1-1-2015

Application and Refinement of a Zero-Length Chemical Cross-Linking and Mass Spectrometry Method to Examine Native Protein Structures

Roland Fernando Rivera-Santiago
University of Pennsylvania, roland.f.rivera.santiago@gmail.com

Application and Refinement of a Zero-Length Chemical Cross-Linking and Mass Spectrometry Method to Examine Native Protein Structures

Abstract
Chemical cross-linking and mass spectrometry (CX-MS) is a structural technique that has been gaining in popularity with the advent of instruments that provide increased access to high-resolution MS/MS data. The Speicher laboratory is interested in zero-length cross-links, which provide the highest-quality data, but are the most difficult to identify and analyze. A MS sample preparation and data analysis pipeline, including a software package called ZXMiner, has been developed to collect and analyze zero-length CX-MS data to use in molecular modeling experiments. I used this pipeline to examine the peroxiredoxin-6 (PRDX6) enzyme, which prevents oxidative damage in the lung. I found the identified cross-links did not fit the published crystal structure of a catalytic intermediate, which suggests PRDX6 undergoes conformational changes as part of catalysis. A solution structure of PRDX6 was created using this CX-MS data. Next, I performed optimization experiments in order to adapt the ZXMiner data analysis pipeline from the LTQ Orbitrap XL instrument it was created for to a more state-of-the-art Q Exactive Plus instrument in order to expand its ability to probe complex samples. Combined with a new version of ZXMiner, I was able to determine several parameters that improved the quality of cross-link data for erythrocyte membrane white ghost (WG) samples. I combined the cross-links identified in those samples with immunoprecipitation experiments and review of the pertinent literature to determine a topology model for the anion exchanger 1 (AE1) protein, which plays critical roles in CO2 transport and the erythrocyte membrane skeleton. Using this topology model and published crystal structures for the N-terminal domain of AE1, I was able to perform combinatorial modeling experiments which resulted in a model of the full-length protein that resolved several controversies in the field. These results show that zero-length CX-MS is a powerful technique to determine structural information of proteins difficult to interrogate through conventional means.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biochemistry & Molecular Biophysics

First Advisor
David W. Speicher

Keywords
Chemical cross-linking, Mass spectrometry, Native structure, Structural biology

Subject Categories
Biochemistry | Biophysics

APPLICATION AND REFINEMENT OF A ZERO-LENGTH CHEMICAL CROSS-LINKING AND MASS SPECTROMETRY METHOD TO EXAMINE NATIVE PROTEIN STRUCTURES

Roland F. Rivera-Santiago

A DISSERTATION

in

Biochemistry & Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

_____

David W. Speicher

Professor

Graduate Group Chairperson

_____

Kim Sharp, Associate Professor

Dissertation Committee

Benjamin Garcia, Associate Professor

Dennis Discher, Professor

Gregory Van Duyne, Professor

Ravi Radhakrishnan, Professor

James Bruce, Professor

Kathryn M. Ferguson, Associate Professor

APPLICATION AND REFINEMENT OF A ZERO-LENGTH CHEMICAL CROSS-LINKING AND MASS SPECTROMETRY METHOD TO EXAMINE NATIVE PROTEIN STRUCTURES

Dedication page

To my beloved parents, without whom none of this would be possible.

To my dearest girlfriend Johanna, whose love and unwavering support carried me through.

To the rest of my family, especially my cousin Papo and my aunts Miriam and Nancy, who spurred me on with their belief.

To Dr. David Speicher, for giving me a second chance.

To Kaye Speicher, for being caring and supportive.

To Dr. Kate Ferguson and Dr. Joshua Wand, for believing in me when I was doubting myself.

To the Speicher lab and Wistar Proteomics Core facility, for making me feel at home.

To my good friend Dr. Christopher Skipwith, for never giving up on me.

To my friends Yakov Feygin, Brian Cole, Leksa Nall, Mike Allegrezza, Ian Magil, Josh Varrone, Anthony Amato, Nicole Kline, Sam Jordan, Kenneth Jackson, Chun Lai, Vince Chin, Albert Ng, Jesse Holz, David Napiorski, and many others, for making sure I never felt alone.

Acknowledgment

I would like to take this opportunity to acknowledge Mr. Peter Hembach and Ms. Sandra Harper of the Speicher lab, who have helped me tremendously with the biochemistry required across the course of this thesis. I would also to acknowledge other members of the Speicher Lab and the Proteomics Core Facility for their invaluable support throughout the entirety of this process. I would also like to thank the Structural Biology Training Grant (T32GM008275) and the Philadelphia Health Care Trust Fellowship for providing me with funding throughout my graduate studies.

Abstract

Application and Refinement of a Zero-Length Chemical Cross-linking and Mass Spectrometry Method to Examine Native Protein Structures

Roland F. Rivera-Santiago

Dr. David W. Speicher

Chemical cross-linking and mass spectrometry (CX-MS) is a structural technique that has been gaining in popularity with the advent of instruments that provide increased access to high-resolution MS/MS data. The Speicher laboratory is interested in zero-length cross-links, which provide the highest-quality data, but are the most difficult to identify and analyze. A MS sample preparation and data analysis pipeline, including a software package called ZXMiner, has been developed to collect and analyze zero-length CX-MS data to use in molecular modeling experiments. I used this pipeline to examine the peroxiredoxin-6 (PRDX6) enzyme, which prevents oxidative damage in the lung. I found the identified cross-links did not fit the published crystal structure of a catalytic intermediate, which suggests PRDX6 undergoes conformational changes as part of catalysis. A solution structure of PRDX6 was created using this CX-MS data. Next, I performed optimization experiments in order to adapt the ZXMiner data analysis pipeline from the LTQ Orbitrap XL instrument it was created for to a more state-of-the-art Q Exactive Plus instrument in order to expand its ability to probe complex samples. Combined with a new version of ZXMiner, I was able to determine several parameters that improved the quality of cross-link data for erythrocyte membrane white ghost (WG) samples. I combined the cross-links identified in those samples with immunoprecipitation experiments and review of the pertinent literature to determine a topology model for the anion exchanger 1 (AE1) protein, which plays critical roles in CO_2 transport and the erythrocyte membrane skeleton. Using this topology model and published crystal structures for the N-terminal domain of AE1, I was able to perform combinatorial modeling experiments which resulted in a model of the full-length protein that resolved several controversies in the field. These results show that zero-length CX-MS is a powerful technique to determine structural information of proteins difficult to interrogate through conventional means.

Table of Contents

Table of Contents

List of Tables

List of Illustrations

Chapter 1: Introduction

In this chapter, I will introduce the underlying concepts necessary to frame the context for the rest of this project. Specifically, I will discuss the theory behind structural mass spectrometry and the literature of the field of chemical cross-linking combined with mass spectrometry. I will then focus on zero-length chemical cross-linking combined with mass spectrometry, including the studies that led up to the creation of the methods I used and optimized as a member of the Speicher lab. Most of the content in this chapter has been published as a manuscript in Methods (Rivera-Santiago, Sriswasdi et al. 2015). In this study, Dr. Sira Sriswasdi contributed the sections that describe the ZXMiner software package, as well as the figures that describe the effect of high-resolution MS-MS spectra and XlinkInspector. He and Ms. Sandra Harper published the papers that were referenced in the development of this method. I then wrote this manuscript with input from the previous authors and Dr. David Speicher.

1.1. Structural MS

The field of structural mass spectrometry (MS) uses high-resolution MS data to interrogate structural parameters of complex biological molecules. It is typically used as a complement to more established structural methods, such as X-ray crystallography and NMR. Recent studies have demonstrated its usefulness for proteins and protein complexes that were difficult to examine by crystallography or NMR (Lasker, Forster et al. 2012; Greber, Boehringer et al. 2014; Olson, Tucker et al. 2014; Sriswasdi, Harper et al. 2014; Belsom, Schneider et al. 2015; Politis, Schmidt et al. 2015; Rivera-Santiago, Harper et al. 2015). I will first discuss techniques other than chemical cross-linking combined with mass spectrometry (CX-MS) then focus on CX-MS and advances that the

Speicher laboratory and other groups have made in this field, with an emphasis on "zero-length" CX-MS.

Structural MS techniques can be broadly classified as either top-down or bottom-up methods. Top-down methods derive structural insights from MS analyses of intact (or large fragments of) proteins and protein complexes. Bottom-up methods typically involve MS analyses of peptide mixtures after proteolytic digestion to identify specific amino acid residues that have been chemically modified or cross-linked. These data are then translated into structural information.

For example, ion-mobility MS (IM-MS) (Ruotolo, Benesch et al. 2008; Konijnenberg, Butterer et al. 2013; Lanucara, Holman et al. 2014) is a top-down technique that uses the relationships between the shapes and sizes of protein complexes and their flight time induced by an electric field in the presence of a buffer gas to provide collisional cross-sections of protein complexes. These cross-sections can then be used to distinguish between independently generated 3D structure models. Some advantages of IM-MS include its abilities to quantify the relative abundance of simultaneously existing conformations and to analyze proteins under native conditions (Konijnenberg, Butterer et al. 2013). One of its major limitations is that it requires stable transfer of an intact protein or protein complex into the gas-phase for analysis by MS, which is not always feasible or efficient.

Bottom-up methods are generally more flexible than top-down methods, but they suffer from high sample complexity, challenges in MS/MS data analysis, and difficulty in detecting low-abundance modified peptides. Hydroxyl radical footprinting (Kiselar and Chance 2010; Wang and Chance 2011; Monroe and Heien 2013; Vahidi, Stocks et al.

2013; Maleknia and Downard 2014; Yan, Chen et al. 2014), also known as oxidative footprinting, is an example of a bottom-up technique that modifies solvent-accessible amino acid side chains with hydroxyl radicals from hydrogen peroxide or irradiated water molecules. Changes in the pattern of modified residues reflect changes in protein surface topology. These changes can then be used to probe protein-protein and protein-ligand binding interfaces, or to infer the folding process (Poor, Jones et al. 2014). However, the fact that almost every amino acid residue can be modified in this manner (Xu, Takamoto et al. 2003; Xu and Chance 2004; Xu and Chance 2005; Maleknia and Downard 2014) greatly complicates downstream peptide identification.

Hydrogen-deuterium exchange (HDX) (Wales and Engen 2006; Konermann, Pan et al. 2011; Wei, Mo et al. 2014) utilizes a similar premise as oxidative footprinting, but one key difference is that it measures solvent accessibility of backbone hydrogen atoms via exchanges of deuterium atoms from deuterium-containing water (D2O) instead of chemical reactivity. HDX is a versatile technique that can be used as either a bottom-up approach or a top-down approach in order to visualize complex-wide exchange patterns (Wang, Abzalimov et al. 2013). Back-exchange, which is defined as the deuterated site reverting back to hydrogen, is a major concern for this technique, and special sample preparation steps coupled with rapid MS analysis need to be taken in order to ameliorate its effects (Wei, Mo et al. 2014).

1.2. CX-MS

CX-MS (Back, de Jong et al. 2003; Sinz 2006; Leitner, Walzthoeni et al. 2010; Rappsilber 2011; Paramelle, Miralles et al. 2013) is a bottom-up structural MS technique that provides information in the form of 3D spatial constraints between reactive amino acid side chains, which are tied together using a cross-linking reagent. This basic

concept has been used to probe proteins since the 1970s (Fasold, Klappenberger et al. 1971), but the difficulty of identifying these peptides and specific cross-linked residues in a protein sequence has severely limited its use. Interest in chemical cross-linking has increased steadily over the past decade primarily due to the great improvements in MS resolution and mass accuracy, which enable the precise detection of cross-linked peptides. The development of cross-linking reagents designed specifically for MS analysis, and development of specific data analysis tools and strategies have also contributed. Even large multi-subunit protein complexes and expansive protein-protein interaction networks can be probed with this structural MS technique (Chen, Jawhari et al. 2010; Herzog, Kahraman et al. 2012; Lasker, Forster et al. 2012; Leitner, Joachimiak et al. 2012; Walzthoeni, Leitner et al. 2013). Moreover, applications of CX-MS are not limited to the use of the canonical tryptic peptides typical of bottom-up structural MS techniques, as it has also been demonstrated to be useful in the analysis of intact proteins (Sinz 2006).

Most modern cross-linking reagent designs follow a basic template and contain features that belong to four broad aspects: cross-linking reactivity (required), an enrichment tag (optional), an isotopic label (optional), and an MS-cleavable site (optional) (Paramelle, Miralles et al. 2013). These optional features are typically found in the reagent's spacer arm, which is a set of atoms that is chemically inserted between the two cross-linked amino acid side chains. The cross-linker's reactivity is defined by the amino acid side chains that are amendable to cross-linking for a given chemistry. Some of the commonly utilized reactivities are amine-to-amine, carboxyl-to-amine, cysteine-to-cysteine, and cysteine-to-non-specific (photo-activated). Some of these cross-linkers also involve enrichment tags such as biotin (Petrotchenko, Serpa et al. 2011). Cross-

linking reagents that employ isotopic labels are commercially available as 1:1 mixtures of a single cross-linking reagent that have “heavy” and “light” versions of the spacer arm. By performing a cross-linking reaction with these reagents, all cross-linked peptides should have similar yields of peptide (precursor) ions that differ in mass by an amount equal to the mass difference of the heavy and light linker regions (and an m/z value of that mass difference, divided by the charge). In an LC-MS/MS experiment, signals corresponding to cross-linked peptides can be differentiated from the uncross-linked background by this distinctive heavy-light pattern of co-eluting peptides.

MS-cleavable sites are labile bonds that dissociate upon exposure to the low-energy fragmentation within the mass spectrometer. These sites are either attached to a reporter ion as a marker of cross-link-specific MS/MS spectra (Back, Hartog et al. 2001), or located within the spacer arm of the cross-linker to allow the two peptides forming the cross-link to be separated for individual MS^3 analysis (Tang, Munske et al. 2005; Kao, Chiu et al. 2011). However, these isotopic tags usually requires a spacer arm of >7Å in length. The incorporation of enrichment tags and MS-cleavable tags make the cross-linker spacer arm even bulkier. This can result in residues whose spatial locations in the structure of the protein or protein complex are not near each other to be cross-linked, and therefore greatly reduces the stringency of the derived distance constraints.

1.3. Structural analysis gaps addressed by CX-MS

CX-MS is at its most valuable when providing a basis upon which to base a medium-resolution structural model that could not have been completed using traditional structural means. Because traditional structural biology techniques such as X-ray crystallography and NMR are limited by the availability of pure, highly concentrated protein samples, any technique that is capable of providing structural insights in the

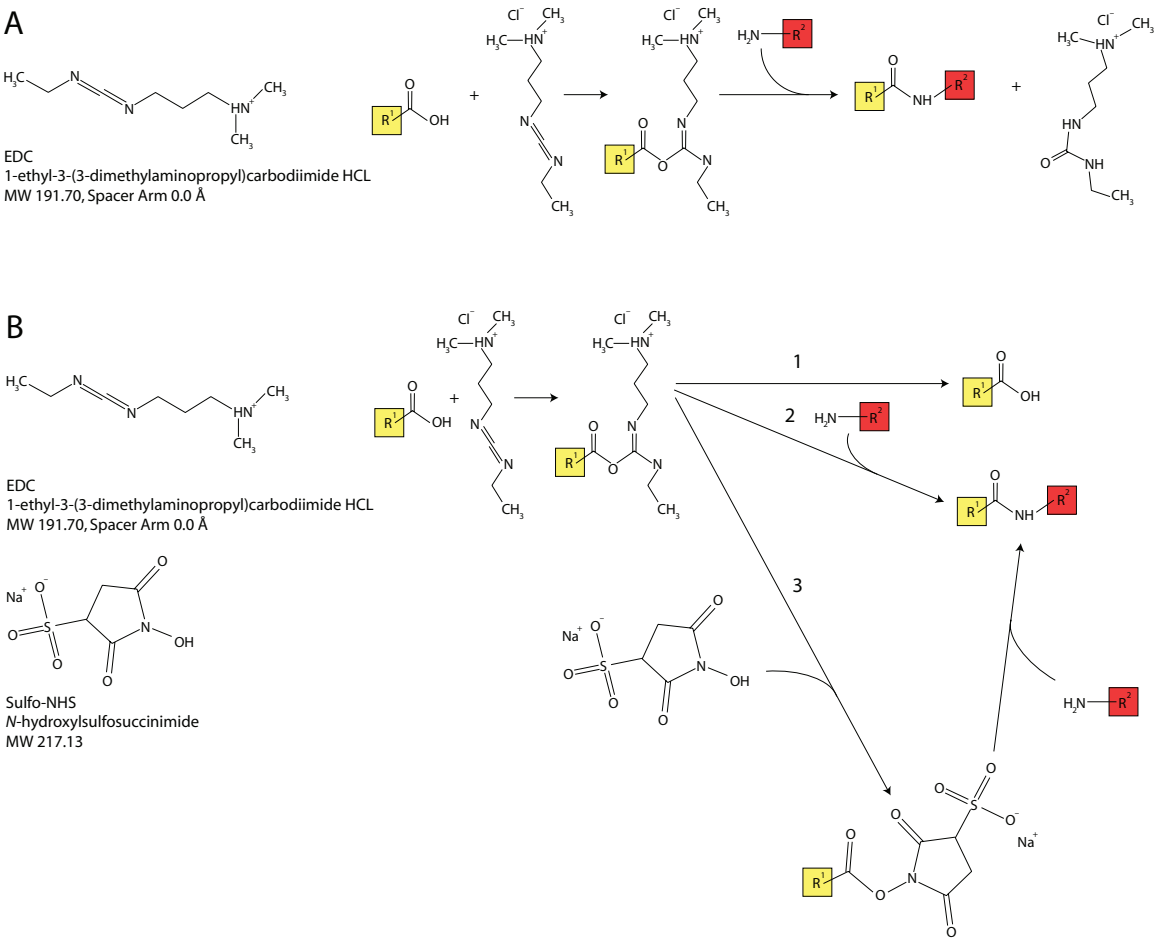
event that obtaining the aforementioned samples is not possible will fill some gaps in the traditional structural biology paradigm.

Moreover, X-ray crystallography and NMR tend to be limited in terms of size as to what proteins and protein complexes they can survey. This is most notably the case for NMR, as the spectra get exponentially more complicated with increased protein size. CX-MS does not have this size limitation, and properly designed experiments can gather valuable insights about the conformation of proteins in close to an *in vivo* setting, as will be discussed in Chapter 4.

1.4. Zero-length cross-linking and its role in improving structural models

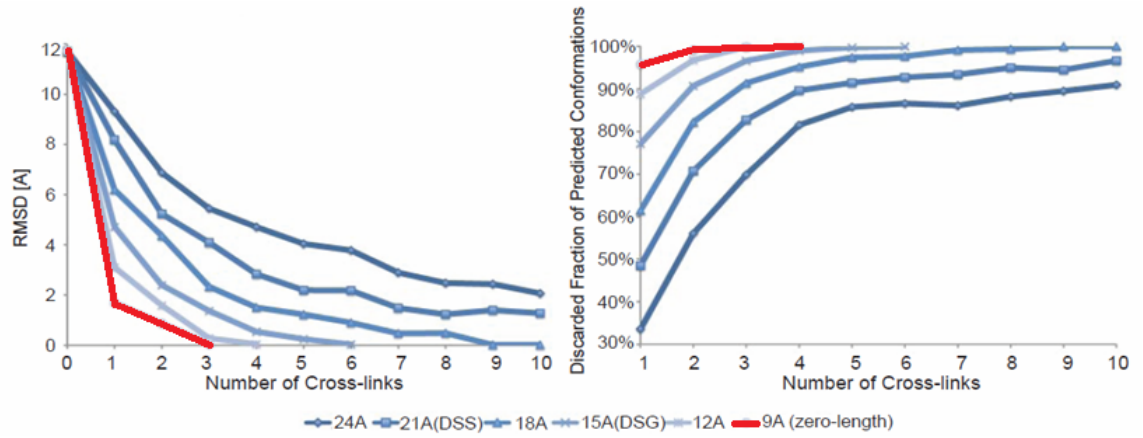
Among the groups of cross-linking reagents discussed in Section 1.2, one group stands out as different. Zero-length chemical cross-linkers are cross-linking reagents that do not feature a spacer arm between their reactive groups. The most common example of a zero-length cross-linker is 1-ethyl-3-(-3-dimethylaminopropyl) carbodiimide hydrochloride (EDC). This cross-linker creates a highly reactive but unstable acylisourea with a short half-life upon interaction with a carboxyl amino acid residue such as Asp, Glu, or the C-terminal carboxyl group of the protein. This intermediate ester can then interact with a primary amine (Lys or protein N-terminal amine) to form a peptidyl bond with the elimination of a water molecule. Ordinarily, this intermediate species has a lifetime of several seconds. However, addition of a second reagent, N-hydroxysuccinimide (NHS) or its water-soluble sulfo-derivative (sulfo-NHS), greatly extends the lifetime of this reactive intermediate, which in turn magnifies the extent of cross-linking (Staros, Wright et al. 1986; Grabarek and Gergely 1990). For further details on the chemistry of this reaction, please see Fig. 1-1.

Fig. 1-1: Zero-length cross-linking reaction chemistry. Reproduced with permission from (Rivera-Santiago, Sriswasdi et al. 2015). (A) Chemical formula for EDC (left) and forward cross-linking reaction involving an acylisourea intermediate (right) (B) Chemical formulas for EDC and sulfo-NHS (left) and reaction chemistry for zero-length cross-linking reaction, including: (1) the reverse reaction, (2) cross-link formation, and (3) stabilization of the acylisourea intermediate via substitution for an amine-reactive sulfo-NHS ester, followed by cross-link formation.



As was demonstrated via a systematic *in silico* analysis (see Fig. 1-2) (Leitner, Walzthoeni et al. 2010), zero-length cross-links provide the most specific and powerful distance constraints for the refinement and verification of homology models. This is because zero-length cross-linkers do not incorporate atoms from a spacer arm between the reactive sites, which implies that only residues whose side chains are within salt bridge distance can interact, and these shorter distances give a researcher or a structural prediction program greater resolving power between potential models.

Fig. 1-2: Results of *in silico* analysis of cross-link distance restraints and their effect on molecular models. Adapted with permission from (Leitner, Walzthoeni et al. 2010). Red line denotes zero-length cross-linker distance restraints.



Moreover, because the reaction can only occur between residues that are roughly within salt bridge distance of each other at the time of reaction, any residue interactions captured in this way are more likely to describe direct contact sites, rather than simply sites that are in close proximity (Zybailov, Glazko et al. 2013). Zero-length cross-linking reagents are also less likely to generate “dead-end” products. These products occur when only one of the reactive groups is able to react with a site on a

protein and the second site reacts either with water or a quenching reagent. EDC is less likely to produce these than other cross-linkers, partly because the intermediate ester is relatively unstable and reverts back to the unmodified carboxyl if it does not react with an amine. As a result, EDC dead-end adducts have only been observed at much high concentrations of cross-linker than those typically used for zero-length CX-MS (Lopez-Alonso, Diez-Garcia et al. 2009). In contrast, dead-end products are common and can often have higher stoichiometries than the actual true cross-links for other cross-linking chemistries. These dead-end products complicate the analysis of non-zero-length cross-link experiment data by enlarging the search space a database must consider. Zero-length cross-link experiments are also less likely to generate "self-linked" products that occur when interacting reactive groups are on the same peptide (Zybailov, Glazko et al. 2013), as both reactive groups would have to be on the same digestion enzyme peptide and also directly interact in a salt bridge for that to happen. Dead-end and self-linked products of non-zero-length cross-linkers are problematic because they must be accounted for in database searches.

The major challenge with zero-length cross-linkers for CX-MS experiments is how difficult it is to identify these sub-stoichiometric cross-linked peptides in the presence of the far more numerous and more abundant linear peptides. This is because adding MS-cleavable bonds to generate MS^3 fragmentation spectra (Kao, Chiu et al. 2011) or direct incorporation of differential isotopic labels as part of the cross-linking reaction are not feasible, as discussed above. However, there is an alternative isotopic labeling method can be used with zero-length cross-linkers. It involves tagging the N- or C-termini of peptides following the cross-linking reaction, which helps distinguish cross-linked peptides as they contain two N- or C-termini instead of one each (El-Shafey, Tolic

et al. 2006). Unfortunately, with the exception of strategies that introduce an isotopic tag during proteolysis, such as ^{18}O (Back, Notenboom et al. 2002), interpretation of N- or C-terminal tags can be complicated by the presence of internal reactive groups in linear peptides.

Another challenge with zero-length CX-MS, which also applies to its non-zero-length counterparts, is that the number of possible cross-links increases exponentially with the amount of unique sequences in the target protein or protein complex, potentially making the search space unmanageably large (Leitner, Walzthoeni et al. 2010). For the purposes of database search space, the limiting factor is the amount of unique primary sequence, rather than the molecular weight of the total protein complex. This sequence complexity problem is particularly challenging for zero-length cross-linking experiments. This is due in large part to the absence of an MS-based method to differentiate between modified and unmodified peptides.

As with most other types of cross-linkers, the specificity of zero-length cross-linkers (such as EDC) is a double-edged sword. Because its cross-linking reaction only involves lysine, aspartic acid, glutamic acid residues, and unmodified protein termini, EDC cannot be used to interrogate regions that lack any of these residues. Furthermore, the reactive groups themselves need to be accessible in aqueous solution. Hence, any reactive groups that are buried within a protein or macromolecular complex, or buried within the lipid bilayer will be refractory to cross-linking. Additionally, very large and very hydrophobic cross-linked peptides are generally very difficult to detect in a mass spectrometer tuned for peptide analysis. Therefore, regions of protein that contain reactive sites but no conveniently located proteolytic sites will be difficult to interrogate by CX-MS, although use of an alternative protease may minimize this problem.

1.5. Other groups' efforts

Due in large part to the disadvantages mentioned above, zero-length CX-MS has typically been a low-usage technique in the CX-MS field. In fact, I surveyed a series of CX-MS reviews published since 2012, and I saw that zero-length CX-MS was either not mentioned (Hyung and Ruotolo 2012; Serpa, Parker et al. 2012; Stengel, Aebersold et al. 2012; Merkley, Cort et al. 2013; Walzthoeni, Leitner et al. 2013; Budayeva and Cristea 2014) or only superficially discussed (Pacholarz, Garlish et al. 2012; Paramelle, Miralles et al. 2013). Despite all of these caveats, several studies by other groups have productively utilized zero-length CX-MS in various capacities.

One of these studies took a qualitative approach to zero-length CX-MS by utilizing it as a low-impact stabilizing reagent for protein macrocomplexes, which were then analyzed by matrix-assisted laser desorption ionization (MALDI) time of flight (TOF)-MS (Lepvrier, Doigneaux et al. 2014). This study did not identify specific cross-linked peptides. It instead used cross-linking to stabilize quarternary protein complexes of up to 388 kDa in size, which contained around 126.6 kDa in unique protein sequence.

Another study combined MALDI-TOF MS with zero-length cross-linking in an effort to gain insights on the interactions composing the photosystem II (PSII) in green algae (Nagao, Suzuki et al. 2010). This study found that the proteins known as PsbO, PsbP, and PsbQ form close interactions with each other, as well as with several other proteins in PSII. Additionally, an interaction between PsbP's N-terminus and PsbQ's C-terminus was identified via manual analysis of uncross-linked and cross-linked MS/MS spectra.

Several studies focused on mapping protein-protein interactions and used zero-length cross-links to determine the precise interaction sites between two purified proteins known to associate with each other (Schmidt, Kalkhof et al. 2005; Marekov 2007; Bumpus and Hollenberg 2010). One study focused on the interaction between the cytochrome P450 2B6 enzyme and NADPH-cytochrome P450 reductase (Bumpus and Hollenberg 2010). This study used ^{18}O labeling and de novo MS/MS sequencing to identify and analyze cross-links between a synthetic peptide mimicking the C-helix of cytochrome P450 2B6 and the connecting domain of the 76.7 kDa P450 reductase.

Another study used Fourier transform ion cyclotron resonance (FT-ICR) MS analysis to probe the interface between calmodulin and adenylyl cyclase 8 (Schmidt, Kalkhof et al. 2005). A similar technique with a variety of cross-linkers, including EDC, was later used to map the interaction between calmodulin and the skeletal muscle myosin light chain kinase M13 (Kalkhof, Ihling et al. 2005). The resulting cross-links were manually identified with the aid of the General Protein Mass Analysis for Windows (GPMAW) (Peri, Steen et al. 2001) and the Automatic Spectrum Assignment Program (ASAP) (Young, Tang et al. 2000) software packages, along with manual verification. These cross-links revealed an interaction between a 26-residue peptide corresponding to the N-terminus of M13 and the EF-hand 2 domain in calmodulin, which is in agreement with an NMR structure of the complex.

Zero-length CX-MS was also utilized in combination with data from other structural techniques to assist homology modeling of the interaction interface between proteins. One such study explored the interface between cytochrome P450 2E1 and cytochrome b5 using FT-ICR MS, and was able to generate a model of the interaction surface of the roughly 70kDa 1:1 complex (Gao, Doneanu et al. 2006). In another study, the binding

interface of the transiently PSII-associated protein known as Psb27 was probed via zero-length CX-MS (Liu, Huang et al. 2011), and the resulting cross-links were identified using MassMatrix (Xu and Freitas 2009). This experiment showed that Psb27 associates with the chlorophyll binding protein CP43 at two distinct sites, and a model of for the 61 kDa complex (with 61 kDa of unique sequence) was constructed. An interesting study combined HDX and zero-length CX-MS to re-examine the dimerization interface of the 14-3-3ζ protein (Haladova, Mrazek et al. 2012). These analyses uncovered novel contacts in the 56kDa dimer interface and resolved ambiguous salt bridge interactions that were previously noted.

Several groups have attempted to improve zero-length CX-MS data analysis and to reduce the reliance on manual verification. One study proposed a program, named Popitam. Popitam was based on the premise that a fragmentation pattern of a zero-length cross-linked peptide can be approximated as a mixture of two peptides, each having a modification of unknown size (corresponding to the other peptide) attached. This allows the MS/MS spectra of cross-linked peptides to be semi-automatically annotated via conventional peptide identification schemes (Singh, Shaffer et al. 2008). The algorithm was tested on a model system of the cytochrome P450 2E1-cytochrome b5 complex. A later study by the same group employed this protocol as part of a de novo protein modeling experiment in conjunction with previously published cryo-EM data in order to determine the structure of the roughly 20 kDa gpE viral capsid protein for bacteriophage lambda (Singh, Nakatani et al. 2013). Another zero-length CX-MS analysis strategy modified the database search algorithm of a popular MS data analysis software package, known as SEQUEST (Eng, McCormack et al. 1994). This modified algorithm considered all possible products from a cross-linking reaction and the

subsequent tryptic digest (cross-linked peptides, adducts, and linear peptides), generated theoretical spectra, then matched them to the observed spectra (McIlwain, Draghicescu et al. 2010).

As the above studies illustrate, applications of zero-length CX-MS were mostly limited to fairly small proteins or protein complexes because of the limitations in detection capacity for zero-length cross-links. Many of these studies also either relied on prior knowledge of the interaction interace or reported protein-protein interactions without identifying specifically cross-linked residues. Those that did identify specific cross-linked residues typically involved extensive manual analysis of MS/MS spectra, which make the data analysis tedious and unfeasible.

1.6. The Speicher laboratory's initial enhancements to zero-length CX-MS

The Speicher laboratory also performed several early zero-length CX-MS studies of red cell spectrin. These were done using recombinant protein constructs that preserved functionally important interfaces instead of the full-length protein. Initially, a combination of the SEQUEST (Eng, McCormack et al. 1994), GPMAW (Peri, Steen et al. 2001), and ASAP (Young, Tang et al. 2000) software packages, along with extensive manual curation of MS/MS spectra, was used to develop a structure for the spectrin heterodimer initiation site. The recombinant proteins used here were approximately 100 kDa in size, with an equivalent amount of unique protein sequence (Li, Tang et al. 2008).

A subsequent study utilized Rosetta Elucidator (Rosetta Biosoftware, Seattle, WA), GPMAW (Peri, Steen et al. 2001), and the homology modeling program MODELLER (Sali and Blundell 1993) to determine a model of the spectrin tetramerization interface using a "mini-spectrin" recombinant construct. This construct

formed a 180 kDa “tetramer” complex with 90 kDa of unique sequence (Li, Harper et al. 2010).

While these studies resulted in useful structural insights, similar limitations to the above studies by other groups were encountered. As mentioned above, the data analysis methods used were tedious and time-consuming, and the number of high-confidence cross-links identified was modest. These limitations led to the development of a new data analysis pipeline specifically tailored for zero-length CX-MS analysis, which is described in below.

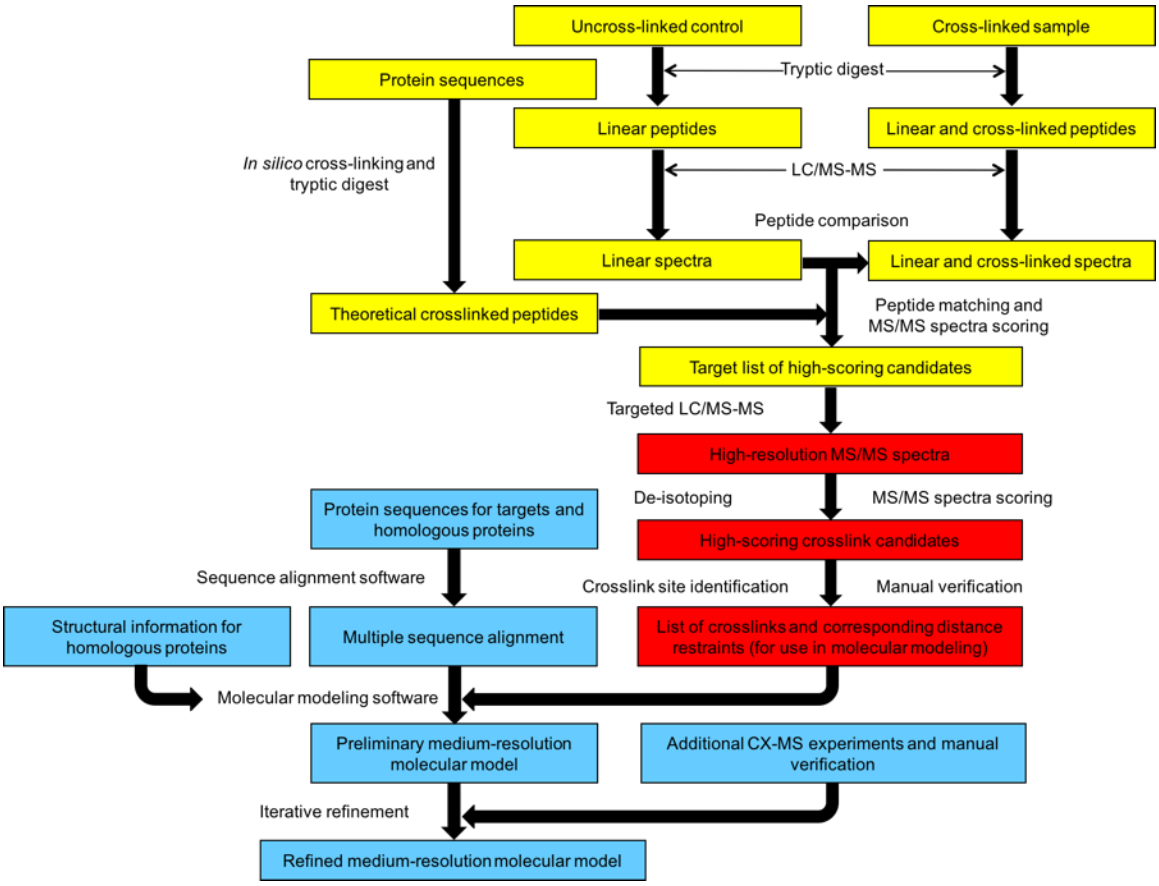
1.7. Development of ZXMiner and multi-stage MS analysis of zero-length CX-MS peptides

Because zero-length cross-linkers do not possess any of the special properties discussed above that can aid in cross-linked peptide identification, the Speicher laboratory developed an LC-MS/MS data analysis method coupled with an open-source software tool named Zero-Length Cross-link Miner (ZXMiner). The data acquisition method and the software were specifically optimized for identification of zero-length cross-links.

There are two key components to this strategy (see Fig. 1-3). First, parallel LC-MS/MS analyses of a cross-linked sample and an otherwise identical uncross-linked control are performed, and a label-free comparison of the resulting LC-MS datasets is performed to identify putative cross-linked precursors. This process was developed for use with LTQ-Orbitrap XL mass spectrometers (Thermo Scientific, Waltham, MA), which can conduct MS scans in the high-resolution orbitrap and MS/MS scans in the low-resolution ion trap in parallel to achieve a high duty cycle. However, the duty cycle was

too low when MS/MS scans were also conducted in the orbitrap in order to obtain high resolution, high mass accuracy MS/MS data in unbiased discovery runs. This approach resulted in very few identified cross-linked peptides.

Fig. 1-3: Diagram for the zero-length CX-MS cross-linking protocol, as optimized for a LTQ Orbitrap MS instrument. Reproduced with permission from (Rivera-Santiago, Sriswasdi et al. 2015). The protocol is separated into 3 major categories: sample preparation, label-free LC-MS comparison, and comparison to all possible theoretical cross-linked peptides (in yellow), database search of candidate spectra and cross-link identification (in red), and incorporation of distance restraints into homology modeling experiments (in light blue).



In order to counteract this problem, the Speicher laboratory employed a two-tiered MS analysis approach (Li, Harper et al. 2010; Harper, Sriswasdi et al. 2013; Sriswasdi, Harper et al. 2014; Sriswasdi, Harper et al. 2014). Specifically, as shown in Fig. 1-3, the initial comparison of the control and cross-linked is achieved by a “discovery run”. This is defined as a high-speed, low-resolution acquisition run of MS/MS data in the ion trap. This is followed by a “targeted run” with low-speed, high-resolution acquisition of MS/MS data in the orbitrap. This two-step procedure required additional LC-MS/MS runs for most experiments, but it allowed acquisition of high-resolution MS/MS data for all putative cross-linked peptides without compromising the method’s depth of analysis. Some key details of the method and representative results are described below.

In an ideal cross-linking experiment, each molecule of protein or protein complex should contain only one, or at most a few, cross-links in order to minimize the risk of over-cross-linking and altering the native structures of interest. Because of this, cross-linked peptides typically constitute much less than 1% of the peptides in a tryptic digest of a cross-linked sample. Because uncross-linked control samples should contain all detectable linear peptides, a quantitative label-free comparison of LC-MS patterns of a control and cross-linked sample can be used to remove these peptides from further consideration. In an earlier study [ref], a 10-fold intensity enrichment (cross-link to control) cutoff was used to compensate for the limited ability of the Elucidator software that was being used by the Speicher laboratory at that time to distinguish isotopic envelope patterns of low-intensity cross-linked precursors from overlapping noise.

To overcome this limitation, the Speicher laboratory developed a label-free comparison software module for ZXMiner whose purpose was determine the correct precursor masses and charge states of all ions and aligned two LC-MS patterns

(manuscript in preparation). This module operates under similar concepts to those utilized by the proteome analysis software known as MaxQuant (Sriswasdi, Harper et al. 2014). Namely, it used the correlation of intensity profiles across LC-MS scans to improve the detection and de-convolving of isotopic envelopes into monoisotopic species prior to performing label-free comparisons. This is a critical improvement when compared to Elucidator, which matches discrete MS signals regardless of whether they are part of a peptide isotopic envelope. The current label-free comparison module greatly diminishes random matches between cross-linked peptide signals and spurious signals in the control sample, as illustrated in Table 1-1. This increases the specificity of label-free comparison, and greatly decreases the number of candidate cross-linked peptides for subsequent targeted high-resolution MS/MS analyses.

Table 1-1: Distinguishing Interference in Control Samples Using Label-Free Comparison Software. Reproduced with permission from (Sriswasdi, Harper et al. 2014).

ID	Charge State	m/z	Cross-linked Peptide Sequences	Intensity in Control (Elucidator)	Intensity in Control (LFA)
1	3	819.0875	YEEHLY[E]R-{MSPILGYW[K]IK	1.9e+05	Absent
1	4	614.5674	YEEHLY[E]R-{MSPILGYW[K]IK	3.7e+05	Absent
1#	5	495.0544	YEEHLY[E]R-{M#SPILGYW[K]IK	4.7e+03	Absent
2	4	877.9311	FELGLEFPNLPYYIDG[D]VK-HNMLGGCP[K]ER	3.6e+05	Absent
2#	3	1175.5707	FELGLEFPNLPYYIDG[D]VK-HNM#LGGCP[K]ER	9.3e+04	Absent
3	4	970.4894	HNMLGGCP[K]ER-NKKFELGLEFPNLPYYIDG[D]VK	5.1e+04	Absent
3#	4	974.4881	HNM#LGGCP[K]ER-NKKFELGLEFPNLPYYIDG[D]VK	2.9e+04	Absent
4	3	1212.9374	KFELGLEFPNLPYYIDG[D]VK-HNMLGGCP[K]ER	6.2e+04	Absent
4	4	909.9549	KFELGLEFPNLPYYIDG[D]VK-HNMLGGCP[K]ER	4.3e+05	Absent
4#	3	1218.2691	KFELGLEFPNLPYYIDG[D]VK-HNM#LGGCP[K]ER	5.7e+04	Absent
4#	4	913.9536	KFELGLEFPNLPYYIDG[D]VK-HNM#LGGCP[K]ER	1.7e+04	Absent

5	3	854.4874	LLL[E]YLEEK-IEAIPQID[K]YLK	1.4e+04	Absent
5	4	641.1174	LLL[E]YLEEK-IEAIPQID[K]YLK	Absent	6.05e+06 [a]
6	4	680.1426	LLL[E]YLEEK-RIEAIPQID[K]YLK	7.0e+03	Absent
7	3	912.1363	LLLEYLE[E]K-YIAD[K]HNMLGGCPK	Absent	Absent
7#	3	917.4679	LLLEYLE[E]K-YIAD[K]HNM#LGGCPK	1.0e+04	Absent
7#	4	688.3527	LLLEYLE[E]K-YIAD[K]HNM#LGGCPK	4.3e+04	Absent
8	3	683.3577	[D]F[E]TLK-IAYS[K]DFETLK	1.1e+05	Absent
8	4	512.7701	[D]F[E]TLK-IAYS[K]DFETLK	1.2e+05	Absent
9	5	664.5565	YIAWPLQGWQATFGGG[D]HPPK-I[K]GLVQPTR	1.9e+04	Absent
10	4	464.7658	LLL[E]YLEEK-YL[K]SSK	Absent	Absent
11	3	1023.5399	LP[E]MLK-[K]FELGLEFPNLPYYIDGDVK	1.1e+05	Absent
11#	3	1028.8716	LP[E]M#LK-[K]FELGLEFPNLPYYIDGDVK	2.3e+05	Absent
12	3	1248.3078	YIAWPLQGWQATFGGGDHPP[K]}-V[D]FLSKLPEMLK	1.2e+05	Absent
13	3	1012.5001	MFE[D]R-[K]FELGLEFPNLPYYIDGDVK	1.7e+05	Absent

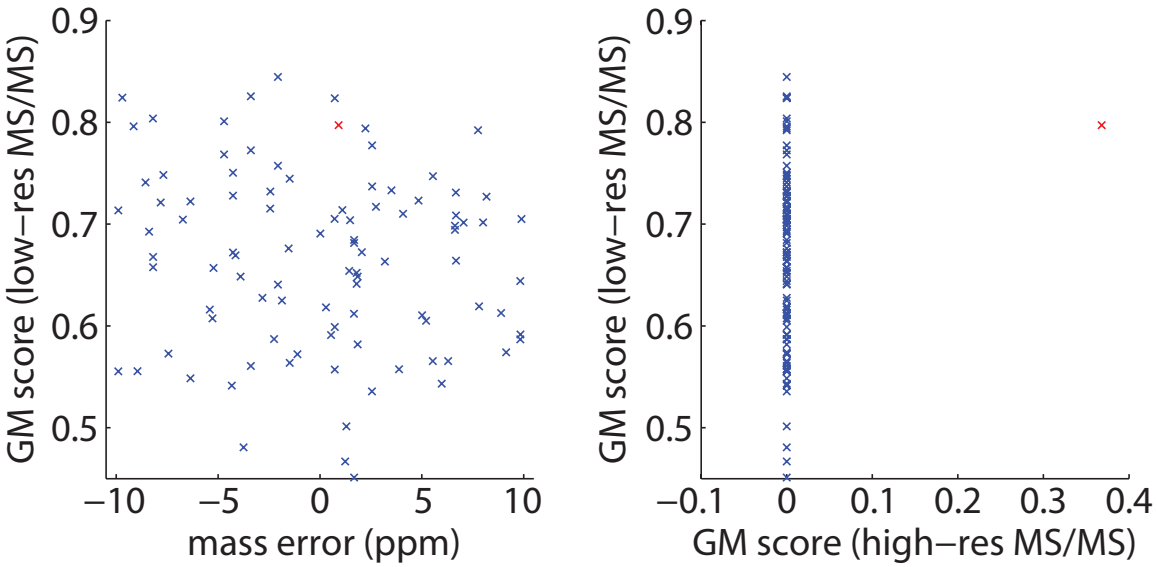
[a] Further inspection confirms that a precursor with similar m/z and charge state is present in the control sample. The intensity level of this precursor was 16 times lower than that of the cross-link sample and did not trigger an MS/MS scan. Elucidator could not detect this precursor.

The capability of modern mass spectrometers to produce high-resolution, high-mass-accuracy MS data has been instrumental in advancing the field of proteomics. Until recently, acquisition of high-resolution MS/MS spectra was less common (Frank, Savitski et al. 2007; Pan, Park et al. 2010). To identify linear peptides, low-resolution MS/MS spectra with mass accuracy of about ± 0.5 Da, such as those obtained in a linear ion trap, were reasonably adequate for common search engines such as SEQUEST (Eng, McCormack et al. 1994), MASCOT (Perkins, Pappin et al. 1999), etc.

However, because cross-linked peptides are much larger and typically feature higher charge states than their linear counterparts, their MS/MS spectra are much more complex. Multiple fragmented ions, often with different charge states, will frequently occur within a ± 0.5 Da mass tolerance window with substantial frequency. Furthermore, the combinatorial expansion in search space when considering cross-linked samples from n peptides to the order of n^2 cross-linked peptides dramatically increases the frequency where different theoretical cross-linked peptides having precursor masses within 5-10 ppm of one another by chance. Because of these factors, low-resolution

MS/MS spectra can be relatively ineffective at distinguishing between alternative cross-linked peptide assignments (see Fig. 1-4).

Fig. 1-4: High-resolution MS/MS spectra are required for high-confidence identification of cross-linked peptides. Reproduced with permission from (Rivera-Santiago, Sriswasdi et al. 2015). In a zero-length cross-linking study of a large 526 kDa spectrin heterodimer using an Orbitrap XL mass spectrometer, a typical MS/MS scan can be matched to as many as 97 distinct theoretical cross-linked peptides within a 10-ppm mass tolerance range. The red X highlights the correct assignment. The geometric mean (GM) score (Sriswasdi, Harper et al. 2014) represents the quality of a match between a cross-link sequence and an MS/MS spectrum. (A) Analysis of low-resolution MS/MS data cannot distinguish between the 97 alternative theoretical cross-linked peptides. (B) With high-resolution MS/MS data, only a single correct assignment stands out with a non-zero score.



The ZXMiner data acquisition strategy overcomes the aforementioned drop in duty cycle that occurs on hybrid ion trap instruments when MS/MS scans are acquired at high resolution by acquiring LC-MS/MS data in two stages. LC-MS data from the discovery analyses was subjected to the label-free comparisons described above and the low-resolution MS/MS associated with putative cross-linked spectra were evaluated using ZXMiner (Sriswasdi, Harper et al. 2014) with low stringency scoring parameters. These two analyses significantly narrow down the list of candidate cross-linked peptide precursors. As a result, only a few parallel targeted LC-MS/MS runs were needed to obtain high-resolution, high-accuracy MS/MS spectra for all candidate precursors.

In these targeted analyses, both MS and MS/MS scans were conducted in the Orbitrap and for complex samples, lists of candidates were split among multiple LC-MS/MS runs to ensure several attempted MS/MS analyses of all targeted peptides across each chromatographic peak. The number of runs required can be estimated by plotting the expected number of target precursors during each chromatographic time interval (the precursor retention times are extracted from the discovery runs). The minimum signal threshold for triggering an MS/MS scan was set at 30,000 ion counts (as compare to the typical setting of 1,000 ion counts to trigger MS/MS scans in the ion trap) to compensate for the reduced sensitivity of the orbitrap mass analyzer. Additionally, the monoisotopic precursor selection option was turned on, and precursors with charge states of +1 and +2 were screened out, as they would rarely correspond to cross-linked tryptic peptides.

The most critical component in the development of an optimized zero-length cross-linking analysis pipeline was the creation of ZXMiner. This software tool was developed to address three major goals: increase the throughput of zero-length cross-

linking experiments by automating data analysis, increase the number of assigned cross-links, and improve the confidence of cross-link assignments. The Speicher laboratory tested ZXMiner against other cross-linked peptide analysis software. Only one previously developed cross-link analysis software package, pLink (Yang, Wu et al. 2012), could utilize high-resolution MS/MS data effectively. Additionally, none of the software packages that had been previously utilized in zero-length CX-MS studies were capable of interpreting the aforementioned data in a high-mass accuracy, high-throughput fashion, which is what prompted development of ZXMiner. A direct comparison shows that ZXMiner significantly outperforms pLink (Zhu, Smith et al. 2010), StavroX (Gotze, Pettelkau et al. 2012), Crux (Park, Klammer et al. 2008), and MassMatrix (Xu and Freitas 2009) when used for zero-length cross-link detection, as shown in Table 1-2. The data for the true positives as compared to the false positives (assigned by comparison to the crystal structure) suggests that the maximum distance cutoff that can be expected for an EDC cross-link is roughly 16Å. Another interesting observation is that the the actual false discovery rate (FDR) is typically much higher than an estimated FDR for all methods when tested on standard proteins. It is also noteworthy that at the unique sequence level, ZXMiner could identify the largest number of true cross-links with no false positives.

Table 1-2: Comparison of GST crosslink identifications using alternative software. Reproduced with permission from (Sriswasdi, Harper et al. 2014). For each peptide sequence, charge states of the cross-linked peptide identified by each software package are indicated. Estimated false discovery rates (FDRs) were derived from the decoy data that each software package provided (not available from MassMatrix). Number of

identifications, false positives, and the actual FDRs were calculated directly from this table.

Unique Sequence ID	Peptide Sequence	Cα-Cα Distance	ZXMiner		pLink	StavroX	Crux[a]	Mass Matrix[b]
True Positives								
1	YEEHLYER-{MSPILGYWKIK	5.5	3,4	3,4,5	3	3,4,5	3,4	5
1#	YEEHLYER-{M#SPILGYWKIK	5.5	5	4,5	-	3,4	N/A	5
2	FELGLEFPNLPYYIDGDVK-HNMLGGCPKER	8.7	4	4	-	-	3,4	4
2#	FELGLEFPNLPYYIDGDVK-HNM#LGGCPKER	8.7	3	3	-	-	N/A	3
3	HNMLGGCPKER-NKKFELGLEFPNLPYYIDGDVK	8.7	4	4,5	-	-	4,5	4
3#	HNM#LGGCPKER-NKKFELGLEFPNLPYYIDGDVK	8.7	4	4	-	-	N/A	5
4	KFELGLEFPNLPYYIDGDVK-HNMLGGCPKER	8.7	3,4	3,4	-	-	3,4	3
4#	KFELGLEFPNLPYYIDGDVK-HNM#LGGCPKER	8.7	3,4	3,4	-	-	N/A	5
5	LLLEYLEEK-IEAIPQIDKYLK	9.2	3,4	3,4	3	3,4	3,4	-
6	LLLEYLEEK-RIEAIPQIDKYLK	9.2	4	4	-	4	-	-
7	YEEHLYER-{M#SPILGYWK	10.0	-	-	-	3,4	N/A	N/A
8	LLLEYLEEK-YIADKHNMLGGCPK	11.4	3	3	3	-	3	3
8#	LLLEYLEEK-YIADKHNM#LGGCPK	11.4	3,4	3,4	-	-	N/A	-
9	DFETLK-IAYSKDFETLK	11.6	3,4	3,4	3	3	3,4	-
10	YIAWPLQGWQATFGGGDHPPK}-IKGLVQPTR	11.8	5	5	-	5	5	5
11	LPEMLK-KFELGLEFPNLPYYIDGDVK	12.4	3	3,4	-	3,4	3,4	3
11#	LPEM#LK-KFELGLEFPNLPYYIDGDVK	12.4	3	3	-	4	N/A	4
12	LLLEYLEEK-YLKSSK	12.4	4	4	-	3,4	4	-
13	AEISMLEGAVLDIRYGVSR-YIADKHNM#LGGCPK	13.8[c]	-	-	-	-	N/A	4
14	YIAWPLQGWQATFGGGDHPPK}-VDFLSKLPEMLK	14.6[d]	-	5	5	-	4,5	5
14#	YIAWPLQGWQATFGGGDHPPK}-VDFLSKLPEM#LK	14.6[d]	3	3	-	3,4	N/A	5
15	MFEDR-KFELGLEFPNLPYYIDGDVK	15.8[c,d]	3	3	3	-	-	-
15#	M#FEDR-KFELGLEFPNLPYYIDGDVK	15.8[c,d]	-	-	-	3	N/A	3
False Positives								
1	LLLEYLEEK-LVCFKK	16.3	-	-	-	-	3	-
2	IKGLVQPTR-DEGDK	16.4	-	-	3	4	3	-
3	IKGLVQPTR-DEGDKWR	16.4	-	-	-	3	-	-
4	DEGDKWR-LPEMLK	18.5	-	-	-	3,4	3	-
5	DEGDKWRNK-LPEMLKMFEDR	18.5	-	-	-	-	3	-
6	VDFLSKLPEMLKMFEDR-ER	19.0	-	-	-	-	-	3
6#	VDFLSKLPEMLKM#FEDR-ER	19.0	-	-	-	3	N/A	-
7	IKGLVQPTR-LLLEYLEEK	19.1	-	-	-	-	-	3
8	{MSPILGYWKIK-ERAEISMLEGAVLDIR	20.9	-	-	-	4	-	-
9	ERAEISMLEGAVLDIRYGVSR-WRNKK	21.5	-	-	-	-	3	-
10	IKGLVQPTR-DFETLK	22.7	-	3	3	-	3	3
11	HNMLGGCPKER-LVCFKK	23.9	-	-	-	-	3	-
12	LLLEYLEEK-DEGDKWR	25.2	-	3	-	3	-	-
13	YEEHLYERDEGDKWR-AEISM#LEGAVLDIR	27.2	-	-	-	-	N/A	4
14	VDFLSKLPEMLKMFEDR-YIADKHNMLGGCPK	28.5	-	-	-	-	5	-
15	IEAIPQIDKYLK-MFEDR	29.6	-	-	3	-	-	-
16	LLLEYLEEK-LPEMLKMFEDRLCHK	33.4	-	-	-	-	3	-
17	DFETLK-LVCFKK	34.2	-	-	-	3	-	-
	Unique Precursor Level	**Estimated FDR:**	< 1%	< 5%	< 5%	< 5%	< 5%	N/A
		True Positives:	25	30	6	20	19	15
		False Positives:	0	2	3	8	9	4

	Actual FDR:	0%	6.3%	33%	29%	32%	21%
Unique Sequence Level	**True Positives:**	13	15	6	10	11	10
	False Positives:	0	2	3	7	9	4
	Actual FDR:	0%	12%	33%	41%	45%	29%

[a] Crux's default setting cannot identify peptides with variable modifications.

[b] MassMatrix cannot identify cross-links involving the protein terminus as one of the cross-linked sites.

[c] Cross-links between subunits.

[d] Cross-links involving flexible regions (as reflected by the elevated B-factors and/or loops).

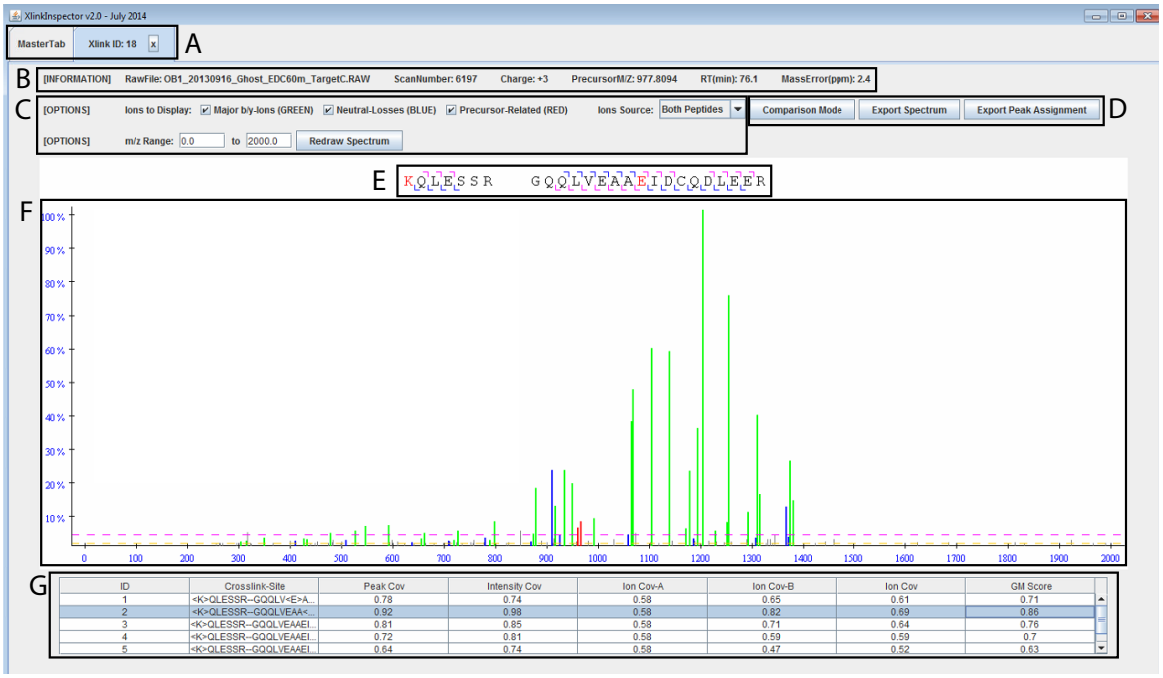
ZXMiner accepts input MS data in mzXML format, along with a list of candidate MS/MS scans generated by a label-free comparison tool, and databases of pertinent protein sequences in FASTA format. The software first performs an *in silico* digestion of the proteins with the specified enzyme (usually trypsin), which is then followed by production of a list of all possible theoretical cross-linked peptides. The masses of these theoretical cross-linked peptides are calculated as the sum of the theoretical peptide masses after accounting for the mass change caused by a cross-linking reaction, which is defined by the user. These masses are then compared to those of the candidate cross-linked peptide precursors, resulting in matches between theoretical cross-link sequences and candidate MS/MS spectra. As the LC-MS data was obtained at high mass accuracy, a stringent mass tolerance, typically 5 or 10 ppm, is applied here. For each match, ZXMiner processes the MS/MS spectrum, performs an *in silico* fragmentation of the putative cross-link sequence to generate a list of theoretical MS/MS ions, and calculates "coverage" scores that represent how well the two sets of data overlap. Individual scores, as well as their geometric mean (GM), which is an overall indicator of cross-link match quality, are reported. When analyzing high-resolution MS/MS spectra, ZXMiner de-convolutes all isotopic envelopes and collapses them into monoisotopic peaks. This use of de-convoluted high-resolution MS/MS data is a critical

factor that helps ZXMiner outperform other software packages and achieve a low false discovery rate (FDR) without sacrificing cross-link coverage, as previously shown (Sriswasdi, Harper et al. 2014) (see Table 1-2).

Another important feature of ZXMiner is that it creates intermediate files throughout a run's progression that record the output of its raw search results and subsequent interpretation. These files can contain the scores of all evaluated MS/MS spectra, cross-linked peptides, and even alternative cross-linked sites on each peptide. Hence, users can re-specify search parameters such as minimum peptide length, minimum scores, or precursor mass tolerance and use ZXMiner to rapidly recalculate the FDR for each identified cross-link without having to re-process LC-MS/MS and sequence data. XlinkInspector, which is part of the ZXMiner package, can be used to inspect the quality of each identified cross-link and confirm the exact cross-linked site (see Fig. 1-5). In recognition that frequently only a few weak MS/MS ions distinguish between alternative cross-linked sites within a peptide, the final decision regarding alternative cross-link site assignments is left to the user, instead of simply having ZXMiner output the highest-scoring candidates.

Fig. 1-5: XlinkInspector is a graphical interface within ZXMiner that aids verification of cross-linked peptide assignment and determination of cross-linked sites. Reproduced with permission from (Rivera-Santiago, Sriswasdi et al. 2015). (A) The tabbed interface allows quick navigation between individual cross-link assignments for visual inspection of data quality. (B) The header displays basic information about the selected MS/MS spectrum. (C) This more detailed interface provides flexibility for spectrum annotation and plotting. (D) The list of identified fragmented ions can be exported for further inspection. The annotated spectrum can also be exported in

publication-ready Scalable Vector Graphics format (SVG). (E) Visual representation of the identified b- and y-ions on the two peptide sequences. The assigned cross-linked residues are highlighted in red. (F) Color-coded plot of MS/MS data. Identified major b- and y-ions are shown in green. Neutral losses are shown in blue. Precursor-related ions are shown in red. Yellow and pink cutoff lines indicate minimum intensity values required to be designated as a peak or scored in the GM scoring algorithm, respectively. (G) Alternative cross-linked sites are listed along with their respective coverage scores. Annotation of the MS/MS spectrum dynamically changes in panel F when different cross-link sites are selected in panel G.



1.8. Previous studies done by the Speicher laboratory using ZXMiner

The Speicher laboratory has employed ZXMiner and zero-length CX-MS in several projects. The most common use of these zero-length CX-MS IDs are to uncover or confirm protein-protein interactions and to contribute to structural determination of

proteins and protein complexes. For this latter application, high confidence cross-link assignments from ZXMiner are then applied as distance constraints to molecular modeling programs, such as MODELLER (Sali and Blundell 1993). The distance constraint utilized for such experiments is typically tied to the location of the α-carbon atoms of the cross-linked residues. The reason why the Speicher laboratory chose this as the restriction method is because the α-carbon backbone of a protein usually provides more reliable information for the purposes of model-building. This distance restraint is typically set at 12Å, but the Speicher laboratory has reported observing distances of up to 16Å for high-confidence cross-links in areas likely to exhibit conformational flexibility when evaluating ZXMiner on solved crystal structures (Sriswasdi, Harper et al. 2014). This upper boundary is further illustrated in Fig. 1-6, which also highlights the importance of having high-mass accuracy MS/MS spectra. The most notable studies employing ZXMiner aimed to elucidate structural insights for human red cell spectrin tetramers, as well as large conformational differences between “closed” dimers and “open” dimers (Sriswasdi, Harper et al. 2014). A 90 kDa fused heterodimer construct with the intact tetramerization site was used as a simplified surrogate for the full-sized 526 kDa heterodimer for these zero-length CX-MS experiments (Harper, Li et al. 2010). CX-MS analysis of mini-spectrin tetramers using ZXMiner revealed important additional cross-links not previously identified using older data acquisition and data analysis methods (Li, Tang et al. 2008; Li, Harper et al. 2010), which enabled determination of a more refined structure for the spectrin tetramer region (see Fig. 1-7).

Fig. 1-6. Analysis of GST cross-links using ZXMiner. Reproduced with permission from (Sriswasdi, Harper et al. 2014). (A) Locations of identified cross-links on the crystal structure of the GST homodimer (PDB ID: 1GTA). Lys residues are highlighted in blue

and Glu and Asp are in red. The black lines connect the two α-carbons of each cross-link. Cross-links between residues whose Cα-Cα distances are significantly larger than 12 Å were highlighted in orange. (B) Scatter plot showing the relationship between GM scores derived from high-resolution MS/MS data and Cα-Cα distances for all cross-linked peptide candidates in the GST data set. A few cross-links located in regions likely to exhibit increased flexibility, such as loops or inter-subunit interfaces, slightly exceeded the expected 12 Å maximum Cα-Cα distance. (C) ROC curves showing the superior performance of high-resolution MS/MS data (area under the curve = 0.99) compared with low-resolution data (area under the curve = 0.80).

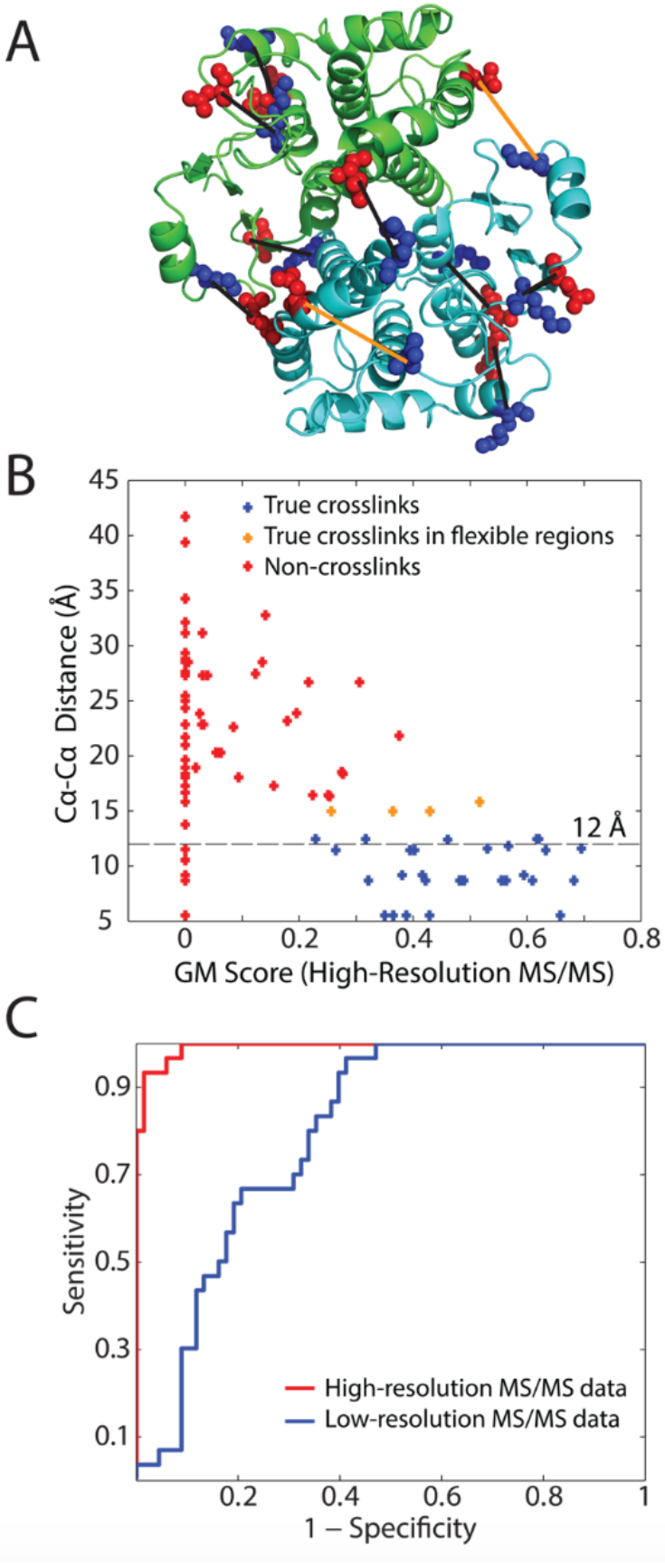
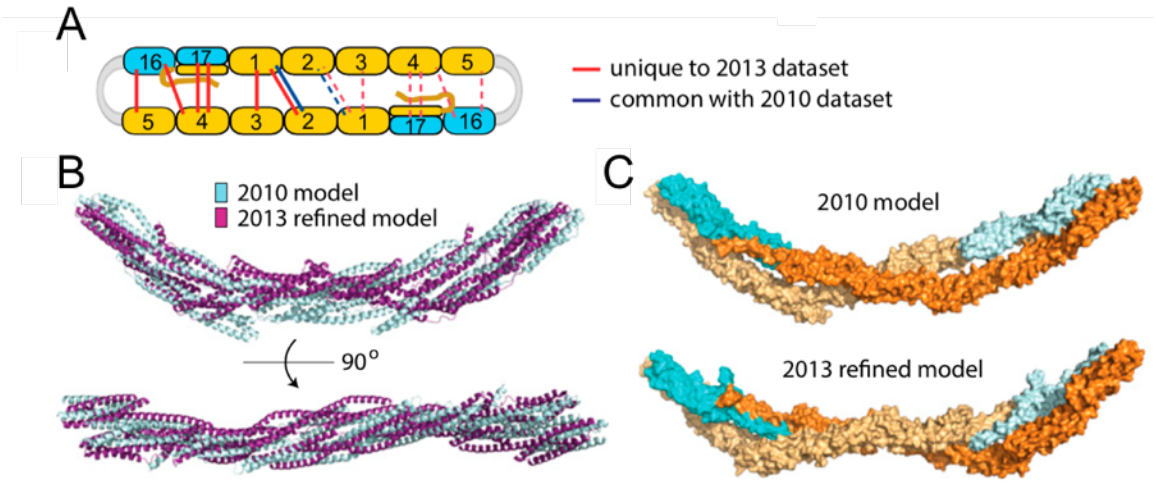
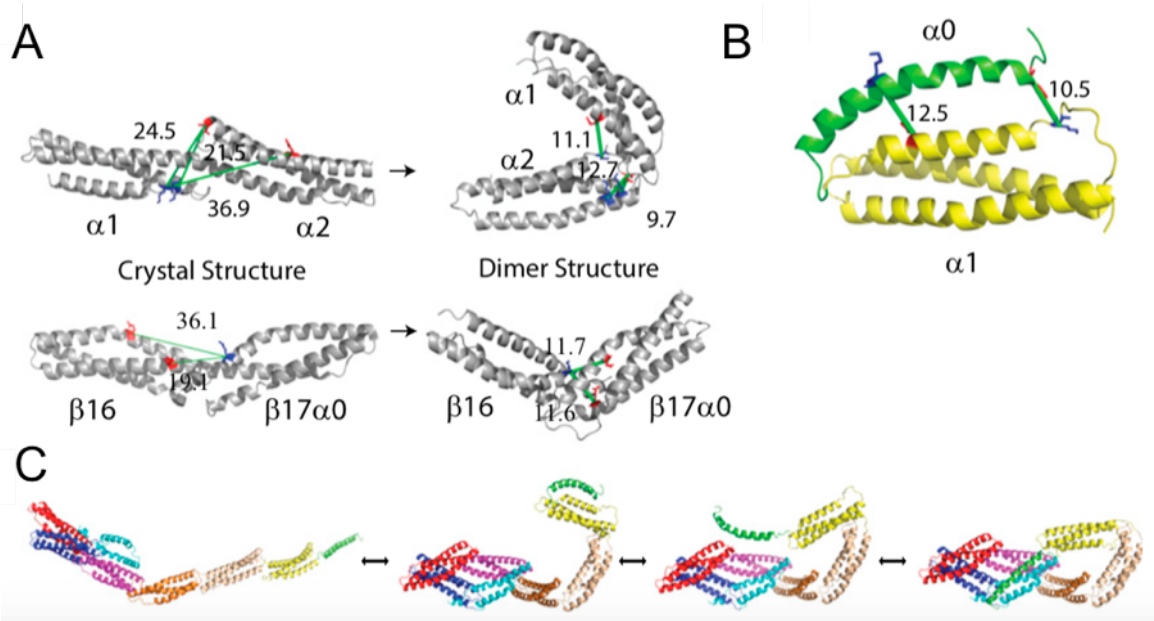

A
B
45
40
35
30
25
20
15
10
5
Cα-Cα Distance (Å)
True crosslinks
True crosslinks in flexible regions
Non-crosslinks
12 Å
0
0.2
0.4
0.6
0.8
GM Score (High-Resolution MS/MS)
C
0.9
0.7
0.5
0.3
0.1
Sensitivity
High-resolution MS/MS data
Low-resolution MS/MS data
0.2
0.4
0.6
0.8
1
1 – Specificity

Fig. 1-7. Structures for mini-spectrin tetramer based on zero-length CX-MS data analysis using ZXMiner. Reproduced with permission from (Sriswasdi, Harper et al. 2014). (A) Locations of inter-domain cross-links used to model mini-spectrin tetramer. Blue lines indicate cross-links identified previously (Li, Harper et al. 2010); red lines indicate new cross-links identified using the ZXMiner workflow; dashed lines indicate the same cross-links repeated in the second half of the tetramer. (B) Superimposition of present and previous tetramer structures. (C) Space-filling representations of tetramer models. β-spectrin domains are colored in bright or pale cyan, and α-spectrin domains are colored in bright or pale orange to distinguish the two strands.



Analysis of mini-spectrin dimers revealed several distance constraints that were mutually exclusive and were produced by the open and closed dimer forms of the protein that were in equilibrium and could not be independently isolated. Distance constraints derived from these cross-links allowed the Speicher laboratory to generate the first structures for the open and closed forms of mini-spectrin dimers (see Fig. 1-8) (Sriswasdi, Harper et al. 2014).

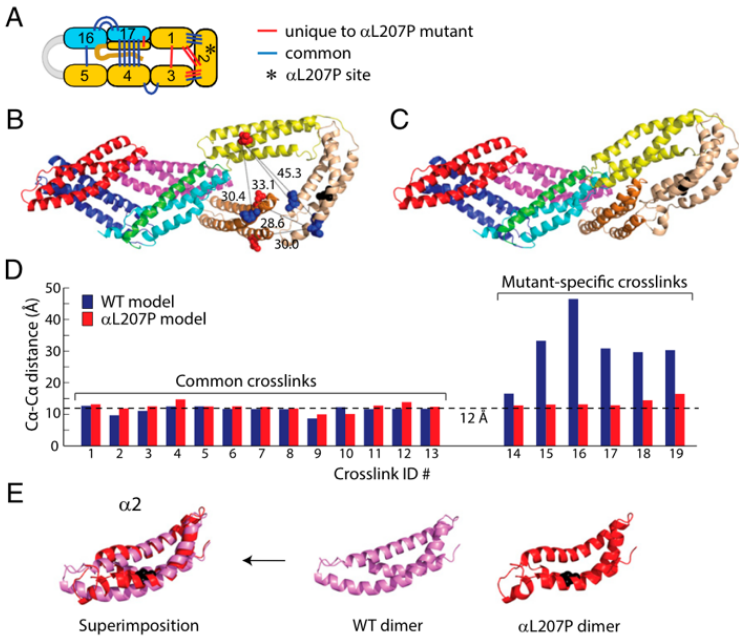
Fig. 1-8. Zero-Length CX-MS enables identification and modeling of large changes in conformation for mini-spectrin dimers. Reproduced with permission from (Sriswasdi, Harper et al. 2014). (A) Open dimer-specific cross-links indicative of nonhelical connectors before (left) and after (right) structural refinement. Lys residues in blue; Glu/Asp residues in red; green lines are cross-links with labeled Cα–Cα distances. (B) Open dimer model supported by two cross-links between α0 and α1 domains. (C) Structures showing the interconversion between fully extended open dimer to closed dimer.



Moreover, subsequent studies utilizing an L207P mutant of the mini-spectrin construct, which is a mutation previously shown to destabilize tetramer formation and which causes hereditary elliptocytosis in patients with this mutation (Gallagher, Tse et al. 1992), revealed cross-links outside the expected ranges of molecular flexibility of the wild-type mini-spectrin dimer. Distance constraints derived from these cross-links were used to elucidate a structure of the mutant dimer (see Fig. 1-9), which in turn illustrated

the allosteric effects of this mutation. Similar studies were also conducted on an L260P mutant that has similar destabilizing effects on spectrin tetramer formation (Glele-Kakai, Garbarz et al. 1996), revealing a similar structural perturbation to that shown in Fig. 1-9 (Harper, Sriswasdi et al. 2013).

Fig. 1-9. Cross-links and structure for the L207P mutant mini-spectrin dimer. Reproduced with permission from (Sriswasdi, Harper et al. 2014). (A) Locations of αL207P inter-domain cross-links; the asterisk indicates the location of the αL207P mutation. (B) Locations of five αL207P mutant-specific cross-links indicative of conformational rearrangements in the α1-α2-α3 region plotted on the WT structure for comparison; blue, Lys; red, Glu/Asp; black, Pro mutation; black lines, cross-links with Cα–Cα distances labeled. (C) Model of the αL207P mutant closed dimer. (D) Cα–Cα distances for inter-domain cross-links identified in the αL207P mutant dimer on the WT and αL207P closed dimer structures. (E) Superimposition of α2 domains from the WT and αL207P closed dimer. The mutated residue is shown in black.



1.9. Conclusion

In summary, CX-MS is an emerging structural technique that is taking advantage of the recent technological advances in MS/MS instrumentation and software. Zero-length cross-links offer the best quality of data it provides for generating and validating structural models, but these pose significant challenges in terms of detection and analysis. The Speicher laboratory is particularly interested in zero-length CX-MS, and has recently developed a powerful method to systematically analyze these samples in ZXMiner (Sriswasdi, Harper et al. 2014). This program has already been employed in several structural studies involving spectrin constructs, and it has proven to dramatically enhance the modeling that can be performed using zero-length CX-MS as a validation mechanism. This means that zero-length CX-MS combined with molecular modeling is an approach that has great potential for further advancement and improvement, as I will detail in the following chapters.

Chapter 2: Solution structure of biologically active human peroxiredoxin-6

In this chapter, I will describe a collaborative project I undertook with other members of the Speicher laboratory, as well as members of Dr. Aron Fisher's laboratory in the University of Pennsylvania. Most of the content in this chapter has been published as a manuscript in The Biochemical Journal (Rivera-Santiago, Harper et al. 2015). In this study, Dr. Suiping Zhou and Dr. Sheldon Feinstein prepared the recombinant protein, Ms. Sandra Harper cross-linked the protein, and I performed the LC-MS analysis and the homology modeling experiments. I then wrote the manuscript with inputs from all the authors, as well as from Dr. Aron Fisher and Dr. David Speicher.

2.1. Background

One major project I embarked upon once I had a grasp of ZXMiner and its requisite skills was to collaborate with Dr. Aron Fisher's group in determining the solution structure of peroxiredoxin-6. Our interest in the protein was driven by its role in regulation of oxidative stress in human lungs, because the presence of reactive oxygen species (ROS) and their by-products can have multiple deleterious effects on cells. The lung is particularly vulnerable to these effects, due to its continual exposure to oxidants via ambient air as well as its extensive capillary networks (Fisher, Forman et al. 1984). Oxidative damage has been notably associated with various disease states in the lung, ranging from acute lung injury (ALI) to chronic obstructive pulmonary disease (COPD) (Lang, McArdle et al. 2002; Rahman and Adcock 2006).

In order to protect against oxidative injury, eukaryotic cells express a number of anti-oxidative stress proteins. Among these is a family of proteins known as peroxiredoxins. Peroxiredoxins differ from others involved in similar functions by working in conjunction with thiol-containing electron donor molecules, as opposed to using redox

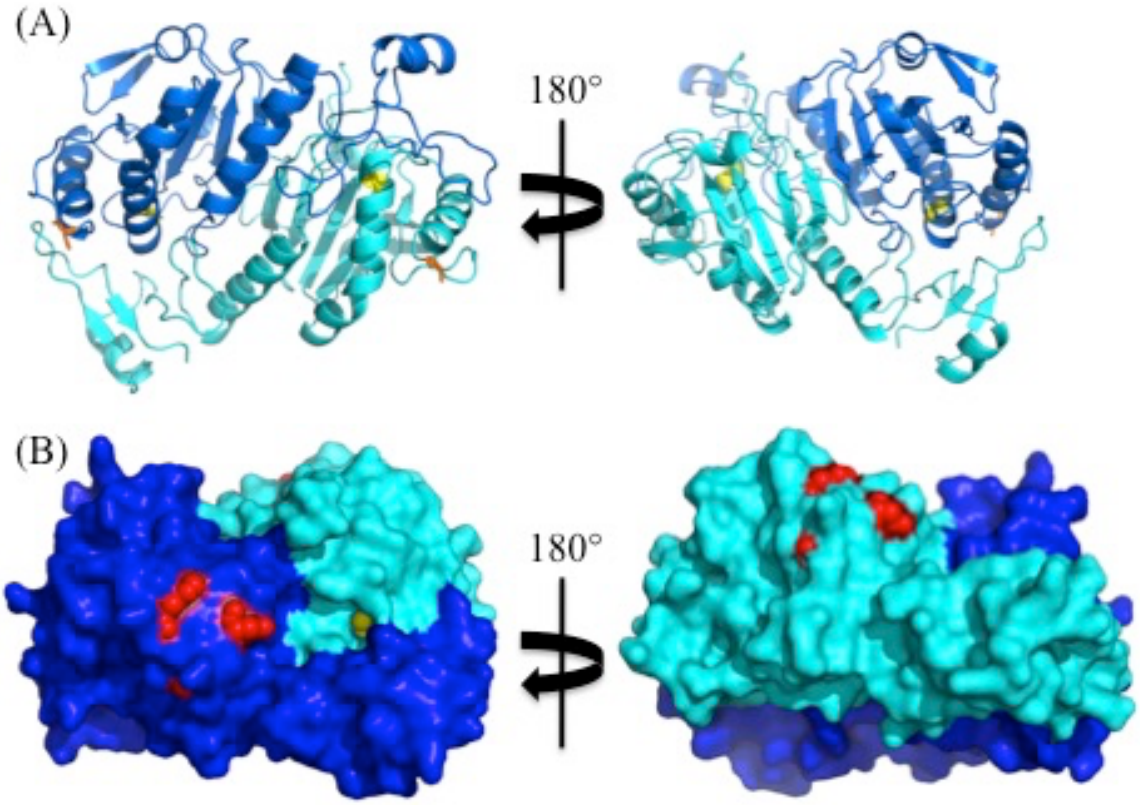
cofactors or prosthetic groups (Wood, Schroder et al. 2003; Rhee, Chae et al. 2005). Most peroxiredoxins employ two conserved Cys residues in a disulfide bond as their electron donor group in the presence of a reductant, typically thioredoxin. This is known as a 2-Cys mechanism (Wood, Schroder et al. 2003; Rhee, Chae et al. 2005).

Peroxiredoxin-6 (PRDX6) is a unique case in the peroxiredoxin family. It is a homo-dimeric enzyme found in mammals, particularly mammalian lungs (Manevich and Fisher 2005), and it features at least two distinct enzymatic activities – the reducing property common to all peroxiredoxins, as well as phospholipase A2 (PLA_2) via a conserved catalytic triad (His-26, Ser-32, Asp-140) (Kang, Baines et al. 1998; Manevich, Reddy et al. 2007). It is also the only peroxiredoxin with the reported ability to reduce phospholipid hydroperoxides (Fisher, Dodia et al. 1999). Moreover, PRDX6 does not feature the two canonical Cys residues that are characteristic of most members of the peroxiredoxin family. For PRDX6, one cysteine residue, Cys-47, is conserved across species. Thus, its peroxidase activity is described as a 1-Cys mechanism that involves oxidation of Cys-47 to a sulfenic acid as a peroxidase catalytic intermediate. The enzymatic cycle is completed by reduction of the active site back to a sulfhydryl by glutathione in conjunction with glutathione S-transferase (Chen, Dodia et al. 2000; Ralat, Manevich et al. 2006). PRDX6's PLA_2 activity is regulated by phosphorylation of a Thr-177 residue, indicating that activation of the PLA_2 activity involves a conformational change (Wu, Feinstein et al. 2009; Rahaman, Zhou et al. 2012). Both the PLA_2 as well as the phospholipid hydroperoxide activities require conformational specificity in the binding of the enzyme to the phospholipid substrate (Manevich, Shuvaeva et al. 2009).

A high-resolution crystal structure for human PRDX6 has already been solved (PDB ID: 1PRX), but several alterations were made to the base protein in order to obtain

it. Most notably, it was apparently necessary to mutate a non-conserved cysteine (Cys-91) and oxidize the active site cysteine to the sulfenic form in order get diffraction-quality crystals (see Fig. 2-1) (Rhee, Chae et al. 2005). Because the active site is in the catalytic intermediate state (active site sulfenic acid), the crystal structure does not seem to completely reflect the solution structure of the reduced form of the protein. For example, a recent study showed that Thr-177, which is totally buried in the crystal structure, can be phosphorylated, indicating solvent exposure under conditions that should include the conformation in solution of the protein with Cys-47 in the reduced state (Wu, Feinstein et al. 2009; Rahaman, Zhou et al. 2012). This phosphorylation induces a further conformational change that increases PLA_2 activity and affinity for liposomes. Other biochemical data indicate that the protein undergoes conformational changes that can modulate enzyme activity, such as the conformational change that results in protein activation when binding with glutathione-S-transferase (GST) (Ralat, Manevich et al. 2006) and the one that occurs upon binding to the phospholipid head group (Manevich, Reddy et al. 2007). Considering the substantial conformational flexibility of the protein and importance of conformational changes on both enzyme activities, elucidation of the solution structure of the reduced form of the enzyme was needed.

Fig. 2-1. Crystal structure of human PRDX6 (PDB ID: 1PRX). Reproduced with permission from (Rivera-Santiago, Harper et al. 2015). Subunit A in blue; Subunit B in cyan. (A) Cartoon representation of the dimer. Cys-47 in yellow sticks; Cys-91 in orange sticks. (B) Solid surface representation of the dimer. Phospholipase catalytic triad (Ser-26, His-32, Asp-140) in red spheres; Cys-47 in yellow spheres.

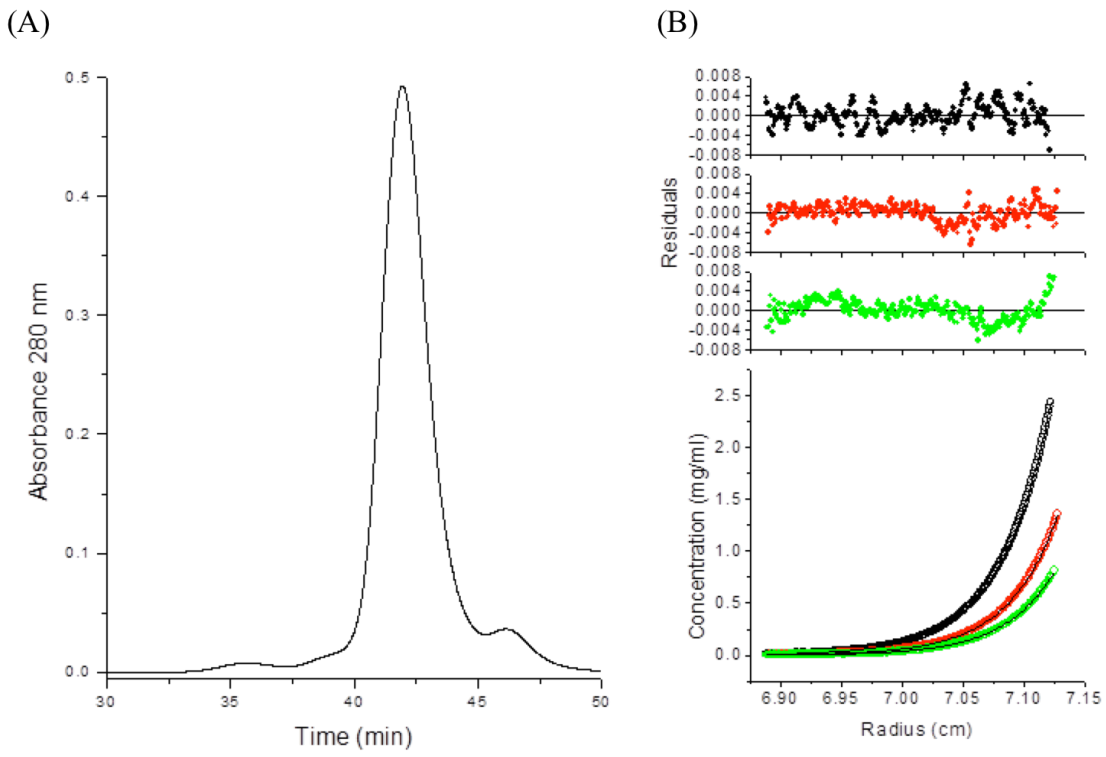


2.2. Methods

2.2.1. Expression, purification, and validation of recombinant PRDX6

As noted above, I received cross-linked PRDX6 samples that had been expressed and purified by the Fisher lab using a previously published protocol (Chen, Dodia et al. 2000; Manevich and Fisher 2005; Manevich, Reddy et al. 2007). These samples had also been validated by Ms. Harper before they were cross-linked, using size exclusion chromatography as well as sedimentation equilibrium experiments. She found that PRDX6 existed in a strong dimeric form in solution and it was in the reduced state, which was according to expectations (Rivera-Santiago, Harper et al. 2015) (see Fig. 2-2).

Fig. 2-2. Oligomer states and conformations of human PRDX6 protein preparations. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015). Experiments performed by Ms. Sandra Harper. (A) HPLC gel filtration of recombinant human PRDX6. (B) Sedimentation equilibrium of the human PRDX6 peak at three different initial loading concentrations (0.8 mg/ml in black, 0.4 mg/ml in red, and 0.2 mg/ml in green) at 26,200 rpm and 30 °C. All samples were prepared and analyzed in 20 mM Tris, 130 mM NaCl, 0.1 mM EDTA, 5% glycerol pH 7.4.

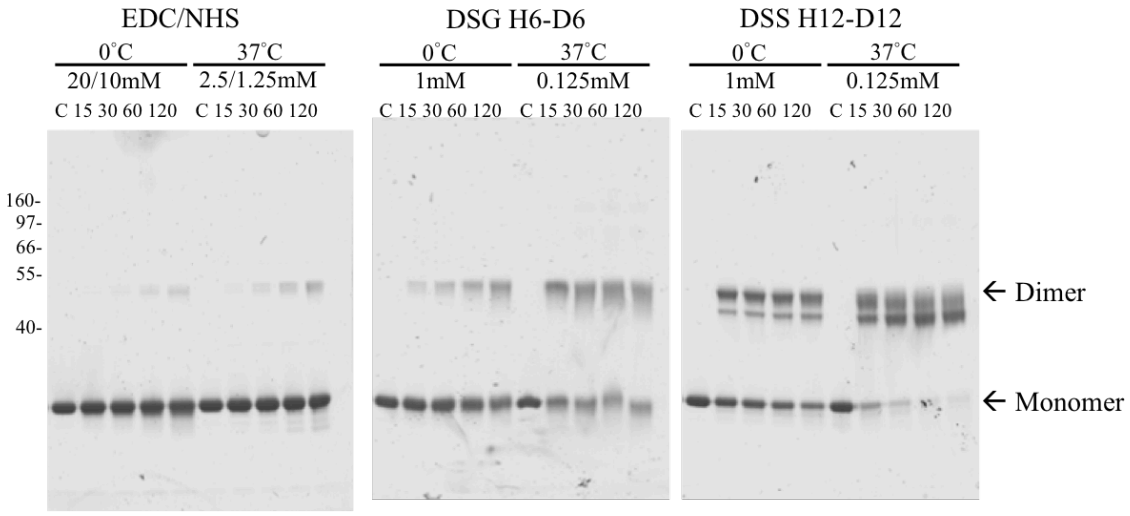


2.2.2. CX-MS analysis of PRDX6

Cross-linked PRDX6 samples for different temperatures (0°C and 37°C), gel bands corresponding to different oligomeric states (monomer vs. dimer), and different cross-linking reagents (EDC/NHS, DSG, DSS) were generated (see Fig. 2-3). Cross-linker concentrations for reactions at 0°C for EDC and sulfo-NHS were 20 mM and 10 mM, respectively, whereas they were 2.5 mM and 1.25 mM, respectively, at 37°C. In the case of DSG and DSS, concentrations were 1 mM at 0°C, and 0.125 mM at 37°C. This adjustment was done in order to account for increased molecular activity at higher temperatures.

Fig. 2-3. Chemical cross-linking of the codon optimized human PRDX6. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015). Experiments

performed by Ms. Sandra Harper. SDS-PAGE of the human PRDX6 after incubation with the indicated chemical cross-linkers. Control (C) and time points (minutes) are indicated above each lane. The position of molecular weight markers are shown on the left, and the migration of monomer and dimer bands are indicated.



The SDS-PAGE and trypsin digestion methods have been previously described (Sriswasdi, Harper et al. 2014). Briefly, samples for the reactions described above were run in SDS-PAGE gels and bands of interest were excised and digested with trypsin.

2.2.3. LC-MS analysis of EDC cross-linked samples

For identification of zero-length cross-links, the cross-linked and control human PRDX6 tryptic digests were analyzed and cross-linked sites were identified as previously described using an LTQ-Orbitrap XL^{TM} mass spectrometer (Thermo Scientific, Waltham, MA) (Sriswasdi, Harper et al. 2014). Briefly, cross-linked and control samples were analyzed in parallel using LC-MS/MS and a label-free comparison was used to identify ions specific to the cross-linked samples. Data analysis and cross-linked peptide identification was performed using the ZXMiner program (Sriswasdi, Harper et al. 2014).

Cross-links for DSG and DSS were analyzed in a similar fashion using xQuest/xProphet (Rinner, Seebacher et al. 2008).

High-confidence cross-links were assigned as intra-subunit, inter-subunit or ambiguous using several criteria. Cross-links confidently identified in the monomer band were considered to be intra-subunit. Cross-links observed in the dimer band were considered to be either intra-subunit or inter-subunit and both distances were calculated. A difference between these two distances of more than 11 Å resulted in assignment to the shorter distance possibility. Finally, for all ambiguous assignments, the crystal structure was visually examined using PyMOL for the presence of any major intervening structural elements, such as critical dimer interface contacts or an intervening helix.

2.2.4. Homology modeling of PRDX6

The PRDX6 protein sequence and the 1PRX crystal structure were submitted to MODELLER 9v11 (Sali and Blundell 1993) to generate and refine human PRDX6 solution structures. All modeling experiments were run as 50-model trials using the “very slow” refinement algorithm and discrete optimized protein energy (DOPE) score as an output. Homology modeling and refinement were performed simultaneously by including known intra- and inter-subunit cross-links as distance restraints between α-carbons imposed at 11.0 +/- 0.1 Å. Each model was subject to 1,000 iterations and 10 optimization repeats. The completed models were then analyzed according to their DOPE score, and the highest-scoring model under this criterion was chosen for further analysis. Molecular graphics were illustrated using Open-Source PyMOL Version 1.3 (Schrödinger, LLC), which also was used to calculate distances between α-carbons of

cross-linked glutamate, aspartate, and lysine. No cross-links were observed involving the N-terminal amine group.

2.3. Results and Discussion

2.3.1. Analysis of CX-MS cross-links on PRDX6 crystal structure

After I had perofmed the LC-MS analysis for the EDC/NHS cross-links via the criteria described in Section 2.2.3, I used ZXMiner (Sriswasdi, Harper et al. 2014) to identify potential zero-length cross-links in the dimeric protein. Once identified, I assigned high confidence cross-links as intra-subunit, inter-subunit or ambiguous using several criteria. I considered cross-links confidently identified in the monomer band to be intra-subunit. I also considered cross-links observed in the dimer band to be either intra-subunit or inter-subunit, and both distances were calculated. A difference between these two distances of more than 11Å resulted in assignment to the shorter distance possibility. Finally, for all ambiguous assignments, the crystal structure was visually examined using PyMOL (Schrödinger, LLC) for the presence of any major intervening structural elements, such as critical dimer interface contacts or an intervening helix.

I was able to assign at least 12 of 15 high confidence zero-length cross-links I detected as either intra-subunit or inter-subunit, with the others remaining ambiguous. Surprisingly, 8 of these 12 observed distances were substantially larger than the expected 12Å or 16Å limits (see Table 2-1). This indicated that the solution structure of the reduced form of the enzyme deviated substantially from the crystal structure (see Fig. 2-1). Therefore, a new solution structure of the PRDX6 protein had to be developed.

Table 2-1: Human PRDX6 cross-links identified using EDC. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015).

Cross-link Group	z[a]	MH+ (Da)[b]	Mass Error (ppm)	Sequence-A[c]	Sequence-B[c]	Distance (Å)[d]	Cross-linked Residues
Inter-Chain Cross-links							
E01	4	1656.93474	4.6	AA[K]LAPEFAK	[E]LPSGK	19.8	56-210
E02	3	2640.42255	4.1	[E]LAILLGM#LDPAEKDEK	ELPSG[K]K	27.5	109-215
E03	3	2726.30843	5.7	DINAYNC(EE)PTEK	GVFT[K]ELPSGK	15.8	92/93-209
E04	5	3040.64670	5.3	LIALSIDSV[E]DHLAWSK	GVFT[K]ELPSGK	22.4	77-209
E05	4, 5	3044.62645	2.9	[E]LAILLGM#LDPAEKDEK	GVFT[K]ELPSGK	33.0	109-209
E06	4	3238.80547	4.2	ELAILLGM#LDPAEK[D]EK	VVISLQLTAE[K]R	28.4	123-173
E07	3, 4	3727.85644	1.8	ELAILLGM#LDPAE[K]DEK	(D)G(D)SVM#VLPTIPEEEAK	37.9	122-183/185
Intra-Chain Cross-links							
E08	3	1243.74389	6.2	[K]LFPK	[E]LPSGK	21.0	200-210
E09	3	1748.92572	5.0	GVFT[K]ELPSGK	YTPQ[P]	18.9	209-224
E10	4	1750.99291	4.9	[K]LFPKGVFTK	YTPQ[P]	12.6	200-224
E11	3	1943.10067	4.8	VVISLQLTAE[K]R	YTPQ[P]	7.0	173-224
E12	3, 4	2609.26208	4.5	AA[K]LAPEFAK	DINAYNC(EE)PTEK	10.3	56-92/93
Ambiguous Cross-links[e]							
E13	3	2514.39782	5.5	[E]LAILLGM#LDPAEKDEK	[K]LFPK	49.9 | 40.0	109-200
E14	5	2770.40894	1.3	FHDFLG[D]SWGILFSHPR	ELPSG[K]K	39.3 | 38.9	31-215
E15	4	2872.47512	5.2	AA[K]LAPEFAK	(D)G(D)SVMVLPTIPEEEAK	20.3 | 26.3	56-183/185

[a] Observed charge states of the cross-linked peptide.

[b] Observed monoisotopic mass of observed peptide.

[c] []: cross-linked residue; (): potential cross-linked residue (ambiguous location); #: methionine oxidation.

[d] All distances are between alpha-carbons of cross-linked residues in the human PRDX6 crystal structure (PDB ID: 1PRX).

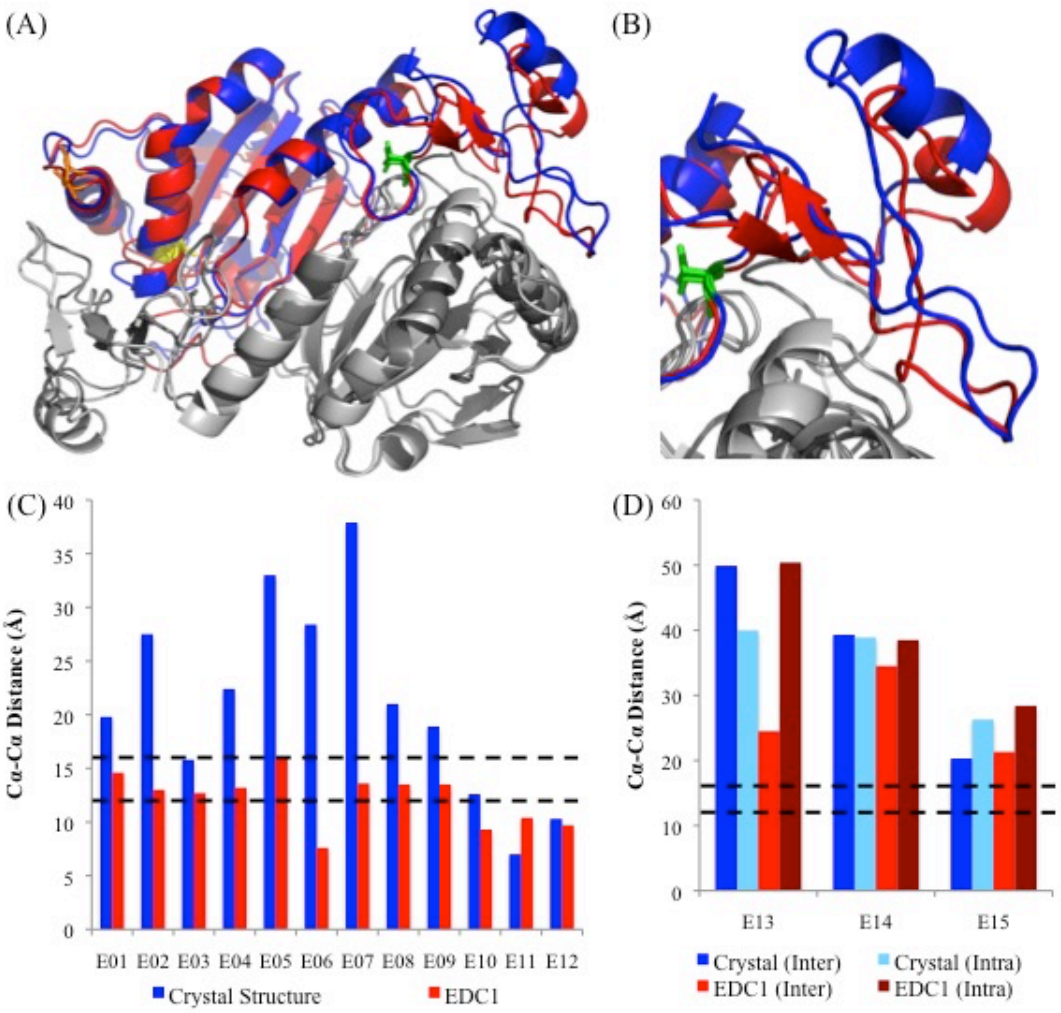
[e] Ambiguous cross-links are cross-links that could not be definitively assigned as inter- or intra-molecular; both distances are described in the table entry, separated by a slash. The inter-chain distance is before the slash, and the intra-chain distance is after the slash.

2.3.2. Model Development and Validation

I used MODELLER as described in Section 2.2.4 to refine the 1PRX crystal structure by imposing α-carbon distance constraints for the 12 unambiguously assigned cross-links. I then compared the resulting preliminary model, hereafter referred to as the EDC1 model, to the 1PRX crystal structure via a superposition (see Fig. 2-4a). I saw that most major structural features in the PRDX6 homodimer remained unchanged, with an overall RMSD of 1.4 Å. Moreover, an analysis of the Ramachandran angles using Coot (Emsley, Lohkamp et al. 2010) for this model also revealed that over 90% of those were found in preferred or allowed conformations for EDC1, reinforcing the validity of the model. The top 5 models generated, as ranked by discrete orbital proximal energy (DOPE) score, were also very similar to one another, indicating convergence on a best model.

Fig. 2-4. Initial Molecular Model of human PRDX6 using EDC Cross-links.

Reproduced with permission from (Rivera-Santiago, Harper et al. 2015). (A) Superposition of the crystal structure (blue) and EDC1 model (red) (RMSD: 1.4Å). For simplicity, only the A chain on both dimers is highlighted; both B chains are in gray. Cys-47 is shown in yellow sticks, Cys-91 is shown in orange sticks, and Thr-177 is shown in light green sticks. (B) Close-up of the C-terminal region on chain A, which shows the largest difference relative to the crystal structure (residues 190-224). (C) α-carbon distances between residues identified using EDC cross-links for the crystal structure and EDC1 model. Expected distance cutoffs are 12Å for well-ordered regions (lower black dashed line), and 16Å for disordered regions (upper black dashed line). (D) α-carbon distances between residues with ambiguous assignments as to whether they are inter-chain or intra-chain.

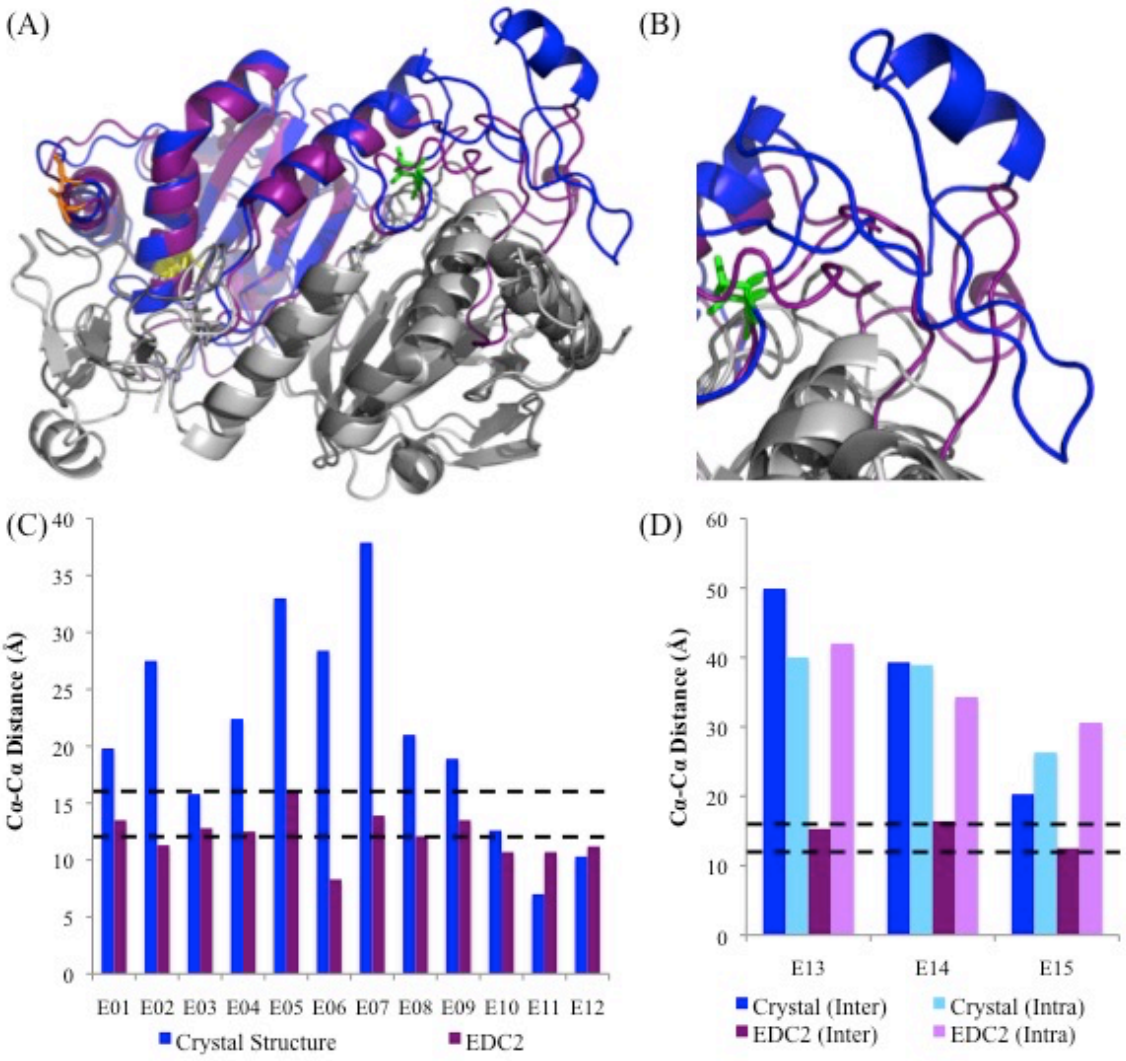


The EDC1 model also pinpointed the area of maximum variability between the two structures as involving residues 190-224 (see Fig. 2-4b), with an RMSD of 2.1 Å. The distances for the cross-links used as distance restraints were all below the 16Å cutoff point (see Fig. 2-4c). Additionally, the α-carbon distances for the three ambiguous cross-links that I did not use in this structural refinement displayed smaller inter-chain distances compared with the intra-chain distances in the EDC1 model (see Fig. 2-4d). This trend, coupled with the observation that all three cross-links had impeding intervening structures between their side chains for a putative intra-chain cross-link, enabled me to assign them as inter-chain cross-links. However, their distance remained

larger than the expected 16Å cutoff, which indicated that I needed to refine my model further.

I then produced a second model, hereafter referred to as the EDC2 model, using MODELLER to refine the 1PRX crystal structure by imposing expected α-carbon distance constraints for all 15 EDC cross-links. Similar to the EDC1 model, most major structural features in the PRDX6 homodimer remained unchanged relative to the crystal structure with an overall RMSD of 1.7 Å (see Fig. 2-5a) and the area of maximum variability between the two structures was residues 190-224 with an RMSD of 2.9 Å (see Fig. 2-5b). I also analyzed this model via Ramachandran plot, and I found it had over 90% of its bond angles in allowed or preferred conformations. As before, the top 5 models by DOPE score were also structurally convergent. In this model, all 15 EDC cross-links were now within the 16Å maximum distance (see Fig. 2-5c-d) indicating that all zero-length cross-link data supported this structure.

Fig. 2-5. Molecular Model of human PRDX6 using all EDC Cross-links. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015). (A) Superposition of the crystal structure (blue) and EDC2 model (purple) (RMSD: 1.7Å). For details on color schemes and highlighted residues, see Fig. 4. (B) Close-up of the C-terminal region on chain A, which shows the largest difference relative to the crystal structure (residues 190-224). (C) α-carbon distances between residues identified using EDC cross-links for the crystal structure and EDC2 model. Expected distance cutoffs are 12Å for well-ordered regions (lower black dashed line), and 16Å for disordered regions (upper black dashed line). (D) α-carbon distances between residues with initially ambiguous assignments as to whether they are inter-chain or intra-chain.



Next, I performed independent CX-MS experiments on human PRDX6 samples that were previously cross-linked as described in Section 2.2.2 using the homo-bifunctional cross-linkers DSG (H6/D6 mixture; 7.7 Å linker) and DSS (H12/D12 mixture; 11.4 Å linker) to further evaluate the PRDX6 solution structure (see Fig. 2-3). Unlike the data analysis for the EDC experiments, these cross-linked peptides were identified using xQuest/xProphet (Rinner, Seebacher et al. 2008). The resulting high confidence cross-links (see Tables 2-2 and 2-3) were mapped onto both the crystal structure and the EDC2 structure (see Fig. 2-6a-b). Because the lengths of the spacer arms between the

reactive groups of these cross-linkers (Paramelle, Miralles et al. 2013), I used maximum α-carbon distances of 22Å for DSG and 26Å for DSS, which is consistent with previous studies using these cross-linkers (Sriswasdi, Harper et al. 2014).

Table 2-2: Human PRDX6 cross-links identified using DSG. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015).

Cross-link Group	z[a]	MH+ (Da)[b]	Mass Error (ppm)	Sequence-A[c]	Sequence-B[c]	Cross-linked Residues
Inter-Chain Cross-links						
D01	4	1898.042	-4.8	AA[K]LAPEFAK	ELPSG[K]K	56-215
D02[d]	4	1955.988	-5.2	DE[K]GMPVTAR	ELPSG[K]K	125-215
D03	4	2302.258	0.5	GVFT[K]ELPSGK	AA[K]LAPEFAK	209-56
D04	4	2376.195	-1.9	GVFT[K]ELPSGK	DE[K]GM#PVTAR	209-125
D05	4	2570.369	-1.5	VVISLQLTAE[K]R	DE[K]GM#PVTAR	173-125
Intra-Chain Cross-links						
D06	4	2630.409	-0.1	L[K]LSILYPATTGR	DE[K]GMPVTAR	144-125
D07	4	2826.431	3.8	DGDSVM#VLPTIPEEEA[K]K	ELPSG[K]K	199-215
D08	3	2883.675	-1.5	L[K]LSILYPATTGR	VVISLQLTAE[K]R	144-173
D09	5	3263.753	-0.6	NV[K]LIALSIDSVEDHLAWSK	LAPEFA[K]R	67-63
Ambiguous Cross-links						
D10	4	2176.226	-1.4	AA[K]LAPEFAK	LFP[K]GVFTK	56-204

[a] Observed charge states of the cross-linked peptide.

[b] Observed monoisotopic mass of observed peptide.

[c] []: cross-linked residue; (): potential cross-linked residue (ambiguous location); #: methionine oxidation.

[d] Multiple variants of this cross-link have been identified in this dataset.

Table 2-3: Human PRDX6 cross-links identified using DSS. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015).

Cross-link Group	z[a]	MH+ (Da)[b]	Mass Error (ppm)	Sequence-A[c]	Sequence-B[c]	Cross-linked Residues
Inter-Chain Cross-links						
S01[d]	4	1940.09	-4.2	AA[K]LAPEFAK	ELPSG[K]K	56-215
S02[d]	4	1998.038	-3.7	DE[K]GMPVTAR	ELPSG[K]K	125-215
S03[d]	4	2344.296	-3.6	GVFT[K]ELPSGK	AA[K]LAPEFAK	209-56
S04[d]	4	3469.896	-0.8	ELAILLGM#LDPAE[K]DEK	L[K]LSILYPATTGR	122-144
S05	5	3523.839	-3.7	ELAILLGM#LDPAE[K]DEKGM#PVTAR	ELPSG[K]K	122-215
Intra-Chain Cross-links						
S06[d]	4	2688.448	-1.6	L[K]LSILYPATTGR	DE[K]GM#PVTAR	144-125
S07	4	2925.727	0	L[K]LSILYPATTGR	VVISLQLTAE[K]R	144-173
S08	4	3830.905	-0.8	DINAYNCEEPTE[K]LPFPIIDDR	AA[K]LAPEFAK	97-56

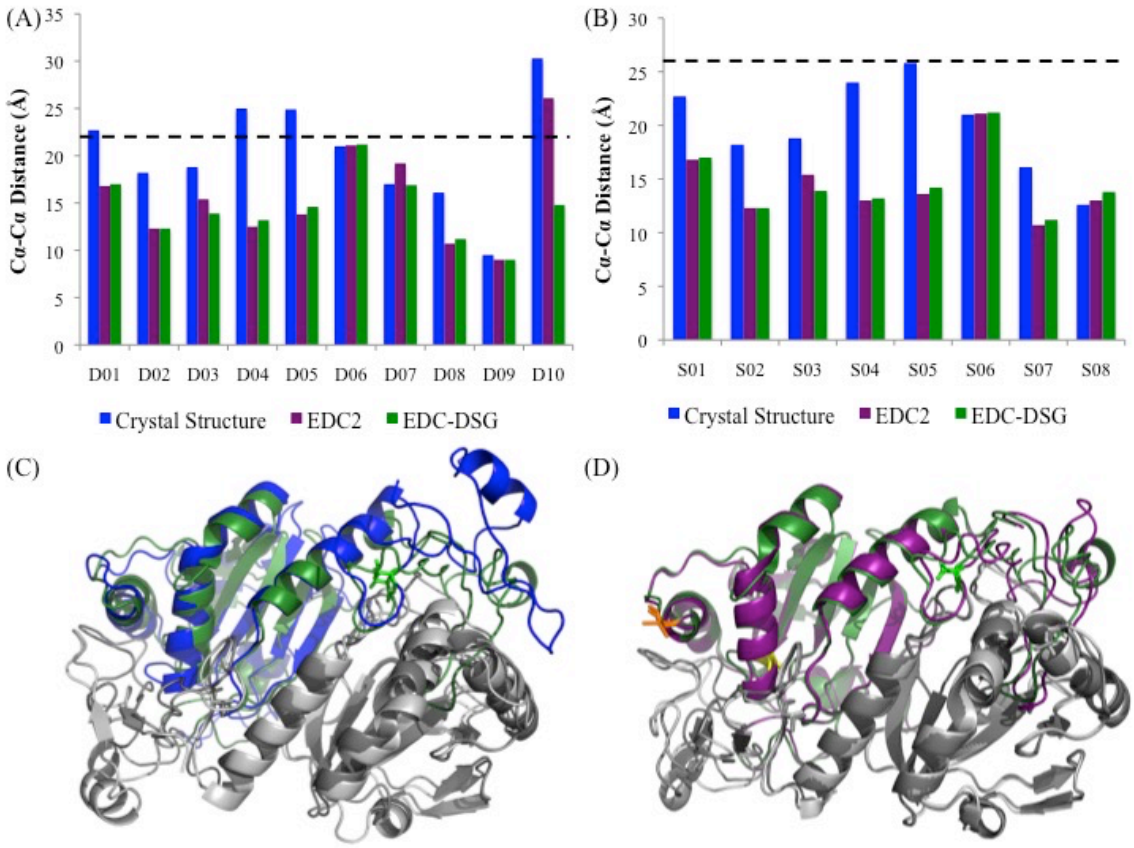
[a] Observed charge states of the cross-linked peptide.

[b] Observed monoisotopic mass of observed peptide.

[c] []: cross-linked residue; (): potential cross-linked residue (ambiguous location); #: methionine oxidation.

[d] Multiple variants of this cross-link have been identified in this dataset.

Fig. 2-6. Evaluation of Solution Structure Using Non-Zero-Length Cross-linkers. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015). (A) Histogram analysis of α-carbon distances between residues identified using DSG cross-links for the crystal structure (blue), EDC2 model (purple), and Final model (forest green). Expected distance cutoff is 22Å (dashed black line). (B) Histogram analysis of α-carbon distances between residues identified using DSS cross-links for the crystal structure, EDC2 model, and Final model. Expected distance cutoff is 26Å. (C) Superimposition of the crystal structure (blue) and Final model (forest green) (RMSD: 1.6Å). For details on color schemes and highlighted residues, see Fig. 4. (D) Superimposition of the EDC2 model (purple) and Final model (forest green) (RMSD: 0.7Å).



When I analyzed the 1PRX crystal structure using these cross-links, 4 of 10 DSG cross-links exceeded the maximum expected distance threshold. This confirms that the solution structure of the reduced protein significantly differed from that of the catalytic intermediate. Interestingly, 3 of these 4 cross-links were within expected limits when the EDC2 model was considered. For DSS, which has an even longer spacer arm, none of the cross-links exceeded the threshold for that cross-linker on any of the structures, which indicated the inability of this longer spacer arm to distinguish between PRDX6's alternative conformations. I then evaluated the single DSG cross-link (hereafter referred to as D10) that continued to exceed the expected distance threshold in the EDC2 structure. Interestingly, I found that this cross-link involved Lys204, which is located within the most variable region of the protein as described above. D10 (see Table 2-2)

was also notable because its assignment as an inter-chain or intra-chain cross-link was ambiguous when compared to the 1PRX crystal structure. This was due to α-carbon distances of 30.3Å and 36.3Å for inter-chain and intra-chain interactions, respectively. There also were intervening structural elements for both possible types of cross-link. However, I resolved this controversy when I examined it on the EDC2 model, as the inter-chain interaction no longer had major intervening structural elements and displayed an α-carbon distance of 26.1Å, as opposed to 42.5Å for the intra-chain interaction.

In order to further explore the impact of the D10 DSG cross-link on the protein structure, I generated a final solution structure using MODELLER to refine the 1PRX crystal structure further by imposing expected α-carbon distance constraints for all 15 EDC cross-links, plus the D10 distance constraint. I found that DSG and DSS cross-links were within expected limits in this model (see Fig. 2-6a-b). Next, I compared the final structure to the crystal structure (see Fig. 2-6c). I once again observed that the overall backbone structure was similar, with an RMSD of 1.6 Å and the RMSD of the C-terminal tail (residues 190-224) of 2.4 Å, which indicated that this region was actually slightly closer to the crystal structure than in the EDC2 model. I also analyzed this model was via Ramachandran plot, and once again I found that over 90% of its bond angles were in preferred or allowed regions. The top 5 models by DOPE score also converged structurally. The final model was also directly compared to the EDC2 model (see Fig. 2-6d). As expected, the differences between the EDC2 model and the final solution structure were minor, with an overall RMSD of 0.74Å and an RMSD of 1.0 for the C-terminal region. Their distributions of Ramachandran bond angles were quite similar overall.

2.4. Conclusions

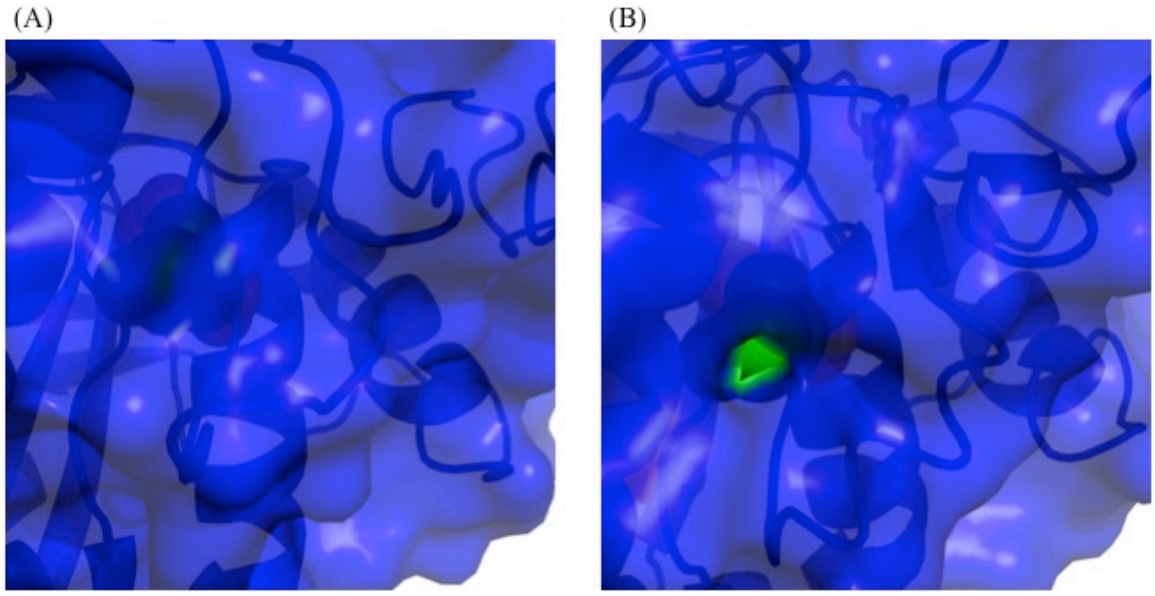
In summary, I was able to use structural refinement with CX-MS distance constraints to determine an experimentally validated solution structure of the reduced form of PRDX6. The final structure fits distance constraints of all high confidence cross-links using reagents with three different cross-link spacer arms (0, 7.7 and 11.4 Å). Consistent with what was observed in prior CX-MS studies by the Speicher lab and others (Leitner, Walzthoeni et al. 2010; Harper, Sriswasdi et al. 2013; Sriswasdi, Harper et al. 2014; Sriswasdi, Harper et al. 2014), the zero-length cross-links proved to be the most useful for identifying conformational differences between the solution structure of the reduced protein and the crystal structure of the catalytic intermediate.

In contrast, the longest cross-linker (DSS) was too imprecise for me to unambiguously identify any discrepancies between the solution structure and the crystal structure. The differences between the solution structure and crystal structure were mostly, but not entirely, located within the C-terminal tail of the protein (residues 190-224). This region is primarily comprised of coil, which probably contributes to its plasticity. However, the fact that all 33 distance constraints can be satisfied by a single structure suggests, but does not prove, that the solution structure of the protein as isolated herein is primarily a single conformation rather than an ensemble of inter-converting structures.

An important note is that significant changes in the PRDX6 protein's structure are not limited to residues 190-224. Another functionally important change is the solvent accessibility of Thr-177. As noted above, this residue is phosphorylated in response to certain stimuli and increases PLA2 activity and affinity for liposomes (Wu, Feinstein et al. 2009), but it is completely buried in the crystal structure. In contrast, this residue is

substantially solvent exposed in my EDC2 model and in the final solution structure (see Fig. 2-7). Taken together, these data suggests to me that the reduced form, should be capable of being phosphorylated, but not the oxidized form of the enzyme.

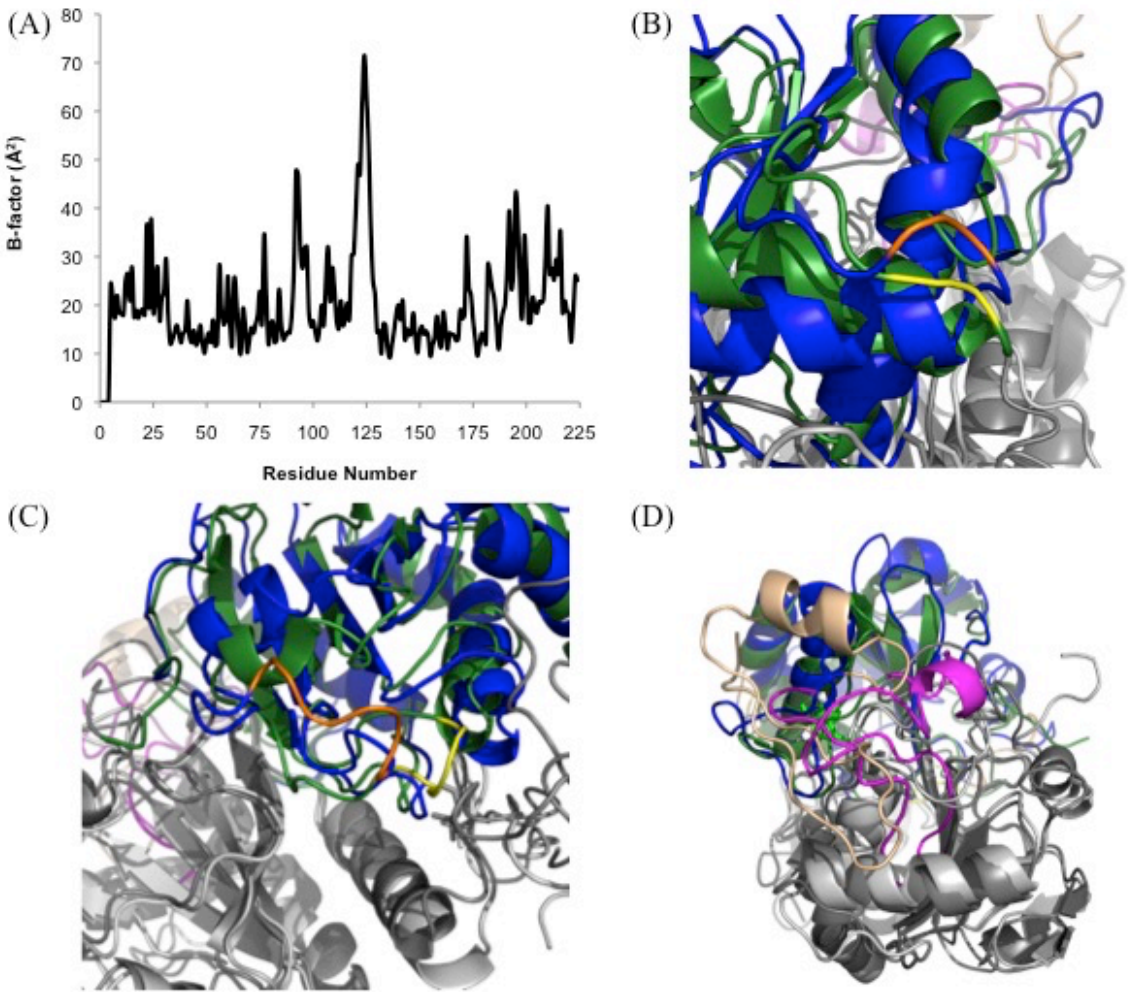
Fig. 2-7. Surface Accessibility of Thr-177 in PRDX6 Crystal and Solution Structures. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015). Closeup of protein surface display for Thr-177 (green spheres) in (A) 1PRX crystal structure, and (B) EDC2 solution structure.



I also evaluated whether the conformational differences between the solution and crystal structures correlated with the most flexible regions in the crystal structure. An analysis of the Debye-Waller temperature factors (B-factors) in the crystal structure (see Fig. 2-8a), shows that residues 92-93 and 121-126 exhibited the highest B-factors. The remainder of the protein had B-factors less than 40, which are indicative of relatively rigid structures. A comparison of residues 92-93 in the crystal structure and the final solution structure show that this coil region is not any more variable than its flanking

regions (see Fig. 2-8b). Residues 121-126 encompass another coil region that also does not exhibit substantially greater variability between structures than other adjacent coil regions.

Fig. 2-8. Relationship of B-factors and Variations Between the PRDX6 Crystal and Solution Structures. Reproduced with permission from (Rivera-Santiago, Harper et al. 2015). (A) Analysis of B-factors for 1PRX crystal structure. (B) Superposition of 1PRX crystal structure (blue) and Final model (forest green) highlighting residues Glu-92 and Glu-93 (orange and yellow for 1PRX and Final, respectively). (C) Superposition of 1PRX crystal structure (blue) and Final model (forest green) highlighting residues Glu-92 and Glu-93 (orange and yellow for 1PRX and Final, respectively). (D) Superposition of 1PRX crystal structure (blue) and Final model (forest green) highlighting residues 190-224 (wheat and magenta for 1PRX and Final, respectively).



Interestingly, however, this region is significantly more compact in the final solution model compared with the crystal structure (see Fig. 2-8c). Finally, as I noted above, the C-terminal region encompassed by residues 190-224 shows the largest variation between the two structures (see Fig. 2-8d), but this region does not have high B-factors in my previous analysis. Interestingly, an analogous region in 2-Cys peroxiredoxins also exhibits conformational change as part of the catalysis process (Wood, Schroder et al. 2003).

However, one key difference is that the 2-Cys C-terminal region experiences localized unfolding in response to the breaking of a disulfide bond, whereas the

analogous region in PRDX6 becomes more compact. Overall, I did not see a strong correlation between regions of high flexibility in the crystal structure and regions involved in conformational changes associated with transition from the reduced form to the peroxidase catalytic intermediate form of the enzyme. That suggests that these conformational changes play an important role in the catalytic mechanism and do not simply reflect variations in intrinsically mobile regions of the protein. It also suggests that these conformational changes associated with the peroxidase mechanism could help regulate PLA2 activity. That is to say, when the enzyme is in the reduced form, but not the catalytic intermediate state, Thr-177 is solvent exposed and can be phosphorylated thereby increasing PLA2 activity and affinity for liposomes.

This study was also a strong showcase for the advantages of using zero-length CX-MS for probing subtle but biologically important conformational changes. I further demonstrated the resolving power of zero-length CX-MS here by identifying a greater number of high-confidence cross-links EDC (15 unique cross-link sites) as compared to DSG (10 cross-link sites) or DSS (8 cross-link sites). This larger number of zero-length cross-links was identified despite less extensive overall apparent cross-linking relative to the DSG and DSS reactions, as indicated by the less extensive formation of covalently linked homodimers on SDS gels (see Fig. 2-3). More importantly, EDC cross-linking represented the most sensitive test for conformational changes. When observed cross-links were compared to the crystal structure, 8 of 12 EDC cross-links (67%) involved residues that were further apart than the maximum likely distance, whereas only 4 of 10 DSG (40%) and 0 of 8 DSS (0%) cross-links were outside expected maximum distances.

2.5. Summary

In summary, I found the solution structure of the reduced form of PRDX6 to be of particular interest, because multiple biochemical studies had identified conformational changes of the protein that are associated with changes in enzyme activity and the only reported high resolution crystal structure is of the peroxidase catalytic intermediate form of the protein (Choi, Kang et al. 1998). My experimentally supported solution structure is consistent with multiple biochemical studies that indicate PRDX6 is a protein with substantial plasticity that can affect both known enzyme activities of this protein. It also yields novel insights regarding the nature of the changes that occur during the peroxidase catalysis and the likely interplay between regulation of the two enzyme activities as indicated by the solvent accessibility of Thr-177. The reported structure provided me with an important structural reference for this protein, as the reduced form of PRDX6 is much more commonly encountered than the peroxidase catalytic intermediate. These insights combined to give me a greater understanding of PRDX6.

Chapter 3: Optimizing analysis of zero-length cross-linked peptides using Q Exactive Plus mass spectrometer

In this chapter, I will discuss efforts made to optimize the zero-length CX-MS method to a more state-of-the-art instrument. Part of the content in this chapter has been published as a manuscript in Methods (Rivera-Santiago, Sriswasdi et al. 2015). In this study, Ms. Sandra Harper expressed, purified, and cross-linked the mini-spectrin protein samples there were used as part of an instrument comparison. I then performed LC-MS and ZXMiner analysis on the cross-links in order to generate the instrument comparison table.

3.1. Background

One of my major goals for the zero-length CX-MS method was to eventually be able to interrogate complex protein systems. The Speicher laboratory, including myself, has had previous success with purified protein samples (Harper, Sriswasdi et al. 2013; Sriswasdi, Harper et al. 2014; Rivera-Santiago, Harper et al. 2015), and the method to analyze those is well-established. However, I would like to expand this capability to examine more complex biologically relevant samples. Furthermore, I wanted to do so in as close to a native setting as possible, in order to maximize the relevance of the structural information I am gathering. In order to do so, I needed to combine an optimal method with the optimal instrument for the job.

The Speicher lab has recently come into the possession of a Thermo Q Exactive Plus mass spectrometer (hereafter referred to as QE+), which is a significantly different instrument from the one that was used to develop the zero-length CX-MS protocol used in previous studies (Harper, Sriswasdi et al. 2013; Sriswasdi, Harper et al. 2014; Sriswasdi, Harper et al. 2014; Rivera-Santiago, Harper et al. 2015), which was the

Thermo LTQ Orbitrap XL mass spectrometer, as described in Chapter 1. One of the major differences lies in the quality of MS and MS/MS data – because the QE+ is a much more state-of-the-art instrument than the LTQ Orbitrap XL, it has the ability to generate significantly more MS/MS spectra per full scan, thus leading to significantly greater depth of analysis. This increased depth of analysis results in the elimination of the need for a two-part discovery and targeted approach, as the QE+ is entirely capable of collecting high-resolution MS/MS spectra on the first run. The impact of the differences in MS data collection speed and quality between the LTQ Orbitrap XL and the QE+ is best demonstrated by a direct comparison analysis between the two instruments. We performed this analysis as part of a previous study, and we used a purified mini-spectrin (Harper, Li et al. 2010) protein sample. The QE+ significantly out-performs the LTQ Orbitrap XL (see Table 3-1) in terms of both instrument time required to produce the data as well as number of total and unique peptides identified, thus validating our interest in optimizing the instrument parameters for it. Moreover, this difference would only be magnified with increasing sample complexity. This means that optimizing the QE+ MS/MS data generation and analysis pipeline ties into the long-term goal of interrogating more complex protein samples.

Table 3-1: Comparison of LTQ Orbitrap and QE+ for zero-length CX-MS experiments using mini-spectrin[a]

Band[b]	Instrument	Number of MS Runs[c]	Total MS Time (h)	Cross-linked Peptides	
				Total[d]	Unique[d]
Dimer	LTQ Orbitrap	12	24	48	22
Dimer	Q Exactive +	3	6	40	23
Tetramer	LTQ Orbitrap	12	24	71	38

Tetramer	Q Exactive +	3	6	107	59

[a] Reproduced with permission from (Rivera-Santiago, Sriswasdi et al. 2015).

[b] The mini-spectrin band used for the in-gel digest.

[c] All MS runs consisted of 2-hour LC gradients and injection sizes of about 0.5 µg of protein.

[d] Peptides have FDR = 0.0.

In spite of the superior high-resolution MS/MS data collection capabilities of the QE+ instrument, some key differences between the QE+ and the LTQ Orbitrap XL have to be taken into account. One such difference is their method of fragmenting ions for MS/MS analysis. The LTQ Orbitrap XL uses its ion trap to fragment ions via collision-induced dissociation (CID). This typically ensures that no ions with a m/z correspondent to that of the precursor ion remain. Because the QE+ does not have an ion trap (it is a hybrid quadrupole-orbitrap mass spectrometer), it instead performs ion fragmentation via higher-energy collisional dissociation (HCD). Because of this, fragmentation is typically based on a normalized collision energy (NCE) setting specified at the start of the data acquisition run. What this typically results in is a distribution of ions that were properly fragmented, ions that were improperly fragmented, usually resulting in MS^3 ions, and ions that were not fragmented at all, thus remaining as intact precursor. These last ions are the most troublesome, as they do not correspond to any b- or y-ions in the fragmentation pattern.

In order to address this issue, I decided to employ a two-pronged optimization strategy. First, I optimized the parameters for the zero-length CX-MS data acquisition method employed by the QE+ instrument in order to maximize the depth of analysis. Secondly, I collaborated with Dr. Sira Sriswasdi, Ms. Sandra Harper, and Dr. David W. Speicher in order to improve the ZXMiner software to address the limitations in analysis of QE+ data. This led to a significant improvement in the quality of zero-length CX-MS

data. The procedures and analysis for both of the approaches that led to these results are described below.

3.2. Methods

3.2.1. Preparation of erythrocyte membrane white ghost (WG) samples

Fresh blood samples were collected from healthy volunteers with informed written consent, using protocols approved by the The Wistar Institute institutional ethical review board. Erythrocyte membranes, commonly called “white ghosts” were prepared by Ms. Sandra Harper as described in Speicher, et al. (Speicher, Weglarz et al. 1992) with several modifications. Briefly, blood was stored at 4 °C for 2 days prior to processing in order to allow reticulocyte maturation. After centrifugation (2000 x g, 20 min., 4 °C) and removal of the serum and buffy coat, the packed red cells were diluted to 50% hematocrit in 10 mM phosphate, 130 mM sodium chloride pH 7.4 and passed through a plasmodipur filter (Accurate Chemical and Scientific Corp, Westbury, NY) to remove any contaminating leukocytes. The filtered sample was centrifuged at 2000 x g for 20 min at room temperature. The supernatant was aspirated and the red cells diluted to 50% hematocrit in 10 mM phosphate, 130 mM sodium chloride at pH 7.4. The red cell suspension was layered onto Lympholyte-H cell separation media (Accurate Chemical and Scientific Corp, Westbury, NY) and centrifuged for 20 min and 740 x g at room temperature to remove any contaminating lymphocytes. The supernatant was aspirated and the red cells were diluted in 10 mM phosphate, 130 mM sodium chloride at pH 7.4. The remaining washing and lysis steps were carried out as previously described. This protocol was performed twice across the course of the WG parameter optimization trials. The subsequent samples will be hereafter referred to as WG-A and WG-B.

3.2.2. Cross-linking of WG samples

WG samples were cross-linked using a previously described protocol (Li, Harper et al. 2010; Sriswasdi, Harper et al. 2014). Briefly, EDC was used in conjunction with sulfo-NHS to cross-link WG samples at 0 °C. For WG-A, aliquots were taken at 30 minutes for 10 mM EDC/5 mM sulfo-NHS. For WG-B, aliquots were taken at 15, 30, 60, and 120 minutes for 10 mM EDC/5 mM sulfo-NHS, and at 15 and 30 minutes for 20 mM EDC/10 mM sulfo-NHS. The reactions were then quenched by adding 20 mM DTT and incubating on ice.

3.2.3. New ZXMiner software

I collaborated with Ms. Sandra Harper, Dr. Sira Sriswasdi, and Dr. David Speicher to evaluate GM scores for ZXMiner spectra using XlinkInspector (see Chapter 1). A new version of ZXMiner (v. 0.2 beta) was developed by Dr. Sriswasdi and used for these studies.

3.2.4. Parameter optimization for QE+ instrument

LC-MS runs were performed on the WG samples described above, then analyzed with ZXMiner v. 0.2 beta using the same parameters as in previous studies (Harper, Sriswasdi et al. 2013; Sriswasdi, Harper et al. 2014; Sriswasdi, Harper et al. 2014; Rivera-Santiago, Harper et al. 2015). Parameters were compared to each other based on their performance in the following categories, prioritized as presented:

1. Peptides with GM scores higher than the highest-scoring false positive (hereafter referred to as high-confidence IDs);
2. Unique peptides with GM scores higher than the highest-scoring false positive (hereafter referred to as unique high-confidence IDs);

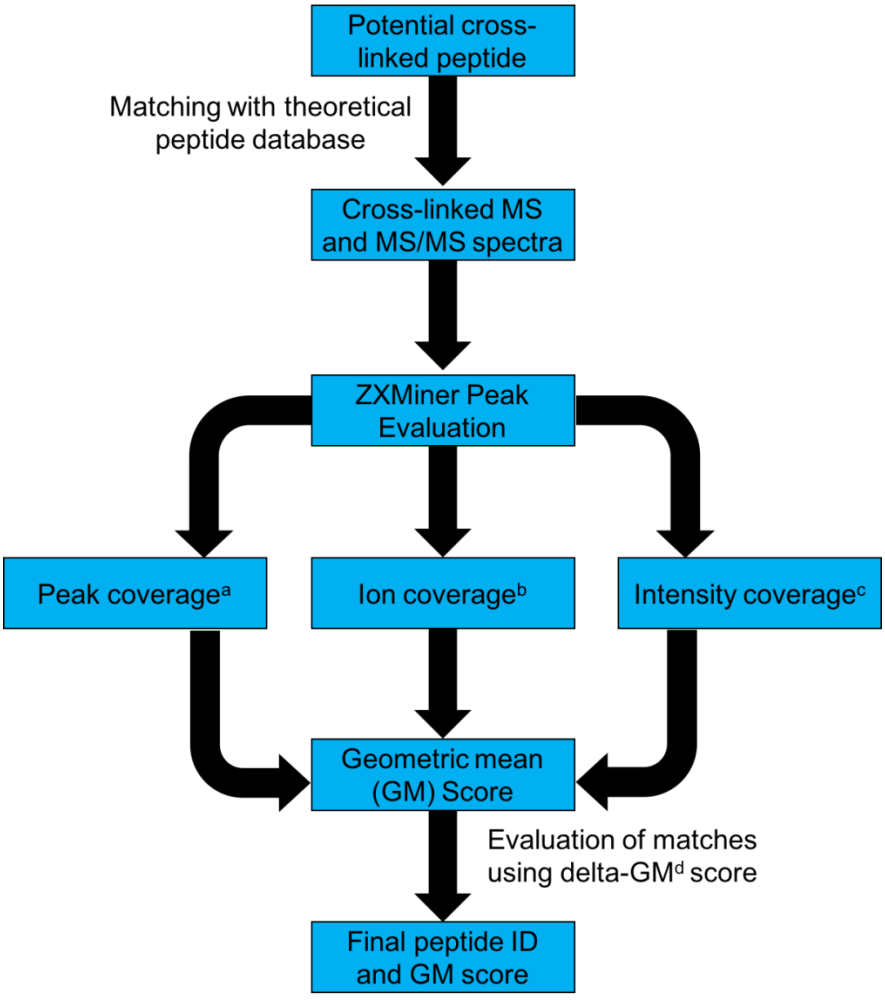
3. Peptides with GM scores lower than the highest-scoring false positive, but higher than the second-highest scoring false positive (hereafter referred to as high- and medium-confidence IDs);
4. Unique peptides with GM scores lower than the highest-scoring false positive, but higher than the second-highest scoring false positive (hereafter referred to as high- and medium-confidence IDs);
5. Peptides with GM scores higher than 0.2 (hereafter referred to as base IDs);
6. Unique peptides with GM scores higher than 0.2 (hereafter referred to as unique base IDs);
7. Total entries.

Parameter optimization experiments for LC gradient length, m/z peak isolation width, amount protein used per injection, and MS/MS AGC were performed on WG-A samples. All other experiments were performed on WG-B samples. All experiments run with a given set of LC-MS parameters for WG-B were considered to be a single experiment in data analysis via ZXMiner.

3.3 Results and Discussion

In order to optimize the parameters on the QE+ for zero-length CX-MS, I collaborated with Ms. Sandra Harper, Dr. Sira Sriswasdi, and Dr. David Speicher to determine how to best account for the differences in LTQ Orbitrap XL spectra and QE+ spectra that contributed to the lower the GM scores for QE+ data. A summary of the process by which these GM scores are generated is presented in Fig. 3-1. We decided that the presence of peaks in the precursor ion region was the most important factor. To address this, Dr. Sriswasdi developed a new version of ZXMiner (version 0.2 beta).

Fig. 3-1: Summary of geometric mean (GM) score determination for ZXMiner data. a. Peak coverage represents score for proportion of peaks assigned as b- or y-ions, precursor-related ions, or neutral losses (scored from 0-1); b. Ion coverage represents proportion of total b- and y-ions assigned for an MS/MS spectrum (scored from 0-1); c. Intensity coverage represents proportion of total peptide intensity assigned as b- or y-ions, precursor-related ions, or neutral losses (scored from 0-1). d. delta-GM score represents the user-assigned parameter that describes the minimum difference between GM scores for the highest-scoring match and the second-highest-scoring match for a given MS/MS spectrum.



I then also consulted with Dr. David W. Speicher as well as with Dr. Craig Dufresne, an Orbitrap specialist at Thermo Scientific who presented a training session on the QE+ instrument that I attended. After consulting with them, I came up with a list of parameters to test by running LC-MS runs of roughly equivalent white ghost (WG) samples under all chosen variations, then comparing the number of various categories of peptides (see Section 3.2.4). This resulted in a total of 25 discrete experiments.

After completing these experiments, the analysis of the various parameters for the QE+ instrument revealed several things. The first was that most of these parameters had a predictable impact on QE+ data. One of the most surprising effects was that the results of the peptide match (PM) setting. PM is a parameter that restricts the selection of precursor ions, and their subsequent MS/MS fragmentations, to ideal isotopic envelopes when in the “on” setting (hereafter referred to as PM(on)). An ideal isotopic envelope is typically one that has a smooth abundance distribution centered on any one peak, and whose charge state can be identified by the instrument. PM being in the “preferred” setting (hereafter referred to as PM(pref)) means that if the preset number of MS/MS scans for the N most intense peptides in a full MS run (commonly referred to as TopN) has not been met, it will choose a peak from those which means that at least some of our zero-length CX-MS IDs are coming from those non-ideal envelopes. This could be due to the fact that low abundance cross-links will typically produce low abundance ions, and thus the envelopes’ shape could be affected by small sample sizes.

Another surprising development was that exclusion lists were only conditionally useful for runs where PM (on) peptide runs, and essentially identical to a run not using the exclusion list whe PM (pref) was used. This could be due to the lack of interference

in MS/MS spectra in the intervening region. Another possibility is that the ions being eliminated by the exclusion list were relatively low in abundance. Lastly, the length of MS/MS fill time had a very small impact, which again suggests that zero-length CX-MS precursor ions are low in abundance.

After this optimal method was determined, I then compared the first full MS/MS White Ghost dataset creating using samples from WG-B (see Section 2.2.1) generated during these parameter optimization trials to the one determined to be most optimal according to the parameters described in Table 3-2. This comparison (see Table 3-3) clearly illustrates that significant improvements have been made in the zero-length CX-MS method using these techniques.

Table 3-2: Parameters used in QE+ method optimization trials. Preferred parameters shown in bold.

Parameter	Iterations Tested
LC gradient length (hours)	**2 hours**[a]; 4 hours
m/z peak isolation width (au)	**1.5 au**[a]; 3.0 au
Amount of protein used per injection(µg)	**1 µg**[a]; 3 µg
Repeat analyses	**Duplicates**[b]
MS/MS AGC (ions)	1 x 10^5 ions; **1 x 10^6 ions**[a]
MS/MS duty cycle	Top-10; **Top-20**[b]
MS/MS fill time (ms)	**120 ms**[b]; 240 ms
Excluded charge states (z)	**+1, +2**[a]; +1, +2, +3
Peptide match (PM) setting	PM(on); **PM(preferred)**[a]
Exclusion lists	**PM (on)**[b]; PM (preferred)
Target lists (LC-MS re-analysis)	All IDs; **positive IDs**[a]

[a] This parameter had a major impact on data quality.

[b] This parameter had a minor impact on data quality.

Table 3-3: Method peptide yield comparison for QE+ zero-length CX-MS optimization trials.

Parameter	Pre-Optimization Method	Post-Optimization Method
Total Entries	174	320
High-Confidence IDs[a]	33	65
Unique High Confidence IDs[a]	28	46
High- and Medium-Confidence IDs[b]	46	87
Unique High- and Medium-Confidence IDs[b]	38	63
Base IDs[c]	105	179
Unique Base IDs[c]	84	146

[a] Peptide IDs with a GM score greater than the highest-scoring false positive in their respective LC-MS experiment.

[b] Peptide IDs with a GM score greater than the second-highest-scoring false positive in their respective LC-MS experiment.

[c] Peptide IDs with a GM score greater than 0.2.

3.4 Summary

In summary, a concerted effort was made to adapt the zero-length CX-MS method to a more state-of-the-art QE+ mass spectrometer, as opposed to the LTQ Orbitrap XL instrument that was used in previous studies (Harper, Sriswasdi et al. 2013; Sriswasdi, Harper et al. 2014; Sriswasdi, Harper et al. 2014; Rivera-Santiago, Harper et al. 2015). In order to do so, differences in MS/MS ion fragmentation were addressed via a new version of ZXMiner. Extensive parameter optimization was also undertaken in order to identify the best parameters to use for the QE+ instrument (see Table 3-2). This

resulted in dramatic improvement of the zero-length CX-MS method (see Table 3-3), which in turn suggests that this method is ready to examine more complex protein samples.

Chapter 4: Determination of structure for full-length anion exchanger 1 (AE1) via zero-length CX-MS

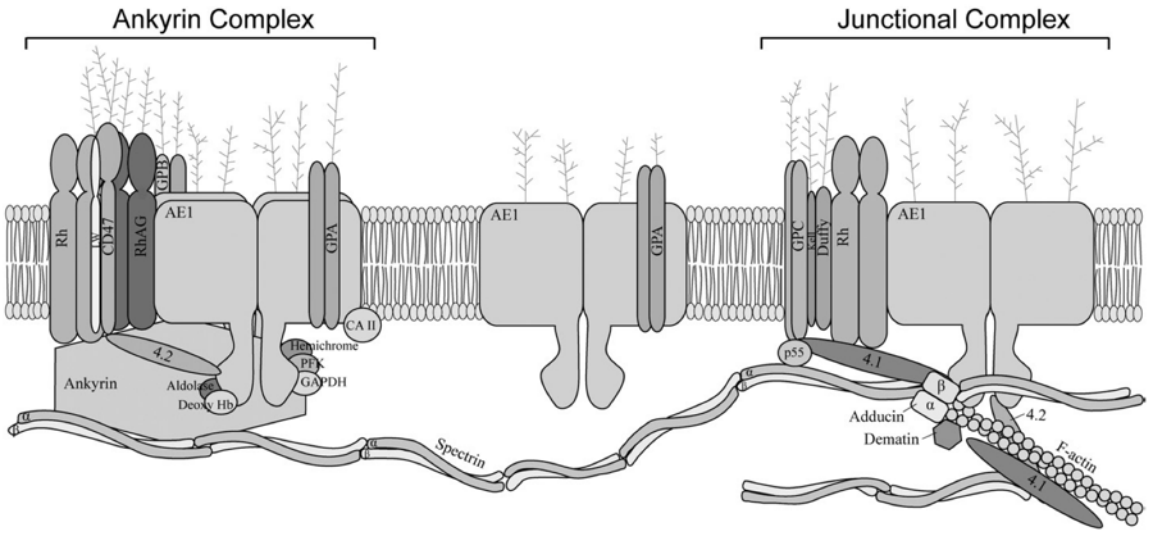
4.1. Background

As mentioned in Chapter 3, one of the major systems I was interested in examining with our optimized zero-length CX-MS technique was the erythrocyte membrane. Specifically, my interests lay in determining the structures of the major protein components of the erythrocyte membrane. These components are critically involved in giving these membranes their trademark elasticity and structural integrity (Fairbanks, Steck et al. 1971). Secondly, because of their high relative abundance, the major structural proteins of the erythrocyte membrane will be likely to produce multiple peptide Identifications. These will in turn provide me with sufficient information to create experimentally validated structural models of key membrane components. This would be novel information, as most previous analyses that have modeled erythrocyte membrane proteins have been composed of speculative models based on very limited structural data, combined with biochemical data from previous studies (Burton and Bruce 2011; Mankelow, Satchwell et al. 2012).

I was particularly interested in the anion exchanger 1 (AE1) protein, in large part because it plays a critical organizing role for all major erythrocyte membrane skeleton complexes (Fairbanks, Steck et al. 1971; van den Akker, Satchwell et al. 2010; Choi 2012). These skeleton complexes, which include the actin junctional complex (Salomao, Zhang et al. 2008; Anong, Franco et al. 2009) and the ankyrin-spectrin complex (Davis, Lux et al. 1989; Grey, Kodippili et al. 2012) (see Fig. 4-1), are unique to the erythrocyte membrane and have been shown to provide it with the flexibility and elasticity that was previously described (van den Akker, Satchwell et al. 2010). Moreover, mutations and

other anomalies in AE1 are also involved in a variety of disease states (Tanner 2002; Stehberger, Shmukler et al. 2007). These include aberrant erythrocyte shapes (Yusoff, Van Rostenberghe et al. 2003; Alper 2009), hemolytic anemia (Eber and Lux 2004; Gallagher 2005; An and Mohandas 2008; Perrotta, Gallagher et al. 2008), and deficiencies or changes in ion transport (Cheung, Cordat et al. 2005; Cheung, Li et al. 2005; Cheung and Reithmeier 2005; Takazaki, Abe et al. 2010; Barneaud-Rocca, Borgese et al. 2011). Finally, because AE1 is the single most abundant protein in the erythrocyte membrane (Fairbanks, Steck et al. 1971), it is very likely to provide the depth of analysis required to generate a structure with a high degree of confidence.

Fig. 4-1: Diagram of the protein complexes in the erythrocyte membrane in relation to AE1. Reproduced with permission from (van den Akker, Satchwell et al. 2010). Ankyrin-spectrin complex (left); actin junctional complex (right).



AE1, also known as band 3, is a 911-residue protein that is primarily expressed in erythrocytes (Fairbanks, Steck et al. 1971) and acid-secretory intercalated kidney cells (Brosius, Alper et al. 1989) and is composed of 2 major domains: a N-terminal region that is involved in recruiting and binding of multiple membrane skeleton protein complexes (residues 1-360), and a C-terminal region involved in anion transport that includes many transmembrane segments (residues 361-911) (Lepke and Passow 1976; Grinstein, Ship et al. 1978; Kopito and Lodish 1985; Choi 2012). Specifically, it neutrally exchanges bicarbonate (HCO_3^-) ions generated by carbonic anhydrase for chloride (Cl^-) ions, thereby performing a critical role in carbon dioxide (CO_2) processing (Passow 1986). It exists as a mix of dimers and tetramers in the erythrocyte membrane, with its oligomeric state being indicative of the protein complexes it is anchoring (van den Akker, Satchwell et al. 2010). Disruption of AE1's ability to form these protein complexes has been implicated in the disease states mentioned above (Tanner 2002).

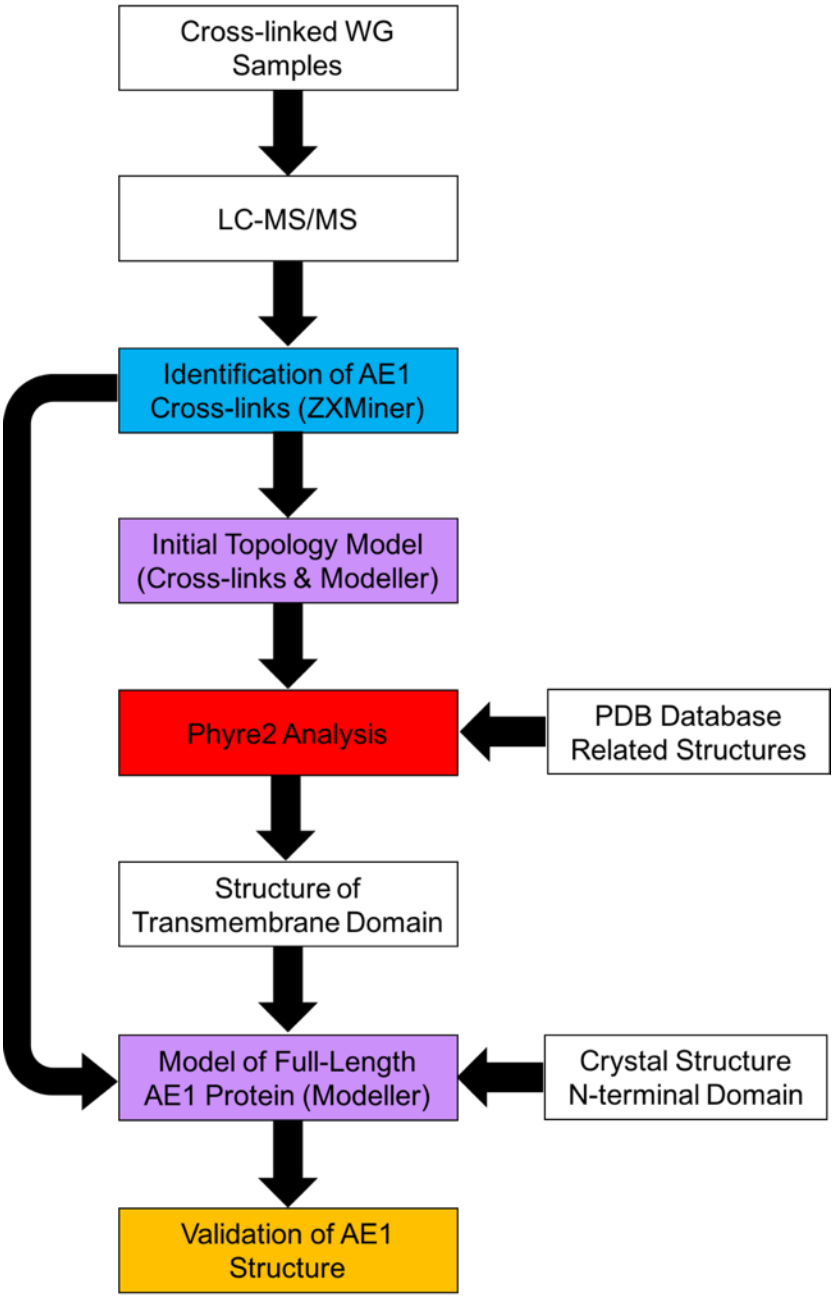
The N-terminal region of AE1, also referred to in the literature as the cytoplasmic domain of band 3, or cdb3, is by far the better-characterized region of the protein. Two crystal structures of this region have been determined, and these describe most of the domain. The only region of cdb3 that is poorly characterized is the extreme N-terminus, which is disordered in solution (Zhang, Kiyatkin et al. 2000; Shnitsar, Li et al. 2013). On the other hand, the structure of the C-terminal region was much less clear when I started this project. A variety of biochemical and structural studies have been conducted on this region (Wang, Kuhlbrandt et al. 1993; Wang, Sarabia et al. 1994; Tang, Fujinaga et al. 1998; Fujinaga, Tang et al. 1999; Popov, Li et al. 1999; Taylor, Zhu et al. 2001; Zhu, Lee et al. 2003; Cheung and Reithmeier 2005; Barneaud-Rocca, Borgese et al. 2011), including a partial structural characterization by cryo-electron microscopy (EM) at a

resolution of 7.5 Å (Yamaguchi, Ikeda et al. 2010). However, the composition of the membrane domain of AE1, including its number of transmembrane spans, was still in contention. Although several studies have theorized that the ion transporter region is composed of 12 transmembrane spans, this is not the consensus in the field. Recently, a study evaluated the AE1 biochemical evidence along with the cryo-EM structure and proposed various topology models (all with 14 helical transmembrane spans) in an effort to consolidate the discussion on the issue (Hirai, Hamasaki et al. 2011). These unresolved controversies combined with the importance of the AE1 protein to make it a very attractive research target.

4.2. Methods

A flowchart of the experimental strategy using chemical cross-linking and molecular modeling to develop a structure of the full-length AE1 protein is summarized in Fig. 4-2.

Fig. 4-2: Flowchart showing the strategy use to generate my full-length AE1 structure. Colors indicate the corresponding programs used in the progression of these experiments, with ZXMiner v. 0.2 (beta) (Sriswasdi, Harper et al. 2014) in blue, MODELLER 9v14 (Sali and Blundell 1993) in purple, and Phyre2 (fold library April 4, 2015) (Kelley, Mezulis et al. 2015) in red.



4.2.1. White Ghost CX-MS

For this study, ZXMiner v. 0.2 beta (Sriswasdi, Harper et al. 2014) analyses were performed on the White Ghost (WG) samples initially used to optimize depth of analysis

on the QE+ mass spectrometer (see Section 3.2.1 for details). Briefly, those experiments involved a variety of cross-linked WG samples run under different instrument parameters (see Table 3-2 for details), which were then evaluated for their magnitude of effect on cross-linked identification. The parameters used for ZXMiner were as described in previous studies (Harper, Sriswasdi et al. 2013; Sriswasdi, Harper et al. 2014; Sriswasdi, Harper et al. 2014; Rivera-Santiago, Harper et al. 2015). Briefly, a precursor mass tolerance of 10ppm, a minimum GM score of 0.2, a minimum ion coverage score of 0.1, and a minimum ΔGM score (which is defined as the minimum difference in GM an identification must have from all other possible identifications in order to be considered a true identification) were imposed.

After the cross-link identifications and corresponding GM scores were generated, the following criteria were applied in order to assign confidence intervals for the identifications:

1. All identifications with a GM score under 0.2 were not considered. This GM score cutoff is in accordance with the minimum base identification score considered in our optimization experiments for the QE+ instrument (see Section 3.2.4).

2. All identifications with a GM score higher than that of the highest-scoring false positive score for the file in which they were generated were considered high-confidence identifications.

3. All identifications with a GM score between the highest-scoring false positive identification and the second-highest-scoring false positive identification were considered medium-confidence identifications.

4. All high- and medium-confidence identifications were analyzed utilizing XlinkInspector version 2.1, and manually reassigned as high- or medium-confidence cross-link identifications based on the following criteria:

A. A high-confidence cross-link identification will have identified at least 50% of the peaks present in its MS/MS spectrum.

B. A high-confidence cross-link identification will have identified the 3 most intense peaks present in its MS/MS spectrum.

C. A high-confidence cross-linked peptide identification will have an unambiguously clear residue assignment of the cross-linked residues. This was defined as an identification that has one cross-linked residue assignment that can be visualized in XlinkInspector that has a GM score higher than all of the alternative possible cross-linked residue assignments for those cross-linked peptides.

5. Once all high- and medium-confidence cross-link identifications were re-evaluated, the cross-links that resulted from these identifications across samples were re-examined. Any cross-link Identifications that only corresponded to medium-confidence identifications were not considered further.

4.2.2. Immunoprecipitation and zero-length CX-MS of AE1

Immunoprecipitation (IP) reactions were designed in collaboration with Mr. Peter Hembach and Ms. Sandra Harper of the Speicher lab. The WG protein consisted of 200 µg of total protein at a concentration of 2 µg/µL in 50 mM Tris, 150 mM NaCl, 1 mM EDTA, and 0.5% SDS at pH = 7.4. This protein had been previously prepared and cross-linked using 10 mM EDC and 5 mM *sulfo*-NHS for 0, 30, 60, and 120-minute intervals at

0 °C. A 10-µg untreated sample (WG Untr) was set aside before the IPs. These WG samples were passed through a Dynabead magnetic bead column with a conjugated polyclonal AE1 antibody (Abcam product number ab78067), then washed with TEA at pH = 11.5. This column was prepared by binding 800mg of antibody to 1.8 mL of Dynabeads Prot. G, then cross-linking and running a blank elution with TEA at pH = 11.5, then neutralizing with 160 mL of 2M Tris at pH = 6.0. 500 µL of Dynabead slurry (at a 1 µg/µL concentration) was used per IP experiment, and the flow-through, wash, and elution fractions were collected by spinning in a JA-20.1 centrifuge at 16,000 rpm for 20 min., then removing the supernatant.

Various elution solvents (glycine + 1% NP-40 at pH 2.5, TEA + 1% NP-40 at pH 11, imidazole + 6M urea + 1% NP-40 at pH 6.0, glycine + NaCl + 6M urea + 1% NP-40 at pH 2.5) were tested in parallel experiments. Each elution was performed in triplicate, then the subsequent elution fractions was then concentrated using a 100kDa ultrafiltration device. The WG Untr, flow-through, wash, and elution fractions were then analyzed via SDS-PAGE using equivalent protein loads, as well as by a Western blot using a monoclonal anti-AE1 antibody (MAb-X-DyBPA) (see Fig 4-3). This Western blot was exposed for 30 seconds, then developed.

Eluent fractions were then analyzed via SDS-PAGE gels run for 2cm to determine what bands would be selected for LC-MS/MS (see Fig. 4-4). After bands were selected, they were analyzed using the most optimal LC-MS protocol determined as part of the parameter optimization trials (see Table 3-2). ZXMiner analyses on these samples used the same parameters as previously described.

4.2.3. Development of AE1 transmembrane model

Phyre2 (Kelley, Mezulis et al. 2015) (fold library April 4, 2015) was used to develop a template structure for the AE1 transmembrane domain, using residues 380-911 from the primary sequence of AE1 as input. This structure was evaluated using the available published biochemical data, as well as the previously generated AE1 topology model (see Section 4.3.3 for details).

4.2.4. Combinatorial modeling

The validated AE1 C-terminal domain template, the AE1 protein sequence, and the 1PRX crystal structure were submitted to MODELLER 9v14 to generate and refine human AE1 a complete structure for full-length AE1. All modeling experiments were run as 50-model trials using the “very slow” refinement algorithm and discrete optimized protein energy (DOPE) score as an output. Homology modeling and refinement were performed simultaneously by including known intra- and inter-subunit cross-links as distance restraints between α-carbons imposed at 11.0 +/- 0.1 Å. Each model was subject to 1,000 iterations and 10 optimization repeats. The completed models were then analyzed according to their DOPE score, and the highest-scoring model under this criterion was chosen for further analysis. Molecular graphics were illustrated using Open-Source PyMOL Version 1.7.x, which also was used to calculate distances between α-carbons of cross-linked glutamate, aspartate, and lysine. No cross-links were observed involving the N-terminal amine group.

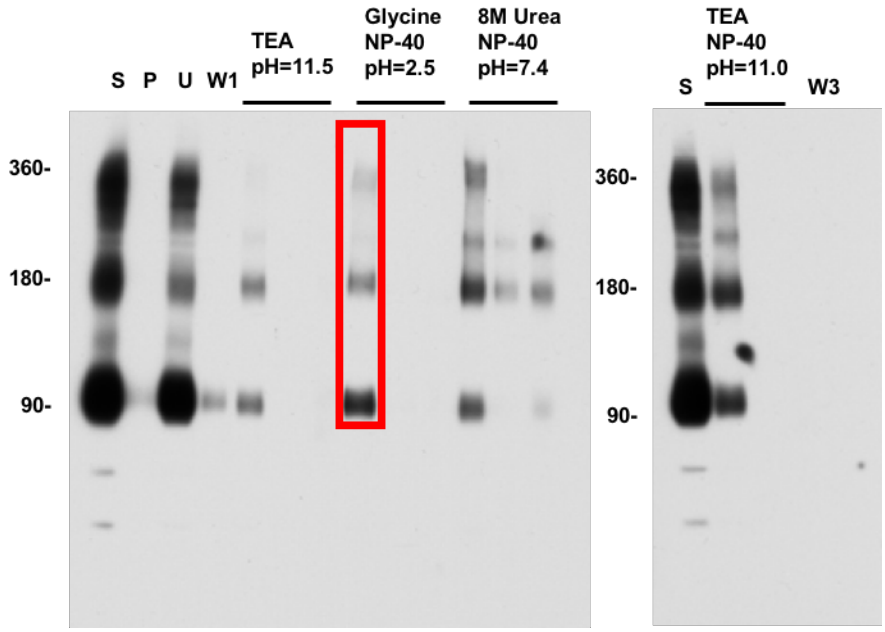
4.3. Results and Discussion

4.3.1. White ghost CX-MS

In order to elucidate additional information about the correlation between identified AE1 cross-links and the protein’s oligomeric, as well as attempt to increase the

depth of analysis for AE1 cross-links in particular, I collaborated with Mr. Peter Hembach and Dr. David Speicher. We designed a series of immunoprecipitation (IP) experiment for AE1 and any proteins cross-linked to it via a magnetic bead column with a conjugated polyclonal AE1 antibody (see Section 4.2.2 for details). The quality of elution would then be detected via Western blots using a monoclonal AE1 antibody (see Fig. 4-3). We settled on a procedure that involved washing the column using TEA, and then eluting with 1% NP-40 and glycine at pH 2.5. The elution fraction was then concentrated using a 100kDa ultrafiltration device. The flow-through, wash, and elution fractions were collected, then analyzed via SDS-PAGE as well as a Western blot using a monoclonal anti-AE1 antibody (see Section 4.2.2 for details).

Fig. 4-3: Immunoprecipitation Western blots for AE1. Elution conditions are described in the relevant lanes. Three alternative elution conditions were testing using replicate immunoprecipitation samples and two sequential elutions were evaluated for each condition to verify the completeness of the elution. S = supernatant for cross-linked WG samples. P = pellet for cross-linked WG samples. U = unbound fraction from WG immunoprecipitation experiments. W1 = initial wash. W3 = Final wash to test column reusability. Red boxed area indicates elution 1 using glycine and 1% NP-40 at pH=2.5 [White Ghost samples were prepared and cross-linked by Sandra Harper. IP experiments and gels were performed by Peter Hembach.].

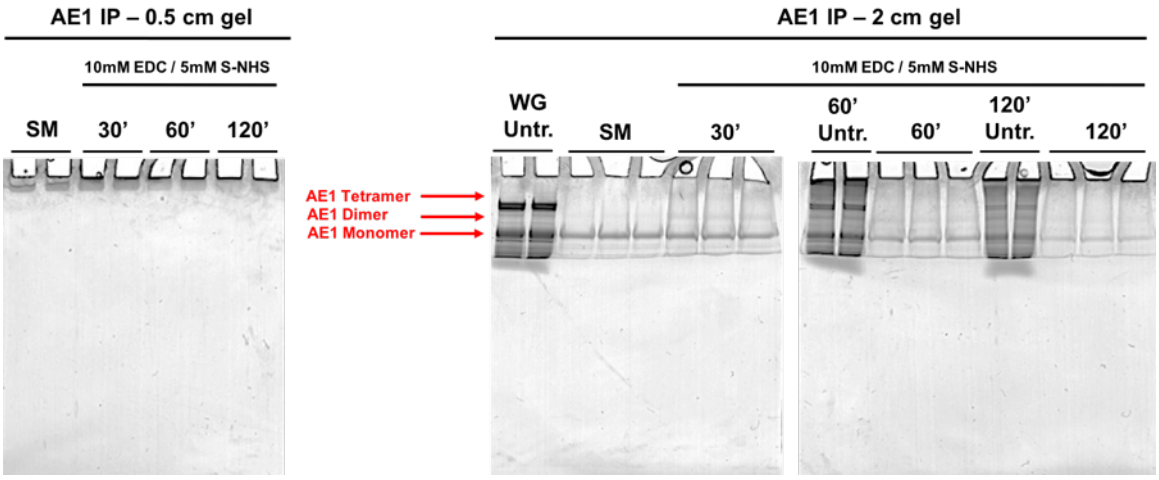


4.3.2. Immunoprecipitation and zero-length CX-MS of AE1

In order to elucidate additional information about the correlation between identified AE1 cross-links and the protein's oligomeric state, as well as attempt to increase the depth of analysis for AE1 cross-links in particular, I collaborated with Ms. Sandra Harper and Mr. Peter Hembach in our laboratory. We designed a series of immunoprecipitation (IP) experiments for AE1 and any proteins cross-linked to it via a magnetic Dynabead column with a conjugated polyclonal AE1 antibody. The quality of elution would then be detected via Western blots using a monoclonal AE1 antibody (see Fig. 4-3). We settled on a procedure that involved washing the column using TEA, and then eluting with 1% NP-40 and glycine at pH 2.5. The elution fraction was then concentrated, then analyzed along with the flow-through and wash fractions via SDS-PAGE as well as by a Western blot using a monoclonal anti-AE1 antibody (see Section 4.2.2 for details). Next, the AE1 samples were run on an SDS-PAGE gel for 2cm, and

the bands that corresponded to AE1 monomers, dimers, and tetramers were cut out and analyzed by LC-MS (see Fig. 4-4). This was a total of 7 experiments.

Fig. 4-4: AE1 Immunoprecipitations used to identify association between cross-linked peptides and different oligomeric states. SM indicates IP of an uncross-linked control white ghost (WG) sample. Lanes labeled WG Untr. are for control samples that were neither cross-linked nor immunoprecipitated. Numbers under cross-linked samples indicate the length of the initial cross-linking reaction. The gel of the left was separated by SDS-PAGE until the tracking dye had migrated 0.5 cm. After trypsin digestion, these samples were used to analyze all components isolated using the IP experiments. The gel on the right was separated 2cm, and regions of the cross-linked samples that corresponded to AE1 monomers, dimers, and tetramers (indicated by red arrows and text) were excised, digested with trypsin, and analyzed by LC-MS/MS.



Upon completing the analysis of the AE1 IP samples, I saw that the majority of AE1 cross-links involving the N-terminal domain and the C-terminal domain were present in all 3 samples corresponding to putative AE1 bands (monomer, dimer, and tetramer), except for cross-link group 5, which was only present in tetramers. A summary

of the peptides found using the combination of the CX-MS of the 25 experiments considered as part of parameter optimizations (see Section 3.3) and digests corresponding to the 7 AE1 IP experiments, and the cross-links found in these 32 total experiments, can be found in Table 4-1. Based on this information and a topology model proposed in a previous study (Hirai, Hamasaki et al. 2011), I was able to develop an AE1 topology model (see Fig. 4-5) that accurately represented an AE1 monomer. With this model as my guide, I could proceed with the modeling of the full-length protein.

Table 4-1: List of Cross-links for AE1 generated using zero-length CX-MS.

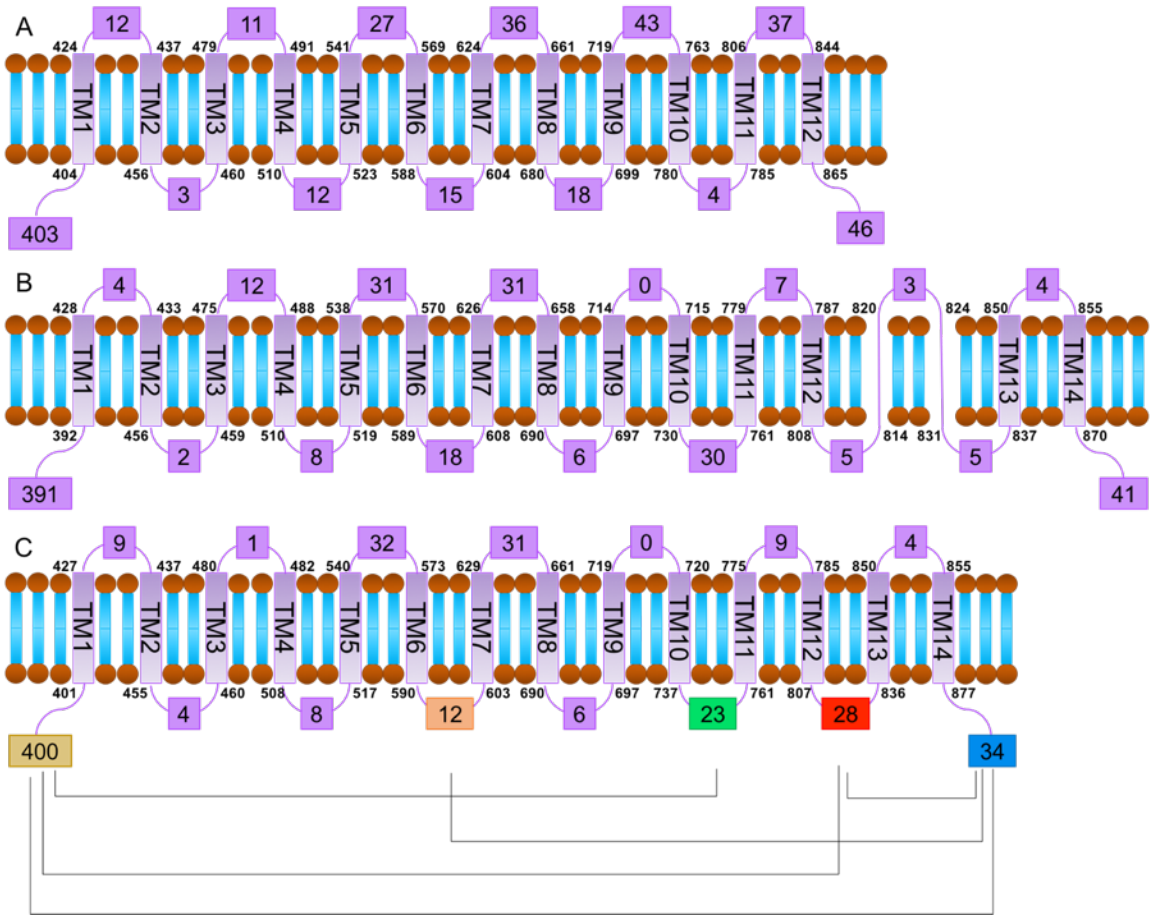
Cross-link Group	Number of Experiments	z[a]	Peptide A Sequence[b]	Peptide A Residues	Peptide B Sequence[b]	Peptide B Residues	Cross-linked Residue A	Cross-linked Residue B
Cross-links Only Involving the AE1 N-terminal Domain								
1	32	3,4,5	VYVELQELVMD[E]K	57-69	HSHAGELELEALGGV[K]PAVLTR	161-180	68	174
2	4	5	WVQL(EE)NLG(E)NGAWGR	81-96	HSHAGELELEALGGV[K]PAVLTR	161-180	85/86/90	174
3	32	4,5	HSHAGELELEALGGV[K]PAVLTR	161-180	LQEAA[E]LEAVELPVPIR	247-263	174	252
4	8	4	LQ(E)AA(E)L(E)AVELPVPIR	247-263	YQSSPA[K]PDSSFYK	347-360	249/252/254	353
5	2	4	FLFDLLGP[E]APHIDYTQLGR	264-283	YQSSPA[K]PDSSFYK	347-360	272	353
6	1	5	I[D]AYMAQSR	296-304	YQSSPA[K]PDSSFYKGLDLNGGPDDPLQQTGQLFGGLVR	347-384	297	353
Cross-links Involving the AE1 N-terminal Domain and Other AE1 Cytoplasmic Segments								
7	8	4	VFT[K]GTVLLDLQETSLAGVANQLLDR	113-138	ATF(DE)EEGR	893-901	116	896/897
8	5	4,5	FIF(ED)QIR	139-146	SVTHANALTVMG[K]ASTPGAAAQIQEVK	731-757	142/143	743
9	25	3,4,5	FIF(ED)QIR	139-146	YHPDVPYV[K]R	818-827	142/143	826
10	12	4,5	ADFL[E]QPVLGFVR	234-246	SVTHANALTVMG[K]ASTPGAAAQIQEVK	731-757	238	743
11	28	3,4,5	ADFL[E]QPVLGFVR	234-246	YHPDVPYV[K]R	818-827	238	826
12	19	4	ADFL[E]QPVLGFVR	234-246	NVELQCLDADDA[K]ATFDEEEGR	880-901	238	892
13	7	4	YQSSPA[K]PDSSFYK	347-360	YHP[D]VPYVK	818-826	353	821
14	32	3,4	YQSSPA[K]PDSSFYK	347-360	ATFDEE[E]GR	893-901	353	899
15	25	3	YQSSPA[K]PDSSFYK	347-	DEY(DE)VAMPV	902-	353	905/906

				360		911		
Cross-links Involving AE1 Cytoplasmic Segments								
16	7	4	F[K]NSSYFPGK	591-600	ATF(DEEE)GR	893-901	592	896/897/898/899
17	30	3,4,5	YHPDVPYV[K]R	818-827	ATFD[E]EEGR	893-901	826	897
18	16	3	YHPDVPYV[K]R	818-827	DEYD[E]VAMPV	902-911	826	906

[a]: Charge states the cross-link group was detected in.

[b] []: cross-linked residue; (): potential cross-linked residue (ambiguous location)

Fig. 4-5: Alternative topology models for AE1 relative to the plasma membrane. A. Topology model for AE1 reported by UniProt (http://www.uniprot.org/uniprot/P02730, UniProt ID P02730). B. Topology model for AE1 as suggested by Hirai, et al., based on cryo-EM data (Yamaguchi, Ikeda et al. 2010) and the assumption that AE1 is the result of a duplication event of a protein that originally had six transmembrane (TM) segments (Hirai, Hamasaki et al. 2011). C. Topology model for AE1 based on our zero-length CX-MS data. The number of amino acids in loops and segments outside the lipid bilayer are indicated by colored boxes, and the N-terminal domain is on the far left side of each model.



Another important note is that surprisingly this IP analysis did not result in significant additional depth of analysis for the AE1 protein. This could be due to the fact that a maximum depth of analysis had been achieved as part of the optimization trials described in Chapter 3; that is, all cross-linked peptides in the samples had been identified. A more likely reason is that ZXMiner was designed to compare closely matched control and experimental samples using a label-free correlation of LC-MS signals. However, dimer and tetramer regions of IPs from cross-linked samples will not completely match the monomer band of AE1 from IPs of control samples. These inevitable differences in protein content have apparently impeded the ZXMiner

software's ability to identify cross-linked peptides in these simplified samples significantly.

4.3.3. Development of AE1 transmembrane model

I then used Phyre2 to generate a model of the AE1 C-terminal domain described in 4.3.2, using the primary protein sequence as my input. Briefly, Phyre2 generated a model which shares great protein fold similarity with the uracil transporter UraA (Lu, Li et al. 2011), which is a protein that potentially shares some evolutionary characteristics with AE1 despite its poor sequence homology (see Fig. 4-7). Specifically, it is believed that both UraA and AE1 are the result of an internal duplication event, which is why their membrane-bound domain architecture seemingly consists of 2 very similar sets of 7 helices (see Fig. 4-6) (Hirai, Hamasaki et al. 2011; Lu, Li et al. 2011; Vastermark and Saier 2014).

Fig. 4-6: Phyre2-generated model for the AE1 C-terminal domain. The PDB database used to generate this model is described in Fig. 4-7. From left to right: cytoplasmic view, membrane view, extracellular view.

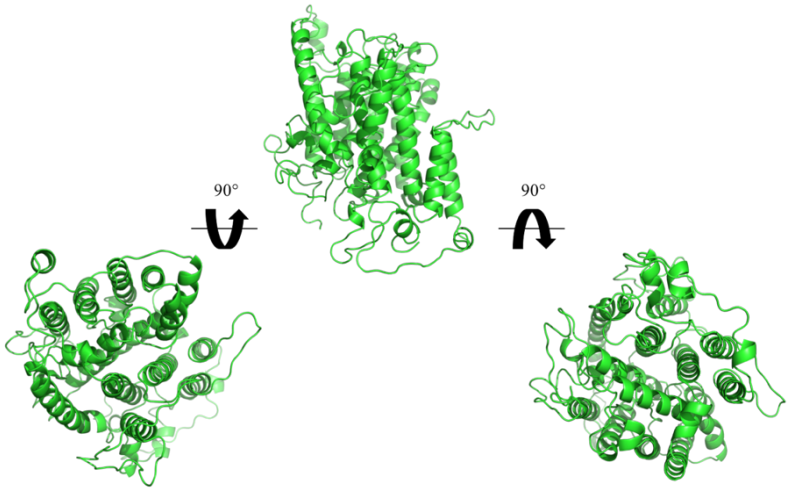
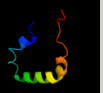
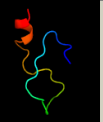
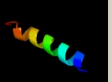
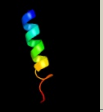
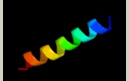


Fig. 4-7: The ten most relevant models from the PDB database used by Phyre2 to create the AE1 transmembrane domain template structure. Confidence and percent sequence identity (%i.d.) scores are color-coded from grey (least) to red (most). Alignment indicates what section of the transmembrane segment was modeled using that structure.

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c3qe7A_	Alignment		100.0	15	**PDB header:**transport protein **Chain:** A: **PDB Molecule:**uracil permease; **PDBTitle:** crystal structure of uracil transporter--uraa
2	c1bzkA_	Alignment		99.4	100	**PDB header:**transport protein **Chain:** A: **PDB Molecule:**protein (band 3 anion transport protein); **PDBTitle:** structural studies on the effects of the deletion in the2 red cell anion exchanger (band3, ae1) associated with3 south east asian ovalocytosis.
3	c1bh7A_	Alignment		96.6	97	**PDB header:**membrane protein **Chain:** A: **PDB Molecule:**band 3; **PDBTitle:** a low energy structure for the final cytoplasmic loop of2 band 3, nmr, minimized average structure
4	c1btrA_	Alignment		90.8	100	**PDB header:**anion transport **Chain:** A: **PDB Molecule:**band 3 anion transport protein; **PDBTitle:** the solution structures of the first and second2 transmembrane-spanning segments of band 3
5	c1btqA_	Alignment		90.8	100	**PDB header:**anion transport **Chain:** A: **PDB Molecule:**band 3 anion transport protein; **PDBTitle:** the solution structures of the first and second2 transmembrane-spanning segments of band 3
6	c1bttA_	Alignment		70.3	100	**PDB header:**transmembrane protein **Chain:** A: **PDB Molecule:**band 3 anion transport protein; **PDBTitle:** the solution structures of the first and second2 transmembrane-spanning segments of band 3
7	c1btsA_	Alignment		68.0	100	**PDB header:**transmembrane protein **Chain:** A: **PDB Molecule:**band 3 anion transport protein; **PDBTitle:** the solution structures of the first and second2 transmembrane-spanning segments of band 3
8	c4us3A_	Alignment		40.7	12	**PDB header:**transport protein **Chain:** A: **PDB Molecule:**transporter; **PDBTitle:** crystal structure of the bacterial nss member mhst in an2 occluded inward-facing state
9	c3b9yA_	Alignment		37.0	13	**PDB header:**transport protein **Chain:** A: **PDB Molecule:**ammonium transporter family rh-like protein; **PDBTitle:** crystal structure of the nitrosomonas europaea rh protein
10	c4m48A_	Alignment		20.6	10	**PDB header:**transport protein **Chain:** A: **PDB Molecule:**transporter; **PDBTitle:** x-ray structure of dopamine transporter elucidates antidepressant2 mechanism

I then evaluated the AE1 C-terminal model using the available published biochemical data, as well as the previously generated AE1 topology model (see Fig 4-5). The topology model was particularly critical, as it guided me in terms of the number of transmembrane spans as well as their approximate locations. The transmembrane spans in the C-terminal model were counted, and the locations of the sites for important biochemical reactions we examined. These include sites such as Lys-542 and Lys-851, which react with the H_2DIDS ion transport inhibitor (Okubo, Kang et al. 1994), the complex oligosaccharide glycosylation site at Asn-642, and the palmitoylation site at Cys-843 (Mitra, Ubarretxena-Belandia et al. 2004; Li, Takazaki et al. 2006). Other key residues used in this evaluation included those involved in the mutation causing Southeast Asian ovalocytosis (residues 400-408), the residues used in previous studies involving Cys mutagenesis scanning (Gly-742 and Ser-745), and the Lyn kinase phosphorylation sites (His-359 and His-904).

After performing all of these analyses, I found that all of these parameters were in line with expectations. Therefore, I deemed the structure to be a valid template for combinatorial modeling experiments (see Section 4.2.4 for details). Briefly, these experiments took the AE1 C-terminal transmembrane model template and the existing crystal structures of the N-terminal domain, and used the cross-links determined in the CX-MS analyses (see Table 4-1) to constrain the location of the cytoplasmic loops between transmembrane spans. These cross-links also helped determine the position of the N-terminal domain relative to the C-terminal domain, as will be discussed below.

4.3.4. Validation and Analysis of Full-Length AE1 Structure

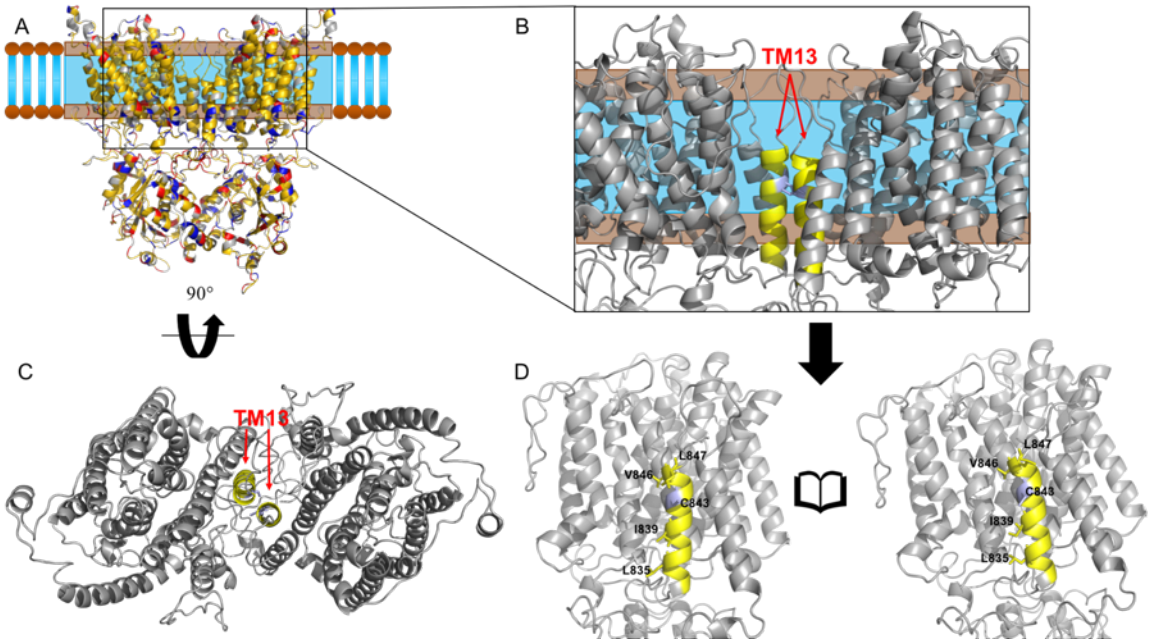
After the final structure for full-length AE1 was determined, I validated it in several ways. Firstly, I examined the distances between the cross-linked residues as

described in Table 4-1 to observe the imposed distance constraints for observed cross-linked sites. These sites should result in α-carbon distances being within 12Å for cross-links in well-ordered regions of the protein, or up to 16Å for disordered ones. I found that all of the cross-links examined were within the expected distances.

I then interrogated the nature of the amino acid residues composing the pore by color-coding the residues according to their hydrostatic character and measured the pore-spanning segment of the protein. This segment was consisted with the expected thickness of the membrane (Mitra, Ubarretxena-Belandia et al. 2004), as shown in Fig. 4-8a. This showed that the transmembrane spans were mostly composed of hydrophobic residues, as expected. I also examined the contact interface for the AE1 dimer within the lipid bilayer. This showed that the AE1 C-terminal domain contributes to the dimerization interface in addi(Eber and Lux 2004)tion to the previously known homodimeric interaction of the cytoplasmic N-terminal domain. Furthermore, it showed that its contribution primarily involves the transmembrane segment T13 (see Fig. 4-8b-d).

Fig. 4-8. Full-length AE1 dimer hydrophobicity and transmembrane domain dimeric interface analysis. A. Full-length AE1 dimer (membrane view). Hydrophobic residues (Ala, Gly, Ile, Leu, Met, Phe, Pro, Val) are in gold, hydrophilic residues (Asn, Cys, Gln, Ser, Thr, Trp, Tyr) are in grey, positively charged residues (Arg, His, Lys) are in blue, and negatively charged residues (Asp, Glu) are in red. B. AE1 transmembrane domain dimeric interface (membrane view). TM13 (yellow) and the palmitoylated Cys-843 (violet) form the entire membrane-embedded dimer interface. C. AE1 transmembrane domain dimeric interface (extracellular view). D. AE1 transmembrane domain dimeric interface (book view). Key interacting residues are shown in sticks are

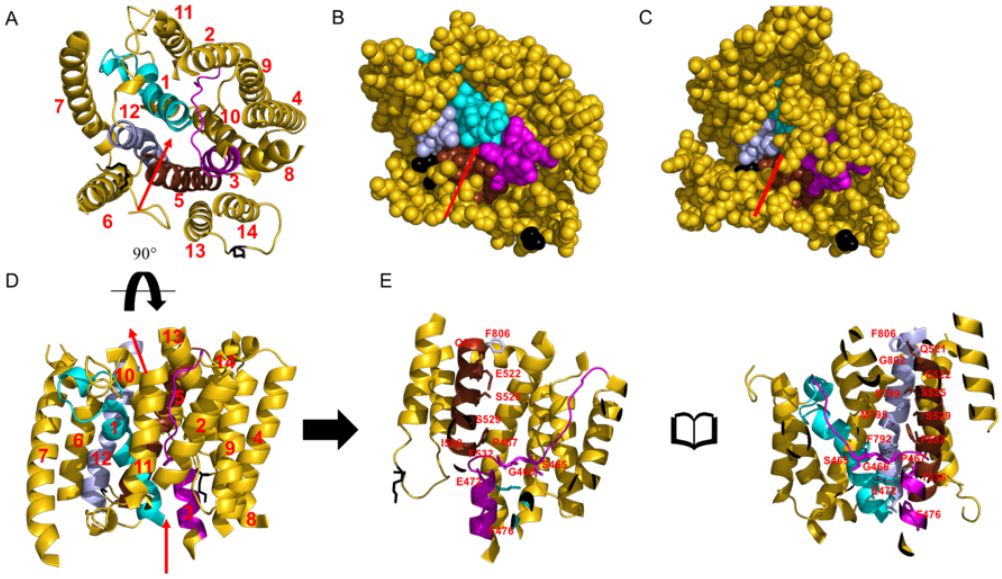
labeled.



Next, I examined the protein channel and determined the arrangement of the helices that comprise it (see Fig. 4-9a). From this analysis, I was able to determine that transmembrane helices T1, T3, T5, and T12 in AE1 were playing a critical role in the formation of the channel (see Fig. 4-9b). Furthermore, display of the H_2DIDS site shows that it is present on the extracellular side as previously determined, and importantly, is near the helices that compose the channel. This is in accordance with previously performed mutagenesis studies that determined that mutation of Lys-542 and Lys-851 does not impede ion transport (Wood, Muller et al. 1992). These residues were previously thought to be loosely associated with the conformational changes the channel undergoes as part of ion transport (Okubo, Kang et al. 1994), but not physically blocking the channel opening. This is in contrast to some of the cytoplasmic loops on the extracellular face of the protein, which seem to be interfering with access to the channel entrance. This can be visualized by an examination of Fig. 4-9c (which has the loops

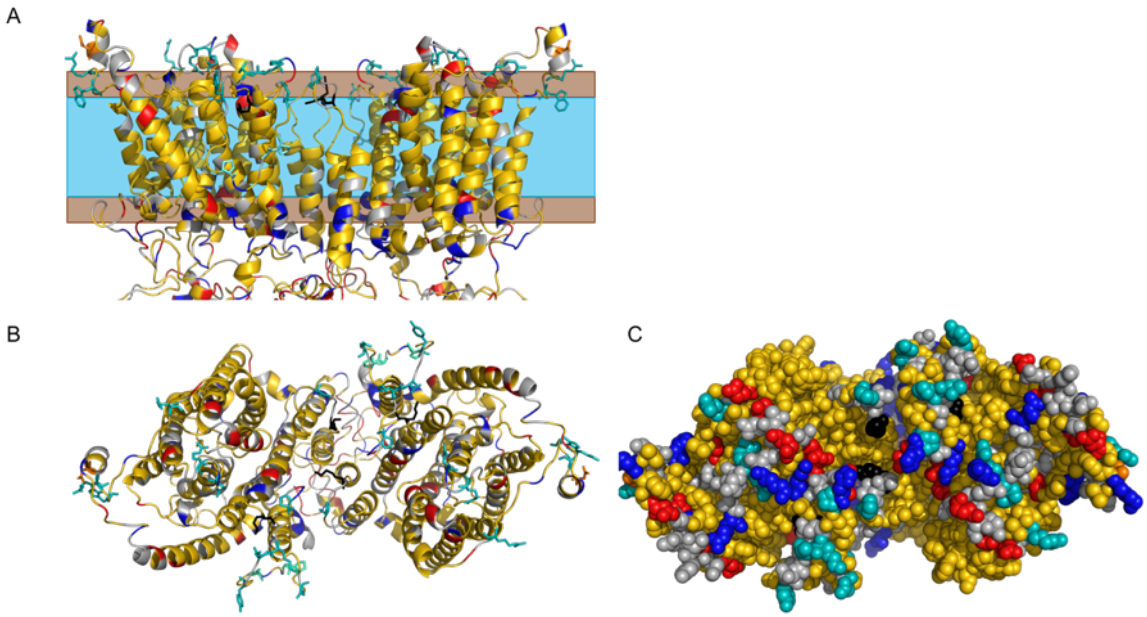
trimmed for clarity) and Fig. 4-9d (which displays the loops fully). Furthermore, examining the inside of the channel (see. Fig. 4-9e and f) allowed me to estimate individual residues that appear to contribute to the channel. Some of these have already been described in the literature, most notably S465, which has been shown to abolish transport when mutated to a residue with a large side chain, such as Ile, Asn, or Asp (Li, Quilty et al. 2000; Barneaud-Rocca, Etchebest et al. 2013; Bonar, Schneider et al. 2013).

Fig. 4-9: The AE1 channel. A. The AE1 channel (extracellular cartoon view). Transmembrane segments are numbered in red. Transmembrane segments critical to the channel are colored in teal (TM1), magenta (TM3), brown (TM5), and violet (TM12). H_2DIDS binding site residues are in black sticks. B. Simplified AE1 channel (extracellular sphere view) without extracellular loops in front of AE1 channel site. C. AE1 channel with full extracellular loops (extracellular sphere view). D. AE1 channel (membrane cartoon view). E. AE1 channel (membrane book view). Potential key residues for ion transport are labeled in red.



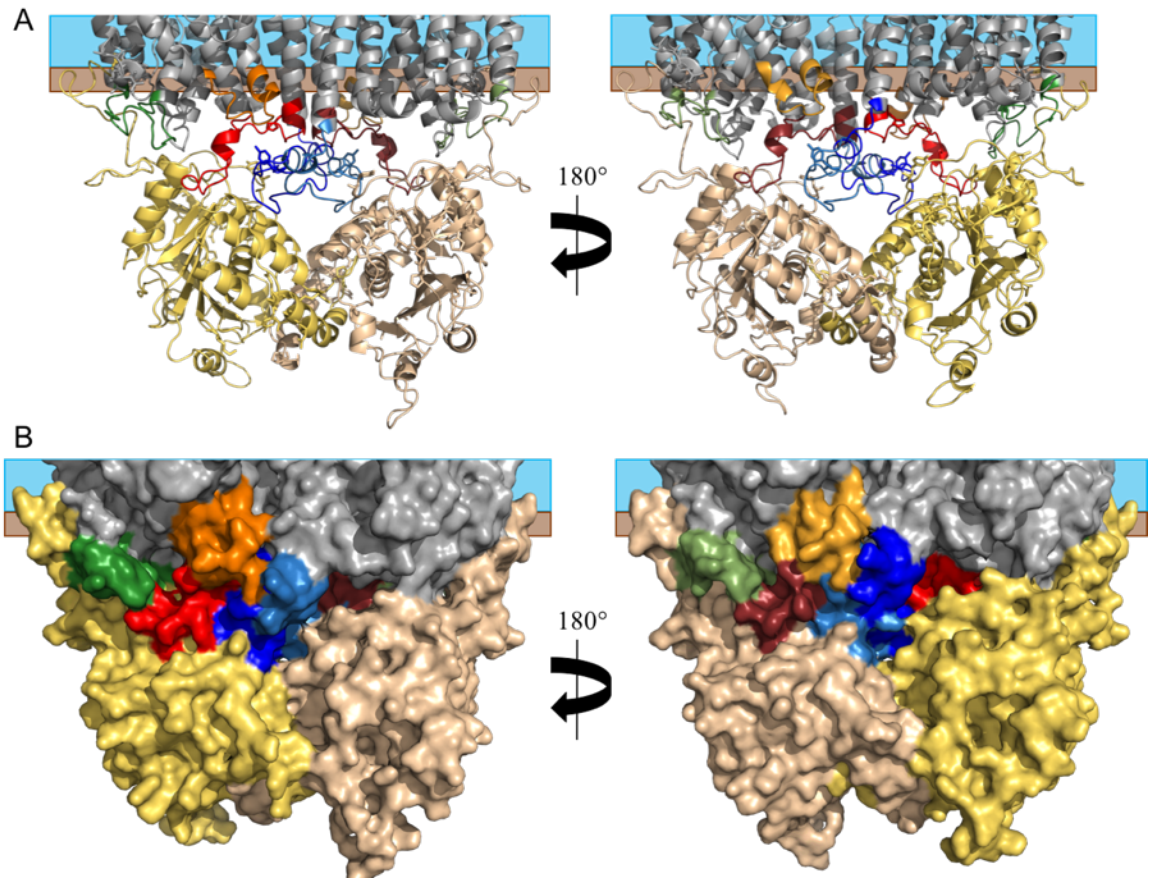
After the channel was examined, I moved on to the extracellular surface of the AE1 protein. Here, I examined the locations of the complex oligosaccharide site at Asn-642 (Li, Quilty et al. 2000; Shnitsar, Li et al. 2013), as well as the sites for previously reported blood antigens on AE1 (Poole 2000) (see Fig. 4-10). I found that the most of these sites were correctly localized on the extracellular face beyond the projected location of the membrane. The few sites that were not located on the extracelluar face were loops with no structural constraints that the modeling software could not unambiguously model. This is due to the fact that these experiments were not conducted by imposing a hydrophobic membrane restraint on residues located within the lipid bilayer, because this is beyond MODELLER's capabilities.

Fig. 4-10: Analysis of AE1 extracellular region. A. AE1 extracellular sites (membrane cartoon view). Hydrophobic residues (Ala, Gly, Ile, Leu, Met, Phe, Pro, Val) are in gold, hydrophilic residues (Asn, Cys, Gln, Ser, Thr, Trp, Tyr) are in grey, positively charged residues (Arg, His, Lys) are in blue, negatively charged residues (Asp, Glu) are in red, blood antigen sites are in teal sticks, H_2DIDS binding site residues are in black sticks, and complex oligosaccharide site (Asn-642) is in orange sticks. B. AE1 extracellular sites (extracellular cartoon view). C. AE1 extracellular sites (extracellular sphere view).



The next step was to examine the cytoplasmic face of AE1 in order to comprehensively evaluate the 3D arrangement of the N-terminal domain in relation to the channel, as well as to the other cytoplasmic loops. In order to do so, I color-coded the C-terminal domain's cytoplasmic loops using the color scheme as shown in the topology model (see Fig. 4-5). I then analyzed the 3D arrangement of these multiple cytoplasmic segments (see Fig. 4-11), and I saw extensive interaction between the C-terminal domain's cytoplasmic loops and the N-terminal domain. Of particular note is that residues 806-835 (in red) are clearly shown as cytoplasmic in localization. This is interesting, as prior publications reached conflicting conclusions regarding whether this loop is cytoplasmic (Erickson 1997; Popov, Li et al. 1999) or extracellular (Popov, Tam et al. 1997), and it has been suggested that its conformation can change in response to stimuli (Jin, Abe et al. 2003). The extreme C-terminal portion of AE1 (residues 877-911, in blue) also seems to be contributing to the dimerization interface, which was not previously known. These insights combine to indicate that the homodimer interface is much more extensive than previously known.

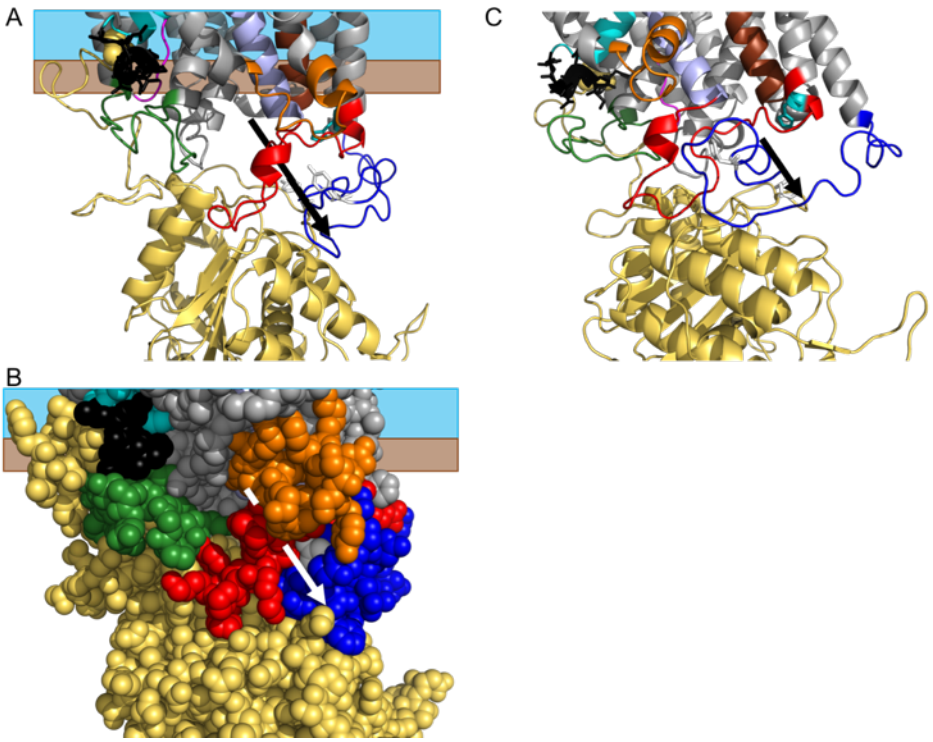
Fig. 4-11: Cytoplasm-membrane interface of an AE1 dimer. A. The N-terminal domain (yellow and wheat) and cytoplasmic loops (membrane cartoon view). Cytoplasmic loops are colored using the color scheme previously shown in Fig. 4-5. B. Surfaces views of the structure shown in A.



Next, I decided to examine the channel's exit location in relation to the N-terminus, as well as examine several important reaction sites on AE1. These include the Lyn kinase phosphorylation sites on Tyr-359 and Tyr-904 (Brunati, Bordin et al. 2000), as well as the DEPC inhibitor site on His-834 (Jin, Abe et al. 2003), and the location of the Southeast Asian ovalocytosis deletion mutant (residues 400-408) (Barneaud-Rocca, Borgese et al. 2011). This analysis showed that AE1's channel exit site seems to be

angled away from where the N-domain region lies, and is instead pointed toward the dimeric interface (see Fig. 4-12). This is consistent with a previous study that examined the N-terminal domain of AE1 to determine whether it had a substrate access tunnel to pass ions through into the cytoplasm, and did not find one (Shnitsar, Li et al. 2013). We also observe that both the phosphorylation sites and the DEPC binding site in our model are near to the proposed channel exit, which is consistent with their roles in ion transport inhibition (Jin, Abe et al. 2003). The Southeast Asian ovaloctyosis deletion mutant site is on opposite face of the molecule from the proposed channel exit site, which superficially could suggest that there is no association between this region and ion transport. However, our model shows that these residues that are probably quite important for protein folding, because they are part of the membrane-entry portion of the TM1 transmembrane segment. Thus, the deleterious effects observed in this deletion mutant (Tanner 2002; Yusoff, Van Rostenberghe et al. 2003; Barneaud-Rocca, Borgese et al. 2011) could be due to channel misfolding.

Fig. 4-12: Arrangement of AE1 cytoplasmic loops and binding sites relative to the ion channel exit site. A. AE1 cytoplasmic loops and proposed channel exit site (membrane cartoon view). Cytoplasmic loops are as shown in Fig. 4-5, and AE1 channel helices are colored as shown in Fig. 4-8. Furthermore, Lyn kinase phosphorylation sites (Tyr-359 and Tyr-904) are shown in white sticks, the DEPC inhibitor binding site is in teal sticks, and the Southeast Asian ovalocytosis mutation site (400-408) is shown in black sticks. The proposed channel exit site is shown as a black arrow. B. A sphere view of the image in panel A, except the proposed channel exit site is highlighted with a white arrow for clarity. C. AE1 cytoplasmic loops and proposed channel exit site (tilted cartoon view).



4.4. Conclusions

In summary, I was able to take zero-length CX-MS data derived from an optimized method for the QE+ instrument on a complex WG sample and generate cross-link identifications for the AE1 protein. I then took these cross-links and combined them with previously reported crystal structures and biochemical data in order to form an experimentally validated model. While this model has some limitations due to the nature of the cross-link localization and the nature of the modeling experiments themselves, I believe it provides a reasonable picture of the structure and properties of this important erythrocyte membrane protein.

However, while preparing this thesis document, I became aware of an accepted publication involving a crystal structure of AE1's C-terminal domain [Arakawa, T., et al.; Structure of the anion exchanger domain of human erythrocyte Band 3; manuscript

accepted], that was accepted in late September 2015. Based on a review of a copy of the accepted manuscript, the crystal structure is generally similar to my AE1 C-terminal domain model. However, it also contains some differences in important details compared to that model. I unfortunately have not yet been able to access the PDB data for this structure in order to perform a more detailed comparison between the two structures.

However, based on review of the manuscript, I can make several preliminary observations. First among these is that both structures have 14 transmembrane segments in their respective C-terminal domains. Secondly, both structures used UraA as a template for the C-terminal domain architecture. Some of the major differences center around the architecture of the channel – the crystal structure's channel is composed of TM1-TM4 and TM8-TM11, whereas my model's channel identified the key transmembrane segments to be TM1, TM3, TM5, and TM12. On the other hand, several cytoplasmic loops that I have identified cross-links for (most notably 742-753) are poorly resolved in the crystal structure, and the crystal structure does not correlate the C-terminal domain with the N-terminal domain.

When the PDB file is publically released, I plan to compare my structure to the crystal structure in precise detail. I will also perform new modeling experiments using the crystal structure as a template to obtain a structure of the entire AE1 protein. I expect this final structure to be highly accurate, given that I will have multiple crystal structure templates coupled with a good density of zero-length cross-link identifications between the N-terminal domain and the cytoplasmic segments of the C-terminal domain. This includes several segments that do not seem to be present in the crystal structure of the C-terminal domain. The combined structure will yield the most definitive insights on full-length AE1's structure and function to date.

Chapter 5: Future directions

5.1. Further improvements in the density of identification zero-length cross-links

5.1.1. CX-MS analysis method improvements

One of the most critical future directions is further development of the zero-length CX-MS technique. Maximizing extent of cross-linking combined with improved depth of analysis results should both improve the confidence of assigned cross-link site identifications and increase the density of the resulting distance constraints to improve the accuracy of molecular modeling. This can be approached in several ways, which I will describe in detail below.

One way to optimize peptide identification lies with improving the software's detection methodology. In addition to the collaborative methodology used to account for the presence of the precursor ion in QE+ MS/MS spectra described in Chapter 3, I have also discussed some problems with in the peak deisotoping algorithm with Dr. Sriswasdi, which results in peaks being assigned a mass that is ±1 Da (nominally) off from their actual mass, thus bringing down the GM scores for these IDs. Dr. Sriswasdi is currently working on software improvements in order to address this problem.

Perhaps the most productive avenue of potential optimization is to further adapt the protocol to higher-throughput instruments. The Speicher lab has recently acquired a Thermo Q Exactive HF instrument (referred to hereafter as QE-HF), which offers increased speed of MS/MS data acquisition when compared to the QE+, and is otherwise quite similar. Testing and analysis of zero-length CX-MS data on the QE-HF should be further explored.

Finally, CX-MS sample preparation can be optimized. While options with regards to traditional sample purification are limited (as described in Chapter 1), the immunoprecipitation (IP) experiments that were described in Chapter 4.2.2 could be employed in the case where an analyte of interest has been previously identified. This could help reduce sample complexity and thus lead to further depth of analysis. One precaution that has to be taken when using this approach is that ZXMiner's (Sriswasdi, Harper et al. 2014) label-free comparison module was designed with the assumption that the uncross-linked control sample was identical to the experimental cross-linked sample with the exception of the introduction of a moderate number of cross-links, but this will not be the case with IP samples. This is most likely the reason why the IP samples for AE1 did not significantly increase depth of analysis, as was discussed in Section 4.3.2. Because of this limitation, it is possible that the improvements made using this approach will be more modest than the other two.

5.1.2. Potential side reactions and their potential interference in cross-linked peptide identification

A potential issue that may need to be addressed in the preparation of zero-length CX-MS samples is the possibility of side reactions. Although the original paper that described the EDC/NHS cross-linking protocol did not report any side reactions (Grabarek and Gergely 1990), some studies have reported seeing EDC adducts at high cross-linker concentrations (Bruce 2012). In order to test for side reactions, I have performed intact LC-MS analysis of intact proteins that were cross-linked using EDC/NHS chemistry on a LTQ Orbitrap-XL with a trap-column-only UPLC setup in 30-minute time intervals. These experiments detected multiple mass modifications of unknown origins, as they did not correspond to the expected mass shifts for the reported

cross-linking intermediates for the EDC carbodiimide and the NHS-ester intermediate. These unknown intermediates were also detected by surveying modifications to linear peptides in MaxQuant using the dependent peptide function (Cox and Mann 2008). This function uses the previously identified peptides found in a database search as the forward database for a second dataset search, which only involves unidentified peptides. This second database search looks for mass differences between identified peptides and unidentified peptides with similar MS/MS peak patterns, and scores the hits based on the similarity of their MS/MS spectra. These peptide hits and corresponding mass differences are then reported, and matched to a mass shift database whenever possible. In the case of the EDC/NHS mass modifications, the precise chemical composition of these intermediates remains unknown. Further studies to characterize the nature of potential intermediates and their approximate yields relative to cross-linked peptides are needed in order to incorporate these side reaction products into the data analysis pipeline.

5.1.3. Alternative cross-linkers

Another potential direction this research project could explore is the use of alternative cross-linkers. In addition to testing one of the more established cross-linkers described in Chapter 1, I am also interested in testing a recently published acidic-residue-specific cross-linker from the Aebersold lab called DMTMM (Leitner, Joachimiak et al. 2014). This cross-linker has been reported to produce zero-length cross-links as well as cross-links with spacer arms of various lengths, so the ensuing cross-linking chemistry has to be further tested and clarified. If a consistent reaction product can be obtained with this reaction, it could be used as a complement to EDC/NHS cross-linking.

It could even be used as a potential replacement if the zero-length cross-linking activity can be isolated and is superior in terms of depth of analysis.

5.1.4. Incorporation of tandem mass tags (TMT)

Another potential alternative approach for the identification of zero-length CX-MS products is to employ tandem mass tags (TMT) in order to aid in their identification. These tags could be applied to cross-linked samples during sample preparation in order to facilitate sample preparation, as well as potentially reduce MS time by requiring less samples in order to achieve the requisite depth of analysis to successfully ID enough peptides to inform modeling experiments. The reason why such a method would require less samples per LC-MS analysis is because a set of TMT tags would allow a number of samples equal to the number of tags in the set (for example, 6 separate isobaric tags in a TMT 6-plex) to be pooled into a single MS run. This single MS run would also not need an uncross-linked control sample, as our current zero-length CX-MS method does. This lack of a control sample could prove useful in terms of reducing the amount of protein required to perform an LC-MS analysis, as well as eliminate some of the previously reported limitations for IP experiments.

5.2. Improvements to modeling method using zero-length CX-MS distance constraints

5.2.1. Alternative modeling approaches

Yet another option to potentially improve the CX-MS method is to reconsider the modeling software used to apply the CX-MS restraints. As I alluded to earlier, MODELLER (Sali and Blundell 1993) is a homology modeling program, and thus needs template structures in order to provide 3D representations of a given system. In order to ensure maximum fidelity and accuracy to the original structure, I strongly prefer to use

structural information corresponding to the protein itself. Alternatively, I could also use a protein that has demonstrated sequence or structural homology in order to make a template, as I did for AE1 (see Chapter 4.2.3). However, structural information of this kind is not always available. I have also found that MODELLER is somewhat limited when faced with regions that are poorly constrained. Two examples of poorly constrained regions are cytoplasmic loops with no identified cross-links, and large protein complexes whose binding interface between components is relatively small. In the event that the cross-linking data or the available template structures are not sufficient to ensure a high degree of confidence from a MODELLER run, a program like the Rosetta software suite (Simons, Bonneau et al. 1999) could be used to try and accomplish these tasks.

5.2.2. Importance of cross-link density in homology modeling experiments

One of the reasons why I am considering alternative modeling approaches in the previous section is because my CX-MS supported homology modeling experiments, particularly the AE1 project discussed in Chapter 4, have enhanced my awareness of the technique's inherent limitations. These limitations are mainly that either a certain degree of cross-link density is required to have a high degree of confidence in the modeling of a given region, or a strongly defined homologous template structure is needed for that region. Without either of these constraints, some regions in modeling experiments are only subject to the modeling software's energy minimization function, which does not necessarily reflect the biologically relevant structure of the protein or protein complex.

The above limitations became readily apparent during the AE1 project discussed in Chapter 4. As was mentioned in Section 4.4, I was recently provided an advance copy

of a manuscript accepted in Science [Arakawa, T., et al.; Structure of the anion exchanger domain of human erythrocyte Band 3; manuscript accepted], which described a crystal structure of the AE1 C-terminal domain. While I do not have access to the PDB file, I was able to make some preliminary comparisons between the structure of the C-terminal domain in my full-length AE1 model and the crystal structure.

These comparisons have shown me that some aspects of my analysis were correct. These include the number of transmembrane spans in the AE1 C-terminal domain, the general residue ranges and hydrophobic nature of these spans, and the structural homology between AE1 and UraA despite their poor sequence similarity. However, other aspects differed significantly from the crystal structure. Foremost among these differences is the nature of the helices that compose the channel region and its critical residues. Unfortunately, because of the soluble nature of the EDC cross-linking reagent, I was not able to obtain any cross-links within the lipid bilayer. Similarly, I did not detect any cross-links between extracellular loops, which could be due to the fact that the EDC cross-linker is not membrane-permeable because of its anionic charge. This lack of distance constraints did not allow me to fully refine the transmembrane segments in order to come up with an accurate structure. This shows that further modeling attempts using the current protocol should have either better high-confidence template structures or have a some cross-linked peptides across all major areas of the protein, in order to ensure maximum accuracy. When the PDB file of the crystal structure for the C-terminal domain of AE1 becomes available, I will incorporate it into my full-length AE1 model. I will revise my structure of the full-length AE1 protein and any functional interpretations made accordingly.

5.3. Further CX-MS analysis of human erythrocyte membranes

Another future direction for the zero-length CX-MS project is to both further mine existing datasets of cross-linked White Ghosts (WG) and conduct new cross-linking experiments. Although Chapter 4 of this document focused extensively on AE1, there are several other major protein components in the erythrocyte membrane that were represented in the cross-linked peptide datasets. These included large oligomeric proteins such as ankyrin and spectrin, as well as actin, and tropomyosin, protein 4.1, protein 4.2, glycophorin-A, α-adducin, and β-adducin. If optimum depth of analysis for CX-MS on WG samples is achieved, there could be enough cross-link identifications to begin building structures of the individual components that are poorly characterized. Alternatively, a cross-link interface between two or more unique components of the erythrocyte membrane that already have partial or full structural information available could be pursued. This would lead to the generation of an experimentally validated model of one of the membrane's major protein complexes, such as the ankyrin-spectrin complex with AE1 or the actin junctional complex (see Fig. 4-1). I believe that both approaches have merit, given that I detected and identified cross-links for many of these proteins as part of the studies involving AE1. Unfortunately, I was not able to gather enough information to comprehensively elucidate new structures of individual proteins or protein complexes. However, I believe that the water-soluble nature of the majority of these proteins make it very feasible to achieve the requisite depth of analysis needed interrogate them further.

5.4. Summary

In conclusion, I believe that there are several methods that can be further optimized in order to improve the CX-MS method. Specifically, ways to improve the

sample preparation, data acquisition, and data analysis protocols have all been described. Moreover, analysis of the proteins and protein complexes that interact with AE1 remain under-characterized. This could be a very promising opportunity for extensive further study.

Bibliography

Alper, S. L. (2009). "Molecular physiology and genetics of Na+-independent SLC4 anion exchangers." Journal of Experimental Biology **212**(Pt 11): 1672-1683.

An, X. and N. Mohandas (2008). "Disorders of red cell membrane." British Journal of Haematology **141**(3): 367-375.

Anong, W. A., T. Franco, et al. (2009). "Adducin forms a bridge between the erythrocyte membrane and its cytoskeleton and regulates membrane cohesion." Blood **114**(9): 1904-1912.

Back, J. W., L. de Jong, et al. (2003). "Chemical cross-linking and mass spectrometry for protein structural modeling." Journal of Molecular Biology **331**(2): 303-313.

Back, J. W., A. F. Hartog, et al. (2001). "A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes." Journal of the American Society for Mass Spectrometry **12**(2): 222-227.

Back, J. W., V. Notenboom, et al. (2002). "Identification of cross-linked peptides for protein interaction studies using mass spectrometry and 18O labeling." Analytical Chemistry **74**(17): 4417-4422.

Barneaud-Rocca, D., F. Borgese, et al. (2011). "Dual transport properties of anion exchanger 1: the same transmembrane segment is involved in anion exchange and in a cation leak." Journal of Biological Chemistry **286**(11): 8909-8916.

Barneaud-Rocca, D., C. Etchebest, et al. (2013). "Structural model of the anion exchanger 1 (SLC4A1) and identification of transmembrane segments forming the transport site." Journal of Biological Chemistry **288**(37): 26372-26384.

Belsom, A., M. Schneider, et al. (2015). "Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology." Mol Cell Proteomics.

Bonar, P., H. P. Schneider, et al. (2013). "Three-dimensional model for the human Cl-/HCO3- exchanger, AE1, by homology to the E. coli ClC protein." Journal of Molecular Biology **425**(14): 2591-2608.

Brosius, F. C., 3rd, S. L. Alper, et al. (1989). "The major kidney band 3 gene transcript predicts an amino-terminal truncated band 3 polypeptide." Journal of Biological Chemistry **264**(14): 7784-7787.

Bruce, J. E. (2012). "In vivo protein complex topologies: sights through a cross-linking lens." Proteomics **12**(10): 1565-1575.

Brunati, A. M., L. Bordin, et al. (2000). "Sequential phosphorylation of protein band 3 by Syk and Lyn tyrosine kinases in intact human erythrocytes: identification of primary and secondary phosphorylation sites." Blood **96**(4): 1550-1557.

Budayeva, H. G. and I. M. Cristea (2014). "A mass spectrometry view of stable and transient protein interactions." Advances in Experimental Medicine and Biology **806**: 263-282.

Bumpus, N. N. and P. F. Hollenberg (2010). "Cross-linking of human cytochrome P450 2B6 to NADPH-cytochrome P450 reductase: Identification of a potential site of interaction." Journal of Inorganic Biochemistry **104**(4): 485-488.

Burton, N. M. and L. J. Bruce (2011). "Modelling the structure of the red cell membrane." Biochemistry and Cell Biology **89**(2): 200-215.

Chen, J. W., C. Dodia, et al. (2000). "1-Cys peroxiredoxin, a bifunctional enzyme with glutathione peroxidase and phospholipase A2 activities." Journal of Biological Chemistry **275**(37): 28421-28427.

Chen, Z. A., A. Jawhari, et al. (2010). "Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry." EMBO Journal **29**(4): 717-726.
Cheung, J. C., E. Cordat, et al. (2005). "Trafficking defects of the Southeast Asian ovalocytosis deletion mutant of anion exchanger 1 membrane proteins." Biochemical Journal **392**(Pt 3): 425-434.
Cheung, J. C., J. Li, et al. (2005). "Topology of transmembrane segments 1-4 in the human chloride/bicarbonate anion exchanger 1 (AE1) by scanning N-glycosylation mutagenesis." Biochemical Journal **390**(Pt 1): 137-144.
Cheung, J. C. and R. A. Reithmeier (2005). "Membrane integration and topology of the first transmembrane segment in normal and Southeast Asian ovalocytosis human erythrocyte anion exchanger 1." Molecular Membrane Biology **22**(3): 203-214.
Choi, H. J., S. W. Kang, et al. (1998). "Crystal structure of a novel human peroxidase enzyme at 2.0 A resolution." Nat Struct Biol **5**(5): 400-406.
Choi, I. (2012). "SLC4A transporters." Curr Top Membr **70**: 77-103.
Cox, J. and M. Mann (2008). "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification." Nature Biotechnology **26**(12): 1367-1372.
Davis, L., S. E. Lux, et al. (1989). "Mapping the ankyrin-binding site of the human erythrocyte anion exchanger." Journal of Biological Chemistry **264**(16): 9665-9672.
Eber, S. and S. E. Lux (2004). "Hereditary spherocytosis--defects in proteins that connect the membrane skeleton to the lipid bilayer." Seminars in Hematology **41**(2): 118-141.
El-Shafey, A., N. Tolic, et al. (2006). ""Zero-length" cross-linking in solid state as an approach for analysis of protein-protein interactions." Protein Science **15**(3): 429-440.
Emsley, P., B. Lohkamp, et al. (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.
Eng, J. K., A. L. McCormack, et al. (1994). "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database." Journal of the American Society for Mass Spectrometry **5**(11): 976-989.
Erickson, H. K. (1997). "Cytoplasmic disposition of aspartate 821 in anion exchanger from human erythrocytes." Biochemistry **36**(33): 9958-9967.
Fairbanks, G., T. L. Steck, et al. (1971). "Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane." Biochemistry **10**(13): 2606-2617.
Fasold, H., J. Klappenberger, et al. (1971). "Bifunctional reagents for the crosslinking of proteins." Angewandte Chemie. International Ed. In English **10**(11): 795-801.
Fisher, A. B., C. Dodia, et al. (1999). "Phospholipid hydroperoxides are substrates for non-selenium glutathione peroxidase." Journal of Biological Chemistry **274**(30): 21326-21334.
Fisher, A. B., H. J. Forman, et al. (1984). "Mechanisms of pulmonary oxygen toxicity." Lung **162**(5): 255-259.
Frank, A. M., M. M. Savitski, et al. (2007). "De novo peptide sequencing and identification with precision mass spectrometry." J Proteome Res **6**(1): 114-123.
Fujinaga, J., X. B. Tang, et al. (1999). "Topology of the membrane domain of human erythrocyte anion exchange protein, AE1." Journal of Biological Chemistry **274**(10): 6626-6633.

Gallagher, P. G. (2005). "Red cell membrane disorders." Hematology Am Soc Hematol Educ Program: 13-18.
Gallagher, P. G., W. T. Tse, et al. (1992). "A common type of the spectrin alpha I 46-50a-kD peptide abnormality in hereditary elliptocytosis and pyropoikilocytosis is associated with a mutation distant from the proteolytic cleavage site. Evidence for the functional importance of the triple helical model of spectrin." Journal of Clinical Investigation **89**(3): 892-898.
Gao, Q., C. E. Doneanu, et al. (2006). "Identification of the interactions between cytochrome P450 2E1 and cytochrome b5 by mass spectrometry and site-directed mutagenesis." Journal of Biological Chemistry **281**(29): 20404-20417.
Glele-Kakai, C., M. Garbarz, et al. (1996). "Epidemiological studies of spectrin mutations related to hereditary elliptocytosis and spectrin polymorphisms in Benin." British Journal of Haematology **95**(1): 57-66.
Gotze, M., J. Pettelkau, et al. (2012). "StavroX--a software for analyzing crosslinked products in protein interaction studies." Journal of the American Society for Mass Spectrometry **23**(1): 76-87.
Grabarek, Z. and J. Gergely (1990). "Zero-length crosslinking procedure with the use of active esters." Analytical Biochemistry **185**(1): 131-135.
Greber, B. J., D. Boehringer, et al. (2014). "The complete structure of the large subunit of the mammalian mitochondrial ribosome." Nature **515**(7526): 283-286.
Grey, J. L., G. C. Kodippili, et al. (2012). "Identification of contact sites between ankyrin and band 3 in the human erythrocyte membrane." Biochemistry **51**(34): 6838-6846.
Grinstein, S., S. Ship, et al. (1978). "Anion transport in relation to proteolytic dissection of band 3 protein." Biochimica et Biophysica Acta **507**(2): 294-304.
Haladova, K., H. Mrazek, et al. (2012). "The combination of hydrogen/deuterium exchange or chemical cross-linking techniques with mass spectrometry: mapping of human 14-3-3zeta homodimer interface." Journal of Structural Biology **179**(1): 10-17.
Harper, S. L., D. Li, et al. (2010). "A fused alpha-beta "mini-spectrin" mimics the intact erythrocyte spectrin head-to-head tetramer." Journal of Biological Chemistry **285**(14): 11003-11012.
Harper, S. L., S. Sriswasdi, et al. (2013). "The common hereditary elliptocytosis-associated alpha-spectrin L260P mutation perturbs erythrocyte membranes by stabilizing spectrin in the closed dimer conformation." Blood **122**(17): 3045-3053.
Herzog, F., A. Kahraman, et al. (2012). "Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry." Science **337**(6100): 1348-1352.
Hirai, T., N. Hamasaki, et al. (2011). "Topology models of anion exchanger 1 that incorporate the anti-parallel V-shaped motifs found in the EM structure." Biochemistry and Cell Biology **89**(2): 148-156.
Hyung, S. J. and B. T. Ruotolo (2012). "Integrating mass spectrometry of intact protein complexes into structural proteomics." Proteomics **12**(10): 1547-1564.
Jin, X. R., Y. Abe, et al. (2003). "Histidine-834 of human erythrocyte band 3 has an essential role in the conformational changes that occur during the band 3-mediated anion exchange." Biochemistry **42**(44): 12927-12932.
Kalkhof, S., C. Ihling, et al. (2005). "Chemical cross-linking and high-performance Fourier transform ion cyclotron resonance mass spectrometry for protein

interaction analysis: application to a calmodulin/target peptide complex." Analytical Chemistry **77**(2): 495-503.
Kang, S. W., I. C. Baines, et al. (1998). "Characterization of a mammalian peroxiredoxin that contains one conserved cysteine." J Biol Chem **273**(11): 6303-6311.
Kao, A., C. L. Chiu, et al. (2011). "Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes." Mol Cell Proteomics **10**(1): M110 002212.
Kelley, L. A., S. Mezulis, et al. (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nat Protoc **10**(6): 845-858.
Kiselar, J. G. and M. R. Chance (2010). "Future directions of structural mass spectrometry using hydroxyl radical footprinting." Journal of Mass Spectrometry **45**(12): 1373-1382.
Konermann, L., J. Pan, et al. (2011). "Hydrogen exchange mass spectrometry for studying protein structure and dynamics." Chem Soc Rev **40**(3): 1224-1234.
Konijnenberg, A., A. Butterer, et al. (2013). "Native ion mobility-mass spectrometry and related methods in structural biology." Biochimica et Biophysica Acta **1834**(6): 1239-1256.
Kopito, R. R. and H. F. Lodish (1985). "Primary structure and transmembrane orientation of the murine anion exchange protein." Nature **316**(6025): 234-238.
Lang, J. D., P. J. McArdle, et al. (2002). "Oxidant-antioxidant balance in acute lung injury." Chest **122**(6 Suppl): 314S-320S.
Lanucara, F., S. W. Holman, et al. (2014). "The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics." Nat Chem **6**(4): 281-294.
Lasker, K., F. Forster, et al. (2012). "Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach." Proceedings of the National Academy of Sciences of the United States of America **109**(5): 1380-1387.
Leitner, A., L. A. Joachimiak, et al. (2012). "The molecular architecture of the eukaryotic chaperonin TRiC/CCT." Structure **20**(5): 814-825.
Leitner, A., L. A. Joachimiak, et al. (2014). "Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes." Proceedings of the National Academy of Sciences of the United States of America **111**(26): 9455-9460.
Leitner, A., T. Walzthoeni, et al. (2010). "Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics." Mol Cell Proteomics **9**(8): 1634-1649.
Lepke, S. and H. Passow (1976). "Effects of incorporated trypsin on anion exchange and membrane proteins in human red blood cell ghosts." Biochimica et Biophysica Acta **455**(2): 353-370.
Lepvrier, E., C. Doigneaux, et al. (2014). "Optimized protocol for protein macrocomplexes stabilization using the EDC, 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide, zero-length cross-linker." Analytical Chemistry **86**(21): 10524-10530.
Li, C., S. Takazaki, et al. (2006). "Identification of oxidized methionine sites in erythrocyte membrane protein by liquid chromatography/electrospray ionization mass spectrometry peptide mapping." Biochemistry **45**(39): 12117-12124.

Li, D., S. L. Harper, et al. (2010). "A comprehensive model of the spectrin divalent tetramer binding region deduced using homology modeling and chemical cross-linking of a mini-spectrin." J Biol Chem **285**(38): 29535-29545.
Li, D., H. Y. Tang, et al. (2008). "A structural model of the erythrocyte spectrin heterodimer initiation site determined using homology modeling and chemical cross-linking." Journal of Biological Chemistry **283**(3): 1553-1562.
Li, J., J. Quilty, et al. (2000). "Processing of N-linked oligosaccharide depends on its location in the anion exchanger, AE1, membrane glycoprotein." Biochemical Journal **349**(Pt 1): 51-57.
Liu, H., R. Y. Huang, et al. (2011). "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 **108**(45): 18536-18541.
Lopez-Alonso, J. P., F. Diez-Garcia, et al. (2009). "Carbodiimide EDC induces cross-links that stabilize RNase A C-dimer against dissociation: EDC adducts can affect protein net charge, conformation, and activity." Bioconjugate Chemistry **20**(8): 1459-1473.
Lu, F., S. Li, et al. (2011). "Structure and mechanism of the uracil transporter UraA." Nature **472**(7342): 243-246.
Maleknia, S. D. and K. M. Downard (2014). "Advances in radical probe mass spectrometry for protein footprinting in chemical biology applications." Chem Soc Rev **43**(10): 3244-3258.
Manevich, Y. and A. B. Fisher (2005). "Peroxiredoxin 6, a 1-Cys peroxiredoxin, functions in antioxidant defense and lung phospholipid metabolism." Free Radic Biol Med **38**(11): 1422-1432.
Manevich, Y., K. S. Reddy, et al. (2007). "Structure and phospholipase function of peroxiredoxin 6: identification of the catalytic triad and its role in phospholipid substrate binding." Journal of Lipid Research **48**(10): 2306-2318.
Manevich, Y., T. Shuvaeva, et al. (2009). "Binding of peroxiredoxin 6 to substrate determines differential phospholipid hydroperoxide peroxidase and phospholipase A(2) activities." Archives of Biochemistry and Biophysics **485**(2): 139-149.
Mankelow, T. J., T. J. Satchwell, et al. (2012). "Refined views of multi-protein complexes in the erythrocyte membrane." Blood Cells, Molecules, and Diseases **49**(1): 1-10.
Marekov, L. N. (2007). "Determination of protein contacts by chemical cross-linking with EDC and mass spectrometry." Curr Protoc Protein Sci **Chapter 19**: Unit 19 16.
McIlwain, S., P. Draghicescu, et al. (2010). "Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs." J Proteome Res **9**(5): 2488-2495.
Merkley, E. D., J. R. Cort, et al. (2013). "Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze." J Struct Funct Genomics **14**(3): 77-90.
Mitra, K., I. Ubarretxena-Belandia, et al. (2004). "Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol." Proceedings of the National Academy of Sciences of the United States of America **101**(12): 4083-4088.
Monroe, E. B. and M. L. Heien (2013). "Electrochemical generation of hydroxyl radicals for examining protein structure." Analytical Chemistry **85**(13): 6185-6189.

Nagao, R., T. Suzuki, et al. (2010). "Topological analysis of the extrinsic PsbO, PsbP and PsbQ proteins in a green algal PSII complex by cross-linking with a water-soluble carbodiimide." Plant and Cell Physiology **51**(5): 718-727.
Okubo, K., D. Kang, et al. (1994). "Red blood cell band 3. Lysine 539 and lysine 851 react with the same H2DIDS (4,4'-diisothiocyanodihydrostilbene-2,2'-disulfonic acid) molecule." Journal of Biological Chemistry **269**(3): 1918-1926.
Olson, A. L., A. T. Tucker, et al. (2014). "Structure and DNA-binding traits of the transition state regulator AbrB." Structure **22**(11): 1650-1656.
Pacholarz, K. J., R. A. Garlish, et al. (2012). "Mass spectrometry based tools to investigate protein-ligand interactions for drug discovery." Chem Soc Rev **41**(11): 4335-4355.
Pan, C., B. H. Park, et al. (2010). "A high-throughput de novo sequencing approach for shotgun proteomics using high-resolution tandem mass spectrometry." BMC Bioinformatics **11**: 118.
Paramelle, D., G. Miralles, et al. (2013). "Chemical cross-linkers for protein structure studies by mass spectrometry." Proteomics **13**(3-4): 438-456.
Park, C. Y., A. A. Klammer, et al. (2008). "Rapid and accurate peptide identification from tandem mass spectra." J Proteome Res **7**(7): 3022-3027.
Passow, H. (1986). "Molecular aspects of band 3 protein-mediated anion transport across the red blood cell membrane." Reviews of Physiology Biochemistry and Pharmacology **103**: 61-203.
Peri, S., H. Steen, et al. (2001). "GPMAW--a software tool for analyzing proteins and peptides." Trends in Biochemical Sciences **26**(11): 687-689.
Perkins, D. N., D. J. Pappin, et al. (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-3567.
Perrotta, S., P. G. Gallagher, et al. (2008). "Hereditary spherocytosis." Lancet **372**(9647): 1411-1426.
Petrotchenko, E. V., J. J. Serpa, et al. (2011). "An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics." Mol Cell Proteomics **10**(2): M110 001420.
Politis, A., C. Schmidt, et al. (2015). "Topological models of heteromeric protein assemblies from mass spectrometry: application to the yeast eIF3:eIF5 complex." Chemistry and Biology **22**(1): 117-128.
Poole, J. (2000). "Red cell antigens on band 3 and glycophorin A." Blood Reviews **14**(1): 31-43.
Poor, T. A., L. M. Jones, et al. (2014). "Probing the paramyxovirus fusion (F) protein-refolding event from pre- to postfusion by oxidative footprinting." Proceedings of the National Academy of Sciences of the United States of America **111**(25): E2596-2605.
Popov, M., J. Li, et al. (1999). "Transmembrane folding of the human erythrocyte anion exchanger (AE1, Band 3) determined by scanning and insertional N-glycosylation mutagenesis." Biochemical Journal **339 (Pt 2)**: 269-279.
Popov, M., L. Y. Tam, et al. (1997). "Mapping the ends of transmembrane segments in a polytopic membrane protein. Scanning N-glycosylation mutagenesis of extracytosolic loops in the anion exchanger, band 3." Journal of Biological Chemistry **272**(29): 18325-18332.

Rahaman, H., S. Zhou, et al. (2012). "Increased phospholipase A2 activity with phosphorylation of peroxiredoxin 6 requires a conformational change in the protein." Biochemistry **51**(27): 5521-5530.
Rahman, I. and I. M. Adcock (2006). "Oxidative stress and redox regulation of lung inflammation in COPD." European Respiratory Journal **28**(1): 219-242.
Ralat, L. A., Y. Manevich, et al. (2006). "Direct evidence for the formation of a complex between 1-cysteine peroxiredoxin and glutathione S-transferase pi with activity changes in both enzymes." Biochemistry **45**(2): 360-372.
Rappsilber, J. (2011). "The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes." Journal of Structural Biology **173**(3): 530-540.
Rhee, S. G., H. Z. Chae, et al. (2005). "Peroxiredoxins: a historical overview and speculative preview of novel mechanisms and emerging concepts in cell signaling." Free Radical Biology and Medicine **38**(12): 1543-1552.
Rinner, O., J. Seebacher, et al. (2008). "Identification of cross-linked peptides from large sequence databases." Nat Methods **5**(4): 315-318.
Rivera-Santiago, R., S. Harper, et al. (2015). "Solution structure of the reduced form of human peroxiredoxin-6 elucidated using zero-length chemical cross-linking and homology modeling." Biochemical Journal.
Rivera-Santiago, R. F., S. Sriswasdi, et al. (2015). "Probing structures of large protein complexes using zero-length cross-linking." Methods.
Ruotolo, B. T., J. L. Benesch, et al. (2008). "Ion mobility-mass spectrometry analysis of large protein complexes." Nat Protoc **3**(7): 1139-1152.
Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol **234**(3): 779-815.
Salomao, M., X. Zhang, et al. (2008). "Protein 4.1R-dependent multiprotein complex: new insights into the structural organization of the red blood cell membrane." Proceedings of the National Academy of Sciences of the United States of America **105**(23): 8026-8031.
Schmidt, A., S. Kalkhof, et al. (2005). "Mapping protein interfaces by chemical cross-linking and Fourier transform ion cyclotron resonance mass spectrometry: application to a calmodulin / adenylyl cyclase 8 peptide complex." Eur J Mass Spectrom (Chichester, Eng) **11**(5): 525-534.
Serpa, J. J., C. E. Parker, et al. (2012). "Mass spectrometry-based structural proteomics." Eur J Mass Spectrom (Chichester, Eng) **18**(2): 251-267.
Shnitsar, V., J. Li, et al. (2013). "A substrate access tunnel in the cytosolic domain is not an essential feature of the solute carrier 4 (SLC4) family of bicarbonate transporters." Journal of Biological Chemistry **288**(47): 33848-33860.
Simons, K. T., R. Bonneau, et al. (1999). "Ab initio protein structure prediction of CASP III targets using ROSETTA." Proteins **Suppl 3**: 171-176.
Singh, P., E. Nakatani, et al. (2013). "A pseudo-atomic model for the capsid shell of bacteriophage lambda using chemical cross-linking/mass spectrometry and molecular modeling." Journal of Molecular Biology **425**(18): 3378-3388.
Singh, P., S. A. Shaffer, et al. (2008). "Characterization of protein cross-links via mass spectrometry and an open-modification search strategy." Analytical Chemistry **80**(22): 8799-8806.
Sinz, A. (2006). "Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions." Mass Spectrometry Reviews **25**(4): 663-682.

Speicher, D. W., L. Weglarz, et al. (1992). "Properties of human red cell spectrin heterodimer (side-to-side) assembly and identification of an essential nucleation site." Journal of Biological Chemistry **267**(21): 14775-14782.
Sriswasdi, S., S. L. Harper, et al. (2014). "Probing large conformational rearrangements in wild-type and mutant spectrin using structural mass spectrometry." Proc Natl Acad Sci U S A **111**(5): 1801-1806.
Sriswasdi, S., S. L. Harper, et al. (2014). "Enhanced identification of zero-length chemical cross-links using label-free quantitation and high-resolution fragment ion spectra." J Proteome Res **13**(2): 898-914.
Staros, J. V., R. W. Wright, et al. (1986). "Enhancement by N-hydroxysulfosuccinimide of water-soluble carbodiimide-mediated coupling reactions." Analytical Biochemistry **156**(1): 220-222.
Stehberger, P. A., B. E. Shmukler, et al. (2007). "Distal renal tubular acidosis in mice lacking the AE1 (band3) Cl-/HCO3- exchanger (slc4a1)." Journal of the American Society of Nephrology **18**(5): 1408-1418.
Stengel, F., R. Aebersold, et al. (2012). "Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes." Mol Cell Proteomics **11**(3): R111 014027.
Takazaki, S., Y. Abe, et al. (2010). "Mutation of His 834 in human anion exchanger 1 affects substrate binding." Biochimica et Biophysica Acta **1798**(5): 903-908.
Tang, X., G. R. Munske, et al. (2005). "Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions." Analytical Chemistry **77**(1): 311-318.
Tang, X. B., J. Fujinaga, et al. (1998). "Topology of the region surrounding Glu681 of human AE1 protein, the erythrocyte anion exchanger." Journal of Biological Chemistry **273**(35): 22545-22553.
Tanner, M. J. (2002). "Band 3 anion exchanger and its involvement in erythrocyte and kidney disorders." Current Opinion in Hematology **9**(2): 133-139.
Taylor, A. M., Q. Zhu, et al. (2001). "Cysteine-directed cross-linking localizes regions of the human erythrocyte anion-exchange protein (AE1) relative to the dimeric interface." Biochemical Journal **359**(Pt 3): 661-668.
Vahidi, S., B. B. Stocks, et al. (2013). "Submillisecond protein folding events monitored by rapid mixing and mass spectrometry-based oxidative labeling." Analytical Chemistry **85**(18): 8618-8625.
van den Akker, E., T. J. Satchwell, et al. (2010). "Band 3 multiprotein complexes in the red cell membrane; of mice and men." Blood Cells, Molecules, and Diseases **45**(1): 1-8.
Vastermark, A. and M. H. Saier, Jr. (2014). "Evolutionary relationship between 5+5 and 7+7 inverted repeat folds within the amino acid-polyamine-organocation superfamily." Proteins **82**(2): 336-346.
Wales, T. E. and J. R. Engen (2006). "Hydrogen exchange mass spectrometry for the analysis of protein dynamics." Mass Spectrometry Reviews **25**(1): 158-170.
Walzthoeni, T., A. Leitner, et al. (2013). "Mass spectrometry supported determination of protein complex structure." Current Opinion in Structural Biology **23**(2): 252-260.
Wang, D. N., W. Kuhlbrandt, et al. (1993). "Two-dimensional structure of the membrane domain of human band 3, the anion transport protein of the erythrocyte membrane." EMBO Journal **12**(6): 2233-2239.

Wang, D. N., V. E. Sarabia, et al. (1994). "Three-dimensional map of the dimeric membrane domain of the human erythrocyte anion exchanger, Band 3." EMBO Journal **13**(14): 3230-3235.
Wang, G., R. R. Abzalimov, et al. (2013). "Conformer-specific characterization of nonnative protein states using hydrogen exchange and top-down mass spectrometry." Proceedings of the National Academy of Sciences of the United States of America **110**(50): 20087-20092.
Wang, L. and M. R. Chance (2011). "Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting." Analytical Chemistry **83**(19): 7234-7241.
Wei, H., J. Mo, et al. (2014). "Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications." Drug Discov Today **19**(1): 95-102.
Wood, P. G., H. Muller, et al. (1992). "Role of Lys 558 and Lys 869 in substrate and inhibitor binding to the murine band 3 protein: a study of the effects of site-directed mutagenesis of the band 3 protein expressed in the oocytes of Xenopus laevis." Journal of Membrane Biology **127**(2): 139-148.
Wood, Z. A., E. Schroder, et al. (2003). "Structure, mechanism and regulation of peroxiredoxins." Trends in Biochemical Sciences **28**(1): 32-40.
Wu, Y., S. I. Feinstein, et al. (2009). "Mitogen-activated protein kinase-mediated phosphorylation of peroxiredoxin 6 regulates its phospholipase A(2) activity." Biochem J **419**(3): 669-679.
Xu, G. and M. R. Chance (2004). "Radiolytic modification of acidic amino acid residues in peptides: probes for examining protein-protein interactions." Analytical Chemistry **76**(5): 1213-1221.
Xu, G. and M. R. Chance (2005). "Radiolytic modification of sulfur-containing amino acid residues in model peptides: fundamental studies for protein footprinting." Analytical Chemistry **77**(8): 2437-2449.
Xu, G., K. Takamoto, et al. (2003). "Radiolytic modification of basic amino acid residues in peptides: probes for examining protein-protein interactions." Analytical Chemistry **75**(24): 6995-7007.
Xu, H. and M. A. Freitas (2009). "MassMatrix: a database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data." Proteomics **9**(6): 1548-1555.
Yamaguchi, T., Y. Ikeda, et al. (2010). "Structure of the membrane domain of human erythrocyte anion exchanger 1 revealed by electron crystallography." Journal of Molecular Biology **397**(1): 179-189.
Yan, Y., G. Chen, et al. (2014). "Fast photochemical oxidation of proteins (FPOP) maps the epitope of EGFR binding to adnectin." Journal of the American Society for Mass Spectrometry **25**(12): 2084-2092.
Yang, B., Y. J. Wu, et al. (2012). "Identification of cross-linked peptides from complex samples." Nat Methods **9**(9): 904-906.
Young, M. M., N. Tang, et al. (2000). "High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry." Proceedings of the National Academy of Sciences of the United States of America **97**(11): 5802-5806.
Yusoff, N. M., H. Van Rostenberghe, et al. (2003). "High prevalence of Southeast Asian ovalocytosis in Malays with distal renal tubular acidosis." Journal of Human Genetics **48**(12): 650-653.

Zhang, D., A. Kiyatkin, et al. (2000). "Crystallographic structure and functional interpretation of the cytoplasmic domain of erythrocyte membrane band 3." Blood **96**(9): 2925-2933.
Zhu, Q., D. W. Lee, et al. (2003). "Novel topology in C-terminal region of the human plasma membrane anion exchanger, AE1." Journal of Biological Chemistry **278**(5): 3112-3120.
Zhu, W., J. W. Smith, et al. (2010). "Mass spectrometry-based label-free quantitative proteomics." J Biomed Biotechnol **2010**: 840518.
Zybailov, B. L., G. V. Glazko, et al. (2013). "Large Scale Chemical Cross-linking Mass Spectrometry Perspectives." J Proteomics Bioinform **6**(Suppl 2): 001.