; font-size:12pt"), placeholder = "for example: YGGFL"),
                    radioButtons("Cys.pp", label=div(em("Carbamiodomethyl on Cys ?"),style="font-family:'marker felt';color:darkgreen; font-size:12pt"), choices = list("Yes","No"), selected = "Yes"),
                    textInput("ptmFormula.pp", label = div(p(em(span("PTM in formula"),style="font-family:'marker felt';color:darkgreen; font-size:12pt"), p(em(span('Note:',style="font-family:'marker felt';color:darkgreen; font-size:10pt"), span('A "SPACE" is required between every element! The allowed elements are',style="font-family:'marker felt';color:darkgreen; font-size:10pt"), span('C, H, N, O, S, P, Br, Cl, F, Si.', style="font-family:'marker felt';color:darkgreen; font-size:10pt")))))),
                    placeholder = "for example: C5 H8 O2 N Br2"),
                    hr(),
                    div(p(strong(em("Residue-specific PTMs"))), style="font-family:'marker felt';color:darkgreen; font-size:12pt"),
                    fluidRow(column(7,uiOutput("PTM1.pp")),
                        column(4, uiOutput("Resi1.pp")),
                    fluidRow(column(7, uiOutput("PTM2.pp")),
                        column(4, uiOutput("Resi2.pp")),
                    fluidRow(column(7, uiOutput("PTM3.pp")),
                        column(4, uiOutput("Resi3.pp"))),
                    checkboxGroupInput("chkbx.pp",label=div(strong(em("extra PTMs"))),style="font-family:'marker felt';color:darkgreen; font-size:12pt"),choices=list("none","Acetyl","Ammonium-add","Ammonium-loss","Dehydrated"), selected = "none")),
            mainPanel(
                tabsetPanel(type="tabs",
                    tabPanel(title=span(strong("Peptide Computations"), style="font-family:'constantia';color:darkgreen;font-size:10pt"),icon = icon("calculator",lib = "font-awesome"),
                        br(),
                )))
    )
))
```
fluidRow(  
  column(5, div(h4(em(p("Peptide Sequence"))), 
    style="font-family:'hannotate tc';color:darkgreen")),
  column(5, div(h4(em(p("Elemental Composition"))), 
    style="font-family:'hannotate tc';color:darkgreen")),
  column(2, div(h4(em(p("Res.#"))), style="font-
    family:'hannotate tc';color:darkgreen"))
  ),
fluidRow(  
  column(5, verbatimTextOutput("text1.pp")),
  column(5, verbatimTextOutput("text2.pp")),
  column(2, verbatimTextOutput("text3.pp"))
  ),
br(),
tableOutput("tb1.pp"),
hr(),
fluidRow(  
  column(3, actionButton("act.pp", label=span(strong("AA 
    Composition"), style="font-
    family:'wawati tc';color:darkgreen; font-
    size:12pt"), icon = icon("cubes",lib="font-
    awesome"))),
  column(8,DT::dataTableOutput("tb2.pp"), offset = 3)
  ),
  tabPanel(span(strong("MS Spectral Simulation"),style="font-
    family:'constantia';color:darkgreen;font-
    size:10pt"), icon=icon("laptop",lib= "font-
    awesome"),
    br(),
div(h4(em(p("Formula (with PTMs)"))), style="font-
    family:'hannotate tc';color:darkgreen"),
    fluidRow(  
      column(6, verbatimTextOutput("text4.pp"))
    ),
    hr(),
    fluidRow(  
      column(4, div(h4(p(em("Charge State to Simulate: "))),
        style="font-family:'hannotate tc';color:darkgreen")),
      column(2, numericInput("ecs.pp", label=NULL,
        value=1,width='80%', min=1, step=1)),
      column(2, actionButton("sim.pp", label =
        span(strong("Simulate"), style="font-family:'wawati tc';color:darkgreen;
        font-size:12pt"), icon=icon("bar-chart",lib="font-
        awesome")))
    ),
    hr(),
    plotlyOutput("Idplot.pp")
  )
)
)
## The 2nd tabPanel is for the protein analysis, see below.
  tabPanel(title=span(strong("Protein Analysis"), style="font-
    family:'luminari';color:purple; font-size:13pt"), icon =
    icon("heartbeat",lib="font-awesome"),
    sidebarLayout(  
      sidebarPanel(  
        textInput("seq", label=div(em("Input protein sequence"),style="font-
          family:'marker felt';color:purple; font-size:12pt")),
      )
    )
  )
)
radioButtons("Cys", label=div(em("Carbamiodomethyl on Cys ?"), style="font-family:'marker felt';color:purple; font-size:12pt"), choices = list("Yes","No"), selected = "Yes"),
fluidRow(
  column(7, textField("ptmFormula", label = div(p(em(span("fixed PTM formula"), style="font-family:'marker felt';color:purple; font-size:12pt")))))
)
)
hr()
div(p(strong(em("fixed Residue-specific common PTMs")))))
)
)

tabPanel(type="tabs",
  tabPanel(span(strong("Computations"), style="font-family:'constantia';color:purple;font-size:10pt"), icon = icon("calculator", lib = "font-awesome")),
  fluidRow(
    column(5, div(h4(em(p("Protein Sequence")))), style="font-family:'hannotate tc';color:purple")),
    column(5, div(h4(em(p("Elemental Composition")))))
  )
)
fluidRow(
  column(5, verbatimTextOutput("text1")),
  column(5, verbatimTextOutput("text2")),
  column(2, verbatimTextOutput("text3"))
)

)

)

)

)

)

tableOutput("tb1")

)

)

)

)
column(8, DT::dataTableOutput("tb2"), offset = 3)
)

),

tabPanel(span(strong("MS Spectral Simulation"), style="font-family:'constantia';color:purple;font-size:10pt"), icon = icon("laptop", lib = "font-awesome"),

br(),

div(h4(em(p("Formula (with PTMs)")))), style="font-family:'hannotate tc';color:purple"),

fluidRow(

column(6, verbatimTextOutput("text4"))
),

hr(),

fluidRow(

column(4, div(h4(p(em("Charge State to Simulate:"))), style="font-family:'hannotate tc';color:purple"),

column(2, numericInput("ecs", label=NULL, value=1,width='80%',

min=1, step=1)),

column(2, actionButton("sim", label = span(strong("Simulate"),

style="font-family:'wawati tc';color:purple; font-size:12pt"),

icon=icon("bar-chart",lib="font-awesome")))
),

hr(),

plotlyOutput("Idplot")
),

}


tabPanel(span(strong(em("in silico")),

strong("Digestion"), style="font-family:'constantia';color:purple;font-size:10pt"), icon = icon("align-right", lib = "font-awesome"),

br(),

fluidRow(

column(3, textInput("cleavSite",

column(2, numericInput("mc", 

label=div(h4(em("Missed"))), style="font-family:'hannotate tc';color:purple;font-size:12pt"),

value=1,min=0, step=1, width = "80%")),

column(3, numericInput("thresh",

label=div(h4(em("Threshold"))), style="font-family:'hannotate tc';color:purple; font-size:12pt"), value=4,min=0, step=1,width = "50%"))
),

actionButton("insilico", label=span(strong("in silico Digest"), style="font-family:'wawati tc';color:purple; font-size:12pt"), icon = icon("sort-amount-desc", lib="font-awesome")),

hr(),

DT::dataTableOutput("tb3")
),

)


tabPanel(span(strong("Targeted m/z"), style="font-family:'constantia';color:purple;font-size:10pt"), icon = icon("map-marker",lib = "font-awesome"),

br(),

fluidRow(

column(3, div(h4(em("Charge from"))), style="font-family:'hannotate tc';color:purple; font-size:12pt", align="center")),
```r

column(2, numericInput("csfrom", label=NULL, value=1, min=1, step=1, width = "80%")),
column(1, div(h4(em("to")), style="font-family:'hannotate tc';color:purple; font-size:12pt")),
column(2, numericInput("csto", label=NULL, value=5, min=2, step=1, width="80%"))
)

fluidRow(
  column(6, selectInput("target.ptm", label =
    div(h4(em("common Variable PTMs to include")),style="font-family:'hannotate tc';color:purple; font-size:14pt"),
    choices=list("none","Oxidation","diOxidation","triOxidation","Carbonyl","Deamidation(N)","Phosphorylation(S,T,Y)"),selected="none",selectize = TRUE, multiple = TRUE)),
  column(3, numericInput("lowmz", label=span(h4(em("Lowest m/z"))), value=150, min=50, max=600, step=1)),
  column(3, numericInput("highmz", label=span(h4(em("Highest m/z"))), value=4000, min=3000, max=5000, step=1))
),

fluidRow(
  column(3, textInput("NM", label=div(em("PTM name"), style="font-family:'hannote tc';color:purple; font-size:12pt"), placeholder="e.g., FMe")),
  column(5, textInput("Fml", label=div(em("PTM formula"), style="font-family:'hannote tc';color:purple; font-size:12pt"), placeholder="e.g., C16 H26")),
  column(3, selectInput("allRes", label=div(em("PTM targets"), style="font-family:'hannote tc';color:purple;font-size:12pt"), choices=strsplit("ACDEFGHILMNPQRSTVWY", split="")[[1]], multiple=TRUE))
),

fluidRow(
  column(4,actionButton("targeted", label=span(strong("Show Targeted List"), style="font-family:'wawati tc';color:purple; font-size:12pt"),icon=icon("tags",lib="font-awesome"))),
  column(4, downloadButton("dl.target", label=span(strong(em("Download Targeted List!")), style="font-family:'wawati tc';color:purple; font-size:12pt")))
),
hr(),
DT::dataTableOutput("tb4")
)
```

120
These content codes of user interface are required to stay within the brackets of layout functions. For example, layout function fluidPage creates a canvas for the interface and utilizes tabPanel function to position the two tabs (i.e., Peptide Analysis, Protein Analysis) on top of the page (Figure 4-2). The position and appearance of user-input widgets and output results (e.g., table, graph, list) are also specified by different layout functions. Although each layout function has its own framework and variables, they all follow the same hierarchical structure where element functions are embraced within a positioning function, which is within a layout function.

The server.R script (see complete code https://github.com/benniu720/Proteomics-Toolkit), on the other hand, plays a vital role in defining instructions for the application while maintaining the application speed by minimizing redundant computations. The entire server.R script is essentially a shinyServer function that incorporates an input object and an output object. It generally involves creating functions that map user inputs to various kinds of output. Objects that change depending on input from widgets in the ui.R are termed “reactive” (e.g., changes in amino-acid sequence input). Each time the input from a widget changes, the value for the reactive object updates to reflect this change. For example, each time the amino acid sequence input changes, the output including the sequence, elemental composition, total residue number, monoisotopic m/z, and amino acid composition change accordingly.

4.4.3 R Shiny Application: HDX Easy Plotting
We present here another R Shiny web-based application, HDX Easy Plotting (URL: https://benniu-wustl.shinyapps.io/hdexaminer-assistant/), which was developed to simplify the plotting of HDX kinetic curves from HDX-MS experiments (Figure 4-6). The raw data obtained from HDX experiments, typically, are submitted to the processing software such as HDExaminer and HD WorkBench, for the automatic quantitation of deuterium uptake of each identified peptide. The output of quantitation results, however, is a massive data table that needs to be manually processed to obtain the HDX kinetic curves (Figure 4-7). The application we developed here is adaptive to the output of HDExaminer, by taking the data table exported from HDExaminer, then reorganizes, computes, and plots the time-dependent curves automatically (Figure 4-8). HDX Easy Plotting allows user-specified plotting parameters (e.g., layout size, line transparency, line color, point size) and supports output of HDX kinetic curves and downloading of formatted data table for further data processing.

The complete code and a short tutorial for HDX Easy Plotting have been uploaded to GitHub (https://github.com/benniu720/HDExaminer-Assistant).
Figure 4-6. The screenshot of HDX Easy Plotting application. Output from HDExaminer software (*.csv format) can be directly uploaded and converted to a table of peptides, kinetic curves, with user-defined experimental and display parameters.

Figure 4-7. Typical workflow of data processing of HDX-MS experiments (yellow arrows). The raw data are submitted to HDExaminer software for quantitation of deuterium uptake for each peptic peptide, followed by an output of a massive data table, from which the data needed for plotting HDX kinetic curves are drawn manually. To reduce the redundant, repetitive work, we developed the HDX Easy Plotting application, which can automatically generate kinetic curves by taking the output from HDExaminer software directly (blue arrows).
Figure 4-8. An example of the HDX kinetic curves generated by using HDX Easy Plotting application. Shown are 9 kinetic curves corresponding to different peptides.

4.4.4 R Shiny Application: MS/MS Spectrum Analysis & Assignment

Mass spectrometry-based proteomics methods largely rely on product-ion (MS/MS) spectra to provide sequence information of peptide [29, 30]. One should, however, scrutinize each product-ion spectrum for accurate identification of product-ion species (e.g., y-ion and b-ion). The identification and annotation of fragment ions in MS/MS can be difficult and laborious, given the large number of peaks in each spectrum, the different identity and location of PTMs, and possible neutral losses. We developed the R Shiny
application, namely, MS/MS Spectrum Analysis & Assignment, to automatically identify and annotate species from any product-ion (MS/MS) spectrum from a bottom-up proteomic experiment. The URL of this application is https://benniu-wustl.shinyapps.io/ms2-autoassign/.

The application of MS/MS Spectrum Analysis & Assignment contains two modules: in silico Fragments, which generates a table of daughter-ion based on user input of amino acid sequence and site-specific modification (Figure 4-9); Peaks Matching, which takes the mass spectrometry data file (in *.mzXML format), displays spectral information with a given scan number (e.g., retention time, precursor m/z, charge state) (Figure 4-10), and annotates the spectrum with user-specified parameters (Figure 4-11). Within the first module, a new table of product ions can be generated by checking the Neutral Loss Search, which allows the input of neutral species being lost and recalculation of the m/z values of y-/b-series ions. In Chapter 7, we utilized this functionality to identify successfully the y-series fragment ions of a peptide with a C-terminal farnesylation PTM, which is typically manifest as a neutral loss of 204 Da (corresponding to $\text{C}_{15}\text{H}_{24}$) [31].
Figure 4-9. A screenshot of MS/MS Spectrum Analysis & Assignment application. Shown is the *in silico* fragments module that computes m/z of the b-/y-series ions, based on the input of an amino acid sequence, location of the modification, formula of the modification, neutral loss identity, etc.

Figure 4-10. By providing a spectrum scan number, the module of Peak Matching can display the corresponding spectral information including retention time, m/z value and charge state of precursor ion.
Figure 4-11. An example of annotated product-ion (MS/MS) spectrum using MS/MS Spectrum Analysis & Assignment application.

4.5 Conclusion and Future Work

R with its many built-in functions and packages is suitable for mass spectrometry-based proteomic analysis. Specifically, R Shiny package greatly facilitates the interactive data analysis and visualization by means of web browsers, allows developments of analytical toolkits that precisely fit users’ needs, and provides conveniences in maintenance and updating. Demonstrated in this chapter are three web-based applications focused on mass spectrometry-based computations and data processing. There are, however, more applications under development in this field, for the purpose of faster, more convenient, and more accurate analysis. For example, there is a need for developing an analytical model to benchmark the curvature and overlapping extent of HDX kinetic curves, from which important binding interfaces are inferred; however, the interpretations of curves by
naked-eye can be different from individual to individual, lab to lab. We are currently developing an R-based model to classify the time-dependent curves from HDX experiments with machine-learning algorithms. The goal is to predict accurately the important binding interfaces of an unknown complex, given a set of HDX kinetic data. The workflow includes the training and optimization of model using sets of HDX data of a ligand-binding protein whose binding interfaces are known; optimized model is then applied to unknown HDX datasets for curve analysis and prediction (Figure 4-12).

**Figure 4-12.** Workflow for prediction of important binding interfaces from a set of HDX kinetic curves. Models with different machine learning algorithms are trained by incorporating sets of HDX data with known binding interfaces, and tested with cross validation. The best model (the one showing the highest prediction accuracy) is then used for the prediction of important binding interfaces for new HDX data set of unknown binding systems.
References

Chapter 5

Fast Photochemical Oxidation of Proteins (FPOP) Locates the Unfolding of ALS-inducing Mutant Variant G93A SOD1 Caused by Destabilization with Trifluoroethanol
5.1 Abstract

Destabilization of Cu, Zn superoxide dismutase (SOD1) mutants may lead to protein aggregation and ultimately amyotrophic lateral sclerosis (ALS). The mechanism and protein conformational changes during this process, however, remain unclear. To study SOD1 at the molecular level and in solution, destabilization coupled with structural studies are needed. Here we destabilized a mutant variant of human SOD1 (G93A) with trifluoroethanol (TFE) and used both top-down native ESI and Fast Photochemical Oxidation of Proteins (FPOP) to characterize the structural changes of G93A SOD1 induced by destabilization. The structural changes upon destabilization include dissociation of SOD1 dimer and local partial unfolding. The use of TFE followed by protein footprinting and native MS should permit continuing inquiries into the high propensity of aggregation of the many mutant SOD1 proteins.

5.2 Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease that specifically targets the motor neurons in the spinal cord, brain stem, and cortex [1]. It is the most common motor neuron disease, with a lifetime risk of approximately 1 in 2000 [2]. Currently there is no cure or effective treatment for this disease, commonly known as Lou Gehrig’s disease in the United States [3]. Approximately 90% of ALS have no known cause and are termed sporadic ALS (sALS). For the remaining 10%, there is a familial history (familial ALS or fALS). Among these fALS cases, about 25% are associated with the CuZnSOD (SOD1) gene on chromosome 21 where numerous site mutations are currently known as ALS-causing mutations [4].
SOD1 is an antioxidant enzyme found in the nucleus, peroxisomes, mitochondrial intermembrane space of eukaryotic cells and in the periplasmic space of bacteria [5]. Its function is to catalyze the dismutation of superoxide anion into H₂O₂ and O₂ [6]. The human enzyme is a 32-kDa homodimer, with one copper and one zinc binding per 153-residue subunit. The X-ray crystal structures of SOD1 from many species have been solved, mostly in the fully metallated state, showing that the structures are highly conserved [7]. Each SOD1 monomer is built upon a β-barrel motif comprising eight anti-parallel β sheets and possesses large functionally important electrostatic and zinc loops (Figure 5-1).

Loop IV or the Zn-binding loop (ZBL, amino acids 49-83) contains His63, His71, and His80 as well as D83 that coordinate Zn²⁺ (Figure 5-1). That binding along with the disulfide formation between Cys57 and Cys146, contribute to the structural stability of SOD1 [8]. The Cu ion, coordinated by His46, 48, 120, and the bridging His63 between Zn and Cu is the heart of the enzymatic activity of SOD1 to catalyze the disproportionation (Eq 5-1) [9, 10]. The two-step process of reduction of cupric ion to form dioxygen (Eq 5-2), and reoxidation of cuprous ion to form hydrogen peroxide (Eq 3). Those reactions (Eq 5-2 and 5-3) have near diffusion-limited rates at physiological pH, and the activity is nearly pH-independent over a wide range [7].

\[
2O_2^- + 2H^+ \rightarrow O_2 + H_2O_2 \quad \text{Eq 5-1}
\]

\[
O_2^- + Cu^{2+}ZnSOD \rightarrow O_2 + Cu^{+}ZnSOD \quad \text{Eq 5-2}
\]

\[
O_2^- + 2H^+ + Cu^{+}ZnSOD \rightarrow H_2O_2 + Cu^{2+}ZnSOD \quad \text{Eq 5-3}
\]
Despite the normal catalytic function of SOD1, multiple studies of both transgenic mice and humans provide convincing evidence for the gain-of-function toxicity of SOD1 upon mutation [11, 12]. Proteinaceous inclusions rich in mutant SOD1 are in tissues from ALS-SOD1 transgenic mice and ALS patients, and in cell-culture models, supporting the hypothesis that SOD1-related fALS is a disorder caused by unfolding and aggregation [13, 14]. Although over 150 different site mutations in the SOD1 gene are linked to fALS [15], it remains unclear how this large distribution of mutations results in the same disease. These mutations, throughout the entire SOD1 sequence, are responsible for dramatically different catalytic activities and metal-binding affinities. Depending on location, these mutations fall into two types: metal-binding-region (MBR) mutations and wild-type-like (WTL) mutations. The former directly compromise the protein’s ability to bind metals, reducing the catalytic activity of SOD1; the latter maintain the metal-binding capability, and these proteins have activity levels similar to that of the wild-type SOD1 [16]. This further confirms the gain-of-function toxicity of mutant SOD1.

Although the toxic function acquired by mutant SOD1 is well-acknowledged, the molecular mechanism and the species responsible for this toxicity are unsettled [1, 4]. As for other neurodegenerative diseases (e.g., Alzheimer’s disease (AD), Parkinson’s disease (PD), Huntington’s disease (HD) [17-19]), the fibrils or insoluble inclusions observed in ALS are unlikely to be toxic because they are formed in a relatively late stage of the disease [20]. Controversy still exists, however, regarding the identity of the cytotoxic species [2, 4, 15]. Clearly monomeric SOD1 plays an important role in its aggregation. For example, SOD1 aggregates found in spinal cords extracted from human A4V SOD1
patients and from various ALS mouse models react with antibodies that specifically recognize the monomer [11, 21]. Further, the monomeric, reduced apo-SOD1 can initiate aggregation of dimeric holo-SOD1 at physiologic conditions [22].

One way to connect the monomer and the aggregates is the misfolding and aggregation hypothesis, which involves a three-state equilibrium (Eq 5-4) among native dimeric SOD1 (D), monomeric SOD1 (M), and unfolded or partially unfolded SOD1 (U) [23-25]. This hypothesis has prompted a variety of biophysical studies of the effects of the mutations on the stability and folding mechanism of SOD1 [26-29].

\[ 2U \rightleftharpoons 2M \rightleftharpoons D \]  \quad \text{Eq 5-4}

Matthews et al. [30] determined the effects of ALS-inducing mutations on the population of folded and unfolded monomeric species (M or U), by perturbing with urea and following with a global stability analysis with far-UV circular dichroism (CD). Quantification of stability changes induced by mutations shows a 2- to 10-fold increase in the population of folded monomeric states and an 80- to 480-fold increase in the population of the unfolded monomeric states, as compared to those formed with wild-type SOD1. This strongly suggests that the disease-provoking property of ALS mutations is likely related to the conformational equilibrium [31]. Other evidence suggests that the ALS-inducing SOD1 mutations manifest a common shift of the folding equilibrium towards denatured and partially unfolded apo monomers by either disrupting the dimer interface, perturbing the monomer structure, or compromising the formation of mature (e.g., holo, disulfide-bonded) dimers [32-35]. This common pattern implicates that the seed for SOD1
cytotoxicity is the immature form of SOD1, which may have increased propensity to self-associate through high-order nucleation reactions [36].

ALS mutations convey the most prominent effect on the immature form of SOD1, destabilizing the metal-free and disulfide-reduced polypeptide and promoting unfolding at physiological temperature; these destabilized states readily form oligomers under physiological concentrations [37]. Indeed, destabilizing the protein can be done in a number of ways, including perturbing the dimer interface [33], adding trifluoroethanol (TFE) [27, 38], reducing disulfide bonds [22], removing metal-ions [30, 39], altering the post-translational modifications (PTMs) [40], or simply heating [27, 41]. Recent studies under destabilizing solution conditions (i.e., upon addition of TFE or increase of temperature) indicate that fALS mutations decrease SOD1 stability and increase the rates of unfolding, forming ALS-like aggregates in vitro with various morphologies by populating the aggregation-prone monomeric species [27] (Figure 2). Therefore, destabilized fALS-associated SOD1 mutants can trigger an equilibrium shift, increasing aggregate formation [27, 37]. The mechanism and molecular basis, however, for shifts in destabilization and protein conformational changes during this process remain unclear.

The TFE-induced SOD1 aggregates, in particular, show greatest resemblance to in vivo ALS aggregates in terms of the tinctorial properties [42], thus we used addition of TFE to destabilize the G93A SOD1 mutant variant, which is among the most extensively studied ALS-inducing mutation as it was the first mutation to be modeled in mice [13]. As described in previous chapters, mass spectrometry is evolving as an alternative tool to study protein high order structures (HOS), with fast turnaround, low sample consumption,
and near-native solution environments [43, 44]. Herein, we adopted both top-down and bottom-up MS approaches to uncover the effect of TFE destabilization of the SOD1 protein, to localize the conformational changes in molecular level, and to shed some light upon the mechanism of aggregates formation.

**Figure 5-1.** Crystal structure of human SOD1 (PDB code 1N18), a native dimer. Shown is only one monomeric subunit, which comprises two major function loops (i.e., electrostatic loop and zinc binding loop) and one β-barrel formed by 8 β-sheet strands.
5.3 Experimental Section

5.3.1 Materials and Methods

All experiments were performed using recombinant apo-SOD1 protein expressed and purified from BL21-Gold(DE3) PLysS E. coli cells (Strategene, Inc., Cedar Creek, TX, USA), as previously described [26]. The two free cysteines Cys6 and Cys111 were
mutated to Ala and Ser, respectively, to avoid intermolecular disulfide scrambling; the glycine to alanine mutation (G93A) was introduced into this pseudo WT background [30].

L-Histidine, L-methionine, 30% hydrogen peroxide, leucine enkephalin acetate hydrate, catalase, trifluoroethanol (TFE), urea, dithiothreitol (DTT), iodoacetamide (IAM), ammonium acetate, formic acid (FA), trifluoroacetic acid (TFA), HPLC-grade solvent, and phosphate buffer saline (PBS) were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). Sequencing-grade trypsin was purchased from Promega Co. (Madison, WI).

5.3.2 Native ESI Measurement of the Intact Protein

A freshly prepared G93A SOD1 protein stock solution (40 µM protein, in 10 mM potassium phosphate, pH 7.2) was stored in 4 °C refrigerator. An aliquot of the stock solution was submitted to buffer exchange by using Vivaspin® 500 (Sartorius, Goettingen, Germany); the desired buffer solution is 200 mM ammonium acetate. After buffer exchange, the resulting solution was aliquoted into two portions; to destabilize the protein, 18% (v/v) TFE was added to one portion. The final concentration of G93A was 5 µM. Samples were subjected to a 1-h incubation at room temperature before submitting the protein to a top-down measurement. To perform top-down native ESI analysis, 5 µL of the protein sample was loaded onto a Borosilicate Emitter with extra coating (Thermo Fisher Scientific, Waltham, MA); the emitter was interfaced with a Thermo Exactive Plus EMR mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The capillary voltage was 1.0-1.3 kV, and the source temperature was 37 °C. The mass spectrometer was operated in full-scan mode with a mass resolving power of 70,000.

5.3.3 FPOP Footprinting
A stock solution of G93A SOD1 was aliquoted, with one aliquot treated with 18% (v/v) TFE for destabilization, and another aliquot untreated. Both TFE-treated and untreated samples were subjected to a 1-h incubation at room temperature, then immediately buffer-exchanged to 10 mM PBS. To perform FPOP, the total volume of 50 µL of solution containing 5 µM G93A SOD1, 8 µM leucine enkephalin (reporter peptide), 20 mM hydrogen peroxide, and histidine as radical scavenger, was loaded onto a 100 µL syringe (Hamilton, Reno, NV) and advanced using a syringe pump (Harvard Apparatus, Holliston, MA). The reporter-incorporated FPOP was conducted as described previously [45]. The excimer laser (GAM Laser Inc., Orlando, FL) was adjusted to 25 mJ/pulse at a frequency of 7.4 Hz and focused to an irradiation window of 2.7 mm. The solution flow rate was adjusted to ensure a 25% exclusion volume fraction to avoid exposing a “plug” of solution to more than one set of radicals [46, 47]. Control samples were handled in the same manner without laser irradiation. To adjust different labeling times and measure the time-dependent response of TFE-untreated G93A SOD1, various histidine concentrations (0, 0.2, 2.0, 40 mM) were used in FPOP. At each histidine concentration, three replicates were drawn from the stock solutions for FPOP labeling. All TFE-treated G93A samples were footprinted at a histidine concentration of 0.2 mM. After labeling, samples were submitted to catalase and methionine treatment to remove any left-over H₂O₂, aliquoted, and submitted to global analysis and bottom-up analysis.

5.3.4 Global Analysis

The protein-level FPOP modification can be determined by using a MaXis 4G Q-TOF (Bruker, Bremen, Germany) coupled to a custom-built, online sample-handling device. In
brief, 10 µL of footprinted G93A SOD1 (~50 pmol) was loaded onto a C8 reversed-phase trapping column (ZORBAX Eclipse XDB, 2 × 15 mm; Agilent Technologies, Santa Clara, CA) for a three-minute desalting with 200 µL/min water containing 0.1% (by volume) TFA. The desalted protein was then eluted from the trapping column with a 10-min linear gradient from 2% to 95% solvent B (acetonitrile with 0.1% FA) at 200 µL/min.

5.3.5 Bottom-up Analysis

All samples were submitted to acetone precipitation for clean-up, and then resuspended with 8 M urea, treated with DTT (10 mM) and IAM (20 mM) sequentially, and diluted with 10 mM PBS buffer (pH 7.5) prior to trypsin digestion. DTT and IAM were used to break disulfide bonds and cap the free cysteines. Samples were then incubated with 20:1 protein: enzyme (by weight) for 12 h (37 °C, 350 rpm). The digestion was quenched by adding 1% (by volume) FA. After digestion, 20 µL digests from each sample were loaded in autosampler vial for bottom-up analysis. The LC-MS/MS analyses were conducted back-to-back with the UltiMate 3000 Nano LC system and a Thermo Q Exactive Plus Hybrid Quadruple-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Aliquots of 5 µL digests were drawn from autosampler vial and loaded onto a C18 reversed-phase trapping column (Acclaim® PepMap100, 100 µm × 2 cm, 5 µm, 100 Å; Thermo Fisher Scientific, Waltham, MA) at 4.5 µL/min with water in 0.1% FA for a 10-min desalting. Peptides were separated using a custom-packed C18 reversed-phase column (Magic, 100 µm × 160 mm, 5 µm, 200 Å; Michrom Bioresource Inc., Auburn, CA) and eluted with an 80-min gradient delivered at 500 nL/min. The mass spectral scan range of precursor ions was set from m/z 380 to 2200 for all samples with high mass resolving
power of 70,000; the mass spectrometer was operated in non-targeted data-dependent-acquisition (DDA) mode (i.e., 18 most abundant precursor ions eluting per scan were subjected to the high-energy C-trap dissociation (HCD) in the HCD cell for product-ion spectra). Dynamic exclusion was activated for 8 s following each scan to enable selection and fragmentation of low abundance peptides. The maximum injection time (IT) for full MS scan was set to 200 ms, and 150 ms for each MS/MS scan.

5.3.6 Data Analysis

The *.raw MS files were searched directly using Byonic (Protein Metrics Inc., San Carlos, CA) against a custom-built database containing the sequence of G93A SOD1 [48]. All known HO• side-chain reaction products were added to the variable modification database for searching for HO•-modified samples [49]. The alkylation of sample with IAM, which adds a carbamidomethyl group (MW = 57.0214 Da) to Cys-containing peptides, was added as fixed modification for searching. The searching tolerance window was 15 ppm for precursor ions, and 50 ppm for product ions. Residues Lys and Arg were selected as fully specific cleavage sites as is the rule for tryptic proteolysis. The quantitation of fraction modified was based on ion abundances from extracted ion chromatograms (XICs). The fractions modified at peptide and residue level were calculated as follows. Time-dependent signals from LC-MS for each unmodified peptide and its modified species were extracted from raw data files by using Xcalibur™ Software (Thermo Fisher Scientific, Waltham, MA) with a mass tolerance of 15 ppm. The peak integrals of both unmodified species ($I_{unMod}$) and modified species ($I_{Mod}$) were readily attainable. The fraction modified for each peptide was determined according to Eq 5-4:
\[ Fraction\ modified = \frac{\Sigma I_{\text{Mod}}}{\Sigma I_{\text{Mod} + \text{unMod}}} \times 100\% \]  
(Eq 5-4)

5.4 Results and Discussion

5.4.1 Native MS Reveals Conformational Changes of SOD1

Native mass spectrometry (sometimes referred to as native electrospray ionization (ESI)) is a soft ionization method extensively used for generation of gas-phase ions of macromolecules [50] for detection by mass spectrometers. Those charged analytes produced by ESI can either be analyzed as an ensemble of molecular ions (typically \( M + H^+ \)), or fragmented by activation in the gas phase, both of which provide important insights on the structural properties of the molecules [51]. In the scenario of analyzing protein complexes (e.g., soluble protein aggregates), one seeks to introduce the charged analytes into the mass spectrometer under mild conditions so that the non-covalent interactions within the complexes can be retained, and the stoichiometry established. Typical ESI spraying conditions (partially organic, acidic solutions with high temperature), however, would denature/unfold the proteins, leaving their basic side chains exposed to solvent, thereby causing a loss of stoichiometric information [52]. In contrast, native ESI (native MS) preserves the native state of proteins by spraying them from aqueous solution containing a volatile salt (e.g., ammonium acetate) under low temperature and neutral pH [53, 54]. As a consequence, the protein, prior to spraying, is “native”, and ESI will introduce the protein, at least for a short time, in its native state (some refer to this as kinetically trapping the native state in the gas phase). The result will render protein ions carrying fewer charges compared to regular ESI, owing to the fewer basic residues
exposed in the folded form. Those ions will manifest themselves as a narrower spread of a charge-distribution envelop than seen in denaturing ESI.

Similar to many of other SOD1 mutant variants, G93A SOD1 is natively a dimer [55]. The native ESI spectrum of G93A SOD1 (Figure 5-3 A) shows three peaks, corresponding to the differently charged dimeric G93A species (i.e., +10, +11, and +12 charge states). We saw no monomeric or higher oligomeric states of G93A, indicating convincingly that G93A SOD1 is predominantly dimeric in the absence of TFE. The native ESI spectrum upon addition of TFE, however, looked dramatically differently (Figure 5-3 B). The expanded charge-state distribution centered at low m/z (therefore, higher charge states) suggests a large increase of protein surface area, exposing more residues that are solvent inaccessible in the native state. Upon deconvolution, we found that these peaks all refer to the monomeric G93A species (Figure 5-3 B). Therefore, we conclude that, by adding TFE, the dimeric G93A SOD1 is destabilized and dissociates, forming monomeric species. To reveal the extent of unfolding or partial unfolding of G93A SOD1 as a result of destabilization, however, a more detailed picture of the monomeric species formed upon dimer dissociation is needed, as illustrated by footprinting and bottom-up analysis that provide localized structural information.
Figure 5-3. Native ESI spectra of G93A SOD1. (A) Without addition of TFE, G93A SOD1 is predominantly a dimer. The observed three peaks correspond to the +10, +11, and +12 charged dimeric G93A; (B) Under the same spraying condition, however, with the addition of TFE, the dimer peaks diminish, all major peaks correspond to monomeric G93A species. The expanded charge state distribution of monomers indicates an increase of solvent-accessible area.

5.4.2 Differential Scavenging of HO• in SOD1 Footprinting

The first obstacle in footprinting SOD1 is to compensate for various radical scavenging reactions that compete with the labeling of the protein itself. In FPOP, hydroxyl radicals
not only react with the protein and radical scavengers (e.g., histidine, phenylalanine, glutamine) [56], but also with other species in the solution including ligands, buffer components, detergents, and, in this case, TFE, which is known to show high reactivity with hydroxyl radicals [57]. The presence of these scavengers, some adventitious, affects the lifetime of hydroxyl radicals, giving rise to unpredictable and misleading extents of modifications. We incorporated leucine enkephalin as a reporter peptide to disentangle the complication of differential scavenging by allowing us to characterize the footprinting (modification extent) at various labeling times (discussed in Chapter 3) [45].

To minimize the scavenging of radicals by TFE and maximize the extent of modifications on the protein, we performed buffer exchange after the destabilization of SOD1 and immediately prior to FPOP footprinting. The MS analyses of intact, FPOP-labeled (using 0.2 mM histidine as radical scavenger) G93A SOD1 show that, without buffer exchange, the protein was poorly labeled in the presence of TFE (Figure 5-4 A). The yield of FPOP under identical conditions increases significantly by performing buffer exchange prior to labeling with 10 mM PBS (Figure 5-4 B). Correspondingly, the reporter peptide fraction modified also increases, from 11% to 25% after removal of the TFE in the solution (Figure 5-4 D). This clearly demonstrates the radical-scavenging role of TFE and the necessity to perform buffer exchange to obtain quantifiable levels of protein modifications. A comparison of reporter peptide fraction modified between the TFE-treated and untreated samples, both of which undergoing buffer exchange prior to FPOP labeling with 0.2 mM histidine (Figure 5-4 D), shows that the reporter in TFE-treated samples had less modifications (25% vs. 38%), indicating a shorter labeling time for the TFE-treated
samples despite buffer exchange. Interestingly, the FPOP modification extent of G93A SOD1 increases at this labeling time compared to that of the TFE-untreated sample (Figure 5-4 B, C, and Figure 5-5 F). This indicates that TFE treatment causes G93A SOD1 to become more solvent-exposed. This phenomenon also implicates that, a direct comparison regarding the FPOP modification extent between TFE-treated and untreated G93A SOD1 will still be biased, even after buffer exchange as there is likely residual TFE in the protein solution, giving different labeling times for apparently the same conditions. To secure a more unbiased comparison, we measured the time-dependent responses of TFE-untreated G93A SOD1, and used the modification extent of reporter peptide to compensate for the differential scavenging.

**Figure 5-4.** Mass spectra of intact, FPOP-labeled G93A SOD1 (+15 charge state) with 0.2 mM histidine as scavenger. Labeling time of each condition is different owing to the presence of TFE, as reported by different fraction modified of reporter peptide. (A) The presence of TFE scavenges hydroxyl radicals, giving low yield of FPOP modifications; (B)
The yield of FPOP modifications increases with removal of TFE by buffer exchange; (C) TFE-untreated sample shows noticeably lower FPOP modifications, indicating a more solvent-accessible structure of G93A SOD1 with TFE treatment; (D) Reporter peptide fraction modified of abovementioned three conditions. The error bars correspond to the \( \pm \) SE of triplicate measurements.

5.4.3 Reporter-peptide Enables Measurements of Footprinting Kinetics

A time-dependent measurement is possible by varying the FPOP labeling times by scavenging the radicals by using different concentrations of histidine. As the concentration of scavenger (His) increases, the radical lifetime becomes shorter; therefore, there will be less modifications on both the reporter peptide and the protein. Such time-dependent modification extents can be readily revealed in protein-level measurements (shown in Figure 5-5 A-E are the mass spectra of intact, TFE-untreated G93A SOD1 (+15 charge state), footprinted by FPOP at different times controlled by using various histidine concentrations (i.e., 0, 0.2, 2.0, 40 mM)). We quantified the modification extent of G93A SOD1 under each condition, and plotted a time-dependent curve (Figure 5-5F). The fraction unmodified of TFE-untreated protein decreases as the time for labeling is increased by lowering the concentration of histidine; the near-linear correlation indicates that the modifications occur normally (Figure 5-5 F). In contrast, the point representing the TFE-treated G93A SOD1 (blue triangle in Figure 5-5 F) is clearly an outlier, with a fraction unmodified at protein level of 29% and reporter fraction modified of 25%. Its deviation from the line indicates that TFE-treated G93A SOD1 has changed conformation [45].
To examine the local conformational changes induced by TFE, we first submit the protein to footprinting followed by proteolysis, compare the FPOP outcomes between TFE-treated and untreated samples, and quantify the extent of modification for both reporter peptide and the G93A digests, at the same labeling times. By plotting the fraction modified of reporter peptide at each histidine concentration versus the corresponding fraction modified of tryptic peptides from the protein digest on the y-axis, we can determine the regional time-dependent FPOP modifications for the TFE-induced G93A at different labeling times. The fraction modified of reporter peptide (x-axis) serves as a “timer” of labeling times, enabling a kinetic output similar to that of HDX curves for peptides [58]. All curves were fitted with a custom algorithm written in R [59]. In this way, we were able to obtain the time-dependent response of FPOP modifications for 10 representative peptides of TFE-untreated samples (native dimer), spanning the entire G93A sequence (Figure 5-6). Although this means of varying footprinting time is simple, the relationship between the modification extent of the reporter peptide and time is more complicated [45]. Most peptides show an apparent linear increase of modification extent as a function of time. Peptides 24-36, 37-69, 116-128, and 123-136 exhibit curvature in their kinetic plots, showing a high rate at short times as would be expected for regions showing relatively higher modification extents. We assign those regions with high FPOP modification extents to be solvent accessible, or to contain residues that are very reactive towards •OH [49], or both. The plots reach a plateau at approximately 40% of reporter modification probably because the concentration of radicals has significantly decreased at long times (Figure 5-6).
Figure 5-5. (A-E) Mass spectra of intact, TFE-untreated G93A SOD1 footprinted by FPOP at different labeling times as adjusted by varied histidine concentrations; (F) Time-dependent curve obtained by plotting the fraction unmodified of TFE-untreated G93A vs. reporter peptide fraction modified (black circles); the point representing TFE-treated G93A (blue triangle) is not on the curve.
Figure 5-6. Kinetic curves of 10 peptides from TFE-untreated G93A SOD1 in a time-dependent FPOP measured in triplicates, performed by adjusting different histidine concentrations. We used the curve of peptide 37-69 as an example to demonstrate that, such time-dependent measurements allow determination of FPOP yields at any given labeling time within the curve. The error bars correspond to the ± SE of triplicate measurements.

5.4.4 FPOP and Protein Flexibility

To test the hypothesis that reactive regions are those that are highly solvent and flexible, we extracted the b-factor values per-atom from SOD1 X-ray structure (PDB code: 3GZP) and averaged those values on a per-residue basis (Figure 5-7 A illustrates the residue-specific b-factor values of SOD1). Regions that exhibit larger b-factor values have higher flexibility in the protein tertiary structure and should be reactive in FPOP [60, 61], and we
see agreement between those b-factors and the measured FPOP labeling extent. For example, region spanning 122-143, which represents one of the major functional loops of SOD1 (i.e., electrostatic loop), shows pronounced FPOP yields (as is the case of peptide 116-128, 123-136 in Figure 5-6). Although this region is replete with FPOP-unreactive residues (e.g., Gly, Ala, Asp, Glu), Lys122, Lys128, and Lys136 are found most reactive and report the high solvent accessibility. The X-ray structure shows that the side chains of these three Lys’s point into solution (Figure 5-7 B), and the large b-factors of this region point to its high flexibility (Figure 5-7 A). Similarly, the high FPOP yields for peptide 37-69 correspond to the high flexibility of the Zn-binding loop (amino acid 50-84), as is indicated by those relatively higher b-factor values (Figure 5-7 A). Interestingly, the region spanning from 23-36, which is relatively protected because it packs together to form a β-sheet (β-strand 3) and is part of the β-barrel, also undergoes a high FPOP yield. Detailed FPOP analysis shows that reactions at Trp32 nearly accounts for all modifications of peptide 23-36 (tryptophan is very reactive towards hydroxyl radicals, with a near-diffusion-rate of \(1.3 \times 10^{10} \text{ M}^{-1}\text{s}^{-1} [49]\)). The crystal structure shows that the side chain of Trp32 is pointed outward from the β-barrel (Figure 5-7 C), making this residue readily available to hydroxyl radicals.

In contrast to these regions represented by peptides with high FPOP yields, those regions at the of N- and C- termini and that of 80-115 undergo low extents of FPOP modification (Figure 5-6), pointing to low solvent accessibility of reactive residues (Figure 5-7 A). These regions undergo protection either due to the formation of a dimer interface or of a
compact β-barrel. Taken together, FPOP footprinting results corroborate b-factors from crystal structure that describe the solvent accessibility of SOD1 molecule.

Figure 5-7. (A) Averaged residue-specific b-factor values plotted versus the residue index. Two regions showing high b-factor values are the zinc binding loop (residues 50-84) and the electrostatic loop (residues 122-143); (B, C) Zoom-in views of the SOD1 crystal structure (3GZP) show that the side chains of Trp32, Lys122, Lys128, and Lys 136 point to solution.

5.4.5 Effects of Destabilization Determined by FPOP

As discussed previously (in Chapter 3), one of the advantages of measuring kinetics in a differential scavenging system is that, it permits comparisons of two time-dependent
curves representing the proteins from two conditions; each curve comprises several different points corresponding to the FPOP yield at various labeling times [45]. It is not possible to take measurements under consistent labeling times, owing to the presence of adventitious scavengers. The time-dependent responses (as shown in Figure 5-6), however, allow us to determine the FPOP yield at any given labeling time, affording the opportunity to compare the FPOP results between TFE-treated and untreated samples, although their labeling times are different.

To compare the FPOP outcome between TFE-treated and untreated samples and to assess the structural differences, quantitation of FPOP yields between the two conditions at a constant labeling time is required. Using the observation that the labeling time of the TFE-treated G93A corresponds to 25% fraction modified of the reporter (Figure 5-4 D), we can readily determine from each kinetic curve the FPOP response of TFE-untreated G93A at the same “time point” (corresponding to 25% modification of the reporter peptide, (Figure 5-6)). Based on this approach, we can assess the FPOP yields between the TFE-treated and untreated G93A SOD1, at a constant labeling time (Figure 5-8 A).
**Figure 5-8. (A)** Comparison of FPOP yields of TFE-treated and untreated G93A SOD1 at peptide level under the same labeling time, as guaranteed by the same reporter fraction modified. Peptides showing differential modification extents are labeled with stars (yellow stars for peptides in dimer interface; blue stars for peptides in β-barrel). The error bars correspond to the ± SE of triplicate measurements; **(B)** FPOP results mapped onto the crystal structure of dimeric G93A SOD1 (PDB code 3GZP), the site mutation is annotated (blue arrow). Regions identified with same FPOP yields are colored red; regions showing differential FPOP yields are colored as described in A.

### 5.4.6 FPOP Confirms Dimer Dissociation

Comparing the regional modification extents between TFE-treated and untreated G93A, we note that regions represented by peptides 24-36, 37-69, 116-128, and 123-136 show statistically similar fraction modified (Figure 5-8 A). For the N-terminus (represented by peptides 1-9, 4-23, and 10-23) and the C-terminus (represented by peptide 144-153), the FPOP yields for TFE-treated G93A are higher than those for TFE-untreated G93A (Figure 5-8 A). The native dimeric SOD1 protein uses both its N-terminus and C-terminus as a dimer interface [62]. The FPOP reactivity changes indicate increased solvent-accessibility of the G93A dimer interface region after adding TFE, consistent with formation of the monomer. This observation is in accord with the results from native MS (Figure 5-3 B), whereby the predominant form of TFE-treated G93A SOD1 is the monomer. Furthermore, Hough et al. [33] showed, with solution X-ray scattering, that there is an increased tendency for SOD1 dimer to open the dimer interface with heat-induced destabilization of SOD1. It is also possible that the decreased interface stability can increase the population of poorly packed and malleable monomeric SOD1 species, elucidating one of the toxic SOD1 misfolding pathways [35].

### 5.4.7 FPOP Locates Unfolding Sites
Besides the dimer interface represented by the four peptides (yellow starred in Figure 5-8 A), a second region located away from the dimer interface and represented by peptides 80-91 and 92-115 (blue stars, Figure 5-8 A), also shows statistically significant increases of FPOP yields upon destabilizing the protein with TFE. The X-ray crystal structure of G93A SOD1 indicates this region is comprised of $\beta$-strand 5 and 6, which forms part of the closely-packed $\beta$-barrel (Figure 5-8 B). Therefore, it should be solvent inaccessible for the untreated protein, and this is in agreement with the low FPOP yields (<7%) of both peptides 80-91 and 92-115 when TFE is not added. The FPOP yields of these regions, however, increase dramatically for the samples treated with TFE. The overall yields at the region represented by peptide 80-91, which incorporates the entire $\beta$-strand 5 and part of the adjoining loop (Figure 5-9 A), increase from 2% to 6% (Figure 5-9 C). This increase suggests a local, destabilization-induced unfolding event, producing a more solvent exposed conformation of the 80-91 region. Examination of the product-ion spectra reveals hydroxyl-radical modifications of +16 on His80, Val81, and Lys91(Figure 5-10). Those differentially modified species are nicely separated as shown by their different retention times in the chromatography (Figure 5-9 B).

The region represented by peptide 92-115 comprises the $\beta$-strand 6 (amino acids 95-101) and the full adjoining loop between strand 6 and 7 (amino acids 102-115, Figure 5-11 D). Its FPOP yields increases significantly from 6% to 32% upon treatment with TFE (Figure 5-11 A), indicating pronounced unfolding and increasing solvent accessibility. Detailed residue-level analysis identifies with high confidence that Val97 and His110 are two of the amino acids being modified (Figure 5-11 B), although there are other residues modified.
at lower yields. The product-ion spectra of peptides containing these latter modified residues do not afford unambiguous sequence information owing to the co-elution of isomeric modified species. Surprisingly, quantitation of the FPOP yields of Val97 and His110 show that the fraction modified of His110 remains nearly unchanged before and after adding the TFE; whereas Val97 exhibits a yield increase of nearly one order of magnitude (Figure 5-11 B). Val97 belongs to strand 6, which is part of the $\beta$-barrel formed along with other seven strands. The crystal structure shows that Val97 is largely buried in the native form of G93A dimer, as its side-chain points toward the inward part of the $\beta$-barrel (Figure 5-11 D). This agrees with the FPOP results as Val97 is only 1.5% modified in the absence of TFE. Nevertheless, the pronounced increase of the Val97 yield strongly indicates the unfolding of the natively protected conformation, exposing Val97 to solvent and making it susceptible to hydroxyl radical attack. His110, on the other hand, is located on the adjoining short loop region between strand 6 and 7, with its side-chain pointing outward, according to the X-ray crystal structure (Figure 5-11 D). Therefore, the solvent accessibility of His110 is only marginally different even upon local unfolding. This difference underscores the ability of FPOP to make comparisons of subtle changes in folding [63].
Figure 5-9. Structure, extracted ion chromatograms, and FPOP quantitation of peptide 80-91. (A) G93A SOD1 crystal structure with zoom-in view of peptide 80-91 (colored in blue) and residues His80, Val81, and Lys91 (colored in yellow); (B) Extracted ion chromatograms (XICs) of unmodified and +16 modified peptide 80-91. The differently modified isomers (modifications on H80, V81, and K91) are separated chromatographically based on their respective hydrophobicity; (C) Quantitation of FPOP yields of peptide 80-91 between TFE-treated and untreated condition shows an overall increase from 2% to 6% with addition of TFE; (D) Quantitation of FPOP yields of residues His80, Val81, and Lys91 before and after TFE treatment. All error bars correspond to the ± SE of triplicate measurements.
Figure 5-10. Product-ion spectra corresponding to +16 modified peptide 80-91 locating the FPOP modification on His80 (A) and Lys91 (B).
Figure 5-11. Structure, extracted ion chromatograms, and FPOP quantitation of peptide 92-115. (A) Quantitation of FPOP yields of peptide 92-115 between TFE-treated and untreated condition shows significant increase from 6% to 32% upon TFE treatment; (B) Quantitation of FPOP yields of Val91 and His110, both residues account for most modifications of peptide 92-115; (C) XICs of unmodified and +16 modified peptide 92-115 of TFE-treated G93A; (D) Crystal structure with zoom-in view of peptide 92-115 (colored in blue) and residues Val91 and His110 (colored in yellow); (E) XICs of unmodified and +16 modified peptide 92-115 of TFE-untreated G93A. All error bars correspond to the ± SE of triplicate measurements.

5.4.8 Regions of Oligomeric Interaction

The comparison of FPOP results between TFE-treated and untreated G93A SOD1 unveils the rupture of dimer interface (N-terminal and C-terminal region, covering β-strand 1, 2, and 8), and the local unfolding of β-strand 5 and 6. Both processes may lead to the exposure of structurally promiscuous interfaces that are normally buried in the protein’s interior, increasing the propensity for self-association, as hypothesized by others.
[30, 33, 35, 39, 64, 65]. Furukawa et al. [65] experimentally identified a protease-resistant core structure within various fALS-inducing mutant SOD1 aggregates, and found three regions (amino acids 1-30, 90-120, and 135-153) that likely form a scaffold core in the aggregates. This finding is in good agreement with the FPOP results, reinforcing that the N-terminal region, C-terminal region, and \( \beta \)-strand 5 and 6 may be essential to aggregation (Figure 5-12).

Contemporaneous with this hypothesis, Nordlund and Oliveberg [64] suggested from folding kinetic analysis and \( \varphi \)-value analysis [66] that the N-terminal strands \( \beta1-3 \) is the seeds for SOD1 aggregation and that the regions of SOD1 most susceptible to unfolding under physiological conditions are strands \( \beta5, \beta6, \) and \( \beta8 \). These latter strands serve as structural protective caps preventing the hydrophobic edges of N-terminal strands from becoming exposed. In several ALS-provoking mutant variants, however, this protection mechanism is completely missing owing to early unfolding of these protective caps. Ding and Dokholyan [39] came to a similar conclusion on the basis of all-atom DMD simulations for SOD1 monomers, localizing strands 5, 6, and 8 as possessing some of the highest propensities for local unfolding. These findings support a hypothesis of sequential local unfolding events, where the strands \( \beta5, \beta6, \) and \( \beta8 \) would undergo earliest unfolding, followed by the exposure of sticky edge strands of N-terminus.
Figure 5-12. Secondary structure representation of SOD1, the eight anti-parallel β strands are shown as arrows with numbers representing the starting and ending residue index of each strand. The β strands that show increased FPOP yields with TFE treatment are colored green.

5.5 Conclusions

SOD1 destabilization is accompanied by not only dimer dissociation but also local unfolding, leading to re-organization of molecular conformations that are prone to aggregation. Our MS-based approaches combine top-down native ESI and protein footprinting (FPOP) coupled with bottom-up analysis to characterize the TFE-destabilized G93A SOD1 at both the protein and peptide/residue levels. The results from the two approaches are consistent with evidence from other studies, allowing us to draw a picture of the TFE-destabilized G93A SOD1 that exposes its dimer interface and undergoes partial unfolding on the β-barrel. Our data are consistent with a previous conclusion that
SOD1 follows the misfolding and aggregation mechanism, whereby the exposure of hydrophobic cores facilitate SOD1 self-association; inhibition of such exposure may be useful in alleviating the ALS symptoms.

Our current FPOP results identify loss-of-structure on both regions without differentiating which one comes first. To dissect the unfolding events, however, a two-laser “pump/probe” platform [67, 68] may be useful, allowing T-jump destabilization of SOD1 proteins followed by hydroxyl radical footprinting to track the conformational changes.
References


37. Furukawa, Y., O'Halloran, T.V.: Amyotrophic lateral sclerosis mutations have the greatest destabilizing effect on the apo- and reduced form of SOD1, leading to unfolding and oxidative aggregation. The Journal of biological chemistry. 280, 17266-17274 (2005)


39. Ding, F., Dokholyan, N.V.: Dynamical roles of metal ions and the disulfide bond in Cu, Zn superoxide dismutase folding and aggregation. Proceedings of the


Chapter 6

Mass Spectrometry and Protein Footprinting in
Characterizing EGF Receptor Intracellular Domain (EGFR-ICD)
6.1 Abstract

The epidermal growth factor (EGF) receptor is regulated by several mechanisms: stimulation by ligand binding and self-phosphorylation, inhibition by heterologous phosphorylation and downregulation by ligand binding. It mediates many cellular response in normal biological processes and in pathological states. The C-terminal tails of all EGF receptor molecules have important regulatory functions. Phosphorylated tails can recruit various adapter proteins via their SH2 (Src homology domain 2) or PTB (phosphotyrosine binding domain) domains. Details about those interactions, however, remain ambiguous, owing to the intrinsic disorder of C-terminal tail. Here, we report protein footprinting techniques (hydrogen-deuterium exchange, hydroxyl radical labeling) coupled to mass spectrometry to characterize the conformations of EGF receptor and complex, by probing their protein surface solvent accessibility. Both HDX and FPOP data agree with the molecular modeling results performed by our collaborators in Washington University Medical School, that the C-terminal tail domain of EGF receptor is largely disordered. We further found that HDX is insufficient to distinguish the Cbl-bound and unbound states of EGF receptor, because the ligand binding kinetics are fast.

6.2 Introduction

The EGF receptor is a transmembrane glycoprotein that constitutes one of the four members of the ErbB family of tyrosine kinase receptor (TKR). Other members of ErbB group include ErbB2, ErbB3, and ErbB4 [1]. All of these tyrosine kinase receptors share the same structural components: an extracellular ligand-binding domain, a single-pass transmembrane domain, an intracellular tyrosine kinase domain, and a C-terminal tail
domain encompassing nine tyrosine residues that can be phosphorylated. In its inactive state, the receptor exists primarily as a monomer [2]; upon binding ligand, however, the receptor undergoes an extensive conformational change that allows the formation of back-to-back dimers by the extracellular domains [3, 4].

The dimerization of the extracellular domains results in the formation of activating asymmetric dimers of the intracellular kinase domains, where the activity of one kinase domain is stimulated to phosphorylate tyrosine residues on the C-terminal tail of its partner subunit [5-7]. The phosphorylated tyrosines on the EGF receptor serve as binding sites for numerous signaling proteins that contain SH2 and/or phosphotyrosine-binding domains [8, 9]. Some of these proteins possess enzymatic activity (e.g., Cbl); others, such as Grb2 or Shc, function as adapter proteins that bridge other proteins with the EGF receptor. For instance, Grb2 recruits Gab1 to the EGF receptor as scaffolding protein; phosphorylation of Gab1 by the EGF receptor permits Gab1 to recruit further additional signaling proteins, such as PI3K-R1 or Shp2 [10, 11]. The recruitment of these signaling proteins eventually trigger the activation of various downstream signaling pathways, which are closely associated with cell proliferation, mobility, apoptosis, etc. The mutation and overexpression of EGF receptor are associated with a variety of cancers, making this prototypical dimerization-activated receptor tyrosine kinase a prominent target of cancer drugs.

The C-terminal tail of EGF receptor, although playing a vital role in the downstream signaling transduction, has thus far eluded definitive structural and mechanistic characterization [12]. No crystal structures are so far available for the C-terminal tail
domain because that region is highly flexible and membrane-associated [13-15]. As an alternative to X-ray crystallography, mass spectrometry (MS) may be a useful means for protein structure characterizations [16-19].

Recent developments have demonstrated that labeling of protein solvent-accessible regions bring insights into the high order structures (HOS) of a protein. This enables the differentiation of flexible regions from rigid regions; for example, a flexible loop region may show more labeling than a closely-packed β-sheet because the loop is more solvent exposed. More importantly, in a two-state comparison, one could localize the protein-ligand interactions, including protein-small molecule, protein-peptide, and protein-protein interactions [13, 20-24]. Chemical footprinting methods coupled to MS can identify peptides/residues with altered reactivity as a result of binding. Various chemical labeling approaches are available (e.g., hydrogen deuterium exchange (HDX) [20, 23], hydroxyl radical footprinting (FPOP) [25, 26], carbene labeling [27], and some other probes that are amino-acid residue-specific [24, 28]).

Keppel et al. [29] recently demonstrated the intrinsic disorder in C-terminal tail of both EGF receptor and HER3 receptor tyrosine kinase via HDX-MS. Numerous molecular dynamics studies have also suggested the high flexibility of C-terminal tail [30, 31]. The intrinsic disorder and extended confirmation of the tail is believed to be important for its functions by increasing the capture radius and reducing the thermodynamic barriers of binding to downstream signaling proteins. One of our major goals with EGF receptor is to localize the binding site(s) of these signaling proteins on the C-terminal tail. The fast dynamics on the disordered tail, however, pose great challenges to the HDX-MS-based
conformational characterizations, because of the relatively long labeling time scale; thus, the labeling samples an ensemble of averaged dynamic conformations [32]. Fast photochemical oxidation of proteins (FPOP), on the other hand, provides opportunities to probe rapidly changing protein conformations. The short-lived hydroxyl radicals enable irreversible labeling of residue side-chains on the microsecond time scale; allowing the mapping of protein surface solvent accessibility in a near-native environment [33-35]. Interfaced with the bottom-up mass spectrometry, FPOP interrogates residue-specific modifications.

Characterizing the EGF receptor intracellular domain with FPOP is challenging in many ways: (i) Sample complexity: numerous species from up-stream processing, including APT, DTT, Tween20, and liposome, coexist in solution with the protein. Therefore, an efficient sample cleanup approach is required prior to the mass spectrometric analysis; (ii) Protein conformation heterogeneity: this tyrosine kinase receptor has nine phosphotyrosine residues on the C-terminal tail domain. Whether or not these residues are phosphorylated, or to what extent each of these residues are phosphorylated, is closely associated with a variety of different tail conformations, and, therefore, with various downstream signaling transductions. (iii) Data analysis: hydroxyl radical labels 14 out of 20 amino acids, this labeling is mostly manifested as post-translational modifications (PTMs) of +16, +32, etc., although other PTMs are also possible (e.g., -30 for Asp, -10 for His) (see Chapter 1 Table 1-1) [36]. The incorporation of phosphorylation group as one extra PTM would greatly complicate the data processing regarding the quantitation of FPOP yields (this will be discussed later in this chapter). Allowing for all
these challenges, we started our research by identifying the phosphorylation sites and quantifying the phosphorylation extent using bottom-up mass spectrometry. Following this we tested the feasibility of using FPOP to probe the conformation of EGF receptor intracellular domain. Our work demonstrates that the phosphorylation extent on the C-terminal tail is heterogeneous, and through FPOP footprinting, we confirmed that the C-terminal tail of EGF receptor is intrinsically disordered.

6.3 Experimental

6.3.1 Materials

Leucine enkephalin acetate hydrate, L-histidine, L-methionine, 30% hydrogen peroxide, catalase, acetone, urea, dithiothreitol (DTT), iodoacetamide (IAM), 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris base), 36.5-38.0% hydrochloric acid, ammonium acetate, formic acid (FA), trifluoroacetic acid (TFA), HPLC-grade solvent, and phosphate buffered saline (PBS) were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). Sequencing-grade trypsin was purchased from Promega Co. (Madison, WI). The EGF receptor protein and signaling protein Cbl were expressed, purified, phosphorylated in the laboratory of our collaborator, Dr. Linda Pike.

6.3.2 FPOP Footprinting

A total volume of 50 µL of solution containing ~2 µM protein, 5 µM Leu-enkephalin, 20 mM H₂O₂, and histidine at various concentrations in PBS, was loaded in a 100 µL syringe (Hamilton, Reno, NV) and advanced using a syringe pump (Harvard Apparatus, Holliston, MA). The excimer laser (GAM Laser Inc., Orlando, FL) was adjusted to 21.5 mJ/pulse at
a frequency of 7.4 Hz and focused to an irradiation window of 2.67 mm. The solution flow rate was adjusted to ensure a 25% exclusion volume fraction to avoid repeated exposure of radicals. Samples after FPOP were submitted to acetone precipitation and trypsin digestion, followed by bottom-up MS analysis.

6.3.3 Bottom-up Analysis

The LC-MS/MS analyses were conducted back-to-back with the UltiMate 3000 Nano LC system and a Thermo Q Exactive Plus Hybrid Quadruple-Orbitrap Mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Aliquots (5 µL) of digests were drawn from autosampler vial and loaded onto a C18 reversed-phase trapping column (Acclaim® PepMap100, 100 µm × 2 cm, 5 µm, 100 Å; Thermo Fisher Scientific, Waltham, MA) at 4.5 µL/min with water in 0.1% FA for a ten-minute desalting. Peptides were separated using a custom-packed C18 reverse-phase column (Symmetry, 100 µm × 200 mm, 5 µm, 100 Å; Waters Inc., Milford, MA) and eluted with a 76-min gradient delivered at 500 nL/min. The scan range of precursor ions was set from m/z 380 to 2200 for all samples with high mass resolving power of 70,000. Dynamic exclusion was activated for 8 s following each scan to enable fragmentation of low abundant peptides. The maximum injection time (IT) for full mass spectral scan was set to 200 ms, and 150 ms for each MS/MS scan.

6.3.4 FPOP Data Analysis

The *.raw MS files were searched directly using Byonic (Protein Metrics Inc., San Carlos, CA) against a custom-built database containing the sequence of EGFR-ICD and Cbl. All known HO• side-chain reaction products were added to the variable modification database for searching of HO•-modified peptides [36, 37]. The quantitation of FPOP
yields was based on ion abundances from extracted ion chromatograms (XICs). For more detailed data analysis, please see Chapter 5.

6.3.5 Hydrogen Deuterium Exchange (HDX)

Stock solution of 8 µM EGFR-ICD and 16 µM Cbl were prepared. Samples were equilibrated at 25 °C for 30 min before HDX analysis (1:2 ratio). Continuous labeling with deuterium was initiated by diluting 10 µL of the stock solution into 40 µL of D₂O. HDX was measured at 5, 10, and 30 s. Two proteases, pepsin (immobilized) and fungal XIII (2 mg/mL), were used for protein digestion to achieve high peptide coverage. The exchange reaction was quenched by mixing with 60 µL of 4 M urea, 2% formic acid at ~4 °C. The mixed solution was then passed through a custom-packed pepsin column (2 x 20 mm) at 200 µL/min. The peptic digests were then trapped onto a reversed phase C8 column (Agilent, Santa Clara, CA, USA) and desalted with a 3-min flow at 200 µL/min of water with 0.1% formic acid, followed by separation using a 2.1 x 50 mm reversed phase C18 column (1.9 µm Hypersil Gold; Thermo Fisher Scientific, Waltham, MA, USA) with an 9.5 linear gradient of 2% to 40% acetonitrile in 0.1% formic acid at 50 µL/min. MS detection was performed on an LTQ-Orbitrap (Thermo Fisher Scientific, Santa Clara, CA, USA) operating in the data-dependent mode, where the six most abundant ions from each scan were selected for MS/MS analysis.

6.3.6 HDX Data Analysis

The *.raw MS files were submitted to MassMatrix for peptide identification and outputted as *.csv files with identified peptides. Both *.csv and *.raw files were input into
HD Examiner software to calculate the centroid masses of isotopic envelopes (m) and deuterium uptake level (D%).

6.4 Results and Discussion

6.4.1 Need for Sample Cleanup

Protein samples commonly contain substances that interfere with downstream application [38]. During our work of bottom-up mass spectrometry to separate spatially each tryptic peptide of EGF receptor, we observed large amounts of polymer-based contaminants coeluting with protein digests. The main component appears to be Tween20 (a polyethylene glycol containing detergent) originally present in the protein sample [39]. The signals corresponding to these polymer contaminants were more intense than the protein digest signals by approximately 1000-fold, indicating that the EGF receptor signals are seriously suppressed in a typical data-dependent MS acquisition mode, resulting in poor protein sequence coverage [40] (Figure 6-1 A shows the total ion chromatogram (TIC) of EGF receptor tryptic digests without sample cleanup). Those intense peaks eluting from 25 min to 40 min correspond to polymeric species with different replicates of PEG units (Figure 6-1 B shows a mass spectrum within that 25-40 min time window). The spectrum contains equally distanced peaks ($\Delta m = 44, -\text{O-CH}_2\text{-CH}_2^-\$) that are characteristic of PEG-containing polymer species that coelute with peptides from the EGF receptor [41]. Besides Tween20, there are other substances that would potentially interfere with the downstream processing as well. As far as we knew, the sample also contains $\beta$-glycerophosphate, DTT, ATP, EDTA, MgCl$_2$, vanadate salts.
To overcome such sample complexity, an effective sample cleanup method was necessary prior to the LC-MS analysis. Several strategies exist for eliminating contaminants from samples. One way of doing it is to exchange sample solutions into appropriate buffers by dialysis or gel filtration (desalting columns) [42, 43]. Another strategy for removing undesirable substances is to add a compound (acetone, trichloroacetic acid, methanol, chloroform, etc.) that causes protein to precipitate [44, 45]. After centrifugation to pellet the precipitated protein, the supernatant containing the interfering substances is removed and the protein pellet is resuspended in a buffer that is compatible with the downstream analysis. Figure 6-1 C shows the TIC of EGF receptor tryptic digests after acetone precipitation. Compared with the TIC before precipitation (Figure 6-1 A), those polymeric contaminants can be eliminated effectively by acetone precipitation. The multiple peaks at different retention times correspond to tryptic peptides from EGF receptor that eluted from the C18 reversed phase column.

Given that protein precipitation would denature the protein, one should perform such sample cleanup after labeling. It would be challenging to use a protocol involving protein precipitation with HDX, as samples are likely to undergo significant back-exchange during sample handling. The irreversible nature of FPOP footprinting, however, is compatible with protein precipitation, as decays of labeling or scrambling are not issues [44, 46]. It should also be noted that protein precipitation is subjected to sample loss; however, we determined that the loss of protein is not significant as the protein signals in LC-MS measurement did not drop noticeably (data not shown).
Figure 6-1. (A) Total ion chromatogram (TIC) of EGF receptor trypsin digests without sample cleanup, loaded amount 2.5 pmol; (B) Mass spectrum containing all scans from 25 to 40 min elution window in the TIC of EGF receptor without sample cleanup; (C) TIC of EGF receptor trypsin digests after acetone precipitation, loaded amount 2.5 pmol.

6.4.2 Identification and Quantitation of Phosphorylation

An activated EGF receptor dimerizes and auto-phosphorylates the C-terminal tail, which contains nine tyrosine residues (Figure 6-2). The phosphorylation status of these tyrosine residues determines the binding of numerous signaling proteins [6, 9, 11, 30]. Besides the nine phosphotyrosine in the C-terminal tail, there are also Thr and Ser residues that are also susceptible to the auto-phosphorylation, although the down-regulation effect of Ser/Thr phosphorylation is believed to be less significant [47]. Herein, we sought to use
a bottom-up mass spectrometry approach to provide a quick identification and quantitation of phosphorylation for EGF receptor intracellular domain (EGFR-ICD). To do this, we prepared solutions of EGFR-ICD in triplicate that were identical in every respect, except for the different phosphorylation time (0 min vs. 10 min). After submitting samples (50 pmol) to acetone precipitation and trypsin digestion, we immediately loaded one aliquot of tryptic digests onto a C18 column for LC-MS/MS analysis. We relied on the product-ion spectra for phosphorylation identification, as the phosphorylation is manifested as a PTM with addition of 79.9663 Da (corresponding to addition of -HPO₃). Figure 6-3 shows the comparison of product-ion (MS/MS) spectra of the unphosphorylated (A) and singly phosphorylated (B) peptide of GSTAENAEYLR (amino acid 1165-1175). We observed a mass shift of 80 Da, starting from the y3 ion, for the entire y-series product ions in the product-ion spectrum of the singly phosphorylated species, pinpointing the phosphorylation at Tyr1173. To quantify the extent of phosphorylation, the ion chromatograms (XICs) corresponding to both unphosphorylated and phosphorylated peptides are extracted. The peak integrals of XICs are used to represent the abundance of each peptide species, and we calculate the extent of phosphorylation as the ratio of these integrals [25, 26] (Figure 6-4 shows the XICs of monoisotopic mass of peptide 1165-1175 (+2 charge state) based on the precursor mass threshold of 15 ppm). The m/z values used for the signal extraction are 605.7886 (unphosphorylated), 645.7717 (singly phosphorylated), and 685.7549 (doubly phosphorylated). We did observe the signal, although it is not intense, corresponding to the doubly phosphorylated species for this peptide. Residues Tyr1173, Ser1166, and
Thr1167 are phosphorylated at low level as well. The peptide after being submitted to 10-min phosphorylation conditions shows phosphorylation extent of approximately 90%; whereas the one from 0-min condition only has a basal level phosphorylation of 1%.

One needs to be cautious about the extent of phosphorylation determined in this manner, because the incorporation of a phospho-group onto a peptide might inevitably alter its ionization efficiency, leading to a biased quantification result [48, 49]. The biased range regarding the ionization efficiency between the phosphorylated and unphosphorylated peptides was previously shown to be within 20% [50], indicating that our quantification results are within a reasonable range. Stable isotope labeling experiments (e.g., SILAC) are available for a more accurate quantitation of phosphorylation extent [51, 52], but we are faced with more serious analytical problems prior to considering such.

A bar-graph representing the residue-specific phosphorylation quantitation result between the 0-min and 10-min phosphorylation EGFR-ICD (Figure 6-5) shows, besides the nine tyrosines in C-terminal tail, that there is also phosphorylation of tyrosines in the kinase domain and also of Ser/Thr. The protein shows an average basal level of phosphorylation of approximately 10%; most residues exhibit no phosphorylation at zero time except Tyr920, S967, Tyr1086 and Tyr1148 phosphorylations are relatively high. We measured the average phosphorylation extent to be ~75% after 10-min of phosphorylation. Thus, we can conclude that the difference regarding the phosphorylation extent between the unphosphorylated and phosphorylated samples is significant.
Figure 6-2. Representation of activated and dimerized EGF receptor. The nine phosphotyrosine residues on the C-terminal tail are annotated; these sites serve as the binding sites for multiple signaling proteins including Cbl, Grb2, Shc, SHP2.
Figure 6-3. Product-ion (MS/MS) spectra of EGFR-ICD tryptic peptide GSTAENAEYLR, amino acid 1165-1175. (A) The y ion series observed for the doubly charged unphosphorylated peptide 1165-1175 of m/z 605.79; (B) the y ion series for the doubly charged singly phosphorylated peptide 1165-1175 of m/z 645.78. All y ions after y3 shows a mass shift of 80 Da that represents phosphorylation at Tyr1173.
Figure 6-4. Comparison of extracted ion chromatograms (XICs) of peptide 1165-1175 between unphosphorylated sample and 10-min phosphorylated sample. Monoisotopic m/z of 605.7886 (unphosphorylated), 645.7717 (singly phosphorylated), and 685.7549 (doubly phosphorylated) were used.

Figure 6-5. Bar-graph showing the differences in phosphorylation extents for major identified residues for 0-min and 10-min reaction.

6.4.3 HDX Not Suitable for Probing the C-terminal Tail Binding

Hydrogen deuterium exchange (HDX) was recently adapted to demonstrate that the C-terminal tail of EGFR-ICD is intrinsically disordered [29]. To test whether HDX-MS can be used to probe the binding between C-terminal tail and signaling proteins, we did a preliminary HDX study on EGFR-ICD binding with Cbl, one of the signaling proteins of MW ~40 kD. The binding of Cbl with the C-terminal tail is believed to be very dynamic; therefore, as a preliminary study we performed HDX at three short time points (5, 10, and
30 s), making single measurements. An overview of the HDX results (Figure 6-6) show the outcome, suggesting, in accord with previously determined results [29], no significant differential uptake between the Cbl-bound and unbound state. Almost all peptides show same level of deuterium uptake for all three time points, with one exception of region 974 to 991, which shows ~15% differential deuterium uptake after 5 s of HDX. This difference attenuates with 10 s HDX, and eventually diminishes at 30 s (Figure 6-6 B, C). The binding between Cbl and EGFR occurs, either directly through the pY-binding module contained in Cbl that binds to pY1045 of the tail, or indirectly through binding of a Grb2:Cbl complex to either pY1068 or pY1086, via the SH2 domain of Grb2 [53-55]. The peptides containing those phospho-tyrosine residues, however, show overlapping HDX kinetic curves. This suggests that the binding between EGFR C-terminal tail and Cbl has faster on-and-off rates than those of HDX [56], indicating the current HDX platform is not fast enough to capture the deuterium uptake differences between the two states.
Figure 6-6. (A-C) HDX result of EGFR-ICD binding to Cbl under three different incubation times (5s, 10s, and 30s). Each point represents a single determination of HDX experiment.

6.4.4 FPOP Probing Solvent Accessibility of EGFR-ICD

To test the feasibility of FPOP to explore the binding between signaling proteins and the C-terminal tail of EGFR-ICD, we first examined the intrinsic disorder of the EGFR-ICD C-terminal tail without any binding using FPOP footprinting. This experiment can show extensive footprinting if the region is indeed flexible, and disordered and it can be useful
in determining the optimal FPOP labeling conditions for any follow-up studies and. We carried out the FPOP experiments as described previously [57]. Given the sample complexity in this study, we incorporated Leu-enkephalin as the reporter peptide to measure the quenching effect of adventitious scavengers (see Chapter 3 for details about reporter peptide) [58]. Phosphorylated EGFR-ICD samples were aliquoted into two portions, with one portion submitted to FPOP labeling directly, and the other to buffer exchange (to 10 mM PBS buffer) prior to FPOP, both experiments used 1 mM His as the radical scavenger. After FPOP, we quantified the fraction modified of reporter peptide based on LC-MS (XICs representing the unmodified and +16 modified Leu-enkephalin are in Figure 6-7 A and B). The reporter peptide shows a huge difference in FPOP yield, indicating that the high radical-scavenging nature of the protein solution when there is no buffer exchange. The concentration of adventitious scavengers clearly is reduced by buffer exchange, as is clearly demonstrated by the dramatic increase of FPOP yields of reporter peptide from 3% to 28% (Figure 6-7 C). Such increases of FPOP yield is also occur on the EGFR-ICD, given that we observed a consistently increased fraction modified of EGFR-ICD tryptic peptides after buffer exchange (Figure 6-8). Our FPOP results suggest that, owing to the existence of adventitious radical scavengers, it is necessary to perform a buffer exchange (or other suitable purification) prior to FPOP footprinting so as to obtain a high and unbiased level of FPOP yields on the EGFR-ICD or, for that matter, on any protein.
Figure 6-7. (A) XICs of unmodified and +16 modified Leu-enkephalin when there is no buffer exchange, the +16 species is minor compared to the unmodified species; (B) Same as (A), but after buffer exchange prior to FPOP labeling; (C) Quantification of FPOP yields on reporter peptide shows dramatic increase from 3% to 28% with buffer exchange.

Figure 6-8. Quantification of FPOP yields for EGFR-ICD tryptic peptides with and without buffer exchange. In agreement with the reporter peptide results, almost all peptides show increased yields after buffer exchange. The red dashed line separates the kinase domain peptides (left) and C-terminal tail peptides (right).
6.5 Data Interpretation

6.5.1 Normalization of Reactivity Towards Hydroxyl Radical

The experimental protocol of protein footprinting includes steps to label the protein (backbones or side chains), then enzymatically digest the protein to peptides, and quantify the extent of labeling per peptide [59]. When proteins bind to ligands (e.g., peptide, metal, another protein), protein footprinting is especially powerful for mapping the interfaces between the domains of macromolecules and mapping conformational changes due to ligand binding [37, 60, 61]. In the absence of a ligand, however, protein footprinting can also be useful to probe the solvent accessibility of proteins. By quantitating and comparing the extent of modifications, one can differentiate those flexible, dynamic regions from those that are protected. For example, HDX-MS can provide local information about different regions of the EGFR-ICD protein and readily distinguish folded or structured, globular regions from intrinsically disordered regions [29]. As is well established, the C-terminal tail domain exchanges very rapidly and reaches a plateau (~80% deuterium uptake) before 5 s of exchange, whereas the kinase domain doesn’t reach a maximum even after 2 h of exchange.

Nevertheless, with hydroxyl radical labeling, such direct comparison of modification extents between one region of a protein and the other does NOT straightforwardly infer the flexibility and rigidity of the structure. As shown in Eq. 6-1, the fraction modified of peptide in FPOP is not only associated with its solvent accessibility but also related to its reactivity with hydroxyl radicals. The reactivity of amino acids towards HO•, however, can
be different by three orders of magnitude, from $E^7$ to $E^{10} \text{M}^{-1}\text{s}^{-1}$ [37]. This indicates that the second term in Eq. 6-1 is highly variable and dependent on the nature of the amino-acid residue when it comes to making comparisons between different regions of a protein, as their amino acid sequences can be quite different. Thus, a proper normalization of the overall reactivity of peptides towards HO• is necessary to compare the solvent accessibility among various regions of a protein based on the respective FPOP yields. It is noteworthy, however, in the typical scenario of two-state comparison experiments (e.g., apo- vs. holo-), such normalization step is not required because the amino acid sequences under comparisons are the same providing the oxidizing potential of the two solution is the same for both solutions (except for the protein under study).

\[ \text{Peptide fraction modified} \propto (\text{Solvent Accessibility}) \ast (\text{Overall Reactivity}) \quad \text{Eq. 6-1} \]

To normalize the overall reactivity of each peptide and obtain the solvent accessibility, we compute the summation of HO• rate constants towards all residues within each peptide, this represents the peptide-specific overall reactivity. The FPOP yields of each peptide is then divided by the overall reactivity (Eq. 6-2). The resulting “solvent accessibility” should be able to reflect the solvent accessibility of peptides (i.e., larger number corresponds to higher level of flexibility).

\[ \text{Solvent accessibility} \propto \frac{\text{Peptide fraction modified}}{\text{Summation of residue reactivity}} \quad \text{Eq. 6-2} \]

We can apply the reactivity normalization to the EGFR-ICD FPOP results (Figure 6-8), and obtain a plot representing the EGFR-ICD peptide-specific solvent accessibility (Figure 6-9). Although the FPOP yields are quite different with and without buffer
exchange, both conditions give very similar solvent accessibility output (Figure 6-9). Nevertheless, we were surprised to find that the C-terminal tail (amino acids 1008-1186) does not show dramatically increased solvent accessibility compared to the kinase domain (amino acids 663-962); as a matter of fact, on the basis of these measurements, the kinase domain appears to have similar flexibility as does the C-terminal tail. This contradicts with the usual idea that the C-terminal tail is largely disordered [6, 29, 30].

![Figure 6-9](image)

**Figure 6-9.** Solvent accessibility output of EGFR-ICD measured by FPOP with the reactivity normalization with buffer exchange (A) and without buffer exchange (B) prior to labeling.

### 6.5.2 Is the Kinase Domain of EGFR-ICD a Radical Sink?
A detailed comparison of the amino acid content of the kinase domain (amino acids 663-962) and C-terminal tail domain (amino acids 1008-1186) shows that the distribution of FPOP-reactive residues is largely uneven (Table 6-1). In particular, kinase domain contains 11 methionines, 6 tryptophans, and 24 arginines, a set of residues that are highly reactive towards HO•. The C-terminal tail of EGFR-ICD consists of residues that are mostly less reactive; that is, it contains no methionines, 1 tryptophan, and only 5 arginines. It is reasonable that in the “micro environment” of a protein surface, the kinase domain competes with the C-terminal tail in reacting with hydroxyl radicals. Consider the bulky size and higher reactivity, the kinase domain can be viewed as an “attached” radical scavenger of the C-terminal tail domain, and the result is a diminished oxidation of the tail. The measured FPOP yields concur with this argument (Figure 6-8). The averaged oxidation extent among the 9 peptides identified in C-terminal tail domain of EGFR-ICD is 5%, which is far lower than 12%, the averaged oxidation extent of kinase domain, despite our understanding that the kinase domain is more compact than the C-terminal tail.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Kinase domain</th>
<th>C-terminal tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Try</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Tyr</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Met</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Phe</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>His</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Arg</td>
<td>24</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 6-1. The numbers of hydroxyl radical reactive residues in kinase domain and C-terminal tail domain of EGFR-ICD. The kinase domain contains a large number of FPOP-reactive residues.
6.5.3 Should Met-Containing Peptides be Excluded for FPOP Comparisons?

Methionine residues, bearing its sulfur-containing side chain, are among the most reactive residues towards HO• [37]. FPOP modification of methionine gives rise to primarily methionine sulfoxide (+16 Da mass shift), which can be further oxidized to methionine sulfone (+32 Da mass shift). Another minor product with a -32 Da mass shift has also been reported, probably due to aldehyde formation at the γ-carbon [37].

From Figure 6-8, we see that all the methionine-containing peptides showed significantly higher modification extents (>30%), compared to those containing no methionine. Product-ion (MS/MS) spectra suggest the high yields come from methionine oxidation. Especially in the kinase domain, methionine oxidations are nearly the predominant modifications accounting for most of the FPOP yields. Among the 18 peptides identified within kinase domain, eight contain methionine and show more oxidation compared to the others (Figure 6-8). Furthermore, even in the control experiments where there is no laser irradiation, we find large amounts of +16 and +32 oxidized peptide species, which are largely due to Met oxidations. Figure 6-10 illustrates the XICs of tryptic peptide 976-1007 from EGFR-ICD, where we saw abundant singly and doubly oxidized peptide species for both samples that were footprinted with FPOP and samples in the no laser control. All major peaks correspond to Met oxidations. The yields for samples submitted to FPOP shows 26% modification extent; nevertheless, that same peptide shows 15% oxidation extent even without turning on the laser! Similar phenomena were nearly exclusively observed for other methionine-containing peptides of EGFR-ICD. Methionine is
susceptible to oxidation by other oxygen-reactive species (including hydrogen peroxide) [62]. Thus, the high-level background oxidation of methionine is likely due to the protein exposure to H₂O₂ prior to the laser irradiation. Therefore, that basal oxidation extent should be excluded when it comes to quantitation of FPOP yields. These methionine oxidations are likely to present a skewed picture of the measured high solvent accessibility of the kinase domain (Figure 6-9).

An alternative explanation for the unusually high methionine oxidations in the kinase domain of EGFR-ICD, is that the precursor hydrogen peroxide binds to the kinase prior to laser irradiation. The catalytic pocket of kinase may be enriched in H₂O₂, and when the laser pulse occurs, the high concentration of H₂O₂ gives also a locally high concentration of methionine. Although there are few studies regarding the binding between H₂O₂ and EGFR kinase domain, it could be instructive by considering the catalase binding with H₂O₂. Multiple researches have suggested that Arg and His of catalase would preferentially form hydrogen bonds with H₂O₂, as the first step of catalysis [63, 64]. Table 6-1 shows that the kinase domain is abundant with these residues, especially Arg; therefore, it is likely that H₂O₂ can form the kinase-H₂O₂ complex, giving rise to local enrichment in the EGFR kinase, which leads to the predominant methionine oxidation in kinase domain.
6.5.4 FPOP Outcome with Removal of Met-Containing Peptides

Perhaps the simplest way to avoid getting confounded by questionable methionine oxidations is to dismiss those methionine oxidations. One of the advantages of hydroxyl radical labeling is its multiple targets, whereby about three quarters of amino acids can be modified [37]. Theoretically speaking, any modified residue should faithfully report the protein solvent accessibility. This has been demonstrated by various residue-specific labeling approaches. For example, with glycine ethyl ester (GEE) labeling, glutamate and
aspartate are the only two residues modified; this approach also probes protein solvent accessibility [13]. The iodine species produced from photolysis of 4-iodobenzoic acid, which selectively is incorporated in His and Tyr residues, also allows interrogation of the high order structures of proteins [28].

Similarly, in FPOP, residues other than methionine can serve as probes of the solvent accessibility. To do this, we performed the normalization steps as mentioned in 6.5.1 by using peptides containing no methionine residues. As opposed to the results in Figure 6-8, where all peptides are selected for comparison, Figure 6-11 shows the solvent accessibility map of EGFR-ICD with removal of methionine-containing peptides. Figure 6-11 indicates, presumably correctly, that the C-terminal tail of EGFR-ICD is more solvent accessible compared to the kinase domain, pointing to its higher flexibility.
Figure 6-11. Solvent accessibility output of EGFR-ICD measured by FPOP with normalization of reactivity of peptides containing no methionine residue, with buffer exchange (A) and without buffer exchange prior to labeling (B). The C-terminal tail domain shows more solvent accessibility in both conditions.

6.6 Conclusions

In this chapter, we show that bottom-up mass spectrometric analysis can be useful in locating phosphotyrosine residues of EGF receptor protein. Besides tyrosine, we found some serine and threonine are also phosphorylated. We quantified the extent of phosphorylation, and found out the average phosphorylation extent of the EGFR-ICD is
~75%, suggesting that an ensemble of different conformations exists for the C-terminal tail.

To confirm the intrinsically disordered C-terminal tail of EGFR-ICD and establish EGFR-ICD binding to Cbl, we performed HDX and compared the Cbl-bound and unbound states. HDX revealed that the C-terminal tail of EGFR-ICD is highly flexible, however, we found that the current HDX platform is too slow to characterize the binding between the tail and Cbl, likely owing to the fast on-and-off rate in the binding. We then applied FPOP and used leu-enkephalin as reporter peptide to optimize the labeling conditions. Interestingly, we found unusually high extent of background oxidation of methionine in the kinase domain, which may be due to the direct oxidation by H₂O₂ prior to laser irradiation. These excessive methionine oxidations bias the comparison of solvent accessibility between the kinase domain and C-terminal tail domain of EGFR-ICD, and could lead to biased conclusions that the kinase domain has more solvent accessibility than the C-terminal tail. We may be able to simply dismiss those methionine-containing peptides, and solely rely on the rest of residues to report the solvent accessibility. In fact, by doing so, the FPOP outcome concurs with previous HDX results that the C-terminal tail domain shows more solvent accessibility and is more disordered than the kinase domain [29] (Figure 6-6).

6.7 Future work

Our next step is to characterize the binding between the EGFR-ICD tail and signaling proteins (e.g., Cbl, Grb2). A variety of labeling techniques can be used. The excessive methionine oxidations of FPOP is less an issue in this context because there is no methionine residue in the C-terminal tail domain of EGFR-ICD. Therefore, we expect
FPOP can readily probe the binding interface and report on conformational changes in the C-terminal domain as a result of binding. The current HDX platform is limited by its time-scale; thus, a faster HDX platform is needed. We should consider the work of Weis and co-workers [65] who developed a simple apparatus for millisecond quench-flow HDX, which is effective for characterizing disordered proteins. An alternative is to use carbene labeling [27, 66], the labeling time scale of which is in nanosecond. Carbenes do not react extensively with Met, and they favorably target Asp and Glu, two residues that are abundant in the tail region. Another possibility is the CF3•, which was recently discovered by the Gross laboratory to be a general footprinting reagent that does NOT modify Met.

**Acknowledgements:** This research is supported by Resource Grant NIH NIGMS 5P41GM103422 and NIH NIGMS 1S10OD016298. I would like to thank Drs. K. Gill, H. Wang, S. Chemuru, H. Rohrs, and D. Rempel for their help in sample preparation, data collection, data analysis, and insightful advice.
References

47. Tong, J., Taylor, P., Peterman, S.M., Prakash, A., Moran, M.F.: Epidermal growth factor receptor phosphorylation sites Ser991 and Tyr998 are implicated in the regulation of receptor endocytosis and phosphorylations at Ser1039 and Thr1041. Mol Cell Proteomics. 8, 2131-2144 (2009)
55. Huang, F., Sorkin, A.: Growth factor receptor binding protein 2-mediated recruitment of the RING domain of Cbl to the epidermal growth factor receptor is essential and sufficient to support receptor endocytosis. Mol Biol Cell. 16, 1268-1281 (2005)


Chapter 7

Protein Footprinting Elucidates the Protein-Membrane Association of Nanodiscs-Incorporated KRAS4b Proteins
7.1 Abstract

KRAS4b protein is a member of the RAS family that regulates signaling pathways essential for cell growth, it becomes functional when interacting with the inner surface of a plasma membrane. This protein is frequently mutated in cancer. Previous studies suggest that KRAS4b primarily interacts with a membrane through its farnesylated hypervariable region (HVR). Because little is known about the structure of the protein-membrane complex and the specific protein-lipid interactions, we describe here incorporating KRAS4b into a Nanodisc, affording a system more amenable to biophysical assays such as Mass Spectrometry footprinting and NMR. We footprinted Nanodiscs-bound and unbound KRAS4b protein by taking advantage of Fast Photochemical Oxidation of Proteins (FPOP) and glycine ethyl ester (GEE) labeling at the residue-level, taking advantage of the irreversible nature of these modifications. The irreversibility provides flexibility in removing the lipids and proteolyzing the protein to provide a coarse-grained picture of this important protein-membrane interaction. One aim is to establish a platform suitable for conformation studies of other membrane proteins.

7.2 Introduction

RAS proteins are members of the small GTPase family that regulate essential cellular functions including proliferation, differentiation, apoptosis, and mobility, by switching between inactive GDP-bound state and active GTP-bound state [1-5]. The three most commonly expressed RAS proteins in human are NRAS, HRAS, and KRAS, among which KRAS is found as two splice variants designated as KRAS4a and KRAS4b [6, 7]. The catalytic domain of RAS proteins (G-domain, amino acids 1-166, Figure 7-1 A) share
highly homologous sequences and structures; their hypervariable regions (HVR, amino acids 167-C termini, Figure 7-1 B), on the other hand, differ markedly in sequence and post-translational lipidation [3, 8] (Figure 7-2). Such lipidation of RAS proteins, including palmitoylation and farnesylation (Figure 7-2), facilitates the membrane binding of RAS, provides a way to reversibly regulate protein localization and interactions [9, 10]. RAS mutations, typically at codons 12, 13, and 61, are associated with approximately 30% of all human cancers [11]; the mutations of KRAS alone, however, account for 85% of all the oncogenic RAS mutations [12, 13]. This makes KRAS a good anticancer target; however, no clinically effective inhibitors are available so far [14].

KRAS4b is found frequently mutated in human colorectal and lung cancers, and in almost all pancreatic cancers [15]. The HVR of KRAS4b, which undergoes post-translational farnesylation, proteolytic cleavage and carboxymethylation at the C-terminal cysteine Cys185 [16], contains a polybasic lysine motif and anchors KRAS4b to anionic lipid rafts on the intracellular side of the plasma membrane [2, 17] (Figure 7-2). This membrane localization is critical for KRAS4b downstream signaling pathways, offering potential therapeutic strategies for inhibition of oncogenic signaling [18, 19]. Although the significance of KRAS4b-membrane association is well acknowledged [2, 20-22], there is no atomic-level structure revealing how KRAS4b interacts with the membrane.

The major challenge is determining the structural and biochemical properties of membrane-associated proteins that are difficult to crystallize. Thus, the membrane-anchored structural information is from some electrostatic calculations [23, 24], or low-resolution analyses including infrared spectroscopies [25] and fluorescence [26]. Within
the last decade, solid-state NMR, as a technique well suited for structural and dynamic studies of membrane-associated peptides and proteins, was first applied to provide the structural basis of membrane insertion of lipidated RAS [27-29]. Recently extensive molecular dynamics simulations have been focusing on both isolated RAS HVR-lipids interactions and full-length, post-translationally modified RAS-membrane interactions [2, 30-33], these studies complement the experimental results and have unraveled some of the basic structural and thermodynamic features of RAS-membrane interactions.

An additional problem for studying membrane proteins is the need to work in aqueous media and solubilize the protein in its native environment (i.e., lipid bilayers). Nanodiscs provide one solution to this problem. They are nano-sized, disc-shaped, phospholipid bilayers wrapped by two molecules of an amphipathic alpha-helical protein [34] (Figure 7-3 A). The use of a Nanodisc as a carrier to incorporate recalcitrant membrane proteins into lipids bilayer can preserve the structure and activity of membrane proteins [35]. Nanodiscs are ideal surrogates for cell membranes and allow lipid compositions to be specified flexibly. Nanodiscs are prepared by mixing together phospholipid/detergent micelles and membrane scaffold proteins (MSP), followed by detergent removal (Figure 7-3 B). The size of Nanodiscs is dependent on the length of the MSP and the stoichiometry of lipids used in the self-assembly process [36]. Incorporating KRAS and membrane proteins in Nanodiscs would make it more amenable to a variety of techniques that have been developed for soluble proteins. In our case, it facilitates the mass spectrometry (MS)-based conformational characterizations of KRAS4b proteins.
The MS-based protein footprinting techniques provide peptide and residue-level information to elucidate protein-protein or protein-ligand interactions. Typical strategies include labeling of a protein in two or more states (e.g., bound versus unbound) followed by enzymatic digestion of labeled proteins, and comparisons of the peptidal fraction modified between states. Decreased fraction modified of bound state relative to that of unbound state is often observed, owing to reduced solvent-accessibility at interacting regions. Two representative MS-based protein footprinting approaches are hydrogen-deuterium exchange (HDX) and hydroxyl radical labeling. HDX has been extensively used for soluble proteins, as it monitors the deuterium exchange extent of amide backbones which reports the solvent accessibility; however, for membrane-bound proteins, HDX becomes challenging owing to low protein solubility, poor digestion efficiency, and increased back-exchange [37, 38]. Besides, as discussed in Chapter 6, it can be challenging to characterize membrane proteins using HDX because these proteins are dynamic. Numerous laboratories have made progresses in overcoming these difficulties [39-41]. Hydroxyl radical labeling, firstly established by Chance and co-workers [42] and elaborated as fast photochemical oxidation of proteins (FPOP) by Hambly and Gross [43], is advantageous because the labeling is fast and irreversible, enabling more rigorous sample handling downstream of the labeling process prior to MS detection, such as sample cleanup, proteolysis, and chromatography. Konermann et al. [44] first demonstrated the feasibility of FPOP on membrane proteins with bacteriorhodopsin in a natural lipid bilayer environment, they observed more oxidative modifications on solvent-accessible loop regions compared to the radical-inaccessible transmembrane regions.
We recently reported the use of FPOP footprinting of a Nanodiscs-incorporated membrane protein light harvesting complex 2 (LH2), to probe its conformation and topology [45].

In this chapter, we describe the use of FPOP to footprint full-length, post-translationally modified KRAS4b protein in a near-native environment facilitated by its incorporation in lipid Nanodiscs. An important feature of FPOP is the adjustable labeling time controlled by different concentrations of radical scavengers [46]. We took advantage of this feature and added a reporter peptide prior to labeling by hydroxyl radicals, to normalize the labeling yield and enable a kinetic-type output. We also employed a residue-specific footprinting, glycine ethyl ester (GEE) labeling, to corroborate the FPOP outcome.
Figure 7-1. (A) Crystal structure of KRAS4b with HVR truncated (amino acid 1-166), PDB 4l8g; (B) Sequence alignment of amino acids in NRAS, HRAS, KRAS4a, and KRAS4b. The catalytic domain (residues 1-166) of the four isoforms shares highly conserved sequence identity (~89%), whereas the HVR domain (residues 167-185/6) has low sequence identity (8%). Cys residues in blue denote the palmitoylated Cys, and the terminal Cys with farnesylation is colored in red. The HVR of KRAS4b is characteristic of its polybasic lysine motif.
Figure 7-2. Post-translationally lipidated cysteine residues in HVR of NRAS, HRAS, KRAS4a, and KRAS4b. The NRAS, HRAS, and KRAS4a contain the palmitoylated cysteine residues besides the terminal cysteine farnesyl methyl ester; the KRAS4b, on the other hand, has a polybasic lysine stretch additional to the terminal cysteine farnesyl methyl ester.

Figure 7-3. (A) A schematic representation of a Nanodisc. The two alpha-helical membrane scaffold protein molecules are shown in blue and green. The polar head groups of the bilayer phospholipids are shown in orange. The hydrophobic tails are in light tan. Reprinted from Shih, A. Y. et al. Nano Lett. 7, 1692-96 (2007); (B) A cartoon representation of the process of Nanodisc assembly. Retrieved from https://www.labome.com/method/Nanodiscs-Membrane-Protein-Research-in-Near-Native-Conditions.html#ref1.

7.3 Experimental and Methods

7.3.1 Material
Leucine enkephalin acetate hydrate, L-histidine, L-methionine, 30% hydrogen peroxide, catalase, acetone, urea, dithiothreitol (DTT), iodoacetamide (IAM), 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris base), glycine ethyl ester (GEE), 1-ethyl-3-(dimethylaminopropyl)-carbodiimide (EDC), 36.5-38.0% hydrochloric acid, ammonium acetate, formic acid (FA), trifluoroacetic acid (TFA), HPLC-grade solvent, and phosphate buffered saline (PBS) were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). Sequencing-grade chymotrypsin was purchased from Promega Co. (Madison, WI). The POPC and POPS lipids were purchased from Avanti Polar Lipids (Alabaster, AL) in chloroform and used without further purification. Lipid concentration was determined by Avanti using total phosphate analysis. Bio-Beads SM-2 was purchased from Bio-Rad (Hercules, CA).

### 7.3.2 Production of Farnesylated and Methylated KRAS4b

Farnesylated and methylated KRAS4b (KRAS-FMe) was expressed and purified as described previously (Gillette et al., 2015, Scientific Reports). Specifically, FNL-Tni cells were infected with baculovirus that was engineered to express farnesyl transferase subunits A and B in addition to His6-maltose binding protein tagged KRAS4b (1-185). Cell pellets were lysed mechanically by three passes at 4,000 psi (Microfluidizer M-110EH, using a high-pressure instrument Microfluidics Corp, Newton, MA). Lysate was clarified by ultracentrifugation (70K x g, 30 min) and stored at -80°C until purification. Clarified lysate was adjusted to 35 mM imidazole and loaded onto an IMAC (immobilized metal affinity chromatography) column. The equilibration buffer for the column was 20 mM HEPES, pH 7.3, 300 mM NaCl, 5mM MgCl₂, 1 mM TCEP, 35 mM imidazole and 1:1000
protease inhibitor. Bound proteins were eluted using a gradient of gradient 35 mM to 500 mM imidazole. Pooled fractions were dialyzed to 20 mM MES, pH 6.0, 200 mM NaCl, 5 mM MgCl₂, 1 mM TCEP. After overnight dialysis at 4°C, the dialyzed sample was diluted 1:1 (v/v) with 20 mM MES, pH 6.0, 5 mM MgCl₂, and 1 mM TCEP (to reduce the NaCl to 100 mM) and immediately loaded to an SP Sepharose column equilibrated in 20 mM MES, pH 6.0, 100 mM NaCl, 5 mM MgCl₂, and 1 mM TCEP. Bound proteins were eluted with a gradient of 100 mM to 650 mM NaCl. Pooled fractions were digested with Tev protease to remove the His6-maltose binding protein fusion tag. The digested sample was loaded on an IMAC column to separate farnesylated and methylated KRAS4b from the His6-MBP and His6 tagged Tev protease. Protein from this step was dialyzed to a final storage buffer of 20 mM HEPES, pH 7.3, 300 mM NaCl, 5 mM MgCl₂, 1 mM TCEP, dispensed in small (250 ul or less) aliquots, flash frozen in liquid nitrogen and stored at -80°C.

7.3.3 Membrane Scaffold Protein (MSP del H5)

*E. coli* BL21 DE3 (pRARE) cells harboring plasmid R712-X04-3 (expressing His6-tev-pMSPdelH5) were grown in Terrific Broth liquid medium at 37°C and induced with 0.5 mM IPTG for 3 hr. Cells were lysed mechanically by three passes at 10,000 psi through a high-pressure instrument (Microfluidizer M-110EH, Microfluidics Corp, Newton, MA). Lysate was clarified by ultracentrifugation (70K x g, 30 min) and stored at -80°C until purification. The protein was captured from the lysate by IMAC, dialyzed to a final buffer of 20 mM Tris, pH 7.5, 100 mM NaCl, and 0.5 mM EDTA, dispensed in 1 ml aliquots, and stored at -80 °C.
7.3.4 Nanodisc Preparation and Incorporation

The Nanodiscs were prepared based on a published protocol with minor modifications [1]. Briefly, POPC and POPS stock lipids in chloroform were mixed together in 70:30 POPC/POPS ratio and slowly dried using a low flow of argon gas while incubating in a bead heat bath at 37 °C. The dried lipid was reconstituted to 65 mM with 130 mM cholate in 20 mM HEPES, 100 mM NaCl, and 0.5 mM EDTA buffer, and mixed with MSP delH5 in a 30:1 lipid to protein ratio. The mixture was mixed on a Nutator at 4 °C for 1 hour, and then 1 g/mL of washed bio-beads was added and incubated at 4 °C for additional 4.5 hours on a Nutator. Afterwards, the Nanodiscs were removed from the biobeads by careful pipetting and further purified using gel filtration chromatography on an AKTA FPLC with a GE Superdex 200 increase column (10x300 mm) running with the buffer containing 20 mM HEPES and 100 mM NaCl at 0.5 mL/min flow rate.

The Nanodiscs and KRAS4b-FMe were dialyzed into PBS buffer (pH 7.4) using Slide-A-Lyzer™ MINI Dialysis Device (MWCO 10kDA, Thermo Scientific, Waltham, MA). Afterwards the Nanodiscs and KRAS4b-FMe in PBS buffer were concentrated to 100 µM using Amicon Ultra 0.5mL centrifugal filters (MWCO 3kDA, EMD Millipore, Billerica, MA). The nanodisc-bound KRAS4b-FMe complex was formed by incubating 100 µM KRAS4b-FMe with an equimolar equivalent of Nanodiscs at room temperature with gentle vertexing for 10 min. A control sample of unbound KRAS4b-FMe was incubating with PBS buffer in the absence of Nanodiscs in the same manner [36].

7.3.5 FPOP Footprinting
Freshly prepared Nanodisc-bound and unbound KRAS4b-FMe were stored in a 4 °C refrigerator and used on the day of footprinting. A total volume of 50 µL of solution containing 5 µM protein, 8 µM Leu-enkephalin, 20 mM H₂O₂, and histidine at various concentrations (0.1, 1.0, 5.0, 20 mM) in PBS, was loaded in a 100 µL syringe (Hamilton, Reno, NV) and advanced using a syringe pump (Harvard Apparatus, Holliston, MA). That volume was flowed and submitted to approximately 1000 laser shots. The FPOP protocol was conducted by incorporating a reporter peptide, as previously described (Chapter 3) [47]. The excimer laser (GAM Laser Inc., Orlando, FL) was adjusted to 23 mJ/pulse at a frequency of 7.4 Hz and focused to an irradiation window of 2.9 mm. The solution flow rate was adjusted to ensure a 25% exclusion volume fraction to avoid repeated exposure of radicals [48, 49]. Each replicate was collected in a low protein-binding Eppendorf containing 50 nM catalase and 70 mM methionine, preventing further oxidation by residual H₂O₂ and other ROS. Control samples were handled in the same manner without laser irradiation. Each collection was aliquoted into two portions, with one portion directly submitted to online desalting followed by MS analysis of the intact (undigested) protein; and the other portion underwent acetone precipitation, dissolution, proteolysis, followed by bottom-up MS analysis as described below [50, 51].

7.3.6 GEE Labeling

Fresh batch of Nanodisc-bound and unbound KRAS4b-FMe were prepared to guarantee sample quality, and stored at 4 °C in a refrigerator until the day of labeling. GEE labeling experiments were conducted as described previously [52]. An aliquot of 10 µL of protein samples (5 µM in PBS) was incubated with 1 µL of GEE (2 M) and 1 µL of EDC (50 mM)
at room temperature for various times (2, 6, and 10 min) to achieve time-dependent carboxylic acid labeling [53]. The reaction was quenched by adding 12 µL of ammonium acetate (1 M). All experiments were performed in triplicate. Directly following modification, samples were precipitated with acetone and submitted to proteolysis.

7.3.7 Proteolysis

Sample pallets from precipitation were resuspended with 8 M urea, treated with DTT (10 mM) and IAM (20 mM) sequentially, and diluted with 50 mM Tris buffer (pH 8.1) prior to chymotryptic digestion. DTT and IAM were used to break disulfide bonds and cap the free cysteines. All samples were incubated with 10:1 protein: enzyme (by weight) for 12 h (37 °C, 350 rpm). The digestion was quenched by adding 2% (by volume) FA. After digestion, 20 µL of digests from each sample was loaded in autosampler vial for bottom-up analysis.

7.3.8 Intact Measurements

FPOP modifications of reporter peptide and that of KRAS4b-FMe at protein level were determined by using a MaXis 4G Q-TOF (Bruker, Bremen, Germany) coupled to a custom-built, on-line sample-handling device. In brief, 10 µL of labeled protein with reporter peptide (~50 pmol) was loaded onto a C8 reversed-phase trapping column (ZORBAX Eclipse XDB, 2 × 15 mm; Agilent Technologies, Santa Clara, CA) for a three-minute desalting with a flow of 200 µL/min water containing 0.1% (by volume) TFA. The desalted protein and reporter were then eluted with a two-step linear gradient at 200 µL/min. Specifically, the gradient was delivered from 2% to 10% solvent B (acetonitrile
with 0.1% FA) for 3 min to elute the reporter peptide, then quickly increased to 100% solvent B in 5 min during which time the protein elutes.

7.3.9 Bottom-up Analysis

The LC-MS/MS analyses were conducted back-to-back with the UltiMate 3000 Nano LC system and a Thermo Q Exactive Plus Hybrid Quadruple-Orbitrap Mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Aliquots of the digests (5 µL) were drawn from autosampler vial and loaded onto a C18 reversed-phase trapping column (Acclaim® PepMap100, 100 µm x 2 cm, 5 µm, 100 Å; Thermo Fisher Scientific, Waltham, MA) at 4.5 µL/min with water containing 0.1% FA for a ten-minute desalting. Peptides were separated by using a custom-packed C18 reversed-phase column (Symmetry, 100 µm x 200 mm, 5 µm, 100 Å; Waters Inc., Milford, MA) and eluted with a 76-min gradient delivered at 500 nL/min. The scan range of precursor ions was set from m/z 380 to 2200 for all samples with high mass resolving power of 70,000. Automatic gain control was used to prevent over-filling of the trap. Two different MS acquisition methods were set for the FPOP-labeled and GEE-labeled samples respectively. For FPOP-labeled samples, MS was operated in non-targeted data-dependent-acquisition (DDA) mode (i.e., the 18 most abundant precursor ions corresponding to peptide elution per scan were subjected to the high-energy C-trap dissociation (HCD) in the HCD cell to obtain product-ion mass spectra. The normalized collision energy was set to 35%. Dynamic exclusion was activated for 8 s following each scan to enable fragmentation of low abundant peptides. The maximum injection time (IT) for a full mass spectral scan was set to 200 ms, and 150
ms for each MS/MS scan. For GEE-labeled samples, MS was operated in targeted DDA mode.

7.3.10 Targeted Data Acquisition

The protein amount used in GEE labeling was 5 times less than that used in FPOP (50 pmol vs. 250 pmol), given that GEE produced relatively simple modified products, a targeted DDA acquisition mode was adopted for GEE-labeled samples to ensure good identification and sequence coverage [54]. To be specific, an inclusion mass list comprising all precursor ions of interest was created and saved as CSV format via a custom-built web application [55]. The file was then imported into the global parent mass list. Details of inclusion list creation are described in Chapter 4. For each full mass spectral scan, the ten top most abundant ions matching to an m/z with a threshold of ±7.5 ppm on the inclusion list were selected for HCD fragmentation. Monoisotopic precursor-ion selection was enabled. The maximum IT for product-ion scans was set to 500 ms. Dynamic exclusion was turned off to maximize the number of product-ion mass spectra obtained for each “matched” precursor-ion.

7.3.11 Data Analysis

The *.raw MS files were searched directly using Byonic (Protein Metrics Inc., San Carlos, CA) against a custom-built database containing the sequence of KRAS4b [16]. All known HO• side-chain reaction products were added to the variable modification database for searching the HO•-modified samples [56]. The two possible forms of modifications from GEE labeling [52] (i.e., amide form (mass shift of 57.0215 Da) and hydrolyzed form (mass shift of 85.0528 Da)), were added to the variable modification database for searching of
GEE-modified samples. The farnesylation and methylation of Cys185, which results in mass increase of 218.2035 Da, was added as variable modifications. The alkylation of the sample with IAM to add a carbamidomethyl group (added mass of 57.0214 Da) to Cys-containing peptides, was included as a fixed modification for searching. The search tolerance window was 15 ppm for precursor ions, and 50 ppm for product ions. Y, F, W, M, L were selected as fully specific cleavage sites in accord with chymotrypsin proteolysis. 

The quantitation of fraction modified was based on ion abundances from extracted ion chromatograms (XICs). Time-dependent signals from LC-MS for each unmodified peptide and its modified species were extracted from raw data files using Xcalibur™ Software (Thermo Fisher Scientific, Waltham, MA) with a mass tolerance of 15 ppm. The peak integrals corresponding to both unmodified species ($I_{unMod}$) and modified species ($I_{Mod}$) are readily attainable. The fraction modified for each peptide was determined as per Eq 7-1:

\[
Fraction \ modified = \frac{\sum I_{Mod}}{\sum I_{Mod} + I_{unMod}} \quad \text{Eq 7-1}
\]

In a few cases where assignment to a single site was ambiguous, either owing to interference from isomers co-elution or incomplete fragmentation in MS/MS, the modification was assigned to a set of possible residues.

7.4 Results and Discussion

7.4.1 Need for Reporter-incorporated FPOP

Hydroxyl radicals are extraordinarily reactive and short-lived [57]. The lifetime of hydroxyl radicals in FPOP, which is closely associated with fraction modified of samples, is
predominantly affected by the identities and/or concentrations of radical scavengers [46]. In the FPOP footprinting of Nanodiscs-bound and unbound KRAS4b-FMe, we used hydroxyl-radical scavengers under identical conditions for both unbound and bound states. Nevertheless, multiple species from Nanodiscs (e.g., MSP, phospholipids) are likely to give “differential scavenging”, affording a less modified fraction for the Nanodisc-bound samples, and consequently a biased comparison between the two states when a simple one-labeling-FPOP experiment was performed. To illustrate and compensate for differential scavenging, we varied the FPOP labeling time by using different concentrations of histidine at 0.1, 1.0, 5.0, and 20 mM, and introduced Leu-enkephalin in FPOP as a reporter peptide (Figure 7-4). The yield of hydroxyl radical under each condition can be reported as the fraction modified Leu-enkephalin (y-axis). Despite identical histidine conditions, the fraction modified of reporter peptide from Nanodiscs-bound samples was constantly less than that from Nanodiscs-unbound samples. This straightforwardly indicates that the dose of HO•, and therefore the radical lifetime, in Nanodiscs-bound samples was intrinsically lower due to the presence of Nanodiscs, clearly indicating a differential scavenging effect that became less pronounced at higher concentrations of histidine. At 20 mM of histidine, where the scavenging by Nanodiscs is overwhelmed by that of histidine, the fraction modified of reporter between the two states is similar (Figure 7-4). Extensive scavenging in FPOP by a high concentration of scavenger, however, is not the solution to differential scavenging as it would provide little labeling of the proteins. The reporter peptide, on the other hand, can “compensate” for the different modification extents for different states as we adjust among various radical
lifetimes. The outcomes are “FPOP kinetic curves” that describe the time-dependent response of FPOP modification for each peptide; here the time domain is proportional in a complex way to the fraction modified of reporter peptide [47].

![Image](image.png)

**Figure 7-4.** Fraction modified of Leu-enkephalin (reporter peptide) under different FPOP labeling times as adjusted by varying scavenger concentrations (histidine concentrations: 0.1, 1.0, 5.0, 20 mM). The error bars correspond to ± SE of triplicate measurements.

### 7.4.2 Analyses of the Intact Reporter and Protein

Following footprinting, we analyzed the Leu-enkephalin and undigested KRAS4b-FMe by LC-MS, integrating their respective TIC signals over an elution window of 0.5 min. The reporter peptide eluted earlier than protein because it is less hydrophobic. To determine the fraction modified of reporter, signals corresponding to the unmodified and singly-oxidized Leu-enkephalin were extracted with a mass tolerance of ±7.5 ppm (Figure 7-5). The resulting XIC of singly-oxidized Leu-enkephalin appeared as one Gaussian-shaped
peak, suggesting co-elution of different isomers of singly-oxidized species, given that Tyr1, Phe4, and Leu5 of Leu-enkephalin have been previously identified to be modified in FPOP (Chapter 3) [47]. We did not, however, observe significant doubly-oxidized Leu-enkephalin signals. Integrals of the resulting two XIC peaks can be obtained using the Data Analysis Software (Bruker, Bremen, Germany). The fraction modified of reporter peptide was calculated per Eq 7-1.

Measurements of the intact KRAS4b-FMe at various histidine conditions demonstrate the effectiveness of adjusting FPOP reaction times. The ESI mass spectra of the most abundant charge state (+27) show clearly the effect of varying the time (Figure 7-6). Within each spectrum, the peaks corresponding to the unmodified protein are followed by that of singly-oxidized species (+15.9949 Da), doubly-oxidized species (+31.9898 Da), etc., as a consequence of one or more substitutions of H for OH. The control experiments show that the protein itself was slightly oxidized before laser irradiation; this portion of background oxidation is subtracted from total FPOP modifications. As the histidine concentration increases, the labeling time decreases, and the fraction modified of KRAS4b-FMe at the protein level decreases accordingly.
Figure 7-5. FPPO labeling extent quantitation of Leu-enkephalin. (A) Extracted ion chromatograms (XICs) of unmodified (dark green) and singly-oxidized (red) reporter peptide. Integrals of XICs of unmodified and modified species allow quantitation of fraction modified of reporter peptide; (B) Mass spectra of unmodified and singly-oxidized reporter peptide.
Figure 7-6. Mass spectra from measurements of intact, FPOP-labeled KRAS4b-FMe (+27 charge) obtained by adjusting histidine concentrations.

7.4.3 Time-dependent FPOP at Peptide and Residue Level

Varied labeling times of FPOP allow the measurements of kinetics. In an effort to normalize the differential scavenging and compare the FPOP outcomes between Nanodiscs-bound and unbound KRAS4b-FMe at peptide/residue level, we plotted the
fraction modified of Leu-enkephalin at each histidine concentration versus the corresponding fraction modified for each chymotryptic peptide of the protein (y-axis). The fraction-modified of reporter peptide acts as a “ruler” of HO- lifetime (i.e., longer times yield more modification), so the resulting FPOP-reporter kinetic plots reveal the time-dependent KRAS4b-FMe modification, enabling an output resembling that of the familiar HDX kinetics data for peptides. It is noteworthy, however, that the dependence between the time and yield of reporter peptide does not follow a simple, linear relationship, as has been elaborated previously [47]. For each peptide, the two curves representing different protein states (Nanodiscs-bound and unbound) should overlap if the solvent-accessibilities of them are the same (Figure 7-7). The shadowed region corresponds to the 0.95 confidence intervals for each curve. Regions where the curves diverge are those for which solvent accessibility and/or conformations change. The FPOP-reporter kinetics reveal, at KRAS-FMe peptide level that, most regions (from N-terminus to 157) share similar solvent-accessibility between the Nanodiscs-bound and unbound states, indicating that a majority of this protein is not involved in an association with a lipid membrane. Nevertheless, decreases in modification extent (solvent-accessibility) do reveal for various labeling times that region spanning from residues 158-185, as represented by three peptides: 158-170, 171-185, and 181-185 are membrane-associated (Figure 7-7). The significantly reduced reactivity in this region demonstrates that it is this region that is in close proximity to the membrane.

The use of LC-MS/MS enables the separation of HO- modified peptides based on their hydrophobicity and the measurement of their differential solvent accessibility.
Furthermore, with MS/MS, we can also identify some amino-acid site-specific modifications. To demonstrate, we show the extracted ion chromatograms (XICs) and precursor-ion spectra of the unmodified and singly-oxidized species of peptide 20-40 (Figure 7-8 A, B). The multiple chromatographic peaks observed for the oxidized species (+15.9949) correspond to the isomers eluting at different time. Manual interpretation of the product-ion spectra localizes modifications on residues H27, F28, and Y32 (Figure 7-8 C, and Figure 7-10). With high confidence, we identified other residue-specific oxidative modifications (Figure 7-9). More than half of these identified residues have side-chains that are either sulfur-containing (M72, M170) or aromatic (H27, F28, Y32, Y64, F78, Y137), not surprising because those residues are highly reactive towards HO•. Other residues including L52, L120, I142, K176-179, K182, K184, and C185 also undergo detectable modifications. These “magnified” residue-specific views improve the spatial resolution of the interacting region. For example, for peptide 158-170, Met170 is the major residue showing differential labeling between the two states. Another region, represented by C-terminus peptide 171-185, containing the farnesylation and methylation on Cys185, shows protection upon Nanodiscs binding. Given the amino acid composition, this peptide is relatively less reactive towards HO•; however, it shows more modification (>10%) in the absence of Nanodiscs, indicating its solvent accessibility (Figure 7-7). We identified oxidative modifications on K176, K177, K178, and K179 for this peptide. The product-ion spectra show mixed signature ion species (Figure 7-10), indicating the co-elution of singly-oxidized isomers. Besides peptide 171-185, a shorter peptide 181-185, resulted from chymotrypsin non-specific cleavage at C-terminal of Lys180, was also identified. The
C-terminal farnesyl group renders this five-residue peptide, which otherwise might be eluted with solvent front, rather hydrophobic, and FPOP results show divergence in between the two states. K182, K184, and C185 were found to be labeled on this peptide.

It is worth mentioning that, the farnesyl group (\(-C_{15}H_{24}\)) often is lost as a neutral (204.1878 Da) in MS/MS experiments of peptides that contain it [58, 59]. Failure to account for this neutral loss could lead to misinterpretation of spectra or even loss of sequence coverage. For example, without accounting for the neutral loss of farnesyl group, we could only identify some b-series ions for the C-terminal peptide 171-185, which contains the farnesylated and methylated Cys185 (Figure 7-11 A). Because all y-series ions are shifted by the neutral loss of farnesyl group, the identification is greatly improved by considering the farnesyl group as neutral loss on Cys185 (Figure 7-11 B), using our custom-built web-application software (See Chapter 4 for details).

Overall, the FPOP results show that the G-domain of KRAS4b-FMe is not involved in the interaction with Nanodiscs; the HVR domain, on the other hand, shows noticeable protection when the protein is incorporated in a Nanodisc. This result is in agreement with the proposal that the HVR of KRAS protein targets the negatively charged phospholipids of the plasma membrane via its polybasic lysine motif [60], and that prenylation of HVR helps anchor the protein to plasma membrane [10, 61]. Recent studies using extensive atomic scale MD simulations of KRAS4b HVRs at an anionic lipid bilayer composed of DOPC and DOPS (mole ratio 4:1) illustrate that the KRAS4b preferentially tethers the membrane in the liquid phase and spontaneously inserts the farnesyl moiety into the loosely packed phospholipid bilayers [62].
Figure 7-7. Response curves of each peptide from Nanodiscs-bound and unbound KRAS4b-FMe in FPOP time-dependent measurements performed by varying the histidine concentrations (0.1, 1.0, 5.0, and 20 mM). The error bars correspond to ± SE of triplicate measurements. The shadowed region of each curve represents the confidence interval region corresponding to one SE.
Figure 7-8. MS/MS of peptide 20-40 illustrates the MS detection of unmodified and modified peptide species, the chromatographic separation of various modified species, and localization of the modified residue. (A) XICs of unmodified and singly-oxidized peptide 20-40; (B) mass spectra of precursor ions (+3 charge) corresponding to unmodified and singly-oxidized peptide 20-40; (C) MS/MS spectrum corresponding to singly-oxidized peptide 20-40 identifies His27 modification.
Figure 7-9. Residue level kinetic curves of Nanodiscs-bound and unbound KRAS4b-FMe. Residues were identified with high confidence and plotted as the fraction modified of each residue vs. the fraction modified of reporter peptide under each histidine concentration. The error bars correspond to ± SE of triplicate measurements. The shadowed region of each curve represents the confidence interval region corresponding to one SE.
Figure 7-10. Product-ion (MS/MS) spectra of various peptides used to locate the sites of FPOP modification.

Figure 7-11. Annotated product-ion (MS/MS) spectra corresponding to the FPOP-modified peptide 171-185 for which the Cys185 is both farnesylated and methylated. (A) Only b-series ions are identified when the neutral loss of farnesylation is not included; (B) All y-ions and molecular ions reveal the neutral loss of farnesylation from the C-terminus.
7.4.4 Glycine Ethyl Ester (GEE) Footprinting

To corroborate the findings in the reporter-incorporated FPOP experiments, we employed GEE footprinting as the second approach. This footprinting specifically modifies solvent-accessible carboxyl side-chain of aspartate (D) and glutamate (E) by an amidation reaction with glycine ethyl ester. GEE footprinting was previously validated and first used to probe protein orientation of a membrane-associated protein [52, 63]. As mentioned earlier, we used the targeted DDA acquisition mode to increase the number of product-ion spectra per peptide. Similar to FPOP, we took time-dependent GEE measurements not only to provide assurance that the modification reactions occur normally but also to add statistical weight to the two-state comparison. We quantified the extent of modification based on verified XIC features corresponding to the unmodified and modified chymotryptic peptides at various labeling times (Figure 7-12 show the fraction modified of each peptide at various labeling times). For most peptides, the fraction modified increases monotonically with labeling, suggesting the absence of GEE-induced protein unfolding [53]. Peptides 83-96 and 158-170, however, showed that the fractions modified reach a plateau with no time-dependence, indicating that these reactions are complete even prior to the earliest labeling time point (i.e., 2 min).

To illustrate better the differential fraction modified in between the Nanodiscs-bound and unbound KRAS4b-FMe, we present the GEE footprinting results as time-dependent plots (Figure 7-13). The curves should overlap for regions that exhibit similar solvent accessibility in the bound vs. unbound states, whereas they will diverge if there are local changes in conformation or solvent accessibility. Most regions of KRAS4b-FMe
(represented by 7 peptides: 29-40, 41-53, 83-96, 97-111, 97-113, 114-137, and 142-157) report no differences for the two states (Figure 7-13), in accord with the FPOP results, indicating nearly identical solvent accessibility with and without the Nanodiscs. These regions must be remote from the lipid bilayer. Peptide 158-170, however, clearly has more protection in the bound state (Figure 7-13), indicating this region is involved in the protein-membrane association, in agreement with FPOP.

Besides the HVR, a second region represented by peptide 54-64 also showed reduced solvent accessibility in the bound state (Figure 7-13). This region is part of the P loop and Switch II loop of KRAS protein (Figure 7-1 A), which plays roles in binding of regulators and effectors [3, 8]. The peptide, abundant with Asp and Glu, undergoes significant modification in the unbound but less so in the bound state. A close look at the XICs and product-ion spectra showed well-resolved modifications on D54, D57, E62, and E63 (Figure 7-14, product-ion spectra not shown). This reduced modification does not occur with FPOP. The difference between FPOP and GEE may be due to changes in orientation of the acidic side changes. Indeed, there are documented residual orientation changes induced by either ligand binding or site mutation [64-66], and the changes here may be transparent to FPOP because acidic residues are relatively unreactive with OH radicals. An alternative explanation is the region represented by this peptide may exist as a crevice formed upon Nanodiscs incorporation between this peptide and membrane, or between this peptide and belt proteins. This crevice region should be more accessible to FPOP, given the small size of hydroxyl radicals, but its openings to solvent may be too small for EDC and the GEE molecules.
With the exception of region 54-64, GEE footprinting largely corroborates FPOP, showing with high confidence that the HVR of KRAS4b-FMe interacts with Nanodiscs. Combining both footprinting results, we conclude at residue level that the interacting region at least encompasses residues E162, E168, M170, K176, K177, K178, K179, and extends to the C-terminus of KRAS4b-FMe (including K182, K184, and C185, Figure 7-15), where the C185 farnesylation allows insertion of prenyl group into membrane. These findings agree well with recent molecular dynamics calculations that pinpoint residue-level interactions of KRAS4b with the POPC/POPS bilayer [17].

![Graph showing GEE labeling results](image)

**Figure 7-12.** Time-dependent GEE labeling results of Nanodiscs-bound (dark grey) and unbound (light grey) KRAS4b-FMe presented as bar-graph, where the x-axis represents the peptides and y-axis represents the GEE modification extents. The error bars correspond to ± SE of triplicate measurements.
**Figure 7-13.** Response curves of each peptide from Nanodiscs-bound and unbound KRAS4b-FMe in a GEE kinetics measurements performed by controlling different labeling times (2, 6, and 10 min). The error bars correspond to ± SE of triplicate measurements. The shadowed region of each curve represents the confidence interval region corresponding to one SE.

**Figure 7-14.** Extracted ion chromatograms (XICs) of unmodified and GEE-labeled peptide 54-64. Each of the four acidic residues is modified by GEE, the corresponding peptide species were separated chromatographically.
Figure 7-15. (A) Representation of Nanodiscs-bound KRAS4b-FMe, where the HVR domain (red) interacts with the membrane bilayers, leaving the G-domain (dark grey) extended outside the Nanodiscs; (B) Zoom-in view of HVR interacting with the Nanodiscs determined by FPOP and GEE labeling. Residues differentially modified by FPOP are labeled blue, residues modified by GEE are labeled green.

7.4.5 Complementary Biophysical Approaches

In order to gain more insights into the structure and orientation of KRAS4b bound to membrane, our collaborator utilized multiple other technologies including neutron reflectivity (NR), paramagnetic relaxation enhancement (PRE)-NMR, SAXS, and MD simulations, in addition to mass spectrometry. The outcome from these measurements are complementary to the footprinting results.

The neutron reflectivity is a technique in which neutrons at thermal energies are used to probe the near surface structure of a system [67]. Neutrons are incident on a surface at a grazing angle less than 3°; at such small angles, the potential for scattering can be approximated by a continuous value called scattering length density (SLD). The very small angles of incidence can be sufficient to cause the neutrons to be totally reflected.
from the surface, and the position of this transition compared to total reflection yields information about the average SLD of the material [68, 69]. In this experiment, the NR measurements were conducted for KRAS4b-FMe bound to tethered lipid bilayers, it provides one dimensional information perpendicular to the bilayer based on the neutron SLD profiles (Figure 7-16). The KRAS4b-FMe is found to have an extended conformation directed away from the lipid bilayer.

Complexes of the element Gadolinium (Gd) are widely used in NMR to enhance the paramagnetism. Because Gd is strongly paramagnetic, owing to its seven unpaired electrons in the 4f shell (Figure 7-17 A) [70], its electrons are not involved with bonding, so paramagnetism is maintained even in molecular complexes [71]. The PRE-NMR of the KRAS4b-Nanodiscs, enabled by assembling the Nanodiscs containing a Gadolinium-lipid chelate (Figure 7-17 B), shows significantly decreased NMR signals of $^{15}$N KRAS4b-FMe, for C-terminal region starting from Met170. This observation indicates the proximity of this C-terminal region to the membrane.

Small angle X-ray scattering (SAXS) is a powerful biophysical method providing information about the sizes and shapes of proteins and complexes in a broad range of molecular weights [72]. Our collaborators used SAXS to measure the scattering profile with variable D$_2$O to highlight lipid and protein components, then fitted the data to determine radius of gyration and protein shape. The SAXS outcome confirms the presence of extended conformation of KRAS4b-FMe on a Nanodisc.

Molecular dynamic (MD) simulations report on the physical movements of atoms and molecules, under the influence of defined physical forces [73]. It gives a view of the
dynamical evolution of the system by allowing the atoms and molecules to interact for a certain period of time [74]. MD simulation is particularly useful and informative when studying membrane proteins because these proteins are generally insoluble and dynamic, limiting the access to high-resolution characterizations by X-ray crystallography and NMR. The ensemble MD simulation results agree qualitatively with the NR and SAXS data that KRAS4b-FMe protein is oriented perpendicular to the membrane with minimal direct contacts to the bilayers.

Figure 7-16. (A) Schematic showing the tethered lipid bilayers prepared for neutron reflectivity experiments; (B) The geometry of a typical reflectivity experiment showing the incident and reflected angles, and the incident and reflected wave factors. Retrieved from https://www.ncnr.nist.gov/programs/reflect/NR_article/.
7.5 Conclusion

Interacting KRAS protein with lipid Nanodiscs allows characterization of proteins in a near-native environment. Unlike reversible HDX, the irreversible nature of FPOP and GEE modifications enables more rigorous sample handling prior to MS detection. This ability to remove the lipids post footprinting is particularly advantageous for membrane-bound proteins as it avoids problems in proteolytic digestion and column separation that must be done fast with HDX.

This mass spectrometry-based platform is suitable for differential studies of unbound and membrane-bound proteins, as enabled by the use of Nanodiscs and protein footprinting (i.e., FPOP and GEE). We were able to make meaningful two-state comparisons (Nanodiscs-bound and unbound KRAS4b-FMe) using time-dependent footprinting approaches. The incorporation of reporter peptide in FPOP normalizes the differential radical dosages between the two states, facilitating the use of FPOP to probe membrane-bound proteins and to allow unbiased FPOP comparisons.

Characterization of membrane-bound protein structures has always been challenging as evidenced by the small fraction of X-ray structures that are available. For KRAS proteins, the HVR is truncated prior to crystallization, removing the important HVR regions responsible for membrane association. Therefore, a platform involving Nanodiscs

Figure 7-17. (A) Electron structure of the Gd atom. Retrieved from http://mri-q.com/paramagnetic-relaxation.html; (B) Schematic of Nanodiscs containing Gd-lipid chelate and $^{15}$N labeled KRAS4b-FMe.
incorporation and irreversible footprinting is promising because it not only provides coarse-grained pictures of protein-membrane interactions, but it also offers opportunities to determine interactions of membrane-bound proteins with other proteins and ligands. We plan future studies to characterize the interactions between Nanodiscs-bound KRAS and its effector proteins; such interactions are pivotal in regulating downstream signaling [2, 16].
References


Chapter 8

Conclusions and Perspectives
The mass spectrometry-based hydroxyl radical footprinting, which has been further elaborated as Fast Photochemical Oxidation of Proteins (FPOP) in Gross’s lab, is advancing rapidly as a method of choice to unravel protein-ligand, protein-protein interactions. Expanding the applicability of FPOP is closely associated with the accompanying developments of methodology, establishing a foundation for a more robust, flexible, and quantitative footprinting platform.

To improve the current FPOP platform, we measured quantitatively by using an isotope-dilution GC/MS approach (reported in Chapter 2) the amount of HO• generated after each laser pulse. The determined initial hydroxyl radical concentration, 0.95 mM per laser pulse, serves as a key input and reference to any kinetic measurements and simulations in the future. We also demonstrated the possibility of the changing of effective dose of HO• by using different identities/concentrations of radical scavengers. This paves the way for the idea of incorporating a reporter peptide in FPOP, to compensate for the adventitious radical scavengers.

In Chapter 3, we described the incorporation of Leu-enkephalin as a reporter peptide in FPOP. The fraction modified of the reporter peptide allows not only the tracking of the labeling time under different sample solution conditions, but also the time-dependent measurements, which enable comparisons of FPOP outcome between two curves, as opposed to two points, adding statistical weights to the analysis. The incorporation of a reporter provides helps to establish unbiased FPOP footprinting comparisons and brings to the method broader applications and more demanding conditions.
FPOP is typically applied in an experiment through a comparative analysis of two or more protein states where differences in labeling extent is indicative of differences in solvent accessibility due to mutations or binding.

In Chapter 5, we compared the FPOP labeling extent of TFE-treated and untreated G93A SOD1, at both protein level and peptide/residue level to elucidate the effect of TFE-induced SOD1 destabilization. We found that upon destabilization structural changes in the protein occur including disruption of its dimer interface and local unfolding, evident by the significant increase of FPOP labeling extent with addition of 18% TFE. We proposed that the exposure of hydrophobic regions upon destabilization promote the self-association of G93A SOD1 molecules, facilitating its gain-of-function cytotoxicity.

In Chapter 6, we utilized both FPOP and HDX to characterize the conformations of EGF receptor proteins and complex. Results from both footprinting methods agree with previous MD simulation results that the C-terminal tail region of EGF receptor is intrinsically disordered. We found, however, that FPOP is a better approach for probing the binding interface between the EGF receptor and adaptor proteins (e.g., Cbl), owing to the fast ligand binding kinetics.

The combination of the FPOP platform and Nanodiscs permits interrogation of membrane protein conformations. In Chapter 7, we reported the marriage of the two techniques, enabling the characterization of the KRAS4b-membrane association. The FPOP results confirmed, with spatial resolution at the peptide and residue level, that the KRAS4b protein interacts with membrane through its C-terminal region, whereas the majority of KRAS4b is outside the membrane. We hope that a thorough understanding of the RAS-
membrane insertion will help decipher the oncogenic properties of RAS proteins and provide insight into the design of cancer therapeutics.

The successful method developments and applications of FPOP are bound with informatics tools that provide a fast, labor-saving, and interactive data-processing platform. Chapter 4 illustrated the potential of using R programming language to develop goal-oriented, user-friendly web applications, aiming at facilitating the output, visualization, and processing of mass spectrometry-based proteomic data.

We envision more methodologies on the basis of the current FPOP platform. For example, to adjust better the FPOP labeling times and to parameterize FPOP, a systematic study with varied parameters (e.g., laser intensity, \( \text{H}_2\text{O}_2 \) concentration, scavenger identity/concentration, buffer components, etc.) should be conducted, based on the dosimetry experiments and reporter-incorporated FPOP. To complement the labeling coverage, new labeling reagents in addition to the hydroxyl radical are to be developed. Those labeling reagents should take advantage of the fast, irreversible nature of FPOP and aim to modify/label FPOP-silent residues including Gly, Ala, Ser, Thr, and Asn.

FPOP combined with other techniques/devices brings in new opportunities and approaches to solve problems in biochemistry and biomedicine. For example, the pump/probe experiment, achieved by using a second IR laser for temperature-jump, should provide useful characterization input in studying the heat-induced destabilization and aggregation of SOD1 proteins. The combination of asymmetrical flow field-flow fractionation (AF4) and FPOP should allow on-line separation of macromolecules prior to
footprinting, affording good potential for characterizing heterogeneous protein aggregates (e.g., Aβ aggregates) with FPOP.