Abstract

ARGO, ANDREW SCOTT. Elucidation of the BRI1/BAK1 Heterodimer Using Chemical Crosslinking and Multidimensional Liquid Chromatography-Tandem Mass Spectrometry. (Under the Guidance of Dr. Michael B. Goshe.)

Chemical crosslinking is a tool amongst many others that can be used to help in the refinement/characterization of protein structure. It becomes an even more powerful tool when combined with other structural techniques such as crystallography or NMR. Membrane proteins have posed particular challenges to crystallography and NMR due to their instability in solution and large size compared to many cytosolic proteins. Chemical crosslinking is not limited by the size of proteins as they are generally digested after crosslinking to be analyzed by mass spectrometry allowing for a large variety of proteins to be studied. Though this technique has numerous advantages, a clear disadvantage is the low population of crosslinked peptides compared to those that are not. This can lead to the inability to detect the crosslinked peptides due to charge competition during electrospray ionization and the low probability of being detected due to their low stoichiometry.

The use of strong cation exchange (SCX) can help in the detection of crosslinked peptides. The crosslinking introduces a minimal +4 charge on the interpeptide crosslinks and this higher charge state causes the crosslinked peptides to bind with higher affinity on SCX columns. Attaching an SCX column on-line with a reversed-phase column on a liquid chromatograph coupled to a mass spectrometer allows for the selective elution of crosslinked peptides based on increasing ionic strength, resulting in an increase in detection of the crosslinked peptides by liquid chromatography-tandem mass spectrometry (LC/MS/MS). The elution of lower charge state peptides due to the application of solvents containing low salt concentrations allows for higher charge state peptides to be enriched on
the SCX column which are then eluted by solvents containing higher salt concentrations, thus increasing the detection of highly charged analytes due to the lowered sample complexity and charge competition.

The ability of chemical crosslinking to characterize proteins in a biologically relevant environment makes it a useful technique to analyze the brassinosteroid-insensitive 1/BRI1-associated kinase 1 (BRI1/BAK1) heterodimer formed by their cytoplasmic domains. With the use of a ThermoScientific Orbitrap Elite Mass Spectrometer implementing SCX-LC/MS/MS a total of 32 interpeptide crosslinks were identified when labeling BRI1/BAK1 with bis(sulfo succinimidyl) suberate (BS3) and disuccinimidyl-succinamyl-aspartyl-proline (SuDP): 12 with BS3 and 20 with SuDP. To model the structure of the heterodimer using the crosslink constraints the HADDOCK protein docking program was used. Multiple static models were created and illustrated how the two proteins could interact as a dimer. Though the static models provide information of how the two kinase domains can interact in solution, it does not take into account the conformational dynamics of these proteins. The crosslinking experiments conducted also involved the use of Flag-BRI1 and GST-BAK1 which are known to homodimerize in solution. To eliminate any higher order oligomers that are simply artifacts of homodimerization, different tags must be used. Further modeling in a more dynamic setting must also be performed to illustrate how these proteins might be activated especially if the steps needed for activation are sequential.
Elucidation of the BRI1/BAK1 Heterodimer Using Chemical Crosslinking and Multidimensional Liquid Chromatography-Tandem Mass Spectrometry

by
Andrew Scott Argo

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Master of Science in Biochemistry

Raleigh, North Carolina
2015

APPROVED BY:

_________________________  __________________________
Dr. Michael B. Goshe       Dr. Steven D. Clouse
Co-Chair of Advisory Committee  Co-Chair of Advisory Committee

_________________________
Dr. Linda Hanley-Bowdoin  Dr. Benjamin Bobay
Biography

Andrew Scott Argo was born in Hopkinsville, KY on April 24, 1989 to Eric and Donna Argo. He lived in Richmond, Indiana and Raeford, North Carolina before moving to Raleigh, North Carolina to attend college at North Carolina State University (NCSU). He originally entered NCSU in 2007 as a Chemical Engineering Major and made a move to the Biochemistry Program in 2009, graduating with his B.S. in the spring of 2012. He then entered into graduate school in the Biochemistry Program for his Master's specializing in mass spectrometry with Dr. Michael B. Goshe. Plans after graduation include entering the Biotechnology Program at NCSU in the fall of 2014 to obtain another Master’s and then pursuing a career in industry.
# Table Of Contents

List of Figures ................................................................................................................................. iv  
List of Tables ................................................................................................................................. v  
Chapter 1 Introduction .................................................................................................................. 1  
Chapter 2 Methods ...................................................................................................................... 47  
Chapter 3 Results and Discussion ............................................................................................... 53  
Chapter 4 Conclusion ................................................................................................................... 88  
Chapter 5 Future Work ............................................................................................................... 89  
Literature Cited ............................................................................................................................ 92
List of Figures

Figure 1.3.1  Crosslinking reaction for lysine specific reagent ........................................ 5
Figure 1.3.2  Double immunodiffusion .................................................................................. 7
Figure 1.3.3  Radiochromatogram of papain digest ............................................................... 8
Figure 1.4.1  Schematic of double focusing magnetic sector .................................................. 10
Figure 1.4.2  Mock-up of the orbitrap Velos Pro ................................................................... 12
Figure 1.4.3  Representation of resolution/resolving power .................................................. 13
Figure 1.5.1  Crosslinking experiment flowthrough ............................................................... 15
Figure 1.6.1  BS2 and DSS crosslinkers .................................................................................. 17
Figure 1.6.2  The zero length crosslinker EDC .................................................................... 19
Figure 1.6.1.1  SuDP cleavable crosslinker ............................................................................ 24
Figure 1.6.2.1  Flowthrough of crosslinked peptide identification .......................................... 26
Figure 1.7.1  Typical cell surface receptor .............................................................................. 28
Figure 1.7.2  Clavata signaling pathway .................................................................................. 30
Figure 1.8.1  Recycling of activated cell surface receptors ..................................................... 33
Figure 1.8.2  Schematic of members of the LRR protein family ............................................. 36
Figure 1.9.1  BRI1/BAK1 cytoplasmic domain ribbon diagrams ........................................... 40
Figure 1.9.2  BRI1/BAK1 extracellular domain ribbon diagrams .......................................... 41
Figure 1.9.3  Model of the BRI1/BAK1 sequential transphosphorylation event ....................... 45
Figure 3.1.1  SCX resin: Polysulfoethyl A ............................................................................. 55
Figure 3.1.2  SCX chromatogram of BSA digest with disulfide bonds intact ......................... 57
Figure 3.1.3  BSA crosslinked with BS3 and separated with SCX chromatography .......... 59
Figure 3.1.4  Overlay of BSA digests at different BS3 amounts ............................................. 60
Figure 3.1.5  Flowthrough for typical online SCX/RP 2D separation ................................... 63
Figure 3.1.6  In-house packed SCX trapping column ............................................................ 64
Figure 3.1.7  Online BSA separation charge state population ............................................... 67
Figure 3.1.8  Charge state population of standard *e. coli* lysate digested with Lys C .......... 68
Figure 3.2.1  Immunoblot of GST-BRI1 and GST-BAK1 ......................................................... 72
Figure 3.2.2  Coomassie stain of GST-mBAK1 ..................................................................... 74
Figure 3.2.3  Immunoblot of BS3 crosslinked BRI1/BAK1 ..................................................... 75
Figure 3.2.4  Anti-Flag and anti-GST immunoblot of SuDP BRI1/BAK1 ............................... 78
Figure 3.2.5  HADDOCK docking simulation of first subset of crosslinks ............................... 84
Figure 3.2.6  HADDOCK docking simulation of second subset of crosslinks ......................... 85
Figure 3.2.7  Electrostatic modeling of the BRI1/BAK1 heterodimer ..................................... 86
Figure 3.2.8  Hydrophobic mapping of the BRI1/BAK1 heterodimer ...................................... 87
List of Tables

Table 3.2.1: Table of BS3 Crosslinks
.................................................................................................................. 81
Table 3.2.2: Table of SuDP Crosslinks ........................................................................................ 82
Chapter 1: Introduction

Proteins and, in particular, the complexes they form have essential and regulatory structural roles in most cellular functions. Characterizing protein structure can help in the explanation of protein functions and how particular functions are performed. It is of no surprise that the study of protein chemistry has become more prevalent as the benefits span from disease treatment and prevention to helping keep highly coveted resources such as gasoline from becoming scarce (Simanshu et al. 2013; Choi & Lee 2013). Protein structure is the basis for protein function; it dictates how and where proteins bind ligands and the faces of domains that participate in protein complexing. Though nature has provided a select few conserved secondary structures that go into forming the tertiary structure, the combinations of amino acids and the number of ways these structures can combine into higher order complexes created the need for structural techniques to determine protein structure to understand its function in biochemical processes.

1.1 NMR

Crystallography and NMR have proven their usefulness in discovering binding sites and how they function in an active cell but limitations do exist for these techniques. NMR has the capability of studying large proteins with dynamic regions in solution. It requires a large amount of pure protein that can be expressed in the presence of isotopically labeled residues or suspended in a deuterated solvent depending on experimental conditions. It has been shown that the average length of
a protein sequence in humans is 375 amino acid residues giving a protein in a molecular weight range that can feasibly be analyzed (Brocchieri & Karlin 2005). Data collection using isotopically labeled residues or deuterated solvents can make analysis easier but isotopically labeled reagents can be costly. The increase in size also causes a large decay in resolution. This occurs in part due to the overlap of resonance signals from the increase in the number of NMR active nuclei and the faster relaxation times that cause line broadening and a decrease in signal selectivity. These size limits have been realized and overcome with techniques such as transverse relaxation-optimized spectroscopy (TROSY), which helps limit the transverse relaxation returning the sensitivity and selectivity to an acceptable level (Garder & Kay 1998). Integral membrane proteins still pose a problem for this technique due to their size and the need for a lipid/lipid-detergent membrane. These add to the complexity of the signals and longer data collection periods as a consequence of the size of the proteins.

1.2 Crystallography

X-ray crystallography arose around the year 1912 and became fully realized during the work of the Bragg father and son team, who were able to elucidate the structure of simple table salt (Bragg et al. 1921). Most structures during the early years of this technique were simple inorganic compounds (Bragg 1914a; Bragg 1914b; Aminoff 1919). A technique once able to produce structures of simple compounds that crystallized easily is now able to elucidate structures of large
proteins as computational power and methods developed. As of 4/16/2014 there have been 88079 structures solved by crystallography, exceeding NMR which has only 10404 (rcsb.org/pdb/statistics/holdings). The problem of protein size which plagues NMR regularly is not as much an issue for crystallography as seen by work done to crystallize and characterize the prokaryotic 50S large ribosomal subunit containing 27 of the 30 proteins that weigh in at a staggering 1.5 million daltons (Wimberly et al. 2000). This amazing work can all be done at single atom resolution. Crystallography has one large glaring disadvantage, as the protein must be crystallized to be analyzed. This can pose a large problem at times and requires months or even years of work trying a myriad of conditions. A technique such as growing protein crystals in the presence of magnets to give more ordered crystal units has been done but this adds another expense. The heating of proteins by the X-rays can also cause perturbations in protein structure giving false data. Experiments utilizing cryogenic temperatures have helped but this has been shown to introduce structural changes as well (Moreno 2009). Although integral membrane proteins perform some of the most powerful chemistry and drive the functions behind a cell such as cell-cell signaling, intracellular signaling and protein anchoring, both NMR and crystallography still have trouble characterizing structural information about them (Lodish et al. 2000). The advancement of chemical modifications and the technique known as chemical crosslinking arose to help bridge the gap left by NMR and crystallography.
1.3 Chemical Modifications

The first known chemical crosslinkers, di-alkyl halides and *bis*-imidoesters, were developed around 1955 (Zahn 1955). They were homo-bi-functional linkers allowing the conjoining of two amino acid residues in a protein as seen in Figure 1.3.1. The use of formaldehyde was practiced earlier, but due to the infancy of the technique in 1945, researchers found it difficult to work with on a large scale due to its high reactivity and unpredictable ability to react with numerous different residues. It was known that formaldehyde reacted with numerous residues and those modifications could affect protein activity (French 1945). The techniques and technology to analyze the data did not exist at that time only allowing for experiments on single amino acids and dipeptides to limit the sample complexity. Work began on regularly available easily purified proteins, such as bovine serum albumin (BSA), to introduce new bonds into the folded tertiary structure (Wold 1961). This work showed that the linkers could introduce another level of stability to the proteins even after being subjected to typical denaturing conditions such heat, urea, and reduction/carboxymethylation. Protein activity was not measured during the experiments but it was shown that the lack of change in sedimentation values indicated some form of structure was being retained. Though the lack of instrumentation slowed the development of structural crosslinking, crude methods were used to obtain very low resolution, protein-protein contact information. For example crosslinking of the 50S subunit of the *E. coli* ribosomal subunit in the presence of the S16 protein of the 30S subunit using formaldehyde allowed for
Figure 1.3.1: Crosslinking reaction for lysine specific reagent (www.thermoscientific.com).
extraction of the two using an antibody directed towards S16 (Sun et al. 1974). Identification of these immunological complexes required the use of the Ouchterlony Double Diffusion (ODD) technique. ODD can take upwards of 48 hours to complete thus requiring a large amount of time for only determining simple information as to whether or not two proteins interact with little to no topological information. As seen in Figure 1.3.2, the significant time required for this technique comes from the diffusion of the antigen and antibody through a gel and precipitating if they do in fact interact. This interaction creates a white line between the two sample wells. This helped provide the first steps in creating a picture of the ribosome.

Providing evidence for amino acid residues within close proximity of one another was the first step towards a more common structural use for crosslinking. This became important in determining the mechanisms behind enzyme function. The cysteine protease papain was known to use a free thiol to help in the catalytic breakdown of peptides. It was thought that an imidazole group of a histidine was present in close proximity thus “activating” the thiol group. A $^{14}$C labeled di-bromoacetone linker was used which irreversibly inhibited papain. The protein digest was analyzed to identify the number of each type of amino acid residue followed by radiochromatography to identify the predicted modified amino acids present in the digest following inhibition by 1,3-dibromoacetone (Husain & Lowe 1968). Figure 1.3.3 shows peaks from the inhibited papain protein digest eluting where it was determined that the modified histidine, 1-carboxymethyl-histidine would elute. The lack of a 3-carboxymethyl-histidine modification and subsequent loss of a histidine
**Figure 1.3.2:** If the antibody and antigen interact precipitation will occur leaving a white precipitate in the gel (www.sbs.utexas.edu).
Figure 1.3.3: The chromatogram shows the elution profile of the acid-digestion of papain with the modified histidines being of significance. Below that is the predicted active site configuration of the serine/histidine residues thought to occupy it. As adapted from (Husain & Lowe 1968).
from the total digest showed strong evidence for the participation of a histidine in range of the active site thiol group. This historical view of crosslinking shows how time consuming and laborious of a technique it was and how the lack of technology for the identification of crosslink modification was the major downfall at the time. The need for crosslinker development, instrumentation and computing power for data analysis was realized and has since been the driving force behind the use of this technique to further refine protein structures.

1.4 Mass Analyzers

The use of instruments such as Edman Sequence analyzers and HPLC systems equipped to derivatize protein digests prior to analysis with o-phthalaldehyde (OPA) and 9-fluoromethyl-chloroformate (FMOC) allowed for sample analyses to be performed quantitative and qualitatively. Though these techniques are still used the development of the mass spectrometer, in particular mass analyzers, has pushed the use of crosslinking forward.

Use of magnetic sectors dates back to 1912 and the beginning of mass spectrometry, but did not allow for the type of resolution needed (Aston 1919). Even with the advent of the double focusing magnetic sector shown in Figure 1.4.1 it has fallen behind due to high vacuum needs and slow acquisition times. Quadrupole and triple quad analyzers allow for the fast analysis needed when trying to detect low population species such as crosslinked peptides while providing the ability for tandem MS. However, these analyzers suffer from low resolution (around 5000) and
**Figure 1.4.1:** The inclusion of an electrostatic field allows for a double focusing magnetic sector to concentrate ions based in kinetic energy first to increase resolution ([www.chromedia.com](http://www.chromedia.com)).
poor mass accuracy (>100 ppm) and the mass range is usually limited (around 3000 m/z). Time-of-flight instruments do not suffer from the limited mass range and when combined with a reflectron to focus ions of similar m/z values the resolution is greatly improved (25,000). The creation of an ion trap type analyzer called the orbitrap, Figure 1.4.2, has brought the type of resolution/resolving power and sensitivity needed for crosslinking experiments. Ions are trapped and forced to travel in an orbit around a center electrode. The instrument records data from the axial movement (end to end) of the ions with longer acquisition times giving better resolution as shown in Figure 1.4.3. These types of instruments have the resolution (R=100,000), mass accuracy (<5 ppm) and sensitivity needed to detect crosslinked peptides in complex mixtures.

The further development of this analyzer into a high field trap analyzer has brought along better detection capabilities with little frequency fluctuation resulting in better accuracy across m/z values. This very brief introduction into instrumentation focusing on mass spectrometry shows the dependence of characterizing protein crosslinking for structural information has on the analytical capabilities of mass analyzers.

1.5 Chemical Crosslinking Considerations

In general single proteins can be manipulated easier than a protein complex as their stability in solution is only contingent on their specific needs. Introducing several proteins requires a compromise of solution conditions. Once multiple
**Figure 1.4.2:** This mock-up of the orbitrap system shows the complexity behind the system combining CID, HCD and ETD along with high resolution mass analyzers to allow for high confidence peak identification using different modes needed for each unique sample (Leitner et al. 2010).
Figure 1.4.3: Representing the relationship between resolution and resolving power and its importance in peak identification (www.thermoscientific.com).
proteins are introduced in solution together the complexity increases as each protein, even present in a complex, has its own pH and salt conditions, and concentration that keeps it stable in solution and in its native form. The DNA sequence of a protein can be transfected into a cell line with its own tag, short or an entire protein domain, for over-expression. This can lead to problems due to the interaction of tags, which can positively or negatively affect complex formation. Non-native concentrations of proteins can also force certain interactions that are not biologically relevant. The advantage of using a tag allows for the removal of proteins from complex lysates to decrease complexity of a crosslinking experiment. It is common practice to use a multitude of crosslinkers when first developing a method due to different reaction rates and ability for crosslinkers to interact with portions of a protein based on hydrophobicity, charge etc (Paramelle 2013). Methods developed generally follow the outline in Figure 1.5.1. This aspect of crosslinking has not changed due to the large differences in proteins in regards to size, reactivity to crosslinker etc. A trade off had to be realized between crosslink yield and conditions that do not over crosslink the complex leading to all components of the complex to be joined and altering protein structure or under crosslinking which leads to under utilization of available protein. These experimental considerations highlight the significant amount of work that must be done even before loading samples onto a mass spectrometer for analysis. Even with the challenges that face crosslinking as a technique its strength in allowing the analysis of proteins at physiological conditions in solution to produce data more relevant to a biological system is not diminished.
**Figure 1.5.1:** Typical crosslinking experiment flow through starting from the isolated protein/complex and ending at modeling of the acquired data (Back et al. 2003).
1.6 The Use of Chemical Crosslinking for Protein Structure Analysis

It has been over a decade since it was first reported that crosslinking, in combination with mass spectrometry, could help in the elucidation of protein structure at low resolution. Crosslinking can help define the proteins in a more biologically relevant manner. Rappsilber et al. used a yeast nuclear pore protein (Nup85p) in combination with a number of crosslinkers, including commonly used BS$_3$ and DSS shown in Figure 1.6.1 (Rappsibler et al. 2000). Analysis with a MALDI-TOF (matrix assisted laser desorption ionization-time-of-flight) system helped explore the topology of a yeast protein sub-complex made up of six proteins. This study, being one of the first of its kind, helped highlight the inherent difficulties that are present in a technique even as “simple” as crosslinking. This method allowed for a low-resolution topological view of the pore complex giving information on neighboring proteins allowing a model to be built. All of this was possible from protein amounts on the femtomole scale. Rappsilber’s findings have held up well and have developed into a technique that has become just as complex as its sister techniques of x-ray crystallography and NMR. The development of mass spectrometry technology, protein purification/analysis, and crosslinker chemistry has brought the idea of crosslinking from a static view of simple complexes and identifying which members neighbor one another to a more dynamic technique. Proteins and the changes that occur due to mutations and protein partner binding can now be viewed, even helping reshape some crystal/NMR structures that were once viewed as complete.
Figure 1.6.1: The non-cleavable cross-linker DSS which is amine specific; BS3 is created with the addition of a sulfo-group to each end of the linker.
Jumping ahead 10 years and not only are protein structures being studied, but the biological implications of those structures and their mutants are being interrogated. All of this is being done in a more physiological relevant way than was previously allowed by techniques such as NMR or crystallography. The study of large, flexible proteins leads to a combination that cannot be tackled by current crystallography or NMR to a lesser extent. This can create a large gap in knowledge as conformational changes in proteins can play key roles in many biological processes (Sriswasdi et al. 2014). Sriswasi et al. examined spectrin, a flexible protein that is a perfect candidate for crosslinking due to its size (1,052 kDa) and its flexible structure not allowing for crystallization of four of its domains. This protein plays key roles in cell membrane shape and mutations are common in diseases such as hereditary elliptocytosis (HE). Its existence as either an open/closed-dimer or tetramer exemplifies the difficulties present in crystallizing the full-length protein complex. The crosslinking method developed by Rappsilber et al. remained almost unchanged except for the addition of concentration/clean up steps showing universality of crosslinking proteins (Rappsibler et al. 2000). The spectrin protein has a highly flexible nature that introduces larger possible error in crosslink identification. Use of a zero length crosslinker, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) shown in Figure 1.6.2, was implemented to reduce this error. These types of crosslinkers, which have been used since the development of the technique, for example formaldehyde, do not contain a spacer arm causing crosslinks to be in the distance range of a salt bridge. Zero-length linkers allow for more confident peptide-
Figure 1.6.2: Zero length cross-linker EDC used commonly in conjunction with NHS-esters to link primary amines.
peptide crosslink identifications. Fewer crosslinks are needed when refining the molecular model when using zero length crosslinkers (Leitner et al. 2010).

Crosslinker type becomes important in the investigation of flexible proteins as longer spacer arms could create crosslinks between portions of the protein that are not necessarily interacting but are in close proximity due to protein movement. This experiment introduces the idea of homology modeling to help build upon the information derived from previous sequences to build a structure. Using hundreds of computer generated models and the root-mean-square-deviation (RMSD) values on top of the crosslinks available, models of best fit can be chosen that closely agree with the data. The study by Sriswasi et al. allowed for a refinement of the previous complex model giving a clearer view of the spectrin protein (Sriswasdi et al. 2014). The detrimental effect a single mutation can have on protein structure was also observed, showing the importance of structural characterization of mutated proteins. The spectrin study shows an example of the development of crosslinking mass spectrometry (XL-MS) as a technique and how it has come into its own as a tool for structural analysis. Development of this technique has allowed it to stand alone for structural studies but it still lends itself well to refining models. The work by Sriswasi et al. shows crosslinking and its maturation into a viable structural technique using the ideas developed by Rappsilber et al. (Seiswasdi et al. 2014; Rappsilber et al. 2000). The differences arise in the use of different crosslinkers all with their own unique chemistries. It has been stated that crosslinking alone only provides low-resolution protein structural information and cannot be used solely to determine a
protein’s structure. Getting past this can be simple if crystallography, NMR or homology modeling data has already been provided.

The work done on the spectrin protein by Sriswasi et al. is a perfect example of this combination of techniques as it uses previous crystallography data to create a model using a homology modeling program (Sriswasdi et al. 2014). The crosslinking data is then overlaid on top of that to further refine the model. The possibility of performing hydrogen/deuterium exchange (HDX) in conjunction with crosslinking can also give important information on already crystallized proteins. A 14-3-3 human protein homodimer interface was explored using HDX along with crosslinking with the EDC crosslinker which activates carboxyl groups to react with primary amines (Haladova et al. 2012). The HDX data allowed for protected portions of the proteins in the complex to be determined and to orient them correctly, and use of a zero length crosslinker helped identify salt bridges present in the dimer.

Arrival of electrospray ionization, a soft ionization technique, allows for the ionization of protein complexes keeping them intact and possibly observing which proteins are interacting (Loo et al. 1993). After protein complexes have been identified along with the number of proteins participating in the complex crosslinking can be applied to narrow down the portions of the protein faces involved in the binding events. Many molecular models can benefit from further resolution, as no single technique is perfect. Crosslinking has shown its strength in helping to resolve protein complexes that are naturally difficult to work with using other structural techniques by providing low-resolution data to help characterize proteins structurally.
1.6.1 Crosslinker Types

Development of novel crosslinkers is an involved process that can involve various reactive ends and spacer regions with different lengths and modifications that allow for specific chemistry. Starting with the types of reactive ends that can be attached to a crosslinker to give the residue specificity and reactivity is an important step in synthesis. This becomes important when working with proteins that require a specific pH range to stay in solution since not all crosslink ends are compatible in all pH ranges. For example, N-hydroxysuccinimide esters are stable compounds when kept in a cool, dry area (Simmerman 1964). When introduced to an aqueous buffer they begin to hydrolyze with a half life of 8 hours at neutral pH and 1 minute in a basic pH ~9 (Besselink et al. 1993). This amine specific group can react with lysines and protein N-termini. When performed at a neutral pH roughly 1% of amino groups are deprotonated allowing for a more controlled crosslinking reaction without over modifying the protein (Paramelle et al. 2013).

Besides NHS esters there are groups that react with other numerous residues in other pH ranges: maleimide (pH 6.5-7, C), thiopyridyl disulfide (pH 7.5, C), dicarbonyl (pH 7, R, K), aldehyde/carbonyl group (pH 5, unnatural residue containing carbonyl group) and numerous others (Paramelle et al. 2013). Other than the reactive ends, all crosslinkers either contain some spacer region of a specified length between them or utilize a zero length crosslinker that does not become part of the crosslinked peptides after the reaction. The spacer region allows for the addition of modifications to further simplify and identify crosslinked peptides from a complex
mass spectrum. A popular method used is the introduction of isotopes $^2\text{H}$ or $^{13}\text{C}$, and crosslinking a protein with 1:1 ratio of light to heavy versions of the linker (Muller et al. 2001). This results, theoretically, in peptides of equal intensity separated by a mass difference related to the number of heavy atoms present in the spacer arm.

Spacers can contain an ion reporter tag that causes the mass spectrometer to dissociate precursor ions and perform further tandem mass spectrometry (MS/MS) analysis on those containing the ion tag (Koning & Jong 2001). Addition of affinity groups to the spacer region can also allow for the simplification of complex crosslinked samples. An example of this is the addition of a biotin group that allows for the conjugation of modified peptides to affinity columns and the removal of contaminating peptides (Kang et al. 2009). Crosslinkers with cleavable spacer regions are the primary types used for the protein complex studies I have performed over the past two years. Originally developed by Soderblom and Goshe, the SuDP crosslinker shown in Figure 1.6.1.1 allows for unequivocal identification of inter- and intra-peptide crosslinks based on the mass difference of the modifications added to the reacted lysine residues (Soderblom & Goshe 2006). These crosslinkers work by utilizing a more labile bond that is preferentially cleaved during a collision-induced dissociation/high energy collision-induced dissociation (CID/HCD) event over the typical amide bond present in the peptides. Individual identification of the peptides is helped by the separation of the two during the dissociation event. Over the years groups have further developed these types of crosslinkers and even created other types of novel crosslinkers.
**Figure 1.6.1.1:** Cleavable cross-linker using the labile bond created between the Asp/Pro residues. Amine reactive ends on both sides. Preferentially cleaved over peptide bonds during a CID event. Similar length as the DSS/BS3 cross-linker (Liu & Goshe 2010).
1.6.2 Analysis of XL-MS Data

After crosslinking and subsequent mass spectrometric analysis of the crosslinked protein has been performed post-acquisition analysis and identification of the crosslinked peptides must be performed. This portion of the experiment may be one of the most challenging aspects of XL-MS as can be seen in Figure 1.6.2.1. The non-trivial task of deconvoluting a mass spectrum and identifying crosslinks requires specialized programs created by bioinformaticians. Knowledge of programming along with a basic understanding of the chemistry and the technology behind mass spectrometry is needed for program development. Numerous programs exist that identify crosslinks based on different methods such as isotopic labeling, CID cleavage linkers, or reporter ions. Scoring is also unique to each program, utilizing custom algorithms to assign whether or not the crosslink is real or just an artifact of misinterpretation of complex spectrum. Programs such as SearchXLinks, X!Link, MS2 Assign all use non-probabilistic means of scoring relying on the number of peaks assigned, the percentage of peaks assigned, or the number of experimental peaks assigned. Probabilistic scoring utilizes a number of probability functions to assign scores (Zhang et al. 2002).

A program that fits the needs of a particular experiment has to be used to obtain crosslinks that make sense when trying to fit them to a model. The programs used during my time include CXLink, which was developed by Fan Liu in the Goshe lab and utilizes Matlab as the program language. The difference between the masses added to each lysine after being modified by the crosslinker is used to
Figure 1.6.2.1: Work Flow of a non-cleavable crosslink identification program. Usually follows a typical peptide identification pathway with differences coming before identification. All possible peptides are created from a theoretical digest along with all possible crosslinked peptides from the digest in respect to the specific crosslinker used. Identified peaks are compared to the generated peptide list and scored (Havlíš & Shevchenko 2004).
identify crosslinked peptides created with a gas cleavable crosslink. This program is still in its infancy so the programs StavroX and MeroX for non-cleavable and cleavable crosslink identification respectively, were used in my studies (Gotze et al. 2012; Gotze et al. 2014). A basic flow through of the StavroX/MeroX program is shown in Figure 1.6.2.1. The development of crosslinking into the structural technique it is today, allowing for elucidation of protein complexes not accessible to crystallography or NMR, has provided the XL-MS platform for my work to characterize the cytoplasmic domain of the BRI1/BAK1 heterodimer. These proteins have been crystallized individually but the structure of the complex still eludes researchers. Development of a functional heterodimer model between these two proteins would be valuable to understand the mechanism of phosphorylation between the proteins and phosphorylation of other substrates and is the focus of my Masters thesis research.

1.7 Plant Receptor-Like Protein Kinases (PRKs)

Cell-surface receptors, similar to those in Figure 1.7.1, are a common feature among animal and plant cells and often contain conserved protein features (The Arabidopsis Initiative 2000). The transmembrane kinases known as receptor-like kinases (RLK) represent a large protein family with at least 610 members encoded in the Arabidopsis thaliana genome (Shiu & Bleecker 2001). The RLK family of proteins, specifically those with an extra-cellular receptor, is of particular interest. Their involvement in numerous developmental processes from organ elongation
Figure 1.7.1: Diagram of typical cell surface interaction with peptides/hormones/pathogens etc and the first steps in translating that signal to the cell. As adapted from (Cock et al. 2002).
(ERECTA) and epidermal specification (CRINKLY4) is the driving force behind this interest (Torii et al. 2000).

An RLK has an extracellular domain, single membrane-spanning domain and the cytosolic kinase domain (The Arabidopsis Intitiative 2000). The kinase domains are conserved among different family members while the extracellular domains exhibit more diversity. The utilization of different protein domains in the extracellular domain allows for different signaling pathways even with structurally similar kinase domains. A few unique domains include S domains, epidermal growth factor (EGF) repeat containing domains, lectin domains and leucine-rich repeats (LRR). Proteins with LRR domains are the focus of my study. More than 230 A. thaliana genes encode LRR proteins, illustrating the importance of characterizing the structures (Shiu & Bleecker 2001).

While evolving independently from animal receptor kinases LRR proteins share many properties such as associating with other proteins to create large signaling complexes. One example represented in Figure 1.7.2 is CLAVATA1 (CLV1), that can be present in a 185 kDa and 450 kDa complex (Trotochaud et al. 1999). It is thought to be composed of an intricate grouping of proteins including two disulfide-linked CLV1-CLV2 dimers plus a CLV3 protein along with other associated proteins such as kinase-associated protein phosphatase (KAPP) and Rop (Rho/Rac-GTPase-related protein) (Jeong et al. 1999). Aside from receptor oligomerization it has been observed that auto- and transphosphorylation events occur to facilitate the activation of the signal transduction pathway. Phosphorylation events specifically in
Figure 1.7.2: Representation of the Clavata signaling pathway. A tetramer composed of two CLV1/CLV2 dimers and CLV3. As adapted from (Jeong et al. 1999).
the BRI1/BAK1 signaling pathway show the importance of trans- and auto-phosphorylation. A plethora of phosphorylation sites (auto- and trans-) were identified with the use of mass spectrometry. Certain residues and their respective phosphorylation are required for complete kinase activity. Mutations in BRI1 including T-1049-A and either S-1044-A or T-1045-A of the activation loop caused complete loss of activity. Mutations T-1039-A and S-1042-A caused a larger effect on substrate (trans-) phosphorylation activity. Interestingly, mutations of residues such as S-838 and T-872 in the juxta membrane domain had little effect on auto-phosphorylation but caused a dramatic decrease in phosphorylation of the BRI3 peptide used to test trans-phosphorylation activity. Similar phenomenon were observed when residues S-1168 and S-1172 in the C-terminal tail of BRI1. The data discussed shows an agreement with the idea that auto-phosphorylation plays an important role in kinase activity along with the possible creation of docking sites for other proteins required for signaling (Wang et al. 2005). The animal receptor kinase family has proteins that act downstream (mitogen activated protein kinase and RAS) to enhance the signal from a ligand bound to the receptor protein. In plants, the Rop protein is thought to activate a downstream MAPK cascade that decreases WUSCHEL expression, a protein known to help in cell differentiation during plant development (Clark 2001).

Regulation of protein receptor kinases (PRKs) contains some parallels with animal receptors in regards to down-regulation. Mechanisms such as kinase inhibitors, phosphatases and endocytosis exist to control these pathways. The
phosphatase mentioned earlier, KAPP, has been shown to interact with and
dephosphorylate the kinase domain of CLV1 and act as a negative regulator
(Williams et al. 1997). Endocytosis mediated by really interesting new gene (RING)
finger proteins has been shown to be a complicated process to understand in plants.
The RING finger protein Cbl in animal cells shares characteristics with the plant
protein ARC1, in that it only interacts with the phosphorylated form of the kinase
domain to cause ubiquitylation and subsequent endocytosis of the protein. The
ARC1 activity becomes difficult to understand after seeing that ARC1 can positively
regulate signal transduction proteins but it has been shown that the animal protein,
Cbl, has both negative and positive roles in its respective pathway (Thien & Langdon
2001). This brief introduction into PRKs shows a large gap in knowledge of plant
pathways including the function they play further downstream.

1.8 Leucine-Rich Repeat Receptor-like Kinases (LRR-RLKs) in A. thaliana

The largest subfamily of trans-membrane kinases in Arabidopsis thaliana is
the LRR-RLK family. These kinases regulate a variety of metabolic processes
ranging from cell proliferation, stem cell maintenance, wound response and
symbiosis. Binding of ligands and translation of that signal throughout the cell is not
fully understood but models such as Figure 1.8.1 are being developed to fully realize
their importance. These proteins are thought to be able to exist as dimers and to
form higher order complexes with other proteins in the LRR-RLK family. The majority
of the kinases act predominantly on serine and threonine residues as opposed to
Figure 1.8.1: The recycling of integral membrane proteins after binding to its respective signal and being activated (Tien & Langdon 2001).
tyrosine residues in animal systems although tyrosine phosphorylation does occur and is important in plants. Mutation of LRR-RLKs that result in loss-of-function show plant phenotypes providing evidence of these proteins participate in development, steroid-hormone responses and disease resistance (Shanmugam 2005; Bishop & Koncz 2002; Jones & Jones 1997). The BRI1/BAK1 signaling pathway is a well studied model that has been shown to exist as homo and heterodimers, phosphorylate serine, threonine as well as tyrosine residues and has been highly mutated to show its importance in plant growth. The plant steroid brassinolide (BL) is a ligand of the BRI1 extracellular domain and is involved with leaf and vascular morphology, senescence and flowering. BL binding allows the association of the BAK1, member of the SERK subfamily of LRR-RLKs, protein known to participate in transphosphorylation of BRI1 fully activating the kinase activity of BRI1. Bak-1 null alleles have shown plant phenotypes similar to a weak bri-1 null phenotype. This phenomenon indicates either a functional redundancy amongst the SERK family of proteins or the ability for BRI1 to utilize a basal level kinase activity to translate the signal across the membrane without associating with a co-receptor such as BAK1. The kinase dead mutant BRI1, K911E, destroyed BL enhanced binding to BAK1 while the kinase dead BAK1, K317E, still showed the ability to associate with BRI1. Due to the functional redundancy of the SERK family the full effects of the bak-1 null mutant could not be observed without further mutating other SERK family members. The importance of kinase activity in regards to BRI1/BAK1 oligimerization has been characterized. When the full-length cytoplasmic domain of BRI1 (BRI1-CD) and
BAK1 (BAK1-CD) are incubated together a drastic increase of peptide substrate phosphorylation is observed; basal levels of peptide phosphorylation were observed when incubated with a kinase dead mutant of BAK1. Pre-incubating BRI1-CD with ATP and BAK1-CD and subsequently removing BAK1-CD, an increase in substrate phosphorylation was observed which was not seen when pre-incubating with a kinase dead BAK1-CD. Mutations of activation loop residues T455 in BAK1 and T1049 in BRI1 cause a significant loss of kinase activity even when mutated to the phosphomimetic residue D. This illustrates the importance of trans-phosphorylation in activation of BRI1 and BAK1 kinase activity. The transgenic lines expressing mutated BRI1 and BAK1 resulting in dramatic morphological changes show the importance of auto- and trans-phosphorylation in signal translation across plasma membranes (Wang et al. 2005; Wang et al. 2008).

The LRR structure is composed of 24 amino acids with conserved leucine residues. The structure revealed that part of the LRR forms a -sheet/-turn motif and when multiple strings of these are attached in tandem a protein binding surface is created (Kobe & Deisnehofer 1993). The general sequence of the LRR is as follows: L--L--L--L--L--N-L-G-IP--; with each dash representing a non-conserved amino acid in the sequence. The non-conserved residues determine the specificity of interactions with other proteins. The number of LRRs in tandem range from 1 to 32 in Arabidopsis proteins. As seen from Figure 1.8.2 the general structure of an LRR-RLK is conserved with LRR groups followed by a transmembrane region and ending with a kinase domain. The binding of specific ligands that recognize certain
Figure 1.8.2: Representations of multiple LRRs that differ slightly in structure depending on the presence of characteristic domains (Torii 2004).
sequence/structural motifs leads to signaling by the correct protein. Plants also encode for proteins referred to as Leucine-rich Repeat Receptor Proteins (LRR-RPs) that have no cytoplasmic kinase domain. They may also play a role in plant development by acting as coreceptors for LRR-RLKs by interacting through their respective extracellular domains.

Specific LRR-RLKs called *BRASSINOSTEROID INSENSITIVE 1 (BRI1)* and *BRI1-ASSOCIATED RECEPTOR KINASE 1 (BAK1)* are involved with steroid hormone signal transduction, in particular the brassinosteroids (BRs) which promote growth in all plants (Torii 2004). The active BR, brassinolide, can activate the BRI1 signaling pathway at nanomolar concentrations. The *BRI1* protein is composed of 1196 amino acids and 25 leucine repeats with an island domain for BR binding (Li & Chory 1997). This protein is very interesting when considering it is a steroid receptor protein that is localized at the plasma membrane. The majority of animal steroid receptors are ligand-activated transcription factors that are later localized to the nucleus. Further studies into plant steroid recognition could lead to the possible discovery and understanding of animal cell-surface steroid binding proteins. LRR-RLKs, including BRI1, also contain paired cysteines that exhibit some functional importance in extracellular domains of certain proteins. These sequence motifs usually consist of two cysteine residues spaced by 6 amino acid residues before the first LRR and immediately after the last LRR. Though their function is not fully understood, point mutations of these cysteines can result in weaker signaling as observed for the cysteine-to-tyrosine mutation in BRI1 (Friedrichesen et al. 2000).
For the CLV1 signaling system the cysteine residues may be implicated in the formation of a dimer (Jeong et al. 1999). Other proteins show no negative response to removal of the cysteine residues, for example the CLV2 maize ortholog, FEA2. This shows that some of the conserved cysteine residues may not be required for activity (Taguchi-Shiobara et al. 2001).

A unique domain possessed by some LRR-RLKs is a binding domain that recognizes steroid/peptide signals outside of the cell. A specific example is the BRI1 protein that contains a 70-amino acid island domain between LRR-21 and LRR-22 (Li & Chory 1997). This domain binds to the plant steroid brassinolide without the help of any steroid binding proteins adding another unique trait to this particular protein (Kinoshita et al. 2005). This structural characteristic of LRR-RLKs is seen as unique because of the LRR domain predominantly being used as a protein-protein interaction site. The co-crystal structure of the FLS2 and BAK1 ectodomains showed interaction between the LRRs of both domains (Sun et al. 2013). However the interaction of the cytoplasmic domains is not known for the FLS2/BAK1 LRR-RLK pair. Thus, the main focus of this research is to characterize the interaction of the cytoplasmic domains of both BRI1 and BAK1 using chemical crosslinking and mass spectrometry to elucidate the structure of the heterodimer that forms upon binding of this pair of proteins.
1.9 BRASSINOSTEROID INSENSITIVE 1 and BRI1-ASSOCIATED RECEPTOR KINASE 1

The kinase domain structures of both BRI1 and BAK1 have been solved using x-ray crystallography, but the heterodimer has not been able to be crystallized (Bojar et al. 2014; Yan et al. 2012). The use of chemical crosslinking and mass spectrometry characterization of the dimer has become a possibility in a biologically relevant experiment. The kinase structure of BRI1 shows an interesting dual specificity kinase activity not usually observed in other receptor kinases. Differential phosphorylation of the activation loop of BRI1 allows it to act as both a Ser/Thr and Tyr kinase. The active loop conformation and phosphorylation sites are specifically conserved among other plant protein kinases also showing possible dual specificity amongst all plant kinases (Oh et al. 2009). BRI1 binds to BRI1 Kinase Inhibitor 1 (BKI1) through its C terminal lobe. Binding of this protein precludes any possible binding by co-receptors such as somatic embryogenesis receptor kinase 1 or 3, suggestive of the presence of a possible protein-docking site present in the C terminal end of the BRI1 kinase domain (Bojar et al. 2014). When looking at the structure of the kinase domains of both BRI1 and BAK1 in Figure 1.9.1 a striking similarity between them can be seen. Hence, it is not surprising that these proteins can create hetero and homo-dimers. It is speculated that the extracellular domains, whose structural differences are visualized in Figure 1.9.2, facilitate heterodimerization of these two proteins in the presence of BL (Bojar et al. 2014). The C-terminal portions of both kinases have been shown to have some regulatory activity
Figure 1.9.1: Ribbon diagrams of BR11 (left) and BAK1 (right) showing the homology present between the two kinase domains. The images were generated using PYMOL 1.7.4 (www.pymol.org) and PDB 3ULZ and 4OA6.
Figure: 1.9.2: Ectodomains of BRI1 (right) and BAK1 (lower left) crystallized as a heterodimer in the presence of brassinolide. The images were generated using PYMOL 1.7.4 (www.pymol.org) and PDB 3RGZ and 4MN8.
as well. In the absence of BAK1, BRI1 can form homo-dimers bringing the C-terminal tails in close proximity, which is believed to keep them in an inactive state. The phosphorylation sites in the CT tail have been mapped to, S1160 and T1180, and mutation of these residues to the phosphomimetic residue D induce more activity in BRI1 (Wang et al. 2005). Similar traits have been observed for other receptor tyrosine kinases where the CT plays a negative role in activity (Niu et al. 2002). The CT of BAK1 has recently been shown to have activity of its own and may have both negative and positive regulatory roles in protein activity (Oh et al. 2014). The CT of BAK1, when truncated, enhanced trans-phosphorylation in situ but in vitro studies showed phosphorylation was severely down-regulated. This suggests that the CT is needed for correct positioning of the kinase relative to BRI1, at least in vitro. A synthetic peptide corresponding to CT residues, 576-615, binds to BRI1. Phosphorylation of the Y610 residue showed increased binding but this could not be attributed simply to charge because phosphorylated S612 did not show increased binding activity (Oh et al. 2014). The data may show the dual specificity nature that is present in the BRI1 kinase domain for Ser/Thr and Tyr residues as BRI1 was able to phosphorylate all three residues, the reasoning behind the dual specificity remains to be understood. BRI1 is known to be active in plant cell differentiation and mutations in this protein causes characteristic deficiencies that could not be rescued by brassinolide treatment. BAK1 has been shown to be more promiscuous, interacting with numerous proteins to participate in immunity response including perception of pathogens (Kemmerling & Nurnberger 2008). Plants with mutated BRI1
were characterized as having a severe dwarf stature, dark green, thickened leaves and reduced apical dominance (Clouse et al. 1996). These mutants show the invaluable nature of BRI1 in plant development.

The kinase activity of BRI1 is known to play a role in its ability to produce a signal and cause plant development. It has been shown that this activity is not necessarily needed and plants expressing loss-of-activity mutants do not always show severe phenotypes; this may be due to the fact that the kinase domain is not directly involved in ligand binding or because there are co-receptors that play specific roles in the signal transduction (Xu et al. 2008; Sun et al. 2013). The role of BAK1 as BRI1 co-receptor has been heavily studied and shown to be important in the activation of BRI1 activity and signal propagation through the formation of a protein-protein heterodimer (Wang et al. 2008; Bucheri et al. 2013; He et al. 2007). The current model of signal propagation describes sequential transphosphorylation events between BRI1/BAK1 to fully realize the BR signal and initiate signal transduction. Binding of the BR signal by the extracellular domain of BRI1 allows for the transphosphorylation of BAK1 and the subsequent phosphorylation of BRI1 by BAK1. Phosphorylation of residues S1044 and T1049 in the BRI1 activation loop is necessary for full activation of its kinase activity. Expression of a bak1-4 null allele results in a weak phenotype similar of a weak bri1 allele showing dimerization of the proteins is required but that binding redundancy, such as SERK-4 which is known to bind BRI1 as well, is also present. Introduction of a bri1-1 mutant shows no dimerization resulting in no transphosphorylation of either BAK1 or BRI1. The bak1-4
null mutant showed dimerization still occurred but the binding redundancy between BAK1 and SERK-4 makes it difficult to fully observe the effects of decreased phosphorylation of BRI1 but it is seen that binding can still occur between BRI1/bak1-4 but to a smaller extent (Wang et al. 2008).

Although the importance of BRI1/BAK1 heterodimerization has been explored extensively, it still needs further refining. Studies utilizing fluorescing tags tracked the endocytosis of BRI1 and BAK1 as either single proteins or together in a complex. The endosomes containing each respective protein or complex can be used to recycle the proteins to the plasma membrane or even be used in further activation of the signal pathway (Bucherl et al. 2013). Mammalian receptors have been shown to follow differing endocytotic pathways to possibly help in signal transduction that utilize lipid rafts. If this scenario could apply to BRI1/BAK1 then endosomes housing a BR-bound-dimer could be sorted through different pathways to be used in signaling. Examining this possibility will require more work, as endosome pathways in plants have not been as well studied as mammalian systems. The current prevailing model behind BRI1/BAK1 binding and subsequent signaling is shown in Figure 1.9.3. The model depicts binding of the BR signal, releasing the BKI1 inhibitor from BRI1 allowing for BAK1 and sequential transphosphorylation between the two proteins propagating the signal. This simplistic model becomes more convoluted when taking into consideration the regulatory systems built into the proteins. The majority of the work done in regards to crosslinking before my project focused mainly on XL-MS method development, e.g. creating the crosslinker and a program to
Figure 1.9.3: Prevailing model of the BR11/BAK1 sequential transphosphorylation events that must occur for full protein activity. As adapted from (Wang et al. 2008).
identify crosslinks robust enough to handle more complex biological samples.

Shifting the work towards proteins with incomplete structural data, including proteins with partial crystallography data, has shown the inherent difficulty of working with truncated membrane proteins as activity and structural integrity can be compromised. Many of the binding events that have been explored are only theoretical with little to no evidence on how exactly these BRI1/BAK1 could come together in a stable heterodimer. Combining XL-MS and molecular modeling has allowed me to develop a model in which these proteins interact in an energetically favorable manner, to provide a more dynamic view and a deeper look into the structure-functional interaction between BRI1 and BAK1.
Chapter 2: Methods

2.1 Synthesis of Bisuccinimidyl-succinamyl-aspartyl-proline (SuDP)

Fmoc-Pro-NovaSyn TGT (http://www.emdmillipore.com) extremely acid-sensitive preloaded resin was used to synthesize the linker region of the SuDP crosslinking reagent. Fmoc-Asp(OtBu)-OH was coupled to the resin-bound amino acid in a batch wise manner using the active ester method. Briefly, following deprotection of the resin-bound amino acid, active esters were formed using Fmoc-amino acid, N-hydroxybenzotriazole, and diisopropylcarbodiimide (www.sigmaaldrich.com) in a 5:5:5.5 molar ratio to the resin-bound amino acid. Following Fmoc deprotection of Fmoc-Asp(OtBu)-OH, succinic anhydride was added to introduce a carboxylate group to the N-terminus of the reagent linker region. The completed linker region was cleaved from the resin with 1% trifluoroacetic acid (TFA) (www.sigmaaldrich.com) in dichloromethane. Following reverse phase high-performance liquid chromatography (rpHPLC) purification, esterification of the dicarboxylic acid was performed using di(N-succinimidyl) carbonate. The esterified product was HPLC purified, dried under vacuum, and stored at -20 °C. Before use, the aspartyl side chain tert-butyl protecting group was deprotected by addition of 90% TFA/5% water/5% DMSO. The mixture was stirred at ambient temperature for 10 min, then dried under vacuum. The linker was then subjected to rpHPLC purification again to remove deprotection contaminants, dried and stored at -80 °C.
2.2 Protein Expression and Purification

Flag-BRI1 (residues 818-1196) and GST-BAK1 (residues 250-615) plasmids were both transformed into BL21 cells per the vendor’s (Promega) instructions for expression in LB media at 37 °C (Wang et al. 2008). The overnight culture was spiked into fresh LB at a 1:10 dilution and grown at 30°C until OD\textsubscript{600} is ~0.7, then the temperature was dropped to 16 °C and expression induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 8 hr. Cells were harvested at 5,000 rpm, 4 °C for 5 min and resuspended in lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 5 mM MgCl\textsubscript{2}, 0.25 mg/mL lysozyme, 0.1 mg/mL DNase, 1 mM PMSF, Complete protease inhibitor tablet, at pH 8) and subjected to gentle agitation at RT for ~30 min then sonicated. The resulting lysate was clarified by spinning at 10,000 rpm, 4 °C for 30 min and filtered through a 0.22 micron filter. The cleared \textit{E. coli} lysate was incubated with M2 anti-Flag resin and glutathione resin and eluted with Flag peptide and glutathione respectively. Samples to be used for crosslinking were dialyzed into working buffer (50 mM HEPES, 150 NaCl, 5 mM MgCl\textsubscript{2} at pH 8).

2.3 Crosslinking and Proteolytic Digestion

Flag-BRI1 and GST-BAK1 proteins in a stock solution (50 mM Tris-HCl, 150 mM NaCl, 5 mM MgCl\textsubscript{2}, 0.25 mg/mL lysozyme, 0.1 mg/mL DNase, 1 mM PMSF, Complete protease inhibitor tablet, at pH 8) were dialyzed for a minimum of 1 h at room temperature (RT) into working buffer. Aliquots of SuDP were dissolved in DMSO and added to the protein solutions to give protein-to-crosslinker molar ratios...
of 1:100, 1:200 and 1:300 keeping the final organic concentration below 1% total volume. Crosslinking was performed at RT with gentle agitation for 1 hr and subsequently quenched for 30 min with 1 M Tris-HCl, pH 7.5 brought to a final concentration of 10 mM. Samples were then analyzed on SDS-PAGE using a 4-20% Bis-Tris gel and proteins visualized with Coomasie stain. Based on the analysis, protein-to-crosslinker molar ratios of 1:200 had bands that were excised from the gel and subjected to in-gel digestion per vendors instructions (thermoscientific).

2.4 Liquid Chromatography-Tandem Mass Spectrometry Analysis

Crosslinked peptide samples were separated using an Easy-nLC system (www.thermoscientific.com) coupled online to an Orbitrap Elite mass spectrometer (www.thermoscientific.com). Reversed-phase separation of the peptides from the protein digest was accomplished using a 75 µm i.d. × 30 cm column packed in-house with 3 µm 200 Å Magic C18AQ stationary phase (Michrom Bioresources Inc., www.michrom.com) coupled to an Acclaim PepMap 100 µm i.d. × 2 cm trap column (www.thermoscientific.com). The mobile phases consisted of (A) 0.1% formic acid in 2% acetonitrile (www.sigmaaldrich.com) and (B) 0.1% formic acid in acetonitrile. A total of 100nmol of peptide digest was loaded onto the reversed-phase trap column for each injection and then separated using a linear gradient of 1% B/min from 5% B to 40% B at a flow rate of 300 nl/min. HCD was performed on the top three intense ions of each full MS scan.
2.5 Data Analysis

Mascot general files were generated from raw data using Proteomic Discoverer (www.thermoscientific) and analyzed by MeroX and StavroX (Gotze et al. 2012; Gotze et al. 2014). These are crosslink identifier programs developed by the Michael Gotze lab at the University of Halle-Wittenberg. Analysis of traditional crosslinkers i.e. BS3 using StavroX required the selection of the crosslinker used and the specifics of the MS experiment such as precursor and fragment ion precision. MeroX for cleavable linkers requires a similar procedure but the data must be ran twice so the modifications of the crosslinker can be added to both peptides of the interpeptide crosslinked species.

2.6 Strong Cation Exchange Online Separation

The LC front end was modified by the addition of an in-house packed strong cation exchange (SCX) 75 μm i.d. × 4 cm packed pre-column with 5 μm 200 Å PolySULFOethyl A stationary phase placed before trap column. The mobile phase conditions were kept the same as previously stated in section 2.4. After initial sample injection subsequent injections of mobile phase (A) at pH 2.0 with varying concentrations of KCl or ammonium formate were loaded: 0, 25, 35, 50, 60, 75, 100, 200 and 500 mM. The various salt concentrations cause a population of peptides to elute from the SCX column and bind to the reverse phase trap for analysis using the method in section 2.4.
2.7 Immunoblotting

Immunoblotting was performed per Abcam (Cambridge, MA) instructions. Briefly, after running samples on a Novex Tris-Glycine 4-12% pre-cast gel the proteins were transferred to a PVDF membrane using an iBlot Dry Blotting System. The membrane was blocked with 5% nonfat milk in tris buffered saline (TBS) (20 mM Tris, 150 mM NaCl, pH 7) for 1 h then washed four times in TBS for 5 min each. A 1:1000 dilution of primary antibody was made with TBS and incubated for one hour at RT or overnight at 4 °C and then washed again for 5 minutes four times. Secondary antibody was prepared at 1:5000 dilution in TBS and incubated for 45 min at RT and washed with TBS as done earlier. To visualize the bands enhanced chemiluminescence (ECL) was used exposing film to the membrane.

2.8 Plasmid Transformation

Plasmid transformation was performed per Qiagen (Santa Clarita, CA) instructions. Briefly, a 5 mL overnight culture of E. coli was grown and DNA was extracted using the Qiagen DNA mini kit. The DNA was diluted if needed to obtain appropriate concentrations for transformation. The DNA was mixed with the competent cells (BL21 and DH5α) and incubated on ice for 30 min then heat shocked for 45 seconds without shaking and placed on ice. LB media was added and the mixture incubated at 37 °C for 1 h. Aliquots of the competent cells were placed on LB agar plates with the appropriate antibiotic and grown overnight at 37 °C. Single colonies were picked from the plate and grown to be used for stock glycerol solutions.
2.9 HADDOCK Modeling Simulation

After loading the crystal structure PDB files a subset of interpeptide crosslinks identified were chosen and used to simulate protein docking. These subsets of crosslinks were chosen based on whether they broke the maximum distance allowed by the linker region and whether they made the most biological sense. 1000 structures were created using the constraints given by the crosslinks and the best 200 were chosen and clustered based off the RMSD values obtained (Dominguex et al. 2003).
Chapter 3: Results and Discussion

3.1 Strong Cation Exchange Separation

The low population of crosslinked peptides present after proteolytic digestion makes identification difficult. This is due to the limitation of chromatographic resolution of the eluting peaks from the LC when trying to separate analytes from a highly complex mixture. To alleviate this problem separation methods and program improvements have been made to increase the observable number of crosslinks and lower the false-discovery rate of identification. The use of a gel-based strategy to help separate components of a protein complex based on size helps enrich for higher molecular weight species that are more likely to produce peptides that are part of a crosslinked pair. This is simple and only requires the addition of an in gel digestion to extract the peptides from the gel. A drawback to this technique is the increased concentration of trypsin that must be used to obtain the same quality of digestion that in-solution delivers (Shevchenko et al. 2006). This can cause a high intensity of autolysis products of trypsin and must be accounted for due to possible false peptide discovery. The loss of peptides is also a possibility when subsequent steps that require sample handling are added to the procedure. It has been shown that 70-90% of peptides can be recovered with this method but the actual yield is largely peptide dependent (Havlis & Shevchenko 2004).

As mentioned Chapter 1 an affinity tag in the spacer region of the crosslink can be utilized. These tags can help with the extraction of peptides from complex samples but are extremely difficult to synthesize and are not readily available
(Leitner et al. 2010). Though these types of crosslinkers do allow for the extraction of crosslinked peptides they do not discriminate against dead-end, intra-peptide, or inter-peptide crosslinked products, with the former two lending little to no information about protein structures besides knowing what residues may be more or less solvent accessible (Zhang et al. 2009).

Size exclusion chromatography (SEC) and strong cation exchange (SCX) chromatography that can enrich for crosslinked peptides based on size and charge respectively, have been utilized off- and on-line to create a 2-D separation. Size-exclusion takes advantage of the increased molecular weight of peptides that have been crosslinked. Many peptides resulting from digestion are on the order of 2kDa or lower and this large differentiation in molecular weight after being crosslinked to another peptide allows for separation even by a crude method such as SEC. Analytical challenges can still remain from this type of separation as larger peptides do not behave “normally” during fragmentation events and begin to deviate from the predicted model resulting in more random cleavages and incomplete fragmentation. This results in the need for a more complex digestion protocol introducing enzymes such as Asp-N, Glu-C and Lys-C in conjunction with trypsin to yield peptides of an optimum length for better fragmentation, thus creating the requirement for further method optimization (Havlis & Shevchenko 2004). The use of SCX allows for the separation of peptides based on charge using resins such as shown in Figure 3.1.1. This technique, like SEC, can be done off- or on-line. A typical tryptic protein digestion yields peptides with a +2 charge state if no missed cleavages occur. After
Figure 3.1.1: Representation of the resin used for SCX. Polymerized aspartamide groups with sulfoethyl chains bound to bind charged peptides (Alpert 1988).
crosslinking this can produce peptides with charge states starting at +4 and greater, if missed cleavages are included. The elution of highly charged peptides from an SCX column requires the use of higher salt concentrations with higher concentrations needed for larger peptides with higher charge states. The efficiency of this method is limited only by missed cleavages present in the peptides, however this can lead to lower chromatographic efficiency as well as peptide fragmentation causing a possible loss in peptide identification and structural information for that protein (Link et al. 1999).

Before performing an online approach an offline method had to be developed for proof of concept. A 150 X 1.0 mm, 5 micron, 300 angstrom PolySULFOETHYL A column (polyLC) was used with an Agilent HPLC system for separation. Mobile phase A was a 20 mM sodium phosphate buffer pH 2.5, mobile phase B was 20 mM sodium phosphate pH 2.5, with 1 M KCl. To obtain initial separation conditions a sample of BSA was loaded that was digested with trypsin but still had all disulfides intact to simulate a crosslinked sample. The gradient method ran 0% to 100% B in 23 min at 0.5 mL/min and stayed at that condition till 39 min and was washed with 100% solvent A for 10 min as shown in Figure 3.1.2. The peaks eluting from 0 to 10 min could represent a number of things such as non-peptide contaminants that cannot bind to the resin in the column. The most likely conclusion though is that those peaks are peptides that could not bind either due to overloading of the column or peptides that could not bind because the flow rate was too fast, thus not allowing adequate time to interact with the binding sites on the resin. The two peaks
Figure 3.1.2: This chromatogram shows loading of a BSA digest with disulfide bonds still intact to simulate a crosslinked sample.
proceeding at ~19 and ~26 min are the lower charged state peptides (i.e. 2+, 3+) showing the depletion required to select for higher charge states later in the elution. The final peak eluting at ~34 min contains the higher charge states (i.e. 3+, 4+, 5+, and greater). This peak is broad showing characteristics of low separation efficiency but this could be due to very high charge state peptides that take longer to elute even at a high concentration of a strong salt. The important thing is that this peak can be resolved from the others indicating that a large portion of this fraction probably contains highly charged peptides. This initial run allowed for preliminary conditions for offline SCX analysis to be obtained. To evaluate the offline SCX method on a chemically crosslinked sample, samples containing BSA protein (2 mg) were reacted with 100, 200 and 500 nmoles of BS3, reduced/alkylated to remove disulfide bonds and digested with trypsin. The first two chromatograms, in Figure 3.1.3 show strong similarities to the uncrosslinked, non-reduced BSA sample. Three characteristic peaks elute close to the same times as the first BSA sample. Similar peaks can be better visualized in Figure 3.1.4. Further proof that the last peak is comprised of the higher charge state peptides is the increased peak height at ~34 minutes with the first two peaks showing little to no increase in the 200 nmol sample. The peak shapes for the higher charge states in the first two profiles do not have the same peak broadness indicating that higher chromatographic resolution is probably due to better digestion after the disulfide bonds were broken to provide trypsin more access to the Arg/Lys residues required for digestion. The chromatogram for the 500-nmol BS3 sample shows a large increase in the final peak but it also shows an
Figure 3.1.3: BSA crosslinked with BS3 and digested with trypsin; A: 100 nmol BS3/2 mg BSA, B: 200 nmol BS3/2 mg BSA, C: 500 nmol BS3/2 mg BSA.
Figure 3.1.4: An overlay of the chromatograms of the BSA crosslinked samples showing similar elution profiles with the use of 100, 200 and 500 nmoles of BS3.
increase in the intensity of the earlier peaks, peak distortion, and a substantial loss in resolution. This phenomenon could be due to over crosslinking of the sample leading not only to a larger population of higher charge state peptides but also a decreased digestion efficiency resulting in charge states much higher than 4+, thus requiring a longer time for the complete elution of all peptides in order to reduce peak broadening. Overlapping of the three chromatograms shows the strong similarities between the 100 and 200 nmol crosslinked samples and the differences they share with the 500-nmol sample. These data supports the possibility of separating crosslinked peptides from the rest of the population of digestion products using SCX along with the consequences of crosslinking conditions that result in over crosslinking of proteins.

Having demonstrated the use of offline SCX to separate charged peptide species such as interpeptide crosslinks, there are still some analytical issues that remain when considering the analysis of low abundant crosslinked peptides. One of the issues is sample loss due to over handling. To obtain the results seen in Figure 3.1.4, which utilized BSA, larger samples of purified BRI1/BAK1 must be used than can reliably be obtained using my currently developed protein purification protocols. The BRI1/BAK1 studies require a minimum of 100 micrograms (1 nmol) of protein per crosslinking reaction, which is 200x less than the BSA used for offline SCX. This offline method requires numerous sample handling steps as well as performing the crosslinking reaction along with a digestion protocol. Samples must then be dried down and resuspended in an appropriate volume for HPLC and the appropriate
elutions collected. Collected samples must then be dried down again and filtered using a desalting column to remove the large amount of salt that will be present and interfere with the electrospray ionization process of the peptides during mass spectrometry. The amount of steps required creates more possibilities for sample loss and decreases chance of detecting crosslinked peptides.

To get past these analytical issues, an online SCX peptide separation method was developed shown in Figure 3.1.5 using an in-house packed SCX trap column. The SCX trap column was packed with PolySulfoethyl A resin, a strong cation exchange resin, which was coupled to a reversed-phase Acclaim PepMap trap column. The basic setup for the modified front end of the mass spectrometer is shown in the Figure 3.1.6. The SCX column can be seen wrapped in green tape, used for support, preceding the trap column. Below that, the fused silica used to house the SCX resin is held in place by fittings that force the flow of solvent through the fused silica leading to the trap column. Only ~1 centimeter of the SCX column itself can be seen with the rest being held inside of the fittings.

Initial conditions had to be established so an E. coli lysate was digested with Lys-C to produce peptides with higher charge states due to cleavage only at lysine residues, unlike trypsin, which cleaves at arginine as well as lysine. This allowed the lysate to mimic a protein sample that had been previously crosslinked before digestion. The initial E. coli lysate digest was dried down and brought back up in 2% acetonitrile (CAN)/0.1% formic acid (FA) pH ~3 and analyzed using the reversed-phase LC method as stated in section 2.4. After the initial sample loading, sequential
**Figure 3.1.5:** Typical flowthrough for an SCX experiment.
Figure 3.1.6: In-house packed SCX trapping column attached on-line to the front end of the mass spectrometer. The SCX column is shown circled in red followed by the RP trapping column in orange. Below the mass spec online set up is a close up of the SCX trapping column.
salt fractions ranging from 0, 25, 35, 50, 60, 75, 100, 200 and 500 mM ammonium formate, pH~3, were loaded on the reversed-phase trap column then separated using the same reversed-phase LC method used with the *E. coli* lysate.

To establish a reliable online SCX method compared to its offline version, a few issues had to be addressed as part of its development. The caveat to the developed online SCX approach is the loss of precise control of elution of the higher charged peptides that can be done with an offline method. Predetermined salt fractions of different concentrations have to be used that will inevitably elute off lower and higher charged state peptides at the same time. To account for this the salt fractions allow for different peptides to elute off of the SCX trap column and flow over the trapping column. Flow rates and salt fractions volumes had to be optimized to allow the salt to interact long enough to slowly elute off the peptides. After the salt cut is loaded and washed through the LC separation method returns to normal using CAN/0.1% FA and water/0.1% FA as base solvents to separate the peptides based on hydrophobicity giving two dimensions of separation for, ideally, better separation and identification.

Due to the online SCX method being coupled to the LC/MS/MS system, additional issues had to be addressed that are not a concern with offline SCX. The smaller diameter of the SCX column in comparison to the trap column on the LC/MS/MS system caused back pressure/flow problems that had to be compensated for by lowering the flow rate. An unforeseen problem arose with the high salt strength of the KCl that was originally used. While flowing through the waste line
leaks were caused possibly due to crystallization of the salt and partial destruction of stators sued for directing flow through different portions of the LC, leading to replacement of those parts. To get past this a volatile salt, ammonium formate, was used. Along with the lower salt strength this particular salt has the added benefit of acting as a buffer with one of its two pKa values being around pH~3 which is the pH needed for salt fractions loaded to keep the high charge states present on the peptides. The final step-wise SCX method using ammonium formate allowed for a gradual elution of lower charge state peptides that dominates the peptide population and enriched for the higher charge state peptides in later elutions, thus lowering the charge competition that is usually observed and leads to lower detection of highly charged peptides.

Using the developed online SCX LC/MS/MS method on BSA crosslinked samples produced an increase in interpeptide crosslink detection. Overall, a depletion of the 2+ and 3+ charge states was seen as they accounted for less of the peptides selected for MS2 experiments as the salt fractionations proceeded. A gradual elution of 4+ and higher was seen which is ideal so as to allow the highest chance of detecting and triggering MS2 events on as many crosslinked peptides as possible, as shown in Figure 3.1.7 and 3.1.8. A large number of peptides spanning across all charge states were seen with the last 500 mM salt fraction. This is simply due to completely cleaning off the SCX trap column to prevent any possible contamination from one sample to the next. To create the largest yield of information per MS2 event high-energy collision induced dissociation (HCD) was used. The
Figure 3.1.7: A standard BSA crosslinking SCX run using HCD. Using the ability to trigger only on 4+ or higher charge states gives the highest probability of observing a crosslinked peptide.
Figure 3.1.8: An *E. coli* standard digested with Lys C to cause it to imitate a crosslinked sample illustrates the problems with the online method. The last three salt fractions could not be obtained due to destruction of the trap column after the first 6 runs. The method works but the hardware is not sturdy enough to withstand the harsh conditions.
advantage of HCD over a typical CID event is the lack of the low molecular mass cut off. As the mass of a peptide increases the number of collisions increases to reach the energy needed to break peptide backbone amide bonds. The ions spend a longer period of time in the collision cell acquiring more collisions leading to a larger amount of energy stored in the bonds and an increased chance of dissociation.

This is important in a crosslinking experiment due to the larger mass range that a crosslinked peptide can inhabit. If a typical CID experiment was used, the risk of incomplete dissociation can occur, thus preventing identification of the peptides. Based on the data obtained and visualized in Figure 3.1.7 an online SCX LC/MS/MS approach is the optimum method to use in crosslink identification. However, a serious problem arises with the trap column when using the low pH/high salt fractions to elute off the peptides. The trap column uses a porous polymeric frit to hold the resin inside of the column. As more concentrated salt fractions are used, the frit loses its structural integrity and begins leaking into the separation column. This can be seen as regularly spaced low intensity peaks that contaminate the mass spectrum that is being obtained. Besides simple contamination the polymers being leaked into the solvent can compete with the peptides of interest for charge. This charge competition inevitably leads to a lower population of peptides entering the gas phase as charged ions and lower intensity of resulting ions giving useful data. This could possibly be worked around with using an in-house packed column reversed-phase trap to replace the trapping column from Thermo. A resin with equal binding capacity would be needed but the issue of bead stability at low pH and high
salt concentration still remains. With this method being possibly cheaper than replacing expensive trap columns after every run, it is an avenue worth exploring to further develop the SCX LC/MS/MS method for 2D separation for enhanced detection of crosslinked peptides.

3.2 BRI1/BAK1 Kinase Domain Homodimer Structural Analysis using XL-MS

Crosslinking of the cytoplasmic domains of BRI1 and BAK1 posed a unique challenge as each protein had very different stability in solution, which led to problems keeping the complex in solution. The fully active Flag tagged BRI1 kinase had no issues with stability in a HEPES buffer containing 150 mM NaCl. Flag-BRI1 precipitated out of solution at high protein concentrations, but a large amount of protein still remained in solution and could be purified. The fully active GST-tagged BAK1 posed the biggest problem in regards to expression and purification. This particular tagged version did not express well at higher temperatures (22-37°C) including room temperature. The highly active protein began to degrade 1-2 hr after expression started and the tag was cleaved off even faster. This low efficiency could be due in part to the high activity of the kinase domain in combination with the GST tag. The highly phosphorylated state of the protein could lead to *E. coli* signaling for the protein to be degraded. It could also be due to do the kinase phosphorylating other proteins indiscriminately before they have finished folding leading to possible aggregation with other proteins and falling out of solution. To get past this problem a
refrigerated incubator was employed and GST-BAK1 was expressed at 16°C over night. This lead to less protein precipitation and a lower population of cleaved tags.

Though the problem with expression was solved, the fully active GST-BAK1 fusion protein could not be purified out of solution in large enough amounts to allow for a complete set of crosslinking experiments to be performed. After incubation with glutathione beads, the majority of the protein remained present in the cleared lysate as determined by an immunoblot. Purification of a GST-tagged protein requires an active GST protein that is able to bind to the glutathione moiety present on the resin. It was thought that the BAK1 may be hyperphosphorylating the GST tag that disturbs the structure and prevents it from binding to the glutathione. This can be seen on the immunoblot as the BAK1 GST runs as a doublet. This same phenomenon was seen after replacing the Flag tag on the BRI1 recombinant protein with the GST moiety shown in Figure 3.2.1. The immunoblot shows the elution profile and flow through of both BRI1 and BAK1 tagged with GST. The enclosed box contains both full-length proteins, and a doublet can be seen running slightly below each band at 70 kDa. The addition of phosphates could cause such a dramatic shift in band location as each phosphate adds ~80 Da to the protein. If a large number of Ser/Thr/Tyr residues are phosphorylated on both the protein and GST-tag, a large shift in molecular weight will be observed. Both 1 mL samples were incubated overnight with 100 µL of glutathione resin that has the capability of binding 0.5-1 mg of protein. The flow through shows a large portion of the protein remains in the lysate, when incubating with glutathione resin again there was no further purification of protein.
**Figure 3.2.1:** Immunoblot comparing BR1 and BAK1 both tagged with GST. Similar issues of expression and purification are observed in both when using the fully active kinase. Yellow box enclosing BR1 and BAK1 with arrows indicating bands of interest. E; elution, L; cleared lysate. The molecular weight markers used during separation were added manually.
This suggests that the residual protein cannot be purified due to interacting with each other to form a soluble aggregate or the large number of phosphate modifications limiting binding to the column. Since the concentration of protein was not high enough for a crosslinking experiment, a Centricon filter was employed to try to reduce the volume of the sample.

The filter resulted in further precipitation or protein binding to the membrane and further loss of GST-BAK1. To get past this problem a mutated GST-BAK1 recombinant protein was used in place of the fully active form. This protein contains a single point mutation, K317D, in BAK1 that eliminates all kinase activity, due to the mutation being present in the active portion required for ATP binding. There should be little to no structural disturbance introduced since the lysine residue is not involved in a structure stabilizing salt bridge leading to similar structure that the wild type would produce although the lack of phosphorylation may significantly lower binding affinity. Figure 3.2.2 shows the dramatic increase in protein production amount and quality of the mutant protein after purification. Few contaminants are present including the free GST tag with GST-mBAK1 being the major component present. The doublet that was observed with the fully active kinase is not observed with the mutated protein, thus providing further support for the theory that hyperphosphorylation is the reason behind the dual bands and lack of purification of GST-BAK1.

Crosslinking of the Flag-BRI1/GST-mBAK1 complex required condition optimization using a standard crosslinker BS3 as shown in Figure 3.2.3. This
Figure 3.2.2: Coomassie stained gel of mBAK1. Dramatic increase in purity and amount pulled out of solution. A large portion of protein is still present in the flow through possibly due to over loading of resin or inaccessible recombinant protein. Arrows indicate BAK1. F.T.; flow through. The molecular weight markers used during separation were added manually.
Figure 3.2.3: Optimization of crosslinking using BS3 as the standard cross-linker. Lanes represent 1:100, 1:200, and 1:300 protein:cross-linker ratio respectively. Lane E3 shows all protein crosslinked in a complex, with possible over crosslinking occurring. E; elution. The molecular weight markers used during separation were added manually.
particular linker would provide a good representation of the reactions that SuDP would undergo as the two are very similar in structure and linker region length. To begin, three nmoles of both Flag-BRI1 and GST-mBAK1 was mixed and dialyzed to remove all Tris and other contaminants with free amines from the purification that could prematurely quench the reaction. This also allowed the introduction of the Mg\(^{2+}\) required for binding and stable formation of the heterodimer. The combined proteins were aliquot into three microfuge tubes to produce one nmole for each protein and 100, 200 or 300 nanomoles of BS3 were dissolved in 5 µl of DMSO and added to each respective sample and allowed to react. This experiment determined the optimum amount of linker to use to produce intra-protein crosslinks and to create a large enough population of inter-protein crosslinks so that detection would be possible on the mass spectrometer. This step also allows for visualization of possible over crosslinking that could occur if there were too much BS3 present. Over crosslinking events could lead to protein structure disruption and a lower efficiency of digestion low protein coverage. The immunoblot in Figure 3.2.3. shows the 100, 200 and 300 nmole reactions. The vial containing 300 nanomole of crosslinker shows almost all of the Flag-BRI1 protein involved in a crosslinked reaction that was removed from the pool of samples to be analyzed due to possible over crosslinking that may have occurred. The other two samples show populations of Flag-BRI1 uncrosslinked, which is ideal to insure the sample is not over crosslinked. The characteristic smear of crosslinking could be due to differing numbers of crosslinks present in each oligomer or because of the presence of contaminating proteins. The
possibility of contaminants required running a gel to help separate the proteins of interest and to separate the crosslinked from the uncrosslinked proteins. This experiment originally used for optimization for crosslinking experiments yielded useful crosslinking information. Due to structural differences in the linkers, BS3 could provide unique information, as it is more hydrophobic than the SuDP linker and contains negatively charged sulfosuccinimidyl ester moieties. Optimization of crosslinking showed that 200 nmol of crosslinker was sufficient for the reaction but to account for possible differences in crosslinker reactivity 100 and 200 nmol of SuDP were used. The immunoblot in Figure 3.2.4 better shows the bands in common between an anti-GST and anti-Flag membrane visualizing portions of the gel that contain Flag-BRI1 and GST-BAK1 as a heterodimer. The bands line up nicely with the predicted molecular mass of a dimer, ~100 kDa, and higher order oligomers. Whether the higher molecular weight species represent a tetramer of Flag-BRI1/GST-mBAK1 or is an artifact due to dimerization of the GST tag attached to the mBAK1 recombinant protein needs to be investigated further. Bands not accounted for by similar bands present in both immunoblots were still excised due to the GST-secondary antibody not luminescing as brightly as the anti-Flag antibody.

After analysis of the data using MeroX and StavroX, cleavable and non-cleavable linker programs respectively, a total of 32 crosslinks were determined with 12 from BS3 and 20 from SuDP. A subset of crosslinks was used to determine the binding regions of the proteins. The subset was determined by selecting the crosslinks that made the most physical sense with the predicted protein interface. To
The bands in the red box are those in common between the anti-Flag (left) and anti-GST (right) western blots corresponding to areas of the gel with BR11/BAK1 crosslinked in complex. The two lanes represent 1:100, 1:200 protein:crosslinker ratio respectively. E; elution. The molecular weight markers used during separation were added manually.
further validate whether or not these faces could bind the crosslinks were used as constraints and entered into a docking program called HADDOCK (Dominguex et al. 2003). This program follows the laws of physics only allowing regions of the protein to come together until they begin to clash due to Van Der Waal radii. The results show binding is favorable using the faces and orientation chosen for each protein.

The problem with this model is that it is a static view of the proteins being pushed together. A second run was performed to give the entire protein full flexibility in order to see if a better fit could be found. The results showed a small fluctuation in certain regions, but overall the model looks similar. To better visualize the flexibility present in the protein complex another subset of crosslinks was added to the original set. This subset was determined by choosing crosslinks that extended further than the maximum distance of the linker causing the protein to flex further. What can be seen is a large fluctuation in structure across different regions. Many loop regions show the ability to span large distances and even the ability for the proteins themselves to flex like a hinge. This type of movement would fit with the predicted structures as the C-terminal lobe of Flag-BRI1, which is thought to act as a binding domain for other proteins. If GST-mBAK1 were bound to that portion, flexibility would be critical to its ability to phosphorylate and fully activate the Flag-BRI1 kinase. This could be possible as the binding of BL shows no structural change, besides the island domain, in the two LRR regions so physically moving the kinase domains may be the mechanism used to further the signal progression. The single pass transmembrane helix could be allowing a large shift to occur after BAK1 LRR binds
to BRI1 LRR with the help of BL to bring the two kinase domains into the correct spatial orientation to allow for further phosphorylation events to occur. The models created by the HADDOCK program can be seen in Figures 3.2.5 and 3.2.6. As they are static models of the best fit of each set of crosslinks they cannot fully capture the complexity of the system. When looking at the distances obtained by fitting the proteins the lysine-to-lysine distance may be larger than the spacer region of the linker allows but this can be attributed to the maximum flexibility of each protein not being fully realized by the program. Alternatively, there may be a conformation that the proteins must go through to become fully active and allow for full flexibility of the loops and the program is trying to fit the entire sequence into one step. Mapping of the electrostatic and hydrophobic surfaces was also performed, as shown in Figures 3.2.7 and 3.2.8 respectively to give an idea as to how they may orient with a lipid bilayer.
Table 3.2.1: Crosslinks identified using BS3. Yellow highlighted rows are part of the first subset of crosslinks used for HADDOCK modeling. Blue highlights are part of the second set off crosslinks used for modeling. K*; lysine with modified with the BS3 crosslinker.

<table>
<thead>
<tr>
<th>Protein 1</th>
<th>Peptide(1)</th>
<th>Residue (1)</th>
<th>Protein 2</th>
<th>Peptide(2)</th>
<th>Residue (2)</th>
<th>Scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAK1</td>
<td>LMDYK*</td>
<td>444</td>
<td>BRI1</td>
<td>DGSAVAIK*KLIHVSGQGDR</td>
<td>911</td>
<td>5784</td>
</tr>
<tr>
<td>BAK1</td>
<td>PK*mSEVV R</td>
<td>555</td>
<td>BRI1</td>
<td>K*LIHVSGQGDR EFmAEmETIGKIK HR</td>
<td>912</td>
<td>5865</td>
</tr>
<tr>
<td>BAK1</td>
<td>GLLK*EKK</td>
<td>514</td>
<td>BRI1</td>
<td>AGVK*LNWSTR</td>
<td>977</td>
<td>4106</td>
</tr>
<tr>
<td>BAK1</td>
<td>GLLKEK*</td>
<td>516</td>
<td>BRI1</td>
<td>DmK*SSNVLLDE NLEARVSDFGMA R</td>
<td>1011</td>
<td>3980-3982</td>
</tr>
<tr>
<td>BAK1</td>
<td>DTHVTTAV RGTIGHIA PEYLSTGK *</td>
<td>468</td>
<td>BRI1</td>
<td>NLVPLLGYBK*</td>
<td>947</td>
<td>6068</td>
</tr>
<tr>
<td>BAK1</td>
<td>GRLADGTV LAVK*R</td>
<td>317</td>
<td>BRI1</td>
<td>PTDSPDFGDNNL VGWVK*</td>
<td>1101</td>
<td>5010-5014</td>
</tr>
<tr>
<td>BAK1</td>
<td>NILGRGGFK*</td>
<td>302</td>
<td>BRI1</td>
<td>QHAK*LRISDVFD PELmK</td>
<td>1105</td>
<td>4936</td>
</tr>
<tr>
<td>BAK1</td>
<td>PK*mSEVV R</td>
<td>555</td>
<td>BRI1</td>
<td>K*LIHVSGQGDR EFmAEmETIGKIK HR</td>
<td>912</td>
<td>5763</td>
</tr>
<tr>
<td>BAK1</td>
<td>GLLK*</td>
<td>514</td>
<td>BRI1</td>
<td>K*LIHVSGQGDR EFmAEmETIGKIK HR</td>
<td>912</td>
<td>5287</td>
</tr>
<tr>
<td>BAK1</td>
<td>LADGTLVA VK*R</td>
<td>317</td>
<td>BRI1</td>
<td>SIEDGGFSTIEMV DmSIK*</td>
<td>1189</td>
<td>5100-5102</td>
</tr>
<tr>
<td>BAK1</td>
<td>K*LIHVSGQ G QGDR</td>
<td>516</td>
<td>BRI1</td>
<td>GLLKEK*KLEALV DVLQGNyK</td>
<td>912</td>
<td>5660</td>
</tr>
<tr>
<td>BAK1</td>
<td>LK*</td>
<td>320</td>
<td>BRI1</td>
<td>PTMVQVMAMFK* EIQAĜsGIDSQST IR</td>
<td>1156</td>
<td>4316-4320</td>
</tr>
</tbody>
</table>
Table 3.2.2: Crosslinks identified with SuDP. Yellow highlight indicate first subset of crosslinks used for modeling. Blue highlight indicate second subset of crosslinks. Mod., modification; Res., residue

<table>
<thead>
<tr>
<th></th>
<th>Peptide 1</th>
<th>Mod. 1</th>
<th>Res. 1</th>
<th>Peptide 2</th>
<th>Mod. 2</th>
<th>Res. 2</th>
<th>Scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAK 1</td>
<td>ERPESQPPLDWPK*R</td>
<td>SuD</td>
<td>389</td>
<td>QHAK*</td>
<td>P</td>
<td>1105</td>
<td>3934</td>
</tr>
<tr>
<td>BAK 1</td>
<td>AFDLARLANDDDVMLLDWVK*GLLK</td>
<td>SuD</td>
<td>510</td>
<td>AILKDGsAVAIK*</td>
<td>P</td>
<td>911</td>
<td>6026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6028</td>
</tr>
<tr>
<td>BAK 1</td>
<td>EK*K</td>
<td>P</td>
<td>516</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAK 1</td>
<td>SSEK*</td>
<td>P</td>
<td>472</td>
<td></td>
<td></td>
<td></td>
<td>5908</td>
</tr>
<tr>
<td>BAK 1</td>
<td>IIHRDVK*</td>
<td>P</td>
<td>418</td>
<td></td>
<td></td>
<td></td>
<td>5303</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5307</td>
</tr>
<tr>
<td>BAK 1</td>
<td>LMDYK*</td>
<td>P</td>
<td>444</td>
<td></td>
<td></td>
<td></td>
<td>933</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4323</td>
</tr>
<tr>
<td>BAK 1</td>
<td>VYK*GR</td>
<td>P</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td>1132</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5251</td>
</tr>
<tr>
<td>BAK 1</td>
<td>AFDLARLANDDDVMLLDWVK*GLLK</td>
<td>SuD</td>
<td>510</td>
<td></td>
<td></td>
<td></td>
<td>911</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5476</td>
</tr>
<tr>
<td>BAK 1</td>
<td>GRLADGTLVAVK*R</td>
<td>SuD</td>
<td>317</td>
<td></td>
<td></td>
<td></td>
<td>1156</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4850</td>
</tr>
<tr>
<td>BAK 1</td>
<td>LANDDDVmLDWVK*GLLK</td>
<td>P</td>
<td>510</td>
<td></td>
<td></td>
<td></td>
<td>1196</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3327</td>
</tr>
<tr>
<td>BAK</td>
<td>ELQVASDNF SNK*NILGR</td>
<td>P</td>
<td>292</td>
<td>BRI1</td>
<td>AILK*</td>
<td>SuD</td>
<td>903</td>
</tr>
<tr>
<td>------</td>
<td>----------------------</td>
<td>----</td>
<td>-----</td>
<td>------</td>
<td>-------</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>BRI1</td>
<td>QHAK*</td>
<td>SuD</td>
<td>1105</td>
<td>BRI1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTDSPDFGD NNLVK*QHAK LRIsDVFDPEL mK</td>
<td>P</td>
<td>1105</td>
<td>4825</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAK</td>
<td>FsLRELQVA SDNFSNK*NILGR</td>
<td>P</td>
<td>292</td>
<td>BRI1</td>
<td>EVPEGK*L</td>
<td>SuD</td>
<td>1196</td>
</tr>
<tr>
<td>BAK</td>
<td>VYK*GRLAD GTLVAVK</td>
<td>P</td>
<td>305</td>
<td>BRI1</td>
<td>PTDSPDFGD NNLGVWVK*</td>
<td>SuD</td>
<td>1101</td>
</tr>
<tr>
<td>BRI1</td>
<td>RK*</td>
<td>P</td>
<td>985</td>
<td>BRI1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KLTFADLLQA TNGFHNDSLI GsGGFGVDVY K*AILK</td>
<td>SuD</td>
<td>899</td>
<td>5625</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAK</td>
<td>PK*MsEVVR MLEGDGLAE R</td>
<td>SuD</td>
<td>555</td>
<td>BRI1</td>
<td>DGSAVAIK*K</td>
<td>P</td>
<td>911</td>
</tr>
<tr>
<td>BAK</td>
<td>BstK*</td>
<td>P</td>
<td>1066</td>
<td>BRI1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAK</td>
<td>LMDYK*</td>
<td>P</td>
<td>444</td>
<td>BRI1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LMSAMDTHL SVSTLAGTPG YVPPEYYQS FRBSTK*</td>
<td>SuD</td>
<td>1066</td>
<td>4645</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAK</td>
<td>IIHRDVK*</td>
<td>P</td>
<td>418</td>
<td>BRI1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAK</td>
<td>GRLADGTLV AVK*R</td>
<td>SuD</td>
<td>317</td>
<td>BRI1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DMK*SSNVLL DENLEARV</td>
<td>P</td>
<td>1011</td>
<td></td>
<td>4617</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>4622</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.2.5: HADDOCK docking simulation using the first subset of cross-linkers highlighted in yellow in table 3.1. BAK1, red; BRI1, tan; lysines, orange; activation loops, blue. The images were generated using PYMOL 1.7.4 (www.pymol.org) and PDB 3ULZ and 4OA6.
Figure 3.2.6: HADDOCK simulation using the second subset of cross-links highlighted in blue in tables 3.2.1 and 3.2.2 BAK1, red; BR11, tan; lysines, orange; activation loops, blue. The images were generated using PYMOL 1.7.4 (www.pymol.org) and PDB 3ULZ and 4OA6.
Figure 3.2.7: Modeling of the surface electrostatics of the BRI1/BAK1 complex. A shows a denser collection of basic charge, possibly showing a preference to orient towards the lipid interface. Left side, BAK1; Right side, BRI1; red, negatively charged; blue, positively charged. These images were generated using PYMOL 1.7.4 (www.pymol.org) and PDB 3ULZ and 40A6.
Figure 3.2.8: Mapping of the hydrophobic surfaces of the BRI1/BAK1 complex. A larger portion of the hydrophobic surface seems to be concentrated on the “top” surface similar to the electrostatics showing a preference towards the lipid bilayer. Orange, hydrophobic. These images were generated using PYMOL 1.7.4 (www.pymol.org) and PDB 3ULZ and 4OA6.
Chapter 4: Conclusions

With the data obtained through the crosslinking experiments a viable structural model has been created for the BRI1/BAK1 complex that could represent the heterodimer formed, specifically when the two proteins are phosphorylating each other’s kinase domain. It was important to use two different crosslinkers to obtain a diverse population of identified crosslinks and take advantage of the improved fragmentation of the gas cleavable SuDP over BS3. Where this model falls short is the lack of a dynamic view that shows how these two domains can flex including their loop regions, which may increase their binding to one another. This may not be the only way these two proteins bind to one another, because the binding predicted by the crosslinks, blocks the active site of BRI1. The active site of BRI1 is needed after full activation to perform further phosphorylation on other protein partners to fully realize the brassinolide signal. Further work must be done to develop the model more in depth and create a dynamic view of the activity of these two proteins. This may be accomplished through molecular modeling using computer programs or using other MS structural techniques such as HDX to explore the binding faces of the BRI1/BAK1 complex and the sequential phosphorylation events that occur.
Chapter 5: Future Work

The work described here illustrates not only the advantages of using crosslinking, especially in conjunction with other methods, but also the difficulties that arise when using this method. Just as NMR can give a more biologically relevant view of the protein complex, crosslinking can achieve the same goal but it still may only be a snapshot of how the proteins interact. To move development of the model further new techniques need to be brought in to create a different view and add to the abundance of data already collected.

The use of HDX in conjunction with crosslinking has been mentioned earlier and can be used to give information about the faces involved in protein-protein interactions. This same experiment could also be used with and without crosslinking to provide insight into protein binding sites in a more flexible (uncrosslinked) and restrained manner (crosslinked). This could reveal how the proteins are interacting and whether or not they are staying bound only by the faces that house the active loops or if they are capable of “opening and closing” like a hinge on a door by binding to one another using their N- and C-terminal regions.

The work done also introduces the idea of possible higher order oligomers that could be present addition to the heterodimer. The problem that arises with this possibility is that the GST tag used on BAK1 to help keep the protein in solution also has the tendency to dimerize with itself. This could lead to crosslinking events happening between the GST dimer holding together two heterodimers and creating an artifact in solution that resembles a protein tetramer. To avoid these problems a
variety of tags must be employed to remove the possibility of unnatural higher order oligomers forming during reactions. This idea has been explored in my work and by Fan Liu, a former student of Dr. Goshe, when first beginning this research. The problem that arises is the lack of BAK1 solubility in solution after removal of the GST tag. The Flag tagged versions showed a large precipitation event immediately after cell lysis with little to no residual protein left in solution. The His-tagged version was also explored but analysis of the crosslink data revealed only homodimers containing the BAK1 protein. This tag must be explored further as it was only used to obtained preliminary data with little to no optimization. There are a number of other tags that could be used including maltose-binding protein (MBP) that plays a similar role as GST by increasing solubility and used as an affinity purification tag. This tag is 42 kDa and could cause problems with dimer formation due to size but could still be a viable option. Other peptide tags such as HA or Myc could be explored as well but the solubility of BAK1 may limit use of these as significant precipitation of the recombinant protein may occur leaving little to no protein in solution. Thus further work will require a large amount of optimization to determine not only the best tag for BAK1 but also conditions to keep the largest amount of protein in solution as crosslinking studies require a much larger amount of protein as compared to kinase activity assays using radioactive ATP which has a very high level of sensitivity.

The gold standard for this protein study would be to obtain the most biologically relevant data to this system. To do this, methods involving formaldehyde crosslinking in planta could be used to link the BRI1/BAK1 dimers together and
purifying them using affinity tags that have been added to their coding sequence. Crosslinking using the SuDP crosslinker could then be used to obtain data on the full-length protein including the extracellular, trans-membrane, and juxta-membrane and kinase domain. This could allow for the combination of all the data obtained from crystallography and previous crosslinking experiments to generate a more complete image. This method presents its own problems as there is a pH difference between the cytosolic and extracellular portion of plant cells. A more acidic environment was needed to help form the dimer between the extracellular domains of the proteins. Although this problem could be side stepped by first applying formaldehyde crosslinking, the data from such an experiment would have to be carefully analyzed so they can be compared with the data already generated from previous experiments.

The few experiments proposed show the gaps in knowledge still present in regards to the BRI1/BAK1 interaction, as well as the signaling pathway in its entirety. The large amount of molecular data present still requires a working model to become more relevant. Until the data is obtained that can fully connect the activities between all domains of the proteins and explain the data obtained from molecular studies, a working model will continue to be just out of arms reach.
Literature Cited


