Summer 8-15-2020

Development and Application of Mass Spectrometry-Based Approaches for Protein Higher Order Structure Analysis and Protein-Protein Interaction Characterization

Mengru Zhang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Chemistry

Dissertation Examination Committee:
Timothy Wencewicz, Chair
Michael L. Gross, Co-chair
Gaya K. Amarasinghe
Joseph A. Fournier
Meredith Jackrel

Development and Application of Mass Spectrometry-Based Approaches for Protein Higher Order Structure Analysis and Protein-Protein Interaction Characterization

by

Mengru Zhang

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

August 2020
St. Louis, Missouri

ORCID: 0000-0001-9082-1272

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 2

Chapter 4

Chapter 5

Chapter 6

Acknowledgments

Foremost, I would like to express my sincere and deepest gratitude to my advisor, Dr. Michael L. Gross, who provided us such a vibrant lab environment and appreciative multidisciplinary collaborations that establish a remarkable platform for me to grow as an independent scientist. Not limiting to his tremendous help in science, I cherish many of his lessons from all aspects. Being caring, hard-working, adventurous, and charismatic, he is my admirable role model of all times. I am so fortunate to be part of his group. These years of memory will be my forever nutrients wherever I go.

I would also like to thank my dissertation advisory and committee members, Prof. Timothy Wencewicz, Prof. Meredith Jackrel, Prof. Joseph A. Fournier, and Prof. Gaya K. Amarasinghe for their valuable advice in my research and dissertation.

In addition, I want to thank my collaborators at WUSTL, Prof. Gaya K. Amarasinghe, Prof. Daisy Leung, Prof. Weikai Li, Dr. Haijun Liu, and at Bristol-Myers Squibb, Dr. Guodong Chen, Dr. Richard Y-C Huang, Dr. Brett R. Beno, Ekaterina G. Deyanova, Dr. Jing Li for their generous scientific input.

I am grateful to all group members in Gross lab for their support and encouragement for the past five years, especially for Dr. Henry Rohrs for many valuable discussions including life and science, and for Don Rempel who helped me greatly with the lasers.

A special thanks to my parents, Fengling Zhang and Jiade Zhang, for always being there for me. Their unconditional love, support and understanding give me the courage and power to fight and thrive.

Last but not least, I want to say thank you to the love of my life, Xiaoran Roger Liu. Thank you for your faithful love and company even in the days of darkness. Today is happened to be our ten-years' anniversary and I am so grateful to have you in my life.

Research in this thesis is supported by the National Institute of Health 2P41GM103422 and the graduate school of Washington University.

Mengru Zhang

Washington University in St. Louis

August 2020

Dedicated to my family.

ABSTRACT OF THE DISSERTATION

Development and Application of Mass Spectrometry Based-Approaches for Protein Higher Order Structure Analysis and Protein-Protein Interaction Characterization

by

Mengru Zhang

Doctor of Philosophy in Chemistry

Washington University in St. Louis, 2020

Professor Timothy Wencewicz, Chair

Professor Michael L. Gross, Co-chair

Proteins, one of the most fundamental biomolecules, adopt unique higher order structures (HOS) to enable diverse biological functions. Deciphering protein HOS is crucial to gain deeper insights of their working mechanisms and to develop biotherapeutics. Mass spectrometry (MS)-based approaches evolved rapidly in the past 30 years and are now playing critical roles in protein HOS characterization. One of those approaches is MS-based footprinting whose principle is to map the solvent accessible surface area (SASA) to deliver structural information. Protein footprinting can be achieved by reversible labeling, e.g., hydrogen-deuterium exchange (HDX), and by irreversible labeling using radical-based reagents or other targeted labeling reagents. Irreversible labeling such as fast photochemical oxidation of protein (FPOP) and cross-linking (XL) delivers the information of the reactive amino acid side chains, whereas HDX allows the analysis of the backbone amides. Information from the two aspects are different yet complementing to each other. In Chapter 1, the two MS-based footprinting methodologies are reviewed and discussed in detail including the fundamental, history, and recent applications.

In the second section, development of the irreversible radical footprinting is the primary focus, specifically the elaboration of FPOP platform. Chapter 2 describes the generation and evaluation of a new radical reagent, the carbonate radical anion, on the FPOP platform; a radical that can selectively label methionine and aromatic residues, complementing others radical reagents in the footprinter “toolbox”. Chapter 3 demonstrates a novel way of elaborating FPOP platform that is to follow protein unfolding by coupling two lasers together: one to induce protein conformational changes by pH jump and the other to label protein with radical reagents. A time delay between the two lasers enables the characterization of the dominant protein conformations at different stages of unfolding.

The third section emphasizes the integration of several footprinting approaches as well as computational methods for comprehensive analysis of protein HOS. In Chapter 4 and 5, HDX, XL-MS and molecular docking are combined to determine protein-protein binding interfaces and to map epitope/paratope of an antigen-antibody complex, respectively. In particular, Chapter 4 discusses the potential of using HDX to adjudicate candidate docking models for quaternary structure elucidation. Besides molecular docking, homology modeling in combination with XL-MS is also a successful marriage to decipher protein structures, an example is demonstrated in Chapter 6. Restraints derived from cross-links help modify and validate a predicted structure of phycolisome, contributing to the first proposed architecture of the protein complex in cyanobacteria.

The six chapters combine to demonstrate the development and application of MS-based footprinting in protein HOS characterization. Given the effectiveness and powerfulness of these methods, significant contributions by MS-related approaches are well to be expected in the field of structural proteomics.

Chapter 1: Introduction to Mass Spectrometry-Based Footprinting for Higher Order Protein Structure Analysis*

* This chapter is based on the following publication: Liu, X. R.; Zhang, M. M.; Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher Order Structure Analysis: Fundamentals and Applications *Chem. Rev.* **2020,** *120*, 4355-4454.

1.1 Abstract

Proteins exhibit diverse higher order structures (HOS) that contribute to their distinctive functionalities. Characterization of protein HOS is crucial to allow deeper understanding of their interactions and enable better designs of biotherapeutics. There are many biophysical approaches to probe protein HOS: traditional optical methods that provide fast but low-resolution information and high-resolution techniques (X-ray crystallography and nuclear magnetic resonance) that deliver atomic-level information. Mass spectrometry-based (MS) approaches offer middle-to-high resolution data and complement the others while providing significant advantages (e.g., fast throughput, sensitive detection, in-solution characterization). One of the essential components of the MS-based approaches is protein footprinting, which utilizes different labeling reagents to map the solvent-accessible area (SASA) of proteins to describe HOS. Reversible footprinting utilizes H-D exchange for a relatively unbiased labeling on the backbone amide hydrogens; however, reagents employed in irreversible footprinting react, often with high specificity, with the side chains of accessible residues. In addition, the labeling time scale, sample preparation and experimental set-up are all different for the two footprinting methods, contributing to different applications. This chapter summarizes and compares these differences in detail including the historical and fundamental aspects, providing a comprehensive background for the thesis.

1.2 Introduction

Proteins are one of the fundamental biomolecules, regulating complex yet delicate bioactivity networks. The diverse functionalities[1] of proteins include (i) providing mechanical support as the cell skeleton; (ii) transporting other biomolecules to the targeted locations; (iii) transmitting

signals to coordinate different types of biological process; (iv) catalyzing cellular reactions to assist new molecules formation; (v) binding to pathogens to provide immune protection. To afford distinctive biological functions, protein three-dimensional (3D) structures are highly diverse, which in turn leads to several mechanisms that are taken to execute their various roles. Therefore, it is crucial to characterize protein higher order structure (HOS), and that insight can not only deepen our understanding of their functions mechanistically but will also aid the design of biotherapeutics along the way.

There are four orders of protein structures, namely (i) primary structure that consists of the amino acids sequence encoded genetically in nucleic acids; (ii) secondary structure that arises by folding of the primary structure to give, for example, α-helices, β-sheets, and β-turns; (iii) tertiary structure resulting from assembly of all local structures into an overall 3D architecture; (iv) quaternary structure that forms when one protein interacts with another to give a protein complex. Characterization of protein HOS typically refers the secondary structure and beyond, and this class of structure can be addressed by many biophysical approaches.

To probe globally the secondary structure content or the conformational changes near a label, CD, FT-IR, FRET and UV can be good candidates. Although these optical techniques provide relatively low resolution, they deliver the results efficiently and rapidly. On the other hand, the approaches that have the highest resolution, X-ray crystallography,[2] nuclear magnetic resonance (NMR)[3-5] and cryo EM[6-7], give atomic-level information but also show some disadvantages. X-ray crystallography relies on the diffraction pattern of a crystalized protein sample, which can be challenging to obtain. In addition, the information derived from a solid-state structure may not be relevant in solution. NMR does probe the protein structures in solution,[8-9] but the required sample amount is large, typically milligrams. Furthermore, the long signal-averaging time and

complicated data analysis also limit its applications. Recently developed cryo-EM also enables the characterization of non-crystalline samples[10-11] with low sample amount and straightforward sample preparation. The approach, however, favors high-molecular-proteins, typically larger than 100 kDa.[12]

Over the past three decades, mass spectrometry-based methods have evolved from a few scattered studies in specific amino acid labeling and HDX to many applications that give middle-to-high resolution. The sensitive detection, fast throughput, low amount of protein sample (nanogram to microgram level) in solution, rapidly developed instrumentation, and advanced data analyzing software greatly promote its adoption to various biological questions and protein systems.

One critical aspect to characterize protein HOS is MS-based footprinting[13], which maps solvent-accessible surface area (SASA) on protein or protein complexes. MS-based footprinting can be further subdivided into two classifications, namely reversible and irreversible footprinting. The most common example of reversible footprinting is based on the H-D exchange on the backbone amide hydrogens; this approach is also known as the hydrogen-deuterium exchange MS (HDX-MS). Irreversible footprinting makes use of targeted-labeling reagents with slower chemistry (mins to hours) and radical-labeling reagents that can react with residue side chains on the micro to millisec time scale. When combined with bottom-up MS analysis, these approaches resolve the structural information from peptide level to residue level.

In this chapter, we review the history, fundamentals, experimental designs, and key applications of both the reversible and irreversible footprinting. The review serves to introduce the development and application of the MS-footprinting approaches as a theme of this thesis.

1.3 Reversible Protein Footprinting: Hydrogen-Deuterium Exchange

1.3.1 History and Fundamentals

Hydrogen deuterium exchange (HDX), a rapid emerging technique in recent years, has been adopted widely to study protein conformations and dynamics. In 1954, Hvidt and Linderstrøm-Lang carried out the first HDX on protein with pork insulin, establishing the connection between the H/D exchange rate and protein dynamics.[14] Later, HDX was coupled with other analytical techniques (e.g., NMR[15-17] and Fourier transform infrared (FTIR) spectroscopy) to achieve better spatial resolution. The marriage of HDX and NMR not only enables residue-level information for small proteins[18] but also allows the determination of exchange rates for individual amide protons[3, 19]. The broad application of HDX-NMR has been largely limited, however, by the complications of handling large proteins. Instead, the integration with mass spectrometry brought forth a more sensitive and broadly applicable approach. Given the sensitive detection of MS and the apparent mass shift when exchanged into deuterium, HDX-MS enabled advances to larger proteins at lower concentration. The first study of HDX-ESI-MS was carried out by Katta and Chait[20] in 1991. After deuterium exchange, they observed a shift of the centroid mass on the isotopic envelope of the global protein. Later, Zhang and Smith[21] coupled a proteolytic approach with HDX to locate the structural changes at regional resolution, a pioneer study setting up the paradigm for the HDX analysis today.

HDX delivers the structural information of proteins by describing the local solvent accessibility and hydrogen bonding networks on amide backbone. The backbone amide hydrogens that are exposed to solvent on a protein surface and involved weakly in hydrogen bonding are more

labile, and they exchange more rapidly than the ones that are buried inside the protein structure and stabilized by H-bonds. The hydrogens on the sidechains, however, are even faster exchangers that are difficult for most experiments to measure. The negligible exchange on side chain hydrogens makes the information from HDX an "unbiased" readout that reports all amino acids containing an N-H hydrogen on backbone. The rate constant of N-H to N-D, k_{ch}, is determined by pD, temperature, and the local environment of the residues. HDX is both acid and base-catalyzed, giving a minimal k_{ch} at pH around 2.5. [22-23] Practically, the apparent HDX exchange rate, k_{HDX}, is much smaller than k_{ch}, mainly affected by the accessibility to the amide sites and the intramolecular interaction of N-H···O=C among peptide bonds.[24] The breathing motion of the H-bonds comes with an opening and a closing transition (expressed as k_{op} and k_{cl}, respectively), which can compete with k_{ch}, contributing to different HDX regimes, EX1 and EX2 (eq 1.1). In EXI where $k_{ch} >> k_{cl}$, the apparent $k_{HDX} = k_{op}$, indicting an immediate exchange to deuterium after the initial opening event. In EX2 where $k_{cl} >> k_{ch}$, the apparent HDX exchange rate is characterized as $k_{HDX} = K_{op}k_{ch}$, in which $K_{op} = k_{op}/k_{cl}$. Generally, the EX2 regime is more prevalent than EX1[25-26], whereas the combination of the two is also common to see[27-28].⇌

$$\text{N-H}_{\text{closed}} \underset{k_{cl}}{\overset{k_{op}}{\longleftrightarrow}} \text{N-H}_{\text{open}} \xrightarrow{D_2O,\ k_{ch}} \text{N-D}_{\text{open}} \underset{k_{op}}{\overset{k_{cl}}{\longleftrightarrow}} \text{N-D}_{\text{closed}} \qquad (1.1)$$

1.3.2 Experimental Setup and Applications

Most HDX experiments are conducted differentially between two states, for example a bound and an unbound protein or a wildtype and a protein mutant. The changes in SASA that result from the binding interactions or conformational alterations are represented by the deuterium uptake extent. In practice, HDX is often coupled with bottom-up analysis to get regional or even residue-level information. Prior to a differential HDX experiment, peptide mapping is an essential step to create a list of peptides that can be monitored to afford spatial resolution. The

main components in peptide mapping is to optimize the quenching condition for a maximal denaturation of the targeted protein and, at the same time, minimal back exchange of the incorporated D to H. Concentration of the denaturant and the reducing regent as well as the incubation time and temperature all require careful consideration. The optimized quenching, digestion protocol, and MS method are then applied in the HDX experiment (workflow is shown in Figure 1.1).

The HDX data obtained are typically analyzed by commercial software (e.g., HDExaminer,[29] HDX Workbench,[30]), where a built-in algorism fits the isotopic distribution of each peptide, giving the centroid of the pattern and the average deuterium uptake compared no exchange.

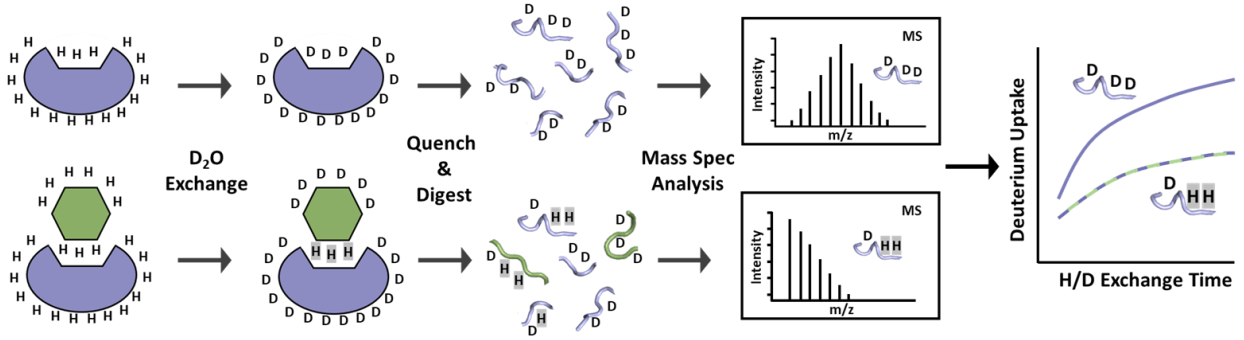


Figure 1.1 Schematic representation of HDX-MS workflow.

HDX-MS being a valuable asset in the characterization toolbox for protein HOS analysis has many advantages. The use of deuterium as a footprinting reagents introduces minimal perturbation on protein structure and provides a "unbiased" labeling on nearly all residues (except Proline), comparing to other protein labeling techniques. In addition, the interface with mass spectrometry enables a fast throughput of data acquisition. Recent developments of HDX data processing software further contribute to more efficient data analysis, promoting wide applications in both academic and industrial research laboratories. Up to now, HDX-MS has been successfully applied to map epitope/paratope of biotherapeutics (an example is discussed in

detail in Chapter 5),[31-34] to characterize protein binding interface with other biomolecules (an example is discussed in detail in Chapter 4),[35-38] to follow protein folding/unfolding dynamics,[39-40] and to address the allosteric regions of protein or protein complexes.[41]

There are challenges for further advancing the approach. To increase the spatial resolution, even to individual amino acid level, researchers have made considerable effort including multi-enzyme digestion to increase overlapping peptides,[42-43] high-pressure digestion for enhanced digestion efficiency,[44-45] and employment of ETD or ECD for orthogonal fragmentations with minimal HD scrambling[46-48]. Furthermore, the demanding adaptability of HDX-MS to membrane proteins is also a concern, wherein the on-line removal of lipids can be troublesome. Although the recent development of using zirconium (IV) oxide to remove detergent and lipids[49] represents a promising solution, a robust methodology is still on the way.

1.4 Irreversible Protein Footprinting: Chemical Cross-linking

The 20 amino acids have different kinds of functional groups on their side chains; most have functional groups (COOH, SH, NH_2, OH, $CONH_2$, aromatic ring) that can be labeled with chemical reagents that react specifically, usually with one or two amino acids. These modifying reactions can be used for footprinting provided the reactivity of these groups depends on SASA of the protein, and the modifications in the early stages do not affect the protein structure, minimizing the biased report during the footprinting itself. Reagents react directly and usually slowly with specific solvent-accessible side chains in contrast to free radicals, which react rapidly. The product contains a characteristic mass tag that can be detected by MS analysis. Although numerous reagents have been developed to react with the amino acid side chains, a

qualified protein footprinting reagent needs to label the protein under physiological conditions with reasonable efficiency and speed. The size and hydrophilicity of the reagent should be close to that of water to ensure its reactivity is an indicator of SASA. Most of these reagents developed to date are highly specific, targeting one or two side chains or functional groups although some can react with more than two.

Residue-specific reagents have a long history of development for various purposes, while their applications in protein footprinting generally started from 1990s. This field is well reviewed in two articles[50-51], hence is not covered in this section. Chemical crosslinkers, which can be viewed as bifunctional footprinters, are discussed in detail.

1.4.1 Assessing Topology and Stoichiometry through Chemical Cross-linking

Chemical cross-linking has developed as a complementary area of research owing to its capability to probe HOS and to locate and define protein/protein interfaces. Our intention, however, is not to review this topic comprehensively, but to describe the workflow, show how XL can be viewed as a means of footprinting (i.e., “double footprinting”), discuss its role in integrated approaches, and highlight some recent developments to unify the methodology described in this work.

Cross-linking mass spectrometry (XL-MS) is effective at assessing protein complex topologies and elucidating protein structures besides capturing protein-protein interactions. Conceptually, XL utilizes bifunctional labeling reagents to cross-link (footprint) the constituent proteins forming an interface, thus providing information on the interacting species and their interfaces. When dealing with large proteins, XL can also report on the overall protein conformation.

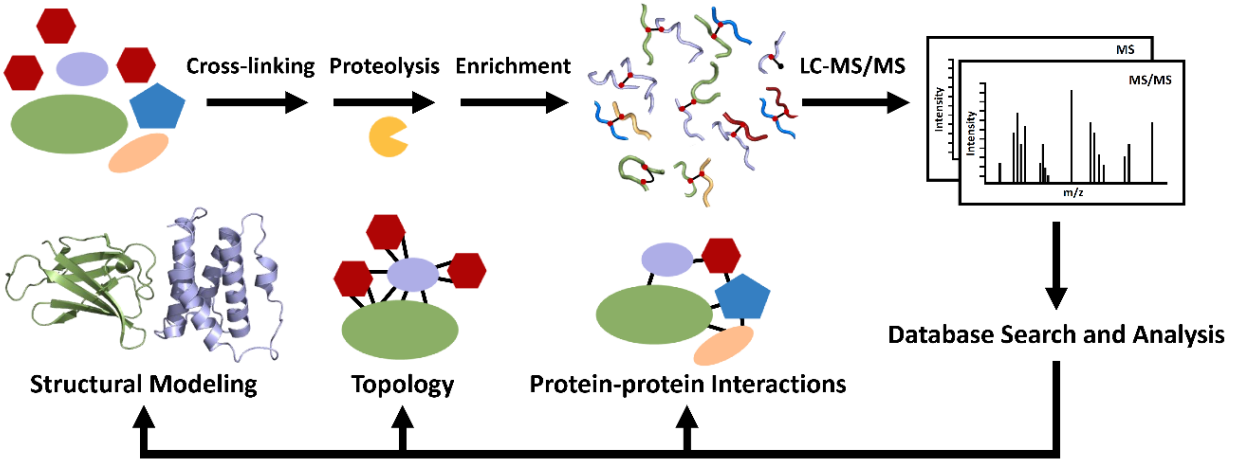


Figure 1.2 Workflow of bottom-up cross-linking mass spectrometry

To date, most of the MS-based XL studies have been by a bottom-up approach, whose workflow is presented in Figure 1.2. Briefly, proteins are first incubated with chosen cross-linkers under the optimized conditions that are established usually by monitoring with techniques simpler than MS (e.g., gel electrophoresis). The tethered proteins are then submitted to enzymatic digestion, enrichment of informative peptides, and LC-MS/MS for separation and detection. The MS data are further analyzed by using search engines to identify cross-linked peptides with an uncertainty specified by mass tolerance and false discovery rate (FDR). Significant advances since 2008 in data analysis have facilitated this approach; new software includes but is not limited to pLink,[52] xQuest,[53] XlinkX,[54] and StavroX[55]. Generated cross-link maps not only identify the connectivity of the adjacent protein subunits but also provide distance restraints as given roughly by the molecular separation between the two reactive functional groups of the cross-linker.

XL-MS alone is a middle-to-low spatial resolution approach because usually there are limited number of reactive residues at or near an interface of two proteins, and few crosslinks form. Young et al.[56] in 2000 showed that by combining XL distance restraints and computational modeling is a compelling way to improve the resolution for elucidating protein structures. The workflows can be adapted for integrative modeling[57-58] and *de novo* structure prediction[59-60]. The

major challenges of computational studies are the sampling and scoring of the generated models[61]. Larger data sets derived from XL-MS effectively decrease the size of the sampling space and benefit the scoring function, thus promoting efforts to find high-confidence models. Given there can be various reactive groups and spacer lengths on cross-linkers, a strategy involving multiple reagents should be taken to insure comprehensive information.

In 2014, Chait and coworkers[58] demonstrated the use of disuccinimidyl suberate (DSS) and a zero-length EDC in characterization of the nuclear pore sub complex of Nup84. Two data sets generated from the use of two reagents delivered different but complementary structural information that significantly benefited the subsequent model construction. Many other studies adopted similar ideas by combining amine-targeting reagents and carboxyl-targeting reagents (e.g., dihydrazides[62-63]) or nonselective cross-linkers[64-65]. In addition, the use of reagent combinations with different spacer lengths is recommended because short cross-linkers yield fewer cross-links but with narrower distance restraints, whereas long cross-linkers afford more cross-links but less structural definition because the distance assignments are over a broader range.[61]

New sample preparation methods also have developed rapidly (e.g., on-bead cross-linking[60]), improving the cross-linking chemistry and providing better analysis sensitivity. Many examples are discussed in recent reviews.[66-68]

By way of contrast, footprinting by targeted and free-radical reagents and/or by HDX maps solvent accessible regions and reflects differences in protein dynamics or binding events. This information provides deeper understanding of the entire protein structure, not just interfaces, and can adjudicate constructed 3D models from XL for in-solution proten.[69-71]

XL is not limited to structural characterization of a single protein complex, but rather it can promote protein-protein interactions (PPIs) studies in proteome-wide investigations.[52, 72-74] To meet the need for understanding the roles, functionalities, and mechanistic behavior of many protein complexes through their PPIs maps, cleavable XL reagents have been developed,[75] simplifying and increasing the accuracy for identification of cross-linked peptides. Incorporation of affinity groups and sophisticated enrichment procedure have led to success; examples are studies performed on E. coli[73] and *Caenorhabditis elegans* (*C. elegans*).[74] For example, the Heck group[54] reported a XL-MS study of whole human cell lysates in which they identified 2179 unique cross-links. These protein complexes (e.g., 80S ribosomal core complex) reveal novel interactions and provide new structural insights.

To summarize, the increasing number of publications based on XL-MS continue to demonstrate its utility in MS-based biophysics, particularly in combination of other techniques. Integrating XL with other methods enables characterization of dynamic biological systems even of heterogeneous systems. The continuing development of cross-linkers will increase applications. Reagents that react more rapidly and target more and more amino acids, instrumentation advances that provide new fragmentation methods in MS/MS (e.g., CID, ETD, EThCD and UVPD), and software are expected for the future. Furthermore, improved separation and enrichment procedures will accommodate the increasing mixture complexity following XL and digestion and allow detection of low abundant yet informative cross-linked species.

1.4.2 Chemical Cross-linkers

Chemical cross-linking occurs via the formation of covalent bonds between interacting proteins and within a protein on adjoining regions.[76-78] This structural proteomics tool often uses a bifunctional chemical reagent and can be viewed as bifunctional footprinting. Many reagents

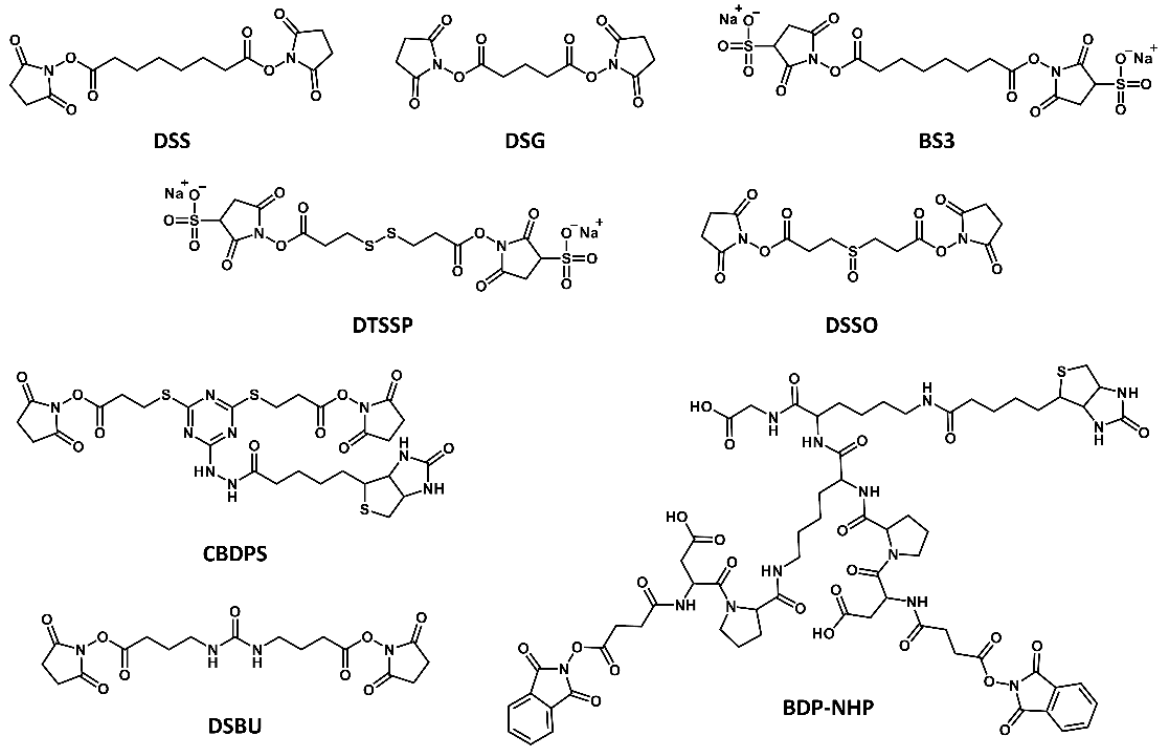
have been developed, promoting more and more applications of chemical cross-linking. The character of a cross-linking reagent is determined by the nature of chemistry, spacer length, and built-in functional groups. Cross-linking can be an effective complement to monofunctional reagents that serve as footprinters, and therefore the subject is briefly covered in this review particularly to make that point. The emphasis is on the commonly used reagents, which fits the theme of footprinting. Indeed, the location of monolinks can be viewed as a footprint. Our discussion is organized around the cross-linking reagents as follows.

1.4.2.1 Amine-reactive Cross-linkers: NHS-Ester and Imidoester

N-hydroxysuccinimide Ester

NHS-esters are the most widely used cross-linkers in field. The ester group undergoes attack by nearby nucleophilic sites (e.g., primary amines on Lys side chains or at the N-terminus of a protein/peptides, hydroxyl groups on Ser or Thr, and even sulfhydryl groups on Met) to form C-X bonds (X = N, O, S). The differing reactivities with various nucleophilic groups, however, depends on reaction conditions and can introduce bias. Primary amines possess the highest reactivity at physiological pH, and their reactivity can be further enhanced when the pH is > 7.[79-80] Hydrolysis of NHS-esters has a half-life of 4-5 hours at pH 7 and 0 °C[81]; the hydrolysis, however, is accelerated under alkaline conditions and at higher temperature.[82-83] Other possible products (i.e., esters and thioesters formed with Ser/Thr or Met) are less stable and undergo more rapid hydrolysis.[83] Although the reaction is favored under acidic conditions (pH = 6.0), cross-linking at pH = 7 still occurs. The current consensus is to consider all possible residues as sites for reactions in simple protein systems, whereas only Lys residues and the N-termini are recommended for cross-linking a large complex or in a whole proteome study to minimize dispersion of cross links and maximize probability for detecting the cross-links.[84]

Scheme 1.1 Common NHS-ester cross-linkers



The first NHS-ester originated in late 1970s as a homo-bifunctional cross-linker.[85] Since then, many NHS-esters possessing useful physical and chemical properties have become suitable for answering a range of biological questions. Designs include cross-linkers with high hydrophobicity and zero charge, being lipophilic and membrane-permeable and ideal for intramembrane cross-linking (e.g., DSS) and disuccinimidyl glutarate (DSG), Scheme 1.1). Other cross-linkers incorporate a sulfonate group, which imparts water-solubility and avoids steps of pre-dissolution in organic solvents that could perturb aqueous conditions and cause some protein denaturation.

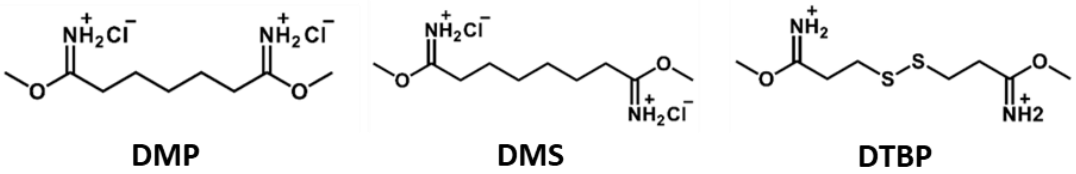
Currently, sulfo-NHS esters have become the dominant cross-linkers in characterizing soluble proteins and their interactions. *bis*(Sulfosuccinimidyl)suberate (BS3, Scheme 1.1), the most extensively used sulfo-NHS cross-linker, can be encoded with deuterium to facilitate better cross-link assignment. The spacer length is 11.4 Å, allowing cross-linking for residues separated

by 27 Å measured between two α carbons in an amino acid residues that are linked (termed Cα).[86]

New developments in NHS-esters resonate well with the rapid growth in proteomics research enabled by advanced MS. More complicated systems with larger-scale protein candidates require better sequencing in MS/MS for a confident cross-linking identification. The disadvantage of using conventional cross-linkers is that the user must sequence species made up of two cross-linked peptides and the cross-linker, introducing complexity into fragmentation and challenging the acquisition of high-quality, readily interpretable MS/MS data. Cleavable cross-linkers are, therefore, designed to address that problem.[68] NHS-esters have incorporated chemical-cleavable motifs (e.g., S-S bond in 3,3'-dithiobis(sulfosuccinimidylpropionate) (DTSSP)[79]) and CID-cleavable functional groups that include C-S bonds (e.g., disuccinimidyl sulfoxide (DSSO)[87] and cyanurbiotindimercaptopropionyl succinimide (CBDPS)[88]) and C-N bonds (e.g., disuccinimidyl dibutyric urea (DSBU)[89] and *N*-hydroxyphthalamide ester of biotin aspartate proline (BDP-NHP)[90]). Chemical structures of these cross-linkers are shown in Scheme 1.1. In addition, some NHS-crosslinkers contain biotin tags (e.g., CBDPS and BDP-NHP), to provide a means of enriching the crosslinked species by affinity purification.

Imidoester

Scheme 1.2 Common imidoester cross-linkers

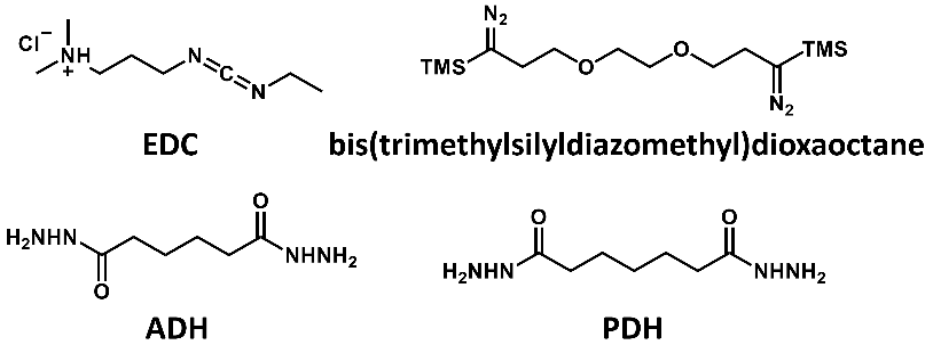


Imidoesters, introduced in 1966 as one of the oldest reagents for protein cross-linking, are water-soluble reagents that react specifically with primary amines (see Lys footprinting section

above).[91] The functional imidate group reacts to form an amide bond via intermediates at an optimized pH range of 8-10.[82] One major advantages of using imidoester cross-linkers is that the reaction product, an amidine, carries one positive charge. Charge removal that occurs with most lysine-targeting cross-linkers may disrupt intramolecular and intermolecular interactions and distort protein conformation, giving a biased result. The lifetime of imidoesters, however, is limited by rapid hydrolysis, being less than 30 min.[92-93] The most commonly used cross-linkers in this class are dimethyl pimelimidate (DMP), dimethyl suberimidate (DMS) and a cleavable analog, dimethyl 3,3′-dithiobispropionimidate (DTBP), as depicted in Scheme 1.2.

1.4.2.2 Carboxylic Acid-reactive Cross-linkers: Carbodiimide and Dihydrazides

Scheme 1.3 Common carbodiimide and dihydrazide cross-linkers



The most widely used carbomiimide is EDC (Scheme 1.3), also known as a “zero-length” cross-linker.[94-95] When this reagent facilitates cross-linking of carboxylate groups and primary amines, there is no spacer chain inserted between the targeted proteins, but rather an amide bond (~ 3 Å) between COOH and NH_2-containing sidechains.[94-95] One critical reaction intermediate, *O*-acylisourea, is not stable in aqueous condition (Scheme 1.6), and it continues to degrade back to the carboxyl group. Therefore, sulfo-NHS is often incorporated in the cross-linking protocol to transform the *O*-acylisourea into a stabilized NHS-ester for more efficient conjugation. EDC cross-linking shows the highest reactivity at pH 4.5 and still a moderate reaction efficiency at neutral pH.[76] 2-(*N*-Morpholino) ethanesulfonic acid (MES) and phosphate buffer are compatible

with carbodiimide reagents; however, more concentrated reagents should be used in the latter buffer owing to the reduced reactivity.

Other commonly used cross-linker for Asp and Glu are dihydrazides. In 2008, Kruppa and Novak[96] first reported its use in combination with EDC activation. The reaction was carried out in acidic conditions (pH = 5.5), which is not physiologically friendly, to give a low cross-linking yield. Later in 2014, Aebersold and coworkers[63] reported a new coupling reagent (other than EDC), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM), and achieved significantly increased reaction yields and better biocompatibility at neutral pH. Recently, Lei and coworkers[97] developed a coupling, reagent-free crosslinker, *bis*(trimethylsilyldiazomethyl)dioxaoctane (Scheme 1.3), that offers selectivity and efficiency under physiological conditions. Isotopic encoded dihydrazides are now commercially available with different spacer lengths (e.g., adipic acid dihydrazide (ADH) and pimelic acid dihydrazide (PDH), Scheme 1.3).

1.4.2.3 Sulfhydryl-reactive Cross-linkers: Maleimide

Maleimide cross-linkers react specifically with sulfhydryl groups at pH 6.5-7.5 to form stable thioethers.[98-99] Under alkaline conditions (pH > 8.5), primary amines are also possible targets but not Tyr and His.[100] One major concern in maleimide cross-linking is to conjugate free sulfhydryl groups, requiring prior reduction of existing disulfide bonds; breaking the -S-S- bond has high potential to distort protein native structure. A more common way of employing maleimide chemistry is in combination with NHS-esters, where both functional groups are coupled onto a heterobifunctional cross-linker. These cross-linkers can be incubated with a protein sample in a stepwise manner, reducing disulfide bonds after reaction with primary amines.[82] In addition, to

ensure the best cross-linking performance, many thiol-containing compounds (e.g., DTT and BME) should be eliminated in the reaction buffer.

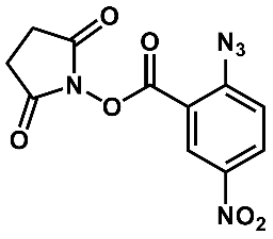
1.4.2.4 Carbonyl-reactive Cross-linkers: Hydrazide

Hydrazides are carbonyl-reactive reagents that exhibit highest reactivity at pH 5-7. Aldehydes and ketones are major targets that can be found in glycoproteins introduced by oxidation of the polysaccharide.[82] The hydrazone bonds thus formed are moderately stable in aqueous solution and can be further secured by reducing the double bond to a secondary amine.

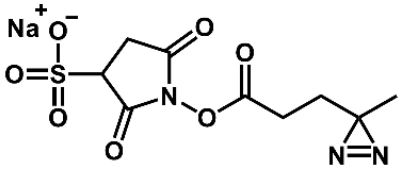
1.4.2.5. Photoreactive Cross-linkers: Aryl Azide and Diazirine

Photoreactive cross-linkers are generally nonspecific owing to the high reactivity of the intermediate species, a nitrene, carbene, or free radical. The most adopted chemistry utilizes nitrenes and carbenes, whose precursors are azides or diazirines, respectively. Among the three major categories, phenyl azides play a dominant role in current applications. Different substituents on the aromatic rings shift their UV absorption dramatically; therefore, they are selected with a biological question in mind.[82] Nitrophenyl azides, which can be activated at 300-460 nm, are compatible with most studies because this long wavelength causes minimal damage to protein molecules. Upon UV activation, the nitrene diradical can insert into most chemical bonds but with a preference for active C-H and N-H sites.[101] Diazirines are a relatively new class, first reported in the 1990s, and they have better photostability than phenyl azides.[102]

Scheme 1.4 Chemical structure of common photoreactive cross-linkers, ANB-NOS and sulfo-SDA



ANB-NOS



sulfo-SDA

Carbene diradicals are usually generated photochemically at ~ 355 nm, and they show reactivity with both single and double bonds. Like the nitrene diradical, heteroatom-H bonds undergo easier insertion. Some diazirine reagents are designed as analogs of amino acids (e.g., photo-Leu and photo-Met[103]) that can be incorporated into the protein sequence during translation for *in-situ* radical generation. High reactivity of carbenes and nitrenes can also cause problems in cross-link identification. The use of homo-bifunctional photoreactive cross-linkers tends to make product analysis complex; therefore, one radical precursor is usually combined with an amine- or sulfhydryl-reactive motif. The commonly used cross-linkers are *N*-5-azido-2-nitrobenzoyloxysuccinimide (ANB-NOS)[104] and sulfosuccinimidyl 4,4'-azipentanoate (sulfo-NHS-diazirine, sulfo-SDA), as seen in Scheme 1.4.

1.4.3 Conclusion and Perspective

Protein footprinting takes advantage of early research that provided many effective chemical reactions for protein labeling. These reagents can now serve as probes to characterize protein SASA and as a basis to reason about protein HOS. The sites of modification can now be efficiently measured by MS, owing to the rapid development of LC separations and hybrid mass spectrometers with MS/MS capabilities. The chemical cross-linker, which is viewed as a bifunctional protein footprinter, when connected to two proteins offers valuable distance restraints between the proteins. These targeted footprinting reagents coupled with MS detection now are utilized to answer several biological questions and are becoming increasingly significant in structural biology.

A limitation of targeted footprinting is that reagents primarily react with one or a few residues that contain functional groups including bases, acids, nucleophiles and aromatic rings, leaving a sizable fraction of the amino acid residues to be “silent”. Bulky reagents may be unable to

penetrate an interface, suggesting that the lack of reaction may not signal the lack of an interface. Moreover, the footprinting reactions are relatively slow, making most targeted footprinters unable to characterize fast protein dynamics (e.g., protein folding and protein aggregation). To overcome these drawbacks requires a new approach that has broader residue coverage and higher reaction rate. Fast footprinting reagents represented by reactive radical species meet these requirements as discussed in section 1.5.

1.5 Fast Labeling Reagents: Reactive Radical Species

1.5.1 Hydroxyl Radical

Hydroxyl radical ($HO^{\bullet}$), a reactive oxygen species (ROS), forms naturally in the cellular environment through partial reduction of oxygen. The ROS levels determine cellular oxidative stress, which mediates the modification of nucleic acids, proteins, and lipids, regulating signaling pathways and contributing to multiple diseases[105-107]. The hydroxyl radical as a footprinting reagent was initiated nucleic acid research[108] and promoted later for protein HOS analysis. Currently, $HO^{\bullet}$ is the most widely used radical-based footprinter given its many advantages. First of all, the size and hydrophobicity of $HO^{\bullet}$ are comparable with the water molecular, making it readily accessible to the exposed protein surface. Secondly, its natural presence in biological fluids demonstrates the comparability of $HO^{\bullet}$ with most of the protein systems. In addition, being a relatively strong oxidant (E_0 = 2.30 V), $HO^{\bullet}$ can react with most of amino acids, providing broad residue coverage and comprehensive SASA information. The reaction rate with different amino acids, however, range over a few orders of magnitude as seen in earlier studies[109-111] with free amino acids (Table 1.1).

Table 1.1 Rate constants for reactions between amino acids and $HO^{\bullet}$[109-110]

Free Amino Acid	Rate Constant ($M^{-1}s^{-1}$)	pH[a]
Cys	3.5×10^{10}	7.0
Trp	1.3×10^{10}	6.5 – 8.5
Tyr	1.3×10^{10}	7.0
Met	8.5×10^{9}	6 – 7
Phe	6.9×10^{9}	7 – 8
His	4.8×10^{9}	7.5
Arg	3.5×10^{9}	6.5 – 7.5
cystine	2.1×10^{9}	6.5
Ile	1.8×10^{9}	6.6
Leu	1.7×10^{9}	~6
Val	8.5×10^{8}	6.9
Pro	6.5×10^{8}	6.8
Gln	5.4×10^{8}	6.0
Thr	5.1×10^{8}	6.6
Lys	3.5×10^{8}	6.6
Ser	3.2×10^{8}	~6
Glu	2.3×10^{8}	6.5
Ala	7.7×10^{7}	5.8
Asp	7.5×10^{7}	6.9
Asn	4.9×10^{7}	6.6
Gly	1.7×10^{7}	5.9

1.5.1.1 Fenton Chemistry

In 1890s, H.J.H Fenton discovered the first oxidation of tartaric acid by hydrogen peroxide under ferrous iron (II) catalysis. [112] The "Fenton-like" reactions were further promoted by other transition metals, for example, Fe(II), Cu(I). [113-116] The well-recognized mechanism of Fenton's reaction is the listed "classical Fenton pathways" (eq 1.2-1.8), proposed initially by Haber and Weiss [117-118] and later revised by Barb and co-workers[119-120]. The chain reactions start from Fe(II) and Fe(III) catalysis (eq 1.2 and 1.3) and terminate by oxidation of Fe (II) to Fe (III) (eq 1.8). The reduction and oxidation of ferrous and ferric ions, together with their reactions with H_2O_2, contribute to the propagating generation of $HO^{\bullet}$ (eq 1.4 – 1.7).

$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + HO^{\bullet} + OH^{-}$ (1.2)

$Fe^{3+} + H_2O_2 \rightarrow Fe^{2+} + HO_2^{\bullet} + H^{+}$ (1.3)

$H_2O_2 + HO^{\bullet} \rightarrow HO_2^{\bullet} + H_2O$ (1.4)

$HO_2^{\bullet} \leftrightarrow O_2^{-\bullet} + H^+$ (1.5)

$Fe^{3+} + HO_2^{\bullet} \rightarrow Fe^{2+} + O_2 + H^+$ (1.6)

$Fe^{3+} + O_2^{-\bullet} \rightarrow Fe^{2+} + O_2$ (1.7)

$Fe^{2+} + HO^{\bullet} + H^+ \rightarrow Fe^{3+} + H_2O$ (1.8)

Hydroxyl radicals generated from Fenton chemistry were first used for footprinting in 1985, when Tullius and Dombroski [121-123] mapped protein binding sites on DNA. The well-designed Fenton system contains H_2O_2, Fe(II)-EDTA complex, which can increase the metal ion concentration and allow favorable formation of Fe(III)-EDTA during propagation, and ascorbate, which reduces Fe(III) back to Fe(II) to complete a catalytic cycle. Such design established a benchmark approach for nucleic acid footprinting and was further expanded with other transition metals[124-129] including Cu(II), Mn(II), Co(II), and Cd(II).

In 1990, the Fenton system was improved by employing a synthetic metal-chelate complex, Fe-(*S*)-1-(*p*-bromoacetimidobenzyl)-EDTA (Fe-BABE) [130-132]. The Fe-BABE tethers the metal and catalyzes $HO^{\bullet}$ formation at specific sites of a protein, for example, the location where cysteine residues reside. In addition, the spacer component in the chelating complex delivers spatial information of the interacting biomolecules, opening a possibility to capture conformational changes. The Fe-BABE system has many advantages compared to the traditional Fenton systems, including higher $HO^{\bullet}$ productivity at neutral pH and the capability of site-directed oxidation. Its limitations, however, are those of all "Fenton-like" reactions: long timescale for the radical chemistry (usually minutes), the possibility that conformational changes occur during labeling, and the large amount of H_2O_2 (30%) [133] that can be detrimental to targeted biomolecules.

1.5.1.2 Synchrotron Radiolysis of Water

Synchrotron X-ray is a powerful source to generate $HO^{\bullet}$ through radiolysis of water. The generated photons transfer energy to electrons, giving hydrated electrons (e_{aq}^-). Then, the ionized water reacts with another to produce $HO^{\bullet}$ (eq 1.9 - 1.11).[110] The generated $HO^{\bullet}$ undergoes self-quenching (eq 1.12) or accepts a hydrated electron to form a hydroxyl anion (eq 1.13).[134]

$2H_2O + e^-$ (ionizing irradiation) $\rightarrow H_2O^{+\bullet} + H_2O^* + e_{aq}^-$ (1.9)

$H_2O^{+\bullet} + H_2O \rightarrow HO^{\bullet} + H_3O^+$ (1.10)

$H_2O^* \rightarrow HO^{\bullet} + H^{\bullet}$ (1.12)

$HO^{\bullet} + e_{aq}^- \rightarrow OH^-$ (1.12)

$2\ HO^{\bullet} \rightarrow H_2O_2$ (1.13)

Chance and coworkers first put the National Synchrotron Light Source at Brookhaven National Laboratory to use for footprinting in both nucleic acids and protein systems. [135-137] In a radiolysis experiment, $HO^{\bullet}$ dosimetry, the effect of beam current, the effect of different buffers and additives need to be carefully considered. A detailed discussion is available in a review article.[110] Hydroxyl radical footprinting enabled by synchrotron radiolysis of water has significant advantages. Water, as the radical precursor, is physiological-related and compatible with almost all bio-systems. In addition, the amount of water (~55 M) is in large excess with respect to the targeted biomolecules, maximizing labeling probability. What's more, the radical dosage can be easily controlled by a shutter electronically[138], enabling the millisecond formation of $HO^{\bullet}$. The disadvantage, however, is the limited access to a synchrotron sources, which can be overcome in part by the development of new beamlines.

1.5.1.3 Laser Photolysis

Upon UV irradiation, homolytic cleavage of H_2O_2 into $HO^{\bullet}$ will occur at the quantum yield of 0.4-0.5. The subsequent chain reactions follow the Haber-Weiss cycle (Table 1.2) and terminate by self-quenching.

Table 1.2 Reaction scheme and rate constant of hydroxyl radical generation from hydrogen peroxide upon laser photolysis

	Reaction	**Rate Constant ($M^{-1}s^{-1}$)**
1	$H_2O_2 + hv \rightarrow 2\ HO^{\bullet}$	$\Phi_{248} = 0.4\text{-}0.5$ [139-140]
2	$HO^{\bullet} + H_2O_2 \rightarrow H_2O + HO_2^{\bullet}$	2.7×10^7 [140]
3	$HO_2^{\bullet} + H_2O_2 \rightarrow H_2O + O_2 + HO^{\bullet}$	3.06×10^5 [140]
5	$HO^{\bullet} + HO^{\bullet} \rightarrow H_2O_2$	7.0×10^9 [141]

Although UV photolysis of H_2O_2 had been accepted in some industrial process (e.g., water treatment) the first attempt for protein footpriting was reported in 2004 by Sharp and his co-workers[142]. They utilized a UV lamp to photolyze 15% H_2O_2 in protein solutions for 5 mins. The mapped SASA of lysozyme and β-lactoglobulin are consistent with literature structures, demonstrating the success as a footprinter. However, the high H_2O_2 concentration and long exposure time are of concern for broader applications.

To overcome the aforementioned limitations, Aye and Sze[143] employed a pulsed Nd:YAG laser at 266 nm and photolyzed H_2O_2 in a static system. Only 0.3% H_2O_2 was added just before the laser photolysis to minimize prior oxidation from the radical precursor. At the same time, Hambly and Gross[141] reported a KrF excimer-laser-based method (248 nm) that can generate $HO^{\bullet}$ and give subsequent labeling in a flow system. In addition, they incorporated a scavenger reagent such as Glu[141] and His[144] to control the primary radical lifetime at micro-seconds level. An experimental setup (Figure 2.1) and detailed discussion are in Chapter 2. In brief, the combination of a flow system and the utilization of a radical scavenger enables the protein of

interest to be irradiated and submitted to radical reactions only once per pulse of the laser, minimizing repetitive labeling on the protein that may be altered in conformation upon the initial oxidation. The oxidized protein samples will then be collected in a quenching buffer that contains Met and catalase to scavenge the H_2O_2 and avoid post oxidation.

The H_2O_2 photolysis has an obvious advantage comparing to synchrotron radiolysis, which is the portable laser source that is more accessible to any laboratory. The limitation, however, mainly relates to the oxidative nature of the H_2O_2 precursor, raising concerns for proteins that are prone to oxidation. Over the years, the fast photochemical oxidation of protein (FPOP) approach has successfully addressed many biological questions such as binding interfaces in a protein/protein or protein/ligand complex[145-146], protein aggregation[147], binding affinity [148-150], and protein dynamics[151-152]. The widespread adoption to industrial research, however, can be limited by associated safety concerns. Efforts are now undergoing to replace the laser source with alternative sources, for example a discharge lamp.

1.5.2 Carbenes

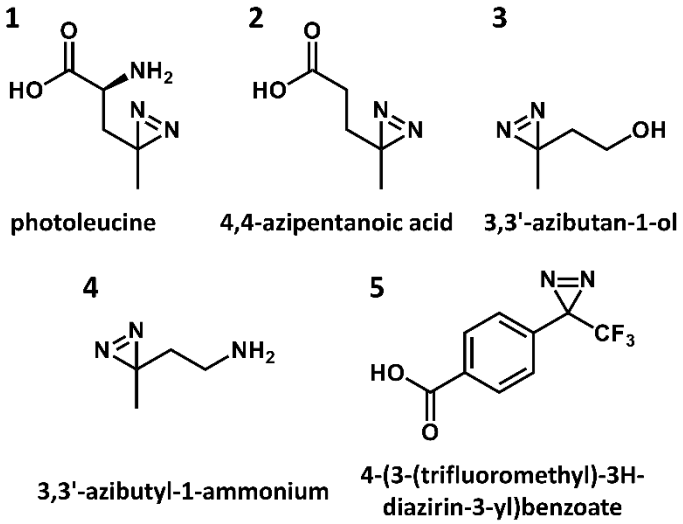
Diazirines, common precursor of carbene diradicals, were synthesized in 1960[153] and emerged as a versatile photoaffinity labeling (PAL) agents in the late 1970s.[154] PAL reagents usually are comprised of a binding motif and a reactive motif. Practically, the binding motif of PAL reagent reversibly binds to the active site of the target protein. Upon photoactivation, reactive motif is activated thus react with adjacent site, after which the PAL reagent is covalently attached to the target protein, serving as a footprint.[155-157] Investigators typically activate diazirines with a UV laser at approximately 350 nm to cause release of N_2 molecules and give an equimolar amount of carbene diradicals. Highly reactive carbenes form irreversible covalent bonds with proteins, allowing stringent downstream affinity purification and target identification.

The use of diazirine-based PAL is an effective strategy to understand protein-drug or other small molecule interactions and identify new drug binding sites. To address specific questions, investigators have designed and synthesized several diazirine analogs to adapt the affinity agent to the protein sample environment. For example, adamantylidene, a lipophilic reagent and an analog of adamantane, was used to obtain topological information of Na/K ATPase[158] and Ca-ATPase[159] in membranes as early as 1983. In addition, H-diaziflurane, an analog of halothane, allowed the examination of binding sites of inhaled anesthetics and their action mechanism.[160]

More recently, reagents containing diazirines were adopted as a new class in photo-activatable cross-linkers, and several were described.[76, 161] The ability to react with many bond types or amino acid residues makes carbenes powerful reagents to capture protein dynamics and intermolecular interactions. Obtaining multiple cross-links magnifies the information by providing more distance restraints on an interacting protein system, and thereby furnishing vital data for molecular simulation or docking to give a more complete description of the system.

Residue Specificity and Proposed Reaction Pathways

Scheme 1.5 Diazirine-based footprinting reagents



Richards et al.[162] first described the carbene diradical as a footprinting reagent in 2000. Methylene was generated from diazirine gas (CH_2N_2) upon UV irradiation and allowed to footprint α-lactalbumin. The labeling yield, however, was low, owing to the limited solubility of gaseous CH_2N_2 in aqueous media. Furthermore, the explosive gas requires conscientious preparation, storage, and safe handling, limiting wide application.

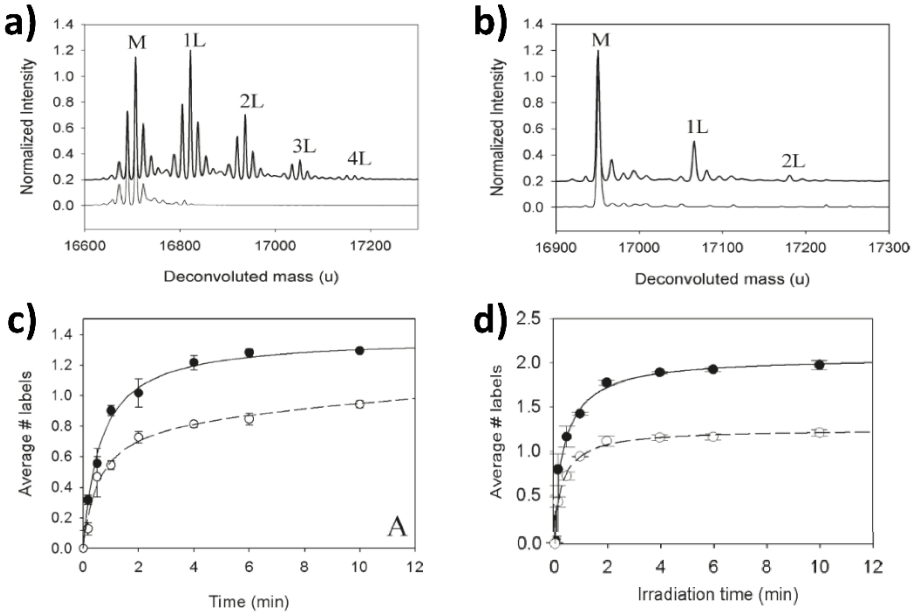


Figure 1.3 Photoleucine-derived carbene footprinting on hMyoglobin. (a) Deconvoluted mass spectra of myoglobin after footprinting with carbenes. (b) Deconvoluted mass spectra of labeled and unlabeled holo-CaM. (c) Ca^{2+}-binding induces conformational change on calmodulin where apo-calmodulin (closed circles) and holo-calmodulin (open circles) were labeled with 100 mM photoleucine in phosphate buffer and monitored as function of time. (d) Free holo-calmodulin (filled circles) is referenced to M13-bound holo-calmodulin (open circles). Reprinted with permission from Ref. [163]. Copyright 2011 American Chemical Society.

Later in 2011, Schriemer and coworkers[163] reported a new diazirine-based reagent, photoleucine (Reagent **1** in Scheme 1.5) that has higher water solubility and stability than CH_2N_2. The reaction platform incorporates a Nd:YAG pulsed laser (355 nm, 1000 Hz) for radical initiation and a 96-well plate cover with lids containing slits, allowing the laser beam to enter the protein solution.

With this experimental setup, the investigators obtained a maximum conversion of the diazirine to the carbene diradical by using an irradiation time of 2 min in the absence of other competitive chromophores. Furthermore, they found that photoleucine does not react with targeted proteins without laser activation. Upon photolysis, irreversibly labeled products exhibit a characteristic +115.03 Da mass shift (Figure 1.3a and 1.5b).

$$\Delta mass_{\text{av}} = \left[\frac{\sum m_i I_i}{\sum I_i}\right]_{\text{labeled}} - M_{\text{unlabeled}} \qquad (1.14)$$

To examine the labeling sensitivity of carbene diradicals for different accessible surfaces of proteins, two calmodulin systems (i.e., with/without Ca^+ and bound/unbound to the peptide M13) were investigated. The average number of labels on each protein molecule can be estimated by determining the difference in the centroid masses between the labeled and unlabeled proteins as seen in a deconvolved mass spectrum of the intact protein (eq. 1.14 and Figure 1.3c, d), where m_i and I_i represent mass and signal intensity, respectively. A reduction in protein surface area resulting from Ca^+ binding or M13 binding is reflected clearly as the measured labeling extent over 2-10 min of irradiation; namely holo-CaM is 45 ± 7% less labeled and M13-CaM is 39 ± 5% less than apo-CaM.

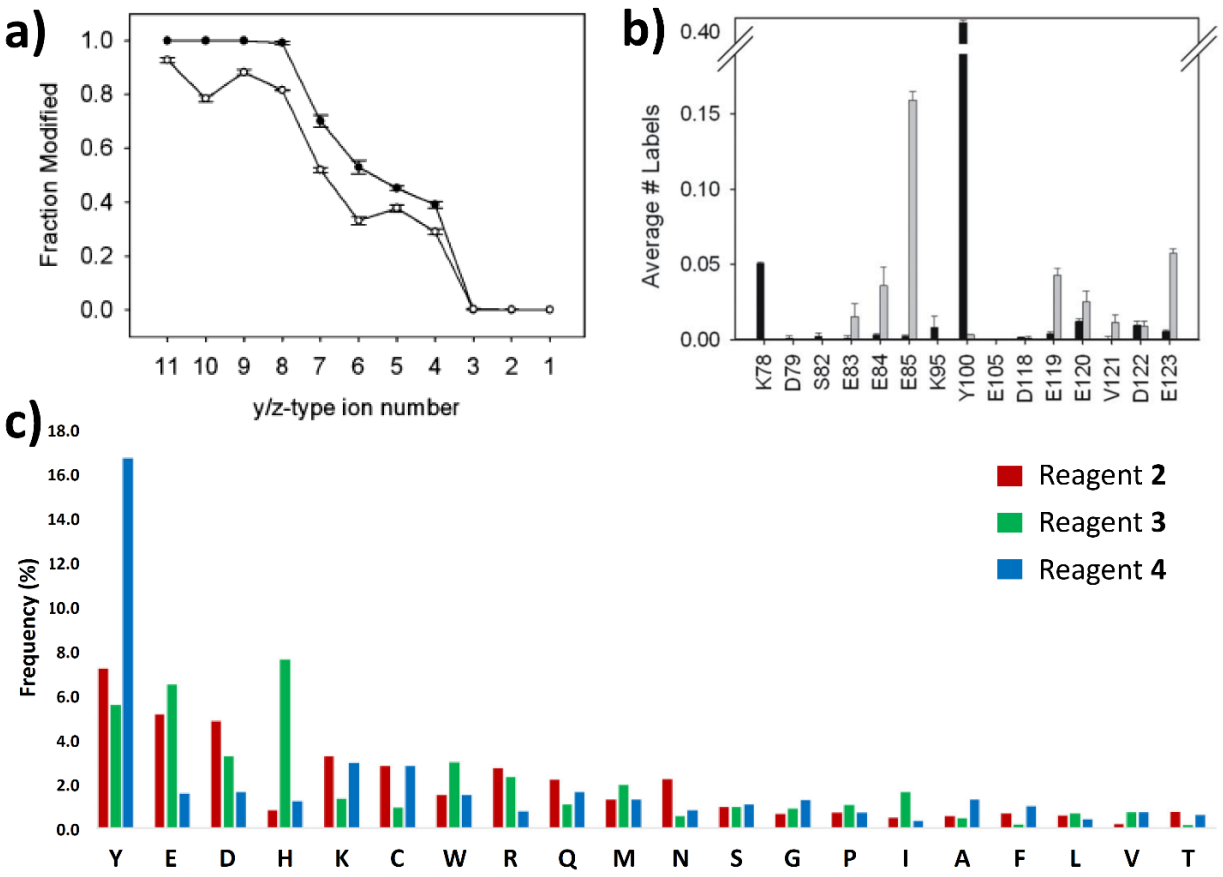


Figure 1.4 Carbene footprinting derived from multiple diazirine reagents. (a) Fractional distributions of carbene label derived from reagent 1 as determined by ETD (z-ions, filled circles) and HCD (y-ions, open circles) for peptide LTDEEVDEMIR (117−127) in CaM. (b) Residue-level reagent incorporation for select residues, based on ETD fragmentation of MKDTDSEEEIR (77−87), VFDKDGNGYISAAELR (92−107) and LTDEEVDEMIR (117−127) with Reagent 1 (gray bars) and Reagent 2 (black bars). Error bars are ± 1 standard deviation. Reprinted with permission from Ref.[164] Copyright 2012 American Chemical Society. (c) Average frequency of carbene insertion at each residue generated from the photolysis of reagent 2, 3, and 4 in the presence of protein digests (777 peptides). Site of the label insertion was located with MS/MS with a Fusion Lumos with EThcD fragmentation and analyzed with Mass Spec Studio software. Reprinted with permission from Ref.[165]. Copyright 2017 Springer Nature.

Residue-level quantification of carbene-induced modification needed attention, and that was discussed in a sequel study.[164] Theoretically, carbene diradicals can insert into X-H bonds and

C=C bonds.[166] Insertion into carboxylic acid functional groups will form labile esters that can be lost in CID fragmentation, thus complicating data interpretation and even losing information. The Shriemer group[164] chose ETD fragmentation as an alternative to CID and compared the two fragmentation methods by reporting the fraction modified for each y/z ion. Although most fragments shared the same trend, the y_{10} ion gave poor precision because it undergoes a neutral loss of the modifying group (Figure 1.4a). Notably, reducing the collision energy reduces the loss, but the abundance of the peptide fragments (product ion in MS/MS) is also reduced. ETD is a superior fragmentation method for retaining labile modifications to afford more comprehensive information than CID. ETD, however, performs poorly when peptides are low in charge and small in size. The best approach might be to use a combination of the two modes of MS/MS.

Another issue is the electrostatic interaction between the carbene precursor and various amino acid side chains; the interaction may concentrate the precursor molecule around the site, promoting more modification and a biased residue preference. Switching to another carbene precursor reagent, 4,4-azipentanoic acid (**2** in Scheme 1.5), that contains no positively charged amine group, shows the effect. Although both reagents give similar results for Tyr, Lys, Glu, and Asp, positive-charged photoleucine shows higher reactivity with the negatively charged aspartic acid (Figure 1.4b). In addition, the ionic strength of the buffer and solution temperature also affect the electrostatic interaction, where increasing ionic strength decreases electrostatic interactions[167] and higher temperatures will lower the dielectric constant[168].

A related approach was also reported by the Gross group[169] in 2015, when they adapted carbene generation in solution on the FPOP flow system. Careful control over the exclusion volume guarantees that photoleucine and the protein CaM are mainly irradiated once, thus diminishing concerns of perturbing the solution equilibrium by the generated nitrogen gas and causing

conformational change with excess labeling. The outcome is less modification for holo-CaM, consistent with its more compact conformation and significant labeling on Try, Asp and Glu, possibly owing to interactions of the protein with photoleucine, concentrating the reagent on the protein surface.

In 2017, Schriemer and coworkers[165] refined the carbene platform to employ a single-shot laser with higher energy (i.e., 150 mJ) to avoid nitrogen perturbation and protein conformational change at the induced air-water interface. Prior to irradiation, the sample solution was snap-frozen by liquid nitrogen before laser irradiation to restrict radical diffusion, to maintain protein HOS, and to minimize quenching of carbene radical and increase modification. To establish residue specificity with carbene chemistry, the investigators, in a Herculean study, footprinted 777 peptides by using three different precursors, a negatively charged precursor **2** (Scheme 1.5), a neutral reagent **3** (3,3'-azibutan-1-ol, Scheme 1.5), and a positively charged reagent **4** (3,3'-azibutyl-1-ammonium, Scheme 1.5). The labeling trends are similar for the three reagents, and the bond insertion propensities generally are a function of side-chain polarity and size. Reagents **2** and **3** both favor Arg, Glu and Asp, whereas, the neutral reagent **4** shows higher reactivity with His (aromatic and neutral) in addition to the three residues, Arg, Glu, and Asp. Remarkably, many hydrophobic amino acids show noticeable modification (Figure 1.4c), supporting the high reactivity of carbenes, even with aliphatic groups. The nature of the carbene precursor also impacts the residue selectivity by participating in complex molecular interactions to increase the local concentration of the precursor ion at the protein surface.

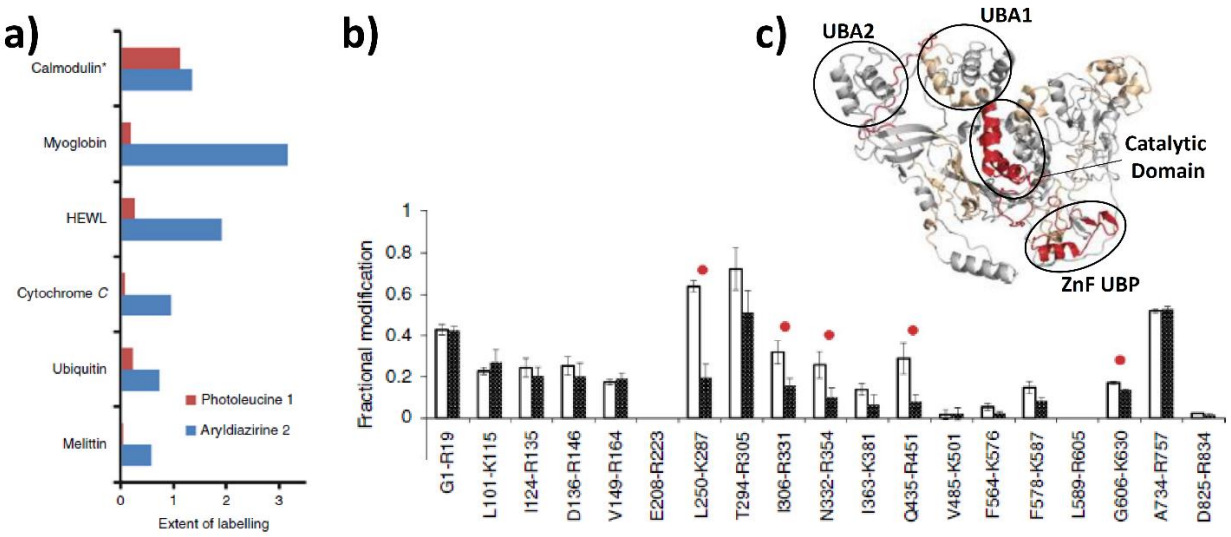


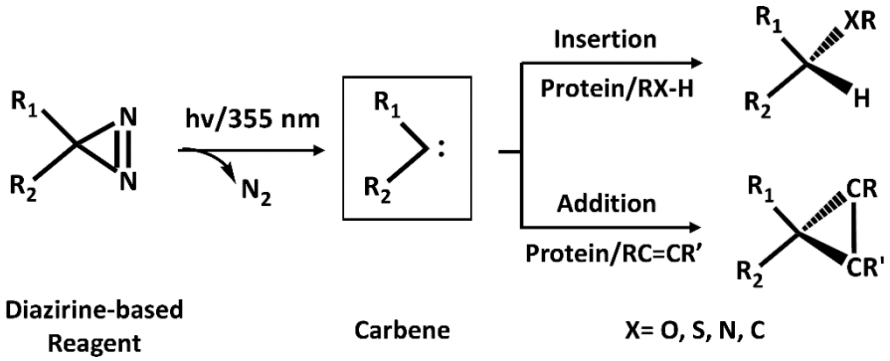
Figure 1.5 Carbene footprinting on FPOP platform. (a) Extent of labelling of a range of proteins with reagent 1 (100 mM, 16 s irradiation), and reagent 5 (10 mM, 4 s irradiation (*1 s in the case of CaM)). (b) Fractional modification by reagent 5 of USP5 peptides in the presence (black bars) and absence (white bars) of di-ubiquitin. Error bars are ± standard deviations (n = 3) and significant differences (Student's t-test, $p < 0.05$) are highlighted with a red dot. (c) Model of USP5 (based on PDB 3IHP) showing the locations of the five peptides (red) that are masked from labelling by di-ubiquitin binding and their relative locations to the ZnFUBP and catalytic domains. Reprinted with permission from Ref.[170] Copyright 2016 Springer Nature.

Manzi et al.[170] tested a carbene from the precursor, 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoate (reagent **5**, Scheme 1.5) that is more reactive than those from reagents **1** – **4.** The investigators' design involved installing an adjoining trifluoromethyl group (Figure 1.5a), leading to the use of less reagent and less irradiation time (i.e., 10 mM for 4 s irradiation). Further, the labeling efficiency improved in comparison to that with photoleucine at 100 mM and 16 s irradiation. The improved reactivity is due to increased stabilization of the carbene radical by the added trifluoromethyl group and the increased hydrophobicity but little change in zwitterionic character. They tested the new reagent on an unknown protein complex (i.e., the deubiquitinating enzyme ubiquitin specific protease 5 (USP5) upon binding with di-ubiquitin (di-Ub)). USP5 is a multi-domain cysteine protease including two ubiquitin associated domains

(UBA) and a Zn-finger ubiquitin-binding domain (ZnF-UBP). Previous studies showed the binding stoichiometry between USP5 and di-Ub to be 1:1, suggesting additional binding sites than those in Znf-UBP. These sites were not identified. The investigators footprinted the C335A mutant of USP5 with carbenes in the presence and absence of one equivalent di-Ub and observed distinct binding regions (Figure 1.5b), which were mapped onto the X-ray structure of USP5 (Figure 1.5c). The catalytic domain was shown to be the other binding site and, additionally, a remote conformational change for the region represented by peptide G606-K630 was found. The design of new, successful radical precursors indicates that there are more opportunities for improvement and application, emphasizing the potential of carbene footprinting as an effective and accurate structural probe for HOS of proteins.

The pathways for carbene chemistry may involve several radical intermediates whose structures and reactivities can be tailored by using different precursors. For example, it is possible to generate by photolysis not only singlet[171] and triplet carbenes[172] but also diazo isomers that further decompose into carbocations.[173,174] Although a singlet carbene preferentially inserts into O-H, N-H and S-H bonds,[173] it is challenging to pinpoint the dominant pathway just from the nature of the modified products. Insertion into Thr can be done by singlet carbene through O-H bond insertion or by triplet insertion into the methylene group located on the amino acid side chain. In addition, the small energy difference (i.e., ~ 2 kcal/mol[175]) between the two states add complications because mixtures of products can form. Because heteroatom-containing residues are usually the favored sites of reaction (e.g., Glu, Asp, Tyr and Arg), the singlet state may be favored; however, blended pathways are more likely. A general route is shown in Scheme 1.6

Scheme 1.6 Proposed carbene reaction pathway



Carbene footprinting has high potential in structural biology. The labeling time is shorter than 10 ns,[176] faster than most protein folding. Irradiation on a flow-system or excitation after snap freezing eliminates deceptive modifications originating from carbene insertions in a protein that has undergone a protein conformational change. Compared to the hydroxyl radical, the short survival time of carbenes owing to reaction with solvent water obviates the need for a scavenger. Carbene generation by diazirines is at a less damaging wavelength to proteins (i.e., ~ 350 nm). In addition, most carbene precursors do not react with proteins prior to laser irradiation (unlike H_2O_2 for $HO^{\bullet}$), and this lack of reactivity minimizes background interference. Once a carbene inserts, there are no reactive biproducts as there are when a radical reacts. Furthermore, the resultant mass shift for carbene modification can be adjusted to be bio-orthogonal by tailoring the precursor design. The physical properties of reagents, however, can favor preconcentration on the surface of the protein, possibly delivering biased residue preference or even information loss. Those properties can also be chosen advantageously to promote binding in lipid membranes, permitting footprinting of transmembrane proteins. As this field develops and more diverse carbene reagents are implemented, a better understanding of interactions will emerge to permit rational design of new carbenes. New footprinting reagents that can target specific residues or provide comprehensive coverage are expected in the future.

1.5.3 Other Radical Reagents

1.5.3.1 Sulfate radical anion

Sulfate radical anion, $SO_4^{-\bullet}$, is a potent oxidant with a standard reduction potential of 2.43 V at neutral pH.[177] Its strong oxidation capability can cause considerable damage to different cellular components including lipids, carbohydrates, proteins and DNA/RNA.[178]

Bridgewater and Vachet [179-181] in 2005-2006 first used sulfate radical anion to determine binding and map the SASA of proteins. In 2010, the Gross group[182] generated sulfate radical anion on the FPOP platform by photolysis of $^{-}OSO_2\text{-}O\text{-}O\text{-}O_2SO^{-}$ at a quantum yield of 0.55. The reactivity and specificity of $SO_4^{-\bullet}$ is similar to those of $HO^{\bullet}$ for residues Met, Trp Glu and Ser; the reactant radical favors His and Tyr . The overall reactivity ranking of $SO_4^{-\bullet}$ is:

Met > Tyr = Trp > Phe = Glu = His > Ser > Pro > Asp = Thr > Lys = Gln > Leu = Val = Ile

1.5.3.2 Trifluoromethyl Radical

Fluorine-containing compounds are extremely rare in biology; only five entities containing F have ever been identified.[183] The most common molecule is fluoroacetate, which is found in many tropical plants as a toxin. Footprinting reactions that insert either fluorine or fluorine-containing substrates may be advantageous because fluorine is the most electronegative[184] (Pauling Electronegativity = 4.0) common substance and has a small radius (1.33 Å), not so dissimilar to that of H,[185] allowing F to be a surrogate for H.

Trifluoromethyl radical as a footprinting reagent, was implemented by Cheng et al.[186] in 2017 to be used on the FPOP platform. The radical precursor is the water-soluble salt, $NaSO_2CF_3$ (Langlois reagent). The $^{\bullet}CF_3$ formation is initiated by the $HO^{\bullet}$, from photolysis of hydrogen peroxide. A likely mechanism is the $HO^{\bullet}$ displaces $^{\bullet}CF_3$ by attack on the S=O bond to form $HOSO_2^{-}$, the conjugate base of sulfurous acid. The $^{\bullet}CF_3$ can react with 18 out of 20 different

amino-acids residues, except Met and Cys, showing its complementary nature with $HO^{\bullet}$, which reacts rapidly with Met and Cys. Combination of $^{\bullet}CF_3$ and $HO^{\bullet}$ in tandem allows more comprehensive characterization of the residues on a targeted protein than does either radical alone, providing a better opportunity to capture subtle structural changes.

1.5.3.3 Iodine Radical

In physiology, iodine plays essential roles in metabolic regulation of thyroid function, especially hormone production.[187-188] The iodine radical has potential to be an effective fooprinting reagent given its specific reactivity towards tyrosine and histidine. This chemistry was initiated on an FPOP platform, as shown by Chen et al.[144] using 4-iodobenzoic acid as the precursor. The $I^{\bullet}$ is formed presumably in concert with a $^{\bullet}C_6H_4COOH$ by photolysis of $I\text{-}C_6H_4\text{-}COOH$ at 248 nm. The carboxylphenyl radical likely abstracts an $H^{\bullet}$ from the OH of Tyr or from the NH of the imidazole ring to give a stabilized radical that is subsequently "capped" by reaction with $I^{\bullet}$ to give an iodinated protein, although there may be other mechanisms.

Compared to $HO^{\bullet}$ footprinting, the unique and larger mass shift of iodination increases the confidence of assigning the modification sites. Although the coverage afforded by the iodide radical is limited, footprinting two targeted residues can answer specific questions with easier data analysis. In addition, the precursor of the iodine radical doesn't react with proteins, which minimizes the background modification and allows simplified post-label sample handling. Unlike specific amino acid labeling, the modifications on the FPOP platform are fast, alleviating concerns about labeling-induced conformational changes.

1.5.4 Conclusion and Perspective

Radical footprinting is an effective means to acquire HOS information for proteins. Radical reactions are much faster (nano to milliseconds) than reactions of a conventional chemical

reagent that modifies one or a few amino acid residues. Therefore, free radical footprinting will deliver fast, broad, and less biased information by largely avoiding questions of labeling-induced perturbation of protein structure. In addition, the short timescale can allow a sensitive report of subtle structural and even dynamical changes. An advantage of the labeling speed is the successful temperature jump, two-laser experiment on the FPOP platform where $HO^{\bullet}$ footprinting can successfully track the folding of barstar at times as short as a few tenths of millisecond.[189] An example of versatility is the tracking of multiple intermediate oligomeric states during amyloid beta aggregation.[190]

Different radicals (e.g., $^{\bullet}CF_3$, $I^{\bullet}$, $SO_4^{-\bullet}$, $HO^{\bullet}$) allow irreversible covalent modifications to occur on the side chain of an amino acid with various residue specificity. To address a particular biological question, a radical reagent with preferred residue selectivity can be an appropriate choice over one with larger residue coverage. For example, using carbene footprinting in the protein systems that are rich in Glu and Asp (e.g., as for calcium-binding proteins) will enable quick and simple data analysis and high throughput. Thus, a footprinting "toolbox" with diverse radical-based reagents is of great interest to our group.

1.6 Conclusion

During the past two decades, we have seen extensive development of the targeted reagents, including chemical cross-linkers, and fast radical labeling reagents, enabled by improved-design. experimental set-ups, well-tailored chemistry, evolving MS instrumentation, and new data processing software.

Chemical cross-linkers are bifunctional protein footprinters, which are governed by similar principles as protein footprinting by targeted reagents. The bi-functional nature of the reagent allows generating distance restraints between two cross-linked residues, and such restraints can

be further utilized to locate protein/protein binding interfaces, to assess protein topologies, to characterize a protein's interactome in a large complex, and to facilitate protein docking (applications in binding interface determination are demonstrated in Chapter 4 and 5) and modeling (an application in structure prediction is shown in Chapter 6).

Fast radical labeling, an approach that not only allows irreversible modification on side chains of amino acid but also permits rapid labeling chemistry, is another essential means to assess protein HOS. Establishing a footprinting "toolbox" with diverse reagents can provide ready availability to researchers to address various biological questions (development of a novel radial reagent is discussed in Chapter 2). In addition, the elaboration of the FPOP platform that integrates two radical footprinters is also an intriguing subject for future research. Taking advantages of the rapid generation and reactions of a radical species, one radical can trigger protein conformational changes, for example by inducing a pH change, and the other can footprint the protein. When changing the time gap between the generation of the two radicals, the folding/unfolding process, therefore, can be captured in time (some method development is discussed in Chapter 3), providing an opportunity to study protein dynamics with even residue-level spatial resolution.

1.7 References

1. Lodish, H.; Berk, A.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Bretscher, A.; Ploegh, H.; Matsudaira, P. Molecular cell biology. Macmillan: 2008.

2. Drenth, J. Principles of protein X-ray crystallography. Springer Science & Business Media: 2007.

3. Wüthrich, K. NMR in biological research: peptides and proteins. North-Holland Amsterdam: 1976.

4. Wuthrich, K. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* **1989**, *243*, 45-50.

5. Wüthrich, K. The way to NMR structures of proteins. *Nat. Struct. Biol.* **2001**, *8*, 923-925.

6. Bai, X.-c.; McMullan, G.; Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **2015**, *40*, 49-57.

7. Cheng, Y. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* **2015**, *161*, 450-457.

8. Williamson, M. P.; Havel, T. F.; Wüthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **1985**, *182*, 295-315.

9. Wüthrich, K. NMR with proteins and nucleic acids. *Europhys. News* **1986**, *17*, 11-13.

10. Yu, X.; Jin, L.; Zhou, Z. H. 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* **2008**, *453*, 415.

11. Zhang, X.; Jin, L.; Fang, Q.; Hui, W. H.; Zhou, Z. H. 3.3 Å Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry. *Cell* **2010**, *141*, 472-482.

12. Thompson, R. F.; Walker, M.; Siebert, C. A.; Muench, S. P.; Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **2016**, *100*, 3-15.

13. Sheshberadaran, H.; Payne, L. G. Protein antigen-monoclonal antibody contact sites investigated by limited proteolysis of monoclonal antibody-bound antigen: protein "footprinting". *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1-5.

14. Hvidt, A.; Linderstrøm-Lang, K. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochim. Biophys. Acta* **1954**, *14*, 574.

15. Saunders, M.; Wishnia, A.; Kirkwood, J. G. The nuclear magnetic resonance spectrum of ribonuclease 1. *J. Am. Chem. Soc.* **1957**, *79*, 3289-3290.

16. Saunders, M.; Wishnia, A. Nuclear magnetic resonance spectra of proteins. *Ann. N. Y. Acad. Sci.* **1958**, *70*, 870-874.

17. Wishnia, A.; Saunders, M. The Nature of the Slowly Exchanging Protons of Ribonuclease. *J. Am. Chem. Soc.* **1962**, *84*, 4235-4239.

18. Glickson, J. D.; McDonald, C. C.; Phillips, W. D. Assignment of tryptophan indole NH proton resonances of lysozyme. *Biochem. Biophys. Res. Commun.* **1969**, *35*, 492-498.

19. Wagner, G.; Wüthrich, K. Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution: Studies with two-dimensional nuclear magnetic resonance. *J. Mol. Biol*. **1982**, *160*, 343-361.

20. Katta, V.; Chait, B. T.; Carr, S. Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 214-217.

21. Zhang, Z.; Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. *Protein Sci.* **1993**, *2*, 522-531.

22. Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **1993**, *17*, 75-86.

23. Smith, D. L.; Deng, Y.; Zhang, Z. Probing the Non-covalent Structure of Proteins by Amide Hydrogen Exchange and Mass Spectrometry. *J. Mass Spectrom.* **1997**, *32*, 135-146.

24. Pauling, L.; Corey, R. B.; Branson, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 205-211.

25. Englander, S. W.; Sosnick, T. R.; Englander, J. J.; Mayne, L. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **1996**, *6*, 18-23.

26. Hvidt, A.; Nielsen, S. O. Hydrogen Exchange in Proteins. In Advances in Protein Chemistry, Anfinsen, C. B.; Anson, M. L.; Edsall, J. T.; Richards, F. M., Eds. Academic Press: 1966; Vol. 21, pp 287-386.

27. Konermann, L.; Tong, X.; Pan, Y. Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *J. Mass Spectrom.* **2008**, *43*, 1021-1036.

28. Konermann, L.; Rodriguez, A. D.; Sowole, M. A. Type 1 and Type 2 scenarios in hydrogen exchange mass spectrometry studies on protein–ligand complexes. *Analyst* **2014**, *139*, 6078-6087.

29. Hamuro, Y.; Coales, S. J.; Southern, M. R.; Nemeth-Cawley, J. F.; Stranz, D. D.; Griffin, P. R. Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry. *J. Biomol. Tech.* **2003**, *14*, 171-182.

30. Pascal, B. D.; Willis, S.; Lauer, J. L.; Landgraf, R. R.; West, G. M.; Marciano, D.; Novick, S.; Goswami, D.; Chalmers, M. J.; Griffin, P. R. HDX Workbench: Software for the Analysis of H/D Exchange MS Data. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1512-1521.

31. Marciano, D. P.; Dharmarajan, V.; Griffin, P. R. HDX-MS guided drug discovery: small molecules and biopharmaceuticals. *Curr. Opin. Struct. Biol.* **2014**, *28*, 105-111.

32. 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*, 95-102.

33. Deng, B.; Lento, C.; Wilson, D. J. Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development – A review. *Anal. Chim. Acta* **2016**, *940*, 8-20.

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

35. Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. Differential hydrogen/deuterium exchange mass spectrometry analysis of protein–ligand interactions. *Expert Rev. Proteom.* **2011**, *8*, 43-59.

36. Gallagher, E. S.; Hudgens, J. W. Chapter Fourteen - Mapping Protein–Ligand Interactions with Proteolytic Fragmentation, Hydrogen/Deuterium Exchange-Mass Spectrometry. In Methods in Enzymology, Kelman, Z., Ed. Academic Press: 2016; Vol. 566, pp 357-404.

37. Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry—A review. *Anal. Chim. Acta* **2012**, *721*, 7-21.

38. Rahman, I. R.; Acedo, J. Z.; Liu, X. R.; Zhu, L.; Arrington, J.; Gross, M. L.; van der Donk, W. A. Substrate Recognition by the Class II Lanthipeptide Synthetase HalM2. *ACS Chem. Biol.* **2020**, *15*, 1473-1486.

39. Englander, S. W.; Mayne, L.; Kan, Z.-Y.; Hu, W. Protein Folding—How and Why: By Hydrogen Exchange, Fragment Separation, and Mass Spectrometry. *Ann. Rev. Biophys.* **2016**, *45*, 135-152.

40. Georgescauld, F.; Wales, T. E.; Engen, J. R. Hydrogen deuterium exchange mass spectrometry applied to chaperones and chaperone-assisted protein folding. *Expert Rev. Proteom.* **2019**, *16*, 613-625.

41. Englander, J. J.; Del Mar, C.; Li, W.; Englander, S. W.; Kim, J. S.; Stranz, D. D.; Hamuro, Y.; Woods, V. L. Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7057-7062.

42. Kan, Z.-Y.; Walters, B. T.; Mayne, L.; Englander, S. W. Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16438-16443.

43. Hamuro, Y.; Zhang, T. High-Resolution HDX-MS of Cytochrome c Using Pepsin/Fungal Protease Type XIII Mixed Bed Column. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 227-234.

44. López-Ferrer, D.; Petritis, K.; Lourette, N. M.; Clowers, B.; Hixson, K. K.; Heibeck, T.; Prior, D. C.; Paša-Tolić, L.; Camp, D. G.; Belov, M. E.; Smith, R. D. On-line Digestion System for Protein Characterization and Proteome Analysis. *Anal. Chem.* **2008**, *80*, 8930-8936.

45. Jones, L. M.; Zhang, H.; Vidavsky, I.; Gross, M. L. Online, High-Pressure Digestion System for Protein Characterization by Hydrogen/Deuterium Exchange and Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 1171-1174.

46. Rand, K. D.; Zehl, M.; Jensen, O. N.; Jørgensen, T. J. D. Protein Hydrogen Exchange Measured at Single-Residue Resolution by Electron Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 5577-5584.

47. Landgraf, R. R.; Chalmers, M. J.; Griffin, P. R. Automated Hydrogen/Deuterium Exchange Electron Transfer Dissociation High Resolution Mass Spectrometry Measured at Single-Amide Resolution. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 301-309.

48. Huang, R. Y.-C.; Krystek Jr, S. R.; Felix, N.; Graziano, R. F.; Srinivasan, M.; Pashine, A.; Chen, G. In Hydrogen/deuterium exchange mass spectrometry and computational modeling reveal a discontinuous epitope of an antibody/TL1A Interaction, MAbs, Taylor & Francis: 2018; pp 95-103.

49. Anderson, K. W.; Gallagher, E. S.; Hudgens, J. W. Automated Removal of Phospholipids from Membrane Proteins for H/D Exchange Mass Spectrometry Workflows. *Anal. Chem.* **2018**, *90*, 6409-6412.

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

51. Liu, X. R.; Zhang, M. M.; Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chem. Rev.* **2020**, *120*, 4355-4454.

52. Yang, B.; Wu, Y.-J.; Zhu, M.; Fan, S.-B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y.-X.; Chen, H.-F. Identification of cross-linked peptides from complex samples. *Nat. Methods* **2012**, *9*, 904.

53. Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L. N.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **2008**, *5*, 315.

54. Liu, F.; Rijkers, D. T.; Post, H.; Heck, A. J. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **2015**, *12*, 1179.

55. Götze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; Krauth, F.; Fritzsche, R.; Kühn, U.; Sinz, A. StavroX—a software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 76-87.

56. Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5802-5806.

57. Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T. The molecular architecture of the nuclear pore complex. *Nature* **2007**, *450*, 695.

58. Shi, Y.; Fernandez-Martinez, J.; Tjioe, E.; Pellarin, R.; Kim, S. J.; Williams, R.; Schneidman-Duhovny, D.; Sali, A.; Rout, M. P.; Chait, B. T. Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol. Cell. Proteom.* **2014**, *13*, 2927-2943.

59. Erzberger, J. P.; Stengel, F.; Pellarin, R.; Zhang, S.; Schaefer, T.; Aylett, C. H.; Cimermančič, P.; Boehringer, D.; Sali, A.; Aebersold, R. Molecular architecture of the 40S· eIF1· eIF3 translation initiation complex. *Cell* **2014**, *158*, 1123-1135.

60. Herzog, F.; Kahraman, A.; Boehringer, D.; Mak, R.; Bracher, A.; Walzthoeni, T.; Leitner, A.; Beck, M.; Hartl, F.-U.; Ban, N.; Malmström, L.; Aebersold, R. Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. *Science* **2012**, *337*, 1348-1352.

61. Hofmann, T.; Fischer, A. W.; Meiler, J.; Kalkhof, S. Protein structure prediction guided by crosslinking restraints–A systematic evaluation of the impact of the crosslinking spacer length. *Methods* **2015**, *89*, 79-90.

62. Gutierrez, C. B.; Yu, C.; Novitsky, E. J.; Huszagh, A. S.; Rychnovsky, S. D.; Huang, L. Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein–Protein Interactions. *Anal. Chem.* **2016**, *88*, 8315-8322.

63. Leitner, A.; Joachimiak, L. A.; Unverdorben, P.; Walzthoeni, T.; Frydman, J.; Förster, F.; Aebersold, R. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9455-9460.

64. Belsom, A.; Mudd, G.; Giese, S.; Auer, M.; Rappsilber, J. Complementary benzophenone cross-linking/mass spectrometry photochemistry. *Anal. Chem.* **2017**, *89*, 5319-5324.

65. Belsom, A.; Schneider, M.; Fischer, L.; Brock, O.; Rappsilber, J. Serum albumin domain structures in human blood serum by mass spectrometry and computational biology. *Mol. Cell. Proteom.* **2016**, *15*, 1105-1116.

66. Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **2016**, *41*, 20-32.

67. Schneider, M.; Belsom, A.; Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **2018**, *43*, 157-169.

68. Yu, C.; Huang, L. Cross-linking mass spectrometry: an emerging technology for interactomics and structural biology. *Anal. Chem.* **2017**, *90*, 144-165.

69. Zhang, M. M.; Beno, B. R.; Huang, R. Y. C.; Adhikari, J.; Deyanova, E. G.; Li, J.; Chen, G.; Gross, M. L. An Integrated Approach for Determining a Protein-Protein Binding Interface in Solution and an Evaluation of HDX Kinetics for Adjudicating Candidate Docking Models. *Anal. Chem.* **2019**, *91*, 15709-15717.

70. Komolov, K. E.; Du, Y.; Duc, N. M.; Betz, R. M.; Rodrigues, J. P. G. L. M.; Leib, R. D.; Patra, D.; Skiniotis, G.; Adams, C. M.; Dror, R. O.; Chung, K. Y.; Kobilka, B. K.; Benovic, J. L. Structural and Functional Analysis of a β2-Adrenergic Receptor Complex with GRK5. *Cell* **2017**, *169*, 407-421.e16.

71. Lin, S.-J.; Chen, Y.-F.; Hsu, K.-C.; Chen, Y.-L.; Ko, T.-P.; Lo, C.-F.; Wang, H.-C.; Wang, H.-C. Structural Insights to the Heterotetrameric Interaction between the Vibrio parahaemolyticus PirAvp and PirBvp Toxins and Activation of the Cry-Like Pore-Forming Domain. *Toxins* **2019**, *11*, 233.

72. Xu, H.; Hsu, P.-H.; Zhang, L.; Tsai, M.-D.; Freitas, M. A. Database Search Algorithm for Identification of Intact Cross-Links in Proteins and Peptides Using Tandem Mass Spectrometry. *J. Proteom. Res.* **2010**, *9*, 3384-3393.

73. Liu, F.; Lössl, P.; Scheltema, R.; Viner, R.; Heck, A. J. R. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **2017**, *8*, 15473.

74. Tan, D.; Li, Q.; Zhang, M.-J.; Liu, C.; Ma, C.; Zhang, P.; Ding, Y.-H.; Fan, S.-B.; Tao, L.; Yang, B.; Li, X.; Ma, S.; Liu, J.; Feng, B.; Liu, X.; Wang, H.-W.; He, S.-M.; Gao, N.; Ye, K.; Dong, M.-Q.; Lei, X. Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. *eLife* **2016**, *5*, e12509.

75. Yu, C.; Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* **2018**, *90*, 144-165.

76. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* **2006**, *25*, 663-682.

77. Leitner, A. Cross-linking and other structural proteomics techniques: how chemistry is enabling mass spectrometry applications in structural biology. *Chem. Sci.* **2016**, *7*, 4792-4803.

78. Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. *Mol. Cell. Proteom.* **2010**, *9*, 1634-1649.

79. Lomant, A. J.; Fairbanks, G. Chemical probes of extended biological structures: synthesis and properties of the cleavable protein cross-linking reagent [35S] dithiobis (succinimidyl propionate). *J. Mol. Biol.* **1976**, *104*, 243-261.

80. Staros, J. V. Membrane-impermeant crosslinking reagents: probes of the structure and dynamics of membrane proteins. *Acc. Chem. Res.* **1988**, *21*, 435-441.

81. Cuatrecasas, P.; Anfinsen, C. B. [31] Affinity chromatography. In Methods in enzymology, Elsevier: 1971; Vol. 22, pp 345-378.

82. Thermo Scientific. Thermo scientific crosslinking technical handbook. Waltham (USA): Thermo Scientific, 2012.

83. Hermanson, G. T.; Giese, R. Bioconjugate techniques. *J. Chromatogr. A.* **1996**, *746*, 303-303.

84. Iacobucci, C.; Piotrowski, C.; Aebersold, R.; Amaral, B. C.; Andrews, P.; Bernfur, K.; Borchers, C.; Brodie, N. I.; Bruce, J. E.; Cao, Y. First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study. *Anal. Chem.* **2019**, *91*, 6953-6961.

85. Bragg, P.; Hou, C. Subunit composition, function, and spatial arrangement in the Ca2+- and Mg2+-activated adenosine triphosphatases of Escherichia coli and Salmonella typhimurium. *Arch. Biochem. Biophys.* **1975**, *167*, 311-321.

86. Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine–lysine distances. *Protein Sci.* **2014**, *23*, 747-759.

87. Kao, A.; Chiu, C.-l.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol. Cell. Proteom.* **2011**, *10*, M110. 002212.

88. Petrotchenko, E. V.; Serpa, J. J.; Borchers, C. H. An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. *Mol. Cell. Proteom.* **2011**, *10*, M110. 001420.

89. Müller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schäfer, M.; Sinz, A. Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal. Chem.* **2010**, *82*, 6958-6968.

90. Weisbrod, C. R.; Chavez, J. D.; Eng, J. K.; Yang, L.; Zheng, C.; Bruce, J. E. In vivo protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *J. Proteom. Res.* **2013**, *12*, 1569-1579.

91. Hartman, F. C.; Wold, F. Bifunctional Reagents. Cross-Linking of Pancreatic Ribonuclease with a Diimido Ester1. *J. Am. Chem. Soc.* **1966**, *88*, 3890-3891.

92. Hunter, M.; Ludwig, M. The reaction of imidoesters with proteins and related small molecules. *J. Am. Chem. Soc.* **1962**, *84*, 3491-3504.

93. Browne, D. T.; Kent, S. B. Formation of non-amidine products in the reaction of primary amines with imido esters. Biochem. *Biophys. Res. Commun.* **1975**, *67*, 126-132.

94. Kunkel, G. R.; Mehrabian, M.; Martinson, H. G. Contact-site cross-linking agents. *Mol. Cell. Biochem.* **1981**, *34*, 3-13.

95. Grabarek, Z.; Gergely, J. Zero-length crosslinking procedure with the use of active esters. *Anal. Biochem.* **1990**, *185*, 131-135.

96. Novak, P.; Kruppa, G. H. Intra-molecular cross-linking of acidic residues for protein structure studies. *Euro. J. Mass Spectrom.* **2008**, *14*, 355-365.

97. Zhang, X.; Wang, J.-H.; Tan, D.; Li, Q.; Li, M.; Gong, Z.; Tang, C.; Liu, Z.; Dong, M.-Q.; Lei, X. Carboxylate-Selective Chemical Cross-Linkers for Mass Spectrometric Analysis of Protein Structures. *Anal. Chem.* **2018**, *90*, 1195-1201.

98. Smyth, D. G.; Blumenfeld, O.; Konigsberg, W. Reactions of N-ethylmaleimide with peptides and amino acids. *Biochem. J.* **1964**, *91*, 589.

99. Gorin, G.; Martic, P. A.; Doughty, G. Kinetics of the reaction of N-ethylmaleimide with cysteine and some congeners. *Arch. Biochem. Biophys.* **1966**, *115*, 593-597.

100. Brewer, C. F.; Riehm, J. P. Evidence for possible nonspecific reactions between N-ethylmaleimide and proteins. *Anal. Biochem.* **1967**, *18*, 248-255.

101. Gilchrist, T. L.; Rees, C. W. Carbenes, nitrenes and arynes. Springer US: 1969.

102. Brunner, J. New photolabeling and crosslinking methods. *Ann. Rev. Biochem.* **1993**, *62*, 483-514.

103. Suchanek, M.; Radzikowska, A.; Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2005**, *2*, 261.

104. Lewis, R. V.; Roberts, M. F.; Dennis, E. A.; Allison, W. S. Photoactivated heterobifunctional cross-linking reagents which demonstrate the aggregation state of phospholipase A2. *Biochemistry* **1977**, *16*, 5650-5654.

105. Murphy, Michael P. How mitochondria produce reactive oxygen species. *Biochem. J.* **2009**, *417*, 1-13.

106. Apel, K.; Hirt, H. REACTIVE OXYGEN SPECIES: Metabolism, Oxidative Stress, and Signal Transduction. *Annu. Rev. Plant Biol.* **2004**, *55*, 373-399.

107. D'Autréaux, B.; Toledano, M. B. ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 813.

108. Ermácora, M. R.; Delfino, J. M.; Cuenoud, B.; Schepartz, A.; Fox, R. O. Conformation-dependent cleavage of staphylococcal nuclease with a disulfide-linked iron chelate. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 6383-6387.

109. Buxton, G. V.; Greenstock, C. L.; Helman, W. P.; Ross, A. B. Critical Review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals (·OH/·O− in Aqueous Solution. *J. Phys. Chem. Ref. Data* **1988**, *17*, 513-886.

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

111. Garrison, W. M. Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chem. Rev.* **1987**, *87*, 381-398.

112. Fenton, H. J. H. LXXIII.—Oxidation of tartaric acid in presence of iron. *J. Chem. Soc. Trans.* **1894**, *65*, 899-910.

113. Masarwa, A.; Rachmilovich-Calis, S.; Meyerstein, N.; Meyerstein, D. Oxidation of organic substrates in aerated aqueous solutions by the Fenton reagent. *Coord. Chem. Rev.* **2005**, *249*, 1937-1943.

114. Kryatov, S. V.; Rybak-Akimova, E. V.; Schindler, S. Kinetics and Mechanisms of Formation and Reactivity of Non-heme Iron Oxygen Intermediates. *Chem. Rev.* **2005**, *105*, 2175-2226.

115. Goldstein, S.; Meyerstein, D.; Czapski, G. The Fenton reagents. *Free Rad. Biol. Med.* **1993**, *15*, 435-445.

116. Sawyer, D. T.; Sobkowiak, A.; Matsushita, T. Metal [MLx; M = Fe, Cu, Co, Mn]/Hydroperoxide-Induced Activation of Dioxygen for the Oxygenation of Hydrocarbons: Oxygenated Fenton Chemistry. *Acc. Chem. Res.* **1996**, *29*, 409-416.

117. Haber, F.; Weiss, J. Über die Katalyse des Hydroperoxydes. *Naturwissenschaften* **1932**, *20*, 948-950.

118. Haber, F.; Weiss, J.; Pope, W. J. The catalytic decomposition of hydrogen peroxide by iron salts. *Proc. R. Soc. Lond.* **1934**, *147*, 332-351.

119. Barb, W. G.; Baxendale, J. H.; George, P.; Hargrave, K. R. Reactions of ferrous and ferric ions with hydrogen peroxide. Part I.—The ferrous ion reaction. *Trans. Faraday Soc.* **1951**, *47*, 462-500.

120. Barb, W. G.; Baxendale, J. H.; George, P.; Hargrave, K. R. Reactions of ferrous and ferric ions with hydrogen peroxide. Part II.—The ferric ion reaction. *Trans. Faraday Soc.* **1951**, *47*, 591-616.

121. Tullius, T.; Dombroski, B. Iron(II) EDTA used to measure the helical twist along any DNA molecule. *Science* **1985**, *230*, 679-681.

122. Tullius, T. D.; Dombroski, B. A. Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 5469-5473.

123. Pogozelski, W. K.; McNeese, T. J.; Tullius, T. D. What Species Is Responsible for Strand Scission in the Reaction of [FeIIEDTA]2- and H2O2 with DNA? *J. Am. Chem. Soc.* **1995**, *117*, 6428-6433.

124. Liochev, S. I. The mechanism of "Fenton-like" reactions and their importance for biological systems. A biologist's view. In Metal ions in biological systems, Routledge: 2018; pp 1-39.

125. Sawyer, D. T. Metal [Fe (II), Cu (I), Co (II), Mn (III)]/hydroperoxide-induced activation of dioxygen (· O2·) for the ketonization of hydrocarbons: oxygenated Fenton chemistry. *Coord. Chem. Rev.* **1997**, *165*, 297-313.

126. Strlic, M.; Kolar, J.; Selih, V.-S.; Kocar, D.; Pihlar, B. A comparative study of several transition metals in Fenton-like reaction systems at circum-neutral pH. *Acta Chim. Slov.* **2003**, *50*, 619-632.

127. Anipsitakis, G. P.; Dionysiou, D. D. Radical generation by the interaction of transition metals with common oxidants. *Environ. Sci. Tech.* **2004**, *38*, 3705-3712.

128. Simpson, J. A.; Cheeseman, K. H.; Smith, S. E.; Dean, R. T. Free-radical generation by copper ions and hydrogen peroxide. Stimulation by Hepes buffer. *Biochem. J.* **1988**, *254*, 519-523.

129. Urbański, N. K.; Beresewicz, A. Generation of OH initiated by interaction of Fe^{2+} and Cu^{+} with dioxygen; comparison with the Fenton chemistry. *Acta Biochim. Pol.* **2000**, *47*, 951-962.

130. Rana, T. M.; Meares, C. F. Specific cleavage of a protein by an attached iron chelate. *J. Am. Chem. Soc.* **1990**, *112*, 2457-2458.

131. Rana, T. M.; Meares, C. F. Iron chelate mediated proteolysis: protein structure dependence. *J. Am. Chem. Soc.* **1991**, *113*, 1859-1861.

132. Rana, T. M.; Meares, C. F. Transfer of oxygen from an artificial protease to peptide carbon during proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 10578-10582.

133. Shcherbakova, I.; Mitra, S.; Beer, R. H.; Brenowitz, M. Fast Fenton footprinting: a laboratory-based method for the time-resolved analysis of DNA, RNA and proteins. *Nucleic Acids Res.* **2006**, *34*, e48-e48.

134. Simic, M. G. [2] Pulse radiolysis in study of oxygen radicals. In Methods in Enzymology, Academic Press: 1990; Vol. 186, pp 89-100.

135. Sclavi, B.; Woodson, S.; Sullivan, M.; Chance, M. R.; Brenowitz, M. Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J Mol. Biol.* **1997**, *266*, 144-159.

136. Chance, M. R.; Sclavi, B.; Woodson, S. A.; Brenowitz, M. Examining the conformational dynamics of macromolecules with time-resolved synchrotron X-ray 'footprinting'. *Structure* **1997**, *5*, 865-869.

137. Sclavi, B.; Woodson, S.; Sullivan, M.; Chance, M.; Brenowitz, M. [19] Following the folding of RNA with time-resolved synchrotron X-ray footprinting. In Methods in Enzymology, Academic Press: 1998; Vol. 295, pp 379-402.

138. Gupta, S.; Sullivan, M.; Toomey, J.; Kiselar, J.; Chance, M. R. The Beamline X28C of the Center for Synchrotron Biosciences: a National Resource for Biomolecular Structure and Dynamics Experiments Using Synchrotron Footprinting. *J. Synchrotron Radiat.* **2007**, *14*, 233-243.

139. Weeks, J. L.; Matheson, M. S. The Primary Quantum Yield of Hydrogen Peroxide Decomposition1. *J. Am. Chem. Soc.* **1956**, *78*, 1273-1278.

140. Dainton, F. S. The Primary Quantum Yield in the Photolysis of Hydrogen Peroxide at 3130 Å. and the Primary Radical Yield in the X- and γ-Radiolysis of Water. *J. Am. Chem. Soc.* **1956**, *78*, 1278-1279.

141. 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*, 2057-63.

142. Sharp, J. S.; Becker, J. M.; Hettich, R. L. Analysis of Protein Solvent Accessible Surfaces by Photochemical Oxidation and Mass Spectrometry. *Anal. Chem.* **2004**, *76*, 672-683.

143. Aye, T. T.; Low, T. Y.; Sze, S. K. Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry. *Anal. Chem.* **2005**, *77*, 5814-5822.

144. 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*, 1306-1318.

145. 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. D.; Tyrniak, 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*, 2250-2258.

146. Yan, Y. T.; Chen, G. D.; Wei, H.; Huang, R. Y. C.; Mo, J. 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*, 2084-2092.

147. 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*, 12090-8.

148. Liu, X. R.; Rempel, D. L.; Gross, M. L. Composite Conformational Changes of Signaling Proteins upon Ligand Binding Revealed by a Single Approach: Calcium-Calmodulin Study. *Anal. Chem.* **2019**, *91*, 12560-12567.

149. Liu, X. R.; Zhang, M. M.; Rempel, D. L.; Gross, M. L. A single approach reveals the composite conformational changes, order of binding, and affinities for calcium binding to calmodulin. *Anal. Chem.* **2019**, *91*, 5508-5512.

150. Liu, X. R.; Zhang, M. M.; Rempel, D. L.; Gross, M. L. Protein-ligand interaction by ligand titration, fast photochemical oxidation of proteins and mass spectrometry: LITPOMS. *J. Am. Soc. Mass Spectrom.* **2018**, *30*, 213-217.

151. 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. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E2596-605.

152. Chen, J. W.; 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*, 18724-18731.

153. Paulsen, S. 3.3 ‐ Dialkyl ‐ diazacyclopropen ‐ (1). *Angew. Chem. Int. Ed.* **1960**, *72*, 781-782.

154. Smith, R. A.; Knowles, J. R. Aryldiazirines. Potential reagents for photolabeling of biological receptor sites. *J. Am. Chem. Soc.* **1973**, *95*, 5072-5073.

155. Wofsy, L.; Metzger, H.; Singer, S. Affinity labeling—a general method for labeling the active sites of antibody and enzyme molecules. *Biochemistry* **1962**, *1*, 1031-1039.

156. Ruoho, A. E.; Kiefer, H.; Roeder, P. E.; Singer, S. J. The Mechanism of Photoaffinity Labeling. *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70*, 2567-2571.

157. Halloran, M. W.; Lumb, J.-P. Recent Applications of Diazirines in Chemical Proteomics. *Chem. Euro. J.* **2019**, *25*, 4885-4898.

158. Nicholas, R. A. Purification of the membrane-spanning tryptic peptides of the. alpha. polypeptide from (sodium and potassium ion)-activated adenosine triphosphatase labeled with 1-tritiospiro [adamantane-4, 1'-diazirine]. *Biochemistry* **1984**, *23*, 888-898.

159. Farley, R. A. Identification of hydrophobic regions of the calcium-transport ATPase from sarcoplasmic reticulum after photochemical labeling with adamantane diazirine. *Int. J. Bio.* **1983**, *15*, 1423-1427.

160. Xi, J.; Liu, R.; Rossi, M. J.; Yang, J.; Loll, P. J.; Dailey, W. P.; Eckenhoff, R. G. Photoactive analogues of the haloether anesthetics provide high-resolution features from low-affinity interactions. *ACS Chem. Biol.* **2006**, *1*, 377-384.

161. Gomes, A. F.; Gozzo, F. C. Chemical cross ‐ linking with a diazirine photoactivatable cross ‐ linker investigated by MALDI ‐ and ESI ‐ MS/MS. *J. Mass Spectrom.* **2010**, *45*, 892-899.

162. Richards, F. M.; Lamed, R.; Wynn, R.; Patel, D.; Olack, G. Methylene as a possible universal footprinting reagent that will include hydrophobic surface areas: Overview and feasibility: Properties of diazirine as a precursor. *Protein Sci.* **2000**, *9*, 2506-2517.

163. Jumper, C. C.; Schriemer, D. C. Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Anal. Chem.* **2011**, *83*, 2913-2920.

164. Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. High-resolution mapping of carbene-based protein footprints. *Anal. Chem.* **2012**, *84*, 4411-4418.

165. Ziemianowicz, D. S.; Bomgarden, R.; Etienne, C.; Schriemer, D. C. Amino acid insertion frequencies arising from photoproducts generated using aliphatic diazirines. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2011-2021.

166. Blencowe, A.; Hayes, W. Development and application of diazirines in biological and synthetic macromolecular systems. *Soft Matter* **2005**, *1*, 178-205.

167. Hu, Y.; Guo, T.; Ye, X.; Li, Q.; Guo, M.; Liu, H.; Wu, Z. Dye adsorption by resins: Effect of ionic strength on hydrophobic and electrostatic interactions. *Chem. Eng. J.* **2013**, *228*, 392-397.

168. Christen, R. P.; Nomikos, S. I.; Smith, E. T. Probing protein electrostatic interactions through temperature/reduction potential profiles. *J. Biol. Inorg. Chem.* **1996**, *1*, 515-522.

169. 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*, 552-555.

170. Manzi, L.; Barrow, A. S.; Scott, D.; Layfield, R.; Wright, T. G.; Moses, J. E.; Oldham, N. J. Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions. *Nat. Commun.* **2016**, *7*, 13288.

171. Irikura, K. K.; Goddard III, W.; Beauchamp, J. Singlet-triplet gaps in substituted carbenes CXY (X, Y= H, fluoro, chloro, bromo, iodo, silyl). *J. Am. Chem. Soc.* **1992**, *114*, 48-51.

172. Horspool, W. M.; Lenci, F. CRC Handbook of Organic Photochemistry and Photobiology, Volumes 1 & 2. CRC press: 2003.

173. Das, J. Aliphatic diazirines as photoaffinity probes for proteins: recent developments. *Chem. Rev.* **2011**, *111*, 4405-4417.

174. Bourissou, D.; Guerret, O.; Gabbai, F. P.; Bertrand, G. Stable carbenes. *Chem. Rev.* **2000**, *100*, 39-92.

175. Richards Jr, C. A.; Kim, S.-J.; Yamaguchi, Y.; Schaefer III, H. F. Dimethylcarbene: a singlet ground state? *J. Am. Chem. Soc.* **1995**, *117*, 10104-10107.

176. Bertrand, G. Carbene chemistry: from fleeting intermediates to powerful reagents. CRC Press: 2002.

177. Davies, M.; Dean, R.; Davies, D. Radical-Mediated Protein Oxidation: From Chemistry to Medicine. Oxford University Press: New York, NY, 1997.

178. Xiao, R.; Liu, K.; Bai, L.; Minakata, D.; Seo, Y.; Göktaş, R. K.; Dionysiou, D. D.; Tang, C.-J.; Wei, Z.; Spinney, R. Inactivation of pathogenic microorganisms by sulfate radical: Present and future. *Chem. Eng. J.* **2019**, *371*, 222-232.

179. Bridgewater, J. D.; Vachet, R. W. Using microwave-assisted metal-catalyzed oxidation reactions and mass spectrometry to increase the rate at which the copper-binding sites of a protein are determined. *Anal. Chem.* **2005**, *77*, 4649-4653.

180. Bridgewater, J. D.; Lim, J.; Vachet, R. W. Using metal-catalyzed oxidation reactions and mass spectrometry to identify amino acid residues within 10 Å of the metal in Cu-binding proteins. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1552-1559.

181. Bridgewater, J. D.; Vachet, R. W. Metal-catalyzed oxidation reactions and mass spectrometry: The roles of ascorbate and different oxidizing agents in determining Cu–protein-binding sites. *Anal. Biochem.* **2005**, *341*, 122-130.

182. 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*, 7821-7827.

183. Chan, K. J.; O'Hagan, D. The rare fluorinated natural products and biotechnological prospects for fluorine enzymology. In Methods in enzymology, Elsevier: 2012; Vol. 516, pp 219-235.

184. O'Hagan, D. Understanding organofluorine chemistry. An introduction to the C–F bond. *Chem. Soc. Rev.* **2008**, *37*, 308-319.

185. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A* **1976**, *32*, 751-767.

186. 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*, 14007-14010.

187. Winkler, R. Iodine—a potential antioxidant and the role of Iodine/Iodide in health and disease. *Nat. Sci.* **2015**, *7*, 548.

188. Cavalieri, R. R. Iodine metabolism and thyroid physiology: current concepts. *Thyroid* **1997**, *7*, 177-181.

189. Chen, J.; Rempel, D. L.; Gross, M. L. Temperature Jump and Fast Photochemical Oxidation Probe Submillisecond Protein Folding. *J. Am. Chem. Soc.* **2010**, *132*, 15502-15504.

190. 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*, 12090-12098.

Chapter 2: A Fast Photochemical Oxidation of Proteins (FPOP) Platform for Free-Radical Reactions: the Carbonate Radical Anion with Peptides and Proteins*

* This chapter is based on the following publication: Zhang, M. M.; Rempel, D. L.; Gross, M. L. A Fast Photochemical Oxidation of Proteins (FPOP) Platform for Generation and Determination of Amino-Acid Reactivity in Peptides and Proteins *Free Radic. Biol. Med.* **2019,** *131,* 126-132.

2.1 Abstract

Fast Photochemical Oxidation of Protein (FPOP), based on a pulsed KrF laser (248 nm) for free-radical generation, is a protein footprinting method that utilizes hydroxyl radicals to footprint proteins in solution. FPOP has been recognized as a biophysical technique in structural proteomics interrogation, including epitope mapping, protein-aggregation characterization, protein-folding monitoring, and binding-affinity determination. The distinct merits of the platform are: i) the use of a scavenger to control radical lifetime and allow fast ("snapshot") footprinting of solvent-accessible residues in protein; ii) the employment of a flow system to enable single-shot irradiation of small plugs of the targeted sample; iii) the incorporation of methionine and catalase after radical oxidation chemistry to prevent post-oxidation with residual oxidizing species; and v) the utilization of mature mass spectrometry-based proteomic methods to afford analysis. In addition to $HO^{\bullet}$, other reactive reagents (e.g., carbenes, iodide, sulfate radical anion, and trifluoromethyl radical) can be implemented on this platform to increase the versatility and scope. In this study, we further elaborate the use of FPOP platform to generate secondary radicals and establish a workflow to answer fundamental questions regarding the intrinsic selectivity and reactivity of radicals that are important in biology. Carbonate radical anion is the example we chose owing to its oxidative character and important putative pathogenic roles in inflammation. This systematic study with model proteins/peptides gives consistent results with a previous study that evaluated reactivity with free amino acids and shows that methionine and tryptophan are the most reactive residues with $CO_3^{-\bullet}$. Other aromatic amino acids (i.e., tyrosine, histidine, and phenylalanine) exhibit moderate reactivity, whereas aliphatic amino acids are inert, unlike with $HO^{\bullet}$. The outcome demonstrates this approach to be appropriate for studying the fast reactions of radicals with proteins.

2.2 Introduction

Fast Photochemical Oxidation of Protein (FPOP), an alternative to radiolysis of water with synchrotron radiation to give $HO^{\bullet}$ [1-2], was initially invented by Hambly and Gross[3]. Hydroxyl radicals, formed by a pulsed KrF laser irradiation of hydrogen peroxide, oxidize protein side chain in situ and footprint the exposed residues depending on their intrinsic reactivity and solvent accessibility. With employment of a scavenging reagent, radical lifetime can be tuned in the microsecond time frame to enable protein structural characterization faster than most conformational changes. In addition, the irreversible modification by $HO^{\bullet}$ allows comprehensive downstream proteomics sample handling, digestion, and mass spectrometry analysis. The FPOP approach, as a valuable biophysical tool in protein structure elucidation, has been recognized for: 1) mapping binding interface in protein-ligand interaction, including small molecules binding events[4] and antigen-antibody interaction[5-6]; 2) revealing multistage protein aggregation[7]; 3) determining binding affinity; 4) identifying hidden conformational changes[8], 5) characterizing overall protein dynamics[9-10]. In addition to $HO^{\bullet}$, other reactive reagents (e.g., sulfate radical anion[11], carbene diradical[12], iodide, and trifluoromethyl radical[13], have been implemented on the FPOP platform to afford broader labeling opportunities, giving more comprehensive coverage of amino acid residues. Now, we wish to elaborate further the utility of the FPOP platform and to demonstrate its capacity to answer more fundamental questions in biological free-radical chemistry. To do this, we address the adaption of the FPOP platform to ascertain the selectivity and reactivity of the carbonate radical anion with model proteins.

In biological fluids, the bicarbonate anion exists at of 24 mM in serum and 14 mM in intracellular media equilibrated with approximately 1.3 mM carbon dioxide[14]. The radical anion can be formed from the abundant bicarbonate-carbon dioxide pair by superoxide dismutase

(SOD)[15], xanthine oxidase (XO)[16], and other enzymes that can play pathogenic roles in many physiological conditions. As a key mediator, carbonate radical anion may cause oxidative damage that eventually leads to vascular disease and neurodegeneration[14]. Its best-known metabolic function is to modulate peroxynitrite activity, a strong oxidant generated from nitric oxide and superoxide anion. The $CO_3^{-\bullet}$/ NO_2 couple, dissociated from nitrosuperoxycarbonate ($ONOOCO_2^-$), promotes protein nitration[17]. In this process, $CO_3^{-\bullet}$ is believed to abstract H from amino acids, typically tyrosine, followed by NO_2 addition to give a nitro-substituted residue.

Carbonate radical anion, a potent one electron oxidant (E^0 = 1.59 V, pH = 7.0), can favorably oxidize electron-rich donors by electron transfer[18]. $CO_3^{-\bullet}$ is also a site-selective oxidant of guanine over other DNA bases as a damaging reagent[19]. The biological consequences of reactions of the carbonate radical anion on proteins are also of concern. Many investigations show, for example, that $CO_3^{-\bullet}$ can oxidize tryptophan-rich lysozyme to afford a ditryptophan form[20], $CO_3^{-\bullet}$ can affect redox signaling pathways by oxidizing tyrosine-containing protein[21] and $CO_3^{-\bullet}$ can also cause activity loss of horseradish peroxidase[22]. Most of the research on $CO_3^{-\bullet}$ relies on a previous radiolysis study[23] carried out in 1973 with free amino acids; this work shows tryptophan and tyrosine to be the preferred amino acids. The reactivity of amino acid residues, however, in peptides and protein is likely different, depending on context and solvent accessibility.

To provide more fundamental information and provide a method for determining the fast reactions of radicals with proteins, we describe a systematic study of the selectivity and specificity of $CO_3^{-\bullet}$ with proteins/peptides as targets. To accomplish this, we extended the FPOP platform for generating secondary radicals. Incorporating hydrogen peroxide into sample aliquots, we generated by pulsed laser irradiation hydroxyl radicals, which then react with other

reagents (here CO_3^{2-}/ HCO_3^-) to generate secondary radicals. The flow system we used overcomes the steric effect of static laser irradiation, ensuring the homogeneity of each labeling event. Specifically, for proteins and peptides dissolved in sodium bicarbonate/bicarbonate buffer, we can successfully generate the carbonate radical anion and monitor the oxidation of selected residues on proteins and peptides, taking advantage of mass-spectrometry-based analysis. For the first time, we successfully demonstrate the reactivity and selectivity of carbonate radical anion towards side chains of amino acid residues in proteins to significantly advance the previous studies of reactions with free amino acids. Moreover, the method and platform reported here can be generalized for other bio-relevant studies of free radicals, providing new possibilities for radical generation and subsequent fundamental investigations of protein/radical reactivity.

2.3 Principles

2.3.1 FPOP Platform

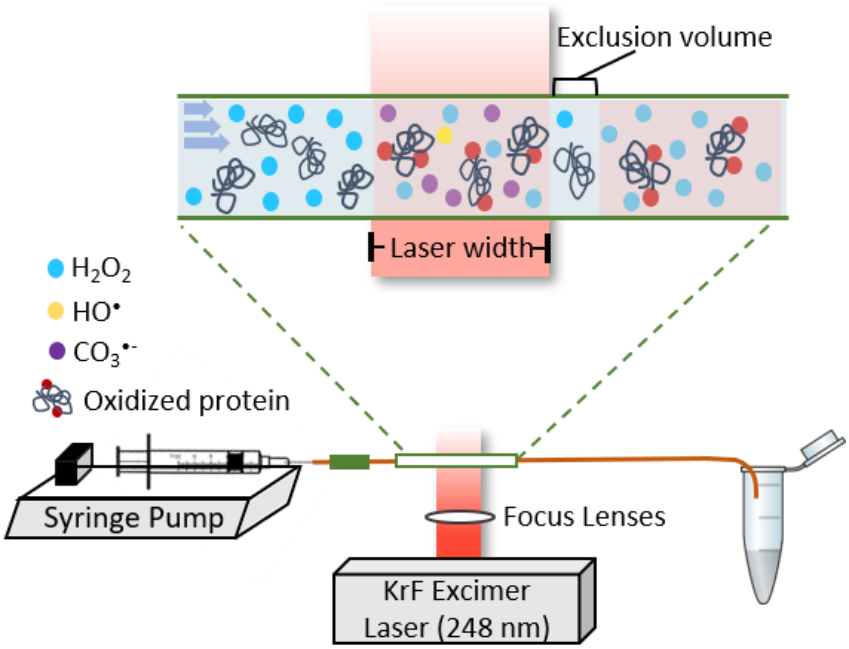


Figure 2.1 FPOP platform. A sample syringe, driven by a syringe pump, is connected to silica tubing, where part of the polyimide coating is removed to afford a transparent window. The other

end of the capillary is inserted into a collection tube containing reagents to deactivate left-over oxidizing agents. A 248 nm KrF excimer laser is placed such that its beam is perpendicular to the flow system and has lenses to focus on the flow tube. Sample proteins will react with the radicals within the laser window. The flow rate is determined such that between every two solution plugs exists an unexposed portion, termed an "exclusion volume", that minimizes "double hits" on the protein solution.

An FPOP platform (Figure 2.1) implements a 248 nm KrF excimer to photolyze cleavable reagents (e.g., hydrogen peroxide). A silica capillary connects both a syringe pump and a collection tube that terminates the flow. The polyimide coating is removed along the tube to give a transparent window for the laser beam. The laser, typically generated at 7.4 Hz by an external pulse generator, is restricted by an iris and focused by two convex lenses to pass through the transparent window. Within the laser window, hydrogen peroxide is photolyzed into hydroxide radicals. These radicals under usual concentration conditions oxidize the protein sample, but under the conditions of this study react with carbonate/bicarbonate anion to generate secondary carbonate radical anions. This approach should be adaptable to formation of other radicals that are important in biology.

To avoid over oxidation on the same solution plug, the flow rate on the syringe pump is carefully tuned. A small volume of solution portion is intentionally excluded to create a barrier between two exposed plugs, termed an exclusion volume. The flow-rate calculation is by equation 2.1, where the exclusion volume is typically ~ 20%. After the oxidation chemistry, the protein sample is collected in the tube containing methionine and catalase to prevent post-oxidation by excess hydrogen peroxide and any remaining oxidative species.

$$Flow\ rate = f_{laser} L_{laser\ width}\ V_{exclusion}\, \pi\left(\frac{i.d.}{2}\right)^2 \quad (2.1)$$

2.3.2 Numerical Simulation of $CO_3^{-\bullet}$ Oxidation Chemistry

Generation of the carbonate radical anion in the laboratory can be by pulse radiolysis of N_2O-saturated aqueous solution containing sodium bicarbonate at alkaline pH[22, 24] or UV photolysis of suitable metal complexes (e.g., $[Co(NH_3)_4CO_3]ClO_4$ [20]). It can also be formed from human superoxide dismutase (hSOD) [25] in phosphate buffer with EPR spin-trapping detection or oxidation of bicarbonate with a sulfate radical anion generated by photolysis of persulfate anion and detected with UV spectroscopy[19, 26]. In the current study, we considered generating the carbonate radical anion as the secondary radical through a cascade process starting with sulfate radical anion, formed by photolysis of sodium persulfate. The persulfate anion, however, reacts moderately with amino acid residues (e.g., tryptophan and methionine) even in the absence of laser irradiation. In addition, the rate constant for carbonate radical anion production from the sulfate radical anion is at least 1000 times smaller than that of protein oxidation by the sulfate radical anion. Furthermore, the carbonate radical anion is a considerably weaker oxidizing agent (E_0 = 1.59 V) than the sulfate radical anion (E_0 = 2.43 V), rendering it difficult to make $CO_3^{-\bullet}$ the predominate oxidative species in the reaction regime. We anticipated these difficulties and chose instead to employ a weaker oxidant, hydroxyl radical (E_0 = 2.30 V), from hydrogen peroxide photodissociation. To select appropriate conditions, we conducted numerical simulation of the cascade chemistry (for rate constants, see Table 2.1).

Table 2.1 Reaction scheme and rate constants for generation of the carbonate radical anion by reactions of hydroxyl radicals from H_2O_2 with carbonate/bicarbonate in a buffer.

	Reaction	**Rate Constant ($M^{-1}s^{-1}$)**
1	$H_2O_2 + hv \rightarrow 2\ HO^{\bullet}$	$\Phi_{248} = 0.4\text{-}0.5$
2	$HO^{\bullet} + CO_3^{2-} \rightarrow OH^- + CO_3^{\bullet -}$	3.0×10^8 [27]
3	$CO_3^{2-} + H_2O \rightarrow HCO_3^- + OH^-$	3.06×10^5 [28]
4	$HO^{\bullet} + HCO_3^- \rightarrow H_2O + CO_3^{\bullet -}$	8.5×10^6 [27]
5	$HO^{\bullet} + HO^{\bullet} \rightarrow H_2O_2$	4.7×10^9 [3]
6	$CO_3^{-\bullet} + CO_3^{-\bullet} \rightarrow CO_3^{2-} + CO_2$	5×10^7 [18]

In a carbonate/bicarbonate aqueous buffer, hydroxyl radicals from photolysis of hydrogen peroxide upon pulsed laser irradiation (Table 2.1, Reaction 1) react with the carbonate (Table 2.1, Reaction 2) or bicarbonate anions (Table 2.1, Reaction 4) to form the carbonate radical anion. Both reactive radicals will decay through recombination (Table 2.1, Reactions 5,6). Using these rate constants, we carried out a numerical simulation using Mathcad 14.0 to predict the outcome from second-order kinetics, as shown in Figure 2.2. We adopted the "Rkadapt" function in MathCad to solve the system of ordinary differential equations that arise from the network of second order reactions. The time interval from 0 to 10 ms was divided into 10 decades, each divided uniformly into 100 steps. The "RKadapt" solves to a "TOL" of 1×10^{-12}.

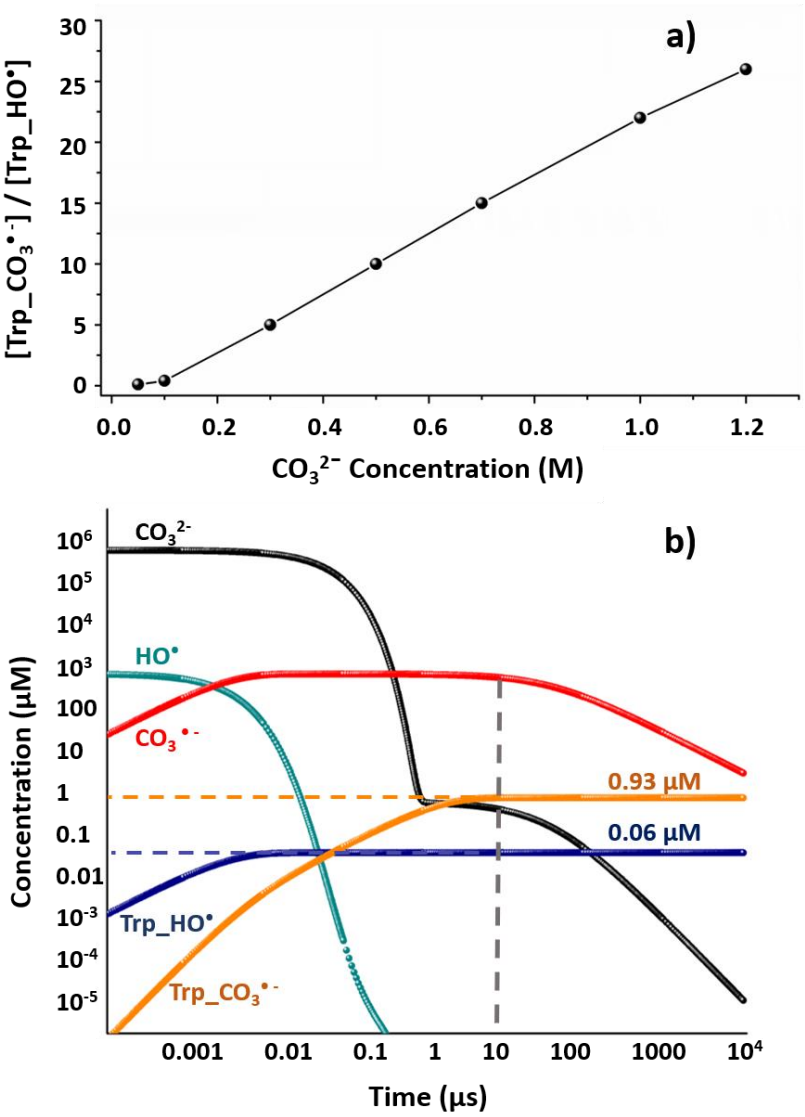


Figure 2.2 Numerical simulation of $CO_3^{-\bullet}$. Numerical simulation of (a) CO_3^{2-} concentration as a function of the ratio of [$CO_3^{-\bullet}$] and [$HO^\bullet$] oxidized tryptophan concentration, where the amino acid tryptophan is taken as a model for reactions of amino acids on a target protein; (b) $CO_3^{-\bullet}$ and other related species as a function of time with 700 mM CO_3^{2-}. Different curves, as denoted by different colors, are plotted on a log scale to represent time-dependent concentrations of each components. The products produced by oxidation of $CO_3^{-\bullet}$ and $HO^\bullet$ are colored in orange and dark blue, respectively. The dashed line represents the time at which the chemistry is optimized (i.e., largely complete in 10 µs).

In the simulation, the initial concentration of hydrogen peroxide is 15 mM and the starting concentration of $HO^\bullet$ after laser irradiation is approximately 1 mM, which is the typical yield as

measured previously[29]. In addition, we used the reaction with tryptophan as a model for the reaction of $CO_3^{-\bullet}$ with a target protein. Under the cascade chemistry regime (Figure 2.2a), the ratios of $CO_3^{-\bullet}$ and $HO^{\bullet}$ oxidized tryptophan concentrations show corresponding increases when [CO_3^{2-}] increases at pH = 10. With 1.2 M CO_3^{2-} as buffer, $CO_3^{-\bullet}$ gives almost 26-fold more oxidized product than that by $HO^{\bullet}$, showing the best result. Taking the practical solubility of CO_3^{2-} and HCO_3^{-} into consideration, however, we chose 700 mM carbonate anion as the condition for the detailed simulation shown in Figure 2.2b. As denoted by the green curve, the hydroxyl radical reacts with carbonate and bicarbonate anion, yielding nearly complete conversion to $CO_3^{-\bullet}$ in less than 1 μs. The concentration of carbonate radical anion reaches a maximum at ~ 0.01 μs and then remains roughly constant until near 100 μs. Ultimately, $CO_3^{-\bullet}$ self quenches, showing a decrease as shown by the red curve. The vertical grey dash line represents approximately the timescale for oxidation chemistry (~ 10 μs). At and beyond this time, the $CO_3^{-\bullet}$ is already the predominant oxidative species as can be seen by comparing with its underlying area with that of $HO^{\bullet}$, which has largely disappeared. The concentrations for $CO_3^{-\bullet}$ and $HO^{\bullet}$ oxidized tryptophan after 10 μs is 0.93 μM and 0.06 μM, respectively. Although there are remaining hydroxyl radicals, the $CO_3^{-\bullet}$ oxidation contributes almost 15-fold more than $HO^{\bullet}$, allowing us to determine the reactivity and selectivity of $CO_3^{-\bullet}$ towards different amino acids contained in a protein.

2.4 Materials

- Bombesin acetate salt hydrate, α-melanocyte stimulating hormone, angiotensin I bovine, leucine enkephalin, bradykinin, glycine, catalase, urea, *L*-histidine, *L*-methionine, 30% hydrogen peroxide, triethylammonium bicarbonate buffer (TEAB), HPLC-grade solvents,

apo-myoglobin from equine skeletal muscle, ubiquitin from bovine erythrocytes (Millipore Sigma Co., St. Louis, MO, USA)

- Sequencing grade trypsin (Promega Co., Madison, WI, USA)
- Silica capillary (150 µm i.d., Polymicro Technologies, Pheonix, AZ, USA)
- C18 reversed-phase desalting column (nanoViper, 100 µm x 2 cm, 5 µm,100 Å; Thermo Fisher Scientific, Waltham, MA, USA)
- Analytical column with C18 reversed-phase material (Magic, 100 µm, 180 mm, 5 µm, 120 Å; Michrom Bioresources, Inc., Auburn, CA)
- Solvent A (water with 0.1% formic acid by volume) and solvent B (80% acetonitrile with 0.1% formic acid by volume).
- C18 NuTip (Glygen Co.)

2.5 Instrumentation

- 248nm KrF excimer laser (GAM Laser Inc., Orlando, FL, USA)
- Syringe pump (Harvard Apparatus, Holliston, MA, USA)
- External pulse generator (B&K Precision, Yorbal Linda, CA, USA)
- Ultimate 3000 Rapid Separation system (Dionex, Thermo Fisher Scientific, Waltham, MA, USA)
- Bruker MaXis 4G quadrupole time-of-flight mass spectrometer (Bruker, Co., Billerica, MA)
- Thermo Q Exactive Plus orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA)

2.6 Protocol

2.6.1 Preparation of Buffer Solution and Stock Solution

(1) Dissolve 0.7819 g sodium carbonate and 0.5361 g sodium bicarbonate in 10 mL distilled water to constitute 700 mM carbonate/bicarbonate buffer to achieve a pH value of 10.0 ± 0.2.

(2) Dissolve 0.376 g glycine, 0.12 g sodium hydroxide in 10 mL distilled water, then add 0.584 g sodium chloride and 0.745 g potassium chloride to constitute 700 mM glycine-NaOH buffer with the same ionic strength at pH = 10 ± 0.2.

(3) Prepare histidine solution by dissolving 0.0233 g histidine in 1 mL glycine-NaOH buffer, then dilute into 50 mM histidine stock solution.

(4) Make 70 mM methionine stock solution in both 1 mL carbonate/bicarbonate buffer and glycine-NaOH buffer with 0.0105 g methionine as scavenger.

(5) Prepare 5 µM catalase solution in 1 mL distilled water by dissolving 0.0012 g catalase and diluting to 500 nM.

(6) Dissolve 0.0050 g ubiquitin and 0.0085 g apo-myoglobin in two different buffers to afford 500 µM protein solutions. Dilute further to 50 µM stock solution. In addition, prepare 0.0010 g bombesin acetate salt hydrate, 0.0010 g α-melanocyte stimulating hormone, 0.0010 g angiotensin I bovine, 0.0010 g leucine enkephalin and 0.0010 g bradykinin solution in two different buffers with concentrations in the range 600 – 780 µM depending on its molecular weight. Dilute all solutions to ~ 500 µM and mix those in the same buffer to give 50 µM a ‘Peptide Cocktail’ stock solution.

(7) Prepare 300 mM hydrogen peroxide solution immediately before the FPOP laser irradiation by mixing 30 µL 30% (v/v) hydrogen peroxide solution with 970 µL of two different buffers

individually. Keep the hydrogen peroxide solution on ice throughout the experimental procedure to avoid decomposition.

2.6.2 Protein Oxidation on FPOP

Oxidation achieved by carbonate radical anion

(1) Warm the 248nm KrF laser and measure the laser energy with a sensor meter. Adjust the laser energy to 24 mJ/pulse at a frequency of 7.4 Hz. The laser width is measured by placing colored tape right behind the transparent silica tubing (width = 2.73 mm) and observing a burn mark. Choose a 20% exclusion volume, giving a flow rate of 26.8 µL/min, which can be calculated based on the equation 1.

(2) Inject 10 µL of 70 mM methionine stock solution in carbonate/bicarbonate buffer and 1 µL 500 nM catalase solution to a low-bind Eppendorf tube to be used as the collection tube. Place the collection tube at the end of the silica tubing.

(3) Mix 5 µL of 50 µM protein solution or 'Peptide cocktail' with 42 □L carbonate/bicarbonate buffer; then add 3 µL of 300 mM hydrogen peroxide solution. Transfer the sample solution to a syringe and insert it in the syringe pump (Figure 1). Start the pump and trigger the laser irradiation. The time window between the addition of hydrogen peroxide and laser irradiation is controlled as 20 s.

(4) After laser irradiation, digest the protein sample with trypsin according to manufacturer's protocol. Desalt the digested peptides and 'peptide cocktail' by using C18 NuTip to eliminate excess salts and remaining reagents.

Oxidation by hydroxyl radicals

Adopt the same experimental procedure as in the above description with the reagents in glycine-NaOH buffer, except add 1 µL of 50 mM histidine stock solution into 46 µL protein before the final addition of hydrogen peroxide solution.

LC-MS/MS analysis

Dilute 5 µL digested protein sample into 45 µL water with 0.1% formic acid and centrifuge for 3 min. Then load the sample solution onto a C18 reversed-phase desalting column at 4 µL/min for 10 min. Use a custom-packed analytical column for sample separation on Ultimate 3000 Rapid Separation with C18 reversed-phase material in silica tubing. Use gradient solvents A (water with 0.1% formic acid by volume) and B (80% acetonitrile with 0.1% formic acid by volume). Control the flowrate at 400 nL/min with the following gradient: 2% B to 60% B in 60 min, increase to 90% in 2 min, maintain at 90% for 5 min, return to 2% B in 1 min and equilibrate at 2% B for 7 min. Use a Thermo Q Exactive Plus orbitrap mass spectrometer (or other suitable proteomics instrument) coupled with a Nanospray Flex source for downstream detection with 2.5 kV spray voltage at 250 °C. Acquire the data in the data-dependent mode, where the 10 most abundant ions are selected for "higher energy" collisional dissociation (HCD).

2.7 Results and Discussion

2.7.1 Generation of Carbonate Radical Anion on FPOP Platform

To demonstrate the oxidative modification by carbonate radical anion, we chose apo-myoglobin(aMb) as the model protein. We conducted the experiments with 15 mM hydrogen peroxide, as the radical precursor, at pH = 10 in 700 mM CO_3^{2-}/HCO_3^- buffer to maximize $CO_3^{-\bullet}$ formation and in 700 mM glycine-NaOH buffer to give only $HO^\bullet$ as a parallel control. In the negative control experiments, the solution composition, including ion strength of the buffer, is maintained as the same. With no laser irradiation (Figure 2.3a), the unmodified protein signal

(at charge state of +20) is the most intense one followed by a small signal for the oxidized species seen with a +15.9949 Da mass shift, showing a basal oxidation. Irradiation by the KrF laser under the same conditions generates $CO_3^{-\bullet}$ and oxidizes aMb to give a more intense signals for singly oxidized species (+15.9949 Da), doubly oxidized species (+31.9898 Da) and triply oxidized species (+47.9847 Da), etc. (Figure 2.3b). As compared to the hydroxyl-radical oxidation in the glycine-NaOH buffer (Figure 2.3c), the two oxidation profiles showed good similarity, demonstrating that $CO_3^{-\bullet}$ oxidation also results in the addition of 'an oxygen' to the target protein. This is consistent with an oxidation mechanism of $CO_3^{-\bullet}$ whereby the formation of protein radical occurs by hydrogen atom abstraction, followed by an addition of dioxygen from dissolved O_2 in solution, reactions that are similar to those of hydroxyl radicals[1].

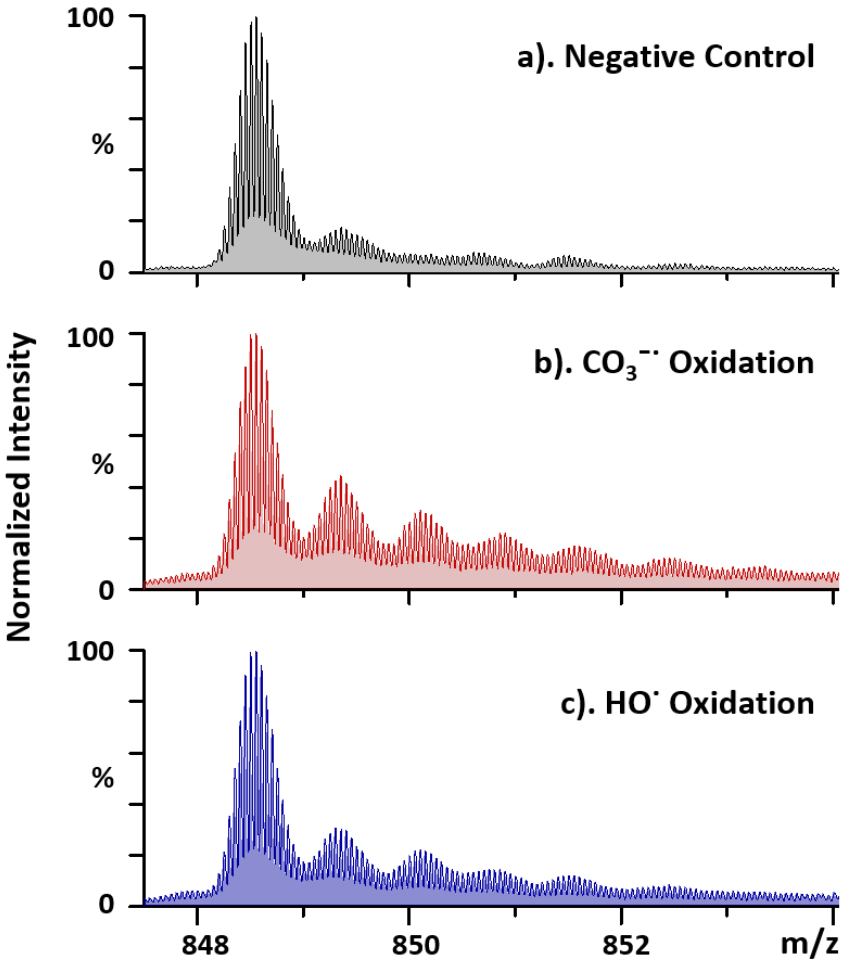


Figure 2.3 Oxidation profiles of apo-myoglobin (aMb). The quadrupole time-of-flight (QTOF) mass spectra of the +20 charge state of aMb with labeling conditions of (a) in CO_3^{2-}/HCO_3^- buffer without laser irradiation (negative control); (b) in CO_3^{2-}/HCO_3^- buffer with laser irradiation; (c) in glycine-NaOH buffer with laser irradiation, oxidized by hydroxide radical at global level.

2.7.2 Selectivity and Reactivity of $CO_3^{-\bullet}$ with Model Peptides

Utilizing the conditions that favor formation of $CO_3^{-\bullet}$ initiation, we employed model peptides with little or no higher order structure for a comprehensive study of its intrinsic residue-level specificity. Along with $CO_3^{-\bullet}$ chemistry, hydroxyl radical oxidation was also performed for batch control and for comparison purposes, as previously described. To examine peptides, we

prepared a “peptide cocktail” containing bombesin acetate salt hydrate, α-melanocyte stimulating hormone (α-Mel), angiotensin I bovine, leucine enkephalin (Leu-enkephalin) and bradykinin and mixed the solutions with hydrogen peroxide in two different buffers at pH = 10. Oxidation percentages of the modified amino acids (Figure 2.4) is shown in the decreasing order upon the reactivity of $CO_3^{-\bullet}$ and has been corrected by subtracting basal oxidation from negative control as shown in Figure 2.3a. Methionine and tryptophan are the most reactive amino acids for both radicals, where nearly 90% becomes oxidized by $CO_3^{-\bullet}$ and 60% by $HO^{\bullet}$. A similar phenomenon occurred for tyrosine where modification by $CO_3^{-\bullet}$ is almost 8-times higher than that by $HO^{\bullet}$. One possible reason is the weaker oxidative character leads to considerably longer lifetime for $CO_3^{-\bullet}$. The more specific reactivity towards selected amino acids for $CO_3^{-\bullet}$ will likely give more oxidation of those reactive residues. Phenylalanine and histidine can also react with $CO_3^{-\bullet}$ with similar reactivity as for $HO^{\bullet}$. Oxidation of other aliphatic amino acids (e.g., proline on bradykinin) occurs with $HO^{\bullet}$ treatment. For the $CO_3^{-\bullet}$ condition, however, only phenylalanine oxidation occurred. From all of the evidence, we conclude that $CO_3^{-\bullet}$ is a more selective oxidant than $HO^{\bullet}$, and the order of inherent reactivity of $CO_3^{-\bullet}$ is:

$M \approx W > Y > H \approx F$.

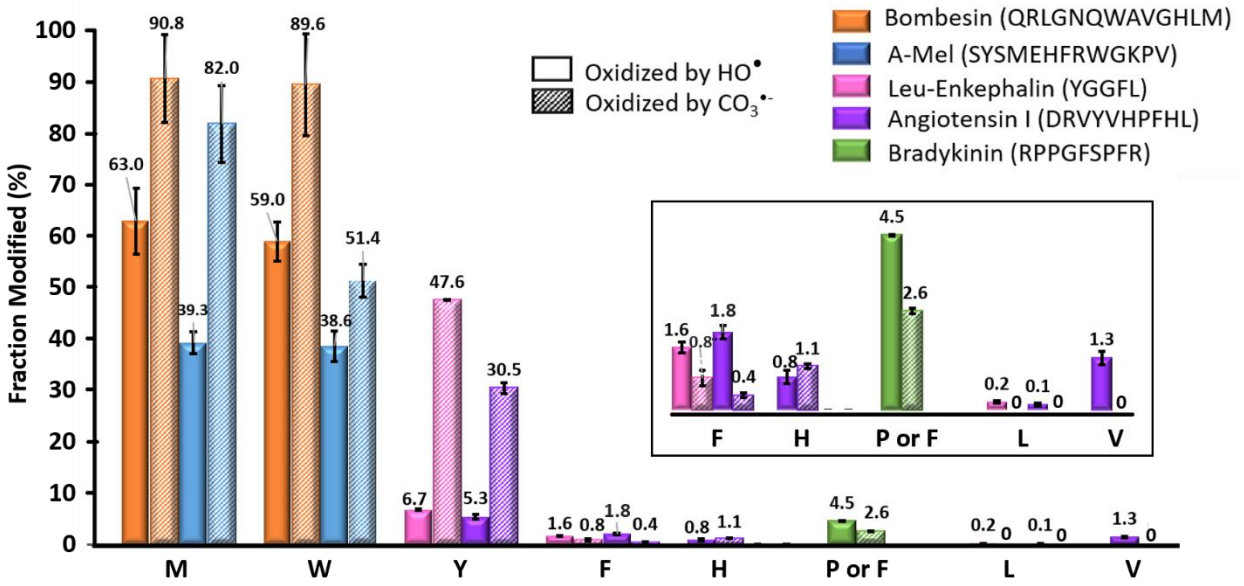


Figure 2.4 Comparison of the residue-level fraction modified (in percentage) of a 'Peptide Cocktail' by $CO_3^{-\bullet}$ and $HO^{\bullet}$. Different color represents different peptides and the sequence are shown correspondingly. Solid bars denote the $HO^{\bullet}$ oxidation extent and the patterned bars denote the $CO_3^{-\bullet}$ oxidation. Presented data has been corrected for negative control. Error bars are the standard deviations of three independent runs. The inset is an enlarged portion of the figure. For bradykinin, only F undergoes oxidation with $CO_3^{-\bullet}$, whereas F and P oxidation are both oxidized but not distinguishable because the chromatograms of the two modified peptides overlap.

The results are consistent with those of a previous investigation[18] where the reactivity of free amino acids in solution was measured, and in which methionine, tryptophan and tyrosine were found to be the most reactive amino acids towards $CO_3^{-\bullet}$ with rate constants of 1.2×10^8 $M^{-1}s^{-1}$, 4.4×10^8 $M^{-1}s^{-1}$ and 2.9×10^8 $M^{-1}s^{-1}$, respectively. Although the reported rate constant for histidine is almost 100-fold larger than that for phenylalanine (i.e., rate constants of 7.0×10^6 $M^{-1}s^{-1}$ and 5.0×10^4 $M^{-1}s^{1}$), the reactivities of the two residues are similar when they are part of proteins and peptides. It is worth noting that the electrostatic interaction between negatively charged carbonate radical anion and the positive charged imidazole group at lower pH may contribute to

an increased reactivity. Therefore, in real biological systems, the reaction rate with histidine may vary depending on the local pH.

2.7.3 Residue-Resolved Modification Measurement by LC-MS/MS

We quantified the oxidation extent at the residue level from the peak areas of the chromatograms. Extracted ion chromatograms (EICs) for each modified and unmodified species were exported from raw files with Qual Browser (Thermo Xcalibur 2.2 software) (Figure 2.5). The (+16 Da) angiotensin I oxidized by the carbonate radical anion (Figure 2.5b) and by the hydroxyl radical (Figure 2.5c) are listed below the unmodified peptides. Different colors represent different residues that are modified. The modification fraction for a residue is calculated with equation 2.2, where A_{ox} is the peak area for the oxidized-species signal and A_{un} is the peak area for the unmodified. Chromatographic peaks are fitted by Gaussian Function and the averaged R^2 is 0.97. For complex with overlapping signals of oxidized residues (Figure 2.5c), the peak representing tyrosine oxidation (purple in Figure 2.5b) is scaled into the lower panel, and the oxidation extent of valine is represented by the grey gaussian peak. In addition, each product-ion (MS/MS) spectrum (e.g., for Figure 2.6) was manually validated to assure an accurate oxidizing residue assignment. Comparing the signals for the fragment ions, we see a clear +16 Da mass shift for b4 and subsequent b ions for the modified peptide as compared with the unmodified peptide, indicating that oxidation took place on the Y4 residue.

$$Oxidation\ extent = \frac{\sum A_{ox}}{\sum A_{ox} + A_{un}} \quad (2.2)$$

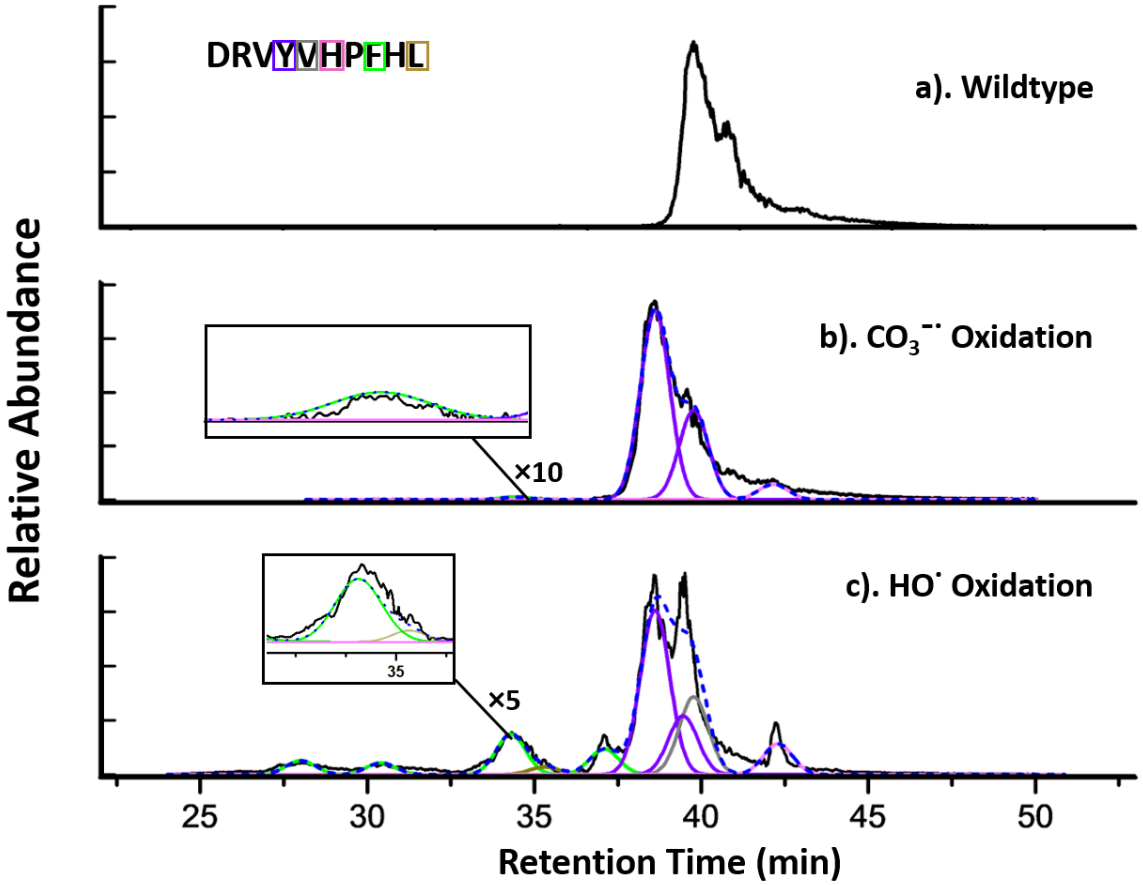


Figure 2.5 Extracted ion chromatogram (EICs) of angiotensin I. Specifically, the wildtype (a), and the oxidized angiotensin I by $CO_3^{-\bullet}$ (b) and $HO^{\bullet}$. (c). Different colors (coded with the chromatogram colors) indicate different oxidized residues as determined by product-ion (MS/MS) spectra. Each peak is fitted by Gaussian Function and the overall fitting curve is presented as blue dotted line. The averaged R^2 for the fitting is 0.97.

From the shown chromatogram (Figure 2.5), it clearly shows that the number of oxidized isomeric peptides by the $CO_3^{-\bullet}$ is significantly less than that by the $HO^{\bullet}$. In the $CO_3^{-\bullet}$ oxidation condition, we resolved at least 4 isomeric peptides, two of which undergo oxidation on Tyr, and two other minor ones involve oxidation on His and Phe. For $HO^{\bullet}$. oxidation, however, a more complex pattern of oxidation occurs presenting at least 6 isomeric products, including oxidations on Val, Leu, Tyr, His and Phe with much more discernible amount.

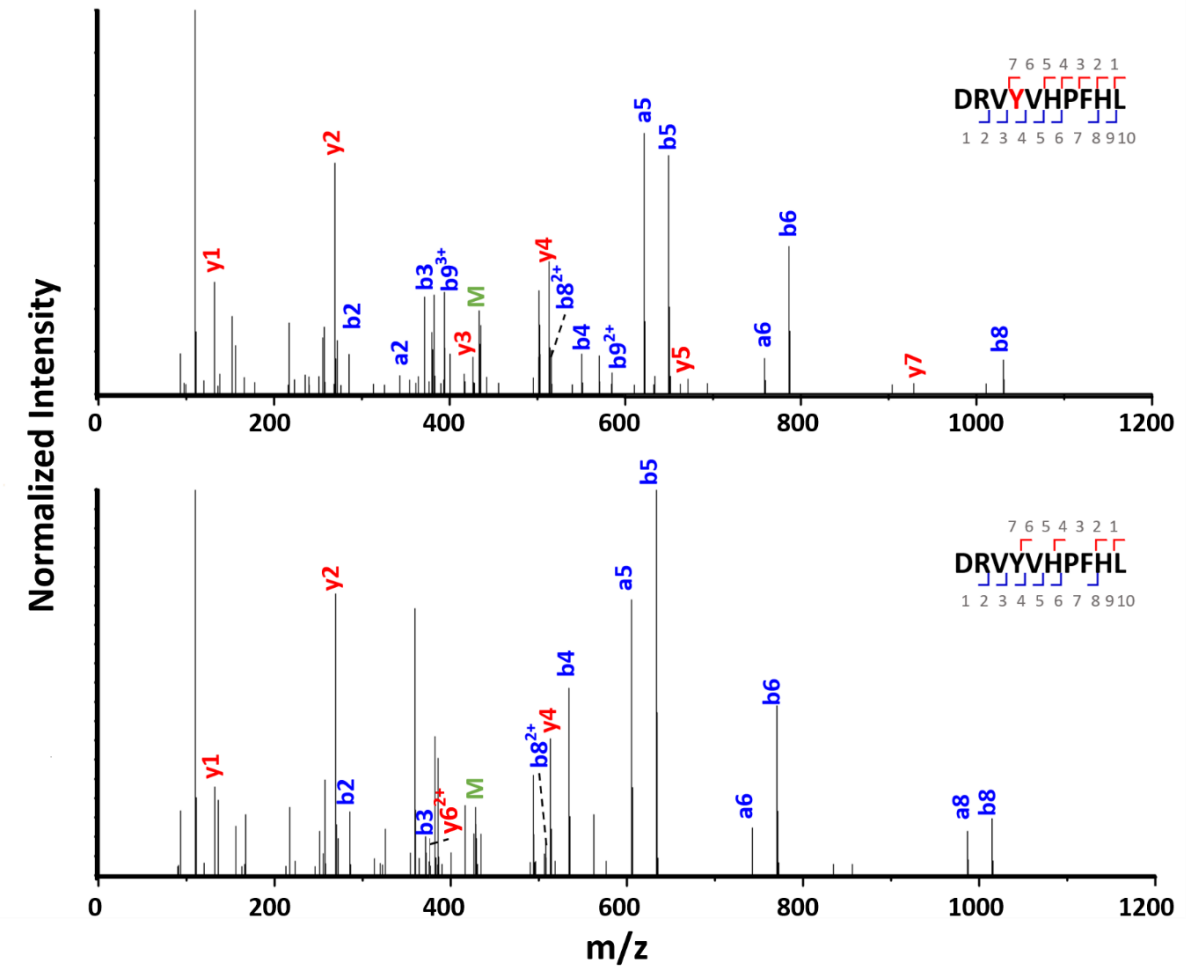


Figure 2.6 Product-ion (MS/MS) spectra of oxidized and wildtype angiotensin I.

2.7.4 Selectivity and Reactivity of $CO_3^{-\bullet}$ with Model Proteins

In addition to peptides mixture, we further investigate the selectivity and specificity of $CO_3^{-\bullet}$ with proteins, where apo-myoglobin and ubiquitin were chosen as the model systems. Experiments were carried out with $CO_3^{-\bullet}$ dominant condition and $HO^{\bullet}$ governing condition individually followed by "bottom-up" proteomics. After integrating the signals obtained through LC-MS/MS, we can list the modified amino acids for each protein (Figure 2.7). From the ubiquitin data set (Figure 2.7a), $CO_3^{-\bullet}$ gives the most oxidation on methionine (M), as does $HO^{\bullet}$, underscoring their oxidation preference towards methionine. Other reactive residues include the aromatic amino acids phenylalanine (F) 4 and 45, tyrosine (Y) 59 and histidine (H)

68 undergo moderate oxidation by $CO_3^{-\bullet}$ as compared to $HO^{\bullet}$. Most other amino acids that are reactive with $HO^{\bullet}$ (e.g., valine (V), leucine (L), glutamine (Q), lysine (K), proline (P)) are unreactive with $CO_3^{-\bullet}$.

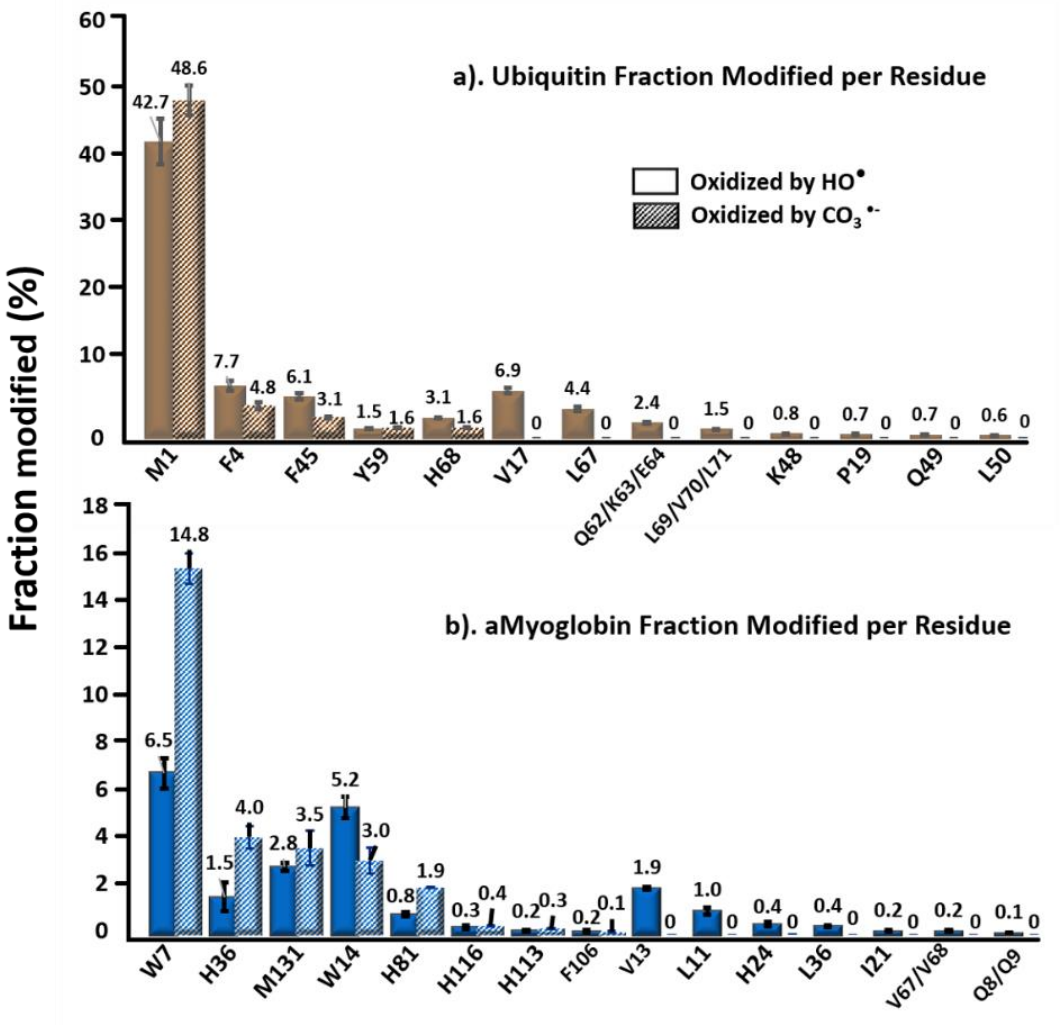


Figure 2.7 Comparison of residue-level fraction modified (in percentage) of (a) ubiquitin and (b) apo-myoglobin modified by carbonate radical ($CO_3^{-\bullet}$) and hydroxyl radical ($HO^{\bullet}$). Solid bars denote the hydroxyl radical oxidation and the patterned bars denote the carbonate radical anion oxidation. Presented data has been corrected for negative control. Error bars are standard deviations of three independent runs.

For apo-myoglobin (Figure 2.7b), tryptophan W7 exhibits the highest oxidation extent on by $CO_3^{-\bullet}$ showing a doubled fraction modified value as compared to that for $HO^{\bullet}$. Histidine also

gives nearly doubled modification fraction by $CO_3^{-\bullet}$ oxidation, for example H36 and H81, indicating its considerable reactivity. The oxidation behavior towards methionine is consistent with the results for ubiquitin, where $CO_3^{-\bullet}$ oxidation slightly favors that of $HO^{\bullet}$, shown as M131. Other potentially reactive residues like H113, H116 and F106 undergo an oxidation extent of less than 0.5%, just as with $HO^{\bullet}$. The reason is likely to be the small solvent accessible area (SASA) for these sidechains on protein surface. In addition, the size difference in $CO_3^{-\bullet}$ and $HO^{\bullet}$ may contribute to the reaction extent on a given amino acid. For example, the oxidation percentage of W14 is 3%, which is nearly 5-times lower than that of W7 (14.8 %). The SASA of W7 and W14 is 15.3 $Å^2$ and 6.8 $Å^2$, respectively by calculation[11]. Thus, for the smaller $HO^{\bullet}$, the oxidation extent is greater than that for the larger $CO_3^{-\bullet}$, which has difficulty penetrating the surface to react with W14. Other residues like V13/67/68, L11/86, I21 and Q8/Q9 are relatively inert to $CO_3^{-\bullet}$ with no shown oxidation.

Compared to the results for the model peptides, the $CO_3^{-\bullet}$ oxidative preference towards Met and Trp over $HO^{\bullet}$ is consistent and so does the negligible reactivity towards aliphatic amino acids. However, for tyrosine, oxidation extent by $CO_3^{-\bullet}$ in ubiquitin (Y59 in Figure 2.7a) is much less than that with model peptides, shown as 1.6% vs 30-50%. The difference arises from the location of the Y59 in the 3D structure and its limited SASA, emphasizing the importance of incorporating peptides with minimal higher order structure to eliminate conformational effect in the discussion of residue-level specificity.

2.8 Advantages and Caveats

We optimized the production of carbonate radical anion on a FPOP platform by using a reaction cascade scheme starting with the hydroxyl radical. This approach enables the study of peptides and proteins to determine the relative reactivities of various amino acid sites in those species.

Although the results for the protein/peptides systems are largely consistent with those of an early study carried out with free amino acids, this new approach provides direct evidence for the radical reactions of proteins (and other biomolecules) with radicals, here the carbonate radical anion, demonstrating clearly those residues or regions of the biomolecule that are reactive.

The approach suggests that a general approach can be achieved by implementing the FPOP platform to generate either primary or secondary biologically relevant radicals (by cascade reactions) and to determine their fundamental reactivities with biomolecules. Primary $HO^{\bullet}$ and $SO_4^{-\bullet}$ radicals can be formed from peroxy-compounds. Generation of secondary radicals (e.g., $CO_3^{-\bullet}$, $NO_2^{\bullet}$, halides) may be optimized by varying buffer contents, pH, and radical precursor and by involving a radical scavenger. In the current study, this was done by using kinetic simulations. For other radicals for which kinetic data do not exist, the conditions must be optimized by using mass spectrometry analysis. Other spectroscopy techniques (e.g., fluorescence microscopy) can also be incorporated in the post-modified identification.

Although we used no scavenging reagents in the current study for the generation of carbonate radical anion, incorporating such reagents can be done in future studies. The use of histidine as a scavenger in the hydroxyl-radical generation not only controls radical lifetime but also intercepts radicals and minimizes protein modification, a point that may be important for physiologically relevant studies. For example, biological systems respond to different oxidative stress, and the oxidizing potential can be varied on the FPOP platform. In addition, the FPOP flow design effectively avoids over-modification of a target and improves labeling homogeneity by controlling flow rate, exclusion volume, and radical lifetime. We suggest that this repurposed FPOP platform coupled with mass spectrometry-based analysis is a new option for studies of radical generation and their reactions with biomolecules.

2.9 Acknowledgments

This study contains combined efforts from all authors. M.M.Z. conducted the experiments and wrote the manuscript. D.L.R. run the simulation. M.L.G edited the manuscript. The study was supported by National Institute of General Medical Sciences grant 5P41GM103422 and by 1S10OD016298 for instrumentation.

2.10 References

1. Xu, G. H.; Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **2007**, *107*, 3514-3543.

2. 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*, 3965-3973.

3. 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*, 2057-63.

4. Li, K. S.; Chen, G. D.; Mo, J. 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*, 7742-7749.

5. 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. D.; Tyrniak, 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*, 2250-2258.

6. Yan, Y. T.; Chen, G. D.; Wei, H.; Huang, R. Y. C.; Mo, J. 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*, 2084-2092.

7. 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*, 12090-12098.

8. Hart, K. M.; Ho, C. M. W.; Dutta, S.; Gross, M. L.; Bowman, G. R. Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* **2016**, *7*, 1-10.

9. 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. *Proc. Natl. Acad. Sci. U. S. A.* **20**14, *111*, E2596-605.

10. Chen, J. W.; 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*, 18724-18731.

11. 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*, 7821-7827.

12. 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*, 552-5.

13. 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*, 14007-14010.

14. Medinas, D. B.; Cerchiaro, G.; Trindade, D. F.; Augusto, O. The carbonate radical and related oxidants derived from bicarbonate buffer. *Iubmb Life* **2007**, *59*, 255-262.

15. Sankarapandi, S.; Zweier, J. L. Bicarbonate is required for the peroxidase function of Cu,Zn-superoxide dismutase at physiological pH. *J. Biol. Chem.* **1999**, *274*, 1226-1232.

16. Hodgson, E. K.; Fridovich, I. The mechanism of the activity-dependent luminescence of xanthine oxidase. *Arch. Biochem. Biophys.* **1976**, *172*, 202-205.

17. Radi, R. Nitric oxide, oxidants, and protein tyrosine nitration. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 4003-4008.

18. Neta, P.; Huie, R. E.; Ross, A. B. Rate Constants for Reactions of Inorganic Radicals in Aqueous-Solution. *J. Phys. Chem. Ref. Data* **1988**, *17*, 1027-1284.

19. Shafirovich, V.; Dourandin, A.; Huang, W. D.; Geacintov, N. E. The carbonate radical is a site-selective oxidizing agent of guanine in double-stranded oligonucleotides. *J. Biol. Chem.* **2001**, *276*, 24621-24626.

20. Paviani, V.; Queiroz, R. F.; Marques, E. F.; Di Mascio, P.; Augusto, O. Production of lysozyme and lysozyme-superoxide dismutase dimers bound by a ditryptophan cross-link in carbonate radical-treated lysozyme. *Free. Radical Bio. Med.* **2015**, *89*, 72-82.

21. Surmeli, N. B.; Litterman, N. K.; Miller, A. F.; Groves, J. T. Peroxynitrite Mediates Active Site Tyrosine Nitration in Manganese Superoxide Dismutase. Evidence of a Role for the Carbonate Radical Anion. *J. Am. Chem. Soc.* **2010**, *132*, 17174-17185.

22. Gebicka, L.; Didik, J.; Gebicki, J. Reactions of heme proteins with carbonate radical anion. *Res. Chem. Intermediat.* **2009**, *35*, 401-409.

23. Chen, S. N.; Hoffman, M. Z. Rate constants for the reaction of the carbonate radical with compounds of biochemical interest in neutral aqueous solution. *Radiat. Res.* **1973**, *56*, 40-7.

24. Boccini, F.; Domazou, A. S.; Herold, S. Pulse radiolysis studies of the reactions of CO3 center dot- and NO2 center dot with nitrosyl(II)myoglobin and nitrosyl(II)hemoglobin. *J. Phys. Chem. A* **2006**, *110*, 3927-3932.

25. Zhang, H.; Joseph, J.; Crow, J.; Kalyanaraman, B. Mass spectral evidence for carbonate-anion-radical-induced posttranslational modification of tryptophan to kynurenfne in human Cu, Zn superoxide dismutase. *Free. Radical Bio. Med.* **2004**, *37*, 2018-2026.

26. Joffe, A.; Geacintov, N. E.; Shafirovich, V. DNA lesions derived from the site selective oxidation of Guanine by carbonate radical anions. *Chem. Res. Toxicol* **2003**, *16*, 1528-38.

27. Buxton, G. V.; Elliot, A. J. Rate-Constant for Reaction of Hydroxyl Radicals with Bicarbonate *Ions. Radiat. Phys. Chem.* **1986**, *27*, 241-243.

28. Schulz, K. G.; Riebesell, U.; Rost, B.; Thoms, S.; Zeebe, R. E. Determination of the rate constants for the carbon dioxide to bicarbonate inter-conversion in pH-buffered seawater systems. *Mar. Chem.* **2006**, *100*, 53-65.

29. 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*, 843-846.

Chapter 3: Free Radical Footprinting to Probe Unfolding of Holo-Myoglobin Triggered by Laser-induced pH jump

3.1 Abstract

Understanding protein folding/unfolding is crucial to gain deeper insights of related biological processes. Commonly adopted characterization methods to follow protein conformational changes are circular dichroism, fluorescence, FT-IT and NMR. Among them, the spectroscopic approaches are the most extensively used, given the fast detection timescale at femtoseconds to microseconds. However, their spatial resolution is low. Mass spectrometry-based approaches enable residue-level resolution, and they may become an asset to probe protein conformational changes. Whereas the major limitation is the time resolution that typically is milliseconds, too long for protein folding/unfolding happening at microseconds. We designed and established a platform by integrating fast photochemical oxidation of protein (FPOP) and a laser-induced pH jump to allow the fast characterization at the microseconds to low milliseconds timescale. Using holo-myoglobin as a model system, we considered different footprinting methods (e.g., hydroxyl radical and carbene diradical footprinting), establishing a foundation for future research.

3.2 Introduction

Protein folding/unfolding is essential for understanding biological processes, reaction mechanisms, and basic interactions that are important in therapeutics [1-3]. To study protein conformational changes in kinetics experiments, investigators commonly use spectroscopic methods such as fluorescence spectroscopy. The basic principle is to measure the fluorescence of some amino acid residues (e.g., tryptophan) that fluoresce differently when buried inside a protein or exposed on its surface. The approach provides usually global information to probe protein folding/unfolding. Infrared spectroscopy, another optical method, can provide some local structural information.

Mass spectrometry-based characterization of protein folding/unfolding started in the early 1990s with the development of pulsed H/D amide exchange and covalent labeling by chemical reagents of equilibrium states. [4-5] The most pronounced advantages of the MS-based approaches is the ability to pinpoint the amino acid residues involved in the protein conformational changes, thereby increasing the spatial resolution achievable by optical approaches. The limitation, however, is time resolution, typically milliseconds, a time scale that is longer than most folding events occurring at microsecond. [6]

One representative study was carried out by Konermann and coworkers[7] who took advantages of the fast photochemical oxidation of protein (FPOP)[8], a platform that can generate hydroxyl radical ($HO^{\bullet}$) through photolysis of hydrogen peroxide on a flow system and label the solvent accessible surface area (SASA) of different protein conformations at low microseconds. In combination with a continuous-flow rapid mixing apparatus, the protein can be denatured from 50 ms to 5 min prior to $HO^{\bullet}$ footprinting. Later, Gross and Chen[9] published a breakthrough study at 2010. They integrated a laser-induced temperature jump and the FPOP platform to probe the folding process of barstar protein at microseconds to low milliseconds timescale. The two lasers, one to perturb protein conformations, and another to form $HO^{\bullet}$ radicals are aligned carefully intersecting at the same transparent window, and the needed time delay between the two lasers is enabled by a signal generator with a delay circuit. By controlling the delay time from 0.2 to 1.0 ms, they traced the structural changes of barstar by analyzing oxidative modifications. The shift of mass centroid was further fitted as a function of delay time by using single-exponential function, providing the rate constant for the transition from the unfolded to the first intermediate state. After bottom-up proteolysis, the key residues associated with barstar folding during the first stage of folding were also resolved, demonstrating the capability of this

approach to achieve enhanced spatial resolution to study protein folding/unfolding at sub-millisecond time scale.[10]

Besides adding a denaturant or introducing a temperature jump, other perturbations can induce protein folding/unfolding. One of many is a laser-initiated pH jump, which was demonstrated frequently by using holo-myoglobin (hMyo) [11-13] with spectroscopic detection. It is known that conformational change of hMyo is determined by the solubility of the heme.[14] The heme group is less soluble in acidic conditions, therefore having a tendency to remain inside the protein's hydrophobic binding pocket. In addition, the heme-protein interaction is thermodynamically more favorable, promoting the persistence of a native-like conformation (closed state). Basic pH, on the other hand, facilitates heme release, pushing the equilibrium towards a relative unfolded conformation (open state). In a detailed study published at 2006, Corrie and coworkers[11] observed the two-state transition from pH 3.5 to 4.5. The pH jump is induced by the irradiation on a widely used photoacid, 2-nitrobenzaldehyde (NBA). The kinetics of hMyo unfolding could be described with a triple exponential function, giving the lifetimes of three dominant intermediates that are 18 ± 1 ms, 0.53 ± 0.01 s and 0.94 ±0.01 s.

In chapter 2, we demonstrated one way to elaborate the FPOP platform, that is to develop new radical footprinting reagents. In this study, we aim to develop the FPOP platform from another aspect. We are interested in advancing the two-laser platform for broader applications, specifically for pH-sensitive proteins. hMyo is a good model system for this purpose. We designed an integrated platform wherein, by tuning the laser frequency, exclusion volume, and flow rate of the sample solution, the delay time between the two lasers can be controlled to capture different unfolded states. In addition to $HO^{\bullet}$ footprinting, we also evaluated carbene labeling as a probe for protein folding/unfolding, establishing the foundation for further research.

3.3 Material and Method

3.3.1 Material

Sodium chloride, tris-buffer, citric acid, 2-nitrobenzaldehyde, methionine, catalase, holo-myoglobin, and apo-myoglobin were purchased from Millipore Sigma (Saint Louis, MO). 1-(2-Nitrophenyl)ethyl sulfate was purchased from ApexBio Technology (Houston, TX). 3-(3-Methyl-3H-diazirin-3-yl) propan-1-ol was purchased from Ambeed (Arlington, IL).

3.3.2 Hydroxyl Radical Footprinting

For the equilibrium conditions, hMyo (4 μM) was dissolved in citrate-phosphate to achieve pHs less than 7, and in tris-buffer for pHs greater than 7. Incubation took about 30 mins prior to constitution of the sample solution which contains 5 mM H_2O_2 and 1 mM histidine. The collection tube includes 10 μL of 70 mM methionine and 1 μL of 500 nM catalase for post-oxidation quenching.

For the pH-jump experiments, deionized water was boiled overnight in a three-neck round bottom flask under nitrogen protection. A syringe of 50 mL with a long needle was used to transfer the degassed water into an Erlenmeyer flask under a continuous flow of nitrogen. A self-built Schlenk line connects a nitrogen tank with the flask and with three others to protect the reagents such as saturated NBA/NaCl solution and protein solution, from absorbing CO_2 throughout the experiments. The NBA reagent and hMyo (or aMyo) powder were kept and dissolved under nitrogen protection to avoid further introduction of CO_2. Specific H_2O_2 and histidine concentration are indicated correspondingly in the discussion. Measurements of pH were performed on a SevenEasy pH meter (Columbus, OH). The experimental for the laser parameters are: ~ 25.00 mJ, 16 kV for KrF laser; ~ 10 mJ for YAG laser with maximized amplification; ~ 1.95 mm focal size; frequency (repetition rate) of 5 Hz; ~ 12.5 μL/min flow rate.

Mass spectrometry characterization was conducted on Bruker MaXis 4G on the global level to measure modification extent at the intact protein level.

3.3.3 Carbene Footprinting

Carbene footprinting was performed similarly as the description above except no methionine and catalase were in the collection tube. The concentration of MDPO in stock solution was 200 mM in NaCl, and the final concentration was 10 mM.

3.4 Experimental Design

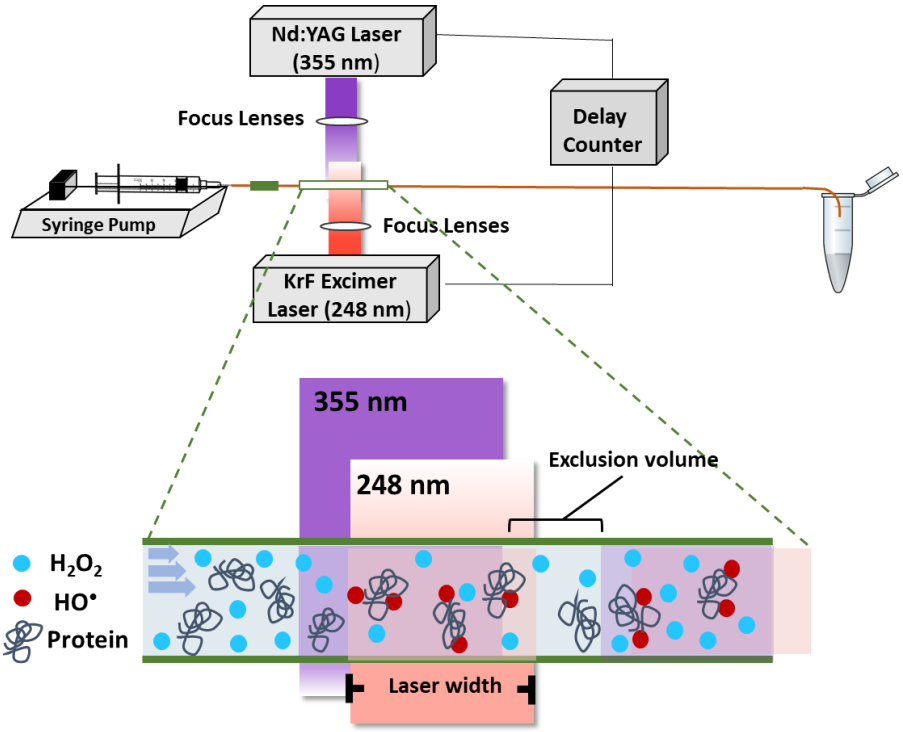


Figure 3.1 Two-laser platform that induces a pH jump at 355 nm followed by hydroxyl radical generation at 248 nm for protein footprinting.

To probe protein conformational changes in-time, we coupled two lasers, namely the KrF excimer laser with a wavelength at 248 nm and the Nd:YAG laser at 355 nm, into an integrated platform (Figure 3.1). Upon irradiation at 355 nm, the photoacid contained in the sample solution is excited and dissociates protons to lower the pH and further induce the folding/unfolding

process of a pH-sensitive protein. The second laser at 248 nm then photolyzes the constituent H_2O_2, generating $HO^{\bullet}$ to map the changing SASA of the protein. A delay counter is configured between the two lasers, allowing $HO^{\bullet}$ footprinting to be executed at several delays (microseconds to milliseconds) after the initial conformational change. The upper limit of the monitoring time is determined by the exclusion volume, a design that avoids double labeling of the targeted protein, and the laser frequency. In this study, the two lasers were both operated at every 200 ms (f = 5 Hz) with an exclusion volume of 20 % at the flow rate of 12 μL/min. The upper limit of the time delay is therefore 40 ms (200×20%). Other settings, including the focus lenses for both lasers, the syringe pump, and the quenching solution in the collection tube, are the same with a typical FPOP experimental setup.

3.5 Results and Discussion

3.5.1. Conformational Changes of Holo-Myoglobin at Equilibrium

Although investigators followed the unfolding of hMyo with spectroscopic methods, the unfolding process, which we seek to characterize by $HO^{\bullet}$ footprinting, is unknown. To understand the conformational changes in the context of SASA changes, we incubated hMyo individually at different pHs from 3.2 to 8.0. All protein solutions, with 19 pH conditions, were equilibrated in buffers for 30 mins prior to the $HO^{\bullet}$ labeling, initiated by 248 nm laser. Each experiment was performed in duplicate. Representative oxidation profiles are shown in Figure 3.2. The first peak represents the unmodified protein (m/z = 848.6 at charge +20), followed by a series of higher-molecular species giving consecutive +16 oxidation. The number of the oxidized proteins are different at various pHs, indicating different extents of oxidation that reflect different protein conformations. For better comparison, we further calculated the modification percentage (eq 3.1) based on the intensity of all the peaks, e.g., +16, +32, +48, etc. We observed

higher oxidation percentages when the conditions are acidic. At pH = 3.6, the fractional oxidation is about 50 %, indicating a relatively open conformation comparing to that in conditions with larger pH values.

$$Oxidation\ percentage = \frac{\Sigma I_{ox}}{\Sigma I_{ox} + I_{un}} \text{ (eq 3.1)}$$

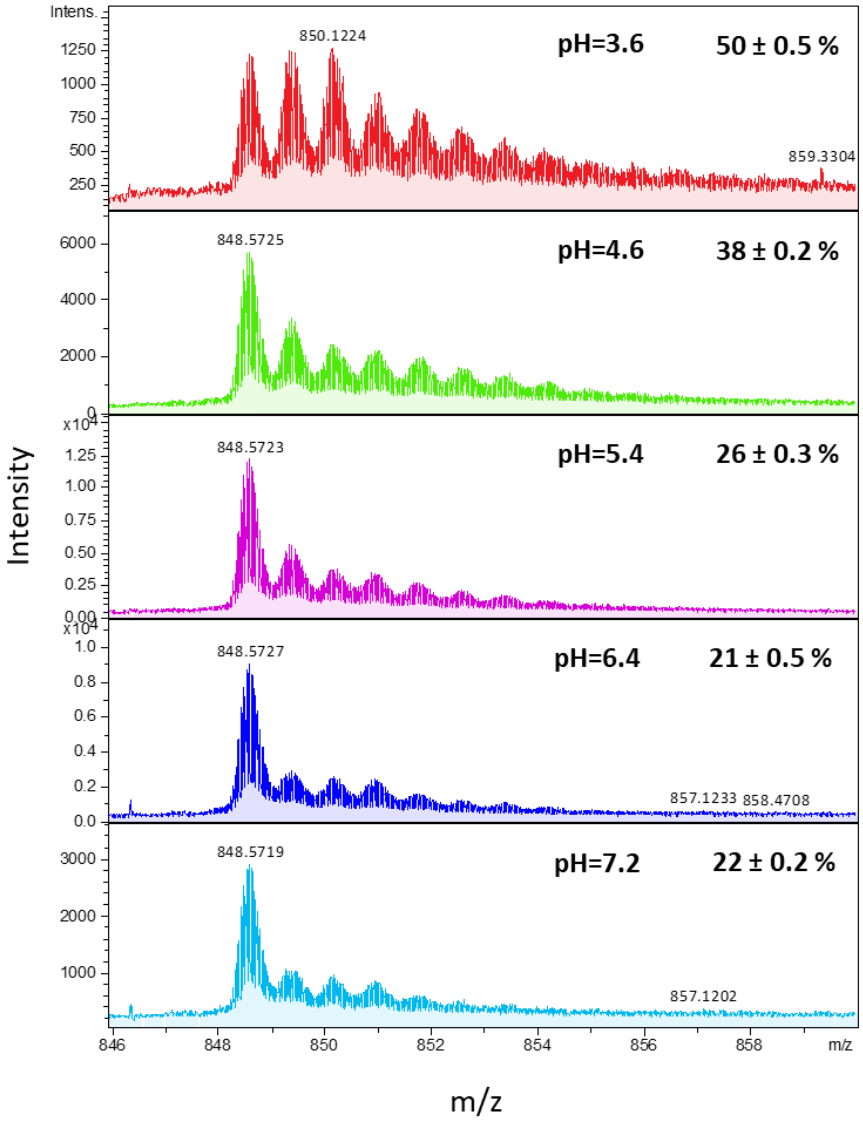


Figure 3.2 Oxidation profiles of hMyo by hydroxyl radical footprinting at several pHs under equilibrium conditions. The peak series represent the protein of charge +20.

We further plotted the oxidation percentages as a function of pH and observed a two-state transition starting from a pH around 3.5 (Figure 3.3). When the pH increased to 5.5, the

oxidation extent dropped significantly by about 25% and then decreased gradually at pH values larger than 5.5. The observed pH range accountable for the conformational transition is similar to that in a previous study as monitored by UV spectroscopy [11].

The successful $HO^{\bullet}$ footprinting of hMyo for equilibrium states laid the foundation for the following investigation, where the laser at 355 nm is incorporated to induce pH changes in advance.

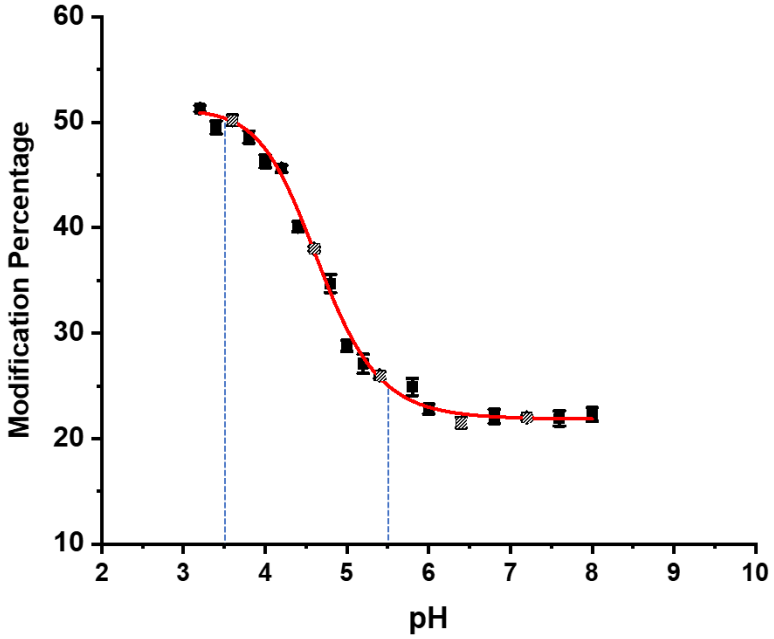


Figure 3.3 Modification percentage of hMyo as a function of pH. The blue dash lines indicate the estimated pH range for the two-state transition. The patterned data points correspond to the representative MS spectrum in Figure 3.2. The fitting is based on a logistic function in Origin, and the correlation coefficient is $R^2 = 0.996$.

3.5.2 Control Experiments at 248 nm

To enable a pH-jump upon irradiation on photoacids, the protein solution should contain no buffers (only NaCl solution). We further screened the concentration of H_2O_2 and histidine that determines the dosage of $HO^{\bullet}$ to achieve comparable oxidation extents with the ones at equilibrium states. In addition, NaCl solution dissolves atmospheric CO_2, giving an acidic pH

around 6.2 at room temperature. Over-night degassing of the stock NaCl solution is needed in the experimental preparation, and the fresh NaCl solution requires nitrogen protection throughout the footprinting experiments. These procedures help maintain the NaCl solution at around neutral pH even after addition of the protein solution and the photoacid. The conditions we adopted are: 1.5 mM H_2O_2, 0.5 mM of His, and 4 uM of hMyo in NaCl, giving a pH of 6.8. In the presence of the saturated NBA solution (~ 8 mM), the pH further dropped to 6.3, but the hMyo is still retained in a native-like conformation (pH > 5.5).

To start with, we examined the oxidation profiles for the open and closed states of hMyo without incorporation of NBA. We prepared two protein solutions, one at pH 6.3 and the other at 3.2. The pHs of the two samples were adjusted with HCl solution to mimic the pH of initial and final situation, respectively, if using NBA for pH jump. The framed oxidation percentages at the closed state of hMyo is 23% (Figure 3.4a) and 48% for the open state (Figure 3.4b), similarly to those values in the calibration curve acquired for equilibrium states.

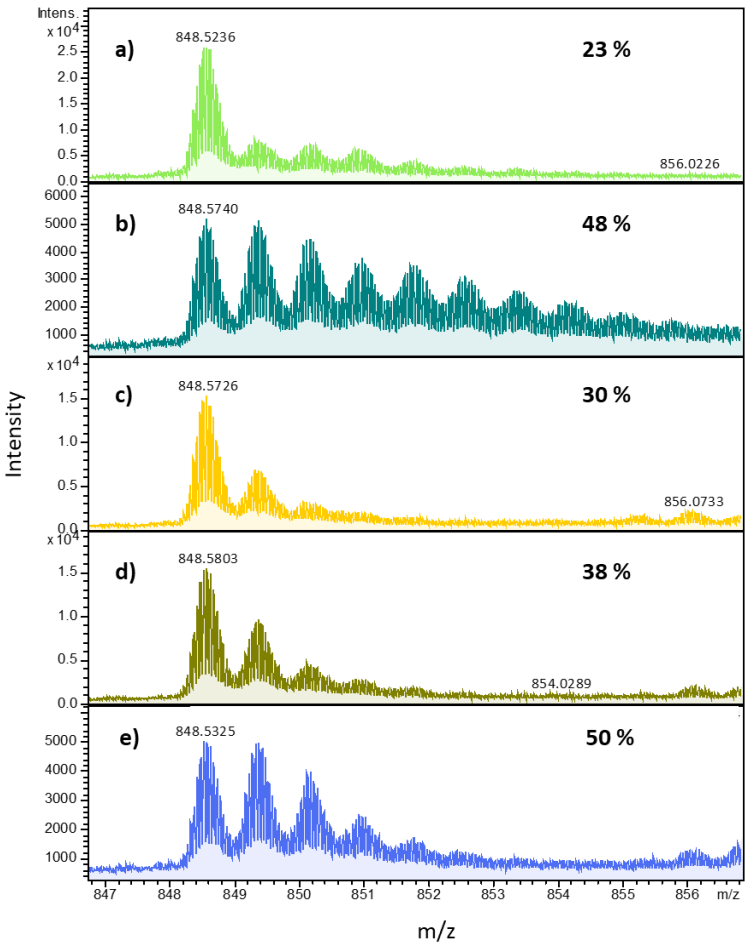


Figure 3.4 Oxidation profiles of hMyo upon only 248 nm laser irradiation under different conditions: a) hMyo (4 µM) with His (0.5 mM) and H_2O_2 (1.5 mM) at pH 6.3 in NaCl; b) hMyo (4 µM) with His (0.5 mM) and H_2O_2 (1.5 mM) at pH 3.2 in NaCl; c) hMyo (4 µM) with His (0.5 mM), H_2O_2 (1.5 mM) and NBA (~ 8 mM) at an initial pH of 6.3 in NaCl; d) hMyo (4 µM) with His (0.5 mM), H_2O_2 (1.5 mM) and NBA (~ 8 mM) at an initial pH of 3.2 in NaCl; e) hMyo (4 µM) with His (0.5 mM), H_2O_2 (7.5 mM) and NBA (~ 8 mM) at the initial pH of 3.2 in NaCl.

Then, we included saturated NBA (~ 8 mM) in the sample solution and evaluated the footprinting upon laser irradiation at 248 nm. Based on our design, the photoacid will be activated for proton release only at 355 nm and will desirably stay inert when $HO^{\bullet}$ is generated under 248 nm, the wavelength used for footprinting. We observed, however, a decrease in pH from 6.8 to 3.2 upon 248 nm irradiation, consistent with the UV absorption profile of NBA that absorbs light from 200 to 400 nm. Although the NBA being activated at 248 nm is not preferred,

the use of NBA in our experimental design may still work. The reason is that in an actual pH-jump experiment, if most of the NBA is exhausted by the first laser irradiation at 355 nm, the pH change caused by the remaining NBA at the second laser can be negligible. To test this, we measured the pH of the protein solution after 355 nm irradiation, obtaining a value of 3.2. The same solution was then recycled and submitted for the second laser at 248 nm. The final pH value is 3.0 ± 0.1, very similar to the value after the treatment at 355 nm, indicating a near-complete consumption of NBA before irradiation by the KrF laser.

We are now in position to test the labeling profiles by $HO^{\bullet}$ footprinting initiated at 248 nm. The closed conformation of hMyo (initial pH at 6.3, Figure 3.4c) gives an oxidation percentage of 30% whereas that of the open conformation (pH 3.2, Figure 3.4d) is only 38%, which is a much smaller modification level than the condition without NBA (Figure 3.4b). NBA appears to be a $HO^{\bullet}$ quencher, and its incorporation requires higher dosage of $HO^{\bullet}$ to produce a convincing difference in the oxidation extents to distinguish the conformational transitions. We further screened the H_2O_2 concentration and determined that we could increase it to 7.5 mM for the following experiments (Figure 3.4e).

3.5.3 Control Experiments at 355 nm

Besides evaluating the system with only 248 nm, we decided to conduct control experiments with only 355 nm. Ideally, only the photoacid NBA will be photolyzed under YAG laser (355 nm) for pH change while the H_2O_2 will stay unreactive only with a burst from the KrF laser. We incorporated either H_2O_2 (without NBA but with hMyo and histidine, Figure 3.5a) or NBA (without H_2O_2 but with hMyo and histidine, Figure 3.5b) in protein solutions and observed only the peaks corresponding to unmodified species, showing the results as per the design. When H_2O_2 and NBA were included together, however, hMyo became oxidized (Figure 3.5c),

indicating that there are reactions occurring between NBA and H_2O_2 and the generated reactive species are likely accountable for the oxidation.

To rule out that this is a protein-dependent chemistry, we performed similar experiments on two other proteins, calmodulin (Figure 3.5d-f) and β-lactoglobulin (Figure 3.5 h-g). Although the relative oxidation extents vary for different proteins, which may be accountable by the number of reactive residues exposed on the surface, all three proteins are clearly oxidized. On the other hand, we did notice that the oxidation percentage on hMyo is only 36% (Figure 3.5c), less than 50% that was observed previously (Figure 3.4e) with same H_2O_2 concentration. The difference between the two values may still provide us the chance to probe conformational changes of hMyo. It is likely that the unknown reactive species label hMyo almost instantaneously upon the proton releases. The labeling process can further induce perturbations in the protein conformations and potentially bias the unfolding and subsequent footprinting readout. Nevertheless, if the oxidized hMyo can still undergo conformational changes, this study may serve as the foundation for setting up a characterization platform to follow protein conformational transitions.

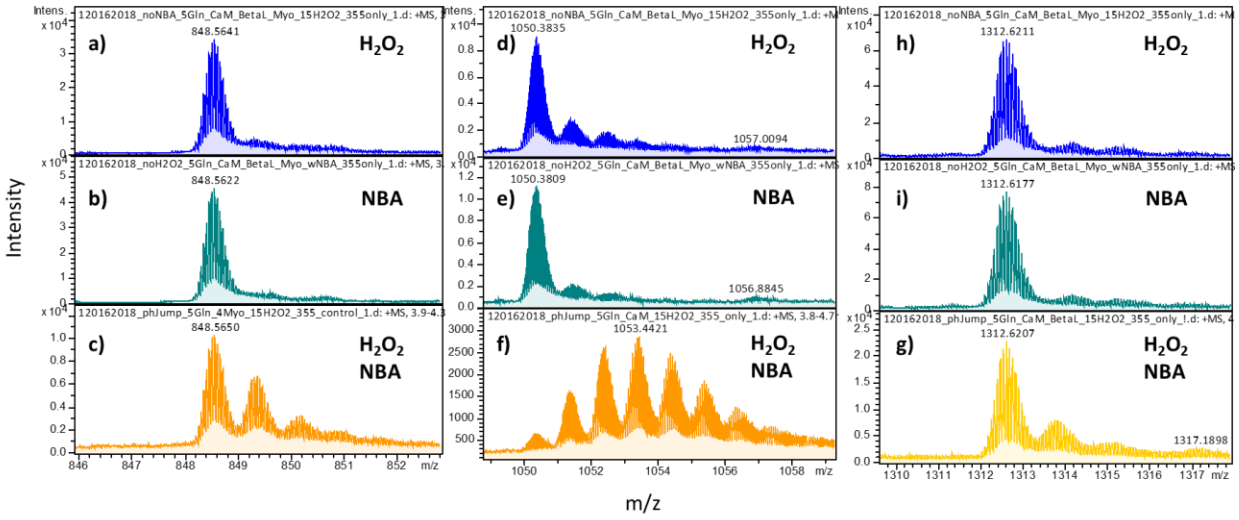


Figure 3.5 Oxidation profiles of hMyo upon only 355 nm laser irradiation under different conditions: a) hMyo (4 μM) with His (0.5 mM) and H_2O_2 (7.5 mM) at pH 6.3 in NaCl; b) hMyo (4 μM) with His (0.5 mM) and NBA (~ 8 mM) at pH 6.3 in NaCl; c) hMyo (4 μM) with His (0.5 mM), H_2O_2 (7.5 mM) and NBA (~ 8 mM) at pH of 6.3 in NaCl. Oxidation profiles of calmodulin are under similar conditions listed from d) to f) and of β-lactoglobulin are listed corresponding from h) to g).

3.5.4 pH Jump with Different Delay Times

To establish whether the oxidized hMyo undergoes further conformational changes, we attempted to footprint the most unfolded protein conformation through a "recycled" experiment. After being labeled with the 355 nm laser, we collected the hMyo solution and let it sit for 3 mins to allow the unfolding to maximize. The protein solution was then recycled for 248 nm irradiation for the upper-limiting oxidation extent (Figure 3.6e). Compared to the oxidation percentage with only 355 nm laser irradiation (Figure 3.6b), the increase in the extent of labeling confirms that a conformational change of hMyo occurs after the initial oxidation, promoting us to probe the conformational transitions during unfolding.

We set the time delays between the two lasers as 0 ms, 1 ms, 5 ms, 20 ms and 40 ms in parallel runs to capture the dominant conformation at different time points. When labeling hMyo simultaneously with two lasers (355 nm - 0 ms - 248 nm), we observed that the second laser irradiation contributed approximately 8% more oxidation (Figure 3.6c) compared to the condition of 355 nm irradiation only (Figure 3.6b). The total oxidation percentage with no time delay is 46%, a value that characterizes the starting conformation during the transition. If we use 40 ms as the time delay (355 nm - 40 ms - 248 ms), the overall oxidation percentage increased to 51% (Figure 3.6d), similar to the value of a completely unfolded conformation (Figure 3.6e). Although the extent of oxidation increased with longer time delays, the absolute difference between the two delay times is only 5%, making it hard to distinguish different conformations

when using a series of smaller delay times. In addition, the reproducibility of this experiment throughout the process was not high. We suggest that the main reason is still the unknown chemistry between NBA and H_2O_2 to yield unspecified reactive radical species. The undefined system leads to uncontrollable oxidations. To overcome this problem, only one of the two chemicals can exist in the pH jump solution.

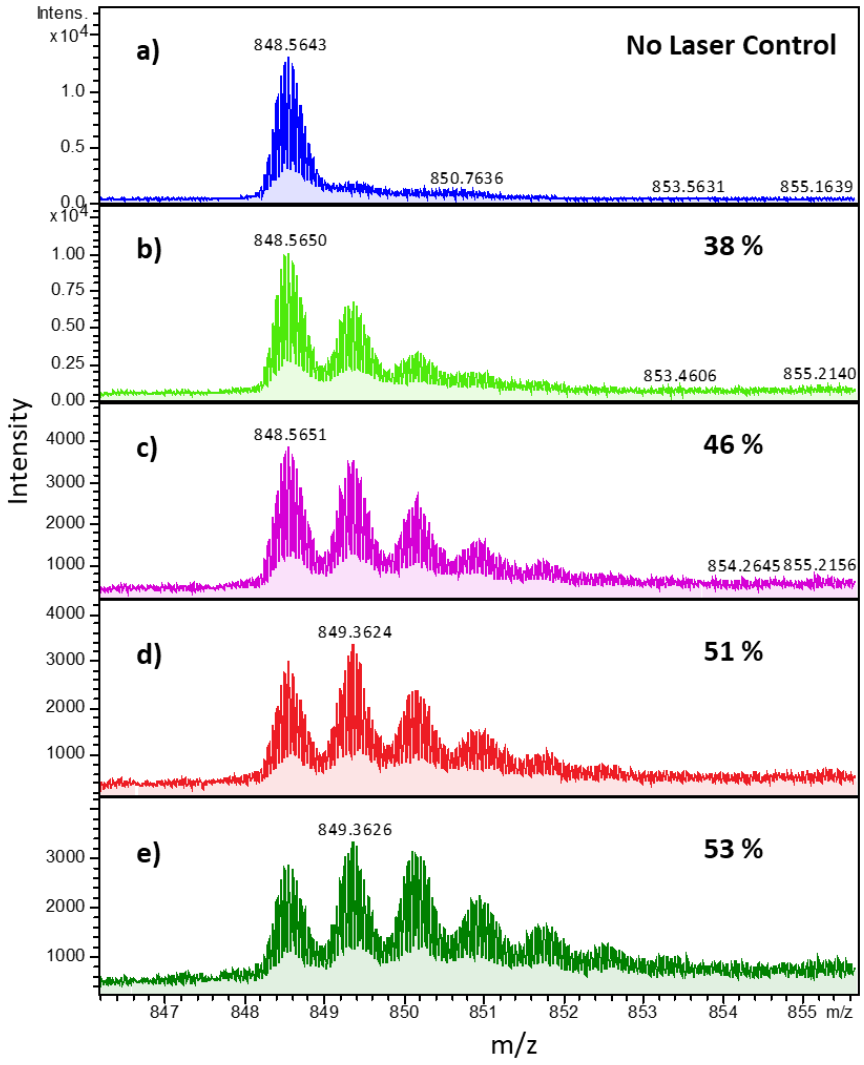


Figure 3.6 Oxidation profiles of hMyo with different time delays between 355 nm and 248 nm laser. Specific conditions are: a) no laser control; b) control with only 355 nm irradiation; c) 0 ms time delay; d) 40 ms time delay; e) "recycled" experiments with 3 mins time delay.

3.5.5 pH Jump with Carbene Radicals

To avoid the simultaneous presence of NBA and H_2O_2, we substituted H_2O_2 with diazirine for carbene footprinting, which is usually initiated with the YAG laser at 355 nm. Given that NBA can also be activated at 248 nm, we can trigger the proton release first with this laser and then fooprint the protein with 355 nm irradiation; therefore, the order of implementing the two lasers needs to be switched (Figure 3.7) compared to the previous design. In addition, we screened several diazirine reagents (Scheme 3.1) for carbene footprinting and succeeded with 3-(3-methyl-3H-diazirin-3-yl) propan-1-ol (MDPO).

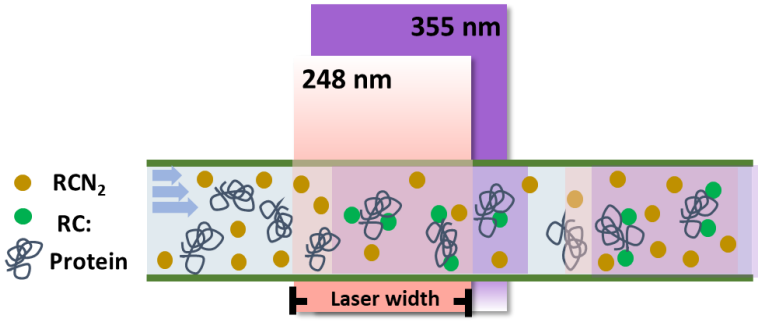
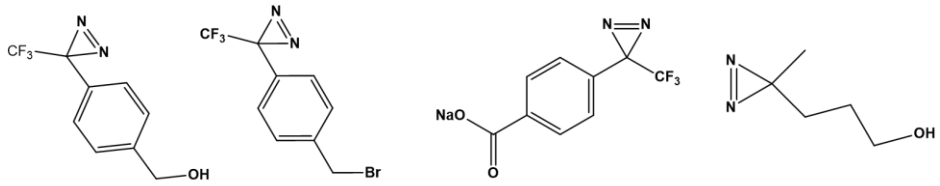


Figure 3.7 Enlarged labeling window using carbene footprinting.



Scheme 3.1 Diazirine reagents for carbene footprinting.

The labeling profiles of hMyo upon 355 nm irradiation of MDPO are shown in Figure 3.8. We first incubated hMyo with MDPO in NaCl solution and observed consecutive + 86.07 Da peaks after the peak for the unmodified protein. The carbene labeling for the closed and open state of hMyo were conducted with initial pHs of 6.8 (Figure 3.8b) and 3.2 (Figure 3.8c), respectively.

More carbene additions were observed under the acidic conditions, consistent with larger SASA in the open conformation. Then, we evaluated the labeling performance of MDPO at 248 nm and observed no + 86 Da peaks (Figure 3.8d), consistent with the experimental design.

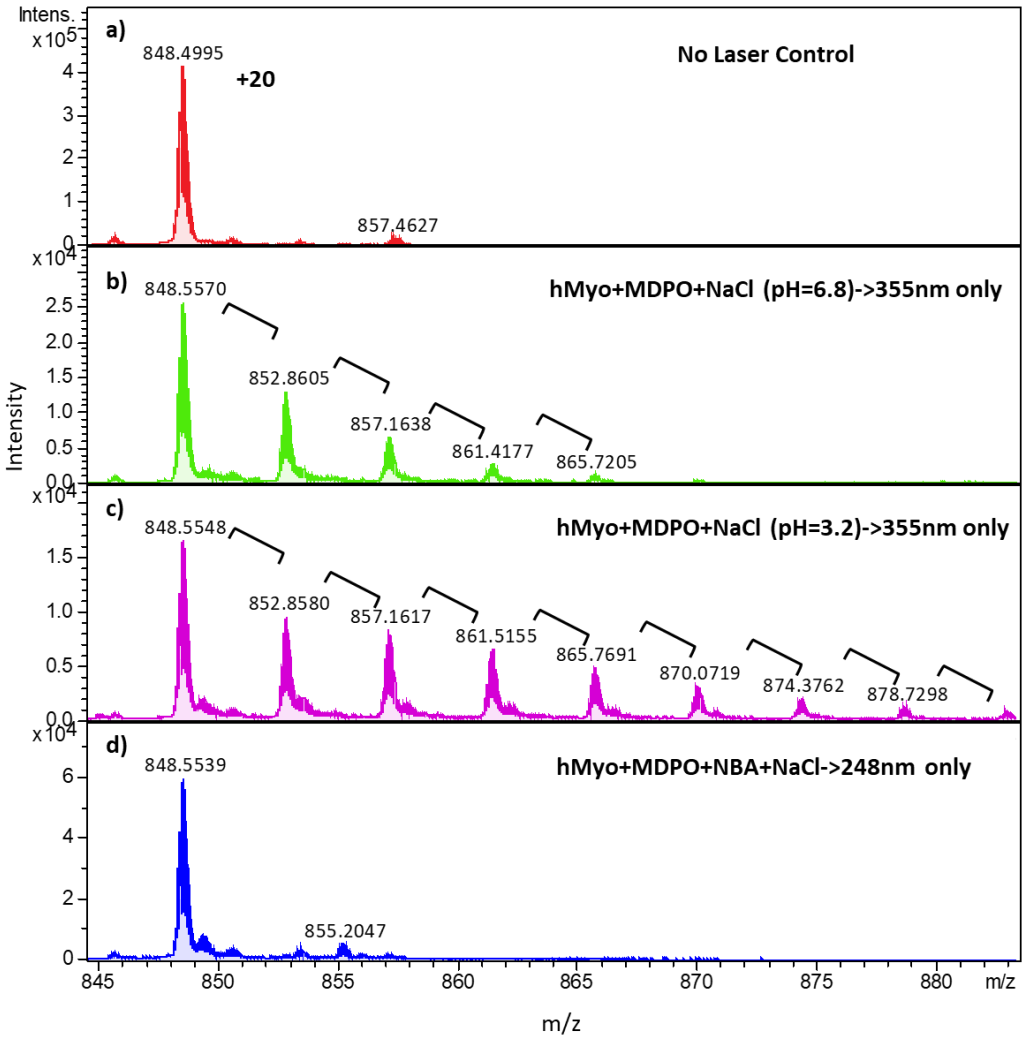


Figure 3.8 Carbene labeling profiles of hMyo under different conditions that are listed in each panel.

In the presence of MDPO and hMyo, the photoacid NBA was also included in the sample solution for YAG laser at 355 nm. The carbene labeling, however, didn't occur, and only a low intensity peak corresponding to oxidation appeared (Figure 3.9a) instead, indicating chemistry between MDPO and NBA. To establish whether it is the NBA or the transformed NBA-acid can

react with MDPO, we performed a similar “recycled” experiment as before. Specifically, we mixed hMyo with only NBA and submitted the solution to 248 nm irradiation, trying to exhaust NBA into the acid form. Then, we added MDPO to the collected solution for carbene footprinting at 355 nm. We observed a similar pattern of showing predominantly oxidation species (Figure 3.9b). In addition to the +16 Da peaks, we noticed two small peaks that represent carbene addition. Therefore, we screened different conditions, trying to promote this carbene chemistry. Even when the concentration of MDPO was tripled, we only observed significant increase in peak intensity for oxidation (Figure 3.9c). Furthermore, by using apo myoglobin (aMyo) as a control, we also excluded the possibility of the heme group being the initiator of this unknown chemistry. The labeling profile still showed the oxidation instead of carbene addition (Figure 3.9d).

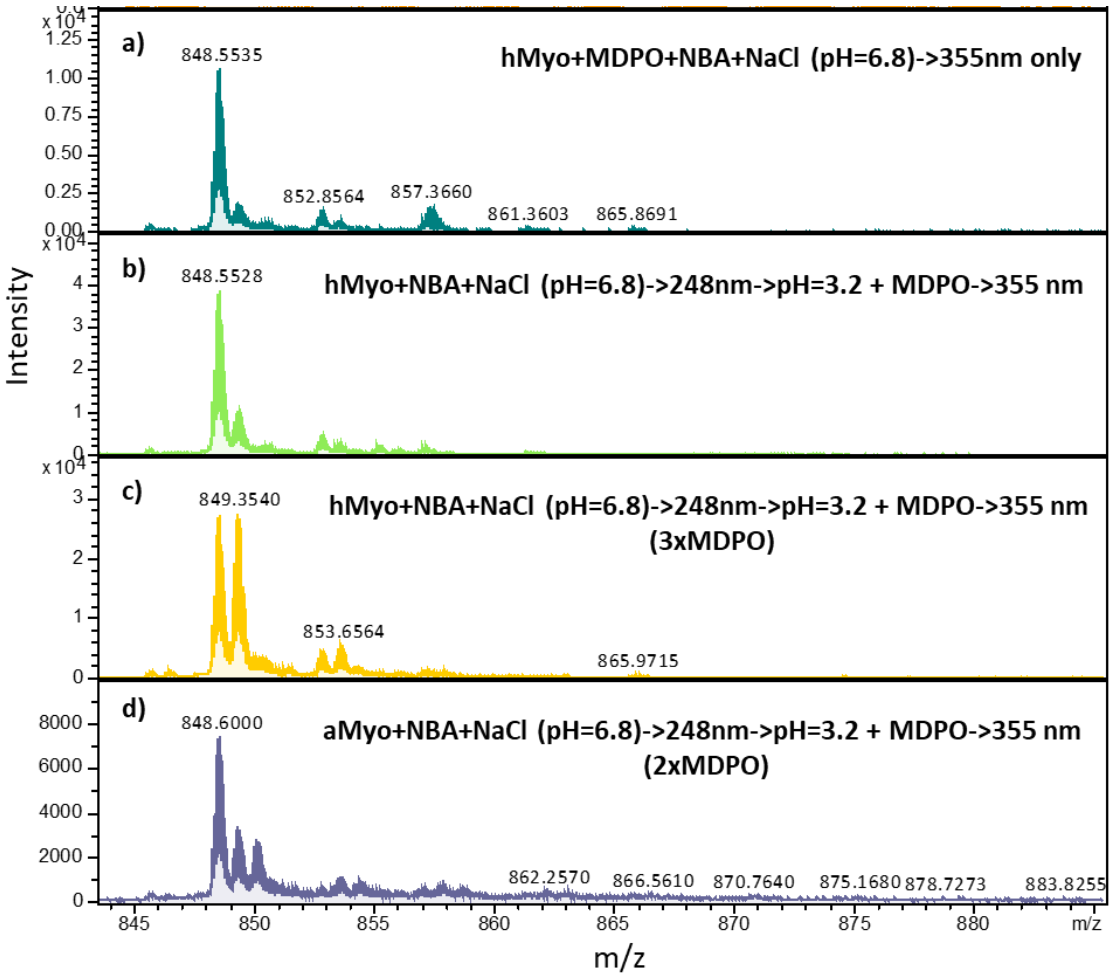


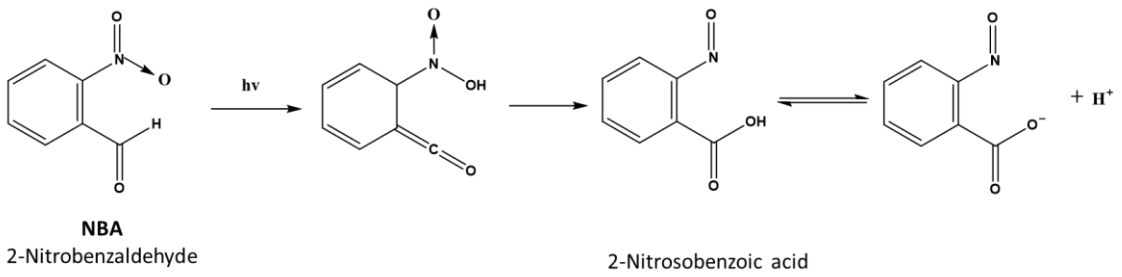
Figure 3.9 Carbene labeling profiles of hMyo and aMyo under different conditions that are listed in each panel.

3.6 Conclusion

The two-laser platform based on $HO^{\bullet}$ footprinting shows preliminary results to probe conformational changes of hMyo. Unknown reactive species that result from chemical reactions between the photoacid NBA and H_2O_2 prohibit control of the system. Unfortunately, replacement of H_2O_2 and $HO^{\bullet}$ footprinting with carbene footprinting was not successful. We also tried to substitute NBA with another widely used photoacid, 1-(2-nitrophenyl)ethyl sulfate (NPE) [13].

However, NPE itself can label protein upon 248 nm or 355 nm irradiation even in the absence of other footprinting reagents.

One of the many reasons of why the chemistry is complex is due to the complication of the proton releasing mechanisms. Other investigators offer several mechanisms, some of which include easily oxidizable intermediates. For example, one proposed mechanism[15] (Scheme 3.2) of NBA rearrangement upon laser activation involves two single nitrogen-oxidation bonds that are susceptible for a strong oxidant like H_2O_2. Therefore, a nitroso-group can be troublesome being part of the photoacid. For further investigations, other types of photoacids such as benzyl ester[16-17] may be a good choice.



NBA
2-Nitrobenzaldehyde

2-Nitrosobenzoic acid

Scheme 3.2 One proposed proton release mechanism of NBA.

For carbene footprinting, it is known that it preferentially reacts with polar motifs such as carbonyl groups. The presence of a carbonyl in a photoacid is usually necessary, given that the pH jump involves transformations into an acidic form. Thus, carbene footprinting may not be a good candidate for this purpose. Other types of radical-based footprinting may be worth to try (e.g., iodine radical and trifluoromethyl radical footprinting).

3.7 Acknowledgments

The author thanks Dr. Rachel Liuqing Shi for her generous help and many valuable discussions, and Don Rempel for his technical support on the construction of the two-laser platform. This work was supported by the NIH 2P41GM103422.

3.8 References

1. Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426*, 884-890.

2. Luheshi, L. M.; Crowther, D. C.; Dobson, C. M. Protein misfolding and disease: from the test tube to the organism. *Curr. Opin. Chem. Biol.* **2008**, *12*, 25-31.

3. Cohen, F. E.; Kelly, J. W. Therapeutic approaches to protein-misfolding diseases. *Nature* **2003**, *426*, 905-909.

4. Kaltashov, I. A.; Eyles, S. J., Mass spectrometry in biophysics: conformation and dynamics of biomolecules. John Wiley & Sons: 2005; Vol. 12.

5. Uzawa, T.; Nishimura, C.; Akiyama, S.; Ishimori, K.; Takahashi, S.; Dyson, H. J.; Wright, P. E., Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *10*, 13859-13864.

6. Yang, W. Y.; Gruebele, M. Folding at the speed limit. *Nature* **2003**, *423*, 193-197.

7. Stocks, B. B.; Konermann, L. Structural Characterization of Short-Lived Protein Unfolding Intermediates by Laser-Induced Oxidative Labeling and Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 20-27.

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*, 2057-63.

9. Chen, J.; Rempel, D. L.; Gross, M. L. Temperature Jump and Fast Photochemical Oxidation Probe Submillisecond Protein Folding. *J. Am. Chem. Soc.* **2010**, *132*, 15502-15504.

10. 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*, 18724-18731.

11. Abbruzzetti, S.; Sottini, S.; Viappiani, C.; Corrie, J. E. Acid-induced unfolding of myoglobin triggered by a laser pH jump method. *Photochem. Photobiol. Sci.* **2006**, *5*, 621-628.

12. Shen, L. L.; Hermans Jr, J. Kinetics of conformation change of sperm-whale myoglobin. I. Folding and unfolding of metmyoglobin following pH jump. *Biochemistry* **1972**, *11*, 1836-1841.

13. Jeong, B.-S.; Dyer, R. B. Proton Transport Mechanism of M2 Proton Channel Studied by Laser-Induced pH Jump. *J. Am. Chem. Soc.* **2017**, *139*, 6621-6628.

14. Sogbein, O. O.; Simmons, D. A.; Konermann, L. Effects of pH on the kinetic reaction. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 312-319.

15. Bouya, H.; Al Rashidi, M.; Roth, E.; Salghi, R.; Chakir, A. Atmospheric degradation of 2-nitrobenzaldehyde: Photolysis and reaction with OH radicals. *Atmos. Environ.* **2017**, *171*, 221-228.

16. Martin, C. J.; Rapenne, G.; Nakashima, T.; Kawai, T. Recent progress in development of photoacid generators. *J. Photochem. Photobiol. C Photochem. Rev.* **2018**, *34*, 41-51.

17. Shirai, M.; Suyama, K.; Okamura, H.; Tsunooka, M. Development of novel photosensitive polymer systems using photoacid and photobase generators. *J. Photopolym. Sci. Tech.* **2002**, *15*, 715-730.

Chapter 4: An Integrated Approach for Determining a Protein-Protein Binding Interface in Solution and an Evaluation of HDX Kinetics for Adjudicating Candidate Docking Models*

* This chapter is based on the following publication: Zhang, M. M.; Beno, B. R.; Huang, R. Y-C.; Adhikari, J.; Deyanova, E. G.; Li, J.; Chen, G.; Gross, M. L. An Integrated Approach for Determining a Protein-Protein Binding Interface in Solution and an Evaluation of HDX Kinetics for Adjudicating Candidate Docking Models *Anal. Chem* **2019,** *91*, 15709-15717.

4.1 Abstract

We describe an integrated approach of using hydrogen−deuterium exchange mass spectrometry (HDXMS), chemical crosslinking mass spectrometry (XL-MS), and molecular docking to characterize the binding interface and to predict the three-dimensional quaternary structure of a protein−protein complex in solution. Interleukin 7 (IL-7) and its α-receptor, IL-7Rα, serving as essential mediators in the immune system, are the model system. HDX kinetics reports widespread protection on IL-7Rα but shows no differential evidence of binding-induced protection or remote conformational change. Crosslinking with reagents that differ in spacer lengths and targeting residues increases the spatial resolution. Using five cross-links as distance restraints for protein−protein docking, we generated a high-confidence model of the IL-7/IL-7Rα complex. Both the predicted binding interface and regions with direct contacts agree well with those in the solid-state structure, as confirmed by previous X-ray crystallography. An additional binding region was revealed to be the C-terminus of helix B of IL-7, highlighting the value of solution-based characterization. To generalize the integrated approach, protein−protein docking was executed with a different number of cross-links. Combining cluster analysis and HDX kinetics adjudication, we found that two intermolecular cross-link-derived restraints are sufficient to generate a high-confidence model with root-mean-square distance (rmsd) value of all alpha carbons below 2.0 Å relative to the crystal structure. The remarkable results of binding-interface determination and quaternary structure prediction highlight the effectiveness and capability of the integrated approach, which will allow more efficient and comprehensive analysis of interprotein interactions with broad applications in the multiple stages of design, implementation, and evaluation for protein therapeutics

4.2 Introduction

Precisely regulated but complex protein-protein interactions are paramount to safeguard homeostasis in the immune system[1]. Protein-binding events initiate vast arrays of signal transduction, communicating critical information to orchestrate multiple health-related immune responses[2]. Comprehensive characterization and impartial understanding of the binding interfaces at high spatial resolution are imperative for both academic and pharmaceutical research. A traditional approach to obtain binding information is X-ray crystallography[3]. The need for fastidious crystallization and long growth times can limit its application. In addition, the delivered solid-state information can raise questions about its relevance to the physiological environment. Nuclear magnetic resonance (NMR), on the other hand, provides a liquid-phase characterization but at the expense of large amounts of isotopically-labelled sample, demanding signal averaging, and complicated data interpretation[4]. Mass spectrometry (MS)-based methods are exceptionally appealing in this context given their low detection limit, fast throughput, and compelling use of native proteins.

Hydrogen deuterium exchange mass spectrometry (HDX-MS)[5,6] is a robust analytical technique for characterizing protein-protein interactions,[7] protein-ligand binding, higher-order structure[8], and conformational dynamics.[9-11] Conformational changes induced by protein-protein binding lead to changes in deuterium uptake kinetics, enabling comparisons between different states (e.g., bound and unbound), thus permitting the determination of binding interfaces and changes in dynamics. Spatial resolution exceeding the peptide-level can be achieved by multi-protease digestion for overlapping peptides, by electron transfer dissociation (ETD)/electron capture dissociation (ECD) fragmentation[12-14], or by combining HDX results with other characterization approaches (e.g., chemical crosslinking and docking, as in this study).

Chemical crosslinking with subsequent MS analysis (XL-MS) has considerable value in the structural biology toolbox.[15-16] It can deliver topological information of protein complexes by bridging the neighboring domains in close proximity, defined by characteristic spacer lengths on cross-linkers. Advanced analysis software[17-19], new crosslinking reagents and protocols[20] have greatly escalated its applications. The obtained information is even more propitious to uncover higher-order quaternary structure when coupled with other methods (e.g., macromolecular docking).

Complementing X-ray crystallography and NMR, protein-protein docking is a computational methodology that also yields complete three-dimensional structural information for protein complexes. Despite rapid and regular increases in computer speed, parallel processing, acceleration of computation through the use of GPUs and the development of improved algorithms and scoring functions, accurate modeling of protein complexes remains challenging, even starting with high-quality structural models (e.g., X-ray crystal structures) of the binding partners[21]. This is due in part to the large number of potential complex structures that must be evaluated in the absence of additional information (sampling problem)[22] and to limitations in the ability to score accurately the models generated (scoring problem)[23-24]. An integrated platform[25-28] that utilizes experiment-derived information (i.e., HDX-MS and XL-MS data) to define restraints that guide the sampling stage of protein-protein docking has the potential to provide models with high accuracy and resolution.

In this study, we chose interleukin 7 (IL-7) as a test case to demonstrate the integrated characterization platform. IL-7 is a representative member of the cytokine family, modulating immune cell physiology through receptor recognition[2, 29]. IL-7 sequentially binds to IL-7Rα, its α-receptor, and γ_c through its extracellular domain to form a ternary-complex that activates the

Janus kinase (JAK) and enables phosphorylation of the signal transducer and activator of transcription (STAT).[30] Downstream signal transduction switches on transcription of the anti-apoptotic genes, promotes cell survival, and causes proliferation of both naïve and memory T cells.[31] Many investigators consider recombinant human IL-7 an ideal treatment agent in cancer immunotherapy. [32] It can promote immune reconstitution[33-34] through peripheral T cell expansion, increase the efficacy of tumor regression,[35] and antagonize the immunosuppressive network.[36] Understanding the critical role of regulating immune functions requires characterization of the involved binding interface, further motivating the investigation of IL-7/IL-7Rα as a model system. The approach, once established, will doubtlessly benefit other protein systems (e.g., other cytokine complexes, antigen/antibody complexes). Deeper understanding will aid biomedical design[37] of protein interface engineering[2] and epitope-based preventive vaccines[38] that are crucial to maintain homeostasis and physiological well-being.

Our goal is to implement and test an analytical approach that is comprised of HDX, chemical crosslinking and protein-protein docking to characterize the binding interface of IL-7/IL-7Rα. With the integrated methods, we generated an accurate three-dimensional model that allows detailed description on the direct inter-protein contacts in the binding region. Furthermore, we investigated how HDX kinetics can adjudicate candidate docking models en route to an ultimate high-confidence complex structure. A previous X-ray crystallographic study[39] of IL-7/IL-7Rα is available for final comparison purposes.

The binding regions assigned by using the approach are in good accord with the published crystal structure and, moreover, indicate another involvement (i.e., C-terminus of helix B in IL-7). Docking results generated with different numbers of cross-links further demonstrate that, with HDX kinetics adjudication, a high-confidence model can be delivered by using only two

intermolecular cross-links. Although HDX-MS has been employed for complementary information in some biological studies[40-44] along with XL-MS, the two methods are usually treated individually. Other studies[27-28] that incorporated XL-based docking have shown that HDX kinetics can be utilized to examine the constructed 3D model; however, none has assessed the process in a systematic way. The integration and evaluation of the three approaches convincingly shows the combination can yield higher-order structural information and provide basic and practical insights for further applications.

4.3 Experimental

4.3.1 Material

BS^3-h12/d12 (bis(sulfosuccinimidyl) suberate) was purchased from Creative Molecules. BS^2G-d4 (*bis*(sulfosuccinimidyl) 2,2,4,4-glutarate-d4), BS^2G (*bis*(sulfosuccinimidyl) 2,2,4,4-glutarate), EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride), Zeba column and colloidal blue staining kit were purchased from Thermo Fisher Scientific (Waltham, MA). Sulfo-NHS (*N*-hydroxysulfosuccinimide sodium salt), MES buffer, dithiothreitol and iodoacetamide were purchased from Millipore Sigma (Saint Louis, MO). Recombinant human IL-7Rα protein was purchased from Sino Biological (Wayne, PA). Recombinant human IL-7 protein was purchased from R&D System (Minneapolis, MN). RapiGest SF Surfactant was from Waters (Milford, MA). Sequencing grade modified trypsin and chymotrypsin were from Promega Co. (Madison, WI). C18 NuTip was purchased from Glygen (Columbia, MD). PNGase(glycerol-free) F was purchased from BioLabs (Hitchin, UK). 4-20% bis-tris precast gels (ExpressPlus™ PAGE) were purchased from GenScript®.

4.3.2 Hydrogen-Deuterium Exchange

The HDX experiments were performed with three states: IL-7 unbound state, IL-7Rα unbound state, and IL-7/IL-7Rα bound state. For the bound state, two proteins were incubated at a 1:1 ratio in the equilibrium buffer (10 mM PBS buffer, pH = 7.4) for at least 45 min at 25 °C. HDX was initiated by diluting 5 µL of the protein solution (20 µM) 5-fold with labeling buffer (10 mM PBS in 90% D_2O, pD = 7), which was substituted into the H_2O buffer for control experiments. Quenching buffer (25 µL, 4 M GdnHCl, 700 mM TCEP, pH = 2.5) was added at different exchange times (i.e., 10, 30, 60, 900, 7200 s and incubated for 3 min). The experiments were in triplicate. The protein mixture was injected immediately into a custom-built HDX platform for digestion by passing through an immobilized pepsin column (2 mm × 20 mm) with 200 µL/min flow rate. A ZORBAX Eclipse XDB C8 column (2.1 mm × 15 mm, Agilent Technologies, Santa Clara, CA) was used for desalting with 0.1% trifluoracetic acid for 3 min. After switching the platform into the analysis mode, the peptide mixture was separated on a Xselect CSH C18 column (2.1 mm × 50 mm, Waters Corporation, Milford, MA) with a 9.5-min-linear gradient (4–40% acetonitrile with 0.1% formic acid) at 200 µL/min flow rate. Peptides were then submitted to a LTQ-FT mass spectrometer (Thermo Fisher, Waltham, MA) for data acquisition at a mass resolving power of 100,000 at *m/z* 400. An electrospray ionization source was operated with the following parameters: spray voltage: 5kV; capillary temperature: 250 C; capillary voltage: 38 V; tube lens: 185 V. The HDX platform, including columns and valves, were submerged in an ice slush bath to minimize back exchange.

After data acquisition, HDX kinetics was analyzed by HDExaminer® (Sierra Analytics, Inc). Per residue deuterium uptake difference for each time point was calculated and exported from the software based on available overlapping peptides using “heavy” smoothing fuction. The

cumulative uptake differences and the associated propagation error was calculated manually. Three times propagation error was chosen to give 99.7% certainty.

4.3.3 BS^3 and BS^2G Crosslinking

Prior to any experiments, IL-7 and IL-7Rα were incubated as 1:1 ratio for 1 h in 10 mM PBS buffer (2.7 mM KCl, 137 mM NaCl, pH = 7.4). The BS^3-h12/d12 crosslinker (1 mg) was dissolved in PBS buffer to make a 25 mM stock solution. The IL-7/ IL-7Rα complex was then mixed with the fresh BS^3-*h12/d12* solution at 1:25, 1:50 and 1:100 ratio, where the concentration of IL-7/ IL-7Rα was 10 µM. The reaction mixture was kept at 25 °C for 45 min. Tris-HCl solution (1 M, pH = 7.4) was added at a final concentration of 50 mM and incubated for 15 mins to stop the crosslinking chemistry. For the BS^2G crosslinking, the experimental procedure was identical except that BS^2G-*d4* and BS^2G-*h4* stock solutions were prepared separately at 25 mM and manually mixed later at 1:1 ratio prior to the XL reactions. Crosslinking experiments were in triplicates for each condition.

4.3.4 EDC Crosslinking

IL-7 and IL-7Rα were incubated in 35 mM MES buffer (15 mM KCl, pH = 6.5) with 1:1 ratio for 1 h. A freshly made EDC/sulfo-NHS stock solution (400 mM EDC/sulfo-NHS in MES buffer, pH = 6.5) was added to the IL-7/IL-7Rα solution for 1 h at 25 °C to allow the XL reactions; the protein concentration was 17 µM and EDC/sulfo-NHS was 5-20 mM in different trials. To quench the reaction, *Tris*-HCl (2M, pH = 7.4) was added to a final concentration of 200 mM, followed by 15 mins incubation. Crosslinking experiments were in triplicates for each condition.

4.3.5 Enzymatic in-solution Digestion

The protein mixture was desalted with a Zeba column and denatured by adding 0.75% (w/v) RapiGest stock solution in Tris-HCl (100 mM, pH = 7.4) at 80 °C for 30 min, with a final concentration of 0.25% (w/v) RapiGest. After cooling to room temperature, reduction and alkylation of disulfide bonds was achieved by incubating sequentially with 7.5 mM DTT at 55 °C for 30 min and 14 mM IAM at dark for 30 min. The protein mixture was then digested and deglycosylated by adding trypsin and PNGase F. After 1 h incubation at 37 °C, chymotrypsin was added for overnight digestion to achieve better coverage. Digestion was stopped under 3% formic acid and incubated at 37 °C for 30 min. The peptide mixture was treated with C18 NuTip for desalting according to the manufacturer's protocol. The eluents were vacuum dried and resuspended in 0.1% formic acid in distilled water, ready for LC-MS/MS analysis.

4.3.6 LC-MS/MS Analysis

A solution of 5 µM peptide digest was loaded onto a C18 reversed-phase desalting column (Acclaim PepMap C18, 100 µm x 2 cm, 5 µm, 100 Å; Thermo Fisher Scientific) at 4 µL/min for 10 min. A custom-packed analytical column with C18 reversed-phase material (Magic, 100 µm x 180 mm, 5 µm, 120 Å; Michrom Bioresources, Inc., Auburn, CA) in silica tubing was used for sample separation on Ultimate 3000 Rapid Separation system (Dionex, Co.). The eluent consisted of solvent A (water with 0.1% formic acid by volume) and solvent B (80% acetonitrile with 0.1% formic acid by volume). The flowrate was controlled as 400 nL/min with a 120 min gradient: 2.5% B to 12% B in 30 min, increased to 60% B for 65 min, ramped to 85% B in 5 min, kept at 85% B for 5 min, returned to 2.5% B in 5 min, and equilibrated at 2.5% B for 10 min. A Nanospray Flex source was mounted on a Thermo Q Exactive Plus orbitrap mass spectrometer for downstream detection. The associated parameters were: spray voltage, 2.5 kV; capillary temperature, 250 °C; full MS maximum injection time, 200 ms; tandem MS maximum

injection time, 100 ms; and charge exclusion, 1, 6, 7, 8. In the BS^3-*h12/d12* and BS^2G-*h4/d4* crosslinked samples, acquisition was performed with user-defined MS-Tag in the data-independent mode. In the EDC crosslinked samples, acquisition was carried out under data-dependent mode, selecting the 20 most-abundant ions for higher energy dissociation (HCD).

4.3.7 Identification of Crosslinked Products

The raw LC-MS/MS files were analyzed by pLink[19, 45] (ver. 2.3.1, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China) and Protein Prospector (ver. 5.22.1, UCSF Mass Spectrometry Facility). IL-7 and IL-7Rα sequences were added manually to the search database. XL information, including monoisotopic linker mass of light and heavy form of BS^3 (156.079 Da and 168.154 Da) and BS^2G (118.057 Da and 114.032 Da), linked sites and composition were all required in pConfig. Trypsin/chymotrypsin digestion was manually defined and added to the library. Search parameters in pLink were: enzyme: trypsin/chymotrypsin; missed cleavage: 3; precursor tolerance: 10 ppm; fragment tolerance: 30 ppm; variable modification: oxidation of M, deamidation of N, Q and N-terminus; fixed modification: carbamidomethyl of C; minimal peptide length: 6 aa; maximal peptide length: 60 aa; minimal peptide mass: 600 Da; maximal peptide mass: 6000 Da. The crosslinked peptides were examined in pLabel to give corresponding summary reports. The identified XL peptides were equal or smaller than a 5% false discovery rate at spectral level with a 10 ppm MS1 filter tolerance. Isotopic doublets were manually confirmed in the raw file, and product-ion (MS/MS) spectra were further validated and compared with the calculated masses of product ions.

4.3.8 Protein-Protein Docking

IL-7/IL-7Rα protein-protein docking analyses were performed with the Rosetta (v. 3.8[46-48]) docking_protocol (RosettaDock) code[49-50] on a cluster of multi-processor Linux servers. For the

docking analyses, the A chain from the X-ray crystal structure of unbound IL-7R extracellular domain (PDB ID: 3UP1) was used as the protein model for IL-7R[51]. Unfortunately, no crystal structure of unbound IL-7 was available either in-house or in the public domain. Thus, the A chain from the IL-7/IL-7R complex crystal structure (PDB ID: 3DI2) provided the model for IL-7[39]. The protein models were prepared for docking in a two-stage process starting with the Rosetta score_jd2 program, which added residue atoms that were not present in the PDB files, added hydrogen atoms, and removed non-protein residues. The second preparation step utilized the Rosetta relax program to perform energy minimization of the structures output by score_jd2 within the context of the Rosetta energy model. With the relax application, 10 models were generated for each structure by using the "constrain_relax_to_start_coords" and "coord_constrain_sidechains" options to limit structural changes, and the model of each protein with the lowest Rosetta total energy score was selected.

The IL-7 and IL-7R models were imported into Maestro (Schrödinger Release 2019-1**:** Maestro, Schrödinger, LLC, New York, NY, 2019.) and manually displaced relative to each other to arbitrary extents and then merged into a single two-chain model that was written to a PDB format file as input for the subsequent protein-protein docking analyses. Docking exercises utilizing all possible unique combinations of 1-5 distance constraints based on the IL-7/IL-7R cross-links identified for residue pairs (IL-7:IL-7R) were performed using the RosettaDock program. The set of five experimental-identified intermolecular cross-links from which the combinations of distance constraints were derived includes only those for which atomic coordinates for Cα atoms for both residues are present in the X-ray structures of IL7 and IL7R (3DI2 and 3UP1). All distance constraints utilized flat-bottomed harmonic potentials which

penalized models in which the Cα-Cα distance(s) for crosslinked residues fell outside of the ranges of 6 – 16 Å[52] and 9 – 30 Å[53] for the EDC and BS^3 crosslinks, respectively.

For each possible combination of 1 – 5 crosslink-derived constraints, 20 docking runs were performed, each yielding up to 2,500 models for a maximum of 50,000 models per constraint combination. For comparison purposes, 20 docking runs, each yielding 5,000 models (100,000 total) were carried out without constraints. For each constraint combination as well as for the unconstrained docking exercise, the top-scoring single model based on total score were identified. In addition, the top-scoring set of 100 models ranked by total score metric was clustered with the Rosetta cluster program using a 2.0 Å cluster radius. Analysis of clusters was performed with in-house shell and Python scripts including a Python/OEChem script that determined the Cα atom-based r.m.s.d. for each clustered model relative to the X-ray crystal structure of the IL-7/IL-7R reference complex (PDB ID: 3DI2). Representative models were superimposed with the X-ray crystal structure of the IL-7/IL-7R complex for difference analysis, and high confidence models were defined as those for which the mean Cα atom r.m.s.d. was less than 2.0 Å.

4.4 Results and Discussion

4.4.1 Hydrogen-Deuterium Exchange

We conducted HDX on the unbound IL-7 and the bound-state in presence of IL-7Rα and collected kinetic plots with 94% sequence coverage and 48 unique peptides (Figure 4.1A for representative HDX kinetic plots and Figure 4.2 for the rest). Regions 17-24 (a), 58-66 (c) and 81-91 (e) of IL-7 show less deuterium uptake across the incubation time up to 2 h, indicating their involvement in the binding interaction. On the other hand, region 35-42 (b) takes up more deuterium, showing exposure when bound to IL-7Rα. Another region, represented by peptide 71-

76 (d), is an example not affected by binding, serving as a negative control. Statistical analysis of the cumulative deuterium uptake differences further supports these observations sometimes up to near residue-level (Figure 4.1B). We consider the differences as significant when they are greater than three times the propagated error, giving 99.7% confidence. We mapped the cumulative HDX differences on the IL-7 structure (Figure 4.1C) to show that the most protected regions lie on helix A (Region a) and helix C (Region e). The C-terminus of helix B (Region c) also shows moderate protection, whereas the loop region (Region b) between helix A and B becomes more flexible upon binding.

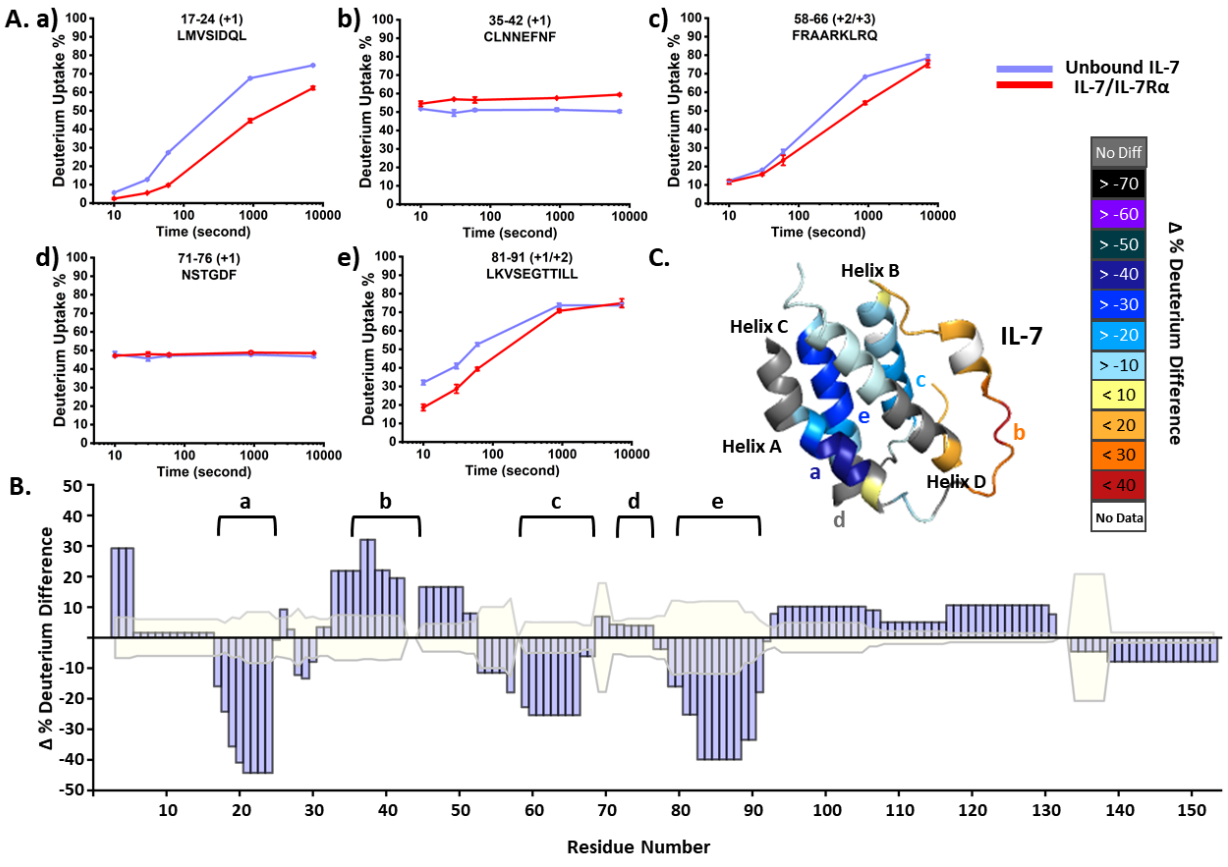
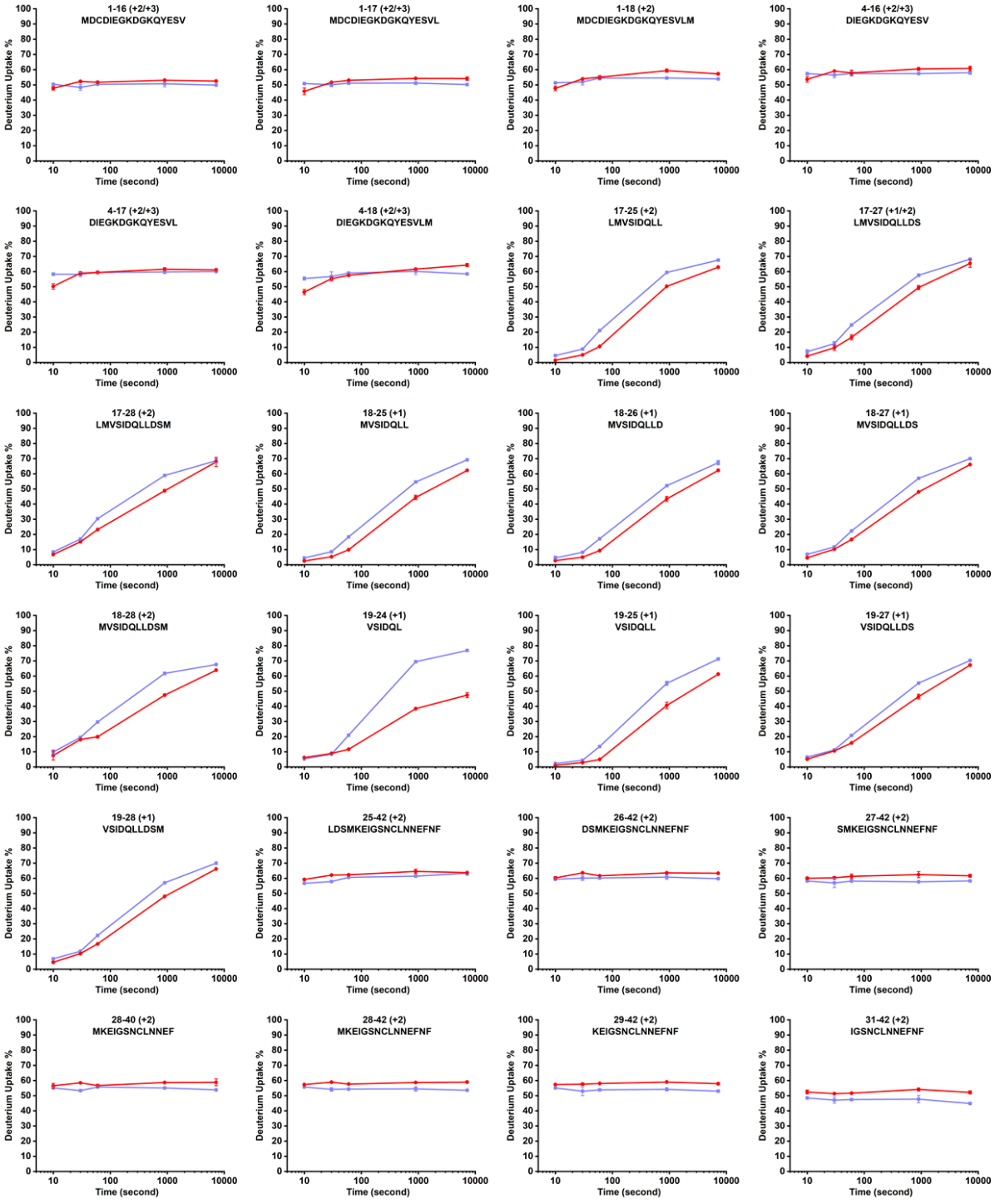


Figure 4.1 Analyzed HDX kinetics of IL-7. (A) Representative HDX kinetics of unbound IL-7 (purple) and of bound with IL-7Rα (red). (B) Statistical analysis of cumulative HDX difference for each residue. Residues are considered being affected upon binding with IL-7Rα when the difference is greater than three times the propagated error (shaded in faint yellow) of all time

points. (C) The cumulative HDX difference of each residue mapped onto the crystal structure of IL-7 (PDB:3DI2).



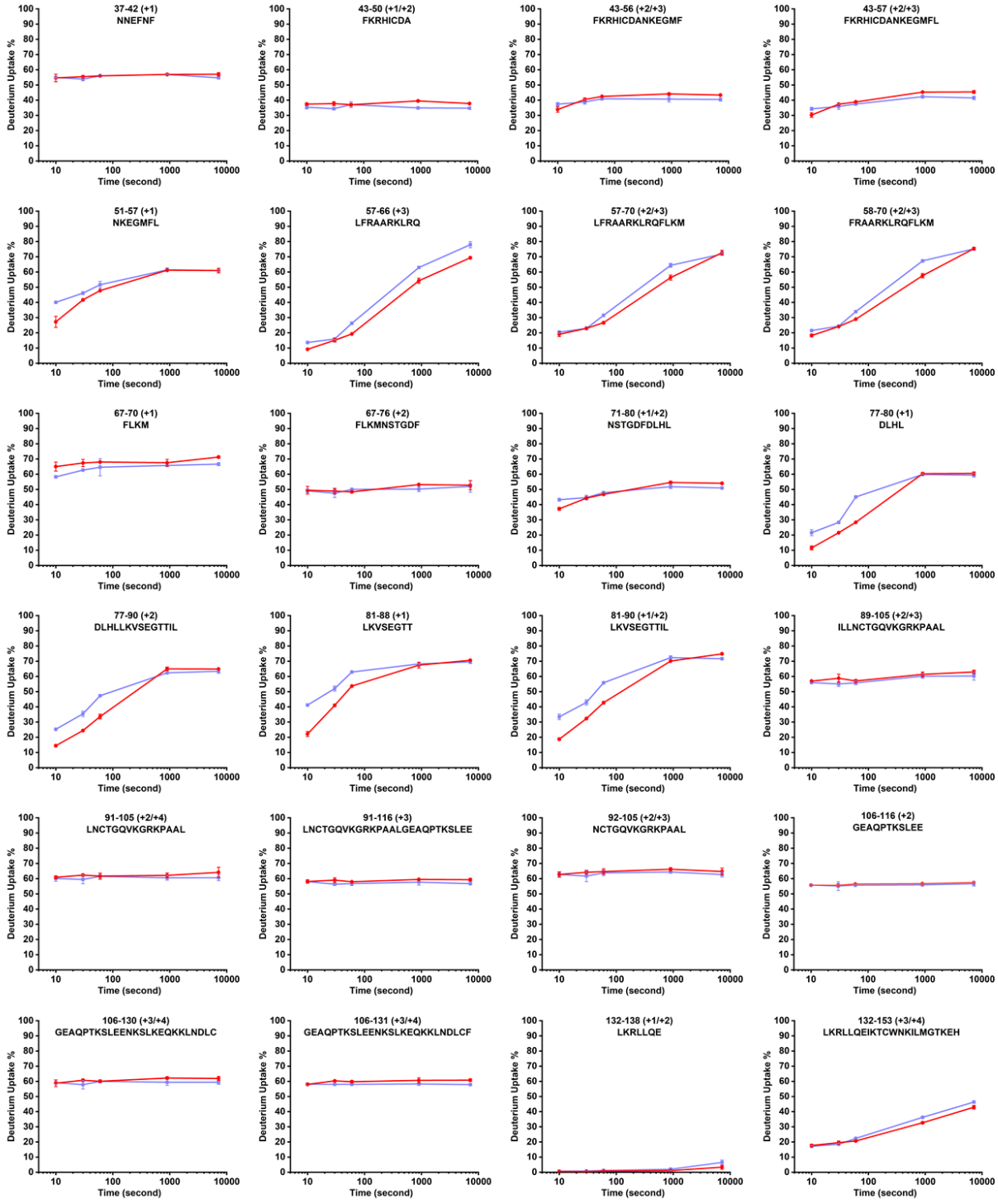


Figure 4.2 HDX kinetic plots of other peptides in unbound IL-7 (purple) and in IL-7/IL-7Rα (red).

To map the binding interface of IL-7Rα, we compared HDX kinetics of its bound and unbound states (Figure 4.3A and Figure 4.4) and obtained 80% sequence coverage and 45 unique peptides. The N-terminal peptide (Region a) shows identical uptake kinetics between the two states. All other peptides, however, present various extents of protection in the HDX profiles, as is clearly demonstrated in the statistical analysis of cumulative deuterium uptake difference (Figure 4.3B). The most protected residues, colored black in the IL-7Rα structure (Figure 4.3C, region c), comprise the elbow region on domain 1 (D1), indicating high probability to be the binding interface. Other regions with pronounced protection are on the other elbow region (Region e) and on one beta-strand (Region d) of domain 2 (D2). The widespread protection indicates that IL-7Rα exhibits a more structured and compact conformation upon binding; however, there is no differential evidence for being the binding interface or undergoing remote conformational changes with the stand-alone HDX. To increase the resolution, we applied crosslinking coupled with MS to obtain complementary information.

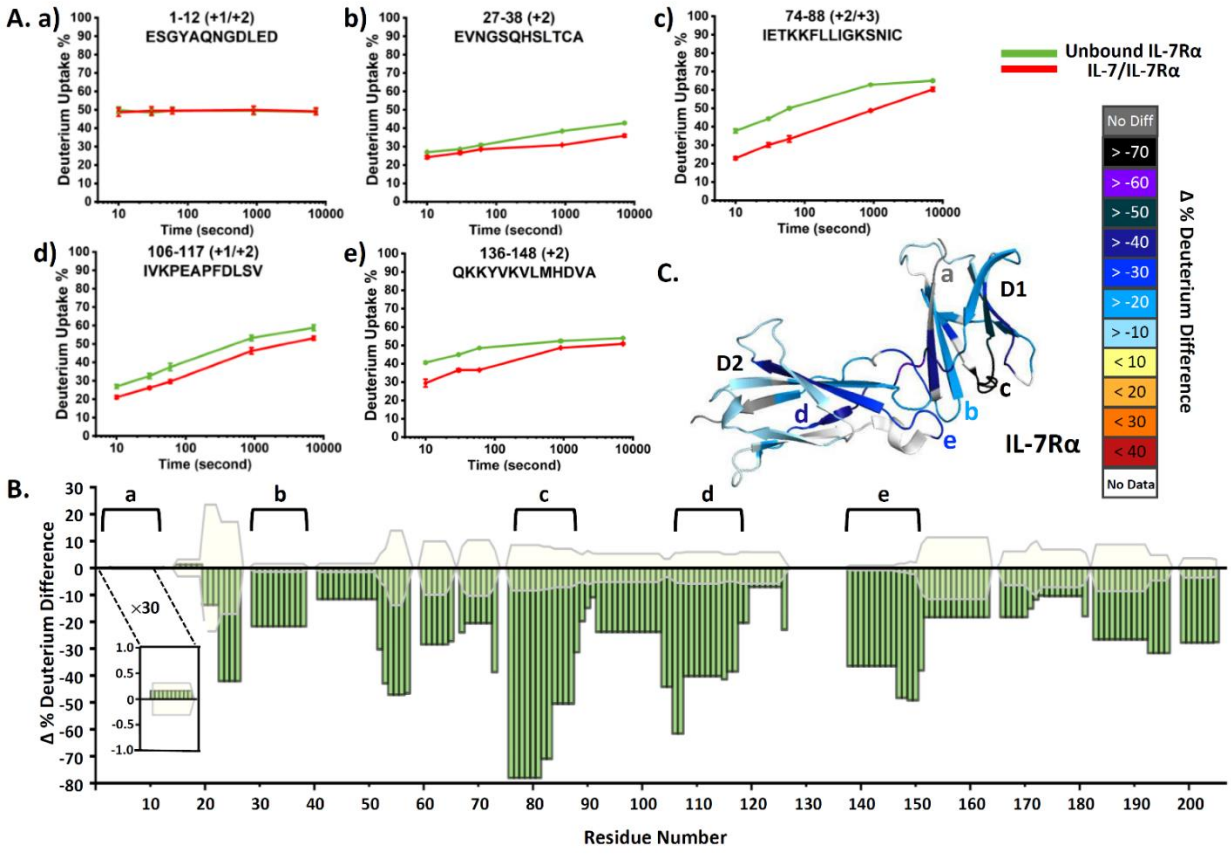
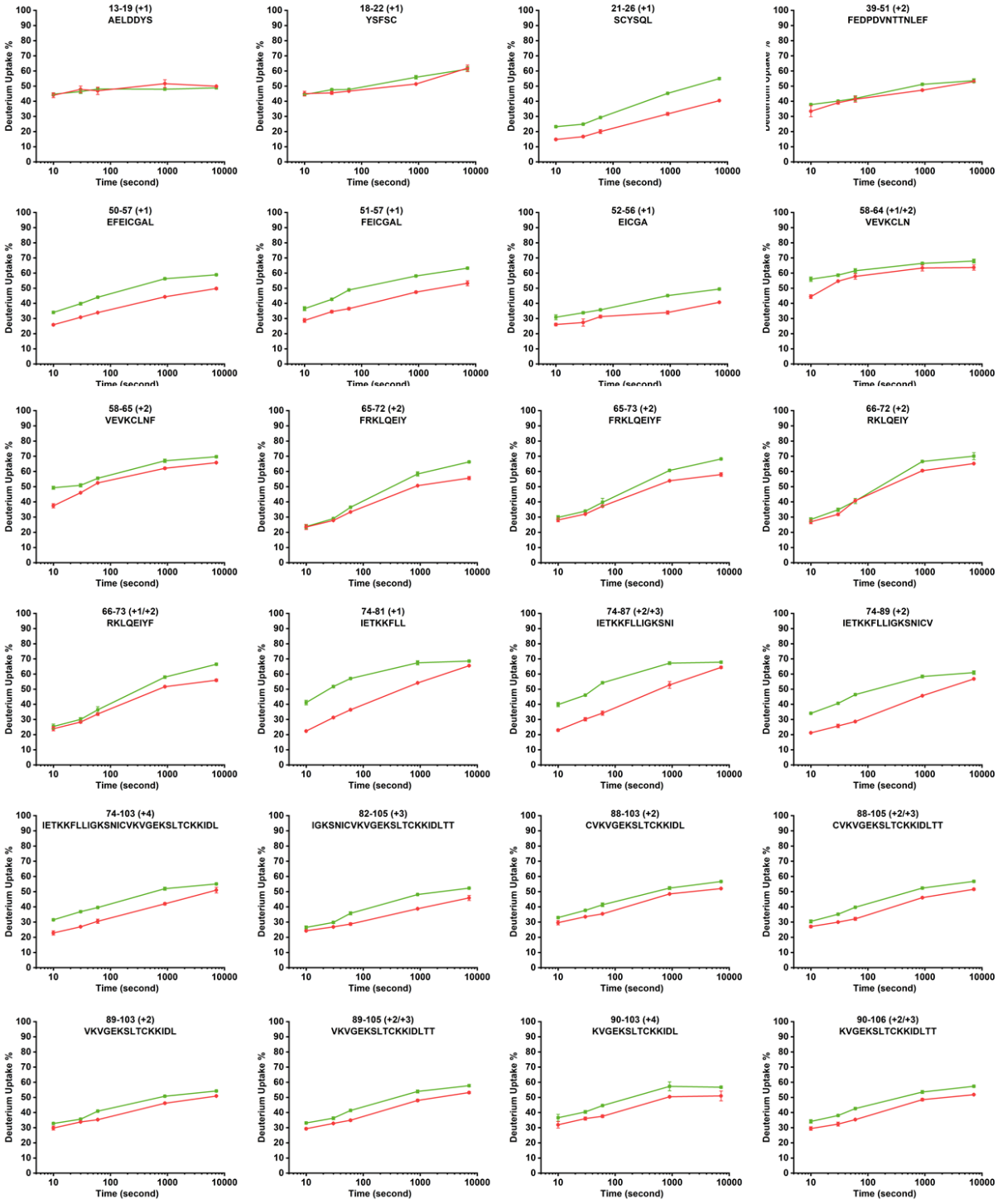


Figure 4.3 Analyzed HDX kinetics of IL-7Rα. (A) Representative HDX kinetics for unbound IL-7Rα (green) and IL-7/IL-7Rα (red). (B) Statistical analysis of cumulative HDX difference upon each residue. Residues are considered affected upon binding with IL-7 when the difference is greater than three times of the propagated error (shaded in yellow) of all time points. (C) The cumulative HDX difference of each residue mapped onto the crystal structure of IL-7Rαs (PDB: 3DI2).


13-19 (+1)
AELDDYS
18-22 (+1)
YSFSC
21-26 (+1)
SCYSQL
39-51 (+2)
FEDPDVNTTNLEF
50-57 (+1)
EFEICGAL
51-57 (+1)
FEICGAL
52-56 (+1)
EICGA
58-64 (+1/+2)
VEVKCLN
58-65 (+2)
VEVKCLNF
65-72 (+2)
FRKLQEIY
65-73 (+2)
FRKLQEIYF
66-72 (+2)
RKLQEIY
66-73 (+1/+2)
RKLQEIYF
74-81 (+1)
IETKKFLL
74-87 (+2/+3)
IETKKFLLIGKSNI
74-89 (+2)
IETKKFLLIGKSNICV
74-103 (+4)
IETKKFLLIGKSNICVKVGEKSLTCKKIDL
82-105 (+3)
IGKSNICVKVGEKSLTCKKIDLTT
88-103 (+2)
CVKVGEKSLTCKKIDL
88-105 (+2/+3)
CVKVGEKSLTCKKIDLTT
89-103 (+2)
VKVGEKSLTCKKIDL
89-105 (+2/+3)
VKVGEKSLTCKKIDLTT
90-103 (+4)
KVGEKSLTCKKIDL
90-106 (+2/+3)
KVGEKSLTCKKIDLTT
Deuterium Uptake %
Time (second)

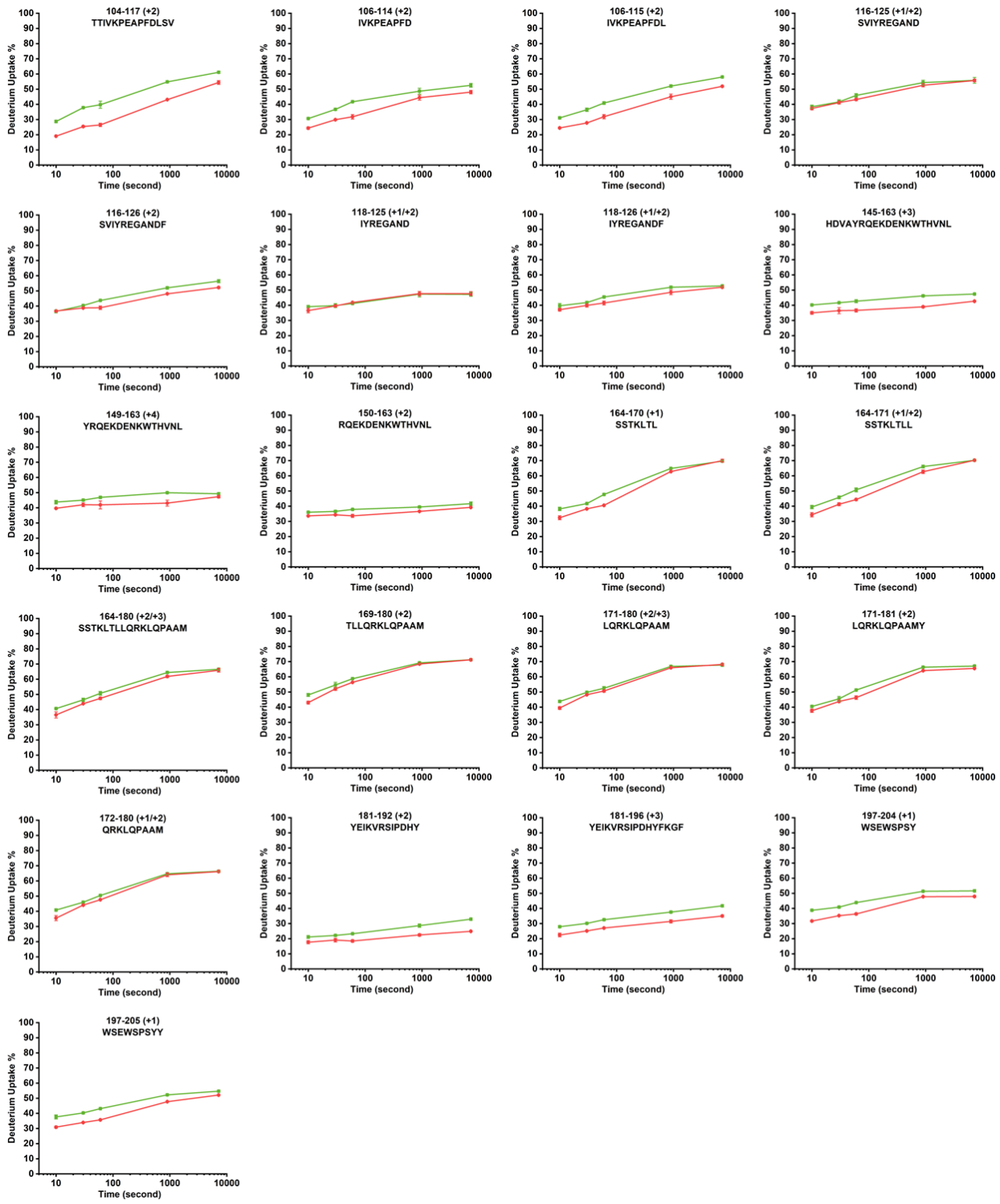


Figure 4.4. HDX kinetic plots of other peptides in unbound IL-7Rα (green) and in IL-7/IL-7Rα (red).

4.4.2 Crosslinking of IL-7/IL-7Rα Complex

To address the interacting regions in the IL-7/ IL-7Rα complex, we chose isotope-encoded BS^3-*h12/d12* (11 Å), BS^2G-h4/d4 (8 Å) and EDC/NHS ("zero-length crosslinker") as the crosslinking reagents. Their spacer lengths cover a range of distance restraints, and their combined use allows interrogation of not only lysine (K) and the N-terminal -NH_2 groups but also glutamic (D) and aspartic acids (E), affording broad coverage. We incubated IL-7Rα and IL-7 at a 1:1 molar ratio to allow complex formation and then mixed them with the cross-linkers at various excess concentrations in separate experiments. We monitored the yields with gel electrophoresis (Figure 4.5). The cross-linked complexes present clearly at approximately 64 kDa, and the yields are estimated according to the intensities of the bands. We found that 100-fold excess BS^3 and BS^2G (1 mM) gives more abundant cross-linked IL-7/ IL-7Rα than does an excess of 25 (0.25 mM) and 50-fold (0.5 mM). For EDC/NHS, use of 10 mM and 20 mM of reagent results in product bands of nearly equal intensity. We submitted the crosslinked IL-7/ IL-7Rα from individual trials to in-solution digestion followed by LC-MS/MS analysis, in which the "light" and "heavy" isotope encoding aids the identification of cross-links. We manually validated the isotope-encoded doublets and the product ions formed in MS/MS from raw files, establishing that the experimental masses agree with their theoretical values.

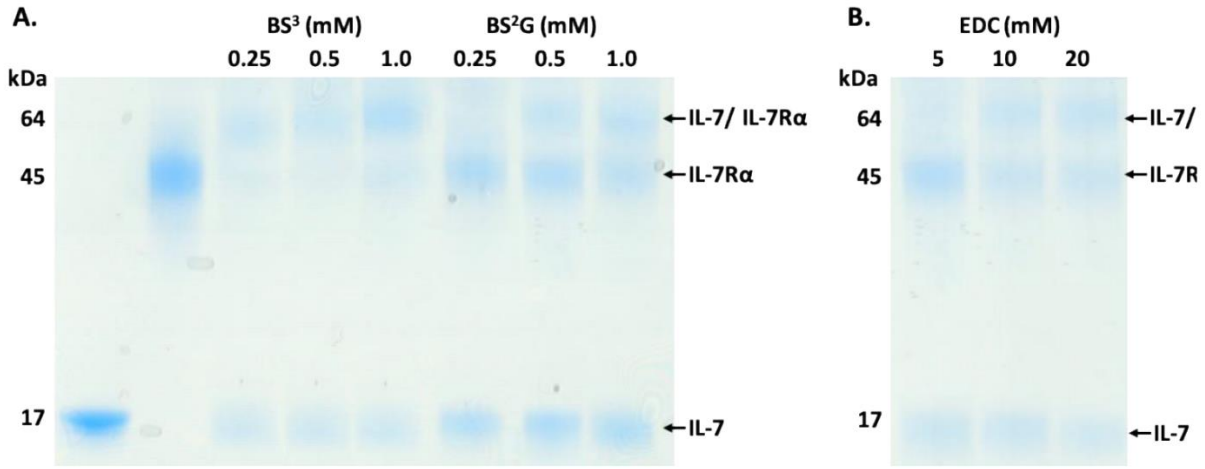


Figure 4.5 Gel electrophoresis of IL-7/IL-7Rα. (A) BS^3 and BS^2G crosslinking of IL-7 (17 kDa) and IL-7Rα (45-50 kDa) are performed at 0.25 mM, 0.5 mM and 1 mM. The crosslinked IL-7/ IL-7Rα band (64 kDa) indicates concentration-dependent production. The first two electrophoresis lanes are shown as control runs. (B) EDC crosslinking is performed at 5 mM, 10 mM and 20 mM. Different species are identified on the right.

In total, we identified 16 intermolecular cross-links. Six of them contain one cross-linked site on a long and flexible loop between helix C and D of IL-7, suggesting the flexibility of the loop region but providing no evidence of the binding interface. This loop can be ruled out as a binding interface, even in the absence of an X-ray structure, because it does not show protection in HDX upon binding. Of the remaining 10 inter cross-links (Table 4.1, corresponding cross-links are shown in Figure 4.6-4.15), those on IL-7 are K11 on helix A (cross-link 1 and 2), D75 on helix C (cross-link 3), K69 on the C-terminus of helix B (cross-link 4), further supporting the previously assigned protected regions in the HDX kinetics study. In addition, we observed several cross-links on a series of residues on the N-terminus (crosslink 6 to 10) and one on helix D (cross-link 5), all of which deliver topological information about the binding complexes.

Table 4.1 Identified inter cross-links with BS^3, BS^2G and EDC crosslinking.

	Il-7		Cross-Linker	Reported Cross-Linkable Cα- Cα Distance, Å	Measured Cα- Cα Distance in the Generated Model, Å
1	K11	K84	BS^3	9-30[53]	14.3
2	K11	K141	BS^3	9-30[53]	13.9
3	D75	K77	EDC	6-16[52]	9.8
4	K69	K78	BS^3	9-30[53]	17.2
5	K152	K141	BS^3	9-30[53]	24.2
6	K8	K84	BS^2G / BS^3	6-26/9-30[53]	12.3[a]
7	N-term	K84	BS^2G / BS^3	6-26/9-30[53]	-
8	D2	K84	EDC	6-16[52]	-
9	D4	K84	EDC	6-16[52]	-
10	E6	K84	EDC	6-16[52]	-

[a] Cross-linked distance is measured at the Cα of D9 instead of K8 in crystal structure (PDB: 3DI2) because the N-terminal region from M1-K8 is missing in the X-ray structure.

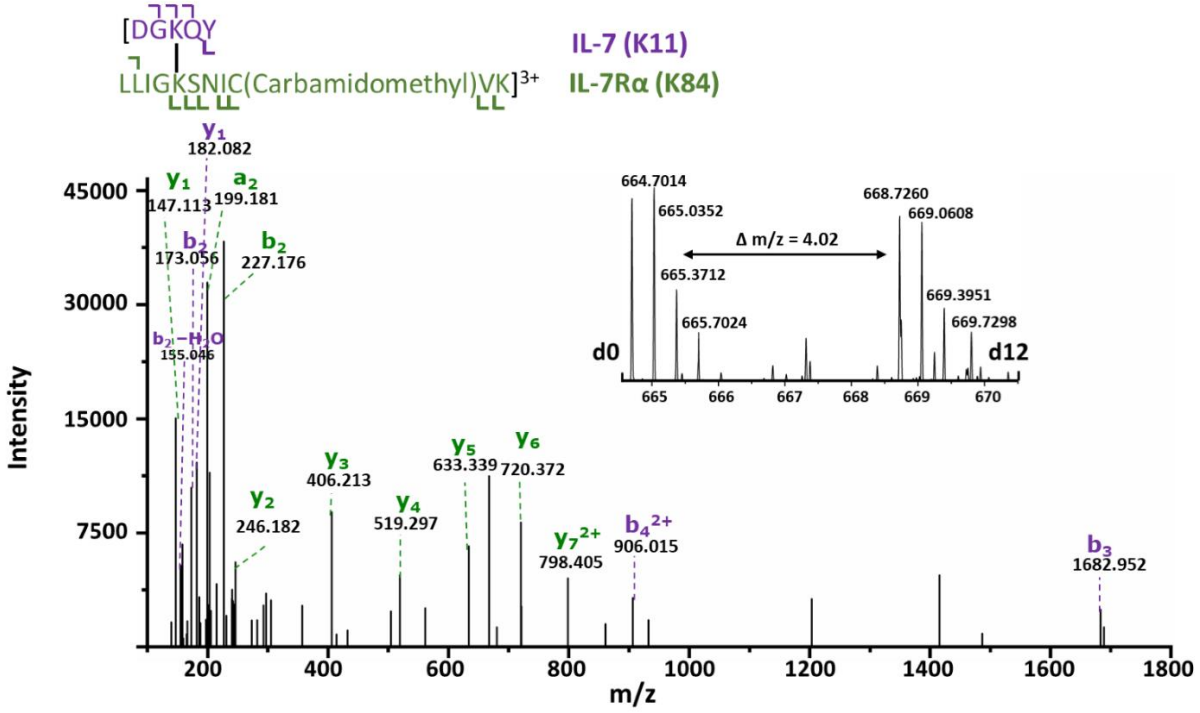


Figure 4.6 Mass spectra of cross-link 1. Extracted precursor mass spectra of cross-link 1 in the inlet, showing the isotopic peaks with equal intensity and shifted by 4 *m/z* unit when the charge state is +3. The product-ion spectrum of the heavy form of crosslinked peptide generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

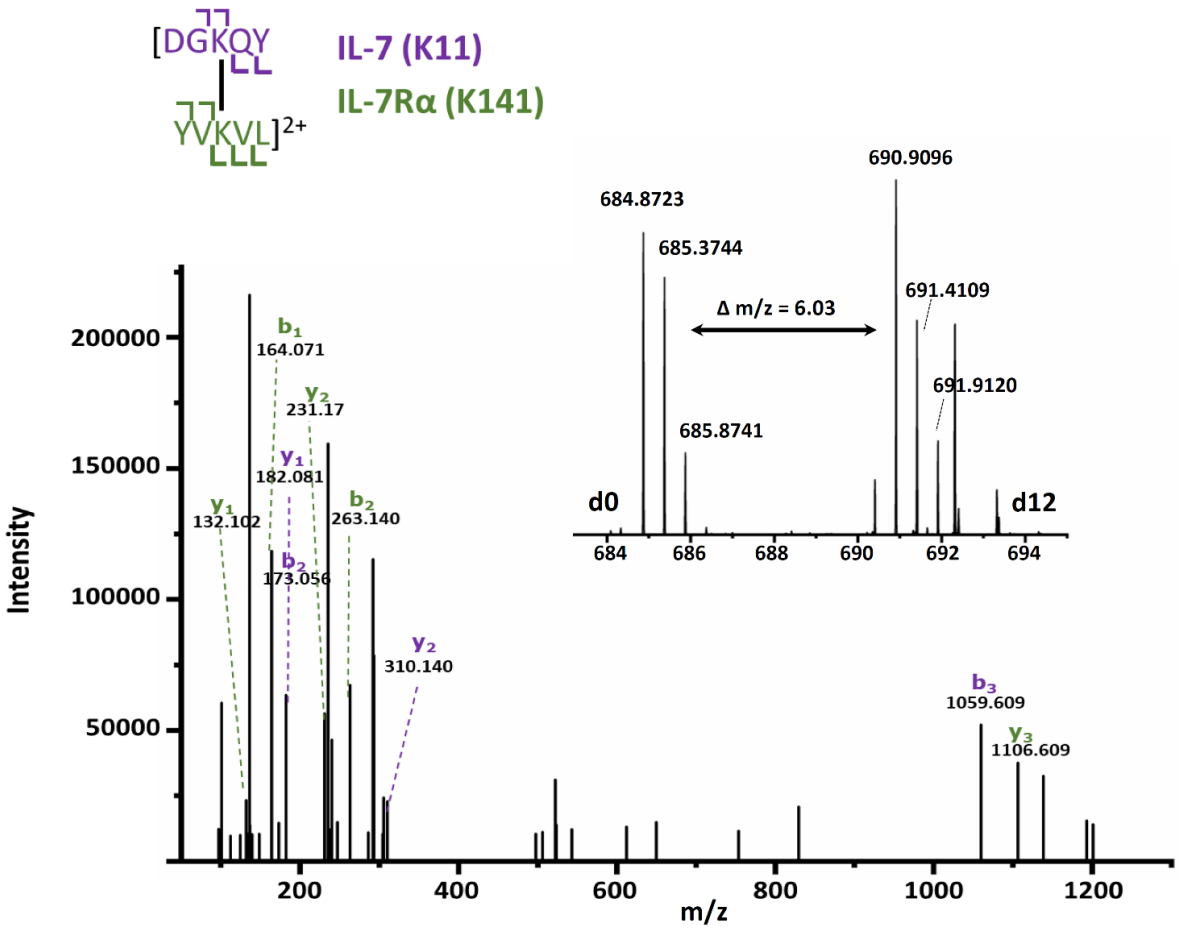


Figure 4.7 Mass spectra of cross-link 2. Extracted precursor mass spectra of cross-link 2 in the inlet, showing the isotopic peaks with equal intensity and shifted by 6 *m/z* unit when the charge state is +2. The product-ion spectrum of the light form of crosslinked peptide generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

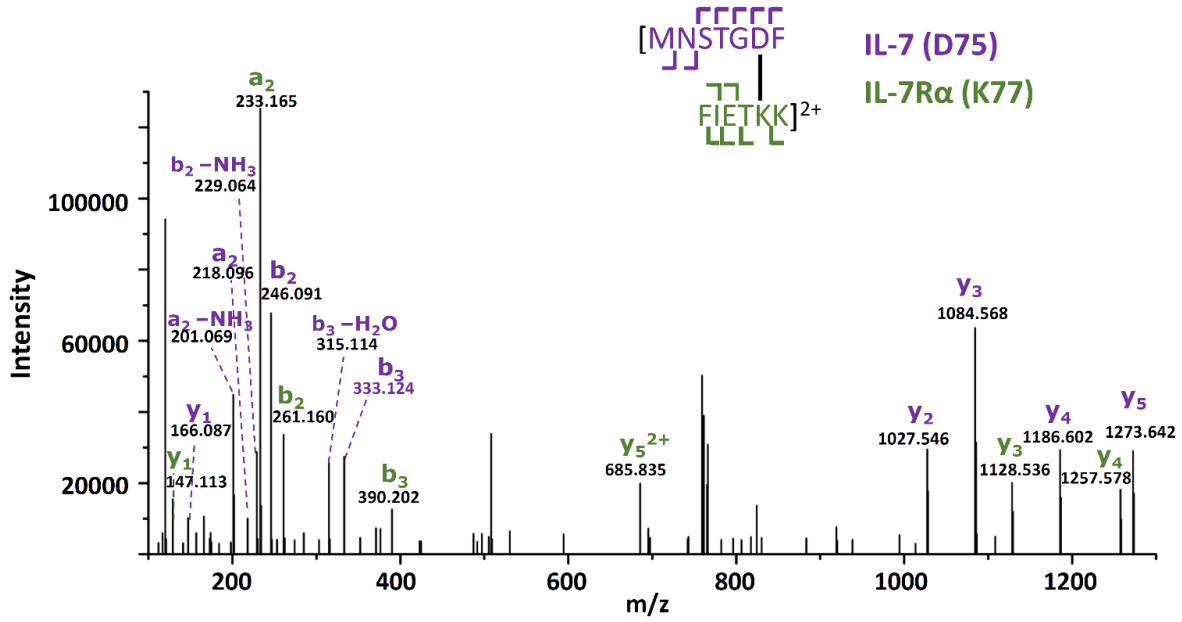


Figure 4.8 Mass spectra of cross-link 3. The product-ion spectrum of the crosslinked peptide (cross-link 3) generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

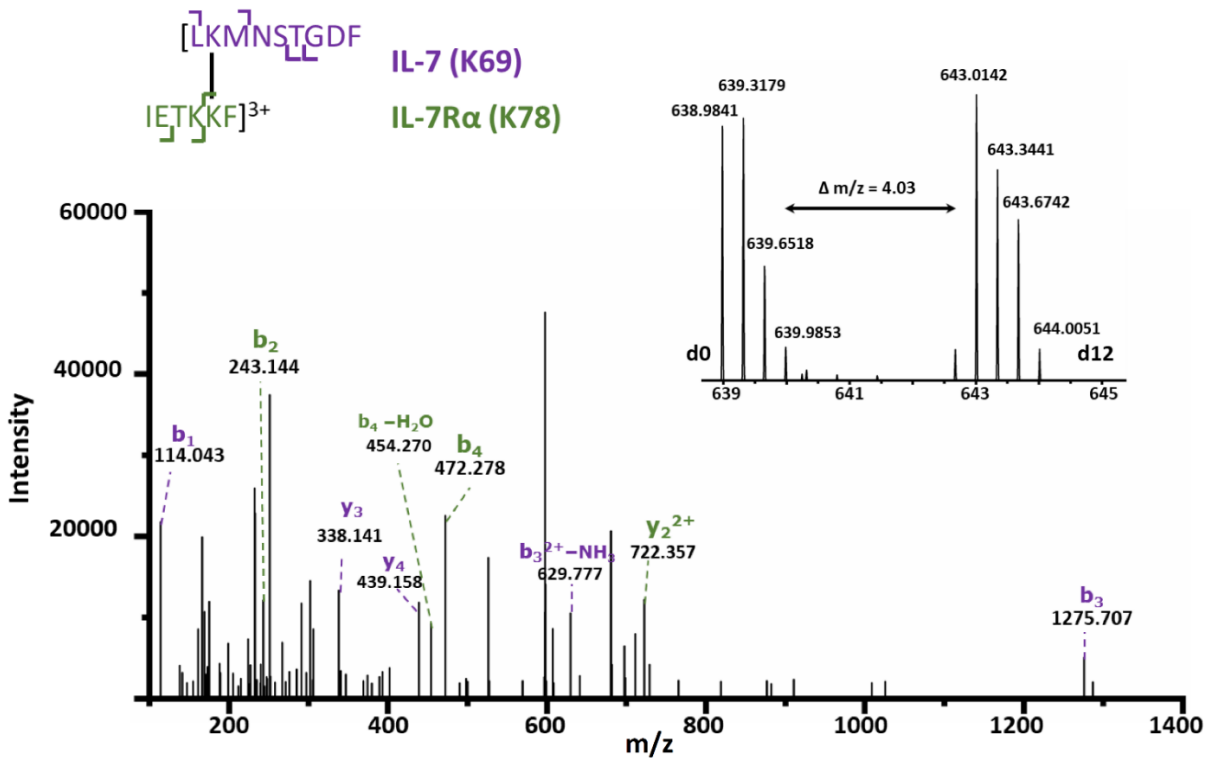


Figure 4.9. Mass spectra of cross-link 4. Extracted precursor mass spectra of cross-link 4 in the inlet, showing the isotopic peaks with equal intensity and shifted by 4 *m/z* unit when the charge state is +3. The product-ion spectrum of the light form of crosslinked peptide generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

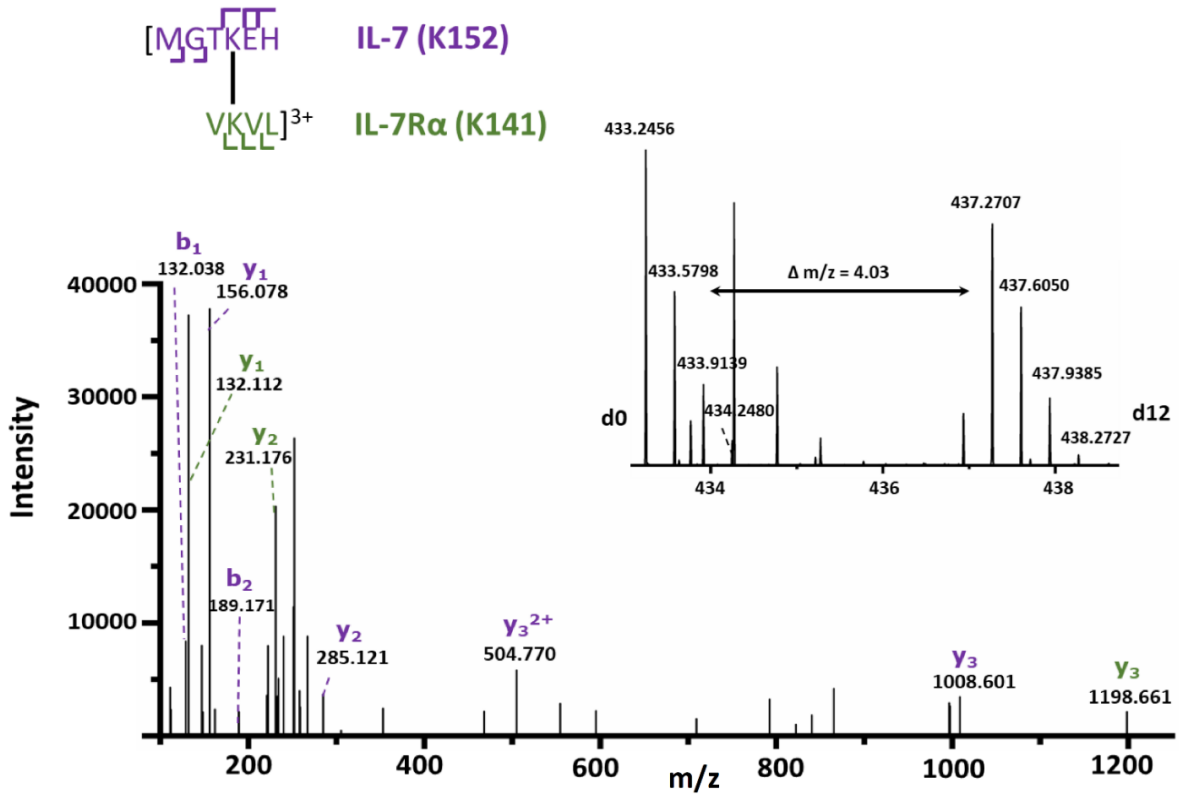


Figure 4.10 Mass spectra of cross-link 5. Extracted precursor mass spectra of cross-link 5 in the inlet, showing the isotopic peaks with equal intensity and shifted by 4 *m/z* unit when the charge state is +3. The product-ion spectrum of the light form of crosslinked peptide generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

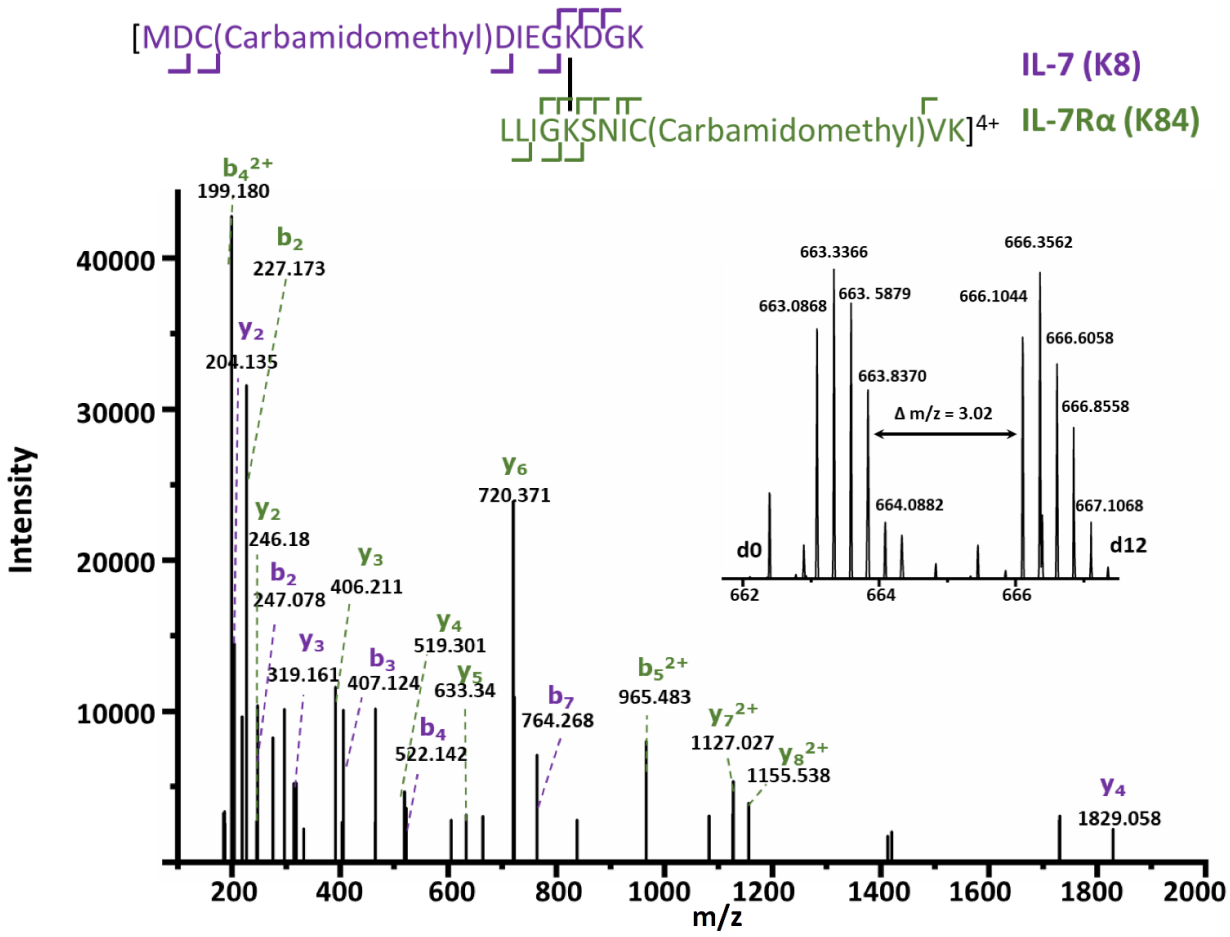


Figure 4.11 Mass spectra of cross-link 6. Extracted precursor mass spectra of cross-link 6 in the inlet, showing the isotopic peaks with equal intensity and shifted by 3 *m/z* unit when the charge state is +4. The product-ion spectrum of the light form of crosslinked peptide generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

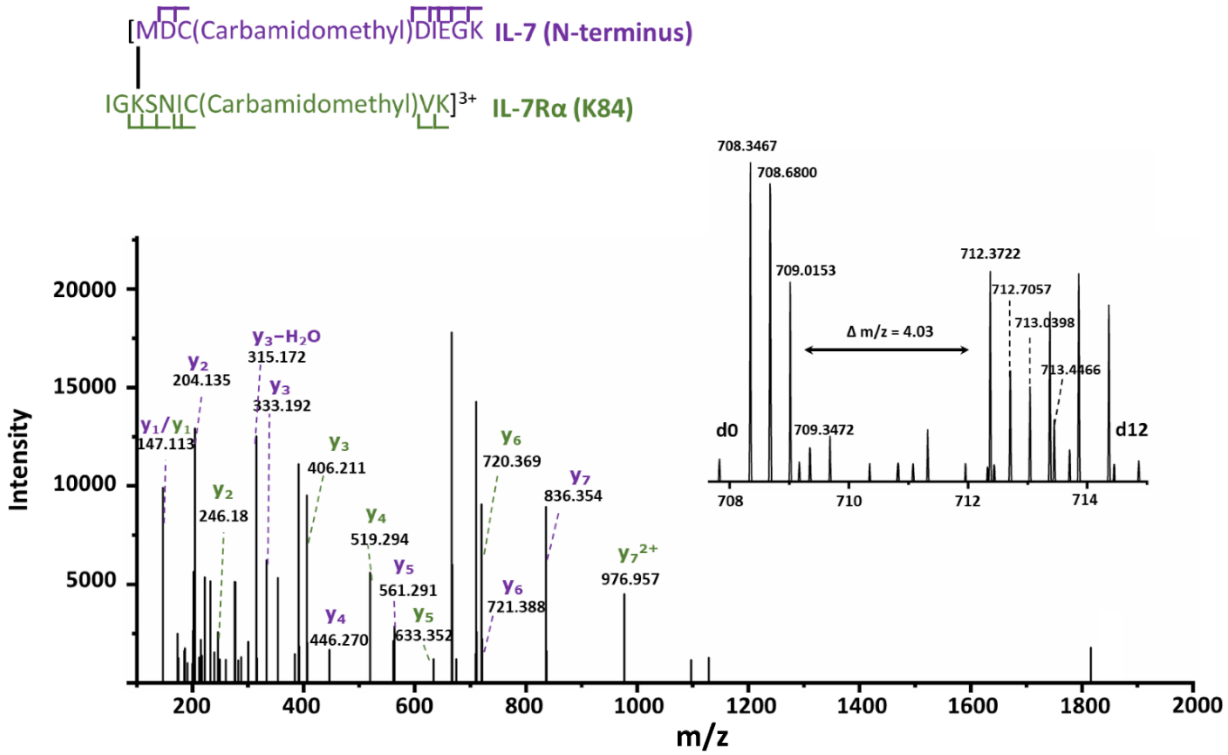


Figure 4.12 Mass spectra of cross-link 7. Extracted precursor mass spectra of cross-link 7 in the inlet, showing the isotopic peaks with equal intensity and shifted by 4 *m/z* unit when the charge state is +3. The product-ion spectrum of the light form of crosslinked peptide generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

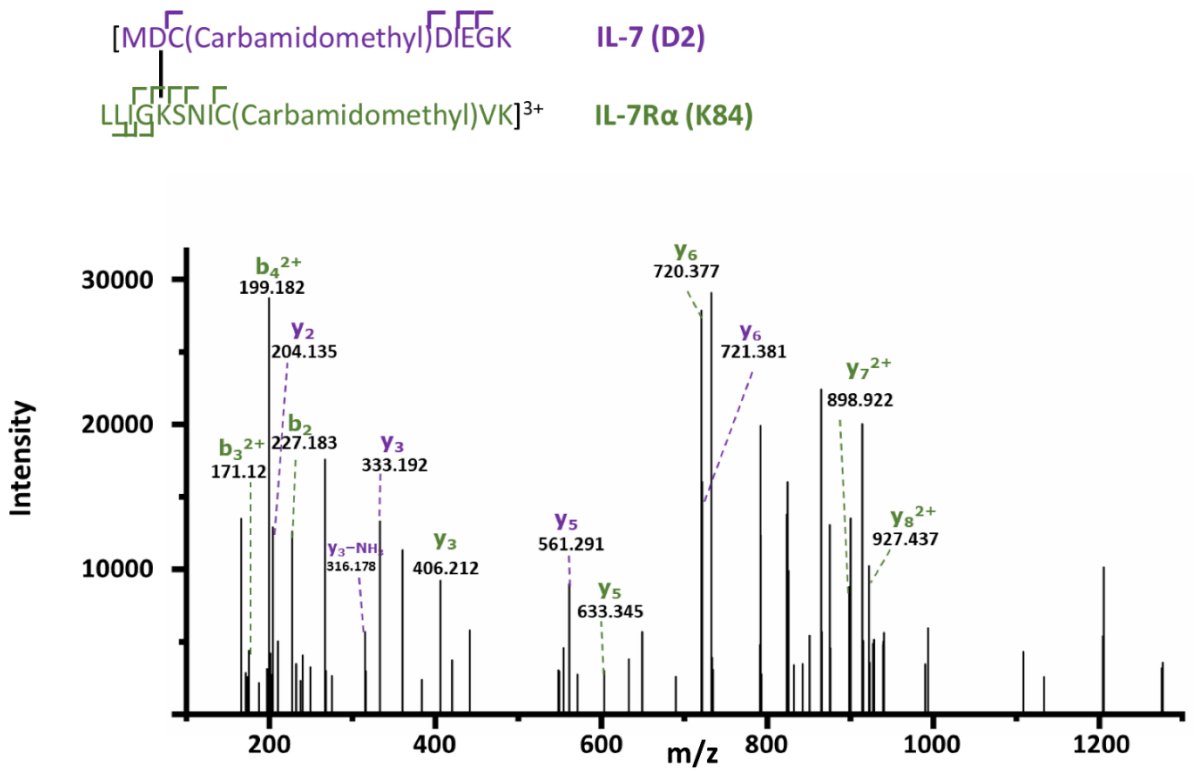


Figure 4.13 Mass spectra of cross-link 8. The product-ion spectrum of the crosslinked peptide (cross-link 8) generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

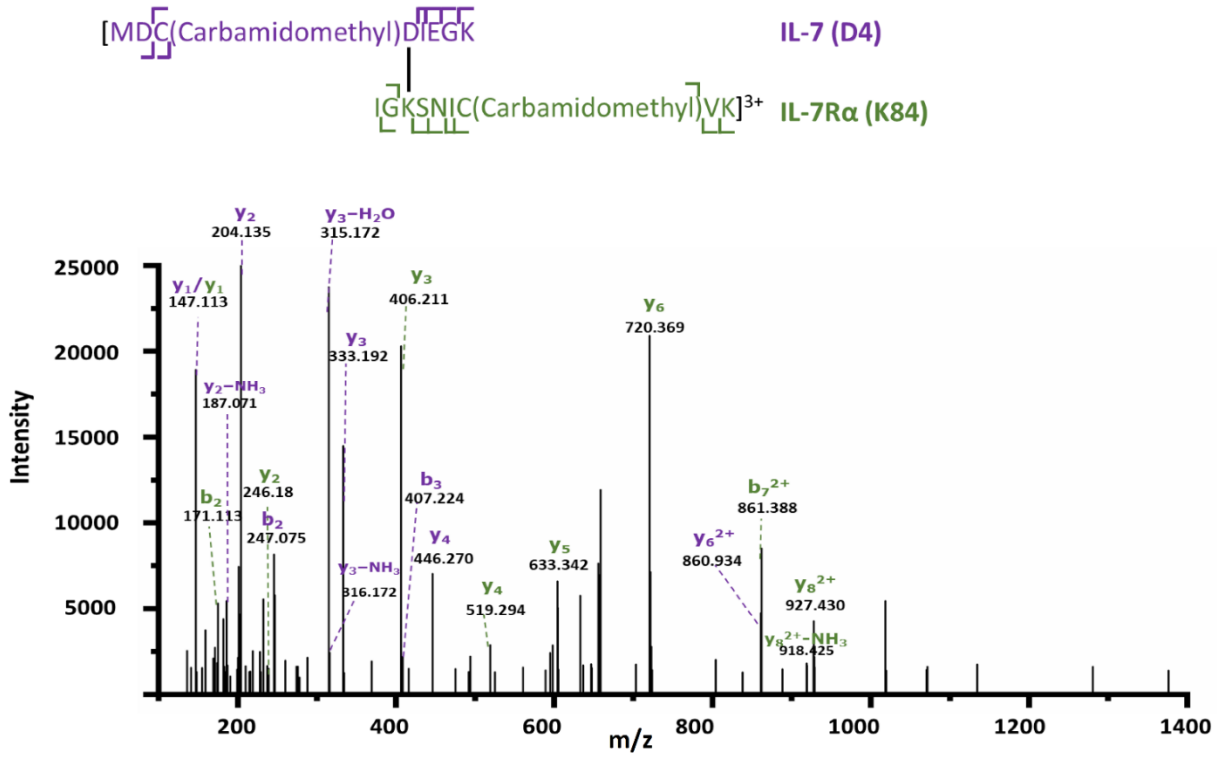


Figure 4.14 Mass spectra of cross-link 9. The product-ion spectrum of the crosslinked peptide (cross-link 9) generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

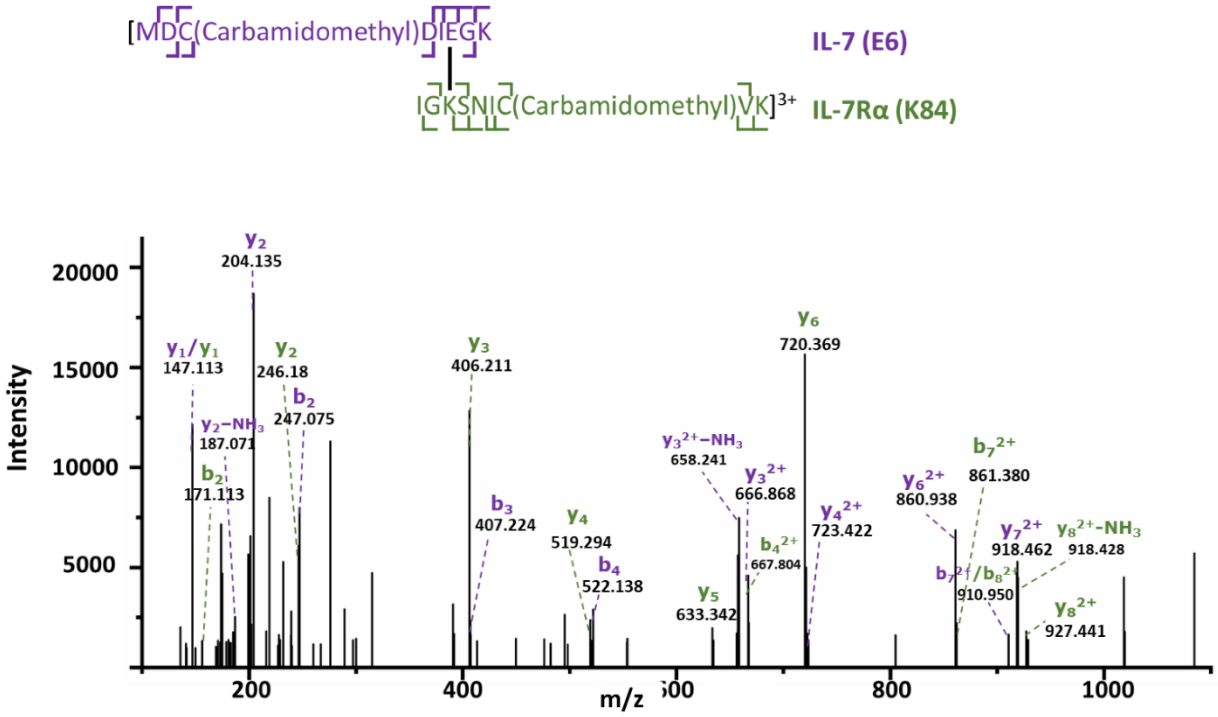


Figure 4.15 Mass spectra of cross-link 10. The product-ion spectrum of the crosslinked peptide (cross-link 10) generated by "higher energy collisional dissociation" (HCD). The assigned fragment ions are in purple for IL-7 and green for IL-7Rα.

For the IL-7Rα, there are four cross-linked residues; three of them, K77, K78 (cross-link 3 and 4) and K84 (cross-link 1 and cross-link 6 to 10) are on the elbow region of D1, and the fourth, K141 (cross-link 2 and 5) is located on the elbow loop of D2, showing discriminating evidence that these two elbow regions constitute the binding interface of IL-7Rα rather than undergoing remote conformational changes. To delineate the interacting regions of the binding complex, we turned to protein-protein docking based on crosslinking distance restraints to elucidate the quaternary structure.

4.4.3 Restraint-based Protein-Protein Docking

Protein-protein docking calculations were performed with the RosettaDock program to discern the likely structure of the IL-7/IL-7Rα complex in aqueous solution. Five representative

intermolecular cross-links (i.e., crosslinks 1-5 in Table 4.1) were selected to define distance restraints. The other five cross-links involve N-terminal residues of IL-7 that are not resolved in the input X-ray crystal structure and, thus, are not included as docking restraints. Initial protein-protein docking experiments with RosettaDock utilized all five crosslink-derived restraints simultaneously. Clustering of the top-scoring 100 IL-7/IL-7Rα complex models (RosettaDock total score) was performed with the Rosetta cluster program employing a 2.0 Å cluster radius. The largest cluster contained 53 similar models. We chose one representative model (Figure 4.16, IL-7 in purple and IL-7Rα in green) and overlaid it on to the X-ray crystal structure (Figure 4.16, black). The r.m.s.d. difference across all alpha carbons is 1.7 Å which is less than 2.0 Å cutoff used to define a high-confidence model.

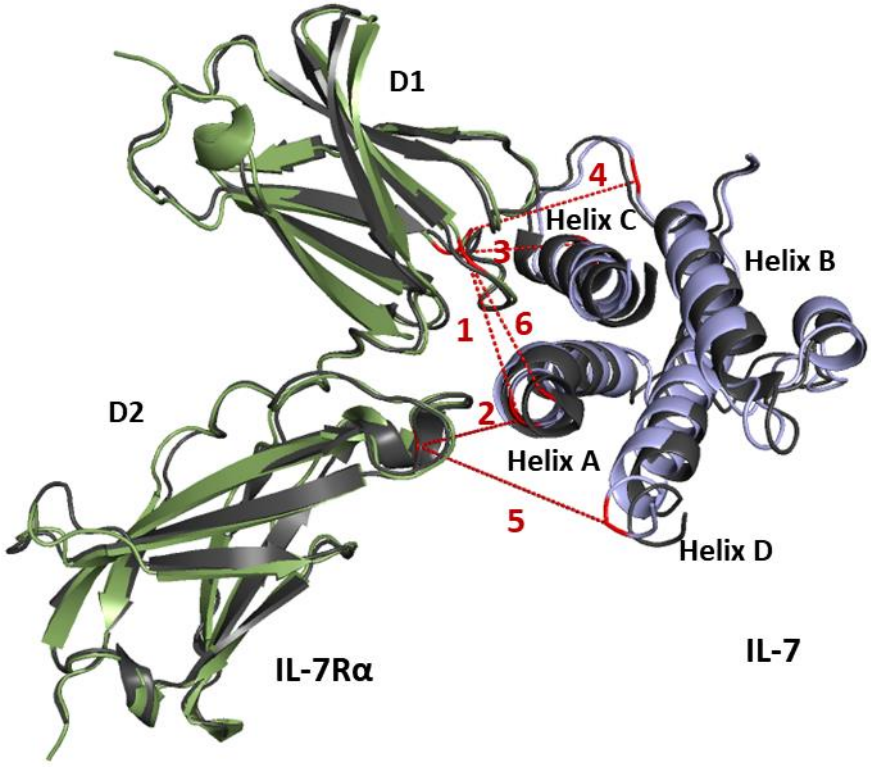


Figure 4.16 X-ray crystal structure of the IL-7/IL-7Rα (black, PDB: 3DI2) overlaid with generated docking models (green for IL-7Rα and purple for IL-7) using five intermolecular cross-links (1-5) as restraints. Cross-linked residues are in red, and dashed lines depict the

corresponding crosslinked spans. The adjoining number corresponds to cross-links shown in Table 4.1.

We further mapped cross-links 1-6 onto the generated model (Figure 4.16). Although the N-terminal region from M1 to K8 in IL-7 is not resolved, cross-link 6 is located more closely than cross-links 7-10 to the resolved residue D9, allowing its Cα-Cα distance to be estimated and taken for mapping. Cross-links 7-10, therefore, are not mapped in the complex model. All of the measured Euclidean Cα- Cα distances are within the reported range[52-53] for each specific cross-linker (Table 4.1). The excellent accord between the model and the X-ray crystal structure demonstrates the validity of solution-based data and the feasibility of using an integrative approach to obtain an accurate structure. In contrast, clustering of the top-scoring 100 models from an unrestrained docking exercise where more extensive sampling was performed (100K models) resulted in identification of only four two-membered clusters, none of which contained models that recapitulated the X-ray structure of the complex.

4.4.4 Binding Interface for IL-7 and IL-7Rα

On D1 of IL-7Rα, the elbow region (Figure 4.3-c) is heavily protected upon binding and considered to interact primarily with helix C of the IL7. Not only the adjacent positions of these two domains, from directly viewing the docking model, but also a short EDC cross-link 3 support their proximity. The observation is consistent with the contacting residues resolved from the crystal structure[39], where K77 on the D1 elbow region of IL-7Rα forms H-bonds with D75 and K82 on helix C of IL-7. Furthermore, HDX kinetics reveals another participating region on the C-terminus of helix B, evidenced by the low deuterium uptake (Figure 4.1A-c) and cross-link 4. In addition, cross-link 6-10 suggest another contact region with the D1 elbow, the N-terminal region of helix A; this region, however, shows no protection in the HDX kinetics (Figure 4.1B),

decreasing the likelihood of being the binding interface. Besides the direct binding domains, adjacent and proximal regions can also allow crosslinking formation, which delivers topological information of the IL-7/IL7Rα complex. This observation shows the limitation of using stand-alone XL-MS and underscores the necessity of incorporating other complementary approaches to achieve an impartial determination of a protein-protein interface. The other elbow region (Figure 4.3-e) on D2 of IL-7Rα is bridged to helix A of IL-7 by cross-link 2, indicating close location of those regions in the generated model. The cross-linked residue, K11, is the nearest linkable residue to Q23, which contributes to H-bond formation with the paired residue K138 in D2[39]. The C-terminus of helix D adjoins the D2 elbow loop, as represented by cross-link 5. The less than 10% cumulative deuterium uptake difference in region 138-152 (Figure 4.1B) and distant location in the docking model, however, suggest involvement but not prominent interaction. The integrated platform of using multiple cross-linkers and of incorporating protein-protein docking significantly increases the certainty and defines better the directly contacting regions.

4.4.5 The Number of Intermolecular Cross-links Needed for a High-Confidence Model

Excellent results were obtained in the protein-protein docking exercise that utilized five crosslink-derived restraints. However, in the context of the integrated approach, it is possible that accurate models might be obtained with fewer restraints. To examine this possibility, we conducted docking runs with all possible combinations of using 1 to 5 cross-links as distance restraints. For each restraint combination, we identified the top-scoring 100 IL-7/IL-7Rα docking models with RosettaDock total score and clustered them with the Rosetta cluster program by using the aforementioned criteria (Cα atom r.m.s.d. $\leq$ 2.0). For an unknown system, the clusters with the largest size are likely to be the most promising ones to choose. Thus, we summarized

and evaluated all the largest clusters for each of the combinations (Figure 4.17). When using a restraint combination, multiple clusters of equal size are possible (e.g., for cross-link 1 and the combination of cross-links 4 and 5). Docking with larger number of restraints generally leads to larger cluster sizes; however, the size also depends on the cross-links chosen. We found that shorter cross-links are more demanding and lead to larger clusters. Supporting this contention is that the largest clusters in each category (e.g., one restraint, two restraints) all contain cross-link 3, the EDC cross-link. This observation further emphasizes the importance of employing a multiple-cross-linker strategy to provide both tight requirements for crosslinking and, at the same time, relatively loose restraints that allow many cross-links to form, thus aiding the downstream analysis.

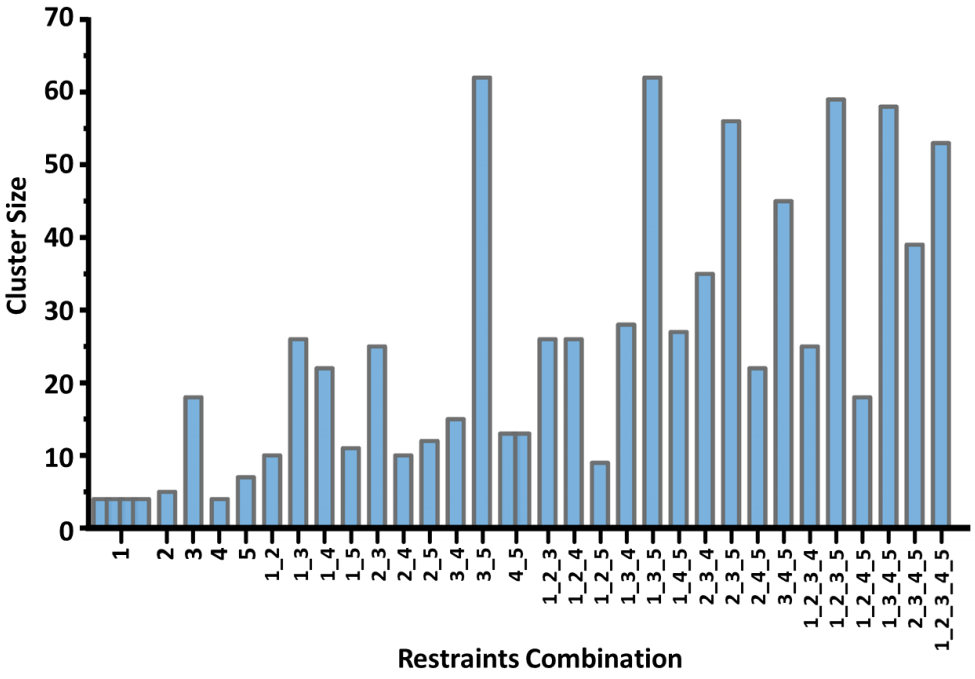


Figure 4.17 Cluster sizes for largest clusters of IL-7/IL-7Rα models generated with each possible combination of 1-5 crosslink-derived restraints from protein-protein docking. Several clusters with nearly equal members are formed for restraint combinations of cross-link 1 and combinations of cross-links 4 and 5.

Table 4.2 Summary of representative models from the largest clusters by using any two cross-links as docking restraints. The solvent accessible area (SASA) of two most protected regions in IL-7, as indicated by HDX kinetics, are calculated for each docking model and for the unbound-state IL-7 in Pymol.

Crosslink-Based Restraints	Size of Largest Cluster	Model Category	SASA of Peptide in Il-7, Å				Matching with HDX (Y/N)	Compared to X-Ray Structure, R.M.S.D, Å
			^{19}VSIDQL^{24}		^{83}VSEGTTIL^{90}			
			Bound IL-7 in Models	Unbound IL-7	Bound IL-7 in Models	Unbound IL-7		
1_2	10	Type 3	230	230	268	319	N	12.3
1_3	26	Type 1	133		220		Y	1.7
1_4	22	Type 2	230		247		N	11.0
1_5	11	Type 1	141		211		Y	1.4
2_3	25	Type 1	158		213		Y	1.9
2_4	10	Type 2	230		269		N	10.9
2_5	12	Type 1	137		209		Y	1.5
3_4	15	Type 1	142		214		Y	1.7
3_5	62	Type 1	160		201		Y	1.7
4_5 **4_5.1**	13	Type 1	127		213		Y	1.3
4_5.2	13	Type 2	230		253		N	10.9

Protein-protein docking with 10 different combinations of 5 crosslink-derived restraints taken two at a time yielded 11 largest clusters in total (Figure 4.17 and Table 4.2). Based on the structural similarity, we grouped and overlaid them into three different types (Figure 4.18). Our approach utilizes HDX results at this stage to evaluate the accuracy of the models. The omnipresent protection of IL-7Rα upon binding make IL-7Rα a poor choice to adjudicate the models. Instead, we chose only IL-7 HDX kinetics for the evaluation and mapped the residue-level cumulative deuterium uptake difference onto its structure as shown before (Figure 4.1). To place the comparison on a quantitative basis, we calculated the solvent accessible area (SASA) of the two most protected regions, ^{19}VSIDQL^{24} and ^{83}VSEGTTIL^{90}, as indicated by HDX, for each generated model (Table 4.2). For the type 1 models (Figure 4.18A), which are the most populated models given by seven different combinations (Table 4.2), the averaged SASA is 143 ± 12 Å and 212 ± 6 Å for ^{19}VSIDQL^{24} and ^{83}VSEGTTIL^{90}, respectively. The values are significantly smaller than 230 Å and 319 Å in the unbound IL-7, suggesting there would be less deuterium uptake in the bound IL-7/IL-7Rα, as is seen in the HDX kinetics.

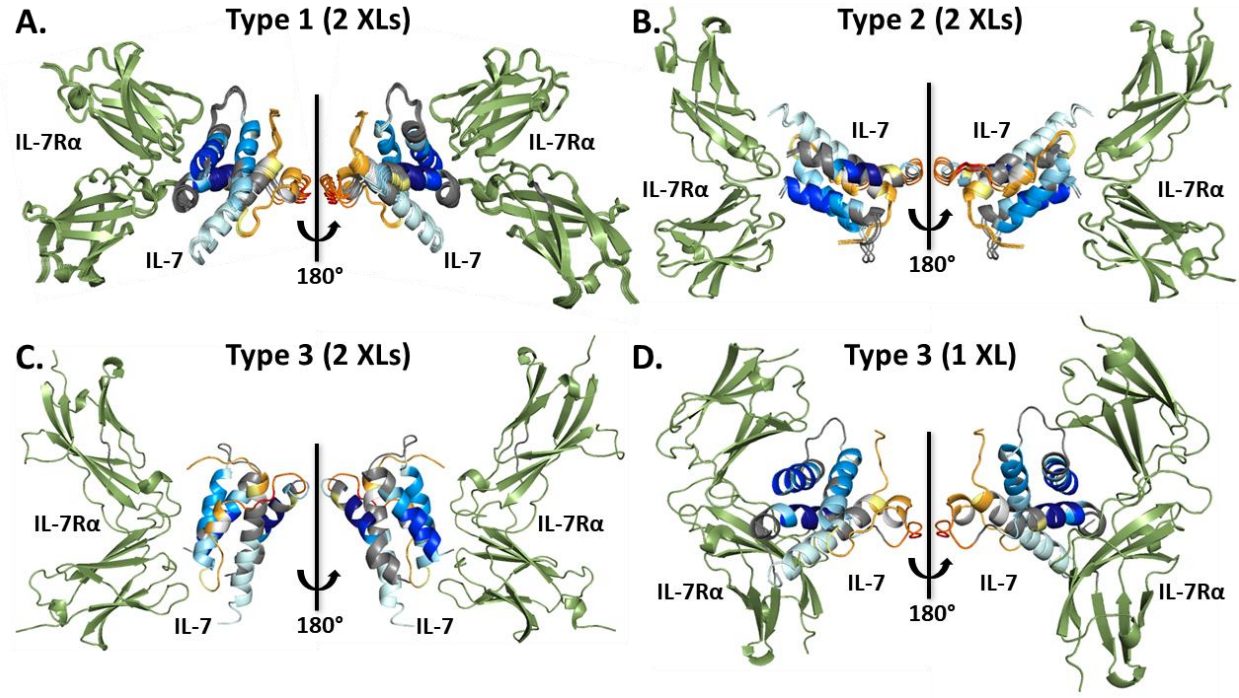


Figure 4.18 Different IL-7/IL-7Rα model-types, including type 1 (A), type 2 (B) and type 3 (C), generated from protein-protein docking with different combinations of two crosslink-derived restraints. IL-7 is mapped with the cumulative HDX uptake difference whereas IL-7Rα is colored in green. Similar models in each type are overlaid. (D) Type 3 IL-7/IL-7Rα model generated using only cross-link 1.

For the type 2 models (Figure 4.18B), however, the two regions are outside the contacting interface and highly exposed, evidenced explicitly by a SASA for the region ^{19}VSIDQL^{24} of 230 ± 1 Å that is identical to the SASA of the unbound IL-7. This identity indicates no binding-induced protection for this region, in contradistinction to the HDX results. This inconsistency is further shown in other regions, for example the loops that connect different helices. Although type 2 models suggest the loops are the binding interface, HDX reports they either become more exposed upon binding or show no differences between two states (Figure 4.1B). The conflicting conclusions from HDX and docking forcefully decrease the likelihood of type 2 models to be correct. From similar reasoning, the type 3 model (Figure 4.18C) is also inconsistent with the HDX kinetics. Although the SASA of ^{83}VSEGTTIL^{90} is 268 Å, smaller than that of unbound IL-

7, the SASA of the other most protected region, ^{19}VSIDQL^{24}, is the same as that of unbound, indicating no involvement in the IL-7/IL-7Rα interaction. Instead, the suggested binding interface is consisted of a mini-helix and a flexible loop, which shows more deuterium uptake (Figure 4.1A-b) in HDX upon binding. Guidance from the HDX data allows us to reasonably rule out dubious docking models.

A subsequent evaluation of the remaining models, i.e. type 1 models, after HDX adjudication and comparison with the published crystal structure reveals only subtle structural differences (Table 4.2), all smaller than 2 Å r.m.s.d for all alpha carbons. These latter models meet the criteria promoted in this study to define a high-confidence model. Using this integrated platform, we can conclude that docking restraints based on two intermolecular cross-links are adequate to identify an accurate quaternary model for this binding complex.

Using three or four inter cross-links as restraints leads to better populated clusters and well-defined IL-7/IL-7Rα models. Only two types of models (Figure 4.19) are generated and the second type can be excluded by considering the HDX results in the decision, where not the SASA values of both two peptides show differences between the bound and unbound IL-7 (Table 4.3). The remaining models are all considered as high-confidence models upon comparison to the X-ray structure. Although using more than two cross-links also gives an accurate model, with no surprise, the successful exercise provides confidence for utilizing integrated methods to obtain higher-order-structure information.

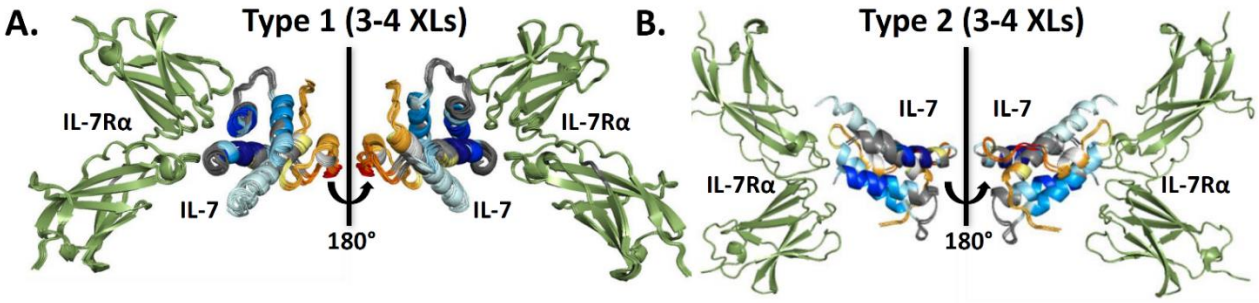


Figure 4.19 Different IL-7/IL-7Rα model-types, including type 1 (A) and type 2 (B), generated from three and four cross-links but with different combinations. IL-7 is mapped with the cumulative HDX uptake difference whereas IL-7Rα is colored in green. Similar models in each type are overlaid.

Table 4.3 Summary of representative models from the largest clusters using three and four cross-links as docking constraints. The solvent accessible area (SASA) of two most protected regions in IL-7 indicated by HDX kinetics are calculated for each docking model and apo-state IL-7 in Pymol.

Crosslink-Based Restraints	Cluster Size	Category	SASA of Peptides in Il-7, Å				Matching with HDX (Y/N)	Compared to X-Ray Structure, R.M.S.D, Å
			^{19}VSIDQL^{24}		^{83}VSEGTTIL^{90}			
			Bound IL-7 in Models	Unbound IL-7	Bound IL-7 in Models	Unbound IL-7		
1_2_3	26	Type 1	151	230	205	319	Y	1.9
1_2_4	26	Type 2	230		252		N	11.0
1_2_5	9	Type 1	128		215		Y	1.7
1_3_4	28	Type 1	148		199		Y	1.5
1_3_5	62	Type 1	138		195		Y	1.5
1_4_5	27	Type 2	230		248		N	11.0
2_3_4	35	Type 1	127		225		Y	1.5
2_3_5	56	Type 1	141		206		Y	1.7
2_4_5	22	Type 1	141		204		Y	1.5
3_4_5	45	Type 1	142		213		Y	1.7
1_2_3_4	25	Type 1	138		208		Y	1.7
1_2_3_5	59	Type 1	137		216		Y	1.7
1_2_4_5	18	Type 2	230		248		N	11.0
1_3_4_5	58	Type 1	130		213		Y	1.6
2_3_4_5	39	Type 1	145		207		Y	1.7

When using only one cross-link as the docking restraint, however, seven different model-types are generated (Figure 4.18 for type 3 and Figure 4.20 for other types). Although some of these can survive HDX adjudication, they still deliver erroneous structures. Type 3 model (Figure

4.18) is an example, showing the two most protected regions interacting with the IL-7Rα. The SASA value for ^{19}VSIDQL^{24} and ^{83}VSEGTTIL^{90} are 65 Å and 182.3 Å, respectively, all smaller than that of the unbound IL-7, suggesting consistent protection, as observed from HDX kinetics. However, due to the mis-orientation of IL-7Rα, the r.m.s.d. difference in comparison with the crystal structure is 14.8 Å (Table 4.4), far exceeding the 2 Å threshold for a good model. A similar situation also applies to type 5 and type 7 models (Figure 4.20 and Table 4.4). Serving as a comparison control, one cross-link is insufficient to assign an accurate quaternary structure; thus, the minimal number of cross-links to fulfill this goal is two.

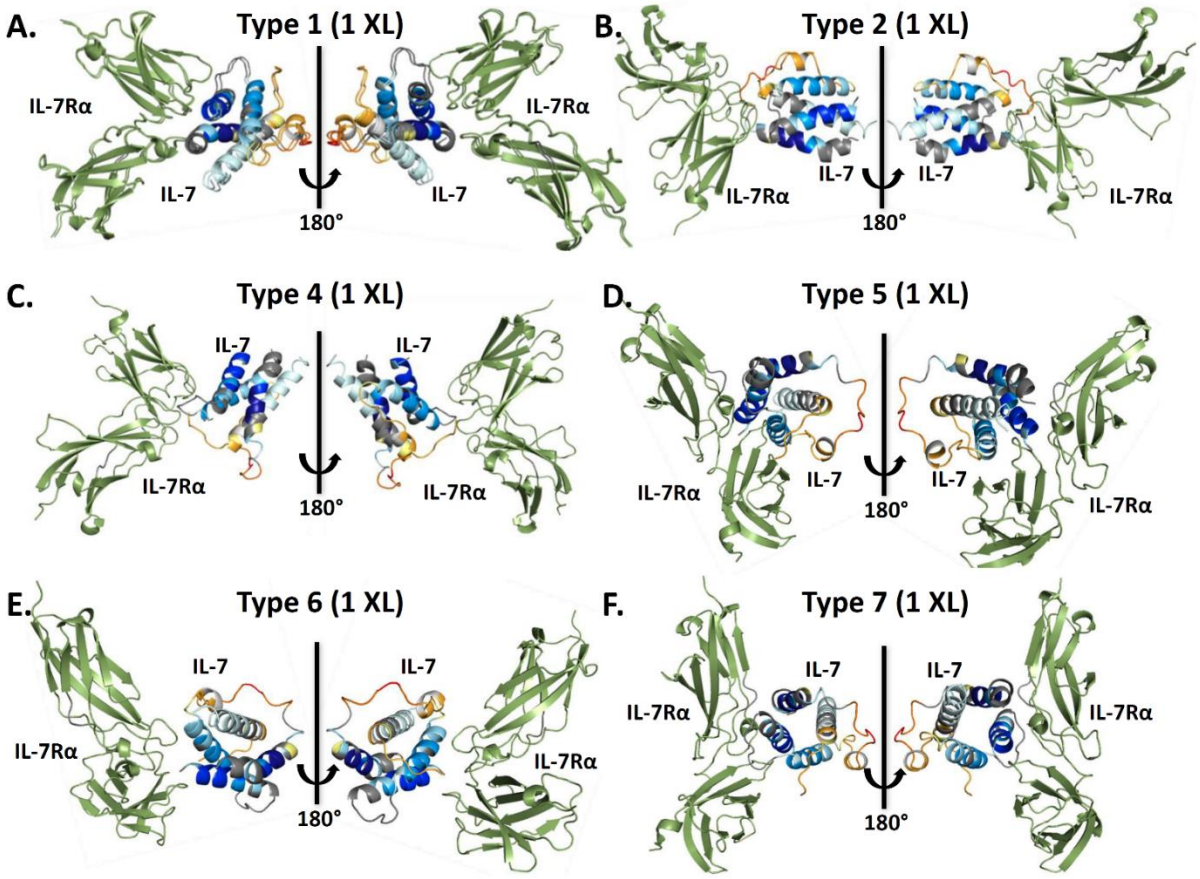


Figure 4.20 Different IL-7/IL-7Rα model-types, including type 1 (A), type 2 (B), type 4 (C), type 5 (D), type 6 (E), type 7 (F), generated using only one cross-link. IL-7 is mapped with the cumulative HDX uptake difference whereas IL-7Rα is colored in green. Similar models in each type are overlaid.

Table 4.4 Summary of representative models from the largest clusters using individual cross-link as docking constraint. The solvent accessible area (SASA) of two most protected regions in IL-7 indicated by HDX kinetics are calculated for each docking model and apo-state IL-7 in Pymol.

Crosslink-Based Restraints		Cluster Size	Category	SASA of Peptides in Il-7, Å				Matching with HDX (Y/N)	Compared to X-Ray Structure, R.M.S.D, Å
				^{19}VSIDQL^{24}		^{83}VSEGTTIL^{90}			
				Bound IL-7 In Models	Unbound IL-7	Bound IL-7 In Models	Unbound IL-7		
1	**1.1**	4	Type 1	137	230	219	319	Y	1.6
	1.2	4	Type 2	230		335		N	15.1
	1.3	4	Type 3	65.0		182		Y	14.8
	1.4	4	Type 4	230		305		N	14.0
2		5	Type 5	230		143		Y	15.1
3		18	Type 1	138		207		Y	1.7
4		4	Type 6	230		222		N	14.0
5		7	Type 7	169		92		Y	11.4

4.5 Conclusion

This work highlights the effectiveness of the integrated approach to deliver quaternary structural information. Although the generality of the number “two cross-links” may vary from system to system, the prospect of generating a high-confidence model based on a few cross-links is clearly demonstrated. More crosslinks-derived restraints increase the likelihood of generating accurate models, but only modestly, and extensive efforts to identify large numbers of cross-links may not be justified for some protein-protein complexes.

The study also demonstrates successful utilization of HDX results in combination with XL-enabled docking studies to identify the IL-7/IL-7Rα binding interface and to predict the quaternary structure of the complex in solution. Use of multiple cross-linkers spanning diverse distances and targeting different amino acids increases the chances of identifying inter-protein contacting regions and the spatial-resolution; restraints derived from shorter crosslinking reagents more effectively focus the docking results towards native-like models. The excellent results achieved for the IL-7/IL-7Rα system may be due in part to the rich concentration of

lysine residues along the binding interface, but for interfaces that are not lysine-rich, employing non-selective crosslinkers, (e.g., sulfo-NHS-SDA), can overcome that limitation.

Implementation of several complementary methods significantly enriches the structural information from tertiary to quaternary, compared to any stand-alone method. The delivery of a high-confidence model with only two experimental cross-links as restraints demonstrate the capability of an integrated platform. It can not only categorize changes arising from either binding-induced protection or remote structural changes but also define the binding interface more precisely with modeling. For many other protein complexes, which require considerable effort to crystalize or don't crystallize at all, the approach is an intriguing alternative to obtain structural information. That information applies to the proteins in solution and can fill in regions that do not diffract well. The established approach will certainly aid the design of protein therapeutics.

4.6 Acknowledgments

This study is a team effort by all authors. M.M.Z. conducted HDX, XL-MS, and wrote the manuscript. B.R.B conducted molecular docking, wrote the corresponding experimental section, and edited the manuscript. R.Y-C., J.A., E.G.D., J.L., G.C., M.L.G. edited the manuscript.

The authors thank Dr. Daniel A Weisz, from department of biology in WUSTL, for his generous help and Dr. Dina Schneidman, Hebrew University of Jerusalem, for helpful discussion. This work was supported by a Research Collaboration with Bristol-Myers Squibb and by the NIH 2P41GM103422. The authors thank Dr. Olafur Gudmundsson and Dr. Lois Lehman-McKeeman from Bristol-Myers Squibb for their support of this project.

4.7 References

1. Rebsamen, M.; Kandasamy, R. K.; Superti-Furga, G. Protein interaction networks in innate immunity. *Trends Immunol.* **2013**, *34*, 610-9.

2. Krohl, P. J.; Ludwig, S. D.; Spangler, J. B. Emerging technologies in protein interface engineering for biomedical applications. *Curr. Opin. Biotechnol.* **2019**, *60*, 82-88.

3. Amit, A. G.; Mariuzza, R. A.; Phillips, S. E.; Poljak, R. J. Three-dimensional structure of an antigen-antibody complex at 2.8 A resolution. *Science* **1986**, *233*, 747-53.

4. Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* **1999**, *38*, 1784-1788.

5. Weis, D. D. Hydrogen Exchange Mass Spectrometry of Proteins. Wiley Online Library: 2015.

6. 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*, 6607-14.

7. Wang, H. L.; Shu, Q.; Rempel, D. L.; Frieden, C.; Gross, M. L. Understanding curli amyloid-protein aggregation by hydrogen-deuterium exchange and mass spectrometry. *Int. J. Mass Spectrom.* **2017**, *420*, 16-23.

8. Huang, R. Y.; Chen, G. Higher order structure characterization of protein therapeutics by hydrogen/deuterium exchange mass spectrometry. *Anal. Bioanal. Chem.* **2014**, *406*, 6541-58.

9. Wang, H.; Shu, Q.; Rempel, D. L.; Frieden, C.; Gross, M. L. Understanding Curli Amyloid-Protein Aggregation by Hydrogen-Deuterium Exchange and Mass Spectrometry. *Int. J. Mass Spectrom.* **2017**, *420*, 16-23.

10. Li, J.; Wei, H.; Krystek, S. R., Jr.; Bond, D.; Brender, T. M.; Cohen, D.; Feiner, J.; Hamacher, N.; Harshman, J.; Huang, R. Y.; 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*, 2250-2258.

11. Chik, J. K.; Schriemer, D. C. Hydrogen/deuterium exchange mass spectrometry of actin in various biochemical contexts. *J. Mol. Biol.* **2003**, *334*, 373-385.

12. Pan, J. X.; Han, J.; Borchers, C. H.; Konermann, L. Characterizing Short-Lived Protein Folding Intermediates by Top-Down Hydrogen Exchange Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 8591-8597.

13. Sterling, H. J.; Williams, E. R. Real-time hydrogen/deuterium exchange kinetics via supercharged electrospray ionization tandem mass spectrometry. *Anal. Chem*. **2010**, *82*, 9050-7.

14. Huang, R. Y.; Krystek, S. R., Jr.; Felix, N.; Graziano, R. F.; Srinivasan, M.; Pashine, A.; Chen, G. Hydrogen/deuterium exchange mass spectrometry and computational modeling reveal a discontinuous epitope of an antibody/TL1A Interaction. *MAbs* **2018**, *10*, 95-103.

15. Sinz, A. Chemical crosslinking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass. Spectrom. Rev.* **2006**, *25*, 663-682.

16. Yu, C.; Huang, L. Crosslinking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* **2018**, *90*, 144-165.

17. Gotze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; Krauth, F.; Fritzsche, R.; Kuhn, U.; Sinz, A. StavroX-A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 76-87.

18. Rey, M.; Sarpe, V.; Burns, K. M.; Buse, J.; Baker, C. A. H.; van Dijk, M.; Wordeman, L.; Bonvin, A. M. J. J.; Schriemer, D. C. Mass Spec Studio for Integrative Structural Biology. *Structure* **2014**, *22*, 1538-1548.

19. Yang, B.; Wu, Y. J.; Zhu, M.; Fan, S. B.; Lin, J. Z.; Zhang, K.; Li, S.; Chi, H.; Li, Y. X.; Chen, H. F.; Luo, S. K.; Ding, Y. H.; Wang, L. H.; Hao, Z. Q.; Xiu, L. Y.; Chen, S.; Ye, K. Q.; He, S. M.; Dong, M. Q. Identification of cross-linked peptides from complex samples. *Nat. Methods* **2012**, *9*, 904-906.

20. Schmidt, C.; Robinson, C. V. A comparative crosslinking strategy to probe conformational changes in protein complexes. *Nat. Protoc.* **2014**, *9*, 2224-2236.

21. Zhang, Q.; Feng, T.; Xu, L.; Sun, H.; Pan, P.; Li, Y.; Li, D.; Hou, T. Recent Advances in Protein-Protein Docking. *Curr. Drug Targets* **2016**, *17*, 1586-1594.

22. Park, H.; Lee, H.; Seok, C. High-resolution protein–protein docking by global optimization: recent advances and future challenges. *Curr. Opin. Struct. Biol.* **2015**, *35*, 24-31.

23. Gromiha, M. M.; Yugandhar, K.; Jemimah, S. Protein–protein interactions: scoring schemes and binding affinity. *Curr. Opin. Struct. Biol.* **2017**, *44*, 31-38.

24. Moal, I. H.; Moretti, R.; Baker, D.; Fernandez-Recio, J. Scoring functions for protein–protein interactions. *Curr. Opin. Struct. Biol.* **2013**, *23*, 862-867.

25. Brodie, N. I.; Popov, K. I.; Petrotchenko, E. V.; Dokholyan, N. V.; Borchers, C. H. Solving protein structures using short-distance crosslinking constraints as a guide for discrete molecular dynamics simulations. *Sci. Adv.* **2017**, *3*, e1700479.

26. Rampler, E.; Stranzl, T.; Orban-Nemeth, Z.; Hollenstein, D. M.; Hudecz, O.; Schlögelhofer, P.; Mechtler, K. Comprehensive crosslinking mass spectrometry reveals parallel orientation and flexible conformations of plant HOP2–MND1. *J. Proteom. Res.* **2015**, *14*, 5048-5062.

27. Lin, S.-J.; Chen, Y.-F.; Hsu, K.-C.; Chen, Y.-L.; Ko, T.-P.; Lo, C.-F.; Wang, H.-C.; Wang, H.-C. Structural Insights to the Heterotetrameric Interaction between the Vibrio parahaemolyticus PirAvp and PirBvp Toxins and Activation of the Cry-Like Pore-Forming Domain. *Toxins* **2019**, *11*, 233.

28. Komolov, K. E.; Du, Y.; Duc, N. M.; Betz, R. M.; Rodrigues, J. P.; Leib, R. D.; Patra, D.; Skiniotis, G.; Adams, C. M.; Dror, R. O. Structural and functional analysis of a β2-adrenergic receptor complex with GRK5. *Cell* **2017**, *169*, 407-421.

29. Elenkov, I. J.; Iezzoni, D. G.; Daly, A.; Harris, A. G.; Chrousos, G. P. Cytokine dysregulation, inflammation and well-being. *Neuroimmunomodulation* **2005**, *12*, 255-269.

30. Walsh, S. T. R. Structural insights into the common gamma-chain family of cytokines and receptors from the interleukin-7 pathway. *Immunol. Rev.* **2012**, *250*, 303-316.

31. Kittipatarin, C.; Khaled, A. R. Interlinking interleukin-7. *Cytokine* **2007**, *39*, 75-83.

32. Gao, J. B.; Zhao, L. T.; Wan, Y. S. Y.; Zhu, B. Mechanism of Action of IL-7 and Its Potential Applications and Limitations in Cancer Immunotherapy. *Int. J. Mol. Sci.* **2015**, *16*, 10267-10280.

33. ElKassar, N.; Gress, R. E., An overview of IL-7 biology and its use in immunotherapy. *J Immunotoxicol* **2010**, *7*, 1-7.

34. Lundstrom, W.; Fewkes, N. M.; Mackall, C. L. IL-7 in human health and disease. *Semin. Immunol.* **2012**, *24*, 218-224.

35. Habibi, M.; Kmieciak, M.; Graham, L.; Morales, J. K.; Bear, H. D.; Manjili, M. H. Radiofrequency thermal ablation of breast tumors combined with intralesional administration of IL-7 and IL-15 augments anti-tumor immune responses and inhibits tumor development and metastasis. *Breast Cancer Res. Treat.* **2009**, *114*, 423-431.

36. Pellegrini, M.; Calzascia, T.; Elford, A. R.; Shahinian, A.; Lin, A. E.; Dissanayake, D.; Dhanji, S.; Nguyen, L. T.; Gronski, M. A.; Morre, M.; Assouline, B.; Lahl, K.; Sparwasser, T.; Ohashi, P. S.; Mak, T. W. Adjuvant IL-7 antagonizes multiple cellular and molecular inhibitory networks to enhance immunotherapies. *Nat. Med.* **2009**, *15*, 819-819.

37. Mackall, C. L.; Fry, T. J.; Gress, R. E. Harnessing the biology of IL-7 for therapeutic application. *Nat. Rev. Immunol.* **2011**, *11*, 330-342.

38. Gershoni, J. M.; Roitburd-Berman, A.; Siman-Tov, D. D.; Tarnovitski Freund, N.; Weiss, Y. Epitope mapping: the first step in developing epitope-based vaccines. *BioDrugs* **2007**, *21*, 145-56.

39. McElroy, C. A.; Dohm, J. A.; Walsh, S. T. R. Structural and Biophysical Studies of the Human IL-7/IL-7R alpha Complex. *Structure* **2009**, *17*, 54-65.

40. Liu, W. H.; Roemer, S. C.; Zhou, Y.; Shen, Z.-J.; Dennehey, B. K.; Balsbaugh, J. L.; Liddle, J. C.; Nemkov, T.; Ahn, N. G.; Hansen, K. C. The Cac1 subunit of histone chaperone CAF-1 organizes CAF-1-H3/H4 architecture and tetramerizes histones. *Elife* **2016**, *5*, e18023.

41. Mysling, S.; Kristensen, K. K.; Larsson, M.; Beigneux, A. P.; Gårdsvoll, H.; Fong, L. G.; Bensadouen, A.; Jørgensen, T. J.; Young, S. G.; Ploug, M. The acidic domain of the endothelial membrane protein GPIHBP1 stabilizes lipoprotein lipase activity by preventing unfolding of its catalytic domain. *Elife* **2016**, *5*, e12095.

42. Zheng, J.; Corzo, C.; Chang, M. R.; Shang, J.; Lam, V. Q.; Brust, R.; Blayo, A.-L.; Bruning, J. B.; Kamenecka, T. M.; Kojetin, D. J. Chemical crosslinking mass spectrometry reveals the conformational landscape of the activation helix of PPARγ; a model for ligand-dependent antagonism. *Structure* **2018**, *26*, 1431-1439.

43. Rimmer, M. A.; Nadeau, O. W.; Artigues, A.; Carlson, G. M. Structural characterization of the catalytic γ and regulatory β subunits of phosphorylase kinase in the context of the hexadecameric enzyme complex. *Protein Sci.* **2018**, *27*, 485-497.

44. Bhat, J. Y.; Miličić, G.; Thieulin-Pardo, G.; Bracher, A.; Maxwell, A.; Ciniawsky, S.; Mueller-Cajar, O.; Engen, J. R.; Hartl, F. U.; Wendler, P. Mechanism of enzyme repair by the AAA+ chaperone Rubisco activase. *Mol. Cell* **2017**, *67*, 744-756. e6.

45. Lu, S.; Fan, S. B.; Yang, B.; Li, Y. X.; Meng, J. M.; Wu, L.; Li, P.; Zhang, K.; Zhang, M. J.; Fu, Y.; Luo, J. C.; Sun, R. X.; He, S. M.; Dong, M. Q. Mapping native disulfide bonds at a proteome scale. *Nat. Methods* **2015**, *12*, 329-U73.

46. Bender, B. J.; Cisneros III, A.; Duran, A. M.; Finn, J. A.; Fu, D.; Lokits, A. D.; Mueller, B. K.; Sangha, A. K.; Sauer, M. F.; Sevy, A. M. Protocols for molecular modeling with Rosetta3 and RosettaScripts. *Biochemistry* **2016**, *55*, 4748-4763.

47. Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **1997**, *268*, 209-225.

48. Simons, K. T.; Ruczinski, I.; Kooperberg, C.; Fox, B. A.; Bystroff, C.; Baker, D. Improved recognition of native‐like protein structures using a combination of sequence‐dependent and sequence‐independent features of proteins. *Proteins* **1999**, *34*, 82-95.

49. Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003**, *331*, 281-299.

50. Chaudhury, S.; Berrondo, M.; Weitzner, B. D.; Muthu, P.; Bergman, H.; Gray, J. J. Benchmarking and analysis of protein docking performance in Rosetta v3. 2. *PloS one* **2011**, *6*, e22477.

51. McElroy, C. A.; Holland, P. J.; Zhao, P.; Lim, J. M.; Wells, L.; Eisenstein, E.; Walsh, S. T. Structural reorganization of the interleukin-7 signaling complex. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 2503-2508.

52. Leitner, A.; Joachimiak, L. A.; Unverdorben, P.; Walzthoeni, T.; Frydman, J.; Forster, F.; Aebersold, R. Chemical crosslinking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9455-9460.

53. Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **2016**, *41*, 20-32.

Chapter 5: Epitope and Paratope Mapping of PD-1/Nivolumab by Mass Spectrometry-Based Hydrogen/Deuterium Exchange, Crosslinking, and Molecular Docking*

* This chapter is based on the following publication: Zhang, M. M.; Huang, R. Y-C.; Beno, B. R.; Deyanova, E. G.; Li, J.; Chen, G.; Gross, M. L. Epitope and Paratope Mapping of PD-1/Nivolumab by Mass Spectrometry-based Hydrogen/Deuterium Exchange, Crosslinking, and Molecular Docking *Anal. Chem.* **2020,** ASAP.

5.1 Abstract

Programmed cell death-1 (PD-1), an antigen co-receptor on cell surfaces, is one of the conspicuous immune checkpoints. Nivolumab, a monoclonal antibody therapeutic approved by FDA, binds to PD-1 and efficiently blocks its pathways. In this study, an integrated approach has been developed to map binding epitope/paratope of PD-1/Nivolumb, including hydrogen deuterium exchange mass spectrometry (HDX-MS) followed by electron-transfer dissociation (ETD), chemical crosslinking and molecular docking. HDX-ETD offers binding sites characterization with amino acid resolution. Chemical crosslinking reveals complementary information on one additional epitope (i.e., the BC-loop) and a potential paratope at the N-terminus of the heavy chain. Furthermore, crosslinking identifies another loop region (i.e., the C'D-loop) that is subject to remote conformational change. The distance restraints derived from the cross-links enable building high-confidence models of PD-1/Nivolumab, evaluated with respect to a resolved crystal structure. This integrated strategy provides an opportunity for comprehensive characterization of other antigen/antibody interactions that would assist in understanding binding mechanisms and designing antibody therapeutics.

5.2 Introduction

Antibodies are key biosensors in the immune system that can neutralize antigens and evoke other biomolecules that fight pathogens.[1] The binding between epitopes and paratopes is exquisitely specific and of high affinity, contributing to numerous applications in biological research, diagnostics, and therapy.[2] Comprehensive description of the epitopes/paratopes, ideally to the residue level, is crucial to understand the binding mechanism and to design future therapeutic agents. Hydrogen deuterium exchange (HDX) coupled with mass spectrometry (MS), an

approach that reflects the local solvent accessible surface area (SASA) and H-bond network of the protein backbone, is a valuable tool for probing protein interfaces.[3-9] Its advantages are the near-native conditions of the experiment, low sample amount and high throughput compared to X-ray crystallography. A major limitation, however, can be the coarse spatial resolution limited by the length of proteolytic peptides generated in the HDX experiment.[10] Besides proteolyzing the protein to smaller fragment peptides, another solution to increase the spatial resolution is electron transfer dissociation (ETD), a fragmentation technique that can locate deuterium on one or a few residues. It utilizes a transferred electron from a radical anion to fragment peptides or protein with minimal scrambling of the amide H and D, in contrast to collision-induced fragmentation that uses many low-energy collisions to induce fragmentation.[11-14] Another potential disadvantage of HDX-MS is the inability to distinguish between the direct binding interaction and remote conformational or allosteric effects. A combination of other complementary methods may overcome this limitation.

Mass spectrometry-based chemical crosslinking (XL-MS) has developed rapidly owing to the increased availability of diverse cross-linkers, advanced analysis software, and improvements in sample handling.[15-17] Observed cross-links deliver information about not only the connectivity of adjacent protein subunits but also the distance ranges between specific amino acid residues as defined by the spacing between functional groups in chosen crosslinking reagents. These features contribute to a wide range of successful applications, including structural elucidation of single proteins[18], topological portrayal of large macromolecular assemblies [19-20], and interaction maps of an entire proteome [20-21]. Mapping epitopes/paratopes by using XL-MS, however, is an underutilized opportunity.[22] In this study, we incorporated XL-MS together with HDX and

HDX-ETD to illustrate an analytical approach for epitope/paratope mapping of an important antigen/antibody system.

Programmed cell death-1 (PD-1)[23], an immune checkpoint, is an antigen-independent co-receptor, located on cell surfaces and expressed predominantly by T-cells.[24] The critical role of PD-1 is to bind with specific ligands, PD-1 ligand 1 (PD-L1)[25] and PD-1 ligand 2 (PD-L2)[26], to maintain immune tolerance by suppressing self-reactive T-cells[27] and preventing pathogenic autoimmunity. The signaling, however, can be utilized by tumor cells to escape immune surveillance.[28-29] Therefore, blockage of the PD-1 pathway has been an appealing target in recent development of immuno-therapeutics.[30-32] Nivolumab , one of two monoclonal antibodies (mAbs) on the market[33], is designed to bind with PD-1, demonstrating immune restoration in multiple tumor conditions[34-36] with impressive clinical efficacy.

Here, we applied HDX-MS to the PD-1/Nivolumab complex to obtain regional binding information, which was further refined by HDX-ETD to specify more closely the critical binding residues. The suggested epitope and paratope regions were subsequently evaluated by XL-MS, revealing complementary binding interfaces, and differentiating remote conformational changes. Utilizing the distance restraints derived from various cross-linkers, we conducted molecular docking to generate high-confidence 3D models and evaluated the strengths and limitations of this approach. A previously resolved X-ray crystal structure of PD-1/Nivolumab Fab [37-38] was employed for final comparison purposes. Besides the successful application in the binding interface determination with IL-7 and IL-7Rα (Chapter 5), the integration of several MS-based approaches also enables precise and detailed characterization of epitopes/paratopes of an antigen/antibody complex, encouraging even broader applications.

5.3 Experimental

5.3.1 Hydrogen-Deuterium Exchange Mass Spectrometry

PD1 (25 μM) and Nivolumab (25 μM), dissolved in non-deuterated PBS buffer (Phosphate Buffered Saline, pH 7.4), were digested individually with an immobilized Enzymate pepsin column, 300 Å, 5 μm, 2.1 mm X 30 mm (Waters Corp., Milford, MA, USA) under different conditions including PD1 alone, Nivo Fab alone, and bound PD1 and Nivo Fab at a molar ratio of 1:2, for selected peptic peptides. Prior to the HDX experiments, PD1 (25 μM) and the Nivo Fab (50 μM) were incubated for 1 h to allow complex formation. HDX was conducted on an HDX PAL robot (LEAP Technologies, Carrboro, NC), in which 5 μL of protein was diluted into 55 μL D_2O buffer (10 mM phosphate buffer, D_2O, pD 7.0) to initiate HDX. Different aliquots were submitted to HDX for several times: 0.33, 1.0, 10, and 240 min. The reaction mixture was then quenched by adding quenching buffer (4 M GdnCl and 0.4 M TCEP, pH 2.5, 1:1, v/v), and 50 μL of quenched sample was injected into a Waters nanoACQUITY UPLC HDX Manager™ system. The deuterated mixture was then digested online by using the same pepsin column at 20 °C for 3 min, and the resultant peptides were desalted with an ACQUITY UPLC BEH C18 VanGuard column (130 Å, 1.7 μm, 2.1 mm × 5 mm) with 0.1% formic acid in water at 100 μL/min for 3 min. The eluted peptides were further separated on a trap column (ACQUITY UPLC BEH C18, 130 Å, 1.7 μm, 1.0 mm × 50 mm), under a 7.5 min gradient: 8-85% acetonitrile in water with 0.1% formic acid at 65 μL/min. Peptides were introduced to a Waters Synapt G2si Q-TOF (Waters, MA) mass spectrometer for mass analysis. The instrument settings were capillary 3.5 kV; sampling cone 35 V; source temperature 80 °C; desolvation temperature 180 °C; and *m/z* range 260-2000.

In the HDX-ETD experiments, 1,3-dicyanobenzene was utilized as the electron carrier for gas-phase fragmentation of selected peptic peptide by using an isolation window of 6 (± 3 Da). The corresponding instrument settings were sampling cone: 20 V; isolation window: ± 3 Da. Experiments were all performed as biological triplicates.

5.3.2 HDX Data Analysis

Peptic peptides were identified through a combination of exact mass analysis and MSE by using ProteinLynx Global Server 3.0.2 (Waters Corp., Milford, MA, USA). Both peptide-level and residue-level (ETD experiments) deuterium uptakes were calculated by using Waters DynamX 3.0^{TM} software. Deuterium uptake differences were calculated from the centroid of the isotopic pattern of the deuterated and non-deuterated peptides ions.

5.3.3 Chemical Crosslinking

PD1 (15 μM, Bristol-Myers Squibb, NY, NY) and 30 μM Nivolumab Fab (Bristol-Myers Squibb, NY, NY) were incubated at room temperature for 40 min prior to initiating the crosslinking reactions. For a buffer, 10 mM PBS (7 mM KCl, 137 mM NaCl, pH = 7.4) was used for NHS-ester cross-linkers, and 50 mM MES buffer (15 mM KCl, pH = 6.5) for EDC crosslinking. For the BS^3-h_{12}/d_{12} (bis(sulfosuccinimidyl) suberate) (Creative Molecules, Canada) and BS^{2}G-$d_0/d4$ (*bis*(sulfosuccinimidyl) 2,2,4,4-glutarate-d_0/d_4) (Thermo Fisher Scientific, Carlsbad, CA) crosslinking experiments, fresh BS^3-h_{12}/d_{12} or BS^{2}G-d_0/d_4 solution were added to the protein mixture to a concentration of 0.75 mM, 1.5 mM and 3 mM, respectively. The solution was incubated at 25 °C with spinning at 500 rpm for 45 min. Tris-HCl (1 μL of 1 M, pH = 7.4) was used to quench the crosslinking reaction after a 15-min incubation. For the EDC (Thermo Fisher Scientific, Carlsbad, CA) crosslinking, EDC and sulfo-NHS (*N*-hydroxysulfosuccinimide sodium salt, Millipore Sigma, St. Louis, MO) were dissolved at a 1:1 molar ratio in 200 mM

MES buffer just before mixing with the protein solution. The reaction was run at 25 °C for 45 min. The EDC/sulfo-NHS concentrations were 5, 10, and 30 mM for different trials. Tris-HCl was added to a final concentration of 200 mM and incubated for 15 min to terminate the crosslinking chemistry.

5.3.4 Gel Electrophoresis

Bis-tris-Precast gels (4-12% SurePAGE) and MES running buffer were purchased from GenScript (Piscataway, NJ). Colloidal Blue Staining kit and the protein marker were from Thermo Fisher Scientific (Carlsbad, CA). For each gel band, approximately 100-200 pmol of each protein was obtained and loaded. The operating voltage was 200 V for 35 mins.

5.3.5 Enzymatic in-solution Digestion

A Zeba column (Thermo Fisher Scientific, Carlsbad, CA) was used to remove access crosslinking reagents. After desalting, a final concentration of 0.25% (w/v) RapiGest (Waters, Milford, MA) was constituted in the reaction mixture and incubated at 80 °C for 30 min for denaturation. The reduction and alkylation of disulfide bonds were conducted with 7.5 mM DTT at 55 °C for 30 min and 14 mM IAM (Millipore Sigma, St. Louis, MO) in the dark for 30 min. Trypsin (Promega, Madison, WI) digestion (10:1), together with PNGase F (glycerol-free) (BioLabs, Ipswich, MA), was performed at 37 °C for 2h, followed by additional chymotrypsin (Promega, Madison, WI) digestion for 4 h to get the best coverage. The peptide mixture so generated was treated with C18 NuTip, the solvent evaporated, and the protein reconstituted in 0.1% formic acid after vacuum.

5.3.6 LC-MS/MS Analysis

The digested peptides were desalted with a C18 reversed-phase column (Acclaim PepMap C18, 100 µm x 2 cm, 5 µm, 100 Å; Thermo Fisher Scientific, Carlsbad, CA) for 10 mins at 4 µL/min.

The resulting peptides were separated on a custom-packed C18 column (Magic Beads, 100 µm x 180 mm, 5 µm, 120 Å; Michrom Bioresources, Auburn, CA) mounted on the Ultimate 3000 Rapid Separation system (Dionex). The HPLC solvent consists of phase A (0.1% formic acid in water) and phase B (0.1% formic acid in 80% acetonitrile). The flowrate was operated at 400 nL/min with a 120-min gradient (2.5% B to 12% B in 30 min, increased to 60% B for 65 min, ramped to 85% B in 5 min, kept at 85% B for 5 min, returned to 2.5% B in 5 min, and equilibrated at 2.5% B for 10 min). A Thermo Q Exactive Plus orbitrap mass spectrometer was used for LC/MS/MS analysis of the digested, cross-linked sample. For isotopic-encoded crosslinking runs, the data-independent mode was set for per charge state with the associated mass tag.

5.3.7 Identification of Cross-links

LC-MS/MS files were submitted to pLink [39-40] (ver. 2.3.1, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China) for cross-link identification. Search parameters were: enzyme: trypsin/chymotrypsin; missed cleavage: 3; precursor tolerance: 10 ppm; fragment tolerance: 30 ppm; variable modification: oxidation of M, deamidation of N, Q and the N-terminus; fixed modification: carbamidomethyl of C; minimal peptide length: 6 aa; maximal peptide length: 60 aa; minimal peptide MW: 600 Da; maximal peptide mass: 6000 Da. The false discovery rate was 5% with 10 ppm MS1 filter tolerance. Isotopic doublets in the mass spectrum was searched and confirmed in the raw file. MS/MS peak assignments were further validated manually on the basis of the product ion mass list.

5.3.8 Molecular Docking with Cross-link Derived Restraints

The Rosetta (v. 3.8) [41-43] docking_protocol (RosettaDock) program [44-45] running on multi-processor Linux servers was used for the protein-protein docking. Protein models for PD1 and

Nivolumab Fab were obtained from the X-ray structure of the PD1/Nivolumab Fab complex (PDB ID: 5WT9) [37]. Coordinates for the Nivolumab Fab chains H and L and PD1 were extracted from the PDB file of the complex, written out as two separate PDB formatted files (Nivolumab Fab and PD1) and then prepared for docking in a two-stage process. First, residue atoms that were not present in the PDB files and hydrogen atoms were added, and non-protein residues were removed with the Rosetta score_jd2 program. Subsequently, energy minimization of the structures output by score_jd2 within the context of the Rosetta energy model was performed with the Rosetta relax program. Sets of 10 models were generated for each structure by using the "constrain_relax_to_start_coords" and "coord_constrain_sidechains" options to limit structural changes, and the model of each protein with the lowest Rosetta total energy score was selected for protein-protein docking.

The processed PD1 and Nivolumab Fab models were imported into Maestro, [46] and the PD1 model was manually displaced and rotated arbitrarily relative to the Nivolumab Fab. Both protein models were merged into a single, three-chain model that was exported to a PDB format file for the protein-protein docking computations.

Docking exercises were performed with the RosettaDock program using eight distance restraints derived from the PD1/Nivolumab Fab chemical cross-links listed in Table 1. The distance restraints were implemented as flat-bottomed harmonic potentials that penalized models where the Cα-Cα distance(s) for crosslinked residues fell outside of the ranges of 6 – 16 Å[47] and 9 – 30 Å[48] for the EDC and BS^2G or BS^3 crosslinks, respectively. For the restrained docking study, 250 runs were performed, each yielding up to 400 models for a maximum of 100,000 models. For comparison purposes, 250 docking runs, each yielding 400 models (100,000 total) were also carried out without restraints. Following completion of the docking calculations, the 100 models

from the restrained docking exercise with the most favorable values of the RosettaDock total_score were identified and used to generate Figure 6. The best scoring 100 models from the non-restrained docking exercise were also identified and compared to the X-ray structure of the PD1/Nivolumab Fab complex. A Python/OEChem[49] script was used to determine the Cα atom-based r.m.s.d. for PD1 in each clustered model relative to the X-ray crystal structure of the PD1/Nivolumab Fab complex (PDB ID: 5WT9) and to analyze models of PD1 docked to Nivolumab Fab.

5.4 Results and Discussion

5.4.1 Epitope and Paratope Mapping by HDX

To map the epitope on PD-1, we performed HDX experiments with unbound PD-1 and PD-1 bound to the antigen-binding fragment of Nivolumab (Nivo Fab) at a molar ratio of 1:2. Deuterium uptake was monitored at 0.33, 1.0, 10, and 240 min on 19 unique peptides, covering 85% of PD-1 sequence. Information on 15% of the sequence was lost likely because the antigen contains complex N-linked glycans at those sites that hamper peptide chromatography and identification. Accumulative deuterium uptake differences across the four time points were calculated (Figure 5.1A), revealing three regions that undergo protection: ^{25}LDSPDRPWNPPTFSPALL^{42}, ^{80}AAFPEDRSQPGQDCRF^{95} and ^{125}AISLAPKAQIKESL^{138}. Specifically, regions ^{125}AISLAPKAQIKESL^{138}, located on the FG-loop of the PD-1 structure (Figure 5.1B), and ^{25}LDSPDRPWNPPTFSPALL^{42}, part of the N-loop (Figure 5.1B), exhibit the most significant decrease in HDX upon binding to Nivo Fab with protection corresponding to decreases of 7 and 4 Da, respectively. Region ^{80}AAFPEDRSQPGQDCRF^{95} containing the C'D-loop (Figure 5.1B) showed less protection corresponding to a decrease of approximately 2 Da. The deuterium uptake differences were similar across all time periods for the three regions

(Figure 5.1C), suggesting stable solvent protection, strong bonding, and associated small off rates in the equilibrium.

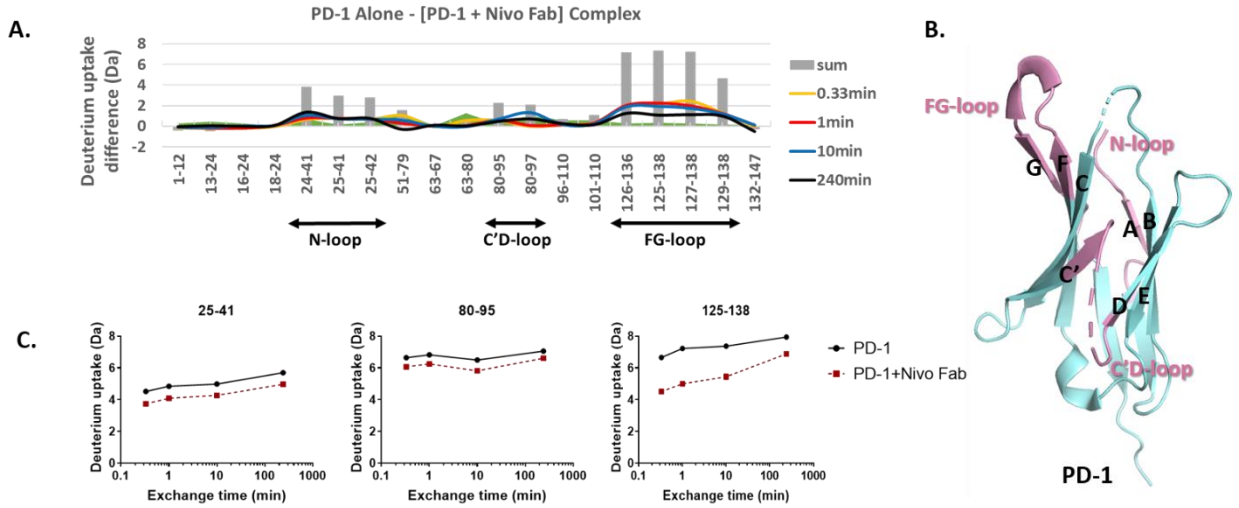


Figure 5.1 Epitope regions on PD-1 indicated by HDX. (A) Differential HDX kinetics plots of PD-1 and PD-1/Nivolumab Fab complex. (B) Epitope regions mapped onto the PD-1 crystal structure (PDB: 3RRQ). (C) HDX kinetics of three peptides corresponding to the epitope regions in PD-1. Unbound PD-1 is colored in black and bound PD-1/Nivolumab Fab is colored in burgundy. The green shaded region is the propagated error across all time points. HDX kinetics results for the remaining peptides are shown in Figure 5.2.

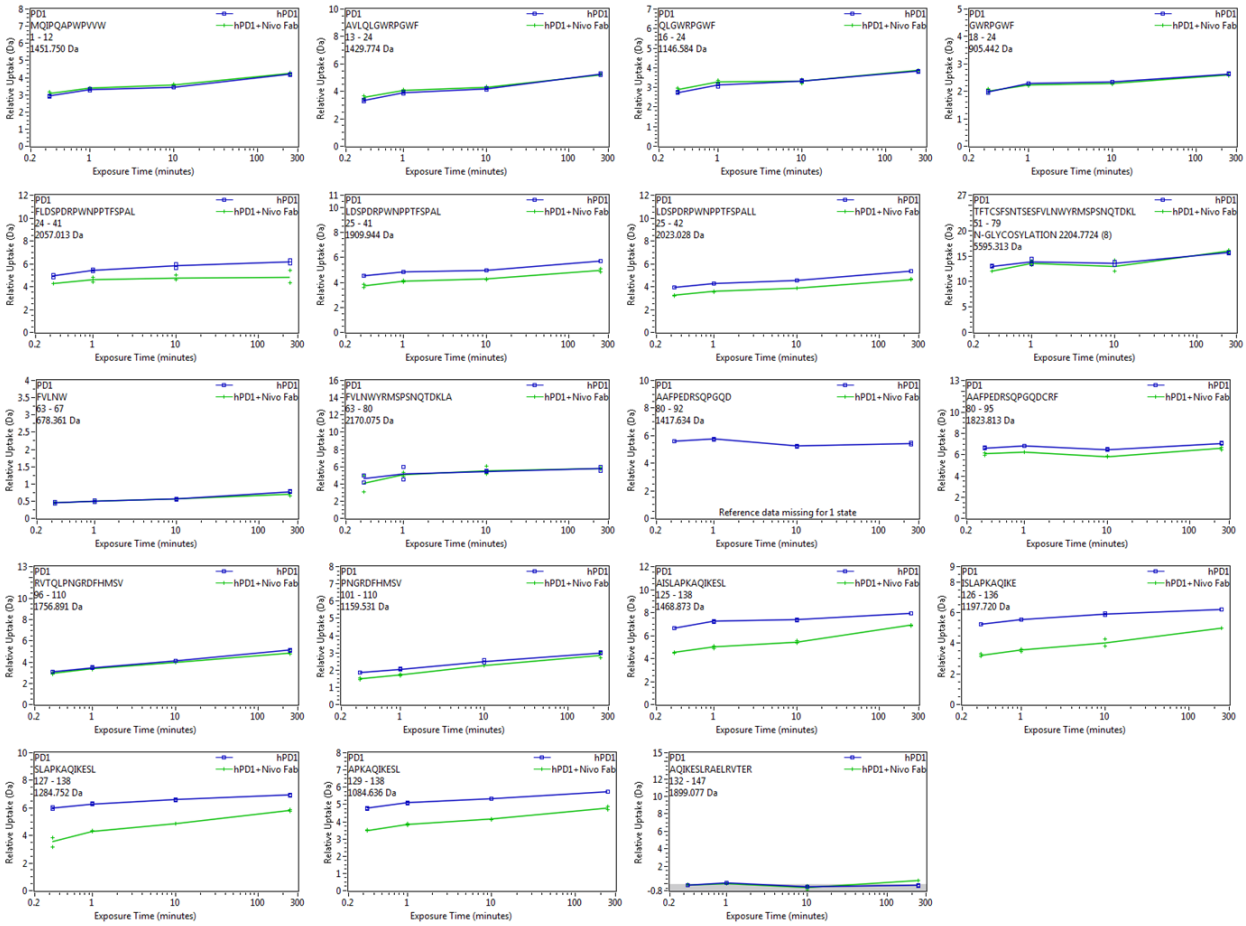


Figure 5.2 HDX kinetic plots of other peptides in PD-1. Unbound PD-1 is colored in blue and bound PD-1/Nivo Fab is colored in green.

We also performed HDX experiments with PD-1 and full-length Nivolumab (Nivo mAb) for comparison (Figure 5.3). One of the peptides, ^{132}KAQIKESLRA ELRVTE^{147}, was not found in PD-1-Nivo mAb complexes, possibly due to the larger number of peptides when using the full Nivo mAb. The larger number of peptides from full-length mAb hampers the acquisition and identification of PD-1 peptides in the fast chromatography used in HDX. The other peptide regions, on the other hand, showed consistent and similar trends regarding the deuterium uptake differences for bound and unbound states, indicating comparable binding behaviors between the full Nivolumab and its Fab.

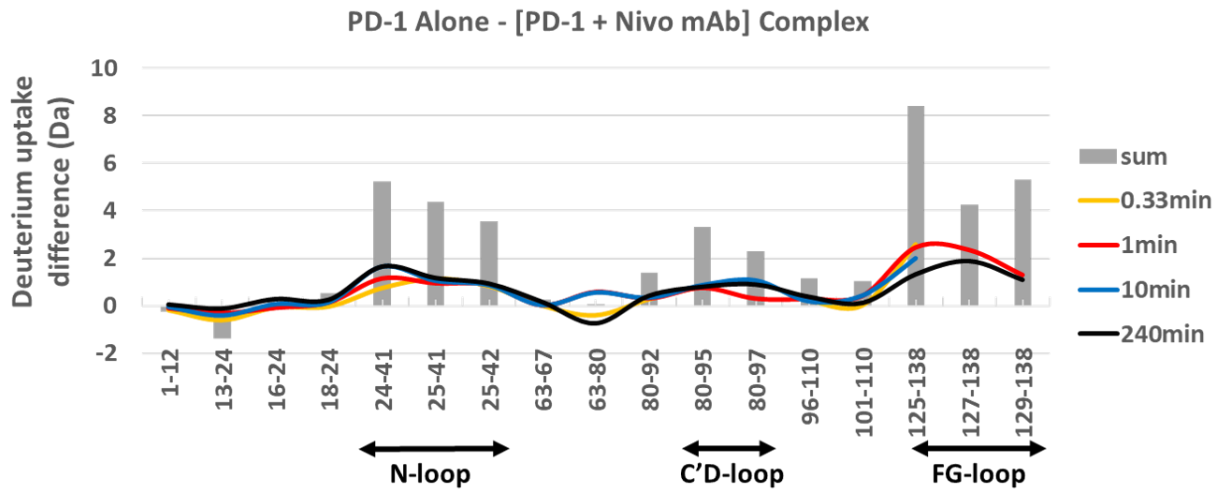


Figure 5.3 Differential HDX kinetics plots of PD-1 and PD-1/Nivo mAb complex.

To map the paratope on Nivolumab, we conducted a series of similar HDX experiments. Non-bound and bound Nivo Fab (PD-1: Nivo Fab at a molar ratio of 2:1) were exchanged with deuterated buffer from 0.33 min to 4 h. The accumulative deuterium uptake differences between the two states confirm the involvement of the CDR regions in PD-1 binding (Figure 5.4). Peptide regions covering CDR-H2 on the heavy chain (Figure 5.4A) and CDR-L3 on the light chain (Figure 5.4B) showed the greatest protection, corresponding to more than 8 Da and a constant difference of HDX as a function of time (Figure 5.4C). Peptides covering CDR-H1, CDR-H3 and CDR-L2 showed smaller changes in HDX upon binding (~ 6 Da for the first and ~ 3 Da for the latter two). The HDX of peptides covering ^{33}SSYLAWYQQKPGQA^{46} exhibited only 1 Da deuterium uptake difference across all HDX time points, and the difference mainly occurred at the longer HDX time (Figure 5.4C). Given that this region only partially overlaps with CDR-L1, ^{27}RASQSVSSYLA^{37}, few amino acids in this region are involved in PD-1 binding, resulting in small differences in HDX. The HDX paratope of Nivolumab, in general, is in accord with what is commonly viewed as CDR regions.

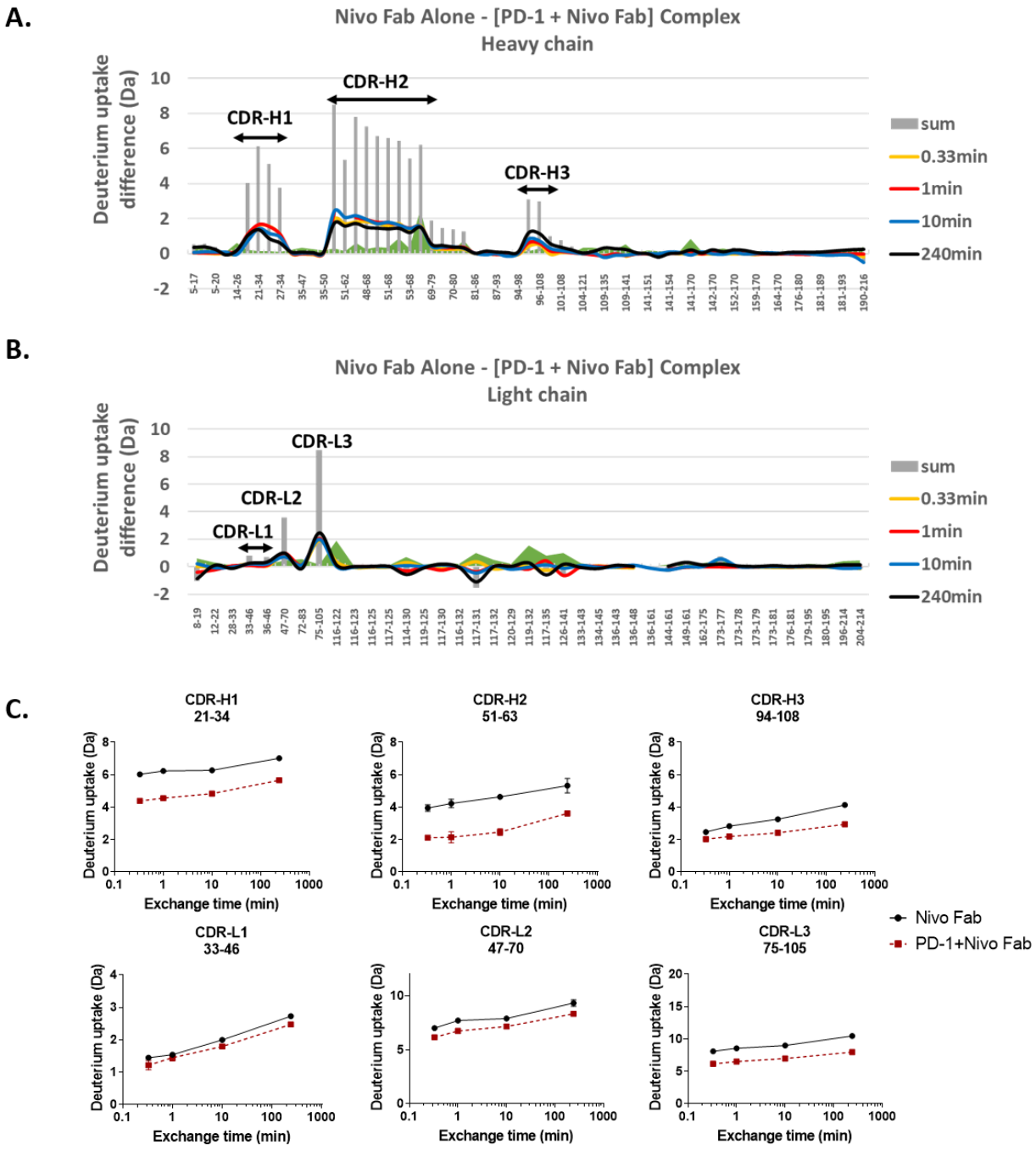
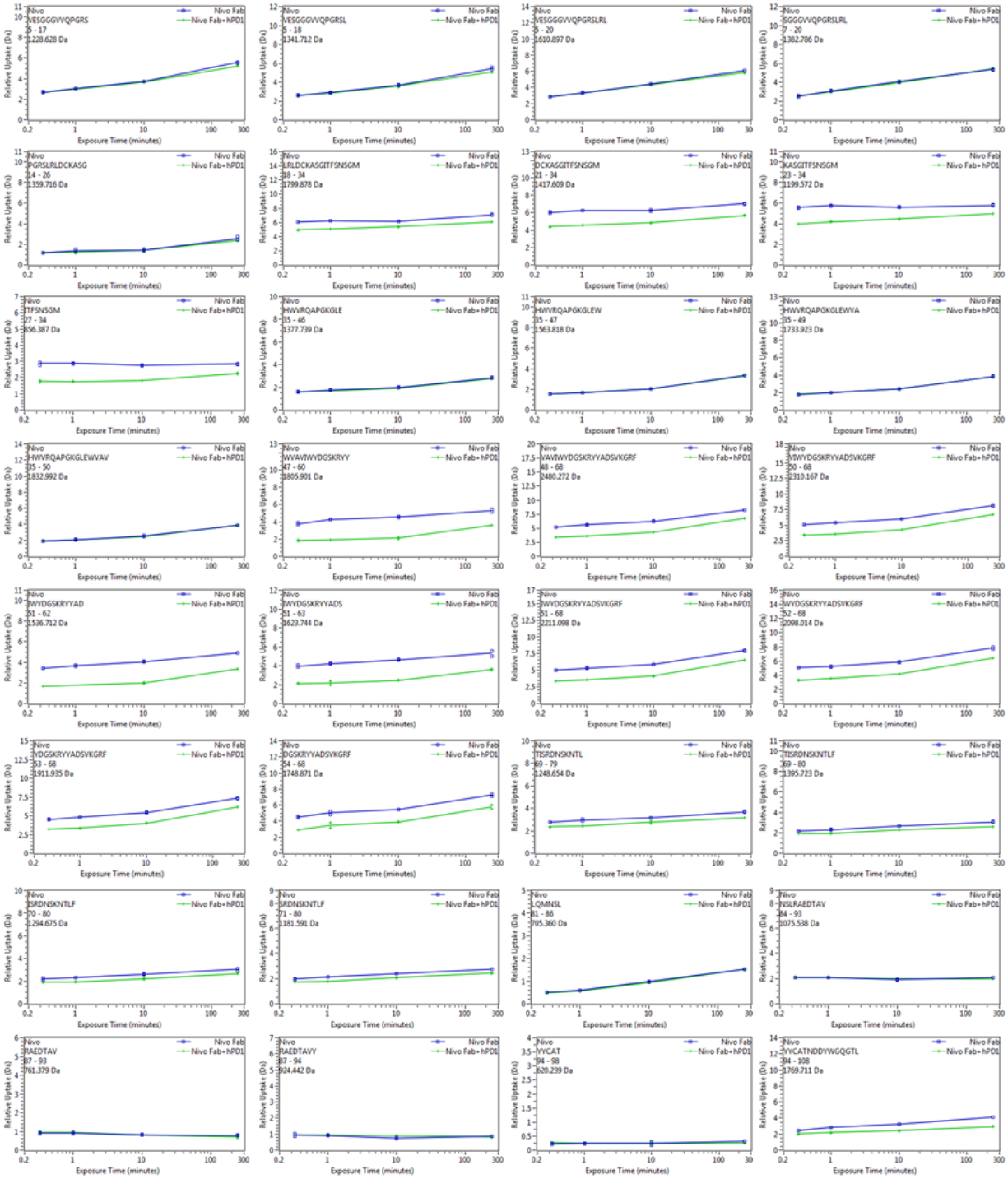


Figure 5.4 Paratope regions on Nivo Fab as determined by HDX. Differential HDX kinetics plots of Nivo Fab vs. PD-1/Nivo Fab complex for (A) heavy chain and (B) light chain, respectively. (C) HDX kinetics of the peptides in the corresponding paratope regions in Nivo Fab (unbound PD-1 is colored in black and bound PD-1/Nivo Fab is in burgundy. The green shaded region is the propagated error across all time points. HDX kinetics plots of the other peptides are shown in Figure 5.5.


Nivo
Nivo Fab+hPD1
Relative Uptake (Da)
Exposure Time (minutes)
VESGGGVVQPGRS
5 - 17
1228.628 Da
VESGGGVVQPGRSL
5 - 18
1341.712 Da
VESGGGVVQPGRSLRL
5 - 20
1610.897 Da
SGGGVVQPGRSLRL
7 - 20
1382.786 Da
PGRSLRLDCKASG
14 - 26
1359.716 Da
LRLDCKASGITFSNSGM
18 - 34
1799.878 Da
DCKASGITFSNSGM
21 - 34
1417.609 Da
KASGITFSNSGM
23 - 34
1199.572 Da
TFSNSGM
27 - 34
856.387 Da
HWVRQAPGKGLE
35 - 46
1377.739 Da
HWVRQAPGKGLEW
35 - 47
1563.818 Da
HWVRQAPGKGLEWVA
35 - 49
1733.923 Da
HWVRQAPGKGLEWVAV
35 - 50
1832.992 Da
WVAVIWYDGSKRYY
47 - 60
1805.901 Da
VAVIWYDGSKRYYADSVKGRF
48 - 68
2480.272 Da
VWYDGSKRYYADSVKGRF
50 - 68
2310.167 Da
WYDGSKRYYAD
51 - 62
1536.712 Da
WYDGSKRYYADS
51 - 63
1623.744 Da
WYDGSKRYYADSVKGRF
51 - 68
2211.098 Da
WYDGSKRYYADSVKGRF
52 - 68
2098.014 Da
YDGSKRYYADSVKGRF
53 - 68
1911.935 Da
DGSKRYYADSVKGRF
54 - 68
1748.871 Da
TISRDNSKNTL
69 - 79
1248.654 Da
TISRDNSKNTLF
69 - 80
1395.723 Da
SRDNSKNTLF
70 - 80
1294.675 Da
SRDNSKNTLF
71 - 80
1181.591 Da
LQMNSL
81 - 86
705.360 Da
NSLRAEDTAV
84 - 93
1075.538 Da
RAEDTAV
87 - 93
761.379 Da
RAEDTAVY
87 - 94
924.442 Da
YYCAT
94 - 98
620.239 Da
YYCATNDDYWGQGTL
94 - 108
1769.711 Da

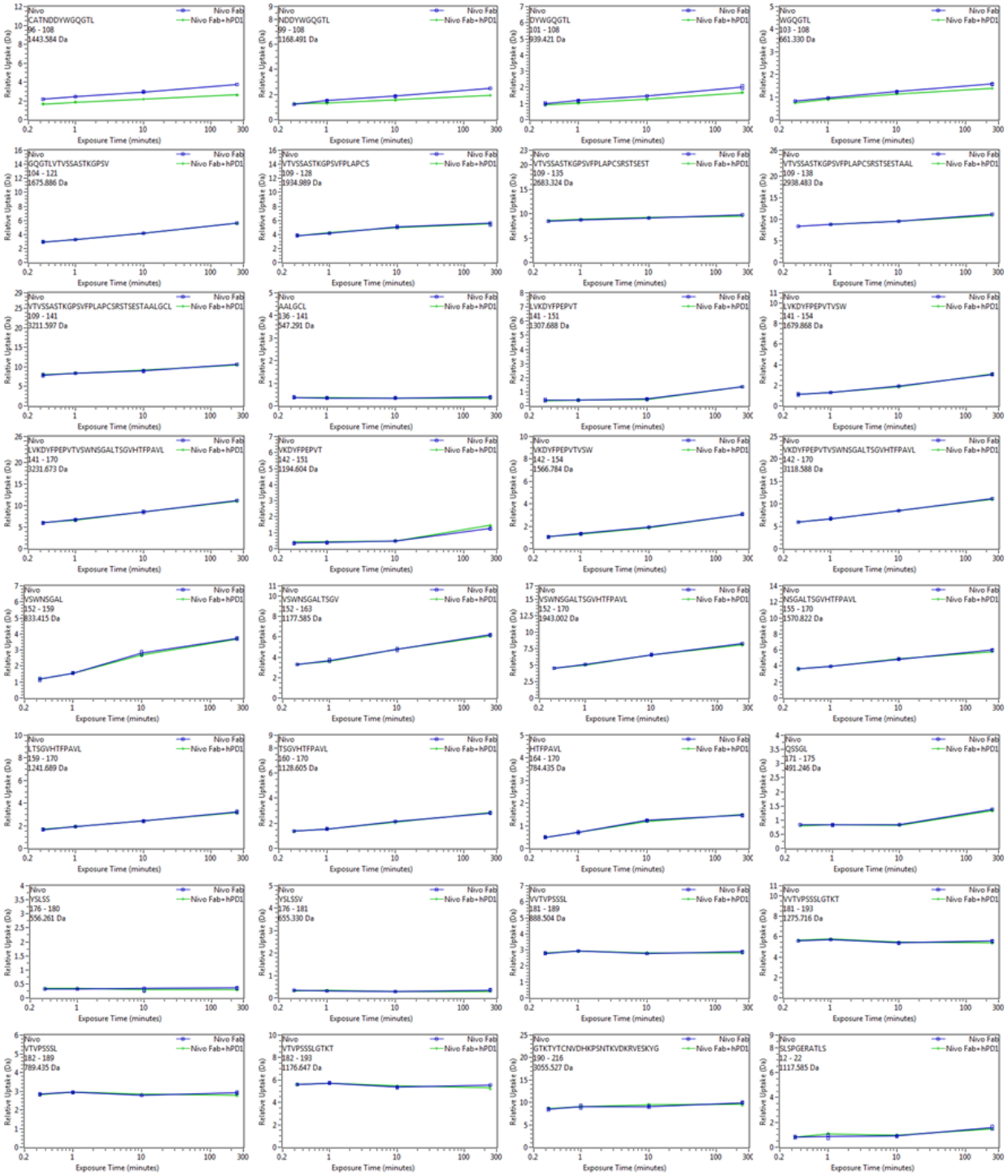


Figure 5.5A HDX kinetic plots of the peptides on the heavy chain of Nivo Fab. Unbound state is colored in blue and bound PD-1/Nivo Fab is colored in green.

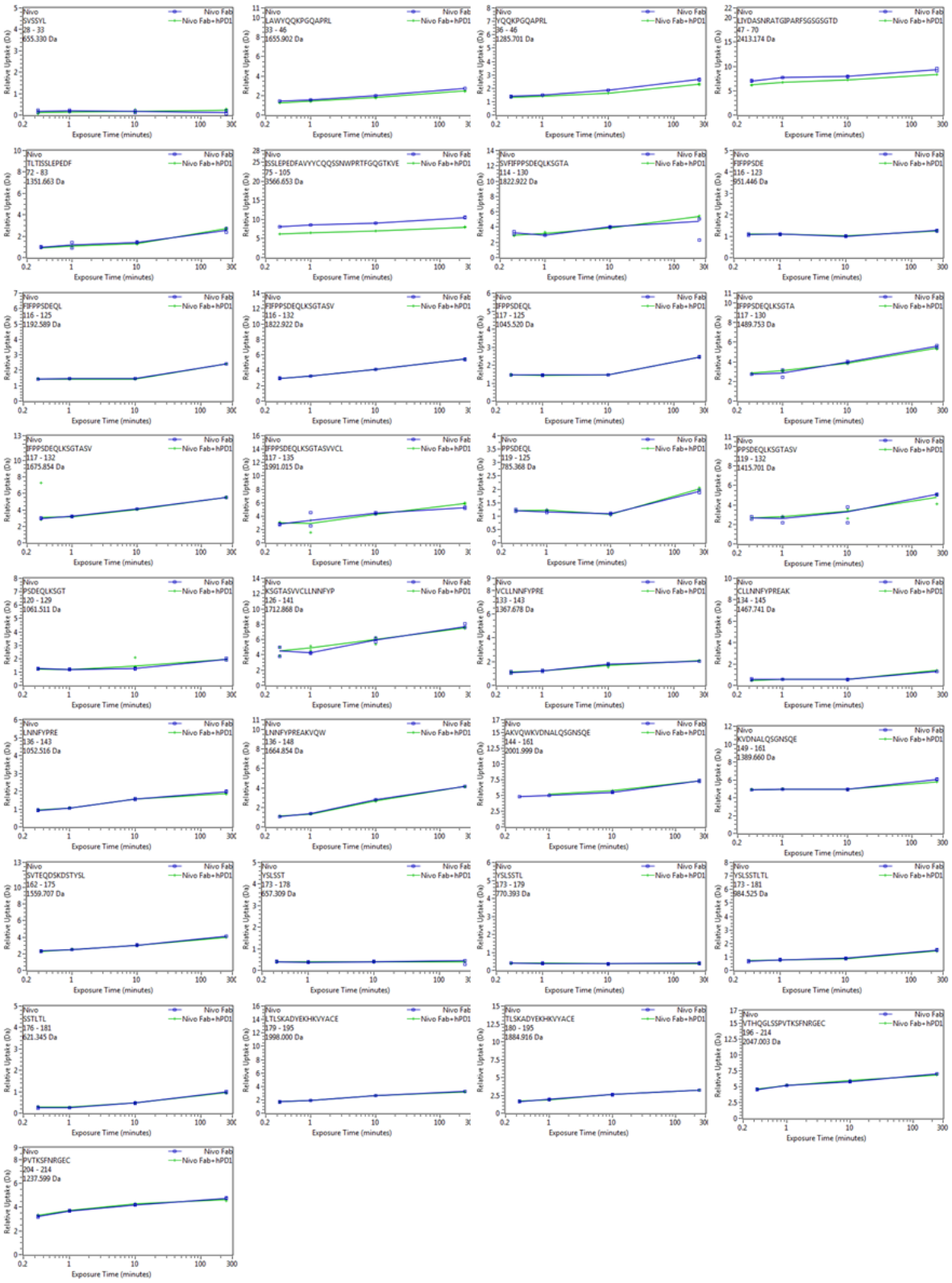


Figure 5.5B HDX kinetic plots of the peptides on the light chain of Nivo Fab. Unbound state is colored in blue and bound PD-1/Nivo Fab is colored in green.

5.4.2 Epitope Refinement by HDX-ETD

Epitope identification by HDX is limited by the size of peptic peptides and by overlapping peptides, providing only regional information. To increase the spatial resolution, we coupled ETD fragmentation with HDX-MS to refine further the protected regions. To set up the ETD measurements, we used a previously published procedure[11] to measure the extent of H/D scrambling of a synthetic peptide, HHHHHHIIKIIK, under several conditions and selected those that show minimal scrambling. In addition, the bound-state PD-1 was achieved with Nivo Fab to produce fewer peptides, lower complexity in the separation step of HDX than that with full-length Nivo mAb, and increased the signal-to-noise ratios for peptide peaks. The incubation time for deuterium labeling was controlled as 1 min, a time point that gives distinct differences in the peptide-level HDX.

We submitted to ETD the three peptides (Figure 5.1C) that cover the epitope regions identified by the HDX kinetics. Not all the peptides could be successfully resolved owing to their nature (e.g., residue composition) and to incomplete fragmentation of the peptide at their available charge states. For example, the doubly charged peptide ^{25}LDSPDRPWNPPTFSPALL^{42}, part of the N-loop, suffers from multiple proline residues in the sequence, showing a limited number of product ions. The peptide ^{80}AAFPEDRSQPGQDCRF^{95} containing the C'D loop has only a moderate difference in HDX between bound and unbound (~ 0.5 Da at 1 min), and that makes it difficult to measure differences in HDX given the large size of the peptide (16 residues) and the experimental error associated with HDX-ETD. On the other hand, a triply charged peptide ^{125}AISLAPKAQIKESL^{138}, located on the FG-loop, successfully produced a series of C fragment ions upon ETD, allowing further epitope refinement for this region.

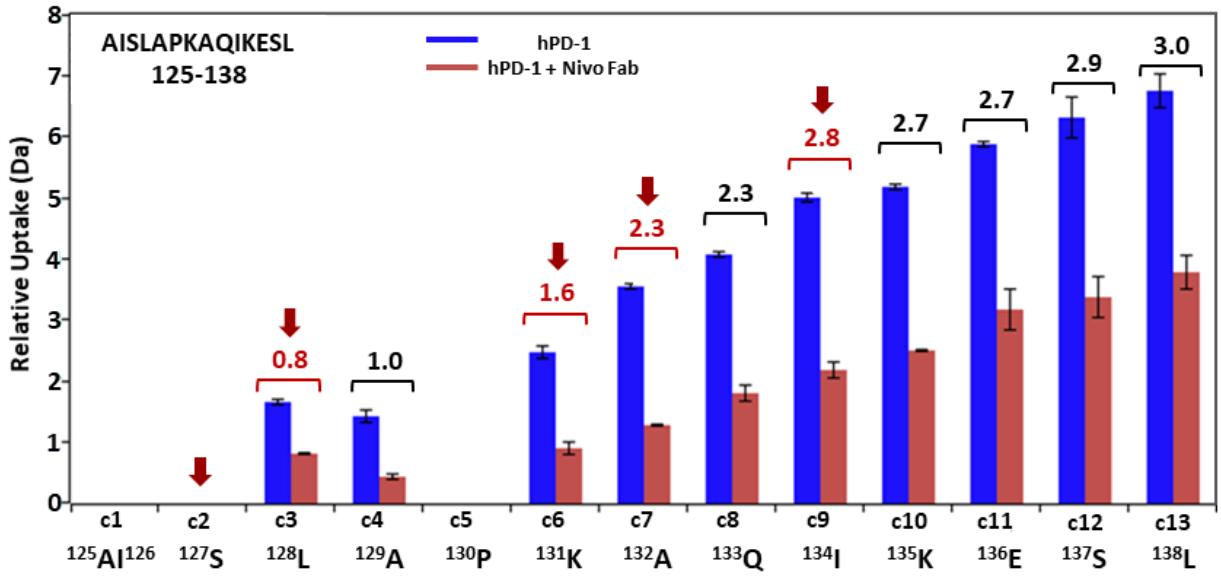


Figure 5.6 Cumulative deuterium uptake plot for C-type ions of ^{125}AISLAPKAQIKESL^{138} in PD-1 and PD-1/Nivo Fab complex by HDX-ETD. Deuterium uptake differences are calculated and labeled for each C-ion. The potential binding residues are indicated with arrows and colored in red.

We plotted the cumulative deuterium uptake for all C-ions and calculated the deuterium uptake difference for each ion between the bound and unbound states of PD-1 (Figure 5.6: note that the deuterium uptake of each C_n-ion represents that occurring on the n + 1 residue). The absence of C_1 and C_2 ions is likely an undesirable consequence of the ionization condition optimization in which we balanced the intensity of the fragment ions and the extent of deuterium scrambling. The deuterium uptake measured on the C_3 ion was reduced (~ 0.8 Da) upon complex formation. Given that deuterons on the first two residues are almost always lost owing to back exchange[50] and the C_2 ion was not resolved in the experiment, protection of residue ^{127}S or ^{128}L or both could account for the observed decrease in deuterium uptake on C_3. Further increases in HDX protection were observed at C_6, showing 1.6 Da reduction in the PD-1/Nivo Fab complex compared to the unbound PD-1; the considerable drop in deuterium uptake pinpoints the protected residue ^{131}K, given the HDX-silent ^{130}P and the similar deuterium uptake difference of

C_4 comparing to that of C_3. In addition, the C_7 ion exhibits additional HDX protection, increasing from 1.6 at C_6 to ~ 2.3 Da. The HDX for C9 shows an additional uptake in protection as at C7, suggesting protection on residues ^{134}I and ^{132}A, respectively, upon binding to the Nivo Fab. Other C_n fragments showed insignificant differences between the two states compared to the adjoining C_{n-1} ions. In summary, HDX-ETD allows epitope refinement to ^{127}S and/or ^{128}L, ^{131}K, ^{132}A and ^{134}I in the region ^{125}AISLAPKAQIKESL^{138}.

Although HDX-ETD successfully reveals several epitope binding residues, it can be challenging to differentiate HDX protection induced by direct binding or by a remote conformational change induced by binding. Additional information that shows the interacting domains is desirable, and that prompts our subsequent investigation with complementing chemical crosslinking.

5.4.3 Chemical Crosslinking of PD-1 and Nivolumab Fab

To achieve better coverage and more comprehensive information of the epitope and paratope regions provided by HDX, we utilized several cross-linkers, including BS^3-H_{12}/D_{12} and BS^2G-H_4/D_4, different in spacer lengths, and EDC/NHS targeting glutamic (D) and aspartic acids (E), complementary to the usual NHS-ester reactive residues (e.g., lysine (K), serine (S), tyrosine (Y) and the N-terminus). We tested several concentrations of cross-linkers with respect to those of the proteins and could monitor the success of crosslinking by the band of PD-1/Nivo Fab on an SDS PAGE gel (Figure 5.7). Individual cross-linked samples were digested in solution followed by LC-MS/MS analysis and crosslinking identification with pLink [39-40]. In total, we identified eight distinct inter-molecular cross-links (Table 5.1; representative mass and product-ion (MS/MS) spectra are shown in Figure 5.8) located on different regions of PD-1 and the Nivo Fab, consistent with the HDX results that the binding interface is discontinuous.

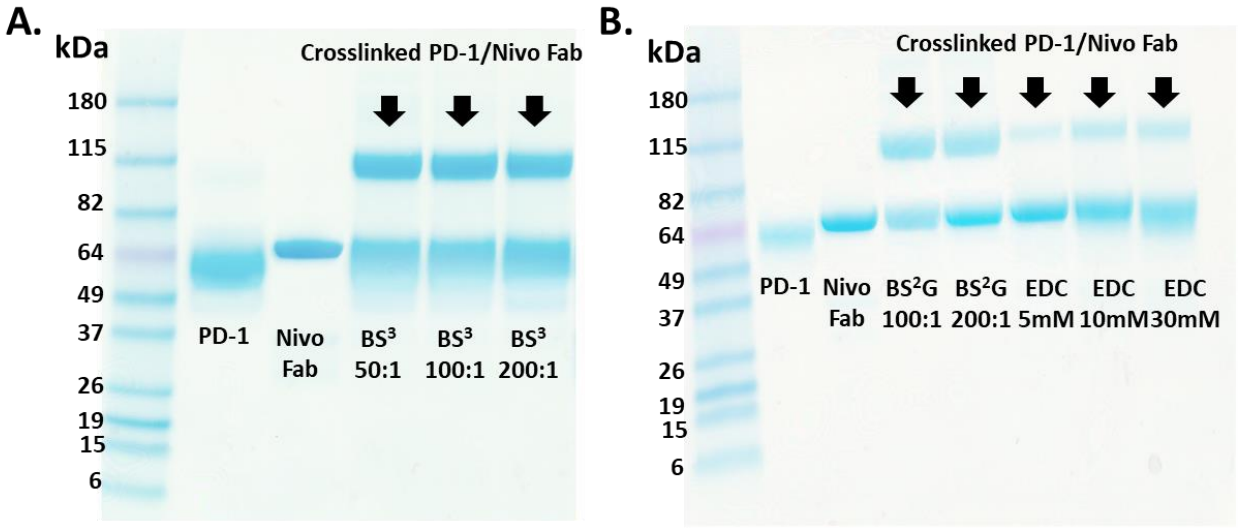


Figure 5.7 Gel electrophoresis of PD-1/Nivo Fab. PD-1 and Nivo Fab cross-linked with (A) BS^3 at the concentration of 0.75 mM (50:1), 1.5 mM (100:1) and 3 mM (200:1), cross-linked with (B) BS^2G at 1.5 mM (100:1) and 3 mM (200:1) and cross-linked with EDC at 5 mM, 10 mM and 30 mM. The first two electrophoresis lanes are control runs. Different species are identified on the lane correspondingly.

Table 5.1. Summary of observed inter cross-links

	PD-1	Nivo Fab	Cross-linker	Epitope	Paratope
1	S27 – K57(H)		BS^2G /BS^3	N-Loop	CDR-H2
2	D26 – K57(H)		EDC		
3	S27 – Y35(L)		BS^3		CDR-L1
4	S62– N-term (H)		BS^3	BC-Loop	N-terminus (H)
5	E61 – N-term (H)		EDC		
6	K135 – K57(H)		BS^3	FG-Loop	CDR-H2
7	K135 – Y35(L)		BS^3		CDR-L1
8	K135 – N-term (H)		BS^3		N-terminal (H)

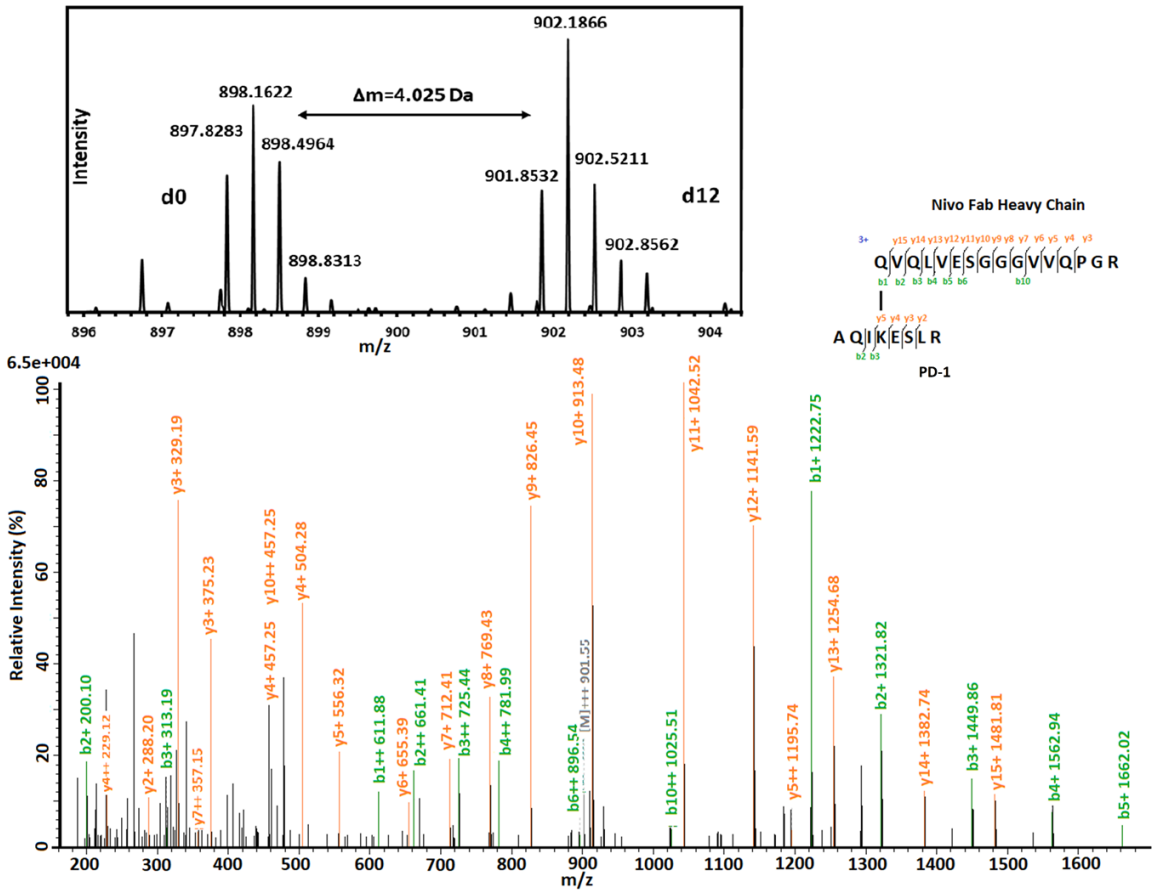


Figure 5.8 Representative mass spectra and product-ion spectra (XL-8)

For PD-1, multiple cross-links were formed on the N-loop (cross-link 1-3) and the FG-loop (cross-link 6-8), two regions that also showed significant protection in the HDX kinetics, consistent with the assignment as epitope regions. XL-MS results not only complement those of HDX but also reveal an additional binding region on PD-1, the BC-loop, identified by cross-links 4-5 by both BS^3 and EDC chemistry. It is worth mentioning that, for EDC crosslinking, the Euclidean distance between the cross-linked atoms is only ~ 3 Å, the length of one amide bond. Thus, this XL reagent locates and defines the binding interfaces with higher spatial resolution than do other reagents. Cross-link 5 emphasizes the vicinity of the BC-loop to the N-terminus of the heavy chain on Nivo Fab, indicating physical contacts between the epitope and paratope.

On the other hand, we observed no cross-links on the C'D-loop, and this could result from the lack of reactive residues or may indicate that it is not a binding interface. Although the sequence of C'D-loop contains several eligible residues for crosslinking (e.g., D and S), their side chains may orient in an unfavorable way for crosslinking. Because the C'D-loop region showed only low extent in the accumulative deuterium uptake in HDX upon binding with Nivo Fab, we suggest this is not a binding region that involves strong interactions. Binding-induced remote conformational changes more likely account for the reduced HDX upon binding.

On the heavy chain of the Nivo Fab, we observed three cross-links involving ^{57}K (i.e., cross-links 1, 2 and 6) located in the CDR-H2 region identified from the HDX kinetics. This region, based on the supporting evidence from both HDX and XL-MS, is confidently assigned to be a paratope. A newly revealed paratope region is the N-terminus of the heavy chain, which affords multiple cross-links not only with the BC-loop of PD-1 but also with the FG-loop. On the light chain of Nivo Fab, we observed only one cross-linked residue, ^{35}Y, on the CDR-L1 peptide. The identified cross-links (i.e., 3 and 7) support the CDR-L1 region as a binding interface with PD-1. For the four other CDR regions, we identified no cross-links, showing the limitations of using stand-alone XL-MS for mapping. Restricted numbers of reactive residues, considering both the intrinsic and low reactivity, side-chain orientation, and complexity of the cross-linked species diminish the possibility of using XL alone to characterize epitope/paratope interfaces. A more confident assignment than either approach alone is integrating HDX and XL-MS, which is validated by a comparison of our MS results and the resolved X-ray crystal structure in the next section.

5.4.4 Epitopes/Paratopes and Comparison with X-ray Crystallography

We conducted HDX to map the epitope and paratope on the peptide-level of PD-1 with both Nivo Fab and full-length Nivolumab, showing consistent results and indicating comparable binding behavior. The residue-level analysis was therefore performed with PD-1/Nivo Fab using HDX-ETD and crosslinking MS.

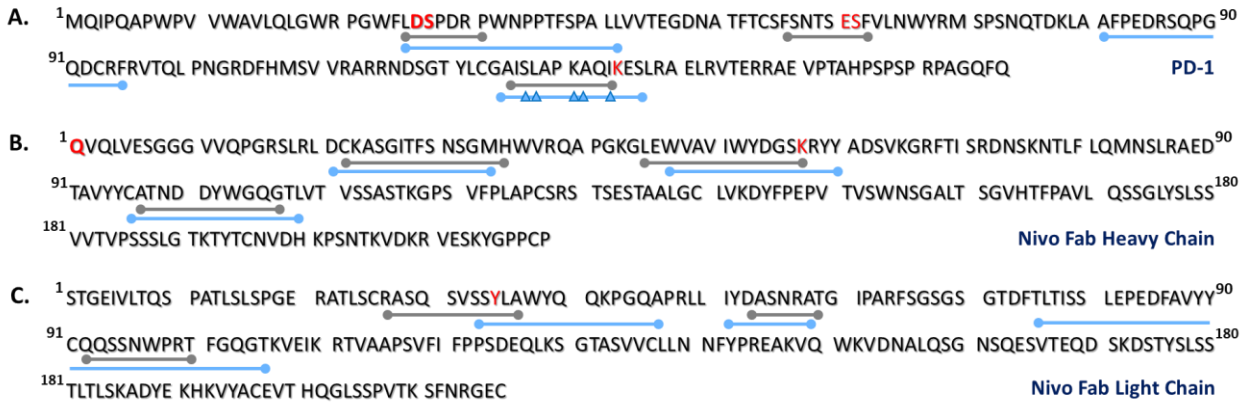


Figure 5.9 Summary of binding regions identified by HDX (blue) and XL-MS (red) for (A) the PD-1 and the Nivo Fab complex including (B) heavy chain and (C) light chain. Critical binding residues indicated by HDX-ETD are pinpointed with triangles. Epitope/paratopes assigned from the crystal structure (PDB: 5WT9) are underlined in grey.

The results from HDX suggest three epitope regions on PD-1, namely the N-loop, the FG-loop and the C'D-loop, whereas XL-MS supported the former two and revealed an additional BC-loop (Figure 5.9). A lack of cross-links and a small HDX difference between bound and unbound make the C'D-loop less likely to be a binding interface but rather a region undergoing a remote conformational change induced by binding elsewhere. These conclusions agree well with the reported epitope regions in the crystal structure of PD-1/Nivo Fab (PDB: 5WT9)[37-38], where the N-loop, BC-loop and FG-loop are identified as epitopes. The C'D loop, however, is not resolved and there are no observed physical contacts between the antigen and any paratope regions. The integrated information from HDX and XL-MS also provide insight on the epitopes/paratopes, in good accord with those indicated in the crystal structure. For example, a resolved H-bond

between ^{25}L on the N-loop of PD-1 and ^{57}K on the Fab heavy chain is consistent with two cross-links (cross-link 1 and 2, Table 5.1) to nearby reactive residues, ^{26}D and ^{27}S. In addition, the stabilization between the N-loop and CDR-L1 is consistent with cross-link 3 and with H-bonding between ^{26}D on PD-1 and ^{35}Y on the Fab light chain. HDX coupled with ETD fragmentation further reveal the binding residues in the FG-Loop, three of which, ^{128}L, ^{131}K and ^{132}A, contact with the Nivo Fab through H-bonding and van der Waals interactions. Assignment of this epitope region is also supported by cross-link 7, resembling the H-bond between ^{131}K on PD-1 and ^{37}A on the Fab light chain. Additionally, we observed that the FG-loop can cross-link with other domains through ^{135}K; those cross-links include the CDR-H2 (cross-link 6) and the N-terminus of the heavy Fab (cross-link 8). The N-terminus also cross-linked with ^{61}E and ^{62}S on the BC-loop of PD-1 (cross-links 4 and 5), suggesting a paratope region that is not seen in the solid-state structure.

In addition, the crosslinking network for the epitopes/paratopes delivers topological information of the PD-1/Nivo Fab complex by providing defined distance ranges. This allows a description of the interaction regions and even of the overall architecture when the information is coupled with other approaches (e.g., protein-protein docking). The 3D-information of the binding complex can provide a foundation for even more accurate determinations of the epitope/paratope.

5.4.5 Protein-Protein Docking of PD-1 and Nivolumab Fab

We conducted a protein-protein docking study with the RosettaDock program[44-45] by starting with the structures of PD-1 and the Nivo Fab extracted from the X-ray structure of the complex (PDB: 5WT9)[37]. Docking of apo Nivolumab and PD-1 was not feasible because many critical residues in apo-PD-1 are not resolved. For each docking run, we separated the two proteins with

the same intial configuration, followed by rotation to arbitrary extents (details in Methods). In Chapter 4[51], we demonstrated that incorporation of multiple cross-link-derived restraints in the protein-protein docking computations can effectively yield high-quality models; thus, the restraints based on all eight of the identified cross-links (Table 5.1) were utilized here. We generated 250 RosettaDock docking runs, each of which gave up to 400 PD-1/Nivo Fab models. The 20 best-scoring models based on the RosettaDock "total_score" metric revealed a tight cluster of models, all of which closely recapitulated the X-ray structure of the PD-1/Nivo Fab complex (Figure 5.10A). For these 20 models, the root-mean-square deviation (r.m.s.d.) of all Cα atoms for PD-1 in the models relative to PD-1 in the X-ray structure ranged from 0.4 – 1.8 Å with a mean of 1.0 ± 0.4 Å, which is less than 2 Å, a threshold that often is used to define a high-confidence model. The successful generation of the complex architecture showing the protein/protein interface enables an in-depth view of the potential epitopes/paratopes.

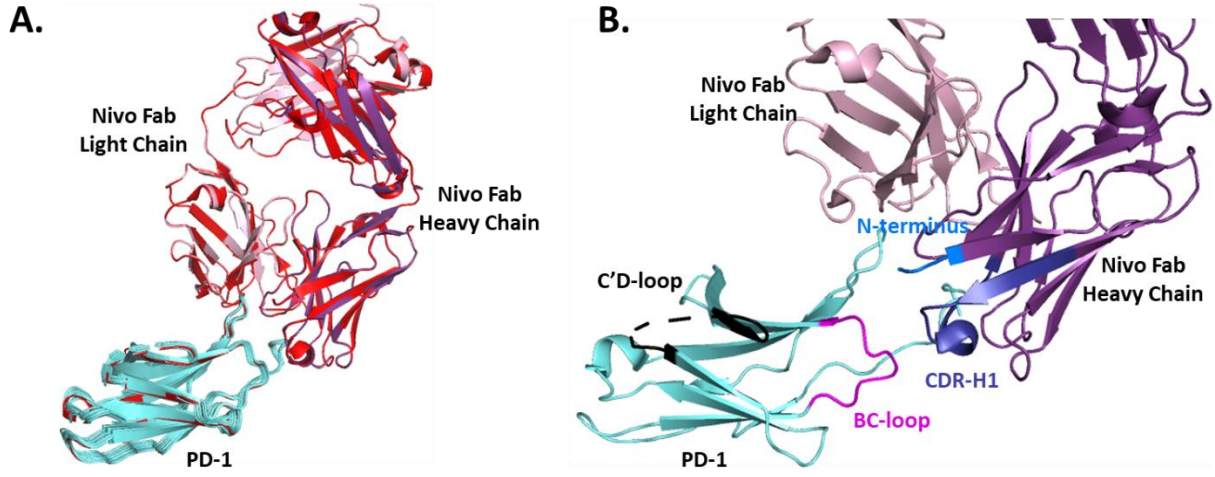


Figure 5.10 Docking models of PD-1/Nivo Fab complex. (A) Twenty best-scoring protein-protein docking models of PD-1/Nivo Fab complex (PD-1 in cyan, the heavy chain of the Nivo Fab in dark pink, the light chain of the Nivo Fab in light pink), superimposed on the X-ray structure (red). (B) A representative model of PD-1/Nivo Fab complex with an enlarged view of the epitopes/paratopes (BC-loop in magenta, C'D-loop in black, N-terminus of Nivo Fab heavy chain in marine, CDR-H1 in purple).

We chose a representative model and enlarged the binding interface between PD-1 and Nivo Fab (Figure 5.10B). The BC-loop of PD-1 locates at similar proximity with CDR-H1 and the N-terminal region of heavy Fab, indicating that the two regions may contribute simultaneously to the binding interaction. The conformation is consistent with the N-terminal region being a paratope. The X-ray structure suggested the BC-loop is in physical contacts with CDR-H1, whereas we identified cross-links on the N-terminus of the heavy chain, complementing the scheme and emphasizing the necessity of examing binding events using solution-based approaches. In addition, another questionable region, the C'D-loop on PD-1, is far from the binding interfaces, minimizing the likelihood of being an epitope. Moreover, unlike it is resolved in the unbound PD-1, the C'D-loop is missing in the bound PD-1 complex, giving supporting evidence of undergoing considerable remote conformational changes companied with different structural dynamics. The quaternary structure of the PD-1/Nivo Fab complex enables more confident assignment of the epitope/paratope regions.

There are limitations, however, in generalizing the integrated method that includes XL-MS and molecular docking to other protein-binding systems. One obstacle we encountered in the docking study of the PD-1/Nivo Fab complex is the dissimilar structures of PD-1 in its unbound state (PDB: 3RRQ) and bound state (PDB:5WT9). Epitopes on PD-1 are mainly loops, some of which cannot be resolved in the X-ray structure owing to their high flexibility in absence of bonding to the Nivo Fab (e.g., the N-loop). Consequently, distance constraints derived from the residues within this region are of little use for downstream docking. More importantly, for loops that are resolved, the orientation may be significantly different than those in the bound state. The conformations of dynamic loop regions are susceptible to the presence of mutations and/or truncations and to crystallization conditions including and ionic strength of the medium; these

can lead to incorrect conformations for the solid-state structure. Given that the docking protocol does not readily accommodate changes in protein tertiary structure, biased initial structures could lead to erroneous 3D models of the binding complex. Incorporation of other computational methods (e.g., discrete molecular simulation) can better accommodate the structural changes. In fact, binding interfaces that mainly contain helices and beta-sheets, which possess relatively fixed high order structures, are preferable inputs for docking studies.

5.5 Conclusion

This study provides convincing evidence that epitope and paratope mapping by HDX, crosslinking, and docking can be effective. Although HDX-MS, as a stand-alone method, has shown fruitful applications in mapping binding interfaces, it can be circumscribed by the limited resolution and ambiguous assignment as epitope or remote conformational change. Using PD-1/Nivo Fab as an example, we demonstrated that integrating HDX-ETD, XL-MS, and molecular docking gives a comprehensive description of epitopes/paratopes. Critical binding residues can be successfully identified from HDX-ETD and chemical crosslinking results, further delineating H-bonds and van der waals interactions along a protein/protein interface. In addition, XL-MS confirms epitopes/paratopes characterized by HDX-ETD and allows assignment of sites showing protection as remote conformational changes. The restrained distances afforded by XL-MS allow building high-confidence 3D models with molecular docking. The integrated platform magnifies the ability of each biophysical method, offering an alternative for other antigen/antibody systems that are difficult to crystallize for X-ray diffraction. It is noteworthy that docking exercises require careful consideration even with the availability of high-quality protein structures, which could be obtained by high-resolution techniques or computational methods. Whereas, even without molecular docking, the combination of HDX-ETD and XL-MS, the latter which is not

often used in epitope/paratope mapping experiments, gives insightful information to deepen our understanding of antigen-antibody binding and to assist the design of antibody therapeutics.

5.6 Acknowledgements

The study represents collaborative efforts by all authors. M.M.Z. conducted XL-MS and wrote the manuscript. R. Y-C. performed HDX-MS, HDX-ETD-MS and edited the manuscript. B.R.B conducted molecular docking, wrote the corresponding experimental section, and edited the manuscript.

The authors thank Dr. Olafur Gudmundsson, Deborah Loughney and Dr. Lois Lehman-McKeeman of Bristol Myers Squibb for their support of this project. The authors would also like to thank Dr. Shrikant Deshpande and Dr. Vangipuram Rangan from Bristol Myers Squibb for technical assistance. This research was supported by the NIH (Grant P41GM103422 to MLG) and by a Research Collaboration with Bristol Myers Squibb.

5.7 References

1. Sela-Culang, I.; Kunik, V.; Ofran, Y. The structural basis of antibody-antigen recognition. *Front. Immunol.* **2013**, *4*, 302.

2. Abbott, W. M.; Damschroder, M. M.; Lowe, D. C. Current approaches to fine mapping of antigen–antibody interactions. *Immunology* **2014**, *142*, 526-535.

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

4. Puchades, C.; Kűkrer, B.; Diefenbach, O.; Sneekes-Vriese, E.; Juraszek, J.; Koudstaal, W.; Apetri, A. Epitope mapping of diverse influenza Hemagglutinin drug candidates using HDX-MS. *Sci. Rep.* **2019**, *9*, 4735.

5. Zhu, S.; Liuni, P.; Ettorre, L.; Chen, T.; Szeto, J.; Carpick, B.; James, D. A.; Wilson, D. J. Hydrogen–Deuterium Exchange Epitope Mapping Reveals Distinct Neutralizing Mechanisms for Two Monoclonal Antibodies against Diphtheria Toxin. *Biochemistry* **2018**, *58*, 646-656.

6. Iacob, R. E.; Chen, G.; Ahn, J.; Houel, S.; Wei, H.; Mo, J.; Tao, L.; Cohen, D.; Xie, D.; Lin, Z. The influence of adnectin binding on the extracellular domain of epidermal growth factor receptor. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 2093-2102.

7. Iacob, R. E.; Krystek, S. R.; Huang, R. Y.; Wei, H.; Tao, L.; Lin, Z.; Morin, P. E.; Doyle, M. L.; Tymiak, A. A.; Engen, J. R. Hydrogen/deuterium exchange mass spectrometry applied to IL-23 interaction characteristics: potential impact for therapeutics. *Expert Rev. Proteom.* **2015**, *12*, 159-169.

8. Huang, R. Y.-C.; Iacob, R. E.; Sankaranarayanan, S.; Yang, L.; Ahlijanian, M.; Tao, L.; Tymiak, A. A.; Chen, G. Probing conformational dynamics of tau protein by hydrogen/deuterium exchange mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 174-182.

9. Weis, D. D. Hydrogen Exchange Mass Spectrometry of Proteins. Wiley Online Library: 2015.

10. 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*, 95-102.

11. Zehl, M.; Rand, K. D.; Jensen, O. N.; Jørgensen, T. J. Electron transfer dissociation facilitates the measurement of deuterium incorporation into selectively labeled peptides with single residue resolution. *J. Am. Chem. Soc.* **2008**, *130*, 17453-17459.

12. Rand, K. D.; Zehl, M.; Jensen, O. N.; Jørgensen, T. J. Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal. Chem.* **2009**, *81*, 5577-5584.

13. Landgraf, R. R.; Chalmers, M. J.; Griffin, P. R. Automated hydrogen/deuterium exchange electron transfer dissociation high resolution mass spectrometry measured at single-amide resolution. *J. Am. Soc. Mass Spectrom*. **2012**, *23*, 301-309.

14. Huang, R. Y.-C.; Krystek Jr, S. R.; Felix, N.; Graziano, R. F.; Srinivasan, M.; Pashine, A.; Chen, G. Hydrogen/deuterium exchange mass spectrometry and computational modeling reveal a discontinuous epitope of an antibody/TL1A Interaction. *MAbs* **2018**, *10*, 95-103.

15. Yu, C.; Huang, L. Crosslinking mass spectrometry: an emerging technology for interactomics and structural biology. *Anal. Chem.* **2017**, *90*, 144-165.

16. Liu, F.; Heck, A. J. Interrogating the architecture of protein assemblies and protein interaction networks by crosslinking mass spectrometry. *Curr. Opin. Struct. Biol.* **2015**, *35*, 100-108.

17. Liu, X. R.; Zhang, M. M.; Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chem. Rev.* **2020**, *120*, 4355-4454.

18. Arlt, C.; Ihling, C. H.; Sinz, A. Structure of full‐length p53 tumor suppressor probed by chemical cross‐linking and mass spectrometry. *Proteomics* **2015**, *15*, 2746-2755.

19. Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; Bukowski‐Wills, J. C.; Nilges, M. Architecture of the RNA polymerase II–TFIIF complex revealed by cross‐linking and mass spectrometry. *EMBO J.* **2010**, *29*, 717-726.

20. Erzberger, J. P.; Stengel, F.; Pellarin, R.; Zhang, S.; Schaefer, T.; Aylett, C. H.; Cimermančič, P.; Boehringer, D.; Sali, A.; Aebersold, R. Molecular architecture of the 40S· eIF1· eIF3 translation initiation complex. *Cell* **2014**, *158*, 1123-1135.

21. Chavez, J. D.; Weisbrod, C. R.; Zheng, C.; Eng, J. K.; Bruce, J. E. Protein interactions, post-translational modifications and topologies in human cells. *Mol. Cell. Proteom.* **2013**, *12*, 1451-1467.

22. Pimenova, T.; Nazabal, A.; Roschitzki, B.; Seebacher, J.; Rinner, O.; Zenobi, R. Epitope mapping on bovine prion protein using chemical cross‐linking and mass spectrometry. *J. Mass Spectrom.* **2008**, *43*, 185-195.

23. Ishida, Y.; Agata, Y.; Shibahara, K.; Honjo, T. Induced expression of PD‐1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J.* **1992**, *11*, 3887-3895.

24. Pedoeem, A.; Azoulay-Alfaguter, I.; Strazza, M.; Silverman, G. J.; Mor, A. Programmed death-1 pathway in cancer and autoimmunity. *Clin. Immunol.* **2014**, *153*, 145-152.

25. Freeman, G. J.; Long, A. J.; Iwai, Y.; Bourque, K.; Chernova, T.; Nishimura, H.; Fitz, L. J.; Malenkovich, N.; Okazaki, T.; Byrne, M. C. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. J. *Exp. Med.* **2000**, *192*, 1027-1034.

26. Latchman, Y.; Wood, C. R.; Chernova, T.; Chaudhary, D.; Borde, M.; Chernova, I.; Iwai, Y.; Long, A. J.; Brown, J. A.; Nunes, R. PD-L2 is a second ligand for PD-1 and inhibits T cell activation. *Nat. Immunol.* **2001**, *2*, 261.

27. Patsoukis, N.; Sari, D.; Boussiotis, V. A. PD-1 inhibits T cell proliferation by upregulating p27 and p15 and suppressing Cdc25A. *Cell Cycle* **2012**, *11*, 4305-4309.

28. Iwai, Y.; Ishida, M.; Tanaka, Y.; Okazaki, T.; Honjo, T.; Minato, N. Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12293-12297.

29. Blackburn, S. D.; Shin, H.; Haining, W. N.; Zou, T.; Workman, C. J.; Polley, A.; Betts, M. R.; Freeman, G. J.; Vignali, D. A.; Wherry, E. J. Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nat. Immunol*. **2009**, *10*, 29.

30. Curiel, T. J.; Wei, S.; Dong, H.; Alvarez, X.; Cheng, P.; Mottram, P.; Krzysiek, R.; Knutson, K. L.; Daniel, B.; Zimmermann, M. C. Blockade of B7-H1 improves myeloid dendritic cell–mediated antitumor immunity. *Nat. Med.* **2003**, *9*, 562.

31. Wang, H. Y.; Lee, D. A.; Peng, G.; Guo, Z.; Li, Y.; Kiniwa, Y.; Shevach, E. M.; Wang, R.-F. Tumor-specific human CD4+ regulatory T cells and their ligands: implications for immunotherapy. *Immunity* **2004**, *20*, 107-118.

32. Zou, W.; Wolchok, J. D.; Chen, L. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations. *Sci. Transl. Med*. **2016**, *8*, 328rv4.

33. Callahan, M. K.; Postow, M. A.; Wolchok, J. D. Targeting T cell co-receptors for cancer therapy. *Immunity* **2016**, *44*, 1069-1078.

34. Topalian, S. L.; Hodi, F. S.; Brahmer, J. R.; Gettinger, S. N.; Smith, D. C.; McDermott, D. F.; Powderly, J. D.; Carvajal, R. D.; Sosman, J. A.; Atkins, M. B. Safety, activity, and immune correlates of anti–PD-1 antibody in cancer. *N. Engl. J. Med.* **2012**, *366*, 2443-2454.

35. Brahmer, J. R.; Drake, C. G.; Wollner, I.; Powderly, J. D.; Picus, J.; Sharfman, W. H.; Stankevich, E.; Pons, A.; Salay, T. M.; McMiller, T. L. Phase I study of single-agent anti–programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *J. Clin. Oncol*. **2010**, *28*, 3167.

36. Motzer, R. J.; Escudier, B.; McDermott, D. F.; George, S.; Hammers, H. J.; Srinivas, S.; Tykodi, S. S.; Sosman, J. A.; Procopio, G.; Plimack, E. R. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N. Engl. J. Med.* **2015**, *373*, 1803-1813.

37. Tan, S.; Zhang, H.; Chai, Y.; Song, H.; Tong, Z.; Wang, Q.; Qi, J.; Wong, G.; Zhu, X.; Liu, W. J. An unexpected N-terminal loop in PD-1 dominates binding by nivolumab. *Nat. Commun.* **2017**, *8*, 14369.

38. Lee, J. Y.; Lee, H. T.; Shin, W.; Chae, J.; Choi, J.; Kim, S. H.; Lim, H.; Heo, T. W.; Park, K. Y.; Lee, Y. J. Structural basis of checkpoint blockade by monoclonal antibodies in cancer immunotherapy. *Nat. Commun*. **2016**, *7*, 13354.

39. Yang, B.; Wu, Y.-J.; Zhu, M.; Fan, S.-B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y.-X.; Chen, H.-F. Identification of cross-linked peptides from complex samples. *Nat. Methods* **2012**, *9*, 904.

40. Chen, Z.-L.; Meng, J.-M.; Cao, Y.; Yin, J.-L.; Fang, R.-Q.; Fan, S.-B.; Liu, C.; Zeng, W.-F.; Ding, Y.-H.; Tan, D. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun*. **2019**, *10*, 3404.

41. Bender, B. J.; Cisneros III, A.; Duran, A. M.; Finn, J. A.; Fu, D.; Lokits, A. D.; Mueller, B. K.; Sangha, A. K.; Sauer, M. F.; Sevy, A. M. Protocols for molecular modeling with Rosetta3 and RosettaScripts. *Biochemistry* **2016**, *55*, 4748-4763.

42. Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **1997**, *268*, 209-225.

43. Simons, K. T.; Ruczinski, I.; Kooperberg, C.; Fox, B. A.; Bystroff, C.; Baker, D. Improved recognition of native‐like protein structures using a combination of sequence‐dependent and sequence‐independent features of proteins. *Proteins* **1999**, *34*, 82-95.

44. Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol*. **2003**, *331*, 281-299.

45. Chaudhury, S.; Berrondo, M.; Weitzner, B. D.; Muthu, P.; Bergman, H.; Gray, J. J. Benchmarking and analysis of protein docking performance in Rosetta v3. 2. *PloS one* **2011**, *6* e22477.

46. Release 4: Maestro, Schrödinger, LLC, New York, NY, 2017. Schrödinger inc 2017.

47. Leitner, A.; Joachimiak, L. A.; Unverdorben, P.; Walzthoeni, T.; Frydman, J.; Förster, F.; Aebersold, R. Chemical crosslinking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9455-9460.

48. Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. *Trends Biochem. Sci.* **2016**, *41*, 20-32.

49. OEChem, T., OpenEye Scientific Software. Inc., Santa Fe, NM, USA 2012.

50. Walters, B. T.; Ricciuti, A.; Mayne, L.; Englander, S. W. Minimizing back exchange in the hydrogen exchange-mass spectrometry experiment. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 2132-2139.

51. Zhang, M. M.; Beno, B. R.; Huang, R. Y.-C.; Adhikari, J.; Deyanova, E. G.; Li, J.; Chen, G.; Gross, M. L. An Integrated Approach for Determining a Protein–Protein Binding Interface in Solution and an Evaluation of Hydrogen–Deuterium Exchange Kinetics for Adjudicating Candidate Docking Models. *Anal. Chem*. **2019**, *91*, 15709-15717.

Chapter 6: Structure of Cyanobacterial Phycobilisome Core Revealed by Structural Modeling and Chemical Cross-linking*

* This chapter is based on the following publication: Zhang, M. M.; Weisz, D. A.; Cheng, M.; Pakrasi, H.M.; Gross, M. L.; Blankenship, R. E.; Liu, H. Structure of Cyanobacterial Phycobilisome Core Revealed by Structural Modeling and Chemical Cross-linking *Sci. Adv.* **2020,** Under Revision.

6.1 Abstract

In cyanobacteria and red algae, the structural basis dictating efficient excitation energy transfer from the phycobilisome (PBS) antenna complex to the reaction centers (RCs) remains unclear. PBS has several peripheral rods and a central core that binds to the thylakoid membrane, allowing energy coupling with Photosystems II (PSII) and Photosystem I (PSI). Here, we integrated chemical cross-linking mass spectrometry with homology modeling analysis to propose a tri-cylindrical cyanobacterial PBS-core structure. Our model reveals a side view crossover configuration of the two basal cylinders, consolidating the essential roles of the anchoring domains comprised of the ApcE PB-loop and ApcD, which facilitate the energy transfer to PSII and PSI respectively. The uneven bottom surface of the PBS-core contrasts with the flat reducing side of PSII. The extra space between two basal cylinders of the PBS-core and PSII provides increased accessibility of regulatory elements, e.g., orange carotenoid protein, which are required for modulating photochemical activities.

6.2 Introduction

In natural photosynthesis, light-harvesting complexes capture and transmit solar energy to the reaction centers (RCs) where photochemistry takes place, leading to long-term energy storage [1-2]. In cyanobacteria and red-algae, phycobilisomes (PBSs) harvest the energy of a spectral range between 450 and 650 nm, which is different than that of Chl *a* (chlorophyll *a*) in the RCs: i.e., photosystem I (PSI) and photosystem II (PSII), and significantly increase the utilization of the solar energy spectrum [3]. Phycobilisomes, with a molecular mass range of 5-20 MDa, are located on the cytoplasmic side (or the stromal side of the red algal chloroplasts) of RCs. PBSs are highly organized assemblies of brightly colored phycobiliproteins and colorless linker

polypeptides [4]. Each phycobilisome consists of a central cylindrical core made of allophycocyanin, from which several outwardly oriented rods radiate. Light energy collected by phycobiliproteins in the peripheral rods is transferred to PSI and PSII, with the PBS-core allophycocyanin acting as a link between phycocyanin and the Chl *a* in RCs. The detailed structural orientation of the PBS and the RCs, however, remains to be determined, probably because of the structurally weak interactions of PBS and RCs, which could be advantageous for easy regulation of the excitation energy transfer under varying light conditions[5-6]. The PBS-core not only serves as a cornerstone for the rods to attach [7-8], but also acts as an anchoring module that dictates efficient energy flow from PBSs to RCs [9-13].

The building blocks of the PBS-core are two homologous proteins, α- and β-phycocyanin subunits, which form a heterodimer, αβ, and further self-assemble into disc-like $(\alpha\beta)_3$ trimers [4]. Several discs of allophycocyanin (ApcA/B) stack into a cylinder, varying in numbers in different organisms. The best-known PBS structure is described as hemidiscoidal PBS [7, 14]. In this group, there are bi-cylinder type of PBS in *Synechococcus sp.* PCC 6301 [15] and penta-cylinder type in *Mastigocladus laminosus* as well [16]. The core structure of *Synechocystis sp.* PCC 6803, however, is tri-cylindrical as determined in 1986 by electron microscopy (EM) [17], Later, advanced EM methodology yielded a high resolution structure at 13 Å using a genetically modified strain in *Synechocystis* 6803 (CK strain) [7] which contains an intact core complex but no PC rods. The EM structure reveals that this PBS-core consists of three stacked core cylinders: two on the bottom and one on the top. The 2D EM map further highlights a two-fold rotational symmetry of the triangular core, indicating that the two basal cylinders are arranged in an anti-parallel fashion. Although the overall architecture is resolved, a detailed molecular model showing the location of

each terminal energy emitter, ApcE, ApcD, and the tomography of the bottom surface of the PBS-core that directly interacts with the thylakoid membranes are not known.

Recently, Zhang *et al.* significantly advanced the understanding of the PBS structure, specifically, in a red alga *Griffithsia pacifica* [18] by charactering a 16.8 MDa PBS (thus GpPBS) using cryo-EM at a resolution of 3.5 Å. GpPBS contains numerous protein subunits, linker proteins, and chromophores, revealing in great detail of the GpPBS architecture. GpPBS is morphologically categorized as a block-shaped type [18], likely an evolutionary derivative of the core of hemidiscoidal PBS [18]. The GpPBS core is characterized by two unique features: the lack of one trimeric ApcA/B disc at the distal end of ApcD in each basal cylinder (Fig. 6.1A) and thus a decreased copy number (by comparing to a four-APC-trimer cylinder) of ApcC (allophycocyanin C, formerly L_C), a small polypeptide linker protein that stabilizes and caps the core cylinders [19]. The top cylinder in GpPBS-core is featured by two APC trimers, instead of four, which are stacked back-to-back (or tail-to-tail). Interestingly, part of the 2nd and 3rd linker domains of ApcE [18], which is usually buried in the core of each cylinder [3], are exposed and do not interact with the APC trimers.

In this study, we build a 3D structure of cyanobacterial PBS-core (cPBS-core) starting with the available crystal structure of cyanobacterial allophycocyanin and the red algal PBS cryo-EM structure in combination with protein structure and function prediction software suites. The generated model represents a cyanobacterial tri-cylindrical core with two basal cylinders and one top cylinder, each of which contains four APC trimers. Other APC subunits, including ApcD, ApcE and ApcF, were all modelled to afford the core architecture. We also performed chemical cross-linking using three types of PBSs isolated from *Synechocystis* 6803, mapped the identified cross-links onto the proposed 3D model, a portrayal of the native cPBS-core, and calculated the

Euclidean Cα- Cα distances between each pair of the cross-linked residues. The proposed cPBS-core was then justified and evaluated by the experimental distance restraints with respect to individual subunits. The described framework reveals a side view X-shape configuration of the two basal cylinders, which only allows the protrusion of the PB-loops on ApcE and ApcD interacting with the thylakoid membranes. The bottom surface of the PBS that touches the thylakoid membrane/RCs is not perfectly flat, in contrast to the flat surface of the reducing side of PSII. The increased accessibilities to the bottom surface of the PBS-core and the reducing side of PSII provide perspectives on the excitation energy regulation and reaction center photoprotection.

6.3 Experimental

6.3.1 Cyanobacteria Culture and PBS Purification

The CK-PBS and CpcL-PBS mutant strains were generous gifts from Dr. Ghada Ajlani[20-21]. Cyanobacterial strains were grown in BG-11 medium at 30 °C with 50 μmol photo m^{-2}. The growth media was supplemented with 20 mM TES (2-[(2-hydroxy-1,1-bis(hydroxymethyl) ethyl) amino] ethanesulfonic acid, N-[tris(hydroxymethyl)methyl]-2 aminoethanesulfonic acid)-KOH (pH 7.5). The harvested cell cultures were resuspended in 0.8 M K-phosphate buffer and incubated with protease inhibitor cocktail (Thermo Fisher Scientific, Waltham, MA) and DNase (Sigma, St. Louis, MO). Cell lysates were obtained through three rounds of a French press at 4 °C and pH 7.5. 2% Triton X-100 (Sigma, St. Louis, MO) was then added following by 30 min incubation at room temperature. The blue supernatant was loaded onto a sucrose gradient for overnight ultracentrifugation (370,000 × g). The purified PBS was then ready for cross-linking.

6.3.2 Chemical Cross-linking and Proteolytic Digestion

CK-PBS, dissolved in 0.4 M K-phosphate buffer at 0.1 μM, and the isotopic-coded BS^3 cross-linker mixture (BS^3-H_{12}/D_{12}, Creative Molecule, Inc) were incubated together for 10 mins in the dark at 25 °C, where the cross-linker was in 10-, 50-, 100-fold excess with respect to the PBS-CK. Tris (1M) was added to give a final concentration of 50 mM to stop the cross-linking chemistry. For quenching and desalting purposes, Zeba spin columns (Thermo Fisher Scientific, Waltham, MA) were employed following the manufacture's protocol. The cross-linked CK-PBS were then desalted and purified by acetone precipitation to prepare them for enzymatic digestion. The protein pellets were dissolved in 20 μl, 8M urea for 30 min at room temperature for denaturation, followed by 30-min incubation with 2.5 mM TCEP (tris(2-carboxyethyl)phosphine) at 37 °C. Iodoacetamide was then added to the reaction sample at a final concentration of 5 mM for 30 min at 25 °C in dark. The first-step digestion was achieved by Lys-C (0.05 μg/μl) for 2-h incubation at 37 °C. Dilution to 1M urea with Tris buffer (100 mM) was needed for the subsequent trypsin digestion, which required overnight incubation at 37 °C with a trypsin:protein ratio of 1:25. The digested mixture was quenched by 0.1% formic acid the next day.

6.3.3 LC-MS/MS

The peptide mixture was loaded onto a C18 trapping column 180 μm × 2 cm, C18 Symmetry, 5 μm, 100 Å, Waters, MA) for desalting with phase A (water with 0.1% formic acid) of Dionex Ultimate HPLC (Thermo Fisher Scientific, Waltham, MA). Peptide samples were then eluted and separated on a reverse-phase C18 column (100 μm × 15 cm, C18 Symmetry, 5 μm, 100 Å, Waters, MA) with a 105-min gradient: increasing phase B (80% acetonitrile, 20% water, 0.1% formic acid) from 2% to 40% for the first 90 mins, from 40% to 95% for next 10 mins and equilibrated at phase A for another 5 mins. The flow was controlled at 250 nL/min and sprayed

through a Nanospray Flex source coupled with a Q Exactive Plus mass spectrometer (Thermo Scientific, Waltham, MA). The mass spectrometer was operated at the following settings: 1.8 kV spray voltage; positive-ion mode; MS1 acquisition at 70,000 resolving power at m/z 200; AGC target 3×10^6; MS/MS acquisition at 17,500 resolving power at m/z 200; AGC target: 1×10^5; maximal injection time of 100 ms. Data dependent acquisition and with "mass tag" were utilized. In the latter method, for charge states of 2 to 7, the delta mass differences were set as ±6.03762, ±4.02508, ±3.01881, ±2.41505, ±2.01254, and ±1.72503 *m/z*, respectively. Singly charged species were excluded and each charge state was acquired independently.

6.3.4 Cross-linked Peptides Identification

MS and MS/MS data were imported into pLink software for identification. Searching parameters were: enzyme was trypsin (up to three missed cleavages) with 20 ppm of precursor tolerance and 60 ppm of fragment tolerance. Variable modifications were: oxidation of M, deamidation of N, Q, and N terminus. The minimum number of peptide length is 6 with peptide mass of 600 Da and maximum number is 60 with peptide mass of 6000 Da. The false discovery rate was equal to or smaller than 5% at spectral level with a 10-ppm filter tolerance. Isotopic pairs were examined manually in raw files to confirm cross-link identification. The theoretical product-ion mass list was calculated in Protein Prospector. Manually validation of the fragments was further performed as a comparison to the pLink assignment.

The cross-linking dataset was also submitted into ICC-Class for cross-linking identification. In the DXMSMSMatch program, the settings were: cross-linker was DSS with DX of 12.07532. The DX mass tolerance was 0.013 Da, retention time tolerance was 60 s. The filtered DX mass tolerance was 5 ppm and the filter DX time window was 60 s. The digestion sites were K and R, including cross-link sites. Missed digest sites was up to 4. Assigned cross-link sites are K and N-

terminal primary amines. Precursor tolerance was 5 ppm and fragment tolerance were 30 ppm. Manually validation of the fragments was carried out as the same way mentioned above.

6.3.5 Homology Modeling and Structural Construction

Homology simulation was performed on I-TASSER server. Detailed information and adopted template are described in homology modeling section.

6.4 Results

6.4.1 Homology Modeling and Construction of a Tri-Cylindrical Cyanobacterial PBS-Core

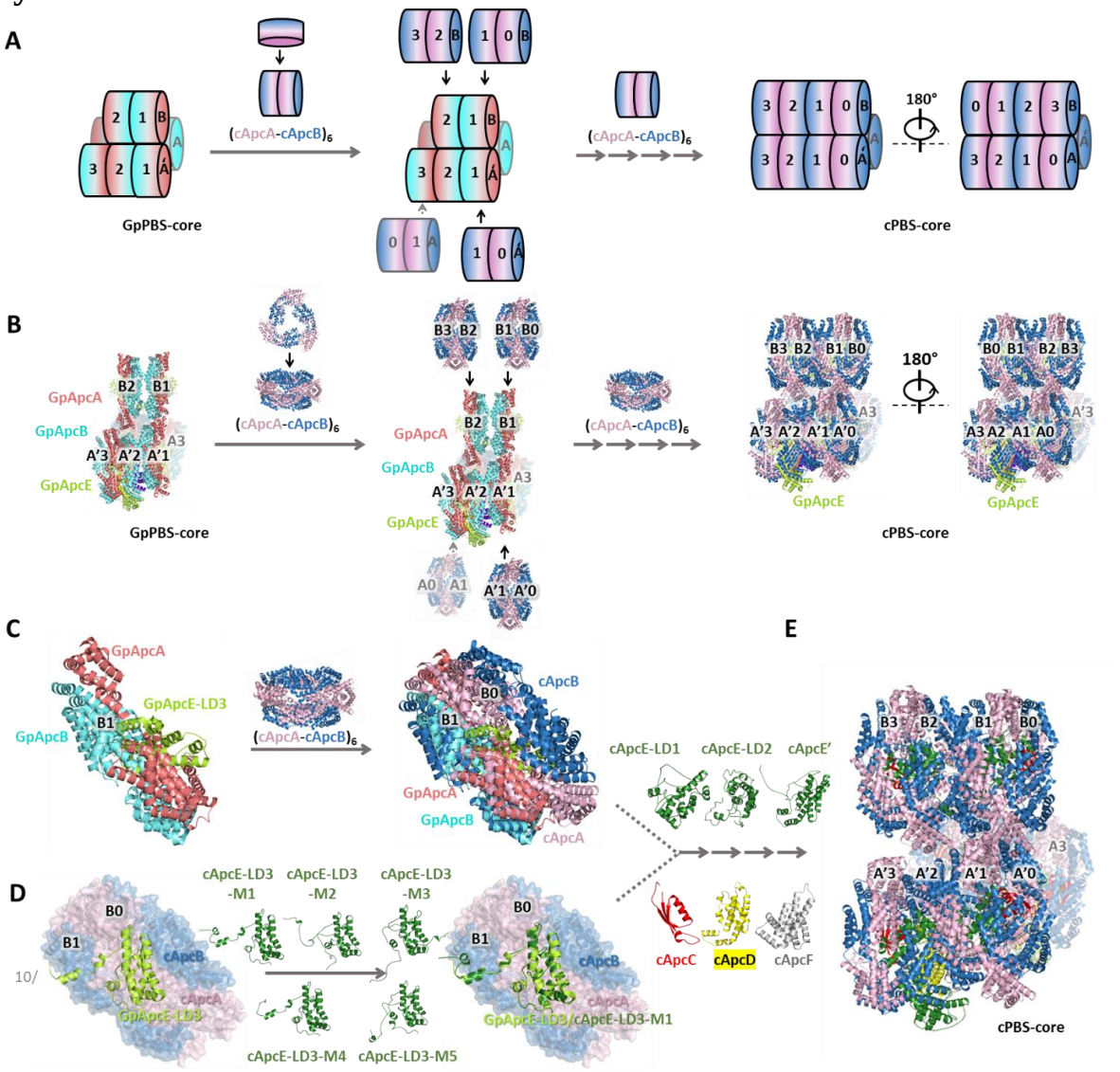


Figure 6.1 Model construction of a tri-cylindrical cyanobacterial PBS-core. (A) Symbolic illustration of disc extension based on GpPBS-core structure (PDBID: 5Y6P) and cyanobacterial structure. GpApcA (salmon), GpApcB (cyan), GpApcE (lime), cApcA (c for cyanobacteria, light pink), cApcB (dark blue). (B) Detailed cartoon representation of disc extension based on GpPBS-core structure (PDBID: 5Y6P) and cyanobacterial structure. (C) Extension of the B1 disc in GpPBS-core using cApcA-ApcB hexamer (PDBID: 4F0U). (D) Determination of the cApcE-LD3 structure based on the evaluation of five predicted cApcE-LD3 models (dark green), given by I-TASSER, with respect to the GpApcE-LD3 (lime). Other APC domains, i.e. cApcE-LD1 (dark green), cApcE-LD2 (dark green), α^{LCM} (dark green), cApcC (red), cApcD (yellow) and cApcF (grey), are constructed similarly. (E) The final core model of cPBS. All structure figures were prepared using PyMOL.

We adopted the high-resolution GpPBS-core structure from red algae *Griffithsia pacifica* (PDBID: 5Y6P) [18] as a building chassis. Although the overall appearance of GpPBS is a block type PBS, it essentially retains the classical two fold symmetry with its symmetrical axis oriented perpendicularly to the thylakoid membranes [18], and importantly, is also considered to be evolutionally derived from a core of hemidiscoidal PBS-core by eliminations of the exterior APC disc (APC trimer) of all three cylinders. This hypothesis is consistent with a hemidiscoidal model proposed in earlier reports [7]. We are interested in building a tri-cylindrical cyanobacterial hemidiscoidal PBS core with four stacked APC trimers per cylinder.

The cryo-EM structure of the GpPBS-core revealed several features: each basal cylinder contains three GpApcA/B trimers, namely disc A1-3 and A′1-3, and the top cylinder contains only two trimers with tail-to-tail orientation (three ApcBs in one disc to three ApcBs in another disc), i.e., disc B1 and B2 (Fig. 6.1A and 6.1B, left, for simplified and detailed cartoon representation, respectively). Each disc contains 3 copies of ApcA and ApcB. This structure is in contrast to that of a model in *Synechocystis* 6803 [7, 11], which indicates that each core cylinder contains four discs, arranged in a face-to-face (three ApcAs in one disc to three ApcAs in another disc), tail-to-

tail, and face-to-face fashion. Using homology modeling, we swapped all the GpPBS-core APC discs by using cyanobacterial APC (cApc) discs. Each cylinder was also extended up to four discs. First of all, a hexameric cyanobacterial APC (PDBID: 4F0U) was generated and used as a building unit. This hexamer contains two discs arranged in face-to-face. To extend the top cylinder, a cyanobacterial APC hexamer was adopted to align with the disc B1. The alignment of the B1 disc results in the replacement of the red algal counterpart and in an extension of one cyanobacterial APC trimer, i.e., the B0 disc (Fig. 6.1A and 6.1B, middle, for simplified and detailed cartoon representation, respectively). Similarly, the top cylinder was extended with the alignment of B2 disc by another hexameric APC (two discs, face-to-face), leading to an addition of disc B3 to make a total of four-disc top cylinder. The replacement and extension for the two bottom cylinders were performed on A1 and A′1 discs, giving additional APC trimers as disc A0 and A′0. The remaining red algae A2/A3 and A′2/A′3 hexamers were further swapped with cyanobacterial APC hexametric units, contributing to a cPBS-core architecture with four cyanobacterial discs on each of the three cylinders (Fig. 6.1A and 6.1B, right, for simplified and detailed cartoon representation, respectively). The selection of cyanobacterial crystal structure (PDBID: 4F0U) took into account of its high resolution structure and high protein sequence homology with those of *Synechocystis* 6803 [22], which was used in our following chemical cross-linking studies. Extension of A0/A′0 not only fulfills the face-to-face geometry in adjacent discs (A1 and A′1 respectively) but also enables the 2nd linker domain ApcE-LD2/ApcE-LD2′ being buried inside the cylinder, as the ApcE-LD1/ApcE-LD1′ is concealed within disc A3/A2 and A′3/A′2 in two basal cylinders respectively [18]. We demonstrated in detail the swapping and elongation of the B1 disc (GpPBS-core) as an example (Fig. 6.1C). A value of less than 1 Å RMSD (root-mean-square deviation) was obtained (Table S6.1), indicative of high structural

similarities of red algal ApcA/B trimer and the cyanobacterial ApcA/B trimer assembly. The protein sequence identities between the cyanobacteria and red alga indeed strongly support this conclusion (Fig. S6.1). The remaining GpApcA/B trimers in each basal cylinder that contain ApcD and ApcE were also substituted by cApcA/B hexamer with consistent low RMSD, contributing to a complete framework of a cPBS core (Fig. 6.1E). For other PBS-core components, such as ApcC, ApcD, ApcE, and ApcF, we either used the available crystal structure (ApcE, PDBID: 4XXI; ApcD, PDBID: 4PO5) [23] or predicted their structures by using protein sequence from *Synechocystis* 6803 on the Zhang Server [24] and then defined their positions by using homology modeling. Their copy numbers were based on the characteristic features of a tri-cylindrical cyanobacterial PBS [7].

ApcE, also termed as Core-Membrane Linker or L_{CM} [25], is a multifunctional protein containing both phycocyanin binding domain, known as PB domain (or α^{LCM}), and several linker domains that are essential for connecting APC trimers and thus for assembly of APC discs into cylinders. The α^{LCM} is located in disc A2 and A′2 in the red algal structure [18]. It was proposed that ApcE linker domain also serves to stitch all the discs and subsequently cylinders together to form the PBS-core [7]. The large cApcE protein in tri-cylindrical PBS in *Synechocystis* 6803 contains four domains, namely, α^{LCM} (PB domain/ α domain in L_{CM}, M^{1}-V^{240}), cApcE-LD1 (D^{247}-V^{427}), cApcE-LD2 (I^{431}-D^{684}) and cApcE-LD3 (K^{685}-G^{896}) (Fig. S6.2), which were treated individually for structure prediction and modeling. For example, when the cApcE-LD3 peptide sequence was submitted to I-TASSER server [24], the homology-modeling algorithm screened through the database and adopted the GpApcE (PDBID: 5Y6P) and another two protein structures (PDBID: 2KY4, DOI: 10.2210/pdb2ky4/pdb; 3OHW, DOI: 10.2210/pdb3OHW/pdb) among others as references, giving top five high-scoring models (Fig. 6.1D, M1-M5, Table S6.1). These models

resemble the overall configuration and helical content which are consistent with protein sequence identity (Fig. S6.2). Each generated model was then aligned with both GpApcE-LD3 domains in the top cylinders, evaluated in terms of RMSD for all corresponding alpha carbons (Table S6.1). The model 1 (M1) and M4 of cApcE-LD3 give two smallest RMSD values, showing major differences at the region K^{685}- E^{703}. This region in cApcE-LD3-M1 sticks towards the basal cylinders, while that of cApcE-LD3-M1 curls back to the top cylinder. Given the fact that cApcE-LD3 bundles basal and top cylinders together as a linker domain, it is believed that the region K^{685}- E^{703} is more towards the bottom cylinders; this hypothesis is also supported by a recently resolved PBS from *prophyridium purpureum* [26]. Therefore, we chose cApcE-LD3-M1 as the representative model. We use both an I-TASSER generated model and the X-ray crystal structure of α^{LCM} (PDBID: 4XXI) [27] in our modeling, since the former model contains the PB-loop that allows model building of the PBS-core-PSII [13].

Other cApcE linker domains, i.e., cApcE-LD1 and cApcE-LD2,) which are composed similarly with a Pfam00427 (PDBID: 2KY4, 2L06, 3OHW, released by Northeast Structural Genomics Consortium of America), and the α^{LCM} were performed for homology modeling in a comparable fashion. Five generated top-scoring models for each domain were evaluated individually by structural alignment in terms of calculated RMSD (Table S6.1). We chose the models with the smallest RMSD value to give the best likelihood of being the most reasonable configuration, namely model 4 for α^{LCM}, model 4 for cApcE-LD1, and model 4 for cApcE-LD2. cApcE-LD1 and cApcE-LD2 are in the cavity of the basal rod-like protrusions between the A2/A3 (or A′2/A′3) and A0/A1 (or A′0/A′1) cApcA/B discs, respectively, stabilizing two basal cylinders. Two copies of α^{LCM} situate at the A2 and A′2 of the basal cylinders with a large loop (PB-loop)

from T^{77} to G^{135}, extending towards the thylakoid membrane and anchoring the whole cPBS assembly to reactive centers [2, 7, 28-29].

GpApcD, an APC α-like subunit that is required for energy delivery from PBS to PSI[30], was substituted by a resolved structure of ApcD at 1.75 Å from a previous study (PDBID: 4PO5) [23]. The two GpApcD subunits are located at the A3 and A′3 discs in the basal cylinders, respectively, both of which are well-aligned with the cApcD and affording RMSD differences that are smaller than 0.65 Å (Table S6.1).

After swapping ApcE and ApcD in the structure, we next performed homology modeling of cApcC, which belongs to the CpcD superfamily and upholds a cylinder in shape [19]. The structure was based on an X-ray crystallographic study of a trimeric ApcC from *Fischerella sp.* PCC 6703 (PDBID: 1B33) [19]. The SPICKER program, an algorithm to discern near-native models from a pool of decoyed protein structures [31], converged the simulation and identified only one large cluster, indicating the high quality of the generated models. The alignment of cApcC onto GpApcC shows little differences in the RMSD value (Table S6.1), indicative of a good model. The copy numbers of ApcC, however, in the characterized GpPBS structure are two per PBS, surprisingly low probably because there is one missing trimeric disc (APC trimer) on each basal cylinder and two discs (APC trimers) missing on the top cylinder where the ApcC subunit usually binds [19]. It has been accepted that in cyanobacteria each core cylinder has two ApcC proteins located at both ends [19]. We substituted the existing GpApcC with cApcC in the cApcA/B trimer (note that GpApcA/B trimer has been swapped by cApcA/B trimer). The complex of cApcC/cApcA/B was then regarded as an independent object and aligned onto the end of each core cylinder, resulting in A0, A′0, B3 and B0 discs containing cApcC.

ApcF, an APC β-like unit, affects state transitions and energy transfer through interactions with α^{LCM} [32]. cApcF is modelled at the same trimeric disc where the α^{LCM} is located (A2 and A′2) with replacement of the GpApcF. Thus, cApcF interacts with α^{LCM}, cApcE-LD1, and potentially with the nearby pigment, α^{LCM} phycocyanobilin [18, 32], to allow energy transmission to PSII [33]. Only one model was generated and its RMSD value is less than 1 Å, indicative of an excellent modeling (Table S6.1). By combining the protein sequence analysis of two ApcFs from GpPBS and *Synechocystis* 6803, we noticed some minor differences in detail (Fig. S6.3): an indel (insertion/deletion) around 139-142 between two structures. In cyanobacteria, this region is comprised of four amino acid residues less than GpApcF. This indel is in a fragment of an α-helix (G-H) in cyanobacterial ApcF. For the red algal PBS (PDBID: 5Y6P), ApcF has an extension at the tip of the G-H helix hairpin [18]. ApcF is considered to be an important factor that fine tunes the spectroscopic properties of the PCB in α^{LCM}. We hypothesize that the combination of an indel and a fragment of an α-helix extension, instead of affecting the spectroscopic properties of α^{LCM}, could more likely help each basal cylinder build up its polarity so that only a trimer (A1/A′1, α3β3) can bind with a back-to-back configuration. In terms of the assembly of trimer A2/A′2, we propose that the linker domain (i.e., ApcE-LD1) plays an important role to recruit ApcF to build (A2/A′2) trimers and to exclude the joining of a β unit. Further biochemical and spectroscopic research on this region by using genetically tractable cyanobacteria may shed light on its function.

6.4.2 Chemical Cross-linking and Strategy

Chemical cross-linking coupled with mass spectrometry (MS) provides information of proximal amino acid residues within proteins and protein complexes. It is particularly useful for protein complexes that are recalcitrant to conventional structural biology studies. Successful observation

of a red algal PBS structure at the atomic level using cryo-EM took advantage of a particularly stable PBS species that allows PBS sample preparation for cryo-EM analysis at low salt condition. The cyanobacterial PBS, however, is much less stable, and the structure at the atomic level remains unclear. Here, we used residue-level chemical cross-linking to provide structural restraints to justify and evaluate a protein complex model that we generated by using bioinformatics methods (Fig. 6.1E).

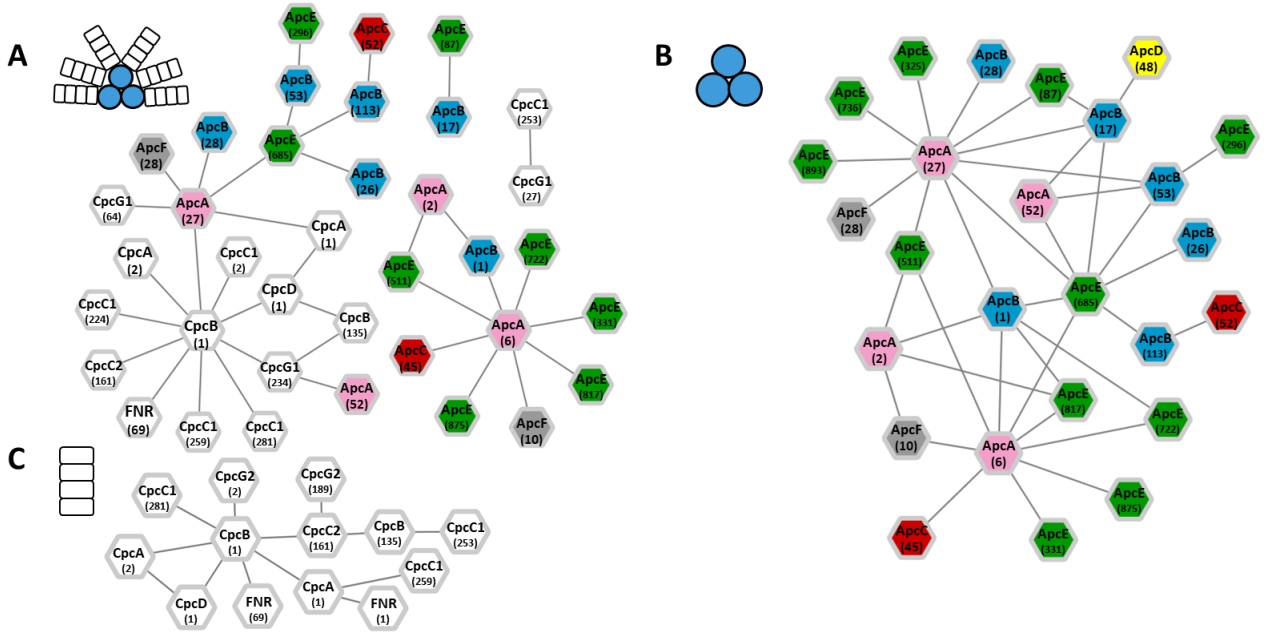


Figure 6.2 Identified cross-links in different types of cPBS. The subunit interaction networks, prepared by Cytoscape [34], are shown for (A) wild type cPBS (WT-cPBS) and two mutants including (B) CK-cPBS and (C) CpcL-cPBS. Cartoon representation of each PBS type is shown in the inset of each panel (top left), where WT-cPBS consists of six rods and a tricylindrical core. CK-cPBS contains only a tricylindrical core and CpcL-cPBS contains only the rods. The numbers represent the location of the cross-linked residues in each subunit. ApcA (light pink), ApcB (dark blue), ApcC (red), ApcD (yellow), ApcE (dark green), and ApcF (grey).

In this experiment, we used wild type PBS of cyanobacteria *Synechocystis* 6803 (WT-cPBS) and two mutant PBS for cross-linking analysis (Fig. 6.2 and Fig. S6.4). In WT-cPBS, we identified a total of 36 cross-links, in which 19 inter-subunit cross-links are within the PBS-core (Fig. 6.2A

and Table S6.2). We also found numerous loop-links and mono-links (data not shown) that provide limited structural information for subunit-subunit interactions. In contrast, when CK-cPBS, a rod-less PBS [20], was cross-linked and submitted to MS interrogation, 38 unique cross-links were identified (Fig. 6.2B, Table S6.3). Besides the 19 inter cross-links identified in the WT-cPBS sample, 19 more cross-links have been found in CK-cPBS. The differences in identified cross-links between WT-cPBS and CK-cPBS are probably due to the shielding effect in WT-cPBS by six outward radiating PC rods that have extensive contacting interfaces and large steric hindrance to limit the accessibility of cross-linkers to core components. A CpcL-cPBS mutant [21] containing only the rods was also used for a negative control. For the CpcL-cPBS mutant (Fig. 6.2C, Table S6.4), we identified 13 cross-links of rod-subunits and associated proteins, especially, ferredoxin-$NADP^+$ oxidoreductase (FNR_L) that has been successfully located in two types of PBSs in our recent research [28]. The obtained cross-linking network allows distinct distance restraints for constructing and validating the cPBS-core model (Fig. 6.1E).

6.4.3 Model Evaluation by Chemical Cross-links

BS^3-H_{12}/D_{12} is one of the most widely used cross-linkers, possessing a spacer length of 11.4 Å. The use of an isotopically encoded cross-linker such as this increases the confidence and accuracy of cross-linked peptide identification, with advantages for studying large protein complexes compared to non-isotope encoded cross-linkers. Each end of the cross-linker can form a covalent bond with a primary amine (lysine side chain or protein N-terminus). If two such functional groups are within reach of ~30 Å, the cross-linker tends to react with either/both and to yield a mono-link or a cross-link, which can be identified by LC-MS/MS, revealing structural information of those two functional groups [35].

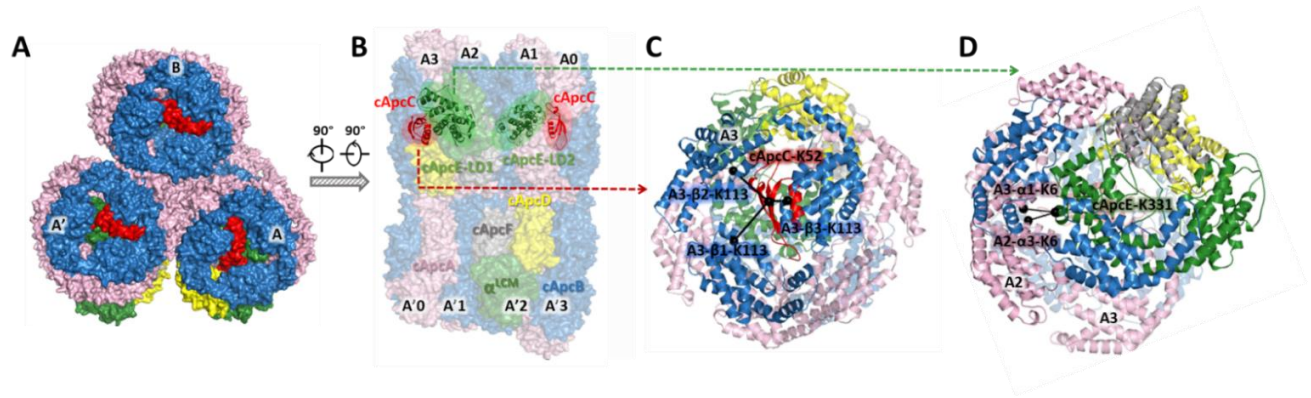


Figure 6.3 The cPBS-core structure mapped with observed cross-links. (A) Side view (cylinder perpendicular to the observer) and (B) bottom view of the cPBS-core, which is further mapped with (C) cross-link cApcC-K52-cApcB-K113, and (D) cross-link cApcE-K331-cApcA-K6. cApcA (light pink), cApcB (dark blue), cApcC (red), cApcD (yellow), cApcE (dark green), cApcF (grey).

In the cross-linked CK-cPBS sample, we successfully identified a total of 38 PBS-core subunit interactions (representative MS and product-ion (MS/MS) spectra of cross-linked peptides are shown in Fig. S6.5 and S6.6). We then mapped them individually to evaluate the PBS-core model. For better visualization, the core structure of cPBS is shown in the side view and bottom view, respectively (Fig. 6.3A and 6.3B). For example, the cross-link between cApcC-K52 and cApcB-K113 (Fig. 6.3C and Table 6.1) can be located at either end APC trimer of each cylinder (i.e., A0, A3, A′0, A′3, B0, B3). We note that cApcC-K52 is surrounded by three cApcB-K113 in the APC trimer (with C_3 rotational symmetry). This arrangement gives reasonable cross-linked distances on all three cApcB copies (i.e., 20.0 Å, 18.5 Å and 11.2 Å (Table 6.1)).

Table 6.1 Summary of two representative cross-links, including the information of possible cross-linking sites located on different copies of cApcA/B discs and the measured Euclidean Cα-Cα distances (< 30 Å)

Inter-molecular Cross-links	Cross-linked Subunit on cApcA/B Discs	Cross-linked Distance, Cα-Cα, Å
cApcC52-cApcB113	A3(or A0/A'3/A'0/B3/B0)-β1	20.0
	A3(or A0/A'3/A'0/B3/B0)-β2	18.5
	A3(or A0/A'3/A'0/B3/B0)-β3	11.2
cApcE331-cApcA6	A3(or A0/A'3/A'0/B3/B0)-α1	19.4
	A2 (or A'2)-α3	12.7

Another example cross-link is between cApcE-LD1 and cApcA, which is formed towards the central cavity of the basal cylinder. There are two possible cross-linked sites with cApcE-K331: one is A3-α1-cApcA-K6 and the other is A2-α3-cApcA-K6 (Fig. 6.3D). (Note the nomenclature (disc (A3 or A2), first (1 or 3) α subunit, lysine (6)). The estimated cross-linked distances fit well within the 30 Å threshold [36] (Table 6.1). Overall, 34 out of the 38 cross-links bridge pairs of amino acids whose Cα-Cα distances are less than 30 ± 1.5 Å , 20 of which exhibit more than one possible cross-linked site owing to the redundancy of ApcA/B in each disc; the multiple chances of forming structurally relevant cross-links in those cases testifies the reliability of the constructed cPBS-core model.

Two of the other four cross-links are associated with ApcE-685, a residue that locates on the loop region connecting ApcE-LD2 and ApcE-LD3. The cross-linked distances are slightly larger than 30 ± 1.5 Å, which could result from the dynamic nature of the loop (Table S6.5). The other two cross-links showing extra-long cross-linked distances are both related to ApcE-K87 in the middle of a flexible loop on cApcE, i.e. cApcE87-cApcA27 in CK-cPBS and cApcE87-cApcB17 in both of WT-cPBS and CK-cPBS (Fig. 6.2A, 6.2B). The cross-linked distance is around 50 Å for the former one and 43 Å for the latter (Table S6.5). The PB-loop, consisting of 58 amino acids with primarily basic residues, has been proposed as an anchoring arm attached to the

thylakoid membrane, possibly through electrostatic interaction [2, 37]. Consequently, the PB-loop can be highly flexible in the absence of the membrane, resulting in the cross-links that cannot form if the PB-loop is anchored on the membrane/reaction centers. In the cryo-EM structure, the PB-loop is not observed [18], possibly owing to the flexible nature of this domain. It should be noted that without the binding partners of the PB-loop, i.e., the membrane/reaction centers, the structural prediction using any program seems uncertain [13]. The cross-linking chemistry and cryo-EM structure may capture and reflect different protein conformers in terms of the loop orientation. To understand its precise location and functionality, various mutants under different chemical environments are needed for adequate cross-linking restraints, which will facilitate the downstream computational simulation. Overall, selection of CK-PBS represented a successful strategy to pinpoint the subunit interaction network. The cross-linked distances supported very well the cPBS-core structure proposed by computational prediction.

6.5 Discussion

The PBS-core in *Synechocystis* 6803 represents a classical tri-cylinder core with C_2-symmetry [7], two of which are basal cylinders that contain various APC units (with copy numbers), i.e. ApcA (32), ApcB (34), ApcD (2), ApcE (2), ApcF (2) and ApcC (6). ApcD, ApcE, ApcF, and ApcC are arranged in mirror positions in two basal cylinders. The symmetrical feature of cPBS correlates well with that of a dimeric PSII, the most active form in cyanobacteria, red algae and higher plants. Captured energy is funneled from the cPBS rod to the cPBS-core unidirectionally and then reaches the terminal energy emitters (TEE), ApcE and ApcD [38], where excitation energy is finally transmitted to Chl*a* in PSII and PSI for photochemical reactions [12]. The interface between the cPBS-core structure and RCs is of great interest in understanding the energy transmission mechanism and dynamic process under varying environmental conditions.

Traditionally, two basal cylinders have been described as antiparallel and evenly covering the thylakoid membrane surface/reaction centers. Previous research techniques indeed have never been able to elucidate the detailed tomography of two basal cylinders relative to the thylakoid membrane surface. Our proposed tri-cylindrical PBS-core model, however, reveals that the two basal cylinders do not adopt a perfect antiparallel configuration, rather, the two basal cylinders are arranged in an acute X-shape at an angle of ~ 24° relative to each other (Fig. 6.4A). We start with our previously proposed PBS-PSII model [13] constrained by the structural proteomic data [2] and further propose that each basal cylinder is tilted 12° relative to the thylakoid membrane (Fig. 6.4B and 6.4C). Specifically, ApcD, from the discs A3/A′3, and α^{LCM}, from the discs A2/A′2, comprise a protrusion that preferentially touches the surface of thylakoid membrane. The distal disc A0/A′0 tilts away from the thylakoid membrane surface (Fig. 6.4A and 6.4C). This orientation leaves an open space between the thylakoid membranes and discs A1/A′1 and A0/A′0 (Fig. 6.4B and 6.4C). The extra discs on both the top cylinder and two basal cylinders may reflect the functional connection difference between rods and PBS-core in the block-shaped PBS (*Griffithsia pacifica*) and the hemidiscoidal PBS (*Synechocystis* 6803). It seems that our cyanobacterial tricylindrical model has increased surface area for attachment of (six) rods, in contrast to that of *Griffithsia*, which has 14 rods attached to a smaller PBS-core, in a much more compact manner. It seems, though, that six rods on the PBS-core in cyanobacteria may have more flexibility relative to each other. This may partially explain why no high resolution cryo-EM structure is yet available in cyanobacteria due to the lack of sample homogeneity.

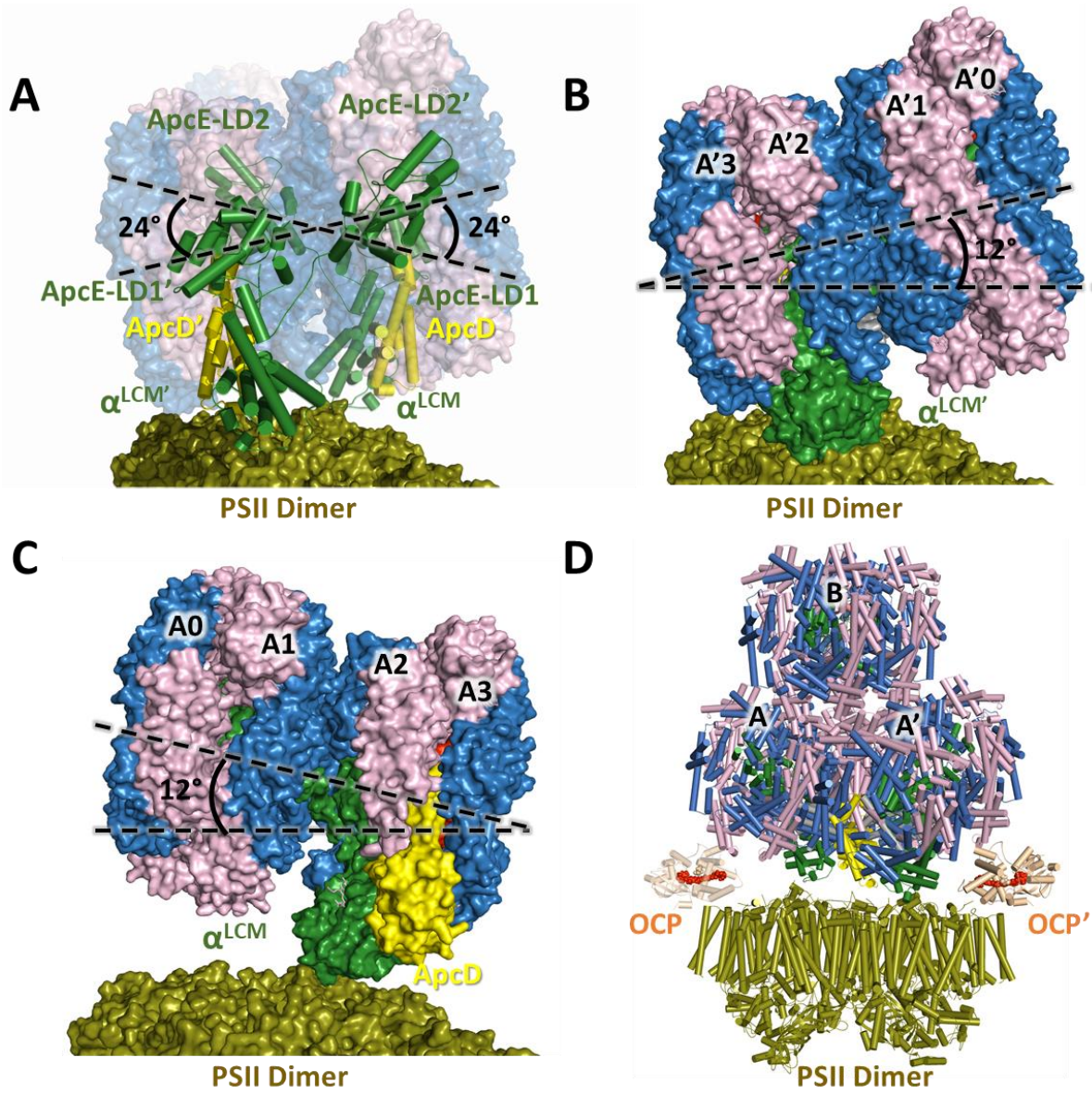


Figure 6.4 The cPBS-core structure reveals an acute X-shape of the two basal cylinders. (A) Cross-over (side view) of the two basal cylinders, front relative to another (back). Dashed lines show the general trend of the two sets of linker domains, adopting an acute 24° angle relative to each other. (B) Front basal cylinder (A′0-A′3) and (C) back basal cylinder (A0-A3) relative to PSII dimer. Dashed lines show the general trend of the cylinder with an acute 12° angle. (D) Side view of cPBS-PSII complex, showing two cavities between uneven PBS basal cylinder and PSII that allows OCPs to interact at the putative geometry. ApcE domain including α^{LCM}, ApcE-LD1 and ApcE-LD2 (dark green), ApcD (yellow), OCP (orange), PSII dimer (split-pea).

A previous study particularly indicated that ApcE-K87 tends to interact with CP47-K227 from PSII [2], and could possibly be one of the critical structural basis dictating excitation energy migration from PBS to the reaction centers, specifically, PSII. With regard to PSI, a chemical cross-link between the N-terminal of PSI and PBS-core has also been reported [2], namely, PsaA-30^{K} - ApcB-17^{K}. Pigment protein complexes involved in light energy capture and chemical conversion are not evenly distributed in the thylakoid membranes [5-6, 39]. In some of thylakoid membrane regions, PSII is enriched, whereas in other regions, PSI is aggregated in patches that allows biochemical isolation and characterization [5]. The reducing side of PSII and PSI have their own characteristics; unlike the flat surface at the PSII reducing side, PsaC, PsaD, PsaE of the PSI reducing side comprise a protrusion collectively protecting the iron sulfur center.

The acute X-shape configuration at the bottom of the PBS-core provides structural accommodations for other regulatory factors on either the PBS-core or the reducing side of PSII. Orange carotenoid protein (OCP), for example, is a critical photoprotective protein that dissipates excess light energy under stress conditions and disarms harmful reactive oxidative species arising from photosynthetic machinery. It is commonly held that the active form of OCP (OCP^{R}) binds to core subunits in cPBS as the primary energy quenching sites [40]. ApcE has been proposed to be a promising target for OCP^{R} [41], given the unique shift of its bilin-binding pocket compared to that of the surrounding ApcA/B subunits, which are all conserved Cys81 bilin sites. The tilted angle of each basal cylinder provides sufficient room for the 34-kDa OCP to move in and bind and, perhaps more importantly, to accommodate the fluorescence recovery protein (FRP) [40], a 14 kDa protein that can accelerate the detachment of OCP from its binding sites followed by subsequent deactivation, and other essential regulators in the photoprotection process. In our model, the extra space between two face-to-face trimers A1/A0 (or A′1/A′0) and

the PSII reducing side can accommodate the entrance of OCP without any spatial conflicts (Fig. 6.4D).

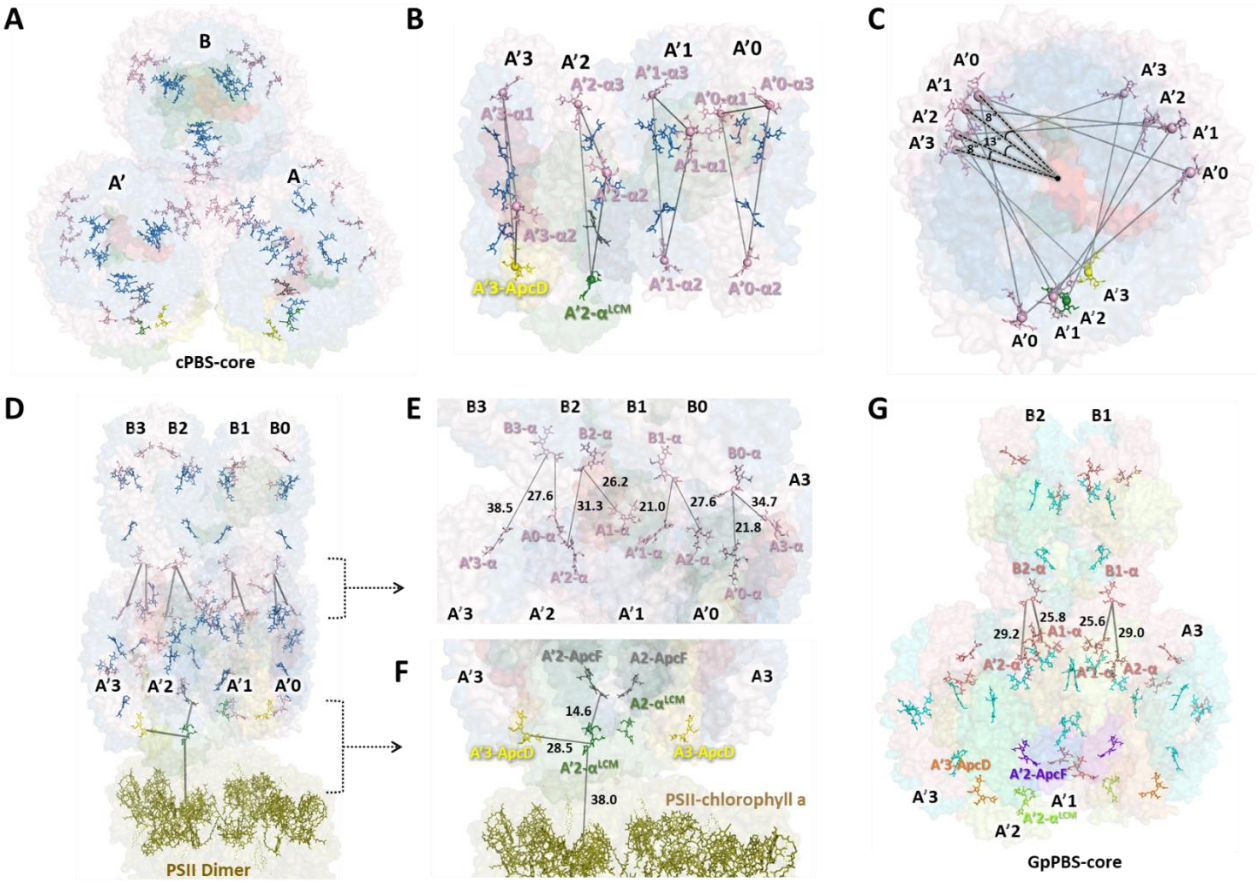


Figure 6.5 cPBS containing bilins and energy transfer pathways. (A) Overview of cPBS-core containing bilins in a side view (cylinder perpendicular to the observer). (B) Enlarged A′ cylinder with bilins twisting in a clockwise fashion, shown in a side view (cylinder parallel to the observer) and (C) in a side view (cylinder perpendicular to the observer). (D) Overview of excitation energy transfer pathway in cPBS-core-PSII complex in a side view (cylinder parallel to the observer). (E) Enlarged interface between B0 and A′ cylinder. (F) Enlarged interface between A′ cylinder and the PSII. (G) Overview of GpPBS-core containing bilins and representative energy transfer pathway [18]. Bilins in cPBS-core: ApcA (light pink), ApcB (dark blue), ApcE domains (dark green), ApcD (yellow) and ApcF (grey). Bilins in GpPBS-core: ApcA (salmon), ApcB (cyan), ApcD (light orange), ApcE domains (limon), ApcF (purple). The PSII dimer is colored in split-pea.

In our model, plausible excitation energy transfer routes can be predicted in cPBS-core and cPBS-core-PSII supercomplex as well. In the red algal structure, there are 48 bilins in the core structure (Fig. 6.5G) [18]. In our structure, four extra trimers were modelled onto the red algal PBS-core (i.e., B0, B4, A0, and A′0 (Fig. 6.1A)), giving a total of 72 bilins in a trycylindrical PBS-core (Fig. 6.5A). The model shows that if the four extra trimers are removed, the excitation-energy-transfer distances between each bilin are comparable to those of red algal. The staggered trimer orientation between two cylinders was retained, similar to that of the red algal structure (Fig. 6.4A). In each basal cylinder, we noticed, interestingly, that each trimer adopts a specific angle twist (clockwise) relative to the previous trimer (Fig. 6.5B and 6.5C), a feature that has not been noticed previously [18]. Each twist angle may seem small (8° of A′2 vs A′3, 13° of A′1 vs A′2, and 8° of A′0 vs A′1); however, the twists accumulate to afford a close orientation of the B0 chromophore (B0-α) to that of A′0 (A′0-α), rather than to the α chromophore on another basal cylinder A3 (Fig. 6.5D and 6.5E). The smallest distance between B0-α and A′0–α is estimated to be 21.8 Å, in contrast to that of B0-α and A3-α, which is 34.7 Å (Fig. 6.5E). Both distances can allow energy transfer to occur; the former, however, may be more efficient. The distance between bilins A′2-β^{ApcF} and A′2-α^{LCM} is 14.6 Å (Fig. 6.5F), the shortest distance between bilin pairs in the core, consistent with the value in red algal PBS [18]. The distance between bilin A′2-α^{LCM} and A′3-ApcD is 28.5 Å. (Note that we measured the conjugation bond system rather than molecule center distance (Fig. 6.5)).

For our previously reported PBS-core-PSII complex, the energy transfer routes can also be predicted from our new model of bilin in PBS to chlorophyll *a* (Chl*a*) in PSII. The shortest distance between the bilin of α^{LCM} and a Chl *a* in CP47 is 38 Å. (Note that the PB-loop in ApcE was not resolved in previous studies, the real distance remains a subject for future research). The

distance between bilin of α^{LCM} and ApcD is 28.5 Å. Interestingly, the distance of bilin of ApcD and Chl *a* in CP47 is in a comparable range as that of the α^{LCM} bilin. It appears that, on one side of the cPBS-core, energy transfers from bilins on B0 to A'0 (B3 to A0) and then horizontally transfers to ApcE where it feeds into the Chl *a* in CP47. Given that PBS-core and PSII share the same *C2* symmetry axis, there is another energy transfer route along B3 to A0, then to A2-α^{LCM} and ends in Chl *a* in CP47, the mirror component in another PSII monomer.

In summary, we constructed the first detailed cPBS-core structure of Cyanobacteria *Synechocystis* PCC6803 to reveal an acute X-shape of the two basal cylinders. Our study provides a structural basis for energy migration and its related regulation associated with the transient association between PBS and RCs. The proximate location of the pigment proteins in the functional complex is essential to promote efficient energy transfer and regulation.

6.6 Acknowledgements

This work is a joint effort by all authors. M.M.Z. analyzed the data and wrote the manuscript, D.A.W. and M.C. analyzed the data, H.B.P., M.L.G., and R.E.B. wrote and edited the manuscript. H.L. and R.E.B designed the research. H.L. performed and analyzed the data.

We thank Ghada Ajlani for the CK and Δ*AB Synechocystis* mutants and helpful discussions. This research was funded by the Photosynthetic Antenna Research Center (PARC), an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Office of Basic Energy Sciences (Grant DE-SC0001035). Support was also provided by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Photosynthetic Systems Program (Grant DE-FG02-07ER15902 to R.E.B and H.L. and Grant DE-FG02-99ER20350 to H.B.P.). Partial support was also provided by the National Institute of

General Medical Sciences of the National Institutes of Health (Grant 2P41GM103422 to M.L.G.).

6.7 References

1. Mullineaux, C. W. Electron transport and light-harvesting switches in cyanobacteria. *Front. Plant Sci.* **2014**, *5*, 7.

2. Liu, H., Zhang, H., Niedzwiedzki, D.M., Prado, M., He, G., Gross, M.L., Blankenship, R.E. Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria. *Science* **2013**, *342*, 1104-1107.

3. Adir, N.; Bar-Zvi, S.; Harris, D. The amazing phycobilisome. *Biochim. Biophys. Acta* **2020**, *1861*, 148047.

4. de Marsac, N. T.; Cohen-bazire, G. Molecular composition of cyanobacterial phycobilisomes. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 1635-9.

5. MacGregor-Chatwin, C.; Sener, M.; Barnett, S. F. H.; Hitchcock, A.; Barnhart-Dailey, M. C.; Maghlaoui, K.; Barber, J.; Timlin, J. A.; Schulten, K.; Hunter, C. N. Lateral Segregation of Photosystem I in Cyanobacterial Thylakoids. *Plant Cell* **2017**, *29*, 1119-1136.

6. Ho, M. Y.; Niedzwiedzki, D. M.; MacGregor-Chatwin, C.; Gerstenecker, G.; Hunter, C. N.; Blankenship, R. E.; Bryant, D. A. Extensive remodeling of the photosynthetic apparatus alters energy transfer among photosynthetic complexes when cyanobacteria acclimate to far-red light. *Biochim. Biophys. Acta. Bioenerg.* **2020**, *1861*, 148064.

7. Arteni, A. A.; Ajlani, G.; Boekema, E. J. Structural organisation of phycobilisomes from Synechocystis sp. strain PCC6803 and their interaction with the membrane. *Biochim. Biophys. Acta* **2009**, *1787*, 272-279.

8. Glazer, A. N. Light guides. Directional energy transfer in a photosynthetic antenna. *J. Biol. Chem.* **1989**, *264*, 1-4.

9. Bald, D.; Kruip, J.; Rogner, M. Supramolecular architecture of cyanobacterial thylakoid membranes: How is the phycobilisome connected with the photosystems? *Photosynth. Res.* **1996**, *49*, 103-18.

10. MacColl, R. Cyanobacterial phycobilisomes. *J. Struct. Biol.* **1998**, *124*, 311-334.

11. Zlenko, D. V.; Krasilnikov, P. M.; Stadnichuk, I. N. Structural modeling of the phycobilisome core and its association with the photosystems. *Photosynth. Res*. **2016**, *130*, 347-356.

12. Mullineaux, C. W. Phycobilisome-reaction centre interaction in cyanobacteria. *Photosynth. Res.* **2008**, *95*, 175-82.

13. Liu, H.; Blankenship, R. E. On the interface of light-harvesting antenna complexes and reaction centers in oxygenic photosynthesis. *Biochim. Biophys. Acta Bioenerg*. **2019**, *1860*, 148079.

14. Sidler, W. A., Phycobilisome and phycobiliprotein structures. In The molecular biology of cyanobacteria, Springer: 1994; pp 139-216.

15. Yamanaka, G.; Glazer, A.; Williams, R. Molecular architecture of a light-harvesting antenna. Comparison of wild type and mutant Synechococcus 6301 phycobilisomes. *J. Biol. Chem*. **1980**, *255*, 11004-11010.

16. Glauser, M.; Bryant, D. A.; Frank, G.; Wehrli, E.; Rusconi, S. S.; Sidler, W.; Zuber, H., Phycobilisome structure in the cyanobacteria Mastigocladus laminosus and Anabaena sp. PCC 7120. *Euro. J. Biochem*. **1992**, *205*, 907-915.

17. Elmorjani, K.; Thomas, J.-C.; Sebban, P. Phycobilisomes of wild type and pigment mutants of the cyanobacterium Synechocystis PCC 6803. *Arch. Microbiol*. **1986**, *146*, 186-191.

18. Zhang, J.; Ma, J.; Liu, D.; Qin, S.; Sun, S.; Zhao, J.; Sui, S.-F. Structure of phycobilisome from the red alga Griffithsia pacifica. *Nature* **2017**, *551*, 57-63.

19. Reuter, W.; Wiegand, G.; Huber, R.; Than, M. E. Structural analysis at 2.2 Å of orthorhombic crystals presents the asymmetry of the allophycocyanin–linker complex, AP· LC7. 8, from phycobilisomes of Mastigocladus laminosus. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1363-1368.

20. Thomas, J. C.; Ughy, B.; Lagoutte, B.; Ajlani, G. A second isoform of the ferredoxin:NADP oxidoreductase generated by an in-frame initiation of translation. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 18368-73.

21. Ajlani, G.; Vernotte, C.; Dimagno, L.; Haselkorn, R. Phycobilisome Core Mutants of Synechocystis Pcc-6803. *Biochim. Biophys. Acta. Bioenerg.* **1995**, *1231*, 189-196.

22. Marx, A.; Adir, N. Allophycocyanin and phycocyanin crystal structures reveal facets of phycobilisome assembly. *Biochim. Biophys. Acta*. **2013**, *1827*, 311-318.

23. Peng, P. P.; Dong, L. L.; Sun, Y. F.; Zeng, X. L.; Ding, W. L.; Scheer, H.; Yang, X.; Zhao, K. H. The structure of allophycocyanin B from Synechocystis PCC 6803 reveals the structural basis for the extreme redshift of the terminal emitter in phycobilisomes. *Acta Crystallogr. D Biol. Crystallogr*. **2014**, *70*, 2558-2269.

24. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725-38.

25. Houmard, J.; Capuano, V.; Colombano, M. V.; Coursin, T.; Tandeau de Marsac, N. Molecular characterization of the terminal energy acceptor of cyanobacterial phycobilisomes. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 2152-6.

26. Ma, J.; You, X.; Sun, S.; Wang, X.; Qin, S.; Sui, S.-F. Structural basis of energy transfer in Porphyridium purpureum phycobilisome. *Nature* **2020**, *579*, 146-151.

27. Tang, K.; Ding, W. L.; Hoppner, A.; Zhao, C.; Zhang, L.; Hontani, Y.; Kennis, J. T.; Gartner, W.; Scheer, H.; Zhou, M.; Zhao, K. H. The terminal phycobilisome emitter, LCM: A light-harvesting pigment with a phytochrome chromophore. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15880-5.

28. Liu, H., Weisz, D, A., Zhang, M.M., Cheng, M., Zhang, B., Zhang, H., Gerstenecker, G.S., Pakrasi, H.B., Gross, M. L., Blankenship, R. E. Phycobilisomes Harbor FNRL in Cyanobacteria. *mBio* **2019**, *10*, e00669-19.

29. Chang, L.; Liu, X.; Li, Y.; Liu, C. C.; Yang, F.; Zhao, J.; Sui, S. F. Structural organization of an intact phycobilisome and its association with photosystem II. *Cell Res.* **2015**, *25*, 726-737.

30. Dong, C.; Tang, A.; Zhao, J.; Mullineaux, C. W.; Shen, G.; Bryant, D. A. ApcD is necessary for efficient energy transfer from phycobilisomes to photosystem I and helps to prevent photoinhibition in the cyanobacterium Synechococcus sp. PCC 7002. *Biochim. Biophys. Acta. Bioenerg.* **2009**, *1787*, 1122-1128.

31. Zhang, Y.; Skolnick, J. SPICKER: a clustering approach to identify near-native protein folds. *J. Comput. Chem.* **2004**, *25*, 865-71.

32. Calzadilla, P. I.; Muzzopappa, F.; Sétif, P.; Kirilovsky, D. Different roles for ApcD and ApcF in Synechococcus elongatus and Synechocystis sp. PCC 6803 phycobilisomes. *Biochim. Biophys. Acta. Bioenerg.* **2019**, *1860*, 488-498.

33. Gindt, Y. M.; Zhou, J.; Bryant, D. A.; Sauer, K. Spectroscopic studies of phycobilisome subcore preparations lacking key core chromophores: assignment of excited state energies to the Lcm, β18 and αAP-B chromophores. *Biochim. Biophys. Acta. Bioenerg*. **1994**, *1186*, 153-162.

34. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. **2003**, *13*, 2498-2504.

35. Sinz, A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Rev. Proteom*. **2014**, *11*, 733-743.

36. Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine–lysine distances. *Protein Sci.* **2014**, *23*, 747-759.

37. Zlenko, D. V.; Krasilnikov, P. M.; Stadnichuk, I. N. Role of inter-domain cavity in the attachment of the orange carotenoid protein to the phycobilisome core and to the fluorescence recovery protein. *J. Biomol. Struct. Dyn.* **2016**, 34, 486-496.

38. Bryant, D. A.; Guglielmi, G.; de Marsac, N. T.; Castets, A.-M.; Cohen-Bazire, G. The structure of cyanobacterial phycobilisomes: a model. *Arch. Microbiol*. **1979**, *123*, 113-127.

39. Folea, I. M.; Zhang, P.; Aro, E. M.; Boekema, E. J. Domain organization of photosystem II in membranes of the cyanobacterium Synechocystis PCC6803 investigated by electron microscopy. *FEBS Lett.* **2008**, *582*, 1749-54.

40. Kirilovsky, D.; Kerfeld, C. A. Cyanobacterial photoprotection by the orange carotenoid protein. *Nature Plants* **2016**, *2*, 16180.

41. Stadnichuk, I. N.; Yanyushin, M. F.; Maksimov, E. G.; Lukashev, E. P.; Zharmukhamedov, S. K.; Elanskaya, I. V.; Paschenko, V. Z. Site of non-photochemical quenching of the phycobilisome by orange carotenoid protein in the cyanobacterium Synechocystis sp. PCC 6803. *Biochim. Biophys. Acta. Bioenerg*. **2012**, *1817*, 1436-1445.

Supplementary Information

Table S6.1 Alignment between different cPBS domains and the corresponding components in GpPBS. All predicted homology models are included. Three cylinders are evaluated independently based on the location of various subunits. The RMSD value is calculated among all Cα in the two corresponded structures given by PyMOL software.

		Alignment with corresponding components in GpPBS, RMSD of all Cα (Å)		
		Cylinder A	**Cylinder A′**	**Cylinder B**
cApcA/B		0.622/0.613	0.787/0.569	0.616/0.745
cApcC		0.621/0.621	0.582/0.582	0.621/0.582
cApcD		0.645	0.644	-
cApcE′ (or α^{LCM})	**M1**	0.379	0.380	-
	M2	0.384	0.386	-
	M3	0.460	0.454	-
	M4	0.309	0.312	-
	M5	0.469	0.471	-
cApcE_LD1	**M1**	0.345	0.359	-
	M2	0.696	0.69	-
	M3	1.0341	1.03	-
	M4	0.277	0.277	-
	M5	1.001	1.018	-
cApcE_LD2	**M1**	0.264	0.283	-
	M2	0.714	0.735	-
	M3	0.249	0.285	-
	M4	0.243	0.277	-
	M5	0.253	0.293	-
cApcE_LD3	**M1**	-	-	0.222/ 0.232
	M2	-	-	1.136/ 1.175
	M3	-	-	1.166/ 1.199
	M4	-	-	0.178/ 0.195
	M5	-	-	1.304/ 1.322
cApcF	**M1**	0.653	0.653	-

Table S6.2 MS-information of the inter cross-links in WT-PBS.

Cross-linked Sequence	Cross-linked Protein	Charge	m/z	Detected peptides in Raw File
MLGQSSLVGYSNTQAANR(1)-MKTPLTEAVSTADSQGR(1)	cpcD (1)-cpcA (1)/	3	1280.666667	072815_PBS_X02_01.22561.22561.3.0.dta
IKAFVTGGAAR(2)-MFDVFTR(1)	ApcA (27)-cpcB (1)/	3	714	072815_PBS_X02_01.25378.25378.3.0.dta
SNKAVIVPFEQLNQTLQQINR(3)-MFDVFTR(1)	cpcC1 (259)-cpcB (1)/	3	1164	072815_PBS_X02_01.33652.33652.3.0.dta
SINPAANTIPKVSAQNINIEASVPR(11)-MKEAALDIVNDPNGITR(2)	cpcG1 (234)-cpcB (135)/	4	1150.75	072815_PBS_X02_01.23144.23144.4.0.dta
SINPAANTIPKVSAQNINIEASVPR(11)-ETIVKQAGDR(5)	cpcG1 (234)-ApcA (52)/	4	965.75	072815_PBS_X02_01.19158.19158.4.0.dta
GPAVNNQVGNPSAVGEFPGSLGAKVFR(24)-SIVTKSIVNADAEAR(5)	ApcE (511)-ApcA (6)/	4	1095.75	072815_PBS_X02_01.24502.24502.4.1.dta
AGNTPAKALGGTVPFGQASK(7)-MFDVFTR(1)	cpcC1 (224)-cpcB (1)/	3	974	073015_PBS_X02_02.14321.14321.3.0.dta
MKTPLTEAVSTADSQGR(1)-IKAFVTGGAAR(2)	cpcA (1)-ApcA (27)/	4	758.25	073015_PBS_X02_03.9412.9412.4.0.dta
EKVLESQLR(2)-IKAFVTGGAAR(2)	cpcG1 (64)-ApcA (27)/	4	583.25	072815_PBS_X02_02.8220.8220.4.0.dta
NAMDELKAYFESGSAR(7)-IKAFVTGGAAR(2)	ApcF (28)-ApcA (27)/	4	754.879	073015_PBS_X02_01.14118.14118.4.0.dta
IKAFVTGGAAR(2)-VDKEVTPR(3)	ApcA (27)-ApcE (685)/	4	543.5555	073015_PBS_X02_01.6911.6911.4.1.dta
SIVTKSIVNADAEAR(5)-KQFFEPFINSR(1)	ApcA (6)-ApcE (331)/	3	1041.883	073015_PBS_X02_02.12424.12424.3.0.dta
MQDAITAVINSADVQGK(1)-SIVTKSIVNADAEAR(5)	ApcB (1)-ApcA (6)/	3	1161.958	073015_PBS_X02_02.15131.15131.3.0.dta
MQDAITAVINSADVQGK(1)-SIVTKSIVNADAEAR(1)	ApcB (1)-ApcA (2)/	3	1157.935333	073015_PBS_X02_02.15156.15156.3.0.dta
DAVTTLIKNYDLTGR(8)-SIVTKSIVNADAEAR(5)	ApcF (10)-ApcA (6)/	3	1130.938333	073015_PBS_X02_02.13022.13022.3.0.dta
APLNQKEIQQYNQILASQGLK(6)-SIVTKSIVNADAEAR(5)	ApcE (817)-ApcA (6)/	3	1366.069	072815_PBS_X02_02.11898.11898.4.0.dta
SIVTKSIVNADAEAR(5)-IMKMGGK(3)	ApcA (6)-ApcC (45)/	3	836.439333	073015_PBS_X02_02.6933.6933.3.0.dta
SINPAANTIPKVSAQNINIEASVPR(11)-MFDVFTR(1)	cpcG1 (234)-cpcB (1)/	3	1223.661333	073015_PBS_X02_02.15086.15086.3.0.dta
MLGQSSLVGYSNTQAANR(1)-MKEAALDIVNDPNGITR(2)	cpcD (1)-cpcB (135)/	3	1297.979333	073015_PBS_X02_02.12620.12620.

SIVTKSIVNADAEAR(5)-QQTKVFK(4)	ApcA (6)-ApcE (722)/	3	867.8326667	073015_PBS_X02_02.8367.8367.3.1.dta
MQDAITAVINSADVQGKYLDGAAMDK(17)-IFTGGSPLSYLEKPVER(13)	ApcB (17)-ApcE (87)/	4	1189.5975	073015_PBS_X02_03.15681.15681.4.0.dta
DITKAYSQSISYLESQVR(4)-AASVISANAATIVKEAVAK(14)	ApcE (296)-ApcB (53)/	4	1010.5395	073015_PBS_X02_03.15334.15334.4.0.dta
YLDGAAMDKLK(9)-VDKEVTPR(3)	ApcB (26)-ApcE (685)/	4	577.0525	073015_PBS_X02_03.7979.7979.4.0.dta
VLNGLKETYNSLGVPISSTVQAIQAIK(6)-IVKVELATGRPGTNAGLA(3)	ApcB (113)-ApcC (52)/	4	1190.929	073015_PBS_X02_03.15349.15349.4.1.dta
LKSYFASGELR(2)-IKAFVTGGAAR(2)	ApcB (28)-ApcA (27)/	4	625.3435	072815_PBS_X02_02.9518.9518.4.0.dta
MKTPLTEAVSTADSQGR(2)-MFDVFTR(1)	cpcA (2)-cpcB (1)/	3	948.7973333	073015_PBS_X02_02.14537.14537.3.0.dta
MLGQSSLVGYSNTQAANR(1)-MFDVFTR(1)	cpcD (1)-cpcB (1)/	3	983.8076667	073015_PBS_X02_02.16771.16771.3.0.dta
NQKTVGFSR(3)-MFDVFTR(1)	cpcC2 (161)-cpcB (1)/	3	701.043	073015_PBS_X02_02.12958.12958.3.0.dta
SIVTKSIVNADAEAR(5)-LYNKLTK(4)	ApcA (6)-ApcE (875)/	4	648.36025	072815_PBS_X02_02.8778.8778.4.1.dta
VEGYEIGSEEKPVVFTTENILSSSDMDNLIEAAYR(11)-VEITAISAPGYPKVR(13)	cpcG1 (27)-cpcC1 (253)/	4	1411.70575	073015_PBS_X02_03.16731.16731.4.2.dta
VLNGLKETYNSLGVPISSTVQAIQAIK(6)-VDKEVTPR(3)	ApcE (685)-ApcB (113)	4	981.7918	072815_PBS_X02_02.16035.16035.4.0.dta
AASVISANAATIVKEAVAK(14)-VDKEVTPR(3)	ApcB (53)-ApcE (685)/	4	724.406	072815_PBS_X02_02.11186.11186.4.0.dta
LGGKVASITPASLS(4)-MFDVFTR(1)	cpcC1 (281)-cpcB (1)/	3	785.0823333	072815_PBS_X02_01.29273.29273.3.0.dta
MFDVFTR(1)-AITTAASR(1)	cpcB (1)-cpcC1 (2)/	3	619.0056667	072815_PBS_X02_01.27051.27051.3.0.dta
MGGKIVSIK(4)-MFDVFTR(1)	FNR (69)-cpcB (1)/	3	662.3526667	072815_PBS_X02_01.25490.25490.3.0.dta
GPAVNNQVGNPSAVGEFPGSLGAKVFR(24)-SIVTKSIVNADAEAR(1)	ApcE (511)-ApcA (2)/	4	1098.8412	072815_PBS_X02_02.13760.13760.4.0.dta

Table S6.3 MS-information of the inter cross-links in CK-PBS.

Cross-linked Sequence	Cross-linked Protein	Charge	m/z	Detected peptides in Raw Files
DITKAYSQSISYLESQVR(4)-AASVISANAATIVKEAVAK(14)	ApcE (296)-ApcB (53)/	3	1350.6667	072815_CK_X02_01.28469.28469.3.0.dta
ETIVKQAGDR(5)-VDKEVTPR(3)	ApcA (52)-ApcE (685)/	3	734.0000	072815_CK_X02_01.8173.8173.3.0.dta
SIVTKSIVNADAEAR(5)-KQFFEPFINSR(1)	ApcA (6)-ApcE (331)/	3	1040.6667	072815_CK_X02_01.22159.22159.3.0.dta
MQDAITAVINSADVQGK(1)-SIVTKSIVNADAEAR(5)	ApcB (1)-ApcA (6)/	3	1160.6667	072815_CK_X02_01.26915.26915.3.0.dta
APLNQKEIQQYNQILASQGLK(6)-MQDAITAVINSADVQGK(1)	ApcE (817)-ApcB (1)/	3	1430.6667	072815_CK_X02_01.27925.27925.3.0.dta
NAMDELKAYFESGSAR(7)-IKAFVTGGAAR(2)	ApcF (28)-ApcA (27)/	3	1010.6667	072815_CK_X02_01.25575.25575.3.0.dta
DAVTTLIKNYDLTGR(8)-SIVTKSIVNADAEAR(5)	ApcF (10)-ApcA (6)/	3	1134.0000	072815_CK_X02_01.23068.23068.3.0.dta
LKSYFASGELR(2)-IKAFVTGGAAR(2)	ApcB (28)-ApcA (27)/	4	628.2500	072815_CK_X02_01.16640.16640.4.1.dta
MQDAITAVINSADVQGK(1)-IKAFVTGGAAR(2)	ApcB (1)-ApcA (27)/	3	1000.6667	072815_CK_X02_01.26095.26095.3.0.dta
APLNQKEIQQYNQILASQGLK(6)-SIVTKSIVNADAEAR(1)	ApcE (817)-ApcA (2)/	3	1370.6667	072815_CK_X02_01.20880.20880.4.0.dta
MQDAITAVINSADVQGKYLDGAAMDK(17)-IAETLAENEKK(10)	ApcB (17)-ApcD (48)/	4	1030.7500	072815_CK_X02_01.23209.23209.4.2.dta
MQDAITAVINSADVQGKYLDGAAMDK(17)-IKAFVTGGAAR(2)	ApcB (17)-ApcA (27)/	4	988.2500	072815_CK_X02_01.25173.25173.4.0.dta
MQDAITAVINSADVQGK(1)-SIVTK(1)	ApcB (1)-ApcA (2)/	3	820.6667	072815_CK_X02_01.28004.28004.3.0.dta
VLNGLKETYNSLGVPISSTVQAIQAIK(6)-IVKVELATGRPGTNAGLA(3)	ApcB (113)-ApcC (52)/	4	1190.7500	072815_CK_X02_01.28355.28355.4.0.dta
ELVVPSFTPVVKVGG(12)-IKAFVTGGAAR(2)	ApcE (893)-ApcA (27)/	3	924.0000	072815_CK_X02_01.24136.24136.3.1.dta
YLDGAAMDKLK(9)-VDKEVTPR(3)	ApcB (26)-ApcE (685)/	3	774.0000	072815_CK_X02_01.14358.14358.3.0.dta
IKAFVTGGAAR(2)-VDKEVTPR(3)	ApcA (27)-ApcE (685)/	4	543.2500	072815_CK_X02_01.11395.11395.4.0.dta
SIVTKSIVNADAEAR(5)-IMKMGGK(3)	ApcA (6)-ApcC (45)/	3	834.0000	072815_CK_X02_01.12626.12626.3.0.dta
[a]MQDAITAVINSADVQGKYLDGAAMDK(17)-IFTGGSPLSYLEKPVER(13)	[a] ApcB (17)-ApcE (87)/	[a] 4	1193.2500	072815_CK_X02_01.29091.29091.4.1.dta

SIVTKSIVNADAEAR(5)-VDKEVTPR(3)	ApcA (6)-ApcE (685)/	3	884.0000	072815_CK_X02_01.13878.13878.3.0.dta
APLNQKEIQQYNQILASQGLK(6)-SIVTKSIVNADAEAR(5)	ApcE (817)-ApcA (6)/	4	1028.2500	072815_CK_X02_01.20848.20848.4.1.dta
MQDAITAVINSADVQGKYLDGAAMDK(17)-ETIVKQAGDR(5)	ApcB (17)-ApcA (52)/	4	998.2500	072815_CK_X02_01.23159.23159.4.1.dta
AASVISANAATIVKEAVAK(14)-VDKEVTPR(3)	ApcB (53)-ApcE (685)/	4	728.2500	072815_CK_X02_01.19943.19943.4.0.dta
SIVTKSIVNADAEAR(5)-QQTKVFK(4)	ApcA (6)-ApcE (722)/	3	867.3333	072815_CK_X02_01.14513.14513.3.0.dta
MQDAITAVINSADVQGK(1)-VDKEVTPR(3)	ApcB (1)-ApcE (685)/	3	947.3333	072815_CK_X02_01.24450.24450.3.2.dta
MQDAITAVINSADVQGK(1)-QQTKVFK(4)	ApcB (1)-ApcE (722)/	3	930.6667	072815_CK_X02_01.25144.25144.3.0.dta
SIVTKSIVNADAEAR(5)-LYNKLTK(4)	ApcA (6)-ApcE (875)/	3	867.3333	072815_CK_X02_01.15684.15684.3.0.dta
AASVISANAATIVKEAVAK(14)-IKAFVTGGAAR(2)	ApcB (53)-ApcA (27)/	4	760.7500	072815_CK_X02_01.22244.22244.4.2.dta
VLNGLKETYNSLGVPISSTVQAIQAIK(6)-VDKEVTPR(3)	ApcB (113)-ApcE (685)/	4	985.7500	072815_CK_X02_01.26883.26883.4.2.dta
GPAVNNQVGNPSAVGEFPGSLGAKVFR(24)-IKAFVTGGAAR(2)	ApcE (511)-ApcA (27)/	4	978.2500	072815_CK_X02_01.22540.22540.4.3.dta
GPAVNNQVGNPSAVGEFPGSLGAKVFR(24)-SIVTKSIVNADAEAR(5)	ApcE (511)-ApcA (6)/	4	1098.2500	072815_CK_X02_02.14524.14524.4.2.dta
IKAFVTGGAAR(2)-LAKSPLYR(3)	ApcA (27)-ApcE (325)/	4	545.7500	072815_CK_X02_02.8677.8677.4.0.dta
[a] IFTGGSPLSYLEKPVER(13)-IKAFVTGGAAR(2)	[a] ApcE (87)-ApcA (27)/	[a] 4	780.7500	072815_CK_X02_02.14261.14261.4.0.dta
IKAFVTGGAAR(2)-VAVKNAIR(4)	ApcA (27)-ApcE (736)/	4	528.2500	072815_CK_X02_02.7717.7717.4.1.dta
AASVISANAATIVKEAVAK(14)-ETIVKQAGDR(5)	ApcB (53)-ApcA (52)/	4	770.7500	072815_CK_X02_02.12149.12149.4.0.dta
GPAVNNQVGNPSAVGEFPGSLGAKVFR(24)-SIVTKSIVNADAEAR(1)	ApcE (511)-ApcA (2)/	4	1095.7500	072815_CK_X02_02.14566.14566.4.4.dta
MQDAITAVINSADVQGKYLDGAAMDK(17)-VDKEVTPR(3)	ApcB (17)-ApcE (685)/	4	3817.9740	072815_CK_X02_01.23156.23156.4.2.dta
MRDAVTTLIKNYDLTGR(10)-SIVTK(1)	ApcF (10)-ApcA (2)/	4	2651.4440	072815_CK_X02_02.12719.12719.4.0.dta

[a] cApcE87 related cross-links that exceed the 30 Å threshold of the BS^3-H_{12}/D_{12} cross-linker.

Table S6.4 Cross-links identified in CpcL-PBS.

Cross-linked Sequence	Cross-linked Protein	Charge	m/z	Detected peptides in Raw File
LGGKVASITPASLS(4)-FDVFTR(1)	cpcC1(281)-cpcB (1)/	2	1177.1190	AB_X02_01.26941.26941.2.0.dta
MGGKIVSIK(4)-MFDVFTR(1)	FNR (69)-cpcB (1)/	3	662.3527	AB_X02_01.23420.23420.3.0.dta
MKEAALDIVNDPNGITR(2)-NQKTVGFSR(3)	cpcB (135)-cpcC2(161)/	4	758.3920	AB_X02_02.9991.9991.4.0.dta
MKEAALDIVNDPNGITR(2)-VEITAISAPGYPKVR(13)	cpcB (135)-cpcC1(253)/	4	899.4800	AB_X02_01.19605.19605.4.0.dta
MKTPLTEAVSTADSQGR(2)-MFDVFTR(1)	cpcA (2)-cpcB (1)/	3	948.7977	AB_X02_01.24858.24858.3.0.dta
MLGQSSLVGYSNTQAANR(1)-MFDVFTR(1)	cpcD (1)-cpcB (1)/	3	983.8073	AB_X02_01.28394.28394.3.0.dta
MLGQSSLVGYSNTQAANR(1)-MKTPLTEAVSTADSQGR(2)	cpcD (1)-cpcA (2)/	3	1279.9830	AB_X02_01.20401.20401.3.0.dta
NQKTVGFSR(3)-MFDVFTR(1)	cpcC2(161)-cpcB (1)/	3	697.0183	AB_X02_01.21905.21905.3.0.dta
SNKAVIVPFEQLNQTLQQINR(3)-MKTPLTEAVSTADSQGR(1)	cpcC1(259)-cpcA (1)/	4	1093.0700	AB_X02_02.15746.15746.4.1.dta
MKTPLTEAVSTADSQGR(2)-MYSPGYVATSSR(1)	cpcA (1)-FNR (1)/	3	1087.3330	AB_X02_01.16412.16412.3.1.dta
MKTPLTEAVSTADSQGR(1)-MFDVFTR(1)	cpcA (1)-cpcB (1)/	3	954.0000	AB_X02_01.24845.24845.3.0.dta
EKLEAIGYFR(2)-NQKTVGFSR(3)	cpcG2(189)-cpcC2(161)/	4	600.7500	AB_X02_02.10144.10144.4.3.dta
TLPLIAYAPVSQNQR(1)-MFDVFTR(1)	cpcG2 (2)-cpcB (1)/	3	912.5000	AB_X02_01.31189.31189.3.2.dta

Table S6.5 Summary of other intermolecular cross-links in CK-PBS, including the information of possible cross-linking sites located on different copies of cApcA/B discs and the measured Euclidean Cα-Cα distances.

Intermolecular Cross-links	Cross-linked Subunit on cApcA/B Discs		Cross-linked Distance, Cα-Cα, Å
cApcA(N-Term)-cApcB2	Any ApcA/B Disc	α1-β1	8.3
cApcA6-cApcB (N-Term)		α1-β1	9.3
cApcA27-cApcB(N-Term)		α1-β1	20
cApcA27-cApcB17		α1-β1	28.7
cApcA27-cApcB28		α1-β1	14.1
cApcA27-cApcB53		α1-β1	29.8
cApcA52-cApcB17		α1-β1	17.3
cApcA52-cApcB53		α2-β3	21.6
cApcC45-cApcA6	A3/A0/A′0/A′3/B0/B3-α3		19.3
cApcD48-cApcB17	A3/A0/A′4/A′1/B0/B3-β3		15.3
[a]cApcE87-cApcA27	[a]A3/A′3-α3		[a]57.6
	[a]A3/A′3-α3		[a]47
[a]cApcE87-cApcB17	[a]A3/A3-β3		[a]42.9
cApcE296-cApcB53	A3/A′3-β1		25.6
	A2/A′2-β1′		16.6
cApcE325-cApcA27	A3/A0/A′0/A′3/B0/B3-α1		29.3
	A2/A1/A′2/A′1/B1/B2-α2		30
cApcE511-cApcA6	A1/A′1-α1		10.7
	A0/A′0-α2		14.9
cApcE511-cApcA27	A1/A′1-α1		24.8
	A0/A′0-α2		26.0
cApcE511-cApcA2	A1/A′1-α1		15.7
	A0/A′0-α2		20.3
cApcE685-cApcA6	A1/A′1-α1		30.5
cApcE685-cApcA27	A0/A′0-α2		[b]36.1
cApcE685-cApcA52	B2/B1- α3		25.8
	A1/A′1-α1		13.2
cApcE685-cApcB(N-term)	A1/A′1-β1		30.1
cApcE685-cApcB17	B2/B1-β3		26.7
	A2/A′2-β3		32.6
	A1/A′1-β1		6.8
cApcE685-cApcB26	A1/A′1-β1		24.7
cApcE685-cApcB53	B1/B2-β2		26.3
	A2/A′2-β2		28.3
	A1/A′1-β3		25.0
cApcE685-cApcB113	A1/A′1-β3		[b]32.8
cApcE722-cApcA6	B2/B1-α3		10.3
	B3/B0-α3		14.0
cApcE722-cApcB(N-term)	B3/B0-β3		18.9
	B2/B1-β3		12.3

cApcE736-cApcA27	B2/B1-α3	27.8
	B3/B0-α3	29.8
cApcE817-cApcA2	B2/B1-α2	15.9
	B3/B0-α1	23.1
cApcE817-cApcA6	B3/B0-α1	28.6
	B2/B1-α2	18.9
cApcE817-cApcB(N-term)	B3/B0-β1	28.0
	B2/B1-β2	11.4
cApcE875-cApcA6	B2/B1-α3	25.0
cApcE893-cApcA27	B2/B1-α2	25.0
	B3/B0-α1	30.2
cApcF10-cApcA2	A2/A′2-α3	10.1
cApcF 10-cApcA6	A3/A′3-α1	25.3
	A2/A′2-α3	13.7
cApcF28-cApcA27	A3/A′3-α1	20.3
	A2/A′2-α3	10.1

[a] cApcE87 related cross-links that exceed the 30 Å threshold of the BS^3-H_{12}/D_{12} cross-linker.

[b] cApcE685 related cross-links that exceed the 30 Å threshold of the BS^3-H_{12}/D_{12} cross-linker.

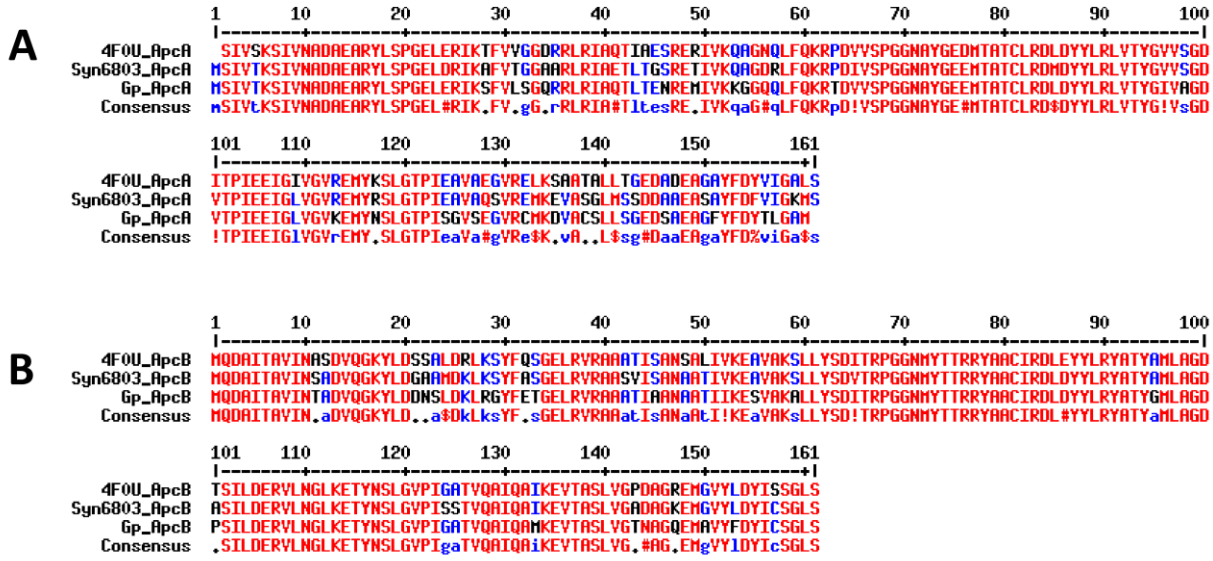


Figure S6.1 Sequence alignments (A) between cApcA (PDBID: 4F0U) and GpApcA (PDBID: 5Y6P) and between (B) cApcB (PDBID: 4F0U) and GpApcB (PDBID: 5Y6P).

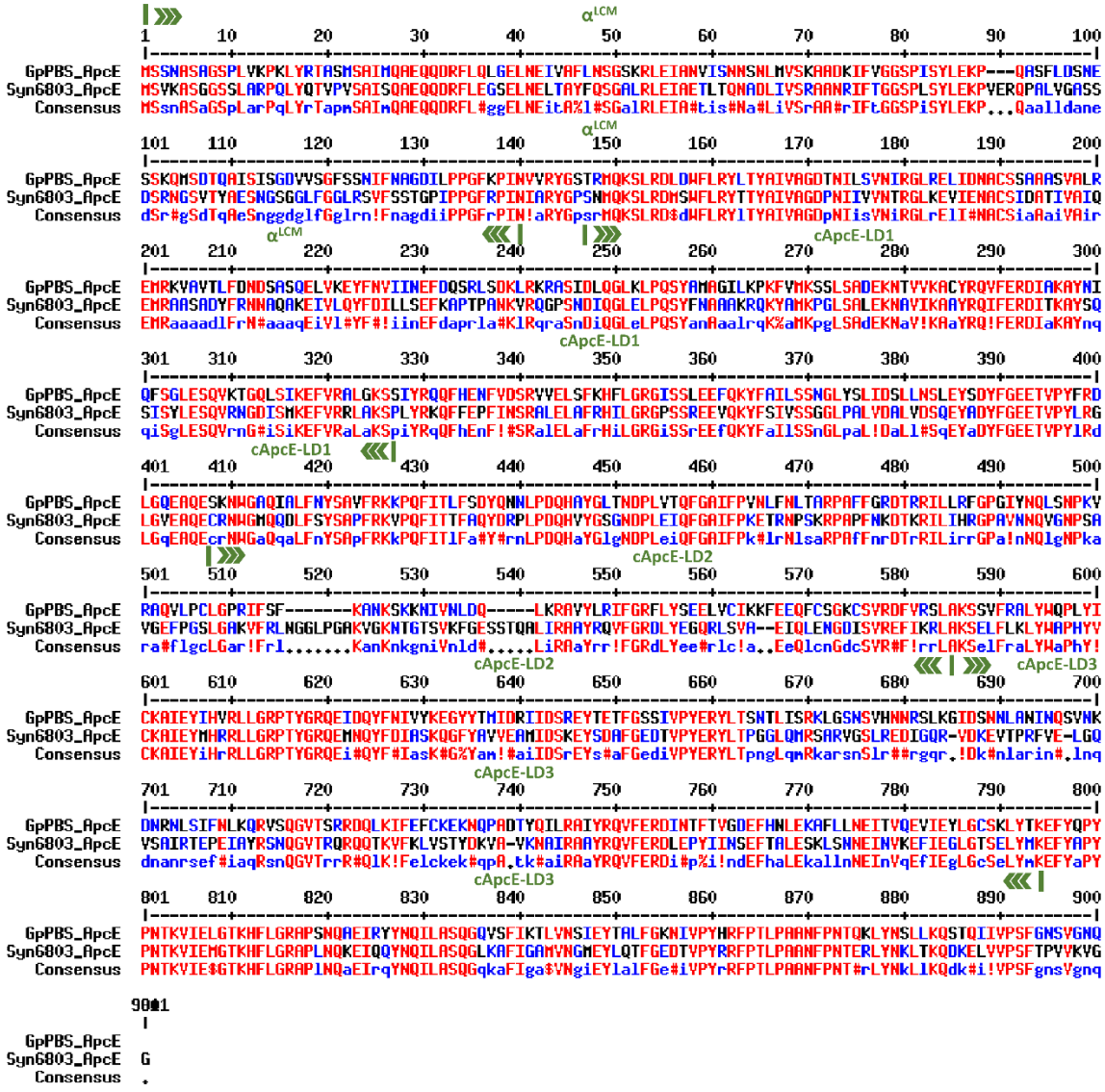


Figure S6.2 The sequence alignment of GpApcE (PDBID: 5Y6P) and cApcE.

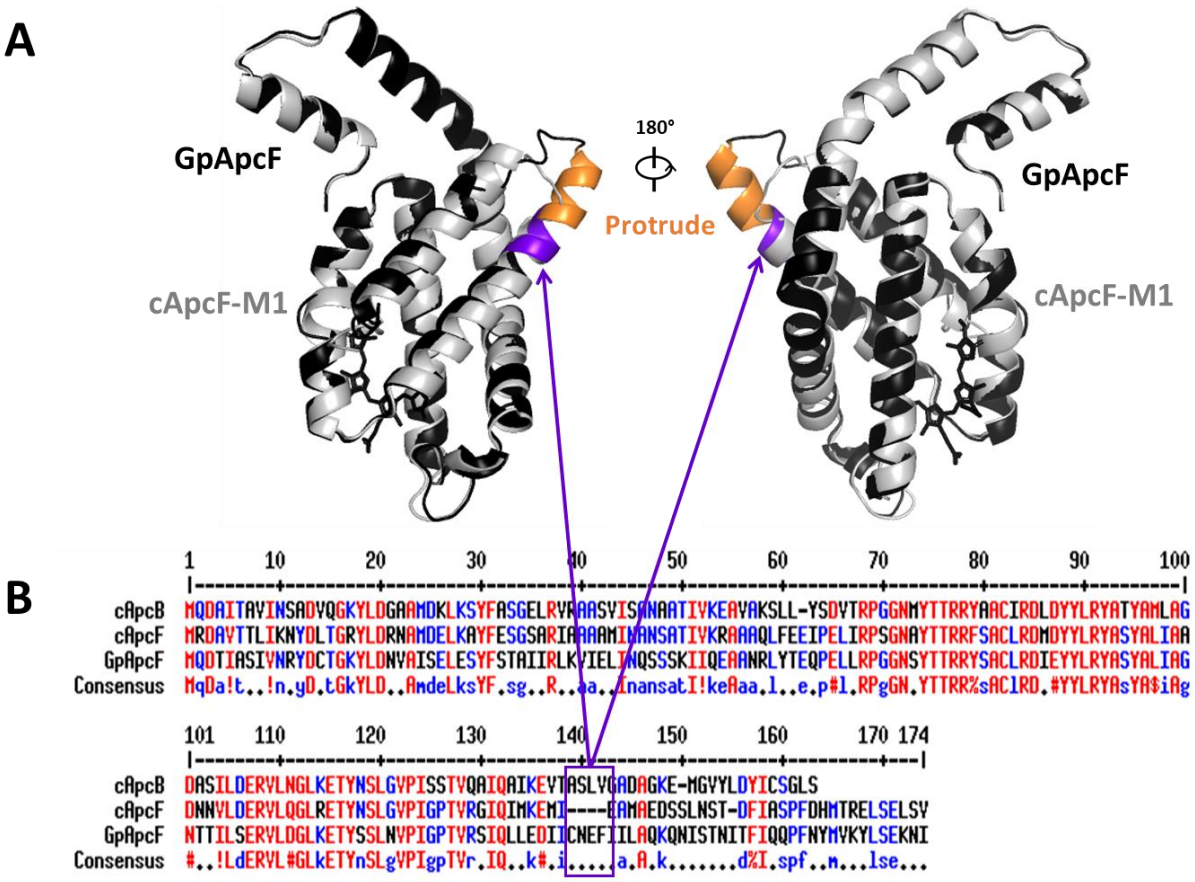


Figure S6.3 Alignment of cApcF and GpApcF. (A) Alignment of the predicted cApcF-M1 (c for cyanobacteria, grey) onto the GpApcF (Gp or *Griffithsia pacifica*, black). Highlighted region (purple) is the indel of residue 139-142. The extension protrude is highlighted in orange. (B) The sequence alignment of GpApcF and cApcF. Main sequence difference is boxed.

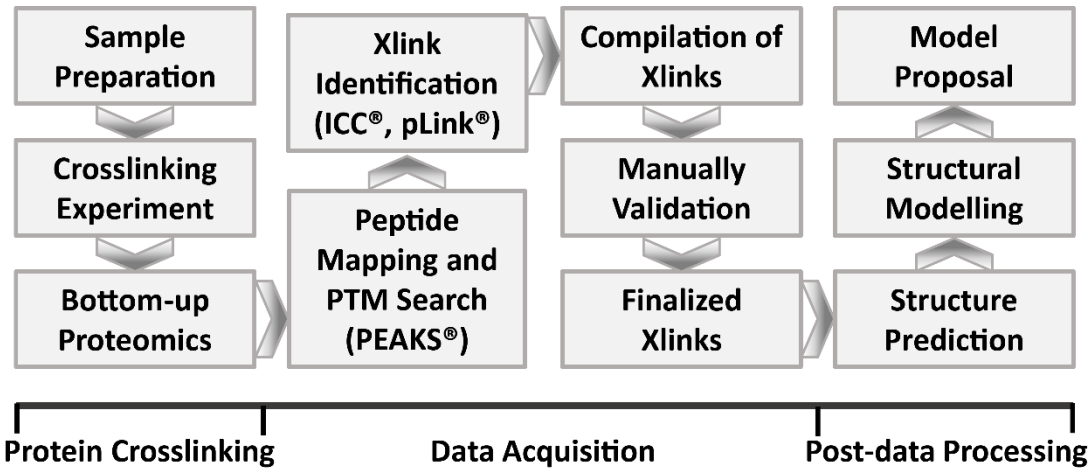


Figure S6.4 Summarized workflow.

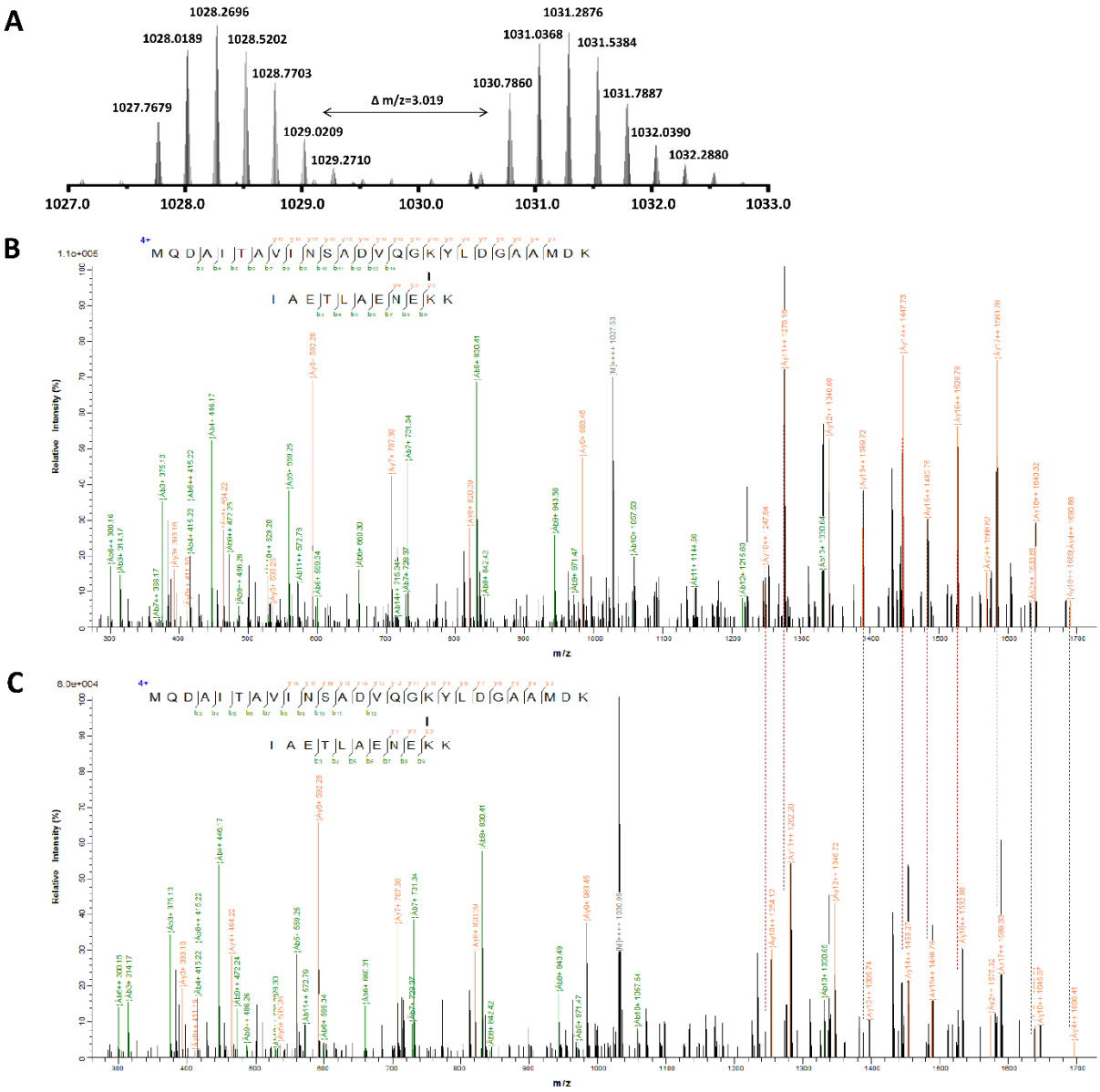


Figure S6.5 MS and MS/MS spectrum of a representative cross-link, ApcB (17)-ApcD (48). (a). The MS1 spectra of the light and heavy cross-linked peptide (BS^3-H_{12}/D_{12}) which are sperated by m/z 3.019 (z=4). (b). MS/MS spectra of the cross-linked peptide (light). (c). MS/MS spectra of the cross-linked peptide (heavy).

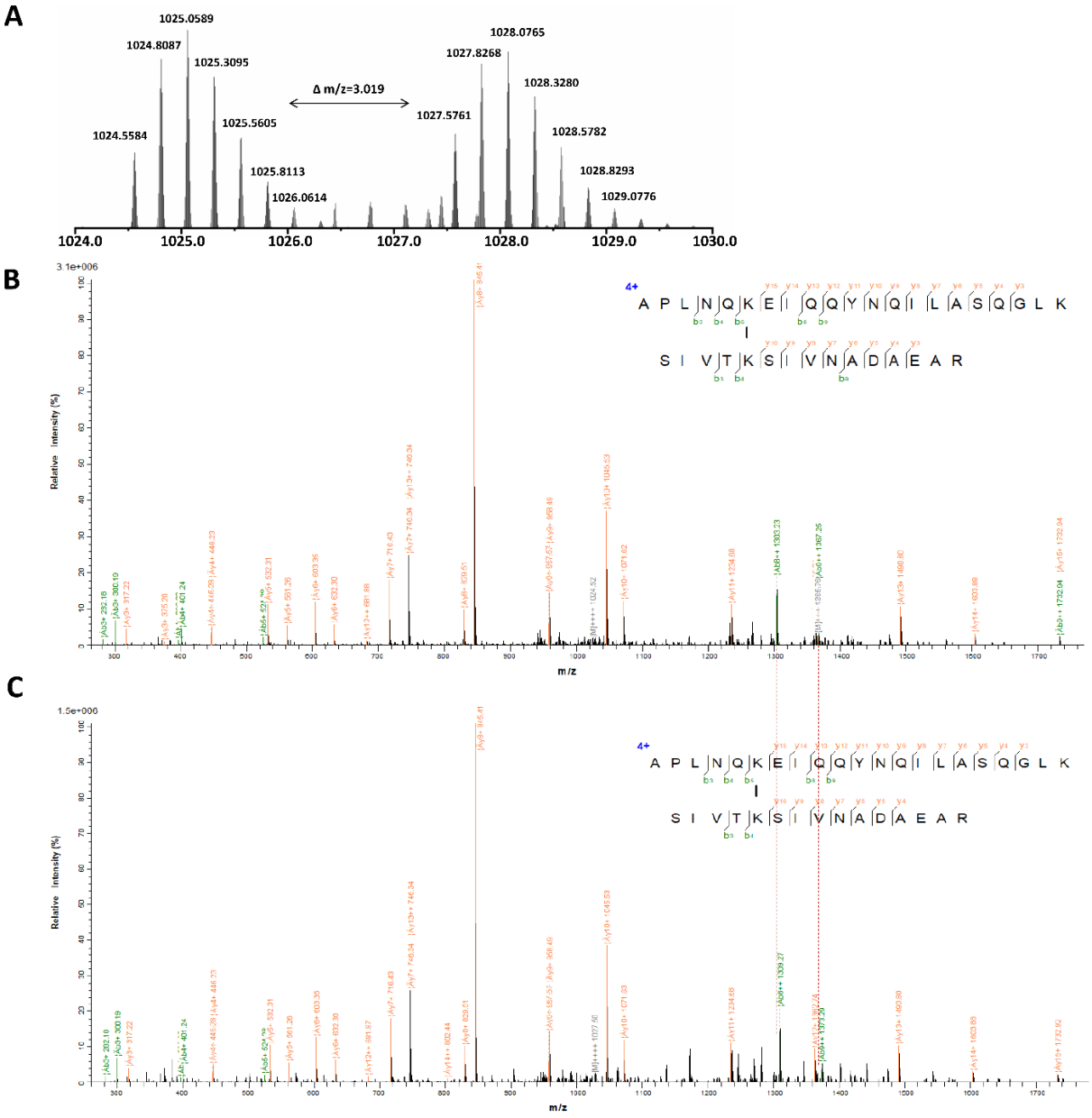


Figure S6.6 MS and MS/MS spectrum of a representative cross-link, i.e. ApcE (817)-ApcA (6). (a). The MS1 spectra of the light and heavy cross-linked peptide (BS^3-H_{12}/D_{12}) which are sperated by m/z 3.019 (z=4). (b). MS/MS spectra of the cross-linked peptide (light). (c). MS/MS spectra of the cross-linked peptide (heavy).

Chapter 7: Conclusions and Outlook

Mass spectrometry (MS)-based approaches have grown rapidly in the past 30 years, becoming an asset to characterize protein higher-order structures (HOS). Comparing to traditional biophysical methods, MS-based protein HOS analysis provides middle-to-high spatial resolution with high throughput and low amount of required sample. MS-based footprinting is one of these approaches, aiming to map solvent accessible surface area (SASA) of protein at different conditions. The focus of research described in this thesis is to develop and apply one or several MS-footprinting approaches to answer specific biological questions. In general, MS-based footprinting includes reversible footprinting (hydrogen-deuterium exchange (HDX)) and irreversible footprinting which can be conducted by radical reagents and other targeted labeling reagents. In this thesis, chapter 1 reviews MS-based footprinting in detail and discusses the difference between its subcategories. The following chapters focus on the development and application of these MS-footprinting approaches.

7.1 Elaboration of FPOP Platform

In 2005, the Gross lab developed fast photochemical oxidation of protein (FPOP) that can generate hydroxyl radicals on a flow system and label protein according to their SASA at the microsecond timescale. In the past decade, FPOP has been adopted in both academic and pharmaceutical labs to study protein conformational changes, protein-protein or protein-ligand interactions, protein aggregation and protein dynamics. For broader applications, elaboration of the FPOP platform is a primary interest in our lab. Chapter 2 describes one way of elaborating FPOP; that is, the development of a new radical labeling reagent, the carbonate radical anion. We generated the carbonate radical anion from hydrogen peroxide in the presence of bicarbonate buffer and investigated its reaction specificity. With selectivity towards methionine and aromatic residues, carbonate radical anion complements other radical reagents in the footprinting

"toolbox". The targeted chemistry of the carbonate radical allows easier data process for protein systems that are rich in those reactive residues. The major limitation of this footprinting approach, however, is the basic pH condition during radical generation; the high pH can potentially limit its applications for pH-sensitive proteins. For future research, radical reagents that can be generated in biological-friendly conditions are to be developed. The ideal condition should constitute of neutral pH, require minimal number of chemicals, and contain inert precursors to avoid pre-labeling modification. In addition, these radicals are preferred to yield bio-orthogonal mass adducts in proteins, in which way that the readout can be distinguished from the mass of PTMs or of degradation products during MS analysis. Furthermore, novel footprinting reagents that target relatively less reactive residues (e.g., alanine and glycine) are also desirable; perspective from organic chemistry regarding C-H bonds activation may provide useful guidance to address this issue. There are also technical concerns of developing a new reagent including the evaluation of its compatibility to different protein systems. In addition, the optimum dosage of the reagent needs a systematic discussion. To investigate residue preferences under different conditions, a series of small peptides that only differ in one amino acid are good model systems, allowing a better methodical comparison.

In chapter 3, we discussed the expansion of FPOP platform from another aspect, which is the integration of two lasers to follow protein unfolding. We designed and established a two-laser platform to allow a pH jump induced by the first laser followed by the radical footprinting triggered by the second one. With a time-delay between the two lasers, we aimed to characterize the protein conformations occurring during unfolding. We evaluated different photoacids and radical reagents for this purpose, setting foundations for future research. In future studies, more photoacids and radical reagents need to be screened to avoid complicated chemistries upon laser

irradiation. Photoacids without nitroso-group (e.g., benzyl ester) can be a good start. Iodine radical footprinting and trifluoromethyl radical footprinting are both worth a try. With the development of more radical footprinters, there are better chances for more candidate reagents to be accommodated on the two-laser platform. The established platform would considerably benefit the study of the gating mechanisms of ion channel proteins, such as potassium ion channel proteins, transient receptor potential channel protein, and the M2 protein, which is a proton transport channel of influenza virus A that is sensitive to pH changes. Structural dynamics of the targeted protein with respect to different lipid environments and substrate binding can further be investigated.

7.2 Integration of HDX-MS, XL-MS with Computational Methods

HDX-MS and irreversible footprinting provide complementary information in terms of elucidating protein HOS. HDX-MS delineates SASA and dynamics of the backbone amides, whereas the irreversible footprinting (e.g., FPOP and chemical cross-linking (XL)) informs on reactive amino acid side chains. Integration of the two methods provides more precise and comprehensive description of protein HOS compared to employing either one.

In Chapters 4 and 5, we showed two examples of adopting this combined approach, specifically with HDX-MS and XL-MS. The implementation of XL-MS successfully differentiates the regions undergoing remote conformational changes from those being protected in the binding events; two scenarios that cannot be distinguished by stand-alone HDX. We further incorporated molecular docking in the protein HOS interpretation, guided by the restraints derived from the cross-links, and acquired quaternary structural information of the binding complex.

Although the integrated platform shows great potential in protein HOS analysis, there are limitations that deserve future attention to overcome. Firstly, molecular docking heavily relies on accurate protein structures to start with. The involved proteins retain their structural rigidity throughout the docking process. This feature largely limits its applications to proteins that have significant numbers of flexible domains. In the future, other computational methods (e.g., discrete molecular simulation) that can better accommodate the structural changes are desirable. In addition, current HDX adjudication is conducted only qualitatively. We envision that the quantized HDX readout can advance the elucidation of protein HOS and ultimately contribute to protein structure prediction in the absence of differential experiments. Furthermore, the integrated MS-based approaches in combination with other biochemical techniques would better advance the characterization platform. For example, peptides representing the suggested binding regions from MS analysis can be synthesized and incorporated in binding assays (e.g., fluorescence polarization assay (FPA), competition FPA, isothermal titration calorimetry, and surface plasma resonance) to ascertain their binding interactions with the targeted protein and to obtain even their binding affinities. In addition, mutagenesis (e.g., alanine shaving) of the potential binding interfaces and the associated analysis of the binding kinetics with mutants are beneficial complements to locate further the critical binding residues that are favored energetically. Subsequent bio-functional assessments such as minigenome assays to study high pathogenic virus using protein or peptide mutants will also provide constructive evidence for a comprehensive analysis of the binding events.

Besides the interleukin complexes and antibody/antigen complexes such as PD-1/Nivolumab, other antibody therapeutics can also be accommodated on this integrated platform. One example of many is the study of bispecific and trispecific antibody systems that contain numerous

designed paratopes and complex targeting antigens. The MS-based footprinting approaches on the integrated platform can not only identify their epitope/paratope regions but can also specify the interacting antigens for each paratope based on the cross-linking results, therefore promoting the mechanistic interpretation of the antibody function and potentially their discovery and development.

Chapter 6 describes a collaboration wherein we used irreversible footprinting and computational methods, the combination of XL-MS and homology modeling. The latter helps to predict the structure of a phycobilisome complex, which is further modified and validated by the obtained cross-links. For future study, other proteins or protein complexes are good opportunities to expand its applications. From a technical point of view, method development in sample preparation is preferred, especially for large protein complexes. Chromatographic separation such as size exclusion chromatography and strong cation exchange chromatography are worth considering. These enrichment methods can distinguish the cross-linked peptides with higher charge states from the linear peptides, sustainably increasing the identification efficiency in the MS analysis. Other separation techniques such as asymmetrical flow field-flow fractionation (AF4) can also be a good candidate, which enables an on-line separation of macromolecules prior to cross-linking, minimizes the heterogeneity of the targeted protein complex when executing the cross-linking reaction, and facilitates subsequent MS analysis. In addition to photosynthetic complexes, other protein systems such as the Mediator complex, a transcriptional coactivator, and the anaphase-promoting complex (APC), which regulates chromosome segregation and anaphase progression, can also be analyzed in such fashion. All in all, marriage of footprinting and computational efforts provides a powerful approach to study large protein

complexes, whose structures are usually hard to obtain from conventional high-resolution structural biology methods including X-ray crystallography and NMR.