Spring 5-15-2017

Mass Spectrometry-Based Structural Analysis of Photosynthetic Protein Assemblies

Yue Lu
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Chemistry

Dissertation Examination Committee:

Robert E. Blankenship, Co-Chair

Michael L. Gross, Co-Chair

Dewey J. Holten

Himadri B. Pakrasi

Gary J. Patti

Mass Spectrometry-Based Structural Analysis of Photosynthetic Protein Assemblies

by

Yue Lu

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

May 2017

St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

List of Abbreviations

APBS: Adaptive Poisson-Boltzmann Solver

CCS: Collisional Cross Section

EIC: Extracted Ion Chromatogram

FPOP: photochemical oxidation of proteins

FRP: Fluorescent Recovery Protein

IM-MS: Ion-Mobility–Mass Spectrometry

LC-MS: Liquid Chromatography–Mass Spectrometry

LH2: Light Harvesting Complex 2

MS: Mass Spectrometry

MS/MS: Tandem mass spectrometry

NTD: N-Terminal Domain

NPQ: Non-Photochemical Quenching

OCP: Orange Carotenoid Protein

PSI: Photosystem I

PSII: Photosystem II

PTM: Post Translational Modification

RC: Reaction Center

Acknowledgments

I would like to thank my two advisors, Prof. Michael L. Gross and Prof. Robert E. Blankenship, for their continual guidance, support and encouragement throughout my studies at Washington University. I thank Bob for the excitement about research and teaching me humility. I thank Dr. Gross for sharing his deep insights in science and teaching me humor.

I would like to thank Drs. Hao Zhang, Weidong Cui, Henry Rohrs and for bringing me into the world of mass spectrometry. I would like to thank Dr. Jing Jiang, Dr. Haijun Liu, Ms. Mindy Prado, Dr. Rafael Sar and Darek Niedzwiedzki for donating their time to teach me all the biological techniques. Special thanks to Drs. Michael Marty, Yelena Grinkova and Steven Sligar at UIUC.

I would like to thank my Dissertation Advisory and Examination Committee members, Prof. Michael L. Gross, Prof. Robert E. Blankenship, Prof. Dewey J. Holten, Prof. Gary J. Patti and Prof. Himadri B. Pakrasi for their help and inspiration regarding my thesis research.

I gratefully acknowledge the financial support received from the U.S. Department of Energy (DOE) and National Institutes of General Medical Sciences of the NIH. Additionally, I thank to the graduate school of art and sciences for dean's fellowship.

Yue Lu

Washington University in St. Louis

May 2017

ABSTRACT OF THE DISSERTATION

Mass Spectrometry-Based Structural Analysis of Photosynthetic Protein Assemblies

by

Yue Lu

Doctor of Philosophy in Chemistry

Washington University in St. Louis, 2017

Professor Robert E. Blankenship, Co-Chair

Professor Michael L. Gross, Co-Chair

This dissertation focuses on using mass spectrometry-based techniques to study photosynthetic protein assemblies. Photosynthesis is a process that converts light energy into chemical energy, the basis of most life on Earth. The two most crucial protein machineries involved in this process are reaction center and light harvesting complexes. They are usually giant protein complexes with different numbers of co-factors. In a more expanded sense, photosynthesis is not just about the utilization of solar energy, the regulation of light energy is also essential as excess light energy is detrimental to photosynthesis organisms. Again, protein assemblies play an indispensable role in this process. The knowledge of the structure and function as well as the molecular mechanism of those protein complexes are desired.

Today, mass spectrometry is being widely used in proteomics studies. Its capabilities include but are not limited to the protein primary structure investigation. The development of MS-based footprinting, native MS and membrane protein MS detection platforms largely benefit the study of photosynthetic proteins. The MS-based footprinting technique can investigate protein conformational change upon its binding to other molecules or under the stimulus of pH change or other factors. Native MS can investigate the conformation and topology of protein complexes

in a near-native environment where the non-covalent interactions are preserved. Membrane proteins are notoriously difficult to study. The development of MS-based membrane protein detection platforms largely benefits the study of photosynthesis, as reaction center and light-harvesting complexes are usually membrane proteins.

In this dissertation, a variety of MS-based techniques were utilized to study reaction center proteins, light harvesting proteins and the proteins involved in the photoprotection process. We utilized top-down MS to study the components as well the primary structure of LH2 from a purple bacterium (*Rb. sphaeroides*), which reveals a new post-translational modification and mutation information. In addition, we developed a MS-based platform to footprint this LH2, investigating its topology in a lipid bilayer. The reaction center from another purple bacterium (*B. viridis*) was studied by both bottom-up and top-down MS and lots of unexpected mutations were identified. We also conducted a native MS study on this reaction center, and the capabilities of retaining the co-factors as well as its collisional cross section in the gas phase are discussed. Lastly, we study the orange carotenoid protein (OCP) and the fluorescence recovery protein, two major players in the non-photochemical quenching process in cyanobacteria. We utilized MS-based techniques to probe the conformation and structure of these two proteins and finally proposed a mechanism for non-photochemical quenching in cyanobacteria.

Chapter 1: Introduction

1.1 Protein machineries involved in photosynthesis

"Photosynthesis is a process in which light energy is captured and stored by an organism, and the stored energy is used to drive energy-requiring cellular processes"[1]. The energy is stored in the form of carbohydrate molecules, and usually oxygen is released as a byproduct. The process starts with the absorption of solar light by pigments associated with the photosynthetic antennas, delivering the energy to the reaction center where charge separation takes place (Figure 1.1).

Various photosynthetic antennas and reaction center systems have been developed in plant, algae, cyanobacteria and other photosynthetic bacteria. For example, light harvesting complex 2 (LH2) from purple bacteria is well known to adopt a "ring" structure, composed of heterodimeric units——α, β apoprotein pairs that serve as a scaffold to bind bacteriochlorophyll *a* and a carotenoid for optimal energy transfer[2]. In cyanobacteria, a giant protein complex called a Phycobilisome (PBS) harvests light and supplies energy for both photosystem II (PSII) and photosystem I (PSI). This giant antenna protein complex, which is composed of various biliproteins and linker bilipeptides, exhibits a "rods and core" structure[3-4].

Although light harvesting antennas execute the capture of solar energy, the primary energy conversion in photosynthesis takes place in the reaction center. The reaction center is a protein complex that incorporates not only pigments—chlorophyll/bacteriochlorophyll and carotenoids—but also includes quinones and iron sulfur centers to carry charge. Among all the co-factors associated with the reaction center protein scaffold, one unique pair of chlorophyll/bacteriochlorophyll located in the hydrophobic core of the reaction center, called the

special pair, can be electronically excited and initiate the charge separation[1]. For example, two mega protein complexes, PSI and II, are the reaction center protein complexes in oxygenic photosynthetic organisms, and function to convert the light energy into chemical energy. PSII is a supramolecular complex embedded in the thylakoid membrane, with a number of protein subunits and various cofactors[5]. The reaction center in most purple bacteria has a comparatively simple structure that includes protein subunits H, M and L as well as many co-factors. Reaction centers from some purple bacteria like *Blastochloris viridis* possess an extra bound cytochrome subunit[6].

Although utilization of light is the essential activity of the photosynthesis process, quenching any excess light energy is also crucial. During periods of excess light energy, excited triplet chlorophyll can induce the production of a high-energy form of oxygen, singlet oxygen. Singlet oxygen is detrimental to the cells, leading to the damage of proteins, pigments, and lipids[7]. Thus, a strategy for regulation of light utilization is needed for photosynthetic organisms. In plants and algae, light harvesting antennas are involved in modulation of light energy—photoprotection[8]. In cyanobacteria, protein machineries, other than light harvesting antennas, are essential to perform photoprotection. Orange carotenoid protein (OCP) is a single carotenoid-binding protein that can quench the excess energy in the light harvesting antenna phycobilisome (PBS), preventing damage of the photosynthetic apparatus by oxidative radicals[9-10]. The activated OCP (red) can accelerate the decay of the excited singlet states of bilin pigments in PBS, affecting the energy transfer from PBS to reaction center. Another protein, called fluorescent recovery protein (FRP), can recover the fluorescence in PBS, just as its name says, by interacting with the OCP[11].

Above all, delicate protein complexes play indispensable roles in photosynthesis, an essential process for almost all life on earth. Tools that can characterize those protein machineries with a

high resolution are desired. Spectroscopic techniques have been widely utilized, owning to the unique absorption and fluorescence characteristics of photosynthetic proteins. Now, mass spectrometry (MS) has quickly developed in protein science and becoming more and more popular, owing to its high speed, sensitivity, accuracy, and high dynamic range, for studies of amino acids to mega-complexes[12-14]. In chapter 1, we will discuss the application of mass spectrometry to photosynthetic protein machineries.

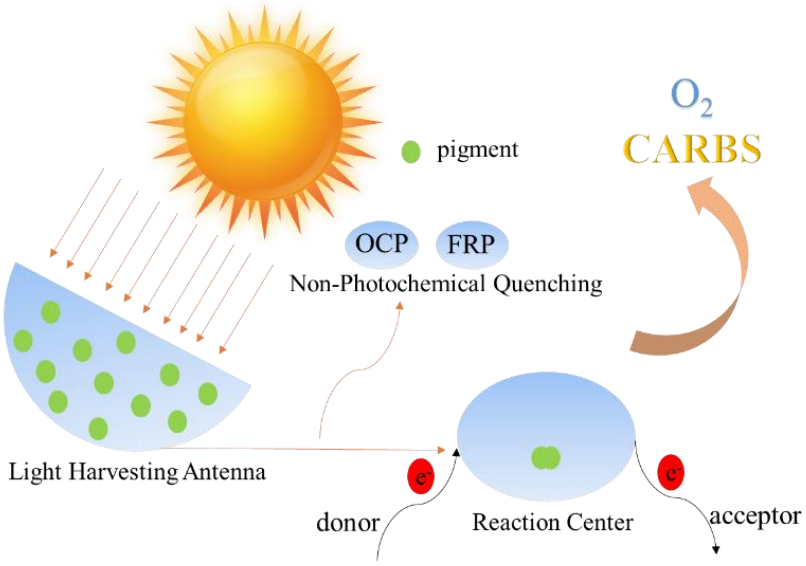


Figure 1.1 Photosynthesis scheme in most photosynthetic organisms. Light blue shapes represent protein assemblies involved in photosynthesis process. OCP and FRP are shown as example proteins that are involved in photoprotection.

1.2 Protein primary structure determination

The information in DNA is encoded by four bases: adenine (A), guanine (G), cytosine (C), and thymine (T). This information is passed on to RNA, in preparation for the final step that

comprises the central dogma of molecular biology. According to the genetic code, every three bases in RNA can be translated into one specific amino acid, subunits composing proteins[15]. Proteins are central parts of all organisms, being essential in cell structure and function. Before the wide availability of gene sequencing, chemical or enzyme methods were being used to investigate the sequence of protein[16]. Edman degradation used to be the dominant method in sequencing amino acids in proteins. The amino acid is labeled from the N-terminus, followed with the releasing of this terminal amino acid derivative to be detected by chromatography, electrophoresis or Ultraviolet–visible spectroscopy (UV-Vis). The process can be repeated, and the whole sequence of the protein can be deciphered. However, the whole process is tedious, and this method is not effective in the presence of a chemically modified N-terminus. Nowadays, the MS-based sequencing (fragmentation) approach is replacing Edman degradations and has become indispensable in probing the primary structure of photosynthetic proteins.

Early mass spectrometers coupled with electrospray ionization (ESI) could only perform one straightforward function, measuring the molecular weights, while no structural information could be obtained[17]. The development of tandem MS enabled the probing of the primary sequence of proteins. The process involves two steps; the first step is the selection of a precursor/parent ion and the second step is the activation of the precursor/parent ion followed with MS detection of fragmented products. One of the activation methods, known as collisional induced activation (CID), is the dominant fragmentation method in probing protein primary structures[18]. During the CID process, the precursor/parent ion is accelerated by an electrical potential to high kinetic energy, allowing it to collide with neutral gas molecules. Nitrogen is the most commonly used neutral gas, and helium and argon are also widely utilized. The kinetic energy is converted into internal energy during this process, leading to chemical bond fragmentation, and partial/complete

sequence of the precursor/ion can be obtained by measuring the MW of fragments. The other CID method, called higher-energy collisional dissociation (HCD), leads to a shorter activation time compared to CID[19]. The fragmentation spectrum of HCD features a predominance of y-ions and b-ions that can be further fragmented into smaller species like a-ions. For HCD, there is no low mass cut-off restriction as there is with ion traps and the method, when coupled with an orbitrap, provides high mass accuracy in the product-ion (MS/MS) spectrum. Thus, HCD is suitable for *de novo* peptide sequencing[20].

In addition to CID and HCD, another fragmentation method called electron capture dissociation (ECD), developed by McLafferty, starts from capturing electrons by multiple charged ions in the gas phase[21]. Unlike CID/HCD, the fragmentation site is not defined by bond strength during the ECD process[22]. Whereas CID/HCD fragmentation occurs by increasing the internal energy of peptide/protein ions, it causes protons to move (i.e., mobile proton model), electron-capture dissociation (ECD) generally breaks bonds near the location of a protonated site that can attract the electrons[23]. As two complementary fragmentation methods, CID/HCD produces b and y series ions whereas ECD produces c and z series ions. Other than CID/HCD and ECD, a variety of fragmentation methods, like electron transfer dissociation (ETD) were developed and utilized in proteomics, providing complementary sequence/PTM information for a certain peptide[19].

Primary structural determination of photosynthetic proteins by MS is centrally important, for verifying amino acid sequence, and identifying PTMs and even the isoforms of proteins. Bottom-up MS proteomics is widely used to identify protein proteolytic digestion products[24]. MS detection in combination with liquid chromatography (LC), the pre-separation before MS detection, is a very useful technique that improves enables the ability to sequence peptides in

complex mixtures (e.g., protein digests). Numerous studies have utilized this technique to gain knowledge of photosynthetic protein primary structure. For example, the MS primary structure elucidation of D1 and D2 proteins from the PS II reaction not only reveals a variety of PTMs, but more importantly, the C-terminal sequence of D1 protein is identified[25]. The C-terminal processing on D1 protein was later found to be essential for the function and assembly of PS II[25-26].

1.3 MS-based Footprinting

Scientists are not content to know just the primary structure of proteins. MS-based footprinting as a biophysical approach is widely utilized to probe protein structure, dynamics, and interactions in different macromolecular assemblies. In general, two types of footprinting strategies are being developed, non-reversible and reversible labeling of amino acid backbones or side chains[27-29], respectively. By comparing the labeling level of the target protein at different states, regions that are involved in interaction or in structural changes can be identified. Utilizing the typical bottom-up approach, we usually measure the labeling extents of two different states (e.g., apo vs. holo). Hydrogen deuterium exchange (HDX) monitored by MS can be used to investigate protein conformational changes when the protein structure is perturbed by a binding partner, pH, denaturation, for example. The targets of an HDX approach are the amide protons of the protein backbone that can undergo exchange with the deuterium in the solution. The extent of exchange reports on H bonding and solvent accessibility of the backbone[30-32.] As a “gentle" but “universal” approach, HDX is one of the most robust and reliable MS-based footprinting methods in probing protein structure. In green-sulfur bacterial, light energy absorbed by chlorosome is transferred to the FMO antenna protein through a baseplate protein. The molecular

details of how the baseplate is coordinating with FMO antenna protein to transfer the energy efficiently were only fully understood. Our group utilized HDX approach to investigate the binding face of the FMO antenna protein and the chlorosome CsmA baseplate protein, as the interacting region on FMO shows significant decrease of deuterium uptake after CsmA binding[33].

Because covalent footprinting by HDX is reversible, efforts have been made to take advantage of bottom-up proteomics for footprinting by developing irreversible labeling. The most common examples are hydroxyl radical and GEE labeling in combination with MS detection are also widely used in studying photosynthetic proteins. Based on the labeling extent on amino acids, structural change of protein assemblies can be deciphered. Oxidative hydroxyl radicals can be generated by X-ray beams, electrical discharge and laser photolysis of hydrogen peroxide in solution, labeling proteins on millisecond and sub-millisecond timescales. The amino-acid site chain reactivity with hydroxyl radicals is broad-based, although the reactivity with various amino acid side chains can vary by three orders of magnitude[34-37]. Another quick and simple footprinting approach targeting carboxyl amino acids, is "GEE labeling", and can also be used to footprint proteins[29]. Carboxyl groups on glutamate (E) and aspartate (D) can be labeled, producing a +57 or +71 Da mass shift[29]. OCP is the essential player during NPQ process in cyanobacteria. It has two states, the red active state and the orange inactive state. The red active state can burrow into PBS and quench the excess light energy. The structural difference of those two states has been an intriguing question. The MS-based footprinting experiments (covalent labeling) successfully identify the global and local structural rearrangement of proteins. The red state was determined to have an open structure whereas the orange state a compact structure. In particular, the hydroxyl radical footprinting method reveal the structural of orange

carotenoid protein change up to the residue level upon photoactivation[38], whereas the results obtained by GEE labeling lead to a similar conclusion with an emphasis on carboxyl amino acids[39].

1.4 CROSSLINKING (XL)-MS

Amino acids pairs that are adjoining in protein complexes can be "snapshotted" by a cross-linker with a certain length, thus identifying regions of proteins that bind. After proteolytic digestion, the cross-linked peptides can be identified by tandem MS. Identification of cross-linked peptides by MS provides insightful information for modelling of protein-protein interactions, protein conformational changes, and protein dynamics[40-43]. As photosynthetic machineries usually contain a variety of protein subunits and co-factors, crosslinking MS (XL-MS) is widely adopted to investigate the organization or topology of photosynthetic protein assemblies during the photosynthesis process, especially in the PSII-centered problems[4, 44-49]. PS II is a giant protein with a complex composition, that is around 20 protein subunits in cyanobacteria and plants as well as accessory light-harvesting proteins. Psb28 is an extrinsic protein of PSII and plays an important role in PSII repair[50]. It protects the RC47 assembly intermediate of PSII. Weisz and co-workers[46] used XL-MS to investigate the transient interaction of Psb28 within PSII, and Psb28 was found to bind to cytochrome b_{559}. XL-MS was also widely used to investigate the interaction between OCP and PBS, another giant protein complex. The OCP binding site on phycobilisome was probed by several groups using XL-MS to provide models for the structure. Although those models don't overlap with each other precisely, all of them suggest the binding site is located on the APC core of the phycobilisome[9].

1.5 Top-down MS

Bottom-up MS approach has been well established and applied in a wide range of proteomics studies, especially the primary structure sequencing and MS-based footprinting as mentioned in previous paragraphs. In the meanwhile, top-down MS proteomics study is emerging in recent years, allowing for the direct molecular weight measurement of intact proteins instead of proteolytic digestion[51]. Usually, fragmentation of intact proteins is carried out during top-down analysis to identify protein sequence and PTM. There are several advantages of top-down MS. Frist of all, no tedious sample preparation procedures like digestion are needed, which not only saves time and effort but also introduces less potential artifacts into the mass spectrometer. Secondly, the data processing of top-down MS is very straightforward. Proteolytic digestion usually results in a varied number of peptides based on the protein size and sequence, while the top-down approach targets only a couple of candidates based on the components of the protein complex. Lastly, crucial sequence information that might get lost by using bottom-up MS approach could be identified by the complementary MS method, top-down MS approach. The very hydrophilic peptides are usually lost during desalting LC and the extremely hydrophobic peptides are reluctant to elute during gradient LC. One example is the MS study on chlorosome proteins CsmA and CsmE, light harvesting apparatus in green bacteria. The presence of truncated versions of CsmA and CsmB were detected by matrix-assisted laser desorption ionization (MALDI) mass spectrometry[52]. In addition, the top-down MS approach has been widely utilized to sequence the integral membrane photosynthetic proteins that are resilient to digestion. MS studies on membrane proteins are discussed in the last part of this chapter.

An innovative form of top-down MS approach called "Native MS" is becoming popular in recent years. The novel aspect of native MS compared to traditional top-down MS is mostly the solvent used to solubilize protein complexes. Traditionally, acid and organic solvent are included in the analysis to increase the ionization efficiency while the protein is no longer in its native state. For native MS analysis, ammonium acetate buffer is used to maintain a neutral pH and a certain salt strength, thus the protein complex can be maintained in a near-native state[53]. This is especially beneficial for photosynthetic proteins, as almost all photosynthetic proteins are non-covalently associated with different co-factors, and photosynthetic protein complexes are usually composed by protein subunits non-covalently associated together. Our group has successfully utilized native MS to characterize multiple photosynthetic protein complexes. For instance, the stoichiometry of pigments in the antenna protein FMO was determined by native MS[54]. Native MS successfully analyzed the OCP protein complex with its pigment binding and further revealed that the OCP undergoes a dimer-to-monomer transition upon light illumination[55]. Later, the concentration effects on oligomerization states of FRP and OCP, two proteins involved in the NPQ process in cyanobacteria, was revealed by native MS[56] (See chapter 5 for this study). All the MS methods that are valuable in studying photosynthetic proteins are shown in Figure 1.2.

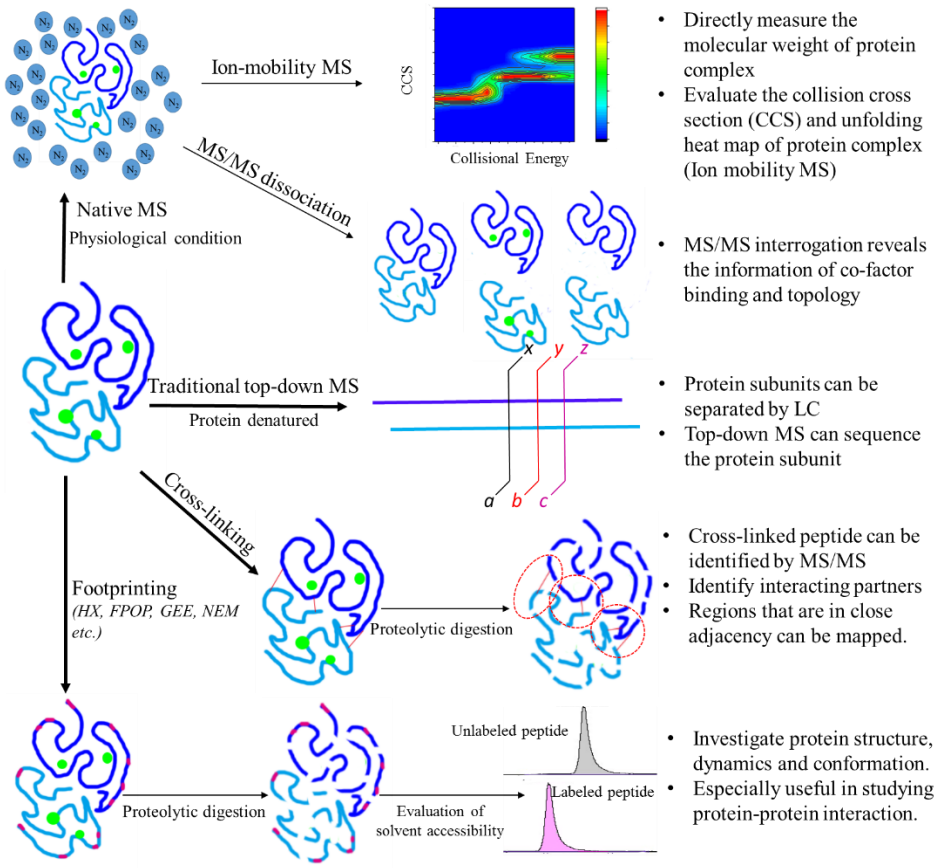


Figure 1.2 Mass spectrometry tools used to study photosynthetic protein complexes. Green dots represent co-factor binding to protein scaffold, such as chlorophyll binding.

1.6 Challenges in photosynthetic membrane protein MS study

The major protein machineries involved in photosynthetic process, reaction center and light harvesting antennas complexes, are mostly hydrophobic membrane proteins[1]. The information of membrane proteomics can be readily obtained with the assistance of bottom-up MS approach and multi-dimensional separation[57]. Considering the hydrophobicity of membrane proteins, detergent or organic solvent are usually added to increase the accessibility of cleavage site during proteolytic digestion[56] (Table 1.1). Alternative enzymes that can function at harsh conditions,

like CNBr, or addition of MS-compatible detergent have been adopted to improve the coverage of intrinsic membrane proteins[58-59]. However, it is difficult to obtain 100% sequence coverage on membrane proteins, especially the hydrophobic regions. As a complementary approach to bottom-up MS approach, top-down MS approach is even more popular in studying transmembrane photosynthetic machineries, considering the complex composition of photosynthetic proteins[60-61]. Instead of ubiquitously used silica-based reverse phase column, the polystyrene/divinylbenzene column was preferred when membrane protein needs to be separated[62] (Table. 1.1). Though *de novo* sequencing is not the strong suit of top-down MS, the composition of complex protein samples can be quickly evaluated with high confidence by top-down MS. For example, eleven integral and five peripheral subunits of a 750 kDa PS II complex from the eukaryotic red alga has been successfully characterized by top-down MS[63]. We identified a new isoform of β subunit in LH2 from purple bacteria by top-down MS coupled with ECD fragmentation (See chapter 2 for this study)[64].

To investigate the structure, conformation and function of photosynthetic membrane protein, native MS studies on membrane proteins are desired, though only a few studies have been conducted. In the native environment, lipids bilayers are essential to stabilize membrane proteins. Detergents are commonly used in membrane proteins study as a mimic of lipid bilayer. One of the major challenges lies in the universal usage of those detergents that interface with most MS detection techniques. The pioneering native MS work on membrane proteins performed by Barrera et al.[65] shows that a membrane protein complex can be kept intact in the gas phase of a mass spectrometer after shaking off detergent micelles. The optimization of backing pressure and desolvation energy was emphasized for membrane protein native MS experiment[66]. For photosynthetic membrane proteins, we successfully observed the intact reaction center

complexes from two purple bacteria, *Rb. sphaeroides* and *B. viridis*, with all the protein subunits as well as several cofactors in the gas phase (See reference 66 and chapter 4). Native MS shows that the bacteriochlorophyll dimer at the core is the tightest bound pigment in both two species, consistent with other observations[67]. In addition, we found the reaction center from *B. viridis* inclines to bind more bacteriochlorophylls, the special pair and the peripheral ones.

MS-based footprinting has successfully been applied on membrane proteins, including HDX, oxidative labeling and other covalent labeling methods[68-70]. The footprinting experiment can be carried out in detergent, Nanodisc or nature lipid environment. In chapter 2, we demonstrated how to utilize MS-based footprinting to study hydrophobic photosynthetic proteins in a near-native environment. After reconstituting LH2 in a Nanodisc vehicle, we successfully utilized fast photochemical oxidative labeling to investigate the structure and topology of LH2 in a lipid bilayer[71].

Table 1.1 MS-centered strategy for studying membrane proteins

Technique	Strategy for Membrane Protein Study
Protein clean-up	Acetone, chloroform/methanol/water[72]
Bottom-up MS (Proteolytic digestion)	Detergent or organic phase added Microwave assisted Alternative enzymes and chemical like thermolysin, chymotrypsin, CNBr and etc[59,73]
Bottom-up MS (Footprinting)	Incorporated in Nanodisc[71, 74] or dissolved in detergent micelle[75]
Top-down MS (Denatured)	Polystyrene/divinylbenzene column for separation[62] Fragmentation schemes: CID, ECD[64, 76]
Top-down MS (Native MS)	Increased backing pressure Increased desolvation energy[66]

Chapter 2: Top-down Mass Spectrometry Analysis of Membrane-bound Light-Harvesting Complex 2 from *Rhodobacter sphaeroides*

This chapter is adapted from the previously peer-reviewed and published first-authored manuscript:

Lu, Y., Zhang, H., Cui, W., Saer, R., Liu, H., Gross, M., and Blankenship, R. Top-Down Mass Spectrometry Analysis of Membrane-Bound Light-Harvesting Complex 2 from *Rhodobacter sphaeroides*. Biochemistry, 2015. 54(49): 7261-7271.

2.1 Abstract

We report a top-down proteomic analysis of the membrane-bound peripheral light-harvesting complex LH2 isolated from the purple photosynthetic bacterium *Rhodobacter (Rb.) sphaeroides.* The LH2 complex is coded for by the *puc* operon. The *Rb. sphaeroides* genome contains two *puc* operons, designated *puc1BAC* and *puc2BA*. Although previous work has shown consistently that the LH2 β polypeptide coded by the *puc2B* gene was assembled into LH2 complexes, there are contradictory reports whether the Puc2A polypeptides are incorporated into LH2 complexes. Furthermore, post-translational modifications (PTM) of this protein offer the prospect that it could coordinate bacteriochlorophyll *a* (Bchl *a*) by a modified N-terminal residue. Here we

describe the LH2-complex components based on electron-capture dissociation (ECD) fragmentation to confirm the identity and sequence of the protein subunits. We found that both gene products of the β polypeptides are expressed and assembled in the mature LH2 complex, but only the Puc1A-encoded polypeptide α is observed here. The methionine of the Puc2B-encoded polypeptide is missing, and a carboxyl group is attached to the threonine at the N terminus. Surprisingly, one amino acid encoded as an isoleucine in both the *puc2B* gene and the mRNA is found as valine in the mature LH2 complex, suggesting an unexpected and unusual post-translational modification or a specific tRNA recoding of this one amino acid.

2.2 Introduction

The capture and utilization of solar energy is one of the most fundamental processes on Earth. Anoxygenic photosynthesis can occur in the absence of air without producing oxygen[1, 77]. The photosynthetic complexes of purple phototrophic bacteria have a rather simple modular construction system that often utilizes two basic types of light-harvesting complexes, called light-harvesting complex 1 (LH1) and light-harvesting complex 2 (LH2). These functions to absorb light energy and to transfer that energy rapidly and efficiently to the photochemical reaction centers where it is trapped by photochemistry. LH2 is composed of heterodimeric units, consisting of α, β apoprotein pairs that serve as a scaffold to bind Bchl *a* and a carotenoid (Car) for optimal energy transfer. Those heterodimers aggregate to produce circular ring structures containing eight or nine heterodimeric units. The LH1 complexes have a similar heterodimeric building block but are comprised of 15 or 16 units that surround the photochemical reaction center complex. The LH1 complex from *Rb. sphaeroides* contains an LH1 dimer with 28 heterodimers[78]. LH2 complexes are adjacent to the LH1-reaction center (RC) core complex, and together, the two complexes effectively capture the light energy that sustains growth of the

organism. The ratio of LH2 complexes to RC is variable and depends on growth conditions[2, 79].

X-ray crystallography revealed structural data to atomic resolution of two types of LH2 complexes from *Rhodopseudomonas (Rps.) acidophil*[80] and from *Phaeospirillum (Ph.) molischianum*[81]. The structure at 100 K of LH2 from *Rh. acidophilus* was refined to 2.0 Å resolution by Papiz et al. [80] The crystal structure of LH2 from *Ph. molischianum* [81] was determined by molecular replacement at 2.4 Å resolution by using x-ray diffraction. In both structures, the modular α, β-heterodimers form a circular ring structure. LH2 from *Ph. molischianum* forms octamers instead of the nonamers observed in the *Rh. acidophilus* structure. Both types of LH2 complexes contain relatively isolated Bchl *a* molecules parallel to the plane of the membrane that absorb light at 800 nm (B800) and closely coupled Bchl *a* dimers that absorb at 850 nm (B850). One of the major differences of the two crystal structures lies in the nature of B800 orientation: aspartate is the Mg ligand in *Ph. molischianum* as opposed to carboxyl-methionine in *Rh. acidophilus.*

Rhodobacter (Rb.) sphaeroides, a member of the α-3 subclass of proteobacteria, is an exemplary model organism for the creation and study of novel protein expression systems as its genome is sequenced, genetic systems are available, and its metabolism is well characterized[82]. Like many purple phototrophic bacteria, the photosynthetic apparatus of *Rb. sphaeroides* is composed of three multimeric transmembrane protein complexes: the LH2 light-harvesting complex, the LH1-reaction-center complex (RC-LH1), and the cytochrome (cyt) bc1 complex[83]. The 3D structure of a dimeric RC-LH1-PufX complex was determined to 8 Å by x-ray crystallography, and a model was built[78]. Although there is no atomic resolution structure of LH2 from *Rb. sphaeroides* available to date, a projection map of this LH2 clearly shows the nonameric organization of the ring[84]. This is also observed for LH2 from *Rh. acidophilus,* but

different from the octamer-ring structure of LH2 from *Ph. molischianum*. Theiler et al.[85] sequenced the apoproteins of LH2 from *Rb. sphaeroides* in 1984 and found a degree of heterogeneity at the N-terminus of β subunit, with some chains starting with threonine and others having an additional methionine residue at the N-terminal position. DNA sequence of the photosynthesis region of *Rhodobacter sphaeroides* 2.4.1 is described by Choudhary et al.[86] and the amino acid sequence predicted by the genome is consistent with the previous protein sequencing result. A few years later, a new operon (designated the puc2BA operon), displaying a high degree of similarity to the original pucBA genes of Rb. sphaeroides, was identified and studied genetically and biochemically by Zeng et al in 2003[87] (Figure 2.1). Employing genetic and biochemical approaches, they obtained evidence that the Puc2B-encoded polypeptide is able to enter into LH2 complex formation, but neither the full-length Puc2A-encoded polypeptide nor its N-terminal 48-amino-acid derivative is able to enter into LH2 complex formation. In contrast, Wang et al.[88] isolated LH2 from mutated strains and found Puc2B and both the N-terminal version and the intact version of the Puc2A-encoded polypeptides. They suggested that the transcription of puc2BA and the assembly of the LH2 complex is independent of the expression of puc1BA, and is only dependent upon the expression of pucC. According to the result from SDS-PAGE, either the first 54 amino-acid residues of the N-terminus or the one containing a 251 residue C-terminal extension of Puc2A encoded polypeptide can be assembled into the LH2 complex. It is possible that the manipulations of genome affect the assembly of LH2 complex, leading to the contradictory results. Later, Woronowicz et al.[89] described a proteomic analysis of the expression levels of the various Puc1BA and Puc2BA operon-encoded polypeptides in the LH2 complexes assembled in *Rb. sphaeroides*. Surprisingly, the Puc2A polypeptide containing a 251 residue C-terminal extension is of major abundance. It was also reported that genomes of *Rh.*

acidophilus and *Rhodopseudomonas (R.) palustris* contain additional, highly homologous copies of the *puc* operon encoding the α, β polypeptides of the LH2 complexes[90-91]. All five copies of the *puc* operon in *R. palustris* were expressed and regulated by incident light intensity, whereas only two copies of the *puc* operon products were detected in LH2 complexes from *Rh. acidophilus.*

```
Puc1a     MNQGKIWTVVDPAVGIPLLLGSVAVTALLVHLAILQNTTWFPAFMQGG---------LKK
Puc2a     MNNSKMWLTVNPNLGVPLLLGSVAVASLVVHGAVLTTTPWIANYYQGSEPWPVAAAPAEE

Puc1a     AAAIVQVVG-----------------------------------------------------
Puc2a     AAAPVEAAAPADEAAAPVEEAAPVAEAAAPAEEAAPAAEAAVPAEEAAPAAEAAAPAEEA

Puc1a     -------------------------------------------------------------
Puc2a     APAAEAAAPAEEAAPAAEAAAPVEEAAPAAEAAAPAEEAAPVAEAAAPAEEAAPVAEPAA

Puc1a     -------------------------------------------------------------
Puc2a     EPAPAAEAAAPVAEVSAPAAELAAPVAMSLVDIAAKLNGLGYSVQSVTKTEGGYVVNMTD

Puc1a     ------------------------
Puc2a     ANGMPVAATLDPVTGLPFVPAAQ
```

Figure. 2.1 Sequence alignment of Puc1A- and Puc2A-encoded polypeptides

We applied mass spectrometry (MS) to this problem because MS is now playing a role in intrinsic-membrane protein (IMPs) analyses, and high-throughput proteomics technology can accelerate the understanding of membrane protein structure/function relationships. Precise characterization of whole intrinsic membrane protein (IMPs), however, remains a challenge despite their essential roles in cell biology. The hydrophobicity of IMPs makes them difficult to be analyzed by traditional bottom-up mass spectrometry owing to its bias toward soluble, hydrophilic peptides that are easily recovered during sample processing and chromatography, and that ionize and dissociate well during mass spectrometry[58]. For example, many membrane

proteins are insoluble under the conditions for enzyme digestion, and subsequent steps in analysis could further lead to precipitation.

In this chapter, we report a top-down MS study of the intact LH2 from wide type *Rb. sphaeroides* to identify and sequence this peripheral antenna system. IMP solubility also challenges the "top down" approach in which intact proteins are introduced directly to the mass spectrometer. In early work from Whitelegge et al.[92], they studied the seven-transmembrane helix protein bacteriorhodopsin and the Dl and D2 reaction-center subunits from spinach thylakoids and demonstrated the potential of top-down analysis of IMPs. Later, they described using top-down high-resolution Fourier transform mass spectrometry with collision-induced dissociation (CID) to study post-translationally modified integral membrane proteins with polyhelix bundles and transmembrane porin motifs[93]. Whereas CID fragmentation occurs by increasing the internal energy of peptide/protein ions and causing protons to move, electron-capture dissociation (ECD) generally breaks bonds near the location of a protonated site that can attract the electrons[21, 51]. The top-down analysis on the c-subunit of ATP synthase (AtpH) shows that thermal activation concomitant with electron delivery increased coverage in the transmembrane domain compared to CID fragmentation[94].

2.3 Material and Methods

2.3.1 LH2 preparation

Rb. sphaeroides wild-type strain ATCC 2.4.1 was grown photosynthetically at RT in 1 L bottles. The membrane-enriched pellet obtained from ultracentrifugation of the sonicated cells was re-suspended in 20 mM Tris (pH = 8.0) to a final concentration of OD (850) = 50,

solubilized by the addition of lauryldimethylamine *N*-oxide (LDAO) to a concentration of 1% (w/v), and allowed to incubate for 30 min at room temperature. Solubilization was stopped by dilution of the mixture with 20 mM Tris (pH = 8.0) to a final LDAO concentration of 0.1%. This mixture was ultracentrifuged once again at 200,000xg for 1 h to remove insoluble debris. The supernatant was collected and loaded onto an anion-exchange column (QSHP resin, GE Healthcare, Uppsala, Sweden) that had been equilibrated with 20 mM tris-HCl, 0.1% (w/v) LDAO (pH = 8.0). After washing extensively, LH2 was then eluted with a linear gradient from 100 mM to 500 mM NaCl. Fractions with the highest A_{850nm}:A_{280nm} ratios (greater than 3.0) were pooled, and the accumulated sample applied to a HiLoad™ Superdex™ 200 prep grade column (GE Healthcare). The further purified LH2 was precipitated with acetone and then solubilized with 20% formic acid before being infused into the MS spectrometer.

2.3.2 Top-down LC-MS analysis of LH2

Resins (PLRP/S, 5 μm, 1000 Å) were packed into 100 μm IntegraFrit capillary (Waters Inc., Milford, MA). A NanoAcuity UPLC (Waters Inc., Milford, MA) was used to separate protein subunits. The gradient was delivered by a NanoAcuity UPLC (0–5 min, 15% solvent B; 5–35 min, 15–90% solvent B. Solvent A: water, 0.1% formic acid; Solvent B: acetonitrile, 0.1% formic acid) at a flow rate 1 μL/min. Two mass spectrometers, a hybrid ion-mobility quadrupole ToF (Synapt G2, Waters Inc., Milford, MA) and a 12 T FTICR mass spectrometer (Solarix, BrukerDaltonics, Bremen, Germany) were operated under normal ESI conditions (capillary voltage 1-2 kV, source temperature ~ 100 °C). The typical ECD pulse length was 0.4 s, ECD bias 0.4 V, and ECD lens 10 V. The ECD hollow cathode heater current was 1.6 A. MS parameters were slightly modified for each individual sample to obtain an optimized signal. For introduction to give ECD fragmentation, an Advion Triversa Nanomate sample robot infused the

sample into the 12 T FTICR. Precursor ions were each isolated over a 10 *m/z* window. Data were processed by using Bruker Daltonics BioTools and Protein Prospector (from the University of California-San Francisco MS Facility web site). Manual data interpretations combined with software tools were adapted to achieve improved sequence coverage. The mass tolerance for fragment ions assignment was 0.02 Da.

2.3.3 puc2B gene identification

Rb. sphaeroides genomic DNA was extracted by Qiagen® DNeasy Plant Mini Kit. *puc2B* genes were PCR-amplified by left primer GCTCCGAGCCCTGATAGTAG and right primer AAGCTGAGCAGAGGGGTCTT. The purified PCR product was cloned and sequenced.

2.3.4 puc2B mRNA identification

Rb. sphaeroides genomic RNA was extracted by TRIzol® Reagent (Life Technologies, Grand Island, NY). Briefly, the cells were broken by ultrasonification in TRIzol® Reagent. After phase separation by chloroform and precipitation by isopropyl alcohol, RNA precipitates were washed by 70% ethanol. TURBO DNA-free™ Kit (Life Technologies, Grand Island, NY) was used to eliminate any remaining DNA contamination. The first strand cDNA was synthesized by RevertAid First Strand cDNA Synthesis Kit (Life Technologies, Grand Island, NY). After synthesis of the first-strand, primer-1 and primer-2 were used to PCR-amplify the *puc2B* sequence. The purified PCR product was cloned and sequenced.

2.3.5 Homology modelling

Homology models of the two subunits were generated by using the Phyre225 online modeling suite. The two subunits were combined, and energy minimization was performed by

Phenix.[95] The top model was aligned to the crystal structure of LH2 from *Rh. acidophilus* [80] by Pymol (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.).

2.4 Results and discussion

2.4.1 Composition of mature LH2 complexes

To study the composition of mature LH2 complexes, we purified the whole complex from a photosynthetically cultured wild type strain ATCC 2.4.1 of *Rb. sphaeroides*. There are three major protein components as seen in the chromatogram of the denatured LH2 protein; these components were later identified as the Puc1B-, Puc2B-, and Puc1A-encoded polypeptides by top-down MS (Figure 2.2). The Puc2A α polypeptide was not detected. The reason might either be the Puc2A α polypeptide is not present in mature LH2 complex or our platform failed to detect Puc2A α polypeptide. The β subunits were eluted earlier than the α subunit as the latter is more hydrophobic. Because there is 94% sequence identity of the two β subunits, the elution times of the two β copies are nearly identical. This result is in accord with those of Zeng et al.[87], who also found that only the Puc1A-encoded α subunit exists in the mature LH2 complex. Wang et al.[88] isolated LH2 from mutated strains and found Puc2B- and both the N-terminal version and the intact versions of the Puc2A-encoded polypeptides. Later, Woronowicz et al.[89] found that the Puc2A-encoded polypeptide containing a 251 residue C-terminal extension is a highly abundant protein of the LH2 complex. The large protein fragment they detected does not contain any apparent membrane-spanning regions. They suggested that this peptide is not part of the functional complex and instead arises from *in vivo* enzymatic cleavage, representing an adventitious co-eluent of CNE and readily detected because mass spectrometers have a bias toward detecting soluble peptides. The existence of certain peptides from Puc2A-encoded

polypeptide indicates, however, it may have some assembly role in the complex. The differences between these results are presently not fully understood. It is possible that the growth conditions and the light intensity can affect the expression and assembly of those polypeptides. Nevertheless, the Puc2A-encoded polypeptide is not a major component according to our results. The experimental MW of each subunit were deconvoluted by Bruker Compass DataAnalysis software and shown in Table 2.1.

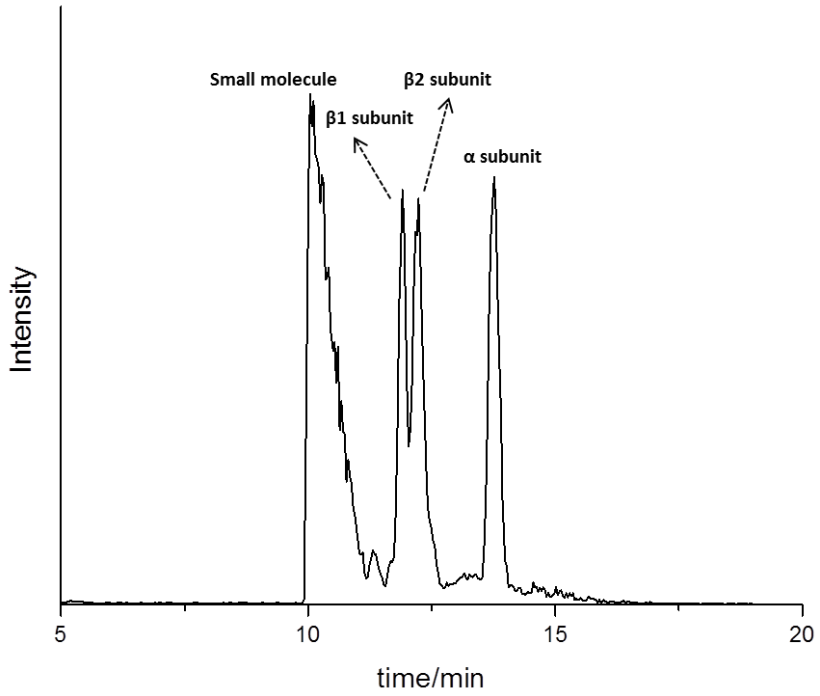


Figure 2.2. Liquid chromatogram of denatured LH2.

Table 2.1. Molecular weight of each subunit from LH2.

Subunit	Theoretical MW(mono)	Experimental MW(mono)
Puc1A polypeptide	5595.0692 Da	5595.0512 Da
Puc1B polypeptide	5456.8854 Da	5456.8706 Da
Puc2B polypeptide	5355.8191 Da	5355.8082 Da

2.4.2 Sequence and post translational modifications

Natural photosynthetic organisms have developed a large variety of light-harvesting strategies that allow them to live nearly everywhere where sunlight can penetrate.[96] Most of the antenna systems are pigment-containing, integral membrane proteins. Detailed sequence information of some of those proteins is still not fully known except when there is a high-resolution crystal structure available. Although the MS-based proteomics characterization of the *Rb. sphaeroides* intra-cytoplasmic membrane assembly was reported by several groups,[89, 97-98] the entire sequence was not identified because 100% coverage was not achieved, which is usually the case for membrane proteins. Those missing regions may play an important role of the function of antenna systems.

We used top-down mass spectrometry to determine the sequence and post-translational modifications (PTM) information of the LH2 complex. Although not every bond of the polypeptides fragmented when we submitted the protein to ECD on a 12 T FTICR mass spectrometer, we found ~ 70% coverage of the sequence. Furthermore, many complementary ions shown in the spectrum and their accurate mass measurements (within a few ppm) provided by the instrument provide high confidence for the results. Unlike CID fragmentation, which occurs by increasing the internal energy of peptide/protein ions until peptide-bond cleavage occurs, electron-based fragmentation (ECD) breaks different bonds near the sites of positive charge where the electron capture occurs, preserving PTM information. The ECD-based top-

down sequencing identifies not only the sequence information of the protein but also the location of the PTMs.

The PTM information of LH2 complex has been of interest for some time. Papiz et al.[80] reported the high-resolution crystal structure of *Rh. acidophilus*, and they found that a carboxyl-modified Met1 of the α subunit is ligated to Mg^{2+} of B800. Our top-down MS investigation of the structure of LH2 and the possible coordination of B800 BChl-*a* shows that the experimental molecular weight of the α subunit is consistent with that predicted from the gene sequence (3 ppm accuracy), clearly indicating there is no carboxylation modification on Met in *Rb. sphaeroides* LH2. We observed some oxidation of methionine, and this was probably introduced during the sample handling. MS/MS with ECD fragmentation provided further evidence that the predicted sequence is correct (Figure 2.3).

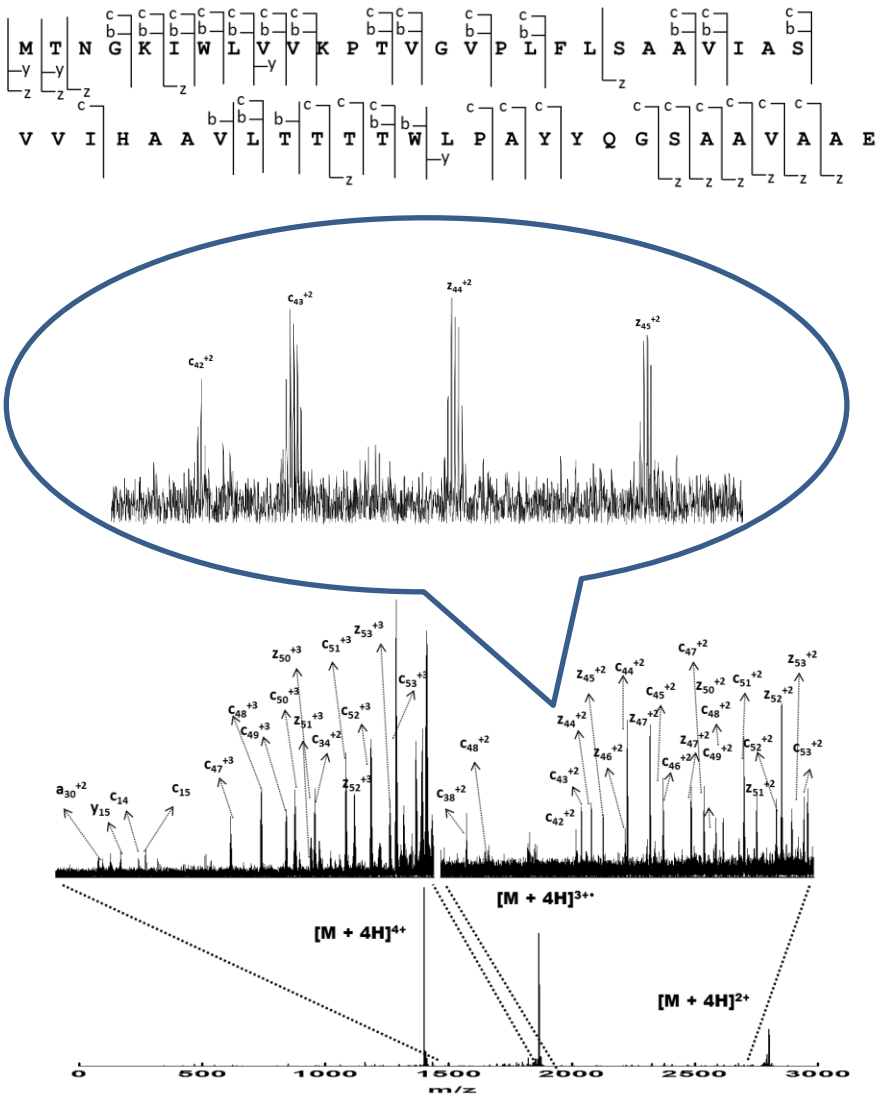
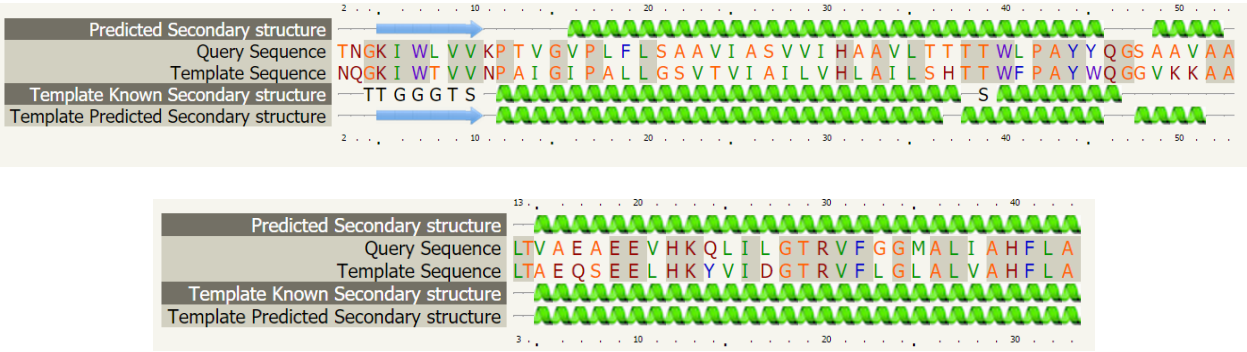


Figure 2.3. Sequence coverage and ECD product-ion spectrum of Puc1A-encoded polypeptides

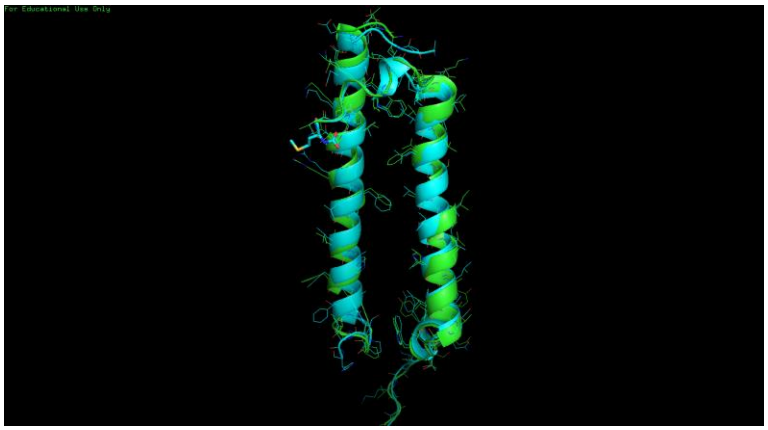
The central Mg^{2+} ion chelation in the core of the Bchl *a* macrocycle helps preserve the planar conformation of the pigment molecule[99]. In principle, oxygen or nitrogen atoms on amino acid side chains (e.g., aspartate, glutamate, asparagine, glutamine, serine, threonine, histidine) or even water can interact with this central Mg atom of BChl *a*. In the LH2 complex of *Rh. acidophilus*, the ligation of B800 BChl *a* is accomplished in part by a carboxyl modification on methionine

whereas in *Ph. molischianum*, the corresponding ligand is aspartate (Asp-6). For *Rb. sphaeroides*, there is no modification of the α-Met, indicating that the N-terminal Met is not the ligand. To obtain a better understanding of the structure, 52 residues (96% of sequence) were modelled with 100% confidence by the single highest scoring template (Figure 2.4A),[100] suggesting that the N terminal region of α subunit of *Rb. sphaeroides* is quite similar to that of *Rh. acidophilus* (Figure 2.4C). The reason they have similar structures but different coordination schemes is not clear. To identify similar regions that may be a consequence of structural relationships, we aligned the sequence of *Ph. molischianum* and *Rh. acidophilus* from RCSB protein data bank to the sequence of *Rb. sphaeroides* we identified. From the sequence alignment result from ClustalW2, the amino acid in *Rb. sphaeroides* is Asp-6 whereas in *Ph. molischianum*, it is asparagine; the former is likely the Mg^{2+} ligand in the LH2 complex (Figure 2.5).

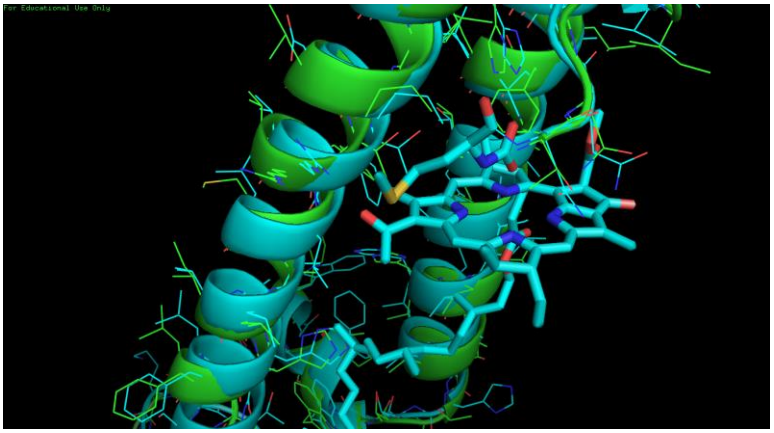
(A)


Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure
TNGKIWLVVKPTVGVPLFLSAAVIASVVIHAAVLTTTTWLPAYYQGSAAVAA
NQGKIWTVVNPAIGIPALLGSVTVIAILVHLAILSHTTWFPAYWQGGVKKAA
TT G G G T S
S
Predicted Secondary structure
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure
LTVAEAEEVHKQLILGTRVFGGMALIAHFLA
LTAEQSEELHKYVIDGTRVFLGLALVAHFLA

(B)



(C)



(D)

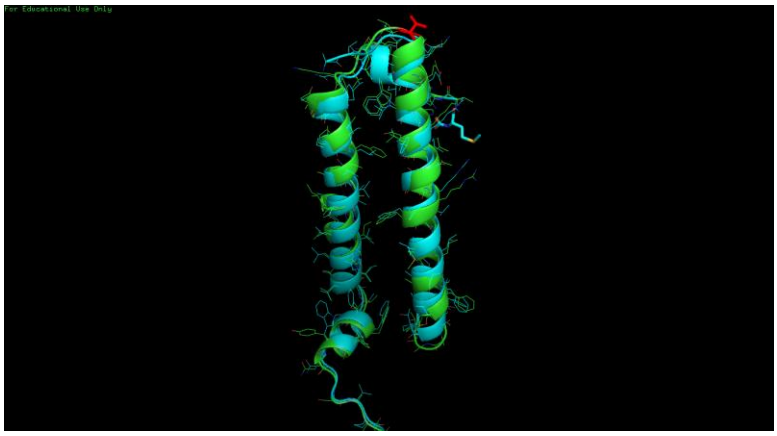


Figure 2.4. Homology model of LH2 from *Rb. sphaeroides*. (A) Sequence alignment to the template (B) LH2 from *Rb. sphaeroides* is shown in green, LH2 from *Rh. acidophilus* is shown in cyan (C) N-terminus structure of two LH2. B800 is coordinated by Carboxyl-Met from *Rh. acidophilus.* (D) Valine from β subunit of *Rb. sphaeroides* is shown in red.

(A)

```
1  ---MNQGKIWTVVNPAIGIPALLGSVTVIAILVHLAILSHT--TWFPAYWQGGVKKAA- 53
2  ---MTNGKIWLVVKPTVGVPLFLSAAVIASVVIHAAVLTTT--TWLPAYYQGSAAVAAE 54
3  SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGFNWIAL---GAAKSAAK 56
        : *** *::*:  :* :   ..: :: :* *:*: .  .*:.    *..  **
```

(B)

```
Puc1B  MTDDLNKVWPSGLTVAEAEEVHKQLILGTRVFGGMALIAHFLAAAATPWLG 51
Puc2B  -TDDPKKVWPSGLTVAEAEEVHKQLILGTRVFGGMALIAHFLAAAATPWLG 50
        *** :**********************************************
```

Figure 2.5. Sequence alignments of LH2. (a) α subunits from different LH2 complexes: (1), *Rh. acidophilus* (2), *Rb. sphaeroides* (3), *Ph. molischianum* (b) Puc1B- and Puc2B-encoded polypeptides from *Rb. sphaeroides.*

Similarly, the molecular weight obtained from the mass spectrum is consistent with the theoretical MW predicted from the amino-acid sequence of the Puc1B-encoded peptide without

any post-translational modifications (3 ppm accuracy) (Table. 2.1 and Figure 2.6). The MW observed for the Puc2B-encoded peptide, however, is 101.0589 Da less than predicted. The mass difference doesn't match a simple modification with or without removal of the N terminal methionine. To address this discrepancy, we analyzed the fragmentation patterns and found that the methionine on the N terminus is removed and a carboxyl group is attached to threonine. In addition, the fourteenth amino acid counting from the N terminus, is a valine instead of the isoleucine that is coded in the gene sequence (Figure 2.7). The fragment ions (C_{13}^{+}, C_{14}^{+}, Z_{36}^{3+}) displayed on the spectrum confirm this assignment.

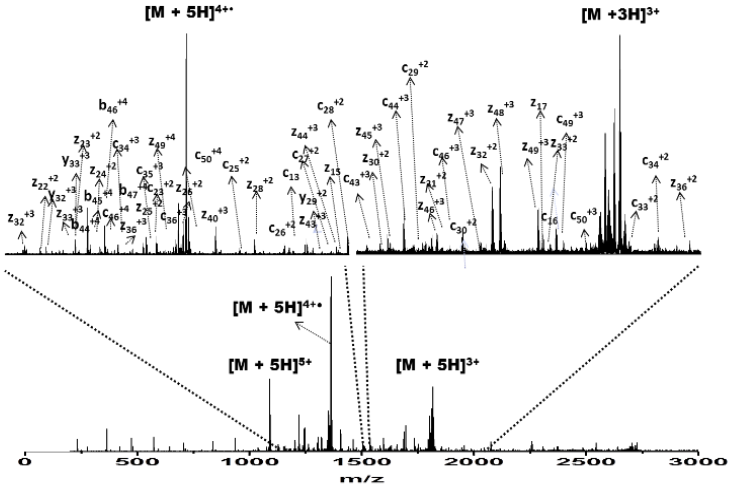


Figure 2.6 Sequence coverage and ECD product-ion spectrum of Puc1B-encoded polypeptides.

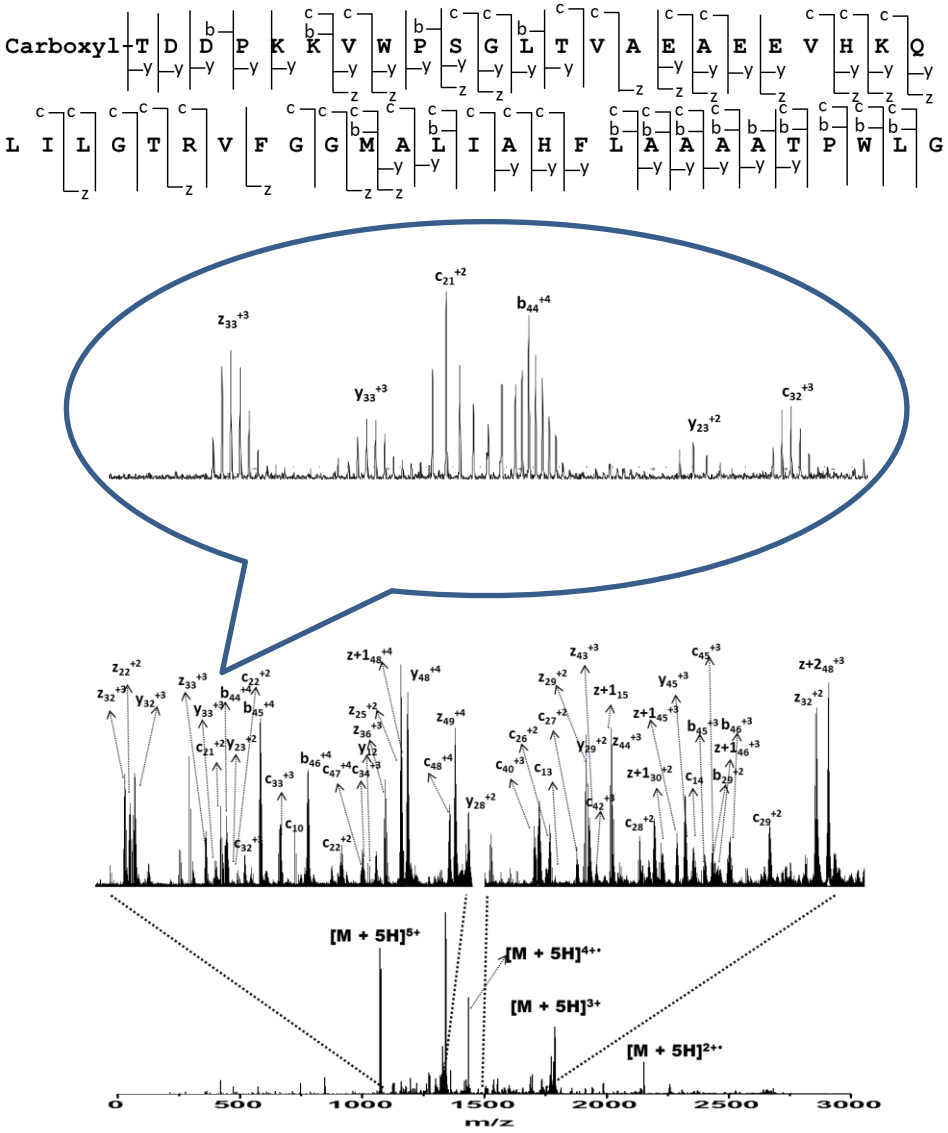


Figure 2.7 Sequence coverage and ECD product-ion spectrum of Puc2B-encoded polypeptides.

We were surprised to find carboxylation on the threonine residue at the N terminus. The high mass fragments (Figure 2.8) clearly show that losses of OH, COOH, COONH, and then threonine from N terminus, consistent with this modification. The simulated isotopic pattern of fragments on the N terminus is also consistent with the experimental patterns (Figure 2.9). To

verify our interpretation of the dissociation spectrum, we undertook bottom-up sequencing. The N terminal peptide has a COOH modification on threonine according to CID (Figure 2.10).

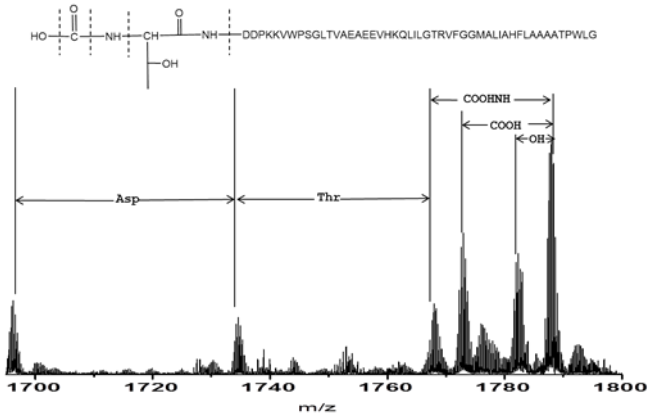
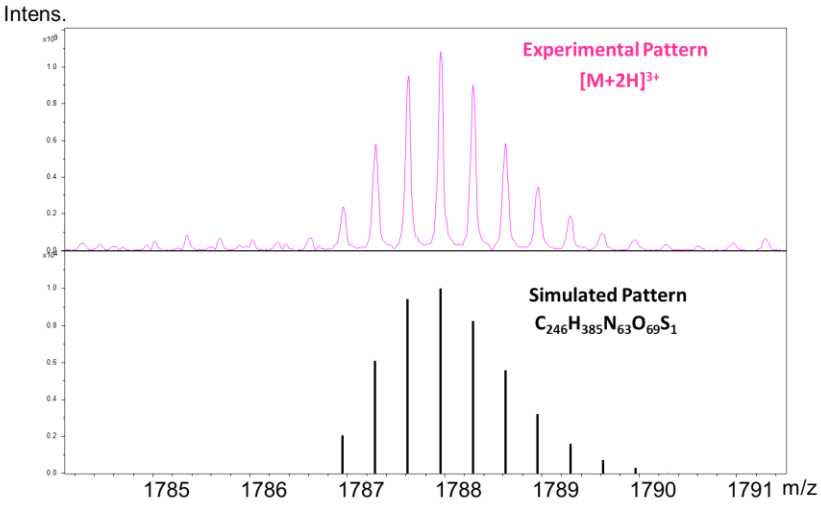
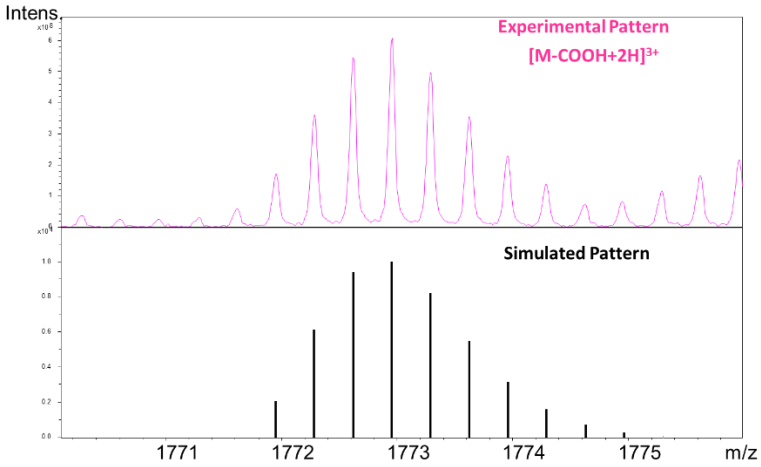


Figure 2.8. High mass fragments of ECD product-ion spectrum of Puc2B-encoded polypeptides.

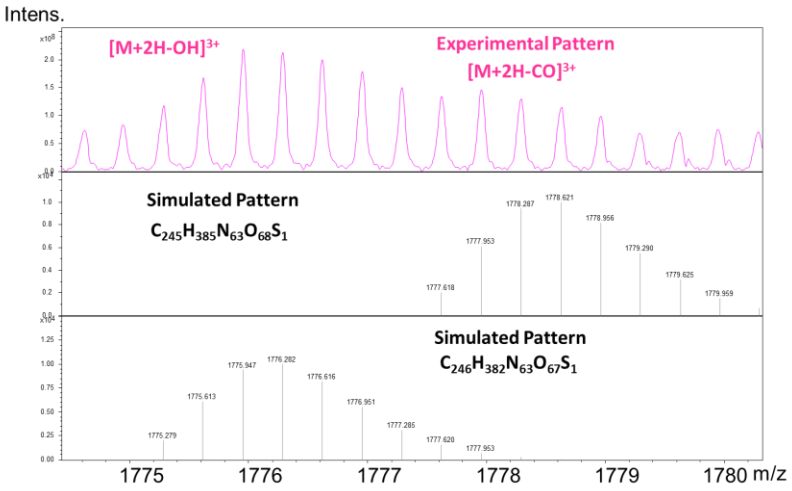


$C_{246}H_{385}N_{63}O_{69}S_1^{3+} \longrightarrow C_{245}H_{384}N_{63}O_{67}S_1^{3+} + COOH$

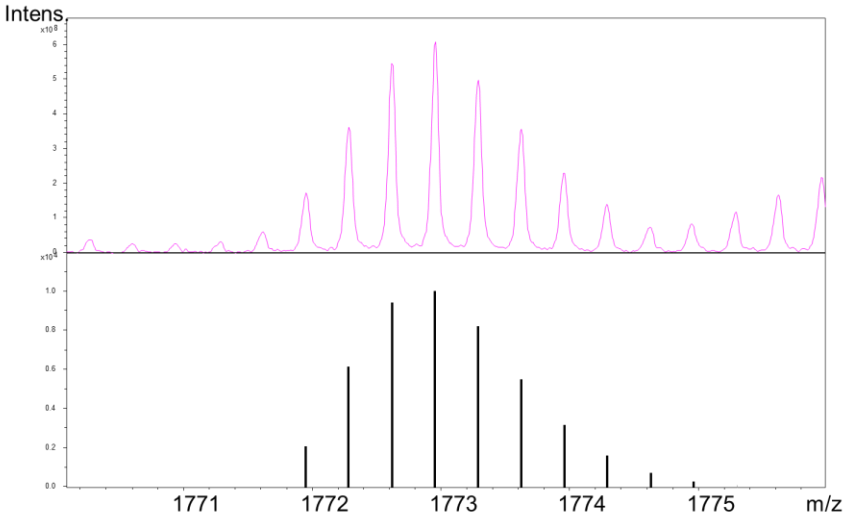


$C_{246}H_{385}N_{63}O_{69}S_1^{3+} \longrightarrow C_{246}H_{382}N_{63}O_{67}S_1^{3+} + OH$

$C_{246}H_{385}N_{63}O_{69}S_1^{3+} \longrightarrow C_{245}H_{385}N_{63}O_{68}S_1^{3+} + CO$



$C_{246}H_{385}N_{63}O_{69}S_1^{3+} \longrightarrow C_{245}H_{384}N_{63}O_{67}S_1^{3+} + COOH$



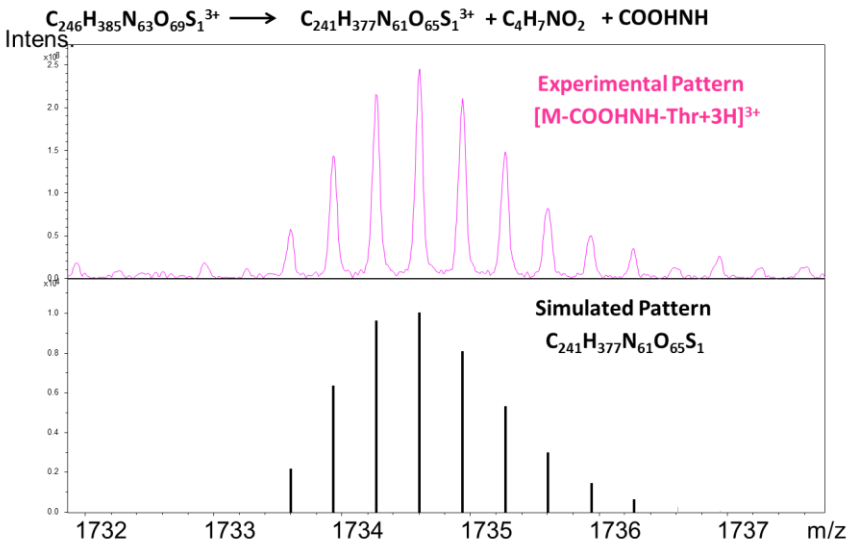


Figure. 2.9. Comparison of simulated and experimental isotopic pattern of parents and fragment ions on N-teriminus of Puc2B-encoded polypeptide

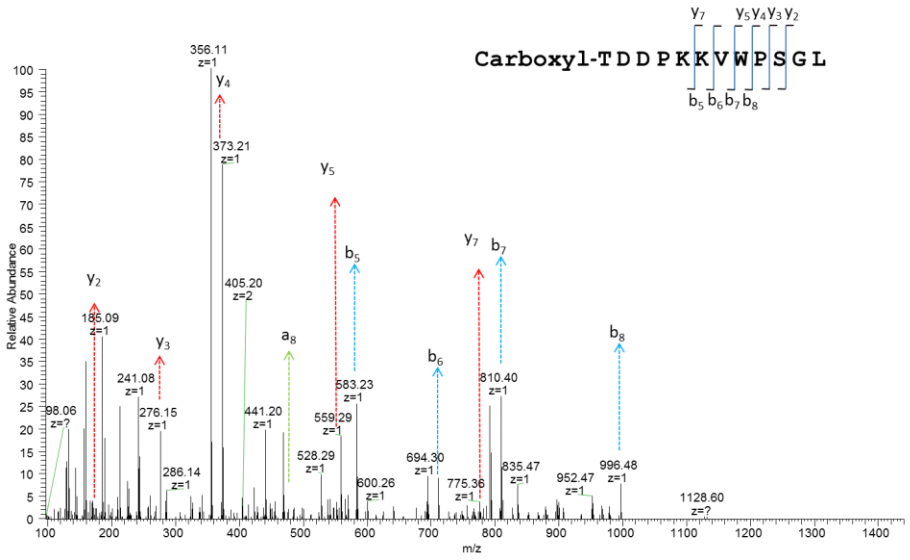


Figure. 2.10 MS/MS fragmentation of N terminus peptide from Puc2B-encoded subunit

Proteins can carry several PTMs, and some proteins may display large numbers of different modifications[101]. New modifications[102] and unexpectedly extensive PTMs[103] can occur, and they are poorly accounted for in existing databases. There are several reports about PTMs of light-

harvesting proteins. The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry revealed gene products with variable post-translational modifications[104]. Michael et al.[105] found acetylation and phosphorylation on spinach light-harvesting chlorophyll protein II aside from the removal of methionine at the N terminus. A proteomics study of the green alga *Chlamydomonas reinhardtii* light-harvesting proteins shows the presence of differentially N-terminally processed forms of Lhcbm3 and phosphorylation of a threonine residue at the N terminus[106]. For *Rh. acidophilus*, the ligation of B800 BChl *a* is accomplished in part by a carboxyl-methionine on the N terminus[80].

Carboxylation generally happens on glutamate residues, which is required for function of factors II, VII, IX, and X, protein C, protein S, and some bone proteins[107-108]. Although our study identified the carboxylation of a threonine residue at the N terminus of Puc2B, the functional role of this PTM is not clear. The assignments of the N terminal fragments further confirm the sequence and PTM information of this polypeptide (Figure 2.8). Zeng et al.[87] found that the ratio of B800 to B850 of the LH2-2 complex in mutant ΔPUC2BA (pUC2ASPhoA) is greater than that of the LH2-1 complex in mutant ΔPUC2BA (0.75 and 0.67, respectively). This result suggests that the Bchl *a* moieties have a slightly altered binding environment in the Puc2B-encoded complex compared to that of the Puc1B-encoded peptide. For the β subunit, 38 residues (75% of sequence) were modelled with 99.9% confidence by the single highest scoring template[100]. The first twelve amino acids are not covered in the homology modeling because there is high sequence discrepancy of the N terminal region of β subunit. It is possible, however, that the carboxyl group serves to coordinate the Mg of B800 in the LH2 complex, increasing the ratio of B800 to B850.

2.4.3 Substitution of Valine for Isoleucine

Initially, we could not match the experimental molecular weight of Puc2B to the molecular weight predicted by the genome, even after removal of methionine and attachment of a carboxyl group on the N terminal threonine. As discussed above, an analysis of the C and Z ions that are produced upon ECD shows that the 14th amino acid from N terminal is valine, not the isoleucine predicted from the gene sequence. The experimental MW matches the theoretical value within 2 ppm, confirming this assignment. Interestingly, we observed two polypeptides, mostly as the "valine version", but there is also a small amount of the "isoleucine version" in a roughly 10:1 ratio (Figure 2.11). The sequence chemically determined by Theiler et al. [85] matches the predicted sequence of the Puc1B protein. It is likely that their samples also contained the Puc2B protein but did not obtain sequence from it due to the blocked N-terminus.

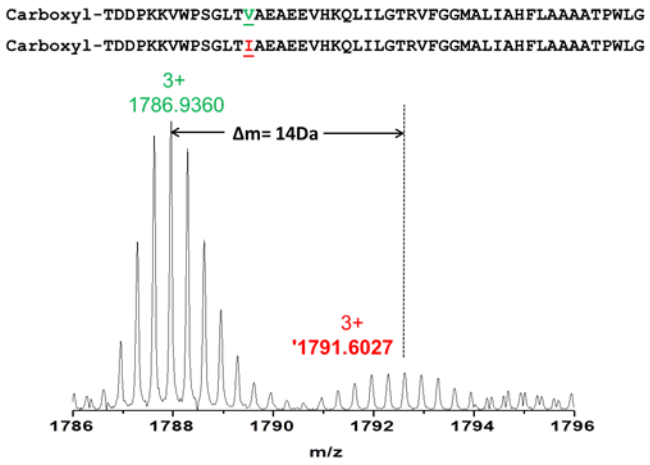


Figure 2.11. Mass spectrum of parent ions of Puc2B-encoded polypeptides.

The question now arises as to the origin of this isoleucine to valine conversion. The spinach chloroplast genome reports a codon for Ser at position 2 whereas Phe was detected at position 2 of PetL of chloroplast-encoded subunits[109]. The authors believe there might be either a DNA

sequencing error or a RNA editing event. To eliminate the possibility of a mutation happening during culturing of cells over many generations, we sequenced the puc2B gene, and we found it to be the same as in the NCBI database (i.e., the 14th codon is ATC, which codes for isoleucine). Another possibility to explain the substitution is RNA editing in which the codon is changed after transcription but before translation. To study the Puc2B-encoded peptide at the transcriptional level, we sequenced the mRNA that is encoded for this region. The result shows that there is no variation at the mRNA level, and the mRNA also indicates isoleucine (See Figure 2.12). Another possibility is a mis-sense error, which results in the substitution of one amino acid for another probably by mischarging the isoleucine tRNA with valine. However, all other isoleucines in the protein, all of which are coded for by the ATC isoleucine codon, appear to be correctly inserted, so the only way that this could be the case is for a context-specific change to be made in just this one place. Another possibility is that the amino acid residue is not valine but its isomer, norvaline, which is sometimes abundant[110-111]. It is also likely that it is a PTM process. The existence of two forms of the polypeptides also suggests that the demethylation is not 100% complete. Thus, we suggest the isoleucine-to-valine conversion is either a posttranslational modification in which a methyl group is cleaved from isoleucine or a norvaline substituion.

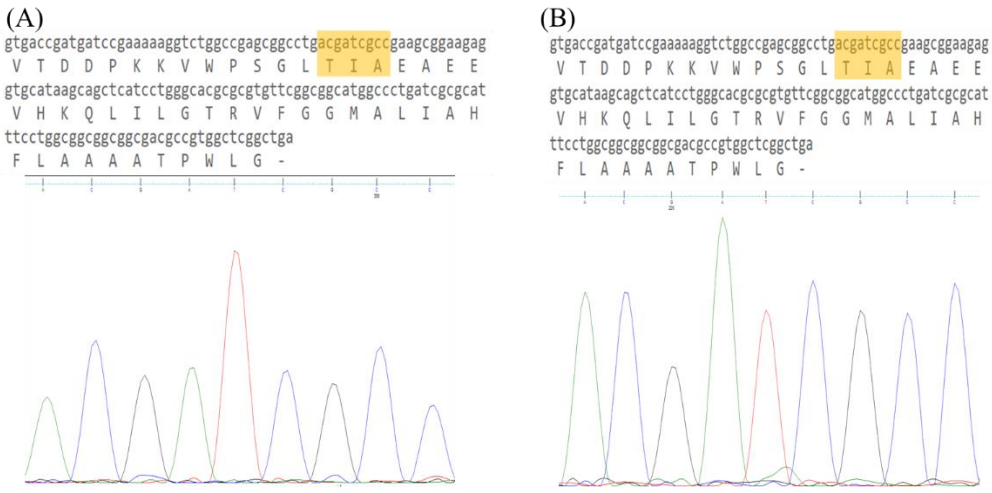


Figure. 2.12. DNA (a) and RNA (b) Puc2B operon sequencing result and trace files around 14th valine region.

Although there are very few protein demethylation cases reported, one well-known example is histone demethylation mediated by the nuclear amine oxidase homolog LSD1[112]. To the best of our knowledge, the post-translational demethylation process of isoleucine to valine has not been reported previously. The norvaline is usually found in place of leucine instead of isoleucine[110]. According to the homology modelling structure, the valine is located in the loop region of the β subunit (Figure 2.4D). The biological significance of this conversion of the Puc2B-encoded polypeptide is not clear at this stage, but given the unprecedented nature of this substitution, more study is needed. After this conversion, three amino acids on the N terminus are different and the rest of the amino acids are the same for the two copies of the β polypeptides.

Although there are very few protein demethylation cases reported, one well-known example is histone demethylation mediated by the nuclear amine oxidase homolog LSD1.41 To the best of our knowledge, the post-translational demethylation process of isoleucine to valine has not been reported previously. The norvaline is usually found in place of leucine instead of isoleucine.42 According to the homology modelling structure, the valine is located in the loop region of the β subunit (Fig. 3D). The biological significance of this conversion of the Puc2B-encoded polypeptide is not clear at this stage, but given the unprecedented nature of this substitution, more study is needed. After this conversion, three amino acids on the N terminus are different and the rest of the amino acids are the same for the two copies of the β polypeptides (Fig. 4B).

2.5 Conclusions

The composition and PTMs of LH2 of purple phototrophic bacteria likely play important roles in absorbing light energy effectively and in allowing the organism to adapt to a changing environment (i.e., light intensity). In this study, a top-down proteomic analysis of the membrane-bound peripheral light harvesting complex LH2 isolated from WT *Rb. sphaeroides* confirms the identity and sequence of these protein subunits. We showed that polypeptide encoded by puc1A is the sole source of the α subunit in the LH2 complex. Consistent with previous reports, this work also shows that both LH2 β polypeptides coded by the puc1B and puc2B gene are assembled into LH2 complexes. Overall, the complex has a similar structure to those in other purple bacteria. Unusual PTMs occur for this protein. For example, a carboxyl group is attached to the N-terminal Thr along with removal of the Met on the Puc2B-encoded polypeptide. The carboxyl group, instead of the carboxyl-methionine in *Rh. acidophilus*, likely coordinates Bchl a in *Rb. sphaeroides*. The unexpected substitution of valine to isoleucine in Puc2B-encoded polypeptide is likely to be a PTM or a norvaline substitution. The biological significance of this conversion is currently not clear and will be the subject of future studies.

Chapter 3: Mapping the Topology and Conformation of an Intrinsic Membrane Protein in a Lipid Bilayer by Fast Photochemical Oxidation of Proteins (FPOP)

This chapter is adapted from the previously peer-reviewed and published first-authored manuscript:

Lu, Y., Zhang, H., Dariusz, M. Niedzwiedki, Jiang, J, Gross, M., and Blankenship, R. Mapping the topology and conformation of an intrinsic membrane protein in a lipid bilayer by fast photochemical oxidation of proteins (FPOP). Analytical Chemistry. 2016. 88 (17): 8827-8834.

3.1 Abstract

Although membrane proteins are crucial participants in photosynthesis and other biological processes, many lack high resolution structures. Prior to achieving a high resolution structure, we are investigating whether MS-based footprinting can provide coarse-grained protein structure by following structural changes that occur upon ligand binding, pH change, and membrane binding. Our platform probes topology and conformation of membrane proteins by combining MS-based footprinting, specifically fast photochemical oxidation of proteins (FPOP), and lipid Nanodiscs, which more similar to the native membrane environment than are the widely used detergent micelles. We describe here results that show a protein's outer membrane regions are

more heavily footprinted by OH radicals whereas the regions spanning the lipid bilayer remain inert to the labeling. Nanodiscs generally exhibit more protection of membrane proteins compared to detergent micelles and also less shielding to those protein residues that exist outside the membrane. The combination of the two technologies, FPOP and Nanodiscs, is a feasible approach to map extra-membrane protein surfaces, even at the amino-acid level, and illuminate intrinsic membrane protein topology.

3.2 Introduction

Membrane proteins are involved in crucial cellular functions, including photosynthesis[1], respiration[113] and signal transduction[114]. They represent ~30% of open reading frames[115] of many genomes, and an increasing number of them are important drug targets[116]. Membrane proteins perform their functions via interaction with other molecules or with themselves to undergo conformational changes important in signaling, for example. Membrane proteins are highly flexible and dynamic, enabling them to perform different tasks with high efficiency but making structure determination difficult. Membrane protein structures are notoriously difficult to resolve compared to water-soluble proteins[117]. Because the membrane proteins are hydrophobic, they are less stable in water than in a membrane, and detergent are needed for solubilizing and stabilizing them. Detergents, however, can affect protein conformation and hinder protein interaction with other molecules. Compared to the large number of soluble proteins, high-resolution structures of membrane proteins are available only for a small fraction of them[118].

Various detergents are used to extract membrane proteins from their native lipid bilayer and to solubilize, stabilize and enclose them in micelles[119]. Unfortunately, detergents are not an ideal mimic of cellular environment. The different micelle sizes and curvature restrictions compromise

protein stability and, in some cases, proper protein functioning[120]. In addition, an excess micellar phase may interfere with the interaction with other molecules and interfere with analytical methods. To overcome these problems, reconstitution of membrane protein in various membranes including monolayers, bicelles, and liposomes have been pursued.

One approach that provides a better mimic of a native environment and controllable stoichiometry of target membrane protein is the lipid-protein Nanodisc[121]. Here, two membrane scaffold proteins (MSP) form a double belt to enclose a lipid bilayer and form a water-soluble "disc" into which target membrane proteins can be incorporated. Under self-assembly conditions, the oligomeric state of a target protein and the nature of the lipids included in the bilayer can be controlled, allowing a membrane protein to be probed from both the cytoplasmic and the periplasmic sides of the "membrane". Thus, a Nanodisc provides a simple and robust means for rendering target membrane proteins in aqueous buffer while keeping the protein in a native-like bilayer environment[122].

Membrane topology can be viewed as "an important halfway house between the amino-acid sequence and the fully folded three-dimensional structure"[123]. Individual transmembrane helices can insert into a lipid bilayer in different ways, and because the proteins are dynamic, they can change conformation and position. Various mass spectrometry (MS)-based labeling methods are now being widely adopted to study those issues; one of them, cross-linking, has become effective in probing protein-protein interactions[43]. As a complementary approach, MS-based footprinting, which includes hydrogen/deuterium exchange (HX) and •OH and other radicals, serve to "label" most of the amino acids and illuminate protein-protein, and protein-ligand interactions.[124] HX is widely used for soluble proteins, and the extent of exchange reports on H bonding and solvent accessibility of the protein backbone[32, 125]. HX can also be used for membrane proteins in the

presence of detergent micelles provided it provides fast isolation, digestion efficiency, and solubility[69, 126-127]. Footprinting, done with the FPOP platform, can label amino acids with OH radicals from photolysing hydrogen peroxide. The amino-acid reactivity with hydroxyl radicals is broad-based although the reactivity with amino acid side chains can vary by three orders of magnitude[128-129]. The dominant product is a +16 adduct, but there are other pathways and products. FPOP probes solvent accessibility of different regions of proteins in a fraction of a second and at the amino acid level[130]. The labeling is carefully controlled so that every fraction or plug of a flowing protein buffer is labeled only once. Compared to the more widely used HDX, the irreversible labeling provides flexibility in digestion as there is no concern for back exchange and good potential for general membrane protein studies, as lipid removal prior to MS analysis is relatively easily accomplished.

To our knowledge, a few different methodologies regarding oxidatively labelling of membrane proteins have been reported previously. Sze et al.[131] adopted a Fenton reaction to oxidize the outer membrane porins and revealed the voltage gating of porin OmpF *in vivo*. Konermann et al.[68] carried out the first FPOP oxidative labeling of a membrane protein, bacteriorhodopsin, in a natural lipid bilayer environment. They only detected oxidative modification of methionines located in solvent-accessible loops that are highly oxidized compared to those located in the transmembrane regions, probably because the protein is rich in Met. A subsequent study revealed the conformational change of denatured bacteriorhodopsin in SDS compared to the native state[132].

More recently, X-ray radiolytic footprinting was also used to study structural water and conformational change of membrane proteins[70, 133]. For example, rhodopsin was dissolved in detergent and radiolysis-produced OH radicals labeled both solvent-accessible and solvent-

inaccessible regions. The labeling of solvent-inaccessible regions may be due to tightly bound structural water molecules that are ionized by the radiation, produce •OH, and label nearby residues[134], "elucidating *in vivo* structural dynamics in integral membrane protein by hydroxyl".

In this study, we applied MS-based FPOP footprinting to a membrane protein complex in a near-native environment. As a model protein, we used the light-harvesting complex 2 (LH2) from *Rhodobacter (Rb.) sphaeroides*, an intrinsic membrane protein with ~18 transmembrane helices[135]. We used Nanodiscs to "house" the LH2 in aqueous buffer prior to and during labeling by hydroxyl radicals. The integrated workflow is shown in figure 3.1. We compared the solvent accessibility of LH2 from FPOP with that in detergent micelles and identified labeling at the residue level. Our results show that Nanodiscs generally provide a better protection of the transmembrane core region of protein and less shielding for the outer membrane region.

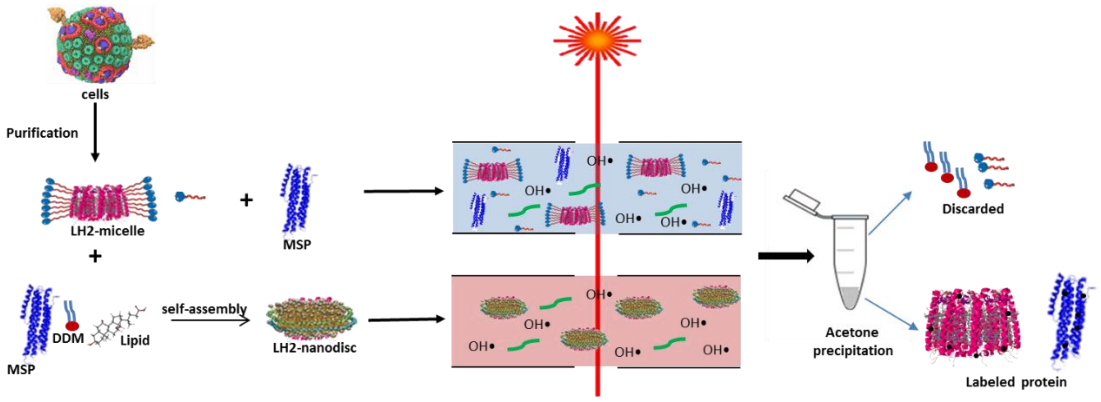


Figure 3.1. The integrated Nanodisc-FPOP MS workflow. Cells are shown by Lamellar chromatophores in *Rhodospirillum photometricum*[136]. LH2 is shown by PDB 1NKZ, MSP is shown by PDB 2A01.

3.3 Materials and Methods

3.3.1 LH2 Preparation.

Rb. sphaeroides wild-type strain ATCC 2.4.1 was grown photosynthetically at room temperature (RT) in 1 L bottles, and LH2 was isolated as previously described[64]. Briefly, lauryldimethylamine N-oxide (LDAO, 1.5%) was added to the resuspended membrane pellets with stirring for an hour at RT. After centrifugation, the supernatant was loaded onto an anion-exchange column (QSHP resin, GE Healthcare, Uppsala, Sweden), and a linear gradient elution was performed. The fractions containing LH2 were loaded onto HiLoad Superdex 200 prep grade column (GE Healthcare). The protein-to-pigment ratio was measured from the absorbances at 850 nm (Q_y absorption band of bacteriochlorophyll *a*) and at 280 nm (protein absorption band); fraction showing a ratio greater than 3.0 were collected.

3.3.2 MSP Preparation.

A pMSP1E3D1 plasmid was purchased from Addgene (Cambridge, MA). Purification was performed as previously described.[137] Briefly, after breaking the cells by ultra-sonication, the supernatant was loaded onto Ni-NTA agarose column (QIAGEN, Valencia, CA). After extensive washing and elution, the purity of MSP was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis.

3.3.3 Self-assembly of Nanodiscs.

1,2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC) dissolved in chloroform (Avanti, Alabaster, Alabama) as dispensed into a disposable glass tube and allowed to dry in a fume hood. The tube was placed in a vacuum desiccator overnight to remove residual solvent. A buffer

containing cholic acid (Affymetrix, Santa Clara, CA) with a concentration twice that of the lipid was added to the lipid film followed by sonication and gentle heating to solubilize the lipid. The MSP and LH2 were added to the lipid buffer in a 2:1:120 ratio and incubated at room RT for 30 min. After adding amberlite XAD-2 (Sigma Aldrich, St. Louis, MO), the mixture was gently shaken for 2 h at RT until the self-assembly process was complete. Finally, the Nanodisc preparation was purified by HPLC with a Superdex 200 prep grade column (GE Healthcare).

3.3.4 Steady-state absorption spectroscopy.

Steady-state absorption spectra of the LH2 were recorded at RT using a Perkin-Elmer Lambda 950 UV-Vis spectrophotometer. Prior to measurements the LH2-Nanodisc sample was dissolved in saline buffer (PBS) containing 10 mM phosphate, 140 mM NaCl and 2.3 mM KCl (Sigma Aldrich, St. Louis, MO). For comparison, a solution of free LH2s was dissolved in the same buffer containing 0.02% n-Dodecyl-β-D-Maltopyranoside (DDM).

3.3.5 Time-resolved fluorescence spectroscopy.

B850 BChl *a* fluorescence decay dynamics were measured using time-correlated single photon counting (TCSPC) setup based on a stand-alone Simple-Tau 130 system (Becker&Hickl, Germany), equipped with a PMC-100-20 detector (GaAs version with full width at half maximum of instrument response function <200 ps), PHD-400 high speed Si pin photodiode (as triggering module), motorized Oriel Cornerstone 130 1/8 m monochromator with ruled 1200 l/mm grating blazed at 750 nm (Newport, USA), and a manual filter wheel. Excitation pulses at 590 nm (Q_x band of BChl *a*) were delivered by ultrafast optical parametric oscillator Inspire100 (Spectra-Physics, USA) pumped with Mai-Tai, an ultrafast Ti:Sapphire laser (Spectra-Physics,

USA), generating ~90 fs laser pulses at 820 nm with a frequency of 80 MHz. The final frequency of the excitation beam was set to 8 MHz (125 ns between excitations) by using a 3980 Pulse Selector from Spectra-Physics. To avoid polarization effects, the excitation beam was depolarized using an achromatic DPU-25 depolarizer (Thorlab, USA). The beam was focused to ~1 mm circular spot at the sample; to assure annihilation-free conditions, the beam intensity was set to ~10^{10} photons/cm^2 per pulse. The signal was collected at right angle to the excitation. The sample absorbances at 850 nm were adjusted to ~0.1.

3.3.6 Dynamic light scattering.

The sample was filtered (0.2 μm filter) prior to analysis by dynamic light scattering using a Malvern Zetasizer Nano S/ZS instrument (Worcestershire, UK.) to estimate the diameter of the Nanodisc. The diameter of empty-Nanodisc was also measured as a control at 25 °C. Data were fitted using the Zetasizer software (Worcestershire, UK) to estimate the diameter of the particles.

3.3.7 Footprinting of Nanodisc-LH2 and detergent micelle-embedded LH2.

The LH2 concentration was estimated by using the molar absorptivity of B850 bacteriochlorophyll *a*[138], and the MSP concentration was calculated according to the predicted molar absorptivity at 280 nm and used for the following experiments[139]. The LH2-Nanodisc sample was dissolved in PBS buffer to the specifications: 2 μM of Nanodisc-LH2, 350 μM of histidine, 5μM of [leu5]-enkephalin (reporter peptide), and 20 mM of H_2O_2. The LH2 dissolved in PBS buffer containing 0.02% DDM, 2μM of LH2, 2μM of MSP, 350 μM histidine, 5μM of [leu5]-enkephalin (reporter peptide), and 20 mM H_2O_2. To minimize any pre-oxidation by H_2O_2, it was added immediately before the laser irradiation. The FPOP experiment was performed as

previously described[140]. The energy of the KrF excimer laser (GAM Laser Inc., Orlando, FL) was adjusted to 22.3 mJ, and sample flow rate was 22 μL/min to ensure a 20% exclusion volume. After laser-induced labeling, each sample was collected in a vial containing 10 mM catalase and 20 mM Met to reduce leftover H_2O_2. Control samples for both Nanodisc- and detergent-LH2 were handled in the same manner without laser irradiation. All experiments were performed in triplicate. For each collection, the buffer was divided into two portions. Formic acid (1%) was added to one portion prior to desalting with a Sep pak C18 (Waters Inco., Milford, MA). The other portion was precipitated with acetone and dissolved in buffer containing 100 mM Tris, 1 mM $CaCl_2$, and 0.02% RapiGest SF. Digestion was at 37 °C for 1 h with chymotrypsin (Promega Corporation, Madison, WI) and subsequently quenched by FA (1%).

3.3.8 MS analysis

Peptide mixtures were trapped by a guard column (Acclaim PepMap100, 100 μm × 2 cm, C18, 5 μm, 100 Å; Thermo Fisher Scientific, Breda, Netherlands) and then fractionated on a custom-packed Magic C18 reversed-phase column. The MS analysis was with a Thermo Scientific™ Q Exactive™ hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen Germany). Peptides were eluted with a 85 min, 250 nL/min gradient coupled to the nanospray source. A 50 min, 250 nL/min gradient was adopted for the reporter-peptide analysis. The default charge state was 2, and the scan range was from *m/z* 380-1500. Mass spectra were obtained at high mass resolving power (70,000, FWHM at *m/z* 200) and the top 15 most abundant ions corresponding to eluting peptides per scan were submitted to CID in the ion trap, with charge-state rejection of unassigned and >8 ions enabled. Precursor ions were added to a dynamic exclusion list for 8 s to ensure good sampling of each elution peak.

3.3.9 Data analysis

The oxidation extent was calculated with the equation shown below. Briefly, the strategy relied on an integrated peak area relation between of un-oxidized and oxidized peptides. The integration of peak area was performed by using Xcalibur™ Software. (Thermo Fisher Scientific) For consistency, only +16 modifications were considered in this study as the abundance of di-oxidized products was either low or below the detection limit of the instrument.

$$\textbf{Normalization factor} = \frac{\textit{Oxidation level of reporter peptide in LH2-Nanodisc}}{\textit{Oxidation level of reporter peptide in LH2-micelle}}$$

$$\textbf{Oxidation level of LH2 peptides in micelle} = \frac{\Sigma I_{ox}}{\Sigma I_{ox} + \Sigma I} \times \textit{Normalization factor}$$

$$\textbf{Oxidation level of LH2 peptides in Nanodisc/reporter peptide} = \frac{\Sigma I_{ox}}{\Sigma I_{ox} + \Sigma I}$$

3.3.10 Sequence alignment, topology prediction and homology modeling

Sequence alignment of LH2 from different purple bacteria was performed by an online web server[141]. The TOPCONS web server was adopted for prediction of LH2 topology[142]. Homology models of LH2 from *Rb. sphaeroides* were generated, as previously described[64]. The heterodimer models (α and β) were based on PDB 1NKZ and processed by Pymol[143] (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.).

3.4 Results and discussion

3.4.1 Characterization of the Nanodisc containing LH2

The model membrane protein in this study is LH2, a protein complex belonging to the photosynthetic antenna family whose primary function is to harvest light and transfer absorbed energy. It is important to characterize the protein in Nanodiscs and detergent prior to footprinting to insure that our comparisons are valid and that the Nanodisc indeed contains the intact complex. An advantage of using LH2 is the Nanodisc-protein can be convincingly characterized by absorption and fluorescence spectroscopies. Ideally, LH2 preparations in Nanodiscs and detergent media should show identical absorption spectra with characteristic well-developed and resolved electronic absorption bands for bound pigments (i.e., B800 and B850 bands at 800 and 850 nm associated with bacteriochlorophyll *a*, and carotenoid (spheroidene) absorption band between 480 and 515 nm). This essential analysis shows that the pigment environments are not fundamentally altered in the two preparations (Figure 3.2).

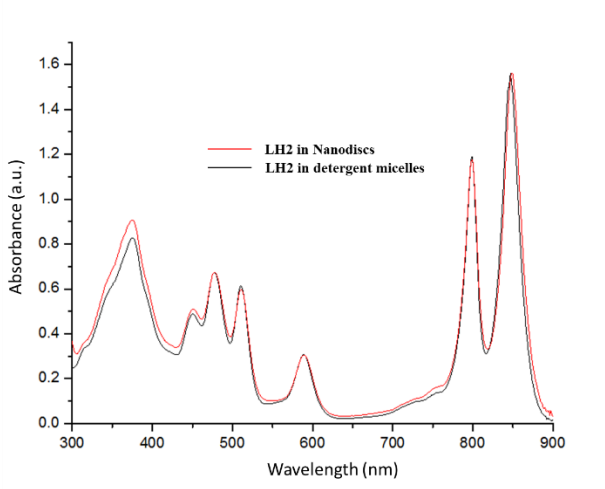


Figure 3.2. Steady-state absorption spectrum of LH2 in detergent micelles and Nanodiscs.

In addition, the intact LH2 complex can be probed by time-resolved fluorescence using the B850 emission. The B850 fluorescence decay lifetime of this LH2 complex is typically ~ 1 ns[144], and any significant variation of this value will indicate perturbation of the B850

bacteriochlorophyll *a* array. Furthermore, different excitonic coupling in the B850 exciton will alter the rate of radiative decay. Alternatively, a significant reduction of fluorescence lifetime of the Nanodisc-LH2 would indicate that the Nanodisc bundles more than one LH2 complex and allows formation of oligomeric LH2 structures. As shown in Figure 3.3, however, the B850 fluorescence lifetimes are essentially the same for both preparations and fit the expected time range for monomeric and not structurally deficient or altered LH2. To add certainty, we also used the A850/A280 ratio as a marker of the LH2 purity. The ratio for LH2 in detergent (3.05) should be higher than in in the Nanodiscs (2.34) because the MSP also contributes to the absorption at 280 nm and lowers the ratio. Furthermore, a calculation based on molar absorptivity indicates that the "Nanodisc" contains one LH2 complex and a certain percentage of "empty" Nanodisc is also present.

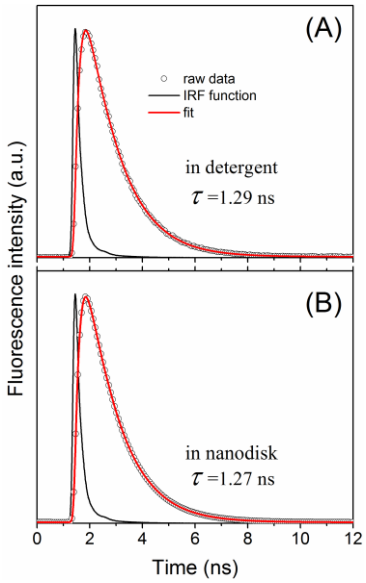


Figure 3.3: Time-resolved fluorescence of LH2. Fluorescence decay of (A) detergent and (B) Nanodisc LH2 preps. Fluorescence was monitored at maximum of the B850 emission band. The fits (red lines) consist of mono-exponential decay function convoluted by instrument response function (IRF). Fluorescence decay lifetimes obtained from data fitting are indicated in graphs.

To complete the characterization of the preparation, we used dynamic light scattering (DLS) to characterize rapidly the particle size of the Nanodisc[145]. The homogeneity of empty- and LH2-embeded Nanodiscs is revealed by the size distribution (Figure 3.4). The result (~10 nm) is consistent with a previous report[146] and demonstrates that the nanoidsc-LH2 has a slightly larger dimension compared to the empty-Nanodisc. This perturbation is probably caused by expansion of the disc induced by the LH2 residing in its middle, and it is consistent with a previous report that shows that addition of target membrane protein into Nanodiscs slightly enlarges their dimensions[147].

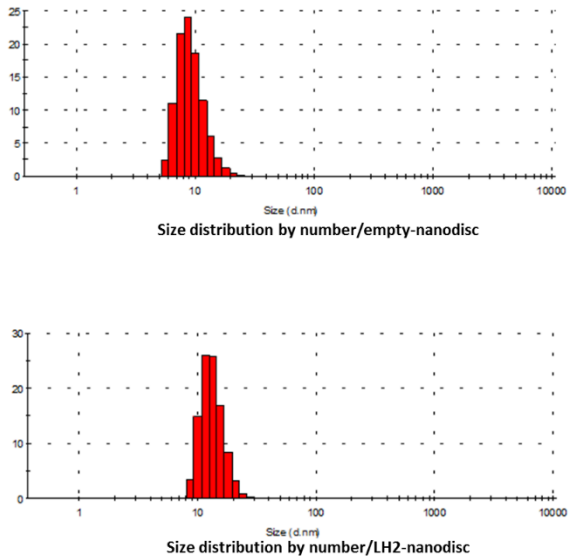


Figure 3.4. Dynamic light scattering of empty-Nanodisc and LH2-Nanodisc

3.4.2 Application of FPOP to membrane proteins

Fuller advantage of Nanodiscs as a convenient and powerful vehicle for membrane proteins, may be achieved by combining the technology with protein footprinting. Now that a variety of membrane proteins have been successfully reconstituted into Nanodiscs[148-151] and characterized

biophysically, we can seek higher resolution structural information to complement data on the size and activity of target membrane proteins. High motivation exists for this goal because no crystal structure exists for many membrane proteins including the LH2 of interest here.

"Footprinting" can be done by NMR and HDX. Stefan et al.[152] showed, for example, that membrane-protein bond orientations in Nanodiscs can be obtained by measuring residual dipolar couplings with the outer membrane protein. Jorgenson, Rand, Engen and coworkers probed the conformational analysis of γ-glutamyl carboxylase by HDX, demonstrating the applicability of HDX to Nanodiscs[153]. Subsequently, Jorgenson and Engen[154] investigated binding of γ-glutamyl carboxylase to a propeptide, employing good separation and HDX. Recently, Adkins and coworkers studied the membrane interactions, ligand-dependent dynamics, and stability of cytochrome P450 in Nanodiscs by HDX[155]. In the HDX studies, it is required to disassemble rapidly the Nanodisc and remove the excess lipid from solution prior to MS analysis while minimizing back exchange, which can be a nagging problem. We are advocating irreversible labeling with the hydroxyl radical to permit, prior to the MS analysis, easier lipid removal such as acetone precipitation, chloroform/methanol extraction, and other forms of offline desalting by use of reversed-phase cartridges. Indeed, we adopted acetone precipitation, and we could remove most of the lipid after processing. Moreover, we could obtain complete coverage (100%) in the digestion of the α and β subunits in LH2 (Figure 3.5) and extend the analysis from the peptide to the amino-acid residue level in some cases. Our results show that the regions that are likely to be in the cytoplasmic or periplasmic space undergo a higher extent of oxidative labeling compared to the regions of the protein deeply embedded in the Nanodisc. The MSP proteins that wrap around the lipid bilayer also get labeled to different extents for different regions.

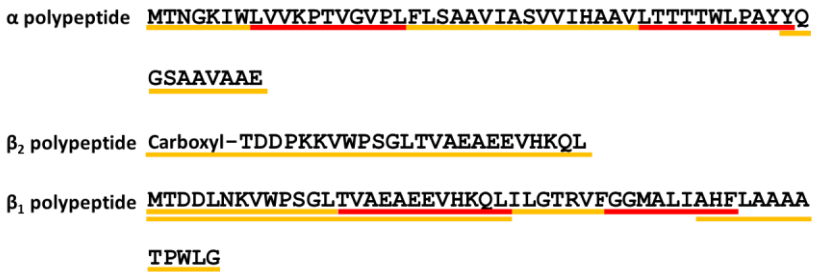


Figure 3.5. Coverage map of LH2.

3.4.3 Membrane protein in detergent micelle vs. Nanodisc

Because both lipids and detergents are prone to oxidative modifications by •OH, we measured the hydroxyl radical reactivity in the two environments normalized to the labeling yield of a reporter peptide (i.e., the five amino-acid leu enkephalin)[156-157]. In this way, we can compensate for any differences in protein reactivity introduced by changing from detergent to Nanodisc. We found that the oxidation level of the reporter peptide in Nanodiscs is 1.67 times greater than with the DDM micelles under the same experimental conditions (Figure 3.6). Due to the lack of high resolution structure, we then used homology modeling with the known structure of LH2 from *Rh. acidophilus* [64] and obtained a result with the high certainty (99.9%) to assist the discussion of the results.

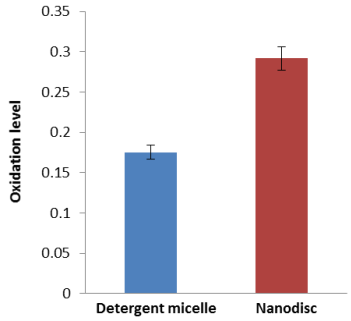


Figure 3.6. Oxidation level of [Leu5]-Enkephalin YGGFL in detergent micelle/Nanodisc environment.

Although large amounts of lipids are present in solution when footprinting a protein in Nanodiscs, their alkyl tails are embedded and not highly available for reaction with the free radicals. Coarse-grain molecular dynamics simulations reveal that the lipids in the Nanodisc have higher acyl tail order than lipids in a lamellar bilayer phase[158]. The detergent, however, exists as a monomer at low concentration, and when its concentration is increased above the critical micelle concentration (CMC), it self-associates to form non-covalent micelles. Although a spherical detergent micelle is often viewed as uniformly packed, they are not, and the octyl glucoside micelles contain a distribution of surfactant molecules. Instead of a static shape as usually assumed, those different size micelles fluctuate between spherical and near-ellipsoidal shapes[159]. And not all hydrophobic tails are buried or point toward the center of the micelle; rather the micelle surface is rough and heterogeneous. Furthermore, the state of detergent micelles and of detergent-protein micelles is relatively dynamic, undergoing rapid exchange of detergent and solvent[160]. Thus, a detergent is more likely to quench •OH than a Nanodisc.

Considering now the protein complex in a Nanodisc, we find, as expected, that the solvent-exposed terminal regions of LH2 undergo greater oxidative modification than the transmembrane regions (Figure 3.7). Although this is the case for both the detergent and the Nanodisc, we expect that the hydrophobic regions of a membrane protein will be protected less in a detergent micelle than in Nanodiscs, and this is seen for all regions of LH2 where the oxidation level is larger in the presence of detergent micelles than of Nanodiscs. Nevertheless, the overall reactivity trends are similar, suggesting that detergent micelles do provide a similar environment to the lipid bilayer. A previous study shows that detergent molecules in a globular micelle can exchange over hundreds of nanoseconds with detergent molecules in a micelle bound to a protein[161]. For example, the dynamic fluctuations of OmpA protein are 1.5 times greater in the micellar environment than in the lipid bilayer, and this increased overall mobility may be attributed to the increased diffusion properties and reduced packing of detergent molecules[162]. The differences we observed are in accord with those results, suggesting that membrane proteins in detergent micelles have more flexibility and solvent accessibility compared to in a lipid Nanodisc. For the methionine residues located in the transmembrane region of LH2, the Nanodisc affords even more protection than detergent micelles. This region is embedded in the lipid bilayer of a Nanodisc but only closely associated with hydrophobic tails in the detergent (Figure 3.8).

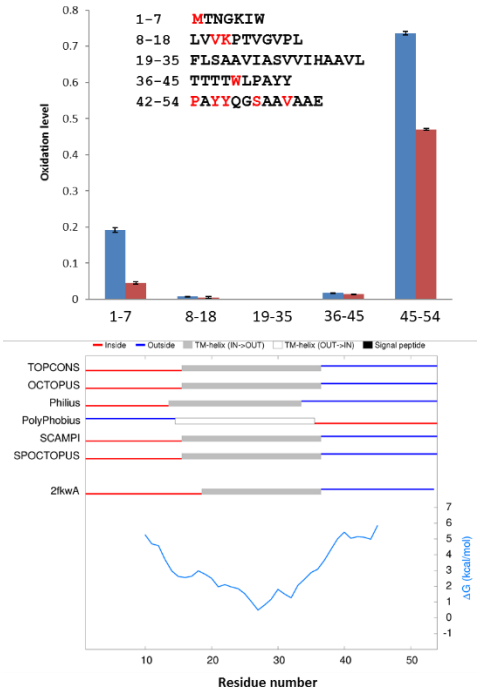


Figure 3.7a. Oxidation level of peptides from the α subunit. Peptides in detergent are shown in blue, and in Nanodiscs are shown in red. The lower panel shows the consensus prediction of membrane protein topology by TOPCONS web server.

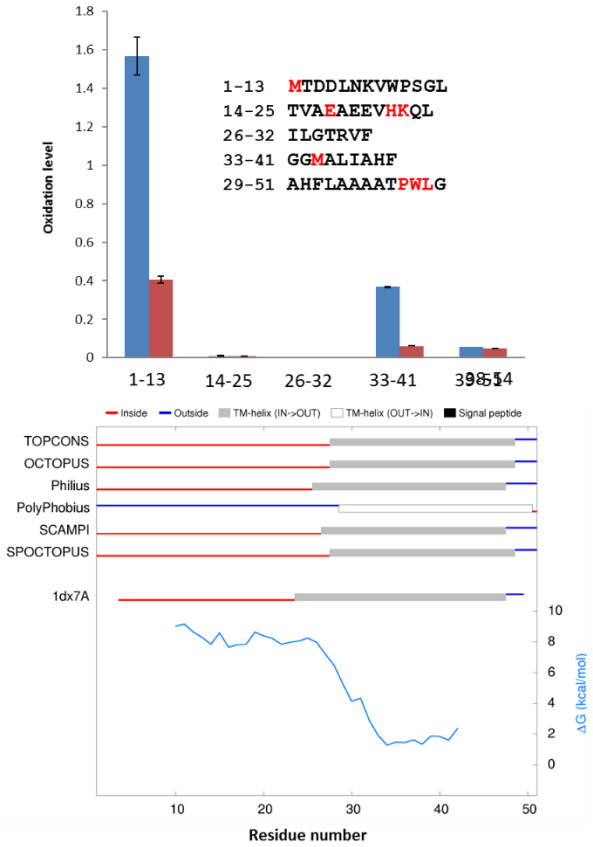


Figure 3.7b. Oxidation level of peptides from β1 subunit. Peptides in detergent are shown in blue, and in Nanodiscs are shown in red. The lower panel shows the consensus prediction of membrane protein topology by TOPCONS web server.

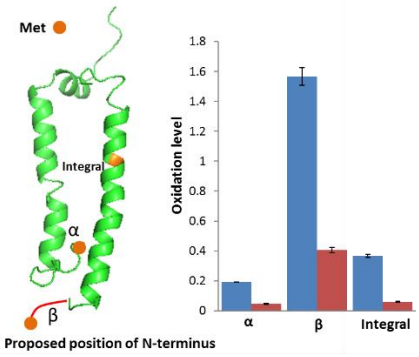


Figure 3.8. Oxidation level of Met in detergent micelle/Nanodisc environment. Met were labeled in orange and the proposed position of Met are shown as orange dots. Proposed N-terminal of β subunit (MTDDLNKVWPSG) is shown as red line. Mets in detergent are shown in blue, Nanodisc are shown in red.

3.4.4 Methionine as a marker in membrane protein labeling by FPOP

A previous study of bacteriorhodopsin reports extensive oxidation of methionines located in its solvent-accessible loops[132]. For LH2 in our study, Met is also relatively reactive, but the protein reactivity occurs on many residues besides Met. For each heterodimer (α and β) composing the ring of LH2, there are three Mets. One is in the transmembrane region, and the other two are near the N-terminus (Figure 3.8). LH2 in a lipid bilayer exhibits higher protection for those regions compared to in a detergent micelle. According to the homology modeling, the N-terminal Met of the α subunit is not protruding into the outer region of the membrane but instead is bent toward the inner region of a lipid bilayer[64] where it coordinates the central Mg^{2+} ion of a nearby bacteriochlorophyll *a*, and preserves the planar conformation of the pigment molecule.

The axial coordination of a central Mg^{2+} ion is crucial for all the photosynthetic chlorophyll-proteins, in terms of both structure and function[163]. A carboxyl modified Met1 of the α subunit from *Rhodoblastus (Rh.) acidophilus* is ligated to Mg^{2+} of B800[80] whereas for *Phaeospirillum (Ph.)molischianum* [81], the corresponding ligand is Asp6. For both these structures, the N-terminal regions of the α subunits exist as a loop structure but are closely associated with the transmembrane regions owing to its coordination with B800. In our homology model, the N-terminal of α subunit of the LH2 from *Rb. sphaeroides* is also closely associated with the transmembrane region (Figure 3.8). The Met located in the transmembrane region of the protein shows a similar oxidative modification to the one located on the N-terminal of the α subunit. The first few amino acids on the N-terminal end of the β subunit are not covered in the homology model (Figure 3.9). We suggest this Met is pushed out of the membrane and undergoes

relatively high oxidative modification. As a reference, the methionine-containing peptides from the MSP of the Nanodisc show a 0.15-0.60 range of oxidation. We conclude that the extent of oxidative modification of the highly reactive Met is a good marker for the topology of transmembrane proteins on the FPOP platform[132].

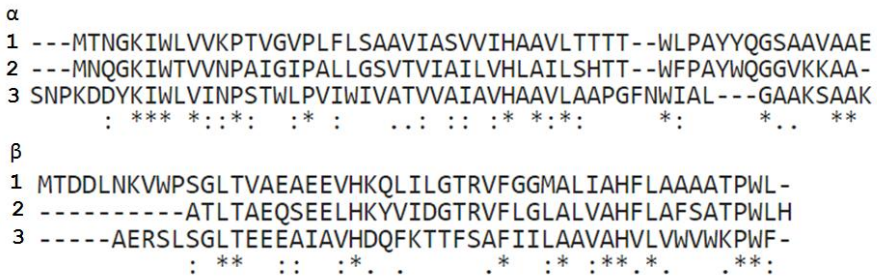


Figure 3.9. Sequence alignments of LH2. α and β subunits from different LH2 complexes: (1), *Rb. sphaeroides* (2), *Rh. acidophilus* (3), *Ph. molischianum*

3.4.5 Locating the membrane protein in the lipid bilayer

Membrane proteins are closely associated with lipid bilayers and integral transmembrane domains are more deeply embedded than exterior regions. It is intriguing to probe the interaction of lipid bilayer with different domains of membrane proteins. The lipid hydrophobic tails are closely associated with the LH2 hydrophobic transmembrane domains, and their lengths determine the thickness of the membrane core (typically ~ 3 nm). The thickness of the polar head of lipids on each side is ~ 1.5 nm[164-165]. Siuda et al.[166] observed that the Nanodisc thickness is smaller near the MSP double belt, owing to the perturbation from boundary lipids. The average thickness of MSP1E3, which is the MSP we used for the LH2-Nanodisc, obtained by applying small-angle X-ray scattering (SAXS), is 4.6 nm. Previous studies showed the Stokes diameter of MSP1E3 is 12.1 nm, whereas SAXS gives a value of 12.8 nm[121]. LH2, mapped by atomic force microscopy, shows average center-to-center distances between complexes within

the dimer as 7.7 nm.[167] The above values show that although the LH2 has a relatively large transmembrane domain, it could be positioned in the middle of the MSP1E3D1 Nanodisc. Those values also strongly suggest that only a single LH2 complex can be incorporated into the Nanodisc. This fixed stoichiometry of LH2 in the disc is also carefully controlled by using an appropriate ratio of LH2 and MSP in the assembly process.

Although an increasing number of high-resolution crystal structures of membrane proteins are published every year, it is necessary to picture the topology of membrane protein sitting in the dynamic membrane bilayer. To do this, we adopted the TOPCONS web server[168] for a consensus prediction of the structural and functional features, membrane-inside and outside (i and o, respectively). In addition, we used a biological hydrophobicity scale to predict the free energy of membrane insertion centered on each position in the sequence[142]. Experimental data, however, are needed to confirm the topology and conformation of membrane proteins. Oxidative labeling shows that the terminal ends of the two transmembrane helices are more heavily solvent-accessible than are the integral regions. The oxidative modification levels are generally in accord with the free-energy trends. For transmembrane regions, little or no oxidative modification occurs. Further, LH2 in the Nanodisc shows a lower level of oxidation compared to the one in detergent, suggesting better protection of LH2 in a lipid bilayer (Figure 3.7).

The Met in the N-terminal end of the α subunit, as discussed above, is pointed inward to coordinate a pigment molecule. Thus, the oxidation level of this Met is not high compared to the other Met residues in the protein assembly (Figure 3.8). The C-terminal end of the α subunit is highly modified even though no highly reactive methionine is present. This region, as modeled

with 99.9% confidence to the crystal structure of the LH2 from *Rh. acidophilus*, and it shows a helix-loop structure sticking out of the "ball" structure (Figure 3.7a). This region has an extended conformation that passes between the β-chains of the neighboring heterodimers, and the large occupancy volume indicates high flexibility[80]. The extracted ion chromatogram (EIC) of the oxidized form (+16) shows one major and two minor peaks (Figure 3.10). The product-ion spectra (see Figure 3.11 a-e) reveal that the first two minor peaks represent the peptide with oxidatively modified Tyr, Ser and Val and the major peak represents the peptide with an oxidized Pro. Because the elution time for peptides containing oxidized Tyr, Ser or Val are overlapping, we cannot differentiate the modification extents of those residues in detergent or in the Nanodisc. The rate constants of Pro, Ser and Val with •OH are of the same order of magnitude but an order of magnitude lower than that for Tyr[128]. The high oxidative extent of Pro and low level of Tyr should be related to the protein conformation of this region. The 2.0 Å crystal structure of the LH2 from the template shows that the glucoside head groups of the rhodopsin glucoside carotenoid molecule (RG1) are located at the cytoplasmic surface whereas the second carotenoid (RG2) is at the periplasmic surface[80] (Figure 3.12). Raman scattering, however, shows no bands that could be attributed to RG2[169]. Later, the authors claimed that this RG2 site is actually occupied by a mix of BOG and LDAO molecules, owing to incomplete detergent exchange. This "RG2" was located adjacent to the Tyr residues, as shown in Figure 6. Our results also suggest that this Tyr site is relatively solvent-inaccessible owing to its association with detergents/lipids. The crystal structure shows Pro is facing inward and the other residues are either shielded by the detergent/lipid molecules (Tyr-Tyr in PDB 1NKZ, Tyr-Tyr in homology model) or adjacent to the C-terminal end of the β subunit (Gly in PDB 1NKZ and Ser in the homology model) with the exception of valine (Figure 3.12a-c). We propose that the

association of transmembrane helices with detergents/lipids in the “hole” or center of the LH2 ring complex is different than outside, similar to the lipids adjoining the MSP in the Nanodisc where the thickness of the disc is smaller. This may be a result of distorted packing of the lipids to minimize any hydrophobic mismatch at the protein-lipid interface[166, 170]. It is interesting that the Pro exhibits lower oxidative modification in Nanodiscs, which is consistent with the behavior of other peptides/residues in LH2. Other residues (Tyr, Ser and Val) in this peptide, however, exhibit slightly higher levels of oxidation in Nanodiscs than in detergent, suggesting that the lipids provide better protection of proline compared to the detergent micelle, while the other residues are slightly more exposed in the lipid bilayer (Figure 3.12d). The detergent micelle might hinder the solvent accessibility of a number of residues on the surface of the C-terminal end of the α subunit, while relatively more regularly packed lipids in the Nanodiscs exhibit lesser blocking.

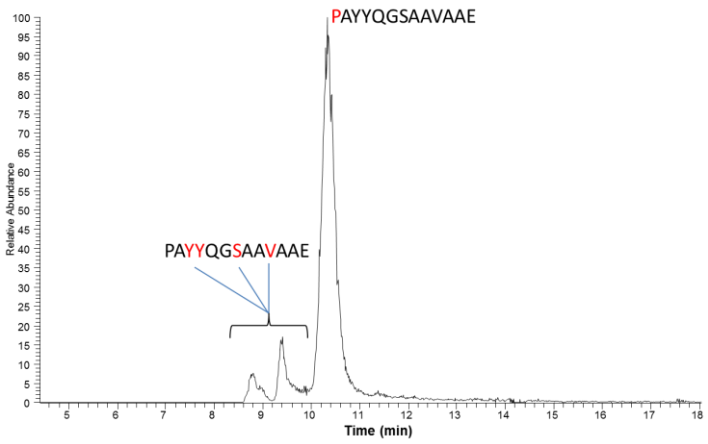
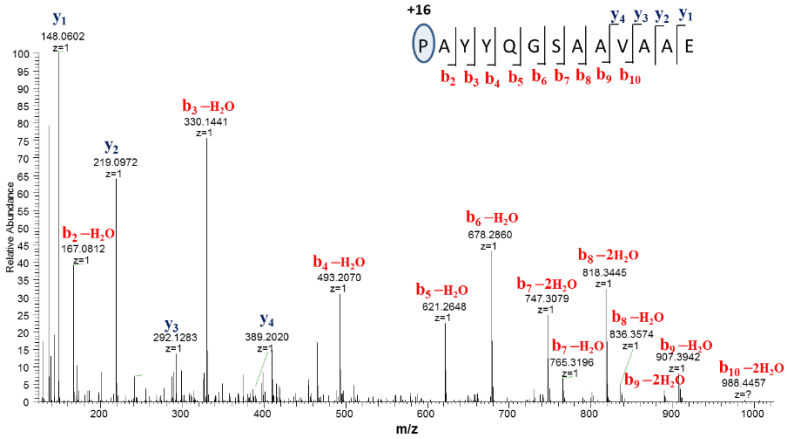
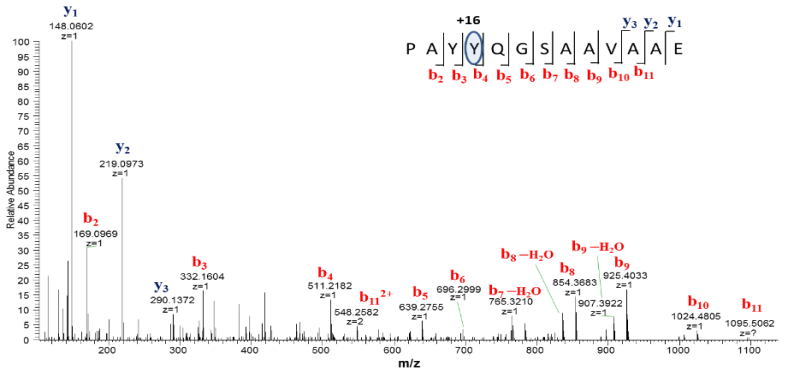


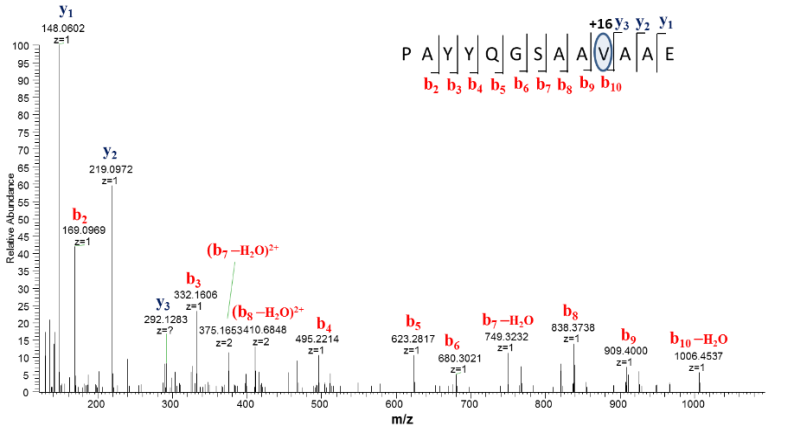
Figure 3.10. EIC of m/z = 649.3074 (PAYYQGSAAVAAE). The oxidized residues are marked in red.


+16
P A Y Y Q G S A A V A A E
y1 148.0602 z=1
y2 219.0972 z=1
y3 292.1283 z=1
y4 389.2020 z=1
b3 −H2O 330.1441 z=1
b2 −H2O 167.0812 z=1
b4 −H2O 493.2070 z=1
b5 −H2O 621.2648 z=1
b6 −H2O 678.2860 z=1
b7 −2H2O 747.3079 z=1
b8 −2H2O 818.3445 z=1
b8 −H2O 836.3574 z=1
b7 −H2O 765.3196 z=1
b9 −H2O 907.3942 z=1
b9 −2H2O
b10 −2H2O 988.4457 z=?
Relative Abundance
m/z

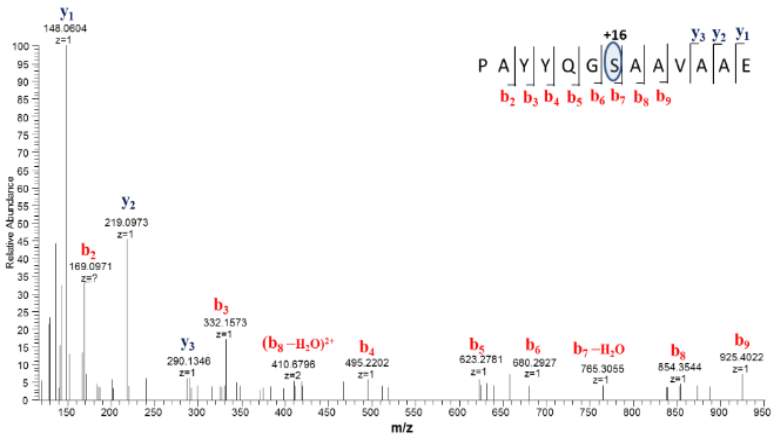
a


+16
P A Y Y Q G S A A V A A E
y1 148.0602 z=1
y2 219.0973 z=1
b2 169.0969 z=1
y3 290.1372 z=1
b3 332.1604 z=1
b4 511.2182 z=1
b11 2+ 548.2582 z=2
b5 639.2755 z=1
b6 696.2999 z=1
b7 −H2O 765.3210 z=1
b8 −H2O
b8 854.3683 z=1
b9 −H2O
b9 925.4033 z=1
b10 1024.4805 z=1
b11 1095.5062 z=?
Relative Abundance
m/z

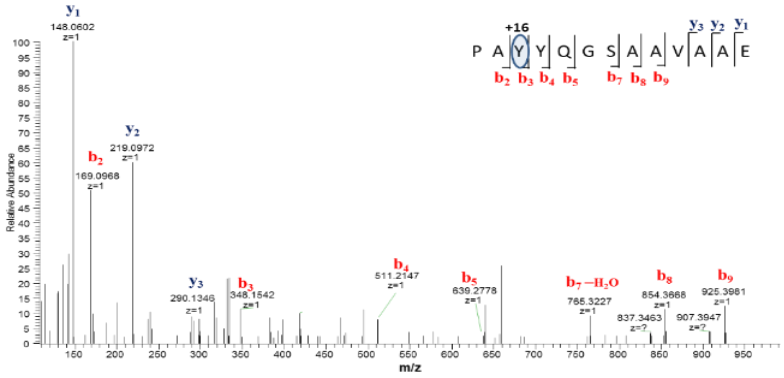
b


+16
P A Y Y Q G S A A V A A E
y1 148.0602 z=1
y2 219.0972 z=1
b2 169.0969 z=1
(b7 −H2O)2+
b3 332.1606 z=1
y3 292.1283 z=?
(b8 −H2O)2+ 375.1653 z=2
410.6848 z=2
b4 495.2214 z=1
b5 623.2817 z=1
b6 680.3021 z=1
b7 −H2O 749.3232 z=1
b8 838.3738 z=1
b9 909.4000 z=1
b10 −H2O 1006.4537 z=1
Relative Abundance
m/z

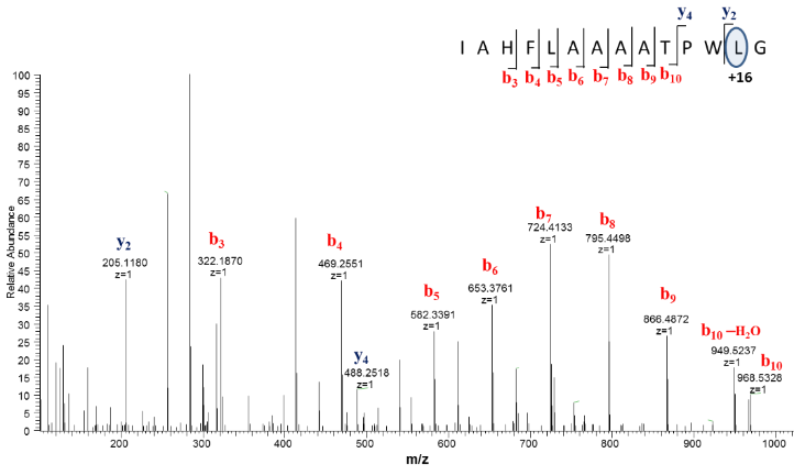
c


P A Y Y Q G S A A V A A E
+16
Relative Abundance
m/z

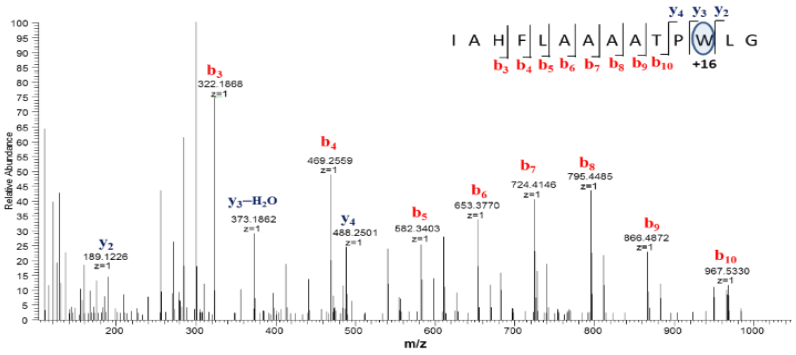
d


P A Y Y Q G S A A V A A E
+16
Relative Abundance
m/z

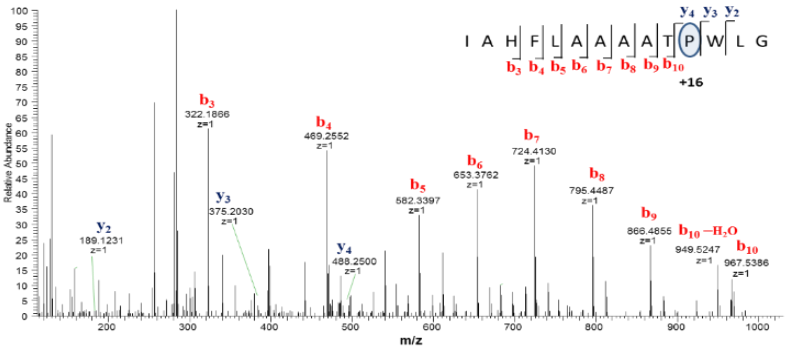
e


I A H F L A A A A T P W L G
+16
Relative Abundance
m/z

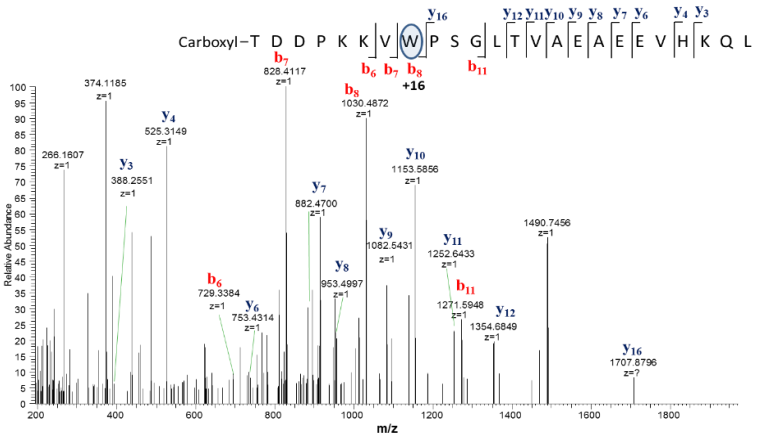
f


I A H F L A A A A T P W L G
+16
Relative Abundance
m/z

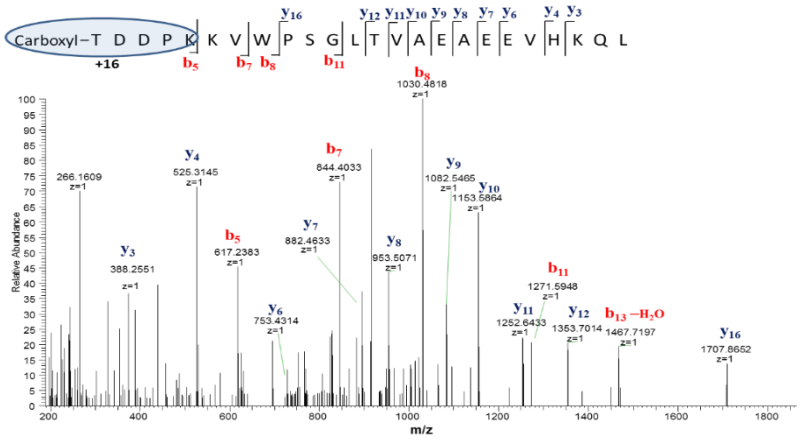
g


I A H F L A A A A T P W L G
+16
Relative Abundance
m/z

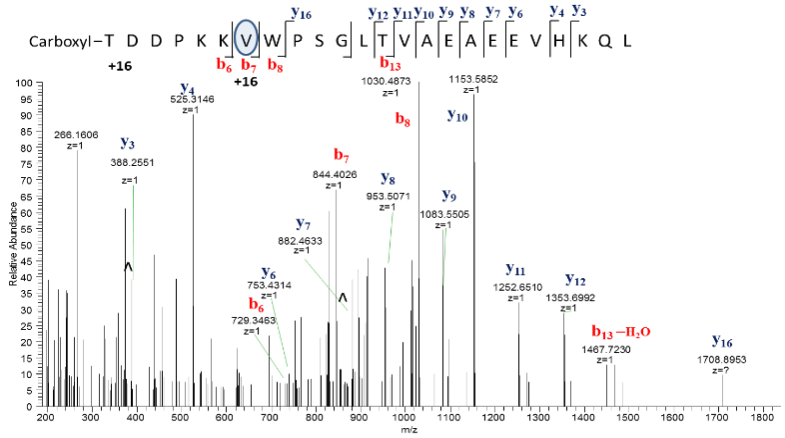
h


Carboxyl−T D D P K K V W P S G L T V A E A E E V H K Q L
+16
Relative Abundance
m/z

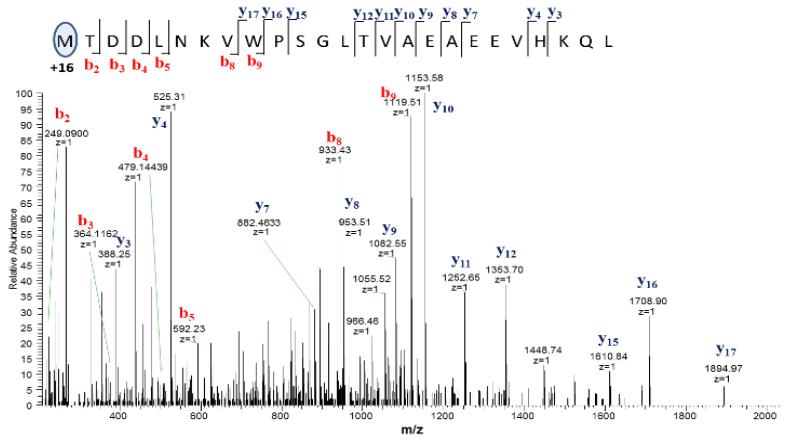
i



j



k



l

Figure 3.11. MS/MS fragmentation spectrums of different peptides as shown in the picture

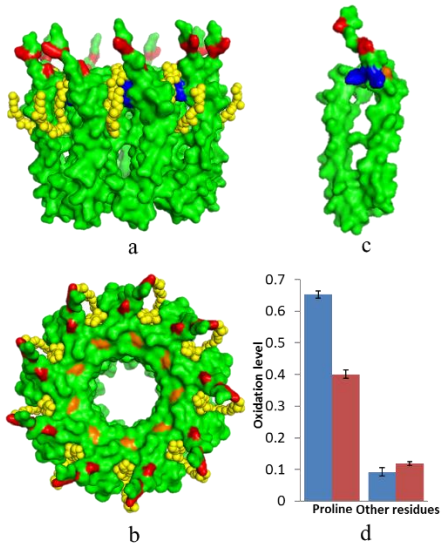


Figure 3.12. The homology model (panel c) and PDB 1NKZ (panel a and b) were used to present the residues being discussed in the paper. The "second carotenoid" (RG2) is shown in yellow as spheres. Pro in both structures are shown in orange; Tyr-Try in PDB 1NKZ and the corresponding Tyr-Tyr in homology model are shown in blue; Gly and Lys in PDB 1NKZ and the corresponding Ser and Val in homology model are shown in red. Panel d shows the oxidation level of proline vs. other residues. Peptides in detergent are shown in blue, and in Nanodiscs are shown in red.

The C-terminal loop domain of the β subunit (AAAATPWLG), does not extend from the membrane but more likely exists at a water/lipid interface and bends inward (Figure 3.13), considering the low oxidatively labeling (Figure 3.7b). Although Trp, a highly reactive residue with •OH, is present in this subunit, no prominent oxidative modification occurs for it (Figure 3.7b). MS/MS shows only the terminal Pro, Trp and Leu are oxidatively modified (Figure 3.11f-h) but not Gly, which is inert to FPOP.

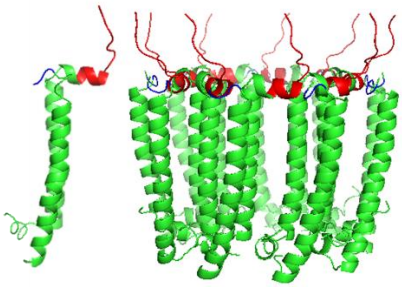


Figure 3.13. Peptides on the C-terminal of α subunits are shown in red in both the homology model (PAYYQGSAAVAAE) and PDB 1NKZ (YWQGGVKKAA). Peptides (PWL) on the C-terminal of β subunits are shown in blue in both the homology model and PDB 1NKZ. (Full length C-terminal cannot be shown here because it is not covered 100% in the homology model)

The consensus from TOPCON also indicates that the N-terminal end is much less hydrophobic than the C-terminal end. The former contains an N-terminal Met that undergoes the highest oxidative modification of all the Mets (Figure 3.7). This Met in the β subunit has higher solvent-accessibility than the others. Although this region is not covered in our homology model[64], it may exist, on the basis of LH2 from *Rh. acidophilus*[80], as an "elongated" peptide region attached to the homology model (Figure 3.8). This region extends beyond the membrane and has good solvent accessibility (Figure 3.14). The remaining part of the LH2 is approximately 5 nm in length (Pymol), a length that is nearly equal to the width of the Nanodisc (4.6 nm)[166] (Figure 3.15).

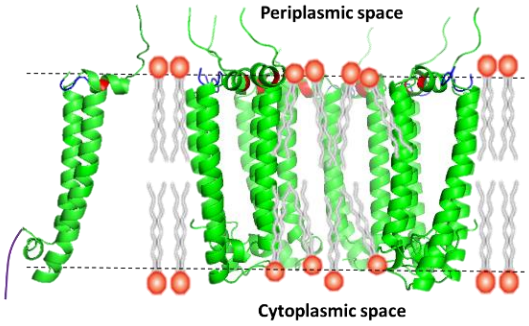


Figure 3.14. Proposed borderline between the solvent accessible domains and the domains that are embedded in hydrophobic tails of lipids. The heterodimer is shown by the homology model and the half ring structure is shown by PDB 1NKZ. Proposed position of β subunit N-terminal (MTDDLNKVWPSG) is shown in purple in the isolated heterodimer, the proline in C-terminal of α subunit is shown in red and the C-terminal of β subunit is shown in blue. The perturbed lipids packing inside the ring is presented here as cartoon.

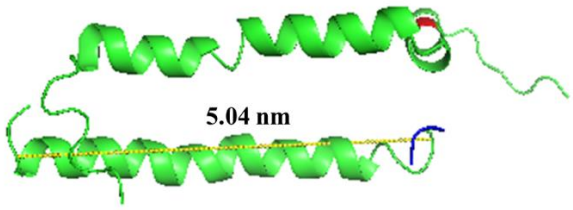


Figure 3.15. Measurement of solvent inaccessible region of LH2 by Pymol. PWL on the C-terminal of β subunit is shown in blue, highly oxidized Pro from α subunit is shown in red.

The final question we want to address is the topology difference of the two β subunits which could be crucial for the function of LH2[87]. The function of the second copy of β subunits as previously investigated by amplifying and cloning the puc2BA operon of *Rb. sphaeroides.* The resulting LH2 is spectroscopically distinct from the Puc1BA encoded LH2 with a blue-shifted B850 absorption band at 846 nm[88]. Another study found that Puc2AB-encoded LH2 are predominant under high light and in the early stages of acclimation to low light[89]. To assess the

solvent accessibility and topology of the two β subunits, we compared the oxidation levels of the N-terminal peptides from the two. We used the longer peptide on the N terminal end for this comparison as the signal intensity of the shorter β2 peptide was too low. Because a highly reactive Met is present on the N-terminal end of the β1 subunit and not on the β2 end, it is difficult to make a fair comparison. No oxidation of Val and Trp occurs for the β1 subunit, whereas ~ 10% of oxidized Val or Trp occurs in the β2 subunit (Figure 3.16). This result suggests that the regions adjacent to Val and Trp of the β2 subunit are more exposed to the cytoplasmic space than is β1. The role of these two copies of β is still not fully understood at this stage, and this study provides another perspective.

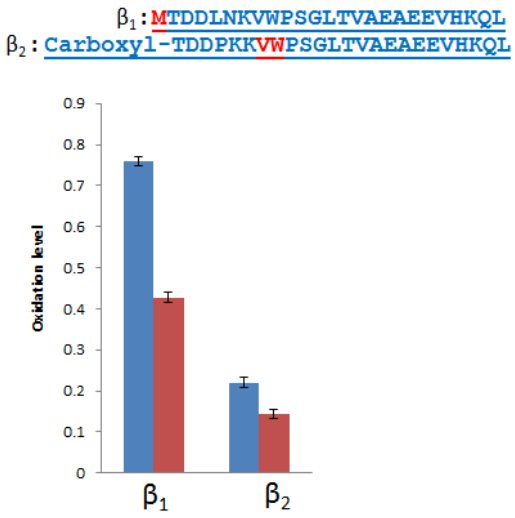


Figure 3.16. Oxidation level of N-terminal peptides from both β subunit. The oxidized residues are shown in red. Peptides in detergent are shown in blue, and in Nanodiscs are shown in red.

3.5 Conclusions

We describe here an MS-based platform to map the topology and conformation of an intrinsic membrane protein complex. To the best of our knowledge, this is the largest transmembrane assembly that has been successfully inserted into Nanodiscs. Although this large protein complex has overall ~18 transmembrane helices and most are embedded in the lipid bilayer, there is only one transmembrane helix for each subunit. As the outer-membrane structure is short, it has little higher order structure, affording an opportunity to understand, for example, steric shielding at the interface of lipid/water. We probed the LH2 topology and conformation in both lipid Nanodics and detergent micelles. On the basis of the oxidative-modification extents of peptides/residues, we conclude that Nanodiscs generally provide better protection for LH2 than do detergent micelles. For the residues that are located at the membrane interface, there is also less shielding in the Nanodisc system. Nevertheless, Met shows high modification propensity and may be a good marker for comparing solvent accessibility of different regions. Sample handling can be more considered with FPOP footprinting than with HDX. Lipid Nanodiscs offer an environment with accessibility from both sides of the membrane and the opportunity to assess topology and conformation of membrane protein in a near native-state. Thus, we think the FPOP-Nanodisc platform is a promising experimental tool for studying membrane protein topology. It opens the door for studying membrane protein interactions with other molecules, for determining the conformational changes induced by various factors, and investigating the lipid influence on membrane proteins.

Chapter 4: Mass Spectrometry characterization of reaction center from *Blastochloris viridis*——the first integral membrane protein complex determined by X-ray crystallography

4.1 Abstract

The reaction center from *Blastochloris viridis* is the first integral membrane protein complex determined by X-ray crystallography and has been studied extensively since then. It is composed of four protein subunits, H, M, L and cytochrome as well as co-factors, including bacteriopheophytin, bacteriochlorophyll, menaquinone, ubiquone-9, carotenoid and Fe. In this study, we utilized mass spectrometry to study this reaction center protein via bottom-up sequencing to top-down ligand-binding analysis. The results show a series of mutations on this protein complex and the unusual alteration and extension on the C-terminus of the M-subunit. This reaction center exhibits not only a strong ability to bind the special pair but also a tendency to preserve the two peripheral bacteriochlorophylls as shown by the native MS result.

4.2 Introduction

Membrane proteins play essential roles in various cellular process. Thirty percent of naturally occurred proteins are predicted to be membrane proteins. However, membrane proteins are not friendly to most of the characterization techniques owing to their hydrophobic properties[171-172]. In early years, researchers worked on predicting the structure of membrane proteins based on their own unique sequences, an algorithms analysis[173]. Nowadays, those tools are easily accessible on the internet and is being routinely used for membrane protein structure prediction[142]. However, experimental results on membrane protein structures are still desired. The first membrane protein structure revealed by its X-ray crystal structure at atomic resolution is the reaction center from *Blastochloris viridis*[174].

With the development of genomics and proteomics, up to date, more than 600 unique membrane protein structures are solved by X-ray crystallography[175]. Crystallography is still the primary method in providing detailed information for the structure of membrane proteins and their important ligand binding. The selection and purity of detergent molecules are crucial for obtaining those atomic resolution structures[176]. Although mass spectrometry (MS) cannot provide an atomic resolution structure, it is generally applicable to all membrane proteins. It is especially useful in determining the isoforms of membrane proteins as crystal structure determinations usually deliver structures of only one form of the membrane proteins. Thus, MS has been widely utilized to study membrane proteins and their protein constituents to obtain the sequence information. For example, our group successfully utilized MS-based techniques to extract the components and PTM information of an integral membrane protein, light harvesting complex 2 from purple bacteria[64].

In recent years, the emergence of native MS endowed the mass spectrometer with the power to decipher the topology, structure and dynamics of membrane proteins[66, 177-178]. After using collisional events to shake off the detergent micelles around a membrane protein complex, the intact membrane protein complex can be detected with a quadrupole time-of-flight (Q-TOF) MS[66]. In this experiment, the membrane proteins are dissolved in ammonium acetate buffer to which has been added a MS-friendly detergent. Then the membrane protein solution can be directly introduced into the electrospray source of the mass spectrometer[66]. The protocol maintains the integrity of the membrane protein complex after shaking off the detergent micelles. In the photosynthetic protein area, this native MS technique is especially desired, as reaction centers and light harvesting complexes, the two-major players in photosynthesis process, are mostly transmembrane protein complexes with many cofactors non-covalently interacting with the protein scaffold. Our team has demonstrated an example using this technique to characterize the reaction center from *Rhodobacter sphaeroides* in lauryl-β-D-maltoside (DDM) micelles[67].

As mentioned in the first paragraph, the reaction center from *Blastochloris viridis* is the first membrane protein obtained with an atomic resolution structure,[174] and 30 X-ray crystallographic structures of this reaction center have now been published. This protein has been a well-accepted model membrane protein to evaluate new crystallographic techniques[179-181]. Here, we describe the use of MS to study this model membrane protein for two purposes. Firstly, we want to verify the components and PTM/mutation information of this protein complex. Secondly, we want to evaluate the native MS platform for studying photosynthetic reaction centers, probing the topology as well as the co-factor binding, laying groundwork for future studies on transmembrane photosynthetic membrane proteins.

Unexpectedly, we identified an unusual extension on the C-terminal of M subunit as well as a series of mutations on all four subunits. In addition, we observed the intact reaction center in the gas phase and showed that increasing the collisional energy gradually striped off the co-factors associated with this protein complex until four bacteriochlorophylls remained, exhibiting a strong binding ability to the protein scaffold.

4.3 Materials and Methods

4.3.1 Cell culture and reaction center preparation

B. viridis strain DSM 133 cells were grown anaerobically in 1:1 YPS/RCV media and harvested[182]. A Branson 450 sonifier was used to break the cells. After sonication, the sample was centrifuged at 10 000 × g for 1 h at 4 °C using a Sorvall SS-34 rotor to pellet the cell debris. The supernatant was spun at 450 000 × rpm for 4 h at 4 °C using a Beckman Type 70 Ti (BT) rotor. After pelleting the membranes, 30% (v/v) lauryldimethylamine N-oxide (LDAO, Sigma) was added to the membrane pellets resuspension. The resuspension was stirred at 1.5% (v/v) final LDAO concentration at room temperature for ~1 hr followed by centrifugation (450 000 × rpm for 1 h at 4 °C using a Beckman Type 70 Ti (BT) rotor). The supernatant was kept and loaded on to HiTrap Q HP anion-exchange column (GE Healthcare) and eluted with a gradient of NaCl-containing buffer (20mM Tris). The separation was further elaborated by elution from a Sephacryl S-200 (GE Healthcare) gel filtration column previously equilibrated with a buffer containing 100 mM NaCl (20mM Tris).

4.3.2 Top-down LC-MS

For intact protein MS analysis, the experiment was conducted by using the protocol described in a previous pulication[64]. Briefly, the purified reaction center was precipitated by

acetone, and the pellet was solubilized in 40% formic acid right before the analysis. After separation on custom-packed capillary column (PLRP/S, 5 µm, 1000 Å, ~10cm) (Waters Inc., Milford, MA), a hybrid ion-mobility quadrupole TOF (Synapt G2, Waters Inc., Milford, MA) was used to analyze the molecular weight of protein subunits.

4.3.3 Bottom-up LC-MS

Cyanogen bromide (CNBr) cleavage of reaction center was performed following a published protocol[93]. The product was vacuum dried and then dissolved in 10 µL 8 M urea in preparation for following enzymatic digestion. Peptide mixtures were trapped by a guard column (Acclaim PepMap100, 100 µm × 2 cm, C18, 5 µm, 100 Å; Thermo Fisher Scientific, Breda, Netherlands) and then fractionated on a ACQUITY UPLC Peptide BEH C18 Column (10 K psi, 130 Å, 1.7 µm, 75 µm X 100 mm, Waters Corporation, Milford, MA). The MS analysis was with a Thermo Scientific™ Q Exactive™ hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen Germany). Peptides were eluted with a 120 min, 250 nL/min gradient coupled to the nanospray source. The default charge state chosen for the MS was 3, and the scan range was from *m/z* 380-1500. Mass spectra were obtained at high mass resolving power (70,000, FWHM at *m/z* 200) and the top 15 most abundant ions corresponding to eluting peptides per scan were submitted to CID in the ion trap, with charge-state rejection of unassigned and > 8 ions enabled. Precursor ions were added to a dynamic exclusion list for 8 s to ensure good sampling of each elution peak.

4.3.4 Native MS analysis of reactions center

Buffer exchange assisted by a 100 kDa MWCO filter (Millipore Amicon Centrifugal Filters, Billerica, USA) was conducted on the purified reaction centers. After 5 cycles of concentration-dilution, 10 μL of the mixture was loaded into an offline electrospray capillary (GlassTip 2 μm ID, New Objective, Woburn, USA). The sample solution was injected to a hybrid ion mobility quadrupole time-of flight mass spectrometer (Q-IM-TOF, SYNAPT G2 HDMS, Waters Inc., Milford, MA). The instrument was operated in the sensitive mode under gentle ESI conditions (ES387, Hudson, New Hampshire, Thermo Scientific, source temperature 37 °C). The sampling cone and extraction cone voltages were adjusted to reach the best signal for protein complexes. The pressure of the vacuum/backing region was 5-6 mbar. The instrument was externally calibrated up to 10000 m/z with a NaI solution. The peak picking and data processing was performed in Masslynx (v 4.2, Water Inc, Milford, MA). For IM experiments, the gas flow rate was 50 mL/min, the ion mobility separation (IMS) wave height was 40 V, and the IMS wave velocity was 700 m/s.

4.3.5 MS data processing

The top-down MS raw data file was combined and smoothed in Masslynx (v 4.2, Water Inc, Milford, MA). The mass list with intensities was exported and saved as a txt file for re-plotting and data analysis. Massign software package was used to assign peaks in reaction center mass spectra[183]. The bottom-up MS data were processed by PEAKS software and searched against the *B. viridis* proteomics database. For ion mobility MS experiment, the drift-time information from native MS ion-mobility experiments was converted into CCS by considering the molecular weight and charge state of the protein assemblies[184-185].

4.4 Results and discussion

4.4.1 Top-down MS analysis of denatured reaction center

Although more than a dozen crystallographic structures of this reaction center are available, we still want to examine the components by top-down MS, to remove any concerns about contaminants during the purification process as well as to examine the possible existence of isoforms. We observed a total of four protein components by top-down LC-MS. All four protein subunits from the reaction center, H, M, L and cytochrome exhibit a discrepancies of the experimental MW to the theoretical one (Figure 4.1) (the calculated MW based on the genome[186] and experimental MWs are listed in Table 4.1). The MW discrepancies on H, L and cytochrome may be ascribed to mutations or the presence of PTMs. Such do not readily explain the more than 600 Da MW discrepancy on the M subunit. Thus, we decided to use bottom-up MS to decipher the unusually large MW discrepancy on the M subunit.

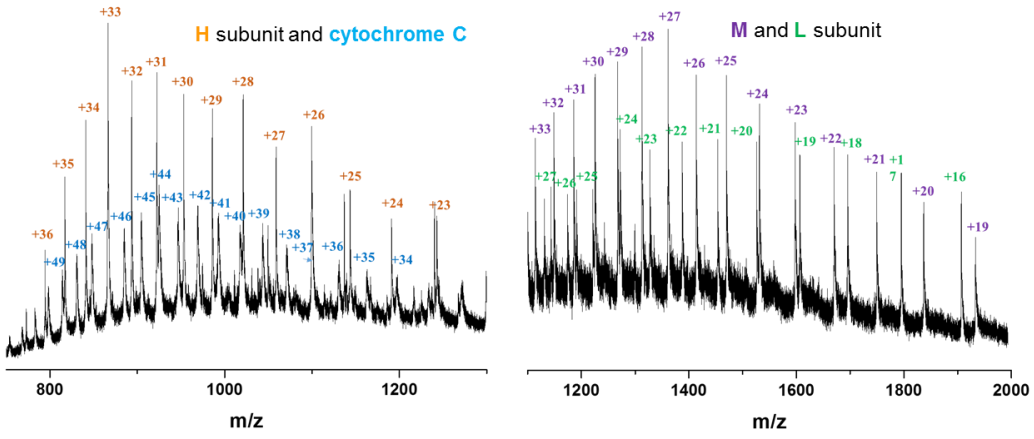


Figure 4.1. Mass spectra of the four subunits from reaction center (*B. viridis*)

Table 4.1. Comparison of molecular weights (MW) determined by top-down MS and genomic sequence

Subunit	Experimental (Da)	Theoretical (Da)
H	28549	28526
M	36713	36035
L	30495	30447
Cytochrome	40612	40443

4.4.2 Bottom-up MS analysis of reaction center

To map those MW discrepancies on amino acid sequence, we utilized a combination of chemical and enzyme-cleavage reagents, CNBr and trypsin, to locate any possible PTMs and mutations. Many mutations are located on all four subunits, and unexpectedly, the C terminal alteration and extension was identified on the M subunit. Rosak and co-worker[187] reported the intraspecies evolution of reaction center over the 14 years in the laboratory from the same strain that we studied. They identified a total of 16 amino acid mutations by comparing the genome sequence of the current active strain and the glycerol stock strain from 14 years ago. In this work, we identified more than 20 mutations by MS sequencing (Table 4.2). One of the typical mutation types is “methylation/demethylation” as seen by the alteration between D and E, and V and L. The reason for those alterations are not clear; it could be the evolution of genome in the laboratory, or a change of codon after transcription but before translation during the RNA editing.

Table 4.2. List of mutations that we identified on the four subunits of reaction center. For example, the forty-fourth amino acid on H subunits is shown as H44.

Position	H44	H55	H126	H128	H158	H163	H186	H215	H216	H252	L66	L152	L187	L250
Predicted	V	D	A	V	V	V	L	S	E	E	S	S	A	A
Experimental	L	E	D	L	L	L	F	A	D	D	A	A	D	S

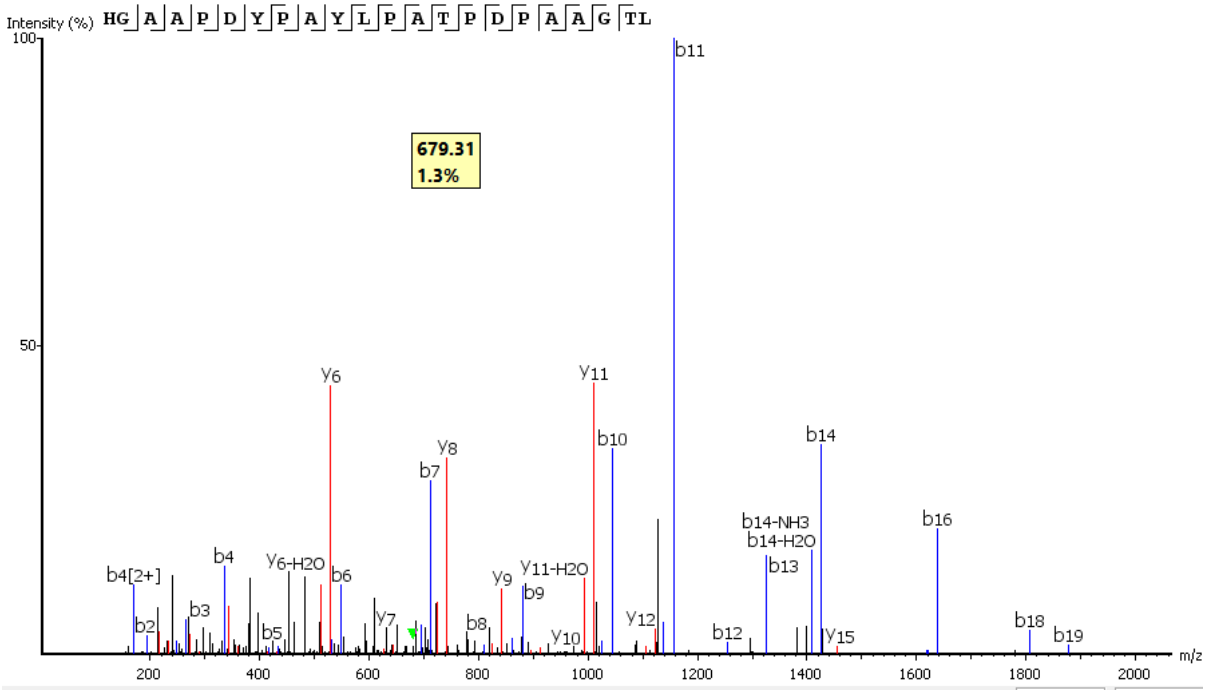
Position	M4	M7	M27	M30	M36	C64	C68	C180	C192	C196	C199	C331
Predicted	Y	I	N	S	S	N	A	T	I	M	N	R
Experimental	W	V	E	L	L	D	P	Q	V	R	T	Q

The most surprising finding is not the large number of mutations, but the seven-amino acid alteration plus seven-amino acid extension on the C-terminus of the M subunit (Figure 4.2). When we identified the large MW discrepancy of M subunit by top-down approach, we posited that PTMs to be the most likely reason. Owing to the hydrophobicity of the M subunit, we first used CNBr to cut the protein into large sections and then used trypsin to cleave those long peptides into smaller pieces. In addition, we considered the report that dilute acid can cleave at aspartic acid[188]. Because of these cleavage processes, we could observe several peptides mapping the whole region of this C-terminal extension (Figure 4.2, See Figure 4.3 for product-ion (MS/MS) spectra).

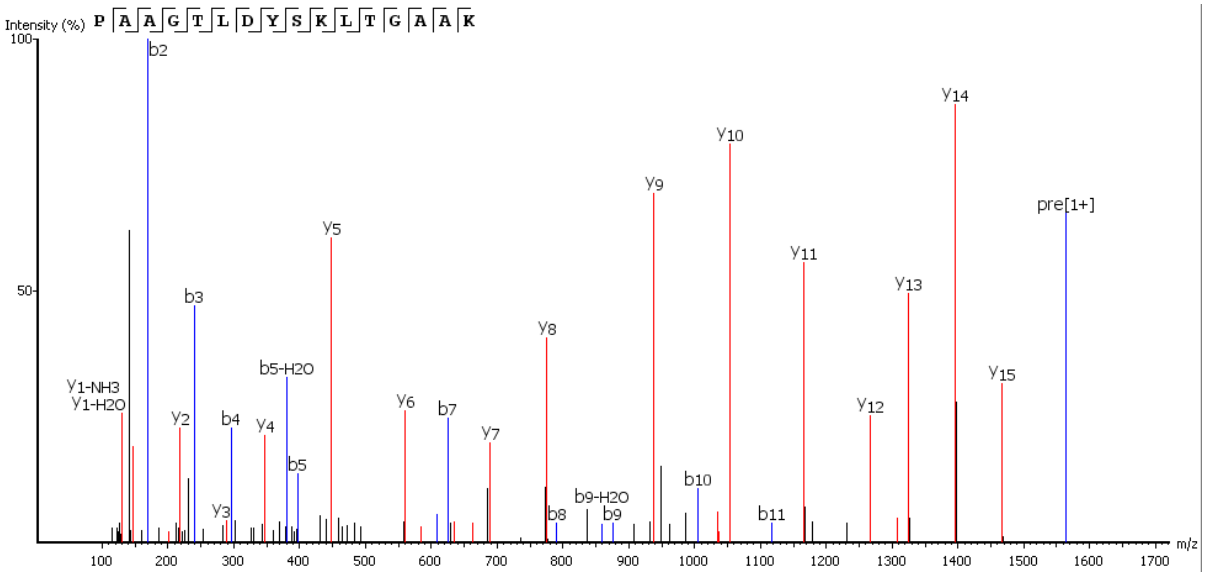
(A)

(B)

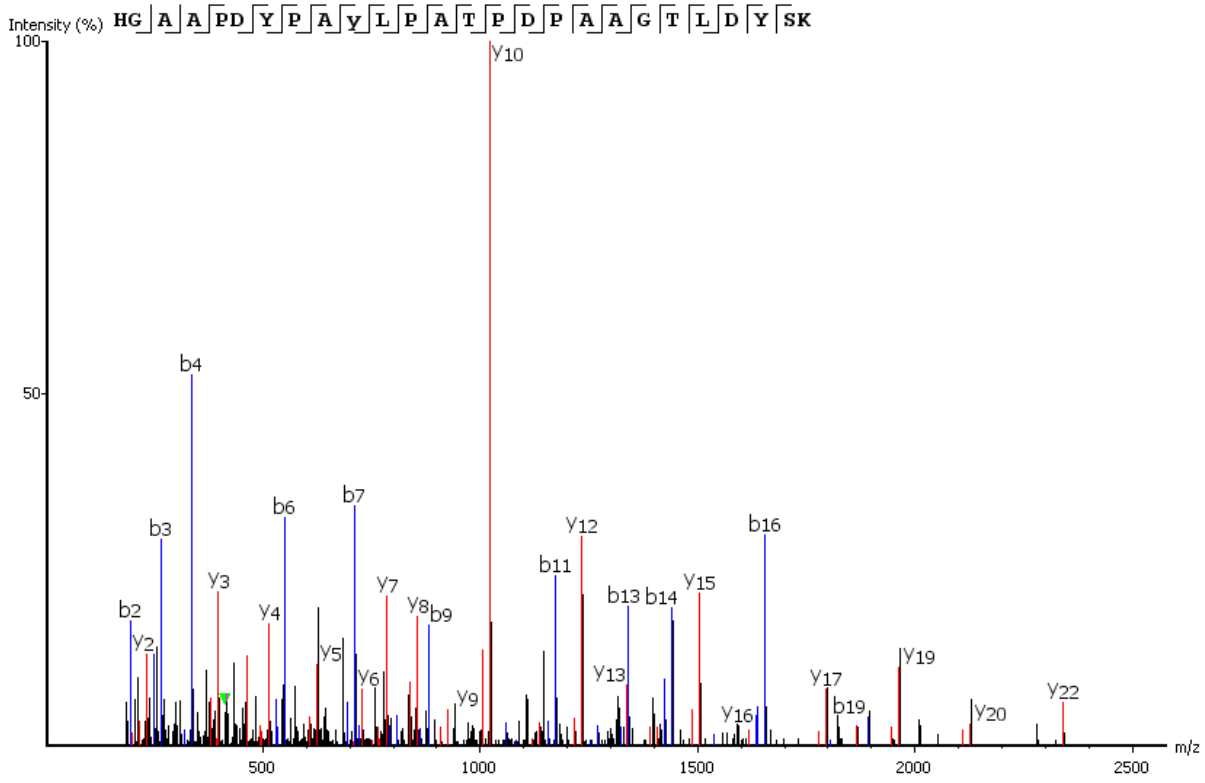
Figure 4.2 (A) C-terminal sequence on the M subunit identified by MS (B) Sequence coverage on the C-terminus of the M subunit.



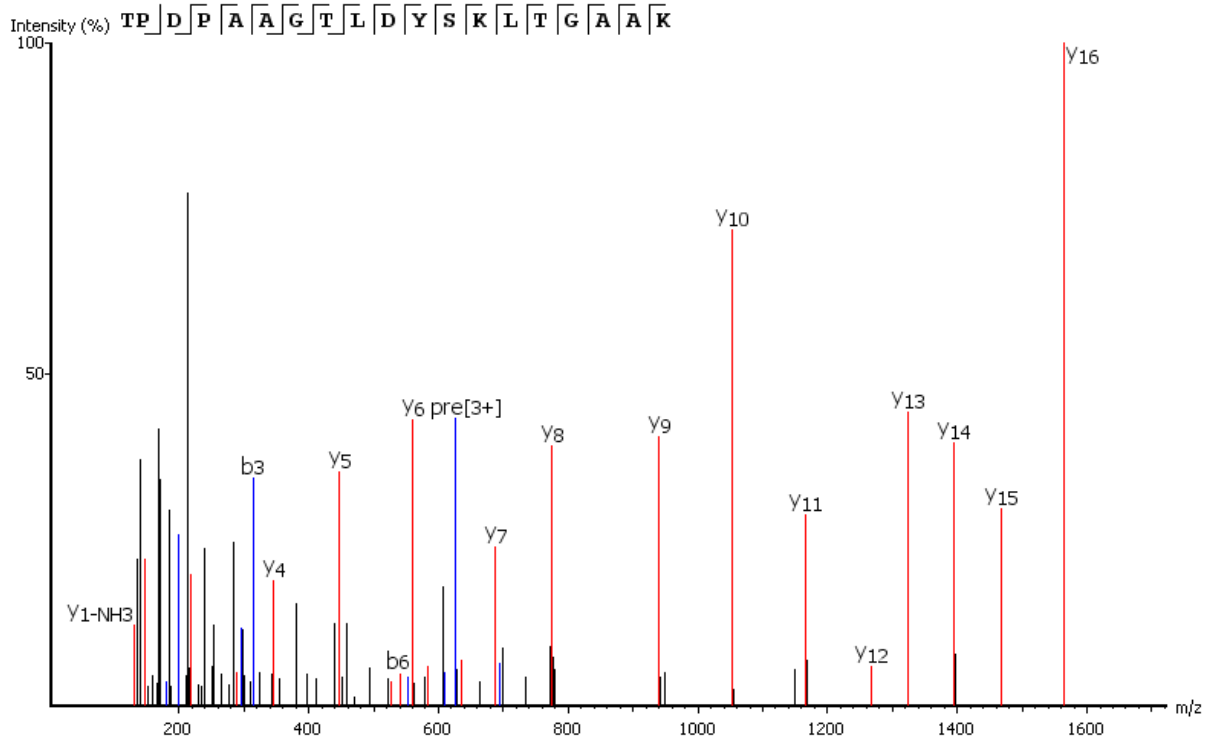
(A)


Intensity (%)
P A A G T L D Y S K L T G A A K
100
50
b2
y14
y10
y9
pre[1+]
y5
y11
y13
b3
y8
b5-H2O
y15
y1-NH3
y1-H2O
y6
b7
y12
y2
b4
y4
y7
b5
b10
y3
b9-H2O
b8
b9
b11
m/z
100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700

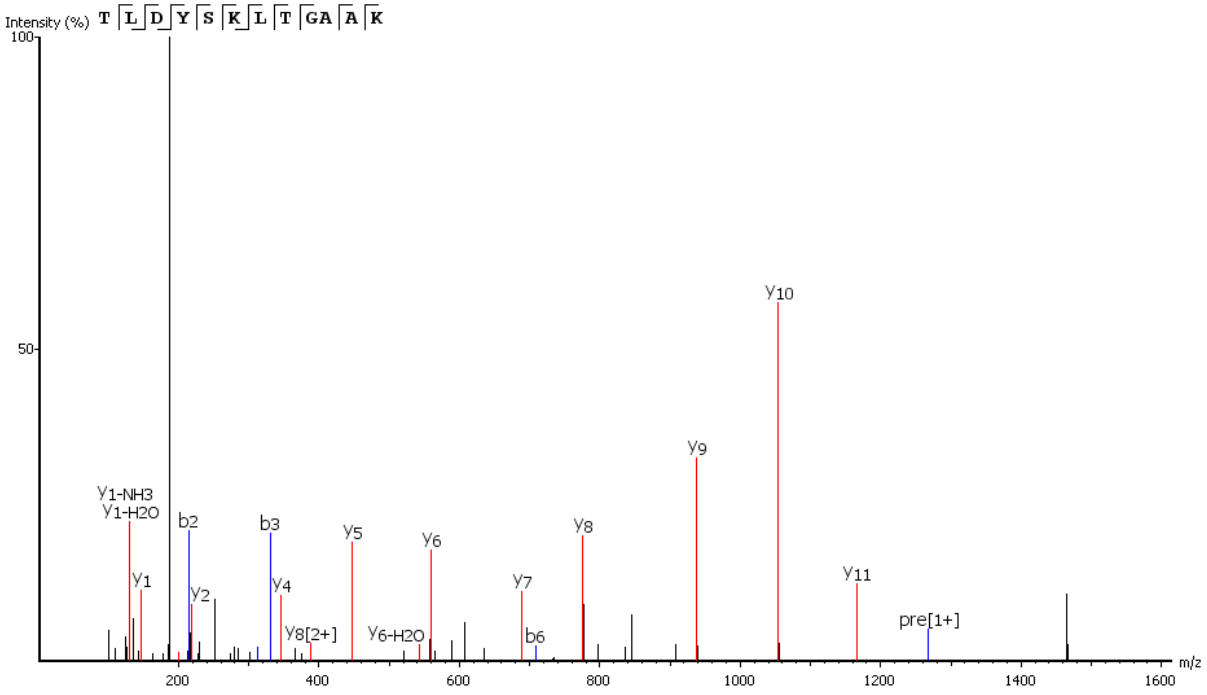
(B)


Intensity (%)
HG A A PD Y P A y L P A T P D P A A G T L D Y SK
100
50
y10
b4
b6
b7
b3
y12
b16
b11
y3
y7
b13
b14
y15
b2
y4
y8
b9
y2
y5
y19
y6
y13
y17
y9
b19
y16
y20
y22
m/z
500 1000 1500 2000 2500

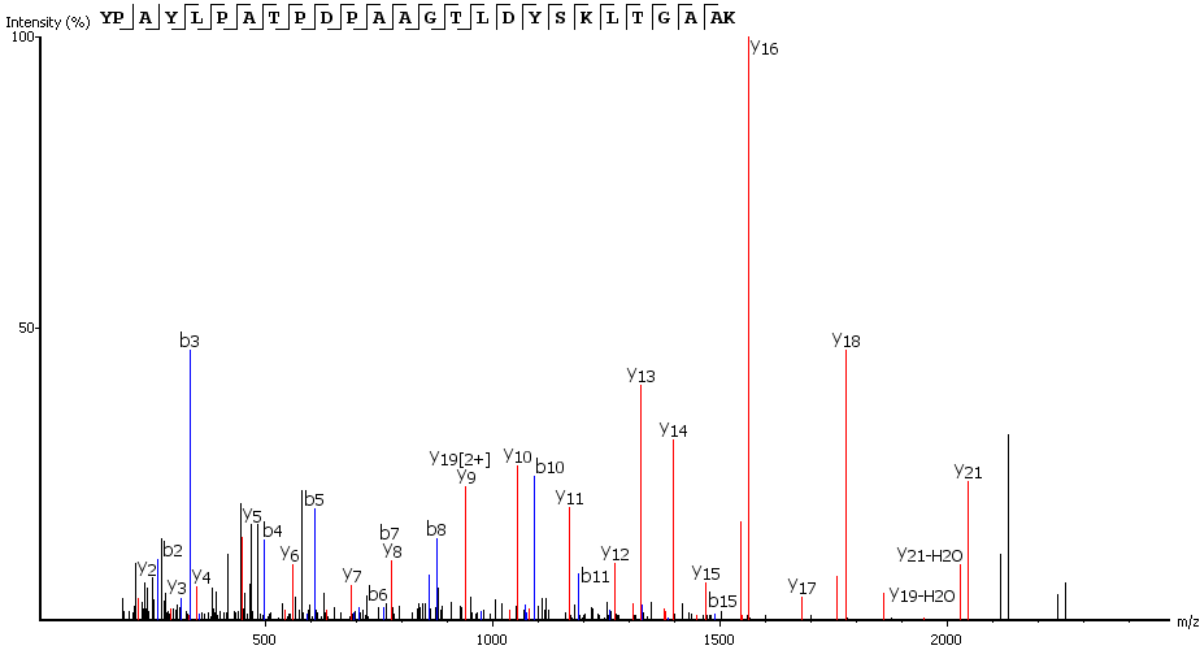
(C)


Intensity (%)
TP D P A A G T L D Y S K L T G A A K
100
50
y16
y10
y13
y14
y15
y9
y8
y6 pre[3+]
y5
b3
y11
y7
y4
y1-NH3
b6
y12
200
400
600
800
1000
1200
1400
1600
m/z

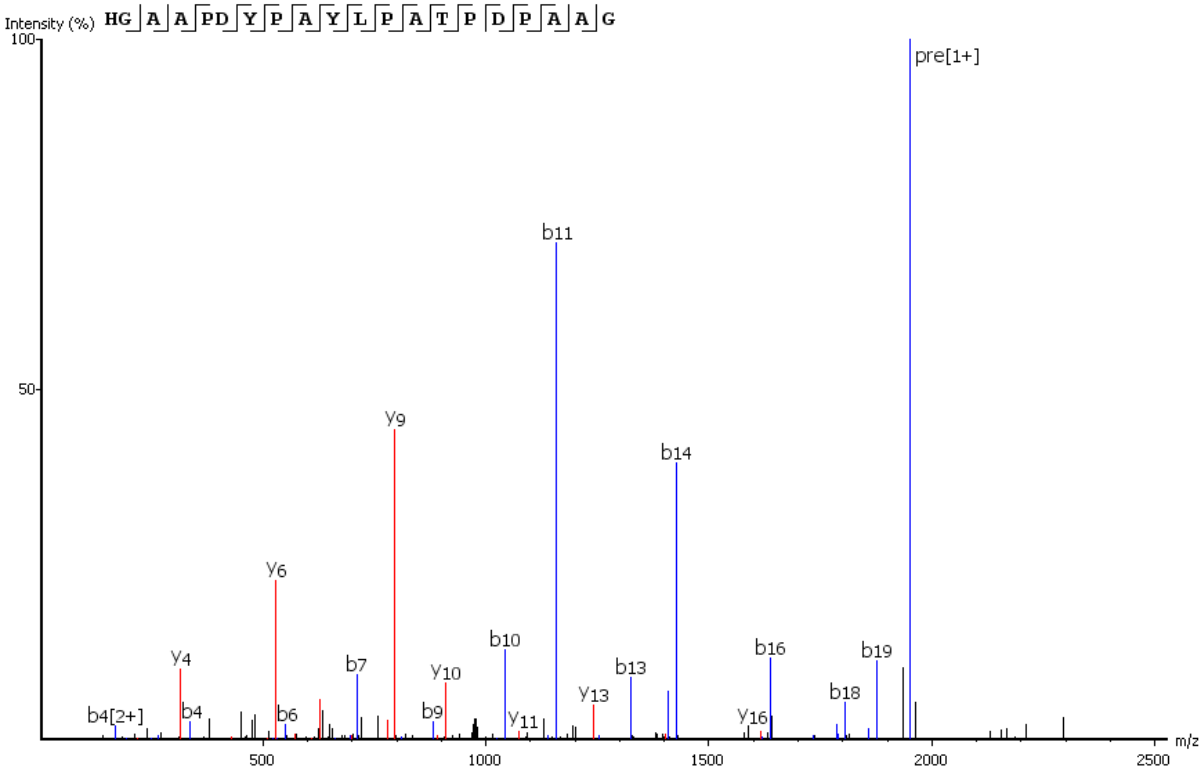
(D)


Intensity (%)
T L D Y S K L T GA A K
100
50
y10
y9
y1-NH3
y1-H2O
b2
b3
y5
y6
y8
y11
y1
y2
y4
y7
pre[1+]
y8[2+]
y6-H2O
b6
200
400
600
800
1000
1200
1400
1600
m/z

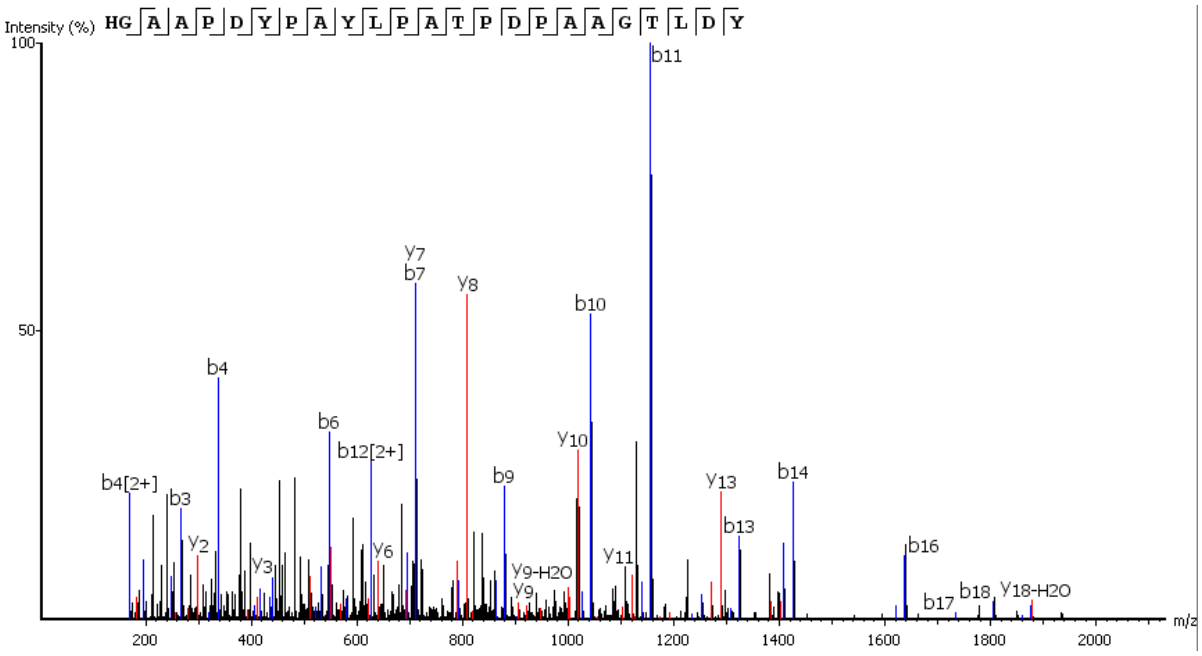
(E)


Intensity (%)
YP A Y L P A T P D P A A G T L D Y S K L T G A AK
100
50
y16
b3
y18
y13
y14
y19[2+]
y9
y10
b10
y21
b5
y11
y5
b4
b7
b8
y6
y8
b2
y12
y21-H2O
y2
y4
y3
b6
y7
b11
y15
y17
b15
y19-H2O
500
1000
1500
2000
m/z

(F)


Intensity (%)
HG A A PD Y P A Y L P A T P D P A A G
100
50
pre[1+]
b11
y9
b14
y6
b10
b16
b19
y4
b7
y10
b13
b18
b4[2+]
b4
b6
b9
y11
y13
y16
500
1000
1500
2000
2500
m/z

(G)



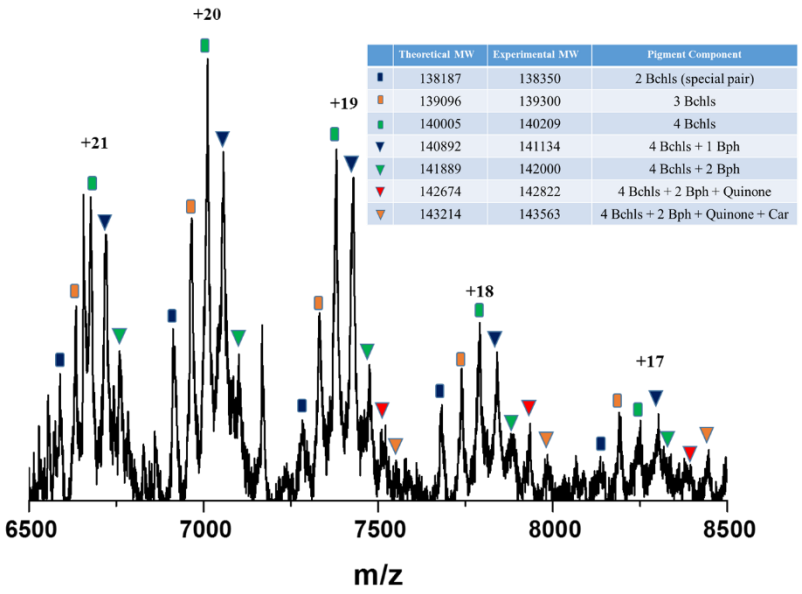
(H)

Figure 4.3 Product-ion (MS/MS) spectra of peptides on the C-terminus of the M subunit. The fragment ion assignment is performed by PEAKS software[189].

The core subunit D1 in reaction center (PSII) from cyanobacteria is firstly synthesized with a C-terminal extension and must be processed by a C-terminal peptidase before incorporation into PSII and final assembly. In higher plants, this core subunit D1 is essential not only for the PSII assembly but also for the formation of supercomplexes[26]. Here, we also found the extension of C-terminal on the reaction center from purple bacteria. The evolutionary relationship of reaction center among cyanobacteria, higher plants and purple bacteria could be an intriguing topic for future studies[190].

4.4.3 Native MS analysis of reaction center

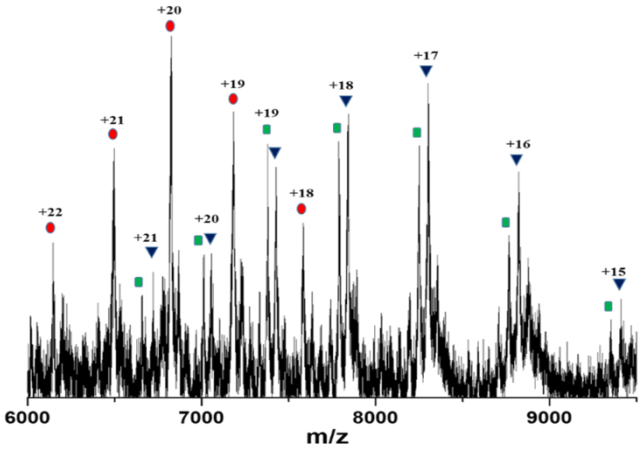
Native MS analysis on intrinsic membrane proteins is an emerging and quickly developing field[177, 191]. Native MS enables us to analyze the reaction center in near-native state[191]. The reaction center is dissolved in ammonium acetate buffer with the addition of DDM. The mass spectra show that the intact RC can be observed in the gas phase with/without co-factors (Figure 4.4). Under medium collisional energy, we can "shake off" the detergent micelles and observe a series of peaks corresponding to intact reaction center protein scaffold binding to different number of co-factors (See spectrum in Figure 4.4A). Owing to the mild analysis conditions, various numbers of water or ammonium can remain with the reaction center, causing small increases in the experimental MW compared to the theoretical one. Increasing the collisional energy in the Trap region of the mass spectrometer leads to the gradual loss of co-factors (Figure 4.5). Under the most vigorous conditions, the reaction center can be completely striped of co-factors, leaving only the protein scaffold (Figure 4.4B). In addition, we found that the reaction center protein scaffold with four bacteriochlorophylls comprises a stable complex in the gas phase, producing the most intense peaks during medium collisional energy and remaining during application of the largest collisional energy available on the instrument (Figure 4.4).



	Theoretical MW	Experimental MW	Pigment Component
	138187	138350	2 Bchls (special pair)
	139096	139300	3 Bchls
	140005	140209	4 Bchls
	140892	141134	4 Bchls + 1 Bph
	141889	142000	4 Bchls + 2 Bph
	142674	142822	4 Bchls + 2 Bph + Quinone
	143214	143563	4 Bchls + 2 Bph + Quinone + Car

(A)

	Theoretical MW	Experimental MW	Pigment component
	136369	136416	0
	140138	140052	4 Bchls
	141160	1400939	4 Bchls + 1 Bph



(B)

Figure 4.4 Native mass spectrum of reaction center under (A) medium collisional energy (The first peak series represents the RC protein binding to the special pair) (B) the highest collision energy that can be achieved

Our previous analysis on a different reaction center from *Rb. sphaeroides* shows that reaction centers can be observed in the gas phase in a near-native state, and they have a strong tendency to preserve the bacteriochlorophyll special pair[192]. Here, we observed that the reaction-center protein from *B. viridis* also strongly binds to the special pair as shown fig 4.4a. This new reaction center investigated here, however, tends to preserve more co-factors, four bacteriochlorophylls, including the special pair and some peripheral ones (Figure 4.4, 4.5). Those two reaction centers must adopt a similar overall architecture and protein-pigment interactions[193]. The different results obtained by native MS analysis might be attributed to the extra cytochrome in the reaction center from *B. viridis*, which stabilizes and protects the co-factors in the protein complex.

Unfolded protein complex ions undergo more collisions with the neutral gas in an ion mobility chamber and, therefore, exhibit a larger collisional cross section (CCS) than the theoretical one[36]. Here, we observed the ion mobility-MS (IM-MS) measurements of CCS on reaction center proteins are larger than the theoretical one. In addition, the measurement exhibit a linear relationship between charge state and protein size, indicating the partial unfolding of the reaction center complex in the gas phase (Figure 4.6). A linear relationship between charge state and CCS has been confirmed by multiple studies[194-196]. In brief, the reaction center we observed in the gas phase might be partially unfolded regarding its crystal structure.

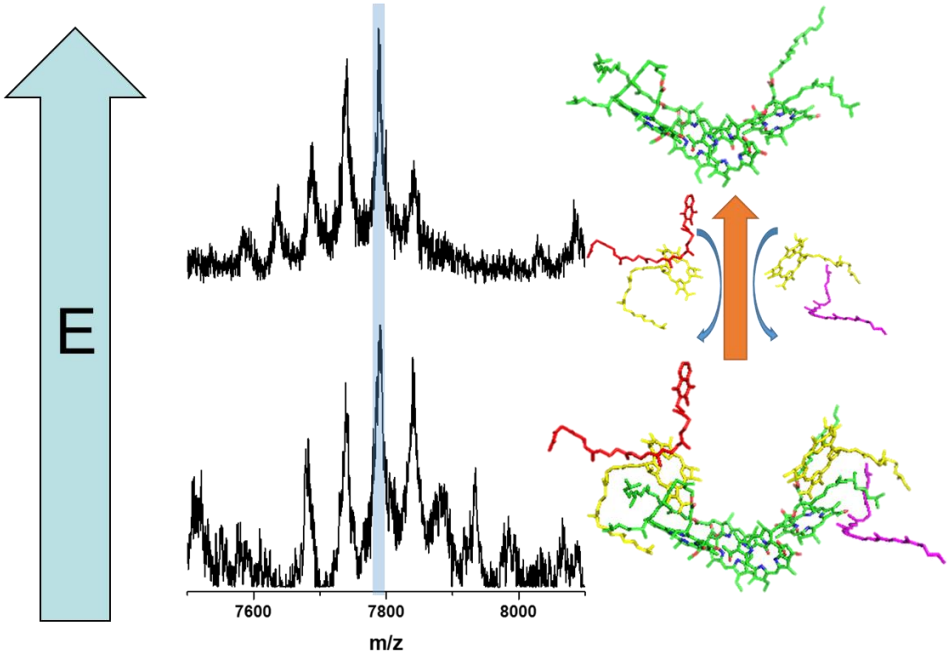


Figure 4.5. Loss of the peripheral pigments, carotenoid, quinone, and Bacteriopheophytin. The highlighted peak represents the intact RC carrying four bacteriochlorophylls. The four bacteriochlorophylls bound to the protein remain as the most abundant component under these conditions.

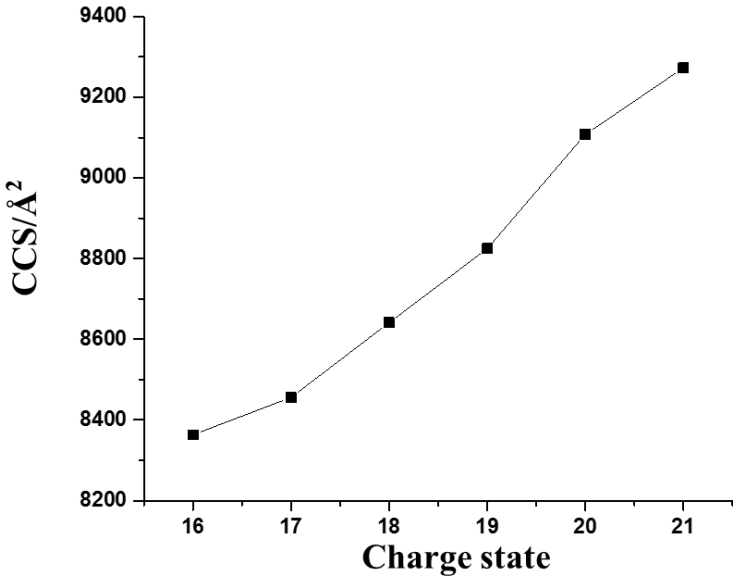


Figure 4.6 Ion mobility MS measurement of CCS of reaction center at different charge states.

4.5 Conclusions

The reaction center from *B. viridis* is one of the most well-studied intrinsic membrane proteins, and it has been utilized as a model membrane protein in the development of new analytical approaches. We found that this membrane protein, however, might not behave exactly as we expected. Many mutations were located on the proteins, including alteration between valine and leucine, aspartic acid and glutamic acid and so forth. In addition, the unusual alteration and extension on the C-terminus of the M subunit occurs, including a total of 14 amino acids. This result is not consistent with the reported crystallographic structures, and the reason for the discrepancy is not clear at this stage. In addition, this reaction center can also be observed in the gas phase at near-native state just like the one from *Rb. sphaeroidses,* exhibiting a tendency of preserving all four bacteriochlorophyll molecules as a "special quartet".

Chapter 5: Native mass spectrometry analysis of oligomerization states of FRP and OCP: Two proteins involved in the cyanobacterial photoprotection cycle

This chapter is adapted from the previously peer-reviewed and published first-authored manuscript:

Lu, Y., Liu, H., Saer, R., Zhang, H., Meyer, C., Li, V., Shi, L., King, J., Gross, M., and Blankenship, R. Native mass spectrometry analysis of oligomerization states of FRP and OCP: Two proteins involved in the cyanobacterial photoprotection cycle. Biochemistry. 2017. 56 (1): 160–166.

5.1 Abstract

The orange carotenoid protein (OCP) and fluorescence recovery protein (FRP) are present in many cyanobacteria, and regulate the essential photoprotection cycle in an antagonistic manner as a function of light intensity. We characterized the oligomerization states of OCP and FRP by using native mass spectrometry, a technique that has the capability of studying native proteins under a wide range of protein concentrations and molecular masses. We found that dimeric FRP (dFRP) is the predominant state at protein concentrations ranging from 3 μM to 180 μM, and that higher order oligomers gradually appeared at protein concentrations above this range. The OCP, however, demonstrates significantly different oligomerization behavior. Monomeric OCP

(mOCP) dominates at low protein concentrations, with an observable population of dimeric OCP (dOCP). The ratio of dOCP to mOCP, however, increases proportionally with the protein concentration. Higher order OCP oligomers form at protein concentrations beyond 10 μM. Additionally, native mass spectrometry coupled with ion mobility analysis allowed us to measure protein collisional cross sections (CCS) and interrogate the unfolding of different FRP and OCP oligomers. We found that monomeric FRP exhibits a roughly one-stage unfolding process, in accord with its C-terminal bent crystal structure. The structural domain compositions of FRP and OCP are compared and discussed.

5.2 Introduction

Oligomerization of proteins is a common phenomenon; 35% or more of the proteins in a cell are oligomeric[197]. This behavior is advantageous for protein evolution because it opens opportunities for new function and control. The strength and duration of association depend on the nature of the protein and various experimental variable (e.g., T, pH, and concentration)[198]. Oligomerization also occurs for photosynthetic proteins, which are involved in solar energy capture and storage. An example is chlorophyll *a*-chlorophyll c_2-peridinin-protein (apcPC), one of the major light harvesting complexes in the dinoflagellate *Symbiodinium.* Size exclusion chromatography (SEC), blue native gel electrophoresis (BN-PAGE), and native mass spectrometry (MS) all demonstrate that it exists as a trimer[199]. Nondenaturing electrophoresis of light-harvesting complex 2 from a purple bacterium revealed the presence of dimers, trimers, and even supercomplexes[200]. These examples provide valuable insights into the higher-order assemblies of and interactions among the different components of a photosynthetic apparatus. The oligomerization states of many photosynthetic proteins, whether they be monomers, oligomers, or co-exist as multi-oligomeric states, however, still remain largely unclear.

Photosynthesis starts with light-energy absorption by pigment-protein antenna complexes and continues with transfer to reaction centers for photochemistry. The regulation of energy transfer is crucial for sustainable photosynthesis because excess energy, if not properly dissipated, may damage the photosynthetic machinery and lead to photoinhibition and cell death[201].

In many cyanobacteria, the excess excitation energy absorbed by phycobilisome (PBSs) antennae is dissipated through a nonradiative pathway. This pathway starts from the absorption of blue light by the orange carotenoid protein (OCP) to induce a color and conformational change of the OCP from an orange (OCP^O) to a red form (OCP^R)[202]. The OCP^R is competent for PBS binding, forming a PBS-OCP^R quenching complex. Quenching is terminated by the action of a fluorescence recovery protein (FRP), which facilitates the conversion of OCP^R back to OCP^O, thereby expelling OCP from the PBS. The OCP is the first photoactive protein identified to use a carotenoid as the photoresponsive chromophore[203]. The crystal structure of OCP reveals an antiparallel homodimer[204-205]. Each monomer is composed of two domains that encompass a keto-carotenoid: an α-helical N-terminal domain (NTD) and an α helix/β sheet C-terminal domain (CTD). Although dimerization of OCP may be an artifact of crystallization[205], native MS, which does not require protein crystallization, reveals the existence of both dimeric and monomeric OCP[55].

The oligomerization of FRP, like that of OCP, is also not well understood. Most of the FRP sequences in cyanobacterial strains contain 106 to 111 amino acids. In *Synechocystis* sp. PCC 6803, the first Met (GTG) encoded by slr1964 gene can be considered as the first Met of the FRP and the fourth Met coincides with the first Met of most of the homologs. Using a series of *Synechocystis* mutants, Gwizdala et al.[206] compared “shorter” and “longer” versions of FRP. The

results suggest that the longer version FRP (beginning at Met1) synthesized in *Synechocystis* is less active than the shorter version, and that the starting Met for the shorter version is Met26. Remarkably, FRP exists in two different oligomeric states in the same crystal unit: dimeric and tetrameric, and the dimer is in two different conformations. Based on co-immunoprecipitation and docking simulation results, one dimeric FRP (dFRP) appears to be the active form whereas the tetrameric form may be inactive[207].

Thus, the oligomeric states of OCP and FRP remain an open question. Here we report the use of native MS to probe the oligomeric state of FRP and OCP in solution, and to investigate further how concentration affects the oligomeric states of those proteins. Native MS is emerging for the characterization of proteins, especially large protein assemblies that are recalcitrant to crystallization[53, 208]. Although native MS does not provide resolution at the atomic level, its broad mass range and ability to analyze multiple species simultaneously makes it a powerful tool for interrogating protein and cofactor stoichiometries, protein topologies, and ligand-protein interactions[209].

One classical example is the characterization of the 4-oxalocrotonate tautomerase protein, which was originally reported as a pentamer[210]. A hexamer, however, was later observed with X-ray crystallography[211], consistent with native MS[212]. Determination of protein oligomeric states by native MS, however, is challenging to interpret when oligomers dissociate during desolvation or form as non-specific adducts in the spray[213]. The former possibility can be avoided by carefully optimizing sample cone and collisional voltages to maintain the complex intact. For the latter, protein concentration seems to be an important factor. For example, native MS identified urease as an $(\alpha\beta)_{12}$ assembly that readily disassembles into $(\alpha\beta)_3$ subunits, supporting an $((\alpha\beta)_3)_4$ architecture, in accord with the crystal structure, which revealed an $(\alpha\beta)_{12}$ assembly. At higher

concentrations, urease forms 24-, 36-, and even 48-mers in the gas phase, probably as non-specific adducts without any biological relevance[214]. Insulin, for example, not only forms well-defined oligomers in its native state, but also aggregates and gives amyloid fibril[215]. Native MS shows that association of insulin is concentration-dependent whereas HDX shows the rapid equilibrium between higher order oligomers and monomers[216].

Additionally, native MS can be coupled with ion mobility (IM) to separate ions based on their size and shape and provide collisional cross section (CCS) to be compared with those from the crystal structure and from theory[217-218]. Increasing the collisional voltage causes protein unfolding as seen by an increase in the CCS; thus, native MS and IM can provide insights in conformational dynamics of protein higher order structures[219-220]. Our goal in this work is to utilize native MS and IM to probe the oligomerization state of FRP and OCP in their native states.

5.3 Materials and Methods

5.3.1 Expression of FRP in *E. coli*

The full-length version of FRP (*Slr1964*) without its stop codon was amplified by using the primers (FRPndeIF and FRPEcoRIR, Table 5.1) from genomic DNA of *Synechocystis* sp. PCC 6803 and cloned into the pET21a vector. The insertion sites are NdeI and EcoRI as indicated by the primer names. The truncated (short) version of FRP starting from the second open reading frame (i.e., MLQTAEA) was generated by using primers (FRPSF and FRPSR, Table S1). Cell culture and protein induction were performed according to previously reported methods[207].

Table 5.1. Primers used in the construction of full length and truncated version of FRP in pET21a

Name	Sequence (5'-3')
FRPndeIF	AGC TGG CAT ATG GTC ATG ATA ATT ACA AAT C
FRPEcoRIR	GCA AAT GAA TTC TCA GTG ATG GTG ATG GTG ATG CAG CCG TGC CAG GGC CTT AA
FRPSF	GTT TAA CTT TAA GAA GGA GAT ATA CAT ATG TTA CAA ACC GCC GAA GCA CC
FRPSR	GGT GCT TCG GCG GTT TGT AAC ATA TGT ATA TCT CCT TCT TAA AGT TAA AC

5.3.2 FRP purification

E. coli cells were lysed by sonication in 20 mM Tris buffer (pH 7.5) buffer A supplemented with protease inhibitor cocktail and DNase (Sigma, St. Louis, MO) and 200 mM NaCl. The cell lysate was clarified by centrifugation at 25,000 x g, and the supernatant was loaded onto a HisTrap HP column (GE healthcare, Marlborough, MA). The His_6-tagged FRP was eluted with buffer A containing 300 mM imidazole, and further purified by gel filtration chromatography by using a HiPrep Sephacryl S-200 HR (GE healthcare, Marlborough, MA) column and an isocratic flow of buffer A. The purity of FRP was confirmed by SDS–PAGE by using a precast gradient gel (Any kD^{TM} Mini-PROTEAN, Bio-Rad, CA). The concentration of FRP was determined by 280 nm absorption on a NanoDrop spectrophotometer (Thermo Scientific, MA, USA).

5.3.3 OCP purification

OCP was isolated from *Synechocystis sp.* PCC 6803 using the procedure of Zhang et al.[55].

5.3.4 Native MS and IM-MS Analysis of FRP

The purified FRP sample was washed with 400 mM ammonium acetate (pH 8.0) in a 10 kDa molecular weight cut off filter (Vivspin, Goettingen, Germany). The original buffer and salts were removed after 10 cycles of washing. The FRP sample was introduced into the ESI source of

a Waters Synapt G2 mass ESI-TOF (Electrospray ionization-quadrupole time of flight, Waters Corporation, Milford, MA) mass spectrometer by using commercial borosilicate emitters with extra coating. (ES387, Hudson, New Hampshire, Thermo Scientific). To investigate the effects of concentration on the oligomeric state of FRP, the concentrations of the introduced FRP were varied. The backing pressure was adjusted to 5 mBar for transferring large protein ions. The sample cone voltage was 20 V. The collisional energy for the trap region was manipulated to observe the dissociation of FRP. For IM experiments, the gas flow rate was 35 mL/min, the ion mobility separation (IMS) wave height was 20 V, and the IMS wave velocity was 500 m/s. The data were output from MassLynx (Waters Corporation, Milford, MA) and plotted by Origin (OriginLab Corporation, Northampton, MA). The IM experiment was calibrated with protein standards (ubiquitin and myoglobin) by using published protocols. The drift-time information from native MS ion-mobility experiments was converted into CCS by considering the molecular weight and charge state of the protein assemblies[184-185].

5.3.5 Native MS of OCP

OCP was analyzed on the same instrument with the same parameters as mentioned above. The samples were washed with 400 mM ammonium acetate solution (pH 8.0) and analyzed at a series of different concentrations. To be consistent with FRP analysis, the pH value of the buffer was adjusted to be 8.0 instead of 6.8 as our previous study[55]. Because the dimer to monomer ratio can vary slightly depending on the instrument conditions; this concentration analysis was conducted in rapid succession.

5.4 Results and discussion

5.4.1 The oligomeric state of FRP

To probe the oligomeric state of FRP, we performed a native MS experiment by using one stock FRP sample diluted in series to concentrations of 3, 5, 10, 40, and 180 µM. (Figure 5.1). Considering the estimated pI (isoelectric point) of FRP is approximately 6.49 from amino acid component analysis[221], we adjusted the pH of NH_4Ac buffer to 8.0 to avoid precipitation.

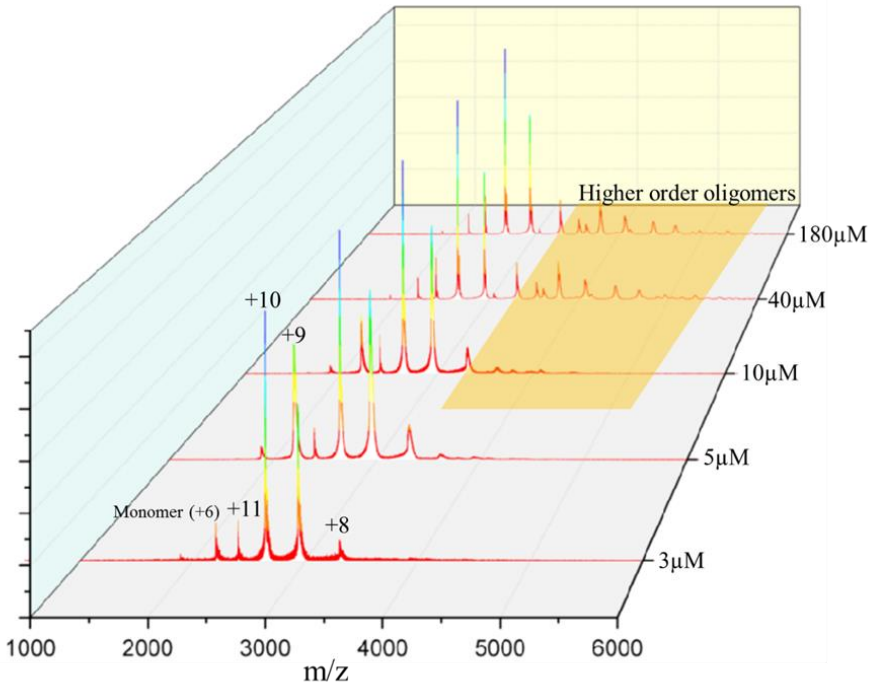


Figure 5.1. Native mass spectrum of FRP diluted in series to concentrations of 3, 5, 10, 40, and 180 µM.

The foundation of native MS is that ionization of protein complexes from aqueous solution, preserves the native structure, at least for the time scale of MS. The sample in the capillary is held at high electric potentials, and droplets containing high charge are drawn out to form the "Taylor cone"[222]. The droplet size is reduced by a potential gradient, aided by the nebulizing gas, until the "Rayleigh limit" is reached[223]. When the protein concentration is high, a significant

number of droplets will contain more than one protein or protein complexes. Adopting low concentrations of proteins for native MS analysis avoids the formation of non-specific adducts. A calculation, based on "18 nm" average droplet size, shows that most droplets are empty and a few contain one protein when the protein concentration is 50 μM[223]. In this study, the lowest concentration we used is 3 μM, as lower concentrations lead to a low signal-to-noise spectrum (data not shown) and a less confident interpretation. A distribution of dFRP carrying different charges was identified at all concentrations. (*m/z* = 2199.65 at +12, *m/z* = 2399.74 at +11, *m/z* =2639.57 at +10, *m/z* = 2932.73 at +9 and *m/z* = 3299.0881 at +8 (experimental MW = 26,385.23 Da). The peak representing the +12 charge state is slightly higher, however, than the +11 charge state of dimeric FRP at 3 μM, inconsistent with a Gaussian distribution of charge states. This suggests the presence of small amounts of monomeric FRP (Figure 5.1) because the peaks representing dFRP carrying +12 charge and monomeric FRP (mFRP) carrying +6 charge overlap. This trace amount of monomer is likely generated during desolvation and transmission to the mass spectrometer, as we could observe no prominent peaks representing mFRP when we decreased the protein concentration. Furthermore, peaks representing mFRP increased when we increased the collisional voltage, confirming that the small amount of mFRP can be generated in the transmission process (Figure 5.2).

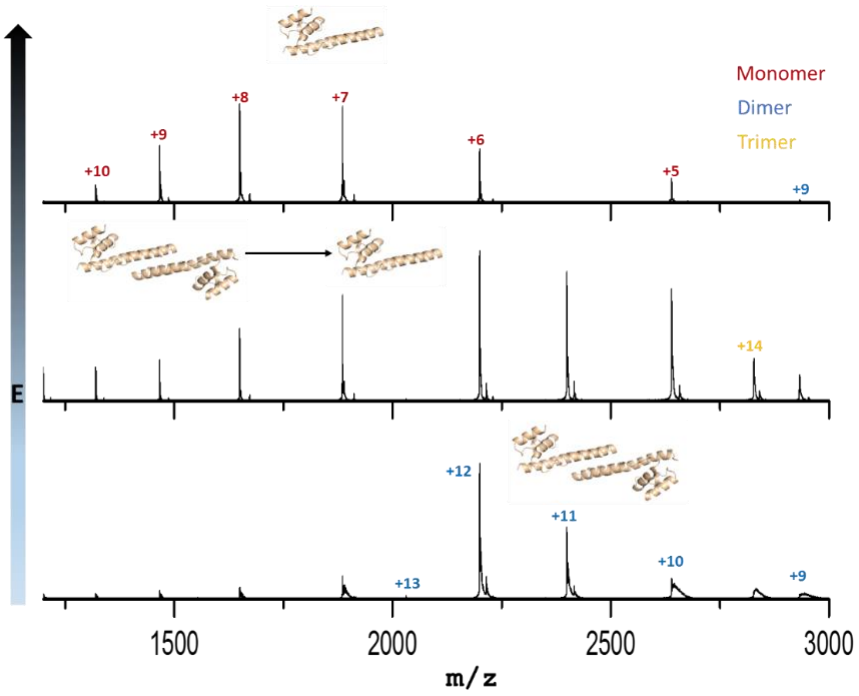


Figure 5.2. Collision induced dissociation of dimeric FRP.

We found that dFRP is consistently the dominant state in native MS at all protein concentrations, and the charge states appear as a Gaussian distribution. Trimeric and tetrameric FRP forms gradually appeared in the spectrum when the concentration increased. Furthermore, higher order oligomers—up to octamers—form at 180 μM (Figure 5.3). Aside from the peaks representing mFRP and dFRP, those for the higher order protein oligomers decease in relative abundance as the oligomer increases in size; tetrameric FRP is no exception. In other words, no special behavior was observed for tetrameric FRP. Although X-ray crystallography favors the tetramer[207], that species is unlikely to be of any biological relevance. Instead, it is concentration-driven and forms more readily at high concentration.

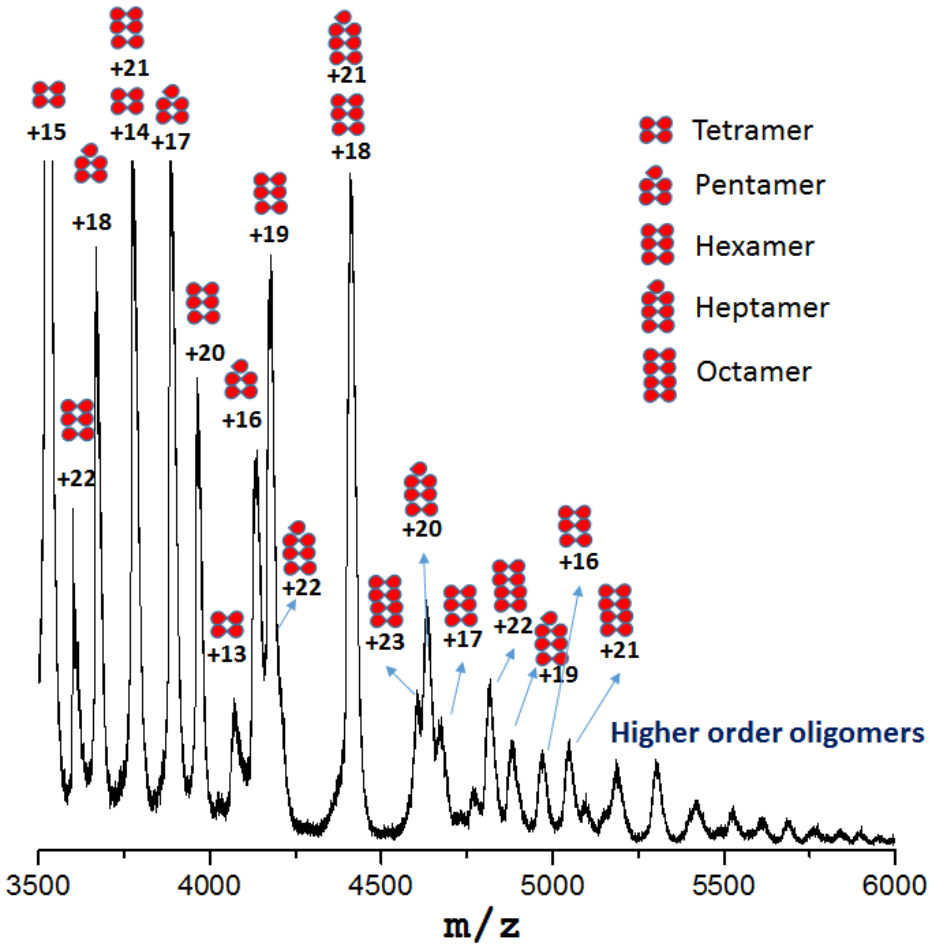


Figure 5.3. High order oligomeric FRP detected at 180 µM.

The MW of the protein complex in the native state, from a calibration curve of SEC, shows FRP exist as a trimer[11]; however, this was later revised to a dimer[207]. Another study on concentration effects on the oligomeric state of FRP suggests partial dissociation of dimeric FRP at low concentration[224]. The discrepancies may be due to the proteins (calibrants or the sample) existing in various conformation, as protein shapes do not necessarily correlation linearly with MW[225]. The FRP crystal structure showed that two conformations exist as dimer and one as a tetramer[207].

5.4.2 Oligomeric states of OCP

Although a dimer is revealed by crystallography, OCP elutes as a monomer by SEC in both its active (red) and inactive (orange) states[38, 205], and the dimerization of OCP may not be biologically relevant, considering the energy of binding and the number of conserved residues.[205] Because determining protein oligomeric states under native conditions can be challenging especially *in vitro* and biases can be introduced by different techniques (e.g., band broadening during SEC) we compared the oligomeric state of OCP at various protein concentrations to that of FRP, as a reference point and follow-up to our previous report[55]. We found Gaussian distributions of mOCP (+9, +10 and +11) and dOCP (+14, +15 and +16), in agreement with our previous report[55]. Moreover, the relative abundance of dOCP to mOCP increased proportionally to the concentration, as shown in Figure 5.4. This result suggests that the dimerization of OCP (also revealed by crystallography) could be the consequence of the high concentrations of protein sample required for protein crystallography[38, 204] and also for some native mass spectrometry experiments[55].

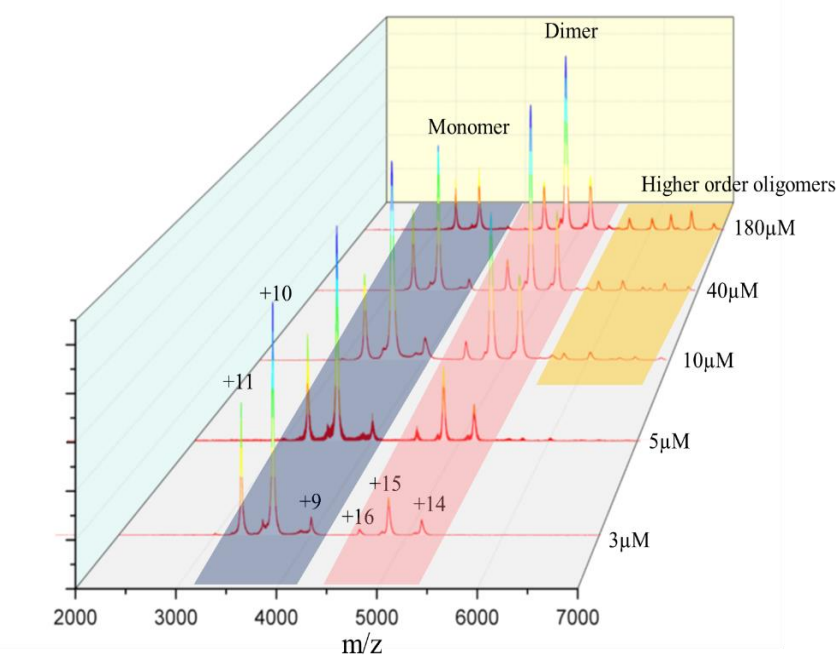


Figure 5.4. Native mass spectrum of OCP diluted in series to concentrations of 3 μM, 5 μM, 10 μM, 40 μM, and 180M.

Thus, OCP has a high tendency to dimerize, depending on the protein concentration *in vitro*, and perhaps *in vivo*. Of note is that OCP can form higher order oligomers just as FRP (Figure 5.5). In our previous report, we found the dimer to monomer ratio is different for the red and orange states of the protein[55]. Whether the functional form of OCP is a monomer or a dimer still remains unclear. In our first report in which we used native MS, we could detect a monomer-to-dimer transition. How this phenomenon is related to physiological function, and how the dimer-monomer transition could benefit biosensor designs are questions for future studies.

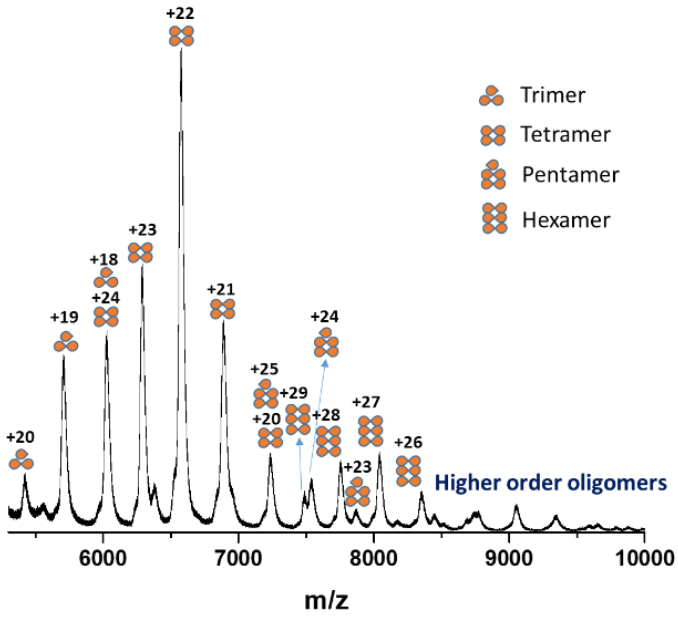


Figure 5.5. High order oligomeric OCP detected at 180 μM.

5.4.3 Ion mobility of FRP

Unfolded or large protein complex ions undergo more collisions with the neutral gas in an IM chamber and, therefore, exhibit a larger CCS than folded or small protein complex ions[36]. To look for folded or unfolded conformers in the gas phase, we calibrated the ion mobility instrument with denatured protein standards according to previous reports (calibration curve is shown in Figure 5.6)[184-185]. We then calculated theoretical CCS values by a projection approximation (PA) method. This method calculates the averaged projected area, ignoring the scattering and long-range interactions between the neutral gas and the ions[226]. Whereas the resulting CCS is usually underestimated by this method[227], an empirically scaled PA, proposed by Ruotolo and Robinson, can correct for this. Its accuracy in predicting protein CCS was previously verified[228-229]. The CCS values for FRP as analyzed by IM-MS in this study, and the

scaled PA values of FRP (PDB 4JDX), along with the measured ones, are listed in Table 5.2. The predicted CCS agree generally with the measured values. It is worth noting that the calculated E/F chain (comprising the tetrameric FRP in PDB 4JDX) dimer is only 0.5% different from the measured one.

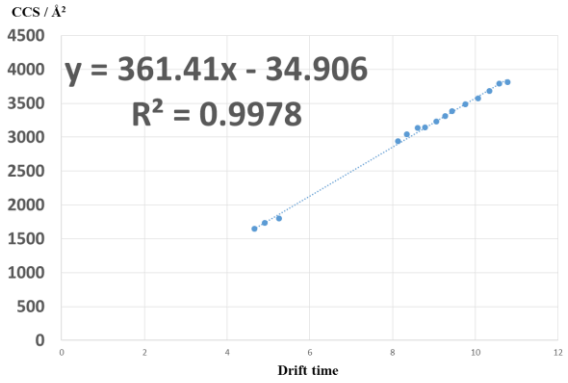


Figure 5.6. Calibration curve CCS vs. New Td.

Table. 5.2. Theoretical (scaled PA) CCS value and experimental CCS value are listed in the table. Experimental CCS value are adopted at 10% peak height.

	Chain	PA/Å²	Difference	Experimental/ Å²
Monomer	A/C	1428	4.3%	1330-1420
	B/D	1433	4.7%	
	E/F	1500	9.5%	
Dimer	AC	2310	4.7%	2170-2240
	BD	2375	7.7%	
	EF	2215	0.5%	
Tetramer	EFEF	3480	-4.2%	3590-3710

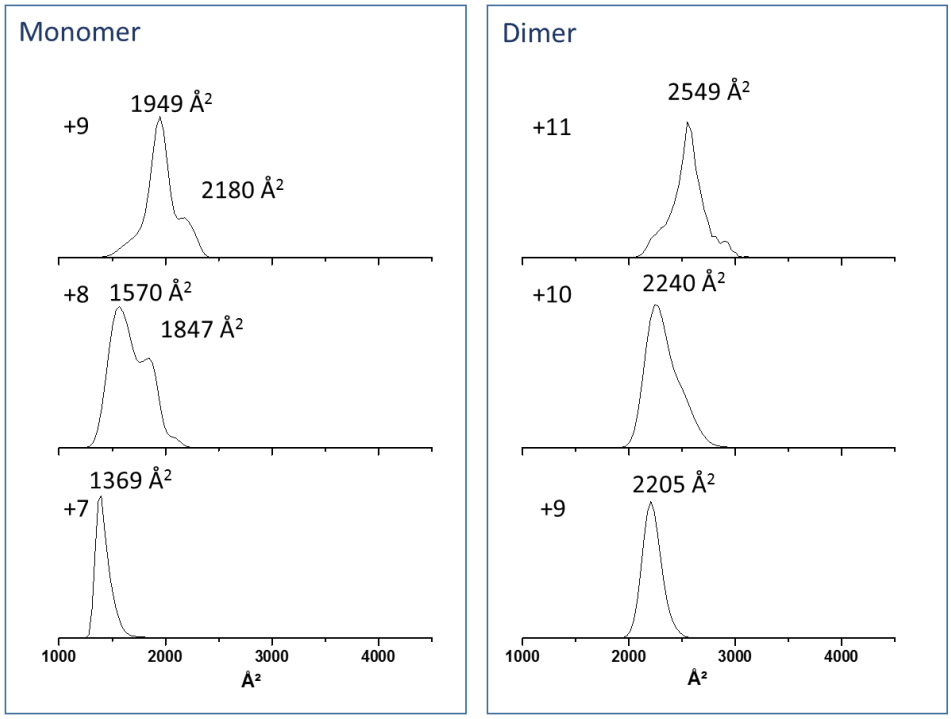
5.4.4 Gas-phase unfolding of FRP under collisional activation

Although the measured CCS agrees well with the theoretical CCS, FRP clearly exhibits an inverse relationship between charge state and CCS, especially for the monomer and dimer (Figure 5.7). When FRP carries extra charges, the protein partially unfolds, and this "enlarged CCS effect" is more obvious for mFRP and dFRP, as a relatively high percentage of domains are affected compared to higher order FRP oligomers. Multiple studies support a linear relationship between charge state and CCS[194-196], which is in agreement with our results.

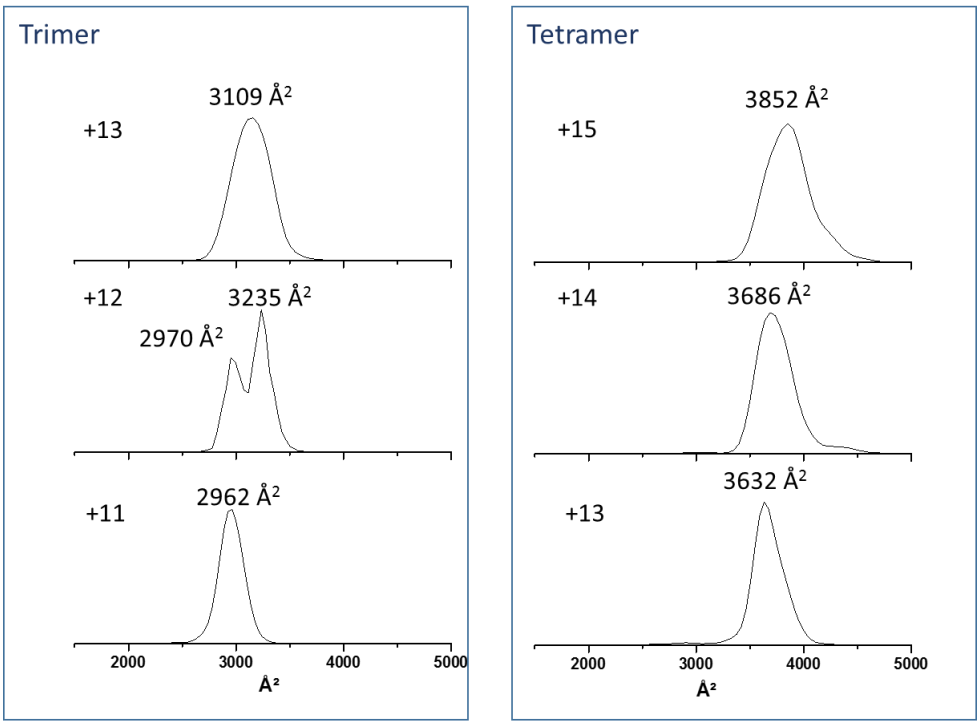
Robinson and co-workers using IM-MS[230] reported experimental evidence of long-lived, unfolded non-covalent complexes. The intermediates in the dissociation pathway can be correlated with increases in CCS indicating that the protein complex undergoes both high order structure refolding and monomer unfolding prior to dissociation. Thus, the unfolding heat map can provide information on the stability and flexibility of protein complex. Here we observed stable unfolding intermediates of all the oligomeric FRP samples in the gas phase of the mass spectrometer (Figure 5.7 and Figure 5.8).

Interestingly, monomeric FRP exhibits a roughly one-stage unfolding process, whereas the dimer undergoes a two-stage process, and the trimer and tetramer a three-stage unfolding process. Usually, the more complex the protein domains are, the more diversified the conformations it can assume before reaching a Coulombic repulsion limit. We propose the one-step unfolding of mFRP to be in the C-terminal region, which is "bent" toward the center chain in the crystal structure. The two-step unfolding process of dFRP then reflects the unfolding of individual FRP C-terminal domains, while maintaining the interface between two subunits. Proteins that fold via a two-state kinetics pathway usually evolve toward a dimer form[231]. This is in accord with our results, as this unfolding process could be viewed as the folding process

performed in reverse. The dramatic change of CCS for different charge states (Figure 5.7) is as large as the intermediates observed during the increasing of collisional voltage (Figure 5.8) and suggests that dFRP retains highly flexible and multiple conformations during the unfolding process.

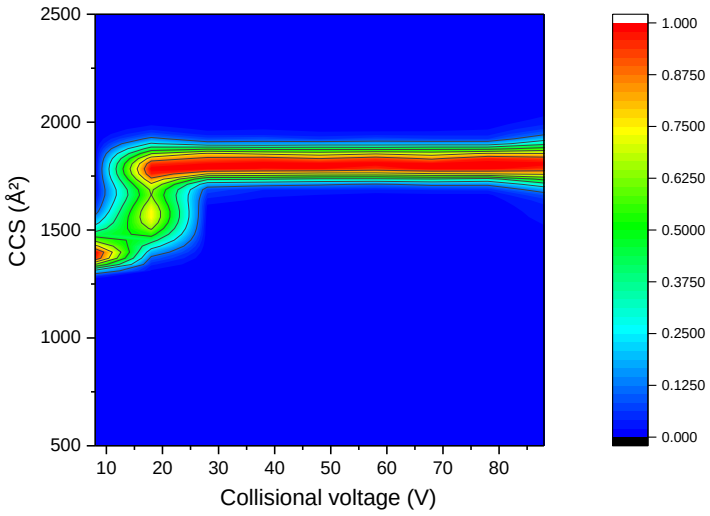


(A)

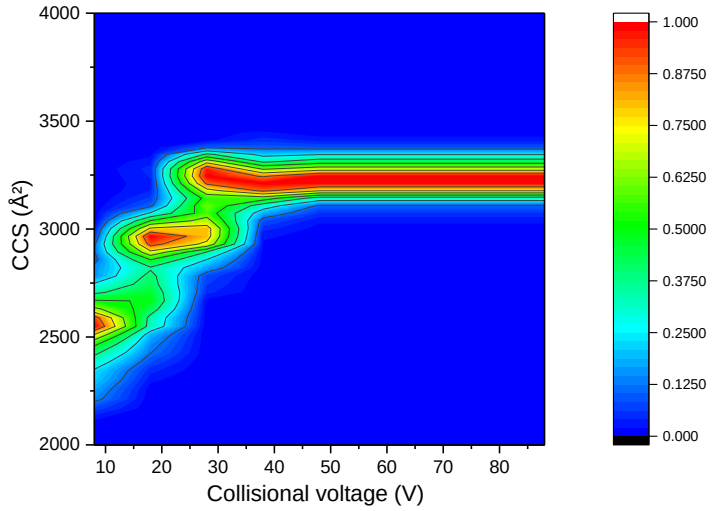


(B)

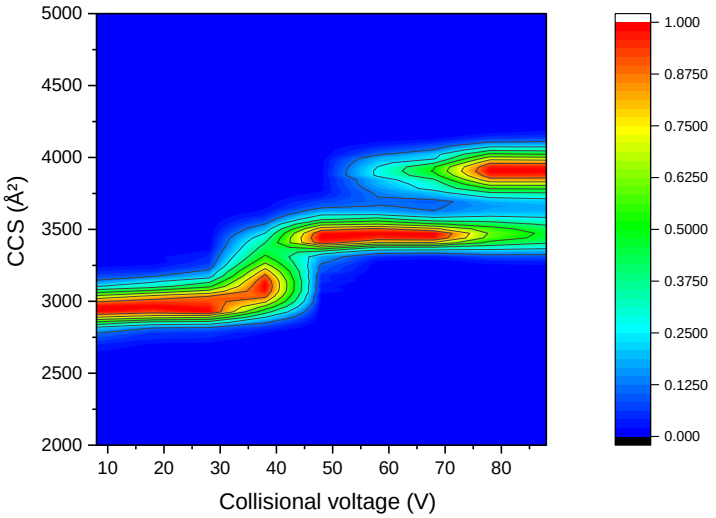
Figure 5.7. CCS of FRP (a) monomer and dimer (b) trimer and tetramer at different charge states.



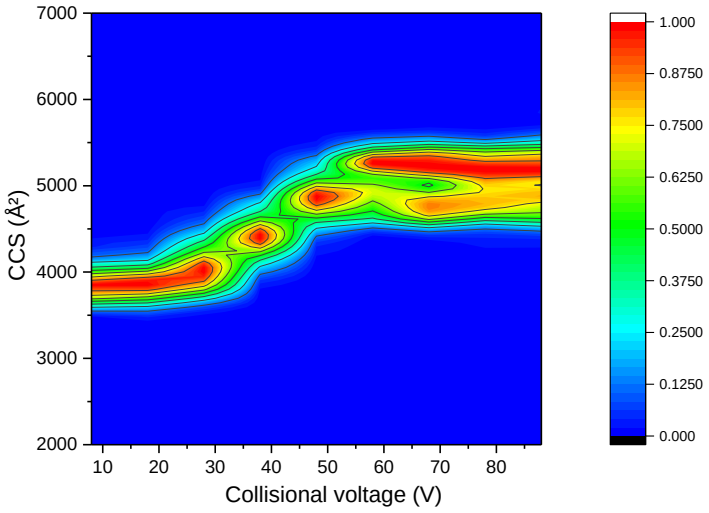
(a)


4000
3500
3000
2500
2000
CCS (Å²)
10
20
30
40
50
60
70
80
Collisional voltage (V)
1.000
0.8750
0.7500
0.6250
0.5000
0.3750
0.2500
0.1250
0.000

(b)


5000
4500
4000
3500
3000
2500
2000
CCS (Å²)
10
20
30
40
50
60
70
80
Collisional voltage (V)
1.000
0.8750
0.7500
0.6250
0.5000
0.3750
0.2500
0.1250
0.000

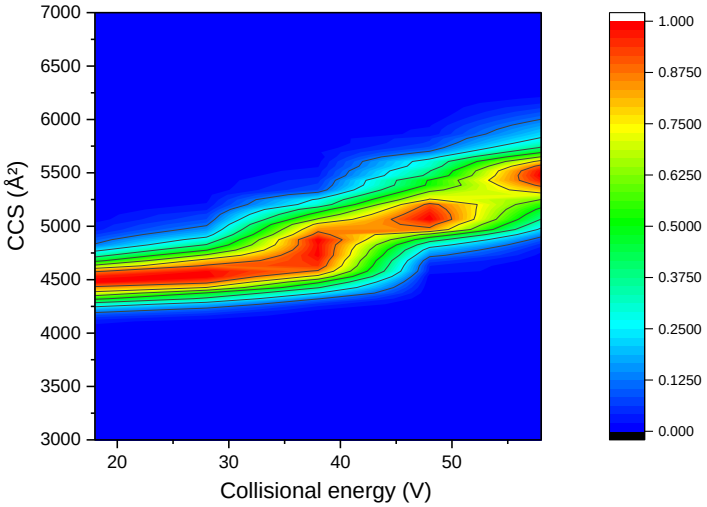
(c)



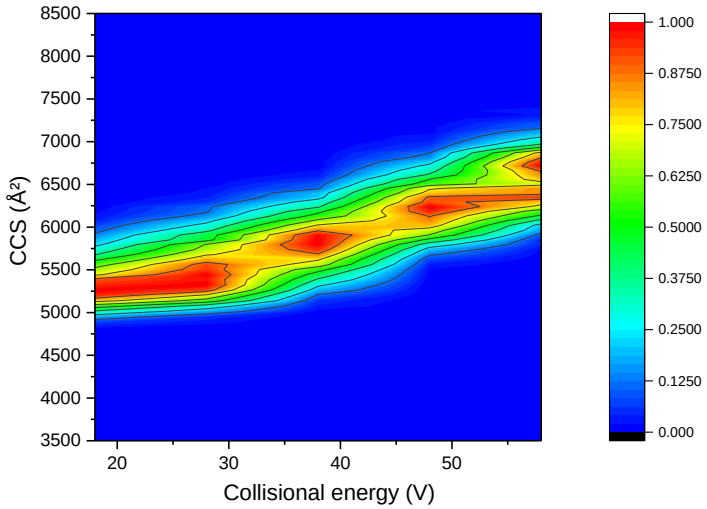
(d)

Figure 5.8.Unfolding heat map of FRP (a) monomer at +4 charge state, (b) dimer at +11 charge state, (c) trimer at +13 charge state, (d) tetramer at +15 charge state.

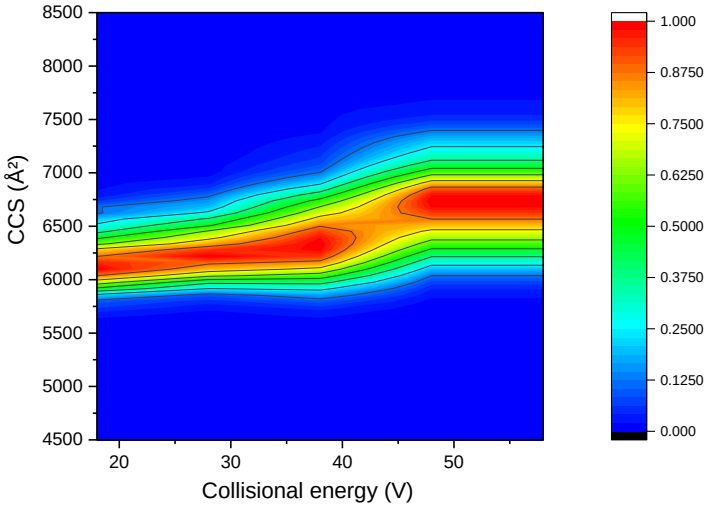
The unfolding heat maps of pentameric, hexameric, heptameric, and octameric FRP (Figure 5.9) show that as the complexity of interacting domains is increased, fewer intermediate states exist. Specifically, heptameric and octameric FRP exhibit rather rigid structures in the gas phase. Dimeric state FRP was proposed to be the functional state[207]; thus, any form of higher order FRP in the native environment, if present *in vivo*, would require some structural flexibility to dissociate into the active dimer. Thus, these higher order FRP oligomers cannot re-dissociate into dimeric FRP owing to the rigid structure. This reinforces our proposal that the higher order oligomers of FRP are artifacts of a high protein concentration.


7000
6500
6000
5500
5000
4500
4000
3500
3000
CCS (Å²)
20
30
40
50
Collisional energy (V)
1.000
0.8750
0.7500
0.6250
0.5000
0.3750
0.2500
0.1250
0.000

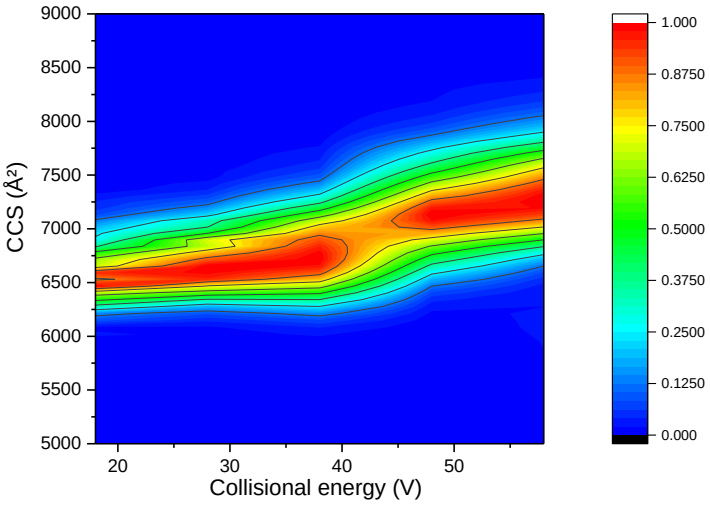
(a)


8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
CCS (Å²)
20
30
40
50
Collisional energy (V)
1.000
0.8750
0.7500
0.6250
0.5000
0.3750
0.2500
0.1250
0.000

(b)



(c)



(d)

Figure 5.9. Unfolding heat map of FRP (a) pentamer at +16 charge state, (b) hexamer at +19 charge state, (c) heptamer at +21 charge state, (d) octamer at +21 charge state

5.5 Conclusions

Protein oligomerization plays important roles in their conformation, function and stability. In this report, we analyzed the structure and oligomeric state of two proteins, OCP and FRP, which are crucial participants in the cyanobacterial photoprotection cycle. FRP is a non-chromophore protein, but its oligomeric state and active form remained largely enigmatic until a crystal structure was determined. Up to three conformations, however, exist in the crystal structure: two separate dimers and a tetramer. In the meantime, the native state of OCP, whether monomeric or dimeric, has also been controversial.

In this study, we used native MS to characterize FRP and OCP and found FRP to be predominantly dimeric independent of the protein concentration, although a small fraction of higher order oligomers forms at higher protein concentration. We analyzed the oligomeric state of OCP in the same manner and observed, in contrast to FRP, that OCP exists as both monomer and dimer, and the relative abundance of dOCP increases with protein concentration. Moreover, the unfolding heat map revealed by IM-MS shows a one-step unfolding process for mFRP, a two-step for dFRP and a three-step for trimeric and tetrameric FRP. The high order FRP oligomers retain a rather rigid structure, strongly suggesting they are of little relevance *in vivo*. This study offers new insights into the biological assemblies of FRP and OCP, especially in terms of their oligomeric states, providing the groundwork for future structure-function analyses of this important photoprotection mechanism.

Chapter 6: A Molecular Mechanism for Non-Photochemical Quenching in Cyanobacteria

Manuscript in preparation:

Lu, Y., Liu, H., Saer, R., Zhang, H., Gross, M., and Blankenship, R. A molecular mechanism for non-photochemical quenching in cyanobacteria. Manuscript submitted.

6.1 Abstract

The cyanobacterial Orange Carotenoid Protein (OCP) protects photosynthetic cyanobacteria from photodamage by dissipating excess excitation energy collected by phycobilisomes (PBS) as heat. Dissociation of the PBS-OCP complex *in vivo* is facilitated by another protein known as the Fluorescence Recovery Protein (FRP), which primarily exists as a dimeric complex. We used a range of mass spectrometry-based techniques to investigate the molecular mechanism of this FRP-mediated process. FRP in the dimeric state (dFRP) retains high affinity to the C-terminal domain (CTD) of OCP in the red state (OCP^r). The site-directed mutagenesis and native MS results suggest the head region on FRP could be a binding candidate to OCP. After attachment to CTD, the conformational changes of dFRP enable dFRP to bridge the two domains together, which facilitates the reversion of OCP^r into the orange state (OCP^o) accompanied with the

structural rearrangement of dFRP. Interestingly, we found a mutual response between FRP and OCP: FRP and OCP^r destabilize each other, whereas FRP and OCP^o stabilize each other. A detailed mechanism of FRP function is proposed based on the experimental results.

6.2 Introduction

Solar energy is utilized by photosynthetic organisms to perform photosynthesis, a process that produces and stores chemical energy, later to be used to power cellular processes. Excess light energy is detrimental, however, to photosynthetic organism, resulting in oxidative stress, and ultimately leading to the damage of photosynthetic apparatus or even the death of the cell. Light regulation strategies, therefore, are necessary for balancing the absorption and utilization of light energy. One of the most important protection mechanisms is called non-photochemical quenching (NPQ), during which excited-state chlorophylls or other pigments are quenched, and excess excitation energy is dissipated as heat[232]. In plants and algae, NPQ is carried out by light-harvesting antennas, a process mediated by a ΔpH across the thylakoid membrane. NPQ is also related to the xanthophyll cycle, which plays an important role in the protection of plants and algae against oxidative stress[232-233]. In cyanobacteria, the major light-harvesting complex is the phycobilisome (PBS), a soluble complex attached to the surface of the membrane, unlike the integral membrane light-harvesting complexes in plants and algae. Thus, a distinct photoprotection mechanism has evolved in cyanobacteria. The orange carotenoid protein (OCP), a single carotenoid-binding protein, functions both as a light sensor and a photoprotective entity[9, 202, 234].

The OCP is composed of an α-helical N-terminal domain (NTD), an α-helix/β-sheet C-terminal domain (CTD), and a flexible linker region that joins them. A keto-carotenoid spans both

domains and is encapsulated by the protein scaffold with almost no solvent exposure. The OCP forms a compact globular structure by a strong interaction between the NTD and CTD via salt bridges as well as by hydrogen bonds. Under high-intensity illumination, these bonds are broken, which leads to the solvent exposure of the major NTD-CTD interface, accompanied by a 12 Å carotenoid translocation; this process converts orange OCP (OCP^o) into its active quenching (red) state (OCP^r) [7, 38, 204-205, 235-237]. It is generally accepted that OCP^r quenches energy and fluorescence via its interaction with the core of the PBS. Efforts have been made to locate the specific binding site on PBS, and several models have been proposed[1, 55, 238-240]. FRP is able to recover the fluorescence from the PBS by interacting with OCP^r. The active OCP^r is metastable, and reverts back to the inactive orange state in the dark. FRP greatly accelerates this reversion process by interacting with the CTD[206-207]. The FRP crystal structure reveals three conformational states: two in the dimeric form and one in the tetrameric form. The dimeric form was proposed to be the functional state[207], and it was found to be the dominant state in solution[56, 207, 224]. The study on OCP apoprotein suggests FRP functions as a general scaffold protein for OCP maturation[237]. The recent model proposed by Thurotte et al.[241] demonstrates that FRP has two distinct activities during the fluorescence recovery process: it first accelerates OCP^r detachment from phycobilisomes[242] and then assist OCP^r relaxation into OCP^o. The detailed molecular mechanism of how FRP mediates the fluorescence recovery process in cyanobacteria is still elusive.

We have chosen mass spectrometry (MS) to investigate this problem. MS is now being widely used to study protein conformation, structure, dynamics, and protein-protein or protein-ligand interactions[43, 209, 243-245]. Native MS, in particular, allows for the detection and analysis of intact protein complexes in their near-native states[53, 246]. Cross-linking complements native MS

because it can reveal amino acid pairs that are positioned in close proximity by linking protein partners constrained by the effect length of the linker[40]. A medium-resolution interaction model can be obtained by mapping the adjacent residues on protein complexes[43]. MS-based isotopic cross-linking can provide quantitative information on protein interactions, giving insights into the binding sites as well as conformation changes[41]. MS-based protein footprinting techniques can further provide detailed information on protein solvent accessibility[13, 29, 34, 247]. All in all, an integrative MS-based tool kit can provide complementary information of protein structures in complexes to afford a more complete description of interaction model.

In this work, we utilized a variety of techniques, especially MS-based ones, to investigate the molecular mechanism of the FRP-mediated OCP^r to OCP^o conversion process. We found that FRP accelerates OCP^r to OCP^o conversion through several distinct steps. A dramatic conformational change of FRP occurs upon interacting with OCP in both the red and orange states. In addition, the head domain on FRP could play an essential role during its binding to OCP. Our study provides novel insights into the FRP-mediated OCP^r to OCP^o process, and a working model is proposed based on the experimental results.

6.3 Materials and Methods

6.3.1 Expression and mutagenesis of FRP

The expression of WT FRP (SGL_RS10235) was performed as previously reported[56]. Site-directed mutants of FRP were constructed by using complementary mutagenic PCR primers. A list of all the primers used is shown in Table S1. Each mutagenic PCR reaction consisted of ~50 ng of plasmid DNA, 10 pmol of each primer, 10 nmol of dNTPs, 0.5 μL of a high-fidelity DNA

polymerase (Phusion, Thermo Fisher Scientific, Waltham MA), and 10 µL of a 5x PCR reaction buffer (Phusion HF buffer, Thermo Fisher Scientific) in a 50 µL reaction. The PCR reaction consisted of an initial denaturation step at 98 °C for 15 sec, followed by 18 cycles of 98 °C for 15 sec, 50 °C for 20 sec, and 72 °C for 20 sec. Because the length of the SGL_RS10235 gene is short, a final extension step was not included in the PCR reaction. Following the mutagenic PCR reaction, 1 µL of DpnI restriction enzyme (New England Biolabs, Ipswich, MA) was added to each reaction in order to digest the original template DNA. Five microliters of this reaction were used to transform chemically competent *E. coli* DH10B cells. Plasmids extracted from the transformants were confirmed by DNA sequencing prior to a subsequent transformation of the closed circular plasmid into *E. Coli* BL21(DE3) cells for protein expression.

Table. 6.1

Forward primers used to construct site-directed mutants of the SGL_RS10235 gene (reverse primers are the reverse complements of the forward primers)

Mutant	Forward primer sequence (5'-3')
R60L	GAAACTCCATGATTTTTTGAGTGCAAAACTGCACGAAATTGATGGC AAGTACG
D64A	GTGCAAAACGCCACGAAATTGCCGGCAAGTACGACGATCGCC
G65D	GTGCAAAACGCCACGAAATTGATGATAAGTACGACGATCGCCAGTC
R70D	GAAATTGATGGCAAGTACGACGATGATCAGTCGGTGATTATTTTTG TTTTTGC
F76D	CGATCGCCAGTCGGTGATTATTGATGTTTTTGCCCAACTGCTCAAGG
K102D	TAGCCGCCGATAAGCAATCTGATATTAAGGCCCTGGCCCGG

6.3.2 FRP and OCP purification

The isolation of OCP, FRP and FRP mutants were performed by using published protocols[55-56]. The OCP^r partial digestion was carried out as previously described[248].

6.3.3 Native MS and IM-MS Analysis of NTD, CTD and FRP complex

The NTD/CTD mixture obtained after OCP^r partial digestion and FRP samples were separately washed with 400 mM ammonium acetate at pH = 8.0 (09689, Sigma-Aldrich, Missouri, USA) in a 5 kDa molecular weight cut off filter (Vivaspin, Goettingen, Germany). The original buffer and salts were removed by10 cycles of washing. The NTD/CTD and FRP were mixed in 4:1 and 1:4 ratios, respectively, and introduced into the ESI source of a Waters Synapt G2 ESI Q-TOF (Electrospray ionization-quadrupole time of flight, Waters Corporation, Milford, MA) mass spectrometer by using commercial borosilicate emitters with extra coating (ES387, Hudson, New Hampshire, Thermo Scientific). The backing pressure was adjusted to 5 mBar for transferring the large protein ions. All the mutant FRP proteins were mixed with the NTD/CTD in a 2:1 ratio, respectively, and analyzed in the same manner. As a reference, intact OCP was also mixed with FRP in a 1:2 ratio to investigate the affinity of FRP to OCP^o. The IM-MS experiment and data processing were carried out as previously described[56].

6.3.4 Activity assays

The OCP was previously photoconverted to the red form by 10 min illumination with 2,000 μmol photons $m^{-2}s^{-1}$ white light at 8 °C. The OCP^r to OCP^o reversion processes in the absence/presence of FRP WT/mutants were monitored in a Lambda 950 (Perkin Elmer UV WinLab) spectrophotometer at 8 °C. One point was recorded per second by monitoring the absorption at 550 nm for 30 min.

6.3.5 Cross-linking and LC-MS

All the cross-linking experiments were carried out in triplicate at 4 °C for 2 h. Illumination with 2,000 μmol photons $m^{-2}s^{-1}$ white light was kept constant during the cross-linking process for OCP^r-included samples. The cross-linker, disuccinimidyl suberate (DSS-d_0, Cross-linking reagent without any deuterium atoms, ProteoChem, UT, USA), was added to FRP (reference 1), OCP^o and OCP^r-FRP samples. DSS-d_4 (Cross-linking reagent with 4 deuterium atoms that provide a 4 Dalton mas shift, ProteoChem, UT, USA) was added to FRP (reference 2), OCP^r and OCP^o-FRP samples. OCP^r-FRP^{d4} and OCP^o-FRP^{d0}; FRP^{d4}, $OCP^{o,\,d4}$ and OCP^o-FRP^{d0}; FRP^{d0}, $OCP^{r,\,d0}$ and OCP^r-FRP^{d4} were mixed, respectively, in equimolar quantities to investigate the structural changes and interacting regions of these proteins. (See Figure 6.1 for a flow chart). The molecular weight of cross-linked complexes was analyzed by SDS-PAGE. Peptides from the digest of cross-linked samples were prepared by acetone precipitation and enzymatic digestion as previously described[249]. Sep Pak cartridges (Waters Corporation, Milford, MA) were used to desalt the sample. An LC-MS experiment was done, as previously described, with some adjustments[71]. Peptide mixtures were trapped by a guard column (nanoACQUITY Trap Column, Waters Corporation, Milford, MA) and then fractionated on an ACQUITY UPLC Peptide BEH C18 Column (10 K psi, 130 Å, 1.7 μm, 75 μm X 100 mm, Waters Corporation, Milford, MA). The MS analysis was performed with a Thermo Scientific™ Q Exactive™ hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen Germany). Peptides were eluted with a 120 min, 250 nL/min gradient coupled to the nanospray source. The default charge state was 3, and the scan range was from *m/z* 380-1500. Mass spectra were obtained at high mass resolving power (70,000, FWHM at *m/z* 200), and the top 15 most abundant ions corresponding to eluting peptides per scan were submitted to collision-induced dissociation (CID) in the ion

trap, with charge-state rejection of unassigned, +1, +2 and >8 ions enabled. Precursor ions were added to a dynamic exclusion list for 8 s to ensure a good sampling of each elution peak.

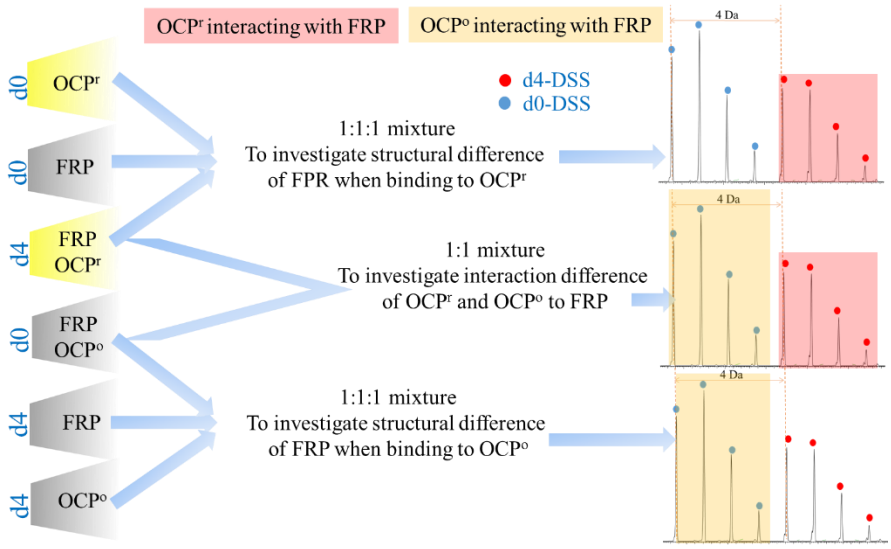


Figure 6.1. Flow chart of isotopic cross-linking experiments on FRP and OCP.

6.3.6 GEE Labeling

The FRP protein was mixed with OCP^r and OCP^o respectively in a 2:1 ratio. The modification reaction was carried out for a time course up to 2 h under either dark or light conditions at 4 °C, using freshly prepared 1.5 M GEE (Sigma, St. Louis, MO) and 0.5 M EDC (Pierce, Rockford, IL) stock solutions. (10 mM PBS, pH 8.0) The reaction was quenched by adding a 1/10 volume of 1 M Tris–HCl (pH 8.0) followed by buffer-exchange using a Zeba™ desalting spin column (Thermo Scientific, Rockford, IL) according to the manufacture's protocol. The FRP-only sample was also labeled by GEE on the same platform as the control. Preparing the peptides and conducting the LC/MS experiments were performed in the same way as for the cross-linked samples[39].

6.3.7 MS data analysis

Byonics[250] and Protein Prospector online server (Baker, P.R. and Clauser, K.R. http://prospector.ucsf.edu) were utilized to identify DSS- and GEE-labeled peptides. The searching parameters were set as follow: Peptide tolerance: 10 ppm, MS/MS tolerance: 0.02 Da, Mass type: Monoisotopic, ^{13}C isotope ions: Yes, Enzyme: Trypsin, Missed cleavages: 2.

6.4 Results and Discussion

6.4.1 FPR accelerates the OCP^r to OCP^o relaxation by bridging NTD and CTD

The OCP is a unique protein that functions as a light sensor, a signal propagator, and an energy quencher. The NTD is a chromophore-containing domain that can burrow into the PBS and thermally dissipate excess excitation energy in the PBS, whereas the CTD can regulate the accessibility of OCP, and hence the activity of the PBS-binding NTD[240, 251-252]. The spontaneous reversion from the red to orange state of OCP is significantly accelerated in the presence of FRP[11, 207, 242]. Under light irradiation, the most exposed cleavage sites for trypsin are located on the linker region of OCP owing to complete domain dissociation[38]. Thus, fragments of the CTD (186-310) and NTD (10-170) can be obtained by partial proteolysis, whereas parts of the linker region and N-terminal arm (flexible loop region located on the N- termini) are missing[248].

Immunoprecipitation, size-exclusion chromatography (SEC), native-gel electrophoresis, and molecular modelling results all suggest that the CTD is the domain that interacts with FRP[207, 239, 253]. Our previous native MS study shows that FRP primarily exists as a dimer[56], although two different oligomeric states were identified by crystallography[207]. In this study, FRP was mixed

with NTD/CTD partial digestion fragments in 4:1 and 1:4 ratios and subjected to native MS analysis. The electrospray ionized protein complex in the gas phase carries a series of charges which usually exhibit a Gaussian distribution. All the peak assignment was determined both manually and by Massign software[183]. The results indicate that a high abundance of the dFRP-CTD complex is formed (Figure 6.2), which is convincing evidence of the high affinity of the CTD to dFRP. Collisional induced dissociation (CID) of this protein complex produces monomeric FRP-CTD (mFRP), suggesting a strong interaction between one FRP monomer and the CTD (Figure 6.3A). A previous ion mobility (IM)-MS analysis shows one intermediate state in the unfolding process of dFRP[56], and none in the unfolding process of the CTD owing to its compact structure[248]. Here, two intermediate states occur during the unfolding process of CTD-dFRP, driven by the higher order structure refolding, suggesting high stability of this complex (Figure 6.4).

Surprisingly, protein complexes containing components of the NTD, dFRP, and CTD were found to co-exist with the CTD-dFRP complex. NTD fragments (sequences from 10-168 and 10-170 with and without carotenoid, respectively) are binding partners to the dFRP-CTD (Figure 6.2A). We then carried out tandem MS to investigate the topology of those protein complexes. When protein complex ions are accelerated to high kinetic energy it usually results in ejection of a single protein subunit[254]. After those ejection events, complexes including the dFRP-NTD and mFRP-NTD remain, giving evidence for the existence of a binding face between FRP and the NTD (Figure 6.3B). Interestingly, no such complexes were detected during an MS1 analysis, when no MS/MS activation was applied. The previous study on FRP and NTD mixture also shows FRP doesn't bind to NTD[253]. It appears that the binding of the CTD to dFRP initiates a conformational change of dFRP, facilitating its binding to the NTD. In another word, the

determining step of OCP^r binding to dFRP event is the attachment of CTD to dFRP. An interface between CTD and NTD might also exist in protein complex CTD-dFRP-NTD, but the interaction is likely to be weak since no NTD-CTD complex can be observed with or without CID.

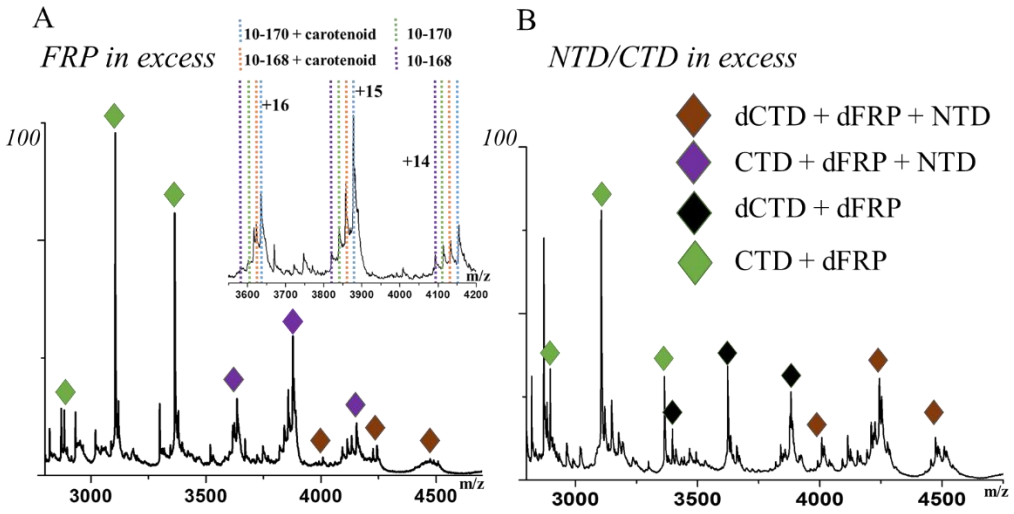


Figure 6.2. Native mass spectra of FRP and the NTD/CTD mixture in a (A) 4:1 or (B) 1:4 ratio. Complexes include the dCTD-dFRP-NTD, CTD-dFRP-NTD, dCTD-dFRP and CTD-dFRP. The inset in A shows the binding of the NTD fragments to the dFRP-CTD.

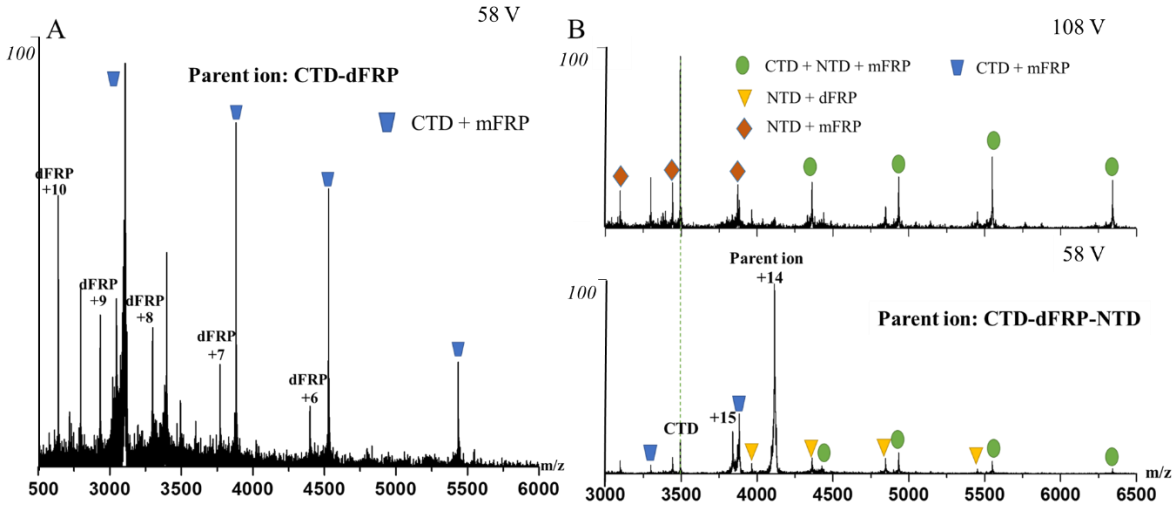


Figure 6.3. (A) Product-ion (MS/MS) spectrum of the CTD-dFRP ion obtained at 58 V collisional voltage. The resolved complex of the CTD-mFRP suggests a primary binding face exists on one of the FRP subunits. (B) Product-ion (MS/MS) spectra of CTD-dFRP-NTD recorded at 58 V and 108 V collisional voltage. The resolved complexes of NTD-dFRP and NTD-mFRP after CID suggest that a binding interface exists between the NTD and FRP.

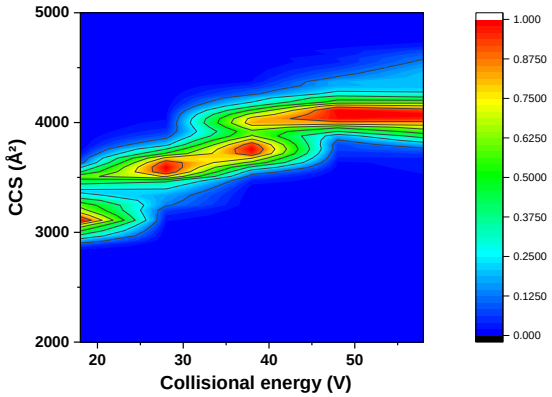


Figure 6.4. IM-MS unfolding heat map of CTD-dFRP complex.

We also observed the formation of complexes consisting of the dCTD-dFRP-NTD and dCTD-dFRP (Figure 6.2). To investigate whether the CTD can form a dimer by itself, the NTD/CTD mixture was submitted to native MS at concentrations ranging from 5 µM to 200 µM. The analysis revealed a noticeable level of dimeric CTD (dCTD), even at low concentrations (Figure 6.5), which is in accord with the previous SEC results[253]. Thus, a complex of dCTD-dFRP is formed owing to the binding of dCTD to dFRP. These results suggest that different regions of the CTD are involved in binding to the other CTD and FRP. Similarly, the NTD could bind to this dCTD-dFRP complex forming dCTD-dFRP-NTD. When the NTD/CTD is in excess, more dCTD-dFRP-NTD complexes are formed (Figure 6.2B).

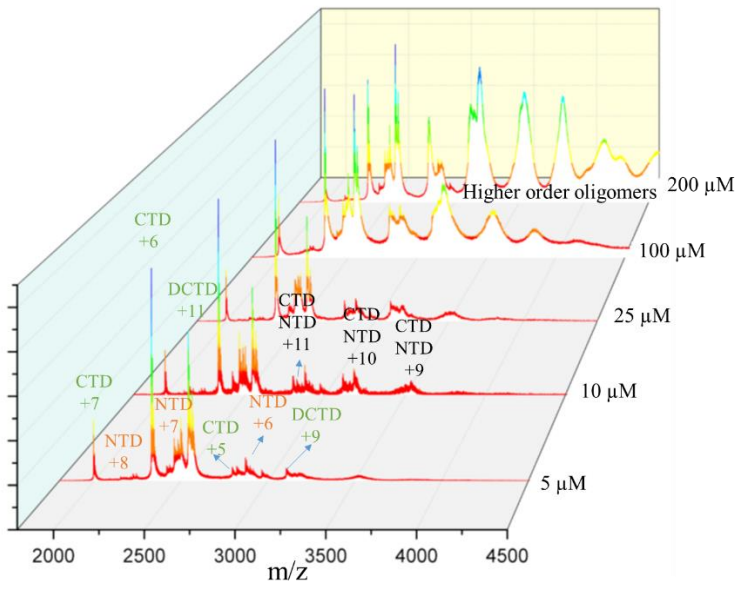


Figure 6.5. Native MS analysis of NTD/CTD mixture ranging from 5 µM to 200 µM.

To describe better the interaction between FRP and OCP, an isotopic cross-linking strategy was adopted to compare quantitatively their interactions under dark or illuminated conditions[41]. SDS-PAGE analysis revealed bands corresponding to the molecular weights of one FRP and one OCP, and two FRPs and one OCP in the red state, while the corresponding bands in the orange state were barely visible (Figure 6.6). It appears that mFRP-OCP is more abundant than dFRP-OCP on the gel image, but we don't see that as a direct evidence of dFRP monomerization. Usually only a very small fraction of interacting protein complex can be cross-linked. In order to observe dFRP-OCP, mFRP needs to be cross-linked to the other mFRP and OCP needs to be cross-linked to mFRP which lower the chance of obtaining them. Nevertheless, among all the lysine-NH_2 groups in OCP (twelve) and FRP (eight), and the N-termini of the two proteins, we identified six cross-links by LC/MS out of the 117 potential cross-links (Figure 6.7A). A residue located at the CTD (K249) was found to be linked to K23 on FRP, further proving the adjacency of CTD to FRP. The linker region and the N-terminal arm of OCP were also found to be cross-linked to FRP (Figure 6.7A, product-ion (MS/MS) spectra are shown in Figure 6.8). Mapping the

cross-linked residues on the FRP and OCP crystal structures, we notice that those residues are located near the interface of two terminal domains (Figure 6.7B, C, D).

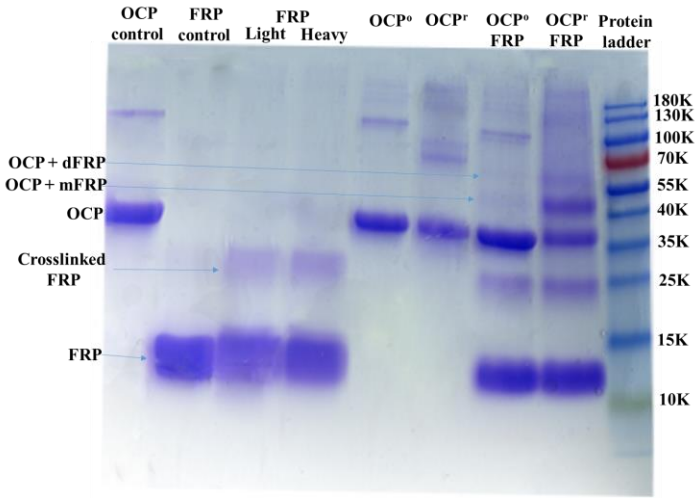


Figure 6.6. SDS-PAGE analysis of cross-linked samples. The content of sample is labeled on top of each lane.

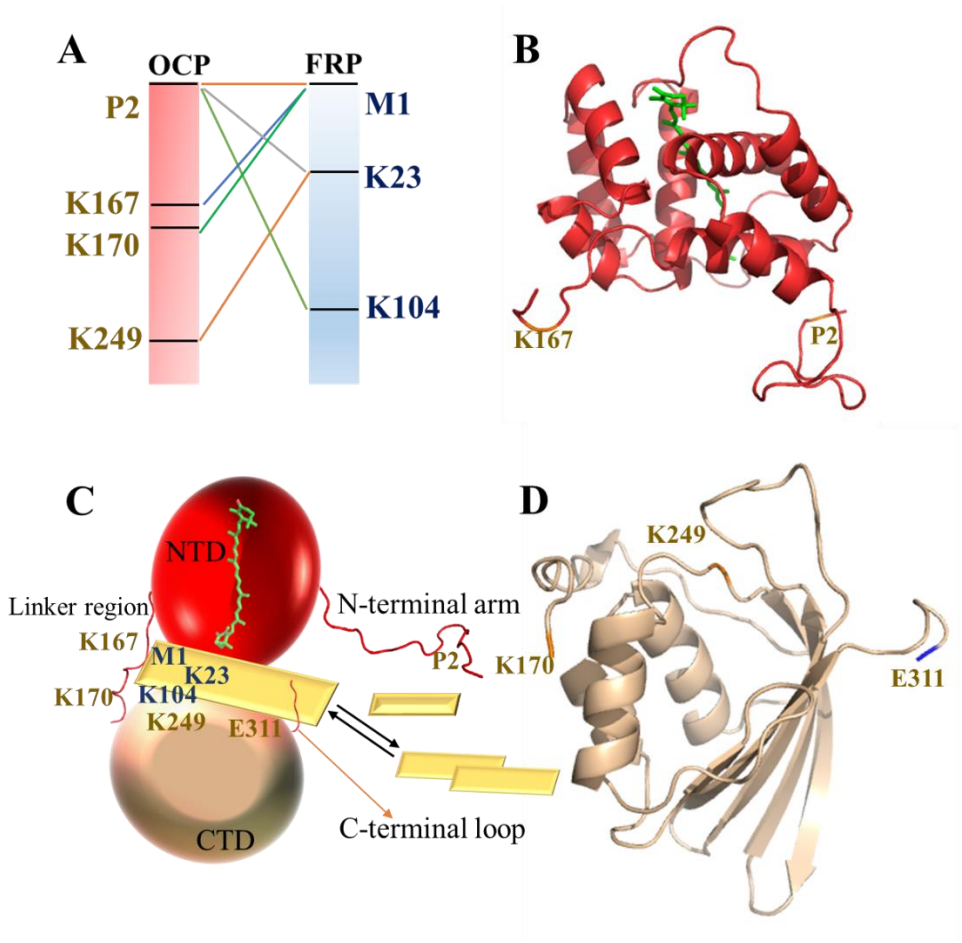
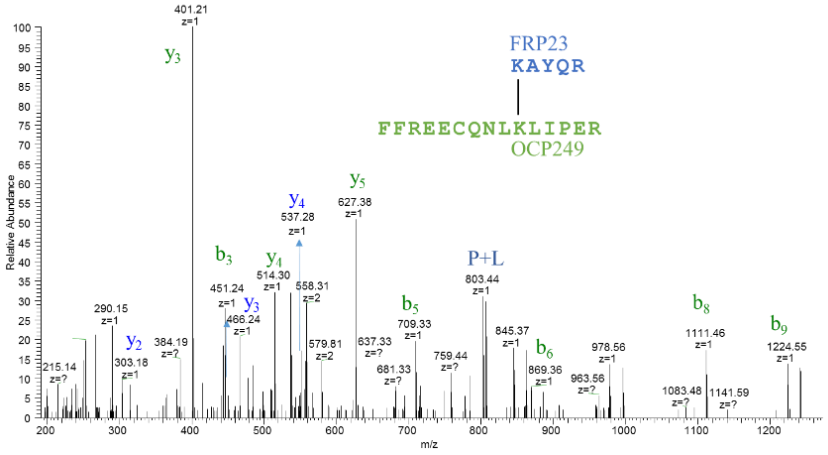
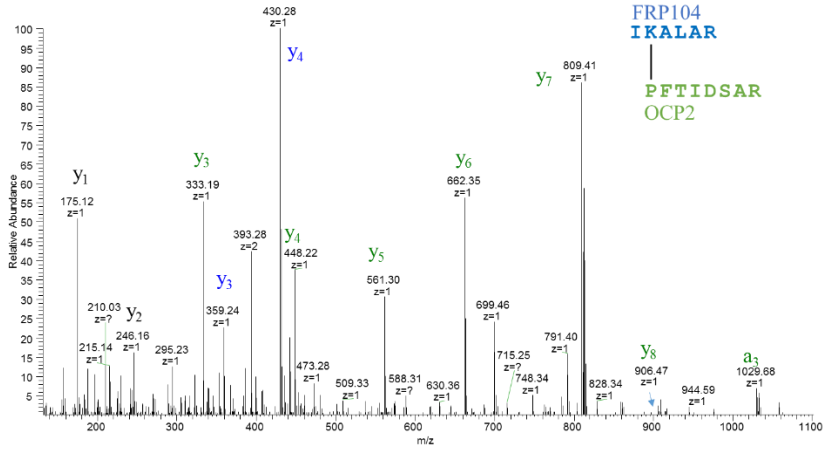


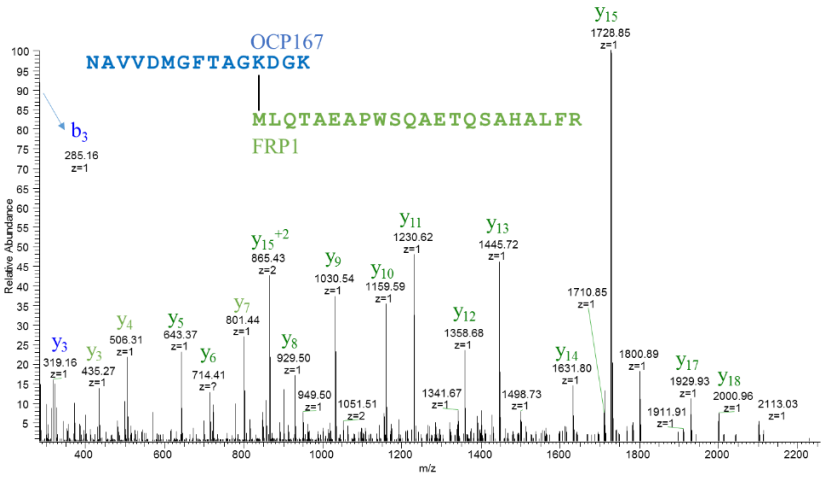
Figure 6.7. (A) Cross-links identified between OCP^{r} and FRP. The corresponding positions are mapped on the cartoon representations of the OCP NTD (B, PDB 4XB5[235]) and CTD (D, PDB 3MG1[205], sequence 170-311) crystal structures. (C) Model of the OCP-FRP interaction from crosslinking data. The missing loop regions in the crystal structure were generated by software available at UCSF[255]. Each mFRP unit is shown by yellow rhombus.


FRP23
KAYQR
FFREECQNLKLIPER
OCP249
P+L
Relative Abundance
m/z

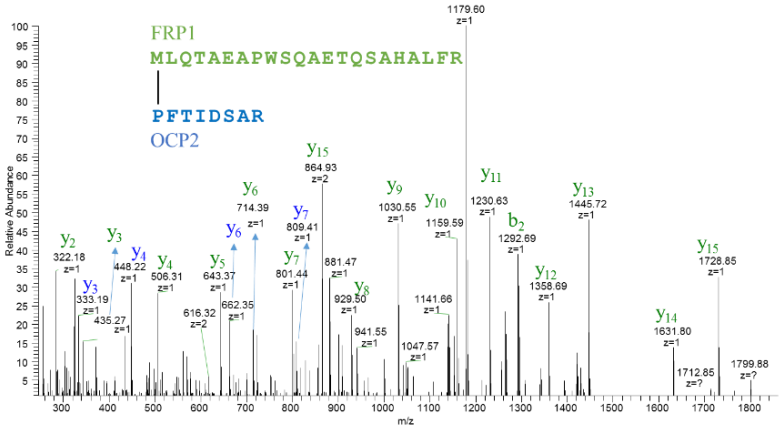
(A)


FRP104
IKALAR
PFTIDSAR
OCP2
Relative Abundance
m/z

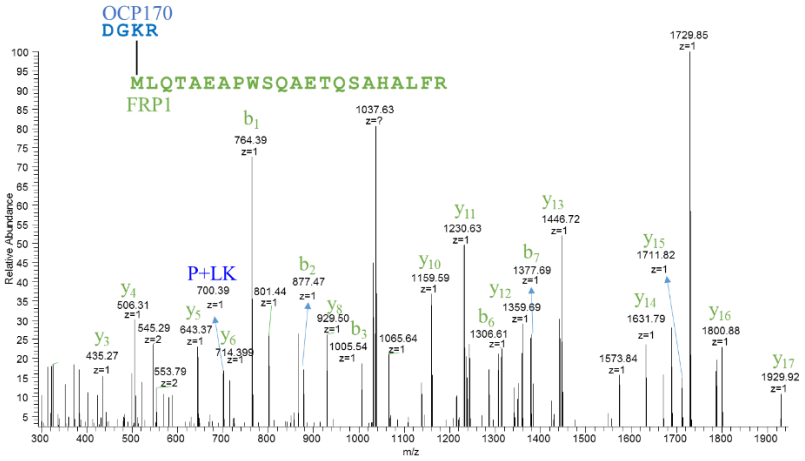
(B)


OCP167
NAVVDMGFTAGKDGK
MLQTAEAPWSQAETQSAHALFR
FRP1
Relative Abundance
m/z

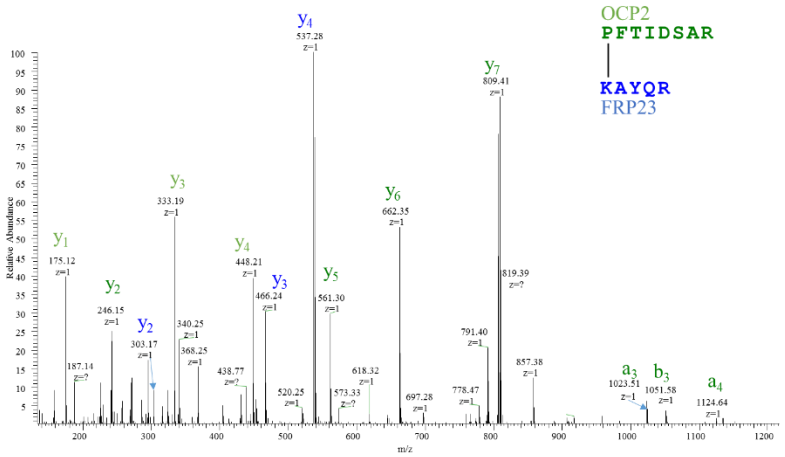
(C)


FRP1
MLQTAEAPWSQAETQSAHALFR
PFTIDSAR
OCP2
Relative Abundance
m/z

(D)



(E)



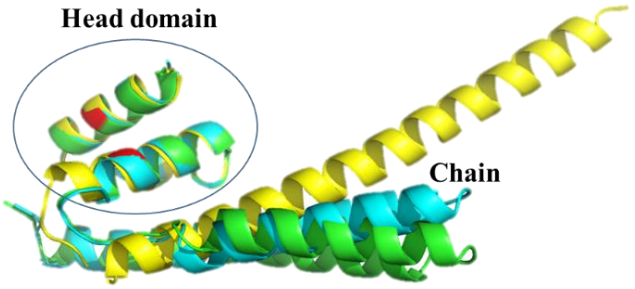
(F)

Figure 6.8. MS/MS ion-product spectra of cross-linked peptides.

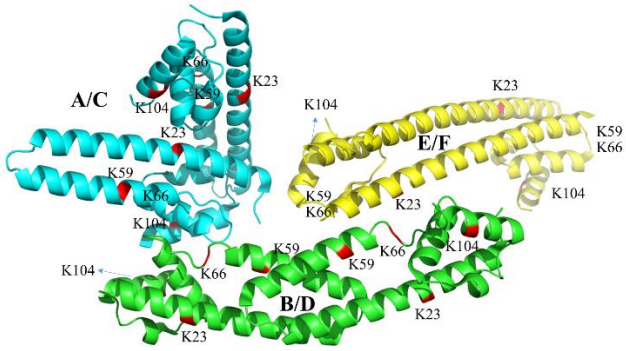
6.4.2 Binding domain on FRP to CTD

To locate the binding site of FRP, we generated a series of FRP site-directed mutants based on conservation analysis[207] and analyzed them with the same native MS platform. The existence of both FRP and NTD/CTD including complexes is seen as direct evidence of the binding affinity between mutant FRP and the NTD/CTD. The peak series (+11, +12 and +13)

representing dFRP-CTD complex dropped to different degree in all mutant samples (Figure 6.10C and Figure 6.11). Thus, the peak series representing the free dFRP (+9 and +10) increased in the two mutant samples mentioned above (Figure 6.10C and Figure 6.11A). The F76D mutant exhibits the most striking decrease in interaction, and the binding of the K102D mutant to the NTD/CTD is also severely affected. Overlapping of the three crystallographic conformation states[207] shows that FRP is composed conservatively of an extended α-helical domain, a small helical cap, while the chain regions are folded in a slightly different way. F76 and K102 are located on the conserved helical cap ("head" region) in all three conformations (Figure 6.9A). Based on the affinity analysis of CTD to dFRP, we propose that this head domain could be the binding face of FRP to the CTD. To be certain of the binding face on FRP, however, more experiments need to be carried out, such like construction of different mutations on the head region.



(A)



(B)

Figure 6.9. (A) Alignment of FRP residues 72-106 on chain A/C (cyan), chain B/D (green) and chain E/F (yellow) from PDB 4JDX[143, 207]. F76 and K102 are colored in red. (B) Chain A/C (cyan), chain B/D (green) and chain E/F (yellow) from PDB 4JDX[143, 207]. The residues being cross-linked are shown in red. K59 and K66 in E/F dimer cannot be labeled because they are located in the missing loop structures.

CTD shows a negatively charged interface that is hidden in the orange state (Figure 6.12A, B), while the binding domain of FRP is positively charged (Figure 6.13A, B). In previous mutagenesis studies on OCP, the "catalytic" ability of FRP was found to be affected by the mutations on the surface of the inter-domain cavity[256]. One proposal is that the positively charged head domain of FRP interacts with the negatively charged interface on the CTD. In addition, the other side of the head, which is originally embedded between the head and chain of FRP, is negatively charged (Figure 6.13D, F). The NTD interface that is hidden in the orange state reveals a positively charged surface (Figure 6.12C, D). One possibility is that, upon binding

of the positively charged head domain of FRP to the negatively charged CTD interface, the head domain on FRP unfolds, exposing the originally embedded negative face that interacts with the positively charged interface on the NTD. In brief, one assumption from the molecular modelling and surface electrostatic analysis is OCP and FRP interacts via an unfolding and bridging mechanism. After the head domain attaching to the CTD, the head domain unfolds and the original hidden face on the head domain attach to the NTD.

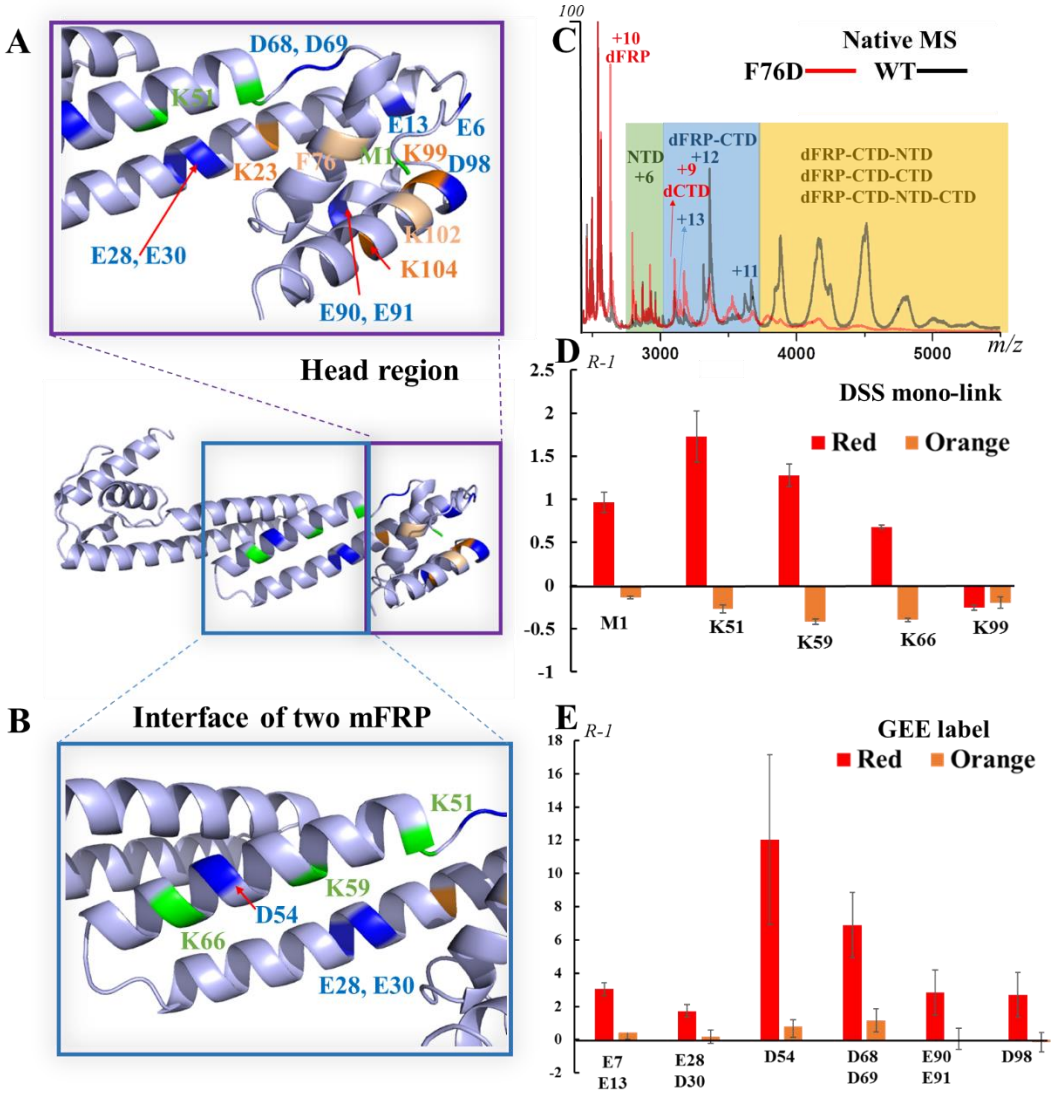
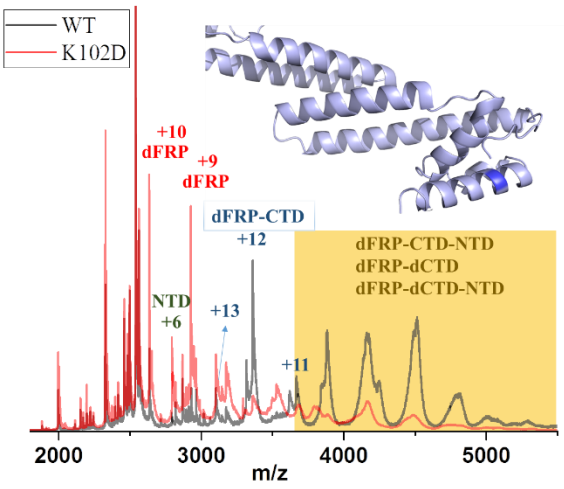
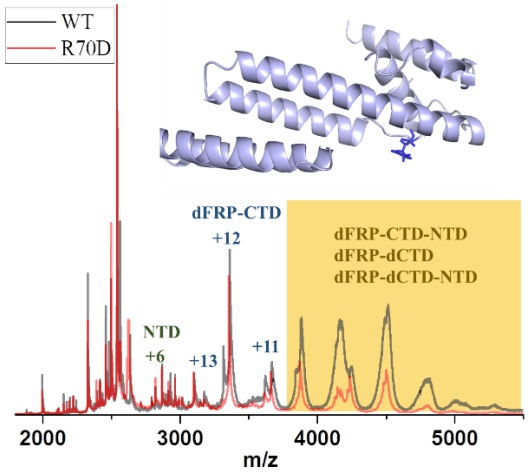


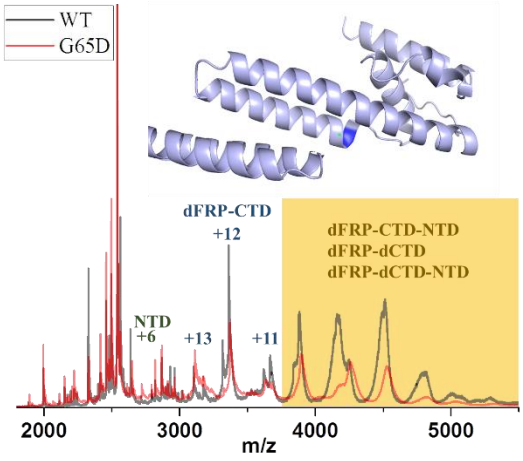
Figure 6.10. Map of residues identified by native MS, cross-linking, and GEE labeling on the FRP head region (A) and dimer-interacting region (B) of chains B/D; structure from PDB 4JDX.[207] (C) Native mass spectra of the NTD/CTD in the presence of WT or F76D FRP. The binding affinity of F76D mutant to NTD/CTD is greatly diminished. (D) A bar graph showing the changing of mono-link extent on FRP in the presence of OCP^r or OCP^o. $R = \frac{\text{Abundance of FRP monolink upon FRP interacting with OCP}}{\text{Abundance of FRP monolink for FRP only sample}}$ (E) Bar graph showing the change in solvent accessibility of FRP residues, as probed by GEE labeling.


WT
K102D
+10
dFRP
+9
dFRP
dFRP-CTD
+12
dFRP-CTD-NTD
dFRP-dCTD
dFRP-dCTD-NTD
NTD
+6
+13
+11
2000
3000
4000
5000
m/z

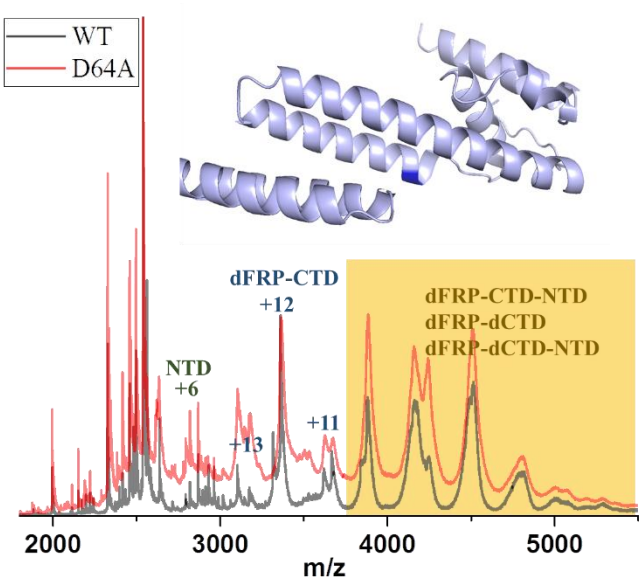
(A)


WT
R70D
dFRP-CTD
+12
dFRP-CTD-NTD
dFRP-dCTD
dFRP-dCTD-NTD
NTD
+6
+13
+11
2000
3000
4000
5000
m/z

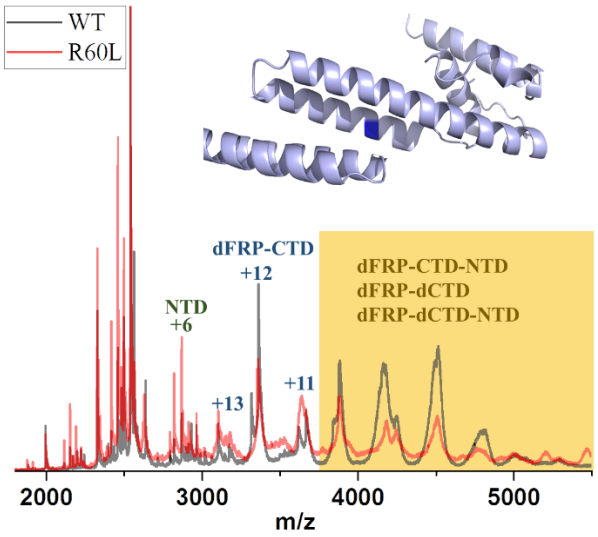
(B)


WT
G65D
dFRP-CTD
+12
dFRP-CTD-NTD
dFRP-dCTD
dFRP-dCTD-NTD
NTD
+6
+13
+11
2000
3000
4000
5000
m/z

(C)



(D)



(E)

Figure 6.11. Native mass spectra of mutant FRP in the presence of NTD/CTD partial digestion fragments (Red). The native MS spectra of WT FRP are shown in black.

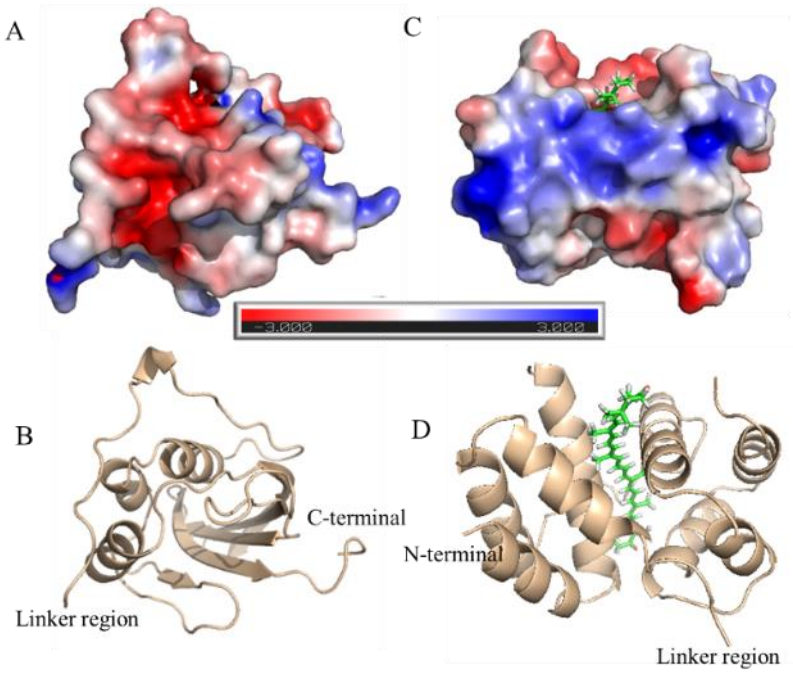


Figure 6.12. The surface electrostatic analysis by APBS (Adaptive Poisson-Boltzmann Solver) shows negatively charged CTD interface (A, the corresponding cartoon structure is shown in B) and positively charged NTD interface (C, the corresponding cartoon structure is shown in D).

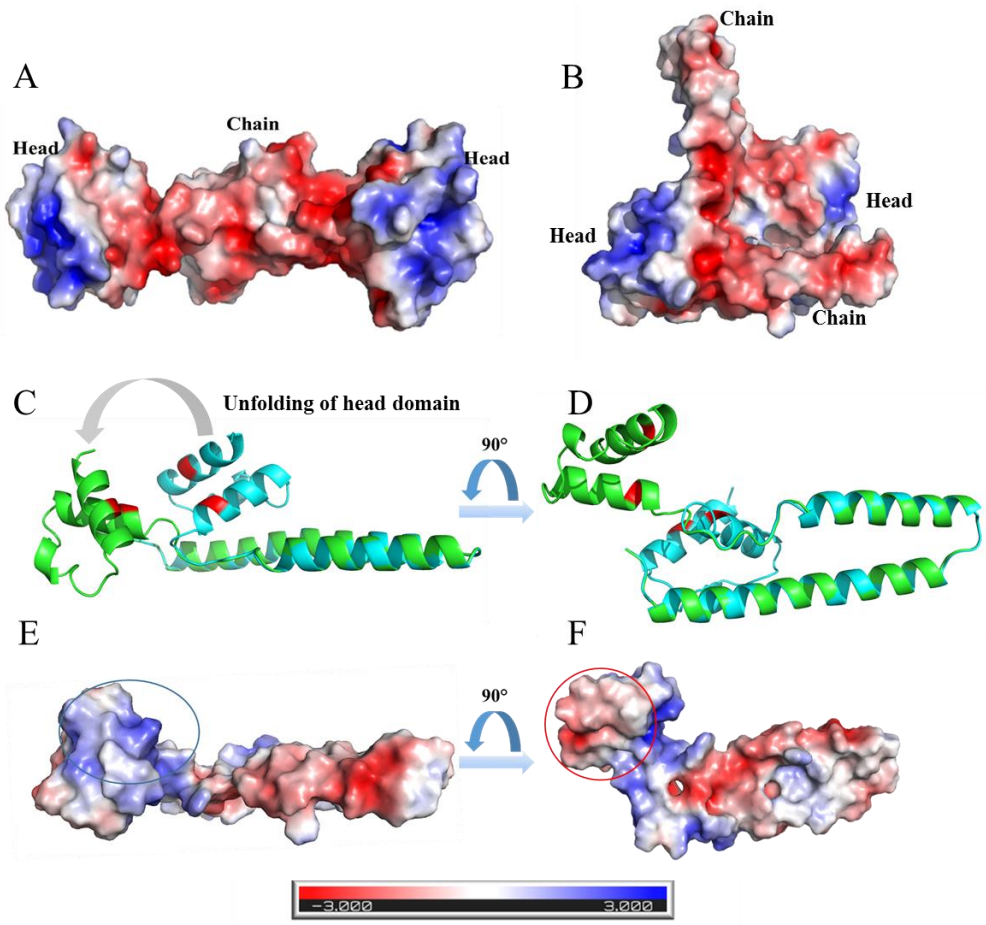


Figure 6.13. The surface electrostatic analysis by APBS shows negatively charged head domains and positively charged chain domain of BD (A) and AC (B) dimer from PDB 4JDX. Chain B (cyan) is aligned to the "unfolded" chain B (Green) as shown in C and D, and the corresponding surface electrostatic analysis of the "unfolded" chain B are shown in E and F. The positively charged binding region is highlighted with blue circle, and the negatively charged region, which in hidden in original state, is highlighted with red circle. The modified structure of chain B (Green) was generated by using the structure building and energy minimization tools in Chimera software[255]. The dihedral angles in the loop region from AA 68-73 (DDRQSV) were modified. Then the produced geometry was energy minimized by applying 1000 steepest descent steps with a step size of 0.02 angstrom (C, D, E, F).

6.4.3 Regions influence FRP function

To evaluate the residues that influence the function of FRP, and not just its binding affinity, we performed a kinetic analysis on the OCP^r to OCP^o relaxation process by monitoring changes in the absorption at 550 nm in the presence of different FRP mutants. The F76D and R60L mutations were found to lose their ability to accelerate OCP^r conversion, and the G65D mutation also greatly affects the FRP function (Figure 6.14). A previous study also demonstrates that the R60L mutant loses its acceleration ability on OCP^r relaxation[207]. The result with the F76D mutant is not surprising, as this mutant loses its affinity to interact with the NTD/CTD, as discussed in the previous section. The R60L mutant can still bind to the NTD/CTD to some degree (Figure 6.11E), although no acceleration capability on OCP^r relaxation is retained (Figure 6.14). We propose that the binding and the acceleration of OCP^r relaxation processes can be decoupled because different amino acids are involved. Native MS analysis is essentially a snapshot of the interaction between the FRP and OCP^r (without the linker region and N-terminal arm), whereas the kinetic analysis requires the FRP to perform its function fully by converting OCP^r into OCP^o. R60 is not located in the binding face on FRP; thus, the binding affinity of FRP is not affected upon mutation. Thurotte and co-workers[241] found that the R60L mutant can still detach OCP from the PBS, but it is unable to accelerate the conversion to the orange state. Both results indicate that the acceleration process can be decoupled into several stages. In addition, the R60L mutant FRP can still perform a conformation change upon binding with the CTD, as the dFRP-CTD-NTD complex was also observed during native MS analysis (data not shown). It is likely that R60 is the crucial residue in the later conformational change, when the two domains are already attaching to FRP.

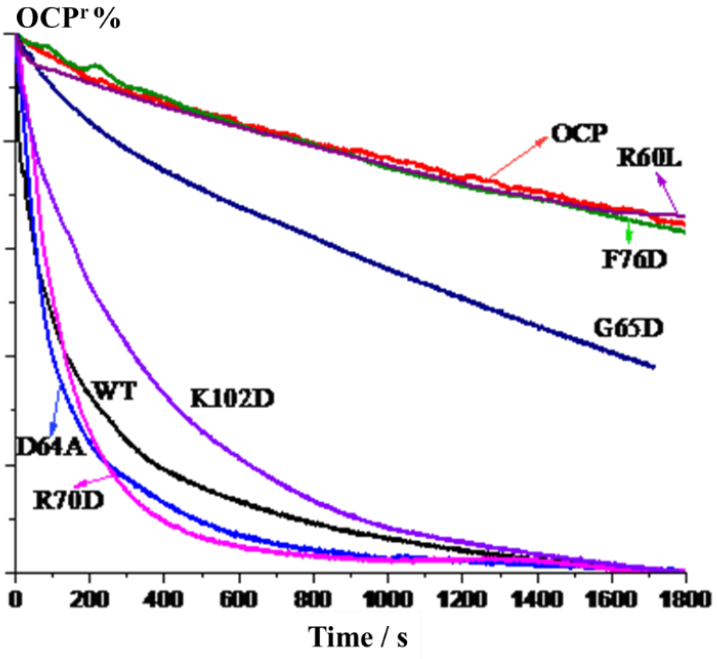


Figure 6.14. Kinetics of the conversion of OCP^r to OCP^o as monitored through changes 550 nm absorption. OCP was either incubated alone, in the presence of WT FRP, or with the FRP mutants R70D, D64A, K102D, G65D, F76D, R60L.

6.4.4 Substantial conformational changes of dFRP after bridging the two domains

To investigate the possible conformational changes that take place upon bringing the two domains together, we utilized intra-molecular cross-links on FRP to evaluate the structural change upon its interaction with OCP. When FRP interacts with OCP^r, the intensity of intra-molecular links increase by several folds, whereas a decrease of intra-molecular cross-link intensity occurs when FRP was incubated with OCP^o (Figure 6.15). The cross-links identified here can either arise from one individual chain, or two subunits of dFRP, although the abundance of cross-linked dFRP is much lower than intra-linked mFRP, as observed by SDS-PAGE (Figure 6.6). The measured distance of the cross-linked amino acid pairs within chains A/C, B/D, and

E/F, as well as dimeric AC, BD and EF (Figure 6.9B) are listed in Table S2. Although the spacer length of DSS is 11.4 Å, a distance constraint of 26–30 Å between Cα atoms is considered to be possible owing to native-state protein dynamics[42]. AC and BD dimers have a higher chance to form the intra-molecular links than the EF dimer when considering the distance constraints. Nevertheless, a substantial structural rearrangement of FRP takes place upon interaction with OCP, especially when OCP is in the red state.

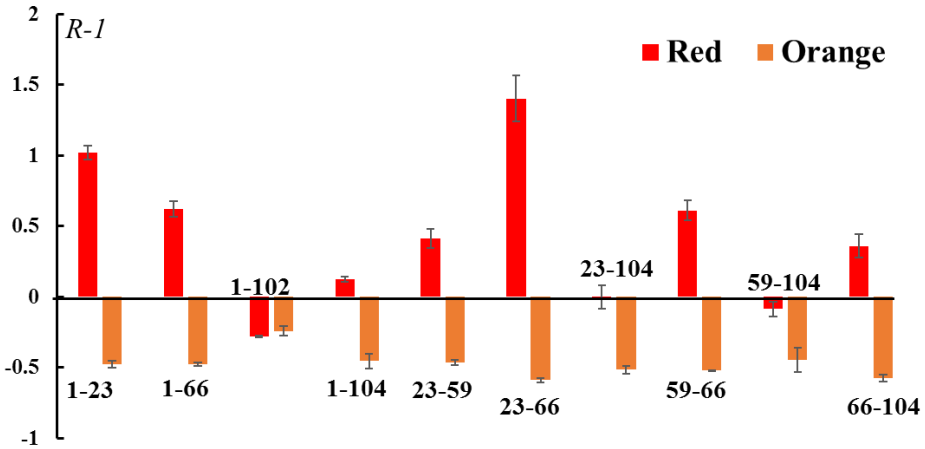


Figure 6.15. Bar graph showing the changes in intra-FRP cross-links in the presence of OCP^r or OCP^o. The numbers on x axis correspond to the position of amino acid residues on FRP.

$$R = \frac{\text{Abundance of intra-FRP crosslink upon FRP interacting with OCP}}{\text{Abundance of intra-FRP crosslink for FRP only sample}}$$

Table. 6.2

Distances of cross-linked lysine/N-termini residues measured by Xwalk server[257] in PDB 4JDX[207] Cross-links involving M1 are not covered in analysis due to the missing N-terminal loop in PDB file.[207] (A) A/C chain (B) B/D chain (C) E/F chain (D) AC dimer (E) BD dimer (F) EF dimer

(A)

Residue 1	Residue 2	Euclidean Distance (Å)	SAS Distance (Å)
K23	59	18.4	27.5
K23	66	13.6	26.4
K23	104	18.2	27.6
K59	66	9.7	12.6
K59	104	25.5	34.2
K66	104	19.3	33.1

(B)

Residue 1	Residue 2	Euclidean Distance (Å)	SAS Distance (Å)
K23	59	19.2	28.4
K23	66	14.8	28.8
K23	104	18.4	28.0
K59	66	10.2	12.9
K59	104	26.5	34.1
K66	104	19.6	32.3

(C)

Residue 1	Residue 2	Euclidean Distance (Å)	SAS Distance (Å)
K23	59	57.6	68.4
K23	66	54.9	61.3
K23	104	43.2	48.6
K59	66	9.2	10.8
K59	104	27.9	48.9
K66	104	23.8	38.0

(D)

Residue 1	Residue 2	Euclidean Distance (Å)	SAS Distance (Å)
K23	59	14.1	16.1
K23	66	13.6	30.3
K23	104	15.8	35.1
K59	66	9.7	12.6
K59	104	19.0	19.3
K66	104	18.7	35.4

(E)

Residue 1	Residue 2	Euclidean Distance (Å)	SAS Distance (Å)
23	59	18.7	28.7
23	66	13.9	26.3
23	104	18.4	28.0
59	66	10.2	12.9
59	104	26.0	34.4
66	104	19.6	32.3

(F)

Residue 1	Residue 2	Euclidean Distance (Å)	SAS Distance (Å)
23	59	23.7	31.2
23	66	16.3	22.9
23	104	19.5	27.2
59	66	9.2	10.8
59	104	27.9	52.4
66	104	23.8	45.1

Interestingly, we identified cross-links between two mFRP subunits: M1 to M1 and K23 to K23. We also found an increased number of M1-M1 cross-links when FRP interacts with OCP^r and, vice versa, when FRP interacts with OCP^o (Figure 6.16B). In addition, the frequency of K23-K23 cross-linking also drops when FRP interacts with OCP^o (Figure 6.16B). One possibility is that the flexible N-terminal loop on each mFRP approaches each other when dFRP interacts with OCP^r, and vice versa when dFRP interacts with OCP^o; the region containing K23 on each mFRP dissociates from each other when dFRP interacts with OCP^o. The other possibility is FRP unfolds into a long alpha helix and thus causes an increase of solvent accessibility. In this new conformation, the N-terminus of each mFRP gets closer compared to the original structure.

A

Dimer	K23--K23 (Å)	M1-M1 (Å)
AC	17.1	48.4 P8--P8
BD	47.2	81.1 P8--A12
EF	41.0	78.3 P8--A12

B

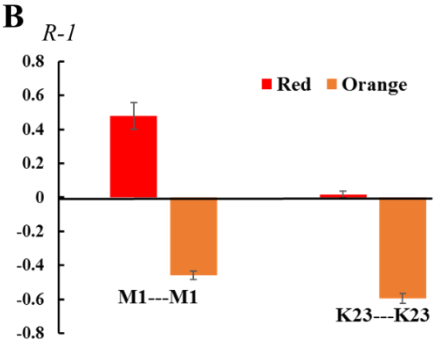


Figure 6.16. (A) Distances between the corresponding residues in PDB 4JDX (A/C chain, B/D chain and E/F chain) are measured by using Pymol software[143]. (B) The bar graph shows the change of inter-mFRP cross-link in the presence of OCP^r or OCP^o.

To investigate further the conformational change of FRP upon interaction with OCP, we probed the solvent accessibility of the protein by evaluating the intensity change of DSS mono-links and GEE-labeled residues in a carboxyl (GEE) footprinting experiment. When FRP interacts with OCP^r, all the mono-links show an increased intensity except for K99, which is located on the binding domain of FRP (Figure 6.10D). To get a more comprehensive or higher resolution picture of solvent accessibility under different conditions, we adopted GEE to footprint carboxyl groups on the amino-acid residues. The reactivity of a carboxyl group on D and E amino-acid side chains is proportionally related to the solvent-accessible surface area of it and the adjacent regions[29]. Similar to that of most of the lysine residues, the solvent accessibility of all the D and E residues on FRP increase dramatically upon interacting with OCP^r (Figure 6.10E) These results strongly support the idea that FRP undergoes a dramatic conformational change during its interaction with OCP^r. The most dramatic change is located on D54, followed by on D68 and D69 (Figure 6.10E). The mutagenesis analysis shows that the amino acids that form a network of hydrogen bonds between the two mFRP chains (R60, W50 and D54) are essential for the enhancement of OCP^r to OCP^o conversion[207]. In addition, two of the previous

studies suggests that FRP monomerizes when interacting with OCP analogs[224, 237]. In this study, a protein complex including mFRP, CTD and NTD was observed after MS/MS dissociation (Figure 6.3B). Thus, the dramatic solvent accessibility change on D54 may be caused by a structural rearrangement of two FRP subunits—perhaps a monomerization of dFRP. However, the increased amount of inter-mFRP M1-M1 cross-links contradicts the monomerization idea of dFRP during its interaction with OCP^r. It is worth mentioning that OCP^r is under constant illumination at low temperature during the labeling process, and the samples including OCP^r still exhibit a red color upon quenching of the reaction. Thus, the FRP hasn't reached the end of its journey of converting the OCP^r into orange. After dFRP attaching to the two domains, the leaving mFRP could be in equilibrium with both the bound mFRP and free mFRP in solution (Figure 6.7C). It is likely that the increasing dynamic motion facilitates the forming of inter-mFRP and intra-FRP cross-links (Figure 6.15, Figure 6.16B). The other possibility is FRP unfolds into long alpha helix and thus causing the increasing of solvent accessibility. And in this new conformation, the N-terminal on each mFRP gets closer compared to the original structure.

6.4.5 Response of OCP^r to FRP

In this part of the study, we identified multiple cross-links between OCP^r-K167, -K170, and -P2 to FRP, suggesting the adjacency of the N-terminal arm and linker region during its interaction with FRP (Figure 6.7). In addition, both domains can attach to FRP, despite the CTD being the major target of FRP, as discussed in the previous section. The two domains of OCP are fully dissociated upon photoactivation, whereas an unstructured loop (the linker region) connecting the two domains becomes significantly exposed[38]. That the OCP photo-activation is not reversible after partial enzymatic digestion indicates that this unstructured loop is crucial for

the conversion of active OCP^r to inactive OCP^o. The N-terminal arm is also a crucial coordinator in the OCP conversion process; this arm not only dissociates from the CTD but becomes disordered during the activation[38-39]. Moreover, the absence of the N-terminal arm largely facilitates the action of FRP on OCP^r and accelerates the detachment of the OCP^r from the PBS[258]. Those previously published findings reinforce our conclusion that the linker-region, N-terminal arm and CTD on OCP^r could be adjacent to the regions that are interacting with FRP.

The extent of both DSS mono-linking and GEE-labeling on OCP^r exhibits no obvious changes in the presence or absence of FRP. An exception is residue E311, which is located at the C-terminal loop of OCP (Figure 6.7D); its decrease in solvent accessibility suggests that E311 could be the binding residue to FRP (Figure 6.17A). The C-terminal loop of OCP is not fully seen in the crystal structures, likely owing to its flexibility[204-205]. Our previous GEE-labeling study shows, upon photoactivation, a marked increase in the extent of E311 labeling[39] as well as a > 2-fold decrease in the labeling of P309/K310 in OCP^r relative to the corresponding amino acids in OCP^o [38]. Both results suggest a movement or structural rearrangement of the C-terminus upon photoactivation. In this study, we found that the C-terminal loop plays a role in the FRP-mediated OCP^r to OCP^o conversion process. In addition, that the FRP burrows into the inter-cavity of the two OCP domains may further separate them, exposing the linker region to a larger extent (increased labeling extent on K167) and slightly increasing the solvent accessibility of K249 on the CTD interface (Figure 6.17C).

6.4.6 Response of OCP^o to FRP

That FRP prevents OCP^o from photoactivation[11, 224] suggests two possible mechanisms: one is that FRP associates with OCP^o and hinders its activation, and the other is that FRP immediately converts OCP^r to OCP^o, preventing the detection of OCP^r. OCP^o was proposed to be weakly or not attached at all to FRP by an immuno-precipitation study[11], and a similar study on the NTD/CTD reveals that only the CTD can bind FRP[207]. The results obtained by Sluchanko et al.[224] suggest a transient interaction between FRP and OCP^o. In this study, we confirmed the interaction between FRP and OCP^o, and found that the linker region is in adjacency with FRP during interaction.

A cross-link was found between the linker region and FRP. Unlike for OCP^r, only one cross-link between OCP^o and FRP could be identified (Figure 6.17D). The intensity of the signal representing this cross-link is much lower compared to that for the red state, suggesting a much smaller interface or binding affinity. The only residues exhibiting a noticeable decrease of solvent accessibility upon binding to FRP are D35 (from GEE labeling) and T15 (from DSS cross-linking) (Figure 6.17B, C). These results suggest that the NTD in the orange state associates with the FRP to some degree.

In addition, the abundance of signals representing intra-molecular links on FRP all decreased when interacting with OCP^o (Figure 6.15). This could be due to the anchoring of dFRP on OCP^o, preventing dynamic motions or transition into other states of FRP. In addition, the formation of OCP-dFRP is barely detectable by native MS (Figure 6.18), and the corresponding bands on SDS-PAGE are barely visible (Figure 6.6). Thus, the interaction between OCP^o and FRP is significantly weaker compared to OCP^r and FRP.

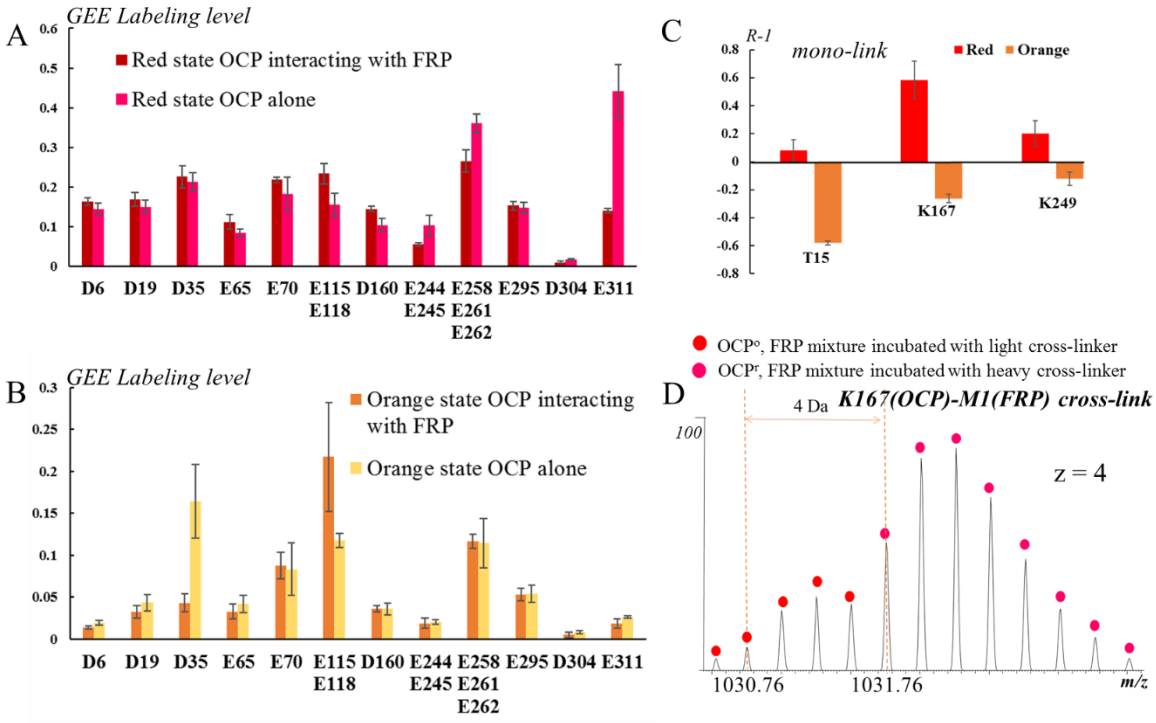


Figure 6.17. (A) GEE labeling extents of OCP^r amino-acid residues in the presence or absence of FRP. (B) GEE labeling level of OCP^o amino acid residues in the presence or absence of FRP.

$$\text{GEE labeling level} = \frac{\text{Peak area of labled peptide}}{\text{Peak area of labled peptide + Peak area of unlabled peptide}}$$ (C) A bar graph showing the changes in OCP mono-links in the presence FRP. $$R = \frac{\text{Abundance of monolink upon interacting with FRP}}{\text{Abundance of monolink for OCP only sample}}$$ (D) An isotope-encoded mass spectrum of crosslinked OCP-FRP showing the K167(OCP)-M1(FRP) cross-link in the orange state is less abundant than in the red state.

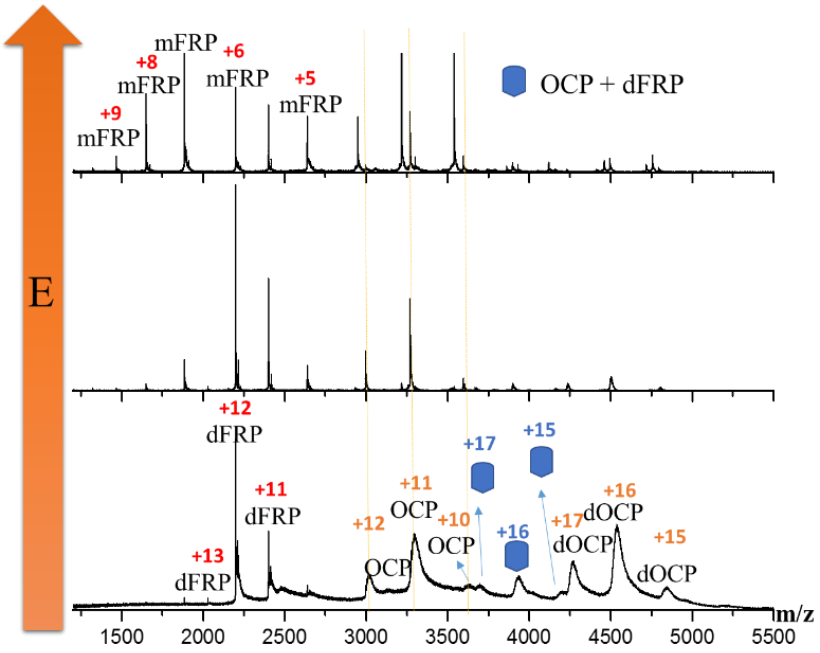


Figure 6.18. Native MS spectra showing the weak association between OCP^o and dFRP. The collisional energy is increased from bottom to top.

6.5 Conclusion

We propose that FRP performs its function through several independent steps, involving various amino acids in the process. The accelerated conversion process starts with dFRP binding to the CTD on OCP^r as suggested by native MS results. The preliminary results obtained by native MS, site-directed mutagenesis (F76, K102) and mono-crosslink (K99) suggest the initial binding domain on FRP to the CTD. The structural rearrangement of dFRP upon binding with the CTD enables its binding to the NTD. To the best of our knowledge, this is the first experimental evidence showing the interaction between the NTD on OCP^r and dFRP. Considering the cross-links found between two domains of OCP and FRP, we can propose a bridging mechanism. This is in accordance with previous molecular modeling results[239]. In that

model, the CTD interacts with both head and chain domains of the FRP, while the NTD is in contact only with the head domain. We also here propose that the head domain interacts with the two domains on OCP. This proposal also explains the low affinity of OCP^o to FRP; that is, the binding face is hidden owing to the compact structure of the orange state. Besides, the previous publication suggests the N-terminal arm on OCP could interfere with the interaction with FRP[49]. The cross-linking results suggest the adjacency of N-terminal arm to FRP, which reinforces this view. To reset OCP^r into a compact structure, the FRP interaction must bring the two domains back together, preparing for the translocation of carotenoid. Native MS results provide the direct evidence that NTD, CTD and dFRP can form a stable complex. The cross-linking results suggest the adjacency of the interface on NTD and CTD to FRP. Either the monomerization of dFRP or dFRP unfolds into a long alpha helix is likely to happen after dFRP bridges the two domains together, enabled by the crucial role of residue R60 in this conformational change. The linker region or N-terminal arm of OCP might be crucial in inducing the structural rearrangement of dFRP. In the orange state, the linker region is associated with dFRP to a lesser degree and the NTD also possibly retains an affinity to dFRP (Fig. 9). In addition, we found an interesting mutual response between FRP and OCP. FRP destabilize OCP^r by accelerating its conversion to orange state, while FRP itself is also "destabilized". FRP can "lock" OCP^o in orange state, while FRP itself is also "stabilized" by OCP^o. Thus, we propose a "stiff" structure when FRP interacting with OCP^o and a dynamic structure when FRP interacting with OCP^r. This study offers new insights into the interaction between FRP and OCP, laying the groundwork for further investigations into the energy-transfer regulation mechanisms of cyanobacterial NPQ.

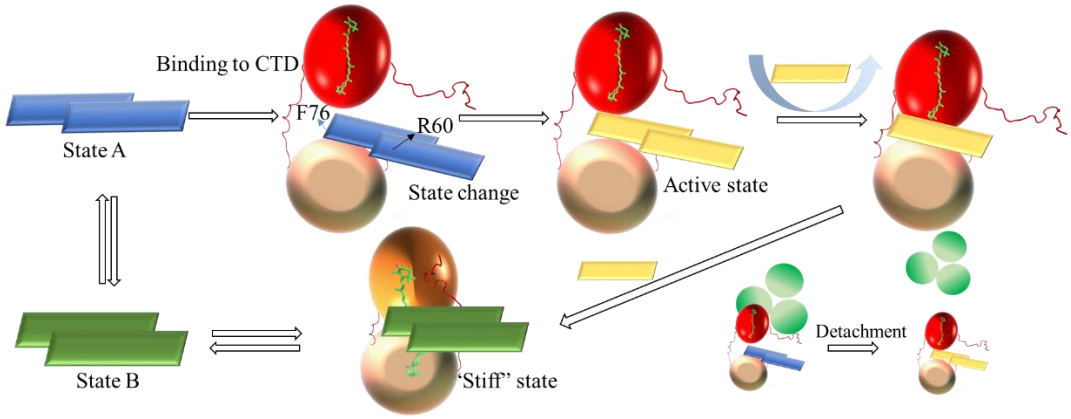


Figure 6.19. Proposed model of the interaction between FRP and OCP. FRP is shown as a rectangle with different colors corresponding to different states; the OCP NTD is shown as a red ball in the red state and an orange ball in the orange state; the OCP CTD is shown as a wheat-colored ball; the APC core from the PBS is shown as three green circles. In the first step, dimeric FRP approaches and binds the CTD, inducing a conformational change in FRP (likely unfolding of the head region) and enabling its binding to the NTD. dFRP forms a stable complex with OCP^r by bridging the two domains. At the stage, OCP^r can be detached from PBS by FRP. The cooperative action of the linker region and the N-terminal arm facilitates the dissociation of the FRP dimer. Monomeric FRP is more flexible and effectively facilitates the closing-up of the two OCP domains. Finally, the accelerated conversion process finishes, and dFRP weakly associates with OCP^o around the linker region. The "capped" N-terminal arm on the CTD inhibits the dFRP from entering the cavity.

Chapter 7: Conclusions and Perspective

Photosynthesis is one of the most important biological processes, supplying energy for almost all life on Earth. Solar energy is stored as chemical energy, and the oxygen content of the atmosphere is produced and maintained by photosynthesis. Understanding the molecular mechanism of photosynthesis is crucial; however, some important details of the molecular processes are still missing. For example, one of the biggest concerns in photosynthesis is the water-splitting process at the oxygen evolution center. Even more noteworthy is that although a variety of protein machineries play critical roles in this process, how those machineries coordinate the photosynthesis process remains an unanswered and intriguing question. To continue to elaborate the details of photosynthesis, protein chemistry must be addressed, and mass spectrometry now plays a central and indispensable role in protein science.

In chapter 2, we utilized the top-down MS and ECD fragmentation to identify the PTM/mutation on LH2, an intrinsic membrane protein. We identified the isoforms of α and β subunit as well as a new PTM and an unexpected mutation. Although the 100% identification of membrane protein structure by MS is usually difficult, especially via bottom-up approaches, the top-down sequencing of photosynthetic membrane proteins is a promising platform, as the issue of low accessibility for enzymatic digestion is avoided, and the advance of different fragmentation methods largely increase the capability of top-down MS. Especially given the fact that many photosynthetic proteins are encoded with more one copy of operon, whether the proteins encoded with those operons are expressed and incorporated into the protein complexes is an intriguing question.

We developed a MS-based membrane protein footprinting platform, using LH2 as a model membrane protein and Nanodiscs as a vehicle to carry this membrane protein, as described in chapter 3. The results indicate that the outer membrane part has a larger solvent accessibility while the regions spanning the membrane have a lower solvent accessibility. This study lays the foundation for future studies of other photosynthetic membrane proteins where the experimentalist can focus on protein-protein interactions, protein-pigment interactions and protein-lipid interactions. For example, whether the properties of lipids influence the topology, and whether the intrinsic lipid binding regions exist on the protein pigment complex.

Chapter 4 describes a combination of MS-based methods to study the reaction center from *B. viridis*, the first membrane protein structure obtained with an atomic resolution structure. The results contradict a crystallographic study. A 7-amino acid alteration plus a 7-amino acid extension were identified on the M subunit of this reaction center. The native MS results show that this reaction center can maintain its integrity in the gas phase, even to the point to preserve four non-covalent bacteriochlorophylls that are contained as part of its higher order structure. We think it is a good example showing that it is necessary to conduct MS analysis on proteins even with the crystal structure available, providing complementary information and even contradictory results.

In chapter 5, we utilized the native MS to investigate the oligomeric state of FRP and OCP proteins, as the oligomeric state of these two proteins has been controversial. We observed that OCP exhibits concentration-dependent behavior in terms of oligomeric state whereas FRP exhibits a dimeric structure consistently. In addition, ion mobility MS reveals the high flexibility of the FRP protein.

Chapter 6 shows how a combination of MS-based methods, including XL-MS, footprinting and native MS, as well as mutagenesis and kinetic studies can reveal the molecular mechanism of non-photochemical quenching in cyanobacteria. FRP functions like an enzyme, "catalyzes" the relaxation of OCP and interacts with both states of OCP. This study lays the ground work for the future study of the three-player system in the cyanobacterial NPQ process. That system includes OCP, FRP and PBS. As OCP burrows into PBS to quench the excess energy, FRP may work as an inducer of conformational change of OCP or as a stronger binder than PBS and, thus, drag the OCP away from PBS. It will be intriguing to draw the global picture as well as the local detail of mechanism in this process.

MS has been routinely used to conduct proteomics analysis, for example, evaluating the proteomics in different cell fractions. This is especially useful for photosynthetic proteins, as the composition is complex, many of them are expressed with PTMs, and more than one isoform may be expressed. Now, mass spectrometers can be employed to achieve different goals: deciphering the structure, dynamics and function of proteins. These advances are beginning to benefit the study of photosynthesis.

We foresee a broad utilization of mass spectrometry to study protein-pigment interactions, reaction center/light harvesting complex dynamics, and structure. For example, the protein subunit organization of PSII is being probed by XL-MS, as shown in collaborative work between the Pakrasi and Gross labs[46], underscoring the utility of crosslinking in the future. In addition, the development of membrane protein MS techniques in recent years also benefits the study of photosynthetic membrane proteins. Our MS-based footprinting platform on membrane proteins lays ground for future membrane protein studies: membrane protein-ligand

interaction, influence of lipids on membrane protein, influence of pH on membrane protein and so forth. One promising direction is the native MS interrogation of photosynthetic proteins incorporated in a Nanodisc. The most native structure of photosynthetic membrane proteins may be probed and protein-lipid and even pigment-lipid interaction can be investigated.

References

1. Blankenship, R. E., *Molecular Mechanisms of Photosynthesis*. Wiley: 2002.
2. Codgell, R. J.; Southall, J.; Gardiner, A. T.; Law, C. J.; Gall, A.; Roszak, A. W.; Isaacs, N. W., How purple photosynthetic bacteria harvest solar energy. *Comptes Rendus Chimie* **2006,** *9* (2), 201-206.
3. MacColl, R., Cyanobacterial Phycobilisomes. *Journal of Structural Biology* **1998,** *124* (2), 311-334.
4. 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* (6162), 1104-1107.
5. Shen, J.-R., The structure of photosystem II and the mechanism of water oxidation in photosynthesis. *Annual review of plant biology* **2015,** *66*, 23-48.
6. Ponomarenko, N. S.; Li, L.; Marino, A. R.; Tereshko, V.; Ostafin, A.; Popova, J. A.; Bylina, E. J.; Ismagilov, R. F.; Norris Jr, J. R., Structural and spectropotentiometric analysis of Blastochloris viridis heterodimer mutant reaction center. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2009,** *1788* (9), 1822-1831.
7. Kirilovsky, D., Photoprotection in cyanobacteria: the orange carotenoid protein (OCP)-related non-photochemical-quenching mechanism. *Photosynthesis Research* **2007,** *93* (1), 7.
8. Szabó, I.; Bergantino, E.; Giacometti, G. M., Light and oxygenic photosynthesis: energy dissipation as a protection mechanism against photo-oxidation. *EMBO Reports* **2005,** *6* (7), 629-634.
9. Kirilovsky, D.; Kerfeld, C. A., Cyanobacterial photoprotection by the orange carotenoid protein. *Nature Plants* **2016,** *2*, 16180.
10. Cogdell, R. J.; Gardiner, A. T., Activated OCP unlocks nonphotochemical quenching in cyanobacteria. *Proceedings of the National Academy of Sciences* **2015,** *112* (41), 12547-12548.
11. Boulay, C.; Wilson, A.; D'Haene, S.; Kirilovsky, D., Identification of a protein required for recovery of full antenna capacity in OCP-related photoprotective mechanism in cyanobacteria. *Proceedings of the National Academy of Sciences* **2010,** *107* (25), 11620-11625.
12. Lane, C. S., Mass spectrometry-based proteomics in the life sciences. *Cellular and Molecular Life Sciences CMLS* **2005,** *62* (7), 848-869.
13. Schermann, S. M.; Simmons, D. A.; Konermann, L., Mass spectrometry-based approaches to protein–ligand interactions. *Expert Review of Proteomics* **2005,** *2* (4), 475-485.
14. Mehmood, S.; Allison, T. M.; Robinson, C. V., Mass spectrometry of protein complexes: from origins to applications. *Annu. Rev. Phys. Chem.* **2015,** *66*, 453-474.
15. Crick, F., Central dogma of molecular biology. *Nature* **1970,** *227* (5258), 561-563.
16. Domon, B.; Aebersold, R., Mass spectrometry and protein analysis. *science* **2006,** *312* (5771), 212-217.
17. Håkansson, K.; Klassen, J. S., Ion Activation Methods for Tandem Mass Spectrometry. In *Electrospray and MALDI Mass Spectrometry*, John Wiley & Sons, Inc.: 2010; pp 571-630.
18. Mitchell Wells, J.; McLuckey, S. A., Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*, Academic Press: 2005; Vol. Volume 402, pp 148-185.
19. Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S., Improved Peptide Identification by Targeted Fragmentation

Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *Journal of Proteome Research* **2011,** *10* (5), 2377-2388.
20. Chi, H.; Sun, R.-X.; Yang, B.; Song, C.-Q.; Wang, L.-H.; Liu, C.; Fu, Y.; Yuan, Z.-F.; Wang, H.-P.; He, S.-M.; Dong, M.-Q., pNovo: De novo Peptide Sequencing and Identification Using HCD Spectra. *Journal of Proteome Research* **2010,** *9* (5), 2713-2724.
21. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society* **1998,** *120* (13), 3265-3266.
22. Syrstad, E. A.; Tureček, F., Toward a general mechanism of electron capture dissociation. *Journal of the American Society for Mass Spectrometry* **2005,** *16* (2), 208-224.
23. Zubarev, R. A.; Haselmann, K. F.; Budnik, B.; Kjeldsen, F.; Jensen, F., Towards an understanding of the mechanism of electron-capture dissociation: a historical perspective and modern ideas. *European Journal of Mass Spectrometry* **2002,** *8* (5), 337-350.
24. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews* **2013,** *113* (4), 2343-2394.
25. Sharma, J.; Panico, M.; Shipton, C. A.; Nilsson, F.; Morris, H. R.; Barber, J., Primary Structure Characterization of the Photosystem II D1 and D2 Subunits. *Journal of Biological Chemistry* **1997,** *272* (52), 33158-33166.
26. Che, Y.; Fu, A.; Hou, X.; McDonald, K.; Buchanan, B. B.; Huang, W.; Luan, S., C-terminal processing of reaction center protein D1 is essential for the function and assembly of photosystem II in Arabidopsis. *Proceedings of the National Academy of Sciences* **2013,** *110* (40), 16247-16252.
27. Guan, J.-Q.; Chance, M. R., Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. *Trends Biochem. Sci* **2005,** *30* (10), 583-592.
28. Deng, B.; Lento, C.; Wilson, D. J., Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development – A review. *Analytica Chimica Acta* **2016,** *940*, 8-20.
29. Zhang, H.; Wen, J.; Huang, R. Y. C.; Blankenship, R. E.; Gross, M. L., Mass spectrometry-based carboxyl footprinting of proteins: method evaluation. *International journal of mass spectrometry* **2012,** *312*, 78-86.
30. Engen, J. R., Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Analytical Chemistry* **2009,** *81* (19), 7870-7875.
31. Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C., Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-A review. *Analytica Chimica Acta* **2012,** *721*, 7-21.
32. Andrew N. Hoofnagle; Katheryn A. Resing; Ahn, N. G., Protein Analysis by Hydrogen Exchange Mass Spectrometry. *Annual Review of Biophysics and Biomolecular Structure* **2003,** *32* (1), 1-25.
33. Huang, R. Y. C.; Wen, J.; Blankenship, R. E.; Gross, M. L., Hydrogen-deuterium exchange mass spectrometry reveals the interaction of Fenna-Matthews-Olson protein and chlorosome CsmA protein. *Biochemistry* **2012,** *51* (1), 187-193.
34. Maleknia, S. D.; Downard, K. M., Advances in radical probe mass spectrometry for protein footprinting in chemical biology applications. *Chemical Society Reviews* **2014,** *43* (10), 3244-3258.
35. Maleknia, S. D.; Downard, K. M., Radical approaches to probe protein structure, folding, and interactions by mass spectrometry. *Mass Spectrometry Reviews* **2001,** *20* (6), 388-401.

36. Hambly, D. M.; Gross, M. L., Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *Journal of the American Society for Mass Spectrometry* **2005,** *16* (12), 2057-2063.
37. Maleknia, S. D.; Ralston, C. Y.; Brenowitz, M. D.; Downard, K. M.; Chance, M. R., Determination of Macromolecular Folding and Structure by Synchrotron X-Ray Radiolysis Techniques. *Analytical Biochemistry* **2001,** *289* (2), 103-115.
38. Gupta, S.; Guttman, M.; Leverenz, R. L.; Zhumadilova, K.; Pawlowski, E. G.; Petzold, C. J.; Lee, K. K.; Ralston, C. Y.; Kerfeld, C. A., Local and global structural drivers for the photoactivation of the orange carotenoid protein. *Proceedings of the National Academy of Sciences* **2015,** *112* (41), E5567-E5574.
39. Liu, H.; Zhang, H.; King, J. D.; Wolf, N. R.; Prado, M.; Gross, M. L.; Blankenship, R. E., Mass spectrometry footprinting reveals the structural rearrangements of cyanobacterial orange carotenoid protein upon light activation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2014,** *1837* (12), 1955-1963.
40. Bullock, J. M. A.; Schwab, J.; Thalassinos, K.; Topf, M., The Importance of Non-accessible Crosslinks and Solvent Accessible Surface Distance in Modeling Proteins with Restraints From Crosslinking Mass Spectrometry. *Molecular & Cellular Proteomics* **2016,** *15* (7), 2491-2500.
41. Schmidt, C.; Robinson, C. V., A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nat. Protocols* **2014,** *9* (9), 2224-2236.
42. 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 Science* **2014,** *23* (6), 747-759.
43. Rappsilber, J., The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of Structural Biology* **2011,** *173* (3), 530-540.
44. Mummadisetti, M. P.; Frankel, L. K.; Bellamy, H. D.; Sallans, L.; Goettert, J. S.; Brylinski, M.; Limbach, P. A.; Bricker, T. M., Use of protein cross-linking and radiolytic footprinting to elucidate PsbP and PsbQ interactions within higher plant Photosystem II. *Proceedings of the National Academy of Sciences* **2014,** *111* (45), 16178-16183.
45. Liu, H.; Huang, R. Y. C.; Chen, J.; Gross, M. L.; Pakrasi, H. B., Psb27, a transiently associated protein, binds to the chlorophyll binding protein CP43 in photosystem II assembly intermediates. *Proceedings of the National Academy of Sciences of the United States of America* **2011,** *108* (45), 18536-18541.
46. Weisz, D. A.; Liu, H.; Zhang, H.; Thangapandian, S.; Tajkhorshid, E.; Gross, M. L.; Pakrasi, H. B., Mass spectrometry-based cross-linking study shows that the Psb28 protein binds to cytochrome b559 in Photosystem II. *Proceedings of the National Academy of Sciences* **2017**.
47. Mummadisetti, M. P.; Frankel, L. K.; Bellamy, H. D.; Sallans, L.; Goettert, J. S.; Brylinski, M.; Bricker, T. M., Use of Protein Cross-Linking and Radiolytic Labeling to Elucidate the Structure of PsbO within Higher-Plant Photosystem II. *Biochemistry* **2016,** *55* (23), 3204-3213.
48. Gao, X.; Xin, Y.; Bell, P. D.; Wen, J.; Blankenship, R. E., Structural Analysis of Alternative Complex III in the Photosynthetic Electron Transfer Chain of Chloroflexus aurantiacus. *Biochemistry* **2010,** *49* (31), 6670-6679.

49. He, G.; Zhang, H.; King, J. D.; Blankenship, R. E., Structural Analysis of the Homodimeric Reaction Center Complex from the Photosynthetic Green Sulfur Bacterium Chlorobaculum tepidum. *Biochemistry* **2014,** *53* (30), 4924-4930.
50. Sakata, S.; Mizusawa, N.; Kubota-Kawai, H.; Sakurai, I.; Wada, H., Psb28 is involved in recovery of photosystem II at high temperature in Synechocystis sp. PCC 6803. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2013,** *1827* (1), 50-59.
51. Cui, W.; Rohrs, H. W.; Gross, M. L., Top-down mass spectrometry: Recent developments, applications and perspectives. *Analyst* **2011,** *136* (19), 3854-3864.
52. Persson, S.; Sönksen, C. P.; Frigaard, N.-U.; Cox, R. P.; Roepstorff, P.; Miller, M., Pigments and proteins in green bacterial chlorosomes studied by matrix-assisted laser desorption ionization mass spectrometry. *European Journal of Biochemistry* **2000,** *267* (2), 450-456.
53. Zhang, H.; Cui, W.; Gross, M. L.; Blankenship, R. E., Native mass spectrometry of photosynthetic pigment–protein complexes. *FEBS Letters* **2013,** *587* (8), 1012-1020.
54. Wen, J.; Zhang, H.; Gross, M. L.; Blankenship, R. E., Native Electrospray Mass Spectrometry Reveals the Nature and Stoichiometry of Pigments in the FMO Photosynthetic Antenna Protein. *Biochemistry* **2011,** *50* (17), 3502-3511.
55. Zhang, H.; Liu, H.; Niedzwiedzki, D. M.; Prado, M.; Jiang, J.; Gross, M. L.; Blankenship, R. E., Molecular Mechanism of Photoactivation and Structural Location of the Cyanobacterial Orange Carotenoid Protein. *Biochemistry* **2014,** *53* (1), 13-19.
56. Lu, Y.; Liu, H.; Saer, R. G.; Zhang, H.; Meyer, C. M.; Li, V. L.; Shi, L.; King, J. D.; Gross, M. L.; Blankenship, R. E., Native Mass Spectrometry Analysis of Oligomerization States of Fluorescence Recovery Protein and Orange Carotenoid Protein: Two Proteins Involved in the Cyanobacterial Photoprotection Cycle. *Biochemistry* **2017,** *56* (1), 160-166.
57. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001,** *19* (3), 242-247.
58. Whitelegge, J. P., Integral Membrane Proteins and Bilayer Proteomics. *Analytical Chemistry* **2013,** *85* (5), 2558-2568.
59. Zhang, X., Less is More: Membrane Protein Digestion Beyond Urea–Trypsin Solution for Next-level Proteomics. *Molecular & Cellular Proteomics* **2015,** *14* (9), 2441-2453.
60. Wei, X.; Su, X.; Cao, P.; Liu, X.; Chang, W.; Li, M.; Zhang, X.; Liu, Z., Structure of spinach photosystem II–LHCII supercomplex at 3.2 Å resolution. *Nature* **2016,** *534* (7605), 69-74.
61. Whitelegge, J., Tandem mass spectrometry of integral membrane proteins for top-down proteomics. *TrAC Trends in Analytical Chemistry* **2005,** *24* (7), 576-582.
62. Whitelegge, J. P.; le Coutre, J.; Lee, J. C.; Engel, C. K.; Privé, G. G.; Faull, K. F.; Kaback, H. R., Toward the bilayer proteome, electrospray ionization-mass spectrometry of large, intact transmembrane proteins. *Proceedings of the National Academy of Sciences* **1999,** *96* (19), 10695-10698.
63. Thangaraj, B.; Ryan, C. M.; Souda, P.; Krause, K.; Faull, K. F.; Weber, A. P. M.; Fromme, P.; Whitelegge, J. P., Data-directed top-down Fourier-transform mass spectrometry of a large integral membrane protein complex: Photosystem II from Galdieria sulphuraria. *Proteomics* **2010,** *10* (20), 3644-3656.
64. Lu, Y.; Zhang, H.; Cui, W.; Saer, R.; Liu, H.; Gross, M. L.; Blankenship, R. E., Top-Down Mass Spectrometry Analysis of Membrane-Bound Light-Harvesting Complex 2 from Rhodobacter sphaeroides. *Biochemistry* **2015,** *54* (49), 7261-7271.

65. Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V., Micelles Protect Membrane Complexes from Solution to Vacuum. *Science* **2008,** *321* (5886), 243-246.
66. Laganowsky, A.; Reading, E.; Hopper, J. T. S.; Robinson, C. V., Mass spectrometry of intact membrane protein complexes. *Nat. Protocols* **2013,** *8* (4), 639-651.
67. Zhang, H.; Harrington, L. B.; Lu, Y.; Prado, M.; Saer, R.; Rempel, D.; Blankenship, R. E.; Gross, M. L., Native Mass Spectrometry Characterizes the Photosynthetic Reaction Center Complex from the Purple Bacterium Rhodobacter sphaeroides. *Journal of The American Society for Mass Spectrometry* **2016**, 1-9.
68. Pan, Y.; Stocks, B. B.; Brown, L.; Konermann, L., Structural Characterization of an Integral Membrane Protein in Its Natural Lipid Environment by Oxidative Methionine Labeling and Mass Spectrometry. *Analytical Chemistry* **2009,** *81* (1), 28-35.
69. Busenlehner, L. S.; Salomonsson, L.; Brzezinski, P.; Armstrong, R. N., Mapping protein dynamics in catalytic intermediates of the redox-driven proton pump cytochrome c oxidase. *Proceedings of the National Academy of Sciences* **2006,** *103* (42), 15398-15403.
70. Orban, T.; Gupta, S.; Palczewski, K.; Chance, M. R., Visualizing Water Molecules in Transmembrane Proteins Using Radiolytic Labeling Methods. *Biochemistry* **2010,** *49* (5), 827-834.
71. Lu, Y.; Zhang, H.; Niedzwiedzki, D. M.; Jiang, J.; Blankenship, R. E.; Gross, M. L., Fast Photochemical Oxidation of Proteins Maps the Topology of Intrinsic Membrane Proteins: Light-Harvesting Complex 2 in a Nanodisc. *Analytical Chemistry* **2016,** *88* (17), 8827-8834.
72. Doucette, A. A.; Vieira, D. B.; Orton, D. J.; Wall, M. J., Resolubilization of Precipitated Intact Membrane Proteins with Cold Formic Acid for Analysis by Mass Spectrometry. *Journal of Proteome Research* **2014,** *13* (12), 6001-6012.
73. Switzar, L.; Giera, M.; Niessen, W. M. A., Protein Digestion: An Overview of the Available Techniques and Recent Developments. *Journal of Proteome Research* **2013,** *12* (3), 1067-1077.
74. Denisov, I. G.; Sligar, S. G., Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **2016,** *23* (6), 481-486.
75. Vadas, O.; Jenkins, M. L.; Dornan, G. L.; Burke, J. E., Chapter Seven - Using Hydrogen–Deuterium Exchange Mass Spectrometry to Examine Protein–Membrane Interactions. In *Methods in Enzymology*, Michael, H. G., Ed. Academic Press: 2017; Vol. Volume 583, pp 143-172.
76. Souda, P.; Ryan, C. M.; Cramer, W. A.; Whitelegge, J., Profiling of integral membrane proteins and their post translational modifications using high-resolution mass spectrometry. *Methods* **2011,** *55* (4), 330-336.
77. Pullerits, T.; Sundström, V., Photosynthetic Light-Harvesting Pigment−Protein Complexes: Toward Understanding How and Why. *Accounts of Chemical Research* **1996,** *29* (8), 381-389.
78. Qian, P.; Papiz, M. Z.; Jackson, P. J.; Brindley, A. A.; Ng, I. W.; Olsen, J. D.; Dickman, M. J.; Bullough, P. A.; Hunter, C. N., Three-Dimensional Structure of the Rhodobacter sphaeroides RC-LH1-PufX Complex: Dimerization and Quinone Channels Promoted by PufX. *Biochemistry* **2013,** *52* (43), 7575-7585.
79. Law, C. J.; Roszak, A. W.; Southall, J.; Gardiner, A. T.; Isaacs, N. W.; Cogdell, R. J., The structure and function of bacterial light-harvesting complexes (Review). *Molecular Membrane Biology* **2004,** *21* (3), 183-191.

80. Papiz, M. Z.; Prince, S. M.; Howard, T.; Cogdell, R. J.; Isaacs, N. W., The Structure and Thermal Motion of the B800–850 LH2 Complex from Rps. acidophila at 2.0 Å Resolution and 100 K: New Structural Features and Functionally Relevant Motions. *Journal of Molecular Biology* **2003,** *326* (5), 1523-1538.
81. Koepke, J.; Hu, X.; Muenke, C.; Schulten, K.; Michel, H., The crystal structure of the light-harvesting complex II (B800–850) from Rhodospirillum molischianum. *Structure 4* (5), 581-597.
82. Jaschke, P. R.; Saer, R. G.; Noll, S.; Beatty, J. T., Chapter twenty-three - Modification of the Genome of Rhodobacter sphaeroides and Construction of Synthetic Operons. In *Methods in Enzymology*, Chris, V., Ed. Academic Press: 2011; Vol. Volume 497, pp 519-538.
83. Scheuring, S.; Nevo, R.; Liu, L.-N.; Mangenot, S.; Charuvi, D.; Boudier, T.; Prima, V.; Hubert, P.; Sturgis, J. N.; Reich, Z., The architecture of Rhodobacter sphaeroides chromatophores. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2014,** *1837* (8), 1263-1270.
84. Walz, T.; Jamieson, S. J.; Bowers, C. M.; Bullough, P. A.; Hunter, C. N., Projection structures of three photosynthetic complexes from Rhodobacter sphaeroides: LH2 at 6 Å, LH1 and RC-LH1 at 25 Å. *Journal of molecular biology* **1998,** *282* (4), 833-845.
85. Theiler, R.; Suter, F.; Zuber, H.; Cogdell, R. J., A comparison of the primary structures of the two B800-850-apoproteins from wild-type Rhodopseudomonas sphaeroides strain 2.4.1 and a carotenoidless mutant strain R26.1. *FEBS Letters* **1984,** *175* (2), 231-237.
86. Choudhary, M.; Kaplan, S., DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2.4.1(T). *Nucleic Acids Research* **2000,** *28* (4), 862-867.
87. Zeng, X.; Choudhary, M.; Kaplan, S., A Second and Unusual pucBA Operon of Rhodobacter sphaeroides 2.4.1: Genetics and Function of the Encoded Polypeptides. *Journal of Bacteriology* **2003,** *185* (20), 6171-6184.
88. Wang, W.; Hu, Z.; Li, J.; Chen, G., Expression characterization and actual function of the second pucBA in Rhodobacter sphaeroides. *Bioscience Reports* **2009,** *29* (3), 165-172.
89. Woronowicz, K.; Olubanjo, O. B.; Sung, H. C.; Lamptey, J. L.; Niederman, R. A., Differential assembly of polypeptides of the light-harvesting 2 complex encoded by distinct operons during acclimation of Rhodobacter sphaeroides to low light intensity. *Photosynthesis Research* **2011,** *108* (2), 201-214.
90. Tadros, M.; Waterkamp, K., Multiple copies of the coding regions for the light-harvesting B800-850 alpha-and beta-polypeptides are present in the Rhodopseudomonas palustris genome. *The EMBO journal* **1989,** *8* (5), 1303.
91. Gardiner, A. T.; MacKenzie, R. C.; Barrett, S. J.; Kaiser, K.; Cogdell, R. J., The purple photosynthetic bacterium Rhodopseudomonas acidophila contains multiple puc peripheral antenna complex (LH2) genes: Cloning and initial characterisation of four β/α pairs. *Photosynthesis research* **1996,** *49* (3), 223-235.
92. Whitelegge, J. P.; Gundersen, C. B.; Faull, K. F., Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Science : A Publication of the Protein Society* **1998,** *7* (6), 1423-1430.
93. Ryan, C. M.; Souda, P.; Bassilian, S.; Ujwal, R.; Zhang, J.; Abramson, J.; Ping, P.; Durazo, A.; Bowie, J. U.; Hasan, S. S.; Baniulis, D.; Cramer, W. A.; Faull, K. F.; Whitelegge, J. P., Post-translational Modifications of Integral Membrane Proteins Resolved by Top-down Fourier Transform Mass Spectrometry with Collisionally Activated Dissociation. *Molecular & Cellular Proteomics* **2010,** *9* (5), 791-803.

94. Zabrouskov, V.; Whitelegge, J. P., Increased Coverage in the Transmembrane Domain with Activated-Ion Electron Capture Dissociation for Top-Down Fourier-Transform Mass Spectrometry of Integral Membrane Proteins. *Journal of Proteome Research* **2007,** *6* (6), 2205-2210.
95. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (2), 213-221.
96. Croce, R.; van Amerongen, H., Natural strategies for photosynthetic light harvesting. *Nat Chem Biol* **2014,** *10* (7), 492-501.
97. Jackson, P. J.; Lewis, H. J.; Tucker, J. D.; Hunter, C. N.; Dickman, M. J., Quantitative proteomic analysis of intracytoplasmic membrane development in Rhodobacter sphaeroides. *Molecular Microbiology* **2012,** *84* (6), 1062-1078.
98. Woronowicz, K.; Harrold, J. W.; Kay, J. M.; Niederman, R. A., Structural and functional proteomics of intracytoplasmic membrane assembly in Rhodobacter sphaeroides. *Journal of molecular microbiology and biotechnology* **2013,** *23* (1-2), 48-62.
99. Fiedor, L.; Kania, A.; Myśliwa-Kurdziel, B.; Orzeł, Ł.; Stochel, G., Understanding chlorophylls: Central magnesium ion and phytyl as structural determinants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2008,** *1777* (12), 1491-1500.
100. Kelley, L. A.; Sternberg, M. J. E., Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols* **2009,** *4* (3), 363-371.
101. Mann, M.; Jensen, O. N., Proteomic analysis of post-translational modifications. *Nat Biotech* **2003,** *21* (3), 255-261.
102. Medzihradszky, K. F.; Darula, Z.; Perlson, E.; Fainzilber, M.; Chalkley, R. J.; Ball, H.; Greenbaum, D.; Bogyo, M.; Tyson, D. R.; Bradshaw, R. A.; Burlingame, A. L., O-Sulfonation of Serine and Threonine : Mass Spectrometric Detection and Characterization of a New Posttranslational Modification in Diverse Proteins Throughout the Eukaryotes. *Molecular & Cellular Proteomics* **2004,** *3* (5), 429-440.
103. Liao, G.; Xie, L.; Li, X.; Cheng, Z.; Xie, J., Unexpected extensive lysine acetylation in the trump-card antibiotic producer Streptomyces roseosporus revealed by proteome-wide profiling. *Journal of Proteomics* **2014,** *106*, 260-269.
104. Gómez, S. M.; Nishio, J. N.; Faull, K. F.; Whitelegge, J. P., The Chloroplast Grana Proteome Defined by Intact Mass Measurements from Liquid Chromatography Mass Spectrometry. *Molecular & Cellular Proteomics* **2002,** *1* (1), 46-59.
105. Michel, H.; Griffin, P. R.; Shabanowitz, J.; Hunt, D. F.; Bennett, J., Tandem mass spectrometry identifies sites of three post-translational modifications of spinach light-harvesting chlorophyll protein II. Proteolytic cleavage, acetylation, and phosphorylation. *Journal of Biological Chemistry* **1991,** *266* (26), 17584-17591.
106. Stauber, E. J.; Fink, A.; Markert, C.; Kruse, O.; Johanningmeier, U.; Hippler, M., Proteomics of Chlamydomonas reinhardtii Light-Harvesting Proteins. *Eukaryotic Cell* **2003,** *2* (5), 978-994.
107. Berkner, K. L.; Runge, K. W., The physiology of vitamin K nutriture and vitamin K-dependent protein function in atherosclerosis. *Journal of Thrombosis and Haemostasis* **2004,** *2* (12), 2118-2132.

108. Lee, T.-Y.; Lu, C.-T.; Chen, S.-A.; Bretaña, N. A.; Cheng, T.-H.; Su, M.-G.; Huang, K.-Y., Investigation and identification of protein γ-glutamyl carboxylation sites. *BMC Bioinformatics* **2011,** *12* (13), S10.
109. Whitelegge, J. P.; Zhang, H.; Aguilera, R.; Taylor, R. M.; Cramer, W. A., Full Subunit Coverage Liquid Chromatography Electrospray Ionization Mass Spectrometry (LCMS+) of an Oligomeric Membrane Protein: Cytochrome b6f Complex From Spinach and the Cyanobacterium Mastigocladus Laminosus. *Molecular & Cellular Proteomics* **2002,** *1* (10), 816-827.
110. Apostol, I.; Levine, J.; Lippincott, J.; Leach, J.; Hess, E.; Glascock, C. B.; Weickert, M. J.; Blackmore, R., Incorporation of Norvaline at Leucine Positions in Recombinant Human Hemoglobin Expressed in Escherichia coli. *Journal of Biological Chemistry* **1997,** *272* (46), 28980-28988.
111. Kisumi, M.; Sugiura, M.; Chibata, I., Biosynthesis of norvaline, norleucine, and homoisoleucine in Serratia marcescens. *The Journal of Biochemistry* **1976,** *80* (2), 333-339.
112. Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y., Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **2004,** *119* (7), 941-953.
113. Krause, F., Detection and analysis of protein–protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (Membrane) protein complexes and supercomplexes. *ELECTROPHORESIS* **2006,** *27* (13), 2759-2781.
114. Pan, J.; Misamore, M. J.; Wang, Q.; Snell, W. J., Protein Transport and Signal Transduction During Fertilization in Chlamydomonas. *Traffic* **2003,** *4* (7), 452-459.
115. Wallin, E.; Heijne, G. V., Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science* **1998,** *7* (4), 1029-1038.
116. Grimm, D.; Bauer, J.; Pietsch, J.; Infanger, M.; Eucker, J.; Eilles, C.; Schoenberger, J., Diagnostic and Therapeutic Use of Membrane Proteins in Cancer Cells. *Current Medicinal Chemistry* **2011,** *18* (2), 176-190.
117. Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S., Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology* **2008,** *18* (5), 581-586.
118. http://blanco.biomol.uci.edu/mpstruc/.
119. Seddon, A. M.; Curnow, P.; Booth, P. J., Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2004,** *1666* (1–2), 105-117.
120. Rice, A. J.; Alvarez, F. J. D.; Davidson, A. L.; Pinkett, H. W., Effects of lipid environment on the conformational changes of an ABC importer. *Channels* **2014,** *8* (4), 327-333.
121. Denisov, I. G.; Grinkova, Y. V.; Lazarides, A. A.; Sligar, S. G., Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *Journal of the American Chemical Society* **2004,** *126* (11), 3477-3487.
122. Shen, H.-H.; Lithgow, T.; Martin, L., Reconstitution of Membrane Proteins into Model Membranes: Seeking Better Ways to Retain Protein Activities. *International Journal of Molecular Sciences* **2013,** *14* (1), 1589.
123. von Heijne, G., Membrane-protein topology. *Nat Rev Mol Cell Biol* **2006,** *7* (12), 909-918.
124. Rajabi, K.; Ashcroft, A. E.; Radford, S. E., Mass spectrometric methods to analyze the structural organization of macromolecular complexes. *Methods* **2015,** *89*, 13-21.

125. Bennett, B. C.; Gardberg, A. S.; Blair, M. D.; Dealwis, C. G., On the determinants of amide backbone exchange in proteins: a neutron crystallographic comparative study. *Acta Crystallographica Section D* **2008,** *64* (7), 764-783.
126. Busenlehner, L. S.; Codreanu, S. G.; Holm, P. J.; Bhakat, P.; Hebert, H.; Morgenstern, R.; Armstrong, R. N., Stress Sensor Triggers Conformational Response of the Integral Membrane Protein Microsomal Glutathione Transferase 1. *Biochemistry* **2004,** *43* (35), 11145-11152.
127. Zhang, X.; Chien, E. Y. T.; Chalmers, M. J.; Pascal, B. D.; Gatchalian, J.; Stevens, R. C.; Griffin, P. R., Dynamics of the β2-Adrenergic G-Protein Coupled Receptor Revealed by Hydrogen−Deuterium Exchange. *Analytical Chemistry* **2010,** *82* (3), 1100-1108.
128. Xu, G.; Chance, M. R., Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chemical Reviews* **2007,** *107* (8), 3514-3543.
129. 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. *Journal of the American Chemical Society* **2012,** *134* (45), 18724-18731.
130. 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). *Journal of The American Society for Mass Spectrometry* **2015,** *26* (5), 843-846.
131. Zhu, Y.; Guo, T.; Park, J. E.; Li, X.; Meng, W.; Datta, A.; Bern, M.; Lim, S. K.; Sze, S. K., Elucidating in Vivo Structural Dynamics in Integral Membrane Protein by Hydroxyl Radical Footprinting. *Molecular & Cellular Proteomics* **2009,** *8* (8), 1999-2010.
132. Pan, Y.; Brown, L.; Konermann, L., Mapping the Structure of an Integral Membrane Protein under Semi-Denaturing Conditions by Laser-Induced Oxidative Labeling and Mass Spectrometry. *Journal of Molecular Biology* **2009,** *394* (5), 968-981.
133. Bavro, Vassiliy N.; Gupta, S.; Ralston, C., Oxidative footprinting in the study of structure and function of membrane proteins: current state and perspectives. *Biochemical Society Transactions* **2015,** *43* (5), 983-994.
134. Angel, T. E.; Gupta, S.; Jastrzebska, B.; Palczewski, K.; Chance, M. R., Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proceedings of the National Academy of Sciences* **2009,** *106* (34), 14367-14372.
135. Cogdell, R. J.; Fyfe, P. K.; Barrett, S. J.; Prince, S. M.; Freer, A. A.; Isaacs, N. W.; McGlynn, P.; Hunter, C. N., The purple bacterial photosynthetic unit. *Photosynthesis Research 48* (1), 55-63.
136. Chandler, D. E.; Strümpfer, J.; Sener, M.; Scheuring, S.; Schulten, K., Light Harvesting by Lamellar Chromatophores in Rhodospirillum photometricum. *Biophysical Journal* **2014,** *106* (11), 2503-2510.
137. Ritchie, T. K.; Grinkova, Y. V.; Bayburt, T. H.; Denisov, I. G.; Zolnerciks, J. K.; Atkins, W. M.; Sligar, S. G., Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods in enzymology* **2009,** *464*, 211-231.
138. Clayton, R. K.; Clayton, B. J., B850 pigment-protein complex of Rhodopseudomonas sphaeroides: Extinction coefficients, circular dichroism, and the reversible binding of bacteriochlorophyll. *Proceedings of the National Academy of Sciences* **1981,** *78* (9), 5583-5587.
139. Sligar, S. G. http://sligarlab.life.uiuc.edu/nanodisc/protocols.html.
140. Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L., Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Analytical Chemistry* **2009,** *81* (16), 6563-6571.
141. Consortium, T. U., UniProt: a hub for protein information. *Nucleic Acids Research* **2015,** *43* (D1), D204-D212.

142. Tsirigos, K. D.; Peters, C.; Shu, N.; Käll, L.; Elofsson, A., The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Research* **2015**.
143. Schrödinger, L., New York., The PyMOL Molecular Graphics System, version 1.7.4.
144. Chen, X.-H.; Zhang, L.; Weng, Y.-X.; Du, L.-C.; Ye, M.-P.; Yang, G.-Z.; Fujii, R.; Rondonuwu, F. S.; Koyama, Y.; Wu, Y.-S.; Zhang, J. P., Protein Structural Deformation Induced Lifetime Shortening of Photosynthetic Bacteria Light-Harvesting Complex LH2 Excited State. *Biophysical Journal* **2005,** *88* (6), 4262-4273.
145. Inagaki, S.; Ghirlando, R.; Grisshammer, R., Biophysical characterization of membrane proteins in nanodiscs. *Methods* **2013,** *59* (3), 287-300.
146. Hagn, F.; Etzkorn, M.; Raschle, T.; Wagner, G., Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *Journal of the American Chemical Society* **2013,** *135* (5), 1919-1925.
147. Inagaki, S.; Ghirlando, R.; White, J. F.; Gvozdenovic-Jeremic, J.; Northup, J. K.; Grisshammer, R., Modulation of the Interaction between Neurotensin Receptor NTS1 and Gq Protein by Lipid. *Journal of Molecular Biology* **2012,** *417* (1–2), 95-111.
148. Boldog, T.; Grimme, S.; Li, M.; Sligar, S. G.; Hazelbauer, G. L., Nanodiscs separate chemoreceptor oligomeric states and reveal their signaling properties. *Proceedings of the National Academy of Sciences* **2006,** *103* (31), 11509-11514.
149. Denisov, I. G.; Baas, B. J.; Grinkova, Y. V.; Sligar, S. G., Cooperativity in cytochrome P450 3A4: Linkages in substrate binding, spin state, uncoupling, and product formation. *Journal of Biological Chemistry* **2007,** *282* (10), 7066-7076.
150. Bayburt, T. H.; Sligar, S. G., Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers. *Protein Science* **2003,** *12* (11), 2476-2481.
151. Bayburt, T. H.; Vishnivetskiy, S. A.; McLean, M. A.; Morizumi, T.; Huang, C. C.; Tesmer, J. J. G.; Ernst, O. P.; Sligar, S. G.; Gurevich, V. V., Monomeric rhodopsin is sufficient for normal rhodopsin kinase (GRK1) phosphorylation and arrestin-1 binding. *Journal of Biological Chemistry* **2011,** *286* (2), 1420-1428.
152. Bibow, S.; Carneiro, M. G.; Sabo, T. M.; Schwiegk, C.; Becker, S.; Riek, R.; Lee, D., Measuring membrane protein bond orientations in nanodiscs via residual dipolar couplings. *Protein Science* **2014,** *23* (7), 851-856.
153. Hebling, C. M.; Morgan, C. R.; Stafford, D. W.; Jorgenson, J. W.; Rand, K. D.; Engen, J. R., Conformational Analysis of Membrane Proteins in Phospholipid Bilayer Nanodiscs by Hydrogen Exchange Mass Spectrometry. *Analytical Chemistry* **2010,** *82* (13), 5415-5419.
154. Morgan, C. R.; Hebling, C. M.; Rand, K. D.; Stafford, D. W.; Jorgenson, J. W.; Engen, J. R., Conformational Transitions in the Membrane Scaffold Protein of Phospholipid Bilayer Nanodiscs. *Molecular & Cellular Proteomics* **2011,** *10* (9).
155. Treuheit, N. A.; Redhair, M.; Kwon, H.; McClary, W. D.; Guttman, M.; Sumida, J. P.; Atkins, W. M., Membrane Interactions, Ligand-Dependent Dynamics, and Stability of Cytochrome P4503A4 in Lipid Nanodiscs. *Biochemistry* **2016,** *55* (7), 1058-1069.
156. Yoon, H.-H.; Lee, M.-S.; Kang, J.-H., Reaction of ferritin with hydrogen peroxide induces lipid peroxidation. *BMB reports* **2010,** *43* (3), 219-224.
157. Repetto, M.; Boveris, A.; Semprine, J., Lipid peroxidation: chemical mechanism, biological implications and analytical determination. **2012**.

158. Debnath, A.; Schäfer, L. V., Structure and Dynamics of Phospholipid Nanodiscs from All-Atom and Coarse-Grained Simulations. *The Journal of Physical Chemistry B* **2015,** *119* (23), 6991-7002.
159. Bogusz, S.; Venable, R. M.; Pastor, R. W., Molecular Dynamics Simulations of Octyl Glucoside Micelles: Structural Properties. *The Journal of Physical Chemistry B* **2000,** *104* (23), 5462-5470.
160. Garavito, R. M.; Ferguson-Miller, S., Detergents as Tools in Membrane Biochemistry. *Journal of Biological Chemistry* **2001,** *276* (35), 32403-32406.
161. Bond, P. J.; Cuthbertson, J.; Sansom, M. S. P., Simulation studies of the interactions between membrane proteins and detergents. *Biochemical Society Transactions* **2005,** *33* (5), 910-912.
162. Bond, P. J.; Sansom, M. S. P., Membrane Protein Dynamics versus Environment: Simulations of OmpA in a Micelle and in a Bilayer. *Journal of Molecular Biology* **2003,** *329* (5), 1035-1053.
163. Kania, A.; Fiedor, L., Steric Control of Bacteriochlorophyll Ligation. *Journal of the American Chemical Society* **2006,** *128* (2), 454-458.
164. Singer, S. J.; Nicolson, G. L., The fluid mosaic model of the structure of cell membranes. *Science* **1972,** *175* (4023), 720-731.
165. Lee, A. G., Lipid–protein interactions in biological membranes: a structural perspective. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2003,** *1612* (1), 1-40.
166. Siuda, I.; Tieleman, D. P., Molecular Models of Nanodiscs. *Journal of Chemical Theory and Computation* **2015,** *11* (10), 4923-4932.
167. Olsen, J. D.; Tucker, J. D.; Timney, J. A.; Qian, P.; Vassilev, C.; Hunter, C. N., The Organization of LH2 Complexes in Membranes from Rhodobacter sphaeroides. *Journal of Biological Chemistry* **2008,** *283* (45), 30772-30779.
168. Hennerdal, A.; Elofsson, A., Rapid membrane protein topology prediction. *Bioinformatics* **2011,** *27* (9), 1322-1323.
169. Gall, A.; Gardiner, A. T.; Cogdell, R. J.; Robert, B., Carotenoid stoichiometry in the LH2 crystal: No spectral evidence for the presence of the second molecule in the α/β-apoprotein dimer. *FEBS Letters* **2006,** *580* (16), 3841-3844.
170. Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K., Coarse Grained Protein−Lipid Model with Application to Lipoprotein Particles. *The Journal of Physical Chemistry B* **2006,** *110* (8), 3674-3684.
171. Tan, S.; Tan, H. T.; Chung, M. C. M., Membrane proteins and membrane proteomics. *PROTEOMICS* **2008,** *8* (19), 3924-3932.
172. Rawson, S.; Davies, S.; Lippiat, J. D.; Muench, S. P., The changing landscape of membrane protein structural biology through developments in electron microscopy. *Molecular Membrane Biology* **2016,** *33* (1-2), 12-22.
173. von Heijne, G.; Manoil, C., Membrane proteins: from sequence to structure. *Protein Engineering, Design and Selection* **1990,** *4* (2), 109-112.
174. Deisenhofer, J.; Michel, H., The Photosynthetic Reaction Center from the Purple Bacterium Rhodopseudomonas viridis (Nobel Lecture). *Angewandte Chemie International Edition in English* **1989,** *28* (7), 829-847.
175. .

176. Moraes, I.; Evans, G.; Sanchez-Weatherby, J.; Newstead, S.; Stewart, P. D. S., Membrane protein structure determination — The next generation. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2014,** *1838* (1, Part A), 78-87.
177. Cong, X.; Liu, Y.; Liu, W.; Liang, X.; Russell, D. H.; Laganowsky, A., Determining Membrane Protein–Lipid Binding Thermodynamics Using Native Mass Spectrometry. *Journal of the American Chemical Society* **2016,** *138* (13), 4346-4349.
178. Konijnenberg, A.; van Dyck Jeroen, F.; Kailing Lyn, L.; Sobott, F., Extending native mass spectrometry approaches to integral membrane proteins. In *Biol. Chem.*, 2015; Vol. 396, p 991.
179. Johansson, L. C.; Arnlund, D.; Katona, G.; White, T. A.; Barty, A.; DePonte, D. P.; Shoeman, R. L.; Wickstrand, C.; Sharma, A.; Williams, G. J.; Aquila, A.; Bogan, M. J.; Caleman, C.; Davidsson, J.; Doak, R. B.; Frank, M.; Fromme, R.; Galli, L.; Grotjohann, I.; Hunter, M. S.; Kassemeyer, S.; Kirian, R. A.; Kupitz, C.; Liang, M.; Lomb, L.; Malmerberg, E.; Martin, A. V.; Messerschmidt, M.; Nass, K.; Redecke, L.; Seibert, M. M.; Sjöhamn, J.; Steinbrener, J.; Stellato, F.; Wang, D.; Wahlgren, W. Y.; Weierstall, U.; Westenhoff, S.; Zatsepin, N. A.; Boutet, S.; Spence, J. C. H.; Schlichting, I.; Chapman, H. N.; Fromme, P.; Neutze, R., Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography. *Nature Communications* **2013,** *4*, 2911.
180. Sharma, A.; Johansson, L.; Dunevall, E.; Wahlgren, W.; Neutze, R.; Katona, G., Asymmetry in serial femtosecond crystallography data. *Acta Crystallographica Section A: Foundations and Advances* **2017,** *73* (2), 93-101.
181. Johansson, L. C.; Arnlund, D.; White, T. A.; Katona, G.; DePonte, D. P.; Weierstall, U.; Doak, R. B.; Shoeman, R. L.; Lomb, L.; Malmerberg, E.; Davidsson, J.; Nass, K.; Liang, M.; Andreasson, J.; Aquila, A.; Bajt, S.; Barthelmess, M.; Barty, A.; Bogan, M. J.; Bostedt, C.; Bozek, J. D.; Caleman, C.; Coffee, R.; Coppola, N.; Ekeberg, T.; Epp, S. W.; Erk, B.; Fleckenstein, H.; Foucar, L.; Graafsma, H.; Gumprecht, L.; Hajdu, J.; Hampton, C. Y.; Hartmann, R.; Hartmann, A.; Hauser, G.; Hirsemann, H.; Holl, P.; Hunter, M. S.; Kassemeyer, S.; Kimmel, N.; Kirian, R. A.; Maia, F. R. N. C.; Marchesini, S.; Martin, A. V.; Reich, C.; Rolles, D.; Rudek, B.; Rudenko, A.; Schlichting, I.; Schulz, J.; Seibert, M. M.; Sierra, R. G.; Soltau, H.; Starodub, D.; Stellato, F.; Stern, S.; Struder, L.; Timneanu, N.; Ullrich, J.; Wahlgren, W. Y.; Wang, X.; Weidenspointner, G.; Wunderer, C.; Fromme, P.; Chapman, H. N.; Spence, J. C. H.; Neutze, R., Lipidic phase membrane protein serial femtosecond crystallography. *Nat Meth* **2012,** *9* (3), 263-265.
182. Magdaong, N. C. M.; Niedzwiedzki, D. M.; Goodson, C.; Blankenship, R. E., Carotenoid-to-Bacteriochlorophyll Energy Transfer in the LH1–RC Core Complex of a Bacteriochlorophyll b Containing Purple Photosynthetic Bacterium Blastochloris viridis. *The Journal of Physical Chemistry B* **2016,** *120* (23), 5159-5171.
183. Morgner, N.; Robinson, C. V., Massign: An Assignment Strategy for Maximizing Information from the Mass Spectra of Heterogeneous Protein Assemblies. *Analytical Chemistry* **2012,** *84* (6), 2939-2948.
184. Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V., Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protocols* **2008,** *3* (7), 1139-1152.
185. Michaelevski, I.; Kirshenbaum, N.; Sharon, M., T-wave ion mobility-mass spectrometry: basic experimental procedures for protein complex analysis. *J Vis Exp* **2010,** *41*.

186. Liu, L.-N.; Faulkner, M.; Liu, X.; Huang, F.; Darby, A. C.; Hall, N., Revised Genome Sequence of the Purple Photosynthetic Bacterium Blastochloris viridis. *Genome Announcements* **2016,** *4* (1).
187. Roszak, Aleksander W.; Moulisová, V.; Reksodipuro, Adhie D. P.; Gardiner, Alastair T.; Fujii, R.; Hashimoto, H.; Isaacs, Neil W.; Cogdell, Richard J., New insights into the structure of the reaction centre from Blastochloris viridis: evolution in the laboratory. *Biochem. J* **2012,** *442* (1), 27-37.
188. Table 1. Selected chemical and enzymic cleavage reagents for protein chains. *Cold Spring Harbor Protocols* **2007,** *2007* (4), pdb.tab1ip13.
189. Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B., PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Molecular & Cellular Proteomics* **2012,** *11* (4).
190. Blankenship, R. E., Origin and Evolution of Photosynthesis. In *Molecular Mechanisms of Photosynthesis*, Blackwell Science Ltd: 2008; pp 220-257.
191. Heck, A. J. R., Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Meth* **2008,** *5* (11), 927-933.
192. Zhang, H.; Harrington, L. B.; Lu, Y.; Prado, M.; Saer, R.; Rempel, D.; Blankenship, R. E.; Gross, M. L., Native Mass Spectrometry Characterizes the Photosynthetic Reaction Center Complex from the Purple Bacterium Rhodobacter sphaeroides. *Journal of The American Society for Mass Spectrometry* **2017,** *28* (1), 87-95.
193. el-Kabbani, O.; Chang, C. H.; Tiede, D.; Norris, J.; Schiffer, M., Comparison of reaction centers from Rhodobacter sphaeroides and Rhodopseudomonas viridis: overall architecture and protein-pigment interactions. *Biochemistry* **1991,** *30* (22), 5361-9.
194. Valentine, S. J.; Clemmer, D. E., Temperature-dependent H/D exchange of compact and elongated cytochrome c ions in the gas phase. *Journal of the American Society for Mass Spectrometry 13* (5), 506-517.
195. Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., Naked Protein Conformations: Cytochrome c in the Gas Phase. *Journal of the American Chemical Society* **1995,** *117* (40), 10141-10142.
196. Hall, Z.; Robinson, C. V., Do Charge State Signatures Guarantee Protein Conformations? *Journal of The American Society for Mass Spectrometry* **2012,** *23* (7), 1161-1168.
197. and, D. S. G.; Olson, A. J., Structural Symmetry and Protein Function. *Annual Review of Biophysics and Biomolecular Structure* **2000,** *29* (1), 105-153.
198. Ali, M. H.; Imperiali, B., Protein oligomerization: How and why. *Bioorganic & Medicinal Chemistry* **2005,** *13* (17), 5013-5020.
199. Jiang, J.; Zhang, H.; Orf, G. S.; Lu, Y.; Xu, W.; Harrington, L. B.; Liu, H.; Lo, C. S.; Blankenship, R. E., Evidence of functional trimeric chlorophyll a/c2-peridinin proteins in the dinoflagellate Symbiodinium. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2014,** *1837* (11), 1904-1912.
200. D'Amici, G. M.; Rinalducci, S.; Murgiano, L.; Italiano, F.; Zolla, L., Oligomeric Characterization of the Photosynthetic Apparatus of Rhodobacter sphaeroides R26.1 by Nondenaturing Electrophoresis Methods. *Journal of Proteome Research* **2010,** *9* (1), 192-203.
201. Nishiyama, Y.; Allakhverdiev, S. I.; Murata, N., A new paradigm for the action of reactive oxygen species in the photoinhibition of photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2006,** *1757* (7), 742-749.

202. Kirilovsky, D., Modulating energy arriving at photochemical reaction centers: orange carotenoid protein-related photoprotection and state transitions. *Photosynthesis Research* **2015,** *126* (1), 3-17.
203. Kerfeld, C. A.; Kirilovsky, D., Structural, mechanistic and genomic insights into OCP-mediated photoprotection. *Advances in Botanical Research* **2013**.
204. Kerfeld, C. A.; Sawaya, M. R.; Brahmandam, V.; Cascio, D.; Ho, K. K.; Trevithick-Sutton, C. C.; Krogmann, D. W.; Yeates, T. O., The Crystal Structure of a Cyanobacterial Water-Soluble Carotenoid Binding Protein. *Structure* **2003,** *11* (1), 55-65.
205. Wilson, A.; Kinney, J. N.; Zwart, P. H.; Punginelli, C.; D'Haene, S.; Perreau, F.; Klein, M. G.; Kirilovsky, D.; Kerfeld, C. A., Structural Determinants Underlying Photoprotection in the Photoactive Orange Carotenoid Protein of Cyanobacteria. *Journal of Biological Chemistry* **2010,** *285* (24), 18364-18375.
206. Gwizdala, M.; Wilson, A.; Omairi-Nasser, A.; Kirilovsky, D., Characterization of the Synechocystis PCC 6803 Fluorescence Recovery Protein involved in photoprotection. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2013,** *1827* (3), 348-354.
207. Sutter, M.; Wilson, A.; Leverenz, R. L.; Lopez-Igual, R.; Thurotte, A.; Salmeen, A. E.; Kirilovsky, D.; Kerfeld, C. A., Crystal structure of the FRP and identification of the active site for modulation of OCP-mediated photoprotection in cyanobacteria. *Proceedings of the National Academy of Sciences* **2013,** *110* (24), 10022-10027.
208. Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V., Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chemical Reviews* **2007,** *107* (8), 3544-3567.
209. Marcoux, J.; Robinson, Carol V., Twenty Years of Gas Phase Structural Biology. *Structure* **2013,** *21* (9), 1541-1550.
210. Chen, L. H.; Kenyon, G. L.; Curtin, F.; Harayama, S.; Bembenek, M. E.; Hajipour, G.; Whitman, C. P., 4-Oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer. *Journal of Biological Chemistry* **1992,** *267* (25), 17716-17721.
211. Roper, D. I.; Subramanya, H. S.; Shingler, V.; Wigley, D. B., Preliminary crystallographic analysis of 4-oxalocrotonate tautomerase reveals the oligomeric structure of the enzyme. *Journal of Molecular Biology* **1994,** *243* (4), 799-801.
212. Fitzgerald, M. C.; Chernushevich, I.; Standing, K. G.; Whitman, C. P.; Stephen, B. H. K., Probing the Oligomeric Structure of an Enzyme by Electrospray Ionization Time-Of-Flight Mass Spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **1996,** *93* (14), 6851-6856.
213. Smith, R. D.; Light-Wahl, K. J., The observation of non-covalent interactions in solution by electrospray ionization mass spectrometry: Promise, pitfalls and prognosis. *Biological Mass Spectrometry* **1993,** *22* (9), 493-501.
214. Heck, A. J. R.; van den Heuvel, R. H. H., Investigation of intact protein complexes by mass spectrometry. *Mass Spectrometry Reviews* **2004,** *23* (5), 368-389.
215. Brange, J.; Andersen, L.; Laursen, E. D.; Meyn, G.; Rasmussen, E., Toward Understanding Insulin Fibrillation. *Journal of Pharmaceutical Sciences* **1997,** *86* (5), 517-525.
216. Nettleton, E. J.; Tito, P.; Sunde, M.; Bouchard, M.; Dobson, C. M.; Robinson, C. V., Characterization of the Oligomeric States of Insulin in Self-Assembly and Amyloid Fibril Formation by Mass Spectrometry. *Biophysical Journal* **2000,** *79* (2), 1053-1065.
217. Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H., Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison

with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Communications in Mass Spectrometry* **2008,** *22* (20), 3297-3304.
218. Wang, S. C.; Politis, A.; Di Bartolo, N.; Bavro, V. N.; Tucker, S. J.; Booth, P. J.; Barrera, N. P.; Robinson, C. V., Ion Mobility Mass Spectrometry of Two Tetrameric Membrane Protein Complexes Reveals Compact Structures and Differences in Stability and Packing. *Journal of the American Chemical Society* **2010,** *132* (44), 15468-15470.
219. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* **2014,** *6* (4), 281-294.
220. Hall, Z.; Politis, A.; Robinson, Carol V., Structural Modeling of Heteromeric Protein Complexes from Disassembly Pathways and Ion Mobility-Mass Spectrometry. *Structure* **2012,** *20* (9), 1596-1609.
221. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A., ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **2003,** *31* (13), 3784-3788.
222. Kebarle, P.; Tang, L., From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. *Analytical Chemistry* **1993,** *65* (22), 972A-986A.
223. Rayleigh, L., XX. On the equilibrium of liquid conducting masses charged with electricity. *Philosophical Magazine Series 5* **1882,** *14* (87), 184-186.
224. Sluchanko, N. N.; Klementiev, K. E.; Shirshin, E. A.; Tsoraev, G. V.; Friedrich, T.; Maksimov, E. G., The purple Trp288Ala mutant of Synechocystis OCP persistently quenches phycobilisome fluorescence and tightly interacts with FRP. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2017,** *1858* (1), 1-11.
225. Hong, P.; Koza, S.; Bouvier, E. S., A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *Journal of liquid chromatography & related technologies* **2012,** *35* (20), 2923-2950.
226. Mack, E., AVERAGE CROSS-SECTIONAL AREAS OF MOLECULES BY GASEOUS DIFFUSION METHODS. *Journal of the American Chemical Society* **1925,** *47* (10), 2468-2482.
227. Shvartsburg, A. A.; Jarrold, M. F., An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters* **1996,** *261* (1–2), 86-91.
228. Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V., Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. *Journal of the American Chemical Society* **2012,** *134* (7), 3429-3438.
229. Benesch, J. L. P.; Ruotolo, B. T., Mass spectrometry: come of age for structural and dynamical biology. *Current Opinion in Structural Biology* **2011,** *21* (5), 641-649.
230. Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V., Ion Mobility–Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. *Angewandte Chemie International Edition* **2007,** *46* (42), 8001-8004.
231. Nussinov, R.; Xu, D.; Tsai, C.-J., Mechanism and evolution of protein dimerization. *Protein Science* **1998,** *7* (3), 533-544.
232. Niyogi, K. K.; Truong, T. B., Evolution of flexible non-photochemical quenching mechanisms that regulate light harvesting in oxygenic photosynthesis. *Current Opinion in Plant Biology* **2013,** *16* (3), 307-314.

233. Latowski, D.; Kuczyńska, P.; Strzałka, K., Xanthophyll cycle – a mechanism protecting plants against oxidative stress. *Redox Report* **2011,** *16* (2), 78-90.
234. Wilson, A.; Ajlani, G.; Verbavatz, J.-M.; Vass, I.; Kerfeld, C. A.; Kirilovsky, D., A Soluble Carotenoid Protein Involved in Phycobilisome-Related Energy Dissipation in Cyanobacteria. *The Plant cell* **2006,** *18* (4), 992-1007.
235. Leverenz, R. L.; Sutter, M.; Wilson, A.; Gupta, S.; Thurotte, A.; Bourcier de Carbon, C.; Petzold, C. J.; Ralston, C.; Perreau, F.; Kirilovsky, D.; Kerfeld, C. A., A 12 Å carotenoid translocation in a photoswitch associated with cyanobacterial photoprotection. *Science* **2015,** *348* (6242), 1463-1466.
236. Maksimov, E. G.; Sluchanko, N. N.; Mironov, K. S.; Shirshin, E. A.; Klementiev, K. E.; Tsoraev, G. V.; Moldenhauer, M.; Friedrich, T.; Los, D. A.; Allakhverdiev, S. I.; Paschenko, V. Z.; Rubin, A. B., Fluorescent Labeling Preserving OCP Photoactivity Reveals Its Reorganization during the Photocycle. *Biophysical Journal* **2017,** *112* (1), 46-56.
237. Moldenhauer, M.; Sluchanko, N. N.; Tavraz, N. N.; Junghans, C.; Buhrke, D.; Willoweit, M.; Chiappisi, L.; Schmitt, F.-J.; Vukojević, V.; Shirshin, E. A.; Ponomarev, V. Y.; Paschenko, V. Z.; Gradzielski, M.; Maksimov, E. G.; Friedrich, T., Interaction of the signaling state analog and the apoprotein form of the orange carotenoid protein with the fluorescence recovery protein. *Photosynthesis Research* **2017**, 1-15.
238. Stadnichuk, I. N.; Krasilnikov, P. M.; Zlenko, D. V.; Freidzon, A. Y.; Yanyushin, M. F.; Rubin, A. B., Electronic coupling of the phycobilisome with the orange carotenoid protein and fluorescence quenching. *Photosynthesis Research* **2015,** *124* (3), 315-335.
239. 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. *Journal of Biomolecular Structure and Dynamics* **2016,** *34* (3), 486-496.
240. Harris, D.; Tal, O.; Jallet, D.; Wilson, A.; Kirilovsky, D.; Adir, N., Orange carotenoid protein burrows into the phycobilisome to provide photoprotection. *Proceedings of the National Academy of Sciences* **2016,** *113* (12), E1655-E1662.
241. Thurotte, A.; de Carbon, C. B.; Wilson, A.; Talbot, L.; Cot, S.; López-Igual, R.; Kirilovsky, D., The cyanobacterial Fluorescence Recovery Protein has two distinct activities: Orange Carotenoid Protein amino acids involved in FRP interaction. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2017,** *1858* (4), 308-317.
242. Gwizdala, M.; Wilson, A.; Kirilovsky, D., In Vitro Reconstitution of the Cyanobacterial Photoprotective Mechanism Mediated by the Orange Carotenoid Protein in Synechocystis PCC 6803. *The Plant cell* **2011,** *23* (7), 2631-2643.
243. Walzthoeni, T.; Leitner, A.; Stengel, F.; Aebersold, R., Mass spectrometry supported determination of protein complex structure. *Current Opinion in Structural Biology* **2013,** *23* (2), 252-260.
244. Konermann, L.; Tong, X.; Pan, Y., Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *Journal of Mass Spectrometry* **2008,** *43* (8), 1021-1036.
245. Sharon, M.; Robinson, C. V., The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. *Annual Review of Biochemistry* **2007,** *76* (1), 167-193.
246. Boeri Erba, E.; Petosa, C., The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. *Protein Science* **2015,** *24* (8), 1176-1192.

247. Zhang, H.; Cui, W.; Gross, M. L., Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Letters* **2014,** *588* (2), 308-317.
248. Zhang, H.; Liu, H.; Lu, Y.; Wolf, N. R.; Gross, M. L.; Blankenship, R. E., Native Mass Spectrometry and Ion Mobility Characterize the Orange Carotenoid Protein Functional Domains. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2016**.
249. Liu, H.; Zhang, H.; Orf, G. S.; Lu, Y.; Jiang, J.; King, J. D.; Wolf, N. R.; Gross, M. L.; Blankenship, R. E., Dramatic Domain Rearrangements of the Cyanobacterial Orange Carotenoid Protein upon Photoactivation. *Biochemistry* **2016,** *55* (7), 1003-1009.
250. Bern, M.; Kil, Y. J.; Becker, C., Byonic: Advanced Peptide and Protein Identification Software. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **2012,** *CHAPTER*, Unit13.20-Unit13.20.
251. Leverenz, R. L.; Jallet, D.; Li, M. D.; Mathies, R. A.; Kirilovsky, D.; Kerfeld, C. A., Structural and functional modularity of the orange carotenoid protein: distinct roles for the N- and C-terminal domains in cyanobacterial photoprotection. *The Plant cell* **2014,** *26* (1), 426-37.
252. Kirilovsky, D.; Kerfeld, C. A., The Orange Carotenoid Protein: a blue-green light photoactive protein. *Photochemical & Photobiological Sciences* **2013,** *12* (7), 1135-1143.
253. Moldenhauer, M.; Sluchanko, N. N.; Buhrke, D.; Zlenko, D. V.; Tavraz, N. N.; Schmitt, F.-J.; Hildebrandt, P.; Maksimov, E. G.; Friedrich, T., Assembly of photoactive Orange Carotenoid Protein from its domains unravels a carotenoid shuttle mechanism. *Photosynthesis Research* **2017**.
254. Popa, V.; Trecroce, D. A.; McAllister, R. G.; Konermann, L., Collision-Induced Dissociation of Electrosprayed Protein Complexes: An All-Atom Molecular Dynamics Model with Mobile Protons. *The Journal of Physical Chemistry B* **2016,** *120* (23), 5114-5124.
255. Goddard, T. D.; Huang, C. C.; Ferrin, T. E., Visualizing density maps with UCSF Chimera. *Journal of Structural Biology* **2007,** *157* (1), 281-287.
256. Wilson, A.; xe; xe; Gwizdala, M.; Mezzetti, A.; Alexandre, M.; Kerfeld, C. A.; Kirilovsky, D., The Essential Role of the N-Terminal Domain of the Orange Carotenoid Protein in Cyanobacterial Photoprotection: Importance of a Positive Charge for Phycobilisome Binding. *The Plant cell* **2012,** *24* (5), 1972-1983.
257. Kahraman, A.; Malmström, L.; Aebersold, R., Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* **2011,** *27* (15), 2163-2164.
258. Thurotte, A.; Lopez Igual, R.; Wilson, A.; Comolet, L.; Bourcier de Carbon, C.; Xiao, F.; Kirilovsky, D., Regulation of Orange Carotenoid Protein activity in cyanobacterial photoprotection. *Plant Physiology* **2015,** *169* (1), 734-47.