Spring 5-15-2018

Mass Spectrometry-based Strategies for Protein Characterization: Amyloid Formation, Protein-Ligand Interactions and Structures of Membrane Proteins in Live Cells

Ke Li
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Chemistry

Dissertation Examination Committee:

Michael L. Gross, Chair

Robert E. Blankenship

Weikai Li

Liviu M. Mirica

Gary J. Patti

Mass Spectrometry-based Strategies for Protein Characterization: Amyloid Formation, Protein-Ligand Interactions and Structures of Membrane Proteins in Live Cells

by

Ke Li

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

May 2018

St. Louis, Missouri

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 4

Chapter 5

Acknowledgments

I would like to express my special appreciation to my PhD advisor, Prof. Michael Gross for his scientific inputs, guidance and support that I received throughout my PhD study. He has set an example of excellence as a researcher, mentor, instructor, and role model. I have been extremely lucky to be his student and benefit from not only his exceptional scientific knowledge but also his extraordinary human qualities.

I would also like to thank my dissertation advisory and examination committee members, Prof. Liviu Mirica, Prof. Gary Patti, Prof. Weikai Li, and Prof. Robert Blankenship for their advice in my research and all the help in my thesis preparation.

I am grateful to my collaborators, Prof. Weikai Li and Dr. Shuang Li from WUSM, Dr. Adrienne Tymiak, Dr. Guodong Chen, Dr. Richard Huang, Ekaterina Deyanova from Bristol-Myers Squibb for our productive collaboration. A special thanks to Dr. Kwasi Mawuenyega for his constructive suggestions in my research in amyloid beta.

I would also like to thank all my colleagues in the lab for their support, encouragement and friendship, especially Don Rempel and Henry Rohrs for their assistance in data analysis and instrumentation. I am truly thankful for their always prompt help whenever I needed.

Finally, and without hesitation, I would like to thank my parents, Baoyin Li and Jianping Wu, to whom this thesis is dedicated for their belief in me.

This thesis research is supported by the National Institute of Health P41GM103422 and the graduate school of Washington University.

Ke Li

Washington University in St. Louis

May 2018

ABSTRACT OF THE DISSERTATION

Mass Spectrometry-based Strategies for Protein Characterization: Amyloid Formation, Protein-Ligand Interactions and Structures of Membrane Proteins in Live Cells

by

Ke Li

Doctor of Philosophy in Chemistry

Washington University in St. Louis, 2018

Professor Michael L. Gross, Chair

Mass spectrometry (MS)-based protein footprinting characterizes protein structure and protein-ligand interactions by interrogating protein solvent-accessible surfaces by using chemical reagents as probes. The method is highly applicable to protein or protein-ligand complexes that are difficult to study by conventional means such as X-ray crystallography and nuclear magnetic resonance. In this dissertation, we describe the development and application of MS-based protein footprinting from three perspectives, including I) protein aggregation and amyloid formation (Chapter 2-3), II) protein-ligand interactions (Chapter 4-5), and III) *in-cellulo* structures and dynamic motion of membrane proteins (Chapter 6). Fast Photochemical Oxidation of Proteins (FPOP) is the main methodology implemented in the work presented in this dissertation. Chapter 1 provides an overview of FPOP and discusses its fundamentals as well as its important applications in both academic research and biotechnology drug development.

In Part I, Chapter 2 covers the early method development of FPOP for monitoring amyloid beta (Aβ) aggregation. In this work, we demonstrated the high sensitivity and spatial resolution of the

method in probing the solvent accessibility of Aβ at global, sub-regional, and some amino-acid residue levels as a function of its aggregation, and revealed Aβ species at various oligomeric states identified by their characteristic modification levels. In Chapter 3, we extended the application of the platform to assess the effect of a putative polyphenol inhibitor on amyloid formation and to provide insights into the mechanism of action of the inhibitor in remodeling Aβ aggregation pathways.

In Part II, we evaluated different protein footprinting techniques, including FPOP, hydrogen-deuterium exchange (HDX), and carboxyl group footprinting, for probing protein-ligand (drug candidates) interaction in the context of a therapeutic development. Chapter 4 focused on protein-protein interaction by investigating the epitope of IL-6 receptor for two adnectins that have similar apparent biophysical properties. In Chapter 5, we probed the hydrophobic binding cavity of bromodomain protein for a small molecule inhibitor. This study serves as an example of interrogating protein-small molecule interactions. The two studies in Part II demonstrate the unique capabilities and limitations of protein footprinting methods in protein structural characterization.

In Part III, we pushed the boundary of MS-based protein footprinting by applying the method to footprint live cells and investigate the dynamic structures/motion of membrane-transport proteins in their native cellular environment. We employed protein engineering, suspension cell expression and isotopic-encoded carboxyl group footprinting to identify salt bridges in two proteins, GLUT1 and GLUT5, that control their alternating access motions for substrate translocation. With functional analysis and mutagenesis, live-cell footprinting provides new insights into the transport mechanism of proteins in the major facilitator superfamily.

The five studies in the dissertation demonstrate the powerful capability of MS-based protein footprinting in protein structural biology and biophysics research. The method also holds great potential in studying more complicated biological systems and solving demanding problems related to protein structure and properties.

Chapter 1: Introduction of Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization*

* This chapter is based on a recent accepted review: Li, K. S.; Shi, L.; Gross, M. L., Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization. *Acc. Chem. Res.* DOI: 10.1021/acs.accounts.7b00593

1.1 Abstract

Assessment of protein structure and interaction is crucial for understanding protein structure/function relationships. Compared to high-resolution structural tools including X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-EM, and traditional low-resolution methods such as circular dichroism, UV-Vis and florescence spectroscopy, mass spectrometry (MS)-based protein footprinting affords medium-to-high resolution (i.e., regional and residue-specific insights) by taking advantage of proteomics methods focused on the primary structure. The methodology relies on "painting" the reactive and solvent-exposed amino acid residues with chemical tags and using the pattern of modifications as footprints from analysis by bottom-up MS-based proteomics to deduce protein higher order structures. The outcome can refer to proteins in solution or even in cells and is complementary to those of X-ray crystallography and NMR. It particularly useful in mapping protein-ligand interfaces and conformational changes resulted from ligand binding, mutation and aggregation.

Fast photochemical oxidation of proteins (FPOP), in its original conception, is a type of hydroxyl-radical-based protein footprinting that utilizes a pulsed KrF laser (248 nm) to trigger hydrolysis of hydrogen peroxide to produce solution hydroxyl radicals, which subsequently modify the protein in-situ. The platform is expanding to adopt other reactive species including carbenes. The reactivity of the probe depends on the intrinsic reactivity of the radical with the residue side chain and the solvent accessibility of the residue as a function of the tertiary/quaternary structures. By introducing an appropriate scavenger to compete with hydroxyl radical self-quenching, the lifetime of the primary radicals is remarkably shortened to ~ μs. Thus, the sampling timescale of FPOP is much faster than hydrogen-deuterium exchange and other covalent labeling methods relying on non-radical reactions.

The short footprinting timescale of FPOP offers two major advantages for protein structure elucidation: 1) it allows the protein to be interrogated in its native or near-native state with minimum structural perturbation; 2) it exhibits high sensitivity toward alterations in protein higher order structures because its sampling time is short with respect to protein conformational changes and dynamic motion. In addition, the covalent and irreversible oxidation by the hydroxyl radical provides more flexibility in the downstream proteomics workflow and MS analysis, permitting high spatial resolution with residue-specific information.

Since its invention in 2005 by Hambly and Gross, FPOP has developed from proof-of-concept to a valuable biophysical tool for interrogating protein structure. In this account, we summarize the principles and experimental design of FPOP that enable its fast labeling, and describe the current and unique capabilities of the technique in protein higher order structure elucidation. Application examples include characterization of amyloid beta self-assembly, protein-ligand interactions with a special emphasis on epitope mapping for protein therapeutics (e.g., antibody, Fab and adnectin), protein folding detailed to residue-specific folding kinetics, and protein flexibility/dynamics. Additionally, the utility of FPOP-based oxidative footprinting should grow with our continuing developments of novel reagents (e.g., sulfate radical anion, carbene diradical and trifluoromethyl radical). These reactive reagents are compatible with the current FPOP platform and offer different reactivity and selectivity towards various types of amino acid residues, providing complementary insights into protein higher order structures for soluble proteins and ultimately for membrane-bound proteins.

1.2 Introduction

Mass spectrometry (MS)-based protein footprinting has become a compelling tool for proteins that are incompatible with or simply too demanding for traditional structural techniques of X-ray crystallography and NMR or not appropriate for cryo-EM. As a highly practical technique that can be applied routinely, MS affords regional and residue-specific information in assessing structure and dynamics of proteins.

For perspective, we can view the commonly applied hydrogen-deuterium exchange (HDX) as a complementary and powerful means of protein footprinting. HDX interrogates protein higher order structures based on the stability, H-bonding, and structural protection of the protein backbone.[1] The labile N-D bond after exchange, however, is prone to back exchange when the protein or its constituent peptides are placed in protic solvents for analysis, limiting the capacity of HDX in structural proteomics applications that require sophisticated sample handling (e.g., purification, proteolysis, buffer exchange) prior to MS analysis.

In contrast, hydroxyl-radical (·OH) footprinting, first introduced by Chance[2-3] who used synchrotron radiation to ionize water and produce ·OH, yields irreversible modifications on the protein. Its elaboration as Fast Photochemical Oxidation of Proteins (FPOP) gives a method complementary to HDX. Hydroxyl radicals label the protein irreversibly, allowing application of rigorous downstream proteomics workflows to report accurately on solvent accessibility of the protein (Figure 1.1A). It accomplishes the labeling more rapidly and provides residue specificity, although with lower generality than HDX because labeling by ·OH requires the presence of reactive side chain(s) in a peptide segment.

FPOP was first developed by Hambly and Gross[4] to utilize ·OH generated from photolysis of hydrogen peroxide (H_2O_2) to monitor solvent accessibility of the protein side chain. Compared to

synchrotron radiolysis of water or use of the Fenton reaction, FPOP permits tuning the timescale of footprinting by introducing a scavenger to control the lifetime of the radical. This allows investigation of the protein structure in its native or near-native state without sampling altered conformations induced by modifications.

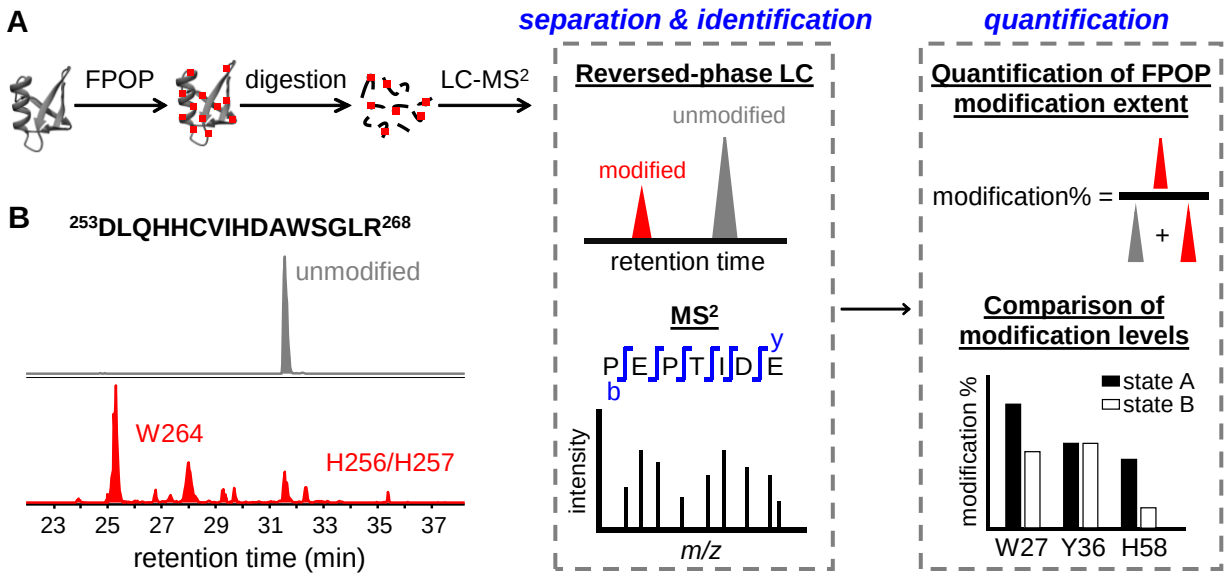


Figure 1.1. A standard workflow of protein footprinting by FPOP. (A) Two different states, A and B, of a targeted protein (gray ribbon) are labeled by FPOP. The hydroxyl radical abstracts a hydrogen from the protein, and a series of follow-up reactions occur subsequently to produce stable covalent modifications (red dots) on the protein. The peptides produced in digestion are shown as black lines. Signals of unmodified and modified peptides observed in the HPLC chromatograms are shown in gray and red, respectively. MS^2 is used to identify the modified residues. The extents of modification are quantified based on the signals of the unmodified and modified species and compared for proteins in different states. (B) Representative extracted-ion chromatograms showing the separation of the unmodified (grey) and FPOP-modified (red) peptides from the tryptic digests of interleukin-6 receptor (adapted with permission from ref. 8.

Copyright 2017 American Chemical Society.). The peptide sequence and numbering are indicated above the chromatogram. The modified residues corresponding to each signal peak are labeled on the chromatogram.

During the past decade, FPOP has been elaborated and applied to protein higher-order-structure characterization in both academic research and biotechnology development labs. Although most applications are occurring in biophysics, there is a significant need for new methods to support the discovery, development and quality control of protein therapeutics. In this review, we discuss the principles of FPOP and highlight selected applications including characterization of amyloid formation;[5] of protein fast/slow folding;[6-7] of protein-ligand interactions (emphasis on epitope mapping);[8-10] and of protein dynamics and the identification of hidden conformations.[11-12] These examples illustrate the power of FPOP to decipher the higher order structures of protein and protein complexes.

1.3 FPOP Fundamentals

1.3.1 FPOP setup

The FPOP apparatus uses a 248 nm KrF excimer laser to cleave H_2O_2 at low concentration (15 mM, 0.04%) (Figure 1.2). We chose 248 nm because the absorptions of water and most proteins are low at this wavelength. The laser beam is focused with convex lenses to give an exposure window of 2.0-3.0 mm wide on a 150 μM i.d. silica tubing (termed the FPOP capillary). The laser affords a high and converging flux of photons, maximizing the yield of ·OH in a small plug of an irradiated solution.

FPOP uses a flow system for protein footprinting, in which the capillary is placed perpendicular to the laser beam (Figure 1.2). Prior to FPOP, the protein is mixed with H_2O_2 and a radical scavenger (usually a free amino acid). The solution mixture is then immediately transported through the FPOP capillary by a syringe pump. When the solution passes the transparent window on the capillary, laser-triggered photolysis of H_2O_2 produces ·OH in nanoseconds. The radicals further undergo Haber-Weiss chain reactions in competition with self-quenching to re-form H_2O_2 (Scheme 1.1).

The hydroxyl radical induces protein oxidation by abstracting a hydrogen from the protein, which results in a protein radical that further undergoes a series of subsequent reactions to generate products with stable oxidation.[2] Rate constants for reaction with individual amino acids range from 10^7 to 10^{10} M^{-1} s^{-1}, serving as approximates of the intrinsic reactivities of protein side chains.[2]

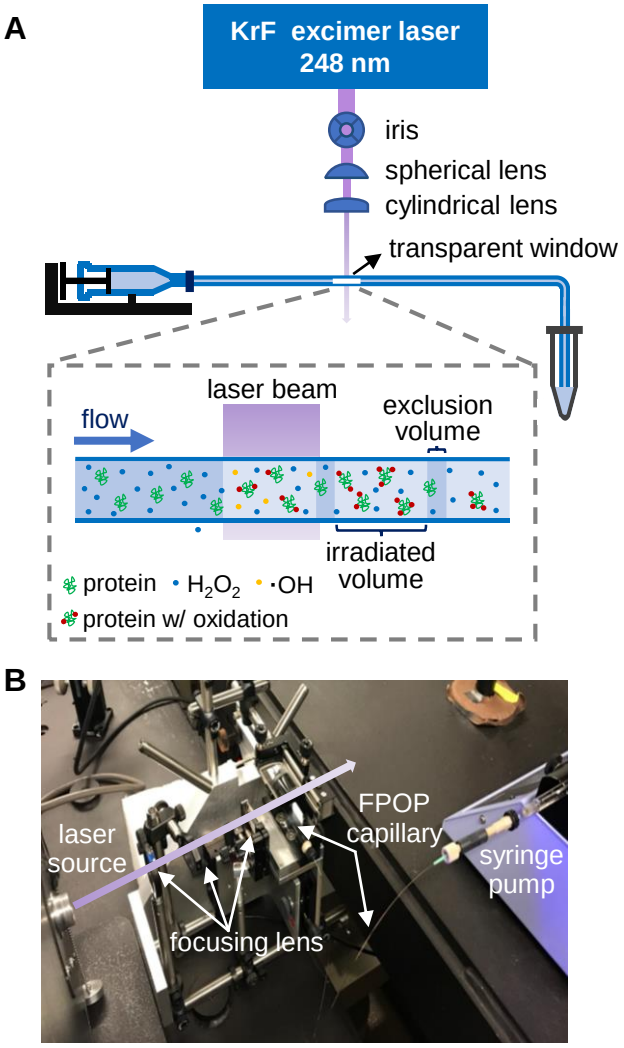


Figure 1.2. FPOP setup. (A) Schematic of the FPOP platform. The laser beam (violet) generated from an excimer laser (blue square) is focused through an iris and then by two convex lenses. The FPOP capillary (blue line) made of silica tubing is connected to a syringe pump. The transparent window without polyimide coating indicates the location of laser irradiation (a blow-up of the transparent window is shown in the dashed box). A tube containing catalase and Met is placed at the end of the FPOP capillary to collect the FPOP-modified sample. (B) Photo of the FPOP apparatus. Primary components are labeled, and the violet arrow represents a visual pathway of the laser beam perpendicular to the FPOP capillary.

$$H_2O_2 \xrightarrow{h\nu} 2HO\bullet \qquad \text{eq. 1}$$

$$HO\bullet + H_2O_2 \longrightarrow H_2O + HO_2\bullet \qquad \text{eq. 2}$$
$$k_1 = 2.7 \times 10^7\ M^{-1}\cdot s^{-1}$$

$$HO_2\bullet + H_2O_2 \longrightarrow H_2O + O_2 + HO\bullet \qquad \text{eq. 3}$$
$$k_2 = 7.0 \times 10^9\ M^{-1}\cdot s^{-1}$$

$$2HO\bullet \longrightarrow H_2O_2 \qquad \text{eq. 4}$$
$$k_3 = 4.7 \times 10^9\ M^{-1}\cdot s^{-1}$$

Scheme 1.1. Chain reactions involved in the generation of ·OH by FPOP. The rate constant is shown below each reaction.[13]

An external pulse generator is incorporated to control frequency of the laser pulse, which together with a proper flow rate of the solution minimizes multiple laser irradiation to the same solution plug. To further avoid "double shooting", a small volume is intentionally excluded from laser exposure to create a "barrier" between each exposed plug (termed an "exclusion volume"). The outlet of the FPOP capillary is placed into a sample-collection tube containing catalase and free methionine in buffer to remove leftover H_2O_2 and to prevent post-footprinting oxidation artifacts from any remaining reactive species. The control sample, used for correcting the background oxidation (< 5%, mainly on Met) of the protein induced by H_2O_2 or in post-FPOP sample handling is prepared the exact same way as the experimental sample, except the laser is omitted during the FPOP workflow.

If there is a concern that significant oxidation will occur when mixing the protein with H_2O_2, then H_2O_2 can be supplied prior to laser irradiation by using a T-shaped micromixer. In this setup, one syringe pump delivers the protein in buffer, and another delivers H_2O_2 and the scavenger for T-mixing prior to laser irradiation. This setup improves reproducibility by minimizing the exposure of the protein to H_2O_2 and hence any non-FPOP oxidation by the peroxide.[14]

1.3.2 Timescale of the FPOP probe

The radical lifetime in FPOP is tuned through varying the type and/or concentration of the scavenger.[4, 15] In the absence of a scavenger, the radical lifetime in FPOP is limited by the recombination reaction and can extend to approximately 100 μs. Allowing a protein to react for this time may lead to unfolding, exposing buried residues to solvent exposure. Addition of a scavenger (e.g., 20 mM Gln) to the protein solution effectively shortens the lifetime of the primary hydroxyl radical to ~ 1 μs, limiting the modification to reactive and solvent-exposed residues in the native conformation. Gau et al.[16] studied the population distribution of the FPOP oxidation (0, +16, +32 Da etc.) for three oxidation-sensitive proteins (β-lactoglobulin, apo calmodulin and lysozyme), and proposed that a fit of the distribution to a Poisson signals conformational homogeneity of the protein during FPOP labeling. That is, protein samples with FPOP properly conducted by including a radical scavenger and suitable quenching of peroxide (catalase and Met) undergo footprinting of a single conformational state without perturbing the native conformation during labeling. These results underscore the indispensable roles of the flow system, the radical scavenger, and the prompt removal of peroxide in FPOP.

Any radical reaction leads to production of a secondary or higher order radicals, which have lifetimes in the millisec range.[17] If these longer-lived radicals modify the protein, the time of labeling is longer than ~1 μs that we claimed.[4] Unfortunately, there is little direct evidence that reactions of secondary radicals like ·OOH modify the protein. Some indirect demonstration of the rapid labeling, however, comes from the application of FPOP in probing protein fast folding (see "Protein folding" section)[7]. The clear time-dependence seen in that study suggests that the radical reactions are not continuing onto the millisecond time range.

Protein modification depends not only on lifetime but also the number of radicals produced by the laser. Using d_0/d_5 Phe as an isotope-encoded dosimeter, for 15 mM H_2O_2, we found that the initial concentration of ·OH is 0.95 mM.[15] Because the amount of the radical scavenger is in a large excess with respect to the solvent-exposed, reactive sites in the protein (e.g., 20 mM free Gln vs. 10 μM protein), the dominating chemistry that controls the lifetime of ·OH in FPOP is the reaction with the scavenger. Therefore, moderate variations in the protein concentration or the presence of a ligand will not significantly affect the overall oxidative yield on the protein or the relative yields on various side chains.[18] From our experience, this remains valid for compounds with a size up to that of an antibody (~ 150 kDa).[19] Thus, observed FPOP variations between the protein in different conformational states (e.g., apo and holo) correlate only with their different solvent accessibilities.

1.3.3 Post-FPOP proteomics workflow

The FPOP-modified protein is usually analyzed by a bottom-up proteomics strategy to obtain sub-regional and residue-specific structural information (i.e., the footprint) (Figure 1.1). Owing to the covalent and irreversible modification by FPOP, many improvements made for analysis and determination of the primary structure (analytical proteomics) can be applied to facilitate the FPOP approach for higher order structure characterization. For example, the irreversible modification (unlike in HDX) offers high flexibility in proteolysis. Digestion with specific (e.g., trypsin, Asp-N), semi-specific (e.g., chymotrypsin) and nonspecific (e.g., pepsin) proteases can be performed in tandem or separately to generate overlapping peptide fragments that give comprehensive coverage of the protein sequence. Unlike identifying primary structure in proteomics, footprinting requires high coverage of the protein sequence, but this is manageable because usually a single, known protein or a simple mixture of proteins is used.

The resultant peptides are analyzed by liquid chromatography and tandem mass spectrometry (LC-MS^2) to identify the modification sites and to quantify the levels. Oxidation by FPOP makes the peptide less hydrophobic, permitting separation of the modified peptide from its unmodified form by reversed-phase LC (see Figure 1.1B for an example). Generally, the modified peptides elute earlier than the unmodified ones, which allows a clear identification of the modified peptide on the basis of its accurate mass and product-ion spectrum (MS^2). We quantify the peptide modification level based on the primary modification species (i.e., products of OH substitution (+15.9949 Da)). To calculate the modification percentage for a specific peptide, signal intensities of the unmodified peptide (I_u) and its modified species (ΣI_{ox}) are taken as peak areas from extracted ion chromatograms (XICs). The modification fraction for a certain peptide is calculated from the following equation: Fraction-modified = $\Sigma I_{ox}/(I_u + \Sigma I_{ox})$. This algorithm is straightforward and sufficiently sensitive to report the solvent accessibility change.[5]

1.4 Applications of FPOP

To show that the method is reliable and versatile, we now review four topics that show the utility of FPOP in biophysics and structural proteomics.

1.4.1 Amyloid protein aggregation

Aggregation of amyloid beta (Aβ) is one of the primary pathogenic events leading to Alzheimer's disease. Characterizing the products of Aβ aggregation, however, remains challenging owing to the vast heterogeneity and transient nature of the intermediate aggregates. We recently used FPOP to monitor amyloid formation and solvent accessibility of $A\beta_{1-42}$, the 42-amino acid form of Aβ, at global, peptide, and amino-acid levels (Figure 1.3).[5] FPOP footprinting immediately captures the solvent accessibilities of $A\beta_{1-42}$ intermediates, filling in the

gap between solution NMR for the early state oligomers and solids NMR and cryo-EM for the final fibrils. The readout from FPOP is clear, because the irreversible modification by FPOP should not be affected by Aβ aggregation and/or conformational change occurring post-footprinting.

We modified $A\beta_{1-42}$ aggregates in various oligomeric states by FPOP. Occurring are extensive modifications for $A\beta_{1-42}$ monomer owing to its intrinsically disordered structure (Figure 1.3A). As $A\beta_{1-42}$ assembles and reorganizes into higher order oligomers, the FPOP modification of $A\beta_{1-42}$ deceases owing to formation of secondary/tertiary structures (Figure 1.3B and C). At the end of aggregation, the modification extent decreases to a modest level (Figure 1.3D), indicating the formation of compact amyloid fibrils with a core structure that resists modification. A fit of a plot of FPOP modifications as a function of incubation time, adopting the conventional nucleation-condensation mechanism (Finke-Watzky), affords a sigmoidal curve that shows sequential formation of the major species of $A\beta_{1-42}$ during aggregation (Figure 1.3E). We also characterized the interconversion of those representative Aβ species by simulating their concentration change as a function of the incubation time (Figure 1.3F).

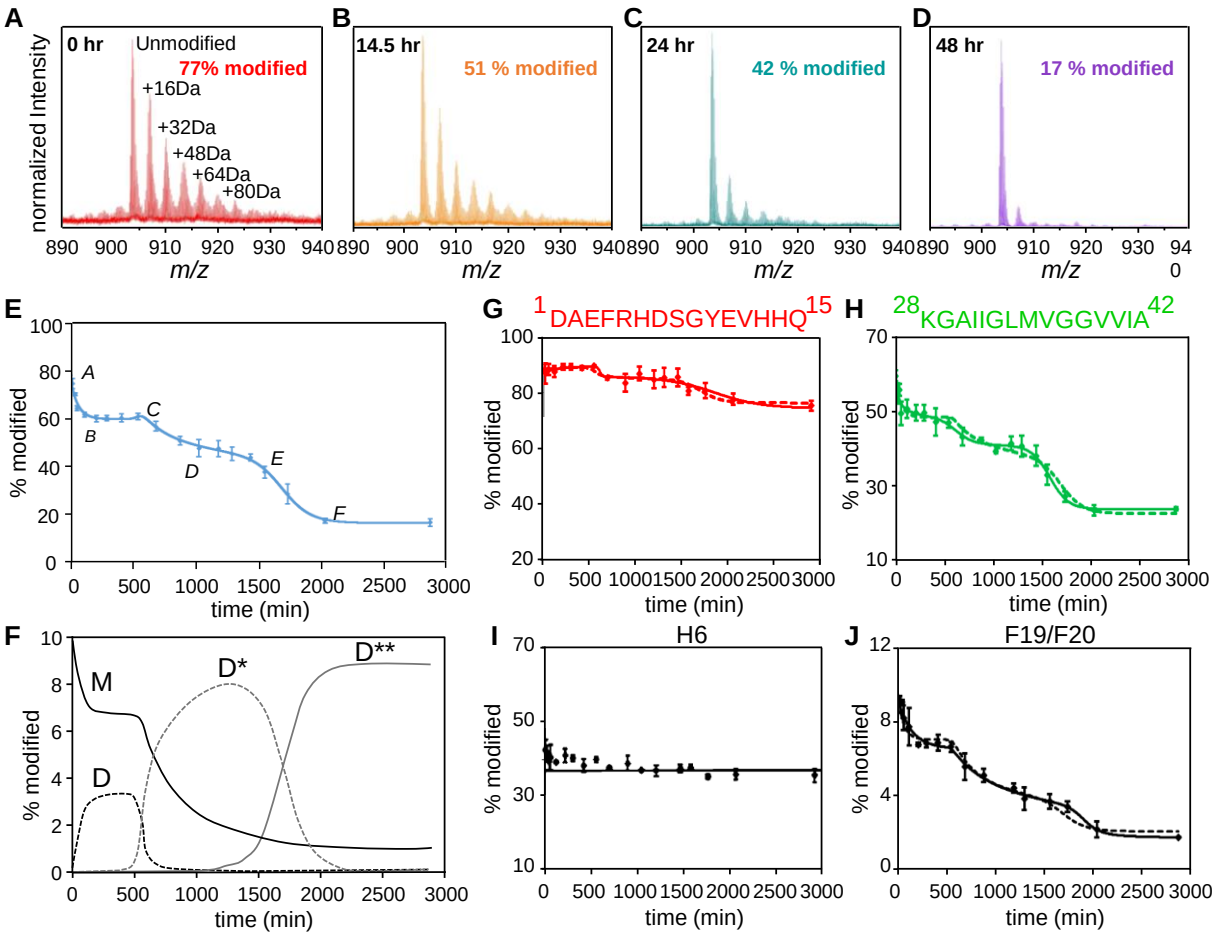


Figure 1.3. FPOP and kinetic modeling characterize the time-dependent aggregation of Aβ-1-42 at protein, peptide and amino-acid residue levels. (A-D) Mass spectra showing extents of FPOP modification for intact $A\beta_{1-42}$ (5+ charge) as a function of the incubation time. The determined modification percentage is shown in each panel. (E) Characterization of $A\beta_{1-42}$ aggregation on the global (full-polypeptide) level by a kinetic simulation. Points represent experimental data, and the solid curve is a model fit based on two auto-catalytic reactions. (F) Concentrations (in monomeric equivalents) change of representative $A\beta_{1-42}$ species (M-monomer, D-paranuclei, D*-protofibrils and D**-fibrils) from kinetic simulation. (G-H) Aggregation curves of N-terminal region 1-15 and C-terminal region 28-42. (I-J) Aggregation curves of representative $A\beta_{1-42}$ residues H6 and F19/F20. In G-J, the solid and dashed curves are model fits

independent of or constrained by the global rates, respectively. Adapted with permission from ref. 5. Copyright 2016 American Chemical Society.

FPOP also provides a high-resolution view of aggregation at regional and amino-acid levels for some residues. We characterized the modification of three $A\beta_{1-42}$ sub regions, including the N-terminal peptide 1-15, central domain 16-27, and C-terminal peptide 28-42 from digestion of the full-length $A\beta_{1-42}$. FPOP modification for the N-terminal peptide remains high (~ 80%) with little variation as $A\beta_{1-42}$ aggregates, indicating that the N-terminus remains structurally flexible and solvent accessible, with little participation in $A\beta_{1-42}$ aggregation (Figure 1.3G). This finding is further supported by NMR results showing that the N-terminus of $A\beta_{1-42}$ is disordered in a variety of oligomers and in the mature fibril. By contrast, the critical roles of the central domain and hydrophobic C-terminus in promoting self-association is highlighted by a dramatic decrease in FPOP modification from monomers to fibrils (Figure 1.3H). Furthermore, measuring the modification at the amino-acid level allows characterization of the aggregation tendencies for $A\beta_{1-42}$ at a residue level. For example, H6 shows relatively constant FPOP modification, during aggregation, indicating little change in its solvent-accessibility and conformation (Figure 1.3I). On the other hand, F19/F20 are residues primarily responsible for the modification change observed for the central domain (Figure 1.3J), and we propose that interactions of F19/F20 contribute a driving force for $A\beta_{1-42}$ aggregation by serving as the hydrophobic nucleation interface.

We envision that this FPOP platform can reveal how interactions of Aβ with other molecules affect site-specific aggregation. Application to drug candidates and natural products that promote/inhibit oligomerization should be particularly fruitful.

1.4.2 Protein folding

Protein folding and stability is crucial in understanding the biological effects of mutations and the pathogenesis of diseases associated with aggregation of misfolded proteins. Protein folding is transient and rapid (usually sub-millisecond), challenging structural characterization of kinetic intermediates formed during folding. Previously, spectroscopic methods, including fluorescence emission, circular dichroism, and electron-spin resonance, have provided insights into the underlying mechanism of protein folding. Approaches having high spatial resolution and sensitivity, however, are still sparse. Because the labeling speed of FPOP is faster than most protein folding, it should be effective in probing folding by measuring protein solvent-accessibility changes accompanying the conformational change. Meanwhile, the high spatial resolution of FPOP allows direct assessment of the folding down to amino-acid residues in some cases.

To explore this opportunity, we designed a two laser-based platform that combines a temperature jump (one laser) and FPOP (second laser) to study protein fast folding (Figure 1.4A).[7] This platform employed a Raman-shifted Nd:YAG laser (1900 nm) to generate a nanosecond temperature jump (~ 20 °C) in the buffer solution and initiate protein (barstar) refolding, followed by a folding time-dependent FPOP footprinting triggered by the KrF excimer laser. Control of the heating and the FPOP laser, and hence the folding times is achieved by two delay circuits, which are used to adjust the time between firing the two lasers.

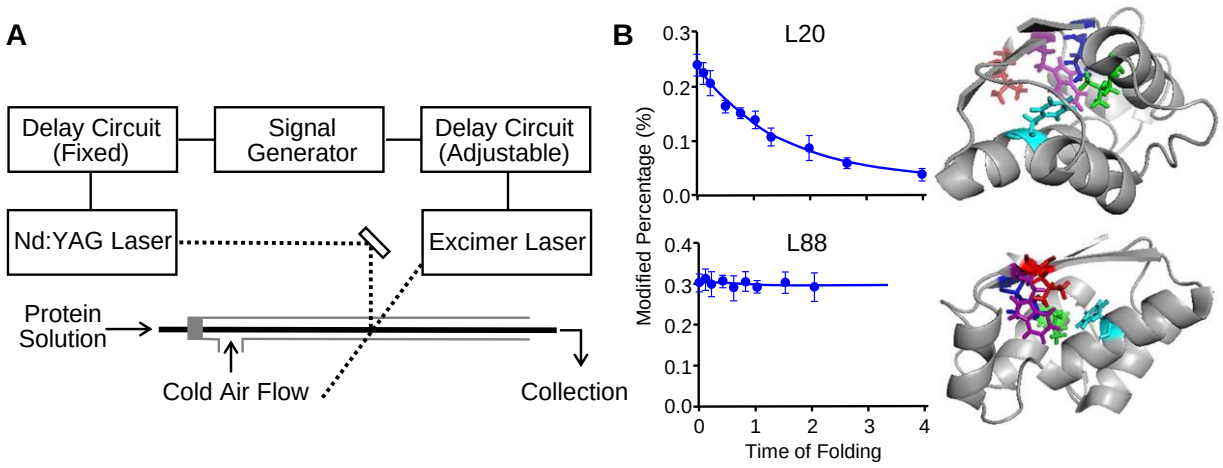


Figure 1.4. Folding of barstar characterized by a T-jump FPOP. (A) Schematic representation of the two-laser FPOP platform. As in Figure 1.2A, the transparent window is located where the two laser beams (dash lines) are incorporated. (B) Left: FPOP modification percentage of two representative residues as a function of the protein (barstar) folding time. Solid lines in the plots are obtained from kinetic fitting. Right: Five critical residues identified by FPOP mapped on native barstar structure. Two views are provided to show the side chains of the amino acid L88 (red), F74 (cyan), I5 (blue), L20 (green), and W53 (purple). Adapted with permission from ref. 7. Copyright 2012 American Chemical Society.

The test protein, barstar, a small globulin protein with a single domain structure composed of four α-helixes and three β-sheets, is a well-characterized model for protein folding. We probed its folding from a fraction of a ms to 2 ms, a time domain spanning its early folding and found residues with distinct kinetic features in terms of their solvent accessibility (Figure 1.4B). Modifications of H17, L20 and L24, all located in helix$_1$, show an exponential decrease in the first 2 ms of folding (a control is represented by L88 in Figure 1.4B). Their modification at 2 ms time is attenuated owing to its folding. These residues showing dramatically decreased FPOP

modification are consistent with folding via hydrophobic interactions in the first 2 ms. Our findings agree with the spectroscopic results[20] but bring spatial resolution, demonstrating that the hydrophobic collapse on $helix_1$ is the initial step in barstar folding.

FPOP can also be used to probe protein slow folding. An example is study of the refolding of a viral fusion protein, parainfluenza virus 5 F (PIV5 F).[6] We triggered the irreversible refolding of PIV5 F by heat, and captured five conformations on the basis of their FPOP footprints: PIV5 F in the pre-fusion state (before heating), three intermediate states (45, 55, and 65 °C), and the post-fusion state. The experiment required the FPOP capillary contiguous to the laser window to be placed in a heated chamber that maintains the temperature of the protein solution. The modification of the PIV5 F decreases upon transition from the pre-fusion to the post-fusion state, indicating that the structure of PIV5 F in the pre-fusion state is more open and solvent-exposed.

The modifications of the PIV5 F peptides afford structural information on regions becoming solvent-exposed or protected along the refolding pathway. This allows us to propose a model describing the dynamic refolding of PIV5 F and to compare the differences in FPOP modification of the pre-fusion and post-fusion PIV5 F with changes in the solvent accessible surface area (SASA) from X-ray structures. There are greater changes in solvent accessibility for nine regions, as reported by FPOP but not predicted by SASA. This indicates an underestimation of the pre-fusion PIV5 F's SASA calculated from the static crystal structure and points to a higher structural flexibility in solution. Interestingly, most of those regions contain the epitopes of PIV5 F for the prefusion-specific neutralizing antibodies, suggesting the importance of those regions for PIV5 F biological function.

1.4.3 Epitope mapping

Understanding antibody action relies on identifying the binding sites of the antibody on the target antigen. Methods for this, called epitope mapping, include X-ray crystallography, NMR, site-directed mutagenesis, and MS-based approaches. Those techniques offer complementary information on the epitope and on binding-induced conformational changes. To illustrate the capability of FPOP for epitope mapping, our lab has collaborated with Bristol-Myers Squibb and Genentech to characterize epitopes for a number of targets upon binding of protein therapeutics.

Early in the development, we applied FPOP to investigate the epitope of human epidermal growth factor receptor (EGFR) for adnectin binding at both the peptide and amino-acid residue levels.[10] The FPOP-determined binding interface involves various amino-acid and peptide regions near the N-terminus of EGFR. Our data correlate well with the previously determined epitopes from the crystal structure. Additionally, successful characterization of epitopes in thrombin[19] and vascular endothelial growth factor [18] serves as other good examples of FPOP's ability to identify both the epitope and regions with remote conformational changes.

We later combined HDX-MS, site-directed mutagenesis, and FPOP to study the epitope of interleukin-23 (IL-23) upon binding of an anti-IL-23 antibody (Figure 1.5).[9] Using HDX-MS, we found five peptide regions on IL-23 showing reduced backbone amide solvent accessibility upon antibody binding. Five different peptides of IL-23 are identified by FPOP, among which three regions are identified by HDX as well. Additional FPOP analysis at residue level allows us to assign potential critical-binding residues.

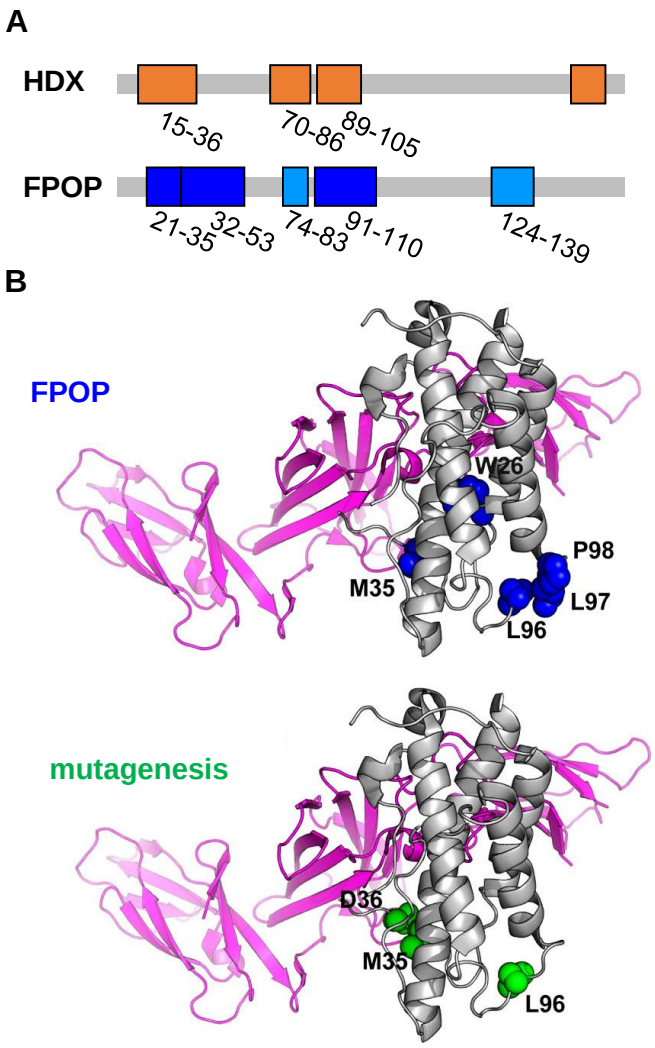


Figure 1.5. Epitope regions of IL-23 identified by FPOP, HDX-MS, and alanine shave mutagenesis. (A) Critical peptides discovered by HDX (top) and FPOP (bottom). These regions are mapped on the linear sequence of IL-23 (in orange for HDX results and blue for FPOP). (B) Critical residues identified by FPOP (top) and alanine-shave mutagenesis (bottom) mapped on the IL-23 crystal structure. Amino acids identified by FPOP are in blue, and by mutagenesis in green. Gray regions suggest no significant difference in the conformation of IL-23 upon binding with Fab. Adapted with permission from ref. 9. Copyright 2017 American Chemical Society.

More recently, we implemented FPOP, HDX, and carboxyl group footprinting to map interactively the epitope of IL-6 receptor (IL-6R) for two adnectins with distinct affinities (K_d, Adnectin1 ~ 6.2 pM vs K_d, Adnectin2 ~ 46 nM).[8] Besides identifying the epitope as a flexible loop that connects two β-strands in the cytokine-binding domain, our results reveal that two loops, located beyond the conserved epitope undergo reduced dynamic motion upon adnectin1 binding. Those local effects on the IL-6R structure are attenuated or not observed in the case of adnectin2 binding. Apparently, binding of adnectin1, the stronger ligand, stabilizes IL-6R by reducing the flexibility of those dynamic regions.

Outcomes from the above studies highlight the capability of FPOP as a sensitive tool for epitope mapping, motivating its further applications in protein discovery and development either as a stand-alone method or in conjunction with orthogonal structural approaches (e.g., X-ray, SAXS, EM) and other MS-based methods (e.g., HDX-MS, ion mobility). In addition, the combined use of MS-based methods provides critical information on therapeutic targets in the absence of an X-ray structure, has high throughput, and may be a major advance for the discovery and development of biopharmaceuticals.

1.4.4 Protein dynamics and hidden conformations

FPOP can report variations in protein dynamics and flexibility because its sampling time is short with respect to local dynamics of the protein. For example, a study of the oligomerization interface of apolipoprotein E3 (ApoE3) reveals that the C-terminal helix of ApoE3 and an unstructured hinge region proceeding the helix undergo decreases in solvent accessibilities in the tetramer compared to a monomeric mutant.[12] The C-terminal helix of ApoE3 has long been hypothesized to be the primary interface for self-association leading to tetramer formation. Evidence for the change in the mobility of the hinge region associated with oligomerization

cannot be seen by HDX and carboxyl group footprinting on ApoE3, done subsequently. Thus, the preceding hinge region is stabilized owing to formation of the oligomerization interface by the C-terminus of ApoE3. The inability to see this by HDX and carboxyl-group footprinting stems from their relatively long labeling time compared to that of FPOP.

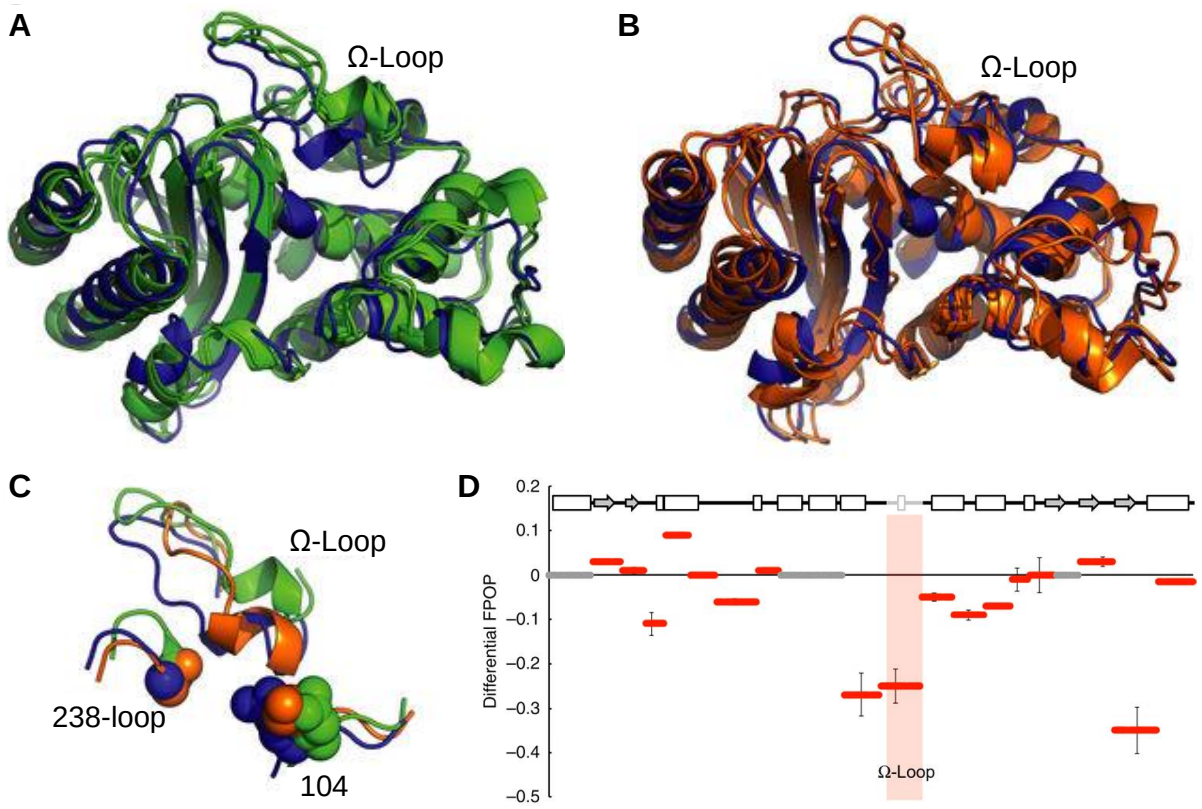


Figure 1.6. Modeling and FPOP reveal restricted motion in the Ω-loop in cefotaximase variant. A crystal structure of TEM-1 (blue) is overlaid with the two most populated structures taken from modeling of (A) the non-cefotaximase states (green, favored by wild type) and (B) cefotaximase states (orange, favored by the E104K/G238S). (C) Large structural rearrangements in the Ω-loop distinguish low-energy non-cefotaximase states from cefotaximase states. (D) FPOP data reveal experimentally the restricted motion in cefotaximase variant predicted from modeling. Hydroxyl labelling of TEM variants are shown as the difference in percent labelled

between cefotaximase variant and wild type. The Ω-loop (164–179) is shaded grey in the sequence and pink in the graph. Adapted with permission from ref. 11. Copyright 2016 Nature Publishing Group.

FPOP also effectively reveals the difference in local dynamics between wildtype (WT) TEM-1 and its cefotaximase variant. TEM β-lactamase is an enzyme responsible for the antibiotic resistance in pathogenic Gram-negative bacteria. The cefotaximase variant of TEM, favored by a E104K/G238S mutant, hydrolyzes cefotaxime 1,400-fold more efficiently than does WT TEM-1.[11] Although the function and specificity of various TEM variants can vary dramatically, no apparent conformational changes are found for those variants (Figure 1.6A and B). To uncover the structural difference between the cefotaximase variant and WT TEM-1, our collaborators, using Markov state models-based simulation, proposed that the decreased flexibility of the Ω-loop in the cefotaximase variant is responsible for its high enzymatic efficiency by stabilizing the binding-competent state (Figure 1.6C). To test, we turned to FPOP and found significantly decreased FPOP modification on the Ω-loop itself and the region preceding the Ω-loop in the cefotaximase variant (Figure 1.6D). In addition, the decreased solvent accessibility pertains to the C-terminus, suggesting that the E104K/G238S substitution not only reduces the flexibility in the Ω-loop, but that of the C-terminus remotely via long-range interaction. In this case, FPOP provides a test for the theory from molecular dynamics simulation and demonstrates its capability in probing protein solution dynamics that are inaccessible by static crystal structures.

1.5 Conclusions and Outlook

FPOP coupled with MS has emerged as a powerful approach to study protein higher order structures. The robustness of the platform has been further enhanced by recent advances in methodology, allowing the approach to provide improved quantitative information and broader coverage in footprinting for various types of amino acid residue.

One example is incorporation of a reporter peptide into the protein sample to resolve discrepancies in the radical dosage. In this scheme, the protein and the reporter peptide are modified by FPOP simultaneously under the same condition. Thus, the modification extent of the reporter peptide correlates with the radical lifetime in the sample. By varying the scavenger concentration (i.e., the radical lifetime), the reporter peptide approach permits a normalized, time-dependent measurement of the modification on the protein.[21] In addition, the FPOP platform has been extended by Jones to footprint live cells,[22] allowing investigation of protein conformation and interactions in the native cellular environment.

The versatility of FPOP to accommodate new reagents in addition to the ·OH is illustrated by the developments of the sulfate radial anion (SO^{4-}·)[23] and iodine radical (I·)[24] as alternative footprinting reagents. We also adapted the carbene radical, a laser-based footprinting reagent developed by Schriemer[25] using a Nd-YAG laser at 355 nm, into the FPOP platform to modify Asp/Glu.[26] More recently, Cheng et al. developed trifluoromethyl radical (·CF_3) as a novel reagent that can modify 18 of 20 amino acids residues, including Gly, Ala, Ser, Thr, Asp and Glu, which have low reactivities with ·OH.[27]

The ability to accommodate new reagents on the FPOP platform allows construction of a “library” of reagents that offer different chemical reactivities and specificities to footprint proteins and distinguish structural changes that occur from apo to holo, bound to unbound, and wild type to

mutant. As these improvements grow, FPOP-based oxidative footprinting approaches will play an increasingly important role by serving as constrains of the macromolecule structure, much like chemical shifts do in NMR. The ultimate goal is to use chemistry and MS-based proteomics to determine coarse-grained structure of proteins and their assemblies. This will require high coverage of the protein residues in footprinting and motivates continued development and application of FPOP.

1.6 References

1. Engen, J. R., Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal. Chem.* **2009,** *81* (19), 7870-7875.

2. Xu, G.; Chance, M. R., Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **2007,** *107* (8), 3514-3543.

3. Maleknia, S. D.; Brenowitz, M.; Chance, M. R., Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry. *Anal. Chem.* **1999,** *71* (18), 3965-73.

4. Hambly, D. M.; Gross, M. L., Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J. Am. Soc. Mass Spectrom.* **2005,** *16* (12), 2057-63.

5. Li, K. S.; Rempel, D. L.; Gross, M. L., Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1–42 Aggregation. *J. Am. Chem. Soc.* **2016,** *138* (37), 12090-12098.

6. Poor, T. A.; Jones, L. M.; Sood, A.; Leser, G. P.; Plasencia, M. D.; Rempel, D. L.; Jardetzky, T. S.; Woods, R. J.; Gross, M. L.; Lamb, R. A., Probing the paramyxovirus fusion (F) protein-refolding event from pre- to postfusion by oxidative footprinting. *Proceedings of the National Academy of Sciences* **2014,** *111* (25), E2596-E2605.

7. Chen, J.; Rempel, D. L.; Gau, B. C.; Gross, M. L., Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level. *J. Am. Chem. Soc.* **2012,** *134* (45), 18724-31.

8. Li, K. S.; Chen, G.; Mo, J.; Huang, R. Y. C.; Deyanova, E. G.; Beno, B. R.; O'Neil, S. R.; Tymiak, A. A.; Gross, M. L., Orthogonal Mass Spectrometry-Based Footprinting for Epitope Mapping and Structural Characterization: The IL-6 Receptor upon Binding of Protein Therapeutics. *Anal. Chem.* **2017,** *89* (14), 7742-7749.

9. Li, J.; Wei, H.; Krystek, S. R.; Bond, D.; Brender, T. M.; Cohen, D.; Feiner, J.; Hamacher, N.; Harshman, J.; Huang, R. Y. C.; Julien, S. H.; Lin, Z.; Moore, K.; Mueller, L.; Noriega, C.; Sejwal, P.; Sheppard, P.; Stevens, B.; Chen, G.; Tymiak, A. A.; Gross, M. L.; Schneeweis, L. A., Mapping the Energetic Epitope of an Antibody/Interleukin-23 Interaction with Hydrogen/Deuterium Exchange, Fast Photochemical Oxidation of Proteins Mass Spectrometry, and Alanine Shave Mutagenesis. *Anal. Chem.* **2017,** *89* (4), 2250-2258.

10. Yan, Y.; Chen, G.; Wei, H.; Huang, R. Y.-C.; Mo, J.; Rempel, D. L.; Tymiak, A. A.; Gross, M. L., Fast Photochemical Oxidation of Proteins (FPOP) Maps the Epitope of EGFR Binding to Adnectin. *J. Am. Soc. Mass Spectrom.* **2014,** *25* (12), 2084-2092.

11. Hart, K. M.; Ho, C. M.; Dutta, S.; Gross, M. L.; Bowman, G. R., Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat Commun* **2016,** *7*, 12965.

12. Gau, B.; Garai, K.; Frieden, C.; Gross, M. L., Mass Spectrometry-Based Protein Footprinting Characterizes the Structures of Oligomeric Apolipoprotein E2, E3, and E4. *Biochemistry* **2011,** *50* (38), 8117-8126.

13. Elliot, A. J.; Chalk River Nuclear, L.; System, C.; Corrosion, *Rate constants and G-values for the simulation of the radiolysis of light water over the range 0-300˚*. System Chemistry & Corrosion Branch, Chalk River Laboratories: Chalk River, Ont., 1994.

14. Zhang, Y.; Rempel, D. L.; Zhang, H.; Gross, M. L., An Improved Fast Photochemical Oxidation of Proteins (FPOP) Platform for Protein Therapeutics. *J. Am. Soc. Mass Spectrom.* **2015,** *26* (3), 526-529.

15. Niu, B.; Zhang, H.; Giblin, D.; Rempel, D. L.; Gross, M. L., Dosimetry Determines the Initial OH Radical Concentration in Fast Photochemical Oxidation of Proteins (FPOP). *J. Am. Soc. Mass Spectrom.* **2015,** *26* (5), 843-846.

16. Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L., Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* **2009,** *81* (16), 6563-6571.

17. Vahidi, S.; Konermann, L., Probing the Time Scale of FPOP (Fast Photochemical Oxidation of Proteins): Radical Reactions Extend Over Tens of Milliseconds. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (7), 1156-1164.

18. Zhang, Y.; Wecksler, A. T.; Molina, P.; Deperalta, G.; Gross, M. L., Mapping the Binding Interface of VEGF and a Monoclonal Antibody Fab-1 Fragment with Fast Photochemical Oxidation of Proteins (FPOP) and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017,** *28* (5), 850-858.

19. Jones, L. M.; B. Sperry, J.; A. Carroll, J.; Gross, M. L., Fast Photochemical Oxidation of Proteins for Epitope Mapping. *Anal. Chem.* **2011,** *83* (20), 7657-7661.

20. Agashe, V. R.; Shastry, M. C.; Udgaonkar, J. B., Initial hydrophobic collapse in the folding of barstar. *Nature* **1995,** *377* (6551), 754-7.

21. Niu, B.; Mackness, B. C.; Rempel, D. L.; Zhang, H.; Cui, W.; Matthews, C. R.; Zitzewitz, J. A.; Gross, M. L., Incorporation of a Reporter Peptide in FPOP Compensates for Adventitious Scavengers and Permits Time-Dependent Measurements. *J. Am. Soc. Mass Spectrom.* **2017,** *28* (2), 389-392.

22. Espino, J. A.; Mali, V. S.; Jones, L. M., In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells. *Anal. Chem.* **2015,** *87* (15), 7971-7978.

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

24. Chen, J.; Cui, W.; Giblin, D.; Gross, M. L., New Protein Footprinting: Fast Photochemical Iodination Combined with Top-down and Bottom-up Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2012,** *23* (8), 1306-1318.

25. Jumper, C. C.; Schriemer, D. C., Mass Spectrometry of Laser-Initiated Carbene Reactions for Protein Topographic Analysis. *Anal. Chem.* **2011,** *83* (8), 2913-2920.

26. Zhang, B.; Rempel, D. L.; Gross, M. L., Protein Footprinting by Carbenes on a Fast Photochemical Oxidation of Proteins (FPOP) Platform. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (3), 552-555.

27. Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L., Laser-Initiated Radical Trifluoromethylation of Peptides and Proteins: Application to Mass-Spectrometry-Based Protein Footprinting. *Angew. Chem. Int. Ed.* **2017,** *56* (45), 14007-14010.

Chapter 2: Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1-42 Aggregation*

* This chapter is based on the following publication: Li, K. S.; Rempel, D. L.; Gross, M. L., Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1–42 Aggregation. *J. Am. Chem. Soc.* 2016, 138 (37), 12090-12098.

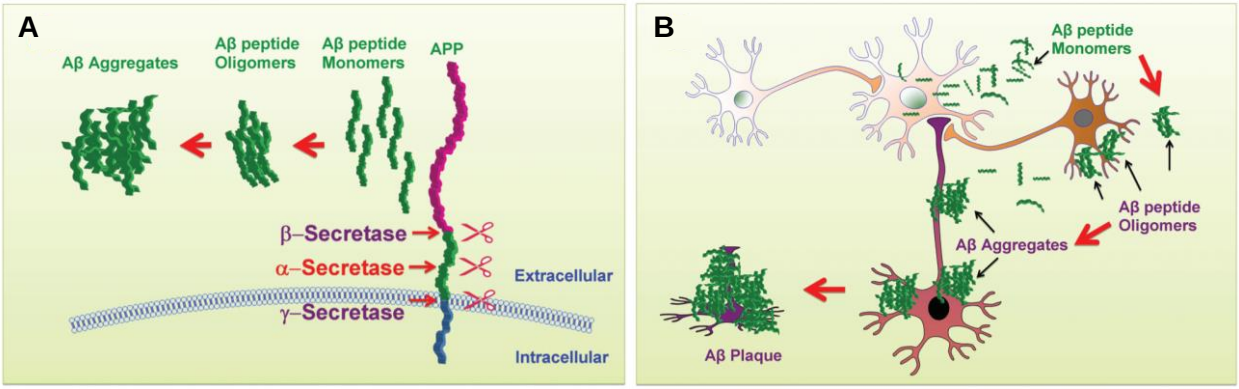
2.1 Abstract

Preventing and treating Alzheimer's disease require understanding the aggregation of amyloid beta 1-42 ($A\beta_{1-42}$) to give oligomers, protofibrils, and fibrils. Here we describe the footprinting of $A\beta_{1-42}$ by hydroxyl radical-based fast photochemical oxidation of proteins (FPOP) and mass spectrometry (MS) to monitor time-course $A\beta_{1-42}$ aggregation. We resolved five distinct stages characterized by two sigmoidal behaviors, showing the time-dependent transitions of monomers–paranuclei–protofibrils-fibrillar aggregates. Kinetic modeling allows deciphering of the amounts and interconversion of the dominant $A\beta_{1-42}$ species. Moreover, the irreversible footprinting probe provides insights into the kinetics of oligomerization and subsequent fibrillar growth by allowing the conformational changes of $A\beta_{1-42}$ at sub-regional and even amino-acid-residue levels to be revealed. The middle domain of $A\beta_{1-42}$ plays a major role in aggregation whereas the N-terminus retains most of its solvent-accessibility during aggregation, and the hydrophobic C-terminus is involved to an intermediate extent. This approach affords an *in-situ,* real-time monitoring of the solvent accessibility of $A\beta_{1-42}$ at various stages of oligomerization, and provides new insights on the site-specific information on $A\beta_{1-42}$ for a sample state beyond the capabilities of NMR or x-ray crystallography.

2.2 Introduction

The aggregation of the amyloid beta (Aβ) into oligomers and fibrils is a key process associated with Alzheimer's disease (AD) (Figure 2.1). Among Aβ isoforms that present in AD, $A\beta_{1-42}$ is generally considered to be the most pathogenic.[1] Extensive effort has focused on characterizing

the conformation, size, and shape of Aβ aggregates (e.g., as dimers, pentamers, dodecamers[2-7] and fibrils[8-13]) (Figure 2.2), and the outcomes show a complex picture of $A\beta_{1-42}$ aggregation. Although soluble oligomers are thought to be the most critical players in the pathology of AD, large aggregates and fibrils may also be toxic.[1, 14] Unfortunately, we know little of their rates and extents of formation.



C

$A\beta_{1-40}$	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
$A\beta_{1-42}$	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA

Figure 2.1. Aβ and AD. (A) Accumulation of Aβ peptide leading to the formation of Aβ plaque. Two major forms of Aβ peptide, $A\beta_{1-40}$ and $A\beta_{1-42}$, are derived from amyloid precursor protein (APP) by enzymatic cleavage. (B) Aβ peptides are implicated in AD pathogenesis. (C) Sequences of $A\beta_{1-40}$ and $A\beta_{1-42}$. (A) and (B) are adapted from Kung, H. F., ACS Medicinal Chemistry Letters 2012, 3 (4), 265-267.

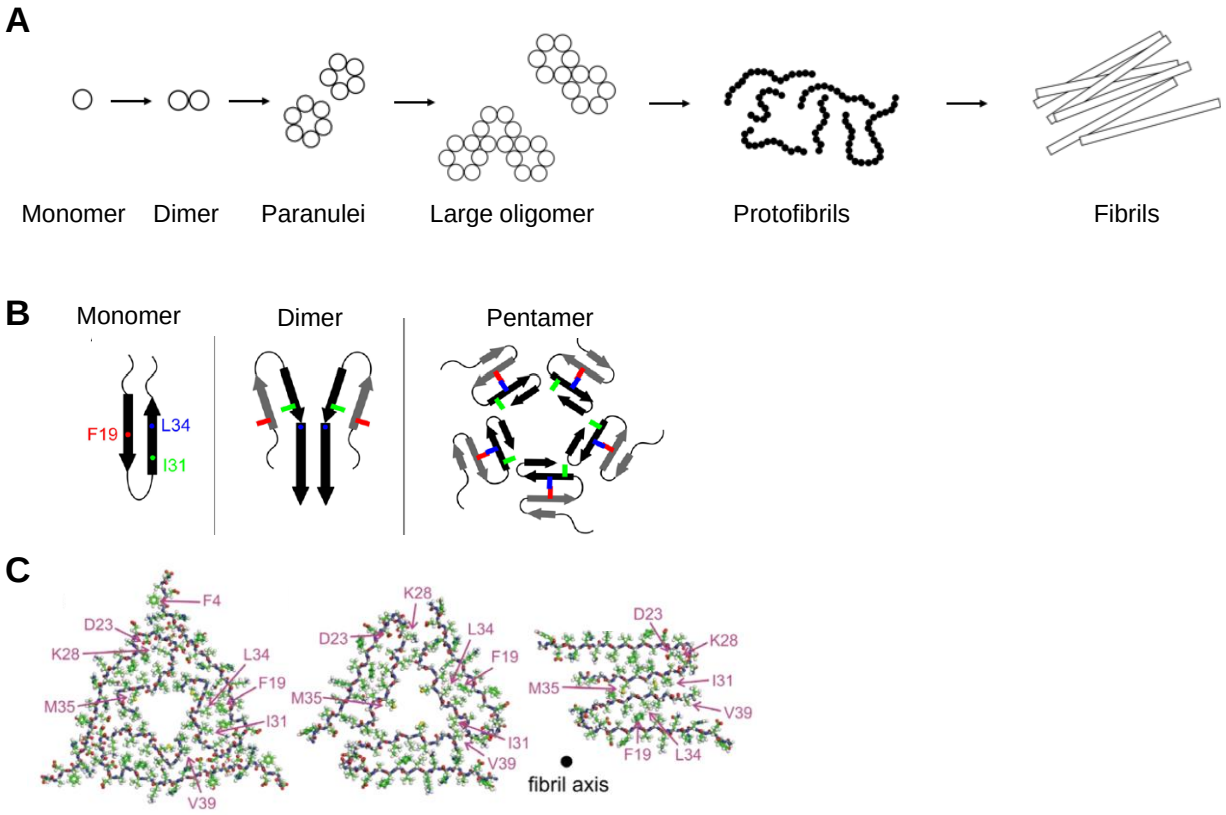


Figure 2.2. Aβ and its aggregates. (A) Aggregation pathway of $A\beta_{1\text{-}42}$ is a complicated multistep process. (B) Some proposed models of Aβ monomer, dimer, pentamer by solution-phase NMR. (C) Three structures of $A\beta_{1\text{-}40}$ fibrils reported by solid-phase NMR. Adapted from ref 6.

The most common approach in methodology for characterizing structure is multiphase NMR and X-ray crystallography, giving an essential, high resolution picture of the final state of Aβ fibrils while addressing some intermediates or oligomerized states that can be retained under non-physiologically relevant conditions. Owing to the vast heterogeneity and high aggregation propensity, however, the low molecular weight (low-MW) Aβ oligomers are not amenable to NMR or X-ray crystallography. Along a similar vein, atomic force microscopy (AFM) visualizes

morphology of the aggregates but provide no site specificity. More recently, Bowers et al.[15] implemented high resolution AFM to provide insight into $A\beta_{1-42}$ and $A\beta_{1-40}$ oligomers formed at early stage. The results with respect to the morphology is clear, but the process used to evaporate the droplet of Aβ solution before the visualization is not free of ambiguity.

Fluorescence may be the most widely adopted approach to follow Aβ aggregation, offering a solution medium that affords more physiological relevance. The signature measurement for amyloid formation is the fluorescence-based Thioflavin T (ThT) labeling assay that shows two regions or stages of aggregation separated by a sigmoid (Figure 2.3)[16]. Recently, Frieden et al.[17] labeled Aβ with tetramethylrhodamine (TMR) as a novel fluorescent reporter for oligomerization. This method reports more details, showing a lag and growth phases in amyloid formation. Although those fluorescence-based methods inform on the states of aggregation, only low structural resolution data is available. Despite their ease of use, they require either adding a dye molecule, which affects the accuracy in measuring amyloid-ligand interactions because anti-amyloid compounds such as polyphenols with strong absorption and fluorescence properties can significantly bias the fluorescence readout,[18] or pre-modifying Aβ, a perilous approach given that addition or subtraction of one amino acid changes significantly the properties of Aβ.

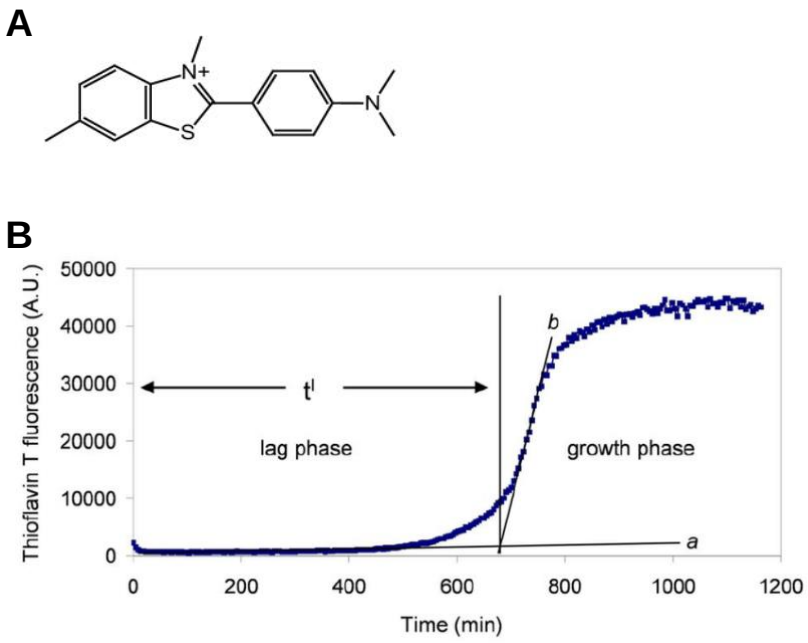


Figure 2.3. Thioflavin T (ThT) fluorescence assay for monitoring amyloid fibrilization. (A) Structure of ThT. (B) A typical ThT fluorescence curve representing two stages in amyloid formation, including a lag phase and a growth phase. Adapted from ref 16.

Mass spectrometry is now playing a role in understanding Aβ oligomerization, principally via measurements of the ion mobility of the gas-phase species.[3] Most ion-mobility work reports on small oligomers and short Aβ fragments, whereas the full protein and its very large soluble oligomers (n~100) likely hold the secret of its debilitating role in the human brain. Contemporaneously with the application of the TMR fluorescence method, we applied MS-based pulsed hydrogen-deuterium exchange (HDX)[19] and obtained similar results as Frieden et al. Although pulsed HDX avoids the problem of pre-modifying Aβ, the application of sophisticated proteomics for downstream protein analysis must be constrained to minimize the back-exchange resulted from hydrogen/deuterium scrambling.[20] As an alternative, Axelsen et al.[21] applied a

synchrotron-based hydroxyl radical footprinting to study $A\beta_{1-40}$ fibril and compared the outcome with those of multiple NMR structural studies. Despite all these efforts, the techniques we have currently, however, do not fully address the intermediate conformational transitions of $A\beta_{1-42}$ during aggregation. Thus, the transient aggregates of oligomerization remain insufficiently characterized, and our tools to characterize the oligomerization are temporarily exhausted.

In this report, we describe a new platform based on fast hydroxyl-radical footprinting (FPOP) and mass spectrometry to follow the aggregation of $A\beta_{1-42}$, which is regarded to be the most pathologically relevant Aβ isoform in AD. Unlike $A\beta_{1-40}$, the folding and assembly of $A\beta_{1-42}$ are highly complicated and multi-step process,[2-3, 6-7, 13-14, 22] (Figure 2.2A). Here, we seek an informative, time-dependent approach that not only follows oligomerization but also can be extended to test the effects of other proteins, lipids, and potential drugs on the monomer/soluble-oligomer transitions. The cutting-edge approach builds on proteomics measurements and MS-based footprinting. It utilizes hydroxyl radical-mediated irreversible reactions[23] initiated on a FPOP platform[24] to footprint $A\beta_{1-42}$ as it undergoes oligomerization. FPOP allows various $A\beta_{1-42}$ oligomeric species to be footprinted rapidly and irreversibly on amino-acid side chains. To locate the modified regions, we implemented Lys-N rapid digestion and LC-MS/MS to characterize $A\beta_{1-42}$ fragments. Our hypothesis is that FPOP modifications sensitively respond to the solvent accessibility changes of $A\beta_{1-42}$ upon self-association (Figure 2.4). This approach uses downstream sample processing independent of protein conformational changes occurring post footprinting to report changes on the $A\beta_{1-42}$ side chains at the global, peptide regional, and even amino-acid levels.

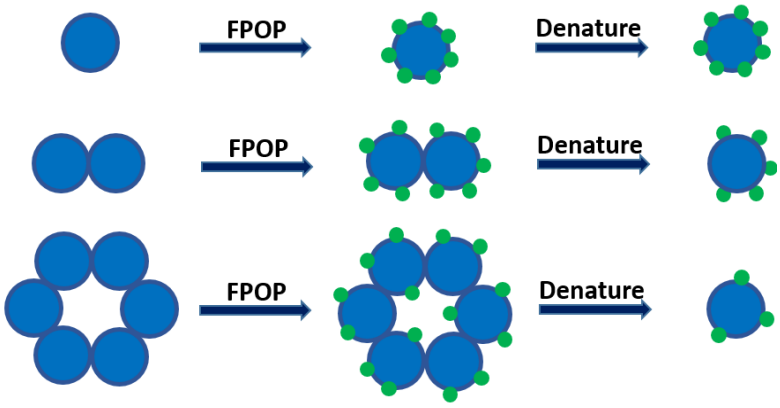


Figure 2.4. Schematic of using FPOP to probe Aβ aggregation. As Aβ (blue sphere) undergoes self-association, its decreased solvent accessibility is reflected by changes in FPOP modification (green sphere). The FPOP-modified Aβ are denatured for measuring the averaged modification extent of the peptide using LC-MS.

2.3 Materials and Methods

2.3.1 Materials

Synthetic wild type human amyloid beta 1-42 ($A\beta_{1-42}$) was from AnaSpec (San Jose, CA). Lys-N protease was from Seikagaku Corporation (Tokyo, Japan). Hexafluoroisopropanol (HFIP), sodium hydroxide (NaOH), *L*-glutamine, *L*-methionine, catalase, hydrogen peroxide (H_2O_2), phosphate buffer saline (PBS, 10 mM phosphate, 138 mM NaCl, 2.7 mM KCl), urea, formic acid and trifluoroacetic acid were from Sigma Aldrich (St. Louis, MO).

2.3.2 Formation of $A\beta_{1-42}$ aggregates

The purchased $A\beta_{1-42}$ was dissolved in HFIP at 1 mM and incubated at room temperature for 1 h to disrupt any pre-existing aggregates. The resulting solution was then aliquoted into tubes, and

the HFIP was evaporated in a fume hood. After solvent evaporation, a clear film of $A\beta_{1-42}$ remained at the bottom of the tube. HFIP treatment was then repeated two more times. Samples were then frozen in -80 ºC for future experiments. Prior to incubation for aggregate formation, HFIP-pretreated $A\beta_{1-42}$ was dissolved in 3 mM NaOH (pH 11.7) and incubated without stirring at room temperature for 3 min followed by sonication for 1 min. Aggregation was initiated upon diluting the $A\beta_{1-42}$ NaOH solution by 20 fold (v/v) with PBS buffer (pH 7.4). The final $A\beta_{1-42}$ concentration after dilution was 10 μM. Different extents of aggregation were achieved by varying the incubation time from 0 to 48 h in PBS buffer (25 ºC or 37 ºC). For each time point, the incubation and analysis was done in triplicate.

2.3.3 FPOP labeling of $A\beta_{1-42}$

At the various incubation times, $A\beta_{1-42}$ was immediately submitted to FPOP labeling. FPOP was performed as previously described.[25] Briefly, a KrF excimer laser (GAM Laser Inc., Orlando, FL, USA) generated laser beam with an excitation wavelength of 248 nM initiated H_2O_2 photolysis to give hydroxyl radicals. The $A\beta_{1-42}$ in PBS was mixed with 20 mM glutamine (scavenger) and 20 mM H_2O_2 just prior to injection into the flow tubing for FPOP. The flow rate was adjusted according to the width of the laser irradiation window to ensure 20% irradiation-excluded volume and to minimize repeated laser exposure. The total time for one sample to pass through the silica tubing was ~2 min. After laser irradiation, the sample was collected in a tube containing 10 mM catalase and 20 mM methionine to decompose leftover H_2O_2 and prevent oxidation-artifacts during storage. For each aggregation time point, $A\beta_{1-42}$ was incubated independently in triplicate and subjected to FPOP. In addition, control samples of $A\beta_{1-42}$ were handled in the same manner in triplicates, but not laser-irradiated. Each FPOP-labeled sample

was transferred into aliquots for intact $A\beta_{1-42}$ characterization and enzymatic digestion, respectively.

2.3.4 Proteolysis

Urea was added to each 20 μL aliquot of the FPOP-labeled $A\beta_{1-42}$ sample to give a final concentration of 6 M. Lys-N was then added with an enzyme to protein ratio of 1:10 (w/w). The concentrated urea facilitated rapid digestion and prevented hydrophobic peptides from self-assembly during digestion. Samples were incubated at 45 ^{o}C for 30 min. The digestion was then quenched by adding trifluoroacetic acid to a final concentration of 1% (by volume).

2.3.5 Mass spectrometry

For intact $A\beta_{1-42}$ characterization, 20 μL of the FPOP-labeled sample was directly submitted to an Agilent 1100/1200 separation system at a flow rate of 200 μL before being admitted to a MaXis 4G quadrupole-time-of-flight (Q-ToF) mass spectrometer (Bruker Daltonics, Billerica, MA). Solvent A was water + 0.1% formic acid, and solvent B was 80% acetonitrile + 0.1% formic acid. The gradient started from 5% B and increased to 15% B in 0.3 min, ramped to 50% B in 5.2 min, increased to 100 % B in 0.5 min, held at 100 % B for 0.5 min, returned to 5% B in 0.1 min and equilibrated at 5 % B for 2 min. The mass spectrometer was operated in the positive-ion electrospray ionization mode at a mass resolving power of 30,000 (*m/z* 400).

For peptide and residue level analysis, 5 μL digested sample was submitted to LC-MS/MS analysis. Samples were pre-concentrated on an Acclaim PepMap C18 column (Thermo Scientific, 100 μm × 2 cm, 5 μm, 100 Å) and desalted for 15 min before elution. Separation was performed on a 15 cm custom-packed C18 column (Magic, 75 μm × 15 cm, 5 μm, 200 Å) maintained at 65 ^{o}C by using a Nano UltiMate 3000 Rapid Separation system (Dionex, Co.).

Solvent A and B were the same as above. Peptides were eluted at a flow rate of 700 nL/min with the following gradient: 2% B to 20% in 1 min, ramped to 70% B in 10 min, increased to 90% in 1 min, held at 90% for 3 min, returned to 2% B in 1 min, and equilibrated at 2% B for 4 min. LC separation was directly coupled to online detection using a Q Exactive Plus hybrid quadrupole orbitrap mass spectrometer with a Nanospray Flex ion source (Thermo Fisher, Santa Clara, CA). The top ten abundant ions seen in the mass spectrum were subjected to higher energy collision dissociation (HCD) for identification and characterization of possible FPOP modifications. The mass resolving power was 70,000 (*m/z* 400) for MS1 and 17,500 (*m/z* 400) for MS/MS.

2.3.6 Data analysis

Whole protein-level analysis with the custom program afforded the fraction of unmodified $A\beta_{1\text{-}42}$ after FPOP for all samples. Signal intensities for the modified (I_{ox}) and unmodified species (I) were integrated from the raw data files, either with the custom program for the intact $A\beta_{1\text{-}42}$ or Thermo Xcalibur for the digested peptide and residue-level analysis. The extent of modification was calculated by using the following equation, as previously described,[25]

Eq. 2.1 $$\%\, modified = \frac{\Sigma I_{ox}}{\Sigma I_{ox} + \Sigma I} \times 100$$

For analysis of $A\beta_{1\text{-}42}$ digested peptides, product-ion spectra obtained with the orbitrap mass spectrometer were searched for peptide identification by using Mascot (Matrix Science, London, UK) software. Mass of unmodified peptides and assigned modifications validated by manual inspection were input into an inclusion list to afford better sampling and identification of FPOP modifications.

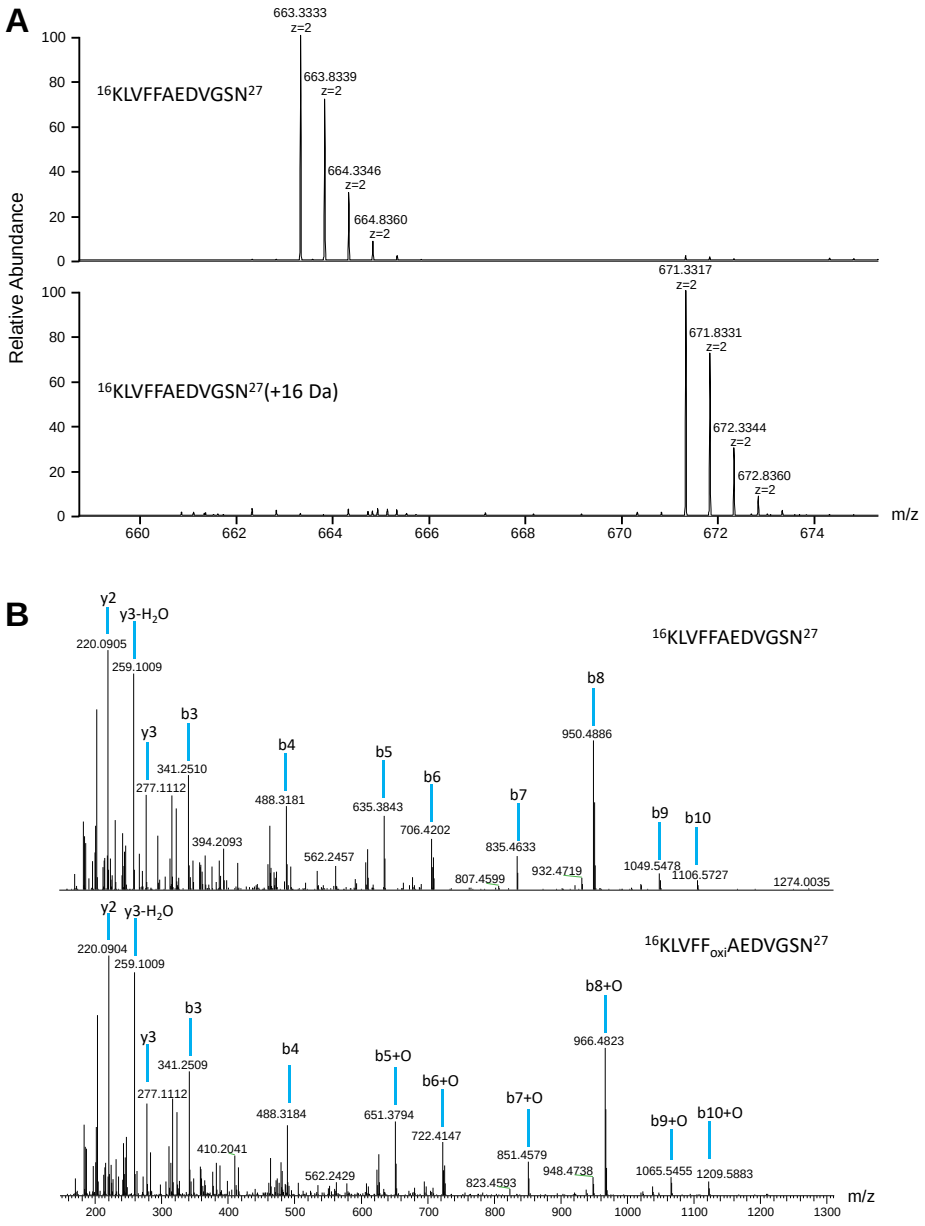


Figure 2.5. An example of FPOP identification. (A) Mass spectra of peptide 16-27 (doubly charged), unmodified peptide (top), and modified peptide (bottom). (B) Product-ion spectra (MS/MS) of the unmodified (top) and modified (bottom) peptide 16-27. The MS/MS spectra of the modified peptide was integrated over the signal with the retention time of 10.2 min and indicates FPOP oxidation on F20 of $A\beta_{1-42}$.

Modification sites on the peptide were assigned on the basis of product-ion (MS/MS) spectra (Figure 2.5). In a few cases, the location of a modification to a single residue was not possible owing to limited fragmentation information from MS/MS or to the presence of interference from co-elution of peptide isomers. In that case, the modification was indicated to occur on a set of possible residues. For any specific residues, the fraction modified was calculated by the following equation as the ratio of the signal of each peptide modified at that residue ($\Sigma I_{ox\ on\ residue\ X}$) to the total intensity of modified and unmodified peptide signal spanning this residues.[26]

Eq. 2.2 For amino-acid residue, $\%\ modified\ = \frac{\Sigma I_{ox\ on\ residue\ X}}{\Sigma I_{ox} + \sum I} \times 100$

2.3.7 Kinetic Modeling

The fraction of FPOP modified molecules was modeled on the basis of Scheme 2.1 by using six rate constants as model parameters (see "Fitting Parameters" in SI). The overall FPOP fraction modified F(t) was computed as shown in Eq. 2.3 from the species concentrations ([D], [D*], [D**]), which varied with time, and the species fraction-modified (F_0, ΔFD, ΔFD*, ΔFD**) that are also model parameters. F_0 is the fraction modified for $A\beta_{1\text{-}42}$ monomer. The monomer concentration [M] is implicated by the experimental starting concentration of the monomer M_T.

Eq. 2.3 $$F(t) = F_0 - \left(\frac{\Delta FD \cdot 2[\mathrm{D}]}{M_T} + \frac{\Delta FD^{*} \cdot 2[\mathrm{D}^{*}]}{M_T} + \frac{\Delta FD^{**} \cdot 2[D^{**}]}{M_T}\right)$$

The model curve fractions modified $F(t)$ were calculated by solving differential equations (see "Fitting Parameters" in SI) representing each reaction in Scheme 2.1 with are time-dependent species concentrations and the postulated species fractions modified (F_0, ΔFD, ΔFD*, ΔFD**). The quantities of ΔFD, ΔFD*, ΔFD** are properties (in term of fraction modified) of the

proposed $A\beta_{1-42}$ species in various oligomeric states (monomeric $A\beta_{1-42}$ as the reference). Note that the definition of the fraction modified (Eq. 2.1) should not be interpreted in terms of intrinsic reactivities with the hydroxyl radical; that is, the differences in the fraction modified signaled the presence of different $A\beta_{1-42}$ species with a characteristic protection level.

The process for determining the model parameters and the kinetic curve was implemented in Mathcad v.14.0 M020 (Parametric Technology Corp.). The "Nonlinear Quasi-Newton" mode was used in the "Minimize" function in a search for the solution model parameters. In each trial, the postulated normalized rate constants were converted to their physical values by multiplication by the initial physical rate constants to accommodate the numerically large range of rate constants. The adaptive step-size fourth order Runge-Kutta "Rkadapt" function was then used to solve the system of ordinary differential equations corresponding to Scheme 2.1 with initial conditions that all concentrations were zero except $[M] = M_T = 10\ \mu M$.

Generally, all ten model parameters were varied in a search to minimize the root-mean of the squares of the residuals between experiment data and model curve as shown in Figure 2.8 for $A\beta_{1-42}$ and the solid curves shown in Figure 2.10 and Figure 2.11. For the dashed curves in Figure 2.10 and Figure 2.11, all rate constants were fixed to the values obtained from the model whose curve is shown in Figure 2.8A for $A\beta_{1-42}$ while the four postulated species fractions modified were varied.

2.4 Results and Discussion

2.4.1 Aggregation of $A\beta_{1-42}$ at the global (protein) level

Given that $A\beta_{1-42}$ aggregation is time-sensitive and continues after footprinting, it is important to utilize an irreversible footprinting "snapshot" to capture accurately the oligomerization intermediates. In our triplicate experiments, formation of $A\beta_{1-42}$ aggregates begins with incubating $A\beta_{1-42}$ monomers in PBS buffer for up to 48 h (Figure 2.6), followed by footprinting of the protein occurring after irradiation by a labeling laser pulse in a flow system. MS analysis gives representative mass spectra of intact $A\beta_{1-42}$ labeled by hydroxyl radical. The oxidative labeling occurs at solvent-accessible $A\beta_{1-42}$ side chains to give variously labeled proteins observed as isotopic clusters (Figure 2.7).

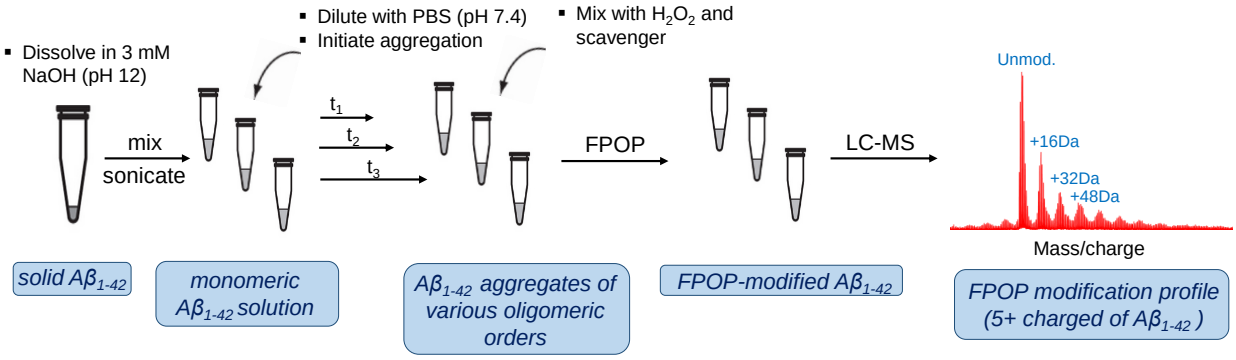


Figure 2.6. Preparation of $A\beta_{1-42}$ aggregates for LC-MS analysis.

Owing to its intrinsically disordered structure, 77% of unfolded $A\beta_{1-42}$ monomers undergo modification (Figure 2.7B) at short times. As $A\beta_{1-42}$ adopts some conformational order by folding into higher-order oligomers, it loses solvent accessibility, gains protection, and undergoes correspondingly decreased FPOP modification (Figure 2.7C-E). At the longest incubation times, $A\beta_{1-42}$ becomes highly resistant to modification, indicating formation of mature

aggregates, presumably with beta-sheet bundles (Figure 2.7F).[10, 27] The FPOP modification pattern for various charge states of $A\beta_{1-42}$ produced in the electrospray ionization are consistent, as expected, with a relative error of less than 2% for the calculated extent of modification of various charge states.

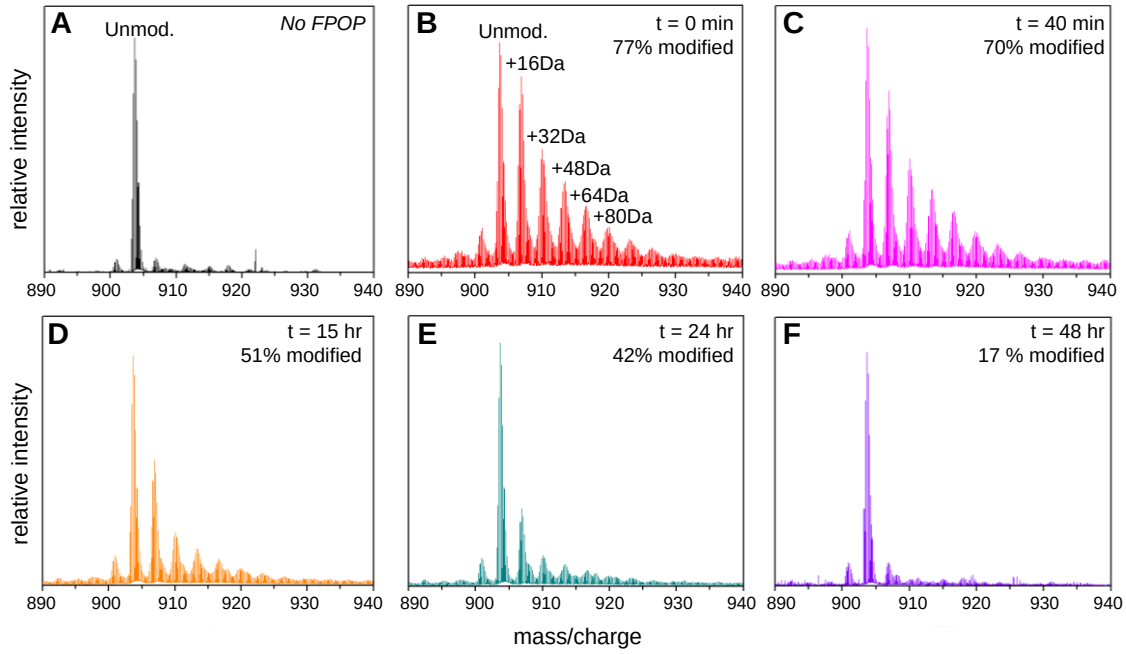


Figure 2.7. Mass spectra showing extents of FPOP modification for intact $A\beta_{1-42}$ (5+ charged) as a function of incubation time. (A) $A\beta_{1-42}$ monomer as a control with all reagents, including H_2O_2, added and flowed through the FPOP tubing but without laser irradiation, (B) extensively hydroxyl radical-modified, unstructured $A\beta_{1-42}$ monomers, (C-E) decreasing FPOP modification extents of $A\beta_{1-42}$ aggregates, reflecting increasing structural protection to FPOP modification, (F) minimal $A\beta_{1-42}$ modification, reflecting solvent-inaccessible, highly ordered core structure of the fibrillar aggregates that resist FPOP modification. In B-F, t represents the time of incubation.

We calculated the fraction modified of the full-length $A\beta_{1-42}$ and its digested peptide (as discussed later) from the ratio of the signal of the modified peptide (I_{ox}) to that of the total amount of the modified and unmodified peptides (I) (Eq. 2.1), as described in the experimental section. By adopting this metric, we can pursue the change in the amount of modified species raised from $A\beta_{1-42}$ conformational change, regardless of the intrinsic reactivity or reaction kinetics of the peptide with hydroxyl radical. The use of the above metric for quantifying the level of modification is not unusual (see Xie and Sharp[28]). Here the fraction unmodified is directly inferred as one minus fraction modified, and for a given peptide or residue it increases monotonically (because it is the product of the fraction unmodified at each reaction site) as the reactivity of the overall peptide or specific residue decreases owning to $A\beta_{1-42}$ association. The sensitivity of this metric to increasing reactivity diminishes, however, as the percent unmodified becomes very small. For small proteins like $A\beta_{1-42}$ and, more importantly, for amino-acid residues, the relative amount of the unmodified species remains high; thus, this is of minor concern under the conditions of our footprinting.

Another possible concern for processing modified peptides is that some residues can decrease in reactivity while others increase. One might suggest as a solution to this concern a metric in which the intensities of oxidized species are weighted by the number of oxygens.[29] This would be analogous to the centroid calculation often used as an algorithm for HDX footprinting of proteins. This proposed solution does not always address the issue[30] and has the potential to magnify uncertainties arising from signals with low signal-to-noise ratio.[31] A similar problem applies to centroid calculations (i.e., weighing the intensities by the number of modifications) in hydroxyl radical footprinting. Therefore, we choose the metrics (Eq.1) that does not compromise

the accuracy in quantification and meanwhile sufficiently provides insights in reporting the conformational change of $A\beta_{1-42}$ as it associates.

The oxidation patterns of $A\beta_{1-42}$ at various incubation times in the FPOP experiment follow a Poisson-distribution;[25] that is, the signal of unmodified $A\beta_{1-42}$ is the most intense, followed by those corresponding to the oxidized species with the mass shifts of +16, +32, +48 Da etc. in a sequentially descending manner. The patterns of various extents of oxidation follow this manner as the overall oxidation level decreasing owning to the folding of $A\beta_{1-42}$.

2.4.2 Structural Rationalization of Multiple Kinetic Phases

The hydroxyl radicals are comparably sized as water molecules, allowing them to modify solvent-accessible side chains and report on the transient states of soluble aggregates. We chose to model the $A\beta_{1-42}$ aggregation kinetics, observed at 25 ^{o}C (Figure 2.8A), on the basis of the nucleation-autocatalytic growth mechanism proposed by Finke et al.[32] with modifications that address the transient stages observed in the present case.[3, 33] This modeling approach is perhaps the simplest model that accounts for the various stages of aggregation. Two nucleation and two autocatalytic reactions, as shown in Scheme 2.1, were used to describe the time-dependent aggregation of $A\beta_{1-42}$, one more stage than seen by fluorescence and pulsed HDX.

With kinetic modeling (see experimental section) based on Scheme 2.1 and support from other published work,[3, 14, 33] we can follow the aggregation through five stages: AB, BC, CD, DE, and EF (Figure 2.8A), instead of the four observed previously by pulsed HDX and fluorescence.[17, 19] In the modeling, our assumption is that the monomer concentration at time zero is 10 μM, which is the concentration upon diluting $A\beta_{1-42}$ into the PBS buffer, and the solution at that time is free of oligomers.

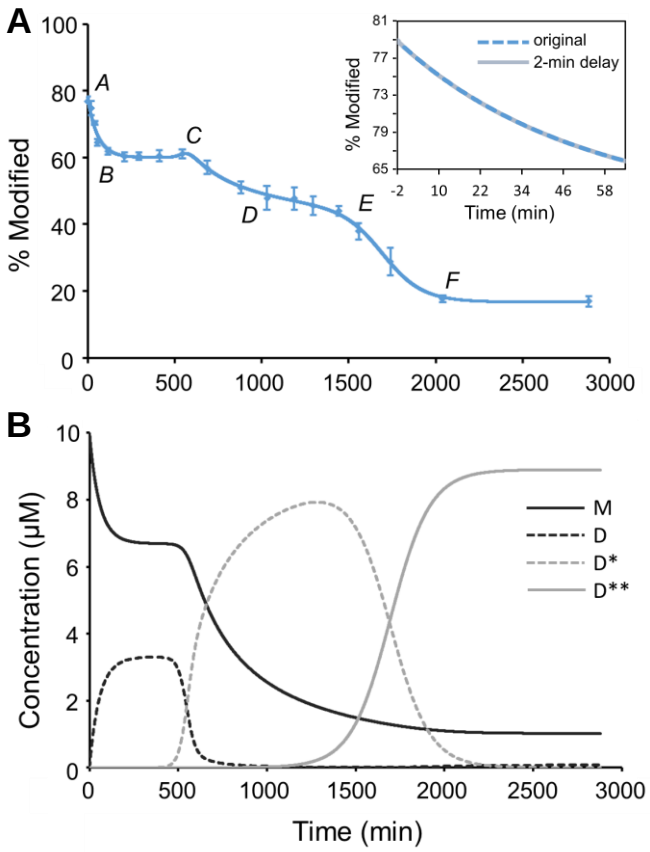


Figure 2.8. Characterization of Aβ$_{1-42}$ aggregation on the global (full-polypeptide) level by a kinetic simulation. (A) points represent experimental data (10 μM, 25 ^{o}C, pH 7.4, no agitation), and the solid curve is a model fit based on two autocatalytic reactions. The inset in (A) shows a comparison of the fitting with and without the consideration of the 2-min FPOP dead time. (B) Time-dependent concentrations of the various Aβ$_{1-42}$ species from the solution to the differential equations of the fitted model in (A). M-monomer; D-dimer; D*-product of 1st catalytic reaction, D**-product of 2nd catalytic reaction. For each species, the concentration is of monomeric Aβ$_{1-42}$ equivalents.

The stage resolved additionally by FPOP is represented by transition from D to F, which leads to the formation of lateral protofibrillar and then to fibrillar aggregates, whereas the transition is a single growth phase in the pulsed HDX and fluorescence platforms. In Figure 2.8A, the first stage corresponds to (1) $A\beta_{1-42}$ monomers assembling rapidly to form dimers and small oligomers. As the monomers continue to oligomerize, their solvent accessibility and concomitant FPOP reactivity decrease in a fast, exponential-like manner (AB). (2) The early-formed small oligomers cooperatively nucleate to provide paranuclei-like "seeds" for subsequent aggregation (transition BC). During this time, the oxidation extent cannot significantly change because the monomers have almost ceased to disappear. (3) As the concentration of "seeds" reaches a critical threshold, larger oligomers form in an autocatalytic fashion (transition CD).[34] (4) Lateral oligomers likely with the beta-strand structure associate, elongate into large aggregates/protofibrils presumably with in-register beta sheets (transition DE),[33, 35] and (5) large aggregates/protofibrils further associate and catalyze the reorganization of other $A\beta_{1-42}$ species into mature fibrils presumably with the well-defined beta-sheet structure (transition EF).[10]

Transition A-B. $M + M \rightleftharpoons D$ *(dimerization)*

Lag stage B-C. $D \rightarrow D^*$ *(1^{st} nucleation)*

Transition C-D. $D + D^* \rightarrow D^* + D^*$ *(1^{st} autocatalytic reaction)*

Lag stage D-E. $D^* \rightarrow D^{**}$ *(2^{nd} nucleation)*

Transition E-F. $D^* + D^{**} \rightarrow D^{**} + D^{**}$ *(2^{nd} autocatalytic reaction)*

M- monomer *D*- product of 1^{st} catalytic reaction*
D- dimer *D**- product of 2^{nd} catalytic reaction*

Scheme 2.1 Proposed kinetic scheme for $A\beta_{1-42}$ aggregation.

We chose Scheme 2.1 to support the observation that there are additional stages reported in the aggregation kinetic curve while applying the principle of parsimony in modeling. The monomer to dimer transition (reaction *A*-*B*) is illuminated by the work of Frieden et al[17], in which early changes in their kinetic curves with different starting Aβ concentrations behaved consistently with a model of dimer and trimer formation. Here we have not deployed the trimer formation component because we argue that the amount of additional protection afforded by the trimer over the dimer is not sufficiently large to justify the additional parameters that the trimer would require in the scheme. Visual inspection of the kinetic curves shows two steps; each step is characteristic of a slow nucleation phase (reaction *B*-*C* and *D*-*E*) followed by a rapid autocatalytic growth phase (reaction *C*-*D* and *E*-*F*). Given there are two steps, we invoked the Finke-Watzky two-step mechanism twice in succession, which might be described as double autocatalytic.[36]

The use of the Finke-Watzky two-step mechanism does have limitations. It does not account for the equilibrium between the $A\beta_{1-42}$ monomer and other species.[37] An alternative and still minimal model that addresses this issue is by Crespo et al[38]. In addition, nucleation and autocatalysis models augmented with fibril growth and fragmentation fail to match correctly the lag time scaling as a function of monomer concentration for a number of fibrilizing protein systems.[39]

The time for the each $A\beta_{1-42}$ passing through the FPOP workflow is 2 min; thus, for each experimental time point, the FPOP fraction modified quantifies an assemble average of $A\beta_{1-42}$ conformations over that 2 min in addition to the reported aggregation time. Although this "dead time" in FPOP will have a minor influence on the samples with long-time incubation, it may affect the early-stage characterization of the curve as the monomer will aggregate to some extent in this short period of time. To test, we repeated the simulation by shifting the actual time of

aggregation by 2 min as a maximum and extrapolated the curve to the actual start point. We found the overall fitting of the aggregation curve is not significantly affected by considering the FPOP dead time. In a zoom-in view of the first 60 min (inset in Figure 2.8A), the "-2 min" on the time axis of the solid grey curve is regarded as the actual starting point of the aggregation that modeled in the testing trial. The outcomes overlap well with curve neglecting the 2 min (dashed blue) even for the early aggregation stage as one can think it as the aggregation curve is right-shifted by 2 min. According to the kinetics revealed by TMR fluorescence, which does provide clearer time resolution for the first a few minutes of $A\beta_{1-42}$ aggregation, the initial exponential decrease in fluorescence intensity representing the monomer-dimer-trimer transition proposed happens over 30-45 min for the $A\beta_{1-42}$ with concentrations ranging from 0.5 μM up to 4 μM with stirring the solution.[17] From the pulsed HDX results, although the initial phase is not well defined, it shows a burst tendency of increasing protection over 30 min for $A\beta_{1-42}$ of 50 μM without any stir.[19] Taken these observations together, we believe that it is unlikely that the dead time from the FPOP labeling has a substantial influence on the aggregation kinetics. If a burst phase occurred so rapidly in the 2 min prior to laser irradiation to initiate FPOP such that we no longer can explain transition AB, the kinetics will need to be addressed by involving a model justifying the additional conformation change in the burst phase as well as transition AB tandem to it.

Although we calculated the reaction rate constants for each stage by solving the differential-equation for each reaction in Scheme 2.1, we are not quoting them because they do not directly relate to the actual rate constants. Each rate constant in the model is condensed and simplified as a summary of many underlying reaction steps.[32] The "dimer" species, for example, is expected to be made up of many species, each formation with its own rate constant. Others noted that

nucleation rate constants for the lag phase are unreliable.[39] Reliable rate constants associated with the growth phases require representation of the maximum slopes during the growth phase, which are not present in the data. Nevertheless, we can specify the species concentrations as a function of the time obtained from the model and describe the disappearance of $A\beta_{1-42}$ monomers (M) and the formation of various oligomer including paranuclei (D), protofibrils (D*) and fibrils (D**) (Figure 2.8B). In the experiment, the initial concentration of $A\beta_{1-42}$ monomers is 10 μM. In the early oligomerization, ~30% of the monomers rapidly assemble into low-MW oligomers (solid black curve), followed by a lag phase during which low-MW oligomers slowly accumulate until the concentration is sufficient to catalyze (seed) the formation of larger oligomers (dashed black curve). As the amount of large aggregates/protofibrils reaches a concentration corresponding to ~80% consumption of the original monomer equivalents (dashed grey curve), a structural reorganization occurs to deplete the monomers and form rapidly mature fibrillar aggregates (solid grey curve). At the end of the aggregation, equilibrium is nearly achieved for fibrils and persistent, leftover monomers (~10%). At this stage, the rapid second autocatalytic reaction almost shuts down the first autocatalytic reaction, and, as a result, the monomer concentration becomes relatively stable.

A previous study shows that $A\beta_{1-42}$ and $A\beta_{1-40}$ have distinct aggregation pathways.[6] The assembly of $A\beta_{1-42}$ involves formation of several distinct transient structures to form paranuclei that assemble further to form superstructures similar to early protofibrils, whereas $A\beta_{1-40}$ mainly forms low-MW oligomers at a much slower rate than that of $A\beta_{1-42}$ at similar concentration.[3, 19] We have preliminary results from the FPOP study of $A\beta_{1-40}$ under the same conditions (10 μM, pH 7.4, 25 ^{o}C, no agitation) as for $A\beta_{1-42}$ that this proteoform shows extensive but relatively constant FPOP modification over the time scale we investigated (0~48 h) (Figure 2.9A),

indicating that it is considerably less reactive than $A\beta_{1-42}$ in term of the aggregation propensity. Further, we examined the aggregation of $A\beta_{1-42}$ at 37 ºC (Figure 2.9B). The higher temperature accelerated $A\beta_{1-42}$ aggregation substantially without showing any lag or intermediate phases. Within the first 8 h of incubation, the FPOP fraction modified decreases sharply and then levels off, suggesting either the nucleation process happens so rapidly that no apparent lag phase is observed or the oligomerization pathway changed from a self-catalytic seeding mechanism.

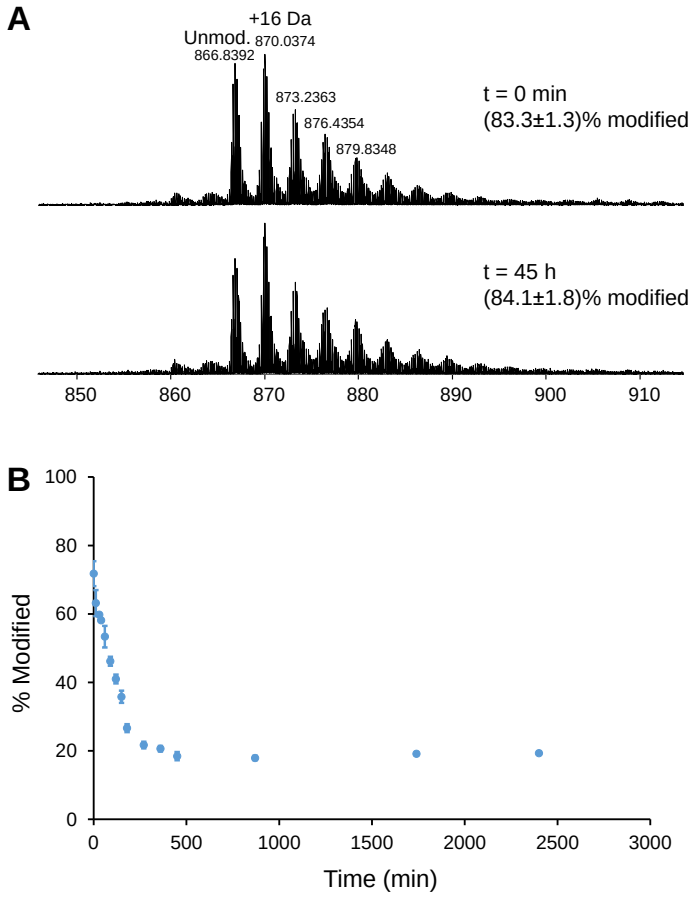


Figure 2.9. Aggregation of $A\beta_{1-40}$ at 25 ºC and of $A\beta_{1-42}$ at 37 ºC. (A) Mass spectra of $A\beta_{1-40}$ (5+ charged) with 0 min and 45 h incubation, 10 μM, 25 ºC, pH 7.4, no agitation. (B) $A\beta_{1-42}$ aggregation curve, 10 μM, 37 ºC, pH 7.4, no agitation.

2.4.3 High Resolution View of Oligomerization

The FPOP approach permits a high-resolution view of the oligomerization down to even the amino-acid level for some residues. To interrogate the conformational changes at a regional level, we proteolyzed the irreversibly footprinted protein and analyzed the resulting peptides by LC-MS/MS. Normally, this is a straightforward process for soluble proteins, but $A\beta_{1-42}$ and its hydrophobic C-terminal proteolytic fragments continue to associate during the digestion, resulting in appreciable resistance to long-term proteolysis, a loss in mass spectral signals, and a possible bias towards those regions that are less prone to association.[19, 27] These compounded issues motivate a kinetic study that targets high structural resolution by effective and rapid digestion of $A\beta_{1-42}$ aggregates. We used Lys-N protease under optimized digestion conditions to address successfully this issue. Taking advantage of Lys-N maintaining its high enzymatic activity in 6 M urea,[40] we reconstituted a post-labeling aliquot of $A\beta_{1-42}$ in 6 M urea and incubated it with Lys-N (enzyme: protein ratio of 1:10 (w/w)) at 45 ^{o}C to digest the $A\beta_{1-42}$ as quickly as 30 min. The high-concentration of urea present in the digestion not only affords rapid proteolysis by denaturing $A\beta_{1-42}$, but minimizes non-covalent association of the hydrophobic peptides. More importantly, Lys-N cleaves $A\beta_{1-42}$ at the amino side of lysine residues, leaving the hydrophilic residue K28 on the highly hydrophobic C-terminal proteolytic fragment, thereby increasing the sensitivity by two orders of magnitude for the peptide covering this region compared to the corresponding peptide formed when trypsin is used (data not shown), improving quantification and accuracy.

2.4.4 Aggregation of $A\beta_{1-42}$ at the Peptide and Residue Levels

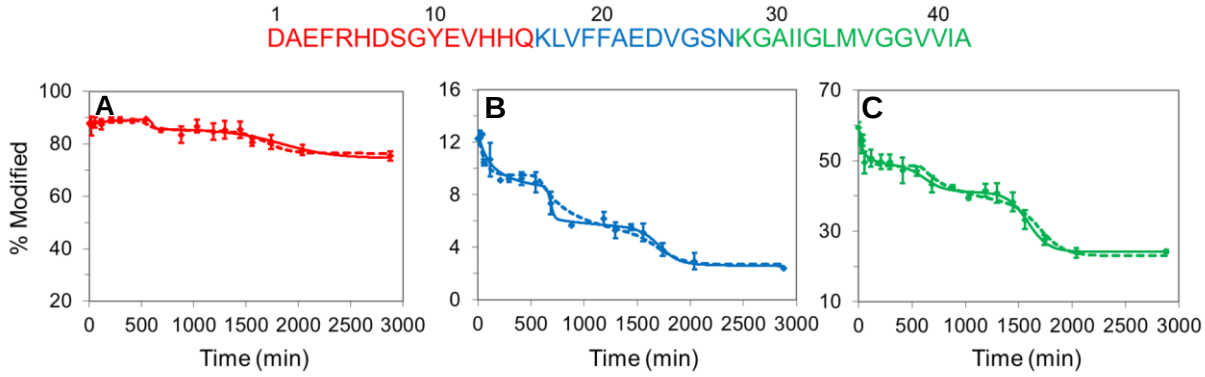


Figure 2.10. Time-dependent FPOP labeling of $A\beta_{1-42}$ incubated at 25 °C, pH 7.4, no agitation and kinetic simulations for Lys-N digested $A\beta_{1-42}$ peptides. Oligomerization of (A) N-terminal region 1-15, (B) middle region 16-27, and (C) C-terminal region 28-42. Solid curves are simulations that afford rates constants for each peptide treated independently, whereas the dashed curves are simulations constrained by global rates.

We characterized further the conformational changes in aggregation for the regions represented by the three peptides resulting from Lys-N digestion (N-terminal region 1-15, middle domain 16-27, C-terminal region 28-42). We simulated the kinetics for each region by using the model described above (Figure 2.10), both independently and constrained by the outcome of the global (protein)-level rate constants. As discussed above, we expect regions that undergo conformational changes and serve as oligomerization interfaces to exclude solvent and show decreased oxidative modification. The different scales of fraction modified for each peptide is related to the distinct reactivities of the corresponding $A\beta_{1-42}$ sub-regions. The reactivity is a function of: 1) the intrinsic reactivity of those sub-regions as contributed by the amino acid residues with respect to its primary sequence, and 2) the protection from the secondary and/or

ternary structure of $A\beta_{1-42}$ for that sub-region. Given that FPOP modification of the N-terminal peptide does not change significantly as a function of aggregation time, this region must remain structurally flexible with little self-association and little loss of solvent accessibility as $A\beta_{1-42}$ associates. This is in accord with solid state-NMR data, indicating that the N-terminal region remains disordered in various $A\beta_{1-42}$ oligomers and fibrils.[2, 9] In contrast, the extent of FPOP modification significant decreases (protection increases) over time for the middle domain peptide 16-27 and the C-terminal peptide 28-42 (6 and 2.5 times decreases in rate from $A\beta_{1-42}$ monomers to fibrils, respectively). This indicates convincingly the importance of those two regions in $A\beta_{1-42}$ aggregation whereby the central region must play a prominent role and serve as the self-association interface to drive aggregation. This is key experimental evidence that the middle region is a nucleation interface, as suggested by molecular dynamics studies.[35, 41]

The peptide-level results are better modelled by allowing the rate constants for each peptide region to be independent (Figure 2.10, solid curves). When the modeling is restricted by the rate constants from fitting for the whole protein, we find relatively pronounced discrepancies between the results and the simulated curves (Figure 2.10, dashed curves). Assuming the discrepancy is not solely contributed by any structural bias in the digestion, we suggest that regions of $A\beta_{1-42}$ act nearly independently in the oligomerization of $A\beta_{1-42}$ and that the local (regional) rates are not predictable from the kinetic results for the whole protein. Notice that in the pulsed HDX, the "half lives" of the peptide-level transitions are also different[19] and show that the center and C-terminal regions are more important in the oligomerization than is the N-terminus.

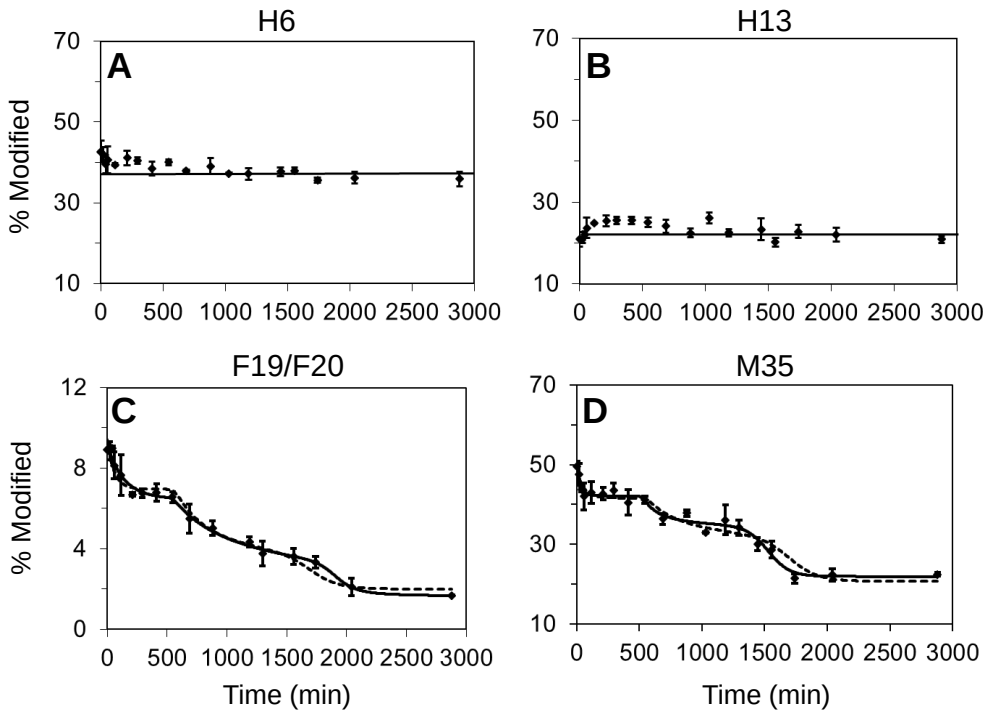


Figure 2.11. FPOP results for Aβ$_{1-42}$ amino acid residues and kinetic simulations. (A) H6 showing little participation in the oligomerization, (B) H13 showing slightly increasing modification initially and then level off, (C) and (D) F19/F20 and M35 showing large changes of solvent accessibility along with oligomerization (FPOP modifications on F19 and F20 were summed to improve accuracy in quantification).

Furthermore, measuring the modification at the amino-acid level allows characterizing aggregation tendencies for Aβ$_{1-42}$ at that level at least for some of the amino-acid residues (Figure 2.11). The oxidative modification on the residue level was assigned by using the product-ion spectra of the modified peptide precursor ion. Their aggregation curves are generated based on measuring the ratio of the signal abundance from that specific oxidized residues to the total amount of the peptide observed that contained that residue (see experimental section). The

modification for H6, H13, F19, F20 and M35 are well resolved chromatographically with quantifiable extent of oxidation at better than a signal-to-noise ratio of 10:1. We also identified FPOP modifications on H12, L17, V18, V39/V40/I41. The extent of modification for those residues, however, is of very low abundance (less than 1%) and not well sampled; and thus in the present case not quantified.

Residues H6 and H13 show an overall constant extent of FPOP modification during aggregation, indicating that few conformational changes occur for these two residues (Figure 2.11). Interestingly, for H13 despite showing little change in the fraction modified, this residue shows a clear trend of an initial increase in FPOP modification within the first 30 min, suggesting this site becomes relatively solvent-exposed and could be a critical nucleation site either to promote or delay the next transition. Note that the absolute FPOP modification extents for residues of the same type are only related to the surrounding steric environment and protection. The modification extent of H6 is 40% and of H13 is 19%, whereas that of H12 is too low to characterize accurately. These results suggest that, although the N-terminal region remains structurally disordered during aggregation, the solvent accessibilities of these residues represent different microenvironments. H6 is the most solvent-exposed among the three, followed by H13 and then H12. For F19 and F20 we summed the FPOP modifications before plotting those values. These two residues show a decrease in FPOP modification extent by 4.5 times in going from monomers to fibrils, and M35 from the C-terminal region showed a similar decrease of 2.5 times (Figure 2.11). For the middle domain of $A\beta_{1-42}$ represented by peptide 16-27, the difference in modification largely occurred on F19/F20, because not only is phenylalanine highly reactive to hydroxyl radical compared to other nearby residues[42], but their high hydrophobicity potentially driving the oligomerization. We envision that the ability of FPOP to reveal the site-

specific aggregation can be applied to understand interactions of Aβ with other molecules, especially drug candidates that affect oligomerization.

2.4.5 Results from Complementary Biophysical Tools

Many available biophysical tools provide complementary results in tracking Aβ assembly. We are limiting our comparison to those that can follow the time course of aggregation under physiologically relevant condition. We are not considering the many approaches that view Aβ as its initial, largely low-MW oligomers (e.g., solution NMR) or in its final aggregated state (e.g., solid state NMR). Among the methods tracking aggregation, ThT dye fluorescence is regularly used to follow amyloid formation. The dye undergoes a large enhancement in fluorescence upon binding to amyloid fibrils, making it a particularly powerful and convenient tool. A typical ThT fluorescence curve reports a single sigmoid including a lag phase and a rapid growth phase to form fibrils. Because ThT primarily interacts with β-sheet structure, this method is less suitable to study the amorphous oligomers lacking well-defined β-strand structure. ThT does not give site-specific information on the oligomerization. Furthermore, because ThT bears structural similarities to many amyloid inhibitors, it is possible that ThT can influence the fibril structure and the formation kinetics.[18]

The recent development of TMR, as an alternative fluorescence dye, affords a more informative sigmoidal curve. In this experimental design, Aβ was pre-modified to carry an extra lysine residue that was covalently linked to the TMR molecule at the N-terminus of Aβ.[17] Unlike in the ThT assay, fluorescence quenching due to the proximity of the TMR molecules was detected as Aβ associates. This assay characterized the early-stage oligomer formation by showing an initial exponential drop in TMR fluorescence intensity before the emergence of the first lag phase.

This, like all the fluorescence methods, however, suffers from the need to modify the Aβ, a polypeptide whose aggregation is highly sensitive to its length and size (consider the difference between $A\beta_{1-42}$ and $A\beta_{1-40}$).

We found similar results with pulsed HDX as with TMR. In the pulsed HDX method, the hydrogen on Aβ backbone amide is in exchange with the deuterium in the solvent, and the level of deuterium-uptake within a certain short time window reflects changes in the conformation of the polypeptide backbone.[19] This approach, although yielding regional information, is not highly sensitive.

More recently, using high resolution AFM, Buratto, Bowers, and co-workers[15] found that $A\beta_{1-42}$ hexamer- and dodecamer-sized structure become dominant in as short as 5 min, and the density of large spherical aggregates termed preprotofibrils grows considerably by 20 min. This result brings new insight, but it may be confounded by surface effects because a subset of protein aggregates may be preferentially deposited on the surface for AFM and the interactions between the protein and the sample surface can affect aggregate morphology and formation kinetics.[43-44] Therefore, comparisons of solution measurements by fluorescence and MS footprinting with measurements from using sample deposition will require careful interpretation. The former approaches map co-populated species in a mixture, and report $A\beta_{1-42}$ conformation quantitatively as an assemble average of structural protection in bulk solution without pointing to individual oligomers; the latter provides a visualization of the oligomer morphology at certain oligomeric orders (e.g. hexamer and dodecamer). Note, as we discussed above, the initial burst phase of the early oligomerization is represented by an exponential decrease in fluorescence intensity in the TMR assay and a rapid increase in the structural protection in pulsed HDX experiment. This phase occurs over a time scale of 30-60 min, with our outcome from the FPOP experiment

falling in the same range. Considering the vast complexity and high sensitivity of $A\beta_{1-42}$ oligomerization towards external conditions, experimental phenomena observed in various approaches are not necessarily proportionate in terms of the time scale, and it is sensible for us to the set our sights on the general trends.

More importantly, caution is needed to correlate the "phases" observed using various available tools. The molecular-level mechanisms of the probes used in each method discussed above, including FPOP, differ significantly. The aggregation of Aβ, especially of $A\beta_{1-42}$ under study, is a process highly sensitive to buffer conditions, concentration, temperature, pH, agitation etc. Therefore, the observed phases uncovered by these methods can be different. For example, using an equivalent $A\beta_{1-42}$ system, the growth phase indicated by ThT fluorescence comes later than that observed in the TMR assay.[17] In addition, we do not expect the fluorescence and HDX data to fit into Scheme 2.1 because we invoked the double-catalytic reactions to address the five distinct stages revealed uniquely by FPOP. In the FPOP approach, the footprinting reagent, here a hydroxyl radical, is comparably sized to water molecules and directly modifies the solvent-accessible amino-acid side chains of $A\beta_{1-42}$. The probe is sensitive to subtle conformational changes involving the transient states, especially those in which $A\beta_{1-42}$ side chains are involved (e.g., several protofibrillar filaments twist to form bundles of mature fibrils, in which case the polypeptide backbone stays relatively unchanged.[10-11]).

2.5 Conclusions

FPOP footprinting provides new, in-depth insights into $A\beta_{1-42}$ conformational changes and its aggregation by affording structural resolution even down to the amino-acid residue level. It not

only defines the early oligomerization stages but also reports subtle conformational changes that occur after early beta-strand formation, allowing the intermediate transitions to be effectively resolved to reveal the multi-step nature of oligomerization. We chose to model the aggregation curve by two autocatalytic reactions based on a modified Finke-Watzky mechanism, being attracted by the simplicity of this model. The successful fit of the observed aggregation kinetics to two-sigmoid model provides more insight on the molecular mechanisms involved in $A\beta_{1-42}$ self-assembly.

The FPOP approach overcomes most of the weaknesses of other methodologies used to study the time-dependent amyloid formation by affording a direct, real-time, fast, and accurate measurement of the solvent accessibility of Aβ. Moreover, applications of the platform can be extended to address the effects of pH, concentration, agitation, and Aβ modification on aggregation. We expect this measurement strategy to be utilized further to guide the design of optimal compounds that inhibit Aβ aggregation and/or toxicity, and to be extended to the evaluation of other amyloidogenic proteins.

2.6 References

1. Klein, W. L.; Stine, W. B., Jr.; Teplow, D. B., *Neurobiol. Aging* **2004,** *25* (5), 569-80.

2. Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O., *Nat. Struct. Mol. Biol.* **2010,** *17* (5), 561-7.

3. Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J. E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T., *Nat Chem* **2009,** *1* (4), 326-31.

4. Carulla, N.; Zhou, M.; Arimon, M.; Gairí, M.; Giralt, E.; Robinson, C. V.; Dobson, C. M., *Proceedings of the National Academy of Sciences* **2009,** *106* (19), 7828-7833.

5. Hashimoto, T.; Adams, K. W.; Fan, Z.; McLean, P. J.; Hyman, B. T., *J. Biol. Chem.* **2011,** *286* (31), 27081-91.

6. Nasica-Labouze, J.; Nguyen, P. H.; Sterpone, F.; Berthoumieu, O.; Buchete, N. V.; Cote, S.; De Simone, A.; Doig, A. J.; Faller, P.; Garcia, A.; Laio, A.; Li, M. S.; Melchionna, S.; Mousseau, N.; Mu, Y.; Paravastu, A.; Pasquali, S.; Rosenman, D. J.; Strodel, B.; Tarus, B.; Viles, J. H.; Zhang, T.; Wang, C.; Derreumaux, P., *Chem. Rev.* **2015,** *115* (9), 3518-63.

7. Parbhu, A.; Lin, H.; Thimm, J.; Lal, R., *Peptides* **2002,** *23* (7), 1265-70.

8. Bertini, I.; Gonnelli, L.; Luchinat, C.; Mao, J.; Nesi, A., *J. Am. Chem. Soc.* **2011,** *133* (40), 16013-16022.

9. Xiao, Y.; Ma, B.; McElheny, D.; Parthasarathy, S.; Long, F.; Hoshi, M.; Nussinov, R.; Ishii, Y., *Nat. Struct. Mol. Biol.* **2015,** *22* (6), 499-505.

10. Luhrs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Dobeli, H.; Schubert, D.; Riek, R., *Proc Natl Acad Sci U S A* **2005,** *102* (48), 17342-7.

11. Lu, J. X.; Qiang, W.; Yau, W. M.; Schwieters, C. D.; Meredith, S. C.; Tycko, R., *Cell* **2013,** *154* (6), 1257-68.

12. Selkoe, D. J., *Nat. Cell Biol.* **2004,** *6* (11), 1054-61.

13. Schmidt, M.; Sachse, C.; Richter, W.; Xu, C.; Fändrich, M.; Grigorieff, N., *Proceedings of the National Academy of Sciences* **2009,** *106* (47), 19813-19818.

14. Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B., *Proc Natl Acad Sci U S A* **2003,** *100* (1), 330-5.

15. Economou, N. J.; Giammona, M. J.; Do, T. D.; Zheng, X.; Teplow, D. B.; Buratto, S. K.; Bowers, M. T., *J. Am. Chem. Soc.* **2016,** *138* (6), 1772-1775.

16. Brorsson, A. C.; Kumita, J. R.; MacLeod, I.; Bolognesi, B.; Speretta, E.; Luheshi, L. M.; Knowles, T. P.; Dobson, C. M.; Crowther, D. C., *Front Biosci (Landmark Ed)* **2010,** *15*, 373-96.

17. Garai, K.; Frieden, C., *Proc Natl Acad Sci U S A* **2013,** *110* (9), 3321-6.

18. Hudson, S. A.; Ecroyd, H.; Kee, T. W.; Carver, J. A., *FEBS J.* **2009,** *276* (20), 5960-5972.

19. Zhang, Y.; Rempel, D. L.; Zhang, J.; Sharma, A. K.; Mirica, L. M.; Gross, M. L., *Proc Natl Acad Sci U S A* **2013,** *110* (36), 14604-9.

20. Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R., *Expert review of proteomics* **2011,** *8* (1), 43-59.

21. Klinger, A. L.; Kiselar, J.; Ilchenko, S.; Komatsu, H.; Chance, M. R.; Axelsen, P. H., *Biochemistry* **2014,** *53* (49), 7724-7734.

22. Fändrich, M.; Meinhardt, J.; Grigorieff, N., *Prion* **2009,** *3* (2), 89-93.

23. Xu, G.; Chance, M. R., *Chem. Rev.* **2007,** *107* (8), 3514-3543.

24. Hambly, D. M.; Gross, M. L., *J. Am. Soc. Mass Spectrom.* **2005,** *16* (12), 2057-63.

25. Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L., *Anal. Chem.* **2009,** *81* (16), 6563-6571.

26. Jones, L. M.; B. Sperry, J.; A. Carroll, J.; Gross, M. L., *Anal. Chem.* **2011,** *83* (20), 7657-7661.

27. Kheterpal, I.; Lashuel, H. A.; Hartley, D. M.; Walz, T.; Lansbury, P. T.; Wetzel, R., *Biochemistry* **2003,** *42* (48), 14092-14098.

28. Xie, B.; Sharp, J. S., *Analytical Chemistry* **2015,** *87* (21), 10719-10723.

29. Li, Z.; Moniz, H.; Wang, S.; Ramiah, A.; Zhang, F.; Moremen, K. W.; Linhardt, R. J.; Sharp, J. S., *The Journal of Biological Chemistry* **2015,** *290* (17), 10729-10740.

30. Zhang, J.; Ramachandran, P.; Kumar, R.; Gross, M. L., *Journal of the American Society for Mass Spectrometry* **2013,** *24* (3), 450-453.

31. Slysz, G. W.; Percy, A. J.; Schriemer, D. C., *Analytical Chemistry* **2008,** *80* (18), 7004-7011.

32. Morris, A. M.; Watzky, M. A.; Agar, J. N.; Finke, R. G., *Biochemistry* **2008,** *47* (8), 2413-27.

33. Pallitto, M. M.; Murphy, R. M., *Biophys. J.* **2001,** *81* (3), 1805-1822.

34. Wolff, M.; Unuchek, D.; Zhang, B.; Gordeliy, V.; Willbold, D.; Nagel-Steger, L., *PLoS ONE* **2015,** *10* (5), e0127865.

35. Cruz, L.; Rao, J. S.; Teplow, D. B.; Urbanc, B., *The Journal of Physical Chemistry B* **2012,** *116* (22), 6311-6325.

36. Besson, C.; Finney, E. E.; Finke, R. G., *Chemistry of Materials* **2005,** *17* (20), 4925-4938.

37. Stravalaci, M.; Beeg, M.; Salmona, M.; Gobbi, M., *Biosens. Bioelectron.* **2011,** *26* (5), 2772-2775.

38. Crespo, R.; Rocha, F. A.; Damas, A. M.; Martins, P. M., *Journal of Biological Chemistry* **2012,** *287* (36), 30585-30594.

39. Eden, K.; Morris, R.; Gillam, J.; MacPhee, C. E.; Allen, R. J., *Biophys. J.* **2015,** *108* (3), 632-643.

40. Taouatas, N.; Heck, A. J. R.; Mohammed, S., *Journal of Proteome Research* **2010,** *9* (8), 4282-4288.

41. Urbanc, B.; Betnel, M.; Cruz, L.; Bitan, G.; Teplow, D. B., *J. Am. Chem. Soc.* **2010,** *132* (12), 4266-4280.

42. Klinger, A. L.; Kiselar, J.; Ilchenko, S.; Komatsu, H.; Chance, M. R.; Axelsen, P. H., *Biochemistry* **2014,** *53* (49), 7724-34.

43. Vácha, R.; Linse, S.; Lund, M., *J. Am. Chem. Soc.* **2014,** *136* (33), 11776-11782.

44. Campioni, S.; Carret, G.; Jordens, S.; Nicoud, L.; Mezzenga, R.; Riek, R., *J. Am. Chem. Soc.* **2014,** *136* (7), 2866-2875.

Chapter 3: Fast Photochemical Oxidation of Proteins Characterizes the Interaction of a Small Molecule Inhibitor on Amyloid Beta 1-42 Aggregation

3.1 Abstract

Significant evidence has linked aggregation of amyloid beta (Aβ) to the pathogenesis of Alzheimer's disease, motivating the development for small molecules that prevent aggregation. Study of anti-amyloid compounds suggest that many small molecules may re-direct the aggregation cascade rather than completely inhibiting it. Here, we describe an application of Fast Photochemical Oxidation of Proteins (FPOP) to evaluate the effect of a polyphenolic compound derived from curcumin in remodeling the aggregation pathway of $A\beta_{1-42}$. We find that the small molecule inhibits aggregation by maintaining $A\beta_{1-42}$ in LMW oligomers, resulting in an extended lag phase prior to the rapid formation of fibrils. The compound presumably interacts with the N-terminus of $A\beta_{1-42}$, "caps" oligomers, and slows down stacking of oligomers, as evidenced by the reduced solvent accessibility of the N-terminal region in the remodeled $A\beta_{1-42}$. The compound also destabilizes the middle region, which has been hypothesized to serve as the nucleation interface, whereas the C-terminal region is minimally affected. FPOP effectively reveals the effect of the compound on aggregation by accurately reporting the solvent accessibility changes of $A\beta_{1-42}$ with peptide and even some amino-acid readouts, providing insights into the pathway of $A\beta_{1-42}$ aggregation and the mechanism of inhibition of anti-amyloids. Our approach overcomes major limitations of the widely-used Thioflavin T fluorescence assay for monitoring amyloid formation, including low spatial resolution and biased readout due to presence of exogenous inhibitors with intrinsic fluorescence (e.g., polyphenols).

3.2 Introduction

Accumulation and aggregation of Aβ peptides in the brain are critical pathogenic events in Alzheimer's disease. The exact pathway of Aβ aggregation, however, is not fully understood. A general view of Aβ aggregation involves formation of dimers and soluble oligomers followed by growth into protofibrils and fibrils via a complex multistep-nucleated polymerization.[1-3] From previous studies in the aggregation mechanism of Aβ as well as from the insights described in Chapter 2, this process involves the following features: (i) nucleation dominated by a fast, short initial phase to form soluble, low-molecular-weight (LMW) oligomers, (ii) assembly of the LMW oligomers into larger oligomeric states with defined beta-sheet structure, (iii) association and elongation of the β-sheet-rich oligomers into higher-order protofibrillar aggregates, and (iv) reorganization of the higher-order intermediate aggregates to form mature fibrils.

Initially, fibrils were considered to be the molecular culprit in AD. The accumulation of amyloid plaque, however, does not correlate well with AD pathogenesis.[4] More recent evidence has supported that the soluble Aβ oligomers or prefibrillar aggregates are the causes of synaptic loss and cognitive impairment.[5-6] Although precise relationships between the oligomeric states of Aβ aggregates and the disease remain to be established, the connection between Aβ aggregation, cellular dysfunction, and AD suggests that anti-AD drugs targeting soluble Aβ oligomers could hold considerable promise and ultimately lead to therapeutics that prevent and/or treat AD.

Over the past decade, significant research effort was made toward discovering and developing of compounds that inhibit aggregation of amyloid proteins. These compounds, known as anti-amyloids, show efficacy in preventing amyloid formation, disaggregating pre-formed fibrils and reducing amyloid-induced cytotoxicity.[7-8] Natural polyphenols, a class of compounds abundant

in wine, tea and spices have captured considerable attention owing to their effective suppression of amyloid formation.[9] Inhibitory effects of several natural polyphenols, such as (-)-epigallocatechin-3-gallate (EGCG), resveratrol, and curcumin (Figure 3.1) have been extensively studied.[10-13] Polyphenol inhibitors exhibit different interaction patterns and inhibition mechanisms, but presumably remodel aggregation by interacting with the β-sheet, maintaining the protein in a non-aggregated, soluble form or yielding nontoxic, off-pathway oligomers or high-molecular-weight (HMW) aggregates.

The heterogeneity of Aβ aggregates, however, represent important challenges in structural and functional characterization. In vitro aggregation assays require tools for detecting Aβ oligomers/fibrils and monitoring the dynamic process in real time. In Chapter 2, we described the development of the FPOP platform to monitor $A\beta_{1-42}$ self-association by probing the solvent accessibility change of the peptide. In this chapter, we describe an extension of the platform by evaluating the effect of a curcumin derivative on $A\beta_{1-42}$ aggregation. Curcumin (diferulomethane) (Figure 3.1), found in the spice turmeric, is a potent antioxidant possessing anti-inflammatory activities. It has been shown that curcumin directly binds to small Aβ species, prevents oligomerization, and protects against Aβ-induced neurotoxicity.[13-14]

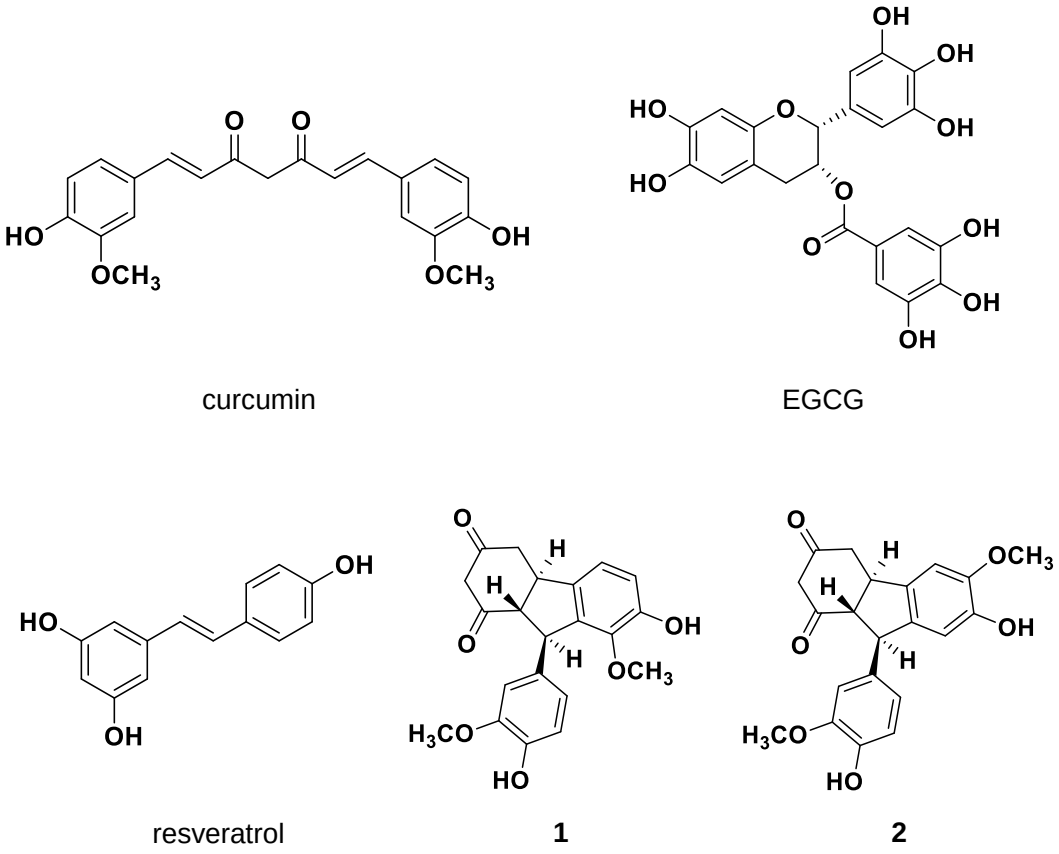


Figure 3.1 Structure of curcumin, EGCG, resveratrol, compound 1, and compound 2.

Curcumin, however, does not have good therapeutic perspective owing to its poor metabolic stability and limited blood–brain barrier penetration.[15] Here, we studied two novel curcumin derivatives, (4aR,9R,9aR)-7-hydroxy-9-(4-hydroxy-3-methoxyphenyl)-8-methoxy-4,4a,9,9a-tetrahydro-1H-fluorene-1,3(2H)-dione (compound1) and (4aR,9R,9aR)-7-hydroxy-9-(4-hydroxy-3-methoxyphenyl)-6-methoxy-4,4a,9,9a-tetrahydro-1H-fluorene-1,3(2H)-dione (compound 2) (Figure 3.1), on $A\beta_{1-42}$ aggregation. Compound 1 and compound 2 are structural isomers obtained from acid-catalyzed cyclization of curcumin. Our collaborator, Dr. George Mathai, hypothesized that treating curcumin in acid would mimic the condition when curcumin is ingested into the stomach. In our study, we first performed Thioflavin T (ThT) fluorescence

assays to identify preliminarily in vitro the most potent inhibitors among curcumin and its two derivatives. We then applied MS-based FPOP to investigate the effect of compound 1, the most potent inhibitor among the three, on $A\beta_{1-42}$ aggregation and reveal changes in the behavior of $A\beta_{1-42}$ sub-regions in the presence of compound 1. This study, together with that detailed in Chapter 2, demonstrates the capability of the FPOP platform as a sensitive tool to monitor amyloid formation and to evaluate crucial factors affecting the aggregation.

3.3 Materials and Methods

3.3.1 Materials

Synthetic human amyloid beta 1-42 ($A\beta_{1-42}$) was from AnaSpec (San Jose, CA). Curcumin, compound 1 and compound 2 were synthesized by our collaborator, Dr. George Mathai. Hexafluoroisopropanol (HFIP), sodium hydroxide (NaOH), ThT, *L*-glutamine, *L*-methionine, catalase, hydrogen peroxide (H_2O_2), phosphate buffer saline (PBS, 10 mM phosphate, 138 mM NaCl, 2.7 mM KCl), urea, formic acid and trifluoroacetic acid were from Sigma Aldrich (St. Louis, MO). Lys-N protease was from Thermo Scientific (Waltham, MA).

3.3.2 ThT fluorescence assay

To investigate the effect of the test compounds on $A\beta_{1-42}$ fibrilization, 5 μL HFIP-pretreated $A\beta_{1-42}$ dissolved 3 mM NaOH was sequentially mixed with 1 μL containing the test compounds in PBS (or 1 μL PBS only for the control) and 1 μL ThT (1.5 mM in water), followed by a 1:20 dilution with PBS buffer to initiate aggregation. The final concentration of $A\beta_{1-42}$ and ThT after dilution were 20 μM and 15 μM, respectively. For each compound, two molar ratios, including

1:1 and 1:5 of $A\beta_{1-42}$ to the small molecule were examined. ThT fluorescence was measured in duplicate using a SpectraMax M2e plate reader (Molecular Devices, San Jose, CA). Excitation and emission wavelengths were 440 and 485 nm, respectively.

3.3.3 Sample preparation for FPOP

HFIP-pretreated $A\beta_{1-42}$ was dissolved in 3 mM NaOH and incubated without stirring at room temperature for 3 min followed by sonication for 1 min. Aggregation was initiated upon diluting $A\beta_{1-42}$ by 10 fold (v/v) with compound 1 in PBS (pH 7.4). The final concentrations of $A\beta_{1-42}$ and compound 1 after dilution were 20 μM and 100 μM, respectively. The solution mixture was then incubated at 25 °C without stirring. For each time point, the incubation and analysis were done in duplicate.

3.3.4 FPOP

At the various incubation times, $A\beta_{1-42}$ was immediately submitted to FPOP labeling. FPOP was performed as previously described,[16] except no scavenger was added. This is because compound 1 is a hydroxyl radical-reactive polyphone and serves as the radical scavenger in FPOP. Briefly, a laser beam from a KrF excimer laser (GAM Laser Inc., Orlando, FL, USA) of wavelength 248 nM was used to initiate H_2O_2 photolysis to give hydroxyl radicals. The $A\beta_{1-42}$ in PBS was mixed with 20 mM H_2O_2 just prior to injection into the flow tube for irradiation with the laser. The flow rate was adjusted according to the width of the laser irradiation window to ensure 20% irradiation-excluded volume and to minimize repeated laser exposure. The oxidatively modified sample was collected in a tube containing 10 mM catalase and 20 mM methionine to decompose leftover H_2O_2 and prevent oxidation-artifacts during storage. Control samples of $A\beta_{1-42}$ in the

presence of compound 1 were handled in the same manner as the experimental ones, but not laser-irradiated.

3.3.5 LC-MS/MS

Digestion of the FPOP-modified $A\beta_{1-42}$ and the subsequent LC-MS/MS analysis were adopted from the previous established protocols described in 2.3.4 and 2.3.5 in Chapter 2.

3.3.6 Data analysis

Quantification of the FPOP modification was adopted from the previous established protocol described in 2.3.6 in Chapter 2.

3.4 Results and Discussion

3.4.1 Potency of compounds active for inhibition of $A\beta_{1-42}$ aggregation

ThT is a benzothiazole dye that undergoes enhanced fluorescence upon binding to amyloid fibrils. It is the most commonly used probe for monitoring fibrillization kinetics in vitro. Using fluorescence spectroscopy with ThT, we performed a global examination of the effects of curcumin, compound1 and compound 2 on the aggregation of 5 μM Aβ1-42 incubated at 25 °C, pH 7.4 (Figure 3.2). In our measurement, the fluorescence intensity at long times decreases owing to precipitation of mature fibrils. Despite this, all the ThT fluorescence curves of $A\beta_{1-42}$ incubated under our conditions follow a characteristic sigmoidal curve representing a lag phase and a growth phase in aggregation, consistent with the nucleation-dependent polymerization. In the presence of curcumin or compound 1, the ThT fluorescence at the final equilibrium level are lower than that in the ligand-free sample (Figure 3.2A and B), presumably due to the effective

inner quenching of the ThT fluorescence by curcumin and compound 1 upon binding and stacking in fibrils. This is, however, not observed in samples incubated with compound 2 (Figure 3.2C).

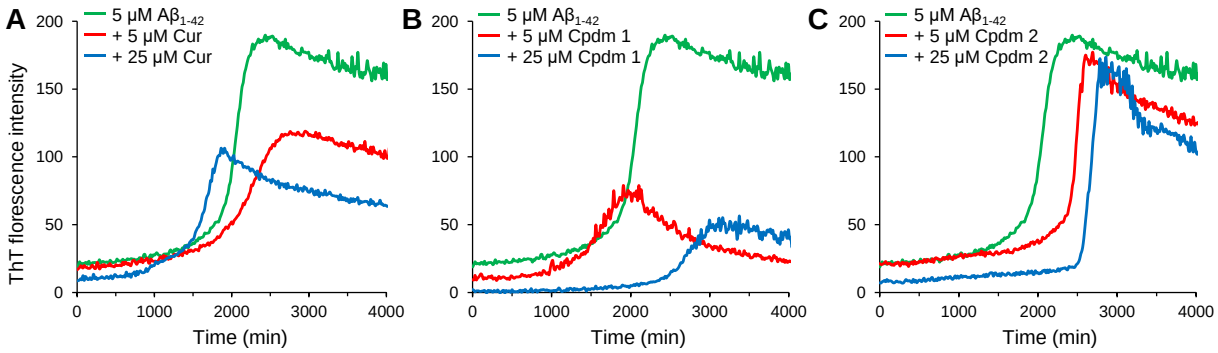


Figure 3.2. ThT fluorescence analysis of Aβ$_{1-42}$ (5 μM) at 25 °C, pH 7.4. Aβ$_{1-42}$ was incubated in the absence and presence of 5 μM or 25 μM (A) curcumin, (B) compound 1, and (C) compound 2. Data points are the mean of two independent measurements.

Despite the reduced fluorescence intensity at equilibrium owing to binding, curcumin did not significantly affect the length of the lag phase prior to fibrilization. This observation is consistent with the previous ThT study on the effect of curcumin on remodeling Aβ aggregation[13]. For the inhibitory effect of curcumin, controversies arise as some results shows curcumin inhibits both oligomers and fibrils formation[14], whereas others found that curcumin inhibits oligmerization but not fibrilization[20]. These variable results are likely to result from the different conditions (e.g., concentration of Aβ, agitation) used in the experiments.

In contrast, Aβ$_{1-42}$ incubated with a 25 μM solution of compound 1 at high stoichiometry (1:5 molar ratio of Aβ$_{1-42}$ to compound 1) displays a prominently extended lag phase compared to those in the ligand-free sample and the sample incubated with 5 μM compound 1 (1:1 molar

ratio) (Figure 3.2B). This suggests that compound 1 is effective in stabilizing soluble oligomers in a dose-dependent manner, resulting in a longer lag phase to achieve fibrilization, but eventually allowing conversion of $A\beta_{1-42}$ into fibrils. In addition, $A\beta_{1-42}$ incubated with 5 μM or 25 μM of compound 2 also undergoes a delay in fibrilization (Figure 3.2 C), but a high concentration of compound 2 at 25 μM does not result in an extension of the lag phase that is as pronounced as that for compound 1 does at the same concentration.

ThT fluorescence, however, is not highly suitable for mechanistic studies primarily owing to its inability to report on early steps of oligomerization that occur prior to fibril formation.[17] In addition, problems arise when relying on ThT fluorescence to assess the anti-amyloid activity of the polyphenol compounds. Polyphenols can be chromophoric or even intrinsically fluorescent because they have strong π-π* electronic transitions via conjugation of the phenolic aromatic rings upon binding to amyloids.[18] Actually, the spectroscopic properties of curcumin significantly interferes with the fibril-associated ThT fluorescence by absorbing both the excitation an emission lights of ThT.[19] We thus turned to FPOP for an accurate characterization of the inhibitory effect, taking advantage of the ability of FPOP to follow Aβ in-vitro without requiring binding of a fluorophore. The method provides a sensitive, clear readout on the changes in Aβ solvent accessibility in the presence or absence of a ligand or inhibitor.

3.4.2 Characterization of the effect of compound 1 on $A\beta_{1-42}$ aggregation by FPOP

Preliminary results from ThT fluorescence indicate an apparent delay in $A\beta_{1-42}$ fibrillation in the presence of compound 1 at 1:5 molar ratio of Aβ42: compound 1. We, thus, chose to characterize the effect of compound 1 at this ratio by FPOP. To allow formation of aggregates, 20 μM $A\beta_{1-42}$

was incubated with 100 μM compound 1 for various times. Upon completion of the incubation, $A\beta_{1-42}$ was immediately mixed with H_2O_2 and submitted to FPOP. The modified $A\beta_{1-42}$ was then digested using Lys-N, the aggregation was followed using the measured modification levels of the three peptide fragments as determined by LC-MS/MS (Figure 3.3 A-C).

The aggregation 20 μM $A\beta_{1-42}$ in the presence of compound 1 (Figure 3.3 A-C) shows significantly different characteristics than in the absence of compound 1 (data taken from our previous FPOP study for a concentration of 10 μM, Figure 3.3 D-F, adapted from Chapter 2). There are two variables, a higher starting concentration of $A\beta_{1-42}$ at 20 μM, which should accelerate the aggregation and shorten the time required for fibrillation. For ease of comparison, we "normalized" the plots to have the same width although the time scales are different in Figure 3.3.

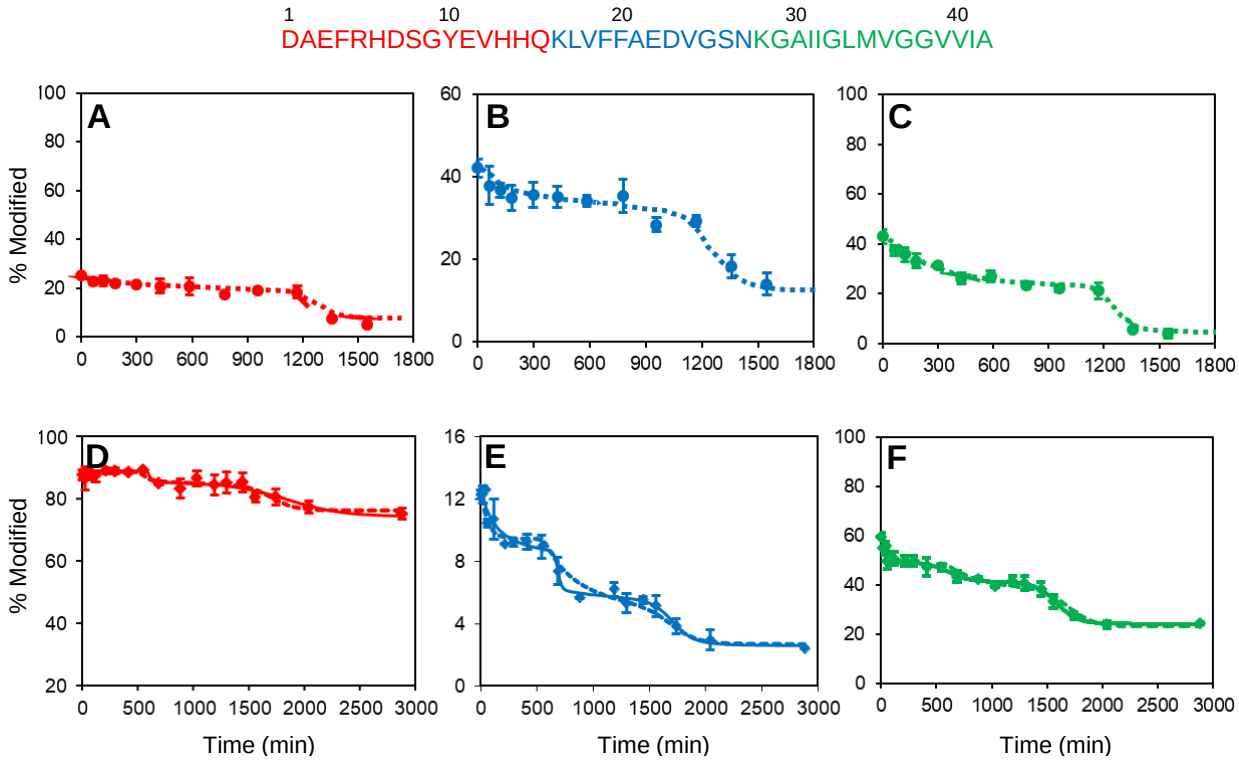


Figure 3.3. Changes in the FPOP modification of $A\beta_{1-42}$ sub-regions as a function of incubation time. Aggregation at 25 °C in the presence (A-C) and absence (D-F) of compound 1 in PBS solution (pH 7.4) without stirring. Three peptide fragments, 1-15, 16-27 and 28-42, obtained from Lys-N digestion of the full-length $A\beta_{1-42}$ are colored in red, blue, and green on the $A\beta_{1-42}$ sequence, respectively. (A-C) $A\beta_{1-42}$ incubated in the presence of compound 1, with initial concentrations of $A\beta_{1-42}$ and compound 1 at 20 μM (equivalent to monomeric $A\beta_{1-42}$) and 100 μM, respectively. (D-F) $A\beta_{1-42}$ incubated in the absence of compound 1 (adapted from Figure 2.5 in Chapter 2), with an initial concentration of $A\beta_{1-42}$ at 10 μM (equivalent to monomeric $A\beta_{1-42}$).

There are modest variations in the parameters of FPOP (e.g., laser energy, focusing lens) and between the two experiments conducted in the presence and absence of compound 1. We first considered to incorporate a reporter into the FPOP experiment to evaluate the dosage of hydroxyl

radicals in each sample. This would provide a mean to normalize the extents of modification between different samples. The method requires spiking in a fixed amount of a short peptide (e.g., Leu-enkephalin) into each protein sample prior to the laser irradiation (for an example of using a reporter peptide to assist FPOP quantification, referred to 5.4.2 in Chapter 5). Adding a reporter peptide into the sample may influence Aβ self-association. Additionally, the reporter peptide exhibits poor reproducibility in terms of its signal intensities in LC-MS and the modification levels, presumably because the reporter peptide interacts and possibly co-aggregates with $A\beta_{1-42}$ and its oligomers. The outcome indicates incorporation of a reporter is not feasible in study of an extremely hydrophobic peptide like $A\beta_{1-42}$.

Despite the small variations in FPOP parameters, we conclude that the modification levels of the C-terminal fragment (peptide 28-42) are nearly the same in the presence and absence of compound 1 (Figure 3.3 C and F). Specifically, the C-terminal region of the compound 1-remodeled $A\beta_{1-42}$ shows nearly the same extent of change in solvent accessibility as in the ligand-free sample during the monomer-to-fibril transition (both showing an absolute decrease of ~ 35% in modification extent from monomers to fibrils) (Figure 3.3 C and F). Those results indicate that the C-terminal region of $A\beta_{1-42}$ is unaffected or only affected to a modest extent by compound 1.

In contrast, the modification extent of the N-terminal region (peptide 1-15) is considerably reduced in the presence of compound 1. This dramatic decrease is likely due to 1) compound 1 binding to the N-terminal region, leading to protection from free-radical reactions, 2) compound 1 interaction of $A\beta_{1-42}$ at sites other than the N terminus, causing remote conformational changes that reduced its flexibility or solvent accessibility, leading to protection, or 3) a combination of these effects. Although the aggregation curves are remarkably displaced, their shapes as $A\beta_{1-42}$ transitions from monomers to fibrils are nearly identical in the presence and absence of

compound 1 (in both showing an absolute decrease of ~ 20% in modification). This phenomenon is more likely if compound directly binds to the N-terminal region of the monomer. This proposed mode of action is similar to that of its precursor curcumin as revealed by atomic force microscopy[21].

On the other hand, the middle domain (peptide 16-27) exhibits an opposite trend as the N-terminus. It shows a significant increase in FPOP modification in the compound 1-remodeled $A\beta_{1-42}$, indicating that binding of compound 1 results in increasing solvent exposure, flexibility and possibly decreasing oligomerization involving this central region.

Moreover, the FPOP aggregation curves indicate that the lag phase for nucleation of LMW oligomers is prominently extended in the presence of compound 1, which is observed for all three regions (Figure 3.3). After an initial decrease in modification during the first 60 min of incubation, the modification extent, a measure of the solvent accessibility of $A\beta_{1-42}$, becomes nearly constant, displacing a long plateau (60 ~ 1100 min) stage. Finally, a sharp decrease in modification takes place at long times, indicating a rapid transition of LMW oligomers to fibrils (Figure 3.3). During this process, there is little evidence for a clear phase of formation of HMW protofibrillar aggregates prior to fibrillation. Thus, we propose that compound 1 remodels $A\beta_{1-42}$ aggregation by stabilizing LMW oligomeric conformers.

We note that these results do not inform us on the exact conformation or biochemical properties of the compound-1-remodeled oligomers, or if those aggregates are on- or off-pathway. This means those oligomers could be similar to or different from the small oligomers formed in the absence of compound 1. In addition, FPOP indicates that under our condition and in the presence of compound 1, $A\beta_{1-42}$ eventually matures structurally as fibrils. This is consistent with the

outcome from the ThT fluorescence (Figure 3.2 B), and the relatively rapid oligomerization is due to the high concentration of $A\beta_{1-42}$ used in those in-vitro assays.

3.5 Conclusions

Compound 1 is capable to remodel $A\beta_{1-42}$ aggregation by maintaining $A\beta_{1-42}$ in LMW oligomers, presumably via interacting with the N-terminal region. Binding of compound 1 also results in deprotection and destabilization of the central domain, which was previously proposed to play an important role in $A\beta_{1-42}$ self-association by serving as a folding nucleus or interface for oligomerization.[22-23] The C-terminal region is minimally influenced by compound 1.

In terms of methodology development, we have extended the FPOP platform for monitoring amyloid formation (Chapter 2) to evaluate the effect of a polyphenolic inhibitor on $A\beta_{1-42}$ aggregation and reveal changes in the behavior of $A\beta_{1-42}$ sub-regions modulated by the inhibitor. This study demonstrates that FPOP is an effective and sensitive tool for probing the effect of potential therapeutic agents on Aβ aggregation in vitro. In addition, FPOP can be combined with other biophysical tools (e.g., atomic force microscopy, fluorescence spectroscopy) and biochemical assays (e.g., SDS-PAGE, immunoblotting) to obtain a comprehensive understanding of the properties of Aβ species and the mechanism of amyloid inhibition. For example, a previous study indicates that small, aromatic compounds lead to remodeling of Aβ soluble toxic oligomers into alternative conformers with reduced toxicity.[24] For our work, neurotoxicity of the intermediate LMW oligomers remodeled and stabilized by compound 1 can be further assessed by using the cytotoxicity assay. Last but not the least, we speculate that our FPOP approach would be applicable for studies of many amyloid inhibitors, not only small molecules, but also

peptides and antibodies against Aβ, providing valuable insights for the rational design of anti-amyloids therapeutics.

3.6 Acknowledgements

We thank Dr. George Mathai for providing curcumin, compound 1 and compound 2 for this study, and Dr. Carl Frieden and Dr. Liviu Mirica lab for their help with the ThT fluorescence assay.

3.7 References

1. Garai, K.; Frieden, C., *Proc Natl Acad Sci U S A* **2013,** *110* (9), 3321-6.

2. Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B., *Proc Natl Acad Sci U S A* **2003,** *100* (1), 330-5.

3. Urbanc, B.; Betnel, M.; Cruz, L.; Bitan, G.; Teplow, D. B., *J. Am. Chem. Soc.* **2010,** *132* (12), 4266-4280.

4. Terry, R. D., *J. Neuropathol. Exp. Neurol.* **1996,** *55* (10), 1023-1025.

5. Glabe, C. G., *J. Biol. Chem.* **2008,** *283* (44), 29639-29643.

6. McLean, C. A.; Cherny, R. A.; Fraser, F. W.; Fuller, S. J.; Smith, M. J.; Beyreuther, K.; Bush, A. I.; Masters, C. L., *Ann. Neurol.* **1999,** *46* (6), 860-6.

7. Yamin, G.; Ono, K.; Inayathullah, M.; Teplow, D. B., *Curr. Pharm. Des.* **2008,** *14* (30), 3231-3246.

8. van Dyck, C. H., *Biological Psychiatry* **2018,** *83* (4), 311-319.

9. Stefani, M.; Rigacci, S., *International Journal of Molecular Sciences* **2013,** *14* (6), 12411-12457.

10. Bieschke, J.; Russ, J.; Friedrich, R. P.; Ehrnhoefer, D. E.; Wobst, H.; Neugebauer, K.; Wanker, E. E., *Proceedings of the National Academy of Sciences* **2010,** *107* (17), 7710-7715.

11. Rezai-Zadeh, K.; Shytle, D.; Sun, N.; Mori, T.; Hou, H.; Jeanniton, D.; Ehrhart, J.; Townsend, K.; Zeng, J.; Morgan, D.; Hardy, J.; Town, T.; Tan, J., *J. Neurosci.* **2005,** *25* (38), 8807-14.

12. Marambaud, P.; Zhao, H.; Davies, P., *J. Biol. Chem.* **2005,** *280* (45), 37377-82.

13. Ono, K.; Hasegawa, K.; Naiki, H.; Yamada, M., *J. Neurosci. Res.* **2004,** *75* (6), 742-50.

14. Yang, F.; Lim, G. P.; Begum, A. N.; Ubeda, O. J.; Simmons, M. R.; Ambegaokar, S. S.; Chen, P. P.; Kayed, R.; Glabe, C. G.; Frautschy, S. A.; Cole, G. M., *J. Biol. Chem.* **2005,** *280* (7), 5892-901.

15. Nelson, K. M.; Dahlin, J. L.; Bisson, J.; Graham, J.; Pauli, G. F.; Walters, M. A., *J. Med. Chem.* **2017,** *60* (5), 1620-1637.

16. Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L., *Anal. Chem.* **2009,** *81* (16), 6563-6571.

17. Xue, C.; Lin, T. Y.; Chang, D.; Guo, Z., *Royal Society Open Science* **2017,** *4* (1), 160696.

18. LeVine, H., 3rd, *Methods Enzymol.* **1999,** *309*, 274-84.

19. Hudson, S. A.; Ecroyd, H.; Kee, T. W.; Carver, J. A., *FEBS J.* **2009,** *276* (20), 5960-72.

20. Necula, M.; Kayed, R.; Milton, S.; Glabe, C. G., *J. Biol. Chem.* **2007,** *282* (14), 10311-24.

21. Fu, Z.; Aucoin, D.; Ahmed, M.; Ziliox, M.; Van Nostrand, W. E.; Smith, S. O., *Biochemistry* **2014,** *53* (50), 7893-7903.

22. Roychaudhuri, R.; Yang, M.; Condron, M. M.; Teplow, D. B., *Biochemistry* **2012,** *51* (19), 3957-3959.

23. Cruz, L.; Rao, J. S.; Teplow, D. B.; Urbanc, B., *The Journal of Physical Chemistry B* **2012,** *116* (22), 6311-6325.

24. Ladiwala, A. R. A.; Dordick, J. S.; Tessier, P. M., *J. Biol. Chem.* **2011,** *286* (5), 3209-3218.

Chapter 4: Orthogonal Mass Spectrometry-Based Footprinting for Epitope Mapping and Structural Characterization: The IL-6 Receptor upon Binding of Protein Therapeutics*

* This chapter is based on the following publication: Li, K. S.; Chen, G.; Mo, J.; Huang, R. Y. C.; Deyanova, E. G.; Beno, B. R.; O'Neil, S. R.; Tymiak, A. A.; Gross, M. L., Orthogonal Mass Spectrometry-Based Footprinting for Epitope Mapping and Structural Characterization: The IL-6 Receptor upon Binding of Protein Therapeutics. *Anal. Chem.* 2017, *89* (14), 7742-7749.

4.1 Abstract

Higher order structure (HOS) is a crucial determinant for the biological functions and quality attributes of protein therapeutics. Mass spectrometry (MS)-based protein footprinting approaches play an important role in elucidating the relationship between protein biophysical properties and structure. Here, we describe the use of a combined method including hydrogen-deuterium exchange (HDX), fast photochemical oxidation of proteins (FPOP) and site-specific carboxyl group footprinting to investigate the HOS of protein and protein complexes. The work focuses on implementing complementary solution-phase footprinting approaches that differ in time scale, specificity for protein residue side chains vs. backbone as well as selectivity for different residue types to map integratively the epitope of human interleukin-6 receptor (IL-6R) for two adnectins with distinct affinities ($K_{d,\ Adnectin1}$ ~ 6.2 pM vs. $K_{d,\ Adenctin2}$ ~ 46 nM), and evaluate the resultant conformation/dynamic change of IL-6R. The suggested epitope, which is conserved for adenctin1 and adenctin2 binding, is a flexible loop that connects two β-strands in the cytokine-binding domain (DII) of IL-6R. We also found that adnectin1, the more strongly binding ligand, induces structural perturbations on two unstructured loops that are distally located beyond the epitope. Those changes are either attenuated or not detected for the case of adnectin2 binding. In addition to providing credibility in epitope determination, utilization of those combined approaches reveals the structural effects that can differentiate protein therapeutics with similar apparent biophysical properties.

4.2 Introduction

In Chapter 2-3, we focus on the development and application of hydroxyl radical footprinting, specifically FPOP to follow protein aggregation. In this chapter, and the following chapter 5, we implement FPOP with another two commonly used footprinting approaches, HDX and carboxyl group footprinting, to gain complementary insights into protein-ligand interaction and protein higher order structures.

Protein higher order structures describes the three-dimensional arrangement of a protein structure required for biological function. Monitoring protein higher order structures is critical for understanding the impact of molecular conformation on biotechnological applications in the protein-discovery pipeline.[1-2] Furthermore, maintaining protein higher order structures presents one of the key challenges for achieving robust and stable formulations of therapeutic proteins.[3] For the design of monoclonal antibodies (mAbs) and other therapeutic protein products, protein higher order structures is essential because binding of the therapeutic to the target is based on the specific recognition of the epitope on the protein. This is not only related to the primary sequence but also to conformation and post-translational modifications.[4]

Although atomic-level mapping of a protein or a protein complex can be achieved by high-resolution X-ray crystallography, the resulting static structure may have limited biological relevance and not reveal solution phase dynamics or long-range protein-protein interactions.[5-6] The complexity and low-throughput of X-ray crystallography restrict its application in the initial research stages where many potential therapeutic protein candidates may be of interest. Spectroscopy-based approaches including circular dichroism (CD),[7] infrared (FTIR)[8] and fluorescence spectroscopy[9] provide quick, global measurements of protein conformation, but the profile obtained from those methods often contains no local or regional structural information. In

contrast, protein footprinting, an evolving bioanalytical tool in structural biology, can reveal coarse-grained structural information relevant to proteins and their complexes. High sensitivity and fast data acquisition recommend MS-based footprinting for characterization of protein structure and macromolecular interactions at regional and even residue-specific levels of detail.[10-12]

Here, we describe a combination of MS-based protein footprinting methods, including hydrogen-deuterium exchange (HDX), fast photochemical oxidation of proteins (FPOP) and carboxyl group footprinting for mapping the extracellular region of human interleukin-6 receptor α-chain (referred as IL-6R hereafter) interacting with adnectins. Interleukin-6 (IL-6) plays critical roles in the pathogenesis of multiple myeloma, autoimmune diseases, and prostate cancer, appearing in abundant IL-6/IL-6R complexes.[13] Inhibition of the IL-6/IL-6R complex is a primary goal to antagonize the action of IL-6 in vivo.[14] The interacting partners of IL-6R selected for this study, adnectins, belong to a class of therapeutic proteins designed based upon the 10^{th} human fibronectin type III domain.[15] The two adnectins (adnectin1 and adnectin2) bind to IL-6R with picomolar and nanomolar affinity, respectively. X-ray structures of the IL-6R/adnectin complexes are not available, however, further motivating protein footprinting.

Among the methods to footprint IL-6R in this work, HDX is already well-established for protein HOS characterization.[16-17] HDX occurs via formation of covalent bonds in a reversible manner.[18] Its sensitivity to structural change is high provided there is little back exchange due to the labile nature of the N-D bond. HDX may be insensitive to subtle differences in conformation or dynamics when the exchange at the local region is low or rapid with respect to the HDX timescale.[19]

As an alternative to HDX, footprinting by incorporating irreversible modifications is emerging because it provides site-specific information by targeting amino acid side chains. Unlike HDX, irreversible labeling can survive extensive sample treatment and digestion. Footprinting by the hydroxyl radical, a common approach, involves irreversible oxidation of surface-accessible amino acid side chains as the primary product formation pathway. The radical probe has high reactivity with many residues, particularly those with sulfur-containing, aromatic, and aliphatic side chains.[20] FPOP, the method used to generate hydroxyl radicals in this work, utilizes pulses of 248 nm KrF laser radiation to induce photolysis of hydrogen peroxide.[21] As discussed in Chapter 1, the lifetime of labeling with primary hydroxyl radicals can be controlled within microseconds by introducing a radical scavenger.[22] To date, FPOP has been implemented for protein higher order structures characterization of several therapeutic targets[23-26] (referred to 1.4.3 in Chapter 1 for more details), showing its suitability for proteins of interest in drug discovery.

As a complement to free-radical footprinting, a variety of chemical reagents that target amino acid residues in a site-specific manner (e.g., *N*-ethylmaleimide[27] and diethylpyrocarbonate[28-29]) can also provide information on site-specific solvent accessibility but react with protein substrates more slowly than do free radicals.[30] In this work, we performed carboxyl group footprinting with glycine ethyl ester (GEE) to corroborate the findings from HDX and FPOP, taking advantage of the presence of the many Asp/Glu residues in the flexible loops of IL-6R. The chemical modification occurs for solvent-accessible Asp/Glu side chains (and the C-terminus) as a result of activation by 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) through formation of an *O*-acylisourea intermediate, which is subsequently displaced by the amine group of a GEE molecule via nucleophilic attack.[31] The labeling product, which is stable

and remains intact in post-sample handling and digestion processes, can be quantified to determine the solvent accessibility at various carboxylic acid sites.[32-33]

In practice, application of protein footprinting to structural characterization projects including epitope mapping requires careful consideration of time and resource allocation because, at this stage, experiments still require considerable instrument and interpretation time. Questions arise whether time is better spent doing many replicates or time points with one approach or instead employing other complementary approaches each in a less rigorous manner. Although an exhaustive evaluation by one approach will often provide an answer, we have chosen an integrative course, and the results presented here demonstrate the value of applying that approach for epitope mapping and HOS characterization.

4.3 Materials and Methods

Recombinant human IL-6R alpha extracellular region (residue 20-358, referred as IL-6R below) was purchased from R&D systems (Minneapolis, MN). Adnectin1 (K_d ~ 6.2 pM) and adnectin2 (K_d ~ 46 nM) were expressed and purified at BMS as previously described.[34] All surface plasmon resonance (SPR) experiments for binding affinity measurement were performed using a Biacore T100 instrument (GE Healthcare) (details of SPR can be found in SI Materials and Methods). To form the IL-6R/adnectin complex, bound state IL-6R was prepared by mixing 50 μM IL-6R with adnectin1 or adnectin2 at a 1:1 molar ratio and incubated at room temperature for 1 h.

4.3.1 H/D exchange

HDX was performed by following a standardized protocol (see SI Materials and Methods). Briefly, a mapping experiment of IL-6R peptic peptides was performed under non-denaturing condition, and the common peptides identified were further monitored for their deuterium uptake levels with a Synapt G2 High Definition mass spectrometer (Waters). HDX was initiated by mixing the labeling buffer (10 mM phosphate buffer in D_2O, pD 6.99) with the protein solution. The labeling reaction was allowed for different periods of times: 20 s, 1 min and 10 min.

4.3.2 FPOP

Prior to injection into the FPOP tubing, the protein sample in PBS was mixed with 20 mM H_2O_2 and 500 μM histidine. The final concentration of IL-6R for FPOP labeling was 10 μM. No dosimeter or reporter peptide was used (see SI for explanation). To avoid repeated laser exposure, the flow rate was adjusted to give ~20% irradiation-excluded volume. The laser beam was from a KrF excimer laser (GAM Laser Inc.), providing an excitation wavelength of 248 nM to initiate H_2O_2 photolysis into hydroxyl radicals. After laser irradiation, the sample solution was collected in a tube containing 10 mM catalase and 20 mM Met to remove leftover H_2O_2 and prevent post-labeling oxidation artifacts. Control samples of IL-6R with all the reagents added (including H_2O_2) were handled in the same manner, but not laser-irradiated. Samples of each state were subjected to FPOP in triplicate.

4.3.3 Carboxyl Group Footprinting

For carboxyl group labeling, glycine ethyl ester (GEE), 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) stock solutions were prepared freshly in PBS buffer. GEE was added to each pre-equilibrated sample to a concentration of 200 mM, followed by adding EDC to a concentration of 50 mM. The final concentration of IL-6R was 10 μM. In this reaction regime,

the by-products of the reaction (e.g., Lys-Asp/Glu cross-links) were essentially eliminated because the excessive amount of GEE dominates the cross-linking reactions.[35-36] Time-dependent labeling was carried out at room temperature and quenched at 1 min, 3 min and 10 min by adding an equal volume of 1 M ammonium acetate. Samples were further desalted using a Zeba column (Thermo Scientific, Rockford, IL).

4.3.4 Mass Spectrometry

FPOP or GEE-labeled protein was deglycosylated using PNGas F (New England Biolabs, Ipswich, MA) and digested using Trypsin/Lys-C or chymotrypsin (Promega, Madison, WI). In LC-MS/MS analysis, peptide fragments were separated on a custom-packed C18 column (CSH, 75 μm × 15 cm, 3.5 μm, 130 Å) using a Nano UltiMate 3000 Rapid Separation system (Dionex Co.) and analyzed with a Q Exactive Plus Orbitrap mass spectrometer (Thermo Scientific). The relative FPOP or GEE modification fraction was calculated by dividing the intensity of modified peptide/residue (I_{ox}) by the summed intensity of modified and unmodified peptides (I_u) (i.e., fraction of modified = $I_{ox}/(I_{ox} + I_u)$). A detailed description of MS methodology and data analysis of FPOP and GEE footprinting is given in SI Materials and Methods.

4.4 Results and Discussion

4.4.1 HDX mapping

HDX is a widely used method for exploring protein conformation and monitoring protein-ligand interactions based on mapping the hydrogen bonding of protein backbone amide; a number of examples of using HDX MS for epitope mapping were reported.[37-40] We first applied HDX to probe the structure of IL-6R and its complexes with the adnectins and found modest changes in the region 130-141 (Figure 3.1A). Although there is usually a more significant change in

protection (reduction in HDX rates) of the amide backbone with epitope binding[37-39], the results clearly suggest an epitope at this site. At this point, we considered repeating the HDX study and extending it over longer times with improved sequence coverage (78% residues mapped by peptic peptides (Figure 3.2)) by using other acid-insensitive proteolysis enzymes.[41] We reasoned, however, that more time points or higher coverage would not alter our conclusion about the 130-141 region or address the distinctive affinities of adnectin 1 and adnectin 2. In fact, HDX kinetic curves often show convergence at longer times owing to fast off rates, suggesting that extending the time for HDX is not productive.[24, 37, 39] Furthermore, all other detected regions follow identical HDX kinetics for the apo and holo forms (see Figure 3.1B as an example). We were thus motived to seek orthogonal footprinting methods to provide corroborating evidence for the epitope determination as well as to uncover conformational effects that may impact the binding of the two adnectins.

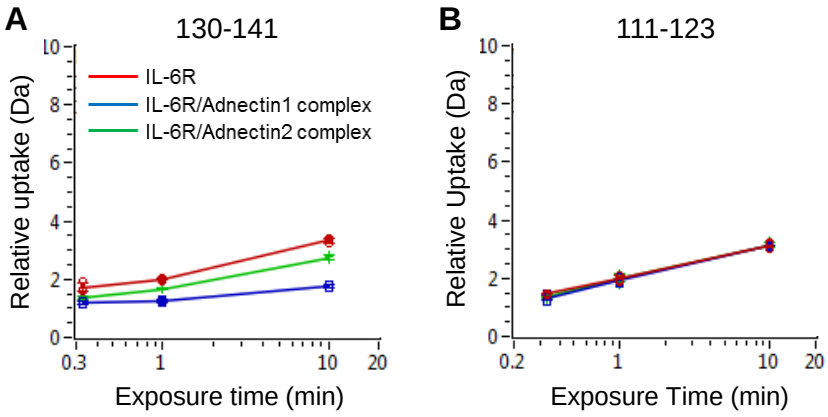


Figure 4.1. Representative HDX kinetics for IL-6R. (A) Region 130-141 shows reduced rates or extents of exchange in the holo (bound) states. (B) Region 111-123 shows no difference providing an example of a region that is not involved in or affected by adnectin binding.

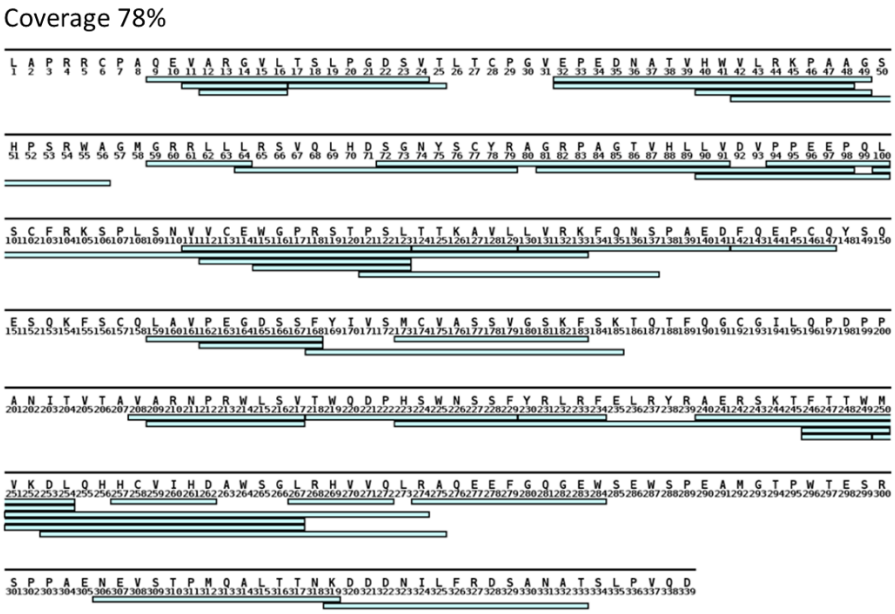


Figure 4.2. Sequence coverage of IL-6R in HDX.

4.4.2 FPOP mapping

FPOP is capable of reporting on protein transient dynamics, including fast folding[42] and alteration in side-chain flexibility[43]. We also showed that FPOP reveals fast fluctuations occurring remotely upon ligand binding, which is undetectable by slower footprinting methods[24]. Thus, we applied FPOP as a probe with high sensitivity to monitor changes in structure and dynamics of IL-6R. To obtain structural resolution spanning the IL-6R sequence, we chose two separate proteolysis experiments with Lys-C/trypsin and chymotrypsin. In the LC-MS/MS analysis, peptides and their modifications were identified by relying on their accurate (< 5 ppm) mass and the product-ion spectra (see Figure 4.3 for an example). In the bottom-up strategy, all tryptic peptides, including those with one missed cleavage site, provide > 90% sequence coverage of IL-6R. By contrast, chymotryptic cleavage is less specific. We found that chymotryptic digestion of IL-6R provides a better overall regional resolution by yielding peptides with shorter average length, the signal intensities for some regions are dispersed among

a greater number of overlapping chymotryptic fragments. With 80% sequence coverage from chymotrypsin digestion, we observed loss in the signal intensities for some chymotryptic peptides. This phenomenon is particularly pronounced for modified peptides of low abundance. Nevertheless, performing two sets of digestion experiments afforded a combined coverage of 96% of IL-6R sequence (Figure 4.4), permitting a detailed investigation into the local structure of IL-6R. In the data analysis, we only selected representative peptides that were relatively short and had desirable signal-to-noise ratios for accurate, label-free quantification (Table 4.1).

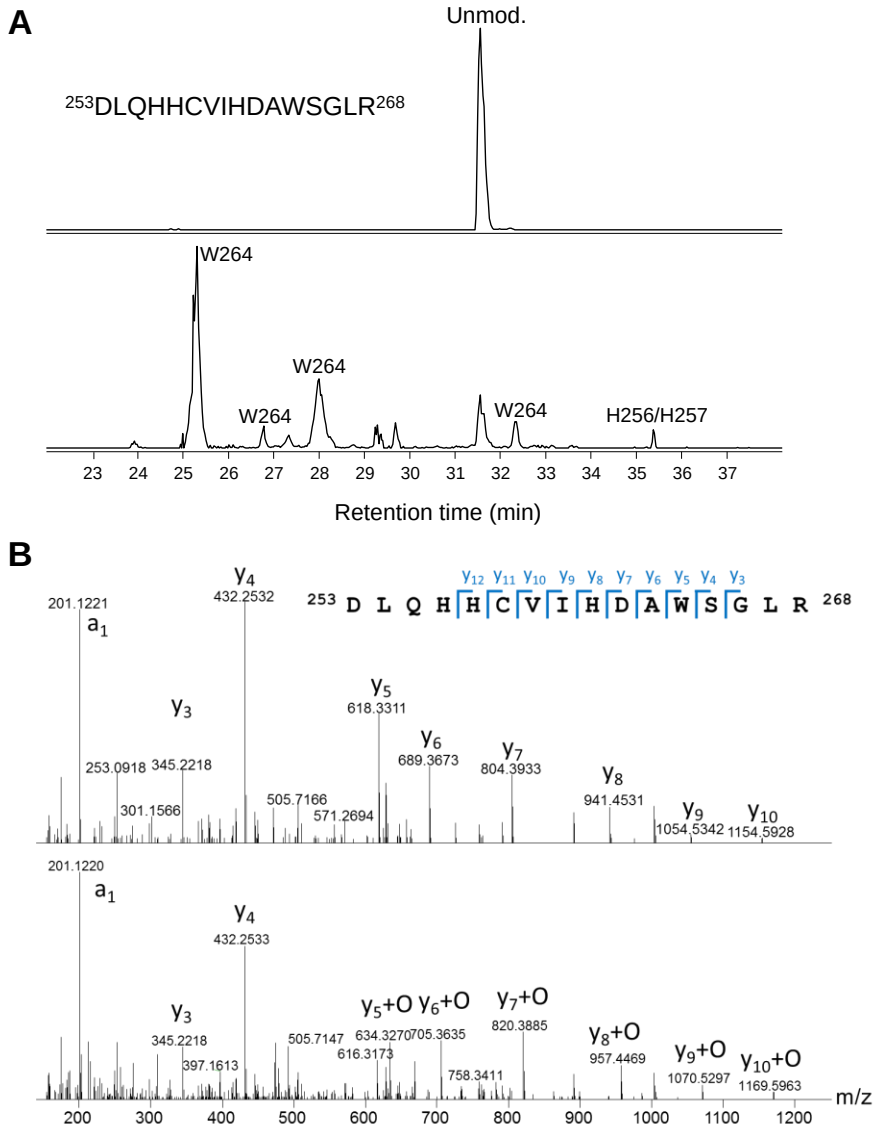


Figure 4.3. Representative LC-MS/MS identification of the peptide and its FPOP modification. (A) Extracted ion chromatogram of the unmodified peptide 253-268 (top, m/z = 486.7378) and its mono-oxidized (+15.9949 Da) form (bottom). Labels on the chromatogram indicate signal of site-specific oxidation determined by the product ions. (B) Product-ion spectra of the unmodified peptide 253-268 (top) and its modified form with oxidation identified on Trp^{264} (bottom).

FPOP clearly shows differential modification of peptides from regions in IL-6R (Figure 4.5A and B). Some regions (e.g., 61-65, 119-126, 232-237) are "FPOP-silent"; that is, we could detect no FPOP modification even though we saw signals from the unmodified peptides. We found small modification extents for other regions that are either shielded in the inner core of IL-6R (PDB: 1N26) and not solvent-accessible, or mainly composed of residues less reactive to hydroxyl radicals (e.g. Glu, Ser, Lys and Thr). By contrast, we observed high levels of modifications (modified fraction > 50%) for the C-terminus (288-296 and 301-319), because this region not only contains highly reactive Met^{292}, Trp^{296} and Met^{312} residues but also diversified loops and random coils that are inherently flexible and expected to be reactive with short-lived radicals.

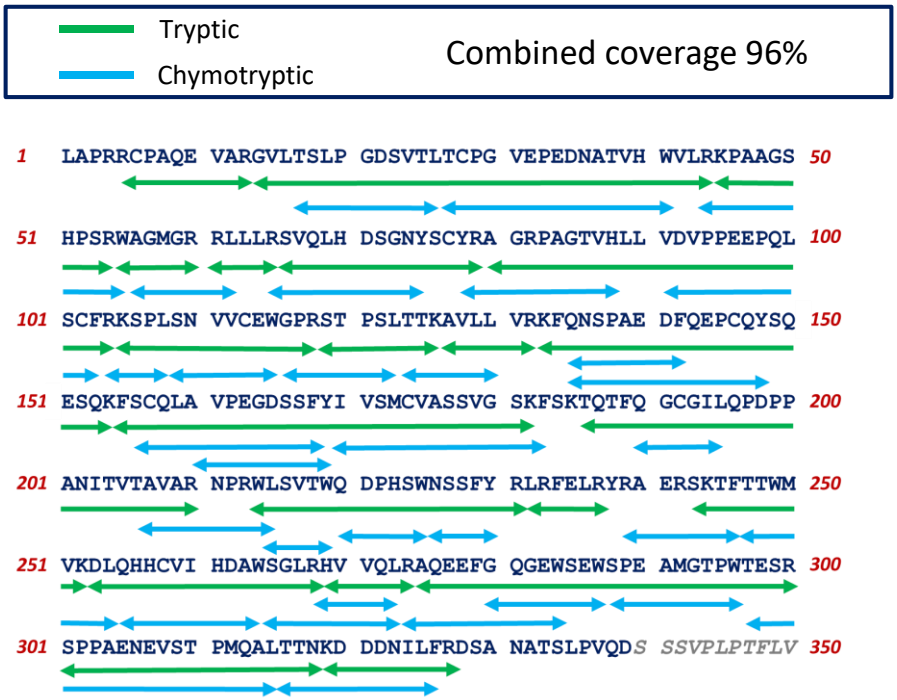


Figure 4.4. Sequence coverage of IL-6R in FPOP. Tryptic peptides are shown in green and chymotryptic peptides are shown in blue.

Peptide	Sequence
5-13	RCPAQEVAR
17-26	TSLPGDSVTL
27-41	TCPGVEPEDNATVHW
45-54	KPAAGSHPSR
55-60	WAGMGR
61-65	RLLLR
66-79	SVQLHDSGNYSCYR
80-104	AGRPAGTVHLLVDVPPEEPQLSCFR
105-118	KSPLSNVVCEWGPR
119-126	STPSLTTK
127-132	AVLLVR
135-148	QNSPAEDFQEPCQY
133-154	KFQNSPAEDFQEPCQYSQESQK
156-168	SCQLAVPEGDSSF
170-183	IVSMCVASSVGSKF
186-210	TQTFQGCGILQPDPPANITVTAVAR
214-231	WLSVTWQDPHSWNSSFYR
232-237	LRFELR
245-252	TFTTWMVK
253-268	DLQHHCVIHDAWSGLR
269-274	HVVQLR
274-284	RAQEEFGQGEW
288-296	SPEAMGTPW
301-319	SPPAENEVSTPMQALTTNK
320-327	DDDNILFR

Table 4.1. Peptides from IL-6R digest selected for FPOP quantification.

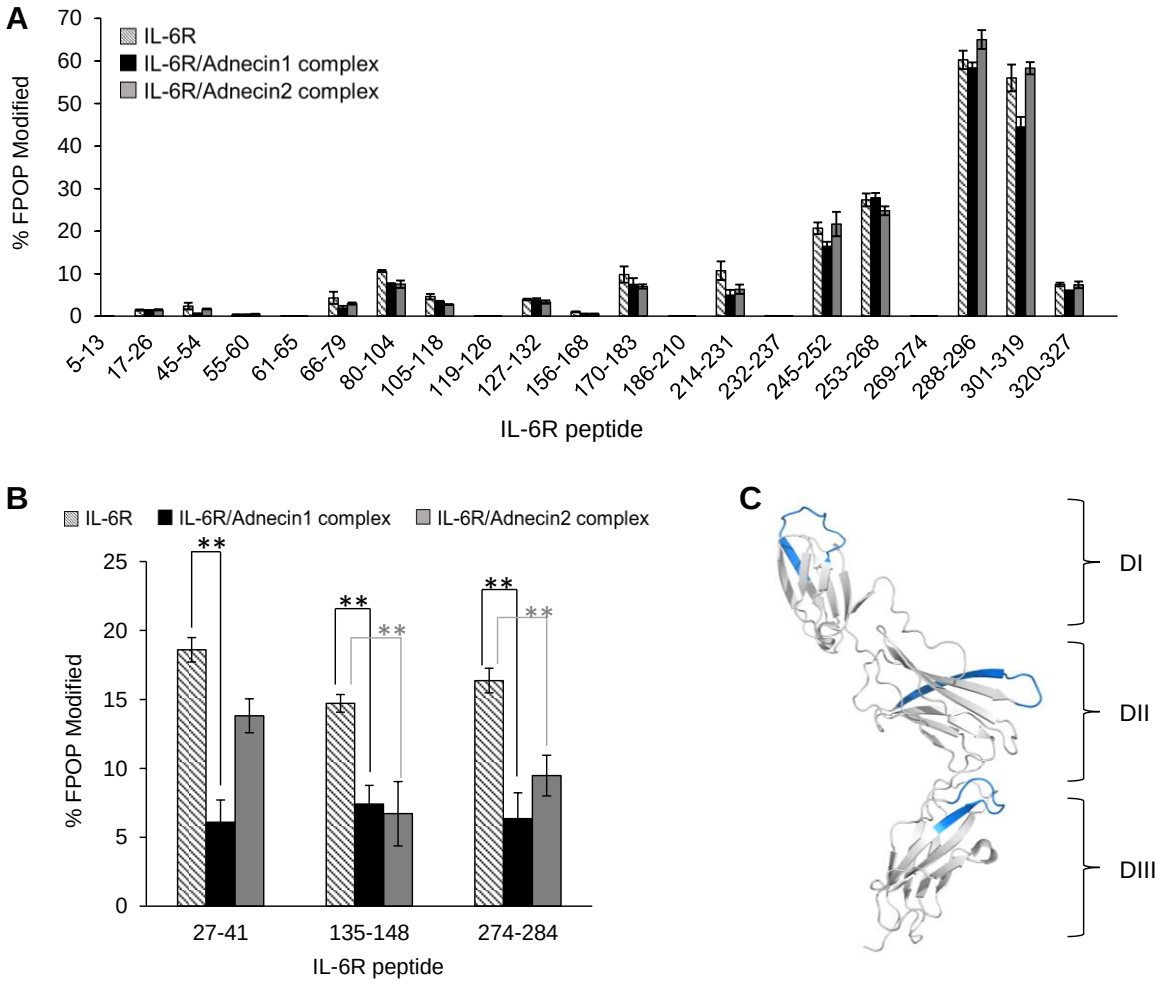


Figure 4.5. FPOP modification extents of IL-6R peptide regions and the locations of the regions showing altered solvent accessibility. (A) Regions of IL-6R without significant change in their extent of FPOP modification. (B) Regions of IL-6R with significantly decreased (**$\Delta_{rel.}$ > 40%, p < 0.005) FPOP modification, indicating reduced solvent accessibility upon adnectin1/adnectin2 binding. (C) Regions in (B) mapped onto IL-6R (PDB: 1N26) are colored blue. The three domains of IL-6R are referred as DI, DII and DIII on the structure.

Regions 27-41, 274-284 and 135-148 in adnectin1-bound IL-6R, and regions 274-284, 135-148 in adnectin2-bound IL-6R show remarkably decreased FPOP modification (relative difference >

40%, $p < 0.005$ in Student's t-test) (Figure 4.5B), suggesting those regions undergo major conformational changes introducing reduced solvent accessibility upon adnectin binding (Table 4.2 summarizes the FPOP modifications of those regions exhibiting statistically significant differences in solvent accessibility upon adnectin binding). We also observed minor differences in FPOP modification for region 301-319 in the adnectin1-bound state, which is likely attributed to minor structural or dynamical perturbation on this region upon adenctin1 binding. The 3D structure of IL-6R consists of three domains including the N-terminal Ig-like domain (DI) and two cytokine-binding domains (DII and DIII). All the regions for which solvent accessibility is significantly altered in FPOP adopt a flexible loop structure (Figure 4.5C). In addition, we found the overlapping tryptic and chymotryptic peptides (e.g., tryptic peptide 133-154 and chymotryptic peptide 135-148) reveal correlated trends of FPOP modification change (Figure 4.6). This indicates that the observed difference is due to structural changes of IL-6R in the holo states, instead of structure-based proteolytic bias caused by the FPOP modification.

Ligand	**IL-6R peptide**	**Sequence**	**Relative difference (%)**	***p*-value**
Adnectin1	27-41	TCPGVEPEDNATVHW	67	0.001
	274-284	RAQEEFGQGEW	61	0.004
	135-148	QNSPAEDFQEPCQY	50	< 0.001
Adnectin2	135-148	QNSPAEDFQEPCQY	54	0.001
	274-284	RAQEEFGQGEW	42	0.004

Table 4.2. Summary of the regions showing significant changes in conformation and/or dynamics identified by FPOP.

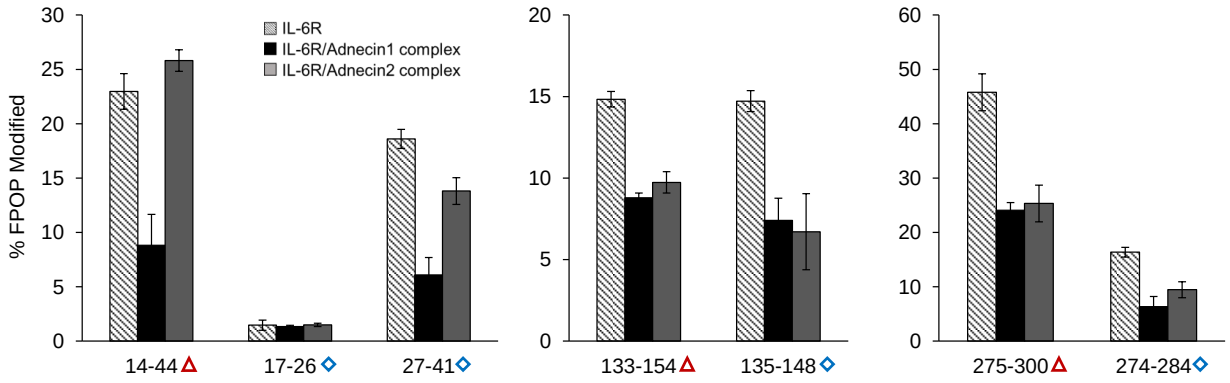


Figure 4.6. Correlation in FPOP quantification for overlapping peptides. The overlapping regions covered by tryptic (red triangle) and chymotryptic (blue diamond) fragments show correlated trend in term of the change in the FPOP modification. For example, region represented by tryptic fragment 14-44 are covered by two chymotryptic fragment 17-26 and 27-41, whereas FPOP quantification on the shorter chymotryptic fragments indicates the decreased solvent accessibility upon adnectin binding is attributed to region 27-41, with the solvent accessibility of region 17-26 remains unchanged.

A significant advantage of FPOP is that it provides residue-level information (Figure 4.7). His^{40} and Trp^{41} are the residues that are predominantly modified by FPOP in region 27-41 (TCPGVEPEDNATVHW). The two residues are located at the front end of a β-stand connected to the loop. The side chain of His^{40} is exposed on the protein surface, and the aromatic side chain of Trp^{41} is largely protected inside the protein core. Note that the amino acid residues from the sequence that compose the loop (^{30}GVEPEDNAT^{38}) on region 27-41 are much less reactive than His and Trp, with the most reactive Val and Pro being ~ 20X and 6X less reactive compared to Trp and His, respectively.[20] Therefore, the hydroxyl radical preferentially modifies His^{40} and Trp^{41} rather than the less reactive ones from the loop. The local selectivity is also pronounced for

some Met containing regions (e.g., 245-252 and 288-296). Moreover, the product-ion spectrum does not definitively indicate what residue is modified on region 274-284, but it does show that the FPOP modification occurs on either Pro^{138}, Phe^{142}, Pro^{145} or Tyr^{148}.

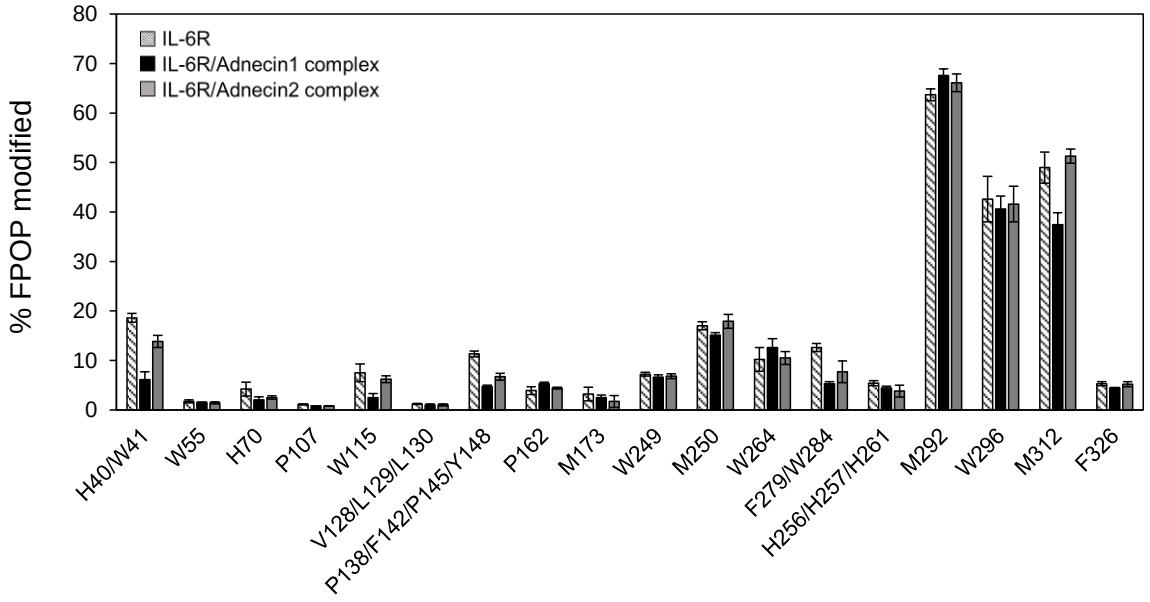


Figure 4.7. FPOP modification level for IL-6R residues.

Although peptide regions in Table 4.2 undergo significantly reduced FPOP modification in the holo states, one should be cautious in interpreting the data because the change in FPOP extent may result either from the epitope binding, or from decreased dynamics and flexibility induced remotely from the adnectin binding site. For the residues with high intrinsic reactivity with the hydroxyl radical, a relatively modest change in their solvent accessibility in response to the dynamic motion can result in a dramatic difference in their FPOP modification.[44-45] Our hypothesis is that dynamic motions occurring within the sub-second time range will be differentiated by fast labeling but not by slow labeling that presents an averaged view over

seconds, and applying orthogonal footprinting can distinguish binding from remote dynamics change.

There are ways to improve the confidence in the FPOP experiment. For example, the footprinting can be done as a function of time to give outcomes similar to those typically obtained for HDX. Assigning differences can be elaborated further with kinetic curves[46] (e.g., five time points) rather than single point, but this requires considerable investments in LC/MS/MS analysis time, data processing, and interpretation as well as more sample. At this point, we decided to turn instead to another type of footprinting.

4.4.3 Carboxyl group footprinting

The effectiveness of protein footprinting to map the epitope and conformational changes depends on whether the reagent-active residues are located on those regions.[12] Although residue-specific labeling provides less structural information due to limited numbers of target residues on the protein surface, fortuitously, the regions of IL-6R with significant changes in the solvent accessibility mapped by FPOP (^{27}TCPGVEPEDNATVHW^{41}, ^{274}QNSPAEDFQEPCQY^{284} and ^{135}RAQEEFGQGEW^{148}) are rich in Asp/Glu as potential GEE modification sites, and we expect carboxyl group footprinting, as an orthogonal approach to provide more site-specific insights for those acidic regions.

In the GEE reaction, carboxyl groups of Asp/Glu located on the surface are readily modified, whereas ones surrounded by hydrophobic amino acids or buried in the interior of the structure undergo less or even no modifications (assuming there is no additional steric restrictions prohibiting the access of EDC and the GEE molecules).[33, 47] To investigate the kinetics of the labeling, we performed sparse, time-dependent GEE footprinting (1, 3 and 10 min) on the apo and adnectin-bound IL-6R. Given that for 16 Asp/Glu-containing chymotryptic peptides, 17-26,

109-134, 247-254, 255-264 and 297-315 have no detectable modification, we performed quantification for the other 11 peptides.

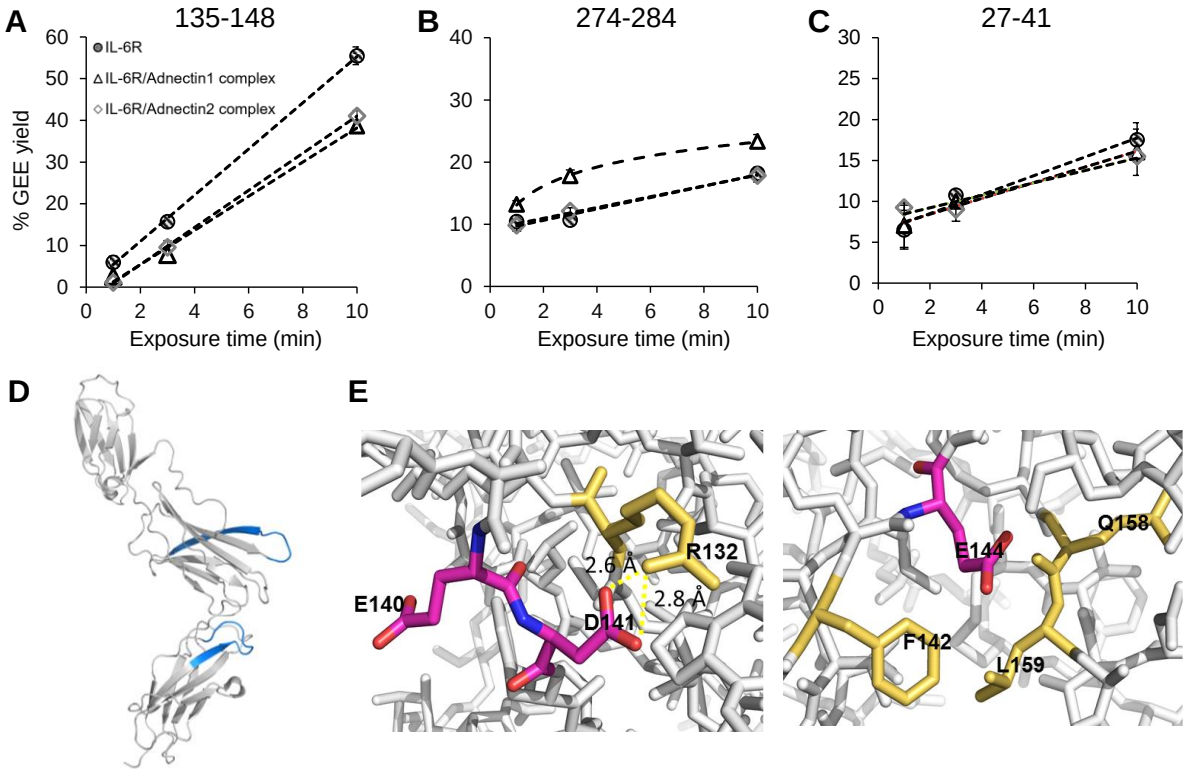


Figure 4.8. GEE labeling kinetics for selected IL-6R peptides in the ligand-free (gridded circle), adnetin1-bound (triangle) and adnetin2-bound state (diamond) state. (A) Region 135-148 shows decreased GEE incorporation upon adnectin1/adnectin2 binding, whereas (B) region 274-284 shows increased GEE modification upon adnectin1 binding. (C) A representative peptide region without differentiable GEE modification extent between apo and holo as the control. Dashed trend curves in (A), (B) and (C) are generated by linear or 2nd-degree polynomial fitting. (D) Region 135-148 and 274-284 mapped onto IL-6R. (E) Microenvironments of Glu^{140}, Asp^{141} and Glu^{144}.

From the time-dependent GEE footprinting, we found two regions 135-148 on DII and 274-284 on DIII show a clear difference upon binding of either adnectin1 or adnectin2 (Figure 4.8A and B). The two regions are mapped onto the IL-6R structure in Figure 4.8D. The rates of GEE incorporation for region 135-148 in the two holo states decrease significantly with a relative change of 30% at 10 min of labeling, and the differences are made more apparent by the time-dependent results, indicating prominently decreased solvent accessibility of the region upon adnectin binding. Region 135-148 contains Glu^{140}, Asp^{141} and Glu^{144} as possible GEE modification sites, but we found the modification is exclusively on Glu^{140}. The structure of IL-6R indicates that the GEE-modified Glu^{140} is on the surface of a loop with a flexible carboxyl group amenable to EDC/GEE reaction (Figure 4.8E). By contrast, Asp^{141} is involved in the front end of a β-strand with its side-chain hydrogen bonded to Arg^{132}, and Glu^{144} is also located on the same β-strand with its side chain occluded by surrounding residues Phe^{142}, Gln^{158} and Leu^{159}. The GEE labeling reaction requires activation of the carboxyl group by EDC, and the sizes of EDC and GEE are larger than the small reagents in FPOP and HDX, suggesting steric requirements for the reaction. We reason that although Asp/Glu as charged residues are often prone to be on the protein surface, their solvent accessibility can be largely diminished or even completely blocked by their microenvironments.

We found region 247-284 to be slightly deprotected in the adnectin1-bound state (Figure 4.8B) as reported for the acidic residues. Due to a steric effect similar to that described above, Glu^{283}, which is located on the unshielded surface loop, is readily modified by GEE. Glu^{277} and Glu^{278}, however, which are located on or close to the end of a β-strand, remain unmodified owing to the protection from the surrounding loops and hydrogen bonds. The increased solvent accessibility can thus be attributed to the side chain of Glu^{283}. Interestingly, residue-level analysis of FPOP

modifications on the same region shows that the Phe^{279}/Trp^{284} side-chain solvent accessibilities are reduced (Figure 4.7), and the analysis and interpretation for the motion of the region will be discussed later in section 4.4.4.

By contrast, a majority of peptides, as represented by 27-41, show nearly identical labeling kinetics for bound and unbound states, supporting their role as controls and suggesting their local conformations remain unchanged upon adnectin binding (Figure 4.8C and Figure 4.7). The labeling extents of those regions at 10 min, at which time the difference in modification is expected to be greatest, are not differentiable (Figure 4.10). Unlike FPOP labeling, which is often performed over a single exposure time, kinetic curves of GEE labeling provide statistical weight by tracking the labeling over a time course. We also found the GEE modification extent generally increases with the reaction time, but the labeling of some peptide regions can occur very rapidly. For example, regions 1-16, 288-296 and 316-326 show bursts in their GEE incorporation at the first 1 min of the reaction (Figure 4.9), indicative of their flexible loop or coil secondary structures.

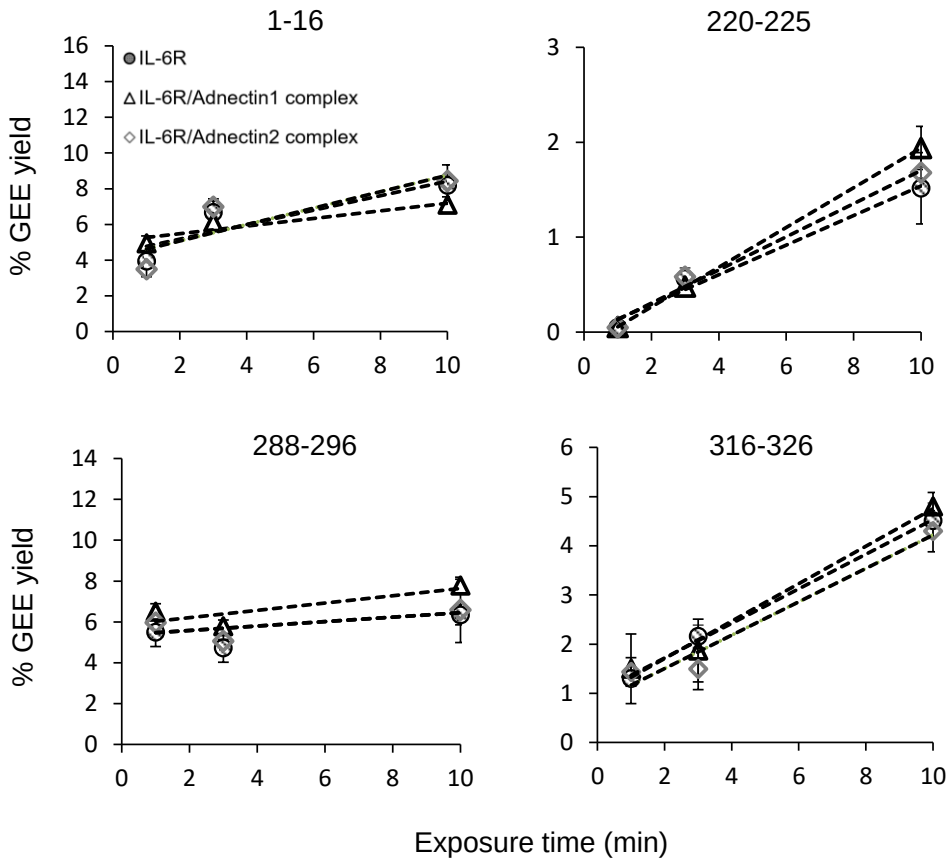


Figure 4.9. Representative GEE labeling kinetics for selected IL-6R peptides without differentiable GEE modification extent. Ligand-free (gridded circle), adnetin1-bound (triangle) and adnetin2-bound state (diamond).

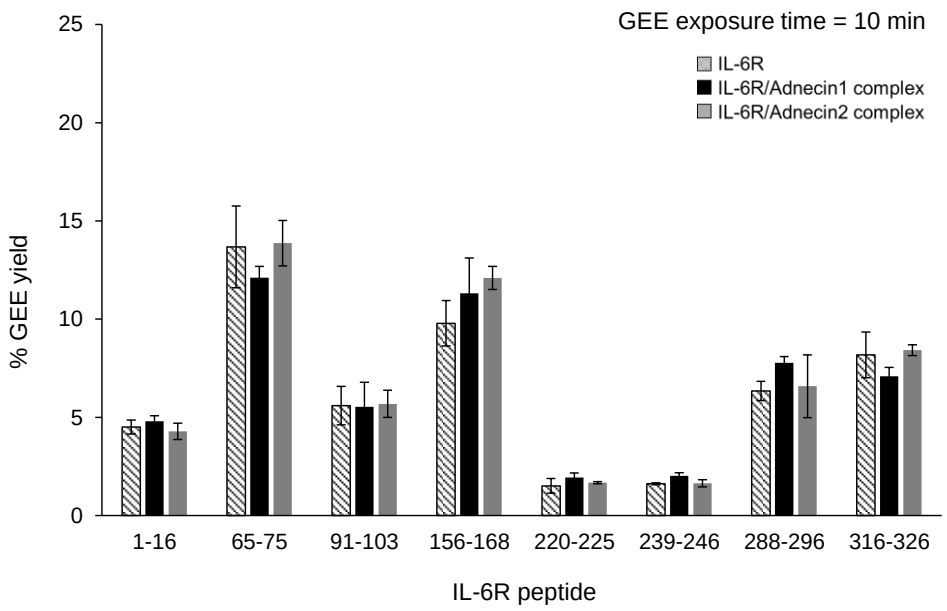


Figure 4.10. Regions without appreciable difference in their modification extent between the ligand-free and adnectin1/adenctin2-bound IL-6R. Samples are labeled with GEE for 10 min.

4.4.4 Structural features from complementary footprinting

HDX rates of exchange of labile amide hydrogens are characteristic of local backbone conformations and related to its hydrogen-bonding pattern and solvent accessibility, which are affected by protein binding. In highly dynamic, unstructured regions, the exchange reaction proceeds on the millisecond to second timescale, whereas amides that are hydrogen bonded will exchange more slowly (minutes to days).[48] With comparable reaction rates to HDX, GEE targets solvent-accessible carboxyl group side chains. Labeling by FPOP, however, is considerably faster than HDX and GEE because the lifetime is microseconds for the primary hydroxyl radical and milliseconds for all radicals.[22, 49] Hydroxyl radicals react with a variety of amino acid residues, and the labeling extent of a particular residue site is a function of the inherent reactivity and solvent accessibility of the amino acid side chain. Clearly the approaches are complementary

in location of footprinting (protein backbone vs. side chain), residue specificity, and rate of reaction. The more rapid approaches offer an opportunity to understand protein dynamics and minor structural fluctuations.[43]

The epitope and critical contacting residues

In FPOP and GEE footprinting, region 135-148 becomes significantly protected upon binding of adnectin1 or adnectin2, whereas HDX focuses on region 130-141 (region 142-147 does not change). Taken together, the integrated outcome strongly suggests the short segment ^{135}QNSPAED^{141} to contain the critical binding epitope. The suggested epitope is conserved for adnectin1 and adenctin2 binding. Mapping the proposed epitope onto the IL-6R structure highlights a loop on the DII domain of IL-6R (Figure 4.11).

A closer examination of the residue-specific data may reveal the contacting residues. Generally, FPOP modification on a residue is assigned when a +15.9949 Da shift is observed in the product-ion series of the modified peptide. For peptide 135-148, each of FPOP modified residues Pro^{138}/Phe^{142}/Pro145/Tyr148 produces multiple +16 Da products as structural isomers (e.g., by o-/m-/p- oxidation of Phe). Those isomers of peptide 135-148 cannot be distinguished by mass and give very slightly different retention times. The complicated chromatogram for oxidized peptide 135-148 makes FPOP quantification of a specific residue difficult. As for the suggested epitope 135-141, however, we identified Pro^{138} to be the modified residue responsible for the changes in FPOP. Carboxyl group footprinting also slows for the proposed epitope, but the information is restricted to Glu^{140} as it is the only residue modified in region 135-148. Although there are Asp^{141} and Glu^{144} as potential GEE modification sites in this region, no modification is detected for the two residues owing to shielding of their microenvironments as discussed above.

Summarizing the information from the complementary methods, we posit that ^{135}QNSPAED^{141} represents the region containing the epitope of IL-6R, with Pro^{138} and Glu^{140} being possible binding residues or closely adjoining the critical binding residues (Figure 4.11 shows Pro^{138} and Glu^{140} in the surface presentation with side chains extruding). At this time, we cannot rule out the residues on region 135-141 that are not mapped by the hydroxyl radical or GEE probe (e.g., Gln ^{135}, Asn ^{136}, Ser^{137} and Ala^{139}), but we can potentially identify some key interacting residues as well as narrow down the epitope to a short segment by FPOP and GEE-reactive residues that are involved in binding. Note that HDX fails to report the backbone solvent accessibility of Pro^{138} owing to its lack of an amide hydrogen atom, which may account for why the putative epitope region 130-141 only shows a modest protection in HDX, despite the two adnectins both being high-affinity ligands against IL-6R.

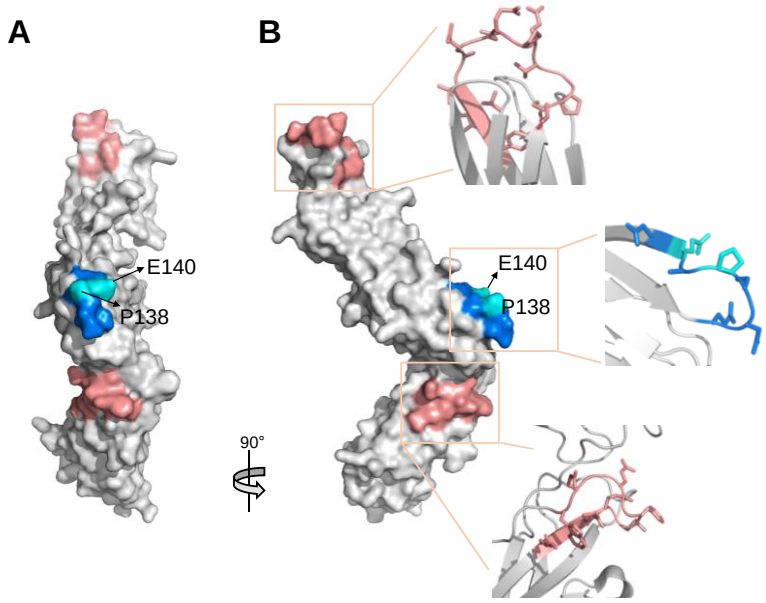


Figure 4.11. Structural change of IL-6R (PDB: 1N26) upon adnectin binding in the surface representation. IL-6R consists of three domains. DI, DII and DIII are located from the top to the bottom on the structure. (A) Front view and (B) side view of the regions that undergo

conformational change in structure or dynamics upon adnectin binding. Close-up views of the region in (A) shows the secondary structure of the region and the side chain orientation. Region 135-141 as the proposed epitope for adentin1/adnectin2 binding are colored in blue with Pro^{138} and Glu^{140} highlighted in cyan. Region 27-41 and 274-284 (located on DI and DIII respectively) whose conformation is affected by adnectin binding are colored in coral.

Conformational dynamics and side-chain motion of IL-6R loop regions

Protein structure fluctuations will be affected by protein binding.[50] If the sampling time of a footprinting reaction is long with respect to the protein dynamics, differences will be averaged and no effect seen. FPOP is capable of reporting on regions showing fast dynamics because its sampling time is short with respect to local motions.[19, 42, 51] In the adnectin1-bound IL-6R, the decreased solvent accessibility reported by FPOP for region 27-41 is not observed by HDX or carboxyl group footprinting, indicative of fast changes in the conformational dynamics of region 27-42 upon adenctin1 binding. Region 27-41 is located on the DI domain of IL-6R as a surface-exposed loop (Figure 4.11B) where changes in dynamics are likely to happen.

For amino-acid residues with high intrinsic reactivities with the hydroxyl radical, drastic changes in their modification occur in response to modest change in the solvent accessibility,[45] whereas differences in modification extent of less reactive residues (e.g., Leu, Pro and Val) is expected to be observed for large conformational changes. His^{40} and Trp^{41} are the only two residues modified by FPOP in region 27-42, and they do exhibit dramatic differences in FPOP modification with a relatively decrease of 67% in the adenctin1-bound state. The observed protection on the two residues with high susceptibility to the hydroxyl radical suggests relatively modest changes in the structure or dynamics upon adnectin-1 binding. Considering the binding

strength of adnectin1 (K_d ~ 6.2 pM) is greater than that of adnectin2 (K_d ~ 46 nM), we posit that the binding of adnectin1 stabilizes IL-6R more by reducing the local flexibility of region 27-41.

Interestingly, in contrast to the decreased FPOP modification of residue Phe^{279}/Trp^{284} in region 274-284 in adnectin1 bound IL-6R, the GEE labeling of Glu^{283} increases slightly for the adnectin1-bound state, but not for adnectin2, whereas no difference in HDX occurs for this region. This suggests that region 274-284 undergoes minor structural perturbation in dynamic and/or sidechain reorientation upon adnectin binding, which may not cause changes in the amide hydrogen-binding pattern. Considering the hydrophobicity of Phe/Trp and the polarity of Glu, it is possible that the binding causes side-chain motions involving inward rotation of Phe^{279}/Trp^{284} with its solvent accessibility decreasing, whereas the side chain of Glu^{283} rotates outward and becomes relatively solvent exposed. These solution-state motions would not be reflected in a single static 3D-structure. Furthermore, region 274-284 spans a flexible loop, where side-chain rotation is more facile than for a more rigid α-helix or β-sheet. This may have implication in the low-affinity binding of IL-6 to IL-6R that precedes the binding to a signal-transducing molecule gp130 to form high-affinity functional complex.[52-53] An investigation into the structure of the IL-6R/IL6/gp130 complex (PDB: 1P9M) further reveals that the loop region represented by peptide 274-284 serves as one of the interfaces with IL-6 in this biologically-relevant complex (Figure 4.12). The structural changes adopted by this region may be inferred by the therapeutic efficacy of the adnectins to block IL-6-mediated signal transduction through inhibiting the binding of IL-6 to IL-6R.

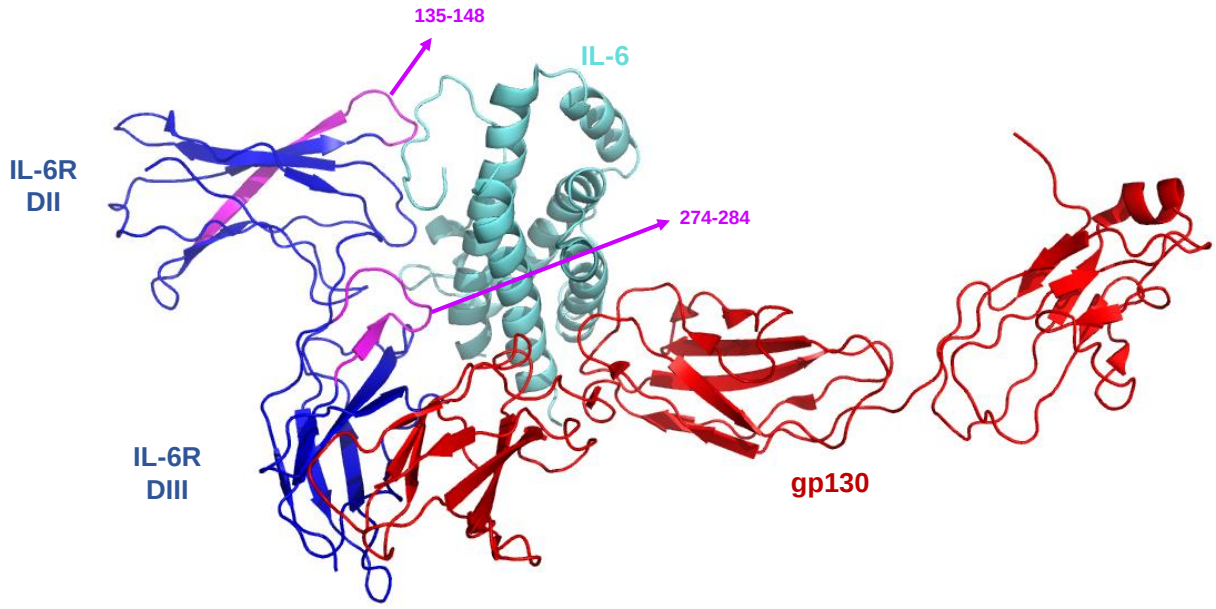


Figure 4.12. Cocrystal structure of IL-6R/IL6/gp130 complex. Region 135-148 and 274-284 (magenta) on IL-6R (blue) both closely interact with IL-6 (cyan) as shown in the 3D-structure of the biologically-relevant human IL-6/IL-6R/gp130 (2:2:2 stoichiometry) hexameric complex. The complex consists of two IL-6, IL-6R, and gp130 hetero-trimers (PDB: 1P9M, only one copy of the hetero-trimer is shown for clarity). IL-6 and the signal transducer glycoprotein 130 (gp130) is in cyan and red respectively. Domain DI of IL-6R is absent in this 3D-structure.

4.5 Conclusions

Solution-state protein-protein interactions and related conformational changes can be interrogated with high spatial resolution by using orthogonal footprinting and structural mapping. We proposed the epitope of IL-6R is in the region 135-141 and concluded that adnectin binding affects fast dynamics and side-chain reorientation of some of IL-6R's flexible loops. Those "hidden" motions in structure and/or dynamics are invisible to a relatively slow footprinting method like HDX. The results support a more comprehensive understanding of IL-6R HOS and

highlight the sensitivity of FPOP towards fast structural changes owing to the short half-life of hydroxyl radicals and higher coverage compared to HDX and site-specific carboxyl group footprinting. Their combined use not only serves to categorize and interpret changes in footprinting as due to protection from binding or to remote structural changes occurring with binding, but also adds confidence to assign the epitope where any stand-alone method has uncertainty. This integrated approach shows great utility for charactering protein and protein complex, which can be applied efficiently to assist understanding and optimizing the design of protein therapeutics.

4.6 References

1. Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianférani, S., Characterization of Therapeutic Antibodies and Related Products. *Anal. Chem.* **2013,** *85* (2), 715-736.

2. Chen, G.; Warrack, B. M.; Goodenough, A. K.; Wei, H.; Wang-Iverson, D. B.; Tymiak, A. A., Characterization of protein therapeutics by mass spectrometry: recent developments and future directions. *Drug Discov. Today* **2011,** *16* (1-2), 58-64.

3. Weiss, W. F. t.; Gabrielson, J. P.; Al-Azzam, W.; Chen, G.; Davis, D. L.; Das, T. K.; Hayes, D. B.; Houde, D.; Singh, S. K., Technical Decision Making With Higher Order Structure Data: Perspectives on Higher Order Structure Characterization From the Biopharmaceutical Industry. *J. Pharm. Sci.* **2016,** *105* (12), 3465-3470.

4. Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N., Strategies and challenges for the next generation of therapeutic antibodies. *Nat. Rev. Immunol.* **2010,** *10* (5), 345-352.

5. Acharya, K. R.; Lloyd, M. D., The advantages and limitations of protein crystal structures. *Trends Pharmacol. Sci.* **2005,** *26* (1), 10-14.

6. Davis, A. M.; Teague, S. J.; Kleywegt, G. J., Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem. Int. Ed.* **2003,** *42* (24), 2718-2736.

7. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **2006,** *1* (6), 2876-2890.

8. Kong, J.; Yu, S., Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochimica et Biophysica Sinica* **2007,** *39* (8), 549-559.

9. Tcherkasskaya, O.; Bychkova, V. E.; Uversky, V. N.; Gronenborn, A. M., Multisite Fluorescence in Proteins with Multiple Tryptophan Residues: APOMYOGLOBIN NATURAL VARIANTS AND SITE-DIRECTED MUTANTS. *J. Biol. Chem.* **2000,** *275* (46), 36285-36294.

10. Huang, R. Y.-C.; Chen, G., Higher order structure characterization of protein therapeutics by hydrogen/deuterium exchange mass spectrometry. *Anal. Bioanal. Chem.* **2014,** *406* (26), 6541-6558.

11. Jones, L. M.; B. Sperry, J.; A. Carroll, J.; Gross, M. L., Fast Photochemical Oxidation of Proteins for Epitope Mapping. *Anal. Chem.* **2011,** *83* (20), 7657-7661.

12. Wecksler, A. T.; Kalo, M. S.; Deperalta, G., Mapping of Fab-1:VEGF Interface Using Carboxyl Group Footprinting Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2015,** *26* (12), 2077-2080.

13. Mihara, M.; Hashizume, M.; Yoshida, H.; Suzuki, M.; Shiina, M., IL-6/IL-6 receptor system and its role in physiological and pathological conditions. *Clinical Science* **2012,** *122* (4), 143-159.

14. Gaillard, J. P.; Mani, J. C.; Liautard, J.; Klein, B.; Brochier, J., Interleukin-6 receptor signaling. I. gp80 and gp130 receptor interaction in the absence of interleukin-6. *European Cytokine Network* **1999,** *10* (1), 43-48.

15. Lipovsek, D., Adnectins: engineered target-binding protein therapeutics. *Protein Eng. Des. Sel.* **2011,** *24* (1-2), 3-9.

16. Cummins, D. J.; Espada, A.; Novick, S. J.; Molina-Martin, M.; Stites, R. E.; Espinosa, J. F.; Broughton, H.; Goswami, D.; Pascal, B. D.; Dodge, J. A.; Chalmers, M. J.; Griffin, P. R., Two-Site Evaluation of the Repeatability and Precision of an Automated Dual-Column Hydrogen/Deuterium Exchange Mass Spectrometry Platform. *Anal. Chem.* **2016,** *88* (12), 6607-6614.

17. Zhang, Z.; Zhang, A.; Xiao, G., Improved Protein Hydrogen/Deuterium Exchange Mass Spectrometry Platform with Fully Automated Data Processing. *Anal. Chem.* **2012,** *84* (11), 4942-4949.

18. Engen, J. R.; Wales, T. E., Analytical Aspects of Hydrogen Exchange Mass Spectrometry. *Annual Review of Analytical Chemistry* **2015,** *8* (1), 127-148.

19. Gau, B.; Garai, K.; Frieden, C.; Gross, M. L., Mass Spectrometry-Based Protein Footprinting Characterizes the Structures of Oligomeric Apolipoprotein E2, E3, and E4. *Biochemistry* **2011,** *50* (38), 8117-8126.

20. Xu, G.; Chance, M. R., Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **2007,** *107* (8), 3514-3543.

21. Hambly, D. M.; Gross, M. L., Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J. Am. Soc. Mass Spectrom.* **2005,** *16* (12), 2057-63.

22. Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L., Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* **2009,** *81* (16), 6563-6571.

23. Zhang, Y.; Wecksler, A. T.; Molina, P.; Deperalta, G.; Gross, M. L., Mapping the Binding Interface of VEGF and a Monoclonal Antibody Fab-1 Fragment with Fast Photochemical Oxidation of Proteins (FPOP) and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, 1-9.

24. Li, J.; Wei, H.; Krystek, S. R.; Bond, D.; Brender, T. M.; Cohen, D.; Feiner, J.; Hamacher, N.; Harshman, J.; Huang, R. Y. C.; Julien, S. H.; Lin, Z.; Moore, K.; Mueller, L.; Noriega, C.; Sejwal, P.; Sheppard, P.; Stevens, B.; Chen, G.; Tymiak, A. A.; Gross, M. L.; Schneeweis, L. A., Mapping the Energetic Epitope of an Antibody/Interleukin-23 Interaction with Hydrogen/Deuterium Exchange, Fast Photochemical Oxidation of Proteins Mass Spectrometry, and Alanine Shave Mutagenesis. *Anal. Chem.* **2017,** *89* (4), 2250-2258.

25. Yan, Y.; Chen, G.; Wei, H.; Huang, R. Y.-C.; Mo, J.; Rempel, D. L.; Tymiak, A. A.; Gross, M. L., Fast Photochemical Oxidation of Proteins (FPOP) Maps the Epitope of EGFR Binding to Adnectin. *J. Am. Soc. Mass Spectrom.* **2014,** *25* (12), 2084-2092.

26. Li, K. S.; Rempel, D. L.; Gross, M. L., Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1–42 Aggregation. *J. Am. Chem. Soc.* **2016,** *138* (37), 12090-12098.

27. Guan, L.; Ronald Kaback, H., Site-directed alkylation of cysteine to test solvent accessibility of membrane proteins. *Nat. Protocols* **2007,** *2* (8), 2012-2017.

28. Borotto, N. B.; Zhou, Y.; Hollingsworth, S. R.; Hale, J. E.; Graban, E. M.; Vaughan, R. C.; Vachet, R. W., Investigating Therapeutic Protein Structure with Diethylpyrocarbonate Labeling and Mass Spectrometry. *Anal. Chem.* **2015,** *87* (20), 10627-10634.

29. Zhou, Y.; Vachet, R. W., Increased protein structural resolution from diethylpyrocarbonate-based covalent labeling and mass spectrometric detection. *J. Am. Soc. Mass Spectrom.* **2012,** *23* (4), 708-17.

30. Mendoza, V. L.; Vachet, R. W., Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **2009,** *28* (5), 785-815.

31. Zhang, H.; Wen, J.; Huang, R. Y. C.; Blankenship, R. E.; Gross, M. L., Mass spectrometry-based carboxyl footprinting of proteins: method evaluation. *Int. J. Mass spectrom.* **2012,** *312*, 78-86.

32. Kaur, P.; Tomechko, S. E.; Kiselar, J.; Shi, W.; Deperalta, G.; Wecksler, A. T.; Gokulrangan, G.; Ling, V.; Chance, M. R., Characterizing monoclonal antibody structure by carboxyl group footprinting. *mAbs* **2015,** *7* (3), 540-552.

33. Liu, H.; Zhang, H.; King, J. D.; Wolf, N. R.; Prado, M.; Gross, M. L.; Blankenship, R. E., Mass spectrometry footprinting reveals the structural rearrangements of cyanobacterial orange carotenoid protein upon light activation. *Biochim. Biophys. Acta* **2014,** *1837* (12), 1955-1963.

34. Ramamurthy, V.; Krystek Jr, Stanley R.; Bush, A.; Wei, A.; Emanuel, Stuart L.; Das Gupta, R.; Janjua, A.; Cheng, L.; Murdock, M.; Abramczyk, B.; Cohen, D.; Lin, Z.; Morin, P.; Davis, Jonathan H.; Dabritz, M.; McLaughlin, Douglas C.; Russo, Katie A.; Chao, G.; Wright, Martin C.; Jenny, Victoria A.; Engle, Linda J.; Furfine, E.; Sheriff, S., Structures of Adnectin/Protein Complexes Reveal an Expanded Binding Footprint. *Structure* **2012,** *20* (2), 259-269.

35. Zhang, H.; Liu, H.; Blankenship, R. E.; Gross, M. L., Isotope-Encoded Carboxyl Group Footprinting for Mass Spectrometry-Based Protein Conformational Studies. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (1), 178-181.

36. Collier, T. S.; Diraviyam, K.; Monsey, J.; Shen, W.; Sept, D.; Bose, R., Carboxyl Group Footprinting Mass Spectrometry and Molecular Dynamics Identify Key Interactions in the HER2-HER3 Receptor Tyrosine Kinase Interface. *J. Biol. Chem.* **2013,** *288* (35), 25254-25264.

37. Chen, E.; Salinas, N. D.; Huang, Y.; Ntumngia, F.; Plasencia, M. D.; Gross, M. L.; Adams, J. H.; Tolia, N. H., Broadly neutralizing epitopes in the Plasmodium vivax vaccine candidate Duffy Binding Protein. *Proceedings of the National Academy of Sciences* **2016,** *113* (22), 6277-6282.

38. Wei, H.; Mo, J.; Tao, L.; Russell, R. J.; Tymiak, A. A.; Chen, G.; Iacob, R. E.; Engen, J. R., Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. *Drug Discov. Today* **2014,** *19* (1), 95-102.

39. Malito, E.; Faleri, A.; Lo Surdo, P.; Veggi, D.; Maruggi, G.; Grassi, E.; Cartocci, E.; Bertoldi, I.; Genovese, A.; Santini, L.; Romagnoli, G.; Borgogni, E.; Brier, S.; Lo Passo, C.; Domina, M.; Castellino, F.; Felici, F.; van der Veen, S.; Johnson, S.; Lea, S. M.; Tang, C. M.;

Pizza, M.; Savino, S.; Norais, N.; Rappuoli, R.; Bottomley, M. J.; Masignani, V., Defining a protective epitope on factor H binding protein, a key meningococcal virulence factor and vaccine antigen. *Proc Natl Acad Sci U S A* **2013,** *110* (9), 3304-9.

40. Zhang, Q.; Willison, L. N.; Tripathi, P.; Sathe, S. K.; Roux, K. H.; Emmett, M. R.; Blakney, G. T.; Zhang, H. M.; Marshall, A. G., Epitope mapping of a 95 kDa antigen in complex with antibody by solution-phase amide backbone hydrogen/deuterium exchange monitored by Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2011,** *83* (18), 7129-36.

41. Zhang, H.-M.; Kazazic, S.; Schaub, T. M.; Tipton, J. D.; Emmett, M. R.; Marshall, A. G., Enhanced Digestion Efficiency, Peptide Ionization Efficiency, and Sequence Resolution for Protein Hydrogen/Deuterium Exchange Monitored by FT-ICR Mass Spectrometry. *Anal. Chem.* **2008,** *80* (23), 9034-9041.

42. Chen, J.; Rempel, D. L.; Gau, B. C.; Gross, M. L., Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level. *J. Am. Chem. Soc.* **2012,** *134* (45), 18724-31.

43. Hart, K. M.; Ho, C. M. W.; Dutta, S.; Gross, M. L.; Bowman, G. R., Modelling proteins' hidden conformations to predict antibiotic resistance. *Nature Communications* **2016,** *7*, 12965.

44. Li, Z.; Moniz, H.; Wang, S.; Ramiah, A.; Zhang, F.; Moremen, K. W.; Linhardt, R. J.; Sharp, J. S., High Structural Resolution Hydroxyl Radical Protein Footprinting Reveals an Extended Robo1-Heparin Binding Interface. *The Journal of Biological Chemistry* **2015,** *290* (17), 10729-10740.

45. Charvátová, O.; Foley, B. L.; Bern, M. W.; Sharp, J. S.; Orlando, R.; Woods, R. J., Quantifying Protein Interface Footprinting by Hydroxyl Radical Oxidation and Molecular Dynamics Simulation: Application to Galectin-1. *J. Am. Soc. Mass Spectrom.* **2008,** *19* (11), 1692-1705.

46. Niu, B.; Mackness, B. C.; Rempel, D. L.; Zhang, H.; Cui, W.; Matthews, C. R.; Zitzewitz, J. A.; Gross, M. L., Incorporation of a Reporter Peptide in FPOP Compensates for Adventitious Scavengers and Permits Time-Dependent Measurements. *Journal of the American Society for Mass Spectrometry* **2017,** *28* (2), 389-392.

47. Wen, J.; Zhang, H.; Gross, M. L.; Blankenship, R. E., Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectrometry-based footprinting. *Proceedings of the National Academy of Sciences* **2009,** *106* (15), 6134-6139.

48. Englander, S. W., Hydrogen Exchange and Mass Spectrometry: A Historical Perspective. *J. Am. Soc. Mass Spectrom.* **2006,** *17* (11), 1481-1489.

49. Vahidi, S.; Konermann, L., Probing the Time Scale of FPOP (Fast Photochemical Oxidation of Proteins): Radical Reactions Extend Over Tens of Milliseconds. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (7), 1156-1164.

50. Keskin, O., Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies. *BMC Struct. Biol.* **2007,** *7* (1), 31.

51. Hart, K. M.; Ho, C. M.; Dutta, S.; Gross, M. L.; Bowman, G. R., Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat Commun* **2016,** *7*, 12965.

52. Boulanger, M. J.; Chow, D.-c.; Brevnova, E. E.; Garcia, K. C., Hexameric Structure and Assembly of the Interleukin-6/IL-6 α-Receptor/gp130 Complex. *Science* **2003,** *300* (5628), 2101-2104.

53. Ward, L. D.; Howlett, G. J.; Discolo, G.; Yasukawa, K.; Hammacher, A.; Moritz, R. L.; Simpson, R. J., High affinity interleukin-6 receptor is a hexameric complex consisting of two molecules each of interleukin-6, interleukin-6 receptor, and gp-130. *J. Biol. Chem.* **1994,** *269* (37), 23286-9.

Chapter 5: Mapping Hydrophobic Interactions of Human Bromodomain with a Small Molecule Inhibitor by Hydrogen-Deuterium Exchange and Hydroxyl Radical Footprinting Mass Spectrometry

5.1 Abstract

Mass spectrometry-based protein footprinting, a valuable structural tool in mapping protein-ligand interaction, has been extensively applied to protein-protein complexes, showing successes in mapping large interfaces. Here we compared hydrogen-deuterium exchange (HDX) and hydroxyl radical footprinting using fast photochemical oxidation of proteins (FPOP), as a proof-of-concept study, to characterize the interaction of human bromodomain-containing protein 4 (BRD4) with a hydrophobic benzodiazepine inhibitor. HDX does not provide strong evidence for the location of the binding interface, presumably because the shielding of solvent by the small molecule is not large. Instead, HDX suggests that BRD4 appears to be stabilized by showing a modest decrease in dynamics caused by binding. In contrast, FPOP points to a critical binding region in the hydrophobic cavity, also identified by crystallography, and, therefore, exhibits higher sensitivity than HDX in mapping the interaction of BRD4 with compound 1. In the absence of or under low concentrations of the radical scavenger, FPOP modifications on Met residues are significantly different with respect to a minor change in protein conformation. This problem can be avoided by using a sufficient amount of proper scavenger, as suggested by the FPOP kinetics directed by a dosimeter of the hydroxyl radical.

5.2 Introduction

Despite the rapid development of protein-based therapeutic biologics, small molecules are still remarkably dominant in development pipelines of the biopharmaceutical industry, making up over 90 percent of the therapeutics in use.[1] Small-molecule drugs have relatively stable chemical property and are mostly non-immunogenic. Their tiny size and chemical composition make them easy to penetrate cell membranes and reach desired delivery destinations. Generally, design of

small-molecule drugs is based on specific macromolecules as targets for inhibition or modulation. Thus, characterization of the interaction between a small molecule and its target macromolecule is important for drug development, as knowledge in their interaction is essential for understanding the mechanism of action. Compared to Chapter 4 discussing protein-protein interaction, this chapter is focused on evaluating MS-based protein footprinting, including FPOP and HDX to probe protein-small molecule interaction. From a method development perspective, this is also the first example of applying FPOP to investigate the direct binding interface of a protein with its small molecule ligand.

MS-based protein footprinting is a valuable tool to characterize protein structure and dynamics. Hydrogen-deuterium exchange (HDX) and hydroxyl radical footprinting are the two most commonly applied protein footprinting methods. HDX reports the stability and structural protection of the protein by measuring the exchange of amide hydrogens with deuterium on the protein backbone.[2] Its utility has been established by extensive applications in studying protein-protein, protein-DNA, protein-membrane interactions,[3-4] showing successes in probing large interfaces in protein-ligand complexes. Using HDX to probe protein-small molecule interfaces is challenging, however, possibly owing to the relatively small protection afforded to the protein by the small molecule. Hernychova et al.[5] applied HDX to the interaction of the protein MDM2 with a small molecule, Nutlin3, and observed reduced HDX kinetics upon ligand binding in regions surrounding the pocket only at relatively high protein-to-ligand ratio (1:4). Wang et al. [6] used HDX with ligand titration to obtain the affinities of a small molecule drug with Apolipoprotein E3 at the peptide-level, obtaining convincing changes in HDX at the binding sites. Despite these successes, direct mapping of a small molecule interaction when the binding is

hydrophobic may be particularly difficult because the binding interface does not involve hydrogen bonding of the protein backbone, explaining the lack of sensitivity in HDX.

By contrast, hydroxyl radical footprinting reports changes in solvent accessibility of amino-acid side chains via covalent and irreversible oxidative modification. Here differences in bound vs. unbound states are less dependent on H bonding. Furthermore, hydrophobic side chains (Phe, Leu, Ile, Val) are reactive with •OH, leading to potentially sensitive indications of binding.[7]

Here we describe the implementation of fast photochemical oxidation of protein (FPOP), which uses laser-induced hydrolysis of hydrogen peroxide to generate hydroxy radicals.[8] Using a radical scavenger, we varied the time scale of labeling, determined by the lifetime of the radical, is tuned to be faster than protein conformational change or unfolding induced by modification.[9] Although FPOP has been applied to protein folding, protein aggregation,[10-12] protein-ligand interactions including epitope/paratope mapping,[13-16] to our knowledge, FPOP, and hydroxyl radical footprinting in general, have not yet been applied to protein-small molecule interactions. Thus, we want to explore this opportunity and test the capability of the method for this application.

We chose to evaluate HDX and hydroxyl radical-based FPOP to characterize binding of a benzodiazepine inhibitor, benzyl (1-methyl-6-phenyl-4H-benzo[f][1,2,4]triazolo[4,3-a][1,4]diazepin-4-yl)carbamate (referred to as compound 1 hereafter), to human bromodomain-containing protein 4 (BRD4). BRDs are protein interaction modules that specifically recognize acetylation motifs, a key event in the reading process of epigenetic marks.[17] Inhibitors targeting BRD have therapeutic efficacy as anti-inflammatory, antiviral, and anticancer agents.[18] Compound 1 is a potent (nM) inhibitor that disrupts the function of the bromodomain family (BRD2, BRD3, and BRD4). A cocrystal structure of the bromodomain 1 of BRD4 and

compound 1 shows that the inhibitor occupies the central hydrophobic cavity of the protein used for acetyl-lysine recognition, thus directly antagonizing the interaction between the bromodomain and the acetylated histone peptides by steric competition.[18] The outcome of this comparative study is not only valuable for understanding the principles of HDX and FPOP but also should provide insights into the analytical capabilities of the methods, guiding the choice of suitable MS-based footprinting methods for investigating protein structure and interactions.

5.3 Experimental

5.3.1 Materials

The bromodomain 1 (residue 41-168) of human BRD4 (refereed to as BRD4 hereafter) and compound 1 were provided by Bristol Myers Squibb. *L*-Glutamine, *L*-methionine, catalase, hydrogen peroxide (H_2O_2), trifluoroacetic acid (TFA), formic acid (FA), phosphate buffer saline (PBS, 10 mM phosphate, 138 mM NaCl, 2.7 mM KCl), urea, dithiothreitol and iodoacetamide were from Sigma Aldrich (St. Louis, MO). Trypsin and chymotrypsin were from Promega (Madison, WI).

5.3.2 HDX

To prepare compound1-bound BRD4, the protein solution and compound 1 were combined to form a 40 µM:48 µM concentration ratio and incubated at 25 °C for 30 min to allow formation of the complex. Aliquots of 3 µL of 40 µM BRD4 with and without the complex were then continuously labeled in 37 µL of 1 × PBS in 99.9% deuterium oxide (pD =7.4), at 25 °C, at 10 s, 30 s, 1 min, 2 min, 15 min, 1 h, 2 h, 4 h, and 6 h time points. The reaction was quenched by adding 60 µL of 3 M urea and TFA (final pH = 3.0). Samples were then flash frozen in liquid nitrogen and stored at -80 °C. Control samples were prepared in like fashion by using 1 x PBS in

water in place of deuterium oxide. Each sample was thawed immediately prior to LC-MS analysis. The protein was digested by pepsin by using a custom-packed pepsin column (2 mm x 20 mm) at a flow rate of 100 µL/min. The resulting peptic peptides were then trapped on a Zorbax Eclipse XDB-C-8 trap column (2.1 x 15 mm, 3.5 µm) and desalted for 3 min, followed by separation of the peptides on a Hypersil Gold C-18 analytical column (2.1 x 50 mm, 2.5 µm) by using a gradient of 4-80 % acetonitrile with 0.1% formic acid at a flow rate of 100 µL/min. Peptides were detected using a LTQ XL Orbitrap mass spectrometer (Thermo Fisher Scientific), at a mass resolving power of 100000 at *m/z* 400. The capillary temperature and voltage for ESI were 275 °C and 49 V, respectively. All experiments were performed in duplicate.

5.3.3 HDX data analysis

To identify the peptic peptides generated from the protein and to provide a list of peptides to be followed during HDX, peptide mapping with a non-deuterated protein sample was conducted. The product−ion mass spectra were acquired in a data-dependent mode, with the six most abundant ions from each scan selected for MS/MS. Peptides, as identified using Mascot (Matrix Science, London, UK), were used to guide the HDX setup. The HDX mass spectra were analyzed with HDX Workbench (Scripps, Jupiter, FL). All experiments were performed in duplicate.

5.3.4 FPOP

A complex of bound BRD4 with compound 1 was prepared by incubating 20 µM BRD4 with compound 1 in a 1:1 molar ratio for 30 min at 25 °C in PBS. Unbound BRD4 was prepared as a 20 µM stock solution in PBS. Just prior to injection into the flow tube for FPOP, H_2O_2 was added to a final concentration of 20 mM. FPOP was performed as previously described[8], expect that for each state of the protein, FPOP was conducted in duplicate each for four different

concentrations of His (0, 0.3, 3, 30 mM) as the radical scavenger. The final concentration of BRD4 for FPOP labeling was 10 μM. After laser irradiation, the sample was collected in a tube containing 10 mM catalase and 20 mM Met to react with any leftover H_2O_2 and prevent oxidation artifacts. In addition, control samples of BRD4 with all the reagents (including H_2O_2) added were handled in the same manner in duplicate, but not laser-irradiated.

5.3.5 LC-MS/MS

For each 20 μL aliquot of the FPOP sample, urea was added to a final concentration of 6 M to denature the protein. Cysteines of the protein were reduced by adding dithiothreitol to a final concentration of 5 mM, and samples were incubated at 37 ^{o}C for 30 min. Iodoacetamide was subsequently added to a concentration of 15 mM to alkylate the free cysteines. Samples were incubated in the dark at room temperature for 30 min. PBS was added to dilute the urea to 1 M. Samples were digested overnight with trypsin or chymotrypsin at 37 ^{o}C with a protease-to-protein ratio of 1:25 (w/w). The digestion was terminated by acidifying the sample with 1% trifluoroacetic acid. The sample was then vacuum-dried on a speedvac, and each peptide digest was reconstituted with 60 μL 0.1% FA in water for LC-MS analysis.

An aliquot of 5 μL of protein digest was submitted to LC-MS/MS analysis. For the elution, solvent A was water with 0.1% (v/v) FA, and solvent B was 80% (v/v) acetonitrile with 0.1% (v/v) FA. Peptides were pre-concentrated on an Acclaim PepMap C18 column (Thermo Scientific, 100 μm × 2cm, 5μm, 100 Å) and desalted for 10 min at a flow rate of 4.5 μL/min by using solvent A. Separation was performed on a custom-packed C18 column (CSH, 75 μm × 15 cm, 3.5 μm, 130 Å) at a flow rate of 400 nL/min with a gradient of 2-65 % solvent B in 90 min using a Nano UltiMate 3000 Rapid Separation system (Dionex, Co.). The solution flow was directed with a Nanospray Flex source coupled for online detection using a Q Exactive Plus

orbitrap mass spectrometer (Thermo Scientific) with a spray voltage of 2.5 kV and a capillary temperature of 250 ^{o}C. In the data-dependent mode, the 15 most-abundant ions were submitted to higher energy collision dissociation (HCD) with an isolation window of *m/z* 1.5 and a normalized collisional energy of 30%. Resolving power was 70,000 (at m/z 400) for MS1 and 17,500 (at *m/z* 400) for product ions in MS/MS.

5.3.6 FPOP data analysis

Identification of the unmodified peptides and assigned modifications were done by using Byologic (Protein Metrics) and further validated by manual inspection. Modification sites were identified based on MS/MS. Signal intensities of the unmodified peptide (I_u) and its modified species (I_{ox}) were integrated using Byologic (Protein Metrics) from the extracted ion chromatograms (XICs). The FPOP modification level of a peptide was calculated using the following equation: % modified = $I_{ox}/(I_{ox} + I_u)\times100$. Quantification of the modified species was based on the dominant FPOP products with oxygen addition/substitution(s) (+15.9949 Da, + 31.9898 Da, etc.).

5.4 Results and Discussion

5.4.1 Mapping by HDX

In our model system, benzyl (1-methyl-6-phenyl-4H-benzo[f][1,2,4]triazolo[4,3-a][1,4]diazepin-4-yl)carbamate (compound 1) (Figure 5.1A) is a BRD competitive inhibitor that possesses a binding cavity for acetyl-lysine recognition (Figure 5.1B).[18] The structure of bromodomain adopts a left-handed four-helix bundle topology. Overlapping the compound 1-bound BRD4 (bound) with the ligand-free protein (unbound) shows the two structures are nearly identical

(Figure 5.2C), indicating no significant structural rearrangement of the protein that is caused compound 1 binding.

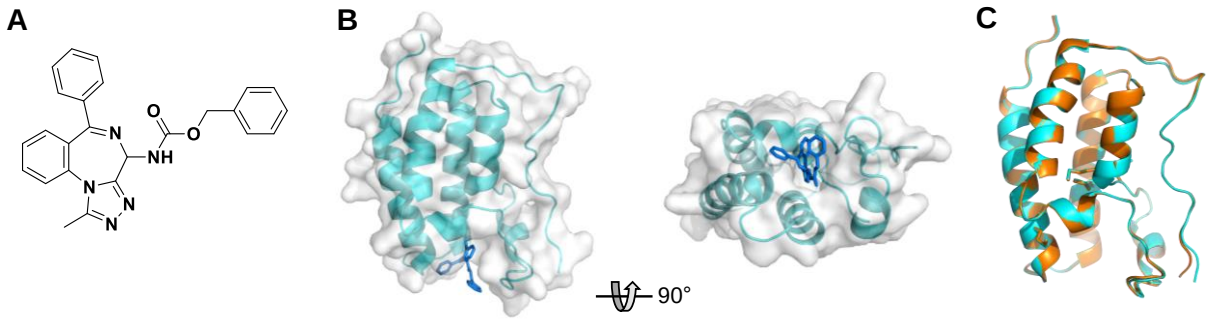


Figure 5.1. Binding of BRD4 to compound 1. (A) Molecular structure of compound 1. (B) Crystal structure of BRD4 in complex with compound 1 (PDB 2YEL), referred as "bound" hereafter. Secondary structures and surface presentation of the protein are shown in cyan and white, respectively. (C) Bound BRD4 in cyan is overlaid with the ligand-free BRD4 in orange (PDB 4LYI), referred as "unbound" hereafter.

In the HDX experiment, the hydrophobic core of BRD4 is expected to show very slow rate of deuterium uptake owing to its solvent accessibility. Thus, we extended the time course of D_2O exposure to 6 h to allow exchange to occur in the hydrophobic cavity. In accord with the nearly identical structures of BRD4 in the unbound and bound states, most regions in BRD4 shows no change in HDX kinetics upon compound 1 binding, as represented by peptides in Figure 5.2A-C. By way of contrast, we found some regions exhibit slightly reduced rates and extents of exchange in the bound state (Figure 5.2E-J). The differences in those regions, however, are still too modest compared to what we would typically expect for a binding interface.[19-21] Instead, HDX shows that BRD4 only undergoes a slight decrease in H-bonding and flexibility upon compound 1 binding, and we conclude that the overall structures of the unbound and bound

BRD4s are very similar. Indeed, we are seeking a binding scenario such as this to test whether protein footprinting is sensitive to small-molecule binding.

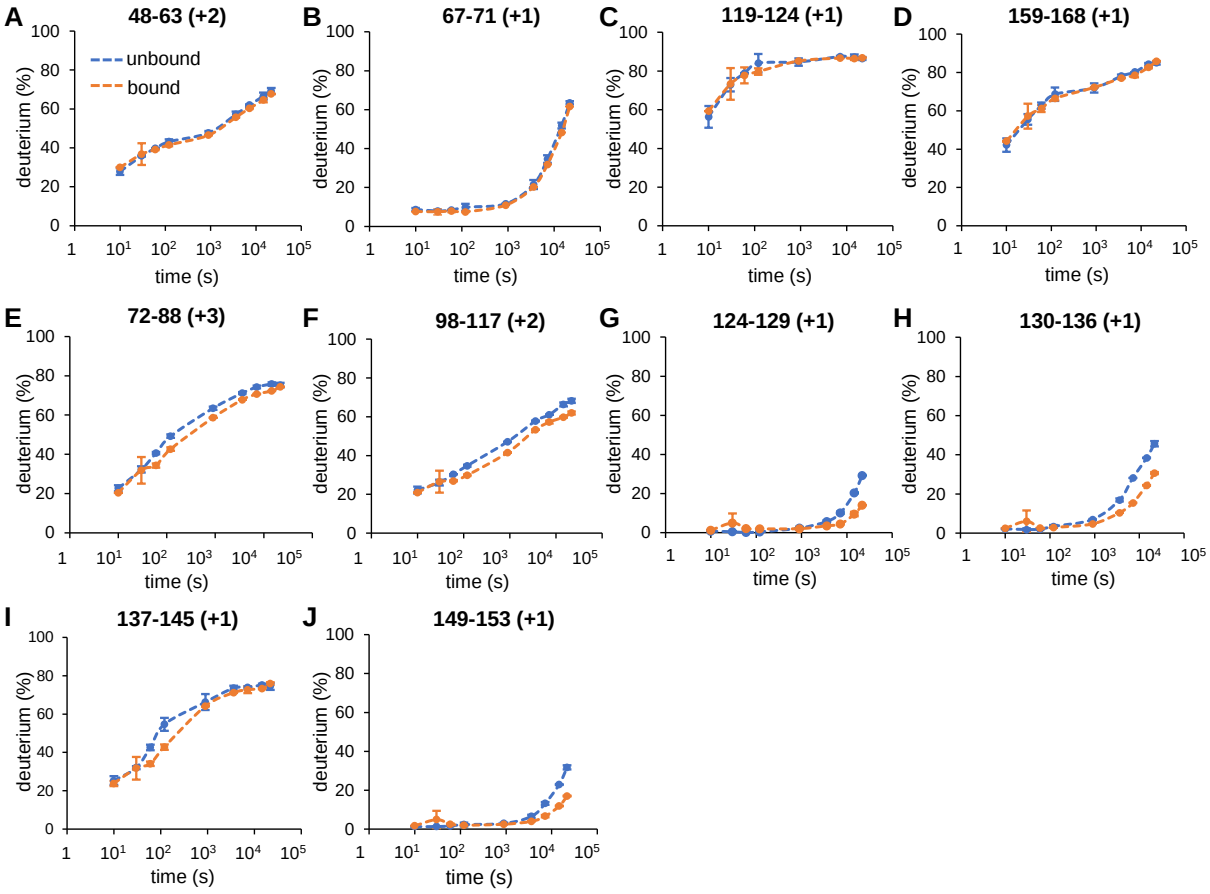


Figure 5.2. HDX kinetics of BRD4 upon binding of compound 1. Comparison between the HDX of unbound (blue) and bound (orange) BRD4 reveals representative peptide regions (A-D) that do not show changes in HDX upon binding, and regions (E-J) showing small changes upon binding to compound 1. Amino-acid numbering and charge state of the peptide from pepsin digestion of BRD4 are shown above each plot.

Given the small changes in HDX resulting from binding, we turned to FPOP to probe the interaction of BRD4 with compound 1. Compared to HDX that interrogates the structure and dynamics of backbone amide, we expect FPOP to show higher sensitivity in mapping the binding

cavity of BRD4 with the compound 1, because the binding interface may involve sidechain interactions and, moreover, many of the hydrophobic residues are reactive with the OH radicals used for footprinting.

5.4.2 Mapping by FPOP using a reporter peptide strategy

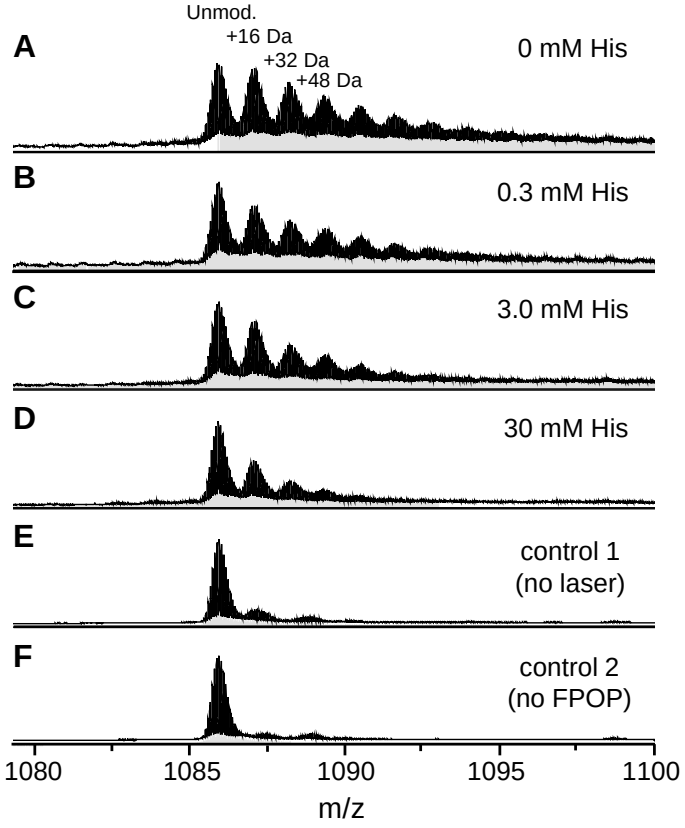


Figure 5.3. Global FPOP modification of BRD4 as a function of the scavenger concentration. Mass spectra of unbound BRD4 (+14 charged). FPOP was done without addition of the scavenger in the protein sample (A), and with 0.3, 3, 30 mM His used as scavenger (B-D). (E) Control sample of BRD4 (with addition of H_2O_2) submitted to the same FPOP workflow as the experimental group except the laser was turned off. (F) starting state of BRD4 in PBS buffer.

To obtain FPOP kinetic curves of the unbound and ligand-bound BRD4, we performed the labeling experiment with various concentrations (0, 0.3, 3, 30 mM) of histidine as the radical scavenger to give different timescales of labeling. Mass spectra of the intact BRD4 shows that the extent of FPOP modification gradually attenuates in respond to increased amount of the scavenger (Figure 5.3). To track the lifetime of the hydroxyl radical in the sample, we used Leu-enkephalin (YGGFL) as a peptide reporter for the hydroxyl radical[22] by mixing each protein sample with a fixed amount of Leu-enkephalin prior to FPOP. Upon laser irradiation, the reporter is competitively labeled with BRD4 by the hydroxyl radical under the same condition, and thus the reporter-% modified can be used to follow the hydroxyl radical lifetime in the sample.

Compound 1, a polyphenol molecule, is likely to be reactive with the hydroxyl radical, and if so this ligand will quench some of the radicals. Thus, there may be a different overall oxidation potential for the bound-sample solution compared to that containing the unbound. The modification of compound 1 will not compromise our efforts to locate the binding provided the count of radicals available to the protein can be normalized for the unbound and bound states. In fact, the "unintended" quenching effect of compound 1 is negligible under high concentrations of scavenger (3 and 30 mM His) as indicated by the equal modification extents of the reporter for the unbound and unbound samples when submitted to FPOP (Table 5.1). In absence of the scavenger (His = 0 mM), however, the reporter for the bound sample is significantly less modified than for the unbound BRD4 (relative difference 32%). In the presence of 0.3 mM His, a small variation in the reporter-% modified is observed. Those results demonstrate that in these FPOP experiments, the modification extent is well-controlled because the dominating chemistry

that controls the lifetime of the hydroxyl radical is the reaction of the radical with the large-excess scavenger rather than that with the protein or ligand.

His (mM)	Averaged % modified ± S.D. of the reporter	
	Unbound hBRD4	Bound hBRD4
0	22.40 ± 0.12	15.31 ± 0.53
0.3	14.92 ± 0.05	12.40 ± 0.11
3	6.69 ± 0.01	6.60 ± 0.36
30	1.85 ± 0.08	1.88 ± 0.02

Table 5.1. FPOP modification levels of the reporter peptide. Extent of modification of the reporter in the unbound and bound BRD4 sample. Scavenger concentrations of His are 0, 0.3, 3 and 30 mM respectively. FPOP labeling was done in duplicate.

To obtain comprehensive coverage across the BRD4 sequence in the bottom-up proteomics strategy used for analysis, we performed digestion with trypsin and chymotrypsin separately. Plotting the % modified of BRD4 peptide as a function of the reporter-% modified provides a "time-course" for the FPOP modification. The curves are similar to the HDX kinetic curves characterized by deuteration extents at different deuterium exposure times (Figure 5.4), curves that are commonly viewed in the field.

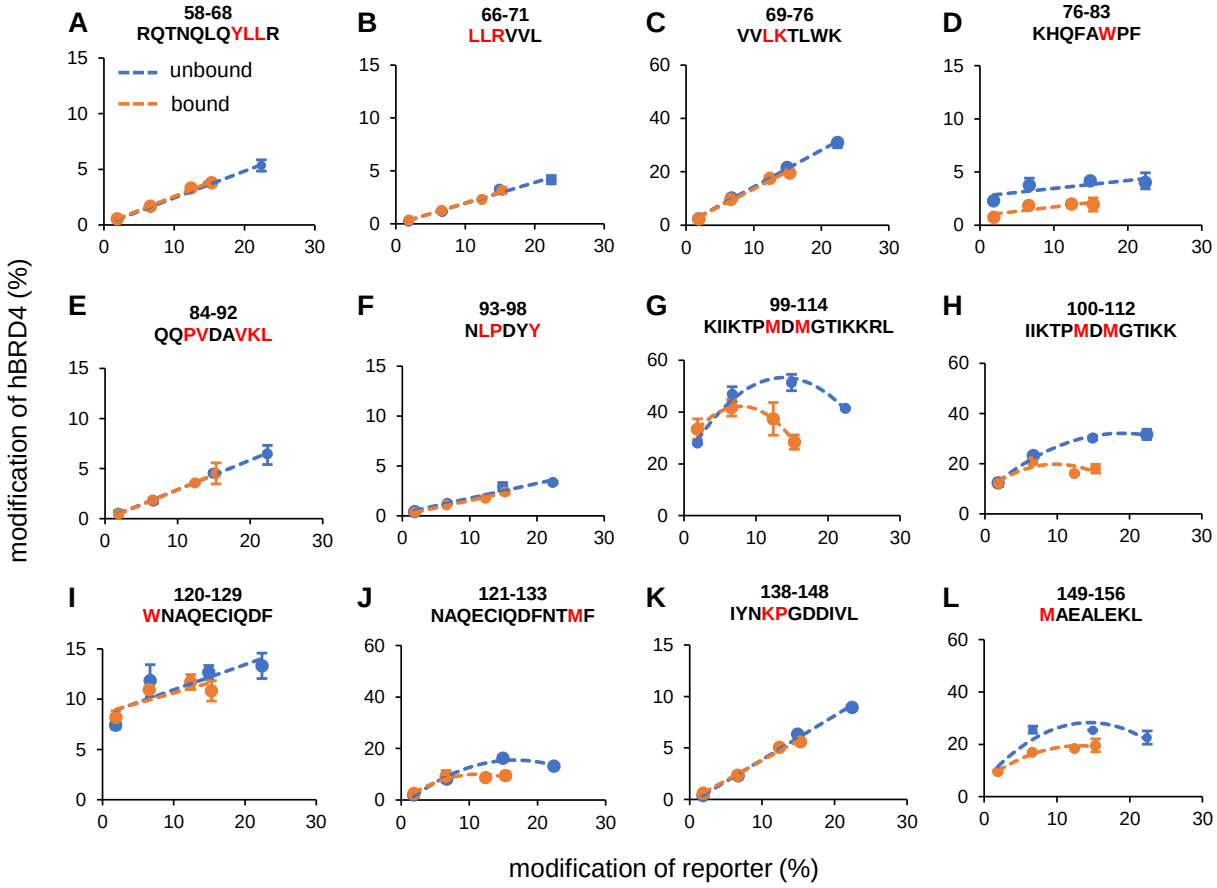


Figure 5.4. FPOP kinetic curves of BRD4. The x-axis of each curve is the % modified of the reporter peptide, which correlates with the lifetime of the hydroxyl radical. The y-axis is the % modified of the peptide from the FPOP-modified BRD4 in unbound (blue) and bound (orange) states, respectively. Numbering and sequence of the peptide from digestion of BRD4 are shown above each plot. Identified residues with FPOP modification are highlighted in red on the peptide sequence.

We find that for peptide regions not containing Met, the % modified for the protein shows a linear correlation with that of the reporter (Figure 5.4 A-F, I and K). Among those regions, 76-83 clearly exhibits reduced modification in the bound state. Therefore, this region is likely part of the binding site.

In contrast, FPOP modifications of peptide regions that contains one or more Met (e.g. 99-114, 100-112, 121-133 and 149-156) saturate and decrease as the lifetime of the hydroxyl radical becomes longer. (Figure 5.4 G, H, J and L). Because the modification on those regions occurs nearly exclusively on Met residue(s), we first hypothesized saturation and turnover of the modification extent are due to a switch of the Met oxidized products to higher oxidation states (e.g., + 32 Da, + 48 Da) as the labeling time increases (peptides with multiple oxidation are very rare, so they are generally not considered in the algorithm for peptide-level quantification). By manual search of the data, we found no detectable signals in LC-MS/MS for FPOP products in higher oxidation states (e.g., +32, +48). Although the reasons for such trends observed for Met-containing peptides is unclear, we are considering two explanations: 1) As the lifetime of the hydroxyl radical becomes longer, oxidation by FPOP on Met produces oxidized species that are not yet known. 2) Met-containing regions in the protein have affinity with H_2O_2, and thus the radicals are generated in the immediate vicinity of Met. In the latter case, the reactions of Met-containing regions with the hydroxyl radical is no longer limited by diffusion of the radical, but by the local concentration of H_2O_2 that surrounds the Met.

5.4.3 Correlating HDX and FPOP mapping with BRD4 structures

A co-crystal structure of the complex indicates that compound 1 binds to the acetyl-lysine recognition pocket of BRD4. Using proximity of hBRD4 residues with compound 1 as a criterion, we considered that residues of the protein within 4 Å from compound 1 could be key contacts. Using this criterion, we identified W81 (3.6 Å), P82 (3.9 Å), F83 (3.7 Å), L92 (3.7 Å) and N140 (3.3 Å) as key residues that directly interact with compound 1 from the crystal structure (Figure 5.5A).

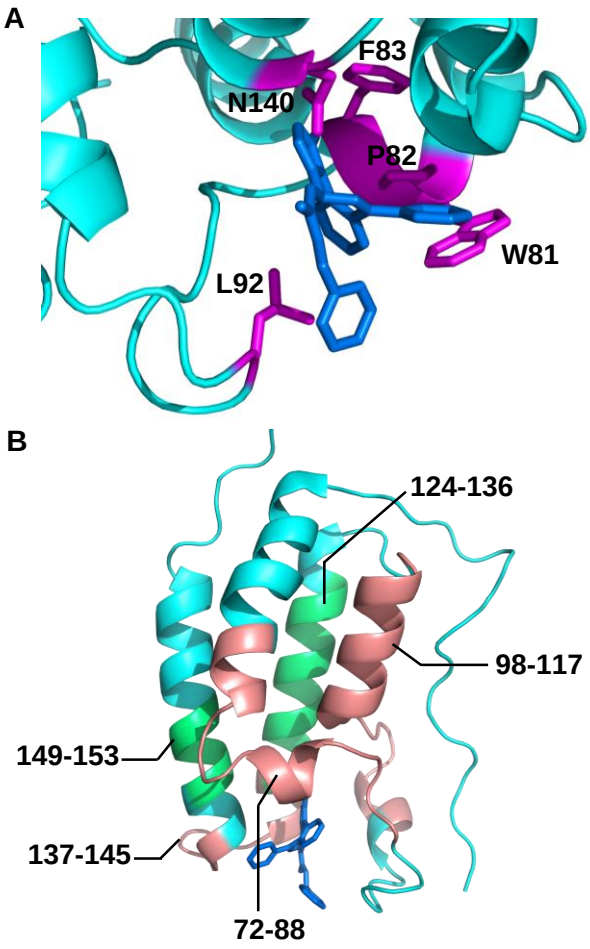


Figure 5.5. Correlation of the footprinting results with BRD4-compound 1 interaction. (A) Key contacts (magenta) of the protein (cyan) in contact with compound 1 (blue) according to the X-ray structure. The distance criterion for a contact was chosen to be 4 Å. (B) Peptide regions that showing reduced rates of HDX in the bound state (Figure 2 E-J) mapped onto the BRD4 structure (PDB 2YEL). Regions with low and intermediate extents of HDX (i.e., relative HDX of the peptide in the bound or unbound state instead of the difference in HDX between the two states) are colored in green and salmon, respectively.

5.4.3.1. HDX: Mapping the peptide regions that show reduced HDX in the bound (Figure 5.2E-J) onto BRD4-compund 1 complex indicates that those regions span the binding pockets and the

surrounding sequence (Figure 5.5B). As discussed earlier, the differences in HDX, however, are subtle and do not provide high confidence that a binding pocket has been identified. This may because 1) the protein only undergoes subtle structural change upon compound 1 binding, as indicated by the nearly identical crystal structures of the unbound and bound protein, 2) the size of the ligand is too small to provide efficient shielding from the solvent at the binding interface, and 3) the binding is not significant involved with peptide bonds but rather with side-chain interactions.

Despite these uncertainties, the kinetics of HDX provides information on the secondary structure elements of the region based on their effects on HDX. For example, regions 124-136 and 149-153 that are comprised of a two α-helix bundles that are structurally rigid (Figure 5B) resist solvent contact and show low extents of HDX (Figure 5.2 G, H and J). By contrast, region 119-124 is a loop located on the surface of the protein. Its flexible structure and solvent exposure allow the region to uptake rapidly deuterium and undergo extensive HDX even at short times (Figure 5.2C). Regions represented by peptides 72-88 and 137-145 cover both fast and slow exchangers and show intermediate rates of HDX (Figure 5.2E and I).

5.4.3.2. FPOP: Differential labeling by FPOP shows that region 76-83, containing key contacts W81, P82 and F83 is decreased for the bound state (Figure 5.4D), clearly indicating a binding interface. Further residue-specific analysis of this region indicates the FPOP modification occurs only on W81, which is at the entrance of the binding pocket and accessible to the solvent (and H_2O_2) (Figure 5.2A). No modifications were detected on P82 and F83 presumably because the two residues are buried deep in the hydrophobicity pocket and have low solvent accessibility (Figure 5.5A). Actually, the three residues W81-P82-F83 constitute the conserved WPF shell motif in the bromodomain family; this domain plays an important role in positioning inhibitors

through hydrophobic van der Waals interactions.[23-24] Our finding confirms that role of W81 is critical for substrate recognition and binding.

In addition, for L92 and N140 that have also been identified to be in contact with compound 1 based on the crystal structure, regions covering the two residues (84-92 and 138-148) show overlapping FPOP kinetic curves in unbound and bound (Figure 5.4 E and K). Although modification was detected on L92, we found no difference in its modification levels between unbound and bound protein. This is likely due to the insufficient solvent protection from compound 1 and the structural flexibility of L92 in a dynamic loop (Figure 5.5A). Meanwhile, owning to the low intrinsic reactivity of Asn with the hydroxyl radical, N140 did not undergo modification in FPOP. Instead, modification was detected on K141 and P142 nearby (Figure 5.4K). In this case, we lack information on N140 because the residue is not reactive with the hydroxyl radical probe.

Interestingly, all the covered Met residues of the protein (e.g., M105, M107, M132 and M149) show reduced rates of FPOP modification in the bound state (Figure 5.4G, H, J and L). According to the x-ray structure, there is no significant alteration in their conformation or side chain orientation upon compound 1 binding (Figure 5.1C). Thus, we indicate that the difference in FPOP labeling associated with those Met residues should NOT be attributed to a major change in protein structure or protection caused by compound 1 binding. Instead, given the similar unbound and bound structures of BRD4, the decreased modifications on Met in the ligand-bound state are more likely to result from the increased stability of the protein upon compound 1 binding, as revealed by reduced HDX in regions containing M105, M107, M132 and M149 (Figure 5.2F, H and J). Those changes in protein dynamics are not necessarily reflected by crystal structures that capture only a single low-energy state of the protein.

Furthermore, previous studies showed that Met tends to exhibit large change in FPOP modification in response to a minor structural perturbation, presumably owing to its high intrinsic reactivity with the hydroxyl radical.[14, 25] Therefore, caution is needed when interpreting changes in modification extents on Met and associating it with a major structural effect. Furthermore, although FPOP can potentially result in a "false positive" in suggesting a conformational change associated with Met, this problem can be avoided by performing the experiment under a high concentration of the scavenger (e.g., ~ mM His), because at short labeling times the modification extents on the Met for unbound and bound are more modest and, importantly, are identical (Figure 5.4G, H, J and L). More research is required to understand this effect of Met on FPOP yields.

5.5 Conclusions

In this chapter, we applied two most commonly used protein footprinting methods, HDX and hydroxyl radical footprinting to interrogate the epitope of BRD4 for a small molecular inhibitor. Results from HDX show very similar kinetics for the unbound and compound-1 bound BRD4, indicating the sensitivity of the method is not sufficient in this case to map directly the intermolecular interface between the protein and small molecule ligand presumably because the shielding of solvent by the small molecule is ineffective. In FPOP, the region containing the conserved WPF shell and residues nearby show a clear decrease in FPOP modification upon compound 1 binding, consistent with the predicted interactions at the binding interface determined by crystallography. Although there may be other residues in the vicinity of the binding, they do not variations in their solvent accessibility reported by FPOP.

In addition, regions containing Met residues display distinct chemical kinetics in FPOP, including saturation in the modification at extended labeling time and a relatively large difference in modification as an effect of a modest change in protein dynamics. Those differences seem to be false positives and require further evaluation.

Although the interactions of the protein BRD4 with the small molecule (compound 1) are difficult to decipher by FPOP and particularly by HDX, the small differences nevertheless provide valuable constraints of how the small molecule is poised with respect to the protein. Those constrains can be used to guide molecular dynamics simulation to pinpoint the binding site especially in the early stages of research or when a high-resolution structure from X-ray/NMR cannot be obtained.

In addition to providing insights into protein structure and properties, this chapter, together with Chapter 4 gives us a better understanding in the sensitivity, spatial resolution, applicability and limitation of each protein footprinting method for protein structure characterization.

5.6 Acknowledgements

We thank Elizabeth Bergman for her effort in the HDX-MS experiment and Bristol Myers Squibb for providing BRD4 and compound 1 used in this study. This research was supported by the NIH (Grant P41GM103422) and by Bristol-Myers Squibb.

5.7 References

1. Cohen, Y. Small Molecules: The Silent Majority of Pharmaceutical Pipelines. https://www.xconomy.com/boston/2015/11/23/small-molecules-the-silent-majority-of-pharmaceutical-pipelines/.

2. Masson, G. R.; Jenkins, M. L.; Burke, J. E., An overview of hydrogen deuterium exchange mass spectrometry (HDX-MS) in drug discovery. *Expert Opinion on Drug Discovery* **2017,** *12* (10), 981-994.

3. Pirrone, G. F.; Iacob, R. E.; Engen, J. R., Applications of Hydrogen/Deuterium Exchange MS from 2012 to 2014. *Anal. Chem.* **2015,** *87* (1), 99-118.

4. Engen, J. R., Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal. Chem.* **2009,** *81* (19), 7870-7875.

5. Hernychova, L.; Man, P.; Verma, C.; Nicholson, J.; Sharma, C. A.; Ruckova, E.; Teo, J. Y.; Ball, K.; Vojtesek, B.; Hupp, T. R., Identification of a second Nutlin-3 responsive interaction site in the N-terminal domain of MDM2 using hydrogen/deuterium exchange mass spectrometry. *Proteomics* **2013,** *13* (16), 2512-25.

6. Wang, H.; Rempel, D. L.; Giblin, D.; Frieden, C.; Gross, M. L., Peptide-Level Interactions between Proteins and Small-Molecule Drug Candidates by Two Hydrogen−Deuterium Exchange MS-Based Methods: The Example of Apolipoprotein E3. *Anal. Chem.* **2017,** *89* (20), 10687-10695.

7. Xu, G.; Chance, M. R., Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **2007,** *107* (8), 3514-3543.

8. Hambly, D. M.; Gross, M. L., Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J Am Soc Mass Spectrom* **2005,** *16* (12), 2057-63.

9. Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L., Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* **2009,** *81* (16), 6563-6571.

10. Li, K. S.; Rempel, D. L.; Gross, M. L., Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1–42 Aggregation. *J. Am. Chem. Soc.* **2016,** *138* (37), 12090-12098.

11. Chen, J.; Rempel, D. L.; Gau, B. C.; Gross, M. L., Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level. *J. Am. Chem. Soc.* **2012,** *134* (45), 18724-31.

12. Zhang, H.; Gau, B. C.; Jones, L. M.; Vidavsky, I.; Gross, M. L., Fast Photochemical Oxidation of Proteins (FPOP) for Comparing Structures of Protein/Ligand Complexes: The Calmodulin-peptide Model System. *Anal. Chem.* **2011,** *83* (1), 311-318.

13. Zhang, Y.; Wecksler, A. T.; Molina, P.; Deperalta, G.; Gross, M. L., Mapping the Binding Interface of VEGF and a Monoclonal Antibody Fab-1 Fragment with Fast Photochemical Oxidation of Proteins (FPOP) and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, 1-9.

14. Li, K. S.; Chen, G.; Mo, J.; Huang, R. Y. C.; Deyanova, E. G.; Beno, B. R.; O'Neil, S. R.; Tymiak, A. A.; Gross, M. L., Orthogonal Mass Spectrometry-Based Footprinting for Epitope Mapping and Structural Characterization: The IL-6 Receptor upon Binding of Protein Therapeutics. *Anal. Chem.* **2017,** *89* (14), 7742-7749.

15. Yan, Y.; Chen, G.; Wei, H.; Huang, R. Y.-C.; Mo, J.; Rempel, D. L.; Tymiak, A. A.; Gross, M. L., Fast Photochemical Oxidation of Proteins (FPOP) Maps the Epitope of EGFR Binding to Adnectin. *J. Am. Soc. Mass Spectrom.* **2014,** *25* (12), 2084-2092.

16. Jones, L. M.; B. Sperry, J.; A. Carroll, J.; Gross, M. L., Fast Photochemical Oxidation of Proteins for Epitope Mapping. *Anal. Chem.* **2011,** *83* (20), 7657-7661.

17. Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S., Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **2012,** *149* (1), 214-31.

18. Chung, C. W.; Coste, H.; White, J. H.; Mirguet, O.; Wilde, J.; Gosmini, R. L.; Delves, C.; Magny, S. M.; Woodward, R.; Hughes, S. A.; Boursier, E. V.; Flynn, H.; Bouillot, A. M.; Bamborough, P.; Brusq, J. M.; Gellibert, F. J.; Jones, E. J.; Riou, A. M.; Homes, P.; Martin, S. L.; Uings, I. J.; Toum, J.; Clement, C. A.; Boullay, A. B.; Grimley, R. L.; Blandel, F. M.; Prinjha, R. K.; Lee, K.; Kirilovsky, J.; Nicodeme, E., Discovery and characterization of small molecule inhibitors of the BET family bromodomains. *J. Med. Chem.* **2011,** *54* (11), 3827-38.

19. Malito, E.; Faleri, A.; Lo Surdo, P.; Veggi, D.; Maruggi, G.; Grassi, E.; Cartocci, E.; Bertoldi, I.; Genovese, A.; Santini, L.; Romagnoli, G.; Borgogni, E.; Brier, S.; Lo Passo, C.; Domina, M.; Castellino, F.; Felici, F.; van der Veen, S.; Johnson, S.; Lea, S. M.; Tang, C. M.; Pizza, M.; Savino, S.; Norais, N.; Rappuoli, R.; Bottomley, M. J.; Masignani, V., Defining a protective epitope on factor H binding protein, a key meningococcal virulence factor and vaccine antigen. *Proc Natl Acad Sci U S A* **2013,** *110* (9), 3304-9.

20. Zhang, J.; Adrián, F. J.; Jahnke, W.; Cowan-Jacob, S. W.; Li, A. G.; Iacob, R. E.; Sim, T.; Powers, J.; Dierks, C.; Sun, F.; Guo, G.-R.; Ding, Q.; Okram, B.; Choi, Y.; Wojciechowski, A.; Deng, X.; Liu, G.; Fendrich, G.; Strauss, A.; Vajpai, N.; Grzesiek, S.; Tuntland, T.; Liu, Y.; Bursulaya, B.; Azam, M.; Manley, P. W.; Engen, J. R.; Daley, G. Q.; Warmuth, M.; Gray, N. S., Targeting Bcr–Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* **2010,** *463*, 501.

21. Brudler, R.; Gessner, C. R.; Li, S.; Tyndall, S.; Getzoff, E. D.; Woods, V. L., Jr., PAS domain allostery and light-induced conformational changes in photoactive yellow protein upon I2 intermediate formation, probed with enhanced hydrogen/deuterium exchange mass spectrometry. *J. Mol. Biol.* **2006,** *363* (1), 148-60.

22. Niu, B.; Mackness, B. C.; Rempel, D. L.; Zhang, H.; Cui, W.; Matthews, C. R.; Zitzewitz, J. A.; Gross, M. L., Incorporation of a Reporter Peptide in FPOP Compensates for Adventitious Scavengers and Permits Time-Dependent Measurements. *J. Am. Soc. Mass Spectrom.* **2017,** *28* (2), 389-392.

23. Flynn, E. M.; Huang, Oscar W.; Poy, F.; Oppikofer, M.; Bellon, Steve F.; Tang, Y.; Cochran, Andrea G., A Subset of Human Bromodomains Recognizes Butyryllysine and Crotonyllysine Histone Peptide Modifications. *Structure* **2015,** *23* (10), 1801-1814.

24. Ember, S. W. J.; Zhu, J.-Y.; Olesen, S. H.; Martin, M. P.; Becker, A.; Berndt, N.; Georg, G. I.; Schönbrunn, E., Acetyl-lysine Binding Site of Bromodomain-Containing Protein 4 (BRD4) Interacts with Diverse Kinase Inhibitors. *ACS Chemical Biology* **2014,** *9* (5), 1160-1171.

25. Li, Z.; Moniz, H.; Wang, S.; Ramiah, A.; Zhang, F.; Moremen, K. W.; Linhardt, R. J.; Sharp, J. S., High Structural Resolution Hydroxyl Radical Protein Footprinting Reveals an

Extended Robo1-Heparin Binding Interface. *The Journal of Biological Chemistry* **2015,** *290* (17), 10729-10740.

Chapter 6: Identify Key Salt Bridges Controlling the Gating of Major Facilitator Superfamily Transporters Using Mass Spectrometry-Based Live-Cell Footprinting

6.1 Abstract

The major facilitator superfamily (MFS) transporters move a wide range of substrates across biological membranes. They function through an alternating access motion, during which the protein progresses through a series of conformational states to accomplish substrate translocation. Those conformational changes are mainly orchestrated by salt-bridge interactions between the N- and C-terminal domains of the transporter. Even with a structure from X-ray crystallography, however, identification of the key salt bridges of MSF transporters remains challenging. Here, we report the development of a MS-based live-cell footprinting approach, combined with suspension cell expression, green florescence protein (GFP) tagging and optimized in-gel proteolysis to probe the structures of MFS transporters in their native cellular environment. As a proof of concept, we applied carboxyl group footprinting to probe the formation and breakage of the salt bridges of human glucose transporters (GLUTs) during the transportation cycle. Substrates/inhibitors stabilizing the inward- or outward-open conformations generate distinct modification patterns, revealing the key salt-bridge interactions that control the gating of those transporters. We identified key salt bridge interactions in GLUT1 and GLUT5 on both its intracellular and extracellular sides. Remarkably, mutations disrupting the salt bridges located at the two opposite sides of membrane have opposite effect to the transporter activity. Mutants (at the cytosolic side) disputing the outward-facing conformation are inactive, suggesting that this conformation need to be stabilized. Conversely, mutants (at the extracellular side) disrupting the inward-facing conformation are hyperactive, suggesting that this alternative conformation needs to be maintained transient, because removal of the salt bridges normally stabilizing this state facilitates the transporter motion. Thus, the rocker-switch motion is asymmetrical in time, a mechanism probably adopted by most MFS transporters. This structural mechanism has evaded

extensive structural investigation of MFS transporters for decades, demonstrating again the powerful capability of our method to investigate the conformations, interaction and dynamic motion of membrane proteins in live cells.

6.2 Introduction

The major facilitator superfamily (MFS) is the largest group of secondary active transporters found in nearly all living organisms. This superfamily represents about one quarter of known membrane transport proteins,[1-3] with 85 subfamilies being identified to date (http://www.tcdb.org).[4] Each functionally characterized subfamily transports a certain set of substrates, including ions, sugars, nucleotides, amino acids, peptides, vitamins, and drugs. This broad range of substrates are transported across the membrane via structural motions of MFS proteins.

Crystal structures and functional analysis show that MFS proteins mediate substrate transport by an alternating-access mechanism. These transporters share a canonical MFS fold consisting of 12 transmembrane helices (TM); the N- and C-terminal halves of the proteins, each containing 6 TMs, form two pseudo symmetrical domains surrounding a translocation pore. The substrate is bound at a single site in this pore, and is transported through a rocker-switch rotation of the two protein halves that alternatively exposes the substrate to promote its binding or release at each side of the membrane. Thus, inward-open and outward-open occur as two major conformations during the transport cycle, with substrate-occluded conformations as intermediate states.

Formation and breakage of critical salt bridges controls the rocker-switch rotation and the gating of the translocation pore.[5] In proton-coupled transporters, such as LacY[6-7], GlpT[5] and POT[8], the

protonation state of key residues and changes in their salt bridges were proposed to induce significant structural rearrangements. The salt bridges provide part of the driving force, or even act as the pivot, to flip the transporters between inward- and outward-facing conformations. In GLUT5,[9] a glucose uniporter, an outward-facing conformation is stabilized by interdomain salt bridges between Glu and Arg residues, which are strictly conserved as part of the sugar porter signature. These residues, however, in the inward-facing conformation are too far apart to form salt bridges. Conversely, disruption of a salt bridge by a Glu329Gln mutation in GLUT1 stabilizes its inward-facing conformation, permitting the determination of its crystal structure.[10]

MFS transporters still hold specific challenges in structural biology compared to soluble proteins. For example, the conformation of MFS transporters is sensitive to extraction, solubilization and purification. Obtaining the crystal structure of the MFS transporters, particularly for eukaryotes is primarily limited by their hydrophobic surface, intrinsic flexibility, and lack of stability. Despite advances in X-ray crystallography[11], identification of alternating salt bridges during the rocker-switch motion remains a challenge. Until 2013, none of the MFS proteins had structure captured in more than one conformational state. Today, structures of 15 MFS proteins are available, including a handful of eukaryotic proteins (mostly glucose transporters), among which GLUT5 is the only one determined in the two alternating conformations. Additionally, crystal structures capture only a single low-energy state, which heavily depends on the crystallization conditions, such as pH[12], mutation (e.g., conformationally restricted E329Q in hGLUT1)[10], presence of an antibody, and crystal packing, all of which may perturb the native state of the protein. Furthermore, the surrounding lipids in their native environment are likely to affect the function, kinetics, or even thermodynamics of the transporter. Owing to these limitations, one can turn to molecular-dynamics simulation to rationalize the formation and breakage of salt

bridges implicated in biochemical studies.[13] Direct evidence of such subtle changes, however, especially in a native cellular environment, remains enigmatic for MFS transporters.

Recently, MS-based methods are showing potential for membrane protein characterization. Membrane proteins and their complexes can be interrogated in gas-phase under non-denaturing condition by intact mass analysis (native MS). The method allows membrane protein complexes to be released from detergents, bicelles or nanodiscs after collisional activation in the gas phase for MS detection. The approach provides insights into the subunit composition and lipid interactions of membrane protein complexes.[14-15] Native MS can also be coupled with ion mobility to obtain information on membrane-protein conformation and dynamics.[16] As an alternative, MS-based "bottom-up" proteomics identifies the proteolytic peptides of a membrane protein.[17] This approach, used with chemical footprinting are usually performed on purified protein in buffer solution, as exemplified by Chapter 2-5. For study of membrane proteins, the protein is usually stabilized by detergent or nanodiscs in buffer for characterization. In this chapter, we develop a novel way to apply protein footprinting by using it to label live cells. This should give results on the structure of the membrane protein and its modifications in the native cellular environment with high spatial resolution detailed to sub regions or even amino-acid residues.

Because salt bridges are formed or broken between basic and acidic residues, the study and identification of exposed carboxyl side chains can be achieved by chemical reactions (footprinting) with a nucleophilic compound, glycyl ethyl ester (GEE). This carboxyl group footprinting approach has been applied and discussed in Chapter 4, and showed its simplicity to be performed as a bench-top method. This is advantageous for studying a complicated biological system like the cell. In the footprinting reaction, the small chemical probe (GEE) irreversibly

modifies Glu and Asp residues, made more quantitative with isotopic encoding.[18] Although the GEE method was originally developed for soluble proteins *in vitro*,[19] we are motivated to conduct this labeling under mild conditions that we find compatible with human cells. Here we describe an *in cellulo* GEE footprinting method using mass spectrometry (MS), with GFP-aided enrichment, to obtain nearly complete sequence coverage of these membrane proteins analysis. This enables us to identify key salt bridges in both GLUT1 and GLUT5, including some missed by crystal structures. Substrates and inhibitors stabilizing the inward- or outward-open conformations generate distinct modification patterns, indicating specific changes of salt bridges. The native cellular state of GLUT1 and GLUT5 appears to contain both conformations. The results demonstrate that GEE MS-based footprinting is a novel and effective approach to study alternating salt bridges, in living cells, that govern the transport of numerous MFS proteins.

6.3 Methods

6.3.1 Cell culture and virus transfection

GLUT1/GLUT5 knockdown HEK293 cells were cultured in MEM medium added with 3mg/L puromycin dihydrochloride from Streptomyces alboniger, 10% FBS, 1% Pen/Strep, 1% L-glutamine (fresh) and 1% non-essential amino acids. Cells were grown at 37 ^{o}C, 5% CO_2, followed by adding 5% (v/v) virus GLUT1/GLUT5 after the density of cells reached 90%. After 8 h incubation, 0.5% (v/v) Na-butyrate was added, and the cells were incubated for another 36 h at 30 ^{o}C in 5% CO_2. Cells were plated into 25 μg/mL PEI (in 150 mM NaCl)-treated 24 wells plate, with 400,000 cells in each well. Cells were then incubated for 24 h at 37 ^{o}C, 5% CO_2 for cells adhesion onto the bottom of the plate.

6.3.2 Live-cell membrane protein footprinting

For each sample, 250 mL suspension cells are required. Cell suspension was centrifuged at 200× g, 4 ^{o}C for 5 min resulting in a cell pellet. After removal of the supernatant, the pellet was resuspended in 10 mL culture media. In separate experiments for the substrate-bound hGLUT, *D*-glucose, *D*-maltose or cytochalasin B were added to the cell media to final concentrations of 200 mM, 200 mM and 100 μM, respectively. Cell suspensions were incubated at 37 ^{o}C for 30 min for substrate binding. Control sample (unbound) was prepared in the same manner as above except no substrate was added. A subsequent centrifugation at 200×g for 5 min resulted in a cell pellet. The pellet was washed with 20 mL ice-cold PBS (10 mM, pH 7.4) and resuspended in 3 mL PBS containing 0.025% (w/v) digitonin. To drive the binding equilibrium and maintain GLUT in its bound form, the substrate was again added to the PBS solution to the same concentrations as above.

For protein footprinting, 500 μL light and heavy GEE mixture (1:1 molar) stock solutions were added to the cells to a total concentration of 300 mM, followed by adding 300 μL EDC to give a final concentration of 50 mM. Cells were labeled (footprinted) for 20 min at 25 ^{o}C while the sample tube was rotated. The reaction was then terminated by adding 3 mL 1 M ammonium acetate to quench the GEE footprinting, and samples were incubated at 25 ^{o}C for 10 min. This quench step simultaneously diluted the concentration of digitonin and allowed the permeabilized plasma membrane to "heal".

6.3.3 Post-labeling protein purification

Immediately after footprinting, GLUT1/GLUT5 were solubilized with 1% (w/v) n-dodecyl β-D-maltoside (DDM) in PBS and incubated for 1 h at 4 ^{o}C. The mixture was centrifuged at 20000×g

for 10 min, and the supernatant containing the solubilized membrane protein was collected. Cobalt affinity resin (300 µL) was added to the supernatant and the mixture was incubated for 30 min at 4 ^{o}C. The cobalt resin bound with the target membrane protein was then loaded onto a pre-equilibrated chromatographic column and washed at 4 ^{o}C with a buffer containing 10 mL PBS and 0.04% (w/v) DDM. The resin was immediately subjected to reducing SDS-PAGE to enrich further the protein. The green florescence of the GFP-tagged GLUT was excited under UV irradiation to reveal the corresponding gel band of the GFP-tagged protein, which was excised for in-gel digestion.

6.3.4 In-gel digestion

Two sets of enzymatic digestions, with chymotrypsin and trypsin, were performed separately to yield complementary sets of peptides, affording high sequence coverage of the membrane protein in LC-MS/MS analysis. Digestion was performed as previously described[20] with modifications. Briefly, gel pieces were digested for 8 h at 37 ^{o}C with 0.02 µg/mL chymotrypsin (Promega) in a buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM $CaCl_2$ and 0.1% RapiGest SF surfactant (Waters), or with 0.025 µg/mL trypsin/Lys-C mix (Promega) in a PBS buffer (pH 7.4) and 0.1% RapiGest SF surfactant. The digestion was quenched by adding formic acid to a final concentration of 1% (by volume). The solutions of peptides released by digestion were directly used for MS and MS/MS bottom-up proteomic analysis.

6.3.5 Mass Spectrometry

Mass spectrometry analysis was performed on an Orbitrap Q Exactive plus mass spectrometer (Thermo Fisher Scientific) coupled to a Nano UltiMate 3000 Rapid Separation system (Thermo Fisher Scientific). The peptides were pre-concentrated on an Acclaim PepMap C18 column

(Thermo Scientific, 100 μm × 2cm, 5μm, 100 Å) and desalted for 10 min. Peptides were then loaded on to a custom-made column (75 μm × 200 mm) packed with reversed-phase C18 material (Symmetry, 100 Å,5 μm) and separated at a flow rate of 500 nL/min by a gradient from 2 to 60% B in 80 min. Solvent compositions for solutions in channels A and B were water with 0.1% (v/v) formic acid and 80% (v/v) acetonitrile with 0.1% (v/v) formic acid, respectively. Full mass scans were performed at a resolution of 70,000 after accumulation to an automated gain control (AGC) target value of 3×10^6 or a maximum injection time of 150 ms. In the data-dependent mode, the 15 most-abundant ions were isolated by using a quadrupole mass filter with an isolation window of 3 *m/z* and an isolation offset of 1 *m/z* and submitted to "higher energy collisional dissociation" (HCD) at a normalized collision energy (NCE) of 30% of the maximum. Dynamic exclusion was set to be 8 s for the selected ions.

6.3.6 Data Analysis in LC-MS/MS

Unmodified peptides and assigned modifications on the peptide were identified by Byonic (Protein Metrics) based on the accurate *m/z* of the peptide and the product-ion spectra (MS/MS) and further validated by manual inspection. Reaction of GEE with the side chain of Asp/Glu results in primary products with a mass shift of +85.0527 Da (light, $+C_4H_7NO$) or +88.0527 Da (heavy, $+C_2{}^{13}C_2H_7{}^{15}NO$). For some modifications, the ester bond in the primary products underwent hydrolysis under the acidic conditions, resulting in secondary products with a mass shift of +57.0214 Da (light, $+C_2H_3NO$) or +60.0251 Da (heavy, $+C^{13}CH_3{}^{15}NO$).

Peptide signals represented by the peak area from the extracted ion chromatogram (XIC) based on the accurate mass were integrated by Byologic (Protein Metrics). In the quantification, total signal (A_{total}) of a certain peptide includes the unmodified form (A_u), the primary (A_{pri}) and

secondary (A_{sec}) products on all modified D/E residues (namely x, y, … if more than one D/E present) in that peptide (i.e., $A_{total} = A_u + \Sigma (A_{pri,x} + A_{sec,x} + A_{pri,y} + A_{sec,y} + \ldots)$. For a certain residue *x*, signals of its primary (A_{pri}) and secondary (A_{sec}) products, each including the light and heavy forms were summed to count as the total modified signal of residue *x*. The fraction modified of residue *x* was then calculated by the following equation: fraction modified = $(A_{pri,x} + A_{sec,x})/A_{total}$ where A refers to an integrated signal intensity.

6.3.7 Substrate uptake assay

The medium for cell culture was first removed from the plate wells, followed by rinsing and resuspending the cells by using glucose/fructose-free HEPES (pH 7.3). Cells were starved for 30 min at 37 ºC. To monitor GLUT1 and its mutant uptake activity, 2-Deoxy-D-glucose (2DOG, cold) and 3H-2DOG (hot) mixture were added to final concentrations of 0.5 mM and 13 μCi/mL, respectively. To monitor the uptake activity of GLUT5 and its mutants, *D*-fructose (cold) and 3H-*D*-frucose (hot) were added to final concentrations of 0.5 mM and 13 μCi/mL, respectively. The plate was incubated for 4 and 2 min, respectively, before measuring the uptake levels of GLUT1 and GLUT5, respectively. The substrate uptake levels of GLUT1/GLUT5 were then measured at 0, 10 s, 20 s, 40 s, 1 min, 2 min, 4 min, 8 min. In the kinetic assays, a series of concentrations of 2DOG/D-fructose (0, 0.625, 1.25, 2.5, 5.10, 20, 40, 80 mM) were added to GLUT1/GLUT5. After incubation of the cells with the substrate, the plate was washed with cold PBS several times and gently knocked to remove the remaining solution. Triton X-100 (1%) in PBS was added to each well. The plate was shaken at room temperature for 20 min for cell lysis. The supernatant was transferred into scintillation-counter fluid for measuring the uptake activity of GLUT1/GLUT5.

6.4 Results and Discussion

6.4.1 Footprinting design of MFS transport in living cells

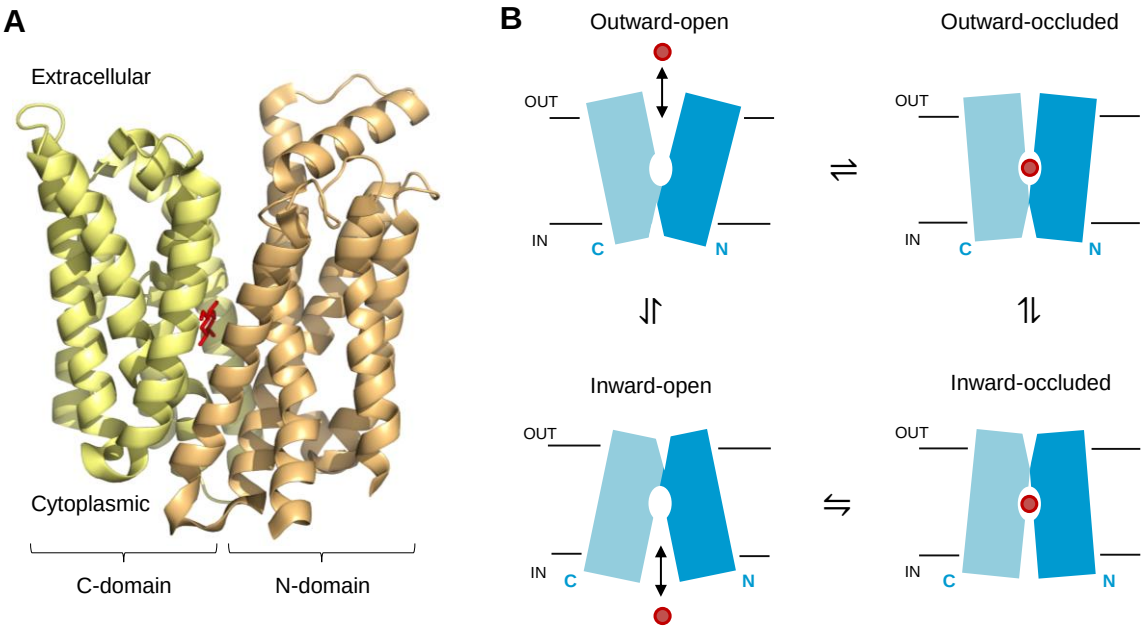


Figure 6.1. The alternating access model of MFS transporters. (A) The overall structure of MFS transporters, exemplified by human GLUT3 (PDB 4ZW9). The N- and C-domain TM bundles are colored in orange and yellow, respectively. The intracellular helical (ICH) domain on the cytoplasmic side is not shown for clarity. The substrate, *D*-glucose is bound in the central cavity formed by residues from both N- and C- domains, and shown by stick structure in red. (B) A schematic illustration of the alternating access (rocker-switch) model of MFS transporters. The motion and conformational change of the N- and C-domain of the protein translocate the substrate (red sphere) across the biomembranes (depicted by black lines).

The structures of MFS transporters are comprised of the N-domain and C-domain, each containing 6 TM helices, centered on a rotation axis that crosses the central substrate-binding site at the interface of the two half domains (Figure 6.1A).[11] Substrate translocation of MFS transport is mediated by an alternate access mechanism known as the 'rocker-switch' rotation. This model incorporates alternate movements of the N-domain and C-domain to create access on two sides of the membrane, allowing uptake of the substrate from one side of the membrane to the binding pocket and then release of the substrate to the other side of the membrane (Figure 6.1B).[2, 21] This alternative-accessing mechanism prevents simultaneous opening of both sides, preserving essential electrochemical gradients across biological membranes. During the transportation cycle enabled by alternating access, the inward-open and outward-open conformations occur as two major conformations, with the ligand-bound occluded conformations as intermediates. The inward-open and outward-open conformations are characterized by interactions, particularly of salt bridges, between the N-domain and C-domain of the transporter.[2, 5, 21] Formation and disruption of salt bridge upon substrate binding may play a role in controlling the conformational change required for substrate translocation in several MFS transporters, including Lac Y, GlpT etc.[5-6]

Choosing the human glucose transporters (hGLUTs) as models to evaluate the effectiveness of MFS transporter footprinting, we designed a MS-based living-cell footprinting approach to probe the structure of MFS transporters in the native cellar environment and to identify the key salt bridges controlling their rocker switch motion. The GLUT family, exemplified by GLUT1-5 belong to the solute carrier 2 family (SLC2). GLUT proteins translocate glucose and other monosaccharides including fructose, myo-inositol, and urate across the biological membrane in mammalian cells.[22] All GLUT proteins appear to share a similar structure, one that typically

contains 12 TM helices for MFS transporters. Recently, a number of crystal structures of GLUTs, including GLUT1 in the inward-open conformation[10, 23], GLUT3 in the outward-open and outward-occluded conformations[24], and GLUT5 in both outward-facing and inward-facing conformations[9] were determined. Those structures serve as valuable models for understanding MFS transporters as well as for bench-marking our protein footprinting results. Among eukaryotic MFS transporters, GLUT5 is the only protein whose structures have been reported in two alternating confirmations.

In our design, we carried out live-cell protein footprinting with MFS protein over-expressed in HEK293 cells (Figure 6.2). The MFS transporter (e.g., GLUT) was engineered to carry a green florescence protein (GFP) tag followed by a poly-His tag on the C-terminus of the protein. The outward- or inward-facing conformation of the protein in solution is favored by binding of a ligand (e.g., substrate or inhibitor), allowing us to interrogate the protein conformation in different functional states. Prior to chemical footprinting, we permeabilized the cell plasma membrane of by treating cells with a low concentration of digitonin. Digitonin is a nonionic detergent that reversibly permeabilizes the plasma membrane of mammalian cells without significantly affecting their functional integrity.[25] Permeabilization of cells allows the chemical reagents for protein footprinting to diffuse passively into the cell and modify the cytoplasmic portion of the membrane protein.

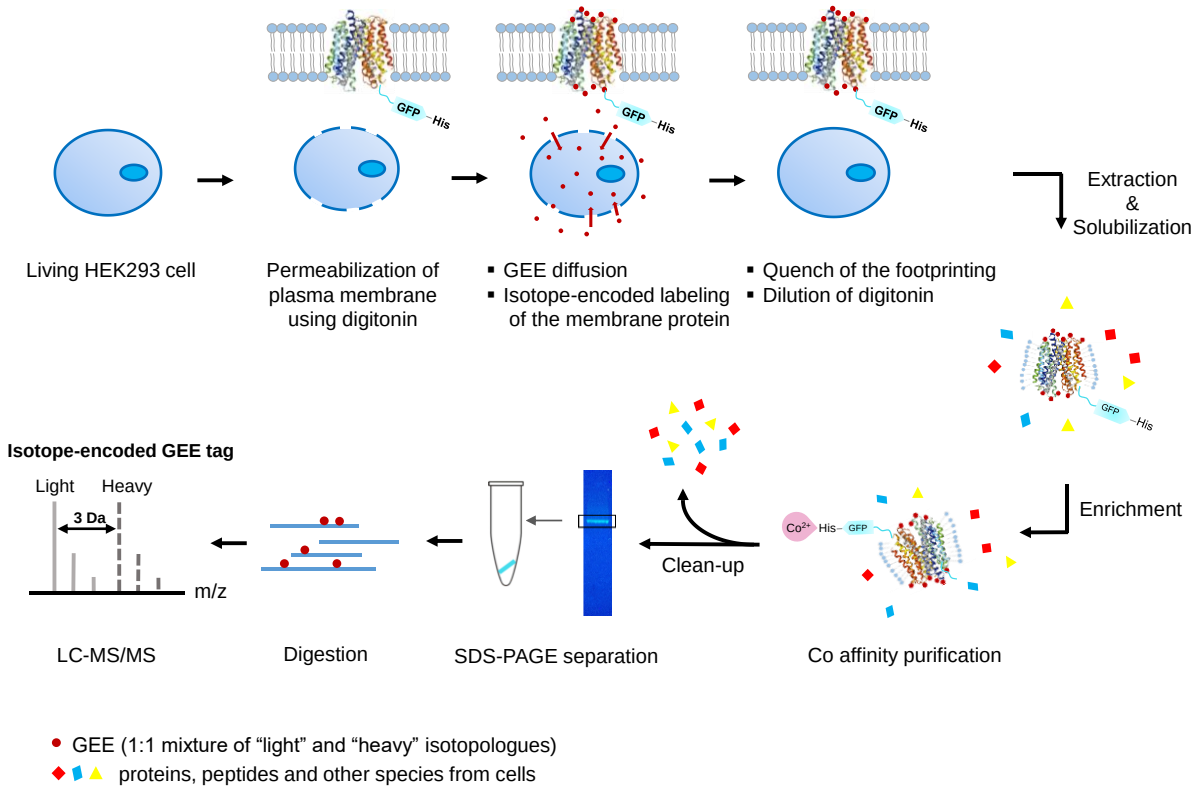
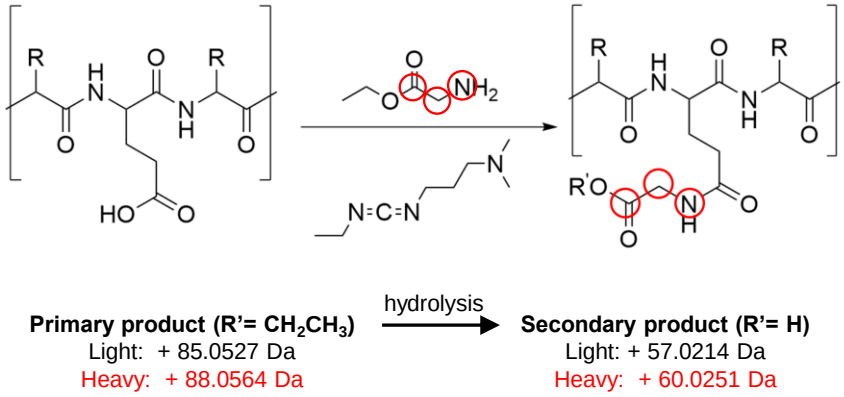


Figure 6.2. Schematics of the MS-based live-cell footprinting of membrane transporters. The MFS protein (e.g., GLUT) is over expressed in HEK293 cell. The protein is engineered to carry a GFP tag followed by a poly-His tag at the C-terminus. For protein footprinting, the plasma membrane of HEK293 cells is permeabilized by treating cells with digitonin, followed by modification of Asp/Glu side chain via the EDC-mediated GEE reaction using an isotopic mixture of a 1:1 molar ratio of "light" and "heavy" GEE. The protein after GEE footprinting is purified and enriched by Co-affinity and SDS-PAGE. Proteolytic peptides of the protein from in-gel digestion are analyzed by LC-MS/MS to quantify the modification extent of the labeled Asp/Glu residues. These extents are compared between protein samples in different functional states.

In protein footprinting, we submitted the cells to chemical modification with an isotopic mixture of "light" and "heavy" (1:1 molar) GEE (Scheme 6.1). GEE labels the acidic residues (Asp and Glu) that may involve a salt bridge. In the reaction, the carboxyl side chain of the acidic residue is first activated by forming an O-acylisourea intermediate with EDC. This intermediate quickly reacts with GEE to produce the primary labeling product with a chemical tag of + C_4H_7NO (+85.0527 Da for "light", or +88.0564 Da for "heavy").[26-27] In the downstream sample handling procedures, hydrolysis of the ester bond in the product may occur, which results in a secondary product with a chemical tag of + C_2H_3NO (+57.0214 Da for "light", or + 60.0251 Da for "heavy").



Scheme 6.1. EDC-mediated GEE carboxyl group footprinting. GEE reacts with Asp and Glu side chains to produce the primary products with a stable chemical tag. Some of the primary products undergo hydrolysis in the down-stream sample handling to produce the secondary products. Both primary and secondary products are encoded by the reaction with a 1:1 mixture of "light" and "heavy (C_{13}×2 and N_{15}×1)" GEE. The locations of the "heavy" isotopes are highlighted in red circles.

After quench of the reaction and dilution of digitonin, we extracted the membrane proteins from the plasma membrane of the cell into a buffer solution for purification and enrichment using Co-affinity and gel electrophoresis (SDS-PAGE) (Figure 6.2). Based on the fluorescence of its GFP tag, the target protein in the gel was immediately visualized by UV irradiation. We digested the proteins in the gel spots and submitted the proteolytic fragments to LC-MS analysis. Identification of the GEE-modified peptide is made as certain as possible by using the accurate mass of ions encoded with characteristic isotopic patterns. By comparing the modification extents of Asp/Glu in distinct functional states of the protein, we are able to determine the solvent accessibility changes regions containing Asp and Glu and map the motion of those residues.

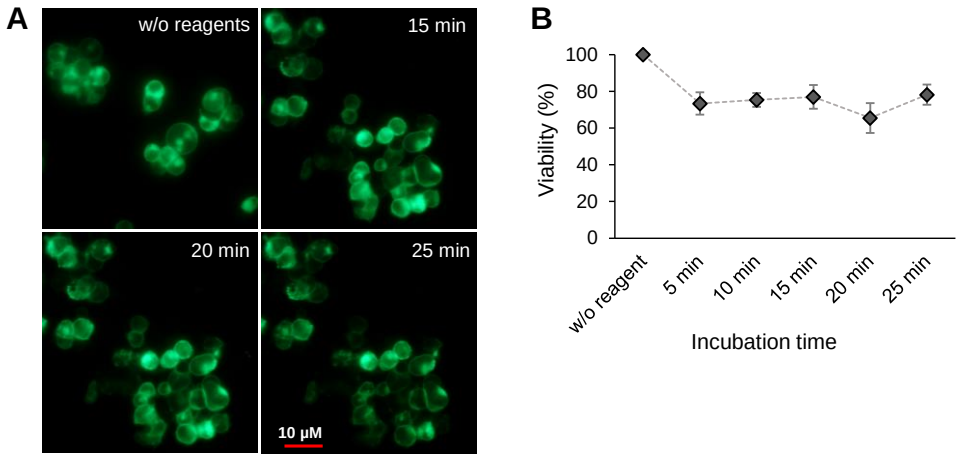


Figure 6.3. Morphology and viability of the cells exposed the carboxyl group footprinting reagents. (A) HEK293 cells exposed to 50 mM EDC and 300 mM GEE in PBS buffer. The cell morphology is depicted by the green fluorescence of the GFP-tagged GLUT on the plasma membrane. (B) Viability of HEK293 cells exposed to 50 mM EDC and 300 mM GEE in PBS buffer. Cell viability was normalized to the control sample in PBS buffer without the addition of EDC and GEE.

The chemical reagents, EDC and GEE are potentially toxic to cells as determined by their concentration and resident time in the cell. Thus, we investigated whether the integrity and viability of the cells are affected during protein footprinting. We treated the HEK293 cells with various concentration of the footprinting reagents, and we found the morphology of cells remain intact in 50 mM EDC and 300 mM GEE for up to 25 min (with observation of fluorescence photobleaching in the sample owing to long-term excitation) (Figure 6.3A). This length of time is sufficient for completing the GEE footprinting. In addition, the viability assay shows that despite a 25% decrease in viability owing to the addition of 50 mM EDC and 300 mM GEE, a majority (75%) of the cells remain alive for up to 25 min (Figure 6.3B). This ensures that the protein conformation and cell activity that we are probing are biologically relevant.

6.4.2 Proteomics strategy for high sequence coverage of MFS in structural analysis

"Bottom-up" proteomics analysis affords high structural resolution by providing details on sub-regions or even amino-acid residues of the protein. Difficulties in characterization of MFSs transporters, however, arise from their high hydrophobicity, a common problem for membrane proteins. MFS transporters have 12 α-helical TM domains, making them poorly soluble in aqueous solution and immune to proteolysis, thus limiting the sequence coverage in "bottom-up" proteomics.

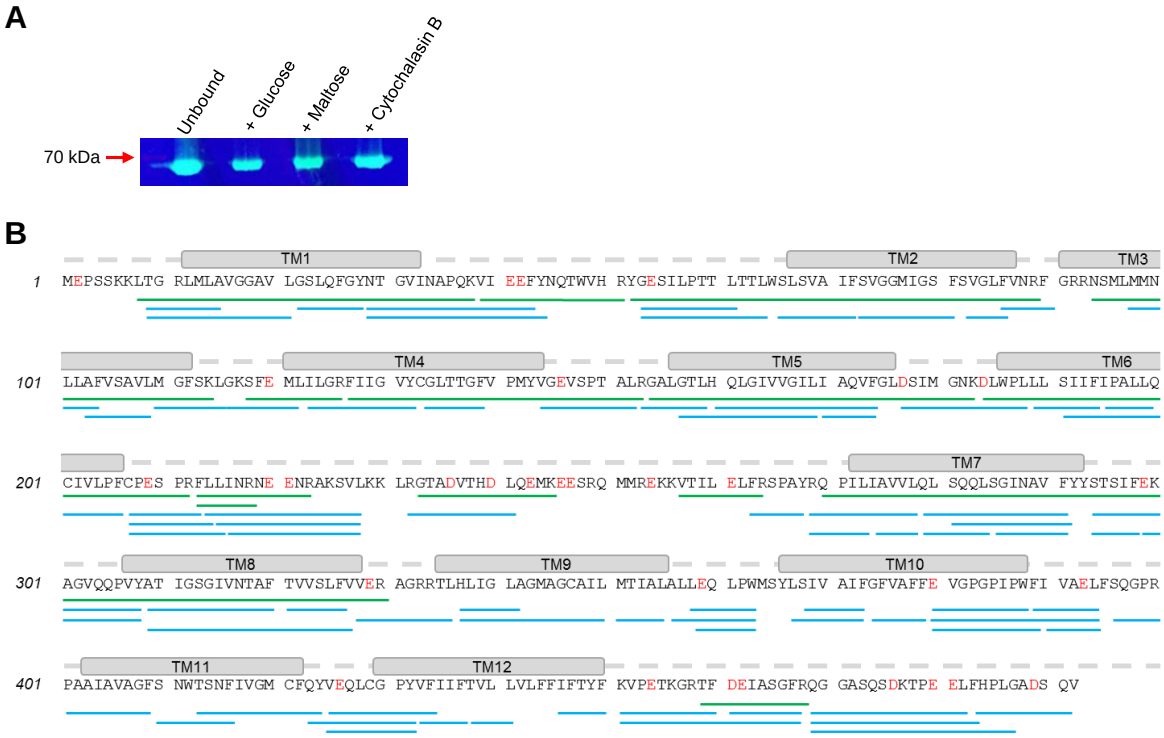


Figure 6.4. Post-labeling purification and proteolysis of hGLUT1. (A) SDS-PAGE separation shows a band near 70 kDa visualized by the inherent fluorescence of the GFP-tagged hGLUT1 under UV irradiation. (B) The tryptic (green bar) and chymotryptic (blue bar) peptides of hGLUT1 detected in LC-MS/MS. The twelve TM domains are indicated in the GLUT1 sequence above. Asp and Glu are in red on the sequence.

We used SDS–PAGE separation and multi-enzyme digestion to improve the sequence coverage of GLUTs in bottom-up proteomics. SDS-PAGE, in addition to Co-affinity purification, provides enrichment of the protein, and removes salts and detergents that interfere with downstream proteolysis and LC-MS analysis. Meanwhile, SDS-PAGE strongly denatures the membrane protein, enhancing its solubility and making the protein more susceptible to proteolysis. Taking

advantage of the GFP tag, we can directly visualize the gel band containing the target MFS transporter (Figure 6.4A).

In a previous LC-MS study, hydrophilic peptides of the membrane protein were not adequately covered.[28] The outcome usually suffers from a lack of cleavage sites (e.g., Arg or Lys for trypsin in the TM regions) and the tendency of hydrophobic peptides to aggregate in aqueous solution after removal of detergents. In our LC-MS analysis, we see some lengthy peptides (> 40 residues) from the TM regions of GLUTs. They are not selected for label-free quantification owing to their poor reproducibility in reversed-phase chromatography and/or the low signal intensity in MS and MS/MS. Proteolysis of the protein with chymotrypsin, which cleaves at Tyr, Phe, and Trp, significantly favors proteolysis in the TM regions due to the frequent presence of hydrophobic residues. Thus, proteolysis of GLUTs, as well as other membrane proteins using trypsin and chymotrypsin offers the advantage of creating overlapping peptides throughout the membrane protein sequence. We found that digestion with Lys-C/trypsin and chymotrypsin separately on the GLUTs generates properly-sized peptides that maximize the sequence coverage of hGLUT1 and hGLUT5 to 95% (Figure 6.4B) and 93%, respectively (Figure 6.5).

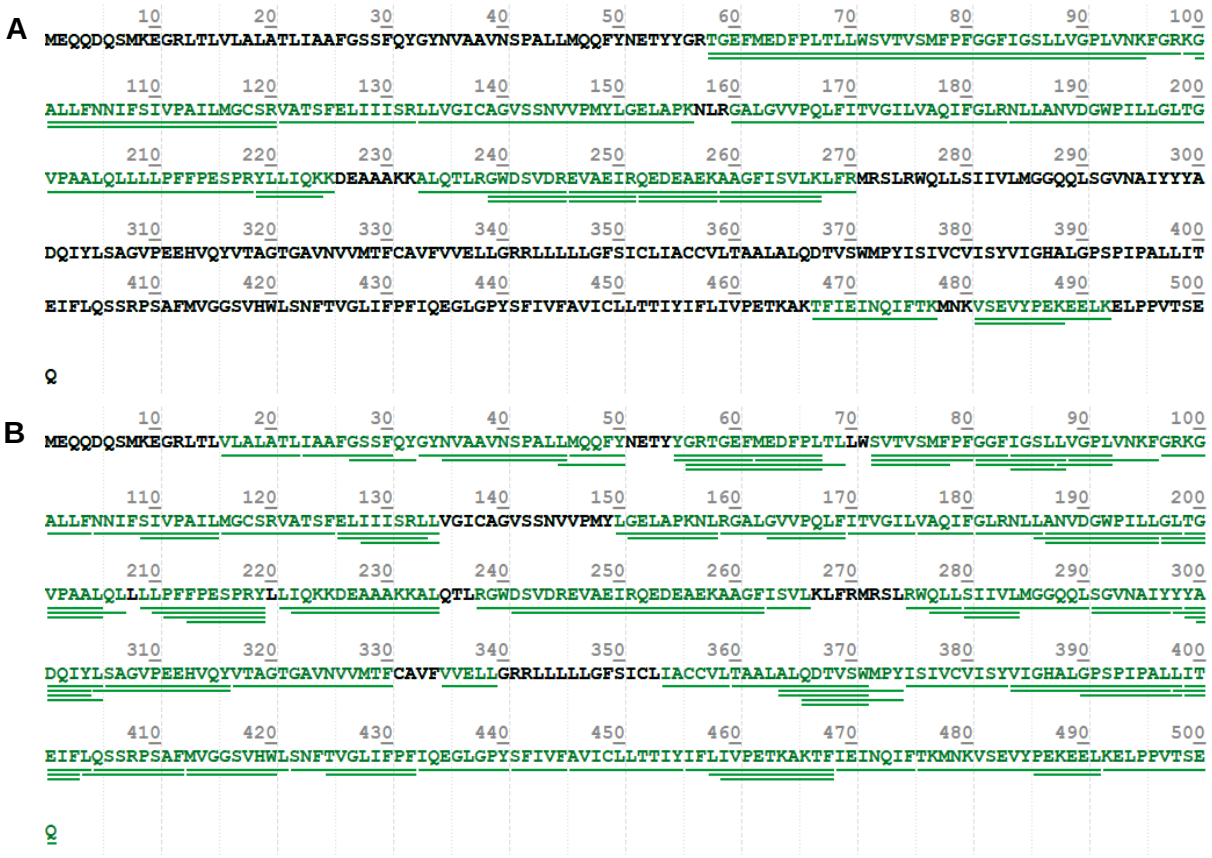


Figure 6.5. Sequence coverage of hGLUT5 achieved in LC-MS/MS analysis. Peptide fragments from digestion with (A) trypsin and (B) chymotrypsin.

6.4.3 Isotope-encoded MS identification of GEE-modified peptides

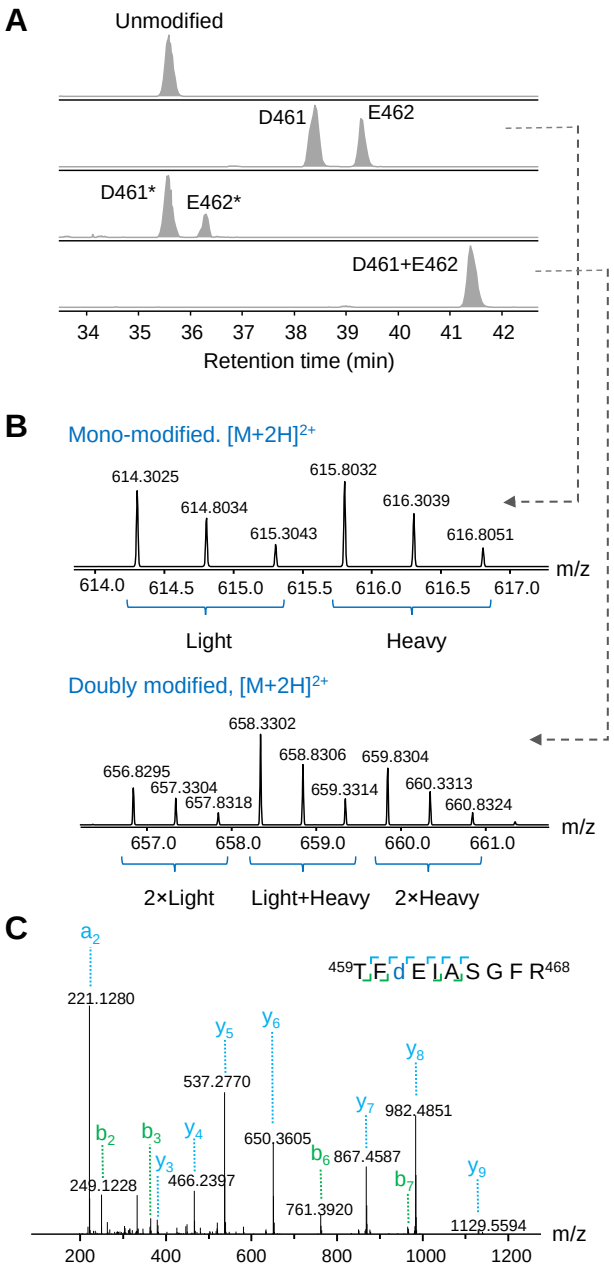


Figure 6.6. Representative LC-MS/MS analysis for GEE footprinting. The tryptic peptide ^{459}TFDEIASGFR^{468} of hGLUT1 is shown as an example. (A) Extracted ion chromatograms (EICs) of (from top to bottom) the unmodified peptide (*m/z* = 571.7774), mono-modified peptides with the GEE tag on either Asp^{461} or Glu^{462} (primary product of *m/z* = 614.3038;

secondary product (*) of m/z = 600.2882), and doubly-modified peptide with GEE tags on both Asp^{461} and Glu^{462} (primary product of *m/z* = 656.8302; the secondary product has low signal, comparable to the noise). All *m/z* values are for $[M + 2H]^{2+}$ ions, and the sites of modification were identified by MS/MS, and the fragment series is labeled above each signal peak. The "heavy" labeled peptides co-elute with their "light" equivalent and are not shown. (B) The mono-modified peptide (top) shows an isotopic doublet of 1:1 ratio in their intensities, and the doubly-modified peptide (bottom) shows an isotopic triplet with 1:2:1 ratio in their intensities. (C) The product-ion (MS/MS) spectrum of the precursor ion of *m/z* = 614.3038 indicates a mass shift of + 85.0527 Da on Asp^{461} that designates the GEE modification onto that site.

Labeling by GEE on Asp and Glu is covalent and stable (irreversible). Thus, the modification made on the protein in live cells is preserved in the downstream sample handling processes of purification, proteolysis and LC-MS analysis. This insures that the modification readouts from MS analysis report on the structure of the protein in its native cellular environment. In LC-MS analysis, a GEE-modified peptide is separated with the unmodified form (i.e., have different retention times (Figure 6.6 shows the LC-MS analysis of peptide region ^{459}TFDEIASGFR^{468} in hGLUT1 as an example). This region contains Asp^{461} and Glu^{462} as two potential acidic sites for GEE modification. We found the modified forms of the region include peptides with one GEE tag on both Asp^{461} and Glu^{462}, and a peptide with double GEE tags on both sites (Figure 6.6A).

Peptides of GLUT from the digestion contains a high content of hydrophobic peptides from the protein TM domains. Those hydrophobic peptides generally exhibit low solubility in the aqueous solution and poor ionization efficiency in electrospray ionization (ESI). Those problems lead to difficulties in identification of the GEE-modified peptide, because the modified forms of those

peptides are even less abundant. To identify unambiguously and to quantify accurately the GEE modification, we employed isotope encoding to enhance the delectability of the low-abundance modified peptide. Carboxyl group footprinting using a mixture of "light" and "heavy" GEE in 1:1 molar ratio produces labelling products that display isotope-encoded signals in MS.

6.4.4 Intracellular salt-bridge interaction stabilizes GLUT1 in the outward-facing conformation

GLUT1 specifically translocases glucose across biomembranes in mammals. We performed live-cell GEE footprinting on hGLUT1 in the ligand-free state as the control, and then in three ligand-bound states with *D*-glucose, maltose and cytochalasin B, respectively. Maltose, a condensation of two *D*-glucoses (Figure 6.7A), functions as an exofacial and competitive inhibitor for glucose transport by GLUTs.[29] The crystal structure of GLUT3 in complex with maltose shows that the one of the constituent glucoses is sufficient to occupy the glucose-binding pocket, causing the interactions between GLUT3 and maltose to be nearly identical to those of *D*-glucose.[24] Therefore, maltose binding, similar to that of glucose, favors the outward-facing conformation that allows substrate binding. In contrast, cytochalasin B (Figure 6.7A), a cell-permeable mycotoxin, binds to GLUT1 in the inward-open state and favors the inward-facing conformation.[23, 29]

LC-MS/MS analysis of the post-labeling hGLUT1 reveals that 28 of 31 Asp/Glu residues are covered in peptide mapping. GEE modification is detected on 19 Asp/Glu residues (Figure 6.7B, D and Figure 8). GEE modification dominantly occurs on residues occupying the periphery of the protein, because Asp and Glu, as polar residues, tend to appear on the protein surface in

contact with aqueous solution. The only acidic residue in the TM domains of hGLUT1 is E380 (TM10), and it is not modified.

In hGLUT1, we found two residues, E299 and E329, that show differential modifications in distinct functional states of hGLUT1 (Figure 6.7B and D). We identified from the crystal structure an intracellular salt bridge between E329 and R333, and an extracellular one between E299 and K38, both located near the translocation pore (Figure 6.7C and E). The outward-facing conformation, favored by glucose or maltose binding, is stabilized by inter-TM bundle salt bridging on the intracellular side involving E329 (TM9), R333 (TM9) on the C-domain and the backbone of G154 (TM5) of the N-domain (Figure 6.7C). Formation of the salt bridge results in decreased GEE modification consistent with increased protection on E329 (Figure 6.7B). As GLUT1 adopts the inward-facing conformation, initiated by cytochalasin B binding, the salt-bridge interactions on the intracellular side are disrupted, E329 becomes more solvent accessible, and it undergoes increased modification (Figure 6.7B). Meanwhile, the inward-facing conformation is stabilized by salt bridging on the extracellular side formed by K38 (TM1) from N-domain and E299 (TM7) from the C-domain (Figure 6.7E). Thus, the changes in modification of E299 on the extracellular side are opposite to those of E329 on the other side of the transporter (Figure 6.7D), verifying the alternating accessing motion of the membrane transporter.

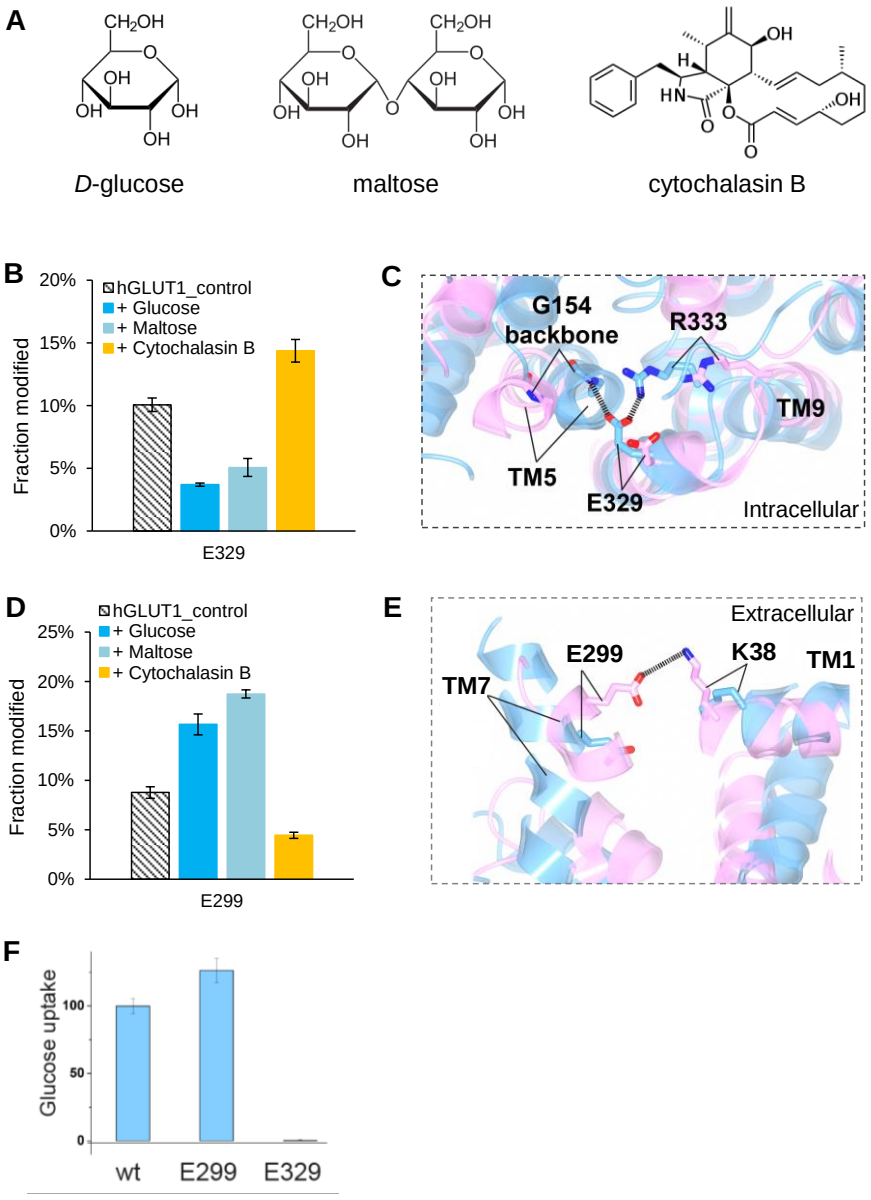


Figure 6.7. Salt bridge interactions in hGLUT1. (A) Chemical structures of the ligands for GLUT1. (B) GEE modification extents of E329, and (C) the associated salt bridge interactions on the intracellular side. (D) GEE modification extents of E299, and (F) the associated salt bridge interactions on the extracellular side. In (C) and (E), the structure of hGLUT1 in the inward- and outward-facing conformation is shown in pink and blue, respectively. (F) Effect of substitution of E299 or E329 to alanine on the glucose-uptake ability of hGLUT1.

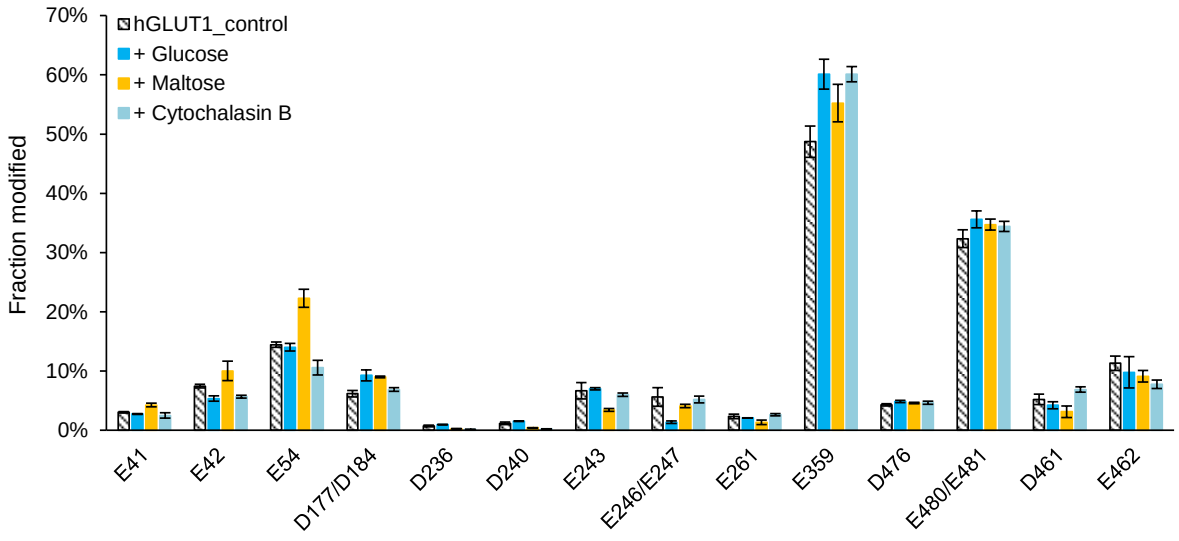


Figure 6.8. Asp/Glu residues in hGLUT1 showing no significant change in the level of GEE modification.

Examining the functional significance of the identified salt-bridge interactions (Figure 6.7F). we found that substitution of E329 to alanine almost inactivate the uptake of D-glucose. However, substitution of E299 slightly increases the uptake of D-glucose, all in comparison to WT. Thus, disruption of the intracellular salt bridge (with E329) weakens the interactions between the N- and C- domains and significantly impedes the transport cycle by disfavoring the outward-facing conformation for substrate binding. Conversely, disruption of the extracellular salt bridge (with E299) results in hyperactivity in substrate translocation, suggesting that during the transport cycle the inward-facing conformation needs to be maintained transient.

The intracellular salt bridge with E329 is seen by X-ray crystallography[10], as disruption of the salt bridge by substituting E329 with Gln arrests GLUT1 in an inward-facing conformation and hence facilitates crystallography. Actually, it was believed that GLUT1 lacks conserved salt bridge on the extracellular side. Nevertheless, we found a salt bridge with E299 on the

extracellular side of GLUT1 by probing the in-cell dynamic structure of the transporter using MS-based footprinting.

6.4.5 Apo GLUT1 contains outward- and inward-facing conformations in equilibrium

It was previously hypothesized that apo GLUT1 prefers an outward-facing conformation because of the extensive interactions between the TM and ICH domain.[30] Our live-cell footprinting results, however, show the modification extents of E299 and E329 in the ligand-free state (control) fall into a medium level between those of the outward- and inward-facing states. This suggests that apo GLUT1 in its the native cellular state contains both outward- and inward-facing conformations in equilibrium.

6.4.6 Intracellular salt bridges stabilize GLUT5 in the outward-facing conformation

GLUT5 is the only GLUT family member identified so far to translocate specifically fructose (Figure 6.9A). We applied the live-cell protein footprinting to study the ligand-free hGLUT5 and and its conformations in presence of glucose, fructose, and cytochalasin B. The GEE modification extents of hGLUT5 upon addition of glucose or cytochalasin B are nearly identical to that of the ligand-free protein. This suggests GLUT5, a fructose-specific transporter, interacts weakly if at all with glucose or cytochalasin B (Figure 6.9B, D and Figure 6.10).

By contrast, two acidic residues, D64 and E337, in the fructose-bound GLUT5 are modified differentially compared to those in the ligand-free protein. Similar to the effect of glucose or maltose on GLUT1, excess fructose favors the outward-facing conformation of GLUT5. X-ray crystal structures of GLUT5 are available for both outward- and inward-facing conformations of

rat (rGLUT5) and bovine (bGLUT5), respectively[9]. rGLUT5 and bGLUT5 share ~81% sequence identity to hGLUT5. On the basis of the rGLUT5 structure, we can conclude that for the outward-facing conformation of rGLUT5, E336 (TM8) forms salt bridges with both R158 (TM5) and R340 (TM9) on the intracellular side. Extrapolating to hGLUT5, we expect equivalent salt bridges to form between the conserved residues E337, R159 and R341. To facilitate structural analysis, we will use the hGLUT5 residue numbers to annotate the conserved residues in rGLUT5 (Figure 6.9C, left). The above hypothesis with hGLUT5 is further supported by a bioactivity assay of hGLUT5 mutants. Compared to the WT protein, single alanine substitutions of R159, E337 and E401 in hGLUT5 strongly reduce the uptake of fructose by ~ 70%, and mutation of R341 reduces the uptake by ~ 60%, indicating that salt bridging on the intracellular side of GLUT5 is essential for stabilizing the outward-facing conformation. In contrast, as the structure of GLUT5 rearranges to the inward-facing conformation (using bGLUT5 for demonstration), the N- and C-domains move apart to release substrate, leading to disruption of the interdomain R159-E401 and the E337-R341 salt bridge formed in the C-domain of GLUT5 (Figure 6.9C, right).

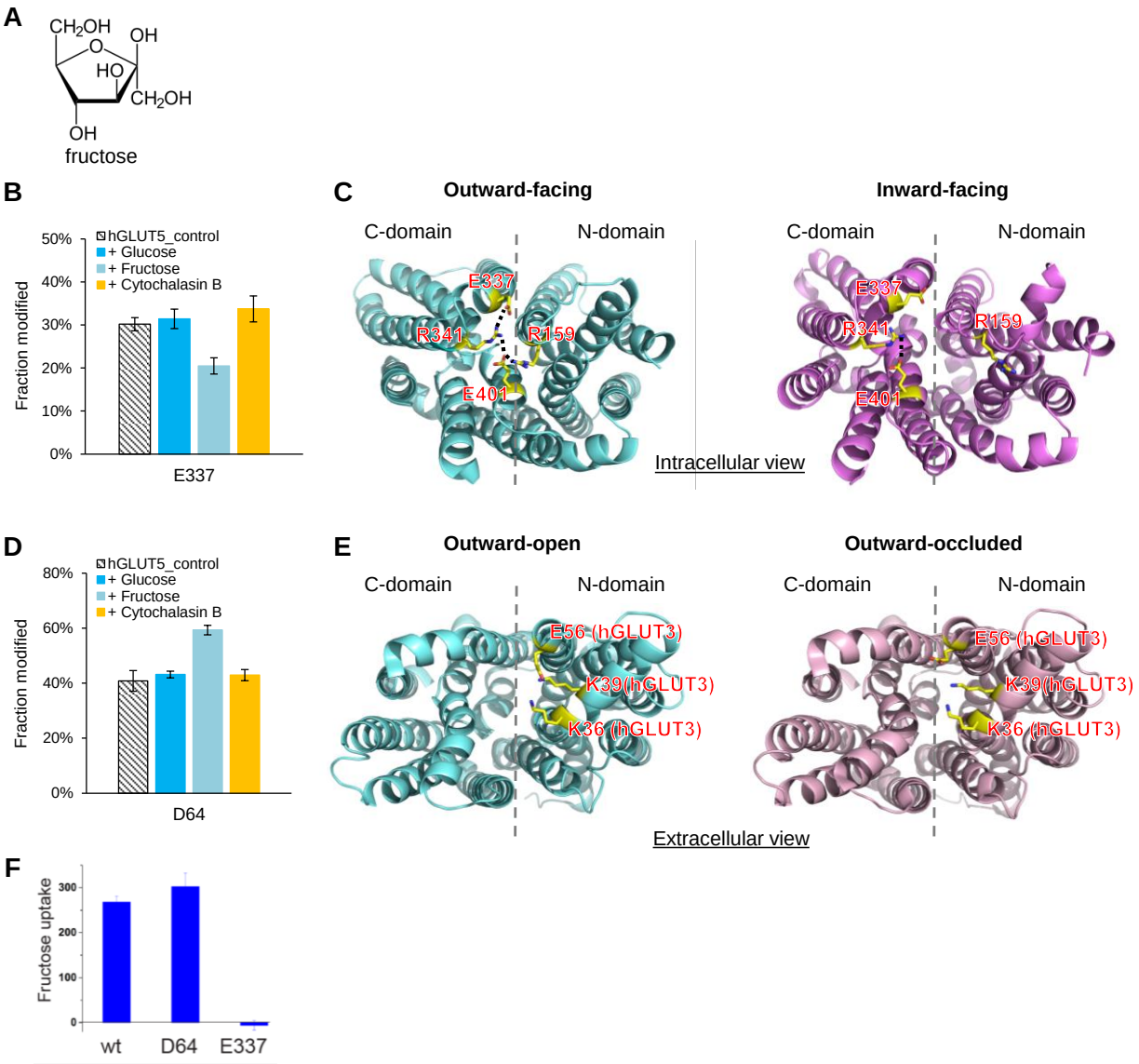


Figure 6.9. Salt bridge interactions in hGLUT5. (A) Structure of fructose. (B) GEE-modification extents of E337 in hGLUT5. (C) Salt bridge interactions associated with E337 mapped onto the structure of rGLUT5 in the outward open conformation (cyan, PDB 4YBQ) and bGLUT5 in the inward-facing conformation (violet, PDB 4YB9). To facilitate structural comparison, rGLUT5 and bGLUT5, residues are labelled with hGLUT5 numbering. Views are from the intracellular side, and ICHs are omitted for clarity. (D) GEE modification extents of D64 in hGLUT5. (E) Orientation of E56 in hGLUT3 (equivalent to D64 in hGLUT5) in the

outward-open (cyan, PDB 4ZWC) and outward-occluded (pink, PDB 4ZW9) conformations. Views are from the extracellular side. (F) Effect of substitution of D64 or E337 to alanine on the fructose-uptake ability of hGLUT5.

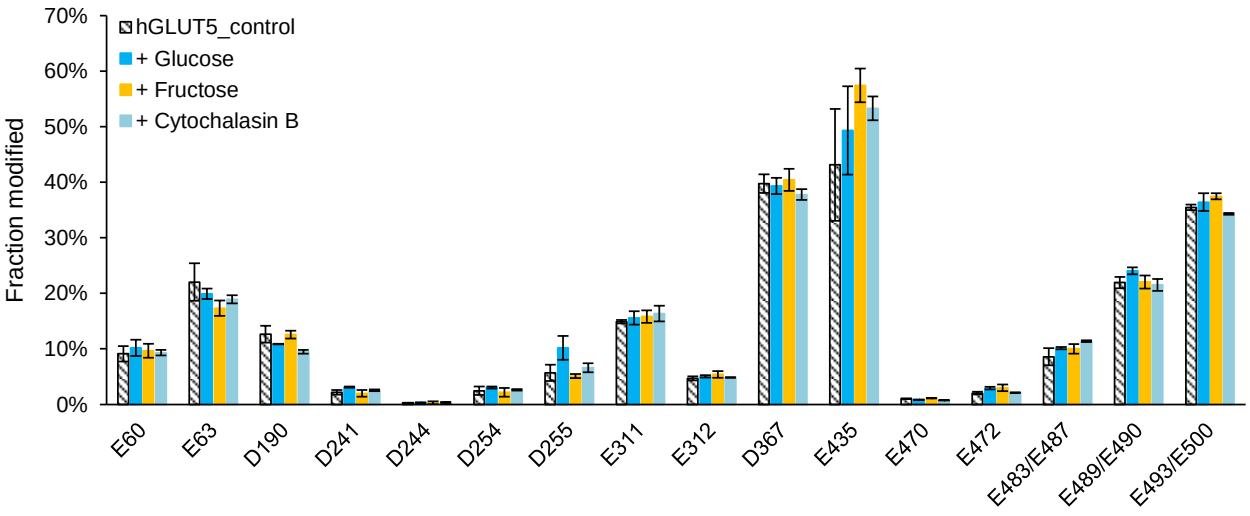


Figure 6.10. Asp/Glu residues in hGLUT5 showing no significant change in the level of GEE modification.

On the extracellular side, D64 undergoes increased GEE modification in the fructose-bound hGLUT5 that presumably favors the outward-facing conformation (Figure 6.9D). The structure of GLUT5 in the outward-open conformation obtained for rGLUT5, however, lacks resolution in the region containing the residue equivalent to D64 in hGLUT5. As a result, we based our analysis on the structures of hGLUT3 that has E56 equivalent to D64 in hGLUT5. Only structures of the outward-facing hGLUT3 are available, including the outward-open and outward-occluded conformations.[24] For inward-facing hGLUT3, E56 (TM3) can salt bridge in the N-domain with both K36 (TM2) and K39 (TM2) on the extracellular side. Figure 6.9E shows

those putative salt bridge residues mapped onto the outward-open and outward-occluded conformations of hGLUT3, in which the distance between the acidic and basic residues are > 9 Å and > 4Å, respectively. These distances indicate salt bridging on the extracellular side is abolished or severely weakened in GLUT3 in the outward-facing conformation. Assuming a similar set of interactions with the conserved residues in hGLUT5 as in hGLUT3, the salt bridge is between D64 (TM3) and K47 (TM2) in the N-domain of hGLUT5. Our protein footprinting result shows that interaction of D64 is weakened in the outward-facing conformation, as evidence by its increased GEE modification for fructose-bound GLUT5 (Figure 6.9D). Comparing the structure of GLUT3 in the outward-open and outward-occluded conformations shows a difference in the side-chain orientation of E56, equivalent to D64 in hGLUT5; the side chain points inward to the translocation pore in the outward-open conformation (Figure 6.9E, left), and rearranges to point to the extracellular side in the outward-occluded conformation (Figure 6.9E, right), indicating that E56 in GLUT3 structurally rearranges in the translocation cycle (we lack its structure in the inward-facing conformation for a direct comparison). Substitution of D64 slightly increases the substrate uptake activity of hGLUT5 (Figure 6.9F), suggesting the role of D64, similar to that of E299 in hGLUT1, on the extracellular side in maintaining a transient inward-facing conformation.

6.5 Conclusions

We implemented a novel live-cell protein footprinting approach, combining suspension cell expression, GFP tagging, isotope-encoded labeling, improved in-gel digestion and mass spectrometry to identify key salt bridges that controls the alternating access motion of MFS

transporters. As a probe of changes in protein solvent accessibility in live cells, MS-based footprinting offers direct evidence for the alternate access model of the membrane transporters, which is consistent with the X-ray structures and site-directed alkylation, tryptophan fluorescence and Förster resonance energy transfer. Our results also provide a confirmation that in the cell, the salt bridges are as predicted from X-ray.

The use of suspension cells provides sufficient amount of a membrane protein that has a low expression level for protein footprinting. Incorporation of isotope-encoding facilitates the detectivity of the modified peptides and allows unambiguous identification of the modification in a complex mixture of peptides. Optimization of the downstream proteomics workflow overcomes the problem of under-representation of peptides from hydrophobic regions by affording high coverage across a membrane protein with multiple TM domains. The outcome is a detailed view of the local structure of the membrane proteins. Using the method, we studied structures of the MFS transporters, exemplified by GLUTs, in their native cellular environment by chemical "footprints", and identified the key salt bridges that act as molecular switches for the conformational changes required for substrate translocation.

MFS transporters specifically, and membrane proteins in general, are heterogenous in term of conformation and post translational modification. Advantageously, the applicability of the live-cell protein footprinting method is minimally affected by this limitation. Furthermore, structure information obtained from the membrane protein in live cell can also be used to facilitate interpretation of X-ray crystallography for an atomic-level characterization of the membrane protein. In addition, adapting chemical probes with different specificities and reactivities with amino-acid residues, we envision continued expansion of the approach to study other function-related structural features of membrane proteins and their interactions with ligands and drugs that

regulate/affect the biological activity. Although the focus of this study is the translocation mechanism of MFS, we suggest that ligand-induced conformational changes, aggregation, and effects of modification of other membrane receptors are now in reach.

6.6 Acknowledgements

We thank Dr. Weikai Li, Dr. Shuang Li, Dr. Yihu Yang, Dr. Weidong Cui for their help and discussion in designing/conducting the live-cell protein footprinting experiment, and Dr. Paul W. Hruz and Monique Heitmeier, Dr. Weikai Li, Dr. Shuang Li for conducting the functional assays. This research was supported by the NIH P41GM103422 (to MLG), and NIH R01HL121718, AHA 14GRNT20310017, ASH scholar award (to WL).

6.7 References

1. Yan, N., Structural Biology of the Major Facilitator Superfamily Transporters. *Annu Rev Biophys* **2015,** *44*, 257-83.

2. Law, C. J.; Maloney, P. C.; Wang, D. N., Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **2008,** *62*, 289-305.

3. Saier, M. H., Jr.; Beatty, J. T.; Goffeau, A.; Harley, K. T.; Heijne, W. H.; Huang, S. C.; Jack, D. L.; Jahn, P. S.; Lew, K.; Liu, J.; Pao, S. S.; Paulsen, I. T.; Tseng, T. T.; Virk, P. S., The major facilitator superfamily. *J. Mol. Microbiol. Biotechnol.* **1999,** *1* (2), 257-79.

4. Reddy, V. S.; Shlykov, M. A.; Castillo, R.; Sun, E. I.; Saier, M. H., Jr., The major facilitator superfamily (MFS) revisited. *FEBS J.* **2012,** *279* (11), 2022-35.

5. Law, C. J.; Almqvist, J.; Bernstein, A.; Goetz, R. M.; Huang, Y.; Soudant, C.; Laaksonen, A.; Hovmoller, S.; Wang, D. N., Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. *J. Mol. Biol.* **2008,** *378* (4), 828-39.

6. Stelzl, L. S.; Fowler, P. W.; Sansom, M. S.; Beckstein, O., Flexible gates generate occluded intermediates in the transport cycle of LacY. *J. Mol. Biol.* **2014,** *426* (3), 735-51.

7. Andersson, M.; Bondar, A. N.; Freites, J. A.; Tobias, D. J.; Kaback, H. R.; White, S. H., Proton-coupled dynamics in lactose permease. *Structure* **2012,** *20* (11), 1893-904.

8. Doki, S.; Kato, H. E.; Solcan, N.; Iwaki, M.; Koyama, M.; Hattori, M.; Iwase, N.; Tsukazaki, T.; Sugita, Y.; Kandori, H.; Newstead, S.; Ishitani, R.; Nureki, O., Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc Natl Acad Sci U S A* **2013,** *110* (28), 11343-8.

9. Nomura, N.; Verdon, G.; Kang, H. J.; Shimamura, T.; Nomura, Y.; Sonoda, Y.; Hussien, S. A.; Qureshi, A. A.; Coincon, M.; Sato, Y.; Abe, H.; Nakada-Nakura, Y.; Hino, T.; Arakawa, T.; Kusano-Arai, O.; Iwanari, H.; Murata, T.; Kobayashi, T.; Hamakubo, T.; Kasahara, M.; Iwata, S.; Drew, D., Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature* **2015,** *526* (7573), 397-401.

10. Deng, D.; Xu, C.; Sun, P.; Wu, J.; Yan, C.; Hu, M.; Yan, N., Crystal structure of the human glucose transporter GLUT1. *Nature* **2014,** *510* (7503), 121-125.

11. Quistgaard, E. M.; Low, C.; Guettou, F.; Nordlund, P., Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat. Rev. Mol. Cell Biol.* **2016,** *17* (2), 123-32.

12. McPherson, A.; Gavira, J. A., Introduction to protein crystallization. *Acta Crystallographica. Section F, Structural Biology Communications* **2014,** *70* (Pt 1), 2-20.

13. Mohan, S.; Sheena, A.; Poulose, N.; Anilkumar, G., Molecular Dynamics Simulation Studies of GLUT4: Substrate-Free and Substrate-Induced Dynamics and ATP-Mediated Glucose Transport Inhibition. *PLOS ONE* **2010,** *5* (12), e14217.

14. Hopper, J. T. S.; Yu, Y. T.-C.; Li, D.; Raymond, A.; Bostock, M.; Liko, I.; Mikhailov, V.; Laganowsky, A.; Benesch, J. L. P.; Caffrey, M.; Nietlispach, D.; Robinson, C. V., Detergent-free mass spectrometry of membrane protein complexes. *Nat. Methods* **2013,** *10*, 1206.

15. Barrera, N. P.; Robinson, C. V., Advances in the mass spectrometry of membrane proteins: from individual proteins to intact complexes. *Annu. Rev. Biochem.* **2011,** *80*, 247-71.

16. Laganowsky, A.; Reading, E.; Allison, T. M.; Ulmschneider, M. B.; Degiacomi, M. T.; Baldwin, A. J.; Robinson, C. V., Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **2014,** *510*, 172.

17. Whitelegge, J. P., Integral Membrane Proteins and Bilayer Proteomics. *Anal. Chem.* **2013,** *85* (5), 2558-2568.

18. Zhang, H.; Liu, H.; Blankenship, R. E.; Gross, M. L., Isotope-Encoded Carboxyl Group Footprinting for Mass Spectrometry-Based Protein Conformational Studies. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (1), 178-81.

19. Wen, J.; Zhang, H.; Gross, M. L.; Blankenship, R. E., Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectrometry-based footprinting. *Proc Natl Acad Sci U S A* **2009,** *106* (15), 6134-9.

20. Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **2007,** *1*, 2856.

21. Huang, Y.; Lemieux, M. J.; Song, J.; Auer, M.; Wang, D. N., Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science* **2003,** *301* (5633), 616-20.

22. Thorens, B.; Mueckler, M., Glucose transporters in the 21st Century. *American Journal of Physiology - Endocrinology and Metabolism* **2010,** *298* (2), E141-E145.

23. Kapoor, K.; Finer-Moore, J. S.; Pedersen, B. P.; Caboni, L.; Waight, A.; Hillig, R. C.; Bringmann, P.; Heisler, I.; Muller, T.; Siebeneicher, H.; Stroud, R. M., Mechanism of inhibition of human glucose transporter GLUT1 is conserved between cytochalasin B and phenylalanine amides. *Proc Natl Acad Sci U S A* **2016,** *113* (17), 4711-6.

24. Deng, D.; Sun, P.; Yan, C.; Ke, M.; Jiang, X.; Xiong, L.; Ren, W.; Hirata, K.; Yamamoto, M.; Fan, S.; Yan, N., Molecular basis of ligand recognition and transport by glucose transporters. *Nature* **2015,** *526* (7573), 391-396.

25. Miyamoto, K.; Yamashita, T.; Tsukiyama, T.; Kitamura, N.; Minami, N.; Yamada, M.; Imai, H., Reversible membrane permeabilization of mammalian cells treated with digitonin and its use for inducing nuclear reprogramming by Xenopus egg extracts. *Cloning Stem Cells* **2008,** *10* (4), 535-42.

26. Zhang, H.; Liu, H.; Blankenship, R. E.; Gross, M. L., Isotope-Encoded Carboxyl Group Footprinting for Mass Spectrometry-Based Protein Conformational Studies. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (1), 178-181.

27. Wen, J.; Zhang, H.; Gross, M. L.; Blankenship, R. E., Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectrometry-based footprinting. *Proceedings of the National Academy of Sciences* **2009,** *106* (15), 6134-6139.

28. Speers, A. E.; Wu, C. C., Proteomics of integral membrane proteins--theory and application. *Chem. Rev.* **2007,** *107* (8), 3687-714.

29. Carruthers, A.; Helgerson, A. L., Inhibitions of sugar transport produced by ligands binding at opposite sides of the membrane. Evidence for simultaneous occupation of the carrier by maltose and cytochalasin B. *Biochemistry* **1991,** *30* (16), 3907-15.

30. Park, M.-S., Molecular Dynamics Simulations of the Human Glucose Transporter GLUT1. *PLoS ONE* **2015,** *10* (4), e0125361.

Chapter 7: Conclusion and Outlook

MS-based protein footprinting is a valuable tool for characterizing protein and protein complexes. The method probes in-solution structure and provide information complementary to that from X-ray crystallography or nuclear magnetic spectroscopy (NMR). The primary interest in Gross lab is focused on the development and application of MS-based structural proteomics to study protein biophysics. The six chapters in this dissertation reflect the method development and application of MS-based protein footprinting to study protein aggregation, protein-ligand interaction and higher order structure of membrane proteins.

7.1 FPOP and MS for monitor amyloid formation

In Chapter 2, we described the application of FPOP to characterize the oligomerization and fibrillation of Aβ by monitoring its solvent accessibility changes that accompany these processes. Compared to a traditional thioflavin T (ThT) fluorescence assay and an HDX approach, the level of FPOP modification sensitively responds to the solvent accessibility of various Aβ species. Therefore, we were able to identify different oligomeric species based on their characteristic modification levels and to resolve a more detailed picture of the transition stages in Aβ aggregation. Furthermore, the irreversible labeling by FPOP permits characterization of aggregation behaviors of Aβ sub-regions and even some amino-acid residues.

We later extended the platform to evaluate the effect of a polyphenol amyloid inhibitor in remodeling Aβ aggregation (Chapter 3) and found that binding of the inhibitor results in different structural effects on Aβ sub-regions. The interaction leads to stabilization of small Aβ oligomers and a significant extension in the lag phase and may provide hints for slowing aggregation in vivo.

For the future, in addition to small molecule inhibitors, other forms of inhibitors, including peptides and antibodies can also be investigated by using the platform. Furthermore, the approach should be useful for the study of Aβ mutant. Other investigations have shown that some mutants (e.g., Tottori (D7N), Flemish (A21G), and Arctic (E22G)) and N-terminal truncated Aβ species (e.g., pyroglutamate-modified Aβ) exhibit different early oligomer distributions and accelerated rates of aggregation. The platform can be readily expanded to study those Aβ species.

Another immediate interest in this field is the interaction of Aβ with lipid membranes. It is known that there is a complex interplay between lipids and the generation, clearance, and deposition of Aβ. Membrane surfaces can serve as catalytic sites and promote formation of toxic Aβ aggregates. Compared to HDX, FPOP is advantageous for studying Aβ in a lipid environment, because the irreversible modifications it imparts allow more flexibility in the post-labeling sample clean-up (e.g., acetone precipitation, chromatography or buffer exchange) to remove lipids that are incompatible with MS analysis. Furthermore, FPOP can be applied to study the interaction of Aβ with apolipoprotein E (Apo E), which binds to Aβ and significantly influences the receptor-mediated Aβ uptake by neurons that ultimately leads to slow Aβ clearance. Last but not the least, the FPOP platform can be applied to study aggregation and misfolding of other amyloidogenic proteins such as α-synuclein and Tau protein.

7.2 FPOP for Protein-Ligand Interaction

In Chapter 4-5, we demonstrated the capabilities of three different protein footprinting techniques, including HDX, FPOP and carboxyl-group footprinting in deciphering protein-ligand

interaction and protein higher order structures. In the case of bromodomain-containing protein interacting with a benzodiazepine inhibitor (Chapter 5), FPOP clearly points to a critical binding region, whereas HDX is not sufficiently sensitive in mapping the binding interface in the protein hydrophobic cavity. This study serves as an example for using FPOP to map directly the binding interface of a protein for a small-molecule ligand that does not induce significant conformational change on protein conformation. In this study, we incorporated a peptide reporter into the FPOP experiment to provide a measure of the hydroxyl radical lifetime. Interestingly, we find the modification on Met residues saturates as the hydroxyl radical lifetime increases (achieved through decreasing the scavenger concentration). This phenomenon is specifically observed for Met and is worth further investigation to help understand the modification mechanism of Met in FPOP.

7.3 Live-cell footprinting for membrane protein characterization

In Chapter 6, we applied protein footprinting, specifically carboxyl group footprinting, to label proteins in a live cell and capture the structure of membrane transport proteins (i.e., hGLUT5 and hGLUT5). We demonstrated a live-cell footprinting methodology combining suspension cell expression, GFP tagging, improved in-gel digestion, and mass spectrometry to study structures of MFS transporters in a native cellular environment. This proof of concept study is a template for using chemical footprinting, a method normally performed on purified protein in buffer solution, to interrogate a complicated, biologically relevant system such as the cell.

Although this example is focused on using glycine ethyl ester to label specifically acidic residues (Asp/Glu) and identify potential key salt bridges controlling the alternating access motion of membrane transport proteins, the method should be readily expandable by adapting other footprinting reagents into the established workflow to study many other biological activities of interest for membrane proteins. For example, development of a hydrophobic precursor that can enter the lipid bilayers will permit footprinting of residues in transmembrane regions. Moreover, the methodology can be applied to characterize dynamic biological processes that happen in the cell. For example, ferroportin, an iron-regulated transporter, undergoes ubiquitination followed by internalization upon binding of its endogenous hormone inhibitor hepcidin. By footprinting ferroportin in the cell at different times after its exposure to hepcidin, we should be able to characterize the ligand-induced conformational change and ubiquitination sites of ferroportin during degradation.